

Appunti del corso SF1-MZ
Teoria di Statistica
Laboratorio di Fisica
aa 2021/22

Michele Doro

28 marzo 2022

Quest'opera è distribuita con licenza Creative Commons "Attribuzione
– Non commerciale – Condividi allo stesso modo 3.0 Italia".



NOTA IMPORTANTE:

Il testo e' in fase iniziale di scrittura e potrebbe contenere diverse imprecisioni e refusi, di cui mi scuso fin da subito, inclusi errori grammaticali. Non mi e' infatti sempre possibile rileggere tutto quanto scritto e modificato. Potrebbe addirittura contenere, cosa ben più grave, errori. Chiedo agli studenti che leggeranno questi appunti di informarmi immediatamente su eventuali ovvi errori o distorsioni rispetto alla lezione in classe.

Michele Doro michele.doro@unipd.it 4 Ottobre 2021

Indice

Introduzione	1
I STATISTICA DESCRITTIVA	5
I.I Metrologia ed Incertezza della misura	7
1 Metrologia e Misura	7
1.1 Grandezze e Misure in fisica	7
1.2 Incertezza ed errore nella misura	9
1.3 Accuratezza e Precisione	12
1.4 Risoluzione e Sensibilità	14
1.4.1 Altri concetti metrologici	15
1.5 Commenti Finali	15
2 Presentazione misure	19
2.1 Cifre significative	19
2.2 Istogrammi	21
2.3 Grafici	24
2.3.1 Andamento temporale delle misure	24
2.3.2 Andamento cumulativo delle misure	25
2.3.3 Grafico di una variabile contro un'altra	26
2.4 Concetti Finali	26
3 Misure e Gaussiana	31
3.1 Misure Singola	32
3.2 Poche misure	33
3.3 Misure Ripetute	34
3.3.1 Terminologia sulle incertezze	36
3.3.2 Altre stime di media	37
3.4 La curva Gaussiana	39
3.5 Commenti Finali	42
I.II Statica Descrittiva di Misure	45
4 Teoria della Probabilità	47
4.1 Evento Aleatorio	47
4.2 Definizioni di probabilità	48
4.3 Operazioni logiche tra eventi aleatori	50
4.3.1 Dipendenza e indipendenza logica, probabilità condizionata	50
4.3.2 Legge della Probabilità Totale, Marginalizzazione	53
4.3.3 Teorema di Bayes	54

4.4	Calcolo Combinatorio	55
4.5	Commenti Finali	57
5	Distribuzioni di probabilità di variabili aleatorie	63
5.1	Variabile aleatoria	63
5.2	Funzioni di massa di probabilità (PMF)	64
5.2.1	Distribuzione discreta uniforme \mathcal{U}	64
5.2.2	PMF discreta (binomiale) di Bernoulli	65
5.2.3	PMF discreta binomiale \mathcal{B}	66
5.2.4	PMF di Poisson \mathcal{P}	67
5.3	Funzioni di densità di probabilità: PDF	69
5.3.1	PDF continua uniforme \mathcal{U} e PDF continua triangolare \mathcal{T}	69
5.3.2	PDF Esponenziale	70
5.3.3	PDF Gaussiana	72
5.3.4	Altre PDF	72
5.4	Funzioni cumulative di probabilità	73
5.4.1	CDF Gaussiana	74
5.5	Momenti delle distribuzioni	75
5.5.1	Media di una PMF/PDF	76
5.5.2	Varianza di una PMF/PDF	77
5.5.3	Momenti di ordine superiore	78
5.6	Quantili delle distribuzioni	80
5.7	Commenti finali	81
6	Relazioni tra variabili aleatorie	87
6.1	Congiunzione logica, Indipendenza e Correlazione tra variabili aleatorie	87
6.1.1	Covarianza e correlazione lineare	90
6.2	PF di funzioni di variabili	92
6.2.1	PF di funzioni di una variabile	92
6.2.2	Caso bivariato e multivariato	94
6.3	Momenti di combinazioni lineari di variabili aleatorie	95
6.4	Formula della propagazione delle incertezze	96
6.4.1	Esempi di propagazione delle incertezza statistiche	99
6.5	Commenti Finali	100
7	Statistica descrittiva dei campioni di misure	103
7.1	Media e altre stime centrali	104
7.1.1	Deviazione standard della media campionaria	105
7.1.2	Varianza e deviazione standard campionaria	105
7.1.3	Incetezza su varianza e deviazione standard campionarie	108
7.1.4	Asimmetria, Curtosi, quantili e mediana dei campioni	108
7.2	Correlazione campionaria	109
7.2.1	Esempi	110
7.3	Propagazione delle incertezze campionarie	112
7.3.1	Esempio	113
7.4	Note aggiuntive	114
II	STATISTICA INFERENZIALE	121
8	Il metodo dei minimi quadrati e la interpolazione (lineare)	123
8.1	Minimi quadrati	123
8.2	Interpolazione lineare	126
8.2.1	Un esempio numerico	128

8.3	Interpolazione caso generale	129
8.4	Errore a posteriori	130
8.5	Correlazione tra i parametri di interpolazione	130
9	Stima dei Parametri con la Massima Verosimiglianza	133
9.1	La funzione di verosimiglianza	133
9.2	Stima di parametri	135
9.2.1	Combinazione di verosimiglianze	140
9.3	Performance di uno stimatore	140
9.4	Incertezza di una stima	141
9.4.1	Limite gaussiano della verosimiglianza	141
9.4.2	Incertezza dal profilo di verosimiglianza	142
9.5	Media Ponderata	143
9.6	Chiusura	144
9.6.1	Quesiti	144
9.6.2	Approfondimenti	145
10	Il Test del χ^2	149
10.1	La PDF del χ^2	149
10.2	La costruzione del test del χ^2	151
10.3	Esecuzione del test del χ^2	153
10.3.1	Interpretazione del test	154
10.3.2	Un esempio: interpolazione lineare	154
10.3.3	Compatibilità tra due stime	156
10.3.4	Un valore di riferimento comune	157
10.4	Limite pratico del test del χ^2	157
10.5	Considerazioni finali	157
11	Test χ^2 Pearson e il Test di Student	161
11.1	Test del χ^2 di Pearson	161
11.1.1	Validità e prassi	163
11.1.2	Un esempio illustrativo: verifica distribuzione normale di un'istogramma	164
11.2	Test di Student	165
11.2.1	Costruzione del test t o di Student	166
11.2.2	Esempi applicazione test Student	168
11.3	Chiusura	171
11.4	Chiusura	172
12	Il confronto di ipotesi	175
12.1	Rapporto delle verosimiglianze	175
12.2	Errori di tipo I e di tipo II	177
12.2.1	Analogia con il parametro di interesse	178
12.3	Potenza del test	179
12.4	Un esempio	180
12.5	Chiusura	181
III	RACCOLTA ESERCIZI	183
13	Quesiti d'Esame	185
13.1	Prima Parte Programma	185
13.2	Seconda Parte Programma	189
13.3	Soluzioni	190

14 Raccolta di Esercizi	195
14.1 Prima Parte Programma	195
14.2 Soluzioni	197

Introduzione

La Fisica e' un ramo della scienza con due forti identita', quella teorica, che si dedica allo sviluppo di teorie sulla base di osservazioni sperimentali (specialmente le anomalie) o di considerazioni puramente teoriche, e quella sperimentale, che si basa sulla ideazione di misure basate su considerazioni teoriche (per lo più innovative). Attenzione! Quanto scritto sopra è chiaramente estremamente riduttivo! Una teoria non e' necessariamente un complesso di difficili equazioni che definiscono un nuovo comportamento della materia, ma anche più umilmente un complesso di idee sul comportamento di un nuovo componente elettronico, o di un classificatore basato su reti neurali, o della risposta di uno strumento, e così via. Analogamente la misura sperimentale non si applica solo strettamente a letture di uno strumento fisico, ma si può anche pensare come la misura della validità di una affermazione basata su certe considerazioni logiche. Fisica teorica e sperimentale vanno a braccetto su moltissimi settori, ma ovviamente - specialmente in tempi moderni - va a finire che la ricercatrice si specializza in un settore molto preciso che richiede delle competenze altissime che lei ha maturato in anni di esperienza sui libri e articoli di settore, e sul campo. Di conseguenza, spesso la ricercatrice darà per scontati alcuni concetti di base del suo settore, adotterà standard talvolta basati sulla consuetudine, e userà acronimi incomprensibili anche per colleghi non troppo distanti. Tuttavia, ogni volta che questa ricercatrice andrà a fare una misura o svilupperà una teoria, e pubblicherà un suo lavoro, userà sempre le stesse consuetudini basate sul *metodo scientifico*, secondo cui ogni analisi deve essere oggettiva, affidabile, verificabile e condivisibile¹. Questo si traduce in sostanza nel fatto che un *report* scientifico, e quindi la misura stessa, deve far sì che la collega che andrà a valutare la validità del risultato, riesca a formarsi un grado di *confidenza* sulla veridicità della affermazioni, esattamente uguale a quello che si è fatta la sperimentatrice nell'ottenerlo. Cercheremo di capire in che senso un risultato e' *vero* in senso sperimentale. Da ogni lavoro scientifico deve emergere chiaramente quindi un *livello di confidenza* che rappresenta uno dei concetti fondamentali di questo corso.

Francesca e Beatrice Avrete forse notato che per parlare dello sperimentatore ho usato termini femminili. La scelta e' di proposito, ma non e' di mia paternità. Si usa molto soprattutto in testi angloamericani. La ragione e' che la carriera di scienziate STEM (acronimo di *science, technology, engineering and mathematics*) e' ancora troppo spesso svantaggiata rispetto alla carriera dei colleghi scienziati. Le cause sono molte, e talvolta non palesi neanche a chi le pratica². Credo che questa pratica possa mantenere alta l'attenzione sul problema. Durante il corso, introdurremo due scienziate: Francesca, studentessa del primo anno della Laurea Triennale in Fisica, e Beatrice, studentessa magistrale di Fisica. Ci aiuteranno talvolta a capire alcuni approcci diversi alla statistica delle misure.

Incerteza e Livello di confidenza Chiediamo subito a Francesca se, date due misure di accelerazione di gravità, ad esempio $g_1 = 9.79 \text{ m s}^{-2}$ e $g_2 = 9.81 \text{ m s}^{-2}$, si possa dire che $g_1 < g_2$? Francesca risponderà forse che le sembra sia vero... Ma e' così? Cosa c'e' di sbagliato? La ragione e' che non esiste (praticamente mai, come vedremo) una misura fisica che non sia misurata con una certa *incerteza*. Questo concetto e' complesso e sara' l'argomento di tutta la prima parte del corso. Ogni misura ha una incerteza e questa e' di solito riassunta riportando un risultato con un \pm e poi la stima

¹Wiki: https://it.wikipedia.org/wiki/Metodo_scientifico

²Si dice *implicit bias* per descrivere quando abbiamo atteggiamenti nei confronti delle persone o associamo loro stereotipi a nostra insaputa.

della incertezza. Supponiamo allora che le due misure precedenti fossero, assieme alle loro incertezze $g_1 = (9.79 \pm 0.05) \text{ m s}^{-2}$ e $g_2 = (9.81 \pm 0.07) \text{ m s}^{-2}$. Cosa si vede? Considerando le incertezze, i due valori sono molto vicini. In termini che vedremo, esse sono ottimamente compatibili. Ma allora $g_1 < g_2$? La risposta corretta non e' nemmeno no. Forse anche Beatrice si sarebbe sbagliata qua se un po' distratta. Come mai? Supponiamo che le due misure siano ora ripetute con due strumenti a maggiore precisione, dando $g_1 = (9.791 \pm 0.001) \text{ m s}^{-2}$ e $g_2 = (9.812 \pm 0.002) \text{ m s}^{-2}$. Allora in questo caso, il ragionamento di prima non funziona piu': le misure sono distanti osservando le loro incertezze. Allora mentre nel primo caso non potevo, a meno di non assumere a torto nulla le incertezze, fare una affermazione scientifica. Nel secondo caso (compatibilita') Beatrice poteva dire :

Le misure sono compatibili entro le incertezze sperimentali

Fate attenzione, questo implica che potrebbero anche non esserlo in realta'! Infine, nel terzo caso, Beatrice dovrebbe dire:

Le misure sono incompatibili con un livello di confidenza del 95%

Cosa si intende per livello (o margine) di confidenza lo vedremo durante il corso, ma per ora anticipiamo alcuni concetti utili. Quanto riportiamo $g_1 = (9.791 \pm 0.001) \text{ m s}^{-2}$ in realta' diamo per implicito che assumiamo che la natura delle incertezze (la cui stima e' 0.001 m s^{-2} e' di tipo Gaussiano (vedremo cosa significa), che e' il caso piu' frequente. Allo stesso modo questa scrittura ci permette di quantificare e comunicare alla collega il *livello di confidenza*. Vedremo infatti che, se non esplicitato diversamente, il livello di confidenza associato a quella misura e' del 68%³. Beatrice sa invece che

Se $g_1 = (9.791 \pm 0.001) \text{ m s}^{-2}$ e le incertezze sono Gaussiane, allora c'e' un livello di confidenza del 68%, ovvero una probabilita' del 68% che il valore vero della misura sia tra 9.790 e 9.792) m s^{-2}

Cosa si intende esattamente per valore vero lo vedremo durante il corso. Attenzione che il concetto e' sfuggente. La frase di Beatrice sopra implica anche che la probabilita' che il valore vero sia proprio 9.7921 m s^{-2} non e' cosi' alta, visto che la probabilita' che stia in 9.790 e $9.792) \text{ m s}^{-2}$ e' del 68%. E infine, nemmeno questo intervallo da poi cosi' garanzia. Se la misura e' Gaussiana allora vedremo che per aumentare questa probabilita', in maniera forse controintuitiva, devo allargare l'intervallo: al 95% la misura si trova tra 9.789 e 9.793 m s^{-2} e al 99% tra 9.788 e 9.793 m s^{-2} . 69,95,99% sono tre diversi livelli di confidenza.

Questo senso concetto di confidenza, vedremo, non si applica solo alla misure, ma anche alle ipotesi. Ad esempio, se prendessimo misure di spazio e tempo osservando un carico che scende lungo un piano inclinato, ci potremo chiedere quale sia la probabilita' che esse seguano un moto uniformemente accelerato. Questa domanda ci sembra piu' che lecita, ma vedremo che la risposta a questa domanda sara' data secondo due linee di pensiero diverse: una detta *frequentista* secondo cui non si puo' rispondere a quella domanda, ma ad una analoga ma profondamente diversa ovvero

quale e' la probabilita' che, assumendo per vera l'ipotesi del moto uniformemente accelerato, le misure supportino questa ipotesi?

Francesca imparera' a rispondere a questa domanda in maniera quantitativa e dettagliata durante questo corso. Secondo un procedimento diverso invece, si potra' in effetti valutare la probabilita' di una ipotesi, secondo la via *bayesiana*, ma attraverso una domanda specifica:

quale e' la probabilita' che l'ipotesi di moto accelerato sia vera alla luce delle misure in questione?

Beatrice ha imparato a rispondere e a capire che questa domanda e' diversa da *quale e' la probabilita' che l'ipotesi di moto accelerato in quanto prevede l'uso delle misure in questione*.

Ma come quantifichiamo queste incertezze sperimentali e questo margine di confidenza? In questo ci aiuta la stastica. Come dice T. Loredo⁴:

³Ora Francesca si chiedera 68% di cosa? E farebbe bene

⁴T. Loredo homepage <http://hosting.astro.cornell.edu/~loredo/>

Il piu' fondamentale principio del paradigma statistico, il suo punto di partenza, e' che l'incertezza puo' essere descritta dalla probabilita'. [T. Lored]

Ovvero, e' possibile descrive in termini probabilistici l'incertezza di una misura. Una volta che poi la eseguiamo quindi, e otteniamo un risultato, abbiamo quindi una guida per confrontare quello che abbiamo rilevato con quello che ci aspettiamo dalla teoria statistica della probabilita'.

Struttura del testo **Prima parte.** Per poter arrivare a formulare tali giudizi, è il caso di cominciare da alcune nozioni di base. Cominceremo con la teoria degli errori sperimentali e nozioni di metrologia nel [Capitolo 1](#). Qui impareremo a usare il vocabolario giusto per definire caratteristiche di uno strumento e di una misura, e delle incertezze legate alla misura. Procederemo nel [Capitolo 2](#) a mostrare quale sia la forma corretta di presentare un risultato numerico, e come graficare le misure. Per introdurre la relazione tra statistica astratta e distribuzioni di dati sperimentali anticiperemo alcuni concetti studiando il comportamento di misure ripetute in condizioni di ripetibilità, osservando come esse si distribuiscano attorno ad un valore centrale in maniera decisamente regolare, aggiustandosi sotto una curva detta Gaussiana. Faremo questo nel [Capitolo 3](#). Passeremo quindi a discutere la Teoria della probabilità nel [Capitolo 4](#), ovvero la teoria matematica sulla quale si basa la statistica, facendo un breve excursus nel calcolo combinatorio, che serve per calcolare facilmente il numero di istanze con determinate qualità rispetto ad un numero di istanze globale. Continueremo con la discussione astratta delle distribuzioni di probabilità nel [Capitolo 5](#), e loro caratteristiche in termini di momenti algebrici e centrali. Continueremo descrivendo la funzioni di questa variabili nel [Capitolo 6](#). Infine, applicheremo tutto quanto visto al caso reale, quello dei campioni di misura nel [Capitolo 7](#). Chiuderemo con esercizi riassuntivi per la prima parte del programma in ?? **Seconda parte.** Cominceremo descrivendo una importante funzioni dei parametri liberi di un modello detta funzione di verosimiglianza. Con questa procederemo alla stima dei parametri che descrivano distribuzione di misure attraverso il metodo della massima verosimiglianza [Capitolo 9](#). Nel [Capitolo 10](#) descriveremo il test del χ^2 per la stima di bonta' di un fit, e nel successivo [Capitolo 11](#) il test di Student per le stime di compatibilità. Infine, parleremo di ipotesi alternative e confronto di ipotesi in [Capitolo 12](#). Chiuderemo con esercizi riassuntivi per la seconda parte del programma in ?? **Laboratorio di Misura** Cominceremo qui tutta la parte relativa alle esperienza di laboratorio. Dopo una introduzione in ??, procederemo nel ?? a parlare della prima esperienza quello sul pendolo semplice e in ?? della guidovia ad aria compressa.

Lecture di riferimento Ci sono innumerevoli testi di riferimento per gli argomenti di questo corso, ed è difficile raccomandarne uno o un altro in maniera da coprire gli argomenti del corso nella maniera che viene qui trattata. In particolare, ce ne sono pochi in italiano al di la di dispense del corso - come questa. Il testo di riferimento per questo corso e' 'Teoria degli Errori e Fondamenti di Statistica. Introduzione alla Fisica Sperimentale' [[Loreti, 2008](#)] scaricabile gratuitamente <https://wwwcdf.pd.infn.it/labo/INDEX.html>.

In aggiunta si consiglia di scaricare due manuali internazionali di riferimento per questa disciplina. Questi sono redatti dal Joint Committee for Guides in Metrology (JCGM) per conto di ISO e sono: 1. Guide to the expression of Uncertainty in Measurement (GUM) e 2. International Vocabulary of Metrology (VIM). I test si possono scaricare gratuitamente in <https://www.bipm.org/en/publications/guides/>.

Infine, si possono citare una larga serie di libri di consultazioni, con il rischio di creare piu' confusione piuttosto che aiutare la studentessa. Rimane comunque la voglia di citare fonti consultate che hanno permesso di dare forma a queste lezioni. Riporto prima quelli in italiano. Per statistica descrittiva e inferenziale: 'Probabilita' in fisica' [[Boffetta and Volpiani, 2012](#)]. Per laboratorio di misura ho trovato molto attuale e completo 'Il Laboratorio di Fisica I' [[Canale and Pietra, 2020](#)], che presenta molti approfondimenti. Ho consultato anche 'Introduzione al Laboratorio di Fisica' [[Ciullo, 2014](#)]. La letteratura in inglese e molto vasta. Riporto qui solo alcuni testi che approfondiscono in maniera diversa i vari argomenti trattati applicando a casistiche diverse : 'Statistical Methods for Data Analysis in Particle Physics' [[Lista, 2016](#)], 'Data reduction and Error Analysis' [[Bevington and](#)

[Robinson, 2003](#)], 'Introduction to Statistics and Data Analysis for Physicists' [[Bohm and Zech, 2010](#)]. Per curiosità date una occhiata ai testi per vedere la vastità del materiale che potenzialmente potrebbe coprire un corso di base di statistica.

Come seguire questo corso Questo corso presenta una difficoltà intrinseca difficilmente risolvibile: cercheremo di descrivere statisticamente campioni finiti di misure con funzione di probabilità, e riferimenti teorici. Come sempre accade in queste situazioni, si incontreranno casi in cui 'non si sa cosa fare', ovvero se applicare un metodo oppure un altro, e alla fine si farà fatica a dare quel giudizio di confidenza tanto discusso. Questo può portare a frustrazione, ma si tratta di una esperienza istruttiva. La ricerca è sempre frutto di frustrazioni e successi. I successi (rari) ripagheranno di anche anni di frustrazione a cercare di comprendere i fenomeni. Se vi siete iscritti a Fisica dovrete essere pronti con questo spirito!

Parte I

STATISTICA DESCRITTIVA

Capitolo 1

Metrologia e Misura

Quest'opera è distribuita con licenza [Creative Commons](#) "Attribuzione – Non commerciale – Condividi allo stesso modo 3.0 Italia".



Abbiamo visto nell'introduzione come una misura fisica sia sempre soggetta a incertezze, che queste vadano sempre riportate assieme alla misura, e che attraverso di queste si costruisce il livello di confidenza sulla misura stessa. In questa sezione entriamo piu' in dettaglio su cosa sono le misure fisiche, su cosa sono le sorgenti di incertezza e come classificarle. Per fare questo, entreremo nel dettaglio del vocabolario della teoria della misura, e del vocabolario della metrologica degli strumenti con cui appunto si fanno le misure.

Vedremo fin da subito che e' tutt'altro che semplice definire chiaramente queste quantità, ma lavoreremo per avere un vocabolario di concetti comune.

1.1 Grandezze e Misure in fisica

Quantità, Misurando e Valore vero di una quantità

Siccome ci torna utile, partiamo subito con un esempio. Pensiamo alla stima dell'accelerazione di gravità attraverso una delle misure piu' semplici da effettuare (anche in casa, fatelo subito! o aspettate la prima esperienza di laboratorio descritta in ??), ovvero attraverso la misura del periodo di oscillazione di un pendolo semplice. Un pendolo semplice e' un oggetto costituito da un punto fisso (ad esempio un chiodo sul muro), un filo sottile che collega il perno ad una massa (ad esempio un bullone) fissata all'altra estremità. E' troppo presto ora per discutere la teoria che porta al calcolo che segue, ma con buona approssimazione se si mette in oscillazione dolcemente il pendolo, e' valida la seguente relazione funzionale tra il periodo di oscillazione singolo del pendolo T , la lunghezza l del filo tra il perno e il bullone e l'accelerazione di gravità g :

$$T \simeq 2\pi \sqrt{\frac{l}{g}} \quad (1.1)$$

Supponiamo di chiedere a Francesca di fare la misura e poi confrontare con il valore *noto* dell'accelerazione di gravità? Francesca si mette a fare la misura, ma mano a mano che ci pensa verranno fuori molti dubbi. Per fortuna Beatrice sa gia' la risposte. Ecco un dialogo possibile:

FRANCESCA: Quale e' il valore di riferimento di g ?

BEATRICE: Hai guardato su Wikipedia?

F: Si, ma non so se va bene?

B: Che precisione di serve raggiungere?

F: Boh, non saprei

- B: Ecco, alla fine devi capire quale potrebbe essere la precisione raggiunta dalla misura. Tutti sanno che $g = 9.81 \text{ m s}^{-1}$ ma se il tuo strumento misurato, diciamo con una precisione di 1/1000, allora devi cercare un valore di riferimento una cifra significativa in piu. In quel caso, devi stare attenta, perche' g puo' variare di qualche permille a seconda del luogo in cui la misuri
- F: Ah, quindi devo cercare g a Padova?
- B: Si, credo sia $g = 9.806 \text{ m s}^{-1}$. Conta che se volessi spingerti oltre con la precisione, ad esempio per una parte su un milione, dovresti considerare un sacco di fattori in piu': il movimento della terra attorno al sole, la presenza di oggetti massivi attorno alla misura, etc.
- F: Ultima domanda: ma a che precisione devo arrivare?
- B: Beh, dipende! Nel caso del laboratorio di misura stai solo imparando, quindi diciamo che devi arrivare alla *massima* precisione che ti puo' dare lo strumento, pero' nulla ti vieta di pensare a possibili soluzioni per aumentare la precisione, ma attenta alla accuratezza della misura anche!

Le domande di Francesca sono molto ragionevoli, e nascondono una difficoltà che sembra quasi un non senso: non e' facile definire cosa si vuol misurare. Vogliamo misurare g ? Si ma a Padova? a Padova il 6 maggio di ogni anno? a Padova sopra la Specola oppure sotto? Ecco, la misura di una grandezza parte dalla definizione di Misurando e di Valore Vero della Quantita' che si vuole misurare. Leggiamo prima le definizioni che da il VIM di quantita, misurando e valore vero:

Quantità (Quantity) VIM 1.1 Proprietà di un fenomeno, corpo o sostanza la cui grandezza si puo' esprimere attraverso un numero ed una unita' di misura.

Misurando (Measurand) VIM 2.3 Quantità particolare soggetta a misurazione

Valore Vero di una grandezza (True Quantity Value) VIM 2.11 Valore quantitativo coerente con la definizione di una determinata quantità.

Vedete allora come il valore vero e' il valore che si otterrebbe sia con una *misura perfetta* che con una *definizione* della grandezza da misurare 'perfetta'. In pratica, i valori veri sono per natura indeterminati. Il misurando differisce dal valore vero in quanto e' non piu' teoricamente, ma praticamente quello che ci prefiggiamo di misurare. Nel caso del misurando quindi vanno date le informazioni utili alla misura, ad esempio, accelerazione locale di gravità a Padova sopra la Specola.

La precisione gioca un ruolo importante. Se facessimo l'esperimento del pendolo semplice a casa, non avrebbe senso dire che il misurando e' 'accelerazione locale di gravità a Padova il 6 maggio 2022', in quanto non saremo in grado di apprezzare variazioni milionesime della quantità. Con il pendolo semplice misuro 'accelerazione locale di gravità'. Beatrice ha anche parlato di accuratezza della misura, mettendo in allarme Francesca. Torneremo su questo tra poco.

Dovrebbe quindi gia' risultare chiaro che e' di fondamentale importanza pensare bene alla procedura di misura che si vuole effettuare per avere un buon risultato . Il VIM raccomanda l'uso di due concetti diversi per descrivere la misurazione:

Metodo di misurazione (Measurement Method) VIM 2.5 Sequenza logica delle operazioni, descritta genericamente, utilizzata nell'esecuzione delle misurazioni

Procedura di Misurazione (Measurement Procedure), VIM 2.6 Serie di operazioni, descritte in modo specifico, utilizzate nell'esecuzione di misurazioni particolari secondo un determinato metodo

Ad esempio, nel caso della stima di g riportata sopra, Francesca avrebbe potuto dire in termini generici:

Il metodo di misura di g si basa sulla misura del periodo di oscillazione di un pendolo semplice

La procedura della misura invece va nel dettaglio ed elenca i passi in maniera che il lettore possa in linea di principio riprodurre l'esperimento. Ad esempio:

'La procedura della misura prevede di mettere in oscillazione ad un angolo di 10 gradi il pendolo senza più toccarlo, misurando attraverso un cronometro digitale con risoluzione di 0.1 s i primi 30 periodi di oscillazione, osservando di volta in volta il punto in cui il pendolo si fermava e inverte la direzione'

Al di là del vocabolario, tenere traccia della procedura di misura è non solo importante, ma tendenzialmente complesso, soprattutto nel caso si facciano diversi test e prove, o le misure siano molte. Di solito gli sperimentatori usano un 'logbook' ovvero un quaderno (reale o digitale) dove annotare tutti i passaggi svolti da usare poi per la stesura del report¹. Torneremo su questo punto più avanti.

Attenzione infatti ad una cosa: non è garantito che il risultato della misura sia proprio il misurando che vi eravate prefissati! Se avete sbagliato qualcosa senza accorgervene, avete misurato un altro misurando!

1.2 Incertezza ed errore nella misura

Quando si riporta una misura di una grandezza fisica si parla sia di incertezza sia di errore. Nel linguaggio parlato incertezza è un concetto più vago mentre la parola errore è più forte e sembra sottintendere di aver sbagliato qualcosa. In realtà, anche se in qualche caso si può effettivamente trattare di un errore, i due concetti si riferiscono entrambi ad una differenza o distanza rispetto ad un certo valore di riferimento, ma in maniera diversa, e per errore non si intende il concetto di 'sbaglio'. È importante capire la differenza tra i due concetti, tuttavia sottolineiamo come poi nel linguaggio parlato tra scienziati e talvolta scritto si usino i concetti di incertezza ed errore spesso erroneamente in maniera intercambiabile. Ecco le definizioni del VIM:

Errore di misurazione (*Measurement Error*) VIM 2.16 Risultato di una misurazione meno il valore di riferimento

Incertezza di misurazione (*Measurement uncertainties*) VIM 2.26 Parametro, associato al risultato di una misurazione, che caratterizza la dispersione dei valori che potrebbero ragionevolmente essere attribuiti al misurando

Cominciamo dal concetto di incertezza. La definizione ci dice che in generale, data una misura (vedremo, fatta con sufficiente sensibilità), non otterremo sempre lo stesso valore, ma valori diversi *dispersi attorno ad un valore di riferimento*. Ad esempio, se misuriamo il periodo di oscillazione del pendolo semplice fatto in casa, se misuriamo con un cronometro con sufficiente sensibilità, ad esempio con una risoluzione di 1 centesimo di secondo, non troveremo lo stesso valore ma valori diversi, tutti (come vedremo) distribuiti attorno ad un valore centrale, ad esempio

1.91, 1.97, 1.94, 1.97, 1.89... [s]

Ma cosa è che fa 'oscillare' questi valori? Pensateci prima di procedere.

Esercizio:

Cosa fa oscillare i valori misurati?

¹Andata a cercare in internet le immagini dei quaderni di appunti di Galileo Galilei, vedrete come non sono molto diversi dai quaderni di appunti moderni di uno sperimentatore!

Prima di passare allo studio delle incertezze, consideriamo anche che molto raramente si fanno esclusivamente misure dirette di certe quantità, ma per la maggior parte di volte il misurando è ottenuto in maniera indiretta da diverse quantità, a partire da un **modello matematico** o una **relazione funzionale**.

Esempio:

Ad esempio, una differenza potenziale / viene applicata ai capi di una resistenza R . La potenza misurata (il misurando) P è descritto dal seguente modello matematico:

$$P = P(V, R_0, \alpha, T) = \frac{V^2}{R} = \frac{V^2}{R_0(1 + \alpha(T - T_0))}$$

In questo caso, V, R sono misure di input, T è un fattore influenzate, α è un parametro (esterno o misurato) e P è il misurando di output.

Le quantità di input X da cui dipende la quantità di output Y possono esse stesse viste come misure e possono esse stesse dipendere da altre quantità, inclusi correzioni e fattori di correzione per effetti sistematici, portando così a una possibile complicata relazione funzionale f che può non essere mai scrivibile esplicitamente. Inoltre, f può essere determinato sperimentalmente o esiste solo come algoritmo che deve essere valutato numericamente.

Sorgenti di incertezza della misura Ci sono diversi sorgenti di incertezza della misura che il GUM (3.3.2) elenca così:

1. **Variazioni in osservazioni ripetute sotto le stesse condizioni.** Questi, di cui parleremo in dettaglio, sono le proprie incertezze casuali.
2. Definizione incompleta del misurando. Per esempio, la definizione di accelerazione locale di gravità come visto sopra senza ulteriori specifiche.
3. Imperfetta realizzazione della definizione del misurando. Per esempio, la misura di g con un pendolo semplice ideale trascura il fatto che il pendolo reale che costruiamo non sarà mai perfettamente ideale.
4. Campione non rappresentativo. Qui l'esempio del pendolo vale meno ma il concetto è facile. Se si vuole sapere se la popolazione italiana ama i Måneskin non si può fare un sondaggio solo su un gruppo di persone tra i 10 e i 20 anni e poi estrapolarlo alla popolazione.
5. Conoscenza inadeguata dell'effetto dei fenomeni ambientali sul misurando. Ad esempio, se il pendolo fosse sostenuto da un filo metallico, questo potrebbe allungarsi o accorciarsi con la temperatura.
6. Valori inesatti dei campioni e dei materiali di riferimento.
7. Valori inesatti di costanti e parametri che intervengono nell'analisi dati.
8. Approssimazioni e assunzioni nel metodo e nella procedura. Ad esempio, non considerando lo smorzamento dovuto all'attrito sul perno.
9. Errori umani nel leggere strumenti analogici. Ad esempio, se leggete il pendolo con un orologio a muro molto di sbieco, potreste sovrastimare o sottostimare la misura facilmente.
10. Risoluzione finita dello strumento o soglia di discriminazione. Supponete che il periodo del pendolo sia 2.05 s esatti e il vostro cronometro abbia solo 0.1 s come risoluzione, allora qualche volta il cronometro segnerà 2.1 s e qualche volta 2.0 s.

Inceteeze Casuali. Le inceteeze casuali sulla misurazione sono quelle del punto (1) della lista sopra. Sono legate alla presenza di diversi fattori. Nella maggior parte dei casi, avvengono a causa di fluttuazioni o nel misurando stesso oppure nello strumento di misura o nella procedura di misurazione. Prendiamo la stima di g attraverso il pendolo semplice. L'inceteeza in questo caso è difficilmente imputabile a variazioni apprezzabili intrinseche del misurando, tuttavia, è ben possibile che piccole correnti d'aria, microscopiche oscillazione longitudinali del filo, se osservate con uno strumento con sufficiente precisione, possano essere apprezzate. Esse faranno 'oscillare' la misura attorno ad un valore

centrale. Queste viste sopra sono incertezze di tipo casuale che fanno oscillare la misura attorno un valore centrale talvolta in eccesso talvolta in difetto. Vedremo che nella maggior parte dei casi reali questo tipo di fluttuazioni segue una legge universale la cui forma funzionale e' una curva Gaussiana. Le incertezze casuali hanno la caratteristica che **non sono eliminabili**. Esse saranno sempre presenti nei dati, possono solo essere ridotte aumentando la precisione della misura. Infine, esse influiscono sul risultato **in maniera simmetrica** ovvero in maniera che la probabilità di fluttuazione positiva o negativa sia uguale a parità di entità.

Incetnze Sistematiche Ad esempio spesso le incetnze (2-8) della lista sopra sono di tipo sistematico, ovvero generano errori di tipo sistematico. Cosa succede se ad esempio la lunghezza del filo e' stata misurata in maniera sbagliata, ad esempio non partendo dallo zero? Tutte le misure saranno sistematicamente in eccesso o in difetto (della stessa quantità). Cosa succede se il filo si allunga durante le misure? Ci sara' un effetto sistematico di deriva delle misure. In linea generale, se le incetnze hanno una maggiore probabilità di aumentare (o diminuire) il valore misurato, allora sono sistematiche. Talvolta la classificazione non e' semplice. Le incetnze sistematiche hanno la caratteristica che **se misurate, sono eliminabili**: nel momento in cui si conosce per via teorica, o sperimentale, la loro entità, allora e' possibile correggere il risultato per la discrepanza trovata. Inoltre, anche se non necessariamente, la maggior parte delle incetnze sistematiche influiscono sul risultato **in maniera asimmetrica**, o alterando il valore in maniera sistematica positivamente o negativamente (del tipo $x = x' + a$, dove x e' il valore misurato, a l'effetto sistematico e x' il valore accurato) oppure anche in maniera progressiva (del tipo $x = ax'$).

Incetnze di Lettura Sono quelle legate solo alla lettura dello strumento, tipo le (9-10) della lista sopra. Cosa succede se il valore e' 9.805 e lo misurate con una precisione di due cifre dopo la virgola? Trovate o 9.80 o 9.81. Oppure se leggete male uno strumento analogico? Oppure se trascurate la risoluzione finita di lettura di uno strumento digitale? Le incetnze di lettura possono portare a loro volta a errori sistematici o fluttuazioni casuali a seconda del loro tipo.

Se per fluttuazione si intende questa variabilità attorno ad un valore di riferimento, la distanza rispetto al valore di riferimento si dice *errore di misura*. Se ad esempio durante un esperimento misurate $g = 9.707 \pm 0.001 \text{ ms}^{-2}$ potreste riportare che 0.001 ms^{-2} e' una stima della incetnze di misurazione mentre se considerate 9.807 ms^{-2} come valore vero di riferimento allora $9.807 - 9.707 = 0.1 \text{ ms}^{-2}$ e' una stima dell'errore compiuto nella misura. In particolare, vedremo come spesso, l'incetnze attorno al valore misurato sia distribuita secondo una curva detta 'gaussiana', centrata attorno ad un valore detto media aritmetica. Vedremo come invece la forma della distribuzione di misure (o della distribuzione di probabilità' delle misure) nel caso di errori sia invece diversa caso per caso.

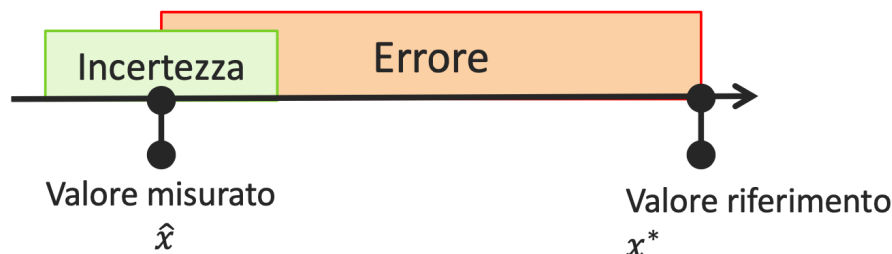


Figura 1.1: Rappresentazione grafica del concetto di incetnze e di errore di misurazione

Indipendentemente dalla loro natura, il BIPM chiarisce che e' superfluo caratterizzarle in una maniera o nell'altra, ma e' preferibile concentrarsi su come si puo' *valutare* la loro entità. Se cio' e' possibile attraverso metodi statistici, ovvero associando alle incetnze una distribuzione statistica di probabilità, quantifica attraverso una grandezza detta *deviazione standard* allora sono di un tipo, altrimenti di un altro. Ecco le definizioni del VIM:

Valutazione della incertezza metodo A VIM 2.28 Incertezza valutabile con metodi statistici (e quantificabile attraverso una deviazione standard).

Valutazione della incertezza metodo B VIM 2.29 Incertezza della misurazione non di tipo A.

Incertezza standard VIM 2.30 Incertezza della misura stimabile attraverso la deviazione standard

Fanno parte della prima categoria tutte quelle fonte di incertezza di cui conosciamo la distribuzione di probabilità, ad esempio, nel caso di misure ripetute in condizioni di ripetibilità dello stesso misurando, ci si aspetta che la distribuzione sia Gaussiana. Nel caso di una misura singola effettuata con uno strumento analogico, ci si aspetta che la distribuzione di probabilità attorno al valore del misurando sia di tipo triangolare e così via. Si tratta di incertezza di tipo A se la valutazione dell'incertezza di una componente è svolta mediante un'analisi statistica dei valori di quantità misurate ottenuti in condizioni di misurazione definite o se il metodo di valutazione dell'incertezza mediante analisi statistica di serie di osservazioni. Nel caso non possiamo stimare questa probabilità, ad esempio, se consideriamo un parametro riportato senza errore da terzi, allora si tratta di un errore di tipo B la cui trattazione è più complessa. Si tratta di incertezza di tipo B se la Valutazione di una componente dell'incertezza di misura è determinata con mezzi diversi da una valutazione di tipo A dell'incertezza di misura².

Esercizio:

Un errore sistematico costituisce sempre una incertezza di tipo B?

1.3 Accuratezza e Precisione

Nel linguaggio parlato i due concetti di precisione e accuratezza sono spesso usati come sinonimi. Nel linguaggio delle misure invece sono concetti diversi ed è opportuno usarli correttamente. Vediamo la definizione che ne dà il GUM.

Accuratezza di misurazione (*Measurement Accuracy* VIM 2.13) Vicinanza dell'accordo tra il risultato di una misurazione e un valore reale del misurando

Precisione della misura (*Measurement Precision* VIM 2.15) Accordo tra misure ripetute dello stesso misurando in condizioni identiche

Anzitutto notiamo che i due concetti sono principalmente qualitativi. In alcuni casi, questi sono caratterizzati numericamente da due valori: l'errore relativo, e la discrepanza o bias. Se una misura è nota con una incertezza $x + \delta_x$ allora per errore relativo si intende il rapporto tra incertezza e valore $\epsilon = \delta_x/x$, espresso numericamente o come percentuale. È un valore spesso usato e di facile comprensione, tuttavia perde senso nel caso in cui il valore misurato sia vicino a zero (in relazione alla sua incertezza). Attenzione che una misura più precisa ↗ ha un valore di errore relativo più basso ↘. Invece è più semplice quantificare la *inaccuratezza* piuttosto che la accuratezza ove possibile attraverso una grandezza chiamata *bias* o *discrepanza*, ovvero la differenza tra il valore misurato e quello di riferimento. Di solito se il valore di riferimento è un valore noto (un valore vero) allora si parla di bias, mentre il concetto di discrepanza è più generico come differenza tra due valori. Nel caso il valore misurato sia superiore al valore aspettato si parla anche di *sovrastima* e nel caso opposto di *sottostima*.

Le misure possono essere precise, imprecise, accurate o inaccurate. Vediamo un esempio grafico nella Fig. 1.2. Beatrice e Francesca sono d'accordo nel dire che è importante che una misura sia accurata e precisa (caso D), ma non sono d'accordo, nel caso grande accuratezza e precisione non

²Si guardi approfondimento in GUM sezione F.2

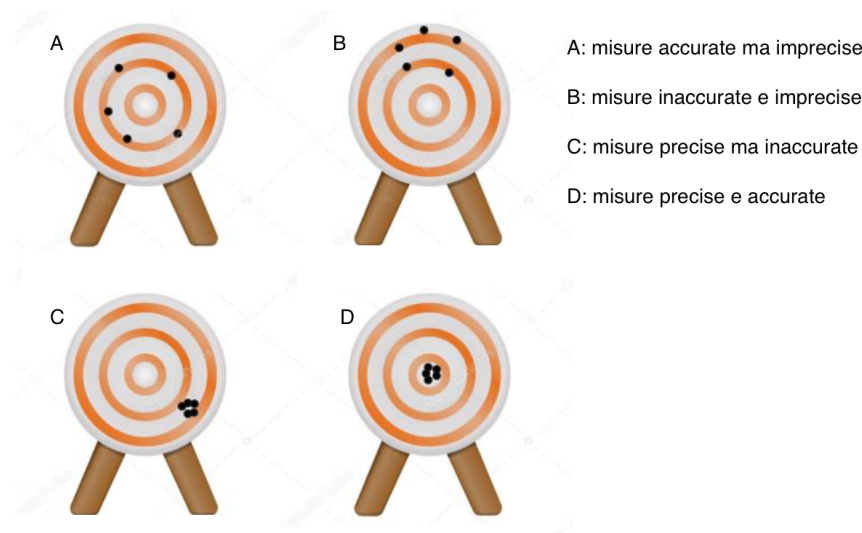


Figura 1.2: I concetti di precisione e accuratezza mostrati graficamente.

possano essere raggiunte entrambe, se sia preferibile avere elevata accuratezza o elevata precisione. Francesca sostiene che sia meglio avere una misura precisa (caso C), perché significa che è stata presa da uno strumento ad alte prestazioni. Beatrice invece la convince che sia meglio, non potendo avere entrambe, ad avere accuratezza della misura. È facile infatti convincersi che la probabilità di avere il valore vero all'interno delle incertezze è maggiore nel caso accurato ma impreciso (A) piuttosto che nel caso preciso ma inaccurato (C). Tuttavia, è importante notare che se nel caso C si riesce a stimare l'inaccuratezza, le misure possono essere corrette per questa inaccuratezza, a questo punto C tornerebbe ad essere sia precisa che accurata. Nel caso B è persino difficile dare una stima di inaccuratezza. La preferenza è quindi: $D \rightarrow A \rightarrow C \rightarrow B$.

In altre parole, la precisione è legata anche alla ripetibilità del risultato della misura di una stessa grandezza mentre l'accuratezza alla capacità di fornire valori corrispondenti a quello realmente posseduto dalla grandezza in esame. Affrontiamo quindi alcuni concetti legati alla ripetizione di una misura. È chiaro che qualsiasi misura sperimentale DEVE poter essere riprodotta per poter essere verificata, migliorata o confutata. A riguardo il VIM:

Ripetibilità della misura (*Measurement Repeatability*) VIM 2.21 Condizione di misurazione che garantisce una serie di condizioni che includono la stessa procedura di misurazione, stesso sistema di misurazione, stesse condizioni operative, stesso luogo, possibilità di ripetere la misura in brevi intervalli di tempo

Riproducibilità della misura (*Measurement Reproducibility*) VIM 2.25 Condizione di misurazione, al di fuori di una serie di condizioni che include posizioni, operatori, sistemi di misurazione diversi e replicare le misurazioni sullo stesso oggetto o su oggetti simili]

Anche in questo caso, a volte le definizioni non sono semplici da applicare al caso reale. In linea di massima, ad ogni modo, il concetto di ripetibilità si riferisce più alla procedura che allo strumento, che appunto non deve cambiare. La riproducibilità si riferisce più alla stabilità del risultato se si cambiano alcuni parametri strumentali. Ad esempio, l'esperimento del pendolo semplice fatto in casa è decisamente riproducibile. Se si usa un filo di canapa o di plastica, sottile o più spesso il risultato cambia poco, e così per la massa appesa. È anche ripetibile in questo caso. Nel caso volessimo invece misurare il limite di rottura di un filo di lana grezza a seguito di una forza applicata, la misura sarebbe poco ripetibile non solo in quanto il filo poi si spezza poi si spezza, ma anche perché non è facile (direi impossibile) trovare un filo di lana grezza esattamente uguale al primo.

Un ulteriore concetto relativo alla misura è quello di **affidabilità**. Per affidabilità si intende la capacità dello strumento di fornire misure uguali della stessa grandezza entro la sua risoluzione, anche in condizioni di lavoro difficili o variabili (vibrazioni, sbalzi di temperatura, ...). In pratica

lo strumento deve risultare ben isolato rispetto agli effetti dell'ambiente esterno, escluso ovviamente l'effetto dovuto alla grandezza in esame. Uno strumento o una misura è affidabile se è stabile e priva di deriva. Per **stabilità** (VIM 4.19) si intende la proprietà di uno strumento di misura, per cui le sue proprietà metrologiche rimangono costanti nel tempo. Ad esempio, un righello è uno strumento affidabile per lungo tempo, potrebbe non esserlo invece una bilancia a molla, in quanto la molla perde le sue caratteristiche nel tempo. Per **deriva strumentale** o *drift* (VIM 4.21) si intende un caso particolare di non stabilità in cui la variazione nella risposta è continua o incrementale nel tempo nell'indicazione, dovuta a cambiamenti nelle proprietà metrologiche di uno strumento di misura. Ad esempio nel caso precedente è probabile che la molla mostri una deriva sperimentale progressiva.

1.4 Risoluzione e Sensibilità

Qui di seguito definiamo ulteriori concetti di metrologia, importanti a costruire un vocabolario pertinente che ci permetta di descrivere accuratamente le misure che ci accingiamo a fare. Ecco alcune

Risoluzione (Resolution VIM 4.14) Il più piccolo cambiamento in una quantità misurata che provoca un cambiamento percepibile nell'indicazione corrispondente

Sensibilità (Sensitivity, VIM 4.12) Quoziente della variazione di un'indicazione di un sistema di misurazione e la corrispondente variazione del valore della quantità misurata

Attenzione che anche in questo caso le definizioni possono dare luogo a diverse interpretazioni. Anzitutto cominciamo a distinguere tra le applicazioni di questi concetti al singolo strumento o poi alla misura nel suo complesso, che può essere costituita da diversi strumenti e diverse fasi. La risoluzione è in effetti univocamente definita solo nel caso di uno strumento digitale, nel quale essa corrisponde al 'digit' meno significativo del lettore digitale che riporta la grandezza. Ad esempio, in un cronometro al millesimo di secondo la risoluzione è $R = 0.001$ s. Nel caso di uno strumento analogico, ad esempio un righello, normalmente si assegna una risoluzione pari alla 'tacca' dello strumento, ad esempio normalmente $R = 1$ mm nei righelli scolastici, ma si può anche considerare il fatto che l'occhio è in grado di stimare facilmente frazioni della tacca, posizioni intermedie, e in questo senso la risoluzione potrebbe essere *stimata* dall'osservatore come $R = 0.5$ mm o $R = 0.25$ mm. Un caso eclatante è quello ad esempio di un orologio con solo la lancetta e senza tacche. Un occhio abilitato è in grado di stabilire l'ora esatta probabilmente con una risoluzione di al più 1-2 minuti. È chiaro che condizione necessaria affinché una misura sia precisa è che gli strumenti usati siano precisi. La condizione non è sufficiente. Ad esempio, se misuriamo il periodo di oscillazione del pendolo semplice casalingo azzerando il cronometro manualmente, anche se questo avesse una precisione del milionesimo di secondo, la nostra misura non potrebbe raggiungere questa precisione. Quale sarebbe quindi la risoluzione della nostra misura? Torneremo più avanti su questo concetto importante mostrando che, almeno nel caso di misure ripetute, è possibile stimare sperimentalmente la risoluzione.

La sensibilità è un concetto che mescola quello di risoluzione a quello della specifica misura che si intende fare. Si può calcolare in due maniere a seconda dei casi, in una leggenda solo allo strumento, e nell'altra anche alla misura. Facciamo un esempio. Se prendiamo una rollina metrica avete risoluzione $R = 0.5$ mm possiamo definire anche la *sensibilità dello strumento* (della rollina metrica) come:

$$S = \frac{\text{tacca}}{\text{aumento}} = \frac{1}{R} = 2 \text{ mm}^{-1} = 2 \times 10^3 \text{ m}^{-1} \quad (1.2)$$

Questa grandezza ha il vantaggio di crescere all'aumentare della performance dello strumento. Se prendiamo in fatti una rollina con $R = 0.05$ mm allora $S = 2 \times 10^4 \text{ m}^{-1}$. Diciamo quindi *lo strumento ha grande sensibilità*.. Possiamo anche definire la sensibilità sulla base della misura oltre che dello strumento. Supponiamo infatti di usare la rollina metrica per valutare l'allungamento di un filo di acciaio a causa della dilatazione termica. È chiaro che *osserviamo* l'allungamento solo se questo comporta una variazione dell'indicazione sulla rollina, ovvero una variazione maggiore della risoluzione della rollina. Supponiamo che nel nostro caso la rollina abbia $R = 0.5$ mm e che osserviamo

un allungamento solo se la temperatura è variata di $\Delta_T = 70$ K. Allora possiamo calcolare la *sensibilità della misura*

$$S = \frac{\Delta K}{R} = \frac{70 \text{ K}}{0.5 \text{ mm}} \sim 140 \text{ K mm}^{-1} = 0.14 \text{ K m}^{-1} \quad (1.3)$$

1.4.1 Altri concetti metrologici

Altri concetti metrologici sono quelli di **intervallo di funzionamento** (*range*) che è dato dal valore massimo (**portata o fondo scala**, *reach*) e il valore minimo (**soglia**, *threshold*) della grandezza in esame che lo strumento è in grado di fornire. La **prontezza** è invece una caratteristica dello strumento legata alla rapidità con cui è lo strumento è in grado di fornire il risultato di una misura cioè al tempo necessario (**tempo caratteristico**) affinché questo risponda ad una variazione della grandezza in esame. Si quantifica meglio attraverso il tempo caratteristico: quanto minore è tempo caratteristico, tanto maggiore è la prontezza. Il **tempo morto** (*dead time*) è invece l'intervallo di tempo in cui uno strumento, a seguito di una misura, torna nella condizione di eseguirne una seconda.

Si parla inoltre di **fattori esterni** o **quantità influenzanti** alla misura. Per fattori esterni si intendono tutte quelle quantità che in qualche maniera intervengono nella misura di una grandezza e che possono portare ad errori di misura. Ad esempio, la temperatura di un micrometro utilizzato per misurare la lunghezza o la luce ambiente in una misura di intensità luminosa. È chiaro che una misura deve essere poco sensibile rispetto ai fattori esterni.

1.5 Commenti Finali

Ricapitolando

In sommario, abbiamo visto che:

- Una misura deve riportare informazioni quantitative per stabilire il grado di affidabilità
- I fattori che influenzano il risultato di una misura sono di tipo casuale o sistematico, influiscono incertezze ed errori del risultato, e in sostanza la precisione e l'accuratezza.
- Per definire uno strumento e una misura si usano diversi concetti di vocabolario metrologico, che talvolta hanno significato diverso rispetto al linguaggio parlato

Concetti chiave

- Con una misura ci prefiggiamo di quantificare un misurando. Questo potrebbe essere diverso dal valore vero se a) non conosciamo il valore vero (scoperta), b) se facciamo un errore nella definizione del misurando o della procedura di misura
- L'incertezza su una misura si può aumentare aumentando la precisione, l'inaccuratezza è più difficile da scovare ma si può correggere
- I parametri metrologici come risoluzione, sensibilità non si riferiscono allo strumento ma a un mix tra strumento metodo e procedura di misura
- Ogni strumento/misura va opportunamente 'settato'

Abbiamo imparato diversi termini chiave come • Valore vero di una quantità • Ripetibilità e riproducibilità di una misura • Incertezza ed errore • Incertezza tipo A e B • Risoluzione (strumento, misura) • Errore relativo • Modello matematico, grandezze di input, output, parametri • Quantità influenzanti • Misurando • Metodo di misurazione, procedura di misurazione • Errore casuale o sistematico • Deviazione e discrepanza • Precisione e accuratezza • Bias • Affidabilità • Stabilità e deriva strumentale • Sensibilità • Intervallo di funzionamento di misura • Prontezza, tempo caratteristico

Quesiti di autovalutazione

□ **Es. 1.1** Come si può diminuire l'incertezza casuale di una misura? Come si può diminuire l'incertezza sistematica di una misura?

- **Es. 1.2** Una misura con uno strumento con ottima risoluzione sarà precisa o accurata?
- **Es. 1.3** Si pensi ad un esempio di una misura precisa ma inaccurata. Si pensi ad un esempio di una misura imprecisa ma accurata
- **Es. 1.4** Quale è la differenza tra errore casuale e sistematico? Essi influenzano precisione o accuratezza di una misura?
- **Es. 1.5** Ogni misura deve avere un ‘goal’ di precisione e di accuratezza. In che senso, perché?
- **Es. 1.6** Perché il valore vero può non essere noto o essere diverso dal misurando?
- **Es. 1.7** Perché è preferibile una misura accurata ad una precisa in linea generale?
- **Es. 1.8** Stimiamo per un amperometro una risoluzione pari alla tacca di 1 mA. Durante una misura singola di corrente misuro 7.5 mA, quale è la precisione della misura? Posso stimare l’accuratezza? Se stimassi la mezza tacca come risoluzione, come cambierebbe la precisione.
- **Es. 1.9** (Preso da Ciullo) In un multimetro al variare del fondo scala varia la risoluzione dello strumento. Nella tabella che segue, nella seconda colonna viene riportata la lettura di un visualizzatore digitale. Completare la tabella fornendo la deviazione assoluta sulla misura, la deviazione standard con ipotesi uniforme misura, e l’incertezza relativa nei due casi.

Fondo Scala	Misura	δ_V	δ_V/V	σ_V	σ_V/V
V	V	V	%	V	%
1000	2				
100	2.1				
10	2.12				

- **Es. 1.10** (Preso da Ciullo) Misurate una distanza con una rollina metrica con risoluzione 1 cm ed ottenete la misura di 750 cm. Misurate il lato di una calcolatrice pari a 70 mm con una riga avente risoluzione 1 mm. Riportate l’incertezza relativa delle due misure. Quale è la più precisa?

Approfondimenti

Appr. 1.1 Grandezze Fisiche, sistemi di misura, unità di misura.

Il seguente testo è riprodotto da M. Loreti

Le grandezze fisiche si sogliono dividere in *fondamentali* e *derivate*. Con il primo di questi nomi si indicavano, originariamente, quelle grandezze misurate con strumenti e metodi sperimentali che richiedessero un confronto diretto con un campione, scelto arbitrariamente come unità di misura; mentre le seconde venivano generalmente determinate in modo indiretto, ovverosia (come appena detto) attraverso misure dirette di altre grandezze ad esse legate da relazioni algebriche: che permettevano non solo di calcolarne i valori, ma ne fissavano nel contempo anche le unità di misura.

In questo modo si sono definiti vari sistemi di misura coerenti, come il Sistema Internazionale (**SI**) attualmente in uso: esso assume come grandezze fondamentali lunghezza, massa, tempo, intensità di corrente elettrica, temperatura, intensità luminosa e quantità di materia; con le rispettive unità metro, chilogrammo, secondo, Ampère, grado Kelvin, candela e mole. Le unità per la misura delle altre grandezze sono poi univocamente determinate dalle relazioni algebriche che le legano a quelle fondamentali.

Se ciascuna unità fondamentale viene ridotta di un certo fattore, il valore della grandezza espresso nelle nuove unità dovrà essere moltiplicato per un prodotto di potenze dei medesimi fattori. Così, per restare nell’ambito della meccanica, se riduciamo l’unità di lunghezza di un fattore L , l’unità di massa di un fattore M e quella di tempo di un fattore T , ed il valore di una grandezza fisica ne risultasse in conseguenza moltiplicato per

$$L^\lambda M^\mu T^\tau \dots,$$

si dirà che la grandezza in questione ha le *dimensioni* di una lunghezza elevata alla potenza λ per una massa elevata alla potenza μ per un tempo elevato alla potenza τ .

Se si alterano ad esempio l'unità di lunghezza moltiplicandola per un fattore $1/L = 1000$ (chilometro), quella di tempo moltiplicandola per un fattore $1/T = 3600$ (ora) e quella di massa moltiplicandola per un fattore $1/M = 1000$ (tonnellata), il valore di qualunque velocità nella nuova unità (chilometro all'ora) risulterà alterato rispetto al precedente di un fattore

$$L^1 M^0 T^{-1} = LT^{-1}$$

e si dice pertanto che le dimensioni fisiche di una velocità sono quelle di una lunghezza divisa per un tempo.

Come altro esempio si consideri l'energia cinetica di un corpo, definita come il lavoro compiuto dalla forza che si deve applicare per arrestarlo; e che è pari numericamente alla metà del prodotto della massa per il quadrato della velocità del corpo stesso:

$$K = \frac{1}{2} mv^2.$$

Essa è pertanto una grandezza derivata, la cui unità di misura nel Sistema Internazionale è l'energia cinetica di un corpo avente massa di 2 kg ed in moto traslatorio con velocità di 1 m/s (unità detta *joule*). Passando al nuovo sistema di unità prima definito (assai inconsueto per un'energia), il valore di K risulta moltiplicato per il fattore $M^1 L^2 T^{-2}$; si dice dunque che un'energia ha le dimensioni di una massa, moltiplicata per il quadrato di una lunghezza e divisa per il quadrato di un tempo.

Queste proprietà di trasformazione sono legate alla cosiddetta *analisi dimensionale* ed alla *similitudine meccanica*, argomenti che esulano da questo corso. Basti qui osservare che il numero di unità indipendenti non coincide necessariamente con quello delle grandezze assunte come "fondamentali"; così l'angolo piano e l'angolo solido sono entrambi privi di dimensioni in termini di grandezze fisiche fondamentali, e come tali dovrebbero avere come unità di misura derivata ($1m/1m$ e rispettivamente $1m^2/1m^2$) lo stesso "numero puro" 1, mentre esistono per essi due diverse unità: il radiante e lo steradiano, quasi essi avessero dimensioni proprie e distinte.

Né vi è alcunché di necessario nella scelta delle grandezze fondamentali quale si è venuta configurando storicamente nel Sistema Internazionale, potendosi definire un sistema coerente anche con l'assegnazione di valori convenzionali alle costanti universali delle leggi fisiche (come proposto agli inizi del secolo da Max Planck): così un sistema di unità "naturali" si potrebbe fondare, in linea di principio, ponendo uguali ad 1 la velocità della luce nel vuoto, il quanto d'azione (o costante di Planck), la costante di gravitazione universale, la costante di Boltzmann ed il quanto elementare di carica elettrica (ovverosia la carica dell'elettrone). Ma, a parte considerazioni di opportunità e consuetudine, ciò che determina in ultima analisi fino a che punto si possa tradurre in pratica un simile programma, e quali grandezze siano quindi da considerare fondamentali, è la riproducibilità dei campioni e la precisione con cui è possibile il confronto diretto tra grandezze omogenee.

Appr. 1.2 Il NIST.

Il Dipartimento del Commercio degli Stati Uniti di America promuove l'Istituto Nazionale degli Standard e della Tecnologia. È istruttivo vedere come pubblicano le misure. Qui una raccolta <https://www.nist.gov/pml/productsservices/physical-reference-data>. Nella figura a lato, notate ad esempio come è quotata la massa dell'elettrone. Vedete applicati i concetti

electron mass energy equivalent in MeV	
$m_e c^2 / \text{MeV} = 10^{-6} (m_e c^2 / \text{J}) / (e / C)$	
Numerical value	0.510 998 950 00 MeV
Standard uncertainty	0.000 000 000 15 MeV
Relative standard uncertainty	3.0 x 10 ⁻¹⁰
Concise form	0.510 998 950 00(15) MeV

Massa Elettrone riportata dal NIST

di valore stimato, deviazione standard, e deviazione relativa standard. La forma concisa la vedremo nel prossimo capitolo. L'unita' di misura MeV sta per megaelectronvolt, ovvero un milione di electronvolt, che e' una unita' di misura di massa se trascura il fattore c^2 (quadrato della velocita' della luce) che spesso e' posto $c = 1$.

Capitolo 2

Presentazione numerica e grafica di misure

Quest'opera è distribuita con licenza Creative Commons "Attribuzione – Non commerciale – Condividi allo stesso modo 3.0 Italia".



A seguito di una certa misura indiretta di lunghezza ho ottenuto come valore del misurando e incertezza standard:

$$x = 12438.0605 \text{ m} \quad \delta_x = 723.6576 \text{ m}$$

Come mostriamo il risultato? Cosa c'è di sbagliato nel riportarlo come $x = 12438.0605 \pm 723.6576 \text{ m}$?

Nella comunicazione scientifica nulla è lasciato al caso, e ogni singola parola e numero va controllato scrupolosamente. Nel caso sopra, vediamo che l'incertezza δ_x è riportata con 7 cifre (significative). Perché non va bene? In questo capitolo dopo aver descritto il concetto di cifra significative vedremo come presentare correttamente un risultato.

Con l'esperimento del pendolo semplice domestico, abbiamo misurato 30 volte il periodo singolo $x = (1.91, 2.01, 1.99, \dots)$ s. Come comunichiamo le nostre misure? Le elenchiamo tutte? O a forse più senso raggrupparle in classi sapendo che il valore centrale avviene più spesso? Vedremo qui come costruire istogrammi di misure.

Pensiamo ora invece di aver preso diverse misure di spazio percorso e tempo trascorso durante la discesa di una massa lungo un piano inclinato. Abbiamo diverse coppie di dati, ad esempio $x = (1.1, 2.1, 3.0, 3.9)$ cm e $y = (0.9, 4.1, 8.8, 15.7)$ s. Quale è la maniera di riportare i dati sapendo in fondo che si tratta probabilmente di un moto uniformemente accelerato? Si elencano? Si mettono in tabella? Come si struttura una tabella? O si mettono in un grafico? E che grafico? In questo capitolo vedremo alcune tecniche di presentazione dei grafici.

2.1 Cifre significative

Una volta scritto un numero, è sempre possibile calcolarne il numero di cifre significative (*significant figures* in inglese) secondo queste regole:

1. La cifra diversa da zero più a sinistra è la cifra più significativa
2. Se il numero è riportato senza decimali, la cifra diversa da zero più a destra è la meno significativa. Viceversa, se il numero è riportato con decimali, la cifra più a destra è la meno significativa
3. Il numero di cifre dalla più alla meno significativa dà il numero di cifre significative

La ragione della differenza nel punto 2 è chiara: nel caso di notazione decimale, è una scelta di chi scrive il numero decidere quante cifre mettere, ad es. 8.120 e non 8.12. Ciò non è possibile nell'altro caso, ad esempio ho una unica maniera di scrivere 8120.

Esempio:

Tutti questi numeri hanno 4 cifre significative:

1234; 123400; 123.4; 1001; 10.10, 0.0001010, 100.0

Vedremo a breve che la scelta del numero di cifre significative con cui si presenta un risultato di una misura e' una scelta dello sperimentatore legata a precise regole. Di conseguenza e' sempre preferibile usare una notazione in cui la scelta del numero di cifre significative appare deliberata, ad esempio, il numero 12,344,000 mm ha 5 cifre significative, tuttavia non e' possibile sapere se cio' e' dovuto ad una scelta deliberata oppure se per caso la 6 e la 7 cifra fossero 0. Per renderla deliberata possiamo riscrivere il numero in notazione scientifica come

1.234×10^7 mm oppure 12.3×10^6 mm oppure 1234.4 m

Nei casi sopra, abbiamo scelto *deliberatamente* il numero di cifre significative sulla mantissa, rispettivamente: 4, 3, 5. La notazione scientifica (primi due casi) e' spesso usata perche' non richiede necessariamente di usare multipli o sottomultipli di una grandezza, perche' a prima vista si coglie l'ordine di grandezza del numero attraverso l'esponente. La scelta del numero di cifre significative della mantissa e' quindi sempre deliberato.

Tuttavia la scelta del numero di cifre significative non si attua a partire dal valore misurato, **ma dalla sua incertezza**, e solo successivamente si aggiusta il valore del misurando sulla base di questa scelta. Per ora ci limitiamo a dire che di norma il numero di cifre significative sulla incertezza deve essere 1 massimo 2. Vedremo piu' avanti ([Capitolo 7](#)) il perche'. Ci basti intuire che se scrivo un numero risultato di una misura sottintendo che l'incertezza su quel numero e' minore della cifra meno significativa. Ad esempio, se scrivo 1.2 senza specificare incertezza sottintendo che la incertezza e' sulla cifra meno significa (2) e al piu' questa incertezza vale meta' della cifra finale, ovvero $< 0.1/2 = 0.05$. Se scrivo 1.12 sottintendo che conosco questo numero con una incertezza al piu' di $< 0.01/2 = 0.005$. Ecco allora spiegato come mai nella nostra misura di lunghezza iniziale non ha senso stimare la incertezza come $\delta_x = 23.3576$ m. Vedremo infatti ([Capitolo 7](#)) che e' decisamente improbabile saper stimare la *incertezza* di una misura di qualche per cento anche avendo misurato una grandezza a una parte su un miliardo. Come possiamo riscrivere l'incertezza con un numero diverso di cifre significative?

$\delta_x = 23.3$ m 3 cifre significative

$\delta_x = 23$ m 2 cifre significative

$\delta_x = 20$ m 1 cifra significativa

Delle tre proposte, la prima contiene 3 cifre significative e NON va usata. Queste scrittura per l'incertezza costituisce un errore formale nel caso le incertezze non siano note per meno di una parte su mille. Nella scrittura con 2 cifre significative abbiamo arrotondato, e cosi' abbiamo fatto nel caso della scrittura con 1 cifra significativa. La procedura di arrotondamento segue la nota regola ($0 \leq d < 5$) $\rightarrow d^* = 0$, ($5 \leq d < 9$) $\rightarrow d^* = 10$).¹

Una volta fissato il numero di cifre significative, in conseguenza si scrive il valore del misurando. Allora a partire dalle stime $x = 12438.0605$ m e $\delta_x = 723.6576$ m possiamo riportare il risultato in maniera corretta nelle seguenti modalita' (tra le altre):

¹Tuttavia e' spesso lecito approssimare per eccesso l'incertezza, in maniera da riportare il risultato in maniera piu' conservativa. Nel caso sopra quindi si poteva riportare l'incertezza ad una cifra significativa come $\delta_x = 30$ m. E' chiaro che questa procedura va valutata anche in base alla precisione corrispondente della misura. Nel caso di buona precisione, arrotondare per eccesso non ha molta influenza. Ad esempio nel caso sopra: $23.3/12438.0605 = 0.00187$; $23/12438.0605 = 0.00185$; $30/12438.0605 = 0.00241$. La precisione rimane in tutti i casi attorno al 2 per mille.

$x = (12438 \pm 23) \text{ m}$	Due cifre significative
$x = 12440 \pm 20 \text{ m}$	Una cifra significativa
$x = (12440 \pm 30) \text{ m}$	Una cifra significativa approssimata per eccesso
$x = 12.438 \pm 0.023 \text{ km}$	Due cifre significative, uso multipli
$x = (12.44 \pm 0.02) \text{ km}$	Una cifra significativa, uso multipli
$x = (1243.8 \pm 2.3) \times 10^2 \text{ m}$	Due cifre significative, notazione scientifica
$x = (1.244 \pm 0.002) \times 10^5 \text{ m}$	Una cifra significativa, notazione scientifica
...	

Come si vede, la parentesi non e' strettamente necessaria tranne nel caso di notazione scientifica. Attenzione. Durante i conti invece se possibile si tiene il numero MASSIMO di cifre significative. Gli "arrotondamenti" vanno fatti solo "alla fine". Ovviamente questo non e' sempre possibile ma e' facile con i computer. Infine, nel caso si sommino addendi con numero diverso di cifre significative, il risultato va espresso con il numero di cifre significative piu' basso tra gli addendi. Invece nel caso di prodotti e' possibile esprimere il risultato con il numero di cifre significative piu' alto tra i fattori, ad esempio $1.1 + 1.1231 = 2.2$; $1.1 \cdot 1.121 = 1.23$. E' sbagliato invece: a) mescolare unita' di misura ($120.2 \text{ m} \pm 3.1 \text{ dm}$), b) non accordare il numero di cifre significative tra misurando e incertezza ($120.2 \pm 3.123 \text{ cm}$); c) riportare un numero eccessivo di cifre significative.

Esercizio:

Quante cifre significative ci sono nei seguenti numeri? Quale e' la cifra piu' significativa e quella meno significativa?

976.45; 84000; 0.0094; 301.07; 4.000; 10; 5280; 400; 4.00×10^2 ; 3.010×10^4

Approssimare con due cifre significative i numeri sopra.

2.2 Istogrammi

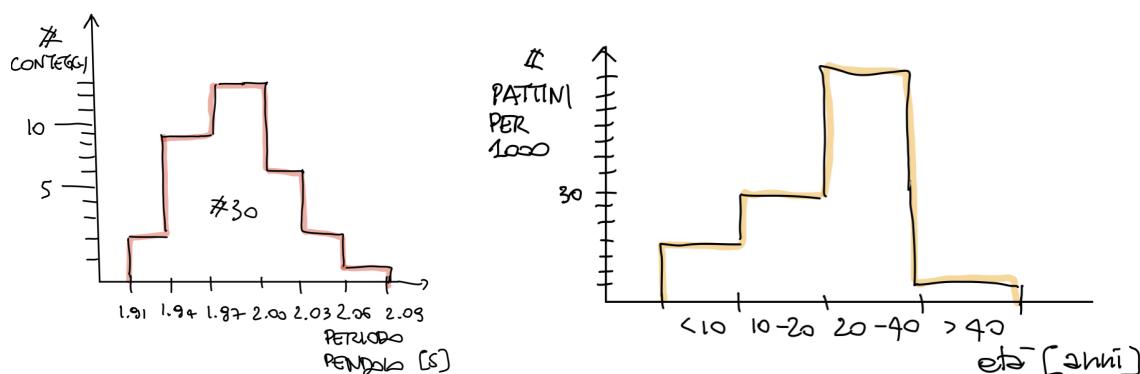


Figura 2.1: Esempio di istogrammi. A sinistra, misure ripetute. A destra, un campionamento generico di una grandezza in classi di frequenza

Avviene spesso che durante un esperimento si prendano diverse misure della stessa quantità, a volte per migliorare la qualità del risultato, come vedremo, altre volte semplicemente per motivi generici di campionamento (vedi Fig. 3.1. In alcuni casi, le misure si concentrano attorno ad un valore centrale, in altri casi, sono distribuite diversamente. Molto spesso, nel caso in cui si abbiano diverse misure si procede a raggruppare in *classi di frequenza (bin)*. Ad esempio, se abbiamo 30 misure di periodo ottenute tra 1.91 e 2.07 secondi, possiamo pensare di raggrupparle in 5, 10, 15 classi di frequenza. Ad esempio, nel caso di 5 classi possiamo immaginare che siano equamente distribuite tra 1.91 e 2.07 ovvero ampie 0.032 secondi. Avremo quindi la prima classe [1.91, 1.942[, la seconda [1.941, 1.974[e così via. Oppure per evitare numeri con troppe cifre potremmo pensare le classi come ampie 0.03 secondi e distribuite a partire da 1.9 ad esempio [1.9, 1.94[, [1.94, 1.97[e così via. Vedremo che in generale e'

decisamente importante che tutte le misure siano considerate, e che si scelga bene il numero di bin, ne' troppi ne' troppo pochi, mentre la scelta della 'centratura' del bin non e' cosi' rilevante. Ad ogni modo, una volta fissato il numero di bin e la posizione, possiamo costruire una tabella con le classi di frequenza ed i conteggi assoluti e relativi, ad esempio:

Bin	Frequenza Assoluta (Conteggi)	Frequenza relativa Conteggi/N
1.91-1.94	2	0.06*
1.94-1.97	9	0.3
1.97-2.00	13	0.43
2.00-2.03	7	0.23*
2.03-2.06	3	0.1
2.06-2.09	1	0.03*
Somma	30	1

Torneremo piu' avanti sulla scelta del numero di conteggi. Per ora vediamo che gia' in forma tabellare, comincia ad apparire chiaro che le misure hanno una maggiore frequenza al centro, attorno al valore 2 s, con delle 'code' di valori maggiore e minori. Tuttavia, per farci questa opinione dobbiamo visionare diversi numeri. Una maniera piu' diretta di riportare questi dati e' quello di farne un *istogramma*. L'istogramma e' un grafico a colonne in cui la base e' la classe di frequenza, e l'altezza e' espressa o come frequenza assoluta (conteggi) e come frequenza relativa. In alcuni casi si puo' riportare anche l'istogramma in forma 'normalizzata' ovvero in maniera che l'area dell'istogramma sia pari a 1. In questo caso l'altezza della barra e' pari alla densita' di probabilita'. Su questo torneremo.

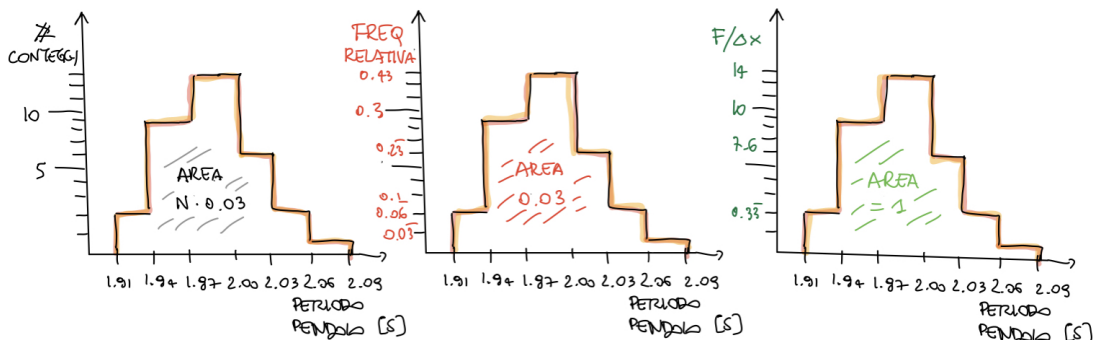


Figura 2.2: Lo stesso istogramma riportato a sinistra con i conteggi (frequenza assoluta), al centro con la frequenza relativa (conteggi normalizzati) e destra con la densita' di frequenza.

In Fig. 3.1 riportiamo gli istogrammi nei 3 casi descritti sopra. Si vede come la forma dell'istogramma non cambia, cambia solo l'asse delle ordinate e quindi il significato dell'altezza della barra. Anzitutto notiamo abbiamo una certa liberta' nel descrivere le classi di frequenza: possiamo riportare sull'asse gli estremi (1.91,1.94,1.97...) oppure il loro valore centrale (1.925, 1.955, ...). In generale, si usa maggiormente la seconda scelta ma entrambe vanno bene. Riguardo alla scelta tra frequenza assoluta, relativa e densita' di frequenza, consideriamo che (bw e' l'ampiezza delle classi di frequenza):

	Somma	Area
Frequenza assoluta	N	$N \cdot bw$
Frequenza relativa	1	bw
Densita' di frequenza	$1/bw$	1

Tabella 2.1: Somma delle altezze dei rettangoli e area racchiusa sotto l'istogramma nei 3 casi analizzati.

Vediamo che attraverso questo grafico possiamo giudicare con l'occhio molto bene la distribuzione dei dati, senza leggere i numeri in dettaglio, ma solo dall'impressione globale. Vedremo che ci sono diverse maniere statistiche per quantificare questa nostra impressione. Si pone l'attenzione ad un

fatto importante: con la costruzione delle classi di frequenza si e' PERSA informazione rispetto alla lista completa di misure, infatti, mentre inizialmente le misure erano note individualmente, ora ne conosciamo solo il numero complessivo dentro la classe. Se la classe e' particolarmente larga, maggiore sara' la perdita di informazione.

Come mai allora si usano gli istogrammi? Ci sono diverse ragioni fondamentali:

- Anzitutto nella maggior parte dei casi la perdita di informazioni rispetto all'elencazione dei singoli valori non influisce sulla determinazione del misurando o sulla precisione della misura
- in secondo luogo, questa perdita di informazioni e' controbilanciata da un aumento di capacita' analitica: mentre non e' possibile per l'uomo giudicare velocemente sequenze di decine di numeri, diventa immediato giudicare il loro istogramma
- Vedremo piu avanti quando parleremo di confronto di ipotesi, che raggruppare i dati in classi di frequenza permette di studiarli attraverso la distribuzione binomiale di probabilità, che e' quella che si applica nel caso di conteggi.

Si fa notare come poi a volte l'istogramma si usi invece solo come 'supporto grafico' e in realta' la analisi dei dati si faccia al computer considerando tutti le misure singolarmente.

Scelta del numero di classi di frequenza (*bin*)

La scelta del numero di classi di frequenza dipende dalla natura della distribuzione di dati, ma in generale viene fatta seguendo una regola piuttosto generica: il numero di bin non deve essere ne' troppo basso ne' troppo alto. Nel primo caso infatti, otterremo un istogramma in cui sara' difficile studiare l'andamento delle code. Nel secondo caso, la maggior parte delle classi sara' vuota o popolata da al piu' una o due misure. Un esempio dell'effetto e' mostrato in [Figura 2.3](#). Istantogrammi di questo tipo NON vanno riportati nei report scientifici.

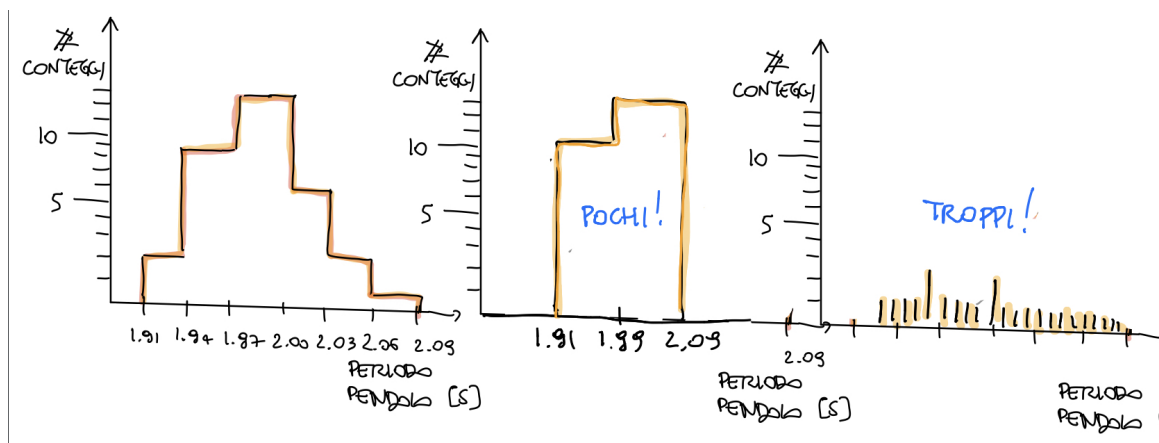


Figura 2.3: Lo stesso istogramma riportato a sinistra con un numero di classi di frequenza ragionevole, al centro con un numero di classi di frequenza troppo basso, e a destra con un numero di classi di frequenza troppo elevato.

Al di la' di questi casi estremi, bisogna poi di norma verificare che la scelta del numero di bin non alteri il messaggio visivo trasmesso. A volte non e' possibile fare troppe prove (ad esempio, si hanno molti istogrammi da generare) o si preferisce non farlo per non influenzare il risultato. In questo caso, c'e' ampia letteratura su come calcolare in maniera automatica il numero di bin. Si veda piu' sotto negli approfondimenti ad esempio le regole di Rice e Sturges basate solo sul numero di dati N o le piu' complesse regole di Scott e Freedman-Diaconis basate anche sulla distribuzione dei dati. Per le nostre applicazioni, una regola semplice che va bene e' partire da un valore di riferimento che potrebbe essere ad esempio la radice quadrata del numero di eventi: $n_{bins} = \sqrt{N}$. Dopodiche' va comunque fatta una ispezione visiva, nel senso che, al variare di poco del numero di bin attorno a \sqrt{N} l'aspetto dell'istogramma non dovrebbe variare: non dovrebbero apparire 'punte' o 'vuoti'. Se cio' avviene, ci potrebbero essere due cause. Si consideri infatti che l'istogramma raggruppa dati vicini che cadono o in un bin o nell'altro adiacente a seconda della scelta del numero di bins. Ecco cosa potrebbe accadere:

- appare una punta o un vuoto puramente casualmente. I dati erano distribuiti in maniera tale che per caso molti sono capitati in un bin piuttosto che in un altro. A volte si chiama questa

distorsioni per effetto del binning. Se possibile, evitare scelte del binning che distorcono l'istogramma. Si noti che piu' numeroso e' il campione minore e' la probabilita' di incorrere in una distorsione casuale

- *non* appare una caratteristica dei dati che in realta' e' presente. Supponiamo infatti che ci sia un valore che si ripete sempre uguale il 20% delle volte. A titolo esemplificativo, si supponga che si stia misurando una frequenza su un circuito. Capita spesso di 'pigliare' anche la frequenza di 50 Hz classica delle reti elettriche. Questa diventa un elemento spurio della nostra distribuzione di dati che si ripete in maniera regolare senza fluttuazioni. E' sempre 50 Hz. Se il binning dell'istogramma e' troppo basso, potrebbe succedere che non ci si accorge di questa serie di dati spuri. Solamente aumentando il binning ad un certo punto appare una 'punta' attorno ai 50 Hz. In questo caso l'istogramma ci permette quindi di individuarla facilmente e nel caso rimuovere quei dati. Si veda [Figura 2.4](#)

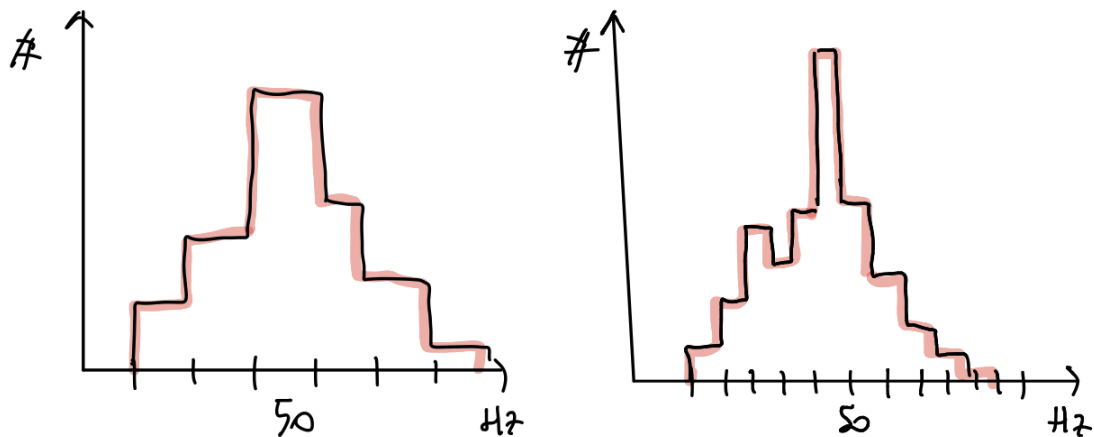


Figura 2.4: Lo stesso istogramma riportato a sinistra con un numero di classi di frequenza ragionevole. A destra, il numero di classi maggiore ha mostrato un eccesso 'anomalo' attorno ai 50 Hz che l'istogramma a sinistra non mostra.

