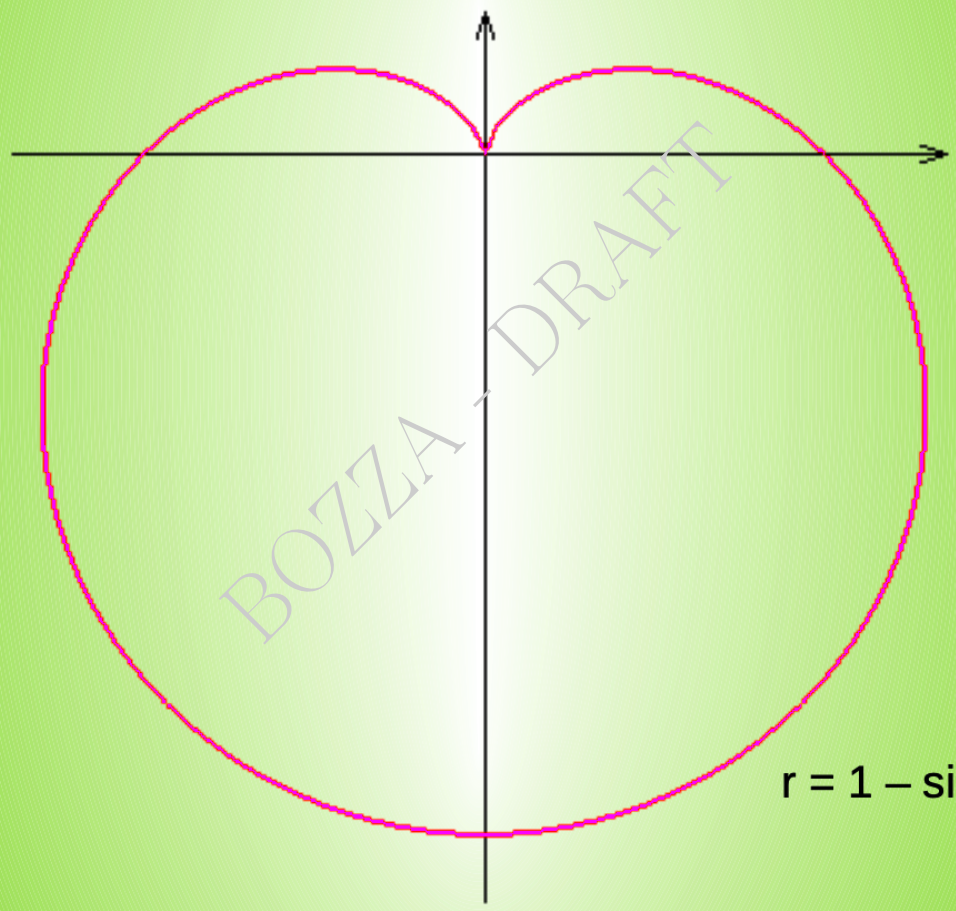
ALESSANDRO SORANZO

Le Compresse di Matematica per le Scienze della Farmacia



BOZZA - DRAFT

Alessandro Soranzo
Ricercatore universitario
Settore Scientifico Disciplinare:
MATH-03/A - Analisi matematica
Gruppo Scientifico Disciplinare:
01/MATH-03 - Analisi matematica, probabilità e statistica matematica

BOZZA - DRAFT

Le Compresse di Matematica per le Scienze della Farmacia

sive (titolo vecchio & alternativo)

Le Pillole di Matematica

per Farmacia e per Chimica e

Tecnologie Farmaceutiche

BOZZA - DRAFT

Alessandro Soranzo

Università degli Studi di Trieste

versione 69 a, GI 24 aprile 2025

"La mente non è un vaso da riempire, ma un legno da far ardere, perché si infuochi il gusto della ricerca e l'amore della verità."

Plutarco

"Non c'è niente di più pratico di una buona teoria"
(Kurt Lewin)

BOZZA - DRAFT

0 Note e Prerequisiti

La Premessa segue questa breve sezione di note.

0.1 Note tecniche sul testo

Nota 1. Questo testo fa una scelta drastica: usa *entrambi* gli standard del punto decimale e della virgola decimale, per abituare lo studente, e, volendosi limitare, considererà solo occasionalmente lo standard del punto a mezza altezza.
Inoltre, seguendo il N.I.S.T. (National Institute of Standards and Technology) statunitense, di solito userà lo spazietto come separatore delle terne di cifre.

BOZZA DRAFT

Nota 2. Attualmente la separazione della parole in fine riga, fatta automaticamente dall'editor di testo, non sempre corrisponde alla grammatica italiana.

Nota 3. SI PREGA DI AVVERTIRE L'AUTORE DI OGNI ERRORE ANCHE SE MINUSCOLO, GRAZIE!

Nota 4. In questo testo è usato il simbolo ≈ per indicare "circa" com'è previsto dallo standard ISO. Ma in futuro sempre di più si userà anche il simbolo ~ tradizionalmente usato in Farmacia.

Nota 5. Per le Lezioni fino alla 31 una revisione è stata fatta dal Dott. Lorenzo Girella. Però poi il testo ha subito nuove variazioni.

Nota 6. Il testo è organizzato
per un corso da 40 ore;
per un corso da 64 ore.
Alla fine ci sono 2 Lezioni soprannumerarie, sperabilmente gradite. Si veda l'Indice ipertestuale delle Lezioni.

Nota 7.
Legenda per gli esercizi

$*$ è richiesto il valore esatto. Può anche essere $+\infty$, $-\infty$, o una frase.

$\approx$ è richiesta una ragionevole approssimazione.

$^{\%}$ è richiesto il valore in percentuale, se serve ragionevolmente approssimato.

(R) è richiesto solo il risultato.

BOZZA - DRAFT

0.2 Valori numerici fondamentali

Alla fine di questo corso ci si aspetta che lo studente conosca a memoria questi numeri:

$\frac{1}{3} = 0.\overline{3} = 0.33333333333...$ (periodico; valore esatto) $\approx 0.333 \approx 0.33$

$\phi = \frac{1+\sqrt{5}}{2} \approx 1.618$ (sezione aurea)

$\pi \approx 3.14$ (pi greco)

$e \approx 2.718$ (Numero di Nepero o di Eulero)

$\lg 2 \approx 0.3$ (logaritmo decimale di 2, cioè $\log_{10} 2$)

$\lg e \approx 0.4343$ (logaritmo decimale di e, cioè $\log_{10} e$)

$\phi_{0.975} \approx 1.96$ (un quantile normale)

≈ 19.68 un quantile del chi quadrato a 11 gradi di libertà delle tavole del chi quadrato, variamente denotato dai vari Autori.

BOZZA - DRAFT

0.3 Prerequisiti

• Geometria elementare minima, dal punto alla sfera. In questo testo c'è comunque un ripasso su quegli argomenti: Lezione 2.

• I numeri primi: 2 (sì anche 2, ma non 1), 3, 5, 7, 11, 13, 17...

• L'algebra dei numeri e con le variabili, per esempio $\frac{15}{20} = \frac{3}{4} = 0{,}75$ e $\frac{x^2+1}{(x^2+1)^3} = \frac{1}{(x^2+1)^2}$. C'è comunque un ampio ripasso: Lezioni 7, 8.

• Le unità di misura, ad un livello basico.

• Semplici operazioni con ore, giorni, settimane, mesi, anni, date. E in particolare:

◦ per esempio 3,5 h (cioè 3 ore e mezza) dopo le 23 sono le 2:30 (di notte).

◦ 1 d = 24 h cioè 1 giorno ha 24 ore, da 0 a 24 o 2 volte da 0 a 12

◦ 1 h = 60' cioè 1 ora ha 60 minuti

◦ 1' = 60" cioè 1 minuto ha 60 secondi

◦ un mese (standard) ha 30 giorni; i mesi da gennaio a dicembre hanno certe specifiche durate

◦ 1 anno (standard) ha 365 giorni, 1 anno bisestile ha 366 giorni

BOZZA - DRAFT

Premessa

Questo è un testo su Statistica e altra Matematica. Soprattutto la Matematica utile per la Statistica. E, il tutto, con particolare riferimento alle Scienze della Farmacia.

Il testo è ipertestuale con collegamenti interni, che funzionano su computer anche offline, e collegamenti a internet.

Grande attenzione è stata posta agli errori più comuni: prevenire è meglio che curare. Esistono errori tipici, contro i quali è utile agire attivamente. Per esempio ritenere che l'arcotangente sia il reciproco della tangente, e invece è l'inversa. Questo errore consueto e altri simili sono dovuti ad ambiguità notazionali della matematica *de facto*, nello stato in cui si trova oggi scritta. Grande rilevanza è stata data a questo aspetto delle ambiguità notazionali. D'altra parte, la situazione oggi è tale, incredibilmente, che nemmeno è chiaro, quando in un testo si trova scritto

BOZZA - DRAFT

2,125

se si intenda quel numero, fra 2 e 3, "due virgola centoventicinque", oppure "duemilacentoventicinque"!
Troviamo per esempio la relazione fra la nuova e la vecchia unità di misura del campo magnetico

1 T = 10.000 G su Wikipedia italiana Link ->
1 tesla = 10,000 gauss su Wikipedia in inglese Link ->

D'altra parte a cosa servirebbe sapere i più sottili teoremi del calcolo infinitesimale, se poi si permetterà ad Excel, software usatissimo nelle scienze applicate, di calcolare

−3^2

in modo *erroneo* – o, volendo essere benevoli, attenendosi alla convenzione tutta sua, di quel software, opposta a quella di praticamente tutta la comunità scientifica? Dice 9 invece che −9. E' meglio

avvertire di questo lo studente. (E pure che un simulatore online di Excel dice... 9, correttamente; tale è lo stato dell'arte).

Nella simbologia ci si è attenuti in generale allo standard ISO 80000-2:2009[1], per esempio tan e non tang nè tg per la funzione tangente. E uno spazietto, eventualmente, seguendo il NIST, come separatore delle migliaia: 2 125.

Ripetutamente si accenna a WolframAlpha: lo Studente dovrebbe abituarsi a utilizzare questa potentissima intelligenza artificiale disponibile gratuitamente on-line, specialistica per la Matematica (ma sa molto altro).

Una certa attenzione è stata rivolta a Wikipedia. Lo Studente dovrebbe abituarsi a consultarla. L'idea che non sia una fonte attendibile, per gli argomenti di questo testo elementare è – in generale, ovvio – senz'altro falsa. (Per questioni "sensibili", invece, alquanto male[2] si potrebbe dire di Wikipedia).

BOZZA - DRAFT

Fonti autorevoli: moltissima Matematica e Statistica Medica si trova in fonti autorevoli presenti online: PubMed, CDC, OMS... In questo testo si fanno molti riferimenti a queste fonti indicando link, familiarizzando con esse lo Studente di Farmacia.

Premesso che non tratteremo degli aspetti *materiali* della raccolta dati (questionari, interviste, eccetera), diciamo subito che la Statistica, così limitata, è una branca della Matematica, astratta come tutta la vera Matematica.
La Statistica verrà suddivisa in Statistica Descrittiva e Statistica Inferenziale.
In qualche modo possiamo dire che tutta la matematica che tratter-

[1] Si veda per esempio https://ja.wikipedia.org/wiki/ISO_80000-2

[2] Parrebbe che molte pagine siano praticamente "blindate" da gruppi variamente ideologizzati – anche su questioni di Medicina. E si veda questo Link->

emo, dai numeri alla Statistica Descrittiva, al Calcolo Infinitesimale e al Calcolo delle Probabilità, è propedeutica per la trattazione della Statistica Inferenziale.

Nelle *Matematiche della certezza*, prima Parte del corso, è compresa la Statistica Descrittiva, che costituisce il Capitolo IV, e nelle *Matematiche dell'Incertezza* è compresa la Statistica Inferenziale, Capitolo XII, che in qualche modo rappresenta l'apice della trattazione.

La seconda Parte del corso riguarda le *Matematiche dell'Incertezza*. Naturalmente, le Matematiche dell'Incertezza non fanno affermazioni incerte, bensì fanno affermazioni certe su fatti incerti, per esempio che conviene scommettere (alla pari) che un dado darà un numero primo piuttosto che un numero quadrato.

BOZZA - DRAFT

La Statistica Inferenziale è poco insegnata a scuola, e ancor meno lo era nei decenni passati.
Per uno statistico, è sempre notevole constatare quanto poco presente sia la mentalità statistica a livello di popolazione italiana. Da una parte alcuni si attaccano al caso singolo senza rendersi conto di quanto poco rappresentativo sia della situazione complessiva – in ciò magistralmente teleguidati dai media al soldo dei loro referenti – e dall'altra qualcuno ti obietta che "4 casi non fanno Statistica", credono che ci vogliano sempre migliaia di casi, e non è così, anzi in una ricerca che esporremo Fisher, "mostro sacro" della Statistica, nel suo pionieristico lavoro fece proprio una brillante esperienza di Statistica con 4 casi; dipende da vari fattori. Certo è vero che neppure migliaia di casi mal scelti bastano a fare buona Statistica, e questo è un errore sempre in agguato nella Statistica Medica e in particolare in Farmacia: fare statistiche solo sui soggetti che portano avanti un esperimento fino alle nostre misurazioni finali; come se fosse la stessa cosa che dei, diciamo, 500, che non completano il trial, siano morti 1 (che magari sarebbe la norma in quel lasso di

tempo senza prendere il nostro prodotto sperimentale) oppure 50, oppure siano tutti andati alle Seichelles a godersi la guarigione: i *drop out* sono una catastrofe nella Statistica Medica, e semplicemente non considerarli è del tutto fuorviante.

Questo testo tratta alcuni argomenti che sono anche individuati nel progetto Agenda 2030. Questo progetto, attentamente considerato, risulta un mero strumento (molto ben pitturato) dell'obiettivo delle élite sovranazionali che di fatto lo promuovono (e negli ultimi anni alquanto esplicitamente dichiarato) di ridurre di 7 miliardi le "bocche inutili" (*useless mouths*) sul pianeta, inutili (per loro) *consumatori di risorse* (in primis idrocarburi) ovvero, come si potrebbe definirli per responsabilizzarli (ma sono esattamente le stesse persone, fra cui lo scrivente e suoi Lettori) *produttori di CO2*. Nondimeno, molti singoli obiettivi dell'Agenda 2030 (che comunque l'Università italiana di fatto viene orientata a divulgare) sono molto condivisibili. In questo testo di Matematica, apprezzandoli, si farà riferimento a questi:

• il punto 1, "Povertà zero": con la trattazione della Curva di Preston ovvero del logaritmo come modello dell'aspettativa di vita in funzione del PIL pro-capite;

• il punto 3, "Salute e benessere": con grafici e statistiche di pertinenza epidemiologica, con particolare riferimento alla pandemia;

• il punto 5, "Uguaglianza di genere": con la trattazione del test statistico su quale sproporzione numerica fra maschi e femmine è statisticamente significativa per ragionevolmente escludere una fluttuazione casuale, e supporre invece l'aborto selettivo e/o infanticidio delle neonate, che è una questione statistica, non una raccolta indiziaria di "si dice", che là le donne fanno così e cosà;

• il punto 10, "ridurre le disuguaglianze": con varianza e altri indici di variabilità e outliers (e non sia mai che qualcuno pensi che quando in una commissione internazionale si discute di disuguaglianze socioeconomiche, lo si faccia confrontando i Paesi mediante sequenze video di case belle e case brutte: i confronti si fanno proprio su in-

BOZZA - DRAFT

dici di Statistica socioeconomica che misurano la variabilità, come il Coefficiente di Gini).

Questo testo vorrebbe giungere a dare un po' di *forma mentis* matematica e statistica.

Il testo è così organizzato:

PARTE A — MATEMATICHE DELLA CERTEZZA

Sezione A1 – Matematiche elementari

Sezione A2 – Calcolo infinitesimale

PARTE B — MATEMATICHE DELL'INCERTEZZA

Sezione B1 – Calcolo delle probabilità

Sezione B2 – Statistica Inferenziale

BOZZA - DRAFT

Il testo è organizzato
per un Corso da 40 ore;
per un Corso da 64 ore.
Il Corso minore prevede molta meno Statistica Inferenziale, e il Corso maggiore comprende, fra l'altro, 3 Lezioni di Matematica col Computer.
Ci sono anche 2 Lezioni soprannumerarie, sulle Serie e sulla Matematica delle Epidemie.
Naturalmente si potranno anche utilizzare solo alcune Lezioni.

Si veda l'Indice ipertestuale delle Lezioni.

Al momento, il testo della Lezione 1 dalle pagine 41 a 56 è organizzato per essere proiettato 2 pagine alla volta, sinistra e destra, 41-42 eccetera.

Il testo, al 2024, è *in fieri* / work in progress.

Indice delle 2 Parti, delle 4 Sezioni e dei 12 Capitoli

PARTE A — MATEMATICHE DELLA CERTEZZA

Sezione A1 – Matematiche elementari

I —— Numeri, Figure, Insiemi, Logica
II —— Funzioni, Algebra e Piano Cartesiano
III —— Funzioni elementari, con equazioni e disequazioni
IV —— Statistica Descrittiva

Sezione A2 – Calcolo infinitesimale

V —— Limiti e Serie Geometrica
VI — Derivate e Integrali

PARTE B — MATEMATICHE DELL'INCERTEZZA

Sezione B1 – Calcolo delle probabilità

VII —— Probabilità assiomatica ed elementare
VIII — Variabili aleatorie discrete
IX —— Variabili aleatorie continue

Sezione B2 – Statistica Inferenziale

Nota basale sulla Statistica
X —— Stimatori puntuali e intervallari
XI — Test statistici

XII — Lezioni soprannumerarie

BOZZA - DRAFT

Segue l'Indice ipertestuale delle 40 o 64 Lezioni.

I numeri della 1° colonna sono per un corso da 40 ore.
I numeri della 2° colonna sono per un corso da 64 ore.

40^h 64^h
00 00 **Note e Prerequisiti**

Indice delle 40 o 64 Lezioni

BOZZA - DRAFT

PARTE A — MATEMATICHE DELLA CERTEZZA

Sezione A1 – Matematiche elementari

I —— Numeri, Figure, Insiemi, Logica
01 01 I numeri

02 02 Ripasso di Geometria Euclidea Elementare
03 03 Logica delle proposizioni

04 04 La Logica con l'Insiemistica
05 05 Calcolo Combinatorio
II ——Funzioni, Algebra e Piano Cartesiano
06 06 Funzioni
07 07 Ripasso di Algebra – I parte
// 08 Ripasso di Algebra – II parte (*) v. nota
08 09 Piano cartesiano – I parte
09 10 Piano cartesiano – II parte

10 11 Piano cartesiano – III parte
// 12 Matematica col Computer – Lezione A

III —— Funzioni elementari, con equazioni e disequazioni
11 13 Funzioni e dis/equazioni ir/razionali
12 14 Esponenziali e logaritmi – I parte

13 15 Esponenziali e logaritmi – II parte

14 16 Esponenziali e logaritmi – III parte
15 17 Esponenziali e logaritmi – IV parte

16 18 Funzioni trigonometriche

IV —— Statistica descrittiva
17 19 Introduzione alla Statistica Descrittiva

18 20 Bar chart, istogrammi, e altri diagrammi
19 21 Quartili e box-plot

20 22 Variabilità, covarianza e correlazione
21 23 Note finali sulla Statistica Descrittiva

Sezione A2 – Calcolo infinitesimale

V —— Limiti e Serie Geometrica
22 24 Limiti di successioni e serie geometrica
// 25 Limiti e continuità per funzioni reali (*) v. nota

VI —— Derivate e integrali
23 26 Derivate – I parte

24 27 Derivate – II parte
25 28 Applicazioni del Calcolo Differenziale - I

// 29 Applicazioni del Calcolo Differenziale - II (*) v. nota
26 30 L'integrale indefinito

// 31 L'integrale definito (*) v. nota
// 32 Matematica col Computer – Lezione B

PARTE B — MATEMATICHE DELL'INCERTEZZA

Sezione B1 – Calcolo delle probabilità

VII —— Probabilità assiomatica ed elementare
27 33 Introduzione al Calcolo delle Probabilità

// 34 Concezione assiomatica della probabilità (*) v. nota

BOZZA - DRAFT

28 35 Sensibilità, specificità, predittività, ROC
// 36 Probabilità combinatoria
VIII — Variabili aleatorie discrete
29 37 Introduzione alle variabili aleatorie

30 38 Variabili aleatorie discrete
// 39 Variabili aleatorie uniformi e geometriche (*) v. nota
// 40 Leggi congiunte e indipendenza
31 41 Variabili aleatorie binomiali e bimodali

32 42 Speranza matematica e varianza
IX — Variabili aleatorie continue
33 43 Uniforme continua, di Cauchy, esponenziale

// 44 Quantili delle variabili aleatorie continue (*) v. nota 22.11
34 45 Speranza matematica, varianza, covarianza
35 46 Densità e variabile aleatoria normale
// 47 Approssimazione di $\Phi(x)$ e ϕ_α (*) v. nota

// 48 Legge dei Grandi Numeri (*) v. nota
// 49 Approssimazione Normale (*) v. nota

// 50 Note finali sul Calcolo delle Probabilità (*) v. nota 25.11
Sezione B2 – Statistica inferenziale
X — Stimatori puntuali e intervallari
36 51 Introduzione alla Statistica Inferenziale
// 52 Stimatori e stimatori non distorti (*) v. nota
37 53 Intervalli di fiducia e caso di μ di $N(\mu, \sigma^2)$

// 54 Intervalli di fiducia per la varianza 28.11
XI — Test statistici
38 55 I Test Statistici
// 56 Errori di I e II Specie (*)
39 57 Test del χ^2 di indipendenza

// 58 Il Test del χ^2, quello basico 6.12

BOZZA - DRAFT

// 59 Il test dei ranghi
40 60 Alcuni Test di Student

// 61 Test di Student per il confronto di medie (*) v. nota
// 62 Matematica col Computer – Lezione C
// 63 Matematica delle Epidemie

// 64 Note finali sulla Statistica 12.12

—— Lezione soprannumeraria
// // Le serie numeriche

(*) Per il Corso da 40 ore:
riassumere parzialmente questa Lezione nelle Lezioni precedente

Segue l'Indice ipertestuale delle Lezioni e dei Paragrafi.

CALENDARIO

ORE 39-40 – LU 9 DIC ULTIMA LEZIONE
– GI 12 DIC RICEVIMENTO STUDENTI supplementare
Sì video. No registrazione presenze

BOZZA - DRAFT

Indice delle Lezioni e dei Paragrafi

Contents

0 Note e Prerequisiti **5**
0.1 Note tecniche sul testo 5
0.2 Valori numerici fondamentali 7
0.3 Prerequisiti 8

1 I numeri **43**
1.1 Perché numeri, se ci interessano farmaci? 43
1.2 Alcuni usi dei numeri in Farmacia 44
1.3 La sconcertante questione della virgola e del punto . 47
1.4 Il punto a mezza altezza 51
1.5 Prima definizione dei numeri: scritture decimali . . . 53
1.6 Seconda definizione dei numeri: punti di una retta . 55
1.7 Terza definizione dei numeri: assiomi e costruzione . 55
1.8 Numeri e loro scrittura 57
1.9 Gli insiemi numerici 58
1.10 Note sugli errori medici e farmaceutici 59
1.11 Errori di calcolo negli articoli scientifici 60
1.12 Numeri romani 61
1.13 Notazione scientifica dei numeri 63
1.14 Varia 64
1.15 Quando pochi decimali uccisero i "buoni" 65
1.16 I decimali sono gratuiti – usiamoli 66
1.17 Cifre significative 67
1.18 Conteggio delle cifre significative 68
1.19 Scrittura percentuale dei numeri 69
1.20 Scrittura reciproca dei numeri 71
1.21 Note sulle percentuali prossime al 100% 73
1.22 Usare numeri approssimati per rendere l'idea 73
1.23 Falsi amici, e pure confusionari 74
1.24 Spirito critico, già coi semplici numeri 76

BOZZA - DRAFT

1.25 Altro 76
1.26 Bottom line – prima dei Complementi 77
1.27 Complementi – Simboli simili 78
1.28 Complementi – Scrittura a mano 80
1.29 Complementi – 7 americano o 1 italiano? 82
1.30 Complementi – Note finali sui numeri 82
1.31 Complementi – Il fantasmagorico microgrammo . . . 84
1.32 Complementi – Numeri e approssimazioni 85
1.33 ESERCIZI SULLA LEZIONE 1 87
1.33.1 Esercizio risolto a – Globuli rossi 87
1.33.2 Esercizio risolto b – Percentuali capricciose . 88
1.33.3 Aneddoto – Il significato ultimo 88

2 Ripasso di Geometria Euclidea Elementare 89
2.1 Perché questo corso ha 1 ora di Geometria? 89
2.2 Questioni pratiche affrontate dalla Geometria 90
2.3 Geometria, ovunque! 91
2.4 Una figura può valere mille parole 93
2.5 Una figura può valere mille bugie 94
2.6 Primi concetti della Geometria Euclidea Elementare 95
2.7 Altri concetti della Geometria Euclidea Elementare . 97
2.8 Bordo ovvero contorno ovvero frontiera 98
2.9 Corde e convessi 99
2.10 Note di Geometria Euclidea 1- 2- 3- dimensionale . . 100
2.11 Decomposizione di una figura piana in più figure . . 101
2.12 Rette parallele e incidenti e sghembe, ed angoli . . . 102
2.13 Geometria Elementare Bidimensionale 102
2.14 Teorema di Pitagora 103
2.15 Cerchio, circolo, circonferenza 105
2.16 Pi greco, lunghezza del circolo, area del cerchio . . . 107
2.17 Sull'orlo dell'abisso in compagnia di pi greco 109
2.18 Geometria Elementare Tridimensionale 111
2.19 Alcune cose supposte note dagli studi precedenti . . 113
2.20 Con WolframAlpha 114

BOZZA - DRAFT

2.21 Le simmetrie, la chiralità e i farmaci 115
2.22 Perchè si studia Geometria? Da millenni? 116
2.23 Complementi – Congruenza 118
2.24 Complementi – Similitudine 118
2.25 Complementi – Quanti circoli esistono? 119
2.26 Complementi – Teoremi di Euclide. 119
2.27 Complementi – Note sugli enti geometrici 119
2.28 Complementi – Definizioni precise in Farmacia 120
2.29 ESERCIZI SULLA LEZIONE 2 122
2.29.1 Esercizio risolto a – Volume di una goccia . . 122
2.29.2 Esercizio risolto b – Volume di una capsula . 123
2.29.3 Problema tanto carino quanto difficile ☺☺☺ . 124
2.29.4 Problema insolubile ☞ 124
2.29.5 Aneddoto – Perché studiare Geometria? . . . 124

3 Logica delle proposizioni 125
3.1 Premessa 125
3.2 Introduzione 125
3.3 Connettivi logici 127
3.4 Connettivo logico non 127
3.5 Connettivo logico et 127
3.6 Connettivo logico vel 128
3.7 Connettivo logico aut 128
3.8 I connettivi logici in Farmacia 129
3.9 Tautologie e contraddizioni 130
3.10 Nota sulla Logica 131
3.11 ESERCIZI SULLA LEZIONE 3 132
3.11.1 Esercizio risolto a – Sintomi e patologia . . . 132
3.11.2 Esercizio risolto b – Finanziare terapie o no . 133
3.11.3 Esercizio risolto c – Sintomatici per il Ministero 133
3.11.4 Esercizio tanto carino quanto difficile 134

4 La Logica con l'Insiemistica 136
4.1 Nozioni di insieme, elementi e appartenenza 137

BOZZA - DRAFT

4.2 Esempio: comorbidità . . . 138
4.3 Esempio: proprietà farmacologiche . . . 139
4.4 Uguaglianza; insieme vuoto; inclusione . . . 140
4.5 Gli insiemi numerici . . . 141
4.6 Cardinalità, insiemi finiti e infiniti . . . 141
4.7 Estensioni. . . . 142
4.8 Insieme delle parti . . . 142
4.9 Operazioni insiemistiche e logiche a confronto . . . 144
4.10 Nota sull'insiemistica . . . 145
4.11 Predicati e insieme di verità . . . 146
4.12 Quantificatori . . . 147
4.13 Esempi sui quantificatori e controesempio. . . . 147
4.14 Regole di negazione . . . 149
4.15 Esempi di negazione . . . 149
4.16 Logica in Matematica e in Farmacia . . . 150
4.17 Il triste caso del monossido di diidrogeno . . . 150
4.18 Prodotto cartesiano . . . 154
4.19 Note sul prodotto cartesiano. . . . 154
4.20 ESERCIZI SULLA LEZIONE 4 . . . 157
4.20.1 Esercizio risolto a – Insieme di elementi chimici 157
4.20.2 Esercizio risolto b – Chimica combinatoria . . . 157
4.20.3 Esercizio risolto c – Schedine . . . 158

5 Calcolo Combinatorio 159
5.1 Un po' di Calcolo Combinatorio in Farmacia . . . 159
5.2 Introduzione a Calcolo e Probabilità Combinatori . . . 160
5.3 Elencazione con conteggio . . . 161
5.4 Cardinalità dell'unione . . . 162
5.5 Fattoriale . . . 164
5.6 Permutazioni . . . 165
5.7 Combinazioni semplici . . . 166
5.8 Disposizioni semplici . . . 166
5.9 Complementi – Triangolo di Tartaglia e potenze . . . 169
5.10 Complementi – Costante di Lieb . . . 170

BOZZA - DRAFT

5.11 ESERCIZI SULLA LEZIONE 5 171
5.11.1 Esercizio risolto a – Gruppo di studenti . . . 171
5.11.2 Esercizio da risolvere – Numeri scelti a caso . 171
5.11.3 Esercizio tanto carino quanto difficile 171

6 Funzioni **173**
6.1 Funzioni Numeriche 173
6.2 Passaggio a opposto, reciproco, e valore assoluto . . 173
6.2.1 Passaggio all'opposto 173
6.2.2 Passaggio al reciproco 174
6.2.3 Valore assoluto 174
6.2.4 Parte intera di un numero non negativo . . . 174
6.2.5 Segno di un numero 175
6.3 Altri esempi di funzioni numeriche 175
6.4 Immagine, controimmagine, composta 176
6.5 Funzione iniettiva, suriettiva, biiettiva, inversa 177
6.6 Crescenza e decrescenza delle funzioni numeriche . . 178
6.7 Ambiguità notazionale dell'esponente −1 179
6.8 Ambiguità notazionali della parentesi graffa grande . 180
6.9 Funzioni definite per numeri naturali: le successioni . 181
6.10 La successione di Fibonacci 185
6.11 Complementi – Fibonacci e Legge di Stigler 186
6.12 ESERCIZI SULLA LEZIONE 6 187
6.12.1 Esercizio risolto – Successione di Fibonacci . 187
6.12.2 Aneddoti spiraleggianti 188

7 Ripasso di Algebra – I parte **189**
7.1 Operazioni unarie (interne) dell'algebra dei numeri . 189
7.2 Operazioni binarie (interne) dell'algebra dei numeri . 189
7.3 Salto dell'uguale di tipo additivo 192
7.4 Salto dell'uguale di tipo moltiplicativo 193
7.5 Horribilia: frazioni miste 194
7.6 Horribilia: scrittura del marketing 194
7.7 La divisione è variamente definita 194

BOZZA - DRAFT

7.8 Divisione euclidea in Z . . . 195
7.9 Proprietà delle 4 operazioni . . . 196
7.10 Precedenze algebriche e parentesi . . . 198
7.11 Diverse opinioni, 1 o 16, su quanto fa 8/2(2+2) . . . 199
7.12 Excel sbaglia ovvero calcola modo suo . . . 200
7.13 Le potenze . . . 201
7.14 Proporzioni . . . 204
7.15 ESERCIZI SULLA LEZIONE 7 . . . 208
7.15.1 Esercizio risolto a – Equazione lineare . . . 208
7.15.2 Esercizio tanto carino quanto capriccioso . . . 208

8 Ripasso di Algebra – II parte 209
8.1 Frazioni generatrici . . . 209
8.2 Un terzo . . . 209
8.3 Le radici . . . 210
8.4 Proprietà dei radicali ovvero radici . . . 211
8.5 La radice quadrata in Medicina e Farmacia . . . 213
8.6 La radice cubica in Medicina e Farmacia . . . 215
8.7 Esempi di calcolo con le radici. . . . 216
8.8 Ordinamento dei numeri . . . 216
8.9 Altre formule classiche dell'algebra . . . 217
8.10 Radici quadrate e cubiche, ed elettrocardiogrammi . . . 218
8.11 La Medicina non è una Scienza Esatta . . . 219
8.12 Il volùbil aspartame . . . 219
8.13 Complementi – Ulteriori formule classiche . . . 221
8.14 ESERCIZI SULLA LEZIONE 8 . . . 222
8.14.1 Esercizio risolto a – Area corporea . . . 222
8.14.2 Esercizio risolto b – Mortalità fra ricoverati . . . 223

9 Piano cartesiano – I parte 224
9.1 Traiamo da Food and Drug Administration, USA . . . 224
9.2 Introduzione, anche con riferimento alla Farmacia . . . 224
9.3 Definiamo l'obesità di animaletti . . . 225
9.4 Premessa definizionale . . . 227

BOZZA – DRAFT

9.5 Assi cartesiani . . . 227
9.6 Punti del piano cartesiano . . . 228
9.7 Grafico di dispersione ovvero scatterplot . . . 228
9.8 Diagrammi e grafici . . . 230
9.9 Rette del piano cartesiano . . . 232
9.10 Funzioni e dis/equazioni di primo grado . . . 235
9.11 Esempio di rette: peso normale secondo Broca . . . 237
9.12 Nota finale sulle rette oblique . . . 239
9.13 Nota sui modelli matematici . . . 239
9.14 I modelli per fare l'interpolazione, spesso bene . . . 239
9.15 I modelli per fare l'estrapolazione, spesso male . . . 240
9.16 Esercizio risolto a – Retta e sangue . . . 242
9.17 Esercizio risolto b – Retta e sangue . . . 243

10 Piano cartesiano – II parte 245
10.1 Iperbole equilatera riferita agli asintoti . . . 245
10.2 Circolo e cerchio . . . 246
10.3 Diagramma bolle e sua ambiguità . . . 246
10.4 La parabola . . . 250
10.5 Equazione della parabola con asse verticale . . . 251
10.6 Le curve a forma di J . . . 252
10.7 Area del segmento parabolico, ed epidemie . . . 253
10.8 Note finali sulle coniche . . . 254
10.9 Risolvere le equazioni con calcoli o disegni? . . . 255
10.10 Altre figure nel piano cartesiano . . . 256
10.11 Esempio farmaceutico. . . . 257
10.12 Le famiglie di curve . . . 258

11 Piano cartesiano – III parte 259
11.1 Sistemi lineari di 2 equazioni in 2 incognite . . . 259
11.2 Esempio di sistema: miscela di 2 soluzioni . . . 260
11.3 Esempio risolto di sistema . . . 261
11.4 Esempio parzialmente risolto di sistema . . . 262
11.5 Sistemi lineari di n equazioni in n incognite . . . 262

BOZZA - DRAFT

11.6 Sistemi di dis/equazioni e sistemi generalizzati 263
11.7 Esempio di sistema non lineare 263
11.8 Parabole e funzioni e dis/equazioni di secondo grado 264
11.9 3 metodi per disegnare una parabola 265
11.10Equazione di secondo grado 266
11.11Fattorizzazione del polinomio di secondo grado . . . 268
11.12Disquazione di secondo grado 268
11.13Valore assoluto, e relative dis/equazioni 269
11.14Errore assoluto, relativo e percentuale 270
11.15La sezione aurea 271
11.16Complementi – Un sistema non lineare 273
11.17ESERCIZI SULLA LEZIONE 11 274
11.17.1 Esercizio risolto a – Cerotto e pillola 274
11.17.2 Esercizio risolto b – Risoluzione grafica 275

12 Matematica col Computer – Lezione A 277
12.1 Introduzione 277
12.2 Programma informatico sulla parte intera 279
12.3 Un'equazione di facile espressione 280
12.4 Un'equazione di difficile espressione 281

13 Funzioni e dis/equazioni ir/razionali 284
13.1 Polinomi e funzioni razionali intere 284
13.2 Funzioni e dis/equazioni razionali intere 284
13.3 Funzioni e dis/equazioni razionali fratte 286
13.4 Una funzione razionale fratta della Farmacia 287
13.5 Le funzioni radice quadrata e cubica 287
13.6 Funzioni e dis/equazioni irrazionali 288
13.7 Una funzione algebrica con grafico sigmoide 288
13.8 Complementi – Un esempio non banale 291
13.9 Complementi – Un esempio difficile 291
13.10Complementi – Dis/equazioni con radice quadrata . 293

14 Esponenziali e logaritmi – I parte 296
14.1 Vecchie funzioni potenza e nuove esponenziali 296

BOZZA - DRAFT

14.2 La sostanziale unicità degli esponenziali 297
14.3 Le 4 basi di maggior interesse 297
14.4 L'esponenziale in Farmacia 299
14.5 Una sigmoide basata sull'esponenziale 300
14.6 Logaritmi 300
14.7 I diagrammi logaritmici 302
14.8 Logaritmi in Farmacia e nelle Scienze Applicate . . . 304
14.9 La curva di Preston 305

15 Esponenziali e logaritmi – II parte 310
15.1 Definizione del logaritmo 310
15.2 Logaritmi naturali e decimali; esponenziale 311
15.3 "Trasforma prodotti in somme" 313
15.4 "Trasforma divisioni in sottrazioni" 314
15.5 "Trasforma reciproco in opposto" 314
15.6 "Trasforma la radice ennesima in divisione per n" . . 315
15.7 "Trasforma l'elevare a potenza in moltiplicare" 315
15.8 Le proprietà viste, con riferimento alle varie basi . . 316
15.9 Prevenire è meglio che curare: errori tipici 318
15.10Esempio di uso del logaritmo: il pH. 319
15.11Esempi di calcoli coi logaritmi 320
15.12Altre proprietà degli esponenziali e dei logaritmi . . 322
15.13ESERCIZI SULLA LEZIONE 15 324

16 Esponenziali e logaritmi – III parte 325
16.1 Conversioni fra logaritmi naturali e decimali 325
16.2 Primo passo per il calcolo dei logaritmi 325
16.3 Calcolare i logaritmi decimali dei numeri da 1 a 10 . 326
16.4 Risoluzione pratica delle equazioni con exp e log . . 327
16.5 Soluzioni spurie ovvero fittizie 332
16.6 Equazioni e disequazioni con exp e log – teoria . . . 334

17 Esponenziali e logaritmi – IV parte 338
17.1 L'esponenziale e la campana gaussiana 338
17.2 Funzione di Gompertz 340

BOZZA DRAFT

17.3 Il logaritmo decimale come contatore di zeri 340
17.4 Ordine di grandezza di un numero positivo 341
17.5 La successione di Fibonacci e gli esponenziali 344
17.6 Emivita di una sostanza nel sangue o nelle ossa . . . 345
17.7 L'intensità sonora: logaritmi e decibel 348
17.8 Equazioni con logaritmi della x e x libera 349
17.9 ESERCIZI SULLA LEZIONE 17 350
17.9.1 Esercizio risolto a – Risoluzione per via grafica 350
17.10 Altri sercizi su logaritmi ed esponenziali 351

18 Funzioni trigonometriche 354
18.1 Generabili nella geometria riappaiono ovunque 354
18.2 Oscillazioni periodiche ovunque 355
18.3 Le funzioni trigonometriche in Farmacia 356
18.4 Le nozioni principali 357
18.5 Definizioni 360
18.6 Gradi e radianti 362
18.7 Gli angoli di legame della Chimica 363
18.8 Alcuni valori notevoli 364
18.9 Funzioni goniometriche inverse 365
18.10 Trigonometria – La goniometria dei triangoli 368
18.11 Complementi – Altre formule goniometriche 369
18.12 Complementi – Il calcolo di queste funzioni 370
18.13 Complementi – Angoli notevoli 370
18.14 Complementi – Trigonometria 372

19 Introduzione alla Statistica Descrittiva 379
19.1 Introduzione 379
19.2 Riassumere i dati: quanti gatti e quanti cani avete? . 380
19.3 Esempio di possibile raccolta dati 382
19.4 Il dataset e le operazioni su esso 383
19.5 Medie 386
19.6 Inesistenza di una media migliore di tutte 390
19.7 Disilludersi sulle medie anche senza outlier 391

ROZZA - DRAFT

19.8 Stratificazione per età in Medicina e Farmacia 392
19.9 Sul trasformare dati ordinali in dati numerici 392
19.10 Le medie creano illusioni percettive 393

20 Bar chart, istogrammi, e altri diagrammi 398
20.1 Diagrammi a torta 398
20.2 Catastrofe Statistica: dati mancanti e drop out . . . 398
20.3 Istogrammi a barre o bar chart, e istogrammi 399
20.4 Eventuali asimmetrie nei dataset: skewness 402
20.5 Funzioni a campana varie 402
20.5.1 Esercizio risolto a – Bar chart 406
20.5.2 Esercizio risolto b – Istogramma 408

21 Quartili e box-plot 410
21.1 Quartili e riassunto dei 5 numeri 410
21.2 Box-plot ovvero diagramma a scatola e baffi 412
21.3 I dati? Quali dati? 416
21.3.1 Esercizio risolto a – Media interquartile . . . 417

22 Variabilità, covarianza e correlazione 419
22.1 Indici di dispersione ovvero variabilità 419
22.2 Covarianza; correlazione; retta di regressione 422
22.3 L'inimicizia fra progresso e variabilità 423

23 Note finali sulla Statistica Descrittiva 426
23.1 Bugie, bugie cattive, e statistica 426
23.2 La scelta dei parametri. Cosa misurare? 427
23.3 Rispetto a quale standard fare le statistiche? 430
23.4 Col ranking facciamo ben figurare il Paraguay 434
23.5 Illusioni percettive nella presentazione dei dati 435
23.6 Di cosa parliamo? 436
23.7 Cherry picking – bias di selezione 437
23.8 Approfittare delle fluttuazioni statistiche 438
23.9 Capziosità verbali 439
23.10 Illustrazione di un triplice esempio reale 440

BOZZA - DRAFT

23.11Conclusioni sulla fallace presentazione di dati veri . . 441
23.12Il valore anomalo: outlier 444
23.13Omissione di dati ritenuti poco significativi 445
23.14"Fatta la legge, trovato l'inganno" 445
23.15Traslazioni e confusioni linguistiche 446
23.16Il problema della discrezionalità 447
23.17Esempio importante: le cause di morte 447
23.18Cose vere in articolo titolato oppostamente 449
23.19La falsificazione dei dati 450

24 Limiti di successioni e serie geometrica 456
24.1 Limite della concentrazione plasmatica 456
24.2 Una triste storia di infermità a lieto fine e una no . . 457
24.3 Un limite banale ma illuminante 458
24.4 Successioni (di numeri reali) 459
24.5 Limite finito delle successioni 460
24.6 Limite nullo delle successioni 461
24.7 I 2 possibili limiti infiniti delle successioni 462
24.8 Sulla rapidissima crescita del fattoriale 463
24.9 Serie geometrica di ragione fra 0 e 1 464
24.10Complementi – Problema aperto: Collatz 466
24.11Complementi – Problema aperto: Feigenbaum 466
24.12Complementi – Esperienza pratica coi limiti 466
24.13Complementi – 4 tipi di definizioni di successioni . . 466
24.14Complementi – Applicazione probabilistica 472

25 Limiti e continuità per funzioni reali 474
25.1 Introduzione ai limiti di funzioni 474
25.2 Limiti verso un numero finito, e continuità 476
25.3 Funzioni limitate, infinite, infinitesime 477
25.4 Teoremi algebrici sui limiti 477
25.5 Esistenza di forme di indecisione dei limiti 479
25.6 Complementi – Forme di indecisione dei limiti 481
25.7 Complementi – Limiti di successioni prolungabili . . 481

BOZZA - DRAFT

26 Derivate – I parte **484**
26.1 Introduzione alle derivate 484
26.2 Notazioni per la derivata prima 484
26.3 Derivate delle funzioni elementari reali 485
26.4 Derivata in un punto e funzione derivata 486
26.5 Epidemie, curve sigmoidi, a campana, derivata 489

27 Derivate – II parte **493**
27.1 Teoremi algebrici sulle derivate 493
27.2 Le funzioni elementari in Farmacia, e i modelli 495

28 Applicazioni del Calcolo Differenziale - I **499**
28.1 Cos'è molto o poco? Migliora o peggiora? 499
28.2 De/crescenza, min/max, concavità, flessi 500
28.3 Retta tangente a un grafico in un punto 502
28.4 Approssimazioni delle potenze presso l'unità 503

29 Applicazioni del Calcolo Differenziale - II **507**
29.1 Regola di de l'Hospital 507
29.2 Asintoti 508
29.3 sup e inf 510
29.4 La "ricetta" per lo studio di funzione 510
29.5 Esempio 1: la campana gaussiana 511
29.6 Esempio 2: il Potenziale di Lennard-Jones 513
29.7 Le equazioni differenziali 517
29.8 Il modello SIR dell'epidemiologia 518
29.9 Esercizi di studio di funzione 519
29.10Riempitivo lorem ipsum 521

30 Integrale indefinito e ODE **522**
30.1 Introduzione all'integrale indefinito 522
30.2 Alcuni integrali indefiniti 523
30.3 La costante additiva dell'integrale indefinito 524
30.4 Le 2 formule della linearità dell'integrale 524
30.5 Un'equazione differenziale della Farmacocinetica . . 526

BOZZA - DRAFT

30.6 Equazioni differenziali risolubili con un'integrazione . 526
30.7 Complementi – Altre formule sull'integrale indefinito 528
30.8 Complementi – Integrale del logaritmo naturale . . . 528
30.9 Complementi – Le equazioni differenziali 529

31 L'integrale definito 531
31.1 La questione dell'area sotto una curva 531
31.2 Teoria ed esempi dell'integrale definito 534
31.3 Altri 3 teoremi sugli integrali 535
31.4 Note finali sulle funzioni elementari 540

32 Matematica col Computer – Lezione B 542
32.1 Ordinare dati con LibreOffice 542
32.2 Visualizzazione grafica con Calc di LibreOffice 543
32.3 La successione di Fibonacci con LibreOffice 548
32.4 Medie varie con LibreOffice 549

33 Introduzione al Calcolo delle Probabilità 554
33.1 Inquadramento della questione 554
33.2 La concezione frequentista della probabilità 555
33.3 La mortalità 556
33.4 La concezione classica della probabità 557
33.5 Concezione soggettiva della probabilità 560

34 Concezione assiomatica della probabilità 564
34.1 Le basi, gli assiomi 564
34.2 L'evento impossibile fa inghippi che fuggiremo 565
34.3 Mortalità e letalità di una malattia o infezione 566
34.4 La letalità apre una finestra su un baratro teorico . . 567
34.5 Indipendenza degli eventi 569

35 Sensibilità, specificità, predittività, ROC 577
35.1 Sensibilità 578
35.2 Specificità 579
35.3 Predittività 580

BOZZA - DRAFT

35.4 Esempio: i test rapidi per il covid-19. 582
35.5 Le curve ROC 582

36 Probabilità combinatoria **587**
36.1 Dismutazioni 587
36.2 I compleanni coincidenti 588
36.3 Il gioco d'azzardo: patologia, probabilità e storia . . 589

37 Introduzione alle variabili aleatorie **593**
37.1 Variabili aleatorie discrete e continue 593
37.2 La campana gaussiana e quella di Cauchy 597

38 Variabili aleatorie discrete **599**
38.1 Alcuni esempi di variabili aleatorie discrete 600
38.2 Forma del grafico di densità e f.r. 601
38.3 Variabili aleatorie uniformi discrete 602
38.4 Variabili aleatorie discrete con infiniti valori 602

39 Variabili aleatorie uniformi e geometriche **604**
39.1 Variabili aleatorie e processo di Bernoulli 604
39.2 Variabili aleatorie geometriche 605

40 Leggi congiunte e indipendenza **611**
40.1 Introduzione alle leggi congiunte 611
40.2 Densità marginali e indipendenza 612

41 Variabili aleatorie binomiali e bimodali **613**
41.1 Le variabili aleatorie binomiali 613
41.2 La binomiale come spiegazione delle campane 615
41.3 Densità in qualche modo bimodali 617
41.4 La natura empirica delle densità bimodali 621
41.5 ESERCIZI SULLA LEZIONE 39 622
41.5.1 Esercizio risolto a - Dado regolare a 8 facce . 622

42 Speranza matematica e varianza **623**
42.1 Speranza matematica (o valore atteso, o media) . . . 623

BOZZA - DRAFT

42.2 Costi e benefici 625
42.3 Varianza 627
42.4 Note su varianza e istogramma della densità 628
42.5 L'aspettativa di vita o speranza di vita 629
42.6 Passeggiate aleatorie 631
42.7 Esercizi sulla speranza matematica di v.a. discrete . 633

43 Uniforme continua, di Cuchy, esponenziale 636
43.1 Variabile aleatoria uniforme continua 638
43.2 Variabile aleatoria di Cauchy. 638
43.3 Variabile aleatoria esponenziale 639

44 Quantili delle variabili aleatorie continue 643
44.1 Definizione semplificata 643
44.2 Mediana di variabile aleatoria continua 643
44.3 Esempio di quantile classico 644
44.4 Definizione più completa di quantile 645
44.5 Proprietà di densità e funzioni di ripartizione 646

45 Speranza matematica, varianza, covarianza 649
45.1 Sulle definizioni di distribuzione e legge 649
45.2 Introduzione e definizioni 649
45.3 Approfondimenti sulle v.a. discrete e continue 650
45.4 Funzione Gamma e Legge Gamma 651
45.5 Densità e leggi del chi quadrato 652
45.6 Densità e quantili di Student, e legge di Cauchy . . . 654
45.7 Legge di $g(X)$ 657
45.8 Standardizzazione di una variabile aleatoria 659
45.9 Complementi – Disuguaglianza di Chebyshev 662
45.10Complementi – Esercizi sulla legge di $g(X)$ 663

46 Densità e variabile aleatoria normale 664
46.1 Introduzione alla densità e v.a. normale 664
46.2 Variabile aleatoria normale standard 666
46.3 Quantili normali, e valore 1.96 668

BOZZA - DRAFT

46.4 Scarti dalla media per v.a. normale e qualunque . . . 669

47 Approssimazione di $\Phi(x)$ e ϕ_α 674
47.1 Grafici . . . 674
47.2 Approssimazioni . . . 675
47.3 Variabile aleatoria log-normale . . . 678
47.4 Confronto fra normale e log-normale, e cigni neri . . . 680

48 Legge dei Grandi Numeri 684
48.1 Inquadramento euristico della situazione . . . 684
48.2 Legge dei Grandi Numeri semplificata . . . 685
48.3 Alcune precisazioni . . . 686
48.4 Limite in probabilità e Legge dei Grandi Numeri . . . 686

49 Approssimazione Normale 691
49.1 Convergenza in legge . . . 691
49.2 Teorema Limite Centrale . . . 693
49.3 Simulazione di una variabile aleatoria uniforme . . . 694
49.4 Simulazione di una variabile aleatoria normale . . . 697
49.5 Sul naturale formarsi delle campane . . . 700
49.6 Lo scarto di una radice quadrata dalla metà . . . 702
49.7 Calcolo sullo scarto della radice quadrata . . . 703
49.8 Lo scarto di mezza radice quadrata dalla metà . . . 704

50 Note finali sul Calcolo delle Probabilità 707
50.1 Sulle problematicità del Calcolo delle Probabilità . . . 707
50.2 Mettiamo in dubbio l'esistenza della probabilità . . . 709
50.3 Sull'applicare le probabilità a singoli casi . . . 710
50.4 Ma alla fine la probabilità verrà considerata? . . . 711
50.5 ...Forse sì, ma dalle intelligenze artificiali . . . 712
50.6 Il Calcolo delle Probabilità è controintuitivo . . . 714
50.7 Conclusioni . . . 717

51 Introduzione alla Statistica Inferenziale 727
51.1 Citazione da un testo importante . . . 727

BOZZA - DRAFT

51.2 Introduzione 728
51.3 Primo esempio completo di Statistica Inferenziale . . 730
51.4 Secondo esempio di Statistica Inferenziale 731
51.5 Riconoscere statisticamente una mancanza di neonate 733

52 Stimatori e stimatori non distorti 736
52.1 Parametri e stimatori; stimatori non distorti 736
52.2 Stimatori non distorti 738
52.3 Stimatori non distorti di media e varianza 738
52.4 Cenno agli stimatori di massima verosimiglianza . . 740
52.5 Riassumiamo: gli stimatori puntuali 742
52.6 Nota dolente finale e prospettive future 743
52.7 Complementi – gli stimatori da un libro classico . . . 744

53 Intervalli di fiducia; media di $N(\mu, \sigma^2)$ 747
53.1 Introduzione agli intervalli di fiducia 747
53.2 Intervalli di fiducia per μ per campioni gaussiani . . 749
53.3 Apprezziamo i CI in un articolo scientifico 752
53.4 Prevenire è meglio che curare: le 2 SD e il 95% . . . 754
53.5 Possibile delusione dagli articoli scientifici 755
53.6 Esercizi 756
53.6.1 Esercizio 756
53.6.2 Esercizio 757

54 Intervalli di fiducia per la varianza 761

55 I Test Statistici 763
55.1 Introduzione ai Test Statistici 763
55.2 Il test della signora del tè 767
55.3 Test a una coda e a due code 768
55.4 Test a una coda facilmente fattibile con successo . . 770
55.5 A due code, facilmente fattibile, ma con insuccesso . 771
55.6 Il p-value, in italiano valore p 772
55.7 Buona Statistica: ipotesi prima dell'esperimento . . . 774
55.8 Nota sul sentire statistico 776

BOZZA - DRAFT

56 Errori di I e II Specie **778**
56.1 Due tipi di sfortune nei test statistici 778
56.2 Nota sul parametro medico su cui si fa il test 782
56.3 Dopo gli errori di I e II specie, le assurdità 785

57 Test del χ^2 di indipendenza **789**
57.1 Inquadramento della questione 789
57.2 Semplice statistica per trial clinico 791
57.3 Sulle condizioni per la validità del test 794
57.4 Test di Fischer esatto 794
57.5 Morti per tutte le cause 794
57.6 Il solito problema: cosa misuriamo? Placebo 795
57.7 Trucchi del mestiere 795
57.8 Duplice esempio da un articolo scientifico 797

58 Il Test del χ^2, quello basico **800**
58.1 premessa sul nome del test 800
58.2 Premessa sul livello del test 800
58.3 La teoria del Test del χ^2, quello basico 800
58.4 Le tavole del chi quadrato 802
58.5 Densità di Benford 808

59 Il test dei ranghi **810**
59.1 Premessa 810
59.2 Testing monotonicity of ranks 810
59.3 Lombardia, 31 March 2020 811
59.3.1 Original data for Lombardia, 31 March 2020 812
59.3.2 Derived data for Lombardia, 31 March 2020 . 813
59.3.3 Ranks 813

60 Alcuni Test di Student **815**
60.1 Premessa: come misurare l'esposizione? 815
60.2 Test di Student per la media $\leq$ e = ... 815
60.3 Note euristiche sui 2 Test di Student 818
60.4 Esempio estremo per fissare le idee 818

BOZZA - DRAFT

61 Test di Student per il confronto di medie **826**

62 Matematica col Computer – Lezione C **829**
62.1 Risolvere un'equazione col computer online 829
62.2 Risolvere un'equazione col computer offline 830
62.3 Grafici cartesiani e polari con MATLAB 830

63 Matematica delle Epidemie **833**
63.1 Alcune cose viste in Statistica Descrittiva 833
63.2 Alcune cose viste in altre Matematiche Elementari . 835
63.3 Alcune cose viste in Calcolo Infinitesimale 836
63.4 Alcune cose in Calcolo delle Probabilità 836
63.5 Alcune cose in Statistica Inferenziale 836
63.6 OR e RR 838
63.7 NNT: il numero di persone da trattare 838
63.8 Il modello logistico e le loglet 839
63.9 La pandemia 2020-2021 – i numeri e le chiacchiere . 840
63.10L'arbitrarietà dei modelli previsionali 841
63.11Il tasso di letalità 843
63.12Mappe coropletiche 847

64 Note finali sulla Statistica **849**
64.1 I numeri parlano da soli? Forse una lingua ignota . . 849
64.2 Quale misurazione è più profittevole? 849
64.3 "I numeri che vi diamo parlano da soli" 851
64.4 Falsificazioni deliberate 852
64.5 Il (doppio) cieco che ci vede benissimo 853
64.6 Il conflitto di interessi 855
64.7 Come avviene l'inversione di significante 857
64.8 I dati? Quali dati? Postilla 858
64.9 Bottom line – Legge pseudo-magica della Statistica . 859

65 Le serie numeriche **860**
65.1 Introduzione alle serie 860
65.2 Il calcolo del logaritmo di 2 della Farmacocinetica . . 861

BOZZA - DRAFT

65.3 Definizioni e notazioni 861
65.4 Il numero di Nepero da una serie esponenziale 863
65.5 Il numero pigreco: la serie ciclometrica 863
65.6 La serie armonica 863
65.7 Complementi – Proprietà delle serie 864

66 APPENDICE – Esperienze statistiche in aula 867
66.1 Questionario per esperienze statistiche in aula 868
66.2 Lez. 55 – Test simile a quello della signora del tè . . 869
66.3 Lez. 52 – Stima del secondo estremo dell'uniforme . 871
66.4 Caccia al Tesoro: usare AI e rete 871

BOZZA - DRAFT

PARTE A — MATEMATICHE DELLA CERTEZZA

BOZZA - DRAFT

Sezione A1 – Matematiche Elementari

BOZZA - DRAFT

I – Insiemistica e logica

BOZZA - DRAFT

1 I numeri

1.1 Perché numeri, se ci interessano farmaci?

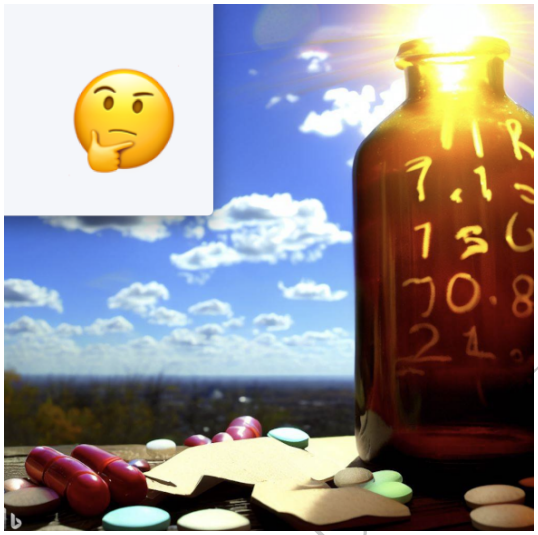


Immagine: Bing Image Creator, rielaborata.

BOZZA

Se in una qualunque tecnologia non si fanno misurazioni non è facile fare miglioramenti

Alla fine di questa trattazione, vedremo che misurarle non basta più, bisogna sottoporre i numeri ad analisi statistica.

Detto molto approssimativamente, la Medicina moderna e con essa la Farmacia, tende sempre più a ridurre
la Medicina a Biologia
la Biologia a Biochimica
la Biochimica a Chimica
la Chimica a Fisica
la Fisica a Matematica.
(*Scientismo*).

Da ciò seguono molti successi. E pure insuccessi, anche perché ad ogni "riduzione" si fa un'approssimazione, più precisamente si sostituisce una realtà con un suo *modello*, inevitabilmente approssimato.

Nondimeno, appare chiarissimo che i modelli matematici riaffiorano costantemente in tutte le sopraelencate Scienze Applicate, *e in tutte le altre*.

1.2 Alcuni usi dei numeri in Farmacia

Ecco alcuni usi dei numeri in Farmacia:

- gestione economica della Famacia;
- gestione logistica della Farmacia;
- dosare sostanze nei farmaci;
- esprimere parametri corporei, come la temperatura o la glicemia;
- confrontare performance di diversi farmaci

che si potrà fare dopo la cosa seguente, di tipo statistico:

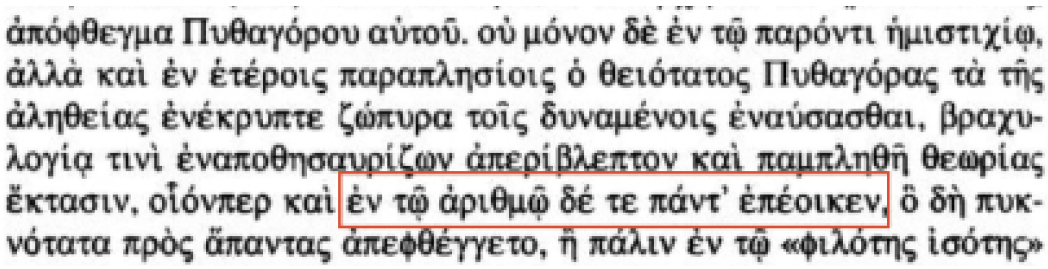
- riassumere molti dati in pochi numeri densi di significatività.

BOZZA - DRAFT

Già Pitagora 25 o 26 secoli fa:

:

"Ogni cosa si adatta al numero"


ἀπόφθεγμα Πυθαγόρου αὐτοῦ. οὐ μόνον δὲ ἐν τῷ παρόντι ἡμιστιχίῳ, ἀλλὰ καὶ ἐν ἑτέροις παραπλησίοις ὁ θειότατος Πυθαγόρας τὰ τῆς ἀληθείας ἐνέκρυπτε ζώπυρα τοῖς δυναμένοις ἐναύσασθαι, βραχυλογίᾳ τινὶ ἐναποθησαυρίζων ἀπερίβλεπτον καὶ παμπληθῆ θεωρίας ἔκτασιν, οἷόνπερ καὶ ἐν τῷ ἀριθμῷ δέ τε πάντ' ἐπέοικεν, ὃ δὴ πυκνότατα πρὸς ἅπαντας ἀπεφθέγγετο, ἢ πάλιν ἐν τῷ «φιλότης ἰσότης»

Detto grezzamente, i numeri saltano fuori dappertutto.

> Int J Hyg Environ Health. 2020 Aug;229:113594. doi: 10.1016/j.ijheh.2020.113594.
Epub 2020 Aug 5.

Metabolites of the fragrance 2-(4-tert-butylbenzyl)propionaldehyde (lysmeral) in urine of children and adolescents in Germany - Human biomonitoring results of the German Environmental Survey 2014-2017 (GerES V)

Aline Murawski [1], Nina Fiedler [2], Maria I H Schmied-Tobies [3], Enrico Rucic [3], Gerda Schwedler [3], Markus Stoeckelhuber [4], Gerhard Scherer [4], Nikola Pluym [4], Max Scherer [4], Marike Kolossa-Gehring [3]

Affiliations + expand

PMID: 32763837 DOI: 10.1016/j.ijheh.2020.113594

Free article

Abstract

The synthetic fragrance 2-(4-tert-butylbenzyl)propionaldehyde, also known as lysmeral, butylphenyl methylpropional, lilial, or lily aldehyde, is widely used in cosmetics, personal care products, laundry detergents, and air fresheners. It is classified as suspected to be harmful to fertility and possibly endocrine disrupting. Its maximum concentration in cosmetics is limited. First-morning void urine samples (N = 2133) were analysed for several metabolites of lysmeral (Chemical Abstract Service (CAS) No.: 80-54-6). Samples were collected in the population-representative German Environmental Survey for Children and Adolescents 2014-2017 (GerES V) from German residents aged 3-17 years. Four main metabolites tert-butylbenzoic acid, lysmerol, lysmerylic acid, and hydroxy-lysmerylic acid were found in quantifiable amounts in 100%, 99%, 40%, and 23% of the samples, respectively, with geometric mean concentrations of 10.21 µg/L (8.658 µg/g_{crea}) for tert-butylbenzoic acid, 1.528 µg/L (1.296 µg/g_{crea}) for lysmerol, and below the limit of quantification of 0.2 µg/L and 0.4 µg/L for lysmerylic acid and hydroxy-lysmerylic acid, respectively. Girls had higher urinary concentrations of lysmeral metabolites than boys. Usage of fragrances, fabric softener, and personal care products, especially perfume, was positively associated with urinary concentrations of lysmeral metabolites. Source identification builds a basis to derive proposals for reduction of exposure. These results can also provide the foundation for developing reference values for urinary metabolite concentrations of lysmeral in children and adolescents in Germany that will facilitate recognising future exposure trends.

Keywords: BMU/VCI project; Butylphenyl methylpropional; Endocrine disruptor; Exposure; Lilial; Lily aldehyde.

Vediamo su PubMed l'abstract di un articolo scientifico sul lilial, una sostanza di cui in media hanno molto parlato, per esempio (14/9/2023) Rainews:

"Sugli scaffali erano pronti per la vendita ma all'interno, contenevano una sostanza vietata dall'Unione Europea dal 2022, perchè nel novero delle sostanze "ritenute cancerogene e tossiche per la fertilità". Il mega sequestro dei finanzieri a Pisa, durante un'ispezione in 7 negozi della città toscana: 2000 confezioni di profumi, bagnoschiuma, shampoo, creme, lozioni, saponi, prodotti di varia natura ma tutti contenenti il Butylphenyl Methylpropional (sulle etichette chiamato "Lilial")."

Si noti **quanti numeri**. (E il fantasmagorico microgrammo.)

Non esiste un foglietto illustrativo di un farmaco,
non esiste un articolo scientifico di Farmacologia,
non esiste bilancio né organizzazione di una farmacia,
che non abbia numeri.

BOZZA - DRAFT

PubMed.

È un servizio di ricerca gratuito di letteratura scientifica biomedica. È su sito governativo statunitense dei NIH, National Institutes of Health:

`https://pubmed.ncbi.nlm.nih.gov/`

"comprises more than 37 million citations for biomedical literature"

Sono strapieni di statistica medica, sia descrittiva che inferenziale.

Articoli scientifici.

La scienza moderna è nata verso il 1600 e fin dal 1665 si è trovata bene con *articoli scientifici* pubblicati su *riviste scientifiche*.

1.3 La sconcertante questione della virgola e del punto

Una caloria termochimica, con simbolo cal,

> secondo i testi in italiano (e altre lingue) equivale a 4,184 J
> secondo i testi in inglese (e altre lingue) equivale a 4.184 J

Quante saranno, circa 4 o circa 4mila?

Nemmeno un organismo del livello della Comunità Europea (mezza miliardata di abitanti), che pure può normare anche il diametro regolamentare delle clementine (minimo 35 millimetri), non può standardizzare questa ambiguità notazionale.

BOZZA - DRAFT

Nella scrittura dei numeri con le cifre decimali (numeri “arabi", di origine indiana)

0, 1, 2...9

ci sono 2 questioni:
– separare la parte intera dalle cifre decimali
– separare eventualmente le terne di cifre per facilitare la lettura.
Per risolvere le 2 questioni vengono *variamente usati* 4 simboli:
• la virgola
• il punto
• lo spazietto
• il punto a mezza altezza ·

Incredibilmente a tutt’oggi non esiste uno standard internazionale per il separatore della parte decimale.

(Ebbene, per la caloria termochimica si tratta di *poco più di 4* Joule.)

Importante

Usando il punto decimale si segue il miglior standard che sperabilmente si ritroverà negli articoli scientifici internazionali di Farmacia, in lingua inglese. (Ma non può qua essere garantito).
Invece in farmacia in Italia in generale si troverà quasi l'opposto:
virgola per separare la parte intera dai decimali: $\pi \approx 3{,}141$
punto per separare le migliaia: 1.250 per milleduecentocinquanta.

BOZZA - DRAFT

Negli stati di lingua inglese e negli articoli scientifici internazionali in inglese di solito si usa il punto decimale, e in Italia spesso la virgola decimale.

Molti Autori che usano il punto come separatore decimale, usano la virgola come separatore delle migliaia, **e viceversa**.

Wikipedia in italiano dice che 1 tesla equivale a 10.000 gauss
Wikipedia in inglese dice che 1 tesla equivale a 10,000 gauss
Ebbene, si tratta di *diecimila* gauss. (Sono unità di misura del campo magnetico, quello della risonanza magnetica).
E su calcolatrici diverse, materiali o virtuali, disponibili qua in Italia, il punto e la virgola hanno proprio quei 2 diversi significati! **Attenzione!**
Nei vari luoghi del mondo ci si può aspettare una varietà di usi.

Spesso si riesce a capire il significato del punto, o della virgola, dal contesto, e, sulla calcolatrice, provando a scrivervi numeri.

BOZZA - DRAFT

Radice quadrata di due milioni, su diverse calcolatrici:

1,414.21356 1.414,21356

(Stranamente la virgola e il punto come separatori delle terne di cifre sono in generale usati solo prima del punto/virgola decimale).

Giustamente la *Raccomandazione per la prevenzione degli errori in terapia conseguenti all'uso di abbreviazioni, acronimi, sigle e simboli* (settembre 2018) del Ministero della Salute italiano ci ricorda che

> **L'uso non standardizzato di abbreviazioni, acronimi, sigle e simboli, può indurre in errore e causare danni ai pazienti**

Importante

Dice il Ministero:

> usare il punto per separare i tre zeri delle migliaia o usare parole come 1 milione per favorire la corretta interpretazione (ad esempio, 1000 unità va scritto 1.000 unità, 10000 unità va scritto 10.000 unità)

Il presente testo fa una scelta drastica: usa entrambi gli standard, del punto decimale e della virgola decimale, per abituare lo studente.

BOZZA DRAFT

Inoltre, seguendo il N.I.S.T. (National Institute of Standards and Technology) statunitense, di solito userà lo spazietto come separatore delle terne di cifre.
Per esempio per la costante di Faraday il N.I.S.T. dà

$$96\,485.332\,12...\ \mathrm{C\ mol^{-1}}$$

(Che veramente parrebbe la scrittura migliore, N.d.S.).
(Altri separano con lo spazietto le cinquine di cifre.)

Excel è il più utilizzato programma per la produzione e gestione di fogli elettronici, ed è usatissimo nelle Scienze Applicate, per archiviare e gestire i dati sperimentali.
Excel settato con la lingua italiana usa la virgola decimale, settato con la lingua inglese usa il punto decimale.

*

Lo scrivente (*vox clamantis in deserto*) apprezza un modo quasi mai usato: l'apostrofo come separatore delle terne di cifre, o almeno (prima del punto/virgola decimale) delle migliaia:

> 10'000 Gauss, finalmente del tutto chiaro.

1.4 Il punto a mezza altezza

Su certi testi tecnico-scientifici, talvolta antiquati ma talvolta modernissimi, $3 \cdot 673$ indicherebbe quel numero minore di 4 che in italiano scriverebbero 3,673: cioè, alcuni usano il punto a mezza altezza come se fosse un punto/virgola decimale: ecco un esempio in articolo scientifico (sul cancro infantile, 2017) su rivista scientifica internazionale di alto livello, The Lancet Oncology: Link->

> Since the 1980s, the global WSR[3] of registered cancers in children aged 0-14 years has increased from 124·0 (...) to 140·6 (...) per million person-years.

Oppure si veda su The Lancet stesso, la "rivista madre", di altissimo livello, nel *comment* (2020) (cosa diversa da un *articolo scientifico* [*peer review*]) a questo Link->, in cui si noti pure lo spazietto separatore:

> projections suggest that even fairly short lockdown measures, combined with severe mobility disruptions and comparatively moderate food systems disruptions, result in most LMICs[4] having an estimated average 7·9% (...) decrease in GNI[5] per capita relative to pre-COVID-19 projections (...) 128 605 (...) additional deaths in children younger than 5 years during 2020

BOZZA - DRAFT

Si veda ancora questo esempio, su come scriveremo in generale:

il punto a mezzo indica la moltiplicazione→ $3 \cdot 675 = 2\,025$ ← *si noti lo spazietto*

Ripetiamo che il presente testo elementare, dovendosi limitare, userà quasi solo il punto decimale e la virgola decimale, riservando pochissimo spazio al punto a mezza altezza con funzione di punto decimale ovvero virgola decimale, lasciando in generale quel simbolo per la moltiplicazione.

[3] Age-standardised incidence rate.
[4] Low-income and middle-income countries.
[5] Gross national income.

Importante

Per risolvere 2 questioni:
– separare la parte intera dalle cifre decimali
– separare eventualmente le terne di cifre per facilitare la lettura.
vengono *variamente usati* 4 simboli:
• la virgola
• il punto
• lo spazietto
• il punto a mezza altezza ·

Si faccia attenzione!

BOZZA - DRAFT

ESERCIZIO$_{\mu 2024}$ ⋆ (R) Completare la parola mancante:
• L'assunzione di
1,125 microgrammi
di selenio in un breve periodo è da considerare pericolosa per la salute, intendendo la virgola come separatore delle migliaia (uso scientifico internazionale).
• L'assunzione di
1,125 microgrammi [6]
di selenio in un breve periodo è da considerare in generale sicura (sebbene qua non possiamo dare garanzie mediche), intendendo la virgola come ... (uso farmaceutico italiano).

virgola decimale

Lo scrivente spera si sia capita la criticità della questione.

[6] Per esempio 9 microgrammi suddivisi in 8 giorni. Per esempio si trova facilmente un multivitaminico Multivitamin Donna con 8.3 microgrammi di selenio per compressa.

1.5 Prima definizione dei numeri: scritture decimali

Un numero è la scrittura decimale valida di un numero; cosa sia scrittura decimale valida di un numero è lunghissimo da specificare (con casi e sottocasi) ma chi ha fatto le scuole ormai lo sa.

Per esempio, usando per adesso lo standard del punto decimale, abbiamo:

numeri	non numeri
3	A
2023	2023\$
-3	± 3
-32.9	∓ 32.9
3.14	$3.1F$
$+3.14$	3.1\$
$3.1\overline{4}$	$3.F2$
$3.81\overline{14}$	3;14

BOZZA - DRAFT

Solo alcune avvertenze:

• 3.2 e 3.20 e 3.200, ecc., in matematica sono uno stesso numero (nelle Scienze Applicate invece indicano 3 quantità note con crescente precisione);

(a) ± 3 e ∓ 3 sono scritture valide, ma di una coppia di numeri;

(b) 3.14 e $+3.14$ sono lo stesso numero;

(c) 0 e $+0$ e -0 sono lo stesso numero e gli ultimi 2 si usano solo in Informatica;

(d) 3.14 e $3.14\overline{0}$ e $3.13\overline{9}$ sono lo stesso numero, e l'ultima scrittura è da evitarsi;

(e) $9.\overline{3}$ e $9.3\overline{3}$ e $9.\overline{33}$ e $9.\overline{333}$ sono lo stesso numero, e solo la prima scrittura è opportuna; similmente per (infinite) altre stranezze;

(f) $\frac{1}{2}$ è anch'esso un numero ma dovrà essere scritto 0.5

(g) $\frac{1}{3}$ è anch'esso un numero ma dovrà essere scritto $0.\overline{3}$

(h) $\sqrt{2}$ è anch'esso un numero ma dovrà essere scritto 1.4142... con infinite cifre e questo si potrà fare, nel nostro mondo reale, solo a un livello teorico; ma non è un problema: $\sqrt{2}$ resta un numero, perchè di esso esiste la sua scrittura decimale valida, anche se nessun essere umano può vederla tutte.

Importante

Ogni scrittura decimale valida è un numero;
e sì, per ogni numero esiste la sua scrittura decimale; ed essa è unica salvo le (*quasi solo per matematici*) ambiguità (c), (d), (e).
Tutto funziona bene.

Ecco spiegato cosa sono i numeri, in un primo modo: scritture!

Esistono anche scritture in base non decimale, in particolare *binaria* ed *esadecimale*; anche con virgola (o punto) decimale.

☺ Esistono 10 tipi di persone: quelle che conoscono il sistema binario e quelle che non lo conoscono.

ES. 0a$_{\mu 2024}$ (R) $*$ Quanti mesi ci sono in 10,5 anni?

126

(Si tratta di 10 anni e mezzo ovvero 10 anni e 6 mesi, ovviamente non 10 anni e 5 mesi).

BOZZA - DRAFT

1.6 Seconda definizione dei numeri: punti di una retta

Un secondo modo di definire i numeri, dopo quello del paragrafo precedente, è:
identificarli coi punti di una retta:
– sulla retta è fissato un orientamento;
– sulla retta è fissato un punto O che diciamo origine e identifichiamo col numero 0;
– sulla retta è fissato un punto U, che diciamo punto unità e identifichiamo con 1.

Questo è un po' il modo antico di vedere i numeri: a Pitagora non passava neanche per la testa che $\sqrt{2}$ fosse qualcosa come 1.4142... com'è naturale per noi moderni: per lui era proprio la lunghezza della diagonale del quadrato unitario; e π la lunghezza del semicircolo di raggio unitario. Inoltre consideravano una semiretta e allora in pratica solo i numeri positivi, mentre modernamente si considera una retta orientata e quindi corrispondentemente coi numeri sia positivi che negativi, e lo 0.
Una serie di costruzioni geometriche permetteva di dividere in 2 parti uguali i segmenti, e in 3, in 4, eccetera, ottenendo geometricamente le lunghezze ovvero i numeri frazionari.
A quel punto allora, sì, Pitagora trovava un'approssimazione di π, come frazione, $\frac{22}{7}$. Poco pratica operativamente, rispetto alla moderna rappresentazione con le cifre decimali.

BOZZA - DRAFT

1.7 Terza definizione dei numeri: assiomi e costruzione

Qualunque approccio si usi nella trattazione dei numeri, qualcosa deve essere ritenuto noto a priori.

Il numero 2 è quel qualcosa che hanno in comune
l'insieme delle orecchie di ogni gatto normale
l'insieme delle ali di ogni piccione normale
l'insieme degli elettroni di ogni atomo di elio
l'insieme dei protoni di ogni atomo di elio

eccetera eccetera, ogni insieme... di *due* elementi.

Similmente con ogni altro *numero naturale*.

I numeri interi possono essere costruiti in un modo complicato con coppie di numeri naturali, oppure in un modo più triviale... anteponendovi un simbolino, il *meno*. (Sì, funziona).

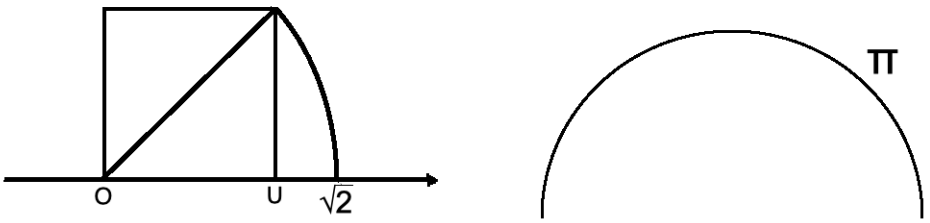
I numeri razionali nascono dalle frazioni, che di fatto sono coppie ordinate di numeri interi, dicendo equivalenti $\frac{2}{3}$ e, per esempio, $\frac{4}{6}$, mediante una formula (formula del prodotto incrociato), e similmente per tutte le altre frazioni.

Poi la costruzione assiomatica dei numeri reali è alquanto complessa.

Importante

Oltre alle scritture decimali, esistono altri 2 importanti modi di definire i numeri:

• Come punti di una retta: valido, ma usato soprattutto nell'antichità:



• Con assiomi: è il metodo scientifico moderno. Per esempio si assiomatizza che

$$\frac{a}{b}+\frac{c}{d}=\frac{a\cdot d+c\cdot b}{b\cdot d}$$

questione che rivedremo.

1.8 Numeri e loro scrittura

Una cosa è un numero, e un'altra è la sua scrittura.

Esempio. 4 rappresentazioni esatte di un numero:

$$\frac{2}{64} \quad = \frac{1}{32} \quad = 0.03125 \quad = 3.125\%$$

e 6 sue rappresentazioni approssimate:

$$\approx 0.0313 \quad \approx 0.031 \quad \approx 0.03$$

$$\approx 3.13\% \quad \approx 3.1\% \quad \approx 3\%$$

Esempio meno banale:

$$\sqrt{9-\sqrt{32}} \qquad 2\sqrt{2}-1 \quad 1.8284...$$

sono lo stesso numero in 3 rappresentazioni diverse. Una semplice calcolatrice porta la prima o la seconda forma nella terza.
(Un difficile calcolo porta la prima forma nella seconda).

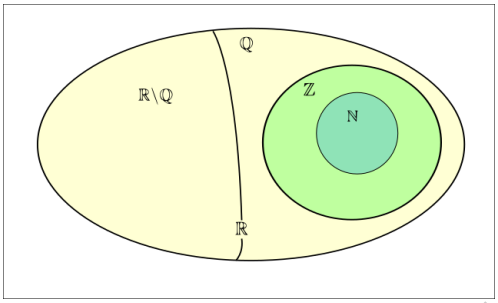
BOZZA - DRAFT

È fondamentale sia saper usare la propria calcolatrice, che applicare le regole algebriche elementari, che sono fra i prerequisiti di questo Corso. Per esempio:

$\frac{2}{64}$ si semplifica in $\frac{1}{32}$. Con la calcolatrice danno 0.03125

Importante

1.9 Gli insiemi numerici



Public Domain Wikimedia

BOZZA - DRAFT

Supponiamo noti dalle scuole:

- $\mathbb{N}$: i numeri naturali: 0, 1, 2...
- $\mathbb{Z}$: i numeri interi: sono di 3 tipi:
  - i numeri naturali diversi da 0, detti *interi positivi*: 1, 2, 3...
  - i loro *opposti*, detti *interi negativi*: -1, -2, -3...
  - lo 0.
- $\mathbb{Q}$: i numeri razionali: sono di 2 tipi, riconoscibili dalla loro scrittura decimale (ma la distinzione è poco significativa perché dipende dalla base 10 scelta per la scrittura dei numeri):
  - i numeri decimali limitati, come $2\,024$ e -3.5
  - i numeri decimali illimitati periodici, come $0.\overline{142\,857} = \frac{1}{7}$, ed eventualmente con *antiperiodo*: $0.08\overline{3} = \frac{1}{12}$

  → entrambi i tipi ammettono una scrittura sotto forma di frazione: $\frac{2024}{1}$, $-\frac{35}{10}$, $\frac{1}{7}$, $\frac{1}{12}$, rispettivamente.
- $\mathbb{R}$: i numeri reali: sono di 2 tipi:
  - i razionali, qua sopra esposti
  - gli irrazionali: la loro scrittura decimale è un numero decimale illimitato non periodico, come

$$\sqrt{2} = 1.4142... \approx 1.41 \qquad \pi \approx 3.14$$

(Esistono poi i numero complessi $\mathbb{C}$ che non tratteremo. E pure altri insiemi "numerici", come i poco noti quaternioni).

1.10 Note sugli errori medici e farmaceutici

Leggiamo sul sito dell'Organizzazione Mondiale della Sanità in http://www.emro.who.int/emhj-volume-17/issue-2/article9.html

> The issue of medication prescribing errors was little discussed until (...) Barker and McConnell in the United States of America (USA) first demonstrated that medication errors occur more frequently than suspected. They estimated a rate of 16 errors per 100 doses (...) errors by pharmacists in dispensing drugs are an important cause of medication error, and many factors have been identified. The reported rate of dispensing errors ranges from 3.8% to 12.4% [Enfasi aggiunta]

SONO PERCENTUALI MOSTRUOSE [N.d.S.], speriamo siano ben state ridotte dagli anni a cui si riferiscono quelle ricerche (1962 e 1991).

Leggiamo riportato su sito governativo statunitense in https://www.ncbi.nlm.nih.gov/pubmed/9559708 (articolo del 1998)

> Forty-two percent of errors were considered to put the patient at risk for a serious or severe preventable adverse outcome. Errors in decimal point placement, mathematical calculation, or expression of dosage regimen accounted for 59.5% of dosage errors. The dosage equation was wrong in 29.5% of dosage errors. [Enfasi aggiunta]

BOZZA - DRAFT

1.11 Errori di calcolo negli articoli scientifici

Ci sono pressoché infiniti errori negli articoli scientifici (e questo è ovvio e fa parte del normale procedere della Scienza) e parecchi sono errori di calcolo (il che è molto meno giustificabile) e verosimilmente moltissimi ci sono nei calcoli statistici che non vengono riportati (è la norma) su dati che pure non vengono riportati (è usuale[7]) e che gli Autori (spesso) non sono nemmeno disponibili a mostrare a Scienziati che glieli chiedessero.

Talvolta gli errori vengono trovati, e talvolta pure corretti in testi intitolati erratum. Per esempio un errore di calcolo (sovradimensionamento di 10 volte) in un articolo[8] scientifico (2016) viene[9] così (2022) corretto:

> There was a decimal placement error in data in the paper on page 566. Instead of "... 10–11 g/(kg·day), and thus an RDA of 12–13 g/(kg·day)", it should instead be "... 1.0–1.1 g/(kg·day), and thus an RDA of 1.2–1.3 g/(kg·day)".

(Enfasi aggiunta)

BOZZA - DRAFT

[7] Scrive (2022, 2025) il Professore Ordinario di Farmacologia di riferimento per lo scrivente (che invece è da trent'anni docente di Statistica) sulla *peer review*:

> NELLA MASSIMA PARTE DEI CASI (quasi sempre) I REVISORI NON HANNO ACCESSO AI DATI GREZZI NE' HANNO ALCUN MODO DI VERIFICARE LA LORO VERIDICITA'. LA REVISIONE CONSISTE IN UNA VALUTAZIONE FORMALE DEL TESTO RISPETTO AGLI STANDARD DELLA SPECIFICA DISCIPLINA, CHE NE VALUTA AL MEGLIO LA VEROSIMIGLIANZA. NIENT'ALTRO.

[8] Phillips SM, Chevalier S, Leidy HJ. Protein "requirements" beyond the RDA: implications for optimizing health. Appl Physiol Nutr Metab. 2016 May;41(5):565-72. doi: 10.1139/apnm-2015-0550. Epub 2016 Feb 9. Erratum in: Appl Physiol Nutr Metab. 2022 May;47(5):615. doi: 10.1139/apnm-2022-0131. PMID: 26960445.

[9] Phillips SM, Chevalier S, Leidy HJ. Correction: Protein "requirements" beyond the RDA: implications for optimizing health. Appl Physiol Nutr Metab. 2022 May;47(5):615. doi: 10.1139/apnm-2022-0131. Epub 2022 Mar 29. Erratum for: Appl Physiol Nutr Metab. 2016 May;41(5):565-72. doi: 10.1139/apnm-2015-0550. PMID: 35349368.

1.12 Numeri romani

Per i numeri interi positivi esiste anche la scrittura in numeri romani.

> Per esempio, arrivando fino al numero VII cioè 7,
> manganese(II) oxide
> manganese(II,III) oxide
> manganese(III) oxide
> manganese(IV) oxide
> manganese(VI) oxide
> manganese(VII) oxide.

(Sarebbe anche bello ricordare il simbolo ss per $\frac{1}{2}$).

E si consideri l'acetato arsenito di rame(II), tossico, un bel pigmento verde usato nell'Ottocento per colorare[10] "carta da parati, sapone, giochi per bambini, vestiti e persino decorazioni per dolci". Tempi oscuri di passata ignoranza.

E si consideri anche l'ossido di titanio (IV) o titanio diossido (il comunissimo colorante E171 di molti farmaci e vernici per muri, ben sospettato di tossicità[11]).

BOZZA - DRAFT

[10] Wikipedia, l'enciclopedia libera, alla voce *Verde di Parigi*, letta l'8 agosto 2024

[11] "Dopo aver condotto una recensione di tutte le evidenze scientifiche disponibili in merito, l'EFSA ha concluso che non sono da escludere timori circa la genotossicità delle particelle di TiO2. Sulla scorta di tali preoccupazioni in termini di genotossicità gli esperti dell'EFSA non ritengono più sicuro il biossido di titanio se usato come additivo alimentare. In altre parole per l'E171 non è possibile stabilire una dose giornaliera accettabile (DGA). (...) Nel loro ruolo di gestori dei rischi la Commissione europea e gli Stati membri rifletteranno ora sul parere fornito dell'EFSA e decideranno in merito a eventuali misure normative o consigli appropriati da dare ai consumatori. (...) L'EFSA sta vietando il biossido di titanio? No. Il compito dell'EFSA si limita alla valutazione dei rischi (...)" Si veda anche, precedente, in https://www.ncbi.nlm.nih.gov/pubmed/28106049 Food-grade TiO2 impairs intestinal and systemic immune homeostasis, initiates preneoplastic lesions and promotes aberrant crypt development in the rat colon. Scientific reports (2017), 7, 40373. Bettini S, Boutet-Robinet E, Cartier C, Coméra C, Gaultier E, Dupuy J, Naud N, Taché S, Grysan P, Reguer S, Thieriet N, Rèfrègiers M, Thiaudiére D, Cravedi JP, Carriére M, Audinot JN, Pierre FH, Guzylack-Piriou L, Houdeau E.

Importante

In questa trattazione elementare ci limitiamo ai primi 12 numeri romani, che si trovano nella numerazione dei capitoli:

1 I ◇ 2 II ◇ 3 III ◇ 4 IV ◇ 5 V ◇ 6 VI ◇
7 VII ◇ 8 VIII ◇ 9 IX ◇ 10 X ◇ 11 XI ◇ 12 XII

Per esempio un cancro al IV stadio è più grave di uno al III stadio.

In Farmacia raramente servono numeri romani maggiori di 12, per esempio:

– nella numerazione dei volumi delle riviste scientifiche

– nella scrittura dell'anno delle riviste scientifiche

– nella numerazione delle pagine fra la seconda di copertina e la pagina 3 di molti libri, per esempio European Pharmacopoeia, eighth edition, volume 1, Council of Europe, Strasbourg, ISBN: 978-92-871-7525-0, online in LINK-> (Attenzione: questa online, del 2013, non è l'ultima versione).

BOZZA DRAFT

Per numeri maggiori di 12 si possono

– studiare le (semplici) regole della numerazione romana

– online su WolframAlpha digitare

`roman number` *numero decimale o romano da convertire*

Sono esistite varie notazioni alternative, in particolare IIII per 4.

1.13 Notazione scientifica dei numeri

Ecco l'età che più spesso hanno gli Studenti all'esame di Matematica, scritta in notazione scientifica:

2E1

ovvero, su testi o software che dispongono di esponenti

$2\text{x}10^{1}$

(Apriamo una parentesi: nelle Scienze Mediche e Farmaceutiche l'età in generale si considera come numero intero di anni compiuti, per cui a 19 anni e 11 mesi si hanno 19 anni).

Su molte calcolatrici e testi tecnico-scientifici, `E` (ma purtroppo anche `e`) indica 10^, con esponente positivo o negativo, per esempio

`6.022E23` = $6.022 \cdot 10^{23} = 602\,200\,000\,000\,000\,000\,000\,000$
`1.38E-23` = $1.38 \cdot 10^{-23} = 0.000\,000\,000\,000\,000\,000\,000\,013\,8$

BOZZA DRAFT

sono Approssimazioni del Numero di Avogadro (della Chimica) e della Costante di Boltzmann (della Termodinamica), private delle unità di misura.
Ci si aspetti di trovare scritto anche `E+23` invece di `E23`, e purtroppo anche `e23`, `e+23`, `e-23`, con possibile equivoco[12] col numero e.

(Ma non si confonda questo simbolo `E` con l'uguale simbolo che alcune calcolatrici danno... in caso di errore di calcolo! Per esempio immettendo `1/0`).

[12] Su testi diversi, $3\,\text{e} - 2$ può indicare $3 \cdot 10^{-2} = 0.03$ oppure $3\,\text{e} - 2 \approx 6.154\,85$. Su WolframAlpha si eviti del tutto la notazione con `E` ed `e` perché entrambi i simboli vengono usati *anche* per il numero e, con facile possibilità di errori; si scriva invece `6.022 10^23`.

Sono scritture equivalenti:

- `2.019E3` e anche `2.019E+3` ma si evitino `2.019e3`, `2.019e+3`
- `2.019` $\cdot$ `10^3` e anche `2.019` $\cdot$ `10^3`
- `2.019*10^3` e anche `2.019*10^3`
- `2.019 10^3` ← scrivete così su WolframAlpha ma non altrove
- `2.019` $\times$ `10^3` e anche `2.019` $\times$ `10^3`

e nell'ultima si badi di non confondere il *per* con una *ics*, specialmente scrivendo a mano un'equazione come

$$2.3 \times 10^3 + 5.4\, x\, 10^2 = 0$$

ESERCIZIO$_\mu$ (R) ≈ Calcolare la media di questi valori, ipoteticamente prodotti da un'apparecchiatura biomedica:

```
1.38E-2
6.01E+2
8.37E-1
```

200.62

oppure

200.617

(Si tratta dei numeri $1.38 \cdot 10^{-2}$, $6.01 \cdot 10^{2}$, $8.37 \cdot 10^{-1}$, da sommare, e poi si divide per 3).

BOZZA - DRAFT

1.14 Varia

Nella pratica, su internet spesso si trovano scritture grezze come `2.019x10-3`, da intendendersi come `2.019E-3` = `2.019` $\times$ 10^{-3}.

Si noti che in certi contesti si usa il simbolo `K` o `k` per indicare le migliaia e cioè esso significa $\cdot 1\,000$, per esempio le 2.5 `K` *views* ossia visualizzazioni della pagina web della farmacia Cuore Integerrimo, ma bisogna essere pronti a trovare con lo stesso significato anche THD, cioè *thousand*.
Si troverà anche MM e perfino mm per "milione".
E quant'altro... – *estote parati*.

1.15 Quando pochi decimali uccisero i "buoni"

☞ Danni costosi o mortali avvengono a causa di approssimazioni numeriche scadenti. Nelle Scienze Applicate si arrotonda e approssima di tutto, sia bene che non bene.

Riportiamo una nota su un'incidente che causò decine di morti: un sistema missilistico computerizzato faceva i calcoli (sostanzialmente) con 7 cifre decimali, che risultarono insufficienti – a causa dell'enorme velocità in gioco – causando una catastrofe ai "buoni" (triste sarcasmo, per riflettere: con più precisione l'ecatombe magari ci sarebbe stata comunque, ma nel campo avverso, ma così va il mondo, fino che non la smettiamo).

BOZZA - DRAFT

On February 25, 1991, during the Gulf War, an American Patriot Missile battery in Dharan, Saudi Arabia, failed to track and intercept an incoming Iraqi Scud missile. The Scud struck an American Army barracks, killing 28 soldiers and injuring around 100 other people. Patriot missileA report of the General Accounting office, GAO/IMTEC-92-26, entitled Patriot Missile Defense: Software Problem Led to System Failure at Dhahran, Saudi Arabia reported on the cause of the failure. It turns out that the cause was an inaccurate calculation of the time since boot due to computer arithmetic errors. Specifically, the time in tenths of second as measured by the system's internal clock was multiplied by 1/10 to produce the time in seconds. This calculation was performed using a 24 bit fixed point register. In particular, the value 1/10, which has a non-terminating binary expansion, was chopped at 24 bits after the radix point. The small chopping error, when multiplied by the large number giving the time in tenths of a second, led to a significant error. Indeed, the Patriot battery had been up around 100 hours, and an easy calculation shows that the resulting time error due to the magnified chopping error was about 0.34 seconds. (The number 1/10 equals 1/24+1/25+1/28+1/29+1/212+1/213+.... In other words, the binary expansion of 1/10 is 0.0001100110011001100110011001100.... Now the 24 bit register in the Patriot stored instead 0.00011001100110011001100 introducing an error of 0.0000000000000000000000011001100... binary, or about 0.000000095 decimal. Multiplying by the number of tenths of a second in 100 hours gives 0.000000095×100×60×60×10=0.34.) A Scud travels at about 1,676 meters per second, and so travels more than half a kilometer in this time. This was far enough that the incoming Scud was outside the "range gate" that the Patriot tracked. Ironically, the fact that the bad time calculation had been improved in some parts of the code, but not all, contributed to the problem, since it meant that the inaccuracies did not cancel. (https://www-users.cse.umn.edu/~arnold/disasters/patriot.html) (E 24 bit corrispondono ad

almeno 7 cifre decimali significative).

1.16 I decimali sono gratuiti – usiamoli

Nella Farmacia che arriva all'utilizzatore finale ben difficilmente si daranno più di 3 cifre significative:

37.8 °C
35%
0,5 g

con rispettivamente 3, 2 e 1 cifra significativa.

Come indicazione generalissima, cercheremo i fare i calcoli con 5 o 6 cifre significative, e di dare i risultati con 3 o 4 cifre significative.

BOZZA - DRAFT

Calcoli successivi possono degradare via via la precisione.

L'operazione che più degrada la precisione è la sottrazione **fra numeri vicini.** Come sono per esempio π e $\sqrt{10}$.

Per esempio

$$\frac{1}{\sqrt{10}-\pi}$$

calcolato come $\frac{1}{3.16-3,14}$ è sbagliato del 3%, sgradevole;
calcolato come $\frac{1}{3.1623-3.1416}$ è sbagliato solo dello 0.07%.

Nei 2 casi gli errori assoluti sono rispettivamente 1.7 e 0.04 circa.

Detto con precisione, l'errore relativo (in valore assoluto) di un numero si conserva abbastanza bene moltiplicando e dividendo con valori esatti, e sommando valori esatti dello stesso segno. Per esempio $\frac{1}{2023\,\pi}$ ha ancora un errore relativo (circa) dello 0.5 per mille quando calcolato, invece che col valore esatto di π, con la sua classica approssimazione 3.14, che ha un errore relativo solo (circa) dello 0.5 per mille, alquanto buona.
Ma la precisione (nel senso dell'errore relativo) si degrada alquanto sottraendo da un valore approssimato un valore esatto, simile e dello stesso segno.

1.17 Cifre significative

In Matematica 0.4 e 0.40 sono uguali a $\frac{2}{5}$, e 0.40 si scrive raramente.

In Farmacia 0.4 e 0.40 possono essere diversi e $\frac{2}{5}$ si scrive raramente.

E non solo in Farmacia: nelle Scienze Applicate
0.4 rappresenta una quantità fra 0.35 e 0.45
0.40 rappresenta una quantità fra 0.395 e 0.405
Si capisce che 0.40 si ritiene conosciuto con maggior precisione.

Per esempio l'articolo scientifico Revised Estimates for the Number of Human and Bacteria Cells in the Body stima che il corpo del "reference man" abbia $3.0 \cdot 10^{13}$ cellule, e si guarda bene dall'impoverire il risultato scrivendo 3 invece di 3.0.

BOZZA - DRAFT

Tutte le cifre diverse da 0 sono significative: 3.14
Gli zeri finali di numeri interi sono (!) un mistero: 8 000 000 000.

2024 ha 4 cifre significative
20.50 ne ha 4
20.05 ne ha 4
2020 ha 3 o 4 cifre significative – ahimè
20.2 ne ha 3
50 ne ha 1 oppure 2 – ohibò
51 ne ha 2
0.51 ne ha 2
0.50 ne ha 2
0.05 ne ha 2
0.510 ne ha 3
0.051 ne ha 2
000.05 ha 1 cifra significativa: 2 zeri aggiunti per impaginare:
234.05 questo numero bene si impagina col soprastante.

1.18 Conteggio delle cifre significative

In Matematica le cifre significative sono argomento poco rilevante, e nelle Scienze Applicate sono avvolte in una nebbiolina definizionale. In questo testo stiamo tentando di dare qualche cenno chiaro, ma non garantiamo che tutti gli Autori facciano ugualmente.

Nota. Le prime 3 cifre significative di e=2.71828... sono 2 7 2.

Leggiamo addirittura degli 0.00% in un articolo[13] scientifico (del gennaio 2021) del prestigioso epidemiologo J.P.A. Joannidis:

> To estimate the infection fatality rate of coronavirus disease 2019 (COVID-19) from seroprevalence data (...) Infection fatality rates ranged from 0.00% to 1.63%, corrected values from 0.00% to 1.54%. Across 51 locations, the median COVID-19 infection fatality rate was 0.27% (corrected 0.23%):

BOZZA - DRAFT

Per sottili variazioni nelle Scienze Applicate si veda questo LINK->

La problematica è complessa perché in generale non si sa la precisione con cui è scritto un dato.

Possono formarsi varie problematiche, per esempio quando si fanno conversioni di unità di misura. Per esempio a proposito del Coronavirus (2020) circola ampiamente il consiglio di tenere una distanza di 1.82 m da ogni altra persona. Addirittura in un articolo scientifico di molto precedente[14] leggiamo, su altro virus:

> Adult volunteers who sat 1.82 m from an infected child did not become infected.

Di fronte a tale precisione viene il sospetto che la rivista scientifica richiedesse – com'è usuale – il Sistema Metrico Decimale, mentre la ricerca era stata fatta misurando 6 piedi, poi trasformati in 1.82 metri – non bene in effetti: corrispondono a circa 1.83 m – con *fittizia* precisione di 3 cifre significative. La "regola dei 6 piedi" si trova per esempio in questo LINK ai presrtigiosi CDC statunitensi, Centers for Disease Control and Prevention: "Close contact is defined as being within 6 feet for at least a period of (...)"

[13] Ioannidis JPA. Infection fatality rate of COVID-19 inferred from seroprevalence data. Bull World Health Organ. 2021 Jan 1;99(1):19-33F. doi: 10.2471/BLT.20.265892. Epub 2020 Oct 14. PMID: 33716331; PMCID: PMC7947934.

[14] Am J Respir Crit Care Med. 2016 Aug 1;194(3):308-16. doi: 10.1164/rccm.201509-1833OC. Evidence of Respiratory Syncytial Virus Spread by Aerosol. Time to Revisit Infection Control Strategies? Kulkarni H, Smith CM, Lee Ddo H, Hirst RA, Easton AJ, O'Callaghan C.
In https://www.atsjournals.org/doi/pdf/10.1164/rccm.201509-1833OC

1.19 Scrittura percentuale dei numeri

Il simbolo % significa $\cdot\frac{1}{100}$ (per un centesimo). Esempio: $5\% = 0.05$

• Dalla scrittura percentuale si ottiene quella decimale togliendo il simbolo % e dividendo per 100, per esempio

75%=0.75 175%=1.75

• Dalla scrittura decimale si ottiene quella percentuale moltiplicando il numero per 100 e posponendovi il simbolo %, per esempio

$$0.75 = 75\% \qquad 1.75 = 175\%$$

Esempi.

$$x \equiv 100\,x\,\% \quad \text{per esempio } \frac{3}{2} = 150\%$$

$$0 = 0\% \quad 1 = 100\% \quad \frac{1}{2} = 0.5 = 50\% \quad \frac{1}{4} = 0.25 = 25\%$$

$$-0.3 = -30\% \quad \frac{4}{3} = 1.\overline{3} = 133.\overline{3}\% \approx 133.3\% \quad 0.05 = 5\%$$

😎 **Nota.** Poichè prevenire è meglio che curare, diciamo esplicitamente che è falso che le percentuali sono numeri fra 0 e 100. Sono numeri qualunque, ovvero, fra $-\infty$ e $+\infty$.

Importante

Talvolta per i numeri puri (cioè privi di unità di misura)
- ∘ e specialmente per i numeri ≥ 0, ma non necessariamente
- ∘ e in particolare per quelli fra 0 e 1, ma non necessariamente,

si trova comodo rappresentare i numeri come *percentuali*:
col simbolo % che significa $\cdot\frac{1}{100}$ (per un centesimo)

Ma si faccia attenzione che nelle Scienze Applicate la scrittura
$a \pm 10\%$
non indica affatto i 2 numeri $a \pm 0.1$ (che sarebbe il significato pedissequo in Matematica) nè l'intervallo $[a-0.1, a+0.1]$ bensì
l'intervallo $[a-0.1\,a, a+0.1\,a]$.

Per esempio $20 \pm 10\%$ indica l'intervallo $[18, 22]$.

E nel caso generale, con una percentuale qualunque fra 0% e 100%,

$$a \pm t\,\% \text{ indica l'intervallo } [(1-t\,\%)\,a, (1+t\,\%)\,a]$$

Naturalmente invece se si tratta di 2 percentuali il significato è quello ovvio:

$$42\% \pm 5\% \text{ indica l'appartenenza all'intervallo } [37\%, 47\%]$$

Non si confondano questi 3 numeri, qua variamente rappresentati:

5%=0.05=1/20 numero classico della Statistica Inferenziale
0.5%=0.005=1/200 o 5 per mille
0.05%=0.0005=1/2000 o 0.5 per mille

Nota. Per le probabilità, che sono numeri fra 0 e 1, la scrittura percentuale è tipica, per esempio

$$P(\text{un dado regolare dà 3}) = \frac{1}{6} \approx 16.7\%.$$

BOZZA - DRAFT

1.20 Scrittura reciproca dei numeri

La scrittura reciproca dei numeri è del tutto indicata per le probabilità minori dell'1%. Per esempio

invece di probabilità 0.0005 ovvero 0.05%

si può dire 1 probabilità su 2000

decisamente più intelligibile.
Similmente 0.0001% corrisponde a 1 (probabilità) su un milione.

Si fa così: divido 0.0001 per 100 e ottengo 0.000001 senza %:

$$0.0001\% = 0.000001$$

e poi calcolo il reciproco con 1: trovando il milione detto.

Tuttavia può presentarsi una problematica praticamente insolubile al livello di questa trattazione elementare:

0.000 7% corrisponderebbe a (circa) 1 (probabilità) su 142 857 ma questo potrebbe far pensare ad una conoscenza del dato originario con una precisione che in realtà non si aveva, proprio come se si fosse avuto 0.000 700 000%, con 6 cifre significative. Diciamo, senza veramente risolvere il problema, "circa 1 (probabilità) su 140mila", come potrebbe trovarsi scritto su un testo divulgativo.

BOZZA - DRAFT

In Farmacia. Se l'utilizzatore finale, il cliente o paziente, legge il contenuto di magnesio per grammo in questi tipici integratori alimentari

Magnesio Citrato 0,160 g

Magnesio Cloruro 0,120 g

Magnesio Gluconato 0,058 g

Magnesio Orotato 0,077 g

potrebbe trovarsi in difficoltà a capire la situazione:

quanto magnesio effettivamente c'è nel prodotto?

Con la scrittura reciproca la situazione può risultare più chiara:

Magnesio Citrato circa 1 parte su 6

Magnesio Cloruro circa 1 parte su 8

Magnesio Gluconato circa 1 parte su 17

Magnesio Orotato circa 1 parte su 13
Questi numeri si calcolano con il passaggio al reciproco: 1:0,160=6,25.

Oppure, invece di dire, molto tecnicamente com'è giusto per un articolo scientifico pubblicato su una rivista scientifica[15] di alto livello,

> Combined, the mRNA vaccines were associated with an excess risk of serious adverse events of special interest of 12.5 per 10,000 vaccinated

che poi magari un profano si spaventa, possiamo dire molto più chiaramente, a livello divulgativo, 1 su 800, così uno sta tranquillo.

Oppure, se una sostanza è pura al 99.5%, il contaminante – chiamiamolo così solo adesso con linguaggio comune seppure tecnicamente impreciso – è lo 0.5%, cioè 1 parte su 200 – decisamente più chiaro.

BOZZA - DRAFT

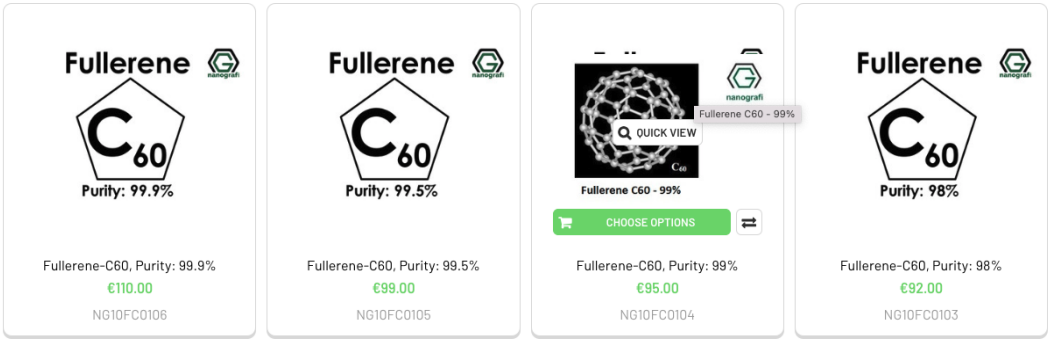
[15] Joseph Fraiman, Juan Erviti, Mark Jones, Sander Greenland, Patrick Whelan, Robert M. Kaplan, Peter Doshi, Serious adverse events of special interest following mRNA COVID-19 vaccination in randomized trials in adults, Vaccine, Volume 40, Issue 40, 2022, Pages 5798-5805, ISSN 0264-410X, https://doi.org/10.1016/j.vaccine.2022.08.036.

1.21 Note sulle percentuali prossime al 100%

Se un tale ha una quota di proprietà di una farmacia del 98%, o del 99%, per lui non cambia quasi niente in termini di guadagni. Per il comproprietario cambia tutto: nel secondo caso guadagna la metà, 1% invece di 2%.

Se una malattia ha *tasso di sopravvivenza* 99% e un'altra 99.9%, la prima ha una *letalità* decupla della seconda.

In questi casi gli arrotondamenti possono essere molto dannosi.



Una nota Azienda vende il fullerene C60 con diversi gradi di purezza, e quello più costoso e puro costa circa il 20% in più del meno costoso: non si paga tanto il prodotto in più che è pochissimo – si passa da 0.98 grammi a 0.999 grammi – ma l'enorme diminuzione della parte che non è garantita fullerene: scende dal 2% allo 0.1%: riduzione di 20 volte.

1.22 Usare numeri approssimati per rendere l'idea

L'Australian Bureau of Statistics afferma che da giugno 2021 a maggio 2022, in Australia ci sono stati 16 467 decessi in più rispetto al normale; 5 619 decessi covid e 10 848 non covid (con tutta la problematica dell'attribuzione ma adesso non ci occupiamo di questo). Essendo

$$10\,848 : 5\,619 \approx 1.93$$

si può senz'altro dire che

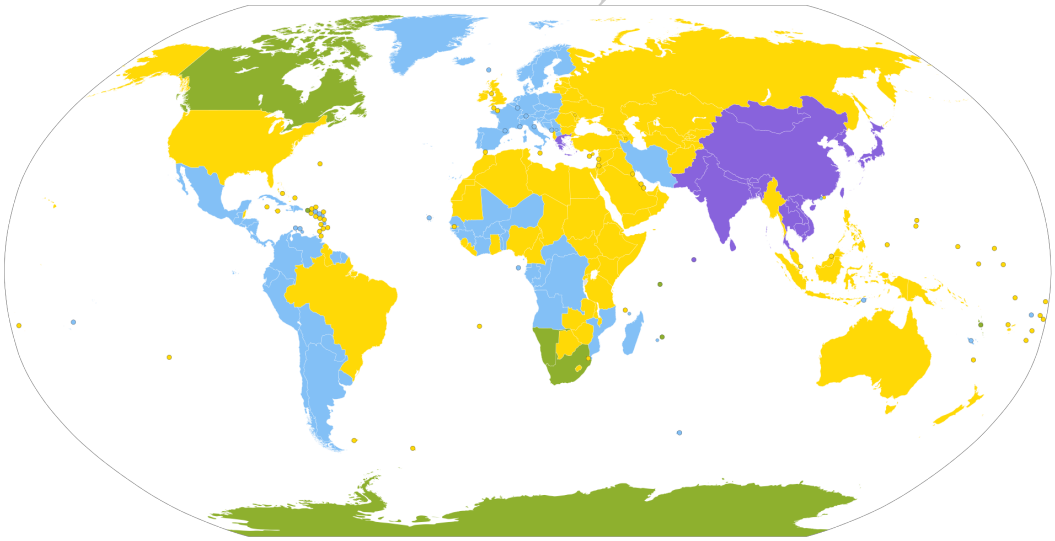
"per ogni decesso Covid
ci sono due decessi inspiegabili non da Covid"

frase decisamente più significativa per l'ascoltatore o lettore medio, che non vorrà ricordare il profluvio di cifre sopra scritte.

1.23 Falsi amici, e pure confusionari

La questione di bilione e billion (10^9 o 10^{12}?), trilione e trillion (10^{12} o 10^{18}?), è talmente ingarbugliata, con variazioni di significato non solo nello spazio ma addirittura nel tempo, che qua oltre ad osservare che attualmente (2021) WordReference online traduce entrambe le parole italiane "bilione" e "trilione" con *trillion* in inglese, ci limitiamo a linkare Wikipedia, l'enciclopedia libera, LINK->, e a... diffidare quando si leggono quei termini.

DRAFT



Nella figura di https://commons.wikimedia.org/wiki/File:Scala_lunga_e_scala_corta.svg, nelle zone in azzurro – in teoria, ma si diffidi – il bilione (nelle sue varianti linguistiche) è 10^{12}, in quelle in giallo 10^9.

Si evitino come la peste le parole "bilione" e "trilione" scrivendo in italiano.

È sconcertante la confusionarietà delle parole "bilione" e "trilione",

attualmente entrambi tradotte con “trillion" da Wordreference: 2 parole da evitare come la peste scrivendo in italiano, però, in Chimica, senza alcuna ambiguità:

ppm = parts per million = parti per milione

ppb = parts per billion = parti per miliardo[16].

BOZZA - DRAFT

[16]Leggiamo su Wikipedia, l’enciclopedia libera: "This notation is not part of the International System of Units (SI) system and its meaning is ambiguous."

1.24 Spirito critico, già coi semplici numeri

Bene Lord Kelvin ha individuato l'*utilità* delle misure numeriche per il progresso delle Scienze, avendo peraltro già Galilei e Pitagora individuato l'universalità – in qualche modo *ontologica* – dei numeri nella realtà. Un'altra affermazione del pur ottimo Lord Kelvin, è che il futuro non è degli aeroplani ma delle mongolfiere: "non ho la minima molecola di fede nella navigazione aerea diversa da quella aerostatica"*Link->*. Non per dire male dell'ottimo scienziato, ma per iniziare questo testo indirizzando lo studente verso lo spirito critico, che già le problematiche della semplice scrittura dei numeri dovrebbero avergli ispirato.

BOZZA - DRAFT

1.25 Altro

Ecco altre cose interessanti, per le quali il lettore interessato troverà molti altri dettagli nella sezione di Complementi.

• Simboli pericolosamente simili, come lo zero e la o; e come la i maiuscola e la elle minuscola di I-doped BiOCl, l'ossicloruro di bismuto drogato con iodio.

• La pericolosità delle ricette scritte male a mano:

> Doctors' sloppy handwriting kills more than 7,000 people annually.

(negli USA, 2006).

• Il fantasmagorico microgrammo, da alcuni indicato con μg, da altri con mcg, e da altri con ug... mentre la predetta Raccomandazione ministeriale esige la parola *microgrammi* completa.

• E altro ancora.

1.26 Bottom line – prima dei Complementi

Le Scienze Applicate e massimamente la Farmacia soffrono della grave ambiguità notazionale della virgola e del punto decimali:

Attenzione a virgola e punto decimali !

I più banali errori con le 4 operazioni e il % nella pratica sono soggetti ad errori potenzialmente catastrofici:

Attenzione ai calcoli, fare e rifare e chiedere aiuto !

Ricordiamo che in Farmacia i decimali sono quasi gratuiti:

Usiamo i decimali !

(Raramente ci servirà qualcosa di più di una semplicissima calcolatrice col display a 6 cifre).

BOZZA - DRAFT

Complementi

1.27 Complementi – Simboli simili

Nella scrittura a mano, ma anche in molti font digitali su carta o su schermo o display (di calcolatrice o macchina diagnostica), è facile confondere – vieppiù se non si vede tanto bene – alcuni simboli, con possibili errori, in particolare:

- numeri:
  ∅ Θ 8 zero barrato, zero barrato, otto
- numeri e lettere:
  0 O zero, o maiuscola; altro font: `0, O`
  `6, b` sei, bi minuscola
  `5 s S` cinque, esse minuscola, esse maiuscola
  `q g 9`, in certi font, con aggravamento per testo sottolineato.
  1 l I uno, elle minuscola, i maiuscola; altro font: `1 l I`

BOZZA DRAFT

Nella predetta nota ministeriale leggiamo

> lasciare uno spazio tra nome e il dosaggio, in modo particolare per quei nomi (anche commerciali se ammessi nella procedura aziendale) che finiscono in l (elle) per evitare interpretazioni errate (ad esempio, Inderal 40 mg al posto di Inderal40mg che potrebbe essere confuso con Inderal 140 mg)

L'uno e la elle sono similissimi fra loro, sì, ma pure alla i maiuscola. Per esempio solo una conoscenza della Chimica permette di riconoscere (ed eventualmente ricopiare correttamente al computer) le lettere i maiuscola ed elle minuscola nel nome chimico[17]

I-doped BiOCl

qua a stento distinguibili ma talvolta proprio uguali sui browser.

[17] Si tratta dell'ossicloruro di bismuto drogato con iodio.

Se il testo ambiguo è in formato digitale, si può risolvere l'enigma copiaincollandolo in https://www.iorl.org/

E se cercherete nel data-base della vostra farmacia i prodotti contenti iilial, prima che arrivino i NAS e li trovino loro, perchè avete letto che lilial è illegale e pensate che la prima lettera sia una i, là scritta maiuscola, magari non ne troverete, perchè era una elle minuscola.

Informatica.

BOZZA DRAFT

In linguaggi di programmazione come C, C++, Java, e molti altri, le variabili sono case-sensitive, il che significa che i minuscola e i maiuscola sono considerati caratteri diversi. Allora potrebbe servirci una i maiuscola (perché stiamo copiando un programma dove c'è tale carattere in un nome di variabile), che non possiamo trascrivere minuscola (sarebbe un'altra variabile), ma adesso che è maiuscola un altro potrebbe leggerla come elle minuscola, e fare errori di programmazione...

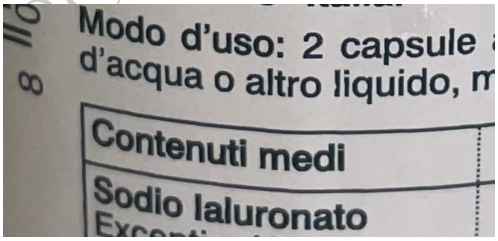
L'Acido Laluronico 😄

Paradossalmente appare che su internet i software automatici,
evidentemente leggendo male con un OCR (o forse male imbeccati da un umano) Acido Ialuronico
come se fosse con la elle minuscola invece che con una i maiuscola, hanno inventato un inesistente

Acido Laluronico:

E qua sorridiamo ma se ci fosse da reperire in magazzino, per un malato grave, un farmaco male archiviato – e ormai praticamente irreperibile in tempi ragionevoli – sarebbe grave.

BOZZA DI



1.28 Complementi – Scrittura a mano

Sull'ulteriore problema della pessima calligrafia dei medici che scrivono le ricette mediche si veda "Poor handwriting remains a significant problem in medicine" del Journal of the Royal Society of Medicine, riportato su sito governativo statunitense: Link->

Citiamo anche dal Time, http://content.time.com/time/health/article/0,8599,1578074,00.html

> Doctors' sloppy handwriting kills more than 7,000 people annually. It's a shocking statistic, and, according to a July 2006 report from the National Academies of Science's Institute of Medicine (IOM), preventable medication mistakes also injure more than 1.5 million Americans annually. Many such errors result from unclear abbreviations and dosage indications and illegible writing

BOZZA - DRAFT

(Enfasi aggiunta).
Con l'informatizzazione del sistema sanitario e farmaceutico è verosimile che quel dato (7000 morti/anno negli USA) diminuisca, e similmente in Italia.
E certo, se dei medici si ostinano a scrivere ,5 per 0,5 "risparmiando" uno zero, anche l'informatizzazione potrebbe non bastare. Tale scrittura è esplicitamente vietata dalla predetta Raccomandazione (settembre 2018) del Mistero della Salute italiano.

In essa leggiamo anche:

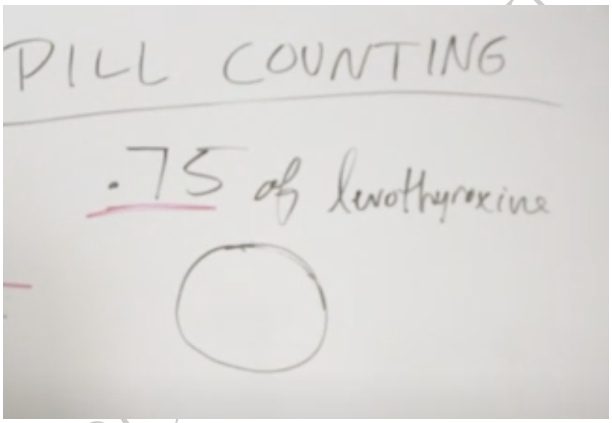
U (significato unità)
può essere erroneamente
interpretato come
"0" (zero)
causando un sovradosaggio
di 10 volte
ad esempio, 4U può essere
interpretato come 40

Se i problemi della scrittura a mano sono in diminuzione, invece con l'incombente anglicizzazione del linguaggio, può presentarsi un nuovo problema, con la lettura al telefono dei nomi dei farmaci. (La parola *ypsilon*, per fare un esempio, appare avere 8 pronunce: Link->).

1.29 Complementi – 7 americano o 1 italiano?

Negli U.S.A. non usano la scrittura italiana, a mano, del numero sette col taglietto orizzontale, e il vostro sette col taglietto qualcuno là potrebbe anche non capire cosa intendete con quel simbolo.
L'immagine è tratta da questo video su calcoli farmaceutici:

https://www.youtube.com/watch?v=pM6NXky4LcE&list=PLTE32pAcpyx_liqfbgbJAzrcBcwuVPwDM&index=7

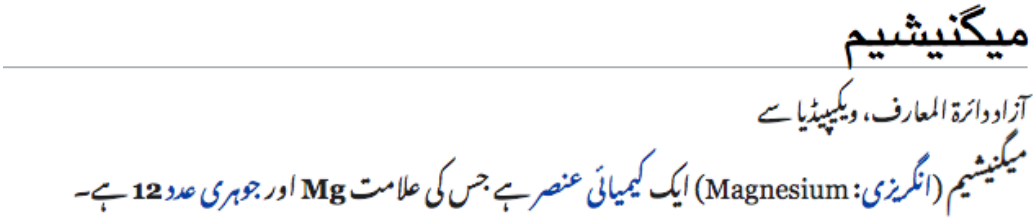


1.30 Complementi – Note finali sui numeri

Praticamente in tutto il mondo per scrivere i numeri si usano le "cifre arabe"

0, 1, 2...9

e questa è una buona notizia. Anche gli arabi, che scrivono da destra a sinistra, scrivono i numeri nel verso "nostro", cioè 243 significa 243 anche in un testo in arabo – buono a sapersi, per interpretare le ricette, per un farmacista che si trovasse a operare in una missione più o meno di pace in una zona di lingua, o anche solo scrittura alfabetica, araba.


میگنیشیم

آزاد دائرۃ المعارف، ویکیپیڈیا سے

میگنیشیم (انگریزی: Magnesium) ایک کیمیائی عنصر ہے جس کی علامت Mg اور جوہری عدد 12 ہے۔

Sreenshot (5 ottobre 2021) da Wikipedia in urdu, lingua (non araba) scritta con alfabeto arabo: il numero che vi appare riportato è proprio dodici, non ventuno, ed è il numero atomico del magnesio, simbolo Mg com'è ben noto.

Superato questo livello minimo di unitarietà, la difformità nella scrittura dei numeri, in questo inizio di XXI secolo, è *sconcertante*.

Un testo italiano sarà sufficientemente scientifico perché il numero 2.718 sia minore di 3, o è di livello così divulgativo che quel numero è... un migliaio di volte più grande?

Veramente, nella pratica, c'è da aspettarsi un po' di tutto: attenzione!

BOZZA - DRAFT

L'intelligenza artificiale online WolframAlpha – di cui in generale si può dire più che bene – usa il simbolo E sia per il numero e, sia per la notazione scientifica dei numeri, con serie possibilità di errori per l'utente:

WolframAplha interpreta l'input 3 E6 come $3 \times 6 \times e$ cioè $18\,e \approx 48.93$

WolframAlpha interpreta l'input 3E6 come $3 \cdot 10^6$ cioè $3\,000\,000$

(WolframAlpha interpreta l'input 2E6 come Groton Municipal Airport; è un eccesso di intelligenza e di interpretazione, si direbbe...)

Non vogliamo qua estendere più di tanto il discorso, ma si noti che se dai *dati grezzi* risulta che un farmaco è stato somministrato dall'11/3 al 12/8, calcoliamo che è stato somministrato

155 giorni se è avvenuto in Italia: dall'11 marzo al 12 agosto (compresi)

36 giorni se è avvenuto negli Stati Uniti: dal 3 novembre all'8 dicembre (compresi).

Altre problematicità sorgono dalle unità di misura, e qua accenniamo soltanto alla seguente.

1.31 Complementi – Il fantasmagorico microgrammo

Da alcuni indicato con μg, da altri con mcg, e da altri con ug... La predetta Raccomandazione ministeriale esige la parola *microgrammi* completa.

Leggiamo su Wikipedia, l'enciclopedia libera:

> The microgram is typically abbreviated "mcg" in pharmaceutical and nutritional supplement labelling, to avoid confusion, since the μ prefix is not always well recognized outside of technical disciplines (...) In the United Kingdom, because serious medication errors have been made from the confusion between milligrams and micrograms when micrograms has been abbreviated, the recommendation given in the Scottish Palliative Care Guidelines is that doses of less than one milligram must be expressed in micrograms and that the word microgram must be written in full [Enfasi aggiunta]

Ma attenzione ai testi antichi, perché, ci avverte la stessa pagina di

BOZZA - DRAFT

Wikipedia,

> The expression "mcg" is also the symbol for an obsolete CGS unit of measure known as the "millicentigram", which is equal to 10 μg

Ancora, la stessa pagina ci avverte in nota che

> The practice of using the abbreviation "mcg" rather than the SI symbol "μg" was formally mandated in the US for medical practitioners in 2004 by the Joint Commission on Accreditation of Healthcare Organizations (JCAHO) in their "Do Not Use" List: Abbreviations, Acronyms, and Symbols because "μg" and "mg" when handwritten can be confused with one another, resulting in a thousandfold overdosing (or underdosing).

BOZZA DRAFT

1.32 Complementi – Numeri e approssimazioni

Nelle Scienze Applicate si preferisce evitare nei risultati finali le frazioni, i decimali periodici, e, π, le radici e le altre funzioni elementari, e tutto si esprime con scritture decimali esatte se si può e merita, o altrimenti approssimate.
Ma in Farmacia è largamente usato il segno ~ per indicare un valore approssimativo invece di ≈.
Inoltre, nelle Scienze Applicate, compresa la Farmacia, spesso si scrive = intendendo ≈, per esempio $1 : 3 = 0{,}333$ o addirittura, con meno buona approssimazione, $1 : 3 = 0{,}33$.

In Matematica	**In questa trattazione.** La scrittura percentuale la daremo solo se ha senso. Esempio: per le probabilità	Nelle Scienze Applicate. La scrittura percentuale si dà solo se ha senso. Esempio: per le probabilità
$= \frac{1}{8} = 0.125$	$= \frac{1}{8} = 0.125 = 12.5\%$	$=0.125$ oppure $= 12.5\%$ ma anche $= 0.1250$ e $= 12.50\%$
$= \frac{1}{3} = 0.\overline{3}$	$= \frac{1}{3} \approx 0.3333 = 33.33\%$	$=0.3333$ oppure $= 33.33\%$
$= \frac{\sqrt{2}}{2} = 0.7071...$	$= \frac{\sqrt{2}}{2} \approx 0.7071 = 70.71\%$	$=0.7071$ oppure $= 70.71\%$
$\frac{6}{\pi} = 1.9098...$	$= \frac{6}{\pi} \approx 0.9099 = 90.99\%$	$=0.9099$ oppure $= 90.99\%$
$\frac{1}{e} = 0.3678...$	$= \frac{1}{e} \approx 0.3679 = 37.79\%$	$= 0.3679$ oppure $= 37.79\%$

Si noti che $= 0.3678...$ è diventato ≈ 0.3679 perché la prima cifra seguente è ≥ 5. (Fatto non conoscibile dalla scrittura $= 0.3678...$). Si noti la scrittura 0.1250 delle Scienze Applicate, nelle quali questa scrittura indica – se usata appropriatamente – che il valore è conosciuto con 4 cifre significative, cioè è compreso fra 0.12495 e 0.12505, mentre 0.125 sarebbe conosciuto con sole 3 cifre significative e allora indicherebbe un numero compreso fra 0.1245 e 0.1255. La questione è sottile, perché richiede di conoscere la precisione dei dati iniziali, che in generale è scarsamente nota. Naturalmente possono considerarsi approssimazioni più precise, o meno.

BOZZA - DRAFT

Nota. Sono assurde scritture come

"mezza tazza di carote contiene 459 mcg di beta-carotene"
"mezza tazza di carote contiene circa 459 mcg di beta-carotene"

perchè l'errore relativo massimo sul numero 459 sarebbe

$$\left|\frac{459 - 458.5}{458.5}\right| \approx 0.1\%$$

un millesimo, precisione impensabile per la "mezza tazza" di carote.

1.33 ESERCIZI SULLA LEZIONE 1

1.33.1 Esercizio risolto a – Globuli rossi

$_{\mu 2021}$ * Prendiamo come valori medi 5 litri di sangue in una persona e 5 milioni di globuli rossi al millimetro cubo. Quanti sono in tutto i globuli rossi in una persona? Si ricordi che 1 litro equivale a un cubo di lato 100 mm. Si esprima il risultato nella notazione scientifica con 2 cifre significative $n.m \times 10^k$ oppure $n.m\,\mathrm{E}\,k$.

SVOLGIMENTO
Ricordando il volume del cubo

$$volume = lato^3$$

per il litro abbiamo l'equivalenza

$$1\ \mathrm{L} = (100\ \mathrm{mm})^3$$

cioè

$$1\ \mathrm{L} = 10^6\ \mathrm{mm}^3$$

allora

$$5\ \mathrm{L} = 5 \times 10^6\ \mathrm{mm}^3$$

e considerando che ci sono 5 milioni di globuli rossi in ogni millimetro cubo

$$NumeroGlobuliRossi = (5 \times 10^6) \times (5 \times 10^6) = 25 \times 10^{12} =$$

$$\boxed{2.5 \times 10^{13}}$$

ovvero

$$\boxed{2.5\,\mathrm{E}\,13}$$

Nota 1. Se volessimo esprimere il risultato a parole, entriamo in un guazzabuglio di equivoci. A rigore sono 25 bilioni, ovvero, in inglese, *25 trillions*, ma di fatto alcuni anche in italiano direbbero "trilioni", con grande confusione; attualmente WordReference online traduce sia "bilione" che "trilione" con *trillion*; in inglese invece *billion* ha l'unico significato di "miliardo". Ma "trilione", per Treccani online, è 10^{18}.
Si evitino come la peste le parole "bilione" e "trilione".
Nota 2. Il numero di tutte le cellule del corpo umano è probabilmente

BOZZA - DRAFT

“di poco" maggiore, secondo alcuni $\approx 3.0 \times 10^{13}$: buona parte delle cellule del corpo umano sono globuli rossi. Si veda https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002533#pbio.1002533.ref001 in cui si legge anche “We estimate the total number of bacteria in the 70 kg "reference man" to be $3.8 \cdot 10^{13}$." Un po’ di più.

1.33.2 Esercizio risolto b – Percentuali capricciose

$_{\mu 2023}$ (R) % Supponiamo che la probabilità di una malattia nelle femmine sia il quadrato della probabilità di quella malattia nei maschi, e che questa sia 8%: qual è la probabilità della malattia nelle femmine?

$$\boxed{0.64\%}$$

(Ovviamente 8% è 0.08 e $0.08^2 = 0.0064$ che è 0.64%. Allora no, non è 64% ☺

Questa più o meno è, semplificando, la situazione del daltonismo, con la percentuale riferita al Nord Europa. Per Genetica – gene recessivo trasportato sul cromosoma X – la probabilità del daltonismo nelle femmine – in una popolazione “stabile" isolata – è il quadrato della probabilità del daltonismo nei maschi

OPPURE,
un modo pessimo di risolvere, è elevare direttamente al quadrato il numero 8, ma allora bisogna elevare al quadrato anche il %, in un modo che non si è mai visto scritto, $(8\%)^2 = 8^2\,\%^2$ e ricordando che % significa esattamente

$$\cdot\frac{1}{100}$$

si conclude come sopra, trovando 64 decimillesimi, 0.0064, ovvero 0.64%).

1.33.3 Aneddoto – Il significato ultimo

WolframAlpha, la più potente intelligenza artificiale matematica gratuita online, alla domanda

`Which is the Answer to the Ultimate Question of Life, the Universe, and Everything?`

risponde... LINK->

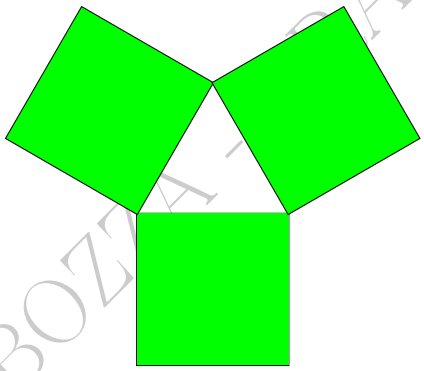
BOZZA - DRAFT

2 Ripasso di Geometria Euclidea Elementare

2.1 Perché questo corso ha 1 ora di Geometria?

Per lo studio della Geometria sono
- motivi formativi/astratti
- motivi pratici.

Problema. In figura si vedono dall'alto 3 cubi uguali appoggiati su un piano orizzontale. Si appoggi un quarto cubo uguale sul buco fra i 3 cubi in modo che un suo vertice vada quanto più in basso possibile. È difficilissimo immaginare la figura costituita dai 4 cubi! Che forma ha vista dall'alto? Cioè, ci chiediamo come si modifica la figura verde qua sotto, quando verrà aggiunto il quarto cubo.



BOZZA - DRAFT

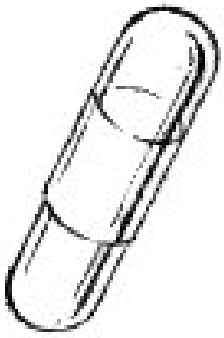
Come mostra il soprastante esempio, la geometria elementare ci fa raggiungere subito i limiti della nostra potenza mentale, facendoci fare un salutare bagno di umiltà.

Questo è un motivo formativo/astratto per lo studio della geometria, e alla fine della Lezione ne vedremo vari altri.

Adesso però vediamo alcuni motivi pratici, direttamente utili per le Scienze della Farmacia.

2.2 Questioni pratiche affrontate dalla Geometria

Impareremo a calcolare il volume utile di una capsula ("pillola") modellizzata con un cilindro e 2 semisfere.



BOZZA - DRAFT

Potremo calcolare anche, fra le altre cose:

– il volume di compresse ellissoidali;

– l'area di cerotti rotondi e rettangolari;

– il volume di un vaso sanguigno, modellizzato con un cilindro;

e, almeno in casi semplici oppure approssimativamente, anche:

– l'area della nostra farmacia

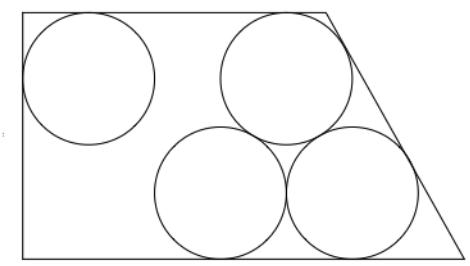
— e il numero di clienti ospitabili, se dipendente dall'area;

– il volume della nostra farmacia e del suo magazzino

— e il numero max di lavoranti là: per legge dipende dal volume;

– il volume di provette, siringhe, matracci, serbatoi...

e tante altre cose.



Problema da semplice a difficile
Supponiamo che per un'epidemia si debba tracciare sul pavimento della nostra bella farmacia circoli di raggio 1 m, uno per cliente ospitabile. Quanti se ne potranno fare, e in quali posizioni?
A seconda della forma dello spazio utile (diciamo: davanti al bancone), il problema può essere da semplice a difficilissimo.
E fa una bella differenza poter ospitare 4 clienti oppure, ottimizzando le posizioni dei circoli, 5 o magari addirittura 6 clienti.

2.3 Geometria, ovunque!

La presenza di forme geometriche nella Scienza e nella Tecnologia è molto pervasiva.
Scrive Galileo Galilei, uno dei fondatori della Scienza moderna:

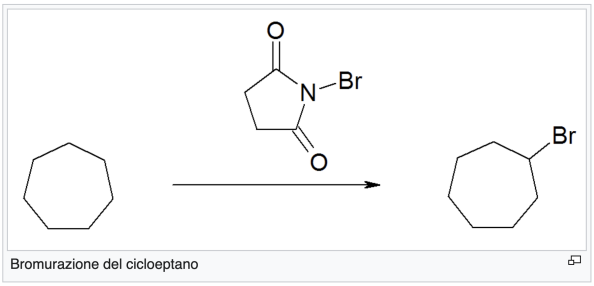
> La filosofia è scritta in questo grandissimo libro che continuamente ci sta aperto innanzi a gli occhi (io dico l'universo), ma non si può intendere se prima non s'impara a intender la lingua, e conoscer i caratteri, ne' quali è scritto. Egli è scritto in lingua matematica, e i caratteri son triangoli, cerchi, ed altre figure geometriche, senza i quali mezi è impossibile a intenderne umanamente parola; senza questi è un aggirarsi vanamente per un oscuro laberinto.
> (https://it.wikiquote.org/wiki/Galileo_Galilei)

BOZZA - DRAFT

Questa profonda osservazione scritta nel XVII secolo si è rivelata vera anche per innumerevoli scoperte scientifiche successive: ovunque nella Scienza ricorrono segmenti, triangoli, quadrati, pentagoni, esagoni, ettagoni, sfere...
In figura la *formula di struttura* del *cicloeptano*, usato nella produzione di farmaci (fotografia da Wikipedia, l'enciclopedia libera), e dell'Aspirina (acido acetilsalicilico), il più antico farmaco (1897) non già esistente in natura (fotografia da WolframAlpha), che tanto buoni risultati sta dando nella terapia del covid-19:

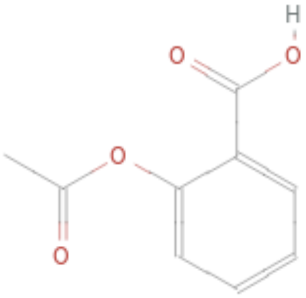
> The results showed that aspirin use was associated with a reduction in COVID-19 mortality (adjusted RR 0.69 (...))[18] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9217117/

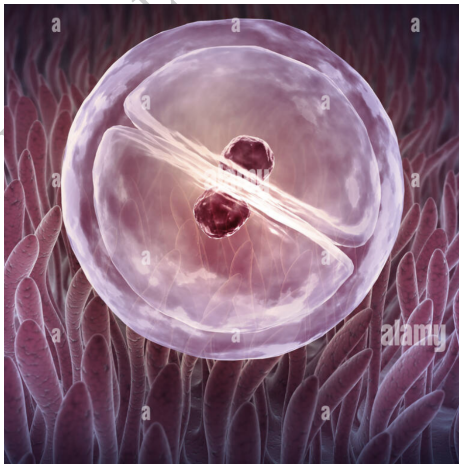
[18] Ma S, Su W, Sun C, Lowe S, Zhou Z, Liu H, Qu G, Xia W, Xie P, Wu B, Gao J, Feng L, Sun Y. Does aspirin have an effect on risk of death in patients with COVID-19? A meta-analysis. Eur J Clin Pharmacol. 2022 Sep;78(9):1403-1420. doi: 10.1007/s00228-022-03356-5. Epub 2022 Jun 22. PMID: 35732963; PMCID: PMC9217117.



Bromurazione del cicloeptano

Sintesi del metilcicloesano



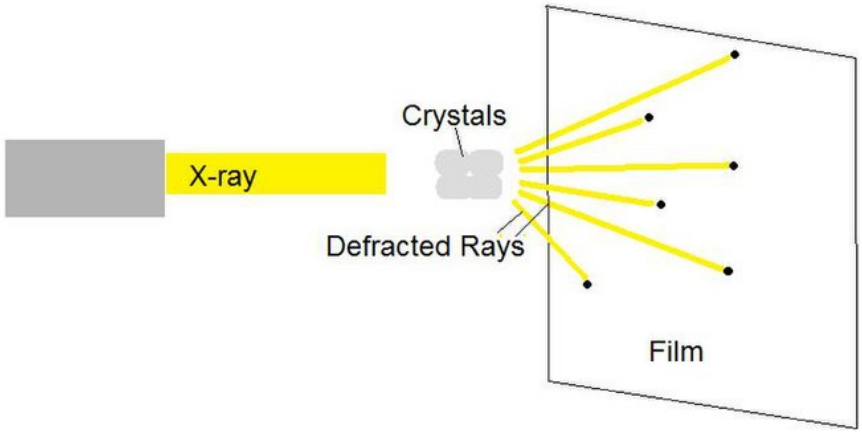


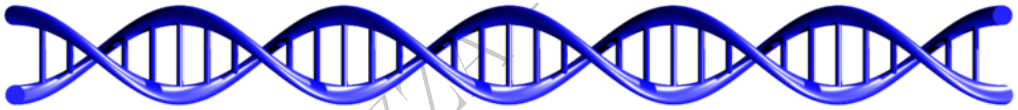
L'approssimativa sfericità della Terra era ben nota al tempo di Galileo Galilei, lo zigote umano è stato scoperto molto dopo.

Si potrebbe obiettare che il corpo della persona adulta non ha più la forma di una sfera, nè è diventato un cilindro o un cono. Ma nonostante la sua forma più complessa, gli si applicano i concetti che vengono elaborati per le figure elementari, come

- distanza fra una sorgente di raggi X e un organo
- area della superficie corporea
- volume del corpo

2.4 Una figura può valere mille parole





2.5 Una figura può valere mille bugie

Ecco le fotografie del mio viaggio in Cechia e rispettivamente in Chad che ho mostrato al mio amico americano (che peraltro non ha idea di dove siano la Cechia e il Chad, come usa colà). Immaginate che idee si è fatto.

Cechia

By Dezidor, in Wikimedia

Chad

By Kayhan ERTUGRUL, in Wikimedia

In effetti la prima è una catapecchia di zingari in Cechia, il secondo è il palazzo del parlamento del Chad.

Questo è uno dei modi in cui vengono condotte le manipolazioni mediatiche su certi argomenti “sensibili", trascegliendo immagini.

Molto più indicativo, proprio potente, è il confronto dei 2 valori di un parametro, fondamentale nelle Scienze Sociali e largamente correlato alla Medicina:

aspettativa di vita alla nascita in Cechia: 78 anni

aspettativa di vita alla nascita in Chad: 55 anni.

Questi 2 numeri sono medie fra milioni di persone, tutti gli abitanti dei 2 stati, invece le immagini sono del tutto fuorvianti fotografie di singole situazioni. (Che è proprio il modo con cui i media spesso mostrano le informazioni, corredandole di “esempi" fotografici... opportunamente scelti, in base agli interessi dei loro referenti).

Ma ncora più straordinariamente informativo di un numero, può essere un grafico, che è una particolare figura geometrica.

2.6 Primi concetti della Geometria Euclidea Elementare

I concetti primitivi di base della Geometria Euclidea sono questi:

il *punto* o spazio 0-dimensionale

la *retta(euclidea)* o spazio 1-dimensionale

il *piano(euclideo)* o spazio 2-dimensionale

la *spazio(euclideo)* o spazio 3-dimensionale

(L'ultimo è lo spazio euclideo propriamente detto ma anche gli altri vengono chiamati spazi euclidei, con la precisazione della dimensione).

Esistono spazi non euclidei, dei quali non ci occuperemo.

È ben possibile – la questione è ampiamente studiata in Fisica – che l'universo non si disponga affatto su uno spazio euclideo tridimensionale. Ma sulla scala umana, in cui avvengono tutte le cose della Farmacia, lo spazio euclideo tridimensionale modellizza benissimo il nostro mondo, in cui siamo immersi, viviamo e operiamo.
In esso possiamo modellizzare capsule e compresse di varie forme, la nostra farmacia e il suo magazzino, i capelli e i vasi sanguigni, e un'infinità di altre cose.

Si deve considerare noto che
– la retta è costituita da punti
– il piano è costituito (in vari modi) da rette
— e quindi da punti
– lo spazio è costituito (in vari modi) da piani
— e quindi da rette
— e quindi da punti.

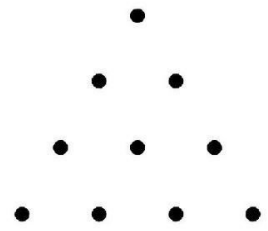
Il punto è il concetto base delle varie Geometrie.

BOZZA - DRAFT

Il punto viene assiomatizzato “infinitamente piccolo" – si potrebbe dare una definizione più rigorosa – e la sua rappresentazione sul piano dovrebbe avere area 0 ma per visualizzarlo lo rappresenteremo più “cicciottello", con un pallino.

Definizione 1. Una certa quantità di punti (in seguito diremo *un insieme* di punti) di uno spazio euclideo si chiama *figura (geometrica, euclidea).*

Per esempio la *tetraktys*, di 10 punti, su cui i seguaci di Pitagora facevano i giuramenti più solenni.

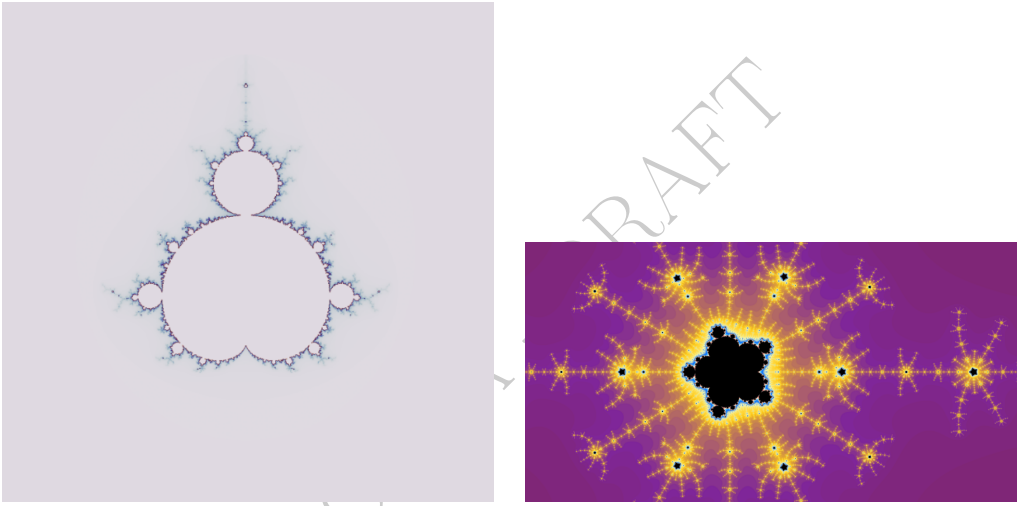


Tuttavia, molti Autori preferiscono quest’altra definizione:
Definizione 1 bis. Una certa quantità di punti (in seguito diremo *un insieme* di punti) di uno spazio euclideo che si dispone con continuità ovvero senza stacchi, si chiama *figura (geometrica, euclidea).*
Con questa definizione modificata, la tetrattide non è una figura.

Le definizioni variano a seconda degli Autori e, in Farmacia, anche degli Enti regolarori.
Le definizioni in Farmacia sono una sorta di game changer. Per esempio se un prodotto verrà classificato vaccino piuttosto che profarmaco – e per i vaccini covid a mRna la questione è sottile – rientrerà in legislazioni e procedure di controllo diverse.
Oltre ad avere prospettive commerciali molto diverse, ovviamente.

Lo studio delle figure è l'oggetto delle Geometrie.

Tuttavia, noi ci occuperemo solo di figure molto "elementari" – seppure non si possa qua definire esattamente e precisamente la questione – non come quelle delle 2 figure qua sotto. Certamente non considereremo figure "frastagliatissime" come i frattali.
(La questione teorica è alquanto complessa: Teoria della misura).

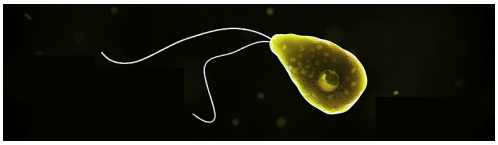


BOZZA DRAFT

A sinistra: il frattale di Mandelbrot. Immagine generata con l'assistenza di ChatGPT 4 di OpenAI, salvo successiva rotazione. A destra: Frattale. Public domain https://commons.wikimedia.org/wiki/File:Mandelbrot_6_Ph%C3%A1o_B%C3%B4ng.png.

2.7 Altri concetti della Geometria Euclidea Elementare

Figure 1- e 2- e 3- dimensionali, e misure 1- e 2- e 3- dimensionali:
Curve – figure 1-dimensionali – fra cui le rette – hanno lunghezza.
Superfici – figure 2-dimensionali – fra cui i piani – hanno area.
Solidi – figure 3-dimensionali – fra cui lo spazio – hanno volume.



Considereremo *concetti primitivi* anche:
la *distanza* di 2 punti;
il *segmento*;
la *retta orientata*, dove i punti si *precedono* o *seguono* fra loro;
il fatto che un punto *origine* divide la *retta* in 2 *semirette*;
il fatto che una retta *origine* divide il piano in 2 *semipiani*;
il fatto che un piano *origine* divide lo spazio in 2 *semispazi*;
per semplicità di trattazione, anche la *curva* e la sua *lunghezza*;
per semplicità di trattazione, anche la *superficie* e la sua *area*.

2.8 Bordo ovvero contorno ovvero frontiera

Supponiamo noto il concetto – la cui definizione sarebbe complicata – di distanza di un punto da una figura.
(In uno spazio euclideo, a qualunque numero di dimensioni).

O.

Si potrebbe definire ma la definizione è complicata. Si immagini la distanza dalla Terra di un atomo in cielo. Oppure di quell'atomo da un'astronave, pur con la sua forma molto più variegata di una sfera. Oppure si ragioni sulla distanza del punto dalla lettera o soprastante.

BOZZA - DRAFT

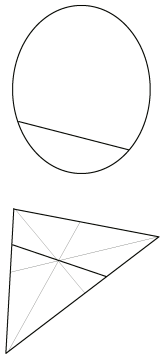
Definizione. Si chiama bordo o contorno o frontiera di una figura di uno spazio euclideo (a qualunque numero di dimensioni) l'insieme dei punti dello spazio che distano 0 dalla figura, e distano 0 dallo spazio privato della figura.

Nei casi più semplici è esattamente quello che nel linguaggio comune viene chiamato bordo della figura. Si pensi per esempio a un "cerchio pieno": la frontiera è il "cerchio vuoto".

Il bordo delle figure del piano euclideo considerate
in Geometria Elementare ha area 0.
Il bordo delle figure dello spazio euclideo considerate
in Geometria Elementare ha volume 0.

2.9 Corde e convessi

Definizione. Un segmento con estremi sul bordo di una figura si dice *corda* della figura.
Per altri Autori la corda è un'altra cosa: le discrepanze definizionali sono comuni in Geometria Euclidea. (Come in Medicina).



Il grandissimo matematico Godfrey Harold Hardy disse, sostanzialmente, a proposito della prima guerra mondiale, che ogni scienza viene apprezzata per quanti morti può fare in una guerra.

Definizione. Se ogni corda di una figura è contenuta nella figura, la figura si dice *convesso*. (Si pensi a un cerchio "pieno", e, come figura non convessa, una stella a 5 punte, anch'essa "piena").
Definizione. Una corda che congiunge un punto medio di un lato di un triangolo col vertice opposto si chiama mediana.
Teorema. Le 3 mediane di un triangolo si incontrano in 1 solo punto, detto baricentro. (Esso modellizza il baricentro della fisica).

La Tomografia Convessa, lontanamente collegata alla Tomografia Medica, si occupa delle lunghezze delle corde delle figure convesse.
Teorema[19]**.** Ogni triangolo è univocamente determinato dalle lunghezze delle sue corde passanti per il baricentro, a meno di una simmetria centrale con centro nel baricentro.

Si potrebbe obiettare che è un risultato inutile. Però:
– Ha bellezza.
– Ancora nessuno è riuscito a utilizzarlo per fare nuove armi, oppure epidemie, magari che colpiscano selettivamente gruppi etnici.

[19] Richard J. Gardner, Alessandro Soranzo, Aljoša Volčič: On the Determination of Star and Convex Bodies by Section Functions. Discret. Comput. Geom. 21(1): 69-85 (1999)

2.10 Note di Geometria Euclidea 1- 2- 3- dimensionale

La Geometria Euclidea Elementare Unidimensionale ovvero 1-dimensionale studia le figure contenute in una retta.
I tipi principali, fra infiniti, sono i seguenti.

∘ Il segmento ha lunghezza pari alla distanza degli estremi.
• Il punto ha lunghezza 0.
⋄ La retta e la semiretta hanno lunghezza $+\infty$.
⋆ figure terribilmente "frastagliate" potrebbero non avere una lunghezza e non verranno considerate nella Geometria Elementare (al livello di questo testo).

• Nel piano il punto, il segmento, la retta e tutte le altre figure (della Geometria Elementare) contenute in una retta hanno area 0.
• Nel piano ogni curva (della Geometria Elementare) ha area 0.
⋄ Il piano e il semipiano hanno area $+\infty$.
⋆ figure terribilmente "frastagliate" potrebbero non avere un'area e non verranno considerate nella Geometria Elementare (al livello di questo testo).

BOZZA - DRAFT

• Nello spazio ogni curva (della Geometria Elementare) ha volume 0.
• Nello spazio ogni superficie (della Geometria Elementare) ha volume 0.
• Nello spazio il punto, il segmento, la retta, il piano e tutte le figure contenute in un piano hanno volume 0.
⋄ Lo spazio e il semispazio hanno volume $+\infty$.
⋆ figure terribilmente "frastagliate" potrebbero non avere un volume e non verranno considerate nella Geometria Elementare (al livello di questo testo).

2.11 Decomposizione di una figura piana in più figure

Se nel piano euclideo una figura F si decompone, salvo al più la sovrapposizione sui bordi, in più figure F_1,...,F_n di aree a_1,...,a_n, allora l'area di F è $a_1 + ... + a_n$. Si pensi alle aree dei 4 triangoli nella figura

Similmente avviene coi volumi nello spazio. Il paradosso di Banach-Tarski non riguarda le figure semplici a cui ci limitiamo ma proprio quelle super-frastagliatissime che abbiamo deciso di ignorare. ☺

DRAFT

Paradosso 1. Decomponendo la prima figura con 14 ballerine e ricomponendola si ottengono 15 ballerine.

Paradosso 2. LINK

2.12 Rette parallele e incidenti e sghembe, ed angoli

Due semirette diverse con stessa origine P dividono il piano in 2 *angoli* di *vertice* P. Se essi sono uguali si dicono *piatti*. Se uno è "minore" dell'altro quello minore si dice *convesso* e l'altro *concavo*. Due rette del piano con intersezione vuota le diremo *parallele*. Rette non complanari non incidenti si dicono *sghembe*. Se l'intersezione ha esattamente 1 punto si dicono *incidenti*. Se l'intersezione coincide con le rette stesse, si tratta di 2 rette *coincidenti* (talvolta dette *parallele coincidenti*): in effetti è 1 sola retta considerata 2 volte. Due rette incidenti possono formare 4 angoli uguali che si dicono *retti*, e in quel caso le rette si dicono *ortogonali* o *perpendicolari*. Angoli "minori" di un angolo retto si chiamano *acuti* e angoli "maggiori" di un angolo retto e "minori" di un angolo piatto si dicono *ottusi*.
Tutte le predette cose tranne le rette sghembe, sono definite o nel piano euclideo, oppure in un piano dello spazio euclideo.

BOZZA - DRAFT

2.13 Geometria Elementare Bidimensionale

La Geometria Elementare Bidimensionale ovvero 2-dimensionale studia le figure contenute in un piano (euclideo).

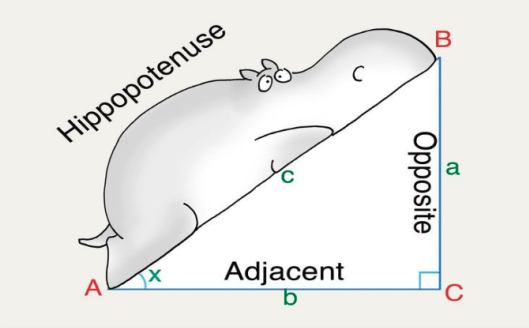
Alcune cose che supponiamo note dagli studi precedenti

poligono
lato del poligono, perimetro del poligono, area del poligono
triangolo: il poligono con 3 lati
lato e base e altezza del triangolo
triangolo rettangolo, isoscele ed equilatero
rettangolo, il poligono con 4 lati e 4 angoli retti
base e altezza del rettangolo
quadrato: il rettangolo con 4 lati uguali; lato del quadrato

- Il triangolo di base b e altezza h ha area $\frac{1}{2}\,b\,h$.
- Il rettangolo di lati a e b ha perimetro $2\,a + 2\,b$ e area $a \cdot b$.
- **Allora, il quadrato di lato a ha perimetro $4\,a$ e area a^2.**

2.14 Teorema di Pitagora

Un triangolo con un angolo retto si chiama triangolo rettangolo; si chiama ipotenusa il suo lato maggiore; cateti gli altri.



Dette nell'ordine i, c_1 e c_2 le lunghezze di quei lati, vale

$$\boxed{i^2 = c_1^2 + c_2^2} \tag{1}$$

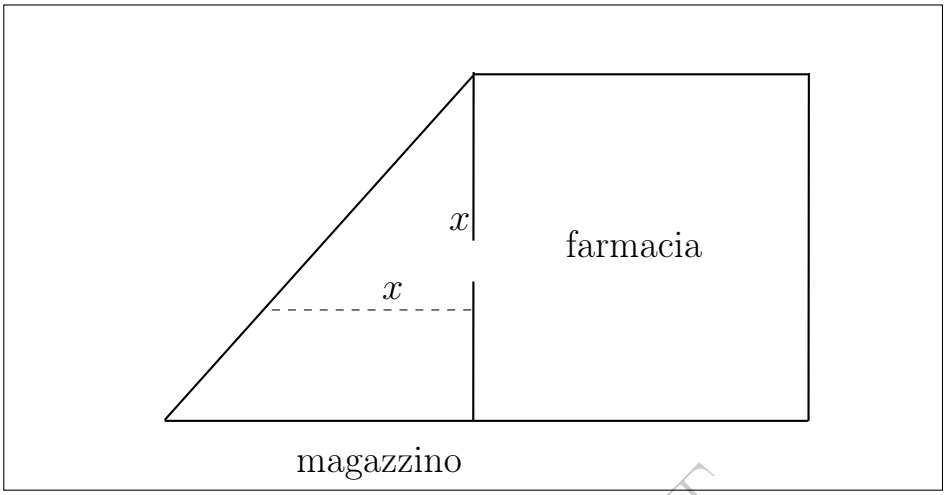
Ecco 2 conseguenze del Teorema di Pitagora:

$$\boxed{\text{la diagonale del quadrato di lato } a \text{ ha lunghezza } \sqrt{2}\,a} \tag{2}$$

$$\boxed{\text{il triangolo equilatero di lato } a \text{ ha altezza } \sqrt{3}\,a}$$

ESERCIZIO$_\mu$ $\approx$ Supponiamo che la Protezione Civile ci imponga di svuotare metà del magazzino della nostra farmacia per urgenti necessità relativamente a una epidemia o terremoto. Il nostro magazzino ha la forma (cioè, in pianta) della metà di un quadrato che si ottiene suddividendolo con una diagonale. (Si tratta di un triangolo rettangolo isoscele). Il lato del quadrato misura 1 (per esempio 1 dm, decametro, 10 metri, ma non ci occuperemo delle unità di misura). Ora noi vogliamo suddividere (per esempio con un nastro adesivo sul pavimento) quel triangolo in 2 parti di ugual area; per esempio con una linea parallela a un cateto, ottenendo un trapezio e un triangolo. Trovare l'altezza x del triangolo. (Ovviamente il nuovo triangolo ha non solo altezza x ma anche, per considerazioni di geometria elementare, base x).

BOZZA - DRAFT



SVOLGIMENTO

L'area del quadrato di lato 1 è 1^2 cioè 1, e allora il magazzino ha area

$$A_{triangolo\,grande} = \frac{1}{2} \qquad (*)$$

Il triangolo nuovo ha sia altezza che base lunghe x e allora l'area è

$$A_{triangolo\,piccolo} = \frac{1}{2} \cdot x \cdot x \qquad (**)$$

Il valore in $(**)$ dev'essere metà del valore in $(*)$:

$$\frac{1}{2} \cdot x^2 = \frac{1}{2} \cdot \frac{1}{2} \qquad \Big/ \cdot 2$$

$$x^2 = \frac{1}{2} \qquad \Big/ \sqrt{\ } \quad (x > 0)$$

$$x = \sqrt{\frac{1}{2}}$$

e con la calcolatrice

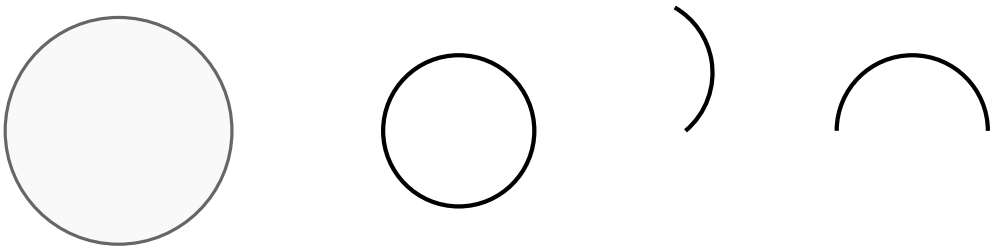
$$\boxed{\approx 0.707}$$

(Per un lato iniziale di misura ipotetica 1 decametro ovvero 10 metri, si tratta di 7 metri e 7 centimetri).

BOZZA - DRAFT

2.15 Cerchio, circolo, circonferenza

In questo paragrafo semplicemente definiremo



cerchio, circolo, circonferenza,
centro, raggio, diametro,
arco di circolo, semicircolo, semicerchio,
notando grandi discrepanze definizionali.

BOZZA - DRAFT

In quest'ambito la nomenclatura è molto ambigua e varia fra i diversi Autori, usandosi in italiano
almeno 6 termini
per 3 sole cose,
e un'altra varietà di termini in inglese scientifico:
cerchio
circolo
circonferenza
disco
palla (2-dimensionale)
lunghezza della circonferenza.

In base alla nomenclatura che adottiamo in questo testo, possiamo immaginare le 3 cose così:

cerchio — affine a un disco di vinile — è una figura
circolo — affine a una fede nuziale — è una figura
circonferenza — lunghezza del circolo — in Matematica è un numero mentre in Fisica ha un'unità di misura ed è una grandezza fisica.

Definizioni. Dati un punto C e numero $r > 0$, l'insieme dei punti che hanno distanza r da C si chiama *circolo* (per altri Autori *circonferenza*) di *centro* C e *raggio* r. È una curva dotata di *lunghezza*, detta *circonferenza*.
Si noti che:
– il circolo appare "vuoto", e il centro non gli appartiene;
– il cerchio appare "pieno", e il centro gli appartiene:
in questa trattazione chiameremo *cerchio* l'insieme dei punti che hanno distanza $\leq r$ da C, *centro del cerchio* (oltre che del circolo).
Ogni segmento nel cerchio lungo $2\,r$ si chiama *diametro*.
Teorema. Ogni diametro è una corda e contiene il centro.

Un diametro divide il cerchio in 2 *semicerchi* e il circolo in 2 *semicircoli*.

Ogni segmento che ha un estremo nel centro e uno sul circolo si chiama *raggio*.

Si noti che le parole *raggio* e *diametro* indicano sia dei segmenti che le loro lunghezze, spesso denotate con r e d.

Ogni corda di un cerchio crea in modo ovvio e ben noto 2 *archi di circolo*.

L'arco di circolo è una curva dotata di *lunghezza*.

BOZZA - DRAFT

2.16 Pi greco, lunghezza del circolo, area del cerchio

Definizione e Teorema.
In tutti i cerchi è sempre uguale il rapporto fra circonferenza e diametro, e il rapporto[20] si chiama *pi greco*:

$$\boxed{\pi := \frac{circonferenza}{diametro}} \tag{3}$$

Da quest'ultima indicando con d il diametro

$$circonferenza = \pi\, d$$

e da quest'ultima indicando con r il raggio (essendo $d = 2\,r$)

$$\boxed{circonferenza = 2\,\pi\, r} \tag{4}$$

La scrittura decimale di π ha infiniti decimali che non si ripetono periodicamente

BOZZA - DRAFT

[20] Equivalentemente, dividendo per 2 il numeratore e il denominatore della (3), si ha che in tutti i cerchi è sempre uguale il rapporto fra semicirconferenza e raggio:

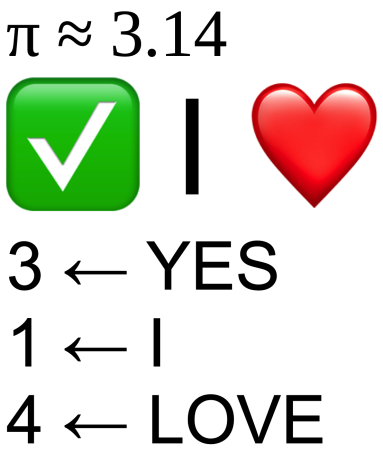
$$\pi := \frac{semicirconferenza}{raggio}$$

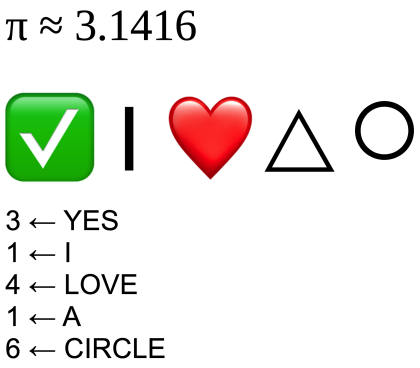
BOZZA - DRAFT

e vale l'approssimazione (con 3 cifre significative, o con 2 decimali esatti)

$$\boxed{\pi \approx 3.14} \tag{5}$$

(e infinite altre approssimazioni con 3 o 4 o più cifre decimali esatte, approssimazioni che non ci proponiamo di imparare a memoria: ≈ 3.14 ci basta).





Mnemonico per 3.14:
contare le lettere delle 3 parole.

Mnemonico per 3.1416:
contare le lettere delle 5 parole.

Per un cerchio di raggio r

$$\boxed{\text{area} = \pi\, r^2}$$

(Da cui segue che un cerchio occupa circa il 78% di ogni suo quadrato circoscritto, e precisamente $\frac{100\,\pi}{4}\%$ ovvero $25\,\pi\,\%$).

2.17 Sull'orlo dell'abisso in compagnia di pi greco

Pi greco è ancora al centro di questioni scientifiche irrisolte.

Per esempio non si riesce ancora a dimostrare se tutte le cifre da 0 a 9 ricorrano infinite volte oppure no.

Moltissimi matematici hanno prodotto approssimazioni di π nel corso dei millenni.

Wikipedia italiana dà un'approssimazione di π con le prime 100 cifre decimali:

3,14159 26535 89793 23846 26433 83279 50288 41971 69399 37510 58209 74944 59230 78164 06286 20899 86280 34825 34211 70679

BOZZA - DRAFT

(scritto con la virgola decimale, e spazietto ogni 5 cifre) e ci racconta altre cose interessanti:

Verso il XX sec. a.C. Babilonesi usavano l'approssimazione $\frac{25}{8}$, cioè 3.125. Archimede nel III sec. a.C. dà l'approssimazione $\frac{211\,875}{67\,441}$, cioè 3.14163...

> Il matematico tedesco Ludolph van Ceulen (1600 circa) calcolò i primi 35 decimali. Era così fiero del suo risultato che lo fece scrivere sulla sua lapide. (Wikipedia, l'enciclopedia libera)

Nel 2020 sono state calcolate 50 000 miliardi di cifre decimali.

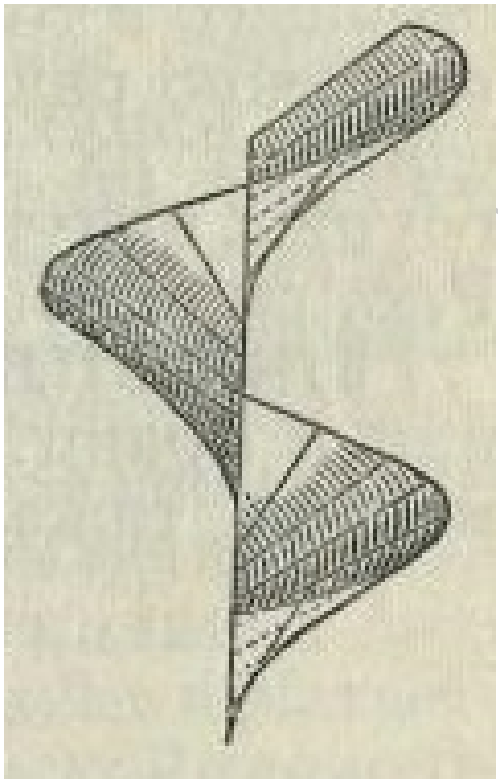
BOZZA - DRAFT

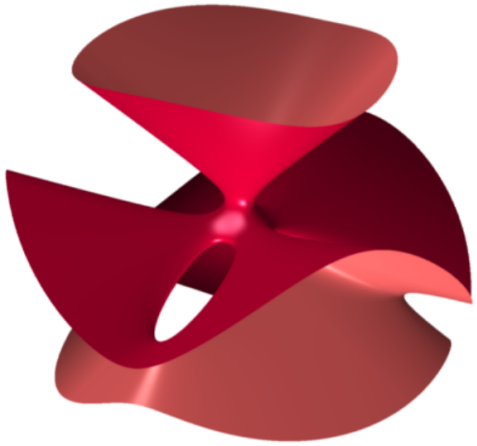
2.18 Geometria Elementare Tridimensionale

La Geometria Elementare Tridimensionale ovvero 3-dimensionale studia le figure contenute nello spazio (euclideo).

Ci sono fenomeni ancora più controintuitivi che in 2 dimensioni. Per esempio la superficie di Clebsh qua sotto a destra (parzialmente, ovvio) rappresentata contiene 27 rette, nostante la sua apparenza alquanto ondulata. È una superficie del tipo di quelle (superfici algebriche) indagate da Ugo Morin a cui è dedicata un'aula a Trieste (dove nacque nel 1901) nell'Università per la quale questo libro di Matematica viene scritto.

- DRAFT





A sinistra: Elicoide. Public domain (?) image in https://commons.wikimedia.org/wiki/File:%D0%91%D0%A1%D0%AD1._%D0%93%D0%B5%D0%BB%D0%B8%D0%BA%D0%BE%D0%B8%D0%B4.jpg. A destra: Superficie di Clebsch. By Fly by Night in https://commons.wikimedia.org/wiki/File:Clebsch_Cubic.png.

◦ **Il parallelepipedo di lati a, b, c ha volume $a\,b\,c$:**

$$V_{\text{parallelepipedo}} = a \cdot b \cdot c \tag{6}$$

e allora il cubo di lato x ha volume x^3.

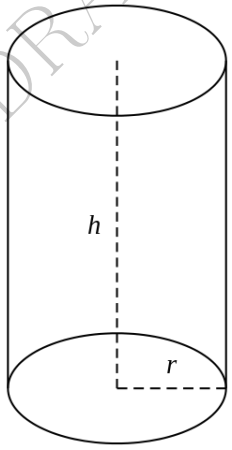
◦ **La sfera di raggio r ha volume**

$$V_{\text{sfera}} = \tfrac{4}{3}\,\pi\, r^3 \tag{7}$$

come da poesia in figura; risultato trovato da Archimede.

- DRAFT

Poesia Futurista Sferica
(Pseudomarinetti)



Cilindro

◦ **Il cilindro ha volume** $area_di_base \cdot altezza$ **ovvero**

$$V_{\text{cilindro}} = \pi\, r^2\, h \tag{8}$$

e area laterale $2\pi\, r\, h$.

Per *cilindro*, in geometria elementare si intende quello del linguaggio comune, come nella figura; più tecnicamente potrebbe essere chiamato *cilindro circolare retto troncato*, e potrebbero considerarsi anche cilindri non retti o non circolari o non troncati.

2.19 Alcune cose supposte note dagli studi precedenti

Una figura si dice *limitata* se esiste un cerchio che la contiene (per esempio i segmenti), altrimenti *illimitata* (per esempio le rette).
L'unione di 2 segmenti AB e BC (da altri denotati $[AB]$ e $[BC]$) non *allineati*, con un estremo B in comune si chiama *linea spezzata* di 2 *lati* AB e BC e *vertice* B, e in modo analogo è definita quella di più *lati*. Non facciamo i dettagli, come pure per le seguenti intuitive denominazioni:

linea spezzata *chiusa*, o altrimenti *aperta*;
linea spezzata *intrecciata*, o altrimenti non intrecciata;
lati *consecutivi*;
lunghezza della linea spezzata.

BOZZA - DRAFT

Supponiamo noto il fatto (non banale da dimostrare) che una linea spezzata chiusa non intrecciata divide il piano in 2 parti. Quella limitata si chiama *poligono (semplice)* (o *non intrecciato*), coi lati e i vertici della linea spezzata, e *perimetro* la sua lunghezza.
Il poligono si dice

equilatero se ha i lati uguali
equiangolo se ha gli angoli uguali
regolare se è equilatero ed equiangolo.

◇ I poligoni con 3 lati si chiamano *triangoli*.

◇ I poligoni con 4 lati si chiamano *quadrilateri*.

Fra i quadrilateri ci sono i *parallelogrammi*, che hanno i lati a 2 a 2 paralleli, e fra essi ci sono

* il *rombo*, che è equilatero,
* il *rettangolo*, che è equiangolo (angoli tutti retti)
* il *quadrato*, che è equilatero equiangolo.

◇ I poligoni con 5 lati si chiamano *pentagoni*.
◇ I poligoni con 6 lati si chiamano *esagoni*.

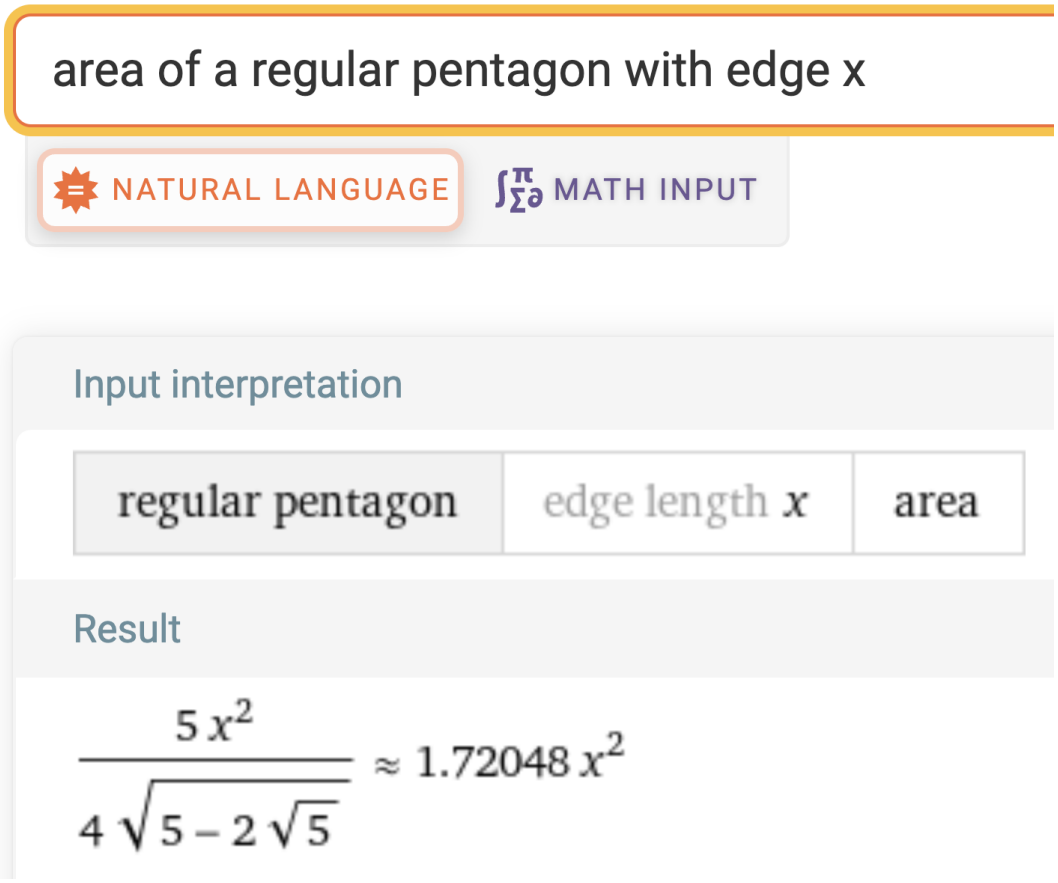
2.20 Con WolframAlpha

Mai sufficientemente bene sarà detto di WolframAlpha, la più potente intelligenza artificiale matematica attualmente disponibile gratuitamente online.

Esempio. Scriviamo

```
area of a regular pentagon with edge x
```

e ci risponde



2.21 Le simmetrie, la chiralità e i farmaci

Ben note al lettore sono le simmetrie, rispetto a un punto, rispetto a una retta, rispetto a un piano. L'ultimo caso corrisponde allo specchio. Leggiamo su Science[21]

> All of life exists on just one side of a mirror. To put it more technically, the biomolecules that comprise living things – DNA, RNA, and proteins – are all "chiral." Their building blocks have two possible mirror-image shapes, but in every case, life chooses just one. At least so far.
>
> Today in Science, researchers report they've made strides toward exploring the other side of the mirror. They re-engineered a workhorse enzyme that synthesizes RNA so it makes the mirror-image form. They then used that enzyme to construct all the RNAs needed to make a ribosome, the cellular machine responsible for constructing proteins. Other components still need to be added, but once completed, a mirror-image ribosome might be able to churn out proteins that could serve as novel drugs and diagnostics and cant readily be broken down in the body.

BOZZA DRAFT

Tragica vicenda della Farmacia, è quella del/la Talidomide, correlata alla chiralità delle molecole. Leggiamo[22] dal CNR, consiglio Nazionale delle Ricerche, il principale ente pubblico di ricerca italiano:

> La talidomide, farmaco ampiamente utilizzato in passato per curare l'influenza, come sedativo e per alleviare i sintomi della nausea da gravidanza, viene spesso menzionato come caso studio per introdurre i concetti di isomeria e chiralità delle molecole. Considerato inizialmente privo di tossicità, in seguito alla sua ampia diffusione

[21] 'Mirror-image' protein factories could one day make durable drugs the body can't break down (27 OCT 2022)

[22] https://almanacco.cnr.it/articolo/11422/simmetria-e-asimmetria-delle-molecole, letto il 7 settembre 2024

> provocò il verificarsi di gravi effetti collaterali, tra questi la nascita di bambini focomelici o con gravissime malformazioni agli arti, figli delle donne che avevano assunto il farmaco durante la gravidanza. La molecola, infatti, esiste in due forme aventi formula identica e con gli atomi legati fra loro nello stesso ordine, l'unica differenza è l'arrangiamento degli atomi nello spazio, per cui una forma è l'immagine speculare dell'altra, ma le due forme non sono sovrapponibili, analogamente a quanto osserviamo confrontando la mano destra con la mano sinistra.
>
> Una forma della talidomide è quella curativa, mentre nell'altra la molecola è tossica teratogena. Questa asimmetria viene definita come la chiralità di una molecola: una molecola è chirale quando è presente in due forme, gli enantiomeri, ciascuno immagine speculare dell'altro, ma non sovrapponibili. Le molecole chirali spesso sono presenti come miscela dei due enantiomeri, dette miscele raceme. Nel caso specifico della talidomide, isolarne la forma "innocua" dal racemo è risultato inefficace, perché il nostro organismo è in grado di convertire un enantiomero nell'altro.

BOZZA - DRAFT

Meno tragicamente, leggiamo sul sito dell'UTIFAR, Unione Tecnica Italiana Farmacisti:

> L'ibuprofene (...) viene attualmente commercializzato in forma racemica, ma i dati di letteratura suggeriscono come soltanto l'enantiomero S sia responsabile dell'azione farmacologica, mentre l'enantiomero R appare privo di tale caratteristica e si accumula nei grassi sotto forma di estere del glicerolo. Alla luce di ciò, risulta importante sviluppare strategie sintetiche che permettano di ottenere solo l'enantiomero S.

2.22 Perchè si studia Geometria? Da millenni?

Alcune cose della geometria sono direttamente utili in Farmacia, talvolta attraverso la Chimica, la Fisica, la Fisiologia.

Ma il motivo per cui da millenni nelle società civilizzate tutti studiano la Geometria Euclidea, anche i prìncipi destinati a diventare re, è molto più profondo.

La Geometria Euclidea è la Scienza razionale per eccellenza, con un valore didattico immenso, sperabilmente formativo per la persona. Segnaliamo alcuni aspetti, tutti poi ricorrenti nella Farmacia:

- la rigorosa precisione delle definizioni
- la ferrea logica delle argomentazioni
- i risultati controintuitivi
- le illusioni percettive
- capire i limiti dell'intelligenza.

Nota sulle definizioni. Le Aziende farmaceutiche devono per legge attenersi rigorosamente a un'infinità di definizioni.
Ma anche per altri motivi ci tengono molto ai nomi e alle questioni definizionali. Si pensi a come una nota azienda ha capitalizzato l'enorme fiducia acquisita presso i Consumatori col Fluimucil, vendendo adesso il Fluimucil Influenza e Raffreddore.
Che però non contiene il principio attivo del Fuimucil.
($ia onore al merito).

BOZZA - DRAFT

Nota sulla ferrea logica. La rigorosa precisione e la ferrea logica, col loro richiedere di "rompersi la testa" su ogni rigo di testo, sono l'opposto della moderna cultura dei 140 caratteri dei tweet, degli slogan, dei *non sequitur* cioè gli errori di ragionamento, che hanno dato immensa prova di forza e pervasività anche durante la pandemia – e adesso (2022-2024) con la guerra. Slogan che sono molto più connaturali all'essere umano della sottil logica – in qualche modo "giustamente": *i lupi sono cattivi* avrà salvato molti bambini in passato – e sempre hanno permesso e adesso permettono facilmente di trascinare le masse

alla guerra, con gli slogan invece della razionalità.

Complementi

2.23 Complementi – Congruenza

Se per 2 figure F ed F' esiste una funzione biiettiva $f : F \to F'$ che conserva le distanze di tutte le coppie di punti corrispondenti

$$\overline{f(P)\,f(Q)} = \overline{P\,Q} \qquad \forall\, P, Q \in F$$

le 2 figure si dicono *congruenti* e si possono immaginare come *sovrapponibili*; ma si faccia attenzione che l'azione della sovrapposizione può necessitare di passare momentaneamente nello *spazio* in cui è contenuto il piano – può in effetti non bastare un movimento di strisciamento sul piano – come per i grafemi d e b. Tutte le rette sono congruenti, e anche le semirette e i semipiani.
Se F ed F' sono il piano euclideo stesso, f si chiama *isometria*.

BOZZA DRAFT

2.24 Complementi – Similitudine

Se per 2 figure F ed F' esistono un numero $\alpha > 0$ e una funzione biiettiva $g : F \to F'$ che moltiplica per α le distanze di tutte le coppie di punti corrispondenti ovvero precisamente

$$\overline{f(P)\,f(Q)} = \alpha \cdot \overline{P\,Q} \qquad \forall\, P, Q \in F$$

le 2 figure si dicono *simili*. In pratica si tratta di ingrandimenti o rimpicciolimenti.
Si dimostra che se un triangolo ha gli angoli uguali a quelli di un altro, essi sono simili. Allora dette a, b, c le misure dei lati del primo, e a', b', c' quelle del secondo, vale la (doppia) proporzione

$$a : a' = b : b' = c : c' \qquad ossia \quad \frac{a}{a'} = \frac{b}{b'} = \frac{c}{c'}\,.$$

2.25 Complementi – Quanti circoli esistono?

Tutti i segmenti sono simili, e anche i circoli, i cerchi, i quadrati e i triangoli equilateri, e perfino – cosa sorprendente e poco nota anche fra i matematici – tutte le parabole.

Più in dettaglio:
Nel piano euclideo esiste sostanzialmente una sola retta e tutte le altre sono solo rotazioni traslate (spostate) di quella.
Nel piano euclideo esiste sostanzialmente un solo circolo e gli altri sono ingrandimenti/rimpicciolimenti di quello, salvo eventualmente ancora una traslazione (spostamento).
Nel piano euclideo – e questo è meno evidente – esiste sostanzialmente una sola parabola, e tutte le altre sono ingrandimenti/rimpicciolimenti di quella, salvo eventualmente ancora una traslazione (spostamento) e una rotazione.
Una cosa analoga seppure non uguale avviene coi grafici delle funzioni esponenziali (vedi 14.2)

BOZZA - DRAFT

2.26 Complementi – Teoremi di Euclide.

Un triangolo con un angolo retto si chiama *triangolo rettangolo*; *ipotenusa* il suo lato maggiore e *cateti* gli altri. Dette nell'ordine i, c_1 e c_2 le lunghezze di quei lati, h quella della distanza dell'ipotenusa dal vertice *opposto* (cioè h è la lunghezza dell'*altezza relativa all'ipotenusa*), e p_1 e p_2 le lunghezze delle *proiezioni dei cateti sull'ipotenusa* (nell'ordine), valgono:

$$c_1^2 = h \cdot p_1, \quad c_2^2 = h \cdot p_2 \qquad \textit{Primo Teorema di Euclide}$$

$$p_1 : h = h : p_2 \qquad \textit{Secondo Teorema di Euclide.}$$

2.27 Complementi – Note sugli enti geometrici

La distanza è un numero reale non negativo.
Lunghezza, area e volume sono $+\infty$ o numeri reali non negativi.
Le altre cose cose finora considerate invece sono figure e quindi insiemi di punti, tranne l'orientazione della retta e della semiretta: l'orientazione è un ente matematico diverso dai numeri, da $+\infty$ e

dalle figure; è più sottile ma possiamo ritenerlo intuitivo. (Ancora più sottile è l'orientazione del piano, che vedremo, e ancora più sottile l'orientazione dello spazio).

2.28 Complementi – Definizioni precise in Farmacia

Le definizioni, parola per parola, hanno implicazioni legali – talvolta di smisurata portata, anche economica, perfino mondiale.

Leggiamo su Facta:

> La definizione (...) al momento in cui scriviamo, recita «Vaccino: Un preparato che è usato per stimolare la risposta immunitaria del corpo contro delle malattie». La stessa pagina il 21 agosto 2021, preservata da Internet Archive, invece definiva il vaccino «Un prodotto che stimola il sistema immunitario di una persona per produrre l'immunità da una specifica malattia, proteggendo la persona da quella malattia».

BOZZA DRAFT

Questo cambiamento di definizione ha permesso di far approvare come vaccini prodotti che prima non sarebbero rientrati nella definizione (i vaccini non possono, come dice Facta, essere "efficaci al 100 per cento"), anche capitalizzando la fiducia guadagnata dal Prodotto presso il Consumatore nei secoli precedenti, ascritta al nome "vaccino".

Non si facevano cento miliardi di dollari inoculando una cosa chiamata per esempio *profarmaco*; con *vaccino* sì.

Sconcertante per uno Statistico è l'incredibile varietà di definizioni usate negli articoli scientifici per "vaccinato", quando pure si degnano di riportare quella che hanno usato per la ricerca. Leggiamo per esempio in un articolo[23] scientifico:

[23] Pálinkás A, Sándor J. Effectiveness of COVID-19 Vaccination in Preventing All-Cause Mortality among Adults during the Third Wave of the Epidemic in Hungary: Nationwide

> Partially vaccinated people who received the vaccine during the follow-up period but did not become protected during the study period (did not receive the booster or died before the 7th day after complete vaccination) were excluded from the investigation.

Altro esempio: una famosa famosa Azienda produttrice, nei trial clinici ha definito come reazione avversa superata l'aborto spontaneo, ottenendo quindi un minor peso complessivo delle reazioni avverse, aumentando le possibilità che le venisse concessa l'autorizzazione per la messa in commercio del suo vaccino.

BOZZA - DRAFT

Retrospective Cohort Study. Vaccines (Basel). 2022 Jun 24;10(7):1009. doi: 10.3390/vaccines10071009. PMID: 35891173; PMCID: PMC9319484.

2.29 ESERCIZI SULLA LEZIONE 2

2.29.1 Esercizio risolto a – Volume di una goccia

$_{\mu 2021}$ * Se un millilitro ovvero centimetro cubo d'acqua corrisponde a 20 gocce, e supponendo sferica una goccia, trovarne il diametro.
L'unità d misura, cm, deve essere scritta esplicitamente nel risultato.
(In Farmacia e Medicina si ritiene in generale che 1 millilitro d'acqua corrisponda a 20 gocce. **Non c'è alcuna garanzia che ciò sia un fatto sempre vero, che siano sempre esattamente 20, ma di questo non ci occupiamo.** Con sostanze diverse dall'acqua distillata il numero 20 può cambiare molto ma non ce ne occupiamo. E naturalmente nel campo gravitazionale la forma non sarà esattamente sferica ma non ce ne occupiamo).

SVOLGIMENTO
Verrà usato lo standard della virgola decimale. (Ma si potrebbe usare quello del punto decimale, a scelta).

BOZZA DRAFT

Il volume di 1 goccia è $\frac{1}{20}$ di centimetro cubo, cioè 0,05 cm^3 ovvero ml, e ricordando il volume della sfera abbiamo l'equazione nell'incognito raggio r

$$\frac{4}{3}\pi r^3 = \frac{1}{20} \qquad / \cdot \frac{3}{4\pi}$$

dove r è in cm, e $\frac{1}{20} = 0{,}05$ in cm^3, ma facciamo i calcoli senza unità di misura

$$r^3 = \frac{3}{4\pi} \cdot \frac{1}{20} \qquad / \sqrt[3]{\ }$$

$$r = \sqrt[3]{\frac{3}{80\pi}}$$

e moltiplicando per 2 per avere il diametro invece del raggio

$$\boxed{2 \cdot \sqrt[3]{\frac{3}{80\pi}}\ \text{cm}}$$

e l'esercizio sarebbe finito.

Tuttavia possiamo esprimere meglio la soluzione trovata, facendo "entrare" il 2 sotto la radice cubica, con l'uguaglianza $2 = \sqrt[3]{8}$, e 8 si semplifica con 80,

$$\boxed{\sqrt[3]{\frac{3}{10\pi}}\ \text{cm}}$$

Nota. Numericamente sono ≈ 0.46 cm, circa mezzo centimetro di diametro.

2.29.2 Esercizio risolto b – Volume di una capsula

$_{\mu 2019 modif}$ ≈ Si consideri una capsula (*vulgo*: pillola) ipotetica ma comunque di apparenza alquanto normale – non viene qua seguito alcuno standard legale o industriale – il cui volume interno abbia la forma di un cilindro di diametro 0,4 cm e lunghezza 1,2 cm completato con due semisfere.
Calcolatone il volume, lo si riduca di ¼.
Si ricordi che cilindri e prismi hanno volume *area di base* × *altezza* qualunque forma abbia la base.

SVOLGIMENTO
Verrà usato lo standard della virgola decimale (come già nel quesito).

