3.3.3 Probabilità condizionata e probabilità composta

La probabilità che si verifichi l'evento E nel caso in cui si sa già che si è verificato l'evento F si indica con il simbolo $p(E|F)$ e si chiama *probabilità condizionata*: si ricava per essa facilmente, usando la terminologia dell'esempio precedente, l'identità

$$f(E|F) = \frac{n_{11}}{n_{11}+n_{21}} = \frac{n_{11}}{N}\frac{N}{n_{11}+n_{21}} = \frac{f(EF)}{f(F)}$$

con l'analoga

$$f(F|E) = \frac{f(EF)}{f(E)} \ ;$$

e vale quindi, passando al limite, la

$$p(EF) = p(F)\cdot p(E|F) = p(E)\cdot p(F|E) \ . \tag{3.3}$$

Nel caso particolare di due eventi casuali tali che il verificarsi o meno dell'uno non alteri la probabilità di presentarsi dell'altro, ovverosia per cui risulti $p(E|F) = p(E)$ e $p(F|E) = p(F)$, questi si dicono tra loro *statisticamente indipendenti*[3]; e per essi vale la seguente legge (della *probabilità composta*):

$$p(EF) = p(E)\cdot p(F) \ .$$

Questa si generalizza facilmente (sempre per induzione completa) ad un evento complesso costituito dal verificarsi contemporaneo di un numero qualsiasi di eventi semplici (sempre però tutti statisticamente indipendenti tra loro); per il quale vale la

$$p(A\cdot B\cdots Z) = p(A)\cdot p(B)\cdots p(Z) \ . \tag{3.4}$$

Più in particolare, gli eventi casuali appartenenti ad un insieme di dimensione N (con $N > 2$) si dicono *tutti* statisticamente indipendenti tra loro quando la probabilità del verificarsi di uno qualsiasi di essi non è alterata dal fatto che uno *o più d'uno* degli altri si sia già presentato.

Come esempio si consideri il lancio indipendente di due dadi, ed i seguenti tre eventi casuali: A, consistente nell'uscita di un numero dispari sul primo dado; B, consistente nell'uscita di un numero dispari sul secondo dado; e C, consistente nell'uscita di un punteggio complessivo dispari. È facile

[3] Il concetto di indipendenza statistica tra eventi casuali fu definito per la prima volta nel 1718 da Abraham de Moivre (purtroppo noto al grosso pubblico solo per aver correttamente predetto il giorno della propria morte servendosi di una formula matematica), nel suo libro "The Doctrine of Chance".

vedere che questi eventi casuali sono, se considerati a due a due, statisticamente indipendenti: A e B per ipotesi, A e C perché $p(C|A) = \frac{1}{2} = p(C)$, ed infine B e C perché anche $p(C|B) = \frac{1}{2} = p(C)$; ma gli stessi tre eventi, se vengono considerati nel loro complesso, *non* sono *tutti* statisticamente indipendenti — perché il verificarsi di A assieme a B rende poi impossibile il verificarsi di C.

3.3.4 Il teorema di Bayes

Supponiamo che un dato fenomeno casuale A possa dare luogo a N eventualità mutuamente esclusive A_j, che esauriscano inoltre la totalità delle possibilità; e sia poi un differente fenomeno casuale che possa condurre o al verificarsi o al non verificarsi di un evento E. Osservando la realizzazione di entrambi questi fenomeni, se E si verifica, assieme ad esso si dovrà verificare anche una ed una sola delle eventualità A_j; applicando prima la legge della probabilità totale (3.2) e poi l'equazione (3.3), si ottiene

$$p(E) = \sum_{j=1}^{N} p(E\cdot A_j) = \sum_{j=1}^{N} p(A_j)\cdot p(E|A_j) \ . \tag{3.5}$$

Ora, riprendendo la legge fondamentale delle probabilità condizionate (3.3), ne ricaviamo

$$p(A_i|E) = \frac{p(A_i)\cdot p(E|A_i)}{p(E)}$$

e, sostituendovi la (3.5), si giunge alla

$$\boxed{p(A_i|E) = \frac{p(A_i)\cdot p(E|A_i)}{\sum_j \left[p(A_j)\cdot p(E|A_j)\right]}} \tag{3.6}$$

L'equazione (3.6) è nota con il nome di *teorema di Bayes*, e viene spesso usata nel calcolo delle probabilità; talvolta anche, come adesso vedremo, quando le A_j non siano tanto eventi casuali in senso stretto, quanto piuttosto *ipotesi* da discutere per capire se esse siano o meno rispondenti alla realtà.

Facendo un esempio concreto, si abbiano due monete: una "buona", che presenti come risultato la testa e la croce con uguale probabilità (dunque pari a 0.5); ed una "cattiva", con due teste sulle due facce. Inizialmente si sceglie una delle due monete; quindi avremo due eventualità mutuamente esclusive: A_1 (è stata scelta la moneta "buona") e A_2 (è stata scelta la moneta "cattiva") con probabilità rispettive $p(A_1) = p(A_2) = 0.5$. Se l'evento casuale E consiste nell'uscita di una testa, ovviamente $p(E|A_1) = 0.5$ e $P(E|A_2) = 1$.

Se ora facciamo un esperimento, lanciando la moneta una volta e ottenendo una testa, quale è la probabilità che nell'effettuare la scelta iniziale si sia presa quella "buona"? La risposta è data dal teorema di Bayes, da cui si ottiene:

$$\begin{aligned} p(A_1|E) &= \frac{p(A_1)\cdot p(E|A_1)}{p(A_1)\cdot p(E|A_1)+p(A_2)\cdot p(E|A_2)} \\ &= \frac{0.5\cdot 0.5}{0.5\cdot 0.5+0.5\cdot 1} \\ &= \frac{0.25}{0.75} \\ &= \frac{1}{3}\ . \end{aligned}$$

Ovviamente, se si volesse progettare un esperimento reale, sarebbe meglio associarlo al lanciare la moneta N volte (con $N > 1$): o si ottiene almeno una croce, ed allora è sicuramente vera A_1; o, invece, si presenta l'evento E consistente nell'ottenere N teste in N lanci. In quest'ultimo caso, $p(E|A_2) = 1$ e $p(E|A_1) = 1/2^N$ se i lanci sono indipendenti tra loro; utilizzando ancora l'equazione (3.6), si ricava che la probabilità di aver scelto la moneta "buona", $p(A_1)$, è data da $1/(1+2^N)$ — e di conseguenza $p(A_2) = 2^N/(1+2^N)$ è la probabilità che si sia scelta la moneta "cattiva".

Qui il teorema di Bayes viene utilizzato per *verificare una ipotesi statistica*: ovvero per calcolare la probabilità che l'una o l'altra di un insieme di condizioni A_j che si escludono a vicenda sia vera, sulla base di osservazioni sperimentali riassunte dal verificarsi di E; ma questo ci risulta possibile solo perché si conoscono *a priori* le probabilità di *tutte* le condizioni stesse $p(A_j)$.

Se, viceversa, queste non sono note, la (3.6) ci dà ancora la probabilità che sia vera l'una o l'altra delle ipotesi A_j se sappiamo che si è verificata la condizione sperimentale E; ma essa non si può ovviamente calcolare, a meno di fare opportune ipotesi sui valori delle $p(A_j)$: ad esempio assumendole tutte uguali, il che è chiaramente arbitrario. Per essere più specifici, non potremmo servirci di un esperimento analogo a quelli delineati e del teorema di Bayes per calcolare la probabilità che una particolare moneta da 1 euro ricevuta in resto sia o non sia "buona": a meno di non conoscere a priori $p(A_1)$ e $p(A_2)$, le probabilità che una moneta da 1 euro scelta a caso tra tutte quelle circolanti nella nostra zona sia "buona" o "cattiva".

3.4 Definizione assiomatica della probabilità

Per completezza, accenniamo infine alla cosiddetta *definizione assiomatica della probabilità*[4], che è matematicamente consistente:

> *Sia S l'insieme di tutti i possibili risultati di un fenomeno casuale, ed E un qualsiasi evento casuale definito su S (ossia un qualsiasi sottoinsieme $E \subseteq S$). Si definisce come "probabilità" di E un numero, $p(E)$, associato univocamente all'evento stesso, che soddisfi alle seguenti tre proprietà:*
>
> 1. $p(E) \geq 0$ *per ogni* E;
> 2. $p(S) = 1$;
> 3. $p(E_1 \cup E_2 \cup \cdots) = p(E_1) + p(E_2) + \cdots$ *per qualsiasi insieme di eventi* $E_1, E_2, \ldots$, *in numero finito od infinito e a due a due senza alcun elemento in comune (ossia tali che* $E_i \cap E_j = \emptyset$ *per ogni* $i \neq j$*).*

Questa definizione, pur matematicamente consistente[5], non dice nulla su come assegnare dei valori alla probabilità; tuttavia su tali valori si possono fare delle ipotesi, verificabili poi analizzando gli eventi reali osservati.

3.4.1 Le leggi della probabilità e la definizione assiomatica

Dalla definizione assiomatica è possibile ricavare, come abbiamo già prima accennato, le stesse leggi cui siamo giunti a partire dalla definizione empirica. Infatti:

- Essendo $S \cup \emptyset = S$, la proprietà 3 (applicabile perché $S \cap \emptyset = \emptyset$) implica $p(S) + p(\emptyset) = p(S)$; da cui ricaviamo, vista la proprietà 2,
$$p(\emptyset) = 0\ .$$
- Se $A \supset B$, essendo in questo caso $A = B \cup (A \cap \overline{B})$, applicando la proprietà 3 (il che è lecito dato che $B \cap (A \cap \overline{B}) = \emptyset$) si ottiene $p(A) = p(B) + p(A \cap \overline{B})$; e, vista la proprietà 1,
$$A \supset B \quad \Rightarrow \quad p(A) \geq p(B)\ .$$

[4] Questa definizione è dovuta all'eminente matematico russo Andrei Nikolaevich Kolmogorov; vissuto dal 1903 al 1987, si occupò principalmente di statistica e di topologia. Fu enunciata nel suo libro del 1933 *Grundbegriffe der Wahrscheinlichkeitsrechnung*.