Non si tema, il calcolo del numero di bin si apprende con l'esperienza e la tipologia di misure che si andra' a svolgere.

2.3 Grafici

Gli studi scientifici fanno ampio uso di grafici, e non e' possibile ne ha senso cercare di presentarne nemmeno i principali. Ci limiteremo a discuterne due o tre di ampio uso durante il corso, e poi per l'analisi dei dati: il grafico che rappresenta l'andamento di una misura in funzione di un'altra, e il grafico che rappresenta l'andamento temporale delle misure. Infine discuteremo anche un grafico di solito meno usato, ma talvolta utile che e' quello dell'andamento cumulativo delle misure.

2.3.1 Andamento temporale delle misure

Nel caso si abbiano decine di misure, ad esempio quelle del nostro pendolo, ha senso chiedersi che andamento hanno *con l'avanzare del tempo*. Possiamo perciò costruire un grafico in cui mettiamo sull'asse delle ascisse o l'effettivo tempo trascorso dall'inizio della presa dati, oppure talvolta piu' semplicemente il numero consecutivo della misura, che e' in rapporto lineare al tempo trascorso se la distanza temporale tra le misure e' costante. Con questo grafico, in cui mostriamo un esempio in Fig. 2.5 possiamo osservare diverse cose tra cui: l'andamento generale dei dati rispetto alla media, la stabilita' dei dati e della loro media rispetto al tempo, la ragionevole presenza di dati anomali. Ad esempio, un effetto di deriva della misura potrebbe essere osservato attraverso l'andamento temporale di queste.

Sull'asse delle ordinate, e' possibile mettere diversi parametri di interesse, dal valore della singola misura, al valore della media, o degli scarti quadratici.

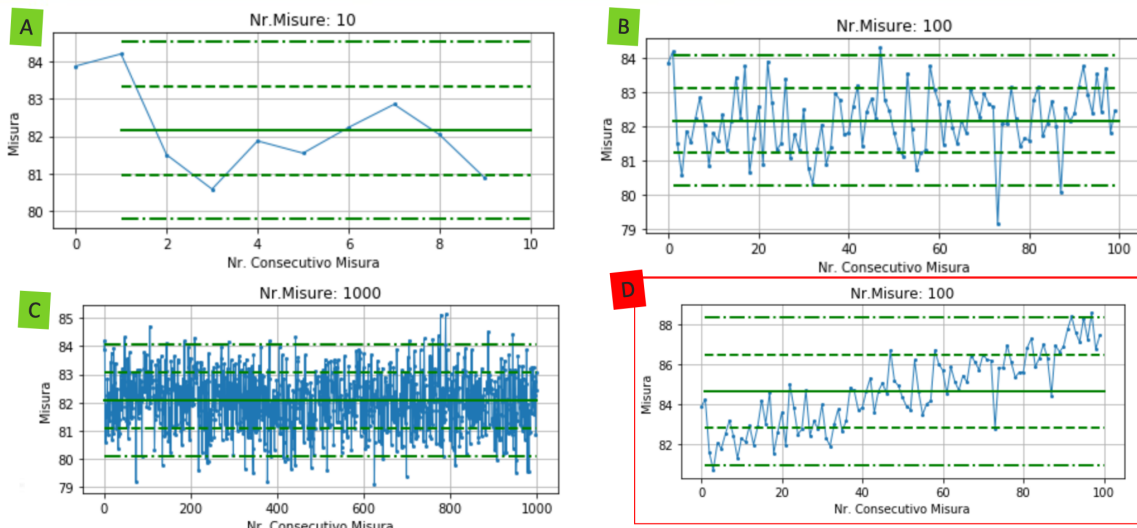


Figura 2.5: Esempi di andamento temporale delle misure ottenuti da estrazioni casuali per campioni con diversa numerosità: 10 misure (A), 100 misure (B), 1000 misure (C). Il caso D rappresenta un campione di misure in cui si è introdotta una deriva temporale di tipo $x = ax'$

2.3.2 Andamento cumulativo delle misure

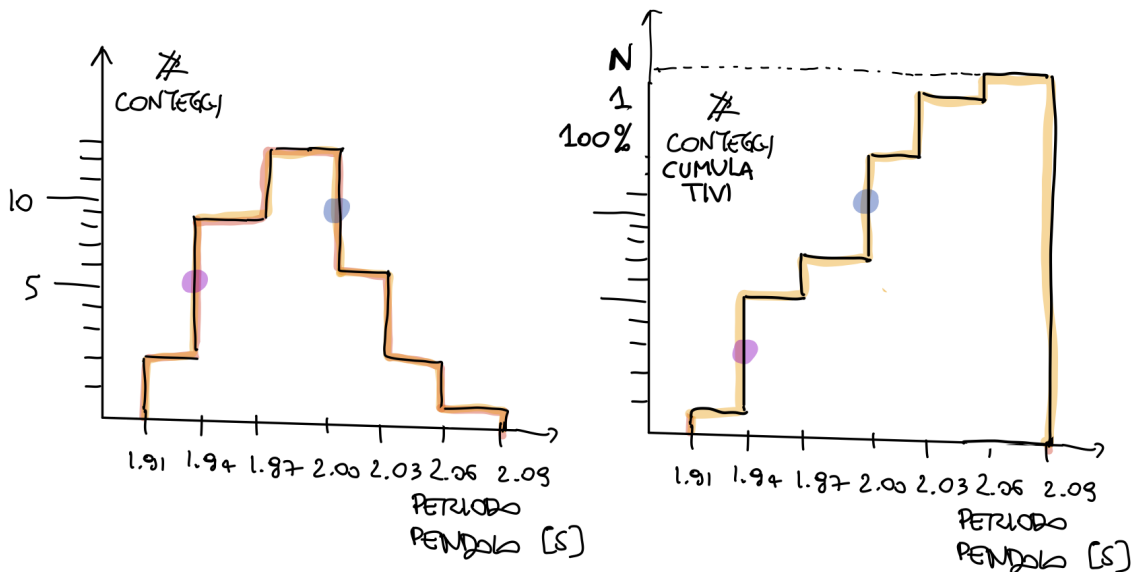


Figura 2.6: Istogramma delle misure (sinistra) e corrispondere andamento cumulativo delle misure (destra). I due segmenti colorati hanno la stessa altezza. L'ordinata del grafico cumulativo può essere espressa in numero dei conteggi, percentuali, o frequenza relativa.

Talvolta ha invece senso chiedersi quanti dati sono minori (o maggiori) di un certo valore? Ad esempio, nella nostra distribuzione delle misure del pendolo, quanti dati stanno sopra i 2 s? Per fare questo bisogna ordinare i dati in ordine crescente, e poi può avere senso riportare nel grafico cumulativo in cui sulle ascisse ci sono le misure, e sulla ordinate il numero (assoluto o relativo) di conteggi inferiori a quella misura². Più avanti vedremo che nel caso si conosca la distribuzione statistica dei dati, allora è nota anche la distribuzione cumulativa dei dati.

²Esistono test statistici che valutano la discrepanza di una distribuzione di dati da una ipotetica distribuzione di probabilità basati sulla distribuzione cumulativa, ad esempio il test di Kolmogorov-Smirnov https://en.wikipedia.org/wiki/Kolmogorov-Smirnov_test che non tratteremo in questo corso.

2.3.3 Grafico di una variabile contro un'altra

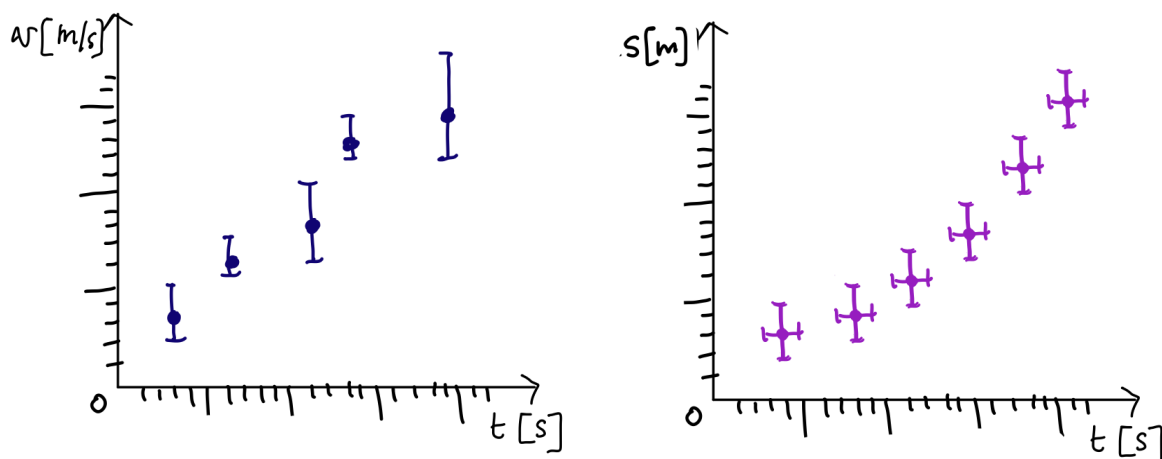


Figura 2.7: Andamento congiunto di due serie di misure. Andamento apparentemente lineare con incertezze solo sull'asse delle ordinate ed errori diversi punto per punto (sinistra). Andamento apparentemente a legge di potenza con incertezze su ascisse e ordinate uguali per tutti i punti (destra)

Spesso e' necessario misura a coppie o a gruppi diverse grandezze di input e da queste calcolare un output sempre che sia nota la relazione matematica. Nel caso di due variabili di input (o di 3 a dire il vero, anche se piu' complesso) e' possibile rappresentare graficamente due serie di misure una *contro* l'altra. Supponiamo ad esempio che si stia stimando la velocità di una massa che cade su un piano inclinato a partire da misure di spazio e di tempo. Ci saranno tanti grafici diversi tante quante relazione tra variabili, ma nel nostro caso sappiamo che ci aspettiamo un andamento parabolico. Per stimare un andamento del genere ha senso *campionare* le due grandezze, ovvero misurare diversi valori di una e poi controllare il valore in corrispondenza dell'altra. Essendo le misure soggette a fluttuazioni casuali, piu' misure abbiamo piu' andremo a sondare la relazione. Se prendessimo un solo punto, potremmo si calcolare $v = x/t$ ma non avremo certo verificato un andamento parabolico o una legge di moto uniformemente accelerato.

Nel rappresentare questa relazione graficamente, almeno nel caso bidimensionale di due variabili ad esempio s, t , allora sul grafico cartesiano riporteremo gli assi coordinati, le grandezze e le unita di misura, le coppie di punti (t_i, s_i) ciascuna con il proprio errore δ_{t_i} relativo alle misure dei tempi e δ_{s_i} relativo alle misure delle distanze. Il punto avra' quindi sempre l'aspetto di una croce con i doppi errori.

Vedremo che la scelta su quale delle due variabili vada messa nelle ascisse non e' fondamentale, ma e' spesso deliberata, mettendo su quest'asse la grandezza meglio nota delle due, ovvero quella con incertezza relativa standard minore. Piu' punti avremo meglio la relazione sara' studiabile. Inoltre, a partire da questo grafico quantificheremo l'importante valore della correlazione lineare tra grandezze. Anche in questo caso, riportare un grafico e' molto piu' efficace di riportare unicamente i singoli valori (t_i, s_i) .

2.4 Concetti Finali

Ricapitolando

Quesiti di autovalutazione

Es. 2.1 Quante cifre significative ci sono nei seguente numeri? Quale e' la cifra piu' significativa e quella meno significativa?

9.7645; 840.00; 0.94; 301.070; 4000;

10001; 5.280; 0.100; 3.009×10^2 ; 3010×10^4

□ **Es. 2.2** Approssimare con due cifre significative i numeri dell'esercizio sopra.

□ **Es. 2.3** Riscrivere in modo corretto le seguenti misure:

$$t = 3.624 \text{ s} \pm 1'25.03''; \quad l = 1.0501 \text{ m} \pm 257 \text{ mm};$$

$$e = 1.6034567 \times 10^{-19} \text{ C} \pm 3.5 \times 10^{-21} \text{ C}; \quad g = 9.807 \text{ m s}^{-2} \pm 5.3 \text{ cm s}^{-2};$$

$$m = 0.0052345 \text{ kg} \pm 1.3 \times 10^{-2} \text{ g}; \quad L = 10^4 \text{ N m} \pm 10^5 \text{ g m}^2 \text{ s}^{-2}$$

Rappresentare sotto forma di istogramma, valutando il numero di classi di frequenza, delle seguenti serie. Valutarne l'andamento temporale e cumulativo:

$$x_1 = [3, 7, 3, 7, 12, 8, 9, 7, 5, 7, 12, 8, 6, 6, 7, 6, 7, 8, 9, 8, 5, 10, 8, 8, 8]$$

$$x_2 = [73, 91, 72, 81, 82, 46, 89, 75, 62, 58, 73, 46, 64, 61, 50, 89, 91, 82, 71.$$

$$76, 69, 70, 82, 90, 63, 70, 94, 44, 100, 88, 56, 94, 51, 79, 63, 87, 54, 100, 72, 81]$$

□ **Es. 2.4** Si rappresenti in un istogramma e in un istogramma cumulativo per la seguente raccolta campionaria di misure di massa (conteggio tra parentesi):

$$15.10(2), 15.11(9), 15.12(13), 15.13(7), 15.14(3), 15.16(1)$$

□ **Es. 2.5** Si rappresenti in un grafico una serie di coppie di punti

$$(1.1, 1.93), (1.9, 4.80), (2.9, 11.11), (4.2, 24.20), (5.1, 36.73)$$

Si considerino i seguenti casi: a) incertezza su ascissa trascurabile e su ordinata pari a $\delta_{y_i} = 0.03 y_i$ b) incertezza su ascissa $\delta_{x_i} = \delta_x = 0.1$ e su ordinata pari a $\delta_{y_i} = \delta_y = 1.5$. Si ragioni su che tipo di relazione possa sussistere tra ascisse e ordinate. Si valuti l'uso della scala logaritmica.

Approfondimenti

Appr. 2.1 Regole per la scelta delle classi di frequenza. Abbiamo visto che il numero delle classi di frequenza va scelto in maniera che non siano ne' troppi ne' troppo pochi e in maniera che non siano 'nascoste' caratteristiche particolari dell'istogramma o viceversa che non siano 'mostrate' caratteristiche strane che in realta' non siano proprie delle misure ma solo una casualità. A questa considerazione si affiancano tuttavia regole numeriche, ma si sottolinea fin da subito come sara' spesso naturale fare alcune prove sul numero di bin in maniera da rappresentare al meglio la distribuzione, mostrandone le proprieta' salienti invece che perdendole. Le prime tre regole non sono molto diverse tra loro, funzionano bene soprattutto per distribuzioni centrali e con molte misure. Si noti come non considerino in nessun modo la distribuzione dei dati, ma solo il loro numero.

- **Radice quadrata** Una prima veloce prova di puo' fare calcolando la radice quadrata del numero di conteggi: $n_{bin} = \sqrt{N}$. Questa approssimazione va bene soprattutto se le distribuzioni di dati sono 'centrali'
- **Regola di Sturges** In questo caso si usa una funzione un po' piu' complessa del numero di misure: $n_{bin} = 1 + \log_2(N)$. Essendo una funzione piu' lenta di N, questo permette di non avere troppi bin nel caso N sia grande.
- **Regola di Rice** Si ha sempre una funzione del numero di misure, ma diversa $n_{bin} = 2 N^{1/3}$

Due ulteriori regole invece prendono in considerazioni la forma della distribuzione dei dati e non solo il loro numero:

- **Regola di Scott:** Il numero di bin e' calcolato in funzione dello scarto quadratico medio s_x dei dati: $n_{bin} = 3.5 s_x N^{-1/3}$

- **Regola di Freedman-Diaconis** Forse la migliore. Il numero di bin e' calcolato a partire da un valore detto $IQR = q_{75} - q_{25}$ pari all'ampiezza dell'intervallo tra il 25esimo e il 75esimo quantile: $n_{bin} = 2 IQR N^{-1/3}$. IQR sta per *interquantile range* e sar  definito in [sezione 5.6](#).

Esercizio:

Per esercizio si disegnano le prime tre curve sopra in funzione del numero di misure N . Assumendo poi la distribuzioni di dati come in tabella sopra, si disegnano anche le due finali.

Appr. 2.2 Binning di larghezza non uniforme. Nella maggior parte dei casi, si usano bin di uguale larghezza. Tuttavia, l'istogramma non si usa solo per misure ripetute, ma ogni volta che serva raggruppare le misure. In alcuni di questi casi si possono usare bin la cui larghezza si 'adatta' a fattori speciali (esempio: voglio che ogni bin contenga lo stesso numero di conteggi del successivo): in questo caso si parla bin adattivo [[Scargle et al., 2013](#)].

Appr. 2.3 Le scale logaritmiche e semi-logaritmiche. L'uso della scala logaritmica e' assai diffuso ed e' importante chiarirlo. Date due misure x, y spesso risulta utile rappresentare graficamente le serie in scala semilogaritmica $x, \log(y)$ o $\log(x), y$ oppure in scala logaritmica $\log(x), \log(y)$. Non ha sempre senso farlo, ma in alcune occasioni e' decisamente funzionale, ad esempio:

- se le misure x o y sono distribuite in maniera disomogenea su ampi intervalli, allora conviene 'compattarle' in scala logaritmica. Ad esempio:

1.2, 1.5, 5.1, 127.1, 13121.0

se disegnati in scala lineare, non si potrebbero apprezzare i primi tre numeri perche' troppo vicini, pero' passando a loro logaritmo:

0.079, 0.17, 0.7, 2.1, 4.11

l'estensione della scala diventa piu' compatta

- se si vuole verificare una relazione tipo la **legge di potenza** $y = a x^k$ spesso conviene passare in scala logaritmica in quanto

$$\log(y) = \log(a x^k) = \log(a) + k \log(x)$$

, allora ponendo $\hat{y} = \log(y), \hat{x} = \log(x)$, il grafico

$$\hat{y} = \log(a) + k \hat{x}$$

appare lineare con pendenza k e intercetta $\log(a)$.

- se si vuole verificare una relazione tipo **esponenziale** $y = a e^{bx}$ spesso conviene passare in scala semi-logaritmica sull'asse delle ordinate, in quanto

$$\log(y) = \log(a e^{bx}) = \log(a) + b x$$

, allora ponendo $\hat{y} = \log(y)$, il grafico

$$\hat{y} = \log(a) + b x$$

appare lineare con pendenza b e intercetta $\log(a)$.

In entrambi i casi, si presti attenzione a come di disegnano gli assi: nelle scale logaritmiche i separatori sono posizionati non equispaziati ma essi stessi in posizioni logaritmiche, e quindi in questo caso, supponiamo che si tratti dell'asse delle ascisse, l'asse riporta numericamente i valori x ma non $\log x$. Si dice quindi che sia *la scala che fa il logaritmo*. L'unita dell'asse insomma e' x . Talvolta, si riporta solo l'esponente del multiplo, allora in questo caso l'unita' da riportare sul nome dell'asse e' $\log x$.

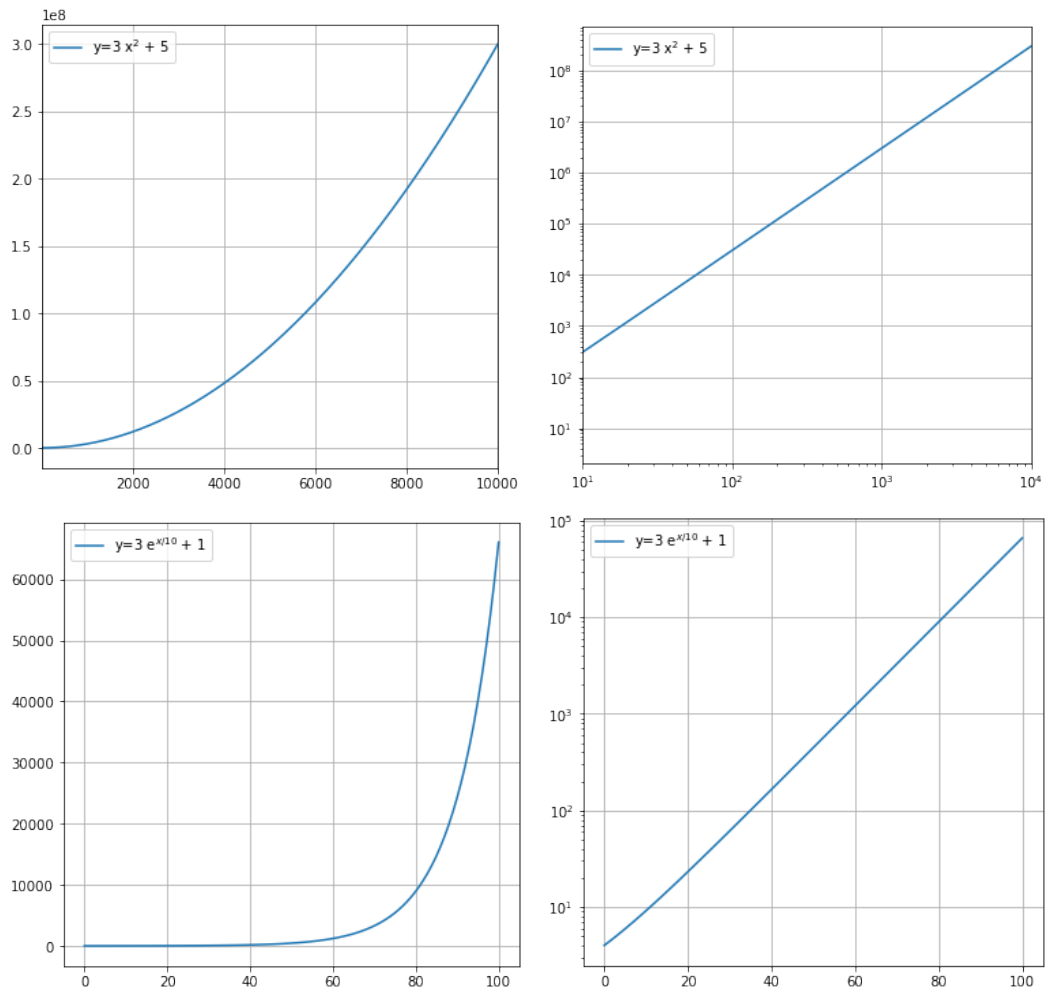


Figura 2.8: Scale logaritmiche e lineari. (Sopra) scala lineare e logaritmica per una legge di potenza. (Sotto) scala lineare e semilogaritmica per una legge esponenziale

Capitolo 3

Misure singole, misure ripetute. La Gaussiana

Quest'opera è distribuita con licenza [Creative Commons](#) "Attribuzione – Non commerciale – Condividi allo stesso modo 3.0 Italia".



Questo capitolo si propone di cominciare la discussione su come si affronta il tema della stima del valore centrale del misurando e della sua incertezza. Cominceremo affrontando il tema in maniera piuttosto intuitiva, mentre per comprendere le ragioni che portano alle conclusioni a cui qui arriveremo, bisogna aspettare di aver introdotto le funzioni di distribuzioni di probabilità, le loro proprietà, e per fare questo la teoria della probabilità. Torneremo quindi su questo argomento più avanti, nel capitolo

Quando si fa una misura si ricordi quali sono le linee guida:

1. Si deve ideare la misura in maniera che lo strumenti misuri effettivamente quello che ci siamo proposti, ovvero che non ci sia discrepanza tra il risultato della misura e il misurando aspettato, e che il misurando a sua volta abbia discrepanza minore possibile rispetto ad un valore vero di riferimento
2. Si deve ideare la misura in maniera che lo strumento dia risultati precisi, dove il livello di precisione richiesto deve essere se possibile valutato prima della misura, in accordo a cosa si vuole ottenere con la misura stessa
3. Si deve dare la miglior stima del misurando, ovvero quella con minor discrepanza rispetto al valore vero
4. Si deve dare la miglior stima dell'incertezza della misura, non nota a priori, ovvero quella intrinseca che si avrebbe se il misurando fosse noto in maniera perfetta e lo strumento fosse perfetto. Sottolineiamo ancora una volta come si tratta di una STIMA e non di una MISURA della incertezza.
5. Una volta stimato il misurando e la incertezza si deve riportare il risultato numerico in maniera appropriata.

Come fare questo *in ogni situazione*? La risposta e' tutt'altro che ovvia, e la disciplina sperimentale e' di fatto una disciplina di apprendimento continuo a seconda della situazione in cui ci si trova, tuttavia, ci sono diverse vie maestre. Si avvisa subito il lettore che, come discuteremo, ci sono diverse ragioni per cui e' assolutamente probabile che non vi sia una maniera *esatta* di arrivare ad una stima, ma ve ne siano diverse *ragionevoli*. E' sempre importante quindi percorrere diverse vie ragionevoli, se tutte portano a risultati simili, allora la stima e' robusta.

Qui di seguito partiamo dal caso di misure singole per poi passare a quello di misure ripetute. A prima vista non e' chiaro perché si faccia questa scelta, lo apparirà durante la narrazione. Per le stime di grandezze, a partire da ora useremo il simbolo 'cappello' sopra il nome della variabile, ad esempio $\widehat{x^*}$.

3.1 Misure Singola

Ci sono diversi casi in cui per ragione o possibilità, si ha a che fare con misure singole. Partiamo dal caso di ragione. Consideriamo che si faccia una pesata con una bilancia digitale con risoluzione di 2kg. La persona che si misura la prima volta ottiene un certa misura, diciamo 82 kg, una seconda volta ottiene ancora 82 kg e così via. La persona ha ripetuto la misura ma cosa ha imparato? Da questa ripetizioni si può dedurre una cosa soltanto: *la risoluzione dello strumento è maggiore delle fluttuazioni casuali della misura*. Cosa significa questo? La frase da per scontato che esistano fluttuazioni casuali nella misura, cosa che abbiamo visto in precedenza, qui si ribadisce solo che lo strumento non è in grado di *apprezzarle*. Se infatti ora cambiamo strumento, e prendiamo una bilancia con una risoluzione di 0.1 kg, ecco che sicuramente vedremo oscillare le pesate attorno a 82 kg: 81.9 kg, 82.0 kg etc. In questo caso appunto, la risoluzione è in grado di apprezzare le fluttuazioni statistiche. Quale delle due misure è migliore? La seconda di sicuro, ma la ragione diventerà evidente solo a breve. Essa sta nel fatto che l'osservazione della distribuzione della fluttuazioni casuali segue una legge ben nota (la Gaussiana) e questo ci permetterà di assegnare una incertezza decisamente inferiore alla stima del misurando.

Torniamo al caso della singola misura. Se non apprezziamo le fluttuazioni casuali, allora si tratta appunto di una singola misura. L'altro caso è quando la misura per vari motivi non è ripetuta o non si può ripetere. È chiaro che in questo caso la stima migliore del valore vero non è altro che la singola misura stessa, diciamo x_1 . Quale è la stima della incertezza? Pensiamo alla risoluzione dello strumento. Una diversa maniera di pensare alla risoluzione è la seguente: *al di sotto della risoluzione, non è possibile apprezzare variazione dell'indicatore della misura*. Ovvero, se le variazioni sono più piccole della risoluzione, non le vediamo. Questo ci guida nella risposta alla domanda. La misura potrebbe variare infatti (certamente lo fa) entro la risoluzione, semplicemente noi non ce ne accorgiamo. Cosa possiamo concludere sulla incertezza? Certamente possiamo concludere che il valore del misurando è *entro la risoluzione* dello strumento. Convertendo questo in formula:

Nel caso di una misura singola ottenuta con uno strumento con risoluzione R , una maniera plausibile di riportare il risultato è

$$\widehat{x^*} = x_1 \pm \frac{R}{2} \quad (3.1)$$

Con la scrittura sopra, stiamo tacitamente dicendo che siamo sicuri (a meno di incertezze sistematiche, che riteniamo nulle in questo momento), che il risultato si trovi entro la risoluzione dello strumento. La certezza si può anche tradurre in questo concetto: la probabilità che il valore vero del misurando si trovi nell'intervallo $[x_1 - R/2, x_1 + R/2]$ è 100%.

Esercizio:

Se la probabilità che la misura si trovi nell'intervallo $[x_1 - R/2, x_1 + R/2]$ è 100%, quale è la probabilità di ottenere esattamente $x = x_1$? Pensateci, ma per la risposta dovrete attendere qualche lezione.

Esercizio:

Se la probabilità che la misura si trovi nell'intervallo $[x_1 - R/2, x_1 + R/2]$ è 100%, quale è il livello di confidenza con il quale il risultato viene mostrato?

Dire che la probabilità che il valore vero del misurando si trovi nell'intervallo $[x_1 - R/2, x_1 + R/2]$ è 100% è analogo a dire che abbiamo il 100% di confidenza che il risultato vero si trovi tra $[x_1 - R/2, x_1 + R/2]$. Per questa ragione, questo intervallo è detto *dispersione Massima*, e il valore $R/2$ è detto *semidispersione massima*. In questa notazione quindi la nostra stima della incertezza è stata fatta attraverso la semidispersione massima.

Perche in Eq. 3.1 abbiamo scritto che si tratta di una maniera plausibile? La ragione sta nel fatto che siamo *certi* che la stima della incertezza attraverso la semidispersione massima e' una *sovrastima* di quella che sarebbe l'incertezza sulla misura nel caso avessimo uno strumento con migliore risoluzione. Per questa ragione spesso (ed e' quello che consigliamo) si da una diversa stima (ridotta rispetto alla semidispersione massima) della incertezza sulla misura singola, *basandosi su una ipotesi riguardo alla probabilità della distribuzione di misure*. Ci spieghiamo subito: nel caso di uno strumento analogico, come un righello millimetrato ad esempio, abbiamo visto come la risoluzione sia la frazione di tacca che possiamo apprezzare con l'occhio. Tuttavia, dopo aver misurato e ottenuto a esempio (21.1 ± 0.2) cm, usando la semidispersione massima, mentre da un lato diciamo che siamo sicuri al 100% che la misura stia tra $[20.9, 21.3]$, d'altro canto pensiamo che sia piu' probabile che sia 21.1 piuttosto che 20.9 e 21.3, gli estremi. Questa ipotesi, come vedremo, si puo' formalizzare dicendo che crediamo che la distribuzioni di probabilità sia piccata (in forma triangolare) sul valore centrale. Se utilizziamo questa ipotesi, vedremo che la miglior stima di incertezza non e' la semidispersione massima. La stessa cosa non si puo' invece fare nel caso di uno strumento digitale. Se avesso un calibro digitale con risoluzione 0.2 cm, e leggessimo 21.1 cm, non avremmo modo di sapere se il valore vero era ad un estremo o ad un altro. In questo senso, la probabilità dei valori e' uniforme tra i due estremi $[20.9, 21.3]$.

Nel caso della misura singola, se lo strumento e' analogico di risoluzione R e assumiamo che la distribuzione di probabilità della misura sia triangolare attorno al valore misurato, allora una maniera plausibile di riportare il risultato e'

$$\widehat{x^*} = x_1 \pm \frac{R}{2\sqrt{6}} \quad (3.2)$$

Se invece si tratta di una misura singola ottenuto con uno strumento digitale di risoluzione R , allora, assumendo che la distribuzione di probabilità sia uniforme attorno al valore misurato, una maniera plausibile di riportare il risultato e':

$$\widehat{x^*} = x_1 \pm \frac{R}{2\sqrt{3}} \quad (3.3)$$

Anzitutto, osserviamo come la stima di incertezza sia ora inferiore a quella ottenuta con la semidispersione massima, di un fattore $\sqrt{3}$ e $\sqrt{6}$ rispettivamente, in linea con il ragionamento sopra. Vedremo che queste stima di incertezza ora cambia nome, si chiameranno *deviazione standard secondo ipotesi triangolare* e *deviazione standard secondo ipotesi uniforme*. Vedremo che l'uso delle deviazioni standard e' ampiamente raccomandabile rispetto a quello della semidispersione massima (che potremo chiamare anche 'deviazione massima'). Ci rimane un dubbio, qual e' la probabilità che il valore vero sia tra $[x - R/(2\sqrt{6}), x + R/(2\sqrt{6})]$ e $[x - R/(2\sqrt{3}), x + R/(2\sqrt{3})]$? Ovvero quale e' il livello di confidenza con il quale riportiamo un risultato? Vedremo piu' avanti che si puo' calcolare esattamente, e anticipiamo che per il caso triangolare, il livello di confidenza entro la deviazione standard e' del 58%, ovvero la probabilità che il valore vero stia entro quell'intervallo e' del 58%, e del 48% nel caso triangolare.

Cominciamo quindi a vedere come vi possano essere diverse via plausibili da percorrere, spesso supportate da scelte ragionevoli dello sperimentatore.

3.2 Poche misure

Continuando con l'esempio precedente, supponiamo ora di ripetere la pesata con una diversa bilancia analogica che abbia la risoluzione $R = 0.1$ kg. Facciamo 5 pesate e otteniamo i valori $x = 81.9, 82.3, 81.7, 81.9, 82.0$. Quale e' la maniera migliore di riportare la stima del nostro peso? Vediamo anzitutto che avendo cosi poche misure non ha senso raggrupparle in classi di frequenza, e meno ancora quindi valutarne la distribuzione con un istogramma. Tuttavia siamo convinti che la bilancia stia funzionando correttamente. In questo senso allora di queste misure ha senso prenderne

la *media aritmetica*. Per ora diamo buona questa scelta, ma vedremo che si puo' dimostrare che e' la scelta migliore che possiamo fare.

Si definisce media aritmetica di un campione $x = x_1, x_2, \dots, x_N$ di N misure il valore:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (3.4)$$

Nel caso sopra, $\bar{x} = 81.96$ kg. Che incertezza attribuiamo a questo valore? Potremo ragionevolmente prendere la dispersione massima delle misure ovvero $[x_{min}, x_{max}] = 81.7, 82.3$ la cui lunghezza e' $D = 0.6$ kg e dire con un livello di confidenza del 100% che i dati sono compresi tra questi due valori, ma questo non sarebbe una stima della incertezza su \bar{x} . Per assegnare questa incertezza potrebbe usare la semidispersione massima come in Eq. 3.1, questa volta assumendo $R = D = 0.6$ kg e quindi scrivendo $\bar{x} = 81.96 \pm 0.30$ kg o $\bar{x} = 82.0 \pm 0.3$ kg (avendo sfruttato la regola delle cifre significative)¹ Vediamo che questa scelta porta a qualche problema. L'intervallo usando $D/2$ come dispersione diventa $[81.66, 82.26]$, e lascia fuori il valore 82.3 kg. In questo senso non sappiamo neppure chiaramente assegnare un livello di confidenza a questa scelta. Oltretutto, la semidispersione massima ottenuta cosi' soffre di un altro svantaggio: se aumentassimo il numero delle misure, molto probabilmente osserveremo valori ancora piu' distanti dalla media, e quindi la semidispersione massima sara' destinata ad aumentare. Vedremo che una condizione per una buona stima di incertezza e' che diminuisca al numero delle misure, non aumenti. Come possiamo fare? Anche in questo caso ci viene in aiuto l'assumere che le fluttuazioni dei dati seguano una certa distribuzione di probabilita'. Se assumiamo infatti, che questa sia gaussiana, allora possiamo usare come stima della incertezza la *deviazione standard secondo ipotesi gaussiana*, che sara' discussa nella prossima sezione.

Prima di chiudere questa sezione, dobbiamo fare alcune ulteriori ragionamenti. Cominciamo confrontando la risoluzione dello strumento con la distribuzione dei dati. La risoluzione della bilancia era stata data come $R = 0.1$ kg. I dati ora sono distribuiti per fluttuazioni casuali tra 81.7, 82.3, possiamo quindi concludere che *la risoluzione dello strumento ci permette di valutare le fluttuazioni casuali della misura*. Continuiamo notando come nella discussione dell'incertezza su \bar{x} abbiamo parlato solo della dispersione D dei dati e non piu' della risoluzione R . Soffermiamoci su questo un momento.

Esercizio:

Quale e' l'incertezza della singola misura nel caso di misure ripetute?

Abbiamo visto che l'incertezza sulla singola misura si puo' calcolare attraverso la risoluzione strumentale R . Questo valore non e' stimato a livello statistico, ma semplicemente stimato attraverso l'osservazione dello strumento. Tuttavia qui non abbiamo singole misure ma ripetute. Vedremo che attraverso la ripetizioni delle misure e' possibile stimare il 'vero' valore della incertezza sulla singola misura. Per ora questo puo' apparire un concetto fumoso, ma riveste una certa importanza. Si puo' pensare che, al ripetere della misura, di volta in volta abbiamo informazione maggiore con la quale *aggiornare* la nostra stima di questa incertezza. Allo stesso tempo, all'aumentare delle misure aggiorniamo anche la stima della incertezza *della media aritmetica*.

3.3 Misure Ripetute

Supponiamo ora di aumentare il numero delle misure ad un valore superiore a $20 - 30$ ². In questo caso, se la risoluzione dello strumento ce lo permette, dovremmo essere in grado di apprezzare le fluttuazioni casuali della misura, e una volta prese, di raggrupparle in classi di frequenza e costruirne l'istogramma. Usando le regole viste in un capitolo precedente, con $N = 30$ misure, usando la regola della radice

¹In questo caso, dato lo scarso controllo sulla stima di incertezza, la prima scelta con due cifre significative era deplorabile.

²Di nuovo, piu' avanti vedremo esattamente perche' di questi numeri

quadrata, di Sturges e di Rice otteniamo sempre un valore di circa $n_{bins} = 5,6$. La costruzione di un'istogramma è importante perché ci permette di dare una prima valutazione visiva sulla natura gaussiana della distribuzione dei dati, che da un lato ci garantisce di poter usare la corrispondente deviazione standard, dall'altro — almeno in alcuni casi — di escludere la presenza di chiari effetti sistematici. Torneremo su questo punto, e più avanti introdurremo anche test più quantitativi per la valutazione di un istogramma.

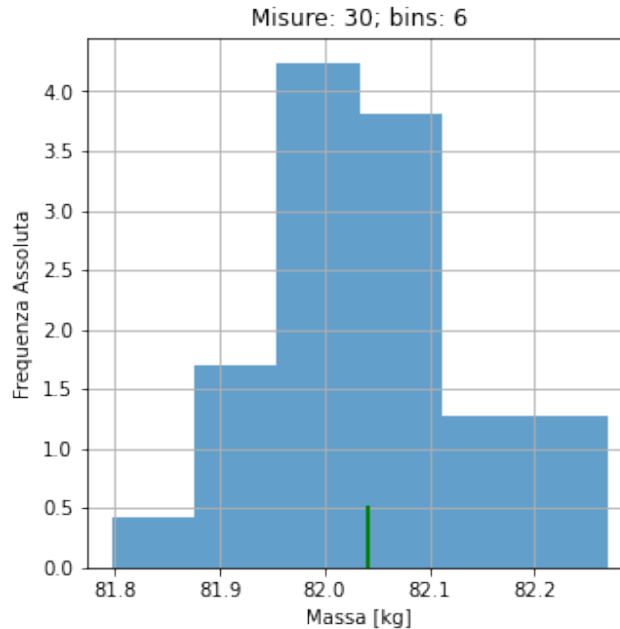


Figura 3.1: Example distribution of 30 randomly generated Gaussian numbers.

Supponiamo di generare 30 misure, con media 82 e dev. standard: 0.1:

```
[82.03870792 82.24619188 82.13126552 82.03909139 81.95630015 81.95440447
81.79729568 82.07202628 81.99524565 82.08471465 82.08881759 81.92855801
82.12694271 81.95460066 81.91967554 81.98715181 81.9969823 81.9786338
82.01965204 82.2105104 81.97681378 82.26936055 82.08201036 82.06814588
81.95803182 82.03167059 82.036482 82.01665959 82.09195434 82.16923924]
```

Usando 6 bin, otteniamo l'istogramma come in Fig. 3.1. Dopo aver valutato l'istogramma, possiamo procedere alla stima del misurando e della sua incertezza causale facendo l'ipotesi della natura gaussiana delle fluttuazioni:

Nel caso di misure ripetute, in cui si sia verificata o comunque assumendo che seguano una distribuzione Gaussiana, allora la miglior stima del valore centrale è dato dalla media aritmetica, mentre la miglior stima dell'incertezza è data dalla *deviazione standard della media*:

$$\hat{x}^* = \bar{x} + s_{\bar{x}} \quad (3.5)$$

dove

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

e

$$s_{\bar{x}} = \frac{s_x}{\sqrt{N}} = \frac{\sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}}{\sqrt{N}} = \sqrt{\frac{1}{N(N-1)} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (3.6)$$

Nell'Eq. 3.6, il termine s_x è detto deviazione standard campionaria secondo la distribuzione gaussiana. Questo termine, di cui parleremo a lungo, è proprio la stima statistica dell'errore della singola

misura, noto con via via maggior precisione al crescere delle misure. Ripetiamo questo fatto importante, indipendentemente dal fatto che la risoluzione dello strumento sia R , se questo valore permette di apprezzare le fluttuazioni statistiche, allora la miglior stime dell'incertezza della singola misura e' s_x .

Esercizio:

A partire dalle definizioni, pensare a come variano in funzione del tempo le stime di media aritmetica, deviazione standard campionaria e deviazione standard della media.

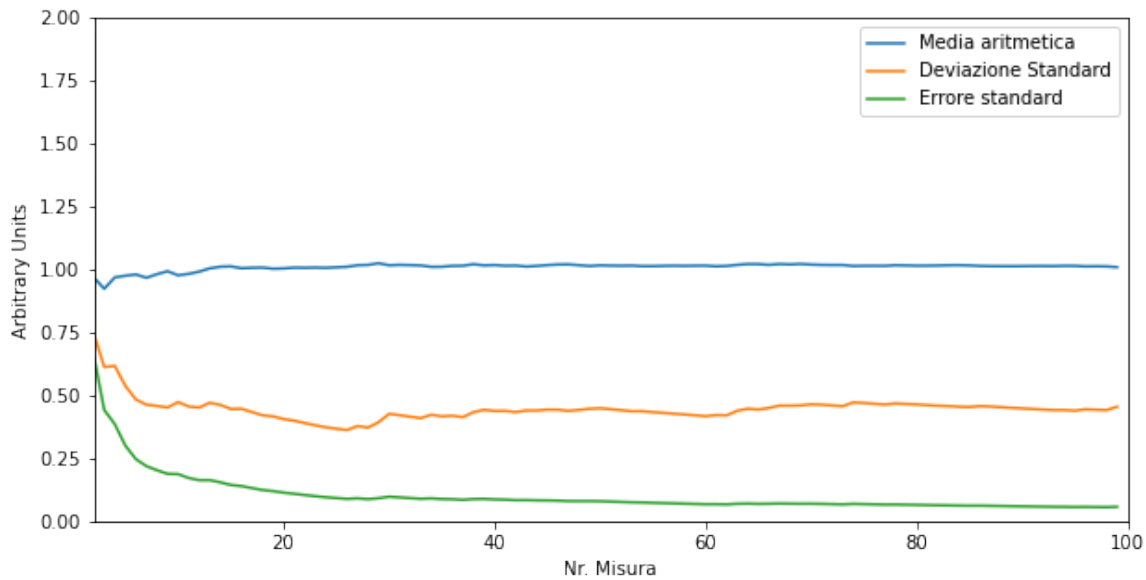


Figura 3.2: Andamento di media aritmetica, deviazione standard campionaria e deviazione standard della media all'aumentare del numero di dati. Il grafico e' ottenuto con estrazione random di dati al computer.

Il concetto di 'aggiornamento della stima all'aumentare della misura' e' molto importante, vediamo rappresentato graficamente. In Fig. 3.2. Come si puo' vedere e come ci si aspetta, la media aritmetica converge all'aumentare delle misure ad un valore stabile con precisione via via maggiore. Forse meno intuitivo, anche la deviazione standard campionaria si stabilizza ad un valore all'aumentare delle misure, che e' proprio la miglior stima dell'errore della singola misura. Invece, a causa della presenza del termine \sqrt{N} al denominatore, la deviazione standard della media diminuisce in maniera quadratica all'aumentare delle misure. In linea di principio, l'incertezza sul valore medio tende a zero per $n \rightarrow \infty$. Si consiglia di meditare su queste tre tendenze.

Esercizio:

Scrivere un programma che estrae in maniera casuale due o piu campioni di dati con un numero di estrazioni via via crescente e ridisegnare i grafici di Fig. 3.2

3.3.1 Terminologia sulle incertezze

Finora abbiamo incontrato le seguente stime di incertezze: deviazione massima (semidispersione massima), deviazione standard campionaria, deviazione standard della media. Essendo il loro uso molto ampio, in letteratura si troveranno diversi nomi e simboli per le stesse quantita'. La qual cosa puo' essere fonte di confusione se non si ha una conoscenza accurata di cosa il simbolo rappresenta. In aggiunta, a volte si pecca nella traduzione da italiano a inglese, lingua oramai unica per la veicolazione scientifica dei risultati. Ad ogni modo, qui di seguito si elencano alcune grandezze e terminologie simili per le stime di incertezza:

- Di solito per i campioni di misure per le deviazioni standard si usa il simbolo s , mentre si riserva il simbolo σ per le deviazioni standard ottenuto da popolazioni di misure o per variabili aleatorie. Spesso tuttavia, si usa il simbolo σ anche per i campioni di misure
- Una grandezza simile alla deviazione standard campionaria e' lo scarto quadratico medio, che in inglese prende il nome di RMSD (Root Mean Square Deviation) o RMSE (Root Mean Square Error) definitio come

$$\text{RMSE/RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (3.7)$$

Lo scarto quadratico medio e' solo diverso (inferiore) alla deviazione standard campionaria, e converge a questa per grandi numeri, o in generale si discosta poco. Essendo tuttavia una stima sistematicamente inaccurata, si deve prediligere la deviazione standard campionaria. Similmente sono definiti il MSD (Mean Square Deviation) o MSE (Mean Square Error) come i quadrati di Eq. 3.7.

- Le deviazioni standard campionarie o delle popolazioni sono chiamate in inglese anche SE (Standard Error), mentre le deviazioni standard della media sono chiamate SEM (Standard Error of the Mean)
- Il calcolo della deviazione standard presuppone una ipotesi sulla natura probabilistica della distribuzione dei dati (triangolare, uniforme, gaussiana). Un valore che misura la ampiezza di una distribuzione senza ipotesi statistica, che ha senso in alcune applicazioni e' il FWHM (Full Width Half Maximum), ovvero la ampiezza di una distribuzione calcolata a meta' altezza.

3.3.2 Altre stime di media

Abbiamo accennato al fatto che la media aritmetica e' la miglior stima di valore centrale di una distribuzione di dati. Tuttavia, non abbiamo ancora chiarito cosa intendiamo per migliore, lo faremo solo aver introdotto i valori di aspettazione di variabili aleatorie. Tuttavia, spesso si ha a che fare con distribuzioni non necessariamente gaussiane, o non simmetriche, o di varia natura. Alcune di questa hanno caratteristiche di 'centralita' (massimo al centro e code esterne), altre no. A seconda della applicazione, possono essere utili alcune diverse stime centrali. Ne impariamo alcune. Si possono dividere in posizionali e algebriche. Le prime sono calcolate guardando la 'posizione' del dato nella lista delle misure, le seconde invece combinando algebricamente tutte le misure. La media aritmetica chiaramente rientra in questa seconda categoria. Le medie posizionali sono la moda e la mediana. Le media algebriche sono, oltre alla media aritmetica troviamo la media pesata, la media quadratica, la media geometrica, la media armonica. E' utile conoscerle, in alcune occasioni possono tornare utili.

Cominciamo con le medie posizionali.

Moda (Mode) Si tratta semplicemente del valore $\overline{x_{mode}}$ che si presenta piu' spesso in una serie di misure. Se le misure sono abbastanza precise, e' difficile che un valore si presenti piu' volte, allora si puo' parlare di moda per la classe di frequenza con maggiori conteggi. E' possibile che una distribuzione di dati sia bimodale o multimodale, se presenta rispettivamente due o piu' valori (o classi di frequenza) piu' frequenti. Invece se il valore piu' frequente e' agli estremi di una distribuzione (ad es. una distribuzione esponenziale), allora non si parla di moda.

Mediana (Median) La mediana $\overline{x_{med}}$ e' il valore che separa in due il campione di misure, nel senso che meta' misure sono maggiori della mediana e meta' sono minori. Per calcolarla una maniera e' quella di ordinare le misure in ordine crescente e prendere quella centrale (se le misure sono dispari) o la media delle due centrali, se le misure sono pari. Esistono anche altri metodi per stimare la mediana ma li vedremo successivamente.

E' intuitivo che se una distribuzione di dati e' simmetrica e sufficientemente numerosa, moda, mediana e media aritmetica devono coincidere. In effetti, calcolare moda e mediana e confrontarle con la media aritmetica potrebbe essere usato come stima di 'asimmetria' di una distribuzione. La moda esprime semplicemente il valore che appare piu' volte, ed e' usata in questo senso. La mediana torna utile invece nel caso di distribuzioni asimmetriche o nel caso in cui sia presente qualche dato lontano dal centro,

potenzialmente anomalo. In questo caso, la presenza di un dato 'lontano' ha un effetto maggiore sulla media aritmetica piuttosto che sulla mediana. Un dato qualsiasi ha il 50% di probabilita di essere maggiore o minore della mediana. Si puo' anche dimostrare che la mediana e' quel valore che minimizza la somma dei moduli degli scarti:

$$\sum_{i=1}^N |x_i - x_{mediana}| \leq \sum_{i=1}^N |x_i - x^*|$$

Esercizio:

Calcolare moda e mediana di questa serie di dati:

$$x = [12, 17, 13, 11, 12, 16, 13, 12, 10, 12, 15, 12, 14, 12, 14, 12, 13, 11, 13, 13, 14, 15, 11]$$

Continuiamo con le medie algebriche.

Media Geometrica (Geometric Mean) La media geometrica e' definita come la radice N -esima del prodotto delle N misure (non negative)

$$\overline{x}_G = \sqrt[N]{\prod_{i=1}^N x_i} = \sqrt[N]{x_1 \cdot x_2 \cdot \dots \cdot x_N}$$

Si puo' dimostrare che la media geometrica e' sempre inferiore alla media aritmetica. Inoltre, e' utile sapere che essa corrisponde all'esponenziale della media aritmetica dei logaritmi naturali della serie di misure:

$$\overline{x}_G = \exp\left(\frac{1}{N} \sum_{i=1}^N \ln x_i\right)$$

Cio spiega come mai la media geometrica sia il punto medio di una decade in scala logaritmica.

Media Armonica (Harmonic Mean) La media armonica di N valori e' definita come il reciproco della media aritmetica dei reciproci

$$\overline{x}_A = \frac{N}{\sum_{i=1}^N 1/x_i}$$

Essendo basata sui reciproci, pesa di piu i valori «vicini» alla media e meno quelli lontani, i cosiddetti 'outliers'. Inoltre, in alcuni casi ha applicazioni dirette, ad esempio dati due resistori R_1, R_2 posti in parallelo, la resistenza equivalente e' data dalla loro media armonica $R_{eq} = 2/(1/R_1 + 1/R_2)$

Media quadratica (Mean Square) E' la Radice della media dei quadrati:

$$\overline{x}_Q = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$$

Ad esempio, la deviazione standard e' calcolata attraverso la media quadratica degli scarti rispetto alla media aritmetica. Nella fisica delle molecole di gas, esiste una relazione tra costante dei gas, temperatura e massa delle particelle che da la velocità media quadratica media.

Media Pesata (Weighted Mean) La media pesata e' una media aritmetica calcolata dando un peso diverso ad ogni misura. E' chiaro che la somma dei pesi deve essere pari a 1:

$$\overline{x}_Q = \sum_{i=1}^N p_i x_i; \quad \sum_{i=1}^N p_i = 1$$

I pesi p_i possono essere calcolati in maniera diversa. Piu' avanti nel corso consideremo il caso della media pesata con gli scarti quadratici rispetto alla media aritmetica. Si noti anche che nel caso i pesi $p_i = p$ siano tutti uguali, la media pesata corrisponde alla media aritmetica.

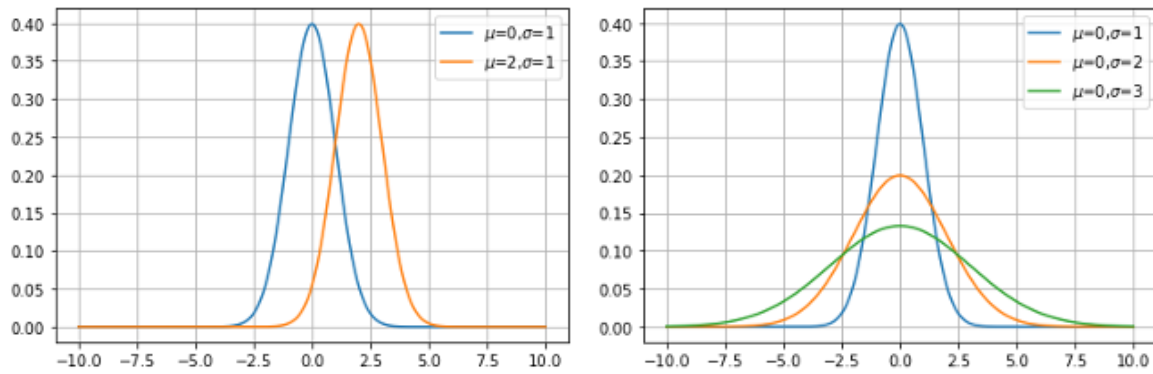


Figura 3.3: (Sinistra) Due curve Gaussianhe con stesso parametro σ ma media diversa. (Destra) Tre curve gaussianhe con stessa media e diverso parametro σ .

3.4 La curva Gaussiana

Proviamo ora ad immaginare come apparirebbe l'istogramma in Fig. 3.1 se avessimo $N \rightarrow \infty$. Anzi-tutto avremmo in linea di principio la possibilità di avere un numero infinito di classi di frequenza. Al limite quindi le altezze dei 'rettangoli' dell'istogramma formerebbero una serie di punti che assieme darebbero l'aspetto di una funzione continua, come mostrato in Fig. 3.3.

La funzione descritta da questi punti avrebbe la formula di una opportuna funzione di Gauss. Si dice funzione di Gauss una funzione del tipo:

$$f(x) = e^{-x^2} \quad \text{per cui:} \quad \int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$$

A partire da questa generica funzione, la funzione di distribuzione di probabilità Gaussiana, che è una funzione di Gauss a) normalizzata a 1, b) centrata in μ , e con ampiezza σ :

Si definisce funzione di densità di probabilità Gaussiana la funzione:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Esercizio:

Si dimostrino le seguenti proprietà della funzione Gaussiana: a) l'integrale indefinito è uguale a 1; b) il punto di massimo è μ e il massimo vale $1/\sqrt{2\pi}$; c) la posizione dei flessi a distanza $\pm\sigma$ da μ . Si veda la parte finale del capitolo per la soluzione.

Dimostrazione. Partiamo osservando che:

$$\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = \int_{-\infty}^{+\infty} e^{-\left(\frac{x}{\sqrt{2}}\right)^2} d\left(\frac{z}{\sqrt{2}}\right) \sqrt{2} = \sqrt{2} \int_{-\infty}^{+\infty} e^{-z^2} dz = \sqrt{2\pi}$$

Calcoliamo ora l'integrale della Gaussiana:

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

eseguendo il cambio di variabile $z = (x - \mu)/\sigma, dz = dx/\sigma$, otteniamo

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} \sigma e^{-\frac{z^2}{2}} dz = \frac{\sqrt{2\pi}}{\sqrt{2\pi}} = 1$$

Per calcolare il massimo eseguiamo la derivata prima rispetto a x :

$$f'(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \left[-\frac{2(x-\mu)}{2\sigma^2} \right]$$

La derivata prima è nulla quindi per $x = \mu$. La derivata seconda è

$$f''(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \left[\left[-\frac{2(x-\mu)}{2\sigma^2} \right]^2 - \frac{1}{\sigma^2} \right]$$

e $f''(\mu) < 0$ pertanto μ è punto di massimo. Il massimo si calcola facilmente $f(\mu) = 1/\sqrt{2\pi}$. La derivata seconda si annulla per $x_1 = \mu + \sigma$ e $x_2 = \mu - \sigma$.

La curva Gaussiana, per quanto sorprendente possa sembrare, è la funzione quindi che approssima la *distribuzione delle fluttuazioni casuali di ciascuna misura*. L'affascinante ragione che sta dietro a questo aspetto è legata a come si ottiene la Gaussiana, ovvero come limite punto a punto del numero di combinazioni possibili tra fluttuazioni positive e negative. Questo aspetto è lasciato agli approfondimenti a fine capitolo. La ragione è quindi puramente di calcolo combinatorio, la funzione Gaussiana descrive questo andamento per un numero infinito di effetti di fluttuazione.

È chiaro che dovremo fare un confronto quindi tra i dati del campione posti in istogramma e la curva Gaussiana che rappresenti il loro andamento nel caso il campione avesse numerosità $N \rightarrow \infty$. Anzitutto, come vedremo più avanti, avendo un campione di dati non conosciamo μ, σ , ma possiamo dal campione stesso stimarli, come visto sopra attraverso la media aritmetica e la SE.

$$\hat{\mu} = \bar{x} \quad \hat{\sigma} = s_x$$

Inoltre, dobbiamo ricordare che l'area sottesa dall'istogramma delle frequenze assolute (conteggi) e $N bw$, di conseguenza:

la distribuzione Gaussiana che meglio approssima un istogramma di frequenze assolute ottenute da un campione di misure è:

$$f(x_i) = \frac{N bw}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)^2}$$

dove N è la numerosità del campione, bw l'ampiezza delle classi di frequenza.

A cosa serve la sovrapposizione di questa funzione? Lo vedremo in dettaglio nella trattazione delle variabili aleatorie, ma anticipiamo che — dal nome stesso — con le funzioni di distribuzione di probabilità possiamo calcolare in maniera quantitativa la probabilità che un valore appartenga ad un certo intervallo attorno alla media aritmetica, o lontano da essa. Attraverso questa conoscenza si possono fare diverse cose, tra cui discuteremo qui di seguito la definizione del livello di confidenza e lo studio degli eventi anomali.

Prima di continuare semplifichiamo la notazione per la Gaussiana, che useremo spesso;

Una notazione compatta di uso comune per la funzione di distribuzione di probabilità Gaussiana e':

$$\mathcal{G}(x; \mu, \sigma^2) = \frac{N}{\sqrt{2\pi\sigma_x}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma_x}\right)^2} \quad (3.8)$$

In alcuni testi si usa invece $\mathcal{G}(x; \mu, \sigma)$ al posto di $\mathcal{G}(x; \mu, \sigma^2)$. La funzione $\mathcal{G}(x; 0, 1)$ è detta Gaussiana standard.

Livello di confidenza Dalla serie discreta di numerosità N , facendo l'ipotesi che questa segua una distribuzione Gaussiana, possiamo stimare i parametri della funzione di distribuzione di probabilità Gaussiana $\mu = \bar{x}, \sigma = s_x$. Abbiamo dimostrato che l'integrale, ovvero l'area sotto la curva, tra $-\infty, \infty$ e' 1. Questo corrisponde numericamente al fatto che ha probabilità 1 (certezza) il fatto che la misura cada tra $-\infty, \infty$. Quale e' la probabilità che la misura cada tra $\mu - \sigma, \mu + \sigma$? Sfruttando un integrale già visto sopra e' facile vedere (dimostratelo) che:

$$\int_{\mu-\sigma}^{\mu+\sigma} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} = 68.3\%$$

Possiamo quindi affermare con —attenzione— facendo l'ipotesi che la distribuzione di dati misurati segua la distribuzione Gaussiana, allora il 68.3% delle volte la misura effettuata cada entro l'intervallo $\bar{x} - s_x, \bar{x} + s_x$, la nostra miglior stima di $\mu - \sigma, \mu + \sigma$. A partire da questo, costruiamo quindi il livello di confidenza *sulla stima del valore vero* del 68.3% con la deviazione standard della media s_x/\sqrt{N} . Nel prossimo riquadro si mostrano altri intervalli utili per la costruzione di livelli di confidenza.

Intervalli Utili della Gaussiana: Si riportano qui diversi utili intervalli centrali:

$$\int_{\mu-\sigma}^{\mu+\sigma} \mathcal{G}(x; \mu, \sigma^2) = 68.3\%; \quad \int_{\mu-2\sigma}^{\mu+2\sigma} \mathcal{G}(x; \mu, \sigma^2) = 95.5\%; \quad \int_{\mu-3\sigma}^{\mu+3\sigma} \mathcal{G}(x; \mu, \sigma^2) = 99.7\%$$

Sono importanti anche i seguenti intervalli esterni:

$$\int_{\mu+1\sigma}^{+\infty} \mathcal{G}(x; \mu, \sigma^2) = 15.9\%; \quad \int_{\mu+2\sigma}^{+\infty} \mathcal{G}(x; \mu, \sigma^2) = 2.3\%; \quad \int_{\mu+3\sigma}^{+\infty} \mathcal{G}(x; \mu, \sigma^2) = 0.3\%;$$

. E' altresì molto utile calcolare il complementare di queste probabilità. Riassumiamo quindi tutto in tabella assieme ad altri utili intervalli:

Intervallo I	$P(I)$	1-P(I)
$[\mu - \sigma, \mu + \sigma]$	68.3%	31.7%
$[\mu - 2\sigma, \mu + 2\sigma]$	95.5%	4.5%
$[\mu - 3\sigma, \mu + 3\sigma]$	99.7%	0.3%
$[-\infty, \mu + 1\sigma]$		15.9%
$[-\infty, \mu + 2\sigma]$		2.3%
$[-\infty, \mu + 3\sigma]$		0.14%
$[-\infty, \mu + 1.96\sigma]$	95%	
$[-\infty, \mu + 2.71\sigma]$	99%	

Esclusione di eventi anomali Durante la presa della misure, è possibile che avvengano degli errori (questa volta il nome e' corretto) che portano a errori di scrittura o trascrittura del valore misurato: urti con l'apparato, glitch elettronici, errori di sbaglio umano, etc. In questo caso é lecito escludere queste misure erronee. Tuttavia, in alcuni casi, l'effetto apportato non e' cosi intenso da permetterci di escludere il dato facilmente. Il dato anomalo si discosta dagli altri ma magari non eccessivamente. Sappiamo d'altro canto ci aspettiamo discostamenti dal valore centrale a causa della fluttuazioni casuali della misura, e sappiamo dalla loro distribuzione teorica Gaussiana che questi spostamenti possono essere anche grandi (la Gaussiana e' definita su tutto l'asse reale) ancorche' via via piu' rari.

Attraverso la Gaussiana possiamo valutare la plausibilità di un evento raro, facciamo un esempio.

Esempio:

Ripeto la misura del periodo del pendolo 80 volte, ottenendo come stima del valore vero $T = \bar{x} \pm s_{\bar{x}} = 2.1 \pm 0.1$ s. Uno dei dati e' 2.45 s. In che misura lo posso considerare un evento anomalo?

Svolgimento: Per risolvere il quesito si avanza cosi. Dalla stima riportata ricaviamo che la miglior stima dei parametri della Gaussiana sia $\hat{\mu} = \bar{x} = 2.1$ s e $\hat{\sigma} = s_{\bar{x}}\sqrt{N}$ ricavando cioe' SE dal SEM. Ora, con 80 misure, quante ce ne aspettiamo oltre i 3σ (con $T > 2.4$) sulla coda di destra? Osservando la tabella sopra, lo 0.14%, che in rapporto alle misure prese da $n_{out} = 0.14\%(N) = 0.014N = 1.12$, ovvero poco piu' di un evento. Ne abbiamo trovato 1 a 2.45 s, neppure troppo distante da $\bar{x} + 3s_{\bar{x}}$. Concludiamo che il dato molto probabilmente non e' anomalo.

Nota: il dato anomalo deve essere vicino all'intervallo per cui si calcola la probabilità. Se il dato anomalo fosse stato ad esempio $T = 2.9$, nonostante me ne aspetassi solo 1 sopra 3 SE, esso ne era molto distante. In questo caso avrei potuto ad esempio calcolare la probabilita' che il dato fosse a piu' di 8 SE nella coda superiore. Avremmo trovato un valore decisamente inferiore ad 1. Si provi a fare il conto.

In conclusione, per la ricerca di dati anomali si procede cosi. Anzitutto si definisce l'intervallo che si ritiene raro (nel nostro caso, e di norma, si considerano 3 SE, ma in altri casi potrebbero essere 2, o talvolta 5). Si stima il numero di misure che, nell'ipotesi che i dati seguano la Gaussiana identificata da \bar{x}, s_x ci si aspetta in una certa coda. Si confronta poi il numero e la posizione di questi dati anomali con questa stima.

Attenzione: dietro ad un dato anomalo potrebbe celarsi una scoperta. A meno che non siate certi di aver commesso un ovvio errore, che sapete spiegare senz'altro, dietro ad un dato anomalo che non sapete spiegare potrebbe celarsi o la scoperta di un problema strumentale serio, oppure la scoperta di qualcosa di nuovo, il sogno di ogni fisico sperimentale! Non buttate via misure alla leggera!

3.5 Commenti Finali

Ricapitolando

Quesiti di autovalutazione

□ **Es. 3.1** La popolazione delle prime 10 città americane in milioni è la seguente. Calcolare la popolazione media (aritmetica) e la popolazione mediana. Discutere la differenza tra i due indici

New York (New York)	9.21	Philadelphia	1.45
Los Angeles (California)	4.05	Dallas (Texas)	1.31
Chicago (Illinois)	2.83	San Diego (California)	1.30
Houston (Texas)	2.01	San Antonio (Texas)	1.24
Phoenix (Arizona)	1.55	San Josè (California)	0.94

□ **Es. 3.2** Data una distribuzione di 120 misure descritta da una distribuzione di probabilità gaussiana di media $\mu = 0$ e deviazione standard $\sigma = 1$, quanti dati ci aspettiamo secondo la statistica gaussiana con valore $x > 4$? e $|x| > 4$?

□ **Es. 3.3** Si supponga di avere due produzioni di tubi di ferro, A e B, ciascuno formato da 100 tubi. Si ipotizza che i tubi siano realizzati con alta precisione. Per verificarlo, si decide di fare un confronto a campione tra le due produzioni. Della prima si estraggono alcuni e si misurano le lunghezze, ottenendo la distribuzione in tabella. Per la seconda serie, si ripetono altrettante misure e si ottiene come stima $L_B = 5.014 \pm 0.004$.

- Stimare media e deviazione standard della prima serie di misure;
- Riportare nella forma corretta la stima della media della prima misura e della incertezza;
- Scrivere la formula analitica della distribuzione gaussiana che si puo' associare alla prima serie di misure;
- Verificare se ci sono dati misurati che si trovano al di la di 3 deviazioni standard dalla media.

L_A [m]	Frequenza assoluta
5,10	2
5,11	9
5,12	14
5,13	8
5,14	3

Approfondimenti

Aprr. 3.1 Proprieta della media aritmetica. La media aritmetica rappresenta la miglior stima del misurando in un campione di misure ripetute per le seguenti proprietà:

1. La somma algebrica (con segno) degli scarti rispetto alla media aritmetica e' NULLA
2. La somma dei quadrati degli scarti rispetto alla media è MINIMA
3. E' la stima del valore vero affetta dalla minima incertezza casuale
4. L'errore statistico della media converge a zero al ripetere delle misure sempre piu volte

Aprr. 3.2 La Gaussiana come combinazioni di fluttuazioni. La curva Gaussiana e' una curva creata dalla Natura! E' data dalle combinazioni con cui diversi errori casuali si possono combinare dando origine alla misura. L'idea di base è che un effetto casuale puo' aumentare o diminuire una grandezza. Ad esempio una molla in una bilancia a molla puo' a volte comprimersi leggermente in meno o in piu. In una misura le incertezze casuali si sommano in forma casuale.

Supponiamo di avere solo 2 effetti casuali, che possono ciascuno aumentare o diminuire la misura di una certa' quantita

- Effetto 1: aumenta o diminuisce di ± 0.5 ;
- Effetto 2: aumenta o diminuisce di ± 0.5 ;

La probabilità che l'effetto 1 o 2 prendano uno dei valori e' la stessa $P(-0.5) = P(0.5) = 0.5$ (vedremo che possiamo chiamare questo evento, evento aleatorio di Bernoulli. Che valori per la misura posso ottenere? $-1, 0, +1$. Tuttavia, non ottengo questi valori lo stesso numero di volte: ottengo 1 volta -1 , 2 volte 0 , 1 volta $+1$. Si noti la sequenza $(1, 2, 1)$. Se ora aumentiamo a 3 effetti casuali indipendenti della stessa intensità otteniamo una sequenza $(1, 3, 3, 1)$, con 4 effetti $(1, 4, 6, 4, 1)$. Se li mettiamo in righe successive:

```

      1
     1 1
    1 2 1
   1 3 3 1
  1 4 6 4 1
 . . . . .

```

Otteniamo il ben noto Triangolo di Pascal (detto Triangolo di Tartaglia in Italia). Ciascun elemento del triangolo di Pascal corrisponde al numero di combinazioni che danno effetti positivi rispetto alla riga n -esima e la posizione k -esima. Ad esempio, se si parte da $n = 0, k = 0$, e' facile verificare che $4 = \binom{n}{k} = \frac{n!}{k!(n-k)!}$ con $n = 4, k = 1$. Se ora aumentiamo il numero di righe, ovvero il numero di effetti casuali, vediamo che i coefficienti combinatori presi nella riga ci daranno via via piu' similmente la forma di una Gaussiana! Ecco quindi spiegato la ragione matematica per cui emerge questa curva: e' legata al numero di possibili combinazioni di effetti casuali che aumentano e diminuiscono la misura. Se questi effetti sono numerosi e indipendente, allora e' *naturale* che le fluttuazioni complessive casuali siano distribuite secondo la Gaussiana.

Appr. 3.3 Teorema del limite centrale.

Il teorema del limite centrale sara' affrontato piu' nello specifico nel [Capitolo 5](#), ma svolge un ruolo fondamentale nella costruzione di un teoria statistica perche' chiarisce molto la natura probabilistica della trattazione. Fu introdotto gia' nel 1718 nel 'Dottrina delle chances' di A. De Moivre, un libro per scommettitori, assieme ad una prima presentazione della distribuzione Gaussiana, che del teorema del limite centrale e' una conseguenza. In sostanza il teorema dice che quando si *sommino* gli effetti di ripetute variabili aleatorie (indipendenti, e distribuite alla stessa maniera), il risultato e' una variabile aleatoria distribuita pressoché Gaussianamente:

$$X_1 + X_2 + \dots X_n \rightarrow \mathcal{G}$$

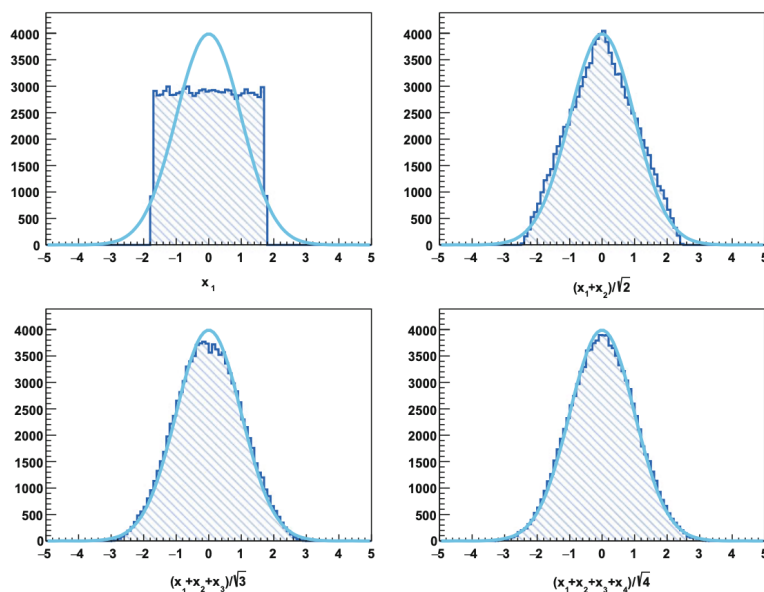


Figura 3.4: Teorema del limite centrale. Al sommare di piu' variabili aleatorie indipendenti e similmente distribuite si ottiene una distribuzione via via piu' simile ad una Gaussiana. (Riprodotta da [Lista, 2016](#))

Ovvero, la distribuzione Gaussiana emerge dalla *combinazione* di variabili aleatorie purché numerose, indipendenti, e similmente distribuite. Il teorema si può leggere in questa maniera per misure sperimentali "l'effetto combinato di fluttuazioni della misura, purché indipendenti tra loro e di entità simile, porta complessivamente ad una distribuzione Gaussiana delle misure, siano esse originalmente Gaussiane o no". Si prendano ad esempio (esempio preso da L. Lista) distribuzioni uniformi X_1 tra $[-\sqrt{3}, \sqrt{3}]$. Nel grafico di Fig. 3.4 si mostrano rispettivamente $X_1, (X_1 + X_2)/\sqrt{2}, (X_1 + X_2 + X_3)/\sqrt{3}, (X_1 + X_2 + X_3 + X_4)/\sqrt{4}$ (dove i denominatori hanno solo una funzione di riscalare l'altezza della distribuzione senza alterarne la forma). Si vede chiaramente come 'pescando in maniera casuale' tra le X_i già dopo qualche interazione la forma complessiva e' quella di una Gaussiana.

Appr. 3.4 Persone.

Johann Friedrich Carl Gauss (1777 –1855) E' stato un matematico, astronomo e fisico tedesco, che ha dato contributi determinanti in analisi matematica, teoria dei numeri, statistica, calcolo numerico, geometria differenziale, geodesia, geofisica, magnetismo, elettrostatica, astronomia e ottica. Nacque a Braunschweig nel ducato di Brunswick-Lüneburg (ora parte della Bassa Sassonia, in Germania), figlio unico di una famiglia di bassa estrazione sociale e culturale. Gauss era un bambino prodigio. Un aneddoto racconta che a 9 anni il suo insegnante, per mettere a tacere i turbolenti allievi ordinò loro di fare la somma dei numeri da 1 a 100. Quasi subito il bimbo Gauss diede la risposta esatta, sorprendendo l'insegnante ed il suo assistente. Non si è certi di quale metodo abbia adottato Gauss; forse mise in una riga i numeri da 1 a 100 e in una riga sotto i numeri da 100 a 1, e vide che ogni colonna dava come somma 101: Carl moltiplicò 100×101 e divise per due, ottenendo il risultato; oppure - ancora più semplicemente - scrisse in fila i numeri da 1 a 50 e in una fila sotto in senso inverso i rimanenti da 51 a 100, ottenendo così per ogni coppia la somma costante di 101: il risultato era quindi 101×50 .



Carl Gauss

Capitolo 4

Teoria della Probabilità

Quest'opera è distribuita con licenza [Creative Commons](#) "Attribuzione – Non commerciale – Condividi allo stesso modo 3.0 Italia".



La Teoria della Probabilità è un costrutto matematico teorico che permette di calcolare le probabilità di eventi detti aleatori, come l'esito del lancio di una moneta, del calcolo di una sezione d'urto per l'interazione tra due particelle elementari, della formazione di un ciclone estivo in un tale luogo. È interessante notare che nonostante il concetto di probabilità sia piuttosto intuitivo e radicato nel ragionamento comune, il *computo* della probabilità per i tre eventi descritti è radicalmente diverso! Nel primo caso, nel lancio della moneta, gli eventi possibili sono solo 2, esce testa o esce croce. È facile calcolare la probabilità che esca testa o croce in uno o più lanci. Ci sembra intuitivo di poterlo fare con un elevato livello di confidenza. Anche se la moneta fosse truccata, possiamo immaginare esperimenti ad esempio ripetendo molte volte il lancio che ci permettano di affermare con un certo livello di confidenza se la moneta è truccata oppure no. Nel secondo caso, al di là della interazione specifica, ci si rende conto che ci saranno più fattori da tenere in conto, l'energia delle singole particelle, la configurazione nell'urto, etc. Il tutto è più complesso rispetto al caso della moneta, ma ci sembra che con sufficiente conoscenza si possa arrivare (si fa) a fare predizioni e stime, e che migliorando il setup sperimentali si possano ottenere risultati con via via maggior livello di confidenza. Nel caso del ciclone, ci rendiamo conto invece che è 'ancora più aleatorio' nel senso che non riusciamo a controllare tutte le variabili, e le predizioni saranno sempre fatte con un minore livello di confidenza. Questi esempi ovviamente non esauriscono le infinite casistiche possibili, ma vogliono solo illustrare la complessità del fatto di usare una unica 'teoria della probabilità' per tutti.

Una differenza importante tra eventi di cui si vuole calcolare la probabilità è se lo spettro di esiti è discreto (numero finito di esiti) o continuo. Ad esempio, il risultato del lancio di una moneta è discreto (testa o croce), mentre il risultato di una misura di una grandezza ha di norma spettro continuo (la misura è in sé un numero reale). Il primo caso è matematicamente più facile da analizzare del secondo. Ce ne occuperemo in questo capitolo. Vedremo come generalizzare questo concetto per eventi a spettro continuo. Tutte le regole della Teoria della Probabilità entreranno poi come regole per il ragionamento sulle distribuzioni di probabilità di misure.

4.1 Evento Aleatorio

Cominciamo con la definizione di evento aleatorio, anche detto casuale o stocastico o *random*.

Un **evento aleatorio** è un evento che può avere diversi esiti e di cui è impossibile sapere quale tra gli esiti avviene, ma per cui è possibile in linea di principio calcolare la probabilità di un esito rispetto ad un altro.

L'esito dell'evento aleatorio e' detto *evento elementare* e_i , e l'insieme degli eventi elementari e' detto *Insieme Universo* o *spazio degli eventi elementari* Ω . Gli eventi e_i di Ω sono mutuamente esclusivi, nel senso che o avviene un esito oppure un altro. L'insieme universo puo' contenere un numero finito o infinito di elementi.

Esempio:

Ad esempio l'insieme universo per l'evento aleatorio lancio di un dado singolo a 6 facce e' l'insieme $\Omega = \{1, 2, 3, 4, 5, 6\}$, nel caso della moneta $\Omega = \{T, C\}$ o del lancio doppio di monete $\Omega = \{TT, CT, TC, CC\}$. Nel caso del calcolo della media aritmetica di una serie di misure e' in linea di principio $\Omega = \mathbb{R}$.

A partire dall'insieme Universo si puo' costruire lo *Spazio degli Eventi*, che possiamo indicare con $\mathcal{F}(\Omega)$. Questo e' l'insieme contenente tutti i possibili eventi aleatori E_i , ottenuti a partire dagli eventi elementari dell'evento aleatorio in questione. Puo' essere pensato (non troppo rigorosamente) come l'insieme di tutte le domande che ci possiamo fare attorno ad un evento aleatorio. Ad esempio, nel lancio singolo di un dado a 6 facce, mentre $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F}(\Omega)$ contiene eventi tipo 'esce un numero pari' o 'esce un numero minore di 3' etc. E' chiaro che l'evento elementare rimane uno e uno solo, esce una delle 6 facce, ma sono anche aleatori gli eventi complessi che abbiamo scritto. Allora nel lancio del dado $\mathcal{F}(\Omega) = \{\{1\}, \{2\}, \dots, \{1 \text{ oppure } 2\}, \dots, \{1 \text{ oppure } 2 \text{ oppure } 3\} \dots\}$ ovvero l'insieme di tutti i sottoinsiemi che si possono formare a partire dagli esiti elementari. Lo spazio degli eventi e' matematicamente detto anche *l'insieme delle parti* nel caso Ω sia finito, mentre richiede formulazioni piu' complesse come quelle delle σ -algebre nel caso continuo. $\mathcal{F}(\Omega)$ contiene sempre l'insieme vuoto \emptyset (ovvero l'evento aleatorio non avviene) e l'insieme universo, ovvero, esce uno qualsiasi degli esiti elementari. Gli eventi E_i di $\mathcal{F}(\Omega)$ non sono piu' ora necessariamente mutuamente esclusivi, lo sono solo gli esiti elementari e_i . o

Esempio:

Ad esempio lo spazio degli eventi $\mathcal{F}(\Omega)$ per l'evento aleatorio lancio singolo di una moneta e' $\Omega = \{\emptyset, \{T\}, \{C\}, \Omega\}$. Nel caso del lancio doppio di monete, indicando con ad es TT l'evento aleatorio che esca testa nel primo e nel secondo lancio, allora $\mathcal{F}(\Omega) = \{\emptyset, \{TT\}, \{CT\}, \{TC\}, \{CC\}, \{TT \text{ o } CT\}, \{TT \text{ o } TC\}, \{CC \text{ o } CT\}, \{CC \text{ o } TC\}, \{CC \text{ o } CT \text{ o } TC\}, \{TT \text{ o } CT \text{ o } TC\}, \Omega\}$

Costruire una teoria della probabilità significa quindi decidere come assegnare a ciascun degli elementi dello spazio degli eventi un numero, che chiamiamo probabilità:

$$E_i \longrightarrow P(E_i)$$

Ma come farlo? Nel caso discreto del lancio del dado o della moneta visti sopra, ci sembra fattibile, basta 'contare le possibilità', ma nel caso continuo? Ad esempio, per una misura possiamo chiederci la probabilità che sia entro un intervallo, o maggiore di un certo valore, etc. Come scriviamo lo spazio degli eventi? E come calcoliamo le probabilità? La moderna teoria di probabilità risponde e questa domanda.

4.2 Definizioni di probabilità

Il concetto di probabilità e' evoluto durante la storia umana, e la sua evoluzione e' decisamente affascinante. Un brevissimo resoconto di questa storia e' riportato in calce a questo capitolo. Mentre fin dall'antichità si e' sempre usato il concetto di probabilità, il concetto di *misura* di probabilità e' decisamente piu' moderno. E' stato sviluppato a partire dal XVII secolo a dirla tutta per calcolare bene la posta in gioco in qualche scommessa da salotti rinascimentale...! Dopodiche, quando si e' cominciata ad applicare in maniera piu' sistematiche alle scienze, si e' incorsi in problemi complessi (come appunto quello di evento aleatorio continuo) e da quel momento in poi si è evoluta attraverso

un percorso accidentato¹ fino ad una versione formale completa ottenuta solo negli anni '30 del XX secolo, ben dopo ad esempio la stesura della teoria dell'elettromagnetismo o della relatività speciale. La ragione sta nel fatto che nessuna definizione di probabilità sembra soddisfare tutti i campi in cui si può applicare. Per questo illustreremo *diverse* definizioni di probabilità che storicamente sono state proposte ma anche praticamente ci permettono di coglierne le sfumature.

Definizione Classica di probabilità La definizione classica di probabilità si fa risalire a Cardano, nel XVI secolo. Possiamo scrivere questa definizione come:

Definizione Classica di Probabilità:

$$P(E_i) = \frac{\text{numero di casi favorevoli per } E_i}{\text{numero di casi possibili}}$$

La definizione classica di probabilità, si costruisce in sostanza a partire dalla elencazione dei casi possibili e al conto dei casi favorevoli. È decisamente intuitiva, si pensi al numero di casi favorevoli in cui esca un numero pari in un dado a 6 facce. È tuttavia applicabile solo ad una ristretta classe di eventi aleatori: gli esiti elementari devono essere equiprobabili, indipendenti ed esaustivi delle possibilità. Addirittura questa definizione soffre di una evidente tautologia: la definizione di probabilità richiede che gli eventi elementari siano *equiprobabili*. Questa definizione non può andare bene quindi in casi come il calcolo della probabilità della formazione di un ciclone, o della misura di una grandezza fisica.

Formulazione Frequentista di probabilità Questa seconda definizione affronta un argomento importante, su cui torneremo in dettaglio, che è quello della *ripetizione* di un evento aleatorio: un modo di dire che 'esce testa' ha una probabilità del 50% di avvenire, a pensarci, è che la testa esce il 50% *delle volte*. Questo concetto è alla base della Legge dei Grandi Numeri discussa in Approfondimento 4.2. Questo si basa sul fatto che implicitamente si ipotizza di poter ripetere l'evento aleatorio tante volte. Se è così, allora si può immaginare di registrare il numero totale di ripetizioni N , e il numero totale di successi n_{E_i} in cui l'evento E_i è avvenuto, il loro rapporto dà la frequenza dei successi $f(E) = n_E/N$. Secondo la definizione frequentista, che si deve a Richard von Mises (1883-1953), la probabilità è dunque il valore limite della frequenza per un numero infinito di ripetizioni.

Formulazione Frequentista di Probabilità:

$$P(E_i) = \lim_{N \rightarrow \infty} f(E_i) = \lim_{N \rightarrow \infty} \frac{\text{numero di successi } n_{E_i}}{\text{numero di ripetizioni } N}$$

Anche questa definizione soffre di alcune patologie importanti. Il limite infatti non è un limite in senso classico, secondo cui è possibile stabilire un valore arbitrariamente piccolo per cui f , a partire da un certo punto, si discosti da P per meno di questo valore. È un limite cosiddetto *debole o statistico* per cui si può solo stabilire un valore arbitrariamente piccolo per cui la *probabilità* che f si discosti da P per più di questo valore non può essere arbitrariamente grande² Anche in questo caso il concetto di limite debole è tautologico.

