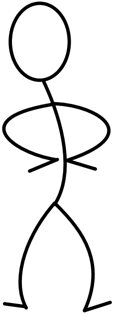
Distribuzioni continue



Definizione

Sia $X : \Omega \to \mathbb{R}$ una variabile aleatoria.

Diremo che X ha **distribuzione continua** se $P(X{=}r) = 0$ per ogni r.

> Per esempio, scegliendo un numero reale con probabilità uniforme tra 2 e 4, la probabilità di indovinare π oppure e, con tutte le loro cifre decimali, è 0. (Anche se non si tratta dell'evento vuoto.)

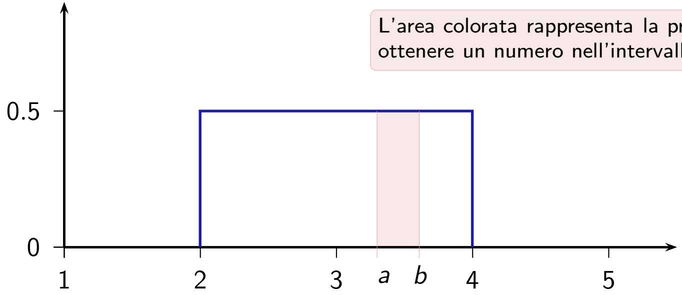
Le distribuzioni continue si rappresentano con una **densità di probabilità**

Distribuzione uniforme

Le distribuzioni continue si rappresentano con una **densità di probabilità**

Per esempio la distribuzione uniforme nell'intervallo $[2,4]$ si rappresenta con la funzione

$$f(x) = \begin{cases} 0 & \text{se } x < 2 \\ 0.5 & \text{se } 2 \leq x \leq 4 \\ 0 & \text{se } 4 < x \end{cases}$$



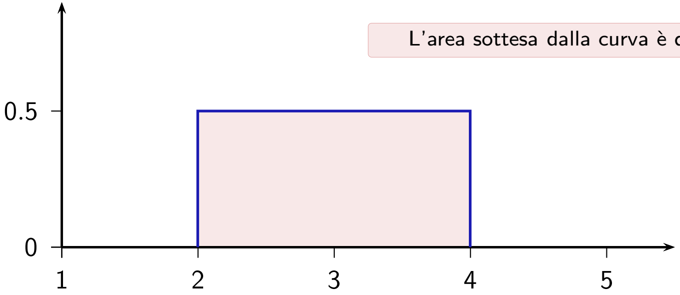
L'area colorata rappresenta la probabilità di ottenere un numero nell'intervallo $[a, b]$.

Distribuzione uniforme

Le distribuzioni continue si rappresentano con una **densità di probabilità**

Per esempio la distribuzione uniforme nell'intervallo $[2, 4]$ si rappresenta con la funzione

$$f(x) = \begin{cases} 0 & \text{se } x < 2 \\ 0.5 & \text{se } 2 \leq x \leq 4 \\ 0 & \text{se } 4 < x \end{cases} \qquad \int_{-\infty}^{+\infty} f(x)dx = 1$$



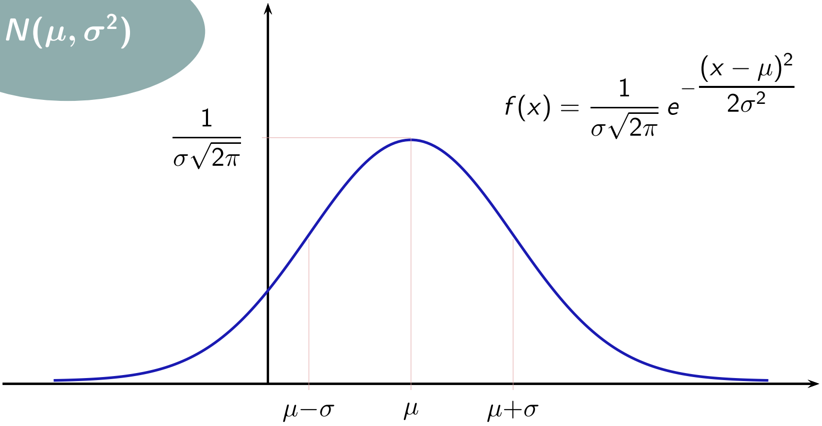
L'area sottesa dalla curva è quindi 1.

