1

Richiami di teoria delle probabilità e di statistica

1.1. Variabile aleatoria (o casuale)

Si dice che una variabile segue la legge del caso (variabile aleatoria o casuale) quando i valori che essa assume dipendono da infinite cause.
Nella realtà non esistono variabili che dipendono da infinite cause, ma variabili che dipendono da moltissime cause.
Si possono distinguere due tipi di variabili aleatorie:

a) variabili aleatorie discrete;

b) variabili aleatorie continue.

1.1.1. Variabili aleatorie discrete

Si consideri, per esempio, l'estrazione a caso di una pallina da un'urna contenente palline bianche e nere. Si ripeta l'esperimento n volte e sia x il numero di palline, per esempio, bianche estratte in un ordine qualsiasi. Ripetendo più volte le n estrazioni il numero x varia in maniera casuale ed essendo un numero intero prende il nome di *variabile aleatoria discreta.*

1.1.2. Variabili aleatorie continue

Si consideri, per esempio, la fabbricazione in serie di un determinato prodotto. Siano $x_1, x_2, \ldots, x_n$ i valori di una prefissata caratteristica del prodotto (dimensionale, fisica, chimica, meccanica, ecc.) misurati con uno strumento a scala continua su n unità estratte a caso da quelle prodotte.
Variando in maniera casuale e potendo assumere teoricamente qualsiasi valore (almeno tutti quelli contenuti in un certo intervallo), x prende il nome di *variabile aleatoria continua.*

1.2. Popolazione e campione

Un intero gruppo di oggetti prende il nome di *popolazione o universo.* La totalità dei mm di pioggia (h) caduti in un anno (min $h = 0$; max $h = \infty$) compresa nell'intervallo $(0 \div \infty)$; il numero dei bulloni difettosi o non difettosi prodotti in una fabbrica in un dato giorno sono esempi di popolazione o universo.
Una popolazione può presentare numerosità *finita o infinita.*

Un esempio di popolazione finita è fornito da un lotto costituito dalle unità prodotte in un determinato periodo di tempo (per esempio tutti i bulloni prodotti in una fabbrica in un dato giorno) o dalle unità da sottoporre all'accettazione di un committente.
Un esempio di popolazione infinita è fornito dalla successione di tutte le possibili uscite (testa, croce) che si possono ottenere in successivi lanci di una moneta.
Molto spesso è impossibile, antieconomico o poco pratico esaminare tutte le unità di una popolazione specie se questa ha un'elevata numerosità.
Si esamina, allora, soltanto una piccola parte di essa che prende il nome di *campione* (1).
Dall'analisi di un campione, rappresentativo di una popolazione, si possono trarre importanti conclusioni circa la popolazione stessa. Tali conclusioni, non essendo certe, sono accompagnate da valutazioni di probabilità.

1.3. Dati, classi, frequenza, frequenza relativa e densità di frequenza (variabili aleatorie continue)

I dati raccolti ma non numericamente ordinati si dicono dati *grezzi*.
Volendo riassumere grandi quantità di dati grezzi, è spesso utile distribuirli in *classi* e determinare il numero di individui appartenenti a ciascuna classe, detto *frequenza della classe*.
Così siano $x_1, x_2, \ldots, x_n$ gli n valori assunti dalla variabile aleatoria x nelle n unità di un campione. Riportati gli n valori su di un asse x (fig. 1.1.):

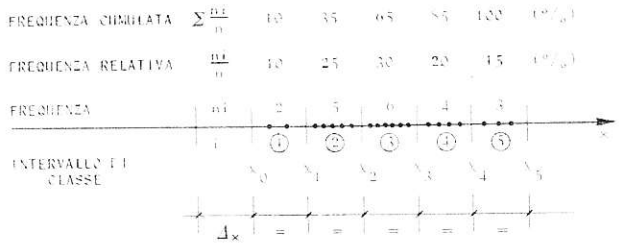


Fig. 1.1.

si suddivida l'intervallo $(x_{\max} - x_{\min})$ in un numero intero m $(5 \div 20)$ d'intervalli di ampiezza

$$\Delta x = \frac{x_{\max} - x_{\min}}{m}.$$

Gli intervalli così ottenuti prendono il nome di *intervalli di classe* mentre la loro ampiezza Δx prende il nome di *ampiezza degli intervalli di classe*. All'interno dell'i-esimo intervallo di classe cadano n_i valori di x. Il numero n_i, come si è detto, prende il nome di *frequenza*. Si definisce *frequenza relativa* f_i' dell'i^{mo} intervallo di classe il rapporto tra la frequenza della classe i^{ma} ed il totale della frequenze di tutte le classi, cioè il rapporto n_i/n $(i = 1, 2, \ldots, m)$, generalmente espresso in %:

$$f_i' = \frac{n_i}{n} \cdot 100$$

(1) In generale quando l'ampiezza dei campioni è > 30 si parla di *grandi campioni*, di contro, per ampiezza dei campioni < 30 si parla di *piccoli campioni*.

