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

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

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):

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

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:

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

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

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

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

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

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

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

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:

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

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, $E(x) = x^*$:

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

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

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

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:

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

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

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

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

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

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

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

Ora, avendo già dimostrato che

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

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).

Capitolo 6

Variabili casuali unidimensionali continue

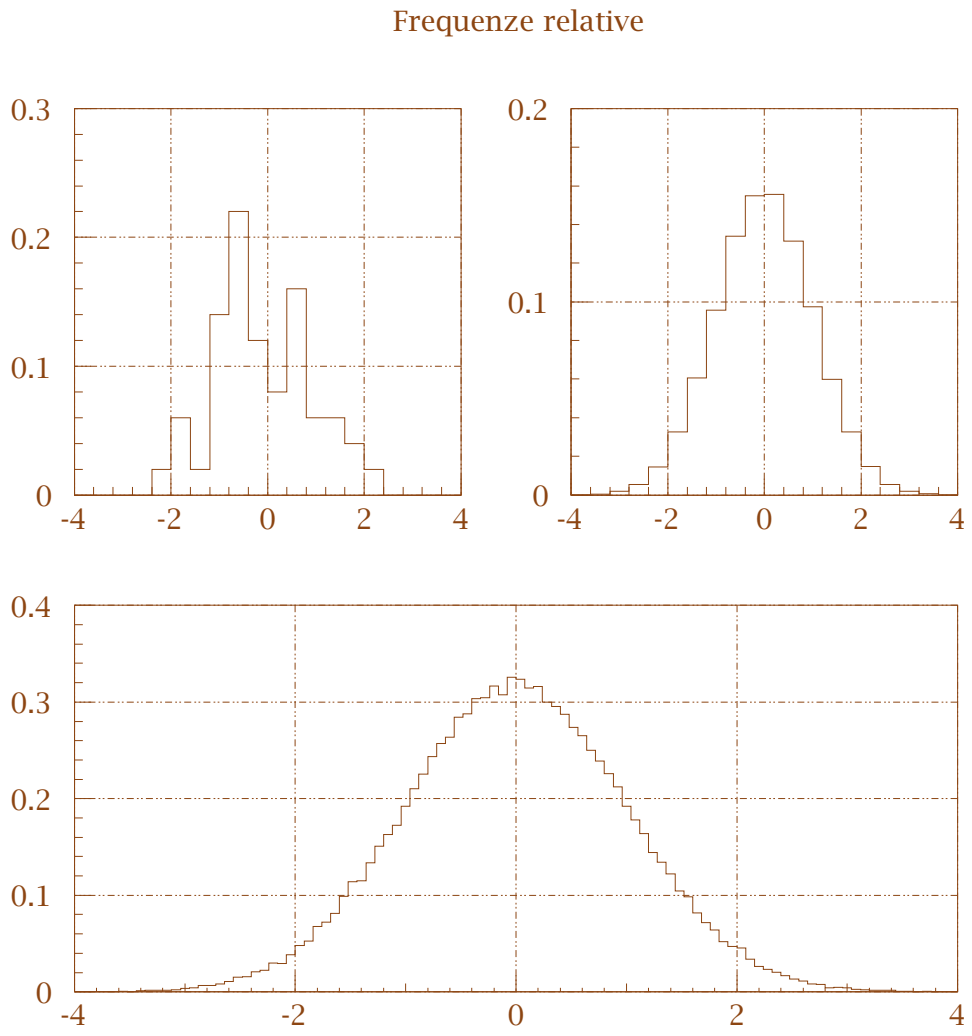
Le definizioni di probabilità che abbiamo finora usato sono adatte solo per una variabile casuale che possa assumere solo valori discreti; vediamo innanzi tutto come il concetto di probabilità si possa generalizzare a variabili casuali continue, variabili che possono cioè assumere tutti gli infiniti valori appartenenti ad un insieme continuo: tali si suppone generalmente siano i risultati delle misure delle grandezze fisiche, per poter applicare ad essi il calcolo differenziale ed integrale.

6.1 La densità di probabilità

Definiamo arbitrariamente delle classi di frequenza, suddividendo l'asse delle x in intervalli di ampiezze che, per semplicità, supponiamo siano tutte uguali; ed immaginiamo di fare un certo numero N di misure della grandezza fisica x. Come sappiamo, possiamo riportare le misure ottenute in istogramma tracciando, al di sopra dell'intervallo che rappresenta ogni classe, un rettangolo avente area uguale alla frequenza relativa[1] con cui una misura è caduta in essa; l'altezza dell'istogramma in ogni intervallo è data quindi da tale frequenza divisa per l'ampiezza dell'intervallo di base, e l'area totale dell'istogramma stesso vale uno.

[1] Non vi è alcuna differenza nell'usare frequenze relative o assolute: essendo esse proporzionali l'una all'altra, l'aspetto dell'istogramma è il medesimo — cambia solo la scala dell'asse delle ordinate.

FIGURA 6a - Nella prima figura, l'istogramma della grandezza x per un numero piccolo di misure; nella seconda, lo stesso istogramma per un numero molto grande di misure; nell'ultima, l'istogramma si approssima alla curva limite quando l'intervallo di base tende a zero.



Se immaginiamo di far tendere all'infinito il numero di misure effettuate, in base alla legge dei grandi numeri ci aspettiamo un "aggiustamento" dell'istogramma in modo che l'area rappresentata sopra ogni intervallo tenda alla *probabilità* che il valore misurato cada entro di esso; le altezze tenderanno quindi al rapporto tra questa probabilità e l'ampiezza dell'intervallo di base dell'istogramma.

Disponendo di un numero infinitamente grande di misure, ha senso diminuire l'ampiezza degli intervalli in cui l'asse delle x è stato diviso, e renderla piccola a piacere. Se l'intervallo corrispondente ad una data classe di frequenza tende a zero, la probabilità che una misura cada in esso tende ugualmente a zero; ma se esiste ed è finito il limite del rapporto tra probabilità $\mathrm{d}p$ ed ampiezza $\mathrm{d}x$ dell'intervallo, l'istogramma tenderà ad una curva continua la cui ordinata sarà in ogni punto data da tale limite.

L'ordinata di questa curva al di sopra di un intervallo infinitesimo $\mathrm{d}x$ vale quindi

$$y = f(x) = \frac{\mathrm{d}p}{\mathrm{d}x}$$

e le dimensioni della grandezza y sono quelle di una probabilità (un numero puro) divise per quelle della grandezza x; la y prende il nome di *densità di probabilità*, o di *funzione di frequenza*, della x.

La variabile continua schematizza il caso in cui i valori osservabili (sempre discreti per la sensibilità limitata degli strumenti) sono molto densi, separati cioè da intervalli molto piccoli, e assai numerosi. In questa situazione la probabilità di osservare uno solo di tali valori è anch'essa estremamente piccola — ed ha interesse soltanto la probabilità che venga osservato uno tra i molti possibili valori della x che cadono in un dato intervallo $[x_1, x_2]$ di ampiezza grande rispetto alla risoluzione sperimentale.

Se dividiamo tale intervallo in un numero molto grande di sottointervalli infinitesimi di ampiezza $\mathrm{d}x$, gli eventi casuali consistenti nell'appartenere il risultato della misura ad una delle classi di frequenza relative sono mutuamente esclusivi; di conseguenza, vista l'equazione (3.2), la probabilità che x appartenga all'intervallo finito $[x_1, x_2]$ è data dalla somma delle probabilità (infinitesime) rispettive $\mathrm{d}p = f(x)\,\mathrm{d}x$: e questa, per definizione, è l'integrale di $f(x)$ rispetto ad x nell'intervallo $[x_1, x_2]$.

Insomma, qualunque sia l'intervallo $[x_1, x_2]$ vale la

$$\Pr\Big(x \in [x_1, x_2]\Big) = \int_{x_1}^{x_2} f(x)\,\mathrm{d}x \quad ;$$

e, in definitiva:

> *Per le variabili continue non si può parlare di probabilità attraverso le definizioni già esaminate. È invece possibile associare ad ogni*

variabile continua x una funzione "densità di probabilità" $f(x)$, da cui si può dedurre la probabilità che la x cada in un qualsiasi intervallo finito prefissato: questa è data semplicemente dall'area sottesa dalla curva nell'intervallo in questione.

Analogamente al concetto sperimentale di frequenza cumulativa relativa, introdotto a pagina 33 nel paragrafo 4.1, si può definire la *funzione di distribuzione* per una variabile continua x come

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t \quad .$$

Essa rappresenta la probabilità di osservare un valore non superiore ad x, e dovrà necessariamente soddisfare la $F(+\infty) \equiv 1$. Quindi deve valere la cosiddetta

CONDIZIONE DI NORMALIZZAZIONE: *l'integrale di una qualunque funzione che rappresenti una densità di probabilità, nell'intervallo $[-\infty, +\infty]$ vale 1.*

$$\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x \;=\; 1 \quad . \tag{6.1}$$

È da enfatizzare come il solo fatto che valga la condizione di normalizzazione, ossia che converga l'integrale (6.1), è sufficiente a garantire che una qualsiasi funzione che rappresenti una densità di probabilità *debba tendere a zero* quando la variabile indipendente tende a più o meno infinito; e questo senza alcun riferimento alla particolare natura del fenomeno casuale cui essa è collegata. Questo non è sorprendente, visto che la disuguaglianza (5.7) di Bienaymé–Čebyšef implica che a distanze via via crescenti dal valore medio di una qualsiasi variabile casuale corrispondano probabilità via via decrescenti, e che si annullano asintoticamente.

