

2. Variabili casuali continue

2.1 Il passaggio dal discreto al continuo

2.1.1 Da cosa si origina il problema?

Nella descrizione di fenomeni che comportino misurazioni, soprattutto nelle scienze naturali, fisiche o nelle scienze economico sociali, si ha l'esigenza di operare con grandezze le cui proprietà sono descritte dai numeri reali. Ciò avviene se si devono sfruttare le proprietà di densità e continuità.

Ma quando $x \in \mathcal{D} \subseteq \mathfrak{R}$ non ha più senso assegnare alla condizione $X = x$ un valore di probabilità perché condizione necessaria affinché la somma di tutte le possibili probabilità faccia 1 è che si tratti di infinitesimi.

Per questa ragione invece di operare con le probabilità si opera con le *densità di probabilità* cioè con funzioni $f(x)$ tali che la probabilità infinitesima che $x < X < x + \delta x$ indicata con $\delta p = f(x) \delta x$.

La funzione $f(x) = \frac{\delta p}{\delta x}$ rinvia direttamente alle derivate mentre la funzione di ripartizione (somma di tanti δp) rinverrà agli integrali definiti.

Fatta questa premessa non si ripercorrerà noiosamente tutto il percorso già svolto sulle definizioni e proprietà delle variabili casuali discrete, perché ogni volta si tratterà di generalizzare le definizioni già date utilizzando gli integrali al posto delle \sum . Poiché sia la derivata, sia l'integrale sono operatori lineari tutte le proprietà di linearità già viste valgono anche nel continuo.

2.1.2 Densità di probabilità e funzione di ripartizione

[def] Una **variabile casuale continua** è una grandezza X caratterizzata da un dominio $\mathcal{D} \subseteq \mathfrak{R}$ a cui viene associata una funzione $f(x)$ detta *densità di probabilità* che deve essere in grado di generare una probabilità con il metodo indicato nel paragrafo precedente. Ciò richiede che

1. $\forall x, x \in \mathcal{D}$ sia $f(x) \geq 0$ e $dp = f(x) dx$ (1.1)
2. $f(x)$ ammetta una *funzione integrale* $F(x)$ in grado di generare la *funzione di ripartizione*¹
3. Poiché l'area estesa all'intero dominio deve dare l'evento certo (cioè si deve avere probabilità 1) deve essere $\int_{\mathcal{D}} dp = \int_{\mathcal{D}} f(x) dx = 1$ (1.2)

Se l'integrale definito (o generalizzato) esiste finito ma vale k basta considerare la funzione $\frac{1}{k}f(x)$ per ottenere una nuova funzione $g(x)$ che soddisfa tutte le richieste. Questa operazione si chiama *normalizzazione*.

[def] Data la variabile casuale continua X con densità di probabilità $f(x)$ indichiamo con m l'estremo inferiore del dominio (eventualmente $-\infty$).

Si chiama *funzione di ripartizione* $F(x) = p(X \leq x) = \int_m^x f(t) dt$ (1.3)

[N.B.] Quando si opera sulle variabili continue non esiste più la distinzione tra $<$ e \leq nelle funzioni di ripartizione perché la probabilità in corrispondenza di un singolo valore è sempre nulla.

[T.] $F(b) - F(a) = p(a \leq X \leq b)$ (1.4)

[dim.] $F(b) - F(a) = \int_m^b f(t) dt - \int_m^a f(t) dt = \int_m^b f(t) dt + \int_a^m f(t) dt = \int_a^b f(t) dt = p(a \leq X \leq b)$

☺

[T.] $p(X > b) = 1 - p(X \leq b)$ (1.5)

¹ Di solito si richiede che $f(x)$ sia continua ma ciò non è indispensabile. E' invece indispensabile la integrabilità con area finita

dim. Basta ragionare sul fatto che $X > b$ è l'evento contrario di $X \leq b$ oppure ragionare geometricamente in termini di area sottesa dai diagrammi.

☺

2.1.3 La media e la varianza

Le definizioni si ottengono generalizzando al continuo le definizioni date nel discreto. Pertanto:

def. *Valor medio* o valore atteso della variabile casuale:

$$\mu_x = \int_{\mathcal{D}} x dp = \int_{\mathcal{D}} x f(x) dx \quad (1.6)$$

Varianza della variabile casuale

$$\sigma_x^2 = \int_{\mathcal{D}} (x - \mu)^2 dp = \int_{\mathcal{D}} (x - \mu)^2 f(x) dx \quad (1.7)$$

Valgono tutte le proprietà già viste per il discreto e che derivano dalla linearità dell'operatore integrale).