Si definisce *densità di frequenza* f_i dell'i-esimo intervallo di classe il rapporto tra la frequenza relativa e l'ampiezza dello stesso intervallo:

$$f_i = \frac{f_i'}{\Delta x} \cong \frac{n_i}{n\,\Delta x}.$$

Nel caso della figura 1.1. è $n = 2 + 5 + 6 + 4 + 3 = 20$.
Si hanno allora i seguenti valori delle frequenze relative per:

$$n_i = 2 \rightarrow f_1' = \frac{2 \times 100}{20} = 10\%$$

$$n_i = 5 \rightarrow f_2' = \frac{5 \times 100}{20} = 25\%$$

$$n_i = 6 \rightarrow f_3' = \frac{6 \times 100}{20} = 30\%$$

$$n_i = 4 \rightarrow f_4' = \frac{4 \times 100}{20} = 20\%$$

$$n_i = 3 \rightarrow f_5' = \frac{3 \times 100}{20} = 15\%$$

(ovviamente $\Sigma f_i' = 100\%$)
ed i seguenti valori delle *densità di frequenze* (essendo $\Delta x = 20/5 = 4$):

$$\text{per } n_i = 2 \rightarrow f_1 = \frac{0{,}10}{4} = 0{,}025 = 2{,}5\%$$

$$n_i = 5 \rightarrow f_2 = \frac{0{,}25}{4} = 0{,}0625 = 6{,}25\%$$

$$n_i = 6 \rightarrow f_3 = \frac{0{,}30}{4} = 0{,}075 = 7{,}5\%$$

$$n_i = 4 \rightarrow f_4 = \frac{0{,}20}{4} = 0{,}05 = 5\%$$

$$n_i = 3 \rightarrow f_5 = \frac{0{,}15}{4} = 0{,}0375 = 3{,}75\%$$

1.4. Istogramma e **poligono della densità di frequenza (variabili aleatorie continue)**

Sono rappresentazioni grafiche di distribuzioni di frequenza. Riportando in ordinate, in corrispondenza del punto medio di ogni intervallo di classe ([2]), i valori della *densità di frequenza* f_i e tracciando i rettangoli che hanno come base Δx e come altezza f_i, si ottiene un diagramma che prende il nome di *istogramma della densità di frequenza* (fig. 1.2.). Congiungendo, invece, con dei segmenti i punti che hanno come ascissa i valori medi degli intervalli di classe e come ordinate le corrispondenti f_i si ottiene una spezzata che prende il nome di *poligono della densità di frequenza* (fig. 1.3.).
Qualora si estenda il poligono della densità di frequenza in modo da interessare i due in-

([2]) Detto anche *valore centrale di una classe* (somma dei limiti superiore ed inferiore di una classe diviso per due).

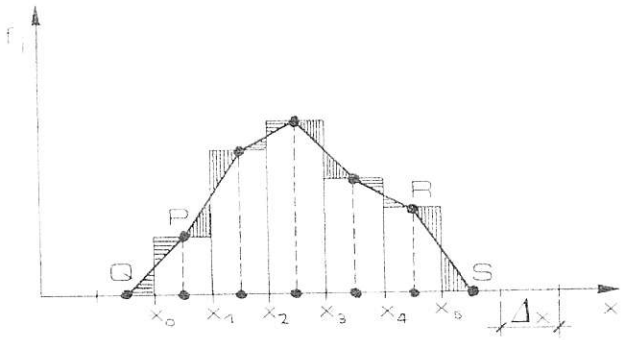


Fig. 1.2.

tervalli terminali di frequenza zero, le aree rispettivamente racchiuse tra l'asse x e l'istogramma e tra l'asse x e il poligono risultano uguali (fig. 1.3.).

In particolare tali aree valgono 1 (100%); infatti per i rettangoli dell'istogramma si ha:

$$\sum_{i=1}^{m} f_i \Delta x = \sum_{i=1}^{m} \frac{n_i}{n} \frac{\Delta x}{\Delta x} = \frac{\sum_{i=1}^{m} n_i}{n} = \frac{n}{n} = 1 = \sum_{i=1}^{m} f_i'$$

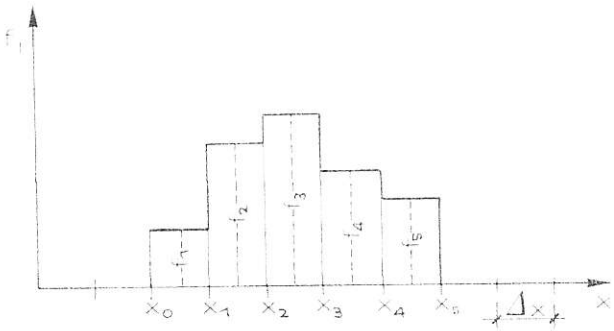


Fig. 1.3.

1.5. Frequenza cumulata e poligoni di distribuzione (variabili aleatorie continue)

La somma di tutti i valori di frequenza relativa degli intervalli di classe per i quali la variabile aleatoria x è minore o uguale al limite superiore x_i ($i = 1, 2, ..., m$) di un dato intervallo di classe prende il nome di *frequenza cumulata* fino a quell'intervallo di classe compreso. La frequenza cumulata si indica generalmente con F e si esprime in %.

In altre parole la frequenza cumulata rappresenta la frequenza relativa dell'intervallo di classe somma degli intervalli di classe fino a quello considerato compreso, ovvero rappresenta la frequenza relativa dell'intervallo $x \leq x_i$.

Così con riferimento alla figura 1.1. la frequenza cumulata fino al 3° intervallo di classe compreso, vale:

$$F(x_3) = 0{,}10 + 0{,}25 + 0{,}30 = 65\%$$

La frequenza cumulata si può calcolare anche a partire dall'istogramma (o dal poligono) della densità di frequenza sommando le aree dei rettangoli fino all'intervallo di classe considerato compreso.
Con riferimento alla fig. 1.2. si ha infatti:

$$x_3 = \sum_{i=1}^{3} f_i \Delta x = \sum_{i=1}^{3} \frac{n_i}{n} \frac{\Delta x}{\Delta x} = \sum_{i=1}^{3} f_i'. \qquad (1)$$

Riportando in ordinate, in corrispondenza del limite superiore x_i di ogni intervallo di classe, i valori della frequenza cumulata F_i si ottiene una spezzata che prende il nome di *poligono di distribuzione della frequenza cumulata o di poligono di distribuzione* (fig. 1.4.).