Distribuzione normale

La distribuzione più frequentemente usata è la **distribuzione normale** detta anche **di Gauss**

$$N(\mu, \sigma^2)$$

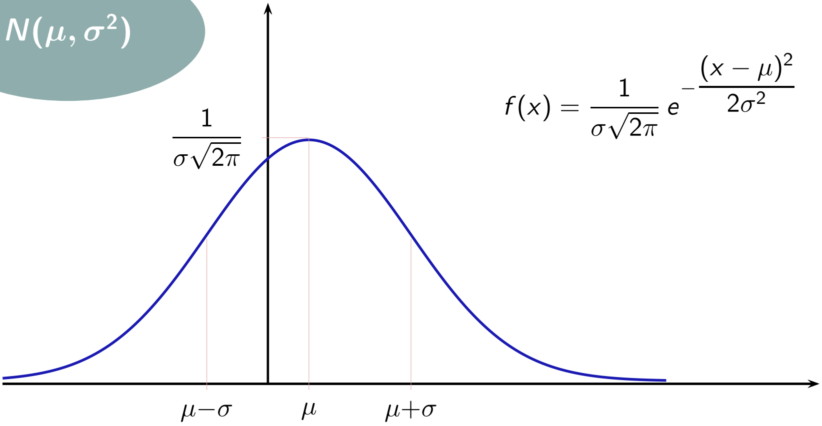
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



La distribuzione più frequentemente usata è la **distribuzione normale** detta anche **di Gauss**

$N(\mu, \sigma^2)$

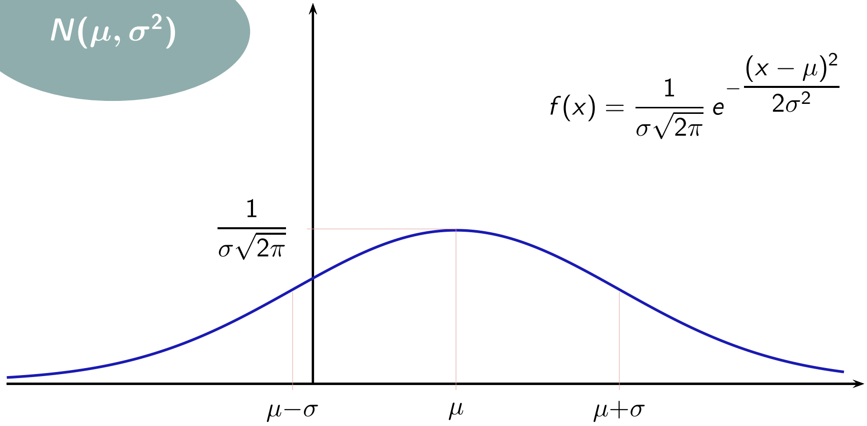
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}$$



La distribuzione più frequentemente usata è la **distribuzione normale** detta anche **di Gauss**

$N(\mu, \sigma^2)$

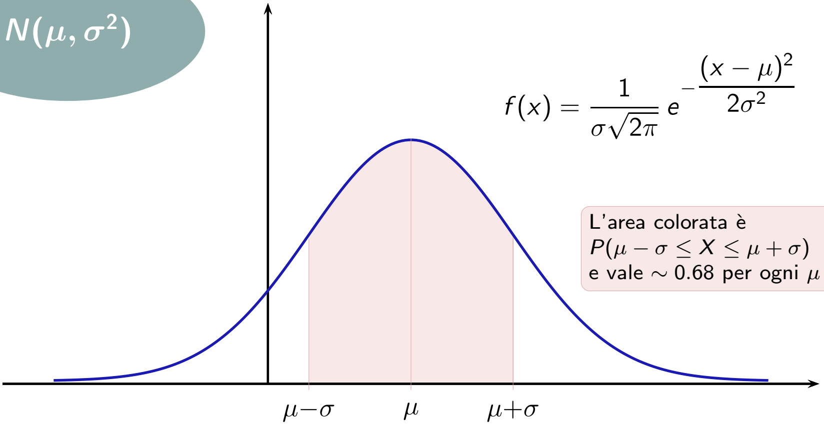
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}$$



