

DISPENSE

Statistica

Probabilità, Laboratorio e Teoria degli Errori

Francesco Marrocco
Riccardo Mordente
Alice Pellegrino
Daniele Nicastro

Layout a cura di TecoSaur

Versione 1.5.0

28 settembre 2023

Indice

5	Capitolo 1 Probabilità
19	Capitolo 2 Teoria degli Errori
45	Capitolo 3 Distribuzioni
84	Capitolo 4 Inferenza
115	Capitolo 5 Altri Concetti
127	Capitolo 6 Test d'ipotesi
133	Capitolo 7 L ^A T _E X
157	Capitolo 8 Appendice

Prefazione

Agli studenti del corso di laboratorio di meccanica (o a chiunque entri in possesso di questo file),

Queste dispense nascono con lo scopo di aiutare un gruppo di matricole a organizzare i concetti chiave e i calcoli necessari per l'esame orale di questa materia, grande scoglio del primo anno della laurea triennale, in modo da avere tutto organizzato secondo un ordine che vada leggermente oltre le "pagine del quaderno sparse per la cameretta". L'utilità individuale che abbiamo trovato in questa organizzazione ci ha fatto poi riflettere su come questo pdf possa essere d'aiuto ad altri nostri colleghi, e siccome siamo convinti dell'idea che l'università debba essere un luogo di massima condivisione del sapere e aiuto reciproco, ci siamo decisi a mandare il nostro lavoro a chi ce lo chiedeva, con un risultato molto positivo (almeno a detta loro).

Sia chiaro, queste pagine sono anni luce lontane dal volere o potere sostituire eventuali libri di testo o le dispense dei professori (che trovate a questo [link](#), e da cui abbiamo ripreso moltissime cose) ma cercano piuttosto di essere un buono strumento di sintesi, da utilizzare dopo aver assorbito le nozioni del corso da un punto di vista concettuale, per fare mente locale e non perdersi tra i tantissimi argomenti affrontati.

Parte fondamentale del file sono i calcoli, spesso lunghi e dispersivi, che è necessario affrontare e saper fare per dimostrare molti aspetti fondamentali (basti pensare alla linearizzazione, ai fit lineari...) e che in modo altrettanto frequente vengono *lasciati al lettore per esercizio*. Ora è bene chiarire che non abbiamo nessun problema con questa frase, siccome farsi due conti non fa mai male a nessuno, e finché non ci si impasta bene le mani con alcune cose non le si comprende a fondo (inoltre se avete passato analisi si presume che li sappiate fare) ma purtroppo capita che non si riesce proprio a risolvere un integrale o una serie, o che la si sa fare ma durante il ripasso molto vasto si perde la memoria di quello che si è fatto, si mischiano i conti, ci si fa prendere dall'ansia o altre mille fattori, e siccome a Fisica ci insegnano a renderci la vita facile, abbiamo deciso di dedicare quanta più attenzione possibile ai calcoli espliciti, con molti passaggi (anche superflui) e metodi per la comprensione facile di alcune dimostrazioni, in modo da non perdere il filo.

Sperando che le nostre dispense possano esservi d'aiuto per lo studio dell'orale di laboratorio di meccanica, vi auguriamo una buona lettura.

Alice, Francesco, Riccardo e Daniele

Per qualsiasi errore nelle dispense, dubbi o altro potete contattarci alla mail riccardo.mordente@gmail.com, francescomatteo03@gmail.com. Il file è graficamente ispirato al layout gratuito a cura di Tecosaur, via <https://github.com/tecosaur/BMC>.

1

Probabilità



Sommario

- 1.1 Introduzione alla Probabilità, 6
- 1.2 Probabilità e scommesse eque, 8
- 1.3 Quote di Scommessa, 9
- 1.4 Classe completa, 11
- 1.5 Probabilità dell'unione di Eventi, 11
- 1.6 Leggi e cose utili, 12
- 1.7 Teorema di Bayes, 14
- 1.8 Paradosso di Monty-Hall, 16
- 1.9 Riassunto, 18

1.1 Introduzione alla Probabilità

In questa primissima sezione analizzeremo dei fondamenti di probabilità che ci aiuteranno nel resto del corso. Il primo tema che vogliamo proporvi sono gli assiomi di Kolmogorov, dei fondamenti per la definizione stessa del concetto di probabilità e della sua nomenclatura.

In breve, affermano che una probabilità è una funzione che assegna un valore tra 0 e 1 ad ogni evento in uno spazio campionario. Lo spazio campionario è l'insieme di tutti i possibili esiti di un esperimento e gli eventi sono sottoinsiemi dell'insieme degli esiti.

Gli Assiomi di Kolmogorov stabiliscono, quindi, che la probabilità deve soddisfare tre proprietà fondamentali:

- La probabilità di ogni evento è non negativa: la probabilità di ogni evento in uno spazio campionario è sempre maggiore o uguale a zero.
- La probabilità dell'intero spazio campionario è uguale a 1: la probabilità che si verifichi uno degli esiti possibili in uno spazio campionario è sempre pari a 1.
- La probabilità di una unione di eventi è la somma delle probabilità dei singoli eventi: se due eventi sono mutuamente esclusivi (non possono verificarsi contemporaneamente), la probabilità dell'unione di questi eventi è la somma delle probabilità dei singoli eventi.

Adesso però riscriviamoli in matematica:

Definizione 1

Dato un evento E_i , chiamiamo $P(E_i)$ la sua probabilità assoluta, ovvero non condizionata da uno stato di informazione. La probabilità condizionata, invece, la indichiamo con $P(E_i|I_i)$.

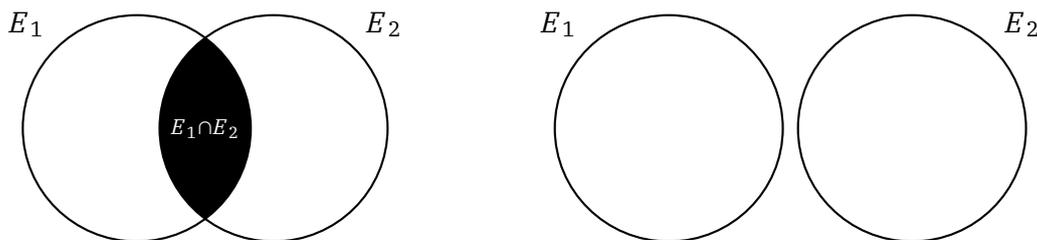
Chiamiamo poi Ω uno spazio campionario, ovvero l'insieme dei possibili outcome di un esperimento, vale quindi che $\Omega = \{P_i\}$. Infine, usiamo i simboli dell'insiemistica per definire la probabilità dell'unione o dell'esclusione di più eventi o spazi campionari, per esempio la probabilità che avvengano due eventi contemporaneamente è $P(E_1 \cup E_2)$.

Ora possiamo riscrivere gli assiomi in una modalità formale, in modo da poter maneggiare problemi di probabilità senza dover scrivere un libro intero (cosa che noi stiamo facendo).

π Teorema 1: Assiomi di Kolmogorov

- $0 \leq P(E|I_s) \leq 1$
- $P(\Omega) = 1$
- $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$

È bene notare come il terzo assioma, che può sembrare teatro di confusione, non ci dà la probabilità che due eventi accadano insieme (anche perchè sopra li abbiamo definiti come mutuamente esclusivi, quindi o accade uno o l'altro) bensì la probabilità che accada l'uno oppure l'altro. È come dire "prendo un insieme in cui sono contenuti entrambi gli eventi, e mi chiedo la probabilità di questo insieme in cui ci sono le probabilità dei due eventi", appare chiaro quindi perchè farne la somma. Se volessi la probabilità che i due eventi accadano insieme mi troverei $P(E_1 \cap E_2)$.



Da qui infatti possiamo definire un'altra regola della probabilità, ovvero la probabilità dell'unione di due eventi non mutuamente esclusivi. Essa segue l'assioma di Kolmogorov, ma dobbiamo sottrarre la probabilità che i due eventi accadano insieme.

π Teorema 2: Legge della Somma

Dati due eventi E_1, E_2 , in cui $P(E_1), P(E_2)$ sono le probabilità associate ai due eventi e in cui $P(E_1 \cap E_2) \neq 0$, ovvero non sono mutuamente esclusivi e possono accadere insieme, allora la probabilità dell'unione dei due eventi vale:

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2) \quad (1.1)$$

1.2 Probabilità e scommesse eque

Avendo chiarito che una probabilità è da ritenersi sempre come probabilità condizionata (perché nel mondo reale funziona così), ci resta ancora da trovare una regola generale e operativa per valutarla, ovvero dare una “definizione” che vada oltre il concetto intuitivo e che non sia limitata ai casi stereotipati degli eventi equiprobabili o delle prove effettuate in condizioni di equiprobabilità.

Un modo di superare l'impasse è quello di far ricorso al concetto di scommessa, percepibile a livello intuitivo da tutte le persone razionali. Finora, quando si è fatto uso del termine scommessa al fine di provocare un giudizio di maggiore probabilità fra più eventi alternativi, si è sempre indicata, o lasciata immaginare, una scommessa alla pari con l'intento di vincere. Ad esempio, considerando il lancio di un dado, dovendo scommettere pro o contro il “6”, si è d'accordo che conviene puntare contro. Il solo problema è ...trovare la persona che accetti la scommessa. Se invece si propone che chi gioca sul “6” punta mille lire e chi contro 99 mila lire si capisce come il problema si rovesci.

In entrambi i casi le scommesse hanno un verso privilegiato. Mentre nel primo caso nessuno era disposto a puntare 1000 lire per vincerne 2000, ora tutti sono favorevoli a giocare la stessa cifra per vincere 100 mila. Chiaramente il segreto non sta nell'importo superiore, perché nessuno punterebbe 50 mila sul sei per vincerne 100 mila. Così nessuno gioca alla pari contro la vittoria in casa della capolista del campionato di calcio nella partita contro l'ultima in classifica, ma potrebbe accettare la scommessa per vincere 10, 100 o 1000 volte la puntata.

Questo ci insegna che la puntata massima A che si è disposti a scommettere per ricevere una somma S se l'evento E si verifica e niente se non si verifica (eventualmente con le dovute condizioni) è proporzionale a quanto si crede che l'evento possa accadere, ossia:

$$A \propto P(E)$$

Ma lo è anche all'importo S che si può vincere

$$A \propto S$$

E allora mettendo insieme le due cose otteniamo la costante di proporzionalità:

$$A = P(E) \cdot S \tag{1.2}$$

Se si ritiene che l'evento sia praticamente impossibile non si è disposti a puntare niente. Se si è assolutamente sicuri si arriva a puntare S , il che è chiaro: chi non punterebbe tutti i soldi su una puntata certa al 100%? Forse qualcuno allergico ai soldi.

1.3 Quote di Scommessa

Talvolta la credibilità di un evento è espressa dal rapporto delle puntate che si ritiene equo scommettere pro o contro. Dalla conoscenza dei rapporti delle puntate è possibile risalire al valore di probabilità stimato. Chiamiamo ad esempio B la puntata che si è disposti a scommettere su \bar{E} , ovvero l'evento che si verifica se non si verifica E : è equivalente a dire la probabilità che non si verifichi E . Si deduce chiaramente, a partire dall'Equazione 1.2 che:

$$B = P(\bar{E}) \cdot S \quad (1.3)$$

Sommando membro a membro, sempre a partire dalla stessa, otteniamo che:

$$A + B = (P(E) + P(\bar{E}))S \quad (1.4)$$

Ricordandoci che la posta S è data dalla somma delle puntate A e B , si ottiene che:

$$P(E) + P(\bar{E}) = 1$$

Questo in realtà è abbastanza intuitivo senza ricorrere a formule: se $P(E)$ è la probabilità che l'evento accada e $P(\bar{E})$ che non accada, è chiaro che per i fondamenti di logica l'evento accadrà oppure no, quindi l'unione delle due probabilità darà un evento certo. Bisogna chiarire però che ciò è valido solo se le probabilità si riferiscono allo stesso stato d'informazione:

π Teorema 3

Dato un evento e chiamato E l'outcome in cui questo evento si verifica, e \bar{E} quello in cui non si verifica, allora vale che

$$P(E|I) + P(\bar{E}|I) = 1 \quad (1.5)$$

Mettendo adesso a rapporto l'Equazione 1.4, con l'Equazione 1.3 si ottiene:

$$\frac{A}{B} = \frac{P(E)}{P(\bar{E})} \quad (1.6)$$

Da cui arriviamo a dire che $P(E) = \frac{A}{B} \cdot P(\bar{E}) = \frac{A}{B} \cdot \frac{B}{S}$ ovvero:

$$P(E) = \frac{A}{A + B} \quad (1.7)$$

Siccome non sempre le scommesse sono eque, è interessante chiamare speranza matematica (di vincita), o valore atteso (di vincita), la somma che si spera di vincere per la probabilità di vincita. Chiamiamo questa quantità $M (= P(E) \cdot S)$, per distinguerla dalla

generica posta A che chi organizza il gioco può richiedere di pagare. Se A è uguale a M il gioco è equo. Quindi il rapporto A/M caratterizza l'equità del gioco. La quantità $M = P(E) \cdot S$ può essere riscritta nel seguente metodo:

$$M = P(E) \cdot S + (1 - P(E)) \cdot 0 = P(E) \cdot S + P(\bar{E}) \cdot 0 \quad (1.8)$$

Mettendo in luce che essa è ottenuta dalla somma degli importi che si possono vincere se si verifica un certo evento, ciascuno moltiplicato per la sua probabilità. Se si eseguono contemporaneamente n scommesse su tanti eventi E_1, E_2, \dots, E_N , la speranza matematica è la somma delle speranze matematiche:

$$M = \sum_{i=0}^n P(E_i) \cdot S_i \quad (1.9)$$

1.4 Classe completa

Adesso introduciamo il concetto di **classe completa di eventi**: ovvero eventi tali che essi siano a due a due mutuamente esclusivi e tali che la loro somma logica costituisca l'evento certo:

Definizione 2

Una classe di eventi $E = \{E_1, E_2, \dots, E_n\}$ si dice una classe completa di eventi se valgono le relazioni:

$$\begin{aligned} \bigcup_{i=1}^n E_i &= \Omega \\ E_i \cap E_j &= \emptyset \end{aligned}$$

Se io volessi trovare la probabilità di un evento H , interno alla classe completa posso utilizzare le proprietà di quest'ultima:

$$P(H) = P(H \cap \Omega) = P\left[H \cap \left(\bigcup_{i=1}^n E_i\right)\right] = P\left[\bigcup_{i=1}^n (H \cap E_i)\right] \longrightarrow \sum_{i=1}^n P(H \cap E_i) \quad (1.10)$$

1.5 Probabilità dell'unione di Eventi

Dobbiamo partire dal considerare uno spazio arbitrario che chiamiamo S , che potrebbe contenere o un numero finito o infinito di elementi. In seguito svilupperemo delle proprietà generali riguardo la probabilità. In questa sezione in particolare studieremo il caso particolare dell'unione di N eventi $A_1 \dots A_N$. Se gli eventi $A_1 \dots A_N$ sono disgiunti, allora sappiamo che:

$$Pr\left(\bigcup_{i=1}^N A_i\right) = \sum_{i=1}^N Pr(A_i) \quad (1.11)$$

1.6 Leggi e cose utili

Mano a mano che continuiamo la nostra trattazione ci stiamo avvicinando sempre di più all'argomento che è ovvero il teorema di Bayes, dunque prima di introdurlo facciamo un excursus per citare alcune leggi utili che si utilizzano per la dimostrazione del teorema, senza preoccuparci troppo della base teorica, ricordandoci che il nostro obiettivo è la sintesi concettuale e matematica degli argomenti, quindi se non state capendo un granchè significa che non avete seguito bene il corso prima di avventarvi nel ripasso, o che noi scriviamo da cani. Entrambi gli eventi hanno una probabilità abbastanza alta.

1.6.1 Legge della Moltiplicazione

Se volessi calcolare la probabilità di N eventi tale che $P(E_1 \cap E_2 \dots \cap E_N)$ ottengo la **Legge della Moltiplicazione**:

$$P(E_1 \cap E_2 \dots \cap E_N) = P(E_1) \cdot P(E_2|E_1) \cdot P(E_3|E_1 \cap E_2) \dots P(E_N|E_1 \cap E_2 \cap \dots \cap E_{N-1})$$

π Teorema 4: Legge della Moltiplicazione

Dati N eventi $E_1 \dots E_N$, la probabilità dell'occorrenza di tutti gli eventi è

$$P\left(\bigcap_{i=1}^N E_i\right) = P(E_1) \cdot P(E_2|E_1) \cdot \dots \cdot P\left(E_N \mid \bigcap_{i=1}^{N-1} E_i\right) \quad (1.12)$$

Se gli eventi sono indipendenti tra loro, invece, la relazione diventa:

$$P\left(\bigcap_{i=1}^N E_i\right) = \prod_{i=1}^N P(E_i) \quad (1.13)$$

1.6.2 Legge della probabilità condizionata

La legge sulla probabilità condizionata è il primo tassello per il calcolo di probabilità non assolute (abbiamo visto concettualmente, ma non numericamente cosa significhi probabilità condizionata), ed è anche estremamente importante per il Teorema di Bayes.

π Teorema 5: Legge della probabilità condizionata

Siano A, B due eventi, allora le loro probabilità condizionate sono:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}; \quad (1.14)$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)}; \quad (1.15)$$

1.6.3 Legge delle alternative

La legge delle alternative è un principio fondamentale della probabilità condizionata che descrive come calcolare la probabilità assoluta di un evento, il quale accade influenzato da più condizioni. Questa si costruisce a partire dall'Equazione 1.10 vista in precedenza, unita alla formula della probabilità condizionata

π Teorema 6: Legge della Alternarive

Sia E un evento visto come somma dei prodotti logici tra E e tutti gli elementi di una classe completa, ovvero:

$$P(E) = \sum_{i=1}^n P(E \cap H_i) \text{ se } \begin{cases} H_i \cap H_j = \emptyset \forall i \neq j \\ \bigcup_{i=1}^n H_i = \Omega \end{cases} \quad (1.16)$$

Allora, per la legge della probabilità condizionata, posso esprimere l'evento come:

$$P(E) = \sum_{i=1}^n P(E|H_i) \cdot P(H_i) \quad (1.17)$$

1.7 Teorema di Bayes

Ed eccoci arrivati al momento più importante di questo capitolo: il Teorema di Bayes. "Perchè?" vi chiederete, e lo facciamo anche noi.

π Teorema 7: Teorema di Bayes

Siano A_1, A_2, \dots, A_n eventi a due a due disgiunti e tali che la loro unione formi l'intero spazio degli eventi Ω . Sia inoltre B un evento tale che $P(B) > 0$. Allora, per ogni $i = 1, 2, \dots, n$, si ha

$$P(A_i|B) = \frac{P(B|A_i) \cdot P(A_i)}{\sum_{j=1}^n P(B|A_j) \cdot P(A_j)} \quad (1.18)$$

✓ Dimostrazione 1

Per dimostrare il teorema di Bayes, utilizzeremo la definizione di probabilità condizionata e alcune proprietà delle probabilità. Innanzitutto, per definizione di probabilità condizionata, abbiamo

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)}$$

Ma abbiamo anche che:

$$P(B|A_i) = \frac{P(A_i \cap B)}{P(A_i)} \implies P(A_i \cap B) = P(B|A_i) \cdot P(A_i)$$

Questo vuol dire che, riorganizzando i termini, abbiamo la formulazione più generica del teorema di Bayes, ovvero:

$$P(A_i|B) = \frac{P(B|A_i) \cdot P(A_i)}{P(B)}$$

Sostituendo la probabilità dell'evento B con la somma delle probabilità condizionate, derivante dalla legge delle alternative, otteniamo che:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^n P(B|A_j)P(A_j)}$$

Che è proprio il teorema di Bayes.

 **Esempio 1:** Il sospetto baro

Facciamo un esempio in cui facciamo ricorso al Teorema di Bayes: immaginiamo di giocare con un nostro amico che potrebbe essere un baro, o semplicemente molto fortunato, siccome vince sempre a testa o croce.

B = la probabilità che il nostro amico sia un baro, posta a $\frac{1}{2}$;

O = la probabilità che sia onesto, ovvero $P(\bar{B}) = 1 - \frac{1}{2} = \frac{1}{2}$;

$\Omega = \{B, O\}$

Il problema mi chiede di calcolare la probabilità che, data la sua vittoria, egli sia un baro:

$$P(B|V) = \frac{P(V|B) \cdot P(B)}{P(V)}$$

Inoltre possiamo trovare, per la legge delle alternative, che

$$P(V) = P(V|B) \cdot P(B) + P(V|O) \cdot P(O)$$

Dunque facendo i calcoli troviamo che

$$P(V) = 1 \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{4}$$

Ovvero la probabilità di vincere. Calcoliamo adesso

$$P(V|B) \cdot P(B) = 1 \cdot \frac{1}{2} = \frac{1}{2}$$

E infine sostituendo nell'equazione iniziale troviamo che :

$$P(B|V) = \frac{1}{2} \cdot \frac{4}{3} = \frac{2}{3}$$

Questo vuol dire che 2 volte su 3, se crediamo al 50% che il nostro amico sia un baro, esso lo sarà effettivamente basandoci sulla sua vittoria.

1.8 Paradosso di Monty-Hall

Il paradosso di Monty-Hall è un interessante problema di probabilità che deve le sue origini ad un gioco a premi americani. Il nome di questo problema deriva proprio dallo pseudonimo del conduttore dello show televisivo. Prende il nome di paradosso perchè la soluzione è controintuitiva ma è logicamente esatta.

Nel gioco vengono mostrate al concorrente tre porte chiuse. Dietro una di queste si nasconde una Nissan Skyline GT-R, dietro le altre due invece ci sono due capre. Il giocatore può scegliere una delle tre porte di cui vincerà il premio corrispondente. Dopo la scelta del giocatore, e prima che venga rivelato il contenuto della porta scelta, il conduttore che sa cosa si trova dietro ogni porta ne apre una delle due rimaste e rivela una capra, offrendo al giocatore la possibilità di cambiare porta. Se non conoscessimo alcuna nozione di probabilità ma soltanto la definizione banale penseremmo: "beh non cambia nulla la probabilità rimane $1/2$ ". Ma non è così e vediamo il perchè.

Precedentemente è stato introdotto oltre al concetto di probabilità condizionata un importantissimo teorema che poi vedremo che sarà utile in tutto il resto del corso, ovvero il teorema di Bayes, che ci tornerà utile per la spiegazione di questo paradosso.

Immaginiamo che il concorrente scelga la prima porta e che il conduttore riveli una capra dietro la porta 2. Adesso la questione è rimanere sulla porta 1 o cambiare scelta: "Cosa ci conviene fare?". Il teorema di Bayes ci permette di calcolare subito la probabilità a posteriori che l'auto si trovi dietro la porta 3:

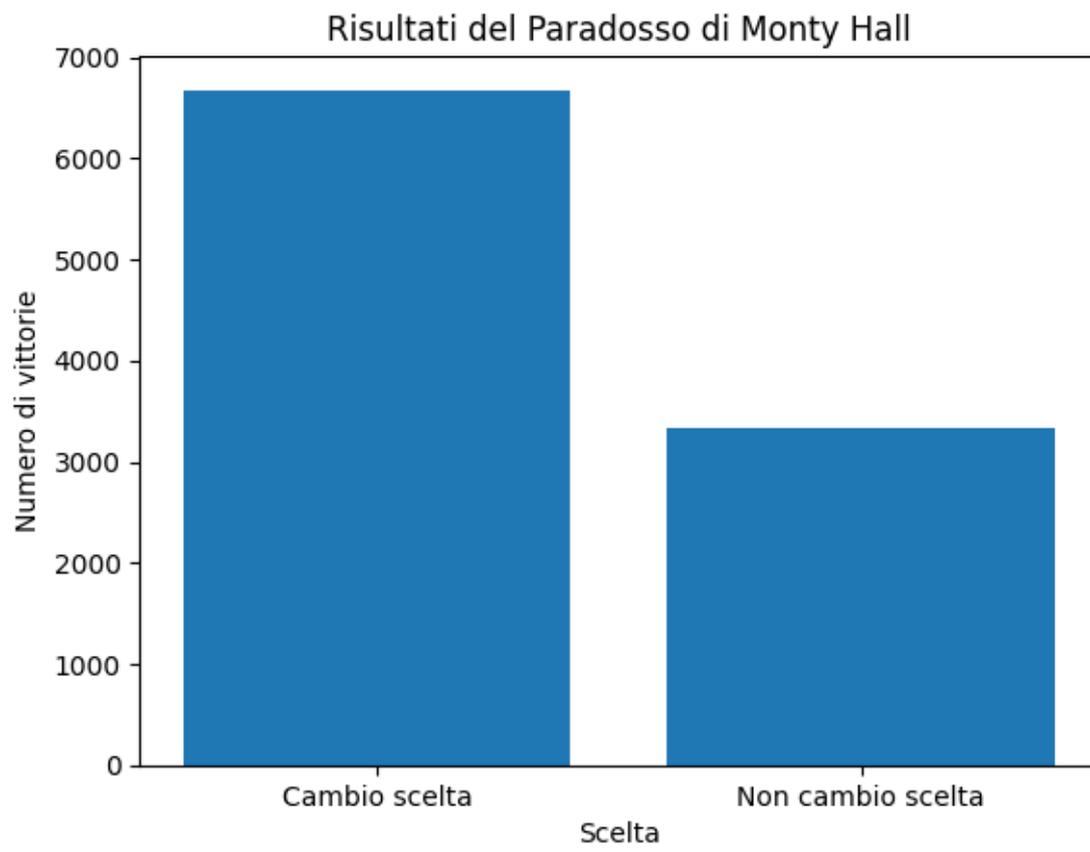
$$P(A3|C2) = \frac{P(C2|A3)P(A3)}{P(C3)} = \frac{(1 \cdot \frac{1}{3})}{\frac{1}{2}} = \frac{2}{3} \quad (1.19)$$

La probabilità a priori che l'auto si trovi dietro la porta 3 $P(A3)$ è $\frac{1}{3}$, poichè in base a quanto sappiamo a priori l'auto ha la stessa probabilità di trovarsi dietro ciascuna porta. La probabilità invece che il conduttore apra proprio la porta 3 è $\frac{1}{2}$, infatti dopo la scelta del concorrente il conduttore può scegliere solo tra due porte. Inoltre la probabilità che aprendo la porta riveli una capra è 1 poichè il conduttore sa dietro quale porta si trova l'automobile e poichè il concorrente ha già scelto la porta 1, l'unica possibilità è proprio la porta 3.

Nel caso in cui il concorrente confermi la sua scelta iniziale allora la probabilità sarebbe:

$$P(A1|C3) = \frac{P(C3|A1)P(A1)}{P(C3)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3} \quad (1.20)$$

Se ancora non foste convinti, qui di seguito è riportato il grafico di una simulazione in python del paradosso, dove si nota come le probabilità di vincita cambiando scelta siano più alte di quando invece non si cambia:



1.9 Riassunto

Tabella Riassuntiva 1

Gli assiomi di Kolmogorov dicono che:

- $0 \leq P(E|I_s) \leq 1$
- $P(\Omega) = 1$
- $P(E_1 \cup E_2) = P(E_1) + P(E_2)$

La legge della Somma dice che

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

Due eventi sono incompatibili se $P(A \cap B) = 0$ e sono indipendenti se vale che $P(A \cap B) = P(A|B) \cdot P(B) = P(A) \cdot P(B)$

La legge della moltiplicazione dice che

$$P\left(\bigcap_{i=1}^N E_i\right) = P(E_1) \cdot P(E_2|E_1) \cdot \dots \cdot P\left(E_N \left| \bigcap_{i=1}^{N-1} E_i\right.\right)$$

Se gli eventi sono indipendenti tra loro, invece, la relazione diventa:

$$P\left(\bigcap_{i=1}^N E_i\right) = \prod_{i=1}^N P(E_i)$$

La legge delle alternative dice che:

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

Il teorema di Bayes dice che:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|\bar{A}) \cdot P(\bar{A})}$$

2 Teoria degli Errori



Sommario

- 2.1 Introduzione, 21
- 2.2 Incertezze di tipo A, 21
- 2.3 Incertezze di tipo B, 21
- 2.4 Variabili casuali e Distribuzioni discrete, 22
- 2.5 Valore Atteso, 23
- 2.6 Varianza e Deviazione Standard, 24
- 2.7 Covarianza, 25
- 2.8 Coefficiente di correlazione, 27
- 2.9 Variabili Casuali Discrete pt.II, 28
- 2.10 Variabili Casuali Continue, 29
- 2.11 Valore atteso, Varianza di combinazioni lineari, 34
- 2.12 Linearizzazione, 35
- 2.13 Bilinearità della covarianza, 39
- 2.14 Convoluzione di due funzioni di probabilità, 40
- 2.15 Cambiamento di Variabile, 43

2.1 Introduzione

Nelle esperienze di laboratorio ci troviamo di fronte al calcolo delle incertezze. In questa prima sezione cercheremo di capire cosa si intende per incertezza e poi analizzeremo più dal punto di vista matematico il calcolo di queste ultime, soprattutto come si propagano da misure dirette a misure indirette.

Per fare questo ci serviremo anche di python, che con i suoi formidabili calcoli e funzioni faciliterà e non poco il nostro lavoro. Questa parte in particolare è trattata nel capitolo dedicato.

2.2 Incertezze di tipo A

Il primo caso di incertezze che introduciamo sono quelle di tipo A, ma cosa si intende per "incertezze di tipo A"? Formalmente è il termine che si ottiene attraverso l'analisi statistica di misure ripetute, più semplicemente sono quelli errori casuali che ripetendo N volte una stessa misura si generano appunto "casualmente" (misurando la stessa quantità, a meno che non si usi uno strumento con poca sensibilità e non si abbia grande fortuna, si prenderanno misure leggermente diverse). Una volta effettuate N misure di una grandezza fisica si prende in esame il valore medio delle misure e come incertezza si prende il rapporto tra lo scarto quadratico medio e la radica quadrata del numero di prove effettuate.

Definizione 3

Chiamiamo **incertezza di tipo A** l'errore o l'incertezza da considerarsi per una misura o una variabile casuale, calcolata a partire dalla ripetizione di una misura o di una variabile x_i , nel momento in cui si ha $x_i \neq x_j$ per qualche valore.

$$\sigma_A = \frac{\sigma}{\sqrt{N}} \quad (2.1)$$

2.3 Incertezze di tipo B

L'incertezza di tipo B è il termine ottenuto tramite informazioni disponibili sulla misura senza effettuare prove ripetute. Esempi di tali informazioni sono: dati acquisiti in misurazioni precedenti, esperienza dell'operatore, informazioni sulla strumentazione utilizzata (se per esempio c'è un offset a uno strumento, o banalmente un tavolo che bascula, è ovvio che farà un errore nel misurare, e devo tenerne conto). Quindi si tratta di un approccio a priori.

2.4 Variabili casuali e Distribuzioni discrete

Per formalizzare i concetti visti fin ora, dobbiamo continuare con la formalizzazione di altri concetti, partendo da quello di variabile casuale e della sua relativa funzione di probabilità che ne descrive appunto la probabilità di assumere un certo valore.

Definizione 4

Chiamiamo X una certa variabile, e chiamiamo P la probabilità che essa ha di assumere un certo valore x_i , in breve $P(X = x_i)$. Tale variabile si dice **discreta** se assume valori finiti o appartenenti a un'infinità numerabile.

Siccome in molte situazioni si può descrivere con una funzione più o meno semplice la probabilità di tutte le occorrenze di X , chiamiamo questa funzione $f(x)$.

Ritorniamo in seguito sulle variabili casuali discrete, ma per ora ci serve solo sapere, oltre a quanto già detto, che la variabile casuale per la costruzione della *funzione di probabilità* deve essere tale che tutte le configurazioni x_i che X può assumere siano una classe completa. Sulla base di ciò, è chiaro che la funzione di probabilità debba avere delle caratteristiche ben precise:

- $0 \leq f(x) \leq 1$;
- $\sum_{i=1}^n f(x_i) = 1$;
- $P(X = x_i \cup X = x_j) = f(x_i) + f(x_j)$.

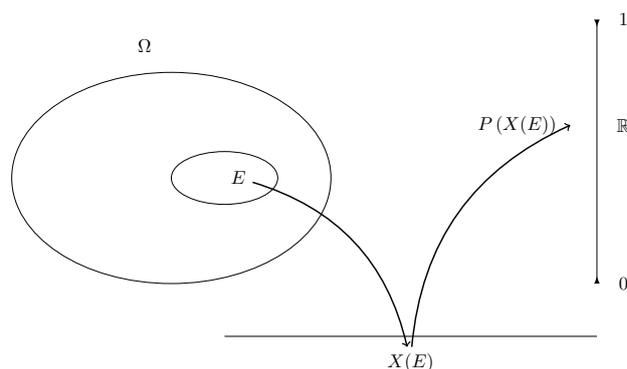


Figura 2.1: Costruzione di una Variabile Casuale

2.5 Valore Atteso

Un operatore interessante da calcolare è il valore atteso $\mathbb{E}[X]$ che semplicisticamente è "una media" delle misure effettuate. Infatti si calcola proprio come una media.

Definizione 5

Si dice **Valore Atteso** di una serie di misure o di variabili casuali x_i , l'operatore lineare definito come

$$\mathbb{E}[X] = \sum_{i=1}^n x_i \cdot f(x_i) \quad (2.2)$$

In cui $f(x_i)$ è la probabilità di ogni valore x_i .

Questo equivale a quello che fin ora abbiamo chiamato x_{best} ovvero la migliore stima di una misura, solo che possiamo usare il valore atteso anche per altre variabili casuali. Oltre a questa notazione si utilizza anche il simbolo greco μ .

2.5.1 Valore atteso di funzione di Variabile casuale

Se X è una variabile casuale, anche la generica funzione $g(X)$ lo sarà, vale che:

$$P(g(x) = g(x_i)) = P(X = x_i) = f(x_i)$$

Allora l'Equazione 2.2 diventa

$$\mathbb{E}[g(x)] = \sum_{i=1}^n g(x_i) \cdot f(x_i)$$

Nel caso in cui $g(x)$ dipendesse linearmente da X avrei che $g(x) = ax + b$ e quindi che:

$$\mathbb{E}[ax + b] = \sum_i (ax_i + b)f(x_i) = \sum_i ax_i f(x_i) + b \sum_i f(x_i) = a\mathbb{E}[x] + b$$

Questa è un'ulteriore conferma del fatto che l'operatore $\mathbb{E}[\dots]$ è di tipo lineare.

2.6 Varianza e Deviazione Standard

Il valore atteso riassume in un solo numero l'informazione relativa intorno al quale probabilmente si realizzerà la variabile casuale. Ma questa non è certa dunque bisogna associarvi un'incertezza. Per fare ciò useremo l'intuito (letteralmente), secondo il quale se la previsione è buona ci aspettiamo un piccolo scarto fra il valore atteso e quello che si verificherà.

$$\Delta = X - \mathbb{E}[X] = X - \mu$$

Una possibile misura dell'incertezza, potrebbe essere il valore atteso dei possibili scarti della previsione stessa, ma si vede subito che questo valore è 0 per ogni misura, siccome

$$\mathbb{E}[X - \mu] = \mathbb{E}[X] - \mu = 0$$

Consideriamo allora la deviazione standard come il valore atteso del quadrato dei possibili scarti:

Definizione 6

Chiamiamo **Varianza** di una misura X il valore atteso del quadrato degli scarti $X - \mu$, quindi

$$\text{Var}[X] = \mathbb{E}[(X - \mu)^2] \quad (2.3)$$

Da qui, diventando pigri e chiamando le variabili casuali semplicemente come x per non mettere ogni volta il Caps, la deviazione standard non sarà altro che la radice della Varianza, ovvero

$$\sigma^2[x] = \text{Var}[x] = \mathbb{E}[(x - \mu)^2] = \mathbb{E}[x^2 - 2x\mu + \mu^2] = \mathbb{E}[x^2] - 2\mu^2 + \mu^2 = \mathbb{E}[x^2] - \mu^2$$

Da qui possiamo definirla con una formula leggermente più elegante (perchè no, d'altronde):

Definizione 7

Definiamo **Deviazione standard** lo scostamento medio di una variabile casuale rispetto al suo valore atteso, ed equivale alla radice quadrata della Varianza:

$$\sigma = \sqrt{\text{Var}[x]} = \sqrt{\mathbb{E}[x^2] - \mathbb{E}^2[x]} \quad (2.4)$$

La varianza inoltre non è un operatore lineare e questo possiamo dimostrarlo :

$$\text{Var}[ax + b] = \mathbb{E} \left[[(ax + b) - \mathbb{E}(ax + b)]^2 \right] = a^2 \left(\mathbb{E}[x] - \mathbb{E}[x^2] \right)$$

2.7 Covarianza

Adesso però dobbiamo introdurre un nuovo concetto (scaga eh?), ovvero quello della covarianza, che molto banalmente è la quantificazione di come le due variabili si discostano dai loro valori medi (quindi vale solo per funzioni in più variabili). Il segno della covarianza permette di dire se le fluttuazioni intorno alla media delle due variabili sono concordi o discordi. Noi la indicheremo come segue:

Definizione 8

Definiamo come **Covarianza** lo scostamento medio di due variabili casuali x, y dai loro valori medi, con x, y variabili di una funzione $f(x, y)$ che rappresenta una variabile casuale.

$$\text{Cov}[x, y] = \mathbb{E}[(x - \mu_x)(y - \mu_y)] = \sigma_{xy} \quad (2.5)$$

Possiamo riscrivere la covarianza anche in un modo un pò più articolato, così, perchè ci piace farci del male da soli:

$$\begin{aligned} \text{Cov}[x, y] &= \mathbb{E}[(x - \mu_x)(y - \mu_y)] = \iint (x - \mu_x)(y - \mu_y) f(x, y) dx dy = \\ &= \iint (xy - \mu_x y - \mu_y x + \mu_x \mu_y) f(x, y) dx dy = \\ &= \mathbb{E} \left[x \cdot y - x \cdot \mathbb{E}[y] - y \cdot \mathbb{E}[x] + \mathbb{E}[x] \cdot \mathbb{E}[y] \right] = \\ &= \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y] - \mathbb{E}[y]\mathbb{E}[x] + \mathbb{E}[x]\mathbb{E}[y] = \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y] \end{aligned}$$

Allora otteniamo una formulazione altrettanto comoda, a cui si poteva arrivare anche con le serie piuttosto che con gli integrali:

$$\text{Cov}[x, y] = \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y]$$

Il valore atteso $\mathbb{E}[xy] = \iint xy f(x, y) dx dy$ non ha in genere metodi di risoluzione

alternativi se non dalla definizione. Si verifica però facilmente che se le due variabili sono indipendenti varrà che

$$f(x, y) = f(x) \cdot f(y)$$

E di conseguenza gli integrali si separano e la Covarianza tra i due vale:

$$\begin{aligned} \text{Cov}[x, y] &= \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y] = \\ &= \int x f(x) dx \cdot \int y f(y) dy - \mathbb{E}[x]\mathbb{E}[y] = \mathbb{E}[x]\mathbb{E}[y] - \mathbb{E}[x]\mathbb{E}[y] = 0 \end{aligned}$$

2.8 Coefficiente di correlazione

Adesso introduciamo un' altro concetto non poco utile, ovvero il **coefficiente di correlazione**.