Prima di tutto osserviamo che nel testo viene dapprima chiamata *lunghezza* del cilindro quella che di solito in geometria elementare, e in particolare nella formula, viene chiamata *altezza* del cilindro: abituiamoci a tale ambiguità.
Ricordiamo la formule dell'area del cerchio e del volume della sfera in funzione dei raggi:

$$\pi r^2 \qquad \frac{4}{3}\pi r^3$$

La base del cilindro è un cerchio di raggio pari alla metà del diametro

$$r = 0{,}2\,\text{cm}$$

e quello è anche il raggio delle semisfere.
Avendosi 2 semisfere, calcoleremo il volume di 1 sfera e lo sommeremo al volume del cilindro, la cui area di base è πr^2, facendo i calcoli dapprima senza unità di misura:

$$V_{sfera\ unione\ delle\ 2\ semisfere} + V_{cilindro} =$$

$$= \frac{4}{3}\pi r^3 + \pi r^2 h =$$

$$\approx \frac{4}{3} \cdot 3{,}14 \cdot 0{,}2^3 + 3{,}14 \cdot 0{,}2^2 \cdot 1{,}2 \approx 0{,}184213$$

che ora riduciamo *di* ¼, come richiesto, con la formula
$x \mapsto x - \frac{1}{4}x$ (da tenere ben distinta dalla $x \mapsto \frac{1}{4}x$ corrispondente a "ridurre *a* ¼")

$$\approx 0{,}184213 - \frac{0{,}184213}{4} \approx 0{,}13816$$

e con ragionevole approssimazione

$$\boxed{0{,}138\ \text{cm}^3}$$

(Sono 138 mm³, approssimativamente $\frac{1}{7}$ di centimetro cubo ovvero millilitro).

BOZZA DRAFT

NOTA. Le scritture frazionarie come $^1\!/_4$ e $\frac{1}{7}$ vengono escluse dalla *Raccomandazione per la prevenzione degli errori in terapia conseguenti all'uso di abbreviazioni, acronimi, sigle e simboli* (settembre 2018) del Ministero della Salute italiano:

> evitare l'uso delle frazioni (ad esempio, $^1\!/_2$ compressa ovvero "metà compressa" può essere frainteso con 1 o 2 compresse) e sostituire, ove possibile, il farmaco con altra forma farmaceutica avente il dosaggio necessario

ma esse continueranno a trovarsi in un'infinità di testi, anche di Farmacia.

2.29.3 Problema tanto carino quanto difficile ☺☺☺

Problema di Apollonio. Trovare un cerchio che tocchi (cioè, *tangente* ai *bordi*) simultaneamente tre cerchi dati.

☺

☺

☺

Ce ne possono essere parecchi ma è molto difficile trovarne i centri e i raggi.

BOZZA - DRAFT

2.29.4 Problema insolubile ☹

Con soli riga e compasso trovare il lato di un cubo che abbia volume doppio di un cubo a assegnato.
(È ben nota la semplice costruzione con riga e compasso di un quadrato avente area doppia di un quadrato assegnato, ma il corrispondente problema col volume dei cubi è insolubile).

2.29.5 Aneddoto – Perché studiare Geometria?

> Si racconta che, nell'antica Alessandria, durante una lezione di geometria, un giovane studente si avvicinò a Euclide, uno dei più grandi matematici di tutti i tempi, e gli chiese: "Qual è il beneficio pratico di studiare la geometria?" Il ragazzo, probabilmente stanco di lavorare su prove e postulati senza vederne un'applicazione immediata, sperava in una risposta che giustificasse il tempo e l'energia spesi nello studio.
> Euclide, senza rispondere direttamente, chiamò un servo e disse: "Dagli tre oboli, perché vuole trarre profitto da ciò che sta imparando." (Un obolo era una piccola moneta dell'antica Grecia.)

(ChatGpt, 2024)

3 Logica delle proposizioni

3.1 Premessa

(Contro)esempi stupefacenti ci indicano che le affermazioni matematiche vanno dimostrate, non basta[24] fare un "mostra e dimostra".

Le illusioni ottiche ci indicano che è inappropriato usare solo figure per fare dimostrazioni.

Le ambiguità linguistiche ci indicano che è inappropriato usare solo testo libero del linguaggio comune per fare dimostrazioni.

La varietà delle lingue ci induce a cercare un **preciso linguaggio matematico simbolico** che prescinda dalle lingue.

Tutto ciò viene fatto ad un livello molto profondo in Matematica ma in questo testo elementare considereremo solo qualche elemento.

BOZZA - DRAFT

3.2 Introduzione

Prima di tutto consideriamo il simbolo = da intendersi nel senso "è lo stesso elemento". (Due rette perpendicolari r e s sono senz'altro *uguali* nel senso del linguaggio comune ma dal punto di vista matematico sono *diverse*: $r \neq s$, non è la stessa retta). Per esempio $area_{\Delta}=\frac{1}{2}\,b\,h$.
In generale scriveremo 3 varianti del simbolo di uguaglianza:

:= nelle definizioni, per esempio $f(x) := x^2$,

$\equiv$ nelle identità, per esempio $(x+1)^2 \equiv x^2 + 2\,x + 1$.

$=^{EQ}$ nelle equazioni, per esempio
$(x+1)^2 =^{EQ} x^2 + 5\,x + 1$, equazione con (unica) *soluzione* $x = 0$.

[24] Per esempio $n^2 + n + 41$ dà un numero primo per $n = 0, 1, 2, ..., 38, 39$ ma non per $n = 40$. E si consideri la tragica disillusione del tacchino induttivista.

Supporemo concetto primitivo (noto) le *proposizioni*. Esempi:
2 è pari
3 è pari (non è molto vero ma non importa, è proposizione)
X ha il diabete
Y è iperteso

Introdurremo alcuni simboli logici rigorosi.

Indicheremo con $\Rightarrow$ l'*allora*, ovvero meglio *implica*:
per esempio
$1009 \cdot 2 = 2018 \Rightarrow$ 2018 è pari.
Indicheremo con $\Leftarrow$ il *perché*:
per esempio
2018 è pari $\Leftarrow 1009 \cdot 2 = 2018$
Si noti che p$\Rightarrow$q si può scrivere q$\Leftarrow$p.

Indicheremo con $\Leftrightarrow$ il *se e solo se*
Per esempio:
X triangolo:
equiangolo $\Leftrightarrow$ equilatero

BOZZA - DRAFT

Supponiamo noti il *vero* e il *falso*.
Li denoteremo V e F ma internazionalmente possono trovarsi indicati T e F e in vari altri modi.

Una **congettura (indimostrata)** è un'affermazione che non è dimostrata vera e non è dimostrata falsa. (Ma è creduta vera da qualche persona significativa, sennò non ne parleremmo).

Esempio. Congettura di Goldbach (Eulero e Goldbach, 1742)

ogni numero pari > 2 è somma di 2 numeri primi.

3.3 Connettivi logici

Vedremo 4 *connettivi logici*: *non*, *et*, *vel*, *aut*: il *non* opera su 1 proposizione e gli altri congiungono fra loro 2 proposizioni.

3.4 Connettivo logico non

Tavola di verità del *non* ovvero $\neg$ ovvero $\tilde{}$ ovvero NOT

p	$\neg p$ scritto anche $\tilde{p}$ o semplicemente *non p*
V	F
F	V

Per esempio p :="3 è pari" ha il *valore di verità* F, $\tilde{p}$ invece V.
$\tilde{p}$ potremo scriverla "non 3 è pari" o meglio "3 non è pari" e in ogni caso è equivalente alla "3 è dispari".

Non è sano. (È malato).

BOZZA - DRAFT

3.5 Connettivo logico et

Continuiamo con l'*et* ovvero $\wedge$ ovvero AND che significa *e*:
"A gennaio sosterrò Matematica e Fisica."
(Se farò solo 1 o 0 di essi mi accuseranno di falsità).
Tavola di verità dell'*et*:

p	q	p *et* q ... proposizione *composta*
V	V	V
V	F	F
F	V	F
F	F	F

È iperteso e ha il diabete.

3.6 Connettivo logico vel

Continuiamo[25] col *vel* ovvero $\vee$ ovvero OR che significa *o* ovvero *oppure*:
"A gennaio sosterrò Matematica o Fisica."
(Se sosterrò entrambi nessuno mi accuserà di falsità).
Questa è la tavola di verità del *vel*:

P	q	$p \vee q$
V	V	V
V	F	V
F	V	V
F	F	F

È iperteso vel ha il diabete. (Potrebbe avere entrambe le patologie).

BOZZA - DRAFT

3.7 Connettivo logico aut

Il *vel* è molto diverso dall'*aut*, l'*o esclusivo*, in italiano (linguaggio comune) ugualmente espresso con *o* oppure *oppure* come il *vel*:
"Mi fidanzerò con Asdrubala o Berenice." Questo *o* è un *aut*.[26]
Questa è la tavola di verità dell'*aut*:

p	q	p *aut* q
V	V	F
V	F	V
F	V	V
F	F	F

In Excel in italiano, i 4 connettivi sono NON, E, O, XOR.

[25] Purtroppo sia il *vel* che l'*et* talvolta vengono resi con una virgola, e *sperabilmente* si capisce il senso dal contesto:

- $\{(x,y)|\, x \geq 0, y \geq 0\}$ qua vale $\wedge$: è il I quadrante, l'insieme dei punti con sia l'ascissa che l'ordinata ≥ 0;
- $\{(x,y)|\, x = \sqrt{2}, \sqrt{3}\}$ qua vale $\vee$ cioè $x = \sqrt{2} \vee x = \sqrt{3}$: sono 2 rette verticali, è ovvio che si richiede che x sia o $\sqrt{2}$ o $\sqrt{3}$, mica può avere sia l'uno che l'altro valore.

[26] L'ambiguità notazionale è notevole per tutti gli operatori logici: per l'*aut* Wikipedia in inglese alla voce *Exclusive or* dà 10 simboli.

3.8 I connettivi logici in Farmacia

Esempio 1. In questo testo (non più presente nell'ottobre 2021 ma non importa) dei prestigiosi CDC statunitensi, Centers for Disease Control and Prevention, troviamo un uso tecnico dell'et, scritto AND, e del vel, scritto OR, per la definizione dei criteri clinici di diagnosi del covid:

Clinical Criteria
At least two of the following symptoms: fever (measured or subjective), chills, rigors, myalgia, headache, sore throat, new olfactory and taste disorder(s)

OR

At least one of the following symptoms: cough, shortness of breath, or difficulty breathing

OR

Severe respiratory illness with at least one of the following:

- Clinical or radiographic evidence of pneumonia, **OR**
- Acute respiratory distress syndrome (ARDS).

AND

No alternative more likely diagnosis

BOZZA - DRAFT

Esempio 2. Operatori logici in ricerca bibliografica in database. Leggiamo in un articolo[27] scientifico:

> Search string used to identify publications from each of the four databases used for potential inclusion in the study.
>
> Search string
>
> "'vaccin* OR innoculat* OR immunization*' AND 'Pfizer OR AstraZeneca OR Moderna OR pfizer-biontech OR pfizerbiontech OR

[27] Reynolds L, Dewey C, Asfour G, Little M. Vaccine efficacy against SARS-CoV-2 for Pfizer BioNTech, Moderna, and AstraZeneca vaccines: a systematic review. Front Public Health. 2023 Oct 24;11:1229716. doi: 10.3389/fpubh.2023.1229716. PMID: 37942238; PMCID: PMC10628441.

> bnt162b2 OR mrna-1273 OR ChAdOx1-S OR Oxford AstraZeneca' AND 'efficacy OR effective*' AND 'COVID-19 OR SARS-CoV-2 OR COVID19* OR COVID OR SARSCOV-2 OR SARSCOV2 OR Severe Acute Respiratory Syndrome Coronavirus 2 OR Severe Acute Respiratory Syndrome Corona Virus 2'".

Naturalmente *vaccin** a sua volta indirizza la ricerca a record contenenti

> vaccine OR vaccinated OR vaccination...

Esempio 3. In questo testo della prestigiosa FDA statunitense, Food and Drug Administration, il *vel* è sottinteso, ed è ovvio che si tratti di un *vel* e non di un *et* o altro:

> An adverse event is any undesirable experience associated with the use of a medical product in a patient. The event is serious and should be reported to FDA when the patient outcome is:
> **Death**
> (...)
> **Life-threatening**
> (...)
> **Hospitalization (initial or prolonged)**
> (...)
> **Disability or Permanent Damage**
> (...)
> (seguono altri casi).

BOZZA - DRAFT

3.9 Tautologie e contraddizioni

Si dice *tautologia* una proposizione sempre vera.
Esempio: $p \vee \neg p$

Si dice *contraddizione* una proposizione sempre falsa.
Esempio: $p \wedge \neg p$

Proposizioni *equivalenti*: hanno la stessa tavola di verità.
Esempio: (p aut q) ≡ (p et non q) vel (q et non p)

3.10 Nota sulla Logica

Lo studio della Logica vorrebbe insegnarci a ragionare correttamente.

Leggiamo su Wikipedia:

> Un caso di "non sequitur" è quello che si indica con l'espressione "*Post hoc ergo propter hoc*" (dopo ciò, dunque a causa di ciò)", ovvero considerare A causa di B solo perché B segue temporalmente A

. È impressionante l'illogicità usata abitualmente dalle persone che hanno studiato poco – ma ancor più colpisce quella usata da persone di alto grado nelle istituzioni, che hanno responsabilità sulla vita di tutti.

BOZZA - DRAFT

Leggiamo il giorno 15 giugno 2020 sul sito del Ministero della Salute italiano[28]

> Mangiare tante arance e limoni previene il contagio perché la vitamina C ha azione protettiva nei confronti del nuovo coronavirus
> Falso
> Non ci sono evidenze scientifiche che provino un'azione della vitamina C sul virus

Ma allora è congettura indimostrata, e quindi forse è falso ma forse è vero, oppure è certamente falso come sopra affermato?

Al 5 ottobre 2021 il testo risultava modificato; nel 2023 sparito.

[28] http://www.salute.gov.it/portale/malattieInfettive/dettaglioNotizieMalattieInfettive.jsp?lingua=italiano&id=4276 letto il 15 giugno 2020.

3.11 ESERCIZI SULLA LEZIONE 3

3.11.1 Esercizio risolto a – Sintomi e patologia

$_{\mu 2018\, modificato}$ * Una certa condizione patologica X, è diagnosticata se:
c'è il sintomo A
e
c'è il sintomo B oppure (vel) il sintomo C
oppure (vel)
c'è il sintomo D ma non il sintomo A.
Cioè, indicando la negazione con la tilde, con ovvio significato dei simboli,

$$(a \wedge (b \vee c)) \vee (d \wedge \tilde{a}).$$

Indicando con V e F il *vero* e il $falso$, si conduca di passaggio in passaggio il calcolo relativo ad un paziente che ha i soli sintomi A, C, D, concludendo la diagnosi.
(Ragionamenti analoghi e più complessi possono venire gestiti da software).

SVOLGIMENTO
Si ha
sintomo A presente: a è vera, V
sintomo B non presente: b è falsa, F
sintomo C presente: c è vera, V
sintomo D presente: d è vera, V
e si calcola

$$(V \wedge (F \vee V)) \vee (V \wedge \tilde{V})$$

$$(V \wedge V) \vee (V \wedge F)$$

$$V \vee F$$

$$V$$

e in conclusione abbiamo la diagnosi

La condizione patologica X è presente

BOZZA - DRAFT

3.11.2 Esercizio risolto b – Finanziare terapie o no

$_{\mu 2018}$ * Una certa terapia verrà supportata da un governo se (e solo se)
è molto efficace e costa molto
OPPURE se
è solo abbastanza efficace e non costa molto.
(Supponendo in qualche modo determinati i significati precisi dei termini).
Con ovvio significato dei simboli

$$(p \wedge q) \vee (r \wedge \neg q).$$

Nel caso specifico di una terapia che sia solo abbastanza efficace e costi molto si attribuiscano i valori di verità V o F a p, q, r, e si valuti di passaggio in passaggio fino al risultato finale il valore di verità dell'espressione soprascritta, concludendo se la terapia verrà supportata o no.
(Ragionamenti analoghi e più complessi possono venire gestiti da software).

SVOLGIMENTO
Si hanno queste 3 proposizioni, coi loro valori di verità nel caso specifico:
F p ="la terapia è molto efficace"
V q ="la terapia costa molto"
V r ="la terapia è solo abbastanza efficace"
Si ha quindi successivamente

$$(F \wedge V) \vee (V \wedge \neg V)$$
$$F \vee (V \wedge F)$$
$$F \vee F$$

F (falso)
La terapia non verrà supportata

BOZZA - DRAFT

3.11.3 Esercizio risolto c – Sintomatici per il Ministero

$_{\mu 2021}$ * Supponiamo che dal Ministero arrivi alle Farmacie una circolare che impone di inviargli una segnalazione se arriva un cliente con

tosse O nelle ultime 48 ore febbre oltre 38,5°C
E
naso che cola O non vaccinato.
Riconosciuto fra questi 4 il calcolo logico da fare

$$(p \vee q) \wedge (\neg(r \vee s))$$

$$(p \wedge q) \vee (\neg(r \wedge s))$$

$$(p \vee q) \wedge (r \vee \neg s)$$

$$(p \wedge q) \vee (r \wedge \neg s)$$

lo si svolga, indicando con V il valore di verità vero, con F quello falso, con ? quello sconosciuto, fino a determinare se la segnalazione va inviata per un cliente con tosse, naso che cola, vaccinato.

SVOLGIMENTO
Viene usato lo standard della virgola decimale.
Chiaramente la terza espressione esprime le condizioni poste.
In effetti non sappiamo se il cliente ha avuto febbre oltre 38,5°C nelle ultime 48 ore o no e allora indichiamo con ? il valore di verità di q:

$V\ p$:="ha la tosse"
$?\ q$:="ha avuto febbre oltre 38,5°C
$V\ r$:="ha il naso che cola"
$V\ s$:="è vaccinato".

Si ha successivamente

$$(V \vee ?) \wedge (V \vee \neg V)$$

Sia $V \vee V$ che $V \vee F$ danno V, vero,

$$V \wedge (V \vee F)$$

$$V \wedge V$$

$$V$$

La segnalazione va inviata

BOZZA - DRAFT

3.11.4 Esercizio tanto carino quanto difficile

Il Problema dei Cappelli.

Tre persone, A, B e C, sono disposte in fila indiana, cioè A può vedere B e C, B può vedere solo C, e C non può vedere nessuno. Ognuna di loro ha un cappello sulla testa, che può essere di colore rosso o blu. Le persone non possono vedere il proprio cappello e non sanno quante persone indossano un cappello di un certo colore. Il giudice comunica alle tre persone che almeno uno di loro ha un

cappello rosso. Quindi chiede a A: "Puoi dire di che colore è il tuo cappello?" A risponde di no. Lo stesso viene chiesto a B, che risponde anche lui di no. Infine, viene chiesto a C che risponde con sicurezza: "Il mio cappello è rosso."
Come ha fatto C a dedurre il colore del suo cappello?

BOZZA - DRAFT

4 La Logica con l'Insiemistica

L'insiemistica dà una concettualizzazione univoca[29] per rappresentare un'infinità di situazioni della realtà, e permette una modellizzazione in grado di rendere chiare relazioni fra enti vari – persone, malattie, farmaci, soldi, di tutto – che sarebbero semplicemente *invisibili* nell'elencazione bruta di, poniamo, 60milioni di casi.

BOZZA DRAFT

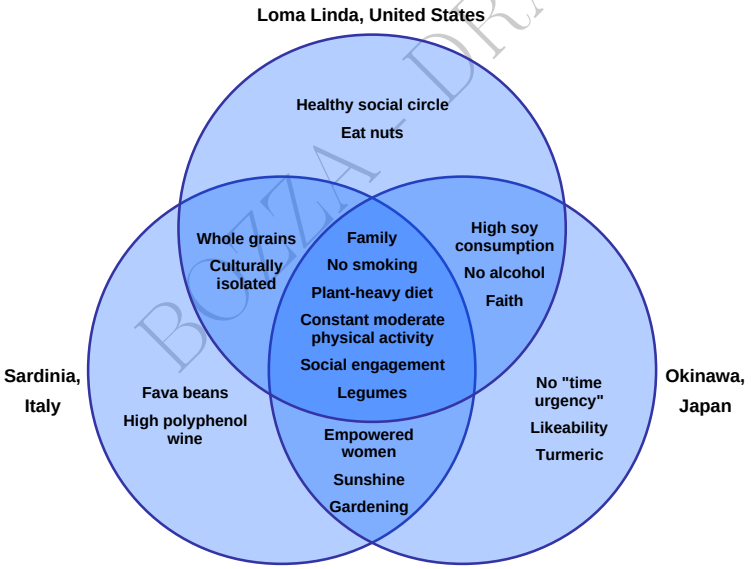


Figure 1: Abitudini in 3 zone del mondo con aspettativa di vita ritenuta molto elevata. Ci sono in tutto 21 elementi, che sono le abitudini considerate. By The RedBurn, https://commons.wikimedia.org/wiki/File:3_blue_zones_venn_diagram.svg

[29] Salvo le attualmente purtroppo numerose ambiguità notazionali, ma questa è altra questione, ora di minor importanza.

4.1 Nozioni di insieme, elementi e appartenenza

In questa trattazione elementare, la nozione di *insieme* è supposta nota, ovvero è data come *concetto primitivo*. Possiamo dire che un insieme è, in qualche modo, una *raccolta* di elementi, o *classe* (di elementi). Ma anche la nozione di *elemento* è supposta nota, ovvero è data come *concetto primitivo*, e anche la nozione di *appartenenza* di un elemento ad un insieme.

Agli insiemi *di solito* si dà nome una lettera latina maiuscola, come A e X, ma per alcuni insiemi si usano grafie particolari, per esempio l'insieme dei numeri naturali viene denotato con $\mathbb{N}$. L'*appartenenza* (concetto primitivo) di un elemento ad un insieme si indica col simbolo $\in$, per esempio con $3 \in I\!N$, e la sua negazione con $\notin$: $-3 \notin I\!N$. Anche, scriveremo equivalentemente prima $\mathbb{N}$ e poi 3, col simbolo $\in$ scritto specularmente, e similmente per $\notin$.

Come *variabili* atte a rappresentare un elemento indeterminato di un insieme, *di solito* si usano lettere latine minuscole: a, x...

È ovvio che un insieme si può considerare assegnato se è univocamente *chiaro*[30] se un elemento vi appartiene o no, per esempio le persone *simpatiche* non costituiscono un insieme. Quelle *sane*, o *povere*, o *diabetiche*, o *ipertese*, sì, se si suppone che da qualche parte siano stati ben definiti quegli attributi.

Un insieme può essere determinato *per elencazione* (*principio di estensione*) elencandone gli elementi fra parentesi graffe, per esempio $\{3, 5, 8\}$, eventualmente con puntini di sospensione per indicare elementi non trascritti ma che si suppone il lettore possa capire quali sono, per esempio $\{0, 1, 2, 3...\}$, l'insieme $\mathbb{N}$ dei *numeri naturali*.

Tipici insiemi sono i *singoletti*, come $\{x\}$, $\{diabete\}$, $\{0\}$, eccetera, con un solo elemento.

Gli insiemi si possono anche rappresentare (*principio di astrazione*) con la *proprietà caratteristica*, cioè *caratterizzante*, degli elementi, per esempio $\{x \in I\!N | x\ pari\}$, che magari talvolta scriveremo solo

[30] La questione della menzionata *chiarezza* non è banale, ma qui non la approfondiremo; si noti comunque che, per esempio, i divisori di $1+2024^{2025}$ costituiscono un insieme, nonostante possa essere improbo stabilire se qualche determinato numero vi appartenga.

BOZZA - DRAFT

$\{pari\}$, che però potrebbe lasciarci nel dubbio, se non è chiaro dal contesto, che si intendesse $\{x \in \mathbb{Z} | x\, pari\}$, comprendendo per esempio -8. Il simbolo $|$ si legge "tale che" e altre volte si scrive : o t.c.. Naturalmente gli insiemi poi possono rappresentarsi graficamente con ovali racchiudenti i loro elementi (*diagrammi di Eulero-Venn*).

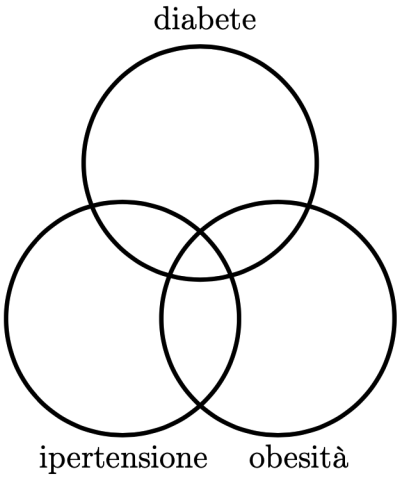
Alcuni Autori distinguono i diagrammi di Eulero da quelli di Venn: si vedano le figure 1 e 8 nell'articolo[31] scientifico
Applying Euler Diagrams and Venn Diagrams to Concept Modeling
Invece in questo testo entrambe le rappresentazioni verranno considerate di Eulero-Venn.

BOZZA - DRAFT

4.2 Esempio: comorbidità

ESERCIZIO$_{\mu}$ $*$ Rappresentare coi diagrammi di Eulero-Venn la situazione che si forma considerando le 3 patologie diabete, ipertensione e obesità, e poi si elenchino i 7 casi possibili di malattia, singola o multipla, che si evidenziano nel diagramma, a partire dal caso peggiore, "ha diabete e ipertensione e obesità", e finendo con "ha solo diabete".

SVOLGIMENTO



ha diabete e ipertensione e obesità
ha diabete e obesità ma non ipertensione
ha diabete e ipertensione ma non obesità
ha ipertensione e obesità ma non diabete
ha solo obesità
ha solo ipertensione
ha solo diabete

Per le 7 diverse categorie di pazienti potrebbero essere previste terapie diverse.

[31] Lyalin D., Bus Rules J. 2020 Feb;21(2):c021. PMID: 35440915; PMCID: PMC9014965.

4.3 Esempio: proprietà farmacologiche

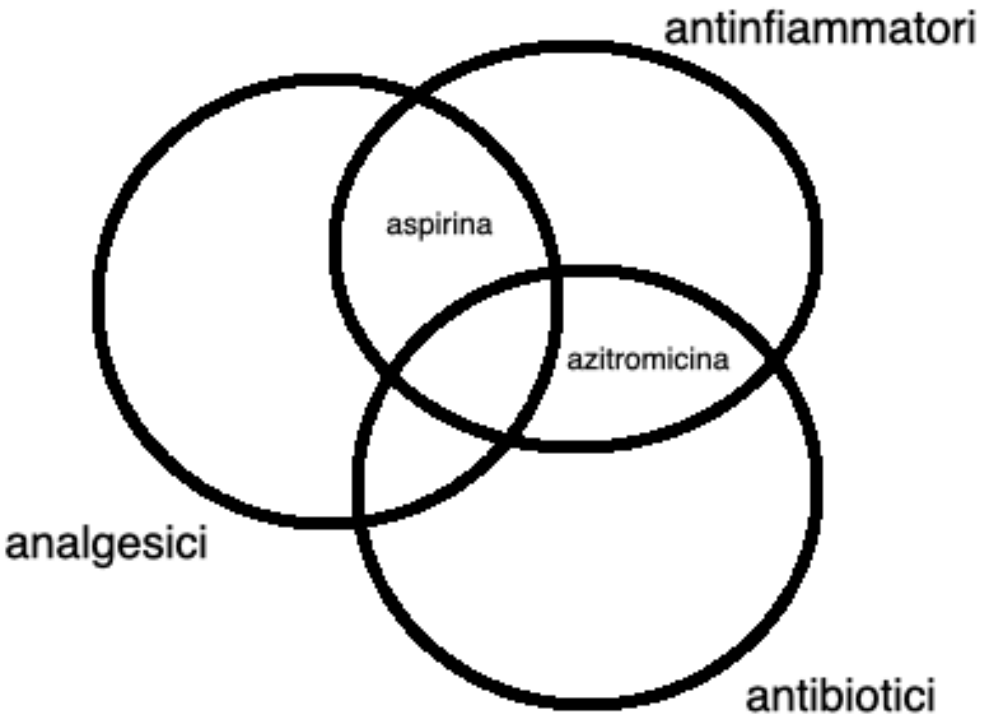
Proprio come sopra, si possono considerare 3 cerchi, ciascuno corrispondente a una di queste proprietà farmacologiche dei vari principi attivi:

antibiotici, antinfiammatori, analgesici.

Per esempio l'azitromicina è un antibiotico antinfiammatorio. Ma alcuni sottoinsiemi sono vuoti: non esistono antibiotici specificamente classificati come analgesici.

Arricchisca il Lettore il sottostante diagramma di Eulero-Venn con altri principi attivi.

DRAFT



4.4 Uguaglianza; insieme vuoto; inclusione

Due insiemi si dicono uguali se hanno gli stessi elementi. Anche se hanno diverse definizioni, per esempio

$$X := \{divisori\ di\ 6\} = \{1, 2, 3, 6\} =$$

$$= Y := \{numeri\ di\ 3\ lettere\ in\ italiano\}$$

Esiste un unico insieme privo di elementi, l'*insieme vuoto*, denotato a stampa $\varnothing$ o, scrivendo a mano, qualcosa come ϕ.

Se ogni elemento di X appartiene a Y si dice che X è *contenuto* in Y:

$$X \subseteq Y \quad \text{(equivalente a } Y \supseteq X\text{, contiene)}$$

Esempi.

{titolari tessera fedeltà farmacia} $\subseteq$ {aventi diritto allo sconto}

{invalidi} $\subseteq$ {aventi diritto allo sconto} $\subseteq$ {umanità}

(Si faccia attenzione che ci possono essere invalidi titolari di tessera fedeltà).

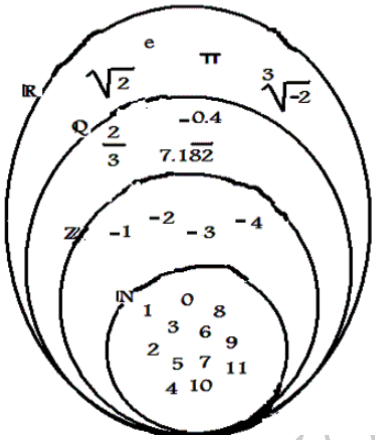
BOZZA - DRAFT

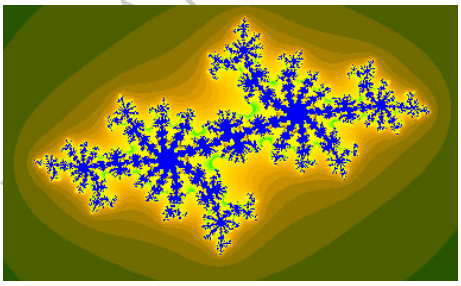
4.5 Gli insiemi numerici

È

$$\mathbb{N} \subsetneq \mathbb{Z} \subsetneq \mathbb{Q} \subsetneq \mathbb{R}$$

e ovviamente $\subsetneq$ indica l'inclusione che esclude l'uguaglianza: per esempio $\mathbb{Q}$ è *contenuto propriamente* in $\mathbb{R}$.





Un *soprainsieme* proprio di $\mathbb{R}$ è l'insieme $\mathbb{C}$ dei *numeri complessi*, dei quali non ci occuperemo: $\mathbb{C} \supsetneq \mathbb{R}$. In $\mathbb{C}$ si formano facilmente "grafici" di spettacolare bellezza, i *frattali*, come l'Insieme di Julia in figura.

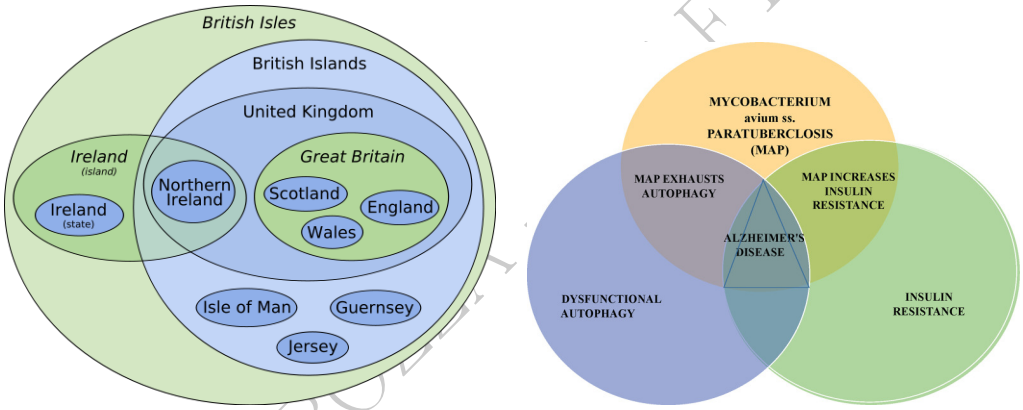
4.6 Cardinalità, insiemi finiti e infiniti

In questa trattazione elementare la *cardinalità* di un insieme finito è il numero dei suoi elementi.
La cardinalità è indicata con $\#E$ (e da altri con card$E$ o $|E|$).
Per esempio $\#\{a, b, c\} = 3$.
Supponiamo noti i concetti di insieme finito e insieme infinito.

4.7 Estensioni.

Il seguente diagramma geografico è molto chiaro ed esplicativo (in azzurro determinazioni giuridiche, in verde determinazioni geografiche) ma non corrisponde esattamente alle definizioni che abbiamo dato: United Kingdom ha sia i 2 elementi Great Britain e Northern Ireland, che i 4 elementi England, Wales, Scotland, Northern Ireland.

Anche il diagramma di argomento medico è alquanto chiaro ma non corrisponde esattamente alle definizioni che abbiamo dato.



Da: Front Immunol. 2021; 12: 714179

By: Coad Thomas Dow

4.8 Insieme delle parti

L'insieme delle parti $\mathscr{P}(A)$ di un insieme A è l'insieme di tutti i sottoinsiemi di A. (Compresi ovviamente l'insieme vuoto e A stesso). In simboli

$$\mathscr{P}(A) := \{X : X \subseteq A\} .$$

L'insieme delle parti di A si chiama anche *insieme potenza* di A.
Trattazione elementare valida per i soli insiemi finiti.
Per esempio se $A := \{a, b, c\}$ allora

$$\mathscr{P}(A) = \{\varnothing, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\} .$$

Se un insieme ha n elementi allora (teorema) il suo insieme delle parti ha 2^n elementi. Cioè, usando il simbolo # della cardinalità, qua nel significato semplice di *numero di elementi*,

$$\#\mathscr{P}(A) = 2^{\#A}\,.$$

Nell'esempio $\mathscr{P}(A)$ ha 8 elementi, cioè 2^3, perché A ha 3 elementi. È possibile una trattazione ad un livello superiore, valida per tutti gli insiemi, finiti e infiniti.

ESERCIZIO$_\mu$ Supponiamo che ad un farmacista, cui normalmente arrivano un centinaio di persone al giorno, una volta ne arrivino un migliaio, lamentando svariati sintomi fra una mezza dozzina di sintomi. Volendo individuare l'insieme di sintomi più caratteristico della nuova situazione, per affrontarla validamente, si propone di elencare tutti i possibili sottoinsiemi di sintomi, da 2 sintomi fino a 6 sintomi. Quanti elementi avrà questa lista di liste di sintomi?

SVOLGIMENTO[32]

I sottoinsiemi dell'insieme di 6 sintomi sono (insieme delle parti) 2^6 ma dobbiamo escludere i 6 insiemi costituiti da 1 solo sintomo e 1 insieme costituito da 0 sintomi (l'insieme vuoto) e allora in tutto

$$2^6 - 6 - 1 =$$

(che naturalmente significa $(2^6 - 6) - 1$: cioè in assenza di parentesi facciamo le operazioni nell'ordine in cui sono scritte, non faremo certo prima $6 - 1$, che produrrebbe un errore, bensì prima $(2^6 - 6)$)

$$\boxed{57}$$

BOZZA - DRAFT

[32] Ecco per completezza alcuni elementi della lista con 57 liste di sintomi, denotati con S1,...,S6:
S1 S2 ← 1^ lista
S1 S3
...
S1 S6
S2 S3
S2 S4
...
S5 S6
S1 S2 S3
S1 S2 S4
...
S1 S2 S3 S4 S5 S6 ← 57^ lista.

4.9 Operazioni insiemistiche e logiche a confronto

Si veda

https://www.shutterstock.com/it/image-vector/sets-theory-basic-opera

Consideriamo gli insiemi come sottoinsiemi di un *insieme universo*, in generale chiamato U ma non necessariamente:

per esempio $\mathbb{N}$, oppure $\mathbb{R}$,

oppure {clienti della nostra farmacia}

oppure {residenti in Italia}.

Insieme complementare di 1 insieme:

$A^C := \{x \in U | x \notin A\}$ cioè $\neg(x \in A)$. Denotato anche A', $\sim A$, $\overline{A}$.

Insieme intersezione di 2 insiemi:

$A \cap B := \{x \in U | x \in A \wedge x \in B\}$

Insieme unione di 2 insiemi:

$A \cup B := \{x \in U | x \in A \vee x \in B\}$

Insieme differenza simmetrica di 2 insiemi:

$A \Delta B := \{x \in U | x \in A \; aut \; x \in B\}$

Si hanno allora queste 4 corrispondenze[33] fra insiemistica e logica:

C $\neg$

$\cap$ $\wedge$

$\cup$ $\vee$

Δ aut

BOZZA - DRAFT

[33] Valgono queste ulteriori corrispondenze fra simboli insiemistici e logici:
2 *involuzioni*:
$(A^C)^C = A$
$\neg(\neg p) = p$
4 *proprietà distributive* (corrispondentisi a coppie)
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
$p \vee (q \wedge r) = (p \vee q) \wedge (p \vee r)$
$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
$p \wedge (q \vee r) = (p \wedge q) \vee (p \wedge r)$
2 proprietà associative, 2 Leggi di De Morgan, che il lettore interessato troverà facilmente, eccetera.

Si definisce poi l'*insieme differenza* di 2 insiemi:
$A \setminus B := \{x \in U | x \in A \wedge x \notin B\}$

Due insiemi si dicono *disgiunti* se hanno intersezione vuota.
Per esempio
$\{italiani\, sani\}$ e $\{italiani\, diabetici\}$
$\{colesterolo \leq 200\}$ e $\{colesterolo > 200\}$ (sottinteso, "persone con", e l'unità di misura è pure sottintesa, e anche l'insieme universo, che potrebbe essere quello degli italiani, degli europei, degli esseri umani... dipende dal contesto considerato).

Si definisce *partizione* di un insieme una sua suddivisione in sottoinsiemi disgiunti. Quella soprascritta dei sani e dei diabetici non dà certo una partizione dell'insieme $\{italiani\}$, ma quella secondo colesterolemia sì, come pure
$\{\{sani\}, \{malati\}\}$
$\{\{poveri\}, \{benestanti\}, \{ricchi\}\}$
$\{\{bambini\}, \{adolescenti\}, \{adulti\}, \{anziani\}\}$
(al solito, supposti ben definiti i termini) tutte di evidente interesse epidemiologico.

BOZZA - DRAFT

4.10 Nota sull'insiemistica

L'insiemistica – che trattatiamo al livello minimo per gli scopi della Farmacia – è materia sottile, che raggiunge vertici di inverosimile complessità. Già migliaia di anni fa sono emersi problemi che a tutt'oggi sono studiati intensamente:

Il barbiere del villaggio, colui che rade coloro
che non si radono da sè, si rade da sè?

Suggerimento: si evitino insiemi che contengono se stessi come elemento:

$$\alpha \in \alpha$$

4.11 Predicati e insieme di verità

Versione breve.
Un predicato è una sorta di proposizione
ma con una o più variabili, le quali potranno essere denotate
con qualunque lettera e in particolare x, y, z..., come per esempio:
$q(x)$:= "x è iperteso".
L'*insieme di verità* di un predicato
è il sottoinsieme del dominio in cui il predicato è vero.

Un predicato è una sorta di proposizione ma con una o più variabili, le quali potranno essere denotate con qualunque lettera e in particolare x, y, z..., come per esempio

$$q(z) := \text{“}z \text{ è sano"} \text{ (si supponga definito il concetto)}$$

$$p(z) := \text{“}z \text{ è una retta obliqua del piano cartesiano"}$$

$$r(z) := \text{“}z \text{ ha diritto allo sconto del } 10\%\text{"}$$

che di per sè non sono nè vere nè false, dipende da z. Per esempio

$$q(\text{Asdrubala}) \text{ è vera (speriamo per lei, è un caso ipotetico)}$$

$$p(\text{asse } x) \text{ è falsa}$$

$$p(\text{bisettrice del I e III quadrante}) \text{ è vera}$$

(34)

L'*insieme di verità* di un predicato è il sottoinsieme del dominio in cui il predicato è vero.[35]

BOZZA - DRAFT

[34] Ecco altri predicati, col loro dominio:
$x \in \mathbb{N}$, $p_1(x)$:="x è un numero pari" ($p(-8)$ non ha senso)
$x \in \mathbb{Z}$, $p_2(x)$:="x è un numero pari" ($p(-8)$ è vera)
$x \in \mathbb{Z}$, $p_3(x)$:="$x^2 = 2$"
$x \in \mathbb{R}^+$, $p_4(x)$:="$x^2 = 2$" ($\mathbb{R}^+$ è l'insieme dei reali positivi)
$x \in \mathbb{R}$, $p_5(x)$:="$x^2 = 2$"
$x \in$ {residenti in Italia}, $q_1(x)$:="x è sano" (si supponga definito)
$x \in$ {residenti in Italia}, $q_2(x)$:="x è diabetico" (come sopra)
$x, y \in \mathbb{Z}$ $t_1(x, y)$:="x divide y"
$x, y \in$ {residenti in Italia} $t_2(x, y)$:="x e y sono coniugi"

[35] Insiemi di verità relativi alla nota precedente:
insieme di verità di p_3: $\varnothing$
insieme di verità di p_4: $\{\sqrt{2}\}$
insieme di verità di p_5: $\{-\sqrt{2}, \sqrt{2}\}$

4.12 Quantificatori

Esistono questi 2 *quantificatori*, dal significato ovvio:
quantificatore *per ogni*, $\forall$
quantificatore *esiste*, $\exists$

Riprendendo la $q(z)$ dell'esser sano, di cui sopra, certamente

$$(\exists\, t \in \{\text{esseri umani}\})q(t)$$

cioè esiste (almeno) un sano.
Riprendendo la $r(z)$ del diritto allo sconto, di cui sopra, potrebbe darsi il caso che

$$(\forall x \in \{invalidi\})r(x)$$

cioè che ogni invalido abbia diritto allo sconto del 10%.
Naturalmente si possono considerare esempi matematici.[36]

Indicheremo con $\exists!$ l'espressione "esiste un unico"

4.13 Esempi sui quantificatori e controesempio.

BOZZA - DRAFT

Versione breve.
Il *controesempio* è un esempio che dimostra falsità.
È molto usato nelle dimostrazioni matematiche.
Esempio:
Proposizione falsa:
Ogni quadrilatero equilatero è equiangolo.
Dimostrazione della falsità:
Si considera il controesempio del rombo non quadrato,

[36] Qua si fa riferimento ai predicati definiti in una nota precedente.

Esempio 1:
"$(\exists x \in \mathbb{R})\, p_5(x)$" (cioè "$(\exists x \in \mathbb{R})\, x^2 = 2$) è proposizione vera
Esempio 2:
"$(\forall x \in \mathbb{R})\, p_5(x)$" (cioè "$(\forall x \in \mathbb{R})\, x^2 = 2$) è proposizione falsa
Esempio 3:
"$(\exists x \in \mathbb{Z})\, t_1(x, y)$" (cioè "$(\exists x \in \mathbb{Z})\, x$ divide y) è predicato in y
(fra l'altro sempre vero, cioè vero $\forall y \in \mathbb{Z}$)
Esempio 4:
"$(\forall x \in \mathbb{Z})\, t_1(x, y)$" (cioè "$(\forall x \in \mathbb{Z})\, x$ divide y) è predicato in y
(fra l'altro sempre falso, cioè falso $\forall y \in \mathbb{Z}$)

il quale è un quadrilatero equilatero non equiangolo.
Abbiamo allora il teorema vero:
Per ogni quadrilatero, equilatero $\not\Rightarrow$ equiangolo.

Per ogni triangolo, equilatero implica equiangolo: teorema vero.
Per ogni triangolo, equiangolo implica equilatero: teorema vero.
Per ogni quadrilatero, equilatero implica equiangolo: teorema falso.
Non dimostreremo i 2 teoremi veri; sono poco utili in Farmacia, e la dimostrazione è addirittura inutile. Ma impariamo la logica.

Dimostriamo invece il teorema falso con un controesempio: consideriamo un rombo non quadrato, ed esso è proprio un quadrilatero equilatero (ma) non equiangolo.

Il *controesempio* è un esempio che dimostra falsità. È molto usato nelle dimostrazioni matematiche.
Esempio.
Proposizione falsa: ogni quadrilatero equilatero è equiangolo. Dimostrazione della falsità: si considera il controesempio del rombo non quadrato, il quale è unquadrilatero equilatero non equiangolo.
Abbiamo allora il teorema vero:
Per ogni quadrilatero, equilatero $\not\Rightarrow$ equiangolo.

Col simbolo *non implica* scriviamo un teorema vero:
Per ogni quadrilatero, equilatero $\not\Rightarrow$ equiangolo.

Esempio in Medicina, anzi in effetti in Statistica Medica:

> the scientific story of COVID-19 conflicts with the widespread view of science as a source of certainties. For the general public, if something is described by an equation, it is exact. Epidemiological modelling is a dramatic counter-example. https://www.nature.com/articles/s42254-020-0188-2.pdf

(Le affermazioni matematiche sono *assolutamente vere per l'eternità*, ma i modelli matematici – che sono equazioni – descrivono una realtà loro propria che solo parzialmente la realtà segue; l'implicazione

BOZZA - DRAFT

"il futuro dell'epidemia è descritto da una formula ⇒ sarà così"
è stata rivelata falsa dal controesempio dell'epidemia del 2020.)

4.14 Regole di negazione

La negazione di
"(tutti) i paperopolesi sono onesti"
non è
"(tutti) i paperopolesi sono disonesti", no, affatto!
bensì
"esiste (almeno) un paperopolese disonesto".

Cioè, in simboli e più in generale,
$\neg((\forall x \in E)p(x)) = (\exists x \in E)\neg p(x)$
e vale anche l'altra negazione
$\neg((\exists x \in E)p(x)) = (\forall x \in E)\neg p(x)$

BOZZA DRAFT

4.15 Esempi di negazione

Facciamo 2 negazioni usando il linguaggio naturale, scrivendo "malato" invece di "non sano" e similmente poi per le malattie "incurabili".

La negazione di
tutti gli italiani sono sani
è
esiste almeno un italiano non sano;
diciamo pure:
esiste qualche italiano malato.

La negazione di
per ogni malattia esiste una cura
è
esiste almeno una malattia tale che non esiste la cura (diciamo pure, incurabile).
@ Questo argomento non è capitato (2022) agli esami ma capiterà

4.16 Logica in Matematica e in Farmacia

Una trattazione completa delle attuali conoscenze di Logica riempirebbe parecchi corsi universitari.
È una branca della Scienza che è studiata fin dall'antichità. Il sillogismo, in particolare, è un tipo di ragionamento dimostrativo che fu teorizzato per la prima volta da Aristotele; ecco un esempio:

(premessa maggiore) Tutti gli uomini sono mortali
(premessa minore) Tutti i greci sono uomini
(conclusione) Tutti i greci sono mortali.

Il sillogismo fu molto approfondito nel medioevo, ammettendo 256 tipi di sillogismi, di cui 4 privilegiate, dette Barbara, Celarent, Darii, Ferio.
Fin dall'antichità sono emersi paradossi logici vari, molto problematici. Nel '900 Gödel ha dimostrato una cosa per certi versi drammatica, che esistono cose vere non dimostrabili. ☹

La trattazione della Logica che si è potuta fare al livello di questo testo elementare vorrebbe insegnarci

a ragionare bene

oppure in subordine almeno

a non ragionare male. (Meglio tacere piuttosto).

BOZZA - DRAFT

4.17 Il triste caso del monossido di diidrogeno

Più del 90% delle migliaia di persone morte a Milano nel dicembre 2015 era entrata in contatto con quantità significative di monossido di diidrogeno nelle 24 ore precedenti la morte, ma **questa affermazione vera non implica alcuna indicazione di pericolosità del monossido di diidrogeno**.
Che infatti nessuno bandisce.[37]

Diffidare dai "ragionamenti a parole".

Non è logica ma pseudo-scienza.

[37] Ci mancherebbe proprio: è l'acqua.

La Scienza moderna, che è sperimentale, inizia idealmente con l'esperimento di Galileo Galilei sulla torre di Pisa:

è ovvio (?) che di una palla di piombo e una di creta di uguali dimensioni, lasciate cadere giù, quella di piombo arriva a terra prima. È ovvio. E infatti così affermava la Scienza antica, aristotelica. Ora, sembrerà anche una cosa ovvia, però non è vera: arrivano a terra insieme, e l'esperimento lo mostra.
Addirittura, su un corpo celeste privo di aria e quindi di attrito con essa, non è nemmeno necessario che i 2 oggetti siano sfere uguali:

https://www.youtube.com/watch?v=ZVfhztmK9zI

È proprio questa esigenza di razionalità che genera l'attuale spinta verso la "Medicina Basata sulle Evidenze", cioè in pratica sulla Statistica Medica.
(Però, come spesso, ogni 10 passi avanti che l'umanità fa, ne fa 9 indietro, e alla fine il progresso è molto minore delle aspettative; nel caso della medicina basata sulle evidenze si potrebbe scrivere un libro intero, ma diciamo molto sinteticamente che è sì molto bello appoggiarsi a statistiche, ma il tutto finisce per funzionare male se ci si attacca a 1 solo parametro per valutare la performance di un atto medico; se statisticamente l'iperglicemico migliora nel 90% dei casi in 1 anno, ma nel 20% dei casi in muore in 10 anni, il vantaggio diventa dubbio).

BOZZA - DRAFT

L'esempio seguente è farmaceutico.

L'evitamento medicale di malattie mortali

Molti credono che evitare farmacologicamente e/o chirurgicamente una malattia che causa morte riduca la mortalità.

E dove sta scritto? È uno pseudoragionamento a parole.

Che effettivamente quella terapia riduca la mortalità andrà di-

mostrato scientificamente, per esempio con uno **studio prospettico randomizzato in doppio cieco.**
Non con il predetto pseudo-ragionamento, ritenuto logico, che evitare una malattia che causa morte riduca la mortalità.
Fra i vari esempi possibili si consideri il seguente, che ha suscitato clamore mediatico anche in riferimento all'attrice Angelina Jolie.

Esistono mutazioni genetiche (BRCA1 e BRCA2) che rendono *enormemente* più probabile il cancro al seno.

Asportare quasi del tutto il seno evita quasi del tutto quel cancro.

Ma si riduce la mortalità? Leggiamo in un articolo scientifico:[38]

> (...) bilateral risk-reducing mastectomy (BRRM) (...) During a mean follow-up of 10.3 years, 722 out of 1712 BRCA1 (...) and 406 out of 1145 BRCA2 (...) underwent BRRM. For BRCA1 mutation carriers hazard ratios were 0.40 for overall mortality (...) For BRCA2 mutation carriers (...) hazard ratio for overall mortality was 0.45

Detto semplificatamente, mortalità doppia fra le operate.

Incidentalmente, si noti che l'impossibilità nel caso specifico di fare uno studio prospettico randomizzato in doppio cieco limita addirittura intrinsecamente la possibilità di raggiungere una certezza sulla questione, se non assoluta almeno del livello di *gold standard* della Medicina e della Farmacia, che è appunto quel tipo di studio.

Ripetiamo: evitare medicalmente la malattia mortale non ha ridotto la mortalità. In quel caso l'ha raddoppiata.

BOZZA - DRAFT

[38] Heemskerk-Gerritsen, Bernadette A M et al. "Survival after bilateral risk-reducing mastectomy in healthy BRCA1 and BRCA2 mutation carriers." Breast cancer research and treatment vol. 177,3 (2019): 723-733. doi:10.1007/s10549-019-05345-2
in https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6745043/

Diffidare dai “ragionamenti a parole".

Non è logica ma pseudo-scienza.

Richiedere invece una dimostrazione razionale.
Per esempio nel caso di malattia prevenibile medialcalmente almeno uno studio prospettico randomizzato in doppio cieco su ampio campione, che mostri se ridurre l’incidenza di una malattia eventualmente mortale prevenibile medicalmente riduca effettivamente la mortalità, e non riduca solo la mortalità per quella specifica malattia al prezzo di aumentare la mortalità per tutte le cause, *all causes mortality*. Com’era nell’esempio sopra riportato.

DRAFT

Figure 2: Pillola blu: vivi tranquillo il tuo lavoro di Medico con la mentalità medica antica non-statistica (curare le malattie senza aggiornarti sugli articoli scientifici che esaminano se la terapia allunga o abbrevia la vita, oltre a curare la malattia, che è il *minimo sindacale*).
Pillola rossa: mentalità della Statistica Medica, ti aggiorni sugli articoli scientifici sull’effettiva utilità di un farmaco, al di là della sola guarigione da una malattia, per esempio l’aritmia che causa morte, la cui cura con un certo farmaco *aumentava* (!) la mortalità *per tutte le cause*.

Si noti che non è compito del Medico e del Farmacista fare studi randomizzati prospettici, e neppure solo osservazionali. Loro potrebbero seguire gli articoli scientifici di statistica medica, ma pochissimi

spiriti superiori lo fanno. Gli altri vengono informati dagli *informatori scientifici*, che gli dicono ogni bene del farmaco, sulla base magari di 1 solo articolo scientifico, scritto dai ricercatori della casa farmaceutica.

4.18 Prodotto cartesiano

Si chiama *prodotto cartesiano* di 2 insiemi X e Y l'insieme denotato con $X \times Y$ delle *coppie ordinate*[39] (x, y) con $x \in X$ e $y \in Y$. In simboli

$$X \times Y := \{(x, y) : x \in X \, et \, y \in Y\} .$$

L'insieme X può essere finito o infinito, e così pure Y.
Se X e Y sono finiti allora (teorema[40], ovvio)

$$\#(X \times Y) = (\#X) \cdot (\#Y)$$

cioè la cardinalità del prodotto cartesiano di 2 insiemi è il prodotto delle cardinalità dei 2 insiemi. Per esempio

$$\{a, b, c\} \times \{\alpha, \beta\} = \{(a, \alpha), (a, \beta), (b, \alpha), (b, \beta), (c, \alpha), (c, \beta)\}$$

ha $2 \cdot 3 = 6$ elementi, che sono coppie ordinate.
Similmente si definisce il prodotto cartesiano di 3 insiemi, composto dalle terne ordinate, e di n elementi, composto dalle n-uple; e si estende banalmente il soprascritto teorema sulle cardinalità.

BOZZA - DRAFT

4.19 Note sul prodotto cartesiano.

Osservato che il prodotto cartesiano di 2 insiemi finiti A e B ha (ovviamente) cardinalità

$$\#(A \times B) = \#A \cdot \#B$$

e con 3 insiemi finiti

$$\#(A \times B \times C) = \#A \cdot \#B \cdot \#C \text{ e similmente con } n \text{ insiemi}$$

[39] Una definizione rigorosa di coppia ordinata (x, y) è $\{\{x\}, \{x, y\}\}$, ma in questa trattazione supponiamo di per sè chiaro il concetto.

[40] Questo teorema è vero anche per insiemi infiniti ma allora la cardinalità risultante è infinita e non ha più il significato elementare qua considerato di *numero di elementi.*

osserviamo che parallelamente, per così dire, se un'azione si può fare in n modi e una seconda azione si può fare in m modi, la sequenza delle 2 azioni si può fare in $n \cdot m$ modi. E la sequenza di 3 azioni in $n_1 \cdot n_2 \cdot n_3$ modi, eccetera.

ESERCIZIO$_{\mu 2018}$

$*$ Calcolare quanti sono i diversi possibili 6-meri dei nucleotidi del RNA ovvero le sequenze ("parole") di 6 lettere dell'alfabeto

$$A, C, G, U$$

(di cui recentemente si è molto scritto riguardo la ricerca contro il cancro).

SVOLGIMENTO

Relativamente alla questione medica, e in futuro eventualmente farmaceutica, si può vedere sul sito governativo statunitense PubMed l'abstract dell'articolo scientifico 6mer seed toxicity in tumor suppressive microRNAs.

BOZZA DRAFT

Per fissare le idee (ma assolutamente non sarebbe necessario) scriviamo alcune delle sequenze/parole:

AAAAAA, AAAAAC, AAAAAG, AAAAAU,
AAAACA, AAAACC, AAAACG, AAAACU,
...
UUUUUA, UUUUUC, UUUUUG, UUUUUU.

Per calcolare il numero osserviamo che

la 1^ lettera può essere scelta in 4 modi
la 2^ lettera può essere scelta in 4 modi
...
la 6^ lettera può essere scelta in 4 modi

e allora la sequenza di 6 lettere può essere costituita in un numero di modi pari a

$$4^6 = 4 \cdot 4 \cdot 4 \cdot 4 \cdot 4 \cdot 4 =$$

$$\boxed{4096}$$

Oppure, si potrebbe anche dire che l'insieme delle sequenze/parole è in corrispondenza biunivoca con il prodotto cartesiano

$$\{A, C, G, U\} \times ...(\text{in tutto 6 volte})... \times \{A, C, G, U\}$$

e allora ha cardinalità

$$\Big(\#\{A, C, G, U\}\Big)^6$$

concludendo come prima.
Tutte le 4096 sequenze sono elencate in https://www.6merdb.org/

ESERCIZIO.$_\mu$ Consideriamo una serie di cassetti in una farmacia etichettati con un codice composto da una lettera (inglese) maiuscola o da una lettera seguita da una cifra (decimale). Quanti sono i possibili cassetti etichettabili?

SVOLGIMENTO
L'insieme dei codici di 2 caratteri è in corrispondenza biunivoca col prodotto cartesiano

$$\{A, B, ..., Z\} \times \{0, ..., 9\}$$

che ha $26 \cdot 10 = 260$ elementi. Considerando anche i 26 codici di un carattere (lettera maiuscola)

si hanno in tutto 286 possibili cassetti

BOZZA - DRAFT

4.20 ESERCIZI SULLA LEZIONE 4

4.20.1 Esercizio risolto a – Insieme di elementi chimici

$_{\mu 2021}$ ∗ Quanti sottoinsiemi non vuoti ha l'insieme {He, Ne, Ar, Kr, Xe, Ra, H_2O}?

SVOLGIMENTO
L'insieme ha 7 elementi (ovviamente in senso matematico, non stiamo conteggiando gli 8 *elementi chimici*) e allora ha $2^7 = 128$ sottoinsiemi di cui 1 è l'insieme vuoto.

$$\boxed{127}$$

BOZZA - DRAFT

4.20.2 Esercizio risolto b – Chimica combinatoria

$_{\mu}$ ∗ Supponiamo in via del tutto ipotetica che una certa molecola possa completarsi in un punto con uno qualunque fra gli elementi con numero atomico da 58 a 71 (lantanoidi diversi dal lantanio) e in altro punto con uno qualunque fra in primi 5 di essi. (Senza voler qua fare Farmacologia, si pensi comunque ai *farmaci chelanti*). In quanti modi può completarsi la molecola? E se aggiungiamo l'ipotesi che in ogni caso non si completa con due atomi uguali?

SVOLGIMENTO
I lantanoidi diversi dal lantanio, con numeri atomici

$$58, 59, ..., 70, 71$$

sono $71 - 57$ cioè 14.
La coppia ordinata di atomi atta a completare la molecola può costituirsi (prodotto cartesiano) in $14 \cdot 5 =$

$$\boxed{\text{a) } 70 \text{ modi}}$$

Se aggiungiamo l'ipotesi che la molecola non può completarsi con due atomi uguali, dobbiamo escludere 5 casi, restandone $70 - 5$ cioè

b) 65 modi

4.20.3 Esercizio risolto c – Schedine

$_{\mu}$ Quante sono le possibili schedine classiche, quelle di 13 risultati 1, 2 o X?

SVOLGIMENTO

L'1-esimo risultato può essere scelto in 3 modi: 1, 2 o X

il 2-esimo risultato può essere scelto in 3 modi: 1, 2 o X

...

il 13-esimo risultato può essere scelto in 3 modi: 1, 2 o X.

In tutto 3^{13} modi cioè

$$1\,594\,323.$$

BOZZA - DRAFT

5 Calcolo Combinatorio

Il Calcolo Combinatorio è costituito da metodi per determinare il numero di elementi di un insieme finito; ed elencarli, eventualmente.

In questa Lezione faremo dei modesti accenni alla Probabilità Combinatoria, basata in modo ovvio sul Calcolo Combinatorio.

POZZA DRAFT

5.1 Un po' di Calcolo Combinatorio in Farmacia

Vediamo alcuni dei molti casi di ricorrenza del Calcolo Combinatorio nella Farmacia.

• Il Calcolo Combinatorio ricorre nella chimica dei farmaci. Una rivista scientifica internazionale è Combinatorial Chemistry Link->. Leggiamo su Wikipedia, l'enciclopedia libera:

> Spesso il ricercatore si imbatte in un composto che dimostra una certa attività biologica, che però non è sufficiente per garantire il successo clinico (e commerciale) del composto. A questo punto inizia un processo di screening "quasi casuale": vengono preparati e testati tutti i possibili composti che mantengono una analogia strutturale per il nucleo fondamentale, ma ne differiscono per i sostituenti collegati.

• Il Calcolo Combinatorio ricorre anche nelle questioni di elencazione, catalogazione e archiviazione dei prodotti e delle attività di una farmacia. Per esempio il Calcolo Combinatorio risponderà alla domanda: in quanti modi posso ordinare 4 prodotti su una brochure pubblicitaria? Sono 24 modi, possono essere elencati facilmente, e poi assoggettati a considerazioni sanitarie e/o di marketing, per sceglierne uno per la stampa finale.

5.2 Introduzione a Calcolo e Probabilità Combinatori

La **probabilità combinatoria elementare** si basa su 2 cose:

• la concezione classica della probabilità, quella della probabilità come *casi favorevoli / casi possibili*;

• il calcolo combinatorio, costituito da metodi per determinare il numero di elementi di un insieme finito E, cioè la sua cardinalità, indicata con $\#E$ (e da altri con card$E$ o $|E|$); conteggiare gli elementi degli insiemi finiti adesso serve proprio per contare i casi favorevoli e i casi possibili di cui sopra.

In questa trattazione elementare, trattiamo il **calcolo combinatorio** – che di per sè sarebbe una matematica elementare – in 3 porzioni:

• una parte l'abbiamo già trattata nell'insiemistica: in particolare
  ◦ prodotto cartesiano
  ◦ insieme delle parti

• un'altra parte in questa lezione con accenni al Calcolo delle Probabilità:
  ◦ elencazione con conteggio
  ◦ cardinalità dell'unione
  ◦ permutazioni
  ◦ combinazioni
  ◦ disposizioni

• una terza parte in una lezione molto successiva, nell'ambito del Calcolo delle Probabilità:
  ◦ dismutazioni.

BOZZA - DRAFT

Per esempio col prodotto cartesiano, e l'elencazione con conteggio, si trova facilmente la probabilità che la somma dei punteggi di 2 dadi sia un numero primo:

$$p = \frac{\#\,casi\ favorevoli}{\#\,casi\ possibili\ equiprobabili} = \frac{15_{\leftarrow elencazione\ con\ conteggio}}{6 \times 6_{\leftarrow prodotto\ cartesiano}} = \\ = \frac{5}{12}$$

5.3 Elencazione con conteggio

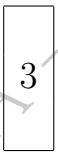
L'elencazione con conteggio è adeguata quando è più semplice dell' applicazione di formule. Per esempio per determinare quanti sono i numeri primi minori di 25

$$2, 3, 5, 7, 11, 13, 17, 19, 23 \qquad \to \textit{sono}\ 9$$

o (col computer) di 10 000.

Esempio di lencazione con conteggio – Anagrammi dorati
ESERCIZIO$_{\mu 2024}$ (R) $*$ Dire quanti sono (a prescindere dal senso eventuale) gli anagrammi della parola *oro*, ovvero in quanti modi si può riordinare il dataset

o, r, o



(L'*elencazione con conteggio* risolve[41] subito il problema, e qua lo applichiamo al dataset considerato:

r, o, o
o, r, o
o, o, r,

e si trova 3 – e similmente considerando gli anagrammi, a prescindere dal senso: *roo*, *oro*, *oor*).

BOZZA - DRAFT

[41] Errato è invece applicare la formula dell'$n!$, ora $3!$, del numero di permutazioni di un insieme di n elementi, perchè qua abbiamo un dataset di 3 elementi ma non un insieme di 3 elementi; gli elementi sono 2 – o, r – e non c'entra nulla, a causa della ripetizione di un elemento; sarebbe anche possibile applicare una formula – *formula delle permutazioni con ripetizione* – per trovare il risultato 3, ma più complessa dell'elencazione con conteggio, e ben poco nota).

Un esempio applicato al Calcolo delle Probabilità. La probabilità che un numero primo minore di 25 scelto a caso sia dispari è

$$\frac{8}{9} \approx 0.889 = 88.9\%$$

perché $\#\{2,3,5,7,11,13,17,19,23\} = 9$ e solo il 2 non è dispari, restandone 8.
Questo sopra, triplice, è il modo in cui in questa trattazione elementare in generale esprimeremo la probabilità, ma si noti che il secondo (0.889), tipico della Matematica, è poco usato nella pratica della Farmacia. Anzi, si faccia attenzione a non confondere 0.889 con 0.889%, che è la centesima parte.

BOZZA - DRAFT

5.4 Cardinalità dell'unione

Per gli insiemi finiti, dei quali soli ora ci occupiamo, vale (teorema) questa (ovvia: si disegnino i diagrammi di Eulero-Venn) formula (che lega le cardinalità dell'unione e dell'intersezione):

$$\#(A \cup B) = \#A + \#B - \#(A \cap B)$$

In particolare se A e B sono disgiunti (l'intersezione ha 0 elementi) la cardinalità dell'unione è la somma della cardinalità.

ESERCIZIO$_{\mu 2018}$
* Supponiamo che di 116 individui
con anticorpo VISα oppure VISγ,73 hanno l'anticorpo VISα, e 53 l'anticorpo VISγ. Quanti hanno entrambi gli anticorpi?

SVOLGIMENTO
Sappiamo che per gli insiemi finiti la numerosità ovvero cardinalità degli insiemi unione e intersezione vale

$$\#(A \cup B) = \#A + \#B - \#(A \cap B)$$

cioè equivalentemente vale

$$\#(A \cap B) = \#A + \#B - \#(A \cup B)$$

e ora con

A = {soggetti con VISα}
B = {soggetti con VISγ}
si ha

numero di soggetti con entrambi gli anticorpi =

$$= \#(A \cap B) = \#A + \#B - \#(A \cup B) =$$

$$= 73 + 53 - 116 =$$

$$\boxed{10}$$

BOZZA - DRAFT

5.5 Fattoriale

Di amplissima ricorrenza nel calcolo combinatorio è il fattoriale.

Figure 3: Bing Image Creator, rielaborata con l'aggiunta di testo.

BOZZA

Si ricordi per fissare le idee

$$3! = 1 \cdot 2 \cdot 3 = 6$$

eventualmente anche col seguente mnemonico:

Nùmer sei fattorial tu sei di tre.

(È un endecasillabo *a maiore* e gioca sul doppiosenso "sei").

☺ *All'università ho scoperto che* $2+4$ *è uguale a* $3!$

Si noti la grande rapidità con cui crescono i fattoriali:
1, 1, 2, 6, 24, 120, 720, 5040....
Addirittura 100! ha già 158 cifre.

Ne riparleremo a proposito delle *successioni*.

5.6 Permutazioni

Ogni ordinamento totale di un insieme finito non vuoto A si dice *permutazione* degli elementi di A. Esso viene identificato con l'unica $(\#A)$-upla di elementi di A che "rappresenta" quell'ordinamento totale. Per esempio dall'unica 5-upla – come (b, d, a, e, c) – che rappresenta un determinato ordinamento totale di A se esso ha 5 elementi.
Il numero delle permutazioni di un insieme di n elementi si indica talvolta con P_n e si pone anche $P_0 := 1$.
Il numero delle permutazioni di un insieme di n elementi è dato dal *fattoriale*[↑] di n, che vale 1 se $n = 0$ e altrimenti $1 \cdot 2 \cdot \ldots \cdot n$:

$$P_n = n! \, .$$

Per esempio 4 prodotti possono essere pubblicizzati ciascuno su 1 di 4 pagine di una brochure in 4! cioè 24 ordini diversi.

BOZZA - DRAFT

Esempio. In quanti ordini diversi si possono mettere in un contenitore di acqua 8 sostanze chimiche diverse?
Che probabilità c'è di ordinarle in un particolare fissato modo mettendole a caso?

È un problema di permutazioni, dell'insieme $\{x_1, \ldots, x_8\}$ o più semplicemente $\{1, \ldots, 8\}$ delle sostanze considerate, che ha 8 elementi. Gli elementi si possono riordinare in $8! = 40\,320$ modi. La probabilità è $1/43\,320 \approx 0.0000231 = 0.00231\%$

Esercizio. In quanti modi diversi si possono allineare 5 prodotti farmaceutici su un espositore?

5.7 Combinazioni semplici

Consideriamo un insieme di $n = 7$ elementi e scegliamone $k = 3$. In quanti modi si può fare? La risposta è data dalle *combinazioni (semplici)* che ora vedremo.
(Ma se anche riteniamo ordinata la terna scelta, si può fare in molti più modi, $3! = 6$ volte tante, e sono le *disposizioni semplici*, considerate nel successivo paragrafo 5.8).

Combinazioni semplici. Dato un insieme E di $n > 0$ elementi (cioè non vuoto) e un numero naturale k tale che $0 \le k \le n$, i sottoinsiemi di E di k elementi si chiamano *combinazioni (semplici)* di n oggetti a k a k (per esempio a 3 a 3) e il numero di esse si chiama *coefficiente binomiale*, e si indica con $C_{n,k}$ o $\binom{n}{k}$ e si pone per convenzione $C_{n,0} := 1$. Questo numero si calcola col Triangolo di Tartaglia$_{\downarrow}$ oppure con una formula di esso

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

BOZZA - DRAFT

Nota. Numeri enormi li lasceremo indicati, per esempio scriveremo $\binom{365}{23} =$

$$= \frac{365!}{23!\,342!}$$

senza procedere ulteriormente nel calcolo.

5.8 Disposizioni semplici

Disposizioni semplici. Dato un insieme E di $n > 0$ elementi (cioè non vuoto) e un numero naturale k tale che $0 \le k \le n$, i sottoinsiemi ordinati di E di k elementi si dicono *disposizioni (semplici)* di n oggetti a k a k (per esempio a 3 a 3), e il numero di esse si indica con $D_{n,k}$ e si pone per convenzione $D_{n,0} := 1$.
è (teorema)

$$D_{n,k} = \frac{n!}{(n-k)!}\,.$$

Mnemonici:

co mbinazioni... **co**n $k!$ e **co**sì **co**mpare numero **co**rto

di spo sizioni... **di**menticati $k!$ e **dispo**ni ordinatamente

Esempi. 3 elementi ordinati si possono scegliere da 7 in $\frac{7!}{(7-3)!} = \frac{1\cdot2\cdot3\cdot4\cdot5\cdot6\cdot7}{1\cdot2\cdot3\cdot4} = 5\cdot6\cdot7 = 210$ modi, e senza riguardo all'ordine in $\binom{7}{3} = \frac{1\cdot2\cdot3\cdot4\cdot5\cdot6\cdot7}{1\cdot2\cdot3\cdot1\cdot2\cdot3\cdot4} = 35$ modi. (Riordinabili in $3!\cdot35 = 210$ modi). Si provi a elencare i 35 modi. (Si fissi per esempio $E := \{1,2,3,4,5,6,7\}$).

Esercizio. In quanti modi si possono scegliere 4 elementi chimici diversi della tavola periodica ("classica") di 92 elementi? E quante sono le quaterne ordinate di 4 elementi?

BOZZA - DRAFT

Per il primo quesito, sono

$$\binom{92}{4} = \frac{92!}{4!\,(92-4)!} = \frac{92!}{4!\,88!} = \frac{92\cdot91\cdot90\cdot89}{1\cdot2\cdot3\cdot4} =$$

$$= 23\cdot91\cdot15\cdot89 = 2\,794\,155$$

(Si provi con Wolframalpha `Binomial[92,4]`).
Per il secondo quesito, sono

$$D_{92,4} = \frac{92!}{(92-4)!} = \frac{92!}{88!} = 92\cdot91\cdot90\cdot89 =$$

$$= 67\,059\,720$$

Esercizi. In quanti modi si possono scegliere 23 giorni diversi del 2018? Con o senza significato, quante "parole" di 4 lettere si possono comporre con le lettere C, R, O, N, I, S, T, A?

Esempio sulle combinazioni semplici. Che probabilità c'è di vincere giocando una cinquina su una ruota del lotto? Per essere sicuri di vincere giochiamo 1 euro su ciascuna di esse. Quanto guadagniamo?
Di casi favorevoli ce n'è 1, e i casi possibili sono le combinazioni

semplici di 90 oggetti a 5 a 5, che sono in numero di $\binom{90}{5}$, e allora la probabilità è

$$p = \frac{1}{\binom{90}{5}} = \frac{1}{\frac{90!}{5!(90-5)!}} = \frac{5!\,85!}{90!} = \frac{1\cdot\ldots\cdot 5\cdot 1\cdot\ldots\cdot 85}{1\cdot\ldots\cdot 90} =$$

$$= \frac{1\cdot 2\cdot 3\cdot 4\cdot 5}{86\cdot 87\cdot 88\cdot 89\cdot 90} = \frac{1}{86\cdot 87\cdot 22\cdot 89\cdot 3} = \frac{1}{43\,949\,268}\,.$$

Abbiamo speso 43 949 268 euro e 1 cinquina da noi giocata vince; poiché si vince 6 milioni di volte la posta, vinciamo 6 milioni di euro, con un guadagno di −37 949 268 euro. (Ciclopica perdita) ☹

Esempio sulle disposizioni semplici. Quanti sono i numeri esadecimali di 5 “cifre" (da 00000 a FFFFF) con le “cifre" tutte diverse?
è un problema di disposizioni semplici. Consideriamo i sottoinsiemi ordinati di { 0, 1,..., 9, A,..., F } di 5 elementi, che – in base al teorema – sono in numero di

$$D_{16,5} = \frac{16!}{(16-5)!} = \frac{16!}{11!} =$$

$$= \frac{1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6\cdot 7\cdot 8\cdot 9\cdot 10\cdot 11\cdot 12\cdot 13\cdot 14\cdot 15\cdot 16}{1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6\cdot 7\cdot 8\cdot 9\cdot 10\cdot 11} = 524\,160.$$

BOZZA - DRAFT

Complementi

5.9 Complementi – Triangolo di Tartaglia e potenze

il *Triangolo di Tartaglia* (o *di Pascal)*, è la *scrittura*

$$
\begin{array}{ccccccc}
 & & & 1 & & & \\
 & & 1 & & 1 & & \\
 & 1 & & 2 & & 1 & \\
1 & & 3 & & 3 & & 1 \\
 & & & \ldots\ldots\ldots & & &
\end{array}
$$

variamente estesa con la Formula di Pascal

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$$

(e cioè ogni numero è la somma dei 2 soprastanti, a parte gli 1 ai margini) dove n è il numero di riga a partire dalla 0-esima e k è la posizione nella riga a partire dalla 0-esima.
Il Triangolo di Tartaglia consente il calcolo dei *coefficienti binomiali* $\binom{n}{k}$ mediante sole somme. Ciò è utile sia nel problema delle combinazioni semplici che nella *potenza del binomio*

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k\, b^{n-k} \qquad \text{ovvero}$$

$$= \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \binom{n}{3}a^{n-3}b^3 + \ldots + \binom{n}{n}b^n$$

(si noti che nella seconda formula gli a e b nel secondo membro sono scambiati fra loro rispetto alla prima formula, come se là fosse scritto $a^{n-k}\,b^k$, ma è equivalente). I vari coefficienti dei monomi si trovano semplicemente sulla n-esima riga del Triangolo di Tartaglia (intendendo come 1-esima quella con due 1), per esempio $(a+b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$.

BOZZA - DRAFT

5.10 Complementi – Costante di Lieb

Vediamo un notevole risultato di Calcolo Combinatorio, applicato alla Chimica.
Traiamo da Wikipedia, l'enciclopedia libera.

> In statistical mechanics, the ice-type models or six-vertex models are a family of vertex models for crystal lattices with hydrogen bonds. The first such model was introduced by Linus Pauling in 1935 to account for the residual entropy of water ice.
>
> An $n \times n$ grid graph (with periodic boundary conditions and $n \geq 2$) has n^2 vertices and $2\,n^2$ edges; it is 4-regular, meaning that each vertex has exactly four neighbors. An orientation of this graph is an assignment of a direction to each edge; it is an Eulerian orientation if it gives each vertex exactly two incoming edges and exactly two outgoing edges.
>
> Denote the number of Eulerian orientations of this graph by $f(n)$.

BOZZA - DRAFT

È ben evidente che $f(n)$ tende all'infinito. Ma quanto rapidamente? Precisamente

$$\lim_{n\to\infty} \sqrt[n^2]{f(n)} = \left(\frac{4}{3}\right)^{\frac{3}{2}} = \frac{8\sqrt{3}}{9} = 1.5396007\ldots$$

che si chiama *costante di Lieb del ghiaccio quadrato*.

5.11 ESERCIZI SULLA LEZIONE 5

5.11.1 Esercizio risolto a – Gruppo di studenti

$_{\mu 2024}$ (R) $*$ In quanti modi diversi si può scegliere un sottogruppo di 4 studenti fra 5 studenti? (O di 4 farmaci fra 5, o di 4 bilance fra 5).

$$\boxed{5}$$

(Le scelte possibili sono tante quanti i modi di scegliere un unico studente da escludere dal gruppo dei 5 per costituire il sottogruppo di 4 e questo si può fare ovviamente in 5 modi; più tecnicamente si tratta del numero di *combinazioni* di 5 elementi a 4 a 4 che sono in numero di

$$\binom{5}{4} = \frac{5!}{4!\,(5-4)!} = \frac{1\cdot 2\cdot 3\cdot 4\cdot 5}{1\cdot 2\cdot 3\cdot 4\cdot 1} = 5$$

e addirittura, in questo caso, anche l'elencazione con conteggio sarebbe *al limite accettabile* come metodo risolutivo).

BOZZA – DRAFT

5.11.2 Esercizio da risolvere – Numeri scelti a caso

Che probabilità c'è che un numero < 10 preso a caso sia primo? E < 25? E nei 2 casi, che sia quadrato? Triangolare? Pari?

5.11.3 Esercizio tanto carino quanto difficile

Che probabilità c'è di vincere giocando il terno 1, 2, 3 su una ruota del lotto, per esempio Venezia?

II – Funzioni, Algebra e Piano Cartesiano

BOZZA - DRAFT

6 Funzioni

Una legge che ad ogni elemento di un insieme D detto *dominio* associa 1 elemento di un insieme C detto *codominio* si chiama *funzione* definita in D a valori in C:

$$f : D \to C$$

Ad ogni funzione daremo un nome, di solito di 1 lettera[42], per esempio f oppure y oppure a, e la *variabile indipendente*, che varia nel dominio, la indicheremo con un nome qualsiasi, di solito di 1 lettera, per esempio scriveremo

$$f(x) \coloneqq x^2$$

e poi sarà per esempio $f(2y) = 4y^2$, e $f(\alpha + 1) = \alpha^2 + 2\alpha + 1$ con variabili y e poi α, mentre la f è la funzione di prima.

Al di fuori della Matematica, per esempio in Chimica e soprattutto in Informatica, è normale dare alle funzioni nomi di più lettere, per esempio pH, `STIPENDIO`, `IVA_2024`.

Esempi di funzioni non da numeri a numeri:

$$f : \text{ essere umano } \mapsto \text{ suo codice fiscale (quello italiano)}$$

$$g : \text{codice fiscale} \mapsto \text{numero di acquisti eseguiti nella farmacia}$$

$$h : \text{codice fiscale} \mapsto \text{volume degli acquisti eseguiti nella farmacia}$$

BOZZA - DRAFT

6.1 Funzioni Numeriche

Le funzioni numeriche – in questa trattazione elementare – sono funzioni con dominio e codominio numerico: cioè a certi numeri associano certi numeri.

6.2 Passaggio a opposto, reciproco, e valore assoluto

6.2.1 Passaggio all'opposto

$a(x) \coloneqq -x$ *passaggio all'opposto*, $x \mapsto -x$. (Definita da $\mathbb{Z}$ in poi). Questo meno non indica affatto negatività ma passaggio all'opposto, per esempio l'opposto di -3 è il positivo 3. (Si noti allora che $-x$ può essere positivo).
Da un punto di vista fisico, il tempo $-2\,h$ significa 2 ore *prima* del tempo 0 dell'inizio di un esperimento.

[42] Una funzione con nome di più lettere, anche con la variabile di più lettere, potrebbe per esempio essere $area(lato) = lato^2$, che dà l'area del quadrato in funzione del lato.

6.2.2 Passaggio al reciproco

$b(x) := \frac{1}{x}$ *passaggio al reciproco*, $x \mapsto \frac{1}{x}$. Notazione deprecabile: x^{-1}. Definita per i numeri diversi da 0 da $\mathbb{Q}$ in poi.

Figure	su WolframAlpha `Link->`	su WolframAlpha `Link->`

6.2.3 Valore assoluto

absolute value

`|x|` oppure `abs(x)` su WolframAlpha

in $\mathbb{R}$ si definisce la funzione

$$|x| := \begin{cases} x & se\ x \geq 0 \\ -x & se\ x < 0 \end{cases}$$

(Definibile in vari modi[43] equivalenti. Si noti che potrebbe essere definita già in Z e Q).

BOZZA - DRAFT

6.2.4 Parte intera di un numero non negativo

floor function

La parte intera $\lfloor x \rfloor$ di $x \geq 0$ si ottiene da x togliendogli tutte le cifre decimali diverse da 0 se ne ha.
Un esempio chiarificatore di tutti i casi:

$$\lfloor \pi \rfloor = 3$$

`floor(Pi)` su WolframAlpha

[43] Ecco 4 modi in cui può essere definito il valore assoluto in $\mathbb{Z}$ o $\mathbb{R}$:

$|x| := \sqrt{x^2}$

$|x| := x\,\mathrm{sgn}(x)$

$|x| := \max\{x, -x\}$

Definizione consigliabile:

$$|x| := \begin{cases} x & se\ x \geq 0 \\ -x & se\ x < 0 \end{cases}$$

La soprastante parentesi graffa (grande) è *"selettiva dei casi"*, ma quel simbolo ha anche altri usi affini.
In $\mathbb{R}$ il valore assoluto è una *funzione elementare*, mentre nell'algebra dei numeri può considerarsi *operazione unaria*.

(Abbiamo tolto le cifre decimali 1,4...).
Questa funzione viene applicata per esempio troncando le età agli "anni compiuti", per cui a 17 anni e 9 mesi ovvero 17.75 anni uno viene classificato come avente 17 *anni compiuti* (non 18, che sarebbe l'arrotondamento).
Una definizione analitica di $\lfloor x \rfloor$ esiste ma non è semplicissima.
Per i numeri negativi la parte intera è fonte di ambiguità fra i vari Autori e non ce ne occuperemo: la parte intera di $-\pi$ sarà -3 o -4?

6.2.5 Segno di un numero

Da $\mathbb{Z}$ in poi si può definire la funzione segno

$$\operatorname{sgn}(x) := \begin{cases} -1 & se\ x < 0 \\ 0 & se\ x = 0 \\ 1 & se\ x > 0 \end{cases}$$

che normalmente però viene definita[44] in $\mathbb{R}$; per esempio è $\operatorname{sgn}(3-\pi) = -1$.

Valore assoluto e segno sono correlate dall'identità $x \equiv |x| \cdot \operatorname{sgn}(x)$.

Figure	su WolframAlpha Link->	su WolframAlpha Link->

BOZZA - DRAFT

6.3 Altri esempi di funzioni numeriche

Ecco alcuni esempi di funzioni definite in $\mathbb{R}$, o, volendo, suoi sottoinsiemi, e a valori in $\mathbb{R}$, definite con le 4 operazioni e altre:

$a(x) := (1 - 20\%)\, x = 0.8\, x$ scontare del 20% un prezzo x

$b(x) := (1 + 20\%)\, x = 1.2\, x$ aumentare del 20% un prezzo x

$f(x) := 3\, x$ triplicare, per esempio il dosaggio, o un prezzo

[44] Con qualche ambiguità: per alcuni non è definita in 0. In questa trattazione è definita in 0 e vi vale 0, che è lo standard ISO.

$g(x) \coloneqq \frac{2}{3}\,x$ ridurre di un terzo ovvero del $33.\overline{3}\%$, $\approx 33.3\%$, per esempio un dosaggio, ovvero scontare un prezzo del 33% (circa)

$h(x) \coloneqq \frac{x}{3}$ ridurre a un terzo ovvero al $33.\overline{3}\%$, $\approx 33.3\%$

$k(x) \coloneqq \frac{4}{3}\,x$ aumentre di un terzo ovvero del $33.\overline{3}\%$, $\approx 33.3\%$

$m(x) \coloneqq 2.4\,x$ portare al 240% del valore iniziale ovvero aumentare del 140%

$r(x) \coloneqq x^3$ elevare alla terza ovvero al cubo

$s(x) \coloneqq \sqrt[3]{x}$ estrarre la radice terza ovvero cubica, che vedremo

$u(x) \coloneqq \sqrt{x}$ estrarre la radice quadrata, dom$u$: $x \geq 0$

BOZZA - DRAFT

Vedendo la difficoltà che hanno alcuni
nel distinguere espressioni come
ridurre di un terzo e *ridurre a un terzo*
ci vorrebbe forse molta cautela nell'usarle
per descrivere una variazione di posologia
farmacologica (N.d.S.)

Ma pensateci bene: preferireste
che una multa venga ridotta *di* un decimo, o
che venga ridotta *a* un decimo?

6.4 Immagine, controimmagine, composta

In questo paragrafo consideriamo sia funzioni numeriche che funzioni non numeriche.

Si definiscono 2 insiemi e 1 funzione:

• $\mathrm{im} f := \{f(x) \in \mathrm{codom} f | x \in \mathrm{dom} f\}$ immagine di f

• $f^{-1}(E) := \{x \in \mathrm{dom} f | f(x) \in E\}$ controimmagine di un sottoinsieme E di $\mathrm{codom} f$ (ma attenzione alla simbologia del $^{-1}$)

• $f(g(x))$ la funzione composta: si "f-izza" la "g-izzazione" degli elementi, per esempio con $a(t) := t^2$ e $b(t) := t + 1$ si ha

$$a(b(t)) = (t+1)^2 \qquad b(a(t)) = t^2 + 1$$

6.5 Funzione iniettiva, suriettiva, biiettiva, inversa

In questo paragrafo consideriamo sia funzioni numeriche che funzioni non numeriche.

BOZZA - DRAFT

Funzione **iniettiva**: "mai 2 vanno in 1". Per esempio il codice fiscale, o il codice cliente della tessera di fidelizzazione di una farmacia.
Tornando a funzioni numeriche, per esempio x^3 e 2^x ma non x^2 intesa come funzione $\mathbb{R} \to \mathbb{R}$ (che è il modo naturale di considerarla).

Figure	su WolframAlpha Link->	su WolframAlpha Link->

Funzione **suriettiva**: "riempie il codominio". Per esempio x^3 ma non x^2 nè 2^x intese come funzioni $\mathbb{R} \to \mathbb{R}$ (che è il modo naturale di considerarle).

Figure	su WolframAlpha Link->	su WolframAlpha Link->

Funzione ***biiettiva***: iniettiva *et* suriettiva. Per esempio x^3 ma non x^2 nè 2^x intese come funzioni $\mathbb{R} \to \mathbb{R}$ (che è il modo naturale di considerarle).

Figure	su WolframAlpha Link->	su WolframAlpha Link->

La funzione biiettiva dà luogo in modo ovvio alla *funzione inversa*. Per esempio $\sqrt[3]{x}$ è la funzione inversa di x^3. E viceversa.

Figure	su WolframAlpha Link->

Una funzione

$$f : D \to C$$

biunivoca si può vedere anche come *relazione* fra dominio e codominio e in quel senso si chiama ***corrispondenza biunivoca***

$$f$$

$$D \leftrightarrow C$$

Due insiemi in corrispondenza biunivoca fra loro possono venire identificati a livello astratto.
Esempi:

una la retta orientata $\leftrightarrow$ i numeri reali cioè $\mathbb{R}$
il piano cartesiano $\leftrightarrow$ le coppie di numeri reali ovvero $\mathbb{R}^2$
i 5034 clienti di una certa farmacia $\leftrightarrow$ i numeri di $\{1, 2, 3, ..., 5034\}$
i 26 532 prodotti in vendita $\leftrightarrow$ le stringhe COD00001,...,COD26532

BOZZA - DRAFT

6.6 Crescenza e decrescenza delle funzioni numeriche

Crescenza (globale) e decrescenza (globale) su intervalli, eventualmente sull'intero dominio di una funzione.
Ci sono 4 casi di cui più importanti il 1° e il 3°.

1) se $x_1 < x_2 \Rightarrow f(x_1) < f(x_2)$ allora f si dice crescente.
Esempi: x^3, arctan, lg, ln, exp, $\sqrt{x}$. Anche x^2 per $x \geq 0$.
"A valore minore fa corrispondere valore minore"
Il grafico non può avere tratti orizzontali.

2) se $x_1 < x_2 \Rightarrow f(x_1) \leq f(x_2)$ allora f si dice non decrescente, o crescente in senso debole (ma alcuni Autori la dicono crescente).
Esempio: $\lfloor x \rfloor$, parte intera di x.
"A valore minore fa corrispondere valore minore o uguale"
Il grafico può avere tratti orizzontali.

3) se $x_1 < x_2 \Rightarrow f(x_1) > f(x_2)$ allora f si dice decrescente.
Esempio: e^{-x}. Anche $\frac{1}{x}$ per $x > 0$. Anche x^2 per $x \leq 0$.
"A valore minore fa corrispondere valore maggiore"
Il grafico non può avere tratti orizzontali.

4) se $x_1 < x_2 \Rightarrow f(x_1) \geq f(x_2)$ allora f si dice non crescente, o decrescente in senso debole (ma alcuni Autori la dicono decrescente).
Esempio: $-\lfloor x \rfloor$
"A valore minore fa corrispondere valore maggiore o uguale"
Il grafico può avere tratti orizzontali.

Nota 1. x^2, $\sin x$ e $\cos x$ non ricadono in alcuna delle 4 categorie.

Teoremi (ovvi)

Crescente su intervallo $\Rightarrow$ iniettiva.

Decrescente su intervallo $\Rightarrow$ iniettiva.

Nota 2. Esiste anche un'altro tipo di crescenza e decrescenza: quella puntuale (invece che sugli intervalli). Essa corrisponde alla crescenza o decrescenza (della funzione rappresentativa) della retta tangente al grafico, questione che vedremo in seguito.

BOZZA - DRAFT

6.7 Ambiguità notazionale dell'esponente −1

😕

Una delle più gravi ambiguità notazionali è quella dell'esponente −1, che ha sostanzialmente 3 significati: reciproco, inversa, controimmagine. Attenzione!

Certamente

a^{-1}

indica inequivocabilmente il reciproco $\frac{1}{a}$ di a se a è un numero, ma se è una funzione – e può ben esserlo – cominciano problemi:

$$f^{-1}(x)$$

in questa trattazione significherà sempre l'inversa di $f(x)$ mentre la reciproca la denoteremo $\frac{1}{f(x)}$ e casomai, volendoci del male, $(f(x))^{-1}$. Il problema maggiore si ha quando – e concretamente ciò viene fatto spesso – in lunghi calcoli per brevità di scrittura si omette di indicare la variabile indipendente, e si scrive f intendendo $f(x)$, lasciando nell'incertezza riguardo la scrittura f^{-1}. Così all'atto pratico ci si ritrova sempre con qualcuno che pensa che l'arcotangente sia il reciproco della tangente, o il logaritmo il reciproco dell'esponenziale (invece è l'inversa).
Resta il fatto che in un testo diverso da questo, la scrittura

$$\ln^{-1}(x)$$

è ambigua, e può indicare sia $\exp(x)$ che $\frac{1}{\ln(x)}$. In questo testo solo $\exp(x)$, che è l'inversa del logaritmo.

Inoltre, l'esponente -1 è usato per indicare anche tutta un'altra cosa, la controimmagine, che non è un elemento ma un insieme.

ROZZA - DRAFT

6.8 Ambiguità notazionali della parentesi graffa grande

😕

La parentesi graffa grande viene usata con almeno 3 diversi significati.

Li illustreremo con 3 esempi.
Sempre c'è un *et* "retrostante", ma non in modo immediato.

Parentesi graffa (grande) *dei sistemi*, con significato puro e semplice di *et*:

$$\begin{cases} x^2 > 9 \\ x + 1 < 0 \end{cases}$$

Parentesi graffa (grande) *"selettiva dei casi"*:

$$|x| := \begin{cases} x & se\ x \geq 0 \\ \\ -x & se\ x < 0 \end{cases}$$

Parentesi graffa (grande) *"onnicomprensiva dei casi"*:

$$\cos(2x) = \begin{cases} \cos^2 x - \sin^2 x \\ 2\cos^2 x - 1 \\ 1 - 2\sin^2 x \end{cases}$$

(Nell'esempio sono affermate con scrittura compatta 3 uguaglianze).

6.9 Funzioni definite per numeri naturali: le successioni

Per la variabile indipendente spesso useremo i nomi n, m, h, k se il dominio è un insieme di numeri interi, per esempio $\mathbb{N}$. In questo caso potremmo dare a tali funzioni nomi $y(n)$ o $a(n)$ ma più spesso useremo le notazioni y_n, a_n e in ogni caso le funzioni definite su $\mathbb{N}$ (o anche su una sua semiretta e cioè per $n \geq n_0$) le chiameremo ***successioni***.

BOZZA DRAFT

Una successione (definita[45] *analiticamente*) già vista (in 5.5) è quella dei fattoriali:

$$1, 1, 2, 6, 24, 120, 720, 5040...$$

Figura su WolframAlpha Link->

[45] Per ogni $n \in \mathbb{N}$, si dice *fattoriale* di n e si indica con $n!$ il numero $1 \cdot 2 \cdot ... \cdot n$ se $n > 0$ e 1 se $n = 0$.
Per esempio $5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$; e $11! = 39\,916\,800$.
Esistono estensioni della definizione ai numeri reali e perfino ai numeri complessi mediante la funzione Gamma:

$$z! := \Gamma(z-1), \quad z \neq -1, -2, -3, ...$$

Valgono le approssimazioni di Stirling

$$x! \sim \sqrt{2\pi x}\left(\frac{x}{e}\right)^x$$

e con maggiore precisione

$$x! \sim \sqrt{2\pi x}\left(\frac{x}{e}\right)^x\left(1 + \frac{1}{12x}\right)$$

per i numeri interi o reali, purché sufficientemente grandi – diciamo maggiori di 8 per avere un'approssimazione all'1% per la prima e maggiori di 1 per avere un'approssimazione all'1 per mille per la seconda.
Si è usato il simbolo $\sim$ invece di $\approx$ perché sono approssimazioni asintotiche – questione sottile.

e si noterà che

si passa da 1 a 2 moltiplicando per 2

si passa da 2 a 6 moltiplicando per 3

si passa da 6 a 24 moltiplicando per 4

e così via:

$$n! := \begin{cases} 1 \cdot 2 \cdot \ldots \cdot n & se \quad n > 0 \\ 1 & se \quad n = 0 \end{cases}$$

ovvero, per ricorrenza, $a_0 := 1$, $a_n := n\, a_{n-1}$ per $n > 0$.

Un'infinità di dati epidemiologici si possono inquadrare come successioni (definite *empiricamente)*, con l'indice che è l'anno, per esempio da 1862 in poi e teoricamente estendibile all'infinito almeno nell'immaginazione, e i valori a rappresentare il numero di nati, di morti, la mortalità infantile, e quant'altro:

$x_{1862} = \ldots$

$x_{1863} = \ldots$

...

BOZZA - DRAFT

Esempio 1. Morti in Italia:

2011 609 000
2012 625 000
2013 611 000
2014 609 000
2015 656 000
2016 627 000
2017 660 000
2018 641 000
2019 645 000
2020 746 000
2021 709 000

Esempio 2. Morti di morbillo in Inghilterra e Galles:

https://www.gov.uk/government/publications/measles-deaths-by-age-gro
measles-notifications-and-deaths-in-england-and-wales-1940-to-2013

Oltre alla successione x_{2011}, x_{2014},... (di fatto inevitabilmente troncata) del numero di morti, molto interessante sarebbe rappresentare anche la successione derivata degli incrementi – eventualmente negativi e cioè di fatto diminuzioni – annuali:

$$z_n := x_n - x_{n-1}$$

Spesso in Medicina una tale differenza viene denotata con Δ, per esempio nell'andamento di un qualche parametro di un'epidemia. Δ 24h può indicare l'incremento (eventualmente negativo)

$$x_{oggi} - x_{ieri}.$$

Spesso sia a livello di rappresentazione grafica che a livello teorico, la successione viene sostituita con una (funzione) interpolante: detto semplificatamente, si "congiungono i puntini" ottenendo una funzione definita su una semiretta di $\mathbb{R}$ invece che su una semiretta di $\mathbb{N}$. (O su un intervallo di $\mathbb{R}$ invece che di $\mathbb{N}$).

BOZZA - DRAFT

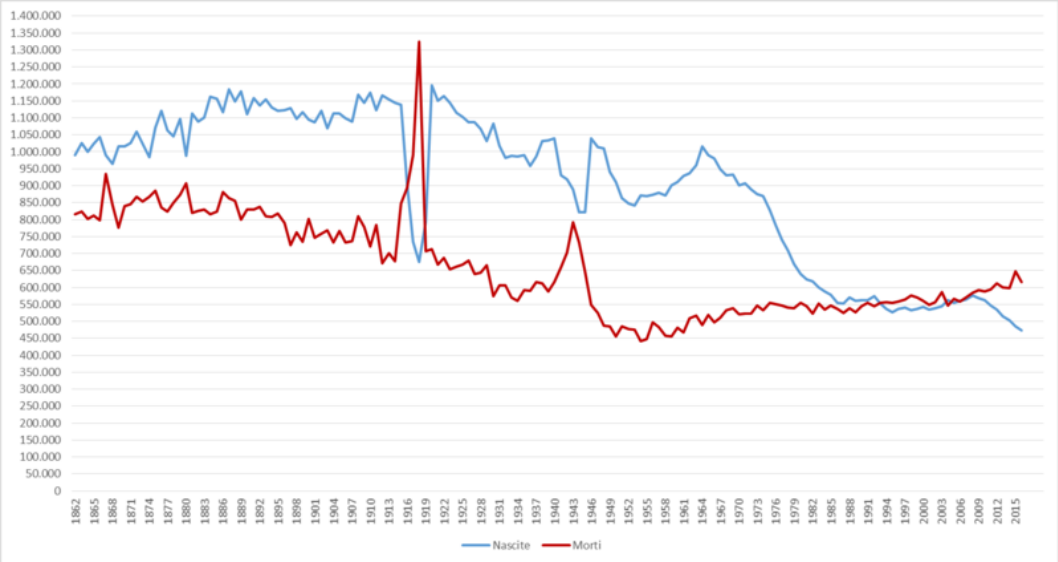


Figure 4: Nascite (blu) e morti (rosso) in Italia tra il 1862 e il 2016. Da https://commons.wikimedia.org/wiki/File:Nascite_e_morti_in_Italia,_1862_-_2016.png

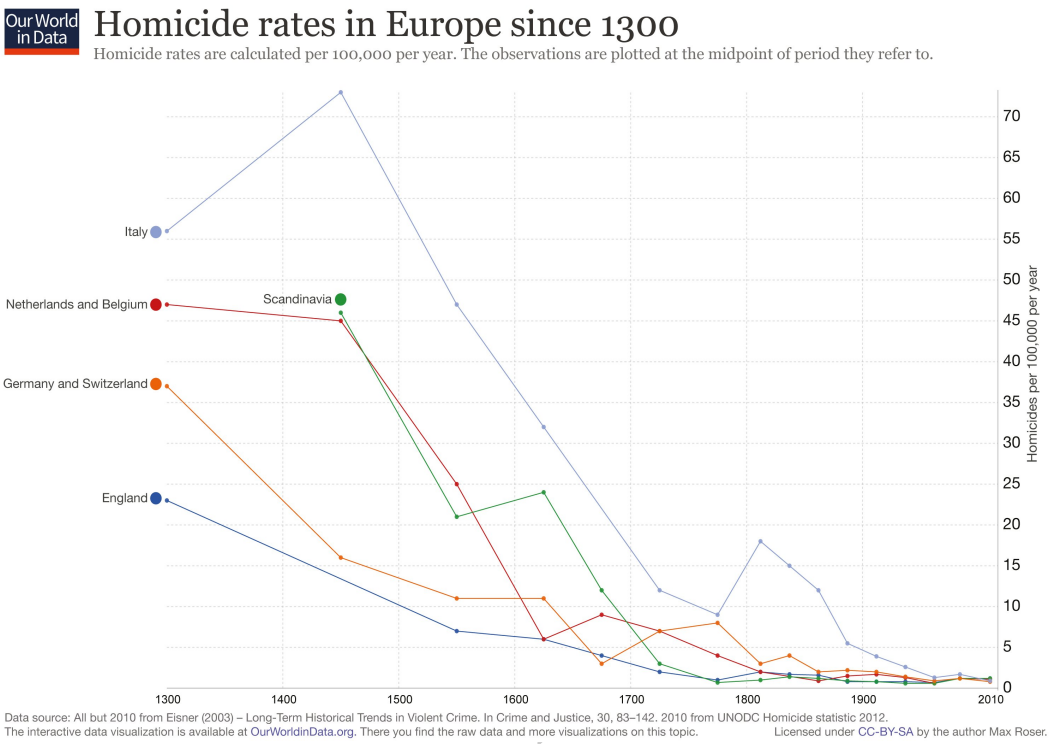


Figure 5: Congetturato tasso di omicidi negli ultimi 7 secoli.

BOZZA

6.10 La successione di Fibonacci

La successione di Fibonacci inizia con 0, 1... oppure con 1, 1... e poi ogni termine è somma dei 2 precedenti:

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55... \text{ oppure} \tag{9}$$

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55... \tag{10}$$

Essa è in qualche modo correlata al tipo di ampliamento di una popolazione di organismi – fu studiata da Fibonacci nel medioevo per modellizzare l'accrescimento di una popolazione di conigli – ovvero anche all'espansione di un'epidemia nella fase iniziale.

Figura | su WolframAlpha `Link->`

BOZZA - DRAFT

È di solito definita per ricorrenza:

$$\begin{cases} a_0 := 0 \\ a_1 := 1 \\ a_n := a_{n-1} + a_{n-2}, \quad n = 2, 3, \ldots \end{cases}$$

Poco nota e meno rilevante questa sua definizione in forma chiusa, non ricorrente:

$$\begin{cases} a_n := \frac{\varphi^n - (-\varphi)^{-n}}{\sqrt{5}} \\ n = 0, 1, 2... \end{cases}$$

con la sezione aurea $\varphi := \frac{1+\sqrt{5}}{2} \approx 1.618$

Sia che la si faccia iniziare con 0 come in (9) sia con 1 come in (10) il primo termine può avere indice 0, oppure 1.

WolframAlpha considera la successione iniziante da 0 e al primo termine dà indice 0, e dà:

0 per `Fibonacci[0]` e lo chiama F_0
1 per `Fibonacci[1]` e lo chiama F_1
1 per `Fibonacci[2]` e lo chiama F_2
2 per `Fibonacci[3]` e lo chiama F_3

6.11 Complementi – Fibonacci e Legge di Stigler

Leggiamo (19 settembre 2023) in
https://it.wikipedia.org/wiki/Legge_dell%27eponimia_di_Stigler:

> La legge dell'eponimia di Stigler descrive alcuni aspetti del processo di scoperta scientifica e, in forma estremamente concisa e semplificata, afferma:
> «A una scoperta scientifica non si dà mai il nome del suo autore»
>
> (...) La stessa legge di Stigler è un esempio della propria tesi: la sua scoperta è infatti attribuita da Stigler a Robert Merton.

La successione di Fibonacci fu descritta (iniziando da 1 e 2) dall'Italiano Fibonacci nel 1202 e da lui prende il nome, però era stata descritta mezzo secolo prima dall'indiano d'India Hemachandra e in effetti, nella sostanza, circa un millennio e mezzo prima (con enorme incertezza sui tempi) da un altro indiano d'India, Pingala (che sostanzialmente fu l'inventore anche del sistema di numerazione binario).

BOZZA DRAFT

> Hemachandra also comments (...) 1, 2, 3, 5, 8, 13, 21, 34 and in this way, afterwards[46]
>
> (LINK ->)

La successione era comparsa nell'ambito del calcolo combinatorio applicato alla poesia (!) e precisamente alla metrica, nell'alternarsi delle sillabe lunghe e brevi, in sanscrito: tutt'altro che la demografia dei conigli di Fibonacci. D'altra parte, questa interessante successione ricorre in vari altri ambiti.

[46] Parmanand Singh, The so-called fibonacci numbers in ancient and medieval India, Historia Mathematica, Volume 12, Issue 3, 1985, Pages 229-244, ISSN 0315-0860, https://doi.org/10.1016/0315-0860(85)90021-7. (https://www.sciencedirect.com/science/article/pii/0315086085900217)

6.12 ESERCIZI SULLA LEZIONE 6

6.12.1 Esercizio risolto – Successione di Fibonacci

$_{\mu 2024}$ * Si scrivano i primi termini della successione di Fibonacci, da $a_0 = 1$ ad $a_{11} = 144$:
1 1 2 ... 144
Calcolare

$$\frac{a_{11}}{a_{10}}$$

A quale notissima costante matematica si avvicina molto?

SVOLGIMENTO
Verrà usato lo standard della virgola decimale. (Ma si potrebbe usare lo standard del punto decimale, a scelta).

I primi 12 termini della successione di Fibonacci, da $a_0 = 1$ ad $a_{11} = 144$, sono, e li scriviamo come richiesto nel quesito,
1 1 2 3 5 8 13 21 34 55 89 144
allora $a_{10} = 89$ e allora

$$\frac{a_{11}}{a_{10}} = \frac{144}{89} \approx 1{,}617\,98$$

vicinissima a questa notissima costante matematica:

sezione aurea

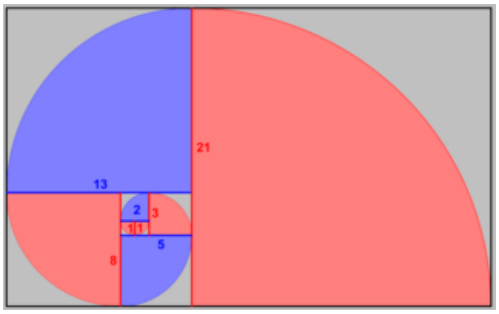
che notoriamente vale circa 1,618 e più precisamente

$$\varphi = \frac{1+\sqrt{5}}{2} \approx 1{,}618\,03$$

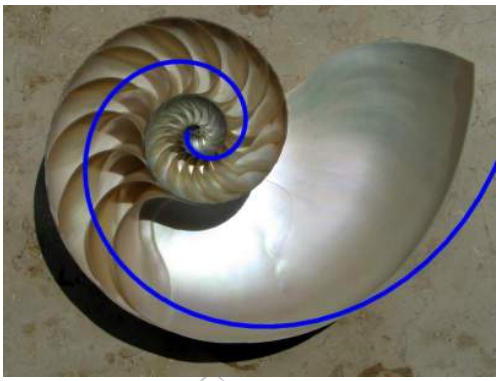
Nota. Il rapporto fra un termine della successione di Fibonacci e il precedente tende, nel senso del limite, esattamente alla sezione aurea φ.
Già $\frac{a_{11}}{a_{10}}$ è molto vicino, come abbiamo sopra visto.

BOZZA - DRAFT

6.12.2 Aneddoti spiraleggianti



Approssimazione di spirale logaritmica, che è connessa alla successione di Fibonacci e alla sezione aurea. By Mabit1 in Wikimedia Commons.



Un vero nautilus con un vera spirale logaritmica. NautilusCutawayLogarithmicSpiral.jpg: Chris 73derivative work: Akkana Peck, CC BY-SA 3.0 via Wikimedia Commons

BOZZA - DRAFT

La successione di Fibonacci compare per esempio nella *spirale logaritmica*,

> indagata estesamente da Jakob Bernoulli, che la definì spira mirabilis, "la spirale meravigliosa", e ne volle una incisa sulla sua lapide

Ma sulla sua pietra tombale fu malamente incisa una banale spirale archimedea, la spirale "infantile" degli scarabocchi.

Lapide di Bernoulli con spirale archimedea, con motto CAMBIATA RISORGO LA STESSA riferito però alla spirale logaritmica, errore *post mortem*.

7 Ripasso di Algebra – I parte

Nel suo progredire nel tempo e approfondirsi, in linea molto generale possiamo dire che ogni Scienza Applicata attraversa successivamente queste fasi:

qualitativa	"bollisci *un po'* di semi per *un po'*"
numeri	"bollisci 12 semi per 1 tazza d'acqua"
operazioni (numeriche)	"50 kg × 3 mg/kg/die = 150 mg/die"
funzioni (numeriche)	"height(cm)=age(yr)×6.5 + 76(cm)" (*)
analisi statistica dei dati	(che vedremo molto più avanti)

(*) La formula sull'altezza si riferisce a bambini cinesi di 2-10 anni in https://pubmed.ncbi.nlm.nih.gov/19573410/).

BOZZA - DRAFT

Vogliamo qua occuparci delle operazioni sui numeri, per esempio a/b, che potrebbe dare la *densità* se a è una massa e b un volume, e che diventa una funzione $f_1(a)$ o $f_2(b)$ se a o rispettivamente b è considerato variabile. (E addirittura una funzione $f_3(a, b)$ di 2 variabili se sono considerati variabili sia a che b, caso di cui ora non ci occuperemo).

7.1 Operazioni unarie (interne) dell'algebra dei numeri

Nei numeri non considereremo operazioni unarie, considerandole funzioni, ma certi software lo fanno.

Le operazioni unarie dell'algebra dei numeri a un numero associano un numero e allora sono funzioni numeriche, e in questa trattazione verranno in generale considerate appunto funzioni e non operazioni unarie, però certi software alcune le consideravano operazioni unarie. Le più tipiche sono il passaggio all'opposto, il passaggio al reciproco, il valore assoluto, la radice quadrata e il fattoriale:

$$-x \qquad \frac{1}{x} \qquad |x| \qquad \sqrt{x} \qquad n!$$

7.2 Operazioni binarie (interne) dell'algebra dei numeri

Le *operazioni binarie* da 2 *operandi* producono 1 *risultato*.
Adesso ne considereremo 5, massimamente classiche:

addizione ovvero somma

sottrazione
moltiplicazione ovvero prodotto
divisione
elevamento a potenza.
Tutte hanno un'infinità di ricorrenze nelle Scienze Applicate. (Un solo esempio: la *densità*, definita con una divisione: massa/volume).

(1) $a+b$, somma, scritto col + e anche col simbolo di sommatoria $\sum$: dati dei numeri a_3, a_4, a_5, la somma

$$a_3 + a_4 + a_5 \quad \text{la scriveremo anche} \quad \sum_{k=3}^{5} a_k$$

e i numeri 3 e 5 possono essere sostituiti da qualsiasi altri, e così pure l'*indice di sommatoria* k e il nome a_k della *variabile indiciata*: una forma comoda da rappresentare somme di molte variabili indiciate: $\sum_{i=n}^{m} x_i$. Per esempio

`ricavoAnnuo=` $\sum_{n=1}^{365}$ `ricavoGiornaliero(n)`

(2) $a - b$, differenza. Che funziona come sommare l'opposto, in Matematica, ma nelle Scienze Applicate è meglio considerare la sottrazione come un'operaziona e sè stante.
Si faccia attenzione che in $a - b - c$ si deve calcolare prima $a - b$.

(3) $a\,b$ ovvero $a \cdot b$; sulle calcolatrici e nei linguaggi informatici anche $*$; lo standard ISO ammette anche la scrittura $a \times b$, che però espone alla confusione con la variabile x. 😐

😐 **Ambiguità.** La notazione $a\,b$ purtroppo dà luogo ad un'ambiguità di scrittura: $y\,(x+1)$ denota sia una funzione y calcolata in $x+1$ che $y \cdot (x+1)$: perciò noi cercheremo di scrivere scrivere, nel caso del prodotto, $(x+1)\,y$ oppure $y \cdot (x+1)$. Inoltre la scrittura $a\,b$ può essere equivocata con una quantità di nome $a\,b$, di 2 lettere, cosa che in questa trattazione succederà raramente, ma nelle Scienze Applicate è normale usare nomi di più lettere. (Si pensi per esempio al pH della Chimica).

BOZZA - DRAFT

(4) $\frac{a}{b}$ ovvero a/b ovvero $a : b$ ovvero $\div$

(5) a^b, sulla calcolatrici $a\text{\^{}}b$ (in certa Informatica anche `a**b`).

😕 **Ambiguità.** In ambito medico e farmaceutico, talvolta per indicare la potenza

non essendo disponibile la scrittura a^b
non essendo disponibile il simbolo ^
non conoscendo il simbolo informatico $**$ per l'elevamento a potenza

si tralascia tutto (!) e si lascia all'intuito del lettore l'intelligenza delle formule.

Per esempio la formula (di Mosteller) che approssima l'area della superficie corporea (usata per il dosaggio dei chemioterapici) da peso W e altezza H

$$\frac{1}{6}(W \cdot H)^{0.5} \qquad \textit{ovvero } \tfrac{1}{6}\sqrt{W \cdot H}$$

si trova trascritta sul sito governativo statunitense PubMed, che riporta l'abstract di un articolo scientifico, in questo modo:

1/6(WH)0.5 LINK->

(e si noti anche l'(1/6) riportato senza parentesi, con seria ambiguità).

Altro esempio, con 2 esponenti, a questo LINK->

BOZZA DRAFT

😕 **Ambiguità.** Un numero tipograficamente ad esponente non sempre indica una potenza.

Per esempio x^0 può denotare sia la variabile x elevata alla 0 sia una variabile di nome proprio x^0.

Così per esempio nel documento sul prezzo equo delle risme di carta (nelle strutture pubbliche, compresi gli ospedali) all'indirizzo https://www.anticorruzione.it/portal/rest/jcr/repository/collaboration/Digital%20Assets/anacdocs/Attivita/Atti/Delibere/2019/2_Allegato_A_aggiornamento%20carta_2019.pdf nella formula $P_{ref}^{2019} = P_{ref}^{2018} * 1.03448$ ovviamente 2018 e 2019 fanno parte dei nomi delle variabili, non indicano potenze.

7.3 Salto dell'uguale di tipo additivo

Icasticamente e come se fossimo a un teatro di burattini:

Qualunque funzione può fare il "salto dell'uguale" cambiando segno
e

da sommata a destra diventa sottratta a sinistra,
da sommata a sinistra diventa sottratta a destra,
da sottratta a destra diventa sommata a sinistra,
da sottratta a sinistra diventa sommata a destra.

Molto chiaramente indicheremo tutto ciò con la scrittura

$$\Big/ + \dots$$

per esempio

$$x^2 + 6 = 5\,x \quad \Big/ + (-5\,x) \text{ o più sveltamente } -5\,x$$

$$x^2 - 5\,x + 6 = 0$$

ZZA - DRAFT

Più tecnicamente: l'equazione, cioè uguaglianza comprendente un'incognita,

$$f(x) = g(x) \qquad \text{per esempio } x^2 + 6 = 5\,x$$

è equivalente (cioè ha le stesse soluzioni) a

$$f(x) + s(x) = g(x) + s(x) \qquad \text{per esempio } x^2 - 5\,x + 6 = 0$$

7.4 Salto dell'uguale di tipo moltiplicativo

Similmente al paragrafo precedente avviene con un moltiplicando, che facendo il "salto dell'uguale" diventa divisore, e viceversa, per esempio nella

$$7\,x = 14$$

il moltiplicando 7 a sinistra diventa divisore a destra

$$x = \frac{14}{7}$$

$$x = 2$$

ovvero con più completa scrittura

$$7\,x = 14 \quad \Big/ : 7$$

eccetera come sopra.
E nella

$$1 = \frac{3}{x}$$

il divisore x diventa moltiplicando (dell'1) trovandosi

$$x = 3$$

(cioè abbiamo fatto $\Big/ \cdot\, x$ richiedendo ovviamente $x \neq 0$).

BOZZA - DRAFT

7.5 Horribilia: frazioni miste

☹ Si faccia attenzione all'ambiguità della "frazione mista" o "numero misto" che si usa per esempio per i voti scolastici, come

$$7\frac{1}{2}$$

Il valore è $7+\frac{1}{2}$ e cioè 7.5, non $7\cdot\frac{1}{2}$ cioè 3.5. Cerchiamo di evitare come la peste tali frazioni miste. Ma potremo trovarle su questionari dove i pazienti valutano $7\frac{1}{2}$ il loro stato di salute (o la soddisfazione del servizio della farmacia), e allora nella raccolta dati trascriveremo sul computer `7.5`. (Non certo `7(1/2)` che manderebbe in errore molti software e da WolframAlpha verrebbe interpretato come `7*(1/2)` invece che `7+(1/2)`).

BOZZA - DRAFT

7.6 Horribilia: scrittura del marketing

☹ Si faccia attenzione alla scrittura *del marketing* 3×2 che non ha nulla a che vedere con la moltiplicazione e rappresenta uno *sconto*, precisamente $m\times n$ significa che di m prodotti non ne pagheremo $m-n$ e allora lo sconto percentuale è

$$sconto = 100\cdot\frac{m-n}{m}\,\%$$

Esempio$_\mu$ Una farmacia vende confezioni di filo interdentale in offerta 4×3: calcoliamo lo sconto.
1 su 4 confezioni non viene pagata e allora lo sconto è $\frac{1}{4}$ cioè 25%.

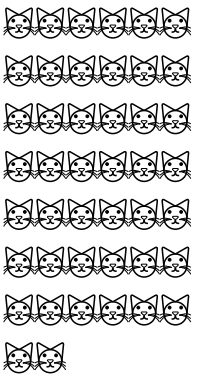
7.7 La divisione è variamente definita

La divisione presenta qualche problematica:
44 diviso 6

- fa $7.\overline{3}\approx 7.333$ in $\mathbb{R}$,
- fa $\frac{44}{6}$ ovvero meglio $\frac{22}{3}$ in $\mathbb{Q}$,
- fa *7 col resto di 2* in $\mathbb{Z}$, come ora approfondiremo.

7.8 Divisione euclidea in $\mathbb{Z}$

Come ci ricorda la canzoncina *Quarantaquattro gatti*,



che si disponevano *in fila per 6*, ma *col resto di 2*, è $44 = 6 \cdot 7 + 2$.[47] Diremo a il *dividendo* e b il *divisore*. Scriviamo per massima chiarezza:

$$dividendo = divisore \times quoziente + resto \quad 0 \leq resto < |divisore|.$$[48]

BOZZA - DRAFT

ESERCIZIO$_\mu$ Una scatola di pillole ha 4 blister di 6×4 pillole. Dopo averne consumate 7 al giorno, ad un certo punto non gliene restano più così tante: quante precisamente?

Le pillole sono in tutto $4 \cdot 6 \cdot 4 = 96$. Con la calcolatrice (o a mano) troviamo

$$96/7 = 13.7...$$

e allora la persona ha consumato 7 pillole al giorno per 13 giorni, in tutto $13 \cdot 7 = 91$. Allora restano alla fine $96 - 91$ cioè 5 pillole. Con WolframAlpha `Remainder[4*6*4,7]`

[47] Più in generale, come leggiamo in Wikipedia, l'enciclopedia libera, alla voce *Divisione euclidea*

> Dati due numero interi a e b con $b \neq 0$ esiste un'unica coppia di interi q ed r detti *quoziente* e *resto* tali che
> $a = b \times q + r$
> $0 \leq r < |b|$

[48] Si noti che il resto è minore del divisore in valore assoluto, ma non necessariamente del quoziente in valore assoluto, per esempio $11 = 4 * 2 + 3$ (ed effettivamente il resto 3 è minore del divisore 4, ma non del quoziente 2), e stiamo dividendo 11 per 4, mentre se dividiamo 11 per 2 abbiamo $11 = 2 * 5 + 1$ (ed effettivamente il resto 1 è minore del divisore 2).
(Si noti ancora che tutto ciò vale anche con numeri negativi, per esempio -21 diviso 9 dà quoziente -3 e resto 6, cioè $-21 = 9 \cdot (-3) + 6$, come troviamo subito online con WolframAlpha con `Quotient[-21,9]` e `Remainder[-21,9]`).

7.9 Proprietà delle 4 operazioni

Con *le 4 operazioni* si intendono + − × /.

Il − non è commutativo: per esempio 6-2 fa 4 ma 2-6 fa -4.
E non è neanche associativo; ovvio ma meglio dirlo:

$11 - 6 - 2$ si calcola nell'ordine $5 - 2$ e infine 3 (il 5 è $11 - 6$)

assolutamente non $11 - 4 = 7$ (computando prima $6 - 2$)

cioè

$$11 - 6 - 2 \quad \text{significa} \quad (11 - 6) - 2.$$

(Le parentesi indicano precedenze nel calcolo).

Il / ha esattamente le stesse problematiche dette per il −, non è commutativo, per esempio 6/3 fa 2 ma 3/6 fa 0.5 ovvero $\frac{1}{2}$, e non è neanche associativo:

$$(x/y)/z \neq x/(y/z).$$

La scrittura $x/y/z$ senza parentesi sarebbe meglio evitarla in Matematica ma è usata in Farmacia.
È da intendersi che quelle 2 operazioni si fanno nell'ordine scritto.
Per esempio l'*indice di massa corporea* si calcola con

$$IMC := BMI := peso(kg)/statura(m)/statura(m)$$

E si consideri anche il classico della Farmacia

$$\text{mg/kg/die}\ ^{(49)}$$

BOZZA - DRAFT

[49] È da utilizzarsi in questo modo:

Y mg/kg/die

si calcola

$Y \times$ (peso [del paziente, o della cavia...] in chilogrammi)

e il risultato è in
mg/die,
cioè milligrammi da assumere ogni giorno. (Da un punto di vista matematico, intendono cioè (mg/kg)/die, implicitamente, come per il − si fanno le operazioni nell'ordine in cui sono scritte, il che è massimamente ragionevole).

Il + e il · sono commutativi e associativi, in tutti gli insiemi numerici (precisazione che in generale ometteremo):

$$x + y = y + x \qquad x\,y = y\,x$$

$$(x + y) + z = x + (y + z) \quad \text{che scriveremo di solito } x + y + z$$

$$(x\,y)\,z = x\,(y\,z) \quad \text{che scriveremo di solito } x\,y\,z$$

(Come detto, le parentesi indicano precedenze nel calcolo).

Valgono 4 proprietà distributive:

$$(x + y)\,z = x\,z + y\,z$$

$$(x - y)\,z = x\,z - y\,z$$

$$\forall z \neq 0 \qquad (x + y)/z = x/z + y/z \quad \textit{ossia} \quad \frac{x+y}{z} = \frac{x}{z} + \frac{y}{z}$$

$$\forall z \neq 0 \qquad (x - y)/z = x/z - y/z \quad \textit{ossia} \quad \frac{x-y}{z} = \frac{x}{z} - \frac{y}{z}$$

ma in generale $\frac{x}{y+z}$ è diverso da $\frac{x}{y} + \frac{x}{z}$, e similmente coi $-$.
Poi

$$-(-x) = x \qquad e \ \forall\, x \neq 0 \ \ \frac{1}{\frac{1}{x}} = x \quad (\textit{involuzioni})$$

e più completamente le 2 "formule delle frazioni a 3 piani"

$$\frac{x}{\frac{z}{w}} = x \cdot \frac{w}{z} \qquad \forall\, z, w \neq 0\,.$$

$$\frac{\frac{x}{y}}{z} = \frac{x}{y} \cdot \frac{1}{z} \qquad \forall\, y, z \neq 0\,.$$

e più completamente ancora la "formula della frazione a 4 piani"

$$\frac{\frac{x}{y}}{\frac{z}{w}} = \frac{x}{y} \cdot \frac{w}{z} \qquad \forall\, y, z, w \neq 0\,.$$

BOZZA - DRAFT

E ancora

$$x + (-x) = x - x = 0 \qquad e \; \forall \; x \neq 0 \;\; x \cdot \frac{1}{x} = \frac{x}{x} = 1$$

$$x + 0 = 0 + x = x \quad (0 \; elemento \; neutro \; rispetto \; al \; +)$$

$$x \cdot 1 = 1 \cdot x = x \quad (1 \; elemento \; neutro \; rispetto \; al \; \cdot)$$

$$x \cdot 0 = 0 \cdot x = 0 \quad (0 \; elemento \; assorbente \; rispetto \; al \; \cdot)$$

$$\frac{0}{x} = 0 \;\; \forall \, x \neq 0 \qquad ma \; \frac{x}{0} \; \textbf{non esiste mai, neppure} \; \frac{0}{0} .😎$$

$\Uparrow$ **Si eviti come la peste "l'orribile fratto zero"** $\Uparrow$

Si dimostra subito

$$\frac{a\,b}{a\,c + a\,d} = \frac{b}{c + d} \quad \text{cioè qua } a \text{ si semplifica}$$

ma (errore comunissimo) 😎

$$\text{da } \frac{a\,b}{a\,c + d} \text{ questa "semplificazione" } \frac{b}{c + d} \quad \text{è } \underline{\text{sbagliata}}.$$

Esempio economico.

Scontare di una certa percentuale il costo totale di 2 articoli, oppure

totalizzare i 2 costi scontati

dà *in teoria* lo stesso risultato, per esempio – per fissare le idee – con uno sconto del 20%

scontare la somma $(x+y)\cdot 0.8 = x\cdot 0.8 + y\cdot 0.8$ sommare gli scontati

Questo è vero se si fanno i calcoli con infiniti decimali. Facendoli con 2 decimali com'è usuale nel commercio in euro o dollari, possono verificarsi discrepanze, in genere piccole. 😎

BOZZA - DRAFT

7.10 Precedenze algebriche e parentesi

Precedenze implicite delle operazioni, da seguire in mancanza di parentesi che le modifichino, dalla precedenza più debole:

+ e − ← precedenza più debole
/ e · anche se non trascritto: in $x + 3\,y$, prima calcolare $3\,y$.
^ anche se non trascritto: in $2\,x^3$, prima calcolare x^3.

Le parentesi possono modificare le precedenze:
$(x + 3) \cdot y$ ci fa prima sommare x e 3. Invece senza parentesi in $x + 3 \cdot y$ prima si moltiplicano 3 e y, poi si fa la somma.

Sempre meglio scrivere parentesi che lasciare incertezze.

Si possono usare quante si vogliano coppie di parentesi "annidate" una dentro l'altra

$$(...(...(...)...)...)$$

ma spesso per chiarezza si usano anche parentesi quadre e graffe:

$$\{...[...(...)...]...\}$$

BOZZA DRAFT

7.11 Diverse opinioni, 1 o 16, su quanto fa 8/2(2+2)

Leggiamo su sito della prestigiosa Università di Harvard

> The question 8/2(2+2) has different answers depending on the rule which is used. One can interpret it as (8/(2(2+2))) = 1 or (8/2)(2+2) = 16 depending on the rule. There is no universally accepted rule as there are several: PEMDAS, BEDMAS, PE(MD)AS. It is not possible to say what is correct and what is incorrect. There are different rules, leading to different results. The expression is not well defined. It appears that most humans naturally give the answer 1 and most computers and programming languages return the answer 16. In order to make the expression unambiguous, one has to put brackets.

Una buona idea: seguire lo standard ISO.
Lo standard ISO esclude le scritture $a/b/c$ e $a/b \cdot c$ senza parentesi.

Senza aspettarsi che l'umanità ci segua, anzi con la certezza che negli articoli scientifici di Farmacia purtroppo scrivono così, non si scrivano espressioni come

$a/b/c$ (si scelga $(a/b)/c$ oppure $a/(b/c)$ a seconda dei casi)
e $a/b\cdot c$ (si scelga $(a/b)\cdot c$ oppure $a/(b\cdot c)$ a seconda dei casi).
(Standard ISO 80000 - 1:2009(E), riportato in LINK->).
Nondimeno nessun matematico sano di mente dubiterà che

$$9\,a^2 : 3\,a$$

faccia $3\,a$. Intendendo (ma chi lo scrive?) $(9\,a^2) : (3\,a)$. Tristezza.
Si veda *I. Relating to the Order of Operations in Algebra.* By N.J Lennes. The American Mathematical Monthly: (1917) Vol. 24, No. 2, pp. 93-95.

La situazione è grave ma non è seria... ...detto col classico aforisma di Flaiano. Attualmente, terza decade del XXI secolo, della situazione con la scrittura delle 4 operazioni si può dire proprio così.

Si trova una calcolatrice

(fotografata per esempio in https://blogdimatematicaescienze.it/il-paradosso-del-pemdas-lavete-risolto/)
che per

$$6 : 2 \times (1+2) \quad \text{e} \quad 6 : 2\,(1+2)$$

dà risultati 9 e 1 rispettivamente.
E ci si preoccupi, se si pensa che la calcolatrice "non sbaglia mai".

BOZZA - DRAFT

Quella calcolatrice dà diverso livello di precedenza alla moltiplicazione se scritta $\times$ oppure indicata implicitamente.

Riprendendo qua a un livello matematico l'imperativo classico *conosci te stesso* diciamo:

CONOSCI LA TUA
CALCOLATRICE

Bottom line
Per ogni coppia di parentesi che mettete, un'ambiguità muore.

7.12 Excel sbaglia ovvero calcola modo suo

Rimarchiamo lo sconcertante fatto che il software Excel, di ampio uso nelle Scienze Applicate, contro le convenzioni

di tutta la comunità matematica, per `-3^2` dà `9` invece che `-9`, come se fosse da intendere $(-3)^2$.
Quindi attenzione se si copia in Excel una formula da un altro software, dove normalmente il risultato sarebbe `-9`, avendo l'elevazione a potenza la precedenza sul meno.
È un caso strano, questo di Excel; addirittura in rete si trova un simulatore di Excel, che dà il risultato opposto di Excel... 😉
E attenzione ancora a questo:

per calcolare -3^2 *che fa -9* in Excel si deve scrivere `-(3^2)`

7.13 Le potenze

Riprendendo per un momento la questione delle ambiguità notazionali, si eviti come la peste la scrittura $a\hat{}b\hat{}c$, specificando con una coppia di parentesi quale potenza si deve calcolare per prima, e se non lo si fa non ci si aspetti unanimità di risultati fra i vari software (e umani).

BOZZA DRAFT

In questa trattazione elementare delle potenze supponiamo nota una conoscenza di base delle radici. Sarebbe possibile fare prima una trattazione elementare delle radici, presupponendo note le potenze. Evitare del tutto queste problematiche è possibile ma al prezzo di complicazioni che qua vogliamo evitare.

1. **Esponente intero positivo.** Scriviamo
   x^2 per $x \cdot x$
   x^3 per $x \cdot x \cdot x$
   x^4 per $x \cdot x \cdot x \cdot x$
   e più in generale definiamo $a \in \mathbb{R}$ *elevato alla* $n \in \mathbb{N}$, $n > 0$, il numero
   $$a^n = a \cdot \ldots \cdot a \quad (n \, volte)$$
2. **Esponente razionale positivo.** Si definisce $a \in \mathbb{R}$ *elevato alla* $q = \frac{n}{m} \in \mathbb{Q}$, $m, n > 0$, il numero
   $$a^{\frac{n}{m}} := \sqrt[m]{a^n}$$

che esiste se $a \geq 0$ *vel* m è dispari. Per esempio $x^{1.5} = x^{\frac{3}{2}} = \sqrt{x^3}$.

3. Esponente reale positivo. Per la potenza con esponente $q \in \mathbb{R}$ si richiede $a \geq 0$, e se q è razionale vale quanto sopra, e se è irrazionale si dà una definizione complessa che si può quasi immaginare sostituendo q con una sua *"straordinariamente buona"* approssimazione razionale: $a^{\sqrt{2}}$ sarà *"circa"* $a^{1.41}$ a sua volta uguale a $\sqrt[100]{a^{141}}$. Ancor meglio $\sqrt[1000]{a^{1414}}$.

4. Esponente negativo. Ponendo per $a \neq 0$

$$a^{-t} := \frac{1}{a^t}$$

la potenza è definita per tutti gli esponenti non nulli; e per l'esponente 0 si pone se $a \neq 0$ (rimanendo non definito 0^0)

$$a^0 := 1 \, .$$

BOZZA - DRAFT

Grafici delle funzioni potenza per base positiva. Ecco su WolframAlpha i grafici di alcune funzioni potenza: Link->

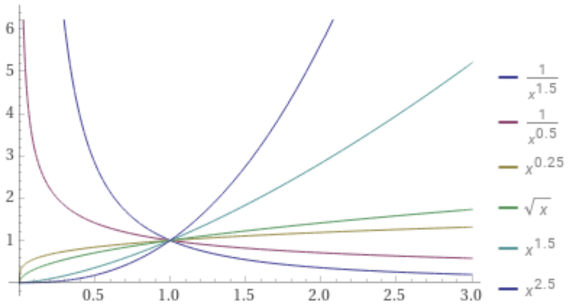


Figure 6: Esponenti −1.5, −0.5, 0.25, 0.5, 1.5. Screenshot da WolframAlpha

Proprietà delle potenze. L'elevamento a potenza non è nè commutativo, per esempio $2^3 = 8 \neq 9 = 3^2$, nè associativo. Valgono invece queste 12 proprietà, di cui la 0-esima già detta:

(0) $x^{-y} := \frac{1}{x^y}$

(1) $x^{y+z} = x^y \cdot x^z$

(1 bis) $x^{y-z} = \frac{x^y}{x^z}$

(2) $x^{y \cdot z} = (x^y)^z$

(2 bis) $x^{\frac{y}{z}} = \sqrt[z]{x^y}$ se $z = 2, 3, 4...$

(3) $(x \cdot y)\hat{}z = (x\hat{}z) \cdot (y\hat{}z)$ $\quad ossia\ (x\,y)^z = x^z y^z$ $\ (distributiva)$

(4) $(x/y)\hat{}z = (x\hat{}z)/(y\hat{}z)$ $\quad ossia\ \left(\frac{x}{y}\right)^z = \frac{x^z}{y^z}$ $\ (distributiva)$

(4 bis) $\left(\frac{x}{y}\right)^{-z} = \left(\frac{y}{x}\right)^z$

(5) $1^x = 1$

(6) $x^1 = x$

(7) $x^0 = 1 \quad \forall\, x \neq 0$

(8) $0^x = 0 \quad \forall\, x \neq 0$

e attenzione poi che

0^0 **non esiste.**

Si noti la notazione, ovvero convenzione sulla notazione:

$$_{usualmente\ scritto \rightarrow}\ x^{y^z} := x^{(y^z)}\ _{\leftarrow raramente\ scritto}$$

per esempio 2^{2^3} è 2^8 cioè 256, non è 4^3 cioè 64, che è $(2^2)^3$.

Esempio sulla proprietà (2).
Volume di un cubo di lato $10^{-10}\,m$. Essendo $volume_{cubo} = lato^3$, si ha $(10^{-10})^3$ cioè $10^{-30}\,m^3$.

Esempio sulle proprietà (2), (4), (5).

BOZZA - DRAFT

Modellizziamo la diluizione omeopatica in acqua con
1 CH : $\frac{1}{100}$ cioè 1 parte di sostanza su 100 totali, e 99 sono acqua
2 CH : $\frac{1}{100}\frac{1}{100}$
e così via, sempreché si riuscissero ad evitare contaminazioni:

$$\text{n CH} : \left(\frac{1}{100}\right)^n$$

(Abbiamo modellizzato solo la questione della diluizione, non l'inevitabile effetto placebo nè l'ipotetica "memoria dell'acqua" affermata dagli omeopati).
A cosa corrisponde 10 CH?

Con le proprietà delle potenze

$$10 \text{ CH} : \left(\frac{1}{100}\right)^{10} =$$
$$= \frac{1^{10}}{100^{10}} = \frac{1}{100^{10}} =$$
$$= \frac{1}{(10^2)^{10}} =$$
$$= \frac{1}{10^{20}}$$

cioè 1 parte su 10^{20} che è

$$100\,000\,000\,000\,000\,000\,000$$

cioè 1 parte su cento miliardi di miliardi. (Ogni 9 zeri c'è un miliardo).

BOZZA DRAFT

7.14 Proporzioni

Si dice *proporzione* la *relazione quaternaria* (cioè fra 4 numeri)

$$\frac{x}{y} = \frac{z}{w} \qquad \text{classicamente scritta} \quad \text{x : y = z : w}$$

ma talvolta si trova anche :: invece del segno =.
Non sottolineeremo qua la corrispondenza del concetto con la realtà sensibile, dovuta essenzialmente alla natura del prodotto e della divisione, supponendola nota dagli studi elementari.

Osserviamo invece che per l'operazione di divisione, oltre alle 4 notazioni già viste,

$$\frac{x}{y} \qquad x/y \qquad x : y \qquad x \div y$$

in ambito farmaceutico se ne usa anche una quinta, la parola (latina) **per**, che noi sempre trascriveremo come frazione.

$$\text{Per esempio:} \quad 12 \text{ mg per ml} \quad \text{scriveremo} \quad \frac{12 \text{ mg}}{1 \text{ ml}}$$

ESERCIZIO Si abbia una boccetta di 10 ml di un farmaco X etichettata "15 mg per ml". Quanti millilitri (ml) bisognerà iniettare per somministrare una dose di 75 mg?

Svolgimento. Riscriviamo l'indicazione in etichetta nella forma

$$\frac{15 \text{ mg}}{1 \text{ ml}}$$

e produciamo la proporzione

$$\frac{15 \text{ mg}}{1 \text{ ml}} = \frac{75 \text{ mg}}{x \text{ ml}} \quad / \cdot \frac{1 \text{ ml} \cdot x \text{ ml}}{15 \text{ mg}}$$

$$x \text{ ml} = \frac{75 \text{ mg} \cdot 1 \text{ ml}}{15 \text{ mg}} =$$

$= 5$ ml (e li abbiamo: la boccetta ne contiene 10).

$$\boxed{5 \text{ ml}}$$

BOZZA - DRAFT

Sul web si ritrovano diffusamente scritte col "per", come "5 mg per 40 ml", e il "per" chiaramente significa / ma a quanto pare in Italia una tale scrittura non si usa, risulta praticamente sconosciuta.

NOTA 1. È ben evidente l'importanza di **calcolare esattamente i dosaggi** in Farmacia. Ci sono stati errori fatali. (Plausibilmente sono numerosissimi e solo in minima parte vengono scoperti).
Ecco cosa consiglia il Royal College of Nursing inglese online:

When you have completed your calculation, remember to check your work. Here's a reminder of the ways you might do this:

- repeat the calculation
- ask a colleague to check your answer
- try to calculate the answer again using a different method
- check against the recommended dose range (e.g. using the British National Formulary)
- look for unusually big or small answers.

NOTA 2. Sebbene esuli dagli obiettivi di questo testo elementare di matematica, si noti che il millilitro (corrispondente al centimetro cubo) può trovarsi indicato **online di fatto** (nei cataloghi di farmaci) sia con ml che mL ma qualcuno scrive anche *diversamente*, in un modo che potrebbe in via ipotetica confondersi con altro multiplo del litro. E si faccia anche ben attenzione a distinguere la l minuscola dal numero 1...

BOZZA - DRAFT

ESERCIZIO$_\mu$
Una pillola contiene una polvere costituita da $2.5\,mg$ di principio attivo e $300\,mg$ di eccipiente. Quanto principio attivo contiene $1\,kg$ della polvere?

SVOLGIMENTO
Una pillola contiene $(300+2.5)mg$ di polvere e allora si ha subito la proporzione

$$2.5\,mg : 302.5\,mg = x\,mg : 1\,kg$$

ovvero con scrittura più moderna

$$\frac{2.5\,mg}{302.5\,mg} = \frac{x\,mg}{1\,kg} \qquad / \cdot \frac{1\,kg}{1\,mg}$$

trovandosi

$$x = \frac{2.5}{302.5} \cdot \frac{1\,mg}{1\,mg} \cdot \frac{1\,kg}{1\,mg} =$$

essendo ovviamente $1\,kg = 1\,000\,g = 1\,000\,000\,mg$

$$= \frac{2.5}{302.5} \cdot 10^6 =$$

$$\boxed{8\,264.46\,mg}$$

(Circa 8 grammi e un quarto).

BOZZA - DRAFT

7.15 ESERCIZI SULLA LEZIONE 7

7.15.1 Esercizio risolto a – Equazione lineare

ESERCIZIO$_\mu$ $\approx$ Risolvere

$$\sqrt{2}\,x - 9 = \frac{7}{5}\,x$$

SVOLGIMENTO

$$\sqrt{2}\,x - 9 = \frac{7}{5}\,x \quad \Big/ + \left(9 - \frac{7}{5}\,x\right)$$

$$\sqrt{2}\,x - \frac{7}{5}\,x = 9$$

$$\left(\sqrt{2} - \frac{7}{5}\right)x = 9 \quad \Big/ : \left(\sqrt{2} - \frac{7}{5}\right) \neq 0$$

$$x = \frac{9}{\sqrt{2} - \frac{7}{5}}$$

che è la soluzione esatta, che adesso approssimiamo numericamente:

$$\approx \frac{9}{1.414214 - 1.4} =$$

$$= \frac{9}{0.014\,214} \approx$$

$$\boxed{633.2}$$

Su WolframAlpha LINK->

BOZZA - DRAFT

7.15.2 Esercizio tanto carino quanto capriccioso

ESERCIZIO$_\mu$ $*$ Calcolare

$$2^{3^4}$$

SVOLGIMENTO Si calcola prima $3^4 = 3 \cdot 3 \cdot 3 \cdot 3 = 81$

$$2^{3^4} =$$

$$= 2^{81} =$$

cioè

$$2 \cdot 2 \cdot ...(\text{c'è in tutto 81 volte il 2})... \cdot 2 =$$

si fa `2^81` con WolframAlpha (oppure si calcola a mano ☺)

$$\boxed{2\,417\,851\,639\,229\,258\,349\,412\,352}$$

8 Ripasso di Algebra – II parte

8.1 Frazioni generatrici

Per scritture decimali limitate la *frazione generatrice* si trova subito: per esempio

$$0.2 \text{ cioè } 2 \text{ decimi, è } \frac{2}{10} \text{ cioè, semplificando, } 0.2 = \frac{1}{5}$$

$$0.25 \text{ cioè } 25 \text{ centesimi, è } \frac{25}{100} \text{ cioè, semplificando, } 0.25 = \frac{1}{4}$$

e similmente coi millesimi si troverà per esempio che 0.125 è $\frac{1}{8}$, e così via.

8.2 Un terzo

Fra le scritture decimali illimitate periodiche, bisognerà conoscere

$$\boxed{\tfrac{1}{3} = 0.\overline{3} = 0.33333333333\ldots \ (\textit{periodico})} \tag{11}$$

BOZZA DRAFT

Un suo valore approssimato, 0.333 oppure 0.3333, si può trovare con la calcolatrice con $1 \div 3$, ma, inversamente, bisogna anche saper riconoscere

$$0.333 \approx \frac{1}{3} \qquad 0.3333 \approx \frac{1}{3}$$

e anzi addirittura, seppure con precisione minore,

$$0.33 \approx \frac{1}{3}$$

Dice per esempio (2022) l'Istat:

> Tra i bambini italiani da 3 a 5 anni, uno su tre è sovrappeso o obeso (33,2% nel 2020 (...))
> https://www.istat.it/storage/rapporti-tematici/sdgs/2022/Rapporto-SDGs-2022.pdf

A un livello giornalistico o divulgativo, addirittura, $0.3 \approx \frac{1}{3}$, e cioè si usa trascrivere 30% come "(circa) uno su 3": LINK->

Nota. Per altre scritture decimali illimitate periodiche, non riusciremo – in questa trattazione elementare – a trovare la frazione generatrice:

$$0.\overline{142\,857} = \frac{?}{?} \qquad \left(\text{in effetti è } \frac{1}{7} \text{ ma non è banale}\right)$$

8.3 Le radici

In questa trattazione elementare, considereremo le radici solo in $\mathbb{R}$ e non nei numeri naturali, interi e razionali.
Sono funzioni (e allora anche operazioni unarie ma non è il modo migliore di considerarle).

• *Radice quadrata* $\sqrt{a}$. Ovvero, ma in generale non scriveremo così, $\sqrt[2]{a}$. (Su WolframAlpha `Sqrt[a]`, in altri software `sqrt(a)`).

$$x \mapsto \sqrt{x}$$

$$\sqrt{\cdot} \ : [0, +\infty[\to \mathbb{R}$$

Per $x \geq 0$ la $\sqrt{x}$ è definita come il numero $y \geq 0$ tale che $y^2 = x$. Cioè **la radice quadrata (di un numero non negativo) è l'unico numero non negativo che elevato al quadrato dà il numero inizialmente considerato.**
Si noti che

$$\sqrt{9} = 3$$

nel modo più assoluto la radice quadrata di 9 non è ± 3. (Come invece affermano testi che seguono un'antiquata definizione di radice quadrata).

BOZZA - DRAFT

• *Radice quarta, sesta, ottava...* $x \mapsto \sqrt[4]{x}$, $x \mapsto \sqrt[6]{x}$... Con considerazioni analoghe alla radice quadrata. In particolare, la radice quarta (di un numero non negativo) è l'unico numero non negativo che elevato alla quarta dà il numero inizialmente considerato. Per esempio la radice quarta di 16 è 2, e quella di 9 è $\sqrt{3}$. Sono definite in $[0, +\infty[$.

• *Radice terza (o cubica), quinta, settima...* $x \mapsto \sqrt[3]{x}$, $x \mapsto \sqrt[5]{x}$...
La $\sqrt[3]{x}$ è definita come il numero y tale che $y^3 = x$, e similmente le altre radici con indice dispari: $\sqrt[3]{\cdot} : \mathbb{R} \to \mathbb{R}$. Per esempio la radice cubica di -8 è -2.

Esempio$_\mu$
Usando la Formula di Keys che, fra i numerosi standard, calcola

$$peso\,ideale\,uomini = (statura\,in\,metri)^2 \times 22.1$$

troviamo quale dovrebbe essere l'altezza corrispondente ai 53 kg:

$$53 =^{EQ} x^2 \times 22.1 \quad \Big/ : 22.1$$

$$x^2 = \frac{53}{22.1} \approx 2.39819 \quad \text{(con la calcolatrice)}$$

$x_1 = -\sqrt{2.39819}$ esclusa, dev'esser positivo (altezza)
$x_2 = \sqrt{2.39819}$
che con la calcolatrice troviamo esser circa 1.55 (in metri).

Nota. La radice quadrata esprime un'infinità di altre cose nelle Scienze Applicate, per esempio il tempo di caduta[50] di un grave. Fra quelle espresse dalla radice cubica citiamo il raggio[51] di una sfera di massa m e densità d.
Per esempio si calcola che una sfera d'oro di 1 kg ha diametro $\approx 4.62\,cm$.
Fra le meno numerose ricorrenze nelle Scienze Applicate della quarta potenza ovvero equivalentemente della radice quarta, citiamo la Legge di Stefan-Boltzmann e la legge di Poiseuille.
Le radici dalla quinta in poi ricorrono moderatamente nelle Scienze Applicate.
Una radice sesta compare in un calcolo relativo al Potenziale di Lennard-Jones della Termodinamica.

BOZZA - DRAFT

8.4 Proprietà dei radicali ovvero radici

Considereremo queste 2 proprietà delle radici

$$x = \sqrt[3]{x^3} \quad e\ similmente\ con\ ogni\ indice\ dispari$$

$$(non\ x) \quad |x| = \sqrt{x^2} \quad e\ similmente\ con\ ogni\ indice\ pari$$

e queste altre 4 proprietà, molto simili fra loro:

$$\forall\, x, y \geq 0 \quad \sqrt{x\,y} = \sqrt{x}\,\sqrt{y} \quad e\ similmente\ con\ ogni\ indice\ pari$$

[50] La formula è $t = \sqrt{\frac{2h}{g}}$.
[51] La formula è $r = \sqrt[3]{\frac{3m}{4\pi d}}$.

$$\forall\, x, y > 0 \quad \sqrt{x/y} = \sqrt{x}\big/\sqrt{y} \quad \textit{e similmente con ogni indice pari}$$

$$\sqrt[3]{x\,y} = \sqrt[3]{x}\ \sqrt[3]{y} \quad \textit{e similmente con ogni indice dispari}$$

$$\forall\, y \neq 0 \quad \sqrt[3]{x/y} = \sqrt[3]{x}\big/\sqrt[3]{y} \quad \textit{e similmente con ogni indice dispari}$$

Esistono anche altre proprietà.[52]

Le radici dei numeri ≥ 0 possono essere intese come potenze con esponente frazionario e precisamente

$$\forall x \geq 0 \quad \left[\sqrt[1]{x} = x\right] \qquad \sqrt{x} = x^{\frac{1}{2}} \qquad \sqrt[3]{x} = x^{\frac{1}{3}} \ldots \quad \textit{e in generale}:$$

$$(\forall\, x \geq 0) \qquad \sqrt[n]{x} = x^{\frac{1}{n}} \qquad n = 2, 3 \ldots$$

In particolare su WolframAlpha calcoliamo la radice n-esima:

$$\textit{numero}\texttt{\^{}(1/n)} \text{ per esempio } \texttt{11\^{}(1/3)} \text{ calcola la } \sqrt[3]{11}.$$

Ma si noti che

$\sqrt[3]{x}$ esiste per ogni x

$x^{\frac{1}{3}}$ esiste solo per ogni $x \geq 0$

(Però si troverà sicuramente qualche Autore che la pensa diversamente e dirà che $x^{\frac{1}{3}} = \sqrt[3]{x}$ per ogni x, che non è una buona idea per sottili motivi).

Da adesso considereremo quasi solo le

[52] Per il lettore interessato:

$$\forall\, x, y \textit{ concordi} \quad \sqrt{x\,y} = \sqrt{|x|}\,\sqrt{|y|} \quad \textit{e similmente con ogni indice pari}$$

(concordi significa aventi ugual segno)

$$\forall\, x, y \textit{ concordi} \quad \sqrt{x/y} = \sqrt{|x|}\big/\sqrt{|y|} \quad \textit{e similmente con ogni indice pari}$$

$$\sqrt[n\cdot m]{x} = \sqrt[n]{\sqrt[m]{x}} \quad \textit{per esempio} \quad \sqrt[6]{x} = \sqrt{\sqrt[3]{x}} = \sqrt[3]{\sqrt{x}}$$

$$\sqrt[nm]{x^m} = \sqrt[n]{x}\,, \quad n, m = 2, 3, 4 \ldots \text{ (si semplifica m)}$$

e di rarissimo uso $\sqrt[n]{x^\alpha} = (\sqrt[n]{x})^\alpha$, $\quad x \geq 0$, $\alpha \in \mathbb{R}$, $n = 2, 3, 4 \ldots$

BOZZA - DRAFT

radici quadrata, cubica e quarta, e per l'ultima vale questa formula di riduzione:

$$\boxed{\sqrt[4]{x} = \sqrt{\sqrt{x}}} \tag{12}$$

(Ovviamente, per $x \geq 0$, altrimenti non è definito nulla).

8.5 La radice quadrata in Medicina e Farmacia

Il significato fondamentale della radice quadrata $\sqrt{x}$ è il lato del quadrato di area x: se per la nostra bella farmacia vogliamo un magazzino quadrato di 400 metri quadrati, deve avere i lati di 20 metri.
Ma vediamo esempi più specifici della Farmacia.

BOZZA - DRAFT

Esempio 1: La Formula di Mosteller per l'area corporea.
Nella chemioterapia per il cancro spesso i dosaggi sono stabiliti non in funzione del peso del paziente – com'è per tanti farmaci – bensì dell'area della superficie corporea.
Una delle formule usate per stimarla è questa di Mosteller, che ha 3 versioni che sono equivalenti per le proprietà delle radici:

$$BSA_{m^2} := area \approx \sqrt{\frac{altezza_{cm} \times peso_{kg}}{3\,600}} =$$

$$= \frac{\sqrt{altezza_{cm} \times peso_{kg}}}{60} = \frac{\sqrt{altezza_{m} \times peso_{kg}}}{6}$$

(che dà circa 2 m^2 per un maschio adulto normale).

Prova tu stesso. Stimare la propria area corporea. 🧍

Esempio 2. La radice quadrata rientra nel calcolo della varianza e dello stimatore della varianza, che vedremo, di amplissimo uso in Statistica Medica.

Esempio 3. Nell'articolo scientifico il cui abstract è in https://www.ncbi.nlm.nih.gov/pubmed/19712746 modellizzano l'accrescimento di una comunità microbica con l'uso della radice quadrata.

Esempio 4. Si voglia moltiplicare per 2 (o rispettivamente per 3, per 4, per qualunque $c > 0$) l'area di un cerotto terapeutico quadrato. Ciò si ottiene moltiplicando il lato per $\sqrt{2}$ (o rispettivamente per $\sqrt{3}$, per 2, per $\sqrt{c}$). Perché da

$$area = lato^2$$

segue, essendo $lato$ un numero positivo,

$$lato = \sqrt{area}$$

e sostituendovi $area$ con $2 \cdot area$ si ottiene

$$lato_{new} = \sqrt{2 \cdot area} = \sqrt{2} \cdot \sqrt{area} = \sqrt{2} \cdot lato$$

e similmente con 3, con 4, con c.

BOZZA - DRAFT

Tutto ciò è vero non solo per il quadrato e il suo lato, ma anche per il cerchio e il suo raggio, e pure per il cerchio e il suo diametro:

il quadrato di area quadrupla ha lato doppio
il cerchio di area quadrupla ha raggio doppio
il cerchio di area quadrupla ha diametro doppio

Con qualche precisazione sul significato delle parole, tali corrispondenze

$$c \times area \leftrightarrow \sqrt{c} \cdot misura\, lineare$$

valgono per qualunque figura geometrica piana, e non solo per il quadrato e il cerchio: una figura – per esempio una macchia d'olio su un liquido – di area quadruplicata rispetto a un'altra, ha misure lineari doppie.

8.6 La radice cubica in Medicina e Farmacia

Il significato fondamentale della radice cubica $\sqrt[3]{x}$ è il lato del cubo di volume x: se per la nostra bella farmacia vogliamo un magazzino cubico di 1000 metri cubi, deve avere i lati di 10 metri.
Ma vediamo esempi più specifici della Farmacia.

Esempio. Supponiamo di voler *ottuplicare* il peso di compresse sferiche (omogenee) che andremo a produrre, rispetto al peso di quelle che già produciamo. Allora il raggio, e quindi il diametro, *raddoppieranno*.

Infatti dalla formula del raggio della sfera di massa m e densità d (che non ci proponiamo di imparare a memoria)

$$r = \sqrt[3]{\frac{3\,m}{4\,\pi\,d}}$$

sostituendo m con $8\,m$ abbiamo

$$r_{new} = \sqrt[3]{\frac{3\cdot 8\,m}{4\,\pi\,d}} = \sqrt[3]{8\cdot\frac{3\,m}{4\,\pi\,d}} =$$

per la quinta ovvero penultima delle proprietà sopra elencate

$$= \sqrt[3]{8}\cdot\sqrt[3]{\frac{3\,m}{4\,\pi\,d}} = 2\,r.$$

BOZZA - DRAFT

Con qualche precisazione sul significato delle parole, un tale raddoppio all'ottuplicare del peso (o equivalentemente del volume) e più in ingenerale la corrispondenza

$$c\times volume \leftrightarrow \sqrt[3]{c}\cdot misura\,lineare$$

(nell'esempio sopra $c = 8$) vale per qualunque figura geometrica, non solo la sfera; per esempio per il cubo relativamente al suo lato, o per una cavia rispetto alla sua lunghezza – ammesso, in via approssimata, che l'accrescersi della cavia nel tempo ne conservi la "forma" (in senso tecnico geometrico si tratta della *similitudine*): la cavia che pesa l'ottuplo, è lunga il doppio. (E il 2 implicato, è la radice cubica dell'8).

Anche un cane o un gatto, nella crescita nel corso del tempo, fatta salva l'approssimazione della forma mantenuta nel tempo:

se ottuplica il volume raddoppia la lunghezza

se moltiplica volume per c, moltiplica la lunghezza per $\sqrt[3]{c}$

Certo se invece di crescere normalmente, piuttosto ingrassa, non è più vero, non conservandosi la forma.
Per due flaconi di medicine di ugual "forma", uno ingrandimento dell'altro, se il primo ha volume ottuplo del secondo, allora ha altezza doppia.

Tutto ciò è vero non solo per i numeri 8 e 2, ma anche per 27 e 3, e per ogni $c > 0$ e $\sqrt[3]{c}$.

BOZZA - DRAFT

8.7 Esempi di calcolo con le radici.

$\sqrt{8} = 2\sqrt{2}$ (perché è $\sqrt{4 \cdot 2} = \sqrt{4} \cdot \sqrt{2}$) e approssimeremo ≈ 2.828
$\sqrt{9} = 3$ (assolutamente non ± 3)
$\sqrt{10} \approx 3.162$ e lo troveremo con la calcolatrice
$\sqrt[3]{-8} = -2$ (perché $(-2)(-2)(-2) = -8$)
$\sqrt[3]{24} = 2\sqrt[3]{3}$ (perché è $\sqrt[3]{8 \cdot 3} = \sqrt[3]{8} \cdot \sqrt[3]{3}$)
$\sqrt[3]{5}$ lo lasceremo come sta. (WolframAlpha con `5^(1/3)` dà 1.7099...).
$\sqrt[4]{4} = \sqrt{2}$ (perché è $\sqrt{\sqrt{4}}$) e approssimeremo ≈ 1.414
$\sqrt[4]{5} \approx 1.495$ (calcolato come radice quadrata della radice quadrata).

8.8 Ordinamento dei numeri

Da $\mathbb{N}$ in poi esiste nei numeri un *ordinamento* (che sostanzialmente supponiamo noto) per cui dati 2 numeri, o il primo è $<$ del secondo, o $>$ del secondo, o sono uguali (*tricotomia*). In modo ovvio si definiscono il *maggiore o uguale* e il *minore o uguale*.

Da $\mathbb{N}$ in poi si può sommare ad ambo i membri di un'uguaglianza,

o di una disuguaglianza, una stessa *quantità* (fissa, cioè un numero, o variabile, come una funzione) conservando l'uguagliaza, o rispettivamente quella stessa disuguaglianza:

$$a < b \Leftrightarrow a + c < b + c \quad e\, analoghe\, con\ \geq, <, \leq, =$$

Da $\mathbb{Z}$ in poi vale questa molteplice relazione fra moltiplicazione e ordinamento dei numeri:

$$a > b \Rightarrow \begin{cases} ac > bc & se\, c > 0 \quad \text{si conserva l'ordinamento} \\ ac < bc & se\, c < 0 \quad \text{si inverte l'ordinamento} \end{cases}$$

e del tutto similmente con $\geq$ e $\leq$

$$a \geq b \Rightarrow \begin{cases} ac \geq bc & se\, c > 0 \quad \text{si conserva l'ordinamento} \\ ac \leq bc & se\, c < 0 \quad \text{si inverte l'ordinamento} \end{cases}$$

Per esempio da $3 > 2$ segue moltiplicando per il numero positivo 10 che $30 > 20$, invece moltiplicando per il numero negativo -10 l'ordinamento si inverte: $-30 < -20$.

BOZZA - DRAFT

8.9 Altre formule classiche dell'algebra

Ecco alcune altre formule classiche dell'algebra:

$$(a+b)^2 = a^2 + 2\,a\,b + b^2$$

$$(a-b)^2 = a^2 - 2\,a\,b + b^2$$

$$a^2 - b^2 = (a+b)(a-b)$$

$$a \cdot b = 0 \Leftrightarrow a = 0 \ \vee\ b = 0 \quad \text{Legge di annullamento del prodotto}$$

E ce ne sono moltissime altre, di cui alcune nella sezione di Complementi a questa Lezione.

8.10 Radici quadrate e cubiche, ed elettrocardiogrammi

In un certo calcolo relativo agli elettrocardiogrammi dal 1920 sono usate alternativamente queste 2 formule

$$\frac{a}{\sqrt{b}} \qquad \text{Formula di Bazett}$$

$$\frac{a}{\sqrt[3]{b}} \qquad \text{Formula di Fridericia}$$

e in seguito ne sono state proposte altre ancora, rimanendo nell'incertezza su quale sia la migliore. Leggiamo infatti in un articolo[53] scientifico (2018) su 5 formule, per neonati:

> There is an almost endless controversy regarding the choice of the (...) formula to be used in electrocardiograms (...) in neonates for screening for long QT syndrome (...). We compared the performance of four commonly used formulae and a new formula derived from neonates.
> (...) Bazett's (...), Fridericia's
> (...) The Bazett's (...) can be used with confidence in neonates

BOZZA_DRAFT

Cioè per i neonati consiglia la formula con la radice quadrata. Ma un'altro[54] studio, fatto su maschi adulti, dice che funziona meglio la radice cubica:

> 881 middle-aged men (...) the cubic root equation might be more accurate than the square root

[53] Stramba-Badiale M, Karnad DR, Goulene KM, Panicker GK, Dagradi F, Spazzolini C, Kothari S, Lokhandwala YY, Schwartz PJ. For neonatal ECG screening there is no reason to relinquish old Bazett's correction. Eur Heart J. 2018 Aug 14;39(31):2888-2895. doi: 10.1093/eurheartj/ehy284. PMID: 29860404.

[54] Puddu PE, Jouve R, Mariotti S, Giampaoli S, Lanti M, Reale A, Menotti A. Evaluation of 10 QT prediction formulas in 881 middle-aged men from the seven countries study: emphasis on the cubic root Fridericia's equation. J Electrocardiol. 1988 Aug;21(3):219-29. doi: 10.1016/0022-0736(88)90096-9. PMID: 3171455.

8.11 La Medicina non è una Scienza Esatta

Leggendo il paragrafo precedente, viste anche le implicazioni farmacologiche, sia un matematico che uno studente di Farmacia o CTF all'inizio dei suoi studi potrebbero restare alquanto sorpresi: radice quadrata e cubica possono dare valori piuttosto diversi, e dai due articoli non si evince un'età soglia in cui cambiare formula, e comunque entrambi gli articoli parlano di esistente incertezza, che cercano di dirimere almeno nelle due classi di età considerate.
Ma la Medicina non è una Scienza Esatta, e simili differenze di opinione fra i vari Medici – in questo caso su quale formula è migliore – non devono sorprendere.
Similmente, durante la pandemia (2020-2023) varie virostar si sono vicendevolmente contraddette – e perfino insultate – proponendo tesi diverse: non è che non sappiano usare la calcolatrice o il microscopio, è che usano modelli diversi, opinabili.

BOZZA - DRAFT

8.12 Il volùbil aspartame

Da (LINK ->, letto il 14 ottobre 2023) Rainews del 14 luglio 2023:

> L'aspartame è "probabilmente" cancerogeno, lo ha stabilito l'OMS

Ancora il 16 ottobre 2023 leggiamo sul sito web dell'EFSA in https://www.efsa.europa.eu/it/topics/topic/aspartame

> L'aspartame è un edulcorante artificiale
> (...) L'edulcorante e i suoi prodotti di degradazione sono autorizzati per il consumo umano da molti anni sulla base di approfondite valutazioni della sicurezza.

Prima però non era nella lista delle sostanze classificate come "probabilmente cancerogeno per l'uomo": è avvenuto un cambiamento di opinione scientifica nel corso del tempo. Sulla base – nelle sue varie forme – della Statistica Medica.

In questo caso però il cambiamento sul piano pratico non è drammatico: continuiamo da RaiNews:

> Nonostante la probabile pericolosità, l'OMS non propone di ridurne la quantità giornaliera consigliata. Esulta l'industria, delusi gli attivisti

Il passaggio da sospetti ad articoli scientifici a pareri non vincolanti a leggi che proibiscano l'uso di determinate sostanze è, diciamo così, molto lento: per l'amianto in Italia molti decenni. Per il biossido di titanio, colorante tossico dei farmaci, nel 2024 siamo ancora in attesa.

BOZZA - DRAFT

* Complementi *

8.13 Complementi – Ulteriori formule classiche

$$(a+b)^3 = a^3 + 3\,a^2\,b + 3\,a\,b^2 + b^3$$

$$(a-b)^3 = a^3 - 3\,a^2\,b + 3\,a\,b^2 - b^3$$

$$a^3 + b^3 = (a+b)(a^2 - a\,b + b^2)$$

$$a^3 - b^3 = (a-b)(a^2 + a\,b + b^2)$$

$$a^4 + b^4 = \left(a^2 - \sqrt{2}\,a\,b + b^2\right)\left(a^2 + \sqrt{2}\,a\,b + b^2\right)$$

(E l'ultima, poco nota, sorprenderà più di qualcuno).

BOZZA DRAFT

8.14 ESERCIZI SULLA LEZIONE 8

8.14.1 Esercizio risolto a – Area corporea

ESERCIZIO$_{\mu 2020}$ % Il dosaggio di particolari farmaci dipende dalla superficie corporea. Una delle formule approssimate (Formula di Mosteller) dà

$$superficie_{m^2} = \sqrt{\frac{altezza_{cm} \times peso_{kg}}{3\,600}}$$

Prendiamo per buona questa formula (senza discuterne eventuali limiti alla validità). Se un fissato individuo conservando la sua altezza raddoppia il peso, di quanto aumenterebbe la sua superficie corporea?

SVOLGIMENTO

$$\frac{superficie\ finale}{superficie\ iniziale} =$$

detta h_0 l'altezza, e p_0 il peso iniziale e allora $2\,p_0$ quello finale

$$= \frac{\sqrt{\frac{h_0 \cdot 2\,p_0}{3\,600}}}{\sqrt{\frac{h_0 \cdot p_0}{3\,600}}}$$

si semplifica tutto tranne $\sqrt{2}$

$$= \sqrt{2} \approx 1{,}41$$

$$\boxed{\approx 41\%}$$

(Nella pratica, appare ragionevole non fare alcun affidamento sull'esattezza della cifra delle unità; diciamo piuttosto che, col modello della Formula di Mosteller, ad un raddoppio del peso conservando l'altezza corrisponde un aumento circa del 40% della superficie corporea).

BOZZA - DRAFT

8.14.2 Esercizio risolto b – Mortalità fra ricoverati

ESERCIZIO$_{\mu 2020}$ $\star$ Supponiamo che di 24391 ricoverati per una nuova malattia muoiano 8123. Si esprima a parole (divulgativamente) la situazione con un'espressione approssimata come

"morti 2 su 3" oppure "morto 1 su 4".

SVOLGIMENTO

Si ha

$$\frac{morti}{ricoverati} = \frac{8123}{24391} \approx$$

con la calcolatrice

$$\approx 0.33303$$

che riconosciamo essere circa $\frac{1}{3} = 0.33333... = 0.\overline{3}$ e allora, a parole e approssimativamente:

morto 1 su 3

BOZZA - DRAFT

9 Piano cartesiano – I parte

9.1 Traiamo da Food and Drug Administration, USA

Fotografia di

Guidance for Industry
Q8(R2) Pharmaceutical Development

a pagina 22:

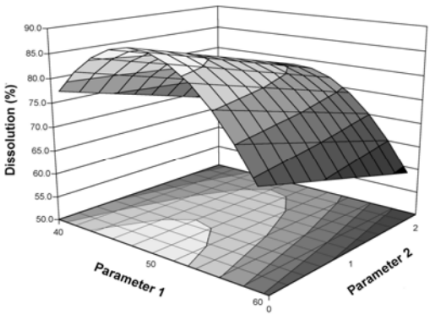


Figure 1a: Response surface plot of dissolution as a function of two parameters of a granulation operation. Dissolution above 80% is desired.

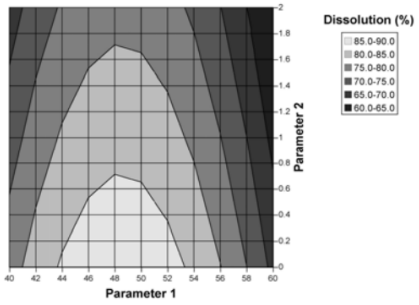


Figure 1b: Contour plot of dissolution from example 1a.

9.2 Introduzione, anche con riferimento alla Farmacia

Già sappiamo che i numeri reali sono identificabili coi punti di una retta orientata su cui sia fissato un punto origine e un punto unità, ciò che costituisce un *sistema di ascisse* in quello che possiamo anche chiamare spazio cartesiano a 1 dimensione.

In questa lezione estenderemo quello spazio unidimensionale a uno spazio bidimensionale, il piano cartesiano, in cui
i punti sono identificabili con le coppie ordinate di numeri reali.

Dei milioni di articoli scientifici nel campo biomedico, innumerevoli contengono rappresentazioni nel piano cartesiano. Si inizi apprezzando il ciclo pressione-volume del ventricolo murino sinistro.

9.3 Definiamo l'obesità di animaletti

Supponiamo che uno sperimentatore abbia 16 animaletti della stessa specie, e di ciascuno
1) misura la lunghezza
2) misura il peso
3) valuta se secondo lui è normopeso, sovrappeso o sottopeso.
Ottiene questi dati, senza qua riportare le unità di misura:

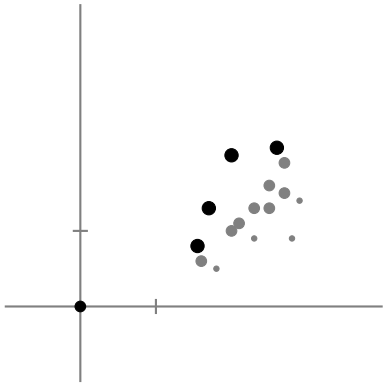
animaletto 1: lunghezza 2, peso 1, normopeso
animaletto 2: lunghezza 2.8, peso 0.9, sottopeso
animaletto 3: lunghezza 2.7, peso 1.9, normopeso
...
animaletto 16: lunghezza 2, peso 2, sovrappeso.

Ricerca interessante, anche costosa (il tempo è denaro), ma risultato oscuro e inconcludente.

L'università di Trieste (2023) alleva:
cavie, maiali, pulcini, pesci zebra;
altri animali da esperimento
vengono comprati da aziende esterne.

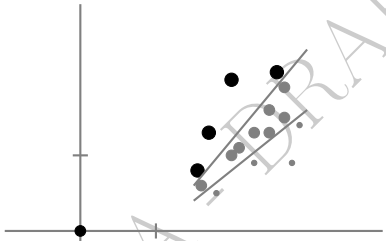
BOZZA - DRAFT

Rappresentiamo le coppie ordinate (lunghezza, peso) di ciascun animaletto come punti del piano cartesiano – adesso considerato in senso intuitivo, e dopo lo definiremo meglio – con colori diversi a seconda della qualifica normopeso, sovrappeso e sottopeso.



Adesso la situazione è molto più chiara: ci sono sostanzialmente 3 regioni del piano in cui cadono i punti corrispondenti rispettivamente ai normopeso, sovrappeso e sottopeso.

A un livello superiore, il ricercatore stabilisce analiticamente dei possibili confini delle 3 regioni, dando di fatto le condizioni che poi – senza fare alcun disegno bensì un semplice calcolo con la calcolatrice – permettono in qualunque Ambulatorio Veterinario – o Laboratorio 😱 – di stabilire, dalla lunghezza e dal peso, se un dato animaletto va considerato normopeso o sovrappeso o sottopeso. Una cosa del genere si fa di fatto per gli esseri umani, in vari modi.



BOZZA - DRAFT

Standard proposto dal ricercatore:

- sottopeso se: $peso < (altezza - 1) \cdot 0.8$
- sovrappeso se: $peso > (altezza - 1) \cdot 1.2$
- altrimenti: normopeso.

Almeno, per $1.5 <$ **altezza** < 3**.**

Le formule delle Scienze Applicate hanno in generale un dominio di validità, spesso non esplicitato, con vaghi confini al di fuori dei quali perdono validità.

Non si possono applicare le formule del sovrappeso a un embrione.

Si noti che nell'esempio soprastante siamo entrati nella Scienza Moderna uscendo dalla medicina preistorica:

Mi sembra che uno con la forma di mio cugino (fa il gesto con le mani) si deve considerare troppo magro.

Adesso abbiamo formule matematiche. Con esse in certi casi si decidono terapie. (protocolli, *evidence based medicine*, ecc.).

9.4 Premessa definizionale

funzione: $f(x)$, p.es. $f(x) := x^2 - 2$, spesso scritta $y = x^2 - 2$;

equazione: $f(x) = g(x)$, p.es. $x^2 - 2 = 0$ ha ***soluzione*** $\pm\sqrt{2}$;

polinomio: $a_n x^n + ... + a_1 x + a_0$, p.es. $x^2 - 2$, ha ***radici*** $\pm\sqrt{2}$;

disequazione in 1 variabile: $f(x) > g(x)$ (o $<$ o $\geq$ o $\leq$), p.es. $x^2 > 2$, ha ***soluzione*** $x < -\sqrt{2} \vee x > \sqrt{2}$.

9.5 Assi cartesiani

Gli assi cartesiani sono una retta orientata detta asse delle ascisse e una retta orientata perpendicolare alla precedente, detta asse delle ordinate, che si intersecano in un punto detto origine e denotato con O. Su ciascuna è fissato un punto unità.
L'orientazione del piano e degli angoli con vertice nell'origine è antioraria. L'asse delle ascisse si sovrappone all'asse delle ordinate con una rotazione di $+90°$.

BOZZA - DRAFT

Spesso l'asse delle ascisse ha nome x, ma t se rappresenta un tempo, e quello delle ordinate y, ma p se rappresenta un peso, eccetera, qualsiasi nome.

Possiamo dire che su ciascun asse il punto unità fissa un'unità di misura, che determina l'ascissa e l'ordinata (distanze con segno dall'origine) dei loro punti. Tuttavia, si ricordi che in Matematica i numeri delle ascisse e ordinate sono numeri puri, ossia numeri, mentre nelle Scienze Applicate possono essere dotati di unità di misura, ed essere grandezze fisiche.

9.6 Punti del piano cartesiano

Equazione del punto[55]:

$$P = (x_0, y_0) \quad ovvero \quad P(x_0, y_0) \quad ovvero \quad (x_0, y_0)$$

Il numero reale x_0 è l'ascissa di P e y_0 l'ordinata.

Da adesso, le relazioni geometriche fra figure diventano relazioni algebriche fra numeri, con enorme vantaggio pratico e applicativo. In particolare, per i computer è normale operare su numeri.

Già il D'Oresme[56] (XIV secolo), iniziatore del metodo, arrivò fino a produrre – sostanzialmente – l'equazione della retta. Oggi noi seguiamo la teoria, più completa, di Descartes (Cartesius, Cartesio, XVII secolo).

BOZZA - DRAFT

9.7 Grafico di dispersione ovvero scatterplot

(Scritto sia scatterplot che scatter plot).

> "Il grafico di dispersione o grafico a dispersione o scatter plot o scatter graph è un tipo di grafico in cui due variabili di un set di dati sono riportate su uno spazio cartesiano.
>
> I dati sono visualizzati tramite una collezione di punti ciascuno con una posizione sull'asse orizzontale determinato da una variabile e sull'asse verticale determinato dall'altra.
> (...)
> solitamente la variabile più importante è sull'asse delle y"
> (Wikipedia, l'enciclopedia libera)

[55] Altri Autori scrivono col punto e virgola: $(x_0; y_0)$

[56] Leggiamo su Wikipedia, l'enciclopedia libera, alla voce Nicola d'Oresme: "matematico, fisico, astronomo ed economista, vescovo, filosofo, psicologo e musicologo francese (...) teologo appassionato, traduttore competente, influente consigliere di re Carlo V di Francia (...) ebbe l'idea di utilizzare ciò che dovremmo chiamare coordinate rettangolari nella terminologia moderna, una lunghezza proporzionale alla longitudo, l'ascissa di un dato punto e una perpendicolare a quel punto, proporzionale alla latitudo, l'ordinata (...) longitudo e latitudo possono variare o rimanere costanti."

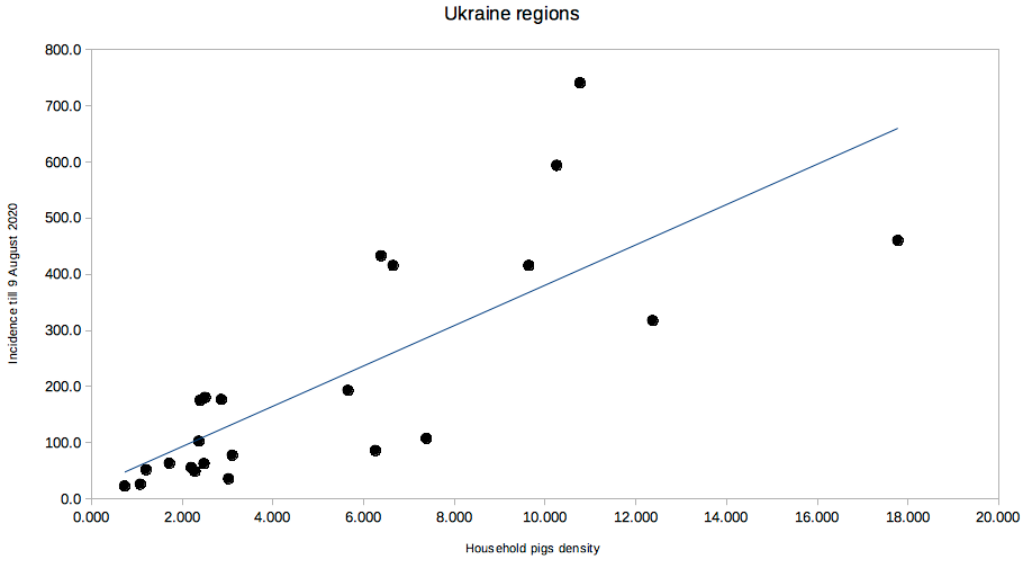


Figure 7: Esempio di scatterplot, che mette in relazione la densità di maiali d'allevamento domestico ed incidenza del covid-19, nelle regioni dell'Ucraina; tratto da: "In the first wave of the 2020 pandemic in several areas more sunlight less pandemic, more pigs more pandemic, and lower correlations with some other livestocks". (December 2020), A. Soranzo. DOI: 10.13140/RG.2.2.29852.31367 Allo scatterplot è aggiunta la *retta di regressione*.

BOZZA DRAFT

Altri esempi in questo LINK->

Lo Studente, magari anche in dipendenza dei suoi interessi, troverà diagrammi interessantissimi cercando online per immagini con qualche motore di ricerca le parole

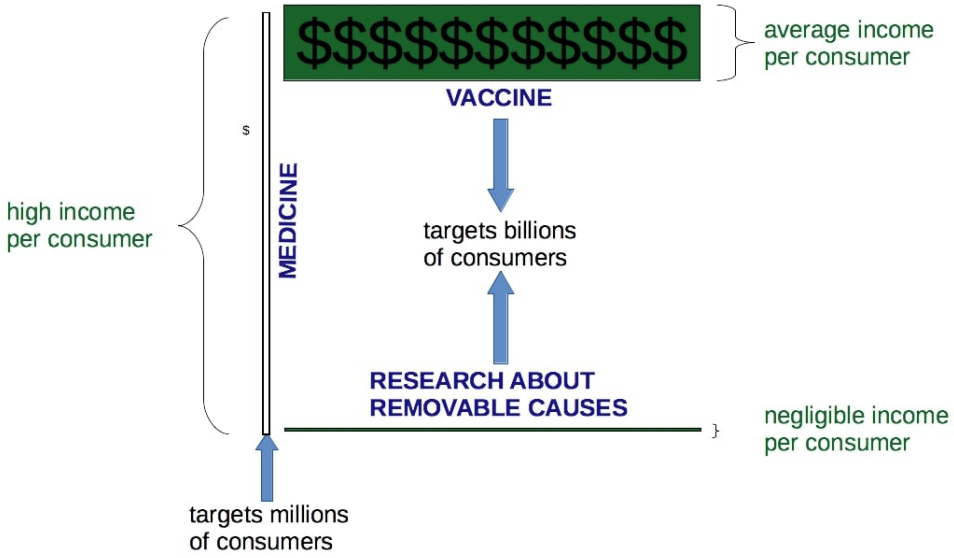
scatterplot qualche_nome_di_malattia_meglio_se_in_inglese

per esempio

scatterplot diabetes

9.8 Diagrammi e grafici

Ogni rappresentazione grafica di dati la diremo *diagramma*.



Esempio di diagramma. Con simili strumenti[57] un'azienda farmaceutica puó decidere in quali ricerche scientifiche investire per ottimizzare i guadagni. Poi, in base alla dottrina economica corrente, *la mano invisibile del mercato* dovrebbe produrre benefici per molti.

BOZZA

Detto ultrasemplificatamente, con numeri riferibili all'Italia per agevole compransione: il farmaco lo date magari a 50mila malati gravi, il vaccino a 50 milioni di sani: sono mille volte di più clienti. I più famosi farmaci (Remdesivir, Paxlovid, Molnupiravir) creati specificamente per il covid (inizialmente propagandati come semi-miracolosi e poi risultati molto meno efficaci) avevano costi dell'ordine dei 500-3000 euro, 1 dose di vaccino sui 4-20 euro: è vero che i farmaci costavano 100 volte più del vaccino, ma i destinatari del vaccino erano 1000 volte di più. Proporzionalmente c'è stato l'interesse delle case farmaceutiche per vaccini e farmaci. Sulla ricerca delle cause rimovibili c'è stato pochissima ricerca scientifica. Il diagramma fa capire chiaramente queste questioni.

[57] Detto icasticamente: il farmaco si darebbe magari a 50mila *casi severi* e il vaccino a 50 milioni di sani, un rapporto 1:1000. E' vero che il farmaco costa più del vaccino, ma non 1000 volte, diciamo circa 100 volte. (Per esempio il famoso Paxlovid, 1400 dollari per un trattamento di cinque giorni).

Fra i diagrammi ci sono i ben noti diagrammi a torte per le percentuali, i grafi usati per le mappe cognitive, gli istogrammi che vedremo, le carte statistiche, perfino interattive.

E **i *grafici* in senso matematico** così definiti:

grafico è un sottoinsieme G di $X \times Y$ tale che

$$(\forall x)\,(\exists!\, y)\,(x,y) \in G$$

(È sottinteso che $x \in X$ e $y \in Y$, e fra la quarta e la quinta parentesi è implicito un "tale che").

Ma nelle Scienze Applicate spesso qualunque diagramma viene chiamato grafico.

Si noti che una retta è un grafico[58] in $\mathbb{R} \times \mathbb{R}$ ma un circolo no, almeno non nel senso matematico sopra detto.

POZZA DRAFT

😇 I matematici fanno la distinzione fra un *grafico*, oggetto astratto che spesso è un insieme illimitato del piano, e un *disegno di un grafico*, necessariamente limitato e quindi in generale incompleto ovvero parziale: nessuno può disegnare tutta una parabola, eppure sia nel linguaggio comune che nelle Scienze Applicate il disegno di una sua parte (quasi) tutti lo chiamano (grafico della) parabola.

Figure	su WolframAlpha `Link->`	su WolframAlpha `Link->`

Nella Scienze Applicate e nella divulgazione scientifica avviene spesso che anche i (disegni dei) grafici in senso matematico vengano arricchiti da notazioni e disegni, diventando diagrammi di vario tipo.

[58] Ovviamente una retta verticale è un grafico di una funzione della y, precisamente una funzione costante del tipo $x = cost$.

9.9 Rette del piano cartesiano

Un'infinità di fenomeni delle Scienze Applicate viene modellizzata, più o meno esattamente ovvero più o meno approssimativamente, con equazioni

$$y = numero \cdot variabile + evantuale\, ltro\, numero$$

i cui grafici nel piano cartesiano sono rette

Esistono
- rette verticali, cioè parallele all'asse y: equazione $x = p$
- rette orizzontali, cioè parallele all'asse x: equazione $y = q$
- rette oblique, non parallele nè all'asse x nè all'asse y

Equazione esplicita della retta passante per l'origine e non verticale, ovvero

retta obliqua od orizzontale, per l'origine:

$$y = m\,x$$

BOZZA - DRAFT

e questa funzione è caratterizzata da
"al raddoppiare di x raddoppia y e viceversa."

La maggior parte delle funzioni di 1 variabile delle Scienze Applicate ha questa forma.
Per esempio la Prima Legge di Gay-Lussac ovvero Legge di Charles con la temperatura assoluta

$$V = V_0\,\alpha\,T \qquad T\ temperatura\ assoluta \qquad retta\ per\ l'origine\ nel\ piano\ (T,V)$$

Equazioni esplicita della retta non verticale, ovvero

retta obliqua od orizzontale:

$$y = m\,x + q$$

per esempio $giorni_dal_concepimento = 365.25\,anni + 273$ (Formula approssimativa).

q ci dice il punto di intersezione con l'asse y, precisamente $(0, q)$.
m: ***coefficiente angolare***, ci indica la pendenza della retta;
se è 0 la retta è orizzontale
se è > 0 la retta "sale"
se è < 0 la retta "scende".

A volte nelle Scienze Applicate l'equazione appare come

$$y = m\,(x + p)$$

che è esattamente come prima con $q = m\,p$ ovvero $p = q/m$.

Per esempio una ricerca scientifica[59] dà questo peso ideale

$$\text{weight_children_aged_1-5_years} = 2 \times (\text{age_in_years} + 5)$$

che nella forma $y = m\,x + q$ sarebbe $\text{weight_children_aged_1-5_years} = 2 \times \text{age_in_years} + 10$.
Addirittura può presentarsi scritta in una forma equivalente

$$y = a\,(b\,x + c)$$

per esempio la Prima Legge di Gay-Lussac ovvero Legge di Charles con i gradi Celsius

$$V = V_0\,(1 + \alpha\,t) \qquad t\ \textit{temperatura}\ ^\circ C \qquad \textit{retta obliqua nel piano}\ (t, V)$$

BOZZA - DRAFT

Equazione implicita della retta: $a\,x + b\,y + c = 0$ con $(a, b) \neq (0, 0)$

per esempio $2\,x - y + 10 = 0$ che è quella appena vista salvo diverse variabili.

Ogni punto $P(x, y)$ le cui coordinate verificano l'equazione della retta appartengono alla retta, e questo sarà un fatto generale, estendibile a tutte le *curve*, in rappresentazione esplicita o implicita.

Per esempio $(1.7, 63.869)$ sta sulla curva del peso ideale maschile secondo Keys, che è la parabola (grafico di) $y = 22.1\,x^2$: altezza in metri 1.70 (ovvero 1.7, matematicamente), peso 63.869 in chilogrammi.

[59] Make your Best Guess: an updated method for paediatric weight estimation in emergencies, by Tinning K, Acworth J., in Emerg Med Australas. 2007 Dec;19(6):528-34.

Equazione della retta per 2 punti:

$$\boxed{\frac{x - x_2}{x_1 - x_2} = \frac{y - y_2}{y_1 - y_2}} \tag{13}$$

Naturalmente con semplici passaggi algebrici la forma soprastante può essere ricondotta ad $a\,x + b\,y + c = 0$, e, se la retta non è verticale, a $y = m\,x + q$.

Formula della distanza di 2 punti:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Il computer non mette un righello da un punto all'altro: usa questa e analoghe formule.
E se un modernissimo computer "vede" i punti con una telecamera, poi la distanza la calcolerà comunque con la formula soprastante, dopo aver trasformato i punti in coordinate numeriche.

BOZZA - DRAFT

2 rette di equazioni $y = m\,x + q$ e $y = m'\,x + q'$

$$\text{sono parallele} \;\Leftrightarrow\; m' = m$$

(e ovviamente 2 rette verticali $x = p$ e $x = p'$ sono parallele).

E si possono considerare molte altre formule.[60]

[60] Per il lettore interessato, ecco alcune altre formule.
2 rette di equazioni $y = m\,x + q$ e $y = m'\,x + q'$

$$\text{sono perpendicolari} \;\Leftrightarrow\; m' = -\frac{1}{m}$$

Formula della distanza punto-retta:

$$d = \frac{|a\,x_0 + b\,y_0 + c|}{\sqrt{a^2 + b^2}}$$

Grafico cartesiano G_f di una funzione $f(x)$ è (il disegno del)l'insieme

$$G_f := \{(x, y) \in \mathbb{R}^2 \mid y = f(x)\}$$

e spesso le funzioni hanno nome $y(x)$ denotato anche semplicemente y, e già abbiamo visto le rette orizzontali e oblique $y = m\,x+q$, che di fatto sono proprio grafici, mentre le rette verticali non sono grafici di funzioni di x. (Ma sono grafici di funzioni di y).

Fra le rette che sono grafici di funzioni di x si distinguono queste:
- $y = x$ bisettrice del I e III quadrante – funzione *identità*
- $y = -x$ bisettrice del II e IV quadrante – passaggio all'opposto
- $y = q$ con $q \in \mathbb{R}$ generica retta orizzontale, e se $q = 0$ si ha
- $y = 0$ asse x

e quelle che non sono grafici di funzioni di x sono precisamente queste:
- $x = x_0$ generica retta verticale, e se $x_0 = 0$ si ha
- $x = 0$ asse y.

BOZZA - DRAFT

9.10 Funzioni e dis/equazioni di primo grado

Per ogni m la funzione $f(x) := m\,x$ si chiama *funzione lineare*, e in questo contesto la scriveremo $y = m\,x$.

> Esempio: l'indice di massa corporea: $BMI := \frac{peso}{altezza^2}$, in cui l'altezza possiamo ragionevolmente considerarla costante – per un fissato individuo – mentre un peso x può variare più facilmente, in pratica abbiamo $y = \frac{1}{altezza^2} \cdot x$; si usino chilogrammi e metri).

È una funzione crescente se $m > 0$, e *decrescente* se $m < 0$, *costantemente nulla* se $m = 0$.
Il grafico è una retta passante per l'origine.
Fissato $m \neq 0$, l'equazione $m\,x = 0$ ha soluzione $x = 0$ (basta dividere per $m \neq 0$), mentre la disequazione in 1 variabile

$$m\,x > 0$$

si risolve dividendo per m ciò che, se e solo se $m < 0$, inverte l'ordinamento. Analogamente si risolve se si aveva $\geq$, $<$ o $\leq$.

