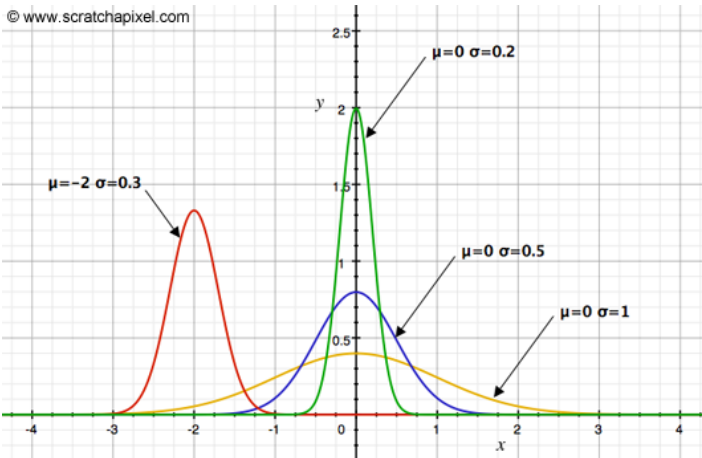


## VARIABILI ALEATORIE NORMALI

Sono le variabili aleatorie continue più importanti: si chiamano normali o gaussiane perchè la densità di probabilità è data dalla funzione di Gauss.

$$\rho(\mathbf{x}) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\mathbf{x}-\mu)^2}{2\sigma^2}}$$

Tale funzione è determinata dai 2 parametri, valore atteso  $\mu = \mathbb{E}(\mathbf{X})$  e deviazione standard  $\sigma$ . La funzione ha un grafico a forma di campana simmetrico intorno al valor medio  $\mu$ . Per questo motivo una variabile gaussiana (normale) con valore atteso  $\mu$  e deviazione standard  $\sigma$  viene indicata con  $\mathbf{N}(\mu, \sigma)$ .



**La distribuzione normale** è completamente individuata dal **valore atteso**  $\mu$  e dello **scarto quadratico**  $\sigma$ . Il valore centrale coincide con  $\mu$ . La **campana** si allarga al crescere di  $\sigma$ .

## QUANDO SI USA IL MODELLO NORMALE

Esempi di situazioni in cui si usa:

- **Misure affette da errore.** La misura di una qualunque grandezza fisica (temperatura, concentrazione di un farmaco,..) può essere vista come la somma del valore esatto della grandezza e dell'errore strumentale. Il fatto che la normale rappresenti bene queste misure fu già messo in evidenza da Gauss.
- **Distribuzione di una caratteristica quantitativa di una popolazione che presenta oscillazioni casuali intorno alla media.** Molte grandezze (statura, peso, ...) relative agli individui di una popolazione omogenea sono ben rappresentate da una variabile aleatoria normale.

**IN GENERALE QUANDO L'ALEATORIETÀ DI UNA MISURA O DI UN FENOMENO È DOVUTA ALLA SOMMA DI DIVERSE CAUSE INDIPENDENTI**

## FUNZIONE DISTRIBUZIONE E PROBABILITÀ

La funzione distribuzione è

$$F(x) = \int_{-\infty}^x \rho(x) dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

che è una funzione strettamente crescente con  $\lim_{x \rightarrow -\infty} F(x) = 0$   
e  $\lim_{x \rightarrow +\infty} F(x) = 1$ .

La probabilità che la variabile **X** sia nell'intervallo **[a, b]** è

$$P(a \leq X \leq b) = \frac{1}{\sqrt{2\pi}\sigma} \int_a^b e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = F(b) - F(a)$$