Formulazione Soggettiva di probabilità Entrambe le formulazioni viste sono quindi incomplete. Esiste una diversa formulazione, più sofisticata delle due sopra, e più complessa da descrivere, che tuttavia considera anche il *grado soggettivo di conoscenza* dell'evento aleatorio, detto anche *grado informato di fiducia*. Questa definizione fu elaborata in particolare dal matematico Bruno de Finetti nel XX secolo. Essendo la quantità di informazione disponibile possibilmente soggettiva, anche la definizione di probabilità lo è. Non ci spingiamo ad ogni modo oltre in questa discussione.

¹Si legga ad esempio 'Ten Great Ideas about Chance' di Diaconis.

²In linguaggio matematico, il limite classico è $\forall \epsilon > 0 \exists M [(N > M) \implies |f(E) - P(E)| < \epsilon]$ Il limite debole o statistico invece dice che: $\forall \epsilon > 0, \delta > 0 \exists M [(N > M) \implies P(|f(E) - P(E)| > \epsilon) < \delta]$

Formulazione assiomatica della probabilità Questa formulazione si deve al matematico russo Kolmogorov, e in generale allo sviluppo della teoria della misura degli insiemi di E. Borel (1871-1956). La formulazione e' detta assiomatica perche' si basa su alcuni assiomi matematici che una grandezza deve avere per poter essere considerata una misura di probabilità. La grandezza in questione non e' definita direttamente, ma attraverso *relazioni tra insiemi in una teoria degli insiemi*, nel nostro caso, tra elementi dello spazio degli eventi. In sostanza, con la notazione precedentemente introdotta:

Definizione assiomatica di probabilità. Sia Ω lo spazio degli eventi elementari e $\mathcal{F}(\Omega)$ lo spazio degli eventi E_i , allora se definiamo una variabile $P(E_i)$ che soddisfi gli assiomi:

- $P(\{E_i\}) \geq 0$
- $P(\Omega) = 1$
- Se $\{E_1\}, \{E_2\}$ di $\mathcal{F}(\Omega)$ sono disgiunti, allora $P(\{E_1\} \cup \{E_2\}) = P(\{E_1\}) + P(\{E_2\})$

Allora P e' una definizione di probabilità' degli eventi aleatori E_i ^a.

^aLa notazione non faccia confusione: E_i e' evento aleatorio in sè, $\{E_i\}$ e' l'elemento di $\mathbb{F}(\Omega)$. Per questo possiamo calcolare la probabilità sia dell'uno che dell'altro: $P(E_i)$ o $P(\{E_i\})$

Dei tre assiomi, i primi due valevano anche per le altre definizioni viste, il terzo invece e' una condizione sulla misura della probabilità che la lega a quella della misura di insiemi. Cerchiamo di capire meglio le operazioni logiche.

4.3 Operazioni logiche tra eventi aleatori

Abbiamo forse gia' intuito che stiamo utilizzando diversi simboli e concetti logici. Ha senso quindi cercare di mettere un po' di ordine. Vediamo di fare chiarezza sulla differenza di simboli tra operazioni logiche, algebriche e insiemistiche.

Iniziamo dicendo che tra *eventi aleatori* facciamo operazioni *logiche*, tra *insiemi* facciamo operazioni *insiemistiche*, tra probabilità facciamo operazioni *algebriche*. Riassumiamo queste in Tab. 4.1. Per aiutarci ricordiamo la differenza tra eventi elementari e_i , eventi aleatori generici E_i , gli elementi $\{E_i\}$ dello spazio degli eventi. Sfruttiamo anche per la rappresentazione grafica un diagramma detto di Venn (Figura 4.1) i cui insiemi e sottoinsiemi rappresentano eventi aleatori. Nel diagramma, gli insiemi A, C e B, C sono disgiunti, A, B non sono disgiunti, D e' sottoinsieme di A . Possiamo associare a questi insiemi eventi aleatori e ragionare sulla loro dipendenza logica.

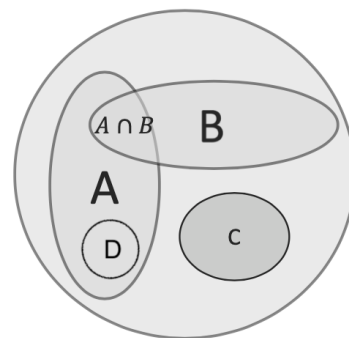


Figura 4.1: Diagramma di Venn

4.3.1 Dipendenza e indipendenza logica, probabilità condizionata

Per dipendenza logica tra due eventi E_1 e E_2 si intende il fatto che il verificarsi di uno ha un effetto sulla probabilità del verificarsi dell'altro. Attenzione che si tratta di eventi non necessariamente elementari. Ad esempio, dato un mazzo di carte da briscola, se $E_1 = \{\text{esce asso}\}$ e $E_2 = \{\text{esce asso di bastoni}\}$, chiaramente la probabilità di E_2 e' diversa se si e' verificato E_1 , passando da $1/40$ (prima di sapere che era uscito un asso) a $1/4$ (dopo averlo saputo).

Questo tipo di relazione logica tra eventi si indica con il simbolo $E_2|E_1$ e si legge anche 'E₂ dopo E₁' anche non si intende una successione temporale ma logica: 'dopo' sta per 'dopo aver saputo che e' avvenuto E₁'. In questo senso, attenzione che $E_2|E_1$ e' diverso da $E_1|E_2$.

Ora, in linea generale, abbiamo due possibilità:

1. L'evento E_1 ha una influenza sulla probabilità che avvenga E_2 . Allora in termini di termini di probabilità $P(E_2|E_1) \neq P(E_2)$. Gli eventi sono quindi **logicamente dipendenti**.
2. L'evento E_1 non ha nessuna influenza sulla probabilità che avvenga E_2 . Allora in termini di termini di probabilità $P(E_2|E_1) = P(E_2)$. Gli eventi sono quindi **logicamente indipendenti**.

Il calcolo della probabilità condizionata Cominciamo dal caso di eventi *logicamente dipendenti*. In questo caso possiamo avere due casi: **a) Implicazione logica** Si ha quando il verificarsi di un evento E_1 implica automaticamente il verificarsi dell'evento di un altro evento E_2 . A livello logico $E_1 \Rightarrow E_2$. Nel caso degli insiemi di [Figura 4.1](#) questo si verifica per insiemi-eventi D, A . E' chiaro che il verificarsi di D automaticamente implica il verificarsi A . Logicamente $D \Rightarrow A$. A livello di insiemi $D \subseteq A$. A livello di probabilità $P(D) \leq P(A)$ e $P(A|D) = 1$. Gli eventi sono **dipendenti** e connessi da *implicazione logica*. Attenzione che l'implicazione logica ha una 'direzione' infatti non e' vero che $A \Rightarrow D$ e possiamo solo dire che $P(D|A) \neq P(D)$. In altre parole, c'e' concomitanza logica tra A, D ma dipendenza logica solo in una direzione $D \Rightarrow A$. **b) Concomitanza logica** Si ha concomitanza logica tra due eventi E_1, E_2 quando si ha dipendenza logica ma non implicazione. Come si calcola quindi $P(E_1|E_2)$? Nel caso di eventi logicamente si hanno due casi. Cominciamo dal caso sopra rispondendo ora a quanto vale $P(D|A)$. In termini di insiemi, sapere che A è avvenuto corrisponde a $P(A) = 1$ e quindi a sua volta cio' corrisponde a *restringere* lo spazio degli eventi da Ω a A . Allora lo spazio degli eventi di D completo e' ora A e quindi

$$P(D|A) = \frac{P(D)}{P(A)} \quad \text{visto che } D \subseteq A \quad (4.1)$$

Se invece ci mettiamo nel caso piu' generale in cui i due insiemi non siano uno sottoinsieme dell'altro ma abbiamo intersezione non nulla, come ad esempio A, B allora, il ragionamento appena visto vale in forma analoga, ma la probabilità condizionata si calcola con la formula generale

Regola probabilità condizionata

Dati due eventi E_1, E_2 vale sempre la regola:

$$P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} \quad (4.2)$$

ovvero $P(A|B) = \frac{P(A \cap B)}{P(B)}$. Si noti come nel caso di D sopra la regola e' ancora valida in quanto $A \cap D = D$.

Esercizio:

Quali tra gli eventi del quadro di Venn sono logicamente dipendenti e indipendenti?

Incompatibilità logica I due eventi D, C sono **logicamente incompatibili** in quanto il fatto che avvenga uno dei due automaticamente implica il NON verificarsi dell'altro. Logicamente $D \Rightarrow \neg C$ e $C \Rightarrow \neg D$ dove con $\neg C, \neg D$ indichiamo l'evento *non avviene C* e *non avviene D*, detti anche **eventi complementari** di C e D rispettivamente. A livello di insiemi, essi sono disgiunti, ovvero $C \cap D = \emptyset$. Attenzione che i due eventi sono **dipendenti** infatti $P(D|C) = 0$ e $P(C|D) = 0$, contrariamente a quanto possa sembrare. Infine, si fa notare che il terzo assioma di Kolmogorov parla proprio di insiemi disgiunti, sostenendo che se $C \cap D = \emptyset$ allora $P(C \cup D) = P(C) + P(D)$. L'insieme $C \cup D$ e' semplicemente l'unione dei due insiemi. Ma a che evento logico corrisponde $C \cup D$ se i due eventi sono logicamente incompatibili? L'evento e' un evento di **somma logica** o **disgiunzione logica**. Si tratta in questo caso del verificarsi dell'evento C oppure dell'evento D , visto che entrambi *contemporaneamente* non possono avvenire.

Somma Logica o Disgiunzione logica In termini generali, si parla di disgiunzione o somma logica il verificarsi dell'evento E_1 , dell'evento E_2 o di entrambi e si indica con $E_1 \vee E_2$. Avremo due casi possibili: gli eventi hanno implicazione o disgiunzione logica. Calcoliamo la somma logica per tutti gli insiemi di [Figura 4.1](#). Uno dei casi e' gia' trattato dall'assioma di Kolmogorov, e si applica ai casi B, C, A, C e D, C . Un secondo caso e' quello di A, B che non sono disgiunti. Il terzo caso e' quello di A, D . Vediamo separatamente:

- Insiemi disgiunti. $A \cup C$ corrisponde all'evento logico $A \vee C$ la cui misura di probabilità e' $P(A \cup C) = P(A) + P(C)$ visto che $A \cap C = \emptyset$
- Insiemi non disgiunti. $A \cup B$ corrisponde all'evento logico $A \vee B$ la cui misura di probabilità e' $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Va quindi sottratta per evitare il doppio conteggio la probabilità $P(A \cap B)$.
- Sottoinsieme. $D \cup A$ corrisponde all'evento logico A quindi $P(D \cup A) = P(A)$.

Congiunzione o Concomitanza logica Due eventi E_1, E_2 sono in congiunzione (o concomitanza) logica se avvengono entrambi, non importa se prima l'uno poi l'altro o contemporaneamente. Nel diagramma quindi abbiamo due sole concomitanze logiche possibili, quella ovvia di implicazione logica tra A, D gia' vista, e quella tra A e B , in cui l'avverarsi di uno non implica necessariamente l'avverarsi dell'altro. L'evento logico e' quindi $A \wedge B$, l'insieme e' $A \cap B$. Ma come calcoliamo le probabilità $P(E_1 \cap E_2)$? Di fatto quello che stiamo facendo logicamente e' una dipendenza logica (simmetrica): possiamo pensare che sia avvenuto E_1 e sulla base di questo aggiorniamo la probabilità di E_2 oppure che sia avvenuto E_2 e aggiorniamo la probabilita di E_1 . La relazione [Equazione 4.2](#) implica anche che:

$$P(E_1 \cap E_2) = P(E_1|E_2)P(E_2) = P(E_2|E_1)P(E_1)$$

Applichiamo al nostro caso:

- Insiemi disgiunti. $A \cap C$ e $B \cap C$ hanno misura nulla, quindi $P(A \cap C) = P(B \cap C) = 0$. Questi due eventi, pur non essendo indipendenti, non possono avvenire in concomitanza.
- Insiemi non disgiunti. $A \cap B$ corrisponde all'evento logico $A \wedge B$ la cui misura di probabilità e' $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$
- Sottoinsieme. $D \cap A$ corrisponde all'evento logico D .

Esempio:

Dato l'evento aleatorio 'estrazione di una singola carta da un mazzo di carte da scopa' si considerino questi eventi: A ="esce un asso", B ="esce bastoni", C ="esce coppe". Si calcoli la probabilita' con le regole logiche viste sopra per l'evento $A + B, B + C, AB$. Si risolva anche con la definizione classica della probabilità.

E' utile aggiungere qualche altro concetto.

Evento irrealizzabile ed evento complementare Si dice evento irrealizzabile l'evento che corrisponde all'insieme \emptyset . Si dice evento complementare di E l'evento logico $\neg E$ o \bar{E} che si verifica se non si verifica E . Si tratta dell'insieme complementare di E in $\mathcal{F}(\Omega)$ la cui probabilità e' $1 - P(E)$. Anche se non intuitivo, talvolta e' piu' semplice calcolare la probabilità dell'evento complementare piuttosto che dell'evento originale. Cio' avviene nel caso l'evento originale sia complesso nel senso che possa avvenire con diverse modalita'.

Esempio:

Supponiamo di lanciare una moneta 3 volte, quale e' la probabilita' che esca almeno una volta testa?

Soluzione: L'evento in questione si realizza attraverso diversi esiti elementari, che sono in totale $2^3 = 8$. Con 3 lanci possiamo enumerarli: $TTT, TTC, TCT, CTT, TCC, CTC, CCT$ e vediamo che sono sette. Tuttavia se i lanci fossero N non saremo piu in grado di elencarli, dovrebbe combinare le probabilita' che testa esca 1 volta in qualsiasi dei lanci, o 2 volte in qualsiasi dei lanci, e cosi via. Tuttavia, logicamente e' molto piu semplice pensare all'evento complementare: $\neg E$ corrisponde all'evento che testa non esce mai, su N lanci: $P(\neg E) = 1/2^N$, allora $P(E) = 1 - P(\neg E) = 1 - 1/2^N$.

Logica	Simbolo	Insiemi	Algebraica
Condizionamento logico (A dopo B)	$A B$		
Implicazione logica (se accade A accade B)	\implies	\subseteq	\leq
Disgiunzione (somma) logica (accade A e anche B)	$\wedge, \cdot, \&$	\cup	\cdot
Congiunzione logica (accade A oppure B o entrambi)	$\vee, +, $	\cap	$+$
Negazione logica (non accade A)	$\neg, !$	A^c	$1 -$

Tabella 4.1: Simboli logici, di insiemi e algebrici per le operazioni logiche con probabilita'.

4.3.2 Legge della Probabilita' Totale, Marginalizzazione

E' spesso possibile partizionare lo spazio degli eventi in un set completo A_1, A_2, \dots, A_n di eventi disgiunti, ovvero:

$$\forall i, j \quad A_i \cap A_j = \emptyset \quad \bigcup_{i=1}^n A_i = \mathcal{F}(\Omega)$$

Ad esempio, si puo' sempre partizionare lo spazio degli eventi attraverso un evento A il suo complementare $\neg A$. Ovviamente ci possono essere diverse partizioni di $\mathcal{F}(\Omega)$, almeno una per ogni evento aleatorio dello spazio degli eventi. Vedremo che queste partizioni torneranno molto utili.

Per esempio, supponiamo di avere un evento E e una partizione n insiemi A_i di $\mathcal{F}(\Omega)$. Allora e' valida la legge della probabilita' totale:

Legge della probabilita' totale:

$$P(E) = \sum_{i=1}^n P(E|A_i)P(A_i)$$

Talvolta e' piu' semplice calcolare separatamente i vari termini $P(E|A_i), P(A_i)$ per ottenere $P(E)$.

Esempio:

Si supponga di avere due urne: la prima contiene 6 palline bianche e 8 nere, la seconda contiene 8 palline bianche e 4 nere; trovare la probabilità che, estraendo a caso una pallina da una delle due urne, la pallina estratta sia nera

Svolgimento Lo spazio universo può essere partizionato almeno in due modi: con l'evento aleatorio N che esca una biglia nera e $\neg N$ che esca una pallina non nera (bianca). Ovvero in un secondo modo, con l'evento aleatorio U_1 che l'estrazione sia stata dalla prima urna e il suo contrario $\neg U_1 = U_2$ che l'estrazione sia avvenuta dalla seconda urna. Applicando la legge della probabilità totale all'evento che esca una pallina nera, possiamo scrivere

$$P(N) = P(N|U_1)P(U_1) + P(N|U_2)P(U_2)$$

Tutti gli addendi e i fattori ci sono dati nel testo: $P(N|U_1) = 8/14$; $P(N|U_2) = 4/12$ e $P(U_1) = P(U_2) = 1/2$. E quindi $P(N) = 19/42 = 0.452 = 45.2\%$

La legge della probabilità totale è usata assieme al concetto di *marginalizzazione di una variabile*. Marginalizzare la variabile E rispetto alla variabile aleatoria A significa calcolare E 'indipendentemente' da A . La domanda infatti che ci siamo posti è analoga a 'quale è la probabilità che esca una pallina nera *indipendentemente* dalla scelta dell'urna?'

4.3.3 Teorema di Bayes

T. Bayes (1702-1761) per primo introdusse una relazione utile tra eventi aleatori che prese poi il nome di formula di Bayes. La formula di Bayes è una relazione valida per sé, ma assume un significato proprio nel contesto del confronto tra dati e ipotesi, dando origine a quella che è nota come scuola statistica Bayesiana in opposizione a quella Frequentista (vedi Approfondimento 4.3). Qui ci limitano a scrivere la relazione, che è valida per ogni coppia di eventi aleatori.

Teorema di Bayes.

Siano H e D due eventi aleatori qualsiasi. Dalla relazione sulla loro congiunzione logica $P(DH) = P(D|H)P(H) = P(H|D)P(D)$ segue che :

$$P(H|D) = \frac{P(H)}{P(D)}P(D|H)$$

La formula lega la probabilità condizionata di H noto D a quella inversa di D noto H .

Esempio:

Le industrie I e II producono pezzi difettosi con probabilità $1/5$ e $1/50$ rispettivamente. Sapendo che la produzione dell'industria I è doppia di quella della II, calcolare: a) la probabilità di trovare un pezzo difettoso; b) la probabilità che un pezzo trovato difettoso provenga dall'industria I.

Soluzione: Indichiamo con $P(I) = 2/3$ e $P(II) = 1/3$ rispettivamente le probabilità che un pezzo sia prodotto dall'industria I oppure II, con $P(d|I) = 1/5$ e $P(d|II) = 1/50$ le probabilità condizionate di avere un pezzo difettoso proveniente dall'industria I oppure II.

a) usando il teorema della probabilità totale, probabilità di un pezzo difettoso

$$P(d) = P(d|I)P(I) + P(d|II)P(II) = 1/5 \cdot 2/3 + 1/50 \cdot 1/3 = 0.14$$

b) usando la formula di Bayes

$$P(I|d) = \frac{P(d|I)P(I)}{P(d)} = \frac{1/5 \cdot 2/3}{0.14} = 0.93$$

4.4 Calcolo Combinatorio

Il calcolo combinatorio ha a che fare con il computo del numero di modi di raggruppare un numero finito di oggetti presi all'interno di un insieme. Nel raggruppare possiamo usare uno schema logico, mostrato in Fig. 4.2 secondo cui valutiamo: se l'ordine degli elementi raggruppati è rilevante, se usiamo tutti o un sottoinsieme degli elementi a scelta, e se ci possono essere ripetizioni.

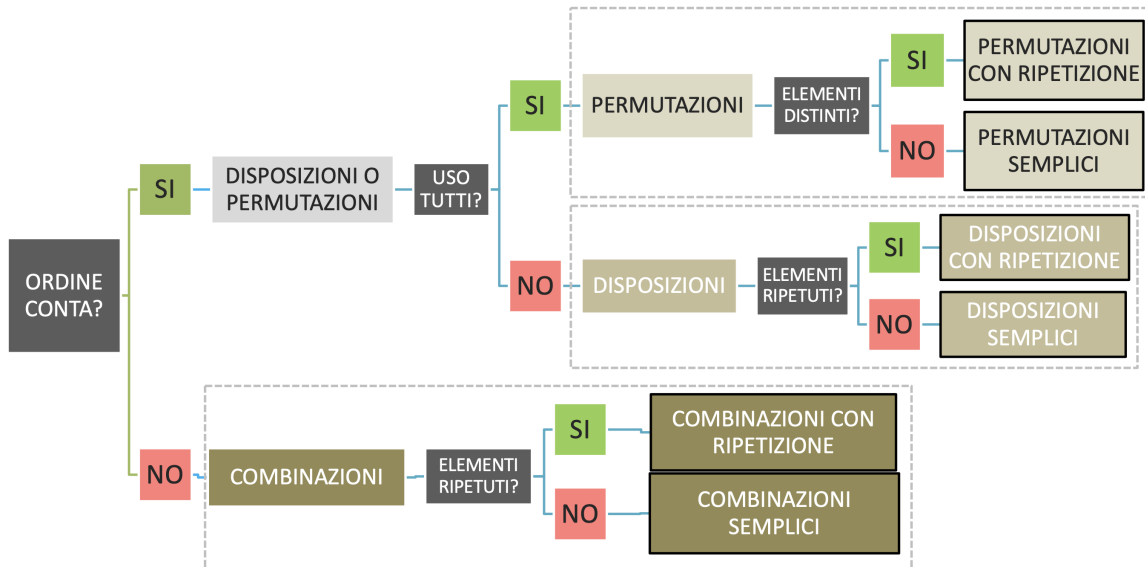


Figura 4.2: Schema logico per la comprensione di tipologia di combinatoria

Permutazioni

Nelle permutazioni, si sistemano tutti gli elementi di un certo insieme in un certo ordine.

PERMUTAZIONI: Dati n oggetti distinti, chiamiamo permutazioni semplici degli n oggetti tutti i gruppi che si possono formare con gli n oggetti dati prendendoli, ogni volta, tutti.

Il numero di permutazioni P_n di un insieme di dimensione n è:

$$P_n = n!$$

PERMUTAZIONI CON RIPETIZIONE: Dati n oggetti dei quali il primo si ripete r_1 -volte, il secondo si ripete r_2 -volte, e così via fino a all'elemento m -esimo ($m < n$) che si ripete r_m volta, si chiamiamo permutazioni con ripetizione tutti i gruppi che si possono formare con gli n oggetti dati prendendoli, ogni volta, tutti. Esse sono in numero di

$$P_n^{r_1, r_2, \dots, r_m} = \frac{n!}{r_1! r_2! \dots r_m!}$$

La relazione è semplice da dimostrare, pensando che per il primo elemento si hanno n possibilità, ma per il secondo rimangono $(n - 1)$ possibilità e così via, ottenendo $n(n - 1)(n - 2) \dots = n!$

È possibile che l'insieme di cui si vuole permutare gli elementi, abbia elementi ripetuti. In questo caso si usano le permutazioni con ripetizione. Si dimostra pensando che permutando i numeri ripetuti il risultato non cambia: $P_n^{r_1, r_2, \dots, r_m} = \frac{P_n}{P_{r_1} P_{r_2} \dots P_{r_m}}$. È ovvio che $P_n > P_n^r$

Esempio:

Quante parole (anche non si senso compiuto) si possono formare con la parola MUONE?
Soluzione La parola muone ha tutte lettere diverse, e quindi permutando le 5 lettere si possono costruire $P_5 = 5! = 120$ parole.

Quante parole si possono formare con la parola ELETTRONE? **Soluzione** La lettera E si ripete 3 volte, e la lettera T due volte, allora

$$P_9^{2,3} = \frac{9!}{2!3!} = 30240$$

Disposizioni

Nelle disposizioni, a differenza delle permutazioni, sistemano solo alcuni elementi in un gruppo in un certo ordine, non necessariamente tutti.

DISPOSIZIONI SEMPLICI Dati n oggetti distinti, si dicono disposizioni semplici di n oggetti di classe k (con $k < n$ e con tutti gli oggetti presi al massimo una sola volta) tutti i raggruppamenti di k oggetti che si possono formare in modo che due raggruppamenti differiscano o per gli oggetti presenti o per l'ordine in cui gli oggetti sono disposti. Il numero $D_{n,k}$ di queste disposizioni semplice e':

$$D_{n,k} = \frac{n!}{(n-k)!}$$

DISPOSIZIONI CON RIPETIZIONE Dati n oggetti distinti, si dicono disposizioni con ripetizione di n oggetti di classe n (k puo' essere maggiore, minore o uguale ad n) tutti i raggruppamenti di k oggetti che si possono formare, anche ripetendo un oggetto fino a k volte, in modo che due raggruppamenti differiscano o per gli oggetti o per l'ordine in cui sono disposti. Il numero $D_{n,k}^r$ di queste disposizioni e':

$$D_{n,k}^r = n^k$$

Di nuovo per la dimostrazione basti pensare che per il primo elemento si hanno n possibili, per il secondo $(n - 1)$ possibilità, ma ci arresta al dopo k volte:

$$D_{n,k} = n(n-1)(n-2) \dots (n-(k-1)) \dots \frac{(n-k)!}{(n-k)!} = \frac{n!}{(n-k)!}$$

dove si e' moltiplicato e diviso per lo stesso numero $(n-k)!$. Vale anche inoltre la relazione $D_{n,n} = P_n$. Per le disposizioni con ripetizioni, basti pensare semplicemente che ogni volta si puo' scegliere tra n elementi, per k volte quindi $n \cdot n \cdot n(k - \text{volte}) = n^k$.

Esempio:

Quanti numeri di targa si possono fare in Italia con la regola attuale? **Soluzione** Le targhe hanno 4 lettere e 3 cifre. Si possono fare $D_{26,4}^r D_{10,3}^r = 26^4 10^3 = 456,976,000$

Combinazioni

Con le combinazioni si sistemano elementi di un gruppo ma ordine degli elementi non conta.

COMBINAZIONI SEMPLICI Dati n oggetti distinti, si dicono combinazioni semplici di n oggetti di classe k (con $k < n$, con tutti gli oggetti presi al massimo una sola volta) tutti i raggruppamenti di k oggetti che si possono formare in modo che due raggruppamenti differiscano per gli oggetti presenti in essi. Il numero di combinazioni semplici è:

$$C_{n,k} = \binom{n}{k} = \frac{n!}{k!(n-k)!} \quad (4.3)$$

COMBINAZIONI CON RIPETIZIONE Dati n oggetti distinti, si dicono combinazioni con ripetizione di n oggetti di classe k (k può essere maggiore, minore o uguale ad n) tutti i raggruppamenti di k oggetti che si possono formare, anche ripetendo un oggetto fino a k volte, in modo che due raggruppamenti differiscano per gli oggetti. Il loro numero è:

$$C_{n,k}^r = \binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}$$

È facile ottenere la formula sopra pensando che si tratta di disposizioni $D_{n,k}$ con le permutazioni sugli oggetti con gli stessi diversi ma ordine diverso P_k :

$$C_{n,k} = \frac{D_{n,k}}{k!} = \frac{n(n-1) \dots (n-k+1)}{k!} = \frac{n!}{k!(n-k)!}$$

Si veda ad esempio la discussione nell'approfondimento 3.2 sul ruolo delle combinazioni di fluttuazioni casuali nella creazione della Gaussiana.

4.5 Commenti Finali

Ricapitolando

Abbiamo visto come una teoria della probabilità sia un costrutto matematico che assegna un numero detto probabilità ad eventi aleatori. Questa assegnazione si può fare in linea di principio in diverse maniere, e una di queste è secondo la definizione assiomatica di Kolmogorov, che legga la stima di probabilità con la misura di insiemi. In questo modo si possono trasporre operazioni logiche con operazioni insiemistiche e algebriche tra probabilità. Importanti i concetti di connessione logica come dipendenza e indipendenza da cui si possono calcolare somma e prodotti logici. Nel caso di spazi discreti, fa comodo il calcolo combinatorio.

Concetti chiave: evento aleatorio * spazio eventi elementari * spazio eventi * definizione classica e frequentista di probabilità * Limite debole o statistico * Definizione Assiomatica di probabilità * Operazioni logiche tra eventi aleatori * Dipendenza e Indipendenza logica * probabilità condizionata * Implicazione Logica * Somma e prodotto logico * Evento impossibile e complementare * Legge probabilità totale * Legge di Bayes * Combinazioni, disposizioni, permutazioni.

Quesiti di autovalutazione

- Es. 4.1** Dato un mazzo di carte da poker (senza jolly) non truccato, quale è la probabilità che esca una figura di cuori? Risolvere in termini di legge della probabilità composta.
- Es. 4.2** [BV] Un dado non truccato è lanciato 2 volte, calcolare la probabilità: a) di avere un solo 6; b) che entrambi i numeri siano pari; c) che la somma dei numeri sia 4;
- Es. 4.3** Si tirano 4 monete. Qual è la probabilità che escano tutte teste? Qual è la probabilità non esca mai croce?
- Es. 4.4** In un gruppo di amici ci sono il 50% di maschi. Supponiamo che il 5% dei maschi e il 10% delle femmine amino mangiare vegano. Si sceglie a caso una persona vegana, qual è la probabilità che sia un maschio?

□ **Es. 4.5** Un gruppo di studenti è stato classificato secondo il voto riportato ad un dato esame e a seconda che l'esame in oggetto fosse stato fatto al primo o secondo appello come in tabella.

	Primo Appello	Secondo Appello
≤ 24	40	15
> 24	45	100

Si estrae a caso dal collettivo uno studente. Si considerino gli eventi A : voto ≤ 24 , B : esame sostenuto primo appello. Calcolare: a) $P(A)$, $P(B)$, b) $P(A \cdot B)$, c) $P(B|A)$. Infine, A e B sono indipendenti?

□ **Es. 4.6** In una certa località, la popolazione delle persone in possesso della patente automobilistica 'e composta per il 47% da uomini e per il restante da donne. Le statistiche mostrano che il 40% degli uomini ha un incidente nell'arco di un anno, mentre la stessa percentuale per le donne scende al 20%. Se un individuo ha avuto un incidente nel dato anno, qual 'e la probabilità che si sia trattato di una donna?

□ **Es. 4.7** [BV] Una moneta non truccata e' lanciata N volte, calcolare la probabilità che: a) il risultato sia testa per la prima volta all' N -mo lancio; b) testa e croce appaiano lo stesso numero di volte (ovviamente N deve essere pari); c) venga testa esattamente 2 volte; d) venga testa almeno 2 volte.

□ **Es. 4.8** Stimare quanti anagrammi anche senza senso si possono formare con la parola *CUORE* senza ripetizione.

□ **Es. 4.9** Stimare quanti anagrammi anche senza senso si possono formare con la parola *MAMMA*.

□ **Es. 4.10** In quanti modi diversi posso far accomodare 5 familiari in un tavolo da 3 sedie numerate?

□ **Es. 4.11** Quanti numeri di 2 cifre si possono formare con le cifre 1, 2, 3?

□ **Es. 4.12** Si supponga che un individuo abbia 3 giacche, 4 pantaloni, 5 camicie e 6 cravatte. In quanti modi diversi si potrebbe vestire?

□ **Es. 4.13** E' necessario costituire 1 commissione giudicatrice per un concorso pubblico che sia costituita da 3 statistici e 2 economisti: si stimi il numero totale di possibilità senza imporre alcun vincolo scegliendo tra un insieme di 10 statistici e 6 economisti rispettivamente.

□ **Es. 4.14** Determinare quante targhe è possibile formare al massimo partendo da tre lettere seguite da 4 cifre numeriche

□ **Es. 4.15** Si organizza un concerto con capienza massima di 180mila persona. Per i biglietti si usano una numerazione fatta di una sequenza di 3 caratteri dell'alfabeto inglese a 26 caratteri e 1 cifra numerica. Si riesce ad avere un numero sufficiente di biglietti con numeri diversi?

□ **Es. 4.16** Si lancia 2 volte una moneta truccata per cui $P(T) = 0.2$ e $P(C) = 0.8$. Considerate la variabile casuale $X =$ numero di teste. Definite la sua distribuzione di probabilità e calcolate il valore atteso e la varianza.

□ **Es. 4.17** [Marchetti] In un laghetto ci sono 10 pesci di cui 2 sono rossi. Pesci a caso senza ripetizione 5 pesci. Qual è la probabilità di pescare 1 pesce rosso? Nell'ipotesi che avessi pescato 5 volte ributtando dentro il pesce, sarebbe cambiata la probabilita?

□ **Es. 4.18** Due ragazzi parlano di avere 10 figli, la loro scelta preferita' e' 4 di un sesso e 6 dell'altro sesso. Verificare con dimostrazione se é piu' probabile questa configurazione o quella in cui nascono 5 maschi e 5 femmine.

□ **Es. 4.19** Si determini il numero di modi in cui si possono scegliere 6 domande su 10 per formare il testo di una prova di ingresso.

- **Es. 4.20** Problema della Tavolozza dei colori: si abbiano 10 gocce bianche e 10 gocce nere. Si mescolano tra loro 10 gocce scelte tra i due colori. Quanti colori si possono formare?
- **Es. 4.21** Una squadra di operai edili deve essere composta da 2 muratori e da 4 manovali, scelti da un totale di 5 muratori e di 6 manovali. Le selezioni dei muratori e dei manovali sono indipendenti. Quante diverse combinazioni sono possibili?
- **Es. 4.22** In una scatola contenente 16 cioccolatini, 4 sono con ripieno al cocco. Qual è la probabilità che scegliendo 4 cioccolatini, nessuno sia con ripieno al cocco?
- **Es. 4.23** Un test a risposta multipla ha 5 domande, ognuna con 5 possibili risposte. Se rispondi sempre a caso, qual è la probabilità di rispondere correttamente a esattamente 3 domande?
- **Es. 4.24** La probabilità che una persona prenda il raffreddore durante l'inverno è 0.4. Si selezionano a caso 10 persone. Qual è la probabilità che esattamente 4 di loro prenderanno il raffreddore?

Approfondimenti

Apr. 4.1 Introduzione storica al concetto di probabilità. La ragione è profonda, e' pressoché impossibile dare una definizione di probabilità che possa abbracciare tutti gli eventi aleatori. La ragione è piuttosto semplice da intuire. Quando si dice che la probabilità che dal lancio di una moneta esca testa è 50%, oppure che su 3 lanci la probabilità che esca testa tre volte è del 1.25%, ovviamente non si sta dicendo che avverrà così. Per fare la predizione ci stiamo immaginando che avvenga l'evento. Dal punto di vista del mondo, i fenomeni avvengono ($P=1$) oppure non avvengono ($P=0$). Alla fine noi elaboriamo il calcolo solo attraverso l'aiuto del nostro bagaglio personale (un infante non ha questa capacità, ripete a volte decine di volte la stessa azione ed è felice nell'osservare la ripetizione, ad esempio, che un biberon lanciato dal tavolo cade sempre verso il basso e mai verso l'alto). Nel lancio della moneta, noi sappiamo che ad esempio, può uscire o testa o croce, non altro. Sappiamo inoltre che mentre non possiamo prevedere se esce testa o croce in un unico lancio, possiamo scommettere di più sul fatto che se lanciamo 1 miliardo di volta la moneta, ci aspettiamo che meta' delle volte esca testa. A partire da questa serie lunga (il miliardo di lanci), facciamo un'inferenza su quello che avverrà in un unico lancio...bel salto!

I primi documenti che si conservano risalgono al XVI secolo e riguardano i giochi d'azzardo. G. Cardano (1501-1576) fu il primo matematico a studiare la probabilità, pubblicando nel 1526 "Liber de Ludo aleae". G. Galilei, nella sua opera "Sopra le scoperte dei dadi" (1630), si occupò di probabilità, stimolato da quesiti posti da nobili fiorentini appassionati del gioco della "zara"³ e fece osservazioni probabilistiche legate alla propagazione degli errori nelle misurazioni. Successivamente, grossi progressi si devono a Antoine Gombaud, Chevalier de Méré (1607 – 1684). Appassionato giocatore, matematico per passione, e amico di Blaise Pascal (1623-1662), con quest'ultimo intraprese una fitta corrispondenza. Nella discussione su come calcolare equamente la posta in un gioco interrotto prima delle conclusioni, si formò la teoria del calcolo combinatorio moderno (problema dei punti). Un'altro esempio importante, noto come "Problema del Cavaliere de Méré" si chiede come mai via sia una differenza tra la probabilità di ottenere almeno un 6 su 4 tiri singoli di un dado a 6 facce, oppure di ottenere almeno un doppio 6 su 24 tiri di due dadi a 6 facce. La formalizzazione della teoria della probabilità comincia nel 1700 con Jakob Bernoulli, uno degli illustri Bernoulli assieme al padre e allo zio. Nel libro «Ars Conjectandi» del 1713 illustrò a fondo il Calcolo combinatorio, formulò la Legge dei grandi numeri (Approfondimento 4.2), e introdusse il concetto di valore di aspettazione. La prima teoria analitica della probabilità si deve a Laplace che nel 1812 formalizzò il procedimento matematico del ragionamento per induzione basato sulla probabilità, che noi oggi riconosciamo come quello del teorema Bayes. Nello stesso anno include un'esposizione del metodo dei minimi quadrati e formalizza la definizione classica di probabilità di A. De Moivre. Ancora non bastava, a cavallo del secondo il grande matematico David Hilbert, durante una famosa conferenza, a cui era stato chiesto di presentare gli 8 problemi per il XX secolo disse, tra altri, «trattare in modo assiomatico quelle parti delle scienze fisiche in cui la matematica gioca un ruolo importante; in particolare la teoria della probabilità e lo

³GIOCO DELLA ZARA: Si gioca con tre dadi: a turno ogni giocatore chiama un numero da 3 a 18, quindi getta i dadi. Vince chi per primo ottiene il punteggio pari al numero chiamato.

sviluppo rigoroso e soddisfacente del metodo delle medie in fisica matematica... ». Da qua alla teoria assiomatica della probabilità, assieme agli avanzamenti dell'algebra e della insiemistica, porto alla formulazione assiomatica di probabilità di Kolmogorov.

Appr. 4.2 Legge dei grandi numeri. La Legge dei Grandi Numeri e' quella che ispira la formulazione frequentista della probabilità. All'aumentare del numero di prove N , la frequenza di un evento tende al limite (statistico o debole, come abbiamo visto) alla probabilità di quell'evento. Analogamente, le proprietà del campione di misure si avvicinano al limite a quella della popolazione di provenienza

Appr. 4.3 Teorema Bayes nella statistica bayesiana. Mentre la formula di Bayes che qui riportiamo

$$P(H|D) = \frac{P(H)}{P(D)}P(D|H)$$

e' valida per ogni variabile aleatoria H, D e usata di frequente in statistica, essa assume un significato particolare quando le variabili aleatorie in questione sono D l'evento aleatorio costituito dall'ottenere un certo set di misure a partire da un certo esperimento (evento Dati) e H l'evento aleatorio costituito dal fatto che una certa teoria spieghi i dati D (evento Teoria). A partire da questo si possono infatti intuire che possono coesistere due vite di interpretazione. Allora $P(H|D)$ rappresenta la probabilità che l'ipotesi H sia vera *alla luce* dei dati sperimentali D , questo termine e' detto anche *posterior likelihood* per questa ragione. $P(H)$ e' la probabilità a priori (cioe' prima di aver preso i dati) che l'ipotesi sia vera; questo termine e' detto *prior* per questa ragione. Anche il termine $P(D)$ e' un prior, ma di solito non ha molta importanza definirlo in quanto semplicemente 'normalizza a 1' la posterior. Infine, un termine che tornera' diverse volte in questo corso e' l'ultimo termine $P(D|H)$ detto *likelihood*. Questo termine indice la probabilità di ottenere un certo set di dati D *assumendo* che l'ipotesi sia vera. Francesca, studente del 1 anno, ragiona cosi: non si puo' assegnare una probabilità alla veridicità di una ipotesi, in fondo in fondo una ipotesi e' o vera $P = 1$ o falsa $P = 0$. Ogni valore intermedio costituisce una affermazione non scientifica come per esempio 'e' piu probabile che vi siano 4 famiglie di neutrini piuttosto che 3'. La frase non e' una sentenza sulla natura ma solo una speculazione. D'altro canto Beatrice ragiona cosi: certamente una ipotesi o e' vera o e' falsa, ma la speculazione fa parte del metodo scientifico. Si va a cercare scientificamente qualcosa dove sembra piu' probabile, aggiornando il concetto di 'piu' probabile' sulla base di informazioni via via acquisite.

Esercizio:

Meditare sui concetti di prior, likelihood, posterior e sulla differenza tra posterior e likelihood.

L'approccio di Francesca e' detto *Frequentista* mentre l'approccio di Beatrice e' detto *Bayesiano*. La letteratura che confronta i due metodo e' vasta e affascinante. Capirne bene le differenze e' importante. In sostanza, non potendo assegnare $P(H)$, nell'approccio frequentista e' possibile solo calcolare la likelihood, ma non la posterior. Vista la definizione di likelihood, l'approccio e' quello di assumere una ipotesi e confrontarla con i dati (e le loro fluttuazioni). Si chiama frequentista, perche' e' fortemente basata sul teorema dei grandi numeri (Approfondimento 4.2) e la definizione frequentista di probabilità: all'aumentare della dimensione del campione di misura, i parametri sperimentali sono noti con maggior precisione perche' aumenta la stima delle incertezze casuali. Si stabilisce quindi una chiara *probabilità dei dati*. Invece nell'approccio Bayesiano, la posterior in sostanza e' una *probabilità sull'ipotesi* che viene costantemente aggiornata dall'aumento della conoscenza attraverso la presa di ulteriori dati. I due approcci portano spesso a risultati simili sulle stime di parametri, ma tendono a dare stime di livelli di confidenza diversa. Abbiamo visto gia' quanto il livello di confidenza sia un concetto in un certo senso debolmente definito o soggetto a considerazioni. I due approcci quantificano a fondo questa differenza.

Appr. 4.4 Proprieta' dei fattoriali e dei coefficienti. binomiali Scriviamo qui di seguito alcune utili relazioni per il fattoriale. Si definiscono $0! = 1! = 1$. Valgono queste proprieta':

$$(n + 1)! = (n + 1)n! \quad n! = n(n - 1)! = n(n - 1)(n - 2)!$$

Per quanto riguarda il coefficiente binomiale, valgono

$$\binom{n}{0} = \binom{n}{1} = \binom{n}{n} = 1$$

$$\binom{n}{k} = \binom{n}{n-k} \quad \binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1} \quad \binom{n}{k+1} = \binom{n}{k} \frac{n-k}{k+1}$$

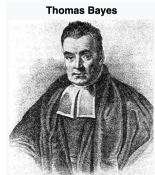
Appr. 4.5 Persone.

Blaise Pascal (1623-1662) Blaise Pascal fu un bambino prodigio: a sedici anni scrisse un trattato di geometria proiettiva e lavorò con Pierre de Fermat sulla teoria delle probabilità. Nel 1644, a 21 anni, costruì la sua prima macchina calcolatrice meccanica, che poi fu chiamata la Pascalina. Nel libro «*Traité du triangle arithmétique*» del 1654, descrisse il triangolo aritmetico, alla base del calcolo combinatorio. Dopo un'esperienza mistica in seguito ad un incidente in cui rischiò la vita, Pascal abbandonò matematica e fisica per dedicarsi alle riflessioni religiose e filosofiche. Oggi un famoso linguaggio di programmazione prende il nome da questo bambino prodigio.



Blaise Pascal

Thomas Bayes (1701-1761) Thomas Bayes fu uno statistico e filosofo inglese. Fu membro della Royal Society inglese, ma tuttavia pubblicò solo due opere: "Benevolenza divina, o tentativo di dimostrare che il fine principale della Provvidenza e del governo divini è la felicità delle sue creature" e "Introduzione alla dottrina dei flussioni e difesa dei matematici contro le obiezioni dell'autore di *The Analyst*" che fu pubblicato postumo assieme al teorema che porta il suo nome. È curioso come un personaggio minore come Bayes abbia poi dato nel corso dei secoli il nome a un intero settore della disciplina statistica che scatena ancora oggi forti argomenti scientifici: l'interpretazione Bayesiana.



Thomas Bayes

Pierre-Simon Laplace 1749-1827 Pierre-Simon Laplace fu un matematico francese. Nel 1769, a 23 anni, si trasferì a Parigi, capitale europea dell'Illuminismo. Con sé portava una lettera di raccomandazione indirizzata a D'Alembert, uno dei matematici più prestigiosi di Parigi. D'Alembert non prestò nessuna attenzione né alla lettera di raccomandazione né a Laplace, che non fu neanche ricevuto. Laplace però non si scoraggiò e scrisse lui stesso una lettera di quattro pagine a D'Alembert, dove dimostrava di conoscere i fondamenti della meccanica e le opere di Newton. A quel punto D'Alembert si accorse della natura di Laplace e gli rispose: «*Signor Laplace, vedete che non prendo molto in considerazione le raccomandazioni. Non ne avete bisogno, vi siete fatto conoscere meglio con questa vostra lettera e questo mi basta. Vi devo il mio appoggio. D'Alembert*». Grazie a questa tenacia studiamo ore molte teoremi e leggi in Fisica che prendono il suo nome.



Pierre-Simon Laplace

Bruno de Finetti (1906-1985) Bruno De Finetti (Innsbruck, 13 giugno 1906 – Roma, 20 luglio 1985) è stato un matematico, statistico e accademico italiano, noto soprattutto per la formulazione della concezione soggettiva operativa della probabilità. Dal 1927 al 1931 lavora all'Istituto nazionale di statistica, appena creato, e viene preposto all'ufficio matematico, diretto da Luigi Galvani. In quel periodo costruisce le tavole di mortalità del 1921 e ricostruisce quelle dei censimenti precedenti. Dal 1931 al 1946 lavora come attuario e statistico presso le Assicurazioni Generali di Trieste, dove è a capo del servizio meccanografico e dell'ufficio razionalizzazione. È probabile che questa esperienza lavorativa lo abbia portato a formulare la sua teoria informata della probabilità. Ha anche insegnato all'Università di Padova. È tra i primi, intorno al 1960, a promuovere fra studenti gare matematiche che possano portare i migliori a partecipare a gare internazionali.



Bruno de Finetti

Andrey Kolmogorov 1903-1987 Preso da <https://l.infn.it/hv>. Di lui, uno dei suoi studenti, V.I. Arnol'd, racconta: «Kolmogorov non spiegava mai niente: poneva problemi. E non li espandeva nemmeno: dava agli studenti completa indipendenza e non forzava nessuno a fare nulla, sempre aspettando di ascoltare dagli studenti qualcosa di eccezionale. Brillava tra i professori che ho incontrato per il suo completo rispetto della personalità dello studente.» Secondo Kolmogorov è essenziale proporre agli studenti problemi che siano alla loro portata e allo stesso tempo richiedano tutta la loro concentrazione e sforzo, e che riescano a stimolare il loro sviluppo più che ottenere risultati utili in sé. Scrisse anche un libello di riferimento “Matematica come professione”, ed organizzò Olimpiadi di Matematica nelle scuole, contribuì negli anni a sviluppare i programmi scolastici di scuole medie e superiori, producendo libri di testo, programmi, definizioni didattiche.



A. Kolmogorov con i suoi studenti.

Capitolo 5

Distribuzioni di probabilità di variabili aleatorie

Quest'opera è distribuita con licenza [Creative Commons](#) "Attribuzione – Non commerciale – Condividi allo stesso modo 3.0 Italia".



In questo capitolo, con il formalismo introdotto della teoria della probabilità e delle operazioni logiche tra eventi aleatori, introduciamo un nuovo formalismo riferito ora a *variabili aleatorie* (*random variable* o *RAV* in inglese) e di *distribuzioni di probabilità*. Quello che vedremo non ha a che fare direttamente con la misura sperimentale, fatta di campioni 'imperfetti' di dati di numerosità finita. Qui lavoreremo a livello di funzioni matematiche rappresentanti le distribuzioni. L'idea di base è che, come vedremo nel [Capitolo 7](#), i dati sperimentali andranno confrontati con queste funzioni teoriche.

5.1 Variabile aleatoria

Finora abbiamo parlato di evento aleatorio o casuale, ovvero di un evento che può avere esiti diversi (lo spazio degli eventi elementari) imprevedibili singolarmente, di cui però si può calcolare la probabilità che si verifichi un esito rispetto ad un altro. Abbiamo visto che possibili eventi aleatori sono 'esce il numero 6 nel lancio singolo di un dado a 6 facce' oppure 'esce un numero pari in un lancio singolo di un dado a 6 facce' oppure 'ottengo una misura di periodo di oscillazione compresa tra $(-1.99, 2.01)$ ' oppure 'ottengo una certa misura di periodo di oscillazione'.

Essendo gli esiti nello spazio degli eventi in un certo senso 'fenomeni', a partire da questi eventi è talvolta più ragionevole costruire *variabili aleatorie* che mappano i fenomeni dello spazio degli eventi a numeri reali. Ad esempio, nel lancio singolo del dado a 6 facce possiamo costruire diverse Variabili aleatorie:

$$X_1 : \mathcal{F}(\Omega) \rightarrow \mathbb{R} \quad X_1 = \begin{cases} 1 & \text{se 'esce pari'} \\ 0 & \text{altrimenti} \end{cases} \quad (5.1)$$

oppure

$$X_2 : \Omega \rightarrow \mathbb{R} \quad X_2 = \begin{cases} 1 & \text{se 'esce numero 1'} \\ 2 & \text{se 'esce numero 2'} \\ 3 & \text{se 'esce numero 3'} \\ \dots & \text{e così via} \end{cases} \quad (5.2)$$

Si può considerare la variabile aleatoria come una funzione avente come dominio lo spazio degli eventi e come codominio lo spazio \mathbb{R} . Gli elementi del codominio sono detti variati (*random variates*). Tuttavia spesso si intercambia l'evento aleatorio con il proprio variato aleatorio.

Vediamo anzitutto che ci sono due tipi di **variabili aleatorie**, **quelle discrete e quelle continue**. Nel caso discreto il numero di elementi dello spazio degli eventi è finito, nel caso continuo,

il numero degli eventi è infinito. Ad esempio, sfruttando gli esempi sopra, l'evento aleatorio dato da 'ottengo una misura di periodo di oscillazione compresa tra $(-1.99, 2.01)$ s' è discreto mentre l'evento aleatorio 'ottengo una certa misura di periodo di oscillazione' è continuo, in quanto il risultato di una misura in linea di principio è un numero reale (se non lo è, è perché lo misuriamo attraverso strumenti con precisione finita, ma intrinsecamente lo sarebbe).

E' possibile ora costruire la distribuzione di probabilità dei variati di variabili aleatorie. La distribuzione di probabilità è intrinseca all'evento aleatorio, è una sua caratteristica. Essa in alcuni casi può essere descritta attraverso una funzione. Le due categorie di eventi aleatorie determinano due classi di distribuzioni di probabilità:

- Per i Variati discreti si costruiscono **Funzioni di Massa di Probabilità (Probability Mass Functions o PMF)** che assegnano ai Variati una probabilità nell'intervallo reale $[0, 1] \subset \mathbb{R}$;

$$f_X : \mathbb{R} \longrightarrow [0, 1]$$

- Per i Variati continui si costruiscono **Funzioni di Densità di Probabilità (Probability Density Functions o PDF)** che assegnano ai Variati una probabilità nell'intervallo reale positivo $[0, \infty) \subset \mathbb{R}$

$$f_X : \mathbb{R} \longrightarrow \mathbb{R}$$

A breve capiremo la differenza tra le due. Prima di farlo rimarchiamo nuovamente come spesso si possa anche parlare di distribuzioni di probabilità di un evento aleatorio senza bisogno di fare il passaggio formale con le variabili aleatorie e i variati.

Si pone fin d'ora la attenzione sui seguenti punti. In primo luogo, una distribuzione di probabilità oltre ad essere funzione del Variato x può essere parametrizzata da uno o più parametri $\lambda_1, \lambda_2, \dots$. Di conseguenza, la maggior parte delle distribuzioni che vedremo sono *famiglie parametriche* di distribuzioni. In secondo luogo, si noti come nel caso di due o più variabili aleatorie i cui eventi sono in una certa concomitanza logica, si possono costruire funzioni di probabilità *bivariate* o *multivariate*. Questo sono funzioni ad esempio del tipo:

$$f_{XY} : \mathbb{R}^2 \longrightarrow \mathbb{R} \quad \text{e} \quad f_{XYZ\dots} : \mathbb{R}^n \longrightarrow \mathbb{R}$$

Avendo già introdotto la Gaussiana 'in una dimensione' si può pensare ad una Gaussiana in 'due dimensioni' o bivariata, come accennato nell'approfondimento 5.3.

Discutiamo ora separatamente il caso discreto e continuo. Vedremo che nonostante tra le due la notazione sia in un certo senso diversa (nel primo caso si usano sommatorie e nel secondo integrali) i due casi sono completamente analoghi.

5.2 Funzioni di massa di probabilità (PMF)

Una funzione di massa di probabilità, deve soddisfare almeno due requisiti fondamentali:

- Deve essere positiva sul dominio S : $f_X(x_i) \geq 0 \quad \forall x_i \in S$
- La sommatoria delle probabilità su tutti gli x_i deve essere unitaria: $\sum_{i=1}^n f_X(x_i) = 1$

Queste sono le prime due condizioni espresse da Kolmogorov nella descrizione assiomatica della probabilità.

5.2.1 Distribuzione discreta uniforme \mathcal{U}

Supponiamo di avere un evento aleatorio con un numero finito n di esiti E_1, E_2, \dots, E_n , elementare e non, che siano 'equiprobabili'. Eventi di questo tipo sono ad esempio l'uscita di una delle sei facce del dado, o l'uscita di un numero pari o dispari del dado, o l'uscita di testa o croce nel lancio della moneta. Non lo sono invece 'uscita di un numero minore di 2 in un dado' o 'una misura compresa tra $(-1.9, 2.1)$ '. Ad ognuno di questi eventi possiamo assegnare un variato x_1, x_2, \dots, x_n attraverso la costruzione della variabile aleatoria X che avrà distribuzione di tipo Uniforme \mathcal{U} :

Si dice **PMF discreta uniforme** \mathcal{U} la funzione $f_X(x)$ della variabile aleatoria X :

$$f_{\mathcal{U}}(x) = \mathcal{U}(x) : \mathbb{R} \rightarrow [0, 1] \quad f_{\mathcal{U}}(x) = P(x) = \begin{cases} \frac{1}{n} & \text{per } x = x_1, x_2, \dots, x_n \\ 0 & \text{altrimenti} \end{cases} \quad (5.3)$$

La PMF assegna ad ogni variato lo stesso valore di probabilità $1/n$. Vediamo che sembra una ‘buona’ distribuzione di probabilità in quanto tutte le probabilità sono maggiori di zero, e sommando tutte le probabilità relative a ciascun variato si ottiene 1.

E’ molto utile avere una rappresentazione grafica delle distribuzioni di probabilità, attraverso un diagramma in cui sull’asse delle ascisse appaiano i Variati e sull’asse delle ordinate la PMF o la PDF. Riportiamo in [Figura 5.1](#) la rappresentazione grafica per la distribuzione discreta uniforme. Nel caso in figura si hanno 5 variati tra a e b e ciascuno ha probabilità $1/n = 1/5$.

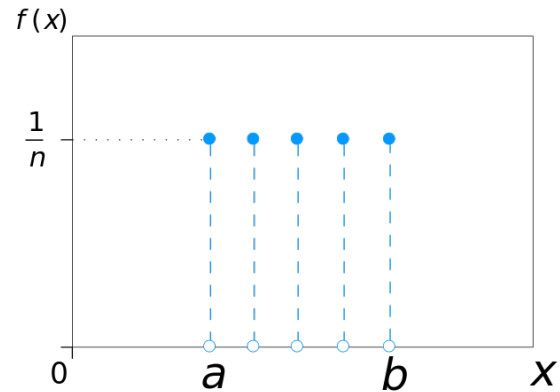


Figura 5.1: Distribuzione di massa di probabilità (PMF) discreta uniforme. Credit: Wikipedia.

Nella definizione [Equazione 5.2.1](#) appare anche la linea $f_X = 0$ ‘altrimenti’. Questa linea logicamente è importante perché certamente esclude ogni altro evento possibile, tuttavia, spesso risulta ovvia e quindi per semplificare la notazione non sarà sempre riportata. Chiudiamo notando che esiste anche il caso distribuzione di probabilità uniforme continua, che sarà trattato in [sottosezione 5.3.1](#).

Esempio:

Scriviamo esplicitamente la PMF per l’evento aleatorio che esca pari nel lancio singolo di un dado a 6 facce:

$$P = \begin{cases} \frac{1}{2} & 1 \leftarrow \text{esce pari} \leftarrow \text{esce } 2 \text{ o } 4 \text{ o } 6 \\ \frac{1}{2} & 0 \leftarrow \text{esce dispari} \leftarrow \text{esce } 1 \text{ o } 3 \text{ o } 4 \\ 0 & \text{altrimenti} \end{cases} \quad (5.4)$$

In questo esempio, i valori 0, 1 sono i variati, ‘esce pari’ ed ‘esce dispari’ sono eventi dello spazio $\mathcal{F}(\Omega)$ e 2, 4, 6 e 1, 3, 5 sono elementi dello spazio degli esiti elementari Ω .

5.2.2 PMF discreta (binomiale) di Bernoulli

Nel caso si abbia un evento aleatorio con al più *due* esiti E_1, E_2 , di norma essi non sono necessariamente equiprobabili. In questo caso la probabilità dei due eventi è descritta da una distribuzione di probabilità detta di Bernoulli (a volta binomiale, anche se la binomiale è più propriamente quella in [sottosezione 5.2.3](#)). In generale è sempre Bernoulliano un evento in cui qualcosa avviene oppure non avviene. La PMF di Bernoulli \mathcal{BE} per una variabile aleatoria X può essere scritta così:

Si definisce **PMF di Bernoulli** la funzione $f_X(x)$ della variabile aleatoria X :

$$f_X(x) : \mathbb{R} \rightarrow [0, 1] \quad f_X(x) = P(x) = \begin{cases} p & \text{per } x = 1 \text{ (evento avviene)} \\ q = 1 - p & \text{per } x = 0 \text{ (evento non avviene)} \end{cases} \quad (5.5)$$

Si noti come il caso di PMF discreta uniforme con due soli variati è anche una PMF di Bernoulli in cui $p = 0.5$. Nel caso di Bernoulli non ha molto senso disegnare il grafico della PMF avendo solo due punti. Si noti comunque che anche in questo caso: $p, q \geq 0$ e $p + q = 1$.

Esempio:

Spesso per uno strumento di misura la cosiddetta *efficienza* che è un termine generico per esprimere una grandezza percentuale. Supponiamo ad esempio di avere uno strumento per rivelare la luce detto fotomoltiplicatore (PMT, <https://it.wikipedia.org/wiki/Fotomoltiplicatore>). Questo strumento, di ampio uso in fisica e non solo, permette di rivelare segnali luminosi molto deboli (fino al singolo fotone) e molto brevi (impulsi anche al di sotto del nanosecondo). Nel fronte del PMT un fotone viene convertito (oppure no) in un elettrone che viene anche detto per questo ‘fotoelettrone’^a. Ora, in questa finestra si ha quindi un evento aleatorio: in alcuni casi l’elettrone effettivamente genera un fotoelettrone, in altri no. L’evento è quindi Bernoulliano. Questa probabilità di conversione è anche detta *efficienza di conversione* $\epsilon = n_e/n_\gamma$. Risulta $0 \leq \epsilon \leq 1$. L’efficienza di un fotomoltiplicatore (come di quasi tutti i sensori di luce) dipende dalla lunghezza d’onda incidente. Quindi



Un fotomoltiplicatore.

$$f_{\text{conversione}} = \begin{cases} \epsilon & \text{per } x = 1 \text{ (avviene conversione } \gamma \rightarrow e^-) \\ 1 - \epsilon & \text{per } x = 0 \text{ (non avviene)} \end{cases} \quad (5.6)$$

Useremo questo esempio in altre applicazioni più avanti.

^aIl fotomoltiplicatore si basa sulla presenza di un fotocatodo, ovvero una finestra di quarzo nella quale avviene la conversione da fotone a elettrone attraverso un processo detto effetto fotoelettrico, che è una delle prime dimostrazioni della natura quantistica delle interazioni elementari.

5.2.3 PMF discreta binomiale \mathcal{B}

Supponiamo di ripetere n -volte un processo di Bernoulli di probabilità p e contare quante volte k osserviamo l’avversarsi del processo. Quale è la probabilità di avere k successi su n ripetizioni?



Si tratta della congiunzione logica di eventi indipendenti. La probabilità di avere (in un certo ordine) k successi con probabilità p su n ripetizioni (ovvero $n - k$ insuccessi di probabilità $q = 1 - p$) e’ il semplice prodotto: $P(n, k) = p^k \cdot (1 - p)^{n-k}$. Ora, nel caso in questione non ci interessa *lordine* in cui avvengono i successi, ma solo il numero numero. La formula appena scritta quindi contiene troppo pochi successi. Se l’ordine non importa dobbiamo considerare il numero di combinazioni di k elementi (successi) presi da un insieme di n elementi senza ripetizione; questo equivale al coefficiente binomiale $\binom{n}{k}$ (si veda [Equazione 4.3](#)). Ecco allora che si per questo tipo di evento aleatorio X si può definire la distribuzione di probabilità binomiale \mathcal{B} :

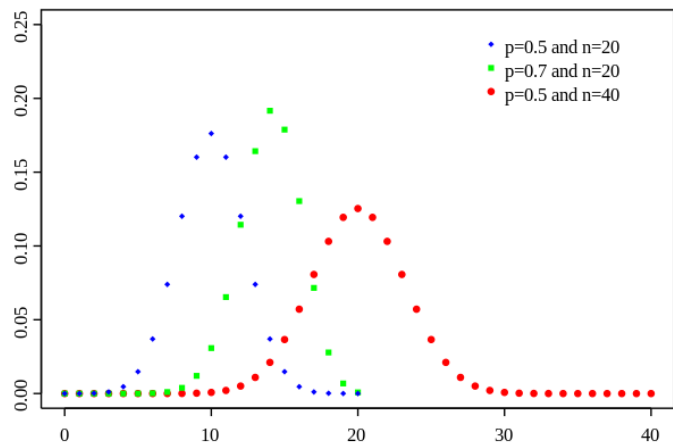


Figura 5.2: Distribuzione di massa di probabilità (PMF) discreta binomiale. Credit: Wikipedia.

Si definisce **PMF discreta binomiale** \mathcal{B} su n ripetizioni la funzione f_X ;

$$f_{\mathcal{B}}(k; n) = \mathcal{B}(k; n) = P(k; n) = \binom{n}{k} p^k (1-p)^{n-k} \quad (5.7)$$

Si noti come in [Equazione 5.7](#), k è il variato di cui si calcola la probabilità e n è un parametro della funzione. Si verifichi anche che per $n = 1$ si riottiene la distribuzione di Bernoulli.

Esempio:

Un sacchetto contiene 10 biglie di cui 3 rosse e le rimanenti bianche. Quale è la probabilità che escano 3 rosse in 4 estrazioni (in cui si rimette la biglia estratta dentro)?

Svolgimento: il fenomeno di estrazione di una biglia rossa dal sacchetto è Bernoulliano con probabilità $p = 3/10 = 0.3$. Questo evento viene ripetuto $n = 4$ volte, ma non ci interessa l'ordine in cui estraiamo le $k = 3$ rosse quindi:

$$P(3; 4) = \binom{4}{3} 0.3^3 0.7^7 = 0.076$$

Diverso sarebbe stato il caso in cui non avessimo riportato la biglia dentro, in quanto allora gli eventi successivi non sarebbero più stati dipendenti. Si provi a calcolare la probabilità in questo caso. Infine, a scopo didattico verifichiamo la probabilità dell'evento complementare, ovvero che 'non escano 3 biglie rosse in 4 lanci'. Qui si tratta di somma logica, allora:

$$\begin{aligned} P(\neg 3; 4) &= P(0; 4) + P(1; 4) + P(2; 4) + P(4; 4) \\ &= 0.2401 + 0.4116 + 0.2646 + 0.0081 \\ &= 0.9244 = 1 - P(3; 4) \end{aligned}$$

Al variare di k e per alcuni valori di p si riporta il grafico della PMF binomiale in [Figura 5.2](#). Si noti come la forma della binomiale, almeno per n sufficientemente grandi sia simile a quella della Gaussiana. Si ricordi in effetti come sia stata costruita la Gaussiana attraverso la combinazione di effetti casuali (vedi approfondimento [3.2](#)). Si può anche vedere come, nel caso $n \gg 1$ and $p \ll 1$ allora la $\mathcal{B} \rightarrow \mathcal{G}$, si veda nell'approfondimento [5.2](#).

La distribuzione di probabilità binomiale è di uso molto frequente, in quando nei vari casi in cui si debbano 'contare' successi rispetto agli insuccessi in un evento fisico, questa è la distribuzione di riferimento.

5.2.4 PMF di Poisson \mathcal{P}

La distribuzione di Poisson è una delle più importanti distribuzioni di probabilità, in quanto un elevato numero di fenomeni fisici sono assimilabili a eventi aleatori appunto di Poisson o Poissoniani.

Un fenomeno aleatorio Poissoniano è un fenomeno:

- Che avviene con una **frequenza media costante** λ (nel tempo, nello spazio, etc.);
- **Raro**, nel senso che la probabilità che si avverino due eventi congiunti (nel tempo, spazio) è nulla¹;
- i cui **esiti sono indipendenti** tra loro.

Ad esempio, supponiamo di mettere un array di contatori Geiger² attorno al punto in cui si fa collidere un fascio di elettroni di alta energia contro un target di piombo. Dalla collisione saranno emesse

¹E' chiaro che non avendo specificato in quanto tempo/spazio è definita la media, è sufficiente prendere un intervallo di tempo/spazio sufficientemente ampio da garantire questa condizione.

²Il contatore Geiger rileva il passaggio di particelle cariche e radiazione di alta energia attraverso attraverso la misura di tensione su un filo conduttore inserito in una camera contenente un gas a bassa pressione https://it.wikipedia.org/wiki/Contatore_Geiger. È stato uno strumento di ampio uso in Fisica (e ne trovate degli esemplari nel Museo della Storia di Fisica del nostro Dipartimento).

in generale un gran numero di particelle elementari come elettroni, positroni e neutrini. Concentriamo per semplicità su elettroni. Il fenomeno di lettura del passaggio di elettroni su un contatore e' Poissoniano: se il fascio accelerato e' mantenuto stabile, ci si aspetta un certo numero medio costante di conteggi nel tempo, ad esempio $\lambda = 4.2$ conteggi/minuto. Inoltre, se il sensore e' abbastanza veloce, e' improbabile che ci siano due conteggi nello stesso tempo. Avendo a disposizione un array di sensori, possiamo anche valutare la distribuzione spaziale. Ipotizzando una certa simmetria, possiamo anche valutare che ci sarà un tasso medio di particelle ad esempio per metro quadro $\lambda' = 12.3$ conteggi/m².

A partire da queste tassi medi costanti, e' possibile costruire la probabilità di ottenere un numero n di conteggi diverso dalla media λ . Allora, se l'evento aleatorio ha le caratteristiche poissoniane viste sopra, allora esso puo' essere descritto da una funzione di massa di probabilità \mathcal{P} detta Poissoniana:

Si definisce la **PMF Poissoniana** $\mathcal{P}(n; \lambda)$ di un evento a media costante λ la probabilità di osservare n conteggi invece che λ . La funzione e':

$$f_{\mathcal{P}}(n; \lambda) = \mathcal{P}(n; \lambda) = P(n; \lambda) = \frac{\lambda^n e^{-\lambda}}{n!} \quad (5.8)$$

Si notino alcuni fatti. Anzitutto il variato in questo caso è n e la poissoniana e' parametrica di parametro λ . Il variato n deve essere un numero naturale, in quanto misura i conteggi assoluti. λ invece puo' essere un numero reale. Si noti anche come il valore numerico del tasso medio dipende dallo specifico valore di tempo/spazio in cui viene calcolato. Sono analoghi infatti $\lambda = 4.2$ conteggi/minuto oppure $\lambda = 0.07$ conteggi/secondo. Tuttavia:

$$\forall n \quad \frac{4.2^n e^{-4.2}}{n!} \neq \frac{0.07^n e^{-0.07}}{n!}$$

Come mai? Bisogna accordare il numero di conteggi n al parametro λ . Si veda l'esempio seguente.

In generale, per evitare equivoci quando applicata a casi fisici, la distribuzione di Poisson puo' essere anche ricordata cosi:

$$f_{\mathcal{P}}(n) = P(n; \lambda t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!} \quad (5.9)$$

dove λ è il tasso medio *per unita' di tempo/spazio/etc* e t è il fattore che considera l'effettivo intervallo richiesto per n .

In [Figura 5.3](#) si mostra la distribuzione di probabilità di Poisson per alcuni valori di λ e in funzione del numero di conteggi n . Anche in questo caso, si noti come per valori di λ sufficientemente alti, la distribuzione di Poisson rassomiglia ad una Gaussiana. La dimostrazione di questo comportamento e' riportata in approfondimento [5.2](#).

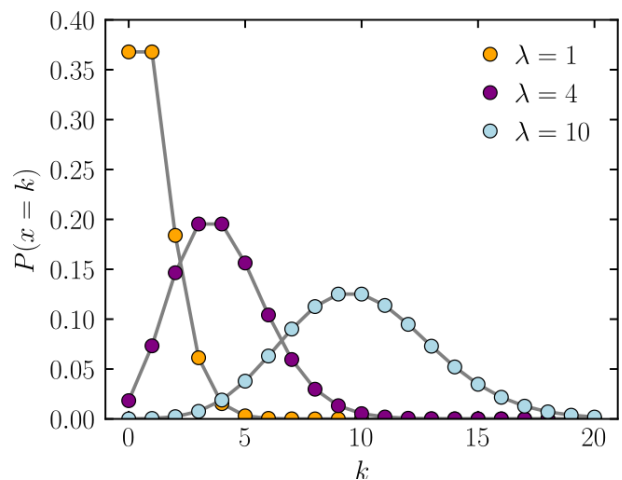


Figura 5.3: Distribuzione di massa di probabilità (PMF) discreta di Poisson. Credit: Wikipedia.

Esempio:

Un fotomoltiplicatore esposto ad una sorgente luminosa costante rileva mediamente 3 fotoni in 2 secondi. Si calcolino le seguenti probabilità: a) probabilità che osservi 3 fotoni in 2 secondi; b) probabilità che osservi 1 fotone in 2 secondi; c) probabilità che osservi 3 fotoni in 3 secondi.

Svolgimento: Si tratta di un evento Poissoniano raro e a media fissa. Anzitutto procediamo con il calcolo del parametro

$$\lambda = \frac{3 \text{ fotoni}}{2 \text{ secondi}}$$

Allora la probabilità di osservare 3 fotoni in 2 secondi puo' essere calcolata:

$$P_{\mathcal{P}}(3; 3) = \frac{3^3 e^{-3}}{3!} \sim 0.224$$

Si noti come il numero n di conteggi debba essere intero. Cio' porta ad alcune considerazioni pratiche. Se da un lato λ è 3 eventi in 2 secondi, si puo' anche affermare che $\lambda = 1.5 \text{ ph s}^{-1}$. Tuttavia, nel conto appena fatto non potremmo calcolare

$$P_{\mathcal{P}}(1.5; 1.5) = \frac{1.5^{1.5} e^{-1.5}}{1.5!}$$

perchè il fattoriale è definito solo per numeri naturali. Dobbiamo quindi sempre utilizzare una media λ valutata in base al conteggio di cui si vuole valutare la probabilità. Ad esempio, la probabilità di osservare 1 fotone in 2 secondi:

$$P_{\mathcal{P}}(1; 3) = \frac{3^1 e^{-3}}{1!} \sim 0.149$$

pero' se ora volessimo calcolare la probabilità di avere 3 fotoni in 3 secondi, dovremmo aggiornare la media: $\lambda' = 4.5$ eventi in 3 secondi e quindi:

$$P_{\mathcal{P}}(3; 4.5) = \frac{4.5^3 e^{-4.5}}{3!} \sim 0.169$$

5.3 Funzioni di densità di probabilità: PDF

Abbiamo visto che nel caso di funzioni di densità di probabilità il variato di una variabile aleatoria è continuo in \mathbb{R} . Essendo continuo, non si puo' calcolare la probabilità la probabilità di un valore specifico $P(x)$, ma solo al piu' per intervalli infinitesimi dx attorno a quel valore: $dP(x)$. Allora in questo caso la funzione appropriata e' detta funzione di densità di probabilità, nel senso che:

$$f_X(x) = \frac{dP(x)}{dx}$$

Le condizioni che deve soddisfare $f_X(x)$ per essere una PDF, analogamente con MPF sono:

- Deve essere positiva sul dominio S : $f_X(x) \geq 0 \quad \forall x \in S$
- L'integrale sul dominio deve essere unitario: $\int_S f_X(x) = 1$

E' da notare che, in virtu' del fatto che si tratta di una densita' di probabilita, localmente il valore della PDF puo' essere maggiore di 1.

5.3.1 PDF continua uniforme \mathcal{U} e PDF continua triangolare \mathcal{T}

. Abbiamo gia' incontrato queste distribuzioni di probabilità quando abbiamo discusso l'incertezza su una misura di tipo digitale e analogica (

Si definisce **PDF Uniforme** $\mathcal{U}(x)$ tra $[a, b]$ la funzione densità di probabilità $f_X(x)$ tale che:

$$f_{\mathcal{U}}(x) = \mathcal{U}(x) = \frac{dP(x)}{dx} = \begin{cases} \frac{1}{|b-a|} & \text{per } x \text{ in } [a, b] \\ 0 & \text{altrimenti;} \end{cases} \quad (5.10)$$

Ad esempio, la distribuzione di probabilità attorno alla misura eseguita con uno strumento digitale segue la distribuzione di probabilità uniforme all'interno della dispersione massima della misura. La distribuzione triangolare continua e' invece quella che puo' essere utilizzata per una misura singola effettuata con strumento analogico. La distribuzione di probabilità triangolare puo' sempre simmetrica o asimmetrica. Qui di seguito ci concentriamo solo su quella simmetrica.

Si definisce **PDF Triangolare Simmetrica** $\mathcal{T}(x)$ tra $[0, 1]$ la funzione densità di probabilità $f_X(x)$ tale che:

$$f_{\mathcal{T}}(x) = \mathcal{T}(x) = \frac{dP(x)}{dx} = \begin{cases} 4x & \text{per } x \text{ in } [0, 1/2] \\ 4(1-x) & \text{per } x \text{ in } [1/2, 1] \\ 0 & \text{altrimenti;} \end{cases} \quad (5.11)$$

Il caso generico $[a, b]$ e il caso non simmetrico sono analoghi ma con una notazione leggermente piu' complessa.

In Fig. [Figura 5.4](#) riportiamo le due PDF.

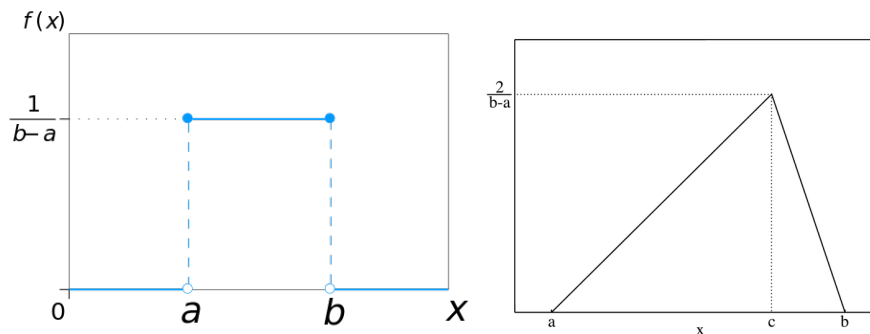


Figura 5.4: Distribuzione di densità di probabilità (PDF) uniforme (sinistra) e triangolare asimmetrica (destra). Credit: Wikipedia.

Esercizio:

Si dimostri che le distribuzioni [Equazione 5.10](#) e [Equazione 5.11](#) soddisfano le proprietà delle distribuzioni di probabilità.

5.3.2 PDF Esponenziale

La distribuzione esponenziale è la distribuzione di probabilità della *durata* di un particolare tipo di fenomeni detti *senza memoria*, ovvero fenomeni in cui una situazione futura è sempre predicibile (in termini di probabilità) dalla situazione ad un tempo dato, senza necessità di conoscere il passato. Ad esempio, se prendiamo un processo di decadimento di nuclei radioattivi di vita media τ risulta che a) il numero di nuclei dopo un tempo pari a τ si ridurra al $1/e \sim 37\%$ ovvero che b) la probabilità di trovare una particella radioattiva *dopo* un tempo τ rispetto al tempo di riferimento si attesterà al $1/e \sim 37\%$. Questi fenomeni senza memoria non sono altro che fenomeni in sé Poissoniani, e quindi la distribuzione esponenziale si può anche pensare come la distribuzione di probabilità dei *tempi tra eventi* di un processo di questo tipo.

Funzione di Densità di Probabilità Esponenziale

Sia λ un parametro fissato, allora si dice PDF esponenziale la funzione

$$f_X(x) = \frac{dP_X(x)}{dx} = \begin{cases} \lambda e^{-\lambda x} & \text{per } x \geq 0 \\ 0 & \text{altrimenti;} \end{cases} \quad (5.12)$$

La PDF esponenziale è rappresentata in [Figura 5.6](#) per alcuni valori di λ .

Se la distribuzione di Poisson risponde alla domanda ‘quale e’ la probabilità di avere n conteggi se il tasso medio di conteggi è λ , l’esponenziale risponde alla domanda ‘qual e’ la probabilità di osservare un conteggio dopo x (secondi, metri, etc.)?’ Ad esempio, supponiamo di prendere un nucleo isotopico radioattivo, ad esempio ^{235}U , l’isotopo usato per arricchire l’uranio naturale. Esso ha una emivita di $\tau \sim (7 \times 10^8)$ anni. Cio’ significa che *in media* si osserva il decadimento ogni τ anni, ovvero in media $\lambda = 1/\tau = 1.4 \times 10^{-9}$ dec yr^{-1} .

L’effetto memoria si può dimostrare: dopo un intervallo $\tau = 1/\lambda$ il valore della funzione è diminuito di un fattore $1/e \sim 0.37$ *indipendentemente da* x :

$$\frac{f_X(x + \tau)}{f_X(x)} = \frac{1}{e} \sim 0.37$$

In effetti, spesso la distribuzione esponenziale viene scritta come

$$f_X(x) = \frac{1}{\tau} e^{-x/\tau}$$

Se τ è un intervallo di tipo temporale viene anche detto *vita media* o *emivita*. Se l’intervallo e’ di tipo spaziale allora τ e’ il cammino libero medio o concetti simili.

Continuiamo con l’esempio di uranio. Se prendiamo un blocco di materiale in cui e’ presente ^{235}U , il numero di nuclei radioattivi decade con emivita τ , allora $N(t) = N_0 e^{-t/\tau}$ ma in maniera probabilistica, nel senso che $P(N(t)) = P(0) e^{-t/\tau} = \lambda e^{-t/\tau}$

Vediamo ora la relazione con Poisson. Se l’esponenziale di parametro λ rappresenta il numero di conteggi dopo un certo intervallo di tempo trascorso, la PDF di Poisson misura la probabilità di avere un certo numero di conteggi in un certo intervallo di tempo. Il legame tra le due è quindi:

$$P_{\mathcal{P}}(n; \lambda t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!} \iff P(t) = \lambda e^{-\lambda t}$$

infatti, nella PDF poissoniana, la probabilità di contare 0 decadimenti nell’intervallo $[0, t]$ è:

$$P_{\mathcal{P}}(0; \lambda t) = e^{-\lambda t}$$

Nella PDF esponenziale, la probabilità di avere conteggi tra $[t, \infty]$ è:

$$P(t < t' < \infty) = \int_t^{\infty} \lambda e^{-\lambda t'} dt' = [-e^{-\lambda t}]_t^{\infty} = e^{-\lambda t}$$

e quindi le due probabilità esprimono lo stesso fenomeno. In sostanza, per un evento Poissoniano di tasso medio λ , il tempo medio tra eventi (anche non consecutivi) è $\tau = 1/\lambda$ e il tempo tra eventi segue la distribuzione esponenziale di parametro λ .

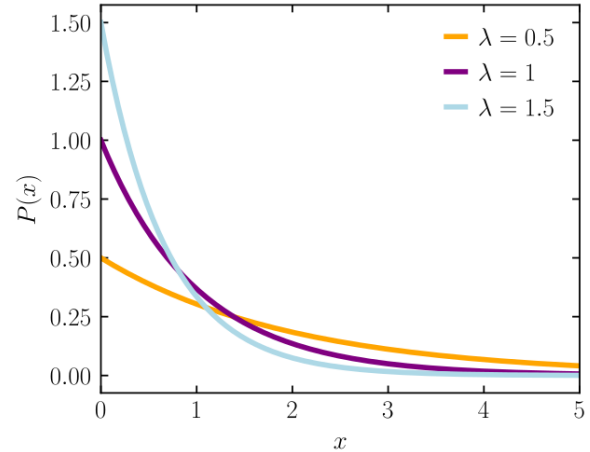


Figura 5.5: Distribuzione di densità di probabilità (PDF) esponenziale. Credit: Wikipedia.

5.3.3 PDF Gaussiana

La funzione di densità di probabilità Gaussiana già introdotta in precedenza, è la prima che osserviamo dipendere non da uno ma da due parametri μ, σ . Essa rappresenta la distribuzione di probabilità delle incertezze casuali, ed è per questo anche chiamata *curva degli errori accidentali*. Essa inoltre è la funzione limite secondo il teorema del limite centrale (Appr. 3.3). Viene anche detta *curva normale* per questa ragione, o *a campana* per la sua forma. Si indica spesso in forma contratta come $\mathcal{G}(x; \mu, \sigma^2), \mathcal{N}(x; \mu, \sigma^2)$.

Funzione di Densità di Probabilità Gaussiana

Dati μ, σ , allora si dice PDF Gaussiana la funzione

$$f_{\mathcal{G}}(x; \mu, \sigma) = \mathcal{G}(x; \mu, \sigma^2) = \frac{dP(x)}{dx} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (5.13)$$

Si usa molto spesso anche la **PDF Gaussiana Standard** con $\mu = 0, \sigma = 1$

$$f_{\mathcal{G}}(x; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad (5.14)$$

Si ricordano qui alcune proprietà già viste: la funzione è simmetrica attorno a μ , definita su tutto \mathbb{R} , μ è anche il punto di massimo, che vale $1/(\sqrt{2\pi}\sigma)$. Il parametro σ individua la ampiezza della curva e la posizione dei due flessi $\mu \pm \sigma$.

Vedremo invece che la Gaussiana standard $\mathcal{G}(0, 1), \mathcal{N}(0, 1)$ tornerà molto spesso in quanto è sempre possibile *standardizzare* una PDF gaussiana per la variabile X in una PDF gaussiana standard per la Variabile X' con l'operazione

$$X \longrightarrow X' = \frac{X - \mu}{\sigma}$$

Abbiamo visto che la PDF binomiale e quella di Poisson in alcune condizioni via via approssimano una PDF Gaussiana (appr. 5.2). Se queste condizioni sono raggiunte, allora la Gaussiana di approssimazione può essere ottenuta sostituendo:

$$\underbrace{\mu = np; \quad \sigma = np(1-p)}_{\text{binomiale}} \quad \underbrace{\mu = \lambda; \quad \sigma = \sqrt{\lambda}}_{\text{poissoniana}}$$

5.3.4 Altre PDF

Il numero di PDF va ben oltre quelle che abbiamo visto. Non ha gran senso farne una lunga elencazione, si ricordi però che esse possono essere raggruppate in due categorie:

- PDF rappresentanti fenomeni fisici aleatori. Abbiamo visto che le PDF rappresentano fenomeni fisici aleatori come la distribuzione delle misure ripetute, il numero di nuclei di decadimento, etc. A seconda del caso fisico, avrete bisogno di usare diverse distribuzioni di probabilità. Tra quelle più note troviamo la *log-normale* che descrive tutti quei fenomeni formati da diversi fattori che si accumulano percentualmente, la *Breit-Wigner o Cauchy* che descrive le risonanze tra particelle elementari, la *Landau* etc. Il lettore più curioso può approfondire in questa raccolta https://en.wikipedia.org/wiki/Probability_distribution
- PDF che rappresentano *test statistici* ovvero che sono costruite a partire dai dati sperimentali per la verifica di ipotesi. Fanno parte di questa categoria ad esempio la PDF del χ^2 quella di *Student*, tra le altre. In un certo senso, anche la Gaussiana fa parte di questa categoria.

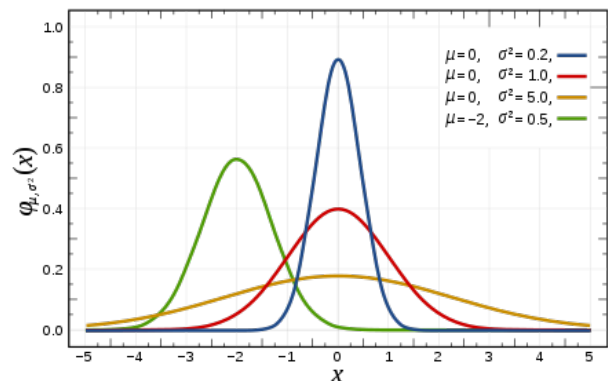


Figura 5.6: Distribuzione di densità di probabilità (PDF) Gaussiana. Credit: Wikipedia.

