Teorema Centrale del Limite

OSS Sia X discreta o continua, si ricava valore atteso di X e varianza di X a partire dalla funzione di distribuzione:

X discreta: $E(X) = \sum x_k P(X=x_k)$ $P(X=x_k) = F_X(x_k) - \lim_{x \to x_k^-} F_X(x)$.

$\rightsquigarrow Var(X)$..

X continua: $E(X) = \int_{-\infty}^{+\infty} x f_X(x)\,dx$ $f_X(x) = F_X'(x)$ (dove esiste)

$\rightsquigarrow Var(X) \rightsquigarrow$

DEF. 2 variabili aleatorie X, Y sono identicamente distribuite se $\boxed{F_X = F_Y}$

$\Rightarrow P(X \in A) = P(Y \in A) \quad \forall A \in \mathbb{R}$
$\Rightarrow E(X) = E(Y), \ Var(X) = Var(Y)$

! X, Y identic. distribuite $\not\Rightarrow X = Y$

DEF. Siano $\{X_n\}_{n \in \mathbb{N}}$ variabili aleatorie.

Si dice che le $(X_n)_{n \in \mathbb{N}}$ sono

INDIPENDENTI e IDENTICAMENTE DISTRIBUITE $\boxed{(i.i.d.)}$ se le variabili sono indipendenti e ident. distribuite:

$$F_{X_0} = F_{X_1} = \dots = F_{X_n} = \dots$$

OSS. Se le $\{X_n\}_n$ sono i.i.d. allora
$\forall n, m \quad E(X_n) = E(X_m)$ e $Var(X_n) = Var(X_m)$

ES. $\{X_i\}_i$ i.i.d. $\quad \mu = E(X_i)$, $\sigma^2 = Var(X_i) \neq 0$

$$E(X_1 + \dots + X_n) = \mu + \mu + \dots + \mu = n\mu$$
$$Var(X_1 + \dots + X_n) = Var X_1 + \dots + Var X_n = n\sigma^2 \to \sigma_{X_1 + \dots + X_n} = \sigma\sqrt{n}$$

La normalizzata di $X_1 + \dots + X_n$ è
[Normalizzata di X: $\frac{X - E(X)}{\sigma_X}$]

Normalizzata di X: $\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}}$: valore atteso 0, varianza = 1

Teorema Centrale del Limite

$\{X_n\}_{n \geq 1}$ $\boxed{\text{i.i.d.}}$, $\mu := E(X_i)$; $\sigma^2 := Var(X_i)$

$$P\left(\frac{X_1 + \dots + X_n - n\mu}{\sqrt{n}\,\sigma} \leq x\right) \xrightarrow[n \to +\infty]{} \Phi(x)$$

$F_{\frac{X_1 + \dots + X_n - n\mu}{\sqrt{n}\,\sigma}}(x)$ $\qquad$ distribuzione della normale standard.

INTERPRETAZIONE: convergenza in distribuzione.

$$\boxed{\forall x \in \mathbb{R} \quad \lim_{n \to +\infty} F_{\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}}}(x) = \Phi(x)}$$

Utilizzo in pratica.

$$P(X_1 + \dots + X_n \leq y) = P(X_1 + \dots + X_n - n\mu \leq y - n\mu)$$

$$= P\left(\frac{X_1+\dots+X_n-n\mu}{\sigma\sqrt{n}} \le \frac{y-n\mu}{\sigma\sqrt{n}}\right)$$

T.C.L.: se n è "grande" $\approx \boxed{\Phi\left(\frac{y-n\mu}{\sigma\sqrt{n}}\right)}$

Consideriamo la variabile normale con
media $E(X_1+\dots+X_n)=n\mu$
varianza $Var(X_1+\dots+X_n)=n\sigma^2$: $n\mu+\sigma\sqrt{n}Z$ $Z\sim N(0,1)$

$$P(n\mu+\sigma\sqrt{n}Z\le y) = P(\sigma\sqrt{n}Z \le y-n\mu)$$
$$= P\left(Z\le\frac{y-n\mu}{\sigma\sqrt{n}}\right) = \boxed{\Phi\left(\frac{y-n\mu}{\sigma\sqrt{n}}\right)}$$

Corollario: se $\{X_n\}_n$ i.i.d. Allora

$$\forall y: \quad P(X_1+\dots+X_n\le y)\approx P(n\mu+\sqrt{n}\sigma Z\le y) \text{ con}$$
$$Z\sim N(0,1), \quad \mu=E(X_i), \quad \sigma^2=Var(X_i).$$

$X_1+\dots+X_n$ si approssima in distribuzione con una variabile NORMALE di stessa MEDIA e stessa varianza.

OSS: Vantaggio: approssimare $P(X_1+\dots+X_n\le x)$
con $P(n\mu+\sigma\sqrt{n}Z\le x) = \Phi\left(\frac{x-n\mu}{\sigma\sqrt{n}}\right)$:
$Z\sim N(0,1)$
per calcolare $P(X_1+\dots+X_n\le x)$ basta guardare la tabella della variabile NORMALE STANDARD.