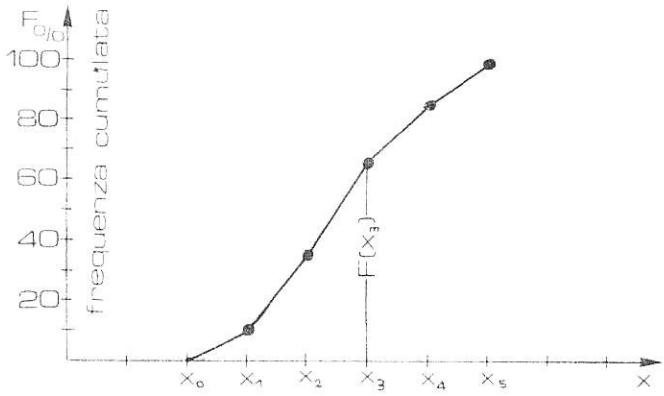


Fig. 1.4.

1.6. Funzione densità di probabilità e funzione di distribuzione (variabili aleatorie continue)

Per n molto grande (teoricamente infinito) è possibile scegliere intervalli di classe di ampiezza Δx molto piccola (teoricamente infinitesima), tale, però che in ciascuno di essi cada ancora un numero sufficiente di valori della variabile aleatoria continua x.
Il poligono della densità di frequenza, formato in questo caso da un numero molto grande di piccoli segmenti, si può approssimare con una curva continua detta *curva della densità di frequenza* (fig. 1.5.).

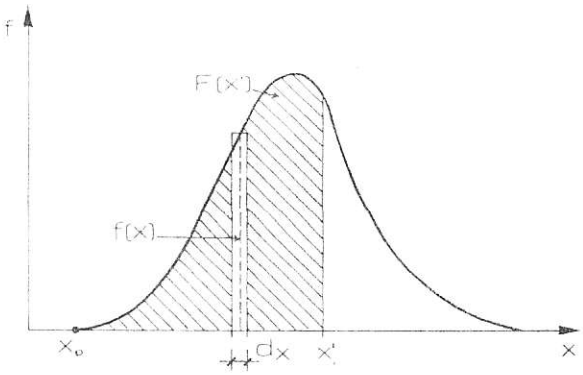


Fig. 1.5a.

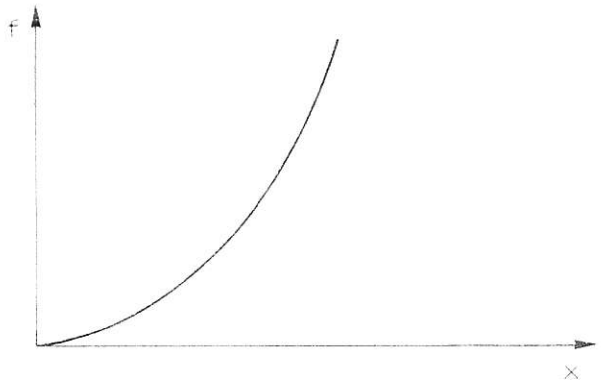


Fig. 1.5b.

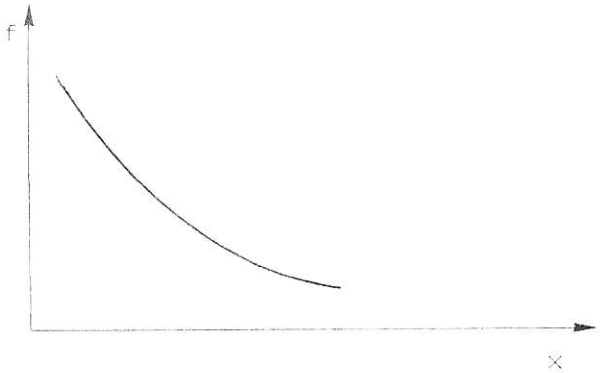


Fig. 1.5c.

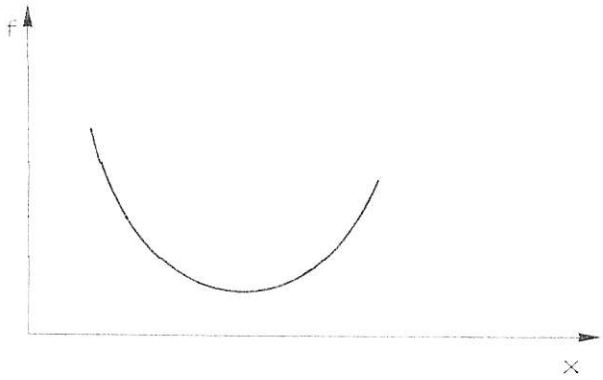


Fig. 1.5d.

Le curve di densità di frequenza che si possono ottenere in pratica si avvicinano alle forme caratteristiche riportate in figura 1.5.
La curva 1.5*a*.), simmetrica, caratterizza la circostanza che le osservazioni equidistanti dal massimo centrale hanno la stessa frequenza.
Le curve 1.5*b*.) e 1.5*c*.), ascendente e discendente rispettivamente, sono caratterizzate dalla frequenza del valore massimo ad una delle estremità.

La curva 1.5*d.*), ad U, ha un massimo ad entrambe le estremità. La curva della figura 1.5*e.*), bimodale, ha due massimi. La curva di figura 1.5*f.*), plurimodale, ha più di due massimi. Le curve di figg. 1.5*g.*) e 1.5*h.*) sono dette obliqua sinistra e obliqua destra se hanno rispettivamente l'estremità più lunga a destra o a sinistra.
Analogamente il poligono di distribuzione, anch'esso formato da un numero molto grande di piccoli segmenti, si può approssimare con una curva continua detta *curva di distribuzione* (fig. 1.6.).