Al lettore attento non sarà sfuggito il fatto che, per introdurre il concetto di densità di probabilità, ci si è ancora una volta basati sul risultato di un esperimento reale (l'istogramma delle frequenze relative in un campione); e si è ipotizzato poi che la rappresentazione di tale esperimento si comporti in un determinato modo quando alcuni parametri (il numero di misure e la sensibilità sperimentale) vengono fatti tendere a limiti che, nella pratica, sono irraggiungibili.

Questo è in un certo senso analogo all'enfasi che abbiamo prima posto sulla definizione empirica della probabilità, in quanto più vicina all'esperienza reale di una definizione totalmente astratta come quella assiomatica; per un matematico la densità di probabilità di una variabile casuale continua è

invece definita semplicemente come una funzione non negativa, integrabile su tutto l'asse reale e che obbedisca alla condizione di normalizzazione. Il passo successivo consiste nell'associare ad ogni intervallo infinitesimo $\mathrm{d}x$ la quantità $\mathrm{d}p = f(x)\mathrm{d}x$, e ad ogni intervallo finito $[x_1, x_2]$ il corrispondente integrale: integrale che, come si può facilmente controllare, soddisfa la definizione assiomatica di probabilità.

6.2 La speranza matematica per le variabili continue

Possiamo ora determinare l'espressione della speranza matematica di una generica variabile casuale continua x; questa grandezza, che avevamo già definito nell'equazione (5.1) come

$$E(x) = \sum_i p_i x_i$$

per una variabile discreta, si dovrà ora scrivere per una variabile continua

$$E(x) = \int_{-\infty}^{+\infty} x \cdot f(x)\,\mathrm{d}x \quad ;$$

dove per $f(x)$ si intende la funzione densità di probabilità della variabile casuale x.

Per ricavare questa formula, basta pensare di aver suddiviso l'asse delle x in un numero grandissimo di intervalli estremamente piccoli di ampiezza $\mathrm{d}x$, ad ognuno dei quali è associata una probabilità anch'essa estremamente piccola che vale $\mathrm{d}p = f(x)\,\mathrm{d}x$; e sostituire poi nella formula per variabili discrete. In base al teorema di pagina 57 (il teorema di Čebyšef), le medie aritmetiche dei campioni finiti di valori della grandezza x tendono proprio a questo $E(x)$ all'aumentare indefinito di N.

La speranza matematica di una qualsiasi grandezza $W(x)$ funzione della variabile casuale x sarà poi

$$E[W(x)] = \int_{-\infty}^{+\infty} W(x) \cdot f(x)\,\mathrm{d}x \quad . \tag{6.2}$$

6.3 I momenti

Per qualunque variabile casuale x si possono definire, sempre sulla popolazione, i cosiddetti *momenti*: il momento di ordine k rispetto all'origine,

λ_k, è la speranza matematica di x^k; ed il momento di ordine k rispetto alla media, μ_k, è la speranza matematica di $[x - E(x)]^k$. In formula (con ovvio significato dei simboli):

$$\lambda_k \;=\; E(x^k) \;=\; \sum_i p_i \, {x_i}^k$$

e

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

per una variabile discreta (analogamente, usando le frequenze, si possono definire i momenti rispetto all'origine ed alla media aritmetica di un campione); oppure

$$\lambda_k \;=\; E(x^k) \;=\; \int_{-\infty}^{+\infty} x^k \, f(x) \, \mathrm{d}x$$

e

$$\mu_k \;=\; E\left\{[x - E(x)]^k\right\} \;=\; \int_{-\infty}^{+\infty} [x - E(x)]^k \, f(x) \, \mathrm{d}x$$

per una variabile continua.

Chiaramente, se la popolazione è costituita da un numero infinito di elementi (quindi, in particolare, per le variabili continue), non è detto che i momenti esistano; inoltre $E(x) \equiv \lambda_1$ e $\mathrm{Var}(x) \equiv \mu_2 \equiv \lambda_2 - {\lambda_1}^2$. Dalla definizione consegue immediatamente che, per qualsiasi popolazione per cui esista $E(x)$,

$$\begin{aligned}
\mu_1 &= \int_{-\infty}^{+\infty} [x - E(x)] \, f(x) \, \mathrm{d}x \\
&= \int_{-\infty}^{+\infty} x \, f(x) \, \mathrm{d}x - E(x) \int_{-\infty}^{+\infty} f(x) \, \mathrm{d}x \\
&= E(x) - E(x) \\
&\equiv 0 \;\; .
\end{aligned}$$

È poi facile dimostrare che, per popolazioni simmetriche rispetto alla media, tutti i momenti di ordine dispari rispetto ad essa, se esistono, valgono zero: basta considerare come, negli integrali, i contributi infinitesimi di ognuno degli intervallini si possano associare a due a due in modo che si annullino vicendevolmente. Il valore del momento del terzo ordine rispetto alla media aritmetica può quindi essere considerato una sorta di misura dell'asimmetria di una distribuzione.

In pratica però si preferisce usare, in luogo di μ_3, un parametro adimensionale; definendo il cosiddetto *coefficiente di asimmetria* (o *skewness*, in inglese) come

$$\gamma_1 = \frac{\mu_3}{\left(\sqrt{\mu_2}\right)^3} = \frac{\mu_3}{\sigma^3}$$

(dove $\sigma = \sqrt{\mu_2}$ è la radice quadrata della varianza); γ_1 è nullo per densità di probabilità simmetriche rispetto alla media, oppure ha segno positivo (o negativo) a seconda che i valori della funzione di frequenza per la variabile casuale in questione si trovino "sbilanciati" verso la destra (o verso la sinistra) rispetto al valore medio.

Dal momento del quarto ordine rispetto alla media si può ricavare un altro parametro adimensionale talvolta usato per caratterizzare una distribuzione: il *coefficiente di curtòsi* γ_2', definito come

$$\gamma_2' = \frac{\mu_4}{{\mu_2}^2} = \frac{\mu_4}{\sigma^4} \tag{6.3}$$

e che è ovviamente sempre positivo. Esso misura in un certo senso la "rapidità" con cui una distribuzione di probabilità converge a zero quando ci si allontana dalla zona centrale in cui essa assume i valori più alti (individuata dal valore di $E(x) \equiv \lambda_1$): o, se si preferisce, l'importanza delle sue "code" laterali; infatti, quanto più rapidamente la funzione converge a zero in queste code, tanto più piccolo sarà il valore di γ_2'. Come si potrebbe ricavare integrandone la funzione di frequenza (che troveremo più avanti nel paragrafo 8.2), il coefficiente di curtosi della distribuzione normale calcolato usando la (6.3) vale 3; per questo motivo si preferisce generalmente definirlo in modo differente, usando la

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 \ .$$

Questo fa sì che esso valga zero per la funzione di Gauss, e che assuma poi valori di segno negativo o positivo per funzioni che convergano a zero nelle code in maniera rispettivamente più "rapida" o più "lenta" della distribuzione normale.

6.4 Funzione generatrice e funzione caratteristica

La speranza matematica della funzione e^{tx} per una variabile casuale continua x prende il nome di *funzione generatrice dei momenti* della variabile stessa; la indicheremo nel seguito col simbolo $M_x(t)$. Il motivo del nome è

che risulta, indicando con $f(x)$ la densità di probabilità di x:

$$\boxed{M_x(t) \;=\; E(e^{tx}) \;=\; \int_{-\infty}^{+\infty} e^{tx} f(x)\,\mathrm{d}x} \tag{6.4}$$

(per una variabile continua, oppure

$$M_x(t) = \sum_i p_i\, e^{tx_i}$$

per una variabile discreta); e, ricordando sia lo sviluppo in serie di McLaurin della funzione esponenziale

$$e^{tx} \;=\; \sum_{k=0}^{\infty} \frac{(tx)^k}{k!}$$

che la definizione dei momenti rispetto all'origine, *se questi esistono tutti fino a qualsiasi ordine* risulta anche

$$M_x(t) \;=\; \sum_{k=0}^{\infty} \frac{t^k}{k!}\lambda_k$$

da cui

$$\left.\frac{\mathrm{d}^k M_x(t)}{\mathrm{d}t^k}\right|_{t=0} \;=\; \lambda_k$$

e, in definitiva, derivando successivamente la funzione generatrice si possono ricavare tutti i momenti della funzione di frequenza da cui essa discende. Se interessa invece uno dei momenti non rispetto all'origine, ma rispetto al valore medio λ, basta considerare l'altra funzione

$$\overline{M}_x(t) \;=\; E\left[e^{t(x-\lambda)}\right] \;=\; e^{-t\lambda} M_x(t) \tag{6.5}$$

e si trova facilmente che risulta

$$\left.\frac{\mathrm{d}^k \overline{M}_x(t)}{\mathrm{d}t^k}\right|_{t=0} \;=\; \mu_k \;.$$

La speranza matematica della funzione e^{itx} si chiama invece *funzione caratteristica* della variabile casuale x, e si indica con $\phi_x(t)$:

$$\boxed{\phi_x(t) \;=\; E(e^{itx}) \;=\; \int_{-\infty}^{+\infty} e^{itx} f(x)\,\mathrm{d}x} \tag{6.6}$$

(per una variabile continua, e

$$\phi_x(t) \;=\; \sum_k p_k\, e^{itx_k} \tag{6.7}$$

per una variabile discreta); e, se esistono i momenti di qualsiasi ordine rispetto all'origine, risulta anche

$$\phi_x(t) \;=\; \sum_{k=0}^{\infty} \frac{(it)^k}{k!}\,\lambda_k \tag{6.8}$$

dalla quale si ricava

$$\left.\frac{\mathrm{d}^k \phi_x(t)}{\mathrm{d}t^k}\right|_{t=0} \;=\; i^k \lambda_k \;\;. \tag{6.9}$$

Queste funzioni sono importanti in virtù di una serie di teoremi, che citeremo qui senza dimostrarli:

- I momenti (se esistono fino a qualunque ordine) caratterizzano univocamente una variabile casuale; se due variabili casuali hanno gli stessi momenti fino a qualsiasi ordine, la loro densità di probabilità è identica.