[5] Volendo essere del tutto rigorosi, questa definizione risulta valida solo se l'insieme dei possibili risultati è composto da un numero finito o da un'infinità numerabile di elementi; la reale definizione assiomatica della probabilità è leggermente differente (ed ancora più astratta).

- Dati due insiemi A e B, visto che qualunque essi siano valgono le seguenti identità:

$$A = (A \cap B) \cup (A \cap \overline{B})$$
$$B = (A \cap B) \cup (\overline{A} \cap B)$$
$$(A \cup B) = (A \cap B) \cup (A \cap \overline{B}) \cup (\overline{A} \cap B)$$

  e applicando a queste tre relazioni (dopo aver verificato che gli insiemi a secondo membro sono tutti disgiunti) la proprietà 3 e sommando e sottraendo opportunamente i risultati, si ottiene la *legge della probabilità totale* nella sua forma più generale:

$$p(A \cup B) = p(A) + p(B) - p(A \cap B) \ .$$

Definendo poi $p(E|A)$ (con $p(A) \neq 0$) come

$$p(E|A) = \frac{p(E \cap A)}{p(A)} \ , \tag{3.7}$$

è facile riconoscere che anche essa rappresenta una probabilità: essendo $p(E \cap A) \geq 0$ e $p(A) > 0$, $p(E|A)$ soddisfa alla proprietà 1; essendo $S \cap A = A$, $p(S|A) = p(A)/p(A) = 1$, e $p(E|A)$ soddisfa alla proprietà 2; infine, se $E_1, E_2, \ldots$ sono insiemi a due a due disgiunti,

$$\begin{aligned} p(E_1 \cup E_2 \cup \cdots | A) &= \frac{p[(E_1 \cup E_2 \cup \cdots) \cap A]}{p(A)} \\ &= \frac{p[(E_1 \cap A) \cup (E_2 \cap A) \cup \cdots]}{p(A)} \\ &= \frac{p(E_1 \cap A)}{p(A)} + \frac{p(E_2 \cap A)}{p(A)} + \cdots \\ &= p(E_1|A) + p(E_2|A) + \cdots \end{aligned}$$

e $p(E|A)$ soddisfa anche alla proprietà 3. Dalla (3.7) si ottiene infine la *legge della probabilità composta* nella sua forma più generale,

$$p(A \cap B) \ = \ p(A|B) \cdot p(B) \ = \ p(B|A) \cdot p(A) \ .$$

3.5 La convergenza statistica

Difetto della definizione empirica di probabilità, oltre a quello di essere basata su di un esperimento, è quello di presupporre a priori una convergenza della frequenza relativa f, al crescere di N, verso un valore ben definito: valore che si assume poi come probabilità dell'evento.

Qualora si assuma come definizione di probabilità quella assiomatica, è effettivamente possibile dimostrare (come vedremo più avanti nel paragrafo 5.6, ed in particolare nel sottoparagrafo 5.6.3) come, al crescere del numero di prove, la frequenza relativa di un *qualunque* evento casuale converga verso la probabilità dell'evento stesso.

È tuttavia assai importante sottolineare come questa legge (*legge dei grandi numeri*, o *teorema di Bernoulli*) non implichi una convergenza esatta nel senso dell'analisi: non implichi cioè che, scelto un qualunque numero positivo ϵ, sia possibile determinare in conseguenza un intero M tale che, se si effettuano N prove, per ogni $N > M$ risulti *sicuramente* $|f(E) - p(E)| < \epsilon$. Si pensi in proposito alla chiara impossibilità di fissare un numero M tale che, quando si lanci un dado più di M volte, si sia *certi* di ottenere almeno un sei: al crescere di M crescerà la *probabilità* del verificarsi di questo evento, ma non si potrà mai raggiungere la certezza.

Nella legge dei grandi numeri il concetto di convergenza va inteso invece in senso *statistico* (o *debole*, o *stocastico*); si dice che all'aumentare del numero di prove N una grandezza x tende statisticamente al limite X quando, scelta una qualsiasi coppia di numeri positivi ϵ e δ, si può in conseguenza determinare un numero intero M tale che, se si effettua un numero di prove N maggiore di M, la probabilità che x differisca da X per più di ϵ risulti minore di δ. Indicando col simbolo $\Pr(E)$ la probabilità di un evento E, la definizione di convergenza statistica è

$$\forall \epsilon, \delta > 0 \quad \rightarrow \quad \exists M : \quad N > M \ \Rightarrow \ \Pr\Big(|x - X| \geq \epsilon\Big) \leq \delta \ . \tag{3.8}$$

Nel paragrafo 5.6 vedremo che, dato un qualunque evento casuale E avente probabilità $\Pr(E)$ di manifestarsi, si può dimostrare che la sua frequenza relativa $f(E)$ su N prove converge statisticamente a $\Pr(E)$ all'aumentare di N; o, in altre parole, come aumentando il numero di prove si possa rendere tanto improbabile quanto si vuole che la frequenza relativa e la probabilità di un qualunque evento casuale E differiscano più di una quantità prefissata.

Capitolo 4

Elaborazione dei dati

In questo capitolo si discute dell'organizzazione da dare ai dati sperimentali, e su come si possano da essi ricavare quantità significative.

4.1 Istogrammi

Una volta che si disponga di un insieme di più misure della stessa grandezza fisica (nella statistica si parla in genere di un *campione* di misure), è opportuno cercare di organizzarle in modo che il loro significato risulti a colpo d'occhio evidente; la maniera più consueta di rappresentare graficamente le misure è quella di disporle in un *istogramma*.

Essendovi una corrispondenza biunivoca tra i numeri reali ed i punti di una retta orientata, ognuna delle nostre misure può essere rappresentata su di essa da un punto; l'istogramma è un particolare tipo di diagramma cartesiano in cui l'asse delle ascisse è dedicato a tale rappresentazione. Tuttavia è facile rendersi conto del fatto che non tutti i valori della variabile sono in realtà permessi, perché gli strumenti forniscono per loro natura un insieme discreto di valori essendo limitati ad un numero finito di cifre significative.

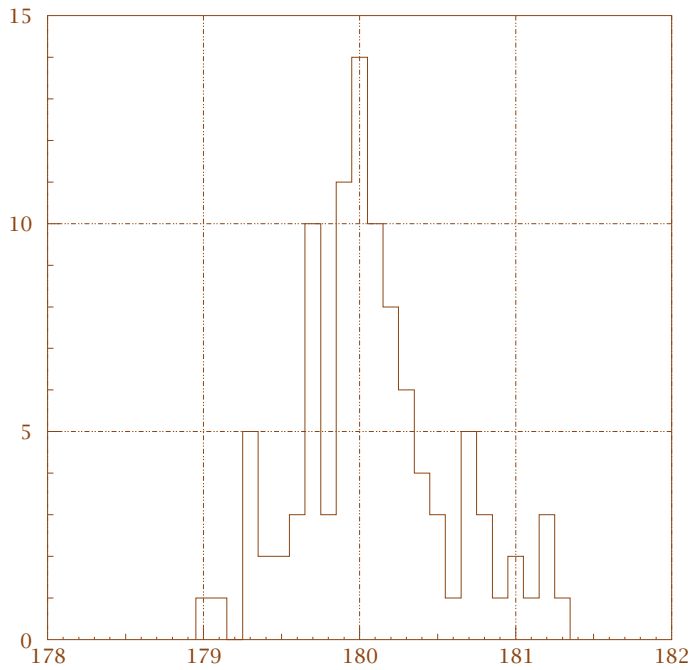
Conviene allora mettere in evidenza sull'asse delle ascisse tutti i possibili valori che possono essere ottenuti da una misura reale; cioè punti separati da un intervallo che corrisponde alla cifra significativa più bassa dello strumento, o comunque alla più piccola differenza apprezzabile con esso se l'ultima cifra deve essere stimata dall'osservatore (ad esempio il decimo di grado stimato ad occhio su un goniometro avente scala al mezzo grado).

Nelle ordinate del diagramma si rappresenta poi la frequenza assoluta

con la quale i diversi valori si sono presentati; questo si fa associando ad ognuna delle misure un rettangolo avente area unitaria, che viene riportato con la base al di sopra dell'intervallo appropriato ogni volta che uno dei possibili valori è stato ottenuto.

Nel caso consueto in cui l'asse delle ascisse venga diviso in intervalli aventi tutti la stessa ampiezza, tutti questi rettangoli avranno ovviamente la stessa altezza: di modo che è possibile, dall'altezza di una colonna di rettangoli unitari sovrapposti, risalire al numero di dati del campione aventi un determinato valore.

FIGURA 4a - Esempio di istogramma (100 misure ripetute della somma degli angoli interni di un triangolo).



Se le frequenze assolute risultassero troppo piccole, può essere opportuno raggruppare le misure in *classi di frequenza*; ciascuna classe corrispondendo ad un intervallo multiplo opportuno del più piccolo rappresentabile discusso sopra.

Anziché costruire l'istogramma riportandovi un risultato per volta, si possono contare prima le frequenze in ciascuna classe e disegnare sopra ognuna di esse un rettangolo avente area corrispondente alla frequenza ivi osservata. L'area dell'istogramma sopra ad un qualsiasi intervallo è proporzionale alla frequenza assoluta con cui si è osservato un valore che cade entro di esso; uguale, se si assume come unità di misura per le aree quella del rettangolo di altezza unitaria. L'area totale sottesa dall'istogramma è, sempre rispetto a tale unità, pari al numero di osservazioni N.