La distribuzione più frequentemente usata è la **distribuzione normale** detta anche **di Gauss**

$$N(\mu, \sigma^2)$$

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



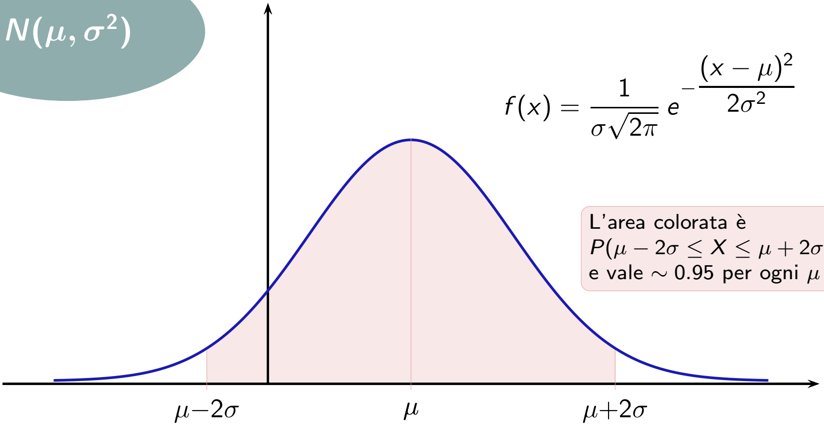
L'area colorata è
$P(\mu - \sigma \leq X \leq \mu + \sigma)$
e vale ~ 0.68 per ogni μ e σ

Distribuzione normale

La distribuzione più frequentemente usata è la **distribuzione normale** detta anche **di Gauss**

$N(\mu, \sigma^2)$

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



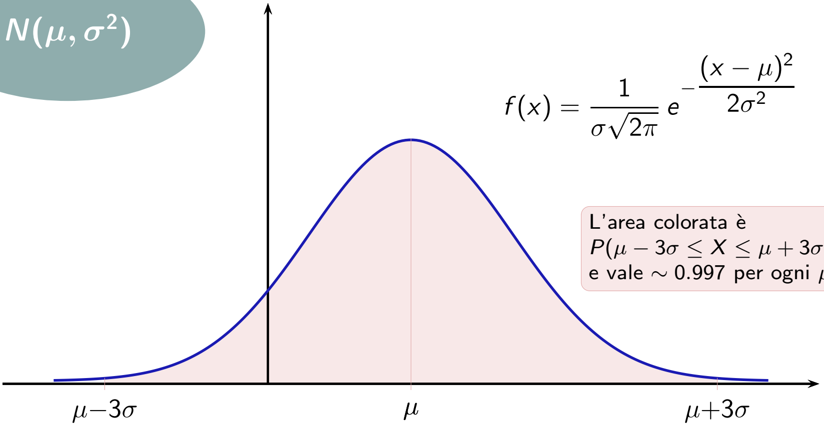
L'area colorata è
$P(\mu - 2\sigma \le X \le \mu + 2\sigma)$
e vale ~ 0.95 per ogni μ e σ

Distribuzione normale

La distribuzione più frequentemente usata è la **distribuzione normale** detta anche **di Gauss**