5.4 Funzioni cumulative di probabilità

Per una funzione di massa o di densità di probabilità, si può costruire la corrispondente *funzione cumulativa di distribuzione di probabilità* (Cumulative Distribution Function CDF) o *funzione distribuzione di probabilità*.

Funzione distribuzione cumulativa di probabilità CDF di una variabile aleatoria X discreta o continua la funzione:

$$F_X(x) = \begin{cases} \sum_{x_{min}}^x f_X(x_i) & \text{nel caso } X \text{ discreta} \\ \int_{x_{min}}^x f_X(x) dx & \text{nel caso } X \text{ continua} \end{cases} \quad (5.15)$$

dove $f_X(x)$ è la funzione di massa o densità di probabilità della Variabile X e x_{min} il valore minimo assunto dal Variato x .

La CDF di una PDF continua è quindi il suo integrale indefinito. La CDF di una PMF è una sommatoria. La CDF valutata in x , corrisponde quindi alla probabilità che il variato assuma un valore minore di un x :

$$F_X(x) = P(X < x)$$

È importante notare come nel caso discreto la CDF sommi un certo numero di probabilità, mentre nel caso continuo la CDF integri un certo intervallo di densità di probabilità.

L'uso delle CDF è molto importante nel calcolo degli intervalli di confidenza delle PDF e nel calcolo delle probabilità per intervalli, come vedremo di seguito. Tuttavia, non riportiamo la forma esplicita delle CDF per tutte le PMF e PDF viste in precedenza per due ragioni principalmente: a) non ne faremo sempre uso per ognuna di quelle, e non vogliamo appesantire la discussione, b) in diversi casi, o il calcolo della CDF è piuttosto semplice, e quindi si può svolgere ogni volta che serve, oppure è decisamente complesso (soprattutto per le PDF), dovuto al fatto che l'integrale di una curva può risultare matematicamente complesso in certi casi. Qui di seguito vedremo solo un caso di interesse, quello della CDF della Gaussiana. Altri casi saranno trattati come esercizi.

Esempio:

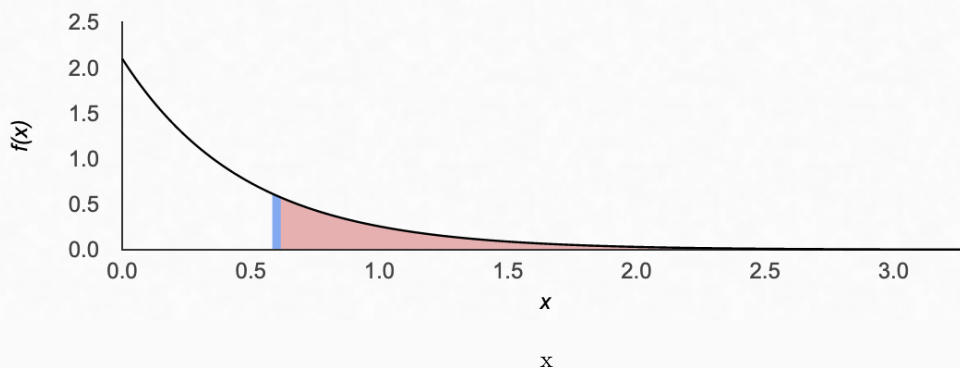
Si trovi la CDF di una distribuzione esponenziale con $\lambda = 2.1$. Si calcoli poi la probabilità che $x > 0.6$

La PDF esponenziale $f_X(x) = \lambda e^{-\lambda x}$ definita su tutto \mathbb{R} . Allora $x_{min} = -\infty$. Ecco che:

$$F_X(x) = \int_{-\infty}^x \lambda e^{-\lambda x'} dx' = \left[-e^{-\lambda x'} \right]_{-\infty}^x = 1 - e^{-\lambda x}$$

A partire dalla CDF è ora molto semplice calcolare le probabilità. La probabilità che $x > 0.6$ è infatti:

$$P(x > 0.6) = 1 - F(0.6) = e^{-3 \cdot 0.6} \sim 0.28$$



5.4.1 CDF Gaussiana

La funzione di distribuzione cumulativa Gaussiana trova un uso più frequente. L'integrale si calcola attraverso una funzione detta *funzione di errore* (*error function*), i cui valori si trovano spesso tabulati, oppure è risolvibile tramite calcolatore.

Funzione cumulativa di distribuzione di probabilità Gaussiana

Per una Gaussiana di parametri μ, σ e per una Gaussiana standard la CDF risulta

$$F_G(x) = \begin{cases} \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x-\mu}{\sqrt{2}\sigma} \right) \right] & \text{per } \mathcal{G}(x; \mu, \sigma^2) \\ \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x}{\sqrt{2}} \right) \right] & \text{per } \mathcal{G}(x; 0, 1) \text{ standard} \end{cases} \quad (5.16)$$

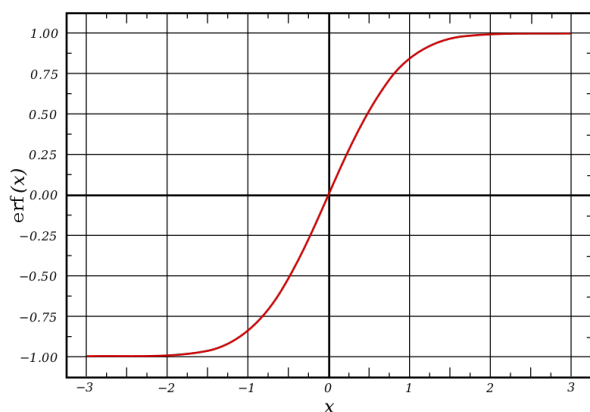


Figura 5.7: Grafico della funzione dell'errore. Si consideri la variabile z e non x secondo la nostra notazione. Credit Wikipedia.

Anzitutto riportiamo in [Figura 5.7](#) il grafico della funzione degli errori. Come si può vedere la funzione è antisimmetrica ($\operatorname{erf}(-z) = -\operatorname{erf}(z)$). Come detto, il calcolo di

$$\operatorname{erf}(z)$$

è numerico. La sua definizione esatta

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$$

Notiamo anzitutto anche questa utile proprietà algebrica. Se vogliamo calcolare la probabilità che il variato sia tra $\mu - \delta$ e $\mu + \delta$ osserviamo che vale:

$$\int_{\mu-\delta}^{\mu+\delta} \mathcal{G}(x) dx = F_G(\mu+\delta) - F_G(\mu-\delta) = \operatorname{erf} \left(\frac{\delta}{\sqrt{2}\sigma} \right) \quad (5.17)$$

Esercizio:

Dimostrare la [Equazione 5.17](#).

$\delta = x - \mu$	$z = \frac{x-\mu}{\sqrt{2}\sigma}$	$\operatorname{erf}(z)$	$1-\operatorname{erf}(z)$
1σ	$1\sqrt{2} = 0.71$	0.683	
2σ	1.414	0.955	
3σ	2.121	0.997	0.03
4σ	2.828		$6.3 \cdot 10^{-5}$
5σ	3.536		$5.7 \cdot 10^{-7}$
1.96σ	1.386	0.90	0.10
2.58σ	1.824	0.99	0.01
3.29σ	2.326	0.999	0.001

Tabella 5.1: Alcuni valori di $\operatorname{erf}(z)$. La sezione in alto rappresenta valori multipli di σ , l'ampiezza di una gaussiana, mentre in basso l'attenzione è su valori 'interi' di probabilità.

Si riportano in [Tabella 5.1](#) e ?? alcuni dei valori tabulati³. Un occhio attento può forse aver notato che i valori della colonna di destra corrispondono ai livelli di confidenza per $1, 2, 3\sigma$ già visti per la Gaussiana. Ed è proprio questa la ragione del nome 'errore' per questa funzione. In particolare riveste un ruolo importante il valore corrispondente a 5σ come descritto in [Appr. 5.4](#). Questo valore è infatti usato come valore di riferimento in molti esperimenti fisici per stabilire una *scoperta*.

³Si veda anche https://en.wikipedia.org/wiki/Error_function#Table_of_values

Tabulation of Error Function Values

z	$\text{erf}(z)$	z	$\text{erf}(z)$	z	$\text{erf}(z)$
0	0	0.55	0.5633	1.3	0.9340
0.025	0.0282	0.60	0.6039	1.4	0.9523
0.05	0.0564	0.65	0.6420	1.5	0.9661
0.10	0.1125	0.70	0.6778	1.6	0.9763
0.15	0.1680	0.75	0.7112	1.7	0.9838
0.20	0.2227	0.80	0.7421	1.8	0.9891
0.25	0.2763	0.85	0.7707	1.9	0.9928
0.30	0.3286	0.90	0.7970	2.0	0.9953
0.35	0.3794	0.95	0.8209	2.2	0.9981
0.40	0.4284	1.0	0.8427	2.4	0.9993
0.45	0.4755	1.1	0.8802	2.6	0.9998
0.50	0.5205	1.2	0.9103	2.8	0.9999

Figura 5.8: Tabulato dei valori di $\text{erf}(z)$

Esercizio:

Si calcoli la probabilità di trovare un valore al di fuori dell'intervallo di 3σ attorno al valore medio di una gaussiana usando la CDF.

5.5 Momenti delle distribuzioni

Le distribuzioni di probabilità possono essere studiate attraverso la loro definizione formale $f_X(x)$ o la loro distribuzione cumulativa $F_X(x)$. Esiste tuttavia una ulteriore maniera di caratterizzare le distribuzioni di probabilità, ovvero attraverso i loro momenti. Matematicamente, i momenti sono calcolati attraverso delle particolari medie sulla distribuzione, detti 'valori di aspettazione'. In generale, a livello matematico, la media dei valori di x su un intervallo (a, b) pesati con una funzione $f(x)$ e' pari a (analogo caso discreto non riportato):

$$\frac{\int_a^b x f(x) dx}{\int_a^b f(x) dx}$$

ma nel caso delle PDF, se l'integrazione avviene su tutto il dominio, il denominatore e' unitario per definizione. Nel caso dei momenti non faremo necessariamente la media dei valori di una funzione, ma a volte anche degli 'scarti' dei valori della funzione rispetto ad un valore 'c', o addirittura di una potenza di essi. Quindi una

definizione piu' generale di momenti di $f_X(x)$ di ordine n rispetto a c e':

$$m_n = E[(X - c)^n] = \begin{cases} \int (x - c)^n f_X(x) dx & \text{caso } f_X \text{ continua} \\ \sum (x - c)^n f_X(x) & \text{caso } f_X \text{ discreta} \end{cases} \quad (5.18)$$

dove abbiamo usato l'operatore matematico E (*valore di aspettazione (expected value)* o *speranza matematica*) per esprimere questa operazione. I momenti del tipo [Equazione 5.18](#) sono detti *momenti centrali di ordine n* e li indichiamo anche con il simbolo m_n . Nel caso si prenda il modulo $|x - c|^n$ sono detti momenti centrali di ordine n in modulo. Con questa notazione anticipiamo che di solito $c = E[X]$ o $c = 0$. I momenti del tipo $c = 0$

$$\mu_n = E[X^n] = \begin{cases} \int x^n f_X(x) dx & \text{caso } f_X \text{ continua} \\ \sum x^n f_X(x) & \text{caso } f_X \text{ discreta} \end{cases} \quad (5.19)$$

sono detti *momenti algebrici di ordine n*. Nel caso si prenda il modulo $|x|$ sono detti momenti algebrici in modulo. Li indichiamo con il simbolo compatto μ_n . Non tutti i momenti di tutti gli ordini sono importanti allo stesso modo. In particolare lo sono il primo momento algebrico μ_1 ; il secondo, terzo e quarto momento centrale m_2, m_3, m_4 . Per la loro importanza essi hanno assunto nomi specifici come media, varianza, asimmetria e curtosi rispettivamente. Cominciamo.

5.5.1 Media di una PMF/PDF.

Si dice *media (mean)* (o anche valore di aspettazione o speranza matematica senza specificare l'ordine), di una distribuzione di probabilità f_X , la seguente grandezza:

$$\mu = E[X] = \begin{cases} \sum_{i=1}^n x_i f_X(x_i) & \text{caso } f_X \text{ discreta} \\ \int x f_X(x) dx & \text{caso } f_X \text{ continua} \end{cases} \quad (5.20)$$

Dove l'integrale si estende al dominio della f_X

Come mai si chiama 'media' questo valore? Osserviamolo in dettaglio. Il caso discreto forse e' piu' semplice da visualizzare, ma il caso continuo e' solo una sua estensione. Si stanno sommando i prodotti tra il valore del Variato x_i e il valore della probabilità del Variato $P_i = f_X(x_i)$ (caso discreto). Proviamo ora a riprendere la definizione di media aritmetica, essa e' definita come:

$$\frac{x_1 + x_2 + \dots + x_N}{N} = \frac{1}{N}x_1 + \frac{1}{N}x_2 + \dots + \frac{1}{N}x_N = \sum_{i=1}^n x_i \frac{1}{N}$$

La formula sopra e' analoga a [Equazione 5.24](#) prendendo $f_X = 1/N$ ovvero una distribuzione discreta uniforme di probabilità. Nel caso della media aritmetica di N misure infatti non ne preferiamo una rispetto all'altra, sono tutte equiprobabili per il calcolo della media aritmetica. Inoltre, nel caso della media pesata (??) si noti come si abbiano invece pesi diversi.

Una interessante proprietà del valore di aspettazione e' che e' **la speranza matematica e' l'unico parametro per che azzera gli scarti rispetto al valore stesso**. In formula:

$$E[(X - E[X])] = \int (x - E[X]) f_X(x) dx = 0 \quad (5.21)$$

Infatti $E[(X - E[X])] = \int x f_X(x) dx - E[X] \int f_X(x) dx = E[X] - E[X] = 0$ ■.

Calcoliamo i valori di aspettazione delle PMF/PDF piu' importanti.

Valore di aspettazione di PMF di Poisson \mathcal{P} Dalla PMF Poissoniana $\mathcal{P} f_{\mathcal{P}}(n; \lambda) = \lambda^n e^{-\lambda}/n!$, applicando la definizione:

$$\begin{aligned} E[\mathcal{P}] &= \sum_{n=0}^N n \frac{\lambda^n e^{-\lambda}}{n!} \\ &\text{Considerando che } n! = n(n-1)! \text{ e lo sviluppo in serie di Taylor di } e^\lambda \\ &= \sum_{n=1}^N n \frac{\lambda^n e^{-\lambda}}{n!} = e^{-\lambda} \sum_{n=1}^N \frac{\lambda^{n-1} \lambda}{(n-1)!} = \lambda e^{-\lambda} e^\lambda \\ &= \lambda \end{aligned}$$

Allora il valore di aspettazione, ovvero il valore medio della poissoniana e' $E[\mathcal{P}] = \lambda$. Ora, il parametro λ di una Poissoniana e' anche il tasso medio di conteggi nell'intervallo di riferimento. Ecco che il risultato ci torna.

Valore di aspettazione di PDF Uniforme Dalla definizione della PDF uniforme in $[a, b]$ \mathcal{U} , $f_{\mathcal{U}}(x) = 1/(b - a)$ otteniamo

$$E[\mathcal{U}] = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \underbrace{\left[\frac{x^2}{2} \right]_a^b}_{b^2 - a^2 = (a+b)(a-b)} = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2} \quad (5.22)$$

E di nuovo abbiamo dimostrato che $E[\mathcal{U}] = (a + b)/2$ e' il valore medio dell'intervallo. Si noti che qui la distribuzione e' uniforme, quindi tutti i valori sono equiprobabili. Ciononostante, il valore di aspettazione e' il valore medio centrale.

Valore di aspettazione di PDF Gaussiana Analogamente a quanto fatto sopra

$$\begin{aligned} E[\mathcal{G}] &= \int x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \\ &\quad \text{cambio Variabile } z = x - \mu \text{ e } dz = dx \\ &= \int (z + \mu) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{z}{\sigma}\right)^2} dz = \int z \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{z}{\sigma}\right)^2} dz + \mu \int \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{z}{\sigma}\right)^2} dz \\ &\quad \text{primo termine nullo perche' funzione dispari, secondo termine e' } \mathcal{G}(0, \sigma^2) \\ &= \mu \end{aligned} \quad (5.23)$$

E di nuovo troviamo come il valore di aspettazione sia il valore medio.

Non ci dilungheremo ulteriormente sulle dimostrazioni dei valori di aspettazione per altre distribuzioni di probabilità. Alcune dimostrazione sono riportate in Appr. 5.1. I valori di aspettazione delle principali PMF/PDF sono inoltre raccolti in [Tabella 5.2](#).

Si noti anche che spesso per il valore di aspettazione di una qualsiasi PDF si usa il simbolo $\mu = E[X]$. Si prenda attenzione a non confondere questo simbolo con il parametro μ della distribuzione di probabilità Gaussiana. Solo nel caso della Gaussiana infatti i due parametri coincidono.

5.5.2 Varianza di una PMF/PDF

Si dice *Varianza (Variance)* di una distribuzione di probabilità f_X , il secondo momento centrale rispetto alla media, ovvero il valore di aspettazione degli scarti quadrati rispetto alla media:

$$\text{VAR}[X] = E[(X - E[X])^2] = \begin{cases} \sum_{i=1}^n (x_i - E[X])^2 f_X(x_i) & \text{caso } f_X \text{ discreta} \\ \int (x - E[X])^2 f_X(x) dx & \text{caso } f_X \text{ continua} \end{cases} \quad (5.24)$$

Dove l'integrale si estende al dominio della f_X

La Varianza e' quindi per costruzione il valore medio del quadrato degli 'scarti' rispetto al valore di aspettazione $E[X]$. Abbiamo visto come il valore medio degli scarti rispetto al valore di aspettazione sia infatti nullo e quindi la Varianza e' il primo momento centrale non nullo. Genericamente la Varianza indica la ampiezza della distribuzione rispetto al valore medio. Vedremo piu' avanti la relazione con la incertezza della misura.

Varianza della Gaussiana Applicando la definizione e considerando che $E[X] = \mu$

$$\begin{aligned} \text{VAR}[\mathcal{G}] &= \int (x - \mu)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \\ &\quad \text{cambio Variabile } z = \frac{x - \mu}{\sigma} \text{ e } dz = dx/\sigma \text{ e poi considero integrale di Gauss} \\ &= \int (z\sigma)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}z^2} \sigma dz = \frac{\sigma^2}{\sqrt{2\pi}} \int z^2 e^{-\frac{1}{2}z^2} dz = \frac{\sigma^2}{\sqrt{2\pi}} \sqrt{2\pi} \\ &= \sigma^2 \end{aligned} \quad (5.25)$$

La Varianza e' quindi equivalente al quadrato del parametro σ , che ricordiamo identificare i punti di flesso della Gaussiana.

Si lascia alle appendici la dimostrazione di altre varianze. I risultati sono pero' raccolti in Tab. [Tabella 5.2](#). Sono lunghi ma non complessi i passaggi per il calcolo del valore di aspettazione e della varianza della distribuzione binomiale (www.riani.it), e della varianza di una poissoniana (https://proofwiki.org/wiki/Variance_of_Poisson_Distribution), e triangolare.

Esercizio:

Si derivino per esercizio le varianze delle distribuzioni di Bernoulli, uniforme, esponenziale. Soluzioni Appr. [5.1](#)

La varianza possiede una proprietà matematica interessante, che ci tornerà utile in alcune dimostrazioni e per alcuni calcoli formali. Vale infatti:

$$\text{VAR}[X] = \text{E}[X^2] - (\text{E}[X])^2 \tag{5.26}$$

ovvero la varianza di X e' la differenza tra il valore di aspettazione E del quadrato di una variabile aleatoria X^2 e il quadrato del valore di aspettazione di quella variabile. Infatti (dimostriamo solo per il caso continuo):

$$\begin{aligned} \text{VAR}[X] &= \int (x - \mu)^2 f(x) dx = \int (x^2 - 2\mu x + \mu^2) f(x) dx \\ &= \int x^2 f(x) dx - 2\mu \int x f(x) dx + \mu^2 \int f(x) dx = \text{E}[X^2] - 2\mu \text{E}[X] + \mu^2 \\ &= \text{E}[X^2] - (\text{E}[X])^2 \quad \blacksquare \end{aligned} \tag{5.27}$$

Nel caso il Variato sia espresso nella grandezza $[M]$ ad esempio metri o secondi, la Varianza risulta essere espressa in m^2, s^2 . Per questa ragione si definisce:

Deviazione Standard di una PMF/PDF

Si dice deviazione standard di una PMF/PDF la radice quadrata della Varianza di quella PMF/PDF e si indica con la lettera σ :

$$\sigma_X := \sqrt{\text{VAR}[X]}$$

In virtu' del fatto che la deviazione standard di una distribuzione di probabilità si indica con la lettera σ , spesso la varianza viene indicata con il simbolo

$$\text{VAR}[X] = \sigma_X^2$$

Si faccia attenzione a non confondere questo simbolo con il parametro σ di una PDF gaussiana. Solo nel caso gaussiano le due grandezze coincidono!

5.5.3 Momenti di ordine superiore

Finora abbiamo visto il primo momento algebrico (speranza matematica) e i primi due momenti centrali (il primo e' nullo e il secondo e' la Varianza). Abbiamo visto che la Varianza misura la distribuzione dei dati attorno al valore medio. Ci possiamo chiedere ora se ha senso valutare anche ordini superiori. Il secondo momento algebrico e' la media geometrica, ma di rado si calcola in questo contesto generico. Invece si usano maggiormente:

PMF/PDF	$f_X(x)$	Media $\mu = E[X]$	Varianza $s\sigma_X^2 = \text{VAR}[X]$	SE σ_X
Gaussiana	$\mathcal{G}(\mu, \sigma)$	μ	σ^2	σ
Uniforme	$\frac{1}{b-a}$	$\frac{b+a}{2}$	$\frac{(b-a)^2}{12}$	$\frac{b-a}{\sqrt{12}}$
Triang. simm		$\frac{b+a}{2}$	$\frac{(b-a)^2}{24}$	$\frac{b-a}{\sqrt{24}}$
Bernoulli	$p, 1-p$	p	$p(1-p)$	$\sqrt{p(1-p)}$
Binomiale	$\binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$	$\sqrt{np(1-p)}$
Poisson	$\frac{\lambda^n e^{-\lambda}}{n!}$	λ	λ	$\sqrt{\lambda}$
Esponenziale	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{1}{\lambda}$
Chi quadro χ_k^2	-	k	$2k$	$\sqrt{2k}$
Student t_k	-	0	$\frac{k}{k-2}$ se $k > 2$	$\sqrt{\frac{k}{k-2}}$ se $k > 2$

Tabella 5.2: Media, Varianza e Deviazione Standard per alcune PMF/PDF.

- Terzo momento centrale $m_3 = E[(X - E[X])^3]$ in quanto e' collegato alla asimmetria di una distribuzione, e
- Quarto momento centrale $m_4 = E[(X - E[X])^4]$ in quanto e' collegato alla curtosi di una distribuzione.

Asimmetria (Skewness) Una distribuzione di probabilità puo' essere piu' o meno asimmetrica o non esserlo. Ad esempio, la Gaussiana e' simmetrica, cosi' come la distribuzione di Student e la uniforme. Non lo posso la Poissoniana (almeno per valori bassi di λ), la binomiale (almeno per valori bassi di np), l'esponenziale o il χ^2 (almeno per valori bassi del numero di gradi di liberta'). In Fig. [Figura 5.9](#) si riporta un visualizzazione di distribuzioni asimmetriche.

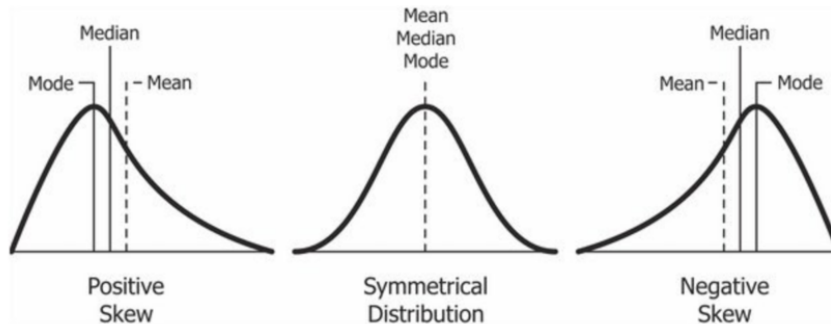


Figura 5.9: Esempio di distribuzioni con asimmetria positiva o negativa. Credit: Wikipedia.

Una stima di asimmetria si fa a partire dal momento m_3 , che non scriviamo esplicitamente. In realta', essendo m_3 un numero con unita' di misura il cubo della misura di X si preferisce calcolare il valore standardizzato:

$$\widetilde{m}_3 = E \left[\left(\frac{X - \mu}{\sigma} \right)^3 \right] = \frac{\mu_3}{\sigma^3} = \frac{E[(X - \mu)^3]}{(E[(X - \mu)^2])^{3/2}} = \frac{m_3}{m_2^{3/2}}$$

Si presti attenzione al fatto che mentre un coefficiente \widetilde{m}_3 positivo o negativo indica asimmetria, un coefficiente nullo potrebbe non indicare asimmetria in qualche caso patologico. Tuttavia, una distribuzione simmetrica ha sempre $\widetilde{m}_3 = 0$. Ad esempio, una distribuzione uniforme e' simmetrica $\widetilde{m}_3 = 0$, una distribuzione di Poisson $\widetilde{m}_4 = \lambda^{-1/2}$, etc. In questo corso ci interessa relativamente poco quantificare la asimmetria, ma useremo nel caso il linguaggio in maniera qualitativa. Evitiamo ulteriori esempi nel caso di distribuzioni e torneremo sull'argomento nel caso di campioni di misure, vedi cap 7.

Curtosi, ed eccesso di Curtosi (Curtosis) Per curtosi si intende invece una grandezza legata al rapporto tra le code di una distribuzione e la sua altezza di una distribuzione, come mostrato in Fig. [Figura 5.10](#). La Gaussiana ha curtosi 3 per questo di solito si parla di eccesso di curtosi rispetto a questo valore. Il coefficiente di curtosi si stima attraverso il quarto momento centrale standardizzato:

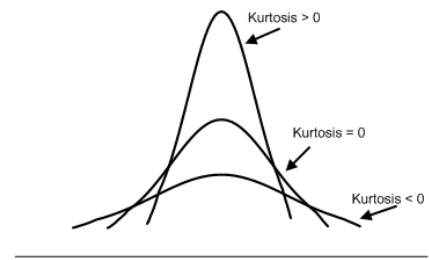


Figura 5.10: Esempi di distribuzioni con curtosi in eccesso

$$\widetilde{m}_4 = E \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right] = \frac{E [(X - \mu)^4]}{(E [(X - \mu)^2])^2} = \frac{m_4}{\sigma^4}, \quad (5.28)$$

Le distribuzioni curtotiche hanno nomi particolari che qui non riportiamo. Una distribuzione ha curtosi superiore a 3 (eccesso positivo di curtosi) se ha le code piu' strette relativamente alla punta, e ha curtosi inferiore a 3 (deficit di curtosi o eccesso negativo di curtosi) se ha le piu' larghe rispetto alla punta. Ad esempio, una distribuzione uniforme ha eccesso di curtosi $\widetilde{m}_4 = -1.2$, la distribuzione di Poisson $\widetilde{m}_4 = \lambda^{-1}$, etc. In questo corso ci interessa relativamente poco quantificare la curtosi, ma useremo nel caso il linguaggio in maniera qualitativa.

5.6 Quantili delle distribuzioni

Spesso diventa interessante dividere una distribuzioni di probabilità in intervalli equiprobabili. Cio' viene fatto attraverso l'uso dei quantili. Se il numero di suddivisioni e' k allora in ciascuna suddivisione rientreranno $1/k$ eventi. Indicando i valori che separano separano questi intervalli con Q_1, Q_2, \dots, Q_k allora risulta:

$$\int_{-\infty}^{Q_1} f_X(x)dx = \int_{Q_1}^{Q_2} f_X(x)dx = \dots = \int_{Q_{k-1}}^{+\infty} f_X(x)dx = 1/k$$

A partire da questo si costruiscono i quantili di ordine α ($0 \leq \alpha \leq 1$) o in percentuale ($0 \leq \alpha \leq 100$) detti anche q_α in maniera che

$$\int_{-\infty}^{q_\alpha} f_X(x)dx = F(q_\alpha) = \alpha$$

Ad esempio, il quantile di ordine $\alpha = 0.1$ è il valore tale che il 10% di valori sia al di sotto.

Siccome i quantili dividono il dominio della funzione di probabilità in intervalli non della stessa lunghezza (tranne che per la PMF/PDF uniforme), allora in generale ad esempio $q_{50} - q_{25} \neq q_{75} - q_{50}$. Questi intervalli sono detti *interquantili*. Abbiamo gia' incontrato un interquantile utile: $IQR = q_{75} - q_{25}$ usato nella calcolo del numero di classi di frequenza di un istogramma nel metodo di Fredman-Diaconis (Appr. 2.1). Si veda come esempio alcuni quantili per la distribuzione di Gauss in [Figura 5.11](#).

A seconda del numero di volte in cui si divide l'intervallo di definizione della distribuzione di probabilità i quantili possono prendere nomi (ovvi) come quartili, cinquili, decili, centili, etc. Con questa nomenclatura, ad esempio, il secondo quartile indica: $\int_{-\infty}^{Q_2} f_X(x)dx = 0.5$ e cosi' via.

Nel caso si conosca la funzione di distribuzione cumulativa di probabilità, e' possibile calcolare anche i valori (del variato) per cui si ottiene un certo quartile. Per calcolarli dobbiamo quindi invertire la funzione di distribuzione di probabilità. In alcuni casi si puo' fare in maniera esatta, in questi casi la funzione e' detta *funzione quantile*. Cio' non e' sempre facile nel caso di distribuzioni di probabilità, tuttavia e' piu' facilmente affrontabile nel caso di campioni di misure. A volte si parla come di quartile 0 come del minimo valore della distribuzione (o del minimo valore misurato di un campione) e di ultimo quantile come del valore massimo della distribuzione o del valore massimo misurato. La mediana corrisponde sempre a $q_{0.5}$.

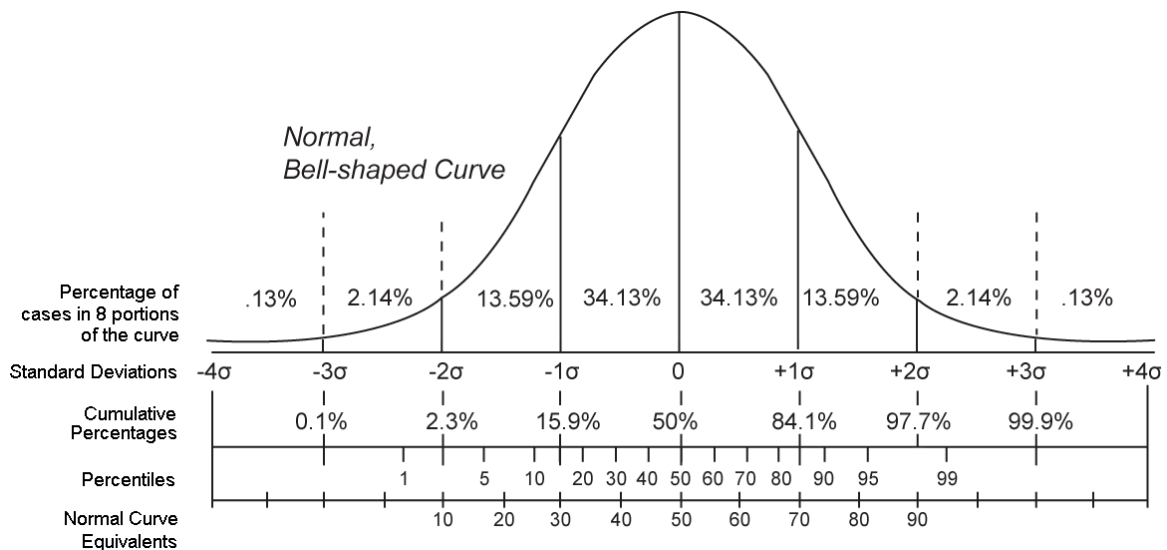


Figura 5.11: Gaussian quantiles. By Chris53516 at English Wikipedia

Si prenda una Gaussiana standard, la funzione quantile e'

$$Q(p) = \sqrt{2} \operatorname{inverf}(2p - 1).$$

dove inverf e' la funzione inversa della funzione di errore, normalmente tabulata o disponibile nelle principali librerie di calcolo. Ad esempio, i quartili della gaussiana sono identificati dai valori $q_0 = -\infty$; $q_{0.25} = -0.68$; $q_{0.5} = 0$; $q_{0.75} = 0.68$ $q_4 = +\infty$

5.7 Commenti finali

Ricapitolando

Esercizi

□ **Es. 5.1** Sia X una Variabile aleatoria di valori x continua e $f(x)$ una funzione analitica definita:

$$f(x) = \begin{cases} -x^2 + a & \text{per } x \in I = [0, \sqrt{a}], \\ 0 & \text{altrimenti} \end{cases}$$

Esprimere le relazioni analitiche cui deve soddisfare $I = [0, a]$ affinché $f(x)$ sia una densità di probabilità. Calcolare la funzione cumulativa di probabilità di $f(x)$. Successivamente esprimere $E[x]$ e $\operatorname{VAR}[x]$ e calcolare la probabilità che x appartenga all'intervallo $[0, 1]$.

□ **Es. 5.2** Sia X una Variabile aleatoria di valori x continua e $f(x)$ una funzione analitica definita:

$$f(x) = \begin{cases} x^2 + 1 & \text{per } x \in I = [-c, c], \\ 0 & \text{altrimenti} \end{cases}$$

Esprimere le relazioni analitiche cui deve soddisfare c affinché $f(x)$ sia una densità di probabilità. Successivamente esprimere $E[x]$ e $\operatorname{VAR}[x]$ e calcolare la probabilità che x appartenga all'intervallo $[0, c/2]$.

□ **Es. 5.3** Il numero di meteoriti superiori a 1 metro di diametro che colpiscono la Terra è in media 3/anno, quale è la probabilità che in un anno si osservino 4 meteoriti di questo tipo? Quale è la probabilità che in un anno si osservino 3 o più meteoriti?

□ **Es. 5.4** Si consideri la distribuzione di probabilità normale $G = N(0, 1.2)$. Usando la tabella della funzione di errore, si calcoli A) la probabilità che un evento si trovi nell'intervallo $[-1.5, 1.5]$, B) la

probabilità che un evento si trovi nell'intervallo $[0, 1.5]$ e C) il numero aspettato di eventi nel caso di misure ripetute che seguano la distribuzione G compresi tra $[-2, 1]$ in un campione di 700 misure ripetute. Si veda [Figura 5.8](#).

□ **Es. 5.5** Sia X una variabile aleatoria continua e sia $f(x)$ una funzione analitica definita:

$$f(x) = \begin{cases} -x^2 + a x & \text{per } x \in I = [0, 1], \\ 0 & \text{altrimenti} \end{cases}$$

- 3.1 Esprimere le relazioni analitiche cui deve $f(x)$ affinché sia una funzione densità di probabilità.
- 3.2 Calcolare il valore di aspettazione e la varianza della distribuzione di densità di probabilità
- 3.3 Calcolare la probabilità che x appartenga all'intervallo finito $[E[x], 1]$.

□ **Es. 5.6** Sia x una variabile aleatoria continua definita su $[0, A]$ dove A è l'ultima cifra della matricola. Si consideri la funzione

$$f(x) = \frac{k}{x + A}$$

Si risolvano i seguenti quesiti:

1. Si dica per quali condizioni $f(x)$ può essere una funzione di densità di probabilità
2. Si calcoli il valore di aspettazione $E[X]$ di $f(x)$
3. Si calcoli la varianza $VAR[X]$ di $f(x)$
4. Si calcoli la probabilità che x appartenga all'intervallo $[0, A/2]$
5. Si stimi la mediana di $f(x)$.

Si usi un numero ragionevole di cifre significative.

Siano x e y due variabili casuali continue definite in \mathbb{R} . Sia $y = (x + 2)^2$. Esprimere $E(y)$, $var(y)$, covarianza $cov(x, y)$ e coefficiente di correlazione di Pearson $\rho(x, y)$ sapendo che $E(x) = 0$ e $var(x) = 1$ (laddove alcune grandezze non siano esprimibili numericamente, lasciare indicato il valore letterale).

□ **Es. 5.7** Al semaforo di una strada secondaria arrivano mediamente, distribuite nel tempo in modo casuale, $\lambda = 2$ macchine al minuto. Supponiamo che inizialmente il semaforo sia sgombero e che rimanga rosso per $80 + A$ secondi. a) Quante auto vengono stoppate mediamente dal semaforo durante il rosso? b) Che probabilità che vengano fermate 4 o più auto? c) Che probabilità c'è che si fermino esattamente 4 auto? Si indichi quale statistica è stata applicata per i calcoli.

Approfondimenti

Appr. 5.1 Dimostrazioni di alcuni momenti. Il calcolo di aspettazione e varianza per la **PMF di Bernoulli** si svolge così:

$$\mu = E[X] = \sum_{i=1}^2 x_i p_i = 1 \cdot p + 0 \cdot (1 - p) = p \tag{5.29}$$

$$\sigma^2 = VAR[X] = \sum_{i=1}^2 (x_i - p)^2 p_i = (1 - p)^2 p + (0 - p)^2 (1 - p) = p(1 - p) \tag{5.30}$$

Per la **PMF Binomiale**, si può risolvere considerando che gli eventi costituenti B sono eventi di Bernoulli indipendenti $B_1, B_2 \dots B_n$. Se è così allora l'evento binomiale è l'evento di somma logica (avviene uno, l'altro o entrambi) e quindi $P(B_1 \cup B_2) = P(B_1) + P(B_2)$. Allora:

$$\mu = E[B] = np \tag{5.31}$$

$$\sigma^2 = VAR[B] = np(1 - p) \tag{5.32}$$

La dimostrazione completa per la **PMF di Poisson** è la seguente.

$$\begin{aligned}
 \mu = E[\mathcal{P}] &= \sum_{i=0}^{\infty} x_i p_i = \underbrace{\sum_{i=0}^{\infty} n_i \frac{\lambda^n e^{-\lambda}}{n!}}_{\text{primo termine e' 0}} = \underbrace{\sum_{i=1}^{\infty} n_i \frac{\lambda^n e^{-\lambda}}{n!}}_{\text{divido sopra e sotto per } n} = \underbrace{\sum_{i=1}^{\infty} \frac{\lambda^n e^{-\lambda}}{(n-1)!}}_{\text{Porto fuori termini}} \\
 &= \lambda e^{-\lambda} \underbrace{\sum_{i=1}^{\infty} \frac{\lambda^{(n-1)}}{(n-1)!}}_{\text{Serie Taylor di } e^\lambda} = \lambda^n e^{-\lambda} e^\lambda = \lambda
 \end{aligned} \tag{5.33}$$

$$\sigma^2 = \text{VAR}[\mathcal{P}] = \lambda^2 \dots \tag{5.34}$$

Per la **PDF di Uniforme** abbiamo visto in che il valore di aspettazione $\mu = (a + b)/2$ e' il punto medio. Per la varianza:

$$\text{VAR}[\mathcal{U}] = \underbrace{\int_a^b (x - \mu)^2 \frac{1}{b-a} dx}_{\text{Cambio variabile } y=x-\mu; dy=dx} = \frac{1}{b-a} \int_{(a-b)/2}^{(a+b)/2} y^2 dy = \frac{1}{b-a} \left[\frac{y^3}{3} \right]_{(a-b)/2}^{(a+b)/2} = \frac{(b-a)^2}{12} \tag{5.35}$$

da cui deriva che la deviazione standard per una misura con risoluzione R e' $\sigma_{\mathcal{U}} = R/\sqrt{12}$. Si noti anche che la frazione di dati contenuta in $[\mu - 1\sigma, \mu + 1\sigma]$ per la distribuzione Uniforme e' pari a 0.58. Per dimostrarlo, si prenda una uniforme con media $\mu = (a + b)/2 = 0$ e $b - a = \sqrt{12} \sim 3.46$ cosi che $\sigma = (b - a)/\sqrt{12} = 1$, allora:

$$E[\mathcal{U}] = \int_{-1}^1 \frac{1}{\sqrt{12}} dx = \frac{1}{\sqrt{12}} [x]_{-1}^1 = \frac{2}{\sqrt{12}} \sim 0.58 \tag{5.36}$$

Si verifichi che cambiando il dominio di \mathcal{U} il risultato non cambia.

Nel caso delle **PDF Esponenziale** $\lambda e^{-\lambda x}$ è necessario integrare per parti. Si ponga $y = \lambda x$, allora $dy = \lambda dx$.

$$\begin{aligned}
 E[X] &= \int_0^{\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda} \int_0^{\infty} y e^{-y} dy = \frac{1}{\lambda} \left[[-y e^{-y}]_0^{\infty} + \int_0^{\infty} e^{-y} dy \right] \\
 &= \frac{1}{\lambda} [-y e^{-y} - e^{-y}]_0^{\infty} = \frac{1}{\lambda} [-e^{-y}(1 + y)]_0^{\infty} = \frac{1}{\lambda}
 \end{aligned} \tag{5.37}$$

Per la varianza si procede in maniera analoga ponendo $y = x^2$ e applicando la regola $\text{VAR}[X] = E[X^2] - (E[X])^2$:

$$\begin{aligned}
 E[X^2] &= \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx = \frac{1}{\lambda^2} \int_0^{\infty} y^2 e^{-y} dy = \frac{1}{\lambda^2} [-2e^{-y} - 2y e^{-y} - y^2 e^{-y}]_0^{\infty} = \frac{2}{\lambda^2} \\
 \text{VAR}[X^2] &= E[X^2] - (E[X])^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}
 \end{aligned} \tag{5.38}$$

Appr. 5.2 Il rapporto tra Binomiale, Poisson e Gaussiana. Abbiamo visto che in alcune condizioni, le PMF Binomiale \mathcal{B} e Poisson \mathcal{P} rassomigliano alla PDF continua della Gaussiana \mathcal{G} . La ragione e' profonda, anche se forse sorprendente visto che sembrano misurare cose molto diverse, e due delle tre sono discrete, l'altra continua. Ne parleremo poi. Il confronto puo' essere fatto tra le varianze delle tre distribuzioni, anche perche' nel caso Binomiale e Poisson il valore di aspettazione e' calcolabile dalla Varianza stessa. Le tre varianze sono:

$$\text{VAR}[\mathcal{B}] = np(1 - p); \quad \text{VAR}[\mathcal{P}] = \lambda; \quad \text{VAR}[\mathcal{G}] = \sigma;$$

i valori medi sono:

$$E[\mathcal{B}] = np; \quad E[\mathcal{P}] = \lambda; \quad E[\mathcal{G}] = \mu;$$

Allora, le condizioni affinché $\mathcal{B} \rightarrow \mathcal{G}$ e $\mathcal{P} \rightarrow \mathcal{G}$ sono

$$\text{VAR}[\mathcal{B}] \geq 9 \quad \& \quad p \ll 1; \quad \text{VAR}[\mathcal{P}] \geq 9;$$

I due limiti si traducono nel fatto che per la binomiale, il numero di ripetizioni n sia piuttosto alto, e per la Poissoniana, che l'evento non sia troppo raro ovvero quindi che il numero di conteggi medi aspettati sia abbastanza elevato. La ragione per cui la binomiale e' rassomiglia alla Gaussiana puo' essere intuita dalla definizione stessa di Gaussiana come combinazione (binomiale) di effetti casuali indipendenti, in numero elevato, che aumentano o diminuiscono il valore medio. In realtà, se pensiamo a tutti questi effetti come casuali, quella che fa la Poissoniana e' contare la probabilità che ne avvenga un certo numero data una media fissa. C'e' da pensarci un po' su, ma poi risulta chiaro. E' anche chiaro infine che alle stesse condizioni, la binomiale puo' essere approssimata da una Poissoniana, piuttosto semplice da scrivere.

Questo limite si puo' anche dimostrare matematicamente. Ho preso questa dimostrazione da Bevington.

$$\binom{n}{k} p^k (1-p)^{n-k} = \frac{1}{k!} \frac{n!}{(n-k)!} p^k (1-p)^n (1-p)^{-k}$$

Ora, assumendo $\text{VAR}[\mathcal{B}] \geq 9, p \ll 1 \rightarrow n \gg 1$ e eguagliando le medie $\lambda = np$ si ottengono le seguenti approssimazioni:

$$\frac{n!}{(n-k)!} = n(n-1) \cdots (n-(k+1)) \sim n^k; \quad \text{se } n \gg 1$$

$$(1-p)^{-k} \rightarrow \text{al primo ordine secondo Taylor} \rightarrow 1 + pk \rightarrow 1 \text{ se } p \ll 1$$

$$\lim_{p \rightarrow 0} (1-p)^n = \lim_{p \rightarrow 0} (1-p)^{\lambda/p} = \lim_{p \rightarrow 0} \left((1-p)^{1/p} \right)^\lambda = e^{-\lambda} \text{ espansione di Taylor}$$

Allora

$$\mathcal{B}_{n,k} = \binom{n}{k} p^k (1-p)^{n-k} \rightarrow \frac{1}{k!} n^k p^k e^{-\lambda} (1) = \frac{\lambda^k e^{-\lambda}}{k!} = \mathcal{P}(k; \lambda)$$

Per curiosità, si guardi anche al seguente link https://en.wikipedia.org/wiki/Relationships_among_probability_distributions#/media/File:Relationships_among_some_of_univariate_probability_distributions.jpg una mappa piu' completa della relazioni tra le principali variabili aleatorie.

Appr. 5.3 La Gaussiana bivariata.

E' facile intuire che il concetto di PMF/PDF di una variabile aleatoria X si puo' estendere al concetto di PDF di diverse variabile aleatorie X, Y, \dots . In questo senso la variabile aleatoria mappa la probabilità che $x = x^*, y = y^*, \dots$ per variabili discrete o $x \in [x_1^*, x_2^*], y \in [y_1^*, y_2^*], \dots$ per variabili continue. Esse mappano cioe' la probabilità congiunta delle due o piu' variabili aleatorie. E' chiaro che il caso piu' comune e' questo. Ad esempio, se una stima sperimentale $z = f(x, y)$ allora la distribuzione di probabilità di Z dipende da quelle di X, Y . Tuttavia, spesso e' troppo complesso, o non serve, costruire l'intera distribuzione di probabilità per Z , come vedremo. In alcuni casi ha senso. Ad esempio, una distribuzione bivariata X, Y Gaussiana, con variabili X, Y *indipendenti*, si scrive:

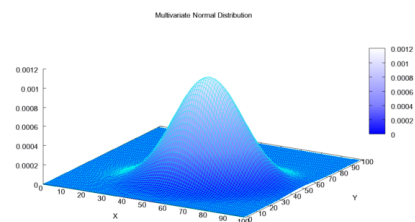


Figura 5.12: Distribuzione di probabilità della Gaussiana bivariata. Credit: Wikipedia.

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y} e^{-\frac{1}{2} \left[\left(\frac{x-\mu_X}{\sigma_X} \right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 \right]} \quad (5.39)$$

Appr. 5.4 La regola dei 5 sigma. La regola dei 5 sigma viene usata in molti ambienti fisici (non in altre discipline) per stabilire il margine di confidenza per poter affermare (e pubblicare) una scoperta scientifica. Spesso, evidenze sperimentali 'al di sotto dei 5 sigma' non vengono accettati dai giornali scientifici. Cosa si intende? Il significato sara' chiaro quando studieremo il test del χ^2

(Capitolo 10). Per ora ci limitiamo a dire che la regola stabilisce che la probabilità che la scoperta sia dovuta semplicemente a fluttuazioni statistiche della misura (outliers li abbiamo anche chiamati) sia inferiore a $5.7 \cdot 10^{-7}$ ovvero meno del $5.7 \cdot 10^{-5}\%$ delle volte. Il valore deriva da Tab. 5.1 ed è preso dall'ultima colonna in corrispondenza di $\delta = 5\sigma$. Una scoperta fatta tra 4 e 5σ è detta avere evidenza marginale. Al di sotto dei 4σ si considera ancora probabile la fluttuazione statistica.

Se la scoperta è ancora più evidente, allora si riporta non solo il fatto di aver superato 'i 5 sigma' ma anche lo si quantifica, calcolando in numero di sigma di scostamento. Ad esempio, una scoperta a '10 sigma' sarà molto più evidente e indiscutibile di una a 5 sigma. Si prenda un articolo di scoperte e si noterà come si esprime la evidenza della scoperta quasi sempre con questo linguaggio⁴.

Rimane da capire questi 'sigma' di distanza rispetto a che cosa siano. Vedremo in Capitolo 10 che si tratta di deviazioni standard rispetto alla ipotesi (detta nulla) in cui non ci sia evidenza per una scoperta!

Appr. 5.5 Persone.

Matt Bognar sito e app Un professore Statounitense, Matt Bognar ha creato una pagina web: <https://homepage.divms.uiowa.edu/~mbognar/> e un'app per smartphone e tablet gratuita che permettono una rapida visualizzazione di distribuzioni di probabilità discrete e continue, e il conto di alcune probabilità. Si dia una occhiata!



Credit: Wikipedia.

Siméon Denis Poisson (1781-1840) Fu un grande matematico e fisico francese, con contributi enormi in molte discipline. Nel corso di Fisica vedrete quante cose portano il suo nome. Fu di origini modeste, ma venne incoraggiato agli studi e poi divenne docente dell'École Polytechnique di Parigi anche grazie al sostegno di Laplace e Fourier. I suoi compagni di scuola avevano fatto l'abitudine di fargli visita dopo una lezione difficile per sentirlo ripetere e spiegare. Divenne un amato e bravo professore. Tra le sue citazioni appassionate, ci piace questa: *"Life is good for only two things: doing mathematics and teaching it"*.

⁴Ad esempio <https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.116.061102> per la scoperta di onde gravitazionali, con margine di confidenza di 5.1σ .

Capitolo 6

Relazioni tra variabili aleatorie

Quest'opera è distribuita con licenza [Creative Commons](#) "Attribuzione – Non commerciale – Condividi allo stesso modo 3.0 Italia".



Supponiamo di avere una variabile aleatoria X legata ad una seconda variabile aleatorio Z dalla relazione $z = g(x)$. La domanda che ci poniamo e', data la distribuzione di probabilità $f_X(x)$ e' possibile conoscere $f_Z(z)$ nota la relazione funzionale $z = g(y)$:

$$z = g(x) \quad f_X(x) \quad \rightarrow \quad f_Z(z)?$$

La risposta e' positiva, ma distingueremo il caso in cui questa operazione si puo' fare in maniera *esatta*, ad esempio per le $g(x)$ di tipo lineare, e altri casi in cui si potra' fare solo in maniera approssimata. Vedremo inoltre che spesso non e' neppure necessario conoscere tutta la $f_Z(z)$ ma e' sufficiente determinare in forma approssimata alcuni dei suoi momenti, in particolare media e varianza.

Il caso sopra rappresenta il caso di una funzione di un'unica variabile. Analogamente, accenneremo al caso piu' generico di due variabili X, Y ed una variabile Z formalmente legata alle prime due da $z = g(x, y)$ (multivariata).

$$z = g(x, y) \quad f_X(x), f_Y(y) \quad \rightarrow \quad f_{XY}(x, y) \quad f_Z(z)?$$

In questo caso infatti dovremo non solo conoscere le FP individuali $f_X(x), f_Y(y)$ che rappresentano le distribuzioni di probabilità che X, Y avvengano singolarmente, ma anche le loro congiunzione logica $f_{XY}(x, y)$ ovvero la distribuzione di probabilità che esse avvengano 'in congiunzione'. E' chiaro il collegamento con il caso di misure: una generica misura z e' output di misure di input x, y secondo un modello matematico $g(x, y)$.

In questo capitolo non duplicheremo sempre la notazione per il caso PMF e PDF per non appesantire la lettura. Si invita lo studente a provare sempre a ottenere il caso mancante. Inoltre, useremo la notazione contratta PF per le funzioni di probabilità quando la formula e' valida sia per le PMF che per le PDF.

6.1 Congiunzione logica, Indipendenza e Correlazione tra variabili aleatorie

In teoria della probabilità ([Capitolo 4](#)) abbiamo visto due eventi aleatori A, B sono indipendenti se $P(A|B) = P(A)$. Questo ci aveva portato a calcolare la probabilità condizionata di eventi indipendenti come $P(AB) = P(A|B)P(B) = P(A)P(B)$.

Questo stesso concetto si puo' tradurre direttamente per le PMF ed e' valido anche per le densità di probabilità delle PMF:

Due variabili aleatorie sono indipendenti se le loro PF soddisfano:

$$f_{XY}(x, y) = f_X(x) \cdot f_Y(y) \tag{6.1}$$

ovvero che la PF congiunta è uguale al prodotto delle PF individuali.

Si noti come nel caso PMF [Equazione 6.1](#) riguardi probabilità mentre nel caso PDF densità di probabilità. L'indipendenza esprime in sostanza il fatto che la probabilità di trovare un valore $x \in [x_1, x_2]$ di X non influenza la probabilità di trovare un valore $y \in [y_1, y_2]$ di Y e viceversa¹. Nel caso quindi in cui le due variabili aleatorie siano indipendenti, di fatto scrivere la funzione di probabilità congiunta equivale a scrivere un prodotto di fattori indipendenti.

Cerchiamo di vedere in dettaglio la forma della distribuzione di probabilità congiunta $f_{XY}(x, y)$. Prendiamo ad esempio due distribuzioni gaussiane $\mathcal{N}(X; 0, \sigma_1^2), \mathcal{N}(Y; 0, \sigma_2^2)$ di media 0 e ampiezza σ_1, σ_2 . Abbiamo visto che la $f_{XY}(x, y)$ è una PDF bivariata. Ne abbiamo già parlato in [Appr. 5.3](#).

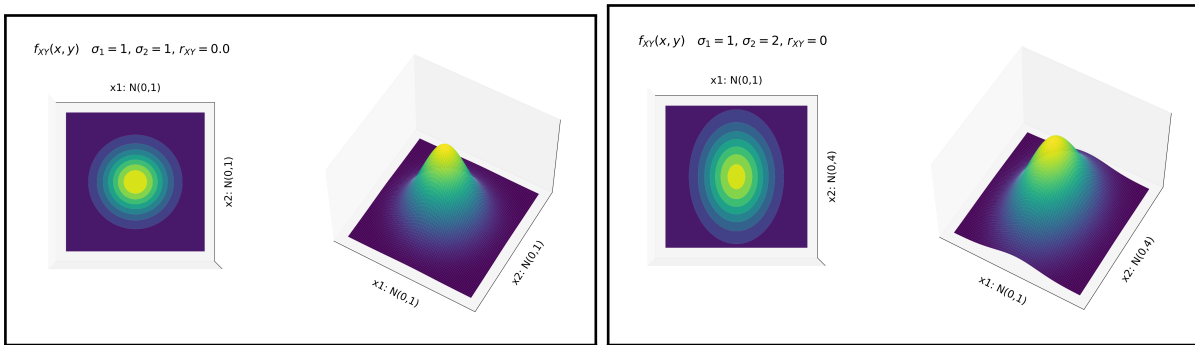


Figura 6.1: (sinistra) PDF bigaussiana con $X = \mathcal{N}(0, 1), y = \mathcal{N}(0, 1)$ e X, Y indipendenti (destra) Come a sinistra ma con $Y = \mathcal{N}(0, 4)$

Se X, Y sono indipendenti, la PDF congiunta $f_{XY}(x, y)$ appare come in [Figura 6.1](#). Nella figura mettiamo due casi, in cui facciamo variare unicamente l'ampiezza σ_2 della seconda gaussiana. Nel caso a sinistra $\sigma_1 = \sigma_2 = 0$, la $f_{XY}(x, y)$ sia una bi-gaussiana, ovvero una gaussiana in 2 dimensioni. Essa è simmetrica per rotazione attorno all'asse centrale. Nel caso a destra $\sigma_1 \neq \sigma_2$, si tratta ancora di una bigaussiana ma non c'è simmetria. In entrambi i casi vale $f_{XY}(x, y) = f_X(x) f_Y(y)$, cioè significa per ogni punto (x, y) la densità di probabilità è il prodotto delle densità separate.

Che cosa esprime la forma della PDF bigaussiana? Ricordiamoci (si veda anche l'Approf. [3.3](#)) che quando parlo di X variabile aleatoria intendo dire che il valore x che può uscire varia all'interno del dominio di X con una certa probabilità data da $f_X(x)$. Si sta cioè 'pescando casualmente' uno dei valori possibili della variabile aleatoria. Nel caso di [Figura 6.1](#) (sinistra) ad esempio, tutti i valori a distanza $r = \sqrt{x^2 + y^2}$ sono equiprobabili tra loro, e la gaussiana è di fatto analoga ad una gaussiana unidimensionale con r invece di x, y . Di fatto questa gaussiana bivariata è simmetrica per rotazione rispetto all'asse centrale. Se si fissa un $x = x^*$ specifico, e si controlla la forma della $f_Y(y|x^*)$ si vede che questa corrisponde a $f_Y(y)$. Nel caso di [Figura 6.1](#) (destra), le due variabili di partenza 'pesano diversamente' ma la situazione è simile alla precedente tranne la totale simmetria tra le due variabili. In questo caso, sono più probabili valori vicini alla media di X che di Y . Nuovamente, la scelta di un $x = x^*$ non influisce su $f_Y(y)$ e viceversa, perché le due variabili sono *indipendenti*. Se ragioniamo in termini di probabilità otteniamo che per l'intervallo $[0, 1]$

$$P_Y(y \in [0, 1]) = \frac{1}{2} P_X(x \in [0, 1])$$

¹Si presti attenzione a non confondere i simboli delle distribuzioni con quelli delle funzioni tra variabili. Ad esempio potrebbe essere

$$z = g(x, y) = a \sin(2x) + \log y \quad \text{e } f_X(x) \text{ normale e } f_Y(y) \text{ uniforme.}$$

La forma della PF bi-gaussiana per X, Y indipendenti e':

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y} e^{-\frac{1}{2}\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right)} \quad (6.2)$$

Tuttavia, oltre alla indipendenza, esistono diverse forme di dipendenza. La dipendenza logica e' totalmente descritta dalla forma della $f_{XY}(x, y)$ e ancora una volta puo' essere pensata in termini di variabili aleatorie dicendo che la la probabilita' di trovare un valore $y \in [y_1, y_2]$ di Y e' influenzata dalla probabilita' di trovare un valore $x \in [x_1, x_2]$ di X o viceversa, o in altre parole, la probabilita' di trovare un valore $y \in [y_1, y_2]$ di Y va *aggiornata* nel caso si conosca cosa e' successo all'evento aleatorio X .

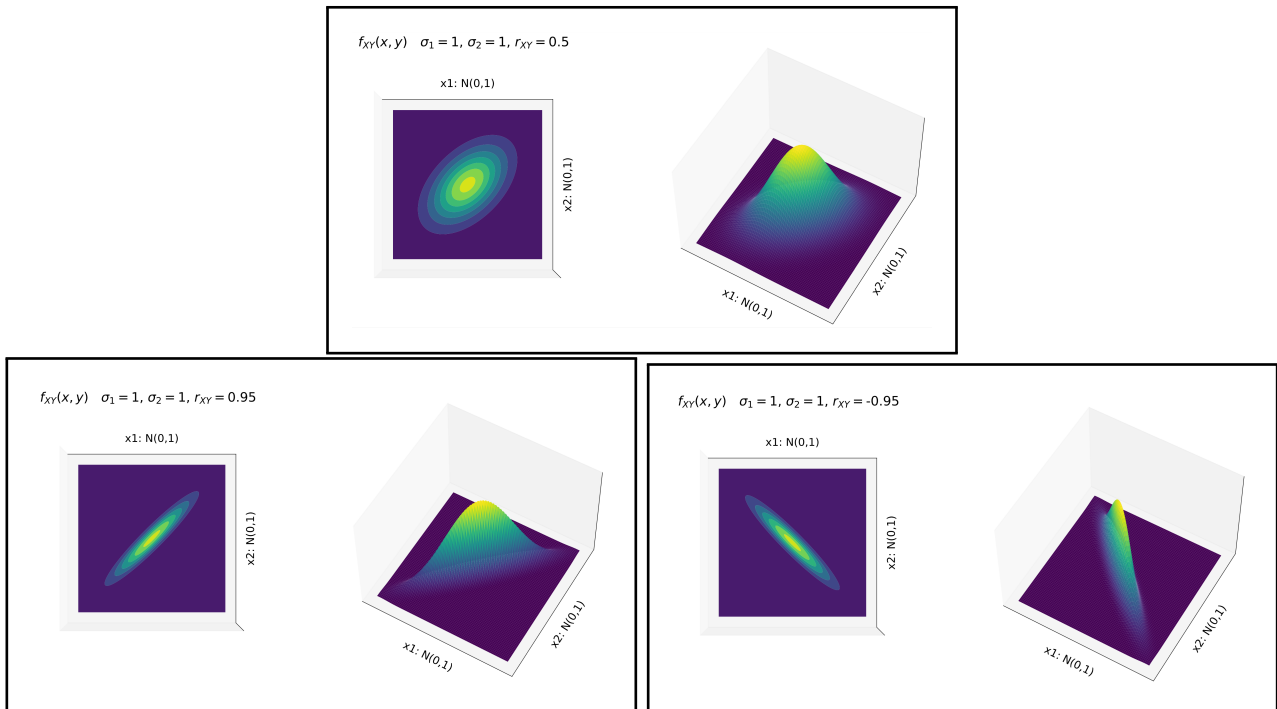


Figura 6.2: (sinistra) PDF bigaussiana con $X = \mathcal{N}(0, 1), y = \mathcal{N}(0, 1)$ e X, Y indipendenti (destra) Come a sinistra ma con $Y = \mathcal{N}(0, 4)$

Cosa succede quindi a $f_{XY}(x, y)$ se X, Y sono dipendenti? In **Figura 6.2** vediamo l'effetto di dipendenza logica tra variabili. In **Figura 6.2** (alto) abbiamo una dipendenza di media 'intensita'. In **Figura 6.2** (basso) abbiamo una dipendenza di forte intensita', ma di tipo diverso. Vediamo come. Nel primo caso, si avra' che non solo i valori di X, Y non sono equiprobabili, ma anche che *la probabilita' di avere un valore alto di X aumenta la probabilita' di avere un valore alto di Y* . Questo equivale a dire che $P(Y|X) \neq P(Y)$ che non e' altro che la condizione di dipendenza logica. Il secondo caso e' analogo al primo solamente piu' vistoso: un valore via via maggiore di X porta con *minore incertezza* ad un valore via via maggiore di Y . Nel terzo caso invece abbiamo un comportamento simile al secondo ma invertito: un valore via via maggiore di X porta con grande certezza ad un valore via via minore di Y . I primi due casi sono detti di *correlazione tra le due grandezze*² mentre il terzo caso di *anticorrelazione tra le due grandezze*.

Nell'ambito delle PF, la correlazione tra due variabili in sostanza qualifica lo scostamento dall'indipendenza, secondo lo schema:

$$f_{XY}(x, y) \begin{cases} = f_X(x)f_Y(y) & \text{Indipendenza} \\ \neq f_X(x)f_Y(y) & \text{Varianza congiunta (covarianza)} \end{cases} \quad (6.3)$$

²Il termine fu coniato da R. Galton (1822-1911), cugino di R. Darwin e inventore dell'eugenetica. A lui si deve anche la Galton board mostrata a lezione

La covarianza congiunta e' totalmente definita dalla forma di $f_{XY}(x, y)$, tuttavia, questa dipende dalle PF di partenza e dalla relazione funzionale tra X, Y . Ad esempio, nel caso di [Figura 6.2](#) la forma funzionale che descrive la curva e':

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}} e^{-\frac{1}{2(1-\rho_{XY}^2)} \left[\left(\frac{x-\mu_X}{\sigma_X} \right)^2 - 2\rho_{XY} \left(\frac{x-\mu_X}{\sigma_X} \frac{y-\mu_Y}{\sigma_Y} \right) + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 \right]} \quad (6.4)$$

dove rispetto a [Equazione 6.2](#) si ha un termine in piu' legato al parametro ρ_{XY} che vedremo ([Equazione 6.11](#)) essere legato al gradi di covarianza tra le variabili. Inoltre, l'effetto di questa covarianza (ovvero l'effetto di conoscere una variabile e poi calcolare probabilita' dell'altra) non e' lo stesso per ogni punto della funzione, ma varia di intervallo in intervallo: l'entita' dell'effetto su $P((y_1, y_2))$ dipende a seconda che $x \in [x_1, x_2]$ o $x \in [x_3, x_4]$. Di conseguenza non e' possibilmente quantificare 'globalmente' la covarianza. E' tuttavia possibile calcolare una 'covarianza media' attraverso il concetto appreso dei valori di aspettazione.

6.1.1 Covarianza e correlazione lineare

Usando la notazione vista in [sezione 5.5](#), per quantificare un grado di correlazione medio e' sufficiente calcolare il valore di aspettazione $E[XY]$ del prodotto logico delle due variabili. In particolare, avvera' che:

$$E[XY] \begin{cases} > E[Y]E[X] & \text{Covarianza positiva (Figura 6.2 caso 1 e 2)} \\ < E[Y]E[X] & \text{Covarianza negativa (caso 3 in Figura 6.2)} \\ = E[Y]E[X] & \text{Covarianza nulla (casi in Figura 6.1)} \end{cases} \quad (6.5)$$

Se definiamo percio' una variabile che chiamiamo covarianza come

$$COV[X, Y] = E[XY] - E[X]E[Y] \quad (6.6)$$

allora:

$$COV[X, Y] \begin{cases} > 0 & \text{Covarianza positiva - correlazione} \\ < 0 & \text{Covarianza negativa - anticorrelazione} \\ = 0 & \text{Covarianza nulla - non correlazione} \end{cases} \quad (6.7)$$

La covarianza e' quindi la differenza tra il valore di aspettazione del prodotto, e il prodotto dei valori di aspettazione. Allora la correlazione e' positiva nel caso la fluttuazione positiva (rispetto alla media) di una variabile implichi una fluttuazione positiva (rispetto alla media) nell'altra variabile. Essa e' negativa nel caso contrario in cui una fluttuazione positiva (rispetto alle media) di un variabile implichi in media una fluttuazione negativa (rispetto alle media) nell'altra variabile. Le fluttuazioni nella teoria delle distribuzioni di probabilita' non sono altro che gli scarti $X - E[X], Y - E[Y]$, allora e' possibile pensare alla covarianza anche come il valore di aspettazione dei prodotti degli scarti $COV[X, Y] = E[(X - E[X])(Y - E[Y])]$ di nuovo il risultato e' positivo, negativo o nullo, come sopra.

Vale (usando $E[X] = \mu_X$ e μ_X e' numero e analogo per Y):

$$\begin{aligned} COV[X, Y] &= E[(X - E[X])(Y - E[Y])] = E[(X - \mu_X)(Y - \mu_Y)] \\ &= E[XY - Y\mu_X - X\mu_Y + \mu_X\mu_Y] = E[XY] - E[Y\mu_X] - E[X\mu_Y] + E[\mu_X\mu_Y] \\ &= E[XY] - \mu_X\mu_Y + \mu_X\mu_Y - \mu_X\mu_Y = E[XY] - E[X]E[Y] \quad \blacksquare \end{aligned}$$

Le due definizioni sono analoghe, e quindi definiamo:

Correlazione lineare tra due variabili aleatorie X, Y la grandezza detta Covarianza di X, Y espressa come

$$\begin{aligned} \text{COV}[X, Y] &= E[(X - E[X])(E[Y - E[Y]])] \quad \text{ovvero} \\ &= E[XY] - E[X]E[Y] \end{aligned} \quad (6.8)$$

Esplicitiamo ora il calcolo della covarianza per le PDF e PMF:

Covarianza in forma estesa:

$$\text{COV}[X, Y] \text{ PDF : } \left(\int_X \int_Y x y f_{XY}(x, y) dx dy \right) - \int_X x f_X(x) dx \int_Y y f_Y(y) dy \quad (6.9)$$

$$\text{COV}[X, Y] \text{ PMF : } \sum_{i=1}^N \sum_{j=1, i \neq j}^N x y f_{XY}(x, y) - \sum_{i=1}^N x f_X(x) dx \sum_{j=1}^N y f_Y(y) \quad (6.10)$$

Inoltre, si utilizza anche il simbolo

$$\sigma_{XY} = \text{COV}[X, Y]$$

La covarianza varia positivamente e negativamente su tutto \mathcal{R} in base ai valori di X, Y . Come già visto per la varianza e i momenti di ordine superiore, e' possibile 'normalizzare' questo valore. Si definisce allora:

Coefficiente di correlazione lineare di Pearson-Bravais (o piu' semplicemente coefficiente di Pearson o coefficiente di correlazione lineare) il parametro:

$$\rho_{XY} = \frac{\text{COV}[X, Y]}{\sqrt{\text{VAR}[X] \text{VAR}[Y]}} \quad (6.11)$$

Si puo' dimostrare (disuguaglianza di Schwartz) che $1 \leq \rho_{XY} \leq 1$. A seconda di questo valore si definiscono i seguenti casi:

$$\rho_{XY} : \begin{cases} \rho_{XY} > 0 & \text{correlazione positiva, } \rho_{XY} = 1 & \text{correlazione totale} \\ \rho_{XY} < 0 & \text{anticorrelazione o corr. negativa, } \rho_{XY} = -1 & \text{anticorrelazione totale} \\ 0 < |\rho_{XY}| < 0.3 & \text{(anti)correlazione debole} \\ 0.3 < |\rho_{XY}| < 0.7 & \text{(anti)correlazione intermedia} \\ 0.7 < |\rho_{XY}| < 1 & \text{(anti)correlazione forte} \end{cases}$$

Ad esempio, nel caso di [Figura 6.2\(alto\)](#) abbiamo una correlazione intermedia ($\rho = 0.5$). In [Figura 6.2\(basso,sx\)](#) abbiamo una correlazione forte ($\rho = 0.9$) e in [Figura 6.2\(basso,dx\)](#) abbiamo una anticorrelazione forte ($\rho = -0.9$).

Limite di validità di $\text{COV}[X, Y]$ e ρ_{XY} . Relazione covarianza indipendenza. In realtà la covarianza e il coefficiente di Pearson rappresentano qualità media, e in questo hanno due principali limitazioni:

- non quantificano direttamente il grado di dipendenza logica tra variabili aleatorie, ma solo la dipendenza *lineare*. In un certo senso quindi, solo il primo ordine di dipendenza logica.
- ci sono diversi casi in cui se la covarianza e' nulla, cio' non implica indipendenza.

In particolare, sottolineiamo come mentre l'indipendenza tra variabili aleatorie implica covarianza nulla, non e' vero il viceversa:

$$\underbrace{f_{XY}(x, y) = f_X(x) f_Y(y)}_{\text{Indipendenza}} \rightarrow \text{COV}[X, Y] = 0$$

$$\text{COV}[X, Y] = 0 \not\Rightarrow \underbrace{f_{XY}(x, y) = f_X(x) f_Y(y)}_{\text{Indipendenza}}$$

Altri casi di limiti di validità sono riportati nell'Approf. 6.1.

Il ruolo della dipendenza funzionale Torniamo ora al ruolo della dipendenza funzionale tra le variabili, ad esempio $y = a + bx$. In questo caso sono correlate X, Y ? E' chiaro che lo sono, all'aumentare di X aumenta Y . Quanto vale la correlazione tra le due variabili?

$$\rho[Y, X] = \rho[a + bX, X] = \begin{cases} +1 & b > 0 \\ -1 & b < 0 \end{cases} \quad (6.12)$$

La relazione si dimostra facilmente con le proprietà che dimostreremo in [sezione 5.5](#). Notiamo come il grafico della PDF congiunta $F_{XY}(x, y)$ in questo sapere molto simile a [Figura 6.1](#) (basso sinistra o destra) solamente 'ancora piu' stretto'. Si noti che il coefficiente $\rho_{XY} = 1$ indipendentemente dalla pendenza b della retta, quindi rette a pendenza diversa (purche' non orizzontale) danno sempre $\rho_{XY} = 1$.

In generale nel caso di variabili aleatorie, se non rappresentano specifiche grandezze misurabili, la relazione funzionale e' nota a priori. Nel caso sia nota allora e' possibile costruire la PF congiunta. In altri casi in cui non fosse nota, non lo si puo' fare. Vedremo che questo concetto e' declinato in maniera specifica nel caso dei campioni finiti dove *stimeremo* sperimentalmente la covarianza campionaria a partire dalle misure stesse.

Il caso di piu' variabili aleatorie Nel caso in cui si abbiamo diverse variabili aleatorie, ad esempio nel caso si voglia valutare una $g(X, Y, Z)$, allora si devono valutare le intercorrelazioni *coppie* di variabili. Si valuterà quindi ad esempio $\rho_{XZ}, \rho_{YZ}, \rho_{XZ}$. Tutti questi termini vengono poi raccolti nella cosiddetta *matrice di covarianza*, presentata in Appr. ?? e che vedrete in dettaglio al II anno.

Incertezza sulla coefficiente di correlazione Nel caso di distribuzioni di probabilità, la covarianza tra due variabili aleatorie e' nota senza incertezza:

$$\text{VAR}(\rho_{XY}) = 0 \quad (6.13)$$

Si noti come non sara' cosi' nel caso dei campioni (Sec. 7.2).

6.2 PF di funzioni di variabili

Torniamo al caso della costruzione della funzione di probabilità di una variabile costruita come funzione di una o piu' variabili: $f(g(x))$ o $f(g(x, y, \dots))$. La domanda che ci poniamo quindi e': se conosciamo la relazione funzionale, le distribuzioni di partenza $f_X(x), f_Y(y), \dots$ e la correlazione tra le variabili, possiamo stimare $f(g(x))$ o $f(g(x, y, \dots))$?

Affronteremo il problema a passi di complessità via via maggiore. In particolare, vedremo come in alcuni casi potremo risolvere il problema in forma esatta, ad esempio nel caso la relazione tra le variabili sia lineare, e poi in maniera approssimata nel caso generico. Partiremo poi anzitutto dal caso univariato (dipendenza di una sola variabile) per poi andare al caso multivariato.

6.2.1 PF di funzioni di una variabile

Si consideri la dipendenza funzionale:

$$y = g(x) \text{ con } x \text{ distribuita secondo } f_X(x)$$

Per calcolare $f_Y(y)$ si deve mappare $x \rightarrow y$ attraverso il cambio di variabile:

$$x = g^{-1}(y)$$

Allora per fare un mappamento delle probabilità si prosegue con:

$$f_Y(y) = \frac{dP_Y(y)}{dy} = \frac{dP_X(x = g^{-1}(y))}{dx} \left| \frac{dx}{dy} \right| = \left| \frac{dx}{dy} \right| f_X(x = g^{-1}(y))$$

Risulta quindi che per poter mappare la probabilità e' necessario che la funzione $g(x)$ sia invertibile. Sono invertibili (o invertibili a tratti) le funzioni monotone crescenti o decrescenti (anche a tratti).

Caso dipendenza monotona lineare Ad esempio, sia

$$y = g(x) = a + bx \quad g(x) \text{ e' monotona e } f_X(x) = \mathcal{G}(x; \mu, \sigma^2)$$

Per calcolare $f_Y(y)$ ragioniamo attraverso il cambio di variabile:

$$x = (y - a)/b; \quad dx = dy/b$$

Allora usando la formula per il mappamento:

$$f_Y(y) = \frac{1}{|b|} f_X\left(\frac{y - a}{b}\right)$$

Non e' finita, allora se $f_X(x) = \mathcal{G}(x; \mu, \sigma^2)$ risulta:

$$f_Y(y) = \frac{1}{|b|} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{\frac{y-a}{b} - \mu}{\sigma}\right)^2} \rightarrow \frac{1}{\sqrt{2\pi} |b| \sigma} e^{-\frac{1}{2}\left(\frac{y - (a + b\mu)}{\sigma|b|}\right)^2}$$

dove la formula a destra e' ottenuta attraverso alcuni passaggi algebrici. Se proviamo a interpretare la nuova $f_Y(y)$ notiamo come essa sia ancora una gaussiana. In particolare, risulta intuitivo come il valore medio segua l'andamento della funzione, ma si noti anche come varia la ampiezza della Gaussiana:

$$\mu \rightarrow a + b\mu \quad \sigma \rightarrow |b|\sigma$$

La trasformazione quindi e' stata:

$$f_X(x) = \mathcal{G}(x; \mu, \sigma^2) \rightarrow f_Y(y) = \mathcal{G}(y; a + b\mu, (b\sigma)^2)$$

Deduciamo che nel caso della relazione lineare la distribuzioni di probabilità cambia forma ma non natura. Questo risultato e' importante e sarà utilizzato successivamente durante la approssimazione di funzioni generiche in funzioni lineari attorno ad un punto.

Si pensi al caso della misura del periodo multiplo del pendolo nella prima esperienza del laboratorio di Fisica ??.

Caso dipendenza monotona esponenziale Si supponga di avere $y = g(x) = e^x$, allora $x = \ln(y)$, $dx = dy/y$ e quindi se $f_X(x) = \mathcal{G}(x; \mu, \sigma^2)$:

$$f_Y(y) = \frac{1}{y} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{\ln(y) - \mu}{\sigma}\right)^2}$$

Si nota come ora $f_Y(y)$ sia la PDF cosiddetta lognormale! In questo caso quindi la distribuzione di probabilità e' cambiata!

Si dimostri attraverso il cambio di variabile opportuno come da una Gaussiana si arriva ad una Gaussiana standard. Si calcoli la PDF di $y = 3x + 2$ sapendo che X e' distribuita secondo una esponenziale di media λ .

Funzione monotona a tratti Se la relazione $y = g(x)$ e' monotona a tratti, allora e' facile vedere che la formula sopra puo' essere valida se scritta come:

$$f_Y(y) = dy = \underbrace{\left| \frac{dx}{dy} \right| f_X(g^{-1}(y))}_{\text{tratto 1}} + \underbrace{\left| \frac{dx}{dy} \right| f_X(g^{-1}(y))}_{\text{tratto 2}} + \dots$$

Si calcoli la PDF di $y = x^2$ sapendo che X e' distribuita secondo una normale standard usando la formula della monotonia a tratti.

Nel caso generico, una relazione $g(x)$ potrebbe non essere invertibile. Come ci si puo' comportare in questi casi? In linea generale, e' importante notare che nel mappare $f_X(x) \rightarrow f_Y(y)$ non siamo realmente interessati necessariamente a tutti i valori di X . E' piu' probabile che sia sufficiente considerare ad esempio un sottoinsieme di valori di X in cui la funzione sia sufficientemente regolare da essere invertibile. Ma quale regione? Una proposta ragionevole potrebbe essere la regione attorno al valore di aspettazione di X ovvero μ_X . E potrebbe sempre essere ragionevole prendere come ampiezza di questa regione la stima di dispersione di misura attorno a μ_X , ovvero σ_X , la deviazione standard. Allora in questo caso dovremmo mappare unicamente $f_X(\mu_X \pm \sigma_X) \rightarrow f_Y(y)$.

In queste condizioni e' possibile sviluppare la funzione $g(x)$ come un polinomio attorno a μ_X secondo lo sviluppo in serie di Taylor (uso $\mu = \mu_X$):

$$y = g(x) \simeq \underbrace{g(\mu)}_{\text{ordine 0}} + \underbrace{\left(\frac{dg(x)}{dx} \right)_{\mu}}_{\text{ordine 1}} (x - \mu) + \underbrace{\frac{1}{2!} \left(\frac{d^2g(x)}{dx^2} \right)_{\mu}}_{\text{ordine 2}} (x - \mu)^2 + \mathcal{O}(3) \quad (6.14)$$

Nella formula sopra, $|dg(x)/dx|_{\mu}$ indica la derivata prima calcolata in μ , si tratta quindi di un numero e non di una funzione. E' chiaro che l'approssimazione e' tanto migliore quanto piu' piccolo e' l'intervallo attorno a cui viene valutata, ovvero tanto piu' piccolo e' $x - \mu$. Essendo ora concentrati attorno al valore di aspettazione di X e alla sua dispersione σ_X , non ha piu' molto senso calcolare l'intera $f_Y(y)$, ma sara' sufficiente calcolarne i primi momenti $\mu_Y = E(g(x))$ e $\sigma_Y = \sqrt{\text{VAR}(g(x))}$. Procederemo in questa maniera in [sezione 6.3](#)

6.2.2 Caso bivariato e multivariato

Nel caso in cui si abbia una dipendenza multi-variata del tipo $z = g(x, y)$ non si puo' piu' pensare ad una semplice inversione di coordinate. La $f_Z(z)$ sara' in generale influenzata dalla probabilita' congiunta di X, Y se le variabili sono correlate. In generale, il problema e' complesso e non sempre risolvibile analiticamente. Come appena discusso, talvolta non conviene costruire l'intera $f_Z(z)$ ma solo i suoi momenti. Tuttavia esistono alcuni casi in cui il problema e' risolvibile, in particolar modo nel caso di variabili indipendenti e di relazione lineare del tipo $z = x + y$. Mostriamo un esempio di questo caso in [Appr. 6.3](#). In generale, procederemo come a sviluppare la funzione $g(x, y)$ come un polinomio in serie di Taylor attorno a μ_X, μ_Y . Nel caso la funzione sia bivariata, l'espansione in serie di Taylor procede in maniera analoga attorno alle medie delle $\mu_x = E[X], \mu_y = E[Y]$ delle due distribuzioni. In questo si deve pero' usare le derivate parziali. Usiamo infine una notazione ancora piu' compatta definendo $x^* = \mu_x = E[X]$ e analogo per Y):

$$z = g(x, y) \sim \left(\frac{\partial g(x, y)}{\partial x} \right)_{(x^*, y^*)} (x - x^*) + \left(\frac{\partial g(x, y)}{\partial y} \right)_{(x^*, y^*)} (y - y^*) + \mathcal{O}(2) \quad (6.15)$$

Continueremo a discutere come usare questa approssimazione piu' sotto in [sezione 6.4](#). Per poterlo fare, dobbiamo prima notare alcune proprietà algebriche dei momenti di combinazioni lineari di variabili. Questo e' fatto in [sezione 6.3](#). Interrompiamo per il momento questa discussione e riprendiamola quindi poi in [sezione 6.4](#).

6.3 Momenti di combinazioni lineari di variabili aleatorie

Consideriamo la relazione lineare $Y = aX + bY + c$, allora dimostriamo che valgono le seguenti proprietà:

$$\begin{aligned} E[aX + bY + h] &= a E[X] + b E[Y] + h \\ \text{VAR}[aX + bY + h] &= a^2 \text{VAR}[X] + b^2 \text{VAR}[Y] + 2ab \text{COV}[X, Y] \end{aligned}$$

Dimostriamo quanto sopra a passi. Cominciamo con la media:

$$E[a] = a \quad : \quad \sum a P(a) = a \cdot 1 = a \quad (6.16)$$

$$E[aX] = a E[X] \quad : \quad \int a x f(x) dx = a \int x f(x) dx = a E[X] \quad (6.17)$$

$$E[X + h] = E[X] + h \quad : \quad \int (x + h) f(x) dx = \int x f(x) dx + \int h f(x) dx = E[X] + h \quad (6.18)$$

$$\begin{aligned} E[X + Y] = E[X] + E[Y] \quad : \quad & \iint (x + y) f_{XY}(x, y) dx dy = \iint x f_{XY}(x, y) dx dy + \iint y f_{XY}(x, y) dx dy \\ & = \int x f_X(x) dx + \int y f_Y(y) dy = E[X] + E[Y] \end{aligned} \quad (6.19)$$

dove nell'ultimo passaggio si è sfruttata la legge della probabilità totale e il concetto di marginalizzazione (vedi anche Appr. 6.4). Passiamo ora alla varianza:

$$\text{VAR}[a] = 0 \quad : \quad \sum (a - E[a])^2 P(a) = 0 \quad (6.20)$$

$$\begin{aligned} \text{VAR}[aX] = a^2 \text{VAR}[X] \quad : \quad & \int (ax - E[aX])^2 f(x) dx = \int (ax - a E[X])^2 f(x) dx \\ & = \int a^2 (x - E[X])^2 f(x) dx = a^2 \text{VAR}[X] \end{aligned} \quad (6.21)$$

$$\begin{aligned} \text{VAR}[X + h] = \text{VAR}[X] \quad : \quad & \int ((x + h) - E(x + h))^2 f(x) dx \\ & = \int ((x + h) - (E(x) + h))^2 f(x) dx \\ & = \int (x - E(x))^2 f(x) dx = \text{VAR}[X] \end{aligned} \quad (6.22)$$

E infine $\text{VAR}[X + Y] = \text{VAR}[X] + \text{VAR}[Y] + 2 \text{COV}[X, Y]$ infatti

$$\begin{aligned} \text{VAR}[X + Y] &= E \left[((X + Y) - E(X + Y))^2 \right] \\ &= E \left[((X - E(X)) + (Y - E(Y)))^2 \right] \\ &= E[(X - E(X))^2] + E[(Y - E(Y))^2] + 2 E[(X - E(X))(Y - E(Y))] \\ &= \text{VAR}[X] + \text{VAR}[Y] + 2 \text{COV}[X, Y] \end{aligned} \quad (6.23)$$

Da cui si ricavano le altre.

