



# Vittorio Casella

Laboratorio di Geomatica - DICAR - Università di Pavia

email: [vittorio.casella@unipv.it](mailto:vittorio.casella@unipv.it)



## La variabile casuale di Gauss

# License/Licenza



This document is © 2013 **Vittorio Casella, University of Pavia, vittorio.casella@unipt.it**, available under the **creative commons 3.0 license**.

You are free:

to Share — to copy, distribute and transmit the work

to Remix — to adapt the work

to make commercial use of the work.

Under the following conditions:

**Attribution** — You must attribute the work in the manner specified by the author (see the red text above) or licensor (but not in any way that suggests that they endorse you or your use of the work).

**Share Alike** — If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar license to this one.

See <http://creativecommons.org/licenses/by-sa/3.0/> for details.



Questo documento è ©2013 **Vittorio Casella, Università di Pavia, vittorio.casella@unipv.it**, disponibile sotto la **licenza creative commons 3.0**.

Tu sei libero:

di riprodurre, distribuire, comunicare al pubblico, esporre in pubblico, rappresentare, eseguire e recitare quest'opera

di modificare quest'opera

di usare quest'opera per fini commerciali

Alle seguenti condizioni:

**Attribuzione** — Devi attribuire la paternità dell'opera nei modi indicati dall'autore (vedo testo in rosso sopra) o da chi ti ha dato l'opera in licenza e in modo tale da non suggerire che essi avallino te o il modo in cui tu usi l'opera.

**Condividi allo stesso modo** — Se alteri o trasformi quest'opera, o se la usi per crearne un'altra, puoi distribuire l'opera risultante solo con una licenza identica o equivalente a questa.

Per dettagli: <http://creativecommons.org/licenses/by-sa/3.0/deed.it>



## 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](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  $\mu$  e  $\sigma$  è indicata anche con