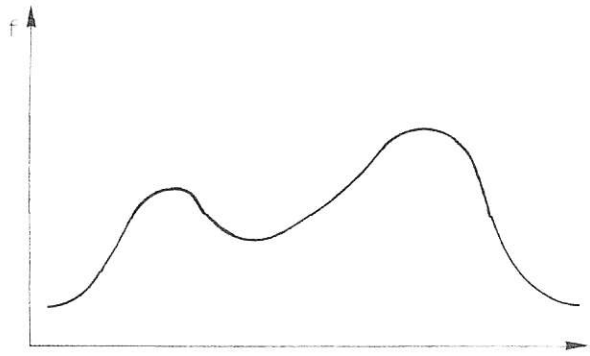


Fig. 1.5e.

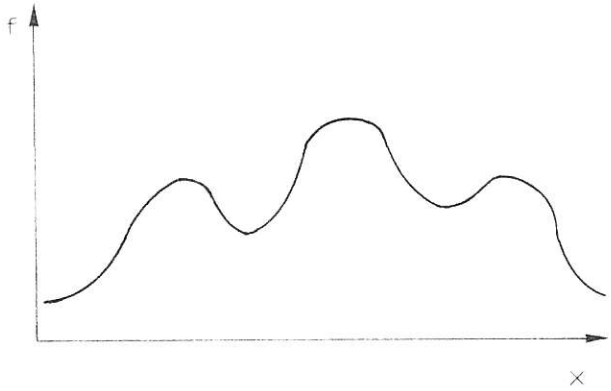


Fig. 1.5f.

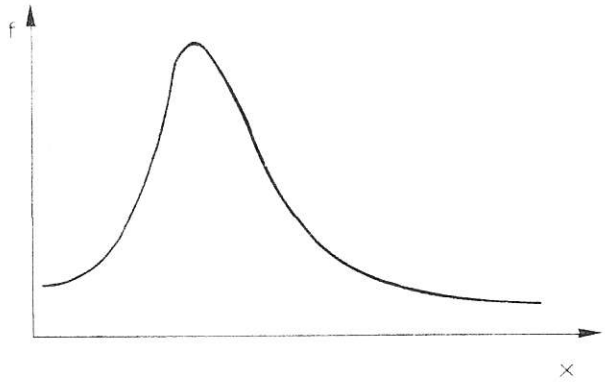


Fig. 1.5g.

Robert von Haláz
La Prefabbric
nell'edilizia ir
Pagg. 269. Figg. 3
L. 14.000 (IVA escl
Antonio Castagne
I calcestruzz
Caratteristiche -
zione
Pagg. 312. Figg. 3
L. 20.500 (IVA con
Edmondo Andre
Giovanni Mendit
Il calcolo de
in calcestru
agli stati lin
Pagg. 89. Figg. 1
L. 5.500 (IVA con
Stanislaw Pere
Con la collabor
Giulio Luigi Fa
Jacek Sokalski
Guida alla
Strutture t
prefabbric
in calcestr
Pagg. 432. Figg
L. 28.000 (IVA c

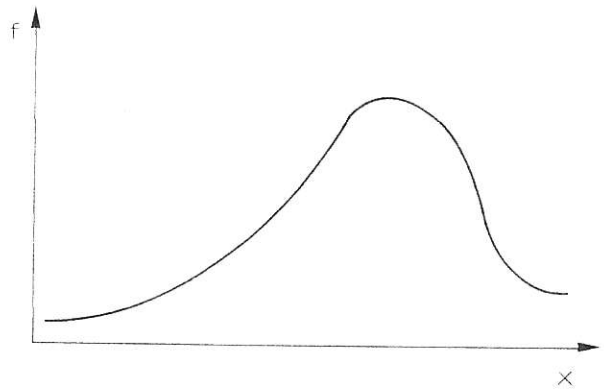


Fig. 1.5h.

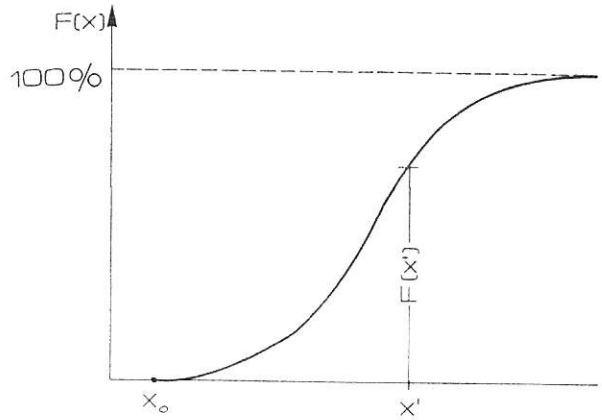


Fig. 1.6.

La F è funzione di x e si definisce anche come la *funzione di distribuzione* della va aleatoria x.
Sostituendo nella (1) ai valori f_i la funzione $f(x)$, a Δx il dx ed al segno di s toria il segno di integrale si ottiene:

$$F(x') = \int_{x_0}^{x} f(x)\, dx$$

che rappresenta l'area compresa tra la curva $f(x)$, l'asse x e la retta $x = x'$ (teggio in fig. 1.5.).
La $F(x)$ è, pertanto, la funzione integrale della $f(x)$; dalla (2) discende che:

$$f(x) = \frac{dF(x)}{dx}.$$

1.7. Definizione di probabilità basata sulla frequenza relativa

La probabilità, oltre che mediante la frequenza relativa (von Mises), può defi — in maniera assiomatica (teoria della misura);

— a priori come rapporto tra le alternative favorevoli ed il numero totale delle alternative (teoria classica);

— come misura di un convincimento (ragionamenti induttivi).