Il coefficiente di correlazione è una misura specifica usata nell'analisi della correlazione per quantificare la forza della relazione lineare tra due variabili. Il coefficiente di correlazione è un indice sempre compreso tra -1 ed 1. Se il coefficiente di correlazione è positivo, le due variabili sono direttamente correlate, ovvero se una variabile aumenta anche l'altra aumenta. Se il coefficiente di correlazione è negativo, le due variabili sono inversamente correlate, ovvero se una variabile aumenta l'altra diminuisce.

Definizione 9

Definiamo **Coefficiente di correlazione** il numero $\rho(x, y)$ tale che $-1 \leq \rho \leq 1$ che indica la relazione lineare tra due variabili x, y .

$$\rho(x, y) = \frac{\text{Cov}[x, y]}{\sigma_x \cdot \sigma_y} \quad (2.6)$$

Possiamo calcolare a mano questo valore (come possiamo calcolare a mano il numero 7^{123}) ma non per questo dobbiamo, piuttosto come ogni cosa che riguarda la statistica e soprattutto la teoria degli errori, impariamo un sacco di teoria per poi far fare tutto a Python.

```

1
2 import numpy as np
3
4 #Dati di esempio
5 x = np.array([1, 2, 3, 4, 5])
6 y = np.array([5, 4, 3, 2, 1])
7
8 #Calcolo del coefficiente di correlazione
9 corr = np.corrcoef(x, y)[0][1]
```

Codice 2.1: Esempio di codice per calolare il coefficiente di correlazione

2.9 Variabili Casuali Discrete pt.II

Abbiamo già definito una variabile casuale discreta. Ora ne definiamo la **Funzione Cumulativa**, ovvero la probabilità che la variabile X assuma un valore più piccolo di un certo x_i :

$$F(x_i) = P(x \leq x_i) = f(x_1) + f(x_2) + \dots + f(x_i) = \sum_i^n f(x_i)$$

Quindi, in modo più elegante (sempre per lo stesso motivo):

Definizione 10

Sia X una variabile casuale discreta. La sua funzione cumulativa $F(x_i)$ è definita come

$$F(x_i) = \sum_i^n f(x_i) \quad (2.7)$$

Anche qui, come per [Sezione 2.4](#), tutti i valori di X formano una classe completa di eventi, dunque la funzione cumulativa ha le seguenti proprietà:

- $0 \leq F(x) \leq 1$;
- $\lim_{x \rightarrow +\infty} F(x) = 1$;
- $\lim_{x \rightarrow -\infty} F(x) = 0$.

Di cui la prima relazione è ovvia a partire dal fatto che, se sto sommando una serie di probabilità di una classe completa, avrò al più come risultato 1. Per la seconda, sviluppanola si ha che $\lim_{x \rightarrow \infty} F(x) = F(x_n) = P(x \leq x_n) = \sum_{i=1}^n f(x_i) = 1$, e allo stesso modo facendo il limite nell'altra parte del dominio otterrò 0.



Figura 2.2: Esempio di N misure rispetto a μ e σ

2.10 Variabili Casuali Continue

Fin ora ci siamo occupati di variabili che possono assumere valori discreti (nel regime di \mathbb{N}), ma nella pratica i valori casuali sono, il più delle volte, variabili continue (come se misuro una lunghezza, un tempo...).

Nel caso continuo non abbiamo più una distribuzione di probabilità che ci da la probabilità puntuale, ma $f(x)$ ci da la probabilità che la variabile casuale assuma un certo valore in un intervallo. Questo perché essendoci in un intervallo arbitrario un numero infinito di punti, la probabilità di ogni punto dovrebbe essere nulla, ragionando con il concetto di casi favorevoli su casi possibili (appunto infiniti!). Si considera quindi un intervallo di probabilità siccome è chiaro che

$$P(a \leq x \leq b) > 0$$

Varrà allora che all'aumentare dell'intervallo aumenterà la probabilità, ovvero

$$\Delta P \propto \Delta x$$

E quando l'intervallo diventa infinitesimo (a noi che ce frega dell'intervallo grande, minimizziamo l'intervallo fino ad arrivare a un valore di probabilità che ci soddisfi per un valore singolo o abbastanza piccolo che non sia 0 come dicevamo prima):

$$dP \propto dx$$

Ma è anche vero che nel caso in cui la probabilità non sia uguale in ogni punto, il rapporto delle probabilità attorno a due punti è proporzionale ai loro gradi di fiducia:

$$\frac{dP(x_1 \leq X \leq x_1 + dx)}{dP(x_2 \leq X \leq x_2 + dx)} = \frac{f(x_1)}{f(x_2)}$$

Cioè

$$dP \propto f(x)dx \implies dP = f(x)dx$$

Siccome abbiamo trovato la costante di normalizzazione della proporzionalità di prima. Parleremo dunque di **DENSITA' DI PROBABILITA'**. Allora otteniamo che:

$$P(a \leq x \leq b) = \int_a^b f(x)dx$$

In poche parole, e com'è anche facilmente intuibile, tutte le funzioni viste nel caso discreto si trasformano in quello continuo (per esempio le somme diventano integrali e così via):

$$f(x_i) \longrightarrow f(x) dx$$

$$\sum_{i=1}^n f(x_i) \longrightarrow \int_{-\infty}^{+\infty} f(x) dx$$

Allo stesso modo la funzione cumulativa si modifica:

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$$

Il valore atteso di x e $g(x)$ diventano:

$$\mu_x = \mathbb{E}[x] = \int_{-\infty}^{+\infty} x f(x) dx$$

$$\mathbb{E}[g(x)] = \int_{-\infty}^{+\infty} g(x) f(x) dx$$

La Varianza invece:

$$\text{Var}[x] = \int_{-\infty}^{+\infty} = \mathbb{E}[x - \mathbb{E}(x)]^2 = \int_{-\infty}^{+\infty} (x - \mathbb{E}[x]) f(x) dx$$

Come già dimostrato in precedenza il valore atteso è un operatore lineare, ma riproponiamo lo stesso la dimostrazione:

$$\mathbb{E}[ax + b] = \int_{-\infty}^{+\infty} (ax + b) f(x) dx = \int_{-\infty}^{+\infty} ax f(x) dx + \int_{-\infty}^{+\infty} b f(x) dx = a\mathbb{E}[x] + b$$

2.10.1 Caso discreto in due variabili

Consideriamo adesso il caso in cui abbiamo una funzione di due variabili del tipo:

$$f(x, y) = P(X = x, Y = y) = P((X = x) \cap (Y = y))$$

La condizione di normalizzazione sarà del tipo:

$$\sum_{i,j=1}^n f(x_i, y_j) = 1$$

Mentre la funzione cumulativa è data da

$$F(x_i, y_j) = \sum_{i,j=1}^n f(x_i, y_j)$$

Volendo, seguendo questa logica, si può considerare una qualsiasi variabile casuale X che dipende da n valori, anch'esse variabili casuali, e la si può scrivere come una funzione $f(x_1, x_2 \dots x_n)$. È poi importante calcolare la deviazione standard di una funzione di questo tipo, che comprenderà la variazione di entrambe le variabili e la loro covarianza (vedi [Sezione 2.7](#)). Consideriamo il caso generale di una funzione lineare, quindi

$$f(x_1, x_2 \dots x_n) = \sum_{i=1}^n a_i x_i$$

A questo punto calcoliamo il valore atteso:

$$\mathbb{E}[f] = \mathbb{E} \left[\sum_{i=1}^n a_i x_i \right] = \sum_{i=1}^n a_i \cdot \mathbb{E}[x_i]$$

E questo per la linearità del valore atteso che abbiamo già discusso. Per quanto riguarda la varianza, ci tocca calcolarci sia $\mathbb{E}[f^2]$ che $\mathbb{E}^2[f]$ e iniziamo proprio da questo che è più facile:

$$\mathbb{E}^2[f] = \sum_{i=1}^n a_i \mathbb{E}[x_i] \cdot \sum_{j=1}^n a_j \mathbb{E}[x_j] = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \mathbb{E}[x_i] \mathbb{E}[x_j]$$

Per quanto riguarda l'altro valore è invece:

$$\mathbb{E}[f^2] = \mathbb{E} \left[\sum_{i=1}^n \sum_{j=1}^n a_i a_j x_i x_j \right] = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \mathbb{E}[x_i \cdot x_j]$$

Uniamo adesso le due cose, e utilizziamo una sola sommatoria per comodità, e troviamo:

$$\text{Var}[f] = \mathbb{E}[f^2] - \mathbb{E}^2[f] = \sum_{i,j=1}^n a_i a_j \mathbb{E}[x_i \cdot x_j] - \sum_{i,j=1}^n a_i a_j \mathbb{E}[x_i] \mathbb{E}[x_j] = \sum_{i,j=1}^n a_i a_j \cdot \underbrace{\left(\mathbb{E}[x_i \cdot x_j] - \mathbb{E}[x_i] \mathbb{E}[x_j] \right)}_{=\text{Cov}[x_i, x_j]}$$

Quindi concludiamo che:

$$\text{Var}[f] = \sum_{i,j=1}^n a_i a_j \text{Cov}[x_i, x_j] \quad (2.8)$$

Questa formula va bene in ogni caso, anche nel momento in cui le variabili siano indipendenti tra loro e quindi abbiano covarianza pari a 0, siccome sviluppando i primi termini vediamo che:

$$\sum_{i,j=1}^n a_i a_j \text{Cov}[x_i, x_j] = a_1 a_1 \underbrace{\text{Cov}[x_1, x_1]}_{=\sigma_{x_1}^2} + a_1 a_2 \underbrace{\text{Cov}[x_1, x_2]}_{=0} + \dots$$

Quindi sopravvivono solo i termini con indice uguali, cioè nel caso di variabili indipendenti abbiamo:

$$\text{Var}[f] = \sum_{i=1}^n a_i^2 \sigma_{x_i}^2 \quad (2.9)$$

2.10.2 Caso continuo in due variabili

Similmente a prima, nel caso continuo in due variabili la funzione cumulativa di densità di probabilità diventa del tipo:

$$F(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f(x, y) dx dy$$

La condizione di normalizzazione invece sarà del tipo:

$$\iint_{-\infty}^{+\infty} f(x, y) dx dy = 1$$

Dalla funzione cumulativa otteniamo che:

$$f(x, y) = \frac{\partial F(x, y)}{\partial x} \frac{\partial F(x, y)}{\partial y} = P \left[(x \leq X \leq x + dx) \cap (y \leq Y \leq y + dy) \right]$$

Il teorema di Bayes nel caso di variabili continue si trasforma nel seguente modo:

$$f(x|y) dx = \frac{f(y, x) dx dy}{f(y) dy} \rightarrow f(x|y) = \frac{f(y|x)f(x)}{\int_{-\infty}^{+\infty} f(x, y) dx}$$

Il valori attesi (che in questo caso sono due, chiaramente) saranno dati da:

$$\begin{aligned} \mathbb{E}[x] &= \iint x f(x, y) dx dy = \int x dx \int f(x, y) dy = \int x f(x) dx \\ \mathbb{E}[y] &= \iint y f(x, y) dx dy = \int y dy \int f(x, y) dx = \int y f(y) dy \end{aligned}$$

2.10.3 Distribuzioni Marginali

Da una distribuzione di probabilità congiunta per due variabili, possiamo ottenere la distribuzione di solo una delle due, sommando (o integrando) le probabilità di tutte le possibilità delle altre variabili per un certo valore della variabile d'interesse. Nel caso continuo sarebbe:

$$f(x) = \int f(x, y) dy \quad (2.10)$$

$$f(y) = \int f(x, y) dx \quad (2.11)$$

2.11 Valore atteso, Varianza di combinazioni lineari

Iniziamo subito con la valutazione del valore atteso di una combinazione lineare che risulta essere un procedimento piuttosto banale:

$$\begin{aligned}\mathbb{E}[x \pm y] &= \iint (x \pm y)f(x, y) dx dy = \\ &= \iint x f(x, y) dx dy \pm \iint y f(x, y) dx dy = \mathbb{E}[x] \pm \mathbb{E}[y]\end{aligned}$$

Una volta calcolato quest'ultimo, ci interessiamo dell'incertezza della previsione:

$$\begin{aligned}\text{Var}[x \pm y] &= \mathbb{E}\left[\left((x \pm y) - \mathbb{E}[x \pm y]\right)^2\right] \rightarrow \mathbb{E}\left[\left((x - \mathbb{E}[x]) \pm (y - \mathbb{E}[y])\right)^2\right] \rightarrow \\ &\rightarrow \mathbb{E}\left[(x - \mathbb{E}[x])^2 \pm (y - \mathbb{E}[y])^2 \pm 2(x - \mathbb{E}[x])(y - \mathbb{E}[y])\right] \rightarrow \text{Var}[x] \pm \text{Var}[y] \pm 2 \text{Cov}[x, y]\end{aligned}$$

Abbiamo concluso, in maniera anche abbastanza sensata, che:

$$\text{Var}[x \pm y] = \text{Var}[x] \pm \text{Var}[y] \pm 2 \text{Cov}[x, y]$$

2.12 Linearizzazione

Consideriamo adesso una funzione del tipo:

$$Y = y(x_1, x_2, \dots, x_n)$$

Calcolare valore atteso e varianza di una funzione di questo tipo è follia, siccome a partire dalla definizione non è umanamente pensabile di fare determinati calcoli se la funzione non è di tipo lineare ma ha potenze, logaritmi... al suo interno. Per questo motivo *linearizziamo* la funzione, e per farlo possiamo fare lo sviluppo di Taylor al primo ordine e ipotizzandola *abbastanza lineare* nell'intorno di ogni X_i ottenere dunque:

$$\begin{aligned} Y &\approx y(\mu_1, \mu_2, \dots, \mu_n) + \sum_i^n \frac{\partial y}{\partial x_i}(x_i - \mu_i) \Big|_{\mu_i} \\ &\approx k + \sum_i^n \frac{\partial y}{\partial x_i}(x_i - \mu_i) \Big|_{\mu_i} \end{aligned}$$

In cui abbiamo posto $y(\mu_1, \mu_2, \dots, \mu_n) = k \in \mathbb{R}$ per sottolineare che questo sia un valore costante, non una variabile (sarebbe la funzione calcolata nel punto μ_i per ogni x_i . Il valore atteso di y è circa uguale alla funzione y calcolata sulle previsioni delle x_i (le derivate restano calcolate nei valori attesi ma non le scriviamo per comodità):

$$\begin{aligned} \mathbb{E}[y] &\approx \mathbb{E} \left[y(\mu_1, \mu_2, \dots, \mu_n) + \sum_i^n \frac{\partial y}{\partial x_i}(x_i - \mu_i) \right] \\ &\approx y(\mu_1, \mu_2, \dots, \mu_n) + \mathbb{E} \left[\sum_i^n \frac{\partial y}{\partial x_i}(x_i - \mu_i) \right] \\ &\approx y(\mu_1, \mu_2, \dots, \mu_n) + \sum_i^n \frac{\partial y}{\partial x_i} \cdot \underbrace{\mathbb{E}[(x_i - \mu_i)]}_{=0} \\ \implies \mathbb{E}[y] &\approx y(\mu_1, \mu_2, \dots, \mu_n) \end{aligned}$$

La varianza invece è pari a:

$$\sigma_y^2 = \sum_{i,j=1}^n \frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j} \sigma_{ij}$$

Vediamo come ci si arriva a questo risultato, partendo dalla definizione di varianza:

$$\sigma_y^2 = \mathbb{E}[y^2] - \mathbb{E}^2[y] \approx \mathbb{E} \left[\left(y(\mu_1, \mu_2 \dots \mu_n) + \sum_i^n \frac{\partial y}{\partial x_i} (x_i - \mu_i) \right)^2 \right] - [y(\mu_1, \mu_2 \dots \mu_n)]^2$$

Adesso calcoliamoci $\mathbb{E}[y^2]$ separatamente per evitare di perderci:

$$\begin{aligned} \mathbb{E}[y^2] &\approx \mathbb{E} \left[[y(\mu_1, \mu_2 \dots \mu_n)]^2 + \sum_i^n \sum_j^n \frac{\partial y}{\partial x_i} (x_i - \mu_i) \cdot \frac{\partial y}{\partial x_j} (x_j - \mu_j) + 2y(\mu_1, \mu_2 \dots \mu_n) \sum_i^n \frac{\partial y}{\partial x_i} (x_i - \mu_i) \right] \\ &\approx [y(\mu_1, \mu_2 \dots \mu_n)]^2 + \sum_{i,j}^n \frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j} \cdot \underbrace{\mathbb{E}[(x_i - \mu_i)(x_j - \mu_j)]}_{=\text{Cov}[x_i, x_j]} + 2y(\mu_1, \mu_2 \dots \mu_n) \cdot \sum_i^n \frac{\partial y}{\partial x_i} \cdot \underbrace{\mathbb{E}[(x_i - \mu_i)]}_{=0} \\ &\approx [y(\mu_1, \mu_2 \dots \mu_n)]^2 + \sum_{i,j}^n \frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j} \cdot \sigma_{i,j} \end{aligned}$$

Adesso mettiamo tutto insieme e otteniamo la formula iniziale:

$$\begin{aligned} \sigma_y^2 &\approx \cancel{[y(\mu_1, \mu_2 \dots \mu_n)]^2} + \sum_{i,j}^n \frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j} \cdot \sigma_{i,j} - \cancel{[y(\mu_1, \mu_2 \dots \mu_n)]^2} \\ &\approx \sum_{i,j}^n \frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j} \cdot \sigma_{i,j} \end{aligned}$$

Ovviamente, e in modo del tutto simile a quanto succedeva per la deviazione standard della combinazione lineare di variabili casuali (ovvero [Sottosezione 2.10.1](#)), se i termini (tutti o alcuni) sono indipendenti, la covarianza vale 0 nella formula e si salvano solo quei termini indipendenti che vengono elevati al quadrato perché gli indici sono uguali:

$$\sigma_y^2 = \sum_{i=1}^n \left(\frac{\partial y}{\partial x_i} \right)^2 \cdot \sigma_{x_i}^2 \quad (2.12)$$

Quello che stiamo facendo, per cercare di dare un senso a tutti questi conti stratosferici, è trovarci un modo relativamente comodo (o quantomeno fattibile, rispetto a integrali impossibili) di calcolarci la varianza e il valore atteso di una funzione a più variabili che possono avere dipendenza o meno, relazioni strane e chi più ne ha più ne metta, e lo facciamo linearizzandola (utile quando per esempio calcoliamo una grandezza in modo indiretto). Approssimiamo quindi la funzione (che qui sotto è rappresentata in una dimensione per facilità grafica) con lo sviluppo al primo ordine di Taylor e calcoliamo la variazione su quella retta piuttosto che sulla funzione in sé:

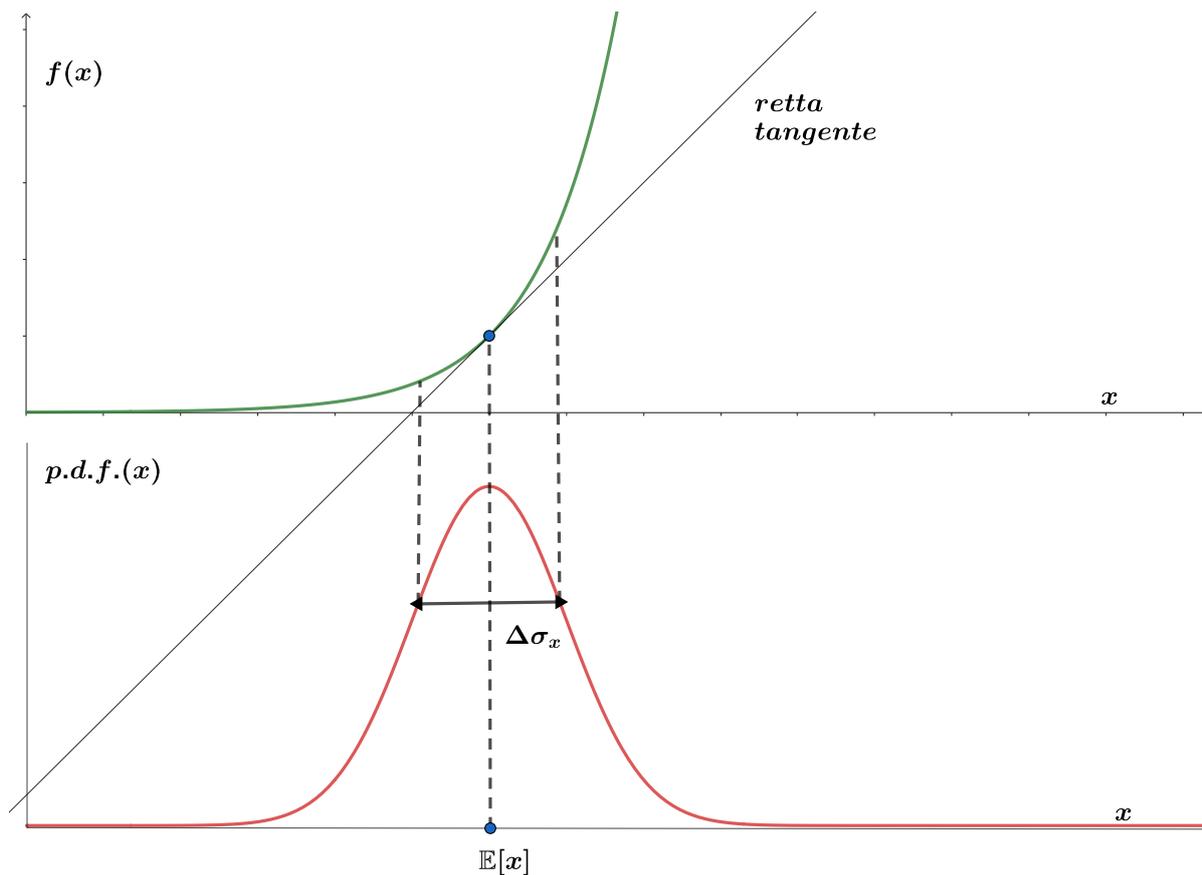
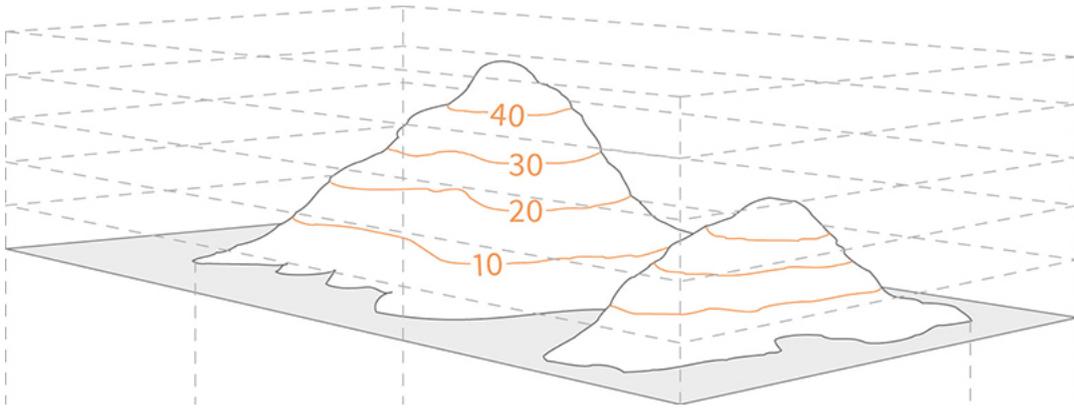


Figura 2.3: Semplificazione del processo di Linearizzazione

Nel momento in cui questa approssimazione dovesse fallire (se spostassimo di poco a destra il valore di $\mathbb{E}[x]$ andremmo in contro alla crescita rapida della funzione e a una maggiore differenza tra la σ calcolata sulla funzione e sulla tangente) possiamo *semplicemente* usare la formula teorica per la deviazione

$$\sigma^2(y) = \int (y - \mathbb{E}[y])^2 f(y) dy = \int (f(x) - \mathbb{E}[f(x)])^2 f(x) dx$$



La figura sopra riportata, mette in evidenza le curve di livello di una montagna. Immaginiamo adesso di approssimare la funzione in un punto che prendiamo essere $f(x_1, x_2)$, allora usiamo Taylor:

$$f(x_1, x_2) = F(\mathbb{E}[x_1, x_2]) + \frac{\partial f(x_1, x_2)}{\partial x_1} \cdot (x_1 - \mathbb{E}[x_1]) + \frac{\partial f(x_1, x_2)}{\partial x_2} \cdot (x_2 - \mathbb{E}[x_2])$$

Da cui ricaviamo due dati importanti:

- $F(\mathbb{E}[x_1, x_2])$ è l'approssimazione del valore atteso in quel punto della funzione;
- $\frac{\partial f(x_1, x_2)}{\partial x_1} \frac{\partial f(x_1, x_2)}{\partial x_2} \cdot \text{Cov}[x_1, x_2]$ indica la varianza.

Dunque se voleste scalare una montagna, in modo molto ma molto banale calcolate il gradiente della funzione così almeno prendete la via più veloce, ma veloce non vuol dire sicura.

2.13 Bilinearità della covarianza

La covarianza è una misura della relazione lineare tra due variabili casuali. La bilinearità della covarianza si riferisce al fatto che la covarianza è una forma bilineare simmetrica positiva. In altre parole, la covarianza soddisfa le seguenti proprietà:

- La covarianza di una costante con una variabile casuale è zero.
- La covarianza di due variabili casuali indipendenti è zero.
- La covarianza è simmetrica, ovvero la covarianza tra X e Y è uguale alla covarianza tra Y e X.
- La covarianza soddisfa la proprietà di bilinearità, ovvero la covarianza di una combinazione lineare di due variabili casuali è uguale alla combinazione lineare delle loro covarianze.

Immaginiamo di avere due funzioni e di voler calcolare la Covarianza di queste due funzioni:

- $y_a = a_1x_1 + a_2x_2$
- $y_b = b_1x_1 + b_2x_2$

$$\begin{aligned} \text{Cov}[y_a, y_b] &= \text{Cov}[a_1x_1 + a_2x_2 + b_1x_1 + b_2x_2] = \\ &= a_1 \cdot \text{Cov}[x_1, b_1x_1 + b_2x_2] + a_2 \cdot \text{Cov}[x_2, b_1x_1 + b_2x_2] = \\ &= a_1b_1 \cdot \text{Cov}[x_1, x_1] + a_2b_2 \cdot \text{Cov}[x_2, x_2] = \\ &= a_1b_1\sigma_{x_1}^2 + a_2b_2\sigma_{x_2}^2 + a_1b_2 \cdot \text{Cov}[x_1, x_2] + a_2b_1 \cdot \text{Cov}[x_2, x_1] \end{aligned}$$

Quanto ottenuto possiamo scriverlo anche come:

$$\text{Cov}[y_a, y_b] = \sum_{i=1}^n \sum_{j=1}^n a_i b_j \text{Cov}[x_i, x_j] = a^T V b$$

Dove V non è nient'altro che la matrice di covarianza composta in questo modo:

$$V = \begin{bmatrix} \text{Cov}(X_1, X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \text{Cov}(X_2, X_2) & \cdots & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \cdots & \text{Cov}(X_n, X_n) \end{bmatrix}$$

2.14 Convoluzione di due funzioni di probabilità

Consideriamo la variabile Z come somma di due variabili indipendenti X, Y , varrà che

$$f(z) = \sum_{x,y} f(x,y) = \sum_x f(x, z-x)$$

E questo perchè vale sempre la condizione $z = x + y$ quindi scelta una variabile l'altra è "vincolata". Nel caso di variabili indipendenti però essa si riduce a

$$f(z) = \sum_x f(x)f(z-x)$$

Che rappresenta il *caso discreto di convoluzione* di due funzioni di probabilità indipendenti. Possiamo esprimere la convoluzione sia dal punto di vista discreto che continuo (non utile nel corso in realtà), ma quello che importa veramente è il senso che si cela dietro queste formule.

Definizione 11: Convoluzione discreta

La **Convoluzione** è un'operazione matematica utilizzata in teoria dei segnali, sistemi, probabilità...

Date due funzioni $f(x)$ e $f(y)$ definite su \mathbb{Z} , allora la loro convoluzione è

$$[f(x) * f(y)] := \sum_{x=-\infty}^{+\infty} f(x)f(z-x) \quad (2.13)$$

Definizione 12: Convoluzione continua

La **Convoluzione** per due funzioni $f(t)$ e $g(t)$ definite in \mathbb{R} è:

$$(f * g)(t) := \int_{-\infty}^{+\infty} f(\tau)g(t-\tau) d\tau \quad (2.14)$$

Esempio 2

Prendiamo come esempio quello del lancio di due dadi, nel caso in cui vogliamo sapere la probabilità che ci esca una certa somma, ovvero dati $P(x)$ e $P(y)$ le probabilità dei lanci singoli, vogliamo sapere $P(x + y)$, che assomiglia molto al processo di convoluzione fatto prima per la somma di due variabili.

A livello "bruto" potremmo fare una tabella con tutti i possibili outcome, e ci accorgeremo che sulle diagonali dei valori avremmo tutte le possibili combinazioni per quel numero:

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

E quindi sapendo che ci sono 36 possibili outcome, basta contare quante volte esce una somma e la sua probabilità sarà $P(x + y) = n/36$. Un altro modo per fare questa operazione però (nonostante ce ne siano altri ancora) è proprio la convoluzione. Seguendo la formula otteniamo quindi, per calcolarci per esempio la probabilità congiunta che esca il 6:

$$\begin{aligned}
 P(x + y)_6 &= \left(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6} \right) * \left(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6} \right) = \sum_{x=0}^6 p(x)p(6-x) = \\
 &= \left(\frac{1}{6} \cdot \frac{1}{6} \right) + \left(\frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} \right) + \dots = \left(\frac{1}{36}; \frac{2}{36}; \frac{3}{36} \dots \right)
 \end{aligned}$$

E quindi ovviamente la probabilità cercata è il quinto elemento del vettore, cioè $P(x + y = 6) = 5/36$. In questo caso l'operazione in sé è relativamente semplice, ma il discorso vale a prescindere. Ecco un grafico del problema in questione (pagina seguente):

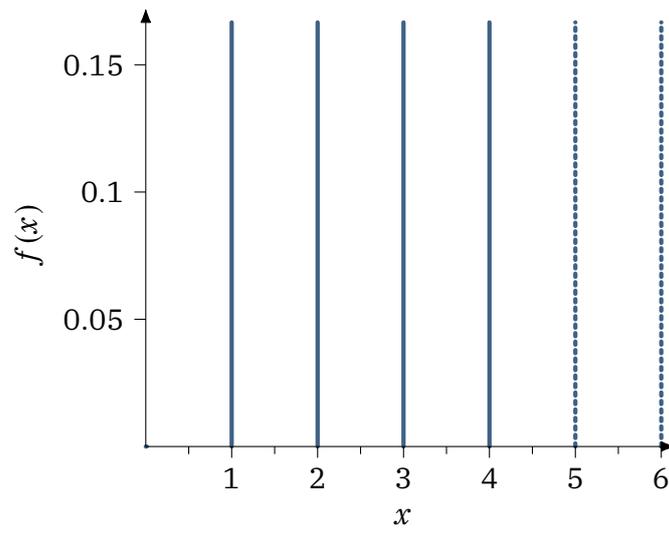


Figura 2.4: Funzioni di probabilità dei due lanci

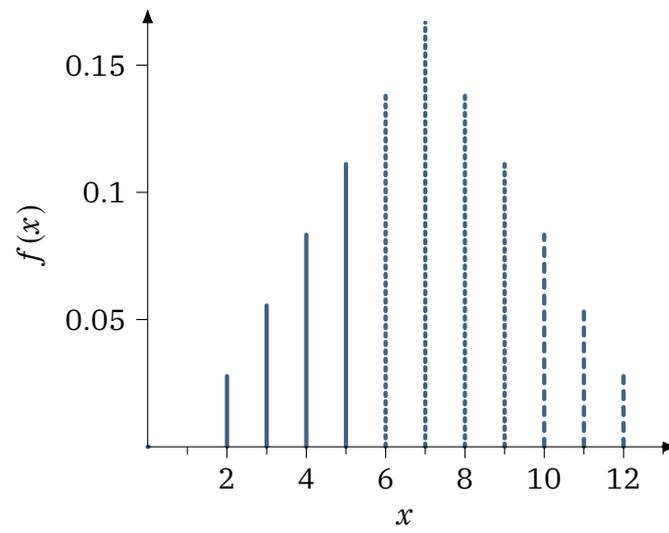


Figura 2.5: Funzione di probabilità di $P(x + y)$

2.15 Cambiamento di Variabile

Affrontiamo adesso un argomento tanto discusso quanto teatro di errori nella prova di statistica, ovvero quello del cambiamento di variabile. È in realtà relativamente banale, e basterebbe usare la formula, ma cerchiamo di capire il ragionamento che ci fa ottenere il risultato finale e quando è più intelligente adottare questo cambio.

Partiamo dalla variabile X con densità di probabilità $f_X(x)$ e consideriamo una funzione crescente $Y = g(X)$. Siccome la funzione è crescente possiamo scrivere che:

$$P(y \leq Y \leq y + dy) = P(g(x) \leq g(X) \leq g(x) + dx) = P(x \leq X \leq x + dx)$$

Tenendo a mente che ovviamente $Y = g(X)$ e $y = g(x)$. Ricordandoci il discorso sulle p.d.f. di variabili continue, questo equivale a dire:

$$\begin{aligned} f_Y(y)dy &= f_X(x)dx \\ &= \frac{f_X(x)dx}{dy} \\ &= \frac{f_X(x)}{\frac{dy}{dx}} \\ f_Y(y) &= \frac{f_X(x)}{g'(x)} \end{aligned}$$

Dove va sempre ricordato che $g(x)$ sarebbe la derivata di $g(X)$ (due cose diverse eh) ma calcolata in un Y tale che, appunto, $g(x) = y$, e quindi per quasi ovvietà $x = g^{-1}(y)$. A questo punto possiamo esprimerci tutto in funzione di y che apparentemente sembra la cosa più logica da fare e otteniamo che:

$$f_Y(y) = \frac{f_X(x)}{|dy/dx|} = \frac{f_X(g^{-1}(y))}{|g'(g^{-1}(x))|} \quad (2.15)$$

Esempio 3

Facciamo un esempio (metodo infallibile per la comprensione degli argomenti) in cui abbiamo una variabile che segue una distribuzione normale, e un'altra variabile come combinazione lineare:

$$\begin{cases} X \sim N(x|\mu_x, \sigma_x) \\ Y = aX + b \end{cases}$$

📄 Esempio 3 continued

Ora i più furbi potrebbero pensare che non ci sarebbe bisogno di farsi formule con indici uguali che ci confono e usare semplicemente le formule che ci siamo ricavati per le combinazioni lineari, e avreste ragione, ma siccome è un argomento del corso utilizziamo l'Equazione 2.15:

$$\begin{aligned}
 f_Y(y) &= \frac{f_X\left(\frac{y-b}{a}\right)}{|a|} \\
 &= \frac{1}{|a|\sigma_x\sqrt{2\pi}} \cdot e^{-\frac{(y-b-a\mu_x)^2}{2a^2\sigma^2}}
 \end{aligned}$$

E quindi da qui possiamo ricavarci valore atteso e varianza della nostra nuova variabile e lavorare *comodamente* con questa.

3

Distribuzioni



Sommarrio

- 3.1 Distribuzione Uniforme, 47
- 3.2 Distribuzioni Triangolari, 49
- 3.3 Distribuzione di Bernoulli, 50
- 3.4 Distribuzione Geometrica, 52
- 3.5 Distribuzione Binomiale, 55
- 3.6 Distribuzione Multinomiale, 58
- 3.7 Distribuzione di Pascal, 61
- 3.8 Distribuzione di Poisson, 63
- 3.9 Distribuzione Normale, 66
- 3.10 Distribuzione del χ^2 , 72
- 3.11 Distribuzione Esponenziale, 77
- 3.12 Distribuzione del tempo di attesa del primo successo, 78
- 3.13 Distribuzione di Erlang, 80
- 3.14 Distribuzione t-student, 82

Diversi tipi di misure hanno diverse distribuzioni limite. Non tutte hanno la classica forma a campana, ad esempio esistono la distribuzione di Poisson, binomiale, esponenziale...

In questo capitolo andremo a esplorare vari tipi di Distribuzione e i loro specifici valori attesi, varianze...

3.1 Distribuzione Uniforme

La distribuzione uniforme è la più semplice delle distribuzioni continue, in cui la variabile che assume valori in un certo intervallo $X = [a; b]$ ha sempre la stessa probabilità k , quindi in breve

$$f(x|k) = k$$

È chiaro che non può succedere che $x \in \mathbb{R}$, poichè avrei una probabilità totale infinita, mentre sappiamo che l'integrale sul dominio deve fare 1. Proprio con questo presupposto, siamo in realtà in grado di calcolarci il valore esatto di k , siccome:

$$\begin{aligned} 1 &= \int_{\Omega} f(x) dx = \int_a^b k dx \\ &= k \cdot \int_a^b dx = k \cdot x \Big|_a^b = k \cdot (b - a) \\ k &= \frac{1}{b - a} \end{aligned}$$

Possiamo quindi dare una definizione completa di questa distribuzione:

Definizione 13

La **Distribuzione uniforme** è la più semplice distribuzione continua, in cui $x \in [a; b]$ e la probabilità è costante.

$$f(x|k) = \frac{1}{b - a} \quad (3.1)$$

A questo punto calcolarsi valore atteso e varianza è molto semplice, si tratta di due integrali banali:

$$\begin{aligned} \mathbb{E}[x] &= \int_{\Omega} x \cdot f(x) dx = \int_a^b x \cdot \frac{1}{b - a} dx \\ &= \frac{1}{b - a} \cdot \int_a^b x dx = \frac{b^2 - a^2}{2(b - a)} = \frac{\cancel{(b - a)}(b + a)}{2\cancel{(b - a)}} \\ \mathbb{E}[x] &= \frac{b + a}{2} \end{aligned}$$

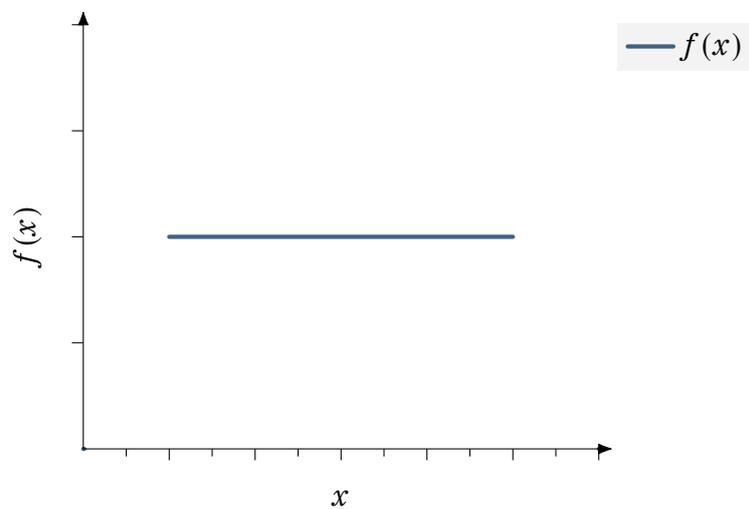
Mentre per quanto riguarda la deviazione standard, calcoliamoci prima $\mathbb{E}[x^2]$:

$$\begin{aligned}\mathbb{E}[x^2] &= \int_{\Omega} x^2 \cdot f(x) dx = \int_a^b x^2 \cdot \frac{1}{b-a} dx \\ &= \frac{1}{b-a} \cdot \int_a^b x dx = \frac{b^3 - a^3}{3(b-a)} = \frac{(b-a)(b^2 + ab + a^2)}{3(b-a)} \\ \mathbb{E}[x^2] &= \frac{b^2 + ab + a^2}{3}\end{aligned}$$

E quindi:

$$\sigma[x] = \mathbb{E}[x^2] - \mathbb{E}^2[x] = \frac{b-a}{\sqrt{12}} \quad (3.2)$$

Una distribuzione uniforme ha un grafico molto lineare:



3.2 Distribuzioni Triangolari

Ragionevolmente nella stragrande maggioranza dei casi la variabile casuale è definita in modo tale che i gradi di fiducia decrescono linearmente dal centro fino a raggiungere gli estremi, si ha così la distribuzione di Simpson. Molto spesso questa distribuzione è utile rispetto a quella uniforme perchè più realistico come modello. Adesso ragioniamo su due casi:

- Il valore atteso della distribuzione si trova esattamente al centro di questa.
- Il valore atteso della distribuzione è x_0 ma non corrisponde al centro dell'intervallo.

