

La distribuzione normale come approssimazione della distribuzione binomiale

In precedenza la distribuzione di Poisson è stata introdotta come approssimazione della distribuzione binomiale quando n è grande e contemporaneamente p è piccolo.

Allora si disse che un'altra distribuzione fornisse una buona approssimazione per n grande a prescindere dal valore di p . *La distribuzione normale è la distribuzione che gode di questa proprietà.*

Prima di indagare sulla natura di questa approssimazione consideriamo un esempio numerico:

sia $n = 12$, $p = \frac{1}{3}$ e costruiamo il grafico della corrispondente distribuzione binomiale.

Questo valore di n non è certo grande, cosicché non ci si dovrebbe aspettare qui una buona approssimazione della distribuzione normale.

Poiché la $f(x)$ si deve calcolare per tutti i valori di x da 0 a 12, è più facile calcolare ciascun valore dopo il primo da quello precedente anziché calcolarlo da solo.

In questo caso la funzione di densità binomiale è la seguente:

$$f(x) = \frac{12!}{x!(12-x)!} \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{12-x}$$

Per questa funzione si vede facilmente che $f(x+1) = \frac{12-x}{x+1} \frac{1}{2} f(x)$, dopo aver calcolato $f(0)$

quest'ultima relazione è stata usata per ottenere i restanti valori riportati in tabella 1 (vedi fotocopia).

Il grafico di questa distribuzione binomiale è mostrato in figura 1:

Fig 1) vedi fotocopia

Sembra che questo istogramma si possa approssimare abbastanza bene con la curva normale appropriata.

Poiché una curva normale è completamente determinata dalla sua media e dallo scarto quadratico medio, la curva normale da usare qui è quella con la stessa media e lo stesso scarto quadratico medio della distribuzione binomiale.

$$\text{Quindi scegliamo } \mu = np = 12 \frac{1}{3} = 4 \text{ e } \sigma = \sqrt{npq} = \sqrt{12 \frac{1}{3} \frac{2}{3}} = \sqrt{\frac{8}{3}} = 1.63$$

Come prova in questo caso dell'accuratezza dell'approssimazione della curva normale e come esempio d'uso dei metodi che si basano sulla curva normale per approssimare le probabilità binomiali considereremo alcuni problemi riferiti alla figura 1.

Problema

Se la probabilità che un tiratore colpisca un bersaglio è $1/3$ e se tira 12 colpi, si vuole calcolare la probabilità che egli colpisca almeno 6 bersagli. La risposta esatta si ottiene sommando i valori della $f(x)$ da $x=6$ a $x=12$, che usando la tabella 1, è uguale a 0.178, valore corretto a tre cifre decimali,

$$\text{ovvero } P\{X \geq 6\} = \sum_{x=6}^{12} f(x) = 0.178.$$

Geometricamente questo valore rappresenta l'area di quella parte dell'istogramma di figura 1 che giace alla destra di $x = 5.5$.

Perciò per approssimare questa probabilità servendosi della curva normale è semplicemente necessario trovare l'area sottesa da quella parte della curva normale approssimante che giace alla destra di $x = 5.5$.

Poiché la curva approssimante ha $\mu = 4$ e $\sigma = 1.63$ segue che $z = \frac{(x - \mu)}{\sigma} = \frac{(5.5 - 4)}{1.63} = 0.92$, quindi

con questi valori di μ e σ , il cambiamento di variabile $z = \frac{(x - \mu)}{\sigma}$ fornisce che



$$\int_{5.5}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \int_{0.92}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = 0.179 \quad , \text{ ma come si può osservare nelle apposite}$$

tabelle, l'area alla destra di 0.92 è uguale a 0.179 che confrontata con il valore corretto 0.178 è sbagliata soltanto dello 0.56%, infatti l'errore relativo è

$$\frac{\Delta P}{P} = \frac{0.179 - 0.178}{0.178} = \frac{0.001}{0.178} = 0.0056 = 0.56\%$$

Per verificare l'accuratezza dei metodi che si basano sulla curva normale su un intervallo più breve, calcoliamo la probabilità che il tiratore abbia successo precisamente in 6 colpi su 12.

Dalla tabella 1 la risposta corretta a tre cifre decimali è $f(6) = 0.111$.

Per approssimare questo valore è necessario trovare l'area sottesa dalla curva normale

approssimante fra $x_1 = 5.5$ e $x_2 = 6.5$. In questo caso si ha così che $z_1 = \frac{x_1 - \mu}{\sigma} = \frac{5.5 - 4}{1.63} = 0.92$ z_1

$$\text{e } z_2 = \frac{x_2 - \mu}{\sigma} = \frac{6.5 - 4}{1.63} = 1.53.$$

Dalle apposite tavole si può osservare che l'area sottesa dalla curva normale standard fra $z = 0$ e $z_1 = 0.92$ è $\rightarrow A_1 = 0.3212$ e l'area sottesa fra $z = 0$ e $z_2 = 1.53$ è $\rightarrow A_2 = 0.4370$, perciò l'area richiesta è uguale a $A_2 - A_1 = 0.116$ che sbagliata della 4.5%, infatti l'errore relativo è

$$\frac{\Delta P}{P} = \frac{0.116 - 0.111}{0.111} = \frac{0.005}{0.111} = 0.045 = 4.5\%$$

Da questi due esempi, sembra che i metodi che si basano sulla curva normale siano abbastanza accurati anche per alcune situazioni come quella considerata qui, in cui n non è molto grande. Gli esempi precedenti sono stati dati per rendere credibile un famoso teorema che garantisce una buona approssimazione della curva normale alla distribuzione binomiale se n è sufficientemente grande. Questo teorema, che enunceremo senza dimostrazione, è il seguente:

Teorema

Se in n esecuzioni indipendenti di un esperimento, la variabile X rappresenta il numero di successi di un evento per cui p è la probabilità di successo in una singola esecuzione, allora la

variabile $Z = \frac{X - np}{\sqrt{npq}}$ ha una distribuzione che tende alla distribuzione normale con media

uguale a 0 e scarto quadratico medio uguale a 1, quando n tende a ∞ .

Questo teorema giustifica l'uso dei metodi che si basano sulla curva normale per approssimare le probabilità di un evento relative ad esecuzioni successive di un esperimento quando n è grande.

L'esperienza indica che l'approssimazione è abbastanza buona purché $np > 5$ quando $p \leq \frac{1}{2}$ e

purché $nq > 5$ quando $p > \frac{1}{2}$.

Le figure 2 e 3, indicano come rapidamente tende alla normalità la distribuzione della variabile

$\frac{X - np}{\sqrt{npq}}$, quando $p = 1/3$ e rispettivamente $n = 24$ e $n = 48$. La scala dell'asse y , per questi due

grafici è approssimante 17 volte quella dell'asse x . Le due approssimazioni della distribuzione binomiale, cioè la distribuzione di Poisson e quella normale, sono sufficienti per permettere di risolvere tutti i problemi più semplici che richiedono il calcolo di probabilità binomiali. Se invece n



è piccolo si usa direttamente la funzione di densità binomiale, perché i calcoli sono allora del tutto semplici.