Si assume come stima della probabilità di un *evento a* (probabilità empirica) la frequenza relativa del presentarsi dell'evento $f' = n_a/n$, dove n_a è il numero di volte che si è verificato l'evento a ed n è il numero totale delle osservazioni.
Si definisce *probabilità* P di un evento a il limite della frequenza relativa quando il numero delle osservazioni cresce indefinitamente (n tendente all'infinito) [3]:

$$P(a) = \lim_{n\to\infty} f' = \lim_{n\to\infty} \frac{n_a}{n}. \tag{3}$$

Dato che la funzione di distribuzione $F(x')$ rappresenta il limite, per n tendente all'infinito, della frequenza relativa dell'intervallo $x \leq x'$, dalla (3) si ottiene che $F(x')$ rappresenta la probabilità dell'evento $\{x \leq x'\}$:

$$P\{x \leq x'\} = F(x'). \tag{4}$$

Pertanto anche l'area a tratteggio in figura 1.5. rappresenta la probabilità dell'evento $\{x \leq x'\}$ e la funzione f si definisce anche come la *funzione densità di probabilità* della variabile aleatoria x.
Nel caso di variabile aleatoria discreta si ha, con ovvio significato dei simboli:

$$f'(x_i) = P\{x = x_i\} \tag{5}$$

$$F(x_i) = \sum_{i=1}^{n} f'(x_i) = \sum_{i=1}^{n} P\{x = x_i\} = P\{x \leq x_i\}. \tag{6}$$

1.8. Valore atteso e dispersione

La funzione di distribuzione $F(x)$ offre una descrizione completa di una variabile aleatoria x. In molti problemi pratici, però, interessano solo alcuni parametri della funzione

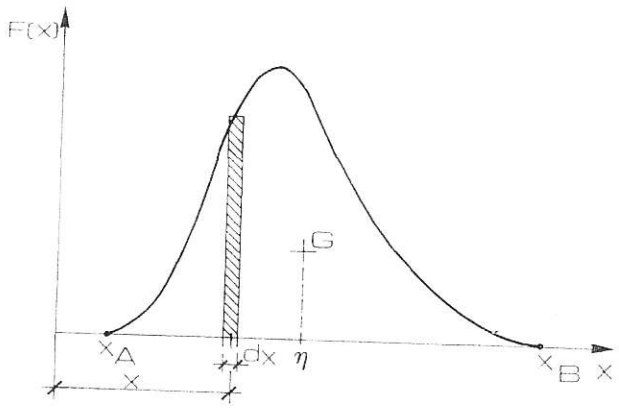


Fig. 1.7.

[3] Questa definizione statistica di probabilità crea delle difficoltà dal punto di vista matematico poiché in realtà può non esistere un numero da assumere come limite.

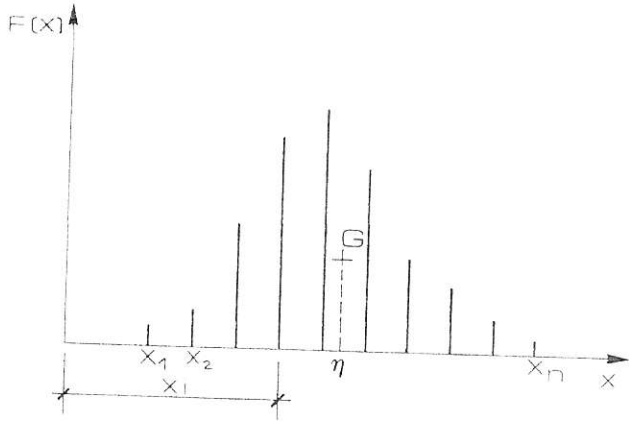


Fig. 1.8.

di distribuzione. Tra i più importanti si ricordano:

— il valore atteso,

— la dispersione.

Il *valore atteso* o *valor medio* o *media* di una variabile aleatoria x si indica con uno dei seguenti simboli: $E\{x\}$ o con η_x o con η e si definisce come segue:

— se x è una variabile aleatoria continua ($x_A \leq x \leq x_B$; fig. 1.7.)

$$\eta = \int_{x_A}^{x_B} x f(x)\, dx \tag{7}$$

— se x è una variabile aleatoria discreta ($x_1 \leq x_i \leq x_n$; fig. 1.8.)

$$\eta = \sum_{i=1}^{n} x_i f'(x_i). \tag{8}$$

Il valore atteso si può interpretare come l'ascissa del baricentro G di un sistema di masse disposto lungo l'asse x e avente massa totale uguale a uno ([4]).

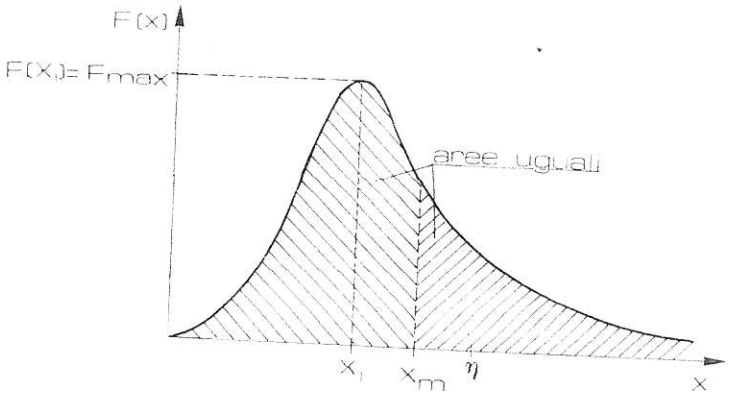


Fig. 1.9.

([4]) Per definizione l'ascissa x_G del baricentro G di una distribuzione di masse m è uguale al rapporto tra il momento statico della distribuzione di masse rispetto all'asse y e la massa totale.