Un'altra rappresentazione, che è poco usata ma vantaggiosa perché non richiede la previa (e in qualche misura arbitraria) definizione delle classi di frequenza, è quella della *frequenza cumulativa*, assoluta o relativa. Essa è definita, per ogni valore dell'ascissa x, dal numero (assoluto o relativo) di volte per cui il risultato della misura è stato minore o uguale a x: si tratta dunque di una funzione monotona non decrescente con uno scalino pari rispettivamente ad 1 o a $1/N$ in corrispondenza di ognuno degli N valori osservati. Risulta inoltre

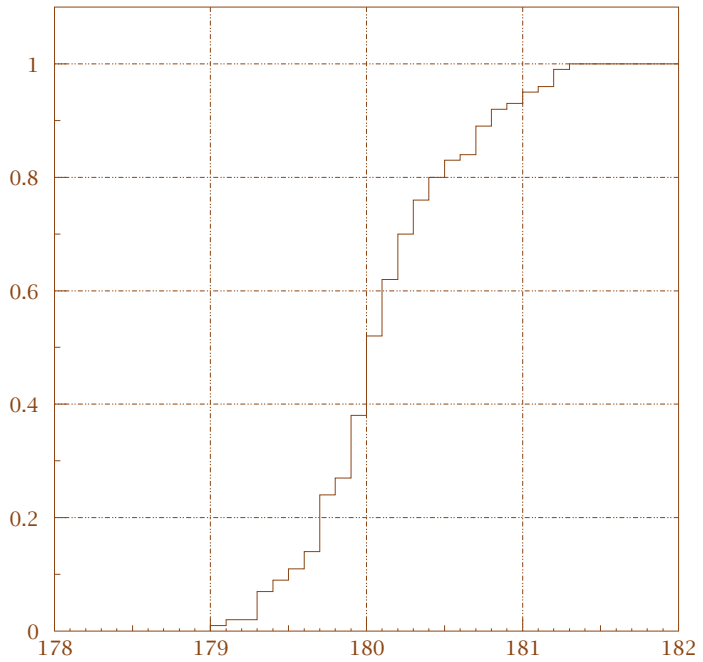
$$0 = F(-\infty) \leq F(x) \leq F(+\infty) = \begin{cases} N & \text{(ass.)} \\ 1 & \text{(rel.)} \end{cases}$$

4.2 Stime di tendenza centrale

In presenza di N valori osservati di una grandezza fisica (che non siano tutti coincidenti), si pone il problema di definire un algoritmo che fornisca la stima migliore del valore vero della grandezza osservata; cioè di determinare quale, tra le infinite funzioni dei dati, ha la maggiore probabilità di darci il valore vero.

Ora, se supponiamo di avere eliminato tutti gli errori sistematici, è intuitivo come il valore di tale stima debba corrispondere ad una ascissa in posizione centrale rispetto alla distribuzione dei valori osservati; sappiamo infatti che gli errori casuali hanno uguale probabilità di presentarsi in difetto ed in eccesso rispetto al valore vero e, *se il numero di misure è sufficientemente elevato*, ci aspettiamo (sulla base della legge dei grandi numeri) che la distribuzione effettiva delle frequenze non si discosti troppo da quella teorica delle probabilità. Dunque ci si attende che i valori osservati si distribuiscano simmetricamente rispetto al valore vero.

FIGURA 4b - Frequenza cumulativa relativa per le stesse misure della figura 4a.



4.2.1 La moda

Nella statistica esistono varie stime della cosiddetta *tendenza centrale* di un campione; una di queste stime è il valore corrispondente al massimo della frequenza, cioè il valore che si è presentato il maggior numero di volte (ovvero la media dei valori *contigui* che presentassero tutti la medesima massima frequenza): tale stima (se esiste) si chiama *moda* del campione, e si indica con il simbolo $\hat{x}$.

In generale però la distribuzione potrebbe non avere massimo (distribuzioni *amodali*), oppure averne più d'uno in intervalli non contigui (distribuzioni *multimodali*); anche se questo non dovrebbe essere il caso per le distribuzioni di misure ripetute. Talvolta si dice che la distribuzione non ha moda anche se il massimo esiste, ma si presenta ad uno degli estremi dell'intervallo che contiene le misure; non essendo in tal caso la moda, ovviamente, una stima di tendenza centrale.

Per tutti questi motivi la moda non è di uso molto frequente, e non è opportuna in questo contesto anche per ragioni che saranno esaminate più avanti.

4.2.2 La mediana

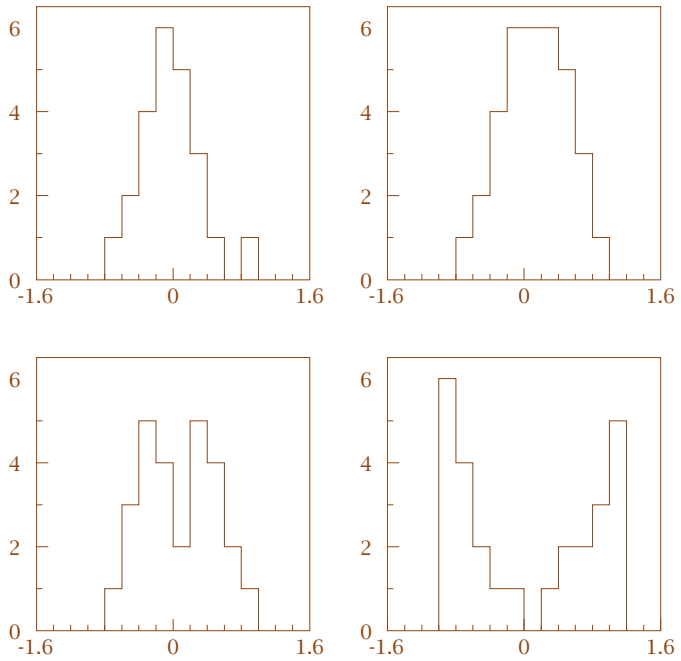
Un'altra stima di tendenza centrale di uso frequente nella statistica (anche se non nella fisica) è la *mediana* di un campione: indicata col simbolo $\tilde{x}$, è definita come quel valore che divide l'istogramma dei dati in due parti di uguale area[1]; in termini meno precisi, la mediana lascia un uguale numero di dati alla propria sinistra ed alla propria destra[2]. Usando questa forma della definizione, per trovare la mediana di un insieme di valori tutti distinti basta disporli in ordine crescente e prendere il valore centrale (per un numero dispari di misure; si prende la semisomma dei due valori centrali se le misure sono in numero pari).

Al contrario della moda, la mediana esiste sempre; nel diagramma della frequenza cumulativa relativa è definita dall'ascissa corrispondente all'ordinata del 50%. Si può dimostrare anche che la mediana $\tilde{x}$ è quel valore di x che rende minima la somma dei valori assoluti degli scarti delle nostre

[1] Il valore della mediana di un insieme di dati, così definito, dipende dalla scelta delle classi si frequenza; per questo motivo la mediana in genere non si adopera tanto per i campioni sperimentali di dati, quanto per le distribuzioni teoriche.

[2] Basta applicare le due definizioni ad un insieme di dati composto dai tre valori $\{0, 1, 1\}$ per rendersi conto della differenza.

FIGURA 4c - Due distribuzioni unimodali (in alto), una bimodale (in basso a sinistra), una senza moda (in basso a destra); quest'ultima distribuzione simula il campionamento a istanti casuali dell'elongazione di un pendolo.



misure x_i da x; cioè tale che

$$\min\left\{\sum_{i=1}^{N}|x_i - x|\right\} = \sum_{i=1}^{N}|x_i - \tilde{x}| \ .$$

4.2.3 La media aritmetica

La stima di gran lunga più usata della tendenza centrale di un campione è la *media aritmetica* $\bar{x}$ dei valori osservati, definita attraverso la

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \ . \tag{4.1}$$

Proprietà matematiche della media aritmetica sono le seguenti:

PROPRIETÀ 1: *la somma degli scarti di un insieme di valori dalla loro media aritmetica è identicamente nulla.*

Infatti dalla definizione risulta

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

ed infine

$$\sum_{i=1}^{N}(x_i - \bar{x}) \equiv 0 \ . \tag{4.2}$$

PROPRIETÀ 2: *la media aritmetica $\bar{x}$ di un insieme di dati numerici $x_1, x_2, \ldots, x_N$ è quel valore di x rispetto al quale risulta minima la somma dei quadrati degli scarti dalle x_i; cioè quel numero per il quale è verificata la*

$$\min\left\{\sum_{i=1}^{N}(x_i - x)^2\right\} = \sum_{i=1}^{N}(x_i - \bar{x})^2 \ .$$

Infatti abbiamo

$$\begin{aligned}\sum_{i=1}^{N}(x_i - x)^2 &= \sum_{i=1}^{N}\left[(x_i - \bar{x}) + (\bar{x} - x)\right]^2 \\ &= \sum_{i=1}^{N}\left[(x_i - \bar{x})^2 + (\bar{x} - x)^2 + 2(x_i - \bar{x})(\bar{x} - x)\right] \\ &= \sum_{i=1}^{N}(x_i - \bar{x})^2 \;+\; \sum_{i=1}^{N}(\bar{x} - x)^2 \;+\; 2(\bar{x} - x)\sum_{i=1}^{N}(x_i - \bar{x}) \;\;;\end{aligned}$$

da qui, sfruttando l'equazione (4.2), si ottiene

$$\sum_{i=1}^{N}(x_i - x)^2 = \sum_{i=1}^{N}(x_i - \bar{x})^2 \;+\; N(\bar{x} - x)^2 \tag{4.3}$$

e finalmente

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

4.2.4 Considerazioni complessive

Oltre le tre stime citate di tendenza centrale ne esistono altre, di uso però limitato a casi particolari e che non hanno frequente uso né nella statistica né nella fisica; per soli motivi di completezza citiamo qui:

- la *media geometrica*, g, definita come la radice N-esima del prodotto degli N valori rappresentati nel campione:

$$g^N = \prod_{i=1}^{N} x_i \;\;;$$

- la *media armonica*, h, definita come il reciproco del valore medio dei reciproci dei dati:

$$\frac{1}{h} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{x_i} \;\;;$$

- la *media quadratica*, q, definita come la radice quadrata del valore medio dei quadrati dei dati:

