Vittorio Casella

DIET – Università di Pavia

email: vittorio.casella@unipv.it

La variabile casuale di Gauss

Dispense

Gauss

Carl Friedrich Gauss (Braunschweig, 30 aprile 1777 – Gottinga, 23 febbraio 1855) è stato un matematico, astronomo e fisico tedesco, che ha fornito contributi determinanti all'analisi matematica, teoria dei numeri, calcolo numerico, geometria differenziale, geodesia, magnetismo e ottica.

[http://it.wikipedia.org/wiki/Carl_Friedrich_Gauss]

La gaussiana – forma funzionale

Fra le infinite f_X, ognuna rappresentante una vc, più o meno interessante, esiste una famiglia a 2 parametri

$$f_N(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

dove

$$\mu \in \mathbb{R}$$

$$\sigma \in \mathbb{R}, \sigma > 0$$

E' detta vc *normale o di Gauss*. La vc normale di parametri μ e σ è indicata anche con

$$\mathrm{N}(\mu,\sigma)$$

E’ detta normale perché le vc che seguono altre distribuzioni costituiscono un’eccezione.

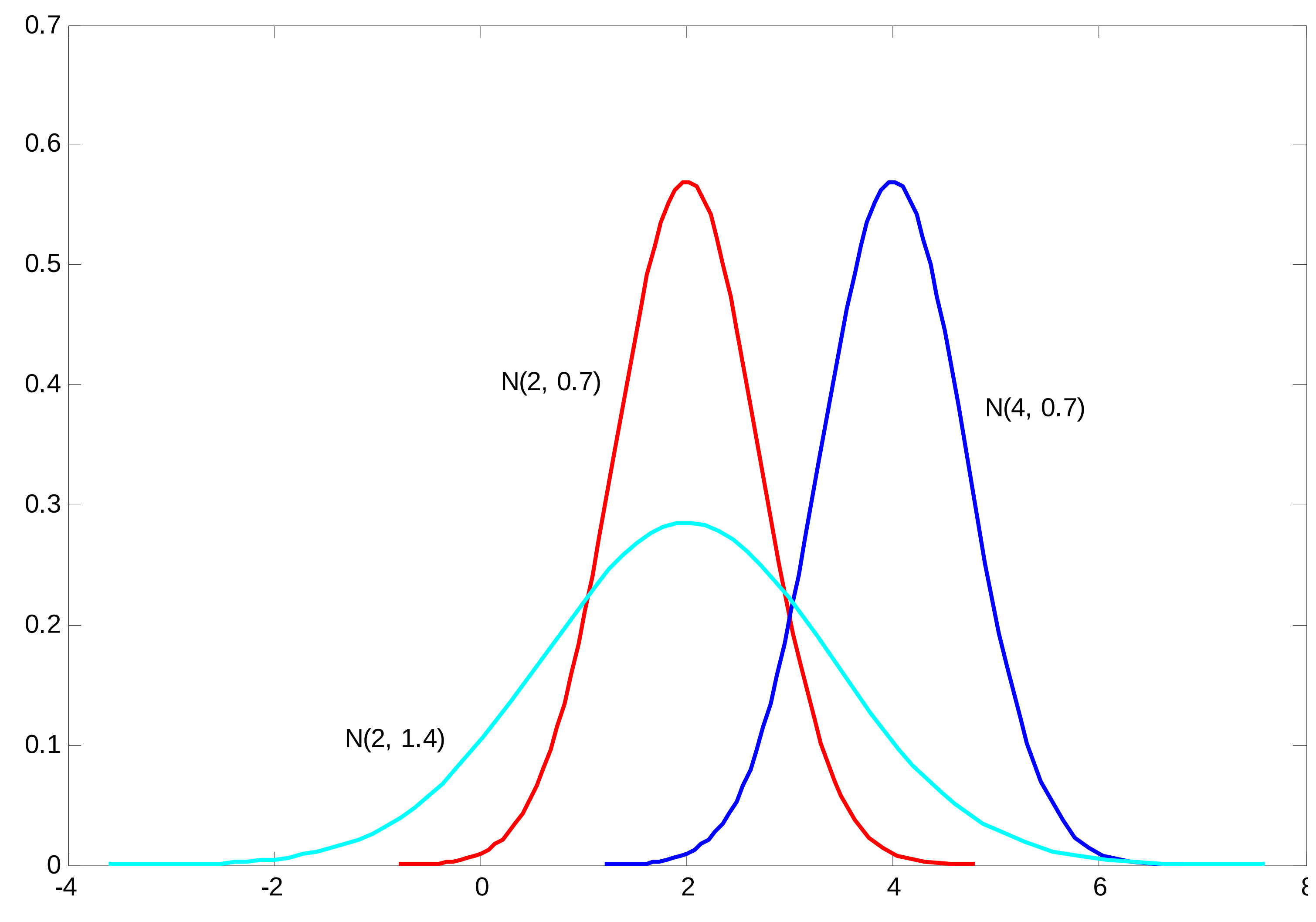
Terminologia

μ - media

σ - deviazione standard

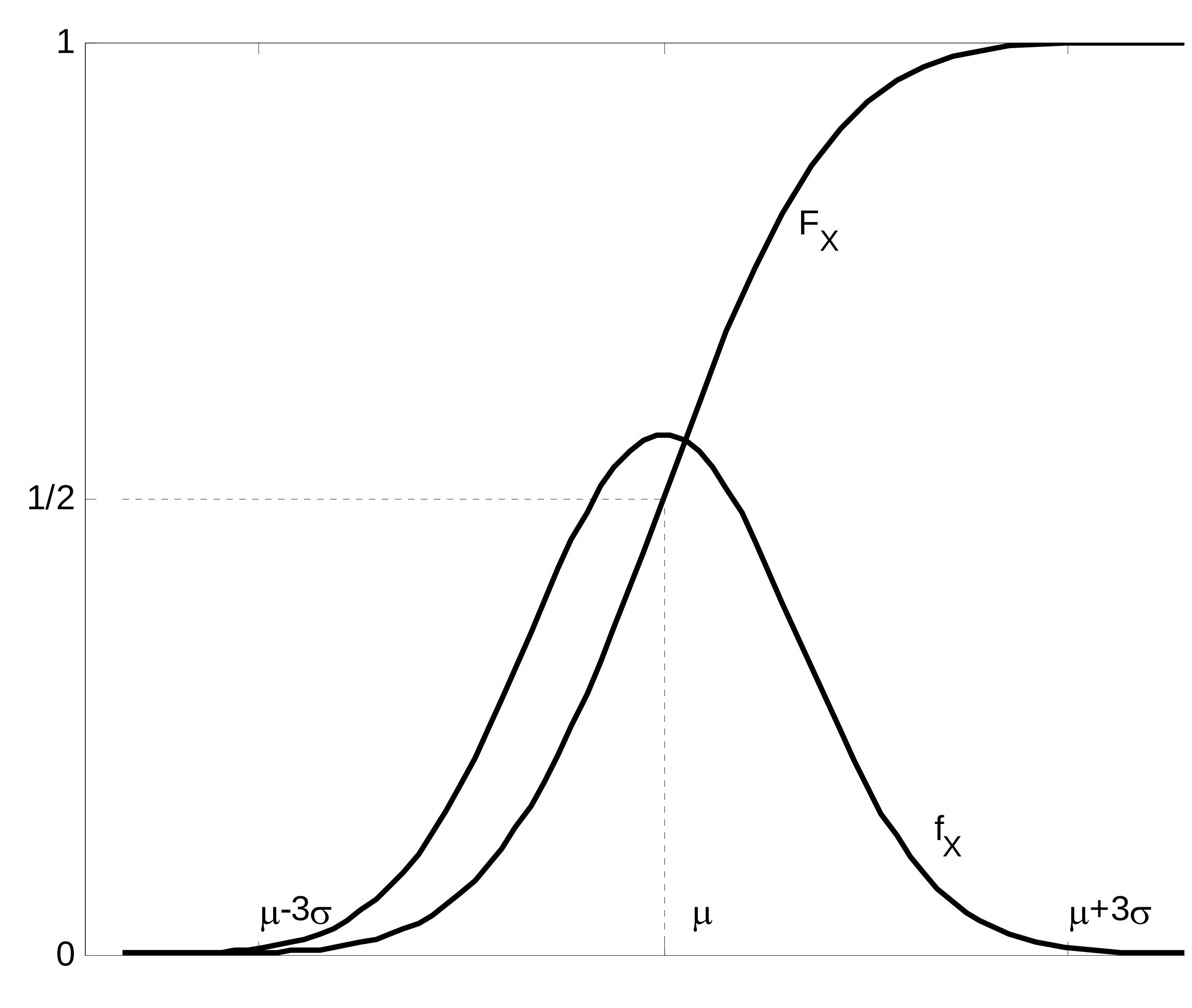
σ^2 - varianza

La gaussiana – grafico



- Tipica forma a campana
- Il picco si trova in corrispondenza della media μ
- Il parametro σ controlla la larghezza della campana