$N(\mu, \sigma^2)$

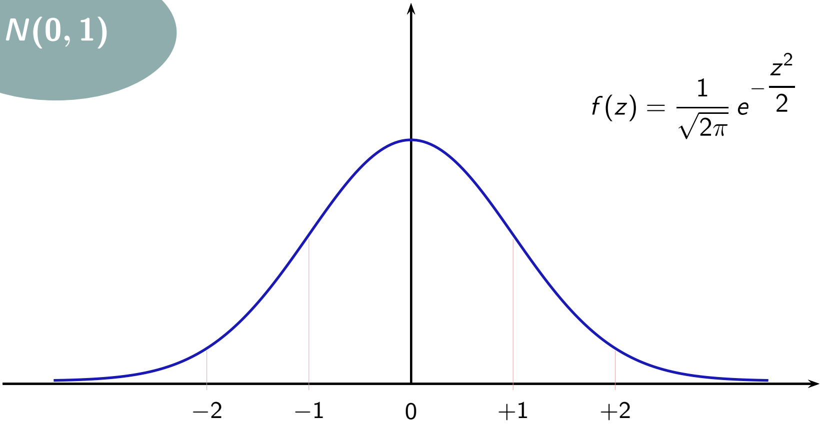
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}$$



L'area colorata è
$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma)$
e vale ~ 0.997 per ogni μ e σ

Se $\mu = 0$ e $\sigma = 1$ diremo che la normale è **standard**.

$N(0,1)$



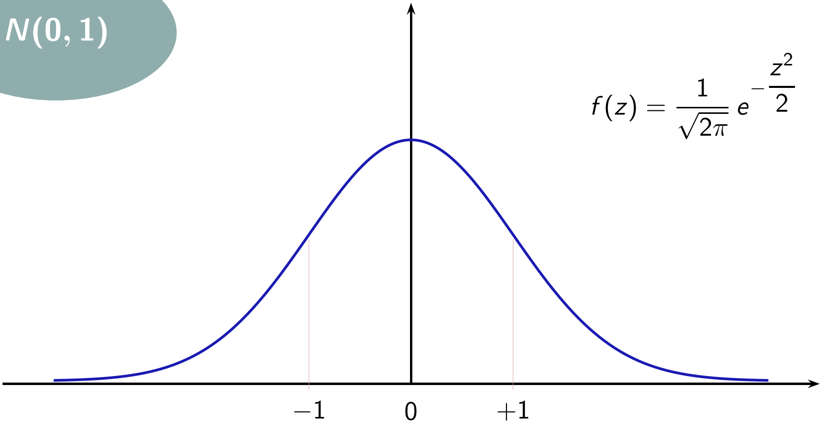
$$f(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}}$$

Standardizzazione

N.B. Se X è una v.a. $N(\mu, \sigma^2)$ allora $Z = \dfrac{X - \mu}{\sigma}$ è normale standard.

$\boldsymbol{N(0,1)}$

$$f(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}}$$



Se $X_1, \ldots, X_n \sim N(\mu_i, \sigma_i^2)$ sono v.a.i. allora

$$\sum_{i=0}^{n} X_i \text{ è una v.a. di tipo } N(\mu, \sigma^2) \text{ dove } \mu = \sum_{i=1}^{n} \mu_i \text{ e } \sigma^2 = \sum_{i=1}^{n} \sigma_i^2.$$

Se $Z_1, \ldots, Z_n \sim N(0,1)$ (quindi, normali standard) sono v.a.i. allora

$$Q \;=\; \sum_{i=1}^{n} Z_i^2 \text{ è una v.a. di tipo } \boldsymbol{\chi_n^2}$$

