

Appunti sull'elaborazione statistica dei dati sperimentali

F. Della Valle

Dipartimento di Fisica, e-mail:dellavalle@ts.infn.it

http://www.ing.univ.trieste.it/dispense/dellavalle/AnalisiDati_2004.pdf

ottobre 2004

Queste note si propongono di fornire un'introduzione non completamente rigorosa e certamente non completa, ma abbastanza autoconsistente, all'elaborazione statistica dei dati sperimentali. La scelta degli argomenti e il taglio sono pensati per un'esposizione di 8-10 ore di fronte a studenti del primo anno di Ingegneria. Sarò grato a chi segnalerà errori o vorrà indirizzarmi commenti.

Indice

1	Generalità	3
2	Misure dirette non ripetute; errore di sensibilità	4
3	Incertezze di misura e cifre significative. Errore relativo e percentuale	6
4	Misure indirette non ripetute; propagazione dell'errore massimo	9
5	Strumenti di misura	11
6	Probabilità e frequenza	14
7	Distribuzione di probabilità	18
8	Misure dirette ripetute: istogramma; legge normale degli errori	21
9	Distribuzioni derivate	24
10	Media e varianza della media	26
11	Metodo della massima verosimiglianza; minimi quadrati	29

1 Generalità

Lo scopo della Fisica è quello di dare descrizioni quantitative delle leggi che regolano i fenomeni fisici, di formulare cioè espressioni matematiche che legano le grandezze fisiche implicate nei fenomeni naturali.

Cominciamo col definire una grandezza fisica come una caratteristica di un sistema individuata da un'operazione di misura il cui risultato rappresenta, per così dire, l'intensità con cui la grandezza è presente, nelle condizioni date, nel sistema in esame. Si noti come la definizione di grandezza fisica è rimandata direttamente a quella di misura: sono cioè grandezze fisiche tutte e sole le proprietà di un sistema per le quali sia possibile stabilire una procedura di misura non ambigua. Tali sono ad esempio le masse, le dimensioni spaziali, gli intervalli di tempo, le velocità, etc. Siamo quindi portati a chiederci come si effettui una misura. Per il momento ci limitiamo a considerare la *misura diretta*, che è la più semplice e sta a fondamento di tutte le altre.

Misurare direttamente una grandezza significa confrontarla quantitativamente con un suo campione a cui viene attribuita convenzionalmente intensità pari a uno e che prende il nome di *unità di misura*. Questa definizione, come vedremo meglio più avanti, implica che esistano, per la grandezza in esame, un criterio operativo di confronto (uguaglianza e disuguaglianza) e una procedura di somma. La scelta dell'unità di misura è largamente arbitraria e dettata solo dalla convenienza: le unità di misura di tutte le grandezze sono cambiate più volte nel corso degli anni per adattarsi alle esigenze sociali, della scienza e della tecnologia.¹ Per una descrizione dei diversi sistemi di unità di misura esistenti si rimanda al corso di Fisica Generale. In ogni caso, per la misura si usano multipli e sottomultipli dell'unità di misura, vale a dire campioni della grandezza le cui intensità, rispetto all'unità di misura, stanno nel rapporto di numeri interi. La descrizione dettagliata del processo di misura diretta sarà l'oggetto del prossimo paragrafo; altri tipi di misura saranno considerati più avanti.

Se una stessa grandezza può essere definita in più modi diversi, questi for-

¹Prima dell'attuale definizione del metro nel Sistema Internazionale come la distanza percorsa dalla luce nel vuoto in un intervallo di tempo di $1/299792458$ secondi, che è stata adottata nel 1983, nel 1791 l'Accademia delle Scienze francese preferì definirlo come la decimilionesima parte del meridiano terrestre di Parigi (distanza tra il Polo e l'Equatore) piuttosto che la lunghezza di un pendolo con un certo periodo di oscillazione, per ovviare alla variabilità dell'accelerazione di gravità sulla superficie terrestre. Rispetto alla definizione, il primo campione risultò più corto di 0.2 mm. Nel 1889 la definizione di metro non faceva più riferimento alle dimensioni terrestri ma soltanto ad un certo campione di platino-iridio mantenuto a pressione atmosferica alla temperatura di 0°C e sostenuto da ben determinati sostegni meccanici. Nel 1960 fu deciso che la definizione di metro fosse legata alla lunghezza d'onda di una certa radiazione atomica del krypton 86.

niscono misure tra loro proporzionali. A titolo di esempio, schematicamente, e senza nessuna pretesa di completezza, notiamo che le lunghezze possono essere ragionevolmente misurate per confronto diretto solo in un intervallo di valori compreso approssimativamente tra 1 mm (10^{-3} m, millimetro) e 1 km (10^3 m, kilometro); per valori più piccoli di questi si ricorre a sofisticati sistemi meccanici (calibri, micrometri, fino a circa $1 \mu\text{m}$ o 10^{-6} m, micrometro), ottici (microscopi, interferometri, fino a circa 1 nm o 10^{-9} m, nanometro), a esperimenti di Fisica Atomica (fino a circa 1 pm o 10^{-12} m, picometro), di Fisica Nucleare (fino a circa 10^{-15} m o 1 fm, femtometro), o di Fisica sub-Nucleare (fino a circa 10^{-18} m o 1 am, attometro; per valori più grandi di 1 km si possono usare gli odometri (fino a circa 10^6 m, 1 Mm, megometro), si passa poi a sistemi elettroottici (fino a circa 10^9 m, 1 Gm, gigometro) o a triangolazioni ottiche (fino a circa 10^{12} m, 1 Tm, terometro); al di là di questo valore rimangono i metodi dell'Astrofisica. Nei casi in cui si possono effettuare le misure usando più di una tecnica si può verificare l'omogeneità delle diverse definizioni di lunghezza.

2 Misure dirette non ripetute; errore di sensibilità

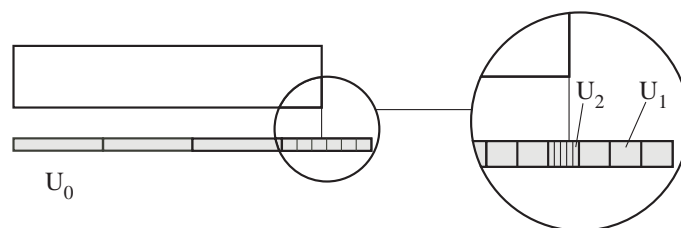


Figura 1: Misura diretta di lunghezza.

Sia G la grandezza in esame e U_0 la sua unità di misura. Per fare un esempio concreto faremo riferimento ad una misura di lunghezza, ad esempio la dimensione del piano superiore di un tavolo. L'unità di misura potrebbe essere, in questo caso, un regolo sottile di metallo, che andrà orientato secondo la direzione dello spigolo del tavolo con un'estremità allineata con il bordo del tavolo. Si marca quindi il tavolo in corrispondenza della posizione della seconda estremità e si sposta il regolo unitario parallelamente a sé stesso fino a farne coincidere la prima estremità con questo segno. Si ripete l'operazione finché è possibile, determinando così quante unità entrano nella larghezza del

tavolo. È evidente che, nel caso generale, alla fine di questa fase della misura il secondo bordo del tavolo non si troverà in corrispondenza della estremità del regolo campione. Definiamo quindi un sottomultiplo U_1 , indicato in questo caso da una serie di tacche equispaziate sul regolo campione. Anche in questo caso il bordo del tavolo non capiterà in coincidenza con una delle tacche. Si potrà allora definire un secondo e più piccolo sottomultiplo dell'unità suddividendo l'intervallo tra le tacche in parti uguali. Il processo può essere ripetuto più volte, e il risultato della misura viene allora costruito in linea di principio come una somma di termini:

$$\begin{aligned} G &= G_0U_0 + G_1U_1 + G_2U_2 + \dots \\ G_0 &= \text{int} \left(\frac{G}{U_0} \right) \\ G_1 &= \text{int} \left(\frac{G - G_0U_0}{U_1} \right) \\ G_2 &= \text{int} \left(\frac{G - G_0U_0 - G_1U_1}{U_2} \right) \\ &\vdots \end{aligned}$$

dove la funzione $\text{int}()$ rappresenta la parte intera della frazione. Dalla descrizione del processo di misura appare evidente che il valore del singolo rapporto

$$\frac{G - \sum_{j=0}^{i-1} G_j U_j}{U_i}$$

non è mai noto: tutto quello che in ciascun confronto si determina è il numero massimo di volte che il sottomultiplo U_i entra nella quantità $G - \sum_{j=0}^{i-1} G_j U_j$. Nel processo di misura evidentemente non occorre dunque mai fare troncamenti: ciascun passo del processo di misura fornisce di già un numero intero.

È anche evidente che in pratica la somma verrà fermata ad un certo ordine i_{max} . A differenza degli altri termini $G_{i_{max}}$ è *arrotondato* e non troncato:

$$G_{i_{max}} = \text{int}' \left(\frac{G - \sum_{i=0}^{i_{max}-1} G_i U_i}{U_{i_{max}}} \right)$$

dove la funzione $\text{int}'()$ rappresenta l'intero più vicino al valore della frazione. Chiameremo la quantità $U_{i_{max}}/2$ *risoluzione* della misura, e la citeremo sempre insieme al risultato della misura. Il risultato di una singola misura diretta è dunque espresso come l'abbinamento di due numeri:

$$G_{mis} \equiv \sum_{i=0}^{i_{max}} G_i \frac{U_i}{U_0}, \quad \Delta G \equiv \frac{1}{2} \frac{U_{i_{max}}}{U_0} \quad [\text{in unità } U_0]. \quad (1)$$

Con questa scrittura si intende dire che l'intensità di G può differire da $G_{mis}U_0$ per una quantità più piccola di $U_{i_{max}}/2$. Per descrivere questa situazione si usa affermare che la quantità $U_{i_{max}}/2$ rappresenta l'incertezza legata alla risoluzione con cui si è effettuata la misura ovvero - con termine forse improprio ma che è ormai passato nel gergo, - l'*errore di sensibilità*. Qui come più avanti nella teoria della misura il termine "errore" non ha il valore che il linguaggio corrente gli assegna, dal momento che nella maggioranza dei casi non sta a suggerire nessuna possibile correzione del risultato della misura, ma è usato come sinonimo di incertezza. In altre parole, la presenza di un "errore" ΔG associato alla misura diretta indica che, nell'eventualità in cui si ripeta la stessa misura migliorandone la risoluzione, si troverà il nuovo valore della misura di G in un punto qualsiasi dell'intervallo $[G_{mis} - \Delta G, G_{mis} + \Delta G]$. A priori, infatti, se ci si basa solo sull'espressione (1), tutti i punti di questo intervallo sono *equiprobabili*. Si noti che la risoluzione di una misura è dimensionalmente omogenea alla grandezza G .