Facciamo vedere, per esempio che

$$\sigma_x^2 = \mu_{x^2} - (\mu_x)^2$$

$$\sigma_x^2 = \int_{\mathcal{D}} (x - \mu)^2 f(x) dx = \int_{\mathcal{D}} (x^2 + \mu^2 - 2\mu x) f(x) dx = \int_{\mathcal{D}} x^2 f(x) dx + \mu^2 \int_{\mathcal{D}} f(x) dx + \int_{\mathcal{D}} -2\mu x f(x) dx = \mu_{x^2} +$$

$$\mu^2 \cdot 1 - 2\mu \int_{\mathcal{D}} x f(x) dx = \mu_{x^2} + \mu^2 - 2\mu \mu = \mu_{x^2} - (\mu_x)^2$$

☺

2.2 Esempi di densità di probabilità

2.2.1 La variabile aleatoria con densità costante

Sia $X = x$ con $x \in [a, b]$ e $f(x) = k$

Per la normalizzazione deve essere $k = \frac{1}{b-a}$

$$\mu_x = \int_a^b k x dx = k \frac{x^2}{2} \Big|_a^b = \frac{1}{2} k (b^2 - a^2) = \frac{1}{2} (b+a)$$

$$\int_a^b (x - \mu)^2 k dx = k \frac{(x - \mu)^3}{3} \Big|_a^b = \frac{k}{3} \left[\left(\frac{b-a}{2} \right)^3 - \left(-\frac{b-a}{2} \right)^3 \right] = \frac{1}{12(b-a)} (b-a)^3 = \frac{(b-a)^2}{12}$$

Si consiglia di ripetere il calcolo, per esercizio, usando la proprietà generale della varianza.

2.2.2 La variabile aleatoria con densità esponenziale decrescente

Sia $X = x$ con $x \in \mathcal{R}^+$ e $f(x) = k \exp(-bx)$

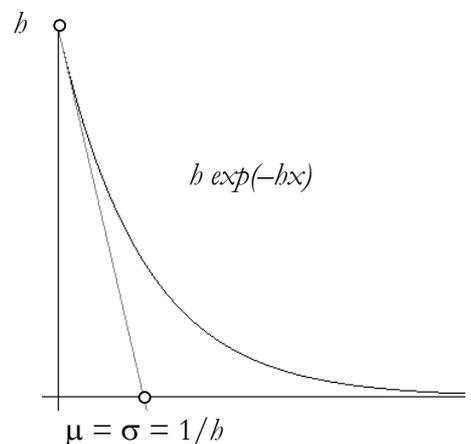
La costante k viene determinata dalla condizione di normalizzazione.

$$\int_0^{+\infty} k \exp(-bx) dx = k \int_0^{+\infty} \frac{-1}{b} \exp(-bx) d(-bx) = -\frac{k}{b} \exp(-bx) \Big|_0^{+\infty} = -\frac{k}{b} (0 - 1) = \frac{k}{b} = 1 \text{ da cui } k = b$$

[T.] Dunque $f(x) = b \exp(-bx)$ è una densità di probabilità.

Nella figura qui a lato sono rappresentate la funzione e la sua derivata in $x = 0$.

I prossimi due esercizi vengono proposti come un utile esercizio sugli integrali e riguardano il calcolo del valor medio e della varianza.



Per definizione: $\mu_x = \int_0^{+\infty} b x \exp(-bx) dx$ e questo integrale si fa per parti dopo aver posto $z = bx$ si ha $dz = b dx$ (si conserva l'intervallo di integrazione).

$$\mu_x = \frac{1}{b} \int_0^{+\infty} z \exp(-z) dz = \frac{1}{b} \int_0^{+\infty} z d(\exp(-z)) = \frac{1}{b} (z \exp(-z) + \exp(-z)) \Big|_0^{+\infty} = \frac{1}{b} \exp(-z) (z + 1) \Big|_0^{+\infty} = \frac{1}{b} (0 - 1) = \frac{1}{b}$$

[T.] Dunque l'intercetta della retta tangente è il valor medio. ²

Per calcolare la varianza passeremo attraverso il calcolo di μ_{x^2} perché ciò consente di procedere in modo più spedito nei calcoli.

$\mu_{x^2} = \int_0^{+\infty} b x^2 \exp(-bx) dx$ questa volta useremo la sostituzione $z = -bx$ con $dz = -b dx$ (si ottengono gli opposti nell'intervallo di integrazione).

² La dimostrazione è identica a quella con cui si dimostra che la vita media di un materiale radioattivo è l'inverso della costante di decadimento.

$$\begin{aligned} \mu_{x^2} &= b \int_0^{-\infty} \frac{x^2}{b^2} e^{-x/b} \left(-\frac{1}{b} dx\right) = -\frac{1}{b^2} \int_0^{-\infty} x^2 d(e^{-x/b}) = -\frac{1}{b^2} \left[x^2 e^{-x/b} - \int 2x e^{-x/b} dx \right]_0^{-\infty} = -\frac{1}{b^2} \left[x^2 e^{-x/b} - 2 \left(x e^{-x/b} - e^{-x/b} \right) \right]_0^{-\infty} = \\ &= -\frac{1}{b^2} e^{-x/b} (x^2 - 2x + 2) \Big|_0^{-\infty} = -\frac{1}{b^2} (0 - 2) = \frac{2}{b^2} \end{aligned}$$

Dunque:

$$\sigma^2 = \mu_{x^2} - \mu_x^2 = \frac{2}{b^2} - \frac{1}{b^2} = \frac{1}{b^2} \text{ e } \sigma = \frac{1}{b}$$