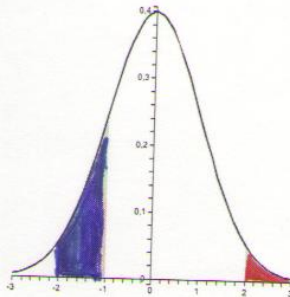
In particolare

$$P(X \leq b) = P(-\infty < X \leq b) = F(b)$$

$$P(X \geq a) = P(a \leq X < \infty) = 1 - F(a)$$

## La probabilità

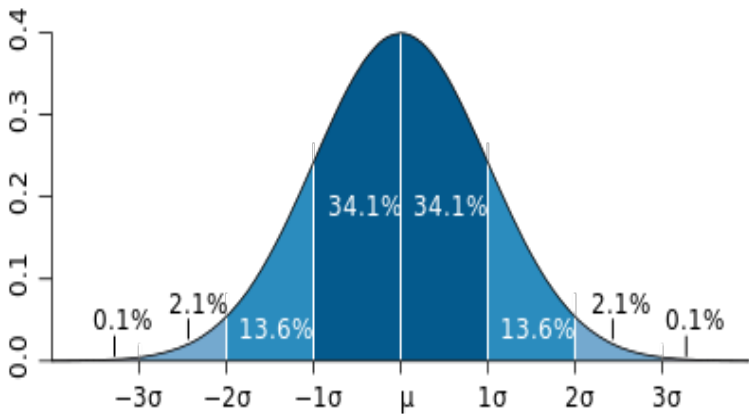
Dobbiamo pensare sempre alle probabilità come aree: ad esempio



in azzurro  $\mathbb{P}(-2 \leq X \leq -1)$ , in rosa  $\mathbb{P}(X > 2)$ .

## È importante tenere presente che

- L'area compresa tra l'asse delle ascisse, il grafico della densità di probabilità e le rette  $x = \mu - \sigma$  e  $x = \mu + \sigma$  è pari al **68.3%** dell'area totale (che è 1).  $P(\mu - \sigma < X < \mu + \sigma) = 0.683$
- L'area compresa tra l'asse delle ascisse, il grafico della densità di probabilità e le rette  $x = \mu - 2\sigma$  e  $x = \mu + 2\sigma$  è pari al **95.4%** dell'area totale (che è 1).  $P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.954$
- L'area compresa tra l'asse delle ascisse, il grafico della densità di probabilità e le rette  $x = \mu - 3\sigma$  e  $x = \mu + 3\sigma$  è pari al **99.7%** dell'area totale (che è 1).  $P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.997$  (si discosta dalla media più di  $3\sigma$  con probabilità bassissima = 0.3 %).



Rappresentazione grafica in termini di aree da cui si vede (al netto delle approssimazioni sui decimali) che

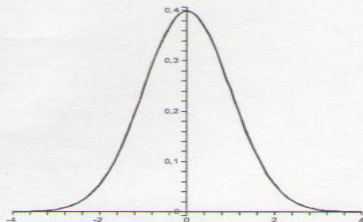
$$P(\mu - \sigma < X < \mu + \sigma) = 0.341 + 0.341 = 0.683 \text{ e}$$

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.341 + 0.341 + 0.136 + 0.136 = 0.954.$$

Una variabile aleatoria normale con **media**  $\mu = 0$  e **scarto**  $\sigma = 1$  si dice **variabile normale standard**. Per abbreviare si scrive  $N(0,1)$ .

La densità di probabilità è  $\rho(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ .

### Proprietà della densità di $\mathcal{N}(0, 1)$



- Il massimo è in  $x = 0$  ( $\max = \frac{1}{\sqrt{2\pi}} \approx 0.3989$ ).
- È simmetrica rispetto all'asse delle  $y$ .
- Va a zero a  $\pm\infty$ .
- $\mathbb{P}(|X| > 3) = 0.0027$ .
- $\mathbb{P}(|X| > 4) = 0.000063$ .
- $\mathbb{P}(-3 < X < 3) = 0.9973$ .

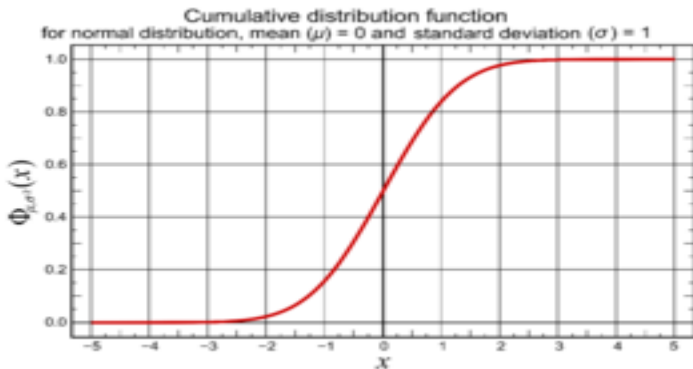
Di fatto per le applicazioni si considera che la probabilità di assumere valori  $> 3$  o  $< -3$  sia trascurabile.