In maniera simile, osserviamo che per tre variabili X, Y, Z vale

Valgono le seguenti proprietà:

$$\text{COV}[aZ, bX + cY + h] = ab \text{COV}[Z, X] + ac \text{COV}[Z, Y]$$

Anche in questo caso, dimostriamo a passi attraverso alcune proprietà. La relazione $E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$ e' stata dimostrata sopra. Procediamo poi passo passo con altre:

$$\text{COV}[X, Y] = \text{COV}[Y, X] : \quad \text{segue da definizione} \quad (6.24)$$

$$\begin{aligned} \text{COV}[aX, Y] = a \text{COV}[X, Y] : \quad & \text{COV}[aX, Y] = E[aXY] - E[aX]E[Y] \\ & = aE[XY] - aE[X]E[Y] = a \text{COV}[X, Y] \end{aligned} \quad (6.25)$$

$$\begin{aligned} \text{COV}[X + h, Y] = \text{COV}[X, Y] : \quad & \text{COV}[X + h, Y] = E[(X + h)Y] - E[X + h]E[Y] \\ & = E[XY] + E[hY] - (E[X] + h)E[Y] \\ & = E[XY] + hE[Y] - hE[Y] - E[X]E[Y] \\ & = \text{COV}[X, Y] \end{aligned} \quad (6.26)$$

$$\text{COV}[X, X] = \text{VAR}[X, X] : \quad \text{segue da definizione} \quad (6.27)$$

E infine $\text{COV}[X + Y, Z] = \text{COV}[X, Z] + \text{COV}[Y, Z]$, infatti:

$$\begin{aligned} \text{COV}[X + Y, Z] &= E[(X + Y)Z] - E[X + Y]E[Z] \\ &= E[XZ] + E[YZ] - (E[X] + E[Y])E[Z] \\ &= \text{COV}[X, Z] + \text{COV}[Y, Z] \end{aligned} \quad (6.28)$$

Da cui si ricavano le altre.

Esempio:

Data una variabile x distribuita secondo $\mathcal{G}(x; 5; 1)$, si calcoli la PDF propagata di Y dove $y = 2e^x + 1$. Di questa PDF si stimi media e varianza. Successivamente, attraverso le operazioni tra momenti, si calcoli di nuovo media e varianza.

Abbiamo visto che

Esempio:

Si dimostri che per una generica X normale $E[X^3] = \mu^3 + 3\mu\sigma^2$. Si utilizzi il fatto che se $z = \frac{x-\mu}{\sigma}$ allora $E[Z^3] = 0$.

D: Se X normale $Z = \frac{X-\mu}{\sigma}$ e' normale standard: $\mu = E[Z] = 0$, $\sigma^2 = \text{VAR}[Z] = 1$. Inoltre essendo simmetrica vale $E[Z^3] = 0$. Allora

$$E[X^3] = E[(\mu + \sigma Z)^3] = E[\mu^3 + 3\mu^2\sigma Z + 3\mu\sigma^2 Z^2 + Z^3] = \mu^3 + 3\mu\sigma^2 \quad \blacksquare$$

6.4 Formula della propagazione delle incertezze

Riprendiamo ora le linearizzazioni delle PF del caso univariato [Equazione 6.14](#) e multivariato [Equazione 6.15](#). Procediamo ora a calcolare $E(Y) = E(g(X))$ e $\text{VAR}(Y) = \text{VAR}(g(X))$. Qui di seguito, per alleggerire la notazione, useremo i simboli $\mu_X = E(X)$ e $\sigma_X = \text{VAR}(X)$ e analogo per Y .

Caso univariato

Riportiamo [Equazione 6.14](#) qui per comodità

$$g(\mu) + \left(\frac{dg(x)}{dx}\right)_{\mu} (x - \mu) + \frac{1}{2!} \left(\frac{d^2g(x)}{dx^2}\right)_{\mu} (x - \mu)^2 + \mathcal{O}(3) \quad (6.29)$$

Notiamo che possiamo riscrivere questa equazioni *arresentandoci al primo ordine* e scomponendo i termini come:

$$g(x) \simeq g(\mu_X) + \left(\frac{dg(x)}{dx}\right)_{\mu_X} x - \left(\frac{dg(x)}{dx}\right)_{\mu_X} \mu + \mathcal{O}(2) \quad (6.30)$$

In questa forma linearizzata, diventa ora piuttosto semplice, applicando le regole di [sezione 6.3](#), calcolare valore di aspettazione e varianza.

Valore di aspettazione Per quanto riguarda valore di aspettazione, notiamo che

$$E(g(\mu)) = g(\mu); \quad E(X - \mu) = 0 \quad E((X - \mu)^2) = \sigma_X^2$$

e quindi

$$\mu_Y = g(\mu) + \frac{1}{2} \left(\frac{dg(x)}{dx} \right)_{\mu}^2 \sigma_X^2 + \mathcal{O}^2 \quad (6.31)$$

Il risultato puo' sorprendere perche' ci dimostra che il valore di aspettazione di $Y = g(X)$ non e' quello calcolato direttamente come $g(\mu_X)$, ma non deve turbare. Esso riporta il fatto che la relazione $y = g(x)$ in qualche maniera deforma la PF. Di quanto e' questo discostamento? Dipende da σ_X^2 . Ora in linea di principio, nel caso di misure fisiche (che affronteremo in [Capitolo 7](#)), lavoreremo con misure ad alta precisione, almeno migliore di qualche parte per cento. In questa ipotesi, l'incertezza relativa di σ_X^2/μ_X e' sempre piccola, di qualche parte per mille, e quindi la differenza tra μ_Y e $g(\mu_X)$ e' piccola.

Varianza Notiamo che i termini $g(\mu), (dg(x)/dx)_{\mu}, \mu$ sono parametri numerici, allora $\sigma^2(k) = 0$ e quindi:

$$\sigma_Y^2 \simeq \left(\frac{dg(x)}{dx} \right)_{\mu_X}^2 \sigma_X^2 + \mathcal{O}(2) \quad (6.32)$$

Abbiamo quindi trovato che, attorno la varianza attorno a μ_Y e' pari alla varianza attorno a μ_X moltiplicata per la derivata della funzione $g(x)$ rispetto a x calcolata in μ_X . Allora la stima migliore del valore vero di Y e' con ottima approssimazione:

$$\hat{y} = g(\mu_X) \pm \sqrt{\left(\frac{dg(x)}{dx} \right)_{\mu_X}^2 \sigma_X^2}$$

questa e' la **formula di propagazione delle incertezze di X a Y**.

Esempio:

Sia $y = \tan(x)$. Sia X distribuita con valore medio 5 e varianza 1, si calcoli la miglior stime di Y attorno al valore medio di X .

Formula propagazione delle incertezze - caso bivariato

Nel caso la funzione sia bivariata, l'espansione in serie di Taylor procede in maniera analoga attorno alle medie come visto in [Equazione 6.15](#) che riportiamo qui per comodità.

$$z = g(x, y) \sim g(x^*, y^*) + \left(\frac{\partial g(x, y)}{\partial x} \right)_{(x^*, y^*)} (x - x^*) + \left(\frac{\partial g(x, y)}{\partial y} \right)_{(x^*, y^*)} (y - y^*) + \dots \quad (6.33)$$

Come fatto sopra, riscriviamo [Equazione 6.33](#) separando gli addendi e arrendandoci al primo ordine. Ora e' possibile calcolare valore di aspettazione a varianza agilmente.

Valore di aspettazione Come fatto sopra,

$$\mu_Z \simeq g(\mu_X, \mu_Y) + \frac{1}{2} \left(\frac{dg(x)}{dx} \right)_{\mu}^2 \sigma_X^2 + \frac{1}{2} \left(\frac{dg(x)}{dy} \right)_{\mu_Y}^2 \sigma_Y^2 + \mathcal{O}^2 \quad (6.34)$$

e di nuovo l'approssimazione $\mu_Z \simeq g(\mu_X, \mu_Y)$ e' spesso sufficiente.

Varianza Allora:

$$\begin{aligned} \sigma_Z^2 \simeq & \left(\frac{\partial g(x, y)}{\partial x} \right)_{(x^*, y^*)}^2 \sigma_X^2 + \left(\frac{\partial g(x, y)}{\partial y} \right)_{(x^*, y^*)}^2 \sigma_Y^2 \\ & + 2 \left(\frac{\partial g(x, y)}{\partial x} \frac{\partial g(x, y)}{\partial y} \right)_{(x^*, y^*)} \rho_{XY} \sigma_X \sigma_Y + \mathcal{O}(2) \end{aligned} \quad (6.35)$$

Questa e' la **formula di propagazione delle incertezze** σ_X, σ_Y .

Esempio:

Si calcoli VAR con $z = 3x^2 + x e^y$ attorno a $x_0 \dots$

Formula propagazione delle incertezze - caso generale

Nel caso di tre variabili, allora la formula di propagazione si riscrive considerando le correlazioni per coppie di variabili (oppure con la matrice di correlazione)

$$\begin{aligned} \sigma_Z^2 \simeq & \left(\frac{\partial g}{\partial x} \right)_{(***)}^2 \sigma_X^2 + \left(\frac{\partial g}{\partial y} \right)_{(***)}^2 \sigma_Y^2 + \left(\frac{\partial g}{\partial z} \right)_{(***)}^2 \sigma_Z^2 \\ & + 2 \left(\frac{\partial g}{\partial x} \frac{\partial g}{\partial y} \right)_{(***)} \rho_{XY} \sigma_X \sigma_Y + 2 \left(\frac{\partial g}{\partial x} \frac{\partial g}{\partial z} \right)_{***} \rho_{XZ} \sigma_X \sigma_Z + 2 \left(\frac{\partial g}{\partial y} \frac{\partial g}{\partial z} \right)_{***} \rho_{YZ} \sigma_Y \sigma_Z \\ & + \mathcal{O}(2) \end{aligned} \quad (6.36)$$

Riscriviamo infine [Equazione 6.36](#) nel caso generico di un vettore di variabili aleatorie \vec{X} . Si presti attenzione a evitare il doppio conteggio dei termini con le covarianze.

Formula propagazione caso generale La formula (approssimata) di propagazione delle incertezze nel caso generale di X_1, X_2, \dots variabili di input sulla variabile aleatoria di variato $z = g(x_1, x_2, \dots)$ si puo' quindi scrivere:

$$\sigma_g^2 = \sum_i \left(\frac{\partial g}{\partial x_i} \right)_{\vec{x}^*}^2 \sigma_{X_i}^2 + \sum_i \sum_{j \neq i} \left(\frac{\partial g}{\partial x_i} \frac{\partial g}{\partial x_j} \right)_{\vec{x}^*} \rho_{X_i X_j} \sigma_{X_i} \sigma_{X_j} + \mathcal{O}(2) \quad (6.37)$$

dove le derivate parziali sono calcolati nelle miglior stime dei valori di aspettazione delle X_1, X_2, \dots (indicato con \vec{x}^*).

Termine di covarianza Inoltre, in alcune situazioni, il termine di covarianza potrebbe essere complessivamente negativo: basta che le derivate parziali siano positive e vi sia anticorrelazione, o che via sia correlazioni ma una delle due derivate parziali sia negativa. In questo caso, il termine di covarianza negativo potrebbe essere preponderante, e addirittura portare tutta la varianza ad avere segno negativo, cosa senza senso. La situazione non deve stupire, in quanto la formula di propagazione si basa su una linearizzazione di funzione che puo' non essere valida in alcune situazioni. Cio' accade di norma se le incertezza che appaiono nella [Equazione 6.37](#) sono grandi. Ad ogni modo, in questo caso la maniera migliore di procedere e' assumere nullo il termine di covarianza, e motivarlo nella discussione, se non fosse possibile ottenere stime migliori delle incertezze. Si trova qualche esempio di questa situazione negli esercizi di ??.

Livello di confidenza di σ_g^2 Si e' visto come $\sigma_X, \sigma_Y, \dots$ rappresentino i quadrati delle deviazioni standard, e siano quindi riconducibili al concetto di livello di confidenza su X, Y, \dots . A seguito della propagazione approssimata, che livello di confidenza rappresenta σ_g^2 ? La risposta a questa domanda e' complessa e in questo corso la affronteremo solo superficialmente. Essa sarà affrontata poi nei corsi successivi. Per ora basti pensare che nel caso l'approssimazione lineare sia buona e le incertezza su X, Y siano gaussiane, allora, anche σ_g^2 rappresenterà un margine di confidenza gaussiano.

Formula propagazione nel caso correlazione nulla e totale E' interessante notare che nel caso non si conosca la correlazione tra variabili, si possono sempre assumere due valori estremali, ovvero correlazione nulla e correlazione (o anticorrelazione) massima. Nel caso di correlazione nulla, la formula di propagazione delle incertezze è anche detta *combinazione in quadratura delle incertezze* in quanto:

$$\sigma_g^2 \simeq \sum_{i=1}^n \left(\frac{\partial g}{\partial x_i} \right)_*^2 (\sigma_{X_i})^2 \quad (6.38)$$

$$\sigma_g \simeq \sqrt{\left(\frac{\partial g}{\partial x_{1*}} \sigma_{X_1} \right)^2 + \left(\frac{\partial g}{\partial x_{2*}} \sigma_{X_2} \right)^2 + \dots} \quad (6.39)$$

$$(6.40)$$

Nel caso di correlazione totale si formano dei quadrati perfetti in [Equazione 6.37](#) e quindi:

$$\sigma_g \simeq \sum_{i=1}^n \left(\frac{\partial g}{\partial x_i} \right)_{\mu_i} \sigma_{X_i} \quad (6.41)$$

Si veda anche la discussione della propagazione delle incertezze massime piuttosto che di quelle standard in [Appr. 6.5](#).

6.4.1 Esempi di propagazione delle incertezza statistiche

Function	Variance
$g = aX$	$\sigma_g^2 = a^2 \sigma_X^2$
$g = aX + bY$	$\sigma_g^2 = a^2 \sigma_X^2 + b^2 \sigma_Y^2 + 2ab \sigma_{XY}$
$g = aX - bY$	$\sigma_g^2 = a^2 \sigma_X^2 + b^2 \sigma_Y^2 - 2ab \sigma_{XY}$
$g = XY$	$\sigma_g^2 \approx g^2 \left[\left(\frac{\sigma_X}{X} \right)^2 + \left(\frac{\sigma_Y}{Y} \right)^2 + 2 \frac{\sigma_{XY}}{XY} \right]$
$g = \frac{X}{Y}$	$\sigma_g^2 \approx g^2 \left[\left(\frac{\sigma_X}{X} \right)^2 + \left(\frac{\sigma_Y}{Y} \right)^2 - 2 \frac{\sigma_{XY}}{XY} \right]$
$g = aX^b$	$\sigma_g^2 \approx \left(abX^{b-1} \sigma_X \right)^2 = \left(g \frac{b\sigma_X}{X} \right)^2$
$g = a \ln(bX)$	$\sigma_g^2 \approx \left(a \frac{\sigma_X}{X} \right)^2$
$g = a \log_{10}(bX)$	$\sigma_g^2 \approx \left(a \frac{\sigma_X}{X \ln(10)} \right)^2$
$g = ae^{bX}$	$\sigma_g^2 \approx g^2 (b\sigma_X)^2$
$g = a^{bX}$	$\sigma_g^2 \approx g^2 (b \ln(a) \sigma_X)^2$
$g = a \sin(bX)$	$\sigma_g^2 \approx [ab \cos(bX) \sigma_X]^2$
$g = a \cos(bX)$	$\sigma_g^2 \approx [ab \sin(bX) \sigma_X]^2$
$g = a \tan(bX)$	$\sigma_g^2 \approx [ab \sec^2(bX) \sigma_X]^2$
$g = X^Y$	$\sigma_g^2 \approx g^2 \left[\left(\frac{Y}{X} \sigma_X \right)^2 + (\ln(X) \sigma_Y)^2 + 2 \frac{Y \ln(X)}{X} \sigma_{XY} \right]$
$g = \sqrt{aX^2 \pm bY^2}$	$\sigma_g^2 \approx \left(\frac{X}{g} \right)^2 a^2 \sigma_X^2 + \left(\frac{Y}{g} \right)^2 b^2 \sigma_Y^2 \pm 2ab \frac{XY}{g^2} \sigma_{XY}$

Tabella 6.1: Formula propagazione delle incertezze per alcuni funzioni. $\sigma_{XY} = 0$ per X, Y indipendenti o non correlate.

In [Tabella 6.1](#) si riportano alcuni casi classici. Dimostriamo alcune di queste relazioni. Tra tutte queste, che si dimostrano con le proprietà descritte in [sezione 6.3](#) verifichiamo esplicitamente la seguente:

Propagazione per prodotto $Z = aX^\alpha Y^\beta$ con X, Y Nel caso del prodotto di due variabili non correlate X, Y , risulta:

$$\frac{\partial g}{\partial x} = a \alpha x^{\alpha-1} y^\beta = \alpha \frac{g}{x}$$

$$\frac{\partial g}{\partial y} = a \beta x^\alpha y^{\beta-1} = \beta \frac{g}{y}$$

Allora:

$$\sigma_g^2 = \left(\alpha \frac{g}{x}\right)^2 \sigma_X^2 + \left(\beta \frac{g}{y}\right)^2 \sigma_Y^2 \rightarrow \left(\frac{\sigma_g}{g}\right)^2 = \alpha^2 \left(\frac{\sigma_X}{x}\right)^2 + \beta^2 \left(\frac{\sigma_Y}{y}\right)^2 \quad (6.42)$$

E' facile vedere che nel caso siano correlate:

$$\left(\frac{\sigma_g}{g}\right)^2 = \alpha^2 \left(\frac{\sigma_X}{x}\right)^2 + \beta^2 \left(\frac{\sigma_Y}{y}\right)^2 + 2\alpha\beta\rho_{XY} \left(\frac{\sigma_X}{x}\right) \left(\frac{\sigma_Y}{y}\right) \quad (6.43)$$

In sostanza si sommano in quadratura le precisioni (pesate con il coefficiente) delle variabili.

Nella maggior parte dei casi, si procede quindi al calcolo formale delle derivate parziali prime, alla valutazione del termine di correlazione, e alla calcolo diretto della propagazione usando [Equazione 6.37](#).

Esempio:

Si propagano le incertezza delle misure L, T su g secondo la relazione $g = 2\pi^2 \frac{L}{T^2} = k \frac{L}{T^2}$ Si consideri $\sigma_{LT} = 0$.

Usiamo due metodo. Procedendo con le derivate:

$$\frac{\partial g}{\partial L} = \frac{k}{T^2} \quad \frac{\partial g}{\partial T} = -\frac{2kL}{T^3}$$

allora

$$\sigma_g^2 = \frac{k^2}{T^4} \sigma_L^2 + \frac{4k^2 L^2}{T^6} \sigma_T^2$$

Possiamo usare direttamente questa formula. Tuttavia osserviamo che se ora dividiamo entrambi i termini per g^2 troviamo:

$$\frac{\sigma_g^2}{g^2} = \frac{k^2 T^4}{k^2 L^2 T^4} \sigma_L^2 + \frac{4k^2 L^2 T^4}{k^2 L^2 T^6} \sigma_T^2 = \frac{\sigma_L^2}{L^2} + 4 \frac{\sigma_T^2}{T^2}$$

Alla stessa formula potevamo arrivare direttamente pensando alla relazione come al prodotto: $g = k \frac{L}{T^2} = k L T^{-2}$ di due variabili con esponenti 1, -2 rispettivamente.

Allora il valore di incertezza, calcolato nella maniera piu' semplice, da attribuire alla stima di g risulta

$$\sigma_g = g \sqrt{\left(\frac{\sigma_L}{L}\right)^2 + 4 \left(\frac{\sigma_T}{T}\right)^2}$$

6.5 Commenti Finali

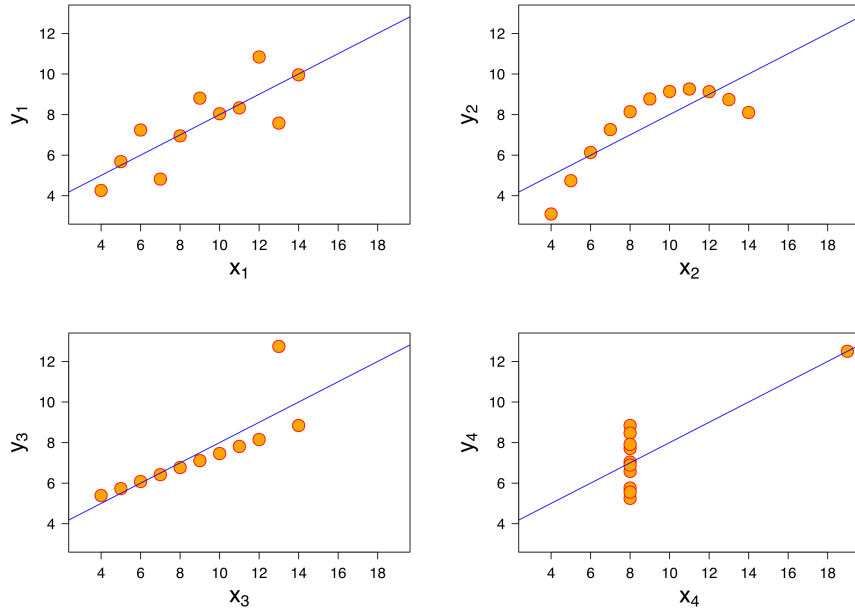
Ricapitolando

Esercizi

Approfondimenti

Appr. 6.1 Limiti della covarianza. Discutiamo qui alcuni casi in cui il concetto di covarianza visto non quantifica sempre correttamente la dipendenza.

- I quattro grafici di [Figura 6.3](#) (set di Anscombe) hanno tutti stessa media (7.5), stessa varianza (4.12), stesso coefficiente di correlazione (0.816) e stessa retta interpolante ($y = 3 + 0.5x$).


 Figura 6.3: Quattro set di dati con lo stesso indice di correlazione $r = 0.816$.

E qui di seguito riportiamo due esempi in cui correlazione nulla non implica indipendenza.

- Si supponga \mathcal{U}_X PDF definita tra $(-1, 1)$, allora $E[X] = 0$. Le due variabili $X, Y = X^2$ sono chiaramente dipendenti, eppure esse hanno correlazione nulla, infatti $COV[X, Y] = E[XY] - E[X]E[Y] = E[X^3] - 0E[Y] = 0$ visto che $f_X(x) = x^3$ e' simmetrica.
- (preso da [Canale and Pietra \[2020\]](#)) Si supponga $\mathcal{U}_X, \mathcal{U}_Y$ PMF definite in $(1, 2, 3, 4, 5, 6)$. Si prenda $S = X + Y$ e $D = X - Y$, allora $E[D] = 0$

$$COV[S, D] = E[SD] - E[S]E[D] = E[SD] - 0 = E[X^2 - Y^2] = E[X^2] - E[Y^2] = 0$$

quindi la covarianza e' nulla, ma certamente S, D non sono indipendenti, infatti S e' sempre pari e D sempre dispari e dunque stocasticamente dipendenti.

Appr. 6.2 Matrice Covarianza. La matrice di covarianza e' una matrice quadrata i cui elementi riportano le covarianza tra coppie di variabili aleatorie. Per questa ragione la matrice e' simmetrica, positiva e sulla diagonale contiene le varianze (che sono le correlazioni di una variabile aleatoria con se stessa). La matrice di covarianza e' anche detta matrice di auto-covarianza, di varianza o di dispersione.

$$COV_{\mathbf{XX}} = \begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1X_2} & \cdots & \sigma_{X_1X_n} \\ \sigma_{X_2X_1} & \sigma_{X_2}^2 & \cdots & \sigma_{X_2X_n} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{X_nX_1} & \sigma_{X_nX_2} & \cdots & \sigma_{X_n}^2 \end{bmatrix} \quad (6.44)$$

Appr. 6.3 Un esempio di PDF completa bivariata. Consideriamo la relazione funzionale $f_{XY} = f_X(x)f_Y(y) = f_X(x)f_Y(z-x)$ Allora

$$f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(x, z-x)dx = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx$$

Prendiamo il caso $f_X(x) = \mathcal{G}(x; \mu_1, \sigma_1^2)$ $f_Y(y) = \mathcal{G}(y; \mu_2, \sigma_2^2)$ otteniamo:

$$f_Z(z) = \int \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} \cdot \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2}\left(\frac{(z-x)-\mu_2}{\sigma_2}\right)^2} dx = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left(\frac{z-(\mu_1+\mu_2)}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)^2}$$

che e' ancora una gaussiana, ovvero $f_Z(z) = \mathcal{G}(z; \mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. Il caso piu' generico $z = ax + by + c$ puo' essere ricavato da questo ricordando che $\mu_1 \rightarrow \mu_1/a, \sigma_1 \rightarrow \sigma_1/a$ e $\mu_2 \rightarrow \mu_2/b, \sigma_2 \rightarrow \sigma_2/b$ e c non influisce.

Supponiamo invece che $f_X(x), f_Y(y)$ siano PDF esponenziali indipendenti di media λ_1, λ_2 . Allora:

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx = \int_0^z \lambda_1 e^{-\lambda_1 x} \lambda_2 e^{-\lambda_2(z-x)} dx \\ &= \lambda_1 \lambda_2 e^{-\lambda_2 z} \int_0^z e^{(\lambda_2 - \lambda_1)x} dx = \frac{\lambda_1 \lambda_2}{\lambda_2 - \lambda_1} (e^{-\lambda_1 z} - e^{-\lambda_2 z}) \end{aligned} \quad (6.45)$$

Esempio:

Si dimostri che $f_Z(z)$ è una PDF. La si disegni per $\lambda_1 = 1, \lambda_2 = 2$. Si calcoli il valore di aspettazione.

Una procedura simile si puo' applicare al caso bivariato con relazione funzionale piu' complessa. Il caso multivariato e il caso con correlazione necessitano di una trattazione piu' avanzata.

Appr. 6.4 PDF marginale. Si supponga di avere una PDF congiunta $f_{XY}(x, y)$ allora, in totale analogia al caso di eventi aleatori descritto in [sottosezione 4.3.2](#) si puo' pensare di calcolare la PDF di una della due variabili aleatorie dopo aver marginalizzato sull'altra. Ad esempio, sia f_{XY} la bigaussiana di [Equazione 3.8](#) allora:

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy \quad \text{e} \quad f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx$$

Si ricorda che la procedura di marginalizzazione corrisponde anche al calcolo della (densita' di) probabilità di X *indipendentemente* da Y , ovvero prendendo qualsiasi caso di Y pesato con la propria probabilità $f_Y(y)$ (e viceversa per X).

Appr. 6.5 Propagazione delle incertezze massime. L'uso statistico della formula [Equazione 6.37](#) si basa sulle deviazioni standard. In linea di principio e' possibile anche propagare la deviazione massima ([Capitolo 2](#)). In questo caso l'incertezza massima su g ovvero Δ_g assume il valore, calcolato in base alla relazione funzionale $g(x_1, x_2, \dots)$ e alle incertezze massime Δ_{X_i} sulle singole variabili:

$$\Delta_g \simeq \sum_{i=1}^n \left| \frac{\partial g}{\partial x_i} \right|_{\mu_i} \Delta_{X_i} \quad (6.46)$$

Si veda [appr. 7.3](#) per un esempio sui campioni di misure.

Capitolo 7

Statistica descrittiva dei campioni di misure

Quest'opera è distribuita con licenza [Creative Commons](#) "Attribuzione – Non commerciale – Condividi allo stesso modo 3.0 Italia".



In questo capitolo (finalmente...!) useremo le conoscenze fin qui raccolte per *descrivere statisticamente* delle misure reali. In questo ambito, le distribuzioni di probabilità del [Capitolo 5](#), sono le distribuzioni di riferimento di una certa serie di misure, ovvero le distribuzioni di dati che risulterebbero se la misura fosse ripetuta nelle stesse condizioni un numero indefinito di volte. In questo caso, il campione di misura si identificherebbe con la *popolazione*¹ di misure. La distribuzione di probabilità della popolazione è quella della variabile aleatoria corrispondente e sulla base di questa si stiamo l'incertezza attraverso la deviazione standard corrispondente della PMF/PDF di riferimento.

Il problema è quindi quello di inferire le proprietà statistiche di una popolazione a partire da un suo campione finito. È chiaro che il campione deve essere *rappresentativo* della popolazione. Ovvero, le sue proprietà estrapolate per numerosità che aumenta devono combaciare con quelle della popolazione. È chiaro che non si può sapere quanto rappresentativo sia il campione, ma è pur vero che impareremo diverse tecniche per quantificare la probabilità che non lo sia. Inoltre, dobbiamo capire quale sia la distribuzione di probabilità di riferimento.

Un concetto importante per comprendere la relazione tra misure e variabili aleatorie è quello di misure (e variabili aleatorie) **Indipendenti e Identicamente Distribuite** (IID, *Independent Identically Distributed*). Prendiamo ad esempio un campione di 100 misure della stessa quantità eseguite in condizioni di ripetibilità. Allora diciamo che abbiamo un vettore di misure $\vec{x} = (x_1, x_2, \dots)$. Ciascuna delle misure x_i è a tutti i sensi un campione di numerosità 1 della variabile aleatoria X associata alla misura. Inoltre, le misure individuali sono nella maggior parte dei casi indipendenti, nel senso che una misura non influenza la successiva (nel caso avvenga così bisogna considerarlo a parte). Allora le misure ripetute sono IID. Con questa definizione, possiamo pensare di calcolare la distribuzione di probabilità congiunta per la serie di misure \vec{x} . Essa sarà, in regola con le regole della teoria di probabilità ([Capitolo 4](#)):

$$f(\vec{x}) = f_X(x_1) \cdot f_X(x_2) \cdot \dots = \prod_{i=1}^n f_X(x_i) \quad \text{se } x_i \text{ sono IID.}$$

Il nostro compito è quindi ora quello di prendere una serie di misure IID e stimarne i vari momenti. Cominciamo dal momento algebrico di ordine zero, il valore di aspettazione in [sezione 7.1](#) per poi passare ai momenti centrali di ordine due, la varianza.

¹I termini campione e popolazione derivano dal fatto che storicamente sono stati introdotti relativamente alla valenza statistica di stime fatte su sottogruppi di persone (campioni) e non sull'intera popolazione, per esempio legate alla validità di una certa cura

7.1 Media e altre stime centrali

La definizione di valore di aspettazione e media per le variabili aleatorie (popolazioni) che abbiamo visto in [Equazione 5.24](#) e' quella della media pesata dei valori secondo la loro distribuzione di probabilità. Se ora abbiamo una serie di misure IID, esse hanno tutta la stessa distribuzioni di riferimento, essendo campioni di numerosità 1 di quella popolazione. Supponiamo che questa popolazione abbia valore di aspettazione μ (si noti che non stiamo facendo ipotesi che sia Gaussiana, quindi $\mu = E[X]$). Il valore di aspettazione e' il valore che ci aspettiamo *in media*. In questo senso, se prendiamo una misura x_i ci aspettiamo in media di misurare μ . Allora possiamo scrivere:

$$E[x_i] = \mu \quad \forall x_i$$

Allora perche' ripetiamo le misure? E perche' di queste prendiamo la media aritmetica? Vedremo durante la discussione che e' una questione di precisione della misura. Ma continuiamo ragionando sulla media. Avevamo definito la media aritmetica N misure genericamente come:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

Come x_i e' un campione di misure unitario della stessa popolazione X allora possiamo pensare a \bar{x} come un campione unitario di una variabile aleatoria \bar{X} definita come combinazione di variabili aleatorie:

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

Ovvero \bar{x}_i e' distribuito secondo $f_{\bar{X}}(\bar{x})$. Conosciamo ora il calcolo dei momenti di combinazioni lineari di variabili aleatorie, come visto in [Capitolo 6](#). Calcoliamo quindi il valore di aspettazione di \bar{X} :

$$E(\bar{X}) = \frac{1}{N} \sum_{i=1}^N E(X_i) = \frac{1}{N} N E(X) = E(X) = \mu \quad (7.1)$$

dove abbiamo sfruttato il fatto che $E(X_i) = E(X)$. Questo ci dice che il valore di aspettazione della popolazione delle medie aritmetiche e' lo stesso delle singole IID. In altre parole, *in media* ci aspettiamo di ottenere come media aritmetica $\bar{x} = \mu$. Attenzione che $f_{\bar{X}}(\bar{x}) \neq f_X(x)$, esse hanno solo lo stesso valore di aspettazione.

Esempio:

Sono state effettuate le seguenti misure di tassi di decadimento (numero di decadimenti nell'unità di tempo dec/s) da un materiale radioattivo:

$$44, 52, 57, 49$$

Si caratterizzi la distribuzioni di probabilità che seguono le misure.

La popolazione di misure segue una statistica Poissoniana (lo assumiamo vero per ora) $\mathcal{P}(n; \lambda) = \lambda^n e^{-\lambda} / n!$. Sappiamo che $E(\mathcal{P}) = \lambda$, e che (ancora senza averlo dimostrato) la media aritmetica e' una buona stima del valore di aspettazione. Allora

$$E(\mathcal{P}) = \lambda = \frac{1}{N} \sum_{i=1}^N x_i = \frac{44 + 53 + 57 + 49}{4} = 50.5 \text{ dec/s}$$

e quindi la distribuzione di probabilità delle misure e':

$$\mathcal{P}(n; 50.5) = \frac{50.5^n e^{-50.5}}{n!}$$

Esempio:

Sono state effettuate le seguenti misure ripetute di durata di un campione di batterie (in ore):

$$13, 12, 15, 14, 12, 13, 11$$

Si caratterizzi la distribuzioni di probabilità che seguono le misure.

Si può assumere che la popolazione di misure segua una statistica Gaussiana (non possiamo a rigore dimostrarlo dai dati, ma lo assumiamo vero per ora) $\mathcal{G}(x; \mu, \sigma)$. Sappiamo che $E(\mathcal{G}) = \mu$, e che $E(\bar{X}) = E(X)$ allora

$$E(X) = \lambda = \frac{1}{N} \sum_{i=1}^N x_i \sim 12.86 \text{ h}$$

Questo valore ci basta per stabilire il parametro μ , ma come stabilire il parametro σ ?

7.1.1 Deviazione standard della media campionaria

Come mai usiamo allora usiamo la media aritmetica \bar{x} invece che la misura singola x_i se tanto $E(\bar{X}) = E(X)$? La risposta è chiara: mentre hanno lo stesso valore di aspettazione, l'incertezza sulla media è diversa dall'incertezza sulla singola misura, come abbiamo già discusso in [Capitolo 3](#) senza dimostrarlo. Facciamolo ora, ma prima soffermiamoci su questo fatto. L'incertezza sul valore di aspettazione di una popolazione è nullo:

$$\text{VAR}(E(X)) = 0 \quad \text{infatti } E(X) = \mu \quad \text{e} \quad \text{VAR}(\mu) = 0 \quad \blacksquare$$

cio' non è vero invece per *la sua stima* ovvero la media aritmetica, infatti:

$$\text{VAR}(\bar{X}) = \text{VAR} \left[\frac{1}{N} \sum_{i=1}^N (X_i) \right] = \frac{1}{N^2} \sum_{i=1}^N \text{VAR}(X_i) = \frac{1}{N^2} N \text{VAR}(X) = \frac{\text{VAR}(X)}{N} \quad \blacksquare \quad (7.2)$$

Si noti bene cosa abbiamo verificato. Usando la definizione di IID, come visto per il valore di aspettazione, anche $\text{VAR}(X_i) = \text{VAR}(X)$. In questo senso la varianza della popolazione può essere considerata come *incertezza della singola misura*. Se ora abbiamo N misure distribuite IID secondo X allora abbiamo verificato che l'incertezza sulla media aritmetica è minore della incertezza sulla singola misura per un fattore N pari alla numerosità del campione. Ci manca da capire come calcolare $\text{VAR}(X)$ ovvero la incertezza della singola misura di un campione di misure. Si ribadisce nuovamente che X è una generica RAV.

7.1.2 Varianza e deviazione standard campionaria

La varianze della principali distribuzioni con cui lavoreremo sono riportate in [Tabella 5.2](#). In particolare, vediamo da quella tabella che per la maggiore parte dei casi la varianza è legata al valore di aspettazione da qualche relazione, ad esempio per Poisson $\text{VAR}[\mathcal{P}] = E[\mathcal{P}] = \lambda$. Esiste un solo caso in cui la varianza è legata ad un secondo parametro indipendente, ed' il caso della Gaussiana. Allora dovremo trattare i due casi separatamente. Per la prima classe di casi, il calcolo della varianza campionaria sarà immediato. Ad esempio

Esempio:

Sono state effettuate le seguenti misure di tassi di decadimento (numero di decadimenti nell'unità di tempo dec/s) da un materiale radioattivo:

44, 52, 57, 49

Si stimi la varianza delle misure.

Abbiamo visto prima che

$$\text{VAR}(\mathcal{P}) = \lambda = \frac{1}{N} \sum_{i=1}^N x_i = \frac{44 + 53 + 57 + 49}{4} = 50.5 \text{ dec/s}$$

Una simile procedura si puo' applicare per le altre distribuzioni di probabilità. Ma come fare per la Gaussiana? Sappiamo che la varianza della popolazione e' il valore degli aspettazioni degli scarti al quadrato. Questo concetto, trasposto al caso del campione, ci porterebbe a stimare come incertezza la (radice della) media degli scarti quadratici, ovvero il MSE:

$$MSE = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

Ora, il quadrato del valore sopra e' quindi una stima della varianza della popolazione di riferimento. Tuttavia, come fatto sopra, se pensiamo in termini di IID allora possiamo pensare ai vari MSE come campioni di dimensioni 1 della popolazione ottenuta con tutti i campioni di numerosità N . In questa condizione avrebbe senso calcolare il valore di aspettazione di questo parametro. Facciamolo.

$$\begin{aligned} E(MSE) &= \frac{1}{N} E \left[\sum_{i=1}^N (X_i - \bar{X})^2 \right] = \frac{1}{N} E \left[\sum_{i=1}^N (X_i + (\mu - \mu) - \bar{X})^2 \right] = \frac{1}{N} E \left[\sum_{i=1}^N ((X_i - \mu) + (\mu - \bar{X}))^2 \right] \\ &= \frac{1}{N} \sum_{i=1}^N \underbrace{E[(X_i - \mu)^2]}_{\text{VAR}[X]} + \frac{1}{N} \sum_{i=1}^N \underbrace{E[(\mu - \bar{X})^2]}_{\text{VAR}[\bar{X}]} + \frac{2}{N} \sum_{i=1}^N \underbrace{E[(X_i - \mu)(\mu - \bar{X})]}_{-\text{COV}[X_i, \bar{X}]} \\ &= \frac{N}{N} \text{VAR}(X) + \frac{N}{N} \text{VAR}(\bar{X}) - \frac{2}{N} \frac{N}{N} \text{VAR}(\bar{X}) \\ &= \text{VAR}(X) - \text{VAR}(\bar{X}) \end{aligned} \quad (7.3)$$

dove abbiamo sfruttato il fatto che le X_i sono IID e quindi hanno la stessa varianza, e che vale

$$\text{COV}[X_i, \bar{X}] = \underbrace{\text{COV} \left[X_i, \sum_i \frac{X_i}{N} \right]}_{\text{COV}[X_i, X_j]=0} = \text{COV} \left[X_i, \frac{X_i}{N} \right] = \frac{1}{N} \text{VAR} X_i = \frac{1}{N} \text{VAR} X$$

Osserviamo allora come MSE sia sistematicamente inferiore (una sottostima) della varianza della popolazione per un termine costante pari a $\text{VAR}(\bar{X})$. Quale e' allora la stima migliore di deviazione standard campionaria?

Partendo da [Equazione 7.3](#) e usando [Equazione 7.2](#) otteniamo che

$$E(MSE) = \text{VAR}(X) - \text{VAR}(\bar{X}) = \text{VAR}(X) - \frac{\text{VAR}(X)}{N} = \frac{N-1}{N} \text{VAR}(X) \quad (7.4)$$

allora

$$\text{VAR}(X) = \frac{N}{N-1} E(MSE) \quad (7.5)$$

si ricordi che quindi questo e' un comportamento in media, allora la miglior stima della varianza della popolazione è:

$$s_x^2 = \frac{N}{N-1} MSE = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 = \frac{N}{N-1} \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (7.6)$$

e di conseguenza la miglior stima campionaria di incertezza e' pari alla deviazione standard campionaria per un insieme di misure ripetute con comportamento normale:

$$s_x = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (7.7)$$

Esempio:

Sono state effettuate le seguenti misure ripetute di durata di un campione di batterie (in ore): 13, 12, 15, 14, 12, 13, 11. Si caratterizzi la distribuzioni di probabilità che seguono le misure. Abbiamo visto prima che la media e' pari a $\bar{x} = 12.86$ h. La deviazione standard campionaria:

$$s_x = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2} \sim 1.34$$

Allora la popolazione di riferimento e' $\mathcal{G}(x; 12.86, (1.34)^2)$ e la miglior stima del valore vero è:

$$x = (12.9 \pm 1.3) \text{ h}$$

Nel caso di un'altra PF ad esempio quella poissoniana:

Esempio:

Sono state effettuate le seguenti misure di tassi di decadimento (numero di decadimenti nell'unità di tempo dec/s) da un materiale radioattivo: 44, 52, 57, 49. Si stimi la varianza delle misure. Abbiamo visto prima che la miglior stima del valore medio e della varianza è: $\lambda = 50.75$ dec/s Allora la miglior stima di incertezza per la misura e' (Equazione 7.2):

$$s_{\bar{x}} = \frac{s_x}{\sqrt{N}} = \frac{\sqrt{50.5}}{\sqrt{4}} \sim 3.55 \text{ dec/s}$$

Allora la nostra miglior stima del valore medio e':

$$\lambda = (50.5 \pm 3.6) \text{ dec/s}$$

Si noti che se avessimo preso solo la prima misura, allora la miglior stima del valore sarebbe stata: $\lambda = (44.0 \pm 6.6) \text{ dec/s}$ in quanto $\sqrt{44} \sim 6.6$.

E' interessante notare, rispetto all'esempio precedente, cosa sarebbe successo se avessimo misurato una sola volta per 4 secondi invece che 4 volte per 1 secondo. Supponiamo di averlo fatto, avremmo trovato 202 conteggi. Quale sarebbe stato l'errore? Sarebbe stato la radice quadrata di quel conteggio ovvero $\sqrt{202} \sim 14.2$ dec/4s. Sfruttando le regole per le relazioni di variabili [Capitolo 6](#) avremo calcolato la stima di decadimenti *al secondo* usando la relazione $y = g(x) = x/4$ e quindi $\sigma_Y = 1/4\sigma_X$ ovvero:

$$\lambda = (50.5 \pm 3.6) \text{ dec/s}$$

Esempio:

Si pensi ora a misure ripetute di oscillazione del pendolo. Si confronti il caso in cui si misurano 4 periodi singoli e 1 singola misura di 4 periodi. Le due misure hanno stessa precisione?

L'incertezza della singola misura e' uguale nei due casi, diciamo che vale T . Nel primo caso di misure ripetute, l'incertezza della media sar :

$$s_{\bar{T}} = \frac{sT}{\sqrt{4}}$$

Nel secondo caso, l'incertezza sulla singola misure, ottenuta secondo la relazione $T_1 = T_4/4$ e', usando la propagazione delle incertezze:

$$s_{T_1} = \frac{sT}{4}$$

La seconda misura e' piu' precisa (si noti, nel caso siano sempre nulli gli effetti sistematici sulla misura!)

Si noti che una formulazione alternativa di della deviazione standard campionaria e'

$$s_x^2 = \frac{1}{N-1} \sum_{i=1}^N x_i^2 - N\bar{x} \tag{7.8}$$

7.1.3 Incertezza su varianza e deviazione standard campionarie

Abbiamo visto che per la media di una popolazione di misure μ l'incertezza e' pari a 0. Cosa li e' per la varianza o la deviazione standard della popolazione. Ma cosa accade per quella campionaria?

Attraverso le formule viste in precedenza, e notando che la varianza campionaria puo' essere pensata come una combinazione lineare di variabili aleatorie e parametri, allora si puo' dimostrare che

$$s_{s_x^2} = \sqrt[4]{\frac{2}{N-1}} s \quad s_{s_x} \simeq \frac{s_x}{\sqrt{2(N-1)}} \quad s_{s_{\bar{x}}} \simeq \frac{s_{\bar{x}}}{\sqrt{2N}} \tag{7.9}$$

L'ultima relazione sopra ci permette ora di validare la ragione per cui si suggerisce di usare una unica o al piu' due cifre significative per la incertezza della media campionaria. Infatti, se calcoliamo l'errore relativo della media

$$\frac{s_{s_{\bar{x}}}}{s_{\bar{x}}} \simeq \frac{1}{\sqrt{2N}}$$

Allora l'incertezza relativa per $N = 10, 50, 100$ e' rispettivamente del 7%, 1.4% e 0.7%. Con 50 misure si e' quindi al limite dell'1% di precisione. Non ha quindi senso con meno di 50 misure di riportare due cifre significative, visto che la seconda (per cento) non e' nota accuratamente.

Si rifletta su come di nuovo al limite della popolazione ($N \rightarrow \infty$) queste incertezza tendano a 0, come ci aspettavamo.

7.1.4 Asimmetria, Curtosi, quantili e mediana dei campioni

Con ragionamenti e procedure simili a quanto visto sopra si possono calcolare asimmetria e curtosi campionarie. Il calcolo algebrico e' piuttosto laborioso e di solito viene fatto attraverso il calcolatore. Si rimanda per una discussione all'appr. 7.1.

La mediana campionaria invece e' definita come il quantile $q_{0.5}$ o il 50esimo percentile e quindi:

$$\tilde{x} = \begin{cases} x_{N/2+1/2} & \text{se } N \text{ pari} \\ \frac{x_{N/2} + x_{N/2+1}}{2} & \text{se } N \text{ dispari} \end{cases} \tag{7.10}$$

Per quanto riguarda il computo dei quantili per campioni, essendo definiti come quei valori che separano in parti uguali la distribuzione dei dati, sono facilmente calcolabili *ordinando* i dati per valore crescente e considerando i dati che corrispondono alle posizioni N/k -esime (N numerosita' campione, k ordine del quantili).

Esercizio:

Si prenda la serie di misure

$$3, 6, 7, 8, 8, 10, 13, 15, 16, 20$$

e si calcolino i quartili. Essendo le misure 10, il primo quartile e' il valore $10/4 = 2.5$ -esimo, che possiamo arrotondare al terzo, il secondo e' il $2 \cdot 10/4 = 5$ (quinto), il terzo e' $3 \cdot 10/4 = 7.5$ -esimo che possiamo arrotondare all'ottavo. Ovvero: $q_0 = 3$; $q_{0.25} = 7$; $q_{0.5} = 10$; $q_{0.75} = 15$; $q_4 = 20$. Chiaramente questo e' un caso poco interessante.

In linea di principio si potrebbe anche mediare tra dati successivi, ad esempio per il primo quartile, potremmo stimare la media del secondo e terzo dato: $q_{0.25} = (6 + 7)/2 = 6.5$. Questo modo ha lo svantaggio del fatto che 6.5 non e' uno dei valori presi. Si presti attenzione nel caso i dati siano distribuiti in classi di frequenza. Si vedano ad esempio gli esercizi di ??.

7.2 Correlazione campionaria

Nel caso dei campioni, la stima della covarianza della popolazione σ_{xy} e' data dalla covarianza campionaria che indichiamo con la lettera s_{xy} .

Covarianza campionaria

$$s_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) \tag{7.11}$$

Il fattore $1/(N-1)$ si ottiene in analogia con lo stesso fattore della varianza (vedi dimostrazione in Appr.

Coefficiente di (correlazione lineare campionaria di) Pearson (Bravais)

$$r_{xy} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \tag{7.12}$$

Facile da verificare. Si riporta inoltre una maniera alternativa di calcolare la covarianza campionaria usando una analoga relazione per la covarianza delle variabili aleatorie gia' vista in precedenza:

$$s_{xy} = \frac{1}{N-1} \sum_{i=1}^N x_i y_i - \frac{1}{N(N-1)} \sum_{i=1}^N x_i \sum_{i=1}^N y_i \tag{7.13}$$

Incertezza sulla coefficiente di correlazione Il coefficiente di correlazione ρ_{XY} e' calcolato a partire da diverse combinazioni (sottrazioni, prodotti, rapporti) di valori di aspettazione di variabili aleatorie. Esso stesso e' quindi una variabile aleatoria, per la quale e' possibile calcolare la distribuzione di probabilita' $f_{\rho_{XY}}$. Non ci interessa tutta la funzione, ma per ora ci basti ricordare che e' possibile calcolare la varianza, e da questa la deviazione standard e per il campione di misure lo SE. Useremo questo valore piu' avanti, per cui e' bene riportarlo:

$$SE_{\rho_{xy}} = \sqrt{\frac{1 - r_{XY}^2}{N - 2}} \tag{7.14}$$

dove N e' la numerosità del campione. Si ricorda come il coefficiente di correlazione per variabili aleatorie e' invece noto senza incertezza come discusso in [Equazione 6.13](#).

7.2.1 Esempi

Tuttavia vedremo ora che la stima di correlazione campionaria e' argomento che merita un approfondimento. Anzitutto partiamo da una considerazione, e ci basiamo su un esempio concreto. Supponiamo di avere due grandezze che crediamo essere correlate, ad esempio velocità e tempo nel modo uniformemente accelerato. Se indichiamo con X la variabile aleatoria riferita al tempo t e Y la variabile aleatoria riferita alla velocità, allora $v(t) = v_0 + at$ dove a e' l'accelerazione. Abbiamo visto in [Equazione 6.12](#) che se la dipendenza e' lineare allora $\rho_{XY} = \pm 1$. Ora noi prendiamo dei campioni di misure di X, Y ad esempio N misure congiunte di spazio e tempo. Cosa ci aspettiamo? Ci aspettiamo che siano allineati su una retta *se l'esperimento e' riuscito!* Ovvero quello che si vuole fare è *verificare* che ci sia dipendenza lineare. Nel caso del moto uniformemente accelerato lo sappiamo a livello teorico e l'unica maniera in cui questa risultato non torni e' quello in cui ci siano *errori sperimentali*. In altri casi invece (piu' interessanti) potremmo non saperlo! Ad esempio peso e altezza di un campione di persone sono correlati linearmente? Con che grado di correlazione lineare.

Un ulteriore esempio comune e' quello 'multivariato' di una misura tipo $z = g(x, y)$ ad esempio il rischio r di infarto è una funzione dell'altezza a , del peso p e dei livelli di colesterolo c : $r = r(a, p, c)$. Se si prendono misure di altezza, peso e colesterolo in un campione di persone per valutare r si deve tenere in conto del fatto che altezza e peso sono correlate nella stima della incertezza s_r ([Equazione 6.36](#)). In questo caso tuttavia non abbiamo una teoria dietro la correlazione tra altezza e peso, dobbiamo misurarla sperimentalmente. A complicare il tutto c'e' anche che una seconda correlazione tra il peso e i livelli di colesterolo.

Si ricordi quindi che [Equazione 7.12](#) è una *stima* del coefficiente delle popolazioni. Vediamo ora diversi casi in cui questa stima puo' essere fatta in maniera pratica.

Caso singola misura Supponiamo di prendere una unica misura di tempo t_1 e velocità v_1 , oppure una unica misura di altezza peso e colesterolo a_1, p_1, c_1 . Chiaramente in questo caso non potrei calcolare [Equazione 7.12](#) e quindi non posso stimare il grado di correlazione campionario. Potrebbe essere che le popolazioni lo fossero ma non sono in grado di apprezzarlo. In questo caso non solo non posso stimare le incertezza sul modello ma non posso nemmeno costruirlo. Per un punto passano infinite rette nel primo caso. E con un unico individuo non posso calcolare il rischio.

Caso misure ripetute In questo caso prendiamo campioni con numerosità N ad esempio di t, v nel primo caso e a, p, c nel secondo caso. Ora possiamo calcolare [Equazione 7.12](#). Per praticità facciamo un esempio numerico del primo caso (del secondo sarebbe piu' complesso). Le misure di tempo e velocità sono riportate in tabella. Si considerano le misure di tempo con incertezza trascurabile mentre le misure di velocità hanno tutte dispersione standard 0.05 ms^{-1} . [Figura 7.1](#) mostra le distribuzioni di dati per due incertezze standard di v : $s_v = 0.5$ (sinistra) e $s_v = 1.2$ (destra).

t_i [s]	v_i [ms^{-1}]
1.1	1.2
1.9	2.1
2.8	2.5
3.9	3.5
4.9	5.5

In questo caso $r_{tv} \sim 0.9703$ ad indicare un alto grado di correlazione campionaria. L'incertezza su questo valore e' $s_r \sim 0.019$. Si noti che l'incertezza s_v non entra nel calcolo del coefficiente di correlazione. Questo rappresenta un limite per il calcolo di correlazione campionaria, perchè naturalmente ci rendiamo conto che nel caso con maggiore incertezza (a destra in [Figura 7.1](#)) ci possiamo aspettare che il margine di confidenza sul fatto che i dati siano allineati attorno ad una retta sia maggiore. Questo margine di confidenza sara' costruito in [Capitolo 9](#) e [Capitolo 10](#). Si noti che nonostante ciò l'incertezza s_r e' comunque piccola in proporzione.

In conclusione, si noti quindi come nonostante la popolazione di riferimento avesse correlazioni $\rho = +1$, a seguito di campionamenti, e degli errori casuali legati al campionamento, abbiamo stimato un valore inferiore di correlazione campionaria. Si noti anche come della correlazioni si puo' solo fare un sottostima (in valore assoluto) e mai una sovrastima, essendo questa superiormente limitata a 1.

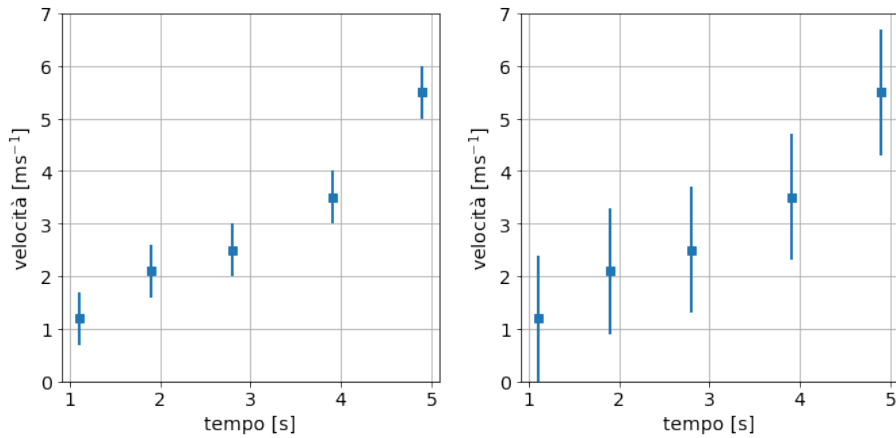


Figura 7.1: Distribuzione della serie di misure in tabella per due incertezze sulla velocità: $s_v = 0.5$ (sinistra) e $s_v = 1.2$ (destra).

Caso generale Si osservi [Figura 7.2](#). Nella figura si ha un campionamento di due variabili aleatorie (due grandezze) X, Y con diversi gradi di correlazione. Prendiamo le prime due righe di grafici. Troviamo diversi casi di correlazione o anticorrelazione totale, $\rho = \pm 1$. Questi corrispondono grossomodo ai casi di [Figura 6.2](#) (seconda riga). Si tratta quindi di variabili linearmente correlate del tipo $y = a + bx$. Ma a cosa corrisponde il caso $\rho = 0.8$ o $\rho = 0.4$? In base alla discussione sopra si ricordi che per il calcolo della correlazione campionaria o non sappiamo quale sia la vera correlazione della popolazione o vogliamo verificarlo. Allora diciamo che i casi di correlazione intermedia della prima riga con $-0.8 < r_{XY} < 0.8$ rappresentano casi in cui *non possiamo affermare con grande margine di confidenza* che vi sia correlazione lineare. Questo però può significare diverse cose: *a)* le due serie non sono effettivamente correlate *b)* le due serie non sono linearmente correlate ma lo sono in altra maniera *c)* le due serie sono effettivamente linearmente correlate ma le incertezze di misura sono così elevate (o il campionamento troppo basso) da non permetterci di osservare questo effetto.

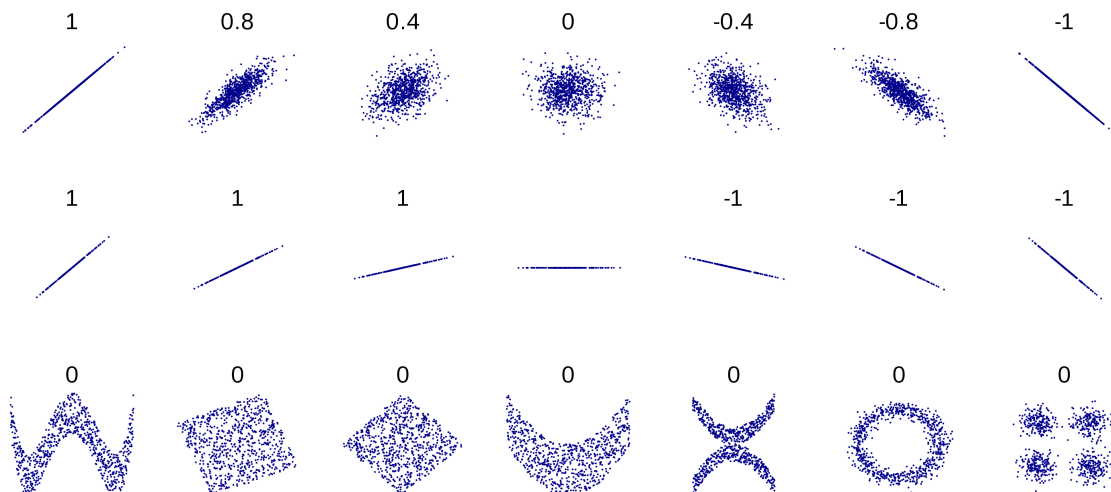


Figura 7.2: Esempi di correlazione lineare positiva, negativa e nulla tra due variabili aleatorie. Credit: Wikipedia.

Nel caso di [Figura 7.2](#) per la riga inferiore invece si ha il caso già discusso di campioni con una più o meno chiara dipendenza logica, e tuttavia con correlazione nulla, come discusso in [sezione 6.1](#).

L'importanza del coefficiente di correlazione campionario è legata principalmente, ma non solo, al computo della propagazione delle incertezze che discutiamo qui di seguito.

7.3 Propagazione delle incertezze campionarie

In [Capitolo 6](#) abbiamo discusso come calcolare media e varianza di una funzione di variabili aleatorie che potesse essere linearizzata con buona approssimazione attorno al valore di aspettazione. Come si ragiona ora in termini di campioni? Riguardo la formula Eq. ??, ricordiamo come nel caso dei campioni nè il valore di aspettazione nè la varianza è nota con assoluta precisione. Queste sono solo stimate. Inoltre, abbiamo ora una interpretazione diversa della formula stessa, essendo $\text{VAR}(X), \text{VAR}(Y)$ una stima della dispersione delle misure, allora attraverso Eq. ?? si stima la dispersione di Z attorno al valore centrale $\text{VAR}(Z)$. La formula quindi prende il senso di *formula di propagazione delle incertezze*. Allora nel linguaggio dei campioni di misure possiamo scrivere

Formula di propagazione degli errori casuali attorno al punto (x_0, y_0)

$$s_{z_0}^2 \simeq \left(\frac{\partial g}{\partial x}\right)_{(x^*, y^*)}^2 s_{x_0}^2 + \left(\frac{\partial g}{\partial y}\right)_{(x^*, y^*)}^2 s_{y_0}^2 + 2 \left(\frac{\partial g}{\partial x} \frac{\partial g}{\partial y}\right)_{(x^*, y^*)} r_{XY} s_{x_0} s_{y_0} + \mathcal{O}(2) \tag{7.15}$$

dove (x^*, y^*) rappresentano i valori di aspettazione di X, Y , ovvero i valori attorno ai quali le deviazioni standard sono calcolate.

Dove si è usato il simbolo x^*, y^* come miglior stima di μ_x, μ_y , s_{x^*}, s_{y^*} come stima di σ_x, σ_y , e r_{xy} come stima di ρ_{XY} . Si noti ora che se si tratta di una misura singola allora $x_0 = x_1, y_0 = y_1$ e s_x, s_y sono le incertezze SE di queste misure. Se invece si tratta di misure ripetute allora $x_0 = \bar{x}, y_0 = \bar{y}$ e $s_{x^*} = s_{\bar{x}, s_{y^*} = s_{\bar{y}}}$ sono le incertezze SEM dei valori medi. La formula è valida se l'approssimazione in serie di Taylor è valida. È chiaro che i termini del tipo $(x - x^*)^n$ rappresentano i discostamenti rispetto alla miglior stima del valore vero x^* . Se consideriamo questo intervallo come l'intervallo di incertezza attorno al valore vero, questi intervalli rappresentano le incertezze di una misura. È chiaro che se la precisione è alta, l'errore relativo è basso, e quindi se questo è del 10% (scarsa precisione) già il termine $(x - x^*)^2$ è dell'ordine dell'1%, e $(x - x^*)^3$ dell'0.1%. Questo ci permette di affermare che se la precisione è migliore dell'1%, per quanto riguarda la propagazione delle incertezze, le due funzioni $g(x)$ e $\left(\frac{df(x)}{dx}\right)_{x^*} (x - x^*)$ differiscono di meno dell'1%.

Ricordiamo infine di cosa intendiamo qui per correlazioni? Intendiamo correlazioni nelle incertezze nelle variabili. Queste possono derivare da due fonti diverse.

Innanzitutto, gli errori di misurazione possono essere correlati anche se i valori veri sottostanti non lo sono. Ad esempio, l'evaporazione da un campione di acqua aumenterà le concentrazioni di tutti i soluti (e quindi creerà una correlazione nelle concentrazioni misurate), anche se le concentrazioni nel flusso non sono correlate.

In secondo luogo, quando i valori sottostanti sono correlati in una popolazione, le incertezze nelle medie del gruppo saranno correlate, anche se le incertezze nei valori individuali non lo sono. Ad esempio, il rischio di infarto è una funzione dell'altezza, del peso e dei livelli di colesterolo. Altezza e peso sono correlati tra gli individui, così come il peso e i livelli di colesterolo. Quindi le incertezze nell'altezza media, nel peso e nei livelli di colesterolo di un gruppo di persone saranno correlate a causa dell'errore di campionamento. Cioè, se il peso medio del gruppo campionato è maggiore della media reale per la popolazione, anche le altezze medie e i livelli di colesterolo saranno probabilmente superiori alle medie reali e la correlazione dovrebbe essere presa in considerazione nel valutare l'incertezza nel rischio medio di infarto per un gruppo. Pertanto, le propagazioni degli errori per le incertezze nelle medie di gruppo devono tenere conto di entrambi i tipi di correlazioni. Fortunatamente, un grafico a dispersione dei singoli valori misurati mostrerà gli effetti combinati di entrambe le correlazioni. Ma se stiamo cercando di stimare l'incertezza nel rischio di infarto per un singolo individuo, allora solo il primo tipo di correlazione conta. L'altezza, il peso e il colesterolo di un individuo sono correlati tra loro, ma le incertezze non lo sono. Poiché queste misurazioni vengono effettuate in modo indipendente, gli errori di misurazione di per sé non sono correlati.

7.3.1 Esempio

Esempio:

Si vuole misurare una resistenza in un circuito usando la relazione $R = \frac{V}{I} \cos(\phi)$. Si prendono 5 misure ripetute, riassunte in tabella.

V [V]	I [A]	ϕ [rad]
5.007	0.019663	1.0456
4.994	0.019639	1.0438
5.005	0.01964	1.0468
4.99	0.019685	1.0428
4.999	0.019678	1.0433

Date le misure, calcolare la miglior stima di R e della sua incertezza.

Ci sono due maniere di risolvere il problema. La prima, usando la propagazione degli errori, la seconda, usando la distribuzione dei risultati.

Primo metodo: propagazione incertezza delle medie In questo caso, per il calcolo del valore centrale::

$$R = \frac{\bar{V} \cos \bar{\phi}}{\bar{I}} = 127.73 \Omega$$

Per il calcolo della incertezza, si usa la formula di propagazione delle incertezza [Equazione 6.36](#), usando la relazione $R = g(V, I, \phi)$ e chiamando $*** = \bar{V}, \bar{I}, \bar{\phi}$:

$$s_R^2 \simeq \left(\frac{\partial g}{\partial V}\right)_{***}^2 s_V^2 + \left(\frac{\partial g}{\partial I}\right)_{***}^2 s_I^2 + \left(\frac{\partial g}{\partial \phi}\right)_{***}^2 s_\phi^2 + 2\left(\frac{\partial g}{\partial V} \frac{\partial f}{\partial I}\right)_{***} r_{VI} s_V s_I + 2\left(\frac{\partial g}{\partial V} \frac{\partial f}{\partial \phi}\right)_{***} r_{V\phi} s_V s_\phi + 2\left(\frac{\partial g}{\partial I} \frac{\partial f}{\partial \phi}\right)_{***} r_{I\phi} s_I s_\phi$$

dove $r_{I\phi}, r_{V\phi}, r_{IV}$ sono i coefficienti campionari di Pearson.

Passiamo ora a calcolare i diversi termini. Le derivate parziali calcolate nei punti medi sono:

$$\frac{\partial g}{\partial V} = \frac{\cos \phi}{I} = 25.6; \quad \frac{\partial g}{\partial I} = -\frac{V \cos \phi}{I^2} = -6496.7; \quad \frac{\partial g}{\partial \phi} = -\frac{V \sin \phi}{I} = -219.8$$

Le deviazioni standard delle medie: $s_{\bar{V}} = 3.21 \cdot 10^{-3}$ V; $s_{\bar{I}} = 9.47 \cdot 10^{-6}$ A; $s_{\bar{\phi}} = 7.52 \cdot 10^{-4}$ rad. Le covarianze campionarie $s_{VI} = -5.4 \cdot 10^{-8}$ VA; $s_{V\phi} = 1.035 \cdot 10^{-5}$ V rad; $s_{I\phi} = -2.298 \cdot 10^{-8}$ A rad, da cui le correlazioni $r_{VI} = -0.355$; $r_{V\phi} = 0.858$; $r_{I\phi} = -0.645$ Da cui $s_R = 0.0711 \Omega$ e quindi la nostra stima e'

$$R = (127.732 \pm 0.071) \Omega \quad (\text{Metodo 1})$$

Si noti che nel caso avessimo trascurato i termini di correlazione, avremo stimato $s_z = 0.195 \Omega$.

Metodo diretto Con questo dataset avremmo anche potuto calcolare in maniera diretta i singoli valori R_i per ogni set di misure e fatta la media,

	V [V]	I [A]	ϕ [rad]	$R = V \cos \phi / I$ [Ω]
1	5.007	0.019663	1.0456	127.67
2	4.994	0.019639	1.0438	127.89
3	5.005	0.01964	1.0468	127.51
4	4.99	0.019685	1.0428	127.71
5	4.999	0.019678	1.0433	127.88
			Media	127.73

Come si vede, avremmo ottenuto lo stesso valore che nel primo metodo. Avremo poi potuto facilmente calcolare la deviazione standard della media dei 5 valori ottenendo $s_{\bar{R}} = 0.0713 \Omega$. Come si vede, c'è solo una piccola discrepanza rispetto al caso sopra. Avremo comunque stimato lo stesso valore

$$R = (127.732 \pm 0.071) \Omega \quad (\text{Metodo 2})$$

Identico entro gli errori. Attenzione, questo secondo metodo è esatto in questo caso, tuttavia, abbiamo un caso particolare in cui abbiamo tre set di dati campionati in contemporanea per cui possiamo calcolare R_i . Ciò non avviene ad esempio se campioniamo indipendentemente le tre grandezze di input oppure se abbiamo numerosità diverse nelle diverse grandezze di input.

7.4 Note aggiuntive

Quesiti

□ **Es. 7.1** Una misura è riportata con una precisione del 2.2% a seguito di 80 ripetizione. Quante ripetizioni devo fare per portare la precisione sotto l'1%?. Quale è l'incertezza sulla deviazione standard?

□ **Es. 7.2** Calcolare l'indice di correlazione (lineare) campionario e il coefficiente di correlazione lineare di (Bravais-)Pearson tra le serie di dati

X	Y
1.1	1.21
1.9	3.61
2.8	7.84
3.9	15.21
4.9	25

□ **Es. 7.3** Si supponga di avere due produzioni di tubi di ferro, A e B, ciascuno formato da 100 tubi. Si ipotizza che i tubi siano realizzati con alta precisione. Per verificarlo, si decide di fare un confronto a campione tra le due produzioni. Della prima si estraggono alcuni e si misurano le lunghezze, ottenendo la distribuzione in tabella. Per la seconda serie, si ripetono altrettante misure e si ottiene come stima $L_B = 5.014 \pm 0.004$.

- Stimare media e deviazione standard della prima serie di misure;
- Riportare nella forma corretta la stima della media della prima misura e della incertezza;
- Scrivere la formula analitica della distribuzione gaussiana che si può associare alla prima serie di misure;
- Verificare se ci sono dati misurati che si trovano al di là di 3 deviazioni standard dalla media.
- Stimare la compatibilità con il valore L_B

$L_A[m]$	Frequenza assoluta
5,10	2
5,11	9
5,12	14
5,13	8
5,14	3

Es. 7.4.1 . [Sada] Siano x e y variabili casuali continue definite tra i numeri reali positivi non nulli: sia $z = \frac{2x - y}{y}$. Esprimere formalmente la varianza di z in $x = x_0$ e $y = y_0$.

Si effettuano 100 misure ripetute della grandezza x e y ottenendo un valor medio $\bar{x} = (1,00 \pm 0,01)$ e $\bar{y} = (1,00 \pm 0,02)$ rispettivamente: esprimere la varianza di z per $x_0 = 2$ e $y_0 = 1$ considerando x e y statisticamente indipendenti.

Es. 7.4.2 . Siano X, Y due variabili aleatorie indipendenti con supporto in \mathcal{R} ciascuna. Sia definita la relazione funzionale $z = 2\sqrt{x} + \sin y$ tra x, y .

Delle variabili x, y abbiamo delle misure ripetute: 100 misure ripetute di x danno come stima 1.03 ± 0.02 e 50 misure ripetute di y danno come stima 0.00 ± 0.04 rad. 2.1) Scrivere la formula della propagazione degli errori per il calcolo della deviazione standard della variabile aleatoria z in un punto, nella sua forma generale, completa al primo ordine, che includa il termine di eventuale covarianza, 2.2) stimare la varianza di z attorno alle due medie $(x_0, y_0) = (\bar{x}, \bar{y})$ trovate; 2.3) stimare la varianza di z attorno ai punti $(x_0, y_0) = (1.00, 0.00)$;

Es. 7.4.3 . Siano X, Y due variabili aleatorie definite in $[0, A]$ dove A è l'ultimo numero di matricola. Si consideri la variabile aleatoria Z definita come

$$z = x^3 + y^2 x$$

Le due serie X, Y sono correlate con covarianza $COV[X, Y] = 0.1$. Si vuole studiare la propagazione delle incertezze su x, y su z su due punti: sul punto medio di x, y ottenuti entrambi con 50 misurazioni e aventi come risultato

$$\bar{x} = A/3 \pm 0.05 \quad \bar{y} = A/3 \pm 0.03$$

e poi su un punto della stessa serie, ottenuto con misura singola in una successiva misura, pari a:

$$x_0 = A/2 \quad y_0 = A/2$$

Con le informazioni sopra a) Si scriva in forma generica la legge della propagazione degli errori b) Si calcoli la miglior stima di z e la sua incertezza nel punto x_0, y_0 c) Si calcoli la miglior stima di z e la sua incertezza nel punto \bar{x}, \bar{y} . Si discuta la validità dell'uso della legge di propagazione nei due casi

Es. 7.4.4 . Sia

$$W(x, y, z) = Aye^x + yz$$

una grandezza dipendente da x, y e z grandezze fisiche definite in \mathbb{R} , A sia parametro dato. Si effettuano 20 misure ripetute di x e z e 15 misure ripetute di y ottenendo come valori medi

$$\bar{x} = 1.00 \pm 0.01; \quad \bar{y} = 2.00 \pm 0.02; \quad \bar{z} = 3.00 \pm 0.03$$

con x, y, z covarianti tali che $COV[X, Y] = 0.01$ e $COV[X, Z] = 0.001$. Si riporti: a) la miglior stima di W a partire dalle misure effettuate; b) l'incertezza casuale attesa per W in un generico punto x_0, y_0, z_0 entro la relativa deviazione standard dalla miglior stima di x, y e z sia con che senza i termini di covarianza, discutendone il risultato; c) l'incertezza da associare a $W(1.01, 2.00, 3.00)$ nel caso in cui x e z siano adimensionali, y abbia le unità di misura di m/s

Approfondimenti

Appr. 7.1 Asimmetria e Curtosi campionaria. Con ragionamenti del tutto analoghi a quanto visto per la varianza si puo' calcolare che una stima accurata a della asimmetria m_3 della popolazione. Questa, con il linguaggio di questo capitolo e':

$$m_3 = \lim_{N \rightarrow \infty} \frac{1}{(s^2)^3} \frac{\sum_{i=1}^N (x_i - \mu)^3}{N} = \frac{1}{(s^2)^3} E(x-\mu)^3 \quad (7.16)$$

allora una sua stima campionaria e':

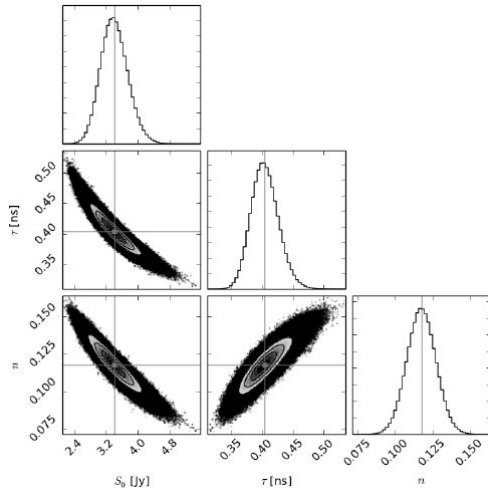
$$a = \frac{1}{s_x^3} \frac{\sum_{i=1}^N (x_i - \bar{x})^3}{\frac{(N-1)(N-2)}{N}} \quad (7.17)$$

e analogamente una stima accurata c dell'eccesso di curtosi m_4 della popolazione e':

$$c = \frac{1}{s_x^4} \left(\frac{\sum_{i=1}^N (x_i - \bar{x})^4}{\frac{(N-1)(N-2)(N-3)}{N}} - \frac{3 \left(\sum_{i=1}^N (x_i - \bar{x})^2 \right)^2}{(N-2)(N-3)} \right) - 3 \quad (7.18)$$

data la lunghezza algebrica (ma non complessità) del calcolo, solo raramente lo faremo manualmente, ma useremo librerie di calcolo nella macchine di calcolo.

Appr. 7.2 Diagramma a triangolo.



Esempio di diagramma a triangolo.

Si consideri il diagramma a lato in cui si hanno 3 variabili aleatorie rappresentate da n, τ, S_0 . Il significato dei simboli non e' importante in questo contesto. Il diagramma a triangolo mostra sulla diagonale le distribuzioni di probabilità delle singole variabili mentre sulla parte interna del triangolo mostra a due a due le correlazioni tra le suddette variabili. Ad esempio, nella prima colonna troviamo in alto il grafico di correlazione tra τ e S_0 . Osserviamo come si abbia una forte anticorrelazione. Subito sotto, troviamo il grafico di correlazione tra n e S_0 . Osserviamo anche qui come si abbia una forte anticorrelazione. Nella seconda colonna in basso troviamo il grafico di correlazione tra n e τ . In questo caso si tratta di buona correlazione.

Appr. 7.3 . Propagazione errori massimi Nelle misure indirette, se ho solo errori massimi, posso comunque usare una formula di propagazione? Si. Il risultato sara' comunque una sovrastima rispetto all'errore statistico, come visto in precedenza. Quello che non si puo' fare e' combinare incertezze massime e standard. La ragione e' che combinando questo tipo di incertezze non si e' in grado di computare il margine di confidenza.

Per la propagazione degli errori massimi passo comunque attraverso una espansione di Taylor attorno al punto di interesse arrestandomi al primo ordine:

$$z = g(x, y) = f(x_0, y_0) + f'_x(x_0, y_0)(x - x_0) + f'_y(x_0, y_0)(y - y_0) + \mathcal{O}^2$$

Se ora considero direttamente $\Delta_x = x - x_0, \Delta_y = y - y_0$ come errori massimi allora $\Delta_g = g(x, y) - g(x_0, y_0)$ e' errore massimo di $z = g(x, y)$. Nella propagazione devo prendere valori assoluti delle derivate per evitare compensazioni tra termini positivi e negativi

$$\Delta_g \simeq |f'_x(x_0, y_0)| \Delta_x + |f'_y(x_0, y_0)| \Delta_y$$

Rispetto alla propagazione degli errori statistici risultera' sempre $\Delta_g \geq \sigma_g$.

Ad esempio per somma o differenza $z = x$ allora

$$\Delta_g \simeq |f'_x(x_0, y_0)| \Delta_x + |f'_y(x_0, y_0)| \Delta_y = \Delta_x + \Delta_y$$

Esempio:

Ho misurato un mattone di lati: $x = 10 \pm 1$ cm; $y = 20 \pm 1$ cm; $z = 5 \pm 1$ cm (dove 1 cm errore massimo). Quanto vale errore massimo sul volume del mattone? Risposta: $V = xyz = (10 \cdot 20 \cdot 5) = 1000$. Errore:

$$\Delta_V = \left| \frac{dV}{dx} \right| \Delta_x + \left| \frac{dV}{dy} \right| \Delta_y + \left| \frac{dV}{dz} \right| \Delta_z = |yz| \Delta_x + |xy| \Delta_y + |xz| \Delta_z = 350$$

E quindi

$$V = 1000 \pm 350 \text{ cm}^3 = 1.00 \pm 0.35 \text{ dm}^3$$

Appr. 7.4 Disuguaglianza di Bienaymé–Čebyšef. Il seguente testo e' riprodotto da M. Loreti

Sia una variabile casuale x , e siano $E(x)$ e σ^2 la speranza matematica e la varianza della sua popolazione; vogliamo ora determinare la probabilità che un valore di x scelto a caso differisca (in valore assoluto) da $E(x)$ per più di una assegnata quantità (positiva) ϵ . Questa è ovviamente data, in base alla legge della probabilità totale (??), dalla

$$\Pr(|x - E(x)| \geq \epsilon) = \sum_{|x_i - E(x)| \geq \epsilon} p_i$$

dove la sommatoria è estesa solo a quei valori x_i che soddisfano a tale condizione. Ora, sappiamo che

$$\sigma^2 = E\{[x - E(x)]^2\} = \sum_i p_i [x_i - E(x)]^2;$$

se si restringe la sommatoria ai soli termini x_i che differiscono (in modulo) da $E(x)$ per più di ϵ , il suo valore diminuirà o, al massimo, rimarrà invariato: deve risultare insomma

$$\sigma^2 \geq \sum_{|x_i - E(x)| \geq \epsilon} p_i [x_i - E(x)]^2 \geq \sum_{|x_i - E(x)| \geq \epsilon} p_i \epsilon^2 = \epsilon^2 \sum_{|x_i - E(x)| \geq \epsilon} p_i$$

e da questa relazione si ottiene la *disuguaglianza di Bienaymé-Čebyšef*

$$\Pr(|x - E(x)| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2} \quad (7.19)$$

e, se si pone $\epsilon = k\sigma$,

$$\Pr(|x - E(x)| \geq k\sigma) \leq \frac{1}{k^2} \quad (7.20)$$

(se nella dimostrazione si sostituissero le frequenze relative alle probabilità e la media aritmetica ad $E(x)$, si troverebbe che una analoga relazione vale anche per ogni campione di valori sperimentali x_i rispetto alla media aritmetica \bar{x} ed alla varianza del campione s^2).

La (7.20) fissa un *limite superiore* per la probabilità esaminata, limite che deve valere per *qualsiasi* variabile casuale; con $k \leq 1$ non si ottiene alcuna informazione significativa da essa, ma con $k > 1$ si vede che il maggiorante della probabilità tende a zero all'aumentare di k . In particolare, per *qualsiasi* variabile casuale la probabilità di uno scarto dal valore medio non inferiore in valore assoluto a 2σ non può superare $\frac{1}{4} = 25\%$; e quella di uno scarto non inferiore in valore assoluto a 3σ non può superare $\frac{1}{9} \approx 11.1\%$.

Si deve notare che non si è fatta alcuna ipotesi sulla distribuzione, a parte l'esistenza della sua varianza σ^2 e della sua speranza matematica $E(x)$; in termini così generali il limite superiore (7.20) non può essere ridotto, ma non è escluso che (per una particolare distribuzione) la probabilità per la variabile da essa descritta di differire dal suo valore medio sia più piccola ancora di quella fissata dalla disuguaglianza di Bienaymé-Čebyšef. Ad esempio, se esiste finita la quantità

$$\mu_4 = E\{[x - E(x)]^4\}$$

(momento del quarto ordine rispetto alla media), con passaggi analoghi si troverebbe che

$$\Pr\{[x - E(x)]^4 \geq \epsilon\} \leq \frac{\mu_4}{\epsilon^4}$$

e, quindi, che

$$\Pr\{[x - E(x)]^4 \geq k^4 \sigma^4\} \leq \frac{\mu_4}{k^4 \sigma^4}.$$

Imponendo altre condizioni (anche non molto restrittive) alla distribuzione di probabilità, si potrebbe ridurre ulteriormente (in quantità anche notevole) il limite superiore stabilito in generale dalla (7.20); e stimare così anche la probabilità di uno scarto della variabile casuale dal suo valore medio inferiore a σ . Risale ad esempio a Gauss (1821) la dimostrazione che per una variabile continua avente distribuzione unimodale (con massimo in x_0), e per la quale esista finita la quantità $\sigma_0^2 = E[(x - x_0)^2]$, la probabilità di uno scarto dalla moda x_0 non inferiore in valore assoluto ad una quantità prefissata non può superare la frazione $\frac{4}{9}$ del limite di Bienaymé-Čebyšef:

$$\Pr\{|x - x_0| \geq k\sigma\} \leq \frac{4}{9k^2}.$$

Se la distribuzione è anche *simmetrica*, moda e media coincidono entrambe col centro di simmetria; e σ_0 è uguale alla deviazione standard σ . Per distribuzioni di questo genere, quindi, il limite superiore per la probabilità di uno scarto che non sia inferiore a k volte l'errore quadratico medio scende a $\frac{4}{9} \approx 44.4\%$ per $k = 1$; a $\frac{1}{9} \approx 11.1\%$ per $k = 2$; ed a $\frac{4}{81} \approx 4.9\%$ per $k = 3$ (e vedremo poi nel paragrafo ?? che per le misure affette da errori puramente casuali i limiti superiori sono ancora più stringenti di questi).