Esercizio. Si effettui la misura diretta di Fig. 1.

Svolgimento:

$$\begin{array}{rcl}
 G_0 = 3, & G_1 = 2, & G_2 = 3 \\
 \frac{U_1}{U_0} = \frac{1}{6}, & \frac{U_2}{U_0} = \frac{U_2}{U_1} \frac{U_1}{U_0} = \frac{1}{30} & \\
 G_{mis} = 3 + \frac{2}{6} + \frac{3}{30} = 3.433 & \Delta G = \frac{1}{60} = 1.7 \cdot 10^{-2} & \text{[in unità } U_0\text{]}
 \end{array}$$

per cui si può scrivere

$$G = (3.433 \pm .017) U_0$$

Scegliendo di usare una sola cifra significativa per ΔG scriveremo invece

$$G_{mis} = 3.43, \quad \Delta G = 2 \cdot 10^{-2} \quad \text{[in unità } U_0\text{]}$$

Sulle cifre significative si dirà meglio nel prossimo paragrafo.

3 Incertezze di misura e cifre significative. Errore relativo e percentuale

Vedremo più avanti che la conclusione raggiunta sopra non è legata alla particolare misura descritta, ma vale per tutti i tipi di misura, anche se la definizione (e in parte anche il significato) di ΔG possono risultare abbastanza diversi: il risultato di una qualsiasi misura non è un numero che rappresenta

il valore “vero” della grandezza in studio, ma piuttosto una stima, della cui bontà ΔG è un indice. Si può avere anzi ragione di sostenere che un tale “valore vero” non esista affatto. Pensiamo di nuovo alla misura di lunghezza descritta sopra; supponiamo ad esempio, in via di pura ipotesi, di poter disporre di uno strumento di misura con una risoluzione pari alla distanza tra due atomi ($\sim 10^{-10}$ m). Scopriremmo *a*) che gli atomi del tavolo sono in continuo movimento attorno ad una posizione di equilibrio, e che quindi la loro posizione non può essere definita con assoluta precisione; *b*) che se si tiene fisso un estremo del tavolo la posizione dell’altro estremo, osservata con risoluzione atomica, cambia con la temperatura e che variazioni infinitesime - e di fatto incontrollabili - della temperatura danno luogo a variazioni di lunghezza molto più grandi delle distanze interatomiche; *c*) che, a livello atomico, risulta impossibile stabilire se gli atomi che compongono gli strati più esterni della superficie del tavolo appartengano al tavolo o all’ambiente che lo circonda: di fatto vi sono continuamente atomi che si staccano dalla superficie e altri che vi aderiscono, provenendo dall’atmosfera o da altri oggetti che vengono in contatto con la superficie.

Ma ben prima di arrivare ad una tale situazione estrema, l’esperienza mostra che quando la sensibilità non è troppo scarsa (cioè l’“errore” troppo grande) la ripetizione delle misure porta a risultati che differiscono tra loro per più della risoluzione dello strumento di misura. Prendendo nuovamente ad esempio la misura di lunghezza discussa prima, è facile che da una misura all’altra l’allineamento del regolo alle tacche e la tracciatura delle stesse possa essere fatta in modo diverso per motivi che possono avere a che fare, per esempio, con il fatto che la tacca non può mai avere larghezza nulla, che la sua direzione può non essere perfettamente ortogonale allo spigolo del tavolo, che durante l’operazione di tracciatura delle tacche il regolo può spostarsi inavvertitamente, etc. Il complesso di tutte le circostanze che influiscono sul risultato della misura, anche in un caso così semplice, sfugge quindi al completo controllo dello sperimentatore. Anche in questo caso, come vedremo in dettaglio più avanti, si può solo stimare il miglior valore della misura e associare ad esso un numero che descrive la dispersione dei dati.

L’esistenza di una incertezza sperimentale, di qualsiasi natura essa sia, pone delle limitazioni al numero di *cifre significative* (vale a dire di cifre non precedute o seguite da soli zeri) con cui viene presentato il risultato della misura. Si usa scrivere il valore di ΔG con non più di due cifre significative, e si riportano le cifre del valore di G_{mis} solo fino alla posizione dell’ultima cifra di ΔG . Nel caso in cui manchi l’indicazione di ΔG , si assume per questo valore metà dell’unità dell’ultima cifra del valore di G_{mis} .

Gli errori di sensibilità e la dispersione delle misure non rappresentano le uniche fonti di incertezza associate ad una misura. Se ad esempio il campio-

ne dell'unità di misura di lunghezza è costituito di un materiale soggetto a dilatazione termica in misura diversa dall'oggetto della misura, si otterranno risultati differenti operando a temperature diverse. L'incertezza derivante da questa situazione prende il nome di *errore sistematico*, dove il termine "sistematico" sta a ricordare che in principio questo tipo di errore può essere previsto ed eventualmente, a differenza di quello di sensibilità, corretto. Si può pensare, ad esempio, di registrare la temperatura a cui si effettuano le misure e, dalla conoscenza dei coefficienti di dilatazione termica dei materiali in gioco, di dedurre dai dati misurati effettivamente quelli che si sarebbero misurati ad un'altra temperatura. In alternativa si può provvedere a stabilizzare la temperatura del sistema. In tutti i casi l'incertezza sulla misura della temperatura si traduce in un'incertezza (o "errore") sistematica sulla misura di lunghezza. Occorre però dire che, a parte l'esempio banale che è stato fatto, la correzione degli errori sistematici non è sempre facile da effettuare, e certamente la loro presenza non può essere messa in evidenza finché si usi un solo strumento di misura (un solo campione nel caso della misura diretta) o non si varino le condizioni della misura.

Un modo alternativo di presentare il dato dell'errore di misura è quello dell'*errore relativo* cioè del rapporto tra l'incertezza di misura ΔG e il valore della misura,

$$\Delta G_{rel} \equiv \frac{\Delta G}{G_{mis}}$$

L'errore relativo permette di confrontare misure di grandezze fisiche differenti. Ad esempio, si può facilmente vedere che la misura di una massa di 2 tonnellate con $\Delta M = 5 \text{ kg}$ ($\Delta M_{rel} = 2.5 \cdot 10^{-3}$) è affetta da un'incertezza minore della misura di una lunghezza di 30 cm con un $\Delta L = 1 \text{ mm}$ ($\Delta L_{rel} = 3.3 \cdot 10^{-3}$). A differenza dell'errore di misura, l'errore relativo è un numero puro, e viene talvolta anche presentato - moltiplicato per 100 - come percentuale. Si usa indicare l'errore relativo con lo stesso numero di cifre significative dell'errore assoluto.

Esercizio. Si correggano le cifre significative della seguente tabella di misure; si espliciti l'errore dove è sottinteso; si calcoli l'errore relativo della misura [nella tabella il testo dell'esercizio è in grassetto, lo svolgimento usa un carattere normale]:

G	ΔG	G_{mis}	ΔG_{rel}
1.234	0.12	1.23	$9.8 \cdot 10^{-2}$
1.234	.1	1.2	$8 \cdot 10^{-2}$
1.23456	.0001	1.2346	$8 \cdot 10^{-5}$
1	.012	1.000	$1.2 \cdot 10^{-2}$
1234	120	$1.23 \cdot 10^3$	$9.8 \cdot 10^{-2}$
1.2	.05	1.20	$4 \cdot 10^{-2}$
1	.5	1.0	.5
2	.5	2.0	.25

4 Misure indirette non ripetute; propagazione dell'errore massimo

Consideriamo ora una grandezza F legata a G da una relazione funzionale:

$$F = F(G).$$

È chiaro che effettuare una misura di G equivale ad effettuare una determinazione del valore di F :

$$F_{mis} = F(G_{mis}).$$

Ad esempio, si può ottenere la lunghezza L di una circonferenza misurando il diametro D del cerchio e calcolando la relazione $L = \pi D$. Oppure stimare il volume V di un cubo misurando la lunghezza L dello spigolo e usando la formula $V = L^3$.