- La funzione generatrice esiste solo se esistono i momenti fino a qualsiasi ordine; e anch'essa caratterizza univocamente una variabile casuale, nel senso che se due variabili hanno la stessa funzione generatrice la loro densità di probabilità è identica.

- La $\phi_x(t)$ prima definita si chiama anche *trasformata di Fourier* della funzione $f(x)$; anch'essa caratterizza univocamente una variabile casuale nel senso su detto. Le proprietà che contraddistinguono una funzione che rappresenti una densità di probabilità implicano poi che la funzione caratteristica, a differenza della funzione generatrice dei momenti, *esista sempre* per qualsiasi variabile casuale; la (6.9) è però valida solo se i momenti esistono fino a qualsiasi ordine. Inoltre, se è nota la $\phi_x(t)$, *la si può sempre invertire* (riottenendo da essa la f) attraverso la

$$\boxed{f(x) \;=\; \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-ixt}\,\phi_x(t)\,\mathrm{d}t} \tag{6.10}$$

(*trasformata inversa di Fourier*).

Vogliamo infine ricavare una relazione che ci sarà utile più avanti: siano le N variabili casuali continue x_k (che supponiamo tutte statisticamente indipendenti tra loro), ognuna delle quali sia associata ad una particolare funzione caratteristica $\phi_k(t)$; il problema che vogliamo affrontare consiste nel determinare la funzione caratteristica della nuova variabile casuale S, definita come loro somma:

$$S = \sum_{k=1}^{N} x_k \quad .$$

Il valore di ogni x_k sarà univocamente definito dai possibili risultati di un qualche evento casuale E_k; per cui la S si può pensare univocamente definita dalle possibili *associazioni* di tutti i risultati di questi N eventi — associazioni che, in sostanza, corrispondono alle possibili posizioni di un punto in uno spazio cartesiano N-dimensionale, in cui ognuna delle variabili x_k sia rappresentata su uno degli assi.

Visto che i valori x_k sono (per ipotesi) tra loro tutti statisticamente indipendenti, la probabilità di ottenere una particolare N-pla è data dal prodotto delle probabilità relative ad ogni singolo valore: e, se indichiamo con $f_k(x_k)$ la funzione densità di probabilità della generica x_k, la probabilità di ottenere un determinato valore per la S è data da

$$\mathrm{d}P \equiv g(S)\,\mathrm{d}S = \prod_{k=1}^{N} f_k(x_k)\,\mathrm{d}x_k$$

($\mathrm{d}S$ rappresenta un intorno (ipercubico) infinitesimo del punto S, di coordinate cartesiane $\{x_k\}$ nello spazio N-dimensionale prima descritto, corrispondente agli N intorni unidimensionali $\mathrm{d}x_k$ dei valori assunti dalle N variabili casuali x_k); e la densità di probabilità per la S vale quindi

$$g(S) = \prod_{k=1}^{N} f_k(x_k) \quad .$$

La funzione caratteristica di S è, dall'equazione di definizione (6.6),

$$\begin{aligned}\phi_S(t) &= \int_{-\infty}^{+\infty} e^{itS} g(S)\,\mathrm{d}S \\ &= \int_{-\infty}^{+\infty} \prod_{k=1}^{N} e^{itx_k} f_k(x_k)\,\mathrm{d}x_k\end{aligned}$$

ed infine

$$\boxed{\phi_S(t) = \prod_{k=1}^{N} \phi_k(t)} \tag{6.11}$$

Quindi *la funzione caratteristica della somma di N variabili casuali statisticamente indipendenti è pari al prodotto delle loro funzioni caratteristiche.*

6.4.1 Funzioni caratteristiche di variabili discrete

Invece della funzione caratteristica definita attraverso la (6.7), e che è una funzione complessa di variabile reale, talvolta, per variabili casuali discrete, viene usata una rappresentazione equivalente ricorrendo alla variabile complessa

$$z = e^{it} \quad .$$

Sostituendo questa definizione di z nella (6.7) si ottiene la *funzione caratteristica di variabile complessa*

$$\phi_x(z) \;=\; \sum\nolimits_k p_k\, z^{x_k} \;=\; E(z^x) \quad ,$$

che ha proprietà analoghe a quelle della funzione caratteristica di variabile reale $\phi_x(t)$. In particolare, definendo una variabile casuale w come somma di due altre variabili x e y discrete e tra loro indipendenti, la funzione caratteristica di variabile complessa $\phi_w(z)$ è ancora il prodotto delle due funzioni caratteristiche $\phi_x(z)$ e $\phi_y(z)$: infatti

$$\begin{aligned}
\phi_w(z) &= \sum\nolimits_{jk} \Pr(x_j)\ \Pr(y_k)\, z^{(x_j+y_k)} \\
&= \sum\nolimits_j \Pr(x_j)\, z^{x_j} \cdot \sum\nolimits_k \Pr(y_k)\, z^{y_k} \\
&= \phi_x(z) \cdot \phi_y(z) \quad ;
\end{aligned}$$

e, generalizzando per induzione completa, la somma S di un numero prefissato N di variabili casuali discrete e tutte tra loro indipendenti

$$S = \sum_{k=1}^{N} x_k$$

è anch'essa associata alla funzione caratteristica di variabile complessa

$$\phi_S(z) = \prod_{k=1}^{N} \phi_{x_k}(z) \quad .$$

Nel caso particolare, poi, in cui le N variabili provengano dalla stessa popolazione,

$$\phi_S(z) = [\phi_x(z)]^N \quad . \tag{6.12}$$

Cosa accade se il numero N di variabili casuali da sommare non è costante, *ma è anch'esso una variabile casuale* (ovviamente discreta)? In altre parole, vogliamo qui di seguito trovare la rappresentazione analitica della funzione caratteristica della *somma di un numero casuale di variabili casuali discrete, indipendenti ed aventi tutte la stessa distribuzione.* Supponiamo che la N sia associata ad una funzione caratteristica

$$\phi_N(z) \;=\; E(z^N) \;=\; \sum\nolimits_N \Pr(N)\, z^N \;; \tag{6.13}$$

la probabilità di ottenere un determinato valore per la S vale

$$\Pr(S) = \sum\nolimits_N \Pr(N)\ \Pr(S|N)$$

e di conseguenza la funzione caratteristica di variabile complessa associata alla S che, per definizione, è data dalla

$$\phi_S(z) \;=\; E(z^S) \;=\; \sum\nolimits_S \Pr(S)\, z^S$$

si potrà scrivere anche

$$\begin{aligned}\phi_S(z) &= \sum\nolimits_S z^S \cdot \sum\nolimits_N \Pr(N)\ \Pr(S|N)\\ &= \sum\nolimits_N \Pr(N) \cdot \sum\nolimits_S \Pr(S|N)\, z^S\\ &= \sum\nolimits_N \Pr(N) \cdot [\phi_x(z)]^N \;.\end{aligned}$$

Nell'ultimo passaggio si è sfruttato il fatto che la sommatoria su S rappresenta la speranza matematica di z^S *condizionata dall'avere assunto N un determinato valore*; rappresenta quindi la funzione caratteristica della S quando N ha un valore costante prefissato, che appunto è data dalla (6.12).

Ricordando poi la (6.13), la funzione caratteristica cercata è infine data dalla *funzione di funzione*

$$\boxed{\phi_S(z) = \phi_N[\phi_x(z)]} \tag{6.14}$$

È immediato riconoscere che, se N non è propriamente una variabile casuale e può assumere un unico valore N_0, essendo tutte le $\Pr(N)$ nulle meno $\Pr(N_0) = 1$,

$$\phi_N(z) = z^{N_0}$$

e

$$\phi_S(z) \;=\; \phi_N[\phi_x(z)] \;=\; [\phi_x(z)]^{N_0}$$

e la (6.14) ridiventa la meno generale (6.12).

6.5 Cambiamento di variabile casuale

Supponiamo sia nota la funzione $f(x)$ densità di probabilità della variabile casuale x; e sia y una nuova variabile casuale definita in funzione della x attraverso una qualche relazione matematica $y = y(x)$. Ci proponiamo di vedere come, da queste ipotesi, si possa ricavare la densità di probabilità $g(y)$ della nuova variabile y.

