

Distribuzione normale o gaussiana

Senza dubbio il modello che si è dimostrato il più utile di tutte le distribuzioni di probabilità di una variabile casuale continua è la distribuzione normale o gaussiana.

In generale essa è definita come segue:

$$(4) f(x) = ce^{-\frac{1}{2}\left(\frac{x-a}{b}\right)^2}$$

dove a, b e c sono parametri che fanno della $f(x)$ una funzione di densità di probabilità, quindi questi

parametri devono essere tali che $\int_{-\infty}^{+\infty} f(x) = 1$, vedremo fra poco che ci sono in effetti soltanto due

parametri indipendenti che determinano questa funzione di densità.

Nella (4) è chiaro che la curva è simmetrica rispetto alla retta $x = a$, quindi la media per simmetria deve essere $\mu = a$.

Momenti

Calcoliamo ora i momenti indirettamente tramite la funzione generatrice dei momenti. Inoltre poiché è più facile in questo caso calcolare i momenti dalla media anziché quelli dall'origine, consideriamo il calcolo $M_{X-\mu}(\vartheta)$.

Dalla forma generalizzata della funzione generatrice dei momenti, con $g(X) = X - \mu$ si ottiene

$$M_{X-\mu}(\vartheta) = \int_{-\infty}^{+\infty} e^{\vartheta(x-\mu)} e^{-\frac{1}{2}\left(\frac{x-\mu}{b}\right)^2} dx.$$

$$\text{Ponendo } \begin{cases} z = \frac{x-\mu}{b} \\ dx = b dz \end{cases} \text{ si ha } M_{X-\mu}(\vartheta) = bc \int_{-\infty}^{+\infty} e^{\vartheta bz - \frac{z^2}{2}} dz$$

Riscriviamo ora l'esponente sotto il segno di integrale come segue

$$\vartheta bz - \frac{z^2}{2} = -\frac{1}{2}(z - \vartheta b)^2 + \frac{1}{2}\vartheta^2 b^2, \text{ e quindi si può scrivere}$$

$$M_{X-\mu}(\vartheta) = bce^{\frac{1}{2}\vartheta^2 b^2} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(z - \vartheta b)^2} dz$$

$$\text{Ponendo infine } \begin{cases} t = z - \vartheta b \\ dz = dt \end{cases}$$

$$\text{Si ha che } M_{X-\mu}(\vartheta) = bce^{\frac{1}{2}\vartheta^2 b^2} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt$$

Tenuto conto che l'integrale al secondo membro è un integrale noto che è uguale a $\sqrt{2\pi}$ si ha che

$$(5) M_{X-\mu}(\vartheta) = \sqrt{2\pi} bce^{\frac{1}{2}\vartheta^2 b^2}$$



Ricordando che $M_X(g) = 1 + \frac{g}{1!}\mu_1' + \frac{g^2}{2!}\mu_2' + \dots + \frac{g^k}{k!}\mu_k' + \dots$ si ha che per qualsiasi funzione generatrice dei momenti $M(0) = 1$ e quindi dalla (5) per $g = 0$ si ha $\sqrt{2\pi}bc = 1$ e quindi

$$(6) M_{X-\mu}(g) = e^{\frac{1}{2}g^2b^2}$$

Sviluppando ora l'esponenziale al secondo membro in serie di potenze si ha che

$$M_{X-\mu}(g) = 1 + b^2 \frac{g^2}{2} + b^4 \frac{g^4}{4} + \dots$$

Esaminando lo sviluppo al secondo membro si può osservare che mancano le potenze dispari di g e quindi che i momenti dispari della variabile X dalla sua media devono essere nulli.

In questo sviluppo il coefficiente di $\frac{g^2}{2!}$ è il momento del secondo ordine della variabile X dalla sua media, perciò $b^2 = \mu_2 = \sigma^2, b = \sigma$.

Poiché si è trovato che la $\sqrt{2\pi}bc = 1$ si ha che $c = \frac{1}{\sigma\sqrt{2\pi}}$ quindi la distribuzione normale definita dalla (4) si può scrivere come segue

$$(7) f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Questo risultato mostra che una distribuzione normale è completamente determinata specificando la sua media e il suo scarto quadratico medio.