Se l'errore di sensibilità nella misura di G è piccolo, l'ampiezza dell'intervallo associato all'errore di sensibilità sulla misura di F sarà dato da

$$|F(G_{mis} + \Delta G) - F(G_{mis} - \Delta G)| \approx 2 \left| \left(\frac{dF}{dG} \right)_{G=G_{mis}} \right| \Delta G.$$

Si può assumere quindi

$$\Delta F = \left| \left(\frac{dF}{dG} \right)_{G=G_{mis}} \right| \Delta G.$$

Per i due casi di prima si avrebbe

$$\begin{aligned} L_{mis} &= \pi D_{mis} & \Delta L &= \pi \Delta D \\ V_{mis} &= L_{mis}^3 & \Delta V &= 3L_{mis}^2 \Delta L \end{aligned}$$

Si capisce facilmente che nel caso di funzioni di più variabili, se cioè

$$F = F(A, B, C, \dots)$$

allora

$$F_{mis} = F(A_{mis}, B_{mis}, C_{mis}, \dots)$$

e

$$\Delta F = \left| \frac{\partial F}{\partial A} \right| \Delta A + \left| \frac{\partial F}{\partial B} \right| \Delta B + \left| \frac{\partial F}{\partial C} \right| \Delta C + \dots \quad (2)$$

Questa definizione, che usa il valore assoluto delle derivate parziali, corrisponde a riconoscere che il caso peggiore, quello in cui gli errori di sensibilità nelle singole misure primarie concorrono a dare un errore di sensibilità su F che sia massimo, ha probabilità uguale a qualsiasi altro. Effettuando, ad esempio, una misura della velocità media V di un corpo misurando lo spazio L percorso e il tempo T impiegato a percorrerlo, se le incertezze sulle misure sono rispettivamente ΔL e ΔT , l'incertezza sulla misura di velocità è data da

$$\Delta V = \frac{1}{T} \Delta L + \left| -\frac{L}{T^2} \right| \Delta T = \frac{1}{T_{mis}} \Delta L + \frac{L_{mis}}{T_{mis}^2} \Delta T.$$

La propagazione dell'errore massimo nelle espressioni monomie prende una forma particolarmente semplice se ci si riferisce all'errore relativo massimo:

$$G = A^a \cdot B^b \cdot C^c \cdot \dots$$

$$\begin{aligned} \Delta G_{rel} = \frac{\Delta G}{G} &= |a| \frac{\Delta A}{A} + |b| \frac{\Delta B}{B} + |c| \frac{\Delta C}{C} + \dots = \\ &= |a| \Delta A_{rel} + |b| \Delta B_{rel} + |c| \Delta C_{rel} + \dots \end{aligned}$$

Ad esempio, per il caso precedente della velocità media $V = L/T$ si avrebbe semplicemente ($|a| = |b| = 1$)

$$\Delta V_{rel} = \Delta L_{rel} + \Delta T_{rel}$$

Esercizio. Si calcoli l'incertezza ΔS con cui è nota l'area della superficie S di un rettangolo di lati $A = 20.2$ cm e $B = 7.3$ mm.

Svolgimento. Calcoliamo l'errore relativo

$$\Delta S_{rel} = \frac{\Delta S}{S} = \Delta A_{rel} + \Delta B_{rel} = \frac{\Delta A}{A} + \frac{\Delta B}{B} = \frac{.05}{20.2} + \frac{.05}{7.3} = 9.3 \cdot 10^{-3}$$

da cui

$$\Delta S = \Delta S_{rel} S = \Delta S_{rel} A B = .14 \text{ cm}^2$$

Esercizio. Si vuole misurare la profondità h di un pozzo lasciandovi cadere dentro un sasso e misurando il tempo t intercorso tra la partenza

del sasso e il l'istante in cui lo sperimentatore sente il rumore dell'urto sul fondo del pozzo. Si mostri che l'espressione

$$h = \frac{1}{2}gt^2$$

è affetta da un errore sistematico dovuto al fatto che il suono si propaga a velocità v , la si corregga e si calcoli l'errore di sensibilità Δh in funzione della risoluzione Δt .

Si calcolino le espressioni ottenute per $v = 300 \text{ m/s}$, $t_{mis} = 2.1 \text{ s}$, $g = 9.8 \text{ m/s}^2$, $\Delta t = 0.05 \text{ s}$.

Svolgimento. L'espressione corretta per h è data da

$$h = \frac{1}{2}g \left(t - \frac{h}{v} \right)^2$$

che è un'equazione di secondo grado in h , la cui soluzione è

$$h_{mis} = \frac{v^2}{g} \left(1 - \sqrt{1 + 2g t_{mis}/v} \right) + v t_{mis} = 20.2 \text{ m}$$

$$\Delta h = \left| \frac{dh}{dt} \right|_{t=t_{mis}} \Delta t = v \left(1 - \frac{1}{\sqrt{1 + 2g t_{mis}/v}} \right) \Delta t = 0.9 \text{ m}$$

La formula semplificata avrebbe invece dato

$$h_{mis} = \frac{1}{2}g t_{mis}^2 = 21.6 \text{ m}$$

con una differenza (di natura sistematica) per eccesso di 1.4 m. Un'altra fonte di incertezza sistematica in questa misura è la resistenza dell'aria, che porta anch'essa a sovrastimare la profondità.

5 Strumenti di misura

La gran parte delle misure viene effettuata con apposite apparecchiature che permettono il confronto della grandezza in esame con l'unità di misura e che prendono il nome di *strumenti di misura*. A ciascun valore della grandezza corrisponde un numero se lo strumento è *digitale*, la posizione (angolare o lineare) di un indice su una scala graduata su cui generalmente sono marcati i valori della grandezza in esame se lo strumento è *analogico*. La risoluzione di una misura strumentale è quindi legata alla risoluzione con cui viene letta la scala graduata dalla legge di propagazione degli errori in una misura indiretta. Sia G la grandezza in esame e Z quella realmente osservata. Valgono relazioni del tipo

$$G = G(Z) \quad \Delta G = \left| \frac{dG}{dZ} \right| \Delta Z$$

Generalmente il valore di ΔG può essere desunto direttamente dalla scala graduata di uno strumento analogico, ovvero corrisponde alla metà della cifra meno significativa usata da uno strumento digitale. La quantità $|dG/dZ|^{-1}$ è una caratteristica dello strumento che prende il nome di *sensibilità*; un valore alto della sensibilità significa che a piccole variazioni della grandezza da misurare corrispondono grandi cambiamenti della grandezza osservata Z .



Figura 2: Nonio.

Talvolta, la sensibilità di lettura di una scala analogica può essere migliorata mediante l'uso di un *nonio*, un piccolo regolo scorrevole, graduato in modo tale che n divisioni della scala del nonio corrispondano a $n - 1$ divisioni della scala principale. Questo significa che la distanza d fra le tacche tracciate sulla scala del nonio è più piccola della distanza D tra le tacche della scala principale per una quantità pari a D/n . Con riferimento alla figura, supponiamo di effettuare una misura diretta di lunghezza; come si è già visto, in generale, il secondo estremo dell'oggetto da misurare verrà a cadere tra due tacche della scala principale; il nonio, che va posizionato di seguito all'oggetto, permette di stimare, con sensibilità D/n , la frazione di tacca residua. Infatti, se chiamiamo x la quantità di cui lo zero del nonio è spostato verso destra rispetto ad una tacca della scala principale, la prima tacca del nonio dista dalla successiva tacca della scala principale di $x - D/n$, la seconda di $x - 2D/n$, e così via. Se la k -ma divisione del nonio coincide (o quasi) con la posizione di una tacca principale si ha, evidentemente, $x = kD/n$. Nel caso riportato in figura, la misura dell'oggetto ottenuta con l'uso del nonio è di 16.7 unità della scala principale.

Oltre alla risoluzione e alla sensibilità, altre caratteristiche di uno strumento di misura sono la *giustezza*, la *ripetibilità*, la *prontezza*, e la *portata*, che è il massimo valore della grandezza misurabile con lo strumento.

La prontezza è misurata dal valore di un tempo che caratterizza la velocità di risposta dello strumento.

La giustezza (detta anche *accuratezza*) esprime la massima deviazione possibile del risultato della misura dal risultato di una misura ideale, non affetta cioè da errore sistematico. La giustezza è legata quindi al massimo

errore sistematico che lo strumento può introdurre nella misura. Può assumere valori diversi a seconda del valore della misura: questo è evidente per tutti gli strumenti che possono lavorare con portata variabile, ma vale anche all'interno di un'unica scala (non linearità dello strumento). Essendo espressione dell'errore sistematico, la giustezza gioca quindi lo stesso ruolo della risoluzione, e ne prende il posto quando sia più grande di questa. Questa situazione è quella che si presenta più frequentemente: generalmente le caratteristiche dei singoli componenti di uno strumento sono note solo per essere comprese all'interno di un'intervallo massimo di variabilità. In linea puramente teorica si può pensare di eliminare questo effetto (o ridurlo al di sotto della soglia di sensibilità) attraverso una taratura dello strumento, da effettuarsi, immediatamente prima o dopo ciascuna misura, con un campione della grandezza G di valore noto. Un'errata o mancata taratura dello strumento può portare a veri e propri errori di misura. Come è stato già sottolineato, l'errore sistematico non è evidenziabile finché si usi un solo strumento di misura non tarato, ma può solo essere stimato (incertezza sistematica), generalmente a partire dalle informazioni fornite dal costruttore dello strumento. Esistono anche strumenti detti *assoluti*, che non richiedono taratura in quanto il loro funzionamento fornisce la misura basandosi su una legge fisica; un esempio di questo tipo di strumenti è il manometro di MacLeod, per una descrizione del quale si rimanda ai corsi di Fisica Tecnica.

La ripetibilità (detta anche *precisione*) esprime la capacità di uno strumento di dare risposte uguali nelle stesse condizioni. Indipendentemente da quanto lo strumento sia giusto, è possibile ed anzi normale che misure successive di una stessa grandezza, pur effettuate con lo stesso strumento, diano valori diversi. I motivi più vari possono determinare un tale comportamento: può essere l'attrito dell'ago dell'indicatore negli strumenti analogici, o un'interferenza elettromagnetica nella circuiteria, o una variabilità della resistenza di un contatto elettrico, etc.. Quello che accomuna tutte queste situazioni è che esse sfuggono al controllo dell'operatore. Se così non fosse, la causa potrebbe essere individuata e neutralizzata; o, anche, la misura potrebbe essere corretta per l'influenza della perturbazione, trattata quindi come un errore sistematico. Questo non è il caso generale.

Si noti, poi, che nella progettazione degli strumenti di misura si tende ad avere un bilanciamento delle prestazioni in sensibilità e in ripetibilità: infatti, dal momento che la precisione determina sostanzialmente il prezzo dello strumento, non è conveniente avere strumenti molto ripetibili e poco sensibili, giacché ripetibilità significa avere una piccola dispersione dei risultati, per sfruttare appieno la quale occorre abbinarla con una piccola risoluzione e quindi con un'alta sensibilità dello strumento. Di contro, strumenti ad alta sensibilità ma scarsa ripetibilità danno incertezze di misura non corrispon-

denti alla sensibilità, bensì determinate dalla dispersione dei dati. Come nel caso precedente, lo strumento avrebbe delle capacità che nella pratica non possono essere sfruttate, e il tempo e i soldi investiti nella progettazione di un tale strumento non sarebbero affatto ben spesi. Occorre quindi, invece, che le due caratteristiche risultino opportunamente bilanciate. Per questo motivo è il più delle volte inutile cercare di leggere su una scala graduata piccole frazioni della sua risoluzione: molto probabilmente la risoluzione nella lettura di Z non è stata migliorata dal costruttore perché la ripetibilità non lo consente.