Nel primo caso abbiamo che il massimo della distribuzione sarà $\frac{1}{\Delta}$ il valore atteso x_0 e l'incertezza sarà:

$$\sigma = \frac{\Delta}{\sqrt{6}} \quad (3.3)$$

Nel secondo caso invece abbiamo che se la distribuzione è definita in un intervallo a e b , allora abbiamo che:

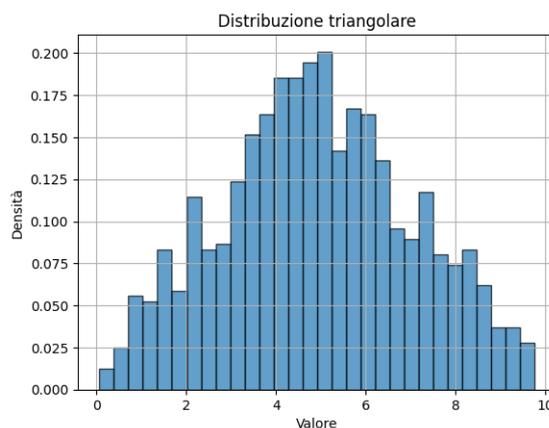
$$\Delta_+ = x_0 - a \quad (3.4)$$

$$\Delta_- = b - x_0 \quad (3.5)$$

il cui valore atteso sarà:

$$E[X] = x_0 + \frac{\Delta_+ - \Delta_-}{3} \quad (3.6)$$

$$\sigma^2 = \frac{\Delta_+^2 + \Delta_-^2 + \Delta_+ \Delta_-}{18} \quad (3.7)$$



3.3 Distribuzione di Bernoulli

La distribuzione di Bernoulli è una distribuzione di probabilità che permette di studiare eventi che hanno solo due possibili esiti: successo o insuccesso (fido). Dato che ci sono due possibili eventi, la variabile casuale associata ha solo due possibilità, per questo viene anche chiamata binaria, o Bernoulliana.

La variabile casuale X , associata a questi due parametri può assumere solo due valori:

$$\begin{cases} P(X = 1) = p \\ P(X = 0) = 1 - p \end{cases}$$

Dunque possiamo definire la distribuzione attraverso la sua funzione di probabilità come segue:

Definizione 14

La distribuzione di Bernoulli è una distribuzione di probabilità di una variabile discreta x che ha come unici valori possibili 0 e 1.

$$B(x|p) = p^x(1-p)^{1-x}, \begin{cases} 0 \leq p \leq 1 \\ x \in [0, 1] \end{cases} \quad (3.8)$$

Per verificare che la funzione appena scritta sia proprio quella che rappresenta la distribuzione di Bernoulli, si può semplicemente sostituire il valore $X = 1$ o $X = 0$, per ritrovare quanto scritto pocanzi. Trovare il valore atteso per questa distribuzione è banale, infatti sappiamo che $\mathbb{E}[x]$:

$$\mathbb{E}[x] = 1p + 0(1-p) = p \quad (3.9)$$

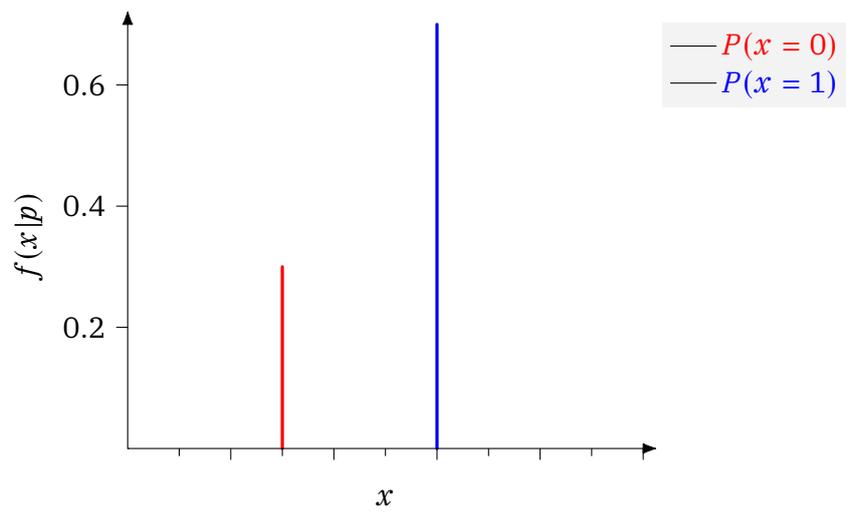
Questo perchè, riprendendo la definizione di valore atteso, ovvero l'Equazione 2.2, la calcoliamo semplicemente così:

$$\mathbb{E}[x] = \sum_{i=0}^n x_i \cdot f(x_i) = \left[0 \cdot (1-p) \right] + \left[1 \cdot p \right] = p$$

Possiamo anche calcolarne la varianza, che da antiche conoscenze sulla cultura sumera sappiamo essere:

$$\text{Var}[x] = \mathbb{E}[x^2] - \mathbb{E}^2[x] = p(1-p) \quad (3.10)$$

Una distribuzione di questo tipo avrà un grafico del genere:



3.4 Distribuzione Geometrica

La distribuzione geometrica descrive il numero di insuccessi necessari affinché si verifichi il primo successo in una successione di esperimenti bernoulliani indipendenti ed identicamente distribuiti. Questa distribuzione, che ha quindi come supporto l'insieme dei numeri naturali, è completamente descritta specificando il parametro p dell'esperimento bernoulliano.

Tenendo conto dell'indipendenza tra le ripetizioni dell'esperimento bernoulliano, richiedere che il primo successo sia avvenuto esattamente dopo i insuccessi equivale a calcolare la probabilità di i insuccessi successivi. Pertanto la distribuzione geometrica può essere definita come segue:

Definizione 15

La distribuzione Geometrica descrive il numero di insuccessi necessari affinché si verifichi un successo di una serie di eventi di Bernoulli indipendenti. La sua p.d.f. è:

$$\mathcal{G}(x|p) = p(1-p)^{x-1}; \quad x \in \mathbb{N} \quad (3.11)$$

Dimostrazione 2: p.d.f. di $\mathcal{G}(x|i, p)$

Ovviamente la probabilità di ogni evento preso esclusivamente è p , quello che cambia è la probabilità che capitino il secondo successo una volta capitato il primo e così via, infatti si vede facilmente che:

$$\mathcal{G}(X = 1) = P(E_1) = p$$

$$\mathcal{G}(X = 2) = P(E_2|\bar{E}_1) \cdot P(\bar{E}_1) = p(1-p)$$

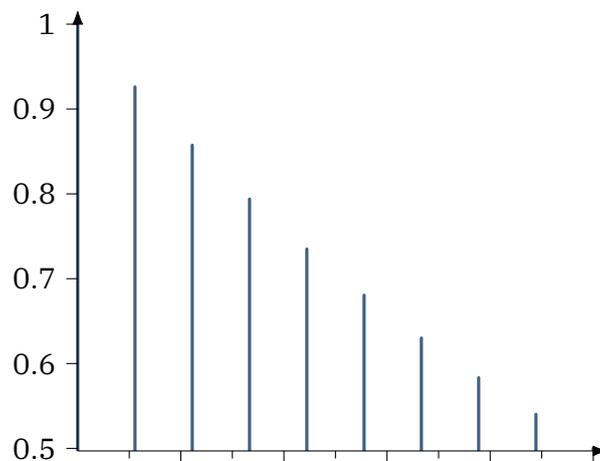
$$\vdots$$

$$\mathcal{G}(X = x) = P(E_x|\bar{E}_1 \dots \bar{E}_{x-1}) \cdot \prod_{i=1}^{x-1} P(\bar{E}_i) = p(1-p)^{x-1}$$

Si verifica facilmente che sommando i valori di densità di probabilità si ottiene come risultato:

$$\sum_{x=0}^{+\infty} \mathcal{G}(x|p) = \sum_{x=0}^{+\infty} p(1-p)^{x-1} = p \sum_{x=0}^{+\infty} (1-p)^{x-1} = p \cdot \frac{1}{1-(1-p)} = 1$$

E questo è anche il motivo per cui si chiama geometrica: ovvero segue il comportamento di una serie geometrica, infatti graficamente una distribuzione geometrica è qualcosa del tipo:



3.4.1 Proprietà della distribuzione Geometrica

Per trovarci il valore atteso basta fare qualche manipolazione algebrica:

$$\begin{aligned}\mathbb{E}[x] &= \sum_{x=1}^{+\infty} x \cdot \mathcal{G}(x|i, p) = \sum_{x=1}^{+\infty} x \cdot p \cdot (1-p)^{x-1} = \\ &= p \cdot \sum_{x=1}^{+\infty} \frac{d}{dq}(q^x) = p \cdot \frac{d}{dq} \left(\sum_{x=1}^{+\infty} q^x \right) = \\ &= p \cdot \frac{d}{dq} \left(\frac{1}{-p} \right) = \frac{p}{p^2} = \frac{1}{p}\end{aligned}$$

Concludiamo quindi che

$$\mathbb{E}[x] = \frac{1}{p} \quad (3.12)$$

La Varianza invece è:

$$\text{Var}[x] = \frac{1-p}{p^2} \quad (3.13)$$

3.4.2 Mancanza di memoria

La mancanza di memoria della distribuzione geometrica è una proprietà interessante: essa afferma che una misura qualsiasi all'interno della successione E_i non ha memoria degli eventi prima di esso, quindi il suo comportamento futuro è uguale a quello di un evento "nuovo", il che non vuol dire che la probabilità è sempre la stessa, ma che i tempi di attesa lo sono, e quindi il comportamento della funzione è sempre lo stesso:

Come dicevamo proprietà interessante, ma ridondante da dimostrare e non troppo utile ai fini del corso, ma lo faremo lo stesso:

$$\begin{aligned}P(X \geq x) &= P(X \geq |x|) = P(X \geq |x| - 1) = (1-p)^{|x|} \\ P(X \geq x+y | X \geq x) &= \frac{P(X \geq x+y, X \geq x)}{P(X \geq x)} = \frac{(1-p)^{|x|+|y|}}{(1-p)^{|x|}} = (1-p)^{|y|} = P(X \geq y)\end{aligned}$$

3.5 Distribuzione Binomiale

Prima di definire la Distribuzione Binomiale facciamo un esempio:

Esempio 4

Supponiamo adesso di intraprendere un "esperimento", che consiste nel lanciare tre dadi e contare il numero di 6 usciti. I possibili risultati dell'esperimento sono le uscite 0,1,2,3 facce 6. Se ripetiamo l'esperimento un enorme numero elevato di volte, allora troveremo la distribuzione limite, che ci dirà la probabilità che in ogni lancio si ottengano $v = 0, 1, 2, 3$.

Supponendo che il dado non sia truccato, la probabilità di ottenere un 6 quando gettiamo un dado è $1/6$. Lanciamo adesso i tre dadi contemporaneamente e calcoliamo la probabilità di ottenere tre 6 $v = 3$, sarà:

$$P(v = 3) = \left(\frac{1}{6}\right)^3$$

Calcoliamo adesso la probabilità che escano due 6, questo risulterà essere più complicato. Poiché possiamo ottenere due 6, da parecchie combinazioni, ragioniamo prima sul caso in cui nei primi due dadi esca un 6 e nel terzo no:

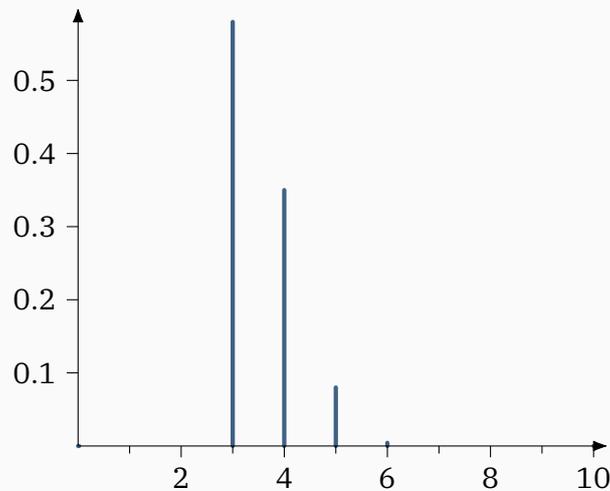
$$P(A, A, \bar{A}) = \left(\frac{1}{6}\right)^2 \cdot \left(\frac{5}{6}\right)$$

La probabilità per due 6 in qualche altro ordine definito è la stessa. Infine, ci sono tre diversi ordini in cui potremmo ottenere i nostri due 6. Così la probabilità totale di ottenere due 6 in qualsiasi ordine è:

$$P(2 \text{ "6" in 3 lanci}) = 3 \cdot \left(\frac{1}{6}\right)^2 \cdot \frac{5}{6} \approx 6.9\%$$

📄 Esempio 4 continued

Il grafico della distribuzione in questione è del tipo:



3.5.1 Definizione della binomiale

Dopo aver ottenuto questi simpaticissimi dati, proviamo ad abbozzare la probabilità di ottenere v successi in n prove. Questa probabilità come vedremo è data dalla cosiddetta distribuzione binomiale:

$$P(v) = B_{n,p}(v) = \frac{n(n-1)\dots(n-v+1)}{1 \cdot 2 \cdot \dots \cdot v} \cdot p^v \cdot q^{n-v} \quad (3.14)$$

L'Equazione 3.14 è detta distribuzione binomiale, per il suo legame con lo sviluppo binomiale:

$$\binom{n}{v} = \frac{n(n-1)\dots(n-v+1)}{1 \cdot 2 \cdot \dots \cdot v} = \frac{n!}{v! \cdot (n-v)!} \quad (3.15)$$

dove si nota l'introduzione del fattoriale $n! = 1 \times 2 \times 3 \times \dots \times n$. Dunque introdotti questi elementi possiamo riscrivere la funzione di distribuzione binomiale e definirla:

📖 Definizione 16

La distribuzione Binomiale è una distribuzione discreta che sancisce il numero di successi in un processo di Bernoulli.

$$B(x|n, p) = \binom{n}{x} \cdot p^x \cdot q^{n-x} \quad (3.16)$$

3.5.2 Proprietà della distribuzione Binomiale

Vogliamo adesso calcolare la media e la varianza della distribuzione binomiale come abbiamo fatto anche con le altre distribuzioni:

✓ Dimostrazione 3

Il **Valore Atteso** della distribuzione binomiale la otteniamo:

$$\mathbb{E}[x] = \sum_{x=0}^N x \cdot \frac{n!}{x!(n-x)!} \cdot p^x \cdot q^{n-x}$$

Sapendo che $n! = n(n-1)!$ allora

$$\mathbb{E}[x] = \sum_{x=0}^N x \cdot \frac{n(n-1)!}{x(x-1)! \cdot [n-1-(x-1)]!} \cdot p^x (1-p)^{n-x}$$

Adesso poniamo $y = x - 1$, e troviamo:

$$np \cdot \sum_{y=0}^{n-1} \frac{(n-1)!}{y! \cdot (n-1-y)!} \cdot p^y (1-p)^{n-1-y}$$

Tenendo presente che:

$$\sum_{y=0}^{n-1} \frac{(n-1)!}{y!(n-1-y)!} p^y (1-p)^{n-1-y} = 1$$

Poichè è la somma delle probabilità di una distribuzione binomiale con parametri $n-1$ e p , seguendo si ottiene:

$$\mathbb{E}[X] = np \tag{3.17}$$

La varianza risulta molto più lunga e noiosa da calcolare, quindi lasciamo il calcolo esplicito in [Sezione 8.3](#) e ci limitiamo semplicemente a riportarne il valore:

$$\sigma^2 = p \cdot (1-p) \cdot n \tag{3.18}$$

3.6 Distribuzione Multinomiale

La distribuzione multinomiale è una distribuzione di più variabili discrete, generalizzazione della distribuzione binomiale dove l'evento può realizzarsi in più di due modalità. In generale, avendo m possibili modalità di realizzazione in ciascuna prova, chiameremo p_i la probabilità associata a ciascuna modalità. Naturalmente vale la condizione

$$\sum_{i=1}^m p_i = 1$$

Le prove totali effettuate sono sempre n , come nelle binomiale, ma cambia che avendo possibili modalità di uscita non ci sono più $1^n = 1$ sequenze possibili ma bensì m^n in genere non equiprobabili. Chiamiamo a questo punto X_1, X_2, \dots, X_m le variabili casuali che rappresentano il numero di volte in cui si è osservato un determinato outcome dei possibili. È chiaro che $X_1 = x_1, X_2 = x_2, \dots, X_m = x_m$ in cui $\sum_{i=1}^m x_i = n$, ovvero immaginiamo gli n lanci come un vettore con m coefficienti, alla quale associamo ad ognuno le variabili casuali discusse prime.

Con queste premesse teoriche che possono sembrare ostiche, le sequenze diverse che si possono produrre con stessa probabilità sono:

$$p_1^{x_1} \cdot p_2^{x_2} \cdot \dots \cdot p_m^{x_m}$$

Allo stesso modo in cui avevamo costruito la binomiale, consideriamo ora il numero delle sequenze che produce la stessa configurazione, che sono

$$\frac{n!}{x_1! x_2! \dots x_m!}$$

E questo perchè avrei $\binom{n}{x_1}$ combinazioni per la prima modalità, $\binom{n-x_1}{x_2}$ per la seconda e così via, quindi si semplificano:

$$\frac{n!}{\cancel{(n-x_1)!} x_1!} \cdot \frac{\cancel{(n-x_1)!}}{\cancel{(n-x_1-x_2)!} x_2!} \cdot \dots$$

E quindi arriviamo finalmente alla definizione di questa distribuzione:

Definizione 17

La **Distribuzione Multinomiale** è una distribuzione in più variabili discrete $x_i \in \mathbb{N} \forall i = 1 \dots m$ ed è una generalizzazione della binomiale con un evento che si può verificare in più modalità.

$$\mathcal{M}(\vec{x}|m, n, \vec{p}) = \frac{n!}{x_1!x_2!\dots x_m!} \cdot p_1^{x_1} \cdot p_2^{x_2} \cdot \dots \cdot p_m^{x_m} \quad (3.19)$$

In cui ovviamente $\vec{x} = \{x_1, x_2 \dots x_m\}$ e $\vec{p} = \{p_1, p_2 \dots p_m\}$

Da qui non ci calcoliamo il valore atteso globale ma quello di una singola variabile x_i che possiamo trattare come una binomiale semplice, verrà fuori che

$$\mathbb{E}[x_i] = np_i \quad (3.20)$$

$$\text{Var}[x_i] = np_i(1 - p_i) \quad (3.21)$$

Si aggiungono poi in questo caso Covarianza e coefficiente di correlazione chiaramente, che valgono esattamente:

$$\text{Cov}[X_i, X_j] = -np_i p_j \quad (3.22)$$

$$\rho(X_i, X_j) = -\sqrt{\frac{p_i p_j}{(1 - p_i)(1 - p_j)}} \quad (3.23)$$

Vediamo subito come ci si arriva:

Dimostrazione 4

Partiamo da un caso più semplice, siccome al fine del corso a quanto pare non è importante il conto preciso (e meno male): sappiamo che la binomiale è della forma

$$B(x|n, p) = \frac{n!}{x!(n-x)!} \cdot p^x \cdot (1-p)^{n-x}$$

Se adesso operiamo delle sostituzioni ci ritroviamo con una multinomiale a due

✓ Dimostrazione 4 continued

dimensioni che in realtà dipende da un parametro solo:

$$\begin{aligned}x &\rightarrow x_1; \\n - x &\rightarrow x_2; \\p &\rightarrow p_1; \\1 - p &\rightarrow p_2\end{aligned}$$

Ci ritroviamo con

$$f(x_1, x_2) = \frac{n!}{x_1!x_2!} \cdot p_1^{x_1} \cdot p_2^{x_2}$$

Quando diciamo che in realtà dipende da un solo parametro intendiamo che x_1, x_2 sono linearmente (anti)-correlate, cioè soddisfano sempre la condizione $x_1 + x_2 = n$ e quindi scelto un valore, l'altro è obbligato ed è anche chiaro che se la loro correlazione è massima, ovvero $\rho(x_1, x_2) = -1$ Da qui calcolare la covarianza è semplice, a partire dalla formula del coefficiente di correlazione

$$\text{Cov}[X_1, X_2] = \rho(X_1, X_2) \cdot \sigma_{x_1} \sigma_{x_2} = -np_1p_2$$

Siccome si può verificare che quest'ultima è valida anche nel caso generale, a partire dalla generica covarianza ci ricaviamo il coefficiente per il caso generico:

$$\text{Cov}[X_i, X_j] = -np_i p_j$$

$$\rho(X_i, X_j) = -\sqrt{\frac{p_i p_j}{(1 - p_i)(1 - p_j)}}$$

3.7 Distribuzione di Pascal

Abbiamo in precedenza studiato la distribuzione geometrica, legata al numero aleatorio “tentativo per il quale si registra il primo successo”, quando si considerano tanti processi di Bernoulli di uguale p . Il caso più generale è quello che descrive la distribuzione di probabilità del “tentativo per il quale si registrano esattamente k successi”. Essa è nota come distribuzione di Pascal.

Negli $x - 1$ tentativi precedenti si devono essere verificati $k - 1$ successi e $x - k$ insuccessi, indipendentemente dall'ordine. La probabilità di questo evento si ottiene dalla distribuzione binomiale di parametri:

$$P(x|p, k) = B(k - 1|p, x - 1) = \binom{x - 1}{k - 1} \cdot p^{k-1}(1 - p)^{x-k}$$

La definiamo quindi in questo modo:

Definizione 18

La distribuzione di Pascal descrive la probabilità per il quale si verificano esattamente k successi in una serie di processi di Bernoulli indipendenti di uguale probabilità.

$$P(x|p, k) = \binom{x - 1}{k - 1} \cdot p^{k-1}(1 - p)^{x-k} \quad (3.24)$$

Siccome ci sono voluti x tentativi per avere k successi, possiamo vedere la distribuzione di Pascal in questione come k variabili indipendenti che seguono una geometrica di parametro x (ed ecco perchè abbiamo lasciato la confusa scrittura della dipendenza da x), ovvero

$$P(x|p, k) = k \cdot \mathcal{G}(x|p)$$

Allora il valore atteso, la varianza e la deviazione standard vengono direttamente dalla geometrica, basta moltiplicare per k :

$$\mathbb{E}[x] = k \cdot \frac{1}{p} = \frac{k}{p} \quad (3.25)$$

$$\text{Var}[x] = k \frac{q}{p^2} = \frac{kq}{p^2} \quad (3.26)$$

$$\sigma_x = \frac{\sqrt{kq}}{p} \quad (3.27)$$

Dal punto di vista grafico la distribuzione di pascal è del tipo:

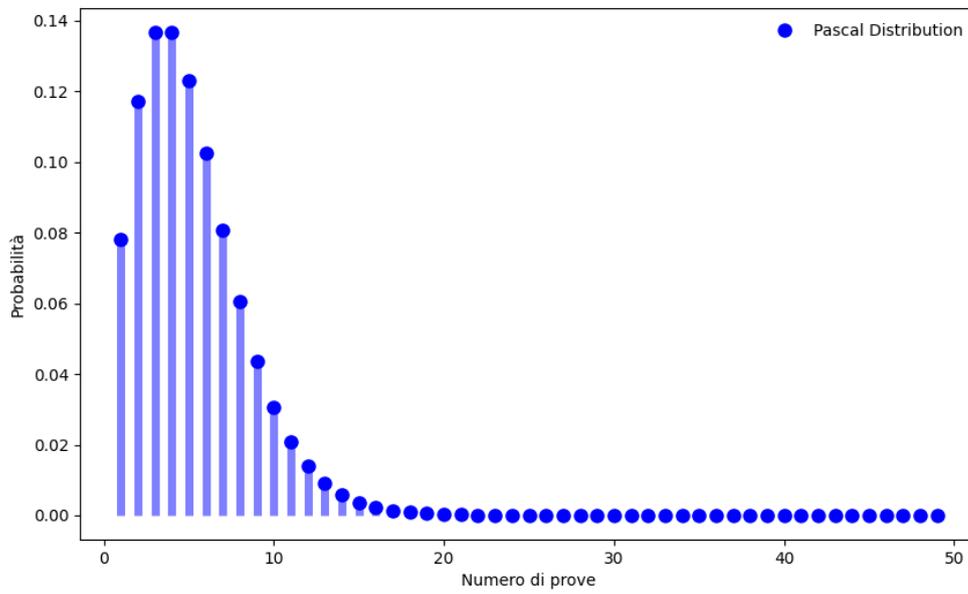


Figura 3.1: Distribuzione con $k = 5$ successi richiesti e $p = 0.5$

3.8 Distribuzione di Poisson

Mettiamoci adesso nel caso limite di una distribuzione Binomiale, ovvero abbiamo che $n \rightarrow +\infty$ e al contrario la probabilità $p \rightarrow 0$, questo è il caso ad esempio del numero di nuclei all'intero di un atomo radioattivo circa 10^{20} , rapportato con la probabilità di osservare un decadimento in un certo intervallo ΔT che risulta essere all'incirca 10^{-20} s. Il valore atteso di questa distribuzione è 1, infatti $n \cdot p = 1$. È interessante notare che la distribuzione non dipenda dagli esatti valori di n e di p , ma solo dal prodotto.

È allora possibile valutare la distribuzione di probabilità senza conoscere né n né p e per fare questo mettiamoci nel caso limite per cui n tende ad infinito e p a 0.

 **Dimostrazione 5:** Distribuzione di Poisson come limite della Binomiale

$$f(x|B_{n,p}) = \binom{n}{x} \cdot p^x \cdot q^{n-x}$$

Sostituiamo adesso $p = \frac{\lambda}{n}$ e otteniamo:

$$f(x|B_{n,p}) = \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \cdot \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

Ora consideriamo il limite per $n \rightarrow \infty$:

$$\lim_{n \rightarrow \infty} \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \cdot \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

Portiamo fuori dal limite λ^k e $x!$ (quest'ultimo dallo sviluppo del binomio di Newton):

$$\rightarrow \frac{\lambda^x}{x!} \cdot \lim_{n \rightarrow \infty} \binom{n}{x} \left(\frac{1}{n}\right)^x \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-x}$$

E sviluppiamo il binomio di Newton:

$$\frac{\lambda^x}{x!} \lim_{n \rightarrow \infty} \frac{n(n-1)(n-2)\dots(n-x+1)\cancel{(n-x)!}}{\cancel{(n-x)!}} \left(\frac{1}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x}$$

Semplifichiamo adesso $(n-x)!$ al numeratore e al denominatore e otteniamo:

$$\frac{\lambda^x}{x!} \cdot \lim_{n \rightarrow \infty} \frac{n^x + \dots}{\underset{\rightarrow 1}{n^x}} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x}$$

Adesso è rimasto il limite notevole $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n$ che risulta essere uguale a $e^{-\lambda}$, dunque otteniamo una distribuzione di probabilità caratterizzata dal solo parametro $\lambda \in \mathbb{R}^+$.

Definizione 19

La distribuzione di Poisson è una distribuzione che indica la probabilità di x conteggi o eventi in un dato intervallo Δt sapendo che in media se ne verificano λ . La sua pdf è

$$P(x|\lambda) = \frac{\lambda^x}{x!} \cdot e^{-\lambda} \quad (3.28)$$

Di questa distribuzione possiamo trovare il valore atteso, la varianza e il "rate" (dimostrazioni in [Sezione 8.5](#)) che risulteranno essere:

$$\mathbb{E}[X] = \lambda \quad (3.29)$$

$$\sigma[x] = \sqrt{\lambda} \quad (3.30)$$

$$v = \frac{1}{\lambda} \quad (3.31)$$

La poissoniana gode inoltre della proprietà riproduttiva (dimostrazione in [Sezione 8.7](#)), ovvero:

Teorema 8: Riproduttività della Poissoniana

Date Y_1, Y_2 tali che $Y_i \sim P(x_i|\lambda_i)$, allora la variabile definita come somma $Y = Y_1 + Y_2$ seguirà una poissoniana con $\lambda = \lambda_1 + \lambda_2$, e in generale date n variabili casuali con le stesse caratteristiche, una variabile somma qualunque segue la stessa proprietà, quindi

$$\sum_{i=1}^n Y_i = Y \sim P(x|\lambda); \quad \lambda = \sum_{i=1}^n \lambda_i$$

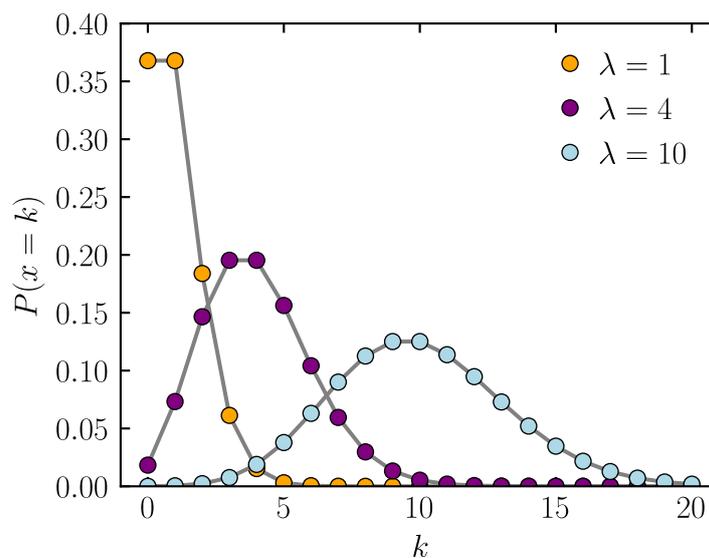


Figura 3.2: Esempi di Distribuzione di Poisson

3.8.1 Applicazioni

Ciascuno dei fenomeni descritti dalla poissoniana può verificarsi nel dominio del tempo o dei conteggi, per rendere il tutto più facile negli esercizi nella stragrande maggioranza dei casi si parla proprio di problemi nel dominio del tempo. Ci interessiamo dunque al numero di conteggi in un determinato intervallo di tempo che indicheremo come ΔT ovvero la variabile aleatoria X è definita come: "numero di conteggi fra 0 e t ".

Adesso supponiamo che la probabilità che si verifichi un conteggio sia proporzionale a ΔT :

$$p = r \Delta T$$

Dove r è il tasso di conteggi in un determinato intervallo di tempo. La probabilità che in Δt si verifichino più di 1 eventi sia trascurabile in confronto a quella che se ne verifichi esattamente 1:

$$P(\text{più di 1 evento}) \approx 0 \quad \text{rispetto a} \quad P(\text{esattamente 1 evento})$$

Il numero di conteggi in un intervallo finito sia indipendente dal numero di conteggi che si verificano in un altro intervallo, se i due intervalli sono disgiunti:

$$P(\text{conteggi in } \Delta t_1 | \text{conteggi in } \Delta t_2) = P(\text{conteggi in } \Delta t_1)$$

Consideriamo n intervallini disgiunti, ciascuno di durata Δt , tali che $t = n\Delta t$, ovvero $\Delta t = \frac{t}{n}$. Quando n tende ad infinito, $\Delta t \rightarrow 0$ e di conseguenza $p \rightarrow 0$. Consideriamo inoltre che:

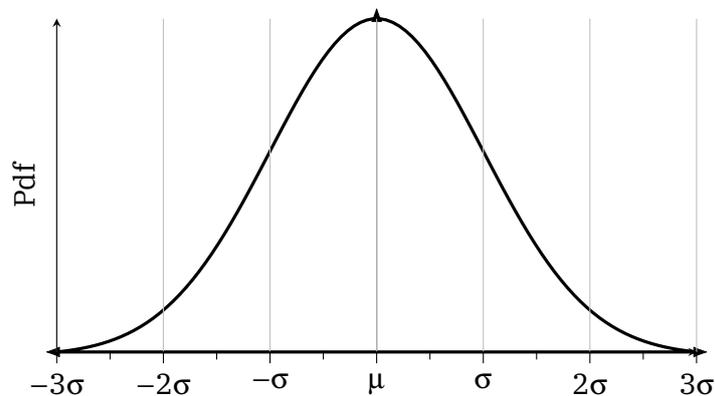
- In ogni intervallino, l'evento "accade un conteggio" può essere considerato come un processo di Bernoulli indipendente dagli altri;
- Essendo r costante e quindi p costante, i n processi di cui ci interessiamo, relativi al numero di successi, danno luogo a una binomiale;
- La condizione $n \rightarrow \infty$ (e conseguente $p \rightarrow 0$) rende la binomiale approssimabile da una poissoniana:

$$f(x|B(n, p)) \rightarrow f(x|P(\lambda)),$$

Con $\lambda = np = rt$, dove r rappresenta il numero atteso di conteggi per unità di tempo, ovvero l'intensità del processo.

3.9 Distribuzione Normale

La distribuzione normale si basa sulle misure che si distribuiscono su una curva a campana e questa sarà centrata sul valore vero di x . Se le nostre misure hanno errori sistematici apprezzabili, allora non dovremmo aspettarci che la distribuzione limite sia centrata sul valore vero. Gli errori casuali spingono con la stessa probabilità le nostre letture sopra o sotto il valore vero. Se gli errori sono casuali, dopo molte misure ci saranno tante osservazioni al di sopra del valore vero, quante al di sotto e la nostra distribuzione sarà centrata sul valore vero. Gli errori sistematici però, come quello provocato da un metro a nastro, spingono tutti i valori in una direzione, e portano la distribuzione lontana dal valore vero. In questa parte assumeremo allora che gli errori sistematici siano trascurabili.



In primis dobbiamo chiederci cos'è il valore vero di una grandezza, domanda alla quale non abbiamo una risposta soddisfacente poiché di nessuna misura si può determinare esattamente il valore atteso di una variabile continua. Tuttavia è conveniente assumere che ogni misura abbia un valore vero, al fine quanto meno di rappresentarla con una distribuzione continua (ricordiamoci che "variabile continua" è un termine che ha poco a che fare con la realtà e tanto con la matematica, ma è il miglior modello disponibile per descrivere certe situazioni, e come ogni cosa nella statistica, per funzionare ha bisogno di tante approssimazioni e assunzioni campate per aria).

Denoteremo i valori veri delle grandezze misurate x, y con le rispettive lettere maiuscole μ_x, μ_y . Se le misure di x sono soggette a molti piccoli errori casuali, ma trascurabili errori sistematici, allora la loro distribuzione sarà una curva a campana, centrata sul valore vero μ_x .

La funzione matematica che descrive la curva a campana è chiamata distribuzione normale o funzione di Gauss. Il prototipo di questa funzione è:

$$e^{-\frac{x^2}{2\sigma^2}} \quad (3.32)$$

Dove σ è un parametro fisso che chiameremo parametro di larghezza. Quando $x = 0$ la funzione di Gauss è uguale ad 1. La funzione è simmetrica intorno ad $x = 0$, dal momento che assume lo stesso valore per x e $-x$. Quando x si allontana dallo zero in ognuna delle due direzioni, $x^2/2\sigma^2$ aumenta: rapidamente se σ è piccolo, più lentamente se è grande. Man mano che la x si allontana dall'origine, la funzione decresce fino a tendere a zero.

Per ottenere una curva a campana centrata attorno a qualche valore $x = X$, sostituiamo la x nella 3.32 con $x - \mu_x$. Così la funzione ha il massimo per $x = X$:

$$e^{-\frac{(x-\mu_x)^2}{2\sigma^2}} \quad (3.33)$$

La funzione 3.33 non è però ancora normalizzata, ovvero non soddisfa la condizione:

$$\int_{-\infty}^{+\infty} f(x) dx = 1 \quad (3.34)$$

Per fare questo poniamo

$$f(x) = N \cdot e^{-\frac{(x-\mu_x)^2}{2\sigma^2}}$$

Dobbiamo scegliere accuratamente N , in modo tale che $f(x)$ sia normalizzata come nell'Equazione 3.34, allora ci calcoliamo:

$$\int_{-\infty}^{+\infty} f(x) dx = \int_{-\infty}^{+\infty} N e^{-\frac{(x-\mu_x)^2}{2\sigma^2}} dx$$

Per calcolare questo tipo di integrale è sempre una buona idea cambiare le variabili per semplificarlo. Così possiamo porre $x - \mu_x = t$ e otteniamo:

$$\int_{-\infty}^{+\infty} f(x) dx = N \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2\sigma^2}} dt$$

In seguito poniamo $\frac{t}{\sigma} = z$ ed otteniamo:

$$N\sigma \int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}} dz$$

L'integrale ottenuto è uno degli integrali standard della fisica matematica. I dettagli non sono interessanti, il procedimento è spiegato nel dettaglio in [Sezione 8.1](#). Il risultato è:

$$\int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}} dz = \sqrt{2\pi} \quad (3.35)$$

Ritornando alle 3.32 e 3.33 troviamo quindi che:

$$\int_{-\infty}^{+\infty} f(x) dx = N\sigma\sqrt{2\pi} = 1$$

Dal momento che questo integrale deve essere uguale ad 1, dobbiamo scegliere il fattore di normalizzazione in modo che:

$$N = \frac{1}{\sigma\sqrt{2\pi}}$$

Con questa scelta per il fattore di normalizzazione, arriviamo alla forma definitiva per la funzione di distribuzione di Gauss che denotiamo con $G(x|\mu_x, \sigma)$ e possiamo darne una definizione abbastanza accurata e giustificata.

Definizione 20

La funzione di distribuzione normale, o di Gauss normalizzata, è una funzione che segue l'andamento di un parametro x attorno a un suo valore vero μ , dato il parametro di larghezza σ . Essa si esprime come:

$$G(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (3.36)$$

Sappiamo che la conoscenza della distribuzione limite per una misura ci permette di calcolare il valore medio \bar{x} dopo un gran numero di prove. La media attesa per la distribuzione di Gauss è:

$$\bar{x} = \int_{-\infty}^{+\infty} x \cdot G(x|\mu, \sigma) dx \quad (3.37)$$

E questo perchè sappiamo dall'[Equazione 2.2](#) nel caso continuo che $\mathbb{E}[x] = \int x \cdot f(x) dx$.