Se $Z \sim N(0,1)$ e $Q \sim \chi_{n-1}^2$

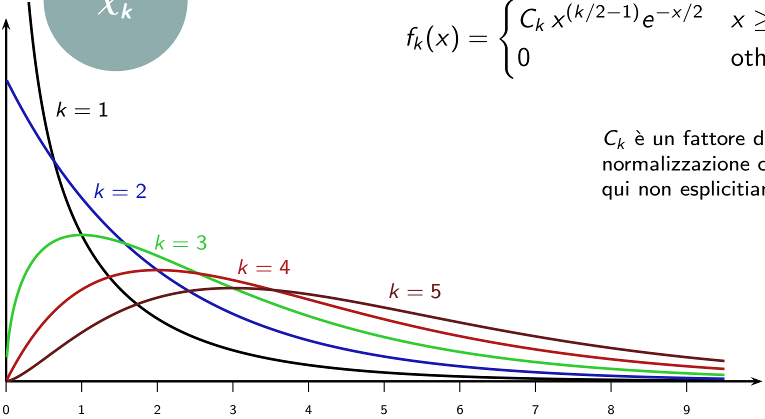
$$T \;=\; \frac{Z}{\sqrt{\dfrac{Q}{n-1}}} \text{ è una v.a. di tipo } \boldsymbol{t_{n-1}}$$

$n-1$ è indicato con ν o con df (abbreviazione per degrees of freedom)

Le distribuzioni chi-quadrato

Se $Z_1, \ldots, Z_k \sim N(0,1)$ sono v.a.i. allora $Q = \sum_{i=1}^{k} Z_i^2$ ha distribuzione

χ^2_k

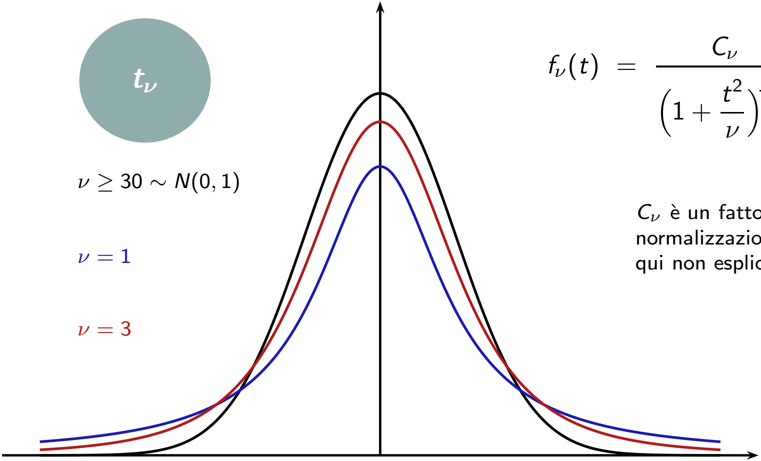


$$f_k(x) = \begin{cases} C_k\, x^{(k/2-1)} e^{-x/2} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

C_k è un fattore di normalizzazione che qui non esplicitiamo.

Anche qui una famiglia di distribuzioni (ν è un qualsiasi intero positivo)

t_ν



$$f_\nu(t) \;=\; \frac{C_\nu}{\left(1 + \dfrac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}}$$

C_ν è un fattore di normalizzazione che qui non esplicitiamo.

Se $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ sono indipendenti allora

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \text{ è una v.a. di tipo } t_{n-1}$$

dove

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2} \qquad \text{(lo stimatore della deviazione standard)}$$

La dimostrazione è laboriosa e viene omessa.

Valore atteso e varianza di una v.a. continua

Sia X è una v.a. continua con densità di probabilità $f(x)$.

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{+\infty} x\, f(x)\, dx &&= \mu \\
Var(X) &= \int_{-\infty}^{+\infty} (x-\mu)^2\, f(x)\, dx &&= E(X^2) - E(X)^2
\end{aligned}
$$

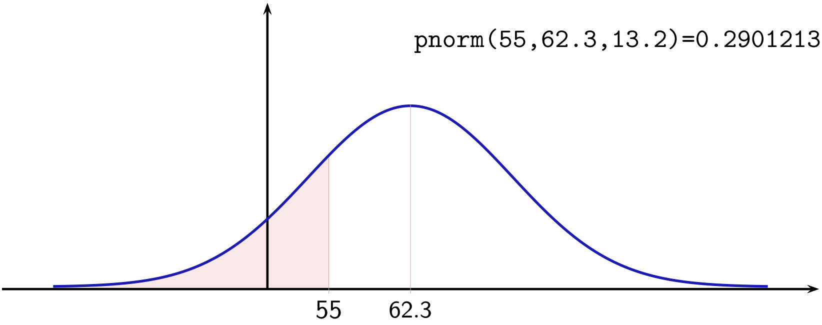
Esempio

La distribuzione di peso di una data popolazione è approssimativamente normale con media $\mu = 62.3$kg e deviazione standard $\sigma = 13.2$kg.