Per ogni $m, q \in \mathbb{R}$ la funzione $f(x) := mx + q$ si chiama funzione *affine*, deprecabilmente detta lineare, e in questo contesto la scriveremo $y = mx + q$.
È una funzione crescente se $m > 0$, e *decrescente* se $m < 0$, *costante* se $m = 0$.
Il grafico è una retta, che *intercetta* (interseca) l'asse y in $(0, q)$.
Fissati $m \neq 0$ e q, l'equazione

$$mx + q = 0$$

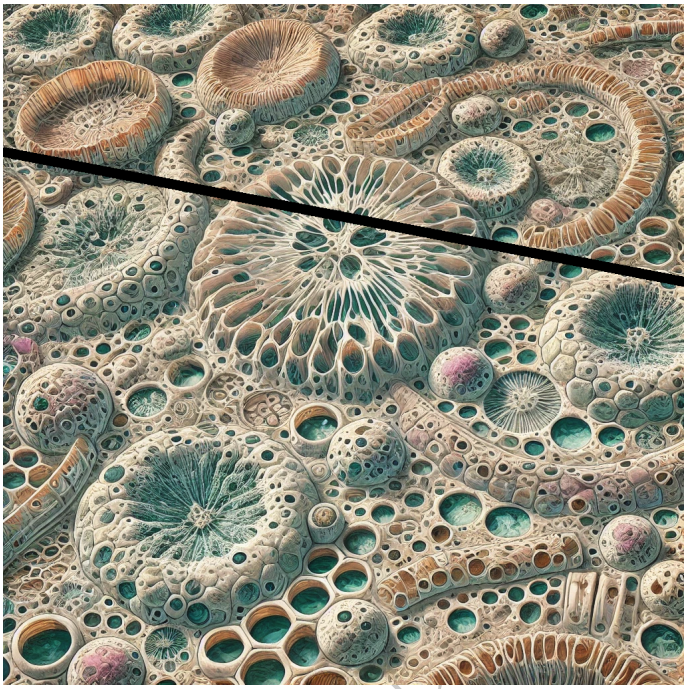
ha soluzione $x = -\frac{q}{m}$. (Si sommi $-q$ e si divida per $m \neq 0$).
Le 4 disequazioni con $>$, $\geq$, $<$, $\leq$ si risolvono sommando $-q$ e poi dividendo per m invertendo l'ordinamento se $m < 0$.
Fissati m e q, la disequazione in 2 variabili

$$y \geq mx + q$$

BOZZA DRAFT

rappresenta il *semipiano chiuso* "sopra" la retta $y = mx + q$, compresi i punti della retta. Col $>$, il *semipiano aperto*, esclusi i punti della retta.
Con $\leq$, e con $<$, si va "sotto" la retta, compresa (semipiano chiuso) o rispettivamente esclusa (semipiano aperto).
Ciò vale anche per la $y > mx$ e le 3 analoghe, che hanno $q = 0$.

Elaborazione digitale di immagini (anche medicali). I computer, nelle elaborazioni digitali, non vedono con gli occhi se un puntolino bianco in una radiografia sta da una parte o dall'altra di una linea: fanno – ultrasemplificando – un calcolo numerico che si conclude con una verifica numerica del tipo maggiore o minore.



9.11 Esempio di rette: peso normale secondo Broca

La più nota e popolare formula per il peso ideale

peso uomini: altezza meno 100

(taluni modificano 100 in 104 per le sole donne) dà una retta del piano cartesiano dei punti $(altezza, peso)$.
E ammessa nel peso una tolleranza del 10%, si ottiene una regione del piano cartesiano, costituita da punti $(altezza, peso)$, delimitata da 2 rette.
(E 2 segmenti, a voler essere pignoli, e il segmento di destra non è ben definito: fino a quale altezza massima la formula è da ritenersi valida? Non vorremo mica definire il peso ideale di una persona ipoteticamente alta 3 metri, quindi caso unico gravemente malata?)

> Con l'altezza h (in centimetri) e il peso p (in chilogrammi) la Formula di Broca definiva (oggi esistono vari altri standard)

BOZZA - D

$peso\,normale = (h - 100) \pm 10\%$ (secondo alcuni, $h \geq 130$)

ovvero, con la corretta interpretazione del ±10%, in senso intervallare,

$peso\,normale \in$

$$[(h-100)-(h-100)10\%, (h-100)+(h-100)10\%]$$

e qua – a conti fatti – riconosciamo 2 rette e 3 regioni del piano Ohp:

sottopeso $p < 0.9\,h - 90$ (sotto la retta)

sovrappeso $p > 1.1\,h - 110$ (sotto la retta)

e la terza regione, dei normopeso, è quella fra le 2 rette. (Secondo certi Autori, ±20% invece che ±10%).

BOZZA - DRAFT

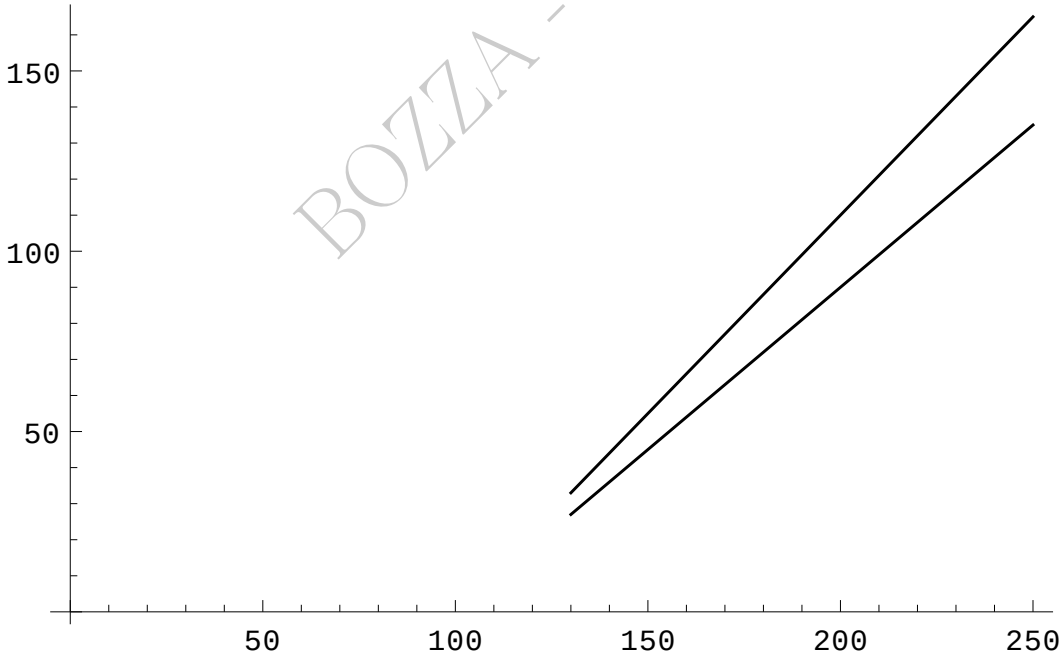


Figure 8: Le 2 rette estreme per i normopeso secondo la Formula di Broca con la tolleranza del 10%, limitatamente ad altezze fra 130 e 250 cm.

Come sarebbe la situazione di 53 kg e 1.70 m? (Si converta in centimetri).

9.12 Nota finale sulle rette oblique

Le rette oblique con pendenza verso l'alto sono grafici di funzioni che "crescono sempre ugualmente".

Per esempio il volume V di liquido in una provetta al crescere del livello h oltre la parte "semisferica": $V = m\,h + q$, dove q è il volume della parte "semisferica" della provetta.

9.13 Nota sui modelli matematici

Il fatto che la formula di Broca per il peso ideale certamente smetta di avere senso in qualche punto imprecisato – non esiste alcun peso ideale per persone alte un chilometro e anzi è dubitabile già per 3 metri – è un caso specifico dell'osservazione generale seguente.

BOZZA - DRAFT

Molte formule delle Scienze Applicate hanno un qualche dominio di ragionevolezza in cui modellizzano bene la realtà sensibile, divenendo sempre meno sensate verso imprecisati estremi.

Leggiamo per esempio in un articolo scientifico (https://pubmed.ncbi.nlm.nih.gov/22211780/)

> The Mosteller formula underestimates BSA in the paediatric population and must be used with precautions because of low precision, most pronounced in neonates and infants.

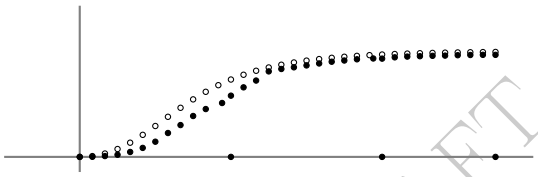
(E certo non calcolerà bene l'area della superficie corporea di un embrione di forma pressochè sferica).

9.14 I modelli per fare l'interpolazione, spesso bene

La semplice funzione (modello)

$$v(t) = \frac{14}{1 + t^{-2}\,\mathrm{e}^{-t}} \qquad 0 < t \leq 2.75$$

approssima per i 2.75 anni dal 1 dicembre 2020 (tempo 0) all'inizio del settembre 2023 (tempo 2.75, due anni e 9 mesi ovvero tre quarti d'anno) il numero totale di dosi di vaccino somministrate, a livello mondiale, in miliardi.
(Vedasi https://ourworldindata.org/covid-vaccinations con opzione `cumulative`).



Per esempio per $t = 2$ (cioè a 2 anni dall'inizio) si trova 13.5 (cioè 13 miliardi e mezzo; il valore vero era 13 miliardi al 1 dicembre 2022).

Con la formula data si può ottenere un valore per esempio per il, poniamo, 15 settembre 2022, che se non esatto sarà comunque verosimilmente vicino al vero: non avranno mica sospeso a livello mondiale le vaccinazioni dal 1 al 15 settembre per poi fare nella seconda metà di settembre quelle di tutto il mese, no?
L'interpolazione riesce *spesso* bene.

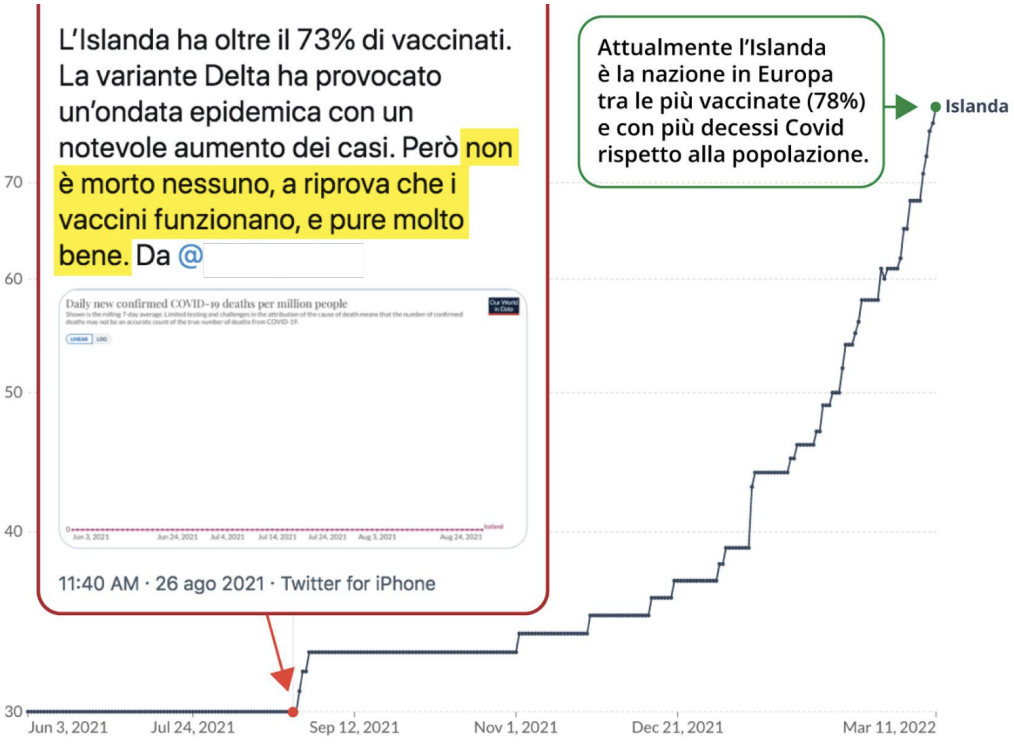
BOZZA - DRAFT

9.15 I modelli per fare l'estrapolazione, spesso male

Se un valore evolve nel tempo secondo una legge empirica (sperimentale) ci viene naturale azzardare una previsione per il futuro. Questo processo è molto meno garantito dell'interpolazione.
Se abbiamo di giorno in giorno, per un fenomeno, i valori

0 0 0 0 0 0 0 0 0

è ovvio che ci viene naturale – in assenza di altre informazioni – azzardare il valore 0 per il giorno successivo.


L'Islanda ha oltre il 73% di vaccinati. La variante Delta ha provocato un'ondata epidemica con un notevole aumento dei casi. Però non è morto nessuno, a riprova che i vaccini funzionano, e pure molto bene. Da @
Daily new confirmed COVID-19 deaths per million people
11:40 AM · 26 ago 2021 · Twitter for iPhone
Attualmente l'Islanda è la nazione in Europa tra le più vaccinate (78%) e con più decessi Covid rispetto alla popolazione.
Islanda
70
60
50
40
30
Jun 3, 2021
Jul 24, 2021
Sep 12, 2021
Nov 1, 2021
Dec 21, 2021
Mar 11, 2022

BOZZA

ESERCIZI SULLA LEZIONE 9

9.16 Esercizio risolto a – Retta e sangue

$_{\mu 2018}$ * Supponiamo che nel sangue di una persona una certa sostanza abbia alle 8:00 di mattina concentrazione 50 mg/dl e alle 6:00 di sera concentrazione 82 mg/dl. Esprimendo i tempi in ore, dallo 0 della mezzanotte, i valori dati individuano 2 punti nel piano cartesiano con i tempi t sull'asse delle ascisse e le concentrazioni y sull'asse delle ordinate (con o senza unità di misura).
Trovare la retta per i 2 punti espressa in forma esplicita, senza unità di misura. (Con essa si potrebbe ipotizzare la concentrazione in qualunque orario intermedio, supponendo una variazione lineare, che in assenza di altre indicazioni appare la più ragionevole approssimazione; ma qua non ne faremo nulla).

Svolgimento
Osservato che le 6:00 di sera sono le 18:00, cioè a 18 ore dal tempo 0, i valori dati individuano i 2 punti del piano cartesiano

$$(8\,h, 50\,mg/dl), \quad (18\,h, 82\,mg/dl) \quad \text{ovvero meglio} \quad (8, 50), \quad (18, 82)$$

e ricordando la formula della retta per 2 punti troviamo

$$\frac{t - t_2}{t_1 - t_2} = \frac{y - y_2}{y_1 - y_2}$$

$$\frac{t - 18}{8 - 18} = \frac{y - 82}{50 - 82}$$

$$\frac{t - 18}{-10} = \frac{y - 82}{-32} \quad \Big/ \cdot (-1) \cdot 10 \cdot 32$$

$$32\,(t - 18) = 10\,(y - 82)$$

$$32\,t - 576 - 10\,y = -820$$

$$-10\,y = -32\,t - 244$$

$$\boxed{y = 3.2\,t + 24.4}$$

anche esprimibile con

$$\boxed{y = \tfrac{16}{5}\,t + \tfrac{122}{5}}$$

(Da cui p.es. a mezzogiorno, $t := 12$, la concentrazione ipotetica 62.8 mg/dl).

BOZZA - DRAFT

9.17 Esercizio risolto b – Retta e sangue

$_{\mu 2018}$ * Supponiamo che nel sangue di una persona una certa sostanza abbia alle 6:00 di mattina concentrazione 70 nmoli/L e alle 9:00 di sera 150 nmoli/L. Esprimendo i tempi in ore, dallo 0 della mezzanotte, i valori dati individuano 2 punti nel piano cartesiano con i tempi t sull'asse delle ascisse e le concentrazioni y sull'asse delle ordinate (con o senza unità di misura, h e poi nmoli/L).
Con la retta per i 2 punti si può ipotizzare la concentrazione in qualunque orario intermedio, supponendo una variazione lineare, che in assenza di altre indicazioni appare la più ragionevole approssimazione. Con l'equazione esplicita di quella retta, senza unità di misura, calcolare l'ora in cui la concentrazione è salita a ≥ 110 (nmoli/L, unità di misura che non esprimiamo per semplicità).

SVOLGIMENTO
Osservato che le 9:00 di sera sono le 21:00, cioè a 21 ore dal tempo 0, i valori dati individuano i 2 punti del piano cartesiano

$$(6\,h, 70\,nmoli/L), \quad (21\,h, 150\,nmoli/L) \quad \text{ovvero meglio} \quad (6, 70), \quad (21, 150)$$

e ricordando la formula della retta per 2 punti troviamo

$$\frac{t - t_2}{t_1 - t_2} = \frac{y - y_2}{y_1 - y_2}$$

$$\frac{t - 21}{6 - 21} = \frac{y - 150}{70 - 150}$$

$$\frac{t - 21}{-15} = \frac{y - 150}{-80} \quad \Big/ \cdot (-1) \cdot 15 \cdot 80$$

$$80\,(t - 21) = 15\,(y - 150)$$

$$80\,t - 1680 - 15\,y = -2250$$

$$-15\,y = -80\,t - 570$$

e dividendo per -15

$$y = \frac{16}{3}\,t + 38$$

e ora risolviamo la disequazione della concentrazione $y \geq 110$

$$\frac{16}{3}\,t + 38 \geq 110$$

$$\frac{16}{3}\,t \geq 110 - 38$$

$$\frac{16}{3}\,t \geq 72$$

BOZZA - DRAFT

$$t \geq 72 \cdot \frac{3}{16} = 9 \cdot \frac{3}{2} = \frac{27}{2} = 13.5$$

cioè a 13.5 ore dalla mezzanotte, cioè 13 ore e mezza, cioè alle

$$\boxed{13{:}30}$$

BOZZA - DRAFT

10 Piano cartesiano – II parte

10.1 Iperbole equilatera riferita agli asintoti

Il grafico della funzione *passaggio al reciproco*, già vista,

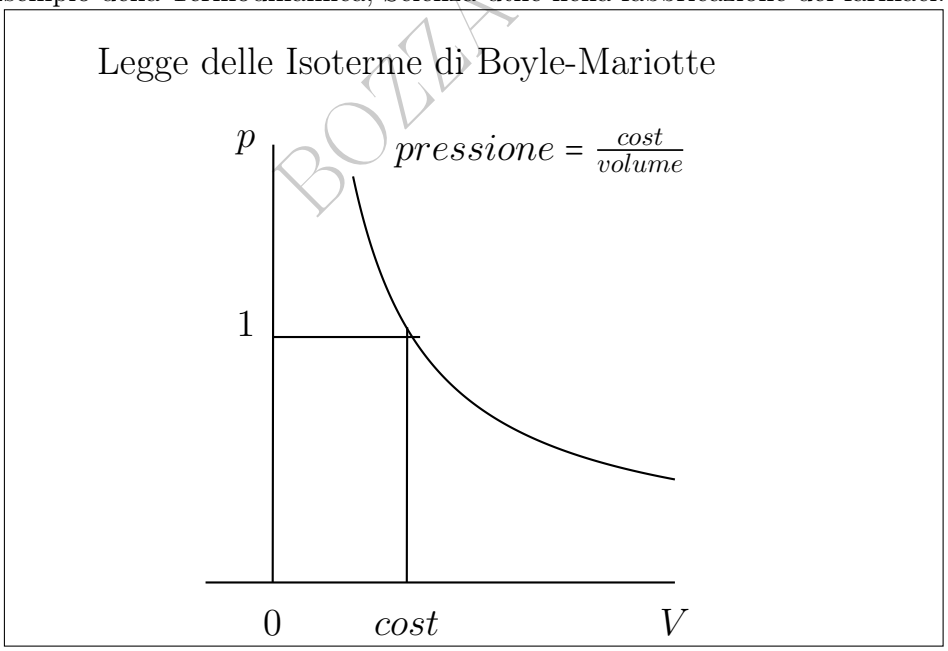
$$y = \frac{1}{x}$$

è una ***curva***[61] e più in generale

$$y = \frac{cost}{x} \qquad cost \neq 0$$

ha grafico una curva detta ***iperbole equilatera (riferita agli asintoti)*** e questa è una delle formule di più ampia ricorrenza nelle Scienze Applicate.

Esempio della Termodinamica, Scienza utile nella fabbricazione dei farmaci.



Queste funzioni $y(x) = \frac{cost}{x}$ sono caratterizzate da

"al raddoppiare di x dimezza y e viceversa"

[61] Una definizione rigorosa di *curva* è alquanto complessa, e in questa trattazione elementare supporremo noto il concetto. Le curve più semplici sono le rette.

BOZZA - DRAFT

10.2 Circolo e cerchio

Tutti i grafici sono *curve*,
e hanno un'equazione *esplicita* $y = f(x)$,
ma esistono curve che non sono grafici,
e hanno un'equazione *implicita* $f(x, y) = 0$.

In particolare il ***circolo*** di raggio r e centro (x_0, y_0):

$$(x - x_0)^2 + (y - y_0)^2 = r^2$$

Altri Autori lo chiamano *cerchio* ma in questo testo il ***cerchio*** sarà la regione "dentro" un circolo.
Questa *disequazione*

$$(x - x_0)^2 + (y - y_0)^2 \leq r^2$$

se verificata rappresenta lo *stare dentro il circolo* ovvero *appartenere* al corrispondente *cerchio*.
Ricordiamo
la lunghezza del circolo $2\pi r$
l'area del cerchio πr^2.

BOZZA - DRAFT

10.3 Diagramma bolle e sua ambiguità

L'area del cerchio può essere usata per rappresentare un valore numerico: i ***diagrammi a bolle*** possono rappresentare 1 o 2 o 3 liste di valori corrispondenti, come vedremo.

Esempio 1 – unidimensionale. Ecco un esempio semplicissimo, in cui 1 lista di 6 numeri viene rappresentata con 6 cerchi, variamente colorati per identificarli. Il cerchio rosso, il più grande, rappresenta i morti di covid-19 in Italia nella prima ondata del 2020.

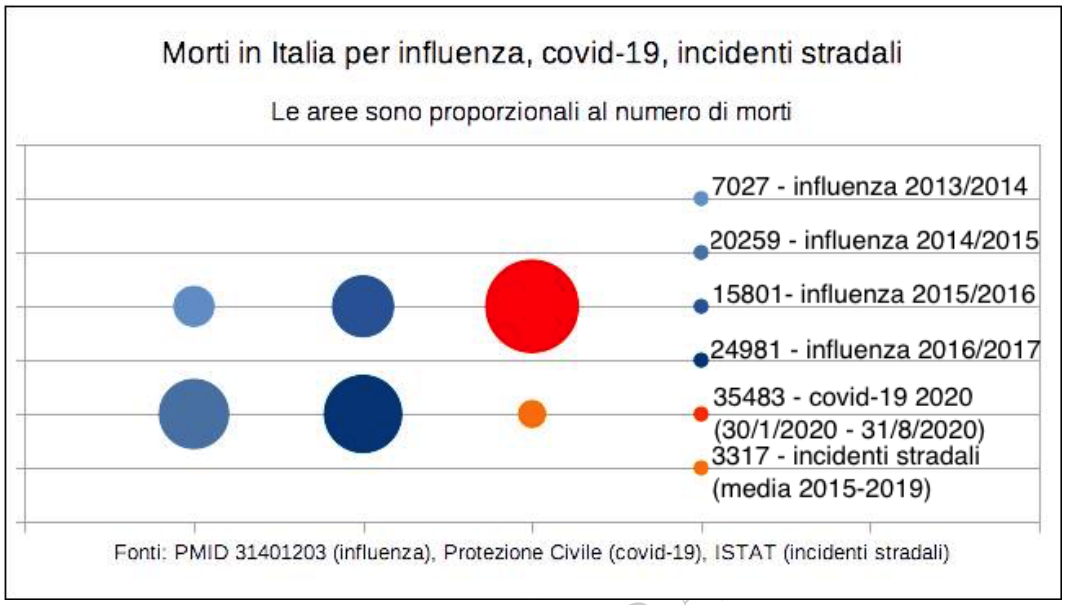


Figure 9: Esempio di rappresentazione grafica di 6 dati numerici

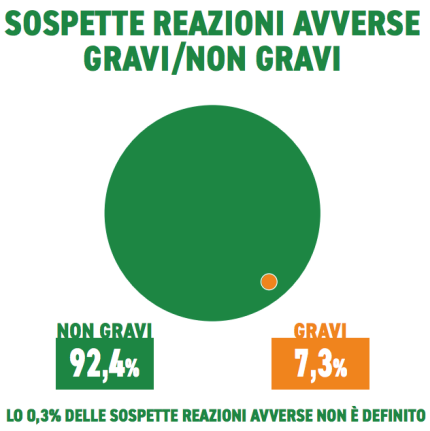
BOZZA - DR

(Si rassicuri l'automunito lettore: nonostante la ciclopica perdita di anni di vita umana causata dai veicoli, con incidenti causati magari da 1 che uccide 2 o 3, anche bambini o giovani con aspettativa di vita residua di 50 anni, nulla di simile alle limitazioni generate dal covid sta per avvenire contro i veicoli, che muovono un mercato mondiale smisurato, con un numero di veicoli a 4 ruote dell'ordine di 10^9 che consumano carburante dell'ordine di 10^{12} litri all'anno).

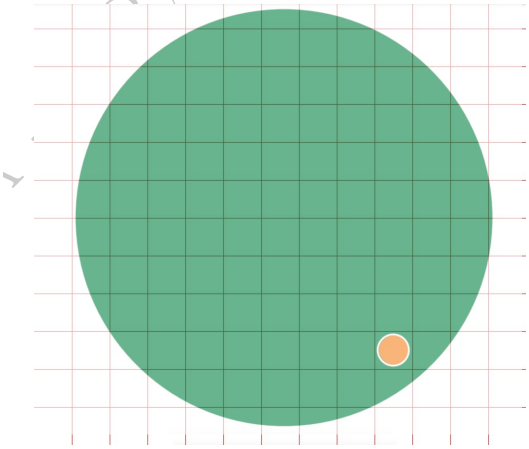
Purtroppo secondo altri Autori non l'area dei cerchi bensì il diametro rappresenta il terzo valore di ogni terna di valori, per cui i diagrammi a bolle di fatto possono essere molto fuorvianti guardandoli, e facendoli bisogna specificare lo standard usato.
E ancor peggio coi diagrammi a bolle tridimensionali.

Esempio 2 – unidimensionale problematico. Con il covid-19 c'è stato un fiorire di fake news sui vaccini, spesso di antivaccinisti, e questo si sa. Ma vediamo in questo esempio la mostruosa sproporzione fra l'area del cerchio arancio e quello verde: mentre si

afferma che l'arancio rappresenta il 7,3% la sua area è meno dell'1%. Guardando bene, è possibile che abbiano usato lo standard di rappresentare il dato con il diametro invece che con l'area, ma allora avrebbero dovuto accostare i cerchi e non sovrapporli perchè sovrapponendoli qualunque persona ragionevole riterrà che il dato sia rappresentato dall'area.
È ovvio che falsi di questo genere generano diffidenza verso i vaccini.



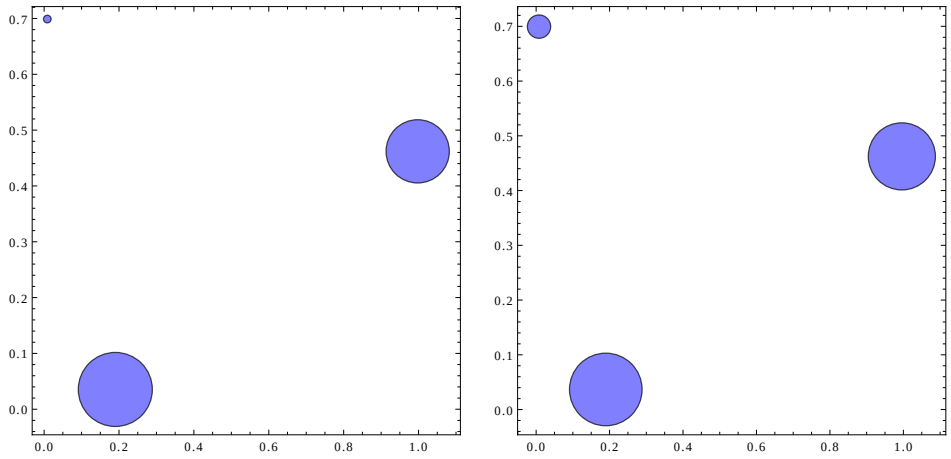
Fonte: AIFA, p. 5 in: Link



Con quadrettatura sovrapposta

Edit. Una serie di inchieste ha poi rivelato che era una cosa deliberata: nelle indicazioni che furono mandate ai grafici, l'AIFA scrive che l'area del cerchio delle manifestazioni gravi non sia proporzionale, "potrebbe essere più piccolo".

Esempio 3 – problematico.



Esempio di 2 rappresentazioni delle stesse 3 terne di numeri coi diagrammi a bolle, a sinistra con le aree e a destra coi diametri.

BOZZA DRAFT

Problematicità ulteriore del caso tridimensionale. Ovviamente il diagramma a bolle non funziona bene se 2 o più bolle si sovrappongono esattamente (o quasi esattamente se le bolle vengono rappresentate piene), e in quel caso il diagramma ottenuto è **fuorviante**.

Esercizio$_\mu$ Si verifichi se il punto $(2,2)$ sta nel circolo

$$x^2 + y^2 + 2\,x - 10\,y + 1 = 0$$

Esercizio$_\mu$ Esplicitare i 2 semicircoli costituenti il circolo.
(Si troverà $y = 5 \pm \sqrt{-x^2 - 2x + 24}$).

Altre curve in equazione implicita sono queste, con $a, b > 0$:

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \qquad \text{un'}\textit{iperbole}$$

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = -1 \qquad \text{un'altra } \textit{iperbole}$$

e le rette $y = \pm\frac{b}{a}\,x$ si chiamano *asintoti* dell'iperbole;

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \qquad \textit{ellisse} \text{ (in } \textit{forma canonica}^{(62)})$$

[62] *Canonico* significa, con parola internazionale di uso più moderno, *standard*.

e la disequazione

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1$$

se verificata rappresenta lo *stare dentro l'ellisse*. Se $a > b$ l'ellisse appare "ribassata" e se $b > a$ l'ellisse appare "allungata verso l'alto".

Esercizio$_\mu$ Con $a := 3$ e $b := 4$ si scrivano, esplicitino (per archi) e disegnino le 2 iperboli e l'ellisse.

10.4 La parabola

La parabola è una curva del piano euclideo (e poi cartesiano) con la forma più o meno della lettera

U

ovvero del simbolo $\subset$, che si estende indefinitamente, allungandosi molto più che allargandosi (comunque si allarga indefinitamente). Nel piano cartesiano viene di solito orientata con asse di simmetria verticale, più o meno così: $\cup$ oppure $\cap$.
Nelle varie applicazioni scientifiche, può interessarne una parte piuttosto che un'altra (raramente tutta: è un'astrazione matematica, come la retta).

BOZZA - DRAFT

Nel piano euclideo tutte le parabole sono *simili*, cioè sovrapponibili salvo ingrandimento, come i circoli ma è molto meno evidente.

Esempio, duplice. Si hanno – oggi si usa di meglio – le Formule di Keys per il peso ideale y dell'uomo e della donna, data l'altezza x:

$$y = 22.1\, x^2$$

$$y = 20.6\, x^2$$

(ragionevolmente considerabili per $x \geq 1.30$, e fino a un x_{max} ragionevolmente fissabile; si usano i metri e i chilogrammi).

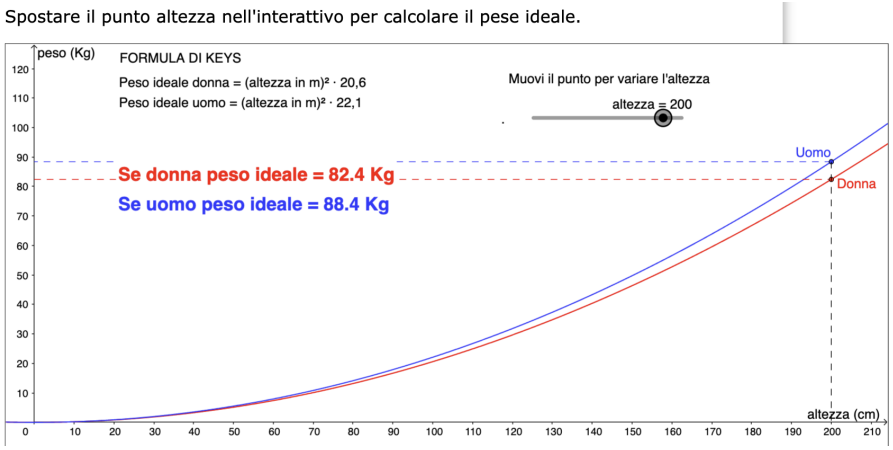


Figure 10: Screenshot da https://www.isissvalleseriana.it/Matematica/matecomeprezzemolo/peso_ideale_formula_di_keys.html

BOZZA - DRAFT

10.5 Equazione della parabola con asse verticale

La generica **parabola** con asse verticale ha equazione

$$y = a\,x^2 + b\,x + c \qquad a \neq 0$$

La disequazione

$$y \geq a\,x^2 + b\,x + c$$

rappresenta lo stare *sopra* la parabola (in *senso debole*, per avere il *senso forte* si ponga >, e questo è un fatto generale).

E ovviamente

$$x = a\,y^2 + b\,y + c \qquad a \neq 0$$

rappresenta la generica parabola con asse orizzontale.

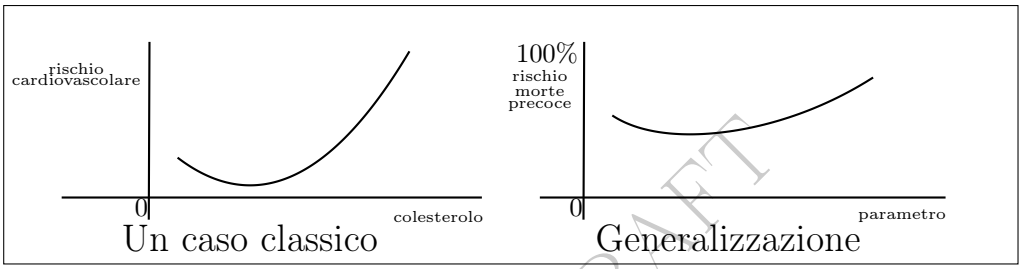
La parabola è simmetrica rispetto ad una retta detta *asse* che interseca la parabola in un punto detto *vertice* della parabola.

Esempio. in uno studio scientifico si è stabilito

$$weight \leq -3{,}767 + 89.11 * length + 1.237 * length^2 \qquad 40 \leq lenght \leq 55$$

(con la lunghezza in centimetri e il peso in chilogrammi) rappresenta bene una soglia per il 90% dei neonati (Hospital Israelita Albert Einstein, San Paolo, Brasile, 1995-1998). (Con 50 cm dà 3781 grammi). (Si noti che usano il punto decimale, e la virgola come separatore delle migliaia), e lo capiamo dal numero in 89.11).

10.6 Le curve a forma di J



BOZZA - DRAFT

La parabola ci offre un semplice modello del concetto, più generale, di *curva a forma di J*. Traiamo da Wikipedia, l'enciclopedia libera, alla voce *J curve*:

> In medicine, the 'J curve' refers to a graph in which the x-axis measures either of two treatable symptoms (blood pressure or blood cholesterol level) while the y-axis measures the chance that a patient will develop cardiovascular disease (CVD). It is well known that high blood pressure or high cholesterol levels increase a patient's risk. What is less well known is that plots of large populations against CVD mortality often take the shape of a J curve which indicates that patients with very low blood pressure and/or low cholesterol levels are also at increased risk.

È nozione comune che il colesterolo troppo alto possa aumentare fare male, ma poco considerato è il caso del colesterolo troppo basso, e similmente per la pressione.
Senza voler qua fare Medicina, citiamo comunque da un articolo[63] scientifico:

[63] https://www.thelancet.com/journals/lanwpc/article/PIIS2666-6065(23)00192-X/fulltext

> Both low and high HDL-C were associated with increased mortality risk. We recommended 50–79 mg/dL as the optimal range of HDL-C among Chinese adults.

È ben possibile che tali curve a forma di J siano tanto diffuse quanto poco considerate

BOZZA - DRAFT

10.7 Area del segmento parabolico, ed epidemie

Mentre l'area del triangolo è notoriamente $\frac{1}{2}\,b\,h$, l'area del segmento parabolico – la cui definizione lasciamo all'intuitivo lettore – è

$$A = \frac{2}{3}\,b\,h$$

Questo ci permette di stimare a colpo d'occhio il numero totale di morti di un'ondata epidemica disponendo di un istogramma a barre (bar chart) dei morti giornalieri, se – come spesso succede – esso disegna più o meno un segmento parabolico:

$$\boxed{\text{morti di un'ondata epidemica} = \frac{2}{3}\,durata \times picco} \qquad (14)$$

Ovviamente la durata va intesa in giorni e sarà bene non prendere il valore di picco effettivo ma mediarlo coi valori dei giorni vicini, anche a occhio, o ancor meglio usare, se disponibile, il massimo della *media mobile a 7 giorni*.

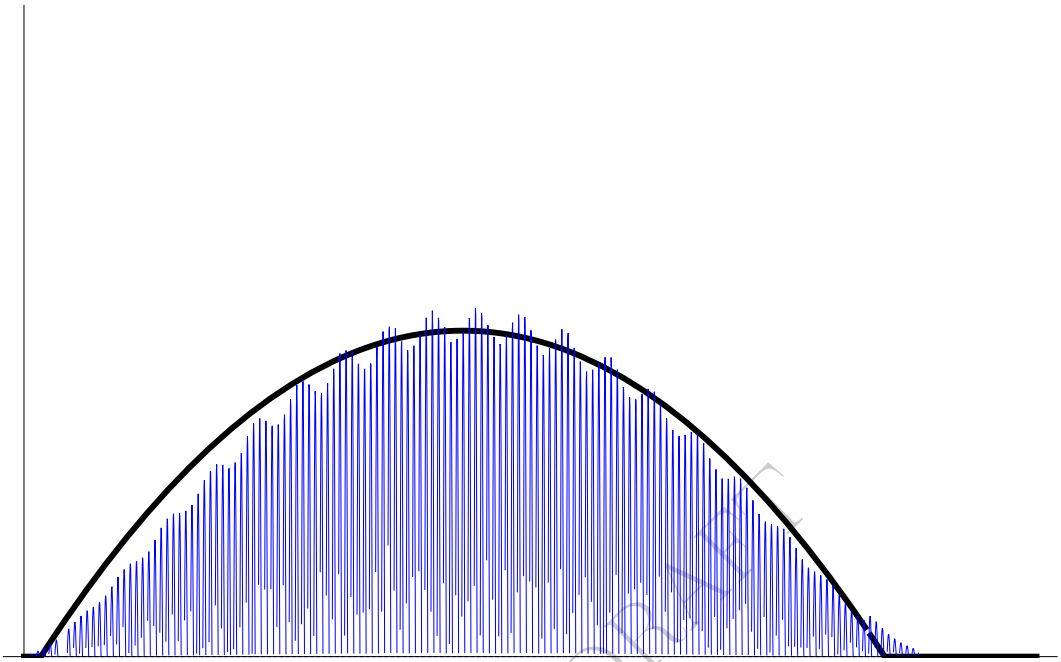


Figure 11: Modellizzazione del numero giornaliero di morti di un'epidemia con una parabola

BOZZA - DRAFT

In generale una migliore approssimazione dell'andamento del numero dei morti delle epidemie si ha con curve "a campana" che crescono poco all'inizio e dotate di una "coda destra".

10.8 Note finali sulle coniche

L'ellisse – in generale, anche se spostata in qualunque punto del piano, e allora con equazione diversa – è il *luogo*[64] dei punti del piano che ha costante la somma delle distanze da 2 punti detti *fuochi*.

Un'analoga proprietà, con la differenza, vale per le iperboli.

La parabola è luogo dei punti per i quali è costante la distanza da un punto – detto *fuoco* – e da una retta – detta *direttrice*. nel *vertice*.

Parabole, ellissi e iperboli – disposte in qualunque modo nel piano – si chiamano **coniche**, e comprendono i circoli (che sono ellissi con $a = b$, se il centro è O).

[64] Insieme. Termine usato in geometria.

Sono intersezione del piano cartesiano con un *cono completo*.

10.9 Risolvere le equazioni con calcoli o disegni?

Il piano cartesiano permette molte volte di risolvere graficamente equazioni e sistemi di equazioni.

Come norma generale bisogna risolvere analiticamente le equazioni e i sistemi di equazioni ed eventualmente disequazioni, cioè coi calcoli, per quanto possibile, perchè esistono casi in cui il disegno inganna, perfino se fatto col computer.

BOZZA - DRAFT

Tuttavia fare i disegni è comunque utilissimo:

(a) per verificare la correttezza dei risultati trovati;

(b) per una comprensione complessiva della situazione;

(c) per presentare divulgativamente i risultati, o per presentarli in modo chiaro in relazioni tecniche;

(d) per un tentativo seppure incerto di soluzione almeno approssimata quando il calcolo analitico è impossibile o proibitivamente difficile.

Esempio. Con rifermento al punto (d) sopradetto, si consideri questo *sistema di 2 equazioni in 2 incognite di* 6° *grado*, molto difficile da risolvere analiticamente:

$$\begin{cases} x^3 + y^3 - 3\,x\,y = 0 \\ x^2 + y^2 = 1 \end{cases}$$

Dal disegno (sperabilmente, e, in effetti, in questo caso non ingannevole) vediamo che ci sono 4 soluzioni, e abbiamo anche loro approssimazioni, tanto migliori quanto più grande e preciso è il disegno.

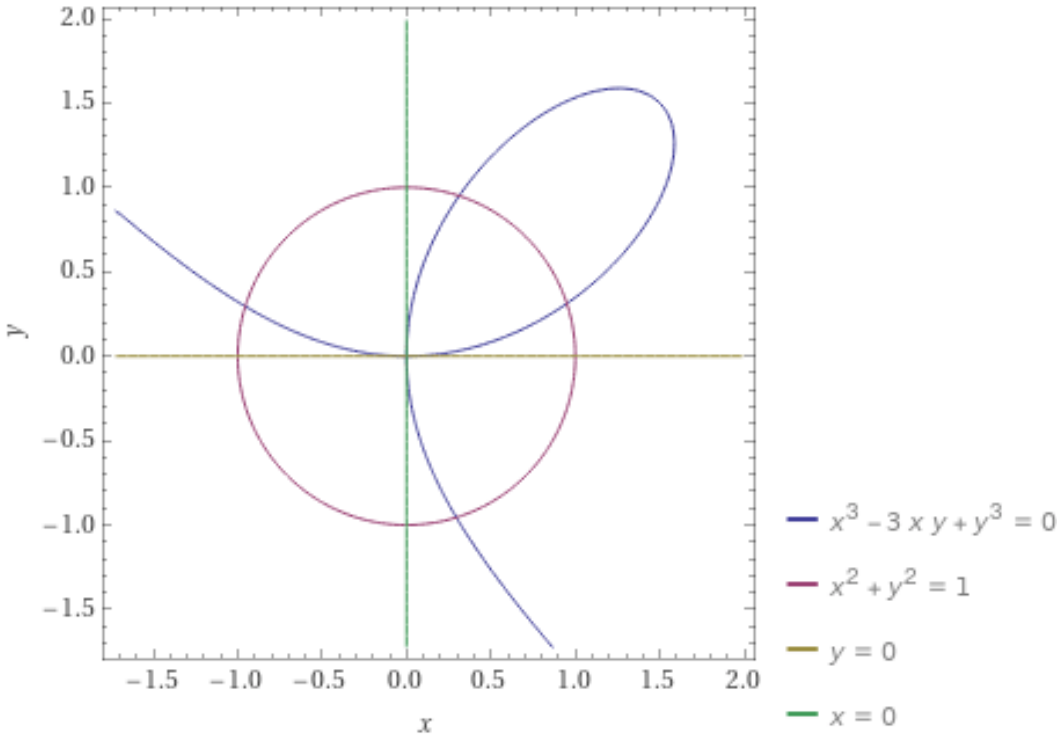


Figure 12: Intersezioni del folium di Cartesio col circolo unitario di centro l'origine

BOZZA

La soprastante figura è stata ottenuta con la più potente intelligenza artificiale matematica attualmente online gratuitamente, WolframAlpha, (qua utilizzata per ben piccola cosa) con cui il futuro professionista farà bene a familiarizzare. Si segua questo LINK.
Si noti che WolframAlpha dice che non ci sono soluzioni, ma questo è dovuto al fatto che per fare un disegno chiaro sono state aggiunte alle 2 equazioni date le 2 equazioni degli assi cartesiani, e togliendo queste ultime dà le 4 soluzioni cercate, approssimate con parecchi decimali: LINK.

10.10 Altre figure nel piano cartesiano

Con le rette (non verticali) e le coniche, e i grafici di $\operatorname{sgn}(x)$ e $|x|$, abbiamo visto grafici che ricorrono un'infinità di volte nelle Scienze Applicate, e ce ne sono infiniti altri, per esempio (quello

della funzione)

$$y = \sqrt[3]{x}$$

e, oltre alle rette verticali, infinite altre curve in rappresentazione implicita, per esempio il *folium di Cartesio*

$$x^3 + y^3 - 3\,a\,x\,y = 0 \qquad a > 0$$

la cui esplicitazione è proibitivamente difficile al livello di questo testo elementare; ma si verifichi la potenza e l'utilità di WolframAlpha scrivendovi

x^3+y^3-3xy=0

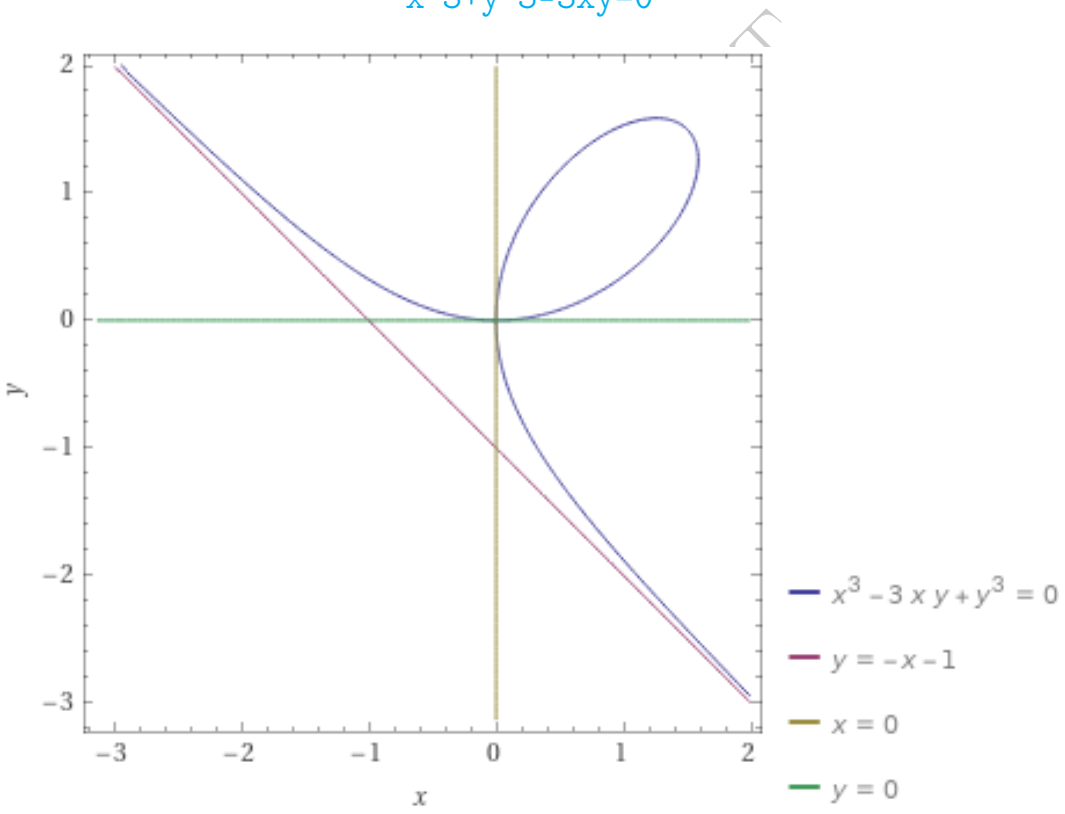


Figure 13: Folium di Cartesio e suo asintoto $y = -x - 1$

(avendo scelto $a := 1$ per fare un esempio concreto).

10.11 Esempio farmaceutico.

In questo documento*link->*della Food and Drug Administration (FDA), ente statunitense internazionalmente considerato autorevolissimo,

ci si impegni a capire le ultime 2 righe a pagina 22 e le prime 6 a pagina 23, guardando attentamente le figure, riguardanti la produzione dei farmaci.
Si apprezzino poi le figure a pagina 24.

10.12 Le famiglie di curve

Un esempio di ***famiglia*** (insieme) di curve in rappresentazione implicita ci viene dalla Termodinamica, con la Legge di van der Waals

$$\left(p + a\,\frac{n^2}{V^2}\right)(V - n\,b) = n\,R\,T$$

nel piano (V, p), con R, n, a, b costanti, al variare di T.

Si consideri come ulteriore esempio di *famiglia* (insieme) di curve in rappresentazione implicita il *folium di Cartesio*

$$x^3 + y^3 - 3\,a\,x\,y = 0 \qquad a > 0$$

la cui esplicitazione è proibitivamente difficile al livello di questo testo elementare; ma si verifichi la potenza e l'utilità di WolframAlpha scrivendovi

```
x^3+y^3-3x y=0
```

(avendo scelto $a := 1$ per fare un esempio concreto).

BOZZA - DRAFT

11 Piano cartesiano – III parte

I *sistemi lineari*, che andiamo a trattare, ricorrono un'infinità di volte nelle Scienze Applicate.

11.1 Sistemi lineari di 2 equazioni in 2 incognite

@ Questo argomento non è capitato agli esami ma capiterà (2025)

Il sistema lineare (ovvero di primo grado) di 2 equazioni in 2 incognite x e y

BOZZA - DRAFT

Esempio

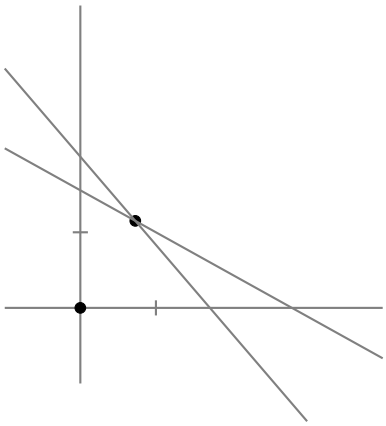
$$\begin{cases} a\,x + b\,y + c = 0 \\ a'\,x + b'\,y + c' = 0 \end{cases} \qquad \begin{cases} 7\,x + 6\,y - 12 = 0 \\ 5\,x + 9\,y - 14 = 0 \end{cases}$$

se $(a, b) \neq (0, 0)$ e $(a', b') \neq (0, 0)$ rappresenta geometricamente l'intersezione di 2 rette nel piano cartesiano. Potrebbero non avere intersezione, se esse sono parallele e distinte, o infinite soluzioni, se esse sono coincidenti. Questi 2 casi sono caratterizzati dal *determinante* nullo

$$\det = a\,b' - b\,a' = 0$$

e si risolveranno con cautela analiticamente e contemporaneamente facendo un disegno. Non ce ne occuperemo.

Ipotizziamo ora $\det \neq 0$, che è il caso generale ed è l'unico che ci interessa. Allora abbiamo 2 rette incidenti in 1 punto, le cui coordinate x_0 e y_0 costituiscono la soluzione del sistema di equazioni.



Ognuna delle eventuali equazioni mancanti di 1 delle 2 variabili si risolve subito, e il valore di x o y trovato si sostituisce nell'altra equazione, se in essa c'è quella variabile, e così quell'equazione diventa un'equazione di primo grado in una sola variabile, e si risolve subito anch'essa.

Ipotizziamo ora che ci siano sia la x che la y in entrambe le equazioni: esplicitiamo y dalla prima e sostituiamola nella seconda, che diventa così un'equazione di primo grado nella sola x. Trovata la x, si trova la y dalla sua espressione precedentemente esplicitata dalla prima equazione.

BOZZA - DRAFT

11.2 Esempio di sistema: miscela di 2 soluzioni

Immagina di avere due soluzioni con differenti concentrazioni a e b di un principio attivo, e vuoi ottenere una miscela con una concentrazione u desiderata. Devi determinare quali volumi x e y di ciascuna soluzione usare per ottenere un volume desiderato v che abbia la concentrazione desiderata u.
Con un procedimento logico non semplicissimo[65] si ottiene questo

[65] Per fissare le idee, per produrre il sistema consideriamo questo esempio:
Soluzione 1 ha una concentrazione di farmaco a = 10 mg/mL
Soluzione 2 ha una concentrazione di farmaco b = 2 mg/mL
Vuoi ottenere un volume totale
v = 100 mL con una concentrazione desiderata di

sistema:

$$\begin{cases} x + y = v \\ a\,x + b\,y = u\,v \end{cases}$$

11.3 Esempio risolto di sistema

Si risolva

$$\begin{cases} F = C & \text{(stesso valore numerico)} \\ F = 1.8\,C + 32 & \text{(formula di conversione)} \end{cases}$$

(che determina il valore numerico – numero *puro*, senza unità di misura – della temperatura che ha lo stesso valore numerico nelle scale Celsius e Fahrenheit).

BOZZA - DRAFT

Esplicitiamo F dalla prima, che in effetti "è già risolta", e sostituiamo nella seconda ottenendo

$$C = 1.8\,C + 32 \qquad \Big/ + (-1.8\,C)$$

$$-0.8\,C = 32 \qquad \Big/ : 0.8$$

$$C = 32 : (-0.8)$$

$$\boxed{C = -40}$$

(40 gradi sotto 0, sia Celsius che Fahrenheit).

---

u = 5 mg/mL
Incognite:
x: Volume della soluzione 1 (in mL).
y: Volume della soluzione 2 (in mL).
Sistema di equazioni:
Equazione del volume totale:
x+y=100 (Il volume totale deve essere 100 mL).
Equazione della concentrazione:
10x+2y=5x100
(La quantità totale di farmaco deve essere 500 mg, ovvero
5mg/mL x 100mL).

11.4 Esempio parzialmente risolto di sistema

Abbiamo una provetta contenente 2 liquidi non miscibili di volumi 7 ml e 6 ml con pesi specifici incogniti e peso complessivo 12 g (si pesa la provetta e sottrae il peso del vetro) e un'altra provetta contenente gli stessi liquidi di volumi 5 ml e 9 ml rispettivamente e peso complessivo 14 g.

Per trovare i pesi specifici incogniti x e y risolviamo il sistema lineare (di 2 equazioni in 2 incognite)

(x g/ml) 7 ml+ (y g/ml) 6 ml=12 g

(x g/ml) 5 ml+ (y g/ml) 9 ml=14 g

ovvero

$$\begin{cases} 7\,x + 6\,y - 12 = 0 \\ 5\,x + 9\,y - 14 = 0 \end{cases}$$

che ha

$$\det = 7 \cdot 9 - 6 \cdot 5 = 72 - 30 \neq 0$$

e si troveranno i pesi specifici (qua dati senza l'unità di misura g/ml) $x = \frac{8}{11} \approx 0.727$ e $y = \frac{38}{33} \approx 1.151$.

Si tratta delle 2 rette della soprastante figura.

BOZZA - DRAFT

11.5 Sistemi lineari di n equazioni in n incognite

Il metodo visto si estende ai sistemi lineari di 3 equazioni in 3 incognite[66] x, y e z e poi anche di più ma può essere malagevole.

Si esplicita z dalla prima equazione e si sostituisce nella seconda e la terza, che insieme costituiscono a questo punto un sistema lineare di 2 equazioni in 2 incognite x e y. Tuttavia, i sottocasi particolari, di 0 e infinite soluzioni, iniziano a diventare molteplici, e non sempre facilissimi da trattare, e dipendono da vari determinanti.

Analogo e via via sempre più complesso il caso dei sistemi lineari di n equazioni in n incognite.

[66] Geometricamente rappresenta l'intersezione di 3 piani nello *spazio cartesiano* $\mathbb{R}^3$.

11.6 Sistemi di dis/equazioni e sistemi generalizzati

Definizioni.

Un sistema di n equazioni in 1 incognita x, con la parentesi graffa (grande) che vale *et*, è il predicato (con opportune funzioni)

$$\begin{cases} f_1(x) = g_1(x) \\ \dots \\ f_n(x) = g_n(x) \end{cases}$$

e sostituendo gli = con segni di disuguaglianza si ha un sistema di n disequazioni in 1 incognita.
Naturalmente possono considerarsi sistemi con 2 o più incognite (per esempio il sistema lineare di 2 equazioni in 2 incognite) e anche sia con equazioni che disequazioni.

Non esiste un metodo generale per risolverli tutti.

BOZZA DRAFT

I predicati del tipo $f(x) \neq g(x)$, che potremmo chiamare *"inequazioni"*, non li nomineremo per nome considerandoli implicitamente compresi nelle disequazioni data l'equivalenza con $(f(x) - g(x))^2 > 0$.

Definizione.

Predicati come (19) e (20) che seguono, che sono come quelli qua sopra considerati ma in più ammettono anche il *vel*, si potrebbero chiamare *sistemi generalizzati di equazioni ed eventualmente disequazioni*.

11.7 Esempio di sistema non lineare

Esercizio$_\mu$
La Formula di Keys – fra i numerosi standard – calcola

$$peso\,ideale\,uomini = (statura\,in\,metri)^2 \times 22.1 \qquad (*)$$

Per quale statura (realistica) essa dà – per gli uomini – lo stesso peso della Formula di Broca

$$peso\,ideale = (statura\,in\,centimetri - 100) \pm 10\%$$

intesa senza la tolleranza del 10%?

Svolgimento
La Formula di Broca intesa senza la tolleranza del 10% è ovviamente

$$peso\,ideale = statura\,in\,centimetri - 100$$

ovvero

$$peso\,ideale = (statura\,in\,metri - 1) \times 100 \qquad (**)$$

da mettere a sistema con la $(*)$:

$$\begin{cases} p = 22.1\,h^2 & \text{che è la } (*) \\ p = 100\,(h-1) & \text{che è la } (**) \end{cases}$$

che è un sistema (non lineare) di 2 equazioni in 2 incognite, e uguagliando si ha successivamente

$$\begin{gathered} 22.1\,h^2 = 100\,(h-1) \\ 22.1\,h^2 - 100\,(h-1) = 0 \\ 22.1\,h^2 - 100\,h + 100 = 0 \\ h_{1,2} = \frac{50 \pm \sqrt{50^2 - 22.1 \cdot 100}}{22.1} \\ h \approx 3.03 \quad \vee \quad h \approx 1.49 \end{gathered}$$

e ovviamente la prima soluzione non ha senso dal punto di vista della realtà sensibile (è ovvio che tutte queste formule hanno un certo dominio entro il quale funzionano bene, e poi perdono significato nella realtà).

BOZZA DRAFT

$$\boxed{\approx 149\,cm}$$

(Il peso ideale corrispondente sarebbe $\approx 49\,kg$).

11.8 Parabole e funzioni e dis/equazioni di secondo grado

Come detto, ogni parabola con asse verticale ha un'equazione

$$y = a\,x^2 + b\,x + c$$

con qualche $a, b, c \in \mathbb{R}$, $a \neq 0$.

La parabola è un oggetto geometrico a cui sono associati – in particolare – 2 oggetti algebrici:
l'equazione di secondo grado e
la disequazione di secondo grado,
che andiamo a esaminare.

11.9 3 metodi per disegnare una parabola

(a) Un disegno approssimativo (del grafico) della parabola (di equazione)

$$y = a\,x^2 + b\,x + c$$

(essendo a, b, c valori numerici noti) si ottiene per punti, e i 7 punti con queste coordinate *spesso* basteranno:

$x = -3; y = y(-3)$ ← calcolare con l'equazione soprastante
$x = -2; y = y(-2)$
$x = -1; y = y(-1)$
$x = -0; y = y(-0)$
$x = 1; y = y(1)$
$x = 2; y = y(2)$
$x = 3; y = y(3)$

(Oppure altri punti, se ritenuti più idonei).

BOZZA - DRAFT

(b) Oppure con i seguenti 7 passaggi, in effetti 8 in casi "sfortunati".

1) Si calcola il valore $-\frac{b}{2a}$.
2) Si segna quel valore sull'asse x.
3) Si disegna la retta verticale per il punto or ora segnato: quello è l'asse di simmetria della parabola.
4) Si calcola l'ordinata del vertice

$$y_V = y\left(-\frac{b}{2a}\right)$$

5) Si segna il vertice V sull'asse di simmetria, con l'ordinata or ora calcolata.
6) Si segna il punto $Q = (0, c)$ che ovviamente appartiene alla parabola.
7) Se V e Q sono 2 punti diversi, con l'aiuto dell'asse di simmetria potete disegnare approssimativamente la parabola facendola passare per quei 2 punti, con sorta di "vaschetta" in V.
8) Se invece $V = Q$, tenete anche conto che:
se $a > 0$ la parabola è rivolta verso l'alto e se $a < 0$ verso il basso.

(c) Ovviamente usare WolframAlpha per disegnare le parabole è più semplice, affidabile, e professionale.

11.10 Equazione di secondo grado

Consideriamo per $a \neq 0$ l'equazione detta di secondo grado

$$a\,x^2 + b\,x + c = 0$$

e la sua parabola associata, avente asse verticale,

$$y = a\,x^2 + b\,x + c$$

per esempio l'equazione e la parabola, associate fra loro,

$$2\,x^2 - x + 1 = 0 \qquad y = 2\,x^2 - x + 1$$

Le equazioni di secondo grado hanno una grande ricorrenza nelle Scienze. Per esempio in Chimica, quando si prepara una soluzione tampone, il pH della soluzione dipende dalla concentrazione degli acidi e delle basi e dalla costante di dissociazione acida K_a, per esempio 10^{-5}: in alcuni casi, per calcolare il pH è necessario risolvere quest'equazione di secondo grado

$$x^2 + K_a\,x - K_0\,C_a = 0$$

essendo C_0 la concentrazione iniziale dell'acido, per esempio 0.1 M. In nota[67] un esempio della Fisica.

BOZZA - DRAFT

Si definisce il discriminante

$$\boxed{\Delta := b^2 - 4ac} \tag{15}$$

Nel piano cartesiano ci sono

6 tipi di parabole con asse verticale,

di generica equazione $y = a\,x^2 + b\,x + c$,

a seconda

di sign(a) (positivo o negativo, 2 casi) e

di sign(Δ) (positivo o nullo o negativo, 3 casi).

[67] Una funzione di secondo grado dà lo spazio

$$s(t) = 9.81\,\frac{t^2}{2} + v_0\,t$$

percorso al tempo t da un corpo partito verso il basso con velocità v_0, nel campo gravitazionale terrestre, da altezza non troppo elevata, supposta la forma tale da rendere insignificante l'attrito con l'aria.

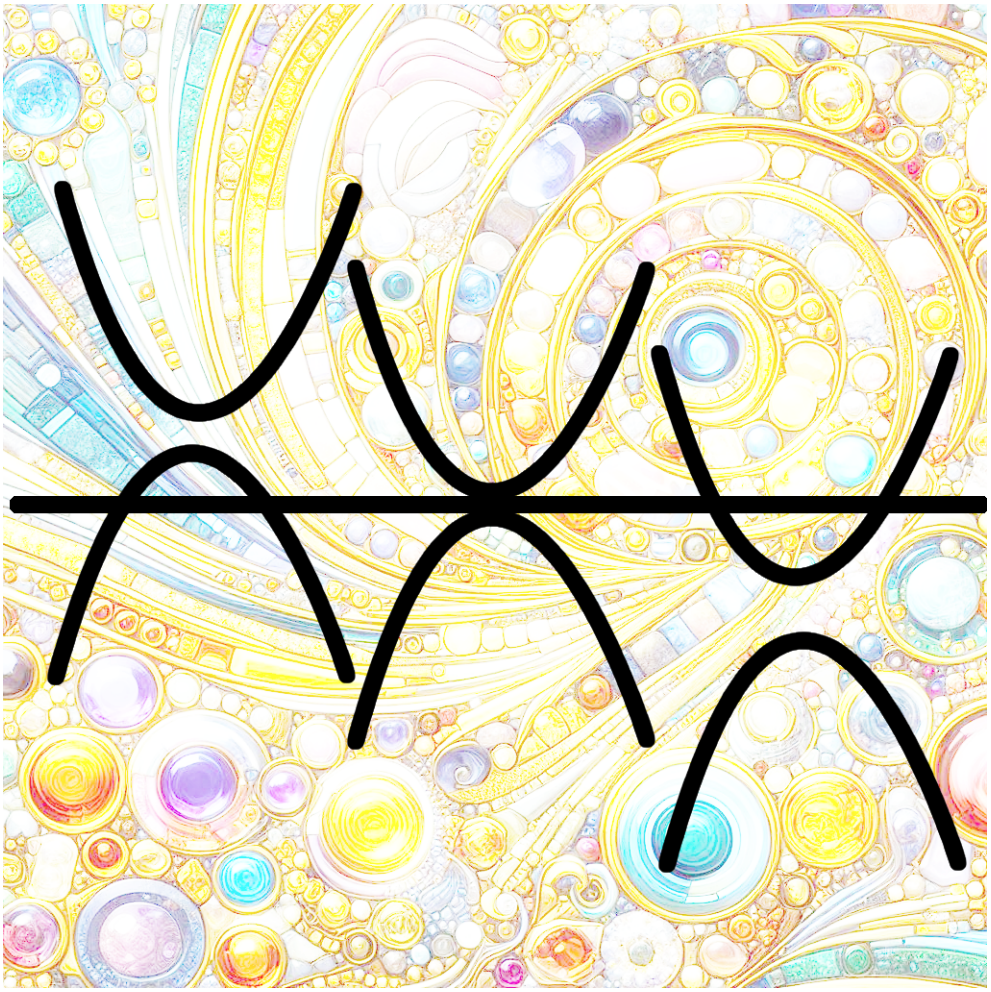


Figure 14: Bing Image Creator, rielaborata

La parabola interseca[68] l'asse x in

$$\boxed{x_{1,2} = \frac{-b \pm \sqrt{\Delta}}{2a}} \tag{16}$$

se $\Delta \geq 0$ e altrimenti mai, e sono le soluzioni dell'equazione di secondo grado associata.

[68] Vale anche questa *formula ridotta*

$$x_{1,2} = \frac{-\frac{b}{2} \pm \sqrt{\frac{\Delta}{4}}}{a} \qquad \frac{\Delta}{4} = \left(\frac{b}{2}\right)^2 - a\,c$$

utile se b è intero pari.

11.11 Fattorizzazione del polinomio di secondo grado

Se x_1 e x_2, dati dalla soprastante formula risolutiva dell'equazione di secondo grado, sono
• le *radici* (eventualmente coincidenti) del *polinomio* $a\,x^2 + b\,x + c$,
• ovvero le *soluzioni* dell'*equazione* $a\,x^2 + b\,x + c = 0$,
• ovvero se $(x_1, 0)$ e $(x_2, 0)$ sono le *intersezioni* della parabola con l'asse x,
allora vale questa
Formula di fattorizzazione del polinomio di secondo grado con $\Delta \geq 0$

$$\boxed{a\,x^2 + b\,x + c \equiv a\,(x - x_1)(x - x_2)} \tag{17}$$

11.12 Disquazione di secondo grado

Risolviamo per esempio

$$-2\,x^2 - 2\,x + 12 > 0 \qquad \Big/ : (-2) < 0$$

$$x^2 + x - 6 < 0$$

e adesso che il *coefficiente principale* (cioè quello di x^2) è 1 applichiamo la (17) con le soluzioni $x_1 = 2$ e $x_2 = -3$ dell'*equazione associata*, trovate con le (15) e (16):

$$(x - 2)(x + 3) < 0 \tag{18}$$

Risolviamo, in ogni caso col $>$, le 2 disequazioni di primo grado
$x - 2 > 0$ ... $x > 2$
$x + 3 > 0$ ... $x > -3$
e allora nella regione intermedia, $-3 < x < 2$, il prodotto è negativo, e in conclusione, siccome in questo caso avevamo il < 0 in (18),

$$\boxed{-3 < x < 2}$$

Questo procedimento del *prodotto dei segni* si può usare anche per più di 2 termini moltiplicati, magari con uno *schema di prodotto dei segni*, che non approfondiamo.

(Altri Autori risolvono la disequazione di secondo grado col disegno della parabola).

BOZZA DRAFT

11.13 Valore assoluto, e relative dis/equazioni

Il valore assoluto in $\mathbb{R}$, già definito,

$$|x| := \begin{cases} x & se\ x \geq 0 \\ -x & se\ x < 0 \end{cases}$$

verifica queste proprietà:

$$|x| \geq 0$$

$$|-x| = |x|$$

$$|x|^2 = x^2$$

$$|x \cdot y| = |x| \cdot |y| \text{ e similmente } \Big(\forall\, y \neq 0\Big)\ |x/y| = |x|/|y|$$

$$|f(x)| = 0 \Leftrightarrow f(x) = 0$$

$$\Big(\forall\, a > 0\Big) \quad |f(x)| = a \Leftrightarrow f(x) = -a \vee f(x) = a$$

$$\Big(\forall\, a \geq 0\Big) \quad |x| > a \Leftrightarrow x < -a \vee x > a \ \ e\, analoga\, con\ \geq$$

$$\Big(\forall\, a > 0\Big) \quad |x| \leq a \ \Leftrightarrow\ -a \leq x \leq a$$

ovvero

$$\begin{cases} x \geq -a \\ x \leq a \end{cases}$$

e molte altre[69].

BOZZA - DRAFT

[69] Per il lettore interessato, ecco alcune altre formule sul valore assoluto:

$$\|x\| = |x|$$

$$|x| = x \cdot \operatorname{sgn}(x)$$

$$x = |x| \cdot \operatorname{sgn}(x)$$

$$|x^n| = |x|^n \qquad \forall\, x \in \mathbb{R}, \quad \forall\, n \in \mathbb{Z}$$

$$|x^\alpha| = |x|^\alpha \qquad \forall\, x \in \mathbb{R}^+, \ \forall\, \alpha \in \mathbb{R}$$

$$\Big(\forall\, a \geq 0\Big) \quad |f(x)| > a \Leftrightarrow f(x) < -a \vee f(x) > a \ \ e\, analoga\, con\ \geq$$

$$\Big(\forall\, a > 0\Big) \quad |f(x)| \leq a \ \Leftrightarrow \begin{cases} f(x) \geq -a \\ f(x) \leq a \end{cases}$$

$$|f(x)| = g(x) \ \Leftrightarrow \ \begin{cases} f(x) \geq 0 \\ f(x) = g(x) \end{cases} \vee \begin{cases} f(x) < 0 \\ -f(x) = g(x) \end{cases}$$

Il valore assoluto ricorre un'infinità di volte nelle Scienze Applicate. Nel prossimo paragrafo ne vedremo alcune, e per intanto osserviamo che $|t|$ rappresenta una distanza temporale dall'istante 0, sia dopo che prima di 0.

11.14 Errore assoluto, relativo e percentuale

Nelle successive 3 formule, si noti che bisogna conoscere il valore esatto di una grandezza considerata.

Si definiscono

$$\text{errore assoluto [rispetto l'esatto]} := |approx - esatto|$$

$$\text{errore relativo [rispetto l'esatto]} := \frac{|approx - esatto|}{esatto}$$

$$\text{errore percentuale [rispetto l'esatto]} :=$$

$$\text{errore relativo [rispetto l'esatto]}:=$$

$$\text{errore relativo [rispetto l'esatto] [in forma percentuale]} := \frac{|approx - esatto|}{esatto} \cdot 100\%$$

BOZZA - DRAFT

Quest'ultimo errore relativo, e quello della seconda formula, sono lo stesso numero, ma espresso in 2 modi diversi.

Nelle Scienze Applicate ci vorrà grande cautela nell'applicare queste 3 formule perché i *valori misurati* vanno intesi in generale come *approssimati*.

---

e i 3 = possono sostituirsi con tutti e 4 i segni di disuguaglianza.

$$|x + y| \leq |x| + |y|$$

e l'ultima si chiama *disuguaglianza triangolare.*

Esempio 1.$_\mu$ All'inizio i 46 cromosomi umani furono erroneamente conteggiati come 44 con errore (percentuale) dell'ordine del 4%:

$errore\ assoluto\,[rispetto\,l'esatto]:\ 2$

$errore\ relativo\,[rispetto\,l'esatto]:\ \frac{1}{23}$

$errore\ percentuale\,[rispetto\,l'esatto]:\approx 4.3\%$ (diciamo pure 4%).

Esempio 2$_\mu$ L'approssimazione 3.14 del valore di π ha errore (percentuale) dell'ordine dello 0.5 per mille:

$$\text{errore percentuale rispetto l'esatto} := \frac{|3.14-\pi|}{\pi}\cdot 100\% =$$

$$= \frac{|3.14-3.14159265...|}{3.14159265...}\cdot 100\% \approx$$

$$\approx \frac{|3.14-3.14159265|}{3.14159265}\cdot 100\% \approx 0.051\% \approx 0.05\%$$

(Meno dell'1 per 1000, anzi circa lo 0.5 per mille).

BOZZA - DRAFT

11.15 La sezione aurea

La sezione aurea è l'unico numero $x > 0$ che sia *medio proporzionale* fra 1 e $1+x$, cioè soluzione di

$$1 : x = x : (1+x) \qquad (*)$$

che ora calcoleremo.
Riscriviamo la proporzione $(*)$ in questa forma

$$\frac{1}{x} = \frac{x}{1+x} \qquad \Big/ \cdot x \cdot (1+x)$$

$$1 + x = x^2$$

$$x^2 - x - 1 = 0$$

$$x_{1,2} = \frac{1 \pm \sqrt{1+4}}{2}$$

ed escludendo la radice negativa, troviamo la soluzione che denotiamo con φ

$$\varphi = \frac{1+\sqrt{5}}{2} \approx 1.618$$

che è la *sezione aurea*.
Questo numero φ tende a ricorrere nelle Scienze Naturali.
È molto collegato alla successione di Fibonacci: si veda 6.12.1.

Curiosità.

Sottraendo 1 alla sezione aurea se ne ottiene il reciproco: $\varphi - 1 = \frac{1}{\varphi} \approx 0.618$.

BOZZA - DRAFT

Complementi

11.16 Complementi – Un sistema non lineare

Esempio oltre il livello che ci prefiggiamo.
Questo ***sistema di equazioni e disequazioni*** in 2 incognite

$$\begin{cases} x\,y = 0 & \leftarrow \textit{rappresenta 2 rette, } x = 0 \textit{ e } y = 0 \\ x^2 + y^2 - 2x - 2y + 1 \le 0 & \leftarrow \textit{rappresenta 1 cerchio, } C(1,1),\ r = 1 \end{cases}$$

equivale a

$$\begin{cases} x = 0 \vee y = 0 \\ (x-1)^2 + (y-1)^2 \le 1 \end{cases} \tag{19}$$

$$\Leftrightarrow \quad \begin{cases} x = 0 \\ (0-1)^2 + (y-1)^2 \le 1 \end{cases} \vee \begin{cases} y = 0 \\ (x-1)^2 + (0-1)^2 \le 1 \end{cases}$$

$$\Leftrightarrow$$

$$\begin{cases} x = 0 \\ (y-1)^2 \le 0 \end{cases} \vee \begin{cases} y = 0 \\ (x-1)^2 \le 0 \end{cases} \tag{20}$$

ha soluzione, come anche si vedeva subito graficamente,

$$\boxed{\Big(x = 0 \wedge y = 1\Big) \vee \Big(x = 1 \wedge y = 0\Big)}$$

ROZZA – DRAFT

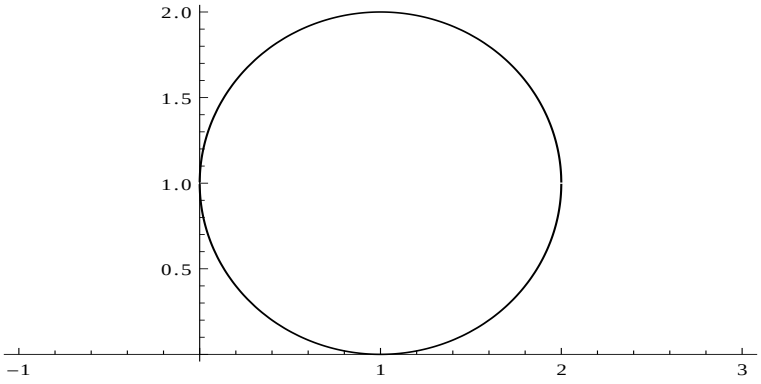


Figure 15: Il cerchio ("pieno") e l'unione delle 2 rette del sistema

11.17 ESERCIZI SULLA LEZIONE 11

11.17.1 Esercizio risolto a – Cerotto e pillola

ESERCIZIO$_{\mu 2023}$ $*$ Un cerotto e una pillola costano insieme €1,10 e la pillola costa 1 euro in più del cerotto. Quanto costa il cerotto?

SVOLGIMENTO
Viene usato – già nel quesito – lo standard della virgola decimale. (Che è fissato per legge in Italia per gli euro).

Ci sono vari modi di risolvere, e qua risolviamo in questo modo, con ovvio significato dei simboli:

$$cerotto + pillola = 1{,}10$$
$$pillola = cerotto + 1$$

e sostituendo la seconda equazione nella prima, si ha successivamente:

$$cerotto + (cerotto + 1) = 1{,}10$$

$$cerotto + cerotto + 1 = 1{,}10$$

$$2 \cdot cerotto + 1 = 1{,}10 \qquad / - 1$$

$$2 \cdot cerotto = 0{,}10 \qquad / : 2$$

$$cerotto = 0{,}05$$

cioè

$$\boxed{5 \text{ centesimi}}$$

ovvero

$$\boxed{€\,0{,}05}$$

Nota. Formulato con una palla e una mazza invece che con un cerotto e una pillola, questo è un problema classico, che si trova in *Pensieri lenti e veloci*, di Daniel Kahneman, vincitore del premio Nobel per l'Economia nel 2002 per il suo lavoro sui pregiudizi cognitivi. (A intuito molti rispondono erroneamente 10 centesimi).

BOZZA - DRAFT

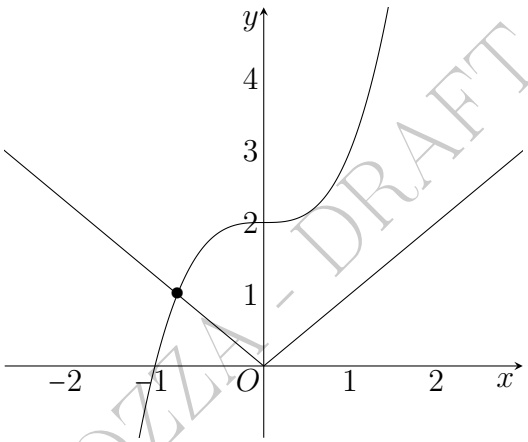
11.17.2 Esercizio risolto b – Risoluzione grafica

ESERCIZIO$_{\mu 2024}$ $*$ Risolvere per via grafica il sistema di equazioni

$$\begin{cases} y = |x| \\ y = x^3 + 2 \end{cases}$$

(dando la soluzione nella forma $x = ...$, $y = ...$).

SVOLGIMENTO



BOZZA - DRAFT

Già bastano questi pochi (valori delle coordinate di) punti dei 2 grafici

x	$\lvert x\rvert$
-2	2
-1	1
0	0
1	1
2	2

x	$x^3 + 2$
-2	-6
-1	1
0	2
1	3
2	10

e disegnando (vedi figura) sullo stesso piano cartesiano grafici (approssimativi), vediamo subito che c'è un punto di intersezione $(-1, 1)$, e non esistono altri per

le proprietà elementari delle funzioni $|x|$ e x^3, e allora la soluzione del sistema di equazioni è data dalle coordinate x e y dell'unico punto di intersezione dei grafici:

$$\boxed{x = -1,\ y = 1}$$

BOZZA - DRAFT

12 Matematica col Computer – Lezione A

12.1 Introduzione

È del tutto impossibile prevedere con quali tipi di software ed eventualmente linguaggi di programmazione avrà a che fare il Laureato in una sua futura professione.

In questo testo elementare praticamente in ogni Lezione viene usato almeno un po' WolframAlpha, che è semplicemente ottimo, ma ci limita a programmi di una sola istruzione.

In queste Lezioni A, B e C specificamente dedicate alla Matematica al Computer, una varietà di software e linguaggi verrà considerata.

Nelle scienze applicate il software più usato per trattare dati è Excel, che fa parte del pacchetto più ampio Microsoft Office, che ha il suo costo.
Un'alternativa gratuita a Microsoft Office e LibreOffice che ha al suo interno il modulo Calc che funziona in modo simile a Excel.

BOZZA - DRAFT

In generale, se si avrà un problema da risolvere col computer
– di tipo numerico, come medie o risoluzione di equazioni, oppure
– di visualizzazione grafica di dati
(1) prima opzione dovrebbe essere WolframAlpha.
Se non si potrà
– perché una riga non ci basta, oppure
– perchè si vuole una cosa che funzioni offline
(2) seconda opzione dovrebbe essere Excel o LibreOffice.
Se non si potrà perché Excel e LibreOffice non risolvono bene il problema, se non con molto complessi artifici,
(3) terza opzione dovrebbe essere realizzare un programma in un linguaggio di programmazione come
Java
Pascal

Fortran

Ormai ottant'anni di storia mostrano che se si vuole fare un software stabile, che continui a girare per lungo tempo sui nuovi computer che verranno – salvo motivate scelte aziendali che pure potranno esistere in casi specifici mentre qua facciamo un discorso generale – è opportuno stare alla larga da linguaggi ultramoderni di grido, che magari verranno dimenticati in breve tempo, e per i quali non verranno più prodotti compilatori funzionanti sui computer delle successive generazioni. Lo scrivente ricorda quanto a suo tempo il Lisp sembrasse il linguaggio del futuro con cui i computer avrebbero acquisito l'intelligenza artificiale, che infatti poi è stata realizzata, ma il Lisp chi se lo ricorda?

BOZZA - DRAFT

Il Java è molto valido ma è adatto per software enormi realizzate da team di programmatori, e certamente vi si potranno fare software di un milione di linee. Per un piccolo programmino non è l'ideale a meno che uno già non lo usi abitualmente. Studiare il Java per fare un programmino non ha senso, è il tipico caso dello "sparare alle mosche con un cannone".

Il terzo linguaggio di programmazione contenuto nella soprastante lista, è il più stabile nel tempo, quello che più tende all'eternità, per così dire, avendo avuto il suo primo standard nel 1954, e molti aggiornamenti successivi compatibili. Possiamo essere certi che in qualunque azienda lavorerete, potrete usarlo, e salvo catastrofi planetarie continuerà a girare nei decenni seguenti, mentre la grande maggioranza dei programmi informatici scritti in passato nei vari linguaggi sono ormai di fatto inutilizzabili: sui nuovi computer non girano i vecchi compilatori e nuovi non ne vengono più fatti per linguaggi in disuso. Questo linguaggio è il corrispondente del latino nella tassonomica: dopo più di tre secoli i nomi di nuovi animali e piante vengono ancora assegnati in latino o latinizzati, e non si è affatto passati al francese nel XIX secolo nè all'inglese nel XX,

e possiamo stare certi che non si passerà al russo o al marziano almeno per tutto il XXI, a meno che non vengano marziani prepotentissimi: fu fissato il latino e latino rimane nei secoli.
In questo linguaggio sono state scritte un'infinità di routine che risolvono i più vari problemi matematici, e nessuno osa tradurle nell'ultimo linguaggio di grido, e nemmeno un byte viene modificato. Un'infinità di linee di codice a tutt'oggi gira nei software di climatologia, checchè ne dicano i suoi detrattori propagandisti dell'ultimo linguaggio di grido, che seppelliremo insieme al Lisp.

BOZZA - DRAFT

12.2 Programma informatico sulla parte intera

Abbiamo valori positivi a_1,...,a_{40}, per esempio le età dei partecipanti a un trial clinico, e vogliamo le loro parti intere b_1,...,b_{40}, nel senso di $\lfloor\,.\,\rfloor$, funzione che in questo linguaggio informatico si chiama `INT()` e com'è ovvio `DO`...`CONTINUE` esegue un'iterazione.
In più, questo programma visualizza i valori trovati, seguiti ciascuno dal valore originale.
Quale di queste è la linea mancante nel programma?

a) `A(N)=INT(B(N))`

b) `A(N)=INT(A(N))`

c) `B(N)=INT(B(N))`

d) `B(N)=INT(N(N))`

e) `B(N)=INT(A(N))`

f) `B(N)=INT(N(A))`

(Con ovvia simbologia: per esempio a_1 si chiama `A(1)`).

```
      PROGRAM FLOOR
C     PONE NEL VETTORE B LE PARTI INTERE DEI VALORI DEL VETTORE A
      DIMENSION A(40), B(40)
      DO 1, N=1,40
        ???????  linea mancate da trovare fra quelle sopra ???????
        WRITE(*,*)B(N),' ',A(N)
1     CONTINUE

      DATA A/
     * 86.1226, 70.4775, 77.8931, 72.3801, 75.9830, 65.8364, 64.1100,
     * 28.4700, 56.7235, 47.3091, 42.0964, 34.1011, 87.1678, 57.7191,
     * 51.5723, 66.4351, 74.8770, 69.9438, 65.0818, 91.4392, 38.2574,
     * 55.4153, 25.2473, 37.8699, 81.9343, 69.6649, 68.6506, 80.1273,
     * 70.3054, 75.5321, 99.4785, 41.0719, 39.9702, 98.6395, 43.0819,
     * 21.7130, 29.7820, 42.9840, 69.4161, 49.9248/

      END
```

BOZZA - DRAFT

(È evidente che con 40 valori faremmo prima a mano che scrivendo il programma, ma con 40mila no).

12.3 Un'equazione di facile espressione

La trentina di soluzioni (approssimate, ovvio) dell'equazione

$$\frac{x^2}{20} - \sin\left(\frac{1}{x + \mathrm{e}^{-4.5}}\right) = 0 \qquad 0 \le x \le 10$$

possono essere trovate online con WolframAlpha: LINK->

12.4 Un'equazione di difficile espressione

```
 ************************************************************
 Ricerca degli zeri di F(X) fra A e B con A e B da inserire.
 ************************************************************

 Estremo A dell'intervallo, con A>1 e A<21:
5
 Estremo B, con B>A e B<21:
12

 ATTENZIONE

 Di 2 eventuali zeri più vicini di    9.99999975E-06
 potrebbe venirne trovato solo 1, essendo quasi indistinguibili.

 Ecco nell'intervallo di estremi
   5.00000000
   12.0000000
 gli zeri di F(X), salvo l'avvertenza sopradetta:

   5.42253017
   5.57747984
   6.17480993
   6.82520008
   7.21897984
   7.78102970
   8.50000000
   8.50000954
   9.07131958
   9.92868996
   10.1958504
   10.8041697
   11.3476791
   11.6523304
```

Questo sopra è l'output del programma che segue, in cui da input sono stati immessi i valori 5 e 12 per cercare nell'intervallo $[5, 12]$ gli zeri della funzione

$$|x - \lfloor x \rfloor - 0.5|^{q_{\lfloor x \rfloor}} - 0.2 \qquad 1 < x < 21$$

che dipende da 20 costanti numeriche $q_1 ... q_{20}$ scritte in DATA.

(Qualche sospetto destano le soluzioni 8.50000000 e 8.50000954).

```
      PROGRAM ZERIF
C     Trova gli zeri di una funzione F(X), purchè non più vicini di D, in un
C     intervallo di estremi A e B modificabili da input, con A fra 1 e 21 e B fra A
C     e 21. La F(X) e il parametro D sono modificabili in questo file sorgente.
      DIMENSION Q(20)
      F(X)=(ABS(X-INT(X)-0.5))**Q(INT(X))-0.2
      D=1E-5
      WRITE(*,*)
      WRITE(*,*)
      WRITE(*,*)
     *  '************************************************************'
      WRITE(*,*)
      * 'Ricerca degli zeri di F(X) fra A e B con A e B da inserire.'
      WRITE(*,*)
     *  '************************************************************'
      WRITE(*,*)
1     CONTINUE
      WRITE(*,*)'Estremo A dell"intervallo, con A>1 e A<21:'
      READ(*,*)A
      IF(A.LE.1.OR.A.GE.21)GOTO 1
2     CONTINUE
      WRITE(*,*)'Estremo B, con B>A e B< 21:'
      READ(*,*)B
      IF(B.LE.A.OR.B.GE.21)GOTO 2
      WRITE(*,*)
      WRITE(*,*)'ATTENZIONE'
      WRITE(*,*)
      WRITE(*,*)
     *  'Di 2 eventuali zeri più vicini di ',D
      WRITE(*,*)
     *  'potrebbe venirne trovato solo 1, essendo quasi indistinguibili.'
      WRITE(*,*)
      WRITE(*,*)'Ecco nell"intervallo di estremi'
      WRITE(*,*)A
      WRITE(*,*)B
      WRITE(*,*)'gli zeri di F(X), salvo l"avvertenza sopradetta:'
      WRITE(*,*)
      X=A
      DO 3, N=1, INT((B-A)/D))
      Y=F(X)
      X=A+N*D
      IF(F(X)*Y.LE.0)WRITE(*,*)X
3     CONTINUE
      DATA Q / 1.61642, 1.31813, 1.54686, 0.286281, 0.629239, 1.43274,
     *  1.26795, 0.0965641, 1.90009, 1.35224, 0.855288, 1.51489, 1.37894,
     *  1.45307, 1.88376, 1.0769, 0.497219, 0.900547, 0.514088, 1.62761 /
      END
```

BOZZA - DRAFT

III – Funzioni elementari, con equazioni e disequazioni

BOZZA - DRAFT

13 Funzioni e dis/equazioni ir/razionali

13.1 Polinomi e funzioni razionali intere

Un polinomio è una *scrittura* come
$x^2 - 4$ oppure
$2\,x^7 - \frac{3}{2}\,x^5 + \sqrt{3}\,x^2 - x + 2$ e in generale

$$a_0 + a_1\,x + a_2\,x^2 + \ldots + a_n\,x^n$$

e n è il *grado* se $a_n \neq 0$.
Il polinomio definisce una funzione *razionale intera* o *polinomiale*

$$f(x) := a_0 + a_1\ + a_2\,x^2 + \ldots + a_n\,x^n$$

Naturalmente f e x solo solo 2 degli infiniti nomi possibili, possono anche trovarsi anche g, y, $temperatura$, $tempo$...
Per esempio
$g(t) := t^4 + t + 4$
$temperatura(tempo) := 2 \cdot tempo^3 + 100$
$tempo(temperatura) := \frac{1}{2} \cdot temperatura - 30$
che hanno gradi 4, 3 e 1 rispettivamente.

BOZZA - DRAFT

Esistono similmente polinomi in più variabili, come $x^2 + x^8\,y^2 - 1$, del cui grado (che comunque in questo esempio è 10) non ci occuperemo.

13.2 Funzioni e dis/equazioni razionali intere

Limitandosi ad 1 variabile, sia essa x, le
equazioni razionali intere
e le
disequazioni razionali intere
sono predicati con uguaglianza o disuguaglianza fra polinomi, che dopo opportune riduzioni che popolarmente si dicono "portare le x a sinistra", assumono una di queste forme, essendo $P(x)$ un polinomio:

• disequazioni:
◦ $P(x) > 0$
◦ $P(x) < 0$ idem
◦ $P(x) \geq 0$ idem
◦ $P(x) \leq 0$ idem
— e ci limiteremo solo a disequazioni di
  ◇ I grado, già trattate, e
  ◇ II grado, già trattate.

• equazione:
◦ $P(x) = 0$
— e ci limiteremo a:
  ◇ equazioni di I grado, già trattate, ed
  ◇ equazioni di II grado, già trattate, e
  ◇ x^n·(polinomio di I o II grado)= 0 **unico caso nuovo** (*)

e quest'ultimo, molto semplicemente, si risolve aggiungendo lo 0 alle radici del polinomio nella parentesi. Faremo un esempio pratico, poco sotto.

BOZZA–DRAFT

Naturalmente si potrebbero considerare casi più complessi: si vedano i Complementi a questa Lezione.

Per la generica equazione di III e IV grado esistono complicate formule, trovate da matematici italiani nel XVI secolo, che non tratteremo. Ci accontentiamo di x^n·(polinomio di I o II grado)= 0, come sopra detto, che può avere anche grado maggiore di 4, ma è molto particolare.

> **Esempio**$_\mu$ La Formula di Livi per il peso ideale è
> $peso\,ideale = (2.37 \times altezza\,in\,metri)^3$
> La Formula di Keys per il peso ideale per le donne è $pesoidealedonne = (altezza\,in\,metri)^2 \times 20.6$
> Per quali altezze (realistiche) danno lo stesso peso, e quale?
>
> Detta h l'altezza in metri – unità di misura che metteremo nella soluzione ma ometteremo nei calcoli – abbiamo l'equazione di 3^

grado

$$(2.37\,h)^3 = 20.6\,h^2$$

e successivamente

$$2.37^3 \cdot h^3 - 20.6\,h^2 = 0$$

che è equazione *cubica*, ovvero *razionale intera di terzo grado*, ma del tipo (*)

$$h^2 \cdot (2.37^3 \cdot h - 20.6) = 0$$

$$h = 0\,(non\,realistica) \quad 13.312053\,h = 20.6$$

e allora

$$h = \frac{20.6}{13.312053}$$

$$p = \left(\frac{20.6}{13.312053}\right)^2 \cdot 20.6$$

$$\boxed{\begin{array}{c} h \approx 1.55\,m \\ p \approx 49.3\,kg \end{array}}$$

BOZZA - DRAFT

13.3 Funzioni e dis/equazioni razionali fratte

Ovviamente l'*equazione razionale fratta*

$$\begin{matrix} polinomio \to \\ polinomio \to \end{matrix} \quad \frac{N(x)}{D(x)} = 0$$

ha la stessa soluzione dell'equazione razionale intera

$N(x) = 0$

tolte le eventuali radici del denominatore.

La *disequazione razionale fratta*

$$\begin{matrix} polinomio \to \\ polinomio \to \end{matrix} \quad \frac{N(x)}{D(x)} > 0$$

ha la stessa soluzione di $N(x) > 0$ se sappiamo che il denominatore è positivo per ogni x, per esempio se il denominatore è $x^2 + 1$.
Se invece non sappiamo se il denominatore è sempre positivo, si ha un caso più complicato, di cui non ci occuperemo (e sicuramente rarissimo in Farmacia).

13.4 Una funzione razionale fratta della Farmacia

Mostriamo una funzione razionale fratta di qualche interesse farmaceutico, almeno teorico, in quanto si è mostrato in un articolo scientifico (su un campione di 363 bambini) che, almeno per l'età pediatrica, approssima abbastanza bene i risultati della ben più nota formula di Mosteller per l'area della superficie corporea (dato di interesse per esempio nel dosaggio dei chemioterapici), che è più complessa da calcolare coinvolgendo sia peso che altezza (invece che solo peso) e una radice quadrata (invece che solo "le 4 operazioni"):

$$AreaSuperficieCorporea = BSA = \frac{4w + 7}{90 + w}$$

essendo w il peso in chilogrammi: https://pubmed.ncbi.nlm.nih.gov/19430073/: "It can safely replace Mosteller formula and dispense the need for time-consuming calculations."

BOZZA - DRAFT

13.5 Le funzioni radice quadrata e cubica

Abbiamo visto nel ripasso di algebra cosa sono la radice quadrata di un numero $x \geq 0$ e la radice cubica di un numero x qualunque.

Al variare di x si ottengono le funzioni radice quadrata e radice cubica. Sono entrambe crescenti. Per $x \geq 0$ queste 2 funzioni hanno grafici qualitativamente simili che si intersecano in $(0, 0)$ e $(1, 1)$ ma per x sufficientemente grande [70] la radice cubica cresce più lentamente della radice quadrata. (Valgono entrambe $+\infty$ in $+\infty$ nel senso del limite che vedremo, ma la radice cubica ci arriva, per così dire, più lentamente).

La radice cubica è dispari e allora ha grafico simmetrico rispetto all'origine.

[70] In effetti già per $x > 0.087791...$

13.6 Funzioni e dis/equazioni irrazionali

• Le radici di indice dispari (cubiche, quinte...) sono crescenti su $\mathbb{R}$ e allora l'equazione o disequazione

$$\sqrt[dispari]{f(x)} \ \ {}_{qualunque\ segno\ di\ dis/uguaglianza} \ \ g(x)$$

si può affrontare semplicemente elevando ambo i membri all'indice (dispari) della radice (che così scompare). Allora si hanno queste 5 semplici equivalenze, che si ricavano subito appunto elevando a potenza conservando l'uguaglianza o l'ordinamento, e allora in definitiva non c'è niente da veramente "imparare a memoria":

$$\begin{array}{ccccc} \sqrt[3]{f(x)} & = & g(x) \Leftrightarrow f(x) & = & g^3(x) \\ & > & & > & \\ & \geq & & \geq & \text{in tutte} \\ & < & & < & \text{l'ordinamento} \\ & \leq & & \leq & \text{si conserva} \end{array}$$

e il 3 della radice può sostituirsi con ogni indice dispari.

BOZZA - DRAFT

• Il caso delle radici di indice pari (radici quadrate, quarte...) è meno semplice perché sono definite solo sui numeri non negativi.

Vale questa formula di riduzione

$$\sqrt{f(x)} = a \geq 0 \quad \Leftrightarrow \quad f(x) = a^2$$

e altre che si trovano nei Complementi di questa Lezione.

13.7 Una funzione algebrica con grafico sigmoide

Definiamo funzione sigmoide o sigmoidea, in via semplificata, una funzione limitata e cioè

$$a < f(x) < b$$

la quale prima ha la concavità verso l'alto e poi verso il basso. Non è assolutamente una definizione standard.

Il suo grafico lo chiameremo curva sigmoide o sigmoidea.

Il prototipo è la funzione arcotangente; si veda il grafico di
`arctan(x)`
su WolframAlpha. In base alla soprastante definizione, non $\sqrt[3]{x}$.

Le funzioni sigmoidi sono semplici modelli per innumerevoli fenomeni delle Scienze Applicate, e in particolare per il numero cumulativo di morti, o di casi, eventualmente per milione, di un'ondata epidemica, nel mondo o in uno stato o regione.

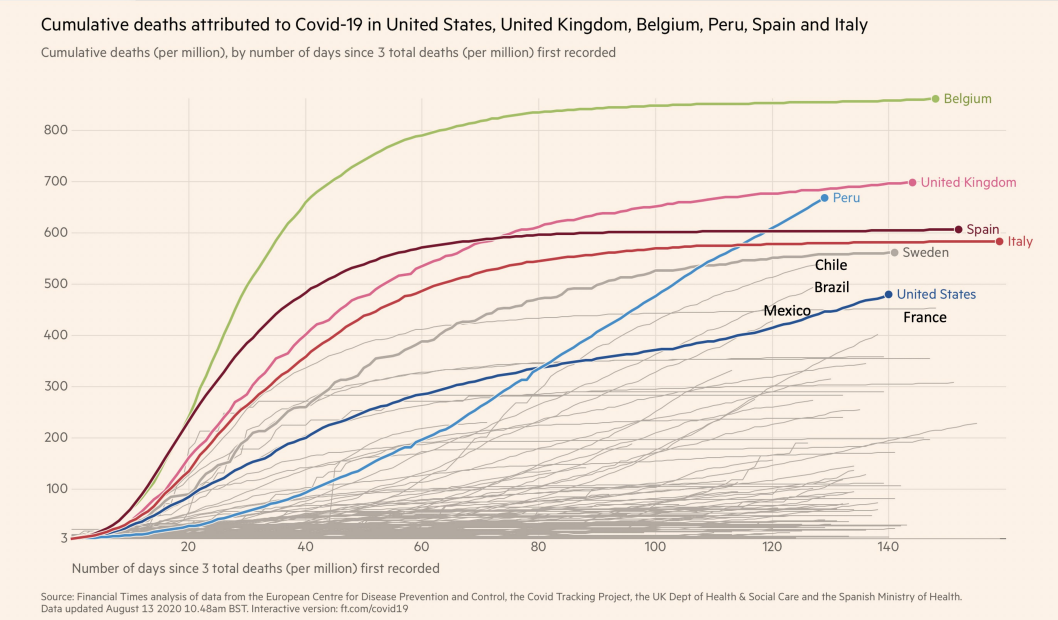


Figure 16: Numero cumulativo di decessi per milione di abitante della pandemia del 2020 verso la metà di agosto 2020, in alcuni stati. (Lo 0 temporale è diverso stato per stato, cioè, per esempio, il giorno 100 dell'epidemia non avviene lo stesso giorno nei vari stati, perchè dipende da quando l'epidemia è iniziata in ogni singolo stato).

Adesso vediamo una funzione sigmoide definita solo con le 4 operazioni e la radice quadrata,

$$f(x) := \frac{x}{\sqrt{x^2+1}}$$

che cresce da -1 a 1, e quest'altra sua correlata

$$g(x) := \frac{1}{2} + \frac{1}{2}\frac{x}{\sqrt{x^2+1}}$$

che cresce da 0 a 1. Essa potrebbe per esempio rappresentare, al variare del tempo x con una qualche unità di misura temporale, il numero cumulativo di morti di covid per mille abitanti, in una regione geografica, in una singola ondata pandemica. (Con ondate successive, la curva si complica).
(Il Belgio nell'agosto 2020 aveva circa 0.9 morti per mille abitanti, in crescita, ormai lieve, come si vede nella figura).
La ricerca scientifica ha comunque trovato molto migliori funzioni, e più complesse, per modellizzare le epidemie.

BOZZA - DRAFT

Complementi

13.8 Complementi – Un esempio non banale

$2\,x^4(-x^2-2x+8)(2-x)(1-x+x^2)=0$ o, disequazione, >0,

oppure altra disequazione con $\geq$ oppure $<$ oppure $\leq$.
Per risolvere *fattorizziamo* il polinomio in fattori di primo grado e/o di secondo grado *irriducibili* ovvero con discriminante negativo. E poi si risolve con uno schema di prodotto dei segni.
Ecco qualche dettaglio per il lettore interessato. Il monomio $2\,x^4$ si fattorizza in $2\,x\cdot x\cdot x\cdot x$ ma in effetti i fattori che sono monomi x^n conviene lasciarli stare come sono. Il polinomio di secondo grado $-x^2-2x+8$, con discriminante positivo, abbiamo visto che è $-1\cdot(x-2)(x+4)$. Il fattore di primo grado $2-x$ è già a posto. E $1-x+x^2$ è irriducibile avendo discriminante negativo, e allora non ha zeri. Si ha allora la fattorizzazione

$$2\,x^4(-1)(x-2)(x+4)(2-x)(1-x+x^2)\,.$$

L'equazione con $=0$ equivale, considerato che $x^4=0\Leftrightarrow x=0$,

$$x=0\ \vee\ x-2=0\ \vee\ x+4=0\ \vee\ 2-x=0$$

che dà $x\in\{-4,0,2\}$. La disequazione si fa con lo schema di prodotto dei segni visto in precedenza trovandosi (per ≤ 0) la soluzione $x\leq -4\vee x=0\vee x=2$. Per >0 si troverebbe $-4<x<0\vee 0<x<2\vee x>2$.

BOZZA - DRAFT

13.9 Complementi – Un esempio difficile

L'equazione e le 4 disequazioni

$$2\,x^9-2\,x^8-22\,x^7+56\,x^6-56\,x^5+32\,x^4\quad =0,\ >0,\ \geq 0,\ <0,\leq 0$$

si risolvono come in 13.8 una volta che si riconoscesse che questo è proprio il polinomio di prima, ma ora è molto lontano dalla fattorizzazione. Il problema di fattorizzare un polinomio può essere

facile, come è in questo caso, o difficile, o impossibile.
Prima di tutto *raccogliamo* il fattore x^4, subito visto:

$$x^4\,(2\,x^5 - 2\,x^4 - 22\,x^3 + 56\,x^2 - 56\,x + 32)$$

e ci resta da fattorizzare il polinomio $P_5(x)$ di 5^ grado. In questo caso si può fare abbastanza facilmente ma in questa trattazione ci limiteremo solo ai casi in cui, arrivati a questo punto, si abbia un polinomio di grado al più 2 (invece qua è 5) e allora si procede come prima.
Per il lettore interessato: prima si raccoglie 2

$$2\,(x^5 - 2\,x^4 - 11\,x^3 + 28\,x^2 - 28\,x + 16)$$

in modo da avere un polinomio *monico*, cioè con *coefficiente principale* 1, poi si procede con la Regola di Ruffini, se si può. Cercheremo qua *solo* fattori $x-m$ con m un divisore intero del termine costante: ±1, ±2, ±4, ±8, ±16. Allora basta cercare un numero m fra quelli, il quale annulli il polinomio. È $P_5(2) = 0$ e allora dividiamo per $x - 2$ con la Regola di Ruffini:

i coefficienti →	+1	−1	−11	+28	−28	+16
la radice → +2	↓	+2	+2	−18	+20	−16
	+1	+1	−9	+10	−8	0 ← sempre così

che dà i coefficienti del quoziente, e allora la fattorizzazione

$$(x - 2)(x^4 + x^3 - 9x^2 + 10x - 8)$$

e riprendendo il monomio x^4 e il fattore 2

$$2\,x^4(x - 2)(x^4 + x^3 - 9x^2 + 10x - 8)\,.$$

Si continua cercando un fattore $x - m'$ con m' fra i divisori interi di −8, e si trova che 2 annulla il polinomio di quarto grado $P_4(x)$, e allora lo si divide per $x - 2$ con la Regola di Ruffini trovando

$$2\,x^4(x - 2)(x - 2)(x^3 + 3x^2 - 3x + 4)$$

BOZZA - DRAFT

poi cercando fra i divisori interi di 4 si trova l'annullamento di $P_3(x)$ in -4 da cui con la Regola di Ruffini applicata a $P_3(x)$

$$2\,x^4(x-2)(x-2)(x+4)(x^2-x+1)$$

che è come prima (un fattore è opposto ma congloba il (-1)).

13.10 Complementi – Dis/equazioni con radice quadrata

Per le equazioni valgono queste formule di riduzione

$$\sqrt{f(x)} = a \geq 0 \Leftrightarrow f(x) = a^2$$

$$\sqrt{f(x)} = g(x) \Leftrightarrow \begin{cases} g(x) \geq 0 \\ \\ f(x) = g^2(x) \end{cases}$$

Per le disequazioni valgono queste formule di riduzione

$$\sqrt{f(x)} > a \Leftrightarrow f(x) > a^2 \; se\, a > 0, similmente \geq \qquad (21)$$

$$\sqrt{f(x)} < a \Leftrightarrow 0 \leq f(x) < a^2 \; se\, a > 0, similmente \leq \qquad (22)$$

e altre formule su cui non insisteremo. Per il lettore interessato ecco altre formule limitatamente alla radice quadrata.
Prima di tutto il caso che diremo **"radice minore"**

$$\sqrt{f(x)} < g(x) \Leftrightarrow \begin{cases} f(x) \geq 0 \\ g(x) > 0 \\ f(x) < g^2(x) \end{cases}$$

Poi il caso **"radice minore o uguale"**, similissimo a quello "radice minore":

$$\sqrt{f(x)} \leq g(x) \Leftrightarrow \begin{cases} f(x) \geq 0 \\ g(x) \geq 0 \\ f(x) \leq g^2(x) \end{cases}$$

e poi 2 casi (similissimi) **"radice maggiore"** e **"radice maggiore o uguale"**:

$$\sqrt{f(x)} > g(x) \Leftrightarrow \begin{cases} g(x) \geq 0 \\ f(x) > g^2(x) \end{cases} \vee \begin{cases} g(x) < 0 \\ f(x) \geq 0 \end{cases}$$

BOZZA - DRAFT

$$\sqrt{f(x)} \geq g(x) \iff \begin{cases} g(x) \geq 0 \\ f(x) \geq g^2(x) \end{cases} \vee \begin{cases} g(x) < 0 \\ f(x) \geq 0 \end{cases}$$

Esercizio 1_{μ} $\sqrt{x^2-9} \geq x-1$.

Esercizio 2_{μ} $x > \pi + \sqrt{x^2-10}$.

Esempio$_{\mu}$ Useremo la Formula di Mosteller per l'area della superficie corporea, in questa versione:

$$area \approx \frac{\sqrt{altezza_m \times peso_{kg}}}{6}$$

Considerato un soggetto per un certo tempo, in cui possiamo ritenere costante la sua altezza, sia essa h, può ben variare il suo peso, sia esso x.
Per quel soggetto abbiamo allora una funzione

$$area(x) = \frac{1}{6}\sqrt{h \cdot x}$$

essendo h l'altezza (in metri, costante),

x il peso (in chilogrammi),

$area(x)$ l'area della superficie corporea (in metri quadrati).
Per quali pesi la superficie corporea sarebbe minore di 2 metri quadrati in questo modello?

BOZZA - DRAFT

Abbiamo la disequazione

$$\frac{1}{6}\sqrt{h \cdot x} < 2 \qquad \Big/ \cdot 6 > 0 \text{ (conserva l'ordinamento)}$$

$$\sqrt{h \cdot x} < 12$$

che risolviamo come in (22)

$$\Leftrightarrow \begin{cases} h \cdot x \geq 0 & \Big/ : h > 0 \text{ (conserva l'ordinamento)} \\ h \cdot x < 12^2 & \Big/ : h > 0 \text{ (conserva l'ordinamento)} \end{cases}$$

$$\Leftrightarrow \begin{cases} x \geq 0 & \text{(ovvio, è un'altezza)} \\ x < \frac{144}{h} \end{cases}$$

$$\big(0 \leq\big)\ x < \frac{144}{h}$$

Per esempio con un'altezza di 1.71 m, $x < 84.2$. (Sotto gli 84 kg, circa).
Per esempio con un'altezza di 1.82 m, $x < 79.1$. (Sotto i 79 kg, circa).

ESERCIZIO$_{\mu 2018}$
Si risolva la seguente disequazione:

$$x > \mathrm{e} + \sqrt{x^2 - 8}$$

SVOLGIMENTO
Essendoci una radice quadrata, cercheremo di ricondurci al caso “radice maggiore" oppure al caso “radice minore", e in effetti in questo caso

$$x > \mathrm{e} + \sqrt{x^2 - 8} \quad \Big/ + (-\mathrm{e})$$

$$x - \mathrm{e} > \sqrt{x^2 - 8}$$
$$\sqrt{x^2 - 8} < x - \mathrm{e}$$

proprio al caso “radice minore", e allora la disequazione equivale al sistema

$$\begin{cases} x^2 - 8 \geq 0 \\ x - \mathrm{e} > 0 \\ x^2 - 8 < (x - \mathrm{e})^2 \end{cases}$$

ovvero

$$\begin{cases} x \leq -\sqrt{8} \vee x \geq \sqrt{8} \\ x > \mathrm{e} \\ x^2 - 8 < x^2 - 2\,\mathrm{e}\,x + \mathrm{e}^2 \end{cases}$$

ovvero

$$\begin{cases} x \leq -2\sqrt{2} \vee x \geq 2\sqrt{2} \approx 2.82 \\ x > \mathrm{e} \approx 2.718 \text{ che sappiamo a memoria} \\ 2\,\mathrm{e}\,x < \mathrm{e}^2 + 8 \quad \Leftrightarrow \quad x < \frac{\mathrm{e}^2+8}{2\,\mathrm{e}} \approx 2.831 \text{ con la calcolatrice dal soprastante} \end{cases}$$

e mettendo in ordine crescente i capisaldi (ovvero valori delimitanti) trovati

.....$-2\sqrt{2}$.....e.....$2\sqrt{2}$.....$\frac{\mathrm{e}^2+8}{2\,\mathrm{e}}$.....
.....|.....|.....|.....|..... schema d’intersezione
~~~~~ – – – – – – – – ~~~~~~~~~~
– – – – – – – – ~~~~~~~~~~~~~~~~ ∩
~~~~~~~~~~~~~~~~~~~~~ – – – –

otteniamo il soprastante schema d’intersezione che ci dà la soluzione

$$\boxed{2\sqrt{2} \leq x < \frac{\mathrm{e}^2+8}{2\,\mathrm{e}}}$$

BOZZA - DRAFT

14 Esponenziali e logaritmi – I parte

14.1 Vecchie funzioni potenza e nuove esponenziali

Già sappiamo il significato della

$$\text{potenza} \quad 2^3 \quad \text{e in generale } b^a$$

di base $b > 0$ ed esponente $a \in \mathbb{R}$

(Se $a > 0$ allora esiste anche per $b = 0$ e vale 0).

Notoriamente, se "liberiamo" la base abbiamo le **funzioni potenza**

$$x^a \qquad \text{per esempio } x^2,\ x^3,\ x^{10},\ x^{\frac{1}{2}}$$

Ma se "liberiamo" l'esponente abbiamo le **funzioni esponenziali**

$$b^x \qquad \text{per esempio } 2^x,\ 3^x, 10^x, \left(\frac{1}{2}\right)^x$$

di base 2, di base 3, eccetera;
e questa funzione è caratterizzata da

- *"al crescere di 1 della x la b^x si moltiplica per b"*
- *"al diminuire di 1 della x la b^x si divide per b"*

per esempio al crescere di 1 della x la 2^x raddoppia, e la 10^x decuplica.

Le funzioni esponenziali modellizzano un'infinità di processi delle Scienze Applicate, con crescenza sempre più rapida se la base è $b > 1$, e decrescenza verso 0 se $0 < b < 1$.
Si vedano online su WolframAlpha i grafici per $x \geq 0$ delle funzioni esponenziali di basi $\frac{1}{3}$, $\frac{1}{2}$, 2, 3, che passano tutti per $(0, 1)$ (illustrando la $c^0 = 1$), e le funzioni potenza con esponenti quegli stessi numeri, che passano tutti per $(1, 1)$ (illustrando la $1^c = 1$).

BOZZA - DRAFT

14.2 La sostanziale unicità degli esponenziali

In buona sostanza, esiste "solo una" funzione esponenziale, e tutte le altre si ottengono riscalando l'argomento. Per esempio la funzione 3^t, essendo t un tempo corrispondente ai mesi (senza unità di misura: numero puro), dà esattamente gli stessi valori della funzione 9^t con t misurato in bimestri (senza unità di misura: numero puro): la prima al mese 4 vale 81 come la seconda al bimestre 2, che è lo stesso tempo nella realtà sensibile.

14.3 Le 4 basi di maggior interesse

Le **basi** più significative sono queste 4, e vediamo alcuni esempi:

- 2: la 2^n dà il num. di sottoinsiemi di un insieme di n elementi;
- φ: la $\frac{\varphi^n}{\sqrt{5}}$ approssima per intero $n \gg 0$ la successione di Fibonacci,

$$\text{e } \varphi \text{ è la } \textit{sezione aurea, } \varphi := \frac{1+\sqrt{5}}{2} \approx 1.618; \tag{23}$$

- 10: per la notazione scientifica dei numeri, $x \cdot 10^n$, con n intero;
- e: e^x è la **funzione esponenziale** (senza dire "di base e") ed

  e è il numero di Nepero o numero di Eulero, in inglese Euler's number:

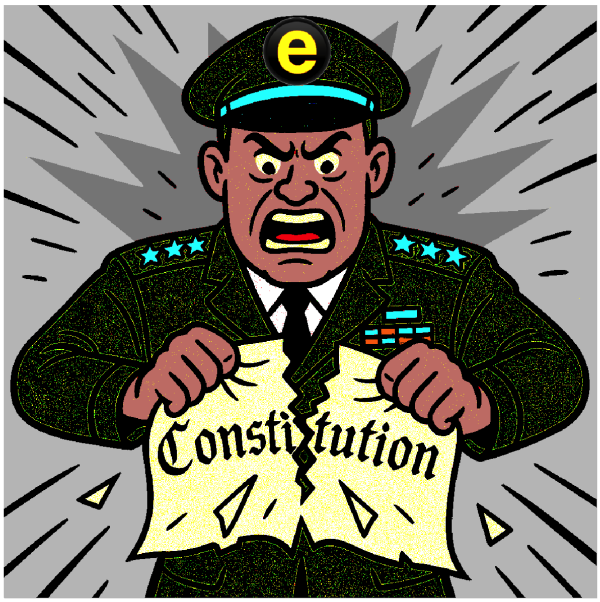
$$\boxed{\text{e} \approx \mathbf{2.718}} \tag{24}$$

Per ricordare queste cifre, sono stati realizzati molti mnemonici, frasi[71] in cui le lunghezze delle parole corrispondono alle cifre (delle approssimazioni) di e.

BOZZA - DRAFT

[71] **Mnemonici per le cifre di e.**
Associando ordinatamente ad ogni parola di certe frasi il loro numero di lettere, si ottengono le prime cifre di e. Dello scrivente, in inglese, per esempio:
Of
Eulerus
e
constant.

In figura (in inglese) uno mnemonico (dello scrivente) per ricordarne le cifre (contando le lettere delle parole).



e ≈ 2.718 HE BECOMES A DICTATOR

```
2 ← HE
7 ← BECOMES
1 ← A
8 ← DICTATOR
```

Dello scrivente, in italiano:

> *"La loquela è vincente* ← 2, 7, 1, 8
> *ma talvolta è migliore il silenzio,* ← 2, 8, 1, 8, 2, 8
> *come disse Aristocle."* ← 4, 5, 9

i numeri delle lunghezze delle parole, si ottengono le prime cifre significative di e: $2, 7, 1, 8, 2, 8, 1, 8, 2, 8, 4, 5, 9$.
Fra le molte molte frasi analoghe, una famosa con le stesse cifre è

> *"Ai modesti o vanitosi*
> *ai violenti o timorosi*
> *do, cantando gaio ritmo,*
> *logaritmo."*

Altre notazioni:

$$\mathrm{e}^x \text{ si denota } \exp x \qquad \text{o meglio } \exp(x).$$

Raramente, e in questo testo quasi mai, si scrive anche

$$\exp_b a \text{ per intendere } b^a, \text{ per esempio } \exp_2 3 = 8$$

* **Nota sul valore esatto di e.** Questo numero è parente di pi greco, e come quello è irrazionale (e, come quello, pure *trascendente*). Il suo valore esatto è la "somma infinita" (serie) dei reciproci dei fattoriali dei numeri naturali:

$$\mathrm{e} := \frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \ldots$$

14.4 L'esponenziale in Farmacia

L'esponenziale ha una ricorrenza enorme nelle Scienze Applicate.

Ecco un'equazione classica della Farmacia, e precisamente della Farmacocinetica, che coinvolge l'esponenziale:

$$C = \frac{\dot{m}}{K} + \left(C_o - \frac{\dot{m}}{K}\right) \mathrm{e}^{-\frac{K \cdot t}{V}} \qquad (72)$$

L'esponenziale modellizza bene sia la fase iniziale di un'epidemia, che la fase iniziale dell'espansione di una popolazione microbica in coltura. La successione di Fibonacci, che sostanzialmente è ben approssimata da $\frac{1}{\sqrt{5}}\varphi^n$ che è una funzione esponenziale, è un buon modello:

1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89...

Un buon modello dello sviluppo successivo è la logistica, definita pur essa con l'esponenziale, che viene illustrata nel paragrafo seguente.

BOZZA - DRAFT

[72] Si veda https://en.wikipedia.org/wiki/Clearance_(pharmacology). Il simbolo $\dot{m}$ indica la *derivata di m rispetto al tempo*, si veda la Lezione 26.

14.5 Una sigmoide basata sull'esponenziale

Una questione interessante è quella della *logistica standard*

$$f(t) := \frac{1}{1 + e^{-t}}$$

un'altra funzione sigmoide ovvero con grafico una curva sigmoide.

È correlata allo sviluppo di una popolazione (animale, microbica...) quando si consideri che, ad un certo punto – e a differenza di quel che si considera nel modello ultra-semplificato della successione di Fibonacci – i vari organismi iniziano ad ostacolarsi a vicenda.
Si calcolino i suoi valori in -2, -1, $-\frac{1}{2}$, 0, $\frac{1}{2}$, 1, 2, e si disegni un grafico approssimativo.
Si veda il grafico su Wikipedia.

Ricordiamo che questo le curve sigmoidee hanno un'ampia ricorrenza nelle Scienze Applicate e in Calcolo delle Probabilità.
Per evidenze in Epidemiologia si cerchino in rete immagini con le parole chiave: "first wave" "cumulative".

BOZZA - DRAFT

14.6 Logaritmi

La funzione inversa dell'esponenziale in base b si chiama logaritmo in base b e la vedremo bene nella prossima lezione.

$$\log_{10} 1000 = 3 \text{ perchè } 10^3 = 1000$$

Per intanto si noti che tutti gli esponenziali in 0 valgono 1, e allora tutti i logaritmi in 1 valgono 0.
Si disegnino i grafici degli esponenziali e simmetricamente – rispetto alla bisettrice del I e III quadrante – dei logaritmi, fissando le basi 2 e poi 10.

Si verifichi poi online su WolframAlpha.

Sul grafico di WolframAlpha non si vede la simmetria rispetto alla bisettrice del I e III quadrante perché la scale sugli assi x e y sono diverse.

La funzione logaritmo in base 10 è caratterizzata da

- *"al decuplicare della x il $\log_{10} x$ cresce di 1"*

e similmente

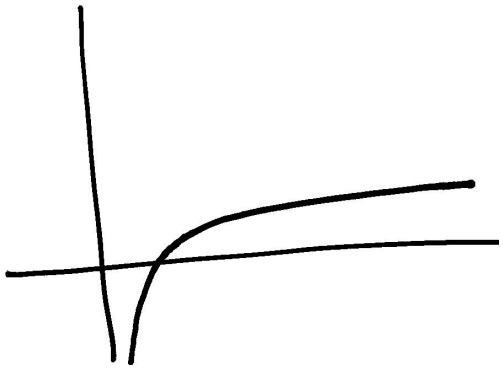
- al raddoppiare della x il $\log_2 x$ cresce di 1
- al dimezzare della x il $\log_2 x$ decresce di 1.

Cresce pianissimo! *("Striscia come un logaritmo").*

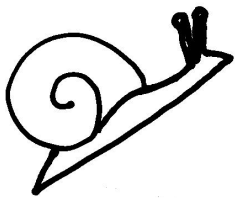
$$\log_{10} 1 = 0 \quad \log_{10} 10 = 1 \quad \log_{10} 100 = 2 \quad \log_{10} 1000 = 3$$

DRAFT

In pratica il logaritmo decimale di 1, 10, 100, 1000, 10000... è il numero di zeri; rimaniamo per ora un po' incerti su quanto valgano il logaritmo decimale degli altri numeri, e gli altri logaritmi.



Striscia come un logaritmo



Il lentissimo crescere del logaritmo ha profonde implicazioni in Analisi Matematica, Teoria dei Numeri, Complessità Computazionale, Fisica, Chimica...

14.7 I diagrammi logaritmici

BOZZA DRAFT

La lentissima crescenza del logaritmo in base 10 permette la costruzione dei diagrammi in scala logaritmica: invece di rappresentare nel piano cartesiano i punti, poniamo,

$$(1,1),\ (10,10),\ (100\,000,100\,000),$$

ciò che molto verosimilmente produrrebbe l'indistinguibilità pratica dei primi 2 punti, usiamo i logaritmi di quei numeri, cioè in pratica, in questo caso, il numero di zeri: prima $(0,0)$, poi $(1,1)$, poi $(5,5)$: e vengono punti ben distinguibili all'occhio. Naturalmente bisogna avvertire il lettore della scala logaritmica usata.

La rappresentazione in scala logaritmica può essere applicata a 1 o 2 degli assi del piano cartesiano (ci sono 3 casi denominati lin-log, log-lin e log-log). Per esempio in questo diagramma (semplificato) delle fasi dell'acqua, la scala logaritmica è applicata ai valori della pressione sull'asse delle ordinate.

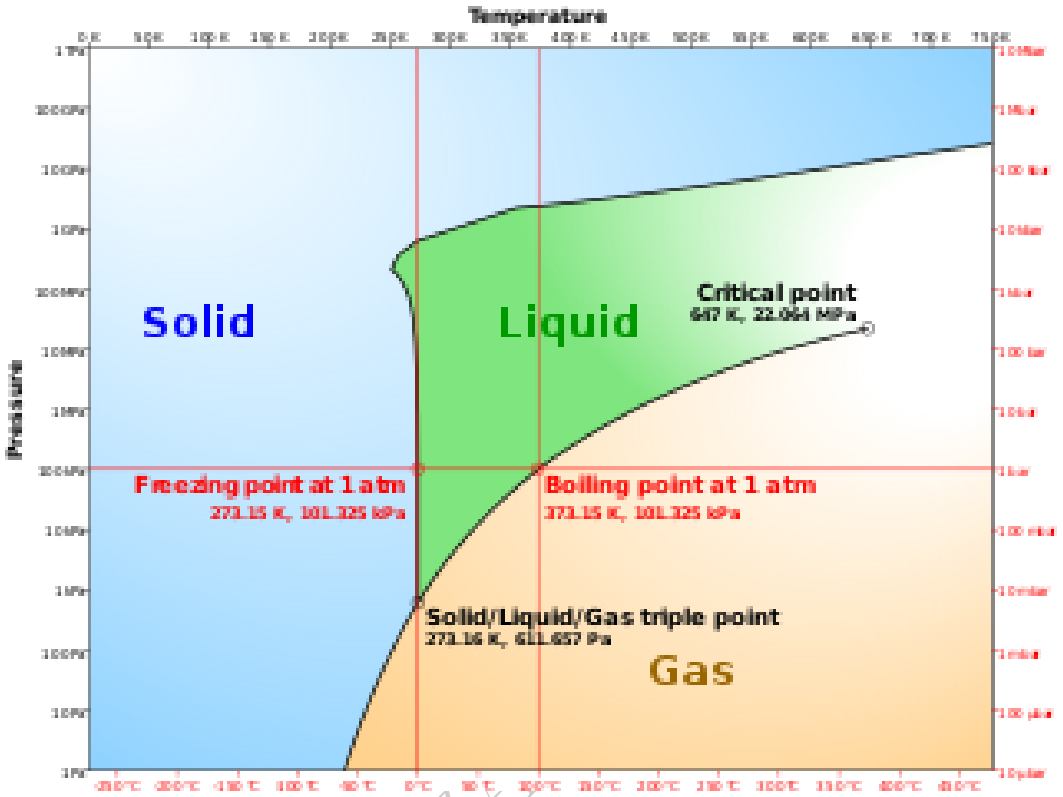


Figure 17: Diagramma di fase dell'acqua, semplificato. By Cmglee, in Wikimedia.

BOZZA

Si vedano ancora per esempio le (veramente appropriate per l'argomento che trattiamo) figure 1a (*linear-log plot*) e 1b (*log-log plot*) in un articolo sui farmaci anti-HIV.

Grafici epidemici in scala logaritmica. Nella figura sottostante sono considerati tutti gli stati confinanti con l'Italia settentrionale (Francia, Svizzera, Austria, Slovenia) e in più quelli di una vastissima area, dal Friuli Venezia Giulia fino ai confini di Russia e Bielorussia.

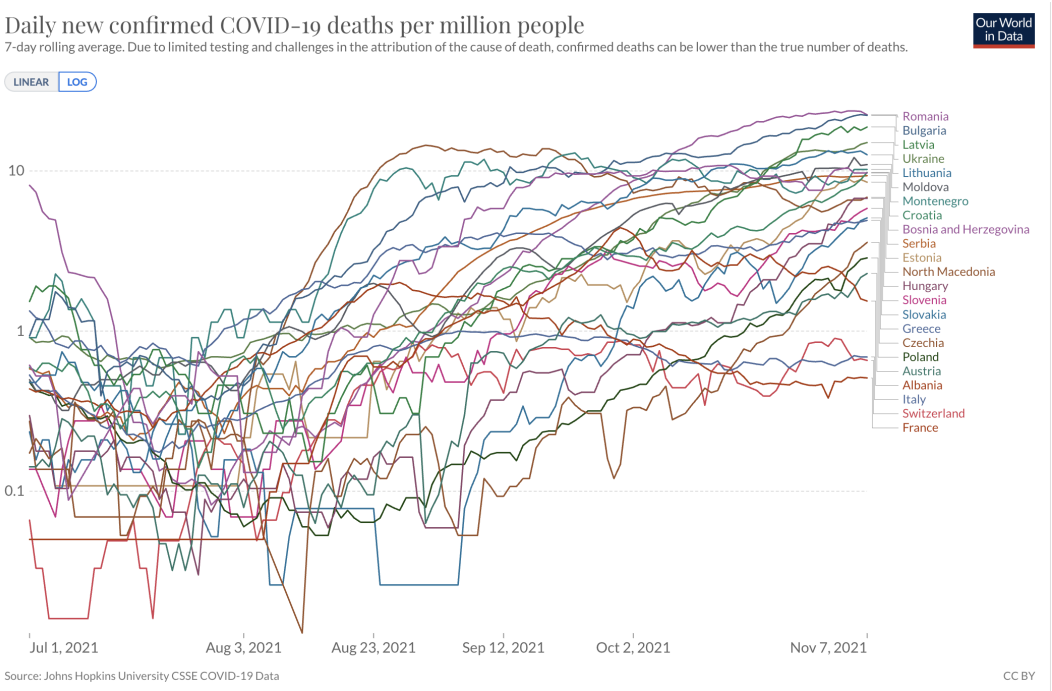


Figure 18: Decessi covid-19 giornalieri, media mobile a 7 giorni

In un normale diagramma cartesiano sarebbero praticamente indistinguibili i grafici di Italia, Francia e Svizzera, ben distinti nel log plot.

Questi 3 stati hanno mortalità covid giornaliera, all'inizio di novembre 2021, sotto l'1 per milione, gli altri stati considerati sopra 1 e alcuni di molto.

Come si verifica facilmente, i primi 3 stati considerati hanno avuto una drammatica prima ondata della pandemia nella primavera del 2020, mentre nessuno degli altri l'ha avuta, iniziando ad avere grandi numeri solo dopo l'estate 2020.

14.8 Logaritmi in Farmacia e nelle Scienze Applicate

I logaritmi hanno una ricorrenza enorme nelle Scienze Applicate.

Ecco 5 ricorrenze dei logaritmi nella Farmacia, che vedremo:

1) i diagrammi in scala logaritmica, sopra accennati;
2) il pH della Chimica;
3) il decibel usato nella misurazione dell'udito;
4) la clearance dei farmaci;

5) La Curva di Preston, del prossimo paragrafo

14.9 La curva di Preston

L'aspettativa di vita alla nascita è uno dei principali indicatori di sviluppo di uno stato (correlato anche a Medicina e Farmacia) insieme alla mortalità infantile (a un anno, e in subordine a 5 anni) e al reddito pro-capite.
Il primo e l'ultimo di questi indicatori, y e x, appaiono da molto tempo correlati a livello mondiale da questo modello empirico:

$$y = 6.6354 \log_e x + 10.754 + error$$

con *error* positivo o negativo ma di solito non grande, il cui grafico – tolto l'*error* – si chiama *curva di Preston*. Ecco la curva e lo scatterplot.

- DRAFT

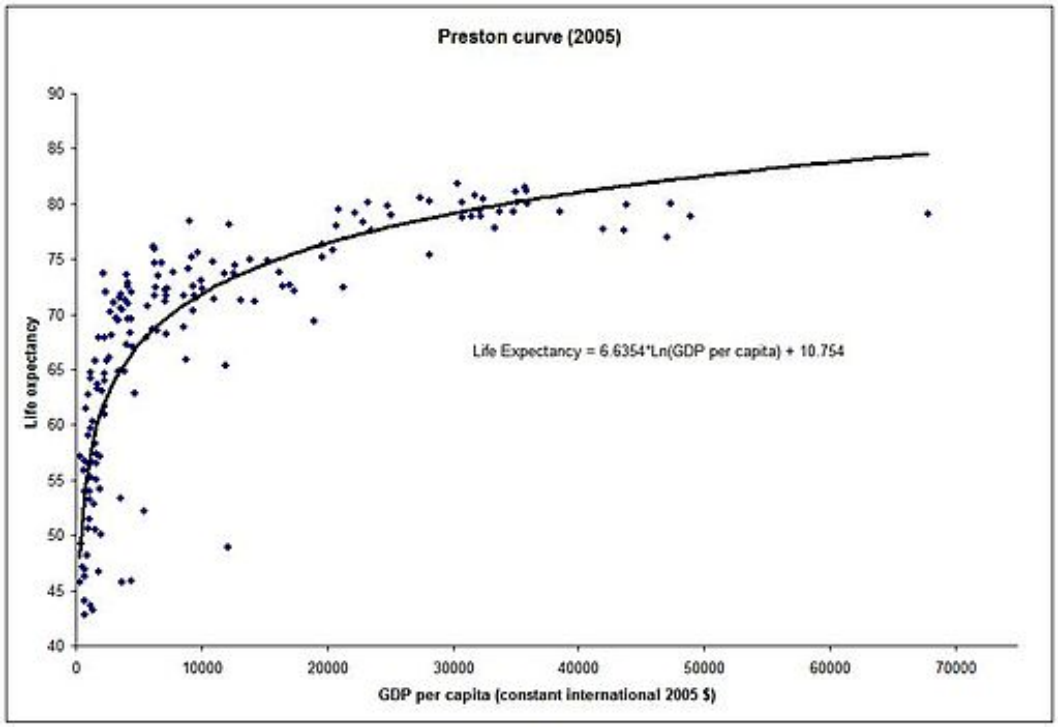


Figure 19: Curva di Preston. Da Wikimedia Commons.

Nota. Sebbene il diagramma illustri una correlazione e non dimostri – in senso stretto – una causalità, induce a riflettere: di quanto si può impoverire una comunità con lockdown e limitazioni ad attività commerciali, per prevenire una malattia che causa morte, senza causare un numero maggiore di morti per impoveriimento? Icasticamente e semplificatamente – senza nessuna pretesa di esattezza numerica – possiamo dire:

"LA FAMIGLIA BENESTANTE IMPOVERITA RINUNCIA A METÀ DELLE AUTOMOBILI, LA FAMIGLIA POVERA IMPOVERITA A METÀ DEL CIBO, CHE A QUEL PUNTO POTREBBE NON BASTARE PER SOPRAVVIVERE."

Di fatto,dall'inizio del covid sono emersi aumenti di mortalità enormi, in vari Stati, solo in minor parte imputabili direttamente al covid. Bisognerà studiare attentamente le cause a cui sono dovuti: impoverimento, mancata cura di altre malattie, riduzione del movimento fisico... Ci sono varie ipotesi, e ognuna avrà il suo peso, positivo o negativo, da determinare scientificamente.

BOZZA - DRAFT

Limitandoci qua fortemente, anche per non farci il sangue troppo, troppo amaro, sulla gestione della pandemia del covid citiamo[73] comunque un editoriale del 2020 del British Medical Journal, una delle più importanti riviste mediche del mondo:

> "Science is being suppressed for political and financial gain. Covid-19 has unleashed state corruption on a grand scale, and it is harmful to public health. (...) When good science is suppressed, people die."

Naturalmente all'impoverimento di miliardi di persone ha corrisposto l'arricchimento di alcuni.

Dal punto di vista della Statistica, la pandemia – ovvero chi aveva il potere di indirizzo sulle comunità umane, non il virus *in persona* – ha *spostato* smisurate quantità di ricchezza da molti a certi pochi, in tempi prima inconcepibilmente

[73] Covid-19: politicisation, "corruption," and suppression of science BMJ 2020; 371 doi: https://doi.org/10.1136/bmj.m4425 (Published 13 November 2020) Cite this as: BMJ 2020;371:m4425

rapidi.
Trascriviamo da Oxfam in
https://www.oxfamitalia.org/la-disuguaglianza-non-conosce-crisi/

> Nel biennio pandemico '20-'21 l'1% più ricco ha visto crescere il valore dei propri patrimoni di 26.000 miliardi di dollari, in termini reali, accaparrandosi il 63% dell'incremento complessivo della ricchezza netta globale (42.000 miliardi di dollari), quasi il doppio della quota (37%) andata al 99% più povero della popolazione mondiale.
> (...) Per la prima volta in 25 anni aumentano inoltre simultaneamente estrema ricchezza ed estrema povertà.

Mnemonico per il nome Preston

FATE **PRESTO**OOOOO....

BOZZA - DRAFT

ESERCIZIO$_{\mu}$ $\approx$ Risolvere l'equazione $x\,\mathrm{e}^{-\frac{3}{2}} = 1$.

SVOLGIMENTO

Verrà usato lo standard del punto decimale. (Ma si potrebbe usare lo standard della virgola decimale).

Dividiamo ambo i membri per uno stesso numero diverso da 0 (essendo il valore di un esponenziale)

$$x\,\mathrm{e}^{-\frac{3}{2}} = 1 \qquad \Big/ : \mathrm{e}^{-\frac{3}{2}} \neq 0$$

$$x = \frac{1}{\mathrm{e}^{-\frac{3}{2}}} =$$

che è già la soluzione esatta (seppure non bene espressa) e poi per le proprietà delle potenze

$$x = \mathrm{e}^{\frac{3}{2}} =$$

e già abbiamo la soluzione esatta ed espressa bene, che ora approssimeremo; per le proprietà delle potenze

$$= \left(\mathrm{e}^{3}\right)^{\frac{1}{2}} =$$

per le proprietà delle potenze e delle radici

$$= \sqrt{\mathrm{e}^{3}} = \sqrt{\mathrm{e} \cdot \mathrm{e} \cdot \mathrm{e}} \approx$$

ricordando il valore approssimato di e, e poi con la calcolatrice

$$\approx \sqrt{2.718 \cdot 2.718 \cdot 2.718} \approx \sqrt{20.079} \approx$$

$$\boxed{\approx 4.481}$$

ma senza voler azzardare tanti decimali (comunque giusti, come si può verificare con una buona calcolatrice o col computer) diciamo con maggior sicurezza, e comunque buona accuratezza,

$$\boxed{4.48}$$

Esercizio$_{\mu dispensa}$ (R) $\approx$ Quanto vale, con 2 decimali dopo la virgola, la media geometrica di pi greco e il numero di Nepero ovvero Eulero?

$$\boxed{2{,}92}$$

BOZZA DRAFT

(È la radice quadrata del prodotto di 3,14 e 2,718. Con queste approssimazioni le calcolatrici danno 2,9213... che magari ci verrebbe la tentazione di approssimare con 2,921 ma molto opportunamente il testo del quesito ci previene dal farlo chiedendo 2 decimali dopo la virgola: infatti la terza cifra decimale dopo la virgola in 2,921 non è corretta relativamente al quesito posto (formulato per π ed e, non per loro valori approssimati). La cifra decimale corretta, terza dopo la virgola, sarebbe il 2 di 2,9222... – ma appunto non ci riguarda – e questo succede perchè le classiche approssimazioni che si ritengono usualmente a memoria, 3,14 e 2,718, non sono i valori esatti di π ed e).

BOZZA - DRAFT

15 Esponenziali e logaritmi – II parte

15.1 Definizione del logaritmo

Nella formula seguente

– il 2023 può essere sostituito con qualunque numero positivo;

– il 7 può essere sostituito con qualunque numero positivo $\neq 1$;

la formula resta vera: e metà della teoria del logaritmo è spiegata.

$$2023 = 7^{\log_7 2023} = 7^{3{,}911\,967\ldots}$$

Se interessa anche il valore numerico del logaritmo del nuovo numero nella nuova base, il $3,911\,967\ldots$ andrà sostituito col nuovo valore numerico $\log_b a$ che si può trovare per esempio col computer.

Una terzina di ottonari:

Logaritmo ad esponente,
Se si mette, la sua base
Lui si mangia, evanescente.

BOZZA - DRAFT

Il logaritmo in base b di x è l'esponente da dar a b per aver x.

(Ma) dev'essere $x > 0$, e $0 < b \neq 1$.

In tutta questa lezione col simbolo LOG intenderemo un logaritmo in qualunque base.

E ovviamente in una formula in cui compaiono più LOG è da intendersi che sono tutti in una stessa base.

15.2 Logaritmi naturali e decimali; esponenziale

Come abbiamo anticipato, la funzione inversa dell'esponenziale in base b si chiama logaritmo in base b.
Facciamo un passo indietro, verso l'algebra.
Per introdurre i logaritmi algebricamente, possiamo vedere

$$\log_2 8 = 3 \qquad \text{come una } \textit{scrittura alternativa} \text{ di} \qquad 2^3 = 8.$$

Ecco due esempi scritti come implicazioni:

$$\log_2 8 = 3 \Leftarrow 2^3 = 8 \qquad \log_{10} 100 = 2 \Leftarrow 10^2 = 100.$$

Considereremo quasi solo 2 basi: di grande importanza applicativa per secoli, e ancor oggi usato in Chimica,

il *logaritmo decimale*, in base 10,

$$\lg x \qquad \textit{oppure} \ \log_{10} x \qquad (\text{ma purtroppo anche } \log x),$$

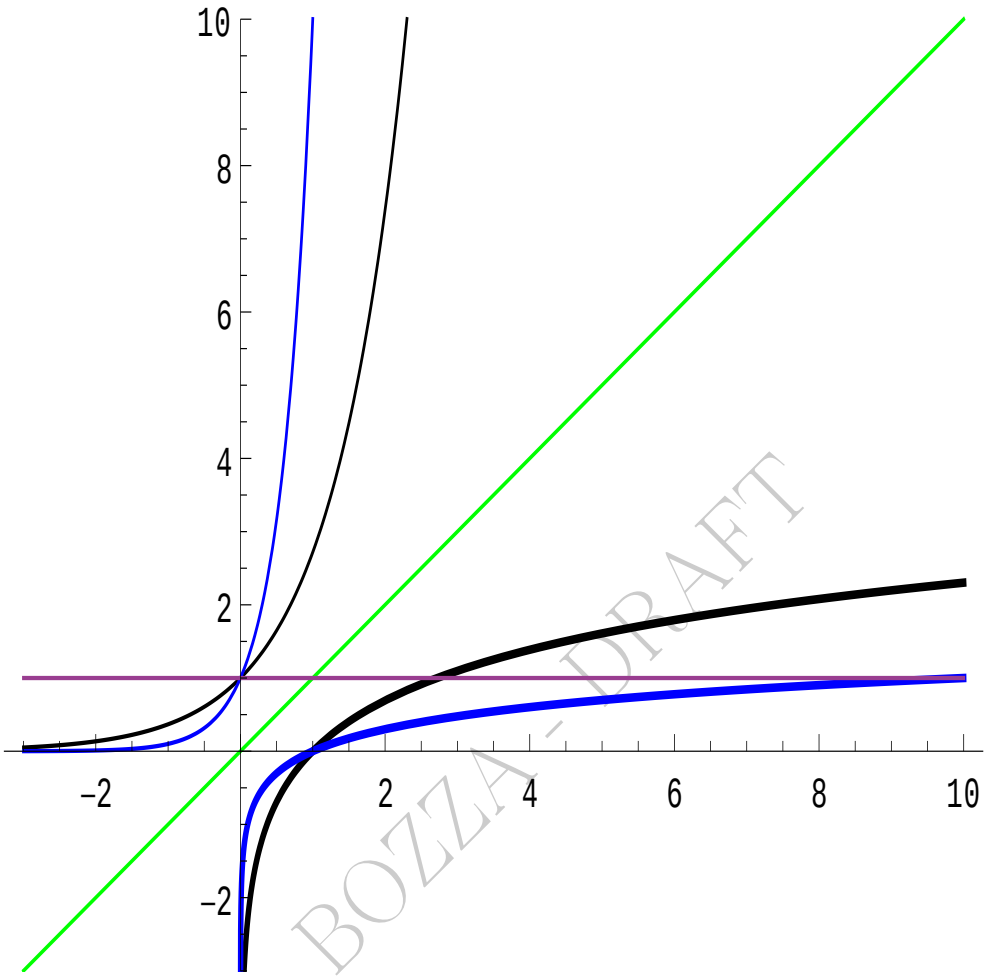
e il *logaritmo naturale*, in base e,

$$\ln x \qquad \textit{oppure} \ \log_e x \qquad (\text{ma purtroppo anche } \log x).$$

Insomma la scrittura log x è ambigua e può riferirsi a logaritmi in base e e in base 10, e anche in altre basi. 😕

La base 10 è stata di grande importanza applicativa per secoli, e ancor oggi è usata in Chimica.
La base e è oggi preferita in Matematica e Fisica.

BOZZA - DRAFT



BOZZA - DRAFT

Figure 20: $\log_e x$ (**nero**), e^x (nero), $\log_{10} x$ (**blu**), 10^x (blu)

Ecco due valori che ci proponiamo di imparare a memoria:

$$\boxed{\lg 2 \approx \mathbf{0.3}} \tag{25}$$

O meglio ≈ 0.301. Il numero 0.3 ha errore (percentuale rispetto l'esatto) $<0.5\%$.

$$\boxed{\lg \mathrm{e} \approx \mathbf{0.4343}} \tag{26}$$

Tutti i logaritmi in 1 valgono 0. I logaritmi in base $b > 1$ sono funzioni crescenti ("verso $+\infty$"), e in base $0 < b < 1$ decrescenti ("verso $-\infty$"). In ogni caso il crescere in valore assoluto avviene con straordinaria lentezza, per esempio $\log_{10} 1000 = 3$. I limiti in 0 sono rispettivamente $-\infty$ e $+\infty$. In ogni caso è $\log_b : \mathbb{R}^+ \to \mathbb{R}$.

(Per ogni possibile base b) il logaritmo in base b è una funzione biiettiva, e la sua inversa è la funzione esponenziale in base b, cioè b^x.
L'inversa del logaritmo naturale è la *funzione esponenziale*, ed è definita su tutto $\mathbb{R}$, cioè $\exp : \mathbb{R} \to \mathbb{R}$:

$$\exp = \ln^{-1} \qquad \ln = \exp^{-1},$$

ed è la funzione esponenziale di base e:

$$\exp x = \mathrm{e}^x.$$

Questi $^{-1}$ denotano l'inversa, assolutamente non il reciproco.

(L'inversa di $\ln x$ come detto è $\exp x$, la reciproca di $\ln x$ è $\frac{1}{\ln x}$, una funzione completamente diversa, neanche definita per $x < 0$).

L'esponenziale in 0 vale 1, ed è crescente "verso $+\infty$" con grande rapidità, per esempio, $\exp(20)$ è più o meno mezzo miliardo.

BOZZA - DRAFT

Ripetiamo.
In tutta questa lezione col simbolo LOG intenderemo un logaritmo in qualunque base.
E ovviamente in una formula in cui compaiono più LOG è da intendersi che sono tutti in una stessa base.

15.3 "Trasforma prodotti in somme"

La proprietà fondamentale dei logaritmi:

$$\boxed{\mathrm{LOG}(x \cdot y) = \mathrm{LOG}(x) + \mathrm{LOG}(y)} \tag{27}$$

Per esempio

$\lg 2000 =$
$= \lg(2 \cdot 1000) =$
$= \lg 2 + \lg 1000 \approx$
$\approx 0.3 + 3 =$
$= 3.3$ (valore approssimato).

Un valore più preciso è 3.301.

Con l'estrema lentezza della crescita del logaritmo, il logaritmo decimale di 2022 non si allontana molto: si trova infatti ≈ 3.306.

15.4 "Trasforma divisioni in sottrazioni"

$$\boxed{\mathrm{LOG}\,(x/y) = \mathrm{LOG}(x) - \mathrm{LOG}(y)} \tag{28}$$

BOZZA - DRAFT

Per esempio

$\lg 5 =$
$= \lg \frac{10}{2} =$
$= \lg 10 - \lg 2 \approx$
$\approx 1 - 0.3 =$
$= 0.7$ (valore approssimato).

15.5 "Trasforma reciproco in opposto"

$$\boxed{\mathrm{LOG}\left(\frac{1}{y}\right) = -\mathrm{LOG}(y)} \tag{29}$$

che si ottiene dal caso del precedente con $x = 1$: $\mathrm{LOG}(1/y) = \mathrm{LOG}(1) - \mathrm{LOG}(y) = 0 - \mathrm{LOG}(y)$.

Per esempio

$\lg 0.5 =$
$= \lg \frac{1}{2} =$
$= -\lg 2 \approx -0.3$

Esempio. In Spettroscopia, l'*assorbanza* è[74] l'opposto del logaritmo decimale della *trasmittanza*, ovvero il logaritmo decimale del suo reciproco:

$$\alpha = -\lg \tau = \lg \frac{1}{\tau}$$

[74] Si veda Wikipedia, l'enciclopedia libera, alla voce *Trasmittanza*

15.6 "Trasforma la radice ennesima in divisione per n"

$$\boxed{\mathrm{LOG}(\sqrt[n]{x}) = \tfrac{1}{n}\,\mathrm{LOG}(x)} \tag{30}$$

In particolare

"Trasforma l'estrazione di radice quadrata nel dimezzamento":

$$\boxed{\mathrm{LOG}(\sqrt{x}) = \tfrac{1}{2}\,\mathrm{LOG}(x)} \tag{31}$$

Per esempio

$$\lg\sqrt{10} = \frac{1}{2}\lg 10 = \frac{1}{2}\cdot 1 = 0.5$$

Osservando che $\sqrt{10} \approx 3.16 \approx \pi$ troviamo

$$\lg\pi \approx 0.5$$

e con la calcolatrice scientifica troviamo infatti ≈ 0.497.

BOZZA - DRAFT

15.7 "Trasforma l'elevare a potenza in moltiplicare"

Logaritmo della potenza.

"Tira giù gli esponenti"

$$\boxed{\mathrm{LOG}(x^{\alpha}) = \alpha\,\mathrm{LOG}(x)} \tag{32}$$

per ogni α e x per cui ha senso[75] la scrittura, e da quest'ultima discendono alcune delle precedenti proprietà, ponendo α uguale a -1 e poi $\frac{1}{n}$ e poi $\frac{1}{2}$.

Da quell'uguaglianza, con facili argomentazioni che comunque non facciamo, si trova che

"Tira giù l'esponente 2 e aggiunge il valore assoluto":

$$\boxed{\mathrm{LOG}(x^2) = 2\,\mathrm{LOG}(|x|)} \tag{33}$$

e più in generale per n intero

"Tira giù gli esponenti pari e aggiunge il valore assoluto":

$$\boxed{\mathrm{LOG}(x^{2n}) = 2n\,\mathrm{LOG}(|x|)} \tag{34}$$

[75] Precisamente: $\forall\,\alpha \in \mathbb{R}, \forall\, x > 0$.

Naturalmente anche senza valore assoluto le ultime 2 formule sono vere, ma col valore assoluto valgono anche per $x < 0$.

Come esempio della proprietà del logaritmo della potenza, considerando 1.024 vicino a 1, si ottiene:

$$\frac{1024}{1000} = 1.024$$
$$\frac{1024}{1000} \approx 1 \qquad \Big/ \lg$$
$$\lg\left(\frac{1024}{1000}\right) \approx \lg 1$$
$$\lg 1024 - \lg 1000 \approx 0$$
$$\lg 1024 \approx \lg 1000$$
$$\lg 2^{10} \approx \lg 10^3$$
$$10 \lg 2 \approx 3$$
$$\lg 2 \approx \frac{3}{10}$$

che spiega l'approssimazione $\lg 2 \approx 0.3$ che ci siamo proposti di imparare a memoria.

BOZZA - DRAFT

15.8 Le proprietà viste, con riferimento alle varie basi

Come già detto

"Trasforma prodotti in somme":

formule per il *logaritmo del prodotto* ed *esponenziale della somma*)

$$\lg(x\,y) = \lg x + \lg y, \ \ \forall\, x, y > 0 \qquad 10^{x+y} = 10^x \cdot 10^y$$
$$\ln(x\,y) = \ln x + \ln y, \ \ \forall\, x, y > 0 \qquad \mathrm{e}^{x+y} = \mathrm{e}^x \cdot \mathrm{e}^y$$
$$\log_b(x\,y) = \log_b x + \log_b y, \ \ \forall\, x, y > 0 \qquad b^{x+y} = b^x \cdot b^y$$

Come già detto

"Trasforma divisioni in sottrazioni":

(formule per il *logaritmo del quoziente* ed *esponenziale della differenza*)

$$\lg(x/y) = \lg x - \lg y, \ \ \forall\, x, y > 0 \qquad 10^{x-y} = 10^x/10^y$$
$$\ln(x/y) = \ln x - \ln y, \ \ \forall\, x, y > 0 \qquad \mathrm{e}^{x-y} = \mathrm{e}^x/e^y$$

$$\log_b(x/y) = \log_b x - \log_b y, \quad \forall\, x, y > 0 \qquad\qquad b^{x-y} = b^x / b^y \,.$$

Come già detto

"Tira giù gli esponenti":

(formule per il *logaritmo della potenza*)

$$(\forall\, \alpha \in \mathbb{R})(\forall\, x > 0)\ \lg x^\alpha = \alpha \lg x$$

$$(\forall\, \alpha \in \mathbb{R})(\forall\, x > 0)\ \ln x^\alpha = \alpha \ln x$$

$$(\forall\, \alpha \in \mathbb{R})(\forall\, x > 0)\ \log_b x^\alpha = \alpha \log_b x$$

e da quest'ultima discendono tutte le successive proprietà di questo paragrafo, ponendo α uguale a -1 e poi $\frac{1}{2}$ e poi $\frac{1}{n}$.

E per esponenti pari anche

$$(\forall\, n \in \mathbb{Z})(\forall\, x \neq 0)\ \lg x^{2n} = 2n \lg |x|$$

$$(\forall\, n \in \mathbb{Z})(\forall\, x \neq 0)\ \ln x^{2n} = 2n \ln |x|$$

$$(\forall\, n \in \mathbb{Z})(\forall\, x \neq 0)\ \log_b x^{2n} = 2n \log_b |x|$$

BOZZA DRAFT

Come già detto

"Trasforma il passaggio al reciproco nel passaggio all'opposto:"

$$\lg \frac{1}{x} = -\lg x$$

$$\ln \frac{1}{x} = -\ln x$$

$$\log_b \frac{1}{x} = -\log_b x$$

Come già detto

"Trasforma l'estrazione di radice quadrata nel dimezzamento:"

$$\lg \sqrt{x} = \frac{1}{2} \lg x$$

$$\ln \sqrt{x} = \frac{1}{2} \ln x$$

$$\log_b \sqrt{x} = \frac{1}{2} \log_b x$$

e questo si estende alle radici di qualunque indice:

$$\lg \sqrt[n]{x} = \frac{1}{n} \lg x$$

$$\ln \sqrt[n]{x} = \frac{1}{n} \ln x$$

$$\log_b \sqrt[n]{x} = \frac{1}{n} \log_b x \, .$$

Esercizio$_{\mu 2021}$ $\approx$
Risolvere l'equazione

$$\lg \mathrm{e}^{2x} = 8$$

SVOLGIMENTO

$$\lg \mathrm{e}^{2x} = 8$$

$$2x \lg \mathrm{e} = 8 \quad \not / : 2 \lg \mathrm{e}$$

$$\Rightarrow x = \frac{8}{2 \lg \mathrm{e}} \approx \frac{8}{2 \cdot 0.4343} \approx$$

$$\boxed{\approx 9.21}$$

BOZZA - DRAFT

15.9 Prevenire è meglio che curare: errori tipici

Questa formula già vista

$$\mathrm{LOG}(x \cdot y) = \mathrm{LOG}(x) + \mathrm{LOG}(y)$$

è vera ogni volta che ha senso (e cioè per x e y positivi) e riguarda il logaritmo del prodotto.
Errori tipici sono scambiare fra loro il + col $\cdot$ e simili[76] cose.

[76] In generale la sequenza $\cdot$ + della formula non si può sostituire con
+ $\cdot$
$\cdot$ $\cdot$
+ +.
Detto altrimenti:

$$\mathrm{LOG}(x \cdot y) \equiv \mathrm{LOG}(x) + \mathrm{LOG}(y)$$

Il logaritmo
Vispo volpino LOG
Con un suo ritmo
Fa uno spuntino.

Un bel prodotto LOG(X•Y)
Vorace ingoia,
Poi che lo ha rotto,
Che il punto muoia. •
Proprio soccomba:
Lascia una tomba. +

E lui si sdoppia: LOG(X) + LOG(Y)
Forma una coppia.

BOZZA - DRAFT

15.10 Esempio di uso del logaritmo: il pH.

Il pH è una complessa questione chimica, per la quale si rinvia comunque ai testi specialistici. Qua vogliamo illustrarne qualche proprietà matematica.
Leggiamo su Wikipedia, l'enciclopedia libera, alla voce pH:

> pH is defined as the decimal logarithm of the reciprocal of the hydrogen ion activity, a_{H+}, in a solution

$$\mathrm{pH} = -\log_{10}(a_{H+}) = \log_{10}\left(\frac{1}{a_{H+}}\right)$$

è un'identità, cioè un'uguaglianza sempre vera (per gli x e y per cui ha senso, cioè positivi) invece per esempio

$$\mathrm{LOG}(x+y) =^{EQ} \mathrm{LOG}(x) + \mathrm{LOG}(y)$$

è un'equazione, vera solo per particolari valori (x, y), per esempio $(2, 2)$:

$$\mathrm{LOG}(2+2) = \mathrm{LOG}(4) = \mathrm{LOG}(2^2) = 2\,\mathrm{LOG}(2) = \mathrm{LOG}(2) + \mathrm{LOG}(2).$$

Ma non succede certo con $(3, 3)$ o $(2, 3)$ o $(3, 2)$.

Ripetiamo la formulazione giusta:
logaritmo del prodotto = somma dei logaritmi. (Ovviamente, nella stessa base).

Se il pH diminuisce di 1 questo equivale al decuplicare di a_H+:

$$pH - 1 = \log_{10}\left(\frac{1}{a_{H+}}\right) - 1 =$$

$$= \log_{10}\left(\frac{1}{a_{H+}}\right) + (-1) =$$

$$= \log_{10}\left(\frac{1}{a_{H+}}\right) + \log_{10}\frac{1}{10} =$$

$$= \log_{10}\left(\frac{1}{a_{H+}} \cdot \frac{1}{10}\right) =$$

$$= \log_{10}\left(\frac{1}{10\,a_{H+}}\right)$$

Così il suo diminuire di 2 corrisponde al centuplicare di a_{H+}.

BOZZA - DRAFT

15.11 Esempi di calcoli coi logaritmi

Esempi con le proprietà 1, 2 e 4 del logaritmo decimale.

$\lg 4 = \lg(2 \cdot 2) = \lg 2 + \lg 2 \approx 0.3 + 0.3 = 0.6$

$\lg 8 = \lg(2 \cdot 2 \cdot 2) = \lg 2 + \lg 2 + \lg 2 \approx 0.3 + 0.3 + 0.3 = 0.9$

$\lg 5 = \lg(10/2) = \lg 10 - \lg 2 \approx 1 - 0.3 = 0.7$

$\lg 7 = \lg\sqrt{49} = \frac{1}{2}\lg 49 \approx$
sostituiamo approssimatamente 49 con 50
$\approx \frac{1}{2}\lg 50 = \frac{1}{2}\lg(5 \cdot 10) = \frac{1}{2}(\lg 5 + \lg 10) \approx \frac{1}{2}(0.7 + 1) \approx 0.85$

(Ci mancano 3, 6 e 9, che vedremo; 1, 2 e 10, e ora 4, 5, 7, 8, li abbiamo: provate a disegnare un grafico e stimare così i 3 valori mancanti; poi verificherete le stime).

Esempio: calcolo del $\log_{10}(50\,000)/2$ dei prodotti omeopatici.
Leggiamo su Wikipedia (inglese)[77] alla voce *Homeopathic dilutions*:

[77] https://en.wikipedia.org/wiki/Homeopathic_dilutions

> Several potency scales are in use (...) If a dilution is designated as q on the Q scale, and c on the C scale, c/q=$\log_{10}(50{,}000)/2$

cioè vale una sorta di formula di conversione fra scale

$$c = q \times \frac{\lg 50\,000}{2}$$

e ora vogliamo calcolare quel numero, col "nostro" $\lg 5 \approx 0.7$:

$$\frac{\lg 50\,000}{2} = \frac{\lg 5 \cdot 10^4}{2} = \frac{\lg 5 + \lg 10^4}{2} \approx \frac{0.7 + 4}{2} = 2.35$$

e infatti dice, e ora lo capiamo un po' meglio,

> A given dilution on the Q scale is roughly 2.35 times its designation on the C scale.

(La formula soprascritta c/q=$\log_{10}(50{,}000)/2$ si trova facilmente[78]).

Alcuni Autori e Produttori scrivono LM invece di Q (il significato inteso sarebbe 50 e 1000 in numeri romani, ma LM non significa 50 000 nei numeri romani). E alcuni purtroppo lm, minuscolo, con difficile interpretazione della elle minuscola iniziale (sul web dove si vendono tali prodotti, pare una i maiuscola).
La formula di conversione si riferisce solo alla *teorica* diluizione *in base al modello* usato, e darebbe per esempio l'equivalenza di 20 Q con 47 C ovvero CH, ma i due preparati sono stati ottenuti con procedura diversa e non vengono considerati equivalenti dal punto di vista terapeutico.

BOZZA - DRAFT

[78] La formula viene da

$$100 \cdot ...(c\ volte)... \cdot 100 = 50\,000 \cdot ...(q\ volte)... \cdot 50\,000$$

$$100^c = 50\,000^q$$

$$(10^2)^c = 50\,000^q$$

$$10^{2c} = 50\,000^q \quad \Big/ \lg$$

$$2c \lg 10 = q \lg 50\,000$$

$$2c = q \lg 50\,000 \quad \Big/ \cdot \frac{1}{2q}$$

$$\frac{c}{q} = \frac{\lg 50\,000}{2}$$

dove la prima equazione rappresenta le diluizioni centesimali che avanzano di 100 volte in 100 volte sulla scala C, e le diluizioni cinquantamillesimali che avanzano di 50 000 volte in 50 000 volte sulla scala Q.

15.12 Altre proprietà degli esponenziali e dei logaritmi

In questo paragrafo, b rappresenta qualunque numero reale che possa essere base di un logaritmo, cioè

$$b > 0 \wedge b \neq 1 \, .$$

Un'attenta considerazione di quanto detto finora dà

$$(\forall x > 0, x \neq 1) \qquad \log_x x = 1$$

e queste 6:

"exp mangia log e log mangia exp"

$$\lg 10^x = x, \ \forall x \qquad 10^{\lg x} = x, \ \forall x > 0.$$

$$\ln \exp x = \ln \mathrm{e}^x = x, \ \forall x \qquad \exp \ln x = \mathrm{e}^{\ln x} = x, \ \forall x > 0$$

$$\log_b b^x = x, \ \forall x \qquad b^{\log_b x} = x, \ \forall x > 0.$$

BOZZA - DRAFT

Tutte le successive proprietà degli esponenziali ci sono già note dall'Algebra, e alcune proprieà dei logaritmi le abbiamo viste da poco; altre sono nuove.

I logaritmi in 1 valgono 0 e gli esponenziali in 0 valgono 1:

$$\lg 1 = 0 \qquad 10^0 = 1$$

$$\ln 1 = 0 \qquad \mathrm{e}^0 = 1$$

$$\log_b 1 = 0 \qquad b^0 = 1$$

Altri esempi con le proprietà del logaritmo decimale.

$\lg 3 = \lg \sqrt[5]{3 \cdot 3 \cdot 3 \cdot 3 \cdot 3} = \lg \sqrt[5]{243} \approx$
sostituiamo approssimatamente 243 con 250

$$\approx \lg \sqrt[5]{250} = \frac{1}{5} \lg 250 = \frac{1}{5} \lg \frac{1000}{4} =$$

$$= \frac{1}{5}\left(\lg 1\,000 - \lg 4\right) = \frac{3 - 0.6}{5} = 0.48$$

$\lg 6 = \lg(3 \cdot 2) = \lg 3 + \lg 2 \approx 0.48 + 0.3 = 0.78$

$\lg 9 = \lg(3 \cdot 3) = \lg 3 + \lg 3 \approx 0.48 + 0.48 \approx 0.96$ ☹

ma per eccesso di approssimazioni quest'ultimo fallisce sul secondo decimale, infatti troviamo con calcolatrice scientifica o WolframAlpha

$$\lg 9 \approx 0.95$$

Adesso abbiamo i logaritmi - in effetti approssimazioni in generale – dei numeri 1, 2, 3, 4, 5, 6, 7, 8, 9, 10; aggiungiamo

$\lg 0.1 = \lg \frac{1}{10} = -\lg 10 = -1$

e disegniamo un grafico!

BOZZA - DRAFT

15.13 ESERCIZI SULLA LEZIONE 15

ESERCIZIO$_{\mu 2019}$
Consideriamo il pH della Chimica:

$$-\log_{10} y$$

dove y viene spesso scritto $[H^+]$ e letto *"concentrazione degli H^+"*, ma tutto ciò richiederebbe precisazioni di Chimica di cui qua non ci occupiamo.
A cosa corrisponde la diminuzione di 0,5 nel pH? Con grossolana approssimazione, si esprima infine la soluzione a parole, come "circa dimezzare y" o "circa decuplicare y" o analoghi.

SVOLGIMENTO
Viene usato lo standard della virgola decimale. (Già nel quesito).

$$-\log_{10} y - 0{,}5 =$$

ricordando che $x \equiv \log_b b^x$

$$= -\log_{10} y - \log_{10} 10^{0,5} =$$

e osservando che $0{,}5 = \frac{1}{2}$ e ricordando che $y^{\frac{1}{2}} = \sqrt{y}$

$$= -\log_{10} y - \log_{10} \sqrt{10} =$$

per algebra delle parentesi

$$= -\Big(\log_{10} y + \log_{10} \sqrt{10}\Big) =$$

per la proprietà del logaritmo del prodotto

$$= -\log_{10}\Big(y \cdot \sqrt{10}\Big)$$

cioè a dire, y viene moltiplicata per $\sqrt{10} \approx 3{,}16$, e con grossolana approssimazione, come richiesto,

circa triplicare y

BOZZA - DRAFT

16 Esponenziali e logaritmi – III parte

16.1 Conversioni fra logaritmi naturali e decimali

$$\ln x \approx \frac{\lg x}{0.4343}$$

da cui (e ci basterà ricordare questa seconda[79] per riottenere all'occorrenza la prima)

$$\boxed{\lg x \approx 0.4343 \ln x} \tag{35}$$

(cioè il logaritmo decimale è *vagamente* la metà del naturale e con maggior precisione il 43% e ancor meglio 43.43%).

$$\Big(\text{Le formule esatte sono:} \quad \ln x = \frac{\lg x}{\lg \mathrm{e}} \qquad \lg x = (\lg \mathrm{e}) \ln x\Big).$$

BOZZA - DRAFT

Esempio$_\mu$

$$\ln 2 \approx \frac{\lg 2}{0.4343} \approx \frac{0.3}{0.4343} \approx 0.69$$

costante spesso approssimata a 0.7 in Farmacologia nella trattazione dell'emivita dei farmaci, come vedremo in un paragrafo successivo.

16.2 Primo passo per il calcolo dei logaritmi

Dato un numero positivo x, lo si esprima in notazione scientifica

$$x = t \cdot 10^n \qquad 1 \le t < 10$$

e allora vale

$$\lg x = \lg t + n \qquad 1 \le t < 10 \tag{36}$$

e allora ci basta saper calcolare approssimatamente il logaritmo decimale dei numeri fra 1 e 10.

Per esempio

$$\lg 2022 = \lg\big(2.022 \cdot 10^3\big) = 3 + \lg 2.022$$

[79] Entrambe le ultime 2 formule hanno circa $\frac{1}{79\,000}$ di errore relativo.

16.3 Calcolare i logaritmi decimali dei numeri da 1 a 10

Considereremo 9 metodi:

1) con questa formula

$$\lg t \approx \frac{\sqrt{\sqrt{t}-1}}{\frac{\sqrt{t}}{4}+0.33} \qquad 1 \le t < 10 \qquad errore^{(80)} < 1\% \qquad (37)$$

per esempio con la sezione aurea viene $\lg \varphi \approx 0.209$ (si verifichi)

2) con artifici come quelli usati per i logaritmi di 3, 4, 5...

3) con una calcolatrice scientifica reale

4) con un'app di calcolatrice scientifica virtuale sul telefonino

5) online con WolframAlpha con `Log[`*base*`,`*numero*`]`

6) con le tavole cartacee (non sono affatto obsolete, anzi sono utili perché mostrano i valori intorno ad uno cercato, cosa che la calcolatrice non fa)

7) con le *logarithm tables* in rete (come sopra)

8) con altre approssimazioni[81]

9) con l'obsoleto *regolo calcolatore*. "gli astronauti (...) si servivano di regoli calcolatori (...) durante le missioni Apollo", da Wikipedia, l'enciclopedia libera, alla voce *Regolo calcolatore*.

BOZZA DRAFT

Per i logaritmi in altre basi si può calcolare il logaritmo decimale con le stesse formule (36) e (37) e poi usare la formula di cambiamento di base $\log_b x = \frac{\lg x}{\lg b}$.

Esercizio$_{\mu 2019}$
Di quanto aumenta il pH se la concentrazione degli H^+ diminuisce del 45%?

[80] Errore percentuale rispetto l'esatto.

[81] Un'approssimazione calcolabile con le sole 4 operazioni e la radice quadrata è

$$\log_{10} x \approx (x^{1/256} - 1) \times 111$$

che su una calcolatrice che operi con almeno 7 cifre (decimali) significative dovrebbe avere non più di 0.004 di errore relativo (cioè 0.4%) e assoluto per $1.1 \le x \le 11$. Naturalmente $x^{1/256}$, si ottiene estraendo 8 volte consecutive la radice quadrata. Per i numeri $y < 1.1$ o $y > 11$, li si moltiplichi per un adeguato 10^m con $m \in \mathbb{Z}$ affinché l'ottenuto $y = 10^m x$ ricada in $[1.1, 11]$ e poi

$$\log_{10} y = \log_{10}(10^m x) = m + \log_{10} x$$

per esempio $\log_{10} 1265 = \log_{10} 1.265 + 3$.

Svolgimento.
Verrà usato lo standard del punto decimale. (Ma si potrebbe usare quello della virgola decimale, a scelta).

Diminuire del 45% significa moltiplicare per 0.55:

$$\mathrm{pH}_{new} =$$

$$= \mathrm{pH}(y \cdot 0.55) = -\lg(y \cdot 0.55) = -\lg y - \lg 0.55 =$$

$$= \mathrm{pH}(y) - \lg(5.5 \cdot 10^{-1}) = \mathrm{pH}_{old} - (-1 + \lg 5.5) \approx$$

con la formula approssimata (37) di cui sopra

$$\approx \mathrm{pH}_{old} - (-1 + 0.745) = \mathrm{pH}_{old} - (-0.255) =$$

$$\approx \mathrm{pH}_{old} + 0.26$$

cioè

il pH aumenta di ≈ 0.26

(Con la calcolatrice ≈ 0.259637).

BOZZA - DRAFT

16.4 Risoluzione pratica delle equazioni con exp e log

Nella pratica le equazioni con esponenziali e logaritmi si possono risolvere abbastanza meccanicamente applicando di passaggio in passaggio regole formali, purché le equazioni siano sufficientemente semplici.
Ne vedremo adesso alcune.

Consideraiamo un'equazione

$$A = B$$

in cui almeno 1 delle espressioni A e B contiene l'incognita, che chiameremo x, e almeno 1 delle 2 espressioni contenga logaritmi o esponenziali.

(1) – ***Calcolo effettivo del logaritmo*** con la sua definizione: *l'esponente da dare alla base per avere l'argomento.* Per esempio

$$2\,x - \log_4 64 = 0$$

$$\Rightarrow 2\,x - 3 = 0$$

$$\Rightarrow x = \frac{3}{2}$$

(2) – ***Passaggio al logaritmo*** in base 10 oppure e oppure altra base, di entrambi i membri dell'equazione. Per esempio

$$2^{3\,x} = 8 \quad \Big/ \log_2$$

$$\log_2(2^{3\,x}) = \log_2 8$$

$$3\,x = 3$$

$$x = 1$$

(3) – ***Sostituzione approssimata di* lg2 *con 0,3***. Per esempio

$$10^x = 2 \quad \Big/ \lg$$

$$x = \lg 2$$

$$\Rightarrow x \approx 0{,}3 \quad (\textit{standard della virgola decimale})$$

(4) – ***Sostituzione approssimata di* lg e *con 0,4343***. Per esempio

$$10^x = \mathrm{e}$$

$$x = \lg \mathrm{e} \approx 0{,}4343 \quad (\textit{standard della virgola decimale})$$

(5) – ***Calcolo approssimato del logaritmo*** con uno dei metodi visti. Per esempio

$$10^x = 500 \quad \Big/ \lg$$

$$x = \lg 500$$

$$= \lg \frac{1000}{2} =$$

$$= \lg 1000 - \lg 2 =$$

BOZZA - DRAFT

$$= 3 - \lg 2 \approx$$

$$\approx 3 - 0.3$$

$$\Rightarrow x \approx 2{,}7 \quad (\textit{standard della virgola decimale})$$

Adesso vedremo su un esempio altri 6 possibili passaggi.

Consideriamo questa equazione:

$$2\,(x^2+1)\,\ln(2\,x)^2 + (x^2+1)\,\lg x^4 + x\,\ln\frac{1}{|x|} + \frac{2}{3}\,\ln|x|^x + x\,\ln\sqrt[3]{|x|} = 0 \tag{38}$$

Ecco ora altri 6 possibili passaggi formali fattibili, da scegliere accuratamente.

(6) – ***Logaritmo di una radice***

$$2\,(x^2+1)\,\ln(2\,x)^2 + (x^2+1)\,\lg x^4 + x\,\ln\frac{1}{|x|} + \frac{2}{3}\,\ln|x|^x + x\cdot\frac{1}{3}\,\ln|x| = 0$$

(7) – ***Logaritmo del reciproco***[82]

$$2\,(x^2+1)\,\ln(2\,x)^2 + (x^2+1)\,\lg x^4 - x\,\ln|x| + \frac{2}{3}\,\ln|x|^x + x\cdot\frac{1}{3}\,\ln|x| = 0$$

(8) – ***"Tirare giù gli esponenti e se pari mettere il valore assoluto"***:

$$2\cdot(x^2+1)\cdot 2\cdot\ln|2\,x| + (x^2+1)\cdot 4\cdot\lg|x| - x\,\ln|x| + \frac{2}{3}\cdot x\,\ln|x| + x\cdot\frac{1}{3}\,\ln|x| = 0$$

e gli ultimi 3 addendi si elidono, e poi, con passaggio algebrico,

$$4\,(x^2+1)\,\ln|2\,x| + 4\,(x^2+1)\,\lg|x| = 0 \quad \Big/ : (4\,(x^2+1)) \neq 0$$

[82] Si potrebbe dire il passaggio del logaritmo del reciproco, non è necessario perchè può essere sostituito dall'applicazione di (1) e (9):

$$\mathrm{LOG}\frac{1}{z} =^{(9)} \mathrm{LOG}(1) - \mathrm{LOG}(z) =^{(1)} 0 - \mathrm{LOG}(z) = -\mathrm{LOG}(z)$$

BOZZA - DRAFT

$$\ln|2x| + \lg|x| = 0 \ \ (segue\ un\ passaggio\ algebrico)$$

$$\ln(|2| \cdot |x|) + \lg|x| = 0 \ \ (segue\ un\ passaggio\ algebrico)$$

$$\Rightarrow \ln(2|x|) + \lg|x| = 0 \tag{39}$$

(9) – ***Passare dalla base e a 10 o viceversa***, approssimatamente con la $\lg a \approx 0{,}4343 \ln x$ (oppure esattamente con la $\lg a = (\lg \mathrm{e}) \ln a$, che comunque non ci proponiamo di imparare a memoria):

$$\Rightarrow \ln(2|x|) + (\lg \mathrm{e}) \ln|x| = 0 \tag{40}$$

(10) – ***Logaritmo del prodotto***:

$$\Rightarrow \ln 2 + \ln|x| + (\lg \mathrm{e}) \ln|x| = 0 \tag{41}$$

$$\ln 2 + \ln|x| + (\lg e) \ln|x| = 0 \quad \Big/ + (-\ln 2) \quad (passaggio\ algebrico)$$

$$\ln|x| + (\lg \mathrm{e}) \ln|x| = -\ln 2 \quad (segue\ un\ passaggio\ algebrico)$$

$$(1 + \lg e) \ln|x| = -\ln 2 \quad \Big/ : (1 + \lg \mathrm{e}) \neq 0 \quad (passaggio\ algebrico)$$

$$\ln|x| = \frac{-\ln 2}{1 + \lg \mathrm{e}} \tag{42}$$

(11) – ***Esponenziazione*** in base 10 oppure e oppure altra base:

$$\ln|x| = \frac{-\ln 2}{1 + \lg \mathrm{e}} \quad \Big/ \exp$$

$$|x| = \exp\left(\frac{-\ln 2}{1 + \lg \mathrm{e}}\right) \quad (segue\ un\ passaggio\ algebrico)$$

$$\boxed{x = \pm \exp\left(\tfrac{-\ln 2}{1+\lg e}\right)}$$

Altre 4 possibili tecniche risolutive
che comunque non ci proponiamo di imparare a memoria

BOZZA - DRAFT

(12) **Formule di cambiamento di base**

$$\log_b x = \frac{\log_c x}{\log_c b} \quad \text{(83) (84) in particolare} \quad = \frac{\lg x}{\lg b} = \frac{\ln x}{\ln b}$$

da cui $\ln x = \frac{\lg x}{\lg \mathrm{e}}$ e con la (26) si ottiene la formula approssimata vista per la conversione fra logaritmi naturali e decimali.

(13) – ***"Unicità dell'intersezione dei grafici di*** a^t ***e*** b^t ***"***.

$$(\forall\, a > 0,\, b > 0,\, a \neq b) \qquad a^t = b^t \quad \Rightarrow \quad t = 0$$

Per esempio:

$$\mathrm{e}^{x^3} = 10^{x^3}$$
$$x^3 = 0$$
$$x = 0$$

(14) – ***"Unicità dell'intersezione dei grafici di*** $\log_a t$ ***e*** $\log_b t$ ***"***.

$$(\forall\, a > 0,\, b > 0,\, a \neq 1,\, b \neq 1,\, a \neq b) \qquad \log_a t = \log_b t \quad \Rightarrow \quad t = 1$$

(15) **Una formula di apparenza sorprendente** che raramente si trova:

$$x^{\mathrm{LOG}(y)} = y^{\mathrm{LOG}(x)} \qquad \forall\, x, y > 0$$

in particolare per il logaritmo naturale

$$\boxed{x^{\ln y} = y^{\ln x} \qquad \forall\, x, y > 0}$$

(che si dimostra – sostanzialmente – applicando ad ambo i membri il logaritmo nella base considerata; e si veda che WolframAlpha la dichiara `True`).

BOZZA - DRAFT

Bottom line. *I passaggi visti sono sufficienti per risolvere molte equazioni con esponenziali e logaritmi, unitamente all'algebra:*

– calcolo (a mano o con la calcolatrice o in altri modi) e proprietà delle 4 operazioni e del valore assoluto e del fattoriale e delle radici e delle potenze;

[83] Anche $= (\log_c x) \cdot (\log_b c)$.

[84] Per il lettore interessato, si ricavano subito anche $\log_b x = \frac{1}{\log_x b}$, e $\log_{1/b} x = -\log_b x$.

– sostituzione (coi loro valori approssimati) delle costanti conosciute (in forma approssimata) a memoria, come e, π, eccetera.

Ovviamente tutto ciò non basta per risolvere tutte le equazioni con gli esponenziali e/o i logaritmi: per esempio non la $e^x = x + 1$, che comunque ovviamente ha unica soluzione $x = 0$. (Si disegnino i 2 grafici, di $f(x) := e^x$ e di $g(x) := x+1$).

Ma si veda ancora il paragrafo seguente per una possibile problematica.

16.5 Soluzioni spurie ovvero fittizie

In qualche equazione coi logaritmi per così dire "capricciosa" può capitare che qualche passaggio incauto introduca soluzioni spurie ovvero fittizie, che assolutamente non sono soluzioni ma sembrano esserlo. Al livello di questa trattazione elementare non è possibile dare un metodo generale per evitare sempre e comunque le soluzioni fittizie.

BOZZA - DRAFT

Per esempio con questi passaggi

$$\lg x + \lg(2\,x) = 0$$

$$\lg(x \cdot 2 \cdot x) = 0$$

$$\lg(2\,x^2) = 0 \quad \Rightarrow 2\,x^2 = 1 \quad \Rightarrow x^2 = \frac{1}{2} \quad \Rightarrow x = \pm\sqrt{\frac{1}{2}}$$

troviamo una soluzione spuria ovvero fittizia negativa che assolutamente non è soluzione dell'equazione iniziale perchè lg non esiste per argomenti negativi.

Al limite estremo, in un'equazione particolarmente capricciosa anche la banale sostituzione di lg2 con l'approssimazione 0,3 potrebbe introdurre soluzioni fittizie, o di π con 3,14.
Molta parte della problematica si può evitare verificando se le soluzioni trovate verificano l'equazione iniziale, ed

eventualmente quindi escludendone qualcuna.
Oppure – e può essere meno facile – trovando il dominio dell'equazione, ed escludendo le eventuali soluzioni al di fuori del dominio.

BOZZA - DRAFT

* Complementi *

16.6 Equazioni e disequazioni con exp e log – teoria

L'esponenziale in ogni base positiva diversa da 1 è iniettivo e allora si può applicarlo ad ambo i membri di un'equazione ottenendo un'equazione *equivalente*, cioè con le stesse (eventualmente esistenti) soluzioni:

$$f(x) = g(x) \Leftrightarrow 10^{f(x)} = 10^{g(x)}$$

$$f(x) = g(x) \Leftrightarrow \mathrm{e}^{f(x)} = \mathrm{e}^{g(x)}$$

$$f(x) = g(x) \Leftrightarrow b^{f(x)} = b^{g(x)} \qquad \forall\, b > 0,\, b \neq 1$$

BOZZA - DRAFT

Si può anche applicare l'esponenziale in base b positivo diverso da 1 a tutti e 4 i tipi di disuguaglianza e l'ordinamento si conserva se e solo se $b > 1$, che è il caso di maggior interesse:

$$f(x) > g(x) \Leftrightarrow 10^{f(x)} > 10^{g(x)}$$

$$f(x) > g(x) \Leftrightarrow \mathrm{e}^{f(x)} > \mathrm{e}^{g(x)}$$

$$f(x) > g(x) \Leftrightarrow b^{f(x)} > b^{g(x)} \qquad \forall\, b > 1$$

e tutte similmente con $<$, $\geq$, $\leq$.

Il logaritmo è iniettivo e definito per ogni numero positivo e allora

$$0 < f(x) = g(x) \Leftrightarrow \lg f(x) = \lg g(x)$$

$$0 < f(x) = g(x) \Leftrightarrow \ln f(x) = \ln g(x)$$

$$0 < f(x) = g(x) \Leftrightarrow \log_b f(x) = \log_b g(x)$$

cioè si può applicare il logaritmo in qualunque base ad ambo i membri di un'equazione $f(x) = g(x)$ ottenendo un'equazione *equivalente*, cioè con le stesse (eventualmente esistenti) soluzioni, se

$f(x) > 0$ nel suo dominio, oppure $g(x) > 0$ nel suo dominio (è equivalente), in particolare se una delle 2 funzioni è un esponenziale in qualche base eventualmente moltiplicato per una costante positiva, che è il caso più comune nella pratica.

Di più: il logaritmo decimale e quello naturale e ogni logaritmo in base $b > 1$ sono crescenti[85] e allora si può applicarli conservando l'ordinamento:

$$0 < f(x) \le g(x) \iff \lg f(x) \le \lg g(x) \quad e\, anche\, con \quad <$$

$$0 < f(x) \le g(x) \iff \ln f(x) \le \ln g(x) \quad e\, anche\, con \quad <$$

$$\forall\, b > 1 \quad 0 < f(x) \le g(x) \iff \log_b f(x) \le \log_b g(x) \quad e\, anche\, con \quad <$$

BOZZA - DRAFT

[85] Invece coi logaritmi in base minore di 1 si inverte l'ordinamento:

$$(\forall\, b,\, 0 < b < 1) \quad 0 < f(x) < g(x) \iff \log f(x) > \log g(x)$$

$$(\forall\, b,\, 0 < b < 1) \quad 0 < f(x) \le g(x) \iff \log f(x) \ge \log g(x)\,.$$

ESERCIZI SULLA LEZIONE 13

ESERCIZIO$_{\mu}$ $\approx$

Risolvere

$$\lg \ln x^2 = 0$$

SVOLGIMENTO

Verrà usato lo standard della virgola decimale. (Ma si potrebbe usare quello del punto decimale, a scelta).

$$\lg \ln x^2 = 0 \quad \big/ 10^{\wedge}$$

(passaggio che non introduce soluzioni spurie ovvero fittizie)

$$\ln x^2 = 10^0$$

$$\ln x^2 = 1 \quad \big/ \exp$$

(passaggio che non introduce soluzioni spurie ovvero fittizie)

$$x^2 = \mathrm{e}$$

$$x = \pm\sqrt{\mathrm{e}}$$

e ricordando che e $\approx 2{,}718$

$$\boxed{\approx \pm 1{,}649}$$

BOZZA - DRAFT

Nota 1. In qualche equazione coi logaritmi per così dire "capricciosa" può capitare che qualche passaggio incauto introduca soluzioni spurie ovvero fittizie, che assolutamente non sono soluzioni ma sembrano esserlo. Per esempio con questi passaggi

$$\lg x + \lg(2\,x) = 0$$

$$\lg(x \cdot 2 \cdot x) = 0$$

$$\lg(2\,x^2) = 0 \quad \Rightarrow 2\,x^2 = 1 \quad \Rightarrow x^2 = \frac{1}{2} \quad \Rightarrow x = \pm\sqrt{\frac{1}{2}}$$

troviamo una soluzione spuria ovvero fittizia negativa che assolutamente non è soluzione dell'equazione iniziale perchè lg non esiste per argomenti negativi. Ma i passaggi svolti per risolvere l'esercizio iniziale, $\lg \ln x^2 = 0$, sono di quelli "sicuri", che non introducono soluzioni spurie ovvero fittizie.

Nota 2. Non è una buona idea cercare di risolvere l'equazione con la proprietà del logaritmo della potenza

$$(\forall\, \alpha \in \mathbb{R})(\forall\, x > 0)\ \mathrm{LOG}(x^\alpha) = \alpha\, \mathrm{LOG}(x)$$

(ove LOG indica 2 volte 1 stesso logaritmo in qualunque base) perchè essa, come si vede scritto, vale per $x > 0$, e di fatto, incautamente applicandola in questo esercizio, fa perdere la soluzione negativa. (Per x negativo la formula $\mathrm{LOG}(x^\alpha) = \alpha\, \mathrm{LOG}(x)$ non ha senso).
Piuttosto, si potrebbe applicare la proprietà degli esponenti pari, aggiungendo il valore assoluto.

BOZZA - DRAFT

17 Esponenziali e logaritmi – IV parte

17.1 L'esponenziale e la campana gaussiana

Nelle Scienze Applicate un'infinità di fenomeni vengono bene modellizzati da funzioni con grafico più o meno a campana.

Talvolta queste campane sono simmetriche, o almeno più o meno simmetriche, altre volte sono lontane dalla simmetria.

Occupiamoci adesso solo delle campane simmetriche.

POZZA - DRAFT

La più "pura" delle campane simmetriche, diciamo il modello ideale in qualche modo, è la campana gaussiana, che ha un'equazione basata sull'esponenziale

$$y = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$

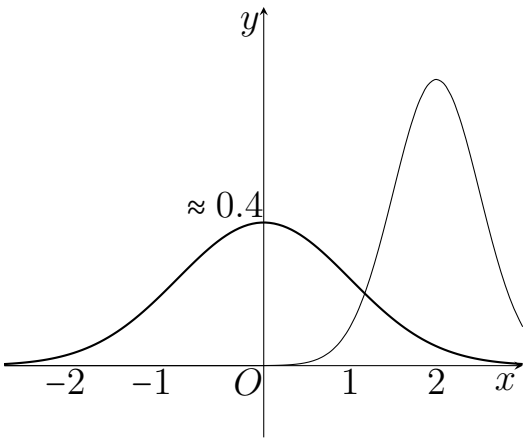
chiamata *densità normale standard*.
Ha asse di simmetria $x = 0$, cioè l'asse y, e l'asse di simmetria può venir spostato in $x = m$ qualsiasi, con l'equazione

$$y = \frac{e^{-\frac{(x-m)^2}{2}}}{\sqrt{2\pi}}$$

ma avvertiamo il Lettore che invece della lettera m di solito si usa la corrispondente lettera greca μ.
Può essere resa più o meno ribassata ovvero "ripida" con un ulteriore parametro $\sigma > 0$ ottenendosi l'equazione generale

$$y = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\,\sigma}$$

chiamata *densità normale di parametri* μ *e* σ.



– Densità normale di parametri $\mu = 0$ e $\sigma = 1$ ovvero densità normale standard

– densità normale di parametri $\mu = 2$ e $\sigma = 0.5$

Grafici di questo genere si otterranno, fra l'altro, con le altezze, e le glicemie, colesterolemie, sideremie... degli Italiani.
E non solo degli italiani: anche degli europei, e anche solo degli Studenti di Farmacia.

BOZZA - DRAFT

Curve a campana alquanto asimmetriche modellizzano bene il numero di morti (o casi o terapie intensive...) giornaliero (o settimanale o anche mensile) delle ondate epidemiche.
Equazioni per modellizzarle sono ancora basate sull'esponenziale.

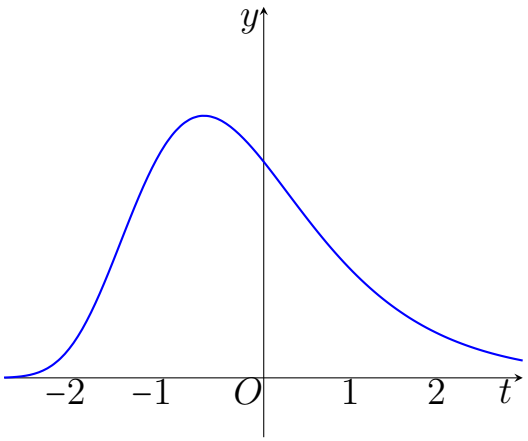


Grafico di

$$10\,\mathrm{e}^{-x-0.5\,\mathrm{e}^{-x}}$$

Il tempo t potrebbe essere in mesi, e lo 0 all'ultimo giorno di un certo anno.

17.2 Funzione di Gompertz

Si chiama funzione di Gompertz ogni funzione della famiglia (insieme) di funzioni

$$f(x) := a\,\mathrm{e}^{-b\mathrm{e}^{-c\,t}}$$

I grafici di queste funzioni sono sigmoidi. Si sono rivelate ottime per la modellizzazione delle epidemie: numero cumulativo dei morti, dei casi, eccetera.

