

## EXPECTED VALUE

**Definition.** The **expected value** (or mean or average ) of the continuous random variable  $X$ , indicated by  $\mathbb{E}(X)$  è is the number

$$\mu = \mathbb{E}(\mathbf{X}) = \int_{-\infty}^{\infty} \mathbf{x}\rho(\mathbf{x})d\mathbf{x}$$

### Example 1

Consider the probability density uniformly equal to 1 in the interval  $[0, 1]$  and vanishing elsewhere (we have already seen that it is correctly normalized to 1), we have

$$\mathbb{E}(\mathbf{X}) = \int_{-\infty}^{\infty} \rho(\mathbf{x})\mathbf{x}d\mathbf{x} = \int_0^1 \mathbf{x}d\mathbf{x} = \left[ \frac{\mathbf{x}^2}{2} \right]_0^1 = \frac{1}{2}.$$

## Example 2

Consider a probability density which equals  $\frac{x}{2}$  in the interval  $[0, 2]$  and vanishes elsewhere We preliminary verify the normalization,:

$$\int_{-\infty}^{\infty} \rho(\mathbf{x}) d\mathbf{x} = \frac{1}{2} \int_0^2 \mathbf{x} d\mathbf{x} = \frac{1}{2} \left[ \frac{\mathbf{x}^2}{2} \right]_0^2 = 1,$$

then the expected value is

$$\mathbb{E}(\mathbf{X}) = \int_{-\infty}^{\infty} \rho(\mathbf{x}) \mathbf{x} d\mathbf{x} = \frac{1}{2} \int_0^2 \mathbf{x}^2 d\mathbf{x} = \left[ \frac{\mathbf{x}^3}{6} \right]_0^2 = \frac{4}{3}.$$

The properties of expected value are the same of the sample mean,  
in fact, if  $a$  and  $b$  are constant,

$$\begin{aligned} E(\mathbf{aX} + \mathbf{b}) &= \int_{-\infty}^{\infty} \rho(\mathbf{x})(\mathbf{ax} + \mathbf{b})d\mathbf{x} = \\ &\mathbf{a} \int_{-\infty}^{\infty} \rho(\mathbf{x})\mathbf{x}d\mathbf{x} + \mathbf{b} \int_{-\infty}^{\infty} \rho(\mathbf{x})d\mathbf{x} = \mathbf{aE(X)} + \mathbf{b} \end{aligned}$$

Moreover, for two variables  $X$  and  $Y$  (not necessarily independent)  
one always has

$$E(\mathbf{X} + \mathbf{Y}) = E(\mathbf{X}) + E(\mathbf{Y})$$

## VALORE ATTESO DI UNA FUNZIONE DI X

Sia  $g$  una funzione, allora  $g(X)$  è una variabile aleatoria ed il suo valore atteso è dato da:

$$\mathbb{E}(g(\mathbf{X})) = \int_{-\infty}^{\infty} g(\mathbf{x})\rho(\mathbf{x})d\mathbf{x}$$

**Esempio:** si consideri di nuovo la densità di probabilità dell'esercizio (uniformemente uguale a 1 nell'intervallo  $[0, 1]$  e nulla al di fuori di esso) e sia  $g(x) = x^2$ .

$$\mathbb{E}(g(\mathbf{X})) = \mathbb{E}(\mathbf{X}^2) = \int_{-\infty}^{\infty} \rho(\mathbf{x})g(\mathbf{x})d\mathbf{x} = \int_0^1 x^2 d\mathbf{x} = \left[ \frac{x^3}{3} \right]_0^1 = \frac{1}{3}$$

## VARIANZA E DEVIAZIONE STANDARD

**Definizione.** Sia  $X$  una variabile aleatoria con valore atteso  $\mu = \mathbb{E}(X)$ , la **varianza di  $X$**  denotata con  $\text{var}(X)$  è

$$\text{var}(\mathbf{X}) = \mathbb{E}\left((\mathbf{X} - \mu)^2\right) = \int_{-\infty}^{\infty} (\mathbf{x} - \mu)^2 \rho(\mathbf{x}) d\mathbf{x}$$

Notate che la varianza è sempre non negativa ed è uguale a 0 solo se la variabile è certa (può assumere un solo valore).

Lo scarto quadratico medio o deviazione standard è invece:

$$\sigma = \sqrt{\text{var}(\mathbf{X})}.$$

Notare che le tutte le definizioni sono le stesse delle variabile aleatorie discrete, con le somme sostituite dagli integrali.