T. La deviazione standard è uguale al valor medio

2.3 La funzione gaussiana

2.3.1 Studiamo l'andamento della curva a campana $y = f(x) = e^{-1/2 x^2}$

La funzione $y = f(x) = e^{-1/2 x^2}$ e le funzioni ad essa correlate rivestono una notevole importanza in statistica perché le *medie campionarie* quale che sia la loro origine tendono a disporsi secondo il tipi andamento a campana.

dominio il dominio è \mathbb{R} ; inoltre si tratta di una funzione pari e pertanto limiteremo il C.d.S. (campo di studio) a \mathbb{R}^+

comportamento agli estremi

$$f(0) = 1$$

$$\lim_{x \rightarrow +\infty} e^{-1/2 x^2} = (e^{-\infty}) = 0^+ \text{ asse } x \text{ asintoto orizzontale}$$

creștere e decrescere

$$f'(x) = -x e^{-1/2 x^2} \geq 0 \Leftrightarrow x = 0$$

La funzione è sempre decrescente e presenta in $x = 0$ un punto a tangente orizzontale (Massimo assoluto)

Inoltre per l'esistenza dell'asintoto orizzontale presenterà almeno un flesso discendente

concavità

$$f''(x) = -(e^{-1/2 x^2} - x e^{-1/2 x^2}) = e^{-1/2 x^2} (x - 1) \geq 0 \Leftrightarrow x \geq 1$$

La funzione è concava verso il basso in $[0,1]$ e concava verso l'alto in $[1,+\infty]$, ha un flesso F in $(1, e^{-1/2})$ con inclinazione della retta tangente $f'(1) = -e^{-1/2}$

normalizzazione

Nonostante $e^{-1/2 x^2}$ sia una funzione continua e definita in maniera molto semplice non esiste alcuna funzione esprimibile in forma elementare che rappresenti la funzione integrale.

Il calcolo dell'area viene ottenuto come caso particolare di un integrale più generale, quello della funzione:

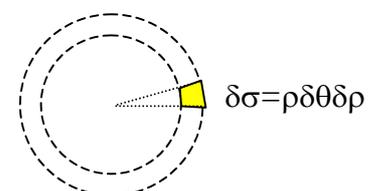
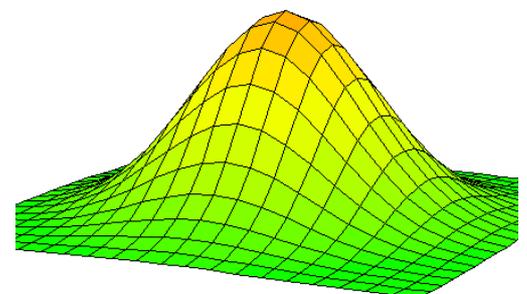
$z = \exp(-x^2 - y^2)$ di cui abbiamo rappresentato l'andamento qui a lato. Si tratta ancora della curva a campana ma questa volta con una simmetria di rotazione intorno all'asse z .

L'integrale $I_R = \int_{\mathcal{D}} \int \exp(-x^2 - y^2) d\sigma$ viene eseguito sul dominio \mathcal{D}

rappresentato dal cerchio di equazione $x^2 + y^2 = R^2$ dopo aver eseguito un passaggio in coordinate polari che consente di sfruttare meglio la simmetria della funzione:

$x^2 + y^2 = \rho^2$ mentre l'area elementare è un rettangoloide di lati $\rho d\theta$ e $d\rho$.³

$$I_R = \int_0^{2\pi} \left(\int_0^R \exp(-\rho^2) \rho d\rho \right) d\theta = -1/2 \int_0^{2\pi} \exp(-\rho^2) \Big|_0^R d\theta = -1/2 2\pi (\exp(-R^2) - 1) =$$



³ Viene esposta la sostanza dei ragionamenti senza badare agli aspetti di natura formale.

$$\pi(1 - \exp(-R^2)).$$

$$I_\infty = \lim_{R \rightarrow +\infty} I_R = \pi$$

Si sfrutta ora la perfetta simmetria tra x e y per osservare che $I = \int_{-\infty}^{+\infty} \exp(-x^2) dx = \int_{-\infty}^{+\infty} \exp(-y^2) dy$ ma

$$I^2 = \int_{-\infty}^{+\infty} \exp(-x^2) dx \cdot \int_{-\infty}^{+\infty} \exp(-y^2) dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp(-x^2 - y^2) dx dy = \pi$$

e dunque

$$I = \int_{-\infty}^{+\infty} \exp(-x^2) dx = \sqrt{\pi}$$

Basta ora eseguire un cambio di variabile $\frac{x}{\sqrt{2}} = X$ da cui $\frac{dx}{\sqrt{2}} = dX$ per ottenere

$$\int_{-\infty}^{+\infty} \exp(-1/2 x^2) dx = \sqrt{2} \int_{-\infty}^{+\infty} \exp(-X^2) dX = \sqrt{2\pi}$$

☺

2.3.2 La funzione normale standardizzata di Gauss

[T.] Per quanto è stato appena dimostrato

$$y = f(z) = \frac{1}{\sqrt{2\pi}} \exp(-1/2 z^2) \quad (3.1)$$

rappresenta una *densità di probabilità* e per ragioni che vedremo tra breve essa è detta *funzione normale standardizzata di Gauss* e la corrispondente variabile Z è detta variabile gaussiana standardizzata.