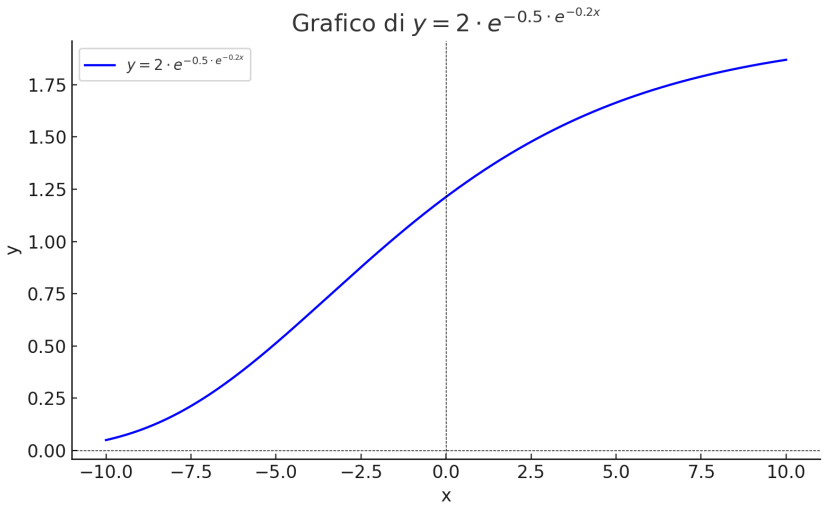


Figure 21: Grafico di $y = 2 \cdot e^{-0.5 \cdot e^{-0.2x}}$. (By Vispo Volpino with ChatGPT).

17.3 Il logaritmo decimale come contatore di zeri

Logaritmi decimali delle potenze di 10:

−3 0.001 tre zeri ma con punto decimale
−2 0.01
−1 0.1
0 1 zero zeri
1 10
2 100
3 1000 tre zeri

17.4 Ordine di grandezza di un numero positivo

L'ordine di grandezza è un concetto tanto utile nelle Scienze Applicate, quanto mal definito nelle stesse Scienze Applicate, come anche altre volte succede (per esempio con le cifre significative).
Qua facciamo una scelta definizionale.

Ci sono 2 questioni:
– definire l'ordine di grandezza di 1 numero positivo;
– dire quando 2 numeri positivi hanno LO STESSO ordine di grandezza.

Per la seconda questione, l'uso generale è dire che 2 numeri positivi hanno LO STESSO ordine di grandezza se il rapporto del più grande e il più piccolo sta fra 1 e 10. (Non si dichiara quale sia l'ordine di grandezza, si dice solo che è lo stesso).
(Ma, dopo che avremo definito l'ordine di grandezza, numeri con rapporto 2 o anche 1.1 potranno non avere UGUALE ordine di grandezza, per esempio 95 e 103 ☹ comunque la consolazione è che i 2 ordini di grandezza non distano mai più di 1).

BOZZA - DRAFT

Come abbiamo già detto, ogni numero positivo x ammette una scrittura

$$x = a \cdot 10^n$$

con $1 \le a < 10$ e n intero: ...−3, −2, −1, 0, 1, 2, 3...
Adesso chiamiamo n ordine di grandezza di x e aggiungiamo:

$$\boxed{\textbf{ordine di grandezza } \text{di } x: \quad n := \lfloor \lg x \rfloor} \tag{43}$$

cioè

$$\textit{ordine di grandezza di } x \;=\; \textit{parte intera di } \lg x$$

Ma altri Autori definiscono con condizioni su a diverse da $1 \le a < 10$.
(Wikipedia in inglese ne dà altre 2).
Qualunque definizione – ragionevole – di ordine di grandezza si usi per 2 numeri positivi

se hanno ugual ordine di grandezza

il massimo non supera il decuplo del minimo

cioè sono simili almeno molto molto vagamente, e

se hanno ordini di grandezza m e $m+1$
il massimo non supera il centuplo del minimo
(ma in generale sarà molto molto vagamente il decuplo)

5 è l'ordine di grandezza del numero di capelli di una persona
11 è l'ordine di grandezza del numero di neuroni di un uomo
11 è l'ordine di grandezza del numero di stelle della Via Lattea
13 è l'ordine di grandezza del numero di globuli rossi di un uomo (3.0×10^{13})
Quindi il numero di globuli rossi di un uomo è molto vagamente il centuplo dei suoi neuroni (2 ordini di grandezza di differenza) e del numero di stelle della Via Lattea.

BOZZA DRAFT

L'ordine di grandezza è molto usato nelle Scienze applicate per fissare le idee sulla grandezza approssimativa di un numero, per agili confronti, semplificandolo all'estremo in un singolo numero intero, spesso fra −20 e 20 (e quasi sempre fra −30 e 30).
Ma la grossolanità dell'approssimazione è mostrata dal fatto che

9 è l'ordine di grandezza del numero di persone del mondo

come se fossero un miliardo e invece sono circa 8 miliardi.

Magari in novembre l'amministrazione dell'università non conosce ancora il numero esatto di quelli che saranno iscritti a Farmacia e Medicina ma può considerare plausibile che i secondi saranno in numero di un ordine di grandezza superiore ai primi.
Questo tipo di grossolanissime approssimazioni sono utilissime nella gestione pratica della realtà.

Orders of Magnitude in Biology

Fill in the blanks in the table using the information given.

Nanometre (nm)	1000000 x smaller than a millimetre
Micrometre (µm)	1000000 x smaller than a metre
Millimetre (mm)	1000 x smaller than a metre
Metre (m)	
Kilometre (km)	1000 x bigger than a metre

$10^2 = 10 \times 10$ This is the same as moving the decimal place one place to the right. 10.0 → 100.0

$10^3 = 10 \times 10 \times 10$ This is the same as moving the decimal place two places to the right. 10.0 → 1000.0

$10^{-1} = 1 \div 10$ This is the same as moving the decimal place one place to the left. 1.0 → 0.1

$10^{-4} = 1 \div 10 \div 10 \div 10 \div 10$ This is the same as moving the decimal place four places to the left. 1.0 → 0.0001

Atom	DNA	Virus	Prokaryotic cell	Eukaryotic cell	Flea	Woodlouse	Peach	Child	Whale	Giant Redwood	School Drive
0.00000000 1m	0.00000 001m	0.0000001m		0.00001-0.0001m			0.1m	1m	10m		
			10^{-6}m			10^{-3}m		1m	10m	10^{3}m	
		100nm		10-100µm	1mm		10cm	1m	10m		1km

Creative-Commons-CC-BY-SA image retrieved (2024) in

https://physics-fiction.fandom.com/wiki/Scale_%26_Magnitude_Order?file=Order_of_Magnitude.png

BOZZA

17.5 La successione di Fibonacci e gli esponenziali

La successione di Fibonacci, che modellizza l'espansione almeno – nelle fasi iniziali – di una popolazione animale o microbica, e le epidemie, è sì definita per ricorrenza (e solo su $\mathbb{N}$) ma è approssimata con una funzione esponenziale in base φ, la sezione aurea:

$$a_n \approx \frac{1}{\sqrt{5}}\,\varphi^n \text{ per } n \text{ sufficientemente grande} \tag{44}$$

per esempio $a_{10} = 55 \approx 55.0036$.
Dopo quanti mesi (se n indicia i mesi) la popolazione raggiunge il valore 135 301 852 344 706 746 049, oppure anzi, uscendo dall'esattezza ideale del problema, un valore $1.35 \cdot 10^{20}$?

Abbiamo l'equaziona approssimata

$$\frac{1}{\sqrt{5}}\,\varphi^n \approx 1.35 \cdot 10^{20} \qquad / \lg$$

$$\lg \frac{1}{\sqrt{5}} + n \lg \frac{1+\sqrt{5}}{2} \approx \lg(1.35 \cdot 10^{20}) \qquad \Big/ + \lg\sqrt{5}$$

$$n \lg \frac{1+\sqrt{5}}{2} \approx \lg(1.35 \cdot 10^{20}) + \lg\sqrt{5} \qquad \Big/ : \lg \frac{1+\sqrt{5}}{2}$$

$$\approx \frac{\lg(1.35 \cdot 10^{20}) + \lg\sqrt{5}}{\lg \frac{1+\sqrt{5}}{2}}$$

$$\approx \frac{\lg(1.35 \cdot 10^{20}) + \lg 2.23607}{\lg 1.61803}$$

e insomma ci servono i valori (approssimati) dei logaritmi decimali di 3 numeri (fra 1 e 10).
Con l'approssimazione prima vista si ha

$$\approx \frac{20 + 0.130 + 0.35054}{0.2089} \approx 98.04$$

e calcolando con pazienza si troverà che il 98-esimo numero di Fibonacci è proprio 135 301 852 344 706 746 049. (Oppure con WolframAlpha: `Fibonacci[98]`).
Nella figura a sinistra si vedono:

(blu) la successione (n, a_n) essendo a_n la successione di Fibonacci
(verde) $\frac{1}{\sqrt{5}}\,\varphi^x$
(magenta) e^x
(nero) 10^x

BOZZA - DRAFT

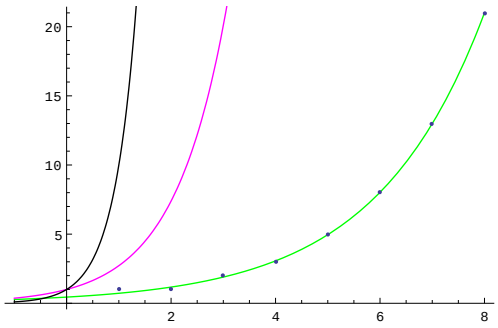


Figura A

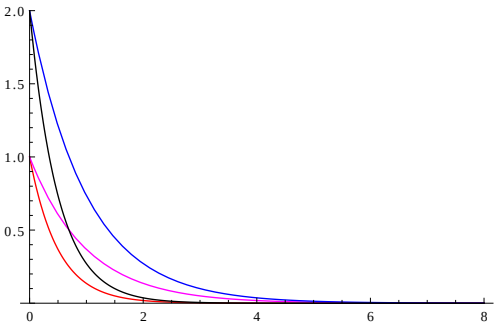


Figura B

Nella figura B, a destra, si vedono grafici di funzioni del tipo $a\,\mathrm{e}^{ct}$: questo tipo di funzioni internazionalmente vengono talvolta chiamate impropriamente *exponential functions* anche se $a \neq 1$.

BOZZA_DRAFT

17.6 Emivita di una sostanza nel sangue o nelle ossa

Leggiamo in *Emivita (farmacologia)*:

> L'emivita ($t_{1/2}$) è un parametro farmacocinetico che indica il tempo richiesto per ridurre del 50% la concentrazione plasmatica di un farmaco (...) è inversamente proporzionale alla sua clearance (Cl) e direttamente proporzionale al suo volume di distribuzione (V_d) secondo la formula
>
> $$t_{1/2} = 0.7\frac{V_d}{Cl}$$
>
> dove la costante 0,7 è in realtà un'approssimazione del logaritmo naturale di 2 (in effetti, più in particolare $\ln(2) = 0,69...$).

Per curiosità elenchiamo[86] senza qua dare garanzie alcune emivite di farmaci:

noradrenalina 2 minuti
morfina 2-3 ore
donepezil 3 giorni (70 ore)
bedaquilina 165 giorni

e di questo elemento, che non è un farmaco:

cadmio nelle ossa 30 anni. (Pile ricaricabili, colori per artisti, eccetera).

Il piombo nelle ossa ha un'emivita – supponiamo esponenziale il suo scomparire, com'è ragionevole in questi casi – di (circa) 10 anni. In quanto tempo si riduce del 75%? E del 90%? E del 95%? E del 97%?

[86] https://en.wikipedia.org/wiki/Biological_half-life

La riduzione del 75% è la riduzione al 25%, cioè ai $\frac{25}{100}$ ovvero $\frac{1}{4}$ e allora corrisponde a 2 dimezzamenti: allora ci vogliono 20 anni.
Per il 90% è meno immediato.
Illustriamo una famiglia (insieme) di funzioni con l'esponenziale, che ricorre amplissimamente nelle Scienze Applicate, ed è espressa dai vari Autori in 2 modi diversi, con una costante negativa (che qua sarà k) oppure con una costante positiva (che qua sarà r) preceduta dal segno meno:

$$u(t) = u_0\,\mathrm{e}^{k\,t} \quad \text{essendo } u_0 = u(0) \qquad k < 0$$

$$u(t) = u_0\,\mathrm{e}^{-r\,t} \quad \text{essendo } u_0 = u(0) \qquad r > 0$$

e la costante k ovvero r dipende da caso a caso.
Questa è l'equazione che esprime in funzione del tempo la concentrazione di piombo, che si attenua; e un'infinità di altri fenomeni.
Nella figura B, precedente, a destra, si vedono 4 casi, con diversi valori di u_0 e r:

(rosso) $\mathrm{e}^{-2\,x}$
(magenta) e^{-x}
(nero) $2\,\mathrm{e}^{-2\,x}$
(blu) $2\,\mathrm{e}^{-x}$

E in questo caso possiamo calcolare la costante r dall'emivita di 10 anni:

$$u(emivita) =^{DEF} \frac{u(0)}{2} =^{EQ} u(0)\,\mathrm{e}^{-r\,10\,anni} \quad \Big/ : u(0)$$

$$\frac{1}{2} = \mathrm{e}^{-r\,10\,anni} \quad \Big/ \ln$$

$$\ln\frac{1}{2} = -r\,10\,anni$$

$$-\ln 2 = -r\,10\,anni \quad \Big/ : (-10\,anni)$$

$$r = \frac{\ln 2}{10\,anni}$$

ottenendosi così l'equazione dello scomparire del piombo nelle ossa

$$u(t) = u_0\,\mathrm{e}^{-\frac{\ln 2}{10\,anni}\,t} \quad \text{essendo } u_0 = u(0)$$

BOZZA - DRAFT

Ora ci si chiede quando sarà ridotto del 90% ovvero al 10%, cioè $u(t) = 0.1\,u(0)$:

$$0.1\,u(0) \overset{EQ}{=} u(0)\,\mathrm{e}^{-\frac{\ln 2}{10\,anni}\,t} \quad \Big/ : u(0)$$

$$0.1 = \mathrm{e}^{-\frac{\ln 2}{10\,anni}\,t} \quad \Big/ \lg$$

$$-1 = \lg \mathrm{e}^{-\frac{\ln 2}{10\,anni}\,t}$$

$$-1 = -\frac{\ln 2}{10\,anni}\,t\,\lg \mathrm{e}$$

$$t = \frac{10\,anni}{\lg \mathrm{e} \cdot \ln 2} \approx \frac{10\,anni}{\lg \mathrm{e} \cdot \frac{\lg 2}{\lg \mathrm{e}}} = \frac{10\,anni}{\lg 2} \approx \frac{10\,anni}{0.3}$$

≈ 33.3 anni (valore che, nella pratica, possiamo stare sicuri che verrà riportato come 33 anni).

Si risolvano gli altri quesiti posti associatamente a questo appena risolto. Con 95% si dovrà calcolare $\lg 0.05 = \lg\left(\frac{1}{2} \cdot 0.1\right) = \lg \frac{1}{2} + \lg \frac{1}{10} = -\lg 2 - \lg 10 \approx -0.3 - 1 = -1.3$.

BOZZA - DRAFT

17.7 L'intensità sonora: logaritmi e decibel

In farmacia misuriamo l'udito... capendo l'intensità sonora.

L'intensità sonora in decibel è definita con un logaritmo decimale, in dipendenza dall'energia sonora, e una serie di calcoli[87] porta a concludere che

ad un raddoppio dell'energia corrisponde un aumento di ≈ 3 decibel (45)

unica cosa che ci proponiamo di imparare di questo argomento. Il 3 viene, coi calcoli, dallo 0.3 dell'approssimazione di lg2.

BOZZA - DRAFT

[87] Per il Lettore interessato: l'intensità sonora in decibel è definita sostanzialmente come

$$\text{intensità sonora} := 10\Big(\lg \frac{\text{energia passante in 1 s per 1 m}^2}{\text{una prefissata energia passante in 1 s per 1 m}^2}\Big)\, dB$$

dove la prefissata energia passante in 1 s per un m^2 è la (piccolissima) $1pW/m^2$.
(L'argomento dei logaritmi e degli esponenziali è sempre adimensionale, cioè un numero puro, senza unità di misura: non esiste il logaritmo di 3 metri o 4 secondi).
Allora come funzione dell'energia è del tipo

$$a \lg \frac{x}{c} \quad = a\,(\lg x - \lg c) \quad = a \lg x + d \quad (\text{con } d := -a \lg c)$$

(e si faccia ben attenzione a distinguere la soprastante scrittura dalla $a \lg(x+d)$). Il grafico è proprio come quello di $\lg x$, riscalato da a sull'asse delle ordinate e traslato verticalmente di d.

Calcoliamo di quanti decibel aumenta l'intensità sonora raddoppiando l'energia.

Detta E l'energia considerata e $2\,E$ il suo doppio, si ha

$$\text{intensità_sonora}(2E) - \text{intensità_sonora}(E) =$$

$$= 10 \lg \frac{2\,E}{1pW/m^2} - 10 \lg \frac{E}{1pW/m^2} =$$

ricordando che $\lg(x/y) = \lg x - \lg y$

$$= 10\Big(\lg(2\,E) - \lg E\Big) = \quad (\text{mentre } \lg(1pW/m^2) \text{ si elide})$$

$$= 10 \lg \frac{2\,E}{E} = 10 \lg 2 \approx 10 \cdot 0.3 = 3$$

cioè quanto detto: ad un raddoppio dell'energia corrisponde un aumento di ≈ 3 decibel.

17.8 Equazioni con logaritmi della x e x libera

Questione interessante è quella delle equazioni in cui compare un $\mathrm{LOG}(x)$ e anche una x fuori dai logaritmi. Per esempio per

$$1 - \lg x - x = 0$$

si può fare

$$1 - x = \lg x$$

e poi risolvere per via grafica il sistema

$$\begin{cases} y = 1 - x \\ y = \lg x \end{cases}$$

Ma questo tipo di equazioni possono diventare più complicate, per esempio

$$2 - \lg x - x = 0$$

La risoluzione di quest'equazione richiede l'uso di una *funzione speciale dell'analisi matematica*.
Precisamente la funzione W di Lambert, e in casi più complessi nemmeno quella basta.

BOZZA - DRAFT

Su questa funzione lo scrivente ci ha passato un bel po' di vita, e si trovano in rete articoli scientifici cercando il cognome e Lambert e function.

17.9 ESERCIZI SULLA LEZIONE 17

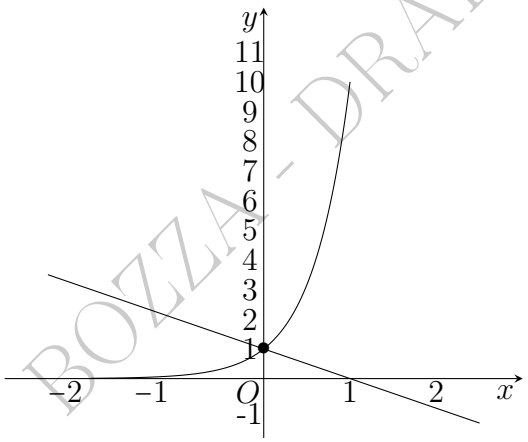
17.9.1 Esercizio risolto a – Risoluzione per via grafica

ES. 1$_{\mu 2024}$ $*$ Risolvere per via grafica il sistema di equazioni

$$\begin{cases} y = 1 - x \\ y = 10^x \end{cases}$$

(dando la soluzione nella forma $x = ...$, $y = ...$).

SVOLGIMENTO



BOZZA - DRAFT

Già bastano questi pochi (valori delle coordinate di) punti dei 2 grafici

x	$1 - x$
-2	3
-1	2
0	1
1	0
2	-1

x	10^x
-2	$\frac{1}{100}$
-1	$\frac{1}{10}$
0	1
1	10

e disegnando (vedi figura) sullo stesso piano cartesiano grafici (approssimativi), vediamo subito che c'è un punto di intersezione $(0,1)$, e non esistono altri per le proprietà elementari delle funzioni $1-x$ e 10^x, e allora la soluzione del sistema di equazioni è data dalle coordinate x e y dell'unico punto di intersezione dei grafici:

$$\boxed{x=0,\ y=1}$$

17.10 Altri sercizi su logaritmi ed esponenziali

ESERCIZIO$_{\mu 2020}$
$*\approx$ Risolvere l'equazione

$$e^{2x^3} - e^{2\lg^6(2\lg 100)} = 0$$

SVOLGIMENTO
Ovviamente $\lg 100 = 2$

$$e^{2x^3} = e^{2\lg^6(2\cdot 2)} \quad \Big/ \ln$$

$$2x^3 = 2\lg^6 4 \quad \Big/ :2$$

$$x^3 = \lg^6 4 \quad \Big/ \sqrt[3]{\cdot}$$

$$x = \lg^2 4 =$$

che è la soluzione esatta, e infine con le proprietà dei logaritmi e la nota approssimazione $\lg 2 \approx 0{,}3$

$$= \lg^2 2^2 = (\lg 2^2)^2 = (2\lg 2)^2 \approx (2\cdot 0{,}3)^2 = 0{,}6^2 = 0{,}36$$

$$\boxed{\lg^2 4 \approx 0{,}36}$$

(Un valore più preciso è 0,3625, come si può verificare per esempio online su WolframAlpha con `Log[10,4]^2`).

ESERCIZIO$_{\mu 2018}$
$*$ Si risolva quest'equazione:

$$\log_{10} x + \log_{10} 200 - \log_{10}(1+x^2) = \log_3 9$$

SVOLGIMENTO
Prima di tutto osserviamo che dev'essere

BOZZA - DRAFT

$$\underline{x > 0}$$

perché x è argomento di un logaritmo. (Ovviamente $1 + x^2 > 0$ sempre). Ricordando le $\log_b u + \log_b v = \log_b(u\,v)$ e $\log_b u - \log_b v = \log_b \frac{u}{v}$ e $3^2 = 9$

$$\log_{10} \frac{x \cdot 200}{1 + x^2} = 2 \qquad \Big/ 10^{\wedge} \quad \leftarrow \textit{potenza in base } 10\textit{; elimina il logaritmo in base } 10$$

$$\frac{200\,x}{1 + x^2} = 100 \qquad \Big/ \cdot \frac{1 + x^2}{100}$$

$$2\,x = 1 + x^2$$

$$x^2 - 2\,x + 1 = 0$$

$$(x - 1)^2 = 0$$

$$\boxed{x = 1}$$

(Che è > 0).

ESERCIZIO$_{\mu 2018}$
$*\approx$ Si risolva l'equazione

$$\log_{10} \frac{x^2}{x^2 + 10} + \log_{10} \frac{x^2 + 10}{10^2} = 1$$

SVOLGIMENTO
Prima di tutto osserviamo che deve essere

$$\underline{x \neq 0}$$

(altrimenti si avrebbe l'inesistente $\log_{10} 0$). (Altre condizioni da richiedere non ci sono perché per $x \neq 0$ gli argomenti dei logaritmi sono entrambi evidentemente positivi).
Ricordando che la somma dei logaritmi (in una stessa base) è il logaritmo (in quella stessa base) del prodotto, si ha

$$\log_{10} \Big(\frac{x^2}{x^2 + 10} \cdot \frac{x^2 + 10}{10^2} \Big) = 1$$

e semplificando

$$\log_{10} \frac{x^2}{10^2} = 1 \quad \Big/ 10^{\wedge} \quad \leftarrow \textit{potenza in base } 10\textit{; elimina il logaritmo in base } 10$$

$$\frac{x^2}{10^2} = 10^1 \quad \Big/ \cdot 10^2$$

$$x^2 = 10^2 \cdot 10^1$$

$$x = \pm\sqrt{10^2 \cdot 10} = \pm\sqrt{10^2}\,\sqrt{10}$$

e in conclusione (senza che la condizione $x \neq 0$ ci faccia escludere qualche valore)

BOZZA - DRAFT

$$\boxed{x = \pm 10\sqrt{10} \approx \pm 31.623}$$

oppure con un decimale in meno

$$\boxed{x = \pm 10\sqrt{10} \approx \pm 31.62}$$

ESERCIZIO$_{\mu}$

$*$ ≈ Risolvere l'equazione $4\lg(100\,x) = 9$.

SVOLGIMENTO

$$4\lg(100\,x) = 9 \qquad \Big/ : 4$$

$$\lg(100\,x) = \frac{9}{4}$$

e per la proprietà del logaritmo del prodotto

$$\lg 100 + \lg x = \frac{9}{4}$$

$$\lg 10^2 + \lg x = \frac{9}{4}$$

$$2 + \lg x = \frac{9}{4}$$

$$\lg x = \frac{9}{4} - 2$$

$$\lg x = \frac{9-8}{4}$$

$$\lg x = \frac{1}{4} \qquad \Big/ 10\text{^}$$

$$x = 10^{\frac{1}{4}}$$

ovvero

$$\boxed{\sqrt[4]{10} \approx 1{,}778\,28}$$

(La radice quarta si calcola con la calcolatrice come radice quadrata della radice quadrata: $\sqrt[4]{z} = \sqrt{\sqrt{z}}$)

BOZZA - DRAFT

18 Funzioni trigonometriche

18.1 Generabili nella geometria riappaiono ovunque

$m\,x+q$, x^2, $\sqrt{x}$, $\exp x$, $\ln x$ si possono definire geometricamente, poi appaiono ovunque nelle Scienze Applicate. Ugualmente avviene per le funzioni trigonometriche che andiamo a definire.

Un'infinità di problemi geometrici con **lunghezze** e **aree** relativamente a **poligoni** possono venir risolti con le funzioni lineari

$$y = m\,x + q \qquad \text{(funzione meglio detta affine)}$$

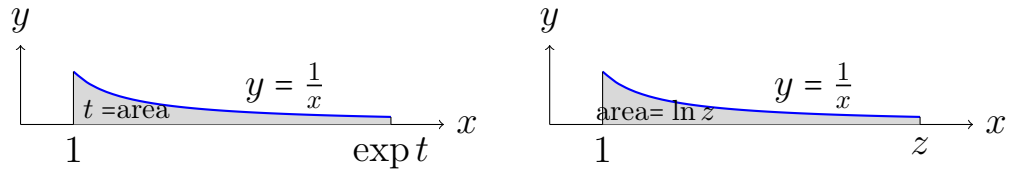
e con la radice quadrata e la sua inversa, il quadrato:

$$\sqrt{x} \qquad x^2$$

BOZZA DRAFT

Queste funzioni, del tutto costruibili nella geometria, si rivelano ottimi modelli di un'infinità di fenomeni delle Scienze Applicate.

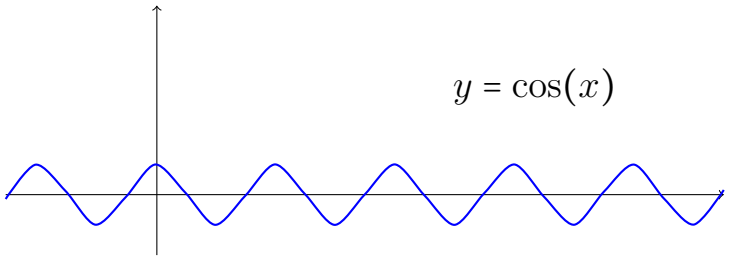
Nuove funzioni compaiono considerando **coniche** invece che poligoni, e, di nuovo, si rivelano ottimi modelli di un'infinità di fenomeni delle Scienze Applicate.
Precisamente l'**iperbole equilatera** genera (in modi non banali: integrali) l'**esponenziale** e la sua inversa, il **logaritmo** naturale:



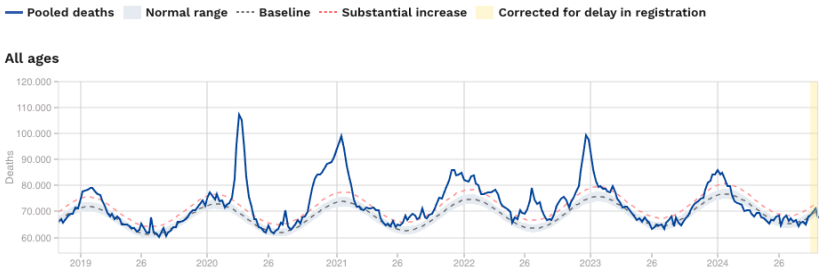
(Storicamente l'esponenziale e il logaritmo non furono definiti in questo modo geometrico ma si può fare).

E il **circolo** genera le **funzioni trigonometriche ovvero goniometriche** che considereremo in questa Lezione: seno, coseno, ecc.

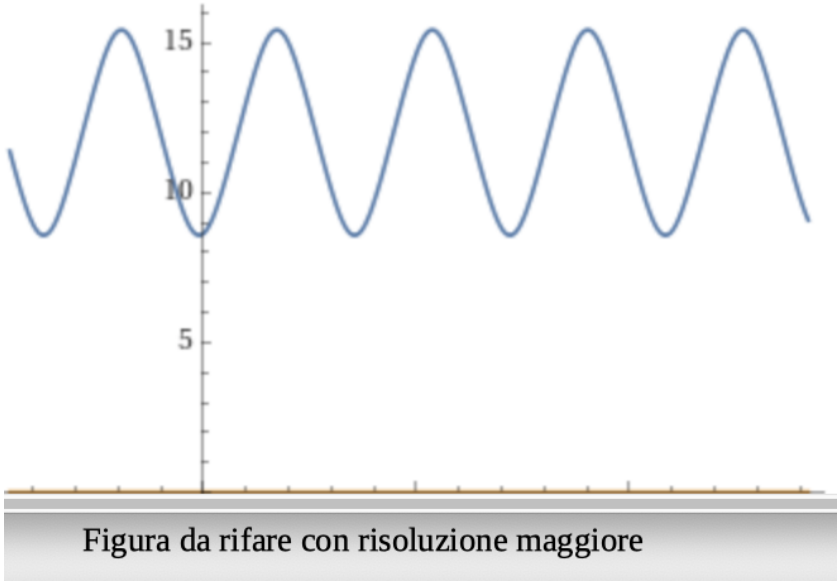
18.2 Oscillazioni periodiche ovunque



Si veda sul prestigioso sito online EuroMOMO, con statistiche sugli eccessi di mortalità in decine di stati europei, la *baseline*, cioè in pratica il numero di morti attesi in base alla media degli anni passati: oscilla, con molti più morti nella stagione fredda che in quella calda, in modo del tutto affine alla cosinusoide sopra disegnata.



Questa è la durata delle ore di luce alla latitudine di 45° nello stesso periodo, in stretta corrispondenza.



In Europa le persone muoiono di più intorno a gennaio che intorno ad agosto, per varie ragioni mediche, fra cui le seguenti.
C'è meno sole: diminuisce la temperatura, favorendo malattie.
C'è meno sole: si riduce la produzione di vitamina D nel corpo.

Guardando in dettaglio il grafico di EuroMOMO, per i 27 stati considerati (2024, prima erano meno) troviamo ogni anno circa 80'000 morti alla settimana in nei mesi di gennaio e 68'000 in luglio, diciamo una media annuale di 64'000 morti alla settimana con un'oscillazione di ±6'000.

BOZZA - DRAFT

18.3 Le funzioni trigonometriche in Farmacia

Le funzioni goniometriche ovvero trigonometriche ricorrono un'infinità di volte nelle Scienze Applicate, in particolare in Fisica.

Le applicazioni in Farmacia e Chimica Farmaceutica sono rare, ma quelle vicine ad esse sono molte, e in Fisica moltissime.

Qua di seguito qualche ricorrenza vicina alla Farmacia.

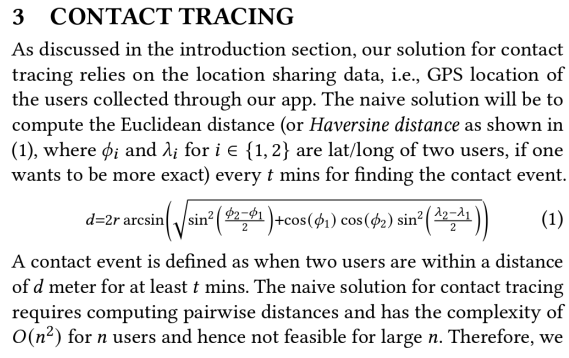
- Si può ricercare sul web
  `"Sinusoidal heart rate"`

dove viene in pratica solo considerata la forma[88] sinusoidale.

- Si può guardare la formula che dà la distanza fra 2 punti sulla Terra (identificati da latitudine e longitudine) che ipoteticamente può venire considerata nel tracciamento dei contatti covid, come si vede in[89] figura, formula con 3 funzioni goniometriche:

[88] Leggiamo per esempio in Sinusoidal heart rate pattern: Reappraisal of its definition and clinical significance, di Modanlou HD, Murata Y, su The journal of obstetrics and gynaecology research 2004 Jun;30(3):169-80: Sinusoidal heart rate (...) was called 'sinusoidal' because of its sine waveform (...) A true SHR is an ominous sign of fetal jeopardy needing immediate intervention.

[89] Screenshot da arXiv:2105.15030 [cs.HC] https://doi.org/10.48550/arXiv.2105.15030 "Digital Contact Tracing for Covid 19", by Chandresh Kumar Maurya, Seemandhar Jain,



3 CONTACT TRACING

As discussed in the introduction section, our solution for contact tracing relies on the location sharing data, i.e., GPS location of the users collected through our app. The naive solution will be to compute the Euclidean distance (or *Haversine distance* as shown in (1), where ϕ_i and λ_i for $i \in \{1, 2\}$ are lat/long of two users, if one wants to be more exact) every t mins for finding the contact event.

$$d = 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\phi_2 - \phi_1}{2}\right) + \cos(\phi_1)\cos(\phi_2)\sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right) \quad (1)$$

A contact event is defined as when two users are within a distance of d meter for at least t mins. The naive solution for contact tracing requires computing pairwise distances and has the complexity of $O(n^2)$ for n users and hence not feasible for large n. Therefore, we

Le formule trigonometriche stanno dietro la grafica co0mputerizzata 3D. Si veda per esempio questo spettacolare visualizzatore interattivo di molecole:

https://molview.org/

• Sono alla base (con l'Analisi di Fourier) di una branca scientifica in qualche modo tangente alle Scienze Biomediche: l'Elaborazione di Immagini.

BOZZA - DRAFT

• Gli angoli di legame della Chimica che accenneremo in 18.7.

• Nelle determinazioni posizionali dell'MRI, *magnetic resonance imaging*.

• La trigonometria tocca marginalmente la Farmacia anche con la Legge di Bragg e la Legge di Snell, utili nella Cristallografia a raggi X, usata per l'analisi dei cristalli di sostanze chimiche solide. LINK-> LINK-> LINK->

• La questione trigonometrica più interessante in questa trattazione elementare è la funzione arcotangente perché dà una distribuzione di probabilità (di Cauchy) simile alla ben più importante distribuzione normale, ma per certi versi più semplice.

18.4 Le nozioni principali

– Grafici:

Vishal Thakre (Submitted on 22 May 2021)

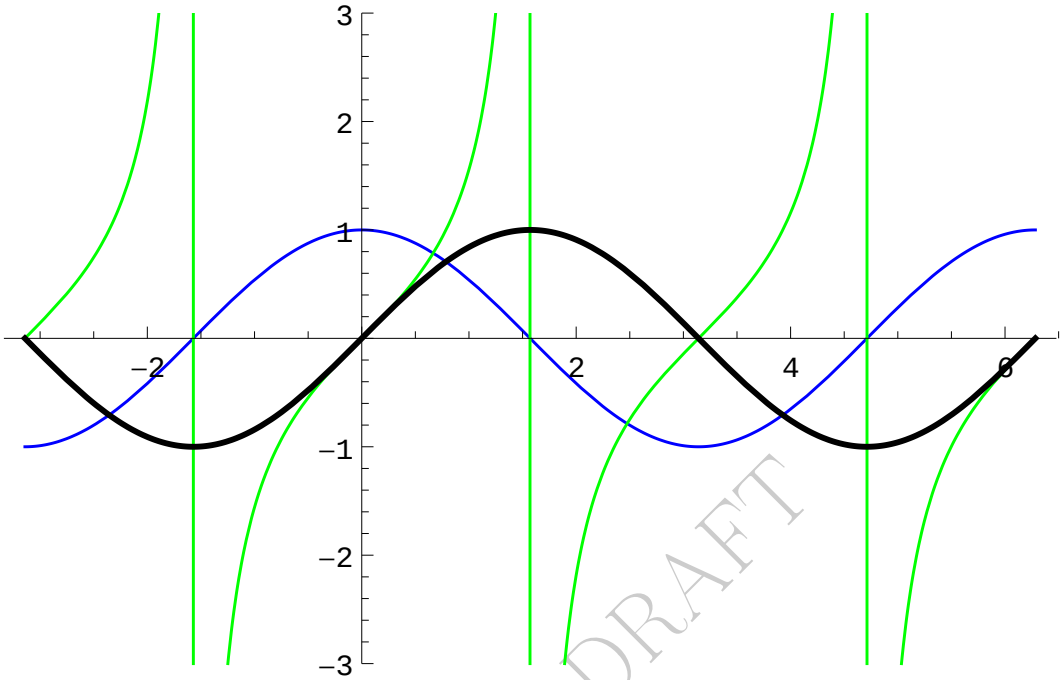


Figure 22: (nero) sinusoide, (blu) cosinusoide, (verde) tangentoide e suoi asintoti

BOZZA - DRAFT

– Il seno e il coseno sono funzioni definite su tutto $\mathbb{R}$ a valori in $\mathbb{R}$.

– Si tratta sostanzialmente della stessa funzione, ovvero più precisamente l'una è "traslata" dell'altra:

$$\sin(x) = \cos\left(x - \frac{\pi}{2}\right)$$

– Seno e coseno sono legate[90] anche dall'**identità goniometrica fondamentale:**

$$\boxed{\sin^2 x + \cos^2 x = 1} \tag{46}$$

– Da esse si definisce una terza funzione, la tangente:

$$\boxed{\tan(x) := \frac{\sin(x)}{\cos(x)}, \qquad x \neq \frac{\pi}{2} + k\pi\,,\ k \in \mathbb{Z}} \tag{47}$$

[90] Questa è chiaramente una relazione *algebrica* fra seno e coseno, mentre la precedente è un'identità funzionale.

– Sono tutte funzioni periodiche:

$$\boxed{\begin{aligned} &\sin(x + 2k\pi) = \sin(x) \quad k \in \mathbb{Z}, \\ &\cos(x + 2k\pi) = \cos(x) \quad k \in \mathbb{Z} \\ &\tan(x + k\pi) = \tan(x) \quad k \in \mathbb{Z} \end{aligned}} \tag{48}$$

– Impariamo, oppure ricaviamo dal circolo goniometrico (vedi dopo) che il seno è dispari, il coseno è pari, la tangente è dispari:

$$\boxed{\begin{aligned} &\sin(-x) = -\sin x \\ &\cos(-x) = \cos x \\ &\tan(-x) = -\tan x \end{aligned}} \tag{49}$$

BOZZA - DRAFT

– Formula di addizione del seno:

$$\sin(x + y) = \sin x \cos y + \cos x \sin y \tag{50}$$

– Da quest'ultima con $y = x$ segue la formula di duplicazione del seno:

$$\boxed{\sin(2\,x) = 2\,\sin x\,\cos x} \tag{51}$$

– Formula di addizione del coseno:

$$\cos(x + y) = \cos x \cos y - \sin x \sin y$$

– Da quest'ultima con $y = x$ segue una formula di duplicazione del coseno:

$$\cos 2\,x = \cos^2 x - \sin^2 x$$

– Da quest'ultima e dall'identità goniometrica fondamentale segue quest'altra formula di duplicazione del coseno:

$$\boxed{\cos(2\,x) = 2\,\cos^2 x - 1} \tag{52}$$

– In $\frac{\pi}{4}$ seno e coseno hanno questo stesso valore:

$$\boxed{\sin\frac{\pi}{4} = \cos\frac{\pi}{4} = \frac{\sqrt{2}}{2}} \tag{53}$$

– Da quest'ultima, applicando le formule di duplicazione si ottengono i valori di seno e coseno in $\frac{\pi}{2}$, poi in π, poi in $2\,\pi$:

$$\boxed{\sin\tfrac{\pi}{2} = 1 \quad \cos\tfrac{\pi}{2} = 0} \tag{54}$$

$$\boxed{\sin\pi = 0 \quad \cos\pi = -1} \tag{55}$$

$$\sin 2\,\pi = 0 \quad \cos 2\,\pi = 1 \tag{56}$$

BOZZA - DRAFT

e da quest'ultima per la $2\,\pi$-periodicità

$$\boxed{\sin 0 = 0 \quad \cos 0 = 1} \tag{57}$$

18.5 Definizioni

Nel piano cartesiano di assi X, Y e origine O consideriamo il *circolo goniometrico* che è il circolo di centro O e raggio 1, rappresentato in Figura.

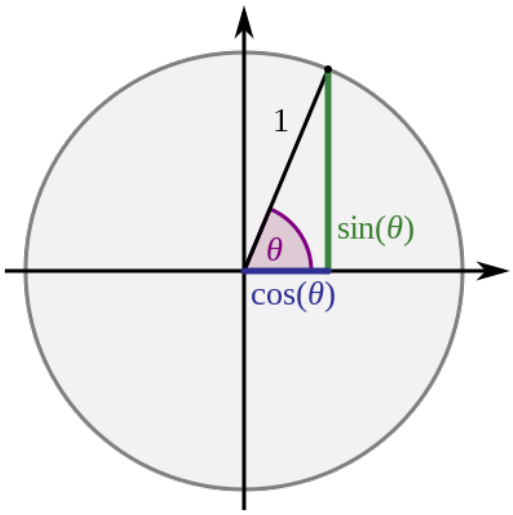


Figure 23: Il circolo goniometrico. Wikimedia, public domain image, by Stephan Kulla (2017)

La lettera greca θ, spesso usata come misura di angoli, come nella soprastante figura, è la *theta*, traslitterata *th*. È la lettera di *arithmos*, il numero, della citazione pitagorica all'inizio della Lezione 1, che ritroviamo nelle parole inglesi *arithmetic* e *logarithm*.

BOZZA DRAFT

Consideriamo un arco (del circolo) che inizia in $(1,0)$ procedendo in senso antiorario. Esso determina un *angolo goniometrico* di vertice O. La lunghezza x dell'arco si dice *misura in radianti* dell'*angolo goniometrico*. E si possono considerare misure negative con archi orientati in senso in senso orario. Chiamiamo $P(x)$ il secondo estremo dell'arco. La misura in radianti a priori $\in [-2\pi, 2\pi]$, e poi in effetti *facendo anche più di 1 giro*, $x \in \mathbb{R}$. Si definiscono le funzioni *seno* e *coseno*, $\sin : \mathbb{R} \to \mathbb{R}$, $\cos : \mathbb{R} \to \mathbb{R}$, e la *tangente*:

$$\cos(x) := ascissa(P(x)) \tag{58}$$

$$\sin(x) := ordinata(P(x)) \tag{59}$$

e la tangente è l'ordinata del punto d'intersezione della retta $OP(x)$ con la retta tangente al circolo goniometrico in $(1,0)$.

Per esempio si ha (banalmente le prime 2, difficilmente la terza)

$$\sin \pi = 0 \qquad \cos \pi = -1 \qquad \sin 1 \approx 0.84 \tag{60}$$

18.6 Gradi e radianti

In questa trattazione considereremo il seno e il coseno quasi solo per $0 \le x \le \pi$ e la tangente principalmente per $-\frac{\pi}{2} < x < \frac{\pi}{2}$.
La misura dell'angolo piatto di π radianti, in un altro sistema di misura – più antico – si dice di 180°, e vale la proporzione

$$\pi_{rad} : 180^\circ = x_{rad} : x^\circ \tag{61}$$

che consente di passare da un sistema all'altro.
In particolare

angolo giro	360°	$2\,\pi$ radianti
angolo piatto	180°	π radianti
angolo retto	90°	$\frac{\pi}{2}$ radianti

(62)

BOZZA - DRAFT

Ecco alcuni degli *angoli notevoli* che si considerano a un livello molto elementare:

x	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	$\frac{2}{3}\pi$	$\frac{3}{4}\pi$	$\frac{5}{6}\pi$	π	$\frac{3}{2}\pi$	2π
x°	0°	30°	45°	60°	90°	120°	135°	150°	180°	270°	360°

e un qualche interesse ha anche 1 rad ≈ 57°, e tutte queste corrispondenze si calcolano con la (61). Quasi mai[91] si scrive *rad*.

[91] Per il lettore interessato, alcuni valori numerici interessanti:

$\frac{\pi}{4} \approx 0.785$ (che di solito useremo nel senso di $\frac{\pi}{4}\,rad \approx 0.785\,rad$)
$\frac{\pi}{3} \approx 1.05$ (che di solito useremo nel senso di $\frac{\pi}{3}\,rad \approx 1.05\,rad$)
$\frac{\pi}{2} \approx 1.57$ (che di solito useremo nel senso di $\frac{\pi}{2}\,rad \approx 1.57\,rad$).
$\frac{\sqrt{2}}{2} \approx 0.707$ (per esempio $\cos\frac{\pi}{4}$)
$\frac{\sqrt{3}}{2} \approx 0.866$ (per esempio $\cos\frac{\pi}{3}$)
$\frac{\sqrt{3}}{3} \approx 0.577$ (per esempio $\tan\frac{\pi}{3}$)

Primi e secondi. La metà di un angolo di $\pi/4$ radianti ha $\pi/8$ radianti. La metà di quello stesso angolo, espresso in gradi, cioè di 45°, è 22.5° ma spesso si scrive diversamente: come

per 22 ore e mezza scriviamo $22^h30'$

cioè

si usano i 60 minuti primi che suddividono 1 ora, similmente
si usano i 60 primi (di grado) che suddividono 1 grado

così

per 22.5° (22 gradi e mezzo) si può anche scrivere $22°\,30'$.

E ci sono anche 60 *secondi* in un *primo* (questione che non approfondiamo).

Le conversioni fra le 2 scritture delle misure in gradi

con le cifre decimali, come 22.5°
coi primi e secondi, come $22°\,30'$

è in generale calcolo difficile; ma in questo caso è stata semplice.

Si tratta esattamente degli stessi valori della *latitudine* e *longitudine*, altra questione che non approfondiamo; diamo solo a titolo di esempio le coordinate di Trieste: 45° 39' 01" N 13° 46' 13" E.
Si tratta naturalmente di angoli nello spazio tridimensionale, come quelli del paragrafo seguente.

BOZZA, DRAFT

18.7 Gli angoli di legame della Chimica

Angolo di legame del metano: $\approx 109°28'$.
E si veda l'eptafluoruro di iodio, IF_7, a questo LINK->
Altri angoli di legame a questo LINK->
Si noti che per il metano dà, correttamente, 109.5°. (I 28' di 109° 28', prima detti, corrispondono, salvo approssimazione, a 0.5°).

ESERCIZIO$_\mu$
Qual è la misura in gradi di un angolo di $\frac{2}{3}\pi$ radianti? (È l'angolo di legame del trifluoruro di boro BF_3).

SVOLGIMENTO
Sappiamo che $180°$ sono π radianti da cui la proporzione

$$\pi_{rad} : 180° = x_{rad} : x°$$

e adesso

$$\pi_{rad} : 180° = \frac{2}{3}\pi_{rad} : x°$$

$$\frac{\pi_{rad}}{180°} = \frac{\frac{2}{3}\pi_{rad}}{x°} \qquad / \cdot \frac{180° \cdot x°}{\pi_{rad}}$$

$$x° = \frac{2}{3} \cdot 180°$$

$$\boxed{120°}$$

BOZZA - DRAFT

18.8 Alcuni valori notevoli

Impariamo a leggere questo diagramma, che, senza imparare a memoria niente, ci dà seni e coseni di molti *angoli notevoli*.

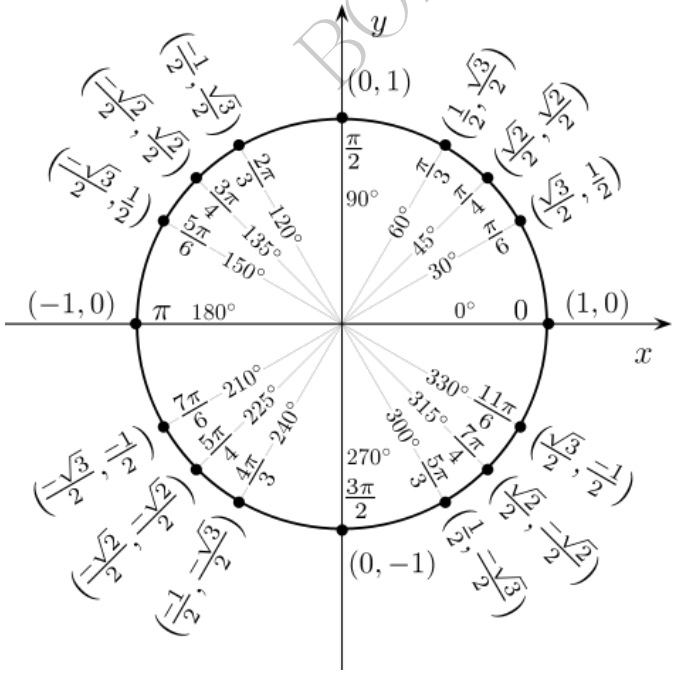


Figure 24: Angoli notevoli e corrispondenti valori di seno e coseno

18.9 Funzioni goniometriche inverse

Definizione classica di arcotangente.

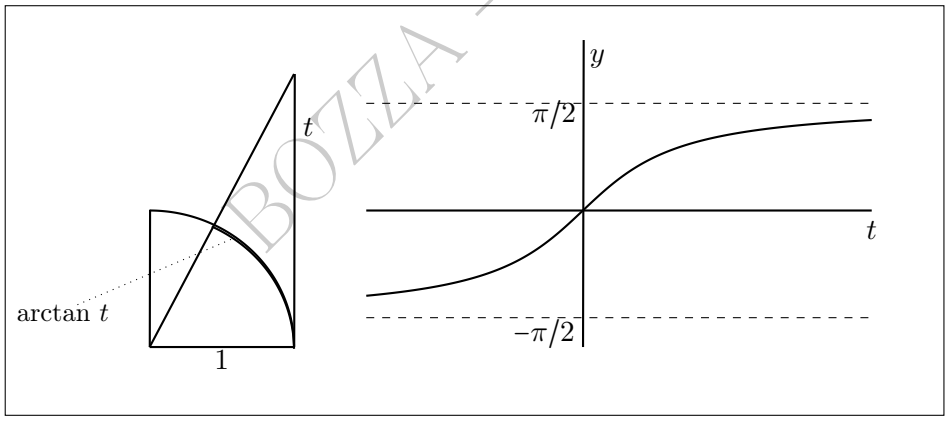
$$\boxed{\text{arctan è l'inversa della restrizione della tangente a } \left]-\frac{\pi}{2},\frac{\pi}{2}\right[} \tag{63}$$

Si prenda allora il secondo arco curvilineo verde nella Figura 22 e lo si simmetrizzi rispetto alla bisettrice del primo e terzo quadrante, e si otterrà il grafico della funzione arcotangente, rappresentato qua sotto nella figura di destra.
In simboli:

$$\arctan := \left(\tan\big|_{]-\pi/2,\pi/2[} \right)^{-1}$$

Detto altrimenti,
$\arctan x$ **è l'angolo** α **la cui tangente è** x, $\qquad \alpha \in]-\pi/2, \pi/2[$**.**

BOZZA - DRAFT



Definizione geometrica di arctan. (Senza usare tan).
Con riferimento alla figura a sinistra, si chiama arcotangente di $t > 0$ la lunghezza dell'arco di circolo intersezione del quarto di circolo col triangolo rettangolo (pieno) di cateti 1 e t.
Per $t < 0$ si pone

$$\arctan(-t) = -\arctan t$$

e ancora si pone

$$\arctan 0 = 0$$

Al variare di t in $\mathbb{R}$ si ottiene così la funzione arcotangente, di cui si vede in figura un disegno (di una parte) del grafico.

Teorema. (Ovvio)
(a) arctan è dispari:

$$\boxed{\arctan(-x) = -\arctan(x)} \tag{64}$$

(b) arctan è crescente;
(c) "$\arctan(-\infty) = -\frac{\pi}{2}$", come limite, valore mai raggiunto;
(d) "$\arctan(+\infty) = \frac{\pi}{2}$", come limite, valore mai raggiunto;
(e) arctan ha grafico sigmoide.

Teorema. (Di difficile dimostrazione)
(f) arctan non è nè razionale nè irrazionale (tecnicamente si dice funzione *trascendente*, come anche sono i *logaritmi*);
(g) arctan ha grafico senza *punti angolosi*, dotato di *retta tangente* in ogni punto (funzione *derivabile*).

BOZZA - DRAFT

Definizioni (che non diamo). Similmente all'arcotangente, vi sono poi anche[92] l'arcoseno (arcsin) e l'arcocoseno (arccos).

Esempio di Chimica. Per esempio

$$\pi - \arctan(2\sqrt{2}) =$$
$$= \arccos(-1/3) \approx$$
$$\approx 1.910633$$

che convertendo da radianti in gradi è $\approx 109°28'$: è l'angolo di legame del metano.
Con WolframAlpha

[92] **Arcoseno:**
$\arcsin x$ è l'angolo α il cui seno è x, con $\alpha \in [-\pi/2, \pi/2]$.

Arcocoseno:
$\arccos x$ è l'angolo α il cui coseno è x, con $\alpha \in [0, \pi]$.
In simboli:

$$\arcsin := \left(\sin|_{[-\pi/2,\pi/2]}\right)^{-1}$$
$$\arccos := \left(\cos|_{[0,\pi]}\right)^{-1}$$

`arccos(-1/3)`

dà

`1.9106332...`(eccetera, moltissimi decimali non qua riportati)

e poi

`convert 1.910633 rad into degrees`

dà

`109.471° (degrees)`

`109 degrees 28 arc minutes 16 arc seconds`

Nota 1.
L'arcotangente è variamente denotata dai vari Autori e software.
In particolare

arctan (che è lo standard ISO)

$\tan^{-1}$ (compatibile con lo standard ISO)

E ancora

arctang, arctg, `ArcTan, Atn, atg, atan (e datan), rad(atan(.))`...
(Non è la funzione `atan2`, comunque correlata).

BOZZA - DRAFT

Nota 2. Ritroveremo l'arcotangente in Calcolo delle Probabilità (nella *densità di Cauchy*, in 37.2). L'arcotangente appare anche nella Cristallografia a raggi X.

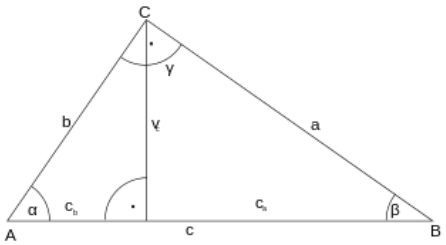
Nota 3: alcuni valori notevoli. Dalle definizioni si ha subito

x	0	$\frac{1}{2}$	$\frac{\sqrt{3}}{3}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1	$\sqrt{3}$
arctan	0		$\frac{\pi}{6}$			$\frac{\pi}{4}$	$\frac{\pi}{3}$
arcsin	0	$\frac{\pi}{6}$		$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	$\nexists$
arccos	$\frac{\pi}{2}$	$\frac{\pi}{3}$		$\frac{\pi}{4}$	$\frac{\pi}{6}$	0	$\nexists$

I valori mancanti in tabella non si considerano *valori notevoli*, ma esistono. (Per esempio $\arcsin\frac{\sqrt{3}}{3} \approx 0.61548$).

Altri valori si ottengono dalla disparità dell'arcotangente (e dell'arcoseno).

18.10 Trigonometria – La goniometria dei triangoli



In un triangolo rettangolo di lati a, b, c, con angoli rispettivamente opposti di misure α, β, $90°$ i lati di misure a e b si chiamano *cateti* e l'altro *ipotenusa* ed è

$$\frac{a}{b} = \tan\alpha \tag{65}$$

$$a = c\,\sin\alpha \tag{66}$$

$$a = c\,\cos\beta \tag{67}$$

BOZZA - DRAFT

Esistono innumerevoli altri teoremi e formule di Trigonometria, e un paio sono nella sezione di Complementi a questa Lezione.

Nota. Di questa Lezione ci si dovrebbe aspettare che lo studente conosca a memoria le formule *contemporaneamente* numerate *et* riquadrate, oltre a tutto l'inquadramento teorico della questione.
Non ci proponiamo di imparare a memoria nessun ALTRO valore di seno, coseno, tangente e arcotangente, ma solo di saperli ricavare disponendo di tabelle analoghe a quelle sopra esposte, oppure dalla Figura 18.8 o analoga.

* Complementi *

18.11 Complementi – Altre formule goniometriche

Formule di addizione e sottrazione:

$$\begin{aligned}
&\sin(x+y) = \sin x \cos y + \cos x \sin y \quad (vista\ prima)\\
&\sin(x-y) = \sin x \cos y - \cos x \sin y\\
&\cos(x+y) = \cos x \cos y - \sin x \sin y \quad (vista\ prima)\\
&\cos(x-y) = \cos x \cos y + \sin x \sin y
\end{aligned}$$

Formule di duplicazione:

$$\sin(2\,x) = 2\,\sin x \cos x \qquad (vista\ prima)$$

$$\cos(2\,x) = \begin{cases} 2\,\cos^2 x - 1 & (vista\ prima)\\ \cos^2 x - \sin^2 x & (vista\ prima)\\ 1 - 2\,\sin^2 x \end{cases}$$

$$\tan(2\,x) = \frac{2\,\tan(x)}{1-\tan^2 x}, \qquad x \neq \frac{\pi}{4} + k\,\frac{\pi}{2}, \quad k \in \mathbb{Z}$$

BOZZA - DRAFT

Formule di prostaferesi:

$$\begin{aligned}
&\sin x + \sin y = 2\,\sin\tfrac{x+y}{2}\,\cos\tfrac{x-y}{2}\\
&\sin x - \sin y = 2\,\cos\tfrac{x+y}{2}\,\sin\tfrac{x-y}{2}\\
&\cos x + \cos y = 2\,\cos\tfrac{x+y}{2}\,\cos\tfrac{x-y}{2}\\
&\cos x - \cos y = -2\,\sin\tfrac{x+y}{2}\,\sin\tfrac{x-y}{2}
\end{aligned}$$

Formule di bisezione della tangente:

$$\tan\frac{x}{2} = \begin{cases} \frac{\sin x}{1+\cos x}, & x \neq (2\,k+1)\,\pi\\ \\ \frac{1-\cos x}{\sin x}, & x \neq k\,\pi \end{cases} \qquad k \in \mathbb{Z}$$

Approssimazione del seno per piccoli valori dell'argomento:

$$\sin x \approx x \qquad |x| \ll$$

Esiste poi questa relazione *funzionale* fra seno e coseno:

$$\cos x = \sin\left(x + \frac{\pi}{2}\right)$$

che si ottiene subito dalla (50) (ponendovi $y = \frac{y}{2}$).
Ed esiste una *relazione differenziale* fra seno e coseno, che esprimiamo a parole:
la retta tangente alla sinusoide nel punto di ascissa x ha coefficiente angolare $\cos x$
(e che in formule si esprime con $D \sin x = \cos x$ che spiegheremo nella lezione sulle derivate).

BOZZA - DRAFT

Vi sono poi le *formule parametriche razionali*, e molte altre.

18.12 Complementi – Il calcolo di queste funzioni

Molti valori delle funzioni goniometriche si trovano nelle tavole che abbiamo dato in questa lezione. Numerosissimi altri valori si trovano in forma approssimata su tavole cartacee ora riprodotte su internet. Valori (approssimati) si calcolano con le calcolatrici scientifiche, reali o virtuali, e online con WolframAlpha.

Questione interessantissima – veramente di fondamentale importanza in Matematica – è come fanno i software a calcolare quei valori, e come fecero i nostri antenati nei secoli passati a calcolare i valori scritti nelle tavole cartacee.

Ebbene, quei valori (approssimati ma con quanti decimali esatti si voglia) si calcolano con *serie*, ovvero in pratica somme infinite, questione che approfondiremo in una lezione successiva.

18.13 Complementi – Angoli notevoli

Nella figura del paragrafo sui valori notevoli, si possono apprezzare molti angoli notevoli, espressi in gradi e radianti, e il loro cor-

rispondente punto goniometrico, con coordinate il coseno e rispettivamente il seno di quell'angolo. Per esempio

$$\cos 30^\circ = \cos \frac{\pi}{6} = \frac{\sqrt{3}}{2}$$

$$\sin 30^\circ = \sin \frac{\pi}{6} = \frac{1}{2}$$

Semplici considerazioni di geometria elementare danno subito

			$-\pi$	$-\frac{\pi}{2}$	
x	0	$\frac{\pi}{2}$	π	$\frac{3}{2}\pi$	2π
x°	0°	90°	180°	270°	360°
sin	0	1	0	-1	0
cos	1	0	-1	0	1
tan	0	$\nexists$	0	$\nexists$	0
	$*$	$*$			

In questa e nelle successive 4 tabelle le colonne asteriscate sono le più importanti.

BOZZA - DRAFT

Non ci proponiamo di imparare a memoria nessun valore di seno, coseno, tangente (e poi arcotangente), ma solo di saperli ricavare disponendo di queste o analoghe tabelle, oppure dalla Figura 24.

Caveat. Nello schema soprastante è da intendersi che

$$\sin\left(-\frac{\pi}{2}\right) = \sin\left(\frac{3}{2}\pi\right) = \sin(270^\circ)$$

assolutamente non si intenda che $-\frac{\pi}{2}$ sia uguale a 270°.

La considerazione del quadrato inscritto nel circolo goniometrico e coi lati paralleli agli assi, che per il Teorema di Pitagora ha lato $\sqrt{2}$, dà subito, con analogo *caveat* come sopra (e anche dopo),

x	$\frac{\pi}{4}$	$\frac{3}{4}\pi$	$-\frac{3}{4}\pi$	$-\frac{\pi}{4}$
$x°$	$45°$	$135°$	$225°$	$315°$
sin	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{2}}{2}$	$-\frac{\sqrt{2}}{2}$	$-\frac{\sqrt{2}}{2}$
cos	$\frac{\sqrt{2}}{2}$	$-\frac{\sqrt{2}}{2}$	$-\frac{\sqrt{2}}{2}$	$\frac{\sqrt{2}}{2}$
tan	1	-1	1	-1
	$*$			

La considerazione dei 2 esagoni regolari inscritti nel circolo goniometrico, con un vertice in $(0,1)$ e rispettivamente in $(1,0)$, dà subito, dividendoli in triangoli equilateri di lato 1 e altezza $\frac{\sqrt{3}}{2}$,

x	$\frac{\pi}{6}$	$\frac{5}{6}\pi$	$-\frac{5}{6}\pi$	$-\frac{\pi}{6}$
$x°$	$30°$	$150°$	$210°$	$330°$
sin	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$
cos	$\frac{\sqrt{3}}{2}$	$-\frac{\sqrt{3}}{2}$	$-\frac{\sqrt{3}}{2}$	$\frac{\sqrt{3}}{2}$
tan	$\frac{\sqrt{3}}{3}$	$-\frac{\sqrt{3}}{3}$	$\frac{\sqrt{3}}{3}$	$-\frac{\sqrt{3}}{3}$
	$*$			

x	$\frac{\pi}{3}$	$\frac{2}{3}\pi$	$-\frac{2}{3}\pi$	$-\frac{\pi}{3}$
$x°$	$60°$	$120°$	$240°$	$300°$
sin	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{3}}{2}$	$-\frac{\sqrt{3}}{2}$	$-\frac{\sqrt{3}}{2}$
cos	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$
tan	$\sqrt{3}$	$-\sqrt{3}$	$\sqrt{3}$	$-\sqrt{3}$
	$*$			

La considerazione di decagoni regolari darebbe ancora altri valori.

BOZZA - DRAFT

18.14 Complementi – Trigonometria

In un triangolo di lati a, b, c, con angoli rispettivamente opposti α, β, γ valgono i seguenti 2 teoremi.

Teorema dei Seni.[93]

$$\frac{a}{\sin\alpha} = \frac{b}{\sin\beta} = \frac{c}{\sin\gamma}.$$

Teorema del Coseno.[94]

$$c^2 = a^2 + b^2 - 2\,a\,b\,\cos\gamma.$$

[93] Nasir al-Din al-Tūsī, persiano, XIII sec. d.C. "He is often considered the creator of trigonometry as a mathematical discipline in its own right", Wikipedia, l'enciclopedia libera.

[94] Euclide, greco, III sec. a.C., ben noto, e Jamshīd al-Kāshī, persiano, XIII sec. d.C. "Much of al-Kāshī's work was not brought to Europe, and much, even the extant work, remains unpublished in any form", Wikipedia, l'enciclopedia libera.

Ed esistono innumerevoli altri teoremi e formule di Trigonometria.

BOZZA - DRAFT

ESERCIZI SULLA LEZIONE 18

ESERCIZIO$_\mu$
Qual è la misura in radianti di un angolo di 150°?

SVOLGIMENTO
Sappiamo che 180° sono π radianti da cui la proporzione

$$\pi_{rad} : 180^\circ = x_{rad} : x^\circ$$

e adesso

$$\pi_{rad} : 180^\circ = x_{rad} : 150^\circ$$

$$\frac{\pi_{rad}}{180^\circ} = \frac{x_{rad}}{150^\circ} \qquad \Big/ \cdot 150^\circ$$

$$\frac{\pi_{rad}}{180^\circ} \cdot 150^\circ = x_{rad}$$

$$x_{rad} = \frac{15}{18}\pi_{rad}$$

$$\boxed{\frac{5}{6}\pi}$$

BOZZA - DRAFT

ESERCIZIO$_\mu$
$*$ La concentrazione sanguigna di una certa sostanza sia modellizzata nelle 24 ore del giorno da

$$u_a(t) := 2\,a + a \sin\frac{\pi\, t}{12} \qquad 0 \le t \le 24$$

essendo a un parametro fissabile farmacologicamente. In quale orario la concentrazione è ≥ 2.5 per $a := 1$?

Svolgimento.
Seppure non sia necessario per risolvere il problema, vediamo su WolframAlpha il grafico della concentrazione per $a := 1$ scrivendovi

```
plot 2+Sin[Pi t/12], t from 0 to 24
```

Ora (con $a := 1$) abbiamo la disequazione

$$2 + \sin\frac{\pi\, t}{12} \ge 2.5 \qquad \Big/ + (-2)$$

$$\sin\frac{\pi\, t}{12} \ge 0.5$$

$$\text{poniamo } x := \frac{\pi\, t}{12} \text{ ottenendo } \sin x \ge \frac{1}{2}$$

e ci interessano solo i tempi

$$0 \le t \le 24 \qquad \Big/ \cdot \frac{\pi}{12}$$

$$0 \le \frac{\pi\, t}{12} \le 2\,\pi \text{ cioè } 0 \le x \le 2\,\pi$$

e allora in $[0, 2\,\pi]$ dobbiamo risolvere $\sin x \ge \frac{1}{2}$, che si fa disegnando il circolo goniometrico, trovando subito (sono valori notevoli)

$$\frac{\pi}{6} \le x \le \frac{5}{6}\pi$$

cioè

$$\frac{\pi}{6} \le \frac{\pi\, t}{12} \le \frac{5}{6}\pi \Big/ \cdot \frac{12}{\pi}$$
$$2 \le t \le 10$$

e in conclusione

Fra le 2 e le 10 di mattina

BOZZA - DRAFT

ESERCIZIO$_{\mu 2018}$ Risolvere la seguente equazione goniometrica:

$$\sin x = \frac{1}{2}$$

Svolgimento
Si risolve subito disegnando un circolo goniometrico e segnandovi gli angoli $\frac{\pi}{6}$ e $\frac{5}{6}\pi$, corrispondenti appunto a $\sin x = \frac{1}{2}$:

$$x = \tfrac{\pi}{6} + 2k\pi \ \vee \ x = \tfrac{5}{6}\,\pi + 2k\pi, \quad k \in \mathbb{Z}$$

(Si notino le periodicità).

ESERCIZIO$_{\mu}$ Approssimare cos(1) usando le (60) e (46).
Si verifichi poi su WolframAlpha con `Cos[1]`

ESERCIZIO$_{\mu 2018}$ Risolvere la disequazione

$$\frac{\sqrt{3}}{2} - \cos x < 2\,\cos\left(x + \frac{2\pi}{3}\right)$$

Svolgimento
Con la formula di addizione del coseno

$$2\Big(\cos x\,\cos\Big(\frac{2\pi}{3}\Big)-\sin x\,\sin\Big(\frac{2\pi}{3}\Big)\Big)=$$

e coi valori notevoli ben conosciuti

$$=2\Big((\cos x)\Big(-\frac{1}{2}\Big)-(\sin x)\Big(\frac{\sqrt{3}}{2}\Big)\Big)=-\cos x-\sqrt{3}\,\sin x$$

e con il secondo membro così riscritto la disequazione equivale a

$$\frac{\sqrt{3}}{2}-\cos x<-\cos x-\sqrt{3}\,\sin x$$

$$\frac{\sqrt{3}}{2}<-\sqrt{3}\,\sin x$$

$$-\sqrt{3}\,\sin x>\frac{\sqrt{3}}{2}\quad\Big/:(-\sqrt{3})<0$$

$$\sin x<-\frac{1}{2}$$

che si risolve subito disegnando un circolo goniometrico e segnandovi l'arco da $-\frac{5}{6}\pi$ fino a $-\frac{\pi}{6}$, estremi esclusi, corrispondente appunto a $\sin x<-\frac{1}{2}$:

$$\boxed{-\tfrac{5}{6}\pi+2k\pi<x<-\tfrac{\pi}{6}+2k\pi,\quad k\in\mathbb{Z}}$$

equivalentemente scritta

$$\boxed{\tfrac{7}{6}\pi+2k\pi<x<\tfrac{11}{6}\pi+2k\pi,\quad k\in\mathbb{Z}}$$

BOZZA - DRAFT

IV – Statistica descrittiva

Senza dati
sei solo un'altra persona con un'opinione
(William Edwards Deming,
autore di "The New Economics:
For Industry, Government, Education")

La potenza dei dati numerici, specie quando sono confrontati almeno 2, è enorme, e può indurre a riflettere, e, avendone la possibilità, a prendere una decisione piuttosto che un'altra. Con i conseguenti spostamenti di soldi e di potere: per questo veniamo inondati di statistiche dai media; che hanno padroni quelli privati, e "referenti", diciamo così, quelli pubblici – e scelgono quali dati mostrarci, fra infiniti.

BOZZA - DRAFT

Esempio: alcuni dati numerici

> Per la ricerca sul cancro, che causa ogni anno 150 mila morti, l'Italia spende circa 225 milioni di dollari, mentre ne destina 30 miliardi per le spese militari. (2010, Umberto Veronesi, che era stato Ministro della Sanità)

La Statistica Descrittiva è ottima per queste cose, quando usata bene, ed è pessima quando usata male, per errore o dolo.

> *Ci sono tre livelli di gravità delle menzogne: le bugie ufficiose ("non ho studiato perchè il criceto stava male"), le bugie calunniose, e infine la statistica.*

La Statistica in Farmacia

Amplissimo in Farmacia è l'uso della Statistica Descrittiva (oggetto di queste Lezioni) e della Statistica Inferenziale (che vedremo in Lezioni molto successive).
Per esempio:

– Gestione: per la conduzione della farmacia come azienda;
– Ricerca: per selezionare farmaci via via migliori;
– Farmacovigilanza: per ritirare dal commercio quelli poi risultati più nocivi che utili.

Leggiamo:[95]

> Non è mai venuta meno quell'attenzione al farmaco anche se in uso da anni, che diventerà l'odierna farmacovigilanza, che purtroppo non riesce ancora ad evitare casi come la Talidomide della tedesca Grünenthal usato dal 1957 come sedativo, antinausea che provocherà neonati focomelici e verrà ritirato dal commercio alla fine del 1961 (in Italia con sei mesi di ritardo) o il Vioxx antinfiammatorio della Merck messo in commercio nel 1999 e ritirato brutalmente in campo mondiale il 30 settembre 2004 a causa dei gravissimi danni cardiologici riscontrati: casi purtroppo non isolati dovuti ad un accorciamento dei tempi obbligatoriamente molto lunghi e molto costosi che pretende una ricerca etica.
> [Enfasi aggiunta]

BOZZA - DRAFT

[95] Giorgio Du Ban: La storia del Farmaco, in: Atti del 1^o Convegno Nazionale – Italia ed Europa: Storia della Medicina e della Croce Rossa – in onore di LORIS PREMUDA

19 Introduzione alla Statistica Descrittiva

19.1 Introduzione

La Statistica Descrittiva opera su dati.
Per avere i *dati*, è necessaria la *raccolta dati*, operazione preliminare – in generale complessa e fonte di innumerevoli errori e distorsioni – che tecnicamente non fa parte della Statistica Descrittiva.
Accenneremo alla raccolta dati, e cosa molto bella è fare esperienze in aula.

La Statistica Descrittiva si occupa essenzialmente di:

- **riassumere dati, tipicamente molti,**
  (se sono pochi può essere meglio elencarli)
- **presentare il riassunto,**
  il più comprensibilmente possibile: medie, diagrammi, e altro.

BOZZA - DRAFT

È cosa ben diversa, più semplice e storicamente antica, dalla Statistica Inferenziale che verrà trattata fra le Matematiche dell'Incertezza.

In una trattazione elementare della Statistica Descrittiva non si distingue fra *popolazione* e *campione*: abbiamo i dati che abbiamo, e quelli consideriamo popolazione, anche se in effetti sono un campione di una popolazione più ampia. La distinzione diventa invece essenziale nella Statistica Inferenziale.

In questo testo elementare, verrà tralasciata la questione della rilevazione materiale dei dati – con tutta la sua problematicità.

E in questa lezione **verrà trattata solo la sintesi** dei dati, ovvero la rappresentazione dei dati in una forma umanamente comprensibile e *trattabile*, cosa particolarmente utile se i dati sono più di una dozzina.

Si vuole riassumere i dati con 1 diagramma oppure 1 o pochi valori, per

comprenderli

confrontarli ← cosa fondamentale per i farmaci

divulgarli

e, a livello eticamente discutibile, presentarli manipolativamente (cosa che viene fatta amplissimamente: un fenomeno in *diminuzione* si trova sempre come presentarlo, almeno a livello di titolo riassuntivo, come *aumento*, se si vuole, usando artifici statistici formalmente legittimi).

19.2 Riassumere i dati: quanti gatti e quanti cani avete?

Dati rilevati a Farmacia a Trieste il 17/11/2023, per gatti e cani:

```
G: 0 0 0 1 0 0 0 0 0 1 0 1 1 0 1 0 0 0 0 1 0 0 0 0 2 3 2 0 1 0
C: 1 1 0 1 0 2 1 0 0 0 0 0 0 2 4 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0
```

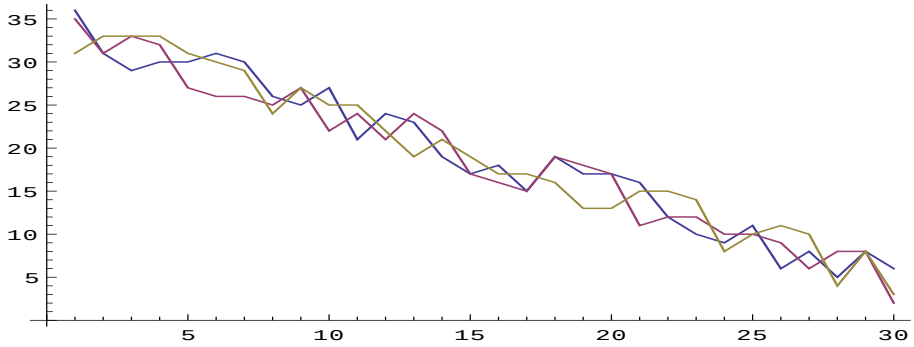
Gli studenti di Farmacia partecipanti alla rilevazione
hanno da 0 a 3 gatti;
hanno da 0 a 4 cani.
Si tratta delle statistiche del minimo e del massimo, di calcolo abbastanza semplice, eppure già molto espressive della situazione, di non così immediata comprensione semplicemente guardando il dataset, specialmente se avesse 100 numeri, come in questo esempio ipotetico:

2, 0, 3, 0, 3, 2, 1, 1, 2, 0, 2, 2, 2, 1, 1, 2, 1, 0, 2, 0, 3, 4, 0, 1, 4, 1, 2, 3, 0, 1, 0, 2, 1, 3, 3, 2, 0, 1, 0, 2, 0, 2, 3, 2, 1, 1, 2, 3, 2, 2, 4, 2, 1, 3, 3, 3, 2, 0, 2, 3, 0, 1, 2, 2, 2, 1, 4, 0, 2, 0, 0, 1, 2, 2, 1, 1, 1, 3, 0, 4, 0, 1, 2, 2, 1, 1, 2, 4, 1, 2, 3, 4, 2, 0, 0, 3, 2, 2, 3, 2

E con più calcoli potremmo trovare il numero medio di gatti, e di cani. E vedremo molte altre forme di Statistica Descrittiva.

BOZZA - DRAFT

Esempio. Consideriamo un fenomeno in evoluzione da un giorno 1 a un giorno 30, tanto più negativo per la società quanto più sono grandi 3 parametri a, b, c.



Un *ingenuo* direbbe che la situazione sta migliorando: i parametri negativi diminuiscono. Invece i media del posto scrivono, a beneficio del *popolo bue*:

Giorno 2 Aumenta il tal parametro!
Giorno 3 Aumenta il tal parametro!
Giorno 4 Aumenta il tal parametro!
Giorno 5 Il Governo valuta la preoccupante situazione
Giorno 6 Aumenta il tal parametro!
Giorno 7 Il Governo valuta la preoccupante situazione
Giorno 8 Il Governo valuta la preoccupante situazione
Giorno 9 Aumenta il tal parametro!
Giorno 10 Aumenta il tal parametro!
Giorno 11 Aumenta il tal parametro!
Giorno 12 Aumenta il tal parametro!
Giorno 13 Aumenta il tal parametro!
Giorno 14 Aumenta il tal parametro!
Giorno 15 Il Governo valuta la preoccupante situazione
Giorno 16 Aumenta il tal parametro!
Giorno 17 Il Governo valuta la preoccupante situazione
Giorno 18 Aumenta il tal parametro!
Giorno 19 Il Governo valuta la preoccupante situazione
Giorno 20 Il Governo valuta la preoccupante situazione
Giorno 21 Aumenta il tal parametro!
Giorno 22 Aumenta il tal parametro!
Giorno 23 Il Governo valuta la preoccupante situazione
Giorno 24 Il Governo valuta la preoccupante situazione
Giorno 25 Aumenta il tal parametro!
Giorno 26 Aumenta il tal parametro!
Giorno 27 Aumenta il tal parametro!
Giorno 28 Aumenta il tal parametro!
Giorno 29 Aumenta il tal parametro!
Giorno 30 Il Governo valuta la preoccupante situazione

BOZZA - DRAFT

BOZZA - DRAFT

19.3 Esempio di possibile raccolta dati

Fare in aula una raccolta dati è interessante ma bisogna fare estrema attenzione al rispetto per la persona (non si possono chiedere il peso e l'altezza, nemmeno in questionari anonimi) e alla possibilità concreta che gli studenti conoscano il dato richiesto (non si può chiedere quanta vitamina D o rame hanno nel sangue perchè in generale non lo sanno). La raccolta dati e successive analisi statistiche sono estremamente valevoli didatticamente e piacciono molto agli Studenti. Alcune possono essere fatte a voce, altre anonimamente.

Ecco una possibile raccolta dati, fattibile addirittura pubblicamente a voce in aula: Quanti animali domestici possiedi?

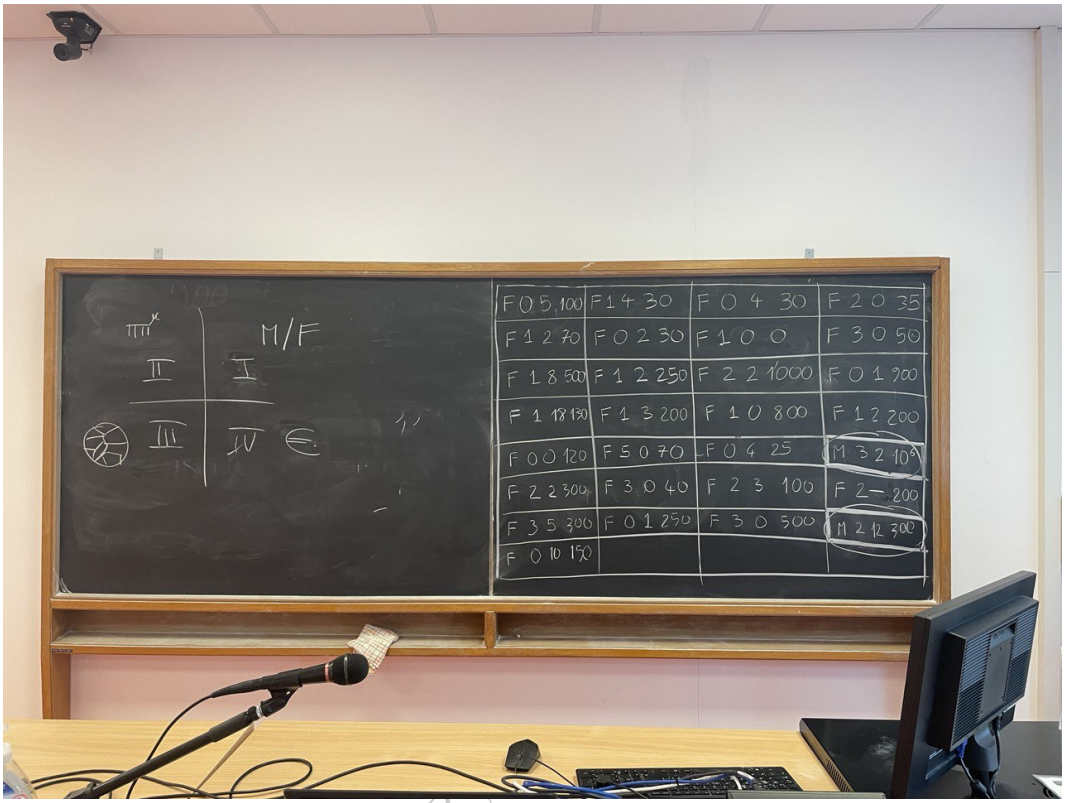


Figure 25: Raccolta dati, Matematica per Farmacia a Trieste, in un anno e con 4 domande che non specifichiamo. Si era consegnato un foglio bianco ad ogni presente in aula e le risposte andavano scritte nei 4 quadranti ideali del foglio. Raccolta dati anonima, risposte lette via via ad alta voce da una volontaria, e trascritte alla lavagna con successiva distruzione dei fogli. Il valore anomalo (*outlier*) un milione è trascritto come 10^6. Un trattino orizzontale indica un dato mancante: un partecipante ha scritto che non sa invece di rispondere con un numero, e forse intendeva 0 o proprio che non sa rispondere, in ogni caso è un dato mancante classificabile come "non sa/non risponde" oppure come dato mancante per qualsiasi causa.

19.4 Il dataset e le operazioni su esso

L'insieme dei dati $x_1, ..., x_n$ (che spesso sono numeri ma non sempre) si chiama *dataset* e non è un insieme in senso matematico perché le ripetizioni dei valori sono ammesse e non vanno elise.

Si considerano dataset, via via più trattabili matematicamente:

• *nominali*: per esempio con valori in { pari, dispari }
pari, pari, dispari, pari, dispari, dispari (è un dataset)
oppure con valori in { integro, spezzato, bifido }
integro, spezzato, spezzato, bifido, integro, integro (è un dataset)

• *ordinali*: per esempio, per valutare un farmaco antidepressivo potremmo considerare questi valori non numerici
morte
grande peggioramento
medio peggioramento
lieve peggioramento
stabile
lieve miglioramento
medio miglioramento
grande miglioramento
e dopo il *trial clinico* avere un dataset di – per esempio – 172 valori di quel tipo.

Oppure, altro caso:
sono completamente d'accordo,
sono d'accordo,
sono parzialmente d'accordo,
non sono d'accordo,
non sono assolutamente d'accordo.

BOZZA - DRAFT

• *numerici*: 3, 1, 2, 4, 6, 1000, 6. Quelli che più ci interessano.

Anche nei dataset ordinali si potrà fissare una corrispondenza con numeri, ma non avrà alcun senso fare la media fra 3 soggetti con: lieve miglioramento, medio miglioramento, morte. Se invece la "perdita infinita" viene esclusa, una media in qualche modo si potrà anche fare, con opportuna trasformazione in valori numerici. (E alla fin fine – di necessità in virtù – qualcosa riusciranno a fare anche con la perdita infinita).

Fra le variabili nominali si distinguono quelle *dicotomiche* ovvero *binarie*, con 2 soli valori, come pari/dispari, vivo/morto, successo/fallimento, 0/1: Link->

Considereremo 2 capitoli della Statistica Descrittiva:

◇ le *rappresentazioni grafiche*: Lezioni 20, 21;

◇ le *statistiche di sintesi*, funzioni $f(x_1, ..., x_n)$ con un numero n a priori indeterminato di argomenti. Inizieremo da queste.

Ecco alcune statistiche di sintesi per un dataset:

• il minimo: $\min(x_1, ..., x_n)$ ← per un dataset almeno ordinale;

• il massimo: $\max(x_1, ..., x_n)$ ← per un dataset almeno ordinale;

• la somma: $x_1 + ... + x_n$ ← per un dataset numerico.

Il significato delle soprastanti statistiche di sintesi è ovvio. Si calcolino quei 3 valori per questo dataset: $\frac{1}{\pi}$, $\frac{2}{7}$, 0.3, $\frac{1}{\sqrt{10}}$, e, e^{-1}, $\frac{1}{3}$.

Per esempio (ipotetico, di farmacologia veterinaria) è immensamente più significativo dire che fra 100 000 ostriche sono state trovate perle con un minimo di 0 e un massimo di 4 in ogni esemplare per un totale di 41 320 perle, piuttosto che elencare il numero di perle 0, 0, 1, 0, 2, 0, 0, 0, 0, 1... di ciascuno degli esemplari. Abbiamo riassunto 100 000 numeri con 3, minimo massimo e somma, e capiamo perfino meglio la situazione che con pagine di numeri! E soprattutto possiamo meglio *confrontarla* con la situazione di un'altra vasca di 100 000 ostriche, magari nutrite/trattate diversamente, che analogamente riassumeremo con quei 3 indici.

BOZZA - DRAFT

Più in dettaglio considereremo queste altre statistiche di sintesi:

• gli indici di posizione:

  ◦ quelli che vorrebbero riassumere in 1 solo valore "medio" il complesso dei dati, e saranno l'argomento di questa Lezione 19.10;

  ◦ quelli che vorrebbero riassumere in 5 valori o corrispondentemente 1 diagramma il complesso dei dati, Lezione 21;

• gli indici di dispersione ovvero variabilità, Lezione 22, che vorrebbero quantificare con 1 numero la non omogeneità di dati numerici.

◇ Esiste anche 1 indice numerico, la *skewness*, che con una complicata formula misura quantitativamente l'asimmetria dei dati, che però noi tratteremo solo qualitativamente.

Ed esistono altri indici che non tratteremo (come la *curtosi*).

19.5 Medie

Media (aritmetica). Da ora consideriamo un dataset $\{x_1, ..., x_n\}$.

$$M(x_1, ..., x_n) := \frac{x_1 + x_2 + ... + x_n}{n}$$

Esempio.
Sul covid-19 in Italia, leggiamo in
https://www.epicentro.iss.it/coronavirus/bollettino/Report-COVID-2019_4_ottobre.pdf

> Dati al 7 settembre 2020 (...) da 4190 deceduti per i quali è stato possibile analizzare le cartelle cliniche. Il numero medio di patologie osservate in questa popolazione è di 3,4

Facile, ottimo anche per i voti, ma il reddito medio di questi

$$3, 1, 2, 4, 6, 1000, 6$$

è 146, non poi così espressivo della situazione globale, a causa dell'*outlier* 1000, valore anomalo ovvero aberrante.

BOZZA - DRAFT

Tratteremo in seguito la questione dei valori anomali ma anticipiamo che talvolta vengono semplicemente *eliminati*.

Media geometrica. Per n numeri positivi (invece la media aritmetica non lo richiede), è la radice n-esima del loro prodotto:

$$GM(x_1, ..., x_n) := (x_1 \cdot ... \cdot x_n)^{\frac{1}{n}} .$$

Nell'esempio numerico soprastante ≈ 7.05; con WolpframAlpha
`geometric mean 3,1,2,4,6,1000,6`

Nota 1. Per quanto possa interessare il lettore, l'opinione dello scrivente è che più della metà delle volte in cui negli articoli scientifici si fa una media aritmetica, la media geometrica sarebbe stata più pertinente; e plausibilmente, in progresso di tempo la media geometrica sostituirà largamente la media aritmetica. Certo, la media

geometrica è meno semplice calcolarla, almeno a mano.

Nota 2. Alcuni Autori ritengono che la media geometrica sia meno sensibile agli outlier della media aritmetica, ma questo non può essere affermato con certezza matematica, dipende da ogni singolo dataset.
Riguardo alle applicazioni non lontane dalle Scienze Biomediche, leggiamo in Wikipedia, l'enciclopedia libera, alla voce "Geometric mean":

> starting from 2010 the United Nations Human Development Index did switch to this mode of calculation, on the grounds that it better reflected the non-substitutable nature of the statistics being compiled and compared:
>
> > The geometric mean decreases the level of substitutability between dimensions [being compared] and at the same time ensures that a 1 percent decline in say life expectancy at birth has the same impact on the HDI as a 1 percent decline in education or income. Thus, as a basis for comparisons of achievements, this method is also more respectful of the intrinsic differences across the dimensions than a simple average.

BOZZA DRAFT

In pratica, se un Paese porta l'aspettativa di vita da 40 a 60 anni, questo verrà ben rilevato facendo la media geometrica con il reddito pro-capite; se invece si facesse la media aritmetica, l'incremento di reddito da 4 000 a 4 040 dollari peserebbe di più: l'incremento è di 40 invece che 20, ma è solo l'1%, mentre da 40 a 60 è del 50%. La media geometrica appare migliore quando dobbiamo riassumere dati che variano su scale molto diverse. E anche in altri casi.

Media (aritmetica) ponderata (o pesata). Dati dei *pesi* $a_1, ..., a_n$, di somma 1, la media pesata del dataset $\{x_1, ..., x_n\}$ è $a_1x_1 + ... + a_nx_n$.
(Detto semplificatamente) pesi a_k maggiori di $\frac{1}{n}$ sono associati a dati che si vuol far "pesare" di più nella considerazione finale.

Per esempio nella media pesata

$$0.4\,voto_Matematica + 0.4\,voto_Chimica + 0.2\,voto_Marketing$$

il voto nell'esame di Marketing pesa/conta metà di ciascuno degli altri voti.

Per il lettore interessato: esistono varie altre "medie".

Mediana. Il numero centrale dei dati riordinati, adesso 4: $1, 2, 3, \underline{4}, 6, 6, 1000$. E se i dati sono in numero pari, si considera la media dei 2 centrali. La mediana è definita anche per valori *ordinali* almeno se il numero di dati è dispari.

Leggiamo (11 novembre 2022) in https://www.epicentro.iss.it/coronavirus/sars-cov-2-decessi-italia

> al 10 gennaio 2022 (...) Complessivamente, le donne decedute dopo aver contratto infezione da SARS-CoV-2 hanno un'età più alta rispetto agli uomini (età mediane: donne 85 anni – uomini 80 anni).

Moda. Il valore più frequente. Nel nostro esempio è 6 ma in generale non è definita perché i numeri sono tutti diversi o perché 2 o più valori sono ripetuti ugualmente. Ecco per esempio un dataset *bimodale*: $6, 6, 6, 6, 7, 7.5, 7.5, 8, 8, 8, 8, 8.5, 8.5, 9, 9.5, 10, 10$.
La moda è definita anche per dati *nominali*, neppure ordinali, per esempio in ogni nazione c'è un cognome più frequente.

BOZZA - DRAFT

Confronto fra media, mediana e moda.
Su Wikipedia, l'enciclopedia libera, alla voce "Mode (statistics)" troviamo questo significativo esempio:
1, 2, 2, 3, 4, 7, 9; media=4, mediana=3, moda=2.

Leggiamo (11 novembre 2022) in https://www.epicentro.iss.it/coronavirus/sars-cov-2-decessi-italia

> al 10 gennaio 2022 (...) L'età media dei pazienti deceduti e positivi a SARS-CoV-2 è 80 anni (mediana 82 (...)).

Non ordinamento fisso di media, moda e mediana
Media, mediana e moda sono 3 valori che possono presentarsi in qualunque ordine fra loro: dipende dal dataset.

Media interquartile. Leggiamo su Wikipedia, l'enciclopedia libera, alla voce Central tendency:

> The method is best explained with an example. Consider the following dataset:
> 5, 8, 4, 38, 8, 6, 9, 7, 7, 3, 1, 6
> First sort the list from lowest-to-highest:
> 1, 3, 4, 5, 6, 6, 7, 7, 8, 8, 9, 38
> There are 12 observations (datapoints) in the dataset, thus we have 4 quartiles of 3 numbers. Discard the lowest and the highest 3 values:
> ~~1~~, ~~3~~, ~~4~~, 5, 6, 6, 7, 7, 8, ~~8~~, ~~9~~, ~~38~~
> We now have 6 of the 12 observations remaining; next, we calculate the arithmetic mean of these numbers:
> xIQM = (5 + 6 + 6 + 7 + 7 + 8) / 6 = 6.5
> This is the interquartile mean.
> For comparison, the arithmetic mean of the original dataset is
> (5 + 8 + 4 + 38 + 8 + 6 + 9 + 7 + 7 + 3 + 1 + 6) / 12 = 8.5
> due to the strong influence of the outlier, 38.

BOZZA - DRAFT

(Che 38 vada considerato outlier è opinabile).

Con artifici si definisce per un numero di dati non quadruplo.

Detto grezzamente, consideriamo il reddito medio delle persone medie, al netto di mendicanti e ricconi, che quelli ci sono comunque dappertutto.

ES. 2$_{\mu 2019}$

≈ Il carbonio risulta avere – almeno secondo alcuni Autori: non possiamo qua garantirlo in forma assoluta e fare Chimica o Fisica; si veda Wikipedia in inglese alla voce *Isotopes of carbon* – 12 isotopi non reperibili in natura (oltre a 3 reperibili in natura) con queste emivite approssimative:

$3{,}5 \times 10^{-21}$ s, 126,5 ms, 19,3 s, 20,364 min, 2,45 s, 0,747 s,
193 ms, 92 ms, 46,2 ms, 16 ms, <30 ns, 6,2 ms.

Dopo aver convertito minuti, millisecondi e nanosecondi in secondi con le note

formule

1 min=60 s, 1 ms=0,001 s, 1 ns=$0,000\,000\,001$ s,

determinare la media interquartile delle emivite.

SVOLGIMENTO

Con la conversione in secondi le emivite sono, nell'ordine iniziale dei dati,

$3,5 \times 10^{-21}$ s, 0,126 5 s, 19,3 s, 1 221,84 s, 2,45 s, 0,747 s,

0,193 s, 0,092 s, 0,046 2 s, 0,016 s, $< 0,000\,000\,030$ s, 0,006 2 s.

Ovvero, in ordine crescente, omettendo l'unità di misura,

$3,5 \times 10^{-21}$ → questo e il seguente potrebbero doversi scambiare

$< 0,000\,000\,030$ → vedi nota alla linea precedente

0,006 2

0,016

0,046 2

0,092

0,126 5

0,193

0,747

2,45

19,3

1 221,84.

I 12 valori ordinati sono 4 terne di valori, ed eliminate la prima terna (coi valori più bassi) e l'ultima (coi valori più alti), i 6 valori centrali sono

0,016

0,046 2

0,092

0,126 5

0,193

0,747

e la loro media è il valore cercato, la media interquartile, in secondi:

$$\mathrm{IQM} = \frac{0,016 + 0,046\,2 + 0,092 + 0,126\,5 + 0,193 + 0,747}{6} = \frac{1,220\,7}{6} =$$

$$\boxed{\approx 0,203 \text{ s}}$$

(IQM (acronimo di InterQuartile Mean) è un simbolo classicamente usato per la media interquartile; anche iqm e x_{IQM}).

19.6 Inesistenza di una media migliore di tutte

Non esiste una media che sia la migliore di tutte in ogni caso.

BOZZA - DRAFT

Eppure avremo sempre bisogno di una qualche media, mica possiamo fermarci a dare liste di migliaia o milioni o miliardi di numeri: è necessario quel passo avanti che riassume i dati in 1 numero medio.

Se veramente siamo interessati a produrre
(1) un valore in qualche modo "medio" dei dati
(2) che prescinda dai casi/valori *eccezionali*
allora
- mediana e
- media interquartile

– sono decisamente valide
+ ma solo la mediana è *sempre* facilmente calcolabile
∗ invece la media aritmetica richiede di escludere gli outlier:
  ◇ se i dati sono pochi si può fare a mano ma lasciandoci dubbi
  ◇ se i dati sono moltissimi non si può proprio fare a mano però
  > si può fare informaticamente con varie regole che definiscono gli outlier, ma in diversi modi a seconda degli Autori.

BOZZA - DRAFT

19.7 Disilludersi sulle medie anche senza outlier

Non ci si illuda: non esiste alcun tipo di statistica di sintesi che riassuma bene con 1 solo numero ogni dataset, neppure in assenza di outlier. Spesso la mediana rappresenta bene il soggetto "tipico", ma non sempre: in un cortile con 6 cani e 6 galline, la media e la media interquartile e la mediana del numero di zampe per animale è 3. Che non rappresenta alcun caso tipico. (E la media geometrica è ≈ 2.83, di male in peggio). Nemmeno l'*unimodalità* rappresenta una garanzia per la mediana, pur in generale così valida: con 1 grillo, 5 cani e 6 galline

$$6,\ 4,\ 4,\ 4,\ 4,\ 4,\ 2,\ 2,\ 2,\ 2,\ 2,\ 2 \quad (moda = 2)$$

la mediana è ancora 3.

Così, il numero medio di tumori alla prostata e all'utero, di I, II e III grado, in un determinato gruppo di soggetti non ha nessun valore statistico: bisognerà fare statistiche separate per genere.

Nota 3. WolframAlpha calcola online
la media con `mean` seguito dalla lista di numeri: link->
la mediana con `median` seguito dalla lista di numeri: link->
la media geometrica con `geometricmean` seguito dalla lista di numeri: link->

19.8 Stratificazione per età in Medicina e Farmacia

Nelle statistiche di Medicina e Farmacia, i dati relativi all'essere umano talvolta vengono variamente stratificati per età e tipicamente sono presentati in:
0-4
5-9
10-14
...
oppure a decenni
10-19
20-29
...
o anche, a ventenni,
20-39
40-59
60-79
...
Si considerano gli anni compiuti. e allora chi avesse 19 anni e 11 mesi sta nella classe 10-19.

BOZZA - DRAFT

19.9 Sul trasformare dati ordinali in dati numerici

La trasformazione di dati ordinali in numerici introduce elementi di arbitrarietà. Vediamo un esempio.

Supponiamo che un'organizzazione raccolga dati sulla soddisfazione degli utenti sulle prestazioni di un suo organo: per esempio un'università, sull'insegnamento di un suo docente – caso reale in Italia. L'utente può scegliere, per valutare il suo grado di soddisfazione, fra

no
più no che sì
più sì che no
sì

Un'università trasforma i valori ordinali in numerici con lo schema

2
5
7
10

e un'altra con quest'altro, altrettanto legittimo:

1
2
3
4

BOZZA - DRAFT

Consideriamo i giudizi sui famosi Professori Pinco e Pallino, dati dai loro 2 studenti per ciascuno:

Pinco: più no che sì, sì: 5+10=15 oppure 2+4=6

Pallino: più sì che no, più sì che no: 7+7=14 oppure 3+3=6.

Col primo schema Pinco figura meglio di Pallino, 15 a 14, e col secondo sono pari, 6 a 6.

19.10 Le medie creano illusioni percettive

In https://www.epicentro.iss.it/coronavirus/bollettino/Report-COVID-2019_4_ottobre.pdf leggiamo

> Dati al 7 settembre 2020 (...) L'età media dei pazienti deceduti e positivi a SARSCoV-2 è 80 anni

e coi dati al 5 ottobre 2021 è ancora 80 anni. Eppure dall'aprile 2021 al settembre 2021 l'età media dei morti è molto più bassa, come si vede in Figura 3 di https://www.epicentro.iss.it/coronavirus/bollettino/Report-COVID-2019_5_ottobre_2021.pdf. Addirittura, dal grafico vediamo che nel giugno 2020 l'età media era circa 84 anni e nel giugno 2021 (poco meno di) 76. Niente di strano: i morti del periodo estivo sono pochi nel 2020 e molti di più ma comunque pochi nel 2021, la media la fanno principalmente i morti di marzo-aprile 2020 e poi novembre-maggio.

BOZZA - DRAFT

ESERCIZI SULLA LEZIONE

ESERCIZIO$_{\mu 2018}$
$*$ Considerato il seguente dataset

19.68 19.20 19.63 18.94 18.81 18.10 18.63 18.85 0.01 19.51 19.54

che possiamo supporre misurazioni di parametri corporei, si determini la mediana dopo avere eliminato un outlier.

SVOLGIMENTO
Chiaramente 0.01 è l'outlier preannunciato. (Potrebbe ragionevolmente provenire da un momentaneo malfunzionamento di una macchina che ha prodotto i dati).
I 10 dati rimanenti riordinati in modo crescente sono

18.10 18.63 18.81 18.85 18.94 19.20 19.51 19.54 19.63 19.68.

I 2 centrali ovvero mediani sono il 5° e il 6°, eliminando 4 da ogni parte:

18.94 19.20

la cui media aritmetica è il risultato cercato:

$$\boxed{19.07}$$

BOZZA - DRAFT

(Se non si eliminasse l'outlier – talvolta non è poi così evidente che qualche valore vada scartato – la mediana sarebbe alquanto simile, 18.94, mentre nei 2 casi le medie aritmetiche sono molto più diverse fra loro, rispettivamente 19.089 e 17.3536; la mediana ha il pregio di risentire poco degli outlier, il che in casi complessi con migliaia o milioni di dati, e magari nessuna certezza su quanti e quali sarebbero da considerare outlier, è molto significativo).

ESERCIZIO$_{\mu 2019}$
$\approx$ Si supponga di avere questi dati di un ospedale in anni successivi, relativi ai consumi di un certo farmaco:

1907 4257
1908 3956
1909 3936
1910 4183
1911 4114
1912 4525
1913 4188
1914 4111

1915 4404
1916 4180
1917 0
1918 0
1919 4361
1920 4035

Dopo aver eliminato gli outlier, determinare la media interquartile dei dati sul consumo del farmaco. (Si approssimi all'intero più vicino).

SVOLGIMENTO

Gli outlier sono i due 0 (verosimilmente dovuti alla guerra).

Ordiniamo il dataset rimanente, che ha 12 elementi:

3936, 3956, 4035, 4111, 4114, 4180, 4183, 4188, 4257, 4361, 4404, 4525

Eliminiamo i primi 3 e gli ultimi 3 valori

$$\Big[3936, 3956, 4035, \Big]4111, 4114, 4180, 4183, 4188, 4257\Big[, 4361, 4404, 4525\Big]$$

e la media dei 6 rimanenti è

$$\approx 4172.17$$

e approssimiamo come richiesto:

$$\boxed{4172}$$

BOZZA - DRAFT

Nota. Mostriamo la straordinaria insensibilità della media interquartile agli outlier rispetto alla media aritmetica. I 12 valori non nulli provenivano da una variabile aleatoria più o meno normale, e poi erano stati aggiunti 2 zeri in corrispondenza ad anni di guerra. Ora, in questo caso è ben evidente che si tratta di outlier per una serie di motivi:

molto diversi dagli altri valori
valori esattamente 0
corrispondenza con anni di guerra.

Ma in altri casi non è così evidente quali valori considerare outlier, per eliminarli "a mano". Allora potremmo fare la media aritmetica di tutti i valori, che però verrebbe fortemente influenzata dagli outlier. Invece la media interquartile dei 12 valori degli ultimi 12 anni, dal 1909 al 1919, senza escludere gli zeri,

1909 3936
1910 4183
1911 4114
1912 4525
1913 4188

1914 4111
1915 4404
1916 4180
1917 0
1918 0
1919 4361
1920 4035

è 4135, alquanto simile a quella di prima. Gli zeri sono finiti nelle porzioni scartate.

Insoma la media interquartile ci mostra di cogliere il *vero valore "medio"* della variabile aleatoria retrostante al fenomeno quando esso avviene *normalmente*, e lo fa in un modo *automatico*, con una formula, che non richiede l'esclusione a mano degli outlier, cosa impossibile se i dati sono milioni.

La media aritmetica 3503 invece è molto minore (essenzialmente a causa dei 2 zeri) della media 4188 dei 12 valori non nulli iniziali (e qua si vede la sensibilità della media aritmetica agli outlier).

Sarebbe bello mostrare che similmente avverrebbe considerando tutti i 14 valori iniziali, ma sfortunatamente la definizione di media interquartile per campioni con un numero non quadruplo di elementi non è facilissima per un calcolo a mano.

BOZZA – DRAFT

20 Bar chart, istogrammi, e altri diagrammi

20.1 Diagrammi a torta

Il *diagramma a torta* ha un ovvio significato: gli angoli o equivalentemente le aree sono proporzionali contemporaneamente ai valori assoluti o alle frazioni percentuali da rappresentare. (La proporzionalità agli uni è equivalente alla proporzionalità alle altre). Si veda in questo link-> un diagramma a torta di interesse farmaceutico coi valori assoluti riportati (le percentuali no, sono lasciate all'occhio del lettore che vede le fette della torta; può essere un'utile esercizio calcolarne qualcuna).
(Più complessa è la situazione in cui oltre alle percentuali relative a 2 o più casi, il che necessita di 2 diagrammi a torta, si voglia rappresentare anche un altro dato per ciascuno dei casi, e allora si faranno diagrammi a torta di diverse dimensioni).
Gli angoli si misurano col goniometro; ma anche a occhio si può fare qualcosa.
Il diffuso software Excel fa i diagrammi a torta.
Funzionano veramente bene solo se i valori da rappresentare sono molto pochi, magari 2 o 3, e specialmente se sono molto diversi fra loro, come 66% e 32% e 2% – ma non di più di 2 ordini di grandezza: valori come 90%, 9.9%, 0.07% e 0.03% (l'ultimo ha solo circa 1 decimo di grado e il penultimo 2 e mezzo) non saranno concretamente apprezzabili su un normale foglio a stampa, o schermo di computer.

BOZZA - DRAFT

Spesso si raggruppano in una classe i valori più piccoli, o più grandi.
Si veda un esempio in https://www.epicentro.iss.it/coronavirus/sars-cov-2-decessi-italia-2 dove in una stessa classe sono i casi "3 o più".

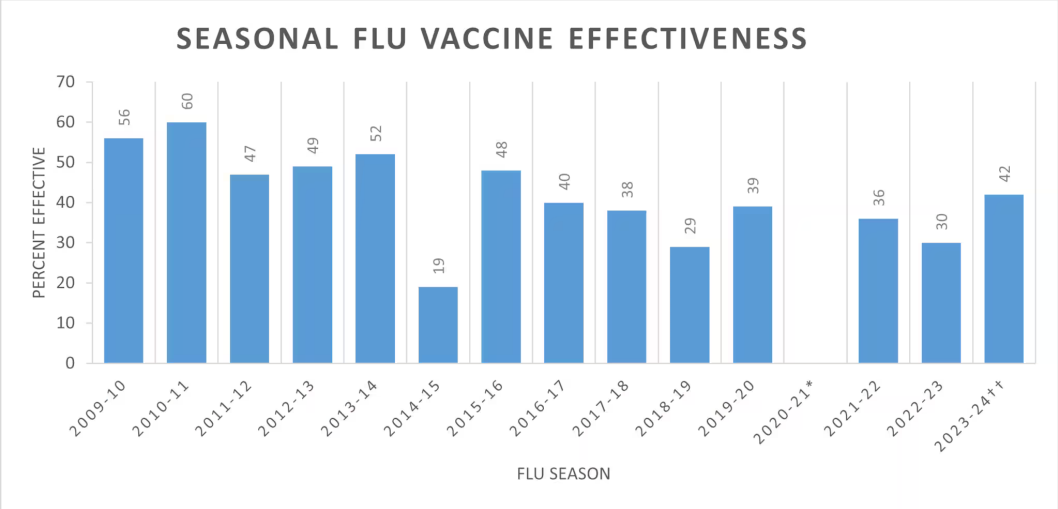
Si evitino come la peste gli ingannevoli diagrammi a torta tridimensionali in rappresentazione prospettica.

20.2 Catastrofe Statistica: dati mancanti e drop out

Una catastrofe della Statistica è il fenomeno dei dati mancanti.
Si veda un esempio nella prima figura del paragrafo seguente.

Un caso particolare di dati mancanti è quello dei *drop out*, partecipanti che abbandonano un trial clinico prima del suo completamento: sono una catastrofe della Statistica Medica. Ragioni il Lettore: se un cosmetico sperimentale è stato dato a 1000 persone, e dopo un mese rispondono a un questionario 400 persone, e magari risulta che mediamente sono soddisfatte, che valore ha quel risultato statistico? Magari sono alle Seichelles a godersi i risultati, o sono in gravi condizioni e non ci rispondono più al telefono; o sono morti?

20.3 Istogrammi a barre o bar chart, e istogrammi



https://www.cdc.gov/flu-vaccines-work/php/effectiveness-studies/index.html

* 2020-2021 flu vaccine effectiveness was not estimated due to low influenza virus circulation during the 2020-2021 flu season.

BOZZA - DRAFT

I diagrammi a colonne ovvero bar chart ovvero istogrammi a barre, e gli istogrammi – purtroppo con ambiguità nominalistiche nei vari testi – sono diagrammi per la visualizzazione di dati. Nei bar chart l'altezza di ogni colonna – o la lunghezza se disposta orizzontalmente – rappresenta un valore, negli istogrammi l'area rappresenta un valore:

bar chart – altezza ovvero lunghezza – valore

istogramma – area – valore

Spesso in italiano chiamano istogrammi quelli che in questo testo più precisamente chiamiamo bar chart.

Mostriamo un istogramma, nel senso di questo testo (area=valore), e poi ci occuperemo maggiormente di bar chart: LINK->
Si noti che la grandezza sull'asse delle ascisse è continua e non potrebbe essere che così.

Torniamo dunque ai bar chart.

Esempio. Dati rilevati a Chimica e Tecnologie Farmaceutiche di Trieste il 20/11/2023, sul numero di gatti e di cani che gli studenti (presenti in aula) hanno:

```
G: 3 2 0 0 2 0 0 0 0 1 0 0 1 0 0 0 0 1 1 1 0 0 0 0 0 0 1 0 0 3 1
C: 0 0 2 0 1 2 0 1 0 0 0 0 0 1 2 0 1 0 1 0 2 0 0 0 1 2 1 2 2 2 1
```

Istogramma a barre delle frequenze per il dataset dei gatti di CTF:

```
  0 gatti:  studenti 22 | ......................
   1 gatto:  studenti 6 | ......
   2 gatti:  studenti 2 | ..
   3 gatti:  studenti 2 | ..
```

BOZZA - DRAFT

Si noti che è possibile che una colonna rappresenti, per così dire, un valore medio di dati già considerati in altre colonne, per esempio in questo LINK->

Le varie barre possono avere colori che le associano fra loro: LINK->

Ma si faccia attenzione nell'interpretare le statistiche perché molto dipende da come si raggruppano i dati. Nel linkato diagramma sulle cause di morte il cancro compare poco perché è stato disaggregato in vari tipi di cancro. Considerandoli tutti, il cancro è la seconda causa di morte, dopo le malattie cardiocircolatorie. Addirittura qua disaggregando i tumori in tanti tipi diversi fa apparire il covid come prima causa di morte in Italia, datato 3 maggio 2020: LINK->

If you torture the data long enough, it will confess to anything.
(In questo LINK-> è in articolo su PubMed).

Si faccia attenzione nell'interpretare i bar chart in scala logaritmica perché le altezze rappresentano sì i valori ma non in modo proporzionale ad essi: LINK->

I bar chart possono rappresentare anche valori negativi: LINK->

Naturalmente si faccia attenzione a distinguere i valori assoluti dai valori relativi; per esempio di vari stati o regioni si possono considerare i morti di covid come numero assoluto, o come morti per milione o centomila abitanti. Vediamo un esempio in cui i valori sono già opportunamente scalati sul migliaio di abitanti in questi spettacolari diagrammi a bolle interattivi: LINK->

Si vedano vari tipi di bar chart di interesse specificamente farmaceutico in questo LINK->
Quella sorta di pelucchi che si vedono su certi bar chart rappresentano le deviazioni standard, concetto che vedremo, e che in sostanza rappresentano la variabilità della grandezza rappresentata, all'interno del campione da cui proviene.

BOZZA - DRAFT

ESERCIZIO$_{\mu 2019}$
≈ In un articolo scientifico[96] della rivista Neuropsychopharmacology leggiamo

> Patients were treated two to three times per week with high-frequency MST (i.e., 100 Hz) (N=24), medium frequency MST (i.e., 60 or 50 Hz) (N=26), or low-frequency MST (i.e., 25 Hz MST) (N=36)

Si disegni un istogramma a barre (bar chart) coi valori N_i denominando N l'asse delle ascisse, con barre denominate (lo si scriva entro esse)

high-frequency MST

medium frequency MST

[96] *Magnetic seizure therapy (MST) for major depressive disorder*, Neuropsychopharmacology (5 September 2019), Zafiris J. Daskalakis, Julia Dimitrova, Shawn M. McClintock, Yinming Sun, Daphne Voineskos, Tarek K. Rajji, David S. Goldbloom, Albert H. C. Wong, Yuliya Knyahnytska, Benoit H. Mulsant, Jonathan Downar, Paul B. Fitzgerald & Daniel M. Blumberger

low-frequency MST
e all'estremità di ogni barra si scriva la frequenza relativa, cioè N_i/N_{tot}, espressa percentualmente, con 1 decimale.

- Considerato il dataset

$$2, 3, 1, 4, 1, 5, 3, 4, 7, 1, 2, 4, 5, 9, 4, 8$$

se ne rappresenti il bar chart. Poi si rappresenti l'istogramma con intervalli $[0, 2.5[$, $[2.5, 5[$, $[5, 7.5[$, $[7.5, 10[$, e poi con intervalli $[0, 2.5[$, $[2.5, 5[$, $[5, 10[$.

20.4 Eventuali asimmetrie nei dataset: skewness

Una distribuzione coi dati più "addensati" verso i valori bassi che quelli alti si dice *right skewed*, e si intuisce com'è una distribuzione *left skewed*. (Queste *non* sono definizioni rigorose[97] ma permettono di decidere nella generalità dei casi non particolarmente "capricciosi").
Si provi con un diagramma a colonne coi dati

12.2 20%,
12.4 30%
12.6 25%
12.8 12.5%
13.0 5%
13.2 7%
13.4 3.5%
13.6 2%

BOZZA - DRAFT

20.5 Funzioni a campana varie

Facendo un bel po' di istogrammi a barre di dati presi dalla realtà sensibile, si troverà che spesso le barre si dispongono a formare più o meno una campana.

Praticamente dietro quasi ogni fenomeno della realtà sensibile c'è in qualche modo una funzione più

[97] Il lettore interessato cercherà sulla rete la *formula*, piuttosto complessa, che quantifica la skewness.

o meno a campana. Imparare a riconoscere queste configurazioni a campana aumenta enormemente la comprensione della realtà.

La campana può avere 2 significati principali: una densità, la classica
“distribuzione più o meno a campana",
coi suoi casi estremi rari e quelli medi più comuni,
oppure,
se in ascissa abbiamo il tempo, talvolta rappresenta (quantitativamente) la classica
“parabola della vita",
valida anche per una comunità microbica, per la potenza dell’Impero Romano, per la concentrazione di un farmaco immesso nel sangue, il numero di malati in un’epidemia, o quant’altro: sorgere, ascendere, declinare, finire.

BOZZA - DRAFT

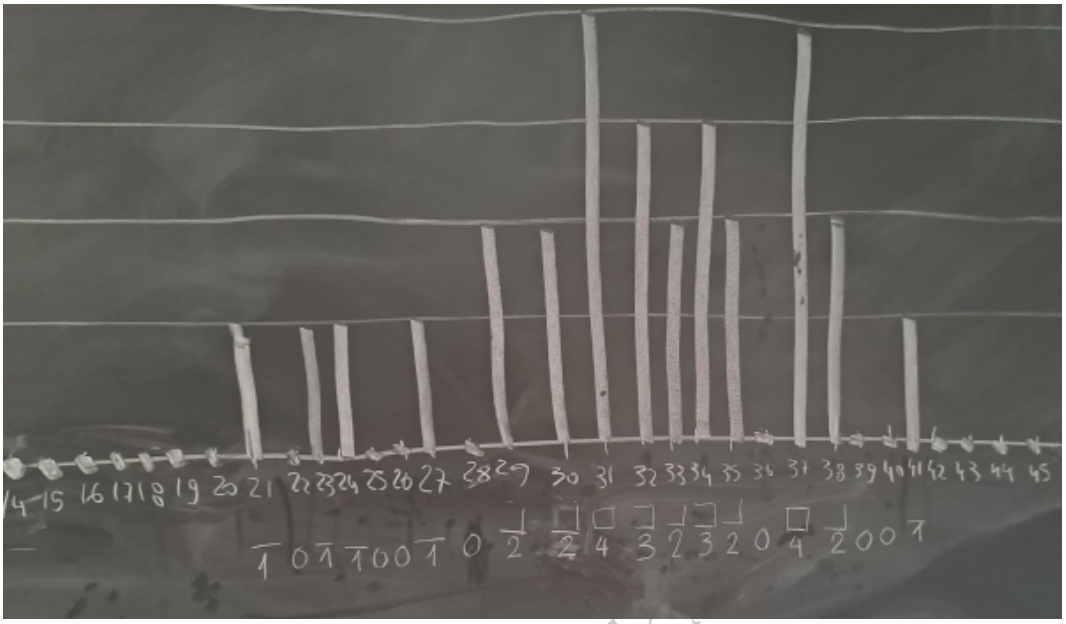


Figure 26: Qual era l'età della madre al momento della tua nascita? Anni compiuti, numero intero. Rilevazione in aula (2023) con questionario anonimo. Anni madre – frequenza: 21 1, 22 0, 23 1, 24 1, 25 0, 26 0, 27 1, 28 0, 29 2, 30 2, 31 4, 32 3, 33 2, 34 3, 35 2, 36 0, 37 4, 38 2, 39 0, 40 0, 41 1.

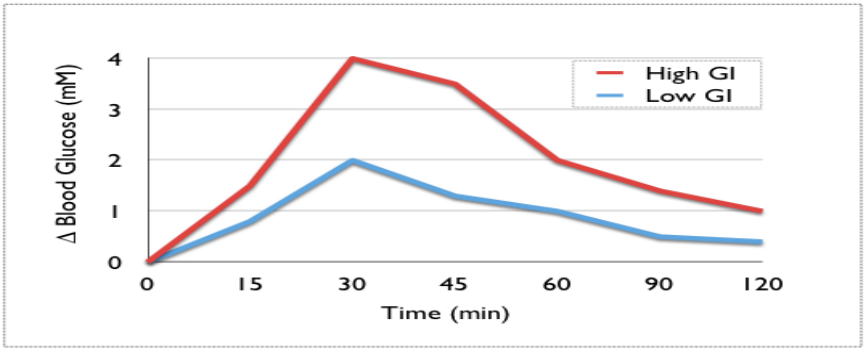
BOZZA DR

Qua stiamo parlando di curve "più o meno a campana" mentre con "curva a campana" *di solito* si intende proprio la *campana gaussiana*, grafico della *densità normale standard* $\frac{1}{\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{t^2}{2}}$, la forma più "pura" di curva a campana del primo tipo.

Curve più o meno a campana del primo tipo, cioè densità, (della Microbiologia) sono in questo link->.

Curve più o meno a campana del secondo tipo, cioè evoluzioni di una quantità nel tempo, sono in questa figura.[98]

[98] Immagine di pubblico dominio tratta da Wikimedia Commons.



Altre, relative ad epidemie di influenza, sono queste[99]
?????FIGURA MANCANTE
Un'altra, (della Microbiologia) è in questo link-> e un altro (della Farmacocinetica) in questo link-> e un altro a pagina 44 di questo link->.

BOZZA - DRAFT

Leggiamo nel Fedone, XXXIX (IV secolo a.C.), di Platone, citato in Wikipedia, l'enciclopedia libera, alla voce *Variabile casuale*:

> Credi forse che sia tanto facile trovare un uomo o un cane o un altro essere qualunque molto grande o molto piccolo o, che so io, uno molto veloce o molto lento o molto brutto o molto bello o tutto bianco o tutto nero? Non ti sei mai accorto che in tutte le cose gli estremi sono rari mentre gli aspetti intermedi sono frequenti, anzi numerosi?

Nota. Naturalmente nelle Scienze Applicate ricorrono anche grafici ben lontani dall'avere una forma a campana, pur così ubiqua. Già abbiamo visto sigmoidi e sinusoidi... Ma la realtà sensibile è ancora più ricca, per esempio si veda il grafico a questo LINK->.

[99] Tratte da Mid-season real-time estimates of seasonal influenza vaccine effectiveness in persons 65 years and older in register-based surveillance, Stockholm County, Sweden, and Finland, January 2017. Euro Surveill. 2017 Feb 23;22(8). pii: 30469. doi: 10.2807/1560-7917.ES.2017.22.8.30469. By Hergens MP et al. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5356437/. "This is an open-access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0) Licence. You may share and adapt the material, but must give appropriate credit to the source, provide a link to the licence, and indicate if changes were made." LINK->

ESERCIZI SULLA LEZIONE 20

20.5.1 Esercizio risolto a – Bar chart

$_{\mu}$ *

Consideriamo il modello di un fenomeno, e per fissare le idee diciamo che è un'epidemia. (Anche se quello che considereremo non è il miglior modello proprio per le epidemie). Il numero di morti dell'epidemia sia modellizzato da

$$x_n = \left\lfloor \frac{20\, n^2}{n\,!} \right\rfloor \qquad n = 1, 2, ...$$

essendo n il numero di giorni dall'inizio ed indicando il simbolo $\lfloor ... \rfloor$ la parte intera (quella che "toglie i decimali" ai numeri positivi, per esempio $\lfloor 3.14 \rfloor = 3$). Fare un istogramma a barre fino all'ultimo giorno con morti (o per meglio dire fino al giorno precedente del primo giorno con 0 morti).

BOZZA - DRAFT

SVOLGIMENTO

Viene lo standard del punto decimale. (Nello svolgimento e già nel testo del quesito, in cui si trova il numero 3.14).

$$x_1 = \left\lfloor \frac{20 \cdot 1 \cdot 1}{1} \right\rfloor = 20$$

$$x_2 = \left\lfloor \frac{20 \cdot 2 \cdot 2}{1 \cdot 2} \right\rfloor = 40$$

$$x_3 = \left\lfloor \frac{20 \cdot 3 \cdot 3}{1 \cdot 2 \cdot 3} \right\rfloor = 30$$

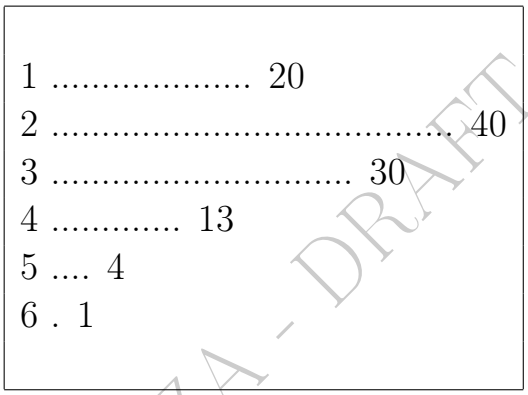
$$x_4 = \left\lfloor \frac{20 \cdot 4 \cdot 4}{1 \cdot 2 \cdot 3 \cdot 4} \right\rfloor = \left\lfloor \frac{40}{3} \right\rfloor = \lfloor 13.33... \rfloor = 13$$

$$x_5 = \left\lfloor \frac{20 \cdot 5 \cdot 5}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} \right\rfloor = \left\lfloor \frac{100}{24} \right\rfloor = \lfloor 4.16... \rfloor = 4$$

$$x_6 = \left\lfloor \frac{20 \cdot 6 \cdot 6}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6} \right\rfloor = 1$$

$$x_7 = \left\lfloor \frac{20 \cdot 7 \cdot 7}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7} \right\rfloor = \left\lfloor \frac{140}{720} \right\rfloor = \lfloor 0.19... \rfloor = 0$$

e abbiamo raggiunto lo 0 previsto. Il numero di morti x_n calcolato dal modello poi non si risolleva da 0 com'è dimostrabile con passaggi facili però laboriosi, ma questa osservazione non ci serve perchè nel testo è detto di fermarsi al giorno precedente del primo giorno con 0 morti, che è il giorno numero 7, come abbiamo trovato. Allora l'epidemia dura 6 giorni, e possiamo produrre l'istogramma a barre.



BOZZA - DRAFT

(Che molto meglio sarà rappresentare con le aste disposte verticalmente).

Nota. Ecco una soluzione online su WolframAlpha:

LINK -> (con l'istruzione `DiscretePlot`)

Rispetto a quello dato nella risoluzione dell'esercizio, questo risultato di WolframAlpha

– ha il pregio della precisione

– ha il pregio delle barre disposte verticalmente

– ha il difetto che non scrive sulle barre i 6 valori effettivi.

Ecco un'atra soluzione online su WolframAlpha:

LINK -> (con le istruzioni `BarChart` e `Table`)

Rispetto a quello dato nella risoluzione dell'esercizio, questo risultato di WolframAlpha

– di nuovo, ha il pregio della precisione

– di nuovo, ha il pregio delle barre disposte verticalmente

– ha il pregio che le barre hanno forma di rettangoli e non segmenti

– di nuovo, ha il difetto che non scrive sulle barre i 6 valori effettivi

– ha il difetto che è piuttosto complessa da trovare, richiedendo l'istruzione `Table`

20.5.2 Esercizio risolto b – Istogramma

$_{\mu 2018}$ [*disegno*]
In relazione all'elemento sodio (Na), definite le cellule

ipo-sodiche: con contenuto di sodio fra 0 e < 1 000 unità

normo-sodiche: con contenuto di sodio fra 1 000 e < 10 000 unità

iper-sodiche: con contenuto di sodio fra 10 000 e < 20 000 unità

(unità che non specifichiamo, essendo ininfluente) un macchinario esamina 52 000 cellule (tutti numeri arrotondati per semplicità) trovandovi

6 000 ipo-sodiche, 36 000 normo-sodiche, 10 000 iper-sodiche.

Rappresentare la situazione con un istogramma. (Non è il bar chart).

BOZZA - DRAFT

SVOLGIMENTO

Nell'istogramma – che purtroppo alcuni Autori scambiano col bar chart ma qua veniamo avvertiti della differenza – le grandezze sono proporzionali alle aree dei rettangoli (invece nei bar chart alle lunghezze delle aste, eventualmente strisce rettangolari).

Avremo quindi 3 rettangoli, con basi che si estendono su un asse orizzontale

da 0 a 1000 lunga 1 000

da 1 000 a 10 000 lunga 9 000

da 10 000 a 20 000 lunga 10 000.

(Non vi è alcuna questione problematica fra valori compresi o esclusi).

Le altezze le otteniamo dividendo le aree

6 000, 36 000, 10 000 (o numeri ad essi rispettivamente proporzionali)

per le lunghezze delle basi corrispondenti, ottenendo:

6 unità
4 unità
1 unità

Come unità si può prendere per esempio 1 centimetro, o su carta a quadretti 1 o 2 o 3 quadretti, o anche altre unità.
Ecco l'istogramma, coi 3 rettangoli ed una legenda:

```
A----------
A---------- A=ipo-sodiche, 6 000
ABBBBBBBBB------ B=normo-sodiche, 36 000
ABBBBBBBBB------ C=iper-sodiche, 10 000
ABBBBBBBBB------
ABBBBBBBBBCCCCCCCCCC
!!----!-----!
0 1000      10000          20000
```

BOZZA - DRAFT

21 Quartili e box-plot

Vogliamo riassumere molti numeri in un modo più ricco rispetto alla sola media o mediana: aggiungeremo alla mediana 4 numeri, o, corrispondentemente, con quei 5 numeri faremo 1 diagramma.

BOZZA - DRAFT

21.1 Quartili e riassunto dei 5 numeri

Si abbia il dataset numerico $x_1, .., x_n$ e lo si riordini in modo crescente: $y_1, ..., y_n$ *sono gli stessi numeri ma in ordine crescente*.

Il valore intermedio,

o la media dei 2 intermedi se n è pari,

sappiamo che è la mediana,

che da adesso chiameremo anche *secondo quartile*,

e lo indicheremo con

$q_{0.5}$ oppure

$q_{1/2}$ oppure

$q_{50\%}$.

Diviso così il dataset in 2 parti uguali,

la mediana della prima parte si chiama *primo quartile*,

e lo indicheremo con $q_{1/4}$ oppure

$q_{0.25}$ oppure

$q_{25\%}$,

e la mediana della seconda parte si chiama *terzo quartile*,

e lo indicheremo con $q_{3/4}$ oppure

$q_{0.75}$ oppure

$q_{75\%}$.

Si ha poi il quartile di indice 0, che è il minimo del dataset,

q_0 oppure

$q_{0\%}$

e quello di indice 1, che è il massimo del dataset:

q_1 oppure

$q_{100\%}$

Si hanno così 5 quartili. Ma purtroppo la loro definizione non è perfettamente univoca nei vari testi e software[100] e inoltre, di fatto, spesso si trova denominato *quartile* l'insieme di valori *fra* 2 quartili – come sopra definiti – e in questo senso, i quartili sono 4.

I 5 numeri detti, $q_{0\%}$,...,$q_{100\%}$, sono una statistica di sintesi (vettoriale[101]) che si chiama *riassunto dei 5 numeri* (o *five number summary*).

Due esempi tratti da Wikipedia, l'enciclopedia libera, riscritti:

per un dataset $x_1, ..., x_{10}$, riordinato in $y_1, ..., y_{10}$,
i 5 numeri (quartili) sono nell'ordine, y_1, y_3, $\frac{y_5+y_6}{2}$, y_8, y_{10};

per un dataset $x_1, ..., x_{11}$, riordinato in $y_1, ..., y_{11}$,
i 5 numeri (quartili) sono nell'ordine, y_1, y_3, y_6, y_9, y_{11}.

POZZA - DRAFT

Applicazione: un valore è tanto o poco? Secondo Scimago, la rivista scientifica *Journal of Pharmacy and Pharmacology* ha – in un determinato momento – l'H index, che è un indicatore bibliometrico, di 107: è tanto o poco? Se avessimo la lista di tutte le riviste scientifiche catalogate per Pharmacology da Scimago, e ciascuna col suo H index, avremmo la risposta, salvo che sarebbe scarsamente leggibile per eccesso di numeri. Un riassunto dei 5 numeri ci permetterebbe di contestualizzare bene quel valore 107. Scimago fa una cosa del genere, magari non esattamente con l'H index (usa un suo indice interno, il *SCImago Journal Rank* ovvero *SJR indicator*) ma la sostanza è la stessa. Salvo che chiama primo quartile quello più alto, cioè dal terzo quartile (numero) al quarto quartile (numero). Scopriamo così che la rivista considerata, per la disciplina scientifica Pharmacology è solo nel terzo quartile nel 2018, non poi

[100] Leggiamo in questo link accademico: "Molti software hanno diversi algoritmi per calcolare i quantili."

[101] *Vettoriale* perché produce una cinquina di numeri e non un solo numero. È solo una questione definizionale.

così bene (e invece per la categoria Pharmaceutical Science è nel secondo, molto meglio). Invece la rivista *British Journal of Pharmacology* risulta nel primo quartile, il migliore, dal 1999 al 2018.

21.2 Box-plot ovvero diagramma a scatola e baffi

Il *box-plot* ovvero *diagramma a scatola e baffi* è un diagramma molto usato in ambito Biomedico per rappresentare tutti i quartili di un dataset, ovvero il riassunto dei 5 numeri, e magari anche gli outlier.

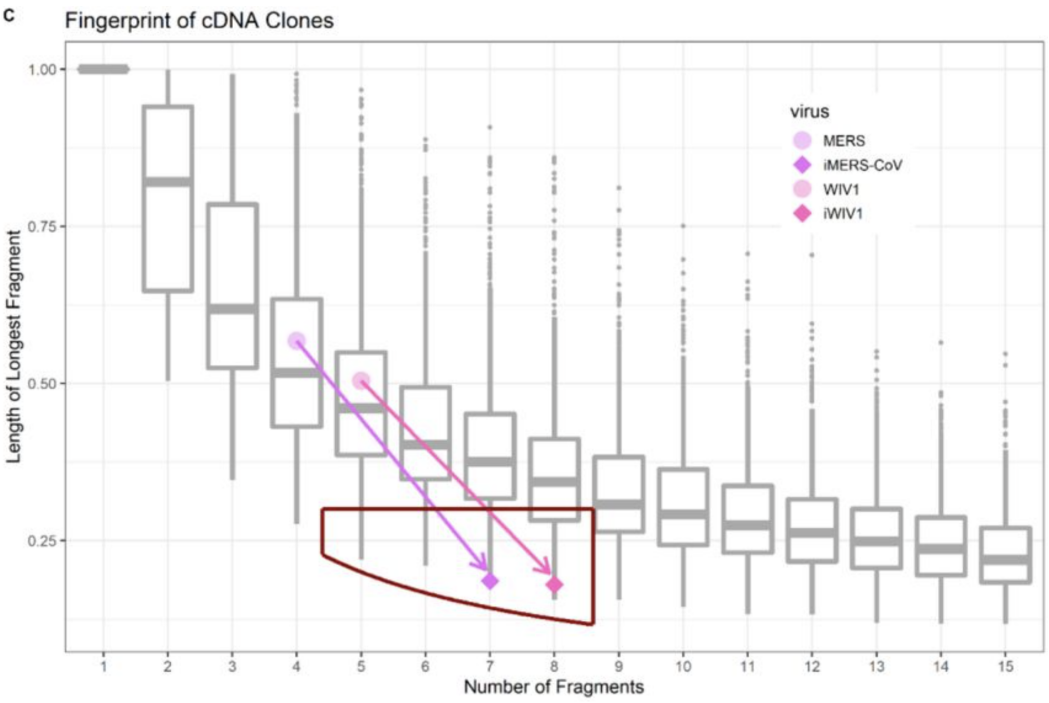


Figure 27: Figura 2 c in "Endonuclease fingerprint indicates a synthetic origin of SARS-CoV-2", by Bruttel, Valentin and Washburne, Alex and VanDongen, Antonius: in bioRxiv (2022), DOI 10.1101/2022.10.18.512756

Vediamo sulla rete alcuni esempi di pertinenza delle Scienze Mediche e Farmaceutiche seguendo questo `link->`, con immagini tratte da PubMed.

Precisiamo subito che di questa rappresentazione grafica
• esistono varianti, sulla lunghezza dei "baffi"[102], ma non solo[103]
• considereremo (quasi) solo la versione più semplice
• chi pubblica un box plot dovrebbe specificare[104] i dettagli su come intenderlo.

Leggiamo su Wikipedia, l'enciclopedia libera, alla voce "Diagramma a scatola e baffi":

> il diagramma a scatola e baffi (o diagramma degli estremi e dei quartili o box and whiskers plot o box-plot) è una rappresentazione grafica [...] Viene rappresentato (orientato orizzontalmente o verticalmente) tramite un rettangolo diviso in due parti, da cui escono due segmenti. Il rettangolo (la "scatola") è delimitato dal primo e dal terzo quartile, $q_{1/4}$ e $q_{3/4}$, e diviso al suo interno dalla mediana, $q_{1/2}$. I segmenti (i "baffi") sono delimitati dal minimo e dal massimo dei valori. In questo modo vengono rappresentati graficamente i quattro intervalli ugualmente popolati delimitati dai quartili.

BOZZA - DRAFT

Esercizio$_\mu$ Si disegni il box-plot dei primi 10 numeri primi, e poi dei primi 11.

Talvolta gli outlier vengono rappresentati isolati. Si faccia così per il dataset 1, $\frac{1}{2}$, $\frac{1}{3}$,..., $\frac{1}{11}$, $\sqrt{7}$; e poi senza $\frac{1}{11}$.

Su Nature.com, un sito web scientifico di altissimo livello, troviamo qui troviamo una descrizione dei box plot, in cui leggiamo

[102] Si veda per esempio questo link.
[103] Per esempio in questo link vediamo una crocetta: possiamo ragionevolmente ipotizzare che rappresenti un decesso.
[104] Per esempio in questo link leggiamo "the whiskers show the smallest and highest value within 1.5 box lengths from the box", che è una variante comune.

> The core element that gives the box plot its name is a box whose length is the IQR and whose width is arbitrary (...) A line inside the box shows the median, which is not necessarily central. The plot may be oriented vertically or horizontally (...) Whiskers are conventionally extended to the most extreme data point that is no more than 1.5 × IQR from the edge of the box (Tukey style) or all the way to minimum and maximum of the data values (Spear style). (...) The 1.5 multiplier corresponds to approximately (...) 99.3% coverage of the data for a normal distribution. Outliers beyond the whiskers may be individually plotted.

L'IQR sopradetto, *interquartile range*, che è $q_{0.75} - q_{0.25}$, la vedremo meglio in seguito; e la *normal distribution*, distribuzione normale, è la più tipica distribuzione delle Scienze Biomediche, e la vedremo in seguito, ma anticipiamo che la sua densità è la già introdotta campana gaussiana.

BOZZA - DRAFT

Applicazione. Abbiamo 1000 soggetti ciascuno con 1 valore fisiologico – glicemia, peso, qualunque cosa – su cui vogliamo intervenire, e gli diamo un farmaco. Diciamo, la glicemia.
Una settimana abbiamo 1000 valori nuovi, o piuttosto, in generale – nella realtà concreta – un po' meno (qualcuno può essere morto, o irraggiungibile, o non ne vuole più sapere di noi), diciamo 980.
La media dei valori iniziali e la media dei valori iniziali, certo sono un'indicazione importante.
Per fissare le idee supponiamo che il parametro si voleva diminuirlo.
Supponiamo che la media sia diminuita: bene!
Gli altri 4 numeri ci indicheranno più chiaramente come si è evoluta la situazione. (Del solo parametro controllato, ovvio, e questo è un problema generale della Medicina e della Farmacologia: *cosa* stiamo misurando? La glicemia, la mortalità, il benessere, i soldi... possono variare *indipendentemente*).
Per esempio potremmo avere una situazione iniziale i questo tipo

```
.......-------!OOOOOOO!OOOOOOOO!-----
```

e una finale di questo primo tipo o quest'altro tipo:

```
.........------!OOOOOO!OOOOOO!---
.....-------!OOOOOOOO!OOOOOOOOOO!------
```

con media ridotta e valori addensati nel primo caso

oppure

con media ridotta ma valori diradati nel secondo caso.
Non faremo qua Medicina, ma è ovvio che la seconda può essere pericolosa, nonostante il miglioramento della media.
Si noti – e questo è un problema generale – che molta parte di verità può sfuggirci nascosta nei soggetti che non abbiamo potuto misurare la seconda volta. Si potrebbe dire: bene, nel primo diagramma rappresentiamo solo i soggetti di cui abbiamo anche la seconda misurazione. Sì, ma il problema resta, perché la mancanza della seconda misurazione potrebbe essere causata in tutto o in parte proprio dal farmaco: morte, irreperibilità per cure all'estero, grave delusione del soggetto. (In parte, è il problema dei *morti per altra causa*).

Una nota finale nella pagina successiva.

BOZZA - DRAFT

21.3 I dati? Quali dati?

In Italia si usa raccogliere pochi dati, e parecchi di quei pochi nasconderli – per poi decidere su impressioni, ideologicamente.

Scrive il direttore Marco Cattaneo in un editoriale (settembre 2023) di Le Scienze (edizione italiana di Scientific American):

> era nata la campagna #DatiBeneComune, «per chiedere al Governo italiano di pubblicare in maniera aperta i dati sulla gestione della pandemia di COVID-19». Oggi la campagna, che ha superato i 60.000 firmatari, si rivolge ancora alle istituzioni per disporre di dati trasparenti sui principali temi di interesse dei cittadini, a partire dal Piano nazionale di ripresa e resilienza (PNRR).
>
> I dati, sempre i dati, croce di questo paese. Non dico che si debba avere l'attenzione spasmodica per i numeri degli Stati Uniti, dove c'è una maniacale ossessione per le statistiche persino nel baseball. Ma i numeri, i dati trasparenti e con il massimo grado di dettaglio possibile, dovrebbero essere lo spirito guida delle decisioni pubbliche, soprattutto quando ci sono in gioco investimenti importanti e strategie per il futuro.
>
> (...) L'analisi delle fonti a disposizione, a livello nazionale, mostra lacune che si sarebbe tentati di definire sconfortanti.
>
> «Il problema, peraltro, non è solo di mancanza del dato, ma anche, nei casi in cui esiste, di una ritrosia a rilasciare i dati», scrivono Faiella e Lavecchia. (Enfasi aggiunta)

BOZZA_DRAFT

La "sconfortante" "ritrosia a rilasciare i dati" sulla mortalità e lo stato vaccinale impedisce (2025) a tutt'oggi (!) a più di 4 anni dalle vaccinazioni di calcolare quanti giorni o mesi o anni di vita hanno (sperabilmente) guadagnato con la vaccinazione covid-19 coloro che si sono vaccinati, rispetto ai no-vax – il che genera sospetti nei *malfidenti*.

ESERCIZI SULLA LEZIONE 21

Esercizio$_{\mu}$ Si disegni il box-plot del dataset

$$x_k := \frac{1}{\pi - k} \qquad 1 \leq k \leq 10$$

rappresentando con un pallino l'outlier, escludendolo quindi dal calcolo del massimo.
Similmente poi con $1 \leq k \leq 11$.

21.3.1 Esercizio risolto a – Media interquartile

$_{\mu 2023} \approx$ Calcolare la media interquartile del dataset
54,071 117,512 19,097 2,345 282,012 0,112 0,023 0,016 2,150 630,013 2.848 3,410

SVOLGIMENTO
Viene usato lo standard della virgola decimale. Questo si vede dal numero 0,112 e altri. (Allora il punto in 2.848 è separatore delle migliaia, e questo della virgola decimale e del punto separatore delle migliaia è esattamente lo standard imposto in Farmacia dal Ministero della Salute italiano).

Il dataset ordinato in modo crescente, e scritto senza separatore delle migliaia, è
0,016 0,023 0,112 2,150 2,345 3,410 19,097 54,071 117,512 282,012 630,013 2848 Esso ha $n = 12$ cioè $4 \cdot 3$ elementi

••• ••• ••• •••

ed eliminando il primo e ultimo "quartile" dà
2,150 2,345 3,410 19,097 54,071 117,512
e la media aritmetica dei soprastanti 6 numeri

$$\frac{1}{6}(2{,}150 + 2{,}345 + 3{,}410 + 19{,}097 + 54{,}071 + 117{,}512) =$$

BOZZA - DRAFT

$$= \frac{198{,}585}{6} \approx$$

è la media interquartile cercata:

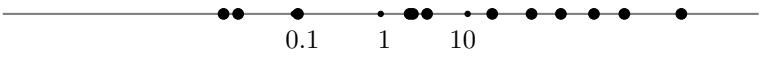
$$\boxed{\approx 33{,}0975}$$

o con minor precisione

$$\boxed{\approx 33{,}1}$$

Nota 1. Salvo l'ordine il dataset è stato ottenuto da una variabile aleatoria log-normale di parametri $\mu = 1$ e $\sigma^2 = 25$, che non sono assolutamente media e varianza della variabile aleatoria. Precisamente sono stati tratti gli esponenziali dei valori della variabile aleatoria normale usata nell'esercizio 5: gli esponenziali dei valori di una variabile aleatoria normale sono distribuiti log-normalmente. Ecco un diagramma cartesiano in scala logaritmica dei 12 valori (ma 2 sono quasi sovrapposti). Si noti l'uguaglianza, salvo i numeri scritti all'asse delle ascisse, col diagramma dell'esercizo 53.6.1, che non è in scala logaritmica.

BOZZA, DRAFT



Il pallino più a destra corrisponde al 2848, il massimo del dataset,

Altro indice importante è in https://it.wikipedia.org/wiki/Coefficiente_di_Gini

22 Variabilità, covarianza e correlazione

22.1 Indici di dispersione ovvero variabilità

[...] seconno le statistiche d'adesso
risurta che te tocca un pollo all'anno:
e, se nun entra nelle spese tue,
t'entra ne la statistica lo stesso
perché c'è un antro che ne magna due. [Trilussa]

Gli indici dispersione ovvero variabilità cercano di quantificare con 1 numero la non omogeneità di dati numerici.

BOZZA - DRAFT

Sono statistiche di sintesi $f(x_1, ..., x_n)$ da $\mathbb{R}^n$ in $\mathbb{R}$, con n a priori indeterminato, proprio come le varie medie, viste in precedenza. È ovvio che dataset diversi possono avere diversi gradi di "disomogeneità", che vogliamo quantificare, anche se hanno la stessa media e/o mediana. Per esempio i 2 dataset, qua già ordinati,

$$2, 4, 6, 8, 10, 12, 14 \qquad 5, 6, 7, 8, 9, 10, 11$$

hanno uguale media, e perfino mediana, sempre 8, ma è chiaro che i valori del secondo dataset variano meno, ovvero sono più addensati intorno alla media. Si pensi a dei redditi per esempio.

Campo di variazione, o *range*.

$$range(x_1, ..., x_n) := \max(x_1, ..., x_n) - \min(x_1, ..., x_n) .$$

Che vale 12 e rispettivamente 6 nei 2 dataset considerati.
Il range è molto sensibile agli eventuali outlier.

Il range come sopra definito è un numero, ma altri Autori chiamano range l'intervallo $[\min, \max]$.

Differenza interquartile. Ben poco sensibile agli outlier:

$$iqr(x_1, ..., x_n) := q_{3/4} - q_{1/4} .$$

Varianza, deviazione standard, coefficiente di variazione. Detta $\bar{x}$ la media aritmetica del dataset $x_1, ..., x_n$ si definiscono

$$\boxed{\textit{population variance} \quad \sigma_X^2 = Var(x_1, ..., x_n) := \frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^2} \tag{68}$$

population standard deviation

$$\boxed{\sigma_X = sd(x_1, ..., x_n) := \sqrt{Var(x_1, ..., x_n)}} \tag{69}$$

$$\textit{ovvero} \quad = \sqrt{\frac{\sum_{k=1}^{n}(x_k - \bar{x})^2}{n}}$$

coefficiente di variazione ovvero deviazione standard relativa

$$\boxed{se\ \bar{x} \neq 0: \quad \sigma_X^* = cv(x_1, ..., x_n) := \frac{\sigma_X}{\bar{x}}} \tag{70}$$

e dell'ultima si può avere la forma percentuale moltiplicando per 100%, ottenendo la *deviazione standard percentuale*.
Alcuni Autori dividono per $|\bar{x}|$ invece che $\bar{x}$.

BOZZA - DRAFT

Esempio. Sul portale web di Our World in Data, generalmente affidabile (in particolare per aggiornamenti statistici sulla pandemia del covid) e consigliato al lettore, leggiamo[105]

> As an aggregate of the regions with available data – Europe, North America, Australia, and East Asia – they found the mean male height to be 178.4 centimeters (cm) in the most recent cohort (born between 1980 and 1994). The standard deviation was 7.59 cm. (...) Women were smaller on average, with a mean height of 164.7 cm, and standard deviation of 7.07 cm.

e la maggior variabilità dell'altezza maschile (7.59 vs 7.07) si rivela solo apparente perchè dividendo per le altezze medie si ottiene, salvo decimali, lo stesso coefficiente di variazione del 4.3%.
(E lo scrivente si permette di ritenere che la piccola differenza sarebbe stata ancora più piccola se tutti i calcoli fossero stati fatti usando i logaritmi delle altezze – ovviamente privati delle unità di misura – arrivando a qualcosa di forse meno intelligibile ma più intrinseco della cosa).

Si trova scritto sia σ che sd che SD; e sia σ_X^* che cv che CV.

[105] https://ourworldindata.org/human-height. Letto il 15 dicembre 2022.

Questi indici sono tutti sensibili agli outlier.

La varianza ha un'unità di misura diversa da quella del dataset.

La deviazione standard, di gran lunga preferita nelle Scienze Applicate alla Varianza che invece si preferisce in Matematica, ha la stessa unità di misura del dataset, per esempio litri invece che... litri quadrati.

Si noti il fattore $\frac{1}{n}$ nella varianza, e allora poi negli altri indici da essa dipendenti. Certi testi e software danno $\frac{1}{n-1}$. La questione sarà approfondita in Statistica Inferenziale, ma anticipiamo la questione, peraltro numericamente poco influente per n grande:

> If the values instead were a random sample drawn from some large parent population (for example, they were 8 marks randomly and independently chosen from a class of 2 million), then one often divides by 7 (which is $n-1$) instead of 8 (which is n) in the denominator of the last formula. In that case the result of the original formula would be called the sample standard deviation. Dividing by $n-1$ rather than by n gives an unbiased estimate of the variance of the larger parent population. This is known as Bessel's correction.[106]

BOZZA - DRAFT

Nota. Grande varianza ovvero deviazione standard: valori dispersi.

Piccola varianza ovvero deviazione standard: valori addensati presso la media.

La deviazione standard può essere considerata piccola oppure grande relazionandola alla media.

Per la varianza, esistono varianze più piccole (valori più addensati) o più grandi di altre (valori più dispersi) ma non ha molto senso considerare grande o piccola una singola varianza. Similmente la deviazione standard.

Esempio 1.

[106] da Wikipedia (in inglese), l'enciclopedia libera, alla voce Standard deviation, in https://en.wikipedia.org/wiki/Standard_deviation, il 24 ottobre 2019

> the average height for adult men in the United States is about 70 inches (177.8 cm), with a standard deviation of around 3 inches (7.62 cm).[107]

Esempio 2. In un articolo scientifico[108] hanno confrontato, mediante le deviazioni standard, la variabilità dell'ora dell'andare a dormire con l'ora dell'alzarsi in popolazioni di tipo primitivo, trovando maggior variabilità nel primo orario:

> The SD of sleep onset times exceeded the SD of sleep offset times

Il confronto fra due deviazioni standard di grandezze omogenee è ben sensato.

• Per i dataset visti in precedenza si calcolino i 5 indici sopradetti.

22.2 Covarianza; correlazione; retta di regressione

Dati 2 dataset numerici di uguale numerosità

$$X : x_1, ..., x_n \qquad Y : y_1, ..., y_n$$

si definisce la loro *covarianza* (ovvero *covarianza di 2 n-uple di numeri*, ovvero *covarianza osservata*)

$$Cov(x_1, ..., x_n; y_1, ..., y_n) := \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + ... + (x_n - \bar{x})(y_n - \bar{y})}{n}$$

dove $\bar{x}$ e $\bar{y}$ sono le medie dei 2 dataset. Si trova denotata anche anche $\sigma_{X,Y}$, oppure con *cov*(... minuscolo. Si definisce anche l'*indice di correlazione di [Bravais-]Pearson*, o *indice di correlazione lineare* (e al posto della parola *indice* altri scrivono *coefficiente*)

$$\rho_{X,Y} := \frac{Cov(x_1, ..., x_n; y_1, ..., y_n)}{\sqrt{Var(x_1, ..., x_n) Var(y_1, ..., y_n)}} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}$$

BOZZA - DRAFT

[107] da Wikipedia (in inglese), l'enciclopedia libera, alla voce Standard deviation, in https://en.wikipedia.org/wiki/Standard_deviation, il 24 ottobre 2019

[108] "Natural sleep and its seasonal variations in three pre-industrial societies", by Yetish G, Kaplan H, Gurven M, Wood B, Pontzer H, Manger PR, Wilson C, McGregor R, Siegel JM8. Curr Biol. 2015 Nov 2;25(21):2862-2868. doi:https://it.overleaf.com/project/6179b2b3c248087503c542f9 10.1016/j.cub.2015.09.046. Epub 2015 Oct 17.

e questo dà un'idea di come varino le y_j relativamente alle x_i: un indice prossimo a 1 suggerisce che al crescere delle x_i crescano linearmente le y_i, un indice prossimo a -1 suggerisce che al crescere delle x_i decrescano linearmente le y_i, e un indice prossimo a 0 indica la non presenza di tali relazioni approssimate. In ogni caso

$$-1 \le \rho_{X,Y} \le 1 \,.$$

Coi dataset X : 2, 7, 4, 3, e Y : 3, -1, 0, 2 si calcoli l'indice di correlazione, e si rappresentino sul piano cartesiano i punti (x_i, y_i). In un senso non banale, una retta $y = mx + q$ si avvicina meglio ai punti se ha coefficienti

$$m := \frac{\sigma_{X,Y}}{\sigma_X^2} \qquad q := \bar{y} - m\bar{x} \quad \text{equivalentemente: } m := \frac{\sigma_Y}{\sigma_X}\rho_{X,Y}$$

e in questo caso si chiama *retta di regressione* (dietro, vi è il *metodo dei minimi quadrati*, che non approfondiamo).
Per esempio per i punti $(2,1)$, $(1,3)$, $(3,-1)$, ovvero per i dataset $X: 2,1,3$ e $Y: 1,3,-1$, si trova $y = -2x + 5$, che passa *esattamente* per i punti. è $\bar{x} = 2$, $\bar{y} = 1$, $\sigma_X^2 = \frac{2}{3}$, $\sigma_{X,Y} = -\frac{4}{3}$. E con $y_2 := 3.14159265$ si trova $y \approx -2.07x + 5.19$. Si facciano i disegni.

BOZZA DRAFT

Il coefficiente di correlazione e la retta di regressione sono fra i più usati strumenti matematici delle Scienze Biomediche. Si segua questo LINK->

È la retta che in assoluto meglio riassume uno scatterplot?
No, tant'è vero che facendo la retta di regressione $y = m\,x+q$ si trova in generale una retta diversa dalla retta di regressione $x = m'\,y+q'$, eppure lo scatteroplot è lo stesso. Bisogna fissare la variabile che riteniamo indipendente, quella in questo paragrafo denotata con x. La questione è sottile e non la approfondiamo.

22.3 L'inimicizia fra progresso e variabilità

Nominate la variabilità a un Ingegnere, in pratica colui che progetta quasi tutte le cose materiali che vedete intorno a voi, e gli

si drizzeranno i capelli in testa: per l'Ingegneria la riduzione della varianza nei processi produttivi è obiettivo prioritario.

Se guardate i flaconi di vetro in un museo di antichità romane, vedete che sono alquanto variabili nella forma: sono fatti a mano uno per uno, di vetro soffiato. Se una farmacia ne comprava una certa quantità per mettervi un certo farmaco, erano tutti diversi.
Se guardate in una farmacia odierna i flaconi di un qualunque prodotto, sono tutti uguali, o, per meglio dire, la variabilità – che a livello molecolare comunque esisterà – è ridottissima.

Questo sopradetto è un fatto generalissimo del progresso.

Tutti gli oggetti di uso tecnico e/o scientifico, dai flaconi alle automobili, sono prodotti con una variabilità piccolissima, e la sostanziale inesistenza di variabilità – comunque esistente a un livello microscopico – è uno dei principali indicatori della bontà di un processo produttivo. Ancora all'inizio del '900 auto e aerei erano costruiti uno per una.
Potete accettare che i le clementine e i limoni, prodotti nei campi, abbiano una certa variabilità nelle dimensioni, ma non certo

- le bottiglie d'acqua minerale
- le siringhe
- i flaconi di un certo farmaco
- le compresse o capsule di un certo farmaco
- i chip dei computer, eccetera eccetera.

E comunque anche per i limoni, e gli altri prodotti agricoli, col tempo i vari standard legislativi sono sempre più restrittivi riguardo la variabilità: i limoni di *size code* 7, i più piccoli, devono avere diametro fra 45 e 52 mm, e comunque anche senza quel codice non si possono vendere come limoni dei limoni di diametro minore di 45 mm. (Non stiamo qua negando che siano limoni ma non potete venderli come limoni; ve li accetteranno forse come materiale per compost). LINK ->

È in atto un secolare processo che tende a ridurre sempre più la variabilità degli oggetti prodotti in una determinata serie.

Icasticamente:

BOZZA - DRAFT

il progresso odia la variabilità

L'uniformità ovvero bassissima variabilità, misurata su qualche parametro (per esempio il diametro dei limoni) consente

– risparmio economico
– risparmio di tempo

In campo farmaceutico, il danno della variabilità è evidente nella necessità di avere dosaggi garantiti uniformi in ogni flacone, in ogni capsula, in ogni compressa, eccetera. È ovvio.
Fra 2 compresse di aspirina di una stessa marca c'è da aspettarsi una variabilità infinitesimale.

DRAFT

Figure 28: Anche il farmaco più miracoloso che l'Uomo avesse mai prodotto potrebbe talvolta fare male se – per ipotesi assurda – il prodotto che arriva all'utenza non è esattamente uguale a quello previsto.

23 Note finali sulla Statistica Descrittiva

Ad un livello basico, nella Statistica Descrittiva **si vuole riassumere molti valori (*dataset*) con 1 valore o pochi valori, e/o con un diagramma**.

Quest'opera di sintesi non è banale.

Viene fatta per comprendere i dati, divulgarli, confrontarli. Alcuni lo fanno perfino per presentarli manipolativamente; allora noi dovremo attrezzarci conoscitivamente per non essere tratti in inganno.

Questo è centrale in Farmacia. Basta ridefinire 1 soglia, e qualche miliardata di persone da sane potrebbero trovarsi riclassificate come malate, verosimilmente bisognose di farmaci.
Leggiamo su un articolo[109] del British Medical Journal:

> There's a lot of money to be made from telling healthy people they're sick. Some forms of medicalising ordinary life may now be better described as disease mongering: widening the boundaries of treatable illness in order to expand markets for those who sell and deliver treatments. Pharmaceutical companies are actively involved in sponsoring the definition of diseases and promoting them to both prescribers and consumers. The social construction of illness is being replaced by the corporate construction of disease.

BOZZA - DRAFT

23.1 Bugie, bugie cattive, e statistica

Da più di un secolo si dice "lies, damned lies, and statistics". Ai numeri si può far dire quel che si vuole o quasi.

[109] Moynihan R, Heath I, Henry D. Selling sickness: the pharmaceutical industry and disease mongering. BMJ. 2002 Apr 13;324(7342):886-91. doi: 10.1136/bmj.324.7342.886. PMID: 11950740; PMCID: PMC1122833.

Attenzione!

PARTE I – DATI VERI

23.2 La scelta dei parametri. Cosa misurare?

In questa trattazione elementare siamo sempre partiti da dati numerici, ma a monte c'è il problema di quali numeri considerare in una data situazione.

Se consideriamo un quadrato di lato 2 ovvero area 4, e 2 quadrati di lato 5 ovvero area 25, e volessimo rappresentare un "quadrato medio", facendo la media dei lati abbiamo 4 (con area 16), invece facendo la media delle aree abbiamo 18 (con lato $3\sqrt{2} \approx 4.24$).

BOZZA - DRAFT

Qual è il senso di quelle tabelle che mettono ai primi posti il peperoncino per contenuto di vitamina C? Il peperoncino si mangia, eventualmente, a grammi, mica a etti come certa frutta, che, certo, contiene meno vitamina C *all'etto*, ma ne contiene di più *per porzione* (*ragionevole*, per quanto ciò possa significare). La scelta di un diverso parametro cambia completamente l'impostazione della questione.

"Il Molise ultimo per decessi covid"

sì ne aveva solo 31, ad una certa data, meno di ogni altra regione, ma ha anche pochi abitanti. Considerando i decessi per 100mila abitanti ne aveva 10, più della Basilicata che ne aveva 8, pur avendo un numero maggiore di decessi. Non è più ultimo.

Se leggiamo

La regione X prima per tumori

potremmo farci l'idea che la regione X sia più inquinata, o meno assistita medicalmente, magari peggio governata da quel noto partito che ha la maggioranza nella regione X – associazioni d'idee che magari sono il motivo reale per cui è stato pubblicato quell'articolo sui media – ma il contenuto informativo di quel titolo è seminullo, e quelle argomentazioni che ci sorgono istintivamente sono

praticamente insensate. Potrebbe ben essere che nella regione X i tumori siano 10 volte meno probabili *per ogni singolo residente* che nella regione Y, ma la regione X ha più del decuplo di abitanti della regione Y, per cui *il numero assoluto di casi* della regione X sopravanza quello della regione Y, molto più inquinata e meno assistita medicalmente.
Nella situazione considerata, altrettanto legittimo sarebbe stato il titolo

La regione Y prima per tumori

e in un caso numerico estremo potrebbe essere vero addirittura

Nella regione X la ***minima*** *incidenza di tumori*

(Per una disamina del termine *incidenza*: ->Link).

Anche qua la scelta di un diverso parametro cambia completamente l'impostazione della questione.

Numero assoluto dei morti: non dice nulla su inquinamento, malasanità...

Tasso di mortalità: serio indizio, su varie possibilità.

BOZZA - DRAFT

Il numero assoluto, spesso espresso da cifre ad alto impatto emotivo ("milioni di tumori"), da un punto di vista medico e farmaceutico ci dice meno della frequenza relativa, che in generale si potrà misurare in casi per 1 000 residenti, o per 100 000, o per milione. (Per evitare scritture come 0.000077, scarsamente leggibili: molto più chiaro, come nell'esempio del link sopra riportato, 7.7 casi per 100 000).
Si noti che la frequenza relativa corrisponde alla probabilià, per 1 singolo residente preso a caso. Per esempio 50 casi ogni 1000 abitanti dà una probabilità di $\frac{1}{20}$. Si tratta della concezione frequentista della probabilità.
La frequenza assoluta ha una diversa importanza: corrisponde alla *dimensione economica* complessiva della questione: è ovvio che se in un certo luogo ci sono più casi di una malattia usualmente trattata con un certo farmaco, là c'è un maggior *mercato* per quel farmaco.

In ambito medico e farmaceutico. Ragionando con la logica comune – sbagliata – trovare una riduzione nelle aritmie che causano morte è un buon risultato per un farmaco, e si verifica abbastanza facilmente; verificare la sopravvivenza a lungo termine è molto più oneroso (e in passato ha dato risultati opposti).

La sopravvivenza a 5 anni dalla diagnosi di cancro è un indice di altissimo valore ma richiede molto tempo e denaro; la riduzione della massa tumorale in un breve tempo si verifica prima.

Cosa misuriamo è un serio problema generale della Medicina e della Farmacologia.
Vogliamo massa tumorale ridotta? A 2 mesi o a 6 mesi o a 1 anno dall'inizio del trattamento, o quando? O vogliamo diminuita mortalità, ma, di nuovo, quando misurata? O vogliamo aumentata sopravvivenza complessiva? Oppure aumentata *progression-free survival*? O migliore *response rate*, tasso di risposta del tumore alla terapia valutato misurando qualche *biomarker*? O diminuita spesa? (Che anche quello conta). O migliorata qualità della vita? (Sottoproblema: misurata come?) O un po' tutto, con una media pesata con coefficienti da decidere? In ognuno di quei casi si otterranno risultati in generale diversi.
Leggiamo sul Journal of Clinical Oncology[110]

> produce results that are clinically meaningful to patients (ie, significantly improved survival, quality of life [QOL], or both)

(La riduzione della massa tumorale, così tanto usata, appare per certi versi la cosa che *meno* possa interessare il paziente).
Lo stesso discorso vale praticamente per ogni patologia.

Icasticamente:

L'operazione è riuscita ma il paziente è morto

Leggiamo in un articolo scientifico:[111]

BOZZA - DRAFT

[110] American Society of Clinical Oncology Perspective: Raising the Bar for Clinical Trials by Defining Clinically Meaningful Outcomes. By: Lee M. Ellis et al. DOI: 10.1200/JCO.2013.53.8009 Journal of Clinical Oncology 32, no. 12 (April 20, 2014) 1277-1280.

[111] Progression-Free Survival: Meaningful or Simply Measurable? By: Christopher M. Booth, Elizabeth A. Eisenhauer. DOI: 10.1200/JCO.2011.38.7571 Journal of Clinical On-

> The last few years have seen an increase in the number of randomized controlled trials (RCTs) of new agents in metastatic solid tumors using progression-free survival (PFS) as the primary end point. Some trials showing improvement in PFS, without a corresponding increase in overall survival (OS), have led to approval of new drugs and/or changes in standard of care. This suggests a growing belief in the oncology community that delaying progression in metastatic disease is a worthy goal, even if OS is not improved. But is a new treatment that improves PFS really an advance for patients? Or is it only lowering the bar to declare active some of our much-heralded new molecular targeted therapies?

23.3 Rispetto a quale standard fare le statistiche?

BOZZA - DRAFT

Una fonte di problematicità è quale standard scegliere. Se per esempio volessimo vedere se la carenza di una certa sostanza nel sangue è statisticamente correlata ad una malattia, avremmo – oltre al problema della diagnosi della malattia nei soggetti, problema medico che non considereremo – il problema di quali soggetti considerare carenti e quali no, in base alle analisi chimiche del sangue: infatti ci sono molti standard per i vari livelli delle sostanze nel sangue, e in generale per i parametri biomedici, come l'essere sovrappeso, e come la mortalità infantile e l'aspettativa di vita alla nascita, standard che variano nel tempo e nelle grandi entità scientifiche, sanitarie e/o geopolitiche ragionevolmente considerabili: Ministero della Salute italiano, Comunità Europea, OMS (Organizzazione Mondiale della Sanità), ONU, USA (NIH, CDC...), la rivista scientifica The Lancet e molte altre...

L'aspettativa di vita delle donne in Italia, e il ranking dell'Italia per essa, sarebbe[112]

86.49 – 3^ secondo l'ONU (luglio 2015), dati 2010-2015
85.6 secondo l'OECD (2016), che considera solo stati OECD
85.1 – 12^-13^ secondo The World Factbook (2017)

cology 30, no. 10 (April 01, 2012) 1030-1033. https://ascopubs.org/doi/full/10.1200/JCO.2011.38.7571

[112] Si veda Wikipedia, l'enciclopedia libera, alla voce List of countries by life expectancy.

84.8 – 7^ secondo l'OMS (maggio 2016), dati 2015
83.90 – 7^ secondo (un articolo di) The Lancet (2012).
Abbastanza variabilità per fare affermazioni molto diverse, pure autorevoli, specialmente usando il dato del ranking, molto fuorviante: per esempio San Marino viene considerato solo nella lista del World Factbook.

Gli standard variano nel tempo, e a seconda delle agenzie che li fissano, pur autorevoli.

Esempio. Attualmente (giugno 2019) leggiamo su sito governativo statunitense a proposito dell'anilina, in https://www.atsdr.cdc.gov/substances/toxsubstance.asp?toxid=79

> Cancer Classification: EPA: Probable human carcinogen. IARC: Not classifiable as to carcinogenicity to humans. NTP: Not evaluated

(Si veda per esempio il Link ->).

Altro esempio: leggiamo[113]

> several factors influence the likelihood of progression from prediabetes to Type 2 diabetes, including how we define prediabetes (there is no internationally agreed-upon standard)

BOZZA - DRAFT

Varia nel tempo e nei luoghi cosa sia un farmaco...

...o piuttosto una droga (come l'eroina, "venduta liberamente dalla multinazionale farmaceutica Bayer dal 1899"[114] per varie patologie), o un cancerogeno[115] (come la formaldeide: "Fino al momento

[113] Online text in https://www.levelshealth.com/blog/how-to-reverse-prediabetes read December 15, 2022.

[114] https://it.wikipedia.org/wiki/Eroina:

> L'intento era quello di ottenere una molecola più efficace della codeina nel sedare la tosse, la tubercolosi e le patologie respiratorie. (...) In breve tempo l'impiego terapeutico si ampliò alle più disparate patologie pneumologiche, ma anche neurologiche, ginecologiche, o a semplici dolori; si diffusero pertanto svariate preparazioni farmaceutiche acquistabili liberamente, e questo fece sì che l'eroina divenisse velocemente uno dei farmaci più venduti in assoluto.

[115] https://it.wikipedia.org/wiki/Formaldeide:

> L'Associazione Internazionale per la Ricerca sul Cancro (AIRC) sin dal 2004 ha inserito la formaldeide nell'elenco delle sostanze considerate con certezza

in cui il prodotto è stato bandito, era uno degli ingredienti delle pastiglie Formitrol per il mal di gola"[116]).

Le conoscenze scientifiche variano nel tempo, e con esse i vari standard, seppure in parte contraddittori, e in uno stesso tempo e luogo esistono opinioni mediche opposte.

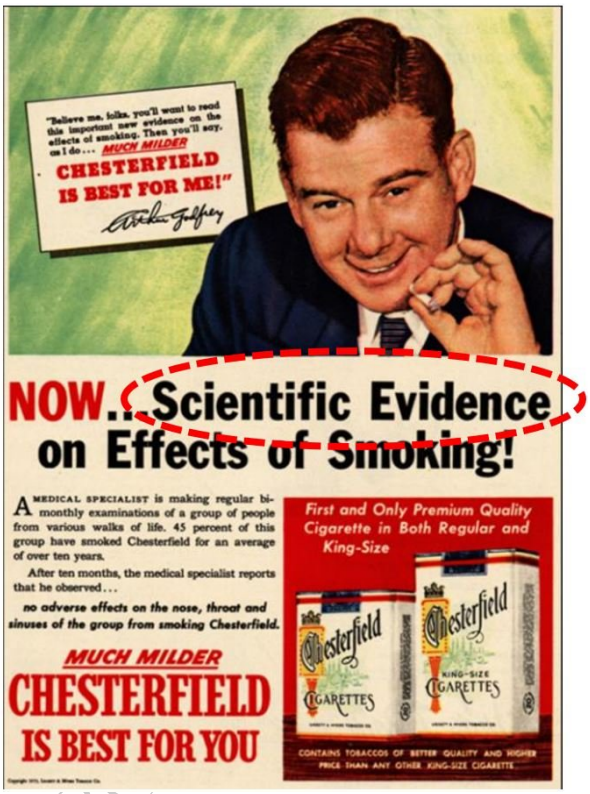
cancerogene per la specie umana. (...) Dal 1° gennaio 2016, la formaldeide è passata dalla classificazione di "sospettato di provocare il cancro" a "può provocare il cancro" (Regolamento (UE) N. 605/2014).

[116] https://it.wikipedia.org/wiki/Formaldeide

Un orbitoclasto è uno strumento chirurgico utilizzato per eseguire lobotomie transorbitali. Esso fu inventato dal dottor Walter Freeman nel 1948 come sostituto all'unica forma di leucotomo allora utilizzato per gli interventi di tal genere. Lo strumento in questione è a forma di punteruolo rompighiaccio, con alcuni segni incisi sul manico ed una struttura a cuneo sull'estremità esterna. L'operazione consiste nell'inserire lo strumento dietro la palpebra del paziente, perforare il sottile strato osseo che è posto inferiormente ad essa, e attraverso colpi di martello far penetrare lo strumento fino al lobo frontale. L'orbitoclasto viene poi mosso in varie direzioni al fine di distaccare i lobi frontali dal talamo. Nel 1948, Freeman perfezionò la procedura introducendo un profondo taglio frontale ed un movimento ulteriore della punta nel lobo, al fine di evitare - come succedeva occasionalmente in precedenza - che l'orbitoclasto si rompesse nella testa del paziente, necessitando così una rimozione chirurgica. In più, Freeman riuscì a commissionare la produzione di orbitoclasti di materiale più resistente ed affidabile.

Poiché durante i primi interventi venivano usati veri e propri punteruoli da ghiaccio, la procedura fu soprannominata "lobotomia con punteruolo da ghiaccio" ("ice pick lobotomy"). Nel 1949 fu assegnato il Premio Nobel per la Medicina a Egas Moniz,

"per la sua scoperta del valore terapeutico della lobotomia in alcune psicosi". (Da Wikipedia, l'enciclopedia libera.



23.4 Col ranking facciamo ben figurare il Paraguay

Supponiamo che abbiamo questi dati, relativi a una cosa qualunque che i lettori di un giornale percepiscono in generale come *buona*:

USA 3723
Canada 3518
Paraguay 18
India 16
eccetera

Oggi vogliamo dire bene del Paraguay, a beneficio di qualche potentato, ovvero interesse geopolitico – il lettore acuto può trovarne uno – a cui il nostro giornale, per il quale scriviamo, risponde. Che fare, con quel misero valore 18? Potremmo dire che

il Paraguay è terzo dopo USA e Canada, davanti all'India.

23.5 Illusioni percettive nella presentazione dei dati

In questo paragrafo usiamo sia lo standard del punto decimale che della virgola decime.

La dose giornaliera raccomandata di vitamina C è stata portata da 60 milligrammi a 90 milligrammi:

possiamo dire che adesso è del 50% maggiore

possiamo che prima era del 33% minore.

Ovviamente col primo modo diamo l'impressione di una variazione maggiore, almeno al lettore poco esperto di numeri.

L'epidemia di influenza del 2016/17 ha fatto circa 25 000 morti, circa il 28% in meno della prima ondata del covid-19 nel 2020 che ha fatto circa 35 000 morti, il 40% in più...

In linea generale, i titoli faranno più conseguenze dei testi sottostanti.

BOZZA - DRAFT

Se il partito dei Vispi Volpini un anno riceve lo 0,6% e l'anno dopo lo 0,9% (è un partito piccolo...) la televisione dirà che è cresciuto dello 0,3%, ma il Corriere dei Vispi Volpini dirà che il loro elettorato è aumentato del 50%, il che è vero.
(Magari anche un telegiornale lo dirà, chissà chi è il direttore...)

N.d.A. Ricordo, decenni fa, un esempio di presentazione fuorviante dei risultati elettorali del tipo "il partito A scende dal ...% al ...%, il partito B prende il ...% contro il ...% della volta precedente" dove nel profluvio di cifre e decimali l'ascoltatore medio ricorda solo che A *scende* e B *prende*, ma in effetti B scendeva più di A...

Fare raffronti con dati collaterali può permettere di titolare "aumento" un articolo che poi parla di un fenomeno in diminuzione, ma moltissimi si limiteranno a leggere il titolo.

In ambito medico, affermare che con diagnosi precoci si allunga la speranza di vita dalla diagnosi può essere molto fuorviante: “Una volta il paziente viveva mediamente due anni dalla diagnosi, adesso dodici anni"... ma facendo la diagnosi mediamente quanti anni prima?

Attenzione alle fallacie della memoria: vi pare di ricordare che secondo l’Organizzazione Mondiale della Sanità ogni anno 12.6 milioni di persone muoiono a causa dell’inquinamento, e adesso leggete che ogni anno 4.2 milioni di persone lasciano questo mondo a causa dell’inquinamento dell’aria? Basta rileggere con attenzione.

Attenzione soprattutto alle rappresentazioni grafiche.
Iniziare un diagramma a colonne o un grafico da un certo tempo invece che da un tempo precedente può permettere di fare affermazioni completamente diverse, magari entrambe formalmente vere, ma che genereranno emozioni e poi reazioni completamente diverse.
Grafici che sull’asse delle ordinate non iniziano da 0 possono dare l’impressione di variazioni enormi, anche raddoppi o dimezzamenti, a variazioni minuscole.
Diagrammi a torte tridimensionali in rappresentazione prospettica possono far apparire più grandi certi dati rispetto agli altri.

BOZZA DRAFT

23.6 Di cosa parliamo?

Diamo solo un cenno di un argomento ciclopico. Ogni persona vive immersa in un flusso di informazioni che gli arrivano, in base al quale costruisce la sua percezione della realtà, che lui crede vera.
Anche se tutte quelle informazioni fossero vere (per ipotesi assurda, ovvio) l’immagine complessiva che l’individuo si forma è in generale (molto!) distorta: una *diversa scelta* di informazioni *altrettanto vere*, da far arrivare coi media – dal livello più basso fino agli articoli scientifici – creerebbe in quella stessa persona un’immagine del

mondo completamente diversa, per molti versi quasi opposta.

Supponiamo qua in via ipotetica che sul pianeta Kepler-22 b, simile alla Terra, il danno, in qualche modo misurato, delle sigarette si disponga a campana gaussiana, con un'ampia zona mediana di un danno medio per un grande generalità di fumatori. E che il danno delle sigarette elettroniche si disponga analogamente, ma più a sinistra sull'asse delle ascisse, cioè con un danno mediamente minore. Per quanto il danno sia mediamente minore, in moltissime persone è grave, per molte addirittura gravissimo, per alcune perfino maggiore di quello medio delle sigarette. Ecco, dai, invece di fare alcuni milioni di articoli su quelli mediamente meno danneggiati (impossibile a farsi), e invece di dire che in media il danno è minore (chi mai ci obbliga?) facciamo un bell'articolo titolando "ragazzo fuma sigarette elettroniche e ha i polmoni devastati".
La gente dirà: ecco, vedi, non serve, soldi buttati, meglio continuare con le sigarette normali!
È impressionante quante persone ragionino per casi singoli, appresi dai media, senza garanzia che siano rappresentativi di situazioni medie invece che eccezionali.

BOZZA - DRAFT

23.7 Cherry picking – bias di selezione

Affine alla questione precedente. Scegliere le prove favorevoli a una tesi e ignorare quelle contrarie. Se un proprietario di media volesse sostenere la tesi che bisogna chiudere le scuole per evitare il propagarsi del covid, potrebbe cercare di citare affermazioni favorevoli alla sua tesi, al livello autoritativo maggiore possibile, trascurando le prove contrarie; e similmente se vuole far passare la tesi opposta. Troverà articoli scientifici per qualunque dei 2 desiderata. L'articolo di review sistematica in https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8375447/ considera decine articoli scientifici di alto livello (dopo averne scartati migliaia di altri) e verifica che metà trovano che la chiusura delle scuole ha ridotto la diffusione

del contagio, e metà hanno trovato che non l'ha ridotto.

23.8 Approfittare delle fluttuazioni statistiche

BOZZA DRAFT

Mentre le funzioni elementari di base della Matematica hanno andamenti molto regolari, come x^2 o x^3, e similmente le successioni propriamente dette come quella di Fibonacci e quella dei fattoriali, le successioni di valori numerici rappresentativi di fenomeni reali normalmente presentano nel tempo fluttuazioni, alternando crescite a diminuzioni.
Esempio, morti in Italia:

2013 600 000
2014 599 000 −1 000 rispetto all'anno precedente
2015 653 000 +53 000
2016 616 000 −37 000
2017 650 000 +34 000
2018 632 000 −18 000
2019 645 000 +13 000

Consideriamo come ulteriore esempio il grafico *Sperimentazioni autorizzate dall'Autorità competente* a pagina 10 del documento La

Sperimentazione Clinica dei Medicinali in Italia – 17° Rapporto Nazionale – Anno 2018 dell'AIFA, Agenzia Italiana del Farmaco. L'ultimo dato è del 2017 e relativamente ad esso potremmo dire *veridicamente*

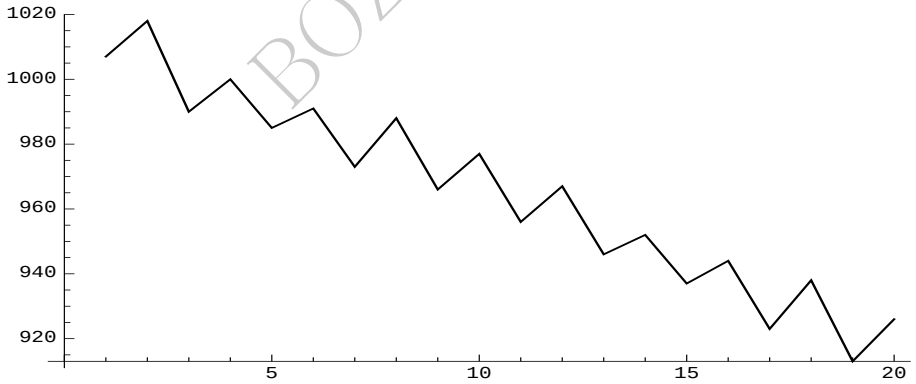
aumentate rispetto al 2000

aumentate rispetto a 15 anni prima

diminuite rispetto all'anno precedente

al livello minimo da 10 anni

ottenendo nel lettore emozioni diverse, che potranno supportare interventi decisionali diversi.
Si noti che anche in situazioni di palese diminuzione sostanziale sul lungo periodo, tali fluttuazioni possono permettere in vari anni di parlare di aumento. E negli anni in cui proprio non si può dirlo... si può parlare della situazione dell'anno precedente, che magari vedeva un aumento rispetto all'anno precedente – mica il lettore sa qual è l'ultimo dato rilasciato, in generale. Agendo così, si potrà *ogni anno* parlare di *aumento* in questo esempio di evidente *sostanziale diminuzione*:



BOZZA – DRAFT

23.9 Capziosità verbali

Un'intera enciclopedia si potrebbe scrivere, di infiniti volumi, sulle capziosità verbali, e, tranne che per la Medicina e la Farmacia, in cui c'è da piangere, sarebbe molto divertente.

Leggiamo per esempio in un articolo[117] online:

(Titolo)

> Perché le trivelle non c'entrano niente con il terremoto nelle Marche

(Sottotitolo)

> Le insinuazioni dei contestatori dell'estrazione di gas sul territorio italiano e la spiegazione dell'esperto

(Corpo dell'articolo)

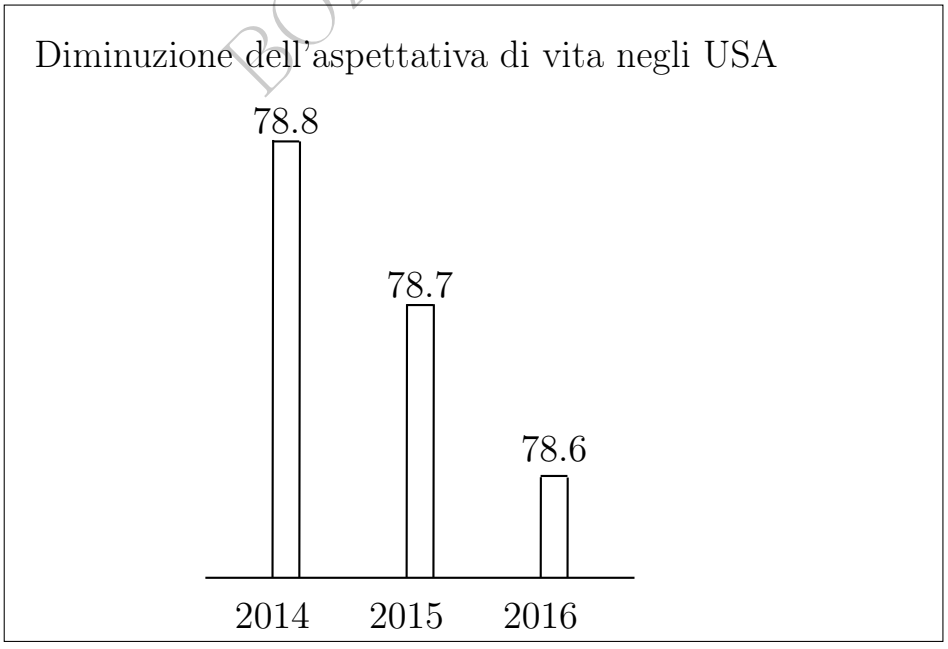
> (...) ha chiarito che "la trivellazione di per sé non comporta alcuno scuotimento sismico, sono semmai le attività di estrazione e stoccaggio che possono farlo. (...)

Insomma i contestatori delle trivelle sono fuori strada: che si trivelli pure, come *divertissement*, casomai si contestino le successive "attività di estrazione e stoccaggio".

BOZZA - DRAFT

23.10 Illustrazione di un triplice esempio reale

Aumento dell'aspettativa di vita negli USA: link<-



[117] https://www.today.it/cronaca/trivelle-terremoto-oggi-9-novembre-2022-marche.html letto il 24 novembre 2022

Aumento del tasso di mortalità negli USA:

```
2009 ..... 7.9
2010 ..... 8.0
2011 ..... 8.1
2012 ..... 8.1
2013 ..... 8.2
2014 ..... 8.2
2015 ..... 8.4
2016 ..... 8.5
2017 ..... 8.6
```

Capire questi diagrammi richiede:
1) conoscere la Statistica Descrittiva, notando in particolare che
il diagramma considera solo gli ultimi 3 anni;
il diagramma ha una linea base posta a circa 78.5 anni.
1 bis) conoscere la Demografia, che possiamo considerare semplificatamente una branca specialistica della Statistica Descrittiva. Questo ci farebbe capire che non è per nulla strano che si alzino da tempo i tassi di mortalità, anche prima della riduzione dell'aspettativa di vita. Il tasso di mortalità può ben alzarsi mentre l'aspettativa di vita si alza, proprio perché la popolazione mediamente invecchia, in compresenza di un fattore demografico completamente diverso: la scarsità di nuovi nati.
2) Conoscere la situazione reale il che è completamente diverso dal guardare questi numeri o grafici.
Per una parziale spiegazione della situazione reale dietro alla diminuzione dell'aspettativa di vita negli USA si segua questo `link<-`.

BOZZA - DRAFT

23.11 Conclusioni sulla fallace presentazione di dati veri

Pensare di aver capito una situazione perché si è vista una qualche statistica o lista di numeri è in generale *ingenuo*, e i grafici sono, per le persone non istruite, perfino più fuorvianti. Le parole, poi, possono essere ancora più ingannevoli anche se sembrano esprimere dati oggettivi, numerici: aumento, diminuzione...

È necessario acquisire una buona competenza statistica per non essere tratti in inganno. Sia da errori fatti in buona fede, che da manipolazioni volontarie, fatte da soggetti interessati.

Superato il problema della comprensione formale di statistiche e grafici, il problema

delle statistiche

e dei grafici,

e perfino dei dati grezzi/completi,

è che ce ne possono sempre essere altri, che non stiamo vedendo ovvero non ci stanno mostrando, che cambierebbero completamente la nostra prospettiva sui fenomeni considerati.

In linea generale, le statistiche economiche dei media non valgono la carta su cui sono scritte. Per le statistiche scientifiche, non è vietato sperare una maggiore oggettività.

ROZZA - DRAFT

Molti testi divulgativi affermano che *la maggior parte* dei contagi da toxoplasmosi non avvengono dai gatti ma dalla sporcizia. Anche se questa statistica – praticamente impossibile da verificare – fosse vera, quegli Autori magari trascurano di dire, nei loro articoli divulgativi, che sebbene il toxoplasma si trovi in tanti animaletti, si riproduce solo nei felini, per cui se tutti i gatti venissero inceneriti (orrore!) la toxoplasmosi semplicemente sparirebbe dall'Italia, dove praticamente l'unico felino che frequenta l'uomo è il gatto.

I gattofili non costituiscono una lobby molto pericolosa e allora qua possiamo prendercela scherzosamente con loro senza rischi, per spiegare cosa vuol dire comunicare una statistica fuorviante perché omette fatti che cambierebbero radicalmente la prospettiva.
Ma la triste verità per uno statistico è che le notizie di statistica medica che arrivano al grande pubblico dalla televisione e dai media *lasciano a desiderare*, diciamo così.

Ulteriori problematiche si presentano nella Statistica Inferenziale.

If you torture the data long enough, it will confess to anything. (In

questo link<- è in articolo su sito governativo statunitense).

Esempio finale. Da La Peste, aprile 2022, su Facebook:

> Il primo gennaio 2015 sono decedute 126 persone di 95-99 anni, mentre il primo gennaio 2021 ne sono decedute 198; 72 in più? Per fare il confronto corretto dobbiamo considerare come è variata la numerosità della popolazione di persone di 95-99 anni. Queste erano 88.799 al primo gennaio 2015, ma sono diventate 158.581 al primo gennaio 2021. Con una variazione così imponente non si può fare un confronto tra le cifre assolute. Il problema si può risolvere considerando la mortalità. Il primo gennaio 2015 sono morte lo 0,142% delle persone di 90-95 anni mentre il primo gennaio 2021 ne sono morte lo 0,125%. La mortalità è un po' inferiore.

BOZZA - DRAFT

A seconda del risultato che si vuole ottenere, si potrà dire che ne sono morti di più, o di meno, dicendo sostanzialmente il vero.

PARTE II – DATI QUASI VERI (QUASI FALSI)

Già abbiamo visto in 19.9 le problematiche del trasformare dati ordinali in dati numerici. Nella pratica, se abbiamo già il dataset ordinale ma siamo ancora liberi nel fissare le corrispondenze numeriche, possiamo largamente manipolare il risultato.

23.12 Il valore anomalo: outlier

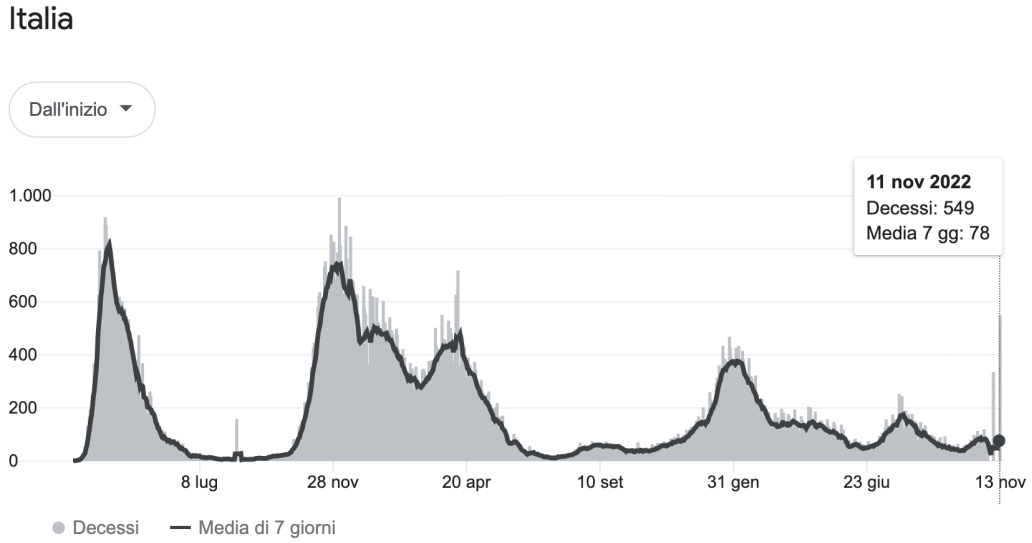


Figure 29: In questa peraltro autorevole statistica – screenshot da Google – l'esorbitante numero di 549 morti dell'11 novembre 2022 è dovuto al fatto che il Ministero ha cambiato modo di comunicare i dati: quei morti non sono di un solo giormo bensì 7.

Per un'infinità di motivi, in un dataset ci può essere qualche valore che il ricercatore ritiene *assurdo*, e tenderà ad eliminarlo prima di procedere con qualunque analisi dei dati stessi, per non comprometterla.

Questi *outlier*, valori anomali ovvero aberranti, provengono da risposte scherzose a questionari, truffe ("reddito zero"...), malfunzionamento di apparati, confusione fra punto decimale e virgola decimale, confusione fra milligrammi e microgrammi, eccetera. Ma talvolta sono dati *veri* benché *eccezionali*.