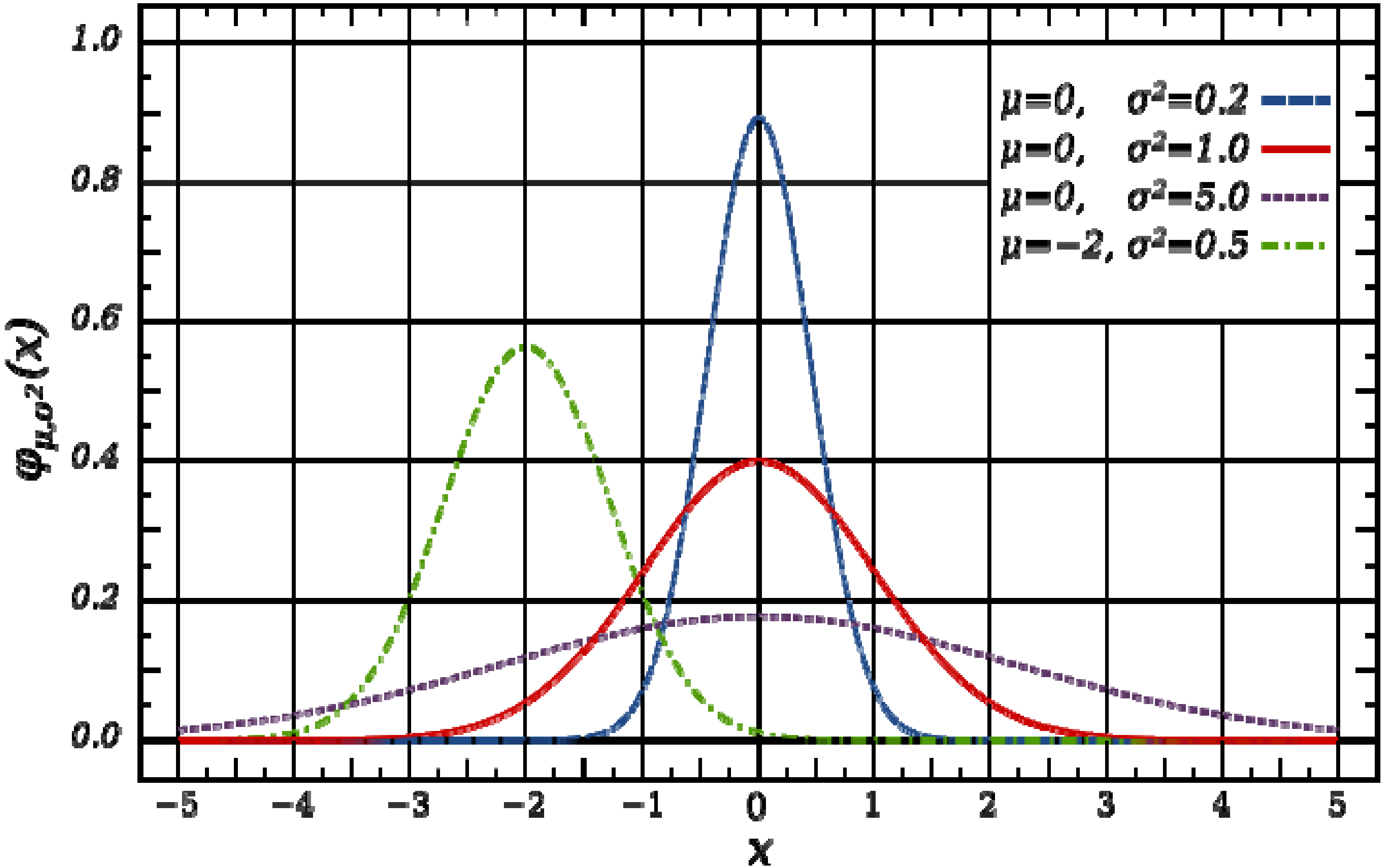
La gaussiana – grafico - 2



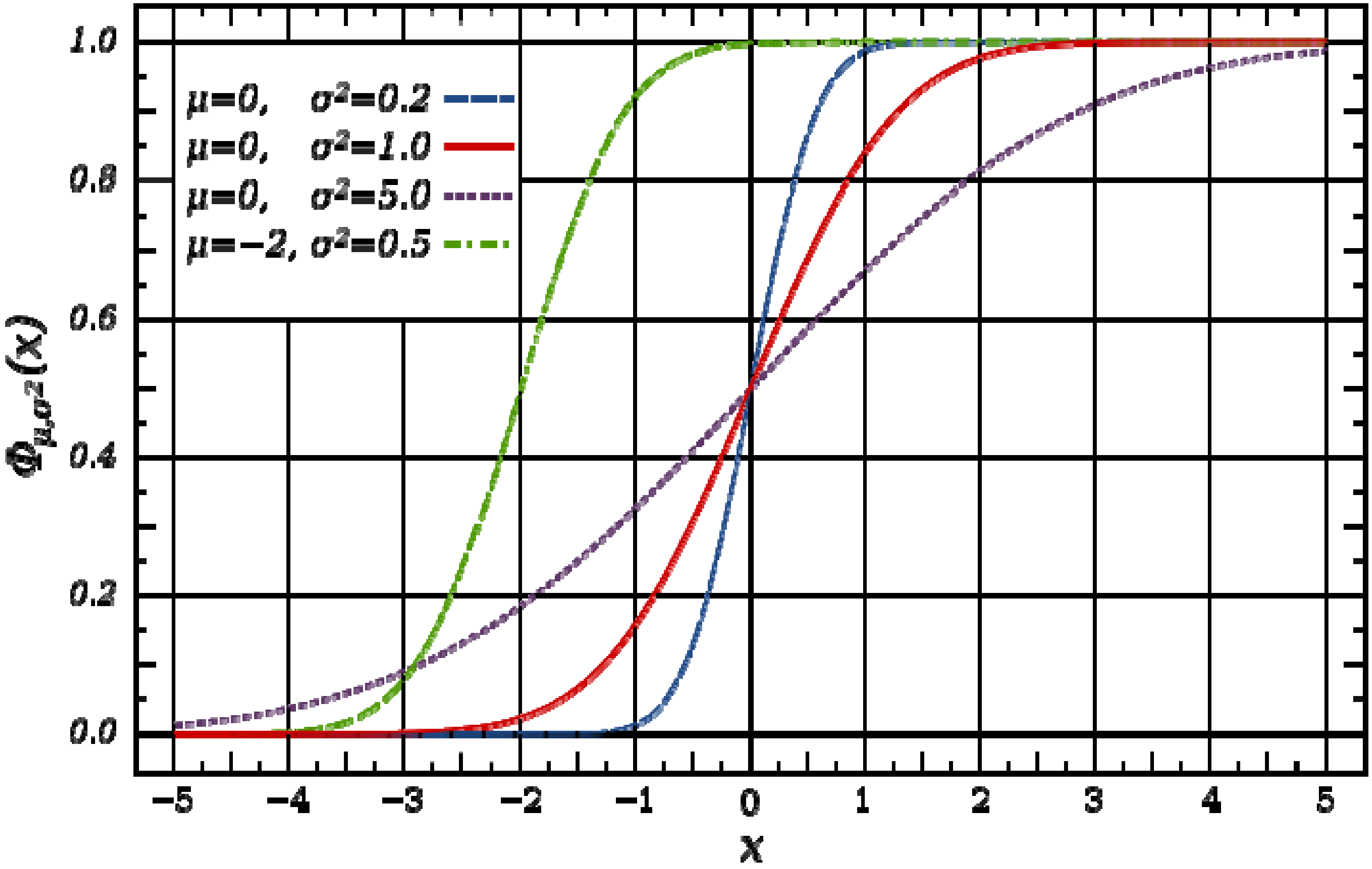
Comparazione della f_X e della corrispondente F_X

In prossimità del valor medio μ, F_X vale $1/2$

Esempi di normale – 1



Esempi di normale – 2



La gaussiana e le misure di precisione

Una misura di precisione è un fenomeno aleatorio che si comporta come previsto dalla distribuzione di Gauss.