Supponiamo dapprima che la corrispondenza tra le due variabili continue sia *biunivoca*: ossia che la $y = y(x)$ sia una funzione monotona in senso stretto, crescente o decrescente, e di cui quindi esista la funzione inversa che indicheremo con $x = x(y)$; ed inoltre supponiamo che la $y(x)$ sia derivabile. Questo, dovendo risultare $y'(x) \neq 0$ in conseguenza dell'ipotesi fatta, implica che sia derivabile anche la $x(y)$ e che risulti

$$x'(y) = \frac{1}{y'[x(y)]} \ .$$

L'asserita biunivocità della corrispondenza tra le due variabili assicura che, se la prima è compresa in un intervallo infinitesimo di ampiezza $\mathrm{d}x$ centrato sul generico valore x, allora e solo allora la seconda è compresa in un intervallo di ampiezza $\mathrm{d}y = |y'(x)|\,\mathrm{d}x$ (il valore assoluto tiene conto del fatto che la $y(x)$ può essere sia crescente che decrescente) centrato attorno al valore $y = y(x)$. Questo a sua volta implica che le probabilità (infinitesime) degli eventi casuali consistenti nell'essere la x o la y appartenenti a tali intervalli debbano essere uguali: ossia che risulti

$$f(x)\,\mathrm{d}x = g(y)\,\mathrm{d}y = g(y)\,|y'(x)|\,\mathrm{d}x$$

identicamente per ogni x, il che è possibile soltanto se

$$g(y) = \frac{f(x)}{|y'(x)|} = \frac{f[x(y)]}{|y'[x(y)]|} = f[x(y)] \cdot |x'(y)| \ . \tag{6.15}$$

Se la relazione che lega y ad x non è invece biunivoca, i ragionamenti sono più complicati e devono essere fatti tenendo conto della natura della particolare funzione in esame; ad esempio, se

$$y = x^2 \qquad \text{e quindi} \qquad x = \pm\sqrt{y}$$

un particolare valore per la y corrisponde a *due* eventualità (mutuamente esclusive) per la x; perciò

$$g(y)\,\mathrm{d}y = [f(-\sqrt{y}) + f(\sqrt{y})]\,\mathrm{d}x$$

e quindi

$$g(y) = [f(-\sqrt{y}) + f(\sqrt{y})] \cdot |x'(y)| = \frac{f(-\sqrt{y}) + f(\sqrt{y})}{2\sqrt{y}} .$$

Per quello che riguarda la funzione generatrice dei momenti e la funzione caratteristica associate a variabili casuali definite l'una in funzione dell'altra, se ci limitiamo a considerare una *trasformazione lineare* del tipo $y = ax + b$, vale la

$$\begin{aligned} M_y(t) &= E\left(e^{ty}\right) \\ &= E\left[e^{t(ax+b)}\right] \\ &= e^{tb}\, E(e^{tax}) \end{aligned}$$

da cui infine ricaviamo la

$$\boxed{M_y(t) = e^{tb}\, M_x(at)} \tag{6.16}$$

per la funzione generatrice dei momenti; e potremmo ricavare l'analoga

$$\boxed{\phi_y(t) = e^{itb}\, \phi_x(at)} \tag{6.17}$$

per la funzione caratteristica (si confronti anche la funzione (6.5), prima usata per ricavare i momenti rispetto alla media, e che si può pensare ottenuta dalla (6.4) applicando alla variabile casuale una traslazione che ne porti il valore medio nell'origine).

6.6 I valori estremi di un campione

Sia x una variabile casuale continua, di cui siano note sia la funzione di frequenza $f(x)$ che la funzione di distribuzione $F(x)$; e sia disponibile un campione di dimensione N di valori *indipendenti* di questa variabile casuale. Supponiamo inoltre, una volta ottenuti tali valori, di averli disposti *in ordine crescente*: ovvero in modo che risulti $x_1 \leq x_2 \leq \cdots \leq x_N$. Vogliamo qui, come esercizio, determinare la funzione di frequenza *del generico di questi*

valori ordinati, x_i: funzione che verrà nel seguito identificata dal simbolo $f_i(x)$.

Supponiamo che x_i sia compreso nell'intervallo infinitesimo $[x, x + \mathrm{d}x]$; la scelta di un certo i divide naturalmente il campione (ordinato) in tre sottoinsiemi, ovvero:

1. x_i stesso, che può essere ottenuto (dall'insieme non ordinato dei valori originariamente a disposizione) in N maniere differenti; si sa inoltre che è compreso nell'intervallo $[x, x + \mathrm{d}x]$ — evento, questo, che avviene con probabilità $f(x)\,\mathrm{d}x$.

2. I primi $(i - 1)$ valori: questi possono essere ottenuti, dagli $N - 1$ elementi restanti dall'insieme non ordinato dei valori originari, in C_{i-1}^{N-1} modi distinti[2]; ognuno di essi è inoltre minore di x, e questo avviene con probabilità data da $F(x)$.

3. I residui $(N - i)$ valori: questi sono univocamente determinati dalle due scelte precedenti; inoltre ognuno di essi è maggiore di x, e questo avviene con probabilità $[1 - F(x)]$.

In definitiva, applicando i teoremi della probabilità totale e della probabilità composta, possiamo affermare che risulta

$$f_i(x)\,\mathrm{d}x \;=\; N \binom{N-1}{i-1} [F(x)]^{i-1} [1 - F(x)]^{N-i} f(x)\,\mathrm{d}x \;\; ; \tag{6.18}$$

in particolare, i valori estremi x_1 e x_N hanno densità di probabilità date da

$$f_1(x) \;=\; N\,[1 - F(x)]^{N-1} f(x)$$

e da

$$f_N(x) \;=\; N\,[F(x)]^{N-1} f(x) \;\; .$$

[2] C_K^N è il numero delle combinazioni di classe K di N oggetti; si veda in proposito il paragrafo A.6.

Capitolo 7

Variabili casuali pluridimensionali

Può avvenire che un evento casuale complesso E sia decomponibile in N eventi semplici E_i, ognuno dei quali a sua volta sia descrivibile mediante una variabile casuale x_i (che supporremo continua); le differenti modalità dell'evento E si possono allora associare univocamente alla N-pla dei valori delle x_i, ossia alla posizione di un punto in uno spazio cartesiano N-dimensionale.

7.1 Variabili casuali bidimensionali

Nel caso multidimensionale più semplice, $N = 2$, se supponiamo che la probabilità $\mathrm{d}P$ per la coppia di variabili casuali x ed y di trovarsi nell'intorno (infinitesimo) di una certo punto dello spazio bidimensionale sia proporzionale all'ampiezza dell'intorno stesso e dipenda dalla sua posizione, possiamo definire la *densità di probabilità* (o *funzione di frequenza*) *congiunta*, $f(x, y)$, attraverso la

$$\mathrm{d}P = f(x, y)\,\mathrm{d}x\,\mathrm{d}y \quad ;$$

e, analogamente a quanto fatto nel caso unidimensionale, definire poi attraverso di essa altre funzioni. Ad esempio la *funzione di distribuzione congiunta*,

$$F(x, y) = \int_{-\infty}^{x} \mathrm{d}u \int_{-\infty}^{y} \mathrm{d}v\; f(u, v)$$

che dà la probabilità di ottenere valori delle due variabili non superiori a quantità prefissate; le *funzioni di frequenza marginali*

$$g(x) = \int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}y \qquad \text{e} \qquad h(y) = \int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}x$$

che rappresentano la densità di probabilità di ottenere un dato valore per una delle due variabili *qualunque sia il valore assunto dall'altra*; ed infine le *funzioni di distribuzione marginali*

$$G(x) = \int_{-\infty}^{x} g(t)\,\mathrm{d}t = F(x,+\infty) \qquad \text{e} \qquad H(y) = \int_{-\infty}^{y} h(t)\,\mathrm{d}t = F(+\infty,y) \ .$$

La *condizione di normalizzazione* si potrà poi scrivere

$$F(+\infty,+\infty) = 1 \ .$$

Per un insieme di due variabili si possono poi definire le funzioni di frequenza *condizionate*, $\pi(x|y)$ e $\pi(y|x)$; esse rappresentano la densità di probabilità dei valori di una variabile quando già si conosce il valore dell'altra. Per definizione deve valere la

$$f(x,y)\,\mathrm{d}x\,\mathrm{d}y \ = \ g(x)\,\mathrm{d}x \cdot \pi(y|x)\,\mathrm{d}y \ = \ h(y)\,\mathrm{d}y \cdot \pi(x|y)\,\mathrm{d}x$$

per cui tra probabilità condizionate, marginali e congiunte valgono la

$$\pi(y|x) = \frac{f(x,y)}{g(x)} \qquad \text{e la} \qquad \pi(x|y) = \frac{f(x,y)}{h(y)} \ .$$

Due variabili casuali sono, come sappiamo, statisticamente indipendenti tra loro quando il fatto che una di esse abbia un determinato valore non altera le probabilità relative ai valori dell'altra: ovvero quando

$$\pi(x|y) = g(x) \qquad \text{e} \qquad \pi(y|x) = h(y) \ ; \tag{7.1}$$

e questo a sua volta implica che

$$f(x,y) = g(x) \cdot h(y) \tag{7.2}$$

Non è difficile poi, assunta vera la (7.2), giungere alla (7.1); in definitiva:

> *Due variabili casuali continue sono statisticamente indipendenti tra loro se e solo se la densità di probabilità congiunta è fattorizzabile nel prodotto delle funzioni marginali.*