La decisione di quali valori siano da considerare outlier ed eventualmente da eliminare, può essere

soggettiva

oppure dettata da formule, che comunque non considereremo,

e in ogni caso è problematica, anche perché ce ne sono sono varie. In una ricerca seria *et* onesta, si *dovrebbe* avvertire se sono stati eliminati outlier.

Leggiamo su Wikipedia, l'enciclopedia libera, alla voce "Outlier", e non è chi non veda le pericolose conseguenze di ciò:

> There is no rigid mathematical definition of what constitutes an outlier; determining whether or not an observation is an outlier is ultimately a subjective exercise. There are various methods of outlier detection. (...) Deletion of outlier data is a controversial practice

BOZZA DRAFT

23.13 Omissione di dati ritenuti poco significativi

Dati ritenuti poco significativi vengono talvolta omessi, per esempio San Marino da molte statistiche; facendoli rientrare si possono ottenere affermazioni molto diverse, ma per certi versi effettivamente meno significative.

23.14 "Fatta la legge, trovato l'inganno"

Fissato uno standard per valutare statisticamente un qualche aspetto della realtà, qualcuno trova conveniente, diciamo così, fare azioni che migliorano di molto la valutazione statistica con scarso o nullo miglioramento della realtà. Se per esempio una legge attribuisce ai *"comuni montani"* particolari finanziamenti, fissando ad una certa altitudine del municipio la denominazione di "comune montano", state sicuri che qualche comune sposterà la sede del municipio, più a monte... *et voilà*, il comune – esattamente lo stesso di prima – diventa "montano" (e ottiene i finanziamenti).

In ambito scientifico e biomedico in particolare, a rischio, dicia- mo così, sono le statistiche che valutano la *credibilità percepita* ("impact factor") delle riviste scientifiche; si veda per esempio *A user's guide*

to inflated and manipulated impact factors, di John P. A. Ioannidis e Brett D. Thombs, European Journal of Clinical Ivestigation (17 giugno 2019) Link->

23.15 Traslazioni e confusioni linguistiche

Spesso le statistiche vengono presentate con quelle che potremmo considerare “traslazioni di linguaggio".
Per necessità di sintesi, soprattutto nei titoli – a non voler pensare male.

I cani non sono la terza causa di morte nel mondo, no. Sono, più precisamente, la terza causa di morte *fra quelle causate da animali*, dopo zanzare e serpenti.

Titolo (vero, al momento ancora reperibile in rete, ottobre 2020):

Piene al 150% le rianimazioni in Abruzzo, prima Regione in Italia ad esaurire i posti letto

ATTUALITÀ / REGIONE ABRUZZO Venerdì 16 Ottobre 2020

BOZZA - DRAFT

Più discretamente la prestigiosa agenzia di stampa ANSA, nel testo di un articolo, non nel titolo, e con qualche parola in più:

> Secondo il report settimanale dell'Alta Scuola di Economia e Management dei Sistemi Sanitari dell'Università Cattolica, campus di Roma (Altems) l'Abruzzo ha saturato il 150% dei posti letto aggiuntivi implementati

Dunque, vediamo che hanno saturato dei posti letto *aggiuntivi*, non quelli ordinari che avevano già prima della crisi.
Un'attenta analisi delle fonti che nessun normale lettore farà rivela che si tratta di 10 posti in terapia intensiva – non sappiamo quanti eventualmente per covid, di cui parlano gli articoli considerati, o incidenti stradali o altro – su 7 di certi posti definiti “aggiuntivi",

aggiunti per la crisi covid, mentre *molti più di 100 restavano non occupati*.

È stato tremendo, ho fatto 3 ore di straordinario
(Nelle 8 ore lavorative ordinarie riposavo)

Naturalmente questi titoli e queste notizie, oltre all'incalcolabile danno di ansia e paura che possono procurare alle persone fragili – in un determinato tempo e luogo nell'epidemia covid del 2020 si è constatato un *triplicamento* degli infarti, forse almeno in parte dovuti alla paura di morire – manipolano il consenso popolare verso una o l'altra azione politica, che decreterà questo o quell'altro *spostamento di risorse* o, addirittura, determinate limitazioni alla vita delle persone.

BOZZA - DRAFT

23.16 Il problema della discrezionalità

Nella rilevazione dei dati può essere presente un ampio margine di discrezionalità: classificare al microscopio cellule in "regolari/ir- regolari", "simmetriche/asimmetriche"...

Esempio lampante di discrezionalità:[118]

> "Moribund larvae were considered dead and included in the analyses."

23.17 Esempio importante: le cause di morte

Il referto sulle cause di morte presenta una certa discrezionalità.

Le persone in generale giungono alla morte con diverse patologie concomitanti (si veda *comorbidità*).
E alcune muoiono pure dissanguate in incidenti automobilistici, in realtà causati dalle loro patologie (non è il dissanguamento la "vera" causa, a monte). Si pensi per esempio all'incidente automobilistico

[118] Nartey, Rita et al. "Use of Bacillus thuringiensis var israelensis as a viable option in an Integrated Malaria Vector Control Programme in the Kumasi Metropolis, Ghana." Parasites & vectors vol. 6 116. 22 Apr. 2013, doi:10.1186/1756-3305-6-116

da crisi ipoglicemica.

Freddie Mercury è morto di polmonite (causa formale) o di AIDS (causa sostanziale)?
Anche alcol e droga forse non gli hanno fatto molto bene. Ma la causa di morte non viene ripartita in percentuali.

Il problema diventa critico per concomitanti patologie entrambe registrabili come causa di morte. Se sul piano di verità entrambe hanno favorito la morte, nelle statistiche ufficiali una sola l'ha causata, e quelle statistiche hanno conseguenze sulle politiche sanitarie e farmaceutiche – magari con la mediazione dei media.

Si pensi all'enorme difficoltà che può esserci dietro questa scelta:

morto per leucemia ma con morbillo
morto con leucemia ma per morbillo.

La decisione di cosa scrivere spetta a persone che sono generalmente molto competenti, tuttavia la situazione talvolta può essere complessa.

BOZZA - DRAFT

A proposito del Coronavirus un virologo avrebbe (2020) affermato:

> Tutte e 7 le vittime avevano anche altre patologie, quindi sarebbe più opportuno parlare di Coronavirus come concausa non come causa diretta.[119]

Molto problematico da un punto di vista statistico questo passaggio di Repubblica:

> Si tratta in gran parte inoltre - viene precisato - di pazienti contagiati dal virus nei mesi scorsi, nel frattempo negativizzatisi, ma che su indicazione del Ministero della Sanità vanno registrati comunque come soggetti con infezione da Covid.

Per capire meglio la complessità della questione, su un sito internet del Ministero della Salute leggiamo[120]

[119] Matteo Bassetti, direttore della clinica di malattie infettive dell'ospedale policlinico San Martino di Genova, citato in https://www.primocanale.it/notizie/vittime-coronavirus-l-esperto-con-l-influenza-probabilmente-sarebbero-morti-lo-stesso--216491.html, letto il 28 febbraio 2020

[120] In https://www.epicentro.iss.it/influenza/sorveglianza-mortalita-influenza, accesso il 9 marzo 2020.

> il virus influenzale non viene identificato o perché **non ricercato** o perché il decesso viene attribuito a polmoniti generiche.

(Enfasi aggiunta).
L'attribuzione di una causa di morte piuttosto che un'altra può avere grandi ripercussioni nelle decisioni di politica sanitaria, sostenuta dalle istanze dei cittadini, mediate o indirizzate dai media.

*

Non per mettere in dubbio nulla, però leggiamo, a proposito di un notissimo caso di cronaca di Trieste, di come possano cambiare i risultati da un'autopsia all'altra. Inizialmente, per la vittima, trovata in un bosco racchiusa in vari[121] sacchi, si era ipotizzato il suicidio.
Leggiamo[122]

> nuova autopsia, eseguita dopo la riesumazione del corpo, che indicano la presenza di lesioni compatibili con un'aggressione da parte di terzi:
>
> • Frattura alla lamina della seconda vertebra toracica, risalente a poco prima della morte.
>
> • Lesioni al volto e un ematoma sulla mano destra.

Questi elementi contrastano con le conclusioni della prima perizia medico-legale, che aveva escluso l'intervento di terze persone.

Leggiamo[123] altrove:

> L'avvocato (2024) Gentile: "Quadro incompatibile con un suicidio".[124]

BOZZA - DRAFT

23.18 Cose vere in articolo titolato oppostamente

In un articolo divulgativo, per esempio su un quotidiano, si possono riportare correttamente i risultati di uno studio scientifico, così nessuno può obiettare, mettendo però un titolo che dice l'opposto, così nessuno può obiettare comunque, perchè il titolo l'ha messo il titolista, mica l'autore che firma l'articolo – e senz'altro gliela si

[121] "si rafforza la tesi del suicidio (...) Dna di Liliana dappertutto. Sui sacchi neri in cui era infilato il corpo, sui sacchetti di nylon che le avvolgevano la testa" in https://necrologie.ilpiccolo.gelocal.it/news/137558

[122] https://www.triestecafe.it/it/news/cronaca/caso-resinovich-a-porta-a-porta-l-analisi-di-vladimir-luxuria-sul-mistero-della-sua-morte.html

[123] https://www.fanpage.it/attualita/liliana-resinovich-cosa-e-emerso-dalla-seconda-autopsia-frattura-a-una-vertebra-e-altre-lesioni/

[124] Precedenti (2022) osservazioni critiche del medesimo avvocato in https://www.triesteprima.it/cronaca/liliana-resinovich-associazione-penelope.html

perdona. Ovviamente il titolo farà immensamente più effetto del corpo dell'articolo.

Covid, effetto boomerang; a ogni reinfezione conseguenze più gravi

di Fabio Di Todaro

Il lavoro di un gruppo di ricercatori dell'Università di Washington pubblicato s possibili conseguenze di infezioni plurime da coronavirus: ad ogni nuovo cont incontro a complicanze

11 NOVEMBRE 2022 AGGIORNATO ALLE 09:44

In assoluto determinano quadri meno gravi rispetto a quelli relativi al primo contagio, ma le reinfezioni da Sars-CoV-2 non è detto che evolvano sempre senza particolari conseguenze. Ragion per cui continuare ad adottare le misure di protezione non

Figure 30: Fotografia di la Repubblica online, 11 novembre 2022. Si legga la prima riga dell'articolo.

BOZZA

PARTE III – DATI FALSI

23.19 La falsificazione dei dati

Finora abbiamo considerato solo dati *veri*, magari presentati con omissioni o in modo fuorviante ma comunque veri: nella realtà pratica, esiste anche la falsificazione dei dati.

Falsificazioni negli articoli divulgativi

Per dare solo una vaga idea del livello stratosferico di menzogna continua – costantemente monitorata dallo scrivente – della generalità dei media italiani (l'Italia è internazionalmente classificata nel 2022 al 58esimo posto nel mondo per libertà di stampa) durante la pandemia del covid-19, si consideri il caso della Svezia, che

ha tenuto una politica anticovid diversissima dall'Italia, e in particolare non ha mai (2020 - OTT 2022) chiuso le scuole: la Svezia veniva presentata come un luogo in cui il covid faceva strage, molto più che in Italia; il lettore interessato troverà in rete che la politica anticovid della Svezia veniva addirittura assimilata a un sacrificio umano. La verità fattuale è opposta:

– la Svezia ha costantemente avuto una mortalità covid molto più bassa dell'Italia;

– questo fatto non può essere liquidato dicendo che là non hanno contato bene i morti per covid, perchè anzi, la Svezia ha avuto un eccesso di mortalità *per tutte le cause* rispetto agli anni precedenti minore dell'Italia.

DRAFT

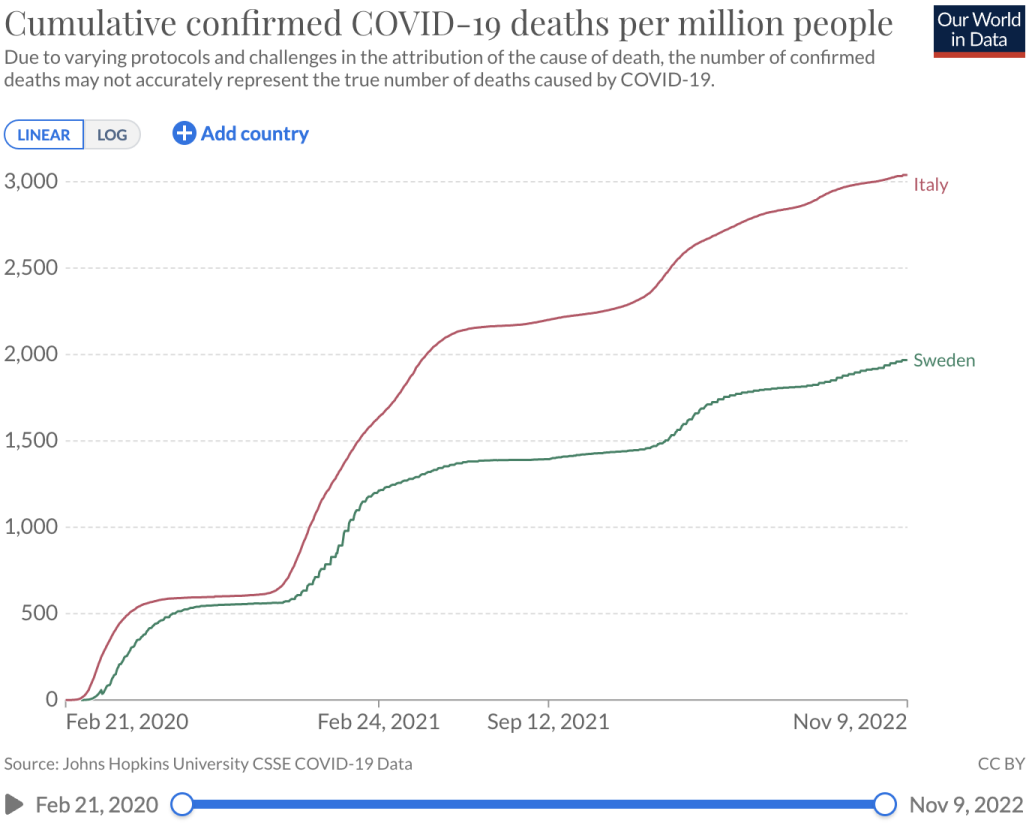


Figure 31: https://ourworldindata.org/covid-deaths?country=ITA~SWE, screenshot.

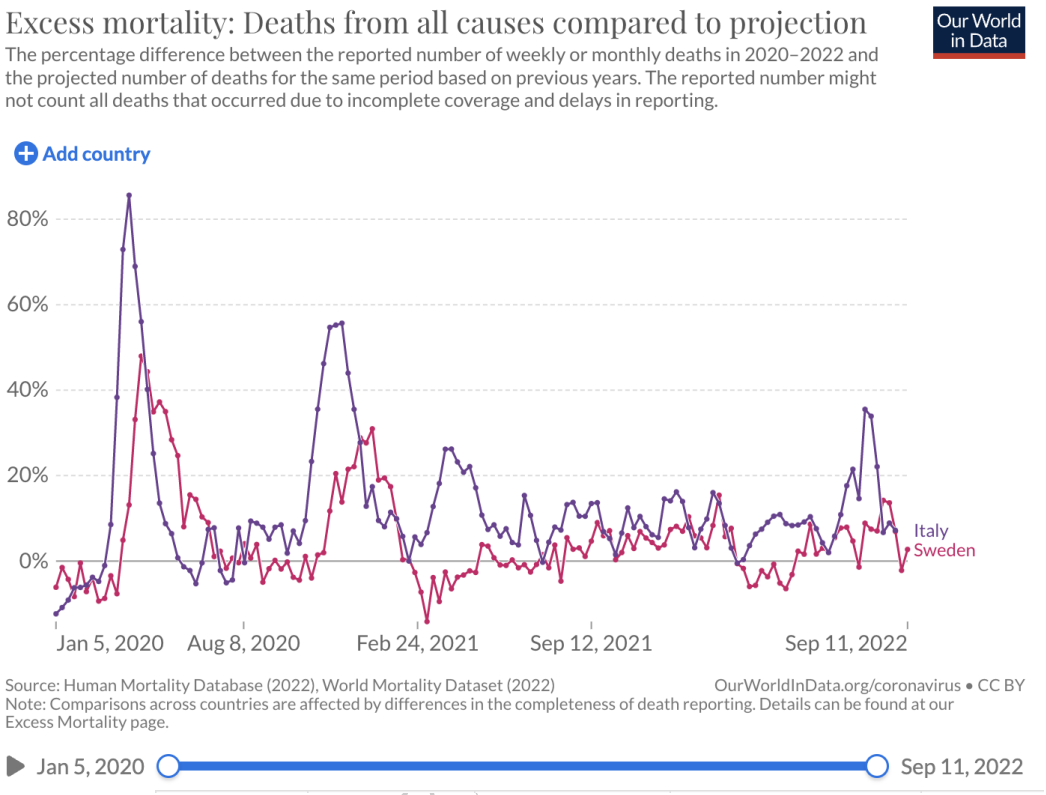


Figure 32: https://ourworldindata.org/excess-mortality-covid, screenshot.

BOZZA

E negli articoli scientifici è legale mentire?

Leggiamo[125]

> Giugno 2019 [...] Scienziati impegnati nella ricerca sul cancro hanno manipolato le immagini dei loro studi, riuscendo così a ottenere successo, carriera, nuovi fondi per le loro ricerche. La Procura (...) ha appena concluso un'indagine lunga e complessa, che fornisce un quadro devastante

e volendo troviamo[126] molti dettagli fra cui

[125] In https://www.ilfattoquotidiano.it/in-edicola/articoli/2019/06/30/ricerca-\sul-cancro-risultati-ritoccati-per-ottenere-milioni/5291204/

[126] https://codacons.it/ricerca-sul-cancro-risultati-ritoccati-per-ottenere-\milioni/

la procura deve archiviare (...) non è reato

Dal punto di vista giuridico, le cose stanno nei termini seguenti.
Le *bugie* possono essere eventualmente immorali, ma non sono illegali, tranne in alcune fattispecie elencate dalla legge: il falso in atto pubblico, il falso in bilancio, il falso ideologico, eccetera, ma assolutamente non è compreso il falso in articolo scientifico, che quindi è assimilabile a una bugia fra privati cittadini – anche se l'articolo contiene descrizioni di esperimenti mai avvenuti, cioè non semplici *errori* ma falsità scritte con dolo.
Naturalmente, un cittadino che ritenesse che dalla pubblicazione di affermazioni false in un articolo scientifico gliene sia conseguito un danno economico, o un danno biologico, potrebbe ipotizzare una querela, o una denuncia – ma si capisce bene che in generale sarà impresa ardua, specialmente per articoli scientifici che non dicono bene di farmaci specifici, ma facciano *Scienza di Base*.
Tuttavia, l'autore truffaldino rischia pur sempre qualcosa.
Se egli è, come avviene quasi sempre, dipendente di un'istituzione scientifica (per esempio un'università), o anche solo affiliato scientificamente ad essa, è ben probabile che la pubblicazione di dati deliberatamente falsificati sia contraria al codice deontologico dell'istituzione. Allora al falsario l'istituzione può comminare una sanzione disciplinare in base al proprio regolamento, per violazione del codice deontologico (comunque solo in base a regolamenti e non a leggi, e quindi ad un livello molto più basso: le pene detentive sono escluse per la violazione di regolamenti).
Inoltre se la falsificazione diviene di dominio pubblico, l'istituzione può rivalersi contro il truffaldino per avere il risarcimento per danno di immagine - con una denuncia o querela (e quindi al livello di legge).
Si noti comunque che mentre la Giustizia tende a perseguire le violazioni delle leggi (cioè i reati), l'istituzione scientifica avrebbe in generale tutto l'interesse a occultare o almeno a minimizzare la violazione del regolamento (avvenuta con la falsificazione dei dati) per evitare un danno d'immagine, e ciò fa intuire come verosimilmente andrebbero le cose.

BOZZA - DRAFT

Esercizio$_{\mu}$ Si afferma che un intervento del 1845 ha causato una rapida diminuzione del parametro X, e si presenta questo diagramma a colonne:

```
1845 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
1846 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
1847 XXXXXXXXXXXXXXXXXXXXXXXXX
1848 XXXXXXXXXXXXXXXXXXXX
1849 XXXXXXXXXXXXXXXXX
1850 XXXXXXXXXXXXXX
```

È plausibile quell'affermazione? Ecco un dataset più completo:

1842 7.68
1843 6.14
1844 4.92
1845 3.93
1846 3.19
1847 2.58
1848 2.09
1849 1.69
1850 1.37

(Risposta: no, non è plausibile. Infatti fino al 1845 il parametro diminuiva circa del 20% annuo e poi circa del 19%).

BOZZA - DRAFT

Sezione A2 – Calcolo Infinitesimale

BOZZA - DRAFT

24 Limiti di successioni e serie geometrica

24.1 Limite della concentrazione plasmatica

La successione

$$y_n = 8 \cdot 2^{-3\,n}$$

potrebbe bene rappresentare la concentrazione decrescente di un farmaco nei giorni 0, 1, 2... con qualche unità di misura. (Precisiamo: magari somministrato alle 6:00 di mattina del giorno 0 e misurato al mezzogiorno di ogni giorno).

Per la concentrazione del farmaco nel sangue sopra considerata sarà

$$\lim_{n\to+\infty} y_n = 0$$

come già possiamo intuire.

ZZA - DRAFT

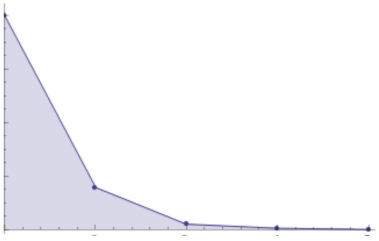


Figure 33: Screenshot da WolframAlpha. Le linea e l'ombreggiatura servono solo per migliorare la lettura del diagramma, ma solo i puntini sono significativi.

Si noti che il soggetto sarà già morto, e quindi privo di concentrazione sanguigna, molto prima non solo dell'inesistente giorno $+\infty$, ma anche del milionesiomo giorno: nondimeno, il fatto che la concentrazione di quella sostanza *tende a zero*, è molto significativo.

Ambiguità notazionale. Altri $n \to \infty$ invece di $n \to +\infty$

24.2 Una triste storia di infermità a lieto fine e una no

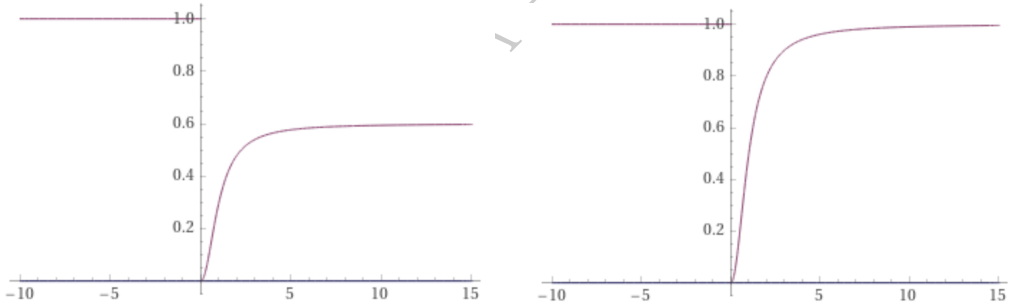
Alcibiade, soggetto A, alle ore 00:00 si sega un dito della mano destra con la sega elettrica, ed essa cade sullo stesso dito della mano destra di Berengario, soggetto B.
Detto t il tempo, per esempio in giorni, e 0 è l'istante (giorno) dell'incidente,

la funzionalità della mano destra di A, sia essa $\alpha(t)$

e

la funzionalità della mano destra di B, sia essa $\beta(t)$

che per $t < 0$ erano 1 ovvero 100%, precipitano entrambe a 0.
Ma nel tempo B tende a recuperare al 100% la funzionalità della mano infortunata, mentre ad A, che pure recupera via via molto nei giorni successivi all'incidente, tende a rimanere sul lungo periodo un'invalidità residua permanente (avvicinandoci qua alla Medicina Legale), diciamo del 40%:



$$\lim_{t\to+\infty} \alpha(t) = 0.6 \qquad \lim_{t\to+\infty} \beta(t) = 1$$

Tutto questo rimane vero considerando 1 singolo valore per giorno, alle 00:01, ottenendosi successioni a_n e b_n invece delle funzioni $\alpha(t)$ e $\beta(t)$

$$\lim_{n\to+\infty} a_n = 0.6 \qquad \lim_{n\to+\infty} b_n = 1$$

e i grafici verrebbero "a puntini" invece che con archi di curva.

24.3 Un limite banale ma illuminante

Consideriamo la funzione, in effetti successione (funzione definita in $\mathbb{N}$ o una sua semiretta destra), che potremmo indicare $f(n)$ variando n nei numeri naturali, ma proprio per questo motivo preferiamo indicare con a_n o y_n o simile, come si usa per le successioni,

$$y_n := \text{valore della costante di gravitazione universale al giorno } n$$

essendo n il numero ordinale del giorno, per esempio a partire dal 1 gennaio (calendario giuliano) dell'anno 1, oppure in qualunque altro calendario.
Per quanto attualmente se ne sa (ma ci sono varie ipotesi, di cui non ci occupiamo assolutamente), il valore della costante di gravitazione universale è, per l'appunto, costante nel tempo, in tutti i giorni della storia dell'universo, e fatte salve le unità di misura che non riportiamo, vale nel Sistema Internazionale delle unità di misura $6.67... \times 10^{-11}$. (È una delle costanti universali della Fisica meno precisamente determinata, cosa che non ci interessa).
Ora vorremmo chiederci (in quest'ipotesi della costanza) quanto vale al giorno $+\infty$. Certo, un giorno con tale numero ordinale mai esisterà, perchè $+\infty$ non è un numero; nondimento, chiunque capisce il senso reale della domanda: quanto varrà in un tempo "infinitamente futuro"? Ovviamente, in quest'ipotesi della costanza, ancora $6.67... \times 10^{-11}$, proprio come il 1 gennaio (calendario giuliano) dell'anno 1, e perfino... oggi.
Abbiamo allora, per ogni costante reale c:

$$\boxed{\lim_{n\to+\infty} c = c} \tag{71}$$

Nota. Finora abbiamo trattato i limiti di successioni e funzioni senza cautele col linguaggio comune, e dal prossimo paragrafo formalizziamo la questione dei limiti delle successioni.

BOZZA DRAFT

24.4 Successioni (di numeri reali)

Mentre passaggio al quadrato, exp, sin, cos... a numeri reali associano numeri, il fattoriale SOLO A NUMERI NATURALI associa numeri. Cioè, esistono 2.5^2 ed $e^{2.5}$ ma $2.5!$ non[127] esiste.

Questo tipo di funzioni numeriche definite solo su $\mathbb{N}$ o su una sua semiretta (per esempio $\{1, 2, 3...\}$ senza lo 0) si chiamano

SUCCESSIONI (di numeri reali).

Invece che indicarle con simboli come $f(x)$ oppure $g(n)$ si preferisce indicarle con a_n oppure y_m oppure z_k e simili, per esempio $y_n := n!$ e il fattoriale è definito così:

$$n! := \begin{cases} 1 & \text{se } n = 0 \\ 1 \cdot 2 \cdot \ldots \cdot n & \text{se } n > 0 \end{cases}$$

Un'altra è la successione di Fibonacci:

$$a_n ... \text{ successione di Fibonacci}$$

e poi abbiamo queste altre, dei reciproci e dei quozienti:

$$r_n := \frac{1}{a_n} \quad a_n \text{ successione di Fibonacci, } n = 1, 2...$$

$$q_n := \frac{a_{n+1}}{a_n} \quad a_n \text{ successione di Fibonacci, } n = 1, 2...$$

Naturalmente si possono ottenere successioni anche restringendo il dominio di funzioni elementari ai soli numeri naturali, per esempio

$$z_n := \frac{e^n + e^{-n}}{2}$$

ma non vogliamo occuparci di queste in questa Lezione.

BOZZA - DRAFT

[127] A un livello superiore viene definito, con la funzione gamma.

24.5 Limite finito delle successioni

Ecco alcuni valori delle successioni a_n e q_n del paragrafo precedente:

Indice n di a_n	Fibonacci a_n	Quoziente $q_n = a_{n+1}/a_n$	Scrittura decimale esatta o approx
0	0		
1	1	1/1	1
2	1	2/1	2
3	2	3/2	1.5
4	3	5/3	1.66667
5	5	8/5	1.6
6	8	13/8	1.625
7	13	21/13	1.61538
8	21	34/21	1.61905
9	34	55/34	1.61765
10	55	89/55	1.61818
11	89	144/89	1.61798
12	144	233/144	1.61806
13	233	377/233	1.61803
14	377	610/377	1.61804
15	610	987/610	1.61803

BOZZA - DRAFT

Ci viene l'idea che q_n "tende" a 1.618... per n che "tende" all'infinito. In effetti tende alla sezione aurea $\varphi = \frac{1+\sqrt{5}}{2} = 1.618...$ ma non possiamo dimostrarlo al livello di questo testo elementare.

Scriveremo

$$\lim_{n\to+\infty} q_n = \frac{1+\sqrt{5}}{2} = 1.618...$$

e leggeremo: "limite per n che tende a $+\infty$ di q_n uguale", eccetera.

La teoria dei limiti viene svolta in Analisi Matematica a un livello piuttosto complesso ma noi ci accontenteremo di RAGIONAMENTI EURISTICI: in pratica di ragionamenti approssimativi basati sul guardare i numeri e i grafici, e su poche formule.

24.6 Limite nullo delle successioni

Ecco alcuni valori delle successioni a_n e r_n del paragrafo 24.4:

Indice n di a_n	Fibonacci a_n	Reciproco $r_n = 1/a_n$	Scrittura decimale esatta o approx
0	0		
1	1	1/1	1
2	1	1/1	1
3	2	1/2	0.5
4	3	1/3	0.333333
5	5	1/5	0.2
6	8	1/8	0.125
7	13	1/13	0.0769231
8	21	1/21	0.047619
9	34	1/34	0.0294118
10	55	1/55	0.0181818
...	...	...	...
25	75025	1/75025	$1.33289 \cdot 10^{-5}$
26	121393	1/121393	$8.23771 \cdot 10^{-6}$
27	196418	1/196418	$5.09118 \cdot 10^{-6}$
28	317811	1/317811	$3.14652 \cdot 10^{-6}$
29	514229	1/514229	$1.94466 \cdot 10^{-6}$
30	832040	1/832040	$1.20187 \cdot 10^{-6}$
31	1346269	1/1346269	$7.42794 \cdot 10^{-7}$
32	2178309	1/2178309	$4.59072 \cdot 10^{-7}$

BOZZA - DRAFT

Ipotizziamo ed è vero e si potrebbe dimostrare che

$$\lim_{n \to +\infty} r_n = 0$$

Naturalmente il limite 0 è un caso particolare del caso del limite finito considerato nel paragrafo precedente. Tuttavia, in vista di varie formule che vedremo, il limite 0 va considerato separatamente dagli altri limiti finiti. Similmente

$$\lim_{n \to +\infty} \frac{1}{n!} = 0$$

24.7 I 2 possibili limiti infiniti delle successioni

Ci sono 2 limiti infiniti, il $+\infty$ e il $-\infty$, che adesso spieghiamo.

Guardando i valori a_n della successione di Fibonacci nella tabella del precede paragrafo, e magari anche alcuni successivi, ipotizziamo ed è vero e si potrebbe dimostrare che essi diventano definitivamente QUANTOMAI GRANDI, e più precisamente che fissata una qualunque soglia M, da un certo punto in poi $a_n > M$.
Diremo che il limite è $+\infty$ e scriveremo

$$\lim_{n\to+\infty} a_n = +\infty \qquad a_n \text{ successione di Fibonacci}$$

Similmente avviene coi fattoriali:

$$\lim_{n\to+\infty} n! = +\infty$$

Analogamente si definisce anche il limite $-\infty$, per esempio

$$\lim_{n\to+\infty} \lg \frac{1}{n!} = -\infty$$

in particolare il logaritmo decimale di $1/70!$ è già circa -100.

BOZZA DRAFT

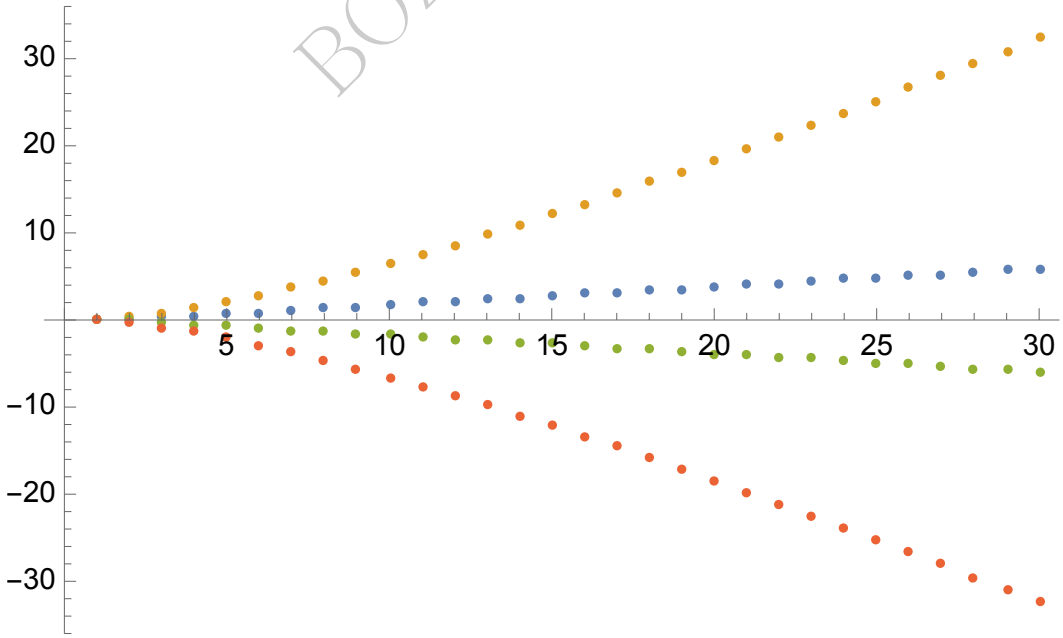


Figure 34: Dall'alto verso il basso, in scala logaritmica, grafici dei valori di $n!$, a_n, $1/a_n$ e $1/n!$ essendo a_n la successione di Fibonacci.

24.8 Sulla rapidissima crescita del fattoriale

Il fattoriale va all'infinito

$$\lim_{n\to+\infty} n! = +\infty$$

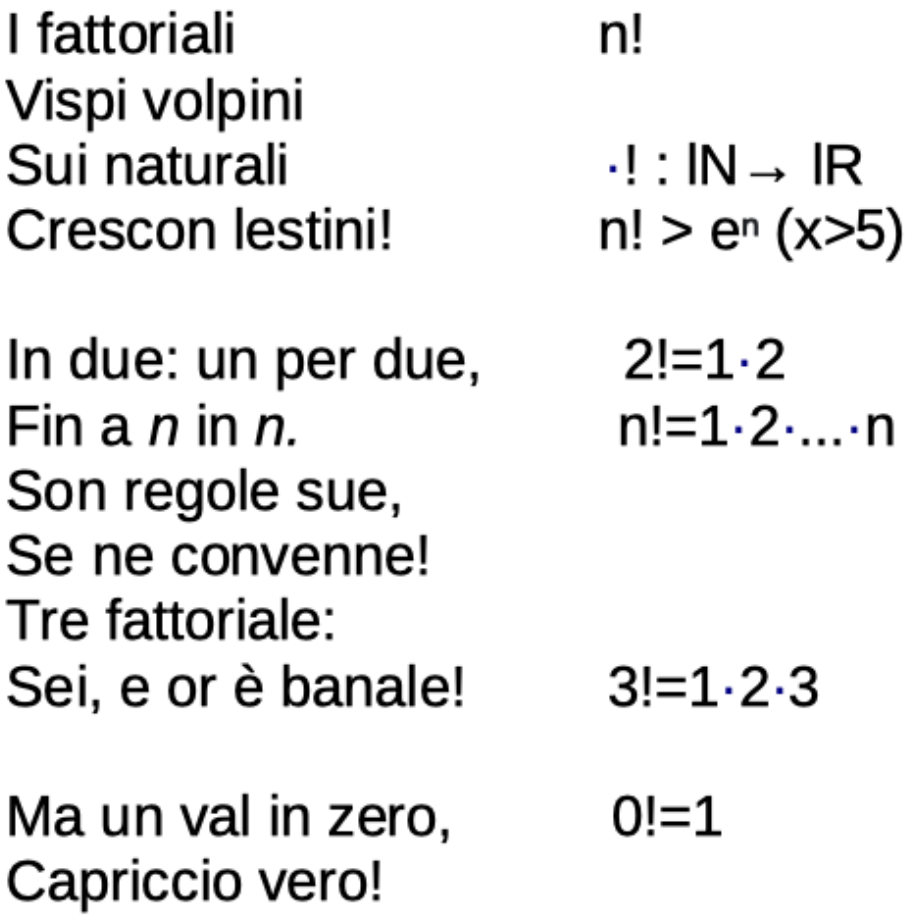
rapidamente, perfino più dell'esponenziale, e per $n \geq 6$ supera e^n.

(Da $n = 6$ in poi il fattoriale sta sopra l'esponenziale

$$\forall\, n > 5 \qquad n! > e^n$$

$$\text{e anzi addirittura } \lim_{n\to+\infty} \frac{n!}{e^n} = +\infty$$

cioè $n!$ da un certo punto in poi è più che doppio di e^n, poi più che triplo, poi più che quadruplo, eccetera).



24.9 Serie geometrica di ragione fra 0 e 1

Le *serie* possono essere viste come una sorta di "somma infinita", ottenuta come limite delle somme finite iniziali.

Ci interessa, per $a, r \in \mathbb{R}$, il limite di questa successione numerica:

$$
\begin{aligned}
&a \\
&a + a\,r \\
&a + a\,r + a\,r^2 \\
&a + a\,r + a\,r^2 + a\,r^3 \\
&\ldots \\
&a + a\,r + a\,r^2 + a\,r^3 + \ldots + a\,r^k
\end{aligned}
$$

BOZZA DRAFT

che definisce la **serie geometrica** di *ragione* r (mentre a potremmo chiamarlo "fattore moltiplicativo").

Ci occuperemo approfonditamente della sola *serie geometrica di ragione r compresa fra 0 e 1 esclusi*, che ricorre ampiamente nel Calcolo delle Probabilità:

$$
\boxed{\begin{array}{l} a + a\,r + a\,r^2 + \ldots + a\,r^k + \ldots = \sum_{k=0}^{\infty} a\,r^k = \ \frac{a}{1-r} \\ a \in \mathbb{R} \\ 0 < r < 1 \end{array}} \tag{72}
$$

(La formula vale anche per $-1 < r < 1$ ma ci interessano solo i valori $0 < r < 1$, e poi il caso $r = 0$ è ovvio).
Nel Calcolo delle Probabilità è sempre $a > 0$, e comunque questo è il caso più significativo in tutto l'ambito Farmaceutico e Biomedico.

Esempio.

$$
3 + \frac{3}{2} + \frac{3}{4} + \frac{3}{8} + \ldots + \frac{3}{2^k} + \ldots \text{ ovvero } \sum_{k=0}^{\infty} 3\left(\frac{1}{2}\right)^k \tag{73}
$$

è una (particolare) *serie geometrica* di *ragione* $\frac{1}{2}$. E in base alla (72) con $a = 3$ (la sua somma) vale

$$\frac{3}{1 - 1/2}$$

cioè 6.

BOZZA - DRAFT

Complementi

24.10 Complementi – Problema aperto: Collatz

Dato un qualsiasi numero intero $n > 1$, considera la seguente successione:

Se n è pari, allora dividilo per 2.

Se n è dispari, allora moltiplicalo per 3 e aggiungi 1.

Ripeti questo processo indefinitamente per ogni nuovo valore di n. Si arriverà sempre al numero 1 prima o poi?

Nota. La *Congettura di Collatz* risponde sì, ma non si ancora.

BOZZA - DRAFT

24.11 Complementi – Problema aperto: Feigenbaum

Una complessa costruzione analitica definisce una successione che numericamente, col computer, si vede tendere al numero

$$\delta = 4.6692...$$

detto costante di Feigenbaum, che ricorre ampiamente nelle Scienze, in particolare nelle Teorie del Caos e dei Frattali. Tuttavia, non si è ancora trovata una forma chiusa per quel numero, che attualmente può solo calcolarsi sperimentalmente, per così dire, e potrebbe racchiudere misteri notevoli della Matematica.

24.12 Complementi – Esperienza pratica coi limiti

Si provi con la calcolatrice a fare la radice quadrata di un numero positivo scelto a piacere, poi a fare la radice quadrata del risultato, e così via: a quale limite tende la successione dei valori?

24.13 Complementi – 4 tipi di definizioni di successioni

Semplificando al massimo, le successioni di possibile interesse Farmaceutico e Biomedico hanno definizioni di 4 tipi che ora vedremo

con riferimento al loro eventuale limite.

Esempio 1 del tipo I: successioni ricorrenti. Consideriamo ora la successione a_n di Fibonacci, a suo tempo definita per ricorrenza:

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144...$$

Ci sembra, ed è vero e si potrebbe dimostrare, che i valori supereranno qualunque soglia, crescendo indefinitamente. Scriveremo

$$\lim_{n\to+\infty} a_n = +\infty \quad (a_n \text{ successione di Fibonacci})$$

e il senso del limite $+\infty$ è

$$\lim_{n\to+\infty} a_n = +\infty \quad \Leftrightarrow \quad (\forall\, M)\,(\exists\, \bar{n})\; n > \bar{n} \Rightarrow a_n > M.$$

Consideriamo ora lo spazio b_n che ha a disposizione ogni microbo nel volume iniziale fisso, diciamolo 1 in qualche opportuna unità di misura, per esempio il centimetro cubo, il pollice cubo, o una unità di misura da noi inventata in modo che valga proprio 1:

$$b_n := \frac{1}{a_n}$$

BOZZA - DRAFT

I valori sono

$$\frac{1}{1}, \frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \frac{1}{5}, \frac{1}{8}, \frac{1}{13} \ldots \quad (a_n \text{ successione di Fibonacci})$$

Ci *sembra*, ed è vero e si potrebbe dimostrare (quando si definisce come più sotto), che b_n *tende a 0 per n che tende all'infinito*:

$$\lim_{n\to+\infty} b_n = 0\,.$$

Più in generale si dà il caso, con altra successione y_n, e $L \in \mathbb{R}$,

$$\lim_{n\to+\infty} y_n = L \qquad (\text{o invece di n, qualunque variabile, p. es. } k).$$

Il significato è di un indefinito avvicinamento, con o senza raggiungimento del limite L. Più precisamente

$$(\forall\, \varepsilon > 0)\,(\exists\, \bar{n})\; n > \bar{n} \Rightarrow |y_n - L| < \varepsilon\,.$$

Cioè, fissato un numer(ett)o positivo, chiamiamolo ε, da un certo punto in poi y_n dista dal limite meno di ε.

Esempio 2 del tipo I: successioni ricorrenti – il Modello Malthusiano. La popolazione – di microbi o quant'altro – si espande nel tempo scandito da n, proporzionalmente alla sua numerosità (detto semplificatamente: a molti microbi seguono quegli stessi microbi più molti figli di microbi, proporzionalmente al tasso di accrescimento):

$$a_{n+1} := a_n + c \cdot a_n \qquad a_0 := \text{popolazione iniziale}$$

con c il *tasso di accrescimento*. Ecco per esempio il caso $c := 2$:

$$a_0,\ 3\,a_0,\ 9\,a_0,\ 27\,a_0 ...$$

per esempio con $a_0 := 1000$

$$1\,000, 3\,000, 9\,000, 27\,000 ...$$

Si può dimostrare che la successione a_n ammette una rappresentazione *in forma chiusa*

$$a_n = (c+1)^n\, a_0$$

(che è funzione "di tipo esponenziale", in inglese si parla di *exponential functions*) e se $c > 0$ si ha un vero e proprio accrescimento verso $+\infty$:

$$\lim_{n\to+\infty} a_n = +\infty \qquad \text{se } c > 0$$

e il senso del limite $+\infty$ è quello prima detto.
Invece se $-1 < c < 0$ il comportamento è molto diverso, per esempio con $c := -\frac{1}{2}$ abbiamo i valori

$$a_0,\ \frac{1}{2}a_0,\ \frac{1}{4}a_0,\ \frac{1}{8}a_0 ...$$

Ci *sembra*, ed è vero e si potrebbe dimostrare (quando si definisce come più sotto), che a_n *tende a 0 per n che tende all'infinito*:

$$\lim_{n\to+\infty} a_n = 0\,.$$

e il senso del limite 0 è quello prima detto per L, numero.

> Si provi con stessa successione del Modello Malthusiano a scrivere i primi 7 valori numerici con $a_0 := 100\,000$, prima con $c := -0.3$ e poi con $c := 0.3$.

E ovviamente con $c := 0$ abbiamo un equilibrio fra nati e morti ovvero tasso di accrescimento nullo e in quel caso la popolazione resta costante:

$$\lim_{n\to+\infty} a_n = a_0 \qquad \text{se } c = 0.$$

BOZZA - DRAFT

Nota sull'accrescimento di Fibonacci o Malthus. Una popolazione microbica o altra, che si espandesse con quelle leggi, tenderebbe ad un'espansione infinita; nella realtà ad un certo punto interverranno meccanismi che di fatto sospenderanno la validità della legge di Malthus o di Fibonacci nel rappresentare la numerostà della popolazione. Questo avviene, se non altro, perché gli organismi da un certo punto in poi inevitabilmente si ostacolano a vicenda in modo significativo, anche per mancanza di spazio, come visto nell'Esempio 1 del Tipo II.
Così, piuttosto che tendere all'infinito esponenzialmente, ad un certo punto l'accrescimento in generale rallenterà, producendo un tratto di sigmoide. Ed eventualmente poi decrescerà, fino ad estinguersi o quasi, producendo un grafico più meno a campana: si veda la figura a questo link->.

BOZZA - DRAFT

Esempio del tipo II: successioni definite "coi puntini". Il fattoriale

$$n! := \begin{cases} 1 & \text{se } n = 0 \\ 1 \cdot 2 \cdot \ldots \cdot n & \text{se } n > 0 \end{cases}$$

assume i valori

$$1, 1, 2, 6, 24, 120, 720, 5040...$$

Si ha

$$\forall\, n > 0 \qquad n! = 1 \cdot 2 \cdot \ldots \cdot n \geq n$$

e già n tende a $+\infty$, e allora anche $n!$ che come appena visto gli sta sopra:

$$\boxed{\lim_{n \to +\infty} n! = +\infty} \tag{74}$$

e il senso del limite $+\infty$ è quello sopra esposto, valido per qualunque successione.
(Il fattoriale ricorre ampiamente nel Calcolo Combinatorio, nel Calcolo delle Probabilità, nelle Serie Numeriche...).
Naturalmente altre successioni definite coi puntini tendono ad altri

valori, per esempio

$$\lim_{n\to+\infty} \frac{1}{1+2+...+n} = 0$$

com'è ovvio.

Esempio 1 del tipo III: successioni con $(-1)^n$. La successione $(-1)^n$ ha i valori

$$+1, -1, +1, -1, +1...$$

e non esiste il limite:

$$\boxed{\nexists \lim_{n\to+\infty} (-1)^n} \tag{75}$$

Questa successione consente nella programmazione informatica – per esempio con l'onnipresente in campo commerciale e scientifico Excel – di distinguere i numeri pari dai numeri dispari. (Esistono anche altri modi). In particolare

$$\frac{3+(-1)^n}{2}$$

produce, a partire da $n := 1$, i valori

$$1, 2, 1, 2, 1, 2, 1...$$

che può consentire per esempio di produrre i turni in farmacia di Aldo e Bianca nei vari giorni dell'anno, alternandoli secondo il numero ordinale del giorno (per esempio al 1 febbraio corrisponde il numero 32, pari: lavora Bianca). Oppure per programmare con un software l'apertura automatizzata di sportellini 1 e 2, a giorni alterni, in un allevamento di animaletti.

> **Esempio informatico.** Si consideri il display ai LED programmabile della farmacia *Al Cuore Vispo* che dispone della funzione
>
> `ordinalNumberOfDay`
>
> e vuole esporre a giorni alterni i messaggi

BOZZA - DRAFT

```
STRING(1):=Buongiorno, allegro ti sia il giorno!
STRING(2):=I nostri clienti campano cent'anni!
```

Ciò si potrà ottenere programmando qualcosa come

```
display(STRING((3+(-1)^ordinalNumberOfDay)/2))
```

Esempio 2 del tipo III: successioni con $(-1)^n$. La successione $(-1)^n \cdot n$ ha i valori

$$+1, -1, +2, -3, +4, -5, +6, -7, +8, -9, +10...$$

e alcuni Autori dicono che va a ∞, l'infinito senza segno, ma in questa trattazione elementare diremo invece che il limite non esiste

$$\nexists \lim_{n\to+\infty} (-1)^n \cdot n$$

e similmente per successioni che oscillano senza "assestarsi".

Esempio 3 del tipo III: successioni con $(-1)^n$. La successione $\frac{(-1)^n}{10^n}$ ha i valori, per $n \geq 1$,

$$-0.1, +0.01, -0.001, +0.0001...$$

e si capisce bene che

$$\lim_{n\to+\infty} a_n = 0.$$

Similmente

$$\lim_{n\to+\infty} \frac{(-1)^n}{n} = 0 \quad \text{e} \quad \lim_{n\to+\infty} \frac{(-1)^n}{\sqrt{n}} = 0$$

e similmente il limite è 0 se il numeratore è *limitato* (cioè si mantiene fra 2 numeri), e il denominatore $\to +\infty$.

Successioni del tipo IV: prolungabili ai numeri reali. Per successioni come

$$a_n := \frac{1}{n} \qquad b_n := 2^{-n} \qquad c_n := \left(1 - \frac{1}{n}\right)^n$$

che potrebbero essere prolungate ovvero "estese" ai numeri reali

$$\alpha(x) := \frac{1}{x} \qquad \beta(x) := 2^{-x} \qquad \gamma(x) := \left(1 - \frac{1}{x}\right)^x$$

BOZZA - DRAFT

(si noti che ciò non si può fare banalmente per i tipi I, II e III) il limite a $+\infty$ sarà trattato nella prossima lezione; per quanto sia già ora ovvio, in base a tutta la trattazione svolta, che

$$\lim_{n\to+\infty} 2^{-n} = 0 \qquad \lim_{n\to+\infty} \frac{1}{n} = 0.$$

Invece il limite di c_n e di $\gamma(x)$ non è ovvio per niente, e lo vedremo.

24.14 Complementi – Applicazione probabilistica

Seppure il Calcolo delle Probabilità lo tratteremo in seguito, ritenendo noto qualche concetto dagli studi scolastici vediamo un esempio di applicazione della serie geometrica, per mostrarne l'utilità.

BOZZA - DRAFT

ESERCIZIO$_{\mu 2018\sharp}$
$\star \approx \%$ Una certa affezione, quando insorge ha una durata in giorni (interi) con probabilità

$\frac{1}{2}$ probabilità di durare 1 giorno
$\frac{1}{4}$ probabilità di durare 2 giorni
$\frac{1}{8}$ probabilità di durare 3 giorni
...
$\frac{1}{2^n}$ probabilità di durare n giorni.

(Che si chiama *densità geometrica iniziante da 1 di parametro* $\frac{1}{2}$).
Qual è la probabilità che duri un numero pari di giorni?

Svolgimento
Detta X la durata dell'affezione, scriviamo (usando la notazione delle variabili aleatorie, che tratteremo in seguito, ma che comunque qua appare di interpretazione ovvia)

$$X = \begin{pmatrix} 1 & 2 & 3 & 4 & \dots & n & \dots \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{16} & \cdots & \frac{1}{2^n} & \dots \end{pmatrix}$$

$$P(X \text{ pari}) = P(X \in \{2, 4, 6 \dots\}) =$$

$$= P(X = 2) + P(X = 4) + P(X = 6) + ... =$$
$$= \frac{1}{4} + \frac{1}{16} + \frac{1}{64} + ...$$

che si riconosce esser serie geometrica di ragione $r = \frac{1}{4}$ (perché ogni termine è pari al precedente moltiplicato per $\frac{1}{4}$)

$$a + ar + ar^2 + ar^3 + ... + ar^n + ...$$

con anche $a = \frac{1}{4}$ e allora somma

$$\frac{a}{1-r} = \frac{\frac{1}{4}}{1-\frac{1}{4}} = \frac{\frac{1}{4}}{\frac{3}{4}} =$$

$$\boxed{\frac{1}{3} \approx 0.333 = 33.3\%}$$

(Si noti che allora è più probabile che duri un numero dispari di giorni. Questo rimane vero con qualunque valore di $p \in]0, 1[$, non solo $\frac{1}{2}$, e perfino con qualunque densità decrescente definita su 1,2,3... A Trieste si dice che il vento bora tende a durare un numero dispari di giorni).

BOZZA - DRAFT

25 Limiti e continuità per funzioni reali

25.1 Introduzione ai limiti di funzioni

Ci occupiamo del comportamento di una funzione $f(x)$

$$f : I \to \mathbb{R}$$

essendo I

intervallo (p.es. $]a,b[$ o $[a,b[$ con eventualmente b infinito) oppure unione finita di intervalli (per esempio l'unione di $]-\infty,0[$ e $]0,+\infty[$) oppure

unione infinita non troppo "capricciosa" di intervalli (per esempio $\operatorname{dom}\tan x$)

quando $x \to +\infty$, com'era nelle successioni, oppure

$x \to -\infty$ oppure

$x \to x_0 \in \mathbb{R}$, cioè un "numero finito":

$$\lim_{x \to qualcosa} f(x)$$

BOZZA - DRAFT

Esempi.

$\lim_{x\to+\infty} \frac{10}{x} = 0$ esempio termodinamico: $V = \frac{cost}{p}$ (isoterma)

al tendere all'infinito della pressione il volume andrebbe a 0, se il gas restasse ideale.

$\lim_{x\to0^+} \frac{10}{x} = +\infty$ esempio termodinamico: $p = \frac{cost}{V}$ (isoterma)

al tendere a 0 del volume la pressione andrebbe all'infinito, se il gas restasse ideale.

La scrittura $x \to x_0^+$, qua sopra, significa che x tende a 0 da destra ovvero da valori maggiori di x_0. E similmente esiste $x \to x_0^-$.

Dai grafici – e per tutte esisterebbe una vera dimostrazione – apprendiamo che

$\lim_{x\to+\infty} x^2 = +\infty$ e similmente	$x^1, x^3 \ldots x^\alpha$ *con* $\alpha > 0, \sqrt{x}, \sqrt[n]{x}$
$\lim_{x\to+\infty} 10^x = +\infty$	$\lim_{x\to+\infty} \mathrm{e}^x = +\infty$
$\lim_{x\to+\infty} b^x = +\infty, \quad b > 1$	e 0 se $0 < b < 1$
$\lim_{x\to-\infty} 10^x = 0$	$\lim_{x\to-\infty} \mathrm{e}^x = 0$
$\lim_{x\to-\infty} b^x = 0, \quad b > 1$	e $+\infty$ se $0 < b < 1$
$\lim_{x\to+\infty} \lg x = +\infty$	$\lim_{x\to+\infty} \ln x = +\infty$
$\lim_{x\to 0} \lg x = -\infty$	$\lim_{x\to 0} \ln x = -\infty \quad (0^+$ in effetti)
$\lim_{x\to-\infty} x^{\textit{dispari positivo}} = -\infty$	$\lim_{x\to-\infty} \sqrt[\textit{dispari positivo}]{x} = -\infty$
$\lim_{x\to 0^-} \frac{1}{x} = -\infty$	$\lim_{x\to 0^+} \frac{1}{x} = +\infty$
$\lim_{x\to-\infty} \arctan x = -\frac{\pi}{2}$	$\lim_{x\to+\infty} \arctan x = \frac{\pi}{2}$
$\nexists \lim_{x\to+\infty} \sin x$	similmente $-\infty$, *e* cos *e* tan

(76)

BOZZA DRAFT

Esempio 1. Guardando il grafico, per la campana gaussiana vale

$$\lim_{x\to+\infty} \frac{\mathrm{e}^{-\frac{x^2}{2}}}{\sqrt{2\pi}} = 0$$

e similmente per $x \to -\infty$

$$\lim_{x\to-\infty} \frac{\mathrm{e}^{-\frac{x^2}{2}}}{\sqrt{2\pi}} = 0$$

e scriveremo i 2 limiti insieme in questo modo:

$$\lim_{x\to\pm\infty} \frac{\mathrm{e}^{-\frac{x^2}{2}}}{\sqrt{2\,\pi}} = 0$$

Esempio 2. La concentrazione del piombo nelle ossa

$$u(t) = u_0 \cdot \mathrm{e}^{-\frac{\ln 2}{10\,anni}\,t} \quad \text{essendo } u_0 = u(0)$$

al tendere del tempo t all'infinito:

$$\lim_{t\to+\infty} u_0 \cdot \mathrm{e}^{-\frac{\ln 2}{10\,anni}\,t} =$$

per proprietà delle potenze

$$= \lim_{t\to+\infty} u_0 \cdot \left(\left(\mathrm{e}^{\ln 2}\right)^{-1}\right)^{\frac{t}{10\,anni}} =$$

per proprietà dei logaritmi

$$= \lim_{t\to+\infty} u_0 \cdot \left(2^{-1}\right)^{\frac{t}{10\,anni}} =$$

per definizione della potenza con esponente intero negativo

$$= \lim_{t\to+\infty} u_0 \cdot \left(\frac{1}{2}\right)^{\frac{t}{10\,anni}} =$$

ed essendo la base fra 0 e 1 esclusi, e l'esponente va a $+\infty$,

$$= 0$$

Cioè la concentrazione va a 0 al tendere del tempo t all'infinito.

BOZZA - DRAFT

25.2 Limiti verso un numero finito, e continuità

Consideriamo il limite di f per $x \to x_0$ con $x_0 \in \mathbb{R}$, anche distinguendo x_0^+ e x_0^-, e ci sono 2 casi:
$\diamond$ $x_0 \notin \mathrm{dom} f$, p. es. $\lim_{x\to 0^-} \frac{x}{|x|} = -1$, $\lim_{x\to 0^-} \frac{1}{x} = -\infty$
$\diamond$ $x_0 \in \mathrm{dom} f$ e ci sono 2 sottocasi:

$$\boxed{\bullet \quad \lim_{x\to x_0} f(x) = f(x_0) : \text{ la funzione si dice continua}} \tag{77}$$

e si noti che

$$\boxed{\text{le funzioni elementari sono continue nei loro domini}} \qquad (78)$$

e allora per esse sempre limite = valore calcolato, per esempio

$$\lim_{x\to 0} \mathrm{e}^x = 1$$

- $\lim_{x\to x_0} f(x) \neq f(x_0)$: la f si dice discontinua in x_0 e questo può essere con funzioni non elementari come $\operatorname{sgn}(x)$ e $\lfloor x \rfloor$.

Allora per le funzioni elementari sono significativi solo i limiti dove la funzione "smette di esistere": gli estremi, finiti o no, degli intervalli che compongono il dominio, non appartenenti ad esso:

$$\lim_{x\to 1} \frac{1-\sqrt{x}}{1-x} = \text{ [forma 0 su 0] } \lim_{x\to 1} \frac{1-\sqrt{x}}{(1-\sqrt{x})(1+\sqrt{x})} =$$

$$= \lim_{x\to 1} \frac{1}{1+\sqrt{x}} \text{ [funzione elementare, limite=valore] } = \frac{1}{2}\,.$$

BOZZA - DRAFT

25.3 Funzioni limitate, infinite, infinitesime

Si dice *limitata* una f tale che esistono 2 numeri M e N tali che $M < f(x) < N$, per esempio $\sin x$. Limitata $\not\Rightarrow \exists$ limite.

Si dice *infinita* una funzione che tende a $+\infty$ o $-\infty$ e *infinitesima* una che tende a 0 (per $x \to u_0$ finito o no).
Per esempio sono infinite per $x \to +\infty$ le b^x con $b > 1$, e infinitesime per $0 < b < 1$; viceversa per $x \to -\infty$. Infinite le $\log_b x$ per $x \to 0^+$ e $x \to +\infty$.

25.4 Teoremi algebrici sui limiti

Tutti questi teoremi algebrici sui limiti valgono anche per le successioni, e per buona chiarezza sarebbe meglio riscriverli con n al posto di x.

Naturalmente nel caso delle successioni n può tendere solo a $+\infty$ mentre nel caso delle funzioni la x può tendere a $+\infty$, $-\infty$ oppure a un numero.

Risolveremo i limiti un po' in modo intuitivo, come abbiamo già ampiamente fatto, anche guardando il grafico, e anche esaminando "a pezzetti" le espressioni delle funzioni; in questa trattazione elementare non possiamo fare una teoria dei limiti molto approfondita. Tuttavia qua formalizziamo alcune "regole" di calcolo dei limiti, che in parte abbiamo già usato.

Con le definizioni dei limiti (con $\varepsilon, M...$) si dimostra che:

BOZZA - DRAFT

(a)	$L_1, L_2 \in \mathbb{R}:$	$tende\ a\ L_1+$		
		$+tende\ a\ L_2 \to L_1 + L_2$	$(p.es\ \frac{1}{x} + \arctan x$	$per\ x \to +\infty)$
(b)	$L_1, L_2 \in \mathbb{R}:$	$tende\ a\ L_1 \cdot$		
		$\cdot tende\ a\ L_2 \to L_1 \cdot L_2$	$(p.es\ \frac{1}{x} \arctan x$	$per\ x \to +\infty)$
(c)	$L \in \mathbb{R} \setminus \{0\}:$	$\frac{1}{tende\ a\ L} \to \frac{1}{L}$	$(p.es.\ 1/\arctan x$	$per\ x \to +\infty)$

(79)

Per quanto gli infiniti non siano numeri e le espressioni seguenti siano scorrette matematicamente, valgono come mnemonici:

(a)		$+\infty + \infty \to +\infty$	$(p.es.\ x^3 + \mathrm{e}^x$	$per\ x \to +\infty)$
(b)		$+\infty \cdot (+\infty) \to +\infty$	$(p.es.\ x^3\, \mathrm{e}^x$	$per\ x \to +\infty)$
(c)	$L \in \mathbb{R}:$	$+\infty + L \to +\infty$	$(p.es.\ x^2 + \arctan x$	$per\ x \to +\infty)$
(d)	$L \in \mathbb{R}^+:$	$+\infty \cdot L \to +\infty$	$(p.es.\ x^3 \arctan x$	$per\ x \to +\infty)$
(d)	$-L \in \mathbb{R}^+:$	$+\infty \cdot L \to -\infty$	$(p.es.\ -2\,x^3$	$per\ x \to +\infty)$
(e)		$\frac{1}{+\infty} \to 0$	$(p.es.\ 1/x^3$	$per\ x \to +\infty)$
(f)		$limitata + (+\infty) \to +\infty$	$(p.es.\ \sin x + x^3$	$per\ x \to +\infty)$
(g)		$\frac{limitata}{+\infty} \to 0$	$(p.es.\ \sin x/x^3$	$per\ x \to +\infty)$

(80)

E anche:

• reciproca d'infinitesima > 0, $\to +\infty$, p.es. $\lim_{x \to \pi^-} \frac{1}{\sin x} = +\infty$

• reciproca d'infinitesima < 0, $\to -\infty$, p.es. $\lim_{x\to 1^-} \frac{1}{\ln x} = -\infty$

Esempio: logistica. Consideriamo il caso ultra-semplificato:

$$\lim_{t\to+\infty} \frac{1}{1+\mathrm{e}^{-t}} =$$

$$= \lim_{t\to+\infty} \frac{1}{1+\frac{1}{\mathrm{e}^{t}}} = \frac{1}{1+0} = 1.$$

Una forma più generale dà, con costanti positive K, q, r,

$$\lim_{t\to+\infty} \frac{K}{1+q\,\mathrm{e}^{-r\,t}} = K$$

(Si veda Wikipedia alla voce *Equazione logistica*).

BOZZA DRAFT

25.5 Esistenza di forme di indecisione dei limiti

Restano fuori queste **forme di indecisione**, da risolvere caso per caso: $+\infty - \infty$, $\frac{0}{0}$, $\frac{\infty}{\infty}$, $0\cdot\infty$, ∞^0, 1^∞, 0^0. Si veda nella sezione di Complementi.

Note. Tutte le funzioni considerate in questa lezione sono definite in singoletti o intervalli o unione (finita o non troppo "capricciosa" se infinita) di singoletti e/o intervalli.

I singoletti non hanno rilevanza per quanto riguarda i limiti.

Sono significativi solo gli intervalli del dominio delle funzioni.

Per esempio la funzione $\frac{1}{x}$ è definita nell'unione dei 2 intervalli $]-\infty, 0[$ e $]0, +\infty[$. Considereremo i limiti a quegli estremi: $-\infty$, 0^-, 0^+, $+\infty$.

Similmente consideremo quei 4 limiti, e soprattutto quelli a 0^+ e $+\infty$, per

$$\frac{1}{x^2}$$

Questa sopra è una funzione che ricorre ampiamente in Fisica, per esempio l'attrazione fra 2 cariche di segno opposto è[128]

$$F = C\,\frac{q_1\cdot q_2}{r^2}$$

[128] In modulo, non vettorialmente.

essendo q_1 e q_2 i valori, senza segno, delle 2 cariche, ovvero come funzione della sola distanza (cioè per cariche fissate)

$$F(r) = C \cdot q_1 \cdot q_2 \cdot \frac{1}{r^2}$$

(E dal punto di vista fisico è significativa solo per $r > 0$). Si ha

$$\lim_{r \to +\infty} C \cdot q_1 \cdot q_2 \cdot \frac{1}{r^2} = 0 \quad \text{forza nulla a distanza infinita}$$

$$\lim_{r \to 0^+} C \cdot q_1 \cdot q_2 \cdot \frac{1}{r^2} = +\infty \quad \text{forza infinita a distanza nulla}$$

Nota. Con la teoria svolta, lo studente volonteroso potrà risolvere questi esercizi. Per $x\sqrt[3]{x}$, $\log_2 |2^x - 1|$, $\ln |x|$, $\sin 2^x$, $\cos \pi x$,

$$\frac{x^3 - 8}{x^2 - 2x}, \quad \frac{4 + 3x - x^2}{x^2 - 2x - 8} \text{ questi 8 limiti: } \lim_{x \to \pm\infty}, \lim_{x \to -2^{\pm}}, \lim_{x \to 0^{\pm}}, \lim_{x \to 4^{\pm}}.$$

Gli stessi 8 limiti per queste 3 funzioni: $\dfrac{2^x - 2^{-x}}{3^x \pm 3^{-x}}$, $\quad \lg \left| 1 - \dfrac{2}{x} \right|$.

DOZZA - DRAFT

Nota. Di questa Lezione ci si dovrebbe aspettare che lo studente conosca a memoria le formule *contemporaneamente* numerate *et* riquadrate, oltre a tutto l'inquadramento teorico della questione.

Complementi

25.6 Complementi – Forme di indecisione dei limiti

Restano fuori queste **forme di indecisione**, da risolvere caso per caso: $+\infty - \infty$, $\frac{0}{0}$, $\frac{\infty}{\infty}$, $0 \cdot \infty$, ∞^0, 1^∞, 0^0. Per esempio

$$\lim_{x \to +\infty} \left(\mathrm{e}^x - x\right) =$$

$$\left[= +\infty - \infty\right]$$

$$= \lim_{x \to +\infty} x \left(\frac{\mathrm{e}^x}{x} - 1\right) =$$

$$= \left[+\infty \cdot \left(+\infty\right) =\right]$$

$$= +\infty.$$

BOZZA - DRAFT

25.7 Complementi – Limiti di successioni prolungabili

Di successioni come

$$\ln n \qquad 2^n$$

$$\varphi^n \quad \left(1 - \tfrac{1}{n}\right)^n$$

cioè prolungabili ovvero "estendibili" da $\mathbb{N}$ a $\mathbb{R}$, per il limite a $+\infty$ considereremo il corrispondente limite delle funzioni prolungate

$$\ln x \qquad 2^x$$

$$\varphi^x \quad \left(1 - \tfrac{1}{x}\right)^x$$

cioè considereremo il limite in $\mathbb{R}$, ottenendo lo stesso risultato perché (teorema) è

$$\text{se } \lim_{x \to +\infty} f(x) = u \quad \text{allora} \quad \lim_{n \to +\infty} f(n) = u$$

Per esempio esiste un limite classico, di difficile dimostrazione matematica e ampia ricorrenza in Calcolo delle Probabilità,

$$\lim_{n\to+\infty}\left(1+\frac{1}{n}\right)^n = \lim_{x\to+\infty}\left(1+\frac{1}{x}\right)^x = \mathrm{e}$$

spesso detto, in una delle 2 forme soprascritte (Primo, oppure Secondo) Limite Fondamentale, e da esso con complicati calcoli segue

$$\lim_{n\to+\infty}\left(1-\frac{1}{n}\right)^n = \lim_{x\to+\infty}\left(1-\frac{1}{x}\right)^x = \frac{1}{\mathrm{e}}$$

che ritroveremo, nella prima forma, in questioni di Farmacia.

BOZZA - DRAFT

Esempio 1.

$$\lim_{x\to+\infty}(7x^2 - x + 1) =$$

$$= \lim_{x\to+\infty} x^2\left(7 - \frac{1}{x} + \frac{1}{x^2}\right) =$$

che per "grandi" x è il prodotto di un numero "grande" per un numero prossimo a 7

$$= +\infty.$$

Esempio 2.

$$\lim_{x\to+\infty}(x - \sqrt{x}) = \lim_{x\to+\infty}\sqrt{x}(\sqrt{x} - 1) =$$

prodotto di 2 numeri "grandi"

$$= +\infty$$

oppure $= x\,(1 - \frac{1}{\sqrt{x}})$, prodotto di un numero "grande" per un numero prossimo a 1: in ogni caso il limite è $+\infty$.

Tutto ciò costituisce il *calcolo dei limiti*, che comunque può raggiungere più alti livelli di sottigliezza:

$$\lim_{x\to+\infty}\frac{3x^4 - 5x^3 + 6 - \mathrm{e}\sqrt{x}}{3x^3 - \frac{1}{x} - 2x^4} = \lim_{x\to+\infty}\frac{3 - \frac{5}{x} + \frac{6}{x^4} - \frac{\mathrm{e}}{x^3\sqrt{x}}}{\frac{3}{x} - \frac{1}{x^5} - 2} = -\frac{3}{2}\,.$$

$$\lim_{x\to-\infty}\frac{\sin x}{x} = 0 \qquad \lim_{x\to-\infty}\frac{3x^4 - 5x^3 + 6 - \mathrm{e}\sqrt{x}}{3x^3 - \frac{1}{x} - 2x^4} = \ldots = -\frac{3}{2}$$

Si dimostrano

$$\lim_{x\to+\infty} \frac{e^x}{x} = +\infty$$

$$\lim_{x\to+\infty} \frac{10^x}{x} = +\infty$$

e in effetti $\lim_{x\to+\infty} \frac{b^x}{x} = +\infty \quad b > 1$

$$\lim_{x\to+\infty} \frac{\ln x}{x} = 0$$

$$\lim_{x\to+\infty} \frac{\lg x}{x} = 0$$

e in effetti $\lim_{x\to+\infty} \frac{\log_b x}{x} = 0 \quad 0 < b \neq 1.$

BOZZA - DRAFT

26 Derivate – I parte

26.1 Introduzione alle derivate

Molto operativamente, data una funzione numerica $f(x)$, se non è troppo capricciosa da essa si calcola una funzione DERIVATA $f'(x)$, che per ogni x ha il valore della pendenza m della tangente $y = m\,x + q$ al grafico nel punto $(x, f(x))$.

Anticipiamo un esempio chiarificatore.
La derivata di x^2 è $2\,x$; scriveremo $Dx^2 = 2\,x$; anche $(x^2)' = 2x$.
In 1 la $2\,x$ vale 2 che è la pendenza del grafico di x^2 in 1.
In -1 la $2\,x$ vale -2 che è la pendenza del grafico di x^2 in -1.
In 0 la $2\,x$ vale 0 che è la pendenza di x^2 in 0.
Per $x > 0$ la $2\,x$ è > 0 dal che là la x^2 cresce (in ogni punto).
Per $x < 0$ la $2\,x$ è < 0 dal che là la x^2 decresce (in ogni punto).

BOZZA - DRAFT

26.2 Notazioni per la derivata prima

Ecco le principali notazioni usate per la derivata prima di $f(x)$:

$$f'(x) \qquad D\,f(x) \qquad \frac{d}{d\,x}\,f(x) \qquad \dot{f}(x)$$

L'ultima è usata per derivata rispetto al tempo.

26.3 Derivate delle funzioni elementari reali

Tutte queste formule[129] valgono nell'intersezione dei domini della funzione derivanda e della funzione derivata. Per esempio $D \ln x = \frac{1}{x}$ vale per $x > 0$ sebbene $\frac{1}{x}$ esista anche per $x < 0$.

$\forall c$	$D\,c$	$=$	0	
	$D\,x$	$=$	1	
	$D\,x^n$	$=$	$n\,x^{n-1}$, n intero,	in particolare $Dx^2 = 2x$ e $D\frac{1}{x} = -\frac{1}{x^2}$
	$D\,x^\alpha$	$=$	$\alpha\,x^{\alpha-1}$ con $x > 0$	e α reale non intero,
			in particolare	$D\sqrt{x} = \frac{1}{2\sqrt{x}}$
	$D \sin x$	$=$	$\cos x$	
	$D \cos x$	$=$	$-\sin x$	
	$D \tan x$	$=$	$\frac{1}{\cos^2 x}$	$= 1 + \tan^2 x$
	$D \arctan x$	$=$	$\frac{1}{1+x^2}$	
	$D\,\mathrm{e}^x$	$=$	e^x	
	$D\,10^x$	$=$	$\frac{10^x}{\lg \mathrm{e}}$	
	$D \ln x$	$=$	$\frac{1}{x}$	e vale anche $D \ln \lvert x\rvert = \frac{1}{x}$
	$D \lg x$	$=$	$\frac{\lg \mathrm{e}}{x}$	e vale anche $D \lg \lvert x\rvert = \frac{\lg \mathrm{e}}{x}$
	$D\,\Phi(x)$	$=$	$\phi(x)$	ma $\Phi(x)$ non è elementare.

(81)

BOZZA - DRAFT

[129] Per lo studente interessato, ad un livello superiore si considerano anche:

$\forall a > 0$	$D\,a^x$	$=$	$a^x \ln a$	
$\forall\, 0 < b \neq 1$	$D \log_b x$	$=$	$\frac{1}{x} \cdot \frac{1}{\ln b}$	
	$D \operatorname{cotan} x$	$=$	$-\frac{1}{\sin^2 x}$	$= -1 - \operatorname{cotan}^2 x [= -\operatorname{cosec}^2 x]$
	$D \arcsin x$	$=$	$\frac{1}{\sqrt{1-x^2}}$	
	$D \arccos x$	$=$	$-\frac{1}{\sqrt{1-x^2}}$	
	$D \sinh x$	$=$	$\cosh x$	
	$D \cosh x$	$=$	$\sinh x$	
	$D \tanh x$	$=$	$\frac{1}{\cosh^2 x}$	$[= 1 - \tanh^2 x = \operatorname{sech}^2 x]$
	$D \operatorname{arsinh} x$	$=$	$\frac{1}{\sqrt{x^2+1}}$	
	$D \operatorname{arcosh} x$	$=$	$\frac{1}{\sqrt{x^2-1}}$	
	$D \operatorname{artanh} x$	$=$	$\frac{1}{1-x^2}$	$[= D \operatorname{arcoth} x]$
	$D \lvert x\rvert$	$=$	$\frac{x}{\lvert x\rvert} = \frac{\lvert x\rvert}{x}$	$[= \operatorname{sgn} x,\ \forall\, x \neq 0]$

26.4 Derivata in un punto e funzione derivata

In questa trattazione elementare considereremo solo le derivate delle funzioni reali di variabile reale.

Essenzialmente, la derivata $f'(x)$ di una funzione f in un punto x, è un numero oppure $+\infty$ o $-\infty$ che se esiste rappresenta la pendenza del grafico in $(x, f(x))$.E se non esiste vuol dire che in quel punto il grafico non ammette retta tangente, per qualche sorta di sua non liscezza. La derivata positiva o $+\infty$ corrisponde (con precisazioni che faremo) alla crescenza di f nel punto x e quella negativa o $-\infty$ alla decrescenza.

Al variare di x nel dominio di f si ottiene una funzione $f'(x)$ ovvero $Df(x)$ che si chiama [funzione] derivata.

BOZZA - DRAFT

La derivata $f'(x)$ in un punto x uguaglia contemporaneamente,
— se finita:

- il coefficiente angolare m della retta tangente in $(x, f(x))$
- la tangente (goniometrica) $\tan\alpha$ dell'angolo della (retta) tangente con l'asse x:

$$\tan\alpha = f'(x) = m$$

— e sia se finita che infinita:

- questo limite, che usualmente viene preso per definirla:

$$f'(x) := \lim_{z\to x} \frac{f(z)-f(x)}{z-x} \quad ovvero := \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$$

per ogni x per cui i limiti (equivalenti) esistono, anche se infiniti. (Questa è la definizione della derivata intesa come valore).

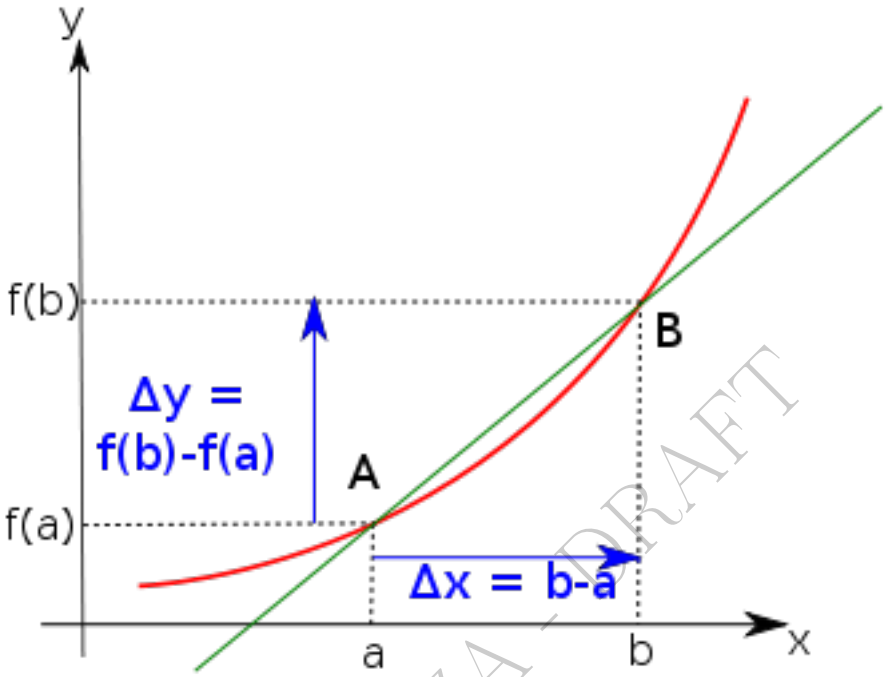


Figure 35: By Xzapro4 (2009), in Wikimedia Commons

BOZZA - DRAFT

La *funzione* derivata ha la stessa definizione ma solo per ogni x per cui i limiti esistono *finiti*.

L'argomento del secondo limite qua sopra si chiama *rapporto incrementale* di f in x.

(La maggioranza dei testi, trattando più separatamente le derivate come numero e come funzione, per il primo caso scrive $f'(x_0) := \lim_{x\to x_0} \frac{f(x)-f(x_0)}{x-x_0}$ ovvero $:= \lim_{h\to 0} \frac{f(x_0+h)-f(x_0)}{h}$, derivata in x_0).

Si dimostra (teorema) che se in $(x, f(x))$ esiste la tangente al grafico essa forma con l'asse x un angolo orientato α tale che $\tan\alpha = f'(x)$.
Da ciò segue, lasciando per adesso all'intuizione il significato di funzione crescente o decrescente in un punto, che

$f'(x_0) > 0 \Rightarrow f$ **crescente** in x_0

$f'(x_0) < 0 \Rightarrow f$ decrescente in x_0

(si tratta della crescenza e decrescenza **puntuali**, non quelle globali del paragrafo 6.6) e lasciando per adesso all'intuizione il significato di funzione "liscia" e di punto di massimo relativo e di minimo relativo

x_0 punto di massimo relativo di funzione liscia $\Rightarrow\ f'(x_0) = 0$
x_0 punto di minimo relativo di funzione liscia $\Rightarrow\ f'(x_0) = 0$

La derivata di $f'(x)$ è la *derivata seconda* $f''(x)$ o $f^{(2)}(x)$ eccetera.

Lasciando all'intuizione il significato di *concavità di una funzione*

fz liscia con $f''(x) > 0$ su un intervallo $\Rightarrow$ concavità verso l'alto
fz liscia con $f''(x) < 0$ su un intervallo $\Rightarrow$ concavità verso il basso

Detto flesso un punto in cui si inverte la concavità, com'è l'origine per x^3,

x_0 punto di flesso di funzione liscia $\Rightarrow\ f''(x_0) = 0$

infatti la derivata seconda di x^3 si vedrà esser $6\,x$ che in 0 è 0.

Esempi.
Si trova facilmente che la derivata di $|x|$ è $\frac{x}{|x|}$ corrispondentemente all'inesistenza della tangente al grafico in 0 (ovvero, in $(0,0)$).

La funzione $\sqrt[3]{x}$ ha derivata $+\infty$ in 0, e corrispondentemente in $(0,0)$ il grafico ha tangente verticale.

BOZZA - DRAFT

Nota. Dopo $Dx^2 = 2x$, che lo studente è invitato a calcolare col limite del rapporto incrementale, non calcoleremo più le derivate col limite del rapporto incrementale: ciò è già stato[130] fatto, secoli fa: oggi usiamo tabelle e regole di derivazione.

26.5 Epidemie, curve sigmoidi, a campana, derivata

Le curve sigmoidi e le curve a campana , che così bene modellizzano in un'epidemia il numero cumulativo di morti e rispettivamente il numero di morti giornaliero, sono strettamente legate fra loro dalla derivata. Detto ultrasemplificatamente,
la derivata di una curva sigmoide è una curva a campana
(affermazione che per essere matematicamente rigorosa richiederebbe varie ipotesi aggiuntive).

DRAFT

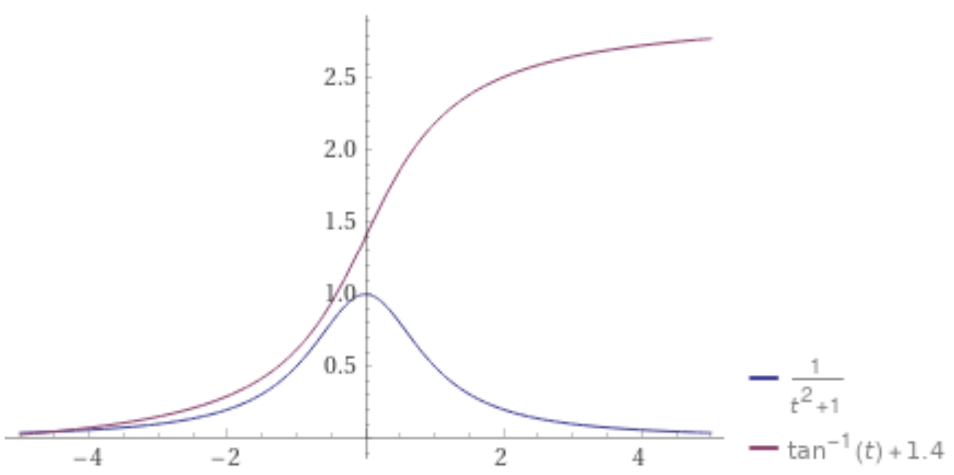


Figure 36: Una curva sigmoide e la sua derivata a campana (WolframAlpha)

In una curva sigmoide sufficientemente regolare riconosciamo:

crescita sempre più rapida, con la *concavità* verso l'alto	$f'(x) > 0$	$f''(x) > 0$
flesso	$f'(x_0) > 0$	$f''(x_0) = 0$
crescita sempre più lenta, con la *concavità* verso il basso	$f'(x) > 0$	$f''(x) < 0$

In una curva a campana sufficientemente regolare riconosciamo:

[130] per le funzioni base di utilità pratica in Farmacia, non per tutte le funzioni, che sono infinite.

crescita sempre più rapida, con la *concavità* verso l'alto	$f'(x) > 0$	$f''(x) > 0$
flesso	$f'(x_1) > 0$	$f''(x_1) = 0$
crescita sempre più lenta, con la *concavità* verso il basso	$f'(x) > 0$	$f''(x) < 0$
la fz raggiunge un max (ordinata) in un punto di max (ascissa)	$f'(x) = 0$	$f''(x) < 0$
di decrescita sempre più rapida, con *concavità* verso il basso	$f'(x) < 0$	$f''(x) < 0$
flesso	$f'(x_2) < 0$	$f''(x_2) = 0$
decrescita sempre più lenta, con *concavità* verso l'alto	$f'(x) < 0$	$f''(x) > 0$

BOZZA - DRAFT

Complementi

Esercizio$_{\mu 2023}$ ≈ Calcolare un'approssimazione della derivata di $\lfloor x \rfloor^x$ (che è definita con la parte intera) in 2.5 con la formula approssimata

$$f'(x) \approx \frac{f(x+h) - f(x)}{h} \qquad h \ll \text{ (cioè molto piccolo)}$$

(approssimazione della formula esatta, col limite, $f'(x) \coloneqq \lim_{h \to 0} \frac{f(x+h)-f(x)}{h}$) usando il valore $h \coloneqq 0.25 = \frac{1}{4}$ (che non può essere considerato "molto piccolo" in assoluto ma in questo caso garantisce un'approssimazione entro il 10%).
Nell sovolgimento si potrebbe incontrare il numero $2^{2.5}$ che ovviamente è $2^2 \cdot 2^{0.5} = 4\sqrt{2}$ (oppure equivalentemente $2^{5/2} = \sqrt{2^5} = \sqrt{32}$).

SVOLGIMENTO
Viene usato lo standard del punto decimale.

Con i valori dati $x = 2.5$ e $h = 0.25$

$$f'(2.5) \approx \frac{f(2.5 + 0.25) - f(2.5)}{0.25} =$$

$$= \frac{f(2.75) - f(2.5)}{0.25} =$$

con la funzione data $f(x) \coloneqq \lfloor x \rfloor^x$

$$= \frac{\lfloor 2.75 \rfloor^{2.75} - \lfloor 2.5 \rfloor^{2.5}}{0.25} =$$

e con la parte intera (che ai numeri positivi "toglie i decimali")

$$= \frac{2^{2.75} - 2^{2.5}}{0.25} =$$

per le proprietà delle potenze

$$= \frac{2^{2.5}\,(2^{0.25} - 1)}{0.25} =$$

ricordando, come è ovvio e anche ci viene detto nel quesito, che $0.25 = \frac{1}{4}$,

$$= \frac{2^{2.5}\,(2^{\frac{1}{4}} - 1)}{0.25} =$$

BOZZA - DRAFT

che per le proprietà delle potenze e delle radici, come anche ci viene detto nel quesito, è

$$= \frac{4\sqrt{2}\,(2^{\frac{1}{4}} - 1)}{0.25} =$$

per le proprietà delle radici riconosciamo una radice quarta

$$= \frac{4\sqrt{2}\,(\sqrt[4]{2} - 1)}{0.25} =$$

e con la nota formula di riduzione della radice quarta $\sqrt[4]{a} = \sqrt{\sqrt{a}}$

$$= \frac{4\sqrt{2}\,(\sqrt{\sqrt{2}} - 1)}{0.25} \approx$$

con una calcolatrice (dotata di radice quadrata)

$$\boxed{\approx 4.281}$$

$$\boxed{\approx 4.28}$$

$$\boxed{\approx 4.3}$$

BOZZA - DRAFT

(Il valore esatto è $2^{2.5} \ln 2 = 4\sqrt{2} \ln 2 \approx 3.921$).

27 Derivate – II parte

27.1 Teoremi algebrici sulle derivate

In tutta questa Lezione, la somma è somma di funzioni, e così il prodotto e il quoziente: per esempio $(f+g)(x) := f(x)+g(x)$.
La derivata della somma [di 2 funzioni] è la somma delle derivate [di quelle 2 funzioni]:

$$(f+g)' = f' + g'$$

che potremmo anche scrivere $D(f+g) = Df + Dg$ ma continueremo con la notazione dell'apice per la derivata prima.
La derivata del prodotto non è il prodotto delle derivate bensì

$(f\cdot g)' = f'\cdot g + f\cdot g'$. Da cui subito $(cf)' = cf'$, c costante .

Derivata della [funzione] reciproca:

$$\left(\frac{1}{f}\right)' = -\frac{f'}{f^2}$$

e da questa e dalla precedente si ricava subito la derivata del quoziente:

$$\left(\frac{f}{g}\right)' = \frac{f'\cdot g - f\cdot g'}{g^2}\ .$$

Derivata della funzione composta:[131]

$$(f(g(x)))' = f'(g(x))\cdot g'(x)$$

Esempi. Con la prima, terza e quinta formula, ricordando la derivata dell'arcotangente, si troverà $D\left(\operatorname{arctg} x + \operatorname{arctg}\frac{1}{x}\right) = 0$.
Dalla quinta formula $Df^\alpha = \alpha\cdot f^{\alpha-1}\cdot f'$ per ogni $\alpha\in\mathbb{R}$.

[131] Per lo studente interessato, quella formula consente la derivazione di f^g derivando l'equivalente $e^{g\cdot\ln f}$:

$$(f^g)' = f^g\cdot\left(g'\cdot\ln f + \frac{g\cdot f'}{f}\right)$$

ricordando la soprastante formula di derivazione del prodotto, e la derivata della funzione esponenziale e della funzione logaritmo.

BOZZA - DRAFT

Supponiamo di avere 2 funzioni V(x) e A(x), per esempio il volume e l'area laterale della sfera in funzione del raggio x.
Supponiamo di considerare il loro prodotto VA e di cercare la sua derivata

$$(VA)^{I}=$$

che per la formula di derivazione del prodotto è

$$=V^{I}A+VA^{I}$$

che vagamente ricorda la scritta VIA VAI.

BOZZA

Esempi di calcolo di derivate

$$D\sqrt{1+x^2} =$$

$$= D\left(1+x^2\right)^{\frac{1}{2}} =$$

$$= \frac{1}{2}(1+x^2)^{\frac{1}{2}-1} \cdot D(1+x^2) =$$

$$= \frac{1}{2}(1+x^2)^{-\frac{1}{2}} \cdot 2\,x =$$

$$= \frac{x}{\sqrt{1+x^2}}$$

$$D\left(-\frac{1}{x^2}\right) =$$

$$= -(1) \cdot D\frac{1}{x^2} =$$

$$= -D\,x^{-2} =$$

$$= 2\,x^{-3} =$$

$$= \frac{2}{x^3}$$

BOZZA - DRAFT

27.2 Le funzioni elementari in Farmacia, e i modelli

L'uso delle funzioni elementari nella ricerca farmaceutica e in generale biomedica è *enorme*, per **modellizzare** i fenomeni.

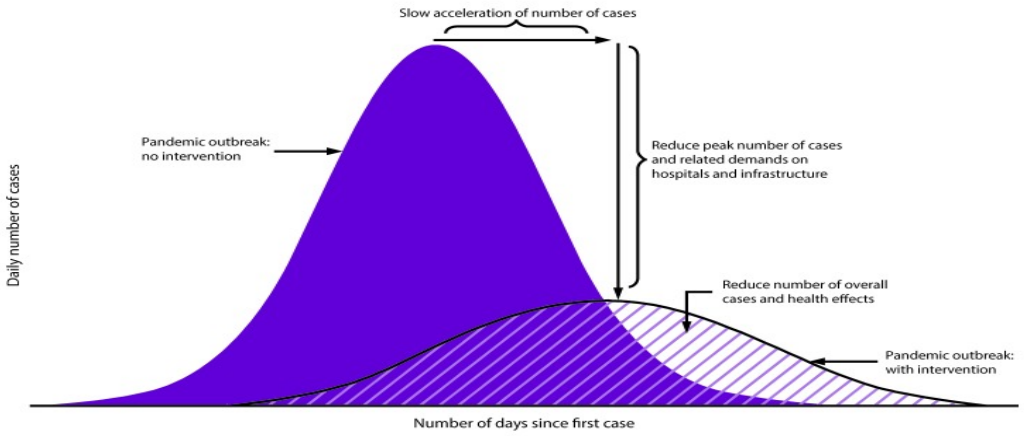
Avere in mano dei numeri, risultato di un esperimento, è *qualcosa*. Avere una funzione che li modellizza – anche se non rende conto *esattamente* di tutti i casi riscontrati – è *immensamente di più*. Per esempio perchè nel modello ci saranno dei parametri, come la concentrazione di qualche sostanza in un liquido di coltura, per fare un esempio, che si potrà ipotizzare di modificare, in un successivo esperimento, e così si procede.
Oppure, consideriamo un altro caso: in una regione hanno raccolto circa 20mila dati (peso, età in mesi) di bambini. Possiamo avere

un file Excel con 2 colonne di dati, che potrà servire per future ricerche, ma sarà inutilizzabile in farmacia per valutare quanto è normale il peso di un bambino di quella regione che entri in quella farmacia. Un enorme tabulato cartaceo sarebbe ancora più inutilizzabile. Invece è più utile una formula, col suo dominio di validità, come

$$height(\text{cm}) = 78 \times age(\text{months}) + 76(\text{cm}) \quad (0 \leq age(\text{months}) \leq 72)$$

che approssimi tutti quei valori, seppure imperfettamente.

Ecco per esempio nella figura una modellizzazione di epidemie:



From: Qualls N, Levitt A, Kanade N, Wright-Jegede N, Dopson S, Biggerstaff M, Reed C, Uzicanin A; CDC Community Mitigation Guidelines Work Group. Community Mitigation Guidelines to Prevent Pandemic Influenza - United States, 2017. MMWR Recomm Rep. 2017 Apr 21;66(1):1-34. doi: 10.15585/mmwr.rr6601a1. PMID: 28426646; PMCID: PMC5837128.

Come abbiamo già detto, nel suo progredire nel tempo e approfondirsi, in linea molto gene-
rale possiamo dire che ogni Scienza Applicata attraversa successivamente queste fasi:

qualitativa ("bollendo *un po'* otterrai *un po' di* precipitato")
numeri
operazioni (numeriche)
funzioni (numeriche)← **qua siamo**

analisi statistica dei dati (numerici)

È allora fondamentale conoscere il comportamento – verrebbe quasi da dire la *psicologia* se non l'*anima* – delle varie funzioni elementari, il loro *modo* di crescere e decrescere e tendere all'infinito eccetera... Limiti e derivate ne sanno molto, dell'anima delle funzioni.

Rivediamo una questione di importanza fondamentale.

Praticamente dietro quasi ogni fenomeno della realtà sensibile c'è in qualche modo una funzione più o meno a campana. Imparare a riconoscere queste configurazioni a campana aumenta enormemente la comprensione della realtà.

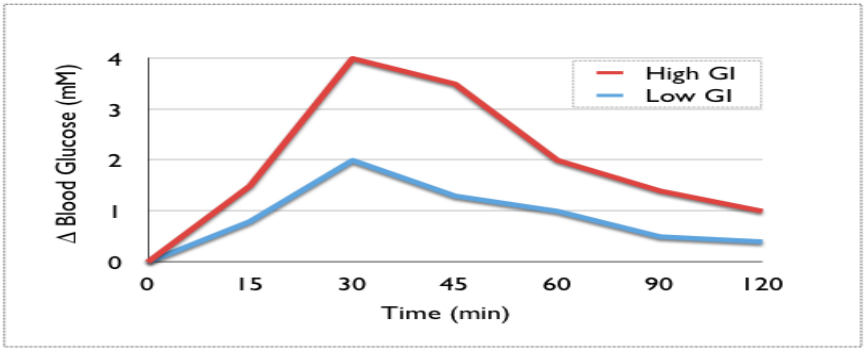
BOZZA - DRAFT

La campana può avere 2 significati principali: una densità, coi suoi casi estremi rari e quelli medi più comuni, oppure, se in ascissa abbiamo il tempo, talvolta rappresenta (quantitativamente) la classica "parabola della vita", valida anche per una comunità microbica, per la potenza dell'Impero Romano, per la concentrazione di un farmaco immesso nel sangue, o quant'altro: sorgere, ascendere, declinare, finire.

Qua stiamo parlando di curve "più o meno a campana" mentre con "curva a campana" *di solito* si intende proprio la *campana gaussiana*, grafico della *densità normale standard* $\frac{1}{\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{t^2}{2}}$.

Curve più o meno a campana del primo tipo, cioè densità, (della Microbiologia) sono in questo link->.

Curva più o meno a campana del secondo tipo, cioè evoluzioni di una quantità nel tempo, è in questa figura



Un'altra, (della Microbiologia) è in questo link-> e un altro (della Farmacocinetica) in questo link-> e un altro a pagina 44 di questo link->.

BOZZA

28 Applicazioni del Calcolo Differenziale - I

28.1 Cos'è molto o poco? Migliora o peggiora?

Il significato profondo delle derivate è che di una grandezza, sia essa $f(t)$, che varia (con continuità) nel tempo t, conta sia il valore, $f(t)$ appunto, sia il fatto che stia aumentando o piuttosto diminuendo, cioè in pratica se la derivata $f'(x)$ sia positiva o negativa.

L'uomo di un paese in via di sviluppo che prima viveva con 100 dollari all'anno e poi 110 e poi 120 e poi 130, magari è pure soddisfatto, nonostante la sua misera situazione economica. E su un piano più oggettivo, certamente se qualcosa del genere avviene non solo per lui ma per milioni di concittadini, qualcosa di buono può essere detto dei governanti di quella regione, le cui politiche di sviluppo stanno avendo un buon successo.

BOZZA_DRAFT

Una situazione di BMI 40.5 è gravissima considerata in sè ma un buon successo – eventualmente in parte anche farmacologico – se alla misurazione precedente era 41. (Per adesso è ancora un'*obesità di III classe*). (A livello personale potrà rattristarsi per la situazione o rallegrarsi per il miglioramento).

Se nella regione col tasso di positività al covid-19 più alto d'Italia quel valore diminuisce di giorno in giorno, qualcosa sta andando bene e (media permettendo) la gente magari ne sarà pure contenta – seppure la situazione vada ancora relativamente male (in confronto alle altre regioni).
Reciprocamente, nella regione col tasso più basso ma in aumento, la gente potrà venir spaventata dai media – per esempio affinchè mantenga il distanziamento sociale – insistendo sul peggioramento.
In questa riproduzione parziale di una tabella della prestigiosa Fondazione[132]

[132] "si tratta di una fondazione nata nel 1996 con l'obiettivo di diffondere in Italia la medicina basata su prove di efficacia, attraverso iniziative di formazione, editoria e ricerca." (https:

Gimbe, l'Emilia Romagna ha una situazione piuttosto negativa rispetto ad altri relativamente al parametro considerato nella prima colonna, ma è segnata in verde perchè, come spiegato in nota, quel parametro è in miglioramento.

Tabella 1. Indicatori regionali: settimana 27 ottobre-2 novembre 2021

Regione	Casi attualmente positivi per 100.000 abitanti	Variazione % nuovi casi	Posti letto in area medica occupati da pazienti COVID-19	Posti letto in te intensiva occup pazienti COVID
Abruzzo	149	21,3%	5%	5%
Basilicata	148	43,3%	7%	0%
Calabria	164	2,0%	10%	3%
Campania	160	34,9%	8%	4%
Emilia Romagna	170	23,7%	4%	3%
Friuli Venezia Giulia	187	70,0%	6%	10%
Lazio	169	22,4%	7%	6%
Liguria	82	45,8%	5%	5%
Lombardia	98	16,0%	5%	3%

Figure 37: Da Gimbe, https://coronavirus.gimbe.org/emergenza-coronavirus-italia/monitoraggio-settimanale.it-IT.html (screenshot 9 novembre 2021).

BOZZA - DRAFT

Bisogna imparare a vedere le grandezze che variano nel tempo in modo dinamico: sia valore che trend.

28.2 De/crescenza, min/max, concavità, flessi

Definizioni. Si dice che f *è crescente in* x_0 se in un intervall(in)o a sinistra di x_0 vale meno che in x_0 e in un intervall(in)o a destra di x_0 vale più che in x_0. (In simboli è alquanto complicato). Con ovvi mutamenti si definisce la funzione decrescente in un punto. Si dice che f è *crescente in un insieme* E se $x_1 < x_2 \Rightarrow f(x_1) < f(x_2)$. Se E è un intervallo ciò equivale (si dimostra) alla crescenza in ogni punto di E. Simile equivalenza con la decrescenza, $x_1 < x_2 \Rightarrow f(x_1) > f(x_2)$, sugli intervalli.

Il punto x_0 si dice di *punto minimo relativo* per f se essa in un intervall(in)o a sinistra di x_0 e in un intervall(in)o a destra di x_0

//it.wikipedia.org/wiki/Medicina_basata_sulle_evidenze)

vale *più* che in x_0. Se *meno*, si ha un *punto di massimo relativo*.
Se $\forall\, x \in \mathrm{dom} f$ è $f(x) \geq f(x_0)$, $f(x_0)$ si dice *minimo assoluto* di f, e *massimo assoluto* se $\forall\, x \in \mathrm{dom} f$ è $f(x) \leq f(x_0)$.
Attenzione a distinguere x_0 punto di massimo relativo da $f(x_0)$ massimo relativo. E similmente coi minimi.
Per esempio per il coseno 0 è punto di massimo relativo (e anche assoluto) e 1 è massimo relativo (e anche assoluto).

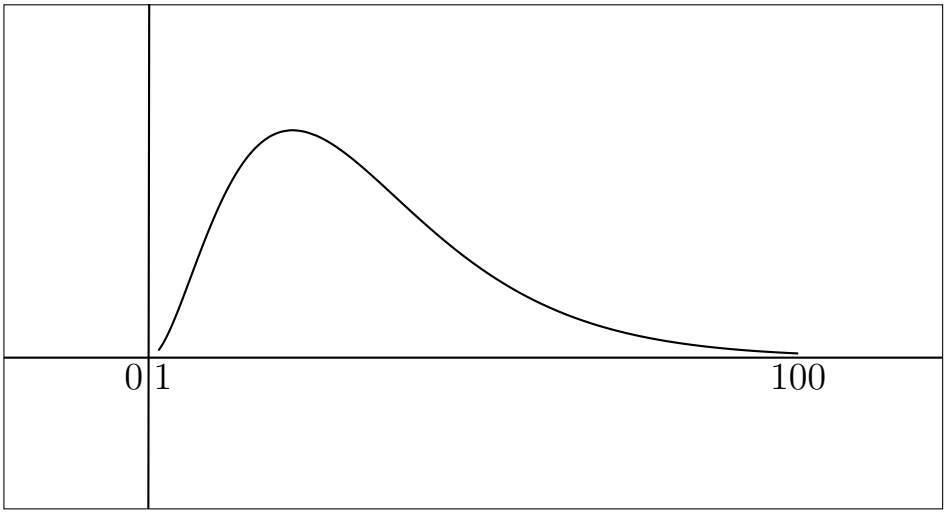
Il punto x_0 si dice di *punto di flesso* per f se essa in un intervall(in)o a sinistra di x_0 *volge la concavità verso l'alto* e in un intervall(in)o a destra verso il basso, oppure viceversa; $(x_0, f(x_0)))$ si dice *flesso*.
Non diamo la complicata definizione analitica di *volgere la concavità*, concetto comunque ovvio se riferito al grafico.

BOZZA - DRAFT

ESERCIZIO$_{\mu 2021}$ $*$ Il numero di morti di un'epidemia viene modellizzato approssimatamente da

$$d(t) := 20\, t^2\, \mathrm{e}^{-\frac{t}{10}} \qquad 1 \leq t \leq 100$$

essendo t il tempo, in giorni, da 1 a 100. In quale giorno si ha il massimo numero di morti? (Più realisticamente, *intorno* a quale giorno; ma usiamo proprio il modello semplificato assegnato). (Ovviamente sono usati i nomi d come *deaths* e t come *time*; ma è irrilevante).
Si risolva con le derivate, sebbene si potrebbe risolvere calcolando i 100 valori di $d(t)$ per t da 1 a 100. Sarà utile ricordare che $D\,\mathrm{e}^{ct} = c\,\mathrm{e}^{ct}$ (che comunque si troverebbe subito con la regola di derivazione della funzione composta).



SVOLGIMENTO
Ricordando la derivazione del prodotto

$$(f \cdot g)' = f' \cdot g + f \cdot g'$$

e ricordando la derivazione di t^2 che dà $2t$, deriviamo la funzione $d(t)$ che è il prodotto $(20\,t^2) \cdot \left(\mathrm{e}^{-\frac{t}{10}}\right)$ ottenendo la disequazione $d'(t) > 0$

$$d'(t) = (40\,t) \cdot \left(\mathrm{e}^{-\frac{t}{10}}\right) + (20\,t^2) \cdot \left(-\frac{1}{10}\,\mathrm{e}^{-\frac{t}{10}}\right) > 0$$

$$\Big/ : t > 0 \qquad \text{(dividiamo per } t > 0 \text{ nel dominio)}$$

$$40\,e^{-\frac{t}{10}} - 2\,\mathrm{e}^{-\frac{t}{10}} > 0 \qquad \Big/ : \mathrm{e}^{-\frac{t}{10}} > 0$$

$$40 - 2\,t > 0$$

$$40 > 2\,t$$

$$t < 20$$

allora il numero di morti $d(t)$ in $[1, 100]$ cresce per $t < 20$ e decresce per $t > 20$ e allora raggiunge il massimo

al giorno 20

BOZZA - DRAFT

Nota. Il modello darebbe circa 1083 morti in quel giorno, calcolando $d(20)$, più precisamente 1082.68: ovviamente la funzione modellizza approssimatamente il numero di morti, che in generale sarà da aspettarsi con ampie fluttuazioni intorno al valore teorico.

28.3 Retta tangente a un grafico in un punto

Teorema 1. La **tangente** al grafico di $f(x)$ in $(x_0, f(x_0))$ è

$$y = f(x_0) + f'(x_0)(x - x_0)$$

e allora si ha l'*approssimazione lineare*

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) \qquad \forall\, x \approx x_0$$

ovvero con $x = x_0 + h$ per h piccolissimo in valore assoluto

$$f(x_0 + h) \approx f(x_0) + f'(x_0) \cdot h \qquad |h| \ll .$$

Per esempio per $x \approx 0$ è $\sin x \approx x$, $\ln(1 + x) \approx x$, $\sqrt{1 + x} \approx 1 + \frac{x}{2}$.

Teorema 2. $f'(x_0) > 0 \Rightarrow f$ **crescente** in x_0

$$f'(x_0) < 0 \Rightarrow f \text{ decrescente in } x_0 .$$

(Se $f'(x_0) = 0$ allora f in x_0 può essere crescente o decrescente o nè crescente nè decrescente: si considerino x^3, $-x^3$, x^2).

Teorema 3. Se f è crescente prima di x_0 e decrescente dopo x_0 allora x_0 è un punto di massimo relativo. Similmente scambiando *crescente* e *decrescente* si ha un **punto di minimo relativo**.

Teorema 4. Se $f''(x) > 0$ in un intervallo, in esso $f(x)$ *volge la* ***concavità*** *verso l'alto*, e verso il basso se < 0.

BOZZA - DRAFT

28.4 Approssimazioni delle potenze presso l'unità

L'approssimazione lineare prima detta

$$f(x_0 + h) \approx f(x_0) + f'(x_0) \cdot h \qquad |h| \ll$$

applicata in $x_0 = 1$

$$f(1 + h) \approx f(1) + f'(1) \cdot h \qquad |h| \ll$$

per le potenze della x

$$x^\alpha \text{ che in 1 vale 1} \quad \text{Derivata: } \alpha\, x^{\alpha-1} \quad \text{che in 1 vale } \alpha$$

dà

$$(1 + h)^\alpha \approx 1 + \alpha \cdot h \qquad |h| \ll$$

ottenendosi queste approssimazioni dei quadrati e dei cubi, e avanti

x_0 quadrato cubo

1.01 1.02 1.03
1.001 1.002 1.003
1.0001 1.0002 1.0003
1 parte su n 2 parti su n 3 parti su n.
Più in generale una funzione

$$c\,x^{\alpha}$$

in cui la x si assoggetti a un aumento percentuale di *pochi* punti percentuali, diciamo m punti percentuali, per esempio 2,

aumenta circa di $\alpha\,m$ punti percentuali

per α piccolo in valore assoluto. Similmente con le diminuzioni. Per esempio $\frac{4}{3}\,\pi r^3$ aumenta (o rispettivamente diminuisce) circa del 12% all'aumentare (o rispettivamente al diminuire) del 4% della r.
Tutto questo vale anche con α negativo, salvo che un aumento, per esempio, del -6%, è in effetti una diminuzione; per esempio $c\,x^{-2}$ diminuisce circa del 6% all'aumentare del 3% della x.

L'approssimazione è tanto migliore quanto più m e $|\alpha|$ sono piccoli. Con $m = 10$ e $\alpha = 3$ dà 30 invece di 33.1.
Tutto similmente coi "per mille", con maggior precisione.

Per piccolo m
aumentando/diminuendo del $m\%$ il lato di una farmacia quadrata, l'area aumenta/diminuisce circa del $2\,m\%$

Per piccolo m
aumentando/diminuendo del $m\%$ il raggio ovvero il diametro di una compressa, il volume ovvero il peso aumenta/diminuisce circa del $3\,m\%$

Per piccolo m
aumentando/diminuendo del $m\%$ l'area di una farmacia quadrata, il lato aumenta/diminuisce circa del $\frac{1}{2}\,m\%$ (radice quadrata, $\alpha = 1/2$)

BOZZA - DRAFT

Per piccolo m
aumentando/diminuendo del $m\%$ il volume di una compressa sferica, il diametro ovvero il raggio aumenta/diminuisce circa del $\frac{1}{3}\,m\%$ (radice cubica, $\alpha = 1/3$)

Fissato α coi valori usuali 2, 3, 1/2 oppure 1/3, e i loro opposti, tutto questo rimane vero coi "per mille", e ogni altra proporzione, purché la variazione sia complessivamente piccola; con 1 parte su 20 funziona bene, e diventano

2 parti su 20 (meglio detto 1 parte su 10) col quadrato,

3 parti su 20 col cubo,

1 parte su 40 con la radice quadrata,

1 parte su 60 con la radice cubica.

Per esempio, aumentando il volume di una sfera di 1 parte su 20, il raggio, che è proporzionale alla radice cubica del volume, aumenta di circa 1 parte su 60.
Invece con 1 parte su 3 l'approssimazione non funziona bene. Diciamo – tanto per dire – da 1 parte su 10 "in poi", che corrisponde al 10%, per potenze in valore assoluto non superiori a 3; e già qua la bontà dell'approssimazione è dubbia: col 10% e $\alpha = 3$ dà 30% invece di 33.1%.

BOZZA - DRAFT

Tutto questo rimane vero, salvo modificazioni linguistiche, anche per forme diverse, purché si conservi la similitudine geometrica, cioè tutte le dimensioni variano allo stesso modo.
Così un flacone di medicinale, all'aumentare del 10% delle sue dimensioni lineari (diametro, altezza) conservando la forma, vedrà aumentare il suo volume circa del 30%. Similmente per il diminuire. Un calcolo più preciso per il 10%, che non è poi tanto piccolo numero, darebbe circa 33%. Con 1% viene abbastanza bene 3%.

E non solo i flaconi: anche capsule "a forma di pillola" (invece che compresse sferiche), e anche cavie e altri animali: un gatto del

5% più lungo di una gatto “di ugual forma" pesa circa il 15% in più. Ma un gatto disegnato richiede il 10% di inchiostro in più. 😸

BOZZA - DRAFT

29 Applicazioni del Calcolo Differenziale - II

29.1 Regola di de l'Hospital

In condizioni che non[133] approfondiamo, certi limiti del tipo

$$\frac{0}{0} \quad \text{oppure} \quad \frac{\infty}{\infty}$$

si possono calcolare sostituendo
il numeratore con la sua derivata
e (anche!)
il denominatore con la sua derivata.

Ripetiamo: quoziente delle derivate. (Non derivata del quoziente).

BOZZA - DRAFT

[133] Per lo studente interessato, qualche dettagio.
Detti u_0 e l due numeri o $+\infty$ o $-\infty$, se per $x \to u_0$

$$\begin{array}{|ll} & f(x) \to 0 \\ \wedge & g(x) \to 0 \\ \wedge & \frac{f'(x)}{g'(x)} \to l \end{array} \quad \vee \quad \begin{array}{|ll} & f(x) \to \pm\infty \\ \wedge & g(x) \to \pm\infty \\ \wedge & \frac{f'(x)}{g'(x)} \to l \end{array}$$

si dimostra (teorema detto Regola di de l'Hospital) che

$$\lim_{x \to u_0} \frac{f(x)}{g(x)} = l$$

se f e g sono "sufficientemente regolari" (come sono in genere le funzioni che ricorrono nelle Scienze Applicate, e negli esercizi elementari, anche di questa trattazione).
Si noti che l'eventuale inesistenza del limite del rapporto delle derivate non esclude l'esistenza del limite del rapporto iniziale.
Esempi.

$$\lim_{x \to 0} \frac{\sin x}{x} \left(= \frac{0}{0} \right) <=^{H} \lim_{x \to 0} \frac{\cos x}{1} = \frac{\cos 0}{1} = 1$$

che si chiama **secondo limite fondamentale** (e per esso non vale *come dimostrazione* il calcolo soprastante, esiste una dimostrazione specifica; e proprio da quel limite si dimostra che $D \sin x = \cos x$, che quassù viene utilizzato).

$$\lim_{x \to +\infty} \frac{x}{\ln x} \left(= \frac{\infty}{\infty} \right) <=^{H} \lim_{x \to +\infty} \frac{1}{1/x} = \lim_{x \to +\infty} x = +\infty \,.$$

Da questi 2 limiti si ha questa sequenza di funzioni "sempre più infinite" in $+\infty$, cioè tali che il rapporto di una di esse con una precedente tende a $+\infty$: (lentissima) $\ln x$, x, e^x (velocissima). Il $\lim_{x \to +\infty} \frac{x^2 + 5\mathrm{e}^x + \pi}{7x^3 + 3x^2 + 1}$ si risolverà con 3 applicazioni successive del teorema. Invece $\lim_{x \to +\infty} \frac{2x + \sin x + \cos x}{5x - \sin x + 3\cos x}$ semplificando per x.

Esempio.

$$\lim_{x\to+\infty} \frac{\mathrm{e}^x}{x+2} \left(= \frac{\infty}{\infty} \right) <=^H \lim_{x\to+\infty} \frac{\mathrm{e}^x}{1} = +\infty \,.$$

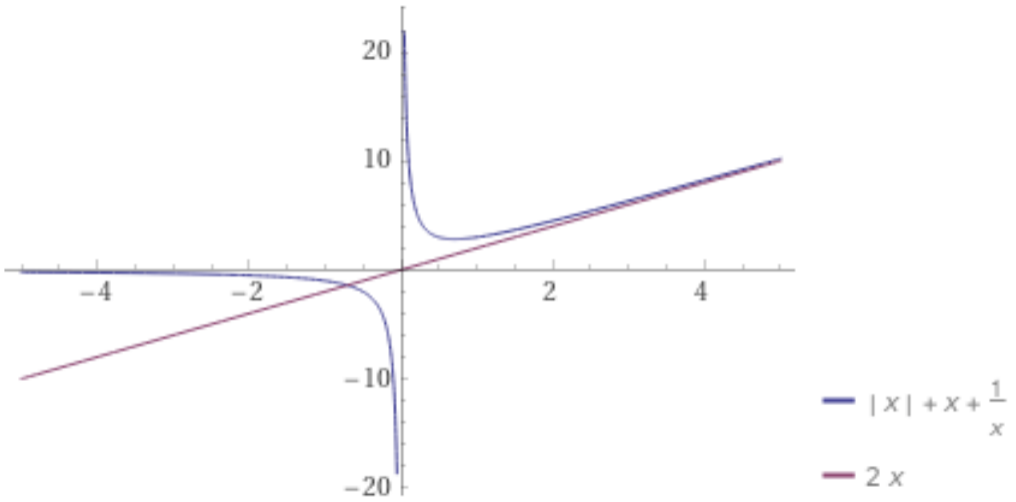
29.2 Asintoti

L'asintoto è una sorta di retta tangente a un grafico in punti infinitamente lontani dall'origine.

Impariamo a riconoscerli per via grafica (con tutte le incertezze del caso). In questo disegno (screenshot da WolframAlpha) del grafico di $|x| + x + \frac{1}{x}$ vediamo

- l'asintoto verticale $x = 0$,
- l'asintoto orizzontale sinistro $y = 0$
- l'asintoto obliquo destro $y = 2\,x$.

DRAFT



per Corso da 40 ore * → Paragrafo 29.3

Definizione 1.[134] Se x_0 è un numero [finito] e

$$\lim_{x\to x_0^-} f(x) = \pm\infty \quad \vee \quad \lim_{x\to x_0^+} f(x) = \pm\infty$$

allora la retta verticale $x = x_0$ si dice ***asintoto verticale*** per f.

[134] Alcuni Autori ritengono inutile, non significativa, questa definizione.

Definizione 2. Se

$$\lim_{x\to-\infty} f(x) = cost$$

allora $y = cost$ si dice **asintoto orizzontale sinistro** per $f(x)$, o per $x \to -\infty$, e similmente si definisce quello **destro**, ovvero per $x \to +\infty$.

La definizione analitica di **asintoto obliquo** è più complessa[135] e non la diamo. Riconosciamo – tentativamente – gli asintoti obliqui dal (disegno del) grafico. Ma con, per esempio, $x + \ln x$ il rischio di sbagliare è notevole (perchè contrariamente all'impressione che in generale si ricava disegnandone un grafico, non ha asintoto obliquo destro). Il riconoscimento "a occhio" degli asintoti obliqui (e anche orizzontali) è problematico.

BOZZA - DRAFT

Esempi.
La funzione $f(x) := \frac{1}{x}$ ha asintoto verticale $x = 0$ e asintoto orizzontale $y = 0$: si trovano coi limiti a $-\infty$, $+\infty$, e 0^+ oppure 0^-.

La funzione $\tan x$ ha gli infiniti asintoti verticali $x = k\frac{\pi}{2}$ per $k \in \mathbb{Z}$.

Si troverà che $y = \frac{\pi}{2}$ è asintoto orizzontale destro per $\arctan x$, e $y = -\frac{\pi}{2}$ è asintoto orizzontale sinistro.

Si troverà che $y = 1$ è asintoto orizzontale destro per

$$\tanh x := \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

[135] Per il lettore interessato: se esistono 2 numeri [finiti] m e q tali che

$$\lim_{x\to+\infty} \frac{f(x)}{x} = m, \qquad \lim_{x\to+\infty} (f(x) - mx) = q$$

allora la retta $y = mx+q$ si dice ***asintoto destro*** (oppure: *per* $x \to +\infty$) per f, in particolare *obliquo* se $m \neq 0$ e *orizzontale* se $m = 0$. Con $x \to -\infty$ si definisce l'eventuale ***asintoto sinistro***.
Si trova che $y = 0$ è asintoto orizzontale sinistro per $\ln(1 + e^x)$ e $y = x$ è asintoto obliquo destro.
Per $f(x) := \sqrt{x}$ si troverebbe $m = 0$ ma q infinito e allora non esiste asintoto destro. (E sinistro non c'è perché $\mathrm{dom} f = [0, +\infty[$).
Con la Regola di de l'Hospital per $\ln x$ si trova $m = 0$ ma q infinito e allora non esiste asintoto destro. (Sinistro escluso dal dominio).
Esercizi. Si trovino i 2 asintoti di questa funzione considerata da Wikipedia, l'enciclopedia libera, alla voce *Asymptote*:

$$\frac{2x^2 + 3x + 1}{x}$$

e con essi e qualche valore si disegni un grafico approssimativo. Similmente per la reciproca, che ha 3 asintoti.

e $y = -1$ è asintoto orizzontale sinistro.

La funzione $f(x) := e^{\frac{1}{x}}$ ha asintoto verticale $x = 0$ che si trova col limite per $x \to 0^+$, e asintoto orizzontale [destro e sinistro] $y = 1$.

29.3 sup e inf

Se una funzione ha un massimo assoluto $\max f$, esso è l'*estremo superiore* $\sup f$. Tuttavia, consideriamo la funzione $\arctan x$. Essa per $x \to +\infty$ tende a $\frac{\pi}{2}$ ma senza mai raggiungerlo: vi si avvicina indefinitamente rimanendo sempre minore. In questo caso $\frac{\pi}{2}$ non è certo max – infatti per essere max ci vorrebbe un x_0 tale che arctan là vale prorio quel valore, il che non succede mai – ma si dice che è estremo superiore:

$$\sup \arctan = \frac{\pi}{2}$$

Similmente, si potrebbe definire l'estremo inferiore $\inf f$; in questo caso $\inf f = -\frac{\pi}{2}$. Ovvio poi è il significato di $\sup f = +\infty$, e di $\inf f = -\infty$. Per esempio

$$\sup \tan = +\infty$$

Le definizioni formali di sup e inf sono complicate e non le daremo.

BOZZA - DRAFT

29.4 La “ricetta" per lo studio di funzione

Fermo restando che una funzione molto “capricciosa" non può essere studiata con metodi elementari, è però vero che un'infinità di funzioni del tipo di quelle che tendono a capitare nelle Scienze Applicate può validamente studiarsi coi metodi visti finora, venendosi in pratica a creare una sorta di “ricetta fissa" per il cosiddetto *studio di funzione*.

per Corso da 40 ore * → Lezione 30

Per quanto possibile si cercherà di seguire questa "ricetta":

1) Dominio
2) Simmetrie e periodicità
3) Zeri [equazione $f(x) = 0$]
4) Segni [disequazione $f(x) > 0$]
5) Limiti [negli "estremi" (in senso lato) del dominio]
6) Asintoti
7) Derivata prima
8) Limiti della derivata prima [non se $\lim_{x \to x_0} f(x) = \pm\infty$]
9) Crescenza/decrescenza, max/min, sup/inf [diseq. $f'(x) > 0$]
10) Derivata seconda, e magari suo studio [talvolta proibitivo]
11) Valori calcolati, ed eventualmente tangente in essi
12) Disegno del grafico

Le funzioni 2π-periodiche, si studino in $[0, 2\pi[$ o meglio $]-\pi, \pi]$. Analoghe riduzioni del dominio si attuino per altre periodicità.

BOZZA - DRAFT

29.5 Esempio 1: la campana gaussiana

$$f(x) := \frac{1}{\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{x^2}{2}}$$

1) Dominio

   $\mathbb{R}$

2) Simmetrie e periodicità

   $f(-x) = f(x)$, allora la funzione è pari e allora il grafico è simmetrico rispetto all'asse y.

(Da ora basterebbe studiarla per $x \geq 0$, volendo "risparmiare").

3) Zeri [equazione $f(x) = 0$]

   La funzione non ha zeri

4) Segni [disequazione $f(x) > 0$]

   $f(x) > 0 \ \forall\, x \in dom\, f$

5) Limiti [negli “estremi" (in senso lato) del dominio]

$$\lim_{x\to\pm\infty} f(x) = 0$$

6) Asintoti

$y = 0$ asintoto orizzontale per $x \to \pm\infty$

7) Derivata prima

$$f'(x) = D\,\frac{1}{\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{x^2}{2}} = \frac{1}{\sqrt{2\,\pi}}\,D\,\mathrm{e}^{-\frac{x^2}{2}} = \frac{1}{\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{x^2}{2}}\,D\left(-\frac{x^2}{2}\right) =$$

$$= -\frac{x}{\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{x^2}{2}}$$

8) Limiti della derivata prima [non se $\lim_{x\to x_0} f(x) = \pm\infty$]

$$\lim_{x\to\pm\infty}\left(-\frac{x}{\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{x^2}{2}}\right) = \lim_{x\to\pm\infty}\frac{-x}{\sqrt{2\,\pi}\,\mathrm{e}^{\frac{x^2}{2}}} = \left(=\frac{\infty}{\infty}\right) <=^{H}$$

con de L’Hospital

$$= \lim_{x\to+\infty}\frac{-1}{\sqrt{2\,\pi}\,\mathrm{e}^{\frac{x^2}{2}}\cdot x} = 0$$

9) Crescenza/decrescenza, max/min, sup/inf [diseq. $f'(x)>0$]

$$-\frac{x}{\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{x^2}{2}} > 0 \quad \Big/ \cdot\frac{\sqrt{2\,\pi}}{\mathrm{e}^{-\frac{x^2}{2}}} > 0$$

$$-x > 0$$

$$x < 0$$

‘schema formale di crescenza e decrescenza"

. . . . . 0 . . . . . .
++++++++++ | ----- $f'(x) > 0$
. . . . . ↗ . . . | . . . ↘ . . . . .

$f'(x) > 0$ per $x < 0$, funzione crescente
$f'(x) < 0$ per $x > 0$, funzione decrescente
$x = 0$ punto di max rel. e ass.

BOZZA - DRAFT

$\max f = f(0) = \frac{1}{\sqrt{2\pi}}$
$\inf f = 0$

10) Derivata seconda, e magari suo studio [talvolta proibitivo] Con qualche calcolo si trova

$$f''(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}(x^2 - 1)$$

Si risolva la corrispondente disequazione col > 0

$$\Big/ : \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} > 0$$

$$x^2 - 1 > 0$$

producendo lo schema di concavità e convessità e trovando l'unico flesso in $x \geq 0$ ovvero, su tutto $\mathbb{R}$, i 2 flessi simmetrici.

11) Valori calcolati, ed eventualmente tangente in essi

12) Disegno del grafico
Lasciato al lettore. Si verifichi poi su WolframAlpha.

BOZZA - DRAFT

29.6 Esempio 2: il Potenziale di Lennard-Jones

(È una funzione della Termodinamica, collegata alla Legge di van der Waals).

$$f(x) := \frac{1}{x^{12}} - \frac{2}{x^6}$$

(In effetti rappresenta un potenziale solo per $x > 0$).
1) Dominio
$x \neq 0$
2) Simmetrie e periodicità
$f(-x) = f(x)$, allora la funzione è pari e allora il grafico è simmetrico rispetto all'asse y.
Da adesso basta studiarla per $x \geq 0$.
3) Zeri [equazione $f(x) = 0$]

Equazione

$$\frac{1}{x^{12}} - \frac{2}{x^6} = 0$$

$$\frac{1}{x^{12}} = \frac{2}{x^6} \quad \Big/ \cdot x^{12} \neq 0 \, nel \, dominio$$

$$1 = 2\,x^6 \quad \Big/ : 2$$

$$x^6 = 2^{-1} \quad \Big/^{\wedge \frac{1}{6}} \text{ Attenzione: in } \mathbb{R}^+$$

$$x = \left(2^{-1}\right)^{\frac{1}{6}}$$

$$x = 2^{-\frac{1}{6}}$$

4) Segni [disequazione $f(x) > 0$]

$$\frac{1}{x^{12}} - \frac{2}{x^6} > 0$$

$$\frac{1}{x^{12}} > \frac{2}{x^6} \quad \Big/ \cdot x^{12} > 0 \, per \, x > 0$$

$$1 > 2\,x^6 \quad \Big/ : 2$$

$$x^6 < 2^{-1} \quad \Big/^{\wedge \frac{1}{6}} \text{ Attenzione: in } \mathbb{R}^+$$

$$x < \left(2^{-1}\right)^{\frac{1}{6}}$$

$f(x) > 0$ per $0 < x < 2^{-\frac{1}{6}}$ (e poi sarà simmetricamente per $x < 0$)

5) Limiti [negli "estremi" (in senso lato) del dominio **(ora** $\mathbb{R}^+$**)]**

$$\lim_{x \to 0^+} \left(\frac{1}{x^{12}} - \frac{2}{x^6}\right) (= +\infty - \infty) =$$

$$= \lim_{x \to 0^+} \frac{1}{x^6} \cdot \left(\frac{1}{x^6} - 2\right) = +\infty$$

$$\lim_{x \to +\infty} f(x) = 0$$

6) Asintoti

BOZZA - DRAFT

$x = 0$ asintoto verticale

$y = 0$ asintoto orizzontale per $x \to +\infty$

7) Derivata prima

$$D\left(\frac{1}{x^{12}} - \frac{2}{x^6}\right) =$$

$$= D\left(x^{-12} - 2\,x^{-6}\right) =$$

$$= -12\,x^{-13} - 2 \cdot (-6)\,x^{-7} =$$

$$= -\frac{12}{x^{13}} + \frac{12}{x^7}$$

8) Limiti della derivata prima [non se $\lim_{x \to x_0} f(x) = \pm\infty$]

$$\lim_{x \to 0^+} \left(-\frac{12}{x^{13}} + \frac{12}{x^7}\right) (= -\infty + \infty)$$

$$= \lim_{x \to 0^+} \frac{12}{x^7} \cdot \left(-\frac{1}{x^6} + 1\right) = -\infty$$

$$\lim_{x \to +\infty} \left(-\frac{12}{x^{13}} + \frac{12}{x^7}\right) = 0$$

9) Crescenza/decrescenza, max/min, sup/inf [diseq. $f'(x) > 0$]

$$-\frac{12}{x^{13}} + \frac{12}{x^7} > 0 \quad \Big/ \cdot \frac{x^{13}}{12} > 0 \; per \; x > 0$$

$$-1 + x^6 > 0$$

$$x^6 > 1 \quad \Big/^{\wedge \frac{1}{6}} \text{ Attenzione: in } \mathbb{R}^+$$

$$x > 1$$

"schema formale di crescenza e decrescenza"

```
0.........1.........
|-----|++++++++++ f'(x) > 0
|.....↘...|...↗.....
```

$f'(x) < 0$ per $0 < x < 1$, funzione decrescente

$f'(x) > 0$ per $x > 1$, funzione crescente

$x = 1$ punto di min rel. e ass.

BOZZA DRAFT

$\min f = f(1) = \frac{1}{1} - \frac{2}{1} = -1$
$\sup f = +\infty$

10) Derivata seconda, e magari suo studio [talvolta proibitivo]

$$f''(x) = D\,f'(x) = D\left(-12\,x^{-13} + 12\,x^{-7}\right) =$$

$$= -12\,(-13)\,x^{-14} + 12\,(-7)\,x^{-8} =$$

$$= 156\,x^{-14} - 84\,x^{-8} =$$

$$= \frac{156}{x^{14}} - \frac{84}{x^8}$$

Disequazione $f''(x) > 0$

$$\frac{156}{x^{14}} - \frac{84}{x^8} > 0 \quad \Big/ \cdot \frac{x^{14}}{12} > 0 \; per\; x > 0$$

$$13 - 7\,x^6 > 0$$

$$13 > 7\,x^6 \quad \Big/ : 7 > 0$$

$$\frac{13}{7} > x^6 \quad \Big/^{\wedge\frac{1}{6}} \text{ Attenzione: in } \mathbb{R}^+$$

$$x < \left(\frac{13}{7}\right)^{\frac{1}{6}} \text{ (è una radice sesta)}$$

BOZZA - DRAFT

"schema formale di concavità e convessità"

$0 \ldots\ldots\ldots \left(\frac{13}{7}\right)^{\frac{1}{6}} \ldots\ldots\ldots\ldots$
$|{+}{+}{+}{+}{+}{+}{+}{+}{+}{+}|{-}{-}{-}{-}{-}\; f'(x) > 0$
$|\ldots\ldots \cup \ldots |\ldots \cap \ldots\ldots$

$f''(x) > 0$ per $0 < x < \left(\frac{13}{7}\right)^{\frac{1}{6}}$, concavità verso l'alto
$f''(x) < 0$ per $x > \left(\frac{13}{7}\right)^{\frac{1}{6}}$, concavità verso in basso
$x = \left(\frac{13}{7}\right)^{\frac{1}{6}}$ punto di flesso

11) Valori calcolati, ed eventualmente tangente in essi

$$f(2) = \frac{1}{2^{12}} - \frac{2}{2^6} =$$

$$= \frac{1 - 2 \cdot 2^6}{2^{12}} = \frac{1 - 128}{4\,096} = -\frac{127}{4\,096}$$

$$f(1/2) = f(2^{-1}) = \frac{1}{2^{-12}} - \frac{2}{2^{-6}} =$$

$$= 2^{12} - 2 \cdot 2^6 = 4\,096 - 128 = 3\,968$$

12) Disegno del grafico.

Lasciato al lettore. Poi verificare su WolframAlpha.

BOZZA - DRAFT

29.7 Le equazioni differenziali

Un'equazione differenziale è un'uguaglianza fra funzioni, di cui almeno una, sia essa $y(t)$, è incognita, e in tutta l'espressione compare almeno una volta $y'(t)$ *vel* una derivata di ordine superiore.
Eccone 5:

$y' = 9.81$ che risolveremo in una lezione sugli integrali

$y' = t$

$y' = -k\,y$ un'equazione classica della Farmacocinetica

$y'' = -y$ un'equazione classicissima in Fisica.

$y''' = t\,y'^2$

Internazionalmente si usa l'acronimo ODE per Ordinary Differential Equation (equazione differenziale ordinaria) se di funzioni incognite ce n'è solo una. Solo di queste ci occuperemo in questo paragrafo, e solo a un livello infimo. (L'argomento potrebbe riempire più di un corso universitario). Nel prossimo paragrafo vediamo un esempio più complesso, il modello SIR dell'epidemiologia.

29.8 Il modello SIR dell'epidemiologia

Ecco un classico sistema di equazioni differenziali dell'epidemiologia:

$$\begin{cases} \dfrac{dS}{dt} = -\dfrac{\beta IS}{N}, \\ \dfrac{dI}{dt} = \dfrac{\beta IS}{N} - \gamma I, \\ \dfrac{dR}{dt} = \gamma I, \end{cases}$$

nelle 3 funzioni incognite $S(t)$, $I(t)$, $R(t)$, *susceptible*, *infectious*, *removed*, in italiano detti anche (ma non è l'esatta traduzione) Suscettibile, Infetto, Risolto (ma in Risolto stanno sia i guariti che i deceduti). Naturalmente $\frac{dS}{dt}$ è solo una scrittura alternativa per S', ovvero $S'(t)$.

> Questo modello è ragionevolmente predittivo per le malattie infettive che vengono trasmesse da uomo a uomo e in cui il recupero conferisce resistenza duratura (Wikipedia, l'enciclopedia libera, letto il 10 novembre 2021).

BOZZA - DRAFT

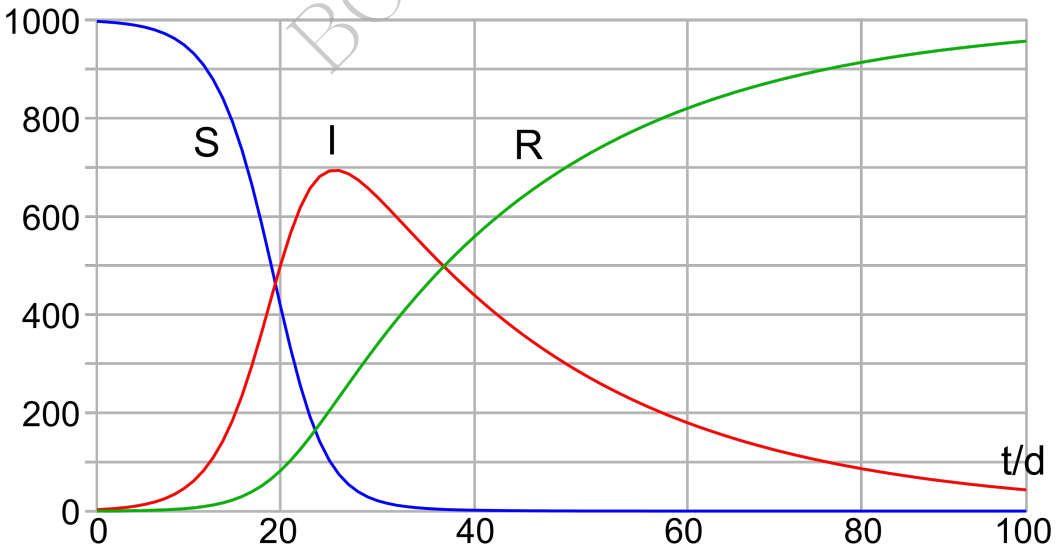


Figure 38: https://commons.wikimedia.org/wiki/File:SIR-Modell.svg, By Klaus-Dieter Keller. Valori iniziali $S = 997$ e $I = 3$, $R = 0$ (cioè 3 infetti su 1000), con parametri $\beta = 0.0004$, $\gamma = 0.04$.

29.9 Esercizi di studio di funzione

Studiare queste 3 **Funzioni Iperboliche**:

$$\begin{array}{ll} \sinh x := \frac{e^x - e^{-x}}{2} & \textit{seno iperbolico} \\ \cosh x := \frac{e^x + e^{-x}}{2} & \textit{coseno ip.} \\ \tanh x := \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}} & \textit{tangente ip.} \end{array}$$

Le funzioni iperboliche hanno un interesse di per sè, tendendo a ricorrere nella Fisica, e hanno anche uno speciale interesse nel Calcolo Differenziale.

Studiare questa funzione di interesse nel Calcolo delle Probabilità:

- $$f(x) := \frac{1}{\pi} \cdot \frac{1}{1 + x^2}$$

Studiare queste altre funzioni:

- $$f(x) := \frac{x}{2 + x^2}$$
- $$f(x) := \ln(x^2 - 1)$$

BOZZA - DRAFT

VI – Integrali e serie numeriche

BOZZA - DRAFT

29.10 Riempitivo lorem ipsum

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

BOZZA - DRAFT

30 Integrale indefinito e ODE

30.1 Introduzione all'integrale indefinito

Abbiamo visto come rispondere[136] alla domanda

Qual è la derivata di questa funzione?

Ma nelle Scienze Applicate e in particolare in Fisica[137] talvolta si deve rispondere a questa domanda in qualche modo inversa:

Di quale/i funzione/i è derivata questa funzione?

Per esempio

accelerazione = D velocità

entrambe espresse in funzione del tempo.

Se conosciamo la velocità, derivando troviamo l'accelerazione.

Ricorrente problema in Fisica è invece trovare la velocità dall'accelerazione.

L'operazione inversa della derivazione è l'integrazione indefinita.

Se conosciamo l'accelerazione, integrando troviamo la velocità, con qualche precisazione che vedremo.

BOZZA - DRAFT

[136] In generale e salvo funzioni capricciose: già $|x|$ crea problemi, non avendo derivata in 0.
[137] Meno spesso in Chimica e poi in Farmacia.

L'integrale indefinito è l'"*anti-derivata*":

$$\int \frac{1}{1+x^2}\,dx = \arctan x + c$$

$$\Leftarrow \quad \frac{1}{1+x^2} = D\left(\arctan x + c\right).$$

Possiamo quasi dire che il segno d'integrale può fare il "salto dell'uguale" trasformandosi in derivata, come un ln fa il "salto" trasformandosi in exp, per esempio $\ln x^2 = 3$ equivale a $x^2 = \mathrm{e}^3$.

L'integrale indefinito risponde alla domanda: "Di cos'è la derivata?"

Il termine "$+c$" ci ricorda che non solo $\arctan x$ ha derivata $\frac{1}{1+x^2}$, ma anche $\arctan x + 99$ o più qualunque numero reale c. Allora l'integrale indefinito è un insieme di funzioni.
Ciascuna di quelle funzioni si chiama *primitiva* di f.

BOZZA - DRAFT

30.2 Alcuni integrali indefiniti

Dalle note derivate bbiamo subito questi integrali indefiniti:

	$\int \frac{1}{1+x^2}\,dx$	=	$\arctan x + c$	
$\forall\,\alpha$	$\int \alpha\,dx$	=	$\alpha x + c$	in particolare: $\int 0\,dx = c$
	$\int \mathrm{e}^x\,dx$	=	$\mathrm{e}^x + c$	
	$\int \frac{1}{x}\,dx$	=	$\ln\lvert x\rvert + c$	
$\forall\,\alpha \neq -1$	$\int x^\alpha\,dx$	=	$\frac{x^{\alpha+1}}{\alpha+1} + c$	in particolare $\int x\,dx = \frac{x^2}{2} + c$ $\int x^2\,dx = \frac{x^3}{3} + c$ $\int \frac{1}{x^2}\,dx = -\frac{1}{x} + c$ $\int \sqrt{x}\,dx = \frac{2}{3}x^{\frac{3}{2}} + c$
	$\int \cos x\,dx$	=	$\sin x + c$	*Nota: non diamo alcun*
	$\int \sin x\,dx$	=	$-\cos x + c$	*significato al termine dx.*

30.3 La costante additiva dell'integrale indefinito

Al posto del $+c$ delle sopra dette formule d'integrazione definita, altri Autori scrivono $+k$ o $+cost$, o quant'altro, e senza errare troppo immaginiamo che vi sia 0, pur scrivendo formalmente $+c$. (La questione tecnicamente è più sottile).

Quella simbologia va intesa come la somma di una qualunque costante in ognuno degli intervalli massimali[138] contenuti nel dominio dell'integranda, che è questione un po' sottile, ma senza errare troppo immaginiamo che sia 1 stessa costante su tutto il dominio, liberamente scelta in $\mathbb{R}$. Per esempio 3, o -5, e spesso 0.

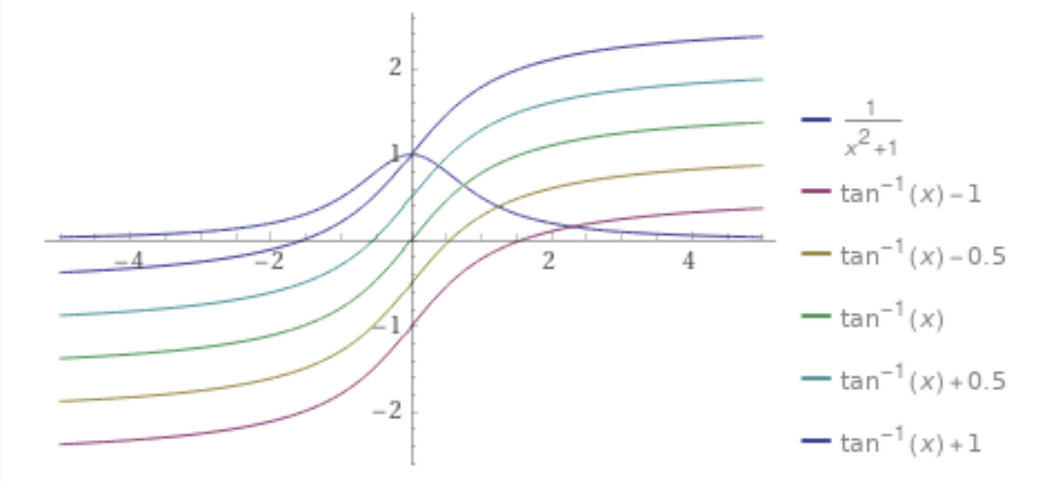


Figure 39: Disegno del grafico di $\frac{1}{1+x^2}$ e di alcune delle funzioni del suo integrale indefinito $\arctan x + c$, cioè alcune primitive. (Screenshot da WolframAlpha).

30.4 Le 2 formule della linearità dell'integrale

Queste 2 formule costituiscono la *linearità dell'integrale*:

$$\boxed{\int c\, f(x)\, dx = c \int f(x)\, dx \qquad \forall\, c \in \mathbb{R}} \tag{82}$$

Cioè: Le costanti moltiplicative escono dall'integrale.

Esempio: $\int 5 \cos x\, dx = 5 \int \cos\, dx = 5 \sin x + c$

[138] Per esempio l'integrale indefinito di $\frac{1}{x}$ è $\ln|x| + c$ e il dominio dell'integranda è costituito dai 2 intervalli $]-\infty, 0[$ e $]0, +\infty[$ e per ciascuno di essi si avrà una costante da sommare a $\ln x$, in modo indipendente). Per esempio $\ln(-x) + 3$ per $x < 0$ e $\ln x + 5$ per $x > 0$.

$$\boxed{\int (f(x)+g(x))\,dx = \int f(x)\,dx + \int g(x)\,dx} \qquad (83)$$

Cioè: L'integrale della somma è la somma degli integrali.

Esempio: $\int(5+\cos x)\,dx = \int 5\,dx + \int \cos x\,dx = 5\,x + \sin x + c$

Nota 1 per il Lettore interessato. Nei Complementi si trovano altre formule, in particolare la bella *formula d'integrazione per parti*:

$$\int f' \cdot g = f \cdot g - \int f \cdot g'$$

Nota 2 per il Lettore interessato. Le sosprastanti proprietà (teoremi) degli integrali indefiniti e quelle riportate nella corrispondente sezione di Complementi seguono (non tutte in modo ovvio) dalle proprietà (teoremi) delle derivate.

BOZZA - DRAFT

Esempio 1

$$\diamond \qquad \int (x^3 + 2\,x^2 + 3\,x)\,dx =$$

$$=^{(83)} \int x^3\,dx + \int 2\,x^2\,dx + \int 3\,x\,dx =$$

$$=^{(82)} \int x^3\,dx + 2\int x^2\,dx + 3\int x\,dx =$$

$$= \frac{1}{4}\,x^4 + \frac{2}{3}\,x^3 + \frac{3}{2}\,x^2 + c\,. \qquad \text{[Verificare derivando]}$$

Esempio 2

$$\diamond \qquad \int (\sin x + 3\cos x + 7)\,dx =$$

$$=^{(83)} \int \sin x\,dx + \int 3\cos x\,dx + \int 7\,dx =$$

$$=^{(82)} (-\cos x + c_1) + 3\int \cos x\,dx + (7x + c_2) =$$

$$= (-\cos x + c_1) + 3\,(\sin x + c_3) + (7x + c_2) =$$

chiamiamo c la somma delle 3 costanti d'integrazione c_1, c_2 e c_3

$$= -\cos x + 3\sin x + 7x + c\,. \qquad \text{[Verificare derivando]}$$

30.5 Un'equazione differenziale della Farmacocinetica

Supponendo che un farmaco lasci il sangue solo per via urinaria e ciò avvenga proporzionalmente alla sua concentrazione nel sangue (quando ce n'è molto, molto esce) si ha l'equazione differenziale nell'incognita concentrazione $y(t)$

$$y'(t) = -k \cdot y(t)$$

essendo k una costante positiva che dipende essenzialmente dal farmaco (e avrà poi una variabilità ulteriore, da soggetto a soggetto).

Non mostriamo come, che è complesso, ma si sappia questo: si risolve con un'integrazione.

BOZZA - DRAFT

30.6 Equazioni differenziali risolubili con un'integrazione

Talvolta nelle Scienze Applicate si è interessati a trovare una funzione del tempo, sia essa $y(t)$, di cui per motivi fisici (in senso lato) si conosce la derivata, sia essa $g(t)$:

$$y'(t) = g(t) \quad y\ incognita \quad g\ nota$$

Se g è sufficientemente regolare ovvero non "capricciosa", ovviamente le infinite soluzioni dell'*equazione differenziale* sopra scritta costituiscono l'integrale indefinito di g:

$$y(t) = \int g(t)\,dt$$

che, detta $G(t)$ una qualunque primitiva di $g(t)$, sono le infinite

$$= G(t) + cost$$

Per esempio per un oggetto che cade verticalmente presso la Terra in assenza di aria si sa che la derivata della velocità $v(t)$, cioè l'accelerazione, è ≈ 9.81:

$$v'(t) = 9.81$$

$$\Rightarrow v(t) = \int 9.81\,dt$$
$$\Rightarrow v(t) = 9.81\,t + c$$

e ponendo in quest'ultima $t = 0$, cioè il tempo d'inizio della caduta, si trova

$$c = v(0)$$

che magari vorremo indicare con $v_{iniziale}$ ottenendosi infine

$$v(t) = 9.81\,t + v_{iniziale}$$

e similmente sulla Luna, già di suo priva di aria,

$$v(t) = 1.62\,t + v_{iniziale} \qquad \text{(Luna)}$$

(e ovviamente entrambe le formule sono inevitabilmente approssimate, non essendo esatte le 2 costanti numeriche).

BOZZA - DRAFT

Nota. Di questa Lezione ci si dovrebbe aspettare che lo studente conosca a memoria le formule *contemporaneamente* numerate *et* riquadrate, oltre a tutto l'inquadramento teorico della questione, e le formule del paragrafo 30.2.

Complementi

30.7 Complementi – Altre formule sull'integrale indefinito

Integrazione per parti:

$$\int f'(x)g(x)\,dx = f(x)g(x) - \int f(x)g'(x)\,dx \tag{84}$$

Esempio: $\int \mathrm{e}^x\,x\,dx = \mathrm{e}^x\,x - \int \mathrm{e}^x\cdot 1\,dx = \mathrm{e}^x\,x - \mathrm{e}^x + c$

Integrazione per sostituzione $y := mx + q$:

$$\int f(m\,x+q)\,dx = \frac{1}{m}\Big(\int f(y)\,dy\Big)_{y=m\,x+q} \quad \forall\, m,q\in\mathbb{R}, m\neq 0. \tag{85}$$

(Una formula più complicata che non diamo permette una sostituzione più generale $y := g(x)$, ed è là che si dà senso al dx).

BOZZA - DRAFT

Esempio:

$$\diamond \int \cos(2\,x+3)\,dx =^{(85)}$$
$$= \frac{1}{2}\Big(\int \cos(t)\,dx\Big)_{t=2\,x+3} =$$
$$= \frac{1}{2}\Big(\sin(t)+k\Big)_{t=2\,x+3} =$$
$$= \frac{1}{2}\sin(2\,x+3)+c\,. \quad \text{[Verificare derivando]}$$

30.8 Complementi – Integrale del logaritmo naturale

Con la formula di integrazione per parti, intendendo $\ln x$ come $1\cdot \ln x$, si può trovare l'integrale di $\ln x$:

$$\int \ln x\,dx = \quad \text{moltiplichiamo per 1}$$

$$= \int 1 \cdot \ln x \, dx = \quad \text{riconosciamo } D\,x = 1$$

$$= \int (D\,x) \cdot \ln x \, dx =$$

con la (84)

$$= x \, \ln x - \int x \cdot \frac{1}{x} \, dx =$$

$$= x \, \ln x - \int 1 \, dx =$$

$$= x \, \ln x - x + c\,. \quad \text{[Verificare derivando]}$$

Con questo procedimento e con la formula di cambiamento di base e con la (82) si calcoli l'integrale del logaritmo decimale.

BOZZA - DRAFT

30.9 Complementi – Le equazioni differenziali

Al contrario di quanto visto nel paragrafo (29.7), le equazioni differenziali di solito non si possono risolvere così semplicemente con un'integrazione indefinita.

Già sappiamo risolvere la

$$(1 + t^2)\, y'(t) = 1 \qquad \text{altrimenti scritta } (1 + t^2)\, y' = 1$$

cioè la

$$y'(t) = \frac{1}{1 + t^2}$$

trovando

$$y(t) = \arctan t + c$$

Con WolframAlpha: `Solve[(1+t^2)y'=1]` dà $y(t) = c_1 + \tan^{-1}(t)$

Esercizio$_\mu$ Risolvere la $t^3\, y' - 5 = 0$.

SVOLGIMENTO

È un'equazione differenziale e si risolve con questi passaggi:

$$t^3\, y' = 5$$

$$y'(t) = \frac{5}{t^3}$$

$$y(t) = \int 5\, t^{-3}\, dt = 5 \int t^{-3}\, dt = 5\, \frac{t^{-3+1}}{-3+1} + c = -\frac{5}{2\,t^2} + c$$

Con WolframAlpha: `Solve[t^3y'=5]` dà $c_1 - \frac{5}{2t^2}$

Note sulle notazioni. Nell'ambito della teoria delle equazioni differenziali, è abbastanza comune usare il nome x per la funzione incognita, il che potrebbe generare fraintendimenti, per esempio

$$x'(t) = x(t) \qquad \text{ovvero } x' = x$$

Ed è abbastanza comune indicare la derivazione col punto soprascritto, per esempio

$$\dot{x}(t) = x(t) \qquad \text{ovvero } \dot{x} = x$$

e 2 punti soprascritti per la derivata seconda: $\ddot{x}$.
Anche il nome u è molto usato per la funzione incognita.
Classicissima in Fisica è l'equazione differenziale del pendolo

$$\ddot{u} = -u$$

e si capisce che è una soluzione è $\cos t$, e più in generale $\cos(t + c)$.

BOZZA - DRAFT

31 L'integrale definito

31.1 La questione dell'area sotto una curva

In un'infinità di questioni interessa conoscere l'area compresa fra il grafico di una funzione $f(x)$ e l'asse delle ascisse, da un numero a a un numero b su quell'asse. (Al limite, pure valori a o b infiniti).

Questo problema verrà risolto dall'integrale definito: $\int_a^b f(x)\,dx$.

Nel Calcolo delle Probabilità il concetto è di uso amplissimo.[139]

È usato nella Termodinamica, scienza utile nella fabbricazione dei farmaci: `link->`

Veniamo alla Farmacia. Leggiamo su Wikipedia[140] (in italiano), l'enciclopedia libera, alla voce *Area Under the Curve*,

BOZZA - DRAFT

[139] Nel Calcolo delle Probabilità l'area sotto il grafico di una *densità di probabilità di una variabile aleatoria* X, concetti che definiremo, rappresenterà la probabilità che $a \le X \le b$:

$$P(a \le X \le b) = \int_a^b f(t)\,dt$$

Si può cominciare a farsi un'idea con la *densità normale standard* $\phi(t) := \frac{1}{\sqrt{2\,\pi}} e^{-\frac{t^2}{2}}$, considerando il suo grafico. Per esempio con WolframAlpha

```
Integrate[Exp[-t^2/2]/Sqrt[2Pi],{t,-2,2}]
```

ci dà circa 95%, e un'interessante figura.

[140] Letto il 3 febbraio 2020

> L'area sotto la curva concentrazione/tempo o AUC (dalla dicitura inglese area under the time/concentration curve, ovvero area sottesa alla curva) è un parametro farmacocinetico dato dall'integrale in un grafico concentrazione/tempo (...)
> Tale parametro è fondamentale per poter descrivere l'effetto dei farmaci poiché riflette l'esposizione dei tessuti al farmaco nel tempo.
> L'AUC (da zero a infinito) rappresenta l'esposizione totale al farmaco in funzione del tempo (...)
> Per indicare l'AUC riferita ad un particolare intervallo temporale si utilizzano i pedici, ad esempio AUC_{4-8h} indica l'area sotto la curva nell'intervallo di tempo che va da 4 a 8 ore.

BOZZA - DRAFT

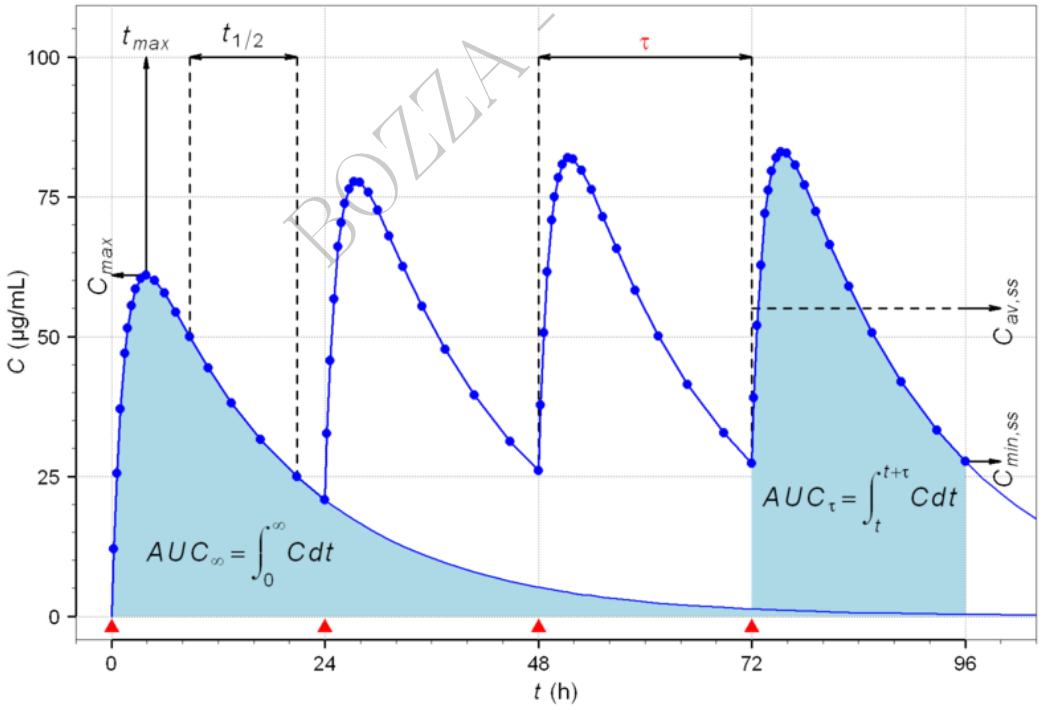


Figure 40: https://commons.wikimedia.org/wiki/File:Linear_PK_Example.png. By Helmut Schütz.

Cioè, e si veda la figura sulla citata pagina di Wikipedia,

$$AUC_\infty := \int_0^{+\infty} concentrazione(t)\,dt$$

Un'applicazione molto vicina alla Farmacia è l'area sotto il grafico che esprime la concentrazione di glucosio nel sangue all'avanzare del tempo dopo l'assunzione di un pasto; si veda questa figura di Wikimedia Commons: link->; ciò serve a definire l'indice glicemico degli alimenti:

> The glycemic index of a food is defined as the incremental area under the two-hour blood glucose response curve (AUC) following a 12-hour fast and ingestion of a food with a certain quantity of available carbohydrate (usually 50 g). (Da Wikipedia (in inglese), l'enciclopedia libera alla voce *Glycemic index*.)

Quel grafico viene certo da analisi del sangue, ma può anche essere modellizzato matematicamente con funzioni, per ulteriori ricerche: link->. (Il quel testo, si noti che *in silico* = col computer; è un aggiornamento dei classici *in vitro* e *in vivo*).

BOZZA DRAFT

Similmente l'indice insulinico:

> Glucose (glycemic) and insulin scores were determined by feeding 1000 kilojoules (239 kilocalories) of the food to the participants and recording the area under the glucose/insulin curve for 120 minutes then dividing by the area under the glucose/insulin curve for white bread. The result being that all scores are relative to white bread. . (Da Wikipedia (in inglese), l'enciclopedia libera alla voce *Insulin index*.)

Ma le applicazioni sono veramente innumerevoli.[141]

[141] In ambito Biomedico, si consideri la funzione $f(t)$ che esprime in millilitri all'ora il flusso di un liquido in un tubo, anatomico o artificiale: l'area in questione, da t_1 a t_2 in ore, rappresenta la quantità di liquido fluito, in millilitri; e ovviamente millilitri e ore possono essere cambiati con altre unità di misura di volume e tempo. Quella quantità sarà proprio $\int_{t_1}^{t_2} f(t)\,dt$.
Si consideri, ancora, la determinazione delle aree geometriche, di figure definite dal loro bordo inteso come grafico di una funzione: ciò può servire nella progettazione di protesi anatomiche (Ingegneria Biomedica) o di reattori chimici per la produzione di farmaci.

31.2 Teoria ed esempi dell'integrale definito

Siano $F(x)$ e $f(x)$ due funzioni definite fra a e b, cioè sull'intervallo $[a,b]$ se $a \le b$ e sull'intervallo $[b,a]$ se $a > b$.

Se $F(x)$ è una primitiva di $f(x)$, cioè $F' = f$, e tale primitiva si può ottenere dall'integrale indefinito di f ponendo $c = 0$ o qualunque altro valore, in questa trattazione elementare – diversamente dall'uso consueto in Analisi Matematica – **definiremo l'*integrale definito* di $f(x)$ da a a b in questo modo**

$$\int_a^b f(x)\,dx := F(b) - F(a)_{\textit{scritto anche}} \, [F(x)]_a^b \leftarrow \textit{incremento di } F \textit{ da } a \textit{ a } b$$

(ed è ovvio che qualunque primitiva si scelga, ovvero qualunque c, si ottiene lo stesso valore: c si elide nella sottrazione).

Si dimostra (teorema, complicato) che esso uguaglia l'area (consueta) fra l'asse x e il grafico di f fra a e b se $a < b$ e su $[a,b]$ è $f(x) \ge 0$. *(Sottografico).*

BOZZA - DRAFT

A un livello superiore, l'integrale definito viene definito sostanzialmente proprio con le aree, e questo consente di averlo anche per funzioni prive di primitiva come $\operatorname{sgn} x$, ma noi non considereremo tali integrali.

Si noti che
- il sottografico è una figura piana (insieme di punti del piano),
- l'integrale definito è un numero,
- la primitiva è una funzione,
- l'integrale indefinito è un insieme di funzioni.

Esempio 1. Calcoliamo l'area sotto una "campata" della sinusoide, da 0 a π:

$$\int_0^\pi \sin x\,dx = [-\cos x]_0^\pi = -\cos\pi - (-\cos 0) = -(-1) - (-1) = 2.$$

Esempio 2. Calcoliamo l'area del sottografico dell'iperbole equilatera di equazione $y := \frac{1}{x}$ da 1 a 7:

$$\int_1^7 \frac{1}{x}\,dx = \left[\ln x\right]_1^7 = \ln 7 - \ln 1 = \ln 7$$

ed è ben evidente che il numero 7 si può sostituire con qualunque numero t, ottenendosi il significato geometrico del logaritmo. Se $0 < t < 1$ il logaritmo naturale è negativo, e infatti l'area del sottografico, da intendersi come *area con segno*, è negativa perché la base viene percorsa da 1 a t in verso contrario all'orientazione dell'asse x. Si disegni l'iperbole e si stimi (dall'area) $\ln 10$.

Esempio 3. Per un processo termodinamico isotermico reversibile vale

$$Lavoro = -\int_{V_{iniziale}}^{V_{finale}} \frac{n\,R\,T}{V}\,dV =$$

per linearità dell'integrale definito, come pure indefinito,

$$= -n\,R\,T \int_{V_{iniziale}}^{V_{finale}} \frac{1}{V}\,dV =$$

$$= -n\,R\,T\left[\ln V\right]_{V_{iniziale}}^{V_{finale}} =$$

$$= -n\,R\,T\left(\ln V_{finale} - \ln V_{iniziale}\right) =$$

$$= -n\,R\,T\,\ln\frac{V_{finale}}{V_{iniziale}}$$

BOZZA - DRAFT

Leggiamo su Wikipedia (in inglese), l'enciclopedia libera, alla voce *Isothermal process*:

> By convention, work is defined as the work on the system by its surroundings. If, for example, the system is compressed, then the work is positive and the internal energy of the system increases. Conversely, if the system expands, it does work on the surroundings and the internal energy of the system decreases.

31.3 Altri 3 teoremi sugli integrali

• Teorema fondamentale del Calcolo Integrale

$$D\int_a^x f(t)\,dt = f(x)$$

• **Scambio degli estremi di integrazione.**
Dalla definizione – come l'abbiamo data – segue subito che

$$\int_b^a f(x)\,dx = -\int_a^b f(x)\,dx$$

• **Regola di Chasles.**
Per ogni $a, b, c \in \text{dom } f$, in qualunque ordine,

$$\boxed{\int_a^b = \int_a^c + \int_c^b} \tag{86}$$

(dove si intendano completati i 3 simboli di integrale con $f(x)dx$, che per focalizzare il significato di questo teorema, non è stato scritto esplicitamente).

Esercizio. Con la Regola di Chasles si calcoli $\int_{-1}^{2} |x|\,dx$, si faccia un disegno, e si ricalcoli quell'integrale con le aree della geometria elementare. (Si noti che di $|x|$ non abbiamo dato una primitiva, ma quella funzione coincide con $-x$ fino a 0, e con x da 0 in poi).

BOZZA - DRAFT

ESERCIZIO$_{\mu 2018*}$
$*$ Supponiamo che in un tubo (per esempio di un reattore chimico per la produzione di farmaci) passi un liquido nella misura di

$$p(t) := |t|\,dl/h$$

nel tempo t fra -1 e 2 (h, ore, unità di tempo; quella negativa indica un tempo anteriore a un tempo detto 0; si può ipotizzare, per esempio, che lo 0 sia una certa mezzanotte, e allora stiamo considerando il tempo dalle 23 alle 2 di notte, in effetti del giorno successivo). Calcolare

$$\int_{-1}^{2} p(t)\,dt$$

che da un punto di vista fisico – ma non ce ne occuperemo – rappresenta la quantità totale di liquido fluito nel tempo considerato, e ha unità di misura dl, cioè decilitri, ma per semplicità si facciano i calcoli e si dia la soluzione senza unità di misura.

SVOLGIMENTO
Ricordando che

$$|x| := \begin{cases} x & se\ x \geq 0 \\ -x & se\ x < 0 \end{cases}$$

ora (eliminando le unità di misura)

$$\int_{-1}^{2} p(t)\,dt = \int_{-1}^{2} |t|\,dt =$$

per la Regola di Chasles

$$= \int_{-1}^{0} (-t)\,dt + \int_{0}^{2} t\,dt =$$

per la linearità dell'integrale (cioè per le formule (82) e (83))

$$= -\int_{-1}^{0} t\,dt + \int_{0}^{2} t\,dt =$$

ricordando la primitiva elementare $\int t\,dt = t^2/2 + c$

$$= -\left[\frac{t^2}{2}\right]_{-1}^{0} + \left[\frac{t^2}{2}\right]_{0}^{2} = -\left(0 - \frac{1}{2}\right) + \left(\frac{4}{2} - 0\right) = \frac{5}{2} =$$

esattamente

$$\boxed{2.5}$$

oppure anche validamente (ma in effetti meno bene, visto il significato fisico)

$$\boxed{\tfrac{5}{2}}$$

(Si tratta di due decilitri e mezzo, dal punto di vista fisico, molto meglio espressi – almeno come risultato finale – come 2.5 che $\frac{5}{2}$).

Esercizio$_{\mu}$ Si consideri la funzione

$$f(t) := \frac{1}{\pi} \cdot \frac{1}{1+t^2} \quad \text{(che si chiama } \textit{densità di Cauchy}\text{)}$$

e si calcoli

$$\int_{0}^{1} f(t)\,dt$$

e si troverà $\frac{1}{4} = 0.25 = 25\%$ che rappresenta – anticipando un argomento di Calcolo delle Probabilità – la probabilità che una variabile casuale (o meglio detta aleatoria) che ha quella densità assuma un valore fra 0 e 1. Si faccia un disegno della funzione ombreggiando il sottografico relativo all'area trovata. Con analogo calcolo, o semplicemente per simmetria, ne viene che

$$\int_{-1}^{1} f(t)\,dt = \frac{1}{2} = 0.5 = 50\%$$

BOZZA - DRAFT

che rappresenterà la probabilità che una variabile aleatoria che ha quella densità assuma un valore fra −1 e +1.
Si calcoli poi

$$\int_0^{+\infty} f(t)\,dt$$

e ovviamente l'estremo $+\infty$ va inteso nel senso del limite[142]. Si troverà $\frac{1}{2} = 0.5 = 50\%$ che rappresenterà la probabilità che una variabile aleatoria che ha quella densità assuma un valore fra 0 e $+\infty$ ovvero non negativo.
Infine si calcoli

$$\int_{-\infty}^{+\infty} f(t)\,dt$$

e si troverà 1 cioè 100% e questo valere 1 dell'integrale su tutto $\mathbb{R}$ avviene per tutte le densità di probabilità (che, poi, hanno anche la caratteristica di essere ≥ 0).

Questa dell'esercizio soprastante grafico di

$$f(t) := \frac{1}{\pi} \cdot \frac{1}{1+t^2}$$

non è la curva a campana che più ricorre nelle applicazioni come densità, che invece è la campana gaussiana già accennata, ma ha il pregio di essere più facilmente trattabile pur esibendo comportamenti in parte simili. (I valori $f(t)$ si possono bene calcolare approssimatamente con le 4 operazioni).

BOZZA - DRAFT

ESERCIZIO$_{\mu 2021}$ ≈ Approssimando $\arctan(20)$ con $\arctan(+\infty)$, calcolare

$$\int_{-20}^{20} f(t)\,dt, \qquad f(t) := \frac{1}{1+t^2}$$

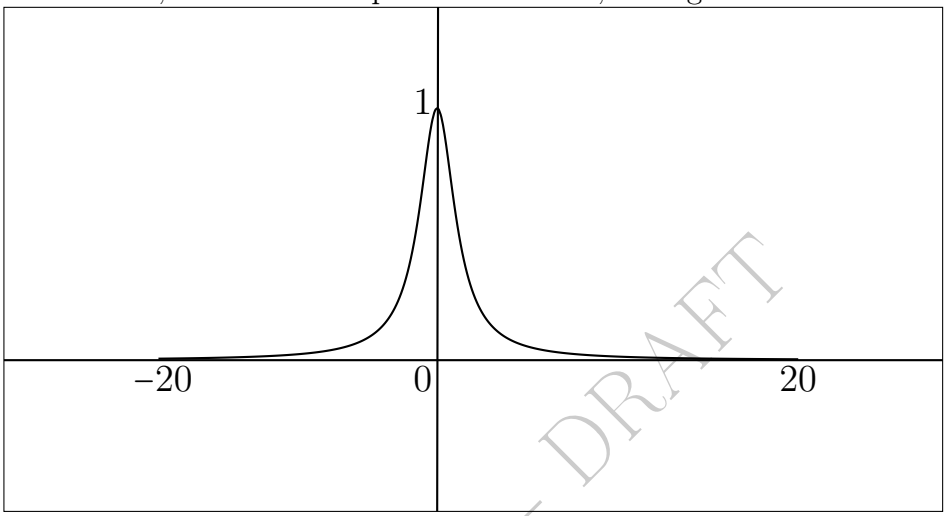
e potrà essere utile questa tavola di valori, dove ovviamente l'infinito va inteso nel senso del limite: somma di una qua

x	0	$\sqrt{3}$	$+\infty$	
$\arctan x$	0	$\frac{\pi}{3}$	$\frac{\pi}{2}$	$(\arctan(-x) = -\arctan x)$

Di questa questione possiamo dare un'interessante interpretazione, seppure non serva per risolvere il quesito. Il numero di morti di un'epidemia, in migliaia,

[142] Per le funzioni molto regolari che ci interessano in pratica, questo non creerà alcun problema, e nei trattati di Calcolo delle Probabilità si scrivono tranquillamente cose come $\arctan(+\infty)$, formalmente scorrette dal punto di vista matematico.

sia approssimatamente modellizzato dalla $f(t)$ considerata, per $-20 \le t \le 20$ essendo t il tempo, dal giorno $t = -20$ (per esempio 11 dicembre 2020) al tempo $t = 20$ (20 gennaio 2021). (Per esempio, vediamo che nel giorno di picco $t = 0 =$ 31 dicembre 2020 il modello dà 1000 morti).
Allora ovviamente l'area espressa dall'integrale considerato dà, molto approssimativamente, il numero complessivo di morti, in migliaia.



BOZZA - DRAFT

SVOLGIMENTO

$$\int_{-20}^{20} \frac{1}{1+t^2}\,dt =$$

$$= \left[\arctan t\right]_{-20}^{20} =$$

$$= \arctan(20) - \arctan(-20) =$$

per la disparità dell'arcotangente ricordata nella tavola

$$= \arctan(20) - (-\arctan(20)) =$$

$$= 2\arctan(20) \approx 2\arctan(+\infty) = 2\,\frac{\pi}{2} = \pi \approx$$

$$\boxed{\approx 3.14}$$

(e nell'interpretazione dell'epidemia corrisponde a circa 3140 morti, diciamo pure "circa 3000 morti" a causa delle grossolane approssimazioni).
Nota 1. È stata un'epidemia molto intensa solo per breve tempo, con un'ascesa verso il picco e una discesa in certo qual modo rapidissime, con buona parte dei morti concentrati in una settimana, molto diversa dalla pandemia attuale

(2020-2021), per la quale – nemmeno per la sola Italia – una formula semplicissima come quella dell'esercizio non può bastare. Ecco alcuni valori dati dalla funzione dell'esercizio:

31 dicembre 1000 morti
1 gennaio 500 morti
2 gennaio 200 morti
3 gennaio 100 morti

Nota 2. Se invece dell'integrale, con un computer (o a mano con molta pazienza) sommiamo i morti previsti dal modello per i 41 giorni dell'epidemia, si ha, molto più precisamente,

$$\sum_{t=-20}^{20} \frac{1}{1+t^2} \approx 3.056$$

cioè ≈ 3056 morti.

BOZZA - DRAFT

31.4 Note finali sulle funzioni elementari

Le regole della Natura sono scritte in linguaggio matematico, ci insegna Galileo Galilei.

Particolarmente efficaci in questo senso sono le *funzioni elementari*: potenze, logaritmi, esponenziali, valore assoluto, funzioni trigonometriche... (Una completa elencazione sarebbe difficile).

È spettacolare quanti fenomeni si possono descrivere con formule fatte di funzioni elementari con le 4 operazioni più l'elevamento a potenza – in particolare la potenza 0.5 che dà la radice quadrata.

Un modello matematico può approssimativamente essere definito come una relazione matematica – diciamo pure in generale una funzione o meglio un'equazione con più variabili – che cattura certi aspetti della realtà sensibile, quantificati da quelle variabili.
Per esempio

$$funzione(variabile_1, ..., variabile_n) = 0$$

come la $pV - nRT = 0$ della Termodinamica.
E, sorprendentemente spesso, quella funzione o equazione è fatta

con funzioni elementari combinate con le 4 operazioni e l'elevamento a potenza. (Talvolta ci sono funzioni *non elementari*).

Entusiasmati dal successo, c'è il rischio di esagerare.
(E "trasformare tutto in un numero").

Osserviamo due limitazioni:

1) la relazione matematica è quasi sempre approssimata;

2) smette di valere al di fuori di certi range (domini) in generale non ben specificati, perdendo progressivamente bontà.

Esempio 0, di Microbiologia: l'accrescimento microbico.
Abbiamo già ben detto, che il numero di microbi in una coltura non cresce per sempre esponenzialmente, ma solo – e comunque approssimativamente – all'inizio. Similmente con qualunque altra popolazione; come pure non c'è epidemia che non finisca.

BOZZA DRAFT

Esempio 1, di Fisica e Chimica: i vasi comunicanti.
Definito un certo sistema fisico, detto dei vasi comunicanti, com'è noto vale la relazione

$$h_1 = h_2$$

Fin dall'inizio c'è l'approssimazione della non perfetta planarità delle superfici liquide nelle immediate vicinanze dei bordi, per complessi fenomeni fisici. Ma poi, è ovvio che tale formula non avrà senso per altezze maggiori del diametro dell'universo. E anche su scale umane, la relazione non vale molto bene se uno dei due vasi è "molto" sottile, per il (complesso) fenomeno fisico della capillarità: via via che uno dei due vasi si assottigliasse, la formula perderebbe precisione. Eppure, la Legge dei Vasi Comunicanti rimane del massimo valore, teorico e applicativo, e il fenomeno è modellizzato proprio dalla semplice equazione sopra scritta. Con tutte le cautele del caso, quindi.

32 Matematica col Computer – Lezione B

Lezione esclusa dal corso breve – 40 ore.

32.1 Ordinare dati con LibreOffice

Nella colonna abbiamo questi nomi (sequenze alfanumeriche):

Pitagora
Gauss
Pascal
Lagrange
Cauchy
Fermat
Tartaglia

Per ordinare questi nomi alfabeticamente, selezioniamo la colonna A ciccando sulla lettera A, dal menu Dati selezioniamo

Ordina in modo crescente

per ordinarli dalla A alla Z e decrescente per ordinarli dalla Z alla A.
Si troveranno così i dati riordinati in questo modo:

Cauchy
Fermat
Gauss
Lagrange
Pascal
Pitagora
Tartaglia

Si può procedere analogamente per ordinare in modo crescente o decrescente una lista di numeri.

BOZZA - DRAFT

32.2 Visualizzazione grafica con Calc di LibreOffice

Vedremo una serie di esempi che giungeranno a produrre le 3 figure di questo paragrafo:

– un grafico più o meno “tradizionale"
– un bar chart
– un diagramma a bolle.

Aprire LibreOffice

Supponiamo che sia impostato in lingua italiana.

Nella colonna “Crea" selezionare Foglio elettronico Calc.

Nella colonna ci sarà una lista di valori di ascisse che in questo esempio produrremo in questo modo:

Nella cella A1 scriviamo 0

Nella cella A2 scriviamo 0,2 (con la virgola decimale perché impostato in lingua italiana, in inglese sarebbe 0.2)

Una volta inserito il secondo valore, selezionare entrambe le celle e dall’angolo in basso a destra della cella A2 trascinare verso il basso per generare la lista di numeri e supponiamo di estendere dalla riga 1 alla riga 33, nella quale comparirà il numero 6,4.

Supponiamo di voler graficare la funzione:

$$10\,\sin(x) + 2\,\sin(x\,\sqrt{8}) \qquad 0 \le x \le 6{,}4$$

e nella cella B1 scriviamo:

=10*SEN(A1)+2*SEN(A1*RADQ(8))

con

BOZZA - DRAFT

– un segno = all'inizio per definire la funzione
– SEN per il seno
– RADQ per la radice quadrata
– asterischi per le moltiplicazioni (non possono essere implicite).

I valori delle ascisse, diciamo pure x, verranno presi dalla colonna A, e nella colonna B verranno calcolate le corrispondenti ordinate, diciamo pure y.

Per farlo basta trascinare l'angolo in basso a destra (come avevamo fatto per la colonna A)fino alla riga 33 dove verra calcolato il valore –0,1943...

(*) (riprenderemo da questo punto un'altra cosa in seguito)

BOZZA - DRAFT

Per creare il grafico selezioniamo entrambe le colonne e poi dal menu Inserisci ... Grafico (sulla stessa barra orizzontale dove si trova il menu File), come tipo di grafico selezioniamo XY (dispersione) e poi Punti e linee (il secondo tipo di grafico).

Il software genererà un'anteprima sul foglio di calcolo, e se ci va bene clicchiamo Fine.

Il grafico generato può essere modificato a nostro piacimento, per esempio modificando colori delle linee, aggiungendo un titolo (possiamo ad esempio inserire l'equazione da cui siamo partiti).

Per farlo clicchiamo con il pulsante destro sul grafico Modifica e poi Inserisci titoli.

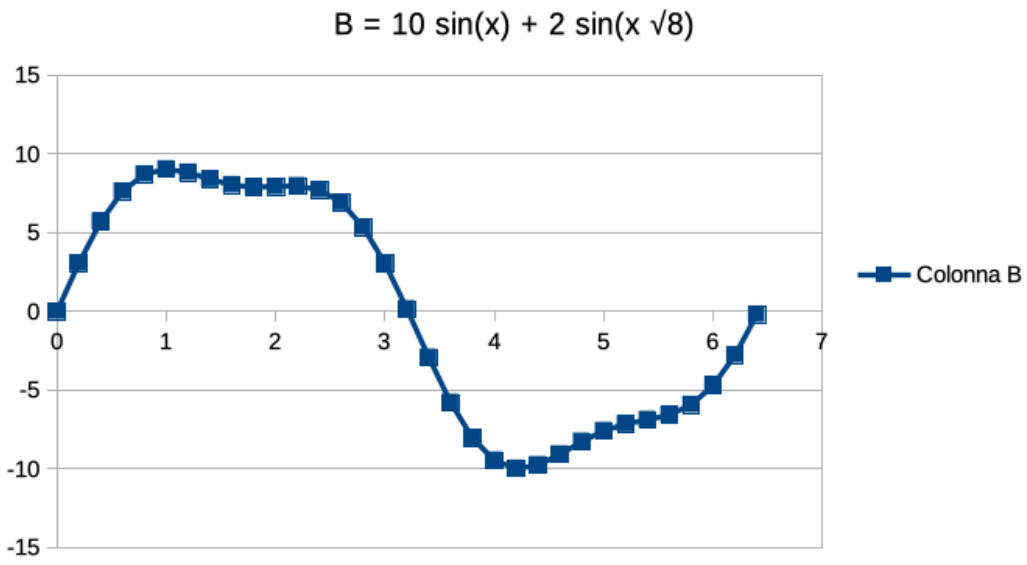
Possiamo inserire come titolo l'espressione analitica della funzione graficata:

B = 10 sin(x) + 2 sin(x $\sqrt{}$ 8)

e il simbolo della radice si trova in rete cercando “unicode symbol square root", ma ovviamente verrà scritto $\surd 8$ e non $\sqrt{8}$.

Una volta ottenuto il grafico desiderato, lo possiamo esportare come immagine per inserirlo, ad esempio, in un nostro scritto: cliccare con il tasto sinistro sul grafico per selezionarlo, poi con il destro selezioniamo Esporta come immagine, diamo per esempio il nome Somma-trigonom e selezioniamo la cartella in cui salvare l'immagine.

Nella figura il risultato finale.



Mostriamo adesso come rappresentare gli stessi dati con un bar chart.

Riprendendo da (*) selezioniamo i dati delle due colonne e dal menu Inserisci ... Grafico selezioniamo Grafico a colonne

Clicchiamo Successivo e poi su “Prima colonna come didascalia"

Adesso mostriamo anche come dare il nome agli assi (e sceglieremo x e y) e come rimuovere la scritta Colonna B, tutte cose che

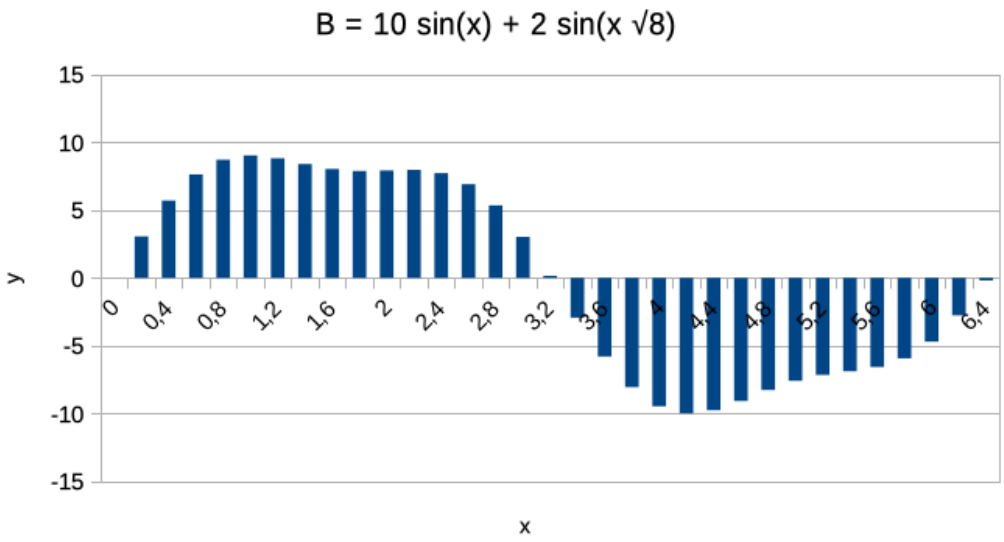
comunque si potevano fare anche nel grafico precedente.

Clicchiamo ancora su Successivo e poi un'ultima volta su Successivo. Inseriamo il titolo del grafico e i nomi degli assi. Concludiamo cliccando su Fine.

Per rimuovere la scritta Colonna B (che prima abbiamo tollerato e adesso impariamo ad eliminare) selezioniamo il grafico cliccando due volte con il pulsante sinistro sul grafico e infine un'altra volta su Colonna B per selezionare questo elemento del grafico e cancelliamola con il tasto Canc, Del o Backspace (il tasto può variare leggermente in base al tipo di tastiera).

DRAFT

Nella figura il risultato finale.



Vedremo adesso un diagramma a bolle, con 11 terne di numeri che produrremo in questo modo:

In un nuovo foglio, inseriamo nella cella A1 la funzione:

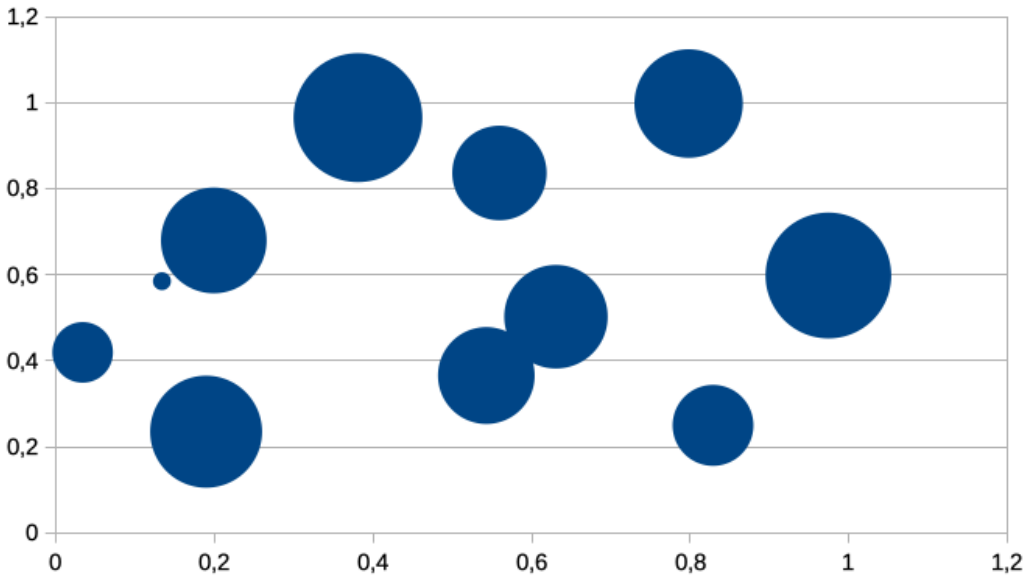
=CASUALE.FISSO()

che genera un numero casuale fra 0 e 1 e dopo averlo generato non lo modifica mentre faremo nuove operazioni sul foglio di calcolo.

La estendiamo prima verso destra fino alla cella A3 e poi verso il basso fino alla cella C11.
Otteniamo così 3 colonne (liste) di 11 numeri casuali, ovvero 11 terne di numeri e ogni terna verra rappresentata da una bolla con l'ascissa nel primo elemento della terna (colonna A), con l'ordinata nel secondo elemento della terna (colonna B) e area (non raggio!) pari al terzo elemento della terna (colonna C).
Creiamo il grafico a bolle dal menu Inserisci ... Grafico, poi clicchiamo su Fine.
Si può ottenere un'immagine simile a questa, ma si consideri che quando il lettore riprodurrà il procedimento otterrà in generale un diagramma a bolle molto diverso per la casualità dei dati.

BOZZA - DRAFT

Nella figura il risultato finale.



Come nei casi precedenti, possiamo esportare come immagine il grafico e inserirla in un nostro scritto.

32.3 La successione di Fibonacci con LibreOffice

Seguiremo lo standard di WolframAlpha:

$x_0 = 0$

(Fibonacci(0), su WolframAlpha, con F maiuscola o minuscola, parentesi tonde o quadre)

$x_1 = 1$ (Fibonacci(1))

$x_n = x_{n-1} + x_{n-2}$

che dà la successione

0, 1, 1, 2, 3, 5, 8, 13...

e ricordiamo che seguendo altri standard si potrebbe invece iniziare da 1, 1... e ancora si potrebbe chiamare il primo valore x_0 oppure x_1.

Aprire LibreOffice

Supponiamo che sia impostato in lingua italiana.

Nella colonna “Crea" selezionare Foglio elettronico Calc.

Nella cella A1 scriviamo 0

Nella cella A2 scriviamo 1

Nella cella A3 inseriamo la formula:
=A1+A2

Una volta inserito il secondo valore, dall’angolo in basso a destra della cella A2 trascinare verso il basso fino alla cella A21 per generare la successione.
Se avete fatto tutto bene troverete che

$$x_{20} = 6765$$

che appare nella cella A21 e ci è confermato da WolframAlpha
https://www.wolframalpha.com/input?i=Fibonacci%2820%29

BOZZA DRAFT

32.4 Medie varie con LibreOffice

Mostreremo con i dati prodotti nel paragrafo precedente come fare medie con LibreOffice.

Scriviamo, solo per chiarezza e non è necessario,
nella cella C1 media aritm.
nella cella C2 media geom.
nella cella C3 mediana.

Media aritmetica dei valori $x_1...x_{20}$ della successione di Fibonacci che sono nelle celle A2...A21 e che poniamo nella cella D1:
=MEDIA(A2:A21)
Se avete fatto tutto bene trovate 885,5.

Media geometrica dei valori $x_1...x_{20}$ della successione di Fibonacci che sono nelle celle A2...A21 e che poniamo nella cella D1:
=MEDIA.GEOMETRICA(A2:A21)
Se avete fatto tutto bene trovate 70,684... (con molti decimali qua omessi)

Mediana dei valori $x_1...x_{20}$ della successione di Fibonacci che sono nelle celle A2...A21 e che poniamo nella cella D1:
=MEDIANA(A2:A21)
Se avete fatto tutto bene trovate 72 che è la media aritmetica fra i valori intermedi 55 e 89.

Si osserva la somiglianza della mediana con la media geometrica e questo fenomeno si presenta non solo considerando i primi 20 termini, ma una qualunque quantità di essi e più in generale questo vale per ogni successione che cresca più o meno esponenzialmente a partire dal valore 1.

Una tavola di valori dell'ISTAT che distingue media e mediana per i redditi degli Italiani si trova – nel (sotto)foglio PROSPETTO 3 – in:

BOZZA - DRAFT

https://www.istat.it/it/archivio/275630
ed è in formato .xlsx tipico di Excel e direttamente utilizzabile in LibreOffice che comunque di suo produce principalmente un formato .ods.

BOZZA - DRAFT

PARTE B – MATEMATICHE DELL'INCERTEZZA

BOZZA - DRAFT

Sezione B1 – Calcolo delle probabilità

BOZZA - DRAFT

VII – Probabilità assiomatica ed elementare

BOZZA - DRAFT

33 Introduzione al Calcolo delle Probabilità

33.1 Inquadramento della questione

Il calcolo delle probabilità è una branca della matematica.
Come tale tratta oggetti astratti, ma, a differenza di altre branche della matematica, gli enti che manipola sono in generale molto vicini o collegati alla realtà sensibile: in particolare troverete affermazioni su dadi e monete, che sono modelli "ideali" , "puliti", perfetti, per poi applicare le tecniche a fenomeni della realtà sensibile, in particolare sanitari e farmaceutici, inevitabilmente molto meno "ideali" e "puliti", come come migliora/peggiora.

BOZZA DRAFT

Sebbene comprenda molti sottocapitoli, il suo prodotto principale è un numero che è *la probabilità di un evento*, come "la probabilità che la somma dei punti di 2 dadi sia 7 è $\frac{1}{6}$", che di per sè potrebbe voler dire poco ma può diventare preziosissimo per *scegliere* fra diverse alternative possibili, confrontandole numericamente; per esempio la probabilità che la somma (dei punti) di 2 dadi sia 8 è $\frac{5}{36}$ che è meno di $\frac{1}{6}$. Non solo per una scommessa sui dadi fra amici, ma anche per delicate scelte mediche o economiche o personali.
Cosa daranno i dadi è incerto ma che convenga scommettere sul 7 piuttosto che sull'8 è certissimo.
Pensate allora al confronto fra 2 terapie farmacologiche.