Ci dobbiamo ora porre il problema di come trattare i risultati di misure ripetute, e di come ricavare da essi un valore per l'incertezza. Prima di fare questo introduciamo i concetti di probabilità e di densità di probabilità.

6 Probabilità e frequenza

Il termine probabilità è comunemente usato per indicare, più o meno correttamente, una stima riguardo ad un evento di cui non si abbiano informazioni sufficienti per prevederne l'esito con certezza. La teoria della probabilità si è sviluppata su basi matematiche a partire dal '600, inizialmente sul tema dei giochi d'azzardo, ma può essere (ed è) utilmente applicata ad un ventaglio molto più vasto di situazioni. Si applica in tutti quei casi in cui una situazione si può presentare con modalità diverse in circostanze apparentemente identiche, quando non si voglia o non sia possibile specificare completamente le condizioni iniziali e i meccanismi secondo cui il sistema evolve. Ad esempio, il lancio di una moneta o di un dado è certamente un evento deterministico, ma di fatto risulta impossibile controllarlo in modo da ottenere un risultato prefissato: la minima variazione delle condizioni iniziali porta a risultati completamente diversi. Esperimenti di questo tipo prendono il nome di esperimenti *casuali*, il cui risultato è cioè affidato al caso, e i loro risultati sono *variabili casuali*.

Pensiamo ad un tale esperimento che possa svolgersi secondo n modalità diverse. Queste modalità siano mutuamente esclusive e siano, per così dire, su un piede di parità: non vi sia cioè, *a priori*, nessun motivo di ritenere che una o più di esse avvenga in via preferenziale. Per avere degli esempi si può pensare al lancio di una moneta ($n = 2$) o al lancio di un dado ($n = 6$), all'estrazione di un numero nel gioco del Lotto ($n = 90$). Sia E una qualsiasi espressione logica che assuma il valore 1 (vero) o il valore 0 (falso) in ciascuna delle n modalità di svolgimento dell'esperimento. Per fissare le idee possiamo pensare all'estrazione di una biglia da una scatola che ne contiene n , ciascuna numerata con un numero diverso. L'evento a cui siamo interessati potrebbe

essere l'estrazione di uno dei numeri x_1, \dots, x_m ($m < n$) su cui abbiamo puntato:

$$E = \{x : x \in \{x_1, \dots, x_m\}\}$$

Definiamo *probabilità a priori* dell'evento E il rapporto

$$P(E) = \frac{n_E}{n}$$

tra il numero n_E di modalità dell'esperimento in cui l'espressione E è verificata, e il numero totale $n = n_E + n_{\bar{E}}$ di casi ugualmente possibili. Nel nostro caso evidentemente n è il numero totale di biglie contenute nella scatola, e n_E vale m . La probabilità vale allora m/n e assume dunque valori tra 0 e 1. Il primo caso ($P = 0$) esprime l'impossibilità *a priori* di E , e il secondo ($P = 1$) la sua certezza (sempre *a priori*). Il primo caso si verifica, ad esempio se, nel gioco del Lotto, abbiamo puntato solo su numeri inferiori a 1 e superiori a 90; il secondo se includiamo nella nostra giocata del Lotto tutti i novanta numeri che possono essere estratti ($m = n$). Analogamente, la probabilità che esca la testa nel lancio della moneta vale 0.5; la croce ha anch'essa probabilità 0.5; la probabilità che esca una data faccia del dado è $1/6$, mentre la probabilità che una data faccia non esca è il complemento all'unità di questo valore, cioè $5/6$.

Dalla definizione risulta immediatamente evidente che le probabilità di eventi indipendenti si sommano:

$$P(E_1 \text{ o } E_2) = P(E_1) + P(E_2)$$

Questa proprietà è già stata implicitamente enunciata sopra a proposito della probabilità dell'estrazione del Lotto.

Supponiamo ora di effettuare N volte l'esperimento. Definiamo *frequenza dell'evento E* il rapporto

$$f(E) = \frac{N_E}{N}$$

tra il numero N_E di volte in cui l'esperimento si è svolto secondo una modalità in cui E è stata verificata, e il numero $N = N_E + N_{\bar{E}}$ totale di prove. È evidente che la frequenza cambia al crescere del numero totale di prove, ma è un fatto sperimentale che al crescere di N le oscillazioni della frequenza diminuiscono di ampiezza, e il suo valore converge a quello della probabilità *a priori*. L'esperienza ha portato a formulare il seguente postulato, noto come "legge empirica del caso": al crescere del numero delle prove la frequenza tende a stabilizzarsi attorno ad un valore limite che coincide con la probabilità *a priori*:

$$\lim_{N \rightarrow \infty} f(E) = \lim_{N \rightarrow \infty} \frac{N_E}{N} = \frac{n_E}{n}.$$

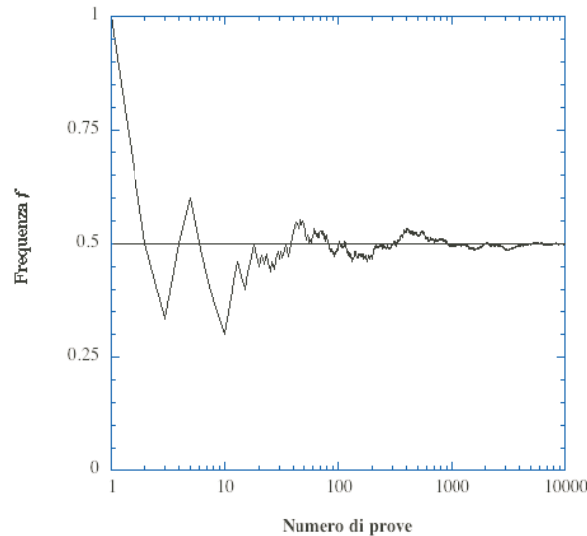


Figura 3: Frequenza del risultato testa (o croce) nel lancio di una moneta, in funzione del numero di prove. Si noti la scala logaritmica dell'ascissa.

In questa enunciazione si è usato il formalismo dell'analisi matematica, secondo la quale una funzione f converge ad un valore f_0 per $x \rightarrow +\infty$ se, scelto un ϵ piccolo a piacere, esiste un x_0 tale che, per qualsiasi $x > x_0$, $|f(x) - f_0| < \epsilon$. È evidente però che la frequenza non è una funzione matematica, ma un risultato sperimentale, per il quale il numero di prove (la x della definizione di limite) rimane sempre un numero finito. Senza volere entrare nelle sottigliezze del ragionamento, si dirà solo che in realtà la legge empirica del caso enunciata sopra significa soltanto che *la probabilità che la frequenza sia un numero sensibilmente diverso dalla probabilità diminuisce al crescere del numero N di tentativi*. Detto in altri termini, *mi aspetto che in media, su n prove, n_E siano favorevoli a E* . Una conseguenza interessante (e molto spesso sottovalutata) della legge empirica del caso è che se al crescere del numero di prove persiste una differenza tra $f(E)$ e $P(E)$, cresce anche la probabilità che sia stata valutata erroneamente $P(E)$. Per esempio, se un certo numero non viene estratto al Lotto per troppo tempo, anziché puntare su questo ritardatario occorrerà prendere in considerazione l'ipotesi che questo numero non sia in effetti nell'urna insieme agli altri. Questo approccio equivale a stimare la probabilità a partire dalle frequenze, cosa che talvolta rappresenta l'unico modo di farlo. Prima di entrare nel vivo dell'argomento dobbiamo però andare avanti con la teoria della probabilità.

Si vuole adesso studiare la probabilità composta. Sia E_2 una seconda espressione logica, indipendente da E_1 , concernente i risultati dell'esperimento, e si voglia calcolare la probabilità che si verifichino entrambi gli eventi. Facendo riferimento ad esempio all'estrazione di due palline numerate si può pensare, affinché gli eventi siano indipendenti, di avere due urne identiche da ciascuna delle quali si estrae una pallina, oppure di avere una sola urna nella quale viene rimessa la prima pallina subito dopo l'estrazione. Infatti, se la seconda pallina viene estratta senza rimettere a posto la prima, le condizioni per la seconda prova non sono identiche a quelle della prima, e le prove non sono più indipendenti; le formule presentate qui non valgono in questo caso, per il quale si rimanda ai corsi di Teoria della Probabilità. È facile convincersi che la probabilità *a priori* che siano verificate contemporaneamente sia E_1 che E_2 è data, nel caso di eventi indipendenti, dal prodotto delle singole probabilità:

$$P(E_1 \text{ e } E_2) = P(E_1) \cdot P(E_2) = \frac{n_{E_1}}{n} \frac{n_{E_2}}{n}.$$

Esercizio. Si stimino le probabilità di ciascuno dei possibili risultati del lancio contemporaneo di due dadi.

Svolgimento. Per contare il numero di casi possibili notiamo che ciascuna faccia del primo dado può essere abbinata, nel lancio, a ciascuna delle sei facce del secondo. Il numero di casi possibili è dunque $n = 36$, e ciascuno di essi ha dunque probabilità $1/36$. Questo numero è anche, correttamente, il quadrato della probabilità del singolo lancio.

Il risultato 2 (risp. 12) può essere ottenuto in un unico caso, cioè se i due dadi mostrano entrambi le facce 1 (risp. 6); la sua probabilità è dunque $1/36$.

Il risultato 3 (risp. 11) può essere ottenuto solo in due casi, se il primo dado mostra la faccia 1 (risp. 6) e il secondo la faccia 2 (risp. 5) o viceversa; la sua probabilità è dunque $2/36$.