Rispetto a quanto è già stato visto nel paragrafo precedente l'unica differenza sta nella presenza del fattore moltiplicativo (che modifica le ordinate dei punti notevoli) $\frac{1}{\sqrt{2\pi}}$ necessario alla normalizzazione della funzione.

Calcoliamo ora gli indicatori statistici che caratterizzano la variabile Z

$$\mathbf{[T.]} \quad \mu_Z = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z \exp(-1/2 z^2) dz = 0 \quad (3.2)$$

[dim.] Per la dimostrazione possiamo evitare qualsiasi calcolo. Infatti la funzione è dispari e poiché gli estremi di integrazione sono simmetrici il risultato dell'integrale è certamente 0.

☺

$$\mathbf{[T.]} \quad \sigma_Z^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z^2 \exp(-1/2 z^2) dz = 1 \quad (3.3)$$

[dim.] Questa volta la simmetria non ci aiuta perché la funzione è pari e dunque eseguiremo il calcolo che richiede una integrazione per parti:

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z^2 \exp(-1/2 z^2) dz &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z d(\exp(-1/2 z^2)) = \frac{-1}{\sqrt{2\pi}} \left[z \exp(-1/2 z^2) - \int \exp(-1/2 z^2) dz \right]_{-\infty}^{+\infty} = \\ &= \frac{-1}{\sqrt{2\pi}} (0 - \sqrt{2\pi}) = 1 \end{aligned}$$

☺

Dunque, nella gaussiana standardizzata il punto di flesso corrisponde alla varianza.

Questa proprietà, come vedremo tra breve, vale anche per la generica funzione gaussiana che ci apprestiamo a studiare.

2.3.3 La funzione normale di Gauss

Consideriamo una variabile x tale che $\frac{x-b}{k} = z$

Per le proprietà generali della media e della varianza avremo che:

$$\mu_z = \frac{1}{k} (\mu_x - b) \text{ e dunque da } \mu_z = 0 \Rightarrow \mu_x = b$$

$$\sigma_z^2 = \frac{1}{k^2} \sigma_x^2 \text{ e dunque da } \sigma_z^2 = 1 \Rightarrow \sigma_x^2 = k^2$$

Calcoliamo infine l'area della funzione $f(x)$ ai fini della necessaria normalizzazione:

$$y = f(x) = \exp \frac{-(x-b)^2}{2k^2}. \text{ Allo scopo eseguiamo la sostituzione } \frac{x-b}{k} = z \text{ da cui } \frac{1}{k} dx = dz$$

$$\int_{-\infty}^{+\infty} \exp \frac{-(x-b)^2}{2k^2} dx = \int_{-\infty}^{+\infty} k \exp(-1/2 z^2) dz = \sqrt{2\pi} k \text{ e questo è il coefficiente da utilizzare per la normalizzazione.}$$

$$\boxed{T.} \text{ La } y = f(x) = \frac{1}{\sqrt{2\pi}k} \exp \frac{-(x-b)^2}{2k^2}$$

detta *funzione normale* o *funzione gaussiana* è una densità di probabilità di valore medio b e di varianza k^2 .

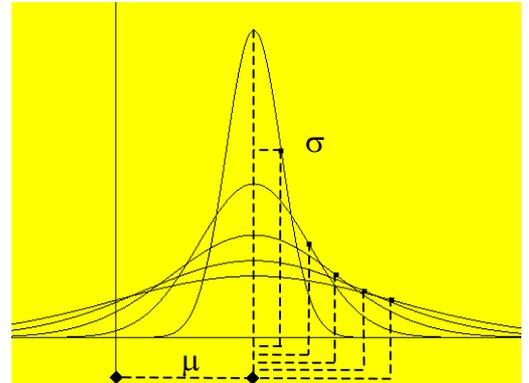
Per questa ragione la si scrive solitamente

$$y = f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \frac{-(x-\mu)^2}{2\sigma^2} \quad (3.4)$$

Il punto di massimo ha coordinate $\left(\mu, \frac{1}{\sqrt{2\pi}\sigma}\right)$ mentre il punto di

flesso ha coordinate $\left(\mu + \sigma, \frac{1}{\sqrt{2\pi}\sigma} \exp(-1/2)\right)$

Nella figura qui a lato sono state raffigurate alcune curve gaussiane con uno stesso valore di μ e con valori della deviazione standard $\sigma =$ rispettivamente $\{0.2\mu, 0.4\mu, 0.6\mu, 0.8\mu, 1.0\mu\}$



2.3.4 La funzione di ripartizione, le sue proprietà, la sua rappresentazione

La funzione di ripartizione della gaussiana standardizzata non è esprimibile in forma analitica ed è resa disponibile in forma tabulata o tramite le apposite funzioni dei programmi statistici.

Nei diversi testi sono riportate due funzioni di ripartizione tra loro strettamente imparentate: quella indicata con $F(z)$ che calcola l'integrale da $-\infty$ e quella indicata con $\Phi(z)$ che lo calcola da 0.