7.1.1 Momenti, funzione caratteristica e funzione generatrice

Analogamente a quanto fatto per le variabili casuali unidimensionali, in uno spazio degli eventi bidimensionale in cui rappresentiamo le due variabili $\{x, y\}$ aventi densità di probabilità congiunta $f(x, y)$, si può definire la speranza matematica (o valore medio) di una qualunque funzione $\psi(x, y)$ come

$$E[\psi(x, y)] = \int_{-\infty}^{+\infty} \mathrm{d}x \int_{-\infty}^{+\infty} \mathrm{d}y \; \psi(x, y) \, f(x, y) \;\; ;$$

i momenti rispetto all'origine come

$$\lambda_{mn} = E\left(x^m \, y^n\right)$$

e quelli rispetto alla media come

$$\mu_{mn} = E\left[(x - \lambda_{10})^m (y - \lambda_{01})^n\right] \;\; .$$

Risulta ovviamente:

$$\begin{aligned}
\lambda_{00} &\equiv 1 \\
\lambda_{10} &= E(x) \\
\lambda_{01} &= E(y) \\
\mu_{20} &= E\left\{\left[x - E(x)\right]^2\right\} = \mathrm{Var}(x) \\
\mu_{02} &= E\left\{\left[y - E(y)\right]^2\right\} = \mathrm{Var}(y) \\
\mu_{11} &= E\left\{\left[x - E(x)\right]\left[y - E(y)\right]\right\}
\end{aligned}$$

La quantità μ_{11} si chiama anche *covarianza* di x ed y; si indica generalmente col simbolo $\mathrm{Cov}(x, y)$, e di essa ci occuperemo più in dettaglio nell'appendice C (almeno per quel che riguarda le variabili discrete). Un'altra grandezza collegata alla covarianza è il cosiddetto *coefficiente di correlazione lineare*, che si indica col simbolo r_{xy} (o, semplicemente, con r): è definito come

$$r_{xy} = \frac{\mu_{11}}{\sqrt{\mu_{20}\,\mu_{02}}} = \frac{\mathrm{Cov}(x, y)}{\sigma_x \, \sigma_y} \;\; ,$$

e si tratta di una grandezza adimensionale compresa, come vedremo, nell'intervallo $[-1, +1]$. Anche del coefficiente di correlazione lineare ci occuperemo estesamente più avanti, e sempre nell'appendice C.

La funzione caratteristica per due variabili, che esiste sempre, è la

$$\phi_{xy}(u, v) = E\left[e^{i(ux+vy)}\right] \quad ;$$

se poi esistono tutti i momenti, vale anche la

$$\left.\frac{\partial^{m+n}\phi_{xy}}{\partial u^m\,\partial v^n}\right|_{\substack{u=0\\v=0}} = (i)^{m+n}\,\lambda_{mn} \quad .$$

La funzione generatrice, che esiste solo se tutti i momenti esistono, è poi definita come

$$M_{xy}(u, v) = E\left[e^{(ux+vy)}\right]$$

e per essa vale la

$$\left.\frac{\partial^{m+n}M_{xy}}{\partial u^m\,\partial v^n}\right|_{\substack{u=0\\v=0}} = \lambda_{mn} \quad .$$

7.1.2 Cambiamento di variabile casuale

Supponiamo di definire due nuove variabili casuali u e v per descrivere un evento casuale collegato a due variabili continue x ed y; e questo attraverso due funzioni

$$u = u(x, y) \qquad \text{e} \qquad v = v(x, y) \quad .$$

Se la corrispondenza tra le due coppie di variabili è biunivoca, esistono le funzioni inverse

$$x = x(u, v) \qquad \text{e} \qquad y = y(u, v) \quad ;$$

se inoltre esistono anche le derivate parziali prime della x e della y rispetto alla u ed alla v, esiste anche non nullo il *determinante Jacobiano*

$$\frac{\partial(x, y)}{\partial(u, v)} = \det\begin{Vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\[2ex] \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{Vmatrix}$$

dotato della proprietà che

$$\frac{\partial(x,y)}{\partial(u,v)} = \left[\frac{\partial(u,v)}{\partial(x,y)}\right]^{-1}$$

In tal caso, dalla richiesta di invarianza della probabilità sotto il cambiamento di variabili,

$$f(x,y)\,\mathrm{d}x\,\mathrm{d}y = g(u,v)\,\mathrm{d}u\,\mathrm{d}v$$

si ottiene la funzione densità di probabilità congiunta per u e v, che è legata alla $f(x,y)$ dalla

$$g(u,v) = f\left[x(u,v), y(u,v)\right] \cdot \left|\frac{\partial(x,y)}{\partial(u,v)}\right| \tag{7.3}$$

7.1.3 Applicazione: il rapporto di due variabili casuali indipendenti

Come esempio, consideriamo due variabili casuali x ed y indipendenti tra loro e di cui si conoscano le funzioni di frequenza, rispettivamente $f(x)$ e $g(y)$; e si sappia inoltre che la y non possa essere nulla. Fatte queste ipotesi, useremo la formula precedente per calcolare la funzione di frequenza $\varphi(u)$ della variabile casuale u rapporto tra x ed y. Definite

$$u = \frac{x}{y} \qquad \text{e} \qquad v = y \ ,$$

la corrispondenza tra le coppie di variabili è biunivoca; e le funzioni inverse sono la

$$x = uv \qquad \text{e la} \qquad y = v \ .$$

Le funzioni di frequenza congiunte delle due coppie di variabili sono, ricordando la (7.2) e la (7.3)

$$f(x,y) = f(x)\,g(y) \qquad \text{e} \qquad \varphi(u,v) = f(x)\,g(y)\left|\frac{\partial(x,y)}{\partial(u,v)}\right|$$

rispettivamente; e, calcolando le derivate parziali,

$$\frac{\partial(x,y)}{\partial(u,v)} = \det\begin{Vmatrix} v & u \\ 0 & 1 \end{Vmatrix}$$

per cui

$$\varphi(u,v) = f(uv)\,g(v)\,|v| \ .$$

In conclusione, la funzione di distribuzione della sola u (la funzione marginale) è la

$$\varphi(u) \ = \ \int_{-\infty}^{+\infty} \varphi(u,v)\,\mathrm{d}v \ = \ \int_{-\infty}^{+\infty} f(uv)\,g(v)\,|v|\,\mathrm{d}v \tag{7.4}$$

7.1.4 Applicazione: il decadimento debole della Λ^0

La particella elementare Λ^0 decade, attraverso processi governati dalle interazioni deboli, nei due canali

$$\Lambda^0 \to p + \pi^- \qquad \text{e} \qquad \Lambda^0 \to n + \pi^0 \ ;$$

il suo decadimento è quindi un evento casuale che può essere descritto dalle due variabili c (carica del nucleone nello stato finale, 1 o 0 rispettivamente) e t (tempo di vita della Λ^0).

La teoria (confermata dagli esperimenti) richiede che la legge di decadimento sia la stessa per entrambi gli stati finali, ovvero esponenziale con la stessa vita media τ; e che il cosiddetto *branching ratio*, cioè il rapporto delle probabilità di decadimento nei due canali citati, sia indipendente dal tempo di vita e valga

$$\frac{\Pr\left(\Lambda^0 \to p + \pi^-\right)}{\Pr\left(\Lambda^0 \to n + \pi^0\right)} = 2 \ .$$

In altre parole, le probabilità marginali e condizionate per le due variabili (una discreta, l'altra continua) devono essere: per la c

$$g(1) = g(1|t) = \frac{2}{3} \qquad \text{e} \qquad g(0) = g(0|t) = \frac{1}{3}$$

o, in maniera compatta,

$$g(c) = g(c|t) = \frac{c+1}{3} \ ;$$

per il tempo di vita t,

$$h(t) = h(t|0) = h(t|1) = \frac{1}{\tau}\,e^{-\frac{t}{\tau}} \ .$$

La probabilità congiunta delle due variabili casuali è, infine,

$$f(c,t) = g(c)\cdot h(t) = \frac{c+1}{3\tau}\,e^{-\frac{t}{\tau}} \ .$$

7.1.5 Applicazione: il decadimento debole K^0_{e3}

I decadimenti K^0_{e3} consistono nei due processi deboli di decadimento del mesone K^0

$$K^0 \to e^- + \pi^+ + \bar{\nu}_e \qquad \text{e} \qquad K^0 \to e^+ + \pi^- + \nu_e \ ;$$

essi possono essere descritti dalle due variabili casuali c (carica dell'elettrone nello stato finale, $c = \mp 1$) e t (tempo di vita del K^0). La teoria, sulla base della cosiddetta "ipotesi $\Delta Q = \Delta S$"), prevede che la funzione di frequenza congiunta sia

$$f(t,c) = \frac{N(t,c)}{\sum_c \int_0^{+\infty} N(t,c)\,\mathrm{d}t}$$

ove si è indicato con $N(t,c)$ la funzione

$$N(t,c) = e^{-\lambda_1 t} + e^{-\lambda_2 t} + 2c\cos(\omega t)\, e^{-\frac{\lambda_1+\lambda_2}{2}t} \quad : \tag{7.5}$$

nella (7.5), le costanti λ_1 e λ_2 rappresentano gli inversi delle vite medie dei mesoni K^0_1 e K^0_2, mentre ω corrisponde alla differenza tra le loro masse.