Adesso applichiamo la (7.19) alla variabile casuale \bar{x} , media aritmetica di un campione di dimensione N di valori che supponiamo essere statisticamente indipendenti:

$$\Pr(|\bar{x} - E(\bar{x})| \geq \epsilon) \leq \frac{\sigma_{\bar{x}}^2}{\epsilon^2}; \quad (7.21)$$

ma valendo, per questa variabile casuale, le

$$E(\bar{x}) = E(x) \quad \text{e} \quad (\bar{x}) = \frac{\sigma^2}{N},$$

sostituendo nella (7.21) otteniamo

$$\Pr(|\bar{x} - E(x)| \geq \epsilon) \leq \frac{\sigma^2}{N\epsilon^2}. \quad (7.22)$$

Ora, scelti comunque due numeri positivi ϵ e δ , si può trovare in conseguenza un valore di N per cui il secondo membro della (7.22) risulti sicuramente minore di δ : basta prendere $N > M = \lceil \sigma^2 / (\delta \epsilon^2) \rceil$. In base alla definizione (??), questo significa che vale il

TEOREMA (DI ČEBYŠEF): *il valore medio di un campione finito di valori di una variabile casuale qualunque converge statisticamente, all'aumentare della dimensione del campione, alla speranza matematica per quella variabile.*

Appr. 7.5 Dimostrazione covarianza campionaria. (Scritto da Michele Allegra) Sia

$$s_{XY} = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})$$

la covarianza campionaria, con $\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$, $\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$. Il valore atteso di s_{XY} è

$$\begin{aligned} E[s_{XY}] &= \frac{1}{N-1} \sum_{i=1}^N E[(X_i - \bar{X})(Y_i - \bar{Y})] = \frac{1}{N-1} \sum_{i=1}^N E[(X_i - \bar{X})(Y_i - \bar{Y})] = \\ &= \frac{1}{N-1} \sum_{i=1}^N \left(E[X_i Y_i] - E[\bar{X} Y_i] - E[\bar{Y} X_i] + E[\bar{X} \bar{Y}] \right) = \\ &= \frac{1}{N-1} \sum_{i=1}^N \left(E[X_i Y_i] - \frac{1}{N} \sum_j E[X_j Y_i] - \frac{1}{N} \sum_j E[X_i Y_j] + \frac{1}{N^2} \sum_{jk} E[X_j Y_k] \right) \end{aligned}$$

Poiché X_i, Y_j sono indipendenti per $i \neq j$, abbiamo $\text{COV}[X_i, Y_j] = \delta_{ij} \text{COV}[X, Y]$ e dunque $E[X_i Y_j] = \delta_{ij} \text{COV}[X, Y] + \mu_X \mu_Y$. Usando quest'ultima formula otteniamo:

$$\begin{aligned}
 E[s_{XY}] &= \frac{1}{N-1} \sum_{i=1}^N \left(\text{COV}[X, Y] + \mu_X \mu_Y - \frac{1}{N} \sum_{j=1}^N (\delta_{ij} \text{COV}[X, Y] + \mu_X \mu_Y) - \right. \\
 &\quad \left. \frac{1}{N} \sum_{j=1}^N (\delta_{ij} \text{COV}[X, Y] + \mu_X \mu_Y) + \frac{1}{N^2} \sum_{j=1}^N \sum_{k=1}^N (\delta_{jk} \text{COV}[X, Y] + \mu_X \mu_Y) \right) = \\
 &= \frac{1}{N-1} \sum_{i=1}^N \left(\text{COV}[X, Y] + \mu_X \mu_Y - \frac{2}{N} (\text{COV}[X, Y] + N \mu_X \mu_Y) + \frac{1}{N^2} (N \text{COV}[X, Y] + N^2 \mu_X \mu_Y) \right) = \\
 &= \frac{1}{N-1} \sum_{i=1}^N \left(\text{COV}[X, Y] + \mu_X \mu_Y - \frac{2}{N} \text{COV}[X, Y] - 2 \mu_X \mu_Y + \frac{1}{N} \text{COV}[X, Y] + \mu_X \mu_Y \right) = \\
 &\quad = \frac{1}{N-1} \sum_{i=1}^N \left(\frac{N-1}{N} \text{COV}[X, Y] \right) = \text{COV}[X, Y]
 \end{aligned}$$

Parte II

STATISTICA INFERENZIALE

Capitolo 8

Il metodo dei minimi quadrati e la interpolazione (lineare)

Quest'opera è distribuita con licenza [Creative Commons](#) "Attribuzione – Non commerciale – Condividi allo stesso modo 3.0 Italia".



Qui di seguito, ci avviciniamo al problema della stima dei parametri, ovvero al problema di determinare quale tra le famiglia di curve parametriche $g(x; \vec{\lambda})$ rappresenti meglio un set di dati misurati. Il problema e' detto genericamente **interpolazione** o *regressione* o *fitting*. Ad esempio, se ci aspettiamo che le misure siano distribuite su una retta $y = a + bx$, quali sono i migliori parametri a, b della retta che descrivono i dati? Affronteremo il problema con una scorciatoia (e solo nel caso particolare della interpolazione lineare) arrivando a definire un metodo particolare, quella dei minimi quadrati, che ci permette di rispondere alla domanda che ci siamo posti. Ritorneremo su questo problema risolvendolo in forma generale in [Capitolo 9](#).

8.1 Minimi quadrati

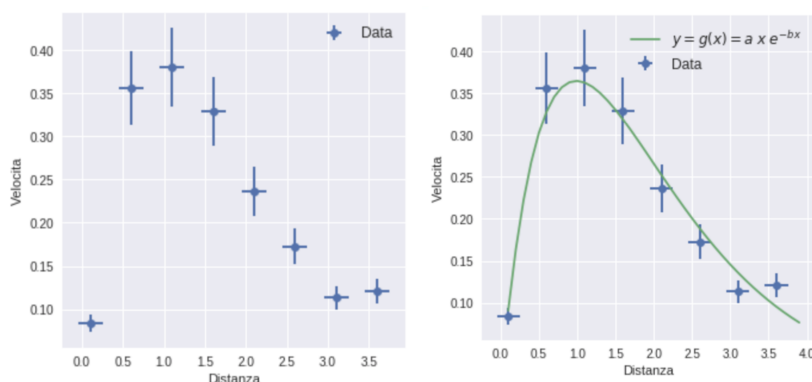


Figura 8.1: Un esempio di distribuzione di dati (sinistra) di cui si vuole fare una interpolazione con un certa funzione (destra)

Si supponga di avere due serie di misure X, Y di cui si vuole valutare se siano legate attraverso la relazione $y = g(x; \vec{\lambda})$. Ad esempio $y = a \sin(x) + b$ oppure $y = a + bx + cx^2$. Se disponiamo i campioni di misure su un grafico cartesiano, avremo una serie di coppie di punti (x_i, y_i) ciascuno con la propria incertezza (s_{x_i}, s_{y_i}) . In linea generale, la natura delle incertezze, normalmente descritta da una deviazione standard, puo' dipendere da una distribuzione di probabilita' o da un'altra. Nel seguito di

questo capitolo, faremo alcune ipotesi importanti, che restringono il campo di azione. Rappresentano comunque il caso piu' frequente:

1. Assumiamo che X, Y siano misure soggette unicamente a fluttuazioni casuali e quindi (s_{x_i}, s_{y_i}) siano deviazione standard della Gaussiana
2. Assumiamo che la misura X sia nota con maggior precisione della misura Y : $\epsilon_Y \gg \epsilon_X$.

La prima ipotesi come abbiamo visto avviene spesso. Cerchiamo infatti dove sia possibile di osservare le fluttuazioni casuali per avere una stima precisa del valore centrale. La seconda ipotesi avviene di norma quando X e' una misura diretta e Y e' una misura derivata. Per le misure derivate si usa infatti la propagazione delle incertezze che di norma, combinando diverse incertezze, ne aumenta il valore finale. La cosa puo' comunque essere controllata date le serie di dati x_i, y_i misurati.

Dopo queste ipotesi abbiamo in sostanza quattro serie di dati $x_i, y_i, s_{x_i}, s_{y_i}$ e ci proponiamo come visto di trovare i parametri della funzione interpolante $y = g(x; \vec{\lambda})$ che meglio approssimano i dati (Figura 8.1). Il cambiare i parametri della funzione (attenzione, non la funzione) ne cambia la forma. A livello intuitivo e' chiaro che la scelta di parametri che *meglio describe* i dati e' quella che fa passare la funzione *piu' vicina* ai dati.

Come misuriamo la vicinanza dei dati con la funzione? Ci sono due punti importanti da comprendere. Diciamo che i punti della funzione sia y_i^* per differenziarli dai punti misurati y_i .

- Non e' importante solo la distanza assoluta (ad esempio sulla Y) $d_i = y_i - y_i^*$ in quanto facendo cosi' si trascura il ruolo della incertezza. Due valori possono avere d_i grande ma avere grande incertezza ed essere compatibili. L'incertezza va sempre tenuta in considerazione e quindi la distanza va pesata con questa $d_i = (y_i - y_i^*)/s_{y_i}$
- Se si hanno incertezze su entrambe le variabili X, Y allora la distanza minima e' funzione sia di X che di Y e corrisponde ad un segmento diagonale sul grafico. Le due grandezze X, Y sono chiaramente correlate, quindi il calcolo della distanza pesata con le incertezze e' complesso. Per questo di norma si trascura l'incertezza sulla misura in ascissa ovvero si pone $s_{x_i} = 0 \forall x_i$. Torneremo su questo alla fine del capitolo.

La regressione quindi si basa sul calcolo di queste distanze e successivamente sul trovare il set di parametri $\vec{\lambda}$ che danno una funzione piu' vicina ai dati. Se si trova allora si ha una *stima dei parametri* della funzione. Per le stime, come gia' visto si usano i simboli $\hat{\lambda}_1, \hat{\lambda}_2, \dots$

Prima di continuare chiediamo a Francesca come mai i punti misurati non sono sulla funzione? Francesca potrebbe rispondere:

Perche' ci sono le fluttuazioni sperimentali della misura...

Giusto! Beatrice aggiungerebbe subito

... si ma anche quelle sistematiche

Esatto anche questo, ma non solo! Avremo anche potuto fare un errore nell'assumere che i dati seguissero questa o quella funzione $y = g(x)$. Il metodo che pero' ora proponiamo funziona ugualmente: e' sempre possibile calcolare la distanza tra punti e linee. Torneremo in [Capitolo 10](#) sulla valutazione di quanto corretto fosse usare l'ipotesi $y = g(x)$ in primo luogo.

Approssimazione incertezze trascurabili nelle ascisse Partiamo dal caso in cui le incertezze sulla variabile in ascissa sono trascurabili. Per il calcolo della distanza minima procediamo ora in maniera analoga al calcolo della compatibilita' (vedi ??) tra i punti e il valore di riferimento. Quale e' il valore di riferimento?

- per le x_i il valore di riferimento e' il valore stesso $x_i^* = x_i$ in quanto sono state assunte senza errore
- per le y_i il valore di riferimento e' quindi $y_i^* = g(x_i; \vec{\lambda})$

e quindi allora la distanza pesata con le incertezze tra punti e retta sarà:

$$d_i = \frac{y_i - y_i^*}{s_{y_i}}$$

il valore d_i sopra può essere positivo o negativo, a seconda del segno del numeratore. E' quindi conveniente prendere il quadrato:

$$d_i^2 = \left(\frac{y_i - y_i^*}{s_{y_i}} \right)^2$$

Se prendiamo ora la somma dei d_i^2 , essa sarà, una volta fissati i valori di riferimento y_i^* un numero:

$$\sum_i d_i^2 = \sum_i \left(\frac{y_i - y_i^*}{s_{y_i}} \right)^2 \quad (8.1)$$

Soffermiamoci però su una considerazione. Abbiamo dato per scontato che y_i^* fossero noti, ma supponiamo che invece i valori di riferimento siano parametrici ovvero $y_i^*(\vec{\lambda})$. Allora [Equazione 10.1](#) può essere pensata come ad una funzione di $\vec{\lambda}$ avendo come parametri x_i, y_i che sono misure già effettuate.

$$Q(\vec{\lambda}; x_i, y_i) = \sum_i \left(\frac{y_i - y_i^*(\vec{\lambda})}{s_{y_i}} \right)^2 \quad (8.2)$$

E' chiaro che valori diversi di $\vec{\lambda}$ danno valori diversi della funzione. In particolare, ci saranno dei valori $\widehat{\lambda}_1, \widehat{\lambda}_2, \dots$ che rendono *minima* la funzione Q . Per questi valori, vista la definizione di Q , risulta minime lo distanze tra i punti misurati e i punti di riferimenti della retta (pesati con incertezze). Ora per trovare i minimi di una funzione (anche di più variabili) e' sufficiente trovare i valori che annullano tutte le derivate parziali prime (e rendono negativa le derivate parziali seconde), ovvero:

$$\begin{cases} \frac{\partial Q}{\partial \lambda_1}(\widehat{\lambda}_1) = 0 \\ \frac{\partial Q}{\partial \lambda_2}(\widehat{\lambda}_2) = 0 \\ \dots \end{cases} \quad (8.3)$$

e

$$\frac{\partial^2 Q}{\partial \lambda_1^2}(\widehat{\lambda}_1) > 0 \quad \frac{\partial^2 Q}{\partial \lambda_2^2}(\widehat{\lambda}_2) > 0$$

...

Se e' nota la funzione Q bastera' quindi fare il calcolo esplicito delle derivate, come vedremo a breve nel caso particolare della interpolazione lineare.

Caso incertezze nelle ascisse non trascurabili In questo caso il problema e' legato al fatto che per il calcolo del valore di riferimento y_i^* anche le incertezze sulle x_i contano. Inoltre le incertezze sulle y_i sono correlate con le incertezze sulle x_i . Una maniera approssimata di risolvere il problema e' di trascurare la correlazione nel calcolo di y_i^* e considerare come stima di distanza, invece che [Equazione 8.1](#)

$$d_i^2 = \left(\frac{y_i - y_i^*}{\sqrt{s_{y_i}^2 + s_{x_i}^2}} \right)^2$$

dove

$$s_i^2 \simeq s_{y_i}^2 + \frac{\partial g}{\partial x} s_{x_i}^2$$

8.2 Interpolazione lineare

Continuiamo quindi lavorando su un caso concreto, quello in cui la funzione e' una retta $y = g(x) = a + bx$, allora

$$Q(a, b; x_i, y_i) = \sum_i \left(\frac{y_i - (a + b x_i)}{s_{y_i}} \right)^2 \quad (8.4)$$

Procediamo per livelli di complessità.

Caso incertezze tutte uguali Cominciamo dal caso in cui tutti gli $s_{y_i} = s_y$ sono uguali. Allora per la stima della miglior intercetta \hat{a} :

$$\frac{\partial Q}{\partial a} = \sum_i -2 \frac{y_i - (a + b x_i)}{s_y^2} = 0 \quad \rightarrow \quad \sum_i y_i - N a - b \sum_i x_i = 0$$

e quindi

$$\hat{a} = \frac{1}{N} \sum_i y_i - b \frac{1}{N} \sum_i x_i = \bar{y} - b \bar{x}$$

Inoltre

$$\frac{\partial^2 Q}{\partial a^2}(\hat{a}) = \frac{2}{s_y^2} > 0$$

Il risultato ci dice che il valore che minimizza la somma degli scarti quadratici e' \hat{a} in effetti tale che $\bar{y} = \hat{a} + b \bar{x}$. Ci manca tuttavia da definire anche il valore di \hat{b} che minimizza anche per il secondo parametro. Allora come prima:

$$\frac{\partial Q}{\partial b} = \sum_i -2 \frac{b x_i y_i - (a + b x_i) x_i}{s_y} = 0 \quad \rightarrow \quad \sum_i x_i y_i - a \sum_i x_i - b \sum_i x_i^2 = 0$$

Sostituendo ad a il valore calcolato prima:

$$\begin{aligned} \sum_i x_i y_i - \left(\frac{1}{N} \sum_i y_i - b \frac{1}{N} \sum_i x_i \right) \sum_i x_i - b \sum_i x_i^2 &= 0 \\ b \left(\sum_i x_i^2 - \frac{1}{N} \sum_i x_i \sum_i x_i \right) &= \sum_i x_i y_i - \frac{1}{N} \sum_i x_i \sum_i y_i \end{aligned}$$

e quindi

$$\hat{b} = \frac{\sum_i x_i y_i - \frac{1}{N} \sum_i x_i \sum_i y_i}{\sum_i x_i^2 - \frac{1}{N} (\sum_i x_i)^2}$$

Si noti che (vedi [Equazione 7.8](#) e [Equazione 7.13](#)) vale anche $\hat{b} = \frac{s_{XY}}{s_X^2}$. Se ora definiamo

$$\Delta = N \sum_i x_i^2 - \left(\sum_i x_i \right)^2 ;$$

Possiamo riscrivere \hat{a}, \hat{b} come:

$$\begin{aligned} \hat{a} &= \frac{1}{\Delta} \left[\sum_i x_i^2 \sum_i y_i - \sum_i x_i \sum_i x_i y_i \right] \\ \hat{b} &= \frac{1}{\Delta} \left[N \sum_i x_i y_i - \sum_i x_i \sum_i y_i \right] \end{aligned}$$

I due valori rappresentano quindi **i valori dei parametri che danno i minimi quadrati** di Q .

Si noti che i due valori sono ottenuti solo con combinazioni di dati x_i, y_i e quindi sempre calcolabili! In quanto combinazioni si possono anche stimare le loro incertezze con le tecniche di [Capitolo 7](#). In questo caso si ottiene (no dim)

$$\sigma_{\hat{a}} = \sigma_y \sqrt{\frac{\sum_i x_i^2}{\Delta}} \quad \text{e} \quad \sigma_{\hat{b}} = \sigma_y \sqrt{\frac{N}{\Delta}}$$

e quindi riassumendo:

Interpolazione lineare: caso I I parametri che meglio approssimano una serie di punti in cui la variabile in ascissa abbia incertezza trascurabile e la variabile in ordinata abbia incertezze gaussiane tutte uguali, la cui distribuzione sia assunta lineare con $y = a + bx$ ottenuti con il metodo dei minimi quadrati:

$$\Delta = N \sum_i x_i^2 - \left(\sum_i x_i \right)^2 ;$$

$$\hat{a} = \frac{1}{\Delta} \left[\sum_i x_i^2 \sum_i y_i - \sum_i x_i \sum_i x_i y_i \right] \quad s_{\hat{a}} = \sigma_y \sqrt{\frac{\sum_i x_i^2}{\Delta}} \quad (8.5)$$

$$\hat{b} = \frac{1}{\Delta} \left[N \sum_i x_i y_i - \sum_i x_i \sum_i y_i \right] \quad s_{\hat{b}} = \sigma_y \sqrt{\frac{N}{\Delta}} \quad (8.6)$$

Incetnze sulle ordinate diverse Il caso in cui $s_{y_i} = s_y$ sono uguali e' poco generico. Il caso piu' generico $s_{y_i} \neq s_{y_j}$ si risolve tuttavia in maniera analoga a quanto visto sopra. Numericamente le formule sono piu' complicate ma sempre calcolabili esattamente dalle misure.

Interpolazione lineare: caso II I parametri che meglio approssimano una serie di punti in cui la variabile in ascissa abbia incertezza trascurabile e la variabile in ordinata abbia incertezze gaussiane diverse, la cui distribuzione sia assunta lineare con $y = a + bx$ ottenuti con il metodo dei minimi quadrati:

$$\Delta = \sum_i \frac{1}{s_{y_i}^2} \sum_i \frac{x_i^2}{s_{y_i}^2} - \left(\sum_i \frac{x_i}{s_{y_i}} \right)^2 ;$$

$$\hat{a} = \frac{1}{\Delta} \left[\sum_i \frac{x_i^2}{s_{y_i}^2} \sum_i \frac{y_i}{s_{y_i}^2} - \sum_i \frac{x_i}{s_{y_i}^2} \sum_i \frac{x_i y_i}{s_{y_i}^2} \right] \quad s_{\hat{a}} = \sqrt{\frac{1}{\Delta} \sum_i \frac{x_i^2}{s_{y_i}^2}} \quad (8.7)$$

$$\hat{b} = \frac{1}{\Delta} \left[\sum_i \frac{1}{s_{y_i}^2} \sum_i \frac{x_i y_i}{s_{y_i}^2} - \sum_i \frac{x_i}{s_{y_i}^2} \sum_i \frac{y_i}{s_{y_i}^2} \right] \quad s_{\hat{b}} = \sqrt{\frac{1}{\Delta} \sum_i \frac{1}{s_{y_i}^2}} \quad (8.8)$$

Incetnze su ascisse e ordinate Nel caso anche s_{x_i} non siano o non si vogliano trascurare, esse influenzeranno il calcolo della distanza tra i punti misurati e la eventuale retta interpolante. In questo caso l'errore efficace e':

$$s_i^2 \simeq s_{y_i}^2 + \frac{\partial g}{\partial x} s_{x_i}^2 = s_{y_i}^2 + b^2 s_{x_i}^2$$

Si pone ora il problema di calcolare b allora si procede cosi: una prima volta di calcola b assumendo $s_{x_i} = 0 \forall i$. Dopodiche' si applica al formula sopra e si calcola:

$$Q(a, b; x_i, y_i) = \sum_i \left(\frac{y_i - (a + b x_i)}{\sqrt{s_{y_i} + b^2 s_{x_i}^2}} \right)^2 \quad (8.9)$$

Questo porta a conti simili al caso precedente ove si sostituisca $s_{y_i} \rightarrow \sqrt{s_{y_i} + b^2 s_{x_i}^2}$. Questo passaggio risulta tuttavia numericamente poco sensato. E' facile vedere che questa sostituzione ha solo un minimo effetto sulla stima di \hat{a}, \hat{b} .

Interpolazione lineare: caso III I parametri che meglio approssimano una serie di punti in cui le variabili abbiano generiche incertezze gaussiane, la cui distribuzione sia assunta lineare con $y = a + bx$, ottenuti con il metodo dei minimi quadrati:

$$\Delta = \sum_i \frac{1}{s_i} \sum_i \frac{x_i^2}{s_i^2} - \left(\sum_i \frac{x_i}{s_i} \right)^2; \quad s_i = \sqrt{s_{y_i} + b^2 s_{x_i}^2}$$

$$\hat{a} = \frac{1}{\Delta} \left[\sum_i \frac{x_i^2}{s_i^2} \sum_i \frac{y_i}{s_i^2} - \sum_i \frac{x_i}{s_i^2} \sum_i \frac{x_i y_i}{s_i^2} \right] \quad s_{\hat{a}} = \sqrt{\frac{1}{\Delta} \sum_i \frac{x_i^2}{s_i^2}} \quad (8.10)$$

$$\hat{b} = \frac{1}{\Delta} \left[\sum_i \frac{1}{s_i^2} \sum_i \frac{x_i y_i}{s_i^2} - \sum_i \frac{x_i}{s_i^2} \sum_i \frac{y_i}{s_i^2} \right] \quad s_{\hat{b}} = \sqrt{\frac{1}{\Delta} \sum_i \frac{1}{s_i^2}} \quad (8.11)$$

8.2.1 Un esempio numerico

Si sono prese 7 coppie di misure come in tabella. Assumiamo diverse condizioni per le incertezze: caso I, tutte le $s_{y_i} = s_y = 0.3$ e casi II e III s_{y_i} tutte diverse.

x_i	y_i	Casi s_{x_i}	(I)	(II)	(III)
			s_y	s_{y_i}	s_i
1.1	2.4	0.1	0.3	0.15	0.21
2.1	3.28	0.1	0.3	0.18	0.23
3.2	4.5	0.1	0.3	0.21	0.25
3.9	5.96	0.1	0.3	0.24	0.28
5.1	7.64	0.1	0.3	0.27	0.31
6.1	8.42	0.1	0.3	0.30	0.33
6.9	10.94	0.1	0.3	0.33	0.36

Gli errori relativi delle due variabili non sono neppure troppo diversi. Quello sulle X varia tra 1.5 – 4% mentre quello sulle Y tra 2.7 – 12.5%. Grafichiamo ora la serie di dati in [Figura 8.2](#) mostrando i casi con le diverse incertezze.

Nonostante questo, trascurando gli errori sulle x_i [Metodo 1] e applicando le [Equazione 8.5](#) otteniamo:

$$a_1 = 0.39 \pm 0.26 \quad b_1 = 1.423 \pm 0.058$$

Considerando le incertezze diverse secondo la tabella

[Metodo 2] otteniamo invece

$$a_1 = 0.63 \pm 0.16 \quad b_1 = 1.357 \pm 0.044$$

E' naturale che questi valori siano diversi dal Metodo 1 in quanto le prime misure hanno incertezza misura e quindi maggior peso. Considerando invece tutte le incertezze [Metodo 3], calcolando gli errori efficaci s_i riportati nell'ultima colonna, e applicando le [Equazione 8.7](#), otteniamo

$$a_3 = 0.56 \pm 0.20 \quad b_3 = 1.377 \pm 0.053$$

Le due stime ottenute con il Metodo 2 e 3 sono diverse ma sono perfettamente compatibili entro le incertezze.

Retta passante per l'origine Se la funzione interpolante e' una retta passante per l'origine del tipo $y = mx$ si ha un unico parametro libero. Procedendo come sopra si ottengono (unicamente caso I, errori ascissa trascurabili):

$$\hat{m} = \frac{\sum_i x_i y_i}{\sum_i x_i^2} \quad \hat{s}_m = \frac{s_y}{\sqrt{\sum_i x_i^2}} \quad (8.12)$$

Si consiglia tuttavia l'uso di questa formula in quanto di norma e' meglio sempre *verificare* che la misure abbiano intercetta passante per l'origine. Per farlo e' sufficiente usare le [Equazione 8.7](#) o [Equazione 8.10](#) e verificare la compatibilita' della intercetta a con l'origine.

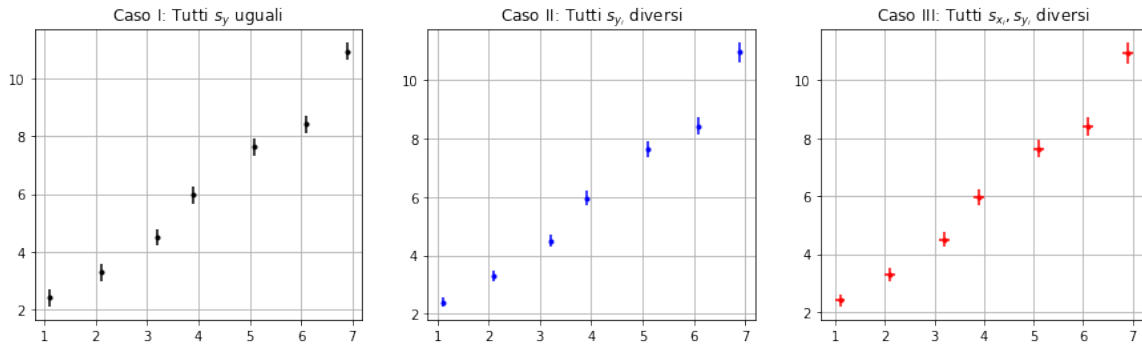


Figura 8.2: I dati della tabella mostrati nei tre casi diversi per le interpolazioni.

8.3 Interpolazione caso generale

Torniamo ora a discutere il caso di una relazione generica tra due variabili del tipo $y = g(x)$. Come ci si puo' comportare in questo caso? E' spesso possibile fare un *cambio scala* che linearizza le scale appunto. Ad esempio:

$$y = \ln(bx) \quad y' = e^y = bx$$

Allora e' possibile fare una regressione lineare sulle misure (x_i, y'_i) , trovare i valori i parametri della retta interpolante e poi rifare il cambio di variabile $y' \rightarrow y$.

Questo metodo e' spesso usato tuttavia ha un inconveniente. Come visto in [Capitolo 6](#), se s_{y_i} sono deviazione standard gaussiane, non e' detto che $s_{y'_i}$ lo siano. Ad esempio, nel caso sopra di $y = \ln(bx)$ le incertezze sono asimmetriche sopra e sotto i punti y' , e quindi le formule di minimizzazione sarebbero diverse. E' anche vero, che per incertezze non troppo grandi numericamente il risultato non cambia troppo velocemente, ma dipende la grado di precisione richiesto.

Interpolazione per distribuzioni di misure . L'interpolazione si fa anche per distribuzioni di misure (istogrammi) e non solo per distribuzioni di coppie di misure. Ad esempio, e' sempre possibile pensare ad un istogramma come ad un grafico in due dimensioni con $(x_i, y_i) = (x_i, f_i)$ ovvero con le frequenze (relative, assolute) come variabile. E' possibile quindi ora fare una regressione assumendo una certa funzione per le f_i ad esempio Gaussiana. Questo vale per ogni distribuzione di probabilita'. Al calcolatore (vedi paragrafo successivo) e' talvolta piu' veloce e altrettanto accurato calcolare i parametri della distribuzione di riferimento in questa maniera.

Esempio:

Si prenda la seguente serie di misure (conteggio assoluto tra parentesi)

$$15.10(2), 15.11(9), 15.12(13), 15.13(7), 15.14(3), 15.16(1)$$

Si calcolino media aritmetica e deviazione standard con le formule analitiche note. Si scriva una routine al calcolatore che esegue una regressione lineare gaussiana sui dati e si confrontino le stime.

Interpolazione al calcolatore In generale, la regressione (lineare o non) viene fatta oggi giorno al calcolatore. I moderni programmi hanno al loro interno delle routine di *minimizzazione*. Queste si preoccupano di tornare i migliori parametri secondo una funzione definita dall'utente. Oltre alla funzione, l'utente puo' definire tutta una serie di informazioni tra cui la precisione richiesta sui parametri, specificare dei valori ragionevoli per i parametri in maniera da ottimizzare l'esecuzione, etc. In alcuni casi, la minimizzazione addirittura puo' fallire (per problemi numerici) oppure richiedere troppo tempo, in quanto di norma viene fatta a passi successivi in cui si fa una prima stima di parametri o poi questa viene raffinata con maggior precisione. La minimizzazione al calcolatore e' molto complessa

e ci sono diverse routine di minimizzazione. Una molto nota nel settore della fisica delle particelle e' MINUIT¹.

Il calcolo manuale dei parametri di regressione e' oramai in disuso, perfino nel caso lineare esatto visto sopra. Si presti pero' attenzione al fatto che e' importante conoscere in linea generale le impostazioni di base delle routine utiizzate ai calcolatori per non avere incertezze sul risultato ottenuto.

8.4 Errore a posteriori

Abbiamo visto come nella stima dei parametri abbiano un ruolo le incertezze delle misure. Ha senso chiedersi *a posteriori* se quelle incertezze misurate fossero ragionevoli *rispetto alla ipotesi di distribuzione dei dati*. A questo puo' aiutare la stima dell'errore a posteriori.

L'errore a posteriori e' la media quadratica degli scarti tra i punti misurati e i punti della funzione di regressione:

$$\delta = \sqrt{\frac{1}{N - k} \sum_i (y_i - g(x; \vec{\lambda}))^2} \tag{8.13}$$

dove k e' il numero di parametri della funzione di regressione $g(x; \vec{\lambda})$

Ad esempio nel caso di interpolazione lineare $y = a + bx$

$$\delta = \sqrt{\frac{1}{N - 2} \sum_i (y_i - (a + bx))^2} \tag{8.14}$$

Nel caso di interpolazione passante per l'origine

$$\delta = \sqrt{\frac{1}{N - 1} \sum_i (y_i - mx)^2}$$

L'errore a posteriori cresce quanto piu' le misure sono disperse lontano dalla funzione interpolante. Piu' nello specifico, il confronto va fatto tra δ e s_{y_i} :

1. se $\delta \simeq s_{y_i}$ allora si presuppone che la regressione ipotizzata sia valida
2. se $\delta \ll s_{y_i}$ [incertezze sperimentali grandi] ci sono motivi per dubitare della relazione lineare che si presume legare x_i, y_i .
3. se $\delta \gg s_{y_i}$ [incertezze sperimentali piccole] ci possono essere diverse spiegazioni, tra cui che abbiamo sottostimato le incertezze casuali o sistematiche.

Vedremo comunque in [Capitolo 10](#) ulteriori modi per quantificare la qualità della regressione lineare.

8.5 Correlazione tra i parametri di interpolazione

Si noti come in tutte le formule di stima dei parametri, essi siano ottenuti come combinazione di misure. Combinazione di misure sono sempre correlate tra loro e per questo in linea generale *i parametri di regressione sono correlati*. Nel caso della regressione lineare lo abbiamo visto esplicitamente calcolando \hat{a} dove compariva il termine b . In linea generale, come per ogni variabile aleatorie si dovrebbe quindi calcolare il coefficiente di correlazione campionario r_{λ_i, λ_j} per ogni coppia di parametri.

Per gli scopi di questo corso, ci interessa la correlazione lineare tra pendenza e intercetta. Essa vale:

$$s_{ab} = \frac{\bar{x} s_y^2}{N s_x^4} \tag{8.15}$$

¹Si veda ad esempio la complessità di MINUIT dal suo manuale <https://root.cern.ch/download/minuit.pdf>

questa relazione sara' usata nella esperienza del viscosimetro.

Le dimostrazioni di queste affermazioni si trovano nel Loreti.

Capitolo 9

Stima dei Parametri con la Massima Verosimiglianza

Quest'opera è distribuita con licenza [Creative Commons](#) "Attribuzione – Non commerciale – Condividi allo stesso modo 3.0 Italia".



In questo capitolo generalizziamo il metodo dei minimi quadrati definendo una funzione dei parametri, la verosimiglianza, di cui cercheremo i massimi. I parametri calcolati in questo modo seguono il metodo della *massima verosimiglianza* o *Maximum Likelihood Estimation (MLE)*.

9.1 La funzione di verosimiglianza

Nel [Capitolo 7](#) abbiamo visto come le singole misure o le misure ripetute siano associate ad una certa distribuzione di probabilità: esse sono campioni unitari o di molteplicità di quella distribuzione di probabilità. Abbiamo anche visto il concetto di IID nel caso di misure ripetute indipendenti della stessa distribuzione. In generale quindi, date N misure indipendenti x_i (che indichiamo con \vec{x}), ciascuna con la propria distribuzione di probabilità $f_{X_i}(x_i; \vec{\lambda}_i)$ funzione dei parametri $\vec{\lambda}_i$, allora, come visto anche in [Capitolo 4](#), e' possibile costruire la (densità di) probabilità congiunta:

$$\prod_{i=1}^N f_{X_i}(x_i; \vec{\lambda}) \quad (9.1)$$

dove in linea di principio le f_{X_i} possono essere tutte diverse fintantoché le X_i sono indipendenti. Ora, siccome le \vec{x} sono misure, se $\vec{\lambda}$ sono definiti (ovvero se conosciamo la distribuzione di riferimento), allora [Equazione 9.1](#) e' una grandezza che viene detta *verosimiglianza dei dati x_i rispetto alle ipotesi $f_{X_i}(\vec{\lambda})$* .

Esempio:

Likelihood per IID Poisson Osserviamo 5 eventi in una unità di tempo e vogliamo calcolare la verosimiglianza relative a tre ipotesi (di distribuzione di probabilità di riferimento): 1) distribuzione di Poisson con valore di aspettazione 2, 2) con valore di aspettazione 9, 3) con valore di aspettazione 20. **R:** Abbiamo un unico evento con $n = 5$ conteggi, le verosimiglianze sono quindi:

$$L_1 = \mathcal{P}(5; 2) = \frac{2^5 e^{-2}}{5!} \simeq 0.036; \quad L_2 = \mathcal{P}(5; 9) \simeq 0.062; \quad L_3 = \mathcal{P}(5; 20) \simeq 0.00005$$

Se osserviamo in una seconda misura nello stesso intervallo di tempo 8 decadimenti, otteniamo la verosimiglianza:

$$\begin{aligned} L_1 &= \mathcal{P}(5; 2)\mathcal{P}(8; 2) = \mathcal{P}(13; 4) \sim 6.4 \cdot 10^{-3} \\ L_2 &= \mathcal{P}(5; 9)\mathcal{P}(8; 9) = \mathcal{P}(13; 18) \sim 5.1 \cdot 10^{-2} \\ L_3 &= \mathcal{P}(5; 20)\mathcal{P}(8; 20) = \mathcal{P}(13; 40) \simeq 6.1 \cdot 10^{-7} \end{aligned}$$

Esempio:

Likelihood per IID Gauss Prendiamo una misura ottenendo $x = 0$. Vogliamo calcolare la verosimiglianza relative a due ipotesi (di distribuzione di probabilità di riferimento Gaussiana): 1) con media 1 e ampiezza 1, 2) con media 2 e ampiezza 2 **R:** Le verosimiglianze secondo le due ipotesi sono quindi:

$$\begin{aligned} L_1 &= \mathcal{G}(1, 1; 0) = \frac{1}{\sqrt{2\pi}} e^{-0.5\left(\frac{0-1}{1}\right)^2} \simeq 0.24 \\ L_2 &= \mathcal{G}(2, 2; 0) = \frac{1}{\sqrt{2\pi}2} e^{-0.5\left(\frac{0-2}{2}\right)^2} \simeq 0.12 \end{aligned}$$

Si noti come nonostante in entrambi i casi la distanza dalla media sia pari a 1σ , la verosimiglianza sia diversa.

Si noti anzitutto che non esiste una verosimiglianza dei dati senza specificare una distribuzione di probabilità di riferimento. Tuttavia sappiamo che c'è un problema nel definire la PF di riferimento a partire dai campioni di misure: questa stima è incerta a causa della fluttuazioni casuali delle misure dovute al campionamento finito. In generale quindi, in riferimento a [Equazione 9.1](#) mentre le x_i sono misure, i parametri $\vec{\lambda}$ delle PF di riferimento sono da stimare! In sostanza allora [Equazione 9.1](#), se i parametri sono incognite, diventa appunto una *funzione dei parametri*:

Si dice **funzione di verosimiglianza** (*likelihood function, likelihood*) dei dati x_i rispetto alle distribuzioni di probabilità parametriche $f_{X_i}(x_i; \vec{\lambda})$, la funzione dei parametri $\vec{\lambda}$:

$$\mathcal{L}(\vec{\lambda}; \vec{x}) = \prod_{i=1}^N f_{X_i}(x_i; \vec{\lambda}) \quad (9.2)$$

Ad esempio, se prendiamo N misure IID distribuite secondo la medesima Gaussiana, allora la funzione verosimiglianza si scrive:

$$\begin{aligned} \mathcal{L}(\mu, \sigma; \vec{x}) &= \prod_{i=1}^N \mathcal{G}(x_i; \mu, \sigma) \\ &= \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_1-\mu}{\sigma}\right)^2} \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_2-\mu}{\sigma}\right)^2} \cdot \dots \end{aligned} \quad (9.3)$$

Osserviamo in [Figura 9.1](#) (sinistra) come varia la forma della \mathcal{L} all'aumentare del numero di misure. Nella figura, σ si intende fissato, e osserviamo la funzione di verosimiglianza per il solo parametro μ in ascissa. Come ci si aspetta, il punto di massimo si sposta all'aumentare della numerosità del campione. Si noti anche come la \mathcal{L} non rappresenti una densità di probabilità in senso stretto in quanto l'integrale sotto le curve varia al variare della numerosità dei dati. Le altre due curve in [Figura 9.1](#) saranno discusse a breve.

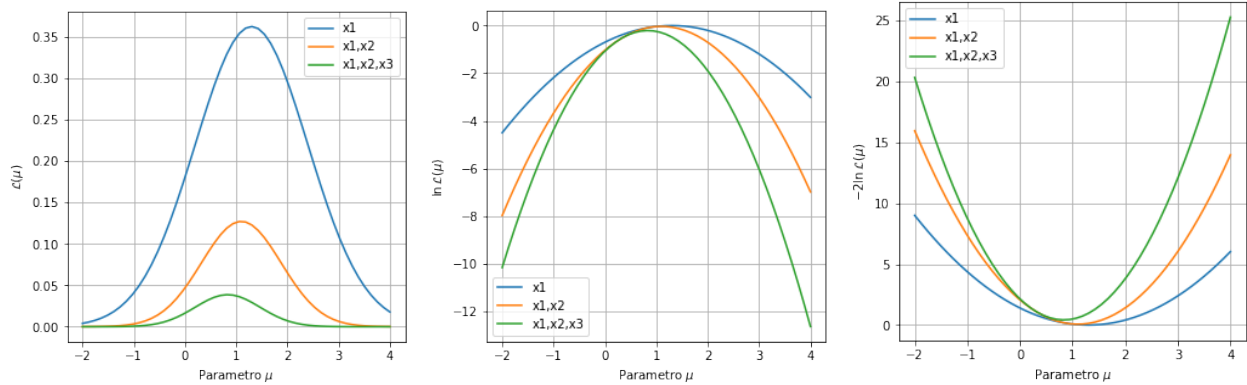


Figura 9.1: Sia $x_1 = 1.3, x_2 = 0.9, x_3 = 0.3$ e $\sigma = 1.3$. (sinistra) Profilo della funzione di verosimiglianza \mathcal{L} per il parametro μ per 1, 2, 3 misure gaussiane; (centro) Profilo della funzione di ln-verosimiglianza $\ln \mathcal{L}$ per il parametro μ per 1, 2, 3 misure gaussiane; (destra) Profilo della funzione di $-2\ln$ -verosimiglianza $-2\ln \mathcal{L}$ per il parametro μ per 1, 2, 3 misure gaussiane.

Prendiamo [Figura 9.1](#) (sinistra) per la curva corrispondente a 3 dati. La curva ci dice che esiste un valore del parametro μ (si ricordi, e' il parametro di interesse da determinare) che rende massima questa funzione. Nella figura, e' un valore leggermente minore di 1 (in realtà 0.83). Gli altri valori di μ danno una verosimiglianza dei dati inferiore a quella calcolate per $\mu = 0.83$.

Si ragioni su quale possa essere il valore μ che massimizza \mathcal{L}

9.2 Il metodo della Massima Verosimiglianza per la stima dei parametri

Ribadiamo come per ogni scelta di $\vec{\lambda}$, [Equazione 9.2](#) ritorna un numero. In questo senso, e' sempre possibile stimare i valori di $\vec{\lambda}$ per cui la funzione $\mathcal{L}(\vec{\lambda}; \vec{x})$ ha un massimo, ovvero quando le derivate prima sono nulle e le derivate seconde negative.

$$\begin{cases} \frac{\partial \mathcal{L}(\vec{\lambda}; \vec{x})}{\partial \lambda_1}(\hat{\lambda}_1) = 0 & \frac{\partial^2 \mathcal{L}(\vec{\lambda}; \vec{x})}{\partial \lambda_1^2}(\hat{\lambda}_1) < 0 \\ \frac{\partial \mathcal{L}(\vec{\lambda}; \vec{x})}{\partial \lambda_2}(\hat{\lambda}_2) = 0 & \frac{\partial^2 \mathcal{L}(\vec{\lambda}; \vec{x})}{\partial \lambda_2^2}(\hat{\lambda}_2) < 0 \\ \dots \end{cases}$$

Questa procedura e' detta del calcolo della **massima verosimiglianza**. I parametri $\hat{\lambda}$ saranno quelli che massimizzano la densità di probabilità congiunta delle N misure. La massima verosimiglianza, che e' un numero puro, a sua volta si indica spesso con il simbolo $\hat{\mathcal{L}}$ ovvero

$$\hat{\mathcal{L}} = \prod_{i=1}^N f_{X_i}(x_i; \hat{\lambda}) \quad (9.4)$$

Come si puo' immaginare, il calcolo di $\hat{\mathcal{L}}$ puo' risultare molto pesante. Ci aiuta il fatto che per applicare MLE non siamo interessati a tutta la funzione, ma essenzialmente alla posizione dei massimi. Visto il nostro interesse, se invece che la funzione di verosimiglianza ne prendiamo il logaritmo $\ln \mathcal{L}$,

essendo il logaritmo funzione crescente nel dominio di \mathcal{L} , abbiamo che i punti di massimo di \mathcal{L} e $\ln \mathcal{L}$ coincidono. Questo varrebbe con qualsiasi funzione crescente, ma il logaritmo ha l'ulteriore vantaggio che trasforma la produttoria di [Equazione 9.2](#) in una ben più maneggevole sommatoria, sfruttando la proprietà algebrica $\ln(ab) = \ln(a) + \ln(b)$. Allora:

Si dice **funzione di log-verosimiglianza** (*log-likelihood*) dei dati x_i rispetto alle distribuzioni di probabilità parametriche $f_{X_i}(x_i; \vec{\lambda})$, la funzione dei parametri $\vec{\lambda}$:

$$\ln \mathcal{L}(\vec{\lambda}; \vec{x}) = \sum_{i=1}^N \ln f_{X_i}(x_i; \vec{\lambda}) \quad (9.5)$$

Log-Verosimiglianza della PDF Gaussiana Ad esempio la log-verosimiglianza della gaussiana risulta, partendo da [Equazione 9.3](#):

$$\ln \mathcal{L}(\mu, \sigma; \vec{x}) = - \underbrace{\frac{1}{2} N \ln(2\pi)}_{\text{non dipende da } \mu, \sigma} - \underbrace{\frac{1}{2} N \ln(\sigma^2)}_{\text{dipende solo da } \sigma} - \underbrace{\frac{1}{2} \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^2}_{\text{dipende da } \mu, \sigma}$$

Si noti che per il calcolo dei massimi, tutti i termini che non dipendono dal parametro in questione non hanno utilità e possono essere non scritti. Ad esempio, la formula sopra si puo' scrivere direttamente

$$\ln \mathcal{L}(\mu, \sigma; \vec{x}) = \frac{1}{2} N \ln(\sigma^2) - \frac{1}{2} \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^2 \quad (9.6)$$

La [Figura 9.1](#) (centro) mostra (una volta fissato il parametro σ), la funzione $\ln \mathcal{L}$ in funzione del parametro μ . Si noti come all'aumentare delle misure prese in considerazioni la $\ln \mathcal{L}$ si stringa attorno al valore massimo. Torneremo piu' tardi su questa proprietà. Procediamo quindi alla ricerca dei massimi di [Equazione 9.6](#). Cominciamo dal primo parametro μ :

$$\frac{\partial \ln \mathcal{L}(\mu, \sigma; \vec{x})}{\partial \mu} = \frac{\partial}{\partial \mu} \frac{1}{2} \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^2 = \frac{1}{2\sigma^2} \sum_{i=1}^N 2(x_i - \mu) = 0$$

dove abbiamo omesso l'addendo con sola dipendenza a σ . Quindi $N\mu - \sum_{i=1}^N x_i = 0$ ovvero

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i = \bar{x} \quad (9.7)$$

ovvero la miglior stima di μ e' la media aritmetica \bar{x} dei valori! La cosa non dovrebbe sorprenderci visto che l'abbiamo assunta per vera gia' per molti capitoli, ma ora ne troviamo dimostrazione. Avendola disegnata, non ci serve strettamente farlo, ma possiamo verificare che \bar{x} massimizza la verosimiglianza infatti la

$$\frac{\partial^2 \ln \mathcal{L}(\mu, \sigma; \vec{x})}{\partial \mu^2} = -\frac{N}{2\sigma^2} < 0$$

In maniera analoga, procediamo ora calcolando σ con il metodo MLE. Matematicamente e' più semplice derivare per σ^2 e cosi' faremo:

$$\begin{aligned} \frac{\partial \ln \mathcal{L}(\mu, \sigma; \vec{x})}{\partial \sigma^2} &= \frac{\partial}{\partial \sigma^2} \frac{1}{2} \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^2 - \frac{1}{2} N \ln(\sigma^2) = \frac{1}{2(\sigma^2)^2} \sum_{i=1}^N (x_i - \mu)^2 - \frac{N}{2\sigma^2} = \\ &= \frac{1}{2\sigma^2} \left[\frac{1}{\sigma^2} \sum_{i=1}^N (x_i - \mu)^2 - N \right] = 0 \quad \text{ovvero} \quad \sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \end{aligned}$$

Di nuovo osserviamo che il valore del parametro che massimizza la verosimiglianza e' lo scarto quadratico medio *rispetto a* μ . Come mai non appare il fattore $N - 1$ al denominatore? Ma perche', come abbiamo visto, questo valore non lo possiamo calcolare dai dati, essendoci μ , e quando al posto di μ usiamo la sua stima \bar{x} allora dobbiamo tenere in conto del bias introdotto e usare la stima:

$$s_x^2 = \frac{1}{N - 1} \sum_{i=1}^N (x_i - \bar{x})^2$$

Infine, e' facile vedere che anche σ^2 è punto di massimo visto che la derivata seconda e' negativa in quel punto.

Minimizzazione della $-2 \ln \mathcal{L}$

Si noti che *per il caso della Gaussiana*, [Equazione 9.6](#) contiene due addendi con i termini $-1/2$ in fronte. Allora, e' prassi consolidata usare invece, *unicamente per il caso della Gaussiana* la funzione $-2 \ln \mathcal{L}$ invece che $\ln \mathcal{L}$. In questo caso quindi:

$$-2 \ln \mathcal{L}(\mu, \sigma; \vec{x}) = N \ln(\sigma^2) + \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^2 \tag{9.8}$$

e si parla allora di *minimizzazione della $-2 \ln \mathcal{L}$* invece che massimizzazione. Si presti attenzione a non fare confusioni di termini. La forma della $-2 \ln \mathcal{L}$ e' riportata in [Figura 9.1](#) (destra).

Si presti ora attenzione a [Equazione 9.8](#). Se consideriamo questa come funzione solamente di μ e non di σ , allora il primo addendo non conta e la funzione diventa equivalente a:

$$-2 \ln \mathcal{L} = \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^2 = \mathcal{Q}(\mu)$$

Questa funzione e' stata gia' incontrata nel [Capitolo 8](#) e chiamata con il nome di funzione dei minimi quadrati \mathcal{Q} (si veda ad esempio [Equazione 8.4](#)). Come mai? Nel caso della interpolazione, il parametro σ era noto, in quanto era un risultato sperimentale (le incertezze sulle ordinate), anzi, in realtà non era necessario che fossero unico per tutte le misure, ma poteva variare misura per misure, ovvero si poteva avere σ_i . Nel caso della interpolazione poi il valore μ di riferimento era proprio il valore y_i^* della retta interpolante, ovvero μ a sua volta era funzione di due parametri: la pendenza e la intercetta della retta. Si mediti su questo punto. In riassunto, il metodo dei minimi quadrati corrisponde al metodo della massima verosimiglianza per variabili IID Gaussiane, ognuna con un valore di riferimento diverso (i punti della retta interpolante).

Esempio:

Scrivere la funzione di massima verosimiglianza per misure (x_i, y_i) in cui y_i siano Gaussiane e x_i note senza incertezza.

Log-Verosimiglianza della PDF Esponenziale Riproponiamo quanto appreso ora per un campione di IID di tipo PDF esponenziale, allora la funzione di verosimiglianza

$$\mathcal{L}(\lambda; \vec{x}) = \prod_{i=1}^N \lambda e^{-\lambda x_i} = \lambda e^{-\lambda x_1} \cdot \lambda e^{-\lambda x_2} \dots$$

diventa

$$\ln \mathcal{L}(\lambda; \vec{x}) = \ln(\lambda^N) - \sum_{i=1}^N \lambda x_i$$

La rappresentazione grafica di \mathcal{L} e $\ln \mathcal{L}$ si trova in [Figura 9.2](#). Derivando [Equazione 9.9](#) rispetto a λ si ottiene la stima di questo parametro:

$$\frac{d \ln \mathcal{L}(\lambda; \vec{x})}{d\lambda} = \frac{N}{\lambda} - \sum_{i=1}^N x_i = 0 \quad \text{ovvero} \quad \lambda = \frac{N}{\sum_{i=1}^N x_i} \quad (9.9)$$

e quindi in questo caso, sapendo che l'esponenziale ha come valore di aspettazione $\mu = 1/\lambda$ e come varianza $\sigma^2 = 1/\lambda^2$ allora per il caso di campioni, la miglior stima del valore medio e della varianza sono:

$$\hat{\mu} = \frac{1}{\lambda} = \frac{1}{\bar{x}} \quad \text{e} \quad \hat{\sigma} = \frac{1}{\lambda^2} = \frac{1}{\bar{x}^2}$$

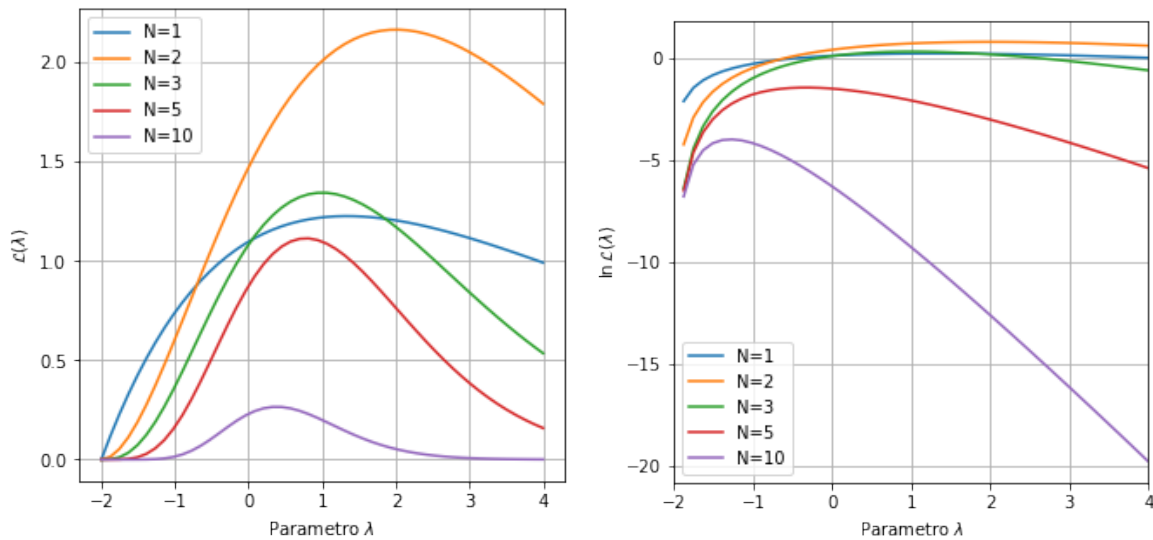


Figura 9.2: (sinistra) Profilo della funzione di verosimiglianza \mathcal{L} per il parametro λ per 1, 2, 3, 5, 10 misure a distribuzione esponenziale; (destra) Profilo della funzione di ln-verosimiglianza $\ln \mathcal{L}$ per le stesse misure. All'aumentare delle misure la verosimiglianza diventa una Gaussiana e la stima $\hat{\lambda}$ diventa piu' precisa.

Esempio:

Sono stati osservati decadimenti da un materiale radioattivo dopo

$$0.0198, 0.0199, 0.197 \text{ s}$$

Si dia la miglior stima del tempo medio di decadimento.

Allora per l'esponenziale siccome per esponenziale $\mu = \sigma = \frac{1}{\lambda}$ la stima MLE del parametro e':

$$\hat{\lambda} = \frac{1}{\bar{n}} = \frac{1}{0.0198} = 50.5 \text{ s}$$

e quindi la distribuzione di riferimento è $f(x) = 50.5e^{-50.5x}$. Il tempo medio:

$$\tau = \frac{1}{\lambda} = \bar{n} = 0.0198 \text{ s}$$

Log-Verosimiglianza della PMF di Poisson Analogamente per la PMF di Poisson possiamo scrivere

$$\mathcal{L}_{\mathcal{P}}(\lambda; \vec{n}) = \prod_{i=1}^N \frac{\lambda^{n_i} e^{-\lambda}}{N_i!} \quad \rightarrow \quad \ln \mathcal{L}_{\mathcal{P}}(\lambda; \vec{n}) = -N \lambda + \ln(\lambda) \sum_{i=1}^N n_i - \sum_{i=1}^N \ln(n_i!)$$

la cui derivata prima rispetto all'unico parametro λ è

$$\frac{d\mathcal{L}_{\mathcal{P}}(\lambda; \vec{n})}{d\lambda} = -N + \frac{1}{\lambda} \sum_{i=1}^N n_i$$

e quindi sapendo che la PMF di Poisson ha come valore di aspettazione $\mu = \lambda$ e come varianza $\sigma^2 = \lambda$ allora per il caso di campioni, la miglior stima del valore medio e della varianza sono:

$$\widehat{\mu}_{\mathcal{P}} = \widehat{\sigma}_{\mathcal{P}}^2 = \frac{1}{N} \sum_{i=1}^N n_i = \bar{n}$$

ovvero la media dei conteggi, come avevamo già visto.

Esempio:

Sono state effettuate le seguenti misure di tassi di decadimento (numero di decadimenti nell'unità di tempo dec/s) da un materiale radioattivo:

$$44, 52, 57, 49$$

Si dia la miglior stima del valore medio con la incertezza.

Allora:

$$\mu = \lambda = \bar{n} = 50.5 \quad \sigma = \sqrt{\lambda} = 7.10$$

e quindi $\sigma_{\bar{x}} = \sigma / \sqrt{N} = 3.6$ dec/s ovvero

$$\lambda = (50.5 \pm 3.6) \text{ dec/s}$$

Log-Verosimiglianza della PMF Binomiale La verosimiglianza per una binomiale corrisponde formalmente alla binomiale stessa (attenzione, tranne che quest'ultima è funzione del parametro p), infatti si tratta già del prodotto di N eventi di Bernoulli indipendenti:

$$\mathcal{L}_{\mathcal{B}}(p; k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \rightarrow \quad \ln \mathcal{L}_{\mathcal{B}}(p; k) = \square + k \ln(p) + (n-k) \ln(1-p)$$

dove \square è un termine che non contiene p . La derivata prima rispetto all'unico parametro p è

$$\frac{k}{p} - \frac{n-k}{1-p} = 0 \quad k - kp - np + kp = 0 \quad k = np$$

Abbiamo quindi trovato che sapendo che la PMF Binomiale ha come valore di aspettazione $\mu = np$ e come varianza $\sigma^2 = np(1-p)$ allora per il caso di campioni, la miglior stima del valore medio e della varianza sono:

$$\widehat{\mu} = k \quad \text{e} \quad \widehat{\sigma}^2 = k \left(1 - \frac{k}{n}\right)$$

Esempio:

Ho misurato l'efficienza del fotocatodo di un fotomoltiplicatore mandando 100 fotoni singoli a 440 nm. Ho letto un segnale in uscita 40 volte. Stimare incertezza.

R: il fenomeno è binomiale: ogni volta il fotone può essere convertito in fotoelettrone e dare segnale. La probabilità che ciò avvenga 40 volte su 100 senza considerare l'ordine è binomiale. Allora la stima della efficienza e della incertezza statistica sulla efficienza è:

$$\epsilon = k \pm \sqrt{k \left(1 - \frac{k}{n}\right)} = 40 \pm 5$$

9.2.1 Combinazione di verosimiglianze

Come vedrete, la funzione di verosimiglianza e' alla base del confronto tra dati e ipotesi. In questo senso, la analisi dati passera' spesso per la costruzioni di una funzione di verosimiglianza e per la sua massimizzazione. Finora abbiamo parlato di verosimiglianza per IID, tuttavia, per come e' costruita la funzione, ovvero come produttorica di verosimiglianze, questo ci permette di combinare a piacere funzioni di verosimiglianza per variabili aleatorie con diversi parametri e diverse PF di riferimento. In realtà, si costruisce una verosimiglianza combinata per tutte le variabili corrispondenti a misure di input.

Ad esempio, se prendiamo una misura di lunghezza S e poi in corrispondenza 5 misure di tempo T supposte gaussiane, allora possiamo costruire la funzione verosimiglianza combinata:

$$\mathcal{L}(\mu_1, \mu_2, \sigma_2; s_1, \vec{t}) = \mathcal{U}(s_1; \mu_1) \prod_{i=1}^5 \mathcal{G}(t_i; \mu_2, \sigma_2) \quad (9.10)$$

e in generale quindi per N misure corrispondenti a variabili aleatorie anche diverse, si puo' costruire la likelihood combinata

$$\mathcal{L}(\vec{\lambda}_1, \vec{\lambda}_2, \dots; \vec{x}, \vec{y}, \dots) = \prod_{i=1}^{N_1} f_X(x_i; \vec{\lambda}_1) \cdot \prod_{j=1}^{N_2} f_Y(y_j; \vec{\lambda}_2) \dots \quad (9.11)$$

e da questa poi la $\ln \mathcal{L}$ se utile.

9.3 Performance di uno stimatore

La stima dei parametri di una PF a partire dai campioni di dati si puo' fare con diverse tecniche, di cui una e' la massima verosimiglianza. In generale quindi si parla di **stimatore** come della funzione $\hat{\Lambda} : \mathbb{R}^N \rightarrow \mathbb{R}$ tale che ad un insieme di misure \vec{x} sia associata una stima $\hat{\lambda}$.

La massima verosimiglianza e' uno dei possibili stimatori di parametri, ma ve ne sono altri, ad esempio il metodo cosiddetto dei momenti. Per questo si definiscono alcune proprietà degli stimatori, che riguardano le caratteristiche delle stime. Non sempre le stime li possono rispettare tutti allo stesso modo, e si dovrà valutare caso per caso. Il metodo della massima verosimiglianza è diffuso perchè è quello piu' efficiente, ovvero a varianza minima: tra tutte le possibili funzioni dei dati che si potrebbero definire per stimare il valore vero λ^* dal campione, quella mediamente più vicina ad esso. Il metodo della massima verosimiglianza, ha soluzioni sono quelle affette dal minimo errore casuale (almeno asintoticamente).

Correttezza e distorsione Un estimatore si dice corretto o non distorto se il valore di aspettazione della stima corrisponde al valore vero della popolazione:

$$E[\hat{\lambda}] = \lambda^*$$

Se questo non avviene lo stimatore (o la stima) sono detti distorti.

Ad esempio, la stima di σ per un campione di IID Gaussiane e' distorto (di un fattore $(N + 1)/N$ come visto). La distorsione se nota puo' essere corretta.

Consistenza Al limite per $N \rightarrow \infty$ la stima tende al valore vero:

$$\lim_{N \rightarrow \infty} \hat{\lambda} = \lambda^*$$

Efficienza Dati due stime $\hat{\lambda}_1, \hat{\lambda}_2$ non distorte, quella con varianza minore e' detta piu' efficiente

$$\text{VAR}[\lambda_1] < \text{VAR}[\lambda_2] \implies \hat{\lambda}_1 \text{ e' piu' efficiente}$$

In particolare esiste un valore minimo di varianza dello stimatore detto di Cramér–Rao¹. Un parametro stimato con minima varianza e' detto *minimum variance unbiased (MVU) estimator*.

¹In appendice G del Loreti trovate una lunga discussione su questa limite con alcune dimostrazioni, che a mio avviso esulano dagli interessi di questo corso.

Robustezza Per robustezza si intende la sensibilità della stima verso valori anomali. Ad esempio, abbiamo visto che la media aritmetica è uno stimatore non distorto del valore di aspettazione di misure ripetute, tuttavia è meno robusto della mediana.

Esempio:

(Tratto da I. Campa, M. Longhi) Determinare lo stimatore di massima verosimiglianza del parametro λ per un campione estratto dalla densità:

$$f(x; \lambda) = \begin{cases} \lambda^2(x-1)e^{-\lambda(x-1)} & x \geq 1 \\ 0 & \text{altrimenti} \end{cases}$$

R: La funzione di verosimiglianza per un campione di n misure prese da $f(x; \lambda)$ è: $L(\lambda; x) = \prod_{i=1}^n f(x_i; \lambda)$. Da questa, la ln-likelihood:

$$\begin{aligned} \ln \mathcal{L}(\lambda; x) &= \log\left(\prod_{i=1}^n f(x_i; \lambda)\right) \\ &= \sum_{i=1}^n \log(f(x_i; \lambda)) = \sum \log(\lambda^2) + \sum \log(x-1) + \sum(-\lambda(x-1)) = \\ &= 2 \sum \log(\lambda) + \sum(-\lambda x) + \sum(\lambda) = 2N \log \lambda - \lambda N \bar{x} + N \lambda \end{aligned}$$

La cui derivata $\frac{d \ln \mathcal{L}}{d \lambda} = 2N \frac{1}{\lambda} - N \bar{x} + N$ si annulla per $\hat{\lambda} = \frac{2}{\bar{x}-1}$

Esempio:

(Tratto da I. Campa, M. Longhi) Si supponga che X_1, X_2, X_3 siano tre elementi di un campione di misure estratto da una distribuzione esponenziale di media λ . Si considerino i seguenti stimatori del parametro :

$$\hat{\lambda}_1 = X_1; \quad \hat{\lambda}_2 = \frac{X_1 + X_2}{2}; \quad \hat{\lambda}_3 = \frac{X_1 + 2X_2}{3}; \quad \hat{\lambda}_4 = \bar{X};$$

Indicare quali sono gli stimatori non distorti di λ e individuare tra gli stimatori non distorti quello con varianza minima.

R: Una VA esponenziale con media λ ha speranza λ e varianza λ^2 . Calcoliamo la speranza dei quattro stimatori:

$$\begin{aligned} E[\hat{\lambda}_1] &= E[X_1] = \lambda \\ E[\hat{\lambda}_2] &= E\left[\frac{X_1+X_2}{2}\right] = \frac{E[X_1]+E[X_2]}{2} = \lambda \\ E[\hat{\lambda}_3] &= E\left[\frac{X_1+2X_2}{3}\right] = \frac{E[X_1]+2E[X_2]}{3} = \lambda \\ E[\hat{\lambda}_4] &= E[\bar{X}] = E\left[\frac{X_1+X_2+X_3}{3}\right] = \lambda \end{aligned}$$

Quindi ognuno è un buon stimatore di λ . Per calcolare la varianza:

$$\begin{aligned} \text{VAR}[\hat{\lambda}_1] &= \text{VAR}[X_1] = \lambda^2 \\ \text{VAR}[\hat{\lambda}_2] &= \frac{1}{4}(\text{VAR}[X_1] + \text{VAR}[X_2] + 2 \text{COV}[X_1, X_2]) = \frac{1}{4} 2 \lambda^2 = \lambda^2/2 \\ \text{VAR}[\hat{\lambda}_3] &= \frac{1}{9}(\text{VAR}[X_1] + 4 \text{VAR}[X_2]) = \frac{5}{9} \lambda^2 \\ \text{VAR}[\hat{\lambda}_4] &= \text{VAR}[\bar{X}] = \frac{1}{3} \lambda^2 \end{aligned}$$

Lo stimatore non distorto con varianza minima è la media campionaria

9.4 Incertezza di una stima

9.4.1 Limite gaussiano della verosimiglianza

La funzione verosimiglianza è ottenuta come una combinazione di misure più o meno complessa. Il teorema del limite centrale ci mostra come la combinazione lineare di variabili IID sia distribuita come una PDF Gaussiana. Cio è perciò valido anche per la verosimiglianza:

$$\lim_{N \rightarrow \infty} \mathcal{L}(\lambda) = \mathcal{G}(\lambda)$$

ovvero, all'aumentare della dimensione del campione del quale si calcola la verosimiglianza, essa assomiglia via via di piu' ad una Gaussiana. Possiamo osservare graficamente come questo avvenga in [Figura 9.1](#) e [Figura 9.2](#). Specialmente per il secondo caso, si vede come gia' a partire da 5 misure indipendenti, la forma della verosimiglianza assomigli gia' molto ad una Gaussiana.

Ma quali sono i valori di aspettazione e la varianza di questa Gaussiana? E' chiaro che la Gaussiana sara' centrata attorno al valore che massimizza la verosimiglianza, ovvero $\hat{\lambda}$. Ma vediamo che risulta facile anche la stima della varianza, infatti

$$\mathcal{L}(\lambda) = \mathcal{G}(\lambda) = \frac{1}{\sqrt{2\pi}\sigma_{\hat{\lambda}}} e^{-\frac{1}{2}\left(\frac{\lambda - \hat{\lambda}}{\sigma_{\hat{\lambda}}}\right)^2}$$

allora, come abbiamo gia' visto:

$$\ln \mathcal{L}(\lambda) = -\ln(\sqrt{2\pi}\sigma_{\hat{\lambda}}) - \frac{1}{2}\left(\frac{\lambda - \hat{\lambda}}{\sigma_{\hat{\lambda}}}\right)^2$$

Derivando due volte rispetto a λ otteniamo:

$$\frac{d^2 \ln \mathcal{L}(\lambda)}{d\lambda^2} = -\frac{1}{\sigma_{\hat{\lambda}}^2} \quad \rightarrow \quad \sigma_{\hat{\lambda}}^2 = -\frac{1}{\frac{d^2 \ln \mathcal{L}(\lambda)}{d\lambda^2}}$$

Ovvero, nell'ipotesi che la verosimiglianza non si discosti troppo da una Gaussiana, si puo' dare una stima della incertezza sul parametro stimato alternativa a quanto visto in [sezione 9.4](#):

$$\hat{\lambda} \pm \sqrt{-\frac{1}{\frac{d^2 \ln \mathcal{L}(\lambda)}{d\lambda^2}}}$$

Esempio:

Ad esempio, nel caso della verosimiglianza per misure distribuite esponenzialmente visto sopra, otteniamo che la derivata seconda della verosimiglianza e':

$$\frac{d^2 \ln \mathcal{L}(\lambda)}{d\lambda^2} = -\frac{N}{\lambda} = -\frac{N}{\bar{x}}$$

e quindi la stima MLE del parametro e della sua incertezza assunta gaussiana e':

$$s_{\hat{\lambda}} = \sqrt{\frac{\bar{x}}{N}} = \frac{\bar{x}}{\sqrt{N}} = \frac{s_x}{\sqrt{N}}$$

ovvero il solito MSE. Si noti tuttavia come da [Figura 9.2](#) questa approssimazione, anche con numerosità del campione pari a $N = 10$ non sia ancora completamente giustificata.

9.4.2 Incertezza dal profilo di verosimiglianza

Esiste una maniera piu' completa di valutare le incertezze associate ad un parametro, attraverso lo studio del profilo della verosimiglianza. Nonostante questo sia la prassi dell'analisi, per questo corso di lascia solo come approfondimento, si veda quindi [Appr. 9.2](#).

9.5 Media Ponderata

La massima verosimiglianza si puo' costruire anche per insiemi di misure che non siano necessariamente IID, ad esempio supponiamo di avere M misure indipendenti di una stessa grandezza, ma questa volta le M misure non hanno caratteristiche diverse, come ad esempio sono a loro volta medie campionarie. Il set di dati sara quindi:

$$x_1 \pm s_{x_1}; \quad x_2 \pm s_{x_2}; \dots$$

Allora se assumiamo che abbiano lo stesso valore di aspettazione x_{mp} , questo si puo' calcolare costruendo la seguente verosimiglianza:

$$\mathcal{L}(x_{mp}; \vec{x}, \vec{s}_x) = \prod_{i=1}^M \frac{1}{\sqrt{2\pi}s_i} e^{-\frac{1}{2} \left(\frac{x_i - x_{mp}}{s_i} \right)^2}$$

$$-2 \ln \mathcal{L}(x_{mp}; \vec{x}, \vec{s}_x) = \sum_{i=1}^M \left(\frac{x_i - x_{mp}}{s_{x_i}} \right)^2 + M \ln(s_{x_i}^2)$$

e derivando:

$$\frac{d}{dx_{mp}} (-2 \ln \mathcal{L}(x_{mp})) = -\frac{2}{s_x} \sum_{i=1}^M \left(\frac{x_i - x_{mp}}{s_{x_i}} \right) \rightarrow x_{mp} \sum_{i=1}^M \frac{1}{s_{x_i}^2} - \sum_{i=1}^M \frac{x_i}{s_{x_i}^2} = 0$$

da cui

$$x_{mp} = \frac{1}{\sum_{i=1}^M \frac{1}{s_{x_i}^2}} \sum_{i=1}^M \frac{x_i}{s_{x_i}^2} \tag{9.12}$$

Si noti come [Equazione 9.12](#) torna ad essere la media aritmetica se tutti i pesi $s_{x_i}^2$ sono uguali. L'incertezza sulla media pesata si puo' ora calcolare con la propagazione degli errori considerando che essa e' una combinazione lineare di variabili aleatorie. Allora se $k = \sum_{i=1}^M 1/s_{x_i}^2$ possiamo scrivere

$$s_{x_{mp}}^2 = \sum_{i=1}^M \left(\frac{1}{k s_{x_i}^2} \right)^2 s_{x_i}^2 = \frac{1}{k} \quad s_{x_{mp}} = \sqrt{\frac{1}{\sum_{i=1}^M \frac{1}{s_{x_i}^2}}}$$

e di nuovo se tutti gli M pesi sono uguali allora $s = s_x$.

Dati M misure $x_i \pm s_{x_i}$ la loro media ponderata vale:

$$x_{mp} = \frac{1}{k} \sum_{i=1}^M \frac{x_i}{s_{x_i}^2} \quad s_{x_{mp}} = \sqrt{\frac{1}{k}} \quad \text{con} \quad k = \sum_{i=1}^M \frac{1}{s_{x_i}^2} \tag{9.13}$$

Esempio:

Due campioni di misure periodo di oscillazione del pendolo danno:

$$\bar{x}_1 = (2.00 \pm 0.02) \text{ s} \quad \bar{x}_2 = (2.02 \pm 0.04) \text{ s}$$

I due valori sono compatibili: $r \simeq 0.4$ allora posso valutare la media pesata. Il secondo valore ha incertezza maggiore quindi la media dovrebbe essere piu' vicina al primo, vediamo. Media ponderata:

$$x_{mp} = \frac{1}{\sum_{i=1}^N \frac{1}{s_{x_i}^2}} \sum_{i=1}^N \frac{\bar{x}_i}{s_{x_i}^2} = \frac{\frac{2.00}{0.02^2} + \frac{2.02}{0.04^2}}{\frac{1}{0.02^2} + \frac{1}{0.04^2}} = 2.004 \text{ s}$$

Che conferma quello che ci aspettavamo. L'incertezza:

$$s_{x_{mp}} = \sqrt{\frac{1}{\sum_{i=1}^N \frac{1}{s_{x_i}^2}}} = \frac{1}{\frac{1}{0.02^2} + \frac{1}{0.04^2}} = 0.00032 \text{ s}$$

e quindi

$$x_{mp} = (2.0040 \pm 0.0003) \text{ s}$$

9.6 Chiusura

9.6.1 Quesiti

□ **Es. 9.1** [da I. Campa, M. Longhi] Si consideri per $\lambda > 0$ la funzione cosi definita:

$$f(x; \lambda) = \begin{cases} \frac{4}{\lambda^2} x & 0 \leq x \leq \lambda/2 \\ \frac{4}{\lambda^2} (\lambda - x) & \lambda/2 \leq x \leq \lambda \\ 0 & \text{altrimenti} \end{cases}$$

- 1) Verificare che per ogni $\lambda > 0$, $f(x; \lambda)$ rappresenta una funzione di densita' di probabilita'
- 2) Sia X_1, X_2, \dots, X_n un campione casuale estratto dalla popolazione di densita' $f(x; \lambda)$, stabilire se

$$\hat{\lambda} = \frac{2}{n} \sum_{i=1}^n X_i$$

è un corretto stimatore di λ .

□ **Es. 9.2** Nel 2018 a Padova, il numero di incidenti stradali in 10 giornate con pioggia scelte a caso è stato di:

$$4, 0, 6, 5, 2, 1, 2, 0, 4, 3,$$

Si usino questi dati per stimare la media e la deviazione standard per quell'anno di incidenti in caso di pioggia, e la frazione di giornate con pioggia con meno di 2 incidenti. Siccome vi è un elevato numero di automobilisti, ciascuno dei quali ha solo una piccola probabilita' di essere coinvolto in un incidente stradale, è ragionevole assumere che il numero di incidenti quotidiani sia una variabile aleatoria di Poisson.

□ **Es. 9.3** [I. Campa, M. Longhi] Sia X_1, X_2, \dots, X_n un campione casuale di ampiezza n estratto da una popolazione con densita' di probabilita':

$$f_X(x; \lambda) = \begin{cases} 5^{-\lambda} \lambda x^{\lambda-1} & 0 < \text{le}q x \leq 1 \\ 0 & \text{altrimenti} \end{cases}$$

con $\lambda > 0$. Determinare lo stimatore $\hat{\lambda}$ con il metodo della massima verosimiglianza.

□ **Es. 9.4** Se prendiamo un insieme di misure che ci aspettiamo distribuite secondo la seguente densità di probabilità:

$$f(x; \lambda) = \begin{cases} \lambda^2(x-2)e^{-\lambda(x-2)} & x \geq 2 \\ 0 & \text{altrimenti} \end{cases}$$

a) Si indichi la formula (senza calcolarla) che fornisce la miglior stima $\hat{\lambda}$ del parametro λ per un campione di misure x_i usando il metodo della massima verosimiglianza; b) Si calcoli $\hat{\lambda}$ definita in [3.1] per un per un campione di misure x_i ; c) Quale e' la moda della distribuzione?

9.6.2 Approfondimenti

Appr. 9.1 Parametri di interesse e di rumore. Nel caso generale, il set di parametri $\vec{\lambda}$ puo' essere vario. Inoltre, alcuni di questi possono essere associati solo ad un sottoinsieme dei prodotti della funzione verosimiglianza, ad esempio, quelli associati ad una variabile aleatoria piuttosto che ad un'altra. In generale, comunque la massimizzazione della verosimiglianza si riferisce ad un *parametro di interesse* (*parameter of interest*) mentre altri parametri possono essere *parametri di rumore* (*nuisance parameters*). Ad esempio, tutti i fattori esterni che influenzano la misura visti in [Capitolo 1](#) possono essere associati a misure e quindi a parametri il cui valore non ci interessa direttamente, ma ci interessa conoscere il loro effetto. A questo punto si puo' procedere separando logicamente le due classi di parametri: $\vec{\lambda} = \lambda(\vec{i}) + \vec{v}$ dove $\lambda(\vec{i})$ sono i parametri di interesse e \vec{v} sono i nuisance parameters. In generale si fissano i valori di \vec{v} in maniera che massimizzino la funzione di verosimiglianza, quindi

$$\mathcal{L}(\lambda(\vec{i}); \hat{\vec{v}}, \vec{x}) = \prod_{i=1}^N f_{X_i}(x_i; \lambda(\vec{i}), \hat{\vec{v}}) \tag{9.14}$$

Appr. 9.2 Incertezza dal profilo di verosimiglianza. Abbiamo visto come una \mathcal{L} tenda asintoticamente ad una Gaussiana all'aumentare della numerosità del campione. Si noti la seguente proprietà matematica della Gaussiana: in corrispondenza dei punti a distanza $\pm 1\sigma, \pm 2\sigma \pm 3\sigma$ la Gaussiana vale rispettivamente $e^{-1/2}, e^{-2}, e^{-9/2}$ meno rispetto al massimo $1/\sqrt{2\pi}\sigma$. Possiamo usare questo fatto per valutare l'incertezza attorno ad un parametro. Infatti, se prendiamo i valori della \mathcal{L} a $e^{-1/2}, e^{-2}, e^{-9/2}$ rispetto al massimo, possiamo, nel limite asintotico, pensare che a questi corrispondano i margini di confidenza Gaussiani del 68.3, 95.5, 99.7%. In termini di $\ln \mathcal{L}$ il calcolo e' piu' semplice, infatti:

$$\begin{aligned} \mathcal{L}(\lambda_1^\pm) &= \frac{1}{e^{-1/2}} \mathcal{L}(\hat{\lambda}) & \text{allora} & \quad \ln \mathcal{L}(\lambda^\pm) = \ln \mathcal{L}(\hat{\lambda}) - 0.5 \\ \mathcal{L}(\lambda_2^\pm) &= \frac{1}{e^{-2}} \mathcal{L}(\hat{\lambda}) & \text{allora} & \quad \ln \mathcal{L}(\lambda^\pm) = \ln \mathcal{L}(\hat{\lambda}) - 2 \\ \mathcal{L}(\lambda_3^\pm) &= \frac{1}{e^{-9/2}} \mathcal{L}(\hat{\lambda}) & \text{allora} & \quad \ln \mathcal{L}(\lambda^\pm) = \ln \mathcal{L}(\hat{\lambda}) - 4.5 \end{aligned}$$

Allora per trovare le incertezza sulla stima e' sufficiente disegnare la \mathcal{L} o la $\ln \mathcal{L}$ in funzione del parametro di interesse, e trovare i punti λ_1^\pm ad esempio per quantificare un margine di confidenza sul parametro di interesse di circa il 68.3. Mostriamo ad esempio questo in [Figura 9.3](#) per i casi di verosimiglianze gaussiane ed esponenziali. Si noti che per la verosimiglianza Gaussiana $-2 \ln \mathcal{L}$ i valori vanno moltiplicati per 2.

In realtà, per ovviare al fatto che talvolta si usa $\ln \mathcal{L}$, talvolta $-\ln \mathcal{L}$ e per la Gaussiana $-2 \ln \mathcal{L}$, spesso si utilizza la $\Delta \ln \mathcal{L}$:

$$\Delta \ln \mathcal{L}(\lambda) = |\ln \mathcal{L}(\lambda) - \ln \hat{\mathcal{L}}| \quad \text{o} \quad -2\Delta \ln \mathcal{L}(\lambda) = 2|\ln \mathcal{L}(\lambda) - \ln \hat{\mathcal{L}}|$$

Allora un margine di confidenza (asintotico) del 68.3% (1σ) corrisponde a

$$\Delta \ln \mathcal{L}(\lambda^\pm) = 0.5 \quad (= 1 \text{ per la } -2 \ln \mathcal{L})$$

un margine di confidenza (asintotico) del 95.5% (2σ) corrisponde a

$$\Delta \ln \mathcal{L}(\lambda^\pm) = 2 \quad (= 4 \text{ per la } -2 \ln \mathcal{L})$$

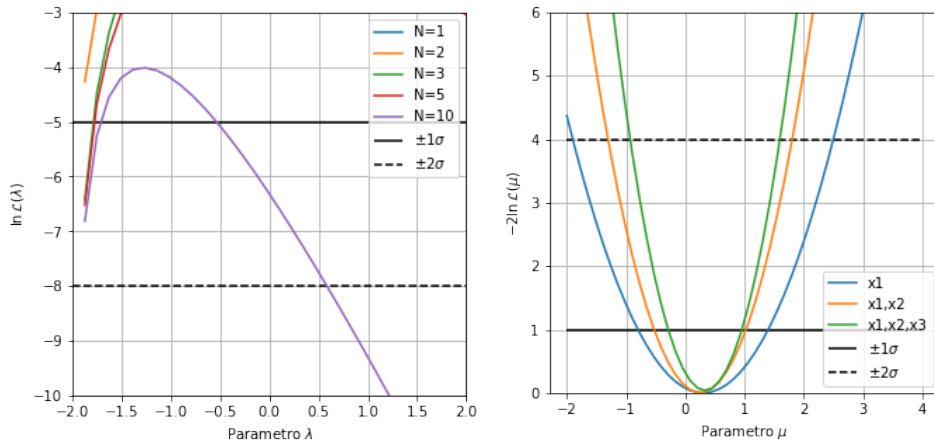


Figura 9.3: Costruzione del margine di confidenza attorno alla miglior stima del parametro

e così via. Si usa spesso anche considerare i margini di confidenza del 95% (1.96σ) e 99% (2.576σ) corrispondono a

$$\Delta \ln \mathcal{L}(\lambda^\pm) = 1.96 \quad (= 3.92 \text{ per la } -2 \ln \mathcal{L})$$

$$\Delta \ln \mathcal{L}(\lambda^\pm) = 2.58 \quad (= 5.1 \text{ per la } -2 \ln \mathcal{L})$$

Si noti che nel caso della Gaussiana gli errori sono simmetrici, e il margine di confidenza è esattamente quello Gaussiano. Non è così nel caso generale, in cui il comportamento Gaussiano è solo asintotico. In questo caso, come si vede chiaramente in [Figura 9.3](#), si avrebbero errori asimmetrici. In questo caso, il parametro di interesse viene tipicamente riportato come:

$$\lambda = \hat{\lambda}^+_{-sx}$$

dove $dx = |\hat{\lambda} - \lambda_1^+|$ e $sx = |\hat{\lambda} - \lambda_1^-|$. Questa scrittura è molto diffusa, ma si noti come sia fonte di errori. Ad esempio, come si propaga la incertezza legata a questo parametro? In un ottimo lavoro, [D'Agostini \[2004\]](#) mostra come in realtà si sarebbe dovuto procedere, avendo a disposizione tutta la funzione di verosimiglianza, a calcolare il valore di aspettazione e la varianza con le tecniche dei momenti apprese nel [Capitolo 5](#). Si noti come in questo caso, il valore di aspettazione del parametro non coincida con la massima verosimiglianza (che rappresenta la moda della verosimiglianza se la intendiamo come distribuzione di probabilità). Tuttavia, l'autore mostra come comunque a partire dalla formulazioni con gli errori asimmetrici sopra si possa, con un buon grado di accuratezza, stimare il valore di aspettazione e la deviazione standard del parametro attraverso correzioni numeriche. Questo argomento, per quanto decisamente importante e spesso poco correttamente usato in letteratura, è troppo avanzato per questo corso e lasciato all'interesse della lettrice.

Appr. 9.3 Interpretazione della verosimiglianza. La lettrice più attenta si è forse resa conto che avevamo già usato il termine verosimiglianza nel contesto della formula di Bayes che riproponiamo qui per facilità. Sia H, D due variabili aleatorie, allora vale:

$$P(H|D) = \frac{P(H)}{P(D)} P(D|H)$$

Avevamo visto come *a)* se D rappresenta l'evento aleatorio di ottenere un certo insieme di dati sperimentali, e H rappresenta l'evento aleatorio (bernoulliano) che una ipotesi sia vera o falsa, allora la formula di Bayes da origine all'interpretazione bayesiana della statistica e *b)* il termine $P(D|H)$ era detto *likelihood*, i termini $P(D), P(H)$ *priors*, e il termine $P(H|D)$ *posterior*. Come mai $P(D|H)$ si chiama così? Ma perché quel termine è proprio la funzione di verosimiglianza descritta in [Equazione 9.2](#). Usando una notazione che li metta in relazione $D = \vec{x}$ e se l'ipotesi è parametrica $H = H(\vec{\lambda})$, allora:

$$P(D|H) = P(D|H(\vec{\lambda})) = P(\vec{x}|H(\vec{\lambda})) = \mathcal{L}(\vec{\lambda}; D) = \mathcal{L}(\vec{\lambda}; \vec{x}) \quad (9.15)$$

Essi sono la stessa cosa infatti in entrambi i casi si tratta di una probabilità condizionata, condizionata al fatto che si verifichi $H(\vec{\lambda})$ che è data per vera. La verosimiglianza infatti è calcolata assumendo come distribuzioni di probabilità una certa distribuzione ipotizzata.