Oltre a questo, possiamo calcolare l'integrale di [Equazione 3.37](#) per la distribuzione di Gauss come segue:

$$\bar{x} = \int_{-\infty}^{+\infty} x \cdot G(x|\mu, \sigma) dx = \frac{1}{\sigma\sqrt{2\pi}} \cdot \int_{-\infty}^{+\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Se facciamo il cambio di variabili $\begin{cases} t = x - \mu \\ x = t + \mu \\ dt = dx \end{cases}$. Così l'integrale diventa la somma di due termini:

$$\bar{x} = \frac{1}{\sigma\sqrt{2\pi}} \left(\int_{-\infty}^{+\infty} t e^{-\frac{t^2}{2\sigma^2}} dt + \mu \cdot \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2\sigma^2}} dt \right) \quad (3.38)$$

Il primo integrale è esattamente zero: lo si può calcolare analiticamente oppure basta accorgersi del fatto che la funzione $f(t) = t e^{-\frac{t^2}{2\sigma^2}}$ è dispari siccome $f(-t) = -f(t)$ e perciò un qualsiasi integrale tra due termini uguali e di segno opposto (anche se improprio) darà 0. Il secondo integrale è quello di normalizzazione incontrato nell' [Equazione 3.9](#), ed ha il valore $\sigma\sqrt{2\pi}$, che si semplifica con il denominatore e resta il risultato atteso:

$$\bar{x} = \frac{1}{\cancel{\sigma\sqrt{2\pi}}} \cdot \mu \cdot \cancel{\sigma\sqrt{2\pi}} = \mu$$

Un' altra grandezza da calcolare è la deviazione standard σ_x , che è data da:

$$\sigma_x^2 = \int_{-\infty}^{+\infty} (x - \bar{x})^2 G(x|\mu, \sigma) dx \quad (3.39)$$

Questo si calcola *facilmente* (dimostrazione in [Sezione 8.2](#)), basta sostituire \bar{x} con μ , poniamo $x - \mu = t$ e $t/\sigma = z$, integriamo per parti e otteniamo:

$$\sigma_x^2 = \sigma^2$$

3.9.1 Giustificazione della media come migliore stima (Extra)

Se $f(x)$ fosse nota, allora potremmo calcolare la media \bar{x} e la deviazione standard σ ottenute dopo infinite misure e conosceremmo anche il valore vero μ . Sfortunatamente non conosciamo la distribuzione limite. In pratica abbiamo un numero finito di valori misurati x_1, x_2, \dots, x_n .

Il nostro problema è arrivare alla migliore stima di μ e di σ , basandoci sugli n valori misurati. Se le misure seguissero una distribuzione normale $G(x|\mu, \sigma)$ e se conoscessimo i parametri μ e σ , potremmo calcolare la probabilità di ottenere una lettura vicino a x_1 in un intervallo molto piccolo dx_1 :

$$P(x_1 < x < x_1 + dx_1) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x_1 - \mu)^2}{2\sigma^2}} dx_1 \quad (3.40)$$

Non siamo però interessati alla dimensione dell'intervallo dx_1 , così abbreviamo l'equazione precedente a:

$$P(x_1) \simeq \frac{1}{\sigma} \cdot e^{-\frac{(x_1 - \mu)^2}{2\sigma^2}}$$

Ci riferiamo all'Equazione 3.9.1 come alla probabilità di ottenere x_1 , sebbene sia la probabilità di ottenere un valore in un intervallo molto piccolo di x_1 , la probabilità di ottenere la seconda lettura x_2 è

$$P(x_2) \simeq \frac{1}{\sigma} \cdot e^{-\frac{(x_2 - \mu)^2}{2\sigma^2}}$$

La probabilità di osservare l'intero insieme di n letture è proprio il prodotto di queste distinte probabilità:

$$P_{\mu, \sigma}(x_1 \dots x_n) = \prod_{i=1}^n P(x_i) \quad (3.41)$$

Dati gli n valori misurati x_1, x_2, \dots, x_n , le migliori stime per μ e σ sono quei valori uguali per i quali gli $x_1 \dots x_n$ assumono la massima probabilità. Cioè, le migliori stime per μ e σ sono quei valori per cui $P_{\mu, \sigma}(x_1 \dots x_n)$ è massima, dato che:

$$P_{\mu, \sigma}(x_1 \dots x_n) \simeq \frac{1}{\sigma^n} \cdot e^{-\frac{\sum_i (x_i - \mu)^2}{2\sigma^2}} \quad (3.42)$$

Ovviamente la 3.42 è massima se la somma nell'esponente è minima. Così la migliore stima per μ è quel valore di μ per cui $\sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^2}$ è minima. Per calcolare questo, facciamo la derivata rispetto ad μ ed imponiamola uguale a 0 ottenendo:

$$\mu \approx \frac{\sum_{i=1}^n x_i}{n} \quad (3.43)$$

Trovare la migliore stima per σ , la larghezza della distribuzione limite è più difficile. Si deve differenziare la 3.42, rispetto a σ e porre la derivata uguale a 0. Questo fornisce la migliore stima per σ che è:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (3.44)$$

Poichè il valore vero di X ovvero μ è incognito dobbiamo sostituire \bar{x} e così arriviamo a stimare che:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (3.45)$$

3.10 Distribuzione del χ^2

3.10.1 Gamma di Eulero

Prima di presentare la distribuzione del χ^2 , introduciamo la funzione gamma di eulero, uno strumento matematico utile: partiamo dalla definizione tramite l'integrale

$$\Gamma(x) = \int_0^{+\infty} t^{(x-1)} \cdot e^{-t} dt = -t^{(x-1)} \cdot e^{-t} \Big|_0^{+\infty} + \int_0^{+\infty} (x-1) \cdot t^{(x-2)} \cdot e^{-t} dt \quad (3.46)$$

Il che equivale a dire che

$$\Gamma(x) = (x-1) \cdot \Gamma(x-1)$$

Vale inoltre che, se sostituisco $x = n$ e integro in modo iterativo per parti ottengo:

$$\Gamma(n) = (n-1)! \quad (3.47)$$

Infatti, sia a livello analitico che logico:

$$\Gamma(1) = \int_0^{+\infty} e^{-t} dt = 1$$

π Teorema 9: Gamma di Eulero

L'integrale Gamma di Eulero è un integrale particolare che vale:

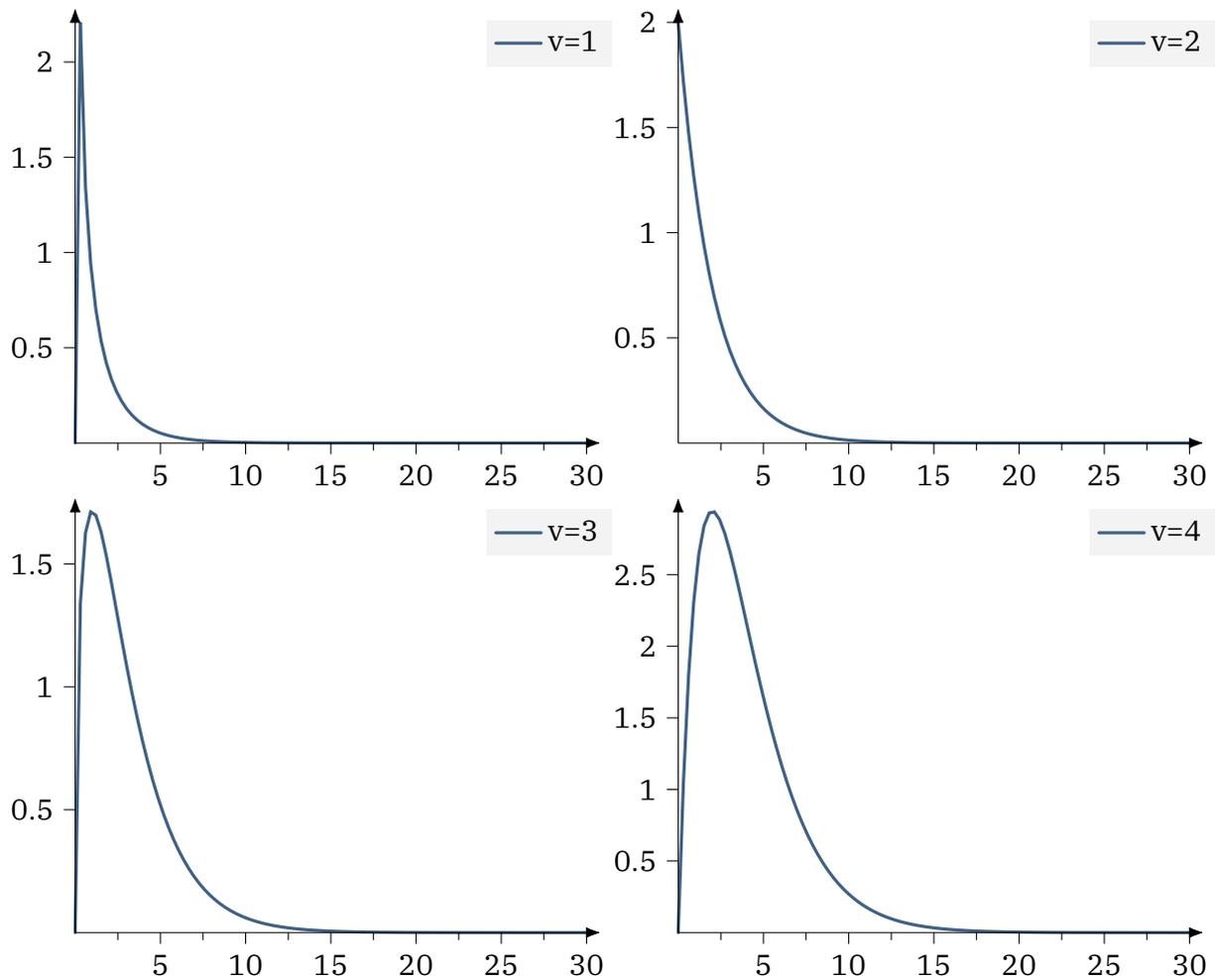
$$\Gamma(x) = \int_0^{+\infty} t^{(x-1)} \cdot e^{-t} dt = (x-1)! \quad (3.48)$$

Adesso possiamo definire la distribuzione del χ^2 :

Definizione 21

$$\chi^2(x|v) = \left[2^{\frac{v}{2}} \Gamma\left(\frac{v}{2}\right) \right]^{-1} \cdot x^{\frac{v}{2}-1} \cdot e^{-\frac{x}{2}} \quad (3.49)$$

Dove v è il numero di gradi di libertà del sistema:



In base a v gradi di libertà abbiamo che:

$$\begin{cases} \text{moda} = 0 \text{ per } v \leq 2 \\ \text{moda} = v - 2 \text{ per } v > 2 \end{cases}$$

Inoltre possiamo definire il valore atteso e la varianza di questa particolare distribuzione come segue (la dimostrazione di $\mathbb{E}[x]$ è [Sezione 8.4](#)).

$$\begin{cases} \mathbb{E}[x] = v \\ \text{Var}[x] = 2v \end{cases}$$

Supponiamo adesso di avere n variabili z_i . Una delle tante proprietà di questa distribuzione è che la somma dei quadrati di variabili normali si comportano come una χ^2 di parametro $v = n$, ovvero

$$X = \sum_{i=1}^n z_i^2 \sim \chi^2(X|v = n)$$

La distribuzione del χ^2 gode inoltre della proprietà riproduttiva:

Definizione 22

Una Distribuzione di qualsiasi tipo gode della **proprietà riproduttiva** se, date due variabili $x_1, x_2 \sim D$ che seguono una distribuzione qualsiasi D , allora $x = x_1 + x_2 \sim D$.

Dimostrazione 6: Riproduttività della distribuzione del χ^2

Costruisco la variabile Y a partire da m variabili: $Y = \sum_{j=1}^m z_j^2$ con $Y \sim \chi^2$ con n gradi di libertà allora:

$$W = X + Y = \sum_{i=1}^n z_i^2 + \sum_{j=1}^m z_j^2 = \sum_{k=1}^{n+m} z_k^2$$

Dunque $W \sim \chi^2$ con $v = m + n$ gradi di libertà.

Supponiamo adesso di avere una variabile gaussiana $X \sim N(\mu, \sigma)$ ed n misure effettuate x_1, x_2, \dots, x_n . Prendiamo la varianza campionaria:

$$s_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}) \cdot \frac{\sigma^2}{\sigma^2} = \frac{\sigma^2}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^2 = \frac{\sigma^2}{n} \sum_{i=1}^n z_i^2$$

Allora otteniamo che:

$$s_n^2 \sim \frac{\sigma^2}{n} \cdot \chi^2(v = n - 1) \quad (3.50)$$

Questo significa che la varianza si distribuisce come una variabile χ^2 con $n - 1$ gradi di libertà, ora dimostriamo che effettivamente sono $n - 1$.

✔ Dimostrazione 7

Partiamo dalla varianza campionaria e la manipoliamo un po', scrivendola però in funzione di μ e non di i :

$$\begin{aligned} \sum_{i=1}^n (x_i - \mu)^2 &= \sum_{i=1}^n [(x_i - \bar{x}) + (\bar{x} - \mu)]^2 = && \text{true} \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 + \underbrace{\sum_{i=1}^n (\bar{x} - \mu)}_{=n \cdot (\bar{x} - \mu)} + 2 \underbrace{\sum_{i=1}^n (x_i - \bar{x})(\bar{x} - \mu)}_{=0} \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2 \end{aligned}$$

Sul perchè dei risultati delle due sommatorie a destra, il conto è abbastanza facile: la prima molto semplicemente è una costante, che viene sommata n volte cioè gli indici della sommatoria, mentre per la seconda essa fa 0 siccome, ricordandosi che la media è $\bar{x} = 1/n \sum_i x_i \implies \sum_i x_i = n\bar{x}$, si ha che:

$$\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - n\bar{x} = n\bar{x} - n\bar{x} = 0$$

✓ Dimostrazione 7 continued

Adesso dividiamo per σ^2 :

$$\begin{aligned} \frac{1}{\sigma^2} \cdot \left[\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2 \right] &= \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^2 + \left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \right)^2 \\ &= \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^2 + \left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \right)^2 \end{aligned}$$

A questo punto, siccome abbiamo moltiplicato a destra dell'uguale, lo abbiamo fatto in realtà anche a sinistra dell'uguale, termine a cui non siamo più tornati. Quel termine iniziale non era infatti la varianza campionaria in sé, ma ora l'abbiamo trovata, e quindi la portiamo dall'altra parte dell'uguale:

$$\begin{aligned} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2 &= \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^2 + \left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \right)^2 \\ &\Downarrow \\ \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^2 &= \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2 - \left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \right)^2 \end{aligned}$$

Ma a questo appunto abbiamo che la prima sommatoria è una chi quadro con n gradi di libertà (è la somma di n variabili gaussiane, proprietà discussa prima), mentre quella a destra ne ha 1, e per la proprietà riproduttiva della distribuzione:

$$\chi^2(v = n) - \chi^2(v = 1) = \chi^2(v = n - 1)$$

Il valore atteso della varianza campionaria allora è:

$$\begin{aligned} \mathbb{E}[s_n^2] &= \frac{\sigma^2}{n} \mathbb{E}[\chi^2(n - 1)] = \frac{\sigma^2}{n} v = \frac{n - 1}{n} \sigma^2 = \frac{n - 1}{n} \cdot \sigma^2 \\ \mathbb{E}[s_{n-1}^2] &= \frac{n - 1}{n - 1} \cdot \sigma^2 = \sigma^2 = \sigma^2 \end{aligned}$$

La varianza invece è:

$$\begin{aligned} \text{Var}[s_n^2] &= \text{Var} \left[\frac{\sigma^2}{n} \chi^2(n - 1) \right] = \frac{\sigma^4}{n^2} \text{Var} [\chi^2(n - 1)] = \frac{\sigma^4}{n^2} 2v = 2 \frac{\sigma^4}{n^2} (n - 1) \\ \text{Var}[s_{n-1}^2] &= \frac{2\sigma^4}{(n - 1)^2} (n - 1) = \frac{2\sigma^4}{n^2} = \frac{2\sigma^4}{n - 1} \end{aligned}$$

La cosa interessante che si cela dietro a questo calcolo è che la variabile casuale della varianza campionaria s_{n-1}^2 ha come valore atteso proprio σ^2 , quindi spesso si preferisce usare questa. Infine, per $n \rightarrow \infty$ si ha che anche $\mathbb{E}[s_n^2] = \sigma^2$ e $\text{Var}[s_n^2] = \text{Var}[s_{n-1}^2] = 0$.

3.11 Distribuzione Esponenziale

Abbiamo visto finora variabili discrete e continue (definite in un intervallo finito). Adesso ci soffermiamo nel caso di variabili continue definite in un intervallo infinito. Quando accade ciò la funzione di densità di probabilità deve avere un andamento decrescente per far sì che il suo integrale sia uguale ad 1. Un caso interessante è quello di una grandezza definita tra 0 e 1 i cui valori decrescono esponenzialmente:

$$f(x) \propto e^{-ax}$$

Il fattore di proporzionalità si ricava dalla condizione di normalizzazione dunque:

$$f(x) = ae^{-ax}$$

Per le applicazioni che faremo di questa distribuzione introduciamo una variabile T con il significato di tempo e un parametro omogeneo a questa $\tau = 1/a$ dunque la nostra funzione diventa:

$$f(t|\varepsilon(\tau)) = \frac{1}{\tau} \cdot e^{-t/\tau}$$

Se ($0 \leq t \leq \infty$) e $\tau \geq 0$, invece:

$$f(t|\varepsilon(\tau)) = 1 - e^{-t/\tau}$$

Dove $0 \leq t \leq \infty$.

Definizione 23

La distribuzione esponenziale, usata per esempio nei tempi di attesa, è:

$$f(t|\varepsilon(\tau)) = \frac{1}{\tau} \cdot e^{-t/\tau} \quad (3.51)$$

Calcoliamoci adesso il valore atteso e la deviazione standard di T :

$$\mathbb{E}[t] = \int_0^{\infty} \frac{t}{\tau} \cdot e^{-t/\tau} dt = -te^{-t/\tau} \Big|_0^{+\infty} + \int_0^{\infty} e^{-t/\tau} dt = \tau \quad (3.52)$$

Prima di trovare il valore di sigma dobbiamo necessariamente trovare il valore di $\mathbb{E}[\tau^2]$ e otteniamo:

$$\begin{aligned} \mathbb{E}[T^2] &= \int_0^{\infty} \frac{t^2}{\tau} \cdot e^{-t/\tau} dt = -t^2 e^{-t/\tau} \Big|_0^{+\infty} + 2 \int_0^{\infty} e^{-t/\tau} dt = 2\tau^2 \\ \sigma_t &= \sqrt{\mathbb{E}[t^2] - \mathbb{E}[t]^2} = \sqrt{2\tau^2 - \tau^2} = \tau \end{aligned} \quad (3.53)$$

3.12 Distribuzione del tempo di attesa del primo successo

Partiamo dal calcolare la probabilità che non si verifichi nessun evento durante un certo tempo finito t . Questo è un evento Poissoniano chiaramente. Chiamiamo allora:

- t il tempo totale, e lo dividiamo in n intervalli;
- T il nome degli intervalli (perchè abbiamo molta fantasia), quindi la loro larghezza sarà $\Delta T = t/n$;
- p la probabilità di conteggio in ogni T , ovvero $p = r \cdot \Delta T$, in cui r è la rate al secondo dei conteggi, e moltiplicandola per l'intervallo di tempo desiderato ottengo la frequenza in quel Δ , ovvero λ .

A questo punto, ci calcoliamo la probabilità $P(T > t)$, ovvero di dover aspettare un tempo maggiore di t per avere un conteggio. Chiaramente la probabilità di non osservare un conteggio nel singolo T è $(1 - p)$, che per la legge della moltiplicazione (vedi [Sottosezione 1.6.1](#)) va appunto moltiplicato per le altre probabilità, che sono uguali e di eventi indipendenti tra loro, sarà quindi:

$$\begin{aligned} P(T > t) &= (1 - p)^n \\ &= (1 - r \Delta T)^n \\ &= \left(1 - r \frac{t}{n}\right)^n \end{aligned}$$

E quando $n \rightarrow \infty$ la probabilità diventa:

$$P(T > t) = e^{-rt}$$

Da cui segue la probabilità cumulativa, ovvero allargando il problema, quanto vale $P(0 < T < t)$ che sarebbe la cumulativa fino a t , ma siccome quella totale è 1 e ci siamo già calcolati la probabilità da t in poi il problema si semplifica:

$$F(t) = 1 - P(T > t) = 1 - e^{-rt}$$

Se adesso volessimo ottenere la funzione di densità di probabilità ci basterà derivare la $F(t)$ per ottenere:

$$f(t) = \frac{dF(t)}{dt} = r e^{-rt}$$

Definizione 24

La **Distribuzione del tempo di attesa del primo successo** segue dalla probabilità di un evento Poissoniano.

$$f(t) = r e^{-rt} \quad (3.54)$$

Chiamando $r = \frac{1}{\tau}$, dove τ è la *costante di tempo*, diventa anche

$$f(t) = \frac{1}{\tau} e^{-t/\tau} \quad (3.55)$$

I suoi valore atteso e deviazione derivano da quelli calcolati per le distribuzioni esponenziali:

$$\mathbb{E}[t] = \tau; \quad (3.56)$$

$$\text{Var}[t] = \tau^2; \quad (3.57)$$

$$\sigma_t = \tau \quad (3.58)$$

3.12.1 Intensità di più processi di Poisson

Apriamo soltanto una piccola parentesi sull'intensità dei processi di Poisson, infatti se abbiamo più processi indipendenti tra di loro, ciascuno con intensità r_i e che producono eventi tra loro indistinguibili, possono essere considerati come un unico processo di poisson:

$$r_{tot} = \sum_{i=1}^n r_i \quad (3.59)$$

3.13 Distribuzione di Erlang

Adesso consideriamo un processo di Poisson con una certa intensità r e interessiamoci alla variabile casuale T che indica il tempo di attesa per cui si verificherà il k -esimo successo. Il caso precedente $k = 1$ è ben descritto dall'esponenziale. Ma adesso consideriamo invece l'evento per cui il k -esimo successo si verifica dopo un tempo $T > t$, ovvero tale per cui si sono già verificati $k - 1$ successi. Dunque poichè il numero di successi nel tempo $T = t$ è dato da una distribuzione di Poisson di parametro $\lambda = rt$ abbiamo:

$$P(E) = f(0|P_{rt}) + f(1|P_{rt}) + \dots + f(k-1|P_{rt}) = F(k-1|P_{rt})$$

Dunque la probabilità che il k -esimo evento si verifichi al tempo $T = t$ è uguale a:

$$P_k(T \leq t) = 1 - P(E) = 1 - F(k-1|P_{rt}) = 1 - \sum_{x=0}^{k-1} \frac{e^{-rt} \cdot (rt)^x}{x!}$$

Si può verificare che quanto abbiamo appena ottenuto è soluzione del seguente integrale:

$$\int_0^t \frac{x^{k-1}}{(k-1)!} \cdot r^k \cdot e^{-rx} dx$$

Da questa ricaviamo facilmente la distribuzione di probabilità che stavamo cercando che prende il nome di distribuzione di Erlang:

$$F(t|Erlang(k, r)) = \int_0^t \frac{t'^{k-1}}{(k-1)!} \cdot r^k \cdot e^{-rt'} dt'$$

Definizione 25

La **Distribuzione di Erlang** serve a trovare la il tempo di attesa da un istante fissato al k -esimo successo, la sua espressione analitica è

$$\mathcal{E}(x_k|r) = \frac{x^{k-1}}{(k-1)!} \cdot r^k \cdot e^{-rx} \quad (3.60)$$

Il valore atteso e la varianza di questa distribuzione valgono:

$$\mathbb{E}[T] = \frac{k}{r} = k\tau \quad (3.61)$$

$$\text{Var}[T] = \frac{k}{r^2} \longrightarrow \sigma[T] = \frac{\sqrt{k}}{r} = \sqrt{k}\tau \quad (3.62)$$

3.13.1 Vita media di Decadimento

Supponiamo adesso di avere un nucleo radioattivo per il quale la probabilità di decadimento è indipendente dal tempo. Come detto in precedenza sappiamo che l'istante di decadimento di un nucleo da un certo istante è descritto da un esponenziale con parametro r . Interessante è calcolare il tempo di dimezzamento della particella (cioè il tempo tale che ci sia il 50% di probabilità che la particella sia già decaduta), che corrisponde alla mediana della distribuzione:

$$P(T \leq t_{1/2}) = 1 - e^{-t/\tau} = 0.5$$

Che banalmente possiamo riscrivere come:

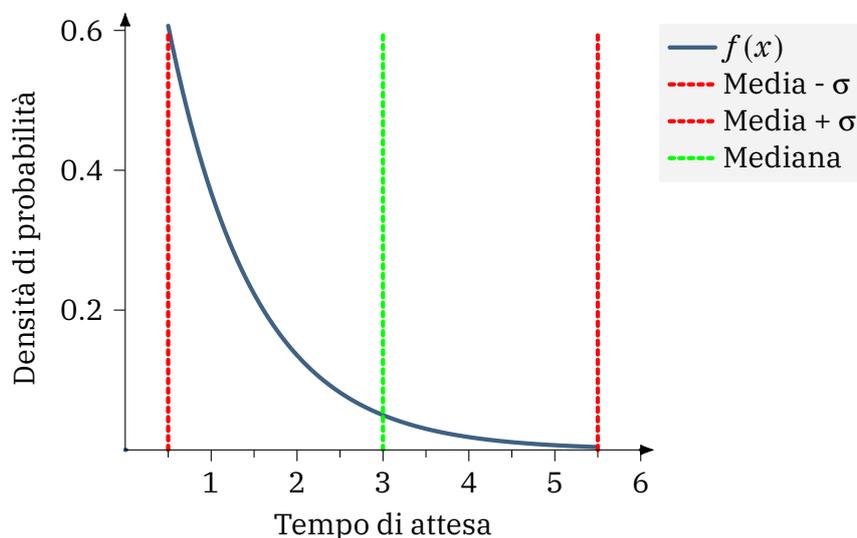
$$t_{1/2} = \tau \cdot \ln 2 \quad (3.63)$$

Per capire meglio questi calcoli immaginiamo di avere N_0 nuclei all'istante $T = 0$ e calcoliamo la previsione di quanti nuclei sono rimasti dopo un tempo t . Questo segue una distribuzione binomiale dove $n = N_0$ e $p = e^{-t/\tau}$ dunque ricaviamo che:

$$\mathbb{E}[N] = N_0 e^{-t/\tau}$$

$$\sigma[N] = \sqrt{N} \sqrt{e^{-t/\tau}(1 - e^{-t/\tau})}$$

Qui di seguito è riportato il grafico del tempo di decadimento di un nucleo radioattivo:



3.14 Distribuzione t-student

La distribuzione t di Student è una distribuzione di probabilità utilizzata principalmente nelle inferenze statistiche quando le dimensioni del campione sono limitate e/o la varianza della popolazione è sconosciuta. La distribuzione t è simile alla distribuzione normale, ma tiene conto dell'incertezza aggiuntiva associata alle stime dei parametri della popolazione basate su un campione più piccolo.

Definizione 26

La **Distribuzione del t-Student** serve nella stima della media di una popolazione che segue la distribuzione normale.

$$f(x|v) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{\pi v} \cdot \Gamma(\frac{v}{2})} \left(1 + \frac{x^2}{v}\right)^{-\frac{v+1}{2}} \quad (3.64)$$

Il valore atteso e la varianza di questa distribuzione sono rispettivamente (v = gradi di libertà):

$$\mathbb{E}[X] = 0 \quad (v > 1) \quad (3.65)$$

$$\text{Var}[X] = \frac{v}{v-2} \quad (3.66)$$

La distribuzione t viene spesso utilizzata per calcolare gli intervalli di confidenza per la media di una popolazione, *test t* per confrontare le medie di due gruppi e per la regressione lineare quando si considerano i coefficienti dei modelli. Possiamo usare la variabile standardizzata, ovvero lo scarto della media di N misure per definire gli intervalli di confidenza per la media di una popolazione sulla base della media \bar{x} e della varianza $\text{Var}[x]$:

$$T = \frac{\bar{x} - \mu}{\sqrt{\frac{S_n^2}{n}}} \quad (3.67)$$

Dunque possiamo riscrivere T come compresa tra due parametri a e b e ottenere:

$$P(a \leq T \leq b) = P\left(\bar{x} - b\sqrt{\frac{S_n^2}{n}} \leq \mu \leq \bar{x} - a\sqrt{\frac{S_n^2}{n}}\right)$$

Scegliamo dei quantili $q_a \leq q_b$

$$B - a = P(q_a \leq T \leq q_b) = P\left(\bar{x} - q_b\sqrt{\frac{S_n^2}{n}} \leq \mu \leq \bar{x} - q_a\sqrt{\frac{S_n^2}{n}}\right)$$

Se adesso ci mettiamo nel caso di intervalli simmetrici allora possiamo utilizzare il parametro z_a e riscrivere il tutto:

$$a = P(|T| \leq z_a) = P(-z_a \leq T \leq z_a) = 2F(z_a) - 1 \rightarrow Z_a = q_{1-\frac{a}{2}}$$

E infine otteniamo che:

$$\left[\bar{x} - z_a \sqrt{\frac{S_n^2}{n}}, \bar{x} + z_a \sqrt{\frac{S_n^2}{n}} \right] \quad (3.68)$$

4

Inferenza



Sommario

- 4.1 Inferenza Probabilistica, 85
- 4.2 Verosimiglianza normale con σ nota, 87
- 4.3 Metodo della massima verosimiglianza, 88
- 4.4 Inferenza simultanea su μ e su σ , 91
- 4.5 Distribuzione Predittiva, 93
- 4.6 Inferenza Poissoniana, 94
- 4.7 Combinazione di Risultati, 97
- 4.8 Inferenza sulla distribuzione binomiale, 98
- 4.9 Inferenza su un esperimento, 100
- 4.10 Inferenza sui parametri di una legge (Fit lineari), 105

L'inferenza in statistica è in poche parole il processo di apprendimento dall'esperienza. Se in genere valutiamo la probabilità a partire dalle cause, il teorema di Bayes ci insegna che possiamo aggiornare le probabilità alla luce di nuove informazioni.

Possiamo quindi modificare un modello, una probabilità, o addirittura dedurre tutto ciò a partire dall'osservazione di dati sperimentali, "invertendo" il processo intuitivo di *causa* \rightarrow *effetto*.

4.1 Inferenza Probabilistica

Partendo proprio dal Teorema di Bayes, che è la base dell'Inferenza statistica, possiamo rileggere la probabilità condizionata $P(E|H)$ come il condizionante H causa dell'evento (effetto) E e il condizionante può provocare vari effetti condizionati, così come un effetto può nascere da varie condizioni (basti pensare alla formula della probabilità totale ovvero $P(E) = \sum_i P(E|H_i) \cdot P(H_i)$).

$$P(E|H) \iff P(\text{Effetto} | \text{Causa})$$

Questo può risultare fuorviante, nel senso che in realtà quando scriviamo $P(E|H)$ ci riferiamo a uno **stato di informazione**, ovvero la probabilità che accada E con l'ipotesi che H sia vera, non dopo che abbiamo verificato che è vera (ecco la sottigliezza alla base dell'inferenza, scambiare "*causa-effetto*" non significa scambiare la temporalità e la causalità di due eventi, ma modificare la nostra teoria, la nostra probabilità... alla luce di nuovi stati di informazione).

4.1.1 Inferenza Bayesiana

Il teorema di Bayes, ovvero [Equazione 1.18](#) che riscriviamo per comodità, dice che:

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)} = \frac{P(E|H) \cdot P(H)}{P(E|H) \cdot P(H) + P(E|\bar{H}) \cdot P(\bar{H})}$$

In questa formula chiamiamo le varie probabilità con dei nomi che ci aiutano a far capire il loro ruolo inferenziale:

- $P(H)$ o meglio $P(H_0)$ è la probabilità a priori di H_0 , detta anche **Prior**;
- $P(E|H_0)$ è la **Funzione di Verosimiglianza**;
- $P(E)$ è la **Probabilità marginale** ovvero quella di osservare E senza informazioni pregresse, ha il ruolo di costante di normalizzazione;
- $P(H_0|E)$ è la **Probabilità a posteriori** di H_0 dato E .

Ricordandoci quindi che $P(E)$ è una costante di normalizzazione, il fulcro dell'inferenza Bayesiana sta nella proporzionalità delle due probabilità a priori e posteriori:

$$P(H|E) \propto P(E|H) \cdot P(H)$$

Ovvero

$$\text{Probabilità a posteriori} \propto \text{verosimiglianza} \times \text{probabilità iniziale} \quad (4.1)$$

Nel caso di passaggio a distribuzioni di probabilità di variabili discrete o continue, ovviamente la p.d.f. fa il ruolo di probabilità di ottenere un certo valore x che equivarrebbe ai *dati*, e utilizzando la formulazione sopra otteniamo

$$f(\mu|\text{dati}, I) \propto f(\text{dati}|\mu, I) \cdot f_{\circ}(\mu|I) \quad (4.2)$$

In cui I è la globalità del nostro stato di informazione della misura, e f_{\circ} è ovviamente la prior.

4.2 Verosimiglianza normale con σ nota

Ci interessiamo all'Inferenza su misure con deviazione nota con l'utilizzo della Gaussiana sia per il teorema del limite centrale sia perchè essa descrive bene gli errori di misura.

Considerando quindi tanti errori e_i di valore atteso nullo (esisterà un valore vero μ della misura, quindi gli errori sono fluttuazioni) e deviazione σ_i . L'errore totale è $e = \sum_{i=1}^n e_i$ con $\sigma_e^2 = \sum_{i=1}^n \sigma_i^2$. Infine, a volte l'inferenza non è fatta a partire da una singola osservazione, ma dalla media di molte osservazioni, come se si trattasse di una singola osservazione equivalente. Anche in questo caso, interviene il teorema del limite centrale che ci fa credere che questa 'osservazione equivalente' possa essere descritta da una gaussiana. Sulla base di tutte queste ipotesi la singola osservazione X sarà descritta da

$$X \sim N(\mu, \sigma_e)$$

Per trovare ora la funzione di densità di probabilità che abbiamo assunto essere gaussiana, utilizziamo il teorema di Bayes e troviamo che

$$f(\mu|X) = \frac{\frac{1}{\sigma_e \sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma_e^2}} \cdot f_{\circ}(\mu)}{\int_{-\infty}^{+\infty} \frac{1}{\sigma_e \sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma_e^2}} \cdot f_{\circ}(\mu) d\mu}$$

Resterebbe il dubbio su cosa usare come $f_{\circ}(\mu)$, ma va ricordato che l'inferenza non è un esercizio matematico bensì l'aggiornamento delle nostre credenze in luce a nuovi fatti: in questa lettura la prior è molto più vaga di $f(x|\mu)$ perchè la prima indica la prior del valore vero mentre l'altra è l'informazione che ci può dare lo strumento (è come se volessimo misurare la temperatura di una stanza per migliorare la nostra conoscenza a riguardo: tutti sono più o meno in grado di percepire una certa temperatura e stimarla, e la nostra sensazione fisiologica è più o meno costante a riguardo). Possiamo quindi approssimare $f_{\circ}(\mu) \approx k \in \mathbb{R}$ e semplificarlo sia sotto che sopra, otteniamo

$$f(\mu|X) = \frac{\frac{1}{\sigma_e \sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma_e^2}}}{\underbrace{\int_{-\infty}^{+\infty} \frac{1}{\sigma_e \sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma_e^2}} d\mu}_{=1}} = \frac{1}{\sigma_e \sqrt{2\pi}} \cdot e^{-\frac{(\mu-x)^2}{2\sigma_e^2}}$$

Da cui segue automaticamente che $\mathbb{E}[\mu] = x$ e $\sigma_{\mu} = \sigma_e$.

4.3 Metodo della massima verosimiglianza

Un altro metodo per inferire μ è quello della *massima verosimiglianza* che evita di calcolarci l'integrale e cerca il μ che massimizza la verosimiglianza con i dati sperimentali.

Abbiamo visto prima come con un prior vaga (o con molti dati) allora $f(\mu|x) \propto f(x|\mu)$ ed è sia matematicamente chiaro che logicamente ovvio che la dipendenza della probabilità finale $f(\mu)$ dalla probabilità a priori (quindi dalla prior) diminuisce sempre di più all'aumentare dei dati sperimentali (si avvalora $f(x|\mu)$).

Trascurando quindi l'integrale di normalizzazione, che vogliamo proprio evitare di usare, $f(\mu|x)$ non è più una p.d.f. ma una funzione di μ con x fissato, che indichiamo con $\mathcal{L}(\mu; x)$. Ora possiamo dare una definizione del principio:

Definizione 27

Il metodo di massima verosimiglianza afferma che la migliore stima per μ è quel $\hat{\mu}$ che massimizza $\mathcal{L}(\mu; x)$.

Senza preoccuparci troppo della dimostrazione (che per i più curiosi è [Sezione 8.6](#)) andiamo a definire il punto di arrivo, ovvero il logaritmo della funzione $\mathcal{L}(\mu; x)$, siccome per massimizzare la funzione di massima verosimiglianza basta massimizzarne il logaritmo che è molto più facile:

$$\log(\mathcal{L}) = - \sum_{i=1}^n \left(\log \sqrt{2\pi} + \log(\sigma) + \frac{(x_i - \mu)^2}{2\sigma^2} \right)$$

Da cui si ricava la miglior stima di μ e della deviazione facendo la derivata e ponendola uguale a 0:

$$\begin{aligned} \frac{\partial \log(\mathcal{L})}{\partial \mu} = 0 &\implies \hat{\mu} \sim \mathbb{E}[\mu] \\ \frac{\partial^2 \log(\mathcal{L})}{\partial \mu^2} \Big|_{\hat{\mu}} &\sim -\frac{1}{\sigma^2[\hat{\mu}]} \end{aligned}$$

Queste relazioni si possono dimostrare ma non le faremo. Riprendendo il logaritmo quindi,

essendo presenti tante costanti la derivata e l'uguaglianza a 0 è molto semplice:

$$0 = \frac{\partial \log(\mathcal{L})}{\partial \mu} = 0 + \frac{1}{\sigma^2} \cdot \sum_{i=1}^n (x_i - \mu) = \frac{1}{\sigma^2} \cdot \left(\sum_{i=1}^n x_i - n\mu \right)$$

$$0 = \frac{1}{\sigma^2} \cdot \left(\sum_{i=1}^n x_i - n\mu \right)$$

$$\hat{\mu} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

Allo stesso modo ci ricaviamo la deviazione sul miglior valore $\hat{\mu}$:

$$\left. \frac{\partial^2 \log(\mathcal{L})}{\partial \mu^2} \right|_{\hat{\mu}} = -\frac{1}{\sigma^2} \cdot \sum_{i=1}^n 1 = -\frac{n}{\sigma^2}$$

E siccome questa si comporta come $-1/\sigma^2[\hat{\mu}]$, ci siamo trovati i valori desiderati:

$$\hat{\mu} = \frac{1}{n} \cdot \sum_{i=1}^n x_i \quad (4.3)$$

$$\hat{\sigma}_{\hat{\mu}}^2 = \frac{\sigma^2}{n} \quad (4.4)$$

Inoltre, derivando rispetto a σ il logaritmo e ponendolo uguale a 0, troviamo invece il miglior valore della deviazione, usato poi in seguito per ricavarsi il miglior valore della deviazione su μ (questo valore è sconosciuto, sennò avremmo usato l'altro metodo per trovarci la funzione):

$$0 = \frac{\partial \log(\mathcal{L})}{\partial \sigma} = -\sum_{i=1}^n \left(\frac{1}{\sigma} - \frac{(x_i - \mu)^2 \cdot 2 \cdot \sigma}{2 \cdot \sigma^4} \right) = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \cdot \sum_{i=1}^n (x_i - \mu)^2$$

$$0 = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \cdot \sum_{i=1}^n (x_i - \mu)^2$$

$$\frac{n}{\sigma} = \frac{1}{\sigma^3} \cdot \sum_{i=1}^n (x_i - \mu)^2$$

$$\frac{\sigma^2}{n} = \sum_{i=1}^n (x_i - \mu)^2$$

$$\hat{\sigma}^2 = \frac{1}{n} \cdot \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}$$

Quindi ci siamo trovati anche il valore della deviazione sconosciuta:

$$\hat{\sigma}^2 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \mu)^2 \quad (4.5)$$

In questo caso in metodo della massima verosimiglianza prende il nome di *metodo dei*

minimi quadrati perchè invece di cercare il massimo di \mathcal{L} abbiamo cercato il minimo di $\sum_i (x_i - \mu)^2$.