Si vede immediatamente che la (7.5) non è fattorizzabile: quindi le due variabili *non sono* tra loro indipendenti. In particolare, le probabilità marginali sono date dalla

$$h(t) = \frac{\lambda_1\lambda_2}{\lambda_1+\lambda_2}\left(e^{-\lambda_1 t} + e^{-\lambda_2 t}\right)$$

e dalla

$$g(c) = \frac{1}{2} + \frac{4\,c\,\lambda_1\lambda_2}{(\lambda_1+\lambda_2)^2 + 4\omega^2} \ ;$$

mentre le probabilità condizionate sono, da definizione, la

$$h(t|c) = \frac{f(t,c)}{g(c)} \qquad \text{e la} \qquad g(c|t) = \frac{f(t,c)}{h(t)} \ .$$

7.1.6 Ancora sui valori estremi di un campione

Come esempio, e ricordando il paragrafo 6.6, calcoliamo la densità di probabilità congiunta dei due valori estremi x_1 e x_N di un campione *ordinato* e di dimensione N; questo sotto l'ipotesi che i dati appartengano a una popolazione avente funzione di frequenza $f(x)$ e funzione di distribuzione $F(x)$ entrambe note.

x_1 può provenire dal campione a disposizione in N maniere distinte; una volta noto x_1, poi, x_N può essere scelto in $(N-1)$ modi diversi; e, infine, ognuno dei dati restanti è compreso tra x_1 e x_N: e questo avviene con probabilità $[F(x_N) - F(x_1)]$. Ripetendo i ragionamenti del paragrafo 6.6, si ricava

$$f(x_1, x_N) \;=\; N\,(N-1)\,[F(x_N) - F(x_1)]^{N-2}\, f(x_1)\, f(x_N) \tag{7.6}$$

che *non è fattorizzabile*: quindi i valori minimo e massimo di un campione *non* sono indipendenti tra loro. Introducendo le variabili ausiliarie

$$\begin{cases} \xi = N \cdot F(x_1) \\ \eta = N \cdot [1 - F(x_N)] \end{cases} \qquad \text{con} \qquad \begin{cases} \mathrm{d}\xi = N\, f(x_1)\, \mathrm{d}x_1 \\ \mathrm{d}\eta = -N\, f(x_N)\, \mathrm{d}x_N \end{cases}$$

ed essendo $F(x_N) - F(x_1)$ identicamente uguale a $1 - [1 - F(x_N)] - F(x_1)$, dalla (7.6) si ricava

$$f(\xi, \eta) \;=\; \frac{N-1}{N} \left(1 - \frac{\xi + \eta}{N}\right)^{N-2}$$

che, ricordando il limite notevole

$$\lim_{x \to +\infty} \left(1 + \frac{k}{x}\right)^{x} \;=\; e^{k} \;,$$

asintoticamente diventa

$$f(\xi, \eta) \quad \xrightarrow{N \to \infty} \quad e^{-(\xi+\eta)} \;\equiv\; e^{-\xi}\, e^{-\eta} \;.$$

Quindi ξ ed η (come anche di conseguenza x_1 e x_N) sono statisticamente indipendenti *solo asintoticamente*, all'aumentare indefinito della dimensione del campione.

7.2 Cenni sulle variabili casuali in più di due dimensioni

Estendendo a spazi cartesiani a più di due dimensioni il concetto di densità di probabilità, possiamo pensare di associare ad un evento casuale E descritto da N variabili continue $x_1, x_2, \ldots, x_N$ una funzione f di tutte queste variabili; la probabilità che, simultaneamente, ognuna di esse cada in un intervallo infinitesimo attorno ad un determinato valore sarà poi data da

$$\mathrm{d}P = f(x_1, x_2, \ldots, x_N)\, \mathrm{d}x_1\, \mathrm{d}x_2 \cdots \mathrm{d}x_N \;.$$

Usando la legge della probabilità totale e la definizione dell'operazione di integrazione, è poi immediato riconoscere che la probabilità dell'evento casuale consistente nell'essere ognuna delle x_i compresa in un determinato intervallo finito $[a_i, b_i]$ è data da

$$P = \int_{a_1}^{b_1} \mathrm{d}x_1 \int_{a_2}^{b_2} \mathrm{d}x_2 \cdots \int_{a_N}^{b_N} \mathrm{d}x_N \cdot f(x_1, x_2, \ldots, x_N) \ .$$

Similmente, poi, se consideriamo il sottoinsieme delle prime M variabili x_i (con $M < N$), la probabilità che ognuna di esse cada all'interno di intervallini infinitesimi attorno ad una M-pla di valori prefissati, *indipendentemente* dal valore assunto dalle altre $N - M$ variabili, è data da

$$\begin{aligned} \mathrm{d}P &\equiv f^M(x_1, \ldots, x_M)\, \mathrm{d}x_1 \cdots \mathrm{d}x_M \\ &= \mathrm{d}x_1 \cdots \mathrm{d}x_M \int_{-\infty}^{+\infty} \mathrm{d}x_{M+1} \int_{-\infty}^{+\infty} \mathrm{d}x_{M+2} \cdots \int_{-\infty}^{+\infty} \mathrm{d}x_N \cdot f(x_1, x_2, \ldots, x_N) \end{aligned}$$

dove gli integrali definiti sulle $N-M$ variabili che non interessano si intendono estesi a tutto l'asse reale; potendosi senza perdere in generalità assumere che tale sia il loro dominio di esistenza, definendo eventualmente la f come identicamente nulla al di fuori del reale intervallo di variabilità se esse fossero limitate.

La f^M definita dalla equazione precedente prende il nome di *densità di probabilità marginale* delle M variabili casuali $x_1, \ldots, x_M$; infine la condizione di normalizzazione si scriverà

$$\int_{-\infty}^{+\infty} \mathrm{d}x_1 \int_{-\infty}^{+\infty} \mathrm{d}x_2 \cdots \int_{-\infty}^{+\infty} \mathrm{d}x_N \cdot f(x_1, x_2, \ldots, x_N) = 1 \ .$$

Definendo, analogamente a quanto fatto nel paragrafo 7.1, la densità di probabilità delle M variabili casuali x_j (con $j = 1, 2, \ldots, M$ e $M < N$) *condizionata* dai valori assunti dalle restanti $N - M$ variabili attraverso la

$$f(x_1, x_2, \ldots, x_M | x_{M+1}, x_{M+2}, \ldots, x_N) = \frac{f(x_1, x_2, \ldots, x_N)}{f^M(x_{M+1}, x_{M+2}, \ldots, x_N)} \tag{7.7}$$

il concetto di indipendenza statistica può facilmente essere generalizzato a *sottogruppi* di variabili: diremo che le M variabili x_j sono statisticamente indipendenti dalle restanti $N - M$ quando la probabilità che le $x_1, x_2, \ldots, x_M$ assumano determinati valori non dipende dai valori assunti dalle $x_{M+1}, x_{M+2}, \ldots, x_N$ — e dunque quando la densità condizionata (7.7) è identicamente uguale alla densità marginale $f^M(x_1, x_2, \ldots, x_M)$.

Esaminando la (7.7) si può facilmente capire come, perché questo avvenga, occorra e basti che la densità di probabilità complessiva sia *fattorizzabile* nel prodotto di due termini: il primo dei quali sia funzione solo delle prime M variabili ed il secondo dei quali dipenda soltanto dalle altre $N - M$; ovviamente ognuno dei fattori coincide con le probabilità marginali, per cui la condizione è espressa matematicamente dalla formula

$$f(x_1, \ldots, x_N) = f^M(x_1, \ldots, x_M) \cdot f^M(x_{M+1}, \ldots, x_N)$$

e, in particolare, le variabili sono *tutte* indipendenti tra loro se e solo se risulta

$$f(x_1, x_2, \ldots, x_N) = f^M(x_1) \cdot f^M(x_2) \cdots f^M(x_N) \quad . \tag{7.8}$$

Nel caso che esista un differente insieme di N variabili y_i in grado di descrivere lo stesso fenomeno casuale E, il requisito che la probabilità di realizzarsi di un qualunque sottoinsieme dei possibili risultati (l'integrale definito, su una qualunque regione Ω dello spazio ad N dimensioni, della funzione densità di probabilità) sia invariante per il cambiamento delle variabili di integrazione, porta infine a ricavare la formula di trasformazione delle densità di probabilità per il cambiamento di variabili casuali nel caso multidimensionale:

$$f(y_1, y_2, \ldots, y_N) = f(x_1, x_2, \ldots, x_N) \cdot \left| \frac{\partial(x_1, x_2, \ldots, x_N)}{\partial(y_1, y_2, \ldots, y_N)} \right| \tag{7.9}$$

dove con il simbolo

$$\left| \frac{\partial(x_1, x_2, \ldots, x_N)}{\partial(y_1, y_2, \ldots, y_N)} \right|$$

si è indicato il valore assoluto del determinante Jacobiano delle x rispetto alle y:

$$\frac{\partial(x_1, x_2, \ldots, x_N)}{\partial(y_1, y_2, \ldots, y_N)} = \det \left\| \begin{matrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_N} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_N} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial x_N}{\partial y_1} & \frac{\partial x_N}{\partial y_2} & \cdots & \frac{\partial x_N}{\partial y_N} \end{matrix} \right\|$$

che esiste sempre non nullo se la trasformazione tra l'insieme delle funzioni x_i e quello delle funzioni y_i è biunivoca; e che gode, sempre in questa ipotesi, della proprietà che

$$\frac{\partial(y_1, y_2, \ldots, y_N)}{\partial(x_1, x_2, \ldots, x_N)} = \left[\frac{\partial(x_1, x_2, \ldots, x_N)}{\partial(y_1, y_2, \ldots, y_N)}\right]^{-1} .$$

Capitolo 8

Esempi di distribuzioni teoriche

In questo capitolo presentiamo alcune funzioni teoriche che rappresentano densità di probabilità di variabili casuali unidimensionali (continue e discrete) che hanno importanza per la fisica.

8.1 La distribuzione uniforme

Il caso più semplice, dal punto di vista teorico, è quello di una variabile casuale x che possa assumere solo valori compresi in un intervallo finito avente estremi costanti prefissati, $[a, b]$; e ivi con probabilità uguale per ogni punto[1].