Il risultato 4 (risp. 10) può essere ottenuto in tre modi diversi: se entrambi i dadi mostrano la faccia 2 (risp. 5) oppure se il primo mostra la faccia 1 (risp. 6) e il secondo la faccia 3 (risp. 4) o viceversa; la sua probabilità è dunque $3/36$.

Il risultato 5 (risp. 9) può essere ottenuto in quattro modi diversi: se il primo dado mostra la faccia 1 (risp. 6) e il secondo la faccia 4 (risp. 3) o viceversa, oppure se il primo dado mostra la faccia 2 (risp. 5) e il secondo la faccia 3 (risp. 4) o viceversa; la sua probabilità è dunque $4/36$.

Il risultato 6 (risp. 8) può essere ottenuto in cinque modi diversi: se entrambi i dadi mostrano la faccia 3 (risp. 4), oppure se il primo dado mostra la faccia 1 (risp. 6) e il secondo la faccia 5 (risp. 2) o viceversa, o anche se il primo dado mostra la faccia 2 (risp. 5) e il secondo la faccia 4 (risp. 3) o viceversa; la sua probabilità è dunque $5/36$.

Il risultato 7 può essere ottenuto in sei modi diversi: se il primo dado mostra la faccia 1 e il secondo la faccia 6 o viceversa, oppure se il primo dado mostra la faccia 2 e il secondo la faccia 5 o viceversa, o anche se il primo dado mostra la faccia 3 e il secondo la faccia 4 o viceversa; la sua probabilità è dunque $6/36$.

Come si vede la somma di tutte le probabilità dà 1.

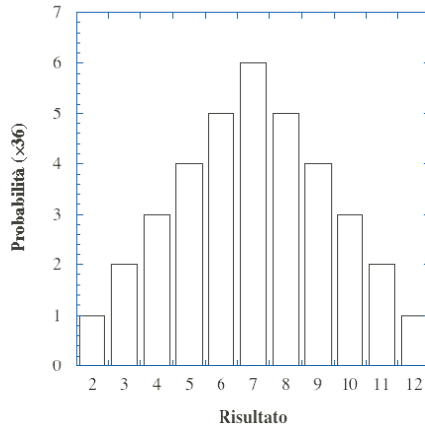


Figura 4: Distribuzione delle probabilità per il lancio di due dadi.

7 Distribuzione di probabilità

La funzione rappresentata sopra è un semplice esempio di una *distribuzione di probabilità*, cioè della corrispondenza tra ciascun evento possibile e la sua probabilità. Nel caso appena descritto si può vedere che la distribuzione è simmetrica attorno al valore 7, dove ha un massimo. Esperimenti diversi hanno distribuzioni di probabilità diverse. Queste devono in principio avere due sole caratteristiche universali: assumere solo valori positivi e che la somma di tutti i valori sia l'unità.

Si vuole ora estendere la teoria fin qui esposta ai risultati delle misure. Infatti, essendo impossibile prevedere l'esatto risultato di una misura, questo può essere considerato una variabile casuale. A differenza delle variabili trattate nel paragrafo precedente, che possono avere solo valori discreti, il risultato di una misura assume valori razionali. Per fissare le idee trattiamo il caso della misura non ripetuta di una grandezza X in cui l'incertezza sia determinata dall'errore di sensibilità Δx . Come si è già detto, tutti i punti dell'intervallo $[x_{mis} - \Delta x, x_{mis} + \Delta x]$ sono da considerarsi equiprobabili rispetto ad una successiva determinazione x'_{mis} della stessa grandezza

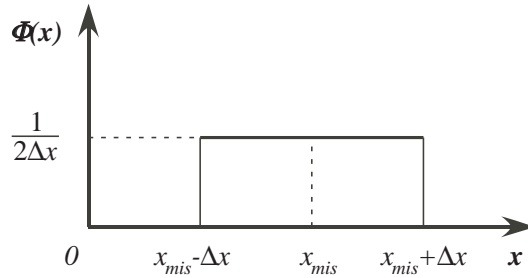


Figura 5: Distribuzione rettangolare di probabilità.

effettuata con uno strumento dotato di miglior risoluzione. Appare subito evidente che, a causa della densità del campo dei numeri razionali, la probabilità associata ad un qualsiasi punto x_0 dell'intervallo è zero. Ha invece senso chiedersi quale sia la probabilità associata ad un qualsiasi intervallo finito compreso nell'intervallo di incertezza. Questa probabilità deve evidentemente dipendere solo dalla dimensione δx dell'intervallo e non dalla posizione del suo centro, e assumere valore unitario quando l'intervallo coincide con $[x_{mis} - \Delta x, x_{mis} + \Delta x]$. Possiamo allora scrivere:

$$dP [x - \delta x/2 \leq x'_{mis} \leq x + \delta x/2] = \frac{\delta x}{2\Delta x} = \Phi(x) \delta x$$

dove

$$\Phi(x) = \frac{1}{2\Delta x}$$

è l'analogo nel continuo della distribuzione di probabilità vista sopra in un caso discreto e prende anche il nome di *densità di probabilità* della variabile x . Ha il senso che l'integrale di $\Phi(x)$ tra x_1 e x_2 rappresenta la probabilità *a priori* che il valore di x sia compreso tra questi due valori:

$$P [x_1 \leq x \leq x_2] = \int_{x_1}^{x_2} \Phi(x) dx.$$

Nel nostro caso

$$P [x_1 \leq x \leq x_2] = \frac{x_2 - x_1}{2\Delta x}$$

Questa densità di probabilità, descritta in Fig. 5, prende il nome di distribuzione rettangolare.

Nel caso generale $\Phi(x)$ non ha ovviamente un'espressione così semplice, ma dipende effettivamente da x . Le uniche condizioni che $\Phi(x)$ deve soddisfare sono, di nuovo:

$$0 \leq \Phi(x) \quad \forall x; \quad \int_{-\infty}^{+\infty} \Phi(x) dx = 1.$$

Quando $\Phi(x)$ non è nota, essa può essere stimata a partire dalla frequenza. Definendo come evento l'espressione $E = \{x - \delta x/2 \leq x_{mis} \leq x + \delta x/2\}$, si ha

$$\Phi(x) = \frac{1}{\delta x} \lim_{N \rightarrow \infty} \frac{N_E}{N}.$$

Come si è già detto il problema centrale della Statistica è quello di ottenere una stima del valore di una grandezza e dell'incertezza che lo accompagna a partire da un certo numero di misure. È evidente che la massima informazione ottenibile in un esperimento è rappresentata dalla distribuzione di probabilità della variabile associata alla misura. Ovviamente questa informazione non è mai disponibile, giacché richiederebbe un numero infinito di misure. Anche nei casi trattati sopra (lancio di monete, dadi, estrazione di biglie), in cui potrebbe sembrare che sia tutto chiaro *a priori*, occorre notare che le distribuzioni di probabilità che sono state discusse in quella sede riguardano situazioni sperimentali "oneste"; niente assicura, *a priori*, che un qualsiasi esperimento reale lo sia (si veda, a questo proposito, alla fine del paragrafo precedente, il *caveat* sui numeri ritardatari del Lotto).

In realtà non è necessario determinare la forma esatta della distribuzione. Si può far vedere che il più delle volte, per caratterizzare la distribuzione di probabilità, basta determinarne due sole grandezze caratteristiche, la *media* μ e la *varianza* σ^2 , che sono sostanzialmente i *momenti* primo e secondo della distribuzione:

$$\mu \equiv \int_{-\infty}^{+\infty} x \Phi(x) dx \quad (3)$$

$$\sigma^2 \equiv \int_{-\infty}^{+\infty} (x - \mu)^2 \Phi(x) dx = \int_{-\infty}^{+\infty} x^2 \Phi(x) dx - \mu^2$$

dove nell'ultimo passaggio si è usata l'unitarietà dell'integrale di Φ . Cogliamo l'occasione per notare che, essendo la varianza, per costruzione, un numero non negativo, vale

$$\langle x^2 \rangle = \int_{-\infty}^{+\infty} x^2 \Phi(x) dx \geq \left(\int_{-\infty}^{+\infty} x \Phi(x) dx \right)^2 = \langle x \rangle^2$$

dove si è introdotta la notazione con le parentesi triangolari $\langle \rangle$ per indicare il *valore atteso* o *aspettato* di una variabile casuale. Notiamo esplicitamente, giacché servirà più avanti, che il valore atteso del momento primo calcolato attorno al valor medio è zero:

$$\langle x - \mu \rangle = 0.$$

Mentre il significato della media è chiaro di per sé, notiamo che la varianza rappresenta una misura della larghezza della distribuzione. Più la varianza

è piccola più alta è la probabilità che le misure si addensino attorno al valor medio. La radice quadrata della varianza, σ , prende il nome di deviazione standard, ed ha le stesse dimensioni fisiche di x e μ . Come sarà chiaro più avanti il significato geometrico preciso della deviazione standard dipende dalla forma della distribuzione.

Esercizio. Si calcolino media e varianza della distribuzione di probabilità rettangolare.

