2.6 Teorema limite centrale

Con questo termine vengono indicati un gruppo di teoremi che risultano indispensabili per la teoria delle distribuzioni necessaria allo sviluppo della statistica inferente.

Questi teoremi costituiscono, in pratica, un modo per «quantificare» la legge dei grandi numeri (vol. I, par. 9.8).

Ricordiamo brevemente le due formulazioni fondamentali di tale legge:

- per prove ripetute indipendenti, dove il risultato di ciascuna prova può essere classificato come successo o insuccesso, si può asserire che (*teorema di Bernoulli*):

  al crescere del numero delle prove la frequenza relativa dei successi converge in probabilità alla probabilità di successo in una prova.

2) per prove ripetute indipendenti in cui il risultato di ciascuna prova è il valore x di una variabile aleatoria X (ad esempio una misura di lunghezza, peso, durata), si può asserire che (*teorema di Cebicev*):

   per un numero sufficientemente grande di prove indipendenti la media aritmetica dei valori osservati di una variabile aleatoria converge in probabilità alla sua speranza matematica.

Tutte le formulazioni della legge dei grandi numeri stabiliscono che i risultati delle singole prove influiscono poco sul risultato medio di un numero elevato di prove: le deviazioni dalla media, inevitabili in una prova singola, si livellano reciprocamente quando il numero di prove è elevato.

Questo significa, in altri termini, che quando il numero di prove è elevato, il risultato medio diventa stabile e quindi può essere previsto.

Le possibilità di effettuare tali previsioni sono rese maggiori dal *teorema limite centrale* che stabilisce quale distribuzione segue la somma di un numero sufficientemente grande di variabili aleatorie.

Tale teorema, detto «centrale» proprio per la sua importanza, permette di definire delle ipotesi e di stimare la loro probabilità di verificarsi.

La formulazione più generale del *teorema limite centrale* è la seguente:

sia S_n *una variabile aleatoria somma di* n *variabili aleatorie indipendenti* X_i *aventi ciascuna la stessa distribuzione di probabilità, speranza matematica* μ *e varianza* σ^2*; al crescere di* n, S_n *tende ad assumere una distribuzione normale con media* $n\mu$ *e varianza* $n\sigma^2$:

$$S_n = X_1 + X_2 + \dots + X_n \tag{2.24}$$

La variabile:

$$Z = \frac{S_n - n\mu}{\sqrt{n}\sigma} \tag{2.25}$$

è la corrispondente variabile normale standardizzata.

Il teorema limite centrale risulta valido, per n sufficientemente grande, qualunque sia la distribuzione delle variabili X_i; inoltre è generalizzabile al caso di variabili aleatorie *con distribuzione di probabilità qualsiasi*, alla sola condizione che ciascuna di queste variabili abbia media e varianza finite e non risulti predominante rispetto alle altre. In questo caso la media di S_n è la somma dei valori medi delle singole variabili e la varianza è la somma delle varianze.

Se le variabili X hanno distribuzione normale con media μ e varianza σ^2 allora la variabile S_n ha sempre distribuzione normale, qualunque sia il valore n, con media $n\mu$ e varianza $n\sigma^2$.

■ ESEMPIO 2.15 ■

Una scatola contiene 50 pastiglie. Il peso di ciascuna pastiglia è una variabile con valor medio 1.2 gr e varianza 0.09 gr.

Determinare la probabilità che la scatola abbia un peso netto:

a) superiore a 62 gr;
b) inferiore a 59 gr;
c) compreso tra 56 e 54 gr.

Risulta:

$$n\mu = 50 \cdot 1.2 = 60 \qquad n\sigma^2 = 50 \cdot 0.09 = 4.5 \qquad \sqrt{n}\sigma = \sqrt{4.5} = 2.12$$

a) In questo caso dalla [2.25]:

$$z = \frac{62 - 60}{2.12} = 0.94$$

$$P(S_n \geqslant 62) = P(Z \geqslant 0.94) = 1 - F(Z = 0.94) = 0.1736$$

b) In questo caso:

$$z = \frac{59 - 60}{2.12} = -0.47$$

$$P(S_n \leqslant 59) = P(Z \leqslant -0.47) = 1 - F(Z = 0.47) = 0.3192$$

c) In questo caso:

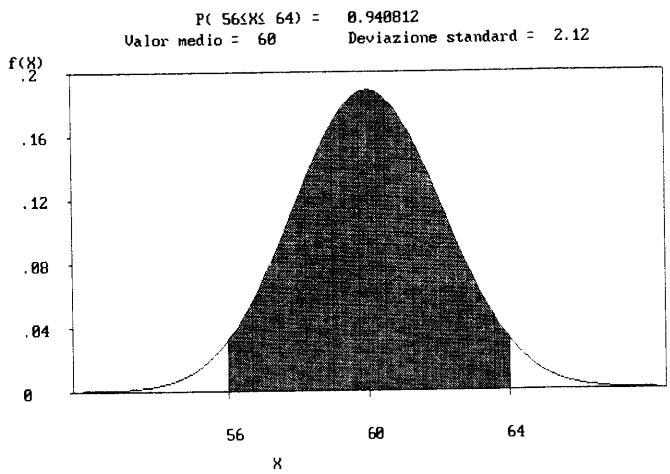
$$z_1 = \frac{56 - 60}{2.12} = -1.89 \qquad z_2 = \frac{64 - 60}{2.12} = 1.89$$

$$P(56 \leqslant S_n \leqslant 64) = P(-1.89 \leqslant Z \leqslant 1.89) = 2F(Z = 1.89) - 1 = 0.94$$

Quest'ultimo caso è rappresentato in fig. 2.29. Il valore della probabilità è stato ottenuto con il programma applicativo riportato nella scheda 4.

■

Fig. 2.29 - Distribuzione normale



Il teorema limite centrale permette di asserire che la distribuzione binomiale tende alla distribuzione normale per valori di n sufficientemente elevati.

Si consideri una prova binomiale, tale che la probabilità di successo sia p e la probabilità di insuccesso sia $q = 1 - p$.

La variabile indicatrice Y assume i valori:

— $Y = 1$ con probabilità p;
— $Y = 0$ con probabilità $q = 1 - p$.

Il valor medio della variabile indicatrice Y risulta:

$$\mu = E(Y) = 1 \cdot p + 0 \cdot q = p$$

La varianza risulta:

$$\sigma^2 = \mathrm{Var}(Y) = (1-p)^2 p + (0-p)^2 q = (1-p)\,p\,(p+q) = pq$$

Si consideri la variabile S_n somma delle variabili indicatrici:

$$S_n = Y_1 + Y_2 + \ldots + Y_n$$

In base al teorema limite centrale la variabile:

$$Z = \frac{S_n - np}{\sqrt{n}\sqrt{pq}} \qquad [2.26]$$

ha distribuzione normale con media np e deviazione standard $\sqrt{npq}$ (che sono. rispettivamente, la media e la deviazione standard della distribuzione binomiale). La somma delle variabili indicatrici S_n indica il numero di successi in n prove.

Nel seguito, per analogia con la notazione utilizzata per la distribuzione binomiale, la somma delle variabili indicatrici verrà generalmente indicata con X:

$$S_n = X$$

In conclusione il teorema limite centrale permette di affermare che, quando il numero delle prove è sufficientemente elevato, la distribuzione binomiale può essere approssimata con la distribuzione normale; in tal modo si ottiene una notevole semplificazione nel calcolo delle probabilità.

■ ESEMPIO 2.16 ■

[illegible] determinare la probabilità che si presenti un numero [illegible] e dalla [2.26] si ottiene:

$$z = \frac{260 - 250}{11.18} = 0.89$$

pertanto:

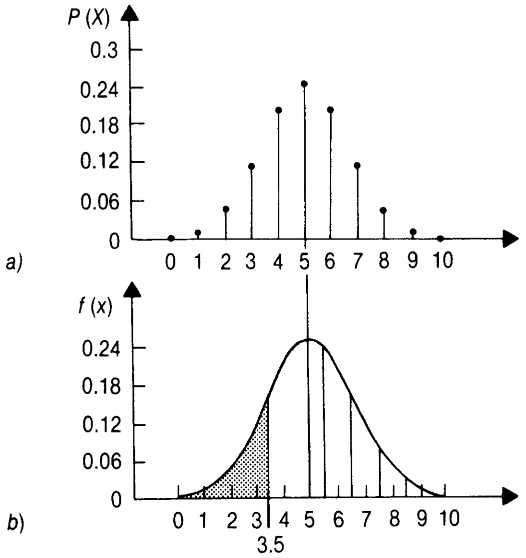
$$P(X \geqslant 260) = P(Z \geqslant 0.89) = 0.1867$$

■

In realtà l'approssimazione della distribuzione binomiale per mezzo di quella normale risulta buona anche quando n non è un valore molto elevato se si tiene conto che la distribuzione normale è una distribuzione continua e si usano le opportune modifiche ai valori, come nell'esempio seguente.

Si considerino 10 lanci di una moneta; si vuole determinare la probabilità che si presenti un numero di teste minore di 4.