Poiché per via della simmetria e della normalizzazione $F(0) = 1/2$ si ha che:

$$F(z) = 1/2 + \Phi(z) \quad (3.5)$$

e dunque è sempre possibile passare dall'una all'altra.

Le tavole riportano i valori della sola variabile standardizzata e si limitano a quelli positivi compresi da 0 e 3 (max 4) perché quelli negativi possono essere calcolati sfruttando la simmetria e quelli maggiori di 4 corrispondono a valori talmente prossimi a 1 da dare scarso significato alla distinzione (si ricordi che la gaussiana viene utilizzata facendo assunzioni di semplificazione che rendono insensato lavorare con più di 4 o 5 cifre significative).

Poiché l'area a sinistra di $-z$ e quella a destra di z sono uguali si ha:

$$F(-z) = 1 - F(z) \quad (3.6)$$

La distribuzione normale consente di precisare in maniera ben delimitata la probabilità che una variabile casuale cada entro intervalli intorno a μ definiti dal valore di σ .

Si cita in particolare la *legge dei 3 σ* che corrisponde a precisare la probabilità che la variabile cada entro intervalli di semiampiezza σ , 2σ , 3σ . In questi casi si ha $z = 1, 2, 3$ e dunque:

$$p(|x - \mu| \leq \sigma) = p(|z| \leq 1) = F(1) - F(-1) = F(1) - (1 - F(1)) = 2F(1) - 1 = 2 \cdot 0.841345 - 1 = 0.68269$$

$$p(|x - \mu| \leq 2\sigma) = p(|z| \leq 2) = F(2) - F(-2) = 2F(2) - 1 = 2 \cdot 0.97725 - 1 = 0.9545$$

$$p(|x - \mu| \leq 3\sigma) = p(|z| \leq 3) = F(3) - F(-3) = 2F(3) - 1 = 2 \cdot 0.99865 - 1 = 0.9973$$

E' questa la ragione per cui nelle scienze sperimentali si ragiona in σ . Parlare di una semidispersione di 3σ vuol dire affermare che la probabilità che la grandezza cada in quell'intervallo è del 99.7% cioè praticamente certa.

In Excel sono disponibili 4 funzioni che hanno a che fare con la gaussiana. Esse sono:

=distrib.norm(x; μ ; σ ;*logico*) che fornisce la densità di probabilità (falso) e la funzione di ripartizione(vero) per una variabile casuale di media μ e varianza σ^2 .

=distrib.norm.st(x) che fornisce la funzione di ripartizione per una variabile gaussiana standardizzata ed è dunque un sottoinsieme della funzione precedente

=inv.norm(p; μ ; σ) che dà la funzione inversa di F(x) per una variabile gaussiana

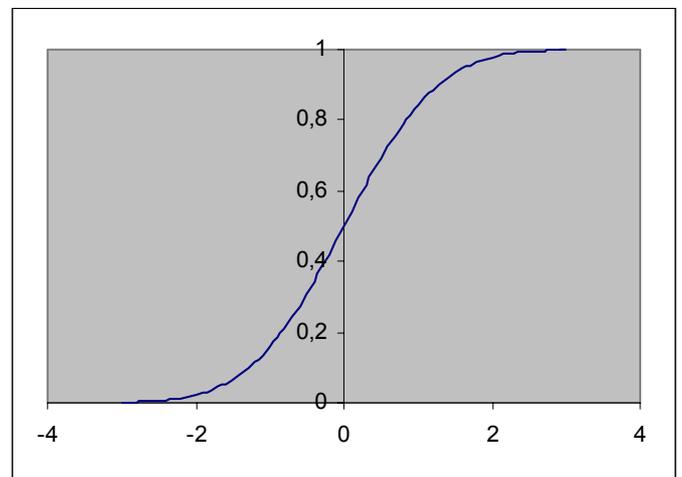
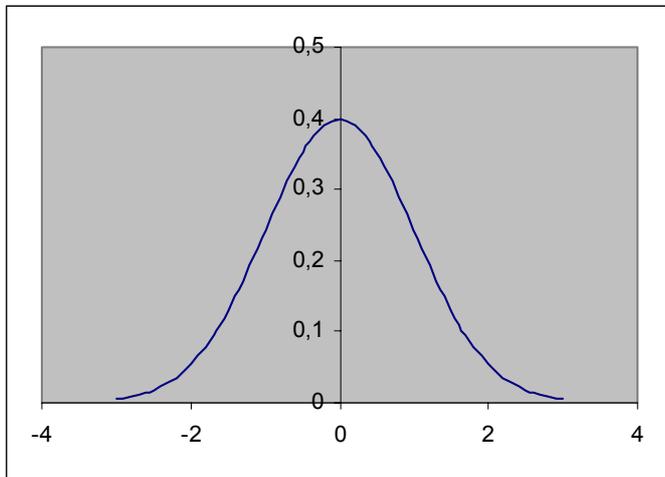
=inv.norm.st(p) che dà la funzione inversa di F(x) per una variabile gaussiana standardizzata

Il loro utilizzo è molto semplice e comodo e il lettore è invitato ad esercitarsi al loro utilizzo.