Questo implica che la densità di probabilità $f(x)$ di questa variabile debba essere definita come

$$\begin{cases} f(x) = 0 & \text{per } x < a \text{ e per } x > b; \\ f(x) = \dfrac{1}{b-a} = \text{cost.} & \text{per } a \leq x \leq b. \end{cases}$$

(il valore costante di $f(x)$ quando $x \in [a, b]$ è fissato dalla condizione di

[1] La frase è intuitiva, ma impropria; si intende qui che la probabilità, per la variabile casuale, di cadere in un intervallino di ampiezza (infinitesima) prefissata $\mathrm{d}x$ e centrato su un qualsivoglia punto del dominio di definizione, ha sempre lo stesso valore.

normalizzazione). La funzione di distribuzione $F(x)$ della x è data da

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \begin{cases} 0 & \text{per } x < a; \\ \dfrac{x-a}{b-a} & \text{per } a \le x \le b; \\ 1 & \text{per } x > b. \end{cases}$$

I valori della media e della varianza della variabile casuale x, come si può facilmente calcolare, valgono

$$\begin{cases} E(x) = \dfrac{a+b}{2} \\ \mathrm{Var}(x) = \dfrac{(b-a)^2}{12} \end{cases} \tag{8.1}$$

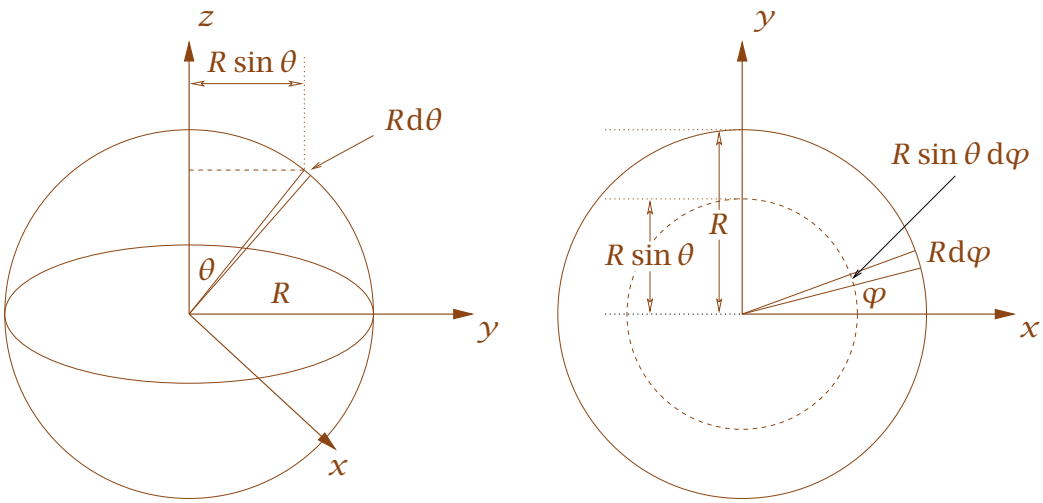
Per vedere una prima applicazione pratica della distribuzione uniforme, supponiamo di misurare una grandezza fisica usando uno strumento *digitale*: ad esempio una bilancia con sensibilità inversa di 1 grammo. Se, per semplicità, escludiamo la presenza di errori sistematici, il fatto che il display digitale indichi (ad esempio) 10 grammi significa solo che la massa dell'oggetto pesato è maggiore o uguale a questo valore e minore di 11 grammi[2]; e tutti i valori interni a questo intervallo ci appaiono inoltre come ugualmente plausibili. Per questo motivo, viste le (8.1), in casi di questo genere si attribuisce all'oggetto pesato una massa di 10.5 g con un errore di $1/\sqrt{12} \approx 0.3$ g.

8.1.1 Applicazione: decadimento del π^0

Esistono, nella fisica, variabili casuali che seguono la distribuzione uniforme: ad esempio, se una particella instabile non dotata di momento angolare intrinseco (come il mesone π^0), originariamente in quiete in un punto (che supporremo sia l'origine degli assi coordinati), decade, i prodotti di decadimento si distribuiscono uniformemente tra le varie direzioni possibili; sostanzialmente per motivi di simmetria, perché non esiste nessuna direzione privilegiata nel sistema di riferimento considerato (ovverosia nessuna caratteristica intrinseca del fenomeno che possa servire per definire uno, o più d'uno, degli assi coordinati).

[2] La maggior parte degli strumenti digitali *tronca* il valore mostrato e si comporta appunto in questo modo; altri invece *arrotondano* il risultato e, se questo fosse il caso, vorrebbe dire che la massa dell'oggetto pesato è maggiore o uguale a 9.5 g e minore di 10.5 g.

Figura 8a - Le aree elementari sulla superficie di una sfera di raggio R (in coordinate polari).



Con riferimento alla figura 8a, pensiamo introdotto un sistema di coordinate polari $\{R, \theta, \varphi\}$: l'elemento infinitesimo di area, $\mathrm{d}S$, sulla sfera di raggio R, che corrisponde a valori della *colatitudine* compresi tra θ e $\theta + \mathrm{d}\theta$, e dell'*azimuth* compresi tra φ e $\varphi + \mathrm{d}\varphi$, è uno pseudorettangolo di lati $R\,\mathrm{d}\theta$ ed $R \sin\theta\,\mathrm{d}\varphi$; quindi, a meno del segno,

$$|\mathrm{d}S| \;=\; R^2 \sin\theta\,\mathrm{d}\theta\,\mathrm{d}\varphi \;=\; -R^2\,\mathrm{d}(\cos\theta)\,\mathrm{d}\varphi$$

mentre l'angolo solido corrispondente vale

$$\mathrm{d}\Omega \;=\; \frac{|\mathrm{d}S|}{R^2} \;=\; \sin\theta\,\mathrm{d}\theta\,\mathrm{d}\varphi \;=\; -\mathrm{d}(\cos\theta)\,\mathrm{d}\varphi \;\;.$$

L'asserita uniformità nell'emissione dei prodotti di decadimento si traduce nella condizione che la probabilità, per essi, di essere contenuti in un qualsiasi angolo solido, sia proporzionale all'ampiezza di quest'ultimo:

$$\mathrm{d}P \;=\; K\,\mathrm{d}\Omega \;=\; K'\,\mathrm{d}(\cos\theta)\,\mathrm{d}\varphi$$

(ove K e K' sono due opportune costanti); ovverosia richiede che le due variabili casuali

$$u = \cos\theta \qquad\qquad \text{e} \qquad\qquad v = \varphi$$

abbiano distribuzione uniforme, e siano inoltre statisticamente indipendenti tra loro (questo in conseguenza dell'equazione (7.8)).

8.1.2 Applicazione: generazione di numeri casuali con distribuzione data

Supponiamo che la variabile casuale x abbia densità di probabilità $f(x)$ e funzione di distribuzione $F(x)$: vogliamo ora dimostrare che la variabile casuale $y = F(x)$ è distribuita uniformemente nell'intervallo $[0, 1]$ *qualunque* siano $f(x)$ e $F(x)$. Chiaramente y può appartenere solo a tale intervallo; ed inoltre, essendo funzione integrale di $f(x)$, è dotata della proprietà di essere continua e derivabile in tutto l'insieme di definizione e con derivata prima data da

$$y' = F'(x) = f(x)$$

così che, ricordando l'equazione (6.15), la densità di probabilità della nuova variabile y è data (ove $f(x)$ non sia nulla) dalla

$$g(y) = \frac{f(x)}{y'(x)} = \frac{f(x)}{f(x)} \equiv 1$$

come volevamo dimostrare.

Supponiamo sia nota la densità di probabilità $f(x)$ di una qualche variabile casuale x; e che si vogliano ottenere dei numeri che si presentino secondo una legge di probabilità data appunto da questa $f(x)$. I moderni calcolatori numerici sono in grado di generare sequenze di numeri casuali[3] che hanno distribuzione uniforme in un intervallo dipendente dall'implementazione dell'algoritmo, e che possono a loro volta essere usati per produrre numeri casuali con distribuzione uniforme nell'intervallo $[0, 1]$; se y è uno di tali numeri, e se si è in grado di invertire, numericamente od analiticamente, la funzione di distribuzione $F(x)$ della variabile casuale x, i numeri