- Qual'è la probabilità che un soggetto selezionato casualmente pesi meno di 55kg?

$P(X \leq 55)$ dove $X \sim N(62.3, (13.2)^2)$

`pnorm(55,62.3,13.2)=0.2901213`



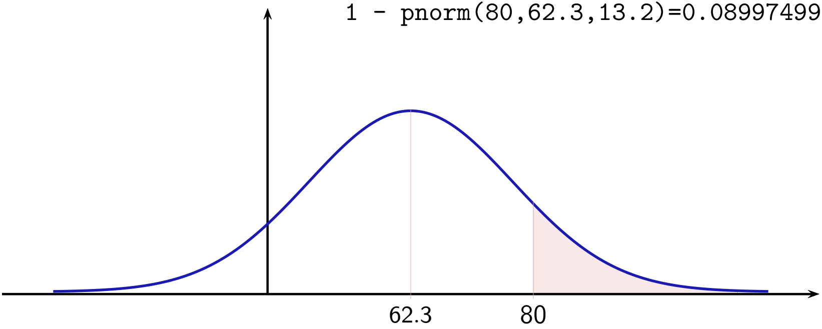
Esempio

La distribuzione di peso di una data popolazione è approssimativamente normale con media $\mu = 62.3$kg e deviazione standard $\sigma = 13.2$kg.

- Qual'è la probabilità che pesi più di 80kg?

$$P(X \geq 80) = 1 - P(X \leq 80) \text{ dove } X \sim N(62.3, (13.2)^2)$$

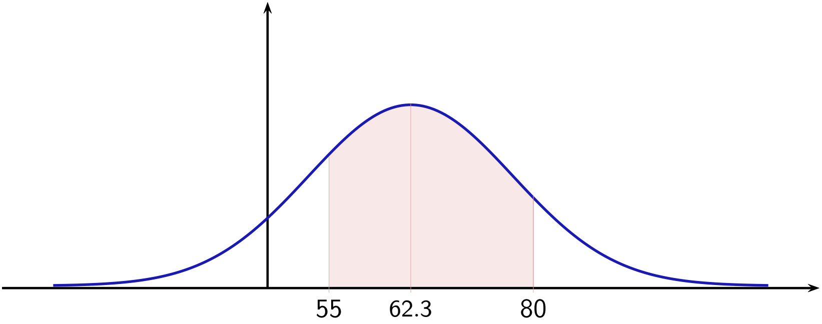
```
1 - pnorm(80,62.3,13.2)=0.08997499
```



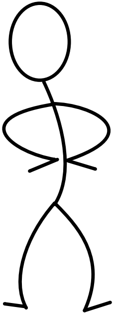
Esempio

La distribuzione di peso di una data popolazione è approssimativamente normale con media $\mu = 62.3$kg e deviazione standard $\sigma = 13.2$kg.

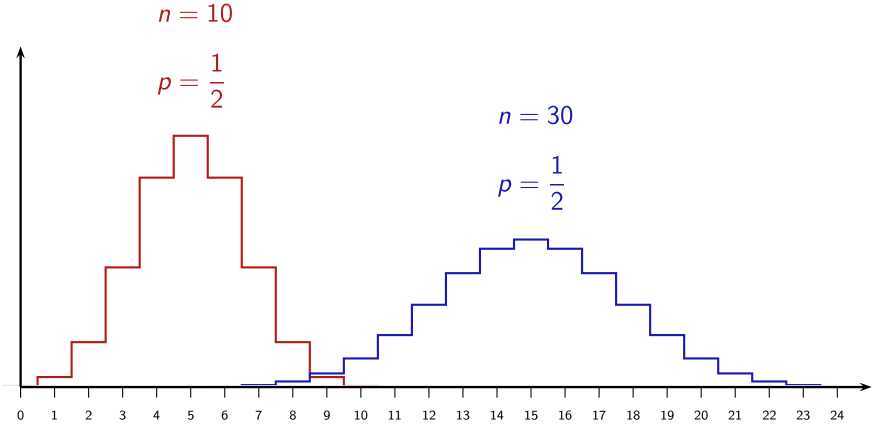
- Qual'è la probabilità che tra 5 soggetti maschi selezionati casualmente dalla popolazione, almeno uno abbia un peso compreso nell'intervallo 55-80kg?