## FORMULA UTILE

Sviluppando il quadrato della varianza  $\text{var}(\mathbf{X}) = \mathbb{E}\left((\mathbf{X} - \mu)^2\right)$  si ha

$$\begin{aligned}\text{var}(\mathbf{X}) &= \mathbb{E}\left(X^2 - 2\mu X + \mu^2\right) = \mathbb{E}(X^2) - 2\mu\mathbb{E}(X) + \mu^2 \\ &= \mathbb{E}(X^2) - 2\mu^2 + \mu^2 = \mathbb{E}(\mathbf{X}^2) - \mu^2\end{aligned}$$

Dunque

$$\text{var}(\mathbf{X}) = \mathbb{E}(\mathbf{X}^2) - \mathbf{E}(\mathbf{X})^2 = \int_{-\infty}^{\infty} x^2 \rho(x) dx - \left( \int_{-\infty}^{\infty} x \rho(x) dx \right)^2$$

Consider also the following property, which is the same as that of the sample variance (in descriptive statistics) and of discrete random variables:

$$\mathbf{var}(\mathbf{aX} + \mathbf{c}) = \mathbf{a}^2 \mathbf{var}(\mathbf{X}),$$

in fact,

$$\begin{aligned} \mathbf{var}(\mathbf{aX} + \mathbf{c}) &= \mathbb{E}((\mathbf{aX} + \mathbf{b})^2) - (\mathbb{E}(\mathbf{aX} + \mathbf{b}))^2 = \\ &\mathbf{a}^2 \int_{-\infty}^{\infty} \mathbf{x}^2 \rho(\mathbf{x}) \mathbf{d}\mathbf{x} - 2\mathbf{a}\mathbf{b} \int_{-\infty}^{\infty} \mathbf{x} \rho(\mathbf{x}) \mathbf{d}\mathbf{x} + \mathbf{b}^2 - \left( \mathbf{a} \int_{-\infty}^{\infty} \mathbf{x} \rho(\mathbf{x}) \mathbf{d}\mathbf{x} + \mathbf{b} \right)^2 \\ &= \mathbf{a}^2 \left[ \int_{-\infty}^{\infty} \mathbf{x}^2 \rho(\mathbf{x}) \mathbf{d}\mathbf{x} - \left( \int_{-\infty}^{\infty} \mathbf{x} \rho(\mathbf{x}) \mathbf{d}\mathbf{x} \right)^2 \right] = \mathbf{a}^2 \mathbf{var}(\mathbf{X}) \end{aligned}$$

Abbiamo visto che se  $X$  ed  $Y$  sono due variabili aleatorie non necessariamente indipendenti allora,

- $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y).$

Inoltre si ha che se  $X$  ed  $Y$  sono due variabile aleatorie **indipendenti** allora,

- $\text{var}(X+Y) = \text{var}(X) + \text{var}(Y).$

Si tenga presente che per definizione due variabili aleatorie  $X$  ed  $Y$  si dicono **indipendenti** se gli eventi  $\{X = x\}$  e  $\{Y = y\}$  sono indipendenti per ogni  $x$ , valore di  $X$  ed  $y$ , valore di  $Y$ .

**Esercizio:** si consideri la densità di probabilità uniformemente uguale a 1 nell'intervallo  $[0, 1]$  e nulla al di fuori di esso e sia  $g(x) = x^2$ . Calcolare varianza e deviazione standard. Abbiamo già calcolato

$$\mu = \mathbb{E}(\mathbf{X}) = \frac{1}{2}, \quad \mathbb{E}(\mathbf{X}^2) = \frac{1}{3}.$$

La varianza, usando la "formula utile" si ottiene nel seguente modo

$$\text{var}(\mathbf{X}) = \mathbb{E}(\mathbf{X}^2) - \mu^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$$

Infine lo scarto quadratico medio o deviazione standard è :

$$\sigma = \sqrt{\text{var}(\mathbf{X})} = \frac{1}{\sqrt{12}} \simeq 0,3.$$

Nelle scienze a volte si usa la notazione  $X = \mu \pm \sigma$ , nel nostro caso  $X = 0,5 \pm 0,3$ .