A titolo di esempio riportiamo i valori della densità e della probabilità per valori da 0 a 3 con incremento 0.1

0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1	1,1	1,2	1,3	1,4
0,398942	0,396953	0,391043	0,381388	0,36827	0,352065	0,333225	0,312254	0,289692	0,266085	0,241971	0,217852	0,194186	0,171369	0,149727
0,50539828	0,57926	0,617911	0,655422	0,691462	0,725747	0,758036	0,788145	0,81594	0,841345	0,864334	0,88493	0,903199	0,919243	

1,5	1,6	1,7	1,8	1,9	2	2,1	2,2	2,3	2,4	2,5	2,6	2,7	2,8	2,9	3
0,129518	0,110921	0,094049	0,07895	0,065616	0,053991	0,043984	0,035475	0,028327	0,022395	0,017528	0,013583	0,010421	0,007915	0,005953	0,004432
0,933193	0,945201	0,955435	0,96407	0,971284	0,97725	0,982136	0,986097	0,989276	0,991802	0,99379	0,995339	0,996533	0,997445	0,998134	0,99865



2.3.5 Esempi di utilizzo

es.1 Una variabile gaussiana X presenta valore medio $\mu = 2.54$ e $\sigma = 1.20$ determinare la probabilità che $1.95 \leq X \leq 2.65$

Per prima cosa occorre procedere alla standardizzazione della variabile tramite la sostituzione $Z = \frac{X - \mu}{\sigma}$

X	Z
1.95	-0.49167
2.65	0,091667

a questo punto si calcola mediante tabella o con Excel il valore di F(z)

$$F(-0.49167) = 1 - F(0.49167) = 0,311478$$

$$F(0,091667) = 0,536519$$

E infine $p(1.95 \leq X \leq 2.65) = F(0,091667) - F(-0.49167) = 0,225041$

es.2 Una variabile gaussiana X presenta una probabilità cumulativa del 46.8%. Quando tale variabile aumenta del 10% la probabilità passa al 53.4%. Inoltre, in corrispondenza del suo valore modale si ha $f(x) = 0.22$. Determinare il valore di x e gli indicatori statistici μ e σ .

Dalla tavole, lette al contrario, o tramite l'uso della funzione $\text{inv.norm.st}(p)$ con $p = 0.468$ si trova il valore z della variabile standardizzata $z_1 = -0.080298$. Analogamente si trova $z_2 = 0,085329$

Il valore modale corrisponde a $x = \mu$ e pertanto, in quel caso, $f(x) = \frac{1}{\sqrt{2\pi}\sigma}$.

Ciò ci permette di calcolare $\sigma = \frac{1}{0.22 \cdot \sqrt{2\pi}} = 1.813$

Se indichiamo con x il primo dei due valori, il secondo sarà $1.1x$ e avremo pertanto

$$x - \mu = 1.813 \cdot (-0.080298) = -0.14561$$

$$1.1x - \mu = 1.813 \cdot (0,085329) = 0.154733$$

Da qui si ottiene $0.1x = 0,300334$ e dunque $x = 3.00334$

$$\mu = 3.00334 + 0.14561 = 3.149$$

Riassumiamo le soluzioni con 3 cifre decimali di precisione

$$\mu = 3.149, \sigma = 1.813, x = 3.003$$

es.3 Una variabile gaussiana ha $p(X \leq x_1) = 0.903291$ con $x_1 = 5.80$. Inoltre con $x_2 = 8.60$ si ha $p(x_1 \leq X \leq x_2) = 0.092450$; determinare μ e σ .

Attraverso la tabella o la funzione inv.norm.st si determina $z_1 = 1.300534$ ⁴

$$F(z_2) = p(x_1 \leq X \leq x_2) + F(z_1) = 0.092450 + 0.903291 = 0.995741$$

Attraverso la tabella o la funzione inv.norm.st si determina $z_2 = 2.630841$

Dunque scrivendo le relazioni di standardizzazione $Z = \frac{X - \mu}{\sigma}$

$$\frac{5.80 - \mu}{\sigma} = 1.300534 \quad \frac{8.60 - \mu}{\sigma} = 2.630841 \text{ e da qui:}$$

$$\sigma = \frac{\Delta x}{\Delta z} = 2.105 \text{ e } \mu = x - \sigma z = 3.063$$

es.4 Entro quale intervallo simmetrico rispetto al valor medio una variabile gaussiana cade con probabilità $p = 0.850$?

Bisogna trovare z | $F(z) - F(-z) = 0.850$.

Ma $F(-z) = 1 - F(z)$ e pertanto deve essere $2F(z) - 1 = 0.850$ ovvero $F(z) = 0.925$

Attraverso la tabella o la funzione inv.norm.st si determina $z = 1.4395$

Dunque l'intervallo richiesto è $|x - \mu| \leq 1.44\sigma$

2.3.6 Perché la gaussiana è importante? Il teorema del limite centrale

L'importanza della gaussiana nasce da due questioni:

- la gaussiana rappresenta con grande accuratezza il modo con cui si distribuiscono gli errori nelle misurazioni ripetute
- la gaussiana è in grado di approssimare le distribuzioni discrete (sia la bernoulliana, sia la poissoniana) quando queste risultano inutilizzabili sul piano pratico (difficoltà di calcolo).