È importante notare come queste sono tra le formule più importanti in assoluto, e infatti sono quelle che più spesso si usano in laboratorio. Per chi non ha un occhio molto attento, il valore vero migliore che ci siamo ricavati è la media, la varianza sui valori misurati è la varianza campionaria e la deviazione standard è la super utilizzata σ/\sqrt{n} :

$$\hat{\mu} = \frac{1}{n} \cdot \sum_{i=1}^n x_i = \bar{x} \quad (4.6)$$

$$\hat{\sigma}^2 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2 = s_n^2 \quad (4.7)$$

$$\sigma_{\hat{\mu}} = \frac{\hat{\sigma}}{\sqrt{n}} = \sigma_{\mu} \quad (4.8)$$

4.4 Inferenza simultanea su μ e su σ

Passiamo adesso al caso generale, immaginiamo di dover inferire non su un singolo valore ignoto ma bensì su due. Per inferire su μ e σ da un insieme n di osservazioni ci basterà applicare il teorema di Bayes e di applicare la distribuzione congiunta sulla variabile che non ci interessa:

$$f(\mu, \sigma | x, I) \propto f(x | \mu, \sigma, I) \cdot f(\mu, \sigma | I) \quad (4.9)$$

Come per μ il modello relativamente più semplice è una distribuzione uniforme per valori positivi di σ . Assumendo implicito il condizionante I otteniamo:

$$\begin{aligned} f(\mu, \sigma | x) &\propto f(x | \mu, \sigma) \\ &\propto \prod_i^n \frac{1}{\sigma} \cdot \exp \left[-\frac{(x_i - \mu)^2}{2\sigma^2} \right] \\ &\propto \frac{1}{\sigma^n} \cdot \exp \left[-\frac{1}{2\sigma^2} \cdot \sum_i^n (x_i - \mu)^2 \right] \end{aligned}$$

Dunque adesso possiamo utilizzare l'uguaglianza:

$$\begin{aligned} \sum_i^n (x_i - \mu)^2 &= \sum_i^n \left[(x_i - \bar{x})^2 + (\bar{x} - \mu)^2 \right] = \\ &= \sum_i^n (x_i - \bar{x})^2 + 2(\bar{x} - \mu) \sum_i^n (x_i - \bar{x}) + \sum_i^n (\bar{x} - \mu)^2 \end{aligned}$$

Sapendo che $\bar{x} = \frac{\sum_i^n x_i}{n} \rightarrow \sum_i^n x_i = n\bar{x}$ troviamo che:

$$\sum_i^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)$$

Ricordandoci che $\sum_i^n (x_i - \bar{x})$ possiamo considerarlo costante e ad esempio lo poniamo uguale a C tornando alla relazione iniziale otteniamo:

$$\propto \frac{1}{\sigma^n} \cdot \exp \left[-\frac{1}{2\sigma^2} \left(n(\bar{x} - \mu)^2 + C^2 \right) \right]$$

Adesso marginalizziamo su σ ed otteniamo:

$$\begin{aligned} f(\mu|x) &= \int_0^{+\infty} f(\mu, \sigma) d\sigma \\ &\propto (n(\bar{x} - \mu)^2 + C^2)^{-\frac{n-1}{2}} \\ &\propto \left(1 + \frac{(\mu - \bar{x})^2}{C^2/n}\right)^{-\frac{n-1}{2}} \end{aligned}$$

Adesso possiamo ricondurci ad una variabile t-student e ci basterà moltiplicare e dividere per $(n-2)$ e otterremo:

$$\propto \left(1 + \frac{(\mu - \bar{x})^2}{(n-2)C^2/n \cdot (n-2)}\right)^{-\frac{(n-2)+1}{2}}$$

Dove possiamo riconoscere:

$$\propto \left(1 + \frac{t^2}{v}\right)^{-\frac{(v+1)}{2}}$$

Da cui possiamo ricavare semplicemente che:

$$\mu = \bar{x} + \frac{C}{\sqrt{n(n-2)}} \cdot t$$

E allora abbiamo che:

$$\begin{aligned} \mathbb{E}[\mu] &= \bar{x} \\ \sigma[\mu] &= \frac{s_n}{\sqrt{n-4}} \end{aligned}$$

In questo modo vediamo come l'incertezza su μ produca un'ulteriore incertezza per σ tanto che per valori di $n < 5$ la varianza alle code della distribuzione è infinita. Dunque adesso proviamo a considerare un caso di prior uniforme in $\log(\sigma)$ per cercare di ridurre le incertezze ottenute dal calcolo precedente. Quindi adesso consideriamo che la nostra prior sia $f_o(\sigma) = \frac{1}{\sigma}$ che diminuirà soltanto di 1 l'effetto della sigma nell'integrando. Dunque abbiamo sempre una variabile t-di student ma con $v - 1$ e nella variabile abbiamo $(\mu - \bar{x})/C^2/(\sqrt{n(n-1)})$. Ricordandoci che:

$$s_{n-1}^2 = \frac{C^2}{n-1} \quad (4.10)$$

Otteniamo allora che:

$$\frac{m - \bar{x}}{s_{n-1}/\sqrt{n}} \quad (4.11)$$

Che segue una distribuzione t di studente con $v-1$ d.o.f.. Adesso possiamo riportare valore atteso e incertezza che sono rispettivamente:

$$\mathbb{E}[\mu] = \bar{x} \quad (4.12)$$

$$\sigma[\mu] = \frac{s_{n-1}}{\sqrt{n}} \sqrt{\frac{n-1}{n-3}} \quad (4.13)$$

E' interessante adesso il caso in cui noi marginalizziamo la funzione congiunta $f(\mu, \sigma)$ rispetto a σ :

$$\begin{aligned} f(\sigma, x) &= \int_{-\infty}^{+\infty} f(\mu, \sigma|x) d\mu \\ &\propto \sigma^{-(n+1)} \cdot \exp\left[\frac{-ns_n^2}{2\sigma^2}\right] \end{aligned}$$

Dove per n molto grandi abbiamo che:

$$\mathbb{E}[\sigma] = s_n \quad (4.14)$$

$$Std[\sigma] = \frac{s_n}{\sqrt{2n}} \quad (4.15)$$

4.5 Distribuzione Predittiva

A questo punto rispondiamo al quesito di voler determinare la probabilità di un valore ignoto. In questo caso $f(x|\mu)$ non basta perchè è un'informazione sulle misure già fatte, è una probabilità di x per ogni ipotesi μ , mentre noi vogliamo a probabilità che tenga conto di tutti i valori possibili di μ , pesati con la loro plausibilità.

$$f(x|I) = \int f(x|\mu, I) \cdot f(\mu|I) d\mu$$

Nel caso in cui la predizione su μ deriva da un valore (o da una serie equivalente) passato x_p con una σ_p , e dato il valore futuro x_f con una σ_f , allora

$$\begin{aligned} f(x_f|x_p) &= \int \frac{1}{\sigma_f \sqrt{2\pi}} e^{-\frac{(x_f-\mu)^2}{2\sigma_f^2}} \cdot \frac{1}{\sigma_p \sqrt{2\pi}} \cdot e^{-\frac{(x_p-\mu)^2}{2\sigma_p^2}} d\mu \\ &= \frac{1}{\sqrt{\sigma_p^2 + \sigma_f^2} \sqrt{2\pi}} e^{-\frac{(x_f-x_p)^2}{2(\sigma_f^2 + \sigma_p^2)}} \end{aligned}$$

In cui valore atteso della misura futura e deviazione valgono

$$\mathbb{E}[X_f] = x_p \quad (4.16)$$

$$\sigma[X_f] = \sqrt{\sigma_p^2 + \sigma_f^2} \quad (4.17)$$

4.6 Inferenza Poissoniana

Passiamo adesso all'inferenza sui conteggi, nei quale vale lo stesso ragionamento (a partire da una serie di conteggi, una rate... faccio inferenza su futuri conteggi o sulla p.d.f.) e più o meno lo stesso approccio.

Se il numero di conteggi è grande, la verosimiglianza è gaussiana, ovvero

$$X \approx N(\lambda, \sqrt{\lambda}) \quad (4.18)$$

$$\lambda \sim N(x, \sigma_x) \quad (4.19)$$

$$\mathbb{E}[\lambda] \approx x \quad (4.20)$$

$$\sigma[\lambda] \approx \sigma[x] \approx \sqrt{\mathbb{E}[x]} \approx \sqrt{x} \quad (4.21)$$

E quindi, per farla breve

$$\lambda \approx N(x, \sqrt{x}) \quad (4.22)$$

Da qui in poi la pratica è relativamente semplice, inferisco le intensità sullo spazio o sul tempo in base al tipo di processo Poissoniano. Se per esempio i conteggi si sono verificati in un T noto avrò che

$$r \sim N\left(\frac{x}{T}; \frac{\sqrt{x}}{T}\right)$$

4.6.1 Caso generale

Nel caso in cui il numero di conteggi non fosse così grande, il ragionamento non cambia, ma il risultato leggermente. Senza usare il teorema del limite centrale inferiamo su $f(\lambda|x, P)$ tramite il teorema di Bayes (tanto o si usa uno o l'altro) e otteniamo:

$$f(\lambda|x, P) = \frac{\frac{\lambda^x \cdot e^{-\lambda}}{x!} \cdot f_{\circ}(\lambda)}{\int_0^{+\infty} \frac{\lambda^x \cdot e^{-\lambda}}{x!} \cdot f_{\circ}(\lambda) d\lambda}$$

Similmente a quanto discusso in [Sezione 4.2](#) approssimiamo la prior vaga a un valore fisso $f_{\circ}(\lambda) = k \in \mathbb{R}$ e la semplifichiamo. Ecco che magicamente i conti si semplificano e otteniamo

$$f(\lambda|x, P) = \frac{\frac{\lambda^x \cdot e^{-\lambda}}{x!}}{\underbrace{\int_0^{+\infty} \frac{\lambda^x \cdot e^{-\lambda}}{x!} d\lambda}_{=1}} = \frac{\lambda^x \cdot e^{-\lambda}}{x!}$$

Da qui in poi possiamo calcolarci il valore atteso in modo molto semplice (ricordandoci che non è più una serie ma un integrale, siccome si che $x \in \mathbb{N}$, ma $\lambda \in \mathbb{R}$ e adesso stiamo calcolando quello):

$$\mathbb{E}[\lambda] = \int_0^{+\infty} \lambda \cdot \frac{\lambda^x \cdot e^{-\lambda}}{x!} d\lambda = \frac{1}{x!} \cdot \int_0^{+\infty} \lambda^{x+1} \cdot e^{-\lambda} d\lambda$$

E riprendendo l'Equazione 3.48 l'integrale diventa molto facile, otteniamo sia il valore atteso che la deviazione (che è la radice del valore atteso):

$$\mathbb{E}[\lambda] = x + 1 \quad (4.23)$$

$$\sigma_\lambda = \sqrt{x + 1} \quad (4.24)$$

Questo risultato, oltre ad essere comodo, ci mostra una cosa importante: i due metodi per fare inferenza sono molto vicini e compatibili tra loro, infatti all'aumentare dei conteggi e quindi di x quel "+1" diventa insignificante e ci ritroviamo nella sfera dell'inferenza con approssimazione Gaussiana.

4.6.2 Inferenza con 0 Conteggi

Un caso critico è quello in cui abbiamo misurato 0 conteggi, ma questo non vuol dire che si debbano necessariamente osservare 0 conteggi in Δt , magari siamo stati sfortunati. In questo caso otteniamo:

$$f(\lambda|x = 0, P) = e^{-\lambda} \quad (4.25)$$

$$F(\lambda|x = 0, P) = 1 - e^{-\lambda} \quad (4.26)$$

E questo banalmente dal fatto che se per un'inferenza generica vale che

$$f(\lambda|x, P) = \frac{\lambda^x \cdot e^{-\lambda}}{x!}$$

Sostituendo $x = 0$ si ottiene $e^{-\lambda}$. Trovandoci il valore atteso e la deviazione (semplicemente sostituendo nell'Equazione 4.24 o facendoci l'integrale) troviamo che

$$\mathbb{E}[\lambda] = 1 \quad (4.27)$$

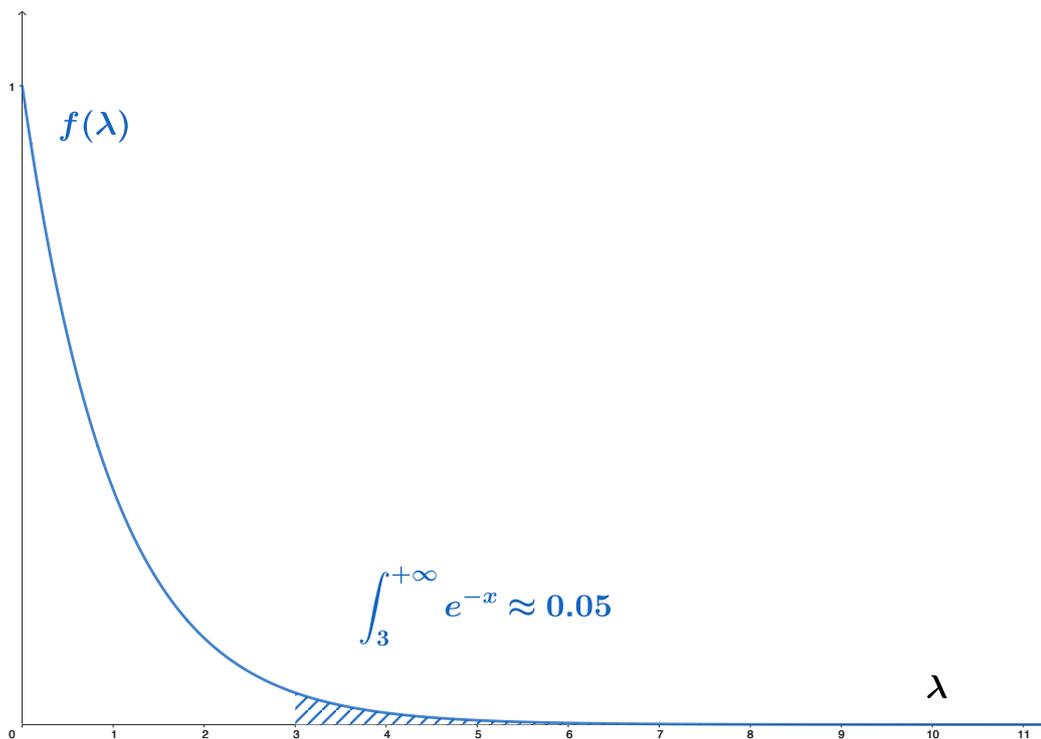
$$\sigma_\lambda = \sqrt{1} = 1 \quad (4.28)$$

Ovvero il valore atteso della frequenza si aggira tra 0 e 2. Questo caso critico, che può sembrare banale siccome basta sostituire, ha un'importanza nel mondo fisico: se cerchiamo di istituire un "limite" superiore, il massimo che λ può assumere sulla base dell'osservazione di 0 conteggi, magari con una confidenza del 95%, troviamo che

$$\int_0^t \frac{\lambda^x \cdot e^{-\lambda}}{x!} d\lambda = 0.95 \implies t \approx 3$$

Cioè con una probabilità del 95% la frequenza non sarà maggiore di 3, il che ha senso perchè se guardiamo il grafico di $f(\lambda)$ vediamo come il valore più probabile sia chiaramente 0 (avendo osservato 0 conteggi) e andando avanti la probabilità diminuisce esponenzialmente. Ironico, no?

L'importanza fisica quindi che ricaviamo da questo caso critico è la capacità di sapersi comportare in una misura del genere: se siamo in laboratorio e misuriamo per 4/5 volte di fila zero conteggi, ma sentiamo bisbigliare dal professore che il risultato dovrebbe essere $\lambda \approx 2$ non dobbiamo spaventarci troppo, le nostre misurazioni sono concordi e possiamo anche scrivere i nostri risultati con sicurezza, o aumentare il numero di misure. Se però sentiamo bisbigliare che il risultato dovrebbe essere $\lambda \approx 4$ ecco che dobbiamo iniziare a preoccuparci.



4.7 Combinazione di Risultati

Immaginiamo di fare n osservazioni di un evento, come ad esempio l'intensità di diversi processi di poisson indipendenti tali che:

$$f(\lambda) = \frac{\lambda^x e^{-\lambda}}{x}$$

Dove abbiamo che $\lambda = r t_i$, e siamo interessati a conoscere la migliore stima di r che tiene conto di tutte le misure effettuate.

Applichiamo allora il metodo della massima verosimiglianza:

$$f(r|x, t) \propto \prod_{i=1}^n (r t_i)^{x_i} e^{-r t_i} \propto \prod_{i=1}^n r^{x_i} e^{-r \sum_i t_i} = r^{\sum_i x_i} e^{-r \sum_i t_i} = r^{x_{tot}} e^{-r t_{tot}}$$

Allora abbiamo che $\mathcal{L} = r^{x_{tot}} e^{-r t_{tot}}$, per il metodo della massima verosimiglianza otteniamo:

$$\mathbb{E}[r] = \frac{\partial \log \mathcal{L}}{\partial r} = 0$$

Calcoliamo prima:

$$\log \mathcal{L} = \log r^{x_{tot}} + \log e^{-r t_{tot}} = x_{tot} \log r - r t_{tot}$$

Adesso facciamo la derivata rispetto ad r ed imponiamola nulla:

$$\frac{\partial \log \mathcal{L}}{\partial r} = \frac{x_{tot}}{r} - t_{tot} = \frac{x_{tot}}{r} = t_{tot} \rightarrow r = \frac{x_{tot}}{t_{tot}}$$

Per calcolare l'incertezza su r calcoliamo la derivata seconda sempre rispetto ad r ed otterremo:

$$\sigma(r) = \frac{\sqrt{x_{tot}}}{t_{tot}} \quad (4.29)$$

4.8 Inferenza sulla distribuzione binomiale

L'inferenza statistica sulla distribuzione binomiale è un processo che consente di fare previsioni o inferenze su una popolazione o su un insieme di dati utilizzando una distribuzione binomiale, può essere utilizzata per determinare le proprietà della popolazione binomiale, come la media e la varianza, e per fare previsioni su eventi futuri, come il numero di successi o fallimenti in un certo numero di prove.

In questa sezione affrontiamo il caso generale dell'inferenza di p dalla conoscenza di n e di $X = x$ e dall'aver assunto un processo di Bernoulli indipendente per ogni esito, sotto condizione che p valga un certo valore potremmo scrivere:

$$f(p|x, n, B)$$

Ricordandoci il teorema di Bayes, sappiamo che non possiamo scrivere che $f(p|x, n, B) = f(x|p, n, B)$ ma dovremmo scrivere:

$$f(p|x, n, B) = \frac{f(x|B_{n,p}) \cdot f_{\circ}(p)}{\int_0^1 f(x|B_{n,p}) \cdot f_{\circ}(p) dp}$$

Dove osserviamo che l'integrale sul dominio di p va da 0 a 1 che sono gli estremi valori che una probabilità può assumere. Possiamo riscrivere il tutto come:

$$\frac{\frac{n!}{(n-x)! \cdot x!} \cdot p^x \cdot (1-p)^{n-x} \cdot \cancel{f_{\circ}(p)}}{\int_0^1 \frac{n!}{(n-x)! \cdot x!} \cdot p^x (1-p)^{n-x} \cdot \cancel{f_{\circ}(p)} dp}$$

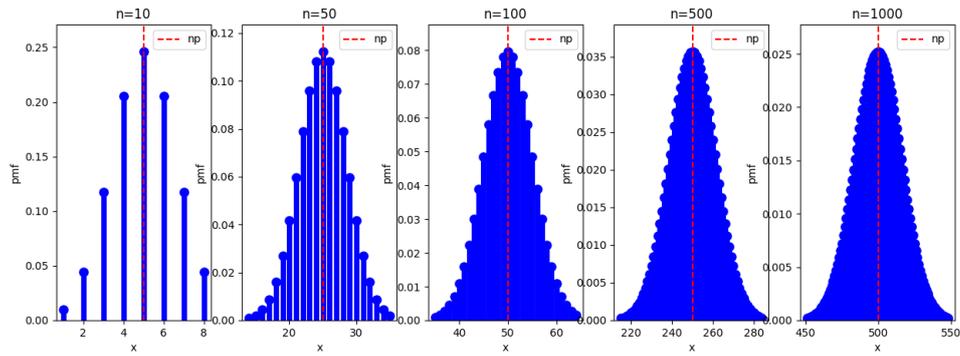
Fatte le opportune semplificazioni otteniamo:

$$\frac{p^x (1-p)^{n-x}}{\int_0^1 p^x (1-p)^{n-x} dp}$$

L'integrale al denominatore è lungo da risolvere, non ve lo chiederà neanche il padre eterno e dunque scriviamo direttamente il risultato:

$$f(p|x, n, B) = \frac{(n+1)!}{x! \cdot (n-x)!} \cdot p^x \cdot (1-p)^{n-x}$$

Possiamo notare che a parità di $x + 1$, al crescere di n si è sempre più sicuri su p . Inoltre per n grande e $x + 1$ lontano da 0 e da 1, la funzione finale $f(A, p)$, ha la forma Gaussiana.



Se adesso volessimo calcolare il valore atteso (non riporto i calcoli perchè non sono necessari ai fini del corso):

$$\begin{aligned} \mathbb{E}[p] &= \int_0^1 p \cdot f(p|m, x) dx = \int_0^1 p \cdot \frac{(n+1)!}{x!(n-x)!} p^x (1-p)^{n-x} dp = \frac{x+1}{n+2} \implies \\ &\implies \mathbb{E}[p] = \frac{x+1}{n+2} \end{aligned} \quad (4.30)$$

Ora definiamo anche la varianza:

$$\text{Var}[p] = \frac{(x+1) \cdot (n-x+1)}{(n+3) \cdot (n+2)^2} = \frac{x+1}{n+2} \cdot \left(\frac{n+2}{n+2} - \frac{x+1}{n+2} \right) \cdot \frac{1}{n+3} = \frac{\mathbb{E}[p](1-\mathbb{E}[p])}{n+3} \quad (4.31)$$

È chiaro che nella sfera di x, n molto grandi si approssima il tutto a una Gaussiana, o come abbiamo imparato dall'Inferenza sulla Poissoniana, quei "+2" o "+3" diventano insignificanti rispetto ai valori molto grandi e possono essere tolti, rimanendo con:

$$\mathbb{E}[p] = \frac{x}{n} \quad (4.32)$$

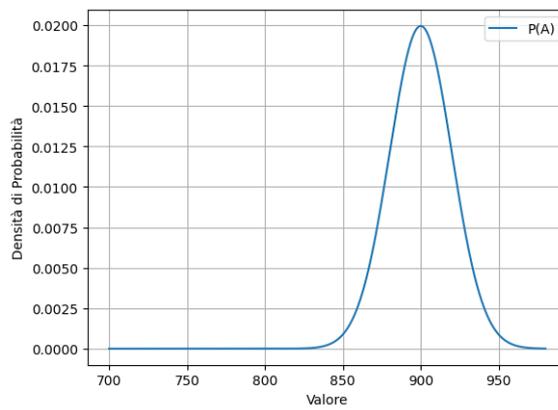
$$\sigma_p = \sqrt{\frac{\mathbb{E}[p](1-\mathbb{E}[p])}{n}} \quad (4.33)$$

4.9 Inferenza su un esperimento

Immaginiamo due scienziati A e B che devono dare la stima di una stessa costante fisica θ . Sappiamo che lo scienziato A è un bravo sperimentatore ed ha grande esperienza con le misure e le stime degli errori. Invece B è meno esperto, da poco ha iniziato a lavorare come ricercatore e dunque immaginiamo che la sua prior sia molto più debole di quella di A. Per prior intendiamo la conoscenza a priori riguardo la costante θ .

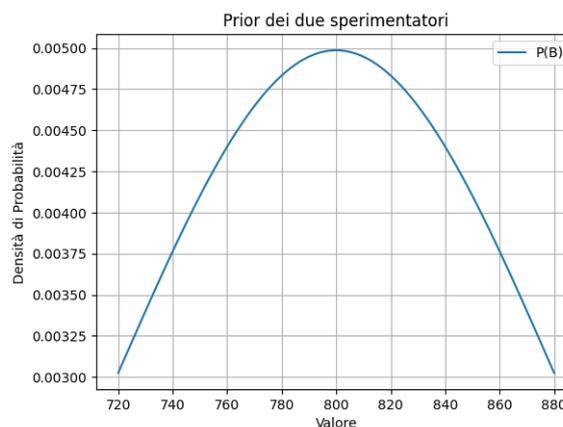
La prior di A si distribuisce secondo una normale con $\mu = 900$ e $\sigma = 20$:

$$P_A(\theta) = \frac{1}{\sqrt{2\pi}20} \cdot \exp\left[-\frac{1}{2}\left(\frac{\theta - 900}{20}\right)^2\right]$$

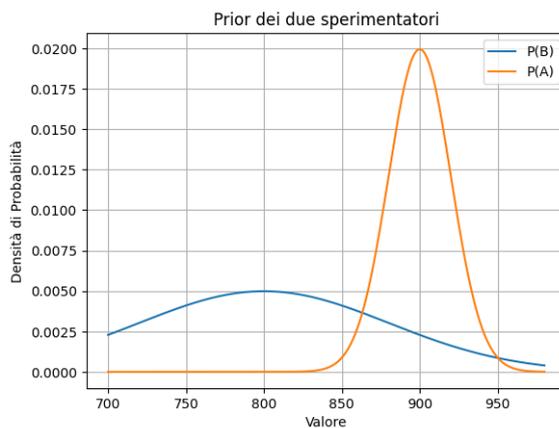


Invece la prior di B si distribuisce secondo una normale con $\mu = 800$ e $\sigma = 80$:

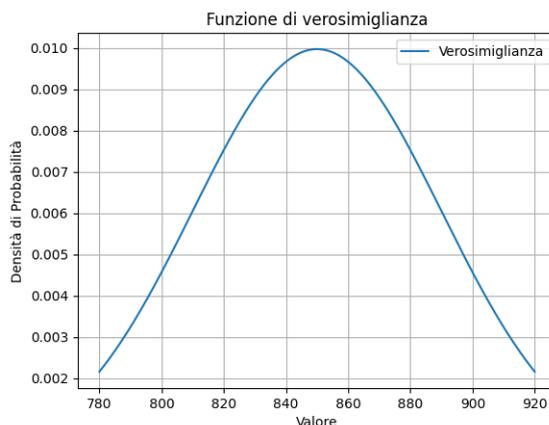
$$P_B(\theta) = \frac{1}{\sqrt{2\pi}80} \cdot \exp\left[-\frac{1}{2}\left(\frac{\theta - 800}{80}\right)^2\right]$$



E' evidente che combinando in unico grafico le due prior quella di A risulta molto più forte di quella di B, che invece è molto vaga:



Supponiamo adesso che esista un metodo imparziale di misurazione sperimentale e che una misura y fatta con questo metodo segua una normale con $\mu = \theta$ e $\sigma = 40$. Ora che abbiamo una misura y la funzione di verosimiglianza si distribuirà come una normale centrata in y e con deviazione standard uguale a 40. Dunque adesso possiamo applicare il teorema di Bayes per mettere in evidenza come le informazioni di ciascun sperimentatore su θ vengano modificate dai dati raccolti sperimentalmente. Supponiamo che y sia uguale ad 850, allora la funzione di verosimiglianza sarà del tipo:



Dunque quello che ci interessa calcolarci tramite il teorema di Bayes è la posterior di A e di B rispettivamente:

$$P_A(\theta|y) = \frac{P_A(y|\theta)P_A(\theta)}{P(y)}$$

$$P_B(\theta|y) = \frac{P_B(y|\theta)P_B(\theta)}{P(y)}$$

Dunque avendo supposto che le distribuzioni a priori siano distribuite secondo una normale di parametro θ e che la funzione di verosimiglianza sia proporzionale ad una funzione normale del tipo:

$$P(\theta|y) \propto \exp \left[-\frac{1}{2} \left(\frac{\theta - x}{\sigma} \right)^2 \right]$$

Dove x è funzione dell'osservazione y . Allora le posteriori dei due sperimentatori diventeranno:

$$P_A(\theta|y) = \frac{P_A(y|\theta)P_A(\theta)}{P(y)} = \frac{f_A(\theta|y)}{\int_{-\infty}^{+\infty} f_A(\theta|y)d\theta}$$

Dove abbiamo che:

$$f_A(\theta|y) = \exp \left\{ -\frac{1}{2} \left[\left(\frac{\theta - 900}{20} \right)^2 - \left(\frac{850 - \theta}{40} \right)^2 \right] \right\}$$

Utilizzando l'identità:

$$A(z - a)^2 + B(z - b)^2 = (A + B)(z - c)^2 + \frac{AB}{A + B} \cdot (a - b)^2$$

In cui

$$c = \frac{1}{A + B} \cdot (Aa + Bb)$$

Noi potremmo scrivere allora:

$$\left(\frac{\theta - 900}{20} \right)^2 + \left(\frac{850 - \theta}{40} \right)^2 = (\sigma_A^{-2} + \sigma_y^{-2}) \cdot (\theta - \bar{\theta})^2 + d$$

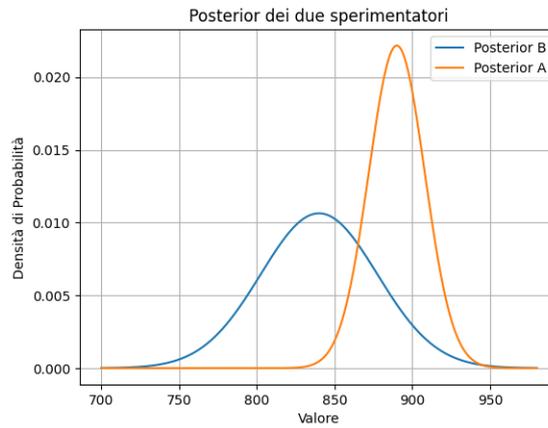
Allora avremmo che:

$$\bar{\theta} = \frac{w_0 \cdot 900 + w_1 \cdot 850}{w_0 + w_1}$$

Dove abbiamo che:

$$\frac{1}{\sigma^2} = w_0 + w_1$$

Dunque calcolando la posterior di A avremo che $f_A(\theta|y)$ si distribuisce come una normale con valore atteso=890 e sigma uguale a 18, invece la posterior di B $f_B(\theta|y)$ come una normale con valore atteso pari a 840 e sigma 35.70. Dunque graficamente avremo che:



Si nota dai due grafici che mentre la prior di A è stata influenzata poco dalla funzione di verosimiglianza, non possiamo dire lo stesso per B che invece ha acquisito importanti informazioni dall'esperimento e si nota come la sigma si sia ridotta di 40 unità e come la distribuzione abbia assunto una forma più definita e a campana, d'altronde notiamo che i valori attesi degli scienziati che prima erano significativamente distanti tra loro, hanno accorciato di molto questa distanza. Allora ci chiediamo se effettuassero N misurazioni cosa accadrebbe??

Supponiamo che vengano eseguite altre 99 misurazioni e che la media $\bar{y} = \frac{1}{100} \sum y_i$ sia pari ad 870. In generale la funzione di verosimiglianza di θ date n misurazioni indipendenti è:

$$f(\theta|y) \propto \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \cdot \exp \left[-\frac{1}{2\sigma^2} \sum (y_i - \theta)^2 \right]$$

Inoltre sappiamo che:

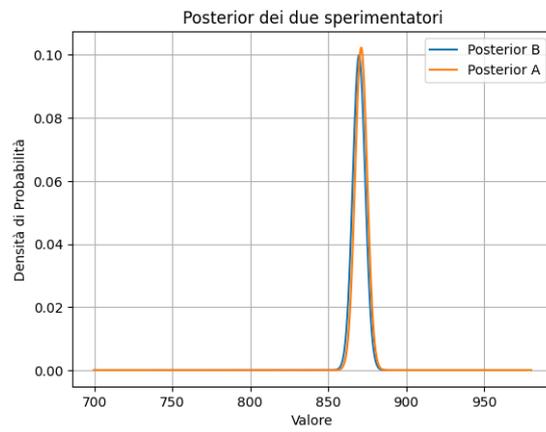
$$\sum (y_i - \theta)^2 = \sum (y_i - \bar{y})^2 + n(\theta - \bar{y})^2 \quad (4.34)$$

Poichè dati i dati, $\sum (y_i - \bar{y})^2$ è una costante fissa, allora la funzione di verosimiglianza sarà:

$$f(\theta|y) \propto \exp \left[-\frac{1}{2} \left(\frac{\theta - \bar{y}}{\sigma/\sqrt{n}} \right)^2 \right]$$

Che è una distribuzione normale centrata in \bar{y} e con deviazione standard σ/\sqrt{n} .

Nel nostro esempio la funzione sarà centrata in 870 con una deviazione standard uguale a 4. Dunque combiniamo come fatto in precedenza la funzione a priori di A e di B con la funzione di verosimiglianza ottenuta facendo N esperimenti e troveremo che:



Si nota subito come all'aumentare delle N misurazioni le due distribuzioni sono arrivate quasi a coincidere questo perchè le informazioni che provengono dall'esperimento sono molto più forti della conoscenza a priori.

4.10 Inferenza sui parametri di una legge (Fit lineari)

Supponiamo di aver misurato delle coppie ordinate (x_i, y_i) e di averne osservato l'andamento graficamente e immaginiamo inoltre di aver ipotizzato che queste grandezze siano legate tra loro da una relazione matematica. Immaginiamo di trovarci nel caso più semplice, ovvero di un andamento lineare del tipo:

$$\mu_y = \mu_x \cdot m + c$$

Indichiamo con x e y i valori osservabili, con m e c i parametri della retta di regressione e il valore vero delle variabili con μ_i , la cui verosimiglianza sarà:

$$f(x_i, y_i | \mu_x, m, c) = f(x_i | \mu_x) \cdot f(y_i | \mu_x, m, c) \propto \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x_i - \mu_x)^2}{2\sigma_x^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{(y_i - \mu_x \cdot m - c)^2}{2\sigma_y^2}}$$

E questo perchè il valore vero di x non ha dipendenza dagli altri parametri (non a caso si chiama variabile indipendente in analisi), mentre il valore vero di y ha dipendenza da tutti gli altri parametri che vanno inseriti quindi nella sua funzione di verosimiglianza. La verosimiglianza inoltre diventa più semplice quando $\sigma_x \rightarrow 0$. In questo modo il valore $\mu_x \approx x_i$, ovvero il valore vero è uguale ai valori osservabili e dunque la verosimiglianza si riduce a quella su y , ovvero:

$$f(x_i, y_i | m, c) = \frac{1}{\sqrt{2\pi}\sigma_y} \cdot \exp\left[-\frac{(y_i - m \cdot x_i - c)^2}{2\sigma_y^2}\right]$$

Questo passaggio che abbiamo fatto è lecito? Domanda intelligente, e la risposta è ni. Come ogni cosa, bisogna sempre ragionare su quello che si ha e su quello che si vuole ottenere: le approssimazioni sono alla base dell'analisi dei dati in alcuni passaggi matematici difficili, così come abbiamo fatto per la semplificazione di prior vaghe in altri tipi di inferenza. Quindi sì, possiamo assumere l'incertezza sui punti x_i pari a 0 nel momento in cui non "tagliamo troppe informazioni", ovvero quando effettivamente la $\sigma_x \ll \sigma_y$.

Andando oltre, abbiamo trovato la funzione di verosimiglianza di una singola coppia di punti. Immaginando di avere tante coppie di valori misurate e che le possibili fluttuazioni intorno alle ordinate siano indipendenti abbiamo, per la regola della moltiplicazione, che:

$$f(\vec{x}, \vec{y} | m, c) = \prod_{i=1}^n f(x_i, y_i | m, c) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_y} \cdot \exp\left[-\frac{(y_i - m \cdot x_i - c)^2}{2\sigma_y^2}\right]$$

Allora possiamo applicare il teorema di Bayes ed ottenere:

$$f(m, c|x, y) \propto \prod_{i=1}^n f(x_i, y_i|m, c) \cdot f(m, c) \quad (4.35)$$

Supponendo come al solito che $f(m, c)$ sia distribuita uniformemente e che la posterior segua una distribuzione normale, ci interessa adesso calcolarci il massimo della distribuzione che corrisponde al massimo della verosimiglianza. Dunque si massimizza il logaritmo della verosimiglianza e si ottiene (i passaggi per l'arrivo a questo risultato sono gli stessi di [Sezione 8.6](#), ovvero il metodo di massima verosimiglianza già affrontato, quindi non riscriviamo tutti i passaggi, basta sostituire l'esponente e ci si ritrova con la stessa cosa):

$$\chi^2 = \frac{(y - mx_i - c)^2}{\sigma_{y_i}^2} \quad (4.36)$$

Prima di applicare a cuor leggero il metodo dei minimi quadrati consideriamo il caso in cui teniamo conto delle possibili incertezze sulle X , facciamo inferenza quindi su μ_x ed otteniamo:

$$f(\mu_x, m, c|x, y) \propto \prod_{i=1}^n f(x_i, y_i|\mu_x, m, c) \cdot f_0(\mu_x, m, c)$$

Adesso marginalizziamo ed otteniamo:

$$f(m, c|x, y) \propto \int f(\mu_x, m, c|x, y) d\mu_x$$

Assumendo che la prior $f_0(\mu_x, m, c)$ si distribuisca in modo uniforme abbiamo che se $\sigma_x \rightarrow 0$ allora le gaussiane che descrivono le probabilità di X_i intorno a μ_{x_i} diventano del tipo $\delta(x_i - \mu_{x_i})$ e allora:

$$f(m, c|x, y) \propto \prod_{i=1}^n \frac{1}{\sigma_{y_i} \cdot \sqrt{2\pi}} \cdot \exp\left[-\frac{(y_i - mx_i - c)^2}{2\sigma_{y_i}^2}\right]$$

Nel caso generale in cui $\sigma_{x_i} \neq 0$ lo trattiamo allo stesso modo ma l'integrale è *un pò più complesso*:

$$f(m, c|x, y) \propto \prod_{i=1}^n \frac{1}{\sqrt{2\pi} \sqrt{\sigma_{y_i}^2 + m^2 \sigma_{x_i}^2}} \cdot \exp\left[-\frac{(y_i - mx_i - c)^2}{2\sigma_{y_i}^2 + m^2 \sigma_{x_i}^2}\right]$$

Si nota in questa seconda relazione che se $\sigma_{x_i} \rightarrow 0$ allora si ritorna al caso generale.