Gli enti teorici di base sono gli eventi e la loro probabilità, definita in 4 modi:

- Concezione frequentista della probabilità
- Concezione classica della probabilità
- Concezione soggettiva della probabilità
- Concezione assiomatica della probabilità

Metodi del calcolo delle probabilità sono l'algebra e la geometria analitica, le funzioni elementari, il calcolo combinatorio, il calcolo infinitesimale.

33.2 La concezione frequentista della probabilità

La concezione frequentista della probabilità è intuitiva e la vediamo con un esempio. Un farmaco è stato somministrato a 1000 persone con una certa diagnosi di malattia, e dopo 5 anni sono vive 700. Per l'uniformità delle condizioni e l'alto numero di *prove* si tende a ritenere che se quello stesso farmaco verrà somministrato ad altre 100 persone con quella stessa diagnosi di malattia, circa 70 saranno vive dopo 5 anni, ovvero, detto altrimenti, che somministrato a 1 persona c'è il 70% di *probabilità* che sia viva dopo 5 anni (ovviamente 70% cioè 0.7 è $\frac{700}{1000}$).
Se per quella stessa diagnosi di malattia un altro farmaco è stato somministrato a 2000 persone e dopo 5 anni sono vive 1200, si tenderà a ritenere che se quello stesso farmaco verrà somministrato ad altre 100 persone con quella stessa diagnosi di malattia, circa 60 saranno vive dopo 5 anni, ovvero, detto altrimenti, che somministrato a 1 persona c'è il 60% di *probabilità* che sia viva dopo 5 anni (ovviamente 60% cioè 0.6 è $\frac{1200}{2000}$), meno del primo farmaco.
Se per quella stessa diagnosi di malattia un altro farmaco è stato somministrato a 4 persone e dopo 5 anni sono vive 3, ne verrebbe una probabilità di sopravvivenza a 5 anni del 75%, cioè 3 su 4, cioè perfino meglio del primo farmaco, che però è stato provato su 1000 persone, mentre questo solo su 4; e qui siamo giunti al limite della validità di questa concezione, se non vengono fatti ulteriori approfondimenti.
Le probabilità così ottenute si chiamano *probabilità a posteriori*.

In pratica, detto semplificatamente, se finora è andata in un certo modo, con una certa quota di "successi" (ma si faccia attenzione che il successo può esere non solo la guarigione ma anche la morte o quant'altro, dipende da come si definisce), riteniamo che potrà continuare ad andare nello stesso modo, "proporzionalmente", e ciò definisce la probabilità (in senso) frequentista.

BOZZA - DRAFT

33.3 La mortalità

Semplificando, il numero, definito per uno stato o una regione geografica, relativamente a 1 anno,

$$m := \frac{\#\, morti}{\#\, abitanti}$$

si chiama mortalità. (E certo, qua semplificatamente consideriamo costante il numero di abitanti nell'anno).
Ultrasemplificatamente, nel terzo millennio è intorno all'1% in molti Stati. Ovvero intorno al 10 per mille.

Riferito al passato, può essere visto nel contesto della Statistica Descrittiva.
Ma in situazioni relativamente stabili – senza guerre o epidemie oltremodo devastanti – può essere visto nell'ottica delle probabilità soggettiva e frequentista:

$$P(\text{un soggetto preso a caso muore entro 1 anno}) = m$$

(E certo un pessimista, nella sua valutazione soggettiva di probabilità, potrà dare valori anche maggiori).

BOZZA - DRAFT

Naturalmente si possono considerare anche mortalità riferite a una singola causa di morte. Per esempio col covid-19 nei 32 mesi e mezzo dall'inizio (fine febbraio 2020) fino a oggi, metà novembre 2022, 3 per mille: circa 180mila morti su circa 60 milioni di abitanti; nel 2020, l'anno peggiore, quello dell'inizio del covid in Italia, 1.3 per mille, e poi meno.

E naturalmente si possono considerare mortalità riferite a fasce d'età e genere. Negli USA molte statistiche sanitarie sono fatte anche per gruppo razziale.

In https://en.wikipedia.org/wiki/List_of_countries_by_mortality_rate si trova (15 novembre 2022) una lista dei tassi di mortalità per

mille dei vari stati nella stima della CIA relativa al 2020. Ecco 3 stati/regioni, che sorprenderanno il lettore poco addentro alla Demografia, che sostanzialmente può essere considerata una branca della Statistica Descrittiva:

Italia 10.70

India 7.30

Uganda 5.30.

Si noti che, alquanto controintuitivamente, un italiano preso a caso aveva nel 2020 probabilità di morire entro un anno doppia di un ugandese preso a caso. Diventa meno sorprendente leggendo in https://en.wikipedia.org/wiki/List_of_countries_by_median_age la stima della stessa CIA per lo stesso anno 2020 delle età mediane, 46.5 anni in Italia e 15.7 in Uganda. (Eppure in Niger l'età mediana è simile all'Uganda, anzi perfino minore, 14.8, e la mortalità è simile all'Italia, 10.20).

Umanità tutta: età mediana 31.0 anni, mortalità 7.70, circa 0.8%.

BOZZA - DRAFT

33.4 La concezione classica della probabità

La concezione classica della probabilità si basa su

un concetto primitivo che bisogna supporre noto: l'*equiprobabilità*;

e

questa definizione, valida solo per i casi equiprobabili:

$$\text{probabilità} := \frac{numero\,dei\,casi\,favorevoli}{numero\,dei\,casi\,possibili}$$

Per esempio i 6 risultati del dado (regolare, non truccato) li riteniamo equiprobabili e allora la probabilità del risultato pari è

$$P(pari) = \frac{\#\{2,4,6\}}{\#\{1,2,3,4,5,6\}} = \frac{3}{6} =$$

$$= \frac{1}{2} = 0.5 = 50\%$$

Come ulteriore esempio, la probabilità che la somma dei punti di 2 dadi sia maggiore di 10 è $\frac{3}{36}$ cioè $\frac{1}{12}$ perché dei 36 casi possibili equiprobabili $(1,1)$, $(1,2)$, $(2,1)$,...,

$(6,6)$ sono 3 quelli favorevoli all'evento considerato (somma maggiore di 10) e cioè $(5,6)$, $(6,5)$, $(6,6)$.

ESERCIZIO$_{\mu 2018}$
$*\approx\%$ Che probabilità c'è di ottenere un numero primo come somma dei risultati di 2 dadi regolari?

SVOLGIMENTO

Dei 36 casi equiprobabili, qua indicati con ovvio significato dei simboli,

```
1+1 1+2 1+3 1+4 1+5 1+6
2+1 2+2 2+3 2+4 2+5 2+6
3+1 3+2 3+3 3+4 3+5 3+6
4+1 4+2 4+3 4+4 4+5 4+6
5+1 5+2 5+3 5+4 5+5 5+6
6+1 6+2 6+3 6+4 6+5 6+6
```

con somme rispettive

```
02 03 04 05 06 07
03 04 05 06 07 08
04 05 06 07 08 09
05 06 07 08 09 10
06 07 08 09 10 11
07 08 09 10 11 12
```

(alcuni zeri sono scritti solo per buon allineamento) sono primi (P: 2, 3 5, 7, 11) e non primi (n: 4, 6, 8, 9, 10, 12) rispettivamente

```
P P n P n P
P n P n P n
n P n P n n
P n P n n n
n P n n n P
P n n n P n
```

avendosi così 15 casi favorevoli su 36 casi equiprobabili e allora

$$p = \frac{\#\,casi\ favorevoli}{\#\,casi\,possibili\,equiprobabili} = \frac{15}{36} = \frac{5}{12}$$

e in definitiva

$$\boxed{\tfrac{5}{12} \approx 0.417 = 41.7\%}$$

oppure, con un decimale in più,

$$\boxed{\tfrac{5}{12} \approx 0.4167 = 41.67\%}$$

BOZZA - DRAFT

(e se ne potrebbero mettere anche di più, se fosse utile o desiderato).

ESERCIZIO$_{\mu 2018}$
$*$ ≈ % Che probabilità c'è di ottenere un numero primo come somma dei risultati di 2 dadi regolari a 8 facce numerate da 1 a 8? (Hanno la forma di ottaedro regolare).

SVOLGIMENTO

Dei 64 casi equiprobabili, qua indicati con ovvio significato dei simboli,

```
1+1 1+2 1+3 1+4 1+5 1+6 1+7 1+8
2+1 2+2 2+3 2+4 2+5 2+6 2+7 2+8
3+1 3+2 3+3 3+4 3+5 3+6 3+7 3+8
4+1 4+2 4+3 4+4 4+5 4+6 4+7 4+8
5+1 5+2 5+3 5+4 5+5 5+6 5+7 5+8
6+1 6+2 6+3 6+4 6+5 6+6 6+7 6+8
7+1 7+2 7+3 7+4 7+5 7+6 7+7 7+8
8+1 8+2 8+3 8+4 8+5 7+6 8+7 8+8
```

con somme rispettive

```
02 03 04 05 06 07 08 09
03 04 05 06 07 08 09 10
04 05 06 07 08 09 10 11
05 06 07 08 09 10 11 12
06 07 08 09 10 11 12 13
07 08 09 10 11 12 13 14
08 09 10 11 12 13 14 15
09 10 11 12 13 14 15 16
```

(alcuni zeri sono scritti solo per buon allineamento) sono primi (P: 2, 3, 5, 7, 11, 13) e non primi (n: 4, 6, 8, 9, 10, 12, 14, 15, 16) rispettivamente

```
P P n P n P n n
P n P n P n n n
n P n P n n n P
P n P n n n P n
n P n n n P n P
P n n n P n P n
n n n P n P n n
n n P n P n n n
```

avendosi così 23 casi favorevoli su 64 casi equiprobabili e allora

$$p = \frac{\#\,casi\ favorevoli}{\#\,casi\ possibili\ equiprobabili} =$$

$$\boxed{\frac{23}{64} \approx 0.359 = 35.9\%}$$

BOZZA - DRAFT

oppure, con un decimale in più,

$$\boxed{\tfrac{23}{64} \approx 0.3594 = 35.94\%}$$

(e se ne potrebbero mettere anche di più, se fosse utile o desiderato).

33.5 Concezione soggettiva della probabilità

La definizione soggettiva della probabilità (di Leonard Jimmie Savage e Bruno[143] de Finetti) la diamo implicitamente attraverso questo esempio. Supponiamo che io possa scommettere sulla vittoria della squadra dei Vispi Volpini della prima partita del campionato e abbia queste idee:

Metto 100 euro e se vince prendo 300 euro: accetto felicissimo
Metto 100 euro e se vince prendo 200 euro: accetto contento
Metto 100 euro e se vince prendo 126 euro: per un pelo accetto
Metto 100 euro e se vince prendo 125 euro: sono indifferente
Metto 100 euro e se vince prendo 124 euro: per un pelo ma rifiuto
Metto 100 euro e se vince prendo 110 euro: rifiuto
Metto 100 euro e se vince prendo 100 euro: ci mancherebbe!

BOZZA - DRAFT

Questo significa che io ritengo che la probabilità di vittoria della squadra è $\frac{4}{5} = 0.8 = 80\%$. In pratica nell'ipotesi intermedia in cui non ho nè interesse a scommettere nè a non scommettere ritengo che se l'evento si verificasse 5 volte, 4 volte la squadra vincerebbe e io avrei preso $4 \cdot 125$ euro cioè 500 euro, esattamente quanto avrei speso nelle 5 scommesse. È ovvio che a 126 euro considero vantaggiosa la scommessa (seppure di poco).

Il pareggiarsi della spesa con l'ipotetica vincita definisce la *probabilità soggettiva* che io attribuisco al verificarsi dell'evento considerato.

[143] Attuario presso le Assicurazioni Generali, lungamente docente all'Università di Trieste, fu anche uno degli ideatori del codice fiscale.

La formula è

$$p = \frac{costo\,della\,scommessa}{vincita\,nel\,caso\,indiffferente}\,.$$

Nell'esempio $\frac{100}{125}$.

Con questa definizione, un esperto attuario può fissare il premio assicurativo per un evento per il quale non sia disponibile una casistica significativa, come l'immissione sul mercato di un farmaco da parte di una nuova azienda, o un viaggio su Marte; o esiste una casistica assolutamente non omogenea.

Il ricorso alla probabilità soggettiva è frequentissimo in Farmacia e Medicina.

Esempio 1.

Leggiamo[144] per esempio sul sito dell'ECDC, European Centre for Disease Prevention and Control:

On the basis of the information currently available, ECDC considers that:

- the likelihood of infection for EU/EEA citizens residing in or visiting Hubei province is estimated to be high;
- the likelihood of infection for EU/EEA citizens in other Chinese provinces is moderate and will increase;
- there is a moderate-to-high likelihood of additional imported cases in the EU/EEA;
- the likelihood of observing further limited human-to-human transmission within the EU/EEA is estimated as very low to low if cases are detected early and appropriate infection prevention and control (IPC) practices are implemented, particularly in healthcare settings in EU/EEA countries;

Il soprastante testo non si spinge fino a dare stime numeriche dei valori di probabilità, ma molto chiaramente parla in termini di probabilità (o verosimigliana: *likelihood*) "alta", "moderata", "da-moderata-ad-alta", "da molto bassa a bassa".

[144] Letto il 4 febbraio 2020 in https://www.ecdc.europa.eu/sites/default/files/documents/novel-coronavirus-risk-assessment-china-31-january-2020_0.pdf

BOZZA - DRAFT

Le valutazioni di probabilità soggettiva prodotte da enti considerati autorevoli possono orientare significativamente i comportamenti, dei singoli e delle comunità, sia direttamente sia influenzando (e di fatto fungendo da base teorica) l'emanazione di regolamenti o leggi, a vari livelli di autorità.

Anche le norme antisismiche italiane a cui devono sottostare i nuovi edifici (coi relativi costi associati), comprese le farmacie e soprattutto gli ospedali, sono basate su valutazioni di probabilità soggettiva, che classificano il rischio sismico nelle varie zone. In particolare la "zona 1" è quella in cui *si ritiene* che probabilità che capiti un forte terremoto sia *alta* (in un prossimo futuro; idealmente, in via semplificata, si pensi a 50 anni). E poi vi sono le zone 2, 3, e 4. Con ovvie differenze nei costi di costruzione.
Mappe che le leggi possono cambiare nel tempo, in base a valutazioni soggettive mutevoli sia per nuovi dati (fra cui i terremoti) che cambiamenti delle composizioni delle commissioni scientifiche che di pensieri di singoli esperti nel corso del tempo, si veda per esempio l'articolo scientifico *Italy's new seismic hazard map is back to square one* (2023)

BOZZA DRAFT

Naturalmente nel corso della storia un'infinità di volte gli *esperti* hanno sbagliato le stime, anche in un modo clamoroso; si veda per esempio lo spassoso libro *La parola all'esperto*. Tuttavia questo fenomeno si riduce se si distinguono i *sedicenti* esperti dagli esperti deputati da organismi già di per loro, in qualche modo, di alto livello. (Eppure anche qua, c'è stato talvolta *da mettersi le mani nei capelli*.)

Se invece scendiamo al livello dei profani relativamente ad un certo argomento specialistico, in generale le loro stime di frequenza o probabilità (il che è relativamente equivalente) in generale valgono, per così dire, *un po' meno di nulla*, con usuali sovrastime di parecchi ordini di grandezza – anche per la potenza della manipolazione

mediatica – delle frequenze (e quindi delle associate probabilità) di fenomeni anche ampiamente considerati e discussi pubblicamente.

Le valutazioni di probabilità soggettiva ovviamente sono alla base della vita di tutte le persone, ma ad un livello basico, senza applicazione di formule, per esempio quando si ritiene più probabile una cosa di un'altra.

Ad un livello matematico più fine, le valutazioni di probabilità soggettiva governano molte scelte di tipo medico e sanitario.

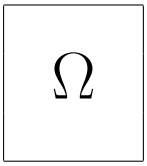
Ad un livello matematico ancora superiore, esse sono alla base di buona parte dell'economia mondiale, governando la finanza speculativa, le assicurazioni, i bonds...
Con specifico riferimento alla Medicina e alla Farmacia, si considerino in particolare gli *ebola bonds*, che traggono il nome dal morbo ebola ma possono applicarsi ad altri, che sono vere e proprie scommesse sulle epidemie. Queste scommesse in teoria sono formulate *positivamente*, perché gli investitori nazionali guadagnano se l'epidemia *non* si verifica – e i danni li pagano organismi sovranazionali come la Banca Mondiale.
Ma distorsioni sono possibili, come vediamo in questo passaggio:

> *"to do so Ebola must kill at least 20 people in at least one other country. So far, neighbouring Uganda has only had two victims. (...) Congolese health authorities have an incentive to let the virus move into Uganda in order to unlock more aid."*[145]

BOZZA - DRAFT

[145] Reuters in https://www.reuters.com/article/us-worldbank-ebola-breakingviews/breakingviews-ebola-bonds-are-wonky-way-to-tackle-pandemics-idUSKCN1V30SW, consultato il 5 marzo 2020.

34 Concezione assiomatica della probabilità



Lettera greca *omega* maiuscola, presente sulle banconote degli euro.

34.1 Le basi, gli assiomi

La concezione assiomatica della probabilità è astratta. Con definizioni e assiomi si crea la probabilità, interpretabile in modo compatibile con le altre 3 concezioni della probabilità viste:

• Ω: *evento certo*.
(Esempi: "il dado fa 1 o 2 o 3 o 4 o 5 o 6"; "sopravvive 5 anni o no").

• $\mathbb{A}$: *σ-algebra degli eventi*: tutti gli eventi *ragionevolmente* considerabili.
Tecnicamente: un sottoinsieme delle parti di Ω (cioè di $I\!P(\Omega)$) sufficientemente regolare, precisamente tale che:
  ◇ $\varnothing, \Omega \in \mathbb{A}$
  ◇ $(\forall A \in \mathbb{A})\ A^C \in \mathbb{A}$ (dove A^C è il complementare di A)
  ◇ $(\forall A_1, ..., A_n, ... \in \mathbb{A})\ \cup_{n=1}^{\infty} A_n, \cap_{n=1}^{\infty} A_n \in \mathbb{A}$.

• La *[funzione] probabilità* $P : \mathbb{A} \to [0, 1]$ tale che
  ◇ $P(\Omega) = 1$ (l'evento certo ha probabilità 100%)
  ◇ $P(A_1 \cup^* A_2 \cup^* ... \cup^* A_n \cup^* ...) = P(A_1) + ... + P(A_n) + ...$
dove $\cup^*$ indica l'unione disgiunta cioè con intersezione vuota.

• Lo *spazio di probabilità*: la terna ordinata $(\Omega, \mathbb{A}, P)$.

• Si definisce la *probabilità condizionata*

$$P(B|A) := \frac{P(A \cap B)}{P(A)} \tag{87}$$

BOZZA - DRAFT

per la quale è apparsa conveniente l'interpretazione di probabilità che si verifichi B *sapendo* che si è verificato A. Per esempio con riferimeno a un dado (regolare)

$$P(\text{“}dispari\text{”}|\text{“}primo\text{”}) = \frac{P(\text{“}dispari\text{”} \wedge \text{“}primo\text{”})}{P(\text{“}primo\text{”})} =$$

$$= \frac{P(\{2,3,5\} \cap \{1,3,5\})}{P(\{2,3,5\})} =$$

$$= \frac{P(\{3,5\})}{P(\{2,3,5\})} = \frac{\frac{2}{6}}{\frac{3}{6}} = \frac{2}{3}$$

• *L'evento impossibile*, che ha probabilità 0.

34.2 L'evento impossibile fa inghippi che fuggiremo

L'impossibilità logica è diversa dall'impossibilità probabilistica. Certi eventi di probabilità 0 *potrebbero* succedere dal punto di vista logico.

Leggiamo nel libro (Zanichelli, 2022) di Andrea Sgarro et. al. Statistica di base:

> Purtroppo se si ammettono le probabilità *estreme* zero e uno (...) il discorso diventa sottile

Non ci addentreremo nella questione. Lo scrivente garantisce personalmente al Lettore che un *prefissato* evento di probabilità 0 non capiterà comunque *mai*. Non se ne preoccupi.

Produciamo qua 2 esempi originali. Se Pierino Pazientino inizia a sommare i cubi degli interi positivi in un qualunque ordine che a lui piace

$1^3 + 1^3$

$1^3 + 2^3$

...

la probabilità che ottenga un numero cubo (come 27, o 1000) entro un anno, a qualunque velocità svolga i calcoli, è 0. Questa è un'impossibilità probabilistica e contemporaneamente logica, perché è stato dimostrato che la somma di 2 cubi interi positivi non può essere un cubo.

BOZZA - DRAFT

Invece che 2 persone che non si vedono scelgano lo stesso punto di un circolo disposto orizzontalmente e rotante – se mai fosse realizzabile tale in effetti non concretamente realizzabile esperimento – è evento possibile dal punto di vista logico ma con probabilità 0, cioè impossibile dal punto di vista probabilistico.

34.3 Mortalità e letalità di una malattia o infezione

Il concetto di probabilità condizionata è sottostante alla distinzione fra *mortalità* e *letalità* di una malattia, qua definibili ultrasemplificatamente così:

$$\text{mortalità} := \frac{morti}{abitanti} = P(morte)$$

$$\text{letalità} := \frac{morti}{malati} = P(morte\,|\,malato)$$

naturalmente in un senso frequentista (cioè in pratica contando i morti), e ovviamente (sia per la mortalità che per la letalità) bisognerà specificare un intervallo di tempo, per esempio dall'inizio alla "fine" di un'epidemia.
Per esempio leggiamo[146] sulla peste del Trecento:

> Una volta giunta in Europa, la peste nera si diffuse rapidamente perdurando tra i sei e i nove mesi nelle aree colpite. Il tasso di mortalità medio fu di circa il 30% nel totale della popolazione, mentre il tasso di letalità fu circa il 60%.

Si noti che per il covid-19 è invalso di fatto lo standard di contare in *casi*, cioè gli infetti riconosciuti, e non quelli che in senso stretto sarebbero i *malati*, cioè malati in senso clinico, cioè con sintomi: sono conteggiati i sintomatici e gli asintomatici, il che li aumenta molto di numero, però semprechè riconosciuti, il che li diminuisce di numero. Una lunga dettagliata dissertazione sulla questione si trova nella pagina web OMS in https://www.who.int/news-room/commentaries/detail/estimating-mortality-from-covid-19,

[146] Da Wikipedia, l'enciclopedia libera, alla voce *Peste nera*.

BOZZA - DRAFT

che infine conclude che... non è possibile stabilire quale sia la letalità del covid-19, più specificamente il *case fatality ratio* (CFR), per varie ragioni che elenca. Per farsi un'idea della distinzione fra IFR, *infection fatality rate*, e CFR, *case fatality rate*, si veda il testo – comunque discutibile – in https://en.wikipedia.org/wiki/Case_fatality_rate. La questione, sia teorica ovvero definizionale, che pratica ovvero numerica, è complessa, e, nel caso del covid, è stata sicuramente influenzata da ciclopici interessi economici, e altri indicibili.

Di fatto indurre nell'umanità la convinzione probabilistica che se ci si contagia molto probabilmente si muore, ha fatto accettare nel 2020 un rapidissimo mutamento nell'umanità, che ha fatto diventare i super-ricchi del pianeta ancora più spropositatamente ricchi – a danno[147] di una parte considerevole degli altri 8 miliardi di esseri umani.

Il grandissimo epidemiologo Ioannidis in un articolo[148] del maggio 2021 stimava in circa lo 0.15% l'IFR del covid-19 al febbraio 2021, "with substantial differences in IFR and in infection spread across continents, countries and locations.".

A metà novembre 2022, in Italia abbiamo quasi 180mila morti e quasi 24 milioni di casi diagnosticati, e il rapporto è 0.75%, 1 "morto covid" ogni 132 casi covid diagnosticati.

BOZZA DRAFT

34.4 La letalità apre una finestra su un baratro teorico

La posizione epistemologica della probabilità è complessa, e vi si trovano molti paradossi (seppure spesso solo apparenti, e cioè in

[147] È questione complessa ma per farsi un'idea si legga: "I miliardari a Davos potranno brindare all'incredibile impulso che le loro fortune hanno ricevuto grazie alla pandemia e all'aumento dei prezzi dei generi alimentari e dell'energia, – ha detto Gabriela Bucher, direttrice esecutiva di Oxfam International – ma allo stesso tempo decenni di progressi nella lotta alla povertà estrema rischiano di essere vanificati con milioni di persone lasciati senza mezzi per poter semplicemente sopravvivere" in https://www.oxfamitalia.org/davos-2022-crescono-le-disuguaglianze/.

[148] Ioannidis JPA. Reconciling estimates of global spread and infection fatality rates of COVID-19: An overview of systematic evaluations. Eur J Clin Invest. 2021 May;51(5):e13554. doi: 10.1111/eci.13554. Epub 2021 Apr 9. PMID: 33768536; PMCID: PMC8250317.

effetti spiegabilissimi).

Prendiamo non solo per buono ma anche valevole ancora adesso il valore – illustrato in 34.3 – di 0.75% della probabilità che ha avuto di morire ("morte covid", comunque sia stata determinata, non apriamo qua un secondo baratro) un qualunque infettato italiano diagnosticato nel periodo dal 2020 al novembre 2022, preso oggi a caso (in effetti attualmente la probabilità di morire se diagnosticati contagiati è molto minore che all'inizio della pandemia, ma ora non consideriamo questo fatto).
Se dunque un italiano oggi si ritrova diagnosticato contagiato, ha lo 0.75% di probabilità di morire "di" questa infezione?
Sì e no. Se del soggetto sappiamo solo che è italiano e che è diagnosticato contagiato, allora sì, in base a quanto sopra detto si deve concludere che ha lo 0.75% di probabilità di morirne. Ma se sappiamo che ha meno di 30 anni, dobbiamo concludere che la probabilità di morte è circa 0% (vedi figura), e se sappiamo che ha più di 90 anni la probabilità risulta circa del 13%. D'altra parte, valori ancora diversi si otterrebbero considerando il genere, la presenza di un tumore attivo, eccetera. La probabilità dipende in modo sostanziale dall'*informazione*, non è un numero che viaggia appiccicato al soggetto, oppure, se lo è, se vogliamo dire diversamente, è tanto inconoscibile quanto tutte le condizioni che potrebbero influenzarlo. Ritorneremo sulla questione con un notevole paradosso.

BOZZA DRAFT

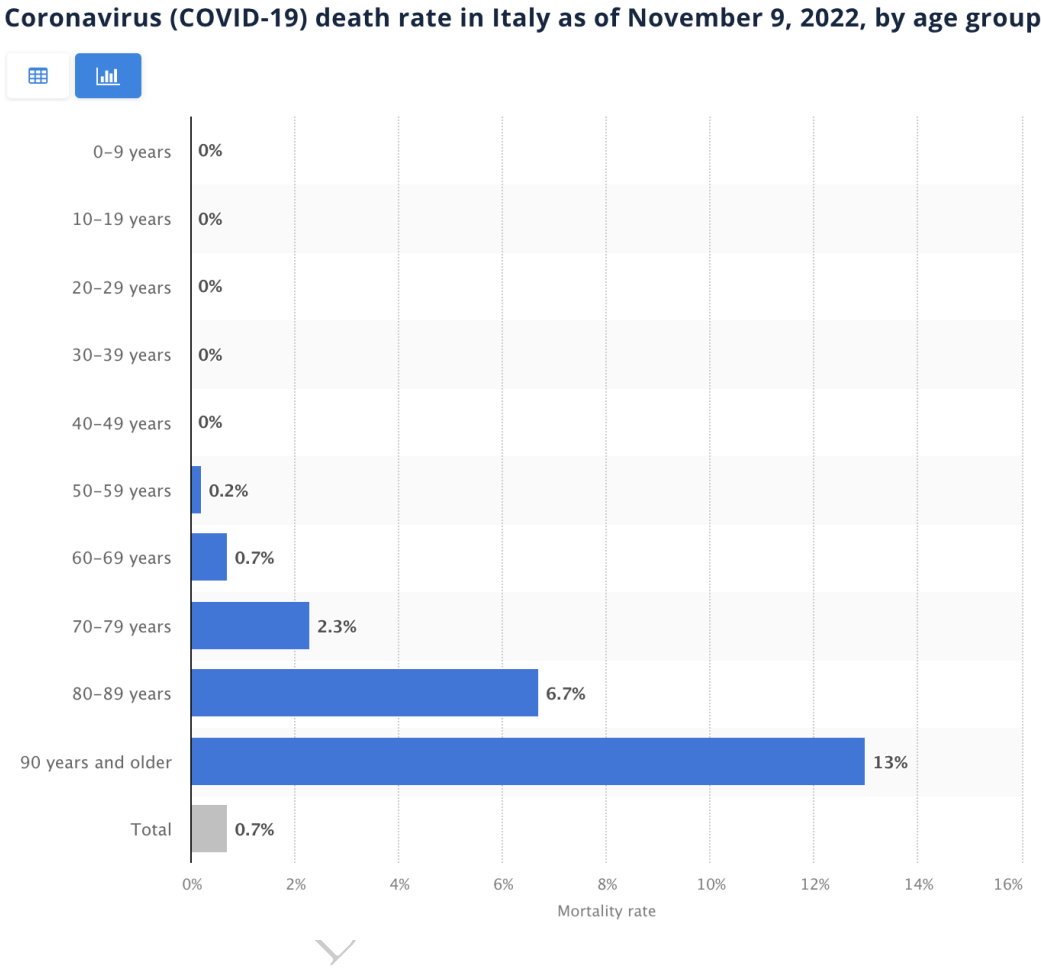


Figure 41: Letalità del covid-19 nelle classi d'età nel mondo. Screenshot da https://www.statista.com/statistics/1106372/coronavirus-death-rate-by-age-group-italy/

34.5 Indipendenza degli eventi

Nel caso che sapere che si è verificato A non muti per noi la probabilità che si verifichi B, cioè $P(B|A) = P(B)$, si ottiene, da quest'equazione e dalla (87),

$$P(A \cap B) = P(A) \cdot P(B) \qquad \text{eventi indipendenti}$$

e quest'equazione definisce assiomaticamente il caso di 2 *eventi indipendenti*. E si estende a 3 o più eventi, ma la questione è

delicata.[149]
Questa degli eventi indipendenti è in qualche modo la formula più importante del Calcolo delle Probabilità fra le non ovvie.

Per esempio

$$P(\text{“la moneta dà testa"} \wedge \text{“il dado dà 4"}) =$$

$$= P(\text{“la moneta dà testa"}) \cdot P(\text{“il dado dà 4"}) =$$

$$= \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12}\,.$$

Invece “dispari" e “primo" non sono indipendenti nel lancio di un dado: $P(dispari) = P(1,3,5) = \frac{1}{2}$ e $P(primo) = P(2,3,5) = \frac{1}{2}$ ma $P(dispari \wedge primo) = P(3,5) = \frac{2}{6} = \frac{1}{3}$ che non è $\frac{1}{2} \cdot \frac{1}{2}$.

BOZZA - DRAFT

- Si dimostra (sono teoremi) che valgono, $\forall A, B, A_1, A_2 \cdots \in \mathbb{A}$:
  - $P(A^C) = 1 - P(A)$ da cui in particolare $P(\varnothing) = 0$
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - $P(A \cup^* B) = P(A) + P(B)$ (unione disgiunta)
  - $P(A_1 \cup A_2 \cup \dots) = 1 - P(A_1^C \cap A_2^C \dots)$
  - $P(B) = P(B \cap A) + P(B \cap A^C)$
  - $A \subseteq B \Rightarrow P(A) \leq P(B)$

Per le precedenti 6 formule i disegni insiemistici chiariscono tutto.

◇ Si dimostra (teorema) la Formula di Bayes: dati un $B \in \mathbb{A}$ e una partizione di Ω, cioè degli insiemi $A_1, \dots A_n$ disgiunti con unione Ω, vale

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_{k=1}^{n} P(A_k)P(B|A_k)}$$

$$i = 1, ..., n.$$

Anche la sola uguaglianza dei denominatori è interessante:

$$P(B) = \sum_{k=1}^{n} P(A_k)P(B|A_k)$$

[149] 3 eventi possono essere a 2 a 2 indipendenti senza che siano tutti indipendenti fra loro.

e in particolare con $n = 2$, scrivendo A invece di A_1 ed essendo necessariamente $A_2 = A^C$ (avendosi una partizione) si ha la

Legge delle Alternative

$$\boxed{P(B) = P(A)P(B|A) + P(A^C)P(B|A^C)} \tag{88}$$

Si faccia un disegno rappresentativo della situazione.

Esempio sulla Legge delle Alternative.

$$P(morte) =$$
$$= P(vaccinato)P(morte|vaccinato) +$$
$$+P(non\,vaccinato)P(morte|non\,vaccinato)$$

ovviamente con riferimento a una popolazione (statistica) e un intervallo di tempo.

BOZZA - DRAFT

Per esempio uno spazio di probabilità si ottiene con un dado regolare, $\Omega = \{1, 2, 3, 4, 5, 6\}$, dove $\{1\}$,...,$\{6\}$ sono gli *eventi semplici*, possiamo fissare molto opportunamente $\mathbb{A} := \mathbb{P}(\Omega)$, e la funzione P vale costantemente $\frac{1}{6}$. Per esempio $A := \{2, 3, 5\}$ è l'evento "esce 1 o 3 o 5", ovvero "[esce] [un numero] dispari".
Si calcolino A^C, $P(A^C)$, $\{1, 2, 6\}^C =: B$, $P(A \cap B)$, $P(A \cup B)$, $P(B)$.
Poi con lo stesso Ω si trovino altri 2 spazi di probabilità.

Esistono anche spazi di probabilità infiniti, per i quali nessuna uniformità è possibile, e in cui nella realtà sensibile possiamo aspettarci che in generale tenderà a ricorrere una *distribuzione* in qualche modo *a campana* – eventualmente troncata da una parte.

Si noti l'immediata (duplice) conseguenza della (87)

$$P(A \cap B) = P(B|A)P(A) = P(A|B)P(B).$$

ESERCIZIO$_{\mu 2018}$
$*$ % Supponiamo che in una certa popolazione la mutazione RX1vis

si presenti con probabilità 25% e, indipendentemente, la mutazione RS2vol con probabilità 40%. Che probabilità c'è che un soggetto di quella popolazione abbia la RX1vis e non abbia la RS2vol?

SVOLGIMENTO

$$P((\text{RX1vis})\, et\, (non\, \text{RS2vol})) =$$

l'avere o non avere la RS2vol è indipendente dall'avere o non avere la RX1vis

$$= P(\text{RX1vis}) \cdot P(non\, \text{RS2vol}) =$$

con l'evento complementare

$$= P(\text{RX1vis}) \cdot (1 - P(\text{RS2vol})) =$$

coi dati numerici

$$= 0.25 \cdot (1 - 0.4) =$$

in definitiva

$$\boxed{0.15 = 15\%}$$

BOZZA - DRAFT

Fissiamo l'attenzione per esempio sullo *spazio di probabilità uniforme* $(\Omega, \mathbb{A}, P)$ del lancio di 1 dado, con 6 *eventi semplici* di probabilità $\frac{1}{6}$, e $64 = 2^6$ eventi nella σ-algebra $\mathbb{A}$ delle parti di Ω:

$$\Omega := \{1, 2, 3, 4, 5, 6\} \qquad P(k) = 1/6,\ k = 1, ..., 6$$

$$\mathbb{A} := \mathbb{P}(\Omega) = \{\varnothing, \{1\}, \{2\}, ...\{6\}, \{1, 2\}, ..., \{1, 2, 3, 4, 5\}, \Omega\}.$$

Per esempio $\{1, 2\}$ è l'evento "esce 1" *vel* "esce 2"; $P(1, 2) = \frac{2}{6} = \frac{1}{3}$. Naturalmente esistono spazi di probabilità finiti non uniformi, per esempio ottenuti con dadi truccati.

Oppure con riferimento a una diagnosi e/o trattamento farmacologico, e un fissato intervallo temporale (per esempio 5 anni), e comunque in un'ottica frequentista, con *scores* da 1=morte a 6=ottimamente, potremmo avere per esempio:

$3\% = P(morte)$ ...
$5\% = P(molto\,male)$ .....
$12\% = P(male)$ .....
$20\% = P(stabile)$ .....
$40\% = P(bene)$ .....
$20\% = P(ottimamente)$ .....

(Si noti la forma più o meno a campana).

Calcoliamo per esempio

$$P(\text{“}migliora\text{”}) = P(\text{“}bene\text{”} \vee \text{“}ottimamente\text{”}) =$$

eventi disgiunti, somma delle probabilità:

$$= P(\text{“}bene\text{”}) + P(\text{“}ottimamente\text{”}) =$$

$$= 40\% + 20\% = 60\%.$$

Per 2 soggetti

$$P(\text{“}entrambi\ morti\text{”}) =$$

$$= P(\text{“}morte\ del\ 1\hat{}\ soggetto\text{”} \wedge \text{“}morte\ del\ 2\hat{}\ soggetto\text{”}) =$$

e nell'ipotesi di indipendenza

$$= P(\text{“}morte\ del\ 1\hat{}\ soggetto\text{”}) \cdot P(\text{“}morte\ del\ 2\hat{}\ soggetto\text{”}) =$$

$$= 0.03 \cdot 0.03 = 0.0009 = 0.09\%$$

(Poco meno dell'1 per mille). Ma se i soggetti sono legati affettivamente, l'ipotesi di indipendenza diventa alquanto inverosimile. (La morte dell'uno in generale aumenta la probabilità di morte dell'altro a prescindere dalla terapia seguita).

Si noti che appena si esce dalla modellizzazione “perfetta" dei dadi e delle monete, si introducono inevitabilmente problematiche che poi di fatto si trasformeranno in incertezze nei risultati statistici. (Dove l'indipendenza, e la normalità gaussiana dei dati – che vedremo – sono spesso

BOZZA DRAFT

ipotesi implicite, di fatto imperfette).

Nota. Il seguente esercizio richiede una calcolatrice con 8 cifre, o una moltiplicazione con carta e penna.

ESERCIZIO$_{\mu 2018}$

$*\approx\%$ Ogni volta che una fissata persona fa una certa cosa, per esempio va in una zona infestata da insetti pericolosi, si espone ad un certo rischio sanitario, per esempio di subire una puntura d'insetto pericoloso, con probabilità 7%. Fissato questo esempio (ma ugualmente si potrebbe ragionare in un caso generale) che probabilità c'è che andando 4 volte nella zona infestata – supponendo indipendenti gli eventi – quella persona venga punta da qualche insetto pericoloso?

SVOLGIMENTO

$$P(punto\,al\,1\hat{}\,viaggio) = ... = P(punto\,al\,4\hat{}\,viaggio) = 7\% = \frac{7}{100}$$

Evento complementare:

$$P(nonpuntoal1\hat{}viaggio) = ... = P(nonpuntoal4\hat{}viaggio) = 93\% = \frac{93}{100}$$

Evento composto:

$$P(mai\,punto\,nei\,4\,viaggi) = \frac{93}{100}\cdot\frac{93}{100}\cdot\frac{93}{100}\cdot\frac{93}{100} = \left(\frac{93}{100}\right)^4$$

Evento complementare:

$$P(punto\,almeno\,1\,volta\,nei\,4\,viaggi) = 1 - \left(\frac{93}{100}\right)^4 =$$

$$= 1 - \frac{93^4}{100^4} = 1 - \frac{74\,805\,201}{100\,000\,000} = \frac{100\,000\,000 - 74\,805\,201}{100\,000\,000} =$$

$$\boxed{\frac{25194799}{100\,000\,000} \approx 0.252 = 25.2\%}$$

BOZZA - DRAFT

ESERCIZIO$_{\mu 2018}$

$*$ Ogni volta che una fissata persona fa una certa cosa, per esempio va in metropolitana, si espone ad un certo rischio sanitario, per esempio di entrare in contatto coi pidocchi, con probabilità p. Fissato questo esempio (ma ugualmente si potrebbe ragionare in un caso generale) che probablità c'è che in 10 viaggi in metropolitana – supponendo indipendenti gli eventi – quella persona sia entrata in contatto coi pidocchi?

SVOLGIMENTO

$$P(\textit{contatto nel } 1\text{^ } \textit{viaggio}) = ... = P(\textit{contatto nel } 10\text{^ } \textit{viaggio}) = p$$

Evento complementare:

$$P(\textit{noncontattonel} 1\text{^}\textit{viaggio}) = ... = P(\textit{noncontattonel} 10\text{^}\textit{viaggio}) = 1-p$$

Evento composto:

$$P(\textit{nessun contatto}) = (1-p)\cdot ... \cdot (1-p) = (1-p)^{10}$$

Evento complementare:

$$\boxed{P(\textit{almeno un contatto}) = 1-(1-p)^{10}}$$

BOZZA - DRAFT

ESERCIZIO$_{\mu 2019}$

$*$ % In via semplificata, consideriamo qua terapie che possono avere solo esito fatale (morte) o successo (non si considerano diversi gradi di successo).

Per un certo paziente si stanno ipotizzando 4 procedure:

terapia T1 e poi terapia T2

terapia T2 e poi terapia T4

terapia T3 e poi terapia T4

terapia T5.

Supponendo l'indipendenza degli eventi, trovare quale procedura conviene avendosi queste probabilità di esito fatale:

T1: 9%, T2: 12%, T3: 4%, T4: 11%, T5: 13%.

SVOLGIMENTO

Con gli eventi complementari, si hanno queste probabilità di successo:

T1: 91%, T2: 88%, T3: 96%, T4: 89%, T5: 87%.

Si ha

$$P(\textit{successo procedura } 1\text{^}) = 0.91 \cdot 0.88 = 80.08\%$$

$$P(\textit{successo procedura } 2\text{^}) = 0.88 \cdot 0.89 = 78.32\%$$

$$P(\textit{successo procedura } 3\text{^}) = 0.96 \cdot 0.89 = 85.44\%$$

e allora conviene l'ultima procedura:

BOZZA - DRAFT

35 Sensibilità, specificità, predittività, ROC

La questione fondamentale affrontata in questa Lezione è

$$P(\text{veramente malato } \textit{sapendo} \text{ test positivo})$$

nel senso della probabilità frequentista, e argomenti affini.

Familiarizziamo con la probabilità condizionata con un esercizio.

Esercizio$^f_\mu$ **risolto.** Un'urna U contiene 20 palline bianche e 40 nere, e un'urna V contiene 5 palline bianche e 6 nere. Si sceglie a caso un'urna e si estrae a caso una pallina che risulta bianca. Che probabilità c'è che sia stata scelta l'urna U?

Svolgimento$_\mu$

$U : 20b + 40n$

$V : 5b + 6n$

A_1: "scelta a caso l'urna U"

A_2: "scelta a caso l'urna V"

B: "estratta a caso una pallina bianca".

A_1 e A_2 costituiscono una partizione dell'evento certo.

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{P(A_1)P(B|A_1)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2)} =$$

$$= \frac{\frac{1}{2} \cdot \frac{20}{60}}{\frac{1}{2} \cdot \frac{20}{60} + \frac{1}{2} \cdot \frac{5}{11}} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{5}{11}} = \frac{\frac{1}{3}}{\frac{1}{3} + \frac{5}{11}} =$$

moltiplicando numeratore e denominatore per $3 \cdot 11$

$$\frac{11}{11 + 5 \cdot 3} = \frac{11}{26} \approx 0.423 = 42.3\%.$$

(Osserviamo che allora con probabilità $\approx 58\%$ era stata scelta l'urna V e questo riflette il fatto che là c'erano proporzionalmente più palline bianche: visto che è venuta una pallina bianca, più probabilmente avevamo scelto quell'urna).

Esercizio$_\mu$ Costruire e risolvere un esercizio analogo.

BOZZA - DRAFT

In questa Lezione si fa riferimento alla concezione frequentista della probabilità, che detto molto semplificatamente ci fa ritenere che se finora le cose sono andate in un certo modo in un gran numero di casi, continueranno ad andare così. Con riferimento alla Farmacologia: se 3 014 soggetti sono morti di 10 000 a 5 anni dal trattamento, riteniamo che la *probabilità* di morte sia circa del 30% (a 5 anni, con quel trattamento, in una popolazione *simile*).

35.1 Sensibilità

"Specificità alta, il malato torna a casa informato"

Traiamo questo paragrafo da Wikipedia italiana, l'enciclopedia libera.

BOZZA - DRAFT

Con il termine sensibilità, in statistica, più precisamente nel campo della epidemiologia, si indica la capacità intrinseca di un test di screening di individuare in una popolazione di riferimento i soggetti malati. Essa è data dalla proporzione dei soggetti realmente malati e positivi al test (veri positivi) rispetto all'intera popolazione dei malati.

Un test sarà tanto più sensibile quanto più bassa risulterà la quota dei falsi negativi (cioè di soggetti malati erroneamente identificati dal test come sani). Un test molto sensibile, in definitiva, ci consente di limitare la possibilità che un soggetto malato risulti negativo al test.

Supponiamo che un test di screening dia come risultato solamente due opzioni: positivo al test e negativo. Essere positivi al test equivale ad essere ammalato secondo quel test, ma indagini diagnostiche successive possono rivelare l'effettiva malattia o meno.
Allora si otterranno 4 tipologie di osservati:
Sani Negativi (veri negativi)
Sani Positivi (falsi positivi) $\leftarrow$ *risultano avere la malattia; ma non ce l'hanno...*

Malati Positivi (veri positivi)
Malati Negativi (falsi negativi) $\leftarrow$ *non scoprono la loro malattia*

Rappresentabili così in tabella:
/////// MALATI - SANI
POSITIVI Veri + Falsi +
NEGATIVI Falsi - Veri -

La sensibilità del test verrà così calcolata:

$$S = \frac{V_+}{totaleMALATI} = \frac{V_+}{V_+ + F_-}$$

ESEMPIO
25 2
4 55
Si troverà 86.2%.

BOZZA - DRAFT

Dalla wikipedia in inglese: *If the goal of the test is to identify everyone who has a condition, the number of false negatives should be low, which requires high sensitivity. That is, people who have the condition should be highly likely to be identified as such by the test. This is especially important when the consequence of failing to treat the condition are serious and/or the treatment is very effective and has minimal side effects.*

35.2 Specificità

"Specificità alta, il sano torna a casa contento"

Traiamo questo paragrafo da Wikipedia italiana, l'enciclopedia libera.

Con il termine specificità, in medicina, si indica la capacità di un

test di dare un risultato normale ("negativo") nei soggetti sani:

$$Sp = \frac{V_-}{totaleSANI} = \frac{V_-}{V_- + F_+}$$

ESEMPIO
25 2
4 55
Si troverà 96.5%.

Dalla wikipedia in inglese: *If the goal of the test is to accurately identify people who do not have the condition, the number of false positives should be very low, which requires a high specificity. That is, people who do not have the condition should be highly likely to be excluded by the test. This is especially important when people who are identified as having a condition may be subjected to more testing, expense, stigma, anxiety, etc.*

BOZZA - DRAFT

35.3 Predittività

Traiamo questo paragrafo da Wikipedia italiana, l'enciclopedia libera.

Per predittività, in medicina, si intende la probabilità che un soggetto positivo ad un test di screening sia effettivamente malato. Il Valore Predittivo Positivo, che esprime numericamente la predittività, si calcola come quota di soggetti veri positivi sul totale dei positivi (veri e falsi positivi).

$$VPP = \frac{V_+}{totalePOSITIVI} = \frac{V_+}{V_+ + F_+}$$

ESEMPIO
25 2
4 55
Si troverà 92.6%.

Vediamo ora il caso in cui la prevalenza (frequenza) della malattia è decisamente minore, aumentando ad esempio di un fattore 100 le persone sane e lasciando inalterato il numero dei malati:

25 200

4 5500

Si troverà 11.1.

Esercizio. Relativamente all'ultima tabella, si calcolino prevalenza, sensibilità e specificità.

ESERCIZIO. (Tratto da Wikipedia, l'enciclopedia libera).
$\approx$ % Per un test diagnostico in una determinata popolazione si abbia

	MALATI	SANI
POSITIVI	25	200
NEGATIVI	4	5500

Calcolare la predittività, ovvero il Valore Predittivo Positivo.

SVOLGIMENTO

Ricordando la definizione

$$\text{predittività} = \text{Valore Predittivo Positivo} = VVP =$$

$$= \frac{\text{veri positivi}}{\text{totale positivi}} = \frac{V_+}{V_+ + F_+}$$

ora abbiamo

$$VVP = \frac{25}{25 + 200} = \frac{25}{225} = \frac{1}{9} \approx$$

$$\boxed{\approx 0.111 = 11.1\%}$$

(Notiamo che il valore è piuttosto basso nonostante la specificità sia alta:

$$Sp = \frac{\text{veri negativi}}{\text{totale sani}} = \frac{V_-}{V_- + F_+} = \frac{5\,500}{5\,500 + 200} = \frac{5\,500}{5\,700} =$$

BOZZA - DRAFT

$$= \frac{55}{57} \approx 0.965 = 96.5\%$$

e questo è dovuto alla rarità della malattia nella popolazione considerata: circa una trentina di malati su poco più di 5 700 soggetti).
Nota di Wikipedia, l'enciclopedia libera:

> Ossia la probabilità che un soggetto positivo al test sia effettivamente malato è pari all'11, 1%, che equivale a dire che il soggetto ha una probabilità dell'88, 9% di essere sano nonostante il test dica il contrario.

35.4 Esempio: i test rapidi per il covid-19.

In *Diagnostic accuracy of rapid antigen tests in asymptomatic and presymptomatic close contacts of individuals with confirmed SARS-CoV-2 infection: cross sectional study*, articolo scientifico del 2021 sul British Medical Journal su due test rapidi per il covid-19 in https://www.bmj.com/content/374/bmj.n1676, leggiamo i valori che gli Autori si prefiggono di misurare:

> **Main outcome measures** Sensitivity, specificity, and positive and negative predictive values of Veritor System (Beckton Dickinson) and Biosensor (Roche Diagnostics) rapid antigen tests

BOZZA - DRAFT

35.5 Le curve ROC

Traiamo questo paragrafo da Wikipedia italiana, l'enciclopedia libera, alla voce *Receiver operating characteristic*.

Nella teoria delle decisioni, le curve ROC (Receiver Operating Characteristic, anche note come Relative Operating Characteristic[1]) sono degli schemi grafici per un classificatore binario. Lungo i due assi si possono rappresentare la sensibilità e (1-specificità), rispettivamente rappresentati da True Positive Rate (TPR, frazione di veri positivi) e False Positive Rate (FPR, frazione di falsi positivi).

In altre parole, si studiano i rapporti fra allarmi veri (hit rate) e falsi allarmi.

La curva ROC viene creata tracciando il valore del True Positive Rate (TPR, frazione di veri positivi) rispetto al False Positive Rate (FPR, frazione di falsi positivi) a varie impostazioni di soglia. Il tasso di veri positivi è anche noto come sensibilità, richiamo o probabilità di rilevazione [2]. Il tasso di falsi positivi è anche noto come fall-out o probabilità di falsi allarmi [2] e può essere calcolato come (1 - specificità). Può anche essere pensato come un diagramma della potenza in funzione dell' errore di tipo I :quando la prestazione viene calcolata da un solo campione della popolazione, può essere considerata come una stima di queste quantità. La curva ROC è quindi il tasso dei veri positivi in funzione del tasso dei falsi positivi. In generale, se sono note le distribuzioni di sensibilità e 1-specificità, la curva ROC può essere generata tracciando la funzione di distribuzione cumulativa (area sotto la distribuzione di probabilità da $-\infty$ alla soglia di discriminazione) della probabilità di rilevamento nell'asse y rispetto alla funzione di distribuzione cumulativa della probabilità di falso allarme sull'asse x.

BOZZA - DRAFT

Il ROC è anche noto come curva Receiver Operating Characteristic, poiché è un confronto tra due caratteristiche operative (TPR e FPR) al cambiare del criterio.[3]

Applicazioni

Le curve ROC furono utilizzate per la prima volta durante la seconda guerra mondiale, da alcuni ingegneri elettrotecnici che volevano individuare i nemici utilizzando il radar durante le battaglie aeree. Recentemente le curve ROC sono utilizzate in medicina,[4][5] radiologia,[6] psicologia,meteorologia [7], veterinaria[8], fisica e altri ambiti, come il machine learning ed il data mining.

Concetto basilare

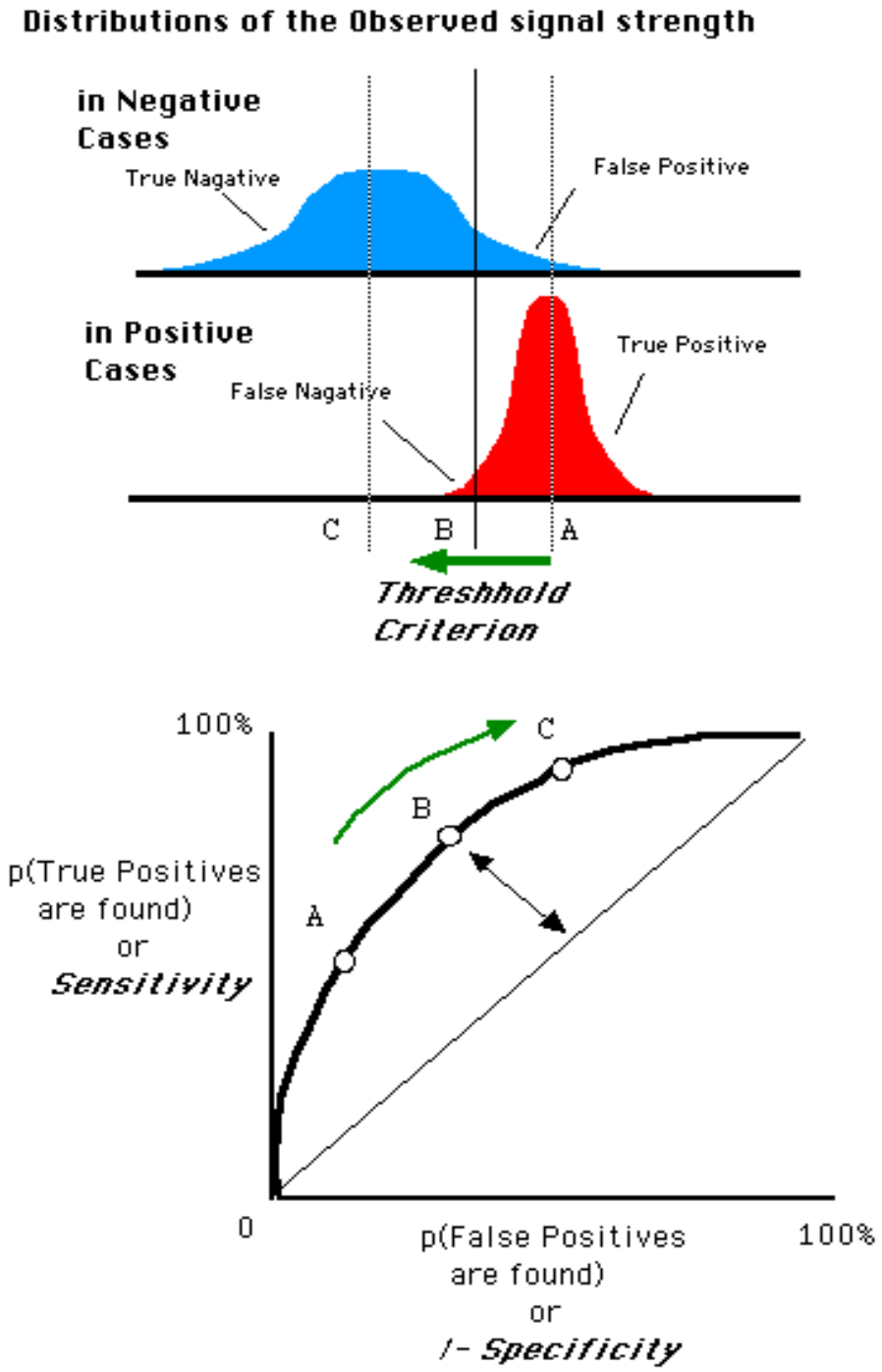


Figure 42: From https://commons.wikimedia.org/wiki/File:ROCfig.PNG. Author (assumed) NekoJaNekoJa~commonswiki

Se si considera un problema di predizione a 2 classi (classificatore binario come da figura: distribuzione rossa e azzurra), scelto un valore di soglia (threshold o cut-off), rispetto a cui decidere il risultato, ovvero se appartenente alla classe positiva (p) o negativa (n), dato che le due curve di distribuzione di probabilità risultano in parte sovrapposte, sono possibili quattro risultati a seconda della posizione del valore di cut-off:

- se il risultato della predizione è positivo p e il valore vero è anche positivo p, viene chiamato vero positivo (true positive - TP);
- se invece il valore vero è negativo, il risultato viene chiamato falso positivo (false positive - FP);
- al contrario, si ha un vero negativo (true negative - TN) quando entrambi, il risultato e il valore vero, sono negativi;
- un falso negativo (false negative - FN) invece si ha quando il risultato è negativo e il valore vero è positivo.

È inoltre possibile rappresentare questo tipo di situazione utilizzando una tabella di contingenza di tipo 2×2, dove le colonne rappresentano la distinzione tra soggetti sani e malati; le righe invece rappresentano il risultato del test sui pazienti. Un risultato qualitativo del test potrebbe essere quello di andare a valutare il numero di falsi positivi e negativi; meno ve ne saranno e maggiormente il test sarà valido.

BOZZA - DRAFT

Una curva ROC è il grafico dell'insieme delle coppie (FP, TP) al variare di un parametro del classificatore. Per esempio, in un classificatore a soglia, si calcola la frazione di veri positivi e quella di falsi positivi per ogni possibile valore della soglia; tutti i punti così ottenuti nello spazio FP-TP descrivono la curva ROC.
Attraverso l'analisi delle curve ROC si valuta la capacità del classificatore di discernere, ad esempio, tra un insieme di popolazione sana e malata, calcolando l'area sottesa alla curva ROC (Area Under Curve, AUC). Il valore di AUC, compreso tra 0 e 1, equivale infatti alla probabilità che il risultato del classificatore applicato ad un individuo estratto a caso dal gruppo dei malati sia superiore a

quello ottenuto applicandolo ad un individuo estratto a caso dal gruppo dei sani.[9]

Le curve ROC passano per i punti (0,0) e (1,1), avendo inoltre due condizioni che rappresentano due curve limite:

• una che taglia il grafico a 45°, passando per l'origine. Questa retta rappresenta il caso del classificatore casuale (linea di "nessun beneficio"), e l'area sottesa AUC è pari a 0,5.

• la seconda curva è rappresentata dal segmento che dall'origine sale al punto (0,1) e da quello che congiunge il punto (0,1) a (1,1), avendo un'area sottesa di valore pari a 1, ovvero rappresenta il classificatore perfetto.

BOZZA - DRAFT

36 Probabilità combinatoria

Lezione esclusa dal corso breve – 40 ore.

La Probabilità Combinatoria è la Probabilità Classica, quella dei casi favorevoli sui casi possibili equiprobabili, entrambi conteggiati coi metodi del Calcolo Combinatorio. Vi abbiamo già accennato nella Lezione sul Calcolo Combinatorio. Vedremo ancora qualcosa.

36.1 Dismutazioni

BOZZA_DRAFT

Si dice *dismutazione* (o *permutazione completa* o *sconvolgimento*) di un insieme ogni sua permutazione in cui tutti gli elementi cambiano posizione. (Si immaginino degli studenti alle loro sedie, si alzano e "rimescolano", si siedono, ma nessuno sulla sedia che aveva prima). Per esempio per l'insieme ordinato { 1, 2, 3, 4, 5 } ovvero per la 5-upla ordinata $(1, 2, 3, 4, 5)$, due permutazioni sono $(4, 5, 1, 2, 3)$ e $(2, 5, 1, 4, 3)$, ma solo la prima è una dismutazione. Per il numero N di dismutazione di un insieme di n elementi, è (teorema) $N \approx \frac{n!}{e}$; l'errore assoluto è sempre < 0.5; e per $n > 5$ l'errore relativo è minore dello 0.05%, che in questo caso possiamo considerare una buona approssimazione, e $\to 0$ per $n \to \infty$:

$$\lim_{n \to +\infty} \frac{N}{n!/e} = 1 \quad \text{ovvero } N \simeq \frac{n!}{\mathrm{e}}$$

in cui $\simeq$ è il simbolo ISO per l'*approssimazione asintotica*.

Esempio. Si abbiano molte persone, e altrettanti foglietti coi loro nomi. Li distribuiamo a caso a quelle persone. Che probabilità c'è che nessuno riceva il foglietto col suo nome? E che qualcuno lo riceva? Ovviamente se non sappiamo il numero n di persone, non possiamo dare una risposta esatta. Tuttavia se, come d'usuale, ci accontentiamo di un valore ben approssimato della probabilità, in pratica possiamo, purché n, che è stato garantito grande, sia ≥ 5 o meglio > 5, e sorprendentemente la risposta non dipende da quel

numero esatto di persone. Consideriamo l'evento che nessuno riceva il foglietto con il suo nome. Così per il primo quesito i casi possibili sono le permutazioni e i casi favorevoli sono le dismutazioni, e allora

$$p \approx \frac{n!/e}{n!} = \frac{1}{e} \approx 0.368 = 36.8\% \quad (\textit{viene questo } \forall n{>}5, \textit{ e per } n{=}5 \textit{ viene } {\approx}0.367).$$

E allora la probabilità che qualcuno riceva il foglietto col suo nome (evento complementare) è $\approx 0.632 = 63.2\%$.

Si potrebbe facilmente calcolare che se in una classe di studenti universitari (di più di 5 elementi) diamo 2 foglietti a testa, uno per numero di matricola e uno per nome e cognome, supponendo che non ci siano omonimie, e poi ridistribuiamo a caso, dando a ciascuno un foglietto di nome e cognome e un foglietto di numero di matricola, c'è una probabilità di

$$\approx 1 - \mathrm{e}^{-2} \approx 86\%$$

che almeno uno riceva almeno il suo nome o la sua matricola.

BOZZA - DRAFT

Esercizio. Elencare le dismutazioni di 2, 3 e 4 elementi.

Nota. Come già detto nella Lezione sul Calcolo Combinatorio, numeri enormi li lasceremo indicati, per esempio scriveremo

$$p = 1 - \frac{365!}{365^{23}\,342!}$$

36.2 I compleanni coincidenti

Con un computer od online con WolframAlpha potremmo trovare il valore approssimato del sopradetto numero $\approx 0.507 = 50.7\%$ ed è la probabilità che fra 23 persone a caso, nate in anni non bisestili, almeno 2 abbiano lo stesso compleanno (approssimando come uniforme la distribuzione delle nascite nell'anno).
Si noti come i fenomeni probabilistici possono essere molto controintuitivi.

Ancora, si noti che per vari motivi le nascite non sono uniformemente distribuite nei vari giorni dell'anno, e si potrebbe dimostrare che questo aumenta la probabilità dei compleanni coincidenti. La non uniformità varia nelle regioni geografiche e nel tempo.

Esempio sulle combinazioni semplici. Che probabilità c'è di vincere giocando una cinquina su una ruota del lotto? Per essere sicuri di vincere giochiamo 1 euro su ciascuna di esse. Quanto guadagniamo?
Di casi favorevoli ce n'è 1, e i casi possibili sono le combinazioni semplici di 90 oggetti a 5 a 5, che sono in numero di $\binom{90}{5}$, e allora la probabilità è

$$p = \frac{1}{\binom{90}{5}} = \frac{1}{\frac{90!}{5!(90-5)!}} = \frac{5!\,85!}{90!} = \frac{1\cdot\ldots\cdot 5\cdot 1\cdot\ldots\cdot 85}{1\cdot\ldots\cdot 90} =$$

$$= \frac{1\cdot 2\cdot 3\cdot 4\cdot 5}{86\cdot 87\cdot 88\cdot 89\cdot 90} = \frac{1}{86\cdot 87\cdot 22\cdot 89\cdot 3} = \frac{1}{43\,949\,268}\,.$$

Abbiamo speso 43 949 268 euro e 1 cinquina da noi giocata vince; poiché si vince 6 milioni di volte la posta, vinciamo 6 milioni di euro, con un guadagno di −37 949 268 euro. (Ciclopica perdita) ☹

BOZZA - DRAFT

Esercizi.
Che probabilità c'è che un numero < 10 sia primo? E < 25? E nei 2 casi, che sia quadrato? Triangolare? Pari?

36.3 Il gioco d'azzardo: patologia, probabilità e storia

Nel 2017 gli italiani hanno speso 55 miliardi per beni durevoli, come automobili ed elettrodomestici, e 88 miliardi in giochi d'azzardo, cioè mediamente circa 1470 euro a testa, diventati 1600 nel 2016; in parte persi, ovvio, mica organizzano i giochi solo per farci divertire; si disse che il lotto è una *tassa per gli stolti*. Bisogna comunque dire che la cinquina del lotto è un caso particolare, gli altri premi del lotto non sono così non equi; e altri giochi sono ancor meno non equi, dando premi più equilibratamente; e in alcuni conta anche

l'abilità, per esempio la schedina del calcio.
È facile verificare che in generale i testi in rete sul gioco d'azzardo tendono a drammatizzare la situazione, parlando del centinaio di miliardi circa giocati ogni anno; molto meno spazio viene dato invece al fatto che la maggior parte di quei soldi ritorna i giocatori nelle vincite:

nel 2017 giocati 102 miliardi
vinti 82 miliardi

e allora perdita netta per i giocatori 20 miliardi. Dividendo per 60milioni di italiani, viene una perdita annua pro capite di circa 333 euro. Circa un euro al giorno, con differenze enormi fra regioni, e fra singoli.
Inoltre c'è da dire che coi giochi d'azzardo lo Stato raccoglie cifre enormi senza suscitare avversione: rinunciando ai giochi lo Stato potrebbe dover esigere quei soldi forzatamente. Circa 8 miliardi nel 2017. Una cosa triste è che a pagare sono in larga misura – difficile da quantificare nei dettagli – i meno abbienti. Altra cosa triste è che per molti diventa una malattia, la **ludopatia**, ma questa può esistere anche indipendentemente dai giochi organizzati dallo Stato, giocando fra privati. Altra questione ancora, è che le sale slot offrono alla malavita un facile modo di riciclare il denaro sporco: incassano denaro pulito e per le vincite danno quello proveniente da illeciti, che si disperde nei mille rivi delle spese personali della gente. Tutto questo, ovviamente, fermo restando che di per sè non è immorale giocare d'azzardo, nè fra amici nè con lo Stato, e può essere una cosa sana, se esercitata con l'opportuna misura. Certo, 333 euro l'anno a testa, persi, danno da pensare. Si potrebbero fare anche molte altre cose con tanti soldi. In un decennio recente, le spese per la cura dei denti sono dimezzate.

> "Millionaires should always gamble, poor men never" (J.M. Keynes, economista).

Una questione molto interessante, invece, è che i giochi d'azzardo forniscono un campo di studi per il calcolo delle probabilità, anzi

BOZZA - DRAFT

ne sono perfino una delle origini storiche, con – fra i relativamente moderni e ben noti – Pacioli, Cardano, Tartaglia, e Galileo Galilei che spiegò perché con 3 dadi il 10 e l'11 siano più probabili del 9 e del 12, che pure sembrano ottenersi in un ugual numero di "modi". Ancora occupandosi di gioco d'azzardo, Pascal e Fermat giunsero al moderno concetto di probabilità.

BOZZA - DRAFT

VIII – Variabili aleatorie discrete

BOZZA - DRAFT

37 Introduzione alle variabili aleatorie

37.1 Variabili aleatorie discrete e continue

Una funzione (che ad eventi semplici associa numeri reali)

$$X : \Omega \to \mathbb{R}$$

si chiama *variabile aleatoria* (purché sia sufficientemente regolare[150] come sono tutte quelle che capitano a un livello elementare).

Per esempio una variabile aleatoria è il risultato del lancio di un dado, che potremo rappresentare così in 2 diversi casi:

$$X := \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}^{\text{dado equilibrato}}_{\text{ovvero regolare}} \qquad Y := \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{12} & \frac{1}{4} & \frac{1}{6} & \frac{1}{6} & \frac{1}{12} \end{pmatrix}^{\text{qua un certo}}_{\text{dado truccato}}$$

ed entrambe potrebbero rappresentare anche le durate di due malattie in giorni, alquanto irrealistiche per la lontananza da distribuzioni a campana – si disegnino i bar chart delle probabilità – come pure quest'altra

$$Z := \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 0.1 & 0.4 & 0.2 & 0.2 & 0.1 \end{pmatrix}$$

decisamente più realistica, che potrebbe anche rappresentare il punteggio da 1 (morte) a 5 (ottimo), esito (per esempio a 3 anni) di un intervento farmacologico.

Si noti che nel caso delle durate in giorni, 5 è effettivamente quintuplo di 1, mentre "ottimo" non è quintuplo di "morte": in quel caso i numeri sono solo simboli senza valore quantitativo.

La variabile aleatoria del dado va vista come il risultato del lancio del dado prima di lanciare il dado. Dopo averlo lanciato invece avremo un numero, che è tutt'altra cosa. E dopo averlo lanciato più volte avremo un dataset, che è un'altra cosa ancora.

BOZZA - DRAFT

[150] Precisamente deve essere $\{\omega | X(\omega) \leq t\} \in \mathbb{A}$ per ogni t reale; questione sottile.

Allora, finora, niente di nuovo nella sostanza, solo una diversa impostazione di fenomeni già considerati.

Un'altra variabile aleatoria, non così facilmente rappresentabile, è il peso in kg del primo bambino che nascerà vivo il 1 gennaio 2020, ora locale del luogo. Potrebbe essere 3.412..., 4.576..., eccetera. Ragionevolmente parlando, ognuno dei singoli valori ha probabilità 0: perché mai il primo bambino dovrebbe pesare *esattamente* (con infiniti decimali) 3.18452785356... kg? Scriveremo

$\{X < 3.5\}$ intendendo l'evento $\{\omega | X(\omega) < 3.5\}$

e similmente con $\leq$ e $>$ e $\geq$.
E ancora con tutte le possibili variazioni di $<$ e $\leq$

$\{2.5 < X \leq 3.5\}$ intendendo l'evento $\{\omega | 2.5 < X(\omega) \leq 3.5\}$

e per questi eventi, e altri del tipo $X \in I \subset \mathbb{R}$, con I insieme sufficientemente regolare (in Farmacia quasi solo intervalli), possiamo invece attenderci probabilità diverse da 0, per esempio
$P(X < 20) = 1$ il bambino peserà sicuramente meno di 20 chili,
$P(X < 0.1) = 0$ il bambino nato vivo non peserà meno di 100 g,
e – ipotizziamo qua – con probabilità 50% peserà meno di 2.1 kg: $P(X \leq 2.1) = 0.5$. E magari ancora $P(X \leq 2.5) = 0.7 = 70\%$.

La funzione

$$F_X(t) := P(X \leq t)$$

si chiama *funzione di ripartizione* della variabile aleatoria X, *cumulative distribution function*:

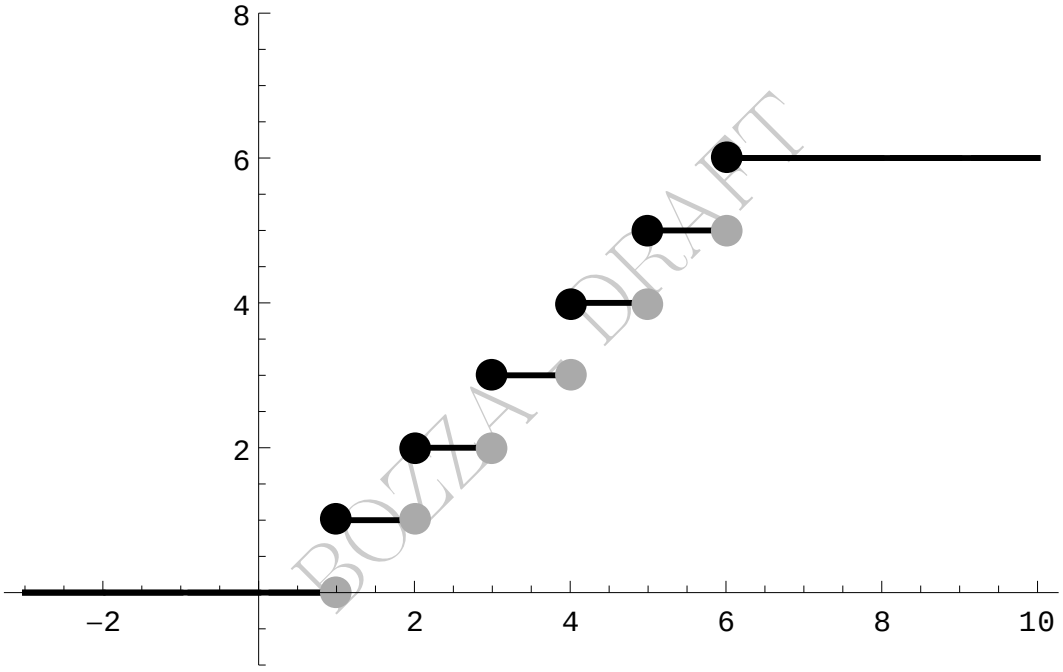
$$F_X : \mathbb{R} \to \mathbb{R}$$

In effetti i valori di essa sono in $[0, 1]$ e il grafico ha più o meno vagamente una forma sigmoide, da 0 in $-\infty$ a 1 in $+\infty$.

BOZZA - DRAFT

Nei grafici (uno è in figura) delle funzioni di ripartizione delle v.a. X e Y inizialmente considerate si osserverà che i salti hanno ampiezze pari alle corrispondenti probabilità. E questo è un fatto generale che succede per tutte le variabili aleatorie discrete:

$$\forall\, x \text{ punto di salto } P(X = x) = salto$$



BOZZA - DRAFT

Figure 43: Funzione di riìpartizione del dado regolare. Ha 6 salti di $\frac{1}{6}$.

La prima v.a., quella del dado, si dice *discreta* (cioè a valori "ben separati", anche eventualmente infiniti) e la seconda *continua*.

Si noti che, almeno per le variabili aleatorie continue, questa impostazione è veramente innovativa, e permette una valida trattazione di casi che altrimenti nel modo precedente si ridurrebbero semplicemente ad affermare che ogni singolo valore ha probabilità nulla

$$\forall\, t \quad P(X = t) = 0 \qquad (X \textbf{ v.a. continua})$$

e sommando o moltiplicando zeri non si otterrebbe niente di significativo.

Teorema (ovvio).

$$\boxed{P\{a < X \le b) = F_X(b) - F_X(a),\ \forall\, a, b} \tag{89}$$

e uno o entrambi gli estremi possono essere infiniti, e allora $F_X(x)$ andrà inteso nel senso del limite: $F_X(-\infty) = 0$ e $F_X(+\infty) = 1$.

La soprastante fondamentale formula darà
per le variabili aleatorie discrete: somme o serie;
per le variabili aleatorie continue: integrali definiti.
Anticipiamo che per le densità delle variabili aleatorie continue è

$$\boxed{\text{funzione di ripartizione: } F(x) = \int_{-\infty}^{x} f(t)\,dt} \tag{90}$$

$$\boxed{\text{(derivata della f.r.) } F' = f \quad \text{(densità)}} \tag{91}$$

$$\boxed{P(a < X \le b) = F(b) - F(a) = \int_{a}^{b} f(t)\,dt} \tag{92}$$

La funzione di ripartizione delle v.a. X, Y, Z... potrà trovarsi denotata con F_X, F_Y, F_Z... ma anche con altri simboli: F, F_1, G...

La densità delle v.a. X, Y, Z... potrà trovarsi denotata con f_X, f_Y, f_Z... ma anche con altri simboli: f, f_1, g...

Vediamo un esempio di applicazione della (92), con primo estremo $a = -\infty$.

ESERCIZIO$_{\mu 2021}$ % Si consideri una variabile aleatoria Z di densità $f(t) = t$ per $1 \le t \le \sqrt{3}$, e 0 altrimenti. Calcolare $P\left(Z < \frac{3}{2}\right)$.

SVOLGIMENTO

Verrà usato lo standard del punto decimale. (Ma si potrebbe usare quello della virgola decimale, a scelta).

$$P\left(Z < \frac{3}{2}\right) =$$

BOZZA - DRAFT

con $<$ oppure $\leq$ vale uguale perchè Z è una variabile aleatoria continua

$$= P\left(Z \leq \frac{3}{2}\right) =$$

$$= \int_{-\infty}^{\frac{3}{2}} \text{densità} =$$

la densità è 0 prima di $t = 1$ e dopo $t = \sqrt{3} \approx 1.73$ ma $\frac{3}{2} = 1.5$ è più piccolo

$$= \int_{1}^{\frac{3}{2}} t\,dt =$$

$$= \left[\frac{t^2}{2}\right]_1^{\frac{3}{2}} =$$

$$= \frac{1}{2} \cdot \left(\frac{3}{2}\right)^2 - \frac{1}{2} \cdot 1^2 =$$

$$= \frac{1}{2} \cdot \frac{9}{4} - \frac{1}{2} =$$

$$= \frac{9-4}{8} =$$

$$= \frac{5}{8} = 0.625 =$$

$$\boxed{62.5\%}$$

BOZZA DRAFT

Tipico delle densità riferite alla realtà sensibile, sia discrete che continue, è avere – spesso ma non proprio sempre – una forma "più o meno a campana".

37.2 La campana gaussiana e quella di Cauchy

La più "pura" delle campane è la *campana gaussiana* della densità normale standard

$$\phi(t) := \frac{1}{\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{t^2}{2}}$$

che però ha il difetto che (quasi tutti) gli integrali di essa non si riescono a calcolare con le funzioni elementari.

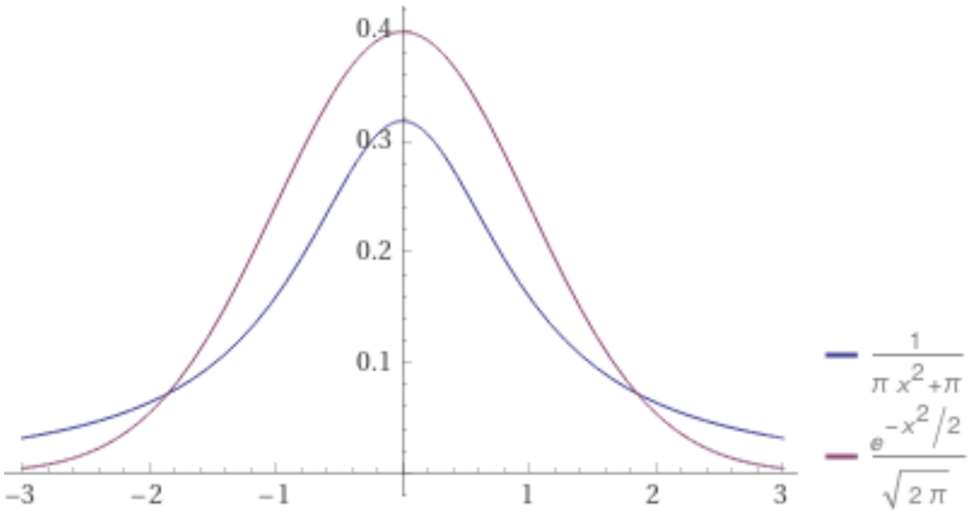


Figure 44: Magenta: densità normale standard (campana gaussiana). Blu: densità di Cauchy. Screenshot da WolframAlpha.

BOZZA - DRAFT

Un modello più semplice di campana è la densità di Cauchy

$$f(t) := \frac{1}{\pi} \cdot \frac{1}{1+t^2}$$

per la quale si ha subito, integrando elementarmente,

$$P(a < X \leq b) = \int_a^b f(t)\,dt = \frac{1}{\pi}(\arctan b - \arctan a)$$

per esempio

$$P(1 < X \leq \sqrt{3}) = \frac{1}{\pi}(\arctan\sqrt{3} - \arctan 1) =$$

$$= \frac{1}{\pi}\left(\frac{\pi}{3} - \frac{\pi}{4}\right) = \frac{4-3}{3\cdot 4} =$$

$$= \frac{1}{12} \approx 0.0833 = 8.33\%.$$ link a WolframAlpha ->

Definizione (implicita): code, e code pesanti.
La densità di Cauchy ha *code* molto più *pesanti* della densità normale.

38 Variabili aleatorie discrete

Data una v.a. discreta, cioè con valori "ben separati", come per esempio la X e la Y della lezione precedente, o queste

$$Z := \begin{pmatrix} 0 & 1 & 2 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix} \quad W := \begin{pmatrix} 1 & 2 & 3 & \dots & k & \dots \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \dots & \frac{1}{2^k} & \dots \end{pmatrix} \begin{matrix} \leftarrow infiniti\ valori\ possibili\ ma\ staccati \\ \\ \leftarrow qua\ si\ forma\ una\ serie\ geometrica \end{matrix}$$

individuiamo dei *valori* x_k, in numero finito o anche infiniti ma comunque "ben separati" (non necessariamente interi) e corrispondentemente le loro probabilità $p_k := P(v.a. = x_k)$:

$$X := \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_k & [\dots] \\ p_1 & p_2 & p_3 & \dots & p_k & [\dots] \end{pmatrix} \begin{matrix} \leftarrow fra\ le\ quadre\ possono\ esserci\ numeri,\ se\ i \\ valori\ sono\ infiniti,\ oppure\ no\ nell'altro\ caso \end{matrix}$$

e ovviamente deve essere, nel senso di una somma o di una serie,

$$somma\ su\ tutti\ gli\ indici \rightarrow \sum_k p_k = 1 \quad (esistono\ serie\ con_somma\ 1,\ non\ solo\ geom.)$$

La funzione di ripartizione $F_X(t) := P(X \le t)$ di una v.a. discreta X è una *funzione a scala (step function)* costante fra x_{k-1} e x_k, con salti pari a p_k in x_1, x_2,...,x_k... e "pallini pieni" a sinistra (continuità a sinistra).

Da adesso consideriamo solo valori interi; allora $p_k = P(v.a. = k)$. La funzione

$$p_k := \begin{cases} P(X = k) & se\ k \in \mathbb{Z} \\ 0 & altrimenti \end{cases}$$

si chiama ***densità*** della variabile aleatoria X. Tenderemo a usare la lettera p coi pedici per le probabilità , mentre per i valori della v.a. possiamo usare x_1, x_2... per una v.a. X e y_1, y_2... per una v.a. Y, eccetera. O anche x_0, x_1... iniziando da 0, o da altro numero, a seconda dei casi. Per esempio per le X e W di prima

$$p_k = P(X = k) := 1/6, \quad k = 1, ..., 6$$

$$p_k = P(W = k) := 1/2^k, \quad k = 1, 2, 3...$$

Le densità si rappresentano graficamente coi bar chart.

BOZZA - DRAFT

38.1 Alcuni esempi di variabili aleatorie discrete

Esempio$_\mu$ (sulla notazione "binomiale" delle variabili aleatorie discrete).
Per una v.a. a 3 valori con $p_1 = \frac{1}{\pi}$ e $p_2 = \frac{1}{\pi^2}$ determinare p_3.
Si tratta di trovare l'unico numero incognito p_3 in

$$\begin{pmatrix} x_1 & x_2 & x_3 \\ \frac{1}{\pi} & \frac{1}{\pi^2} & p_3 \end{pmatrix}$$

e da $\frac{1}{\pi} + \frac{1}{\pi^2} + p_3 = 1$ $\leftarrow$ somma 1 delle probabilità
segue subito $p_3 = 1 - \left(\frac{1}{\pi} + \frac{1}{\pi^2}\right) \approx 0.5804 = 58.04\%$.

Il [risultato del lancio di un] dado (regolare) è una v.a. $X \sim \mathbb{U}\{1,6\}$ e il simbolo $\sim$ lo leggeremo *"con legge"* e il simbolo $\mathbb{U}\{1,6\}$ lo leggeremo *"con legge uniforme discreta di parametri 1 e 6"*:

$$X := \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix} \qquad \textit{equivalentemente}: \ X \sim \mathbb{U}\{1,6\}.$$

E associando a 0 e 1 le facce di una moneta

$$M := \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \sim \mathbb{U}\{0,1\} \text{ rappresenta una moneta } \textit{equilibrata}.$$

Esercizio$_\mu$ Si rappresenti graficamente la densità e la funzione di ripartizione di $\mathbb{U}\{0,1\}$ e $\mathbb{U}\{1,6\}$.

Esercizio$_\mu$ Ecco 3 variabili aleatorie discrete non uniformi

$$Y := \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{12} & \frac{1}{6} & \frac{1}{6} & \frac{1}{12} \end{pmatrix}$$

$$U := \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} & \frac{1}{12} & \frac{1}{12} \end{pmatrix}$$

$$V := \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 0.2 & 0.3 & 0.2 & 0.2 & 0.1 \end{pmatrix}$$

BOZZA - DRAFT

che potrebbero rappresentare la durata in giorni di 3 malattie oppure di una malattia con 3 diverse terapie.
Si disegnino i bar chart delle probabilità, ovvero, delle densità, e i grafici delle funzioni di ripartizione. Si noterà che la prima è alquanto irrealistica, con una forma decisamente non a campana.

38.2 Forma del grafico di densità e f.r.

Sia per variabili aleatorie discrete che continue, la funzione di ripartizione è almeno debolmente crescente (non decrescente) dal valore 0 in $-\infty$, nel senso del limite, al valore 1 in $+\infty$, nel senso del limite, quindi

— il grafico della f.r. ha forma almeno vagamente sigmoide,

seppure non sono esclusi una sorta di "terrazzamenti" ovvero "gradinate".

BOZZA - DRAFT

Invece le densità, che nel caso discreto è opportuno rappresentare come bar chart (eventualmente con infinite colonne) e nel caso continuo con "normali curve", possono avere forme molto variabili, vincolate essenzialmente solo da:

– non negatività, cioè valori ≥ 0

— area del sottografico 1 nel caso continuo

— somma ovvero serie dei valori a somma 1 nel caso discreto

– sufficiente regolarità, molto difficile da definire, ma sempre presente in tutti i casi di interesse nelle Scienze applicate – e quindi concretamente non è mai da dimostrare. E certo, una densità non nulla solo sui punti di una "nuvola di punti" capricciosa come un Insieme di Cantor, no, non va bene – ma interessa solo ai matematici.
Rimane dunque possibile per le densità – a differenza delle funzioni di ripartizione sempre almeno vagamente sigmoidi – un'enorme variabilità, tuttavia nei casi più comuni nelle Scienze Applicate e in particolare nelle Scienze Biomediche, quasi sempre le densità hanno quest'ulteriore caratteristica:

— forma almeno vaghissimamente a campana, 0 in $-\infty$ e $+\infty$, nel senso del limite, quasi sempre unimodali ovvero con un solo punto di massimo assoluto, e spesso perfino con un solo punto di massimo relativo.

38.3 Variabili aleatorie uniformi discrete

Una **variabile aleatoria uniforme discreta** di parametri interi a e b ha n valori interi a, $a+1$, $a+2$,...,$b = a+n-1$, con le corrispondenti probabilità tutte uguali $p_k = \frac{1}{n}$ per $k = 1, ..., n$, e la sua distribuzione ovvero legge verrà indicata con $\mathbb{U}\{a, b\}$.
Per esempio il [risultato del lancio di un] dado (regolare) è una v.a. $X \sim \mathbb{U}\{1, 6\}$ e il simbolo $\sim$ lo leggeremo *"con legge"*:

$$X := \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix} \qquad \textit{equivalentemente}: \ X \sim \mathbb{U}\{1, 6\}.$$

E associando a 0 e 1 le facce di una moneta

$$M := \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \sim \mathbb{U}\{0, 1\} \text{ rappresenta una moneta } \textit{equilibrata}$$

Ecco un'altra, $\sim \mathbb{U}\{a, b\}$ con $a = -2$ e $b = 4$ e allora $n = b - a + 1 = 7$:

$$Y := \begin{pmatrix} -2 & -1 & 0 & 1 & 2 & 3 & 4 \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} \end{pmatrix} \qquad \textit{equivalentemente}: \ Y \sim \mathbb{U}\{-2, 4\}.$$

Esercizio$_\mu$ Si rappresenti graficamente la densità e la funzione di ripartizione di $\mathbb{U}\{0, 1\}$ e $\mathbb{U}\{1, 6\}$.

38.4 Variabili aleatorie discrete con infiniti valori

Una **variabile aleatoria geometrica iniziante da 1** di parametro p (e dev'essere $0 \le p \le 1$) ha valori 1,2,3... con densità

$$p_k = p(1-p)^{k-1}, \qquad k = 1, 2, 3... \quad \textit{somma della serie}: 1$$

per esempio con $p := \dfrac{1}{2}$ si ha quella di prima $\begin{pmatrix} 1 & 2 & 3 & ... & k & ... \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & ... & \frac{1}{2^k} & ... \end{pmatrix}$

BOZZA - DRAFT

$$\text{e con } p := \frac{1}{3} \text{ si ha } \quad \begin{pmatrix} 1 & 2 & 3 \dots & k & \dots \\ \frac{1}{3} & \frac{2}{9} & \frac{4}{27} \dots & \frac{1}{3}\left(\frac{2}{3}\right)^{k-1} & \dots \end{pmatrix}$$

Esercizio$_{\mu}$ Si rappresenti graficamente la densità e la funzione di ripartizione di $Geom_1\left(\frac{1}{2}\right)$, la variabile aleatoria geometrica di parametro $\frac{1}{2}$ (simbologia per nulla standard purtroppo).

Osservazione. La probabilità di "$X = h \ vel \ X = k$", con $h \neq k$, essendo eventi disgiunti, è la somma delle 2 probabilità:

$$P(X = h \ vel \ X = k) = P(X = h) + P(X = k)$$

e similmente con 3 o più valori diversi, e questo vale per le v.a. discrete con qualunque distribuzione.
Per esempio per la v.a. W considerata all'inizio

$$P(W = 2 \vee W = 3) = P(W = 2) + P(W = 3) =$$

$$= \frac{1}{4} + \frac{1}{8} = \frac{2+1}{8} = \frac{3}{8}$$

Nella prossima Lezione vedremo anche la **variabile aleatoria geometrica iniziante da 0**.

BOZZA DRAFT

39 Variabili aleatorie uniformi e geometriche

39.1 Variabili aleatorie e processo di Bernoulli

La variabili aleatoria di Bernoulli o bernoulliana di parametro p, è

$$\begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$$

(per esempio una moneta con testa=1 con $p = \frac{1}{2}$ se regolare) identificabile con

$$\begin{pmatrix} fallimento & successo \\ 1-p & p \end{pmatrix}$$

Denoteremo con $B(p)$ questa legge.
Un **processo di Bernoulli** è una successione $(X_1, X_2...)$ di variabili aleatorie indipendenti aventi la medesima legge di Bernoulli $B(p)$. Leggiamo su Wikipedia, l'enciclopedia libera, alla voce *Processo di Bernoulli*:

> Un processo di Bernoulli può essere considerato come una sequenza di lanci di una moneta (eventualmente anche truccata). Ogni singolo lancio è detto prova di Bernoulli.
>
> In particolare, essendo le variabili indipendenti, vale la mancanza di memoria: la probabilità di una prova di Bernoulli non è influenzata dal risultato delle precedenti (che quindi non possono fornire alcuna informazione sulla nuova prova).

Una sua *realizzazione* è una successione di zeri e uni, con l'1 a denotare il successo, con una sua probabilità p, e lo 0 l'insuccesso: come 0,0,0,0,1,0,1,0,0,0,1...
Anche sopravvive, sopravvive, sopravvive, sopravvive, muore (e non serve continuare). (Si noti che qua il successo è la morte).

BOZZA - DRAFT

39.2 Variabili aleatorie geometriche

Una **variabile aleatoria geometrica iniziante da 1** di parametro p (e dev'essere $0 \le p \le 1$)

$$X \sim Geom_1(p) \qquad \text{(simbologia non standard purtroppo)}$$

ha valori 1,2,3... con densità

$$p_k = p(1-p)^{k-1}, \qquad k = 1, 2, 3... \quad \textit{somma della serie: } 1$$

per esempio con $p := \dfrac{1}{2}$ si ha quella di prima $\begin{pmatrix} 1\,2\,3\,...\;k\;... \\ \frac{1}{2}\;\frac{1}{4}\;\frac{1}{8}\;...\;\frac{1}{2^k}\;... \end{pmatrix}$

e con $p := \dfrac{1}{3}$ si ha $\begin{pmatrix} 1 & 2 & 3\,...\;k & ... \\ \frac{1}{3} & \frac{2}{9} & \frac{4}{27}\;...\;\frac{1}{3}\left(\frac{2}{3}\right)^{k-1} & ... \end{pmatrix}$

Una **variabile aleatoria geometrica iniziante da 0** di parametro p (e dev'essere $0 \le p \le 1$)

$$X \sim Geom_0(p) \qquad \text{(simbologia non standard purtroppo)}$$

ha valori 0,1,2... con densità

$$p_k = p(1-p)^{k}, \qquad k = 0, 1, 2... \quad \textit{somma della serie: } 1$$

per esempio con $p := \dfrac{1}{2}$ si ha $\begin{pmatrix} 0\,1\,2\,...\;k\;... \\ \frac{1}{2}\;\frac{1}{4}\;\frac{1}{8}\;...\;\frac{1}{2^{k+1}}\;... \end{pmatrix}$

e con $p := \dfrac{1}{3}$ si ha $\begin{pmatrix} 0 & 1 & 2\,...\;k & ... \\ \frac{1}{3} & \frac{2}{9} & \frac{4}{27}\;...\;\frac{1}{3}\left(\frac{2}{3}\right)^{k} & ... \end{pmatrix}$.

Supponiamo che vogliamo ottenere un certo risultato con una successione di prove indipendenti, in ciascuna delle quali il successo ha probabilità p. (Per esempio la testa con una moneta e allora $p = \frac{1}{2}$ oppure il 3 con un dado e allora $p = \frac{1}{6}$). Cioè consideriamo un processo di Bernoulli. Allora

$$P(\text{ci vogliono } k \text{ prove per il primo successo}) = p_k = p\,(1-p)^{k-1}$$

$$k = 1, 2, 3...$$

BOZZA - DRAFT

cioè la v.a. “numero di tentativi per avere il primo successo" ha legge $Geom_1(p)$.
Analogamente

$$P(\text{ci capitano } k \text{ insuccessi prima del primo successo}) = p\,(1-p)^k$$

$$k = 0, 1, 2...$$

cioè la v.a. “numero di insuccessi prima avere il primo successo" ha legge $Geom_0(p)$.
Il cosiddetto successo potrebbe per esempio essere un errore in farmacia, il primo della giornata, e in un articolo scientifico hanno effettivamente trovato questo:

> A (...) study involving 50 pharmacies located in six cities across the United States was conducted. A pharmacist trained to detect dispensing errors recorded the number of prescriptions filled by each pharmacy staff member and noted which prescription represented the staff member's first dispensing error (defined as any deviation from the prescriber's order) made during the observation period. (...) the observed cumulative distribution of dispensing errors could have come from a geometric probability distribution that assumed dispensing error rates of 2-3%.[151]

(Fra 2 e 3 percento di errori in farmacia rispetto alle indicazioni dei medici!)

BOZZA - DRAFT

ESERCIZIO$_{\mu 2018}$
$* \approx \%$ Per una variabile aleatoria X geometrica iniziante da 1 di parametro $\frac{1}{3}$ calcolare

$$P(X \geq 3).$$

SVOLGIMENTO.

$$P(X \geq 3) =$$

Con l'evento complementare, che ci evita il calcolo della somma di una serie,

$$= 1 - P(X < 3) =$$

la X è geometrica iniziante da 1, e solo 1 e 2 sono i possibili valori < 3

$$= 1 - P(X = 1 \vee X = 2) =$$

[151] Am J Health Syst Pharm. 2006 Jun 1;63(11):1056-61. Geometric probability distribution for modeling of error risk during prescription dispensing. Carnahan BJ, Maghsoodloo S, Flynn EA, Barker KN.

la probabilità dell'unione di eventi disgiunti è la somma delle probabilità

$$= 1 - \big(P(X=1) + P(X=2)\big) =$$

e ricordando la formula della densità considerata $p_k = p\,(1-p)^{k-1}$, con $p = \frac{1}{3}$

$$= 1 - \frac{1}{3}\Big(1-\frac{1}{3}\Big)^{1-1} - \frac{1}{3}\Big(1-\frac{1}{3}\Big)^{2-1} =$$

e con semplici calcoli aritmetici

$$= 1 - \frac{1}{3}\Big(\frac{2}{3}\Big)^{0} - \frac{1}{3}\Big(\frac{2}{3}\Big)^{1} = 1 - \frac{1}{3}\cdot 1 - 1 - \frac{1}{3}\cdot\frac{2}{3} = \frac{9-3-2}{9}$$

e in conclusione

$$\boxed{\tfrac{4}{9} \approx 0.444 = 44.4\%}$$

BOZZA DRAFT

ESERCIZIO$_{\mu 2018 \neq}$
$* \approx \%$ Si consideri un'entità biologica (come batteri, virus, corpuscoli sanguigni...) che da quando si forma, poi ha una vita (in giorni, interi) che è pari a $7 +$ una variabile aleatoria geometrica X iniziante da 0 di parametro $\frac{1}{2}$. (Quindi può vivere 7 giorni, 8 giorni, 9 giorni... eccetera, con probabilità via via sempre più piccole).
Che probabilità ha di vivere un numero pari di giorni?

SVOLGIMENTO

$$P(\text{vive un numero pari di giorni}) = P(7 + X \text{ è pari}) =$$

$$= P(X \text{ è dispari}) =$$

$$= P(X \text{ è 1 oppure 3 oppure 5 oppure 7...}) =$$

più formalmente

$$= P(X = 1 \vee X = 3 \vee X = 5 \vee X = 7 \vee ...) =$$

e trattandosi di eventi disgiunti

$$= P(X=1) + P(X=3) + P(X=5) + P(X=7) + ... =$$

ricordando la densità $p_k = p\,(1-p)^k$ per $k = 0, 1, 2...$ della variabile aleatoria geometrica di parametro p iniziante da 0, nel nostro caso $\frac{1}{2}\,(\frac{1}{2})^k$,

$$= \frac{1}{4} + \frac{1}{16} + \frac{1}{64} + ...$$

dove si riconosce la serie geometrica di ragione $\frac{1}{4}$ (perché ogni termine è pari al precedente moltiplicato per $\frac{1}{4}$)

$$a + ar + ar^2 + ar^3 + ...ar^n + ...$$

con anche $a = \frac{1}{4}$ e allora somma

$$\frac{a}{1-r} = \frac{\frac{1}{4}}{1-\frac{1}{4}} = \frac{\frac{1}{4}}{\frac{3}{4}} =$$

$$\boxed{\tfrac{1}{3} \approx 0.333 = 33.3\%}$$

oppure, con un decimale in più,

$$\boxed{\tfrac{1}{3} \approx 0.3333 = 33.33\%}$$

(e se ne potrebbero mettere anche di più, se fosse utile o desiderato).

ESERCIZIO$_{\mu 2018}$ $* \approx \%$ Per una variabile aleatoria geometrica Z iniziante da 1 di parametro $\frac{1}{4}$ calcolare

$$P(Z > 2).$$

SVOLGIMENTO

$$P(Z > 2) =$$

con l'evento complementare, che ci evita il calcolo della somma di una serie,

$$= 1 - P(Z \leq 2) =$$

la Z è geometrica iniziante da 1, e solo 1 e 2 sono i possibili valori ≤ 2

$$= 1 - P(Z = 1 \vee Z = 2) =$$

la probabilità dell'unione di eventi disgiunti è la somma delle probabilità

$$= 1 - \big(P(Z = 1) + P(Z = 2)\big) =$$

con la notazione consueta per la densità geometrica iniziante da 1

$$= 1 - (p_1 + p_2) =$$

e ricordando la formula della densità considerata $p_k = p\,(1-p)^{k-1}$, con $p = \frac{1}{4}$

$$= 1 - \frac{1}{4}\Big(1 - \frac{1}{4}\Big)^{1-1} - \frac{1}{4}\Big(1 - \frac{1}{4}\Big)^{2-1} =$$

e con semplici calcoli aritmetici

$$= 1 - \frac{1}{4}\Big(\frac{3}{4}\Big)^0 - \frac{1}{4}\Big(\frac{3}{4}\Big)^1 = 1 - \frac{1}{4} \cdot 1 - \frac{1}{4} \cdot \frac{3}{4} = \frac{16 - 4 - 3}{16}$$

e in conclusione

BOZZA - DRAFT

$$\frac{9}{16} \approx 0.5625 = 56.25\%$$

o con minore ma comunque accettabile precisione

$$\frac{9}{16} \approx 0.563 = 56.3\%$$

BOZZA - DRAFT

Esempi ed esercizi

Esempio$^{f}_{\mu}$ Determinare la v.a. $\mathbb{U}\{0,7\}$.
Si ha subito in base alla definizione di $\mathbb{U}\{a,b\}$

$$U := \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \end{pmatrix} \sim \mathbb{U}\{0,7\}.$$

Esercizio. Rappresentare graficamente densità e funzione di ripartizione della v.a. U del soprastante esempio.

Esempio$^{f}_{\mu}$ Determinare la v.a. $\mathbb{U}\{-3,1\}$.
Si ha subito in base alla definizione di $\mathbb{U}\{a,b\}$

$$V := \begin{pmatrix} -3 & -2 & -1 & 0 & 1 \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \end{pmatrix} \sim \mathbb{U}\{-3,1\}.$$

Esercizio. Rappresentare graficamente densità e funzione di ripartizione della v.a. V del soprastante esempio.

Esercizio. Determinare la v.a. $Z \sim \mathbb{U}\{-4,2\}$ e poi calcolare $P(Z \geq 1)$, $P(Z > 1)$, $P(Z \leq 1)$, $P(Z \geq -\pi)$, $P(Z \geq \sqrt{2})$, $P(-1 \leq Z < 1)$, $P(Z^2 = 1)$, $P(Z^2 \leq 1)$. [L'ultime vale $\frac{3}{7}$].

Esercizio. Determinare le v.a. geometriche dei 2 tipi, di parametri $\frac{2}{3}$ e $\frac{4}{5}$, e verificare le somme delle loro serie.

Esercizi$^{f}_{\mu}$ Rappresentare graficamente densità e funzioni di ripartizione delle v.a Z e W della lezione 38. Calcolare $P(W < 2)$, $P(W > 2)$, $P(W \geq 4)$, e con le serie $P(W > 100)$ e $P(W < 100)$.

Esercizi. Per una v.a. a 3 valori con $p_3 = \frac{1}{e}$ e $p_2 = \frac{1}{e^2}$ determinare p_1. Per una v.a. A a 4 valori 0, 1, 2, 3 con $p_1 = \frac{1}{7}$ $p_2 = \frac{1}{11}$ e $p_0 = \frac{1}{5}$ determinare p_3, e poi $P(A \geq 1)$, $P(A > 1)$, $P(A \leq 1)$, $P(A \geq -\pi)$, $P(A \geq \sqrt{2})$, $P(-1 \leq A < 1)$, $P(A^2 \leq 1)$, $P(A \neq 1)$.

BOZZA - DRAFT

40 Leggi congiunte e indipendenza

Lezione esclusa dal corso breve – 40 ore.

40.1 Introduzione alle leggi congiunte

Considereremo solo coppie di variabili aleatorie discrete ma con qualche attenzione tutto può essere esteso a 3 o più variabili aleatorie. Date 2 variabili aleatorie discrete X e Y

$$X := \begin{pmatrix} x_1\, x_2\, x_3 \ldots x_k\, [\ldots] \\ p_1\, p_2\, p_3 \ldots p_k\, [\ldots] \end{pmatrix} \qquad Y := \begin{pmatrix} y_1\, y_2\, y_3 \ldots y_k\, [\ldots] \\ p'_1\, p'_2\, p'_3 \ldots p'_k\, [\ldots] \end{pmatrix}$$

si può dimostrare che la coppia ordinata (X, Y) è una v.a., detta *2-dimensionale*. La sua densità è la funzione **densità congiunta**

$$p(x, y) := P(X = x \wedge Y = y)$$

per esempio per i 2 dadi, o per la coppia di lanci di 1 dado,

$$(X, Y) = \begin{pmatrix} (1,1) & (1,2) & \ldots & (6,6) \\ \frac{1}{36} & \frac{1}{36} & \ldots & \frac{1}{36} \end{pmatrix}$$

$$p(x, y) = \begin{cases} \frac{1}{36} & per\ x, y = 1, 2, 3, 4, 5, 6 \\ 0 & altrimenti \end{cases} \qquad \text{che scriveremo anche } p_{i,j}.$$

Detto V il prodotto cartesiano degli insiemi dei valori delle v.a.

$$V := \{x_1, x_2, x_3 \ldots x_k\, [\ldots]\} \times \{y_1, y_2, y_3 \ldots y_k\, [\ldots]\}$$

è ovviamente

$$\sum_{(x,y) \in V} p(x, y) = 1 \tag{93}$$

e se $A \subseteq V$, avendosi eventi disgiunti si ha ovviamente

$$P(A) = \sum_{(x,y) \in A} p(x, y) \tag{94}$$

per esempio per i 2 dadi regolari

$$P(X + Y = 8) = p(2,6) + p(3,5) + p(4,4) + p(5,3) + p(6,2) = \frac{5}{36}.$$

BOZZA - DRAFT

40.2 Densità marginali e indipendenza

Date le 2 v.a. discrete X e Y e la **v.a. bidimensionale** (X,Y) prima considerate, le densità p' di X e p'' di Y si chiamano **densità marginali** di (X,Y). Adesso indichiamo con p la densità congiunta. è (teorema)

$$p'_i = \sum_j p_{i,j} \qquad p''_j = \sum_i p_{i,j}$$

Le 2 v.a. X e Y si dicono **v.a. indipendenti** se

$$p_{i,j} = p'_i \cdot p''_j \qquad \forall\, i,j \quad \textit{ovvero se congiunta=prodotto delle marginali.}$$

Per esempio i 2 dadi o il doppio lancio di prima, con le marginali costantemente $\frac{1}{6}$ e la congiunta costantemente $\frac{1}{36}$. Invece questa densità congiunta

$$p_{i,j} := \begin{cases} \frac{1}{30} & \textit{per } i,j = 1,2,3,4,5,6,\ i \neq j \\ 0 & \textit{altrimenti} \end{cases}$$

dà luogo a 2 v.a. (non determinate nei valori per adesso)

$$\begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix} \qquad \begin{pmatrix} y_1 & y_2 & y_3 & y_4 & y_5 & y_6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}$$

che non sono indipendenti perché $p_{1,2} = \frac{1}{30} \neq \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = p'_1 \cdot p''_2$.
Ecco un'altra densità $p_{i,j}$ di v.a. bidimensionale (X,Y)

	y_1	y_2	y_3	y_4	
x_1	$\frac{4}{44}$	$\frac{3}{44}$	$\frac{2}{44}$	$\frac{2}{44}$	$\to \frac{11}{44} = p'_1$
x_2	$\frac{3}{44}$	$\frac{4}{44}$	$\frac{2}{44}$	$\frac{2}{44}$	$\to \frac{11}{44} = p'_2$
x_3	$\frac{2}{44}$	$\frac{2}{44}$	$\frac{5}{44}$	$\frac{5}{44}$	$\to \frac{14}{44} = p'_3$
x_4	$\frac{2}{44}$	$\frac{2}{44}$	$\frac{2}{44}$	$\frac{2}{44}$	$\to \frac{8}{44} = p'_4$
$p_{4,1}\nearrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	
	$\frac{11}{44}$	$\frac{11}{44}$	$\frac{11}{44}$	$\frac{11}{44}$	
	p''_1	p''_2	p''_3	p''_4	

con X e Y non indipendenti:

$$p_{1,2} = \tfrac{3}{44} \neq \tfrac{11}{44} \cdot \tfrac{11}{44} = p'_1 \cdot p''_2.$$

(Lasciamo tutto in 44-esimi sebbene qualche semplificazione sarebbe possibile).

$$X = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ \frac{11}{44} & \frac{11}{44} & \frac{14}{44} & \frac{8}{44} \end{pmatrix} \qquad Y = \begin{pmatrix} y_1 & y_2 & y_3 & y_4 \\ \frac{11}{44} & \frac{11}{44} & \frac{11}{44} & \frac{11}{44} \end{pmatrix}.$$

BOZZA - DRAFT

41 Variabili aleatorie binomiali e bimodali

In questa Lezione tratteremo le variabili aleatorie binomiali e quelle bimodali, due questioni molto diverse, da mettere comunque a confronto.

41.1 Le variabili aleatorie binomiali

Si dice che una variabile aleatoria X è binomiale di parametri n e p (e dev'essere n intero ≥ 1 e $0 \leq p \leq 1$) e si scrive $X \sim B(n,p)$, se ha *densità binomiale*, cioè

$$\boxed{p_k = P(X = k) := \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, ..., n.} \qquad (95)$$

La densità a binomiale rappresenta, al variare di k intero da 0 a n, la probabilità di ottenere k teste in n lanci di una moneta che ha $P(testa) = p$. Più in generale:

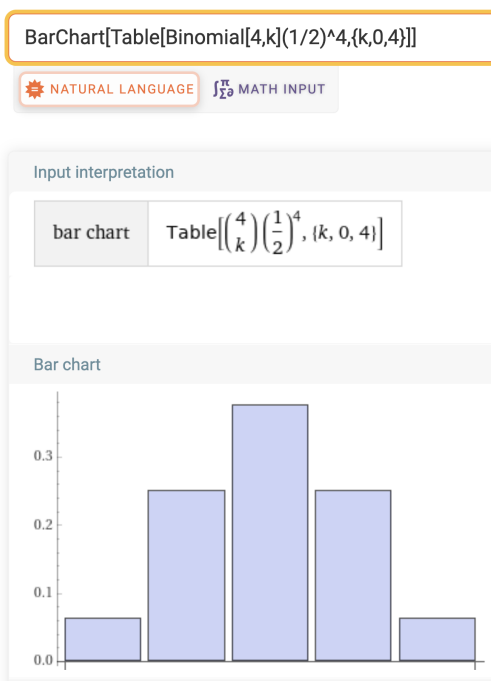
la densità a binomiale dà, al variare di k intero da 0 a n, la probabilità di un certo numero k di successi in uno schema successo-insuccesso con n prove indipendenti avendo il successo probabilità p ad ogni prova.

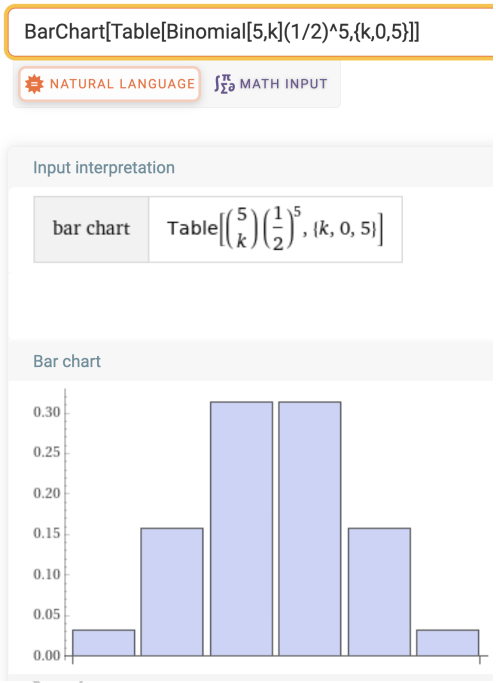
Gli istogrammi a barre rappresentativi della densità binomiale tendono ad avere forme più o meno a campana.

La forma a campana è tanto più riconoscibile quanto più n è grande e p è lontano da 0 e da 1. E la densità è tanto più "decentrata" rispetto a $\frac{n}{2}$ quanto più p è vicino a 0 o a 1; nei 2 casi il punto di massimo si avvicina a 0 e n rispettivamente.

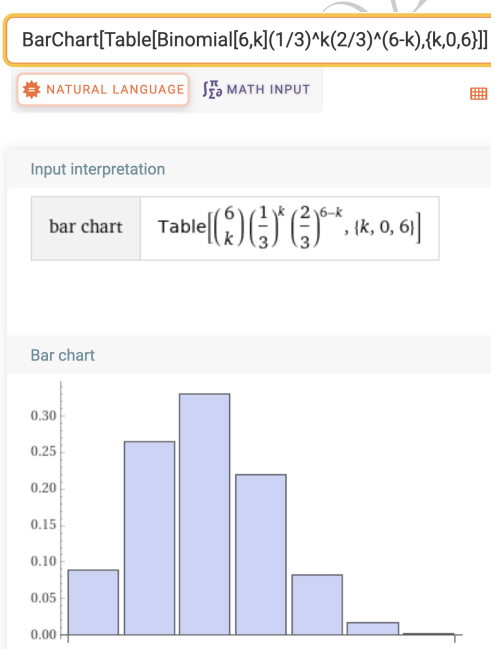
Ecco 2 istogrammi a barre tratti da WolframAlpha, con $n = 4$ e $n = 5$ ed entrambi con $p = \frac{1}{2}$, e quest'ultimo valore causa la simmetria delle campane.

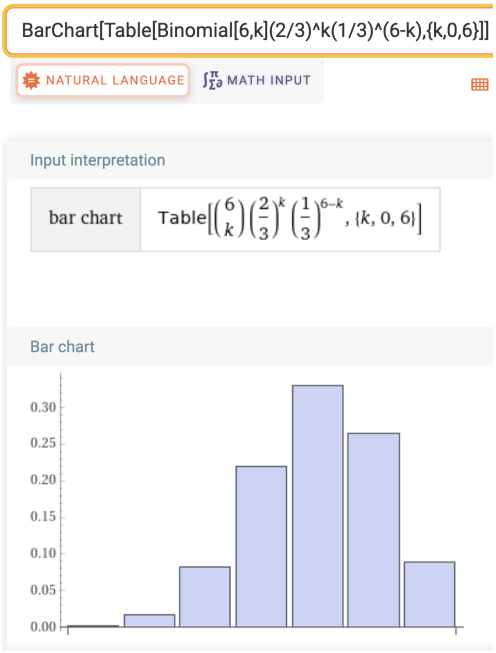
BOZZA - DRAFT





Ecco 2 istogrammi a barre tratti da WolframAlpha, con $n = 6$, e con rispettivamente $p = \frac{1}{3}$ e $p = \frac{2}{3}$, campane asimmetriche, con skewness non nulla.





L'allontanarsi di p dal valore centrale $\frac{1}{2}$ causerà inizialmente asimmetria nella forma a campana (si faccia il bar chart relativo a

$B\left(3, \frac{1}{4}\right)$) ma al crescere di n l'asimmetria tenderà ad attenuarsi sempre più, rimanendo invece la centratura spostata rispetto alla metà $\frac{n}{2}$ di n: la campana tenderà a centrarsi intorno a $n\,p$.

Esempio 1. Con $n = 2$ prove si possono ottenere 0 o 1 o 2 teste (che adesso consideriamo successi) e se la moneta è equilibrata è

$$P(k=0) = \binom{2}{0}\left(\tfrac{1}{2}\right)^0\left(1-\tfrac{1}{2}\right)^{2-0} = \tfrac{1}{4} \quad \leftarrow P(0\,teste)$$

$$P(k=1) = \binom{2}{1}\left(\tfrac{1}{2}\right)^1\left(1-\tfrac{1}{2}\right)^{2-1} = \tfrac{1}{2} \quad \leftarrow P(1\,teste)$$

$$P(k=2) = \binom{2}{2}\left(\tfrac{1}{2}\right)^2\left(1-\tfrac{1}{2}\right)^{2-2} = \tfrac{1}{4} \quad \leftarrow P(2\,teste)$$

Esempio 2. Qual è la probabilità di ottenere [esattamente] 3 volte il numero 1 su 7 lanci di un dado [regolare]?
Il successo ad ognuna delle $n = 7$ prove ha probabilità $\frac{1}{6}$ e allora, con la densità $B\left(7, \frac{1}{6}\right)$, la probabilità cercata è:

$$P(k=3) = \binom{7}{3}\left(\frac{1}{6}\right)^3\left(1-\frac{1}{6}\right)^{7-3} = \frac{7!}{3!(7-3)!}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^4 =$$

$$= \frac{1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6\cdot 7}{1\cdot 2\cdot 3\cdot 1\cdot 2\cdot 3\cdot 4}\cdot\frac{5^4}{6^7} = 5\cdot 7\cdot\frac{625}{279\,936} = \frac{21\,875}{279\,936} \approx 0.0781 = 7.81\%.$$

(Diciamo pure, a parole, circa 8%).

BOZZA - DRAFT

41.2 La binomiale come spiegazione delle campane

Ecco il grafico della densità dell'Esempio 1 visto, che è un bar chart, qua disposto orizzontalmente:

0 XXXX 0.25 qua c'è 1 solo caso su 4: CC cioè croce, croce
1 XXXXXXXX 0.5 qua ci sono 2 casi su 4: TC e CT
2 XXXX 0.25 qua c'è 1 solo caso su 4: TT cioè testa, testa

Si noti che nel soprastante bar chart si vede la distribuzione a campana molto più chiaramente che nel caso di un solo lancio

0 XXXXXX 0.5 (croce)
1 XXXXXX 0.5 (testa)

e si comincia a capire la natura della quasi onnipresente forma più o meno a campana: essa si forma in fenomeni composti da tante "cose che si sommano", parlando in modo (necessariamente) impreciso. Similmente le altezze degli individui di una popolazione, o qualunque altro parametro biometrico, biomedico o fisiologico, tenderanno ad avere una distribuzione con forma più o meno a campana, perché ognuno di quei parametri è risultante da molti altri: per esempio in un individuo altissimo potranno aver concorso nella stessa direzione ("successi") la genetica, l'alimentazione, concause fortuite imponderabili, e tale concordanza sarà rara mentre molti più individui avranno un'altezza media, e di nuovo pochi saranno bassissimi.
La forma a campana diventa ancora più evidente con 3 lanci di moneta, ovvero $B\left(3, \frac{1}{2}\right)$, e sempre più al crescere di n. Si faccia il bar chart di $B\left(3, \frac{1}{2}\right)$ e $B\left(4, \frac{1}{2}\right)$.

POZZA – DRAFT

Prova tu stesso. Leggiamo sul web[152] "Il termometro segna già−41°, ma da queste parti è normale. Siamo nella nella Jacuzia, nel nord-est della Siberia, chiamata anche la regione del "Polo del Freddo Boreale". Qui d'inverno la temperatura non supera mai gli zero gradi, e rimane quasi sempre al di sotto dei −30°C, raggiungendo talvolta i −60°C." Disegna un ipotetico grafico della densità della funzione temperatura in quel luogo, segnando anche sull'asse delle ascisse il dato rilevato. Osserva sul grafico il significato delle parole usate "mai", "quasi sempre" e "talvolta": corrispondono a 3 aree di sottografico, se la temperatura viene considerato un parametro "continuo" – cioè, in pratica, che può assumere valori con inifiniti decimali – e a somme di probabilità delle aste di un bar chart, se le temperature ambientali vengono discretizzate –

[152] http://www.rainews.it/dl/rainews/media/Siberia-la-raccolta-del-ghiaccio-nella-citta-piu-fredda.html Dalle informazioni sappiamo che la funzione densità vale 0 già da 0° in poi, e certamente prima di −273° ma in pratica, usando le conoscenze usuali della realtà sensibile, da −100°.

come si usa – in valori interi. La parola “normale" riferita al valore $-41°$ si riferisce a una quarta area, purché la normalità si riferisca al significato $\leq -41°$, altrimenti nel caso continuo la probabilità dei $-41°$ *esatti* è 0, e nel caso discreto sarà senz'altro *piccola*, tutt'altro che *grande* (“normale").

In mancanza di altre informazioni – che sarebbe Medicina, che qua non vogliamo e non possiamo fare – per le temperature corporee umane possiamo aspettarci similmente un grafico a campana, seppure situato altrove, intorno ai 37°.
Similmente per glicemie, sideremie, colesterolemie, e quant'altro.

Esercizio. Si disegni le funzione di ripartizione della $B\left(2, \frac{1}{2}\right)$ prima considerata, nell'Esempio 1. (Si osserverà che i 3 salti hanno ampiezza pari alle corrispondenti 3 probabilità, e questo è un fatto generale che succede per tutte le variabili aleatorie discrete, come già osservato.)

Abbandoniamo adesso le densità binomiali.

BOZZA - DRAFT

41.3 Densità in qualche modo bimodali

Diremo *bimodale* una densità “come" quella in Fig. 41.3. (96)

Al livello di questo testo elementare non possiamo definire rigorosamente cosa intendiamo per densità discreta con 2 (o eventualmente più, ma qua non ce ne occuperemo) punti di massimo relativo, o addirittura assoluto: è ovvio che in qualche modo bisognerà prescindere da singoli valori eccezionali, ma definirlo matematicamente non sarebbe facile. Comunque in molti casi semplici la *bimodalità* si riconosce a prima vista.

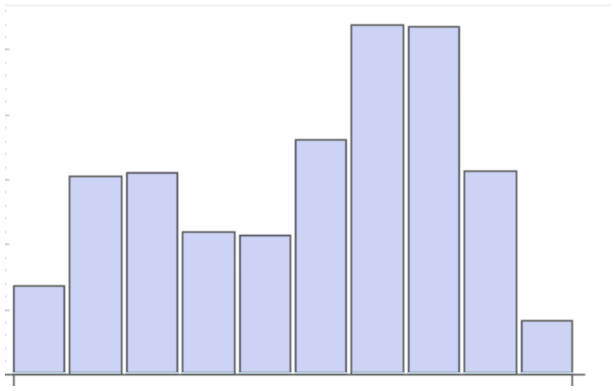


Figure 45: Densità bimodale. Screenshot da WolframAlpha

BOZZA - DRAFT

Distribuzioni “bimodali" fanno ritenere che ´´dietro" vi siano 2 fenomeni distinti.
Per esempio se misuriamo i tempi di impegno ad uno sportello, e dopo aver (sapientemente) riunito i tempi misurati in classi facciamo l'istogramma, e questo presenta 2 massimi nettamente distinti, lontani, è plausibile che allo sportello i clienti stiano facendo 2 operazione diverse: chi un'operazione più rapida, chi una più laboriosa. (Nella zona intermedia fra i 2 massimi sono i veloci a fare l'operazione laboriosa e i lenti a fare quella semplice).

Densità bimodali possono rivelare fenomeni interessanti. Per esempio in certi tempi e certi luoghi, il tempo che intercorre fra il matrimonio e la nascita del primo figlio si distribuisce con una tipica forma “a due gobbe", con un massimo assoluto intorno ai 9 mesi, corrispondente ai figli concepiti nel viaggio di nozze, e un massimo relativo vagamente alquanto prima (intorno ai 6-7 mesi nella statistica in Figura)

Nati vivi in ciascun mese successivo alla celebrazione del matrimonio.
(Australia 1911-1931)

Mesi	Maschi	Femmine	Maschi % Femmine
fino a 1	4528	4174	108,48
1- 2	5621	5211	107,87
2- 3	7933	7509	105,65
3- 4	11067	10146	109,08
4- 5	14223	12780	111,29
5- 6	18459	17139	107,70
6- 7	23804	22611	105,28
7- 8	21290	20778	102,46
8- 9	16538	15529	106,50
9-10	36075	33878	106,49
10-11	30457	29238	104,17
11-12	22348	21020	106,32

Figure 46: Da: *Il rapporto nei sessi nei nati da concepimento antenuziale*, Diego De Castro; Giornale degli Economisti e Rivista di Statistica, Serie quarta, Vol. 73 (Anno 48), No. 10 (ottobre 1933), p. 733. Online in https://www.jstor.org/stable/23229506

BOZZA DRA

In un articolo[153] scientifico si è studiato perchè le valutazioni soggettive di benessere si distribuiscano – o paiono distribuirsi, la questione è complessa e viene studiata nell'articolo – in modo bimodale, con tante persone che danno una variamente bassa valutazione della loro salute, e tante altre che danno una variamente alta valutazione:

> The EQ-5D is internationally one of the most widely used health-related quality-of-life instruments.1 It is therefore very important to obtain a good understanding of the characteristics of EQ-5D data. Many studies of the original 3-level version of the EQ-5D (now known as the EQ-5D-3L) have found that its index scores typically have a nonnormal distribution divided into 2 distinct groups. (This is often described as "bimodal," but we will argue that this label is misleading.) This phenomenon has been found in both patient and general populations, for many conditions, and using different country-specific value sets.

[153] Parkin D, Devlin N, Feng Y. What Determines the Shape of an EQ-5D Index Distribution? Medical Decision Making. 2016;36(8):941-951. doi:10.1177/0272989X16645581

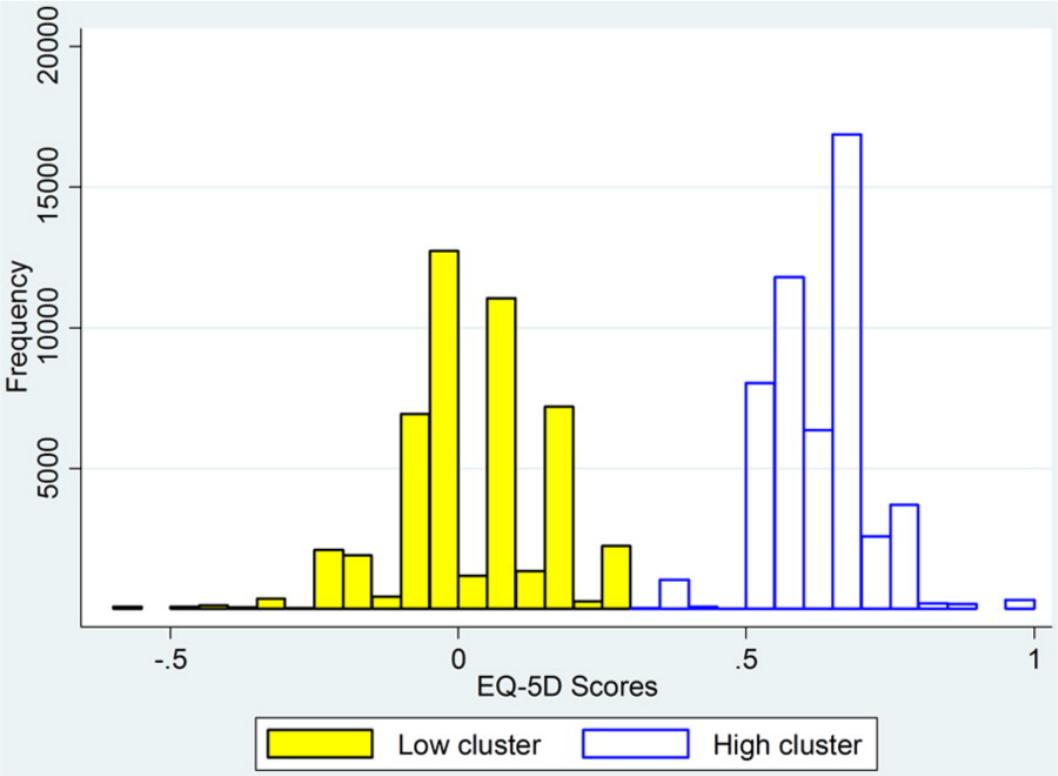


Figure 47: Figura 2 del citato articolo: "Distribution of presurgery EQ-5D index scores for hip replacement patients, using UK-weighted index."

BOZZA

Ancora, leggiamo[154] in https://en.wikipedia.org/wiki/Multimodal_distribution

> Examples of variables with bimodal distributions include the time between eruptions of certain geysers, the color of galaxies, the size of worker weaver ants, the age of incidence of Hodgkin's lymphoma, the speed of inactivation of the drug isoniazid in US adults, the absolute magnitude of novae, and the circadian activity patterns of those crepuscular animals that are active both in morning and evening twilight.

Si noti la – in qualche modo sorprendente – *age of incidence of Hodgkin's lymphom.*

[154] Retrieved 24 november 2022

41.4 La natura empirica delle densità bimodali

Si noti che le distribuzioni bimodali tendono a ricorrere da rilevazioni di dati nella realtà sensibile, e quindi le frequenze corrispondono a probabilità in senso frequentista. Invece non è bimodale nessuno dei pur numerosi tipi di densità discrete e continue che più classicamente si definiscono a livello teorico, con formule (e parecchi sono esposti in questo testo elementare). Leggiamo ancora nella sopra citata pagina di Wikipedia:

> Bimodal distributions, despite their frequent occurrence in data sets, have only rarely been studied

BOZZA - DRAFT

Bottom line. Tuttavia, non solo a livello teorico ma anche empirico ovvero sperimentale, le distribuzioni unimodali, e più specificamente più o meno a campana, sono immensamente più frequenti di quelle multimodali, e dopo questa breve parentesi in questa Lezione, quasi di quelle sole ci occuperemo.

Nota. Di questa Lezione ci si dovrebbe aspettare che lo studente conosca a memoria le formule *contemporaneamente* numerate *et* riquadrate, oltre a tutto l'inquadramento teorico della questione.

41.5 ESERCIZI SULLA LEZIONE 39

41.5.1 Esercizio risolto a - Dado regolare a 8 facce

$\overset{f}{\mu}$ Che probabilità c'è di ottenere almeno 2 volte il numero 7 con 5 lanci di un dado regolare a 8 facce?
Potremmo sommare le 4 probabilità

$$P(k=2)+P(k=3)+P(k=4)+P(k=5)$$

ma è ovvio che conviene considerare l'evento complementare e cioè calcolare

$$1-(P(k=0)+P(k=1))=$$

$$=1-\binom{5}{0}\left(\frac{1}{8}\right)^0\left(1-\frac{1}{8}\right)^{5-0}-\binom{5}{1}\left(\frac{1}{8}\right)^1\left(1-\frac{1}{8}\right)^{5-1}=$$

$$=1-1\cdot 1\cdot\left(\frac{7}{8}\right)^5-5\cdot\left(\frac{1}{8}\right)^1\left(\frac{7}{8}\right)^4=$$

$$=1-\left(\frac{7}{8}\right)^4\cdot\left(\frac{7}{8}-\frac{5}{8}\right)=1-\left(\frac{7}{8}\right)^4\cdot\frac{2}{8}=1-\frac{2\,401}{16\,384}=$$

$$=\frac{13\,983}{16\,384}\approx 0.853=85.3\%.$$

Esercizio$_\mu$ Che probabilità c'è di ottenere 3 volte il numero 4 con 5 lanci di un dado regolare a 6 facce?
Esercizio$_\mu$ Che probabilità c'è di ottenere almeno 2 volte il numero 3 con 5 lanci di un dado regolare a 6 facce?
Esercizio$_\mu$ Che probabilità c'è di ottenere almeno 3 volte il numero 3 con 5 lanci di un dado regolare a 6 facce?
Esercizio$_\mu$ Che probabilità c'è di ottenere almeno 2 volte numeri primi con 4 lanci di un dado regolare a 6 facce?
Esercizio$_\mu$ Che probabilità c'è di ottenere un numero primo di volte testa con 5 lanci di una moneta regolare?
Esercizio$_\mu$ Che probabilità c'è di ottenere un numero primo di volte croce con 6 lanci di una moneta regolare?
Esercizio$_\mu$ Che probabilità c'è di ottenere un numero primo di volte testa con 7 lanci di una moneta regolare?
Esercizio$_\mu$ Che probabilità c'è di ottenere un numero dispari di volte testa con 5 lanci di una moneta regolare?

BOZZA DRAFT

42 Speranza matematica e varianza

La speranza matematica di una v.a. essenzialmente è un numero che congloba
la probabilità degli eventi
col loro
valore numerico (dato da una v.a.) pensabile magari come un costo ovvero guadagno.

La varianza di una v.a. estende il concetto già visto di varianza di un dataset numerico ovvero di una popolazione: è una misura della variabilità.

Consideriamo in questa Lezione variabili aleatorie discrete, e nella Lezione 45 le continue.

BOZZA - DRAFT

42.1 Speranza matematica (o valore atteso, o media)

EN: expected value, expectation, mathematical expectation, mean, average, first moment

Un esperimento ideale

Consideriamo questo esperimento ideale: estrarremo da un bicchier d'acqua distillata un atomo a caso, e sia X la variabile aleatoria che conta i suoi elettroni. L'atomo sarà di idrogeno o di ossigeno, con 1 elettrone o 8 elettroni rispettivamente.
In un'estrazione quanti elettroni possiamo attenderci mediamente? Questo è il significato di speranza matematica della variabile aleatoria X.
La speranza matematica di X è da intendersi come la media di numerosissime sue determinazioni ovvero realizzazioni. Per esempio potremmo trovare questi valori, in estrazioni successive:

1 1 1 8 1 8 1 1 8 1 1 1 8 1 1....

Naturalmente l'1 tenderà a ricorrere doppiamente dell'8 perchè l'idrogeno è doppiamente frequente nel bicchiere di H_2O rispetto all'ossigeno.
È ovvio che la media tenderà a

$$\frac{2}{3} \cdot 1 + \frac{1}{3} \cdot 8$$

cioè $\frac{10}{3} \approx 3.33$.
Ecco, questa è la speranza matematica del numero di elettroni di

un'estrazione; ovvero, il *valore atteso*, termine però fuorviante, perchè non possiamo affatto attenderci di estrarre 3.333... elettroni: ne verranno 8 oppure 1.

Un esempio con un gioco

Supponiamo che ora lanceranno un dado, e io posso scegliere, a costo 0,

— o di ricevere 2 euro se viene pari

— o di ricevere 5 euro se viene 3;

cosa mi conviene scegliere?

• Se spero nel pari, mediamente vincerei metà volte, e allora posso considerare che mediamente vincerò metà dei 2 euro in palio, cioè 1 euro;

• se spero nel numero 3, mediamente vincerei 5 euro 1 volta su 6, e allora posso considerare che mediamente vincerò 5/6 dell'euro euro in palio.

Allora conviene la scommessa sul pari perché 1 è più di 5/6.

BOZZA - DRAFT

Il caso generale

Il concetto è immensamente più generale dei giochi d'azzardo, perché non solo con le assicurazioni, ma anche coi fatti dell'economia e della salute pubblica e individuale, in ultima analisi, la situazione con

spese certe

e ricavi incerti

è analoga ad un gioco d'azzardo.

Anche i farmaci, che pure salvano tante persone e magari in casi acuti (come infarto e incidente stradale) gli regalano decenni di vita ulteriore in passato impossibili, talvolta fanno pure danni, e anche quelli vanno messi in conto, in un bilancio costi-benefici.

Si pensi, per esempio, al rofecoxib, cioè il Vioxx, che tra il 1998, e il 2004, quando fu ritirato, si stima abbia causato morti cardiache in eccesso, quasi 100.000 nell'ipotesi migliore.

Un'attenta considerazione del caso numerico soprastante ci induce a definire come **speranza matematica** (o *media* o *valore atteso*) di una variabile aleatoria discreta $X := \binom{x_1\, x_2\, x_3 \ldots x_k\, [\ldots]}{p_1\, p_2\, p_3 \ldots p_k\, [\ldots]}$

$$\boxed{\mu = E(X) := \sum_k x_k \cdot p_k \qquad \textit{somma su tutti i valori } x_k \textit{ della v.a.}} \tag{97}$$

e questa è una somma finita o una serie, e in quest'ultimo caso a livello teorico si richiede anche una condizione di regolarità, $\sum_k |x_k| p_k < +\infty$ che in questa trattazione elementare diamo ma mai calcoleremo, e comunque è sempre verificata per le variabili aleatorie che considereremo.

BOZZA - DRAFT

42.2 Costi e benefici

La speranza matematica matematizza il concetto di costi e benefici. Un intervento farmaceutico o più in generale medico su una persona ha, semplificando, un costo con una certa probabilità, e un beneficio con un'altra. Entrambi sono difficili da quantificare; in particolare: dopo quanto tempo si esegue la misura? È ovvio che un farmaco per una malattia cronica che dà un certo beneficio per un anno, e magari 0 danni, potrebbe presentare ben peggiori risultati dopo 10 anni.

Ma anche supponendo una misura perfetta, un singolo potrebbe avere preferenze ben diverse dal sistema sanitario. Consideriamo per esempio 2 terapie per una malattia che causa atroce sofferenza:

Terapia 1:

20% morte

80% guarigione

Terapia 2:

10% morte

80% passa da atroce sofferenza a grave sofferenza

10% guarisce

Il sistema potrebbe preferire la Terapia 2, che ha metà rischio di

morte e di solito (90%) migliora la situazione del paziente, ma qualcuno preferirebbe la 1, che gli dà una probabilità dell'80% di guarigione, invece che di passare da sofferenza atroce a grave.

Ancora più complessa la situazione quando il destinatario dell'intervento non è un singolo individuo ma la collettività stratificata in classi che pagano costi diversi a fronte di vantaggi diversi. Per esempio il distanziamento sociale e più in generale tutta la covidizzazione della società fa pagare un prezzo altissimo ai giovani, senza nessun apprezzabile diminuzione di rischio di morte per loro, che in pratica hanno solo costi; l'ipotetico vantaggio è praticamente solo per altre classi di età, più avanzate.

È ben evidente che poi la questione si intreccia coi vantaggi non sanitari, bensì economici (ed eventualmente pure di altro genere), che particolari segmenti della società (numericamente minuscoli), possono ottenere da grandi interventi sanitari.

BOZZA DRAFT

In generale per tutti gli interventi sulla società in ogni caso ci sono costi e benefici, ma quasi mai i benefici sono esattamente per le stesse persone che pagano i costi.
A questo mondo tutto si fa per un presunto beneficio superiore al danno previsto, ma i soggetti destinatari sono sempre diversi.
Coi media quelli che hanno soldi e potere possono convincere la maggioranza che quello che si fa è vantaggioso per tutti, e infatti certamente è vantaggioso – ma più per loro.

Un esempio eclatante, collegato, è il grande fenomeno umano della guerra, che tanto incide sulle cause di morte a livello mondiale. Si fanno delle stime di possibili guadagni e perdite, più o meno probabili, e infine si decide:

"Armiamoci e partite!"

Nessuna legge al mondo vieta di sognare che i fatti della politica, e

delle politiche sanitarie, avvengano in base a criteri diversi.

42.3 Varianza

Se esiste la speranza matematica, si definisce la **varianza** di una v.a. discreta X in 2 modi equivalenti

$$\boxed{\sigma^2 = Var(X) := \sum_k (x_k - E(X))^2 \cdot p_k = E(X^2) - E(X)^2} \qquad (98)$$

e la radice quadrata si dice **deviazione standard**, σ o SD, che ha la stessa unità di misura della variabile aleatoria.

Per esempio le giocate sul pari e sul 3, prima considerate, producono guadagni medi rispettivi

$$E(U) = \frac{1}{2} \cdot 2 = 1 \qquad E(V) = \frac{1}{6} \cdot 5 = \frac{5}{6} \approx 0.833.$$

Ma con diversa varianza, che in qualche modo esprime la tendenza a rischiare molto per avere molto: la si calcoli nei 2 casi con le formule soprastanti.

BOZZA DRAFT

Teorema 1.

Legge	$E(X)$ μ	$Var(x)$ σ^2
$X \sim \mathbb{U}\{a, b\}$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2-1}{12}$
$X \sim B(n, p)$	$n \cdot p$	$np(1-p)$
X geometrica iniziante da 0	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
X geometrica iniziante da 1	$\frac{1}{p}$	$\frac{1-p}{p^2}$

Esercizio$^{f}_{\mu}$ Qual è il valore atteso del punteggio di un dado? E la varianza?
Usando la definizione di speranza matematica si ha

$$X := \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix} \quad E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} =$$

$$= \frac{1}{6} \cdot (1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6} = \frac{7}{2} = 3.5.$$

Usando il teorema sulla speranza matematica di $\mathbb{U}\{a, b\}$

$$X \sim \mathbb{U}\{1, 6\} \Rightarrow E(X) = \frac{1+6}{2} = \frac{7}{2} = 3.5 \, (come\ sopra).$$

Con le 2 definizioni di varianza si ha ugualmente

$$Var(X) = \left(1 - \frac{7}{2}\right)^2 \cdot \frac{1}{6} + \left(2 - \frac{7}{2}\right)^2 \cdot \frac{1}{6} + ... + \left(6 - \frac{7}{2}\right)^2 \cdot \frac{1}{6} = ... = \frac{35}{12}$$

$$X^2 = \begin{pmatrix} 1 & 4 & 9 & 16 & 25 & 36 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix} \quad Var(X) = 1^2 \cdot \frac{1}{6} + ... + 6^2 \cdot \frac{1}{2} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

e con il teorema sulla varianza di $\mathbb{U}\{a, b\}$ si ha ugualmente

$$X \sim \mathbb{U}\{1, 6\} \Rightarrow Var(X) = \frac{(6 - 1 + 1)^2 - 1}{12} = \frac{36 - 1}{12} = \frac{35}{12}.$$

BOZZA - DRAFT

42.4 Note su varianza e istogramma della densità

Grande varianza ovvero deviazione standard: valori dispersi.
Piccola varianza ovvero deviazione standard: valori addensati presso la media.

Per la varianza, esistono varianze più piccole (valori più addensati) o più grandi di altre (valori più dispersi) ma non ha molto senso considerare grande o piccola una singola varianza.
Quanto sopra detto vale per qualunque densità discreta (e in effetti, poi, varrà anche per le continue). (Anche se “bimodale").

Se l'istogramma di una densità discreta (o, in seguito, il grafico di una densità continua) ha una forma più o meno a campana, com'è per le variabili aleatorie binomiali,

a grande varianza corrisponde campana bassa e larga

a piccola varianza corrisponde campana alta e stretta.

Quest'osservazione vale anche per l'uniforme discreta e le geometriche, che, con qualche elasticità, si può pure dire che abbiano densità a forma più o meno a campana – seppure molto vagamente. E similmente varrà per l'uniforme continua, l'esponenziale, e molte altre.

BOZZA - DRAFT

42.5 L'aspettativa di vita o speranza di vita

Consideriamo un esserino robotizzato che viene creato il 1 gennaio, e il 31 dicembre lancia al suo interno un dado e se viene 1 si spegne, muore. Se è sopravvissuto, il 31 dicembre dell'anno dopo fa la stessa cosa e così via. La durata della sua vita è una variabile aleatoria di cui possiamo considerare la speranza matematica. In questo caso semplice, è una variabile aleatoria geometrica iniziante da1 di parametro $\frac{1}{6}$ e allora con speranza matematica 6, che è l'aspettativa di vita dell'esserino, in anni.

Possiamo ipotizzare un esserino automatico che lanci due dadi e una moneta e cessa di vivere solo con 1, 1, testa, che ha probabilità $\frac{1}{72}$, e allora l'esserino ha vita media 72 anni, e l'aspettativa di vita nel mondo è più o meno quella.

È un modello della situazione relativa agli esseri umani, ma con varie semplificazioni:

– l'essere umano può morire ogni giorno, ovviamente in generale con probabilità molto minore di $\frac{1}{72}$;

– la probabilità di morte non è costante negli anni: è molto maggiore nel primo anno che nel secondo, e in vecchiaia aumenta enormemente: un centenario in generale ha probabilità molto minore di $\frac{71}{72}$ di vivere un altro anno;

– la probabilità di morte ad un determinato anno di età varia

globalmente nel tempo: un diciottenne del 2018 aveva molta più probabilità di arrivare a 19 anni dei diciottenni del 1918, soggetti a Prima Guerra Mondiale – e altre difficoltà di quel tempo.
Con riferimento all'ultimo punto, come si fa a calcolare la speranza di vita, se non sappiamo quali guerre eventualmente ci saranno, e altre catastrofi, o più in generale cambiamenti – anche migliorativi, ovvio – della sopravvivenza?
È assolutamente impossibile calcolarla, e non ha neppure alcun senso reale la probabilità che una persona vivente oggi sia viva fra 1 anno, senza sapere neppure se il mondo esisterà ancora. (Come sarebbe definita una tale probabilità? Non ha alcun senso). Quello che viene chiamato aspettativa o speranza di vita, detto molto semplificatamente, è il risultato di un calcolo in cui si considera tutto "a bocce ferme", cioè come se le cose dovessero andare come sono andate finora. (Per i ventenni, i ventunenni, i ventiduenni, eccetera, per tutte le classi di età). *Non è una vera speranza matematica* in senso stretto, ma in qualche modo la "imita".

BOZZA - DRAFT

Oltre all'aspettativa di vita alla nascita, si può considerare l'aspettativa di vita residua, ad una determinata età. Per gli esserini robotizzati, è uguale a quella alla nascita, perché la variabile aleatoria geometrica gode dell'*assenza di memoria*: restano sempre da vivere mediamente 6 o 72 anni nei due casi considerati. (E similmente costante è il tempo da attendere ancora un determinato numero, o ambo, eccetera, del lotto, anche se "in ritardo", contrariamente a quello che pensano ingenui giocatori). Per l'essere umano ovviamente non è così, per il variare della mortalità al progredire dell'età. In particolare in Italia l'aspettativa di vita a 65 anni ha un importante valore legale e infine sociale perché a quel valore, determinato dall'ISTAT, Istituto Nazionale di Statistica, è per legge legata la cosiddetta *età pensionabile*.

Sulla Gazzetta Ufficiale della Repubblica Italiana, Serie Generale - Anno 160, Numero 267, del 14 novembre 2019, nel Decreto 5 novem-

bre 2019 del Ministero dell'Economia e delle Finanze, leggiamo: "Vista la nota del Presidente dell'Istituto nazionale di statistica (ISTAT) n. 2768968/19 del 16 ottobre 2019, con cui si comunica che la variazione della speranza di vita all'età di sessantacinque anni e relativa alla media della popolazione residente in Italia ai fini dell'adeguamento dei requisiti di accesso al pensionamento con decorrenza 1° gennaio 2021 corrispondente alla differenza tra la media dei valori registrati negli anni 2017 e 2018 e il valore registrato nell'anno 2016 è pari a 0,021 decimi di anno".
(Si tratta, ovviamente, di questione connessa anche a fattori medici e farmaceutici, oltre che sociali, politici e geopolitici; la spesa farmaceutica statale è in enorme diminuzione negli ultimi anni, pare ridotta del 70 percento in un decennio).

BOZZA - DRAFT

42.6 Passeggiate aleatorie

Traiamo da Wikipedia, l'enciclopedia libera, versione inglese alla voce "Random walk", in italiano *passeggiata aleatoria*":

> A random walk is a mathematical object, known as a stochastic or random process, that describes a path that consists of a succession of random steps on some mathematical space such as the integers. An elementary example of a random walk is the random walk on the integer number line, $\mathbb{Z}$, which starts at 0 and at each step moves +1 or −1 with equal probability. Other examples include the path traced by a molecule as it travels in a liquid or a gas, the search path of a foraging animal, the price of a fluctuating stock and the financial status of a gambler can all be approximated by random walk models, even though they may not be truly random in reality.
> (...) If a and b are positive integers, then the expected number of steps until a one-dimensional simple random walk starting at 0 first hits b or $-a$ is ab.

Cioè detto T il numero di passi fino al raggiungimento di b o $-a$,

$$\boxed{E(T) = a \cdot b} \tag{99}$$

Traiamo da https://commons.wikimedia.org/wiki/File:Random_Walk_example.svg la seguente figura, che rappresenta (la parte iniziale di) 8 cammini aleatori. ???FIGURA VEDO TEX

Ecco su WolframAlpha una passeggiata aleatoria di 100 passi, nuova ad ogni richiamo del LINK->

Su queste basi ipotizziamo ora un “parametro vitale complessivo", che non riusciamo in questo esempio semplificato a specificare nei dettagli, ma che comunque non deve superare il valore soglia 7, nè scendere sotto il valore soglia -7, variando di anno in anno nell'individuo, salendo o scendendo di 1 con uguale probabilità 50%. Possiamo proprio immaginare un robottino che ogni 31 dicembre lancia una moneta e somma 1 se viene testa o sottrae 1 se viene croce ad un suo “capitale virtuale", partito da 0, e muore ovvero si spegne se sfora i limiti prestabiliti. In base a quanto sopra detto, con $a = b = 8$, la speranza matematica della vita è 64 anni, cioè $8 \cdot 8$. Ma se ora (come in un videogioco si danno le vite supplementari) supponiamo che la prima volta che una soglia (8 o -8) viene raggiunta, un farmaco modifica di 1 o -1 il parametro vitale nella direzione giusta, ecco che al nostro ipotetico soggetto la Farmacia regala altri anni di vita! Superato il momento critico, quando rischiava di andare a 8 o -8 e morire, ricomincia una nuova vita; certo la ricomincia in posizione rischiosa, a 7 o -7, e al passo successivo può morire (e non è detto che gli sia concessa una seconda salvezza), ma non è detto, magari si allontana dal limite invalicabile e sopravvive ancora a lungo.
Questo può essere visto come uno dei significati della Farmacia: la cura urgente del caso acuto (con qualche speranza che il soggetto si allontani poi dalla situazione limite).

BOZZA - DRAFT

Oltre all'aspettativa di vita alla nascita, la mortalità nel primo anno è l'altro grande parametro usato per valutare lo stato di salute complessivo di una popolazione.
Leggiamo su Wikipedia, l'enciclopedia libera:

> Of the 27 most developed countries, the U.S. has the highest Infant Mortality Rate, despite spending much more on health care per capita.

La mortalità infantile in sostanza è una probabilità, in senso frequentista, mentre l'aspettativa di vita alla nascita è una sorta di speranza matematica.

BOZZA - DRAFT

42.7 Esercizi sulla speranza matematica di v.a. discrete

Esercizio$^{f}_{\mu}$ Qual è la vincita netta attesa giocando 1 euro su una cinquina del lotto? (Con la parola "netta" intendiamo che conteggiamo la perdita sicura di 1 euro, il costo della giocata).
Detta V la v.a. "vincita netta", ricordando che si vince 6 milioni di volte la giocata, con probabilità $1/\binom{90}{5} = 1/43\,949\,268$,

$$V = \begin{pmatrix} 5\,999\,999 & -1 \\ \frac{1}{43\,949\,268} & \frac{43\,949\,267}{43\,949\,268} \end{pmatrix}$$

$$E(V) = 5\,999\,99 \cdot \frac{1}{43\,949\,268} - \frac{43\,949\,267}{43\,949\,268} = -\frac{37\,949\,268}{43\,949\,268} =$$

adesso magari semplifichiamo per 2, poi ancora per 2, poi per 3

$$= -\frac{3\,162\,439}{3\,662\,439} \approx -0.8635\,euro$$

cioè una perdita media di circa 86 centesimi di euro. (In generale i giochi d'azzardo sono molto meno svantaggiosi di questo).

Esercizio$^{f}_{\mu}$ Qual è la vincita netta attesa giocando 1 euro su un numero della roulette europea?
La vincita avviene con probabilità $1/37$ perché i numeri equiprobabili sono 37, da 0 a 36, e dà 36 volte la giocata:

$$V = \begin{pmatrix} 35 & -1 \\ \frac{1}{37} & \frac{36}{37} \end{pmatrix}$$

$$E(V) = 35 \cdot \frac{1}{37} - \frac{36}{37} = -\frac{1}{37} \approx 0.027\, euro$$

cioè mediamente perdiamo, molto approssimativamente, 3 centesimo ad ogni giocata di 1 euro.

Nota. Di questa Lezione ci si dovrebbe aspettare che lo studente conosca a memoria le formule *contemporaneamente* numerate *et* riquadrate, e le 4 formule della speranza matematica del Teorema 1, oltre a tutto l'inquadramento teorico della questione.

BOZZA - DRAFT

IX – Variabili aleatorie continue

BOZZA - DRAFT

43 Uniforme continua, di Cuchy, esponenziale

Come abbiamo visto, una v.a. è una funzione $X : \Omega \to \mathbb{R}$ che ad eventi associa numeri e si dice continua se i valori riempiono almeno un intervallo.
Per esempio il peso del primo neonato del prossimo anno, misurato con “infiniti decimali" (numero reale, con precisione infinita), ovviamente da considerarsi prima dell'evento.
Ogni singolo valore ha probabilità nulla: $P(X = t) = 0, \forall\, t$.

In questa trattazione elementare, le v.a. continue che considereremo sono **molto regolari**: **la funzione di ripartizione**

$$F_X(z) := P(X \leq z)$$

– **è continua su tutto** $\mathbb{R}$
– **e derivabile salvo al più in 1 o 2 punti**.

BOZZA - DRAFT

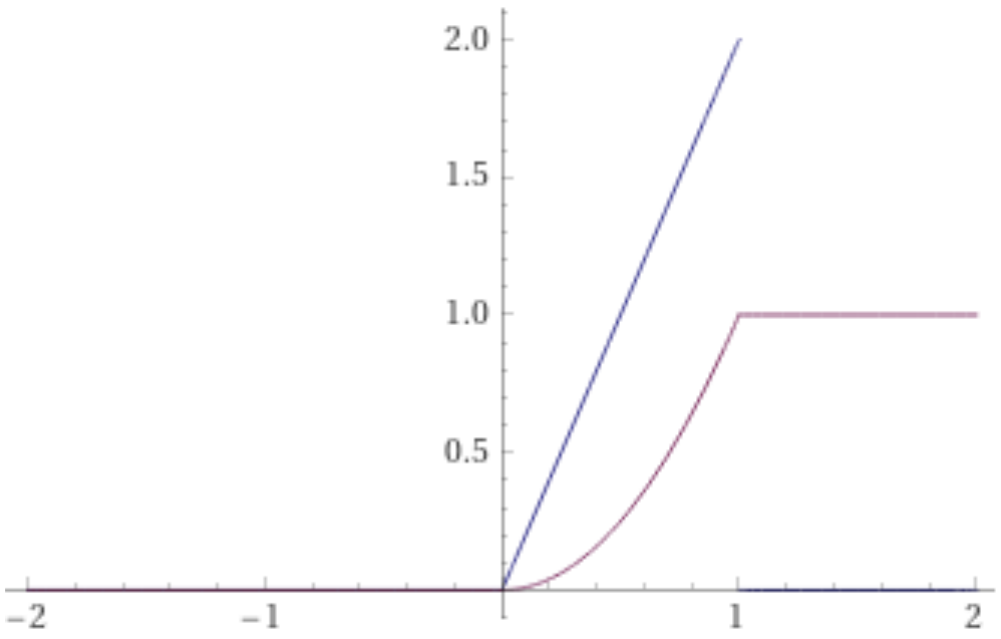


Figure 48: Blu: densità con 1 punto di discontionuità. Magenta: sua funzione di ripartizione con 1 punto di non derivabilità (ma comunque continua). Screenshot da WolframAlpha.

La derivata della f.r., completata ad arbitrio in quegli eventuali punti particolari ovvero *singolari*,

$$f(z) := F'_X(z) = DP(X \le z) \qquad \textit{spesso denotata } f_X\textit{, con argomento z,t,u,x...}$$

si chiama **densità** della variabile aleatoria.

Per fissare le idee nella mente, si immagini

la funzione di ripartizione come una curva sigmoide che cresce da 0 a 1

la densità come una funzione a campana nel senso più lato possibile con asintoto l'asse delle ascisse.

Variazioni sono possibili nei singoli casi ma intanto fissiamo nella mente qualcosa da immaginare per la generalità dei casi:

$$\text{f.r.} \leftrightarrow \text{sigmoide} \qquad \textit{tipo arcotangente ma riscalata fra } 0 \textit{ e } 1$$

$$\text{densità} \leftrightarrow \text{campana}$$

BOZZA - DRAFT

Una funzione f è densità di una v.a. se e solo se

$$f(t) \ge 0 \quad \wedge \quad \int_{-\infty}^{+\infty} f(t)\,dt = 1.$$

Se f e F_X sono rispettvamente densità e f.r. di una variabile aleatoria continua X, allora (teorema)

$$F_X(x) = \int_{-\infty}^{x} f(t)\,dt \qquad \forall\, x$$

$$P(X = x) = 0 \qquad \forall\, x$$

$$P(a \le X \le b) = \int_a^b f(t)\,dt \qquad \forall\, a, b, \quad \textit{ogni} \le \textit{ sostituibile con } <$$

e in quest'ultima a e b possono essere infiniti (col <, ovvio).

43.1 Variabile aleatoria uniforme continua

Consideremo qua la v.a. $\sim \mathbb{U}[a,b]$ oppure $\mathbb{U}(a,b)$, uniforme su $[a,b]$.

$$f(t) := \begin{cases} \frac{1}{b-a} & se\ a \le t \le b \\ 0 & altrimenti \end{cases} \qquad F_X(x) = \begin{cases} 0 & se\ x < a \\ \frac{x-a}{b-a} & se\ a \le x \le b \\ 1 & se\ x > b \end{cases} .$$

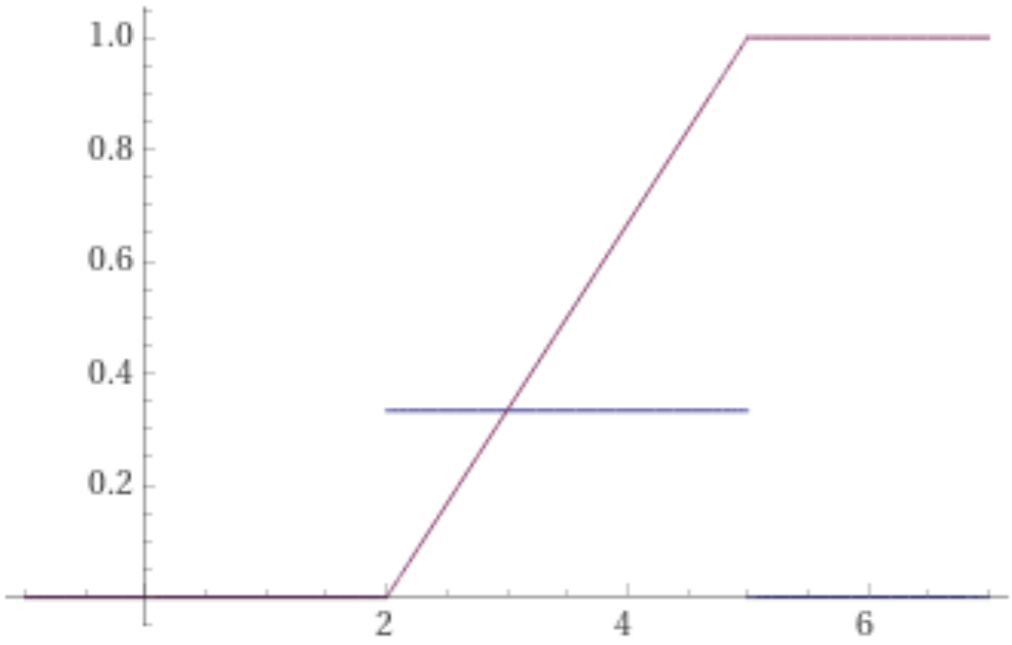


Figure 49: Blu: densità uniforme continua di parametri 2 e 5. Magenta: sua f.r.. Screenshot da WolframAlpha.

La densità ha 2 punti di discontinuità e corrispondentemente la f.r. ha 2 punti di non derivabilità.

43.2 Variabile aleatoria di Cauchy.

$$f(t) := \frac{1}{\pi} \cdot \frac{1}{1+t^2}$$

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan x$$

La distribuzione di Cauchy ha in Farmacia almeno 5 utilità:

– fornisce un modello di densità a campana di cui si possono calcolare i valori con le 4 operazioni

– fornisce un modello di densità a campana di cui si possono calcolare gli integrali con funzioni elementari (l'arcotangente)

– fornisce un esempio di variabile aleatoria priva di speranza matematica

– permette di chiarire, nel confronto con la v.a. normale, la questione delle *code pesanti*

– è usata in Spettroscopia: LINK->

43.3 Variabile aleatoria esponenziale

Queste sono densità e funzione di ripartizione di una v.a. X esponenziale di parametro $\lambda > 0$:

$$f(t) := \begin{cases} \lambda\, \mathrm{e}^{-\lambda t} & \text{se } t \geq 0 \\ 0 & \text{altrimenti} \end{cases} \qquad F_X(x) = \begin{cases} 1 - \mathrm{e}^{-\lambda x} & \text{se } x \geq 0 \\ 0 & \text{altrimenti} \end{cases}$$

ZA - DRAFT

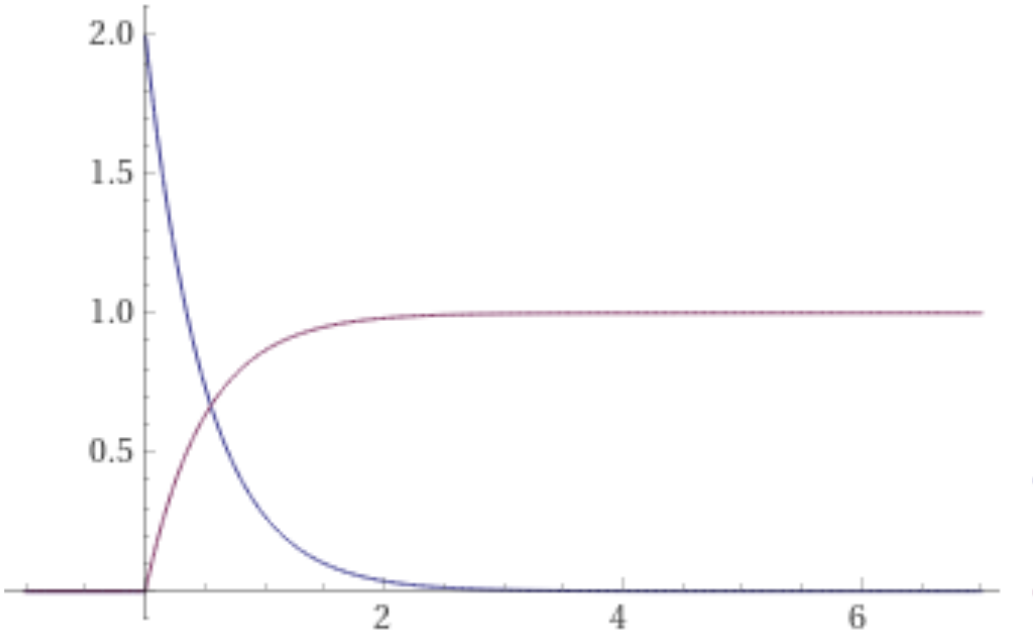


Figure 50: Blu: densità esponenziale di parametro 2. Magenta: sua f.r.. Screenshot da WolframAlpha.

La densità esponenziale modellizza molte cose fra cui l'*intertempo* fra 2 chiamate telefoniche a una farmacia o a un qualunque servizio in un tempo del giorno in cui esse arrivano in ogni minuto con la stessa probabilità: Leggiamo sulla Wikipedia in inglese alla voce *Exponential distribution*:

> if we focus on a time interval during which the rate is roughly constant, such as from 2 to 4 p.m. during work days, the exponential distribution can be used as a good approximate model for the time until the next phone call arrives.

Ecco 11 valori (di una simulazione informatica) della variabile aleatoria esponenziale di parametro 3.5, con 3 cifre decimali:

0.163 0.190 0.695 0.132 0.979 0.157 0.023 0.236 0.291 0.113 0.015

Si notino i valori piccoli e i valori grandi, distribuiti in un modo tipico, caratterizzante appunto la variabile aleatoria esponenziale.

Ecco online su WolframAlpha un po' di determinazioni, ad ogni reload nuove, (di una simulazione informatica) della variabile aleatoria esponenziale di parametro 3.5: `link->`

La distribuzione esponenziale ha un'ampia ricorrenza nelle Scienze Applicate. Traiamo da Wikipedia in inglese, l'enciclopedia libera, alla voce

> The exponential distribution occurs naturally when describing the lengths of the inter-arrival times in a homogeneous Poisson process.
>
> The exponential distribution may be viewed as a continuous counterpart of the geometric distribution, which describes the number of Bernoulli trials necessary for a

BOZZA DRAFT

discrete process to change state. In contrast, the exponential distribution describes the time for a continuous process to change state.

In real-world scenarios, the assumption of a constant rate (or probability per unit time) is rarely satisfied. For example, the rate of incoming phone calls differs according to the time of day. But if we focus on a time interval during which the rate is roughly constant, such as from 2 to 4 p.m. during work days, the exponential distribution can be used as a good approximate model for the time until the next phone call arrives. Similar caveats apply to the following examples which yield approximately exponentially distributed variables:

• The time until a radioactive particle decays, or the time between clicks of a Geiger counter

• The time it takes before your next telephone call

• The time until default (on payment to company debt holders) in reduced form credit risk modeling

Exponential variables can also be used to model situations where certain events occur with a constant probability per unit length, such as the distance between mutations on a DNA strand, or between roadkills on a given road.

In queuing theory, the service times of agents in a system (e.g. how long it takes for a bank teller etc. to serve a customer) are often modeled as exponentially distributed variables. (...)

BOZZA - DRAFT

Per ciacuno di quei fenomeni abbiamo qualcosa del tipo

$$P(\text{dura meno di } x) = 1 - \mathrm{e}^{-\lambda x}$$

con un parametro λ caratteristico di quel fenomeno.
In particolare

$P(\text{dura meno di } 0) = 0$ cioè almeno un istante dura.

Abbiamo introdotto la speranza di vita con la variabile aleatoria geometrica, col robottino che può “morire" solo all’ultimo giorno dell’anno. La variabile aleatoria esponenziale è in qualche modo il corrispondente continuo della variabile aleatoria geometrica, e tiene conto della possbilità di morte – o di guasto per gli apparati tecnologici, dai motori alle lampadine ai fusibili – in ogni momento. Entrambe le variabili aleatorie godono dell’assenza di memoria: resta da vivere sempre la stessa quantità di tempo, mediamente. Esattamente come per la variabile aleatoria geometrica, c’è però da dire che sia per gli esseri viventi che per le macchine, la mortalità non è costante: è molto maggiore presso la nascita, e nella vecchiaia. Entrambi i modelli allora sono molto approssimativi, seppure danno un’idea della questione.

BOZZA - DRAFT

44 Quantili delle variabili aleatorie continue

44.1 Definizione semplificata

Per ogni variabile aleatoria **continua** X di interesse pratico in Statistica Medica e nelle Scienze della Farmacia

il quantile $q_{0.95}$ di ordine 0.95
ovvero ovvero 95%
di X ovvero della sua densità f
è l'unico numero tale che

$$P(X \leq q_{0.95}) = 0.95 \tag{100}$$

e similmente a 0.95 vale per
ogni numero fra 0 e 1 esclusi

BOZZA DRAFT

ed è allora soluzione dell'equazione nell'incognita $q_{0.95}$

$$F_X(q_{0.95}) = 0.95$$

cioè in definitiva ha il valore (teorico, spesso difficile da calcolare)

$$q_{0.95} = F_X^{-1}(0.95)$$

(e similmente a 0.95 vale per ogni numero fra 0 e 1 esclusi).
(Esistono anche i quantili delle variabili aleatorie discrete ma non ce ne occupiamo).

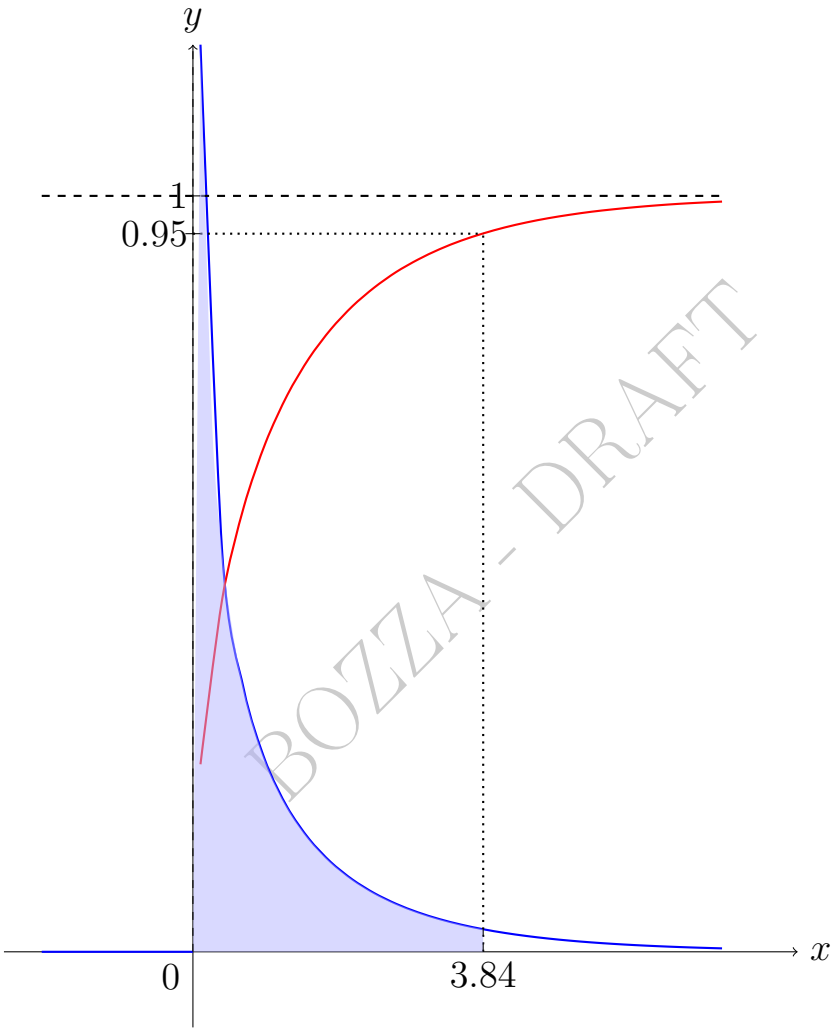
44.2 Mediana di variabile aleatoria continua

$$\boxed{\text{mediana} := q_{0.5}} \tag{101}$$

che in pratica ha il 50% di lasciare alla sua sinistra un valore casuale tratto dalla variabile aleatoria considerata (e ovviamente 0.5=50% ha sostituito nella (100) il valore 0.95).

44.3 Esempio di quantile classico

Considereremo una classica densità (in effetti è la densità del chi-quadrato a 1 grado di libertà, la cui definizione[155] è in nota).



BOZZA - DRAFT

È ombreggiata (purtroppo con una lacuna in alto) l'area da 0 a 3.84, che è il quantile di ordine 0.95, e prima di 3.84 c'è il 95% dell'area 1 del sottografico.
Questo valore 3.84 è un classico della Statistica inferenziale.

[155] Densità del chi quadrato a 1 grado di libertà:

$$f(x) := \begin{cases} \frac{e^{-\frac{x}{2}}}{\sqrt{2\pi}\sqrt{x}} & se\ x > 0 \\ 0 & se\ x \leq 0 \end{cases} \tag{102}$$

Resto della Lezione escluso dal corso breve – 40 ore.

44.4 Definizione più completa di quantile

Se l'insieme dove f è positiva è un intervallo, anche se illimitato, allora (si dimostra) per ogni $\alpha \in]0,1[$ esiste un'unica soluzione dell'equazione

$$F_X(t) = \alpha$$

che si chiama *quantile di ordine* α, indicato talvolta con q_α:

$$P(X \le q_\alpha) = \alpha.$$

Esempio. Per una legge esponenziale di parametro 3 calcolare $P\Big(X \ge \frac{2}{3}\ln 2\Big)$.
La densità è

$$f(x) := \begin{cases} 3\mathrm{e}^{-3x} & \textit{se } x \ge 0 \\ 0 & \textit{altrimenti} \end{cases}$$

e allora

$$P\Big(X \ge \frac{2}{3}\ln 2\Big) = \int_{\frac{2}{3}\ln 2}^{+\infty} f(t)\,dt = \int_{\frac{2}{3}\ln 2}^{+\infty} 3\mathrm{e}^{-3t}\,dt = \star$$

calcoliamo l'integrale indefinito con la (85) con $z := -3t + 0$

$$\int 3\mathrm{e}^{-3t}\,dt = \frac{1}{-3}\Big(\int 3\mathrm{e}^{z}\,dz\Big)_{z=-3t} = -\frac{1}{3}\big(3\mathrm{e}^{z}\big)_{z=-3t} = -\mathrm{e}^{-3t}$$

$$\text{riprendiamo } \star = \Big[-\mathrm{e}^{-3t}\Big]_{\frac{2}{3}\ln 2}^{+\infty} = -\mathrm{e}^{-\infty} - \Big(-\mathrm{e}^{-3\frac{2}{3}\ln 2}\Big) =$$

$$\textit{intendendendo } \mathrm{e}^{-\infty} \textit{ nel senso del limite e allora } 0 \quad = 0 + \mathrm{e}^{-2\ln 2} = \mathrm{e}^{\ln 2^{-2}} = 2^{-2} =$$

$$= \frac{1}{4} = 0.25 = 25\%$$

(e allora $\frac{2}{3}\ln 2$ è il quantile di ordine $\frac{1}{4}$ ossia 0.25 ossia 25%).
Calcolando la probabilità p dell'evento complementare $X < \frac{2}{3}\ln 2$, e poi facendo $1 - p$, si potevano evitare gli infiniti.
Esercizi. Trovare le mediane dell'uniforme e dell'esponenziale. [Si troverà $q_{0.5} = \frac{a+b}{2}$ e $q_{0.5} = \frac{1}{\lambda}\ln 2$ rispettivamente].

BOZZA - DRAFT

Esercizi. Trovare i quartili di ordine $\frac{1}{4}$ e $\frac{3}{4}$ dell'uniforme e dell'esponenziale. Per un'esponenziale di parametro 2 calcolare $P(X \geq 3)$. (Si può calcolare come $1 - P(X < 3)$ evitando l'infinito).

44.5 Proprietà di densità e funzioni di ripartizione

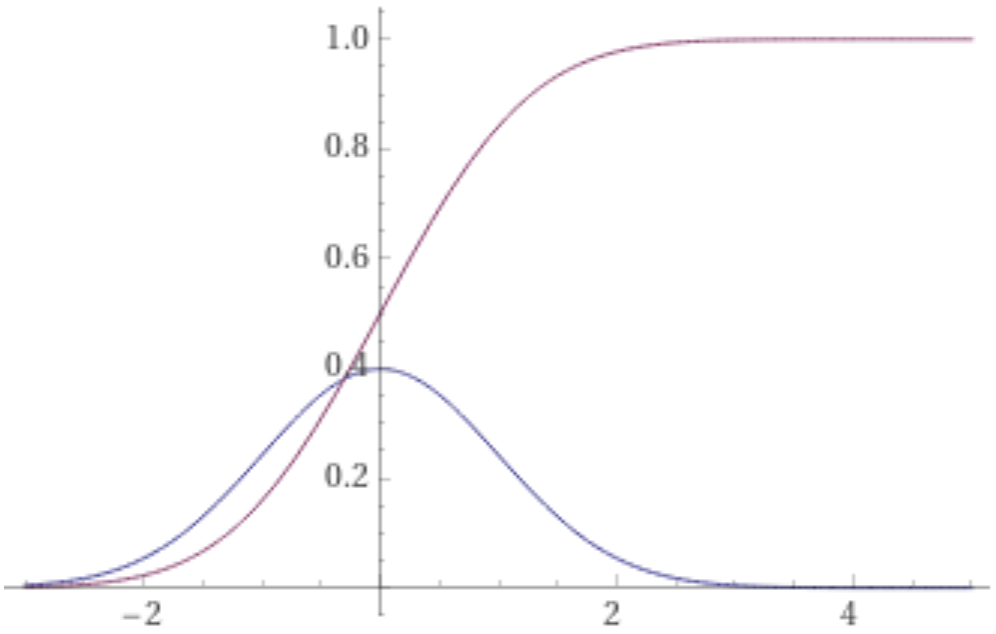


Figure 51: Blu: densità normale standard. Magenta: sua f.r.. Screenshot da WolframAlpha.

Proprietà di ogni f.r. $F(x)$ di v.a. continua:
1) ha valori fra 0 e 1, cioè $0 \leq F(x) \leq 1$;
2) è non decrescente: $x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$;
3) tende a 0 per $x \to -\infty$;
4) tende a 1 per $x \to +\infty$;
5) è continua (in trattazioni di livello superiore continua a destra);
6) dove la densità $f = F'$ è positiva[156] la f.r. è crescente puntualmente.

[156] Dove f vale 0 la f.r. può essere puntualmente crescente oppure no.

Proprietà di ogni densità $f(x)$ di v.a. continua:
1) ha valori ≥ 0, sappiamo;
2) ha integrale 1 su tutto $\mathbb{R}$, sappiamo; e poi si dimostra che
3) è continua salvo al più 1 o 2 punti (in questa trattazione).

3 bis) Tutte le densità che considereremo in questa trattazione tendono a 0 per $x \to -\infty$ e $x \to +\infty$.

Ma densità particolarmente "capricciose", considerabili a un livello superiore di questo testo elementare, potrebbero non avere uno o entrambi quei limiti, con infinite oscillazioni con "campate" sempre più strette, come nella figura sottostante.

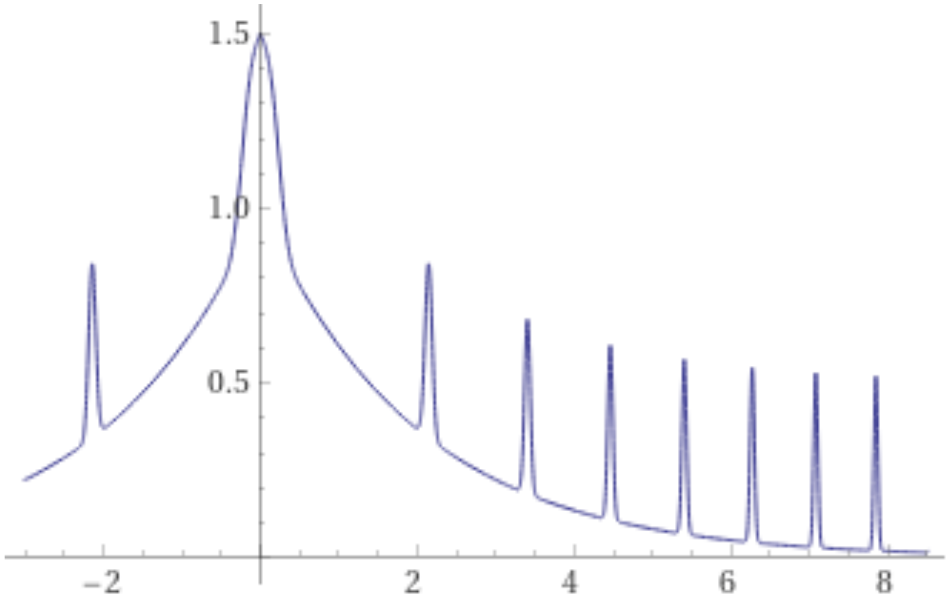


Figure 52: Densità "capricciosa" del tipo che ci proponiamo di non considerare: non tende a 0 per $x \to +\infty$. Screenshot da WolframAlpha.

Esercizi.

•• I principali quantili considerati in statistica sono:
i *quartili* $q_{0.25}$, $q_{0.5}$, $q_{0.75}$, e il secondo è la mediana;
i *decili* $q_{0.1}$, $q_{0.2}$,...,$q_{0.9}$, corrispondenti a 10%, 20%,...,90%;
i *centili* o *percentili* $q_{0.01}$, $q_{0.02}$,...,$q_{0.99}$, corrispondenti a 1%, 2%,...,99%;
due *ventili*, precisamente $q_{0.05}$ e $q_{0.95}$.
Si calcolino relativamente alla densità esponenziale.
Per esempio per i decili si troverà

$$q_{k/10} = \frac{1}{\lambda} \ln \frac{10}{10-k} \qquad k = 1, ..., 9$$

(Si noti che per $k = 5$ si ottiene la mediana).

•• Per una v.a. X di densità esponenziale di parametro 5 calcolare $P(X < 2)$, $P(X^3 > 2)$, $P(X^2 \geq 4)$.

•• Si consideri una v.a. Z di densità

$$f(z) := \begin{cases} 6z(1-z) & se\ 0 \leq x \leq 1 \\ 0 & altrimenti \end{cases} .$$

Senza fare calcoli ma col disegno del grafico stabilire quanto vale la mediana e mettere in ordine crescente

$$P(Z < 0.4), \quad P(Z > 0.9), \quad P(Z \leq 0.2), \quad P(Z \geq 0.7).$$

BOZZA DRAFT

45 Speranza matematica, varianza, covarianza

45.1 Sulle definizioni di distribuzione e legge

Sia per variabili aleatorie discrete che continue, *distribuzione* è sinonimo di *legge*, e indica sia la *densità* che la *funzione di ripartizione* (e in via teorica anche ogni altra funzione ad esse equivalente ma in pratica useremo quelle due).

45.2 Introduzione e definizioni

Definizioni. Consideriamo una v.a. X continua di densità f.

$$\textbf{Speranza matematica: } \mu = E(X) := \int_{-\infty}^{+\infty} x\, f(x)\, dx$$

se l'integrale con $|x|$ esiste finito, che non verificheremo mai.

BOZZA - DRAFT

Del tutto analoga alla definizione per una v.a. discreta:

$$\mu = E(X) := \sum_k x_k \cdot p_k \qquad \textit{somma su tutti i valori } x_k \textit{ della v.a.}$$

Si noti che la sommatoria (finita o serie) del caso discreto è diventata un integrale definito nel caso continuo. E la densità discreta p_k è stata sostituita dalla densità continua $f(x)$; però si faccia attenzione che ogni singolo valore p_k è un valore di probabilità, quindi fra 0 e 1, per esempio $\frac{1}{6}$, mentre i valori di $f(x)$ non sono probabilità e possono essere maggiori di 1. La densità della Figura 43.3 raggiunge il valore 2.

$$\textbf{Varianza: } \sigma^2 = Var(X) := \int_{-\infty}^{+\infty} (x-\mu)^2\, f(x)\, dx = E(X^2) - E(X)^2$$

Del tutto analoga alla definizione per una v.a. discreta:

$$\sigma^2 = Var(X) := \sum_k (x_k - E(X))^2 \cdot p_k = E(X^2) - E(X)^2$$

Sia per v.a. discrete che continue:

$$\textbf{Deviazione standard: } \sigma := \sqrt{Var(X)} = \sqrt{\sigma^2}.$$

$$\textbf{Covarianza: } Cov(X,Y) := E(X \cdot Y) - E(X) \cdot E(Y)$$

e se la covarianza è 0 le v.a. si dicono *incorrelate*.

Diremo **indipendenti** 2 variabili aleatorie X e Y continue se informazioni sui valori assunti da una non modificano le nostre conoscenze probabilistiche sui valori dell'altra, ovvero e per ogni a, x, c, y con $a < x$ e $c < y$

$$P(a < X < x \wedge c < Y < y) = P(a < X < x) \cdot P(c < Y < y)$$

cioè $\{a < X < x\}$ e $\{c < Y < y\}$ sono eventi indipendenti.
Spesso le variabili aleatorie incorrelate sono indipendenti.

Teoremi.

Legge	$E(X)$ μ	$Var(X)$ σ^2
$X \sim \mathbb{U}[a,b]$ ovvero $\mathbb{U}(a,b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
X esponenziale di parametro λ	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$X \sim$Cauchy	$\nexists$	$\nexists$

BOZZA - DRAFT

Detto semplificatamente, la densità di Cauchy non ha la speranza matematica perchè ha le code troppo pesanti, stenta a tendere a 0 a $\pm\infty$ (si veda la Figura 37.2) e ciò rende inesistente l'integrale da $-\infty$ a $+\infty$ di $|x| f(x)$.

45.3 Approfondimenti sulle v.a. discrete e continue

Tutte le cose che diremo qua valgono sia per variabili aleatorie discrete che continue.

Con c intendiamo un qualunque numero reale. E σ := SD := $\sqrt{Var(X)}$.

Relazioni di 1 variabile aleatoria con 1 costante:

$$E(c+X) = c + E(X) \tag{103}$$

$$E(c \cdot X) = c \cdot E(X) \tag{104}$$

$$Var(c+X) = Var(X) \tag{105}$$

$$Var(c \cdot X) = c^2 \cdot Var(X) \tag{106}$$

$$\mathrm{SD}(c \cdot X) = |c| \cdot \mathrm{SD}(X) \tag{107}$$

Relazioni di 2 variabili aleatorie:

$$E(X+Y) = E(X) + E(Y). \tag{108}$$

$$X, Y \text{ indip. } \Rightarrow \text{ incorrelate } \Rightarrow E(X \cdot Y) = E(X) \cdot E(Y) \tag{109}$$

$$Var(X+Y) = Var(X) + Var(Y) + Cov(X,Y) \tag{110}$$

$$\text{indip. } \Rightarrow Var(X + ... + X_n) = Var(X_1) + ... + Var(X_n) \tag{111}$$

BOZZA - DRAFT

Esercizi. Siano X e Y indipendenti di leggi esponenziali di parametri $\log 3$ $\log 4$ rispettivamente. Calcolare la loro covarianza, e la speranza matematica di X, $2Y$, $3+Y$ e $\pi X - Y$.

45.4 Funzione Gamma e Legge Gamma

La **funzione** $\Gamma(x)$, *funzione Gamma*, è una *funzione speciale* dell'Analisi Matematica, cioè, in pratica e semplificando, una funzione non elementare ma di notevole interesse, con una certa definizione che non daremo. I suoi valori approssimati vengono calcolati da vari software e molti si trovano elencati in tavole numeriche. Ma per i numeri $\frac{1}{2}, 1, 2, 3, 4...$ i suoi valori sono semplici:

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}, \qquad \Gamma(x) = (x-1)! \qquad \forall\, x \in \mathbb{N}^+$$

(si faccia un grafico) e quelli soli considereremo, per esempio

$$\Gamma(1) = 0! = 1, \quad \Gamma(2) = 1! = 1, \quad \Gamma(3) = 2! = 2, \quad \Gamma(4) = 3! = 6.$$

La **Legge** $\Gamma(\alpha, \lambda)$, Gamma di parametri $\alpha > 0$ e $\lambda > 0$, ha densità

$$f(t) := \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} \mathrm{e}^{-\lambda t} & se\ t > 0 \\ \\ 0 & se\ t \leq 0 \end{cases} . \tag{112}$$

Si scrivano esplicitamente e grafichino le densità per $\alpha := \frac{1}{2}, 1, 2, 3$ e $\lambda := 2$, e si calcolino e grafichino le funzioni di ripartizione.

Teorema.

Legge	$E(X)$ μ	$Var(X)$ σ^2
$X \sim \Gamma(\alpha, \lambda)$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$

BOZZA - DRAFT

45.5 Densità e leggi del chi quadrato

La legge $\chi^2(n)$ [del] chi quadrato (chi-quadrato, chi quadro, chi-quadro, inglese *chi-square* o *chi-squared*) ovvero χ^2 di parametro n o come si meglio dice *a n gradi di libertà* ha densità

$$f(x) = f(x; n) := \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} \mathrm{e}^{-\frac{x}{2}} & se\ x > 0 \\ 0 & se\ x \leq 0 \end{cases} . \tag{113}$$

Si scrivano esplicitamente e grafichino le densità per $n = 1, 2, 4, 6$, e si calcolino e grafichino le ultime 3 funzioni di ripartizione.

Di questa famiglia (insieme) di leggi, in Statistica sono molto importanti i quantili. I valori approssimati vengono calcolati da vari software e molti si trovano elencati in tavole numeriche.

Il valore di maggior interesse in Statistica Inferenziale è $q_{0.95}$, e in subordine con 0.9 e 0.99:

$$P(X < q_{0.95}) = 0.95 = 95\%$$

$$equivalentemente \quad P(X > q_{0.95}) = 0.05 = 5\%$$

BOZZA - DRAFT

e tale quantile non è un singolo numero perché viene ulteriormente precisato da n=1,2...

In particolare ci proponiamo di imparare a memoria il quantile $q_{0.95}$ relativo a $n = 11$ gradi di libertà, che però di spesso sulle tavole numeriche si trova riferito non a 0.95 bensì al valore 0.05:

$$\approx 19.68$$

di solito riportato con più decimali, per esempio 19.675 con 3 decimali.

Purtroppo c'è confusione nelle notazioni usate, Si veda il paragrafo (58.4).

Ecco alcuni valori:

su alcune tavole →	0.95	0.90	0.10	0.05	0.01
su altre tavole →	0.05	0.10	0.90	0.95	0.99
n				$\chi^2_{0.05}(1)$ *per altri* $\chi^2_{0.95}(1)$ ↓	
1	0.004	0.02	2.71	3.84	6.63
2	0.10	0.21	4.61	5.99	9.21
3	0.35	0.58	6.25	7.81	11.34
4	0.71	1.06	7.78	9.49	13.28
5	1.14	1.61	9.24	11.07	15.09
6	1.63	2.20	10.64	12.59	16.81
7	2.17	2.83	12.02	14.07	18.48
8	2.73	3.49	13.36	15.51	20.09
9	3.32?	4.17	14.68	16.92	21.67
10	3.94	4.87	15.99	18.31	23.21
11	4.57	5.58	17.28	**19.68** ↑ $\chi^2_{0.05}(11)$ *per altri* $\chi^2_{0.95}(11)$	24.73

BOZZA - DRAFT

Ripetiamo che per noi 19.68 è $\approx q_{0.95}$ con 11 gradi di libertà, ma spesso sulle tavole sta sulla colonna 0.05.

45.6 Densità e quantili di Student, e legge di Cauchy

Come le leggi del chi quadrato, la t **di Student** è una famiglia di leggi, con un parametro, di solito indicato con n o ν, ancora detto *gradi di libertà*: $n = 1, 2, 3....$ Scriviamo le prime 2:

$$f(t;1) = \frac{1}{\pi} \cdot \frac{1}{t^2+1} \qquad f(t;2) = \frac{1}{2\sqrt{2}} \cdot \frac{1}{\left(\frac{t^2}{2}+1\right)^{3/2}}$$

e nelle successive $\frac{t^2}{2}$ diventa $\frac{t^2}{3}$, $\frac{t^2}{4}$... e l'esponente $\frac{3}{2}$ diventa $\frac{4}{2}$, $\frac{5}{2}$... e cambia anche la costante moltiplicativa davanti, e ha una espressione che coinvolge la funzione Γ. I grafici un po' si assomigliano: simmetrici rispetto all'asse y, al crescere di n le campane diventano più alte e strette.

La $f(t;1)$ è la **densità di Cauchy**, che sorprendentemente non ha speranza matematica, ma tutte le altre sì, e allora ovviamente è 0 per la simmetria.

Di questa famiglia (insieme) di leggi, in Statistica sono molto importanti **i quantili**. I valori approssimati vengono calcolati da vari software e molti si trovano elencati in tavole numeriche, purtroppo di difficile lettura a causa di varie ambiguità, col serio rischio di confondere un parametro caratterizzante i quantili, α, con $1-\alpha$ oppure $2(1-\alpha)$ oppure $2\alpha-1$.