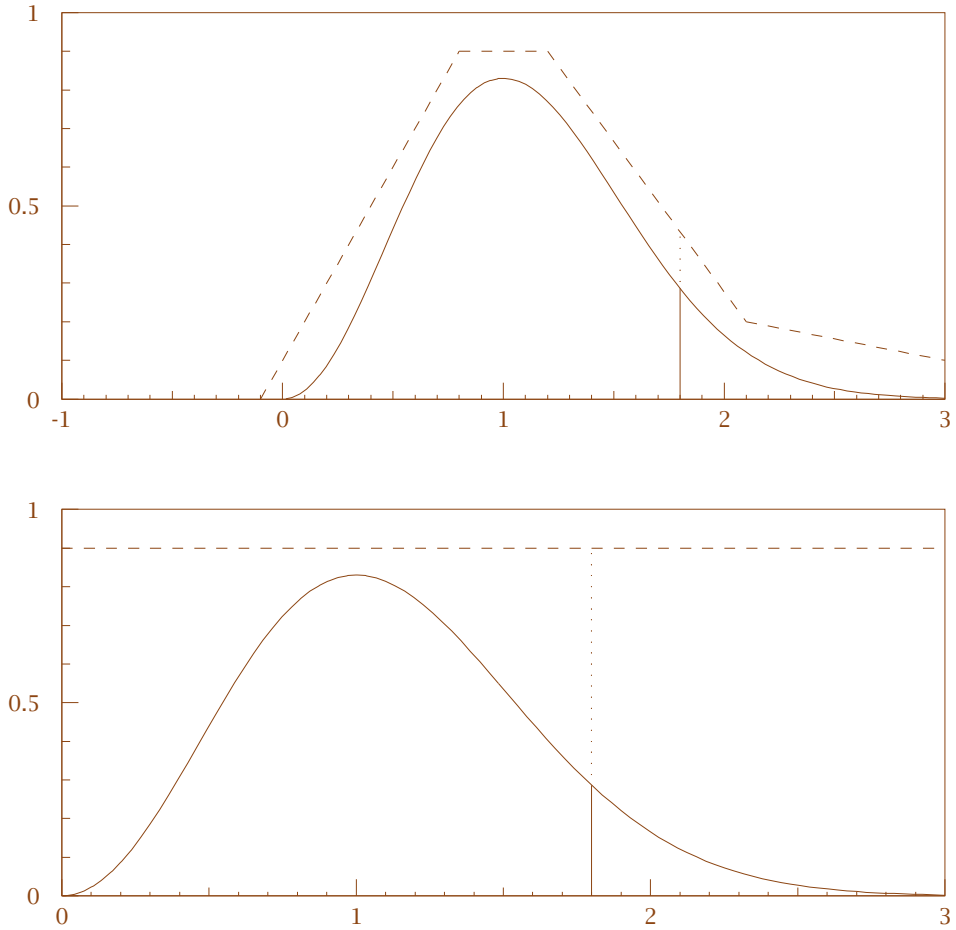
$$x = F^{-1}(y)$$

hanno densità di probabilità data da $f(x)$, come appunto richiesto.

Generalmente le funzioni di distribuzione $F(x)$ non si sanno invertire per via analitica; un metodo numerico spesso impiegato, e che richiede la sola preventiva conoscenza della $f(x)$ (quindi non bisogna nemmeno saper calcolare la $F(x)$, per non parlare della sua inversa) è illustrato qui di seguito (*metodo dei rigetti*). Si faccia riferimento alla figura 8b: sia x limitata in un intervallo chiuso $[x_{\min}, x_{\max}]$ (nella figura, $x_{\min} = 0$ e $x_{\max} = 3$); e si conosca

[3] O meglio *pseudo-casuali*: ovverosia prodotti da un algoritmo ripetibile, quindi non propriamente "imprevedibili"; ma in modo tale che le loro proprietà statistiche siano indistinguibili da quelle di una sequenza casuale propriamente detta.

Figura 8b - La scelta di un numero a caso con distribuzione prefissata mediante tecniche numeriche (la densità di probabilità è la stessa della figura 4d); la funzione maggiorante è una spezzata (superiormente) o la retta $y = 0.9$ (inferiormente).



una funzione $y = \varphi(x)$ *maggiorante* della $f(x)$, ossia una funzione che risulti comunque non inferiore alla f per qualunque $x \in [x_{\min}, x_{\max}]$.

Nel caso si sappia scegliere, sul piano $\{x, y\}$, un punto con distribuzione uniforme nella parte di piano limitata inferiormente dall'asse delle ascisse, superiormente dalla funzione $y = \varphi(x)$, e, lateralmente, dalle due rette di equazione $x = x_{\min}$ ed $x = x_{\max}$, basta accettare tale punto se la sua ordinata risulta non superiore alla corrispondente $f(x)$; e rigettarlo in caso contrario, iterando il procedimento fino a che la condizione precedente non è soddisfatta: le ascisse x dei punti accettati seguono la funzione di distribuzione $f(x)$.

Infatti, i punti accettati saranno distribuiti uniformemente nella parte di piano limitata dalla $y = f(x)$; quindi, in un intervallino infinitesimo centrato su una particolare x, vi sarà un numero di punti accettati proporzionale all'altezza della curva sopra di esso — ovverosia ogni ascissa x viene accettata con densità di probabilità che è proprio $f(x)$.

La scelta, infine, di un punto che sia distribuito uniformemente nella parte di piano limitata dalla funzione $y = \varphi(x)$ si sa sicuramente effettuare se $\varphi(x)$ è stata scelta in modo che si sappia invertire la sua funzione integrale

$$\Phi(x) = \int_{-\infty}^{x} \varphi(t)\,\mathrm{d}t$$

così che si possa associare, a qualsiasi valore A compreso tra 0 e $\Phi(+\infty)$, quella $x = \Phi^{-1}(A)$ che lascia alla propria sinistra un'area A al di sotto della funzione $y = \varphi(x)$ (una scelta banale è quella di prendere come maggiorante una retta, o meglio una spezzata — come illustrato nella figura 8b).

In tal caso basta scegliere un numero A con distribuzione uniforme tra i limiti $\Phi(x_{\min})$ e $\Phi(x_{\max})$; trovare la $x = \Phi^{-1}(A)$ che soddisfa la condizione precedente; ed infine scegliere una y con distribuzione uniforme tra 0 e $\varphi(x)$. Non è difficile rendersi conto che il punto (x, y) soddisfa alla condizione richiesta di essere distribuito uniformemente nella parte del semipiano $y > 0$ limitata superiormente dalla funzione $y = \varphi(x)$: a questo punto non rimane che calcolare la $f(x)$ ed accettare x se $y \leq f(x)$.

Se proprio non si è in grado di effettuare una scelta migliore, anche una retta del tipo $y = \text{cost.}$ può andar bene; basta tener presente che l'algoritmo viene sfruttato tanto più efficacemente quanto più $y = \varphi(x)$ è vicina alla $f(x)$ (in tal caso il numero di rigetti è minore).

Per questo motivo, una scelta del tipo $\varphi(x) = \text{cost.}$ è assolutamente da evitare se la $f(x)$ è sensibilmente diversa da zero solo in una parte ristretta dell'intervallo di definizione (perché in tal caso la scelta uniforme di x all'interno dell'area su detta ci farebbe trascorrere gran parte del tempo ad esa-

minare valori poco probabili rispetto alla $\varphi(x)$, che vengono in conseguenza quasi sempre rifiutati).

8.1.3 Esempio: valori estremi di un campione di dati a distribuzione uniforme

Come ulteriore esempio, applichiamo le conclusioni dei paragrafi 6.6 e 7.1.6 ad un campione di valori proveniente da una distribuzione uniforme. Usando le espressioni per $f(x)$ e $F(x)$ che conosciamo, ed essendo[4]

$$1 - F(x) = \frac{b-x}{b-a} ,$$

la (6.18) diventa

$$f_i(x) = N \binom{N-1}{i-1} \frac{(x-a)^{i-1}(b-x)^{N-i}}{(b-a)^N}$$

e, in particolare, per i due valori minimo e massimo presenti nel campione le densità di probabilità si scrivono

$$f_1(x) = N \frac{(b-x)^{N-1}}{(b-a)^N}$$

e

$$f_N(x) = N \frac{(x-a)^{N-1}}{(b-a)^N} .$$

Come conseguenza, la speranza matematica di x_N vale

$$\begin{aligned} E(x_N) &= \int_a^b x \cdot f_N(x)\,\mathrm{d}x \\ &= \frac{N}{(b-a)^N} \int_a^b [a + (x-a)]\,(x-a)^{N-1}\,\mathrm{d}x \\ &= \frac{N}{(b-a)^N} \left[a\,\frac{(x-a)^N}{N} + \frac{(x-a)^{N+1}}{N+1} \right]_a^b \\ &= a + \frac{N}{N+1}(b-a) \\ &= b - \frac{1}{N+1}(b-a) . \end{aligned}$$

[4] All'interno dell'intervallo $[a, b]$; per brevità ometteremo, qui e nel seguito, di specificare che, al di fuori di questo intervallo, le densità di probabilità sono identicamente nulle e le funzioni di distribuzione valgono o zero od uno.

Allo stesso modo si troverebbe

$$E(x_1) \;=\; a + \frac{1}{N+1}\,(b-a) \;\;;$$

e, per il generico x_i,

$$E(x_i) \;=\; a + \frac{i}{N+1}\,(b-a) \;\;.$$

Dopo gli opportuni calcoli, si potrebbero ricavare anche le varianze rispettive: che valgono

$$\mathrm{Var}(x_1) \;=\; \mathrm{Var}(x_N) \;=\; \frac{N}{(N+1)^2\,(N+2)}\,(b-a)^2$$

e

$$\mathrm{Var}(x_i) \;=\; \frac{i\cdot(N-i+1)}{(N+1)^2\,(N+2)}\,(b-a)^2 \;\;.$$

È immediato calcolare la speranza matematica della semisomma del più piccolo e del più grande valore presenti nel campione

$$d \;=\; \frac{x_1+x_N}{2}$$

che vale

$$E(d) \;=\; \frac{E(x_1)+E(x_N)}{2} \;=\; \frac{a+b}{2} \;\;;$$

come pure quella del cosiddetto *range*,

$$R \;=\; x_N - x_1$$

per il quale

$$E(R) \;=\; E(x_N) - E(x_1) \;=\; (b-a)\left(1-\frac{2}{N+1}\right) \;\;.$$

Per il calcolo delle varianze, invece, si deve ricorrere alla distribuzione congiunta (7.6), dalla quale si può ricavare

$$\mathrm{Var}(d) \;=\; \frac{(b-a)^2}{2\,(N+1)\,(N+2)}$$

e

$$\mathrm{Var}(R) \;=\; \frac{2\,(N-1)}{(N+1)^2\,(N+2)}\,(b-a)^2 \;\;.$$