Svolgimento.

$$\mu = \int_{-\infty}^{+\infty} x \Phi(x) dx = \frac{1}{2\Delta x} \int_{x_{mis}-\Delta x}^{x_{mis}+\Delta x} x dx = x_{mis}$$

$$\sigma^2 = \int_{\mu-\Delta x}^{\mu+\Delta x} (x - \mu)^2 \Phi(x) dx = \frac{1}{2\Delta x} \int_{-\Delta x}^{+\Delta x} x^2 dx = \frac{(\Delta x)^2}{3}$$

8 Misure dirette ripetute: istogramma; legge normale degli errori

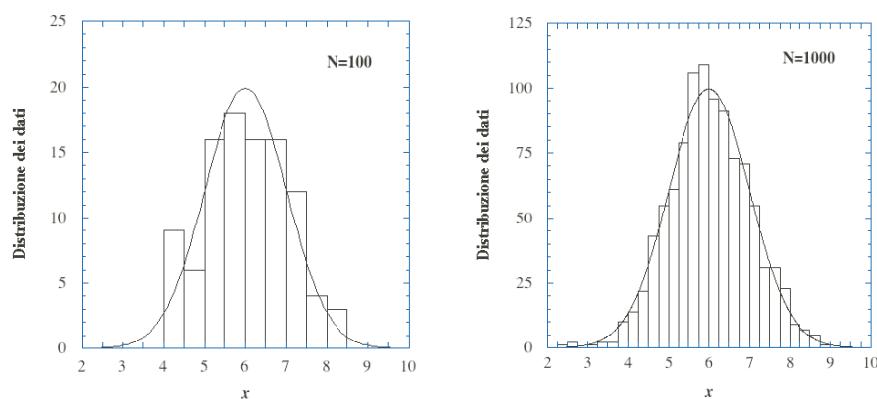


Figura 6: Istogrammi dei risultati di N determinazioni della variabile casuale ottenuta dalla somma di dodici numeri casuali compresi tra 0 e 1. Agli istogrammi è sovrapposta, per confronto, la frequenza prevista, nel limite $N \rightarrow \infty$, da una distribuzione di Gauss con $\xi = 6$ e $\sigma = 1$.

È ben noto che ripetendo una stessa misura, pur nelle stesse condizioni e con il medesimo strumento, i risultati non danno sempre un medesimo valore.

Questa variabilità è dovuta a piccole alterazioni casuali e non controllabili delle condizioni di misura. Strumento di analisi principe per questo caso è l'istogramma. Supponiamo di effettuare N misure di una grandezza X , con N grande, e di ottenere i valori

$$\{x_1, x_2, x_3, \dots, x_N\}.$$

Scegliamo due numeri opportuni x_A e x_B tali che

$$x_A < x_i < x_B \quad \forall i.$$

Dividiamo l'intervallo $[x_A, x_B]$ in n sottointervalli di uguale ampiezza $\delta x = (x_B - x_A)/n$ che chiamiamo *celle*. Il numero n deve essere opportunamente scelto: in prima approssimazione si suggerisce di usare $n \approx \sqrt{N}$. I centri x_{Ci} delle celle hanno ascisse

$$\{x_{Ci} = x_A + (i - 1/2) \delta x\}_{i=1,n}.$$

Su questo insieme discreto definiamo la funzione istogramma $I(x_{Ci})$, una funzione a valori interi che associa a ciascun punto x_{Ci} il numero di misure che cadono nella cella di centro x_{Ci} .

È un dato sperimentale il fatto che al crescere del numero N di misure la funzione $I(x_{Ci})$ converge, una gran parte delle volte, nei punti del dominio, alla funzione

$$\frac{N\delta x}{\sqrt{2\pi}s} \exp \left[-\frac{(x - m)^2}{2s^2} \right]$$

dove m e s sono due parametri opportunamente scelti. Il primo rappresenta l'ascissa rispetto alla quale la funzione è simmetrica, e il secondo ha a che fare con la dispersione delle misure (e con l'altezza della campana). L'espressione scritta sopra definisce la densità di probabilità *gaussiana* o *normale*:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(x - \xi)^2}{2\sigma^2} \right].$$

Questa funzione è di gran lunga la più importante della Teoria della Probabilità, e l'unica che trattiamo in un qualche dettaglio. In un certo senso è una legge della natura, ma spiegare il motivo per cui si presenta così frequentemente per descrivere la distribuzione dei risultati delle misure esula dagli scopi del presente scritto che ha carattere prettamente introduttivo. Si può dimostrare un teorema noto come *teorema del limite centrale della media*, che assicura che una qualunque variabile casuale continua che sia combinazione lineare di un certo numero di altre variabili ha distribuzione di probabilità

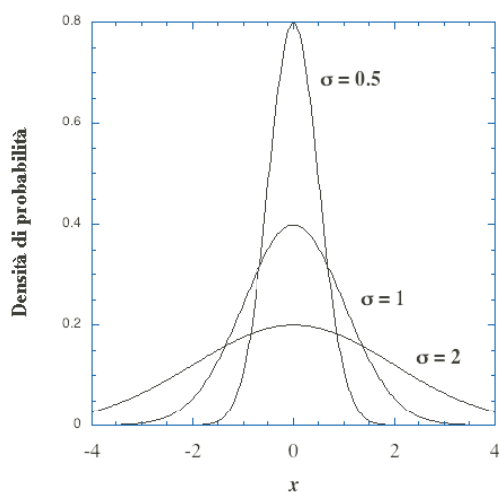


Figura 7: Densità di probabilità gaussiana o normale con $\xi = 0$ e per tre diversi valori del parametro σ .

normale. Per un corollario di questo teorema anche una variabile casuale al cui valore contribuiscano un gran numero di effetti casuali, ciascuno dei quali è responsabile di una piccola variazione rispetto al valore medio, segue la distribuzione di probabilità di Gauss. In quest'ultima categoria ricade una gran parte delle misure. Per la dimostrazione, particolarmente laboriosa, di questo teorema si rimanda ai testi di Calcolo delle Probabilità.

Per verificare che la probabilità totale sia l'unità, si integra la distribuzione sul piano reale invece che sull'asse reale:

$$\begin{aligned}
 & \int_{-\infty}^{+\infty} \Phi(x) dx \int_{-\infty}^{+\infty} \Phi(y) dy = \\
 &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\xi)^2}{2\sigma^2}\right] dx \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y-\xi)^2}{2\sigma^2}\right] dy = \\
 &= \frac{1}{\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-(x^2+y^2)} dx dy = \frac{1}{\pi} \int_0^{2\pi} d\phi \int_0^{\infty} e^{-r^2} r dr = 1.
 \end{aligned}$$

Calcoliamo ora il valore aspettato della variabile x che segue la distribuzione di Gauss (anche se il risultato è già insito nella simmetria della funzione):

$$\langle x \rangle = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\xi)^2}{2\sigma^2}\right] x dx =$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{x^2}{2\sigma^2}\right] (x + \xi) dx = \xi.$$

Analogamente,

$$\begin{aligned} \langle (x - \xi)^2 \rangle &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x - \xi)^2}{2\sigma^2}\right] (x - \xi)^2 dx = \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2} x^2 dx = \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2} dx - \frac{\sigma^2}{\sqrt{2\pi}} \left[x e^{-x^2/2} \right]_{-\infty}^{+\infty} = \sigma^2. \end{aligned}$$

Si noti che la curva ha due punti di flesso per $x = \xi \pm \sigma$; si può infatti facilmente verificare, mediante un calcolo diretto, che in questi punti la derivata seconda cambia di segno. Per una distribuzione gaussiana la probabilità che una misura capiti ad una distanza dal massimo inferiore a $k\sigma$

$$P_k \equiv P[\xi - k\sigma \leq x \leq \xi + k\sigma] = \int_{\xi - k\sigma}^{\xi + k\sigma} \Phi(x) dx$$

è tabulata e vale, per diversi valori di k ,

$$P_1 = 0.6827, \quad P_2 = 0.9545, \quad P_3 = 0.9973, \quad P_4 = 0.9999.$$

9 Distribuzioni derivate

Consideriamo una grandezza z , legata a x da una relazione del tipo $z = g(x)$. Si può ottenere la distribuzione di probabilità $\Gamma(z)$ di z imponendo la seguente equazione tra le probabilità infinitesime:

$$\Gamma(z) dz = \Phi(x) dx$$

da cui

$$\Gamma(z) = \Phi(x) \frac{dx}{dz} = \frac{\Phi(x)}{\dot{g}(x)}.$$

È quindi evidente che, a meno che la funzione $g(x)$ sia lineare la densità di probabilità $\Gamma(z)$ non sarà dello stesso tipo di $\Phi(x)$. Come abbiamo fatto per la variabile x , ci proponiamo di calcolare il valore aspettato $\langle z \rangle$

$$\langle z \rangle = \int \Gamma(z) z dz = \int_{-\infty}^{\infty} \Phi(x) g(x) dx.$$

Generalmente questa equazione è troppo difficile da risolvere, e si preferisce ricorrere a delle approssimazioni. Supponendo che la $\Phi(x)$ sia grande solo in un piccolo intorno di $\langle x \rangle$, si può pensare di restringere l'integrazione a questo intorno. Supponendo che in questo intorno la funzione $g(x)$ non vari apprezzabilmente si può sostituirla con il suo valore al primo ordine in $x - \langle x \rangle \equiv x - \xi$:

$$g(x) \approx g(\xi) + \left(\frac{dg}{dx} \right)_{x=\xi} (x - \xi).$$

In questa approssimazione il valore atteso di z varrà quindi

$$\langle z \rangle \approx g(\xi) + \dot{g}(\xi) \langle x - \xi \rangle = g(\xi) \equiv \zeta.$$

Analogamente

$$\sigma_z^2 \equiv \langle (z - \zeta)^2 \rangle = \langle z^2 \rangle - \zeta^2 \approx [\dot{g}(\xi)]^2 \langle (x - \xi)^2 \rangle = [\dot{g}(\xi)]^2 \sigma_x^2$$

dove abbiamo indicato con σ_x^2 la varianza della distribuzione della variabile x .

Nel caso in cui z sia funzione di più variabili indipendenti:

$$z = g(a, b, c, \dots)$$

la distribuzione di probabilità di z sarà data dalla seguente equazione :

$$\Gamma(z) dz = \Phi_a da \cdot \Phi_b db \cdot \Phi_c dc \cdot \dots$$

e supponendo di potere scrivere, con ovvio significato dei simboli:

$$g(a, b, c \dots) \approx g(\alpha, \beta, \gamma, \dots) + \left(\frac{\partial g}{\partial a} \right)_{\alpha, \beta, \dots} (a - \alpha) + \left(\frac{\partial g}{\partial b} \right)_{\alpha, \beta, \dots} (b - \beta) + \dots$$

si avrà

$$\zeta \equiv \langle z \rangle \approx g(\alpha, \beta, \gamma, \dots)$$

e

$$\begin{aligned} \sigma_z^2 &\equiv \langle (z - \zeta)^2 \rangle \approx \left\langle \left[\left(\frac{\partial g}{\partial a} \right)_{\alpha, \beta, \dots} (a - \alpha) + \left(\frac{\partial g}{\partial b} \right)_{\alpha, \beta, \dots} (b - \beta) + \dots \right]^2 \right\rangle = \\ &= \left[\left(\frac{\partial g}{\partial a} \right)_{\alpha, \beta, \dots} \right]^2 \sigma_a^2 + \left[\left(\frac{\partial g}{\partial b} \right)_{\alpha, \beta, \dots} \right]^2 \sigma_b^2 + \dots \end{aligned} \quad (4)$$

dove l'azzerarsi dei termini misti deriva dall'ipotesi di indipendenza delle variabili a, b, c , etc.. Questa espressione, che descrive la propagazione dell'incertezza statistica, va confrontata con l'equazione (2) a pag. 10, che descrive la propagazione dell'errore massimo.