Nel caso di variabile aleatoria continua la funzione densità di massa (massa per unità di lunghezza) è rappresentata dalla funzione densità di probabilità $f(x)$. Nel caso di variabile aleatoria discreta la distribuzione delle masse è rappresentata dalla funzione $f'(x_i)$. Altre informazioni sulla cosiddetta *tendenza centrale* sono fornite dai parametri chiamati rispettivamente *moda e mediana* (fig. 1.9.).
La *moda* x_i è l'ascissa per la quale si ha il massimo della funzione densità di probabilità.
La *mediana* x_m è l'ascissa per la quale si ha:

$$P\{x \leq x_m\} = F(x_m) = 0{,}5.$$

Media, moda e *mediana* coincidono nel caso di distribuzione simmetrica.
La *dispersione* o *varianza* di una variabile aleatoria x_i viene normalmente indicata con σ^2 e si definisce come segue:

— se x è una variabile aleatoria continua ($x_A \leq x \leq x_B$; fig. 1.7.)

$$\sigma^2 = \int_{x_A}^{x_B} (x - \eta)^2 f(x)\, dx \tag{9}$$

— se x è una variabile aleatoria discreta ($x_1 \leq x_i \leq x_n$; fig. 1.8.)

$$\sigma^2 = \sum_{i=1}^{n} (x_i - \eta)^2 f'(x_i). \tag{10}$$

La *varianza* eguaglia il momento d'inerzia della distribuzione di masse di cui si è detto rispetto a un asse baricentrale e dà una misura della concentrazione o della dispersione della variabile x intorno al valor medio η.
Se $\sigma^2 = 0$ la variabile x è deterministica e coincide con η.
La radice quadrata σ della varianza prende il nome di *deviazione standard o di scarto quadratico medio.* Lo scarto quadratico medio permette di avere un'idea circa la dispersione di grandezze rispetto alla media.
Un'altra misura della dispersione di una variabile aleatoria x è fornita dalla *escursione*:

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

Il rapporto σ/η prende il nome di *coefficiente di variazione o di coefficiente di dispersione.*

1.9. Variabili aleatorie standardizzate

La variabile aleatoria adimensionale:

$$z = \frac{x - \eta}{\sigma} \tag{11}$$

prende il nome di variabile *aleatoria standardizzata* o *variabile ridotta.* La variabile z misura lo scostamento della variabile x dal suo valore medio η in unità della deviazione standard σ.

1.10. Distribuzioni di una variabile aleatoria

La funzione distribuzione può essere descritta raccogliendo in un istogramma i dati ottenuti sperimentalmente. Questo procedimento presenta tuttavia due inconvenienti:

a) i dati sperimentali forniscono buone informazioni relativamente alla distribuzione del campione. Minor precisione si ha allorquando si tende di estrapolare questa informazione alla distribuzione della popolazione;

b) la funzione distribuzione, ottenuta sperimentalmente, è normalmente difficile da utilizzare nelle successive manipolazioni.

Gli inconvenienti sottolineati vengono superati utilizzando distribuzioni matematicamente più note ed i cui valori sono di norma, già tabulati. Pertanto anziché determinare una funzione di distribuzione a partire dai dati sperimentali, si sceglie tra le già note quella che questi dati meglio rappresenta.
Tra le varie «distribuzioni» di una variabile aleatoria x (distribuzione binomiale; di Poisson; ipergeometrica; campionarie; di quadrato; t di Student, ecc.) ci si limita qui a ricordare le seguenti:

1.10.1. Distribuzione normale o gaussiana

È così definita la distribuzione di una variabile aleatoria x continua, per la quale la funzione densità di probabilità è, nella sua forma più generale, la seguente ([5]):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-(x-\eta)^2/2\sigma^2] \tag{12}$$

con $-\infty \leq x \leq +\infty$ e con $\sigma > 0$. I parametri caratteristici della distribuzione normale assumono i valori:

— valor medio:

$$\eta = \int_{-\infty}^{+\infty} x \cdot f(x) \cdot dx;$$

— varianza:

$$\sigma = \sqrt{\int_{-\infty}^{+\infty} (x-\eta)^2 \cdot f(x) \cdot dx}$$

(σ misura la distanza dei punti di flesso della curva dal valor medio η).
Lo studio analitico della (12) mostra che:

per $x \to \infty$, $f(x) \to 0$

$x \to -\infty$, $f(x) \to 0$

$$\frac{df(x)}{dx} = -\frac{1}{\sigma\sqrt{2\pi}} \frac{2(x-\eta)}{2\sigma^2} \exp[-(x-\eta)^2/2\sigma^2] =$$

$$= -\frac{1}{\sigma^2\sqrt{2\pi}} (x-\eta) \exp[-(x-\eta)^2/2\sigma^2] \quad \left(\frac{df(x)}{dx}\right)_{x=\eta} = 0.$$

La funzione $f(x)$ è dunque massima per $x = \eta$.
La funzione $f(x)$ è simmetrica rispetto all'asse passante per $x = \eta$ e parallelo all'asse $f(x)$:

$$\max f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-(\eta-\eta)^2/2\sigma^2] = \frac{1}{\sigma\sqrt{2\pi}}.$$

([5]) La (12) fu data per la prima volta da A. de Moivre nel seicento ma solo nell'ottocento venne sviluppata da C. F. Gauss e Laplace.

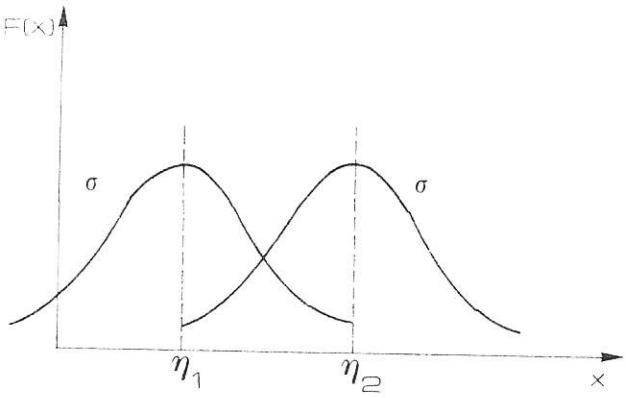


Fig. 1.10.