**Una buona notizia!** per calcolare le probabilità non dobbiamo calcolare le aree (e quindi gli integrali) perché ci sono dei programmi disponibili per ogni tipo di linguaggio (possiamo facilmente fare noi stessi un programma).

I **programmi** riguardano la **distribuzione normale standard**, cioè quella con media 0 e varianza 1. A volte indicheremo tale variabile con  **$N(0, 1)$**  ed altre volte semplicemente con  **$Z$** . Vedremo poi come usare con un "trucco" questi programmi anche per le altre variabili gaussiane.

Indicheremo con  **$\phi(x)$**  la funzione di ripartizione della variabile normale standard  **$Z=N(0,1)$** , ossia  **$\phi(x)=P(Z \leq x)$**  mentre continueremo ad indicare con  **$F(x)=P(X \leq x)$**  la funzione di ripartizione di una generica variabile normale  **$X=N(\mu, \sigma)$** .



Questa è la funzione distribuzione di una variabile gaussiana standard:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{x^2}{2}} dx$$

che è quella comunemente calcolata da software universalmente disponibile.

## CONVERSIONE IN NORMALE STANDARD

Sia  $X$  una variabile aleatoria normale con valore atteso  $\mathbb{E}(X) = \mu$  e varianza  $\text{var}(X) = \sigma^2$ . Definiamo la variabile aleatoria:

$$Z = \frac{X - \mu}{\sigma}$$

Si osservi che per le proprietà del valore atteso

$$\mathbb{E}(Z) = \frac{1}{\sigma} \mathbb{E}(X) - \frac{\mu}{\sigma} = \frac{1}{\sigma} \mu - \frac{\mu}{\sigma} = 0$$

e per le proprietà della varianza

$$\text{var}(Z) = \frac{1}{\sigma^2} \text{var}(X) = 1$$

Inoltre la forma della densità rimane a campana e simmetrica intorno al valore atteso, quindi  $Z$  è una normale standard.

Da questa constatazione si ottiene l'importante risultato

$$\begin{aligned} P(\mathbf{a} \leq \mathbf{X} \leq \mathbf{b}) &= P\left(\frac{\mathbf{a} - \mu}{\sigma} \leq \frac{\mathbf{X} - \mu}{\sigma} \leq \frac{\mathbf{b} - \mu}{\sigma}\right) = \\ &= P\left(\frac{\mathbf{a} - \mu}{\sigma} \leq \mathbf{Z} \leq \frac{\mathbf{b} - \mu}{\sigma}\right) = \phi\left(\frac{\mathbf{b} - \mu}{\sigma}\right) - \phi\left(\frac{\mathbf{a} - \mu}{\sigma}\right) \end{aligned}$$

in particolare ponendo  $a = -\infty$  oppure  $a = +\infty$  si ottiene

$$P(\mathbf{X} < \mathbf{b}) = \phi\left(\frac{\mathbf{b} - \mu}{\sigma}\right) \qquad P(\mathbf{X} > \mathbf{a}) = 1 - \phi\left(\frac{\mathbf{a} - \mu}{\sigma}\right)$$

In conclusione possiamo utilizzare i programmi della normale standard (che calcolano  $\phi(x)$ ) per calcolare le probabilità di una normale generica.

**Esercizio.** Sia  $X$  una variabile aleatoria normale con valore atteso  $\mu = 100$  e deviazione standard  $\sigma = 14.2$ . Calcolare  $P(X > 130)$  e  $P(90 < X < 115)$ .

La variabile  $Z = \frac{X - 100}{14.2}$  ha distribuzione normale standard per cui

$$\begin{aligned} P(X > 130) &= P\left(Z > \frac{130 - 100}{14.2}\right) = P(Z > 2.1127) \\ &= 1 - \phi(2.1127) = 1 - 0.983 = 0.017 \end{aligned}$$

La disuguaglianza  $90 < X < 115$  equivale a

$$\frac{90 - 100}{14.2} < \frac{X - 100}{14.2} < \frac{115 - 100}{14.2}$$

ovvero

$$-0,7042 < Z < 1.056$$

Quindi

$$\begin{aligned} P(90 < X < 115) &= P(-0,7042 < Z < 1.056) = \\ &= \phi(1.056) - \phi(-0,7042) = 0.854 - 0.242 = 0.612 \end{aligned}$$