10 Media e varianza della media

Abbiamo già detto che il primo e principale problema dell'analisi statistica consiste nello stimare i parametri della distribuzione di probabilità. In questo paragrafo mostreremo come stimare questi parametri a partire dall'insieme dei dati sperimentali. Per evitare confusioni, nel seguito indicheremo con le lettere greche μ e σ i parametri delle distribuzioni di probabilità, e con le lettere latine m e s i corrispondenti parametri stimati dall'analisi dei dati.

È intuitivo che, se la migliore approssimazione disponibile alla distribuzione di probabilità è l'istogramma, per analogia con la (3) a pag. 20 una stima della media della distribuzione può essere rappresentata dalla espressione

$$\sum_{i=1}^n f_i x_{C_i}$$

dove le $f_i = I(x_{C_i})/N$ sono le frequenze associate alle celle dell'istogramma e la somma è effettuata su tutte le celle. Supponiamo ora di diminuire la larghezza delle celle fino a far sì che in ciascuna trovi posto un solo dato. L'espressione precedente diventa allora la ben nota formula della media aritmetica di N numeri

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

Vogliamo mostrare che \bar{x} rappresenta una stima di ξ . Innanzitutto, notiamo che possiamo considerare l'equazione precedente come una relazione funzionale lineare tra \bar{x} e N variabili indipendenti x_i :

$$\bar{x} = \bar{x}(x_1, x_2, x_3, \dots, x_N).$$

Sappiamo che le x_i hanno tutte la stessa distribuzione di probabilità:

$$\xi_i = \xi \quad \text{e} \quad \sigma_i = \sigma \quad \forall i.$$

Avremo quindi

$$m \equiv \langle \bar{x} \rangle = \frac{1}{N} \sum_{i=1}^N \langle x_i \rangle = \frac{1}{N} \sum_{i=1}^N \xi = \xi$$

Notiamo che, indipendentemente da se la distribuzione di x sia gaussiana o meno, per il teorema del limite centrale della media (paragrafo 8) la densità di probabilità di \bar{x} è gaussiana. La sua distribuzione di probabilità è caratterizzata dalla varianza

$$\sigma_{\bar{x}}^2 \equiv \langle (\bar{x} - \xi)^2 \rangle = \left\langle \left[\frac{1}{N} \sum_{i=1}^N (x_i - \xi) \right]^2 \right\rangle = \frac{1}{N^2} \sum_{i=1}^N \langle (x_i - \xi)^2 \rangle = \frac{\sigma^2}{N}.$$

Nel penultimo passaggio l'ipotesi di indipendenza delle variabili x_i assicura che per i prodotti misti vale

$$\langle (x_i - \xi)(x_j - \xi) \rangle = 0.$$

Questo risultato, oltre ad assicurare che la media aritmetica rappresenta una stima del valor medio della distribuzione di probabilità delle misure, dice che la distribuzione di probabilità (gaussiana) della media presenta una dispersione minore di quella dei dati sperimentali.

Con un ragionamento analogo a quello fatto per la media potremmo sperare che l'espressione

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

rappresenti una stima della varianza della distribuzione delle x . Dimostriamo invece che l'espressione corretta richiede un fattore $N - 1$ anziché N :

$$s^2 = \frac{1}{N - 1} \sum_{i=1}^N (x_i - \bar{x})^2$$

Calcoliamone infatti il valore atteso:

$$\begin{aligned} \langle s^2 \rangle &= \frac{1}{N - 1} \sum_{i=1}^N \langle [(x_i - \xi) + (\xi - \bar{x})]^2 \rangle = \\ &= \frac{N}{N - 1} \sigma^2 + \frac{N}{N - 1} \sigma_{\bar{x}}^2 + \frac{2}{N - 1} \sum_{i=1}^N \langle (x_i - \xi)(\xi - \bar{x}) \rangle. \end{aligned}$$

Ricordando che $\sigma_{\bar{x}}^2 = \sigma^2/N$ e

$$\sum_{i=1}^N (x_i - \xi)(\xi - \bar{x}) = -N(\bar{x} - \xi)^2$$

si ha

$$\langle s^2 \rangle = \frac{N}{N - 1} \sigma^2 + \frac{1}{N - 1} \sigma^2 - \frac{2}{N - 1} \sigma^2 = \sigma^2.$$

Riassumendo: se di una stessa grandezza x esistono N determinazioni differenti, purché appartengano tutte alla stessa *popolazione statistica* (cioè abbiano tutte la medesima distribuzione di probabilità), si usa citare, come risultato della misura, la media sperimentale dei valori \bar{x} , e come incertezza la deviazione standard sperimentale $\sqrt{s^2/N}$ della media dei dati:

$$x_{mis} = \bar{x} \pm \sqrt{\frac{s^2}{N}} = \frac{1}{N} \sum_{i=1}^N x_i \pm \sqrt{\frac{1}{N(N - 1)} \sum_{i=1}^N (x_i - \bar{x})^2} \quad [P = 68\%]$$

Si noti che il fatto che le misure provengano tutte da una stessa popolazione statistica non può essere provato con certezza, non potendosi dimostrare che le condizioni di misura siano rimaste invariate nel corso delle misure: della difficoltà di rilevare la presenza di errori sistematici e di correggerli si è già detto sopra. Si può pensare di eliminare dati che siano troppo lontani dalla media dell'insieme delle misure, ma rimandiamo ai corsi avanzati la discussione dei criteri da adottare per scartare i dati. Perché un qualsiasi criterio possa essere credibilmente applicato occorre però che il numero N di dati non sia troppo esiguo. L'aumento del numero di rilevazioni sperimentali implica per contro tempi e costi più lunghi, e aggrava il problema di garantire l'uniformità delle condizioni sperimentali su tutto l'insieme delle misure, mentre la diminuzione della deviazione standard della media $\sigma_{\bar{x}}$ al crescere di N è lento: per raddoppiare la precisione di misura occorre quadruplicare il numero dei dati.

Esercizio. Si elabori la seguente tabella delle misure di una variabile x e si dia il risultato.

	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	6.035	0.3376	0.11396
2	5.711	0.0136	0.00018
3	4.237	-1.4604	2.13284
4	5.732	0.0346	0.00120
5	5.821	0.1236	0.01527
6	6.609	0.9116	0.83097
7	5.080	-0.6174	0.38121
8	4.266	-1.4314	2.04898
9	5.264	-0.4334	0.18786
10	6.406	0.7086	0.50208
11	6.778	1.0806	1.16764
12	6.430	0.7326	0.53667
Somme	68.369	0.0002	7.91885

Il risultato della misura è dunque

$$x_{mis} = \bar{x} \pm \sqrt{\frac{s^2}{N}} = 5.70 \pm 0.24 \quad [P = 68\%]$$

dove i due valori sono le stime sperimentali della media ξ e della deviazione standard $\sigma_{\bar{x}}$ della distribuzione di probabilità della media dei valori sperimentali. Si noti che ξ è anche la media della distribuzione delle misure; una stima della deviazione standard σ di questa seconda distribuzione è data da $s = 0.85$. Si noti che, nel caso presentato, la sensibilità delle misure è, incongruamente, un fattore ~ 100 migliore della ripetibilità.

11 Metodo della massima verosimiglianza; minimi quadrati

In questo paragrafo vogliamo discutere un metodo che permette di ottenere l'espressione della migliore funzione di stima di un parametro della distribuzione di una variabile casuale x . Il metodo che si applica è quello cosiddetto della *massima verosimiglianza*: si assume, *a posteriori*, che l'insieme $\{x_i\}_{i=1,N}$ delle misure sia quello *a priori* più probabile. Mostriamo come da questo principio, nel caso di distribuzioni di probabilità gaussiane, si possa dedurre che la miglior stima della media della distribuzione della variabile casuale x sia la media delle misure. Ricordiamo che, essendo ciascuna misura indipendente dalle altre, la densità di probabilità calcolata per i valori delle misure può essere scritta come il prodotto delle singole densità di probabilità. Se le variabili x_i hanno tutte la stessa distribuzione di probabilità normale si ha:

$$\Phi(x_1, x_2, x_3, \dots, x_N) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^N \exp \left[- \sum_{i=1}^N \frac{(x_i - \xi)^2}{2\sigma^2} \right].$$

Affinché al variare di ξ la probabilità sia massima occorre che sia minima la quantità

$$\sum_{i=1}^N (x_i - m)^2.$$

Per questo motivo, nel caso in cui le distribuzioni siano gaussiane, il metodo è detto anche dei *minimi quadrati*. Come già nel paragrafo precedente, si indica qui con m la stima di ξ ottenuta dai valori sperimentali. Per trovare la migliore stima di ξ facciamo quindi la derivata dell'espressione precedente rispetto a m e uguagliamo a zero il risultato:

$$\begin{aligned} 0 &= - \sum_{i=1}^N 2(x_i - m) \\ m &= \frac{1}{N} \sum_{i=1}^N x_i = \bar{x}. \end{aligned}$$

In maniera analoga, nel caso in cui i valori delle singole misure abbiano lo stesso valore atteso ma varianze diverse σ_i^2 (s_i^2) la migliore stima di ξ è ottenuta massimizzando il valore della densità di probabilità calcolato in corrispondenza dell'intero insieme di misure

$$\Phi(x_1, x_2, x_3, \dots, x_N) = \left(\prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_i} \right) \exp \left[- \sum_{i=1}^N \frac{(x_i - \xi)^2}{2\sigma_i^2} \right].$$

Derivando l'esponente rispetto a ξ (m) si ottiene

$$0 = \sum_{i=1}^N \frac{x_i - m}{s_i^2}$$

$$m = \frac{\sum_{i=1}^N x_i/s_i^2}{\sum_{j=1}^N 1/s_j^2} \equiv \bar{x}.$$

Scrivendo

$$a_i = \frac{1/s_i^2}{\sum_{j=1}^N 1/s_j^2}$$

e notando che $\sum_{i=1}^N a_i = 1$, per la varianza $\sigma_{\bar{x}}^2$ si ottiene

$$\sigma_{\bar{x}}^2 \equiv \langle (\bar{x} - \xi)^2 \rangle = \left\langle \left(\sum_{i=1}^N a_i x_i - \xi \right)^2 \right\rangle = \sum_{i=1}^N a_i^2 \langle (x_i - \xi)^2 \rangle = \frac{1}{\sum_{j=1}^N 1/s_j^2}.$$

Queste espressioni descrivono la *media pesata* delle misure.