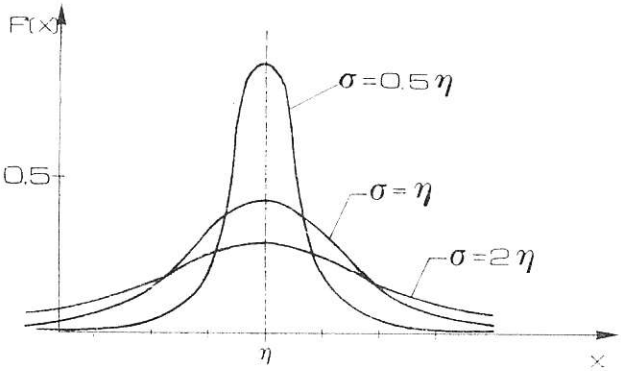


Fig. 1.11.

Effettuando la $d^2f(x)/dx^2$ si individuano 2 punti di flesso simmetricamente posti rispetto a η (fig. 1.10.).
In figura 1.10. sono riportate due distribuzioni aventi la stessa varianza (σ) ma differenti valor medi (η_1 e η_2).
In figura 1.11. sono riportate alcune distribuzioni normali con lo stesso valor medio η e differenti valori della deviazione standard σ.
Se $\eta = 0$ la $f(x)$ è detta «simmetrica» (i valori negativi o positivi sono egualmente probabili).
La variabile standardizzata

$$z = \frac{x - \eta}{\sigma}$$

è ancora una variabile normale che ha valor medio nullo e deviazione standard uguale a 1. In termini di variabile standardizzata la (12) assume la forma cosiddetta standardizzata (fig. 1.12.):

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2). \tag{13}$$

Nella distribuzione normale, media e mediana coincidono, essendo la distribuzione simmetrica. Noti il valor medio η e la deviazione standard σ la distribuzione normale è nota. L'area delimitata dalla curva (12) e dall'asse x vale uno. Quindi l'area sotto la curva

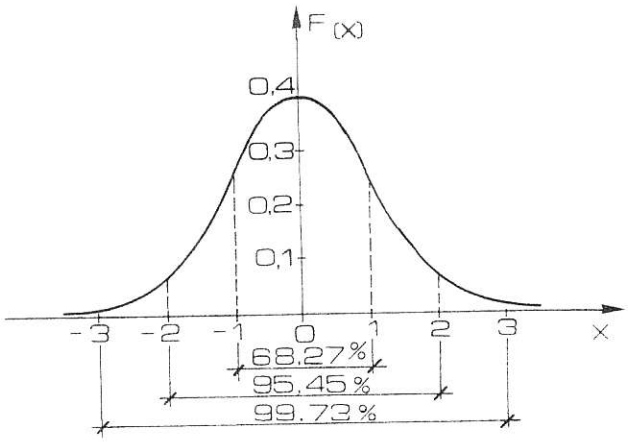


Fig. 1.12.

(12) compresa tra $x = a$ e $x = b$ $(a > b)$ rappresenta la probabilità che x sia compreso tra a e b.
La funzione di distribuzione $F(x)$ è data da:

$$F(x) = \int_{-\infty}^{x} f(x')\, dx' = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp[-(x' - \eta)^2/2\sigma^2]\, dx'. \tag{14}$$

Ne segue che la probabilità che la x sia contenuta in un certo intervallo (a, b) è data dalla relazione:

$$P\{a < x < b\} = F(b) - F(a) = \frac{1}{\sigma\sqrt{2\pi}} \int_{a}^{b} \exp[-(x - \eta)^2/2\sigma]\, dx. \tag{15}$$

Tale valore è dato dall'area tratteggiata nella figura 1.13.
In particolare risulta:

$$\begin{aligned} P\{\eta - \sigma < x < \eta + \sigma\} &= F(x = \eta + \sigma) - F(x = \eta - \sigma) = \\ &= F\left(z = \frac{\eta + \sigma - \eta}{\sigma}\right) - F\left(z = \frac{\eta - \sigma - \eta}{\sigma}\right) = \\ &= F(z = 1) - F(z = -1) = 2F(z = 1) = 2 \times 0{,}3413 = 0{,}6826 = 68{,}26\% \end{aligned}$$

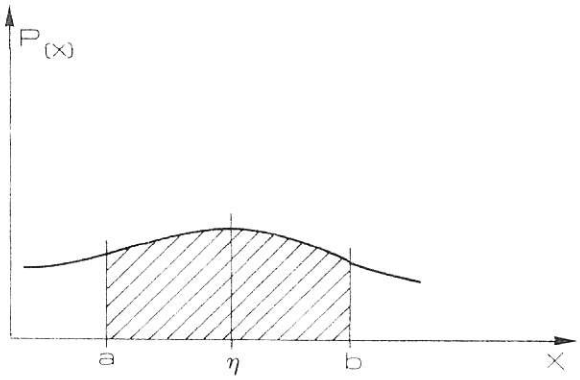


Fig. 1.13.

$$P\{\eta - 2\sigma < x < \eta + 2\sigma\} = F(x = \eta + 2\sigma) - F(x = \eta - 2\sigma) =$$
$$= F\left(\frac{\eta + 2\sigma - \eta}{\sigma}\right) - F\left(\frac{\eta - 2\sigma - \eta}{\sigma}\right) =$$
$$= F(2) - F(-2) = 2F(2) = 2 \times 0{,}4772 = 0{,}9544 = 95{,}44\%$$
$$P\{\eta - 3\sigma < x < \eta + 3\sigma\} = F(x = \eta + 3\sigma) - F(x = \eta - 3\sigma) =$$
$$= F\left(\frac{\eta + 3\sigma - \eta}{\sigma}\right) - F\left(\frac{\eta - 3\sigma - \eta}{\sigma}\right) =$$
$$= F(3) - F(-3) = 2F(3) = 2 \times 0{,}4987 = 0{,}9974 = 99{,}74\%$$

e quindi per la deviazione standard:

$$1 - P = 1 - 0{,}6826 = 0{,}3174$$
$$1 - P = 1 - 0{,}9544 = 0{,}0456$$
$$1 - P = 1 - 0{,}9974 = 0{,}0026.$$

Da ciò si può vedere come, al crescere dell'ampiezza all'intervallo considerato, aumenta la probabilità che la variabile aleatoria x assuma un valore dell'intervallo stesso. Al limite si ha:

$$P\{-\infty < x < \infty\} = \frac{1}{2\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp[-(x - \eta)^2/2\sigma^2]\, dx = 1.$$