Ora la funzione di verosimiglianza è sempre calcolabile a partire dai dati data una certa ipotesi. Questo è quello che fa l'approccio frequentista all'analisi statistica. E qui si ferma. L'approccio bayesiano, attraverso la formula rivista sopra, si spinge più in là e fornisce la probabilità *posterior* $P(H|D)$ ovvero la probabilità che una ipotesi sia vera dopo che sono stati presi i dati. Ci si soffermi sulla differenza tra *posterior* e *likelihood* $P(H|D)$ e $P(D|H)$. Il significato della interpretazione bayesiana è quindi questo: prima di prendere delle misure, ho una idea a priori (prior) della probabilità che una ipotesi sia vera $P(H)$, dopo aver preso i dati, la cui probabilità di occorrere a priori è $P(D)$ (che è un termine di normalizzazione nel caso bayesiano, che può non essere calcolato), posso calcolare la *likelihood* $P(D|H)$ ovvero la probabilità che ottenessi quelle misure assumendo per vera una ipotesi, e con questa *aggiornare* la stima della probabilità della ipotesi dopo aver osservato i dati: $P(H|D)$.

Capitolo 10

Il Test del χ^2

Quest'opera è distribuita con licenza Creative Commons "Attribuzione – Non commerciale – Condividi allo stesso modo 3.0 Italia".



Nei [Capitolo 8](#) e [Capitolo 9](#) abbiamo avuto a che fare con *ipotesi di riferimento*. In entrambi avevamo una certa ipotesi parametrica di riferimento, che davamo per valida, e tra tutte queste sceglievamo quella (ovvero quel set di parametri) che rendeva l'ipotesi piu' vicina alle misure effettuate. In questo capitolo ci occuperemo invece di *valutare la bontà* di questa ipotesi, ovvero il grado di 'vicinanza' ai dati. L'approccio a questo problema, come già' discusso, e' risolto in Statistica attraverso due metodi principali. Il metodo frequentista e il metodo Bayesiano Di quest'ultimo abbiamo parlato in [Approf. 4.3](#) e ci siamo tornati su nel [Capitolo 9](#). Lo studio del metodo Bayesiano non rientra in questo corso. Si consiglia al lettore curioso il testo [?]. Vedremo tuttavia che non valuteremo direttamente la veridicità o falsità di una ipotesi, bensì esclusivamente il grado di vicinanza o lontananza dei dati da una certa ipotesi, detta ipotesi nulla, presa di volta in volta per vera. Il metodo frequentista ci guida valutando come si comportano i dati sul lungo termine, ovvero, nel caso si potesse ripetere un numero infinito di volte l'esperimento nelle stesse identiche condizioni.

10.1 La PDF del χ^2

Nei [Capitolo 8](#) e [Capitolo 9](#) abbiamo incrociato diverse volte, sia nella formula dei minimi quadrati [Equazione 8.2](#) sia nella formula della massima verosimiglianza per una serie di dati con incertezze Gaussiane [Equazione 9.3](#) il termine:

$$d_i = \frac{x_i - \mu}{\sigma},$$

dove x_i era una certa misura, μ un certo valore di riferimento e σ una certa incertezza. Come prima cosa notiamo che, se μ, σ sono il valore di aspettazione e la deviazione standard di x_i e quindi della popolazione X allora si puo' costruire una variabile aleatoria

$$D_i = \frac{X_i - \mu}{\sigma} := \mathcal{G}_D(d; 0, 1)$$

equivalente ad una Gaussiana standard, ovvero con media 0 e varianza 1. Ora, in [Equazione 8.2](#) e [Equazione 9.3](#) questa grandezza appariva al quadrato e si sommava per ogni misura x_i :

$$\mathcal{Q} = -2 \ln \hat{\mathcal{L}} = \sum_{i=1}^N d_i^2 = \sum_{i=1}^N \left(\frac{x_i - x_i^*}{s_{x_i}} \right)^2 := \chi_m^2 \quad (10.1)$$

dove abbiamo generalizzato il fatto che ciascun valore x_i puo' avere diverso valore di riferimento x_i^* e diverso SE s_{x_i} . Trattandosi di una sommatoria di numeri, il risultato sarà un numero che per ragioni che vedremo chiamiamo χ_m^2 . Trattandosi di quadrati di Gaussiane standard, notiamo anzitutto il

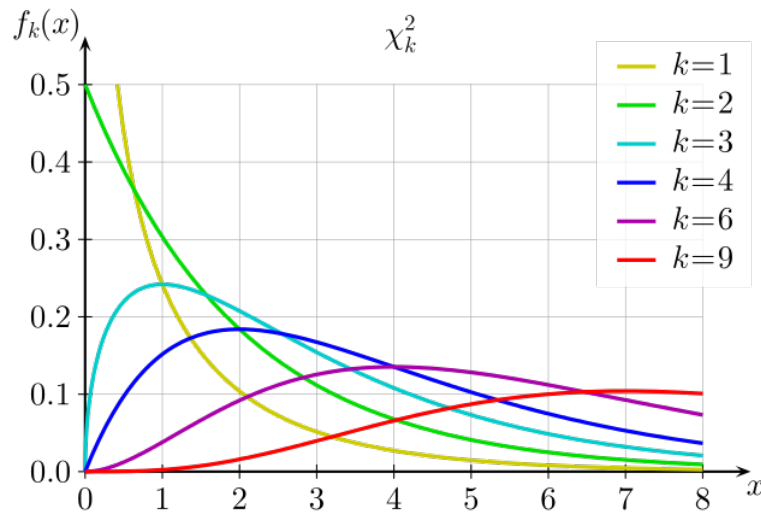


Figura 10.1: PDF del χ^2 per diversi gradi di libertà $k = 1, 2, 3, 4, 6, 9$. Nel grafico il variato e' indicato con x ma in realtà si tratta di $x = \chi^2$.

seguinte fatto. Mentre il valore di aspettazione di $E[D_i] = 0$, il valore di aspettazione di ogni $D_i^2 = 1$ vale:

$$E[D_i^2] = E[D_i]E[D_i] + \text{VAR}[D_i] = 0 \cdot 0 + 1 = 1 \quad (10.2)$$

Torniamo a [Equazione 10.1](#). Essendo una combinazione di variabili aleatorie, χ_m^2 sarà a sua volta associabile ad una variabile aleatoria. Ma quale? Vista la definizione di D_i come quadrati di Gaussiane standard, allora χ_m^2 sarà un variato della distribuzione:

$$f_{\chi^2} = \sum_{i=1}^k \mathcal{G}^2(0, 1) \quad (10.3)$$

La sommatoria si estende a k invece che N per il fatto che, come vedremo, la sommatoria [Equazione 10.1](#) non avrà N termini indipendenti, ma solo $N - \nu$, dove ν è il cosiddetto numero di vincoli, ovvero, anticipando concetti che vedremo, il numero di volte che usiamo i dati stessi per stimare i valori di riferimento x_i^* . Il numero k è noto come *numero di gradi di libertà (GDL, NGDL)* (o *number of degrees of freedom (DOF, NDOF)* in inglese).

Svolgendo la sommatoria in [Equazione 10.3](#) si ottiene una funzione di χ^2 che prende il nome di PDF del χ^2 . La funzione f_{χ^2} ha forma diversa in funzione del numero k di elementi della sommatoria, ed e' riportata in [Figura 10.1](#). Nel caso $k = 1$ la PDF χ^2 ha la forma di una esponenziale, per tutti i casi superiori ha una forma a campana asimmetrica. La forma funzionale non e' particolarmente importante da ricordare, ma e' opportuno vederla una volta scritta:

$$f_{\chi^2}(\chi^2; k) = \begin{cases} \frac{(\chi^2)^{k/2-1}}{2^{k/2} \Gamma(k/2)} e^{-\chi^2/2}, & \chi^2 > 0; \\ 0, & \text{altrimenti.} \end{cases} \quad (10.4)$$

dove $\Gamma(k/2)$ denota la *funzione gamma* che è l'estensione al caso dei numeri reali della funzione fattoriale. Questa funzione si trova nelle principali librerie di calcolo. Il suo legame con il fattoriale per i numeri interi e' $\Gamma[n] = (n - 1)!$. Si noti come il termine $e^{-\chi^2/2}$ è simile al termine analogo della PDF Gaussiana.

La PDF del χ^2 ha le seguenti caratteristiche:

- il variato è χ^2 . Si puo' indicare ovviamente anche con x purché non si faccia confusione tra simboli.
- il valore di aspettazione di f_{χ^2} e' pari a $E[\chi^2] = k$. Infatti, il valore di aspettazioni di ciascun addendo come visto in [Equazione 10.2](#) e' 1.

- La mediana é $\simeq k \left(1 - \frac{2}{9k}\right)^3$ e quindi k non é necessariamente il punto di massimo ma si avvicina progressivamente ad esso per $k \rightarrow \infty$.
- la varianza di f_{χ^2} é pari a $\text{VAR}(\chi^2) = 2k$
- all'aumentare di k , secondo il teorema del limite centrale (Appr. 3.3) allora $f_{\chi^2} \rightarrow \mathcal{G}(k, 2k)$ ovvero la distribuzione del χ^2 tende ad una Gaussiana di media k e varianza $2k$.

Quale é il senso di definire f_{χ^2} ? Qui entra in gioco il ragionamento frequentista. Questo ragionamento si basa sul comportamento *a lungo termine* in cui lo stesso esperimento da cui abbiamo ottenuto la serie di misure, si possa ripetere un numero infinito di volte nelle stesse condizioni. In ciascuna ripetizione, naturalmente si otterrà un insieme diverso di misure, a causa delle fluttuazioni intrinseche della misura, altresì detto del campionamento finito. Si tratta di un comportamento ipotetico in quanto non é sempre possibile ripetere un esperimento, e certamente non é possibile ripeterlo infinite volte. Supponiamo comunque di farlo. In generale ogni campione darà il suo set di misure \vec{x}_i e per ciascun campione potremo quindi calcolare χ_m^2 secondo [Equazione 10.1](#). Ecco qui che entra la PDF del χ^2 : aumentando il numero di campioni a infinito, possiamo calcolare la (densità di) probabilità (ovvero se vogliamo la frequenza assoluta di occorrenza di ciascun χ_m^2) e così facendo otterremo la PDF di [Equazione 10.4](#) e [Figura 10.1](#). Questa curva rappresenta quindi la distribuzione di probabilità di χ_m^2 per campioni di k gradi di libertà (parametro come visto legato alla numerosità n del campione e al numero di vincoli). Si noti come quindi per la maggior parte dei campioni (ovvero degli esperimenti reali), ci aspettiamo di ottenere $\chi_m^2 \simeq k$ attorno al valore di aspettazione, in alcuni casi, via via più rari, possiamo ottenere valori molto piccoli, tendenti a zero, di χ_m^2 . Infine possiamo ottenere valori molto alti di χ_m^2 , e senza un limite superiore, essendo la funzione definita su tutto \mathbb{R}^+ , ma di nuovo questi valori sono via via meno probabili. Sulla base della forma di questa curva costruiremo il test statistico del χ^2 , vediamo come.

10.2 La costruzione del test del χ^2

Per la costruzione del test statistico torniamo a osservare [Equazione 10.1](#). Osserviamo come questa equazione sia composta da un termine che é la misura x_i , un termine che é l'incertezza della misura s_{x_i} e un termine che é un valore di riferimento x_i^* . Si ragioni sul fatto che data una misura, la scelta del valore di riferimento diventa una scelta di una ipotesi sulle misure stesse. Ad esempio, se x_i sono misure ripetute di una grandezza allora il valore di riferimento potrebbe essere il valore di aspettazione della popolazione di misure, ma se x_i sono misure generiche, i valori di riferimento potrebbero essere ad esempio quelli della retta interpolante vista in [Capitolo 8](#). Insomma, la scelta del valore di riferimento rappresenta il fissare di una *ipotesi* sulle misure. Il χ_m^2 quindi é intimamente legato alla scelta di una certa ipotesi: scegliendo diverse ipotesi, si otterranno diversi χ_m^2 .

Ora, siccome conosciamo la PDF allora possiamo calcolare ad esempio la probabilità che il χ^2 sia maggiore di quello ottenuto dalla nostra specifica misura $\chi^2 > \chi_m^2$:

$$P(\chi^2 > \chi_m^2) = \int_{\chi_m^2}^{\infty} f_{\chi^2}(\chi^2) d\chi^2 := p\text{-value} \quad (10.5)$$

questo valore é detto *p-value*. Osservando [Figura 10.2](#), esso equivale a:

- l'area sotto la funzione f_{χ^2} tra χ_m^2 e infinito
- la percentuale di campioni che sul lungo termine (ovvero ripetendo infinite volte l'esperimento) avrebbero restituito un $\chi^2 > \chi_m^2$.
- la probabilità di trovare un campione con $\chi^2 > \chi_m^2$

Si ribadisce quindi che a causa delle fluttuazioni casuali della misura, ci troveremo ad avere di volta in volta χ_m^2 diversi, talvolta grandi talvolta piccoli, tuttavia, essi non sono equiprobabili: valori molto piccoli e molto grandi di χ_m^2 sono improbabili. Sono valori probabili quelli vicino al valore di aspettazione $\chi_m^2 \simeq k$.

Il valore *p-value* é anche detto *significanza statistica* del campione di misure. In linea generale, é piuttosto difficile calcolarlo (si tratta infatti della CDF del χ^2 ovvero del calcolo dell'integrale di

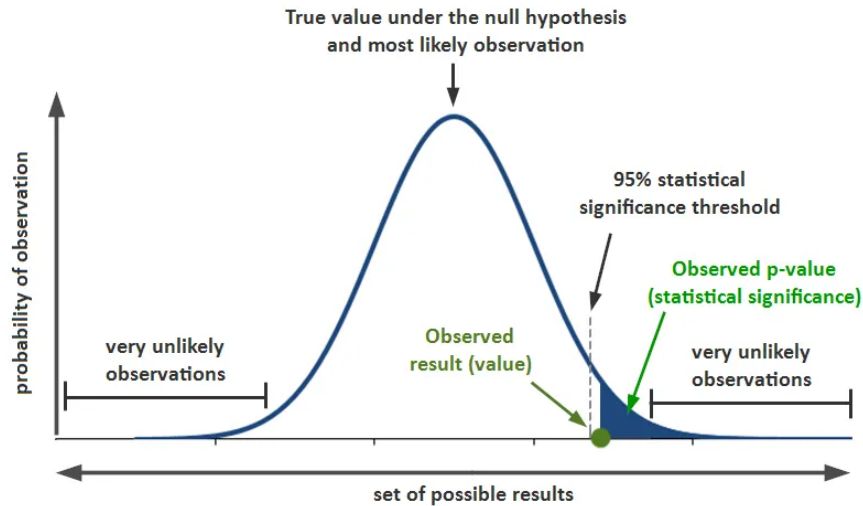


Figura 10.2: Schema della valutazione dell'ipotesi nulla attraverso il p -value e il margine di confidenza.

Chi-Square (χ^2) Distribution										
Degrees of Freedom	Area to the Right of Critical Value									
	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	—	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.299
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.257	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

Figura 10.3: Tabella dei valori critici del χ^2 . Sulle colonne ci sono i livelli di significanza, sulle righe i gradi di libertà.

Equazione 10.5). Questo è fatto normalmente attraverso librerie di calcolo nei diversi linguaggi di programmazione¹. Per questa ragione, spesso si fa riferimento a valori tabulati di *livelli di significanza*, per i quali convenzionalmente si usa il simbolo α . Una di queste tabelle è riportata in Figura 10.3. Nella tabella ogni colonna rappresenta un diverso livello di significanza (di solito si usano i valori delle colonne di destra, ovvero $\alpha = 0.005, 0.01, 0.05, 0.1$). Nelle righe di riportano i diversi valori del χ^2 (detti valori critici o *critical values*) per i diversi livelli di significanza e per i diversi gradi di libertà. Ad esempio, per 9 gradi di libertà, la probabilità di ottenere un $\chi_m^2 > 14.684$ è pari a $\alpha = 0.1 = 10\%$, e così via. Si noti infine che per ogni livello di significanza si può definire il *livello di confidenza* (*confidence level (CL)*) definito come $CL = 1 - \alpha$.

10.3 Esecuzione del test del χ^2

L'esecuzione del test del χ^2 è piuttosto meccanica, come vedremo. Attenzione che invece l'interpretazione del risultato non lo è. Partiamo dal test.

1. **[Formulazione ipotesi nulla]** Si formula anzitutto la ipotesi sui dati, ad esempio "i dati sono allineati su una retta", o "i dati seguono una distribuzione Gaussiana", etc. Questa ipotesi è comunemente detta **ipotesi nulla** e indicata con \mathcal{H}_0 . La ragione, su cui torneremo, sta nel fatto che il test del χ^2 non può confermare o confutare la veridicità di questa ipotesi, ma al più potrà escludere che, assunta vera, i dati la confermino. In questo senso questa ipotesi può solo essere 'nullificata', e da qui il nome. Il test del χ^2 è svolto *assumendo vera* \mathcal{H}_0 .
2. **[Calcolo valori riferimento]** Ora l'ipotesi nulla può essere parametrica o no. Nel caso ad esempio di un andamento lineare, l'ipotesi è parametrica (i due parametri a, b della retta $y = a + bx$), nel caso della distribuzione Gaussiana i due parametri della popolazione. Allora dovremo anzitutto trovare i parametri $\vec{\lambda}$ che *meglio descrivono i dati assunta per vera l'ipotesi*. Uno dei principali procedimenti per trovare i parametri che meglio approssimano i dati si basa sul procedimento della massima verosimiglianza che abbiamo visto in Capitolo 9. A seguito di questo o altri procedimenti otteniamo quindi le migliori stime dei parametri $\hat{\lambda}$ e da queste i valori di riferimento

$$x_i^* = x_i^*(\hat{\lambda})$$

3. **[Calcolo χ_m^2]** A questo punto, calcoliamo i singoli $d_i^2 := \chi_i^2$ di Equazione 10.1 e sommandoli otteniamo χ_m^2 . Ricordiamoci che il valore di aspettazione di ogni χ_i^2 è 1, quindi è importante valutare non solo la somma ma i singoli addendi, prestando particolare attenzioni a quelli troppo grandi o troppo piccoli.
4. **[Valutazione ipotesi]** A questo punto, se abbiamo modo, possiamo calcolare la significanza statistica, ovvero il p -value, che corrisponde al χ_m^2 . Questo calcolo come visto può essere complesso a meno di non aver accesso a routine numeriche al calcolatore. Tuttavia, nella maggior parte delle applicazioni, non è neppure necessario, come vedremo, perché è sufficiente valutare il χ_m^2 rispetto a i valori critici di χ_0^2 di Figura 10.3 in corrispondenza del livello di significanza α . Allora la conclusione del test può avvenire in due maniere:

- (a) Se confrontiamo con un valore di riferimento χ_0^2 (e quindi in riferimento ad un livello di significanza α_0), concludiamo il test dicendo:
 - se $\chi_m^2 > \chi_0^2$, **il test non è superato** con livello di significanza α_0 (o con livello di confidenza $1 - \alpha_0$)
 - se $\chi_m^2 < \chi_0^2$, **il test è superato** secondo un livello di significanza α_0 (o con livello di confidenza $1 - \alpha_0$).
- (b) Riportando semplicemente il p -value, o la significanza statistica corrispondente al χ_m^2 , senza valutazione.

¹Per un calcolatore online si veda ad esempio <http://courses.atlas.illinois.edu/spring2016/STAT/STAT200/pchisq.html>

10.3.1 Interpretazione del test

Per l'interpretazione del test dobbiamo tornare a ragionare sulla ragioni che possono rendere grande o piccolo il χ_m^2 e i suoi singoli addendi:

- affinché χ_i^2 sia grande e' necessario che il numeratore di d_i sia grande in relazione al denominatore. Ciò avviene quando il valore misurato e' *distante dal valore di riferimento per diverse deviazioni standard*, come nel caso di incompatibilità tra una misura e un valore di riferimento.
- affinché χ_i^2 sia piccolo, e' necessario che il numeratore sia piccolo in relazione al denominatore. Ciò avviene di norma quindi quando *l'incertezza della misura e' molto grande*.
- se $\chi_i^2 \simeq 1$ allora numeratore e denominatore sono simili. Cio' puo' avvenire quando la misura è entro una deviazione standard dal valore di riferimento.

Si noti quindi, anche in relazione alla figura [Figura 10.2](#), come siano classificati improbabili sia valori molto alti di χ^2 sia valori molto piccoli. Mentre l'interpretazione per valori alti e' chiara, si ragioni su quella che restituisce valori piccoli.

Allora l'interpretazione del test e' la seguente:

1. se p -value è $p \ll 1$, ovvero $\chi_m^2 \gg k$ o rispetto ad un certo livello di significanza statistica $\alpha_0 < 0.1$ ($\chi_m^2 \gg \chi_0^2$), e quindi quando il test non e' superato, allora la conclusione e'

Assunta vera \mathcal{H}_o , i dati non supportano questa ipotesi con un livello di confidenza di $1 - \alpha_0$

2. se p -value è $p \simeq 1$, ovvero $\chi_m^2 \ll k$ o rispetto ad un certo livello di significanza statistica $\alpha_0 < 0.1$ ($\chi_m^2 \ll \chi_0^2$), il test e' superato, ma la conclusione e'

Assunta vera \mathcal{H}_o , i dati supportano questa ipotesi ma cio' potrebbe essere dovuto ad una sovrastima delle incertezze sperimentali oppure al caso di *overfitting*²

3. se p -value non e' nè troppo vicino a 0 o a 1, ovvero χ_m^2 non troppo distante da k o comunque in relazione ad un certo livello di significanza statistica α_0 (piccolo) si ottiene $\chi_m^2 < \chi_0^2$, allora il test e' superato, e la conclusione e':

Assunta vera \mathcal{H}_o , i dati supportano questa ipotesi con un livello di confidenza di $1 - \alpha_0$

Abbiamo gia' discusso il caso 2) ovvero di incertezze possibilmente sovrastimate. Ragioniamo ancora sugli altri due casi. Il caso 3) ribadiamo ancora una volta, non afferma che l'ipotesi \mathcal{H}_0 sia vera, ma solo che vi sia compatibilità dei dati con questa ipotesi. Ad esempio, potremmo definire una seconda ipotesi e trovare un grado di compatibilità ancora maggiore rispetto a quella ipotesi. Nel caso 1) invece, di nuovo non possiamo affermare che l'ipotesi sia falsa. Le ragioni sperimentali che possono portare ad avere un χ_m^2 alto sono diverse e possono essere legate anche alla presenza di errori sistematici nella misure. Per chiarire tutto questo procediamo con un esempio concreto.

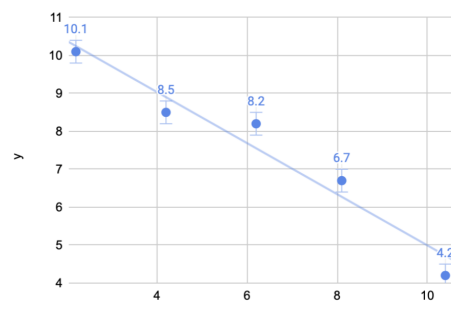
10.3.2 Un esempio: interpolazione lineare

Si consideri il set di misure

$$\vec{x} = (2.2, 4.2, 6.2, 8.1, 10.4)$$

$$\vec{y} = (10.1, 8.5, 8.2, 6.7, 4.2); \quad s_y = 0.3$$

e si considerino le \vec{x} senza incertezza, e le \vec{y} aventi incertezza normale uguale per tutti. Si vuole testare l'ipotesi che queste due serie siano legate da relazione lineare. Per questo, con il metodo della massima verosimiglianza si stimano i coefficienti



²In questa condizione l'accordo tra fit e dati e' fin troppo buono nel senso che e' valido solo per quello specifico campione di dati e non per tutti quelli analoghi ad esso per via della fluttuazioni. Ad esempio, usando come fit una linea curva per passa per tutti i punti.

della retta interpolante, ottenendo la retta $y = 11.7 - 0.674x$. Ci si chiede i dati supportino questa ipotesi con un margine di confidenza superiore al 99%.

Anzitutto, grafichiamo i dati: se ne abbiamo modo, e' sempre la maniera piu' diretta per valutare i dati. Osserviamo come tutto sommato nel complesso i dati sembrino allineati. La covarianza e l'indice di correlazione sono dati rispettivamente: $s_{xy} = -5.6$ e $r_{xy} = -0.97$. Per sviluppare il test del χ^2 dobbiamo anzitutto definire l'ipotesi di riferimento. In questo caso, l'ipotesi di riferimento e' che i punti stiano su una retta. Nel nostro caso, l'ipotesi riguarda solo le \vec{y} poichè le \vec{x} non hanno incertezza. Anzi, il valore di riferimento e' costruito sulla conoscenza delle \vec{x} . Il valore di riferimento e' quindi, punto per punto:

$$y_i^* = 11.7 - 0.674x_i \quad \text{ovvero} \quad \vec{y}^* = 10.25, 8.9, 7.55, 6.27, 4.72$$

mentre diciamo che gli x_i sono gia' i loro valori di riferimento, non avendo incertezza: $x_i = x_i^*$. Allora, il χ_m^2 si calcola:

$$\begin{aligned} \chi_m^2 &= \sum_{i=1}^5 \chi_i^2 = \sum_{i=1}^5 \left(\frac{y_i - y_i^*}{s_y} \right)^2 \\ &= \left(\frac{10.1 - 10.25}{0.3} \right)^2 + \left(\frac{8.5 - 8.9}{0.3} \right)^2 + \left(\frac{8.2 - 7.55}{0.3} \right)^2 + \left(\frac{6.7 - 6.27}{0.3} \right)^2 + \left(\frac{4.2 - 4.72}{0.3} \right)^2 \\ &= 0.24 + 1.78 + 4.64 + 2.02 + 3.05 = 11.75 \end{aligned}$$

Spesso e' conveniente organizzare i dati in una tabella come questa:

x_i	y_i	s_y	y_i^*	$\chi_i^2 = \left(\frac{y_i - y_i^*}{s_y} \right)^2$
2.2	10.1	0.3	10.25	0.25
4.2	8.5	0.3	8.9	1.78
6.2	8.2	0.3	7.55	4.69
8.1	6.7	0.3	6.27	2.05
10.4	4.2	0.3	4.72	3.06
				$\chi_m^2 = 11.75$

Procediamo ora a valutare il test. Anzitutto dobbiamo assegnare il numero di gradi di liberta'. Essi sono pari al numero di misure indipendenti, ovvero al numero di misure meno il numero di vincoli, usati per determinare l'ipotesi di riferimento. In questo caso, abbiamo usato i dati due volte: una volta per calcolare l'intercetta e poi la pendenza. Allora $NDOF = 5 - 2 = 3$. La PDF χ^2 di riferimento avra' 3 gradi di liberta'. Dopodiche', avendo a disposizione un calcolatore, possiamo associare il p -value. Esso vale $p = 0.0082$. Cio' ci dice che, sul lungo termine, solo lo 0.8% delle volte troveremo un $\chi^2 > 11.75$. Possiamo fermarci qua ma ci e' stato chiesto di eseguire il test con un livello di confidenza del 99%. Allora per $\alpha = 0.01$ e per $NDOF = 3$ otteniamo (vedi ??) il valore critico $\chi_0^2 = 11.345$. Essendo $\chi_m^2 > \chi_0^2$ possiamo affermare

Assumendo come ipotesi \mathcal{H}_0 che i dati siano disposti linearmente, escludiamo con un livello di confidenza del 99% che i dati supportino questa ipotesi.

La conclusione non e' molto confortante. Come mai avviene questo nonostante i dati appaiano allineati e ben correlati? A ben guardare, la linea attraversa abbastanza bene tutti i punti tranne il terzo. In effetti, per questo punto il valore del $\chi^2 = 4.69$ e' piuttosto alto. Ricordando la definizione di χ^2 e ricordando che la sua radice e' numericamente pari al rapporto di compatibilita', un po' abusando della notazione, possiamo dire che questo punto stia a $\sqrt{4.69} = 2.1\sigma$ dalla retta. Possiamo quindi ragionare sostenendo che sia proprio questo punto 'colpevole' di rendere alto il χ_m^2 . Quale e' la causa? Potrebbe essere una misura poco accurata a causa di un errore sistematico. Potrebbe essere una misura affetta da una fluttuazioni statistica elevata ma pur sempre plausibile. Potremmo altresì aver sottostimato le incertezze. Oppure potrebbe essere una deviazione dovuta ad un fenomeno fisico che non avevamo considerato (si tratta di una scoperta!). In una situazione reale, dovremmo procedere ad un controllo ulteriore sui dati. Qui di seguito si discutono possibili linee di intervento:

Sottostima delle incertezze Se veramente fossimo convinti di aver sottostimato le incertezze, e a seguito di una valutazione successiva trovassimo che $s_y = 0.5$, procedendo in maniera analoga a sopra otterremo questa volta $\chi_m^2 = 4.2$. In questo caso $p = 0.2407$ e quindi torneremo ad avere una valutazione positiva del test lineare. Tuttavia, chiaramente questo caso avviene solo se in primo luogo si era fatto un errore di sottostima delle incertezze.

Escludiamo il punto sospetto Una procedura possibile sarebbe quella di escludere il punto, e ripartire da capo calcolando la miglior retta interpolante, e poi il χ_m^2 . In questo caso, otterremo $\chi_m^2 = 6$. Attenzione però che ora il numero di gradi di libertà è cambiato: $NDOF = 4 - 2 = 2$. Il p -value diventa ora $p = 0.049$, decisamente migliore rispetto al caso originale.

Proviamo una curva diversa Invece che escludere il terzo punto, avremo potuto procedere sostenendo che l'ipotesi lineare è falsa, e avremo potuto prendere una diversa per vera, ad esempio, che i dati seguano un andamento parabolico del tipo $y = a + bx + cx^2$. In questo caso, la parabola ottenuti con il metodo della massima verosimiglianza sarebbe stata $y = 10.3 - 0.095x - 0.0459x^2$ e i valori di riferimento rispettivamente 9.9, 9.1, 7.9, 6.5, 4.3. In questo caso, i diversi termini del χ^2 sarebbero stati 0.6, 3.9, 0.7, 0.4, 0, 2 per un $\chi_m^2 = 5.8$. In questo caso il numero di vincoli sarebbe stato 3 in quanto una parabola è determinata da 3 parametri, ovvero $NDOF = 5 - 3 = 2$. Allora $\chi_m^2 = 5.8$ corrisponde a $p = 0.055$. Di nuovo, piuttosto accettabile, ma decisamente simile al risultato lineare in cui abbiamo tolto il dato anomalo. Quale sarà il caso allora?

In conclusione, mentre il test viene svolto in maniera meccanica, la successiva interpretazione del test è tutt'altro che meccanica e puramente soggettiva. Si noti che spesso in letteratura scientifica il successo o il fallimento del test viene associato alla veridicità o falsità della ipotesi. Ciò non è assolutamente il caso! Il test, come abbiamo visto, si basa sulla *assunzione di veridicità della ipotesi nulla* e valuta unicamente il grado di accordo tra i dati e questa ipotesi.

10.3.3 Compatibilità tra due stime

Vediamo ora come attraverso il test del χ^2 possiamo capire il test di compatibilità visto discusso in precedenza. Si supponga infatti di voler valutare la compatibilità tra una misura $x \pm s_x$ e un valore di riferimento x^* . Impostiamo il test. Abbiamo un unico elemento quindi:

$$r^2 = \chi_m^2 = \left(\frac{x - x^*}{s_x} \right)^2 \quad (10.6)$$

Numericamente è pari al quadrato dell'indice di compatibilità per come lo avevamo definito:

La PDF di riferimento è un $\chi_2(k = 1)$. È facile calcolare queste probabilità:

- la probabilità che $r = \chi_m^2 < 1$ è 0.683 (p -value = $1 - 0.683$)
- la probabilità che $\chi_m^2 < 4$ e quindi $r < 2$ è 0.9545.
- la probabilità che $\chi_m^2 < 9$ e quindi $r < 3$ è 0.9973.

Ecco allora l'interpretazione che cercavamo:

- $0 < r < 1$ compatibilità ottima: dato il valore di riferimento x^* e l'incertezza del dato s_x allora il 68% delle volte la fluttuazioni è tale da dare $r < 1$, ovvero il 32% delle volte $r > 1$
- $1 < r < 2$ compatibilità buona: come sopra il 95% delle volte la fluttuazioni è tale da dare $r < 2$, ovvero solo il 5% delle volte $r > 2$
- $2 < r < 3$ compatibilità discreta: come sopra il 99.7% delle volte la fluttuazioni è tale da dare $r < 3$ ovvero solo lo 0.3% delle volte $r > 3$
- $r > 3$ sospetta incompatibilità. Meno dello 0.3% delle volte le fluttuazioni danno valori di $r > 3$

I casi di compatibilità tra una misura e un valore di riferimento (noto senza incertezza) e tra due misure con incertezza:

$$r = \frac{x - x^*}{s_x} \quad r = \frac{x_1 - x_2}{\sqrt{s_1^2 + s_2^2}}$$

basta considerare il secondo caso come il caso di compatibilità tra $d = x_1 - x_2$ e 0, considerando $s_d = \sqrt{s_1^2 + s_2^2}$.

10.3.4 Un valore di riferimento comune

Si considerino le seguenti misure (si trattava di misure di massa del quark z in GeV/c^2):

$$x_1 = 91.161 \pm 0.013; \quad x_2 = 91.174 \pm 0.011; \quad x_3 = 91.186 \pm 0.013; \quad x_4 = 91.188 \pm 0.013$$

Con che margine di confidenza questi dati supportano l'ipotesi che misure di una stessa grandezza?

Secondo l'ipotesi, le misure sono della stessa grandezza. Esse sono compatibili, quindi l'ipotesi è plausibile. Allora come stima di questa grandezza, possiamo prendere la loro media pesata, ottenendo

$$x_{mp} = 91.177 \pm 0.006 \text{ GeV}/c^2$$

. Ora dobbiamo valutare la bontà di questa ipotesi. Per farlo, scriviamo il test del χ^2 :

$$\begin{aligned} \chi_m^2 &= \sum_{i=1}^n \left(\frac{x_i - x_{mp}}{s_{x_i}} \right)^2 \\ &= \left(\frac{91.161 - 91.177}{0.013} \right)^2 + \left(\frac{91.174 - 91.177}{0.011} \right)^2 + \left(\frac{91.186 - 91.177}{0.013} \right)^2 + \left(\frac{91.188 - 91.177}{0.013} \right)^2 \\ &= 1.51 + 0.07 + 0.48 + 0.72 = 2.78 \end{aligned}$$

Per valutare χ_m^2 bisogna definire il NDOF. Si hanno 4 stime ma il valore di riferimento è stato calcolato a partire dalle misure stesse, portando quindi a 1 vincolo, e quindi $NDOF = 4 - 1 = 3$. Per 3 gradi di libertà, il valore $\chi_m^2 = 2.78$ corrisponde ad un livello di significanza (p -value) di 0.4268. Ovvero, ben il 43% delle volte avrei potuto ottenere un $\chi^2 > \chi_m^2$. Se non ho a disposizione un calcolatore, posso usare [Figura 10.3](#). Per $NDOF = 3$ e per un livello di significanza anche del 10% ho un valore di riferimento $\chi_0^2 = 6.251$. Essendo $\chi_m^2 < \chi_0^2$ allora il test è passato già con un livello di significanza. Quello che posso affermare quindi è che l'ipotesi di compatibilità dei dati è supportata dai dati.

Si noti come nel calcolo del χ_m^2 al denominatore si sono messe le incertezze misurate delle singole stime. Ci si può chiedere come mai non abbiamo considerato l'incertezza della media pesata $0.006 \text{ GeV}/c^2$. La ragione sta nella natura della ipotesi. L'ipotesi era che ci fosse un valore costante di riferimento per le diverse misure.

10.4 Limite pratico del test del χ^2

La PDF del χ^2 è come visto costruita come sommatoria di quadrati di Gaussiane standard:

$$\chi^2 = \sum_{i=1}^k \mathcal{G}^2(0, 1) = \sum_{i=1}^k \left(\frac{X_i - \mu}{\sigma} \right)^2$$

tuttavia, abbiamo a che fare nel caso reale solo con campioni di misure:

$$\chi_m^2 = \sum_{i=1}^k \left(\frac{x_i - x^*}{s_{x_i}} \right)^2$$

Il metodo quindi funziona sempre che s_{x_i} sia una buona stima di una SE Gaussiana. Come procedere nel caso cioè non sia propriamente garantito, ma sia valido solo in prima approssimazione? Vedremo in [Capitolo 11](#) come utilizzare una versione approssimata del test del χ^2 in caso le incertezze non siano Gaussiane e come risolvere il problema nel caso correggendo la stima della incertezza per un fattore correttivo.

10.5 Considerazioni finali

Quesiti

Esercizio 10.5.1

Frazione atomica di rame (F)	Resistenza (Ω)
0.00	2.30
0.25	2.18
0.50	2.10
0.75	1.88
1.00	1.75

Si considerino i seguenti dati sperimentali che quantificano la resistenza elettrica di vari campioni di lega metallica rame-oro aventi la stessa forma ma diversa concentrazione di rame. Consideriamo che la frazione atomica abbia una incertezza trascurabile mentre la resistenza abbia una incertezza normale con deviazione standard di 0.02Ω . Si ipotizza che la resistenza cambi linearmente con la frazione di rame. a) Calcolare i parametri della retta interpolante, assieme alle loro incertezze, menzionando con quale metodo statistico sono calcolati, e riportando i valori numerici dei termini intermedi necessari alle stime, secondo le formule [Equazione 8.5](#). b) Effettuare il test del χ^2 per l'ipotesi di linearità, con un livello di confidenza del 99% facendo riferimento alla tabella. Riportare il valore del χ_i^2 per i singoli punti e il valore totale oltre al numero di gradi di libertà. Discutere se ci sono dati particolarmente anomali.

□ **Es. 10.1** A seguito di due serie di misure X, Y come in tabella, si vuole cercare se sussiste una relazione lineare tra le due serie. Considerando che a) tutte le misure sono distribuite secondo la statistica normale e sono indipendenti, b) le incertezze sulle misure di X si possono considerare trascurabili, le incertezze sulle misure Y sono tutte uguali e pari a $\sigma_y = \sigma_{y_i} = 0.3$, calcolare:

1.1 I coefficienti della retta interpolante e le loro incertezze.

$$\begin{aligned}
 a &= \frac{\sum_i x_i^2 \cdot \sum_i y_i - \sum_i x_i \cdot \sum_i x_i y_i}{\Delta} & \sigma_a &= \sigma_y \sqrt{\frac{\sum_i x_i^2}{\Delta}} \\
 b &= \frac{N \cdot \sum_i x_i y_i - \sum_i x_i \cdot \sum_i y_i}{\Delta} & \sigma_b &= \sigma_y \sqrt{\frac{N}{\Delta}} \quad \Delta = N \sum_i x_i^2 - (\sum_i x_i)^2
 \end{aligned}
 \tag{10.7}$$

1.2 La covarianza e il coefficiente di correlazione di Bravais-Pearson tra le due serie

1.3 Effettuare il test del χ^2 con un livello di significanza del 5% e discuterlo usando i dati tabulati.

E' possibile (ma non necessario) usare le righe e colonne vuote della tabella.

X	Y						
2.2	10.1						
1-8 4.2	8.5						
1-8 6.2	8.2						
1-8 8.1	6.7						
1-8 10.4	4.2						
1-8							

Approfondimenti

Appr. 10.1 Standard score (Z-score). Spesso in letteratura il p -value e' riportato in una maniera alternativa usando lo *standard score* o *Z-score*. Il conto si svolge considerando che per sufficienti gradi di liberta', la PDF del χ^2 come visto tende ad una distribuzione Gaussiana. In questa ipotesi, il p -value puo' essere riportato *in unita' di σ* (dove *sigma* e' il parametro di ampiezza della Gaussiana). Si osservi ?? per comprendere la affermazione. Il valore in σ -corrispondenti del p -value e' detto *Z-score* o *standard score*. Per il calcolo, di nuovo conviene usare librerie standard per i diversi codici. Io ho trovato anche questo utile calcolatore online <http://courses.atlas.illinois.edu/spring2016/STAT/STAT200/pnormal.html>. Ad esempio, e' facile vedere come il valore ufficialmente riconosciuto come limite per le scoperte in fisica ovvero 5σ , corrisponda ad un $p = 2.867e - 7$. Cio' sta a dire che ripetendo infinitamente volte l'esperimento, solo una volta su $1/p = 3487966$ avremo ottenuto un χ^2 maggiore del misurato.

Appr. 10.2 Errore di tipo I - Falso positivo. Il p -value ha anche una diversa interpretazione, oltre a quella vista. Esso rappresenta la probabilita' che un esperimento *positivo*, ovvero avvenuto nel

caso che l'ipotesi nulla fosse realmente vera, sia dichiarato *falso*. Questo errore e' detto *falso positivo* nel senso che misura il rischio di affermare i dati non supportano l'ipotesi quando in realtà essa era vera. Attenzione quindi che la parola *positivo* si riferisce alla veridicità della ipotesi nulla, non a qualche giudizio positivo. Ad esempio l'ipotesi potrebbe essere *in questi dati non c'e' un segnale*. Si presti attenzione al fatto che questo puo' generare confusione talvolta. Piu' correttamente questi errori sono detti di Tipo I.

Appr. 10.3 Il χ^2 ridotto. Come abbiamo visto il significato del valore χ_m^2 dipende dal numero di gradi di libertà k e il suo valore di aspettazione e' proprio pari a k . Allora spesso si riporta il valore del χ^2 ridotto ovvero diviso per k :

$$\chi_R^2 = \frac{\chi^2}{k} \quad (10.8)$$

E' chiaro quindi che il valore di aspettazione del χ_R^2 e' pari a 1. A volte la notazione usata e' $\chi_R^2 = \chi^2/NDOF$.

Capitolo 11

Test χ^2 Pearson e il Test di Student

Nel [Capitolo 10](#) abbiamo visto come il margine di confidenza sulla valutazione della bontá dell'ipotesi (nel modo visto $P(D|H)$ e non $P(H|D)$) fosse valutato sfruttando il fatto che il χ^2 fosse costruito come sommatoria di quadrati di Gaussiane normalizzate:

$$f_\chi = \lim_{N \rightarrow \infty} \sum_{i=1}^N \mathcal{G}^2(0.1) = \sum_{i=1}^N \left(\frac{X_i - \mu}{\sigma} \right)^2$$

Questo funziona quindi se X_i sono grandezze distribuite normalmente, ma anche se σ é una deviazione standard di tipo Gaussiano. Ora chiaramente questo non avviene sempre e di norma. Come ci si puo' comportare? Spesso, come visto oramai diverse volte in precedenza, l'ipotesi Gaussiana e' all'incirca verificata, e quindi il metodo del χ^2 si puo' usare anche in maniera approssimata, purché questo sia menzionato esplicitamente. In particolare, sempre che si abbia a che fare con conteggi, e quindi con statistiche di tipo Poissoniano o Binomiale, si puo' usare un metodo approssimato detto χ^2 di Pearson, che viene descritto in [sezione 11.1](#). Si noti che ogni volta i dati sono raggruppati in classi di frequenza, a quel punto si ha a che fare con conteggi nella classe di frequenza piuttosto che di singole misure, e quindi il metodo approssimato e' da usare. In un secondo caso invece in cui si cerca la compatibilitá tra due valori, e' possibile usare un metodo esatto che tenga in conto della discrepanza rispetto al caso Gaussiano. Il test e' detto *test di Student* o *t-test* e viene descritto in [sezione 11.2](#).

11.1 Test del χ^2 di Pearson

Si supponga di avere una qualsiasi raccolta di misure che sia organizzata in classi di frequenza, o una raccolta di conteggi. Sono esempi di questo: istogramma di misure ripetute, conteggi per categorie come costo per metro quadro di appartamenti residenziale per regione etc. Quanto si tratta di conteggi, possiamo sempre assegnare a questi conteggi una variabile aleatoria di tipo Binomiale. Si prenda infatti una raccolta di N dati, si puo' pensare ad un evento di tipo Bernoulliano, la probabilitá che un evento cada oppure no nel bin i -esimo. Allora, il conteggio n del bin puo' essere pensato come alla frequenza relativa (e quindi alla probabilitá) che n eventi cadano in quel bin, e quindi il fenomeno e' di tipo Binomiale:

$$f_i = \frac{n_i}{N} \rightarrow p_i = \mathcal{B}(n_i, N)$$

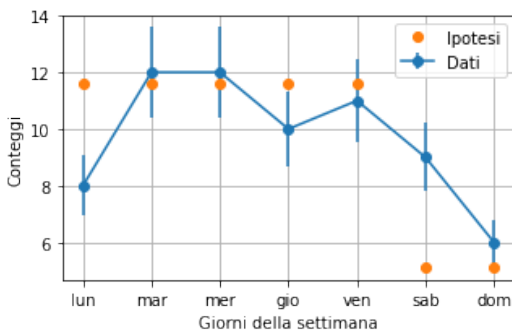
Si supponga ora di fare una ipotesi *sulla distribuzione di dati*. L'ipotesi ha come variati le classi di frequenza e come distribuzione di probabilitá quella che definisce l'ipotesi, ad esempio, che la distribuzione sia uniforme, o gaussiana, esponenziale, etc. Allora in questo caso possiamo *definire* le probabilitá di riferimento per ogni classe di frequenza:

$$p_i^* = p_i(H)$$

Procediamo con un esempio per semplificare la discussione. Si supponga di avere una tabella di misure del numero di nascite di bambini all'ospedale di Padova secondo i giorni della settimana effettuate su un'unica settimana. A fianco delle misure costruiamo le corrispondenti frequenze $f_i = p_i$. Formuliamo

l'ipotesi (che non ha nessun riscontro reale) che nel finesettimana si facciano nascere meno bambini di proposito, mentre negli altri giorni lavorativi cio' non avviene. Raffiniamo l'ipotesi calcolando le p_i^* giorno per giorno. Sulla base di questa ipotesi possiamo costruire i conteggi di riferimento n_i^* . Costruiamo una tabella come segue e grafichiamo dati e ipotesi.

Day	Dati		Ipotesi	
	n_i	$f_i = p_i$	p_i^*	n_i^*
Lunedì	8	0.118	0.17	11.6
Martedì	12	0.176	0.17	11.6
Mercoledì	12	0.176	0.17	11.6
Giovedì	10	0.147	0.17	11.6
Venerdì	11	0.162	0.17	11.6
Sabato	9	0.132	0.075	5.1
Domenica	6	0.088	0.075	5.1
Somma	68	1	1	68



Si noti come la sommatoria dei conteggi misurati e di riferimento sia la stessa e così per le loro frequenze. Che nome diamo all'ipotesi? Non si tratta di una distribuzione standard di quelle che abbiamo visto in [Capitolo 5](#), ma possiamo facilmente costruirla:

$$\mathcal{H}_0 \rightarrow f_X(x) = \begin{cases} 0.17 & \text{se } x \in [1, 2, 3, 4, 5] \\ 0.075 & \text{se } x \in [6, 7] \end{cases}$$

dove abbiamo mappato i giorni della settimana ai variati $x = 1 \dots 7$. Procediamo ora a valutare la bontà dell'ipotesi \mathcal{H}_0 .

Fermatevi a pensare come impostereste questo test e poi andate avanti nella lettura.

Intuiamo che dobbiamo confrontare uno a uno ma anche nel loro insieme le compatibilità tra n_i, n_i^* o tra le loro frequenze p_i, p_i^* , probabilmente non c'è differenza nel risultato. Sappiamo anche che la compatibilità tra due valori è un concetto che si intende se si pesa la distanza con l'incertezza. Passiamo quindi a discutere il caso: che incertezza hanno i valori della tabella? Il fatto di aver raggruppato le misure in classi di frequenza¹ come abbiamo visto ci permette di passare ai conteggi e di definire su questi una variabile aleatoria. Se p_i^* è la probabilità della classe di frequenza i –esima significa che ogni evento degli $N = 68$ ha probabilità p_i^* di cadere in quella classe. Invece la probabilità che n_i eventi capitino in quella classe è data dalla PMF binomiale così costruita:

$$\binom{N}{n_i} p_i^{n_i} (1 - p_i)^{N - n_i}$$

il cui valore medio (di aspettazione) è $E = N p_i$ e la cui varianza è $E = N p_i(1 - p_i)$. Con questa considerazione possiamo quindi definire l'incertezza attraverso la deviazione standard della binomiale sopra costruita. Allora per ogni misura n_i, p_i possiamo calcolare $s_i = \sqrt{N p_i(1 - p_i)}$. Per quanto riguarda l'ipotesi, assumiamo invece che essendo presa per vera per costruzione, non abbia incertezze. Con questi elementi possiamo costruire le compatibilità 'bin per bin' secondo [Equazione 10.6](#):

$$\lambda_i = \frac{|n_i - n_i^*|}{s_i}$$

¹In realtà i bambini nascono a diverse ore del giorno, e avremmo potuto raggruppare i dati (se li avessimo) in maniera diversa considerando l'inizio e la fine dei turni di lavoro, o aumentando le classi di frequenza, etc.

e da questa la sommatoria:

$$\chi_P^2 = \sum_{i=1}^M \left(\frac{n_i - n_i^*}{s_i} \right)^2 = \sum_{i=1}^M \left(\frac{n_i - n_i^*}{\sqrt{Np_i(1-p_i)}} \right)^2 = \sum_{i=1}^M \frac{(n_i - n_i^*)^2}{Np_i(1-p_i)} \quad \text{Binomiale}$$

Ora, se \mathcal{H}_0 fosse vera e s_i fossero tutti gaussiani (deviazione standard della gaussiana), la formula sopra corrisponderebbe al computo del χ^2 . Invece mentre e' vero che se l'ipotesi nulla e' vera il numeratore tende a zero, abbiamo visto che s_i sono binomiali. In questo senso il valore χ_P^2 non e' propriamente un χ_m^2 . Tuttavia come abbiamo visto in Appr. 5.2, sotto alcune condizioni la binomiale (se N grande e $p_i \ll 1$ tende ad una poissoniana e questa (se λ grande) ad una gaussiana. Sappiamo che nel primo dei due casi si puo' approssimare la deviazione standard binomiale a quella poissoniana: $\sqrt{Np_i(1-p_i)} \sim \sqrt{Np_i} = \sqrt{n_i}$. In questo caso

$$\chi_P^2 = \sum_{i=1}^M \left(\frac{n_i - n_i^*}{s_i} \right)^2 = \sum_{i=1}^M \left(\frac{n_i - n_i^*}{\sqrt{n_i}} \right)^2 = \sum_{i=1}^M \frac{(n_i - n_i^*)^2}{n_i} \quad \text{Poisson}$$

In questo senso possiamo pensare che χ_P^2 sia una approssimazione di χ_m^2 . Vediamo numericamente con il nostro esempio:

Dati			Ipotesi		Test Ipotesi			
Day	n_i	$f_i = p_i$	p_i^*	n_i^*	Binomiale		Poisson	
					$\sqrt{Np_i(1-p_i)}$	$\tilde{\chi}_i^2$	$\sqrt{n_i}$	$\tilde{\chi}_i^2$
Lunedì	8	0.118	0.17	11.6	2.7	1.80	2.8	1.58
Martedì	12	0.176	0.17	11.6	3.1	0.02	3.5	0.02
Mercoledì	12	0.176	0.17	11.6	3.1	0.02	3.5	0.02
Giovedì	10	0.147	0.17	11.6	2.9	0.29	3.2	0.24
Venerdì	11	0.162	0.17	11.6	3.0	0.03	3.3	0.03
Sabato	9	0.132	0.075	5.1	2.8	1.95	3.0	1.69
Domenica	6	0.088	0.075	5.1	2.3	0.15	2.4	0.14
Somma	68	1	1	68		4.25		3.71

Dalla tabella osserviamo che in effetti nonostante p_i non siano molto inferiori ad 1, la differenza tra la deviazione standard Poissoniana e quella binomiale non e' cosi' elevata nella maggior parte dei casi, ma cio' comporta una differenza nel $\tilde{\chi}_i^2$ di circa mezza unità.

Come giudichiamo χ_P^2 ? Anche se in prima approssimazione segue la distribuzione del χ^2 , dobbiamo stabilire i gradi di libertà. Come gia' visto, questi si decidono sulla base dei vincoli utilizzati per il calcolo del valore di riferimento p_i^* . In questo caso, per formulare l'ipotesi abbiamo solo utilizzato un dato sperimentale: il fatto che ci fossero N misure in totali. Allora $k = M - 1 = 6$. Il χ^2 a 6 gradi di libertà ha un valore di aspettazione pari a 6. Il χ_P^2 in entrambi i casi e' inferiore a questo valore, e nemmeno di molto, quindi in entrambi i casi concludiamo che i' assunta vera \mathcal{H}_0 i dati sono compatibili con questa ipotesi². Questo test prende il nome di χ^2 di Pearson.

11.1.1 Validità e prassi

Come abbiamo visto, affinché il riferimento alla variabile del χ^2 sia valido e' necessario che la deviazione standard binomiale non si discosti troppo dalla deviazione standard poissoniana e quindi gaussiana. Una prassi che si usa e' quella di fare attenzione alle classi di frequenza con meno di 5 conteggi. Il numero deriva dal fatto che l'approssimazione e' buona quando $\sigma_B^2 = \sqrt{Np_i(1-p_i)} \sim \sqrt{Np_i} \geq 9$ e quindi $Np_i = n_i \geq 3$ elevato a 5 per una minore differenza. Se il conteggi in una classe di frequenza e' inferiore a 5 e' prassi:

²Si noti che in questo caso la presunta compatibilità potrebbe essere legata alla grandezza delle incertezze. Per ridurla dovremo disporre di un numero maggiore di conteggi, ad esempio prendendo piu' conteggi di nascite. In particolare, se i conteggi n_i fossero stati medie di molte serie ripetute di misure, si sarebbe potuto associare incertezza normale e tornare al caso del χ^2 esatto.

- Raggruppare il conteggio di quella classe con una o piu' classi di frequenza vicine fino a superare il valore cumulativo di 5. Il rischio in questo modo e' che la classe di frequenza sia piu' ampia delle altre e quindi di per se meno informativa. Inoltre questa operazione non solo ha lo svantaggio di creare un bias (l'eliminazione di informazioni lo crea sempre) ma anche si riduce il numero di gradi di liberta' per il test gia' approssimato del χ^2 .
- Rimuovere interamente la classe di frequenza. Anche in questo caso si riduce il numero di gradi di liberta'.

11.1.2 Un esempio illustrativo: verifica distribuzione normale di un'istogramma

Si prenda una distribuzione di dati raggruppate in 8 classi di frequenza x_i [m], ciascuna con conteggi (frequenze assolute) n_i . Si valuti la natura gaussiana della distribuzione.

Riportiamo in una tabella, per ogni classe di frequenza x_i oltre ai conteggi n_i anche:

- Le frequenze relative $p_i = n_i/N$, dove $N = 62$ e' la numerosita' del campione
- Le frequenze di riferimento sono $p_i^* = \frac{1}{\sqrt{2\pi}s_x} e^{-\frac{1}{2}\left(\frac{x_i-\bar{x}}{s_x}\right)^2}$ e i conteggi di riferimento sono $p_i^* = \frac{Nw}{\sqrt{2\pi}s_x} e^{-\frac{1}{2}\left(\frac{x_i-\bar{x}}{s_x}\right)^2}$ dove $\bar{x} = 64.24$ m e' la media, $s_x = 2.65$ m e' la deviazione standard, $w = 2$ m la larghezza della classe di frequenza
- I conteggi di riferimento $n_i^* = Np_i^*$
- L'incertezza binomiale dei conteggi per bin $s_i = Np_i(1 - p_i)$ m
- I termini $\chi_i^2 = \left(\frac{n_i - n_i^*}{s_i}\right)^2$

x_i [m]	n_i	n_i^*	p_i	p_i^*	s_i [m]	χ_i^2
58	1	1.1	0.02	0.02	1.0	0.0
60	5	5.1	0.08	0.08	2.1	0.0
62	13	13.0	0.21	0.21	3.2	0.0
64	21	18.6	0.34	0.30	3.7	0.4
66	13	15.0	0.21	0.24	3.2	0.4
68	7	6.9	0.11	0.11	2.5	0.0
70	1	1.8	0.02	0.03	1.0	0.6
72	1	0.3	0.02	0.00	1.0	0.6
Sum	62	61.8	1.000	0.997		2.03

In un primo calcolo abbiamo tenuto tutte le classi di frequenza, anche quelle a numerosita' inferiore a 5. Si e' ottenuto $\chi_P^2 = 2.03$. La PDF del χ^2 di riferimento ha $k = M - 3 = 5$ gradi di liberta' infatti per calcolare n_i abbiamo usato 3 parametri dalle misure: N, \bar{x}, s_x . Il p -value = 0.84 quindi circa l'84% delle volte si avrebbe un $\chi^2 > 2.03$. Il $\chi_R^2 = 0.41$. Il test e' passato e possiamo affermare con un buon livello di significanza che la distribuzione di dati e' gaussiana.

Facciamo ulteriori test:

- Se avessimo escluso i bin con meno di 5 dati ovvero il primo e gli ultimi due, avremmo ottenuto $\chi^2 = 1.4$ per $k = 5 - 3 = 2$ gradi di liberta' corrispondente ad un p -value = 0.49, ancora piu' centrale. Il $\chi_R^2 = 0.7$. Il test sarebbe quindi ancora passato.
- Sommare i bin adiacenti per raggiungere la quota di 5 conteggi rende talvolta piu' complesso il calcolo del valore di riferimento. In questo caso avremmo dovuto calcolare n_i^* considerando di volta in volta la corretta larghezza del bin. Si lascia per esercizio.
- Infine, avremmo potuto anche non usare il raggruppamento dei bin ma considerare tutte le misure una per una. A questo punto avremmo dovuto calcolare

$$\chi_P^2 = \sum_{i=1}^M \left(\frac{x_i - \bar{x}}{s_x} \right)^2 = 24.6$$

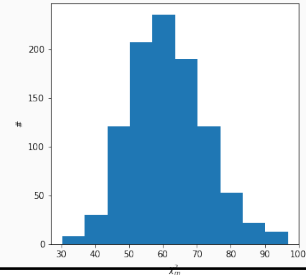
da confrontare ora con $k = N - 1 = 61$ gradi di liberta', in quanto un solo vincolo e' stato usato per definire il valore di riferimento (la media aritmetica). Il p -value = .99 in questo caso ma

il $\chi_R^2 = 0.40$ come nel primo caso. Si noti come in questo caso si abbiano maggiori informazioni che nel primo a dispetto di un calcolo con maggiori termini.

Esempio:

Scrivere un programma che per un elevato numero di volte estrae casualmente da una distribuzione Gaussiana come nell'esempio sopra N valori e ne calcola il χ^2 . Cosa vi aspettate?

R: essendo i valori estratti casualmente da una gaussiana, ci si aspetta che i dati siano distribuiti gaussianamente ovvero che $\chi_m^2 \sim k$ dove k è il numero di gradi di libertà. Se la media è stata impostata non ci sono vincoli e $k = N = 62$. In un test che ho fatto con $N = 1000$ ripetizioni ho ottenuto in media $\overline{\chi_m^2} = 62.19$ e $\overline{\chi_m^2} = 1.003$. Graficando la distribuzione dei χ_m^2 ho ottenuto questo istogramma. Che distribuzione di riferimento si può associare a questo istogramma?



11.2 Test di Student

Il test di Student è un test di verifica di una certa classe di ipotesi, ovvero di compatibilità tra due valori distribuiti normalmente, che si esegue nel caso in cui la stima delle incertezze sia elevata, o a causa di scarso campionamento (bassa numerosità) o per altre ragioni. Il test è anche detto **t-test** perché si basa sulla definizione di un variato t calcolato a partire dai dati. Il test è detto di Student perché Student è lo pseudonimo che usò W. S. Gosset per pubblicare per la prima volta questo metodo (si veda Appr. 11.4 per le ragioni di questa scelta). La variabile t prende origine da un rapporto che abbiamo incontrato già diverse volte: nel caso gaussiano, se si vuole confrontare ad esempio la compatibilità tra un valore x, σ_x (con la sua incertezza) ed un valore di riferimento x^* si studia il rapporto:

$$r = \frac{|x - x^*|}{\sigma_x}$$

Come abbiamo visto e discusso in 10.6 possiamo calcolare la distribuzione di probabilità del variato r^2 nel caso σ_x sia la deviazione standard gaussiana: essa è la PDF del χ^2 con un numero di gradi di libertà pari a 1. Nella condizione appunto in cui σ_x sia la deviazione standard gaussiana allora r è, come visto, un variato di una gaussiana standard $\mathcal{G}(0, 1)$. Tanto più il valore σ_x si discosta dalla deviazione standard gaussiana tanto più l'uso del test del χ^2 per questa compatibilità dà risultati erranei. Gosset si interrogò su come risolvere questo problema e adattare il test al caso in cui si abbia poca conoscenza su σ_x , ad esempio nel caso in cui si sia stimata da un campione di scarsa numerosità o con possibili effetti sistematici. Sappiamo infatti che la deviazione standard campionaria s_x è una stima buona di σ_x solo al limite statistico:

$$\lim_{N \rightarrow \infty} s_x = \lim_{N \rightarrow \infty} \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2} = \sigma_x = \sqrt{\text{VAR}[X]}$$

Un'altra maniera di vedere lo stesso concetto è ricordare che al parametro s_x abbiamo assegnato la incertezza

$$s_{s_x} \sim \frac{s_x}{\sqrt{2(N-1)}}$$

che è piccola solo per grandi numerosità del campione. Gosset osservò che la variabile, che ora chiamiamo t

$$t = \frac{|x - x^*|}{s_x}$$

risulta distribuita non come $\mathcal{G}(0, 1)$ bensì come

$$f_{\mathcal{T}}(t; k) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi} \Gamma(\frac{k}{2})} \left(1 + \frac{t^2}{k}\right)^{-\frac{k+1}{2}}.$$

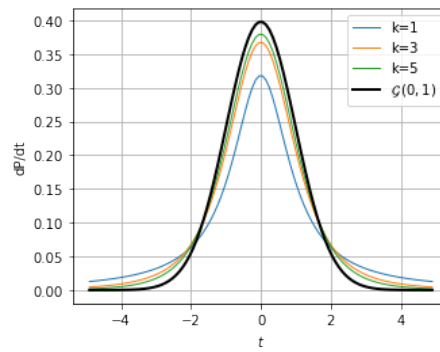


Figura 11.1: La PDF di Student per alcuni grandi di libertà

(11.1)

La funzione é riportata in [Figura 11.1](#). Si può notare come per N che aumenta la PDF assomiglia sempre di più a $\mathcal{G}(0, 1)$ e come per valori bassi di N la distribuzione é maggiormente curtotica (code piu' alte) rispetto alla $\mathcal{G}(0, 1)$. Le proprietà principali di \mathcal{T} sono quindi:

- Media nulla: $\mu = E(\mathcal{T}) = 0$
- Varianza: $\sigma = \text{VAR}(\mathcal{T}) = \frac{k}{k-2}$. Si noti che $\sigma \rightarrow 1$ per $k \rightarrow \infty$
- Al limite di elevati gradi di libertà: $\lim_{k \rightarrow \infty} \mathcal{T} = \mathcal{G}(0, 1)$

11.2.1 Costruzione del test t o di Student

Siccome conosciamo la PDF di t , allora, in maniera del tutto analoga al test del χ^2 , procediamo con la costruzione del test di Student:

1. **[Formulazione ipotesi nulla]** Si formula anzitutto l'ipotesi sui dati, che nel caso del test di Student sono ipotesi di compatibilità (tra due numeri oppure tra una misura e un valore di riferimento). Questa ipotesi può essere chiamata **ipotesi nulla** e indicata con \mathcal{H}_0 , ma vedremo che il test t ha delle differenze rispetto al test χ^2 .
2. **[Calcolo valori riferimento e decisione sulle code del test]** L'ipotesi nulla ci può fornire il valore di riferimento x^* . A questo punto possiamo andare oltre rispetto alla semplice compatibilità in quanto t può assumere valori positivi o negativi a seconda che $x > x^*$ o viceversa. Ci possiamo quindi chiedere:
 - **Test a 1 coda** Il test si dice *ad una coda* se l'ipotesi nulla é «Il valore x é inferiore a x^* », oppure se l'ipotesi é «Il valore x é superiore a x^* »
 - **Test a 2 code** Il test si dice *a due code* quanto \mathcal{H}_0 é «Il valore x é compatibile con x^* »
3. **[Esprimo t]** Il valore esatto di t dipende da caso a caso. Identifichiamo i seguenti sottocasi:
 - **Confronto di una misura con un valore di riferimento** In questo caso il valore di riferimento x_i si intende privo di incertezza. Allora se la misura per il confronto é x , s_x la variabile t si definisce come:

$$t = \frac{x - x^*}{s_x}$$

Se x é una singola misura, s_x é la sua incertezza. In questo caso $N = 1$. Di norma tuttavia il test non viene usato in questa condizione limite in quanto é decisamente poco significativo. Nel caso x sia la media di N misura allora $x = \bar{x}$, $s_x = s_{\bar{x}}$ e la distribuzione di riferimento per t é quindi una variabile di Student con $k = N - V$ dove V é il numero dei vincoli usati per quantificare s_x e quindi la discrepanza rispetto alla Gaussiana standard. In questo caso abbiamo usato una volta i dati per calcolare $s_{\bar{x}}$ come dimostrato in [Appr. 11.5](#).

- **Confronto tra due stime con incertezza.** Si supponga di avere due stime che scriviamo direttamente come medie $\bar{x} \pm s_{\bar{x}}$ e $\bar{y} \pm s_{\bar{y}}$. Supponiamo che la numerosità del primo campione sia N e del secondo M . Ci si può chiedere se \bar{x} è significativamente maggiore (o minore) di \bar{y} (test a 1 coda) o compatibile con \bar{y} (test a due code). Che forma prende la variabile t ? Il test può essere svolto pensando di costruire la variabile $d = \bar{x} - \bar{y}$ e cercarne la compatibilità con 0. Nel farlo si deve stimare s_d . Per la stima di s_d ci sono alcune considerazioni da fare. Se consideriamo \bar{x}, \bar{y} indipendenti allora possiamo calcolare, con le formule della propagazione delle incertezze:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{N} + \frac{s_y^2}{M}}} \quad (11.2)$$

In realtà, l'ipotesi che le due misure siano della stessa popolazione lega i due valori di s_x e s_y . Essi non dovrebbero differire troppo entro le loro fluttuazioni, perché; entrambi sono stime della stessa σ_x della popolazione. Per tenere in conto di questo fatto una valutazione più accurata di t è:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{N+M}{N+M-2} \left[\frac{N-1}{M} \frac{s_x^2}{N} + \frac{M-1}{N} \frac{s_y^2}{M} \right]}} \quad (11.3)$$

La dimostrazione della formula è data in Appr. 11.6. La formulazione in Equazione 11.2 è quindi una sottostima del valore più accurato dato in Equazione 11.3. Si noti che questa formulazione si basa sull'ipotesi che i due campioni siano entrambi rappresentativi della stessa popolazione. Le due varianze non possono essere troppo diverse, e si usa la regola che la compatibilità può essere valutata se $1 < s_x/s_y < 5$ ovvero se le due non differiscono di più di un fattore 5. Il numero di gradi di libertà del test è quindi $k = N + M - 2$.

4. **[Valutazione ipotesi]** La valutazione dell'ipotesi è simile concettualmente a quella del χ^2 ma con alcune differenze. In sostanza, come per il χ^2 si può calcolare il livello di significanza (o p -value). In questo caso si deve fare attenzione al fatto che t può essere positiva o negativa, tuttavia, essendo simmetrica attorno all'origine, ha senso usare $|t|$ e quindi il livello di significanza è sempre inferiore a 0.5. In alternativa si possono usare le tabelle di probabilità di t che come quelle del χ^2 presentano valori critici in funzione dei gradi di libertà e del livello di significanza. La prima differenza è che se il test è svolto a due code, va considerata anche la 'coda di sinistra' e quindi per un certo valore critico corrisponde un livello di significanza *doppio* rispetto al caso a una coda. Dopodiché:
 - Se $|t| > t_0$ allora il test non è superato con livello di significanza di t_0 (controllando se una o due code). L'interpretazione è simile al caso del χ^2 : solo α volte $|t| > t_0$ se ripetessimo l'esperimento molte volte e si trattasse solo di fluttuazioni casuali.
 - Se $|t| > t_0$ allora il test è superato con un margine di confidenza di $1 - \alpha$. Ciò non significa che l'ipotesi nulla sia vera al $1 - \alpha\%$ bensì che i dati sono compatibili con questa ipotesi entro le loro fluttuazioni.

Considerazioni aggiuntive:

- Si osservi Figura 11.2. Scendendo lungo una colonna il valore diminuisce e si stabilizza attorno ad un valore. Ad esempio per $\alpha = 0.05$ (una coda, la colonna segnata) tra 1 e 2 gradi di libertà il valore critico dimezza. Ciò si interpreta (si osservi anche Figura 11.1) pensando che per numerosità basse è probabile che ci siano valori alti di t . Inoltre all'aumentare della numerosità si ha quel comportamento asintotico che porta la t ad assomigliare alla \mathcal{G} .
- Sopra i 20-30 gradi di libertà la PDF del t varia molto lentamente. Essa come sappiamo equivale ad $\mathcal{G}(0, 1)$. In questa condizione si può non svolgere il test t e tornare a quello del χ^2 . Si osservi infatti che a partire da circa 20 per un valore critico pari a 1 si ha un livello di confidenza di $1 - 0.3 = 0.7$. Ciò corrisponde circa a 0.68, il livello di confidenza entro 1σ . Per $t = 2$ si ritrovano i 2σ infatti il livello di confidenza è del 95% in un test a due code.

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
Z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

Figura 11.2: Tabella di valori critici e livelli di significanza per la variabile di Student

- La differenza concettuale tra il test a una e due code non e' grande, e i margini di confidenza si assomigliano (il livello di significanza e' infatti sempre di norma un numero piccolo). Tuttavia in alcuni casi e' piu' corretto chiedersi se un valore non supera (o supera) un altro. Ad esempio, se testiamo un carico di rottura su un campione vogliamo che non sia inferiore a quello di riferimento del produttore, non solamente che sia compatibile. Il test a una coda e' piu' stringente.

11.2.2 Esempi applicazione test Student

Procediamo ora con alcuni esempi illustrativi.

Compatibilità di un campione con un valore di riferimento

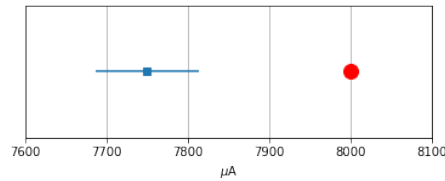
A seguito della modifica di un componente si vuole testare se la corrente di rottura dei fotomoltiplicatori prodotti da uno stabilimento e' compatibile con il valore di riferimento pe quel prodotto di 8000 . E' chiaro che non si vogliono danneggiare troppi fotomoltiplicatori quindi si esegue un primo test su un campione 6 sensori, ottenendo come media $7750 \pm 64 \mu A$. Ci si interroga se questo valore sia compatibile o significativamente inferiore al valore di riferimento. Si utilizzi un livello di significanza del 5%.

R: Essendo la numerosità del campione bassa, l'incertezza sulla deviazione standard della media $64 \mu A$ e' alta. Invece che valutare la compatibilità con il metodo del χ^2 e' piu' corretto usare il test di

Student. Costruiamo la variabile

$$t = \frac{\bar{x} - x^*}{s_{\bar{x}}} = \frac{7750 - 8000}{64} = -3.9$$

e grafichiamola per comodità:



Il numero di gradi di libertà per questo test é $k = 6 - 1 = 5$ dove il vincolo é dato dalla numerosità del campione, usato per stimare s_x . Valutiamo 3 casi:

Test a 2 code con $\alpha = 0.05$ Per $\alpha = 0.05$, se si tratta di compatibilità, utilizziamo un test a due code. Il valore critico allora é $t_0 = 2.571$. Siccome $|t| > t_0$ allora il test non é superato con un livello di confidenza del 95%: le due misure risultano incompatibili con un livello di confidenza del 95%.

Test χ^2 con $\alpha = 0.05$ Se avessimo ignorato la bassa numerosità del campione e assunto che $64 \mu A$ fosse una buona stima di deviazione standard Gaussiana avremmo potuto usare il test del χ^2 . In questo caso, per $\alpha = 0.05$ e 6 gradi di libertà (il valore di riferimento é dato) avremmo ottenuto $\chi_0^2 = 12.6$ e $\chi_m^2 = t^2 = 15.21$. Siccome $\chi_m^2 > \chi_0^2$ avremmo concluso nuovamente per incompatibilità.

Test a 2 code con $\alpha = 0.01$ Si noti che invece per un livello di significanze del 99% il test sarebbe stato superato in quanto $|t| < 4.032$ e le due misure sarebbero state compatibili però *solo* entro un intervallo di confidenza del 99% (di nuovo, ovvero solo l'1% delle volte si avrebbero avuto a causa di fluttuazioni unicamente statistiche discrepanze maggiori).

Test a 1 coda con $\alpha = 0.01$ Se invece ci fossimo chiesti se il valore misurato é significativamente minore del valore di riferimento (con un livello di significanza del 1%) avremmo effettuato un test ad una coda, trovando $t_0 = 3.365$. Questa volta essendo $|t| > t_0$ avremo escluso che il valore fosse significativamente minore e affermato che con un margine di confidenza del 99% il valore misurato é inferiore al valore di riferimento. Si presti attenzione al fatto che in questo ultimo caso l'ipotesi nulla \mathcal{H}_0 non é "Una misura é inferiore all'altra" bensí "Una misura é compatibile all'altra (ma non maggiore o minore)". La differenza tra le due é che solo la seconda é 'nullificabile' mentre la prima non propriamente, anche se poi semanticamente sono analoghe. Si presti quindi attenzione, sulla base della domanda del problema, a definire correttamente \mathcal{H}_0 per il test.

Si noti anche che la PDF di Student [Equazione 11.2](#) é equivalente a:

$$f_{\mathcal{T}}(t) = \frac{\mathcal{G}(0, 1)}{\sqrt{\frac{\chi_{N-1}^2}{N-1}}} = \frac{\mathcal{G}(0, 1)}{\sqrt{\frac{\chi_k^2}{k}}}$$

ovvero la gaussiana standard *corretta* per un valore numerico pari al rapporto tra un χ^2 a $k = N - 1$ gradi di libertà (con N numerosità del campione da cui proviene x) e $N - 1$. Si tratta quindi di un χ_R^2 . La dimostrazione é lasciata in [Appr. 11.5](#).

Compatibilità tra due valori con incertezza.

Un tecnico misura la densità di polvere in una camera pulita prima e dopo un intervento di pulizia. Prima della pulizia avevamo misurato 4 volte (le misure sono in g cm^{-3}):

C'era stato un errore durante la 5a misura ed e' stata scartata. Dopo la pulizia ha ripetuto la misura ottenendo

$$101, 87, 88, 94, 82$$

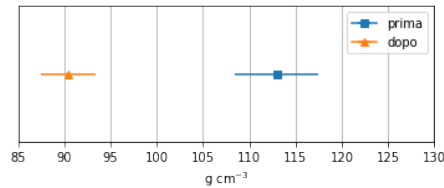
Si nota un significativo miglioramento a seguito della pulizia?

Si comincia calcolando le medie e le deviazioni standard per le due serie. Sia X la prima serie e Y la seconda allora:

$$N = 4 \quad \bar{x} = 113 \quad s_x = 11.015 \quad s_{\bar{x}} = 4.5$$

$$M = 5 \quad \bar{y} = 90.40 \quad s_y = 7.30 \quad s_{\bar{y}} = 2.98$$

Grafichiamo i risultati La valutazione ad occhio é sempre importante. Procediamo ora a definire



l'ipotesi nulla \mathcal{H}_0 : "Le due misure sono compatibili e la seconda é inferiore alla prima". Si tratta quindi di un test ad una coda. Prima di costruire il test, valutiamo il rapporto tra le incertezze. Essendo lo strumento di misura lo stesso ci aspettiamo infatti $s_x \sim s_y$. Confrontando abbiamo $s_x/s_y = 1.5$ e quindi non violano la regola pratica.

Formulazione accurata di t Costruiamo ora:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{N+M}{N+M-2} \sqrt{\frac{N-1}{NM} s_x^2 + \frac{M-1}{NM} s_y^2}}} = 3.7$$

Se svolgiamo ci accontentiamo di un margine di confidenza del 99% otteniamo ($k = N + M - 2 = 4 + 5 - 2 = 7$) $t_0 = 2.998$. Siccome $|t| > t_0$ non passiamo il test e quindi concludiamo che con il 99% di margine di confidenza che c'e' incompatibilita' e che la pulizia ha funzionato. Questo ci dice che solo l'1% delle volte, a partire da queste misure, avremmo erroneamente sostenuto che la pulizia non aveva funzionato. Spingendosi oltre, osservando la tabella si nota come $t_0(\alpha = 0.001) = 4.7$ quindi possiamo affermare che (solo l'1% avremmo compiuti errore di valutazione (tipo 1).

Formulazione veloce di t Se avessimo costruito

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{N} + \frac{s_y^2}{M}}} = 4.2$$

Le conclusioni sarebbero state simili al caso precedente.

Si noti anche che la PDF di Student [Equazione 11.2](#) in questo caso é equivalente a:

$$f_{\mathcal{T}}(t) = \frac{\mathcal{G}(0, 1)}{\sqrt{\frac{\chi_{N+M-2}^2}{N+M-2}}} = \frac{\mathcal{G}(0, 1)}{\sqrt{\frac{\chi_k^2}{k}}}$$

ovvero la gaussiana standard *corretta* per un valore numerico pari al rapporto tra un χ^2 a $k = N+m-2$ gradi di liberta' (con N, M numerosita' dei campioni da cui provengono x, y). Si tratta quindi di nuovo di un χ_R^2 . La dimostrazione é simile a quella in [Appr. 11.5](#).

Valutazione dell'indice di correlazione campionario

Supponiamo di avere due serie di misure

$$x = [2.2, 4.2, 6.2, 8.2, 10.4] \quad y = [10.1, 8.5, 8.2, 6.7, 4.2] \quad s_y = 0.3$$

Se ipotizziamo che siano distribuiti linearmente otteniamo, dopo aver calcolato con il metodo dei minimi quadrati la miglior retta interpolante, un $\chi_m^2 = 11.75$ che confrontiamo con un valore critico per 3 gradi di libertà e un livello di significanza del 0.025 pari a $\chi_0^2 = 12.8$. L'errore a posteriori $\delta_y = 0.27$ simile a quello misurato. Insomma, i dati sembrano ben disposti linearmente. Se calcoliamo il coefficiente di correlazione campionario otteniamo $r_{xy} = -0.973$. Si ricordi che possiamo calcolare la sua incertezza come $s_r = \sqrt{\frac{1-r^2}{N-2}} = 0.13$. Vogliamo valutare quale sia la probabilità che questo valore di correlazione campionaria sia casuale e dovuto alla scarsa numerosità del campione (per due punti passa sempre una retta...). Definiamo quindi come ipotesi nulla $r = 0$. Allora:

$$t = \frac{r_{xy} - 0}{s_r} = 1.4$$

Eseguiamo un test a 1 coda. Il numero di gradi di libertà é $k = N - 2 = 3$. Osservando la tabella, questo corrisponde a circa $\alpha = 0.15$ (1 coda), ovvero solo il 15% delle volte otterrei un $r_{xy} > 1.4$. La possibilità quindi di una correlazione casualmente data dalle fluttuazioni dei dati non e' poi così bassa.

11.3 Chiusura

Quesiti

Es. 11.3.1 . La misura del tempo di percorrenza Δt di un tratto Δs è riportata in tabella. Verificare se la distribuzione dei dati segue un andamento di equazione $\Delta S = a + b\Delta t$ stimando tutti i parametri caratteristici dell'interpolazione lineare, incluso il coefficiente di regressione lineare. Motivare le riposte avvalendosi Chi-quadro. Si usino le formule riportate.

ΔS (mm)	0.11	0.15	0.20	0.17	0.26	0.34	0.50	0.64
Δt (s)	0.20	0.21	0.22	0.23	0.24	0.25	0.26	0.27

Es. 11.3.2 . Un casa editrice ha realizzato 7 versioni grafiche diverse dello stesso libro di Statistica. Si chiede ad un gruppo di persone di scegliere il libro in base alla propria preferenza. Detto i l' i -esimo libro ed f_i la frequenza con cui esso viene scelto, si osserva la distribuzione in tabella. a) Verificare la conformità della distribuzione dei dati con quella prevista teoricamente da una distribuzione uniforme mediante il metodo del χ^2 di Pearson. Motivare la risposta, avvalendosi dei valori tabulati in appendice. b) Ripetere la verifica con il metodo di Neyman. c) Discutere quali dei dati si discostano maggiormente dall'ipotesi di omogeneità

E' possibile (ma non necessario) usare le righe e colonne vuote della tabella.

x_i	1	2	3	4	5	6	7		
f_i	25	18	12	24	18	21	20		
1-10									
1-10									
1-10									
1-10									

11.4 Chiusura

Quesiti

Esercizio 11.4.1 Due giocatori si sfidano ad un videogioco e ottengono rispettivamente i seguenti punteggi:

$$44, 20 + 2A, 49, 27 \quad 18, 24 + A, 46, 11, 33$$

dove A ultima cifra del numero di matricola diversa da 0. *a)* Posso affermare con il 90% di confidenza che il giocatore 1 è mediamente più forte del 2? Si applichi il test del t-Student supponendo che le giocate di ciascun giocatore seguano distribuzioni normali con medie e deviazioni standard incognite. Per risolvere il problema, in particolare si calcoli la media e le stime delle deviazioni standard dei punteggi dei due giocatori, la variabile t- Student associata alla differenza delle medie (giocatore1-giocatore2), i gradi di libertà del test e si dica esplicitamente se l'ipotesi "giocatore 1 mediamente più forte del giocatore 2" sia accettata o rifiutata e perché.

Approfondimenti

Appr. 11.1 Il χ^2 di Neymann. In alcuni casi, ad esempio in cui la numerosità dei conteggi per classi di frequenza in alcuni bin e' bassa, si puo' calcolare una variante del test del χ^2 di Pearson detta di Neymann. In questo caso, l'unica differenza é che l'incertezza che entra al denominatori dei vari termini del $\tilde{\chi}_i^2$ non é la deviazione standard (binomiale o poissoniana) del numero di conteggi misurati, bensí *del valore di riferimento*:

$$s_i = \sqrt{Np_i^*(1 - p_i^*)} \quad \text{o} \quad s_i = \sqrt{n_i^*}$$

Nel caso dell'esercizio nel testo, si sarebbe ottenuto:

Dati			Ipotesi		Test Ipotesi			
Day	n_i	$f_i = p_i$	p_i^*	n_i^*	Pearson	Neymann		
					$\sqrt{Np_i(1 - p_i)}$	$\tilde{\chi}_i^2$	$\sqrt{n_i}$	$\tilde{\chi}_i^2$
Lunedì	8	0.118	0.17	11.6	2.7	1.80	3.1	1.32
Martedì	12	0.176	0.17	11.6	3.1	0.02	3.1	0.02
Mercoledì	12	0.176	0.17	11.6	3.1	0.02	3.1	0.02
Giovedì	10	0.147	0.17	11.6	2.9	0.29	3.1	0.03
Venerdì	11	0.162	0.17	11.6	3.0	0.03	3.1	0.03
Sabato	9	0.132	0.075	5.1	2.8	1.95	2.17	3.22
Domenica	6	0.088	0.075	5.1	2.3	0.15	2.17	0.17
Somma	68	1	1	68		4.25		5.04

E la valutazione del test sarebbe stata simile agli altri casi.

Appr. 11.2 Correzione di Yates. Si puo' dimostrare che i χ_P^2 sono sempre sottostimati per cui talvolta si applica una correzione al calcolo dei vari termini detta di Yates per cui

$$\chi_P^2 = \sum_{i=1}^M \left(\frac{n_i - n_i^* - 0.5}{s_i} \right)^2$$

Nel caso dell'esercizio nel testo, si sarebbe ottenuto:

Dati			Ipotesi		Test Ipotesi		
Day	n_i	$f_i = p_i$	p_i^*	n_i^*	s_i	Pearson $\tilde{\chi}_i^2$	Yates $\tilde{\chi}_i^2$
Lunedì	8	0.118	0.17	11.6	2.7	1.80	2.34
Martedì	12	0.176	0.17	11.6	3.1	0.02	0.00
Mercoledì	12	0.176	0.17	11.6	3.1	0.02	0.00
Giovedì	10	0.147	0.17	11.6	2.9	0.29	0.50
Venerdì	11	0.162	0.17	11.6	3.0	0.03	0.12
Sabato	9	0.132	0.075	5.1	2.8	1.95	1.48
Domenica	6	0.088	0.075	5.1	2.3	0.15	0.03
Somma	68	1	1	68		4.25	4.46

Ma anche in questo caso la valutazione del test sarebbe stata simile agli altri casi.

Appr. 11.3 L'approssimazione Poisson del χ^2 di Pearson e Neyman. La formula

$$\chi_P^2 = \sum_{i=1}^M \frac{(n_i - n_i^*)^2}{n_i}$$

é molto semplice da memorizzare e ricavare, e - al di fuori della fisica - viene molto spesso usata. Una forma in cui spesso viene descritta é

$$\chi_N^2 = \sum_{i=1}^M \frac{(O_i - E_i)^2}{O_i}$$

dove O_i, E_i sta per valori *Observed* e *Expected*. Spesso in applicazione fisiche la formula diviene troppo approssimata, come visto nel testo. Analogamente, il test in versione Neyman si puo' scrivere:

$$\chi_N^2 = \sum_{i=1}^M \frac{(O_i - E_i)^2}{E_i}$$

Appr. 11.4 William Sealy Gosset (1876-1937).