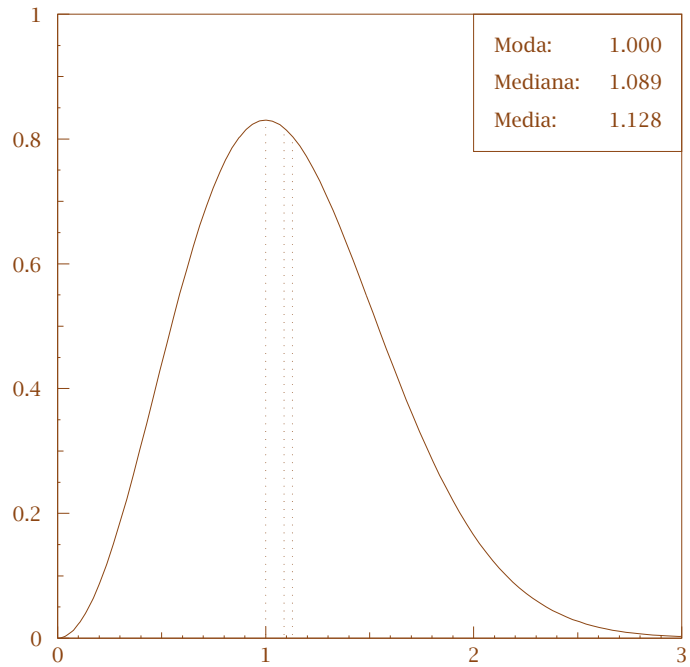
$$q = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \;\;.$$

Se la distribuzione dei dati non è troppo irregolare, le prime tre stime citate per la tendenza centrale (moda, mediana e media aritmetica) non sono molto lontane; esiste una relazione empirica che le lega e che è valida per distribuzioni non troppo asimmetriche:

$$\left|\bar{x} - \hat{x}\right| \;\approx\; 3\left|\bar{x} - \tilde{x}\right| \;\;,$$

cioè la differenza tra media aritmetica e moda è circa il triplo della differenza tra media aritmetica e mediana.

FIGURA 4d - I valori delle tre principali stime di tendenza centrale per la distribuzione di Maxwell–Boltzmann; l'unità per le ascisse è il parametro α che compare nell'equazione (4.4).



Come esempio, nella figura 4d è mostrato l'andamento di una distribuzione di probabilità per una variabile (continua) che è di interesse per la

fisica; e nel grafico sono messi in evidenza i valori per essa assunti dalle tre stime di tendenza centrale considerate. Si tratta della funzione di frequenza detta di *Maxwell–Boltzmann*, e secondo essa sono ad esempio distribuiti, in un gas perfetto, i moduli delle velocità delle molecole: l'equazione della curva è

$$y = f(v) = \frac{4}{\sqrt{\pi}}\,\alpha^{\frac{3}{2}}\,v^2\,e^{-\alpha v^2} \tag{4.4}$$

(in cui α è una costante dipendente dalla massa delle molecole e dalla temperatura del gas).

La scelta dell'uso dell'una o dell'altra stima statistica per determinare la tendenza centrale di un campione di misure ripetute andrà decisa sulla base delle proprietà delle stime stesse; più precisamente sulla base dello studio di come si distribuiscono varie stime che si riferiscano a campioni analoghi, cioè ad insiemi di misure della stessa grandezza fisica presumibilmente affette dallo stesso errore (eseguite insomma in condizioni simili) e composti da uno stesso numero di dati. La stima che sceglieremo dovrebbe essere la migliore, nel senso già usato all'inizio di questo paragrafo 4.2: quella che ha la maggiore probabilità di darci il valore vero della grandezza misurata.

4.2.5 Prima giustificazione della media

La stima di tendenza centrale che è di uso generale per le misure ripetute è la *media aritmetica*: i motivi sono svariati e sostanzialmente legati alle proprietà statistiche della media stessa; di essi ci occuperemo ancora più avanti. In particolare vedremo nel paragrafo 11.3 che la media aritmetica è effettivamente la stima *migliore*, nel senso or ora chiarito di questa frase.

A questo punto possiamo già comunque renderci conto (anche se in maniera *non rigorosa*) che la media aritmetica di più misure dovrebbe avere un errore inferiore a quello delle misure stesse; indichiamo con x^* il valore vero della grandezza x, e con x_i ($i = 1, 2, \ldots, N$) le N determinazioni sperimentali di x: l'errore assoluto commesso in ognuna delle misure x_i sarà dato da $\epsilon_i = x_i - x^*$. L'errore assoluto della media aritmetica è allora dato da

$$\bar{\epsilon} = \bar{x} - x^*$$

e, sfruttando la (4.1),

$$\bar{\epsilon} = \frac{1}{N}\sum_{i=1}^{N} x_i - x^* = \frac{1}{N}\sum_{i=1}^{N}(x_i - x^*) = \frac{1}{N}\sum_{i=1}^{N}\epsilon_i \ .$$

Se gli errori sono solo casuali, saranno ugualmente probabili in difetto e in eccesso rispetto al valore vero; e se le misure sono numerose gli ϵ_i

tenderanno quindi ad eliminarsi a vicenda nella sommatoria, che inoltre è moltiplicata per un fattore $1/N$.

4.2.6 La media aritmetica espressa tramite le frequenze

Siano x_i, con $i = 1, \ldots, N$, gli N valori del campione di cui vogliamo calcolare la media aritmetica; supponiamo che qualcuno dei valori ottenuti sia *ripetuto*, ad esempio che il valore x_1 si sia presentato n_1 volte, x_2 si sia presentato n_2 volte e così via: la media aritmetica si può calcolare come

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \cdots}{N} \qquad (N = n_1 + n_2 + \cdots) \ .$$

Indichiamo con x_j ($j = 1, 2, \ldots, M$) gli M valori distinti di x presenti nel campione; n_j è la frequenza assoluta con cui abbiamo ottenuto il valore x_j nel corso delle nostre misure, ed il rapporto n_j/N è la frequenza relativa f_j dello stesso evento casuale: allora possiamo scrivere

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1}{N}\sum_{j=1}^{M} n_j x_j = \sum_{j=1}^{M}\frac{n_j}{N}x_j = \sum_{j=1}^{M} f_j x_j \ .$$

Formule in cui si sommano valori numerici (qui gli x_j) moltiplicati ciascuno per un fattore specifico (f_j) vanno sotto il nome generico di *formule di media pesata*: ogni valore distinto dagli altri contribuisce infatti al risultato finale con un *peso* relativo dato dal numero f_j.

È bene osservare come si possano definire infinite medie pesate dei valori numerici x_j, corrispondenti alle infinite differenti maniere di attribuire ad ognuno di essi un peso; ed anche che, in genere, con il nome di "media pesata" ci si riferisce a quella particolare formula che permette di calcolare la migliore stima del valore vero di una grandezza fisica sulla base di più misure aventi differente precisione (l'equazione (11.7), che incontreremo più avanti nel paragrafo 11.3), e non alla formula precedente.

Fin qui tale formula si presenta solo come un artificio per calcolare la media aritmetica di un insieme di valori risparmiando alcune operazioni; ma pensiamo di far tendere all'infinito il numero di misure effettuate. In tal caso, se assumiamo che la frequenza relativa con cui ogni valore si è presentato tenda stocasticamente alla probabilità rispettiva, in definitiva otteniamo che la media aritmetica delle misure deve anch'essa tendere ad un limite determinato:

$$\lim_{N\to\infty}\bar{x} = \sum_j p_j x_j \ .$$

In definitiva, se siamo in grado di assegnare in qualche modo una probabilità al presentarsi di ognuno dei possibili valori di una misura, siamo anche

in grado di calcolare il valore assunto dalla media aritmetica di un campione di quei valori nel limite di infinite misure effettuate. Di questa formula ci serviremo più avanti, una volta ricavata appunto (sotto opportune ipotesi) la probabilità di ottenere un certo risultato dalle misure di una grandezza fisica.

4.3 Stime di dispersione

Abbiamo sviluppato il paragrafo 4.2 partendo dall'intuizione (giustificata con l'aiuto delle caratteristiche degli errori casuali e della legge dei grandi numeri) che la tendenza centrale di un insieme di misure è legata al valore vero della grandezza misurata. Così, similmente, si intuisce che agli errori introdotti nell'eseguire le nostre misure è legata un'altra grandezza caratteristica del campione, cioè la sua *dispersione*: ovvero la valutazione della larghezza dell'intervallo in x in cui le misure stesse sono distribuite attorno al valore centrale.

4.3.1 Semidispersione massima e quantili

La più grossolana delle stime statistiche di dispersione si effettua trovando il massimo ed il minimo valore osservato: la *semidispersione massima* è definita come la semidifferenza tra questi due valori,

$$\frac{x_{\max} - x_{\min}}{2} \quad .$$

Essa ha il difetto di ignorare la maggior parte dei dati e particolarmente quelli, generalmente preponderanti, prossimi al centro della distribuzione; inoltre normalmente aumenta all'aumentare del numero di misure, invece di tendere ad un valore determinato. Il doppio della semidispersione massima

$$R = x_{\max} - x_{\min}$$

è anch'esso usato come stima della dispersione di un campione, e viene chiamato *range*.

Grandezze frequentemente usate per caratterizzare una distribuzione nella statistica (non nella fisica) sono i *quartili*, i *decili* ed i *percentili* (collettivamente *quantili*), indicati con Q_i ($i = 1, 2, 3$); con D_i ($i = 1, \ldots, 9$); e con P_i ($i = 1, \ldots, 99$) rispettivamente. Essi sono definiti (analogamente alla mediana) come quei valori della x che dividono la distribuzione rispettivamente in 4, 10 e 100 parti di uguale area; ovviamente vale la

$$Q_2 \equiv D_5 \equiv P_{50} \equiv \tilde{x} \quad .$$

Come stima della dispersione di una distribuzione è usato dagli statistici l'*intervallo semiinterquartilico* $Q = (Q_3 - Q_1)/2$, come pure la differenza $P_{90} - P_{10}$ tra il novantesimo ed il decimo percentile; tali intervalli esistono sempre, ma non sono padroneggiabili agevolmente negli sviluppi teorici.