BOZZA - DRAFT

Nella riga di testa si trovano alcuni valori, talvolta con la specificazione *"one tail"*, e per associarli correttamente ad α, evitando di confonderlo con $1-\alpha$, o $2(1-\alpha)$, o $2\alpha-1$, si cerchi nella prima riga (cioè per $n=1$) il valore ≈ 6.31 corrispondente ad $\alpha = 0.95$:

$X \sim t$ di Student a 1 grado di libertà $\qquad P(X \le 6.31) \approx 0.95.$

Ecco alcuni valori. Qua si intende che α è grande, diciamo > 0.8.

	two tails	$2\alpha-1$	$\rightarrow$	0.9	0.95	0.98	0.99
	two tails	$2(1-\alpha)$	$\rightarrow$	0.1	0.05	0.02	0.01
	one tail	$1-\alpha$	$\rightarrow$	0.05	0.025	0.01	0.005
k	**one tail**	α	$\rightarrow$	0.95	0.975	0.99	0.995
1				6.3134	12.706	31.820	63.657
...				...	...	...	...
100				1.6602	1.984	2.364	2.625

ESEMPIO$_\mu$ Probabilità che una v.a. di legge di Student a 2 gradi

di libertà assuma un valore ≤ 2.52:

$$P(X \leq 2.52) = \int_{-\infty}^{2.52} f(t;2)\,dt =$$

$$= \int_{-\infty}^{2.52} \frac{1}{2\sqrt{2}} \cdot \frac{1}{\left(\frac{t^2}{2}+1\right)^{3/2}}\,dt$$

integrale non facile da calcolare al livello di questa trattazione elementare.
Con Wolframalpha `Integrate (1/(2Sqrt[2]))(1/(t^2/2+1)^(3/2)) from -Infinity to 2.52` dà

$$\approx 0.936 = 93.6\%.$$

BOZZA - DRAFT

Teoremi.

Legge	$E(X)$ μ	$Var(X)$ σ^2
$X \sim \chi^2(n)$ ovvero del χ^2 a n gradi di libertà	n	$2n$
$X \sim t$ di Student a n gradi di libertà	0 per $n > 1$	$\frac{n}{n-2}$ per $n > 2$

Osservazioni. Quelle che abbiamo visto, e le densità normale e log-normale che vedremo, sono fra le principali densità continue. Ma ne esistono infinite altre, alcune con un nome specifico, altre senza. Si considerino per esempio gli esercizi seguenti.

Esercizio 1$_\mu$ Calcolare la speranza matematica della densità

$$g(t) := \alpha\, \mathrm{e}^{-2|t|}$$

naturalmente dopo aver determinato la costante α. (La costante viene determinata dall'integrale unitario fra $-\infty$ e $+\infty$, che per la

parità della densità è 2 volte l'integrale fra 0 e $+\infty$).

Esercizio 2$_\mu$ Calcolare la speranza matematica della densità

$$f(z) := \begin{cases} 0 & se\ z < \frac{1}{2} \\ c & se\ \frac{1}{2} \le z < 1 \\ \frac{c}{z^3} & se\ z \ge 1 \end{cases} \qquad \textit{discontinua!}$$

naturalmente dopo aver determinato c. (Ovviamente bisognerà usare la Formula (14), Regola di Chasles, fatta valere su $]a, b[=]-\infty, +\infty[$, suddividendo 2 volte l'integrale, nei punti $\frac{1}{2}$ e 1).

Esercizio 3$_\mu$ Calcolare la speranza matematica della densità

$$\textit{discontinua!} \quad u(x) := a(2 + \operatorname{sgn}(x))\ per\ -1 \le x \le 2,\ e\ 0\ altrimenti.$$

BOZZA - DRAFT

45.7 Legge di $g(X)$

Se X è una variabile aleatoria discreta allora $g(X)$ è una variabile aleatoria qualunque sia la funzione g.
Per esempio il doppio del risultato di un lancio di dado, $2\,X$.

Per le variabili aleatorie continue invece, nella nostra trattazione elementare, richiediamo la continuità: se X è una variabile aleatoria continua e se g è una funzione sufficientemente regolare, e per noi in particolare continua, anche $g(X)$ è una variabile aleatoria continua.
Si pensi per esempio a $mX + q$, X^2, X^3 e $-X^3$.
Esempio concreto:

$$Y := 453{,}592\,37\,X$$

converte il peso X in libbre del primo bambino che nascerà vivo negli USA il prossimo anno, nel suo peso Y in grammi.

Se poi X e g sono molto regolari, X e $g(X)$ hanno anche speranza matematica.

Esempio 1 di calcolo della legge di $g(X)$ per v.a. continua: $mX+q$.
Troviamo funzione di ripartizione e poi densità di $mX+q$ con $m>0$:

$$F_{mX+q}(t) =$$

per definizione e poi per algebra

$$= P(mX+q \le t) = P(mX \le t-q) =$$

per algebra, essendo $m>0$

$$= P\left(X \le \frac{t-q}{m}\right) =$$

per definizione di funzione di ripartizione

$$= F_X\left(\frac{t-q}{m}\right)$$

e derivando rispetto a t la prima e l'ultima espressione con la regola di derivazione della funzione composta

$$f_{mX+q}(t) = \frac{1}{m} f_X\left(\frac{t-q}{m}\right) \qquad m>0$$

e un analogo calcolo con $m<0$ darà

$$f_{mX+q}(t) = -\frac{1}{m} f_X\left(\frac{t-q}{m}\right) \qquad m<0$$

e le 2 ultime formule si riassumono in questa formula della densità

$$f_{mX+q}(t) = \frac{1}{|m|} f_X\left(\frac{t-q}{m}\right) \qquad m\neq 0$$

Esempio 2 di calcolo della legge di $g(X)$ per v.a. continua: X^2.
Con un analogo tipo di calcoli si troverà

$$F_{X^2}(x) = F_X(\sqrt{x}) - F_X(-\sqrt{x}) \qquad \text{se } g(z) := z^2$$

e poi per tutte la densità si ottiene derivando, in particolare

$$f_{X^2}(x) = \frac{1}{2\sqrt{x}}(f_X(\sqrt{x}) + f_X(-\sqrt{x})) \qquad \textit{se } x>0$$

e ovviamente 0 se $x<0$.

BOZZA - DRAFT

45.8 Standardizzazione di una variabile aleatoria

Definizione. Per una variabile aleatoria X discreta o continua dotata di speranza matematica μ e varianza σ^2, la nuova variabile aleatoria

$$\boxed{Z := \frac{X-\mu}{\sigma} \text{ si dice } \textbf{standardizzazione } X \text{ o suo } \textbf{z score}} \quad (114)$$

(che è $mX+q$ con $m := \frac{1}{\sigma}$ e $q := -\frac{\mu}{\sigma}$).

Teorema. (Che si dimostra con un semplice calcolo).

$$\boxed{Z \text{ standardizzazione } \Rightarrow E(Z) = 0\,, \quad Var(Z) = 1.} \quad (115)$$

(Vale se Z è standardizzazione sia di una v.a. discreta che continua)
La standardizzazione presuppone la conoscenza di media e varianza *veri*, cosa possibile per esempio per un dado: $\frac{X-\frac{7}{2}}{\sqrt{\frac{35}{12}}}$.
La standardizzazione può essere applicata anche ai valori di un dataset:

$$z_i := \frac{x_i - \bar{x}}{\mathrm{SD}}$$

e per calcolare la deviazione standard SD si userà la formula con $\frac{1}{n}$ se il dataset rappresenta tutta la popolazione di interesse, o quella con $\frac{1}{n-1}$ se ne è un campione.

Si noti che lo z score è adimensionale, per com'è definito.

Lo z score tende ad assumere valori intorno a 0, e quelli più vicini a 0 provengono dai valori (originali) più normali e quelli più lontani dai valori più estremi.

Esempio. Per un dado $X \sim \mathbb{U}\{1,6\}$ è $\mu = \frac{7}{2}$ e $\sigma^2 = \frac{35}{12}$, e allora per la somma Z di 2 dadi $E(Z) = 7$ e per l'indipendenza $\sigma^2 = \frac{35}{6}$ ovvero $\sigma = \sqrt{\frac{35}{6}}$.
Prendiamo ora 100 lanci di coppia di dadi virtuali:

BOZZA - DRAFT

12 8 8 4 5 10 5 6 7 7 8 7 2 4 5 9 11 5 5 12 6 8 7 9 6 7 9 8 9 9 8 10 6 5 5 5 3 6 6 5 7 5 10 8 9 9 6 5 4 7 7 8 4 6 6 11 9 7 6 7 9 4 5 7 5 3 7 8 5 7 9 4 5 5 5 3 9 3 9 4 8 10 8 6 9 8 10 7 7 7 9 8 9 8 4 6 6 8 4 7

Possiamo standardizzare

2.42465 0.548712 0.548712 -1.32723 -0.858241 1.48668 -0.858241 -0.389257 0.0797273 0.0797273 eccetera

Si noti nuovamente che lo z score tende ad assumere valori intorno a 0, e quelli più vicini a 0 provengono dai valori (originali) più normali e quelli più lontani dai valori più estremi.
E fa questo in modo "standard": ogni statistico riconosce un valore normalissimo intorno a 0, un po' discosto dalla media intorno a 1 o -1, soprendente intorno a ± 2, un valore eccezionale intorno a ± 3, molto eccezionale intorno a ± 4, eccezionalissimo intorno a ± 5... Sia che la variabile aleatoria iniziale variasse intorno al 3000 com'è per il peso del neonato in grammi, o intorno al 3.5 com'è per il dado.

Esercizio. Si scrivano le standardizzazioni dei 2 dadi di (37.1), con quella stessa notazione. E di una moneta regolare.

Teorema. Come detto, data una variabile aleatoria continua X e una funzione sufficientemente regolare g, anche $g(X)$ è una variabile aleatoria continua. In condizioni di ulteriore regolarità X e $g(X)$ hanno speranza matematica.
Se la v.a. continua X ha densità f_X allora

$$E(g(X)) = \int_{-\infty}^{+\infty} g(t) f_X(t)\, dt.$$

in particolare (con $g(z) := z^n$)

$$E(X^n) = \int_{-\infty}^{+\infty} x^n f(x)\, dx \quad \text{si chiama } \textit{momento } n\text{-esimo.}$$

Il momento primo è proprio la speranza matematica di X.

Esempi. Vediamo 2 esempi tratti da un libro che si distingue nella didattica del Calcolo delle Probabilità e della Statistica: *Calcolo delle probabilità*, di Paolo Baldi, ed. McGraw-Hill. Alle pp. 125-126:

> Supponiamo che X sia uniforme su $[0,1]$. Quanto valgono $E[\sin(2\pi X)]$ e $E[\mathrm{e}^X]$?(...)
>
> $$E[\sin(2\pi X)] = \int_0^1 \sin(2\pi x)\, dx = 0$$
>
> $$E[\mathrm{e}^X] = \int_0^1 \mathrm{e}^x\, dx = \mathrm{e} - 1.$$

BOZZA - DRAFT

(L'uso delle soprastanti parentesi quadre non è conforme alle standard seguito in questa dispensa, ma si tenga presente che le notazioni fra i vari Autori differiscono alquanto).

Esercizi. Si consideri una v.a. X con $f_X(x) \coloneqq x$ fra 0 e $\sqrt{2}$ e 0 altrimenti. Trovare $E(\sin(2\pi X))$, $E(\mathrm{e}^X)$ ed $E(\ln X)$, e anche un'altra speranza matematica con una funzione a scelta.
Poi si cerchino le stesse 4 speranze matematiche con una nuova densità a scelta.

BOZZA - DRAFT

* Complementi *

45.9 Complementi – Disuguaglianza di Chebyshev

La Disuguaglianza di Chebyshev[157] che ora vedremo vale per una **variabile aleatoria discreta o continua qualunque** purché dotata di speranza matematica μ e varianza σ^2.

Riguarda gli *scarti dalla media* $|X - E(X)|$.

Precisamente, la disuguaglianza riguarda la probabilità che gli scarti dalla media siano "grandi" o per meglio dire superino un fissato numero c:

$$P(|X - E(X)| > c) \leq \frac{Var(X)}{c^2} \tag{116}$$

valida per ogni $c \neq 0$ reale; ma utile solo per $c > \sigma$.

Essa ci dice che minore è la variabilità della grandezza aleatoria X, minore è la probabilità che X assuma valori distanti dalla media. Equivalentemente con l'evento complementare

$$P(|X - E(X)| \leq c) \geq 1 - \frac{Var(X)}{c^2} \tag{117}$$

valida per ogni $c \neq 0$ reale; ma utile solo per $c > \sigma$.

BOZZA - DRAFT

Scrivendo da adesso μ per la speranza matematica e σ per lo scarto quadratico medio, coi valori di c multipli interi positivi $m\,\sigma$ di σ, $m = 1, 2, 3...$, osservato che $1 - \frac{\sigma^2}{(m\sigma)^2} = 1 - \frac{1}{m^2}$, la Disuguaglianza di Chebyshev è

$$P(|X - \mu| \leq m\,\sigma) \geq 1 - \frac{1}{m^2} \qquad m = 1, 2, 3...$$

e con $m := 2, 3, 4, 5$ (il caso $m := 1$ non è significativo) risulta

$$P(|X - \mu| \leq 2\sigma) \geq \frac{3}{4} = 0.75 = 75\%$$

$$P(|X - \mu| \leq 3\sigma) \geq \frac{8}{9} = 0.8888... \approx 88.9\%$$

$$P(|X - \mu| \leq 4\sigma) \geq \frac{15}{16} = 0.9375 = 93.75\%$$

$$P(|X - \mu| \leq 5\sigma) \geq \frac{24}{25} = 0.96 = 96\%.$$

[157] Ovvero di Bienaymè-Chebyshev. Miglior traslitterazione dal russo: Čebyšëv.

Queste maggiorazioni valgono – come detto – per distribuzioni qualunque. Molti che sanno qualcosa di Statistica hanno in mente la notissima formula $P(|X-\mu| \le 2\sigma) \approx 95\%$ ma una tale affermazione stringente può essere fatta se si sa che X ha densità normale.

45.10 Complementi – Esercizi sulla legge di $g(X)$

Esercizio. Si trovi densità e f.r. di X^3.

$$P(X^3 \le x) = P(X \le \sqrt[3]{x}) \qquad \text{cioè}$$

$$\boxed{F_{X^3}(x) = F_X(\sqrt[3]{x})}$$

(che si poteva ottenere direttamente dopo aver riconosciuto che $g(x) := x^3$ è crescente suriettiva con inversa $\sqrt[3]{x}$) e derivando, essendo $D\sqrt[3]{x} = D\,x^{\frac{1}{3}} = \frac{1}{3}x^{-\frac{2}{3}} = \frac{1}{3\sqrt[3]{x^2}}$ per $x \neq 0$,

$$\boxed{f_{X^3}(x) = \frac{1}{3\sqrt[3]{x^2}}\, f_X(\sqrt[3]{x}) \quad x \neq 0}$$

e in 0 la densità può essere posta 0 o qualunque altro valore ≥ 0.

BOZZA - DRAFT

Esercizio. Si trovi densità e f.r. di $-X^3$.

$$P(-X^3 \le x) = P(X^3 \ge -x) = 1 - P(X^3 \le -x) = 1 - P(X \le \sqrt[3]{-x}) = 1 - P(X \le -\sqrt[3]{x})$$

cioè

$$\boxed{F_{-X^3}(x) = 1 - F_X(-\sqrt[3]{x})}$$

(che si poteva ottenere direttamente dopo aver riconosciuto che $g(x) := -x^3$ è decrescente suriettiva con inversa $-\sqrt[3]{x}$) e derivando, essendo $D\sqrt[3]{x} = D\,x^{\frac{1}{3}} = \frac{1}{3}x^{-\frac{2}{3}} = \frac{1}{3\sqrt[3]{x^2}}$,

$$\boxed{f_{-X^3}(x) = \frac{1}{3\sqrt[3]{x^2}}\, f_X(-\sqrt[3]{x}) \quad x \neq 0}$$

e in 0 la densità può essere posta 0 o qualunque altro valore ≥ 0.

46 Densità e variabile aleatoria normale

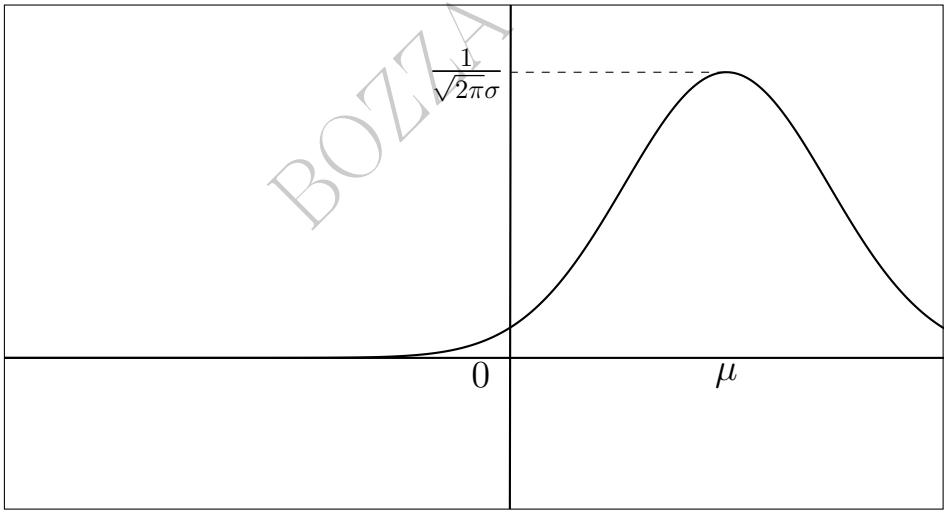
46.1 Introduzione alla densità e v.a. normale

La densità **normale** ovvero **gaussiana** ha un grafico detto "a campana", con limiti 0 a $+\infty$ e $-\infty$, prima crescente e poi decrescente, prima con la concavità verso l'alto, poi verso il basso e infine verso l'alto.

Si tratta in qualche modo della più "pura" e "perfetta" delle densità a campana.

BOZZA - DRAFT

La *moda* (cioè l'eventuale unico punto di massimo di una densità di v.a. continua) esiste e coincide con la media μ, e la densità è simmetrica rispetto alla retta $x = \mu$, e a



causa di questa simmetria anche la mediana è μ: la probabilità di un valore prima di μ è uguale alla probabilità di un valore dopo μ. Per la simmetria la skewness è nulla.

Il massimo assoluto (ovviamente) vale $\frac{1}{\sqrt{2\pi}\sigma}$ e in $\mu \pm \sigma$ ci sono i 2 flessi (che si trovano facilmente con la derivata seconda), e allora

a grande varianza corrisponde campana bassa e larga

a piccola varianza corrisponde campana alta e stretta.

Questa legge è denotata con $N(\mu, \sigma^2)$, ha 2 parametri (come la legge Gamma) ed essi sono proprio la media e la varianza:

densità normale $N(\mu, \sigma^2)$

$$f(t) := \frac{1}{\sqrt{2\pi}\,\sigma} \mathrm{e}^{-\frac{(t-\mu)^2}{2\,\sigma^2}}$$

I parametri sono normalmente considerati μ e σ^2 (non μ e σ ma si faccia attenzione che qualche software invece fa proprio così).

Si noti che è strettamente positiva su tutto $\mathbb{R}$: sono possibili valori grandissimamente positivi o negativi, ma sono pochissimo probabili (globalmente, ovvio: i singoli valori hanno tutti probabilità 0).

Teorema. La somma di 2 (e poi più) variabili aleatorie normali indipendenti è normale con media la somma delle medie e varianza la somma delle varianze.
Cioè, se $X \sim N(\mu_1, \sigma_1^2)$ e $Y \sim N(\mu_2, \sigma_2^2)$ sono indipendenti

$$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

Inoltre per le 2 variabili aleatorie indipendenti vale

$$c + X \sim N(c + \mu_1, \sigma_1^2) \qquad c\,X \sim N(c\mu_1, c^2\sigma_1^2). \tag{118}$$

Per fissare le idee: una $N(200, 2)$ e una $N(300, 3)$ hanno densità 2 campane gaussiane, e la $X+Y$ ha a sua volta una densità a campana gaussiana, ma centrata in 500. (Non ha affatto una densità somma delle 2 campane, con forma "a cammello", come si potrebbe ingenuamente ipotizzare: la densità della somma di 2 variabili aleatorie non è la somma delle densità). E con varianza 5. Si sommano le medie, e si sommano le varianze, e si conserva la normalità, ovvero gaussianità.

BOZZA DRAFT

Nota 1. Il fatto che la somma di 2 variabili aleatorie normali sia una variabile aleatoria normale, è notevole. La somma di 2 variabili aleatorie uniformi continue, per esempio, in generale non è affatto una variabile aleatoria uniforme continua.

Nota 2. Rimarchiamo di nuovo che la densità della somma di 2 variabili aleatorie non è la somma delle densità delle 2 variabili aleatorie.

Nota 3. Anche la f.r. della somma di 2 variabili aleatorie non è la somma delle funzioni di ripartizione delle 2 variabili aleatorie.

BOZZA - DRAFT

46.2 Variabile aleatoria normale standard

(A causa delle (118)) la standardizzazione di una qualunque variabile aleatoria normale è una variabile aleatoria normale $N(0,1)$. Fra una v.a. normale Y e la sua standardizzazione X valgono le relazioni

$$\boxed{\begin{array}{c} X \sim N(0,1) \; standardizzazione \; di \; Y \sim N(\mu, \sigma^2) \\ \\ X = \frac{Y-\mu}{\sigma} \qquad Y = \sigma X + \mu. \end{array}}$$

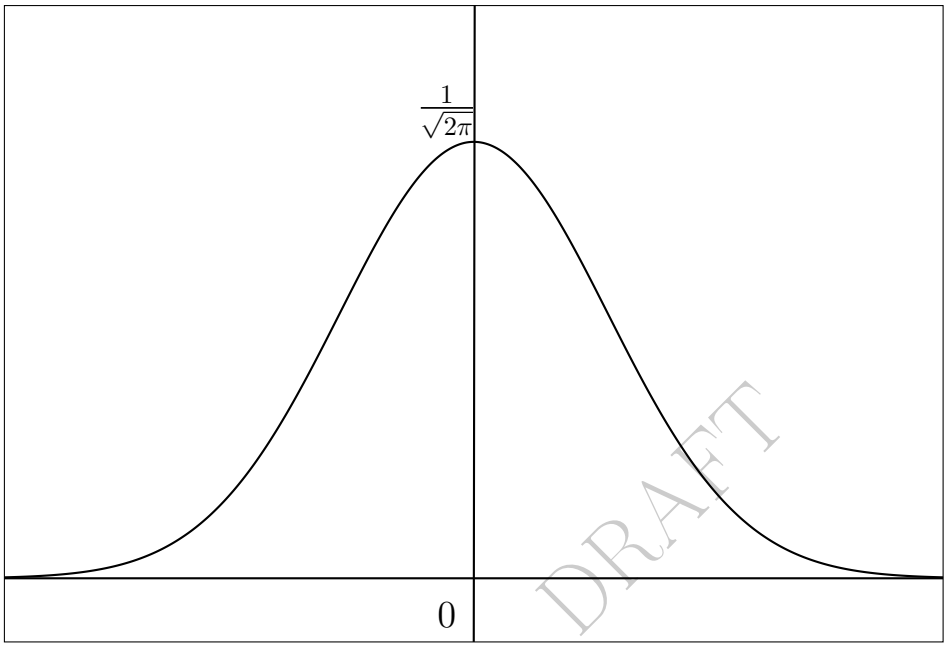
Avendo media 0 e varianza 1, la variabile aleatoria normale standard ha

$$\boxed{\begin{array}{c} \text{densità normale standard } N(0,1) \\ \\ \phi(t) := \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{t^2}{2}} \end{array}} \tag{119}$$

Si noti che la *densità normale standard* sopra scritta viene denotata con $\phi(\cdot)$ piuttosto che con qualche nome generico come $f(\cdot)$ o $g(\cdot)$.

Ha anche moda 0 e mediana 0 e skewness 0.

Il massimo è in 0 e vale $\frac{1}{\sqrt{2\pi}} \approx 0.4$. I punti di flesso sono in ± 1.



BOZZA - DRAFT

Nelle Scienze Applicate tradizionalmente si considera che valga "quasi 0" dopo 3 e prima di –3. Naturalmente non si azzera mai, ma la decrescenza a 0 è rapidissima: ha le *code leggere*.

La sua funzione di ripartizione si indica con $\Phi(x)$

$$\boxed{\Phi(x) \text{ f.r. normale standard}} \tag{120}$$

e si chiama *funzione di ripartizione normale standard*, in Inglese *(standard) normal cumulative distribution function*, e (per la (90) è

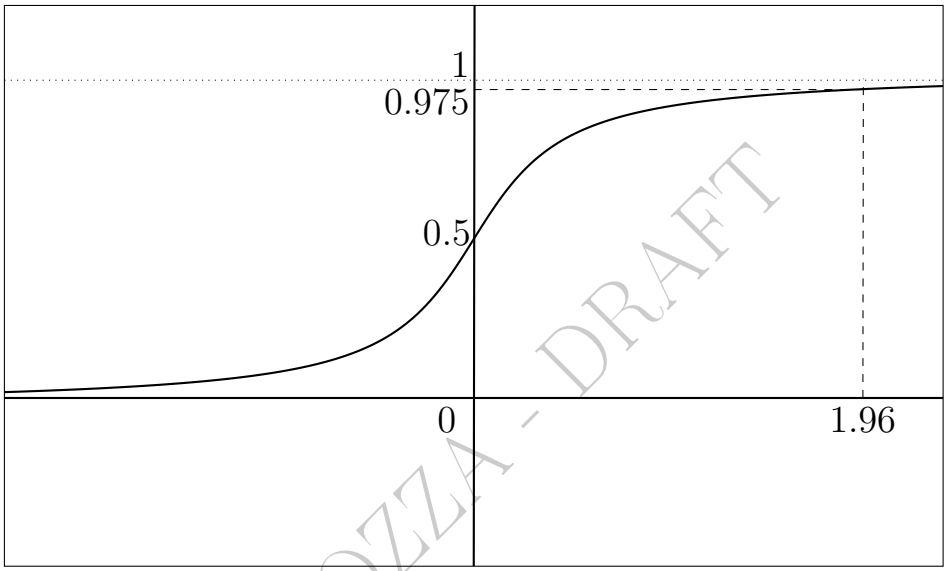
$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt$$

e derivando, col Teorema Fondamentale del Calcolo Integrale,

$$\boxed{\Phi'(x) = \phi(x)} \tag{121}$$

(corrispondentemente a (91)).

L'integrale che definisce questa *funzione speciale (dell'Analisi Matematica)* non può essere risolto in termini di funzioni elementari. Valori numerici (approssimati) di $\Phi(x)$ si ottengono in vari modi che vedremo nella Lezione successiva.



BOZZA - DRAFT

46.3 Quantili normali, e valore 1.96

Il quantile normale ordine α si indica con ϕ_α ed è

$$\phi_\alpha := \Phi^{-1}(\alpha) \tag{122}$$

La costante fondamentale 1.96 della Statistica

$$\boxed{\phi_{0.975} \approx 1.96} \tag{123}$$

Nota. Si veda il (disegno parziale del) grafico di ϕ_α nella prossima Lezione. I quantili normali sono fondamentali in Statistica Inferenziale.

Il grafico della funzione ϕ_α ha dominio $]0,1[$, in 0.5 vale (ovviamente) 0, tende a $-\infty$ in 0 e a $+\infty$ in 1.
Ovviamente quel grafico è il simmetrico di quello visto di $\Phi(x)$ rispetto alla bisettrice del I e II quadrante.
Si disegni quel grafico e su esso si trovi il punto $(0.975, \approx 1.96)$.

46.4 Scarti dalla media per v.a. normale e qualunque

BOZZA - DRAFT

Vediamo qual è la probabilità che una variabile aleatoria disti meno di 1 σ o di 2 σ o di 3 σ dalla sua media.
Se si sa che la variabile aleatoria X è normale si ottengono disuguaglianze, che vedete in tabella, molto più stringenti – e quindi da considerare in generale migliori e più utili – che nel caso generale.

La prima colonna si potrebbe dimostrare con la Disuguaglianza di Chebyshev (paragrafo 45.9).
La seconda colonna si potrebbe dimostrare[158] con calcoli integrali.

[158] Si ha, per ogni $\delta > 0$,

$$P(|X-\mu| \le \delta) = P(|\sigma Y| \le \delta) = P(\sigma|Y| \le \delta) = P\Big(|Y| \le \frac{\delta}{\sigma}\Big)$$

e con facili calcoli si conclude

$$P(|X-\mu| \le c\,\sigma) = 2\,\Phi(c) - 1$$

e coi classici valori (che si trovano per esempio sulle tavole, e vengono da calcoli integrali) si ottengono le approssimazioni.

X v.a. qualunque discreta o continua con media μ e varianza σ^2	X v.a. normale (*) con media μ e varianza σ^2 cioè $X \sim N(\mu, \sigma^2)$
$P(\lvert X-\mu\rvert \le \sigma) \ge 0\%$, inutile	$P(\lvert X-\mu\rvert \le \sigma) \approx 68.3\%$
$P(\lvert X-\mu\rvert \le 2\,\sigma) \ge 75\%$	$P(\lvert X-\mu\rvert \le 2\,\sigma) \approx 95.4\%$
$P(\lvert X-\mu\rvert \le 3\,\sigma) \ge 88.8\%$	$P(\lvert X-\mu\rvert \le 3\,\sigma) \approx 99.7\%$
	$P(\lvert X-\mu\rvert \le 1.96\,\sigma) \approx 0.95 = 95\%$
$P(\lvert X-\mu\rvert \le c\,\sigma) \ge 1-\frac{1}{c^2}$	$P(\lvert X-\mu\rvert \le c\,\sigma) = 2\,\Phi(c)-1$

BOZZA - DRAFT

Nella seconda colonna la quarta è una lieve modificazione della seconda per avere con più precisione 95%.

Migliore approssimazione di 68.3 è 68.26.

Migliore approssimazione di 95.4 è 95.46.

Migliore approssimazione di 99.7 è 99.74.

La possiamo vedere in questa forma molto semplificata:

X v.a. normale con media μ e varianza σ^2 cioè $X \sim N(\mu, \sigma^2)$ (SD = *standard deviation*= σ)	(**)
$P(media - \sigma \le X \le media + \sigma) \approx 68\%$	(sta entro 1 SD)
$P(media - 1.96\,\sigma \le X \le media + 1.96\,\sigma) \approx 95\%$ $P(media - 2\,\sigma \le X \le media + 2\,\sigma) \approx 95\%$	(sta entro 2 SD)
$P(media - 3\,\sigma \le X \le media + 3\,\sigma) \approx 99.7\%$	(sta entro 3 SD)
Per una v.a. qualunque valgono disuguaglianze molto meno stringenti	

BOZZA - DRAFT

Per una normale standard diventano:

Normale standard	$Z \sim N(0,1)$	(***)
$P(-1 \le Z \le 1)$	$\approx 68\%$	
$P(-1.96 \le Z \le 1.96)$	$\approx 95\%$	(o spesso $-2 \le Z \le 2$)
$P(-3 \le Z \le 3)$	$\approx 99.7\%$	

Esempio. Sul portale web di Our World in Data, generalmente affidabile (in particolare per aggiornamenti statistici sulla pandemia del covid) e consigliato al lettore, leggiamo[159]

> Adult heights within a population are approximately normally distributed due to genetic and environmental variance.
>
> (...) The normal distribution of heights allows us to make inferences about the range. Around 68% of heights will fall within one

[159] https://ourworldindata.org/human-height. Letto il 15 dicembre 2022.

> standard deviation of the mean height; 95% within two standard deviations; and 99.7% within three. If we know the mean and standard deviation of heights, we have a good understanding of how heights vary across a population.
>
> (...) As an aggregate of the regions with available data – Europe, North America, Australia, and East Asia – they found the mean male height to be 178.4 centimeters (cm) in the most recent cohort (born between 1980 and 1994).42 The standard deviation was 7.59 cm. This means 68% of men were between 170.8 and 186 cm tall; 95% were between 163.2 and 193.6 cm.

Similmente per le femmine: potrà fare i calcoli il lettore coi valori che in quello stesso testo si trovano, 164.7 e 7.07.

Nota. Di questa Lezione ci si dovrebbe aspettare che lo studente conosca a memoria le formule *contemporaneamente* numerate *et* riquadrate, oltre a tutto l'inquadramento teorico della questione, e ovviamente compresa la (***) (e non necessariamente (**) e (*) che sono decisamente più dettagliate).

BOZZA - DRAFT

Fermiamoci un momento!

Ha avuto ampio risalto sui media un articolo scientifico sull'invenzione di un nuovo tipo di test diagnostico del cancro, che si fa in meno di 10 minuti e non richiede un laboratorio – e allora potenzialmente potrà interessare le farmacie.
È stato pubblicato su una rivista scientifica di altissimo livello, Nature Communications.
Apprezziamo quante cose riusciamo a capire, con lo studio fatto finora, in questa figura dell'articolo: Link->

- le figure a campana
- le curve ROC
- l'AUC, area under the curve, usata per valutare la curva ROC ovvero la bontà del test diagnostico
- specificità, sensitività; e PPV (Positive Predictive Value) è quello che in questa trattazione è stato indicato VPP (Valore Predittivo Positivo, essenzialmente la predittività.
- i box [and wisker] plot – fatti nel modo semplice di questa trattazione:

> "In the box and whisker plots, the middle lines of the boxes represent the median (50th percentile) and the terminal line of the boxes represents the 25th to 75th percentile. The whiskers represent the lowest and the highest value"

- i bar chart

(I "peluzzi" sopra le colonne del bar chart, questione che non abbiamo trattato, si riferiscono alle deviazioni standard dei dataset di misurazioni).

BOZZA - DRAFT

47 Approssimazione di $\Phi(x)$ e ϕ_α

47.1 Grafici

ZA - DRAFT

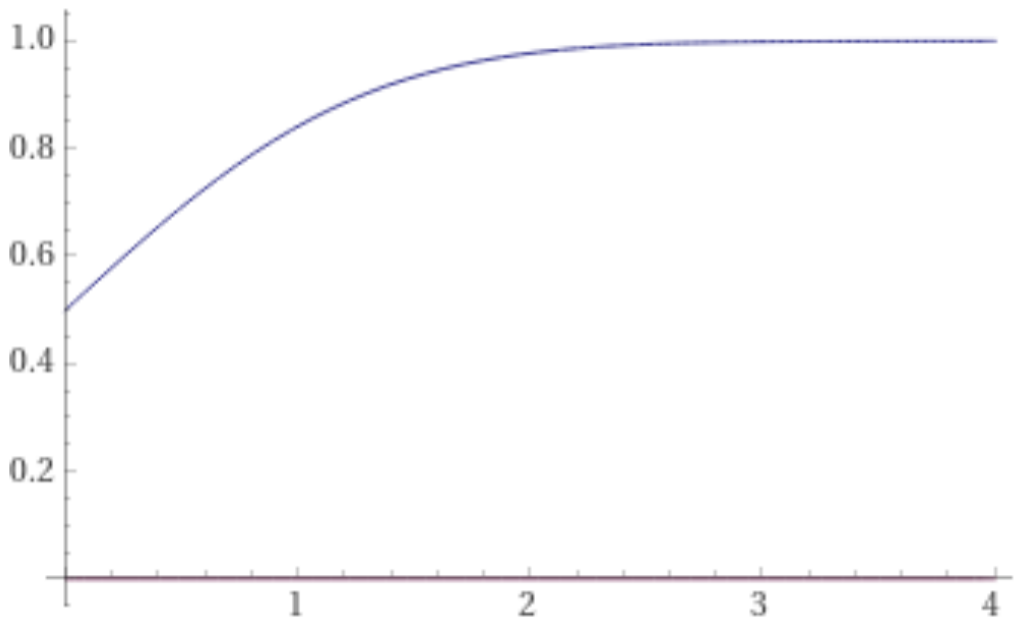


Figure 53: Funzione di ripartizione normale $\Phi(x)$. Naturalmente la funzione ha limite 0 in $-\infty$, ma il grafico della funzione è rappresentato solo per $x \geq 0$. (Screenshot da WolframAlpha).

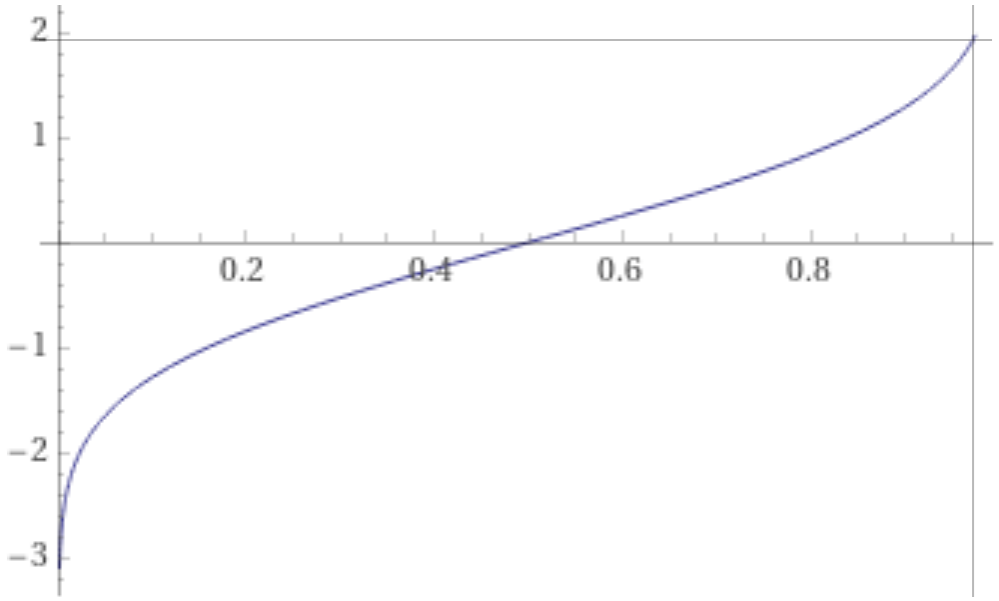


Figure 54: Funzione dei quantili normali ϕ_α. Naturalmente la funzione ha limite $-\infty$ in 0, e $+\infty$ in 1, ma il grafico della funzione è rappresentato solo fino a 0.975 ove essa vale ≈ 1.96. (Screenshot da WolframAlpha).

BOZZA - DI

47.2 Approssimazioni

Valori numerici (approssimati) della funzione di ripartizione normale standard $\Phi(x)$ e dei quantili normali ϕ_α si ottengono:

(1) con quelle (rare) calcolatrici scientifiche che la implementano;

(2) online in www.wolframAlpha.com digitando

`CDF[NormalDistribution[0,1],`*valore di x*`]` per avere $\Phi(x)$,

`InverseCDF[NormalDistribution[0,1],`*valore di α*`]` per avere ϕ_α;

(3) con molti software di manipolazione matematica, fra cui R e Maxima (gratuiti), e Mathematica$^{(R)}$;

(4) magari a memoria per alcuni pochi valori speciali, in particolare

senz'altro $\phi_{0.975} \approx 1.96$ e l'ovvio $\phi_{0.5} = 0$ e magari tutti questi:

x	$\Phi(x) = P(X \le x)$	
0	0.5	il primo è ovvio per simmeteria
≈ 1.64	≈ 0.95	e il terzo vogliamo ricordarlo:
≈ 1.96	≈ 0.975	$\phi_{\mathbf{0.975}} \approx \mathbf{1.96}$ ovvero $\Phi(1.96) \approx 0.975$
≈ 2.58	≈ 0.995	e come sopra scriveremo gli altri.
$(+\infty)$	(1)	Quest'ultimo vale solo come limite.
ϕ_α	α	

(e si faccia attenzione che questi non sono i valori di $P(|X| < x)$);

(5) con le apposite tavole numeriche, che si trovano su internet cercando *normal table* o (e sono essenzialmente le stesse) *normal quantile table*.

(6) con apposite formule di approssimazione. Ne esistono molte decine, più o meno precise e più o meno semplici. Ne considereremo 3:
– questa semplicissima ma poco precisa, err. ass. < 0.04

$$\forall x \ge 0 \qquad \Phi(x) \approx \begin{cases} \sqrt{1 - \frac{(3-x)^2}{12}} & x \le 3 \\ 1 & x > 3 \end{cases}$$

– quella di Shah (1985) di un esercizio seguente, err. ass < 0.006:

$$\Phi(x) \approx \begin{cases} \frac{x\,(4.4-x)}{10} + 0.5 & se\ 0 \le x \le 2.2 \\ 0.99 & se\ 2.2 < x < 2.6 \\ 1 & se\ x \ge 2.6 \end{cases}$$

– questa più precisa[160] con err. ass. < 0.0002

$$\forall x \ge 0 \qquad \Phi(x) \approx 2\text{^}(-22\text{^}(1 - 41\text{^}(x/10)))$$

[160] http://m-hikari.com/ams/ams-2014/ams-85-88-2014/epureAMS85-88-2014.pdf (A. Soranzo, E. Epure – 2014) Ecco la sua inversa (che si ottiene subito ricavando x, che è

BOZZA - DRAFT

che ha un'inversa esprimibile con funzioni elementari che dà ovviamente i quantili normali.

Nota 1. Le prime 2 formule di approssimazione sono definite a tratti, sono poco precise ma hanno il vantaggio che si possono calcolare con le 4 operazioni e la radice quadrata. In particolare la prima, per un uso agevole della calcolatrice la si esprima così, in $[0, 3]$:

$$\sqrt{\texttt{((3-x)\^2)·(-1)/12+1}} \qquad \textit{tutto sotto radice}$$

(Si calcola 3-x, si eleva al quadrato ovvero si moltiplica per se stesso, si moltiplica per -1, si divide per 12, si somma 1, si estrae la radice quadrata).

Nota 2. Sia le tavole numeriche che, di solito, le formule di approssimazione danno (approssimano)

$\Phi(x)$ solo per $x \geq 0$

ϕ_α solo per $\alpha \geq 0.5$

e per i valori di $x < 0$ e $\alpha < 0.5$ si usano le formula di simmetria

$$\boxed{\Phi(-x) = 1 - \Phi(x)} \tag{124}$$

$$\boxed{\phi_{1-\alpha} = -\phi_\alpha} \tag{125}$$

(che seguono dalla parità della densità normale standard).

ESERCIZIO$_{\mu 2018}$

$\approx$ % Per una variabile aleatoria normale standard X calcolare

$$P(-1.2 \leq X \leq 0.8)$$

ϕ_α)

$$\forall \alpha \in [0, 0.5[\qquad \phi_\alpha \approx \frac{10}{\log 41} \log\left(1 - \frac{\log((-\log \alpha)/\log 2)}{\log 22}\right)$$

(che ha errori assoluto e relativo rispettivamente $|\varepsilon(\alpha)| < 5 \cdot 10^{-3}$ $\forall \alpha \in [0.5, 9925]$, $|\varepsilon_r(\alpha)| < 1\%$ $\forall \alpha \in [0.5, 0.99908]$);

BOZZA - DRAFT

usando questa classica approssimazione (Shah 1985)

$$\Phi(x) \approx \begin{cases} \frac{x\,(4.4-x)}{10} + 0.5 & se\ 0 \le x \le 2.2 \\ 0.99 & se\ 2.2 < x < 2.6 \\ 1 & se\ x \ge 2.6 \end{cases}$$

SVOLGIMENTO è

$$P(-1.2 \le X \le 0.8) = P(X \le 0.8) - P(X < -1.2) =$$

trattandosi di densità continua le probabilità con < e ≤ sono uguali

$$= P(X \le 0.8) - P(X \le -1.2) =$$

per definizione di $\Phi(x)$

$$= \Phi(0.8) - \Phi(-1.2) =$$

e con la forumla di simmetria $\Phi(-x) = 1 - \Phi(x)$

$$= \Phi(0.8) - \big(1 - \Phi(1.2)\big) =$$

$$= \Phi(0.8) - 1 + \Phi(1.2) =$$

e con l'approssimazione data

$$\approx \frac{0.8\,(4.4-0.8)}{10} + 0.5 - 1 + \Big(\frac{1.2\,(4.4-1.2)}{10} + 0.5\Big) =$$

$$= 0.788 - 1 + 0.884$$

e in definitiva (recuperando il simbolo ≈ da più sopra)

$$\boxed{\approx 0.672 = 67.2\%}$$

BOZZA - DRAFT

47.3 Variabile aleatoria log-normale

Se X è una variabile aleatoria $N(\mu, \sigma^2)$ allora $Y := e^X$ si dice *log-normale* di parametri μ e σ^2, che però non sono rispettivamente media e varianza della nuova variabile aleatoria.
Inversamente, se Y è log-normale di parametri di parametri μ e σ^2, allora $X := \ln Y$ è $N(\mu, \sigma^2)$.

È evidente che mentre una v.a. normale può assumere qualunque valore reale, una v.a. log-normale può assumere solo valori positivi, data la corrispondenza $Y := \mathrm{e}^X$.

(E quelli proprio piccolissimi sono globalmente improbabili, proprio come per la v.a. normale sono improbabilissimi i valori grandissimamente negativi).

Consideriamo ora solo il caso $\mu = 0$ e $\sigma^2 = 1$:

$$\forall x > 0 \qquad F_{\mathrm{e}^X}(x) = P(\mathrm{e}^X \le x) = P(X \le \ln x) = F_X(\ln x)$$

e per i non positivi, considerando i 2 casi disgiunti $x = 0$ e $x < 0$

$$\forall x \le 0 \qquad F_{\mathrm{e}^X}(x) = P(\mathrm{e}^X \le x) = P(\mathrm{e}^X = x) + P(\mathrm{e}^X < x) = 0 + 0$$

e in definitiva

$$F_Y(x) = F_{\mathrm{e}^X}(x) = \begin{cases} \Phi(\ln x) & se \;\; x > 0 \\ 0 & se \;\; x \le 0 \end{cases} .$$

Derivando troviamo la densità log-normale di parametri $\mu = 0$ e $\sigma^2 = 1$, ricordando che $\Phi'(x) = \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, e naturalmente (derivata della funzione composta) deriviamo anche ln:

$$\forall x > 0 \qquad f_Y(x) = f_{\mathrm{e}^X}(x) = \phi(\ln x) \cdot \frac{1}{x}$$

trovandosi in definitiva la densità log-normale standard

$$\boxed{\text{densità log-norm. standard } f_Y(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\, x} e^{-\frac{\ln^2 x}{2}} & se \;\; x > 0 \\ 0 & se \;\; x \le 0 \end{cases}} \tag{126}$$

e con μ e σ^2 generici la

$$\begin{gathered} \text{densità } Lognormal(\mu, \sigma^2) \\ \text{log-normale di parametri } \mu \text{ e } \sigma^2 \\ f_Y(x) = f_{\mathrm{e}^X}(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\, \sigma\, x} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} & se \;\; x > 0 \\ 0 & se \;\; x \le 0 \end{cases} \end{gathered} \tag{127}$$

???? FIGURA VEDI TEX

BOZZA - DRAFT

(Secondo alcuni Autori i parametri della $Lognormal(\mu, \sigma^2)$ sono μ e σ, secondo altri[161] sono μ e σ^2, e in questa trattazione seguiamo questo secondo standard, ma si faccia attenzione in particolare usando i vari software).

I 3 valori coincidenti per la normale, media moda e mediana, $E(X) = Mod(X)$ =mediana, tutti μ, hanno 3 destini diversi: per la log-normale $Y := e^X$ la mediana e^μ, con $\mu = E(X)$:

$$P(Y \leq e^\mu) = P(e^X \leq e^\mu) = P(X \leq \mu) = \frac{1}{2} \text{ per simmetria}$$

ma media e moda hanno espressioni diverse da e^μ.

Esercizio. Si faccia lo studio di funzione di (127). Quanto vale la funzione di ripartizione in -1, 0, $e^{1.64}$, $e^{2.58}$, 100?

BOZZA - DRAFT

47.4 Confronto fra normale e log-normale, e cigni neri

Si veda nella figura quanto possano assomigliarsi normale e log-normale nella regione intorno alla moda. Sono rappresentate $N(0.95, 0.24)$ e Lognormal$\left(\mu = 0, \sigma^2 = \frac{1}{16}\right)$.

Dati empirici, sperimentali, di una variabile aleatoria per sua natura X log-normale, potrebbe essere facile erroneamente ritenerli provenienti da una variabile aleatoria Y normale. Non cambierà molto nella regione intorno alla media, ma lontano da essa le *code destre* hanno comportamento diversissimo: le code log-normali sono molto più *pesanti*. Tantoché

$$P(X \geq 2) \approx 0.000\,607\,\%$$

$$P(Y \geq 2) \approx 0.28\,\%$$

circa 458 volte più probabile.

[161] Si confrontino per esempio le Wikipedie italiana e in inglese.

Figure 55: Una densità normale e una densità a log-normale

È – solo in parte – la questione dei *cigni neri*, eventi importanti erroneamente ritenuti quasi impossibili, e invece poi si verificano. Possono verificarsi per aver identificato come normale una distribuzione molto simile, non necessariamente log-normale, ma con almeno una coda molto più pesante. O facendo consimili errori. (Il concetto di cigno nero comunque è più ramificato).

Bisogna essere molto cauti nel ritenere normale il modello sottostante i dati empirici, che essendo limitati in numero non evidenziano i casi rarissimi – cosa preoccupante se si tratta di eventi avversi gravi a un medicinale. Se semplicemente lo riteniamo normale, magari confortati da qualche test statistico, e invece è solo quasi nomale, ma ha una coda (tipicamente destra ma può essere sinistra, o entrambe) molto più pesante della normale, finisce che eventi ritenuti tanto rari da non doversene preoccupare, invece poi tanto impossibili non sono.

BOZZA - DRAFT

L'assunzione che i dati siano normali è comunissima nello studio dei farmaci e degli intossicanti, e può essere più o meno perfettamente aderente alla realtà sottostante ai fenomeni, che in generale comunque ci sfuggirà sempre.
Vediamo per esempio come si presenta in un articolo[162] scientifico (2022) l'illustrazione delle analisi statistiche fatte, in cui a questo punto lo studente riconoscerà varie cose viste, e altre ancora le vedremo nelle Lezioni successive:

> Statistical analysis was performed using GraphPad Prism versions 8 and 9 (GraphPad Software, La Jolla, CA, USA). Sample sizes were mainly determined based on our previous experiments and previous lab publications, although the sample size for drug rescue experiments was limited by time availability; no formal randomization was used

[162] Adams, J.W., Negraes, P.D., Truong, J. et al. Impact of alcohol exposure on neural development and network formation in human cortical organoids. Mol Psychiatry (2022). https://doi.org/10.1038/s41380-022-01862-7

> to allocate samples to experimental condition. Results from continuous variables are presented as mean ± standard error of the mean (s.e.m.), and **95% confidence intervals were normal-based**. Means were compared between groups using, where appropriate, unpaired Student's t-test, one-way, or two-way analyses of variance (ANOVA). Outliers were determined using GraphPad criteria. Whenever possible, the investigator was blind to the sample conditions. Tests were performed two-sided with α throughout set as 0.05.

(Enfasi aggiunta).

Nota. Di questa Lezione ci si dovrebbe aspettare che lo studente conosca a memoria le formule *contemporaneamente* numerate *et* riquadrate, oltre a tutto l'inquadramento teorico della questione.

BOZZA – DRAFT

48 Legge dei Grandi Numeri

48.1 Inquadramento euristico della situazione

Supponiamo di lanciare una moneta equilibrata un numero grandissimo di volte, e continuiamo a farlo, conteggiando il numero di teste e il numero di croci. Alcuni ingenui credono che i 2 numeri tendano a diventare sempre più simili, ma questo è falso: è impensabile che lanciando un milione di volte la moneta siano venute esattamente 500mila teste e 500mila croci, o 500 001 o anche 500 002 o simili. Anzi si potrebbe dimostrare che la differenza fra teste e croci tenderà in generale ad essere sempre più grande, non più piccola! Invece quello che tende a succedere è che le proporzioni di teste e di croci tenderanno ad uguagliarsi, tendendo entrambe ad $\frac{1}{2}$. Quello che possiamo effettivamente aspettarci dopo un milione di lanci è una situazione di questo tipo:

BOZZA - DRAFT

teste: $500\,000 \pm$ qualche centinaio: $\#teste = 500\,000 + r := n_0$
croci: $500\,000 \mp$ qualche centinaio: $\#croci = 500\,000 - r := n_1$
r: qualche centinaio in positivo o in negativo, p.es. 424 o -723
frazione di teste: $\frac{500\,000+r}{1\,000\,000} = \frac{1}{2} + \frac{r}{1\,000\,000} \approx 0.5$
frazione di croci: $\frac{500\,000-r}{1\,000\,000} = \frac{1}{2} - \frac{r}{1\,000\,000} \approx 0.5$.

Le *proporzioni empiriche* tendono ad uguagliarsi, non le quantità!

Questo diventerà ancora più evidente al crescere del numero di lanci, cioè l'approssimazione a 0.5 varrà con sempre più decimali, salvo casi sfortunatissimi, comunque possibili.

Questo tendere della proporzione al 50% avviene nonostante che la differenza fra teste e croci tenderà in generale ad essere sempre più grande, non più piccola!

Analogamente avviene non solo con $p = 0.5$ ma ogni p.
Tutto ciò viene generalizzato nel paragrafo seguente.

48.2 Legge dei Grandi Numeri semplificata

Legge dei Grandi Numeri – semplificata
Data X e una successione $X_1, ..., X_n$
di variabili aleatorie di ugual legge,
nell'ipotesi di indipendenza
e salvo rari casi capricciosi
la media empirica/sperimentale
$\frac{X_1+...+X_n}{n}$ tende
alla speranza matematica $E(X)$
cioè alla media "vera" di X.

BOZZA - DRAFT

In pratica: *sperabilmente* troveremo *circa* la speranza matematica di una v.a. di densità sconosciuta – com'è in generale in Statistica, e nella pratica – facendo la media di *molti* valori di quella v.a..

Ancora più terra-terra: abbiamo un astragalo di capra, osso vagamente cubico, segniamo 1 pallino su una faccia e 0 pallini sulle altre: la vera probabilità che esca "pallino" ovvero 1 si trova con

$$\frac{0+0+1+0+1+...+1+0+0}{n} \tag{128}$$

con un grande numero n di lanci, e gli 1 e 0 al posto giusto; e potrebbe essere circa $\frac{1}{6}$ con un buon astragalo piuttosto regolare.

Farmaceuticamente: la "vera" probabilità che un certo farmaco guarisca (per esempio, a 5 anni) è circa (com'è comunque ovvio)

$$\text{P(funziona)} \approx \frac{\text{numero guariti}}{\textbf{grandissimo} \text{ numero trattati}} \qquad (\alpha)$$

e impariamo a memoria solo questa (α) di questo paragrafo.
Che viene dalla (128) coi valori 1 per *vivo* e 0 per *morto*:

$$\frac{morto+morto+vivo+morto+vivo+...+vivo+morto+morto}{n}$$

Resto della Lezione escluso dal corso breve – 40 ore.

48.3 Alcune precisazioni

Similmente a quanto sopra detto avviene per qualunque p_1 fra 0 e 1 che sia la probabilità della testa della moneta (che se $p_1 \neq 0.5$ è non regolare): detto n il numero di lanci, e associato l'1 alla testa e 0 alla croce,

proporzione empirica di teste $\bar{p}_{n,1} = \frac{\#teste}{n} \to p_1$

proporzione empirica di croci $\bar{p}_{n,0} = \frac{\#croci}{n} \to p_0 := 1 - p_1$.

Similmente per un dado avremo 6 limiti p_1,...,p_6, cioè le proporzioni empiriche dei risultati tenderanno alle probabilità *vere* dei vari risultati, per esempio, per un dado regolare, sempre $\frac{1}{6}$:

$$\forall\, k \in \{1, ..., m\} \quad \bar{p}_{k,n} \to p_k = P(X = k) \tag{129}$$

e per il dado $m = 6$, e per una moneta invece $k \in \{0, 1\}$.

Questo tendere però non è quello deterministico, dei limiti delle successioni della matematica: seppure – come si può dimostrare – ha probabilità 0, rimane comunque possibile (!) che un dado *regolare* dia sempre 5, proprio *per sempre*, e allora in quel caso

$$\bar{p}_{5,n} = \frac{\#\, uscite\ del\ 5}{n} = \frac{n}{n} \equiv 1 \not\to \frac{1}{6} = P(X = 5) = p_5.$$

(Si noti però che questo evento possibile ha probabilità 0).

Esercizio. Ipotizzare e graficare le $\bar{p}_{k,n}$ per $n = 100$, $k = 1, ..., 6$.

48.4 Limite in probabilità e Legge dei Grandi Numeri

In quanto detto, resta non definito cosa si intende per il "tendere" ai numeri p_k, e si è ben detto che ci possono essere casi sfortunatissimi. Si tratta di un tendere probabilistico, non deterministico com'è quello dei limiti delle funzioni reali di variabile reale. Esso è precisato e inquadrato dal concetto di *convergenza in probabilità* di una successione di variabili aleatorie X_n, che definiamo senza insistervi particolarmente. Si immagini la X_n di cui parliamo come la proporzione empirica $\bar{p}_{1,n}$ di teste dopo n lanci, che, sì, è

BOZZA - DRAFT

una variabile aleatoria, "prima" di fare i lanci. Il limite della convergenza in probabilità di una successione di variabili aleatorie è esso stesso in generale una variabile aleatoria; solo che nell'esempio prima considerato è la variabile aleatoria discreta

$$X := \binom{0.5}{1} = p$$

che vale 0.5 con probabilità 1. (Variabile aleatoria *costante*). Ma in generale il limite X di una convergenza in probabilità è proprio una variabile aleatoria con una funzione di ripartizione non banale, ed è una variabile aleatoria discreta o continua.

BOZZA DRAFT

Definizione. Diremo che X_n converge in probabilità a X

$$X_n \to^P X \text{ se } \lim_{n\to\infty} P(|X_n - X| \geq \eta) = 0 \quad \forall\, \eta > 0 \tag{130}$$

(o indifferentemente con $> \eta$). (Ha un valore teorico, in questa trattazione elementare la useremo solo una volta fra poco: ci basta conoscerla e capirla).

Legge dei Grandi Numeri. Sia $(X_n)_n$ una successione di variabili aleatorie indipendenti di ugual legge con speranza matematica μ e varianza σ^2. Allora per la *media empirica*

$$\bar{X}_n := \frac{1}{n}(X_1 + \ldots + X_n)$$

$$\text{è} \quad \forall\, \eta > 0 \qquad \lim_{n\to\infty} P(|\bar{X}_n - \mu| \geq \eta) = 0$$

ovvero equivalentemente, **nelle ipotesi dette** (poco stringenti)

nelle ipotesi sopradette (poco stringenti)
$$\bar{X}_n \to^P \mu$$
ovvero:
la media empirica, sperimentale,
tende (in probabilità)
alla media "vera",
la speranza matematica
(β)

Detto altrimenti: *sperabilmente* (qua è il senso probabilistico) troveremo *circa* la speranza matematica di una v.a. di densità sconosciuta – com'è in generale in Statistica, e nella pratica – facendo la media di *molti* valori tratti (indipendentemente, ovvio) da quella v.a. (Il senso del limite è nelle parole "circa" e "molti").

Con $X_h := 1$ per testa e 0 altrimenti per $h = 1, ..., n$, si riottiene il primo caso considerato, con $\bar{X}_n$ la proporzione empirica $\bar{p}_{1,n}$.
Per una moneta regolare la frazione di teste tende in probabilità a $\frac{1}{2}$.
Per un dado regolare la frazione di risultati 3 tende in probabilità a $\frac{1}{6}$.

Possiamo fare i calcoli *esattamente*. La probabilità di k teste è

$$p_k = \binom{n}{k} p^k (1-p)^{n-k}$$

e ipotizzando una moneta regolare ($p = 1/2$, $p^k(1-p)^{n-k} = 2^{-n}$)

$$p_k = \binom{n}{k} 2^{-n}$$

e in particolare la probabilità i fare tante teste quante croci, che è 0 se n è dispari, per n pari è

$$\binom{n}{n/2} 2^{-n}$$

Con n piccolo è possibile fare facilmente il calcolo esatto, per esempio per $n := 6$ la probabilità è $\frac{5}{16}$ e per $n := 20$

$$\binom{20}{10} 2^{-20} \approx 0.176197 \quad \text{circa 1 su 6.}$$

L'esatto pareggio allora è alquanto improbabile[163] già con $n = 6$.

[163] La (131) con $n := 20$ ci dà

$$10 - 2\sqrt{5} \leq X \leq 10 + 2\sqrt{5}$$

BOZZA - DRAFT

Consideriamo n lanci di moneta con probabilità $\frac{1}{2}$ di fare testa.
Con la Disuguaglianza di Chebyshev si trova[164] che con probabilità almeno del 75% il numero di teste verifica

$$\frac{n}{2} - \sqrt{n} \le X \le \frac{n}{2} + \sqrt{n}. \tag{131}$$

Allora con un milione di lanci, almeno al 75% il numero di teste sta fra 499 000 e 501 000.
Nella prossima Lezione vedremo che in effetti quella probabilità è $> 95.4\%$, molto di più.

---

cioè

$$5.527... \le \textit{numero di teste} \le 14.472...$$

ovvero

$$\textit{numero di teste} = 6 \vee 7 \vee 8 \vee 9 \vee 10 \vee 11 \vee 12 \vee 13 \vee 14$$

e questo evento ha probabilità

$$p_6 + ... + p_{14} = 2^{-20}\left(\binom{20}{6} + ... + \binom{20}{14}\right) =$$

$$= \frac{125647}{131072} \approx 0.959 = 95.9\%.$$

Prima si era detto *almeno* 75%, ora si trova *esattamente* 0.95.... (Ma questo calcolo per $n := 1\,000\,000$ è improbo).

[164] Per lo studente interessato:

$$P(|X - E(X)| > c) \le \frac{Var(X)}{c^2} \qquad \forall\, c$$

equivale, con l'evento complementare, a

$$P(|X - E(X)| \le c) \ge 1 - \frac{Var(X)}{c^2}.$$

Se X è il contatore di teste (successi) in n lanci, allora $X \sim B(n,k)$. Essendo per la $B(n,k)$ la varianza $np(1-p)$ e la speranza matematica np, con la moneta regolare $n/4$ e $n/2$ rispettivamente,

$$P(|X - n/2| \le c) \ge 1 - \frac{n/4}{c^2}$$

e fissando $c := 2\sigma = 2\sqrt{VarB(n,k)} = 2\sqrt{n/4} = \sqrt{n}$

$$P(|X - n/2| \le \sqrt{n}) \ge 1 - \frac{\sigma^2}{(2\sigma)^2}$$

cioè

$$P\left(\left(\left|X - \frac{n}{2}\right| \le \sqrt{n}\right) \ge \frac{3}{4} = 0.75 = 75\%\right.$$

e ricordando che $|f(x)| \le g(x)$ equivale a $-g(x) \le f(x) \le g(x)$ si trova la (131).

BOZZA - DRAFT

Si noti che effettivamente $\frac{501\,000}{1\,000\,000} \approx 0.5$, e similmente con 499 000. Detto in altri termini

$$\frac{\frac{n}{2} \pm \sqrt{n}}{n} = \frac{1}{2} \pm \frac{1}{\sqrt{n}} \approx \frac{1}{2} \quad per \quad n \gg .$$

Ripetiamo che **si potrebbe dimostrare che la differenza fra teste e croci tenderà in generale ad essere sempre più grande, non più piccola!** Eppure, la proporzione tende al 50%. (Tende "in probabilità", ora sappiamo).

Nota. Di questa Lezione ci si dovrebbe aspettare che lo studente conosca a memoria le formule (129), (130), (α) e (β), oltre a tutto l'inquadramento teorico della questione.

BOZZA - DRAFT

49 Approssimazione Normale

Solo paragrafo 49.6 nel corso breve – 40 ore.

Oltre alla convergenza in probabilità già vista, esistono altri tipi di convergenza per successioni di variabili aleatorie e in questa trattazione elementare se ne considererà una che ora vediamo.

BOZZA - DRAFT

49.1 Convergenza in legge

Definizione. Diremo che X_n **converge in legge** a X

$$X_n \to^L X \text{ se } \lim_{n\to\infty} F_{X_n}(x) = F_X(x) \tag{132}$$

(con ovvio significato dei simboli) per ogni punto x in cui $F_X(x)$ è continua. (Ha un valore teorico, non la useremo realmente in questa trattazione elementare: ci basta conoscerla e capirla).

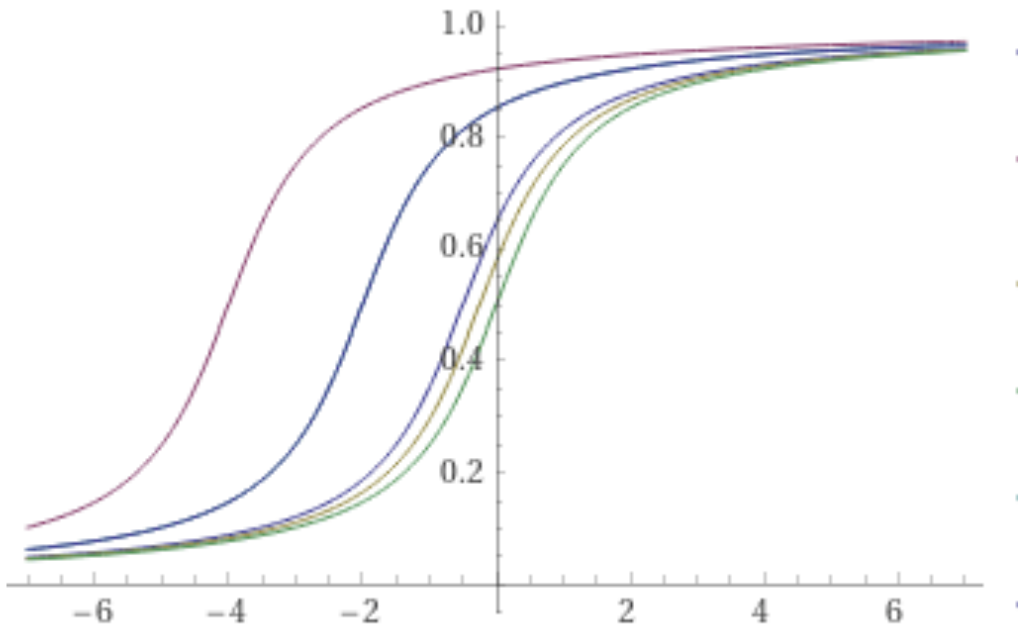


Figure 56: Verde: funzione di ripartizione di una certa v.a. X; altri colori, dall'alto verso il basso: funzioni di ripartizione di X_1, X_2, X_3, X_4; non sono rappresentati gli infiniti grafici delle funzioni di ripartizione di X_5, X_6... ma li si immagini sempre più prossimi alla curva in verde, realizzando la convergenza in legge di X_n a X. Screenshot da WolframAlpha.

BOZZA

In sostanza la convergenza in legge corrisponde alla convergenza delle f.r. di X_n alla f.r. di X. (Nei punti dove essa è continua).

Si noti che le X_n possono essere discrete anche se X è continua. (Si immaginino curve a gradino, coi gradini sempre più piccoli, "fino a tendere" a un grafico liscio).

Rimarchiamo che la convergenza avviene per ogni x fissato (e in effetti neanche in tutti gli x è richiesta ma solo in quelli in cui la funzione di ripartizione $F_X(x)$ di X è continua) e quindi nella figura 56 la convergenza avviene per così dire dall'alto verso il basso e non da sinistra verso destra come magari potrebbe apparire.

49.2 Teorema Limite Centrale

Sia $(X_n)_n$ una successione di variabili aleatorie indipendenti (discrete o continue) di ugual legge con speranza matematica μ e varianza σ^2. Allora la standardizzazione della somma delle prime n tende in legge ad una normale standard. Ciò si esprime in base alla (132) con questa formula

$$\exists\, X \sim N(0,1) \qquad S_n^* := \frac{X_1 + \ldots + X_n - n\,\mu}{\sigma\,\sqrt{n}} \to^L X \tag{133}$$

non semplicissima, di valore teorico. Da un punto di vista più applicativo, per le variabili aleatorie ora considerate vale allora l'

Approssimazione Normale

$$P\Big(X_1 + \ldots + X_n \le x\Big) \approx \Phi\Big(\tfrac{x-n\mu}{\sigma\sqrt{n}}\Big) \tag{134}$$

qua sopra e qua sotto le condizioni da sapere

BOZZA DRAFT

e tradizionalmente questa approssimazione si ritiene sufficientemente buona per $n \ge 30$ (secondo altri Autori $n \ge 50$) in questi casi:

* se le $X_k \sim B(m,p)$ con $m\,p \ge 5$ e $m\,(1-p) \ge 5$

* per tutte le altre distribuzioni degli esercizi scolastici e anche di questa trattazione, e spessissimo anche della pratica.

Di fatto, storicamente si usava considerare quasi normale già la standardizzazione

$$S_{12}^* = X_1 + \ldots + X_{12} - 6 \tag{135}$$

della somma di 12 valori dell'uniforme continua in $[0,1]$.
(La soprastante espressione si trova dalla (133) usando la formula $(a+b)/2$ della media di $\mathbb{U}[a,b]$, che dà 0.5 da moltiplicare per $n = 12$, e la formula $(b-a)^2/12$ della varianza che dà $\sigma^2 = 1/12$, e allora $\sigma = \sqrt{1/12}$, che si semplifica con $\sqrt{n}$).
Proprio così funzionava una vecchia routine FORTRAN che produceva valori asseritamente distribuiti secondo la normale standard.

(Oggigiorno si può fare molto meglio con altre formule).

Osservazione. Allora, la forma a campana e in particolare la variabile aleatoria normale standard è una sorta di “attrattore" per le variabili aleatorie, perché anche se ne sono alquanto diverse, la loro somma standardizzata, a certe condizioni sopra dette, tende proprio alla $N(0,1)$, in legge.

Teorema. La convergenza in probabilità implica quella in legge.

49.3 Simulazione di una variabile aleatoria uniforme

Risulta fondamentale per la simulazione casuale di processi disporre di una v.a. $U_1 \sim \mathbb{U}[0,1[$*, uniforme continua su* $[0,1[$.
(Oppure $U_1 \sim \mathbb{U}[0,1]$, intervallo chiuso, comunque il valore 1 non lo vedremo “quasi sicuramente" mai perchè ha probabilità 0).

BOZZA - DRAFT

In molte applicazioni Biomediche serve simulare virtualmente oggetti reali: si immaginino particelle o microbi o animali o persone o veicoli vaganti, oppure la simulazione del funzionamento di un sistema sanitario prima di modificare quello vero, o la simulazione degli incontri fra contagiati da un virus.
È già illuminante questa citazione

> The random selection was repeated 1000 times to build the reference distance distribution

dall’articolo scientifico *In silico drug repurposing in COVID-19: A network-based analysis*, e *in silico* significa “(simulato) al computer", latinismo moderno che ricalca *in vivo* e *in vitro*.

Si vedano poi le figure che rappresentano la simulazione del girovagare dei cinghiali in città nella Figura 6 (Simulated trajectories of 10 boars during the night (...)) nell’articolo scientifico[165] *Pigs in space: An agent-based model of wild boar (Sus scrofa) movement*

[165] M. Toger et al., (2018) Landscape and Urban Planning, 173, pp.70-80.

into cities, con gli incidenti stradali che possono causare e le malattie che possono portare, fra cui le zecche e forse pure il covid-19 (attualmente nel 2022 appare verosimile[166] un caso di trasmissione all'uomo dal cervo dalla coda bianca).

Esempio di Farmacia Veterinaria. Negli ultimi anni grande apprensione sta suscitando la febbre suina africana, che devasta gli allevamenti di maiali causando danni economici smisurati, e procede dall'Asia verso l'Europa attraverso cinghiali; nel 2022 è stata segnalata nell'Italia continentale e già da anni un ceppo diverso era in Sardegna. Nell articolo scientifico[167] *Simulation of transmission and persistence of African swine fever in wild boar in Denmark* si simulano contatti contagiosi di febbre suina africana fra cinghiali in Danimarca:

> the model runs on a landscape data file containing about 23,000 grid cells, each cell with a size of 4 km^2, representing Denmark, but excluding the Island of Bornholm. The model is initialized in two scenarios of initial number of groups; one scenario with 50 groups of wild boar and the other with 100 groups (Section 2.3.1). Each group was randomly allocated to one habitat cell, reflecting the group home range. (...) one of the 3 cells in that direction is selected randomly (...) Each day dispersing animals may move randomly between 0 and 1 cell
> (Enfasi aggiunta)

BOZZA - DRAFT

Se si vuole scegliere a caso fra 3 celle, come si legge sopra, è sufficiente simulare la variable aleatoria uniforme discreta $D_3 \sim \mathbb{U}\{1,3\}$. Questo si può fare a partire da una v.a. uniforme continua $U_1 \sim \mathbb{U}[0,1[$ (uniforme sull'intervallo $[0,1[$) con la formula

$$D_3 := 1 + \lfloor 3\,U_1 \rfloor \tag{136}$$

[166] https://www.nationalgeographic.com/animals/article/first-potential-case-of-deer-to-human-covid-transmission-found

[167] Halasa T, Boklund A, Bøtner A, Mortensen S, Kjær LJ. Simulation of transmission and persistence of African swine fever in wild boar in Denmark. Prev Vet Med. 2019 Jun 1;167:68-79. doi: 10.1016/j.prevetmed.2019.03.028. Epub 2019 Apr 1. PMID: 31027724.

definita con la parte intera. Similmente sostituendo 3 con 6 si simula un dado (utile per videogiochi e slot mashines).

Risulta quindi fondamentale per la simulazione casuale di processi disporre di una $U_1 \sim \mathbb{U}[0,1[$.

Molti software implementano una funzione, che tentativamente producono un numero a caso fra 0 e 1, con tutta una serie di problematicità.

E no, non è perfettissimamente casuale uniformemente distribuito in $[0,1[$, e uno potrebbe passare la vita a studiare questo argomento complesso; Google nel 2022 trova 334 000 000 di risultati per *random numbers generation*.
Leggiamo in https://en.wikipedia.org/wiki/Middle-square_method

> In the 1949 talk, Von Neumann quipped that "Anyone who considers arithmetical methods of producing random digits is, of course, in a state of sin."

Su Excel `RAND()`, e su WolframAlpha, e magari potresti provarlo,
`RandomReal[]`

Se non si dispone di internet e si vuole un singolo numero casuale fra 0 e 1, un metodo grossolano potrebbe essere procurarsi la data di compleanno di una persona di cui prima non sapevamo la data di compleanno, per esempio 30 marzo, contare i giorni dal 1 gennaio a quella data in un anno non bisestile, e sono 89 nell'esempio considerato, ottenendo un numero più o meno casuale fra 1, 2, 3...365. Sottraendo 1 e poi dividendo per 364 si ottiene un numero reale abbastanza uniformemente distribuito fra 0 e 1 compresi, e nell'esempio detto viene $\approx 0.241\,758$. (Comunque i compleanni non sono proprio uniformemente distribuiti nell'anno). Da esso possiamo simulare un dado con una formula analoga alla (136),

$$D_6 := 1 + \lfloor 6\,U_1 \rfloor$$

e col numero ottenuto viene 2. Abbiamo lanciato un dado virtuale ottenendo 2, oppure abbiamo scelto il gene 2 fra 6. Possiamo così ottenere un numero pseudocasuale anche fra 1 e 7, fra 1 e 8, eccetera: ma non certo fra 1 e 100. (Il lettore curioso e attento potrà studiare

BOZZA - DRAFT

la questione e vedere dove cade il metodo per numeri troppo grandi, diciamo oltre il 30; i numeri casuali, anzi pseudocasuali, sono molto problematici).

49.4 Simulazione di una variabile aleatoria normale

Apposite formule permettono di simulare una gran varietà di variabili aleatorie a partire da una v.a. $U_1 \sim \mathbb{U}[0,1[$.

Se si vuole una singola realizzazione di una variabile aleatoria normale standard, o poche, diciamo 10, il modo migliore è digitare su WolframAlpha

```
10 random numbers normal
```

Se non è disponibile la rete, si può usare la formula (135) e per avere i 12 valori più o meno uniformemente distribuiti in $[0,1]$ si dividano per 30 una dozzina di... date di compleanno di mesi di 31 giorni, diminuite di 1. In pratica, in modo equivalente e più semplice, si sommino le 12 date (nel senso del solo giorno del mese), si sottragga 12, si divida per 30, e si sottragga 6. Il valore che così si sta per ottenere è – prima di ottenerlo ovviamente – una v.a. S approssimativamente di legge $N(0,1)$.
Dal numero $S(\omega)$ ottenuto, determinazione di una v.a. S distribuita (approssimativamente) secondo $N(0,1)$, si può ottenerne uno distribuito secondo $N(\mu,\sigma^2)$ con la nota formula

$$Y = \sigma\, S + \mu$$

Così si possono simulare un'infinità di fenomeni. Per esempio per i neonati italiani di genere femminile viene riferita una distribuzione normale di media 3100 g e deviazione standard 200 g.
In pratica

compleanno 1+
compleanno 2+
...
compleanno 12

BOZZA - DRAFT

= somma
Sottrarre 12
Dividere per 30
sottrarre 6 (e fin qua è più o meno normale standard)
moltiplicare per 200
sommare 3100 → peso bimba

Molto più agevolmente e precisamente, con WolframAlpha
10 random numbers normal mean=3100 sigma=200

BOZZA - DRAFT

La simulazione di una possibile prossima pandemia

Nella simulazione *Catastrophic Contagion* dell'ottobre 2022 è stata simulata una possibile prossima pandemia.

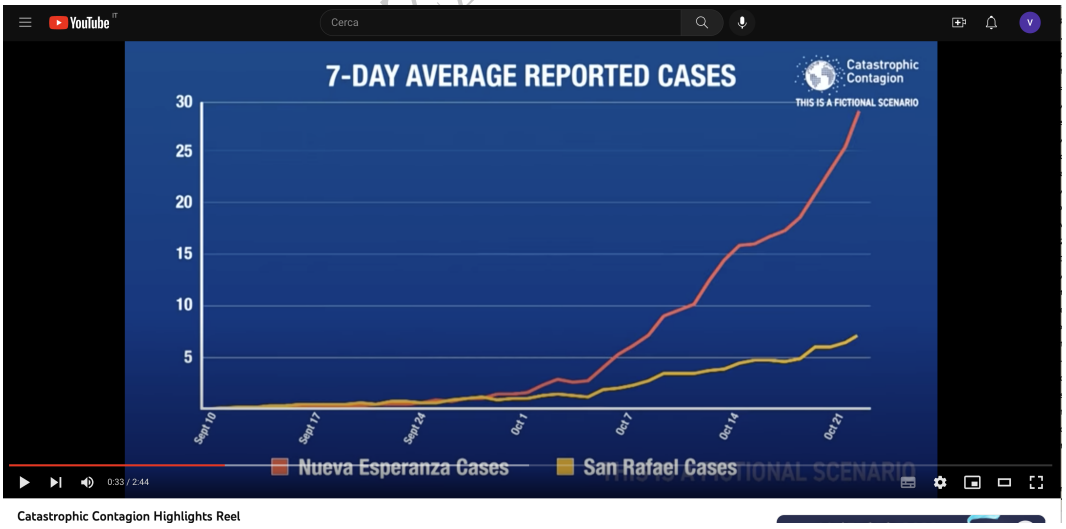


Figure 57: Screenshot da YouTube https://www.youtube.com/watch?v=M-skwhRFWe0 consultato il 12 dicembre 2022, del canale centerforhealthsecurity @biosecuritycntr.

Si può facilmente ipotizzare che in questa simulazione siano stati prodotti moltissimi numeri casuali.

Questa simulazione si è svolta in un assordante silenzio mediatico, almeno relativamente all'Italia, forse perchè la precedente, Event 201 del 18 ottobre 2019, precedette di poco la pandemia del covid-19, e perchè gli esseri umani sono mediamente meschi- ni ed egoisti, non pronti per il meraviglioso esproprio globale – *"Welcome to 2030. I own nothing, have no privacy, and life has never been better"* – propagandato da uno degli organizzatori, il World Economic Forum, che pure scrive*"The pandemic represents a rare but narrow window of opportunity to reflect, reimagine, and reset our world"*, e forse anche meschinamente non capi- scono la filantropia del filantropo Bill Gates, aperto sostenitore della riduzione numerica dell'umanità, un altro dei principali organizzatori, e fisicamente là presente, per il bene di tutti, che lui conosce; ha realizzato *Windows*.

BOZZA - DRAFT

Leggiamo in https://www.centerforhealthsecurity.org/our-work/exercises/2022-catastrophic-contagion/ "The Johns Hopkins Center for Health Security, in partnership with WHO and the Bill & Melinda Gates Foundation, conducted Catastrophic Contagion, a pandemic tabletop exercise at the Grand Challenges Annual Meeting in Brussels, Belgium, on October 23, 2022.

The extraordinary group of participants consisted of 10 current and former Health Ministers and senior public health officials from Senegal, Rwanda, Nigeria, Angola, Liberia, Singapore, India, Germany, as well as Bill Gates, co-chair of the Bill & Melinda Gates Foundation.

The exercise simulated a series of WHO emergency health advisory board meetings addressing a fictional pandemic set in the near future. Participants grappled with how to respond to an epidemic located in one part of the world that then spread rapidly, becoming a pandemic with a higher fatality rate than COVID-19 and disproportionately affecting children and young people.

Participants were challenged to make urgent policy decisions with limited information in the face of uncertainty. Each problem and choice had serious health, economic, and social ramifications."

49.5 Sul naturale formarsi delle campane

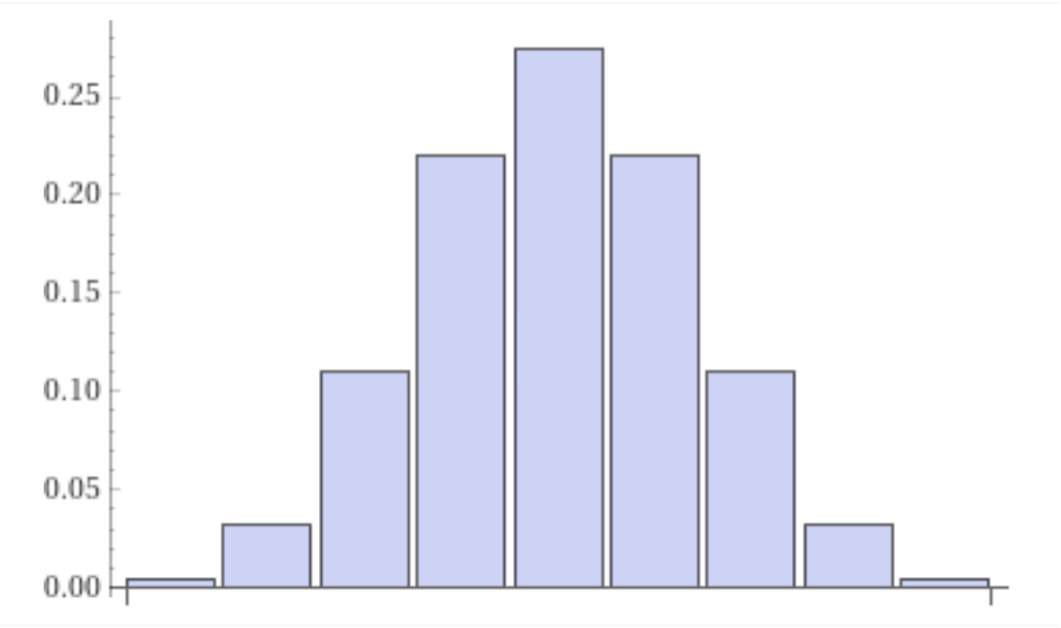


Figure 58: Densità binomiale $B\left(8, \frac{1}{2}\right)$. I possibili valori della variabile aleatoria sono 0,...,8 e il più probabile è 4. Screenshot da WolframAlpha.

BOZZA

Che una variabile aleatoria somma di tante "copie" indipendenti di una stessa variabile aleatoria, per esempio il punteggio totale di 2 lanci di moneta con facce 0 e 1, poi di 3 lanci, poi di 4, eccetera, tenda *spesso* (non sempre: ci sono ben le ipotesi, certamente valide per la moneta) a produrre le forme a campana, è evidente già dalla densità binomiale, che appunto valuta la probabilità di k successi in n prove indipendenti, ovvero, detto altrimenti, è la densità della v.a. che somma gli uni delle teste di n lanci (avendo posto croce=0). Si veda la soprastante figura.

Si noti che la forma è sì a campana, ma lontanissima da quella della densità normale standard, che è simmetrica rispetto all'asse delle ordinate, mentre qua abbiamo una simmetria intorno all'asse verticale $n = 4$. Questo è dovuto al fatto che non si è fatta la standardizzazione della somma, che avrebbe fra l'altro comportato proprio sottrarre 4, che è $n\,\mu$, alla somma (si veda la formula in

nota nel paragrafo sul Teorema Limite Centrale).

Esempio. Calcoleremo la probabilità di ottenere più di 28 teste in 50 lanci di moneta equilibrata.
I risultati X_k hanno legge $B(1, \frac{1}{2})$, con speranza matematica $\mu = \frac{1}{2}$ e varianza $\sigma^2 = 1 \cdot \frac{1}{2}\left(1 - \frac{1}{2}\right) = \frac{1}{4}$, allora $\sigma = \frac{1}{2}$, e allora

$$P\Big(X_1 + ... + X_{50} > 28\Big) = 1 - P\Big(X_1 + ... + X_{50} \le 28\Big) \approx$$

$$\approx 1 - \Phi\Big(\frac{28 - 50 \cdot \frac{1}{2}}{\frac{1}{2}\sqrt{50}}\Big) \approx 1 - \Phi(0.849) \approx 1 - 0.802 \approx 0.2 = 20\%.$$

La classica approssimazione (Shah 1985)

$$\Phi(x) \approx \begin{cases} \frac{x\,(4.4-x)}{10} + 0.5 & se\ 0 \le x \le 2.2 \\ 0.99 & se\ 2.2 < x < 2.6 \\ 1 & se\ x \ge 2.6 \end{cases}$$

in 0.849 darebbe 0.801 invece del più preciso 0.802 che troviamo con Wolframalpha. Ma per il risultato finale ragionevolmente espresso da 20% non cambia nulla.

BOZZA - DRAFT

ESERCIZIO$_\mu$
Calcolare con l'approssimazione normale la probabilità di ottenere con 1 000 000 di lanci di moneta [regolare] un numero di teste compreso fra 499 000 e 501 000 compresi.

SVOLGIMENTO
Per il contatore di teste

$$X_1 + .. + X_{1\,000\,000}$$

cerchiamo

$$q = P(499\,000 \le X \le 501\,000) =$$

$$= P(X \le 501\,000) - P(X < 499\,000) =$$

$$= P(X \le 501\,000) - P(X \le 498\,999) \qquad (137)$$

Il sopradetto contatore di teste è somma di 1 000 000 di variabili aleatorie indipendenti $B\left(1, \frac{1}{2}\right)$, ciascuna con media e varianza

$$\mu = \frac{1}{2} \qquad \sigma^2 = 1 \cdot \frac{1}{2}\left(1 - \frac{1}{2}\right) = \frac{1}{4}$$

e allora

$$P(X_1 + ... + X_{1\,000\,000} \leq x) \approx$$

$$\approx \Phi\Big(\frac{x - n\mu}{\sigma\sqrt{n}}\Big) = \Phi\Big(\frac{x - 1\,000\,000\,\frac{1}{2}}{\frac{1}{2}\sqrt{1\,000\,000}}\Big) =$$

$$= \Phi\Big(\frac{x - 500\,000}{500}\Big)$$

e applicandola alla (139) la probabilità cercata è

$$q = \Phi\Big(\frac{501\,000 - 500\,000}{500}\Big) - \Phi\Big(\frac{498\,999 - 500\,000}{500}\Big) =$$

$$= \Phi\Big(\frac{1\,000}{500}\Big) - \Phi\Big(\frac{-1\,001}{500}\Big) \approx$$

$$\approx \Phi(2) - \Phi(-2) =$$

e con la formula di simmetria

$$\Phi(2) - (1 - \Phi(2)) = 2\,\Phi(2) - 1 \approx$$

e ricordando la $\phi_{0.975} \approx 1.96 \approx 2$

$$\approx 2 \cdot 0.975 - 1 = 0.95 = 95\%$$

BOZZA-DRAFT

49.6 Lo scarto di una radice quadrata dalla metà

Teorema. C'è circa il 5% di probabilità che lanciando una moneta regolare il numero di teste disti dalla metà di un *grande* numero dei lanci più della radice quadrata del numero dei lanci, ovvero, che stia entro quello scarto, circa 95%:

$$\boxed{P\left(\frac{n}{2} - \sqrt{n} \leq \text{ numero di teste } \leq \frac{n}{2} + \sqrt{n}\right) \approx 95\% \qquad n \gg} \quad (138)$$

Un calcolo più preciso troverebbe 95.4 invece di 95 e 4.6 invece di 5.

Insomma, con 10 000 lanci, fra 4900 e 5100: così è chiarissimo.

Questo risultato ci consentirà in 51.4 una spettacolare anticipazione del nucleo fondamentale della Statistica Inferenziale.

49.7 Calcolo sullo scarto della radice quadrata

Qua (per ottenere l'approssimazione 95% e non la più precisa 95.4%) calcoleremo con l'approssimazione normale la probabilità di ottenere con $n \geq 100$ lanci di moneta [regolare] un numero di teste compreso fra $\frac{n}{2} - \sqrt{n}$ e $\frac{n}{2} + \sqrt{n}$ compresi. Per semplicità di calcolo ipotizzeremo n quadrato perfetto pari ma il risultato soprascritto che troveremo vale anche senza quell'ipotesi.

BOZZA - DRAFT

Per il contatore di teste

$$X := X_1 + .. + X_n$$

cerchiamo

$$q = P\left(\frac{n}{2} - \sqrt{n} \leq X \leq \frac{n}{2} + \sqrt{n}\right) =$$

$$= P(X \leq n/2 + \sqrt{n}) - P(X \leq n/2 - \sqrt{n} - 1) \tag{139}$$

Il sopradetto contatore di teste è somma di n di variabili aleatorie indipendenti $B\left(1, \frac{1}{2}\right)$, ciascuna con media e varianza

$$\mu = \frac{1}{2} \qquad \sigma^2 = 1 \cdot \frac{1}{2}\left(1 - \frac{1}{2}\right) = \frac{1}{4}$$

e allora

$$P(X_1 + ... + X_n \leq x) \approx$$

$$\approx \Phi\Big(\frac{x - n\mu}{\sigma\sqrt{n}}\Big) = \Phi\Big(\frac{x - n\,\frac{1}{2}}{\frac{1}{2}\sqrt{n}}\Big) =$$

$$= \Phi\Big(\frac{x - n/2}{\frac{1}{2}\sqrt{n}}\Big)$$

e applicandola alla (139) la probabilità cercata è

$$q = \Phi\Big(\frac{n/2 + \sqrt{n} - n/2}{\frac{1}{2}\sqrt{n}}\Big) - \Phi\Big(\frac{n/2 - \sqrt{n} - 1 - n/2}{\frac{1}{2}\sqrt{n}}\Big) =$$

$$= \Phi\Big(\frac{\sqrt{n}}{\frac{1}{2}\sqrt{n}}\Big) - \Phi\Big(\frac{-\sqrt{n} - 1}{\frac{1}{2}\sqrt{n}}\Big) =$$

$$= \Phi(2) - \Phi\left(-2 - \frac{2}{\sqrt{n}}\right) =$$

per la formula di simmetria

$$= \Phi(2) - 1 + \Phi\left(2 + \frac{2}{\sqrt{n}}\right) = \star$$

ed eliminando senza qua voler approfondire il minuscolo $\frac{2}{\sqrt{n}}$ (minuscolo perchè $n \geq 100$)

$$= \Phi(2) - 1 + \Phi(2) =$$

$$= 2\,\Phi(2) - 1 \approx$$

e ricordando la $\phi_{0.975} \approx 1.96 \approx 2$

$$\approx 2 \cdot 0.975 - 1 = 0.95 = 95\%$$

Volendo fare il calcolo con più precisione, riprendendo da $\star$ (usando la $f(x+h) \approx f(x) + h\,f'(x)$ valida per piccolo h)

$$\star \quad = \Phi(2) - 1 + \Phi(2) + \frac{2}{\sqrt{n}}\,\Phi'(2) =$$

$$= 2\,\Phi(2) - 1 + \frac{2}{\sqrt{n}}\,\Phi'(2) =$$

$$\approx 2 \cdot 0.975 - 1 + \frac{2}{\sqrt{n}}\,\Phi'(2) =$$

$$= 0.95 + \frac{2}{\sqrt{n}}\,\Phi'(2) =$$

che è più del 95%.

BOZZA - DRAFT

49.8 Lo scarto di mezza radice quadrata dalla metà

Nelle stesse condizioni del paragrafo precedente lo scarto dalla metà entro metà della radice quadrata avviene con probabilità $\approx 68\%$:

$$P\left(\frac{n}{2} - \frac{\sqrt{n}}{2} \leq \text{ numero di teste } \leq \frac{n}{2} + \frac{\sqrt{n}}{2}\right) \approx 68\% \qquad n \gg \quad (140)$$

Insomma, con 10 000 lanci, fra 4950 e 5050: così è chiarissimo.

Calcolo. Calcoleremo con l'approssimazione normale la probabilità di ottenere con $n \geq 100$ di lanci di moneta (regolare) un numero di teste compreso fra $\frac{n}{2} - \frac{\sqrt{n}}{2}$ e $\frac{n}{2} + \frac{\sqrt{n}}{2}$ compresi. Per semplicità di calcolo si ipotizzeremo n quadrato perfetto pari ma il risultato soprascritto che troveremo vale anche senza quell'ipotesi.

Per il contatore di teste

$$X_1 + .. + X_n$$

cerchiamo

$$q = P(n/2 - \sqrt{n}/2 \leq X \leq n/2 + \sqrt{n}/2) =$$
$$= P(X \leq n/2 + \sqrt{n}/2) - P(X \leq n/2 - \sqrt{n}/2 - 1) \tag{141}$$

Il sopradetto contatore di teste è somma di n di variabili aleatorie indipendenti $B\left(1, \frac{1}{2}\right)$, ciascuna con media e varianza

$$\mu = \frac{1}{2} \qquad \sigma^2 = 1 \cdot \frac{1}{2}\left(1 - \frac{1}{2}\right) = \frac{1}{4}$$

e allora

$$P(X_1 + ... + X_n \leq x) \approx$$
$$\approx \Phi\Big(\frac{x - n\mu}{\sigma\sqrt{n}}\Big) = \Phi\Big(\frac{x - n\,\frac{1}{2}}{\frac{1}{2}\sqrt{n}}\Big) =$$
$$= \Phi\Big(\frac{x - n/2}{\frac{1}{2}\sqrt{n}}\Big)$$

e applicandola alla (141) la probabilità cercata è

$$q = \Phi\Big(\frac{n/2 + \sqrt{n}/2 - n/2}{\frac{1}{2}\sqrt{n}}\Big) - \Phi\Big(\frac{n/2 - \sqrt{n}/2 - 1 - n/2}{\frac{1}{2}\sqrt{n}}\Big) =$$
$$= \Phi\Big(\frac{\frac{1}{2}\sqrt{n}}{\frac{1}{2}\sqrt{n}}\Big) - \Phi\Big(\frac{-\frac{1}{2}\sqrt{n} - 1}{\frac{1}{2}\sqrt{n}}\Big) =$$
$$= \Phi(1) - \Phi\left(-1 - \frac{2}{\sqrt{n}}\right) =$$

per la formula di simmetria

$$= \Phi(1) - 1 + \Phi\left(1 + \frac{2}{\sqrt{n}}\right) =$$

ed eliminando senza qua voler approfondire il minuscolo $\frac{2}{\sqrt{n}}$

$$= \Phi(1) - 1 + \Phi(1) =$$

BOZZA - DRAFT

$$= 2\,\Phi(1) - 1 \approx$$

e ricordando la $\Phi(1) \approx 0.84$

$$\approx 2 \cdot 0.84 - 1 = 0.68 = 68\%$$

Nota. Di questa Lezione ci si dovrebbe aspettare che lo studente conosca a memoria le formule *contemporaneamente* numerate *et* riquadrate, oltre a tutto l'inquadramento teorico della questione.

BOZZA - DRAFT

50 Note finali sul Calcolo delle Probabilità

50.1 Sulle problematicità del Calcolo delle Probabilità

Per concludere – in qualche modo, drammaticamente – questo minicorso di Calcolo delle Probabilità, consideriamo i seguenti 3 problemi.

1) Che probabilità c'è che domani sorga il sole?

2) Un insettuccio, reale o virtuale, vaga di trattino in trattino su una carta a quadretti illimitata, e ad ogni quadrivio imboocca a caso uno dei 3 trattini diversi da quello appena percorso, e si ferma quando raggiunge un quadrivio già raggiunto precedentemente. Che probabilità cè che il suo vagare termini entro n passi?

3) Chiediamo a molta gente quanto costa il regalo preferito che ipoteticamente vorrebbero avere a Natale, o quanto è estesa in metri quadrati o chilometri quadrati l'isola deserta dei loro sogni, o qualunque altra cosa, anche qual era il primo valore numerico delle loro analisi del sangue, o il secondo, o l'ultimo, o praticamente... qualunque cosa. È più probabile che la prima cifra non nulla del numero detto inizi per 1, 2, 3, oppure per 4, 5, 6, 7, 8, 9?

BOZZA - DRAFT

Il secondo problema è difficilissimo e finora risolvibile per specifici valori di n, non per n generico; comunque il problema è chiarissimo: esiste una funzione $f(n)$ che è risposta al quesito, ma purtroppo attualmente non la conosciamo.
Il prinmo problema più che difficilissimo è impossibile, e nonostante sia evidente che il possibile sorgere del sole domani sia un evento ben definito, ci fa sorgere il dubbio che esista la probabilità di un tale evento. In effetti a tutt'oggi fa discutere ed è sostanzialmente privo di risposta univoca il problema di determinare qual è la probabilità che domani sorga il sole: vedasi https://en.wikipedia.org/wiki/Sunrise_problem. Ma ci sarebbe prima di tutto da chiarire

bene il senso da dare alla domanda.

Come abbiamo visto esistono (almeno) 4 concezioni della probabilità, solo parzialmente sovrapponibili.

Il problema della prima cifra di un numero che esprime un dato della realtà sensibile, “preso a caso" per quanto ciò possa significare, è affascinante e difficile, comunque abbastanza ben risolto, e ne viene che la probabilità complessiva delle cifre iniziali 1, 2 e 3 è circa 60%, delle cifre iniziali 4, 5, 6, 7, 8 e 9 circa 40%. Strano, vero? Uscite in strada e segnatevi i primi 10 numeri che vedete, oppure in un qualunque libro scientifico, pieno di numeri. Oppure, elencate le superfici dei maggiori laghi d’Italia, o d’Europa, o del mondo. Veramente molto strano, eppure è così. La questione, nota come “first digit problem", è stata abbastanza ben chiarita da T.P. Hill e A. Volčič alla fine del secolo scorso.

BOZZA - DRAFT

Alla fine di questa serie di Lezioni sul Calcolo delle Probabilità, osserviamo che esso giunge a livelli tremendamente complessi, lontanissimi dalla banale teoria svolta in questo testo elementare, essenzialmente propedeutica alla Statistica Inferenziale. In https://en.wikipedia.org/wiki/List_of_probability_journals una lista di riviste scientifiche di Probabilità. Senza dubbio il Calcolo delle Probabilità potrebbe riempire un intero corso di studi universitario.

In https://en.wikipedia.org/wiki/Ibragimov%E2%80%93Iosifescu_conjecture_for_%CF%86-mixing_sequences un problema la cui soluzione è ignota da mezzo secolo, e già la formulazione è alquanto complessa, comunque il lettore vi riconoscerà almeno i simboli $N(0,1)$ e $\mathrm{Var}(X_1+...+X_n)$ e le parole *random variables* (variabili aleatorie).

La posizione epistemologica della probabilità è non banale.

I valori della probabilità possono essere molto controintuitivi, come

mostra, fra altri, l'esempio dei compleanni coincidenti.

Esistono molti paradossi probabilistici, seppure spesso si spieghino bene con la teoria, e in effetti sono solo apparentemente dei paradossi: una lista è in https://en.wikipedia.org/wiki/Category:Probability_theory_paradoxes. Il seguente è particolarmente capriccioso.

50.2 Mettiamo in dubbio l'esistenza della probabilità

Esempio che fa dubitare dell'esistenza della probabilità. Alessio e Berto sono due italiani, gemelli identici maschi. Alessio legge che l'1% degli italiani hanno un certo gene. Essendo italiano ritiene di avere l'1% di probabilità di avere quel certo gene. (Il che è sostanzialmente corretto in base alle informazioni che ha). Berto legge che il 10% dei maschi europei hanno quello stesso gene. Essendo un maschi europeo ritiene di avere il 10% di probabilità di avere quel certo gene. (Il che è sostanzialmente corretto in base alle informazioni che ha). La cosa sorprendente è che essendo gemelli identici, entrambi hanno il gene, oppure non ce l'hanno.
Chissà qual è la probabilità che entrambi abbiano quel gene...

BOZZA - DRAFT

Si noti che a seconda che si consideri il soggetto come italiano, o maschio europeo, si hanno valori diversi per la – sfuggente – *probabilità di avere quel gene.*

La valutazione (numerica) della probabilità dipende da ciò che si sa del soggetto, non è qualcosa di associato inesorabilmente al soggetto stesso – quel qualcosa non esiste.

Quel qualcosa non esiste!

50.3 Sull'applicare le probabilità a singoli casi

Nonostante quanto evidenziato nel paragrafo precedente, molte persone leggono statistiche di quel genere, e *realmente* pensano che riguardino *loro stessi*, come se le malattie colpissero *a caso*, come una lotteria – che quella sì non guarda in faccia nessuno – senza relazione con la condizione economico-sociale, lo stile di vita e le dinamiche esistenziali pregresse, l'età, il genere, eccetera eccetera. (Fino a giungere al *singolo caso*).
(E l'alcol e il fumo e le droghe leggere e le droghe pesanti...)

Per fissare le idee si consideri quanto il covid grave è correlato statisticamente all'obesità e alla carenza di vitamina D, prima di autocalcolarsi le pseudo-probabilità di morire di covid così, per caso, tipo lotteria, basandosi su statistiche generiche del tipo "1 contagiato su 100 muore", o anche più dettagliate, per regioni, o per sesso, o per età; si ripensi anche ad Alessio e Berto dell'esempio.

BOZZA DRAFT

Una malattia infantile risulta mortale in, diciamo, 1 caso su n? L'agiato lettore farebbe bene riflettere con *quali* sventurati fa la media un suo eventuale figliolo. Figli di poveri tossicodipendenti che raccolgono da terra siringhe sporche di sangue, e le utilizzano. Bambini denutriti per povertà, anche estrema, con spaventose carenze nutrizionali. Che poi magari quella statistica è fatta a livello europeo, e comprende *milioni* di Rom in Romania, in condizioni di disagio estremo. O bambini non poverissimi ma comunque non curati a bene: si cerchi in rete, della bimba morta *di fame* a Milano, e i genitori il giorno dopo avrebbero comprato un'automobile... Un'altra, recente (2022), è morta di fame – anche questa a Milano – lasciata sola dalla madre andata in vacanza.[168]

[168] Questo è lo *stato dell'arte* del condurre la vita umana, in certuni: RaiNews (19 settembre 2023) https://www.rainews.it/articoli/2023/09/caso-pifferi-davanti-ai-pm-la-37enne-accusata-di-aver-lasciato-morire-di-stenti-la-figlia-diana-f html A chi le chiedeva dove fosse la piccola quando si allontanava per passare serate o weekend con alcuni uomini "dicevo che Diana era al mare con mia sorella oppure con una mia amica".

Non tutti i bambini hanno genitori che li curano ugualmente bene.

LINK ->"neonata era già in astinenza da cocaina"

E per tre sventurati bambini di cui hanno parlato i media, uno statistico con un po' di esperienza può facilmente ipotizzarne moltissimi altri che non giungono all'onore delle cronache, e figurano semplicemente fra i bimbi morti.

La fortuna è cieca ma la sfortuna ci vede benissimo

(e "sa" ove colpire più probabilmente, e figli di genitori poveri o malvagi o tossicodipendenti possono essere in cima alla lista).

Gli agiati borghesi divoratori di notizie mediatiche che credono che se il morbillo uccide una volta su mille questo riguardo i loro figli non hanno capito la statistica, nè, purtroppo, la vita.

> In Italia[169] 200 mila bambini tra 0 e 5 anni di età vivono in povertà alimentare, ovvero in famiglie che non riescono a garantire almeno un pasto proteico ogni due giorni. Rappresentano l'8,5% del totale (...) Quasi un bambino su 10 della stessa fascia d'età, pari al 9,7%, ha sperimentato la povertà energetica cioè ha vissuto in una casa non adeguatamente riscaldata in inverno. Sono alcuni dei dati inediti relativi al 2023 dell'Atlante dell'Infanzia (a rischio) di Save The Children

BOZZA - DRAFT

50.4 Ma alla fine la probabilità verrà considerata?

Si potrebbe ipotizzare che con una conoscenza esatta dei possibili benefici ovvero danni e delle loro probabilità, l'umanità nel suo complesso e magari pure i singoli, farebbero le scelte giuste, più razionali, tali da massimizzare la speranza matematica del vantaggio.

A parte le difficoltà nel realizzare con tale precisione le ipotesi, anche se sono verificate l'essere umano si comporta ben diversamente. Leggiamo per esempio a proposito del premio Nobel Kahneman e di un suo collaboratore:

[169] https://www.rainews.it/articoli/2024/11/in-italia-200mila-bimbi-tra-0-e-5-anni-in-poverta-alim html da Save The Children

> Attraverso numerosi esperimenti di psicologia cognitiva, infatti, Kahneman e Tversky dimostrarono come le scelte degli esseri umani violassero sistematicamente i principi della razionalità economica[170]

Le persone col pallino della razionalità, poche, non giocano mica alle lotterie, in cui la speranza matematica è negativa ovvero in perdita, ma quasi tutta l'umanità, invece, ci gioca. E similmente avviene praticamente per tutto: le scelte umane sono probabilisticamente irrazionali.

BOZZA - DRAFT

50.5 ...Forse sì, ma dalle intelligenze artificiali

Altra questione è quella dell'incipiente ingresso nel nostro mondo delle intelligenze artificiali: il freddo calcolo probabilistico che necessariamente faranno le automobili a guida automatica è diverso dalle attuali scelte operative umane.
Se

non facendo nulla, un carrello fra poco ucciderà 10 persone con probabilità stimata del 90%

ma facendo una cosa, per esempio azionando una leva, una persona può farne morire 1 sola, altra,
nessuna persona ragionevole farà quella cosa, e lascerà gli eventi al loro corso. Un'intelligenza artificiale invece proprio farà quella cosa, perché la speranza matematica è più favorevole:

nel primo caso, speranza matematica -9 (cioè 9 vite umane in meno)

nel secondo caso, speranza matematica -1. Che è meglio di -9.

(Tale questione, classicamente studiata da anni, viene presentata nel film dell'orrore *Alien: Romulus*, del 2024: una sorta automa causa la morte di alcune persone per salvarne un numero maggiore).

[170] Wikipedia, l'enciclopedia libera, alla voce *Teoria del prospetto*.

Naturalmente ci si prospetta un mondo dove le intelligenze artificiali guideranno non solo le automobili, ma praticamente tutti i sistemi. Compreso il sistema sanitario, facendo eventulamente scelte razionalistiche, umanamente folli, se non controlleremo bene il processo.
Attualmente la Legge ancora distingue il corretto modo di operare adatto agli uomini da quello per le cose e gli animali.
A questo proposito si osservi che sopprimere una persona sana ed espiantarne 5 organi vitali può salvare 5 persone da morte imminente certa al lieve prezzo di 1 sola vita umana: un vantaggio enorme. Una terapia molto efficace, un sistema sanitario efficientissimo. Che è esattamente quello che ogni persona normale farebbe se avesse un allevamento di, poniamo, visoni, rimasto momentaneamente senza mangime: se questo può salvare 5 di ogni 6 capi dell'allevamento, senz'altro daremmo da mangiare un sesto dei visoni agli altri cinque sesti.
Eppure (per fortuna, secondo lo scrivente) con le persone è ancora illegale quell'espianto d'organi. Anzi addirittura non è nemmeno obbligatoria la donazione di rene, pur non letale, al fratello/sorella gemello omozigote, morente, magari con altissima probabilità di successo terapeutico.
Proprio l'attuale covidizzazione della società ha messo in luce questa incipiente zootecnizzazione della società umana.
Questo tipo di argomentazioni basate sulla speranza matematica può portare molto lontano se non verrà fermato.

BOZZA - DRAFT

La leggenda del Sedicente Santo Probabilista

Che un dì mentre mangiava un paninozzo, incontrò languenti a terra una donna poverella, il suo marito poverello, e il loro figlio poverello, tutti ormai allo stremo morenti di fame, che lo supplicarono un pezzo di paninozzo. Ei disse "Nol do, chè io cognosco le cose del Mondo, e la Matematica Speranza, e s'io or per certo vi salvo, voi probabilmente prolificherete un altro pargoletto, e non n'abbian poi a morir di fame quattro infelici invecechè tre. Men vo, scevro da tal colpa & libero da tal responsabilità, leggero come vapor sulfureo."

50.6 Il Calcolo delle Probabilità è controintuitivo

Già abbiamo visto il sorprendente fatto dell'aspettativa di vita residua, che sempre "si crea" davanti al robottino considerato in (42.6): gli restano in media da vivere sempre 72 anni. E sarebbe anche possibile fare regole diverse per ll suo spegnimento, fissandole *prima di avviarlo*, in modo che, se sopravvive al primo anno, gliene aspettano da vivere mediamente perfino più di 71, anche 100 o un milione.

Similmente è sempre uguale il numero di estrazioni del lotto da aspettare mediamente per avere una certa combinazione qualunque, indipendentemente dai "ritardi" tanto seguiti da certuni.
Per esempio sempre 18 estrazioni, per il numero singolo, su una determinata ruota.

Abbiamo anche visto il fenomeno del *cigno nero*.

BOZZA - DRAFT

Paradosso del compleanno

> Il paradosso afferma che la probabilità che almeno due persone in un gruppo compiano gli anni lo stesso giorno è largamente superiore a quanto potrebbe dire l'intuito: infatti già in un gruppo di 23 persone la probabilità è circa 0,51 (51%); con 30 persone essa supera 0,70 (70%), con 50 persone tocca addirittura 0,97 (97%), anche se per arrivare all'evento certo occorre considerare un gruppo di almeno 366 persone (367 se si considera l'anno bisestile).[171]

Il Problema di Monty Hall

- Dietro ciascuna di tre porte c'è un'automobile o una capra (due capre, un'automobile in tutto); la probabilità

[171] Wikipedia, l'enciclopedia libera, alla voce *Paradosso del compleanno*

che l'automobile si trovi dietro una data porta è identica per tutte le porte;

• Il giocatore sceglie una delle porte; il suo contenuto non è rivelato;

• Il conduttore sa ciò che si nasconde dietro ciascuna porta;

• Il conduttore deve aprire una delle porte non selezionate, e deve offrire al giocatore la possibilità di cambiare la sua scelta;

• Il conduttore aprirà sempre una porta che nasconde una capra;

• Cioè, se il giocatore ha scelto una porta che nasconde una capra, il conduttore aprirà la porta che nasconde l'altra capra;

• Se invece il giocatore ha scelto la porta che nasconde l'automobile, il conduttore sceglie a caso una delle due porte rimanenti;

• Il conduttore offre al giocatore la possibilità di reclamare ciò che si trova dietro la porta che ha scelto originalmente, o di cambiare, reclamando ciò che si trova dietro la porta rimasta.

BOZZA - DRAFT

Le possibilità di vittoria aumentano per il giocatore se cambia la propria scelta?
(...) La risposta è sì; le probabilità di trovare l'automobile raddoppiano. [172]

[172] Wikipedia, l'enciclopedia libera, alla voce *Problema di Monty Hall*; che continua: "(...) L'obiezione più comune alla soluzione è fornita dall'idea che, per varie ragioni, il passato possa essere ignorato quando si valutano delle probabilità. Dunque, la scelta della prima porta e il ragionamento del conduttore circa quale porta aprire si possono trascurare; dal momento che si può scegliere tra due porte, la probabilità di scegliere quella giusta dovrebbe essere pari al 50%, indipendentemente che si decida di cambiare o mantenere la porta scelta."

Esistono molti paradossi del Calcolo delle Probabilità.
Alcuni mostrano quanto esso sia controintuitivo, come sopra.
Altri mostrano quanto sia problematica la definizione stessa di probabilità. LINK->

Il seguente esercizio ci mostra invece un fatto sorprendente.

ESERCIZIO$_{\mu 2019}$
$\star \approx \%$ Ogni volta che una fissata persona fa una certa cosa, per esempio va in una zona infestata da insetti pericolosi, si espone ad un certo rischio sanitario, per esempio di subire una puntura d'insetto pericoloso, con probabilità $\frac{1}{n}$, con n un numero che per adesso non specifichiamo. Fissato questo esempio (ma ugualmente si potrebbe ragionare in un caso generale) che probabilità c'è che andando n volte nella zona infestata – supponendo indipendenti gli eventi – quella persona venga punta da qualche insetto pericoloso, per n molto grande, diciamo pure per n tendente all'infinito?

SVOLGIMENTO

BOZZA DRAFT

$$P(\text{punto al 1-esimo viaggio}) = \ldots = P(\text{punto all'}n\text{-esimo viaggio}) = \frac{1}{n}$$

Evento complementare:

$$P(\text{non punto al 1-esimo viaggio}) = \ldots = P(\text{non punto all'}n\text{-esimo viaggio}) = 1-\frac{1}{n}$$

Evento composto:

$$P(\text{mai punto negli } n \text{ viaggi}) = \left(1-\frac{1}{n}\right)\cdot\ldots\cdot\left(1-\frac{1}{n}\right) = \quad (n\,volte)$$

$$= \left(1-\frac{1}{n}\right)^n$$

Evento complementare:

$$P(\text{punto almeno 1 volta negli } n \text{ viaggi}) = 1-\left(1-\frac{1}{n}\right)^n$$

e allora al limite per n tendente all'infinito

$$P(\text{punto almeno 1 volta negli } n \text{ viaggi con } n \text{ grandissimo}) \approx$$

$$\approx \lim_{n\to+\infty} \left(1 - \left(1 - \frac{1}{n}\right)^n\right)$$

e ricordando che

$$\lim_{n\to+\infty} \left(1 - \frac{1}{n}\right)^n = \frac{1}{\mathrm{e}}$$

e allora in conclusionela probabilità cercata vale

$$\boxed{1 - \tfrac{1}{\mathrm{e}} \approx 0.632 = 63.2\%}$$

(Per esempio con $n = 10$ si ha $\left(1 - \frac{1}{10}\right)^{10} \approx 0.651$).
Detto molto grossolanamente: se facendo una cosa si prende una malattia con probabilità $\frac{1}{n}$, facendo quella cosa n volte si prende la malattia con probabilità circa del 63%, supponendo l'indipendenza degli eventi ed n molto grande.
Si noti la sorprendente (quasi) indipendenza dal numero n.

BOZZA DRAFT

Un'analoga (quasi) indipendenza l'abbiamo vista con le dismutazioni.

50.7 Conclusioni

Nel mondo perfetto e pulito di monete e dadi e giochi d'azzardo (com'è il Problema di Monty Hall), si presentano fenomeni controintuitivi, ma almeno chiari.

Alcuni esempi visti ci mostrano come sia sottile la questione dell'esistenza stessa di una probabilità in senso oggettivo e univoco.

Nella realtà del mondo ordinario, storico, complessissimo, vi è come una ricorrente tentazione di applicare verso il futuro le frequenze empiriche rilevate, cioè di dare alla probabilità frequentista un valore predittivo su sistemi non isolati come invece sono i dadi e le

monete, per quanto eventualmente non regolari (se sono regolari la probabilità frequentista non serve, ci basta quella classica).
Questo potrà ancora funzionare abbastanza bene per singoli sistemi biologici (una gravidanza umana probabilmente durerà circa 9 mesi, come risulta abbia sempre fatto), **in particolare sull'effetto dei farmaci** (che tende – almeno in parte – a riprodursi nel tempo nei vari soggetti trattati) ma, nei fatti, diventa largamente illusorio per la Storia umana nel suo complesso. Uno stratega militare del passato disse che "le previsioni a tre mesi valgono zero". (Comunque, in tempo di pace le situazioni sono più stabili che in tempo di guerra).
Venendo a tempi più recenti, il Nobel per l'Economia Paul Samuelson ci ricorda che

> The stock market has forecast nine of the last five recessions[173]

BOZZA DRAFT

Le previsioni sul futuro delle farmacie in Italia e del loro mercato, sono molto aleatorie. Chi l'avrebbe mi detto qualche decina d'anni fa, che entrando in una farmacia si sarebbero visti foulard, profumatori d'ambiente, pupazzetti di Babbo Natale, eccetera!

Definitivamente,

> Prediction is very difficult, especially about the future.

Nonostante ricerche scientifiche che sembravano garantire certe previsioni piuttosto favorevoli, l'umanità ha dovuto, poi, alquanto accontentarsi: e cos'altro poteva fare?

[173] In https://en.wikiquote.org/wiki/Paul_Samuelson

Il Messaggero

Salute | Ricerca MODERNA

Pfizer e Moderna, «chi è vaccinato sarà immune per anni». L'ultimo studio degli scienziati

4 Minuti di Lettura

Mercoledì 30 Giugno 2021, 10:53
Ultimo agg. 1 Luglio, 01:09

Secretary Xavier Becerra
@SecBecerra

An updated COVID vaccine can help protect you from the worst outcomes of COVID. If it's been over 2 months since your last dose, make a plan to get one now.

Find updated vaccines at vaccines.gov.

Menu Azienda Sanitaria dell'Alto Adige

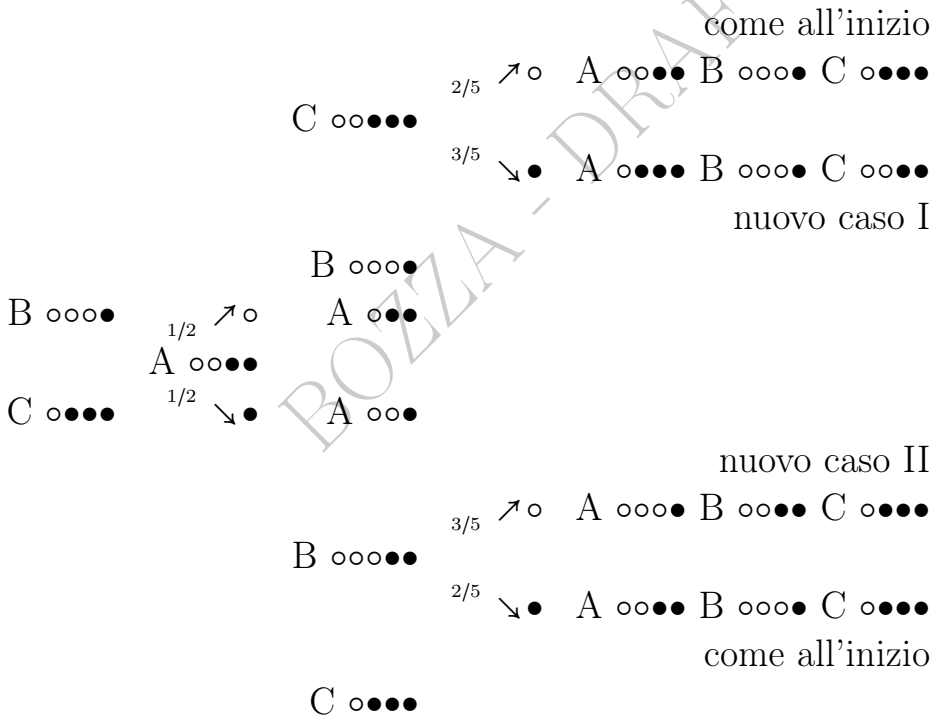
Home | Azienda sanitaria | Strutture | Servizi

Dosi di vaccino somministrate dalla PAB alla popolazione residente: 1.098.772
Prima dose: 399.034 (+5)
Seconda dose: 382.792 (+4)
Terza dose: 274.509 (+91)
Quarta dose: 40.349 (+715)
Quinta dose: 2.087 (+155)
Sesta dose: 1 (+1)

Esercizio risolto$_L$ Tre urne A, B e C contengono palline bianche e nere secondo lo schema

$$A : 2b + 2n \qquad B : 3b + 1n \qquad C : 1b + 3n.$$

Si estrae una pallina da A e si vede il colore:
• se è bianca la si mette in C e poi si estrae una pallina da C e la si mette in A;
• se è nera la si mette in B e poi si estrae una pallina da B e la si mette in A.
Qual è la probabilità di ripristinare la situazione iniziale?
Quali altri casi si possono avere e con quali probabilità?



BOZZA - DRAFT

Ripristino della soluzione iniziale: può avvenire in 2 modi, eventi disgiunti; ciascuno dei 2 addendi è un prodotto perché corrispende ad eventi composti (cioè $\cap$ ovvero *et*) indipendenti; probabilità:
$\frac{1}{2} \cdot \frac{2}{5} + \frac{1}{2} \cdot \frac{2}{5} = \frac{2}{5} = 0.4 = 40\%$.
Nuovi casi e loro probabilità, valendo le osservazioni soprastanti:
I: $(A : 1b + 3n;\ B : 3b + 1n;\ C : 2b + 2n)$, $p' = \frac{3}{10} = 0.3 = 30\%$,
II $(A : 3b + 1n;\ B : 2b + 2n;\ C : 1b + 3n)$, $p'' = \frac{3}{10} = 0.3 = 30\%$.
Esercizio$_\mu$ Costruire e risolvere un esercizio analogo.

B1 – ESERCIZI SUL CALCOLO DELLE PROBABILITà

Esercizio$_\mu$ Che probabilità c'è di ottenere almeno 4 volte il testa con 5 lanci di una moneta regolare?
Esercizio$_\mu$ Che probabilità c'è di ottenere almeno 5 volte il testa con 6 lanci di una moneta regolare?
Esercizio$_\mu$ Che probabilità c'è di ottenere almeno 5 volte il croce con 7 lanci di una moneta regolare?
Esercizio$_\mu$ Che probabilità c'è di ottenere almeno 4 volte il numero 2 con 5 lanci di un dado regolare a 4 facce?
Esercizio$_\mu$ Che probabilità c'è di ottenere almeno 1 volta il numero 3 con 7 lanci di un dado regolare a 4 facce?
Esercizio$_\mu$ Che probabilità c'è di ottenere almeno 4 volte il numero 2 con 5 lanci di un dado regolare?
Esercizio$_\mu$ Che probabilità c'è di ottenere almeno 3 volte il numero 1 con 8 lanci di un dado regolaree?
Esercizio$_\mu$ Che probabilità c'è di ottenere almeno 7 volte il numero 7 con 8 lanci di un dado regolare a 8 facce?
Esercizio$_\mu$ Che probabilità c'è di ottenere almeno 3 volte un numero primo con 8 lanci di un dado regolare?
Esercizio$_\mu$ Che probabilità c'è di ottenere almeno 6 volte un numero primo con 7 lanci di un dado regolare?
Esercizio$_\mu$ Che probabilità c'è di ottenere almeno 6 volte un numero primo con 8 lanci di un dado regolare?
Esercizio$_\mu$ Che probabilità c'è di ottenere un numero primo di volte un numero primo con 6 lanci di un dado regolare?
Esercizio$_\mu$ Che probabilità c'è di ottenere un numero primo di volte un numero primo con 10 lanci di un dado regolare?
Esercizio$_\mu$ Che probabilità c'è di ottenere un numero quadrato di volte un numero primo con 8 lanci di un dado regolare?
Esercizio$_\mu$ Che probabilità c'è di ottenere un numero primo di volte un numero quadrato con 6 lanci di un dado regolare?
Esercizio$_\mu$ Che probabilità c'è di ottenere un numero triangolare di volte testa con 9 lanci di una moneta regolare?

BOZZA DRAFT

Sezione B2 – Statistica Inferenziale

BOZZA - DRAFT

Nota basale sulla Statistica

Una cosa è la Statistica teorica, che tratteremo, fatta sui numeri, che riteniamo veri, e un'altra cosa è la realtà, che con tutti i suoi interessi materiali finisce spesso per inquinare i numeri.

"Oste è buono il vino?" "Buonissimo!"
Un gruppo di amici vuole andare in una delle due enoteche della città e per scegliere quale delle due si affida alla Statistica. Piero afferma che il vino dell'enoteca A è *buono*. L'oste dell'enoteca B afferma che il suo vino è *buonissimo*. Essi pertanto scelgono l'enoteca B, in base alle risultanze della Statistica.

BOZZA - DRAFT

Questa premessa sul **conflitto di interessi** appare vieppiù necessaria oggi, durante la pandemia, in cui la valutazione dell'efficacia e della sicurezza dei vaccini viene affidata principalmente a chi li vende, invece che a enti terzi. (Le varie Agenzie del farmaco non fanno gli studi in doppio cieco). Al 2021, i contratti di acquisto degli Stati – per cifre da capogiro – prevedono che i vari produttori presentino le risultanze degli studi in doppio cieco di lungo periodo entro, chi il 2022, chi il 2023. *I produttori*. (I quali già hanno presentato i loro studi su tempi brevi per ottenere le autorizzazioni in via emergenziale, e i prodotti sono già in uso).

In https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7397951/ si trova la nota sul ritiro di un articolo scientifico sui vantaggi del fumo di sigaretta contro il covid-19, con due autori poi risultati in conflitto di interessi con l'industria del tabacco. (Ma nell'articolo pubblicato – da una prestigiosa rivista – avevano dichiarato di non avere nessun conflitto di interesse).
L'articolo è ancora pubblicato su ResearchGate: "Conflict of Interest statement. None"

Si potrà sperare che siano ben liberi da conflitti di interesse almeno

quelli che lavorano per la FDA (Food and Drug Administration) e le altre Agenzie nazionali e transnazionali, anello intermedio fra i legislatori e i produttori.
Leggiamo sul British Medical Journal:

> "Open Payments reported that [omissis], professor at the University of Michigan School of Public Health and acting chair for the FDA's covid vaccine authorisation meetings, had received over $24 000 (o ((simbolo della sterlina)) 16 970; € 19 650) in payments from drug companies in 2019."
> https://www.bmj.com/content/373/bmj.n1283

Scevri dunque da illusioni (ma sorpresi dalla tolleranza dimostrata da quasi 8 miliardi di persone verso pochi) immergiamoci dunque nello studio dei *numeri* della Statistica, volendoli supporre *veri*, numeri che con la loro purezza sono stati di consolazione ad intere generazioni di matematici, nel corso dei secoli.

BOZZA - DRAFT

Sulla realtà reale invece pesa il

"this is not science, this is business"

citato dall'Editore del British Medical Journal nel 2021. Scrive
https://www.open.online/2021/11/08/covid-19-vaccini-peter-doshi-appell

> Veniamo ora alla frase più controversa di Doshi, che si presta facilmente a manipolazioni. Lui afferma infatti a un certo punto che «questa non è Scienza». Cosa voleva dire esattamente? «È stata estrapolata dal contesto la frase "questa non è Scienza" – spiega Stingi – ma lui in quel caso stava facendo una critica all'aspetto del business. (...)»

Appunto.

> Il dr Szamatolski scrisse: "è mia opinione che la grave condizione della mandibola (di una paziente la cui bocca stava andando in cancrena spontaneamente) è stata causata dall'influenza del radio".

> ...
>
> Era un'idea radicale, tuttavia aveva della scienza a supporto. C'era una considerevole letteratura sui rischi del radio.
>
> ...
>
> Però, l'altro lato della medaglia era tutta la letteratura positiva sul radio.
>
> ...
>
> Ma se si osservavano un po' più attentamente le pubblicazioni positive, c'era un comune denominatore: i ricercatori, tutti, lavoravano per aziende del radio.
>
> ...
>
> L'opinione di Szamatolski, quindi era una voce solitaria e ipotetica contro il ruggito fiammeggiante di una ben finanziata campagna di letteratura pro-radio.
>
> Kate Moore, The Radium Girls, pp 52-53

BOZZA - DRAFT

Il conflitto di interessi non riguarda solo gli Autori degli articoli scientifici, ma anche i decisori degli enti autorizzativi, e perfino i reviewers. Su quest'ultimo punto leggiamo dal prestigioso (1883) Journal of the American Medical Association:

> Between 2020 and 2022, 1155 peer reviewers (58.9%) received at least 1 industry payment[(174)]

[174]Nguyen D, Muramaya A, Nguyen A, et al. Payments by Drug and Medical Device Manufacturers to US Peer Reviewers of Major Medical Journals. JAMA. Published online October 10, 2024. doi:10.1001/jama.2024.17681

X – Stimatori puntuali e intervallari

BOZZA - DRAFT

51 Introduzione alla Statistica Inferenziale

IL PRIMO COMPITO DELLA STATISTICA INFERENZIALE:
DISTINGUERE FRA VARIAZIONI SIGNIFICATIVE E NON SIGNIFICATIVE: PER ESEMPIO PER DISTINGUERE
PLAUSIBILI EFFETTI CASUALI
DA
SPERABILI EFFETTI CAUSALI
DEI FARMACI.

51.1 Citazione da un testo importante

Il presente testo, nonostante le sue centinaia di pagine, è una sorta di "microbo che fluttua in una piastra di Petri".
Il Lettore potrà dover approfondire determinate tematiche, nella sua eventuale futura professione. Esistono moltissimi testi validi.

BOZZA DRAFT

Leggiamo[175] in "Matematica per le scienze della Vita":

> Capire i dati è un processo che richiede diversi passi:
>
> 1. raccogliere i dati;
> 2. riassumere dati;
> 3. analizzare i dati;
> 4. interpretare i risultati e presentarli.
>
> (...) La fase 2 del processo (riassumere i dati) è quella della "statistica descrittiva", in cui l'obiettivo è di astrarre dai dati alcune proprietà per poterli meglio interpretare.
> (...) La fase 3 del processo coinvolge tipicamente l'area della "statistica inferenziale", che consiste nella stima dei parametri e nel testare le ipotesi.

[175] Di Erin N. Bodine, Suzanne Lenhart, Louis J. Gross, a cura di Gabriella Caristi, Maurizio Mozzanica, Giacomo Tommei, 2017, UTET Università, titolo originale Mathematics for the Life Sciences, trad. Patrizia Ferreri

Per quanto riguarda la stima dei parameteri, vedremo gli stimatori, Lezione (52), e gli intervalli di fiducia, Lezioni (53), (54). In Lezioni successive vedremo i test statistici.

51.2 Introduzione

Se abbiamo un dado appena comprato da un rivenditore di cui ci fidiamo, lo riteniamo regolare, cioè riteniamo equiprobabili i suoi risultati, e similmente per una moneta nuova di zecca:

$$X := \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}_{ovvero\, regolare}^{dado\, equilibrato} \qquad Y := \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}_{ovvero\, regolare}^{moneta\, equilibrata}$$

Su queste basi con il Calcolo delle Probabilità possiamo fare moltissime affermazioni certe sui suoi risultati comunque incerti:
– per esempio che conviene scommettere che la somma di 2 lanci del dado sarà 7 piuttosto che 8,
– e conviene scommettere che su un milione di lanci della moneta il numero di teste sarà fra 499 000 e 501 000 piuttosto che no.
Questo è appunto il Calcolo delle Probabilità, e abbiamo visto quante implicazioni ha nella Farmacia.

BOZZA - DRAFT

Tuttavia, gli effetti di un farmaco sono ben lontani da questo tipo di regolarità che hanno il dado e la moneta equilibrati.

Qua entra in campo la Statistica Inferenziale. Mentre nel Calcolo delle Probabilità si parte da una distribuzione e si trovano molte affermazioni conseguenti, come quelle sopra riportate, nella Statistica Inferenziale si suppone noto il tipo di distribuzione, per esempio binomiale o normale o log-normale, ma si ignorano i valori dei parametri della distribuzione, e si vogliono fare affermazioni ragionevoli su essi. Specialmente sulla media e la varianza.

Per esempio potremmo avere un dado di forma irregolare, fatto con un astragalo, o giocare a testa e croce con un frisbee, e ci ritroviamo

con distribuzioni di tipo noto ma con parametri incogniti:

$$X := \begin{pmatrix} 1\,2\,3\,4\,5 & 6 \\ p_1\,p_2\,p_3\,p_4\,p_5 & 1-p_1-p_2-p_3-p_4-p_5 \end{pmatrix}$$

$$Y := \begin{pmatrix} 0 & 1 \\ p & 1-p \end{pmatrix}$$

La statistica inferenziale vuole fare affermazioni ragionevoli sui parametri di distribuzioni per il resto note.

Per esempio, vogliamo affermare ragionevolmente se p è $\frac{1}{2}$ oppure no, che potrebbe corrispondere al farmaco che non fa nè bene nè male, oppure no, qualcosa fa.
Solo ragionevolmente parlando.
La certezza sfugge inesorabilmente alla Statistica Inferenziale.

Tali affermazioni ragionevoli verranno fatte non sulla base della simmetria com'è per dadi e monete, che fa ritenere l'equiprobabilità dei casi possibili, bensì sulla base di molti valori sperimentali ovvero empirici (determinazioni) della variabile aleatoria oggetto di indagine – cioè sulla base di una determinazione (storica, realmente avvenuta nel tempo) di un campione aleatorio. Cioè, lanciamo molte volte l'astragalo o il frisbee, o diamo il farmaco a molti soggetti, e conteggiamo i risultati.

BOZZA DRAFT

Così spiega Wikipedia, l'enciclopedia libera, alla voce *inferenza statistica*:

> Si considereranno principalmente campioni casuali semplici di dimensione $n > 1$, che possono venire interpretati come n realizzazioni indipendenti di un esperimento di base, nelle medesime condizioni. Dal momento che si considera un esperimento casuale, si coinvolge il calcolo delle probabilità. Nell'inferenza statistica c'è, in un certo senso, un rovesciamento di punto di vista rispetto al calcolo delle probabilità. Nell'ambito di quest'ultimo, noto il processo di generazione dei dati sperimentali (modello probabilistico) siamo in grado di valutare la probabilità dei diversi possibili risultati di un esperimento. Nella statistica il processo di generazione dei dati sperimentali non è noto in modo completo (il processo in questione è, in definitiva, l'oggetto di indagine) e le tecniche statistiche si prefiggono di indurre le caratteristiche di tale processo sulla base dell'osservazione dei dati sperimentali da esso generati.

Dovremo però cercare di ragionevolmente distinguere il plausibile effetto casuale, dallo sperabile effetto causale.
Per esempio è ben possibile che una moneta o frisbee pur regolare abbia fatto 3 teste su 4 lanci, ma se dà testa 300 volte su 400, sarà ragionevole ritenere che no, non sia regolare.
Fra il plausibile effetto casuale e lo sperabile effetto causale una certa linea di confine viene demarcata da formule basate sulle funzioni del chi quadrato, della t di student, della gaussiana, e altre.
La linea di confine resta sottile e problematica.

Così ancora spiega Wikipedia, l'enciclopedia libera, alla stessa voce *inferenza statistica*:

> Nell'ambito dell'inferenza statistica, si distinguono due scuole di pensiero, legate a diverse concezioni, o interpretazioni, del significato della probabilità:
>
> • Inferenza classica, o frequentista;
>
> • Inferenza bayesiana.
>
> La prima è legata agli storici contributi di R. Fisher, K. Pearson, e rappresenta la posizione maggioritaria. La seconda, allo stato attuale (2005) ancora minoritaria, ma in crescita, è fondata sull'uso del risultato del teorema di Bayes ai fini dell'inferenza statistica.

BOZZA - DRAFT

Questa trattazione elementare segue il primo modello.

51.3 Primo esempio completo di Statistica Inferenziale

La probabilità che il contatore X_n di teste in n lanci di moneta regolare assuma un valore

$$\frac{n}{2} - \sqrt{n} \leq X_n \leq \frac{n}{2} + \sqrt{n}$$

(che con la Disuguaglianza di Chebyshev si può dimostrare esser $> 75\%$) con calcoli non banali si può dimostrar essere $> 95.4\%$ per ogni $n \geq 4$, cioè

$$P\left(\left|X_n - \frac{n}{2}\right| \leq \sqrt{n}\right) > 95.4\% \qquad \forall\, n > 4. \tag{142}$$

Questo ci permette di fare una prima puntatina nella statistica della Farmacia.

Supponiamo di avere una popolazione di monete regolari, e consideriamo malattia la loro regolarità, cioè il fare testa con probabilità $\frac{1}{2}$.
Ne prendiamo un bel campione per fare un test terapeutico, diciamo 10 000. (Il numero di soggetti ragionevole da considerare è questione statistica non banale; qua scegliamo 10 000).
Le curiamo con una goccia di farmaco (per esempio una goccia di colla al centro dalla parte della testa) e ci chiediamo se la cura ha funzionato, cioè se adesso sono meno regolari. Cioè ci chiediamo se le abbiamo curate dalla loro regolarità.
Lanciamo le 10 000 monete 1 volta ciascuna ottenendo 10 000 risultati indipendenti e diciamo $X_{10\,000}$ il numero di teste.
Noi vorremmo che questo numero sia molto diverso 5 000, cioè la metà di 10 000.
Supponiamo di ottenere 5 100: possiamo essere soddisfatti?
(5100 è proprio $n + \sqrt{n}$).
Sì perché se le monete erano ancora regolari c'era meno del 4.6% di probabilità (100% – 95.4%) che venisse un risultato così estremo, lontano dalla metà. (Siamo soddisfatti ma di poco, se veniva 5 200 o magari 6 000 era molto meglio).
La soglia ordinaria di *significatività* della Statistica è il 5%.
Se ne otteniamo 5 050 non siamo soddisfatti: una tale deviazione è ben compatibile con la regolarità.
Questa è solo un'idea iniziale: nella pratica scientifica seria, bisogna fissare la soglia prima di fare l'esperimento.

BOZZA - DRAFT

51.4 Secondo esempio di Statistica Inferenziale

Durante una guerra viene costruito per 20mila prigionieri un campo di detenzione che ha due ali con ugual numero di detenuti: in una i detenuti bevono l'acqua di un pozzo, e nell'altra ala l'acqua di

un altro pozzo. Non c'è nessun problema etico e nessuna malevolenza nel dare l'acqua dei due diversi pozzi nelle due ali; questo è quello che si chiama un *esperimento naturale*, che si forma da sè. Ad un certo punto sono ormai morti metà dei prigionieri, e nello stesso giorno gli altri vengono liberati e un tecnico del personale analizza l'acqua dei due pozzi e scopre che in uno c'è molto nickel, o qualunque altra sostanza, ma diciamo nickel per fissare le idee, e nell'altro non c'è (in nessuna misura rilevabile). A questo punto il medico del campo ipotizza che il nickel possa influire significativamente sulla sopravvivenza ovvero la mortalità, seppure non sa se favorevolmente o sfavorevolmente. Scopre che 5050 sono morti in un'ala e 4950 nell'altra, e uno potrebbe dire che sì, c'è un forte indizio che il nickel abbia influenzato la sopravvivenza ovvero la mortalità. Questa sarebbe stata la logica conclusione di uno studioso dei secoli passati. Invece la moderna Statistica Inferenziale dice che no, il risultato non è *statisticamente significativo*. Lo scarto di 50 da 5000 è troppo piccolo. Cè una probabilità circa del 5% che il valore cada fuori di $[4900, 5100]$, e ben maggiore del 5% che cada fuori di $4951, 5049$, com'è successo, anche se è $\frac{1}{2}$ la probabilità che ogni singolo deceduto provenisse da una delle due ali del campo.
La moderna Statistica Medica viene fatta in generale proprio così, "al 95%" (di *confidenza*, come si dirà).

BOZZA DRAFT

Nota sui due esempi. Si noti però che in generale nella Farmacia in generale con un farmaco si vuole sbilanciare un parametro in una fissata direzione, (si pensi alla glicemia troppo alta o alla sideremia troppo bassa), non semplicemente smuoverlo in una direzione qualunque.
Questo complicherà le cose. Per intanto abbiamo capito essenzialmente cosa vuol dire distinguere variazioni significative da variazioni non significative: queste ultime *potrebbero ben essere* (qui sta il 95%) *fluttuazioni casuali*.

Con l'approssimazione normale si può dimostrare che, purché n

sia sufficientemente grande, lo sbilanciamento di teste rispetto alla metà in una direzione o nell'altra

$$\left| X_n - \frac{n}{2} \right|$$

con probabilità $\approx 68.3\%$ è $< \frac{1}{2}\sqrt{n}$,
con probabilità $\approx 27\%$ sta fra $\frac{1}{2}\sqrt{n}$ e $\sqrt{n}$,
con probabilità $\approx 4.6\%$ è $> \sqrt{n}$

Riscriviamo l'ultima:

$$P\left(\left| X_n - \frac{n}{2} \right| > \sqrt{n}\right) \approx 4.6\% \qquad \forall\, n \textit{ molto grande} \tag{143}$$

strettamente imparentata alla (142) perché 4.6 è 100-95.4.

BOZZA DRAFT

51.5 Riconoscere statisticamente una mancanza di neonate

Se si ipotizza – questione molto incerta e problematica per vari motivi – che la probabilità del genere maschile e del genere femminile alla nascita sia 50%, uno scostamento eccessivo dalla metà nei nati fa sospettare l'aborto selettivo – a livello di popolazione ovviamente, non di singoli casi: non possiamo entrare nella testa della gente e sapere se ha abortito con riguardo al genere oppure per altri motivi.
Sulla triste questione si può cercare in rete *aborto selettivo*. Ma nella pratica, non si può distinguerlo facilmente da un altro triste fenomeno, collegato ma diverso, l'infanticidio delle neonate (che, in Italia e altri stati, è illegale).

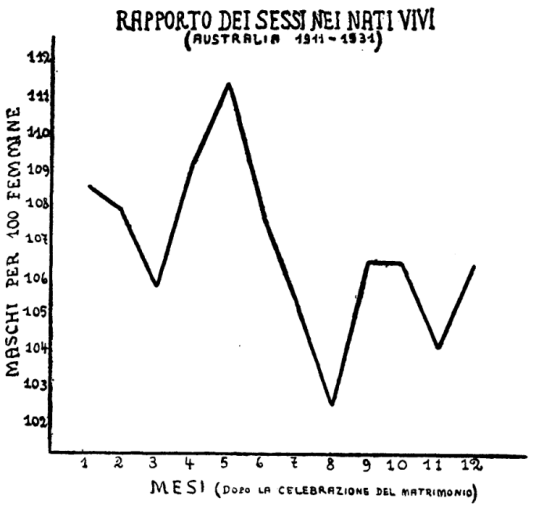
In pratica, detto semplificatamente, nell'ipotesi di equiprobabilità di nascite maschili e femminili, "al 95%" su un milione di neonati il numero di femmine dovrebbe stare fra 499mila e 501mila, ma se è minore di 499mila sospettiamo "al 95%" l'aborto selettivo (oppure, indistinguibilmente, l'infanticidio delle neonate), e con ancor maggiore *livello di confidenza* se è minore di 498mila, e così via.

Tuttavia la questione è molto delicata, perchè se per qualche motivo sconosciuto in una popolazione le femmine tendono per natura – senza aborti – a nascere con probabilità 49,9% invece che 50%, ecco che il valore 499mila diventerebbe del tutto normale, proprio il valore atteso. E dimostrare che in una determinata popolazione non è 49.9% la probabilità naturale, è, come si capisce, ben difficile.

Nella Figura un accenno (relativo all'Australia) da un lontano passato (evitando così di aggiungere alla triste situazione pure stigma a popolazioni attuali). Per la *natimortalità* citata si intenda, in questo caso, sostanzialmente *infanticidio*.

BOZZA - DRAFT

Esistono altre pubblicazioni a me sfuggite? È stato altre volte il fenomeno rilevato dai demografi? È stato spiegato come errore di rilevazione o come fenomeno reale? Per l'Italia sarebbe facile un riscontro. L'Istituto centrale di statistica potrebbe rilevare la notizia che ci interessa senza difficoltà, dato ch'esso raccoglie e pubblica ora le notizie riguardanti le nascite distinte per mesi successivi alla celebrazione del matrimonio, facendo però una sola classe dei nati prima di otto mesi, senza tenere separati i due sessi.



Ammesso che il fenomeno si potesse constatare anche per altri paesi, occorrerebbe poi cercarne le cause. Crederei che esse potessero essere ricondotte a differenze nella abortività e nella natimortalità tra i concepiti prima e i concepiti dopo la celebrazione del matrimonio, differenze che troverebbero la loro ragione di essere in altre cause di carattere sociale. È però inutile il formulare delle ipotesi prima che ulteriori dati consentano di poter con sicurezza affermare che il fenomeno si riscontra in tutti o almeno in molti paesi.

Torino, R. Istituto superiore di scienze economiche e commerciali.

DIEGO DE CASTRO.

Figure 59: Da: *Il rapporto nei sessi nei nati da concepimento antenuziale*, Diego De Castro; Giornale degli Economisti e Rivista di Statistica, Serie quarta, Vol. 73 (Anno 48), No. 10 (ottobre 1933), p. 735. Online in https://www.jstor.org/stable/23229506

Nota. In realtà dovremmo fare un *test statistico* un po' diverso, e più complicato, perchè vogliamo dimostrare lo sbilanciamento in una direzione specifica. Tuttavia, nella pratica Biomedica spesso si fa comunque il *test bilatero* anche se il *test unilatero* sarebbe più appropriato, perchè quello bilatero è più "standard".

Nota. Di questa Lezione ci si dovrebbe aspettare che lo studente conosca a memoria la formula (143), oltre a tutto l'inquadramento teorico della questione.

52 Stimatori e stimatori non distorti

52.1 Parametri e stimatori; stimatori non distorti

Supponiamo che fra poco avremo le altezze di $n := 100$ persone, prese a caso dai 60milioni di italiani, e allora l'altezza di un italiano a caso è una variabile aleatoria X, e gli n numeri che stiamo per avere sono essi stessi variabili aleatorie X_1,...,X_n, e dopo che li avremo saranno numeri $x_1, ..., x_n$, detti *determinazioni* della variabile aleatoria. Chiameremo anche *campione aleatorio* le variabili aleatorie X_1,...,X_n e *campione* i numeri $x_1, ..., x_n$, ma non ci formalizzeremo su questa distinzione.

BOZZA - DRAFT

Ottenere da n numeri una *stima* di un *parametro* incognito di una v.a. ovvero della sua legge, è un *problema di stima*. Ogni funzione

$$h(X_1, ..., X_n)$$

(per esempio $\max(X_1, .., X_n)$) di un campione aleatorio si dice *stimatore*, in generale stimatore di un parametro incognito u della densità della variabile aleatoria, variabile aleatoria che sappiamo esistere ma non conosceremo mai in forma esatta (legge ovvero distribuzione, cioè densità o funzione di ripartizione), ma di cui avremo n determinazioni. Scriveremo

$$\hat{v} := h(X_1, ..., X_n)$$

e diremo che $\hat{v} = h(X_1, ..., X_n)$ (variabile aleatoria) è uno stimatore di v e $\hat{v} = h(x_1, ..., x_n)$ (numero) è una stima di v.

Rimarchiamo che lo stimatore è una variabile aleatoria, funzione di un numero indeterminato n di variabili aleatorie. Certo, dopo che delle variabili aleatorie del campione avremo una determinazione, immediatamente avremo una determinazione dello stimatore, che a quel punto sarà un numero.

Esempio duplice di stimatore. Per esempio per una v.a. $\mathbb{U}[0,a]$ o $\mathbb{U}\{0,a\}$ (uniforme fra 0 e a, continua o rispettivamente discreta)

$$\hat{a} := 2\,\frac{X_1 + ... + X_n}{n} \qquad = 2\,\bar{X}_n \tag{144}$$

è lo *stimatore dei momenti* del parametro a. (Il doppio della media aritmetica).

Per esempio se 3 persone si sono iscritte indipendentemente a una corsa con molti partecipanti e hanno avuto i numeri progressivi n_1, n_2, n_3, il doppio della media – in assenza di altre informazioni – è una stima del numero di iscritti. (Approssimativamente: in effetti i 3 numeri si possono ritenere determinazioni di una v.a. $\mathbb{U}\{1,a\}$ e non $\mathbb{U}\{0,a\}$, e inoltre non c'è una perfetta indipendenza perchè noi sappiamo che $n_2 \neq n_1$, però le 2 imperfezioni non pesano molto perchè ci sono molti partecipanti).
Per esempio per

Quello della formula data non è l'unico stimatore per a, ne esistono infiniti, ma ben pochi sensati. Un altro considerato molto sensato per $\mathbb{U}[0,a]$ è il massimo dei valori del campione, $\hat{a} := \max(X_1, ..., X_n)$, di cui non ci occuperemo.

BOZZA - DRAFT

ESERCIZIO$_{\mu}$ $\approx$
Dato questo campione aleatorio

$$0.5288, 0.0344, 0.6112, 0.072, 0.4584, 0.5104, 0.7296$$

di una variabile aleatoria Z uniformemente distribuita sull'intervallo $[0,a]$, stimare a con lo stimatore dei momenti, con 4 cifre decimali.

SVOLGIMENTO
Viene usato lo standard del punto decimale.

Lo stimatore dei momenti è il doppio della media aritmetica.
(Ovvero in formule, per una v.a. Z, $\hat{a} := 2\,\frac{Z_1+...+Z_n}{n} = 2\,\bar{Z}_n$).

La media degli $n = 7$ numeri è

$$\bar{Z}_7 = \frac{0.5288 + 0.0344 + 0.6112 + 0.072 + 0.4584 + 0.5104 + 0.7296}{7} \approx 0.420686$$

e moltiplicando per 2, si ha, con 4 cifre decimali,

$$\hat{a} = 2\,\bar{Z}_7 =$$

$$\boxed{\hat{a} \approx 0.8414}$$

Nota. (Il campione era stato ottenuto con $a = 0.8$ con l'istruzione

```
v = Table[0.8 RandomInteger[1000]/1000., i, 1, 7]
```

col software Mathematica$^{(R)}$).

Ulteriore esempio. Per una v.a. uniforme $\mathbb{U}[-u, u]$ lo *stimatore dei momenti* di u è

$$\hat{u} := \sqrt{\frac{3\,(X_1^2 + ... + X_n^2)}{n}}$$

BOZZA - DRAFT

52.2 Stimatori non distorti

Esistono molti criteri di bontà di uno stimatore, e noi ne considereremo uno: diremo che lo stimatore $\hat{u}$ del parametro incognito u è *non distorto* (o *corretto*; in inglese *unbiased*) se la speranza matematica di $\hat{u}$ è u, cioè $E(\hat{u}) = u$.

52.3 Stimatori non distorti di media e varianza

La differenza basale fra la *statistica descrittiva* e la *statistica inferenziale* è che la prima opera (in generale) su numeri e la seconda su variabili aleatorie, con il che la prima ricade in quella che abbiamo chiamato *matematica della certezza* e la seconda nella *matematica dell'incertezza.*
Se abbiamo n numeri

$$x_1, x_2, ..., x_n$$

nella statistica descrittiva queste sono la media e la varianza degli n numeri:

$$\bar{x} := \frac{1}{n}\sum_{k=1}^{n} x_k \qquad Var(x_1, x_2, ..., x_n) := \frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^2 \,.$$

Se invece supponiamo che $x_1, x_2, ..., x_n$ siano *determinazioni* di una variabile aleatoria X con una certa media μ e una certa varianza σ^2 *incognite*, dalle precedenti formule possiamo immediatamente definire questi (ragionevoli) stimatori – che sono variabili aleatorie – di μ e σ^2:

$$\hat{\mu} := \bar{X} := \bar{X}_n := \frac{1}{n}\sum_{k=1}^{n} X_k \tag{145}$$

$$W := \frac{1}{n}\sum_{k=1}^{n}(X_k - \bar{X}_n)^2 .$$

Si dimostra che il primo è stimatore *non distorto* della media μ, cioè $E(\bar{X}_n) = \mu$, ma il secondo non è stimatore non distorto di σ^2, cioè la sua speranza matematica è diversa dal parametro che si vuol stimare.
Si dimostra invece che (e si noti la differenza nel denominatore)

$$S^2 := \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2 \ \left[:= \hat{\sigma}^2 \ {}_{\leftarrow \textit{scrittura rara, scriveremo } S^2} \right] \tag{146}$$

è lo *stimatore non distorto della varianza*: $E(S^2) = \sigma^2$.

In Farmacia e Medicina. Per esempio dalle altezze degli studenti in aula si potrebbe ottenere una stima dell'altezza media di popolazioni più ampie – o con la sideremia, o quant'altro. Abbastanza bene, per esempio, dell'altezza delle ventenni triestine, o perfino italiane, meno bene dei ventenni italiani (maschi e femmine), ancora meno bene di tutti i ventenni del mondo, o di tutti gli esseri umani.
Ma di quale popolazione un campione è bene rappresentativo è problema drammatico e non ben risolubile di tutta la Farmacia e la Medicina. Le statistiche mediche spesso si fanno presso 1 solo

BOZZA - DRAFT

ospedale, o 2-3 ospedali, con campioni che non bene rappresentano tutta l'umanità, per vari motivi: socio-economici, etnici, eccetera. Poi, certo, gli Autori degli articoli scientifici se la giustamente cavano con frasi standard come "further studies are required" (20 milioni di risultati di ricerca Google nel settembre 2024), e magari "with larger samples".

52.4 Cenno agli stimatori di massima verosimiglianza

Lo stimatore $\hat{a}$ *di massima verosimiglianza* "cerca di essere" quello che renderebbe massimamente probabile l'uscita di x_1,...,x_n, se a valesse $\hat{a}$. Il metodo per trovarlo si basa sulle derivate e non è semplicissimo.
Ci limitiamo a dare la formula dello stimatore di massima verosimiglianza del parametro λ della legge esponenziale:

$$\hat{\lambda} = \frac{n}{X_1 + ... + X_n} = \frac{1}{\bar{X}_n} \,. \tag{147}$$

(Che non è non distorto).

BOZZA - DRAFT

ESERCIZIO$_{\mu}$ $\approx$
Stimare il parametro λ di una densità esponenziale da questo campione:

16.62, 3.810, 35.97, 4.322, 2.725, 11.44, 0.6671, 14.85, 3.816, 12.54.

Si dia il risultato con 2 cifre significative.

SVOLGIMENTO
Viene usato lo standard del punto decimale.
La media aritmetica degli $n = 10$ valori è $\bar{X}_n = 10.676$.
Lo stimatore di massima verosimiglianza è il reciproco della media

$$\hat{\lambda} = \frac{1}{\bar{X}_n} \approx$$

con 2 cifre significative

$$\boxed{\approx 0.094.}$$

BOZZA - DRAFT

Nota. Salvo approssimare i valori a 4 cifre significative, il campione era stato ottenuto su WolframAlpha con parametro 0.1, abbastanza vicino allo 0.094 trovato con lo stimatore, con l'istruzione

```
10 random numbers exponential distribution lambda=0.1
```

che naturalmente, se richiamata da qua, in generale darà altri 10 numeri, sempre provenienti da quella variabile aleatoria esponenziale simulata, con vero parametro $\lambda = 0.1$. Di volta in volta lo stimatore darebbe diversi valori per $\hat{\lambda}$, plausibilmente "vicini" al vero $\lambda = 0.1$.

Nota. Vale proprio la pena trarre più volte un campione di 10 elementi con questa distribuzione, esponenziale di parametro 0.1, su WolframAlpha:

```
10 random numbers exponential distribution lambda=0.1
```

e osservare la loro natura. Li si pensi come intertempi, in minuti, fra ingressi in Farmacia, o al Pronto Soccorso. Alle volte si aspetta molto, alle volte pochissimo. (E si pensi alle possibili conseguenze in Pronto Soccorso). I Lettori che hanno fatto barista hanno esperienza pratica di questi fatti.

52.5 Riassumiamo: gli stimatori puntuali

Prima di tutto, in questo paragrafo chiameremo stimatori puntuali quelli che finora abbiamo chiamato semplicemente stimatori; esistono poi gli stimatori intervallari, che sono coppie di variabili aleatorie da intendere come estremi di un intervallo, che vedremo.

Per una v.a. discreta o continua X, supponiamo di sapere che:
1) ha distribuzione di un certo tipo noto dipendente da un parametro incognito, sia esso a, per esempio $\mathbb{U}[-b, b]$ o $N(\mu, 1)$ o $N(0, \sigma^2)$ o $\Gamma(\alpha, 1)$ (dove a è rispettivamente b, μ, σ^2 e α);
2) avremo n determinazioni indipendenti x_1,...,x_n di X, che per ora, "prima" di averle, sono n variabili aleatorie X_1,...X_n.

Non sapremo mai quanto vale a ma vogliamo stimarlo con uno *stimatore puntuale* $\hat{a}$ mediante i dati x_1,...,x_n, ovvero (prima di averli) X_1,...,X_n.

BOZZA - DRAFT

Concretamente lo stimatore ci appare come una funzione di un numero arbitrario n di variabili, come $2\,\frac{X_1+...+X_n}{n}$ ovvero $2\,\frac{x_1+...+x_n}{n}$.

Ai fini delle Scienze Applicate, queste formule si trovano semplicemente elencate, sui vari testi, come questo, mentre produrle non è facilissimo. I matematici e in particolare gli statistici le producono, dimostrando le loro varie qualità positive, per esempio l'essere non distorti e altre.

A un livello superiore si considerano parametri 2-dimensionali, come $a := (\mu, \sigma^2)$.

La teoria degli stimatori dei momenti è ricca e interessante e qualcosa si trova nei Complementi a questa Lezione ma nella nostra trattazione elementare ci limitiamo molto.

52.6 Nota dolente finale e prospettive future

La probabilità che lo stimatore dia il valore esatto del parametro può, genericamente e imprecisamente parlando, ritenersi bassa per variabili aleatorie discrete, e nulla per variabili aleatorie continue. Eppure, il valore che danno è molto meglio che niente, e si pensi semplicemente alla media. Questo è dovuto essenzialmente al fatto che anche se non danno il valore esatto del parametro da stimare, è "grande" la probabilità che l'errore sia "piccolo", seppure qua non approfondiamo la questione.
Resta il fatto che quel singolo numero non contiene in sè l'informazione sulla sua plausibile precisione. Questo limite verrà superato dai molto più moderni stimatori intervallari: invece di dare un singolo valore, daranno un intervallo all'interno del quale "sperabilmente" cade il valore da stimare; e ci sarà una sorta di misura per quel livello di sperabilità, il *livello di confidenza*.

Nota. Di questa Lezione ci si dovrebbe aspettare che lo studente conosca a memoria le formule (144), (145), (146), (147), oltre a tutto l'inquadramento teorico della questione.

Naturalmente bisogna conoscere anche le formule di media e varianza, riportate in questa Lezione, che sono state già illustrate in Statistica Descrittiva.

BOZZA - DRAFT

* Complementi *

52.7 Complementi – gli stimatori da un libro classico

Diamo qualche dettaglio sugli stimatori seguendo abbastanza da vicino un libro che si distingue nella didattica del Calcolo delle Probabilità e della Statistica: Calcolo delle probabilità, di Paolo Baldi, ed. McGraw-Hill.

BOZZA - DRAFT

Stimatore dei momenti col momento primo.
Lo stimatore *â dei momenti, col momento primo*, attribuisce al parametro a quel valore che farebbe avere alla densità $f_a(x)$ come media, ovvero speranza matematica ovvero momento primo, il valore che è proprio la media di X_1,...X_n.

Esempio 1. Per una densità $N(\mu, \sigma^2)$ con σ^2 noto (ora $a := \mu$)

$$\hat{\mu} := \frac{1}{n}(X_1 + ... + X_n).$$

Esempio 2. Per una densità esponenziale di parametro λ (ora $a := \lambda$) è

$$E(X) = \frac{1}{\lambda} =^{EQ} \bar{X}_n$$

$$\Rightarrow \hat{\lambda} := \frac{1}{\bar{X}_n} = \frac{n}{X_1 + ... + X_n}\,.$$

Esempio 3. Per una densità $\Gamma(\alpha, \lambda)$ con α noto (ora $a := \lambda$) è

$$E(X) = \frac{\alpha}{\lambda} =^{EQ} \bar{X}_n$$

$$\Rightarrow \hat{\lambda} := \frac{\alpha}{\bar{X}_n} = \frac{n\,\alpha}{X_1 + ... + X_n}\,.$$

Esempio 4. Per una densità uniforme $\mathbb{U}[-u, u]$ è

$$E(X) = \frac{u - u}{2} = 0$$

che non dipende dal parametro incognito $a = u$: il metodo dei momenti col solo momento primo non si può applicare.

Metodo di calcolo dello stimatore dei momenti 1) Si calcola la speranza matematica di $f_a(x)$ come funzione di a;
2) se dipende da a la si uguaglia a $\bar{X}_n$, media di $X_1,...,X_n$;
3) si risolve (se possibile) l'equazione in a;
Quella soluzione trovata, in termini di $X_1,...,X_n$ è lo stimatore dei momenti, inteso come funzione (ed è una variabile aleatoria), per esempio

$$\hat{a} := \frac{n}{X_1 + ... + X_n}$$

e in termini di $x_1,...,x_n$ è lo stimatore dei momenti inteso come valore dello stimatore, nel caso specifico di un campione $x_1,...,x_n$ considerato, per esempio

$$\hat{a} := \frac{n}{x_1 + ... + x_n}$$

e quando a $x_1,...,x_n$ si sostituiscano i loro valori, è un numero, per esempio

$$\hat{a} := \frac{3}{8+7+5} = \frac{3}{20} = 0.15.$$

4) Se $E(X)$ non dipende dal parametro incognito a, si usi allora il momento secondo uguagliandolo alla media quadratica.

BOZZA - DRAFT

Esempio 4 bis. Per una densità uniforme $\mathbb{U}[-u, u]$ è

$$E(X) = \frac{u-u}{2} = 0$$

che non dipende dal parametro incognito $a = u$. Con il momento secondo $\int_{-\infty}^{+\infty} x^2 f_a(x)\,dx$, ora che la densità è $\frac{1}{2u}$ fra $-u$ e u,

$$E(X^2) = \int_{-u}^{u} \frac{x^2}{2u}\,dx = \Big[\frac{x^3}{6u}\Big]_{-u}^{u} = \frac{u^3}{6u} - \frac{-u^3}{6u} = \frac{u^3}{3u} =$$

$$= \frac{u^2}{3} =^{EQ} \frac{1}{n}(X_1^2 + ... + X_n^2) \quad \leftarrow \textit{media quadratica}$$

$$\rightarrow \quad \hat{u} := \sqrt{\frac{3\left(X_1^2 + ... + X_n^2\right)}{n}} \, .$$

BOZZA - DRAFT

53 Intervalli di fiducia; media di $N(\mu, \sigma^2)$

53.1 Introduzione agli intervalli di fiducia

Se X è una variabile aleatoria normale $N(\mu, \sigma^2)$ con μ incognito, e ne avremo un campione aleatorio $X_1,...,X_n$ (variabili aleatorie indipendenti tutte con la stessa legge $N(\mu, \sigma^2)$) oppure ne abbiamo delle determinazioni (numeri) $x_1,...,x_n$, possiamo stimare μ con la media campionaria $\hat{\mu} := \bar{X}_n$ e allora numericamente $\hat{\mu} = \bar{x}_n$. Potrebbe essere per esempio $\bar{x}_n \approx 2.019$ sia se $n = 1$ sia se $n = 100$: la *stima puntuale* ottenuta non distingue i due casi mentre è chiaro che nel secondo caso il valore 2.019 è molto più "sicuro" se non nella sua esattezza (che comunque ha probabilità 0) almeno nella sua vicinanza al valore vero.

Questo della media aritmetica è il modo antico di fare nella Statistica, anche Medica.
Invece il modo moderno è quello di dare un intervallo di valori in cui *sperabilmente* sta il valore vero, come ora mostreremo.

BOZZA - DRAFT

La *stima intervallare* è costituita da 2 stimatori $\hat{u}$ e $\hat{v}$ tali che il parametro incognito, sia ora esso a (per esempio μ di $N(\mu, \sigma^2)$) sia nell'*intervallo aleatorio* $[\hat{u}, \hat{v}]$ con probabilità $\geq 95\%$

$$P(a \in [\hat{u}, \hat{v}]) \geq 95\% \qquad \text{e secondo altri Autori } =$$

fino a che $\hat{u}$ e $\hat{v}$ sono variabili aleatorie dipendenti da $X_1,...,X_n$, cioè prima di venire determinate coi dati numerici $x_1,...,x_n$: dopo, il parametro a o sta o non sta nell'intervallo determinato, e sarebbe arduo anche solo definire in che senso si potrebbe attribuire una probabilità a quel fatto; si potrebbe farlo solo con la concezione soggettiva della probabilità ma con essa si può proporre *qualunque* valore di probabilità. (È un po' come chiedersi che probabilità ha 2019 di essere primo: o è primo o non lo è, non c'è una grande questione probabilistica). La questione è sottile e di fatto diffusissima è la credenza erronea che a stia nell'intervallo di confidenza

con probabilità 95%, o più in generale $(1-\alpha)100\%$. L'articolo scientifico riportato in questo link-> su sito governativo statunitense (associato al PubMed)

> provide an explanatory list of 25 misinterpretations of P values, confidence intervals, and power.

Fissati i valori $x_1,...,x_n$ e conseguentemente i valori di $\hat{u}$ e $\hat{v}$ non si parla più di probabilità ma di *(livello di) confidenza*, p.es. 95%.

Scriveremo per esempio

$$a \in [1.9, 2.1] \quad \alpha = 0.05$$

ma altri Autori scriveranno diversamente e variamente, per esempio

$$\text{C.I.}_{95} = 1.9 - 2.1$$

dove C.I. sta per *confidence interval* e 95 sta per 95% ovvero $\alpha = 0.05$.

BOZZA - DRAFT

Tutto questo si estende mutando la soglia 95% in qualunque altro *livello* $1-\alpha$, e normalmente si usano anche 90% e 99%, e si estende a qualunque parametro incognito di una densità per il resto nota.

Leggiamo in un articolo scientifico[176]

> Although the 95% CI is most often used in biomedical research, a CI can be calculated for any level of confidence. A 99% CI will be wider than 95% CI for the same sample.

Il complemento α di $1-\alpha$ si chiama *(livello di) significatività*, e in generale è 5% = 0.05 o 10% = 0.1 o 1% = 0.01.
Purtroppo α e $1-\alpha$ vengono scambiati fra loro nei vari testi.

[176] Hazra A. Using the confidence interval confidently. J Thorac Dis. 2017 Oct;9(10):4125-4130. doi: 10.21037/jtd.2017.09.14. PMID: 29268424; PMCID: PMC5723800.

Il *(livello di) di significatività* è affine al *p value* dei test statistici, che vedremo in seguito, salvo che quest'ultimo non è prefissato ma si calcola dopo: è l'ultimo valore di significatività possibile coi dati che abbiamo, il valore discriminante. Se avessimo pre-fissato un α più piccolo non saremmo riusciti a dimostrare niente, neppure con l'incerta certezza statistica. La speranza è avere un p value piccolissimo, indice di una forte certezza statistica.

53.2 Intervalli di fiducia per μ per campioni gaussiani

Precisiamo dapprima che *intervallo di fiducia* e *intervallo di confidenza* sono sinonimi.

In questo paragrafo considereremo, per una $N(\mu, \sigma^2)$:
• la stima intervallare di μ essendo noto σ^2;
• la stima intervallare di μ essendo ignota anche σ^2.

Tutte le formule di questo paragrafo valgono per campioni gaussiani e (approssimatamente) anche per campioni non gaussiani se la numerosità è $n \geq 30$ e la densità non è "troppo" asimmetrica.

Per μ con σ^2 nota, intervallo di fiducia bilatero centrato al livello (di confidenza) $1-\alpha$, dove α è "piccolo", come $0.05 = 5\%$ (da non confondersi con $1-\alpha$, "grande"):

$$\left[\bar{X}_n - \frac{\sigma}{\sqrt{n}}\,\phi_{1-\frac{\alpha}{2}},\ \bar{X}_n + \frac{\sigma}{\sqrt{n}}\,\phi_{1-\frac{\alpha}{2}}\right] \quad (\alpha = \ldots) \tag{148}$$

$$\text{ovvero} \quad \bar{X}_n \pm \frac{\sigma}{\sqrt{n}}\,\phi_{1-\frac{\alpha}{2}} \quad (\alpha = \ldots)$$

(ovviamente specificando il valore di α) e in particolare ricordiamo il valore $\phi_{0.975} \approx 1.96$ da cui l'intervallo classico con livello di confidenza del 95%, cioè $\alpha = 0.05$,

$$\left[\bar{X}_n - 1.96\,\frac{\sigma}{\sqrt{n}},\ \bar{X}_n + 1.96\,\frac{\sigma}{\sqrt{n}}\right] \quad (\alpha = 0.05) \tag{149}$$

BOZZA DRAFT

$$\text{ovvero} \quad \bar{X}_n \pm 1.96 \frac{\sigma}{\sqrt{n}} \quad (\alpha = 0.05)$$

Come sopra, per μ con σ^2 <u>non nota</u>, caso più verosimile:

$$\left[\bar{X}_n - \frac{S_n}{\sqrt{n}} t_{1-\frac{\alpha}{2}}(n-1),\, \bar{X}_n + \frac{S_n}{\sqrt{n}} t_{1-\frac{\alpha}{2}}(n-1)\right] \quad (\alpha = ...) \qquad (150)$$

$$\text{ovvero} \quad \bar{X}_n \pm \frac{S_n}{\sqrt{n}} t_{1-\frac{\alpha}{2}}(n-1) \quad (\alpha = ...)$$

(ovviamente specificando il valore di α)
- con (invece che i quantili normali) i quantili di Student, e
- con (invece che σ) la radice quadrata S_n dello stimatore S_n^2 di σ^2.

Si noti che esistono infiniti intervalli bilateri allo stesso livello, non centrati in $\bar{X}_n$, e 2 unilateri; ne mostreremo solo 1 unilatero.

Per μ, un intervallo <u>unilatero</u> al livello di confidenza $1-\alpha$:

$$\left]-\infty,\, \bar{X}_n + \frac{S_n}{\sqrt{n}} t_{1-\alpha}(n-1)\right]. \quad (\textit{Qua ovviamente } \sigma^2 \textit{ non nota}). \qquad (151)$$

(Si specifichi α).

La quantità $\frac{S_n}{\sqrt{n}}$ che ricorre nelle precedenti formule ha un nome:

$$\textit{errore standard} := \frac{S_n}{\sqrt{n}} \qquad (152)$$

e potrebbe trovarsi abbreviato *se*, dall'inglese.

Note sui quantili. I valori dei quantili di Student si trovano sulle tavole, cartacee o sulla rete, di non immediata lettura purtroppo, e vengono calcolati da molti software.
Per grandi valori di n, i quantili di Student possono in pratica sostituirsi coi quantili normali, le cui tavole sono di più facile lettura. Questa sostituzione diventa possibile perché per grandi valori di n

BOZZA - DRAFT

lo stimatore S_n^2 tende a σ^2, cosicché di fatto la varianza diventa "sperabilmente nota" e (circa) uguale a S_n^2.
Preferiremo non fare quest'approssimazione incerta, preferendogli la formula (150), ma ci si aspetti di trovarla nei testi scientifici, in particolare l'incerta formula per il classico 95%

$$\bar{X}_n \pm 1.96 \frac{S_n}{\sqrt{n}} \quad (\alpha = 0.05)$$

che almeno eviteremo come la peste per piccoli n. Certo per $n > 120$ l'approssimazione 1.96 non è cattiva; ecco alcuni quantili di Student per $\alpha = 0.05$ ovvero $1 - \frac{\alpha}{2} = 0.975$:

30 2.042
40 2.021
60 2.000
120 1.980
∞ 1.960

ed eccolo, infine al limite, l'1.96.
D'altra parte, è inutile cercare qua il pelo nell'uovo, quando poi nella pratica l'1.96 stesso viene approssimato spesso con 2, da cui la classica formulazione *pratica*, che troviamo su Wikipedia (in inglese) alla voce Confidence interval:

plus or minus twice the standard error

cioè

$$95\%\ \text{C.I.:} \quad \bar{X}_n \pm 2 \frac{S_n}{\sqrt{n}} \tag{153}$$

di largo uso *pratico*, ma in generale la eviteremo, preferendogli la formula (150). Ma può essere utile per semplici stime a mente.

Tavole (stampabili) dei quantili normali, di Student e del chi quadrato (e altre) si trovano per esempio a questo `link->`

Ecco un esempio di intervallo di fiducia tratto da un articolo scientifico riportato su PMC, sito governativo statunitense:

BOZZA - DRAFT

> The seroprevalence of latent toxoplasmosis in subjects involved in traffic accidents (N = 146) and in the general population living in the same area (N = 446) was compared [...] subjects with latent toxoplasmosis had a 2.65 ($C.I._{95}$= 1.76-4.01) times higher risk of an accident than the toxoplasmosis-negative subjects.

(Si noti che la *stima puntuale* 2.65 non è al centro dell'intervallo $[1.76, 4.01]$, che è bilatero ma non centrato).

53.3 Apprezziamo i CI in un articolo scientifico

Ci sono un'infinità, verosimilmente milioni, di articoli scientifici con i *confidence interval*.

Apprezziamo i CI in un articolo[177] scientifico, e pure i valori p; e seppure non approfondiamo, RR significa *relative risk*:

> The results showed that aspirin use was associated with a reduction in COVID-19 mortality (adjusted RR 0.69; 95% CI 0.50-0.95; P < 0.001). Subgroup analysis found that the low-dose group was associated with a reduced COVID-19 mortality (adjusted RR 0.64; 95% CI 0.48-0.85; P < 0.01). Aspirin use was associated with reduced COVID-19 mortality in Europe and America (crude RR 0.71; 95% CI 0.52-0.98; P = 0.04), and results from cohort studies suggested that aspirin use was a protective factor for COVID-19 mortality (adjusted RR 0.73; 95% CI 0.52-0.99; P = 0.04). Meanwhile, aspirin use was not associated with bleeding risk (crude RR 1.22; 95% CI 0.80-1.87; P = 0.96).

Si noti che gli stimatori puntuali non cadono a metà degli intervalli di fiducia, che non provengono da distribuzioni normali.
Semplificatamente, possiamo dire che "Aspirin use was associated with reduced COVID-19 mortality in Europe and America (crude RR 0.71; 95% CI 0.52-0.98; P = 0.04)" significa che in base a questo studio, in Europa e America l'uso dell'aspirina corrispondeva a una

[177] Ma S, Su W, Sun C, Lowe S, Zhou Z, Liu H, Qu G, Xia W, Xie P, Wu B, Gao J, Feng L, Sun Y. Does aspirin have an effect on risk of death in patients with COVID-19? A meta-analysis. Eur J Clin Pharmacol. 2022 Sep;78(9):1403-1420. doi: 10.1007/s00228-022-03356-5. Epub 2022 Jun 22. PMID: 35732963; PMCID: PMC9217117.

BOZZA - DRAFT

probabilità di morte pari al 71% rispetto al non uso, e per quel valore viene dato questo l'intervallo di confidenza al 95%: $[52\%, 98\%]$. E certo, se il "vero" valore è 98%, ben poco utile sarebbe l'aspirina, ma se è 52% si avrebbe una probabilità di morte praticamente dimezzata; il vero valore non sarà *mai* noto, ma dai dati empirici gli Autori producono la stima puntuale 71% e quella intervallare $[52\%, 98\%]$, e adesso siamo in grado di capirlo.

Questo tipo di espressioni

RR 0....; 95% CI 0....-0....

e altre consimili con OR invece di RR, di significato in qualche modo somigliante, sono diffusissime negli articoli scientifici di Farmacia e Medicina.
Google per

`"pubmed" "RR" "95%"`

dice (2022) di trovare 38.3 milioni di risultati.

BOZZA - DRAFT

Il primo (oggi 15 dicembre 2022) è un articolo[178] della Cochrane Database Systematic Reviews dell'agosto 2021 sull'antivirale remdesivir:

> **Authors' conclusions:** Based on the currently available evidence, we are moderately certain that remdesivir probably has little or no effect on all-cause mortality at up to day 28 in hospitalised adults with SARS-CoV-2 infection.

In seguito probabilmente sono riusciti a farlo funzionare un po' meglio, e attualmente (15 dicembre 2022) il famoso (e costoso) remdesivir è consigliato – detto semplificatamente – tranne che nei casi critici: https://www.bmj.com/content/370/bmj.m3379 (in aggiornamento continuo).

La Cochrane va a fare ciclopiche revisioni su tutti gli articoli scientifici su un determinato farmaco, o malattia. I farmaci quando vengono presentati, dagli studi pubblicati dai produttori, spesso appaiono ottimi. Poi, iniziano a studiarli anche gli altri ricercatori. E magari tirando le somme resta molto meno di quanto annunciato all'inizio.

[178] Ansems K, Grundeis F, Dahms K, Mikolajewska A, Thieme V, Piechotta V, Metzendorf MI, Stegemann M, Benstoem C, Fichtner F. Remdesivir for the treatment of COVID-19. Cochrane Database Syst Rev. 2021 Aug 5;8(8):CD014962. doi: 10.1002/14651858.CD014962. PMID: 34350582; PMCID: PMC8406992.

Il termine "crude RR" si riferisce al "Relative Risk" (RR) calcolato senza aggiustamenti per confonditori o altri fattori. Il crude RR rappresenta il rischio relativo di un evento tra due gruppi, calcolato semplicemente confrontando gli eventi senza considerare altre variabili che potrebbero influenzare i risultati.

Ulteriori Dettagli:
Calcolo: Il crude RR è calcolato come il rapporto tra la probabilità di un evento in un gruppo esposto a un fattore di rischio e la probabilità dello stesso evento in un gruppo non esposto.

Aggiustamenti: Gli RR aggiustati, invece, sono calcolati tenendo conto di potenziali confonditori, il che significa che forniscono una stima più precisa del vero effetto del fattore di rischio.

53.4 Prevenire è meglio che curare: le 2 SD e il 95%

Vediamo di prevenire qua 2 errori tipici su media±2·SD.

Errore 1. *Claim:* Più o meno il 95% dei dati di un dataset sta sempre fra la media meno 2 deviazioni standard e la media più 2 deviazioni standard.

Falso! Questo si può affermare se i dati sono più o meno normalmente distribuiti, altrimenti può succedere o no. Per questo dataset

0.001,0.01,0.1,1, 1.001,1.01,1.1,19,19.001,19.01,19.1,20,20.001,20.01,20.1

la media è ≈ 10.7, la deviazione standard ≈ 9.79 e l'intervallo che si otterrebbe $[0.91, 20.49]$ contiene solo $11/15 \approx 73\%$ dei dati.

Errore 2. *Claim:* Più o meno il 95% di n valori di una variabile aleatoria sta fra la media meno 2 deviazioni standard e la media più 2 deviazioni standard.

Doppiamente falso! Questo si può affermare di solito per n determinazioni di una variabile aleatoria più o meno normalmente distribuita. L'errore comunque è più grave di quello soprastante, perchè la deviazione standard calcolata (come prima) sui dati di un dataset esprime sempre una caratteristica dei valori numerici, mentre per una variabile aleatoria la deviazione standard è un valore teorico, che potrebbe non avere nessuna relazione coi valori di una determinazione empirica di un campione aleatorio. Bene o male, nell'esempio soprastante la deviazione standard 10.7 qualcosa esprime relativamente ai dati, mentre se di una variabile aleatoria per esempio $N(0, 1)$ traiamo i valori sopradetti, il che è improbabilissimo ma possibile, la deviazione standards 1 della variabile aleatoria considerata non ha nessuna relazione coi valori.

BOZZA DRAFT

Short version.

media±2·SD comprende ≈ 95% dei valori per distribuzioni **normali**, sempre nei dataset[179], e spesso per le variabili aleatorie.

53.5 Possibile delusione dagli articoli scientifici

Frequentando gli articoli scientifici, per esempio gli abstract riportati su PubMed, si potrà incorrere in questa delusione: spesso gli intervalli di confidenza, trovati e pubblicati, sono amplissimi.
Per esempio un articolo scientifico potrebbe affermare che col tal farmaco un verme di laboratorio vive mediamente il doppio, ma per il fattore 2, guardando meglio, viene dato un intervallo di confidenza 1.1-4. Allora, sì, mediamente vive il doppio, è questa è la stima puntuale, che in generale verrà pubblicata sui media se il risultato scientifico li raggiunge, ma quello che è stato trovato, nel senso degli intervalli di confidenza al 95%, fa supporre una vita allungata fra il 10% (altro che doppio!) e il quadruplo. (Concretamente, grande incertezza).
Un esempio fra un'infinità, con l'OR, una sorta di confronto del rischio relativamente a 2 situazioni: si ritrova subito in rete la frase

tobacco consumption [OR:2.8 (CI95%:1.3-6.1)]

dove ovviamente in un articolo di divulgazione scientifica si potrà dire *rischio triplicato*, arrotondando il 2.8, ma il valore più piccolo è 1.3, cioè solo +30%, ben diverso dal triplicamento.

E talvolta l'intervallo di confidenza per l'OR contiene il numero 1, perdendosi così la significatività statistica della ricerca statistica fatta: per esempio un OR da 0.5 a 2 indica un rischio da dimezzato a raddoppiato; anche se la media è 0.9 o 1.1. (Risultato alquanto inconcludente).

BOZZA - DRAFT

[179] Si noti che a rigore non esiste un dataset esattamente con distribuzione normale, perchè essa è una densità continua, ma è ovvio che qua ci si riferisce a un dataset con funzione di ripartizione empirica "molto vicina" a quella di una densità normale, e specificare i dettagli non sarebbe agevole.

53.6 Esercizi

53.6.1 Esercizio

$_{\mu 2023}\approx$ Determinare con la formula

$$\overline{X}_n \pm \frac{\sqrt{S^2}}{\sqrt{n}}\, t_{1-\frac{\alpha}{2}}(n-1)$$

(in cui α sarà 0.05) l'intervallo di fiducia (ovvero di confidenza) al 95% per la media della variabile aleatoria normale da cui è stato tratto questo campione:

3,990 4,767 5,642 −3,772 −4,158 −2,190 7,954 0,765 6,446 1,227 2,950 0,852

Il calcolo sarà facilitato sapendo che $S^2 \approx 15{,}747$ (calcolato col computer) e che il quantile di Student necessario è $\approx 2{,}201$ (trovato sulle tavole).

SVOLGIMENTO

Viene usato lo standard della virgola decimale. (Che la virgola non è separatore delle migliaia si vede dai numeri 0,765 e 0,852)

La media aritmetica degli $n = 12$ valori è

$$\overline{X}_{12} = \frac{1}{12}\big(3{,}990 + 4{,}767 + 5{,}642 + (-3{,}772) + (-4{,}158) + (-2{,}190)$$

$$+7{,}954 + 0{,}765 + 6{,}446 + 1{,}227 + 2{,}950 + 0{,}852\big) =$$

$$= \frac{24{,}473}{12} \approx 2{,}039\,417 \approx 2{,}039$$

(e – nel valore più preciso che poi abbandoneremo – abbiamo usato lo spazietto separatore delle terne di cifre, ma stiamo seguendo solo parzialmente il NIST, National Institute of Standards and Technology, statunitense, perchè quello esige il punto decimale, non la virgola).

Applicheremo la formula scritta nel quesito, dell'intervallo di confidenza al 95% bilatero centrato per μ per un campione gaussiano, con σ^2 non nota:

$$\overline{X}_n \pm \frac{\sqrt{S^2}}{\sqrt{n}}\, t_{1-\frac{\alpha}{2}}(n-1)$$

e in questo calcolo siamo molto favoriti perchè nel testo del quesito ci sono dati (approssimatamente) il valore di S^2, alquanto laborioso da calcolare, $\frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X}_n)^2$, e il quantile $t_{1-\frac{\alpha}{2}}(n-1)$, in questo caso $t_{0.975}(11)$, non banale da trovare su normali tavole, gravate da varie ambiguità notazionali:

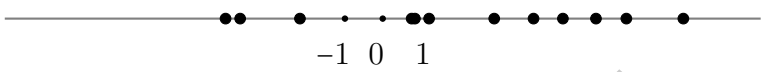
$$2{,}039 \pm \frac{\sqrt{15{,}747}}{\sqrt{12}}\, 2{,}201$$

BOZZA - DRAFT

$$2{,}039 \pm \frac{3{,}968}{3{,}464}\, 2{,}201$$

$$\boxed{\text{CI } 95\%\text{: } 2{,}039 \pm 2{,}521}$$

Nota 1. Ecco un diagramma cartesiano dei 12 valori (ma 2 sono quasi sovrapposti).



Nota 2. In effetti il campione era stato tratto, salvo arrotondamenti, online con WolframAlpha con

`12 random sample, normal distribution, mean 1, variance 25`

(ovviamente seguendo il link il software produce ogni volta un nuovo campione) e possiamo dirci soddisfatti perchè il vero valore $\mu = 1$ sta effettivamente nell'intervallo di confidenza trovato. Un motivo di insoddisfazione è che l'intervallo è molto ampio, ma questo è dovuto anche alla piccolezza del campione.

BOZZA _ DRAFT

53.6.2 Esercizio

$_{\mu 2019}$ ≈ Determinare l'intervallo di fiducia per la media, consueto (bilatero centrato) al 95%, con questo campione gaussiano di varianza 9:

17.65 24.38 22.86 20.09 20.71 25.75 21.85
15.82 22.53 19.66 20.26 18.99 17.09 20.52.

Fra le molte possibili scritture della soluzione, stavolta si usi la

$$\text{CI}_{95} = [a, b]$$

("CI" sta per *confidence interval*; si usano anche "95%CI" e molte varianti).

SVOLGIMENTO

Con la media del campione di rango $n = 14$

$$\bar{X}_{14} = \frac{1}{14}\,(17.65 + 24.38 + 22.86 + 20.09 + 20.71 + 25.75 + 21.85+$$

$$+15.82 + 22.53 + 19.66 + 20.26 + 18.99 + 17.09 + 20.52) = \frac{288.16}{14} \approx 20.582\,86$$

e con la nota formula dell'intervallo di fiducia in questione

$$\text{CI}_{95}: \;\; \bar{X}_n \pm 1.96\,\frac{\sigma}{\sqrt{n}}$$

ora con $\sigma = \sqrt{9} = 3$

$$20.582\,86 \pm 1.96\,\frac{3}{\sqrt{14}}$$

$$20.582\,86 \pm 1.96\,\frac{3}{3.741\,66}$$

$$20.582\,86 \pm 1.96 \cdot 0.801\,784$$

$$20.582\,86 \pm 1.571\,5$$

$$\boxed{\mathrm{CI}_{95} = [19.0, 22.2]}$$

o anche

$$\boxed{\mathrm{CI}_{95} = [19.01, 22.15]}$$

(Salvo un'approssimazione a 2 cifre decimali fatta a mano, i valori erano stati simulati con WolframAlpha con la varianza 9 e la media 20, che effettivamente sta nell'intervallo trovato, seppure in posizione alquanto decentrata. Seguendo questo link potete farvi produrre un altro campione siffatto di 14 elementi, e cliccando di nuovo un altro ancora: https://www.wolframalpha.com/input/?i=14+random+normal+distribution+mean%3D20+variance%3D9; campioni sempre nuovi appena prodotti).

BOZZA DRAFT

ES.ERCIZIO$_{\mu 2019}$
≈ Di 200 soggetti si è misurato un parametro fisiologico producendo un campione che si ritiene gaussiano, e con un foglio di calcolo si è trovato

$$\frac{1}{200}\sum_{k=1}^{200} X_k = 83.21 \qquad \frac{1}{199}\sum_{k=1}^{200}(X_k - \bar{X}_{200})^2 = 1\,405.38$$

Con la grossolana *formula pratica*, di largo uso nelle Scienze Applicate,

$$C.I._{.95}: \ \bar{X}_n \pm 1.96\,\frac{S_n}{\sqrt{n}}$$

trovare il consueto intervallo di fiducia della media, nella forma $C.I._{.95}: [a, b]$.

SVOLGIMENTO
(Il termine $\pm 1.96\,\frac{S_n}{\sqrt{n}}$ ha un errore circa del 2% per $n = 60$ e circa dell'1% per $n = 120$, rispetto al più corretto $\pm\frac{S_n}{\sqrt{n}}\,t_{0.975}(n-1)$, coi quantili di Student, e poi tende a 0; ma tutto ciò non serve per rispondere al quesito).
Ci sono dati

$$\bar{X}_{200} = 83.21 \qquad S^2_{200} = 1\,405.38$$

da cui con la *formula pratica* riportata

$$C.I._{.95}: 83.21 \pm 1.96\,\frac{\sqrt{1\,405.38}}{\sqrt{200}} \approx$$

$$\approx 83.21 \pm 1.96\,\frac{37.4884}{14.1421} \approx$$

$$\approx 83.21 \pm 1.96 \cdot 2.651 =$$

$$= 83.21 \pm 5.1956$$

e infine nella forma richiesta

$$\boxed{C.I._{.95}: [78.0, 88.4]}$$

Nota. Di questa Lezione ci si dovrebbe aspettare che lo studente conosca a memoria le formule (149), (150), (152), oltre a tutto l'inquadramento teorico della questione.
In particolare, con le (148), (151), (153), bisognerebbe saper operare se vengono fornite.

BOZZA - DRAFT

ESERCIZI SULLA LEZIONE 53

ESERCIZIO$_{\mu 2018}$
≈ Conoscendo la varianza 3.2 di una variabile aleatoria normale si determini l'intervallo di fiducia bilatero consueto della media al livello 95% relativamente a questo campione

-1.32 3.78 1.58 1.55 3.96 0.85 4.78 3.74 3.25 0.99

esprimendolo in entrambe le forme usuali: $a \pm b$, e anche $[u, v]$.

Svolgimento.
La media degli $n = 10$ dati é 2.316.
Consideriamo la classica formula dell'intervallo di fiducia al 95% ovvero 5%

$$\mu = \bar{X}_n \pm 1.96 \frac{\sigma}{\sqrt{n}} \quad (\alpha = 0.05)$$

per la quale ora troviamo $2.316 \pm 1.96 \frac{\sqrt{3.2}}{\sqrt{10}}$ e a conti fatti

$$\boxed{\begin{array}{l} \mu = 2.316 \pm 1.109 \quad (\alpha = 0.05) \\ \\ \mu \in \left[1.207, 3.425\right] \quad (\alpha = 0.05) \end{array}}$$

ovvero, con sole 3 cifre significative come nei dati originali,

$$\boxed{\begin{array}{l} \mu = 2.32 \pm 1.11 \quad (\alpha = 0.05) \\ \\ \mu \in \left[1.21, 3.42\right] \quad (\alpha = 0.05) \end{array}}$$

(Si potrebbe obiettare che un corretto arrotondamento a 3 cifre significative di 3.425 é 3.43 e non 3.42; questo é vero ma il 3.425 era giá esso stesso un arrotondamento, di 3.4247..., e arrotondando a 3 cifre significative quest'ultimo, che é il vero valore, si ottiene appunto 3.42).
(I valori erano stati ottenuti simulando $N(2.4, 3.2)$).

BOZZA DRAFT

54 Intervalli di fiducia per la varianza

Il Lettore interessato forse vorrà sapere questo: come esistono intervalli di fiducia per la media, esistono anche per la varianza, e sono basati sui quantili del chi quadrato.

Lezione esclusa dal corso breve – 40 ore.