Approssimazioni continue

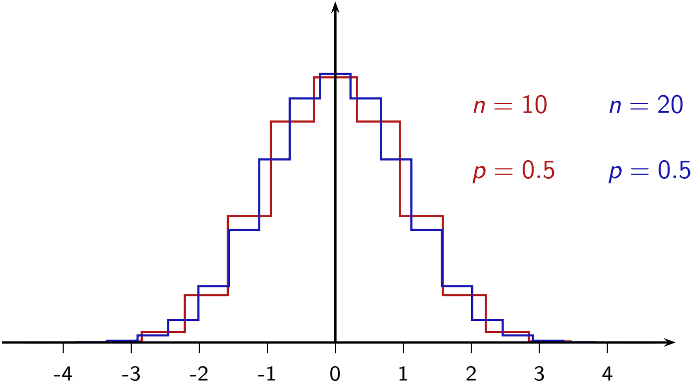


Vogliamo confrontare v.a. di tipo $B(n, p)$ per n diversi.



Approssimazione normale della binomiale.

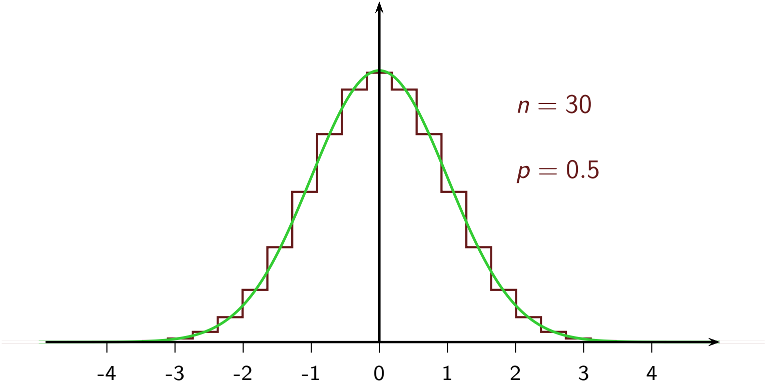
Conviene usare la forma standardizzata $\quad Z = \dfrac{X - np}{\sqrt{np(1-p)}}$



Approssimazione normale della binomiale.

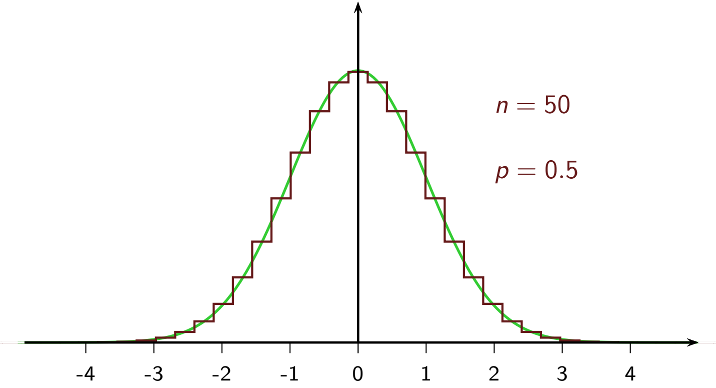
Conviene usare la forma standardizzata $\quad Z = \dfrac{X - np}{\sqrt{np(1-p)}}$

Per n abbastanza grande è ben approssimata da una distribuzione $N(0, 1)$.