4.3.2 Deviazione media assoluta (errore medio)

Altra stima di dispersione è la *deviazione media assoluta* (o *errore medio*), definita come

$$\overline{|x - \bar{x}|} = \frac{1}{N} \sum_{i=1}^{N} |x_i - \bar{x}| \quad ,$$

oppure, meno frequentemente, come

$$\overline{|x - \tilde{x}|} = \frac{1}{N} \sum_{i=1}^{N} |x_i - \tilde{x}| \quad ;$$

ma anch'essa non è facile da trattare a ragione della operazione non lineare costituita dal valore assoluto.

4.3.3 Varianza e deviazione standard

La più importante (e più usata, non solo in fisica) stima di dispersione è la *deviazione standard* (oppure *scarto* o *deviazione quadratica media*) s; che si definisce come la radice quadrata della *varianza*, s^2:

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

Per distribuzioni non troppo asimmetriche la deviazione media assoluta è circa i $\frac{4}{5}$ della deviazione standard, e l'intervallo semiinterquartilico è circa i $\frac{2}{3}$ della stessa.

Per calcolare lo scarto quadratico medio di un campione senza l'aiuto di un calcolatore appositamente programmato, può risultare utile sfruttare la sua seguente proprietà:

$$
\begin{aligned}
N\,s^2 &= \sum_{i=1}^{N} (x_i - \bar{x})^2 \\
&= \sum_{i=1}^{N} ({x_i}^2 + \bar{x}^2 - 2\,\bar{x}\,x_i) \\
&= \sum_{i=1}^{N} {x_i}^2 \;+\; N\,\bar{x}^2 \;-\; 2\,\bar{x} \sum_{i=1}^{N} x_i \\
&= \sum_{i=1}^{N} {x_i}^2 \;+\; N\,\bar{x}^2 \;-\; 2\,N\,\bar{x}^2 \\
&= \sum_{i=1}^{N} {x_i}^2 \;-\; N\,\bar{x}^2
\end{aligned}
$$

da cui la formula

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

che permette un calcolo più agevole di s^2 accumulando successivamente i quadrati dei valori osservati anziché quelli dei loro scarti dalla media.

4.4 Giustificazione della media

Stabiliamo quindi per convenzione che il nostro metodo per misurare la dispersione di un campione di dati è quello del calcolo della deviazione standard; accettiamo anche che *in qualche modo* questo numero sia legato all'errore presumibilmente commesso nel corso delle misure. Una definizione più precisa di ciò che si intende con le parole "errore commesso", ovverosia l'interpretazione probabilistica dello scarto quadratico medio nei riguardi delle misure ripetute, verrà data più avanti nel paragrafo 9.3.

Comunque, una volta assunto questo, possiamo approfondire il discorso già cominciato sulla media aritmetica come stima del centro della distribuzione e quindi del valore vero di una grandezza, nel caso di misure ripetute ed in assenza di errori sistematici. È infatti possibile provare[3] che la media

[3] La dimostrazione risale a Gauss se ci si limita alle sole operazioni lineari sui dati, e solo ad anni recenti per un qualsiasi algoritmo operante su di essi; vedi in proposito l'appendice E.

aritmetica è la stima del valore vero affetta *dal minimo errore casuale*, cioè avente la più piccola deviazione standard.

Riferendosi a quanto prima accennato, ciò significa che le medie aritmetiche di molti campioni analoghi di N misure avranno un istogramma più stretto delle mode, delle mediane e di qualsiasi altra misura di tendenza centrale desumibile dagli stessi campioni; la larghezza di tale istogramma (misurata, come abbiamo assunto, dal suo scarto quadratico medio) sarà messa in relazione con lo scarto quadratico medio delle misure da un teorema di cui ci occuperemo nel seguito. Da esso discenderà anche che l'errore statistico della media aritmetica converge a zero al crescere indefinito del numero di dati N. Per concludere:

1. *Disponendo di più misure ripetute della stessa grandezza fisica, si assume come migliore stima del valore vero di quella grandezza la loro media aritmetica.*
2. *Questa stima è più precisa di quanto non lo siano le singole misure, ed è tanto più attendibile quanto maggiore è il numero delle stesse.*
3. *Come valutazione dell'errore commesso nelle singole misure si assume il loro scarto quadratico medio; o meglio, per motivi che verranno chiariti in seguito, la quantità*[4]

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

[4] La differenza tra questa formula e quella prima citata non è praticamente avvertibile se N non è troppo piccolo.

Capitolo 5

Variabili casuali unidimensionali discrete

Già sappiamo (come osservato nel paragrafo 3.1) che, a causa degli inevitabili errori, la misura di una grandezza fisica può essere considerata un evento casuale; e che il numero reale da noi ottenuto in conseguenza della misura stessa può essere considerato una variabile casuale definita sull'insieme di tutti i possibili risultati.

Un insieme finito di operazioni di misura, i cui risultati costituiscono quello che in linguaggio statistico si dice *campione*, si può pensare come un particolare sottoinsieme formato da elementi estratti a caso dall'insieme di tutte le infinite possibili operazioni di misura che potrebbero essere effettuate sulla stessa grandezza fisica, eseguite col medesimo strumento e sfruttando le medesime procedure.

Quest'ultimo insieme nella terminologia della statistica si dice *universo* o *popolazione*, ed è in effetti una finzione (si pensi all'universo di tutti i possibili lanci di un dado nella teoria dei giochi d'azzardo), nel senso che in realtà esso non è un'entità preesistente alle operazioni effettivamente eseguite; a differenza dell'insieme di tutti gli individui di una vera popolazione, dalla quale si estrae realmente un campione per eseguire una ricerca demografica. Sebbene sia una finzione, questo concetto è tuttavia utile per poter applicare la teoria della probabilità alle caratteristiche di un campione.

In questo capitolo esamineremo il comportamento delle variabili casuali in generale (ed in particolare quello dei risultati delle misure): tra le altre cose, metteremo in evidenza i rapporti tra grandezze statistiche che si riferiscano ad un campione limitato e grandezze analoghe che siano invece

riferite all'intera popolazione (*teoria del campionamento*); e dimostreremo la validità della legge dei grandi numeri.

5.1 Generalità

Riprendiamo ora il concetto di variabile casuale già introdotto in precedenza nel paragrafo 3.1, e consideriamo alcuni esempi: se si associa ad ogni faccia di un dado un numero compreso tra 1 e 6 (il punteggio inciso sulla faccia stessa), si definisce una variabile casuale discreta; se l'evento casuale consiste invece nel lancio di due monete, indicando con E l'apparizione della testa nel lancio della prima e con F l'apparizione della testa nel lancio della seconda, il numero x di teste osservate nell'evento è ancora una variabile casuale discreta, la cui definizione è data dalla tabella seguente:

	x
EF	2
$E\bar{F}$	1
$\bar{E}F$	1
$\bar{E}\bar{F}$	0

e, come si può notare, la corrispondenza tra la variabile casuale e l'insieme dei possibili risultati non è in questo caso biunivoca.

Se l'insieme di definizione è continuo, la variabile casuale $x(E)$ può essere continua; è questo il caso più frequente nella fisica, ad esempio per le misure: ma anche in tal caso, a causa della sensibilità limitata degli strumenti, l'intervallo continuo di definizione della variabile x viene in pratica suddiviso in un numero finito M di intervalli, che vengono rappresentati dai valori centrali x_j della variabile casuale.

Detta ν_j la frequenza assoluta con cui si è presentato il risultato x_j nelle N prove complessive, sarà

$$\sum_{j=1}^{M} \nu_j = N$$

(potendo alcune frequenze ν_j risultare nulle perché i corrispondenti valori x_j non sono stati osservati nelle prove). Indicata con

$$f_j = \frac{\nu_j}{N}$$

la frequenza relativa del valore x_j nelle N prove, dalla prima relazione segue

$$\sum_{j=1}^{M} f_j = \sum_{j=1}^{M} \frac{\nu_j}{N} = \frac{1}{N}\sum_{j=1}^{M} \nu_j \equiv 1$$

esaurendo gli M valori x_j tutti i possibili risultati della misura.

Se il numero delle prove N è molto grande e viene fatto crescere a piacere, ciascuna f_j deve tendere statisticamente al valore p_j (probabilità di osservare il valore x_j), e sarà ancora

$$\sum_{j=1}^{M} p_j \equiv 1$$

come dovevamo ovviamente attenderci ricordando l'equazione (3.1).

5.2 Speranza matematica

Come sappiamo dal paragrafo 4.2.6, il valore medio della variabile x su di un campione finito è dato dall'equazione

$$\bar{x} = \sum_i f_i x_i$$

dove la sommatoria si intende estesa a tutti i valori che la x può assumere, essendo nulle le frequenze di quelli che non si sono effettivamente presentati; definiamo in maniera analoga una nuova grandezza $E(x)$, relativa all'intera popolazione, mediante la

$$E(x) = \sum_i p_i x_i \quad . \tag{5.1}$$

$E(x)$ (che si chiama *speranza matematica* della variabile casuale x) ci appare quindi come una generalizzazione alla popolazione del concetto di media aritmetica e, se si assumesse come definizione di probabilità quella empirica, sarebbe in base ad essa il limite (statistico) del valore medio del campione all'aumentare della sua dimensione; per cui lo chiameremo anche, meno appropriatamente, *valore medio di x sull'intera popolazione*.

È da notare come non ci sia alcuna garanzia dell'esistenza di $E(x)$ se l'insieme dei possibili valori x_i non è finito (in particolare se x è una variabile continua); in effetti esistono delle distribuzioni di probabilità usate anche in fisica (ad esempio la *distribuzione di Cauchy*, che studieremo più avanti nel paragrafo 8.3) per le quali la sommatoria della (5.1) non converge, e che non ammettono quindi speranza matematica.

La speranza matematica per la variabile casuale $[x - E(x)]^2$ (ossia la generalizzazione alla popolazione della varianza di un campione) si indica poi col simbolo $\mathrm{Var}(x)$:

$$\mathrm{Var}(x) = E\left\{[x - E(x)]^2\right\} = \sum_i p_i [x_i - E(x)]^2 \ ,$$

e ad essa ci riferiremo come *varianza della popolazione della variabile casuale* x; come $E(x)$, e per gli stessi motivi, anch'essa potrebbe non esistere per quelle variabili che assumono un numero infinito di possibili valori.