Si fa notare che la sola differenza fra la (4) e la (7) è che i parametri a, b e c nella (4) sono ora ridotti a due soli parametri indipendenti cui è stato dato un preciso significato statistico.

Una formula per $M_X(g)$ espressa in termini di parametri statistici sarà necessaria in seguito, essa si può ottenere dalla (6) sostituendo b^2 con σ^2 (la varianza) e utilizzando la seconda delle due proprietà di una funzione generatrice dei momenti in forma generalizzata e cioè che

$M_{g(X)+c}(g) = eM_{g(X)}(g)$ con $g(X) = X$ e $c = -\mu$, così facendo si può scrivere

$$M_{X-\mu}(g) = e^{-\mu g} M_X(g).$$

Utilizzando ora al primo membro l'espressione data dalla (6) e risolvendo rispetto a $M_X(g)$ si

$$\text{ottiene } M_X(g) = e^{\mu g + \frac{1}{2}\sigma^2 g^2}.$$

Per interpretare lo scarto quadratico medio geometricamente consideriamo i punti di flesso di una curva normale, calcolando le prime due derivate della funzione di densità normale espressa dalla (7)



si ha che $f'(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} * \left[-\left(\frac{x-\mu}{\sigma}\right) * \frac{1}{\sigma}\right] = -\frac{1}{\sigma^2}(x-\mu)f(x)$ e

$$f''(x) = -\frac{1}{\sigma^2}f(x) - \frac{1}{\sigma^2}(x-\mu) \left[-\frac{1}{\sigma^2}(x-\mu)f(x) \right] = -\frac{1}{\sigma^2}f(x) \left[1 - \left(\frac{x-\mu}{\sigma}\right)^2 \right]$$

Da queste derivate è chiaro che c'è soltanto un punto di massimo che si trova in $x = \mu$

Dalla derivata seconda segue che i punti di flesso si trovano in $x = \mu \pm \sigma$

Geometricamente allora, per una distribuzione normale, lo scarto quadratico medio σ è la distanza dei punti di flesso dall'asse di simmetria, ciò implica che la curva normale rivolge la concavità verso il basso fra $\mu - \sigma$ e $\mu + \sigma$, e la concavità verso l'alto all'esterno di questo intervallo, il grafico di una curva normale è illustrato nella figura seguente

Fig 1)

Un ulteriore approfondimento sul ruolo che ha il parametro σ nel determinare la funzione di densità normale si ottiene calcolando l'area sottesa dalla funzione di densità normale negli intervalli mostrati in figura 1, così la probabilità che la variabile X , avente una funzione di densità normale,

cada all'interno dell'intervallo $(\mu - \sigma, \mu + \sigma)$ è data da $\int_{\mu-\sigma}^{\mu+\sigma} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$ ora ponendo

$$z = \frac{x-\mu}{\sigma} \text{ si ha } dx = \sigma dz \text{ e l'integrale diventa } \int_{-1}^{+1} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = 2 \int_0^1 \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} dz = 0.68$$

Fig. 2)

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

Il valore dell'ultimo integrale che si può trovare nelle apposite tavole è 0.3413, quindi moltiplicando per 2 il valore del primo integrale è 0.68

Per l'intervallo $(\mu - 2\sigma, \mu + 2\sigma)$ si può verificare che l'area sottesa è uguale a 0.95, in modo analogo l'area sottesa fra $(\mu - 3\sigma, \mu + 3\sigma)$ è uguale a 0.997, quindi riassumendo circa il 68% della probabilità giace nell'intervallo $(\mu - \sigma, \mu + \sigma)$, circa il 95% nell'intervallo $(\mu - 2\sigma, \mu + 2\sigma)$, e quasi tutta la probabilità, cioè il 99,7%, giace nell'intervallo $(\mu - 3\sigma, \mu + 3\sigma)$.

L'unità di misura data dalla trasformazione $z = \frac{x-\mu}{\sigma}$ si chiama unità standard e

corrispondentemente la distribuzione normale, espressa in funzione di z , e cioè con media uguale a 0 e scarto quadratico medio uguale a 1, prende il nome di **distribuzione normale in unità standard o in forma standard o standardizzata.**