$$N(\mu, \sigma)$$

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

# Terminologia

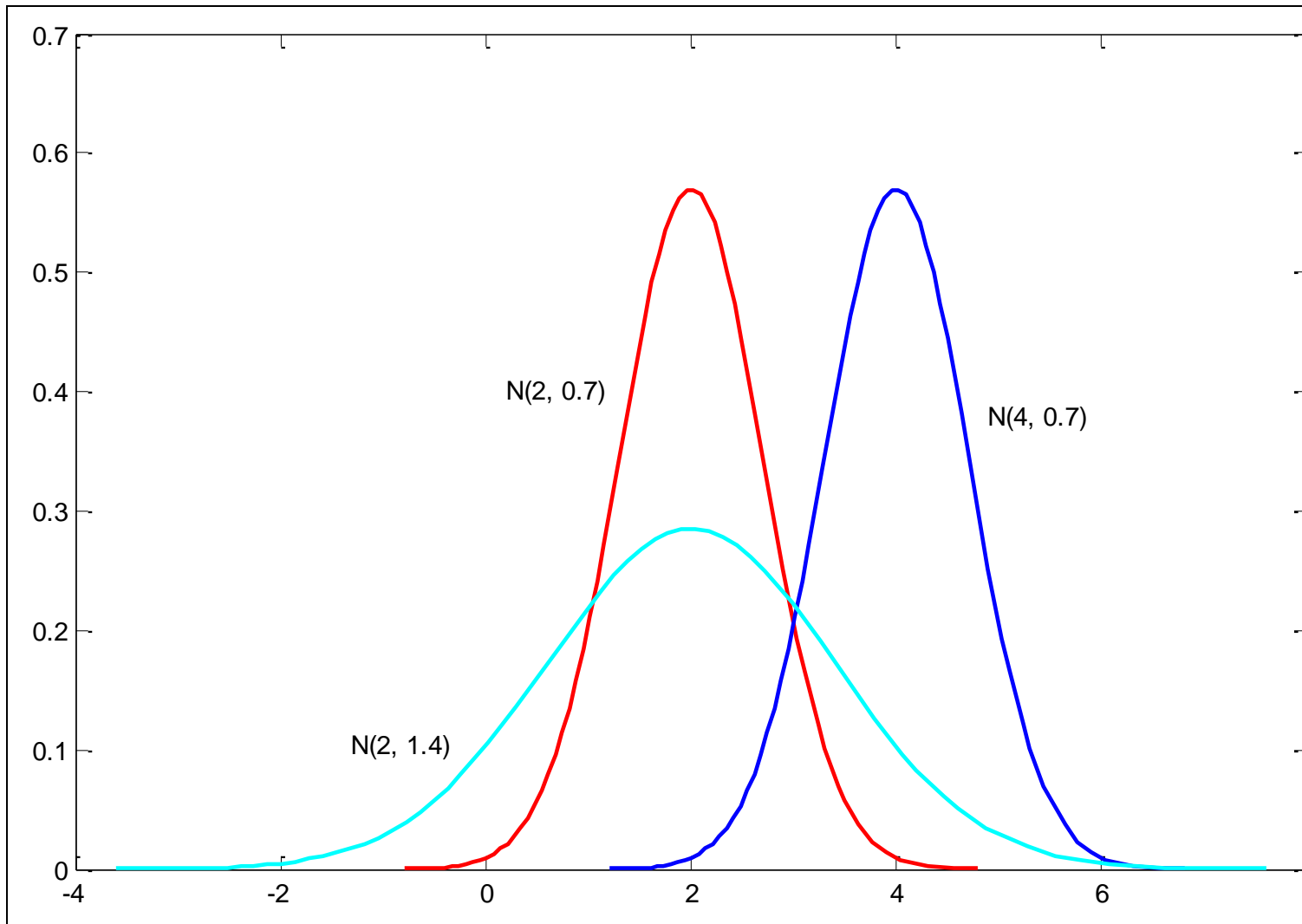
---

$\mu$  - media

$\sigma$  - deviazione standard

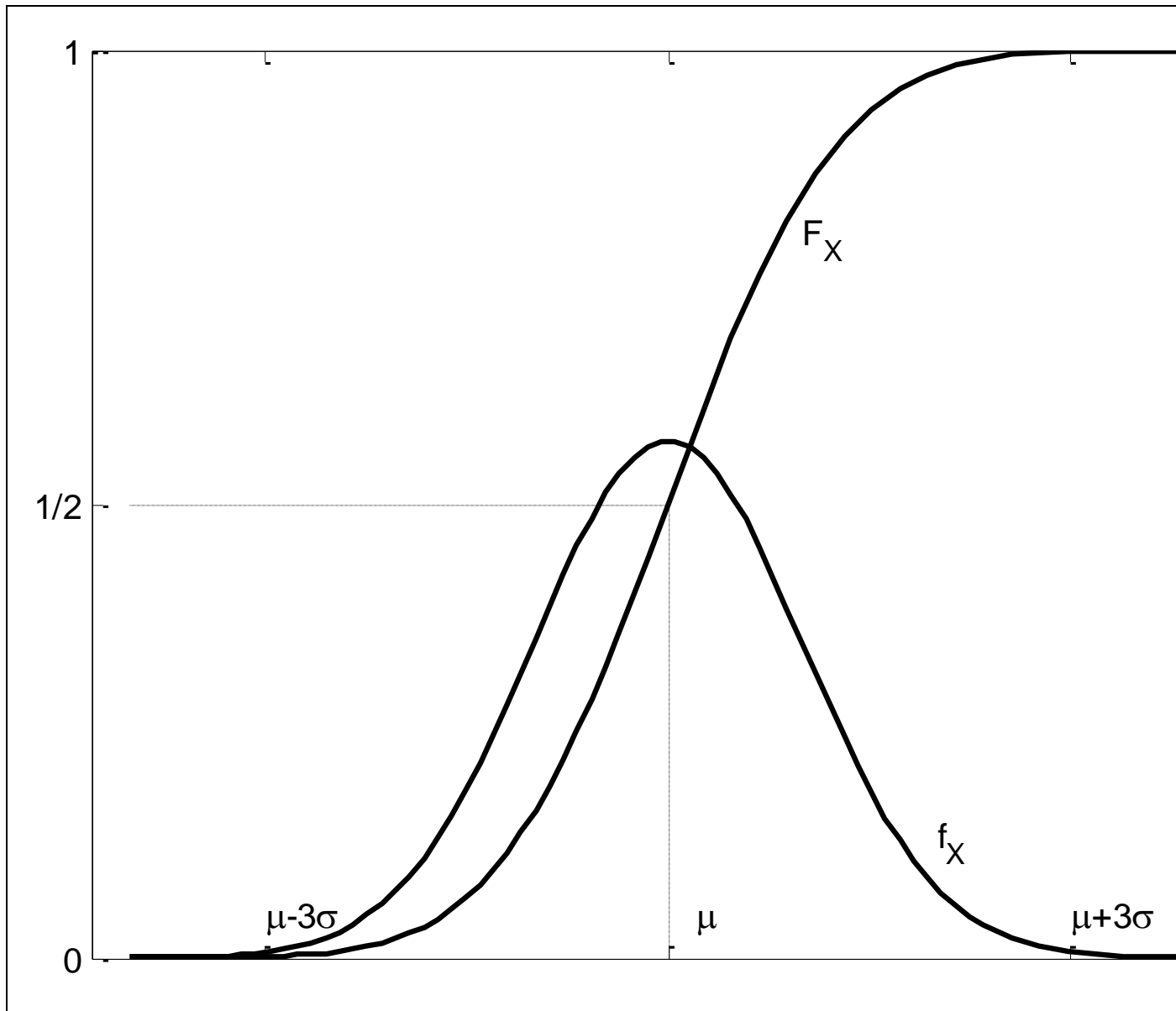
$\sigma^2$  - varianza

# La gaussiana – grafico



- Tipica forma a campana
- Il picco si trova in corrispondenza della media  $\mu$
- Il parametro  $\sigma$  controlla la larghezza della campana