La funzione integranda che compare nella (14) non è una funzione elementare ed il suo integrale non è ottenibile in forma semplice. È perciò necessario ricorrere ad una tabellazione che, per la particolare forma della (14), richiederebbe tante tabelle quante sono le coppie di valori (η, σ).
Tuttavia si osserva che:

$$P\{a < x < b\} = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} \exp[-(z^2\sigma^2/2\sigma)](\sigma\, dz) = \int_a^b \frac{1}{\sqrt{2\pi}} \exp[-(z^2/2)]\, dz$$

ovvero, posto

$$Y = \frac{1}{\sqrt{2\pi}} \exp[-(z^2/2)],$$

si ha:

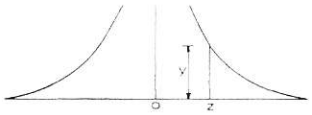
$$P\{a < x < b\} = \int_a^b Y(z)\, dz.$$

Nella tab. 1 sono calcolari i valori di Y e nella tab. 2 è calcolato l'integrale

$$\frac{1}{\sqrt{2\pi}} \int_0^z \exp[-(z^2/2)]\, dz$$

(cosiddetto integrale o funzione di Gauss) che rappresenta le aree sotto la curva (12) compresa tra $z = 0$ e z qualsiasi fino a 3,99. Le aree comprese tra una coppia qualsiasi di valori di z (per $a < z < b$) si possono ottenere tenendo presente che la (12) è simmetrica rispetto all'asse $f(z)$.
La distribuzione normale gioca uno speciale ruolo nella teoria della probabilità e della statistica matematica poiché secondo i teoremi limiti della teoria della probabilità, le distribuzioni di quelle variabili aleatorie che sono formate dalla sovrapposizione di un gran

TABELLA 1. **Valori di $Y = 1/\sqrt{2\pi}\ \exp(z^2/2)$**

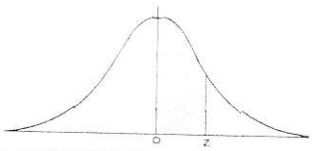


z	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3857	0,3847	0,3836	0,3825
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3725	0,3712	0,3697
0,4	0,3663	0,3668	0,3653	0,3637	0,3621	0,3605	0,3589	0,3572	0,3555	0,3538
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3352
0,6	0,3332	0,3312	0,3292	0,3271	0,3521	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,2468	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0,2131	0,2107	0,2083	0,2059	0,2036	0,2012	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1804	0,1781	0,1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1539	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0,1354	0,1334	0,1315
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,0957
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0804
1,8	0,0790	0,0775	0,0761	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0669
1,9	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0,0584	0,0573	0,0562	0,0551
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0,0440	0,0431	0,0422	0,0413	0,0404	0,0396	0,0387	0,0371	0,0371	0,0363
2,2	0,0355	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0297	0,0290
2,3	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0229
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0180
2,5	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0,0143	0,0139
2,6	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110	0,0107
2,7	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0,0084	0,0081
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0061
2,9	0,0060	0,0058	0,0056	0,0055	0,0053	0,0051	0,0050	0,0048	0,0047	0,0046
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034
3,1	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026	0,0025	0,0025
3,2	0,0024	0,0023	0,0022	0,0021	0,0021	0,0020	0,0020	0,0019	0,0018	0,0018
3,3	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014	0,0013	0,0013
3,4	0,0012	0,0012	0,0012	0,0011	0,0011	0,0010	0,0010	0,0010	0,0009	0,0009
3,5	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007	0,0007	0,0007	0,0006
3,6	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0004
3,7	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003	0,0003	0,0003	0,0003
3,8	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002
3,9	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001	0,0001

numero di fattori indipendenti (o quasi indipendenti), l'influenza di ciascuno dei quali non è significativa, potrà essere approssimata dalla distribuzione normale. Non è quindi da meravigliarsi se la legge di distribuzione normale si trova impiegata per descrivere effetti diversi per esempio la distribuzione di errori di misura aleatoria, la distribuzione di velocità di un vento burrascoso, la distribuzione dei rumori nelle linee di comunicazione, ecc.

Anche in quei casi in cui la distribuzione chiaramente non è normale (per esempio per le caratteristiche meccaniche di materiali che assumono solo valori positivi), viene spesso usata la distribuzione normale per sostituire in modo approssimato l'effettiva legge di distribuzione.

TABELLA 2. **Aree sotto la curva normale standardizzata da 0 a *z***



z	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1064	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2258	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3689	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4949	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4414	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4949	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4983	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

E ciò essenzialmente per i seguenti motivi:

a) la distribuzione normale può essere completamente caratterizzata da due parametri numerici: il valor medio e la varianza;

b) se la variabile aleatoria è distribuita normalmente, la distribuzione rimane normale anche dopo una trasformazione lineare della variabile aleatoria (incluse le operazioni di differenziazione ed integrazione). Le due circostanze ricordate, consentono di trovare leggi di distribuzione di funzioni lineari di variabili aleatorie molto semplicemente se le variabili originali sono distribuite normalmente.