Le considerazioni dei paragrafi seguenti si applicano ovviamente solo a popolazioni di variabili casuali per le quali esista finita la speranza matematica e, qualora la si consideri, la varianza. Inoltre non useremo mai la definizione empirica di probabilità, ma quella assiomatica; e vedremo come, partendo da essa, si possa dimostrare la legge detta "dei grandi numeri" già enunciata nel paragrafo 3.5: ossia la convergenza, all'aumentare del numero di prove effettuate, della frequenza di un qualsiasi evento casuale alla sua probabilità.

5.3 Il valore medio delle combinazioni lineari

Consideriamo due variabili casuali x e y, aventi speranza matematica $E(x)$ ed $E(y)$ rispettivamente; ed una loro qualsiasi combinazione lineare a coefficienti costanti $z = ax + by$. Vogliamo dimostrare ora che la speranza matematica della nuova variabile z esiste, ed è data dalla combinazione lineare delle speranze matematiche di x e di y con gli stessi coefficienti a e b.

Indichiamo con x_j i possibili valori della prima variabile, e con y_k quelli della seconda; indichiamo poi con p_j e q_k le probabilità di ottenere un determinato valore rispettivamente per la x e per la y. Chiamiamo poi P_{jk} la probabilità che simultaneamente si abbia $x = x_j$ ed $y = y_k$; un particolare valore per la x potrà essere associato ad uno qualsiasi dei diversi valori della y, che sono tra loro mutuamente esclusivi: in definitiva, applicando la legge della probabilità totale (equazione (3.2)) risulterà

$$p_j = \sum_k P_{jk} \qquad \text{e} \qquad q_k = \sum_j P_{jk} \ .$$

Per la speranza matematica $E(z)$ di z avremo poi

$$\begin{aligned} E(ax+by) &= \sum_{jk} P_{jk}\,(a\,x_j + b\,y_k) \\ &= \sum_{jk} a\,P_{jk}\,x_j \ + \ \sum_{jk} b\,P_{jk}\,y_k \\ &= a\sum_j \left(\sum_k P_{jk}\right) x_j \ + \ b\sum_k \left(\sum_j P_{jk}\right) y_k \\ &= a\sum_j p_j\,x_j \ + \ b\sum_k q_k\,y_k \\ &= a\,E(x) + b\,E(y) \ . \end{aligned}$$

È immediato poi estendere, per induzione completa, questa dimostrazione alla combinazione lineare di un numero qualsiasi di variabili casuali: se abbiamo

$$F = ax + by + cz + \cdots$$

allora

$$E(F) = a\,E(x) + b\,E(y) + c\,E(z) + \cdots \ . \tag{5.2}$$

Una importante conseguenza può subito essere ricavata applicando l'equazione (5.2) alla media aritmetica $\bar{x}$ di un campione di N misure: essa infatti si può considerare come una particolare combinazione lineare delle misure stesse, con coefficienti tutti uguali tra loro e pari ad $1/N$.

Prendendo dalla popolazione un differente campione di N misure, la loro media aritmetica $\bar{x}$ sarà anch'essa in generale diversa: quale sarà la speranza matematica di $\bar{x}$, ovverosia il valore medio delle varie $\bar{x}$ su un numero molto elevato di campioni di N misure estratti a caso dalla popolazione — e, al limite, su tutti i campioni (aventi la stessa dimensione fissa N) che dalla popolazione è possibile ricavare?

$$\begin{aligned} E(\bar{x}) &= E\left(\frac{1}{N}\sum_{i=1}^{N} x_i\right) \\ &= \frac{1}{N}\sum_{i=1}^{N} E(x_i) \\ &= \frac{1}{N}\cdot N\,E(x) \end{aligned}$$

ed infine

$$E(\bar{x}) = E(x) \tag{5.3}$$

cioè:

Il valore medio della popolazione delle medie aritmetiche dei campioni di dimensione finita N estratti da una popolazione coincide con il valore medio della popolazione stessa.

5.4 La varianza delle combinazioni lineari

Dimostriamo ora un altro teorema generale che riguarda la varianza di una combinazione lineare di più variabili casuali, che supporremo però *stati-*

sticamente indipendenti. Usando gli stessi simboli già introdotti nel paragrafo 5.3, e dette x ed y due variabili casuali che godano di tale proprietà, sappiamo dall'equazione (3.4) che la probabilità P_{jk} che contemporaneamente risulti sia $x = x_j$ che $y = y_k$ è data dal prodotto delle probabilità rispettive p_j e q_k.

Per semplificare i calcoli, dimostriamo questo teorema dapprima nel caso particolare di due popolazioni x e y che abbiano *speranza matematica nulla*; estenderemo poi il risultato a due variabili (sempre statisticamente indipendenti) aventi speranza matematica qualunque. Ciò premesso, la combinazione lineare

$$z = ax + by$$

ha anch'essa speranza matematica zero: infatti applicando l'equazione (5.2) risulta

$$E(z) \;=\; E(ax + by) \;=\; a\,E(x) + b\,E(y) \;=\; 0$$

e si può allora ricavare (indicando con i simboli ${\sigma_x}^2$, ${\sigma_y}^2$ e ${\sigma_z}^2$ le varianze di x, y e z rispettivamente):

$$\begin{aligned}
{\sigma_z}^2 &= E\left\{[z - E(z)]^2\right\} \\
&= E\left\{z^2\right\} \\
&= E\left\{(ax + by)^2\right\} \\
&= \textstyle\sum_{jk} P_{jk}\,(a\,x_j + b\,y_k)^2 \\
&= \textstyle\sum_{jk} p_j q_k (a^2 {x_j}^2 + b^2 {y_k}^2 + 2a\,b\,x_j\,y_k) \\
&= a^2 \textstyle\sum_k q_k \sum_j p_j {x_j}^2 \;+\; b^2 \sum_j p_j \sum_k q_k {y_k}^2 \;+\; 2ab \sum_j p_j x_j \sum_k q_k y_k \\
&= a^2 {\sigma_x}^2 \textstyle\sum_k q_k \;+\; b^2 {\sigma_y}^2 \sum_j p_j \;+\; 2ab\,E(x)\,E(y)
\end{aligned}$$

ed infine

$$ {\sigma_z}^2 \;=\; a^2 {\sigma_x}^2 \;+\; b^2 {\sigma_y}^2 \;. \tag{5.4}$$

Allo scopo di estendere la validità dell'equazione (5.4) appena dimostrata a due variabili casuali x e y aventi speranza matematica anche differente da zero, dimostriamo ora il seguente

> TEOREMA: *due variabili casuali che differiscano per un fattore costante hanno la stessa varianza.*

Infatti, se le due variabili casuali x e ξ soddisfano questa ipotesi, allora deve risultare:

$$\begin{aligned}
\xi &= x + K \\
E(\xi) &= E(x) + K \\
{\sigma_\xi}^2 &= E\left\{[\xi - E(\xi)]^2\right\} \\
&= E\left\{[x + K - E(x) - K]^2\right\} \\
&= E\left\{[x - E(x)]^2\right\} \\
&= {\sigma_x}^2 \;.
\end{aligned}$$

Ora, date due variabili casuali x e y qualsiasi, ed una loro generica combinazione lineare $z = ax + by$, basta definire altre due variabili casuali ausiliarie

$$\xi = x - E(x) \qquad \text{ed} \qquad \eta = y - E(y)$$

(che ovviamente soddisfano l'ipotesi di avere speranza matematica zero): pertanto la loro combinazione lineare $\zeta = a\xi + b\eta$, che differisce anch'essa da z per un fattore costante e pari ad $aE(x) + bE(y)$, avrà varianza che, in conseguenza della (5.4), sarà data dalla

$$ {\sigma_\zeta}^2 = a^2 {\sigma_\xi}^2 + b^2 {\sigma_\eta}^2 \;.$$

Ma per quanto detto, x e ξ hanno la stessa varianza; così y ed η, e z e ζ. Ne consegue come per *qualsiasi* coppia di variabili casuali (purché però statisticamente indipendenti) vale la relazione (5.4), che possiamo enunciare nel modo seguente:

> *Una combinazione lineare, a coefficienti costanti, di due variabili casuali statisticamente indipendenti ha varianza uguale alla combinazione lineare delle rispettive varianze, con coefficienti pari ai quadrati dei coefficienti rispettivi*[1].

È ovvio poi estendere (per induzione completa) questo risultato alla combinazione lineare di un numero finito qualsivoglia di variabili casuali, che siano però sempre tra loro tutte statisticamente indipendenti: se

$$F = ax + by + cz + \cdots$$

[1] O, come si usa dire in sintesi, *gli errori si combinano quadraticamente.* Una formula più generale, che si può applicare a coppie di variabili casuali qualunque, verrà dimostrata nell'appendice C.

allora

$$\sigma_F{}^2 = a^2\sigma_x{}^2 + b^2\sigma_y{}^2 + c^2\sigma_z{}^2 + \cdots \quad . \tag{5.5}$$

5.5 L'errore della media dei campioni

Torniamo ora ad occuparci dello studio delle proprietà statistiche della media aritmetica di un campione di N misure indipendenti estratto da una popolazione,

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

e cerchiamo in particolare di determinarne la varianza. Applicando l'equazione (5.5) appena dimostrata, risulta

$$\begin{aligned}\sigma_{\bar{x}}{}^2 &= \frac{1}{N^2}\sum_{i=1}^{N}\sigma_{x_i}{}^2 \\ &= \frac{1}{N^2}\cdot N\sigma_x{}^2\end{aligned}$$

ed infine

$$\boxed{\sigma_{\bar{x}}{}^2 = \frac{\sigma_x{}^2}{N}} \tag{5.6}$$

In definitiva abbiamo dimostrato che

- *Le medie aritmetiche di campioni di N misure hanno varianza pari alla varianza della popolazione da cui le misure provengono, divisa per la dimensione dei campioni.*

e conseguentemente

- *L'errore quadratico medio della media di un campione è minore dell'analogo errore delle singole misure, e tende a zero al crescere del numero di misure effettuato.*

5.6 La legge dei grandi numeri

Le relazioni (5.3) e (5.6) sono state dimostrate sulla base della definizione di speranza matematica, e senza presupporre la convergenza verso di essa della media dei campioni (né quella delle frequenze verso la probabilità); vediamo ora come la legge dei grandi numeri (cui abbiamo già accennato nel paragrafo 3.5) si possa da esse dedurre.