Conviene usare la forma standardizzata $Z = \dfrac{X - np}{\sqrt{np(1-p)}}$

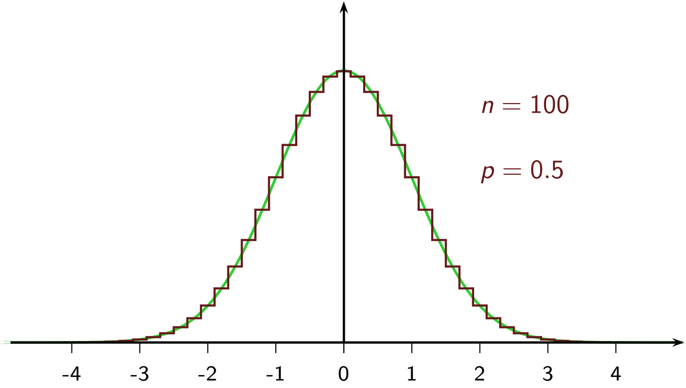
Per n abbastanza grande è ben approssimata da una distribuzione $N(0,1)$.



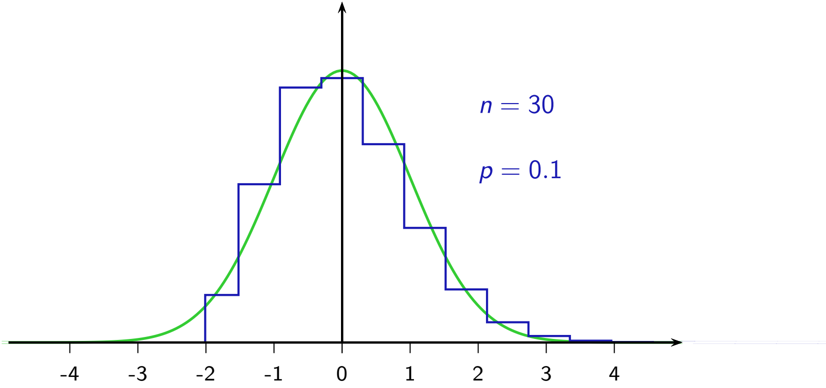
Approssimazione normale della binomiale.

Conviene usare la forma standardizzata $\quad Z = \dfrac{X - np}{\sqrt{np(1-p)}}$

Per n abbastanza grande è ben approssimata da una distribuzione $N(0,1)$.



Ma se p è distante da 0.5, l'approssimazione è meno buona.

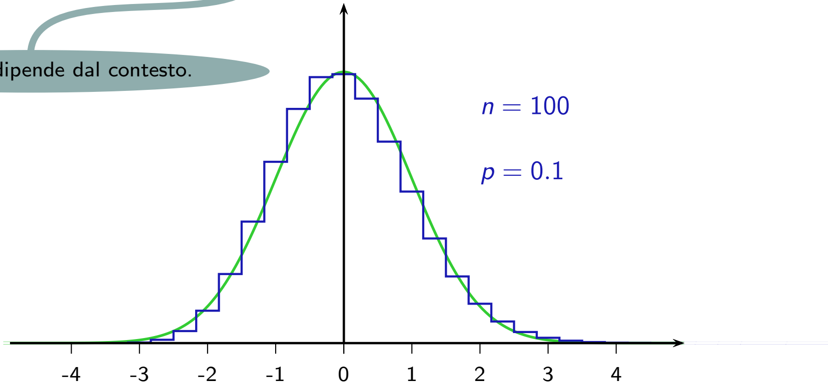


Approssimazione normale della binomiale.

Ma se p è distante da 0.5, l'approssimazione è meno buona.

Per un'approssimazione accettabile si richiede $np \geq 5$ e $n(1-p) \geq 5$

Ovviamente, dipende dal contesto.



Ma se p è distante da 0.5, l'approssimazione è meno buona.

Per un'approssimazione accettabile si richiede $np \geq 5$ e $n(1-p) \geq 5$

Ovviamente, dipende dal contesto.