W. S. Gosset fu uno statistico, chimico e birraio inglese che servì anche come Responsabile del Laboratorio Sperimentale della Guinness e fu un pioniere della statistica moderna. Aprì aperta la strada alla progettazione e all'analisi sperimentale di piccoli campioni con un approccio economico alla logica dell'incertezza. Gosset pubblicò sotto lo pseudonimo di Student e sviluppò la più famosa distribuzione t di Student in quanto vincolato dalla Guinness a non pubblicare scoperte sotto il proprio nome che contenessero le parole 'birra' e 'Guinness'. Il titolo della pubblicazione fu infatti "Sul conteggio degli errori con un emacitometro" per la misura di lieviti. Fu amico di Pearson, che lo aiutò anche con la matematica ma fu più apprezzato e pubblicizzato da Fisher, forse oppositore di Pearson.



W. S. Gosset

Appr. 11.5 Dimostrazione PDF \mathcal{T} . La formula si dimostra:

$$t = \frac{|x - x^*|}{s_x} = t = \frac{|x - x^*|}{s_x} \left(\frac{\sigma_x}{\sigma_x} \right) = \frac{|x - x^*|}{\sigma_x} \left(\frac{\sigma_x}{s_x} \right) = \mathcal{G}(0, 1) \frac{1}{s_x/\sigma_x}$$

Ora il termine s_x/σ_x é riscrivibile come:

$$\frac{s_x}{\sigma_x} = \frac{\sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}}{\sigma_x} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{\sigma_x} \right)^2} = \sqrt{\frac{\chi_{N-1}^2}{N-1}}$$

e si ritrova che per $N \rightarrow \infty$ allora $s_x \rightarrow \sigma_x$.

Appr. 11.6 Test t per due valori con incertezza. Per la formulazione accurata di t nel caso che le incertezza dei due campioni siano compatibili: $1/5 < s_x/s_y < 5$ allora si introduce il concetto di *deviazione standard cumulativa*

$$s = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2 + \sum_{j=1}^M (y_j - \bar{y})^2}{N + M - 2}}$$

. Per la differenza delle *medie* $d = \bar{x} - \bar{y}$ prende la forma:

$$s_d = \sqrt{\frac{s^2}{N} + \frac{s^2}{M}} = \sqrt{\frac{1}{N} + \frac{1}{M}} s = \sqrt{\frac{1}{M} + \frac{1}{N}} \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2 + \sum_{j=1}^M (y_j - \bar{y})^2}{N + M - 2}}$$

e quindi la variabile di Student si scrive:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{1}{M} + \frac{1}{N}} \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2 + \sum_{j=1}^M (y_j - \bar{y})^2}{N + M - 2}}}$$

distribuita con una PDF di Student a $N + M - 2$ gradi di libertà, che si puo' pensare anche come

$$f_{\mathcal{T}}(t) = \frac{\mathcal{G}(0, 1)}{\sqrt{\frac{\chi_{N+M-2}^2}{N+M-2}}}$$

Capitolo 12

Il confronto di ipotesi

Nei capitoli precedenti abbiamo quantificato la bontà di una ipotesi quantitativamente attraverso alcuni test statistici come il χ^2 e il test di Student. I test si basavano sul concetto frequentista che a lungo termine si conosce l'andamento della variabile del test (ovvero si conoscono le PDF del χ^2 e di Student). Questo ci ha permesso di valutare quantitativamente le fluttuazioni casuali delle misure definendo una regione di *accettazione* dell'ipotesi e una regione di *rigetto* della ipotesi. In particolare il significato della regione di rigetto, avente probabilità pari al livello di significanza del test, era interpretabile anche come la probabilità di scartare una ipotesi vera, ovvero di definire 'falso' un positivo. Secondo la notazione vista si trattava di errori di tipo I. Abbiamo anche visto per costruzione il metodo quantifica lo scostamento da una ipotesi ritenuta vera, mentre non quantifica la veridicità della ipotesi stessa.

Tuttavia questo metodo di confronto con una ipotesi non è per niente generico, infatti in diverse situazioni inclusa l'analisi di misure fisiche siamo portati a valutare diverse *ipotesi alternative*. Per ipotesi alternativa si intende una ipotesi mutuamente esclusiva rispetto all'ipotesi nulla, ovvero è vera solo se \mathcal{H}_0 è falsa. Si noti che, nel caso di ipotesi parametriche, la differenza tra ipotesi nulla e alternativa può risiedere in un parametro di interesse, ad esempio: "i dati sono distribuiti normalmente come $\mathcal{G}(0.5, 1.5)$ invece che $\mathcal{G}(0, 1)$ ", "i dati seguono andamento lineare del tipo $y = a_2 + b_2x$ e non $y = a_1 + b_1x$ ". Invece non si intende in questo caso il contrario della ipotesi nulla stessa.

Come valutiamo il maggiore accordo dei dati rispetto all'una o all'altra ipotesi? Abbiamo visto in [Capitolo 9](#) che la definizione di una certa ipotesi corrisponde ad una definizione o scelta di parametri. Tutto ciò è esplicitato dalla funzione di verosimiglianza come visto in [Equazione 9.15](#). Allora possiamo pensare di calcolare la verosimiglianza per diverse ipotesi, *mutuamente esclusive*, e poi confrontarle:

$$\mathcal{L}(D|\mathcal{H}_0) \quad > ? < \quad \mathcal{L}(D|\mathcal{H}_a) \dots$$

Vedremo in questo capitolo come effettuare il confronto di queste verosimiglianze. Inoltre, in virtù del fatto che si è proposta una ipotesi alternativa mutuamente esclusiva, vedremo che emergerà un secondo tipo di errori possibili nella valutazione dei dati. Infatti, mentre avendo a disposizione un'unica ipotesi si potevano fare solo errori di tipo I (ovvero classificare come falsi (rifiutare) casi invece positivi), ora si può anche sbagliare classificando come veri (accettando) casi in realtà falsi. Questo tipo di errori si chiama di seconda specie, o di tipo II. La probabilità di errori di queste specie si indica con il simbolo β e il valore $1 - \beta$ è detto *potenza del test* del confronto delle ipotesi.

Cominciamo dal confronto delle verosimiglianze in [sezione 12.1](#) per poi approfondire il concetto errori di tipo I e II in [sezione 12.2](#) e potenza del test in [??](#). Concludiamo con alcuni esempi in [??](#).

12.1 Rapporto delle verosimiglianze

Facciamo un primo esempio piuttosto irrealistico ma semplice per descrivere il concetto di confronto delle verosimiglianze. Prendiamo una singola misura $x = 0$ e vogliamo confrontarla con due distribuzioni normali (ipotesi) di riferimento. Si noti che non necessariamente il confronto deve essere fatto con la distribuzione che dà la *massima* verosimiglianza, ma in generale può essere fatto con qualsiasi distribuzione. Inoltre, con una singola misura, non siamo nemmeno in grado di stimare la deviazione

standard Gaussiana delle misure¹ Quindi, invece che cercare la massima verosimiglianza, definiamo due verosimiglianze per due diverse ipotesi:

$$H_0 : x \rightarrow \mathcal{G}(1, 1) \quad H_a : x \rightarrow \mathcal{G}(2, 4)$$

Quale delle due ipotesi è piu' probabile? Si noti come $x = 0$ è a distanza 1σ dal valore di aspettazione di entrambe le ipotesi. Calcolando il χ^2 ottengo

$$\chi^2(\mathcal{H}_1) = \left(\frac{0-1}{1}\right)^2 = 1 \quad \chi^2(\mathcal{H}_2) = \left(\frac{0-2}{2}\right)^2 = 1$$

Quindi il metodo del χ^2 non da molta informazione. La verosimiglianza invece e' la probabilità di ottenere un certo set di dati a partire da una certa ipotesi. Calcolando quindi le verosimiglianze per le due ipotesi:

$$\mathcal{L}_0(x=0|\mathcal{H}_0) = \frac{1}{\sqrt{2\pi}1} e^{-\frac{1}{2}\left(\frac{0-1}{1}\right)^2} = 0.24 \quad \mathcal{L}_a(x=0|\mathcal{H}_a) = \frac{1}{\sqrt{2\pi}2} e^{-\frac{1}{2}\left(\frac{0-2}{2}\right)^2} = 0.12$$

La verosimiglianza rispetto a H_0 è piu' grande! E quindi H_0 (che è piu stretta di H_a) è piu' probabile. Per la valutazione delle ipotesi si usa il concetto di *rapporto di verosimiglianze*:

Si definisce **rapporto di verosimiglianza** tra due ipotesi $\mathcal{H}_a, \mathcal{H}_0$ il rapporto:

$$\Lambda = \frac{\mathcal{L}_a}{\mathcal{L}_0} = \frac{\mathcal{L}(\bar{x}|\mathcal{H}_a)}{\mathcal{L}(\bar{x}|\mathcal{H}_0)}$$

Nell'esempio precedente quindi $\Lambda = 0.5$ e si preferisce quindi ipotesi \mathcal{H}_0 . Si notino questi ulteriori due casi:

- Supponiamo di avere ora una singola misura $x = 2$ e le stesse ipotesi di prima, allora

$$\mathcal{L}_0(x=2|\mathcal{H}_0) = \frac{1}{\sqrt{2\pi}1} e^{-\frac{1}{2}\left(\frac{2-1}{1}\right)^2} = 0.24 \quad \mathcal{L}_a(x=2|\mathcal{H}_a) = \frac{1}{\sqrt{2\pi}2} e^{-\frac{1}{2}\left(\frac{2-2}{2}\right)^2} = 0.2$$

e il confronto ritorna $\Lambda = \mathcal{H}_a/\mathcal{H}_0 \simeq 0.83$ quindi, ancora in virtù del fatto che la prima è piu' stringente, nonostante il singolo valore coincida col valore di aspettazione della seconda, la prima è più probabile.

- Se prendessimo le due misure $x = 0, 2$ e confrontassimo $\mathcal{H}_0, \mathcal{H}_a$ otterremo

$$\mathcal{L}_0(x=0, 2|\mathcal{H}_0) = \frac{1}{\sqrt{2\pi}1} e^{-\frac{1}{2}\left(\frac{2-1}{1}\right)^2} \cdot \frac{1}{\sqrt{2\pi}1} e^{-\frac{1}{2}\left(\frac{2-1}{1}\right)^2} = 0.24 \cdot 0.24 = 0.0576$$

$$\mathcal{L}_a(x=0, 2|\mathcal{H}_a) = \frac{1}{\sqrt{2\pi}2} e^{-\frac{1}{2}\left(\frac{0-2}{2}\right)^2} \cdot \frac{1}{\sqrt{2\pi}2} e^{-\frac{1}{2}\left(\frac{2-2}{2}\right)^2} = 0.12 \cdot 0.2 = 0.024$$

e $\Lambda = \mathcal{H}_1/\mathcal{H}_2 \simeq 0.42$ con una ancora maggiore preferenza per \mathcal{H}_0 .

Il criterio per selezionare una ipotesi (ad esempio \mathcal{H}_0 rispetto ad un'altra \mathcal{H}_a) non è semplicemente che $\mathcal{L}_0 > \mathcal{L}_a$. Di norma si definisce un *livello di riferimento* $k \neq 1$ in maniera che:

$$\Lambda = \frac{\mathcal{L}_a}{\mathcal{L}_0} = \begin{cases} < k & \text{accetto } \mathcal{H}_0: \text{ affermo che } \mathcal{H}_0 \text{ è vera e } \mathcal{H}_a \text{ falsa} \\ > k & \text{rifiuto } \mathcal{H}_0: \text{ affermo che } \mathcal{H}_0 \text{ è falsa e } \mathcal{H}_a \text{ vera} \end{cases}$$

La scelta del valore di k , come anticipato nella introduzione a questo capitolo, non dipende ora unicamente dal rischio di errori di tipo I che accettiamo di avere (falsi positivi), ma anche di errori di tipo II. Vediamo di chiarire il concetto.

¹Si potrebbe dire che una prima stima della deviazione standard è l'errore della singola misura, se noto.

Si noti che, in virtù di proprietà già viste, spesso è più conveniente usare il logaritmo del rapporto delle verosimiglianze:

$$\ln \Lambda = \ln \frac{\mathcal{L}_a}{\mathcal{L}_0} = \ln \mathcal{L}_a - \ln \mathcal{L}_0$$

inoltre, spesso questo test viene costruito in maniera^a per cui al denominatore di ha sempre ipotesi di massima verosimiglianza. In questo caso, la differenza sopra sarebbe negativa, allora si preferisce usare

$$-2 \ln \Lambda = -2 \ln \frac{\mathcal{L}_a}{\mathcal{L}_0} = 2(\ln \mathcal{L}_0 - \ln \mathcal{L}_a)$$

che è sempre positivo.

^aAd esempio quando si prevede di usare il teorema di Wilks per il calcolo della bontà del confronto e delle incertezze sui parametri

12.2 Errori di tipo I e di tipo II

Nel confronto di due ipotesi $\mathcal{H}_0, \mathcal{H}_a$ mutuamente esclusive, i casi possibili sono:

- \mathcal{H}_0 è vera, e l'accetto (vero positivo)
- \mathcal{H}_0 è vera, ma la scarto a favore di \mathcal{H}_a (falso positivo, errore di tipo I)
- \mathcal{H}_0 è falsa, e la scarto a favore di \mathcal{H}_a (vero negativo)
- \mathcal{H}_0 è falsa, ma la accetto (falso negativo, errore di tipo II)

Si noti come nella descrizione per dichiarare positivo o negativo si pone l'attenzione su \mathcal{H}_0 , cosa che talvolta può suscitare confusione in termini. Di conseguenza si consiglia di adoperare unicamente il concetto di errori di tipo I e II, che è univocamente definito. Rappresentando il concetto sopra in tabella abbiamo:

	\mathcal{H}_0 vera	\mathcal{H}_0 falsa
\mathcal{H}_0 accettata $\Lambda < k$	(True Positive TP) $1 - \alpha$	(False Negative FN) [Err. Tipo II] β
\mathcal{H}_0 rifiutata $\Lambda > k$	(False Positive FN) [Err. Tipo I] α	(True Negative TN) $1 - \beta$

È importante soffermarsi sul significato degli errori dei due tipi. Anzitutto, si consideri che il significato pratico dipenderà dal caso sotto scrutinio. Esempi:

- Si consideri un test di positività ad una malattia. Il test può dare un risultato corretto in due maniere: il test da positivo e il paziente ha quella malattia (true positive: correttamente identificato un positivo), e il test da negativo e il paziente non ha quella malattia (true negative: correttamente identificato un negativo). Il test può sbagliare in due casi: il paziente non ha la malattia e il test risulta positivo (false positive) e il viceversa in cui il paziente è malato e il test da negativo (falso negativo). In questo caso l'analogia linguistica è ottima.
- Si consideri uno strumento che misura eventi: alcuni sono eventi del segnale che si vuole misurare (ad esempio un decadimento), altri sono eventi cosiddetti di rumore (ad esempio il sensore è scattato a causa di una fluttuazione elettronica). Se si hanno a disposizione dei parametri di qualità, si può costruire un test che abbia due ipotesi alternativa secondo questi parametri. Come sopra il test può funzionare in due casi: correttamente identificare un evento di rumore. Oppure può sbagliare in due modi: scartando un segnale (false positive) o accettando un evento di rumore (false negative)

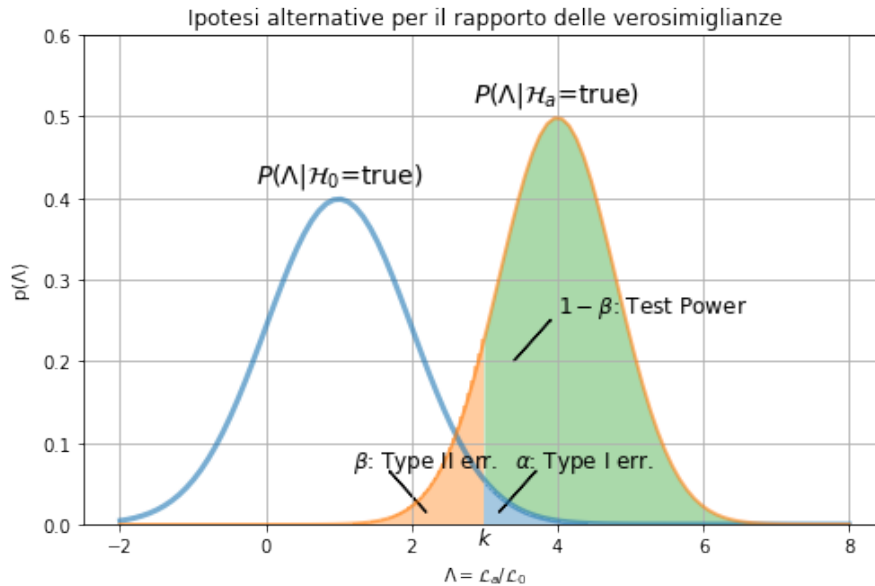


Figura 12.1: Visualizzazione grafica del concetto di errore di tipo I e II. x_0 rappresenta idealmente un set di misure.

A seconda di come e' definita la regione di accettazione o rifiuto si avranno diverse percentuali di α, β e quindi diverse proporzioni di errori. Dipende caso a caso dallo scienziato cosa intende accettare.

Esiste una maniera grafica molto pratica per visualizzare i due tipi di errori, la potenza del test e le regioni di accettazione e rifiuto. Si supponga in fatti di rappresentare in un grafico il valore Λ ($0 \ln \Lambda$) in ascissa e in ordinata le due verosimiglianze $L(\lambda|\mathcal{H}_0) = p(\lambda|\mathcal{H}_0)$ e $L(\lambda|\mathcal{H}_a) = p(\lambda|\mathcal{H}_a)$. Nel grafico le due curve rappresentano i valori della verosimiglianza per le due ipotesi calcolati *per ogni* valore di Λ . Ora, data una serie di misure, e due ipotesi alternative, Λ risulta definito e calcolabile. Si riporta il grafico in [Figura 12.1](#)

Allora, con questa notazione:

$$P_I = \alpha = P(\Lambda > k|\mathcal{H}_0)$$

$$P_{II} = \beta = P(\Lambda < k|\mathcal{H}_a)$$

Un confronto di ipotesi funzionale cerca di minimizzare entrambi gli errori di prima e seconda specie e affinché cio' avvenga le due ipotesi devono essere *sufficientemente distanti*. Spieghiamo questo concetto qui di seguito. Attenzione che di norma le ipotesi si basano sui parametri, e quindi cio' che si fa e' costruire ipotesi (se possibile) tali che il parametro le distingua il piu' chiaramente possibile.

Ad esempio, se vogliamo valutare con un test la qualità delle mele in un cassa possiamo:

- Assaggiarle tutte e valutare la bontà di ciascuna. Forse questo test non é pero' opportuno per il cliente
- Valutarle a occhio: il parametro della ipotesi e' quindi 'giudizio visivo'. Tuttavia se il danno e' sotto la superficie il test avra' un elevato tasso di errori di seconda specie (accetto mela danneggiata)
- Valuto con tecniche IR, UV e ottiche. Combino i parametri delle tre lunghezze d'onda in maniera da avere un parametro combinato di qualità. Anche questo parametro avra' una incertezza. Anche in questo caso posso valutare i valori del parametro per cui accetto un negativo o scarto un positivo.

12.2.1 Analogia con il parametro di interesse

Il rapporto delle verosimiglianze puo' essere usato in linea generale per ipotesi alternative anche con un numero diverso di parametri non noti, ad esempio, ci si puo' chiedere se una distribuzione di dati segua una distribuzione preferenzialmente lineare o quadratica. Tuttavia, siccome l'ipotesi di

riferimento e' una ipotesi *fisica* molto piu' spesso accade di chiedersi se a partire dalla stessa classe di ipotesi parametrica di parametri $\vec{\theta}$ sia piu' probabile un certo valore di un parametro (ad es di θ_1) oppure un altro. In questo caso allora si possono formare due ipotesi alternative ad esempio

- il parametro $\theta_1 = 0$ o $\theta_1 = 0.5$
- il parametro $\theta_1 = 0$ o $\theta_1 > 0.0$
- e cosi via.

Allora, con questa notazione:

$$P_I = \alpha = P(\theta_1 > c | \mathcal{H}_0)$$

$$P_{II} = \beta = P(\theta_1 < c | \mathcal{H}_a)$$

dove c identifica il valore che definisce le regioni di rifiuto e di accettazione.

Queste ipotesi hanno lo stesso numero di gradi di parametri ignoti. In questa condizione l'interesse e' quindi sul parametro visto che la PDF e' la stessa, allora il confronto quantitativo si puo' fare invece che su Λ rispetto al valore di riferimento k sul valore dei parametri stessi. Faremo un esempio di questo in [sezione 12.4](#).

12.3 Potenza del test

In generale siamo nella situazione in cui, data una certa misura sperimentale x e due ipotesi, o esse sono troppo simili per essere discriminate, cosi' diverse che risulti ovvio quale delle due e' vera, e la situazione intermedia. La faccenda e' illustrata in figura [Figura 12.2](#). Normalmente, non si ha estrema liberta' di costruire una ipotesi alternativa. Tuttavia, e' importante notare come nel caso di ipotesi troppo vicine, in sostanza la probabilita' di errori di seconda specie tenda a 1, e quindi il test non abbia piu' senso. Si noti come nel grafico abbiamo messo in ascissa il parametro di interesse piuttosto che il rapporto delle verosimiglianze.

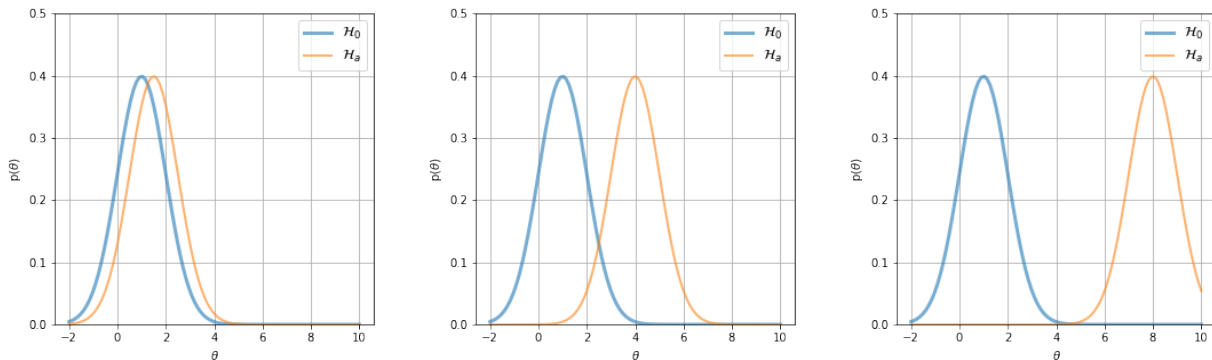


Figura 12.2: Esempio di ipotesi alternativa molto simile (sinistra), discriminabile (centro) e perfettamente separabile (destra).

Considerando che l'ipotesi alternativa e' vera se l'ipotesi nulla e' falsa, per costruzione del test, ha senso definire un parametro detto la potenza del test, che non e' altro che la probabilita' di accettare l'ipotesi alternativa quando questa e' vera, ovvero, numericamente:

$$P = 1 - \beta$$

Si tratta quindi del concetto complementare all'errore di seconda specie. In generale, quando si confrontano ipotesi e' sempre importante definire la potenza del test, in alternativa al concetto di significanza del test nel caso di confronto con la sola ipotesi nulla. La potenza deve essere piu' grande possibile. Il lemma fondamentale di Neyman-Pearson (dimostrato sul Loretì) stabilisce che il rapporto della verosimiglianze tra due ipotesi parametriche semplici (parametri scorrelati) rappresenta il test di verifica piu' potente a livello di significanza α per una soglia k .

Si noti infine che per il primo dei casi in figura, la potenza del test, mediamente e' $P = 0.5$. Si ha cioe' il 50% di probabilita' di accettare una misura scorretta, o di rifiutare una misura vera. Si tratta in sostanza del 50% di probabilita' di sbagliare.

12.4 Un esempio

Si prenda un misura X ad esempio del valore di corrente di fondo di un sensore. La misura e' stata svolta diverse volte dando misure distribuite normalmente con dispersione $s_x = 1$ attorno al valore centrale. Si prenda un campione di $N = 10$ misure ripetute di un certo gruppo di sensori e si ottiene un certo valore $\bar{x} = 0.7$ in unita' arbitrarie. Si vuole confrontare questo valore con due ipotesi: $\mathcal{H}_0 = \mathcal{G}(0, 1)$ e $\mathcal{H}_a = \mathcal{G}(1, 1)$. Nella prima ipotesi, la corrente di fondo ha media nulla e il sensore e' buono, nel secondo caso la corrente di fondo non ha media nulla e il sensore e' difettoso e va scartato. Si vuole calcolare il tasso di errori di prima e seconda specie e la potenza del test.

Partiamo costruendo le due verosimiglianze. Si noti che si tratta di una media aritmetica su $N = 10$ misure.

$$\mathcal{L}_0(\bar{x} = 0.7|\mathcal{H}_0) = \frac{1}{\sqrt{2\pi s_{\bar{x}}}} e^{-\frac{1}{2} \left(\frac{0.7 - \mu_0}{s_{\bar{x}}} \right)^2} = 0.11 \quad \mathcal{L}_a(\bar{x} = 0.7|\mathcal{H}_a) = \frac{1}{\sqrt{2\pi s_{\bar{x}}}} e^{-\frac{1}{2} \left(\frac{0.7 - \mu_a}{s_{\bar{x}}} \right)^2} = 0.11$$

ovvero $\Lambda = \frac{\mathcal{L}_a}{\mathcal{L}_0} = XX$. Ora, per poter definire le regioni di accettazione o rifiuto possiamo usare due vie: o calcoliamo la probabilita' di Λ oppure usiamo il parametro fisico di interesse \bar{x} . Percorriamo entrambe le strade ma partiamo anzitutto osservando che:

$$\begin{aligned} \ln \Lambda &= \ln \frac{\mathcal{L}_a}{\mathcal{L}_0} = \ln \mathcal{L}_a - \ln \mathcal{L}_0 = -\frac{1}{2\sigma^2} \left[-2N\bar{x}(\mu_a - \mu_0) + (\mu_a^2 - \mu_0^2) \right] \\ &= (\mu_0 = 0, \mu_a = 1) := -\frac{1}{2}(2N\bar{x} - N) = \frac{N}{2}(\bar{x} - 1) \end{aligned}$$

Abbiamo quindi un legame algebrico tra il (logaritmo del) rapporto di verosimiglianze e il parametro di interesse, ovvero:

$$\Lambda > k \quad \rightarrow \quad \bar{x} < \frac{N - 2 \ln(k)}{2N} = c$$

Confronto di ipotesi per il parametro di interesse Prendiamo il valore misurato $\bar{x} = 0.7$. Supponiamo di accettare l'ipotesi nulla e dire che il campione di sensori e' buono. Avendo a disposizione l'ipotesi alternativa possiamo calcolare non solo il livello di significanza ma anche la probabilita' di errori di seconda specie:

$$\begin{aligned} \alpha &= P(\bar{x} > 0.7|\mathcal{H}_0) = \int_{0.7}^{\infty} \mathcal{G}_0 dx = 0.0134 \simeq 1.3\% \\ \beta &= P(\bar{x} < 0.7|\mathcal{H}_a) = \int_{-\infty}^{0.7} \mathcal{G}_a dx = 0.17 \simeq 17\% \end{aligned}$$

La potenza del test e' $P = 1 - \beta \simeq 83\%$ quindi accettiamo circa meno del 13% dei sensori difettosi ma e scartiamo circa l'1% dei sensori funzionanti.

Se volessimo accettare meno sensori non funzionanti, dovremmo cambiare il valore di riferimento, prendiamo per esempio $c = 0.4$. In questo caso:

$$\begin{aligned} \alpha &= P(\bar{x} > 0.4|\mathcal{H}_0) = \int_{0.4}^{\infty} \mathcal{G}_0 dx = 0.102 \simeq 10\% \\ \beta &= P(\bar{x} < 0.4|\mathcal{H}_a) = \int_{-\infty}^{0.4} \mathcal{G}_a dx = 0.0289 \simeq 3\% \end{aligned}$$

La potenza del test e' $P = 1 - \beta \simeq 97\%$ quindi accettiamo circa meno del 3% dei sensori difettosi ma e scartiamo circa il 10% dei sensori funzionanti.

Si riportano graficamente i due casi sopra in [Figura 12.3](#).

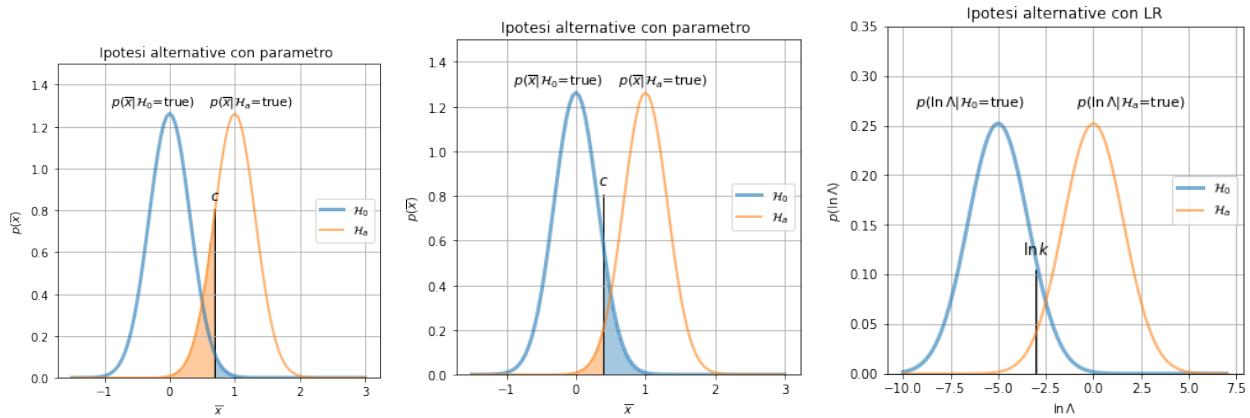


Figura 12.3: Confronto di ipotesi alternative: (sinistra) calcolando la distribuzione di probabilità rispetto al parametro \bar{x} e usando $c = 0.7$ come separatore. (centro) come a sinistra usando $c = 0.4$ come separatore. (destra) calcolando la distribuzione di probabilità in funzione del rapporto delle verosimiglianze e usando $\ln k = -3$ (corrispondente a $c = -0.4$) come separatore

Confronto di ipotesi per il rapporto di verosimiglianza Siccome sappiamo che \bar{x} è una variabile aleatoria Gaussiana di aspettazione \bar{x} e deviazione standard $\sigma_{\bar{x}}$, da questa possiamo costruire le PDF per $\ln \Lambda$:

- $p(\ln \Lambda | \mathcal{H}_0) = \mathcal{L}(\ln \Lambda | \mathcal{H}_0) \equiv \mathcal{G}(-\frac{N}{2}(1 + \mu_0), \frac{N}{2}\sigma_{\bar{x}}) = \mathcal{G}(-5, (5\sigma_{\bar{x}})^2)$
- $p(\ln \Lambda | \mathcal{H}_a) = \mathcal{L}(\ln \Lambda | \mathcal{H}_a) \equiv \mathcal{G}(-\frac{N}{2}(1 + \mu_a), \frac{N}{2}\sigma_{\bar{x}}) = \mathcal{G}(0, (5\sigma_{\bar{x}})^2)$

Allora, se prendiamo ancora $c = 0.4$ otteniamo $\ln k = \ln \Lambda = \frac{N}{2}(1 - c) = -3$ e da questo calcolare:

$$\alpha = P(\ln \Lambda > -3 | \mathcal{H}_0) = \int_{-3}^{\infty} \mathcal{G}(-5, (5\sigma_{\bar{x}})^2) dx = 0.1030 \simeq 10\%$$

$$\beta = P(\ln \Lambda < -3 | \mathcal{H}_a) = \int_{-\infty}^{-3} \mathcal{G}(0, (5\sigma_{\bar{x}})^2) dx = 0.0289 \simeq 3\%$$

come nel caso precedente.

12.5 Chiusura

Quesiti

Es. 12.5.1 . Usando il metodo del rapporto di verosimiglianze, verificare se la singola misura $x = -1$ ha maggior probabilità di appartenere alla distribuzione normale $N_1(x; 0.5, 1.5)$ o $N_2(x; -1.5, 0.5)$

Approfondimenti

Appr. 12.1 Il teorema di Wilks. Il teorema di Wilks semplifica il calcolo della potenza del test e dei margini di confidenza nel caso in cui l'ipotesi nulla e quella alternativa abbiamo lo stesso numero di gradi di libertà. Questa situazione avviene spesso nel caso di *nested sampling* ovvero nel caso in cui l'ipotesi di riferimento sia costruita a partire da diversi parametri, di cui di interesse solo alcuni e gli altri siano di rumore. Supponiamo che vi sia un unico parametro di interesse λ e tutti gli altri siano di rumore \vec{v} allora si può costruire il rapporto di verosimiglianza mettendo al denominatore la verosimiglianza massima su tutti i parametri, e al numeratore, per ogni valore di λ trovare la verosimiglianza avendo fissato i \vec{v} :

$$-2 \ln \Lambda(\lambda) = -2 \ln \frac{\mathcal{L}(\lambda; \hat{\vec{v}})}{\mathcal{L}(\hat{\lambda}; \hat{\vec{v}})} = -2 [\ln \mathcal{L}(\lambda; \hat{\vec{v}}) - \ln \mathcal{L}(\hat{\lambda}; \hat{\vec{v}})] \simeq \chi^2$$

Allora si puo' dimostrare che la funzione sopra ha come limite statistico la PDF del χ^2 con 1 grado di liberta'. Cioe' utile in quanto a questo punto diventa semplice calcolare il valore piu' probabile (moda) del parametro di interesse λ e la sua deviazione standard (radice della varianza)

Parte III

RACCOLTA ESERCIZI

Capitolo 13

Quesiti d'Esame

Quelli di seguito sono tutti esercizi presi da appelli precedenti, talvolta leggermente modificati. Siano di base per una autovalutazione.

13.1 Prima Parte Programma

Esercizi AS. Analisti Statistica Campioni di misure

L'esercizio che segue e' decisamente prototipale. Non ha senso porne diversi perché sono tutti simili. Per risolvere questo esercizio si usano concetti definiti in [Capitolo 2](#) e [Capitolo 3](#) principalmente. Si usa anche il calcolo di compatibilità descritto in [Capitolo 10](#).

Esercizio 13.1.1 Si consideri il seguente campione di misure ripetute di campo magnetico B in cui n_B è la frequenza assoluta, sostituendo ad A l'ultimo numero di matricola diverso da zero.

B [nG]	n_i	B [nG]	n_i
A.00	3	A.04	32
A.01	8	A.05	10
A.02	11	A.06	5
A.03	21	A.07	1

Definire e stimare rispettivamente: a) la numerosità del campione; b) la deviazione standard della singola misura s ; c) la moda e la mediana; d) il valor medio e relativa incertezza, assieme all'errore relativo standard di questo valore; e) si studi la compatibilità di tale valor medio con il valore di riferimento $B_R = (A.040 \pm 0.001)$ nG; f) si calcoli l'intervallo di confidenza centrato sul valor medio e di ampiezza 3 deviazioni standard, a cui applicare il criterio ragionato di reiezione dati; g) si scriva l'espressione analitica della curva normale (Gauss) che meglio rappresenta i dati sperimentali, h) calcolare anche media geometrica, armonica e quadratica. Motivare le risposte.

Cosa da prestare attenzione: uso corretto formule, uso corretto cifre significative e approssimazioni durante i conti, differenza tra deviazione standard campionaria e della media.

Esercizio 13.1.2 Si consideri il seguente campione di misure ripetute di periodo di doppio oscillazione di un pendolo t [s] in cui n_t è la frequenza assoluta.

t [s]	n_i	t [s]	n_i
2.00	2	2.08	29
2.02	9	2.10	9
2.04	13	2.12	4
2.06	23	2.14	1

Definire e stimare rispettivamente: a) la numerosità del campione; b) il valor medio e relativa incertezza, assieme all'errore relativo standard di questo valore la deviazione standard della singola misura

s; c) il 2 e 8 decile; d) si studi la compatibilità di tale valor medio con il valore di riferimento per il singolo periodo $t_R = (1.030 \pm 0.001)$ s; e) si calcoli l'intervallo di confidenza centrato sul valor medio e di ampiezza 3 deviazioni standard, a cui applicare il criterio ragionato di reiezione dati; f) quanti conteggi ci aspettavamo per la classe $t = 2.06$ s. Motivare le risposte.

Esercizio 13.1.3

Si consideri il seguente campione di misure ripetute della lunghezza L di un filo metallico sottoposto a trazione a forza costante fissata in cui f è la frequenza assoluta delle misure effettuate. Definire e stimare rispettivamente: a) la deviazione standard della singola misura b) la moda e la mediana; c) la miglior stima il valor medio e relativa incertezza; d) il valore atteso di f assumendo che la densità di probabilità gaussiana ben rappresenti nella classe 45.20mm. e) indicare se è necessario effettuare la reiezione di dati qualora si assuma il criterio del 3σ ; f) l'intervallo di confidenza centrato sul valor medio e di ampiezza 2σ ; g) assumendo la lunghezza a riposo $L_0 = (A.10 + 45) \pm 0.01$ mm, calcolare l'allungamento medio ΔL subito dal filo e la relativa incertezza (A è parametro fornito dal primo esercizio); h) la compatibilità di tale valore con il valore di riferimento (0.24 ± 0.01) mm. Motivare le risposte nell'elaborato.

L (mm)	f
45.00	7
45.10	12
45.20	14
45.30	5
45.40	8

Esercizi PROB e COMB. Teoria probabilità e Calcolo Combinatorio

Gli esercizi PROB e COMB si basano su operazioni logiche tra eventi aleatori in cui non ci siano distribuzioni di probabilità continue o discrete di riferimento. Si risolvono con le regole del [Capitolo 4](#). Ulteriori esercizi si trovano alla fine di quel capitolo.

Es. 13.1.1 . Nel gioco Malefiz ciascun giocatore possiede una pedina che può entrare nel circuito del gioco solo se, lanciando un dado a sei facce numerate da uno a sei, osserva proprio il numero 6. Dopo esser entrato nel circuito di gioco, quando arriva il suo turno, il giocatore lancia il dado muove la pedina di un numero di passi pari alla faccia del dado uscita. Per raggiungere la meta deve muovere la pedina di almeno 24 passi in totale. a) calcolare qual è il numero minimo di lanci che gli garantiscono di andare in meta; b) supponendo che il giocatore abbia potuto fare 3 lanci del dado complessivamente, calcolare qual è la probabilità che abbia fatto 7 passi; c) calcolare qual è la probabilità che con il quarto lancio arrivi a compiere 12 passi se al terzo lancio si trova ad aver fatto 7 passi. d) stimare se gli eventi descritti ai punti in b) e c) sono statisticamente indipendenti. Motivare le risposte nell'elaborato.

Esercizio 13.1.4 [3 feb 2020] Un'azienda di manutenzione deve definire se il comportamento di termometri è affidabile Per valutarlo utilizza dei sensori. Focalizza perciò l'attenzione su due eventi: A =il sensore segnala che vi è una anomalia; B =il termometro è difettoso. Si indichi con $!A$ il complementare di A e con $!B$ il complementare di B rispettivamente. Sapendo che $P(A|B) = 0.99$ e $P(!A|!B) = 0.98$, calcolare la probabilità che l'anomalia rivelata sia reale, sapendo che $P(B) = 0.3$

Esercizio 13.1.5 [8 luglio 2021] Un naufrago si trova su un'imbarcazione con carburante sufficiente a percorrere $d = 60$ Km. Non ha a disposizione strumenti a parte una radio, la visibilità è nulla e l'imbarcazione è ferma. Riceve un ultimo segnale prima che la radio si spenga definitivamente: si trova a $h = 10$ km da una costa rettilinea e può tenere solo una rotta rettilinea. E' indeciso tra due linee di condotta: 1) Muoversi in linea retta secondo una direzione a caso fino ad esaurire il carburante; 2. Muoversi in linea retta per $d/3 = 20$ km secondo una direzione a caso e, se non si incontra la costa, invertire la rotta e percorrere $2/3 d = 40$ km nella stessa direzione ma nel verso opposto. Stimare la probabilità P di raggiungere la salvezza per ognuno di questi due casi e indicare il metodo migliore. Motivare le risposte.

Esercizi PDF. Applicazioni distribuzioni probabilità

In questo tipo di esercizi, bisogna capire quale sia la distribuzione di probabilità di riferimento relative ad un certo fenomeno casuale. Dopodiché' si procede facendo calcoli di probabilità talvolta condizionate. Si faccia riferimento quindi ai capitoli [Capitolo 4](#) e [Capitolo 5](#).

Esercizio 13.1.6 [8 luglio 2021] In un libro, la probabilità che un carattere sia sbagliato e' pari a $(A + 1) \cdot 10^{-4}$ dove A é l'ultimo numero diverso da zero della matricola. Se in una pagina ci sono 1800 caratteri: *a*) quale e' la probabilità che vi sia una pagina intera senza errori; *b*) quale e' la probabilità che in 2 pagina vi sia esattamente un errore?; *c*) se vi sono 50 pagine nel libro, quante pagine con almeno un errore ci saranno in media?; *d*) che probabilità c'e' di avere esattamente $4A$ pagine con almeno un errore? Si indichino i tipi di distribuzione di probabilità a cui fanno riferimento gli eventi aleatori in questione.

Esercizio 13.1.7 [17 giugno 2021] Nel mese di Aprile a Padova si verificano in media 2 eventi piovosi a settimana (si assumano questi eventi scorrelati, casuali e molto brevi). Si risponda alle seguenti domande: *a*) che probabilità c'e' che non piova per 10 giorni consecutivi? *b*) che probabilità c'e' di avere un giorno di pioggia con almeno un evento piovoso in aprile? *c*) quanti giorni con almeno un evento piovoso ci si aspetta in media nel mese di Aprile? *d*) che probabilità c'e' di avere esattamente $A + 3$ giorni di pioggia in Aprile? Si indichino i tipi di distribuzione di probabilità a cui fanno riferimento gli eventi aleatori in questione. Si ricorda che Aprile ha 30 giorni¹.

Esercizio 13.1.8 [9 settembre 2021] In una focacceria arrivano in media 30 clienti all'ora. Questi comprano una o piu' focacce ciascuno. Si risponda argomentando alle seguenti domande: *a*) a che distribuzione di probabilità di puo' associare il numero di clienti all'ora nella focacceria? *b*) quale e' la probabilità che non entri nessun cliente per 2 minuti di seguito? *c*) quale e' la probabilità che arrivino invece almeno 2 clienti in 2 minuti di seguito?

I clienti comprano 1 singola focaccia in media il $(30 + A)\%$ delle volte. Nessuno esce senza comprare la focaccia. A é l'ultimo numero diverso da zero della matricola. *d*) Calcolare la probabilità che se negli ultimi 10 minuti sono entrati 4 clienti, allora 1 solo di questi abbia comperato una sola focaccia.

Esercizi ASP. Valori aspettazione

Gli esercizi qui di seguito sono tutti molto simili. Si tratta di elaborare sulle condizioni per la quali una funzione possa essere una PDF. Si procede poi nel calcolare alcuni momenti e probabilità. Si faccia riferimento a [Capitolo 5](#).

Esercizio 13.1.9 Sia $f(x)$ la densità di probabilità di una variabile casuale continua x definita come segue:

$$f(x) = \begin{cases} 1 + Ax^2 & \text{per } x \in \left[-\frac{1}{A}, 0\right] \\ k \left(1 - \frac{1}{A}x\right) & \text{per } x \in]0, A] \\ 0 & \text{altrove} \end{cases}$$

dove A è parametro e k è parametro reale da determinarsi. Si definiscano e si stimino *a*) la speranza matematica, *b*) la deviazione standard, *c*) la funzione di distribuzione (CDF), *d*) la probabilità che $x \in [0, A/2]$. Motivare le risposte.

Esercizio 13.1.10 [25 Marzo 2021] Sia $f(x)$ PDF di variabile casuale reale X definita come segue:

$$f(x) = \begin{cases} k(|x| + (A + 1)x^2) & x \in [-1; +1] \\ 0 & \text{altrove} \end{cases}$$

con k costante da determinarsi e A ultimo numero diverso da zero della propria matricola. Si risponda motivando le seguenti domande: *a*) definire k affinché $f(x)$ sia una PDF, *b*) definire implicitamente e calcolare la speranza matematica $E[X]$; *c*) Definire e calcolare la varianza σ_X^2 ; *d*) Ricavare la funzione cumulativa di distribuzione di probabilità $F_X(x)$; *e*) Calcolare la probabilità che $x \in [-1; 1/2]$.

Esercizio 13.1.11 [8 luglio 2021] Sia $f(x)$ PDF di variabile casuale reale X definita come segue:

$$f(x) = \begin{cases} k(x + A) \cos(x) & x \in [0; \pi/2] \\ 0 & \text{altrove} \end{cases}$$

¹Durante l'appello 4 studenti lo avevano dimenticato.

con k costante da determinarsi e A ultimo numero diverso da zero della propria matricola. Si risponda motivando le seguenti domande: *a*) definire k affinché $f(x)$ sia una PDF, *b*) definire implicitamente e calcolare la speranza matematica $E[X]$; *c*) Definire e calcolare la varianza σ_X^2 ; *d*) Ricavare la funzione cumulativa di distribuzione di probabilità $F_X(x)$; *e*) Calcolare la probabilità che $x \in [0; \pi/4]$.

Esercizio 13.1.12 [17 giugno 2021] Sia $f(x)$ PDF di variabile casuale reale X definita come segue:

$$f(x) = \begin{cases} \frac{k}{A+x^2} & x \in [-1; +1] \\ 0 & \text{altrove} \end{cases}$$

con k costante da determinarsi e A ultimo numero diverso da zero della propria matricola. Si risponda motivando le seguenti domande: *a*) definire k affinché $f(x)$ sia una PDF, *b*) definire implicitamente e calcolare la speranza matematica $E[X]$; *c*) Definire e calcolare la varianza σ_X^2 ; *d*) Ricavare la funzione cumulativa di distribuzione di probabilità $F_X(x)$; *e*) Calcolare la probabilità che $x \in [0; 1/2]$.

Esercizio 13.1.13 Sia X una variabile casuale gaussiana centrata in $\mu = 20$ e varianza pari a 4. Si calcoli la probabilità di trovare un valore compreso in $[18; 22]$ e in $[18, 24]$. Assumendo che $y = A(x^2 + x - 1)$ con A il risultato del primo esercizio, calcolare la covarianza $\text{COV}[X, Y]$ per $x = 18$ e $x = 24$.

Esercizi PEC. Propagazione degli Errori Casuali

Per questi esercizi, oltre a conoscere in dettaglio la formula della propagazione degli errori, è importante aver compreso il concetto di correlazione tra variabili aleatorie. Si faccia riferimento a [Capitolo 6](#)

Esercizio 13.1.14 [8 luglio 2021. modificata] Sia X, Y, Z variabili aleatorie continue in \mathcal{R} legate dalla relazione funzionale:

$$z = g(x, y) = x \arcsin(y) + y A^{1/x}$$

dove A ultimo numero diverso da zero della propria matricola. Le variabili X, Y hanno correlazione pari a $\text{COV}[X, Y] = 0.001$. X, Y sono state campionate ripetutamente 1000 volte ciascuna, ottenendo come valori medi e relative incertezze:

$$\bar{x} = 9.0000 \pm 0.0001 \quad e \quad \bar{y} = 0.9949 \pm 0.0002$$

Motivando le risposte: *a*) si scriva la formula generica da utilizzare per la stima della deviazione standard di Z attorno a (\bar{x}, \bar{y}) e se ne calcoli il valore; *b*) si ripeta con e in *a*) ma effettuando la propagazione attorno ad un generico punto (x_0, y_0) che si trova a meno di una deviazione standard da (\bar{x}, \bar{y}) ; *c*) si specifichi la discrepanza (sovrastima/sottostima) che si sarebbe commessa in *a*) trascurando il termine di covarianza

Esercizio 13.1.15 [17 giugno 2021 - modificato] Sia X, Y, Z variabili aleatorie continue in \mathcal{R} legate dalla relazione funzionale:

$$z = g(x, y) = e^{-Ax} \sin(x + y)$$

dove A ultimo numero diverso da zero della propria matricola. Le variabili X, Y hanno correlazione pari a $\text{COV}[X, Y] = -0.01$. X, Y sono state campionate ripetutamente rispettivamente 100 e 50 volte ciascuna, ottenendo come valori medi e relative incertezze:

$$\bar{x} = 0.500 \pm 0.001 \quad e \quad \bar{y} = \left(-\frac{1}{A} - 0.5\right) \pm 0.002$$

Motivando le risposte: *a*) si scriva la formula generica da utilizzare per la stima della deviazione standard di Z attorno a (\bar{x}, \bar{y}) e se ne calcoli il valore; *b*) si ripeta con e in *a*) ma effettuando la propagazione attorno ad un generico punto

$$(x_0, y_0) = \left(0.509, -\frac{1}{A} - 0.509\right)$$

Si discuta se sia il caso di trascurare il termine di covarianza e che implicazioni possa avere.

13.2 Seconda Parte Programma

Esercizi LKL. Massima Verosimiglianza

Esercizi INTERP. Interpolazione lineare

Esercizi CHI2. Test del χ^2

Esercizi STUD. Test di Student

Esercizi LR. Rapporto di verosimiglianze

13.3 Soluzioni

R: 13.1.2 a) la numerosità del campione è pari a $N = 3+8+\dots = 90$; b) il valore medio e' $\bar{a} = 2.0658$ s, la deviazione standard $s_t = 0.0282$ s, la deviazione standard della media $s_{\bar{t}} = s_t/\sqrt{N} = 0.0030$ s. Il valore medio e' quindi $\bar{a} = (2.066 \pm 0.003)$ s oppure $\bar{a} = (2.0658 \pm 0.0030)$ s con precisione $\epsilon = 0.0030/2.0658 = 0.0014 = 0.14\%$; c) Essendo $N = 90$ il secondo quantile si trova in corrispondenza del $N/10 \cdot 2 = 18$ -esimo valore, ovvero $q_{0.2} = 2.04$ s e l'ottavo $q_{0.8} = 2.08$ s; e) la compatibilità con il valore di riferimento va calcolata calcolando $t_1 g = t/2, s_{t_1} = s_t/2$ $r = \frac{|t_1 - t_R|}{\sqrt{s_{t_1}^2 + s_{t_R}^2}} \simeq 1.6$ e quindi

i due valori hanno buona compatibilità. f) l'intervallo di confidenza per tre deviazioni standard è $[\bar{t} - 3s_t, \bar{t} + 3s_t] = [1.9813, 2.1503]$ s, con nessun dato misurato al di fuori, cosa che ci aspettavamo visto che su 90 misure solo 0.27 circa sono aspettate al di fuori; g) il valore di riferimento per la classe $a = 2.06$ [m/s²] e' $\frac{Nw}{\sqrt{2\pi}s_t} e^{\frac{1}{2}\left(\frac{2.06-\bar{t}}{s_t}\right)^2} = 25$ dove $N = 90, w = 0.02$ s

R: 13.1.3 La numerosità del campione e' $N = 46$; La media aritmetica $\bar{L} = 45.1891$ mm; La deviazione della singola misura $s_L = 0.1303$ mm² da cui $s_{\bar{L}} = s_x\sqrt{N} = 0.0192$ mm. La moda, ovvero il valore più ricorrente è = 45.2 mm e la mediana é la mmedia tra la $N/2 = 23$ esima e la 24esima misura, comunque pari a = 45.2 mm. La miglior stima della grandezza e' quindi $L = (45.189 \pm 0.019)$ mm ovvero $L = (45.19 \pm 0.02)$ mm. Per la classe 45.20 il valore aspettato di conteggi é:

$$f^*(45.20) = \frac{Nw}{\sqrt{2\pi}s_L} e^{-\frac{1}{2}\left(\frac{45.20-\bar{L}}{s_L}\right)^2} = 9.81$$

da confrontare con i 5 misurati. $w = 0.1$ e' la larghezza della classe di frequenza. L'intervallo dei 3σ é $[\bar{x} - 2s_L, \bar{x} + 2s_L] = [44.7981, 45.5802]$ mm. Nessuna misura e' da scartare come ci aspettavamo dal fatto che al di fuori di 3σ ci aspettiamo $(1 - 99.7) \cdot 46 = 0.14$ misure. L'intervallo dei 2σ e' invece $[\bar{x} - 2s_L, \bar{x} + 2s_L]$. Dato ora $L_0 = (A.10 = 45) \pm 0.01$ mm calcoliamo

$$\Delta L = \bar{L} - L_0 \quad s_{\Delta L} = \sqrt{s_{\bar{L}}^2 + s_0^2}$$

. A quel punto era da calcolare

$$r = \frac{|\Delta L - 0.24|}{\sqrt{s_{\Delta L}^2 + 0.01^2}} \gg 3$$

e quindi con forte incompatibilità.

R: 13.1.1. a) Il numero minimo di passi per raggiungere metà e' 5: 1 per entrare e 4 per fare 24 passi; b) Per poter avanzare di 7 passi il giocatore deve essere entrato nel circuito. Si tratta quindi di un prodotto logico tra i due eventi aleatori 'E1=il giocatore e' entrato nel gioco' e 'E2=il giocatore ha fatto 7 passi'. Per fare 7 passi servono almeno due lanci, e quindi il giocatore deve entrare necessariamente al primo lancio. La probabilità di E1 e' 1/6 mentre l'evento E2 si puo' realizzare in 6 maniere (1+6,6+1,2+5,5+2,3+4,4+3) su 36, e quindi la probabilità composta é

$$P = \frac{1}{6} \cdot \frac{6}{36} = \frac{1}{36}$$

c) se al terzo lancio ha fatto 7 passi significa che al quarto lancio deve essere uscito la faccia '5', e quindi $P = 1/6$. Se la domanda veniva interpretata come la probabilità composta di entrambi gli eventi allora $P = 1/6 \cdot 1/36$ d) Gli eventi sono statisticamente indipendenti.

R: 13.1.4 Sia A =[sensore segnala anomalia], \bar{A} =[sensore non segnala anomalia], B =[termometro e' difettoso], \bar{B} =[termometro e' funzionante]. Il tema ci fornisce i seguenti dati: $P(A|B) = 0.99$ questa e' la probabilità che il sensore segnali anomalia se in effetti il termometro e' difettoso; $P(\bar{A}|\bar{B}) = 0.98$

²Alcune hanno considerato una deviazione standard basata su ipotesi uniforme, che non poteva essere usata in questo caso

questa e' la probabilita' che il sensore non segnali anomalia se in effetti il termometro e' funzionante; $P(B) = 0.3$ la probabilita' (a priori) che il termometro sia difettoso. Allora la domanda del problema e' $P(B|A)$ ovvero la probabilita' che una anomalia segnata provenga in effetti da un termometro difettoso. Si tratta di una inversione di probabilita' da fare attraverso formula di Bayes: $P(B|A) = \frac{P(AB)}{P(A)} = \frac{P(B)P(A|B)}{P(A)}$ di questo conosciamo gia' i termini al numeratore, mentre per il termine al denominatore possiamo usare, avendo una partizione totale dello spazio degli eventi, la formula della probabilita' totale: $P(A) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B}) = 0.99(0.3) + (1 - 0.98)(1 - 0.3) = 0.311$ e quindi: $P(B|A) = \frac{0.3(0.99)}{0.311} = 0.954$

R: 13.1.5 Nella prima strategia, il naufrago arriva alla costa se prende una direzione entro un angolo pari a $\theta_1 = 2 \arccos(10/60) \simeq 2.8 \text{ rad} \rightarrow P_1 = P(\theta < \theta_1) = \frac{\theta_1}{2\pi} = 0.446$ Nella seconda strategia, il naufrago arriva alla costa se prende una direzione entro due volte un angolo pari a $\theta_2 = 2 \arccos(10/20) \simeq 2.09 \text{ rad} \rightarrow P_2 = 2 P(\theta < \theta_2) = 2 \frac{\theta_2}{2\pi} = 0.667$ La strategia vincente e' al seconda

R: 13.1.6 Assumendo $N = 9$, la probabilita' che un carattere sia battuto male e' $p_1 = (A + 1)10^{-4} = 10^{-3}$. Diciamo $N = 180$ il numero di caratteri in una pagina. a) l'evento e' binomiale e quindi $P_1 = P_B(0; 1800; p_1) = \binom{1800}{0} p_1^0 (1 - p_1)^{1800} = 0.1652$ Si noti che questo e' anche l'evento composto in cui si ha sempre un carattere corretto: $P_1 = (1 - p_1)^N = 0.1652$ Inoltre, essendo l'evento raro p_1 piccola rispetto al numero di ripetizioni, allora l'evento Binomiale tende a quello Poissoniano di media $\lambda = p_1 N = 1.8$: $P_1 = P_P(0; 1.8) = \frac{1.8^0 e^{-1.8}}{0!} = 0.1653$ che si discosta di molto poco rispetto al valore vero Bernoulliano. b) In due pagina abbiamo $N_2 = 3600$ caratteri, allora analogamente a sopra $P_2 = P_B(1; 3600; p_1) = \binom{3600}{1} p_1^1 (1 - p_1)^{3599} = 0.0984$ ovvero, con approssimazione Poissoniana (e quindi con semplificazione dei calcoli) $P_2 = P_P(1; 3.6) = \frac{3.6^1 e^{-3.6}}{1!} = 0.0983$ c) La probabilita' di avere almeno un carattere sbagliato per pagina e' $p_{1+} = 1 - P_1 = 0.835$ e quindi su 50 pagine, il numero medio di pagina con almeno un carattere errato e' $N = 50 p_{1+} = 41.74$ pagine. Infine d) se il libro ha 50 pagine, la probabilita' che $A_4 = 36$ abbiano almeno un errore e': $P_3 = P_B(36; 50; p_{1+}) = \binom{50}{36} 0.035^{36} (1 - 0.835)^{50-36} = 0.016$

R: 13.1.7 Si tratta di un evento a media fissa, e quindi la distribuzione di riferimento e' Poissoniana con $\lambda = 2ev/w$. Per rispondere alle domande dobbiamo di volta in volta ricalcolare il valore di riferimento, ovvero a) in 10 giorni, allora $\lambda_{10} = 2(10)/7 = 2.857 \text{ ev}/10g$ e quindi $P_P(0; \lambda_{10}) = e^{-2.857} = 0.057 = 5.7\%$ c) in Aprile ci sono 30gg. Almeno un evento significa, quindi che ce ne puo' essere 1 oppure 2 o 3 etc. Allora per un giorno di aprile, la probabilita' che non piove e' $P = 1 - P_0$ dove la probabilita' di nessun evento piovoso in un giorno di Aprile e': $P_0 = P_P(0; \lambda_1) = e^{-0.2857} = 0.75$ e quindi il numero aspettato di giorni con almeno un evento piovoso e' $N = 30(1 - P_0) = 7.46 g$ Infine se $A = 3$ allora la probabilita' di avere esattamente $A + 3 = 6$ giorni di pioggia in Aprile e' $P_B(6; 30; 1 - P_0) = \binom{30}{6} 0.25^6 (1 - 0.25)^{24} = 0.15$

R: 13.1.8 Da fare

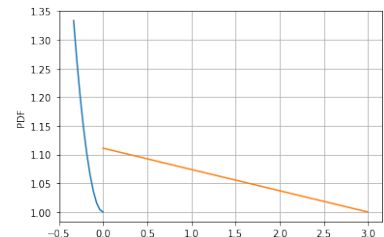
R: 13.1.9. Anzitutto procediamo ad un tentativo di graficare la funzione, che ci puo' aiutare:

La condizione affinché $f(x)$ sia PDF e' che sia positiva sul dominio e l'integrale sul dominio sia pari a 1. La prima condizione e' verificata se $k \geq 0$. Per la seconda, l'integrale per valori minori di 0 e'

$$\int_{-1/A}^0 f(x) dx = \left[x + a \frac{x^3}{3} \right]_{-1/A}^0 = \frac{3A + 1}{3A^2}$$

mentre per la parte a destra basta pensare che si tratta di un triangolo quindi l'integrale (l'area) e' pari a $Ak/2$ e quindi la condizione si puo' scrivere:

$$\frac{3A + 1}{3A^2} + \frac{Ak}{2} = 1$$



Il valore di k é positivo solo per $A > 1$. I valori di k sono riportati sotto. a) Per calcolare il valore di aspettazione bisogna svolgere l'integrale:

$$\mu = E[X] = \int_{-1/A}^A x f(x) dx = \int_{-1/A}^0 (x + Ax^3) dx + \int_0^A k(x - \frac{x^2}{A}) dx = -\frac{1}{2A^2} - \frac{1}{4A^3} + \frac{kA^2}{6}$$

I risultati sono riportati sotto. b) Per calcolare la deviazione standard bisogna calcolare la varianza

$$\text{VAR}[X] = E[X^2] - \mu^2$$

$$E[X^2] = \int_{-1/A}^A x^2 f(x) dx = \int_{-1/A}^0 (x^2 + Ax^4) dx + \int_0^A k(x^2 - \frac{x^3}{A}) dx$$

e poi la sua radice. Vedi sotto per i risultati. c) La funzione di distribuzione si calcola con

$$F(x) = \int_{-1/A}^x f(x') dx' = \begin{cases} \frac{Ax^3}{3} + x + \frac{3A+1}{3A^2} & \text{per } x \in [-\frac{1}{A}, 0] \\ F(0) - \frac{kx^2}{2A} + kx & \text{per } x \in [0, A] \end{cases}$$

d). Per la probabilità si poteva fare il conto esplicito

$$P(x \in [0, A/2]) = \int_0^{A/2} f(x) dx$$

oppure usare la $F(x)$ appena calcolata:

$$P(x \in [0, A/2]) = F[A/2] - F[0] = \frac{3}{8}AK$$

oppure pensare ancora alla forma triangolare della PDF per valori positivi e geometricamente trovare:

$$P(x \in [0, A/2]) = \frac{3}{4} \cdot \frac{AK}{2}$$

Raccogliamo i risultati per i diversi A :

A	2	3	4	5
k	0.417	0.420	0.365	0.315
μ	0.121	0.565	0.937	1.289
σ	0.563	0.800	1.036	1.272
P	0.312	0.472	0.547	0.590

R: 13.1.13 Per il calcolo delle probabilità si osserva come nel primo caso si chiede $P[-1\sigma, +1\sigma] = 0.683$ mentre nel secondo si chiede $P[-1\sigma, +2\sigma] \sim 0.819$. Non serviva quindi usare la tabella della funzione degli errori in quanto i primi margini di confidenza della Gaussiana sono noti. Ad ogni modo, attraverso la tabella il conto sarebbe stato, ricordando che $F_G(x) = \frac{1}{2} [1 - \text{erf}(\frac{x-\mu}{\sqrt{2}\sigma})]$

$$\begin{aligned} F(22) - F(18) &= \frac{1}{2} [\text{erf}(\frac{22-20}{\sqrt{2}}) - \text{erf}(\frac{18-20}{\sqrt{2}})] \\ &= \frac{1}{2} [\text{erf}(\frac{1}{\sqrt{2}}) - \text{erf}(\frac{-1}{\sqrt{2}})] = \frac{1}{2} [2\text{erf}(\frac{1}{\sqrt{2}})] = \text{erf}(0.707) \simeq 0.68 \end{aligned}$$

e analogamente nell'altro caso:

$$F(24) - F(18) = \frac{1}{2} [\text{erf}(\frac{2}{\sqrt{2}}) - \text{erf}(\frac{-1}{\sqrt{2}})] = \frac{1}{2} [0.95 + 0.7] \simeq 0.83$$

Si noti anche che si poteva 'traslare la Gaussiana' mettendo a zero la media e non cambiava il conto.

Per il calcolo della covarianza, che non dipende dal punto in cui e' calcolata, si procedeva con:

$$\text{COV}[Y, X] = \text{COV}[AX^2 + AX - A, X] = \text{COV}[AX^2, X] + \text{COV}[AX, X] - \text{COV}[A, X]$$

$$= A \text{COV}[X^2, X] + A \underbrace{\text{VAR}[X]}_{\sigma^2}$$

Il secondo addendo e' noto, mentre per il primo:

$$\text{COV}[X^2, X] = E[X^3] - \underbrace{E[X^2]}_{\text{VAR}[X]+E[X]^2} \underbrace{E[X]}_{\mu}$$

e quindi:

$$\text{COV}[Y, X] = A E[X^3] - A \mu (\sigma^2 + \mu^2) + A \sigma^2$$

Dove il terzo momento algebrico in realtà (non era richiesto) e' noto algebricamente per la Gaussiana e vale $E[X^3] = \mu^3 + 3\mu\sigma^2$ che si puo' verificare pensando che se X normale $Z = \frac{X-\mu}{\sigma}$ e' normale standard: $\mu = E[Z] = 0$, $\sigma^2 = \text{VAR}[Z] = 1$. Inoltre essendo simmetrica vale $E[Z^3] = 0$. Allora

$$E[X^3] = E[(\mu + \sigma Z)^3] = E[\mu^3 + 3\mu^2\sigma Z + 3\mu\sigma^2 Z^2 + Z^3] = \mu^3 + 3\mu\sigma^2$$

R: 13.1.14 La formula generica della propagazione delle incertezze: $s_{z_0}^2 \simeq \left(\frac{\partial g}{\partial x}\right)_{(x^*, y^*)}^2 s_{x_0}^2 + \left(\frac{\partial g}{\partial y}\right)_{(x^*, y^*)}^2 s_{y_0}^2 + 2 \left(\frac{\partial g}{\partial x} \frac{\partial g}{\partial y}\right)_{(x^*, y^*)} r_{XY} s_{x_0} s_{y_0}$ Per entrambe le domande, vanno calcolate le derivate parziali e valutate in (\bar{x}, \bar{y}) e quindi (uso $A = 9$): $\frac{\partial z}{\partial x} = A \sin(y) + \frac{y A^{1/x} \ln(A)}{x^2} |_{\bar{x}, \bar{y}} = 1.48$ e $\frac{\partial z}{\partial y} = x/\sqrt{1-y^2} + A^{1/x} |_{\bar{x}, \bar{y}} = 90.503$ Ora per quanto riguarda i due casi, nel primo, si tratta del valore medio e quindi $s_{x_0} = s_{\bar{x}} = 10^{-4}$, $s_{y_0} = s_{\bar{y}} = 2 \cdot 10^{-4}$, $r_{XY} s_{x_0} s_{y_0} = \text{COV}[X, Y]/1000 = 10^{-6}$ e quindi $s_z = 0.1810$. Siccome $z(\bar{x}, \bar{y}) = 14.5$ allora la nostra stima di $z = 14.5 \pm 0.2$. Nel caso della misura singola vicina al valore medio $s_{x_0} = s_x = 10^{-4} \sqrt{1000}$, $s_{y_0} = s_y = 2 \cdot 10^{-4} \sqrt{1000}$, $r_{XY} s_{x_0} s_{y_0} = \text{COV}[X, Y] = 10^{-3}$ allora in questo caso $s_z = 5.75$. Se lo avessimo calcolato trascurando la correlazione (per altro bassa) allora $s_z = 5.72$.

Capitolo 14

Raccolta di Esercizi

14.1 Prima Parte Programma

Es. 14.1.1 . Si consideri il seguente campione di misure ripetute di accelerazione a [m/s²] di una slitta su un piano inclinato di 30 gradi in cui n_i è la frequenza assoluta. Sia A ultima cifra non nulla del numero di matricola.

a [m/s ²]	n_i	a [m/s ²]	n_i
4.60	2	5.40	28
4.80	7	5.60	13
5.00	15	5.80	2
5.20	24	6.00	1

Definire e stimare rispettivamente: a) la numerosità del campione; b) il valor medio e relativa incertezza, assieme all'errore relativo standard di questo valore e all'errore della singola misura; c) il terzo e quinto decile; d) mediana e moda e) si studi la compatibilità di tale valor medio con il valore di riferimento per il singolo periodo $g = (9.806 \pm 0.001)$ [m/s²] e si dia una stima dell'eventuale discrepanza tra la misura e il valore di riferimento; f) si calcoli l'intervallo di confidenza centrato sul valor medio e di ampiezza 3 deviazioni standard, a cui applicare il criterio ragionato di reiezione dati; g) quanti conteggi di riferimento ci si aspettava per la classe $a = 5.20$ [m/s²] considerando la popolazione di misure?. Motivare le risposte.

Es. 14.1.2 . Si consideri un misurando W legato a 3 grandezze di input con la relazione:

$$W = g(x, y, z) = x^2 y + x e^z$$

Si misura un grado di correlazione campionario pari a $r_{XY} = -0.3$, $r_{YZ} = 0$, $r_{XZ} = 0.95$. Le tre grandezze di input sono campionate $N = 15$ volte ottenendo i valori:

$$\bar{x} = 1.01 \pm 0.09, \quad \bar{y} = 3.17 \pm 0.16, \quad \bar{z} = 2.11 \pm 0.56$$

a) Calcolare la miglior stima di W a partire dalle misure, b) scrivere la formula implicita per il calcolo della incertezza su W , c) calcolare l'incertezza del valore di W con e senza i termini di correlazione, discutendo il risultato d) riportare la stima con incertezza su W considerando che X sono valori numerici di riferimento e y, e^z e' misurata in Hz., e) riportare le precisioni delle diverse grandezze.

Es. 14.1.3 . Si calcoli la probabilità di trovare un valore compreso in $[-1.55, -0.85]$ in una Gaussiana di media 0 e varianza 4. Si consideri la tabella della funzione di errore di [Figura 5.8](#).

Es. 14.1.4 . Su un tavolo ci sono due dadi. Uno dei due e' truccato, con una probabilità che esca 6 del 22%. Nel lancio di un dado singolo a). Qual e' la probabilità che esca 5? b). Se esce 5, qual e' la probabilità che il dado sia quello truccato?

Es. 14.1.5 . Data la distribuzione di probabilità triangolare asimmetrica

$$\tau(x; 0, b) = \begin{cases} 0 & \text{for } x < 0, \\ \frac{2x}{bc} & \text{for } 0 \leq x < c, \\ \frac{2(b-x)}{b(b-c)} & \text{for } c < x \leq b, \\ 0 & \text{for } b < x. \end{cases}$$

con $0 \leq x \leq b$, si calcoli a) valore di aspettazione, b) moda, c) varianza, d) CDF e) quantile corrispondente a $x = c$.

Es. 14.1.6 . [G. Marchetti] Su 4 famiglie di 2 componenti misuriamo il reddito di Febbraio R e le relative spese per l'alimentazione S . Calcolare la covarianza e il coefficiente di correlazione e discuterli. Rappresentate la relazione tra spesa e reddito con la retta dei minimi quadrati e verificate che questa è: $S = -835 + 0.7R$.

Reddito R	1500	1700	1400	1600
Spese S	200	350	150	300

Es. 14.1.7 . Si consideri uno strumento composto da un plexiglass con trasmissività $\tau = 0.9$ e un fotomoltiplicatore con efficienza $\epsilon = 0.42$. Si immette un fascio formato da 100 fotoni. A) Quale è la migliore stima del numero di fotoni rivelati dal fotomoltiplicatore B) qual' è la incertezza su una singola misura di questo valore? C) quante misure deve ripetere con il fascio da 100 fotoni per poter misurare la precisione del numero di fotoni osservati al di sotto del 10%?

Es. 14.1.8 . Si misura l'efficienza di rivelazione di fotoni in un sensore di luce, ottenendo il valore medio $\epsilon = 0.25$ seguito di 20 misure ripetute nelle stesse condizioni.

4.1) Riportare la stima di efficienza assieme alla sua incertezza

4.2) Quante volte andrebbe ripetuta la misura per portare l'incertezza al di sotto del 10%?

Es. 14.1.9 . In una fabbrica di cioccolatini, si ispezionano due produzioni I e II di cioccolatini fatti prima (I) e dopo (II) un intervento alla macchina che li produce. Il tecnico vuole capire se il peso dei cioccolatini è cambiato, e prende una serie di misure ripetute di N cioccolati estratti a caso tra la produzione II , ottenendo la distribuzione in tabella. Per la produzione I , aveva in precedenza misurato $M_I = 15.014 \pm 0.004$ g. Per la produzione II :

- 1.1 Stimare media aritmetica, mediana, moda e deviazione standard;
- 1.2 Riportare nella forma corretta la miglior stima del valore vero dalle misure ripetute e riportare la precisione finale della stima;
- 1.3 Con il criterio dei 3-sigma, verificare la presenza o assenza di dati anomali. In ogni caso, non procedere al ricalcolo dei momenti della distribuzione di misure.
- 1.4 Disegnare l'istogramma dei dati e scrivere la formula analitica della distribuzione gaussiana che si può associare alla serie di misure;
- 1.5 Stimare la compatibilità con il valore M_{II}

Massa [g]	Frequenza assoluta
15.10	2
15.11	9
15.12	13
15.13	7
15.14	3
15.16	1

Es. 14.1.10 . Si riporta in tabella le classi di frequenza di un campione di misure. (Sia A l'ultimo numero di matricola.)

x [s]	f_x
1.00/A	4
1.01/A	18
1.02/A	22
1.03/A	36
1.04/A	28
1.05/A	11
1.06/A	2

Motivando numericamente e a parole in modo esaustivo e ordinato:

1. Si calcoli la dimensione del campione N
2. Si calcoli la deviazione standard campionaria s_x
3. Si calcoli la deviazione standard della media $s_{\bar{x}}$
4. Si riporti la stima del valore medio e incertezza della serie X
5. Si discuta la precisione delle misure
6. Utilizzando il criterio dei 3σ si discuta la presenza di outliers e una loro eventuale rimozione
7. Si studi la compatibilità della serie X con una serie Y ottenute misurando due volte il tempo X che ha portato come risultato $\bar{y} = 2,048/A \pm 0.010$. La numerosità di Y è simile a quella di X .
8. Si tracci un istogramma dei dati in tabella e si riporti la formula della funzione normale che meglio approssima i dati

14.2 Soluzioni

R: 14.1.1 a) la numerosità del campione è pari a $N = 2 + 7 + \dots = 92$; b) il valore medio e' $\bar{a} = 5.2630$ [m/s²], la deviazione standard $s_a = 0.2700$ [m/s²], la deviazione standard della media $s_{\bar{a}} = s_a/\sqrt{N} = 0.0282$ [m/s²]. Il valore medio e' quindi $\bar{a} = (5.26 \pm 0.03)$ [m/s²] con precisione $\epsilon = 0.0282/5.26 = 0.53\%$; c) Essendo $N = 92$ il terzo quantile si trova in corrispondenza del $N/10 \cdot 3 = 27.6 \simeq 28$ -esimo valore, ovvero $q_{0.3} = 5.20$ e il quinto $q_{0.5} = 5.20$ [m/s²]; d) il valore mediano e' $q_{0.5} = 5.20$ [m/s²] e la moda e' il valore piu' frequente $a = 5.08$ [m/s²]; e) la compatibilità con il valore di riferimento va calcolata calcolando $g = a/\sin(30) = 2a$, $s_g = 2s_a$ $r = \frac{|g-G|}{\sqrt{s_g^2+s_G^2}} \simeq 12.9$ e quindi i due valori sono decisamente incompatibili. La discrepanza tra la misura e' il valore di riferimento è $|g - G| = 0.72$ [m/s²]; f) l'intervallo di confidenza per tre deviazioni standard è $[\bar{a} - 3s_a, \bar{a} + 3s_a] = [4.4530, 6.0731]$ [m/s²], con nessun dato misurato al di fuori, cosa che ci aspettavamo visto che su 92 misure solo 0.3 circa sono aspettate al di fuori; g) il valore di riferimento per la classe $a = 5.06$ [m/s²] e' $\frac{Nw}{\sqrt{2\pi}s_a} e^{\frac{1}{2}\left(\frac{5.2-\bar{a}}{s_a}\right)^2} = 26.05$

R: 14.1.2 a) La miglior stima di W si ottiene calcolando valutando W sui valori misurati di X, Y, Z ottenendo: $w = 11.6$. b) la formula per l'incertezza su W scritta in forma implicita è:

$$s_{w_0}^2 \simeq \left(\frac{\partial w}{\partial x}\right)_{(***)}^2 s_{x_0}^2 + \left(\frac{\partial w}{\partial y}\right)_{(***)}^2 s_{y_0}^2 + \left(\frac{\partial w}{\partial z}\right)_{(***)}^2 s_{z_0}^2 + 2 \left(\frac{\partial w}{\partial x} \frac{\partial w}{\partial y}\right)_{(***)} r_{XY} s_{x_0} s_{y_0} + 2 \left(\frac{\partial w}{\partial x} \frac{\partial w}{\partial z}\right)_{(***)} r_{XZ} s_{x_0} s_{z_0} + 2 \left(\frac{\partial w}{\partial y} \frac{\partial w}{\partial z}\right)_{(***)} r_{YZ} s_{y_0} s_{z_0}$$

dove $*** = (\bar{x}, \bar{y}, \bar{z})$ sono le miglior stime dei valori veri delle rispettive variabili, ovvero le medie aritmetiche, e $s_{x_0}^2, \dots$ sono le miglior stime delle incertezze attorno ai punti medi, ovvero le SEM. c) Per fare il calcolo esplicito serve calcolare le derivate parziali prime e valutarle nei punti medi: $\partial W/\partial x(\bar{x}, \bar{y}, \bar{z}) = 2xy + e^z = 14.734$; $\partial W/\partial y(\bar{x}, \bar{y}, \bar{z}) = x^2 = 1.02$; $\partial W/\partial z(\bar{x}, \bar{y}, \bar{z}) = xe^z = 8.33$. Allora, considerando i vari termini di correlazione: $s_w = 5.17$ mentre senza il termine di correlazione $s_w = 4.85$. d) la miglior stima di W alla luce dei risultati é $W = (11.6 \pm 4.9)$ Hz. e) Le precisioni delle misure sono rispettivamente $x : 9\%$, $y = 5\%$, $z : 27\%$ e $w : 45\%$.

R: 14.1.3 La CDF della Guassiana vale $F(x) = 0.5[1 + \operatorname{erf}[(x - \mu)/(\sqrt{2}\sigma)]$. E' simmetrica rispetto a $x - \mu$. Allora la risposta alla domanda e' $P = 0.5(\operatorname{erf}(0.55) - \operatorname{erf}(0.3)) = 0.116 = 11.6\%$

R: 14.1.4 Sia D_i l'evento aleatorio che si sia usato il dato 1 o 2 e si assuma che il secondo dado sia truccato. Allora, la probabilita' che esca un numero diverso da 6 sul dado D_2 e' $P(5|D2) = (1 - 0.22)/5 = 0.156$. Allora a) $P(5)$ si calcola usando la legge della probabilita' totale ovvero $P(5) = P(5|D1)P(D1) + P(5|D2)P(D2) = (1/6)(1/2) + 0.156(0.5) = 0.161$. b) Dobbiamo calcolare $P(D2|5)$ e usiamo la formula di Bayes: $P(D2|5) = P(D2)P(5|D2)/P(5) = 0.483$

R: 14.1.5 a) Il valore di aspettazione e' si calcola con

$$E(\tau) = \int_0^b x \sqcup(x) dx = \int_0^c \frac{2}{bc} x^2 dx + \int_c^b \frac{2}{(b-c)} x dx + \int_c^b \frac{2}{(b(b-c))} x^2 dx = \dots = \frac{b+c}{3}$$

; b) la moda e' il massimo della funzione, che sappiamo essere raggiunto in c e quindi vale $2/b$;

R 14.1.6 Calcoliamo anzitutto le media $\bar{R} = 1550, \bar{S} = 250$ e le varianze $s_R^2 = 12500, s_S^2 = 6250$. Allora la covarianza e' la media dei 4 prodotti: $(R_i - \bar{R})(S_i - \bar{S})$ danno 2500, 15000, 15000, 2500. Somma : 35000 e quindi $\operatorname{COV}(R, S) = s_{RS} = 35000/4 = 8750$. Per il calcolo della covarianza potevo anche usare $\operatorname{COV}(R, S) = \mu_{RS} - \bar{R}\bar{S} = (1500 \cdot 200 + 1700 \cdot 350 + 1400 \cdot 150 + 160 \cdot 300)/4 - 1550 \cdot 250 = 8750$. Allora il coefficiente di correlazione lineare di Pearson: $r_{RS} = 8750/\sqrt{12500 \cdot 6250} = 0.9899$. Il coefficiente angolare della retta interpolante si calcola con $b = \frac{s_{RS}}{s_R^2} = 8750/12500 = 0.7$. Intercetta e' $a = \bar{S} - b\bar{R} = 250 - 0.7 \cdot 1550 = -835$ e quindi la retta ha equazione: $S = -835 + 0.7R$

R 14.1.7 I processi di trasmissione del fotone sul plexiglass e sul fotomoltiplicatore seguono entrambi una distribuzione statistica di tipo Binomiale, della forma $p_X(x) = [p, 1 - p]$ a seconda che x avvenga o non avvenga. L'aspettazione e' $E[x] = \mu = p$ e la varianza e' $\sigma^2 = p(1 - p)$.

A) Essendo i due processi di attraverso del plexiglass e del fotomoltiplicatore indipendenti, a partire da $n_0 = 100$ fotoni, il numero aspettato di fotoni rilevati dallo strumento e': $n = n_0 \cdot \tau \cdot \epsilon = 37.8$ fotoni.

B) L'incertezza sulla singola misura e': $s_n = n \sqrt{\frac{s_n^2}{\tau^2} + \frac{s_n^2}{\epsilon^2}} = 45.85$ fotoni. C) Se volessimo ottenere una precisione del 10%, ovvero una errore standard di al massimo 3.78 dovremmo ripetere la misura N volte con: $s_{\bar{n}} < \frac{s_n}{\sqrt{N}} = 3.78 \rightarrow N > 147$ misure

R 14.1.8 La rivelazione di un fotone e' un fenomeno che puo' avvenire (con probabilita' p) o non avvenire con probabilita' $1 - p$. Segue quindi la statistica di Bernoulli. L'aspettazione di Bernoulli e' p e la varianza e' $p(1 - p)$. La stima migliore dell'aspettazione nel caso di misure ripetute e' la media aritmetica, che viene data dal problema, quindi $p = 0.25$, la varianza vale quindi $p(1 - p) = \sqrt{0.25(0.75)} = 0.43$. Avendo ripetuto la misura 20 volte, la stima migliore di incertezza e': $s_{\mu_\epsilon} = \frac{s}{\sqrt{N}} = 0.25 \pm \frac{0.43}{\sqrt{20}} = 0.096 \sim 0.1$ L'incertezza e' del $0.1/0.25 = 0.4 = 40\%$. Per portare l'incertezza al di sotto del 10%, assumendo ancora $\bar{\epsilon} = 0.25$ significa $\frac{\sigma/\sqrt{N}}{\bar{\epsilon}} < 0.1 \rightarrow N > \left(\frac{\sigma}{0.1\bar{\epsilon}}\right)^2 = \left(\frac{0.43}{0.1(0.25)}\right)^2 = 296$

Bibliografia

- P. R. Bevington and D. K. Robinson. *Data reduction and Error Analysis*. McGraw Hill, 2003. URL <http://experimentationlab.berkeley.edu/sites/default/files/pdfs/Bevington.pdf>.
- G. Boffetta and A. Volpiani. *Probabilita' in Fisica*. Springer, 2012. URL <https://link.springer.com/book/10.1007/978-88-470-2430-4>.
- G. Bohm and G. Zech. *Introduction to Statistics and Data Analysis for Physicists*. Free distribution, 2010. URL https://www-library.desy.de/preparch/books/vstatmp_engl.pdf.
- V. Canale and M. Delle Pietra. *Il laboratorio di Fisica I*. Edises, 2020. URL https://www.edisesuniversita.it/area_scientifica/canale-laboratorio-di-fisica-i.html.
- G. Ciullo. *Introduzione al Laboratorio di Fisica*. springer, 2014. URL <https://link.springer.com/book/10.1007/978-88-470-5656-5>.
- G. D'Agostini. Asymmetric uncertainties: Sources, treatment and potential dangers. 3 2004.
-]dagostini:2003 Giulio D'Agostini. *Bayesian reasoning in data analysis : a critical introduction / Giulio D'Agostini*. World scientific, New Jersey [etc, [2003]. ISBN 9812383565.
- L. Lista. *Statistical Methods for Data Analysis in Particle Physics*. Springer, 2016. URL <https://link.springer.com/book/10.1007/978-3-319-20176-4>.
- M. Loreti. *Teoria degli errori e fondamenti di statistica*. Free distribution, 2008. URL <https://wwwcdf.pd.infn.it/labo/INDEX.html>.
- Jeffrey D. Scargle, Jay P. Norris, Brad Jackson, and James Chiang. STUDIES IN ASTRONOMICAL TIME SERIES ANALYSIS. VI. BAYESIAN BLOCK REPRESENTATIONS. 764(2):167, feb 2013. doi: 10.1088/0004-637x/764/2/167. URL <https://doi.org/10.1088/0004-637x/764/2/167>.