Fig. 2.30 - Confronto fra distribuzione normale e binomiale



In fig. 2.30 *a*) è riportato l'andamento della distribuzione binomiale con $p = 0.5$ e $n = 10$; in fig. 2.30 *b*) è riportato l'andamento della distribuzione normale con media $np = 5$.

Dal confronto tra le due figure si può osservare che, mentre la funzione di probabilità della distribuzione binomiale ha solo 11 valori (per X che varia da 0 a 10), la funzione di densità di probabilità normale ammette infiniti valori (tutti i numeri reali compresi tra 0 e 10), pertanto per determinare le probabilità si devono considerare anche i valori intermedi.

Nel caso specifico, per determinare la probabilità che si presenti un numero di teste minore di 4 devono essere considerati tutti i valori minori di 3.5.

Con questa approssimazione, detta *approssimazione di continuità*, si può determinare:

$$P(X \leqslant 3.5) = F(X = 3.5) \quad \text{da } z = \frac{3.5 - 5}{\sqrt{10 \cdot 0.5 \cdot 0.5}} = -0.95$$

$$P(X \leqslant 3.5) = F(Z = -0.95) = 0.1712$$

Si confronti questo valore con quello che si ottiene applicando la distribuzione binomiale:

$$P(X \leqslant 3) = \sum_{x_i=0}^{3} \binom{10}{x_i} \left(\frac{1}{2}\right)^{x_i} \left(\frac{1}{2}\right)^{10-x_i} = 0.1719$$

Come si può vedere le due probabilità differiscono di pochissimo (anche se n non è molto elevato).

Quando il numero n delle prove è molto elevato il fatto di utilizzare o meno l'approssimazione di continuità diventa meno importante in quanto, come si è visto nel capitolo precedente, l'incremento della probabilità $P(X)$ tra due valori successivi di X diventa una quantità piccola.

Utilizzando i programmi applicativi si possono confrontare gli andamenti delle due distribuzioni per diversi valori di n.

Un'approssimazione dei valori puntuali della distribuzione binomiale per mezzo di quella normale è data dal *teorema limite locale di De Moivre-Laplace* (storicamente fu il primo teorema limite).

Tale teorema asserisce:

> *se* p *è la probabilità con cui si verifica un successo in una prova ed è costante per tutte le prove, allora la probabilità che si verifichino* X *successi in un numero* n *molto grande di prove tende a*:

$$\frac{1}{\sqrt{npq}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-np}{\sqrt{npq}}\right)^2} \qquad [2.27]$$

Questo teorema significa che, per un numero di prove sufficientemente grande, la funzione di probabilità binomiale può essere approssimata dalla funzione di densità normale:

$$P(X = x) = \binom{n}{x} p^x q^{n-x}$$

tende a

$$\frac{1}{\sqrt{2\pi}\sqrt{npq}} e^{-\frac{1}{2}\left(\frac{x-np}{\sqrt{npq}}\right)^2}$$

In appendice A, tavola II, sono tabulati i valori di $f(z)$:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad [2.28]$$

dove:

$$z = \frac{x - np}{\sqrt{npq}}.$$

Per determinare i valori di $f(x)$ si può osservare che:

$$f(x) = \frac{1}{\sqrt{npq}} f(z) \qquad [2.29]$$

■ ESEMPIO 2.17 ■

Determinare la probabilità che, estraendo con reintroduzione 40 carte da un mazzo di 52, si presentino 12 carte di fiori, utilizzando la funzione binomiale e la funzione di densità normale.

Risulta:

$$p = 0.25 \qquad q = 0.75 \qquad n = 40$$

$$np = 40 \cdot 0.25 = 10 \qquad \sqrt{npq} = 2.74$$

Con la funzione binomiale:

$$P(X = 12) = \binom{40}{12} (0.25)^{12} (0.75)^{28} = 0.1057$$

Con la funzione di densità normale:

$$z = \frac{12 - 10}{\sqrt{40 \cdot 0.25 \cdot 0.75}} = 0.73$$

$$f(X = 12) = \frac{1}{\sqrt{npq}} f(z) = \frac{1}{2.74} \cdot 0.3056 = 0.1115$$

$f(z)$ si può trovare nella tavola II dell'appendice A.

■

Nell'enunciato del teorema limite centrale è specificato che il teorema risulta valido in tutta la sua generalità, solamente quando il numero n delle variabili o delle prove è molto grande.

Nella pratica questa affermazione deve essere quantificata e si assume, di solito, che i teoremi limite siano validi se $n \geqslant 30$.