4.10.1 Metodo dei Minimi quadrati

Il metodo dei minimi quadrati è un metodo per il calcolo delle costanti e delle loro incertezze e correlazioni. Partiamo con il minimizzare la formula del chi quadro rispetto a m e c . Facciamo questo perchè, ricordandoci che seguendo il metodo di massima verosimiglianza avevamo ridotto il logaritmo della funzione a una somma di termini negativi, ovvero:

$$\log[f(m, c|x, y)] = - \sum_{i=1}^n \left[\log(\sigma_y) + \log(\sqrt{2\pi}) + \frac{(y_i - m \cdot x_i - c)^2}{2\sigma_y^2} \right]$$

Questo vuol dire che massimizzare questa funzione equivale a minimizzare l'unico termine che dipende dalle variabili, ovvero l'Equazione 4.36 (da qui *metodo dei minimi quadrati*), e per farlo otteniamo due derivate parziali che dobbiamo porre uguali a 0:

$$\frac{\partial \chi^2}{\partial c} = -2 \sum_{i=0}^n \frac{y_i - mx_i - c}{\sigma_y^2} = -2 (\bar{y} - m\bar{x} - c) \cdot \sum_{i=0}^n \frac{1}{\sigma_y^2} = 0$$

$$\frac{\partial \chi^2}{\partial m} = -2 \sum_{i=0}^n \frac{x_i(y_i - m - x_i - c)}{\sigma_y^2} = -2 (\overline{xy} - c\bar{x} - m\overline{x^2}) \cdot \sum_{i=0}^n \frac{1}{\sigma_y^2} = 0$$

In tutto questo le medie sono pesate con pesi $p_i = 1/\sigma_y^2$, quindi le varie medie sarebbero:

$$\bar{x} = \frac{\sum_{i=0}^n p_i x_i}{\sum_{i=0}^n p_i} \qquad \bar{y} = \frac{\sum_{i=0}^n p_i y_i}{\sum_{i=0}^n p_i}$$

$$\overline{x^2} = \frac{\sum_{i=0}^n p_i x_i^2}{\sum_{i=0}^n p_i} \qquad \overline{xy} = \frac{\sum_{i=0}^n p_i x_i y_i}{\sum_{i=0}^n p_i}$$

A questo punto non resta che risolvere il sistema per trovarsi le costanti m e c :

$$\begin{cases} \bar{y} - c - m\bar{x} = 0 \\ \overline{xy} - c\bar{x} - m\overline{x^2} = 0 \end{cases} \implies \begin{cases} c = \bar{y} - m\bar{x} \\ \overline{xy} - m\overline{x^2} + m\bar{x}^2 - \bar{x} \cdot \bar{y} \end{cases} \implies \begin{cases} c = \bar{y} - m\bar{x} \\ m = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} \end{cases}$$

Risolto il sistema è facile: la migliore stima dei parametri sono i loro valori attesi, e ricordandoci che possiamo usare la media pesata come valore atteso (vedi [Sottosezione 3.9.1](#))

diventa:

$$\hat{m} = \mathbb{E}[m] = \mathbb{E} \left[\frac{\bar{x} \cdot \bar{y} - \overline{xy}}{\bar{x}^2 - \overline{x^2}} \right] = \frac{\mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y]}{\mathbb{E}[x^2] - \mathbb{E}^2[x]} = \frac{\text{Cov}[x, y]}{\text{Var}[x]}, \quad (4.37)$$

$$\hat{c} = \mathbb{E}[c] = \mathbb{E}[\bar{y} - m\bar{x}] = \bar{y} - \hat{m}\bar{x} \quad (4.38)$$

Notare bene che in queste formule $\text{Var}[x]$ non si riferisce allo scarto da x riferito alla sua σ_x , che per l'appunto abbiamo assunto 0, ma bensì a quella dei punti sull'asse delle y .

Ora tocca al momento più importante: calcolarsi le deviazioni standard delle costanti e la loro covarianza. Siccome la nostra è una relazione lineare possiamo usare le formula:

$$\begin{aligned} \text{Var}[m] &= \sum_{i=0}^n \left(\frac{\partial m}{\partial y_i} \right)^2 \cdot \sigma_{y_i}^2 \\ \text{Var}[c] &= \sum_{i=0}^n \left(\frac{\partial c}{\partial y_i} \right)^2 \cdot \sigma_{y_i}^2 \\ \text{Cov}[m, c] &= \sum_{i=0}^n \left(\frac{\partial m}{\partial y_i} \frac{\partial c}{\partial y_i} \right) \cdot \sigma_{y_i}^2 \end{aligned}$$

Riportiamo il conteggio per la varianza di entrambe le misure (sta volta la dimostrazione è lunga, impacciata e pallosa, e non ha a che fare con passaggi logici particolarmente rilevanti come nel caso dell'uso dell'inferenza per arrivare alle formule principali del fit, quindi lasciamo qui le parti principali del conto e quello più completo in ??):

$$\begin{aligned} \text{Var}[\hat{m}] &= \sum_{i=0}^n \left(\frac{\partial m}{\partial Y_i} \right)^2 \cdot \sigma_{Y_i}^2 = \sum_{i=0}^n \left(\frac{x_i}{\sigma_{Y_i}^2} - \frac{\bar{x}}{\sigma_{Y_i}^2} \right)^2 \frac{1}{\text{Var}^2[x]} \frac{1}{\left(\sum_{i=0}^n \frac{1}{\sigma_{Y_i}^2} \right)^2} \sigma_{Y_i}^2 = \\ &= \frac{1}{\text{Var}[x] \cdot \sum_{i=0}^n \frac{1}{\sigma_{Y_i}^2}} \end{aligned} \quad (4.39)$$

$$\begin{aligned} \text{Var}[\hat{c}] &= \sum_{i=0}^n \left(\frac{\partial c}{\partial Y_i} \right)^2 \cdot \sigma_{Y_i}^2 = \sum_{i=0}^n \left(\frac{\bar{x}^2}{\sigma_{Y_i}^2} - \frac{\bar{x}x_i}{\sigma_{Y_i}^2} \right)^2 \frac{1}{\text{Var}^2[x]} \frac{1}{\left(\sum_{i=0}^n \frac{1}{\sigma_{Y_i}^2} \right)^2} \sigma_{Y_i}^2 = [\dots] = \\ &= \frac{\bar{x}^2}{\text{Var}[x] \cdot \sum_{i=0}^n \frac{1}{\sigma_{Y_i}^2}} \end{aligned} \quad (4.40)$$

$$\text{Cov}[\hat{m}, \hat{c}] = -\bar{x} \cdot \text{Var}[\hat{m}] \quad (4.41)$$

4.10.2 σ_y nota e σ_x ignota

Consideriamo il caso in cui σ_y sia nota e supposta costante dunque $\sum_i^n 1/\sigma_{y_i}^2$ diventa: σ^2/n e allora le equazione per stimare le incertezze dei parametri del fit diventano:

$$\sigma_m = \frac{1}{\sqrt{\text{Var}[x]}} \frac{\sigma}{\sqrt{n}} \quad (4.42)$$

$$\sigma_c = \sqrt{x^2} \cdot \sigma_m \quad (4.43)$$

Nel caso in cui le σ fossero ignote non cambia nulla dal punto di vista delle formule. Ma le incertezze devono essere stimate considerando lo scarto tra le ordinate dei punti sperimentali e la retta che meglio approssima i punti. Come per la deviazione standard si calcola lo scarto tra le y_i e le funzioni calcolate in x_i , il problema può diventare piuttosto complesso quando si hanno pochi punti sperimentali. Infatti al denominatore non si usa \sqrt{n} bensì $\sqrt{n-2}$. Quindi la formula finale sarà:

$$\sigma = \frac{\sum_{i=1}^n (y_i - mx_i - c)^2}{\sqrt{n-2}} \quad (4.44)$$

4.10.3 Analisi del Baricentro

Osservando i risultati ottenuti si nota che se $\bar{x} = 0$ si annulla il coefficiente di correlazione, dunque conviene scegliere l'asse delle x nel baricentro dei punti e quindi operare il cambiamento di variabile:

$$x' = x - \bar{x} \quad (4.45)$$

Utile se poi dovessimo estrapolare un valore dal fit lineare oppure per rappresentare le curve di involucro. Con il cambiamento di variabili abbiamo che il fit adesso è sulle coppie ordinate di punti $(x' = x - \bar{x}, y' = y)$ dunque le grandezze trovate in precedenza si trasformano di conseguenza:

$$\mathbb{E}[m'] = \frac{\overline{xy}}{\overline{x^2}} \quad (4.46)$$

$$\mathbb{E}[c'] = \overline{y'} \quad (4.47)$$

$$\text{Var}[m'] = \frac{1}{\sum_{i=1}^n 1/\sigma_{y_i} \overline{x^2}} \quad (4.48)$$

$$\text{Var}[c'] = \frac{1}{\sum_{i=1}^n 1/\sigma_{y_i}} \quad (4.49)$$

Sapendo che :

$$m = m'$$

$$c = c' - m'\bar{x}$$

Allora otteniamo che:

$$\mathbb{E}[m] = \mathbb{E}[m'] = \frac{\overline{x'y'}}{\overline{x'^2}} = \frac{\overline{x - \bar{x}} - \overline{xy'}}{\overline{x - \bar{x}}^2} = \frac{\overline{xy} - \bar{x}\bar{y}}{\text{Var}[x]} \quad (4.50)$$

4.10.4 Estrapolazione

Similmente all'inferenza predittiva, una volta scoperta la retta che approssima meglio i dati sperimentali in nostro possesso, un risultato importante è quello di estrapolare un valore "futuro" della nostra misura che segue sempre la relazione lineare che abbiamo trovato, ma cosa più importante: valutare l'incertezza di tale previsione con questo metodo.

Proprio così, le incertezze stanno sempre in mezzo. Questo deriva principalmente che, esattamente come l'inferenza predittiva, il valore futuro in sé è quasi banale, esso equivale al parametro y_i corrispondente al relativo x_i del punto calcolato sulla retta, o meglio, se la nostra retta è:

$$y = mx + c$$

$$\mathbb{E}[y] = \mathbb{E}[m]x + \mathbb{E}[c] = \hat{m}x + \hat{c}$$

E quindi basta sostituire i valori corrispondenti per trovarci una buona previsione. Per quanto riguarda l'incertezza, invece, che è leggermente meno ovvia, bisogna farsi la propagazione delle incertezze (seguendo il concetto di linearizzazione visto in [Sezione 2.12](#)) tenendo conto della covarianza tra coefficiente angolare e intercetta:

$$\sigma_y^2 = \left(\frac{\partial y}{\partial m}\right)^2 \cdot \sigma_m^2 + \left(\frac{\partial y}{\partial c}\right)^2 \cdot \sigma_c^2 + 2 \left(\frac{\partial y}{\partial m} \frac{\partial y}{\partial c}\right) \cdot \text{Cov}[m, c]$$

$$\sigma_y^2 = \frac{\sigma^2}{n} + \frac{(\mathbb{E}[y] - \bar{x})^2}{\text{Var}[x]} \cdot \frac{\sigma^2}{n}$$

4.10.5 Curve di Involuppo

Il maggior utilizzo della retta di regressione è quello di stimare la migliore risposta per $\mathbb{E}[y]$ dato un particolare valore della variabile di regressione x . Dunque vogliamo calcolare $\mathbb{E}[y]$ dato x_0 che sono i punti su cui è costruita la nostra retta di regressione:

$$\mu_y = m\mu_x + c$$

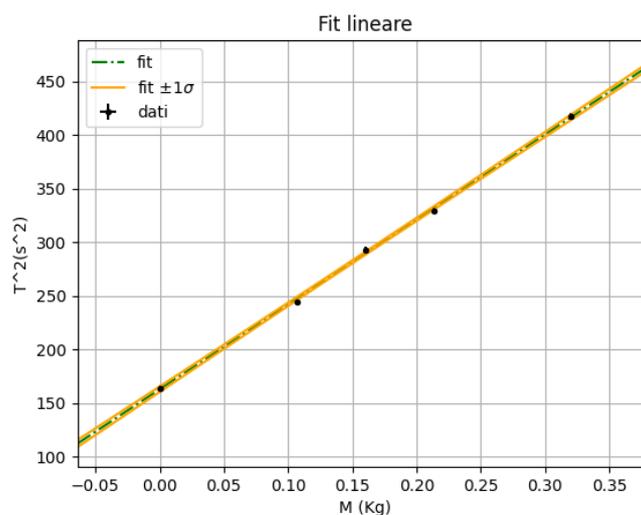
Di questa ne vogliamo calcolare il valore atteso, considerando che $\mu_x \approx x$:

$$\mathbb{E}[\mu_y(x)] = \mathbb{E}[m]\mu_x + \mathbb{E}[c] \quad (4.51)$$

E adesso per definire un intervallo di confidenza che saranno proprio le curve di involuppo calcoliamo la risposta media della variabile y rispetto ad x :

$$\begin{aligned} \text{Var}[\mu_y(x)] &= \mathbb{E} \left[(mx + c - \mathbb{E}[mx + c])^2 \right] = \\ &= \mathbb{E} \left[((mx - \mathbb{E}[m]\mu_x) - (c - \mathbb{E}[c]))^2 \right] = \\ &= \mathbb{E} \left[(mx - \mathbb{E}[m]\mu_x)^2 - 2(mx - \mathbb{E}[m]\mu_x)(c - \mathbb{E}[c]) + (c - \mathbb{E}[c])^2 \right] = \\ &= \text{Var}[m]\mu_x^2 + \text{Var}[c] + 2\mu_x \text{Cov}[m, c] = \\ &= \frac{1}{\sum_{i=1}^n 1/\sigma_{y_i}} \text{Var}[x] [\mu_x^2 + \bar{x}^2 - 2\bar{x}\mu_x + \bar{x}^2 - \bar{x}^2] = \\ &= \frac{1}{\sum_{i=1}^n 1/\sigma_{y_i}} [(\mu_x - \bar{x})^2 + \text{Var}[x]] \end{aligned}$$

Questa è una buona stima dell'intervallo di confidenza di ogni valore x rispetto ad $\mathbb{E}[y]$, che verrà rappresentato attraverso le curve di involuppo che seguiranno un andamento quadratico del tipo:



4.10.6 Analisi dei residui

Una volta eseguito un fit una tra le operazioni di valutazione riguardo la scelta del modello che possiamo fare è analizzare i residui, che ci dicono molto su come sono state eseguite le misure e soprattutto se il modello scelto è coerente con i dati sperimentali.

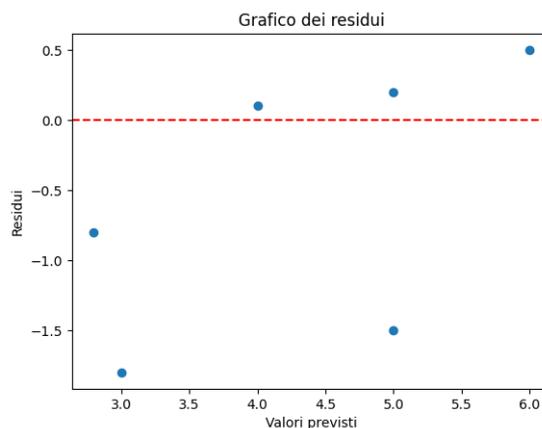
Quando eseguiamo un fit bisogna tenere in mente che questo non coglie tutte le variabilità presenti in un set di dati. La parte di variabilità è proprio il residuo della regressione, ovvero la distanza tra la retta del fit e il dato sperimentale e quindi ci indicano le fluttuazioni dei dati intorno a quest'ultima.

Dunque per valutare se le misure sono state eseguite correttamente e se i dati sperimentali sono in accordo con il modello scelto sicuramente l'analisi dei residui è lo strumento che possiamo utilizzare.

In primis possiamo chiederci: "come possiamo quantificare la bontà dell'accordo tra un modello e le misure effettuate?", sicuramente un metodo valido è il test del χ^2 :

$$\chi^2(x|\mu(x|\mu, \sigma)) = \sum_{i=1}^n \left[\frac{x_i - \mu_{x_i}}{\sigma} \right]^2 \quad (4.52)$$

Sostanzialmente stiamo confrontando le misure con il valore previsto dal modello nell'ipotesi che queste siano distribuite normalmente attorno a $\mu(x)$ con $\sigma(x)$. Una volta calcolato il χ^2 lo si confronta con l'intervallo di accettazione stabilito a priori. Solitamente però il miglior metodo non è quello del χ^2 , ciò ci può dare una buona stima di come sono distribuite le nostre misure a seconda del modello scelto, ma mettiamoci nel caso in cui osservando i residui del fit ci troviamo di fronte ad un caso del genere:



"Come procediamo??"...Beh allora i casi possibili sono 3:

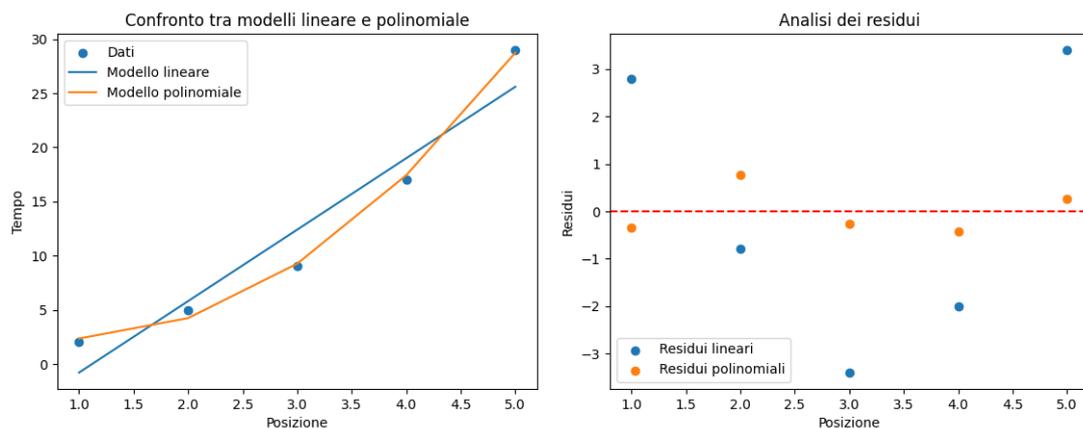
- Sfortuna nel fare le misure.
- Incertezze sottostimate o sovrastimate
- Modello inappropriato

Dunque per scongiurare il primo caso ci rimettiamo a tavolino rifacciamo le nostre misure e realizziamo un grafico dei residui delle nuove misure, possiamo anche realizzare un istogramma e vedere se i residui si distribuiscono come una normale standardizzata con

$\mu = 0$ e $\sigma = 1$. Se una volta eseguite le misure notiamo ancora grande discrepanza da quello che ci aspettiamo e quello che effettivamente abbiamo ottenuto, potrebbe esserci un errore su come abbiamo stimato le incertezze che potrebbero risultare sovrastimate o sottostimate. Quindi una volta ricontrollati e visto che la situazione non cambia possiamo scegliere due vie:

- Un triplo carpiato dal Burj Khalifa
- Possiamo domandarci "Abbiamo scelto il modello appropriato?".

Ora riportiamo un esempio in modo tale da rendere evidente come ci si può rendere conto dell'errore nella scelta del modello dall'analisi dei residui. Immaginiamo di eseguire delle misure sulle posizioni di un corpo che si muove e fare un fit delle posizioni in funzioni del tempo ed applichiamo sia il modello lineare che quello polinomiale per vedere quale dei due si adatta meglio alle nostre misure, riportiamo inoltre l'analisi dei residui per vedere come si distribuiscono questi ultimi intorno al valore medio $\mu = 0$:



Vediamo come il fit polinomiale si adatti senza dubbio meglio ai dati sperimentali rispetto al modello di regressione lineare, si nota anche dall'analisi dei residui dove quelli del fit polinomiale sono distribuiti intorno alla media in un intervallo di $\pm\sigma$ mentre quelli del fit lineare si discostano fino a 4σ dalla media.

5

Altri Concetti



Sommario

- 5.1 Teorema del limite centrale, 116
- 5.2 Legge dei grandi numeri, 118
- 5.3 Disuguaglianza di Markov, 120
- 5.4 Teorema di Čebyšëv, 121
- 5.5 Random Walk, 123
- 5.6 Metodo di Monte Carlo, 126

5.1 Teorema del limite centrale

Il teorema del limite centrale è uno dei risultati fondamentali della teoria della probabilità. Esso afferma che, sotto determinate condizioni, la somma di un gran numero di variabili aleatorie indipendenti ed identicamente distribuite converge ad una p.d.f. normale.

Definizione 28

Se abbiamo una sequenza di variabili aleatorie indipendenti ed identicamente distribuite, con media μ e varianza σ^2 , allora la somma di queste variabili aleatorie, opportunamente normalizzata, tende ad una distribuzione normale standard.

Date n variabili x_i indipendenti di p.d.f. sconosciuta, purchè $\mathbb{E}[x_i]$ e $\sigma[x_i]$ esistano (finiti) e le $\sigma[x_i]$ siano dello stesso ordine di grandezza, allora la variabile

$$y = \sum_{i=1}^n a_i x_i$$

Con a_i coefficienti reali, ha una p.d.f. asintoticamente gaussiana con:

$$\mathbb{E}[y] = \sum_{i=1}^n a_i \mathbb{E}[x_i] \quad (5.1)$$

$$\sigma[y] = \sqrt{\sum_{i=1}^n a_i^2 \cdot \sigma^2[x_i]} \quad (5.2)$$

Dimostrazione 8: NON DA ESAME

Consideriamo una sequenza di n variabili aleatorie indipendenti ed identicamente distribuite $X_1, X_2 \dots X_n$, con media μ e varianza σ^2 . Definiamo la somma $S_n = X_1 + X_2 + \dots + X_n$ e la media campionaria $\bar{X} = \frac{1}{n} S_n$.

La funzione generatrice dei momenti di una variabile aleatoria X è definita come $M_X(t) = \mathbb{E}[e^{tX}]$. Inoltre, se X ha media μ e varianza σ^2 , allora la funzione generatrice dei momenti di X è data da:

$$M_X(t) = \mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx = \exp\left(t\mu + \frac{1}{2}t^2\sigma^2\right) \quad (5.3)$$

Dove $f_X(x)$ è la funzione di densità di probabilità di X . La funzione generatrice dei momenti di S_n è data da:

✓ Dimostrazione 8 continued

$$M_{S_n}(t) = \prod_{i=1}^n M_{X_i}(t) = \left(e^{t\mu + \frac{1}{2}t^2\sigma^2} \right)^n = \exp\left(tn\mu + \frac{1}{2}t^2n\sigma^2 \right)$$

La funzione generatrice dei momenti di \bar{X} è data da:

$$M_{\bar{X}}(t) = \mathbb{E}[e^{t\bar{X}}] = \mathbb{E}\left[e^{\frac{t}{n}S_n} \right] = M_{S_n}\left(\frac{t}{n}\right) = \exp\left(t\mu + \frac{1}{2}\frac{t^2}{n}\sigma^2 \right)$$

Possiamo ora utilizzare la funzione generatrice dei momenti di \bar{X} per dimostrare il teorema del limite centrale. In particolare, possiamo dimostrare che la distribuzione di \bar{X} tende ad una distribuzione normale standard. Per fare ciò, consideriamo la seguente trasformazione:

$$Z_n = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Dove μ e σ sono rispettivamente la media e la deviazione standard di X . Possiamo dimostrare che Z_n converge in distribuzione ad una distribuzione normale standard:

$$\lim_{n \rightarrow \infty} \mathbb{P}(Z_n \leq z) = \Phi(z)$$

Dove $\Phi(z)$ è la funzione di distribuzione normale standard. La dimostrazione di questo risultato può essere fatta utilizzando il teorema di continuità di Levy.

$$\lim_{n \rightarrow \infty} \varphi_{Z_n}(t) = e^{-\frac{t^2}{2}}$$

Dove $\varphi_{Z_n}(t)$ è la funzione caratteristica di Z_n . Infine, utilizziamo il teorema di continuità di Levy, possiamo dimostrare che Z_n converge in distribuzione ad una distribuzione normale standard, ovvero:

$$\lim_{n \rightarrow \infty} \mathbb{P}(Z_n \leq z) = \Phi(z)$$

Dove $\Phi(z)$ è la funzione di distribuzione normale standard. La dimostrazione del teorema del limite centrale con il teorema di continuità di Levy può essere scritta come segue:

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(Z_n \leq z) &= \lim_{n \rightarrow \infty} \int_{-\infty}^z f_{Z_n}(u) du \\ &= \lim_{n \rightarrow \infty} \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \\ &= \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = \Phi(z) \end{aligned} \tag{5.4}$$

Dove abbiamo utilizzato la definizione di funzione di distribuzione, la funzione caratteristica di Z_n e il teorema di continuità di Levy.

5.2 Legge dei grandi numeri

Definizione 29

La legge dei grandi numeri descrive il comportamento della media di una sequenza di n prove di una variabile casuale, indipendenti e caratterizzate dalla stessa distribuzione di probabilità, al tendere ad infinito della numerosità n della sequenza stessa.

In pratica, grazie alla legge dei grandi numeri, possiamo essere ragionevolmente sicuri che la media sperimentale, che calcoliamo a partire da un numero sufficiente di campioni, sia sufficientemente vicina alla media vera, ovvero quella calcolabile teoricamente.

Se, data una successione di variabili casuali X_1, X_2, \dots, X_n indipendenti e identicamente distribuite con media finita μ si considera la media campionaria:

$$\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n}$$

Teorema 10: Legge debole dei grandi numeri

La legge (debole) dei grandi numeri afferma che per ogni $\varepsilon > 0$:

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| < \varepsilon) = 1 \quad (5.5)$$

Ossia la media campionaria converge in probabilità al valore atteso comune X_i

Teorema 11: Legge forte dei grandi numeri

La legge (forte) dei grandi numeri afferma che:

$$P\left(\lim_{n \rightarrow \infty} \bar{X}_n = \mu\right) = 1 \quad (5.6)$$

Ossia lo stimatore media campionaria converge quasi certamente al valore atteso comune delle X_i .

Un modo per testare il funzionamento della legge dei grandi numeri è quello di generare prima 100 poi 1000 poi 10000 e infine 1000000 di punti distribuiti in modo uniforme tra 0 e 1 e calcolarne la media per vedere se all'aumentare delle n prove la differenza tra \bar{x} e il valore atteso diminuisce:

```
1 import random
2
3 def calcola_media(num_prove):
4     somma = 0
5     for _ in range(num_prove):
6         numero_casuale = random.uniform(0, 1) # Genera un numero
           casuale compreso tra 0 e 1
7         somma += numero_casuale
8
9     media = somma / num_prove
10    return media
11
12 valore_atteso = 0.5 # Valore atteso della variabile casuale. In
           questo caso, la media di numeri casuali tra 0 e 1
13 num_prove = 100000000 # Numero di prove da eseguire
14
15 media_calcolata = calcola_media(num_prove)
16 errore = abs(media_calcolata - valore_atteso)
17
18 print(f"Valore atteso: {valore_atteso}")
19 print(f"Media calcolata: {media_calcolata}")
20 print(f"Errore: {errore}")
```

Codice 5.1: Funzione per verifica legge dei grandi numeri

5.3 Disuguaglianza di Markov

Questa disuguaglianza permette di stabilire un limite superiore al valore di probabilità dalla sola conoscenza del valore atteso $\mathbb{E}[X]$, a condizione che la variabile casuale sia definita non negativa.

π Teorema 12

In teoria della probabilità, la disuguaglianza di Markov afferma che per una variabile casuale X non negativa il cui valore atteso esiste:

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a} \quad (5.7)$$

\checkmark Dimostrazione 9

Si definiscano le variabili casuali X ed I come segue:

$$X : \Omega \rightarrow \mathbb{R}_0^+ ; I = \begin{cases} 1 & \text{se } X \geq a \\ 0 & \text{se } X < a \end{cases}$$

Con Ω spazio campionario, $a \in \mathbb{R}_0^+$. Inoltre, per ogni α non nullo, vale che $0 \leq I \leq \frac{X}{\alpha}$. Supponiamo adesso che per la variabile aleatoria X esista $\mathbb{E}[X]$, allora:

$$P(X \geq a) = \mathbb{E}[I]$$

Il valore atteso è definito come somma di tutti i valori che la variabile aleatoria può assumere moltiplicati per la probabilità che tale variabile assuma effettivamente tali valori:

$$\mathbb{E}[I] = 0 \cdot P(I = 0) + 1 \cdot P(I = 1) = P(I = 1)$$

Ma ancora la probabilità che I sia uguale ad 1 è proprio la probabilità che X sia maggiore o uguale ad a :

$$\mathbb{E}[I] \leq \mathbb{E}\left[\frac{X}{a}\right]$$

Il valore atteso mantiene la disuguaglianza degli argomenti (quella detta all'inizio, ovvero che $0 \leq I \leq \frac{X}{\alpha}$) poiché si tratta di una funzione non decrescente, in vista del fatto che gli argomenti sono variabili non negative. Dunque otteniamo per la linearità del valore atteso:

$$\mathbb{E}\left[\frac{X}{a}\right] = \frac{\mathbb{E}[X]}{a}$$

Allora concludiamo che:

$$P(X \geq a) \leq \frac{\mathbb{E}[x]}{a}$$

5.4 Teorema di Čebyšëv

Il teorema di Čebyšëv (*Chebyshev*), o disuguaglianza di Čebyšëv, è un importante risultato teorico in statistica che fornisce una stima generale sull'intervallo di dispersione dei dati in una distribuzione, indipendentemente dalla sua forma specifica. Esso, derivante dal teorema di Markov, stabilisce una relazione tra la deviazione standard e la probabilità che un dato casuale si discosti da un certo valore medio:

π Teorema 13

Per qualsiasi distribuzione con una deviazione standard finita, almeno $1 - 1/k^2$ dei dati si trova entro k deviazioni standard rispetto alla media, con $k > 1$.

In termini più semplici, il teorema di Čebyšëv fornisce una stima sulla percentuale dei dati che si trovano entro un certo numero di deviazioni standard rispetto alla media. Ad esempio, se k è 2, allora almeno il 75% dei dati si troverà entro 2 deviazioni standard dalla media. Il teorema di Čebyšëv è utile perché si applica a qualsiasi distribuzione di dati, indipendentemente dalla sua forma. Questo significa che possiamo ottenere informazioni generali sulla dispersione dei dati anche quando non conosciamo la forma specifica della distribuzione. Tuttavia, è importante notare che la stima fornita dal teorema di Čebyšëv può essere molto conservativa, non ottimale per distribuzioni specifiche.

\checkmark Dimostrazione 10

Partendo dalla disuguaglianza di Markov possiamo ottenere il seguente enunciato:

$$P(|X - \mathbb{E}[X]| \geq k\sigma) \leq \frac{1}{k^2}$$

Con parametro k positivo. Per farlo definiamo una variabile $X : \Omega \rightarrow \mathbb{R}$ e associamo ad essa la variabile aleatoria $Y = (X - \mathbb{E}[X])^2$. Così definita Y è una variabile non negativa, pertanto applichiamo ad essa la disuguaglianza di Markov ottenendo:

$$P(Y \geq k^2\sigma^2) \leq \frac{\mathbb{E}[Y]}{k^2\sigma^2} \iff P((X - \mathbb{E}[X])^2 \geq k^2\sigma^2) \leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{k^2\sigma^2}$$

A destra otteniamo la definizione di varianza $Var = \mathbb{E}[(X - \mathbb{E}[X])^2]$, e sapendo che in generale $\forall x, \mu, a$ vale che:

$$(x - \mu)^2 \geq a^2 \iff |x - \mu| \geq |a|$$

✓ Dimostrazione 10 continued

Otteniamo quanto si voleva dimostrare ovvero:

$$P(|X - \mathbb{E}[X]| \geq k\sigma) \leq \frac{\text{Var}[X]}{k^2 \sigma^2} \quad (5.8)$$

Esempio 5

Da rilevamenti su un grande numero di pozzi risulta che il residuo fisso dell'acqua ha un valore medio di 88.0 mg/l, con una deviazione standard di 25.0 mg/l. Quanto vale la probabilità che un pozzo abbia l'acqua con un residuo fisso maggiore di 180 mg/l?

In questo caso possiamo applicare il teorema di Čebyšëv, non conosciamo la forma della distribuzione, ma ipotizziamo che i dati si distribuiscano secondo una normale con $\mu=88$ e $\sigma=25$. Il testo adesso ci chiede la probabilità di trovare un pozzo che abbia l'acqua con residuo fisso maggiore di 180. Banalmente ci verrebbe in mente di utilizzare la gaussiana standardizzata, trovare la z e calcolare l'area. Questo procedimento è giusto ma non troppo, infatti non abbiamo dati a sufficienza per essere sicuri di essere in regime di approssimazione gaussiano, dunque possiamo applicare quanto appena visto. Una volta calcolato lo z_{score} che risulta essere 3.68 possiamo calcolarci la probabilità $P(X > 180) = 1/z^2$ e dunque otterremo che $P(X > 180) \leq 7.4\%$.

5.5 Random Walk

Una *random walk*, traducibile in italiano come "passo casuale" o "camminata casuale", è un concetto matematico che descrive un percorso costituito da una sequenza di passi casuali. In una random walk unidimensionale, un oggetto o un sistema parte da una posizione specifica e compie passi a destra o a sinistra con probabilità uguali. La direzione di ogni passo viene determinata casualmente. Ad esempio, ad ogni passo potrebbe essere lanciata una moneta e, se esce testa, l'oggetto si sposta di una unità verso destra, mentre se esce croce si sposta di una unità verso sinistra. Il concetto di random walk può essere esteso a dimensioni superiori. In due o tre dimensioni, l'oggetto può muoversi in diverse direzioni (ad esempio su, giù, sinistra, destra, avanti, indietro) con probabilità uguali. Il concetto di random walk può essere applicato per esplorare vari fenomeni come la diffusione, il movimento di particelle. Nel codice seguente in c abbiamo un random walk bidimensionale:

```

1 #include <stdio.h>
2 #include <stdlib.h>
3 #include <time.h>
4 #define NMIN 10
5 #define NMAX 200000
6
7 int inserimento(){
8     int N;
9     printf("Inserisci il numero di passi N ", N);
10    scanf("%d", &N);
11    while (N < NMIN || N > NMAX)
12        {
13        printf("Inserisci il numero di passi N ", N);
14        scanf("%d", &N);
15        }
16    return N;
17 }
18
19 void movimento(int pos[]){
20
21     int m = rand() % 4;
22     if(m==0){
23         pos[0]++;
24     }
25     else if(m==1){
26         pos[1]--;
27     }
28     else if(m==2){

```

```

29     pos[0]--;
30 }
31 else{
32     pos[1]++;
33 }
34 }
35
36 int main(){
37     srand(time(NULL));
38     int pos[2]={0,0};
39     int N= inserimento();
40     int traiettoria[NMAX+1][2]={0,0};
41     for(int i=0; i<N; i++){
42         movimento(pos);
43         traiettoria[i+1][0]=pos[0];
44         traiettoria[i+1][1]=pos[1];
45         printf("%d %d\n", traiettoria[i+1][0], traiettoria[i+1][1] )
46         ;
47     }
48     FILE *fpr;
49     fpr=fopen("walk.dat","w");
50     for(int i=0;i<=N;i++){
51         fprintf(fpr,"%d %d\n",traiettoria[i][0], traiettoria[i][1]);
52     }
53     fclose(fpr);
54 }

```

Codice 5.2: Esempio di Random Walk

Questo codice simula il comportamento di un punto materiale, ne salva la traiettoria su un file txt e in seguito tramite uno script python ne è possibile plottare l' andamento:

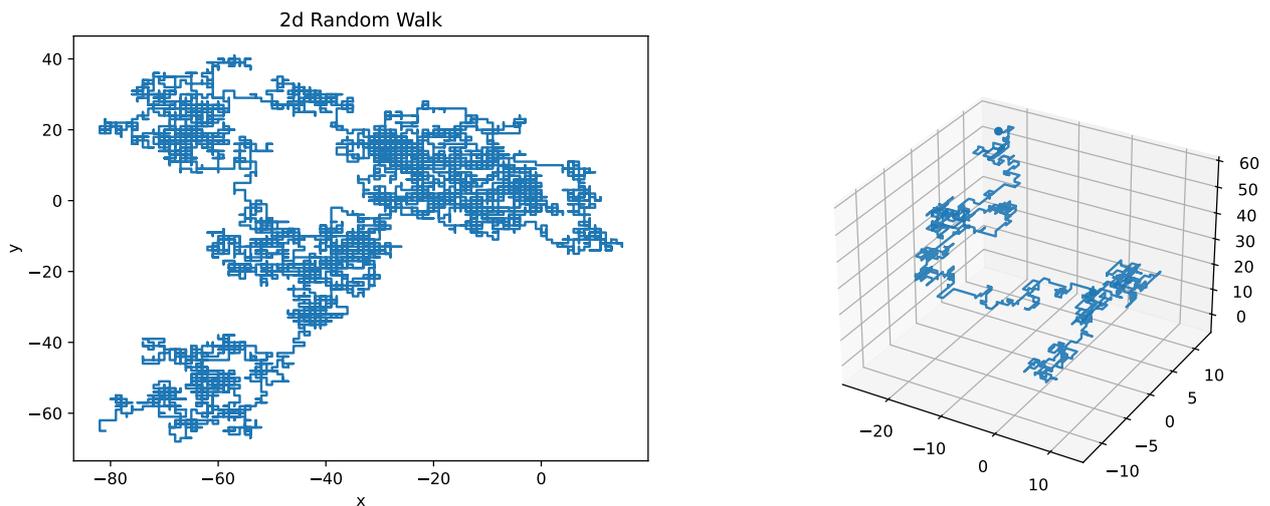


Figura 5.1: Esempio 2d e 3d di random walk per 10'000 punti.

Quello che studieremo noi però è un cammino aleatorio unidimensionale, dunque potremmo muoverci solo avanti o indietro e lo spostamento sarà deciso dal lancio di una moneta. Se dovessimo allora calcolarci il valore atteso otterremo per un n tendente ad infinito il valore 0 poichè la probabilità di muoversi in avanti o indietro è la stessa. Dunque ci interessiamo di calcolare la varianza.

Chiamiamo S la variabile casuale "spostamento in avanti", con X il numero di teste e con Y quello di croci, allora abbiamo che $S = X - Y$ e che $Y = n - X$ e quindi otteniamo:

$$S = 2X - n$$

$$\mathbb{E}[S] = 2\mathbb{E}[X] - n = 0 \quad (5.9)$$

Ora ha senso che con questo ragionamento ci venga che $\mathbb{E}[x] = \frac{n}{2}$ perché abbiamo a che fare con una distribuzione binomiale di parametri $n, 1/2$, motivo per cui la varianza sarà $\sigma_x^2 = npq = n/4$, e quella relativa ai passi in avanti invece:

$$\text{Var}[S] = 4 \text{Var}[X] = 4 \left(n \cdot \frac{1}{2} \cdot \frac{1}{2} \right) \quad (5.10)$$

$$\sigma[S] = \sqrt{n} \quad (5.11)$$

Quindi la deviazione standard cresce come la radice quadrata dei passi compiuti. La funzione di distribuzione di probabilità di S sapendo che X si distribuisce come una binomiale sarà:

$$f(s) = \binom{n}{\frac{s+n}{2}} p^{\frac{n+s}{2}} (1-p)^{\frac{n-s}{2}} \quad (5.12)$$

5.6 Metodo di Monte Carlo

Immaginiamo di voler determinare una quantità θ e di generare n variabili casuali X_1 distribuite con valore atteso $\mathbb{E}[X_1] = \theta$. Facciamo adesso un'altra simulazione e generiamo un'altra variabile casuale X_2 sempre con valore atteso $\mathbb{E}[X_2] = \theta$ e iteriamo il procedimento per k variabili casuali.