## 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  $\mathbf{N}(\mu, \sigma)$ .

It can be shown (it is an exercise) that:

$$\int_{-\infty}^{\infty} \rho(x) dx = 1 \quad \text{it is a probability density!}$$

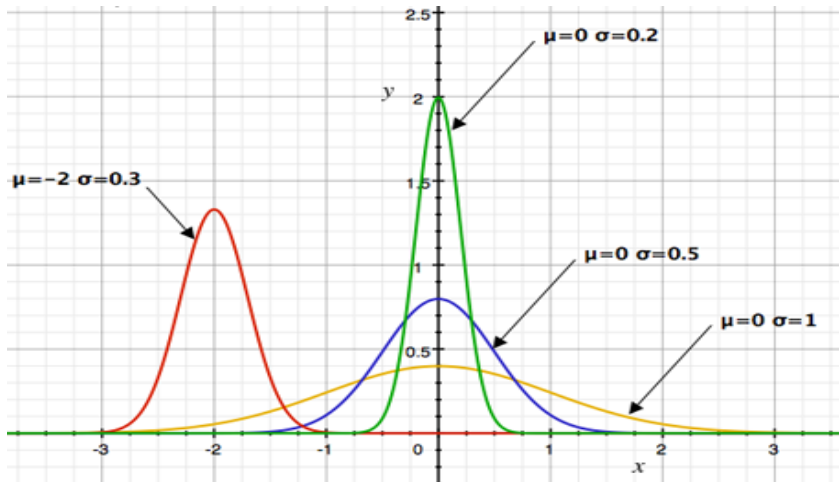
$$\mathbb{E}(X) = \int_{-\infty}^{\infty} \rho(x) x dx = \mu$$

which also indicates the central value,

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} \rho(x) x^2 dx = \mu^2 + \sigma^2$$

$$\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \sigma^2,$$

so that the standard deviation is  $\sigma$  which is a measure of the width of the gaussian.



**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  $\lim_{x \rightarrow +\infty} F(x) = 1$ .

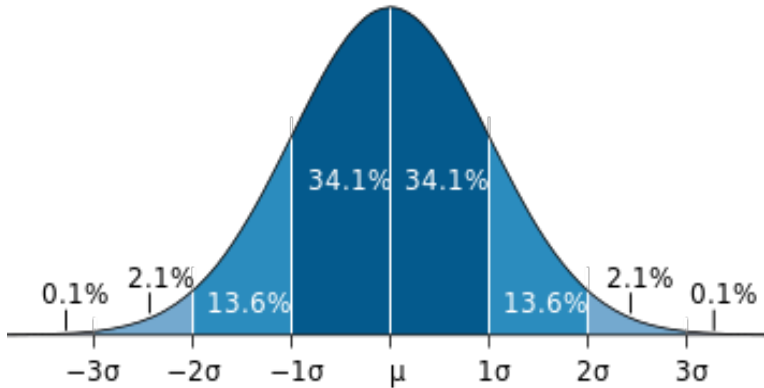
La probabilità che la variabile **X** sia nell'intervallo **[a, b]** è

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

In particolare

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

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



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

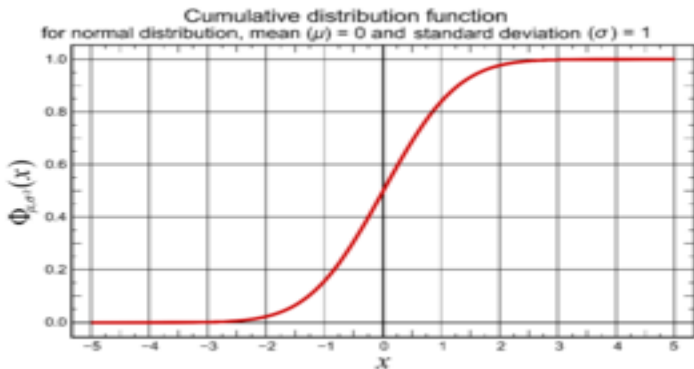
$$P(\mu - \sigma < X < \mu + \sigma) = 0.341 + 0.341 = 0.683 \text{ e}$$

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

**Una 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 (potete facilmente fare voi 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 distribuzione 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:

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

Si osservi che per le proprietà del valore atteso

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

e per le proprietà della varianza

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

Inoltre, si può dimostrare che  $\mathbf{Z} = \frac{\mathbf{X} - \mu}{\sigma}$  direttamente eseguendo la trasformazione della densità di probabilità.

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

$$P(90 < X < 115) = P(-0,7042 < Z < 1.056) =$$

$$= \Phi(1.056) - \Phi(-0,7042) = 0.854 - 0.242 = 0.612$$

Tutto quindi molto semplice se si ha un programma che calcola la funzione di distribuzione della normale standard.