Lo studio di questa curva è dovuto a tre matematici:

De Moivre (1667-1754) nella *Doctrine of Chance* ne inquadra le caratteristiche generali nel quadro dei primi studi sistematici sulla probabilità e fa vedere che la distribuzione bernoulliana è approssimabile tramite la normale;

Pierre Simon Laplace nella *teoria analitica della probabilità* (1812) enuncia il teorema che consente di applicare la distribuzione normale in modo sistematico alla statistica

Karl Friedrich **Gauss** (1777–1855) ne studia le caratteristiche e la applica sistematicamente alla teoria degli errori dal 1809 in poi (studi di topografia e di astronomia).

⁴ Con le tabelle si opera con una precisione inferiore

La teoria degli errori si basa sulla presa d'atto che esiste una regolarità empirica secondo cui gli errori si distribuiscono in modo normale. Ciò consente di dare un senso al *valore vero* di una grandezza fisica affermando che esso cade, con una certa probabilità quantificabile, entro un certo intervallo collocato simmetricamente intorno al valor medio della gaussiana.

Dal punto di vista matematico la gaussiana è interessante perché rappresenta il limite della bernoulliana per $n \rightarrow \infty$ anche nel caso in cui non sia $np = \text{costante}$ (che consente di utilizzare la poissoniana).

Il teorema cui ci si riferisce per l'utilizzo della gaussiana in ambito statistico (teorema che riprenderemo anche in statistica inferenziale quando ci porremo il problema di fare previsioni su una popolazione attraverso il comportamento di campioni estratti da essa) è il *teorema del limite centrale* dovuto a Laplace e del quale De Moivre aveva già dato un enunciato con riferimento a variabili casuali bernoulliane.

[T.] Siano X_1, X_2, \dots, X_n delle variabili casuali indipendenti ed equidistribuite (stessi valori di μ e σ).

Indicata con $S_n = \sum_{i=1}^n X_i$ si consideri la variabile standardizzata $Z = S_n^* = \frac{S_n - \mu_{S_n}}{\sigma_{S_n}} = \frac{S_n - n\mu}{\sqrt{n}\sigma}$ di valore medio 0

e varianza 1 allora:

$$p(S_n^* \leq z) \approx F(z) \quad (3.7)$$

[N.B.] Invece di lavorare su S_n si può, egualmente bene lavorare sulla variabile $Y_n = S_n/n$ con $\mu_Y = \mu$ e $\sigma_Y =$

$$\frac{1}{\sqrt{n}} \sigma \text{ a cui corrisponde la variabile standardizzata } Y_n^* = \frac{(Y_n - \mu)\sqrt{n}}{\sigma}$$

Se ipotizziamo che le X_i siano variabili di tipo bernoulliano con valori 0 e 1 associati alla probabilità p la

variabile standardizzata Z diventa $\frac{S_n - np}{\sqrt{np(1-p)}}$

Per quanto riguarda la approssimazione della binomiale tramite la gaussiana è necessario dare un criterio visto che non ha senso operativo la frase $n \rightarrow \infty$. Concretamente si chiede che n sia abbastanza grande e p e $1-p$ non siano troppo piccole e per farlo si calcola l'intervallo $[\mu - 3\sigma, \mu + 3\sigma]$. Se esso è contenuto in $[0, n]$ si considera accettabile la approssimazione.

[es. 5] In un processo produttivo si presentano pezzi difettosi nel 2.4% dei casi. Calcolare la probabilità che su uno stock di $n = 3'000$ beni prodotti se ne presentino non più di 65 difettosi.

Svolgere l'esercizio utilizzando la approssimazione gaussiana e quindi, utilizzando le potenzialità di calcolo di Excel, determinare il valore non approssimato fornito dalla distribuzione binomiale.

La variabile che descrive i pezzi difettosi è $X_i = \begin{bmatrix} 0 & 1 \\ 0.976 & 0.024 \end{bmatrix}$ con $\mu = p = 0.024$ e $\sigma = \sqrt{p(1-p)} = 0.153$

La variabile S_n che prende valori $x = 0, 1, 2, \dots$ segue la distribuzione binomiale e descrive la probabilità di trovare x pezzi difettosi. Per rispondere alla domanda ci servirà calcolare la relativa funzione di ripartizione.

$$\mu_{S_n} = np = 72.0 \text{ mentre } \sigma_{S_n} = \sqrt{n} \sigma = 8.38$$

Se standardizziamo la variabile avremo $z = \frac{x - 72.0}{8.38}$ e per $x = 65$ $z = -0.835$