Allora come stimatore di θ possiamo prendere la media delle k variabili generate:

$$X = \frac{\sum_{i=1}^k x_k}{N}$$

Poichè $\mathbb{E}[X] = \theta$ ci chiediamo qual è il valore più appropriato di k . Supponiamo dunque di avere n variabili aleatorie indipendenti aventi tutte la stessa distribuzione avremo che:

$$\mathbb{E}[X_i] = \theta$$

$$\text{Var}(X_i) = \sigma^2$$

La media aritmetica viene definita come:

$$X = \frac{\sum_{i=1}^n X_i}{N}$$

Allora avremo che il suo valore atteso sarà:

$$\mathbb{E}[X] = \frac{\sum_{i=1}^n \mathbb{E}[X_i]}{N} = \theta$$

Allora X è un estimatore non distorto e la sua varianza sarà del tipo:

$$\text{Var}[X] = \frac{\sigma^2}{n}$$

Inoltre il metodo di Monte Carlo è utile per il calcolo delle aree generando n punti casuali, infatti si può calcolare una porzione dello spazio di nostro interesse come visto abbondantemente nel corso di laboratorio di calcolo.

6

Test d'ipotesi



Sommario

- 6.1 Introduzione, 128
- 6.2 Test del p-value, 128
- 6.3 Test del χ^2 per un Istogramma, 130
- 6.4 Confronto tra modelli, 131

6.1 Introduzione

I test d'ipotesi, come facilmente intuibile dal nome (e come dovremmo sapere tutti dall'aver seguito almeno una lezione di questo corso) sono strumenti e strategie volti a valutare quale tra le ipotesi (ovvero modelli che rappresentino i dati) sia la migliore, o anche più banalmente se un'ipotesi adottata è adeguata, sulla base della veridicità dei dati.

Per quanto sia marginalmente importante la definizione formale, dividiamo i due approcci con cui eseguiamo i test d'ipotesi:

- **Approccio frequentista:** definisce la probabilità di un'ipotesi come il limite della frequenza di casi favorevoli su possibili, per $cas\ i \rightarrow \infty$ ovviamente.
- **Approccio Bayesiano:** la probabilità è vista come grado di fiducia che un evento si realizzi, e quindi si utilizza il teorema di Bayes e lo scambio di informazione e causalità.

6.2 Test del p-value

Il p-value viene definito in statistica come il livello di significatività osservato e rappresenta la probabilità che il possibile rifiuto dell'ipotesi nulla sia solo dovuto al caso. Generalmente un'ipotesi è un'assunzione che si fa su una popolazione e questa può essere o vera o falsa. I test di ipotesi ci permettono di dire se l'ipotesi è vera o è falsa. Esistono due tipi di ipotesi "Nulla" e "alternativa". Quella nulla è l'affermazione iniziale che non evidenzia nessun effetto significativo e che in genere si cerca di confutare, invece quella alternativa evidenzia effetti particolari. Un esempio può essere ad esempio voler stabilire se una moneta fosse truccata:

Dunque l'ipotesi nulla sarebbe: $P(\text{Testa})=P(\text{Croce})$, mentre l'ipotesi alternativa $P(\text{Testa}) \neq P(\text{Croce})$.

Nel ragionamento statistico ci si chiede: "Cosa succederebbe se ripetessimo tante volte il campionamento?". Si ipotizza dunque che la popolazione da cui è stato estratto il campione abbia una certa caratteristica H_0 e si cerca con i dati di confutarla H_1 .

Il p-value quindi determina se i risultati sono statisticamente significativi. Per convenzione dunque si definisce H_0 l'ipotesi nulla e H_1 l'ipotesi alternativa. Il livello di significatività α definisce l'insieme di valori che appartengono alla regione di rifiuto o di accettazione e generalmente α assume il valore di 0.05. La collocazione di α segue l'ipotesi alternativa:

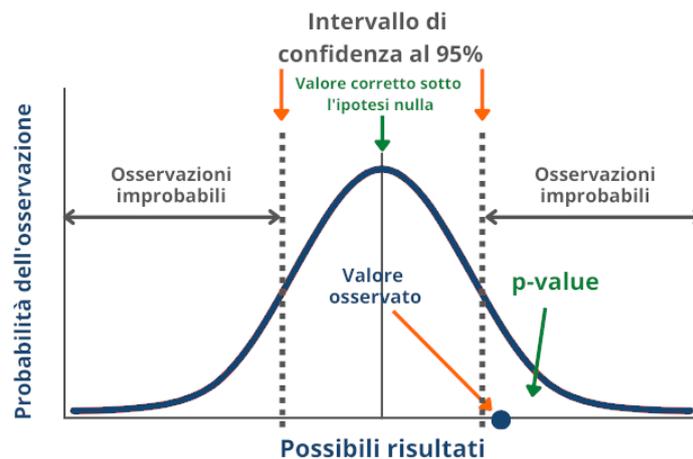
- se nell'ipotesi alternativa troviamo il simbolo \neq posizioneremo metà errore nella coda sinistra e metà in quella destra
- se è presente il $<$ allora α è tutto a sinistra il contrario se abbiamo $>$
- se invece l'ipotesi ricade nell'area $1-\alpha$ non rifiutiamo l'ipotesi nulla.

Dunque il p-value non è altro che la probabilità di osservare un valore θ in una regione di uguale o minore compatibilità con l'ipotesi H :

$$p - value(\theta_i) = \int_{\theta_i}^{\infty} f(\theta|H)d\theta \quad (6.1)$$

il p-value non è altro che la probabilità totale a cui togliamo la cumulativa calcolata in θ_i ovvero:

$$p - value = 1 - F(\theta_i) = 1 - \int_{-\infty}^{\theta_i} f(\theta|H)d\theta \quad (6.2)$$



6.3 Test del χ^2 per un Istogramma

Immaginiamo di confrontare con un modello teorico la probabilità di ottenere delle misure in un certo intervallo, al variare dell'intervallo stesso. Supponiamo di aver fatto N misure e di aver costruito l'istogramma delle occorrenze. Sia n_i il numero di misure che ricadono nell'intervallo Δm_i ed $N = \sum_{i=1}^N b_{in}$ il numero tot delle misure. Se conoscessimo la probabilità che una misura cada in un bin potremmo confrontarla con la frequenza relativa $f_i = \frac{n_i}{N}$ osservata dopo le N misure con probabilità p_i . Allora avremo che:

$$\chi_i^2 = \frac{(n_i - Np_i)^2}{Np_i} \quad (6.3)$$

6.4 Confronto tra modelli

Immaginiamo adesso di avere due modelli possibili che descrivono una serie di dati sperimentali: il primo ha pochissimi parametri, mentre il secondo ha tanti parametri. La differenza sostanziale tra i due è che sicuramente il secondo modello descrive meglio i dati che già ho in possesso, perchè i parametri permettono di essere aggiustati per inserirsi meglio nelle osservazioni, ma è anche vero che il modello più semplice è quello più predittivo, quello che realisticamente valuterà meglio una misurazione futura.

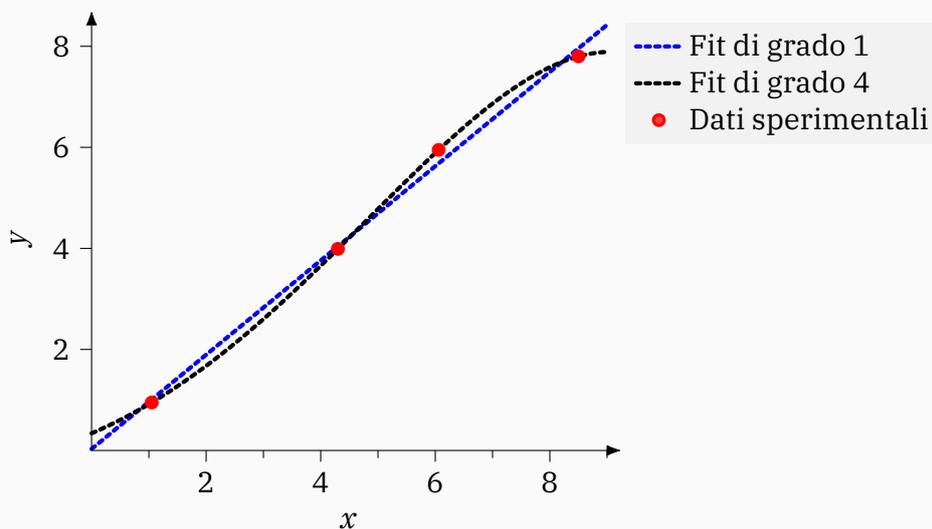
Come fare quindi a scegliere quale modello sia migliore? Il nostro scopo ultimo è l'inferenza predittiva, cioè costruire un modello e usarlo per misure future, ma non è detto che il modello migliore sia il primo, né tantomeno che sia il secondo.

Esempio 6

Facciamo un esempio pratico che fa capire meglio la situazione: immaginiamo di dover fare un fit sulla base di 4 dati. Facciamo due fit e dobbiamo decidere quale sia il migliore: il primo fit è lineare, a un solo parametro, quindi ha più margine di errore sui dati raccontati ma più predittività (soprattutto se mi aspetto una relazione lineare) mentre il secondo è un fit quadratico, precisissimo perchè utilizzando il metodo dei minimi quadrati, ovvero

$$\chi^2 = \sum_{i=0}^n \frac{(y - \mu_i)^2}{\sigma_i^2}$$

Avremo che il fit quadratico è quello che minimizza la differenza, perchè $\chi^2 = 0$.



7 L^AT_EX



Sommario

- 7.1 Introduzione, 134
- 7.2 Comandi principali, 134
- 7.3 Tabelle e Immagini, 137
- 7.4 Scrittura Matematica, 139
- 7.5 Python e Google Colab, 146
- 7.6 Pacchetti utili, 147
- 7.7 Esempio di relazione, 152

7.1 Introduzione

In questo capitolo ci occuperemo di fornire una guida dettagliata e utile della scrittura in L^AT_EX, al fine di scrivere una relazione di laboratorio che rispetti degli standard estetici validi. Con questo non si intende che il contenuto non sia importante o che la forma debba avere la precedenza, tutt'altro, ma siccome è importante anche sapere esprimere i propri risultati, le proprie osservazioni, e scrivere i passaggi logici in maniera coerente e comprensibile, è anche bene occuparsi di questo argomento, a maggior ragione perchè spesso viene preso sottogamba e visto come qualcosa di relativamente inutile. Detto questo, ricordiamo sempre che al fine del corso le relazioni possono essere scritte in genere anche a mano o su Google Colab, ma noi siamo convinti dell'importanza dell'uso di L^AT_EX nell'ambito scientifico, ed ecco perchè consigliamo di dare una letta a quanto segue.

Iniziamo quindi con i comandi più semplici e basilari, in modo che chiunque possa affacciarsi al mondo della scrittura matematica. Possiamo scrivere documenti in L^AT_EX con software come Tex, Visual Studio Code o altri, ma il più semplice è sicuramente <https://www.overleaf.com>, perchè permette di salvare documenti su un cloud e di lavorarci su qualsiasi dispositivo che abbia una connessione internet.

7.2 Comandi principali

La struttura più semplice di un documento L^AT_EX è fatta in questo modo:

```

1 \usepackage{graphicx} %Per l'immissione di immagini
2
3 \title{...}
4 \author{...}
5 \date{June 2023}
6
7 \begin{document}
8
9 \maketitle
10
11 \section{Introduction}
12 Testo qui.
13 \end{document}

```

Codice 7.1: Codice base per un documento in linguaggio TeX

È chiaro quindi che un documento di questo tipo si compone di due parti principali:

- Preambolo: ovvero tutto ciò che viene prima del comando `\begin{document}`, che comprende principalmente i vari pacchetti da usare e che danno la possibilità di usare comandi aggiuntivi, layout personalizzati, e principalmente ogni funzione grafica, matematica e altro che non faccia parte della scrittura a macchina standard di parole e numeri (come grafici, formule...);
- Corpo: ovvero ciò che si vede nel pdf. In questo si pone tutto ciò che si scrive: capitoli, paragrafi...

Per quanto riguarda il lato della scrittura normale c'è relativamente poco da dire: si scrive normalmente come si farebbe con Word o con altri programmi, le uniche differenze sono che per andare a capo si usa il comando `\par` oppure semplicemente `\\`, ovviamente ponendone quattro `\\\\` si va a capo due volte, cioè si lascia uno spazio di una riga per iniziare un nuovo paragrafo. Ci sono poi comandi che permettono di iniziare liste numerate e non, ovvero rispettivamente:

```

1 \begin{enumerate} %Numerizza i vari item
2   \item ...
3   \item ...
4 \end{enumerate}
5
6 \begin{itemize} %Inizia una bullet list
7   \item ...
8   \item ...
9 \end{itemize}

```

Codice 7.2: Codice per inserimento liste

In particolare il comando `itemize` stampa una lista il cui elenco puntato standard è **•**, ma si può cambiare questa scelta scrivendo l'item in questo modo: `\item[%elenco puntato desiderato]`.

Per utilizzare le modalità font *italic*, **bold**, code, sottolineato e altri, esistono i comandi appositi, che sono rispettivamente:

- `\textit{...}` per il testo *italic*;
- `\textbf{...}` per il testo **bold**;
- `\texttt{...}` per il testo code;
- `\underline{...}` per il testo sottolineato;

Per quanto riguarda la ripartizione delle sezioni del documento, possiamo dividere il tutto in capitoli, sezioni, sottosezioni, sottosottosezioni... ognuno con il comando corrispondente (in inglese), quindi avremo per esempio `\section`, `\subsection`... e possiamo far apparire l'indice con il comando `\tableofcontents`, che ovviamente non ha senso porre dopo le sezioni ma andrà posto subito all'inizio del documento (poi se uno vuole può mettere l'indice anche alla fine, la vita è fatta di scelte). Infine, ecco una lista di tutti i principali comandi che servono per scrivere un testo di base in \LaTeX :

<i>Comando</i>	<i>Risultato</i>
<code>\usepackage{...}</code>	Permette di usare i pacchetti
<code>\begin{document}</code>	Inizia il documento
<code>\end{document}</code>	Termina il documento
<code>\section{...}</code>	Inizia una sezione
<code>\subsection{...}</code>	Inizia una sottosezione
<code>\subsubsection{...}</code>	Inizia una sottosottosezione
<code>\newpage</code>	Inizia una nuova pagina
<code>\vspace{...cm}</code>	Lascia dello spazio
<code>\begin{itemize}</code>	Inizia una bullet list
<code>\end{itemize}</code>	Finisce una bullet list
<code>\begin{enumerate}</code>	Inizia una lista numerata
<code>\end{enumerate}</code>	Finisce una lista numerata
<code>\color{...}</code>	Definisce il colore del testo

Tabella 7.1: Comandi principali

7.3 Tabelle e Immagini

Occupiamoci adesso dell'inserimento nel documento di tabelle e immagini, utili per riportare dati e grafici delle nostre misurazioni.

Parliamo come prima cosa delle tabelle: queste vengono aggiunte tramite i comandi `\begin{table}` e `\end{table}`, in mezzo a cui vanno però una serie di opzioni da configurare. Prima tra tutte è il posizionamento della tabella stessa: se scriviamo i comandi senza specificare il posizionamento automatico, questa ci verrà posta in alto all'inizio della pagina successiva. Per evitare ciò, basta scrivere `\begin{table}[h!]` e poi normalmente `\end{table}` e il problema è risolto. Un'altra opzione che possiamo scegliere ma che non viene automaticamente mostrata è la larghezza delle caselle (di default chiaramente 1), che possiamo aumentare o diminuire tramite `\def\arraystretch{...}`. In mezzo al `begin` e `end` va poi messo un altro comando simile di questo tipo, composto da `\begin{tabular}` e `\end{tabular}`, che è dove va effettivamente inserito il testo che vogliamo nella tabella. Prima di ciò però dobbiamo specificare il numero e il tipo di cella di cui sarà composta la nostra tabella. I tipi di cella sono:

- l: una colonna di oggetti allineati a sinistra;
- r: una colonna di oggetti allineati a destra;
- c: una colonna di oggetti allineati al centro.

E inseriamo tante di queste lettere per quante colonne vogliamo (anche di tipo diverso), mettendo inoltre una sbarra "|" tra le lettere se vogliamo una linea che separi le colonne. Adesso siamo pronti per inserire il contenuto: questo si fa scrivendo una linea alla volta con ogni contenuto di cella separato dal carattere "&" e finendo la linea con il comando `\\`. Infine, con il comando `\hline` produciamo una linea orizzontale che separa le righe. Facciamo un esempio di codice di tabella e vediamo il risultato estetico:

```

1 \begin{table}[h!]
2   \centering
3   \def\arraystretch{1.4}
4   \begin{tabular}{|c||c|l|}
5     \hline
6     cella 1.1 & cella 1.2 & cella 1.3 \\
7     \hline
8     cella 2.1 & cella 2.2 & cella 2.3 \\
9     \hline
10    cella 3.1 & cella 3.2 & cella 3.3 \\
11    \hline
12    \end{tabular}
13    \caption{Esempio di tabella}
14 \end{table}

```

Codice 7.3: Codice base per inserimento tabelle

cella 1.1	cella 1.2	cella 1.3
cella 2.1	cella 2.2	cella 2.3
cella 3.1	cella 3.2	cella 3.3

Tabella 7.2: Esempio di tabella

Passando alle immagini, queste hanno una struttura di immissione simile: il tutto è contenuto entro i comandi `\begin{figure}` e `\end{figure}`, entro cui vanno una serie di impostazioni. Similmente a prima, se vogliamo posizionare l'immagine sotto il resto del codice bisogna inserire `[h!]`, altrimenti verrà posto in cima su una nuova pagina. Dobbiamo poi scegliere il file da inserire come immagine tramite `\includegraphics[scale=...]{nomefile.pdf}` in cui la scala serve a diminuire la grandezza mantenendo le proporzioni. È chiaro che, dal momento in cui serve un file da inserire, questo si deve trovare tra i vari file del documento (come il `main.tex` o altri che possono essere aggiunti), e su Overleaf si può inserire un file tramite il tasto upload come mostrato in [Figure 7.1](#). Vediamo il codice tipo di un'immissione immagine e il risultato (che è proprio [Figure 7.1](#)):

```

1 \begin{figure}[h!]
2   \centering
3   \includegraphics{Immagini/Screenshot.png}
4   \caption{Metodo per il caricamento di un file su Overleaf}
5   \label{fig:upload}
6 \end{figure}

```

Codice 7.4: Codice base per inserimento immagini

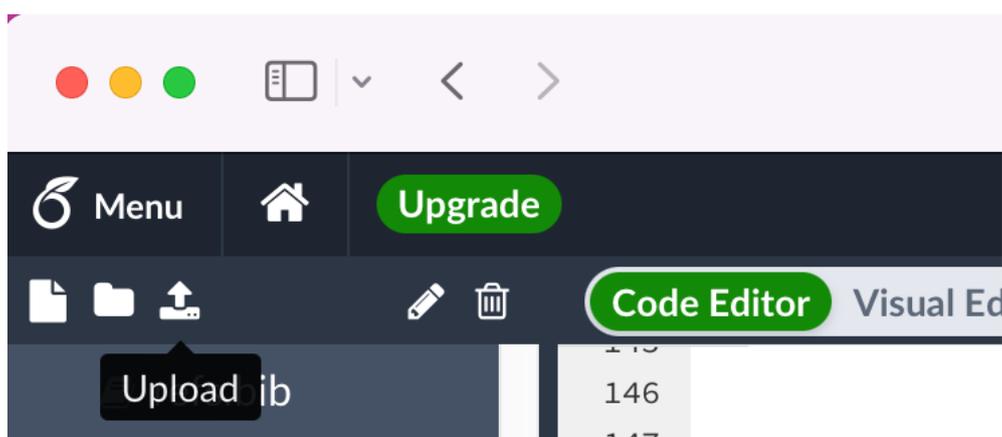


Figura 7.1: Metodo per il caricamento di un file su Overleaf

7.4 Scrittura Matematica

La parte più importante della scrittura in L^AT_EX è probabilmente la `mathmode`, poiché può sembrare a prima vista complicata, ma come tutto è questione di abitudine, e di conoscere i comandi più importanti. Per inserire un'equazione nel nostro documento, dobbiamo innanzitutto usare il pacchetto relativo che è `\usepackage{amsmath}`, e possiamo dunque inserire le nostre formule tra due dollari per avere la modalità `inline` (equazione nella riga) oppure tra quattro, ovvero due dollari seguiti dal nostro codice e seguiti da altri due dollari, per la modalità `display` ovvero al centro a capo con una piccola spaziatura.

`$`modalità inline`$`

`$$`modalità display`$$`

Per quanto riguarda la scrittura matematica in sè, essa è abbastanza intuitiva. I simboli principali (numeri, segno + e -, lettere) si digitano direttamente dalla tastiera, mentre altri hanno un comando che spesso è molto facile da ricordare o addirittura da indovinare, come `\int` per l'integrale o `\sum` per la sommatoria. Facciamo un'esempio di scrittura matematica con qualche simbolo per capire come prendere dimestichezza e vediamo il risultato (nella prossima sottosezione metteremo un elenco dei comandi più importanti):

```
1 Modalita' inline: $43^{12} \approx 4 \cdot 10^{19}$.
2 Modalita' display: $$\int_0^{+\infty} \frac{e^x}{4\pi x^x} dx =
  0.5093...$$
```

Codice 7.5: Esempio di scrittura matematica

Il risultato sono le seguenti righe:

Modalità inline: $43^{12} \approx 4 \cdot 10^{19}$. Modalità display:

$$\int_0^{+\infty} \frac{e^x}{4\pi x^x} dx = 0.5093\dots$$

7.4.1 Lista dei comandi utili

Ecco una lista dei comandi principali per l'uso della mathmode in \LaTeX .

1. Simboli Greci

α <code>\alpha</code>	κ <code>\kappa</code>	ψ <code>\psi</code>	\digamma <code>\digamma</code>	Δ <code>\Delta</code>	Θ <code>\Theta</code>
β <code>\beta</code>	λ <code>\lambda</code>	ρ <code>\rho</code>	\varkappa <code>\varkappa</code>	Γ <code>\Gamma</code>	Υ <code>\Upsilon</code>
χ <code>\chi</code>	μ <code>\mu</code>	σ <code>\sigma</code>	φ <code>\varphi</code>	Λ <code>\Lambda</code>	Ξ <code>\Xi</code>
δ <code>\delta</code>	ν <code>\nu</code>	τ <code>\tau</code>	\wp <code>\wp</code>	Ω <code>\Omega</code>	
ε <code>\varepsilon</code>	\circ <code>o</code>	θ <code>\theta</code>	ρ <code>\varrho</code>	Φ <code>\Phi</code>	
η <code>\eta</code>	ω <code>\omega</code>	υ <code>\upsilon</code>	ς <code>\varsigma</code>	Π <code>\Pi</code>	
γ <code>\gamma</code>	ϕ <code>\phi</code>	ξ <code>\xi</code>	ϑ <code>\vartheta</code>	Ψ <code>\Psi</code>	
ι <code>\iota</code>	π <code>\pi</code>	ζ <code>\zeta</code>	ε <code>\varepsilon</code>	Σ <code>\Sigma</code>	

2 Delimitatori e costrutti

$\frac{abc}{xyz}$ <code>\frac{abc}{xyz}</code>	\overline{abc} <code>\overline{abc}</code>	\overline{abc} <code>\overline{abc}</code>
\overrightarrow{abc} <code>\overrightarrow{abc}</code>	\overleftarrow{abc} <code>\overleftarrow{abc}</code>	\underline{abc} <code>\underline{abc}</code>
\sqrt{abc} <code>\sqrt{abc}</code>	$\sqrt[n]{abc}$ <code>\sqrt[n]{abc}</code>	\widehat{abc} <code>\widehat{abc}</code>
\hat{a} <code>\hat{a}</code>	\widetilde{abc} <code>\widetilde{abc}</code>	\tilde{a} <code>\tilde{a}</code>
\overbrace{abc}^{xyz} <code>\overbrace{abc}^{xyz}</code>	\underbrace{abc}_{xyz} <code>\underbrace{abc}_{xyz}</code>	$ $ <code> </code>
$ $ <code>\lvert</code>	$ $ <code>\lvert</code>	$ $ <code>\Vert</code>
$\{$ <code>\{</code>	$\}$ <code>\}</code>	\langle <code>\langle</code>
\rangle <code>\rangle</code>	\lfloor <code>\lfloor</code>	\rfloor <code>\rfloor</code>
\lceil <code>\lceil</code>	\rceil <code>\rceil</code>	\backslash <code>\backslash</code>

Nota bene: I delimitatori e le parentesi spesso vengono usati con espressioni molto grandi, ma questi non si adattano alla grandezza del testo a meno che non si usino le opzioni di dimensione testo (vedi [Sottosottosezione 7.4.1](#) oppure i comandi `\left...` e `\right...`). Facciamo un esempio:

```
1 $$\left( \sum_{i=0}^n \frac{1}{2^n} \right) \cdot \pi = \pi$$
```

Il risultato è:

$$\left(\sum_{i=0}^n \frac{1}{2^n} \right) \cdot \pi = \pi$$

3. Simboli di grandezza variabile

$$\begin{array}{cccc} \sum_{i=0}^n \backslash \text{sum}_{\{i=0\}^n} & \int_0^1 \backslash \text{int}_{\theta^1} & \prod_{i=0}^n \backslash \text{prod}_{\{i=0\}^n} & \oint_0^1 \backslash \text{oint}_{\theta^1} \\ \iint_0^1 \backslash \text{iint}_{\theta^1} & \bigcap \backslash \text{bigcap} & \bigcup \backslash \text{bigcup} & \biguplus \backslash \text{biguplus} \end{array}$$

4. Funzioni

$$\begin{array}{cccc} \sin \backslash \text{sin} & \cos \backslash \text{cos} & \tan \backslash \text{tan} & \arcsin \backslash \text{arcsin} \\ \exp \backslash \text{exp} & \ln \backslash \text{ln} & \log_{10} \backslash \text{log}_{\{10\}} & \lim_{x \rightarrow 0} \backslash \text{lim}_{\{x \rightarrow 0\}} \end{array}$$

5. Operatori

$$\begin{array}{cccccc} \cdot \backslash \text{cdot} & \times \backslash \text{times} & \circ \backslash \text{circ} & \div \backslash \text{div} & \pm \backslash \text{pm} & \mp \backslash \text{mp} \\ \perp \backslash \text{bot} & \parallel \backslash \text{parallel} & \propto \backslash \text{propto} & \equiv \backslash \text{equiv} & \neq \backslash \text{neq} & \cong \backslash \text{cong} \\ \approx \backslash \text{approx} & \sim \backslash \text{sim} & \simeq \backslash \text{simeq} & \leq \backslash \text{leq} & \geq \backslash \text{geq} & \subset \backslash \text{subset} \\ \subseteq \backslash \text{subsetq} & \in \backslash \text{in} & \gg \backslash \text{ll} & \supset \backslash \text{supset} & \supseteq \backslash \text{supseteq} & \ni \backslash \text{ni} \end{array}$$

6. Freccie

\leftarrow <code>\leftarrow</code>	\rightarrow <code>\rightarrow</code>	\Leftarrow <code>\Leftarrow</code>
\Rightarrow <code>\Rightarrow</code>	\leftrightarrow <code>\leftrightarrow</code>	\Leftrightarrow <code>\Leftrightarrow</code>
\implies <code>\implies</code>	\longleftarrow <code>\longleftarrow</code>	\Leftrightarrow <code>\Leftrightarrow</code>
\Longrightarrow <code>\Longrightarrow</code>	\uparrow <code>\uparrow</code>	\downarrow <code>\downarrow</code>
\Uparrow <code>\Uparrow</code>	\Downarrow <code>\Downarrow</code>	\Updownarrow <code>\Updownarrow</code>
\nearrow <code>\nearrow</code>	\searrow <code>\searrow</code>	\swarrow <code>\swarrow</code>
\nwarrow <code>\nwarrow</code>	\leftrightarrow <code>\leftrightarrow</code>	\nrightarrow <code>\nrightarrow</code>

7. Simboli vari

∞ <code>\infty</code>	∇ <code>\nabla</code>	∂ <code>\partial</code>	\forall <code>\forall</code>	\wp <code>\wp</code>	\exists <code>\exists</code>
\nexists <code>\nexists</code>	\emptyset <code>\emptyset</code>	\cdots <code>\cdots</code>	\vdots <code>\vdots</code>	\ldots <code>\ldots</code>	\ddots <code>\ddots</code>
\Im <code>\Im</code>	\Re <code>\Re</code>	\hbar <code>\hbar</code>	\imath <code>\imath</code>	\jmath <code>\jmath</code>	

8. Font e grandezze

Oltre ai font discussi in [Sezione 7.2](#), ce ne sono altri utilizzati nell'ambiente matematico (per esempio quello per indicare gli insiemi dei numeri, o quello usato per alcuni operatori come il Laplaciano...). Questi sono:

- `\mathcal{...}`: *ABCDEFGHIJKLMNOPQRSTUVWXYZ*;
- `\mathbb{...}`: **ABCDEFGHIJKLMNOPQRSTUVWXYZ**;
- `\mathfrak{...}`: *ABCDEFGHIJKLMNOPQRSTUVWXYZ*;
- `\mathsf{...}`: **ABCDEFGHIJKLMNOPQRSTUVWXYZ**;

Per quanto riguarda le varie grandezze dei font, dobbiamo distinguere tra la `mathmode` e la `textmode`:

Mathmode:	$\int x dx$ $\int x dx$ $\int x dx$ $\int x dx$	<pre>\displaystyle\int xdx \textstyle\int xdx \scriptstyle\int xdx \scriptscriptstyle\int xdx</pre>
-----------	---	---

Textmode:	<pre>\tiny=tiny \scriptsize=scriptsize \footnotesize=footnotesize \small=small \normalsize=normalsize</pre>	<pre>\large=large \Large=Large \LARGE=LARGE \huge=huge \Huge=Huge</pre>
-----------	---	---

9. Ambienti e layout

Gli ambienti sono ulteriori metodi in cui poter utilizzare la scrittura matematica, come ad esempio vettori, matrici, equazioni numerate, ambienti di allineamento del testo... Vediamo i più importanti quali sono:

Per quanto riguarda l'allineamento del testo, l'ambiente in questione è quello `\aling` che si compone di un `\begin{...}` e di un `\end{...}`, andiamo a capo con il solito comando `\\` e allineamo in base al carattere o allo spazio prescelto con `&`. Il tutto senza i classici dollari, il che vale per tutti gli ambienti. Facciamo un esempio:

```
1 \begin{align}
2   \int \left(4x + 7x^2 \right) dx &= \int 4xdx + \int 7x^2dx \\
3   &= 4\int xdx + 7\int x^2dx \\
4   &= 4 + 7x + c
5 \end{align}
```

Il risultato è il seguente:

$$\int (4x + 7x^2) dx = \int 4x dx + \int 7x^2 dx \quad (7.1)$$

$$= 4 \int x dx + 7 \int x^2 dx \quad (7.2)$$

$$= 4 + 7x + c \quad (7.3)$$

Se vogliamo inoltre che le equazioni non siano numerate, basta scrivere l'ambiente come `\begin{align*}` e `\end{align*}`. Possiamo poi decidere lo spazio da lasciare tra una riga e l'altra (anche qui, il tutto vale per OGNI ambiente) con `\\[... pt]`. Simile a questo, l'ambiente `\gather` permette semplicemente di andare a capo senza usare più linee di codice in cui inserire i dollari, tramite il solito `\\` oppure `\\[... pt]`.

Passando all'ambiente equazione, esso ci dà semplicemente la possibilità di numerare un'equazione:

```
1 \begin{equation}
2   i\hbar\{\partial \psi \over \partial t\}\left(r,t\right)=\hat{H}\psi\left(r,t\right)
3 \end{equation}
```

Il risultato è il seguente:

$$i\hbar \frac{\partial \psi}{\partial t}(r, t) = \hat{H} \psi(r, t) \quad (7.4)$$

Ci sono poi vettori e matrici che si scrivono tra un `\begin{matrix}` e un `\end{matrix}`. Ecco i tipi principali di matrici:

```
a \begin{matrix} a \\ b \end{matrix}
```

```
\begin{pmatrix} a \\ b \end{pmatrix}
```

```
\begin{pmatrix} a & b \\ c & d \end{pmatrix}
```

```
\begin{bmatrix} a & b \\ c & d \end{bmatrix}
```

```
\begin{vmatrix} a & b \\ c & d \end{vmatrix}
```

L'ultimo ambiente è quello delle parentesi graffe (usato per esempio nella descrizione di una funzione con condizioni o nello studio delle equazioni cardinali), che si usa tra un `\begin{cases}` e un `\end{cases}`.

7.4.2 Creare nuovi comandi

Apriamo una piccola parentesi non del tutto necessaria ma utile. Tra le tante (tantissime) cose che L^AT_EX ci permette di fare, c'è anche la possibilità di creare comandi per facilitare l'immissione di alcune righe lunghe che altrimenti richiederebbero più tempo per essere scritte. Questo funziona tramite il comando `\newcommand`. Noi stessi abbiamo usato questa funzionalità nel pdf per aiutarci nello scrivere conti con integrali e sommatorie, vediamo in che modo lo abbiamo applicato:

```
1 \newcommand{\intf}{\int_{-\infty}^{+\infty}}
```

In questo modo, per scrivere un'integrale tra meno e più infinito ci basterà digitare `\intf`. Facciamo un esempio di codice di un calcolo relativamente complesso (dal punto di vista della scrittura) che abbiamo usato in questo documento, per vedere come possiamo utilizzare tutte queste funzioni per avere una scrittura di equazioni e codici pulita e ordinata (il calcolo è quello della varianza per la binomiale che si trova in appendice, ma ne riportiamo solo una parte tagliando il testo perchè l'importante adesso è il lato estetico):

```
1 [...]
2   \begin{gather*}
3   \EE[x^2]=\sum_{x=0}^n x^2\cdot \frac{n!}{x!(n-x)!}\cdot p^x\cdot
q^{n-x} = \sum_{x=0}^n \{x^2\cancel{2}^1\}\cdot n\cdot (n-1)\cdot
\cancel{x}\cdot (x-1)!\cdot (n-x)!\cdot p^{x\color{blue}+1-1}\cdot
q^{n-x} = \backslash
4   =np\cdot \sum_{x=1}^n x\cdot \frac{(n-1)!}{(x-1)!(n-x)!}\cdot p^{x-1}\cdot
q^{n-x}
5   \end{gather*}
6 [...]
7   \begin{gather*}
8     \begin{cases}
9       n-1=t\backslash x-1=y
10      \end{cases}\Longrightarrow (t+1)p\cdot \sum_{y=0}^{t+1} \{\color{blue}(y+1)\}\cdot t!\cdot
\over y!(t-y)!\cdot p^y\cdot q^{t-y}
11   \end{gather*}
12 [...]
13   $$\EE[x^2]=(t+1)p\cdot \left[ \sum_{y=0}^{t+1} y\cdot t!\cdot
\over y!(t-y)!\cdot p^y\cdot q^{t-y} + \{\color{blue}\underbrace{\color{black}\sum_{y=0}^{t+1}\{t!\cdot
\over y!(t-y)!\cdot p^y\cdot q^{t-y}}\}_{y=1}}\right]$$
14 [...]
15   $$\EE[x^2]=(t+1)p\cdot (t+1)\Longrightarrow np\cdot \left[ (n-1)p+1\right]$$
16 [...]
```

```

17  $$\sigma^2[x]=\mathbb{E}[x^2]-\mathbb{E}^2[x]=np\cdot\left[(n-1)p+1\right]-
    n^2p^2=np\cdot\left(\cancel{np}-p+1-\cancel{np}\right)=np(1-p)
    )$$
18 [...]
19 \begin{equation}
20 \quad \sigma[x]^2=npq
21 \end{equation}

```

E il risultato è il seguente:

[...]

$$\begin{aligned} \mathbb{E}[x^2] &= \sum_{x=0}^n x^2 \cdot \frac{n!}{x!(n-x)!} \cdot p^x \cdot q^{n-x} = \sum_{x=0}^n \frac{x^2 \cdot n \cdot (n-1)!}{x \cdot (x-1)! \cdot (n-x)!} \cdot p^{x+1-1} \cdot q^{n-x} = \\ &= np \cdot \sum_{x=1}^n x \cdot \frac{(n-1)!}{(x-1)!(n-x)!} \cdot p^{x-1} \cdot q^{n-x} \end{aligned}$$

[...]

$$\begin{cases} n-1=t \\ x-1=y \end{cases} \implies (t+1)p \cdot \sum_{y=0}^{t+1} (y+1) \cdot \frac{t!}{y!(t-y)!} \cdot p^y \cdot q^{t-y}$$

[...]

$$\mathbb{E}[x^2] = (t+1)p \cdot \left[\sum_{y=0}^{t+1} y \cdot \frac{t!}{y!(t-y)!} \cdot p^y \cdot q^{t-y} + \underbrace{\sum_{y=0}^{t+1} \frac{t!}{y!(t-y)!} \cdot p^y \cdot q^{t-y}}_{=1} \right]$$

[...]

$$\mathbb{E}[x^2] = (t+1)p \cdot (tp+1) \implies np \cdot [(n-1)p+1]$$

[...]

$$\sigma^2[x] = \mathbb{E}[x^2] - \mathbb{E}^2[x] = np \cdot [(n-1)p+1] - n^2p^2 = np \cdot (\cancel{np} - p + 1 - \cancel{np}) = np(1-p)$$

[...]

$$\sigma[x]^2 = npq \tag{7.5}$$

7.5 Python e Google Colab

Anche qui apriamo una parentesi, nel senso che non trattiamo l'argomento Python in questo testo poichè viene affrontato largamente a lezione e sono già presenti molti notebook per imparare le basi necessarie ad affrontare il corso. Noi ci limitiamo a mettere a disposizione un notebook di nostra creazione con alcuni esempi (avanzati e non): [qui](#).

7.6 Pacchetti utili

Vediamo adesso una serie di pacchetti utili (ma anche qui, non necessari) per la scrittura delle relazioni.

7.6.1 Lstlisting: inserire codici nei documenti

Il pacchetto `\lstlisting` permette di inserire codici all'interno dei documenti. Ovviamente per usarlo va inserito nel preambolo `\usepackage{listing}`. Esso permette di inserire codici in un rettangolo o in modalità inline (come avete visto fin ora). Il setup del pacchetto è molto complesso, quindi lasciamo di seguito il nostro con cui potete smanettare:

```

1 \usepackage{listings}
2 \lstdefinestyle{mystyle}{
3   backgroundcolor=\color{backcolour},
4   commentstyle=\color{codegreen},
5   keywordstyle=\color{magenta},
6   numberstyle=\small\color{codegray},
7   stringstyle=\color{codepurple},
8   basicstyle=\ttfamily,
9   breakatwhitespace=false,
10  breaklines=true,
11  captionpos=b,
12  keepspaces=true,
13  numbers=left,
14  numbersep=5pt,
15  frame=lines,
16  showspace=false,
17  showstringspaces=false,
18  showtabs=false,
19  tabsize=3,
20  %framexleftmargin=5mm
21 }
22 \lstset{style=mystyle,
23 texcl,
24 }
```

In cui i vari colori sono stati definiti prima tramite:

```
1 \definecolor{codegreen}{rgb}{0,0.6,0}
2 \definecolor{codegray}{rgb}{0.5,0.5,0.5}
3 \definecolor{codepurple}{rgb}{0.58,0,0.82}
4 \definecolor{backcolour}{rgb}{0.95,0.95,0.92}
```

Grazie a questo, siamo in grado di inserire il nostro codice in modo molto semplice, specificando la didascalia e il linguaggio usato:

```
1 \begin{lstlisting}[caption={...}, language{...}]
2     codice...
3 \end{lstlisting}
```

La scrittura inline invece si effettua tramite `\lstinline [language=Tex]` inserendo il codice che vogliamo all'interno delle barre.