5.6.1 La disuguaglianza di Bienaymé-Čebyšef

Sia una variabile casuale x, e siano $E(x)$ e σ^2 la speranza matematica e la varianza della sua popolazione; vogliamo ora determinare la probabilità che un valore di x scelto a caso differisca (in valore assoluto) da $E(x)$ per più di una assegnata quantità (positiva) ϵ. Questa è ovviamente data, in base alla legge della probabilità totale (3.2), dalla

$$\Pr\Big(\,|\,x - E(x)\,| \geq \epsilon\Big) = \sum_{|x_i - E(x)|\geq\epsilon} p_i$$

dove la sommatoria è estesa solo a quei valori x_i che soddisfano a tale condizione. Ora, sappiamo che

$$\sigma^2 \ = \ E\left\{[x - E(x)]^2\right\} \ = \ \sum_i p_i\,[x_i - E(x)]^2 \ ;$$

se si restringe la sommatoria ai soli termini x_i che differiscono (in modulo) da $E(x)$ per più di ϵ, il suo valore diminuirà o, al massimo, rimarrà invariato: deve risultare insomma

$$\sigma^2 \ \geq \ \sum_{|x_i - E(x)|\geq\epsilon} p_i\,[x_i - E(x)]^2 \ \geq \ \sum_{|x_i - E(x)|\geq\epsilon} p_i\,\epsilon^2 \ = \ \epsilon^2 \sum_{|x_i - E(x)|\geq\epsilon} p_i$$

e da questa relazione si ottiene la *disuguaglianza di Bienaymé-Čebyšef* [2]

$$\boxed{\Pr\Big(\,|\,x - E(x)\,| \geq \epsilon\Big) \ \leq \ \frac{\sigma^2}{\epsilon^2}} \tag{5.7}$$

e, se si pone $\epsilon = k\,\sigma$,

$$\Pr\Big(\,|\,x - E(x)\,| \geq k\,\sigma\Big) \ \leq \frac{1}{k^2} \tag{5.8}$$

[2] Irénée-Jules Bienaymé, francese, fu un matematico e statistico vissuto dal 1796 al 1878; Pafnuty Lvovič Čebyšef, matematico russo vissuto dal 1821 al 1894, si occupò di analisi, teoria dei numeri, probabilità, meccanica razionale e topologia.

(se nella dimostrazione si sostituissero le frequenze relative alle probabilità e la media aritmetica ad $E(x)$, si troverebbe che una analoga relazione vale anche per ogni campione di valori sperimentali x_i rispetto alla media aritmetica $\bar{x}$ ed alla varianza del campione s^2).

La (5.8) fissa un *limite superiore* per la probabilità esaminata, limite che deve valere per *qualsiasi* variabile casuale; con $k \leq 1$ non si ottiene alcuna informazione significativa da essa, ma con $k > 1$ si vede che il maggiorante della probabilità tende a zero all'aumentare di k. In particolare, per *qualsiasi* variabile casuale la probabilità di uno scarto dal valore medio non inferiore in valore assoluto a 2σ non può superare $\frac{1}{4} = 25\%$; e quella di uno scarto non inferiore in valore assoluto a 3σ non può superare $\frac{1}{9} \approx 11.1\%$.

Si deve notare che non si è fatta alcuna ipotesi sulla distribuzione, a parte l'esistenza della sua varianza σ^2 e della sua speranza matematica $E(x)$; in termini così generali il limite superiore (5.8) non può essere ridotto, ma non è escluso che (per una particolare distribuzione) la probabilità per la variabile da essa descritta di differire dal suo valore medio sia più piccola ancora di quella fissata dalla disuguaglianza di Bienaymé–Čebyšef. Ad esempio, se esiste finita la quantità

$$\mu_4 = E\left\{[x - E(x)]^4\right\}$$

(momento del quarto ordine rispetto alla media), con passaggi analoghi si troverebbe che

$$\Pr\left\{[x - E(x)]^4 \geq \epsilon\right\} \leq \frac{\mu_4}{\epsilon^4}$$

e, quindi, che

$$\Pr\left\{[x - E(x)]^4 \geq k\,\sigma\right\} \leq \frac{\mu_4}{k^4\,\sigma^4} \ .$$

Imponendo altre condizioni (anche non molto restrittive) alla distribuzione di probabilità, si potrebbe ridurre ulteriormente (in quantità anche notevole) il limite superiore stabilito in generale dalla (5.8); e stimare così anche la probabilità di uno scarto della variabile casuale dal suo valore medio inferiore a σ. Risale ad esempio a Gauss (1821) la dimostrazione che per una variabile continua avente distribuzione unimodale (con massimo in x_0), e per la quale esista finita la quantità ${\sigma_0}^2 = E[(x - x_0)^2]$, la probabilità di uno scarto dalla moda x_0 non inferiore in valore assoluto ad una quantità prefissata non può superare la frazione $\frac{4}{9}$ del limite di Bienaymé–Čebyšef:

$$\Pr\left\{|x - x_0| \geq k\,\sigma\right\} \leq \frac{4}{9\,k^2} \ .$$

Se la distribuzione è anche *simmetrica*, moda e media coincidono entrambe col centro di simmetria; e σ_0 è uguale alla deviazione standard σ. Per distribuzioni di questo genere, quindi, il limite superiore per la probabilità di uno scarto che non sia inferiore a k volte l'errore quadratico medio scende a $\frac{4}{9} \approx 44.4\%$ per $k = 1$; a $\frac{1}{9} \approx 11.1\%$ per $k = 2$; ed a $\frac{4}{81} \approx 4.9\%$ per $k = 3$ (e vedremo poi nel paragrafo 9.3 che per le misure affette da errori puramente casuali i limiti superiori sono ancora più stringenti di questi).

5.6.2 Il teorema di Čebyšef

Adesso applichiamo la (5.7) alla variabile casuale $\bar{x}$, media aritmetica di un campione di dimensione N di valori che supponiamo essere statisticamente indipendenti:

$$\Pr\Big(|\bar{x} - E(\bar{x})| \geq \epsilon\Big) \leq \frac{{\sigma_{\bar{x}}}^2}{\epsilon^2} \ ; \tag{5.9}$$

ma valendo, per questa variabile casuale, le

$$E(\bar{x}) = E(x) \qquad \text{e} \qquad \mathrm{Var}\,(\bar{x}) = \frac{\sigma^2}{N} \ ,$$

sostituendo nella (5.9) otteniamo

$$\Pr\Big(|\bar{x} - E(x)| \geq \epsilon\Big) \leq \frac{\sigma^2}{N\,\epsilon^2} \ . \tag{5.10}$$

Ora, scelti comunque due numeri positivi ϵ e δ, si può trovare in conseguenza un valore di N per cui il secondo membro della (5.10) risulti sicuramente minore di δ: basta prendere $N > M = \lceil \sigma^2/(\delta\,\epsilon^2) \rceil$. In base alla definizione (3.8), questo significa che vale il

> TEOREMA (DI ČEBYŠEF): *il valore medio di un campione finito di valori di una variabile casuale qualunque converge statisticamente, all'aumentare della dimensione del campione, alla speranza matematica per quella variabile.*

5.6.3 Il teorema di Bernoulli

Sia un qualsiasi evento casuale E avente probabilità p di verificarsi; indichiamo con $q = 1 - p$ la probabilità del non verificarsi di E (cioè la probabilità dell'evento complementare $\bar{E}$).

Consideriamo poi un insieme di N prove nelle quali si osserva se E si è o no verificato; ed introduciamo una variabile casuale y, definita come il

numero di volte in cui E si è verificato in una di tali prove. Ovviamente y può assumere i due soli valori 1 (con probabilità p) e 0 (con probabilità q); la sua speranza matematica è perciò data da

$$E(y) = 1 \cdot p + 0 \cdot q = p \ . \tag{5.11}$$

La frequenza relativa f dell'evento E nelle N prove si può chiaramente esprimere (indicando con y_i il valore assunto dalla variabile casuale y nella i-esima di esse) come

$$f = \frac{1}{N} \sum_{i=1}^{N} y_i \ ,$$

ossia è data *dal valore medio della y sul campione di prove*, $\bar{y}$; ma quest'ultimo (per il teorema di Čebyšef[3]) deve convergere statisticamente, all'aumentare di N, alla speranza matematica per y: che vale proprio p. Riassumendo, abbiamo così dimostrato il

> TEOREMA (DI BERNOULLI, O LEGGE "DEI GRANDI NUMERI"): *la frequenza relativa di qualunque evento casuale converge (statisticamente) alla sua probabilità all'aumentare del numero delle prove.*

5.7 Valore medio e valore vero

Anche se non ci basiamo sulla definizione empirica di probabilità, ma su quella assiomatica, possiamo ora presupporre la convergenza della media aritmetica dei campioni di misure alla speranza matematica della grandezza misurata, che ora a buon diritto possiamo chiamare "valore medio del risultato della misura sull'intera popolazione".