$$p(S_n \leq 65) \approx F(-0.835) = 1 - F(0.835) = 0.201848$$

Se volessimo calcolare il valore esatto dovremmo valutare l'espressione

$$p(S_n \leq 65) = \sum_{k=0}^{65} \binom{3000}{k} 0.024^k 0.976^{3000-k}$$

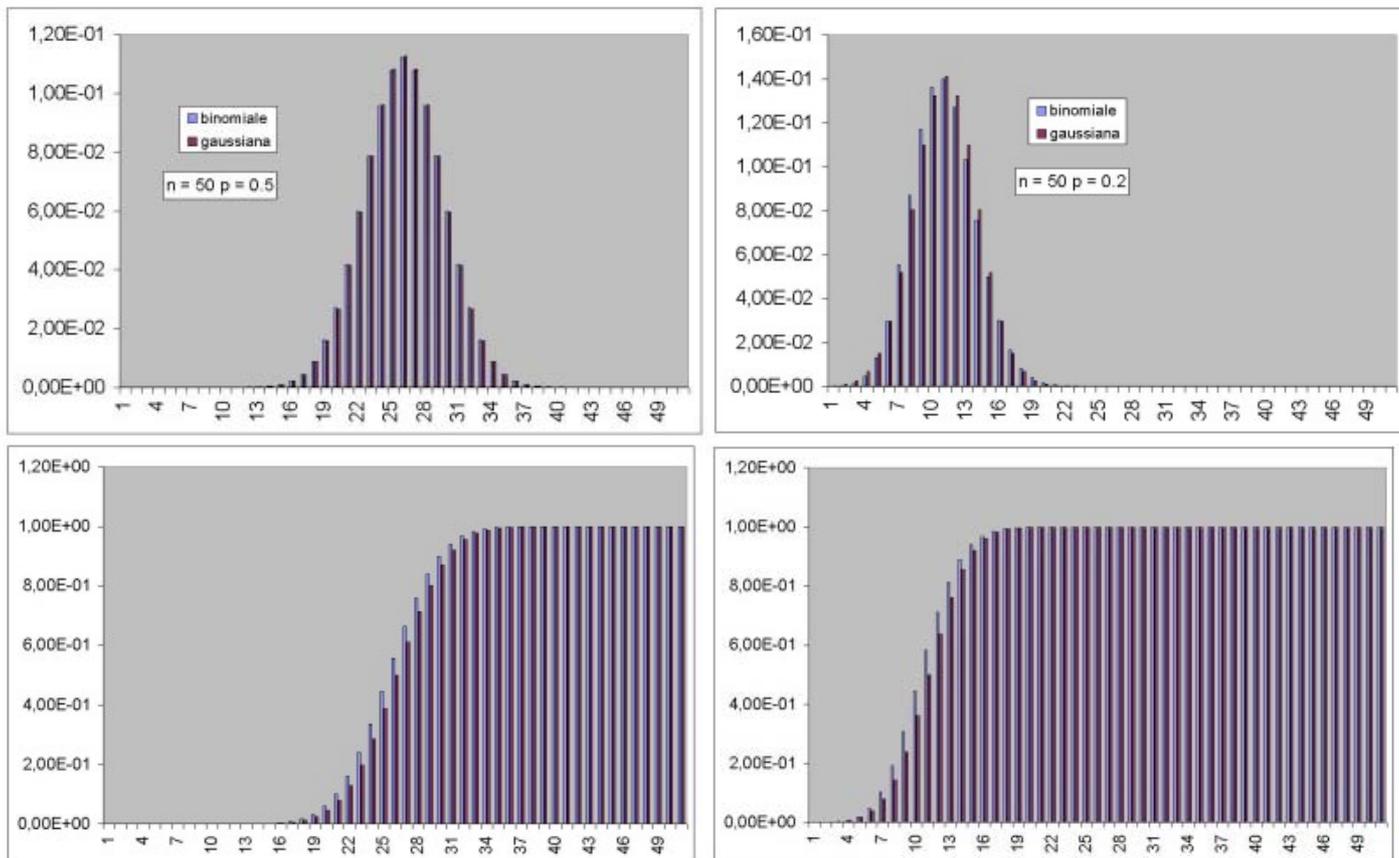
Per farlo utilizziamo la funzione di Excel =distrib.binom($k;n;p;vero$) che ci fornisce la funzione di ripartizione della distribuzione binomiale per n prove e k successi con probabilità p .

Così facendo si ottiene $p(S_n \leq 65) = 0.2212$ con un errore dell'ordine di $2/100$.

Ma c'è di più se si opera con valori di n di poco superiori a 3000 anche Excel va in overflow e dunque, a meno di usare programmi specialistici di matematica, come Derive o Matematica non ci resta che accontentarci delle approssimazioni ricordando che, per altro al crescere di n la capacità della gaussiana di ben approssimare la binomiale aumenta.

Osserviamo infine che, rispetto al criterio di utilizzabilità posto in premessa all'esercizio siamo all'interno dell'intervallo $[\mu - 3\sigma, \mu + 3\sigma]$.

Per dare al lettore un'idea del rapporto tra binomiale e gaussiana abbiamo prodotto due immagini di confronto scegliendo $n = 50$ e assumendo per $p = 0.5$ e 0.2 . Nel primo caso la concordanza delle due curve è molto buona mentre nel secondo caso si notano le prime differenze. Non si sono riprodotte immagini con valori di n più elevati perché l'alto numero di barre dell'istogramma rende disagevole la discriminazione.



Nei due casi considerati si ha rispettivamente:

n	p	μ	σ	$\mu - 3\sigma$	$\mu + 3\sigma$
50	0.2	10	2.8	1.5	18.5
50	0.5	25	3.5	14.4	35.6