7.6.2 hyperref: fare riferimenti

Il pacchetto `hyperref` permette di riferirsi a sezioni, equazioni, grafici o altro all'interno del testo. Per farlo, dopo averlo inserito nel preambolo, bisogna prima di tutto sapere a cosa ci stiamo riferendo: per farlo, inseriamo sotto all'oggetto interessato il comando `\label{...}` e lo denominiamo come vogliamo. Dopo di ciò, con il comando `\autoref{...}` richiamiamo a quanto segnato prima in modo "esteso", nel senso che se ci stiamo riferendo all'*Equazione 1.1* il link si chiamerà proprio *Equazione 1.1*, mentre con il comando `\ref{...}` ci apparirà solo *1.1*.

Per fare un'esempio, riprendiamo l'equazione scritta prima ma sta volta inseriamo il label:

```
1 \begin{equation}
2     i\hbar{\partial \psi \over \partial t}\left(r,t\right)=\hat{H}\psi\left(r,t\right)
3     \label{Schrodinger}
4 \end{equation}
```

Adesso, tramite il comando `\autoref{Schrodinger}` quello che apparirà su schermo è [Equazione 7.4](#). Possiamo inoltre nascondere il bordo rosso che di default appare sui link del pacchetto tramite `\usepackage[hydelinks]{hyperref}`

7.6.3 babel: cambiare la lingua degli input

Di per sè gli input sui comandi nei documenti L^AT_EX sono in inglese, nel senso che ciò che dovrebbe apparire con l'uso di `hyperref` è *Equation 1.1* e non *Equazione 1.1*. Per ovviare a questo problema, usiamo il pacchetto `babel` che traduce molti degli input (ma non tutti). Nel caso in cui qualcosa dovessere ancora essere in inglese basterà aggiungerlo come abbiamo fatto noi di seguito con alcuni riferimenti:

```

1 \usepackage[italian]{babel}
2 \addto\extrasitalian{%
3   \renewcommand{\chapterautorefname}{Capitolo}%
4   \renewcommand{\sectionautorefname}{Sezione}%
5   \renewcommand{\subsectionautorefname}{Sottosezione}%
6   \renewcommand{\subsubsectionautorefname}{Sottosottosezione}%
7   \renewcommand{\equationautorefname}{Equazione}%
8 }
```

7.6.4 tikz: disegnare immagini

Uno dei pacchetti più difficili da utilizzare (e che quindi vi sconsigliamo di usare come motivo per perdere la vostra sanità mentale) è quello `tikz`. Questo permette di disegnare grafici come ad esempio un piano inclinato, un diagramma delle forze e tantissime altre cose. Il problema di questo pacchetto è la grandezza che risiede nella sua complessità e la difficoltà nell'usarlo.

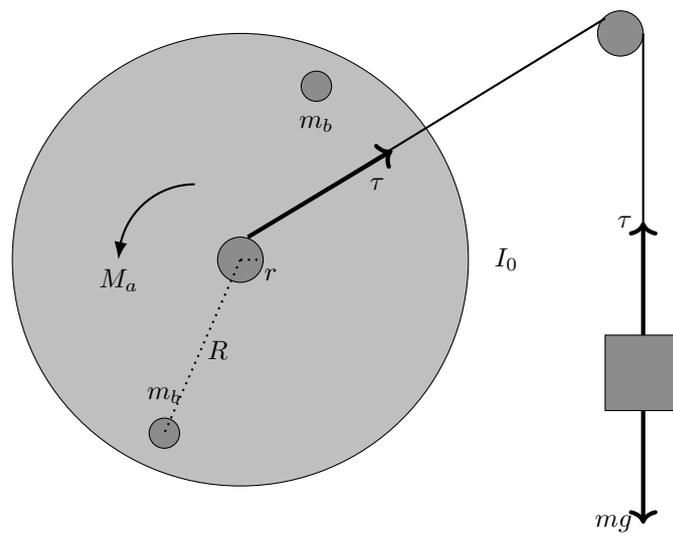
Nel caso in cui voleste tentare l'approccio con questo pacchetto, il consiglio è sempre quello di cercare su internet dei grafici simili a quelli che volete creare e smanettarci un po' su (su StackOverflow o altri siti si trova di tutto). Di seguito mettiamo un esempio di uso del pacchetto, così che voi possiate valutare se usarlo o meno (la prima parte è il setup del pacchetto nel preambolo):

```

1 \usepackage{tikz}
2 \usetikzlibrary{decorations.pathmorphing,patterns}
3 \usetikzlibrary{arrows}
4 \usepackage[utf8]{inputenc}
5 \usepackage{pgfplots}
6 \pgfplotsset{compat=1.15}
7 \usetikzlibrary{positioning}
8 \usetikzlibrary{shapes}
9 \usetikzlibrary{backgrounds, fit}
10 \usetikzlibrary{arrows.meta,bending}
11 \def\leftset{(0,0) circle(2cm)}
12 \def\rightset{(0:2.5cm) circle(2cm)}
13
14
15 \begin{tikzpicture}[auto,node distance=3cm]
16   \begin{scope};
17     \draw [fill=gray!50](0,0) circle (3) node at (3.5,0) {\$I
18     _0\$};
19     \draw [fill=gray!90](0,0) circle (0.3);
20     \draw [fill=gray!90](5,3) circle (0.3);
21     \draw [fill=gray!90](-1,-2.3) circle (0.2) node at (-1,-1.8)
22     {\$m_b\$};
23     \draw [fill=gray!90](1,2.3) circle (0.2) node at (1,1.8) {\$m
24     _b\$};
25     \draw [thick] (0.1,0.3) -- (4.8,3.2);
26     \draw [thick] (5.3, 3) -- (5.3,-1);
27     \draw [fill=gray!90](5.8,-1) rectangle (4.8,-2);
28     \draw [ultra thick, ->] (5.3, -1) -- (5.3, 0.5) node [anchor
29     =east] {\$\tau\$};
30     \draw [ultra thick, ->] (0.1, 0.3) -- (2, 1.45) node at
31     (1.8, 1) {\$\tau\$};
32     \draw [ultra thick, ->] (5.3, -2) -- (5.3, -3.5) node [
33     anchor=east] {\$mg\$};
34     \draw [thick, dotted] (0,0) -- (-1,-2.3) node at (-0.3,
35     -1.2) {\$R\$};
36     \draw [thick, dotted] (0,0) -- (0.3,0) node at (0.4, -0.2)
37     {\$r\$};
38     \draw[-{\Latex[bend], ->},thick] (-0.6,1) arc[start angle=90,
39     end angle=180,radius=0.1cm] node [anchor=north] {\$M_a\$};
40   \end{scope}
41 \end{tikzpicture}

```

E il risultato:



7.7 Esempio di relazione

Scriviamo di seguito il codice per la creazione di una relazione fac simile, in cui però il testo e i numeri sono completamente casuali dal momento in cui non abbiamo modo di farvi vedere una vera relazione di laboratorio. Quello che importa ora comunque è solo il lato estetico (e va anche ricordato che molte cose sono aggiunte in più e sono superficiali).

```

1 \documentclass[11pt,a4paper]article commento per evitare problemi
2 \usepackage{graphicx}
3 \usepackage[bottom=1.2in,margin=1in]{geometry}
4 \usepackage[italian]{babel}
5 \addto\captionsitalian{%
6   \renewcommand{\lstlistingname}{Codice}%
7   \renewcommand\lstlistlistingname{Codicici}}
8 \usepackage{setspace}
9 \usepackage[hidelinks]{hyperref}
10 \usepackage{lipsum}
11 \usepackage{amsmath}
12 \usepackage{bm}
13 \usepackage{array,multitrow}
14 \usepackage{cleveref}
15 \usepackage{chngcntr}
16 \counterwithin{table}{section}
17 \counterwithin{figure}{section}
18 \usepackage{listings}
19 \usepackage{xcolor}
20 \usepackage{cancel}
21   \renewcommand\CancelColor{\color{red}}
22 \definecolor{codegreen}{rgb}{0,0.6,0}
23 \definecolor{codegray}{rgb}{0.5,0.5,0.5}
24 \definecolor{codepurple}{rgb}{0.58,0,0.82}
25 \definecolor{backcolour}{rgb}{0.95,0.95,0.92}
26 \lstdefinestyle{mystyle}{
27   backgroundcolor=\color{backcolour},
28   commentstyle=\color{codegreen},
29   keywordstyle=\color{magenta},
30   numberstyle=\small\color{codegray},
31   stringstyle=\color{codepurple},
32   basicstyle=\ttfamily,
33   breakatwhitespace=false,
34   breaklines=true,
35   captionpos=b,
36   keepspaces=true,

```

```

37 numbers=left,
38 numbersep=5pt,
39 frame=lines,
40 showspace=false,
41 showstringspaces=false,
42 showtabs=false,
43 tabsize=3,
44 %framexleftmargin=5mm
45 }
46 \lstset{style=mystyle}
47 \spacing{1.2}
48 \setlength{\parindent}{0pt}
49 \title{Laboratorio di Meccanica - Canale A-Z \\ \vspace{0.2 cm} \\
      Large{Esperienza di Laboratorio} }
50 \vspace{0.4 cm}
51 \author{Gruppo \emph{Paperino} \\ \\
52 {Pippo \\ Pluto \\ Minnie}
53 }
54 \date{}
55 %%begindocument commento per evitare problemi
56 \counterwithin{lstlisting}{section}
57
58 \maketitle
59 \\
60 \vspace{12 cm}
61 \\
62 \begin{center}
63 \date{11 Maggio 2023}
64 \end{center}
65
66 \newpage
67 \tableofcontents
68 \newpage
69
70 \section{Introduzione}
71 \lipsum[1][1]
72 \begin{table}[h!]
73   \centering
74   \begin{tabular}{|c||c|c|c|}
75   \hline
76     {} & \textit{Pippo} & \textit{Pluto} & \textit{Minnie} \\
77   \hline
78   \hline
79   \textbf{Misure} & {X} & {X} & {X} \\

```

```

80     \hline
81     \textbf{Analisi Dati} & {X} & {X} & {} \\
82     \hline
83     \textbf{LaTeX} & {X} & {} & {X} \\
84     \hline
85     \end{tabular}
86     \caption{Divisione del lavoro}
87     \label{tab:my_label}
88 \end{table}
89 %%subsectionLipsum commento per evitare problemi
90 \lipsum[1][2]
91 \begin{itemize}
92     \item \lipsum[1][3];
93     \item \lipsum[1][4];
94     \item \lipsum[1][5];
95 \end{itemize}
96 %%subsectionFormule commento per evitare problemi
97 \label{section:formule}
98 \lipsum[1][6-10]
99 \begin{figure}[h!]
100     \centering
101     \includegraphics{Piano Inclinato.pdf}
102     \label{fig:Piano Inclinato}
103 \end{figure}
104
105 $$\int_0^1 e^x \ dx \neq 4 \cdot 10^{23}$$
106
107 \newpage
108 \lipsum[2][1-4]
109 $$F_G = G \cdot \frac{m_1 m_2}{r^2}$$
110 \lipsum[2][5-8]:
111 \begin{equation}
112     g = 9.81 \ \text{m/s}^2
113     \label{eq:g}
114 \end{equation}
115 \lipsum[3][1-2] \autoref{eq:g}, \lipsum[3][3-6].
116
117 \vspace{0.5 cm}
118 \begin{lstlisting}[language=Python]
119     if (gidsetsize <= NGROUPS_SMALL)
120         group_info->blocks[0] = group_info->small_block;
121     else {
122         for (i = 0; i < nblocks; i++) {
123             gid_t *b;

```

```

124     b = (void *)__get_free_page(GFP_USER);
125     if (!b)
126         goto out_undo_partial_alloc;
127     group_info->blocks[i] = b;
128     }
129 }
130 return group_info;
131 %\endlstlisting commento per evitare problemi
132 %%\section{Ipsum} commento per evitare problemi
133 \lipsum[5][1-5]:
134 $$\boxed{a=101.4 \pm 0.3}$$
135 \lipsum[5][6-9]
136
137 \newpage
138 \subsection{Dolor}
139 \begin{table}[h!]
140     \def\arraystretch{1.3}
141     \centering
142     \begin{tabular}{|c||c|c|c|c|}
143 \hline
144 {$x_a$} & 1 & 2 & 3 & 4 \\ \hline
145 \hline
146 \hline
147 \hline
148 1 & a & b & c & d \\ \hline
149 2 & a & b & c & d \\ \hline
150 3 & a & b & c & d \\ \hline
151 4 & a & b & c & d \\ \hline
152 5 & a & b & c & d \\ \hline
153 6 & a & b & c & d \\ \hline
154 7 & a & b & c & d \\ \hline
155 8 & a & b & c & d \\ \hline
156 9 & a & b & c & d \\ \hline
157 10 & a & b & c & d \\ \hline
158 11 & a & b & c & d \\ \hline
159 12 & a & b & c & d \\ \hline
160 13 & a & b & c & d \\ \hline
161 14 & a & b & c & d \\ \hline
162 15 & a & b & c & d \\ \hline
163 \vdots & \vdots & \vdots & \vdots & \vdots \\ \hline
164 190 & a & b & c & d \\ \hline
165 191 & a & b & c & d \\ \hline
166 192 & a & b & c & d \\ \hline
167 193 & a & b & c & d \\ \hline

```

```

168 194 & a & b & c & d \\
169 195 & a & b & c & d \\
170 196 & a & b & c & d \\
171 197 & a & b & c & d \\
172 198 & a & b & c & d \\
173  $\mu$  & a & b & c & d \\
174  $\sigma$  & a & b & c & d \\
175 \hline
176     \end{tabular}
177     \caption{abc}
178     \label{tab:dati}
179 \end{table}
180 \newpage
181 %%sectionSin commento per evitare problemi
182 \lipsum[6][1-3]
183  $\theta \approx 45^\circ$ 
184 \lipsum[6][4-9]:
185  $\boxed{b=0.3 \pm 0.1}$ 
186
187 %%sectionRisultati ommento per evitare problemi
188 \lipsum[7][1]
189 \begin{table}[h!]
190     \def\arraystretch{1.3}
191     \centering
192     \begin{tabular}{|c||c|c|c|}
193     \hline
194     \textit{Misura} & \textit{Valore} & \textit{Sigma} & \textit{
Unit di Misura} \\
195     \hline
196     \hline
197     $g$ & $9.81$ & $0.3$ & $m/s^2$ \\
198     $a$ & $101.4$ & $0.3$ & - \\
199     $b$ & $0.3$ & $0.1$ & - \\
200     \hline
201     \end{tabular}
202     \caption{Tabella riassuntiva}
203     \label{tab:my_label}
204 \end{table}
205 \lipsum[7]
206
207 \enddocument ommento per evitare problemi

```

Il risultato di questo codice è questo [documento](#)

8

Appendice



Sommario

- 8.1 Integrale di Gauss, 158
- 8.2 σ_x per la Gaussiana, 159
- 8.3 $\sigma^2[x]$ per la Binomiale, 161
- 8.4 $\mathbb{E}[x]$ per distribuzioni del χ^2 , 163
- 8.5 $\mathbb{E}[x]$ e $\text{Var}[x]$ per la Poissoniana, 164
- 8.6 Dimostrazione metodo di massima verosimiglianza, 166
- 8.7 Riproduttività Poissoniana, 168

8.1 Integrale di Gauss

L'integrale di Gauss ha la seguente forma:

$$I = \int_0^{+\infty} e^{-x^2} dx$$

Siccome non esiste una primitiva di questa funzione, il che si può dimostrare ma non lo faremo, bisogna essere un po' creativi nel risolverlo.

✓ Dimostrazione 11

Iniziamo scrivendoci l'integrale al quadrato, ovvero I^2 , il che equivale a portarlo in due dimensioni:

$$I^2 = \int_0^{+\infty} e^{-x^2} dx + \int_0^{+\infty} e^{-y^2} dy = \iint_0^{+\infty} e^{-(x^2+y^2)} dx dy$$

A questo punto passiamo alle coordinate polari, per le quale vale che:

$$\begin{cases} x = r \cos(\theta) \\ y = r \sin(\theta) \end{cases}$$

Con le condizioni che

$$\begin{cases} 0 < r < +\infty \\ 0 < \theta < \frac{\pi}{2} \end{cases}$$

E siccome per il teorema di pitagora vale che $x^2 + y^2 = r^2$ l'integrale diventa:

$$I^2 = \int_0^{\frac{\pi}{2}} \left(\int_0^{+\infty} r \cdot e^{-r^2} dr \right) d\theta$$

Questo integrale, simile a quello di partenza, ha una primitiva, e quindi posso risolverlo. In particolare l'integrale più interno vale:

$$\int_0^{+\infty} r \cdot e^{-r^2} dr = \frac{-1}{2} e^{-r^2} \Big|_0^{+\infty} = \frac{1}{2}$$

A questo punto, l'integrale più generale diventa:

$$I^2 = \int_0^{\frac{\pi}{2}} \frac{1}{2} d\theta = \frac{\pi}{4}$$

$$\implies I = \frac{\sqrt{\pi}}{2}$$

✓ Dimostrazione 11 continued

A questo punto il gioco è fatto, siccome abbiamo il valore dell'integrale sul dominio superiore, e per calcolarci quello su tutto il dominio, siccome e^{-x^2} è una funzione pari, basta prendere due volte l'integrale appena calcolato, ovvero:

$$\int_{-\infty}^{+\infty} e^{-x^2} dx = 2 \cdot I = \sqrt{\pi}$$

8.2 σ_x per la Gaussiana

Occupiamoci, dopo aver calcolato l'integrale di gauss, di calcolare anche σ_x per una distribuzione di Gauss.

✓ **Dimostrazione 12: NON DA ESAME**

Ricordandoci la formula della varianza, il calcolo che dobbiamo fare è:

$$\text{Var}[x] = \mathbb{E}[x^2] - \mathbb{E}^2[x] = \int_{-\infty}^{+\infty} x^2 \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx - \mu^2$$

Calcoliamoci ovviamente l'integrale separatamente. Per farlo, come al solito, operiamo una sostituzione in cui $t = x - \mu$ e di conseguenza $x = t + \mu$, otteniamo:

$$\rightarrow \mathbb{E}[x^2] = \frac{1}{\sigma\sqrt{2\pi}} \cdot \int_{-\infty}^{+\infty} (t + \mu)^2 \cdot e^{-\frac{t^2}{2\sigma^2}} dt$$

E svolgendo $(t + \mu)^2 = t^2 + \mu^2 + 2t\mu$, otteniamo tre integrali di cui due sono immediati, poichè:

$$\rightarrow \int_{-\infty}^{+\infty} t^2 \cdot e^{-\frac{t^2}{2\sigma^2}} dt + \underbrace{\mu^2 \cdot \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2\sigma^2}} dt}_{=\sigma\sqrt{2\pi}} + \underbrace{2\mu \int_{-\infty}^{+\infty} t \cdot e^{-\frac{t^2}{2\sigma^2}} dt}_{=0}$$

Adesso resta un solo integrale di media difficoltà, ma neanche troppa, considerando che possiamo adoperare una semplice integrazione per parti, che nel caso degli integrali definiti è $\int_a^b f(x)g'(x) dx = f(b)g(b) - f(a)g(a) - \int_a^b f'(x) dx$, e stando attenti a fare i limiti per i valori degli estremi, l'integrale diventa:

$$\int_{-\infty}^{+\infty} t^2 \cdot e^{-\frac{t^2}{2\sigma^2}} dt = \underbrace{\lim_{x \rightarrow +\infty} \left(-\frac{t \cdot e^{-\frac{t^2}{2\sigma^2}}}{\sigma^2} \right)}_{=0} - \underbrace{\lim_{x \rightarrow -\infty} \left(-\frac{t \cdot e^{-\frac{t^2}{2\sigma^2}}}{\sigma^2} \right)}_{=0} + \sigma^2 \cdot \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2\sigma^2}} dt = \sigma^3 \sqrt{2\pi}$$

✓ Dimostrazione 12 continued

A questo punto il gioco è fatto, abbiamo il valore di $\mathbb{E}[x^2]$, che è a partire dai calcoli sovrastanti

$$\frac{1}{\sigma\sqrt{2\pi}} \cdot (\sigma^3\sqrt{2\pi} + \mu^2\sigma\sqrt{2\pi}) = \sigma^2 + \mu^2$$

E ritornando alla formula iniziale della varianza, concludiamo che:

$$\text{Var}[x] = \mathbb{E}[x^2] - \mathbb{E}^2[x] = \sigma^2 + \cancel{\mu^2} - \cancel{\mu^2} = \sigma^2$$

E ci siamo ricavati i valori della varianza e della deviazione standard di una gaussiana normale:

$$\sigma_x = \sigma_x \tag{8.1}$$

$$\text{Var}[x] = \sigma_x^2 \tag{8.2}$$

8.3 $\sigma^2[x]$ per la Binomiale

Per calcolarci la varianza della distribuzione binomiale richiamiamo la funzione di probabilità in questione:

$$B(x|n, p) = \frac{n!}{x!(n-x)!} \cdot p^x \cdot q^{n-x}$$

E anche la definizione di varianza che ormai dovremmo avere stampata in fronte con l'inchiostro indelebile M15:

$$\sigma^2[x] = \mathbb{E}[x^2] - \mathbb{E}^2[x]$$

✓ Dimostrazione 13

Siccome il secondo termine della varianza è molto semplice, ovvero $\mathbb{E}^2[x] = n^2 p^2$, per adesso limitiamoci a calcolare il primo per evitare di fare confusione:

$$\begin{aligned} \mathbb{E}[x^2] &= \sum_{x=0}^n x^2 \cdot \frac{n!}{x!(n-x)!} \cdot p^x \cdot q^{n-x} = \sum_{x=0}^n \frac{x^2 \cdot n \cdot (n-1)!}{x \cdot (x-1)! \cdot (n-x)!} \cdot p^{x+1-1} \cdot q^{n-x} = \\ &= np \cdot \sum_{x=1}^n x \cdot \frac{(n-1)!}{(x-1)!(n-x)!} \cdot p^{x-1} \cdot q^{n-x} \end{aligned}$$

A questo punto, dopo aver fatto partire la sommatoria da 1 perchè tanto il primo termine vale 0, possiamo fare un cambio di variabili:

$$\begin{cases} n-1 = t \\ x-1 = y \end{cases} \implies (t+1)p \cdot \sum_{y=0}^{t+1} (y+1) \cdot \frac{t!}{y!(t-y)!} \cdot p^y \cdot q^{t-y}$$

A questo punto svolgiamo il $(y+1)$ e separiamo le sommatorie, ritrovandoci davanti due risultati abbastanza confortanti:

$$\mathbb{E}[x^2] = (t+1)p \cdot \left[\sum_{y=0}^{t+1} y \cdot \frac{t!}{y!(t-y)!} \cdot p^y \cdot q^{t-y} + \underbrace{\sum_{y=0}^{t+1} \frac{t!}{y!(t-y)!} \cdot p^y \cdot q^{t-y}}_{=1} \right]$$

Diciamo abbastanza confortanti perchè la seconda sommatoria fa 1, essendo la somma su tutto il dominio di una binomiale con parametri t, y , mentre la prima sommatoria non è altro che la definizione di valore atteso di binomiale sempre con gli stessi parametri, quindi senza neanche risolverla possiamo scrivere il risultato e cambiare di nuovo le variabili:

$$\mathbb{E}[x^2] = (t+1)p \cdot (tp+1) \implies np \cdot [(n-1)p+1]$$

✓ Dimostrazione 13 continued

A questo punto non ci resta che usare il risultato appena trovato nella definizione di varianza e includere anche l'altro termine, per trovare che:

$$\sigma^2[x] = \mathbb{E}[x^2] - \mathbb{E}^2[x] = np \cdot [(n-1)p + 1] - n^2p^2 = np \cdot (\cancel{np} - p + 1 - \cancel{np}) = np(1-p)$$

Ed eccoci arrivati al risultato finale:

$$\sigma[x]^2 = npq \tag{8.3}$$

8.4 $\mathbb{E}[x]$ per distribuzioni del χ^2

Ricordandoci che la distribuzione del χ^2 ha come forma:

$$\chi^2(x|v) = \left[2^{\frac{v}{2}} \cdot \Gamma\left(\frac{v}{2}\right)\right]^{-1} \cdot x^{\left(\frac{v}{2}-1\right)} \cdot e^{-\frac{x}{2}}$$

Per calcolarci il valore atteso facciamo l'integrale su tutto il dominio, ovvero

$$\mathbb{E}[x] = \int_{-\infty}^{+\infty} x \cdot f(x) dx$$

Ricordandoci, poichè ne faremo uso e poichè è nella formula, che la funzione gamma di Eulero è:

$$\Gamma(x) = \int_0^{+\infty} t^{(x-1)} \cdot e^{-t} dt = (x-1)!$$

✓ Dimostrazione 14

Riscriviamoci l'integrale usando la definizione della funzione, troviamo che

$$\int_{-\infty}^{+\infty} x \cdot f(x) dx = \int_{-\infty}^{+\infty} x \cdot \left[2^{\frac{v}{2}} \cdot \Gamma\left(\frac{v}{2}\right)\right]^{-1} \cdot x^{\left(\frac{v}{2}-1\right)} \cdot e^{-\frac{x}{2}}$$

Che diventa, portando fuori le costanti:

$$\rightarrow \left[2^{\frac{v}{2}} \cdot \Gamma\left(\frac{v}{2}\right)\right]^{-1} \int_{-\infty}^{+\infty} x^{\left(\frac{v}{2}\right)} \cdot e^{-\left(\frac{x}{2}\right)} dx$$

Adesso, chiamando le costanti $k \in \mathbb{R}$ per comodità, procediamo a risolvere l'integrale semplicemente tramite l'integrale funzione gamma di Eulero, quindi troviamo che

$$\rightarrow k \cdot \int_{-\infty}^{+\infty} x^{\left(\frac{v}{2}\right)} \cdot e^{-\left(\frac{x}{2}\right)} dx$$

A questo punto basta ricondurci quello che è dentro l'integrale alla funzione gamma di eulero:

$$\rightarrow k \cdot 2^{\left(\frac{v}{2}\right)} \int_{-\infty}^{+\infty} \frac{1}{2^{\left(\frac{v}{2}\right)}} \cdot x^{\left(\frac{v}{2}\right)} \cdot e^{-\left(\frac{x}{2}\right)} dx = k \cdot 2^{\left(\frac{v}{2}\right)} \cdot \int_{-\infty}^{+\infty} \left(\frac{x}{2}\right)^{\left(\frac{v}{2}\right)} \cdot e^{-\left(\frac{x}{2}\right)} dx$$

E ora, siccome conosciamo già il risultato dell'integrale, ci basta fare delle semplici osservazioni algebriche e semplificare:

$$\rightarrow \frac{1}{\cancel{2^{\frac{v}{2}}} \cdot \left(\frac{v}{2} - 1\right)!} \cdot \cancel{2^{\frac{v}{2}}} \cdot \frac{v}{2} \cdot \left(\frac{v}{2} - 1\right)! \cdot 2$$

$$\implies \mathbb{E}[x] = v$$

8.5 $\mathbb{E}[x]$ e $\text{Var}[x]$ per la Poissoniana

Dimostriamo il valore atteso e la Varianza per la Distribuzione Poissoniana:

✓ Dimostrazione 15: $\mathbb{E}[x]$

Iniziando dal Valore atteso, per definizione:

$$\mathbb{E}[x] = \sum_{x=0}^{+\infty} x \cdot P(x|\lambda) = \sum_{x=0}^{+\infty} x \cdot \frac{\lambda^x}{x!} \cdot e^{-\lambda} = \sum_{x=1}^{+\infty} x \cdot \frac{\lambda^x}{x!} \cdot e^{-\lambda}$$

A questo punto, dopo aver posto la serie per $x = 1$ poichè il primo termine è 0 (e perchè ci servirà dopo), tiriamo fuori $e^{-\lambda}$ e λ scrivendoci λ^x come $\lambda \cdot \lambda^{x-1}$:

$$\mathbb{E}[x] = \frac{1}{e^\lambda} \cdot \sum_{x=1}^{+\infty} \cancel{x} \cdot \lambda \cdot \lambda^{x-1} = \frac{\lambda}{e^\lambda} \cdot \sum_{x=1}^{+\infty} \frac{\lambda^{x-1}}{(x-1)!}$$

Ora siccome sviluppando i primi termini della serie otteniamo che

$$\sum_{x=1}^{+\infty} \frac{\lambda^{x-1}}{(x-1)!} = \frac{\lambda^0}{1} + \frac{\lambda^1}{1} + \dots = \sum_{x=0}^{+\infty} \frac{\lambda^x}{x!}$$

Ci scriviamo la serie come nel membro a destra dell'uguale, siccome sappiamo anche essere lo sviluppo di Taylor per e^λ , quindi ricaviamo facilmente:

$$\mathbb{E}[x] = \frac{1}{e^\lambda} \cdot \sum_{x=0}^{+\infty} \frac{\lambda^x}{x!} = \frac{\lambda \cdot \cancel{e^\lambda}}{\cancel{e^\lambda}} = \lambda$$

✓ Dimostrazione 16: $\text{Var}[x]$

Per quanto riguarda la varianza, sempre per definizione:

$$\text{Var}[x] = \mathbb{E}[x^2] - \mathbb{E}^2[x]$$

Quindi ci troviamo come prima cosa il valore atteso del quadrato della variabile casuale, ovvero

$$\mathbb{E}[x^2] = \sum_{x=0}^{+\infty} x^2 \cdot \frac{\lambda^x}{x!} \cdot e^{-\lambda} = \sum_{x=1}^{+\infty} x^2 \cdot \frac{\lambda^x}{x!} \cdot e^{-\lambda}$$

Anche in questo caso si fa partire la serie da 1 e non da 0 poichè fra poco ci ritroveremo sempre con $(x-1)!$ che non può assumere valori negativi. Similmente a prima ci

✓ Dimostrazione 16 continued

portiamo fuori sia $e^{-\lambda}$ e λ e rimaniamo con

$$\mathbb{E}[x^2] = \frac{\lambda}{e^\lambda} \cdot \sum_{x=1}^{+\infty} \frac{x\lambda^{x-1}}{(x-1)!}$$

E sempre come prima ci scriviamo la serie in un altro modo, ovvero:

$$\sum_{x=1}^{+\infty} \frac{x \cdot \lambda^{x-1}}{(x-1)!} = 1 + 2\lambda + \frac{3\lambda}{2} + \dots = \sum_{x=0}^{+\infty} (x+1) \frac{\lambda^x}{x!}$$

Svolgendo abbiamo quindi due serie, di cui conosciamo già il risultato: la prima l'abbiamo già svolta prima per il valore atteso, e la seconda è sempre lo sviluppo di Taylor incontrato prima:

$$\mathbb{E}[x^2] = \frac{\lambda}{e^\lambda} \cdot \left(\underbrace{\sum_{x=0}^{+\infty} x \frac{\lambda^x}{x!}}_{=\lambda e^\lambda} + \underbrace{\sum_{x=0}^{+\infty} \frac{\lambda^x}{x!}}_{=e^\lambda} \right) = \frac{\lambda}{\cancel{e^\lambda}} \cdot \cancel{e^\lambda} \cdot (\lambda + 1)$$

Da qui il passaggio è facile, basta svolgere l'equazione iniziale sostituendo i valori che conosciamo già:

$$\text{Var}[x] = \mathbb{E}[x^2] - \mathbb{E}^2[x] = \lambda(\lambda + 1) - \lambda^2 = \lambda$$

8.6 Dimostrazione metodo di massima verosimiglianza

Per dimostrare il metodo di massima verosimiglianza, partiamo con delle osservazioni che ci permettono di calcolare in modo più o meno facile il massimo della funzione $\mathcal{L}(\mu; x)$:

✓ Dimostrazione 17

Poiché la verosimiglianza è positiva, ed in generale data dal prodotto di probabilità indipendenti, è conveniente prenderne il logaritmo (trasformando prodotti in somme).

La funzione di verosimiglianza è:

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

Questo vale per un singolo punto però, mentre noi abbiamo una serie di dati sperimentali, quindi la funzione di verosimiglianza finale sarà il prodotto di tutte le singole $\mathcal{L}(\mu; x)$, per la regola della moltiplicazione:

$$f(\mu|\vec{x}) = \prod_{i=0}^n f(\mu|x_i)$$

Quindi facendone il logaritmo e ricordandocene le proprietà ci troviamo:

$$\begin{aligned} \log\left(\prod_{i=0}^n \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right) &= \sum_{i=0}^n \log\left(\frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right) = \\ &= \sum_{i=0}^n \left(\log\frac{1}{\sigma\sqrt{2\pi}} + \log e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right) = \sum_{i=0}^n \left(\underbrace{\log 1}_{=0} - \log(\sigma\sqrt{2\pi}) - \frac{(x_i-\mu)^2}{2\sigma^2}\right) = \\ &= \sum_{i=0}^n \left(-\log(\sigma) - \log(\sqrt{2\pi}) - \frac{(x_i-\mu)^2}{2\sigma^2}\right) \end{aligned}$$

Siamo arrivati al punto finale. O meglio, alla semplificazione del logaritmo della funzione di verosimiglianza che ci permette di trovare i valori cercati tramite la ricerca del valore che massimizza la funzione. Il massimo di una funzione si fa chiaramente tramite la derivata, e prima di calcolarla, visto che comunque il minimo di f sarà nello stesso punto del minimo di $\log(f)$:

$$\begin{aligned} \frac{\partial \log \mathcal{L}(\mu; x)}{\partial \mu} = 0 &\implies \hat{\mu} \sim \mathbb{E}[\mu] \\ \frac{\partial^2 \log \mathcal{L}(\mu; x)}{\partial \mu^2} \Big|_{\hat{\mu}} &\sim -\frac{1}{\sigma^2[\hat{\mu}]} \end{aligned}$$

✓ Dimostrazione 17 continued

Quest'ultima disuguaglianza super ambigua e di cui si capisce poco deriva dalla disuguaglianza di Cramér-Rao che non dimostreremo (yippiyaye). A questo punto facciamo un piccolo passo concettuale in più e in realtà non necessario, atto a capire meglio il discorso dell'uso del logaritmo: usiamo Taylor per approssimare a una parabola in un intorno del massimo $\hat{\mu}$:

$$\begin{aligned}\log \mathcal{L}(\mu) &= \log \mathcal{L}(\mu) + \frac{1}{2} \frac{\partial^2 \log \mathcal{L}(\mu; x)}{\partial \mu^2} \Big|_{\hat{\mu}} (\mu - \hat{\mu})^2 \\ &= \log \mathcal{L}(\mu) - \frac{(\mu - \hat{\mu})^2}{2\sigma^2}\end{aligned}$$

Da qui si ricava in antitesi:

$$\begin{aligned}\log \mathcal{L}(\hat{\mu} \pm \sigma_{\hat{\mu}}) &= \log \mathcal{L}(\hat{\mu}) - \frac{1}{2} \\ \Delta \log \mathcal{L}(\hat{\mu}) &= \log \mathcal{L}(\hat{\mu} \pm \sigma_{\hat{\mu}}) - \log \mathcal{L}(\hat{\mu}) = -\frac{1}{2}\end{aligned}$$

Ovvero per ricavare l'incertezza si varia μ dal valore che massimizza la funzione finché non si è ridotta di 0.5.

8.7 Riproduttività Poissoniana

Dimostriamo adesso la riproduttività della Poissoniana. Per farlo, siccome stiamo sostanzialmente operando con una variabile $z = x + y$ in cui $x, y \sim P$, utilizziamo il concetto di convoluzione visto in [Sezione 2.14](#).

✓ Dimostrazione 18: NON DA ESAME

Per quanto riguarda la Poissoniana, per la definizione di convoluzione, la probabilità della somma di due variabili è

$$P(z) = \sum_{x=0}^z f(x)f(z-x)$$

Quindi operando con le due poissoniane

$$x \sim P(x|\lambda_1) = \frac{\lambda_1^x}{x!} \cdot e^{-\lambda_1}; \quad y \sim P(y|\lambda_2) = \frac{\lambda_2^y}{y!} \cdot e^{-\lambda_2}$$

Sviluppiamo la somma e otteniamo:

$$\sum_{x=0}^z P(x|\lambda_1)P(z-x|\lambda_2) = \sum_{x=0}^z \frac{\lambda_1^x}{x!} \cdot e^{-\lambda_1} \cdot \frac{\lambda_2^{z-x}}{(z-x)!} \cdot e^{-\lambda_2}$$

Possiamo già portare fuori gli esponenziali che non dipendono da x , che diventando per le proprietà degli esponenziali $e^{-\lambda_1} \cdot e^{-\lambda_2} = e^{-(\lambda_1+\lambda_2)}$, quindi abbiamo già un tassello della Poissoniana. Moltiplichiamo poi per $z!/z!$ e otteniamo

$$e^{-(\lambda_1+\lambda_2)} \cdot \sum_{x=0}^z \frac{\lambda_1^x}{x!} \cdot \frac{\lambda_2^{z-x}}{(z-x)!} \cdot \frac{z!}{z!} = \frac{e^{-(\lambda_1+\lambda_2)}}{z!} \cdot \sum_{x=0}^z \frac{z!}{x!(z-x)!} \cdot \lambda_1^x \cdot \lambda_2^{z-x}$$

A questo punto stiamo costruendo a sinistra della sommatoria la nostra Poissoniana, di cui manca solo $(\lambda_1 + \lambda_2)^z$, e a destra la somma sul dominio di una Binomiale di parametri $x = z, n = z$ e una probabilità che avrà a che fare con entrambi i λ . Adesso infatti facciamo altre due moltiplicazioni per ottenere il risultato finale:

$$\rightarrow \frac{e^{-(\lambda_1+\lambda_2)}}{z!} \cdot \sum_{x=0}^z \frac{z!}{x!(z-x)!} \cdot \lambda_1^x \cdot \frac{\lambda_2^z}{\lambda_2^x} \cdot \left(\frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2}\right)^x \cdot \left(\frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2}\right)^z$$

✓ Dimostrazione 18 continued

A questo punto portiamo fuori $(\lambda_1 + \lambda_2)^z$ e facendo gli opportuni cambi d'esponente rimaniamo con:

$$\frac{(\lambda_1 + \lambda_2)^z}{z!} \cdot e^{-(\lambda_1 + \lambda_2)} \cdot \underbrace{\sum_{x=0}^z \frac{z!}{x!(z-x)!} \cdot \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^x \cdot \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{z-x}}_{=1} = \frac{(\lambda_1 + \lambda_2)^z}{z!} \cdot e^{-(\lambda_1 + \lambda_2)}$$

E chiamando $\lambda = \lambda_1 + \lambda_2$ ci siamo trovati una poissoniana di parametri

$$z = x + y, \lambda = \lambda_1 + \lambda_2$$

$$z \sim P(z|\lambda) = \frac{\lambda^z}{z!} \cdot e^{-\lambda}$$