Per σ^2 con μ non nota, (un) intervallo di fiducia bilatero al livello (di confidenza) $1-\alpha$, dove α è "piccolo", come $0.05 = 5\%$ (da non confondersi con $1-\alpha$, "grande"):

$$\Big[\frac{n-1}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\,S_n^2\,,\ \frac{n-1}{\chi^2_{\frac{\alpha}{2}}(n-1)}\,S_n^2\Big]. \quad (\textit{Non centrato in } S_n^2). \tag{154}$$

Come sopra, un intervallo unilatero:

$$\Big[0\,,\ \frac{n-1}{\chi^2_{\alpha}(n-1)}\,S_n^2\Big]. \tag{155}$$

Nota 1. Le 2 formule soprastanti valgono per campioni gaussiani e (approssimatamente) anche per campioni non gaussiani se la numerosità è $n \geq 30$ e la densità non è "troppo" asimmetrica.

Nota 2. I quantili del chi quadrato, oltre a venire calcolati (approssimatamente) da molti software, si trovano (approssimati) per alcuni tipici valori di α e piccoli valori di n su apposite tavole numeriche, e poi vale (teorema) l'approssimazione

$$\forall\, n \geq 30 \qquad \chi^2_{\alpha}(n) \approx \frac{1}{2}\Big(\phi_\alpha + \sqrt{2\,n-1}\Big)^2. \tag{156}$$

Nota 3. Si possono considerare anche intervalli di fiducia per altri parametri e altri tipi di variabili aleatorie, per esempio$^{\uparrow}$ per p di $B(1,p)$.

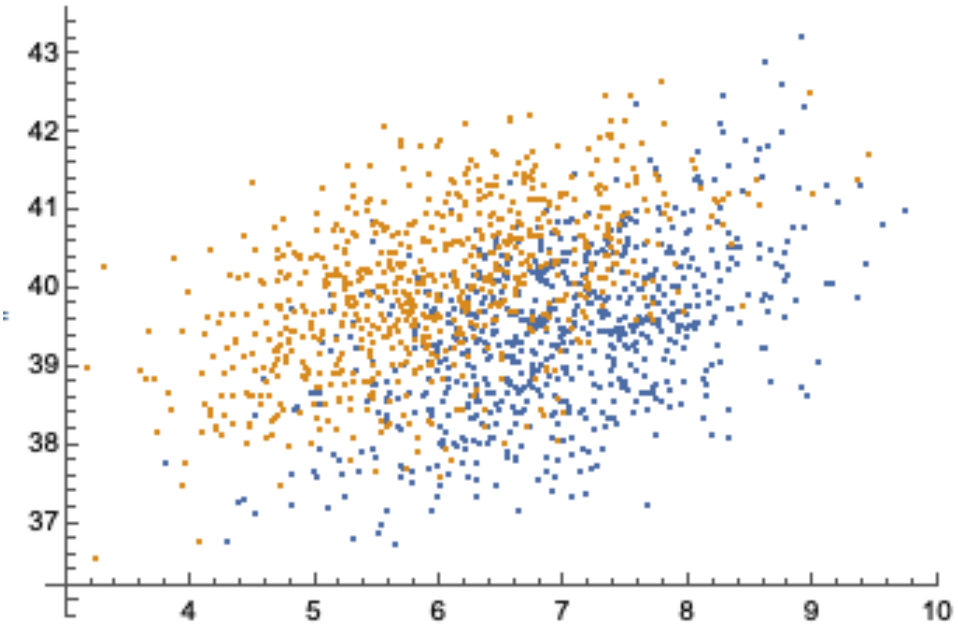
Nota. Di questa Lezione ci si dovrebbe aspettare che lo studente conosca tutto l'inquadramento teorico della questione.
In particolare, con le (154), (155), (156) bisognerebbe saper operare se vengono fornite.

BOZZA - DRAFT

XI – Test Statistici

BOZZA - DRAFT

55 I Test Statistici



Arruoliamo $2n$ persone con BMI di 40.0 e a metà di loro diamo una terapia A e all'altra metà una terapia B.
Alla fine gli chiediamo un giudizio di soddisfazione da 3 a 10, anche coi decimali, salvo il 10, valore massimo, e misuriamo i BMI. Lo scatterplot rappresenta in giallo l'esito della terapia A, e in blu quello della terapia B. La terapia B ha dato più soddisfazione? Mediamente pare di sì, ma non sarà forse una fluttuazione casuale? La terapia B ha diminuito il BMI? Sembrerebbe un po' di sì, ma qua siamo molto più incerti. E la A? Ancora più incerti. Ci servono FORMULE.

BOZZA

55.1 Introduzione ai Test Statistici

Dalle determinazioni $x_1,...,x_n$ di un campione aleatorio $X_1,...,X_n$ tratto da una v.a. X di densità nota salvo un suo parametro a, vogliamo rispondere con "sì" o "no" , con "ragionevole certezza statistica" ovviamente, a una domanda sul parametro incognito.

Per noi a in generale è un numero incognito, come p che poi è

la media μ in $B(1,p)$, ma lasciamo aperta la possibilità, di livello superiore, che sia una coppia di numeri come (μ,σ^2).

La domanda potrebbe essere per esempio

$\mu > 0$? per una v.a. $N(\mu,\sigma^2)$, oppure

$p = \frac{1}{2}$? per una v.a. $B(1,p)$ che nella realtà sensibile può significare per esempio la regolarità di una moneta; da decidersi con ragionevolezza statistica dai risultati di molti lanci.

In Farmacia:

La glicemia è diminuita? (Cioè, così tanto da far ragionevolmente ipotizzare un effetto causale piuttosto che casuale).

Possiamo dire che il farmaco fa più del placebo?

BOZZA - DRAFT

Il test statistico si preordina – prima di avere i dati in mano ovvero prima di fare un esperimento nella realtà sensibile – formulando un'ipotesi statistica, indicata con H o H_0, *ipotesi nulla*, e una ipotesi *alternativa*, indicata con A o rispettivamente H_1, per esempio

$$H : p = \frac{1}{2} \qquad A : p \neq \frac{1}{2}\,.$$

Anticipiamo che l'ipotesi nulla H va identificata in generale col caso che si spera che non sia. ("Vogliamo A!").

Un esempio minimo potrebbe essere così: lanceremo 5 volte la moneta e rifiuteremo l'ipotesi di regolarità se viene testa 0 o 5 volte, perché se la moneta è regolare quei risultati hanno complessivamente probabilità 1/16, un po' pochino.
In realtà la statistica usuale viene fatta "a 1/20", cioè al 5 ovvero 95%; ma largheggiando possiamo fare Statistica "al 90%" e allora respingeremmo l'ipotesi della regolarità con 0 o 5 teste su 5 lanci.

L'insieme $\{0,5\}$ è la *regione critica* ovvero di rigetto dell'ipotesi (nulla). La regione critica di solito viene espressa come un sottoin-

sieme D di $\mathbb{R}$ in cui una certa funzione del campione aleatorio può cadere (e allora rifiutiamo H) o non cadere (non rifiutiamo H). In questi termini, potremmo porre la regione critica $\{0, 5\}$ e verificare se vi cade $X_1 + ... + X_5$ o più usualmente porre $D := \{0, 1\}$ e verificare se vi cade $\bar{X}_5$. In casi più significativi di questo microscopico esempio la regione critica di solito ha una forma del tipo $x > x_0$ (test unilatero) oppure $x < x_1 \vee x > x_2$ (test bilatero).
Vediamo un altro esempio. Sia I la variabile aleatoria che è la glicemia (iniziale) di un soggetto qualunque di un campione di 20 soggetti (persone iperglicemiche) e F la glicemia (finale) dopo la somministrazione di un certo farmaco. Ci potrebbe interessare se mediamente la glicemia diminuisce con quel farmaco cioè se la media (parametro incognito) della variabile aleatoria $X := I - F$ è > 0. (Iniziale grande, finale piccola).
Si formula l'ipotesi nulla: il farmaco non riduce la glicemia:

$$H : \mu \leq 0 \qquad \text{(finale grande come o più dell'iniziale)}$$

essendo μ la media di X.
In realtà spero che riduca la glicemia: ipotesi alternativa:

$$A : \mu > 0 \qquad \text{(finale più piccola dell'iniziale)}$$

Misuriamo la glicemia nei 20 soggetti (campione, o più precisamente determinazione $i_1, \ldots, i_{20}$ di un campione aleatorio $I_1, \ldots, I_{20}$). Diamo ai 20 soggetti il farmaco. Misuriamo di nuovo la glicemia dei 20 soggetti ottenendo così 20 determinazioni $f_1, \ldots, f_{20}$ della variabile aleatoria F. Facciamo 20 sottrazioni ottenendo 20 determinazioni $x_1, \ldots, x_{20}$ della variabile aleatoria X, differenza *prima-dopo* ovvero iniziale-finale. Della variabile aleatoria X vogliamo sapere se la media (speranza matematica) è > 0, come speriamo, oppure no.

L'idea ingenua è fare la media aritmetica dei 20 numeri e concludere che se è > 0 il farmaco ha diminuito la glicemia.
Se fosse così in questo e analoghi casi, la statistica inferenziale non servirebbe, ma non è così: quella verifica non dice di per sè

BOZZA - DRAFT

sostanzialmente nulla perché non distingue l'effetto del farmaco dalle inevitabili fluttuazioni casuali di X, che, non per niente, è da considerarsi una variabile *aleatoria*. (Non possiamo certo aspettarci un effetto *deterministico* del farmaco, che *sempre* riduca la glicemia).
È invece necessario applicare un opportuno test statistico, cioè di fatto applicheremo una non banale formula che ci potrà dire, nel caso che la media aritmetica degli x_i sia > 0, che quell'effetto con ragionevole plausibilità non è casuale. Se invece la media aritmetica viene negatitiva l'esperimento è andato male e la statistica non ci aiuta ulteriormente. Alla fine rifiutiamo o non rifiutiamo l'ipotesi nulla. Speriamo di rifiutarla.
Alcuni dicono "accettare" l'ipotesi nulla ma il modo corretto di vedere le cose è "non rifiutarla". Non abbiamo dimostrato che è vera: semplicemente non siamo riusciti a dimostrarla *verosimilmente falsa*.

BOZZA - DRAFT

Media degli X_i, differenza prima-dopo		
Statistica ingenua:		
La glicemia mediamente non è diminuita; il farmaco non funziona		La glicemia è diminuita; il farmaco funziona
0		
Statistica inferenziale:		
Non rifiutiamo l'ipotesi che la glicemia sia uguale o aumentata ovvero $\mu \leq 0$		Rifiutiamo l'ipotesi che la glicemia sia uguale o aumentata ovvero $\mu \leq 0$
0	*soglia*	È plausibile che il farmaco funzioni.

Insomma X deve essere mediamente ben > 0, non solo > 0, per escludere con ragionevole verosimiglianza la fluttuazione casuale.

Quanto > 0, lo dicono apposite formule che vedremo, che fanno uso dei quantili, di Student in questo caso.

I quantili si trovano e soprattutto si trovavano su tavole numeriche, che permettevano di ottenere la "ragionevolezza al 95%" o ancor meglio al 99%, e anche con altri *livelli di confidenza* tipici.
Diciamo subito che nella pratica si trova scritto indifferentemente "al 99%" o "all'1%" con lo stesso significato, e similmente con 95 e 5, eccetera: purtroppo c'è un'ambiguità terminologica.
Da adesso in questo paragrafo facciamo rifermento al valore "piccolo": non 0.95 ma 0.05, non 0.99 ma 0.01, eccetera.

BOZZA DRAFT

55.2 Il test della signora del tè

Da Wikipedia, l'enciclopedia libera, un esperimento degli anni '30 del secolo scorso, alla base della moderna Statistica:

> The experiment is the original exposition of Fisher's notion of a null hypothesis (...) The lady in question (Muriel Bristol) claimed to be able to tell whether the tea or the milk was added first to a cup. Fisher proposed to give her eight cups, four of each variety, in random order (...) The experiment provides a subject with 8 randomly ordered cups of tea – 4 prepared by first pouring the tea, then adding milk, 4 prepared by first pouring the milk, then adding the tea. The subject has to select 4 cups prepared by one method (...) The null hypothesis is that the subject has no ability to distinguish the teas (...) The critical region for rejection of the null of no ability to distinguish was the single case of 4 successes of 4 possible, based on the conventional probability criterion < 5%. This is the critical region because under the null of no ability to distinguish, 4 successes has 1 chance out of 70 ($\approx 1.4\% < 5\%$) of occurring (...) in the actual experiment the lady succeeded in identifying all eight cups correctly

> (...) in 70 (the combinations of 8 taken 4 at a time). (...) the famous case of the 'lady tasting tea' (...) one of the two supporting pillars ... of the randomization analysis of experimental data.

Il valore $\frac{1}{70}$ viene dalla Probabilità Combinatoria, essendo

$$\binom{8}{4} = \frac{8 \cdot 7 \cdot 6 \cdot 5}{1 \cdot 2 \cdot 3 \cdot 4} = 70$$

i modi di scegliere 4 elementi su 8 (combinazioni di 8 elementi a 4 a 4). L'indovinare solo 3 tazze su 4 invece è un evento molto probabile, $\approx 24.3\%$, nell'ipotesi nulla di tirare a caso ovvero di non avere la capacità affermata, e si calcola in modo non semplicissimo (distribuzione ipergeometrica) analogo al calcolo per il terno al lotto.

BOZZA - DRAFT

55.3 Test a una coda e a due code

Wikipedia ci illustra un test per la regolarità della moneta molto meglio fatto di quello prima esposto con 5 lanci, e ne prevede 20. La regione critica è $\{0, 1, 2, 3, 4, 5, 15, 16, 17, 18, 19, 20\}$, che ha probabilità 0.041 se la moneta è regolare, a cui corrisponde la *regione di accettazione* $\{6, 7, 8, 9, 10, 11, 12, 13, 14\}$, l'insieme complementare. Questo è un *test a due code*.
Il valore 0.41 è la probabilità di un grave allontanamento da 10, metà di 20, speranza matematica del numero di teste nell'ipotesi di regolarità, cioè speranza matematica di $B\left(20, \frac{1}{2}\right)$:

$$P(X \le 5 \vee X \ge 15) = P(X \le 5) + P(X \ge 15) =$$

$$= 1 - P(6 \le X \le 14) = 1 - \sum_{k=6}^{k=14} \binom{20}{k} \left(\frac{1}{2}\right)^k \left(1 - \frac{1}{2}\right)^{n-k} =$$

$$= 1 - 2^{-20} \sum_{k=6}^{k=14} \binom{20}{k}$$

calcolabile con pazienza oppure con WolframAlpha che dà ≈ 0.4139.
Leggiamo in

https://it.wikipedia.org/wiki/Test_di_verifica_d%27ipotesi

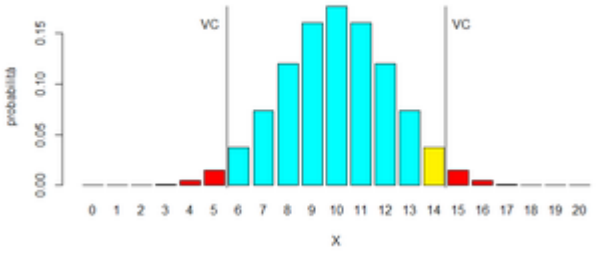


Figure 60: Rosso: regione di rifiuto del test a due code. Giallo: il numero 14 si aggiunge alla coda destra 15,...,20 dando – senza la coda sinistra – la regione di rifiuto del test a una coda. By Ppong.it in wikimedia.

Tale valore p è la probabilità di ottenere un valore altrettanto o più estremo di quello osservato, ammesso che la moneta fosse effettivamente bilanciata. Nel nostro caso è pari a 0,041, ovvero del 4,1%. Giudicando bassa tale probabilità, rigettiamo l'ipotesi di bilanciamento della moneta in esame, ritenendo accettabilmente basso il rischio di compiere un errore di giudizio. La probabilità di rifiutare l'ipotesi sottoposta a verifica, nel caso questa fosse corretta, è pari al massimo valore-p che saremmo stati disposti ad accettare.
(...) supponiamo che noi già prima di fare l'esperimento sospettassimo che fosse sbilanciata verso la testa, in tal caso potremmo dire che l'ipotesi nulla, che noi abbiamo intenzione di smentire, è che la probabilità che esca testa sia minore o uguale a 0,5, anziché necessariamente pari a 0,5. In tal modo evitiamo di rifiutare l'ipotesi nulla se otteniamo un numero di teste basso, ma se, al contrario, contiamo più di 10 teste, calcoliamo il valore-p senza tenere in considerazione i possibili risultati inferiori a 10. Come risultato, la regione di rifiuto perde gli elementi da 1 a 5, ma si allarga sulla destra includendo 14.

BOZZA - DRAFT

L'ipotesi e l'alternativa del test a una coda considerato sono

$$H : p \leq \frac{1}{2} \qquad A : p > \frac{1}{2}\,.$$

L'ipotesi e l'alternativa del test a due code considerato sono

$$H : p = \frac{1}{2} \qquad A : p \neq \frac{1}{2}\,.$$

55.4 Test a una coda facilmente fattibile con successo

Siamo abbastanza sicuri che le presente lezione di Farmacia nel 2022 sarà seguita in aula da più studentesse che studenti (succede ogni anno) ovvero detto altrimenti che il generico studente ha probabilità maggiore di essere di genere femminile piuttosto che maschile:

$$p := P(femmina) > \frac{1}{2} \qquad \text{(crediamo)}$$

Naturalmente potremmo contare e risolvere la questione, ma se la popolazione di riferimento fosse enorme o difficilmente indagabile nella sua totalità (per esempio se volessimo dimostrare che a Roma le gatte sono più numerose dei gatti, o simili cose su zecche o batteri o virus o quant'altro) non potremmo farlo. Faremo invece un test statistico su un campione di 20 soggetti presi in qualche modo a caso, per esempio in base alla prima lettera del cognome, che possiamo ragionevolmente ritenere indipendente dal genere del soggetto. **E la questione dell'indipendenza non sarà mai sottolineata abbastanza.**

BOZZA - DRAFT

Poniamo

$$H : p \leq \frac{1}{2} \qquad A : p > \frac{1}{2}$$

e speriamo di respingere l'ipotesi nulla. In base al paragrafo precedente, se troviamo 14 o più studentesse nel campione, ce l'abbiamo fatta, respingiamo l'ipotesi che siano meno o ugualmente numerose *in tutta la popolazione* considerata. Di solito 20 soggetti bastano perchè la sproporzione è notevole (non è affatto detto che sia così per i gatti di Roma), se fosse molto piccola ci vorrebbe un campione più numeroso per avere buone speranze di avere il desiderato successo col test.

55.5 A due code, facilmente fattibile, ma con insuccesso

È senz'altro vero che la probabilità che un soggetto sia nato in un giorno del mese pari non sia uguale a quella del dispari, ma la differenza è così piccola che quasi sicuramente con un campione di 20 soggetti non riusciremo a respingere l'ipotesi nulla

$$H : p = \frac{1}{2}$$

Comunque possiamo provare per vedere un po' come funziona "sul campo" la Statistica. Arruoliamo quindi 20 soggetti presi a caso, per esempio quelli più vicini al computer sulla cattedra, cosa che possiamo ragionevolmente ritenere indipendente dalla parità del giorno della nascita. **E la questione dell'indipendenza non sarà mai sottolineata abbastanza.**
In base a quanto sopra detto in 55.3, se troviamo meno di 6 o più di 14 nati in data dispari, respingiamo l'ipotesi dell'equiprobabilità, e saremmo contenti (ma difficilmente succederà).
Rivediamo tutto con ordine.
Una persona si convince che non è ugualmente probabile nascere in un giorno con data pari o dispari. Di fatto questo è vero, ma la differenza di probabilità è talmente piccola che il nostro ben difficilmente riuscirà dimostrarlo con qualche decina di dati (verosimilmente riuscirebbe con qualche milione di soggetti).

– Ipotesi nulla che vuol respingere: è equiprobabile nascere in data pari o dispari.

– Pre-fissa il livello di significatività standard 0.05

– Esperimento: intervista 20 persone.

– **Supponiamo per esempio** che trovi 14 pari e 6 dispari (è contento, si illude)

– In un modo o nell'altro calcola $P(pari \leq 6 \vee pari \geq 14) \approx 0.115$

– Non riesce a respingere l'ipotesi nulla al 5% (e neppure al 10%).

Se ne trovava 15, riusciva a respingere l'ipotesi nulla.
Si noti che l'obiezione che lui fa, se non capisce la Statistica, "ho trovato più del doppio, 14, in una classe rispetto all'altra, 6", è ir-

BOZZA - DRAFT

rilevante e fuorviante, su 20 soggetti; potrebbe bastare anche molto meno del +100%, anche solo il +1%, ma su un campione molto più numeroso.

55.6 Il p-value, in italiano valore p

Oggi i quantili vengono calcolati da numerosi software. L'uso di questi software ha permesso nei tempi moderni un passo ulteriore: trovare proprio la soglia discriminante, l'ultimo valore per il quale si passa dal non rifiutare al rifiutare l'ipotesi nulla, e questo valore soglia può ben essere diverso da 0.05 o 0.01, per esempio può essere 0.046, ed è il p-value ovvero valore p, e si trova spesso denotato P sulle riviste scientifiche di Farmacia e Medicina.
Questo 0.046 è proprio il p-value dei 10 000 lanci di moneta con 5 100 teste prima considerato, come si potrebbe calcolare col computer: rifiutiamo l'ipotesi (nulla) di regolarità (per poco) perchè 0.046 è minore di 0.05 (che però a rigore dovevamo fissare prima dell'esperimento, e ciò è tanto più vero per valori di significatività diversi da quello classico 0.05). Ma al livello dell'1%, detto del 99% da altri Autori, non possiamo rifiutarla. (Stiamo cercando di fare un'affermazione troppo categorica ma l'esperimento non ce lo consente; ce lo consentirebbe se fossero venute molte più teste).

BOZZA - DRAFT

L'ideale è trovare un p-value piccolissimo, magari $< 10^{-6}$, ma almeno ≤ 0.05.

Definizione. Formalmente il p-value è definito come la probabilità di ottenere un valore uguale o più estremo di quello ottenuto, nell'ipotesi che sia vera l'*ipotesi nulla*.

Rivediamo questo fatto del valore più estremo nell'esempio:

– Vogliamo "dimostrare" che la moneta non è regolare

– Ipotesi nulla: moneta regolare ovvero $p = 0.5$, probabilità della testa (e questo p non c'entra nulla col p-value che stiamo per intro-

durre, è un'ambiguità notazionale).

– Pre-fissiamo il livello di significatività standard 0.05

– Esperimento: lanciamo la moneta 10 000 volte

– Vengono 5 100 teste

– In un modo o nell'altro calcoliamo $P(teste < 4\,900 \vee teste \geq 5\,100) \approx 0.046$

– Rifiutiamo l'ipotesi della regolarità con P 0.046

Nota finale. Il nostro *trial* è andato come speravamo.

Vengono pubblicati anche articolo scientifici con p value 0.1 ma apprezzatissimi sono p value come 10^{-3} o minori, anche 10^{-6} o più piccoli ancora.
Solo per fare un esempio, p value 0.0375, da un articolo[180] scientifico relativo al Sud Africa:

> The minimum distance between the source location and the medical facility was compared with the malaria case mortality rate to determine the relationship between the two factors (...) The correlation analysis illustrated a Spearman's rank correlation coefficient of 0.3816 with a p value of 0.0375.

BOZZA - DRAFT

Minor distanza minima dagli ambulatori, minor mortalità, appare. (Nota: in Italia il numero degli ospedali pubblici diminuisce da decenni, per *spending review*).

Solo per fare un esempio di p value molto piccoli, fra innumerevoli, il preprint[181] *Solar UV - B/A Radiation is Highly Effective in Inactivating SARS-CoV-2* trova per l'ipotesi nulla (sostanzialmente: che il covid sia indipendente dall'irraggiamento solare locale) dei p value dell'ordine di 10^{-11}. E un altro preprint[182] trova dei

[180] Coetzer RH, Adeola AM. Assessing the Correlation between Malaria Case Mortality Rates and Access to Health Facilities in the Malaria Region of Vhembe District, South Africa. J Environ Public Health. 2020 Dec 2;2020:8973739. doi: 10.1155/2020/8973739. PMID: 33343669; PMCID: PMC7732409.

[181] DOI: https://doi.org/10.1101/2020.06.03.20121392

[182] DOI: 10.13140/RG.2.2.29852.31367

p value anche dell'ordine di 10^{-5} – 10^{-3} per l'ipotesi nulla che nella prima ondata la diffusione del covid sia stata indipendente dall'abbondanza locale di maiali d'allevamento, e perfino, tentativamente, pur nell'incertezza dei dati, 0.005 per l'indipendenza dall'abbondanza locale di cinghiali, per zone in cui è stato possibile trovare allo stesso livello di suddivisione geografica dati sul covid e sulla densità di cinghiali.

In casi semplici il p-value può essere calcolato a mano, per esempio con i 5 lanci di moneta prima considerati, per rifiutare l'ipotesi (nulla) della regolarità, il risultato "0 teste o 5 teste" ha p-value 1/16 = 0.0625 e l'ipotesi non viene rifiutata al livello del 5% ovvero 0.05 ovvero del 95% ovvero 0.95. Il valore $\frac{1}{16}$ viene da

$$= P(T,T,T,T,T \vee C,C,C,C,C) =$$

$$= P(T,T,T,T,T) + P(C,C,C,C,C) =$$

probabilità di 5 successi o 0 successi, casi ugualmente estremi, e di più estremi di quanto ottenuto non ce n'è, con $B(5,\frac{1}{2})$:

$$\left(\frac{1}{2}\right)^5 + \left(\frac{1}{2}\right)^5 = \frac{1}{32} + \frac{1}{32} = \frac{1}{16}\,.$$

BOZZA - DRAFT

Oppure rifiutiamo l'ipotesi della regolarità, al classico livello debole del 10%, invece di quello standard del 5%.

55.7 Buona Statistica: ipotesi prima dell'esperimento

Si capisce che la buona Statistica si fa formulando l'ipotesi prima dell'esperimento, mentre di minor valore è fare affermazioni su dati già noti; sicuramente a partire dai risultati di 10mila lanci, guardandoli con attenzione, si potranno trovare affermazioni strampalate del tipo "i lanci con numero ordinale che scritto in tagalog iniziano per vocale tendono a dare testa", e magari si riuscirà pure a dimostrarle con buon p value; ma se l'affermazione fosse stata fatta prima dei lanci, ben improbabilmente l'esperimento l'avrebbe confermato.

Si provi a formulare qualche ipotesi del genere su questo campione di 100 lanci virtuali di moneta tratto da WolframAlpha, e si capirà quanto delicata sia la questione della statistica. (H = head = testa = 1, T = tail = croce = 0)

T | H | T | H | T | H | T | H | T | H | T | T | T | T | H | H | H | H | H | T | T | T | T | T | T | H | H | H | T | T | T | T | H | T | H | H | H | T | H | T | T | H | T | H | H | T | H | H | H | H | H | T | T | H | T | T | T | T | T | H | T | T | H | T | H | H | T | H | H | H | H | H | T | H | T | T | H | H | T | T | H | H | T | T | T | H | T | H | T | H | T | H | H | T | T | H | T | T | H | T

Ecco per esempio un'affermazione che risulterà statisticamente confermata sul campione del passato, ma in generale non su uno del futuro, e si provi per credere:

in tutti i lanci dal secondo all'undicesimo la differenza assoluta fra teste è croci uscite (a partire dal primo lancio) è sempre minore o uguale a 1.

BOZZA - DRAFT

Nello studio randomizzato controllato (Randomized controlled trial, RCT) l'ipotesi si formula prima. Per esempio che un vaccino

ridurrà i contagiati, oppure

ridurrà i casi sintomatici lievi, oppure

ridurrà i casi gravi, oppure

ridurrà i ricoveri in terapia intensiva, oppure

ridurrà la mortalià, oppure
ridurrà le ospedalizzazioni...

Non si può mica fare il *trial clinico* misurando tutte quelle cose, e poi facendo la statistica magari sull'unica andata bene, non dicendo nulla delle altre. Bisogna dire prima cosa si misurerà, e se tutto va bene si potrà respingere l'ipotesi nulla che non abbia funzionato per quell'*outcome*.

Per esempio per il famoso trial clinico del 2020 della Pfizer per ottenere l'autorizzazione in via emergenziale del vaccino anticovid, benchè il grande pubblico magari pensasse che il *main outcome* misurato fosse la morte (cioè che la statistica mostrasse una riduzione della mortalità da covid) in effetti si misurava la riduzione dei casi sintomatici lievi:

> Many may assume that successful phase 3 studies will mean we have a proven way of keeping people from getting very sick and dying from covid-19. And a robust way to interrupt viral transmission.
> Yet the current phase 3 trials are not actually set up to prove either, says Doshi.
> "None of the trials currently underway are designed to detect a reduction in any serious outcome such as hospitalisations, intensive care use, or deaths. Nor are the vaccines being studied to determine whether they can interrupt transmission of the virus," he writes.
> He explains that all ongoing phase 3 trials for which details have been released are evaluating mild, not severe, disease
> https://www.bmj.com/company/newsroom/covid-19-vaccine-trials-cannot-tell-us-if-they-will-save-lives/

55.8 Nota sul sentire statistico

La Statistica Inferenziale è poco insegnata a scuola, e ancor meno lo era nei decenni passati.
Per uno statistico, è sempre notevole constatare quanto poco presente sia la mentalità statistica a livello di popolazione italiana. Da una parte alcuni si attaccano al caso singolo senza rendersi conto di quanto poco rappresentativo sia della situazione complessiva – in ciò magistralmente teleguidati dai media al soldo dei loro referenti – e dall'altra qualcuno ti obietta che "4 casi non fanno Statistica", credono che ci vogliano sempre migliaia di casi, e non è così, anzi Fisher, "mostro sacro" della Statistica, nel suo pionieristico lavoro fece proprio la famosa brillante esperienza di Statistica con 4 casi; dipende da vari fattori. Certo è vero che neppure migliaia di casi mal scelti bastano a fare buona Statistica, e questo è un errore sempre in agguato nella Statistica Medica e in particolare in Farmacia: fare statistiche solo sui soggetti che portano avanti un esperimento fino alle nostre misurazioni finali; come se fosse la stessa cosa che dei, diciamo, 500, che non completano il trial, siano morti 1 (che magari sarebbe la norma in quel lasso di tempo senza prendere il nostro prodotto sperimentale) oppure 50, oppure siano tutti andati alle Seichelles a godersi la guarigione: i *drop out* sono una catastrofe nella Statistica Medica, e semplicemente non considerarli è del tutto fuorviante.

BOZZA - DRAFT

Questo testo vorrebbe giungere a dare un po' di *forma mentis* sta-

tistica.

BOZZA - DRAFT

56 Errori di I e II Specie

56.1 Due tipi di sfortune nei test statistici

Come detto, prima di eseguire un test statistico dobbiamo formulare un'ipotesi nulla H e la complementare ipotesi alternativa A.
Per esempio

$$H: \quad \mu \leq 0$$

$$A: \quad \mu > 0$$

dove μ potrebbe essere la media di una variabile aleatoria X di cui disporremo di un campione aleatorio $X_1, \ldots, X_n$. I ruoli delle 2 ipotesi non sono interscambiabili: in linea generale come ipotesi alternativa va fissata quella che speriamo vera (e come ipotesi nulla quella che ci avrebbe fatto perdere tempo).
Tratto ovvero prodotto ovvero rilevato il campione x_1,...,x_n, ne calcoliamo l'opportuna funzione che la statistica ci insegnerà a seconda del tipo di test, sia essa ora $g(x_1, .., x_n)$, per esempio la media $\bar{x}_n$, ci sono 4 casi relativamente alla regione critica D, a $g(x_1, .., x_n)$, e alla verità dell'ipotesi H:

BOZZA - DRAFT

1) $g(x_1, .., x_n) \in D$ ed è vera H:
male respingo ipotesi vera: errore di prima specie.
2) $g(x_1, .., x_n) \in D$ ed è vera A:
bene respingo ipotesi falsa: è il caso sperato. (α)
3) $g(x_1, .., x_n) \notin D$ ed è vera H:
non respingo ipotesi vera: ho perso tempo.
4) $g(x_1, .., x_n) \notin D$ ed è vera A:
male non respingo ipotesi falsa: errore di seconda specie.

Esempio di errore di seconda specie. L'errore di seconda specie è quello che si commise quella volta in cui si fece – per davvero – il test indicato in 55.4 trovando 13 femmine su 20 soggetti presi a caso, non riuscendo così a respingere l'ipotesi nulla (che volevamo respingere) che per un generico uditore della lezione fosse:

$$P(maschio) \geq P(femmina)$$

fatto comunque effettivamente falso, fattualmente: male non respingemmo ipotesi falsa, quella della minore o uguale probabilità del genere femminile. Che in aula ci fossero più femmine si vedeva conteggiando tutti i casi in aula, ma questo ovviamente non si sarebbe potuto fare nel caso, per esempio, fossimo stati interessati a dimostrare che in città, in strada, erano più numerose le gatte che i gatti.

Sulla gravità dell'errore di seconda specie.

L'errore di prima specie è considerato molto più grave.

Ad esempio per un farmaco si può mettere l'ipotesi che non curi, sperando di falsificarla con l'esperimento.

La cosa peggiore è diffondere a milioni di persone un farmaco che non cura, solo perché in un *trial clinico* – necessariamente molto più limitato – ha dato buona prova di sè, causando appunto l'errore di prima specie – di fatto sempre possibile.

Diffondere il farmaco a milioni o miliardi di persone, affermandolo migliore del farmaco precedente (o di nessun farmaco) è un fatto gravissimo, magari la gente poteva curarsi col farmaco precedente con migliori risultati.

ROMA – DRAFT

Leggiamo su Wikipedia, l'enciclopedia libera, in cui chiamano *null hypothesis H0* l'*ipotesi nulla* H_0, ma purtroppo *hypothesis* quella che in questa trattazione è chiamata *alternativa*:

> Hypothesis: "A patient's symptoms improve after treatment A more rapidly than after a placebo treatment."
>
> Null hypothesis (H0): "A patient's symptoms after treatment A are indistinguishable from a placebo."
>
> A Type I error would falsely indicate that treatment A is more effective than the placebo, whereas a Type II error would be a failure to demonstrate that treatment A is more effective than placebo even though it actually is

more effective.

Stabiliamo cosa intendiamo per “il farmaco funziona".
Per esempio
aumenta la sopravvivenza a 5 anni rispetto al placebo
riduce la massa tumorale
riduce la glicemia negli iperglicemici.

I 4 casi in dettaglio sono questi.

BOZZA - DRAFT

1) e 3) è vera l'ipotesi H: il farmaco è inutile o dannoso (esempio: sopravvivenza a 5 anni uguale o diminuita).
2) e 4) è vera l'alternativa A: il farmaco è utile (esempio: sopravvivenza a 5 anni aumentata).

1) e 2) La sperimentazione dà esito buono.

1) Per caso i soggetti trattati sono vissuti a lungo e il farmaco dannoso fa bella figura e magari si diffonde: GRAVISSIMO.

2) I soggetti trattati sono vissuti a lungo grazie al farmaco che giustamente fa bella figura

3) e 4) La sperimentazione dà esito cattivo.

3) non respingo l'ipotesi che il farmaco sia inutile o dannoso; il farmaco inutile o nocivo è stato correttamente riconosciuto tale sperabilmente non verrà commercializzato. Ho perso tempo, speravo fosse utile.

4) Per caso i soggetti trattati sono vissuti poco ma il farmaco in generale funziona: il farmaco utile viene purtroppo abbandonato. Peccato.

Gli errori di prima e seconda specie hanno un analogo molto suggestivo negli errori giudiziari. L'ipotesi nulla è l'innocenza:

	innocente	colpevole
condannato	errore di I specie	ottimo
assolto	bene (ma perso tempo)	errore di II specie

Sull'assoluzione del colpevole, errore meno grave della condanna di un innocente, vale il classico: IN DUBIO PRO REO.
(Ovviamente lo scopo dell'apparato giudiziario non è assolvere gli innocenti ma condannare i colpevoli: se facesse questo e non altro realizzerebbe il suo compito, per questo è scritto "perso tempo").

BOZZA - DRAFT

In Medicina e Farmacia, gli errori di I e II specie si presentano anche nella considerazione dei test diagnostici, **ma con una seria duplicità ovvero ambiguità.**

In ambito medico alcuni considerano
ipotesi nulla: la malattia è presente
altri
ipotesi nulla: malattia non presente.

Con riferimento al secondo modo di impostare la questione, si ha allora questo schema:

	sano	malato
positivo	falso positivo: errore di I specie *male respingo ipotesi vera*	vero positivo: ottimo: trovato!
negativo	vero negativo bene ma perso tempo	falso negativo: errore di II specie (condizione non rilevata)

BOZZA - DRAFT

In quest'ottica, il peggio sarebbe curare un sano, il falso positivo a un test diagnostico, a rischio di danneggiarlo: era già sano! PRIMUM NON NOCERE.
Nell'altra ottica il peggio, l'errore di I specie, sarebbe il falso negativo, in pratica non avvertire della malattia il malato.
Attenzione a quest'ambiguità.

56.2 Nota sul parametro medico su cui si fa il test

Per un chemioterapico si può ipotizzare di testare la sopravvivenza a 5 anni (confrontando farmaco e placebo); oppure la riduzione della massa tumorale in un breve tempo; se poi in un tempo doppio il paziente muore, questo non rientra nella statistica. È ovvio che la

sopravvivenza a 5 anni richiede una sperimentazione lunghissima, e costosa.
Ma per quanto di altissimo valore, anche quel solo parametro è poco, nella complessità della realtà: non si è tenuto conto della qualità della vita in seguito alla somministrazione del farmaco.

Essenzialmente si tratta di pre-decidere gli *outcomes* ricercati in un trial clinico, vedi https://en.wikipedia.org/wiki/Randomized_controlled_trial.

Tutt'altra questione: non si è tenuto conto neppure del costo economico: gli stessi soldi il Sistema Sanitario potrebbe spenderli con maggiore beneficio complessivo per la cura di altre malattie. (Resource Management, e in caso di sanità privata anche ROI, Return On Investment).

BOZZA - DRAFT

A questo proposito, su cosa si testa, leggiamo sul British Medical Journal, in https://www.bmj.com/company/newsroom/covid-19-vaccine-trials-cannot-tell-us-if-they-will-save-lives/, datato 21 ottobre 2020, e si noti che qua la prestigiosissima rivista scientifica tende a sostenere un punto di vista scientifico, non *differenti* interessi:

> None of the current trials are designed to detect a reduction in any serious outcome such as hospitalisations, intensive care use, or deaths
>
> Vaccines are being hailed as the solution to the covid-19 pandemic, but the vaccine trials currently underway are not designed to tell us if they will save lives, reports Peter Doshi, Associate Editor at The BMJ today.
>
> Several covid-19 vaccine trials are now in their most advanced (phase 3) stage, but what will it mean exactly when a vaccine is declared "effective"?
>
> Many may assume that successful phase 3 studies will

> mean we have a proven way of keeping people from getting very sick and dying from covid-19. And a robust way to interrupt viral transmission.
>
> Yet the current phase 3 trials are not actually set up to prove either, says Doshi.
>
> "None of the trials currently underway are designed to detect a reduction in any serious outcome such as hospitalisations, intensive care use, or deaths. Nor are the vaccines being studied to determine whether they can interrupt transmission of the virus," he writes.[183]

BOZZA - DRAFT

[183] E continua:

He explains that all ongoing phase 3 trials for which details have been released are evaluating mild, not severe, disease - and they will be able to report final results once around 150 participants develop symptoms.

In Pfizer and Moderna's trials, for example, individuals with only a cough and positive lab test would bring those trials one event closer to their completion.

Yet Doshi argues that vaccine manufacturers have done little to dispel the notion that severe covid-19 was what was being assessed.

Moderna, for example, called hospitalisations a "key secondary endpoint" in statements to the media. But Tal Zaks, Chief Medical Officer at Moderna, told The BMJ that their trial lacks adequate statistical power to assess that endpoint.

Part of the reason may be numbers, says Doshi. Because most people with symptomatic covid-19 infections experience only mild symptoms, even trials involving 30,000 or more patients would turn up relatively few cases of severe disease.

"Hospitalisations and deaths from covid-19 are simply too uncommon in the population being studied for an effective vaccine to demonstrate statistically significant differences in a trial of 30,000 people," he adds. "The same is true regarding whether it can save lives or prevent transmission: the trials are not designed to find out."

Zaks confirms that Moderna's trial will not demonstrate prevention of hospitalisation because the size and duration of the trial would need to be vastly increased to collect the necessary data. "Neither of these I think are acceptable in the current public need for knowing expeditiously that a vaccine works," he told The BMJ.

Moderna's trial is designed to find out if the vaccine can prevent covid-19 disease, says Zaks. Like Pfizer and Johnson and Johnson, Moderna has designed its study to detect a relative risk reduction of at least 30% in participants developing lab-confirmed covid-19, consistent with FDA and international guidance.

Zaks also points to influenza vaccines, saying they protect against severe disease better than mild disease. "To Moderna, it's the same for covid-19: if their vaccine is shown to reduce symptomatic covid-19, they will feel confident it also protects against serious outcomes," Doshi writes.

But Doshi raises another important issue - that few or perhaps none of the current vaccine trials appear to be designed to find out whether there is a benefit in the elderly, despite their obvious vulnerability to covid-19.

If the frail elderly are not enrolled into vaccine trials in sufficient numbers to determine whether there is a reduction in cases in this population, "there can be little basis for assuming

D'altra parte il testo sopra riportato è ormai (2021) superato, e in progresso di tempo molto è stato fatto per valutare i risultati mancanti negli studi iniziali, e ormai noi tutti sappiamo che i vaccini salvano delle vite.

56.3 Dopo gli errori di I e II specie, le assurdità

Gli errori di I e II specie sono essenzialmente sfortune, e, in prima approssimazione, non ha colpa chi fa il test: che colpa potrebbe avere se ha giudicato plausibilmente non regolare (rifiutando l'ipotesi di regolarità) una moneta pur regolare che però ha dato 19 teste su 20 lanci? Caso sfortunato. A un livello superiore rispetto a questa trattazione, comunque, si stabilisce come ridurre il rischio di quegli errori, dopodichè commetterli è parzialmente colpa del ricercatore.

BOZZA - DRAFT

Invece, bisogna fare attenzione – prevenire è meglio che curare – a veri e propri errori che può facilmente fare il non statistico addentro alle cose biomediche.

Supponiamo che 2 test statistici ben fatti (magari ciascuno è addirittura uno studio controllato randomizzato, ovvero studio clinico controllato randomizzato – RCT, dall'inglese randomized controlled trial), trovino con p value piccolissimi, che i soggetti con una certa caratteristica rimovibile, per esempio

- uso di eroina
- abbonamento al mensile *Il cacciatore di gatti con la fionda*

hanno maggior probabilità di fare un incidente stradale (per esempio nell'anno seguente), rispetto a chi manca di quella caratteristica.

---

any benefit against hospitalisation or mortality," he warns.

Doshi says that we still have time to advocate for changes to ensure the ongoing trials address the questions that most need answering.

For example, why children, immunocompromised people, and pregnant women have largely been excluded; whether the right primary endpoint has been chosen; whether safety is being adequately evaluated; and whether gaps in our understanding of how our immune system responds to covid-19 are being addressed.

"The covid-19 vaccine trials may not have been designed with our input, but it is not too late to have our say and adjust their course. With stakes this high, we need all eyes on deck," he argues.

Il profano di Statistica potrebbe pensare che, visto che i p value sono piccolissimi, forte indizio di non casualità delle correlazioni trovate, eliminare dalle persone le 2 caratteristiche rimovibili, l'eroina e l'abbonamento, riduca loro la probabilità fare un incidente stradale nell'anno successivo.
E infatti se riesce a togliergli l'eroina probabilmente gli ridurrà la probabilità di incidente e anche di morte nell'anno successivo (non ne siamo certi ma è verosimile).
Revocandogli l'abbonamento, però, probabilmente non otterrà molto.
Questo succede, molto approssimativamente, perchè la correlazione è cosa ben diversa dalla causalità, e il p value piccolissimo rende molto plausibile solo la correlazione delle vere variabili aleatorie retrostanti (abbonamento, una variabile aleatoria bernoulliana $X \sim B(1, p_1)$, e incidente, un'altra variabile aleatoria bernoulliana $Y \sim B(1, p_2)$) rendendo poco plausibile la casualità (dei valori empirici rilevati). È verosimile che gli abbonati a quel mensile siano molto meno intelligenti della media, ed è questo che causa la loro maggiore mortalità per incidenti stradali, non l'abbonamento. Revocato quello, non è che diventano subito molto più intelligenti; e continuano a fare incidenti.

BOZZA - DRAFT

Naturalmente sono possibili errori più sottili, proprio di tipo medico e non statistico, nell'interpretazione dei risultati degli esperimenti medici. Enfasi aggiunta:

> L'eroina(...) Venne risintetizzata nel 1897 da Felix Hoffmann, un chimico tedesco che lavorava per la Bayer; in quel periodo l'acetilazione era una tendenza diffusa per la ricerca di molecole più attive. Hoffmann realizzò l'acetilazione dell'acido salicilico, ottenendo l'Aspirina, e solamente 11 giorni dopo fece altrettanto con la morfina, producendo appunto l'eroina.
> L'intento era quello di ottenere una molecola più efficace della codeina nel sedare la tosse, la tubercolosi e le patologie respiratorie. Le effettive proprietà sedative sul centro del respiro (le stesse che portano a morte nell'overdose) furono inizialmente male interpretate, ritenendo che la riduzione del ritmo respiratorio dipendesse da una migliorata efficienza respiratoria.
> Fu battezzata commercialmente eroina (dal tedesco "heroisch", "eroico", giacché inizialmente la si credeva priva degli spiacevoli effetti collaterali di dipendenza e assuefazione palesati dalla morfina) e cominciò a essere venduta liberamente dalla multinazionale farmaceutica Bayer dal 1899. In breve tempo l'impiego terapeutico si ampliò alle più disparate patologie pneumologiche, ma anche neurologiche, ginecologiche, o a semplici dolori; si diffusero pertanto svariate preparazioni farmaceutiche acquistabili liberamente, e questo fece sì che l'eroina divenisse velocemente uno dei farmaci più venduti in assoluto.

> A parte qualche farmacologo controcorrente, non la si riteneva in grado di dare dipendenza.
> https://it.wikipedia.org/wiki/Eroina letto il 19 dicembre 2022.

Secondo questo testo – qua non possiamo discutere quanto esatto o inesatto – la riduzione del ritmo respiratorio (constatata senz'altro in un gran numero di casi) non era causata da una migliorata efficienza respiratoria.
E nello stesso sopraripportato testo si noti che si ripropone la questione di cosa misuriamo in un trial clinico: solo la riduzione del dolore, in cui magari l'eroina riusciva bene, o anche l'aumento di mortalità – e comunque, in quali tempi? – o pure la dipendenza indotta?
Si noti anche l'ultima frase: "A parte qualche farmacologo controcorrente, non la si riteneva in grado di dare dipendenza". Un classico: *prima ti ignorano, poi ti deridono, poi ti combattono, infine dicono che loro lo hanno sempre detto.*

Nota. Di questa Lezione ci si dovrebbe aspettare che lo studente conosca a memoria la formula (α) oltre a tutto l'inquadramento teorico della questione.

BOZZA DRAFT

ESERCIZI SULLA LEZIONE 56

ESERCIZIO$_{\mu 2018}$
∗ Supponiamo che per un test statistico, con ipotesi (nulla) H e alternativa A, ad un certo livello α, la regione critica sia $[20.18, +\infty[$ e lo stimatore $T := g(X_1, ..., Xn)$ relativo al test abbia prodotto il valore 19.2, e che sia vera A. Quale delle seguenti affermazioni è vera?

- Si commette un errore di prima specie
- Era il caso in generale sperato
- Si commette un errore di seconda specie
- Si è sostanzialmente perso tempo
- Non è possibile rispondere perché non è specificato il quantile

SVOLGIMENTO
Lo stimatore T vale 19.2 che $\notin$ alla regione critica, e l'ipotesi (nulla) è falsa (perché è vera l'alternativa). Allora "male non respingo ipotesi falsa", cioè

Si commette un errore di seconda specie

BOZZA - DRAFT

ESERCIZIO$_{\mu 2018}$
∗ Supponiamo che per un test statistico, con ipotesi (nulla) H, e alternativa A vera, al livello $\alpha = 0.1$, la regione critica sia $T > 734.66$ e il calcolo dello stimatore del test dia $T = g(x_1, ..., x_n) = 786.45$. Quale delle seguenti affermazioni è vera?

- Si commette un errore di prima specie
- Era il caso in generale sperato
- Si commette un errore di seconda specie
- Si è sostanzialmente perso tempo
- Non ha senso perché α deve essere ≤ 0.05 ossia 5%.

SVOLGIMENTO
Lo stimatore cade nella regione critica, perché $786.45 > 734.66$, e allora l'ipotesi (nulla) viene respinta, ed essa è falsa perché l'alternativa è vera. Siamo nel caso *"bene respingo ipotesi falsa"* che, come è ben noto, per i test statistici

Era il caso in generale sperato

Nota. Di questa Lezione ci si dovrebbe aspettare che lo studente conosca a memoria la formula (α) oltre a tutto l'inquadramento teorico della questione.

57 Test del χ^2 di indipendenza

Troviamo in un articolo[184] scientifico 2 test del chi quadrato, relativi alle stesse 134 persone ma stratificate con 2 diverse soglie per la carenza di vitamina D:

> Our study demonstrates that only severe 25OHD deficiency at the beginning of COVID19 infection was associated with an increased subsequent risk of developing an overt infection

(semplificatamente: avere poca vitamina D non risulta molto associato alla malattia ma averne proprio pochissima sì) e dal medesimo articolo scientifico fotografiamo:

Table 2. Distribution analysis of subsequent cases of COVID-19 infections in healthy subjects during 12 months observation period, stratified according to vitamin D (25OHD) concentrations above or below 12 ng/mL (30 nmol/L) in spring 2020. The difference is statistically significant, $p = 0.0063$, χ^2 test.

COVID-19 Infection	25OHD < 12 ng/mL		
	NO	YES	
NO	89 (for > 12 ng/mL)	14 (for < 12 ng/mL)	103 (76.9%)
YES	20 (for > 12 ng/mL)	11 (for < 12 ng/mL)	31 (23.1%)
Total	109	25	134 (100%)

Table 3. Distribution analysis of subsequent cases of COVID-19 infections during 12 months observation period in healthy subjects, stratified according to vitamin D (25OHD) concentrations above or below 20 ng/mL (50 nmol/L) in spring 2020. The difference was not statistically significant, $p = 0.2787$, χ^2 test.

COVID-19 Infection	25OHD < 20 ng/mL		
	NO	YES	
NO	41 (for > 20 ng/mL)	62 (for < 20 ng/mL)	103 (76.9%)
YES	9 (for > 20 ng/mL)	22 (for < 20 ng/mL)	31 (23.1%)
Total	50	84	134 (100%)

57.1 Inquadramento della questione

In questa trattazione elementare, verrà considerato solo il caso più semplice, con 4 valori/dati. Fissiamo subito le idee su un esempio.

[184] Basińska-Lewandowska M, Lewandowski K, Horzelski W, Lewiński A, Skowrońska-Jóźwiak E. Frequency of COVID-19 Infection as a Function of Vitamin D Levels. Nutrients. 2023 Mar 24;15(7):1581. doi: 10.3390/nu15071581. PMID: 37049423; PMCID: PMC10097275.

Supponiamo di avere 280 monete uguali di sconosciuta probabilità di dare testa, e che a 50 di esse venga fatta meccanicamente una piccola incisione al centro sulla faccia della testa. Le monete trattate hanno tutte un'uguale sconosciuta nuova probabilità di dare testa. (Di maggiore interesse potrebbe essere il trattamento con un farmaco, e l'eventuale morte a 5 anni, ma le monete e i dadi restano modelli generali ottimali, anche perché da essi *a priori ci aspettiamo* di default un comportamento casuale, mentre per un farmaco saremmo portati a supporre causalità – contenendo la tal molecola *dovrebbe* fare un certo effetto – eventualmente inesistenti: nella *statistica medica* dobbiamo appunto verificare se funziona, *contando i morti*, per così dire). Lanciamo una volta ciascuna delle 280 monete e contiamo le teste nelle 2 classi. Supponiamo di avere questa situazione (risultato=viene 1)

	non risultato	risultato
non trattati	188	42
trattati	43	7

BOZZA - DRAFT

Naturalmente $43 + 7$ sono le 50 monete trattate, e $188 + 42$ sono le rimanenti 230, non trattate. Il trattamento ha diminuito la probabilità di dare testa? Verrebbe da dire di sì, perché fra i trattati la frequenza relativa di 1 è $7/50 = 0.14$ mentre fra i non trattati è $42/230 = 0.183$ ma dobbiamo ragionevolmente escludere un effetto casuale: tale miglioramento potrebbe apparire per caso anche se a nostra insaputa il tecnico preposto non avesse affatto compiuto l'incisione dichiarata.
Abbiamo una variabile aleatoria X con 2 valori, "trattate" e "non trattate", e una variabile aleatoria Y con 2 valori, "testa" e "croce". Il test statistico è definito da

$$H : \text{indipendenza} \qquad A : \text{non indipendenza}$$

(vorremmo escludere l'indipendenza, cioè che l'incisione non abbia prodotto risultato).

Come in tutti gli altri test statistici che consideriamo, si fisserà un livello di confidenza α, tipicamente 0.95 = 95%, ovvero un livello di significatività $1-\alpha$, tipicamente 0.05 = 5%.
(Si ritiene che per correttezza scientifica ovvero deontologica, questo valore vada fissato *prima* di fare i calcoli).

Si calcolerà uno *stimatore* (ovvero *statistica*) T, dalla formula alquanto complessa, e lo si confronterà col quantile $\chi^2_{1-\alpha}(k)$ del chi quadrato con

$$k := (m-1)\cdot(n-1) = (\#righe - 1) \times (\#colonne - 1) \tag{157}$$

gradi di libertà, (qua 1 perchè abbiamo 2 righe e 2 colonne) e se T supera il quantile

$$T > \chi^2_{1-\alpha}(k) \tag{158}$$

si respinge l'ipotesi, altrimenti non la si respinge (che alcuni dicono "si accetta"), al livello di significatività $1-\alpha$ ovvero di confidenza α.

Con le tabelle 2×2, all'usuale livello del 95%, il quantile è ≈ 3.84.
Il calcolo di T invece è più complesso e i vari Autori non concordano sulla formula, che per noi sarà la (159).

BOZZA - DRAFT

57.2 Semplice statistica per trial clinico

In pratica per la tabella 2×2 (*tetracorica*) a 4 valori a, b, c, d, ultrasemplificando in *guariti* e *non guariti* i possibili esiti clinici, ecco il calcolo da fare per

	guariti	non guariti
farmaco 1	a	b
farmaco 2	c	d

al livello 0.05 ovvero 0.95 rifiuto l'ipotesi nulla
H: il farmaco 1 equivale al farmaco 2

se (159)

$$T := \frac{(a\,d-b\,c)^2\,(a+b+c+d)}{(a+b)(a+c)(b+d)(c+d)} > 3.84$$

(Altri Autori usano una formula leggermente diversa e più complessa con la *correzione di Yates*).

(Si tratta ovviamente dell'equivalenza dei farmaci dal solo punto di vista delle guarigioni, non del costo o di un'eventuale tossicità a lunga scadenza, e la guarigione andrà ben definita, in particolare nella sua determinazione temporale: lo stesso farmaco che elimina un sintomo in un giorno potrebbe uccidere in un mese, si ripensi all'eroina della Bayer).

BOZZA - DRAFT

Naturalmente il valore 3.84 è il quantile del chi quadrato (a 1 grado di libertà com'è dovuto per la tabella 2×2) al livello 0.05; e si potrà modificarlo per livelli diversi, in particolare 6.63 al livello 0.01, molto più probativo; si veda la tavola del chi quadrato in 58.4.

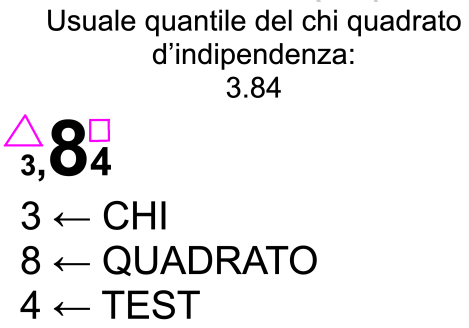


Figure 61: Mnemonico per il numero 3.84: contare le lettere delle parole.

Esempio 1. Per un ipotetico antibiotico per cani, coi valori

	guariti	non guariti
xmicina	52	10
streptomicina	40	21

in http://www.quadernodiepidemiologia.it/epi/assoc/chi_qua.htm trovano $T = 5.46 > 3.84$ e concludono

> consente di ritenere che la differenza fra i due gruppi sia significativa al livello di probabilità 5% ma non al livello di probabilità 1%.

E infatti immettendo i dati per esempio in https://www.socscistatistics.com/tests/chisquare/default2.aspx, dove c'è un calcolatore online del chi quadrato di indipendenza, troviamo che

`chi-square statistic is 5.4607`

e che il valore discriminante fra il rifiuto e il non rifiuto cioè il

p`-value is .019449. Significant at p < .05.` E si noti che 0.019 è sì $< 5\%$ ma non $< 1\%$.

Naturalmente il valore 5.46 si ottiene con la (159), anche con la calcolatrice.

L'intero calcolo può essere fatto online gratuitamente anche per tabelle più grandi in altri siti web, oltreché con vari software statistici usati nelle Scienze Applicate.

BOZZA - DRAFT

Esempio 2. In http://www.biostatistica.unich.it/mat_didattica/Prof_di_nicola/3-test%20del%20chi%20quadrato.pdf invece che un farmaco e la possibile guarigione, vengono considerati un intossicante e una possibile malattia, e ovviamente la teoria è la stessa:

	bronchite	non bronchite
fumo	120	280
non fumo	93	527

$$T = \frac{(120 \cdot 527 - 93 \cdot 280)^2\, 1020}{213 \cdot 807 \cdot 400 \cdot 620} \approx 32.21$$

che supera non solo il quantile 3.84 al 95% ma anche quello 6.63 al 99% e in effetti di molto, e quindi respinge l'ipotesi (nulla) di indipendenza del fumo dalla bronchite. Di per sè quest'affermazione non dice se il fumo sembra favorire o sfavorire la bronchite, ma un'attenta analisi dei dati ci porta a decidere per il favorire.

57.3 Sulle condizioni per la validità del test

Varie condizioni sulla non eccessiva “capricciosità" dei 4 valori vengono fatte dai vari Autori, per garantire la bontà della comunque inevitabile approssimazione del test. La questione è comunque contemporaneamente complessa e largamente arbitraria, e non la tratteremo. Confidiamo piuttosto che il software che il ricercatore poi userà nella pratica, dia un suo avvertimento in casi dubbi.

57.4 Test di Fischer esatto

Si può evitare l'approssimazione che è intrinseca nel test del chi quadrato d'indipendenza col *Test di Fischer esatto*, che ha una formula molto più complicata che viene comunque calcolata in un'istante dai computer, a meno che i numeri non siano esorbitanti.

BOZZA - DRAFT

Il test di Fischer esatto è migliore del test del chi quadrato, che tuttavia è in generale ancora preferito *per tradizione*.

57.5 Morti per tutte le cause

E certo, una terza colonna, con i morti per tutte le cause, *all causes*, sarebbe utilissima per confronto. Per esempio così:

	malattia	non malattia	morti per tutte le cause
vaccino	a	b	c
placebo	d	e	f

Su trial clinici grandi e sufficientemente prolungati nel tempo, i morti ci saranno inevitabilmente, in entrambe le classi: come vago ordine di grandezza, si consideri che in un anno muore circa l'1% di un insieme di persone prese a caso in Stati sufficientemente sviluppati, il che corrisponde a circa 1 morto ogni 1000 persone in 1 mese di trial clinico. Leggiamo in
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8461570/

nell'articolo scientifico[185] *Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine through 6 Months* (2021) sul ciclopico trial clinico di lungo periodo (si arrivò a 6 mesi e poi lo sospesero – dissero – *per motivi etici*) del celeberrimo vaccino covid (2020) della Pfizer, il BNT162b2:

> During the blinded, placebo-controlled period, 15 participants in the BNT162b2 group and 14 in the placebo group died

57.6 Il solito problema: cosa misuriamo? Placebo

Con quale tipo di placebo fare il confronto?
In passato erano comuni trial clinici con placebo tutt'altro che inerti, per esempio un altro vaccino o una soluzione di idrossido di alluminio testando il vaccino HPV. Qua possiamo solo rimarcare lo sconcerto di taluni per questa pratica, senza poter approfondire. Ma nei recenti (2020) trial per i vaccini contro il covid-19 si è salutato con compiacimento l'uso della vera soluzione salina come placebo.

BOZZA - DRAFT

57.7 Trucchi del mestiere

E certo sarebbe una cosa carina che il doppio cieco non venga beffato con pillole... di diversi colori (vedi https://www.theguardian.com/business/2014/apr/10/tamiflu-saga-drug-trials-big-pharma

Ma non ci si illuda troppo. Nel trial clinico che ha portato all'autorizzazione in via emergenziale del vaccino contro il covid-19, la Pfizer ha presentato un testo impresentabile per uno statistico (eppure è stato

[185] Thomas SJ, Moreira ED Jr, Kitchin N, Absalon J, Gurtman A, Lockhart S, Perez JL, Pérez Marc G, Polack FP, Zerbini C, Bailey R, Swanson KA, Xu X, Roychoudhury S, Koury K, Bouguermouh S, Kalina WV, Cooper D, Frenck RW Jr, Hammitt LL, Türeci Ö, Nell H, Schaefer A, Ünal S, Yang Q, Liberator P, Tresnan DB, Mather S, Dormitzer PR, Şahin U, Gruber WC, Jansen KU; C4591001 Clinical Trial Group. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine through 6 Months. N Engl J Med. 2021 Nov 4;385(19):1761-1773. doi: 10.1056/NEJMoa2110345. Epub 2021 Sep 15. PMID: 34525277; PMCID: PMC8461570.

presentato e approvato) in cui il numero di *exclusions* (casi esclusi per un qualche motivo: uno dei grossi problemi della Statistica nel mondo reale) era quintuplo nel gruppo vaccino rispetto al gruppo placebo, senza dare alcuna spiegazione plausibile per questo fatto, inverosimile dal punto di vista statistico: se veramente chi somministrava vaccini e placebo non sapeva, le exclusions avrebbero dovuto trovarsi bilanciatamente nei due gruppi. Si veda sul British Medical Journal https://blogs.bmj.com/bmj/2021/01/04/peter-doshi-pfizer-and-modernas-95-effective-vaccines-we-need-more-

> What is concerning is the imbalance between randomized groups in the number of these excluded individuals: 311 from the vaccine group vs 60 on placebo.

Ma le manipolazioni statistiche andavano ben oltre, si veda https://blogs.bmj.com/bmj/2021/02/05/clarification-pfizer-and-modernas-95-ef riducendosi la dichiarata efficacia del 95% a 2 possibilità

> calculations of 19% and 29% vaccine efficacy

Ma poi, in definitiva, è saltato fuori che c'erano taroccamenti sui dati stessi: https://www.bmj.com/content/375/bmj.n2635, per esempio

> vaccine packaging materials with trial participants' identification numbers written on them left out in the open, potentially unblinding participants

L'importante, comunque, è che il vaccino previene il contagio, la contagiosità, la malattia, la malattia severa, la terapia intensiva, e la morte, nel mondo reale, al di là del trial clinico iniziale: https://www.adnkronos.com/vaccini-moderna-e-pfizer-protezione-per-anni-dal-cov 3IkcynLUR4dzdvJPbVXvyL

E, per buona misura, che la ricercatrice che ha documentato fotograficamente i taroccamenti sia stata licenziata. Queste cose non s'han da fare, perchè potrebbero minare la *fede nella Scienza*.

BOZZA DRAFT

57.8 Duplice esempio da un articolo scientifico

In Medicina c'è una diatriba di lungo corso sull'ipotetica correlazione fra vermi intestinali e bruxismo (digrignamento dei denti nel sonno, in particolare nei bambini.

Un articolo scientifico[186] considera bambini di 3-6 anni e in 50 con bruxismo e 50 senza trova

> Parasitic infections were observed in 19 (11/50 cases and 8/50 controls) children

cioè

	bruxism	without
intestinal parasitic infection	11	8
without	39	42

BOZZA - DRAFT

e conclude

> The frequency of parasitic infections does not indicate any significantly greater frequency rate in the case group.

Verifichiamolo con la Formula (159) del test del chi quadrato d'indipendenza:

$$T := \frac{(a\,d - b\,c)^2\,(a+b+c+d)}{(a+b)(a+c)(b+d)(c+d)} =$$

$$= \frac{(11\cdot 42 - 8\cdot 39)^2\,(11+8+39+42)}{(11+8)(11+39)(8+42)(39+42)} =$$

$$= \frac{(462-312)^2\cdot 100}{19\cdot 50\cdot 50\cdot 81} =$$

[186] Tehrani MH, Pestechian N, Yousefi H, Sekhavati H, Attarzadeh H. The Correlation between Intestinal Parasitic Infections and Bruxism among 3-6 Year-Old Children in Isfahan. Dent Res J (Isfahan). 2010 Summer;7(2):51-5. PMID: 22013457; PMCID: PMC3177368.

$$= \frac{150^2 \cdot 100}{3\,847\,500} = \\ = \frac{22\,500}{38\,475} \tag{160}$$

che non è > 3.84.
Potrebbero concludere così ma forse vogliono proprio dimostrare che il bruxismo è collegato ai parassiti, almeno a quelli più cattivi; distinguono i soggetti coi parassiti in soggetti con parassiti patogenici e soggetti con parassiti non patogenici:

> However, since all the 11/50 of the detected infections in the case group included pathogenic parasites while from the 8/50 parasitic infections found in the controls, only 3/50 involved pathogenic infections, a statistically significant relationship can be established between parasitic infection and bruxism.

BOZZA - DRAFT

Cioè

	bruxism	without
intestinal **pathogenic** parasitic infection	11	3
without	39	47

Verifichiamolo con la Formula (159) del test del chi quadrato d'indipendenza:

$$T =:= \frac{(a\,d - b\,c)^2\,(a+b+c+d)}{(a+b)(a+c)(b+d)(c+d)} = \\ = \frac{(11 \cdot 47 - 3 \cdot 39)^2\,(11+3+39+47)}{(11+3)(11+39)(3+47)(39+47)} = \\ = \ldots \approx 5.316. > 3.84 \tag{161}$$

Perciò nell'abstract possono scrivere, con quasi bizantina sottigliezza.

> Comparison for the frequency distribution of intestinal parasitic infections between the two groups was performed using the chi-square test (...) Parasitic infections were observed in 19 (11/50 cases and 8/50 controls) children. A statistically significant relationship was observed between infection with pathogenic parasites and bruxism ($P < 0.05$).

(Apparentemente sbagliato, vedi (160), ma è giusto, vedi (161), perché nella prima statistica considera i parassiti, nella seconda i parassiti patogenici).

Si faccia attenzione alle singole parole, gli Scienziati quando vogliono bizantineggiano.

Nota. Di questa Lezione ci si dovrebbe aspettare che lo studente conosca a memoria

l'unica

formula *contemporaneamente* numerata *et* riquadrata, oltre a tutto l'inquadramento teorico della questione.

BOZZA - DRAFT

58 Il Test del χ^2, quello basico

58.1 premessa sul nome del test

Il primo test del chi quadrato, dei due che considereremo, lo chiameremo semplicemente *Test del chi quadrato (quello basico).*
Un altro test detto del chi quadrato verrà illustrato nella Lezione successiva.

58.2 Premessa sul livello del test

Come si è detto c'è una notevole ambiguità notazionale nei testi di Statistica, e in particolare
- certi Autori chiamano α il livello di significatività, "piccolo", in generale 0.05,
- altri Autori chiamano α il livello di confidenza, "grande", in generale 0.95, il complemento a 1 del numero precedente.

Per abituare il lettore a entrambe le eventualità

in questa lezione useremo lo standard "dell'α grande".

(Il classico libro di Marco Abate, *Matematica e statistica. Le basi per le scienze della vita*, nel capitolo sul test del chi quadrato usa l'α piccolo e lo chiama *livello di affidabilità*).

BOZZA DRAFT

58.3 La teoria del Test del χ^2, quello basico

Un problema tipo:

> è possibile confutare con ragionevolezza statistica che un dado regolare lanciato 2000 volte abbia dato i risultati 1, 2, 3, 4, 5, 6 rispettivamente
> 388, 322, 314, 316, 344, 316
> volte? (È considerato sul classico libro di Paolo Baldi). Naturalmete, in Farmacia, potremmo avere 2000 campioni biologici, e in ciascuno di essi cercare una variante di 6 geni, e ci chiediamo se le 6 varianti sono equiprobabili.
> Ma si faccia attenzione che talvolta il test viene fatto relativamente a una assegnata densità non uniforme.

Questo test del chi quadrato basico si riferisce a una variabile aleatoria discreta, sia essa X, con un numero finito k di valori che possiamo identificare con i primi k numeri naturali oppure naturali positivi, per esempio 0,1 oppure 1,...,6; per fissare le idee diciamo, in generale, 1,...,k, cioè iniziamo da 1. Nel discorso che segue si possono immaginare, per fissare le idee, $n = 600$ lanci di un dado. Certo, se è regolare possiamo aspettarci circa 100 volte il risultato 1, e il 2, eccetera.

Se abbiamo n determinazioni $x_1,...,x_n$ di X, ce ne sono n_1 che valgono 1, n_2 che valgono 2,..., n_k che valgono k. Sono le *"frequenze assolute (osservate ovvero empiriche)"*. (Dividendole per n si ottengono le *"frequenze relative (osservate ovvero empiriche)"* $\bar{p}_i := \frac{n_i}{n}$, che tendenzialmente dovrebbero assomigliare alle p_i, le probabilità *"teoriche"*, e con esse si ottiene la formula alternativa che daremo in parentesi).

Il test vuole respingere l'ipotesi (nulla) che la densità di X sia data da certi valori $(p_1, ..., p_k)$ assegnati. Per esempio escludere la regolarità di un dado, cioè $H : (p_1, ..., p_6) = (\frac{1}{6}, ..., \frac{1}{6})$, $A : (p_1, ..., p_6) \neq (\frac{1}{6}, ..., \frac{1}{6})$. (Naturalmente gli indici potrebbero essere 0,...,$k-1$).

Fissato *prima* di vedere i dati (come sempre si deve fare in un test statistico serio) il livello di confidenza α, tipicamente 0.95, (in questa Lezione α è "grande") il rifiuto si ha se

$$T_n := \sum_{i=1}^{k} \frac{(n_i - n\,p_i)^2}{n\,p_i} > \chi^2_{1-\alpha}(k-1) \quad et \quad n\,p_i \geq 5 \quad \text{per } i = 1, ..., k \tag{162}$$

e formula alternativa ed equivalente alla (162) è

$$n \sum_{i=1}^{k} \frac{(\bar{p}_i - p_i)^2}{p_i} \quad et \quad n\,p_i \geq 5 \quad \text{per } i = 1, ..., k. \tag{163}$$

Se con l'ultima classe si ha $n\,p_k < 5$, si associano le classi $k-1$ e k, e se col nuovo $\tilde{p}_{k-1} = p_{k-1} + p_k$ si ha $n\,\tilde{p}_{k-1} \geq 5$ si fa il test per

BOZZA - DRAFT

la nuova v.a. a $k-1$ valori. Per esempio si considera un dado "a 5 valori": 1, 2, 3, 4, $5 \vee 6$, con frequenze assolute n_1, n_2, n_3, n_4, n_5+n_6, e nell'ipotesi p_1, p_2, p_3, p_4, p_5+p_6. Similmente se la prima, o più raramente un'altra probabilità, è troppo piccola, $n\,p_m < 5$, si può provare ad associare 2 classi consecutive.

Alcuni Autori richiedono la condizione > 10 invece che ≥ 5.

Esempio di Wikipedia tratto dal già citato libro di Paolo Baldi. Con 2000 lanci di un dado con frequenze assolute osservate 388, 322, 314, 316, 344, 316, si rifiuta, al livello del 5%, ovvero del 95%, la regolarità perché

$$T \approx 12.6 > 11.07 \approx \chi^2_{0.05}(5).$$

(Naturalmente T è T_{2000}).

BOZZA - DRAFT

58.4 Le tavole del chi quadrato

Tavole numeriche dei quantili del chi quadrato si trovano sia in formato cartaceo su molti libri, sia online cercando `chi square table`.

Purtroppo c'è confusione nelle notazioni delle tavole numeriche, con il serio rischio di confondere nella pratica α con $1-\alpha$ usando le tavole. Soprattutto di confondere le colonne relative a 0.95 e 0.05. (E similmente 0.9 con 0.1, e 0.99 con 0.01).
Il quantile q_α di parametro n di solito viene denotato $\chi^2_{1-\alpha}(n)$. In una tavola numerica, si cerchi il valore ≈ 19.68, quantile di ordine $\alpha = 0.95$ relativo al parametro $n = 11$ (in certe tavole denotato r, ν...), **undicesima riga della tavola**, cioè fissiamo l'attenzione su $X \sim \chi^2(11)$:

– se *Link1* 19.68 è l'unicesimo valore della colonna 0.95, intendete

$P(X \leq 19.68) \approx 0.95$ che è l'α di questa Lezione: $q_\alpha = \chi^2_{1-\alpha}(n)$

$$P(X \leq q_\alpha) = \alpha$$

per esempio $q_{0.95} = \chi^2_{0.05}(n)$

(ma si noti che nella notazione con la q non è specificata la dipendenza da n)

e in particolare per $n = 11 \quad q_{0.95} = \chi^2_{0.05}(11) \approx 19.68$

BOZZA - DRAFT

– se *Link2* 19.68 è l'unicesimo valore della colonna 0.05, intendete

$P(X \geq 19.68) \approx 0.05$ (del tutto equivalente a quella sopra) e bisogna calcolare 1– per avere l'α **"grande"**di questa Lezione.
E conseguentemente si intendano tutti i valori della tavola.

Simili problematiche e soluzioni coi vari software.

Attenzione anche al fatto che il valore (approssimato) 19.68 può trovarsi scritto con migliore approssimazione 19.6751, o eventualmente anche altrimenti.

Ecco alcuni valori:

su alcune tavole →	0.95	0.90	0.10	0.05	0.01
su altre tavole →	0.05	0.10	0.90	0.95	0.99
n				$\chi^2_{0.05}(1)$ *per altri* $\chi^2_{0.95}(1)$ ↓	
1	0.004	0.02	2.71	3.84	6.63
2	0.10	0.21	4.61	5.99	9.21
3	0.35	0.58	6.25	7.81	11.34
4	0.71	1.06	7.78	9.49	13.28
5	1.14	1.61	9.24	11.07	15.09
6	1.63	2.20	10.64	12.59	16.81
7	2.17	2.83	12.02	14.07	18.48
8	2.73	3.49	13.36	15.51	20.09
9	3.32?	4.17	14.68	16.92	21.67
10	3.94	4.87	15.99	18.31	23.21
11	4.57	5.58	17.28	**19.68** ↑ $\chi^2_{0.05}(11)$ *per altri* $\chi^2_{0.95}(11)$	24.73

BOZZA - DRAFT

Ripetiamo che per noi 19.68 è $\approx \chi^2_{0.05}(11)$, cioè $\chi^2_{1-\alpha}(11)$ con $\alpha = 0.95$.

In pratica

Se si ha una *normale* tavola del chi quadrato con più righe e colonne,

per l'applicazione del test del chi quadrato, quello basico,

all'usuale livello del 95% ovvero del 5%,

si cerca la colonna che all'undicesima riga ha 19.68 (164)

si va in quella stessa colonna alla riga dell'n che si ha

e vi si trova il quantile da usare nella formula (162).

Nota. Nel prossimo esercizio invece è già stata selezionata la colonna del 19.68, ed è stata scritta in riga.

BOZZA - DRAFT

Eercizio$_{\mu}$ $*$
Testare l'ipotesi nulla dell'uniforme distribuzione di una v.a. X a 3 valori 1, 2, 3, che sono apparsi 18, 35 e 22 volte rispettivamente. Per esempio potremmo chiederci se sono egualmente distribuiti i capelli 1 = biondo/castano, 2 = moro, 3 = rosso, nell'ipotesi semplificata che si possa fare una tale netta tripartizione; oppure potremmo considerare altra caratteristica genetica umana.

TAVOLA DEI QUANTILI DEL CHI QUADRATO — 0.05

1	2	3	4	5	6	7	8	9	10	11
3.84	5.99	7.81	9.49	11.07	12.59	14.07	15.51	16.92	18.31	19.68

SVOLGIMENTO
È usato lo standard del punto decimale.

L'ipotesi nulla H è l'uniforme distribuzione, che si spera di respingere:

$$H: (p_1\, p_2\, p_3) = \left(\frac{1}{3}\,\frac{1}{3}\,\frac{1}{3}\right)$$

$$A: (p_1\, p_2\, p_3) \neq \left(\frac{1}{3}\,\frac{1}{3}\,\frac{1}{3}\right)$$

e A è l'alternativa.
Con $k = 3$, $n_1 = 18$, $n_2 = 35$, $n_3 = 22$ e $n = 18 + 35 + 22 = 75$ dobbiamo verificare se

$$n\,p_i \geq 5 \quad \text{per } i = 1, \ldots, k \;\wedge\; T_n := \sum_{i=1}^{k} \frac{(n_i - n\,p_i)^2}{n\,p_i} > \chi^2_{1-\alpha}(k-1)$$

È $n\,p_i = 75 \cdot \frac{1}{3} = 25 > 5$ per $i = 1, 2, 3$: la condizione è verificata, abbondantemente (e alcuni Autori richiedono la condizione > 10 invece che > 5).
Fissando il classico $\alpha = 0.95$, si trova nella tavola il quantile d'interesse:

$$\chi^2_{1-\alpha}(k-1) = \chi^2_{0.05}(2) \approx 5.99$$

Statistica T_n:

$$T_{75} := \sum_{i=1}^{3} \frac{(n_i - 75 \cdot \frac{1}{3})^2}{75 \cdot \frac{1}{3}} =$$

$$= \frac{(18 - 75 \cdot \frac{1}{3})^2}{75 \cdot \frac{1}{3}} + \frac{(35 - 75 \cdot \frac{1}{3})^2}{75 \cdot \frac{1}{3}} + \frac{(22 - 75 \cdot \frac{1}{3})^2}{75 \cdot \frac{1}{3}} =$$

$$= \frac{7^2 + 10^2 + 3^2}{25} = \frac{158}{25} \approx 6.32 > 5.99$$

Il quantile è superato e allora

si respinge al livello 0.05 l'ipotesi dell'uniforme distribuzione.

(Ovvero, al livello 95%, come altri direbbe). (Si *respinge*, o si *rifiuta*).

BOZZA - DRAFT

ESERCIZIO$_\mu$ Si fa un sondaggio su questo quesito: "Quanto costa il regalo che vorreste a Natale, in euro, approssimato all'intero?"
Si ottiene questo dataset X:

250 700 200 30 69 400
30 50 29 120 110 100
800 9000 90 60 600
500 200 52 500 100
1129 150 0 200 679 20 50
500 3000 198 28 795
124 200 80 580 30 800
50 49 1500 100 1100 3500
180 0 5 0 50 50 900

(a) Eliminati gli outlier nulli, produrre il corrispondente dataset Y ottenuto con la sola cifra iniziale di ogni dato di X, e fare il corrispondente istogramma a barre con le frequenze relative percentuali a destra di ogni barra (orizzontale).
(b) Testare l'ipotesi che la cifra iniziale sua distribuita uniformemente (come molti pensano dovrebbe essere e invece in generale non succede).

SVOLGIMENTO

(a) Dataset Y delle cifre iniziali, dopo l'eliminazione degli outlier nulli:

2 7 2 3 6 4
3 5 2 1 1 1

8 9 9 6 6
5 2 5 5 1
1 1 2 6 2 5
5 3 1 2 7
1 2 8 5 3 8
5 4 1 1 1 3
1 5 5 5 9

Rango del dataset: $n = 50$.
Frequenze assolute, e frequenze relative:

numero di 1, $n_1 = 12$, $f_1 = 12/50 = 0.24$
numero di 2, $n_2 = 8$, $f_2 = 8/50 = 0.16$
numero di 3, $n_3 = 5$, $f_3 = 5/50 = 0.1$
numero di 4, $n_4 = 2$, $f_4 = 2/50 = 0.04$
numero di 5, $n_5 = 11$, $f_5 = 11/50 = 0.22$
numero di 6, $n_6 = 4$, $f_6 = 4/50 = 0.08$
numero di 7, $n_7 = 2$, $f_7 = 2/50 = 0.04$
numero di 8, $n_8 = 3$, $f_8 = 3/50 = 0.06$
numero di 9, $n_9 = 3$, $f_9 = 3/50 = 0.06$

Istogramma a barre con le frequenze relative in percentuale:

```
1 |OOOOOOOOOOOO 24%
2 |OOOOOOOO 16%
3 |OOOOO 10%
4 |OO 4%
5 |OOOOOOOOOOO 22%
6 |OOOO 8%
7 |OO 4%
8 |OOO 6%
9 |OOO 6%
```

(Che non pare molto uniforme).

(b) Test per l'uniformità di

valori	1	2	3	4	5	6	7	8	9
frequenze assolute	12	8	5	2	11	4	2	3	3

$$H: (p_1\ p_2\ p_3\ p_4\ p_5\ p_6\ p_7\ p_8\ p_9) = \left(\frac{1}{9}\ \frac{1}{9}\ \frac{1}{9}\ \frac{1}{9}\ \frac{1}{9}\ \frac{1}{9}\ \frac{1}{9}\ \frac{1}{9}\ \frac{1}{9}\right)$$

$$A: (p_1\ p_2\ p_3\ p_4\ p_5\ p_6\ p_7\ p_8\ p_9) \neq \left(\frac{1}{9}\ \frac{1}{9}\ \frac{1}{9}\ \frac{1}{9}\ \frac{1}{9}\ \frac{1}{9}\ \frac{1}{9}\ \frac{1}{9}\ \frac{1}{9}\right)$$

Dobbiamo verificare se

$$n\,p_i \geq 5 \quad \text{per } i = 1, ..., k \ \textit{et}\ T_n := \sum_{i=1}^{k} \frac{(n_i - n\,p_i)^2}{n\,p_i} > \chi^2_{1-\alpha}(k-1)$$

BOZZA - DRAFT

Con $n = 50$ è $n\,p_i = 50 \cdot \frac{1}{9} > 5$ per $i = 1, ..., 9$. (La condizione è verificata).
Quantile, essendo $k = 9$ e fissando ragionevolmente $\alpha = 0.95$:

$$\chi^2_{1-\alpha}(k-1) = \chi^2_{0.05}(8) \approx 15.51$$

Statistica T_n:

$$T_{50} := \sum_{i=1}^{9} \frac{(n_i - 50 \cdot \frac{1}{9})^2}{50 \cdot \frac{1}{9}} =$$

$$= \frac{(12 - 50 \cdot \frac{1}{9})^2}{50 \cdot \frac{1}{9}} + \frac{(8 - 50 \cdot \frac{1}{9})^2}{50 \cdot \frac{1}{9}} + \frac{(5 - 50 \cdot \frac{1}{9})^2}{50 \cdot \frac{1}{9}} + \frac{(2 - 50 \cdot \frac{1}{9})^2}{50 \cdot \frac{1}{9}} +$$

$$+ \frac{(11 - 50 \cdot \frac{1}{9})^2}{50 \cdot \frac{1}{9}} + \frac{(4 - 50 \cdot \frac{1}{9})^2}{50 \cdot \frac{1}{9}} + \frac{(2 - 50 \cdot \frac{1}{9})^2}{50 \cdot \frac{1}{9}} + \frac{(3 - 50 \cdot \frac{1}{9})^2}{50 \cdot \frac{1}{9}} + \frac{(3 - 50 \cdot \frac{1}{9})^2}{50 \cdot \frac{1}{9}} =$$

$$= \frac{532}{25} = 21.28 > 15.51$$

e superandosi il quantile

si rifiuta al livello 95% l'ipotesi dell'uniforme distribuzione

(Si noti che si riesce a respingerla anche al livello 99% ovvero 1% perché si supera anche il quantile $\chi^2_{0.01}(8) \approx 20.09$. Il p-value, il valore discriminante fra rifiuto e non rifiuto, verosimilmente non è lontano, ma è comunque $p < 0.01$, un buon risultato).

BOZZA - DRAFT

58.5 Densità di Benford

È addirittura possibile che il genere di dataset dell'esercizio precedente segua la *densità di Benford*

$$\left(\lg 2 \;\; \lg\frac{3}{2} \;\; \lg\frac{4}{3} \;\; \lg\frac{5}{4} \;\; \lg\frac{6}{5} \;\; \lg\frac{7}{6} \;\; \lg\frac{8}{7} \;\; \lg\frac{9}{8}\right)$$

ma come sappiamo $\lg 9 \approx 0.95$ e $\lg 8 \approx 0.9$ e allora $\lg\frac{9}{8} = \lg 9 - \lg 8 \approx 0.05$ e con questa p_9 non è verificata la condizione $n\,p_i > 5$ con $n = 50$: troppo pochi dati, non si può eseguire il test, e nemmeno riunendo le ultime 2 classi, perchè è vero che

$$50\,(p_8 + p_9) = 50\left(\lg\frac{8}{7} + \lg\frac{9}{8}\right) \approx 5.46 > 5$$

ma per la classe precedente ancora

$$50 \lg \frac{7}{6} \approx 3.35 < 5$$

Troppo pochi dati, anche se non mancherà qualcuno che tenterà comunque di affermare qualcosa, facendo ulteriori riunioni di classi. Noi ci fermiamo qua.

Una caratteristica della distribuzione di Benford è che la probabilità complessiva delle 3 cifre 1, 2 e 3 è maggiore della probabilità complessiva di tutte le altre 6 messe insieme, e nel nostro dataset avviene che 25 numeri su 50 iniziano per 1, 2 o 3. Un numero "qualunque", ma non iniziante per 0, proveniente dalla realtà sensibile – per quanto ciò possa significare – è più probabile che inizi per 1, 2 o 3 piuttosto che con 4, 5, 6, 7, 8 o 9, contrariamente a quello che si potrebbe pensare.

Nota. Di questa Lezione ci si dovrebbe aspettare che lo studente conosca a memoria le

2

formule *contemporaneamente* numerate *et* riquadrate, oltre a tutto l'inquadramento teorico della questione.

BOZZA - DRAFT

59 Il test dei ranghi

59.1 Premessa

https://en.wikipedia.org/wiki/Pearson_correlation_coefficient#/media/File:Correlation_coefficient.png
https://en.wikipedia.org/wiki/Pearson_correlation_coefficient#/media/File:Correlation_examples2.svg

Seguiremo un testo[187] (pubblicato online il 31 dicembre 2020) che trovava che ad una certa data del 2020 (prima ondata di covid-19) a maggior densità di maiali d'allevamento, in 4 regioni d'Europa, fra cui le province della Lombardia corrispondeva maggior numero di casi (cumulativi, fino a quella data).

BOZZA - DRAFT

59.2 Testing monotonicity of ranks

To test that to more pigs correspond more deaths, or more cases, or whatever, one may test the monotonicity of their respective ranks. Being n the sample size of X and Y and ρ_s the Spearman's rank correlation coefficient, which is exactly the Pearson's correlation index for ranks, the significance may be tested using

$$\rho_s := 1 - \frac{6\Sigma d_i^2}{n\left(n^2-1\right)} \tag{165}$$

where the d are the differences of the corresponding ranks. Just as in https://en.wikipedia.org/wiki/Spearman%27s_rank_correlation_coefficient#Example.
At least if $4 \leq n \leq 30$ a table of critical values for the 1-tail test for increasing monotonicity is available.

H_0: X and Y are independent

H_A: there is a trend associating larger values of X to larger values of Y.

[187] "In the first wave of the 2020 pandemic in several areas more sunlight less pandemic, more pigs more pandemic, and lower correlations with some other livestocks". (December 2020), A. Soranzo. DOI: 10.13140/RG.2.2.29852.31367

The null hypothesis H_0 is refused at significance level α if ρ_s (in that text ρ_s is denoted r_s) is > than the corresponding critical value of the table for α and n.

Here are some values[188] of that table.

n	.001	.005	.010	.025	.050	.100
.						
.						
.						
6	–	.9429	.8857	.8286	.7714	.6000
.						
8	.9286	.8571	.8095	.7143	.6190	.5000
9	.9000	.8167	.7667	.6833	.5833	.4667
10	.8767	.7818	.7333	.6364	.5515	.4424
.						
12	.8182	.7273	.6713	.5804	.4965	.3986
.						
.						
.						
24	.6070	.5200	.4748	.4061	.3435	.2704
.						
.						
.						

BOZZA - DRAFT

59.3 Lombardia, 31 March 2020

The first death in Lombardia, and in Italy, dates[189] 21 February 2020. The fundamental database GitHub[190] starts the historic

[188] Wayne W. Daniel, Chad L. Cross, Biostatistica. Concetti di base per l'analisi statistica delle scienze dell'area medico- sanitaria, Traduttore: M. Attanasio, V. Capursi, Curatore: C. Frigo, E. Perissinotto, Editore: Edises, Terza edizione (2019), p. 737, A-104

[189] https://www.corriere.it/cronache/20_febbraio_21/coronavirus-muore-uomo-77-anni-monselice-dac529f6-54f9-11ea-9196-da7d305401b7.shtml

[190] GitHub database for regions https://raw.githubusercontent.com/pcm-dpc/COVID-19/master/

series from 24 February, with 7 deaths, clearly corresponding to several previous days.

In this section it will be considered Lombardia on 31 March 2020, not considering meaningful a statistic for the last day of February, at the very beginning of the epidemic in Italy.

In **??** the situation will be investigated month by month, following in particular the Pearson's correlation index between pig density (on log scale) and cumulative cases, and the Spearman's rank correlation coefficient ρ.
We may anticipate that the (geographical and then statistical) correlation is very high and that in Autumn it fades. See also **??**.

BOZZA - DRAFT

59.3.1 Original data for Lombardia, 31 March 2020

Province	Area	Human population	Farm pigs	[a]Cases till 31 March
Bergamo	2755	1110457	890418	8803
Brescia	4786	1262135	1358038	8367
Como	1279	599637	2148	1101
Cremona	1770	358578	969149	3869
Lecco	806	337256	4269	1470
Lodi	783	229946	356688	2116
Mantova	2341	411959	1190459	1688
Milano	1576	3233541	77959	8911
Monza	405	871523	3504	2462
Pavia	2969	545611	233078	2133
Sondrio	3196	181249	1834	470
Varese	1198	890418	1018	893

(a) from GitHub[?]. Retrieved 5 December 2020

dati-regioni/dpc-covid19-ita-regioni.csv

59.3.2 Derived data for Lombardia, 31 March 2020

Province	Rank for farm pig density	Farm pig density	Cases per million	Rank for cases per million
Bergamo	8	107.8	7927	10
Brescia	9	283.8	6629	9
Como	3	1.679	1836	2
Cremona	12	547.5	10790	12
Lecco	4	5.297	4359	8
Lodi	10	455.5	9202	11
Mantova	11	508.5	4097	7
Milano	6	49.47	2756	4
Monza	5	8.652	2825	5
Pavia	7	78.50	3909	6
Sondrio	1	0.5738	2593	3
Varese	2	0.8497	1003	1

BOZZA - DRAFT

59.3.3 Ranks

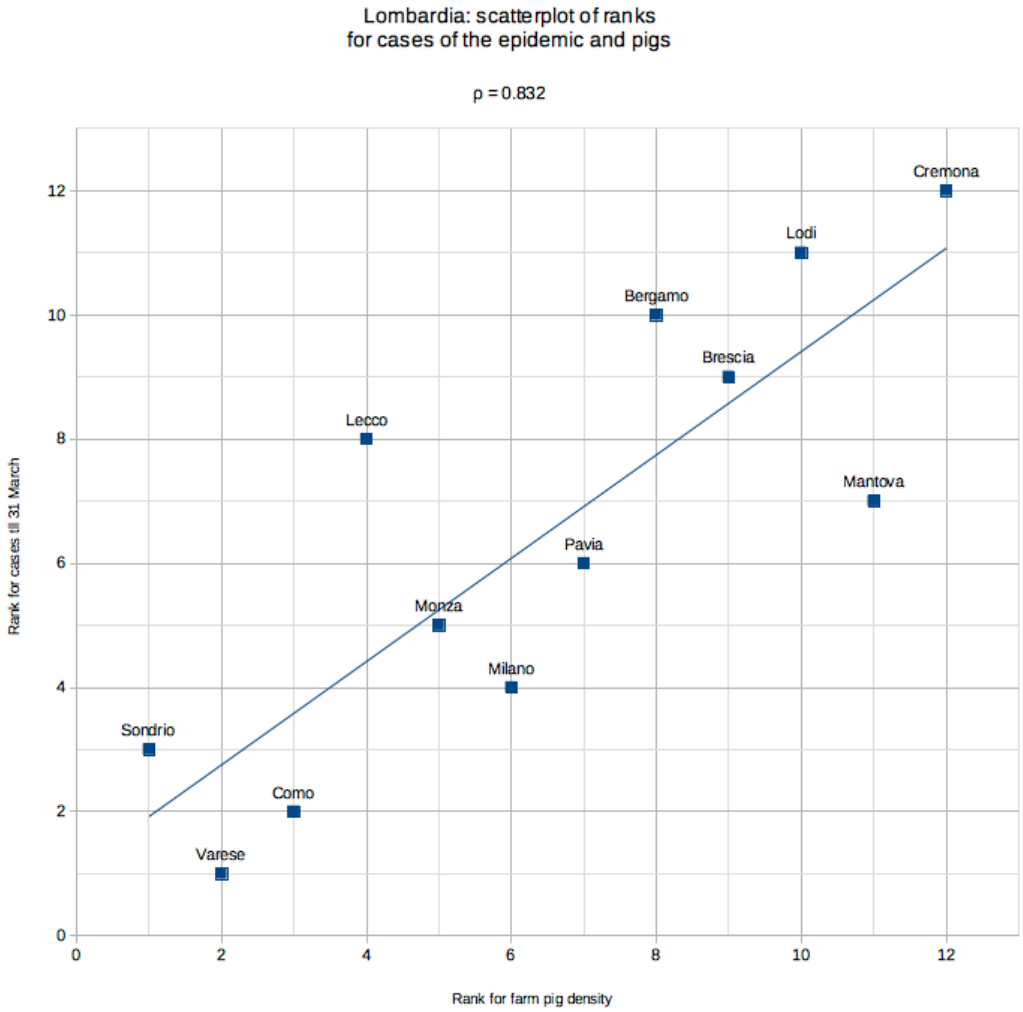
The increasing order is chosen: 1 corresponds to the lowest farm pig density, and to the lowest number of deaths per million.

provinces / *rank for pigs*	P_i	1	2	3	4	5	6	7	8	9	10	11	12
rank for cases / *provinces*	C_{P_i}	3	1	2	8	5	4	6	10	9	11	7	12

Considering the ranks P_i and C_{P_i}, at a glance, it appears that

yes,
generally

more farm pigs, more cases per million.



Final remark: p-value. To test the increasing monotonicity of ranks, with $n = 12$, Formula (165) gives 0.832 (which is in fact ρ_s) which grants, as may be seen in the table in Section **??**,

$$p < 0.001$$

and largely, in fact.

60 Alcuni Test di Student

60.1 Premessa: come misurare l'esposizione?

Inseriamo in questa Lezione una questione di portata molto generale in Medicina e Farmacia.
Supponiamo di voler testare, magari proprio col gold standard, un trial clinico randomizzato in doppio cieco, un farmaco per grandi fumatori, magari per ridurre gli episodi di bronchite, o quant'altro. Ma chi sono i "grandi fumatori", o anche semplicemente i fumatori? Chi fuma più di 10 sigarette al giorno, oppure 20, oppure... quante? Qua si apre un immenso e catastrofico capitolo della Medicina e della Farmacia: come misurare l'esposizione, per esempio all'amianto, oppure al fumo? Considereremo fumatore chi fuma più di 0 o di 10 sigarette al giorno? È chiaro che modulando questi parametri si può largamente pilotare il risultato, eventualmente a favore del finanziatore delle ricerche. E infatti negli articoli scientifici spesso vi è una dichiarazione, eventualmente fasulla, di mancanza di conflitto di interessi. Fumatore, consumatore di carne o pesce o alcol o tranquillanti o qualsiasi cosa... in tutto si potranno scegliere arbitrariamente soglie. Per esempio, qualche anno fa è diventato famosissimo un articolo scientifico che trovava che mangiare molta carne rossa aumenterebbe la mortalità. Ma chi deve essere classificato grande mangiatore di carne rossa, e chi scarso mangiatore di carne rossa, con cui fare il confronto?
E così pure – con meno arbitrarietà perché su queste cose vi sono alcuni standard di enti autorevoli – per la carenza di vitamina D o B12 o quant'altro.

BOZZA - DRAFT

60.2 Test di Student per la media $\leq$ e $=$

In questa Lezione α è un numero positivo e minore di 1 ma "grande", diciamo $> 0.5 = 50\%$, tipicamente $0.95 = 95\%$ (oppure $0.99 = 99\%$).

Diremo i test

al livello di significatività $1-\alpha$,

al livello di confidenza α
e altri diranno, e anche noi talvolta, semplicemente,

al livello α

al livello $1-\alpha$, per esempio 5%, da cui traiamo $\alpha = 0.95$.
Stando attenti a non confondere α e il suo complemento: α è grande in questa Lezione, $1-\alpha$ piccolo; entrambi fra 0 e 1.

Supponiamo di avere dei numeri – per esempio parametri fisiologici, naturali o conseguenti ad una terapia – e di chiederci se la loro media è maggiore di 1492, per esempio.
A questo la Statistica antica, in pratica la Statistica Descrittiva, risponde molto semplicemente: si fa la media aritmetica e si risponde "sì" se è > 1492.
Invece la Statistica moderna, in pratica la Statistica Inferenziale, accetta che la media sia > 1492 in maniera statisticamente significativa, o per meglio dire respinge l'ipotesi che sia ≤ 1492, solo se la media aritmetica è ben più grande di 1492.

BOZZA - DRAFT

Quanto più grande? Fissato il livello di confidenza

– basterà poco se la numerosità n è grande.

– ci vorrà molto se la variabilità S^2 è grande.

Il tutto è risolto da una formula. Essa combina n e S^2 e α coi quantili di Student. E risponde *sì* o *no* in termini di *vero* o *falso*.

Teorema. Sia $X_1,...,X_n$ un campione gaussiano cioè $X_i \sim N(\mu, \sigma^2)$ sono indipendenti (o anche non gaussiano se n è sufficientemente grande), sia μ_0 un numero reale. Allora (coi soliti stimatori $\bar{X}_n$ e S_n^2 di media e varianza) si hanno i seguenti 2 test al livello α (e altri dicono al livello $1-\alpha$; in ogni caso per noi α è "grande", tipicamente 0.95 o addirittura 0.99).

Test a una coda:

$$\boxed{\begin{cases} H : \mu \le \mu_0 \\ A : \mu > \mu_0 \end{cases} \quad rifiuta\ H\ se\ \sqrt{n}\,\frac{\bar{X}_n - \mu_0}{S_n} > t_\alpha(n-1)} \tag{166}$$

Test a due code:

$$\begin{cases} H : \mu = \mu_0 \\ A : \mu \neq \mu_0 \end{cases} \quad rifiuta\ H\ se\ \left|\sqrt{n}\,\frac{\bar{X}_n - \mu_0}{S_n}\right| > t_{\frac{1+\alpha}{2}}(n-1) \tag{167}$$

(La quantità $\sqrt{n}\,\frac{\bar{X}_n-\mu_0}{S_n}$ è uno *stimatore*, ed è la *statistica* del primo test, o *statistica di test*, e spesso è indicata con T. Similmente per il secondo test, e si potrà trovarla indicata T o $|T|$, e ovviamente il valore assoluto può essere messo anche solo al termine $\bar{X}_n - \mu_0$).

Gli autori che usano $1-\alpha$ per α hanno formule diverse. Quanto grande n, nel caso non gaussiano cioè normale? Secondo diversi Autori $n > 30$ oppure $n > 120$. Se normale, va bene per qualunque n.

BOZZA - DRAFT

In pratica. Coi valori $x_1,..,x_n$ si calcola lo stimatore T che è

$$\sqrt{n}\,\frac{\bar{X}_n - \mu_0}{S_n}$$ o nel secondo caso il valore assoluto di questo

e

se supera il quantile si respinge l'ipotesi (vero = sì = rifiuto H) altrimenti "non la si respinge" (falso = no = non rifiuto H).

Esempio del test a due code. Leggiamo (10 gennaio 2023) in https://www.jmp.com/it_it/statistics-knowledge-portal/t-test.html

> La frequenza cardiaca media di un gruppo di persone è uguale a 65

Il test è a 2 code perchè rifiuteremo l'ipotesi nulla dell'uguaglianza sia se troveremo una media troppo minore di 65 sia troppo maggiore.

60.3 Note euristiche sui 2 Test di Student

Consideriamo la formula (166).
– Supponiamo dapprima fissato il livello α.
Con n grande, T tende a superare il quantile anche con piccola superiorità di $\bar{X}_n$ su μ_0.
Con S^2 grande, T stenta a superare il quantile anche con grande superiorità di $\bar{X}_n$ su μ_0. (Dati inconsistenti, si contraddicono).
– Ma anche α partecipa, e infatti è nel quantile.
Più α è piccolo, ovvero più è grande il livello di confidenza $1-\alpha$ ricercato, più cresce il quantile ovvero la soglia da superare (col triplice contributo combinato di grande n, grande sopravanzare di $\bar{X}_n$ su μ_0, e/o piccolo S^2).

Considerazioni analoghe valgono per la formula (167).

BOZZA - DRAFT

60.4 Esempio estremo per fissare le idee

Con $n = 2$, due soli dati, è $n - 1 = 1$ e ci serve $t_\alpha(1)$ da trovarsi con le tavole o i software o addirittura coi calcoli perché $f(t;1)$ è la densità di Cauchy:

$$f(t;1) = \frac{1}{\pi}\frac{1}{1+t^2}$$

$$F(x) = \int_{-\infty}^{x} f(t;1) = \int_{-\infty}^{x} \frac{1}{\pi}\frac{1}{1+t^2}\,dt =$$

$$= \left[\frac{1}{\pi}\arctan t\right]_{-\infty}^{x} = \frac{1}{\pi}\arctan x - \frac{1}{\pi}\arctan(-\infty) =$$

$$= \frac{1}{2} + \frac{1}{\pi}\arctan x$$

e la sua inversa dà i quantili di student $t_\alpha(1)$ risolvendo

$$F(x) \overset{EQ}{=} \alpha \quad > 0.5 \text{ (“grande")}$$

$$\frac{1}{2} + \frac{1}{\pi}\arctan x = \alpha \quad \Big/ + \left(-\frac{1}{2}\right)$$

$$\frac{1}{\pi}\arctan x = \alpha - \frac{1}{2} \quad \Big/ \cdot \pi$$

$$\arctan x = \pi\left(\alpha - \frac{1}{2}\right)$$

un'attenta analisi mostrerebbe che possiamo applicare la tangente

$$\Big/ \tan \qquad \text{(che "mangia" l'arcotangente)}$$

$$x = \tan\left(\pi\left(\alpha - \frac{1}{2}\right)\right)$$

che è il quantile di Student cercato,

$$t_\alpha(1) = \tan\left(\pi\left(\alpha - 0.5\right)\right) \qquad \alpha > 0.5$$

e alcuni valori li avremo in forma esatta:
($t_{0.5}(1) = 0$, basta che $\bar{X}_n = \bar{X}_2 > \mu_0$, ma l'affermazione è irrilevante per troppo basso livello di confidenza, 50%)
$t_{\frac{2}{3}}(1) = \frac{\sqrt{3}}{3} \approx 0.577$
$t_{\frac{5}{6}}(1) = \sqrt{3} \approx 1.73$
($t_1(1) = +\infty$, cioè mai: livello di confidenza voluto troppo alto)
ottendendosi questa tavola dei quantili di Student

α	0.5	0.667	0.833	1
$t_\alpha(1)$	0	0.577	1.73	$+\infty$

che fanno capire chiaramente la situazione. Ma nessuno di quei 4 quantili è usato nella pratica. Più significativamente, con una calcolatrice che ci dia la tangente di

$\pi\left(0.95 - 0.5\right)$
$\pi\left(0.99 - 0.5\right)$
$\pi\left(0.995 - 0.5\right)$

otteniamo questa tavola dei quantili di Student

α	0.95	0.99	0.995
$t_\alpha(1)$	6.31375	31.8205	63.6567

BOZZA DRAFT

Li troviamo confermati nella tavola della Lezione (**??**). Indicati con “one tail", cioè test “a una coda", perché indaghiamo sul $\leq$ e non $\neq$, che sarebbe a 2 code.

Dal sito del NIST in https://www.itl.nist.gov/div898/handbook/eda/section3/eda3672.htm traiamo la seguente più completa tavola dei quantili di Student.

ν	0.90	0.95	0.975	0.99	0.995	0.999
1.	3.078	6.314	12.706	31.821	63.657	318.313
2.	1.886	2.920	4.303	6.965	9.925	22.327
3.	1.638	2.353	3.182	4.541	5.841	10.215
4.	1.533	2.132	2.776	3.747	4.604	7.173
5.	1.476	2.015	2.571	3.365	4.032	5.893
6.	1.440	1.943	2.447	3.143	3.707	5.208
7.	1.415	1.895	2.365	2.998	3.499	4.782
8.	1.397	1.860	2.306	2.896	3.355	4.499
9.	1.383	1.833	2.262	2.821	3.250	4.296
10.	1.372	1.812	2.228	2.764	3.169	4.143
11.	1.363	1.796	2.201	2.718	3.106	4.024
12.	1.356	1.782	2.179	2.681	3.055	3.929
13.	1.350	1.771	2.160	2.650	3.012	3.852
14.	1.345	1.761	2.145	2.624	2.977	3.787
15.	1.341	1.753	2.131	2.602	2.947	3.733
16.	1.337	1.746	2.120	2.583	2.921	3.686
17.	1.333	1.740	2.110	2.567	2.898	3.646
18.	1.330	1.734	2.101	2.552	2.878	3.610
19.	1.328	1.729	2.093	2.539	2.861	3.579
20.	1.325	1.725	2.086	2.528	2.845	3.552
21.	1.323	1.721	2.080	2.518	2.831	3.527
22.	1.321	1.717	2.074	2.508	2.819	3.505
23.	1.319	1.714	2.069	2.500	2.807	3.485
24.	1.318	1.711	2.064	2.492	2.797	3.467
25.	1.316	1.708	2.060	2.485	2.787	3.450
26.	1.315	1.706	2.056	2.479	2.779	3.435
27.	1.314	1.703	2.052	2.473	2.771	3.421
28.	1.313	1.701	2.048	2.467	2.763	3.408
29.	1.311	1.699	2.045	2.462	2.756	3.396
30.	1.310	1.697	2.042	2.457	2.750	3.385
31.	1.309	1.696	2.040	2.453	2.744	3.375
32.	1.309	1.694	2.037	2.449	2.738	3.365
33.	1.308	1.692	2.035	2.445	2.733	3.356
34.	1.307	1.691	2.032	2.441	2.728	3.348
35.	1.306	1.690	2.030	2.438	2.724	3.340
36.	1.306	1.688	2.028	2.434	2.719	3.333
37.	1.305	1.687	2.026	2.431	2.715	3.326
38.	1.304	1.686	2.024	2.429	2.712	3.319
39.	1.304	1.685	2.023	2.426	2.708	3.313
40.	1.303	1.684	2.021	2.423	2.704	3.307

ISCRA - DRAFT

41. 1.303 1.683 2.020 2.421 2.701 3.301
42. 1.302 1.682 2.018 2.418 2.698 3.296
43. 1.302 1.681 2.017 2.416 2.695 3.291
44. 1.301 1.680 2.015 2.414 2.692 3.286
45. 1.301 1.679 2.014 2.412 2.690 3.281
46. 1.300 1.679 2.013 2.410 2.687 3.277
47. 1.300 1.678 2.012 2.408 2.685 3.273
48. 1.299 1.677 2.011 2.407 2.682 3.269
49. 1.299 1.677 2.010 2.405 2.680 3.265
50. 1.299 1.676 2.009 2.403 2.678 3.261
51. 1.298 1.675 2.008 2.402 2.676 3.258
52. 1.298 1.675 2.007 2.400 2.674 3.255
53. 1.298 1.674 2.006 2.399 2.672 3.251
54. 1.297 1.674 2.005 2.397 2.670 3.248
55. 1.297 1.673 2.004 2.396 2.668 3.245
56. 1.297 1.673 2.003 2.395 2.667 3.242
57. 1.297 1.672 2.002 2.394 2.665 3.239
58. 1.296 1.672 2.002 2.392 2.663 3.237
59. 1.296 1.671 2.001 2.391 2.662 3.234
60. 1.296 1.671 2.000 2.390 2.660 3.232
61. 1.296 1.670 2.000 2.389 2.659 3.229
62. 1.295 1.670 1.999 2.388 2.657 3.227
63. 1.295 1.669 1.998 2.387 2.656 3.225
64. 1.295 1.669 1.998 2.386 2.655 3.223
65. 1.295 1.669 1.997 2.385 2.654 3.220
66. 1.295 1.668 1.997 2.384 2.652 3.218
67. 1.294 1.668 1.996 2.383 2.651 3.216
68. 1.294 1.668 1.995 2.382 2.650 3.214
69. 1.294 1.667 1.995 2.382 2.649 3.213
70. 1.294 1.667 1.994 2.381 2.648 3.211
71. 1.294 1.667 1.994 2.380 2.647 3.209
72. 1.293 1.666 1.993 2.379 2.646 3.207
73. 1.293 1.666 1.993 2.379 2.645 3.206
74. 1.293 1.666 1.993 2.378 2.644 3.204
75. 1.293 1.665 1.992 2.377 2.643 3.202
76. 1.293 1.665 1.992 2.376 2.642 3.201
77. 1.293 1.665 1.991 2.376 2.641 3.199
78. 1.292 1.665 1.991 2.375 2.640 3.198
79. 1.292 1.664 1.990 2.374 2.640 3.197
80. 1.292 1.664 1.990 2.374 2.639 3.195
81. 1.292 1.664 1.990 2.373 2.638 3.194
82. 1.292 1.664 1.989 2.373 2.637 3.193
83. 1.292 1.663 1.989 2.372 2.636 3.191
84. 1.292 1.663 1.989 2.372 2.636 3.190
85. 1.292 1.663 1.988 2.371 2.635 3.189
86. 1.291 1.663 1.988 2.370 2.634 3.188
87. 1.291 1.663 1.988 2.370 2.634 3.187
88. 1.291 1.662 1.987 2.369 2.633 3.185
89. 1.291 1.662 1.987 2.369 2.632 3.184
90. 1.291 1.662 1.987 2.368 2.632 3.183
91. 1.291 1.662 1.986 2.368 2.631 3.182

BOZZA - DRAFT

92.	1.291	1.662	1.986	2.368	2.630	3.181
93.	1.291	1.661	1.986	2.367	2.630	3.180
94.	1.291	1.661	1.986	2.367	2.629	3.179
95.	1.291	1.661	1.985	2.366	2.629	3.178
96.	1.290	1.661	1.985	2.366	2.628	3.177
97.	1.290	1.661	1.985	2.365	2.627	3.176
98.	1.290	1.661	1.984	2.365	2.627	3.175
99.	1.290	1.660	1.984	2.365	2.626	3.175
100.	1.290	1.660	1.984	2.364	2.626	3.174
$+\infty$	1.282	1.645	1.960	2.326	2.576	3.090

Si noti nell'ultima riga il limite all'infinito e in particolare il valore

$$1.96 \approx \phi_{0.975} = t_{0.975}(+\infty)$$

Esempio. Dal classico libro *Matematica e Statistica. Le basi per le Scienze della vita*, di Marco Abate.

BOZZA DRAFT

> *Il proprietario dell'azienda vinicola da te preferita teme che il tasso alcolico del suo vino quest'anno possa non essere più pari al 12.5% indicato in etichetta e ti chiede d'investigare. Misurando il tasso alcolico di 6 bottiglie, ottiene i seguenti valori:*
>
> *11.5, 11, 12.5, 13.1, 12.7, 12.4 .*
>
> *Supponendo che il tasso alcolico nel vino segua una distribuzione normale, verifica o smentisci il timore del proprietario.* L'ipotesi nulla è che il tasso alcolico medio del vino sia 12.5%. Siccome abbiamo supposto una distribuzione normale, possiamo applicare il test T con $\nu = 6 - 1 = 5$ gradi di libertà. La media dei campioni è $M_6 = 12.2$ mentre la deviazione standard campionaria è $s_6 \simeq 0.79$. Quindi il valore del test è:
>
> $$|T_5| = \frac{|12.2 - 12.5|}{0.79}\sqrt{6} \simeq 0.93.$$
>
> (...) il valore di soglia al livello di affidabilità 0.1 (QUINDI PER NOI $\alpha = 0.9$) (...) è 2.015, ben più alto del valore che abbiamo ottenuto. Quindi i dati statistici da te raccolti

non smentiscono l'ipotesi nulla e non confermano il timore del proprietario.

Si noti che noi abbiamo agito come un'autorità di controllo: vogliamo dimostrare che non è 12.5 e allora mettiamo come ipotesi alternativa proprio quella. Solo al 90% è stato fatto, poco.
Il valore 2.015 è il quantile $t_{\frac{1+\alpha}{2}}(n-1)$ cioè $t_{\frac{1+0.9}{2}}(6-1)$ cioè $t_{0.95}(5)$.

Esempio. Del libro di Paolo Baldi già citato:

> L'altezza media delle reclute alla visita di leva del 1970 era di 169 cm; 121 reclute vengono scelte a caso nel 1980. I valori di media varianza del campione sono:
>
> $$\bar{X} = 171 \quad \leftarrow \textit{intende } \bar{X}_n \textit{ con } n{:=}121$$
> $$S^2 = 85 \quad \leftarrow \textit{intende } S_n^2 \textit{ con } n{:=}121$$
>
> Si può affermare che l'altezza media delle reclute è aumentata?

BOZZA - DRAFT

(Siccome si vuole dimostrare che l'altezza è aumentata, si metta come ipotesi che non è aumentata, e siccome nulla è detto sul livello del test si usi la classica soglia del 5%).

Applichiamo la (166):

$$\begin{cases} H : \mu \le 169 \\ A : \mu > 169 \end{cases} \qquad rifiuta\ H\ se\ \sqrt{121}\,\frac{171-169}{\sqrt{85}} > t_{0.95}(120)$$

$$11\,\frac{2}{9.22} \approx 2.386 > 1.645 = t_{0.95}(+\infty) \approx t_{0.95}(120)$$

Sulla tavola data prima, abbiamo
$t_{0.95}(100) \approx 1.660$
$t_{0.95}(+\infty) \approx 1.645$
ed entrambi possono essere presi come approssimazione di $t_{0.95}(120)$, concludendo analogamente.

Si respinge al livello 0.05 l'ipotesi che l'altezza non sia aumentata.

Esempio. I test di Student sono usati amplissimamente in Medicina e Farmacia, ma non è facile trovare un test che si possa fare in aula; non possiamo chiedere glicemie o sideremie, sconosciute agli studenti, nè il peso o l'altezza, che potrebbero creare imbarazzo. Tuttavia qualcosa si può tentare. È noto che in Italia il calcio è più seguito dai maschi che dalle femmine, e potremmo cercare di verificarlo anche a livello degli studenti dell'aula.
Per il seguire il calcio, facciamo ragionevolmente così: misureremo l'interesse per il calcio dal numero di anni passati di cui la persona ritiene di conoscere la squadra vincitrice del Campionato di serie A. Per esempio per lo scrivente l'interesse per il calcio, così misurato, è 0: non sa chi ha vinto nemmeno l'ultimo campionato. Ma qualcuno potrebbe sapere anche i risultati degli ultimi 10 o perfino 100 campionati, se è più interessato.
Un'altro possibile indicatore potrebbe essere – e si otterrebbero risultati diversi – il numero di calciatori noti per nome e cognome nella squadra di calcio nazionale italiana. Ma si noti che una ragazza magari è appassionatissima di calcio e conosce tutti i nomi della squadra femminile, ma non di quella maschile. E insomma non se ne viene fuori: ci sono diverse possibili scelte per misurare l'interesse per il calcio, come per l'esposizione al fumo di sigaretta o all'inquinamento o a qualsiasi cosa.
La questione in oggetto potrebbe essere banalmente risolta al livello di Statistica Descrittiva: chiediamo il numero detto a tutti i maschi e facciano la media, e poi similmente con tutte le femmine. Ma invece vogliamo fare Statistica Inferenziale: chiederemo il numero a tutti i maschi presenti (in generale molto pochi), faremo la media e quello sarà il parametro μ_0 di riferimento. Invece per le femmine (in generale molto più numerose) prenderemo un campione, per esempio di $n = 10$ soggetti, trovando lo stimatore $\bar{X}_{10}$ dell'interesse femminile per il calcio nell'aula. Qua c'è un problema perchè il test ha come alternativa $\mu > \mu_0$ e non $<$. Per superare questo prob-

BOZZA_DRAFT

lema, applicheremo il test a 1 coda (166) applicato ai valori *opposti*, cioè l'interesse, poniamo, 3.2, equivale a un disinteresse −3.2 per il calcio:

— rifiuta l'ipotesi nulla che le femmine si interessino al calcio di più dei maschi (ovvero in qualche modo quasi "accetta" che le femmine si interessino di meno dei maschi al calcio, frase più comprensibile anche se meno esatta)

$$se\ \sqrt{10}\,\frac{-\bar{X}_{10}+\mu_0}{S_{10}} > t_{0.95}(10-1) \approx 1.833$$

(Naturalmente possiamo usare WolframAlpha). Con il passaggio ai valori cambiati di segno, se respingeremo l'ipotesi nulla, avremo verificato statisticamente il *minor* interesse delle femmine per il calcio rispetto ai maschi; quello che supponevamo e volevamo dimostrare statisticamente con un campionamento. Il risultato così ottenuto si riferisce solo all'aula, ovviamente, non a tutta Italia, e già là viene solo da un campione e non da tutta la popolazione dell'aula. E poi potrebbe essere verificato con la Statistica Descrittiva, chiedendo a *tutte* le femmine.

BOZZA DRAFT

Nota. Di questa Lezione ci si dovrebbe aspettare che lo studente conosca l'inquadramento teorico della questione trattata e conosca a memoria l'unica formula (166) numerata *et* riquadrata del Test di Student per la media ≤ e sappia usare, se viene data, la (167).

61 Test di Student per il confronto di medie

Un problema classico della Statistica è determinare se 2 popolazioni hanno la stessa media, a partire da campioni di esse.

Per fissare le idee consideriamo il problema:
il tempo di attesa in 2 farmacie è uguale?
Definiamo, in via semplificata 2 variabili aleatorie:
X = tempo di attesa nella farmacia x
Y = tempo di attesa nella farmacia y.
Naturalmente sono variabili aleatorie: non c'è in nessuna delle 2 farmacie un tempo di attesa fisso e costante, con ogni cliente si produce un valore che è una determinazione della variabile aleatoria X o Y rispettivamente.

BOZZA DRAFT

Un altro esempio ci viene da Wikipedia:

> suppose we are evaluating the effect of a medical treatment, and we enroll 100 subjects into our study, then randomly assign 50 subjects to the treatment group and 50 subjects to the control group. In this case, we have two independent samples and would use the unpaired form of the t-test. https://en.wikipedia.org/wiki/Student%27s_t-test, letto il 10 gennaio 2023.

Un altro esempio in
https://www.jmp.com/it_it/statistics-knowledge-portal/t-test.html, letto il 10 gennaio 2023:

> Le frequenze cardiache medie di due gruppi di persone sono uguali

(Fra maschi/femmine, sportivi/non sportivi, fumatori/non fumatori...) Ciascuna delle 2 variabili aleatorie ha una sua distribuzione, che non conosceremo mai esattamente, con una sua speranza matematica, che ugualmente non conosceremo mai esattamente, ma è proprio il confronto statistico di quelle 2 speranze matematiche che ci

interessa. Le chiameremo μ_X e μ_Y.

La Statistica fra poco ci insegnerà come rifiutare o non rifiutare l'ipotesi nulla dell'uguaglianza con ragionevole certezza statistica, purchè disponiamo di un buon numero n di determinazioni di X e un buon numero m di determinazioni di Y. La questione si dirime con uno stimatore da confrontare con un quantile di Student, e ha una formula molto complicata, ma

noi considereremo solo il caso semplice in cui $n = m$

cioè – nell'esempio di prima – misuriamo un ugual numero di tempi di attesa nelle 2 farmacie.
Prima di avere le misurazioni ovvero i valori numerici, abbiamo 2 campioni aleatori

$X_1,...,X_n$
$Y_1,...,Y_n$.

Dopo la sperimentazione, avremo 2 dataset numerici

$x_1,...,x_n$
$y_1,...,y_n$.

BOZZA - DRAFT

Teorema. Siano $X_1,...,X_n$ un campione aleatorio (gaussiano oppure con n sufficientemente grande) di media μ_X e varianza σ_X^2 sconosciute e $Y_1,..,Y_n$ un campione aleatorio (gaussiano oppure con n sufficientemente grande) di media μ_Y e varianza $\sigma_Y^2 = \sigma_X^2 =: \sigma^2$ (uguale a quella del primo campione e sconosciuta), con le X_i e le Y_i indipendenti fra loro.

Si ha questo test (bilatero) al livello $1 - \alpha$ (e altri dicono al livello α; in ogni caso α è "grande", tipicamente 0.95):

$$\begin{cases} H : \mu_X = \mu_Y \\ A : \mu_X \neq \mu_Y \end{cases} \qquad rifiuta\, H\, se \ \left| \frac{\bar{X}_n - \bar{Y}_n}{\sqrt{S_X^2 + S_Y^2}} \sqrt{n} \right| > t_{\frac{1+\alpha}{2}}(2n-2) \tag{168}$$

(La quantità col valore assoluto è la statistica del test, e il suo argomento si potrà trovare indicato con T).

Nota. Purtroppo come si vede c'è la condizione molto limitativa che X e Y abbiano uguali varianze.

Esempio. Non è facile trovare un esempio che si possa fare in un'aula universitaria. Comunque, supponiamo che qualcuno abbia la strana idea che alla nascita le mamme degli studenti maschi e femmine avessero un'età media diversa (cosa relativamente improbabile in Italia) pur rimanendo simili le varianze.

X età mamma dei maschi, determinazioni x_1,...,x_n

Y età mamma delle femmine, determinazioni y_1,...,y_n

Con WolframAlpha si calcolerà T, con `Mean`, `Variance`, eccetera. Quello che c'è da aspettarsi, ragionevolmente, è che le medie campionarie verranno diverse ma non così tanto da respingere l'ipotesi nulla che le "vere" medie siano uguali.

Se per esempio $n = 5$, cioè si considerano 5 maschi e 5 femmine, il quantile da usare al solito livello del 95% è

$$t_{\frac{1+\alpha}{2}}(2 \cdot 5 - 2) = t_{\frac{1+0.95}{2}}(10 - 2) = t_{0.975}(8) \approx 2.306$$

Nota. Di questa Lezione non ci si dovrebbe aspettare che lo studente conosca a memoria alcuna formula ma solo l'inquadramento teorico della questione e sappia quindi usare, se assegnata, la formula numerata.

BOZZA DRAFT

62 Matematica col Computer – Lezione C

62.1 Risolvere un'equazione col computer online

La risoluzione al computer di equazioni in 1 variabile reale è un argomento che si è sviluppato moltissimo nell'ultimo secolo, costituendo una branca dell'Analisi Numerica.

Per il professionista di Farmacia o Scienze e Tecnologie Farmaceutiche, in un tempo seguente all'esame di Matematica, nel caso in cui gli servisse risolvere un'equazione, specialmente se

più complessa di una di primo grado come $2\,x + 3 = 0$,

più complessa di una di secondo grado,

come $2\,x^2 + 3\,x + 4 = 0$,

più complessa di una monomiale,

come $2\,x^3 = 4$,

più complessa di semplici equazioni con esponenziali e/o logaritmi

come $\lg x^2 = 3$

(tutti argomenti sui quali molto si è spiegato nel presente testo)

BOZZA - DRAFT

la cosa più consigliabile è senz'altro

tentare online con WolframAlpha.

Per esempio per risolvere

$$x^2 = \mathrm{e}^x$$

scriviamo su wolframalpha.com

`solve x^2=exp(x)`

e ci dà la soluzione approssimata

≈ -0.703467

Nota. Se ci riuscisse difficile trovare e digitare i caratteri `x^` allora invece di `x^2` possiamo scrivere `x*x`.

Invece di `exp` possiamo scrivere `Exp`, e invece delle parentesi tonde le quadre, e anche `E^x`

62.2 Risolvere un'equazione col computer offline

??? FARE ***

62.3 Grafici cartesiani e polari con MATLAB

Come scritto sulla pagina del Dipartimento di Ingegneria e Architettura dedicata a MATLAB, "UniTS ha sottoscritto una nuova Licenza MATLAB® Campus con funzionalità specifiche per il calcolo scientifico, il data science e la progettazione ingegneristica, che potrà essere scaricata e utilizzata gratuitamente da studenti, professori e staff UniTS sia per attività didattiche che di ricerca pubblica.

Per effettuare il download del software MATLAB ed attivare la licenza occorre accedere al portale:
https://www.mathworks.com/academia/tah-portal/universita-degli-studi-di-trieste-40730846.html "
Nella *Command Window* scriviamo il nostro software, ad esempio

```
x = linspace(0,6*pi);
y = cos(x);
plot(x,y)
```

e otteniamo il grafico della funzione coseno da 0 a $6\,\pi$.
L'immagine può essere ampiamente zoomata/ingrandita, con il mouse o con lo "strumento"indicato in alto a destra da una lente con il segno +.
Mostriamo adesso un grafico in coordinate polari scrivendo nella stessa *Command Window* ad esempio

```
theta = 0:0.01:2*pi;
rho = sin(3*theta).*cos(3*theta);
polarplot(theta,rho)
```

Il parametro 0.01 nella prima istruzione del programma determina la precisione del grafico e se provate a sostituirlo con 0.1 il grafico

BOZZA - DRAFT

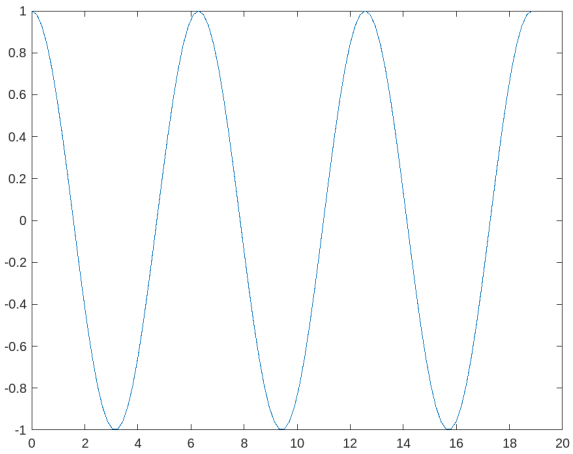


Figure 62: Grafico di $\cos(x)$ in $[0, 6\,\pi]$ ottenuto con il software MATLAB

ZZA - DRAFT

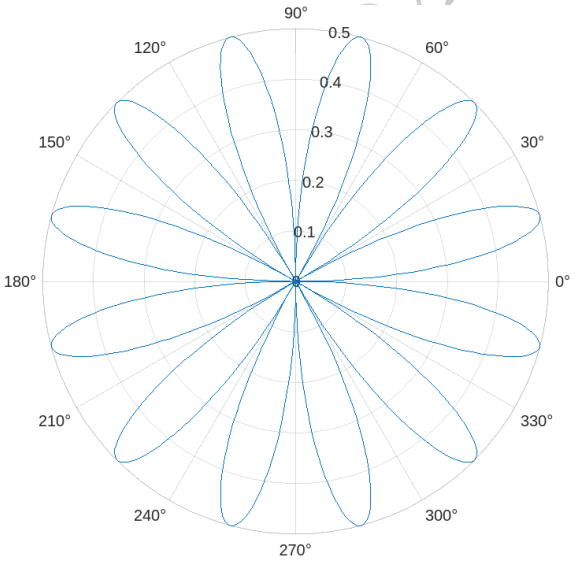


Figure 63: Grafico polare di $\rho = \sin(3\,\theta)$ per $0 \leq \theta \leq 2\,\pi$ ottenuto con il software MATLAB

diventerà molto più impreciso e "spigoloso".

Spieghiamo infine come salvare in locale le immagini prodotte da MATLAB per usarle in seguito in un vostro scritto.
Con il cursore andare sopra alla figura per far comparire l'icona (molto piccola) che indica la condivisione, andare con il cursore sulla piccola icona appena comparsa. Ne comparirà un'altra che indica il salvataggio. Cliccare su questa icona e dare un nome all'immagine che verrà salvata nel cloud di MATLAB. Per salvarla in locale, infine, cliccare con il tasto destro sul nome dell'immagine che sarà sul lato sinistro nel sottogruppo *Files* e selezionare *Download*.

BOZZA - DRAFT

63 Matematica delle Epidemie

Solo il paragrafo 63.7 nel corso breve – 40 ore.

La Matematica delle Epidemie si estende su varie branche della Matematica:
– Statistica Descrittiva
– altre Matematiche Elementari
– Calcolo Infinitesimale
– Calcolo delle Probabilità
– Statistica Inferenziale.

Nella prima parte di questa Lezione (prossimi 5 paragrafi) ricordiamo alcune cose viste in questi settori della Matematica, e poi nella seconda parte ne presentiamo pochissime altre nuove, non già trattate.

BOZZA - DRAFT

PARTE I – ALCUNE COSE GIÀ VISTE

63.1 Alcune cose viste in Statistica Descrittiva

Abbiamo visto i bar chart.

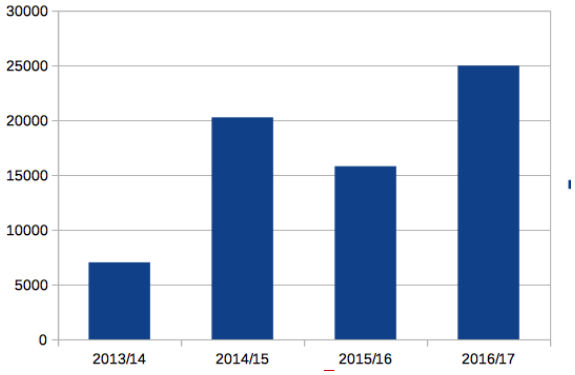


Figure 64: Morti stimati per epidemia di influenza.
Dati in: Rosano A, Bella A, Gesualdo F, Acampora A, Pezzotti P, Marchetti S, Ricciardi W, Rizzo C. Investigating the impact of influenza on excess mortality in all ages in Italy during recent seasons (2013/14-2016/17 seasons). Int J Infect Dis. 2019 Nov;88:127-134. doi: 10.1016/j.ijid.2019.08.003. Epub 2019 Aug 8. PMID: 31401203. https://pubmed.ncbi.nlm.nih.gov/31401203/

Abbiamo visto i grafici.

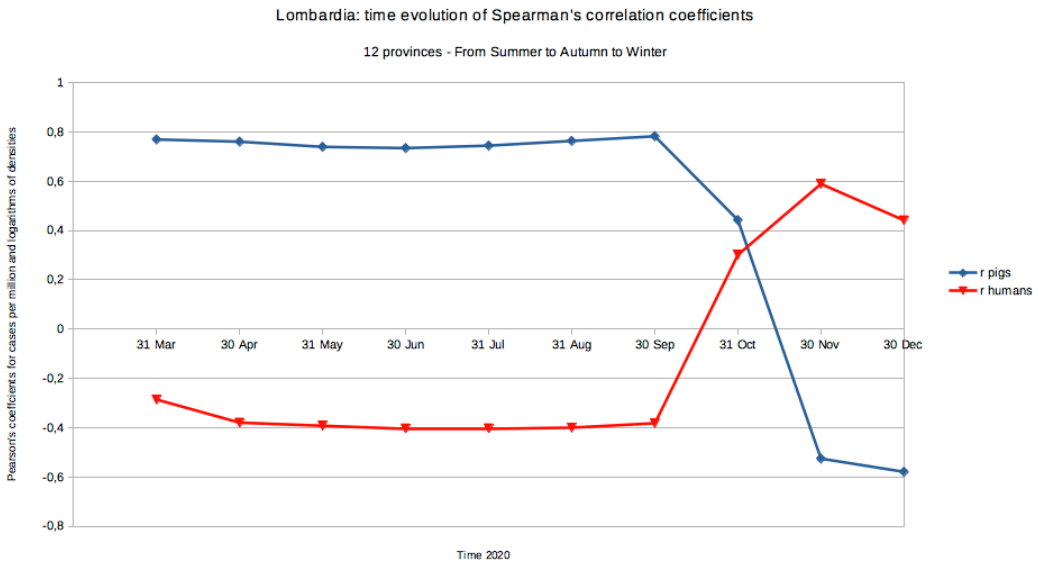
Figure 65: Dal sito web di EuroMOMO: eccessi di mortalità in alcuni stati. In Italia e Paesi Bassi sono ben evidenti gli effetti delle epidemie di influenza (in senso lato: la morte sopravviene per varie cause) e soprattutto della pandemia di covid-19: si osservino in particolare i tremendi picchi della primavera 2020.

BOZZA - DRAFT

Abbiamo visto il coefficiente di correlazione.
Per esempio, qua[191] vengono considerati dati sui casi covid nel 2020, nelle 12 province della Lombardia, in correlazione sia (blu) col logaritmo della densità dei maiali d'allevamento, e poi (rosso) col logaritmo della densità umana:

> At the level of the provinces of Lombardia, beside of the fading of the pig correlation, it may be observed also a very interesting overall increase of the correlation (from Spring to Winter) of the intensity of the pandemic (measured by cases per million) with the (logarithm of) human density, as may be seen in the following graph.

[191] A. Soranzo: In the first wave of the 2020 pandemic in several areas more sunlight less pandemic, more pigs more pandemic, and lower correlations with some other livestocks. (December 2020) ResearchGate, DOI: 10.13140/RG.2.2.29852.31367



Altre cose viste, usate nella Matematica delle Epidemie, sono:
– diagrammi a bolle
– box plot
– medie
– deviazione standard.

BOZZA - DR

63.2 Alcune cose viste in altre Matematiche Elementari

Abbiamo visto la formula dell'area del segmento parabolico da cui segue questa formula che permette spesso di dare una grossolana stima del numero dei morti di un'ondata epidemica – se come spesso succede essa descrive un segmento parabolico – dalla sua durata (in giorni) e dal massimo numero di morti in un giorno (ovvero ancor meglio il massimo della media mobile a 7 giorni):

$$\boxed{\text{morti di un'ondata epidemica} = \frac{2}{3}\, durata \times picco} \tag{169}$$

E abbiamo visto molte altre cose, usate nella Matematica delle Epidemie.

63.3 Alcune cose viste in Calcolo Infinitesimale

Abbiamo visto che le curve sigmoidi e le curve a campana , che così bene modellizzano in un'epidemia il numero cumulativo di morti e rispettivamente il numero di morti giornaliero, sono strettamente legate fra loro dalla derivata. Detto ultrasemplificatamente,
la derivata di una curva sigmoide è una curva a campana

Abbiamo visto il Modello SIR delle epidemie, basato su equazioni differenziali.
È bene dire, però, che nella pandemia iniziata nel 2020 e ancora (2023) in corso, l'enorme lavoro di modellazione teorico fatto ha previsto ben poco ovvero male il futuro. Piuttosto può spiegare bene il passato: coi dati reali in mano (di casi, o morti, o terapie intensive, o quant'altro, si trovano dei parametri che messi nelle varie equazioni riproducono abbastanza bene la realtà avvenuta, ma questo funziona a posteriori. Ne riparleremo nel Paragrafo 63.10.

BOZZA - DRAFT

63.4 Alcune cose in Calcolo delle Probabilità

Abbiamo visto Sensibilità, Specificità, Predittività, Curve ROC.
E abbiamo visto molte altre cose, usate nella Matematica delle Epidemie.

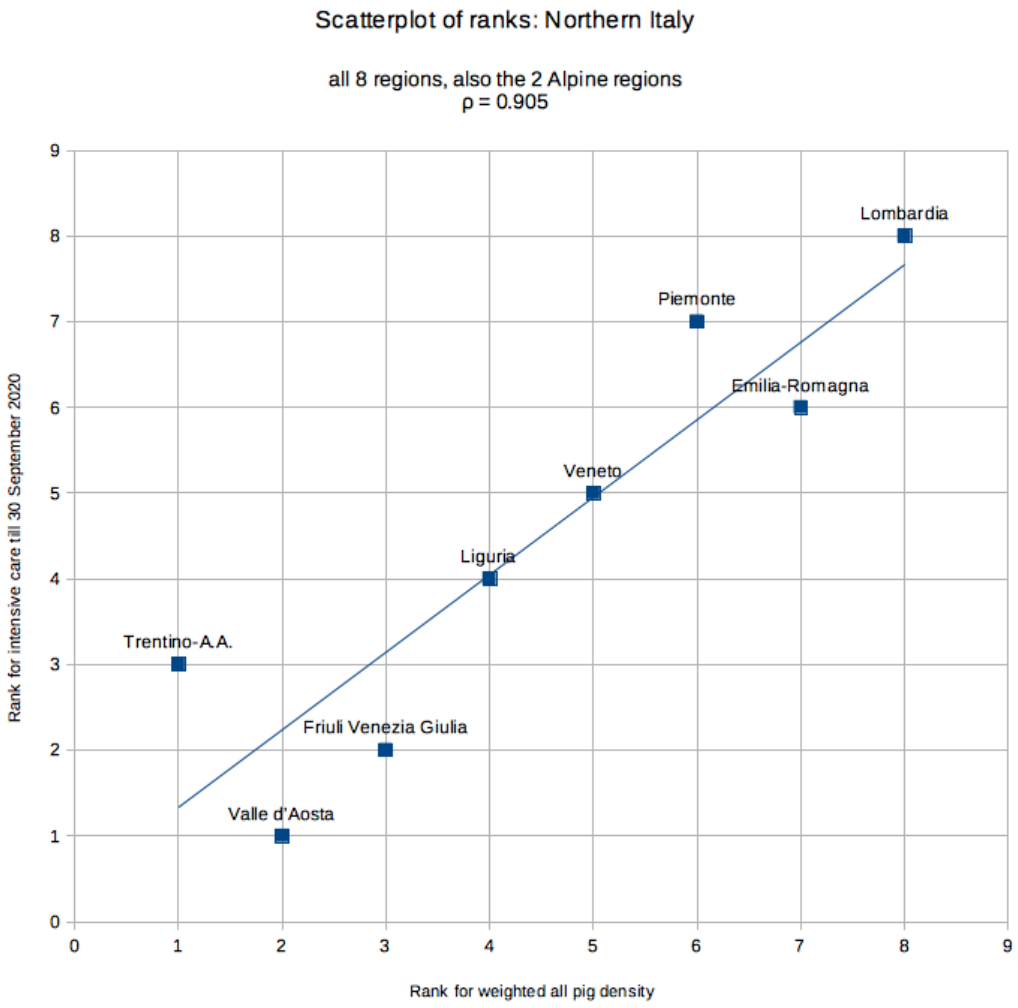
63.5 Alcune cose in Statistica Inferenziale

Abbiamo visto il Test dei Ranghi.

Per esempio, qua[192] vengono considerati dati sulle terapie intensive nella prima ondata del covid nelle 8 regioni dell'Italia Settentrionale, e la densità di tutti i maiali, d'allevamento, e – con una grossolana stima del loro impatto, e ciò si riferisce la parola

[192] A. Soranzo: In the first wave of the 2020 pandemic in several areas more sunlight less pandemic, more pigs more pandemic, and lower correlations with some other livestocks. (December 2020) ResearchGate, DOI: 10.13140/RG.2.2.29852.31367

"weighted" – selvatici ovvero cinghiali in libertà, comunque massivamente presenti solo in Liguria.



Leggiamo:

> To test the increasing monotonicity of ranks, with $n = 8$, Formula (...) gives 0.905 (which is in fact ρ_s) which grants, as may be seen in the table (...)

$$p < 0.005$$

La formula citata è la (165) e la tabella è nel paragrafo 59.2.

Altre cose viste, usate nella Matematica delle Epidemie, sono:
– Stimatori Puntuali
– Intervalli di Confidenza
– Errori di I e II specie
– Test del Chi Quadrato
– Test di Student.

PARTE II – ALCUNE COSE NUOVE

BOZZA - DRAFT

63.6 OR e RR

Una cosa molto interessante sono gli Odd Ratio e i Relative Risk, analoghi, ma il primo rivolto al passato e il secondo al futuro. Il lettore interessato troverà molto in rete. L'OR si potrebbe calcolare per esempio per la tabella 2×2 trovata nella Lezione sul chi quadrato d'indipendenza rilevando per maschi e femmine l'intenzione di restare a Farmacia o di cambiare Corso di Studio.

63.7 NNT: il numero di persone da trattare

> la penicillina salva le vite. Subito. E (quasi) sempre. E' un farmaco che si dà a malati spesso gravi e che guarisce con regolarità cronometrica. Tanto è vero che ha un NNT (https://en.wikipedia.org/wiki/Number_needed_to_treat) prossimo a 1 (uno) per praticamente tutte le sue indicazioni. Che vuol dire che il suo effetto si vede subito, in ogni malato, e ogni malato ne beneficia con certezza. Nessun atto di fede: la penicillina funziona e te lo dimostra.
>
> Per confronto, nelle condizioni migliori, cioè gli studi autorizzativi, i vaccini hanno un NNT di oltre 100, talora anche oltre 200. Ovvero vanno vaccinate 100-200 persone perchè almeno una di queste si risparmi un covid sintomatico di qualunque gravità. Il numero sale per le conseguenze gravi del covid. Per un decesso in meno i dati is-

> raeliani suggeriscono che vanno vaccinate 50-100.000 persone, questo negli adulti. Nei bambini il numero sale ulteriormente al punto da diventare difficilmente stimabile.
> (Marco Cosentino[193] (2023), Professore Ordinario di Farmacologia)

63.8 Il modello logistico e le loglet

Nell'evoluzione temporale quantitativa dei contagi di un'epidemia talvolta si possono riconoscere varie campane logistiche che si sommano, di differenti ampiezze e sfasamenti temporali. Questo è risultato molto evidente col covid-19 in Italia nel 2020-21, vedasi la Figura 66. L'effetto complessivo maschera alquanto le singole campane, le *loglet*.

DRAFT

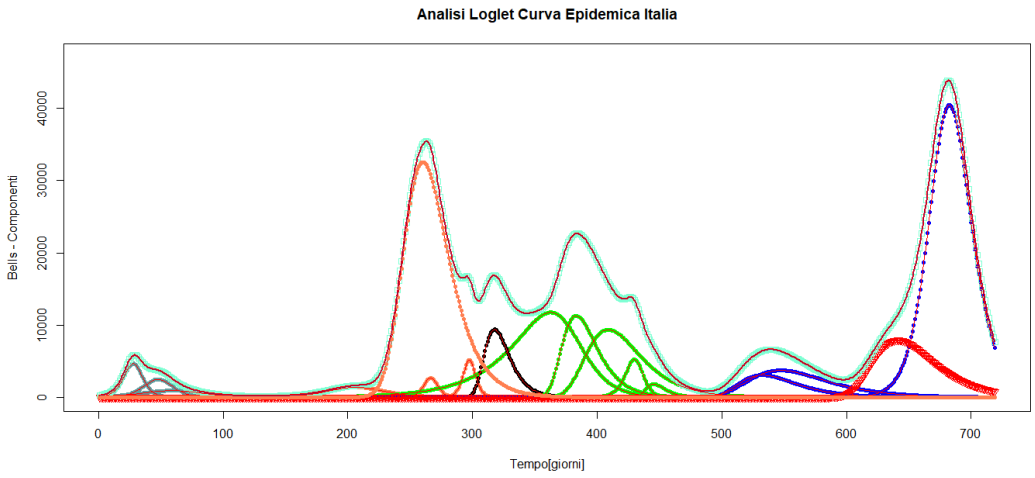


Figure 66: Loglet per contagi SARS-CoV-2, 2020-2021. By Carlo Del Vecchio.

È plausibile (ma non importa qua dimostrarlo) che la campana rossa nel grafico, a destra, sia dovuta essenzialmente alla variante delta del virus SARS-CoV-2, e quella ancora più a destra, tratteggiata in violetto, alla variante omicron. In effetti questo grafico è stato calcolato verso il 22 dicembre 2021 e quindi l'ultimo picco e la successiva discesa rappresentano una previsione, basata appunto sul modello delle loglet. (Per essere precisi nel calcolo, fatto con

[193] Testo assimilabile a *private communication*.

un software, sono stati considerati i “casi medi giornalieri" quindi mediati rispetto al weekend o alle anomalie).

Sulla *decomposizione multilogistica* si veda per esempio *Reproducing country-wide COVID-19 dynamics can require the usage of a set of SIR systems* (di Eugene B. Postnikov, 2021) in https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7797172/):

> the multilogistic decomposition of such epidemiological curves reveals components, which are quite close to the solutions of the SIR model in logistic approximations characterised by different sets of parameters including time shifts. This line of reasoning is confirmed by processing data for Spain and Russia in details and, additionally, is illustrated for several other countries.

BOZZA - DRAFT

63.9 La pandemia 2020-2021 – i numeri e le chiacchiere

Prima di tutto, in un tempo caratterizzato da fake news sulla pandemia (2020-2021) dai livelli più popolari ai livelli istituzionali più alti, segnaliamo alcuni SITI SERI dove reperire DATI e non OPINIONI e CHIACCHIERE.

Italia:
Istituto Superiore di Sanità, dipendente dal Ministero della Salute:
– al 16 dicembre 2020
https://www.epicentro.iss.it/coronavirus/bollettino/Report-COVID-2016_dicembre.pdf
– e in aggiornamento continuo
https://www.epicentro.iss.it/coronavirus/sars-cov-2-decessi-italia
https://www.epicentro.iss.it/coronavirus/sars-cov-2-dashboard

Europa: EuroMOMO:
https://www.euromomo.eu/graphs-and-maps

– qua finisce il mondo dei dati sicuri – *hic sunt dracones* –

Mondo: Worldometer:
https://www.worldometers.info/coronavirus/
Il database Worldometer è ben descritto da Wikipedia:

> Worldometer (...) is a reference website that provides counters and real-time statistics for diverse topics.
> (...)
> is managed by "an international team of developers, researchers, and volunteers"

Di fatto raccoglie i dati dei vari Stati al livello più ufficiale possibile, e di meglio non c'è, per il mondo intero. Ma i dati di molti Stati sono sicuramente fasulli comunque. Leggiamo[194] in un preprint:

> of course in no statistic about the epidemic, data of Turkmenistan may be used, a country declaring (till November 2020) no deaths, no cases, and ranking 180 in the *press freedom index* in 2020, among 180 countries.

BOZZA - DRAFT

63.10 L'arbitrarietà dei modelli previsionali

Leggiamo sul prestigioso Nature.com in https://www.nature.com/articles/s42254-020-0188-2.pdf (letto il 13 giugno 2020)

> At first, epidemiologists can only estimate the parameters that concern the disease itself, such as how readily it is transmitted, and subsequently update the estimations.
> (...)
>
> But even if every model parameter were well-quantified, most predictions won't come true. (...)
>
> In other words, epidemiological modelling is messy. It attempts to squeeze as much useful information as

[194] A. Soranzo: In the first wave of the 2020 pandemic in several areas more sunlight less pandemic, more pigs more pandemic, and lower correlations with some other livestocks. (December 2020) ResearchGate, DOI: 10.13140/RG.2.2.29852.31367

> possible from limited noisy data that describe the complex nonlinear system of an infectious disease interacting with human society. There are no controlled experiments. Thus, although the equations that describe a susceptible-infectious-recovered epidemiological model are simple to write down, forecasting how a disease will spread, even in the short term, is a difficult business.

Esempio. Vediamo un testo dei prestigiosi CDC relativo all'epidemia Covid-19 del 2020, https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios-h.pdf (letto il 20 giugno 2020), tutto pieno di numeri che senz'altro danno al profano l'impressione di grande scientificità.

BOZZA DRAFT

> CDC and the Office of the Assistant Secretary for Preparedness and Response (ASPR) have developed five COVID-19 Pandemic Planning Scenarios that are designed to help inform decisions by modelers and public health officials who utilize mathematical modeling. The planning scenarios are being used by mathematical modelers throughout the Federal government. Models developed using the data provided in the planning scenarios can help evaluate the potential effects of different community mitigation strategies (e.g., social distancing). (...)

Quindi scenari utizzabili per decisioni politiche ("mitigation strategies") di impatto enorme sulle persone (lockdown).
Continuiamo a leggere:

	scenario 1	**scenario 2**	**scenario 3**

.....

Infectiousness of asymptomatic individuals relative to symptomatic individuals Source: Assumption, ASPR and CDC	50%	100%	

L'enorme differenza fra le stime nei 2 diversi scenari mostra l'arbitrarietà dei parametri di base usati poi nelle famose precise equazioni – che saranno pure precise, ma con parametri di ingresso così diversi daranno risultati molto diversi. E d'altra parte, è ben detto che, nonostante testi di appoggio, sono semplicemente ipotesi (scrive "assumption").
Continuiamo a leggere:

Time to seek care (outpatient) Source: Survey of persons with Influenza like illness (ILI), CDC†	≤ 2 days: 35% 3 – 7 days: 50% ≥ 8 days: 25%

La somma delle percentuali è 110%. ☹
Come direbbe qualcuno scherzosamente, *bene ma non benissimo.*
E questi sono i CDC, considerati un riferimento mondiale.

BOZZA - DRAFT

63.11 Il tasso di letalità

Questo paragrafo amplia una questione già trattata

Vediamo ora un concetto fondamentale nella matematica delle malattie e quindi in particolare delle epidemie: il **tasso di letalità** di una malattia – talvolta riportato semplicemente come *letalità* della malattia.
Nella sostanza – ma si vedano le precisazioni successive – esso
è la probabilità di morire in caso di una certa malattia

$$P(muore|ha\,malattia)$$

come leggiamo[195] su Wikipedia:

> [indica] la probabilità di morire se si è afflitti da una specifica malattia

e ovviamente è necessario prefissare un intervallo di tempo.
Naturalmente questa probabilità è da intendere in senso frequentista, cioè in pratica, come leggiamo a quella stessa voce

> indica la proporzione, tipicamente percentuale, di decessi per una determinata malattia sul totale dei soggetti ammalati in un determinato arco temporale.

Per esempio, se muoiono 25 ogni 1000 malati, la letalità è 2.5% ovvero la probabilità di morire è del 2.5%, con tutti i limiti del concetto probabilistico.
Ancora ci avverte quel testo

> Tale valore è estremamente variabile in relazione alla durata dell'osservazione, per questo va contestualizzato in rapporto a un intervallo di tempo. (...) Si usa in particolar modo per le malattie infettive acute, mentre il suo utilizzo nelle condizioni croniche è meno indicativo considerando l'ampiezza della finestra temporale

BOZZA - DRAFT

In pratica, in via semplificata e con molti caveat,

$$\boxed{\text{letalità} = \frac{morti}{casi} = P(muore|malato)} \tag{170}$$

ma con l'avvertenza che in caso di malattia virale, e in particolare per la pandemia del covid-19, "malato" sarà da intendere nel senso di "caso", cioè in pratica contagiato, anche se asintomatico, e quindi non affatto ammalato. Per la pandemia del covid-19, in pratica, è il positivo al SARS-CoV-2, anche se asintomatico e quindi non affatto affetto da covid-19 (che presupporrebbe i sintomi clinici).

[195] Wikipedia, l'enciclopedia libera, alla voce *Tasso di letalità*, letto il 20 giugno 2020.

Nota. Senza voler qua fare Medicina, diamo però qualche indicazione sul fatto che

è un valore difficile da quantificare oggettivamente

perché dipende da chi si definisce affetto dalla malattia e chi no.
Consideriamo ad esempio una malattia virale.
Prima di tutto sgombriamo il campo dall'ipotesi ingenua che si conteggino i *malati* in senso *clinico*, le persone che effettivamente *stanno male* per la malattia considerata.
Internazionalmente si considera *caso* ("case") chi "ha il virus", rivelato da uno specifico test, non chi "sta male". Spesso sta benissimo ("asintomatico").
E qua si capisce bene che non c'è un confine netto fra chi ha il virus e chi non ce l'ha, a meno che non vogliamo considerare contagiato un soggetto che ha in corpo

2 virus soli, o 4 virus soli, o 8 virus, o 16 virus...

che magari non finiranno neppure sul *tampone* con cui si farà il test. È evidente che se – per pura ipotesi – si facessero 2 test a 1 soggetto che ha pochi virus, potrà benissimo succedere che in uno risulti positivo e nell'altro negativo. Non c'è un confine netto fra il *caso* e il *non caso*. E questo è solo un aspetto – di gusto diciamo così *matematico* – della più ampia problematica dei *falsi positivi* e *falsi negativi*.
Concretamente, nel pieno dell'epidemia Covid-19 del 2020 circolavano valori della letalità in Italia in un range vastissimo, forse un po' centrato su 1-2%, ma esteso almeno da 0.8% a 13.7%, a seconda delle fonti – un rapporto 1:17 circa fra minimo e massimo.[196]
D'altra parte, anche facendo la tara di tutte queste problematiche, non c'è dubbio che la letalità del covid-19 in Italia sia molto variata nel tempo, passando grosso modo da un 15% della primavera del 2020 a un 2% abbondante dell'inverno 2021-2022. Calcolare una letalità complessiva sui 2 anni certamente si può fare ma in pratica non significa nulla.

BOZZA - DRAFT

[196] https://www.primocanale.it/notizie/polemica-sui-numeri-del-coronavirus-bonsignore-non-ci-sono-html letto il 20 giugno 2020

Ma, per il singolo, la letalità stessa, anche se ben determinata temporalmente, non significa nulla, data la sua enorme variabilità nelle fase d'età: che muoia 1 contagiato su 50 è mostruosamente pessimistico per un ventenne e molto ottimistico per un novantenne.

Si noti comunque che non solo i diversi stati hanno usato diversi criteri per decidere chi è un caso, ma anche nella stessa Italia le regole sono cambiate ripetutamente, anche a livello locale. Solo un indizio – espressivo – su un argomento che meriterebbe più spazio:

> Riconteggiare i ricoveri per Covid, avendo cura di stralciare dai report i pazienti "negativi" seppur ancora in ospedale e quelli accolti in reparti non dedicati ai casi "acuti". Obiettivo (non scritto): tenere la Lombardia lontano dalle soglie d'allerta che fanno scattare il passaggio in fascia gialla e quindi le restrizioni al commercio e alla mobilità. L'indicazione arriva dalla direzione generale Welfare del Pirellone ed è indirizzata ai direttori generali delle Ats. https://milano.corriere.it/notizie/cronaca/21_novembre_24/covid-lombardia-ricoveri-cambia-conteggio-nshtml?refresh_ce

BOZZA - DRAFT

E poi c'è tutta la questione dei cicli di amplificazione del test: è ovvio che con 45 o 50 cicli si trovano positivi soggetti che non lo risulterebbero con 25; eppure da 2 anni (2020-2021) la questione non è ancora ben definita...

Tasso di letalità per classi di età.
Naturalmente si potrà considerare il tasso di letalità per singole classi di età:

$$P(muore|ha\,malattia\,et\,n < anni < m)$$

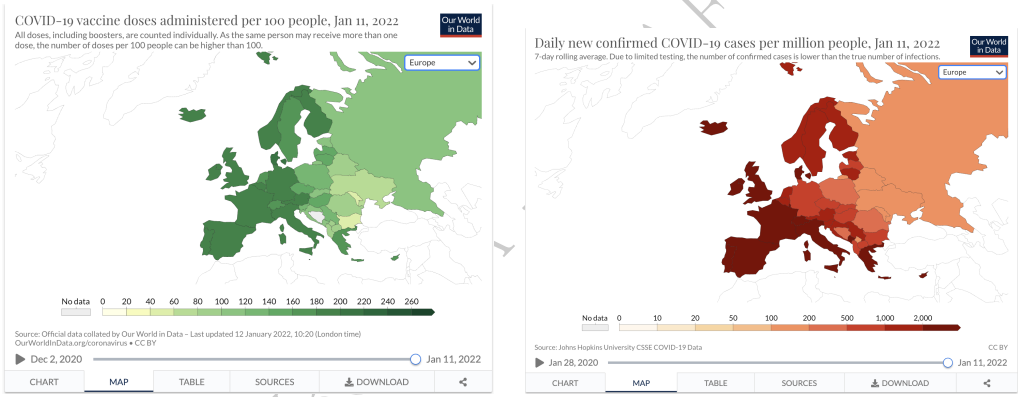
Per esempio, con riferimento alla prima ondata della pandemia in Italia nel 2020 (https://www.epicentro.iss.it/coronavirus/bollettino/Report-COVID-2019_22_luglio.pdf)

$$P(muore|positivo\,SARS-CoV-2\,et\,10 \leq anni \leq 19) = 0.000\,00$$

(Nessun morto fra i casi in quella classe di età, se ce ne fosse stato 1 fra i 34 142 morti si sarebbe avuto 0.000 03).

63.12 Mappe coropletiche

Strumento della Statistica Descrittiva particolarmente utile per la rappresentazione delle incidenze delle malattie e dei tassi di vaccinazione – e moltissime altre cose – sono le *mappe coropletiche*: in esse i dati sono rappresentati con scale di colori. (Però senza uno standard fisso sui colori).



Nelle 2 mappe coropletiche soprastanti sono rappresentati gli ultimi dati disponibili di tassi di vaccinazione e casi covid a livello europeo all'11 gennaio 2022, giorno precedente l'ultima Lezione del Corso di Matematica per Farmacia 2021-22 dell'Università degli Studi di Trieste, per il quale questa dispensa è stata realizzata.

Altre interessantissime mappe coropletiche a questo LINK.

Nota. Di questa Lezione ci si dovrebbe aspettare che lo studente conosca a memoria le

2

formule *contemporaneamente* numerate *et* riquadrate, oltre a tutto l'inquadramento teorico della questione.

* * *

BOZZA - DRAFT

64 Note finali sulla Statistica

> *Ci sono tre livelli di menzogna, via via più gravi: le bugie ufficiose, le bugie malevole, e la statistica.*

64.1 I numeri parlano da soli? Forse una lingua ignota

Si ha spesso l'idea che se di un fenomeno conosciamo i dati oggettivi, i valori numerici, veri, esatti, lo stiamo capendo. Non è così.

Leggiamo su prestigiosa rivista[197] scientifica, *Proceedings of the National Academy of Sciences of the United States of America*:

> This study explores how researchers' analytical choices affect the reliability of scientific findings. Most discussions of reliability problems in science focus on systematic biases. We broaden the lens to emphasize the idiosyncrasy of conscious and unconscious decisions that researchers make during data analysis. We coordinated 161 researchers in 73 research teams and observed their research decisions as they used the same data to independently test the same prominent social science hypothesis (...) research teams reported both widely diverging numerical findings and substantive conclusions despite identical start conditions. Researchers' expertise, prior beliefs, and expectations barely predict the wide variation in research outcomes. More than 95% of the total variance in numerical results remains unexplained even after qualitative coding of all identifiable decisions in each team's workflow. This reveals a universe of uncertainty that remains hidden when considering a single study in isolation. The idiosyncratic nature of how researchers' results and conclusions varied is a previously underappreciated explanation for why many scientific hypotheses remain contested. These results call for greater epistemic humility and clarity in reporting scientific findings.

BOZZA - DRAFT

64.2 Quale misurazione è più profittevole?

Sottolineiamo ancora una volta l'importanza di **cosa si misura** per valutare l'efficacia di un farmaco, o in generale di un trattamento sanitario. Per esempio, ridurre la massa tumorale (facile

[197] "Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty", by Nate Breznau et al., PNAS, Vol. 119, No. 44, October 28, 2022

da verificare) sarebbe certo una buona cosa, a meno che questo non *abbrevi* molto la durata della vita residua (parametro molto più lungo e quindi costoso da valutare), o magari *peggiori* la vita residua (parametro molto più difficile da valutare).

In questo si inquadra il discorso delle procedure accelerate di autorizzazione dei farmaci. Leggiamo in un articolo[198] scientifico:

> In 1992, Congress enacted the "accelerated approval pathway". Drugs and biologics expected to provide a meaningful advantage over available therapies for serious conditions have since been eligible for receiving accelerated FDA approval on the basis of surrogate measures. Surrogate measures are proxies for clinically meaningful outcomes and are "reasonably likely" to predict clinical benefit. An oft-cited example of an established surrogate measure is the magnitude of cholesterol reduction, which has been demonstrated to predict the risk of future fatal and nonfatal heart attacks.6 Such established surrogate measures are frequently used in regular FDA approvals. However, many surrogate measures used for accelerated approval decisions are not established and have a weak empirical association with important outcomes such as overall survival. Hence, agents granted accelerated approval on the basis of surrogate measures are required to conduct studies to confirm the anticipated clinical benefit.
>
> The objective of using surrogate measures as a basis for FDA-accelerated approval is to reduce the evidence requirements and considerably shorten the duration of clinical testing required prior to market entry for agents targeting important conditions. The bar for market entry is

BOZZA - DRAFT

[198] Naci H, Wouters OJ, Gupta R, Ioannidis JPA. Timing and Characteristics of Cumulative Evidence Available on Novel Therapeutic Agents Receiving Food and Drug Administration Accelerated Approval. Milbank Q. 2017 Jun;95(2):261-290. doi: 10.1111/1468-0009.12261. PMID: 28589600; PMCID: PMC5461381.

> therefore substantially lower for agents receiving accelerated versus regular approval. Clinical studies used as the basis of regulatory decisions in the accelerated approval pathway are relatively small, have shorter follow-up durations, often lack comparators, and are less likely to be randomized. Collectively, these study features effectively reduce the research and development timelines for agents in the accelerated approval pathway. According to a recent study, oncology agents receiving accelerated approval entered the market on average 4.7 years earlier than those receiving regular approval.
>
> (...) Therapeutic agents granted accelerated approval have premature data on their clinical benefits and harms at the time of market entry. Substantial research is thus needed to compensate for the limitations of the evidence base and generate meaningful, definitive data confirming clinical benefit following accelerated approval.

BOZZA - DRAFT

64.3 "I numeri che vi diamo parlano da soli"

Ci si tolga dalla testa che se un'affermazione è corredata da numeri (una "statistica", popolarmente) su quell'argomento è detta l'ultima parola, perché:

1) molto più probabilmente di come si potrebbe sperare, è ben possibile che i numeri siano inesatti, per errore o falsificazione;

2) altri numeri che non ci vengono detti (per ignoranza o dolo) potrebbero ribaltare completamente la questione.

Leggiamo in un articolo scientifico riportato su PubMed:

> "there is a good chance that statistical significance will be reached only by increasing the number of hypotheses

> tested in the work. The question is then: is this significant difference real or did it occur by pure chance?
>
> Actually, it is well known that if 20 tests are performed on the same data set, at least one Type 1 error (α) is to be expected. Therefore, the number of hypotheses to be tested in a certain study needs to determined in advance. If multiple hypotheses are tested, correction for multiple testing should be applied or study should be declared as exploratory."

(Il numero 20 corrisponde alla probabilità $5\% = 0.05 = \frac{1}{20}$).

Link->

BOZZA - DRAFT

64.4 Falsificazioni deliberate

Richiamiamo quanto detto nella Parte III della Lezione 23, dove anche sono stati esposti i dettagli giuridici della questione.

Diciamolo a chiare lettere: inviare ad una rivista scientifica un articolo scientifico in cui si parla di esperimenti su 1,500 persone, o topi, mai esistiti, non è reato. È come una bugia fra privati.

Nell'epidemia del 2020 ad un certo punto in Italia e altri Stati è stato (sostanzialmente) vietato l'uso dell'idrossiclorichina, che sembrava funzionare bene, e risalendo indietro nella catena di testi che hanno originato quel fatto c'è 1 singolo articolo scientifico, pubblicato su The Lancet, la più importante rivista scientifica medica del mondo. Esso stabiliva che quel farmaco non solo non serviva per l'epidemia in corso ma addirittura faceva male. Poi l'articolo è stato ritirato.

Citiamo l'agenzia di stampa Adnkronos, con un'enfasi aggiunta. https://www.adnkronos.com/fatti/cronaca/2020/06/05/coronavirus-lancet-ritira-studio-idrossiclorochina_GUFscvcw3DK1PNS4RHoAfL.html (letto l'11 giugno 2020)

> Pubblicato il: 05/06/2020 09:03
>
> Le riviste scientifiche 'The Lancet' e 'New England Journal of Medicine' hanno ritirato due studi basati su dati forniti dall'azienda Usa

> Surgisphere, finita sotto i riflettori del 'Guardian' che ne ha messo in dubbio l'attendibilità. Uno dei due articoli, quello pubblicato da Lancet, lanciava l'allarme su gravi rischi associati all'uso del farmaco antimalarico idrossiclorochina contro il Covid-19.
>
> L'altro, quello apparso sul Nejm, riguardava l'impiego di comuni medicinali antipertensivi nei pazienti con infezione da coronavirus Sars-CoV-2. Entrambi i lavori sono stati ritrattati su richiesta degli autori. "Non siamo più in grado di garantire l'attendibilità delle fonti dei dati", hanno spiegato scusandosi con i lettori e ritirando le firme.
>
> La notizia è rimbalzata nella tarda serata di ieri sui media internazionali. I primi a ritirare le loro firme sono stati 3 autori sui 4 dello studio pubblicato su Lancet: tutti i coautori del paper (Mandeep Mehra del Brigham and Women's Hospital di Boston, Frank Ruschitzka dello University Hospital di Zurigo e Amit Patel della University of Utah), eccetto il fondatore e Ceo di Surgisphere, Sapan Desai, si sono 'sfilati' con una nota congiunta diffusa dalla rivista scientifica. "Siamo entrati in questa collaborazione – hanno precisato – per contribuire in buona fede alla lotta contro la pandemia, in un momento di grande necessità".
>
> Poco dopo è arrivata la ritrattazione del Nejm.

BOZZA - DRAFT

Da The Guardian in https://www.theguardian.com/world/2020/jun/03/covid-19-surgisphere-who-wor
(letto l'11 giugno 2020) apprendiamo che (enfasi aggiunta)

- A search of publicly available material suggests several of Surgisphere's employees have little or no data or scientific background. An employee listed as a science editor appears to be a science fiction author and fantasy artist whose professional profile suggests writing is her fulltime job. Another employee listed as a marketing executive is an adult model and events hostess, who also acts in videos for organisations.
- The company's LinkedIn page has fewer than 100 followers and last week listed just six employees. This was changed to three employees as of Wednesday.

Si veda anche https://www.ilfattoquotidiano.it/2020/06/04/coronavirus-the-lancet-avvia-uninch
5824478/

64.5 Il (doppio) cieco che ci vede benissimo

In https://www.ilfattoquotidiano.it/2014/04/12/farmaci-e-studi-clinici-lo-scandalo-t
949196/?fbclid=IwAR2lsD6W_GaRYkihu_rpL2DY6JyNyrWkW6IfNaaWxifAbJ01d6B6OGSHYMw
(letto il 28 febbraio 2020) troviamo:

> A proposito del ((nome di farmaco)), Goldacre ha rivelato i retroscena della vicenda in un articolo uscito sul Guardian il 10 aprile. Sintetizzo alcuni passaggi che vale la pena sapere.

> Un pediatra giapponese, Keiji Hayashi, lascia un commento sul sito online del gruppo Cochrane (14 mila medici e ricercatori indipendenti che difendono la trasparenza dei dati scientifici nel campo medico e farmaceutico) in cui spiega che la presunta utilità del ((nome di farmaco)) si basa su un riassunto di dieci studi clinici elaborato dalla stessa industria che lo produce, la ((casa farmaceutica)). Ma di questi dieci test, soltanto due sono disponibili nella letteratura scientifica. Degli altri otto le uniche informazioni che si hanno sul metodo usato sono riconducibili al riassunto fatto in casa. E questo non è abbastanza credibile.
>
> Cochrane chiede alla ((casa farmaceutica)) di inviargli i dati mancanti. ((casa farmaceutica)) accetta a patto che la Cochrane reviews, la rivista online del gruppo, firmi un accordo informale che la obblighi a non riferire niente ai lettori. I termini dell'accordo non sono discutibili. Tom Jefferson, il responsabile di pneumologia dell'organizzazione, chiede alla ((casa farmaceutica)) perché c'è bisogno di sottoscrivere un contratto ma non riceve risposta.
>
> Nel 2009 ((casa farmaceutica)) invia sette documenti, di una dozzina di pagine ciascuno, con gli estratti dei dieci studi clinici precedentemente riassunti. Ma è ancora troppo poco: le richieste di Cochrane continuano a rimanere disattese.
>
> Cosa si evince intanto? Che il campione di cavie umane che hanno partecipato agli studi clinici non è abbastanza rappresentativo. E negli esperimenti in "doppio cieco" – quando nè il dottore nè il paziente dovrebbero sapere se si è assunto il placebo o il farmaco – **la pillola placebo e quella vera hanno colori diversi**.

(Enfasi aggiunta).

Non c'è da stupirsi che certa parte dell'opinione pubblica abbia sospetti su certi prodotti farmaceutici, con pillole per test in doppio **cieco** di colori diversi, e un'infinità di altre, diciamo così, problematiche, ben note in ampi settori di consumatori particolarmente attenti ed esigenti, che vogliono prodotti sicuri e garantiti da ricerche scientifiche e non da pubblicità.

Per riconquistare la fiducia del grande pubblico – necessaria per la salute generale della popolazione – sarebbe necessario un drastico repulisti.

BOZZA DRAFT

Ma osta sicuramente la questione seguente.

64.6 Il conflitto di interessi

"Oste, è buono il vino?" "Buonissimo!"

In matematica 2+2 fa sempre 4, e il Teorema di Pitagora è *eternamente* vero.
Ma la Farmacia e la Medicina non sono Matematica, le certezze assolute quasi non esistono, in generale sono temporanee.
Comunque anche per cose di Medicina o Farmacia vere come 2+2 = 4, persone opportunamente pagate – anche oneste – magari senza negare quelle verità, possono insistentemente parlarvi (è il loro lavoro, ripetiamo, onesto) del fatto che 3+3 fa 6, o qualunque altra cosa. E altre persone – talvolta meno oneste – possono usare tutto l'armamentario di tecniche manipolative dei dati illustrate nel Capitolo 23 – quando non le pure e semplici *falsificazioni scientifiche* – e così orientare l'opinione pubblica e di conseguenza le politiche sanitarie e farmaceutiche degli Stati.

BOZZA DRAFT

È il problema del **conflitto di interessi**, enorme in Farmacia.

Il mercato del farmaco mondiale è dell'ordine dei 12 zeri.

L'OMS propaganda prodotti farmaceutici di uno[199] che la finanzia, anzi che, negli ultimi anni, è stato uno dei principali finanziatori, addirittura eventualmente secondo solo agli Stati Uniti.

> "much of the scientific literature, perhaps half, may simply be untrue. Afflicted by studies with small sample sizes, tiny effects, invalid exploratory analyses, and flagrant conflicts of interest, together with an obsession for

[199] Si veda per esempio questo servizio (2020) di Rai 3: LINK ->

> pursuing fashionable trends of dubious importance, science has taken a turn towards darkness."

(Richard Horton, Editore di The [200]Lancet: https://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736%2815%2960696-1.pdf)

Si riveda la Nota basale sulla Statistica.

E si ricordi che:

– praticamente mai ci danno tutti i dati

– con un'opportuna scelta fra i dati, si possono far vedere tante

cose, a scelta.

[200]The Lancet è spesso considerata la più importante rivista di Medicina del mondo.

64.7 Come avviene l'inversione di significante

È usuale che i titoli dei giornali affermino l'opposto della verità.
La verità progressivamente si corrompe in questa trafila, dai fatti ai titoli mediatici:

1) verità fattuale
2) dati rilevati
3) analisi statistica dei dati
— articolo scientifico (4)
(4a) testo
(4b) abstract
(4c) titolo
— divulgazione di livello scientifico (5)
(5a) testo
(5b) sottotitolo o sommarietto
(5c) titolo
— divulgazione mediatica (6)
(6a) testo
(6b) sottotitolo
(6c) titolo – il punto più lontano dalla verità fattuale.

BOZZA - DRAFT

Su questo argomento si potrebbe scrivere un intero libro.

Diamo solo un esempio, di enorme modificazione di significante già nel passare dal punto (5a) al punto (5c) della soprastante lista.
Su sito divulgativo di livello scientifico, dell'Ordine dei Medici Chirurghi e Odontoiatri della Provincia di Forlì Cesena, leggiamo il titolo

Vaccini Moderna e Pfizer: protezione per anni dal covid

del 7 luglio 2021, da poco iniziate le vaccinazioni di massa contro il covid-19. Questo spettacolare risultato, "protezione per anni dal covid", annunciato nel titolo (livello 5c) viene attenuato con un *se* nel testo (livello 5a):

"se questo patogeno con le sue varianti non si evolverà molto oltre la sua forma di oggi."

Comunque il tutto verrebbe, ci dicono, da un articolo scientifico di "un team di scienziati della Washington University School of Medicine di St Louis, negli Usa, in uno studio pubblicato su 'Nature'".
Ecco l'articolo scientifico, che si trova facilmente sulla rete in
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8647067/

Turner, J.S., O'Halloran, J.A., Kalaidina, E. et al. SARS-CoV-2 mRNA vaccines induce persistent human germinal centre responses. Nature 596, 109–113 (2021). https://doi.org/10.1038/s41586-021-03738-2
Articolo il cui titolo – livello (4c) – ovviamente, come si vede, dice ben diversamente. E nell'abstract – livello (4b) – parla di 6 mesi, in alcuni vaccinati:

> "the GC B cell response persists in the draining lymph nodes for at least 6 months in some individuals following vaccination with SARS-CoV-2 mRNA-based vaccines.

BOZZA - DRAFT

64.8 I dati? Quali dati? Postilla

Già abbiamo detto della catastrofica situazione dei dati in Italia, pochi e spesso di fatto secretati.

Consideriamo ora[201] gli USA. Là i cittadini hanno diritto di conoscere i dati esistenti:

> The 1967 FOIA law requires federal agencies to respond to information requests within 20 business days. However, the time it takes to actually get the documents "will vary depending on the complexity of the request and any backlog of requests already pending at the agency," according to the government's central FOIA website

Ma talvolta è necessario pazientare parecchio per ricevere i dati:

> (Reuters) - Freedom of Information Act requests are rarely speedy, but when a group of scientists asked the federal government to share the data it relied upon in licensing Pfizer's COVID-19 vaccine, the response went beyond typical bureaucratic foot-dragging.
>
> As in 55 years beyond.
>
> That's how long the Food & Drug Administration in court papers this week proposes it should be given to review and release the trove of vaccine-related documents responsive to the request.

[201] letto il 7 dicembre 2024 in https://www.reuters.com/legal/government/wait-what-fda-wants-55-years-process-foia-request-over-vaccine-data-2021-11-18/?fbclid=IwY2xjawHBnr9leHRuA2FlbQIxMAABHRA7CzANxFgl0zVwjae8jchbFe16OkfCF4WNAaUJ_tlXx6LobpuXbjp46w_aem_yqY9Q7KQSYMWZF1egZ4Upg

64.9 Bottom line – Legge pseudo-magica della Statistica

Se c'è un interesse che un dato si formi quel dato tenderà a formarsi

Se una nuova legge prevedesse che dal prossim0 anno ci saranno finanziamenti per i *comuni d'alta montagna*, definiti per esempio dall'avere il municipio sopra i 1000 metri, molti comuni potranno ottenere la qualifica... traferendo la sede comunale in un edificio più a monte nello stesso comune.

Il 9 dicembre 2024, nell'ultima lezione di Matematica per Farmacia a Trieste, che ha una frequenza obbligatoria di almeno il 75% delle lezioni, al momento della registrazione delle presenza mediante codice telematico in aula c'erano 31 Studenti e 6 erano online, e in pochi istanti il numero dei presenti ha raggiunto il numero... 42.
Anche per quelli che non hanno prodotto il dato (la qualifica di seguitore della Lezione) ricevendo da qualcuno un messaggio col codice da digitare, il dato va considerato dubitativamente:
– Ognuno in aula, veramente seguiva o qualcuno faceva websurfing?
– Ognuno a casa, veramente seguiva? (V. sopra, aggravatamente).
– Poco dopo la registrazione delle presenze, alcuni Studenti hanno abbandonato l'aula.

Il grandissimo epidemiologo J.P.A. Ioannidis ha pubblicato sulla prestigiosa rivista PLOS Medicine l'articolo scientifico

Why Most Published Research Findings Are False LINK->

BOZZA - DRAFT

65 Le serie numeriche

65.1 Introduzione alle serie

Nella Lezione sulle funzioni trigonometriche abbiamo posto una questione interessantissima: come fanno i software a calcolare quei valori, e come fecero i nostri antenati nei secoli passati a calcolare i valori scritti nelle tavole cartacee.

La risposta sta nelle serie, che andiamo ad esporre.

Delle funzioni trigonometriche diamo solo la serie del seno:

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} + \ldots + \frac{(-1)^n x^{2n+1}}{(2n+1)!} + \ldots$$

(Il primo termine si ottiene per $n = 0$). Naturalmente nessuno può effettivamente sommare gli infiniti termini di questa serie, ma in pratica basta usare una certa quantità dei primi addendi per avere una buona approssimazione.

BOZZA DRAFT

Precisamente, quando calcolata per un numero $x > 0$, la soprastante serie (di funzioni e più in particolare di potenze) diventa una *serie numerica* che ha queste 2 caratteristiche:

– segni alternati
– termini decrescenti in valore assoluto.

Con queste 2 condizioni, l'errore assoluto dell'approssimazione ottenuta dal troncamento della serie è minore del valore assoluto del primo termine che si trascura.

Per esempio il seno di 1 radiante

$$\sin 1 \approx 1 - \frac{1}{6} \approx 0.833$$

con errore assoluto minore di 1/5! a sua volta minore di 0.01.

Essenzialmente le serie sono in qualche modo delle "somme infinite". Da definire.

Estenderemo l'uso del simbolo di sommatoria $\sum$ dal caso finito a quello con infiniti termini: essenzialmente, vogliamo occuparci di $a_0 + a_1 + \ldots + a_n + \ldots$ con infiniti termini.

65.2 Il calcolo del logaritmo di 2 della Farmacocinetica

Come abbiamo visto, nella Farmacocinetica appare la costante $\ln 2$, il logaritmo naturale di 2, solitamente approssimato a 0.7, ma più precisamente 0.693... Quel numero viene dato fondamentalmente da questa serie

$$\ln 2 = 1 - \frac{1}{2} + \frac{1}{3} - ... + \frac{(-1)^{k+1}}{k} + ...$$

che si chiama serie armonica a segni alterni, e anche da serie più "efficienti", nel senso che danno maggior precisione con meno termini.

BOZZA - DRAFT

65.3 Definizioni e notazioni

In questa trattazione elementare

serie è una scrittura come

$a_0 + a_1 + ... + a_k + ...$

oppure equivalentemente

$\sum_{k=0}^{\infty} a_k$ e in entrambi i casi

il primo indice può essere > 0

(171)

L'indice di sommatoria può chiamarsi n oppure k come di solito si usa, o con qualunque altro nome di variabile.

Esempio.

$$1 - \frac{1}{3!} + \frac{1}{5!} + ... + \frac{(-1)^k}{(2k+1)!} + ...$$

che come abbiamo visto prima permette di calcolare (approssimatamente ma con quanta precisione si voglia) $\sin(1)$.

Definizione di somma di una serie.

> Somma della serie è il numero,
>
> se esiste (finito), a cui tende
>
> la somma dei primi termini

(172)

in simboli

$$\sum_{k=0}^{\infty} a_k := \lim_{n\to+\infty} (a_0 + a_1 + ... + a_n) \quad \text{se } \exists \text{ finito} \tag{173}$$

ed estensioni varie possono essere fatte per limiti infiniti.

Questa formula (173) ci dice che la *somma della serie* al primo membro è il **limite**, se esiste (finito o infinito) della

ridotta o *somma parziale* $a_0 + a_1 + ... + a_n$. ← (normale) numero.

BOZZA - DRAFT

Esempio.

$$\sum_{k=1}^{\infty} 3 \cdot 10^{-k} =$$

$$= 0.3 + 0.03 + 0.003 + 0.0003 + ... =$$

$$= 0.33333.... =$$

$$= \frac{1}{3}$$

e già qua ci togliamo il dubbio: sì, può ben essere che aggiungendo indefinitamente quantità positive, non si superi un tanto, e anzi si converga a un limite finito.

Nota sull'ambiguità notazionale. Le 2 scritture equivalenti

$$\sum_{k=0}^{\infty} a_k \qquad \text{e} \qquad a_0 + a_1 + ... + a_n + ... \tag{174}$$

denotano sia la *serie* (scrittura), che l'eventuale sua *somma* (numero oppure $+\infty$ oppure $-\infty$).

65.4 Il numero di Nepero da una serie esponenziale

Il numero e che abbiamo già visto, base dei logaritmi naturali e valore di un Limite Fondamentale, è

$$\mathrm{e} = \frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \ldots \approx 2.718$$

e adesso siamo in grado di comprenderlo come somma di una serie che lo definisce. (In generale, in Analisi Matematica viene invece definito da un Limite Fondamentale, e solo dopo si dimostra la sua serie).

65.5 Il numero pigreco: la serie ciclometrica

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} + \ldots + \frac{(-1)^k}{2k+1} + \ldots$$

65.6 La serie armonica

La *serie armonica*

$$\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{k} + \ldots \text{ ovvero } \sum_{k=1}^{\infty} \frac{1}{k} \tag{175}$$

non ha somma finita (anzi si dice che *diverge a* $+\infty$.)

Per la serie geometrica si veda il paragrafo 24.9.

Nota. Di questa Lezione ci si dovrebbe aspettare che lo studente conosca a memoria le formule *contemporaneamente* numerate *et* riquadrate, oltre a tutto l'inquadramento teorico della questione.

BOZZA - DRAFT

* Complementi *

65.7 Complementi – Proprietà delle serie

• Serie *convergente*, con somma s: il limite delle somme parziali è il numero $s \in \mathbb{R}$. Per esempio la *serie ciclometrica*.

$$1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} + ... + \frac{(-1)^k}{2k+1} + ... = \frac{\pi}{4} . \tag{176}$$

• *Serie non convergenti*: quelle *indeterminate* ($\nexists$ limite) come

$$\sum_{k=0}^{\infty} (-1)^k = 1 - 1 + 1 - 1 + ... + (-1)^k + ... \tag{177}$$

(la somma parziale vale alternatamente 1 e 0), quelle *divergenti a* $+\infty$ come $1 + 1 + ... + 1 + ...$ e quelle *divergenti a* $-\infty$.
⋆ Serie *a termini positivi*: $a_k > 0$ per ogni k. (Similmente le serie a termini *non negativi* se ≥ 0, e negativi, e non positivi se ≤ 0).
⋆ Serie *a termini di segno alternato*, con ovvio significato.

BOZZA - DRAFT

• *Serie geometrica di ragione* $r \in \mathbb{R}$, fissiamo le idee con $a > 0$:

$$a + ar + ar^2 + ... + ar^k + ... = \sum_{k=0}^{\infty} ar^k = \begin{cases} \frac{a}{1-r} & se\ |r| < 1 \\ +\infty & se\ r \geq 1 \\ \nexists & se\ r \leq -1 \end{cases} \tag{178}$$

Nel terzo caso la serie è indeterminata. Per $a<0$ vale il Teor. 1.

• *Serie esponenziale.* $1 + a + \frac{a^2}{2} + ... + \frac{a^k}{k!} + ... = e^a$. (Si dimostra).

• La *serie armonica*, prima vista: diverge a $+\infty$.

Teorema 1. $ca_{n_0} + ca_{n_0+1} + ... + ca_k + ... = c(a_{n_0} + a_{n_0+1} + ... + a_k + ...)$ con l'ovvio significato se la serie in parentesi diverge a $+\infty$.

Teorema 2. Se $a_k \not\to 0$ allora la serie è non convergente.

Teorema 3. Se ogni $a_k \geq 0 \Rightarrow$ la serie converge o diverge a $+\infty$.

Teorema 4. (Di Leibniz). Se in una serie a termini di segno alternato $|a_k| \to 0$ e $|a_k|$ è decrescente anche solo in senso debole, cioè $|a_{k+1}| \leq |a_k|$, allora la serie converge.

In particolare $1 - \frac{1}{2} + \frac{1}{3} - \ldots + \frac{(-1)^{k+1}}{k} + \ldots = \ln 2$, *serie di Leibniz.*

BOZZA - DRAFT

ESERCIZI SULLA LEZIONE 65

ESERCIZIO$_{\mu}$ $\star \approx$ Calcolare $\Sigma_{n=1}^{\infty} 0{,}3^n$.

SVOLGIMENTO

Viene usato lo standard della virgola decimale.

Ricordando la somma della serie geometrica di ragione a fra 0 e 1 esclusi

$$\sum_{k=0}^{\infty} a\, r^k = \frac{a}{1-r} \qquad \Big(a \in \mathbb{R}, \quad 0 < r < 1\Big)$$

si ha, con $a = 1$ e $r = 0{,}3$,

$$\sum_{k=0}^{\infty} 0{,}3^k = \frac{1}{1-0{,}3} = \frac{1}{1-\frac{3}{10}} = \frac{1}{\frac{10-3}{10}} = \frac{10}{7}$$

che è proprio quanto dobbiamo calcolare salvo il termine 0-esimo, e allora

$$\sum_{k=1}^{\infty} 0{,}3^k = -0{,}3^0 + \sum_{k=0}^{\infty} 0{,}3^k = -1 + \frac{10}{7} = \frac{-7+10}{7} =$$

$$\boxed{\tfrac{3}{7} \approx 0{,}428\,571}$$

BOZZA - DRAFT

66 APPENDICE – Esperienze statistiche in aula

BOZZA - DRAFT

66.1 Questionario per esperienze statistiche in aula

1) Genere M/F

2) Quanti animali domestici possiedi? Nel senso di cani, gatti e altri mammiferi, canarini, pappagalli e altri uccelli, pesci rossi e altri pesci. (Non insetti, eccetera). "Possiedi" è da intendere che li ritieni tuoi (anche in condivisione con familiari) e dormono di notte in casa tua. (Non quindi gatti o piccioni a cui dai da mangiare in strada o giardino).

3) Di quanti ultimi anni andando indietro nel tempo conosci la nazione vincitrice dei campionati mondiali di calcio?

4) Quanto costa, approssimato all'euro intero più vicino, il regalo che più vorresti ricevere per Natale? (Oggetto materiale, oppure viaggio/vacanza, o prestazione d'opera come colorazione dei capelli...)

BOZZA - DRAFT

66.2 Lez. 55 – Test simile a quello della signora del tè

Un'esperienza statistica che si avvicina a riprodurre il famoso esperimento della signora del tè è questo il seguente.

Sa in soggetto riconoscere le parole cinesi da quelle giapponesi, quando traslitterate come d'uso, salvo l'omissione degli accenti?

Lo scrivente ha posto a ChatGPT una richiesta di questo tipo per il cinese, e poi similmente per il giapponese:

Mi scrivi 4 nomi di animali in cinese, ma non banali come cane e gatto, bensì di più raro uso?

Le parole sono state private degli accenti e numerate da 1 a 8.
Poi ha fatto questa richiesta:

Mi scrivi una permutazione casuale dei numeri da 1 a 8?

Con la permutazione ha potuto mescolare le 8 parole straniere così:

rakuda
bianfu
kongkue
komori
kujaku
xiniu
kujira
haitun

Se il soggetto trova le 4 parole cinesi (e quindi le 4 parole giapponesi) il test respinge quest'ipotesi nulla:

H_0: *il soggetto non sa distinguere le parole giapponesi da quelle cinesi, quando traslitterate come d'uso salvo l'omissione degli accenti.*

BOZZA – DRAFT

Le parole cinesi erano:

Kǒngquè - Pavone
Xīniú - Rinoceronte
Biānfú - Pipistrello
Hǎitún - Delfino

Le parole giapponesi erano:

Kujaku - Pavone
Kujira - Balena
Rakuda - Cammello
Kōmori - Pipistrello

La permutazione dei numeri da 1 a 8 era:

7, 3, 1, 8, 5, 2, 6, 4.

Similmente si potrà vedere se qualcuno sa distinguere le lingue aborigene australiane Arrernte e Noongar, di cui abbiamo le traduzioni di cane, cavallo, serpente, formica.
Arrernte:

Cane: Irreme

Cavallo: Apwerte arlkweme

Serpente: Tyerrtye

Formica: Tyape

Noongar:

Cane: Dwert

Cavallo: Boornbort

Serpente: Ngarlak

Formica: Nyurnma

Si riordini con la permutazione 5 7 3 6 1 2 8 4.
Su un centinaio di soggetti è ben possibile che qualcuno... indovini.

BOZZA - DRAFT

66.3 Lez. 52 – Stima del secondo estremo dell'uniforme

Si sceglie una città e una via della città con molti numeri civici, per esempio via Fabio Severo a Trieste, e si chiedono agli Studenti alcuni numeri civici per qualche motivo a loro noti: si può così stimare il massimo numero civico di quella via. Ci sono notevoli imprecisioni nel modello: i numeri civici più noti non sono distribuiti uniformemente; inoltre la variabile aleatoria che stiamo ipotizzando in effetti non esiste negli esatti termini detti perché nelle strade non c'è una numerazione da 1 a $n_{\max}$ bensì 2 numerazioni indipendenti dei numeri pari e dei numeri dispari. Nondimeno il risultato che si ottiene è molto meglio che niente. A Trieste in aula a CTF nel 2023 per via Fabio Severo sono stati proposti dagli Studenti 3 numeri, di cui 1 è stato riconosciuto erroneo (outlier da errore), e coi 2 restanti si è ottenuto 154, sicuramente non male.

BOZZA - DRAFT

66.4 Caccia al Tesoro: usare AI e rete

Scrivere un testo di 200 parole sulle relazioni fra scienze farmaceutiche e statistica

Produrre un'immagine digitale con vistosa reazione chimica

Disegnare una spirale di Archimede

Trovare come si chiamava l'intelliigenza artificiale sepolta in units

Dove fu inventata l'elica delle navi?

Il problema non è che le persone siano ignoranti. Il problema è che le persone sono istruite quel tanto che basta per credere a ciò che è stato loro insegnato e non abbastanza istruite per mettere in dubbio qualsiasi cosa da ciò che è stato insegnato loro."

(Richard Feynman – Nobel per la Fisica)

BOZZA - DRAFT