Si può ora meglio precisare la distinzione fra errori casuali ed errori sistematici: i primi, visto che possono verificarsi con uguale probabilità in difetto ed in eccesso rispetto al valore vero, avranno valore medio nullo; mentre errori sistematici causeranno invece per definizione una differenza tra il valore medio delle misure $E(x)$ ed il valore vero. In assenza di errori sistematici assumiamo allora che valore medio e valore vero coincidano: ammettiamo insomma (lo proveremo più avanti per la distribuzione normale) che in tal

[3] Il teorema di Čebyšef vale per tutte le variabili casuali per le quali esistano sia la speranza matematica che la varianza: la prima è espressa dall'equazione (5.11), la seconda sarà ricavata più tardi nell'equazione (8.8) a pagina 109.

caso $E(x)$ esista e sia uguale a x^*. Sappiamo insomma che risulta

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i = \frac{1}{N} \sum_{j=1}^{M} \nu_j x_j = \sum_{j=1}^{M} f_j x_j$$

$$\lim_{N \to \infty} \bar{x} \equiv E(x) = \sum_j p_j x_j$$

e postuliamo che

$$E(x) \equiv x^* \ ;$$

inoltre sappiamo che anche

$$E(\bar{x}) = E(x) \equiv x^* \ .$$

Ossia, non solo $\bar{x}$ converge ad $E(x)$ all'aumentare della dimensione del campione; ma, qualunque sia il valore di quest'ultima grandezza, **mediamente** $\bar{x}$ *coincide* con $E(x)$. Ripetendo varie volte la misura ed ottenendo così più campioni con differenti medie aritmetiche, dando come stima di $E(x)$ la media di uno dei nostri campioni avremo insomma la stessa probabilità di sbagliare per difetto o per eccesso[4].

5.8 Scarto ed errore quadratico medio

L'ultimo punto da approfondire riguarda la relazione tra la varianza s^2 di un campione di N misure e quella σ^2 della popolazione da cui il campione proviene. Ora, s^2 si può esprimere come

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

e possiamo osservare che (per qualsiasi numero x^* e quindi anche per l'incognito valore vero) vale la seguente relazione matematica:

[4] Questo nella terminologia statistica si esprime dicendo che la media dei campioni è una stima *imparziale* della media della popolazione; al contrario della varianza del campione che, come vedremo nel prossimo paragrafo, è una stima *parziale* (o *distorta*) della varianza della popolazione (il concetto verrà poi approfondito nel paragrafo 11.1).

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}(x_i - x^*)^2 &= \frac{1}{N}\sum_{i=1}^{N}\left[(x_i - \bar{x}) + (\bar{x} - x^*)\right]^2 \\
&= \frac{1}{N}\left[\sum_{i=1}^{N}(x_i - \bar{x})^2 + \sum_{i=1}^{N}(\bar{x} - x^*)^2 + 2(\bar{x} - x^*)\sum_{i=1}^{N}(x_i - \bar{x})\right] \\
&= \frac{1}{N}\left[\sum_{i=1}^{N}(x_i - \bar{x})^2 + N(\bar{x} - x^*)^2\right] \\
&= s^2 + (\bar{x} - x^*)^2 \qquad (5.12)
\end{aligned}$$

(si è sfruttata qui l'equazione (4.2), secondo la quale la somma algebrica degli scarti delle misure dalla loro media aritmetica è identicamente nulla; vedi anche l'analoga formula (4.3) nel paragrafo 4.2.3).

Cerchiamo ora di capire come le varianze s^2 dei campioni di dimensione N siano legate all'analoga grandezza, σ^2 o $\mathrm{Var}(x)$, definita sull'intera popolazione, e però calcolata rispetto al valore medio di essa, $E(x) = x^*$:

$$\sigma^2 = E\left\{[x - E(x)]^2\right\} = E\left\{(x - x^*)^2\right\} .$$

Sfruttando la relazione (5.12) in precedenza trovata, si ha

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

e prendendo i valori medi di entrambi i membri (sugli infiniti campioni di dimensione N che si possono pensare estratti in modo casuale dalla popolazione originaria), otteniamo

$$E(s^2) = \sigma^2 - E\left\{(\bar{x} - x^*)^2\right\} .$$

Ricordando come il valore medio del quadrato degli scarti di una variabile (qui $\bar{x}$) dal suo valore medio (che è $E(\bar{x}) = E(x) = x^*$) sia per definizione la varianza della variabile stessa (che indicheremo qui come quadrato dell'errore quadratico medio $\sigma_{\bar{x}}$), si ricava infine:

$$E(s^2) = \sigma^2 - \sigma_{\bar{x}}{}^2 < \sigma^2 . \qquad (5.13)$$

Insomma:

- *Il valore medio della varianza s^2 di un campione è sistematicamente inferiore all'analoga grandezza σ^2 che si riferisce all'intera popolazione.*
- *La differenza tra la varianza della popolazione σ^2 e la varianza di un campione di N misure da essa estratto è* ***in media*** *pari alla varianza della media del campione.*

5.9 Stima della varianza della popolazione

Vediamo ora come si può stimare la varianza dell'insieme delle infinite misure che possono essere fatte di una grandezza fisica a partire da un particolare campione di N misure. Riprendiamo l'equazione (5.13); in essa abbiamo già dimostrato che

$$E(s^2) = \sigma^2 - \sigma_{\bar{x}}{}^2$$

e sappiamo dalla (5.6) che la varianza della media del campione vale

$$\sigma_{\bar{x}}{}^2 = \frac{\sigma^2}{N} .$$

Risulta pertanto

$$\boxed{E(s^2) = \frac{N-1}{N}\,\sigma^2}$$

e quindi

> ***Mediamente*** *la varianza di un campione di N misure è inferiore alla varianza della intera popolazione per un fattore $(N-1)/N$.*

Questo è il motivo per cui, per avere una stima *imparziale* (ossia *mediamente corretta*) di σ, si usa (come già anticipato) la quantità μ definita attraverso la

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

quantità il cui valore medio su infiniti campioni risulta proprio σ^2.

5.10 Ancora sull'errore quadratico medio

Diamo qui un'altra dimostrazione del teorema riguardante la stima corretta dell'errore quadratico medio di una popolazione a partire da un campione, seguendo una linea diversa e più vicina alle verifiche sperimentali che si possono fare avendo a disposizione numerosi dati.

Si supponga di avere M campioni contrassegnati dall'indice j (con j che assume i valori $1, \ldots, M$); ciascuno di essi sia poi costituito da N misure ripetute della stessa grandezza x, contrassegnate a loro volta dall'indice i ($i = 1, \ldots, N$): il valore osservato nella misura i-esima del campione j-esimo sia indicato insomma dal simbolo x_{ij}.

Indicando con x^* il valore vero di x, e con $\bar{x}_j$ la media aritmetica del campione j-esimo, vale la

$$\begin{aligned}(x_{ij} - x^*)^2 &= \left[(x_{ij} - \bar{x}_j) + (\bar{x}_j - x^*)\right]^2 \\ &= (x_{ij} - \bar{x}_j)^2 \;+\; (\bar{x}_j - x^*)^2 \;+\; 2\,(\bar{x}_j - x^*)\,(x_{ij} - \bar{x}_j) \quad .\end{aligned}$$

Ora sommiamo su i tutte le N uguaglianze che si hanno per i valori dell'indice $i = 1, 2, \ldots, N$ e dividiamo per N; se indichiamo con $s_j{}^2$ la varianza del campione j-esimo, data da

$$s_j{}^2 = \frac{1}{N}\sum_{i=1}^{N}(x_{ij} - \bar{x}_j)^2$$

otteniamo alla fine

$$\frac{1}{N}\sum_{i=1}^{N}(x_{ij} - x^*)^2 \;=\; s_j{}^2 \;+\; (\bar{x}_j - x^*)^2 \;+\; \frac{2}{N}(\bar{x}_j - x^*)\sum_{i=1}^{N}(x_{ij} - \bar{x}_j) \quad .$$

L'ultima sommatoria a destra è la somma algebrica degli scarti delle misure del campione j-esimo dalla loro media aritmetica $\bar{x}_j$ che sappiamo essere identicamente nulla. Dunque, per ogni j vale la

$$\frac{1}{N}\sum_{i=1}^{N}(x_{ij} - x^*)^2 = s_j{}^2 \;+\; (\bar{x}_j - x^*)^2$$

e se sommiamo membro a membro tutte le M uguaglianze che abbiamo per $j = 1, 2, \ldots, M$ e dividiamo per M, risulta

$$\frac{1}{M}\sum_{j=1}^{M}\frac{1}{N}\sum_{i=1}^{N}(x_{ij} - x^*)^2 \;=\; \frac{1}{M}\sum_{j=1}^{M}s_j{}^2 \;+\; \frac{1}{M}\sum_{j=1}^{M}(\bar{x}_j - x^*)^2 \quad .$$

Ora supponiamo di avere a disposizione moltissimi campioni e passiamo al limite per $M \to \infty$. Il primo membro (che rappresenta il valore medio, su tutti i dati e tutti gli infiniti campioni, del quadrato degli scarti dal valore *vero*) converge stocasticamente alla varianza della variabile casuale x; il secondo termine a destra (valore medio, su tutti gli infiniti campioni, del quadrato degli scarti della media aritmetica del campione dal proprio valore vero) converge alla varianza delle medie dei campioni di N misure $\sigma_{\bar{x}}{}^2$.

Il primo termine a destra è il valore medio della varianza dei campioni di N misure e, sostituendo, infine si trova

$$\begin{aligned}\sigma^2 &= \lim_{M\to\infty}\frac{1}{NM}\sum_{ij}(x_{ij} - x^*)^2 \\ &= \lim_{M\to\infty}\frac{1}{M}\sum_{j=1}^{M}s_j{}^2 \;+\; \lim_{M\to\infty}\frac{1}{M}\sum_{j=1}^{M}(\bar{x}_j - x^*)^2 \\ &= E(s^2) \;+\; \sigma_{\bar{x}}{}^2 \quad .\end{aligned}$$

Ora, avendo già dimostrato che

$$\sigma_{\bar{x}}{}^2 = \frac{\sigma^2}{N} \quad ,$$

si ricava facilmente

$$\sigma^2 = E(s^2) + \frac{\sigma^2}{N}$$

ovvero

$$E(s^2) = \frac{N-1}{N}\,\sigma^2$$

che è il risultato già ottenuto.

Si noti che mentre molti teoremi della statistica sono validi solo *asintoticamente*, cioè per campioni numerosi o per un numero molto grande di variabili, questo teorema vale per ogni N (≥ 2).