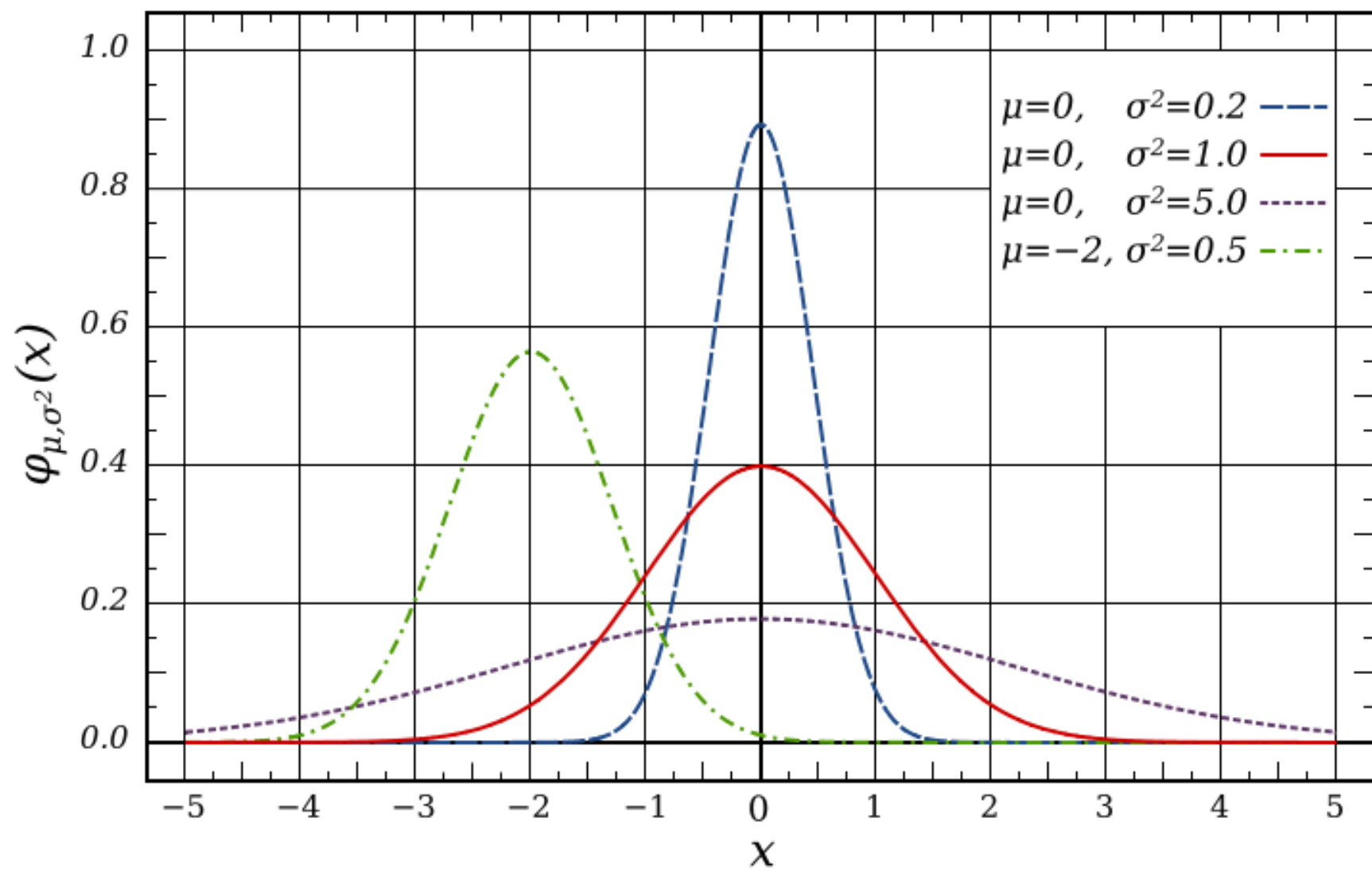
## La gaussiana – grafico - 2



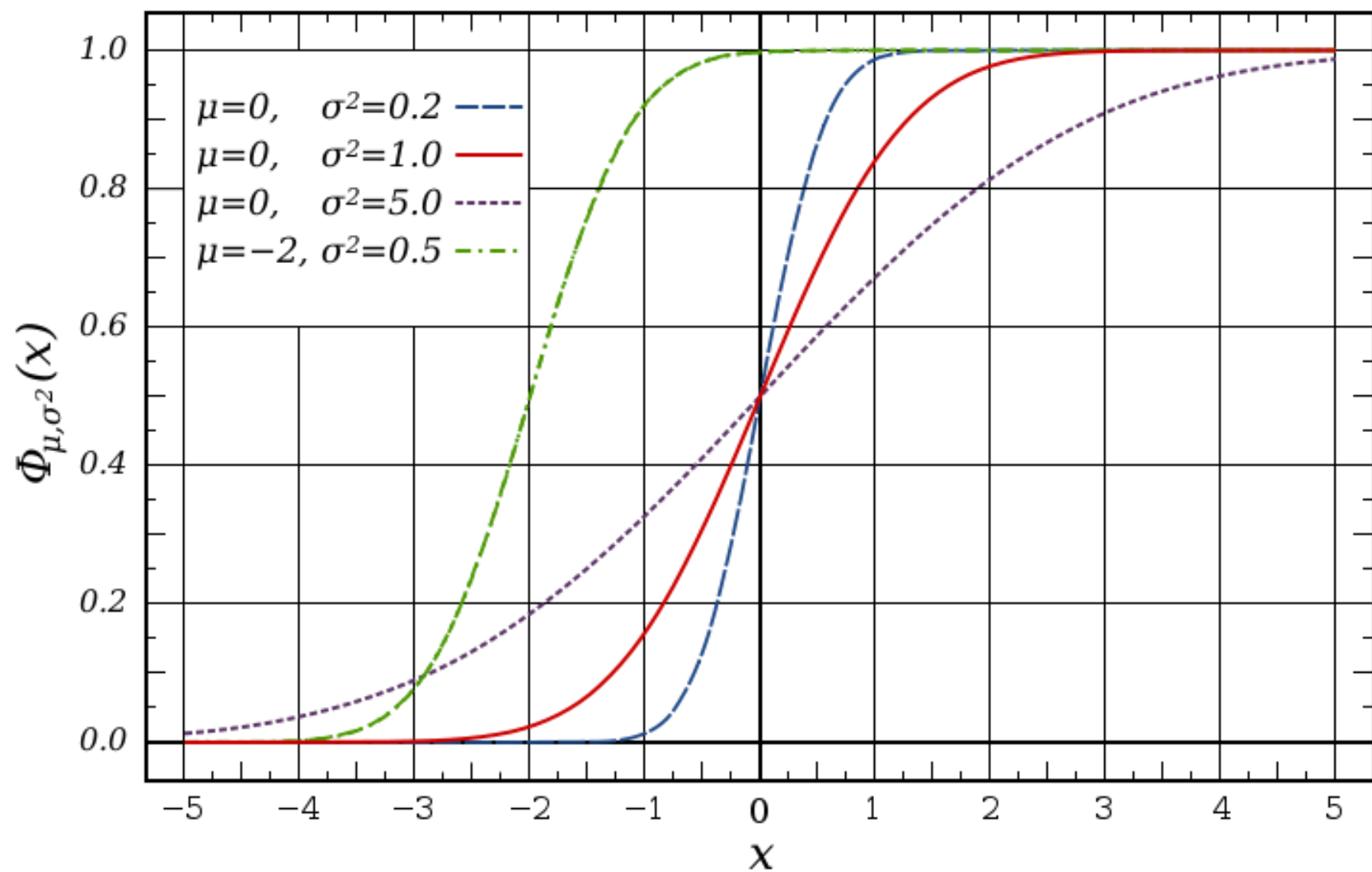
Comparazione della  $f_x$  e  
della corrispondente  $F_x$   
In prossimità del valor  
medio  $\mu$ ,  $F_x$  vale  $1/2$



## 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  $\mu$  rappresenta il valore vero, incognito, attorno al quale oscillano le misure ripetute.

La deviazione standard quantifica la dispersione delle misure

## La normale standardizzata

---

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

$$N(0,1)$$

indicata anche con Z

$$Z=N(0,1)$$

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  $N=N(\mu, \sigma)$  e un intervallo  $[a, b]$ . La probabilità che  $N$  si trovi nell'intervallo è

$$P(N \in [a, b]) = \int_a^b dx f_N(x; \mu, \sigma) \quad (1.1)$$

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).

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

Si può effettuare un cambio di variabile

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

ottenendo

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

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  $n\sigma$

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

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



## Probabilità degli intervalli n-sigma - 2

---

In particolare

$$P(N \in [\mu - \sigma, \mu + \sigma]) = 0.6827$$

$$P(N \in [\mu - 2\sigma, \mu + 2\sigma]) = 0.9545$$

$$P(N \in [\mu - 3\sigma, \mu + 3\sigma]) = 0.9973$$

$$P(N \in [\mu - 4\sigma, \mu + 4\sigma]) = 0.9999$$

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

$$P = \text{normcdf}(2, \theta, 1) - \text{normcdf}(-2, \theta, 1)$$

## DA FARE

---

Sigma come larghezza a metà altezza

Significato di 100 m e sigma di 1 cm in termini di probabilità frequenzistica:  
effettuando 100 ripetizioni, 95 cadono nell'intervallo 2 sigma

Rapporto fra misura vera e misura stimata

$$|\bar{x} - \tilde{x}| \leq 3\sigma \quad P = 0.997$$

Grafico con curva e valore vero su asse x: dove cadono le misure ripetute?