Consideriamo ora una grandezza y che vari al variare di x secondo una legge del tipo

$$y = g(x, a_1, a_2, \dots)$$

dove le costanti a_1, a_2 , etc. sono parametri lasciati liberi dalla teoria. Se si eseguono N determinazioni della grandezza y in corrispondenza agli N valori x_i , e se le variabili x e y hanno entrambe distribuzione gaussiana, i parametri a_1, a_2 , etc. possono essere stimati imponendo che sia minima rispetto ad essi la quantità

$$S^2 = \sum_{i=1}^N \left(\frac{x_i - \xi_i}{\sigma_{x_i}} \right)^2 + \sum_{i=1}^N \left(\frac{y_i - \eta_i}{\sigma_{y_i}} \right)^2$$

dove

$$\eta_i \equiv \langle y_i \rangle = g(\langle x_i \rangle, a_1, a_2, \dots).$$

Generalmente, al fine di non appesantire troppo i calcoli, si assume che le x_i siano note senza incertezza. (In questo senso le x_i sono considerate variabili deterministiche e non casuali.) Le condizioni di minimo si hanno quando valgono simultaneamente le equazioni

$$\frac{\partial S^2}{\partial a_j} = - \sum_{i=1}^N 2 \frac{y_i - g(x_i, a_1, a_2, \dots)}{s_{y_i}^2} \frac{\partial g}{\partial a_j} = 0$$

Queste equazioni permettono di stimare i parametri di un'ampia varietà di funzioni; esse sono però più facilmente risolvibili se la funzione g è lineare; nel seguito ci limitiamo a discutere questo caso:

$$y = g(x, a_1, a_2, \dots) = ax + b$$

In questo caso le equazioni scritte sopra diventano

$$\begin{aligned}\frac{\partial S^2}{\partial a} &= -2 \sum_{i=1}^N \frac{y_i - ax_i - b}{s_{y_i}^2} x_i = 0 \\ \frac{\partial S^2}{\partial b} &= -2 \sum_{i=1}^N \frac{y_i - ax_i - b}{s_{y_i}^2} = 0\end{aligned}$$

Da cui

$$\begin{aligned}a &= \frac{\begin{vmatrix} \sum x_i y_i / s_{y_i}^2 & \sum x_i / s_{y_i}^2 \\ \sum y_i / s_{y_i}^2 & \sum 1 / s_{y_i}^2 \end{vmatrix}}{\begin{vmatrix} \sum x_i^2 / s_{y_i}^2 & \sum x_i / s_{y_i}^2 \\ \sum x_i / s_{y_i}^2 & \sum 1 / s_{y_i}^2 \end{vmatrix}} = \frac{\sum \frac{x_i y_i}{s_{y_i}^2} \sum \frac{1}{s_{y_i}^2} - \sum \frac{x_i}{s_{y_i}^2} \sum \frac{y_i}{s_{y_i}^2}}{\sum \frac{x_i^2}{s_{y_i}^2} \sum \frac{1}{s_{y_i}^2} - \left(\sum \frac{x_i}{s_{y_i}^2} \right)^2} \\ b &= \frac{\begin{vmatrix} \sum x_i^2 / s_{y_i}^2 & \sum x_i y_i / s_{y_i}^2 \\ \sum x_i / s_{y_i}^2 & \sum y_i / s_{y_i}^2 \end{vmatrix}}{\begin{vmatrix} \sum x_i^2 / s_{y_i}^2 & \sum x_i / s_{y_i}^2 \\ \sum x_i / s_{y_i}^2 & \sum 1 / s_{y_i}^2 \end{vmatrix}} = \frac{\sum \frac{x_i^2}{s_{y_i}^2} \sum \frac{y_i}{s_{y_i}^2} - \sum \frac{x_i}{s_{y_i}^2} \sum \frac{x_i y_i}{s_{y_i}^2}}{\sum \frac{x_i^2}{s_{y_i}^2} \sum \frac{1}{s_{y_i}^2} - \left(\sum \frac{x_i}{s_{y_i}^2} \right)^2}\end{aligned}$$

L'incertezza sui coefficienti può essere calcolata usando la (4) a pag. 25:

$$\begin{aligned}s_a^2 &= \sum \left(\frac{\partial a}{\partial y_j} \right)^2 s_{y_j}^2 = \frac{\sum 1 / s_{y_i}^2}{\sum \frac{x_i^2}{s_{y_i}^2} \sum \frac{1}{s_{y_i}^2} - \left(\sum \frac{x_i}{s_{y_i}^2} \right)^2} \\ s_b^2 &= \sum \left(\frac{\partial b}{\partial y_j} \right)^2 s_{y_j}^2 = \frac{\sum x_i^2 / s_{y_i}^2}{\sum \frac{x_i^2}{s_{y_i}^2} \sum \frac{1}{s_{y_i}^2} - \left(\sum \frac{x_i}{s_{y_i}^2} \right)^2}\end{aligned}$$

Nel caso in cui le σ_{y_i} siano tutte uguali

$$\begin{aligned}a &= \frac{N \sum x_i y_i - \sum x_i \sum y_i}{N \sum x_i^2 - (\sum x_i)^2}, & b &= \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{N \sum x_i^2 - (\sum x_i)^2} \\ s_a^2 &= s_y^2 \frac{N}{N \sum x_i^2 - (\sum x_i)^2}, & s_b^2 &= s_y^2 \frac{\sum x_i^2}{N \sum x_i^2 - (\sum x_i)^2}\end{aligned}$$

Occorre sottolineare che il metodo dei minimi quadrati nulla dice riguardo all'attendibilità della relazione funzionale ipotizzata tra le variabili x e y ; altri sono i metodi di analisi che si applicano a questo caso, per una trattazione dei quali rimandiamo ai testi di Statistica e Probabilità.

Esercizio. Si calcolino i coefficienti della migliore retta che passa per i punti dati; si assuma $s_y = 0.85$

x_i	y_i	x_i^2	$x_i y_i$
1	0.47	1	0.47
2	1.29	4	2.58
3	3.26	9	9.78
4	2.27	16	9.08
5	2.68	25	13.40
6	2.39	36	14.34
7	4.42	49	30.94
8	5.73	64	45.84
9	5.24	81	47.16
10	4.59	100	45.90
11	4.72	121	51.92
12	5.57	144	66.84
$\sum x_i = 78$	$\sum y_i = 42.63$	$\sum x_i^2 = 650$	$\sum x_i y_i = 338.25$

Calcoliamo prima la quantità

$$\Delta^2 = N \sum x_i^2 - \left(\sum x_i \right)^2 = 12 \times 650 - 78^2 = 1716$$

Si ha quindi

$$a = \frac{12 \times 338.25 - 78 \times 42.63}{1716} = 0.428, \quad s_a = \sqrt{0.85^2 \frac{12}{1716}} = 0.071$$

$$b = \frac{650 \times 42.63 - 78 \times 338.25}{1716} = 0.77, \quad s_b = \sqrt{0.85^2 \frac{650}{1716}} = 0.52$$

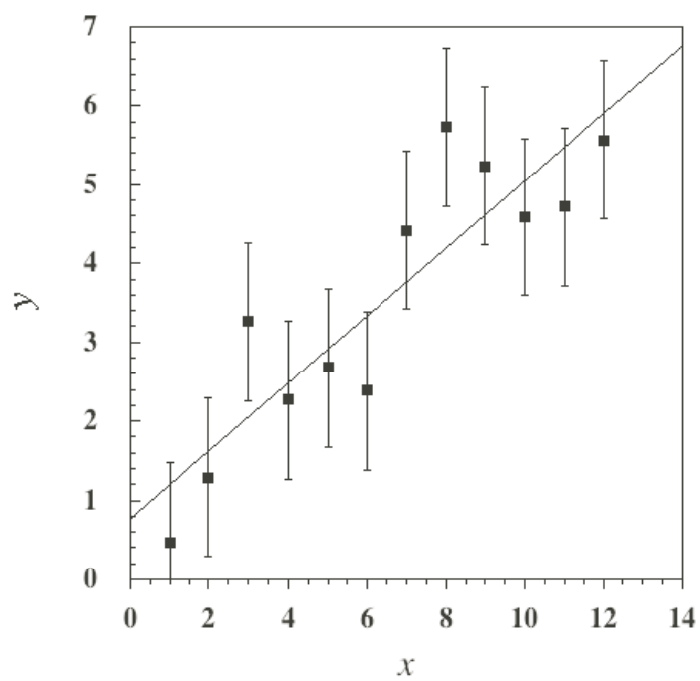


Figura 8: Applicazione del metodo dei minimi quadrati alla determinazione della migliore retta passante per i punti sperimentali, rappresentati con la barra d'incertezza delle ordinate.