La media μ rappresenta il valore vero, incognito, attorno al quale oscillano le misure ripetuto.

La deviazione standard quantifica la dispersione delle misure

La normale standardizzata

Fra le infinite vc normali una particolare è detta *normale standardizzata;* è la

$$\mathrm{N}\left(0,1\right)$$

indicata anche con Z

$$\mathrm{Z}=\mathrm{N}\left(0,1\right)$$

La sua forma funzionale è

$$f_Z(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

La distribuzione normale standardizzata ha una notevole utilità pratica.

Calcolo della probabilità di intervalli

Consideriamo una vc normale N=N(μ,σ) e un intervallo $[a,b]$. La probabilità che N si trovi nell'intervallo è

$$P\left(\mathrm{N}\in[a,b]\right)=\int_a^b \mathrm{d}x\, f_N(x;\mu,\sigma) \qquad (1.1)$$

Che cosa significa esattamente l'espressione "la probabilità che N si trovi nell'intervallo $[a,b]$"?

La probabilità che una estrazione da N fornisca un valore contenuto in $[a,b]$.

Che cos'è un'estrazione?

Nel caso di un dado, il lancio. Nel caso di una misura, l'effettuazione di una misura.

Calcolo della probabilità di intervalli – 2

Sviluppiamo il calcolo (1.1).

$$P(\mathrm{N} \in [a,b]) = \int_a^b \mathrm{d}x \, \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad (1.2)$$

Si può effettuare un cambio di variabile

$$u = \frac{x-\mu}{\sigma} \qquad du = \frac{dx}{\sigma}$$

ottenendo

$$P(\mathrm{N} \in [a,b]) = \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \mathrm{d}u \, \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} = P\left(\mathrm{Z} \in \left[\frac{a-\mu}{\sigma}, \frac{b-\mu}{\sigma}\right]\right)$$

Il calcolo della probabilità di un intervallo qualunque per una normale qualsiasi può essere ricondotto al calcolo della probabilità per un intervallo ausiliario e per la standardizzata.

Calcolo della probabilità di intervalli – 3

Il calcolo della probabilità di un intervallo qualunque per una normale qualsiasi può essere ricondotto al calcolo della probabilità per un intervallo ausiliario e per la standardizzata.

Si possono ridurre i calcoli di probabilità per una normale qualunque a calcoli di probabilità per la sola standardizzata.

Si tratta di un vantaggio perché non si possono calcolare gli integrali (1.2) in forma chiusa, dunque è necessario ricorrere al calcolo numerico, con appositi programmi (Matlab, ma anche Excel).

In alternativa si possono usare tabelle; in questo caso, altro è tabellare una funzione, altro è tabellarne infinite. L'uso delle tabelle era indispensabile nel passato, ma è utile ancora oggi.

Probabilità degli intervalli n-sigma

Data una normale qualunque e calcoliamo la probabilità di un intervallo avente centro nella media e semi-ampiezza pari a $n\sigma$

$$
\begin{aligned}
P\left(\mathrm{N}\in[\mu-n\sigma,\mu+n\sigma]\right) &= P\left(\mathrm{Z}\in\left[\frac{a-\mu}{\sigma},\frac{b-\mu}{\sigma}\right]\right)= \\
&= P\left(\mathrm{Z}\in\left[\frac{\mu-n\sigma-\mu}{\sigma},\frac{\mu+n\sigma-\mu}{\sigma}\right]\right)= \\
&= P\left(\mathrm{Z}\in\left[-n,n\right]\right)
\end{aligned}
$$

La probabilità degli intervalli n-sigma è indipendente della particolare vc considerata

Probabilità degli intervalli n-sigma - 2

In particolare

$$P\left(\mathrm{N}\in[\mu-\sigma,\mu+\sigma]\right)=0.6827$$

$$P\left(\mathrm{N}\in[\mu-2\sigma,\mu+2\sigma]\right)=0.9545$$

$$P\left(\mathrm{N}\in[\mu-3\sigma,\mu+3\sigma]\right)=0.9973$$

$$P\left(\mathrm{N}\in[\mu-4\sigma,\mu+4\sigma]\right)=0.9999$$

Per il calcolo con Matlab della probabilità dell'intervallo 2-sigma, per esempio

```
P=normcdf(2,0,1)-normcdf(-2,0,1)
```