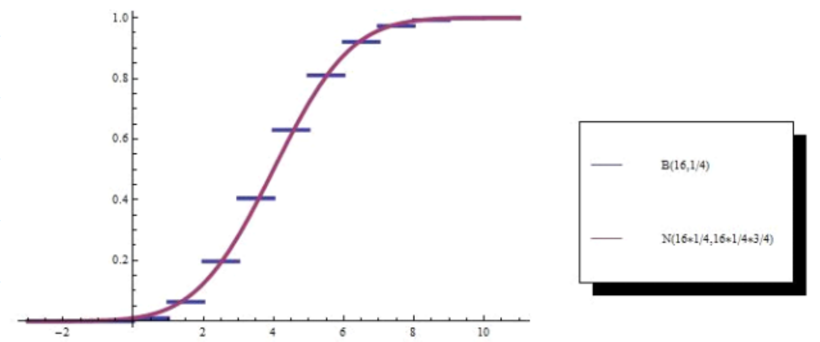
APPLICAZIONI

1) Variabile binomiale $X\sim B(n,p)$
$F_X = F_{X_1+\dots+X_n}$ $X_i\sim Be(p)$ X_i indipendenti

$E(X)=np$; $Var(X)=np(1-p)$

T.C.L $\Rightarrow$ "n grande" $\quad P(X \le x) \simeq P(np+\sqrt{np(1-p)}\,Z \le x) \quad Z\sim N(0,1)$

$$= P\left(Z \le \frac{x-np}{\sqrt{np(1-p)}}\right) = \Phi\left(\frac{x-np}{\sqrt{np(1-p)}}\right)$$



Es. $Y \sim B\left(16, \frac{5}{12}\right) \qquad P(Y\le 7) = \sum_{k=0}^{7} P(Y=k) = \sum_{k=0}^{7} \binom{16}{k}\left(\frac{5}{12}\right)^{k}\left(\frac{7}{12}\right)^{16-k}$

$\simeq 0.6673$

$$P(Y\le 7) = \cancel{\sum_{k=0}^{7} P(X=k)} =$$

$Y = X_1 + \dots + X_{16} \qquad X_i \sim Be\left(\frac{5}{12}\right)$

T.C.L: $P(Y\le 7) \simeq P(np+\sqrt{np(1-p)}\,Z \le 7)$

$p = \frac{5}{12}$, $n=16$, $Z\sim N(0,1)$

$$\simeq P\left(16\times\frac{5}{12} + \sqrt{16\times\frac{5}{12}\times\frac{7}{12}}\,Z \le 7\right)$$

$$= \Phi\left(\frac{7 - 16\times 5/12}{4\times\sqrt{\frac{5}{12}\times\frac{7}{12}}}\right) = \Phi(0.17) \simeq 0.5675$$

La correzione di continuità: se X è a valori in $\mathbb{N}$

$\forall k\in\mathbb{N} \quad P(X\le k) = P(X\le x) \quad \forall x\in[k,k+1[$: in molti casi la scelta più conveniente sembra essere $x = k+\frac{1}{2}$.

Vediamo cosa si troverebbe nell'esempio precedente.

Correzione di continuità:

Correzione di continuità:

$$P(Y\le 7)=P(Y\le 7.5)\approx P\left(np+\sqrt{np(1-p)}\,Z\le 7.5\right)$$

$$=\Phi\left(\frac{7.5-16\times 5/12}{4\times\sqrt{5/12\times 7/12}}\right)=\Phi(0.42)$$

$$\approx 0.6628$$

ESEMPIO. $X\sim B(100,\frac{5}{12})$

$$P(X\le 43)=\sum_{k=0}^{43}\binom{100}{k}\left(\frac{5}{12}\right)^k\left(\frac{7}{12}\right)^{100-k}\approx\boxed{0.6467}$$

T.C.L: $P(X\le 43)\approx P\left(100\times\frac{5}{12}+\sqrt{100\times\frac{5}{12}\times\frac{7}{12}}\,Z\le 43\right)$

$$=\Phi\left(\frac{43-100\times 5/12}{\sqrt{100\times\frac{5}{12}\times\frac{7}{12}}}\right)=\Phi\left(\frac{8}{5\sqrt{35}}\right)=\Phi(0.27)$$

$$\approx\boxed{0.6064}$$

Correzione di continuità: $P(X\le 43)=P(X\le 43.5)$

$$\approx\Phi\left(\frac{43.5-100\times 5/12}{\sqrt{100\times 5/12\times 7/12}}\right)=\Phi(0.37)\approx\boxed{0.6443}$$

ESEMPIO. Si lancia 1000 volte una moneta equilibrata (lanci indipendenti)

$Y=\#$ teste uscite in 1000 lanci.

$$P(Y\ge 480)=\sum_{k=480}^{1000}\binom{1000}{k}\frac{1}{2^k}\frac{1}{2^{1000-k}}$$

$$=\frac{1}{2^{1000}}\sum_{k=480}^{1000}\binom{1000}{k}\approx 0.9026$$

$$P(Y\ge 480)=1-P(Y<480)=1-P(Y\le 479)$$

$$P(Y\le 479)\approx P\left(np+\sqrt{np(1-p)}\,Z\le 479\right)\qquad p=\tfrac{1}{2},\ n=1000$$

$$P(Y \le 479) \approx P(np + \sqrt{np(1-p)}\,Z \le 479) \qquad p = \tfrac{1}{2},\ n = 1000$$

$$P\left(500 + \frac{\sqrt{1000}}{2} Z \le 479\right) = \phi\left(\frac{479 - 500}{\frac{\sqrt{1000}}{2}}\right) = \phi\left(\frac{-21}{5\sqrt{10}}\right)$$

$$= 1 - \phi\left(\frac{+21}{5\sqrt{10}}\right)$$

$$\Rightarrow P(Y \ge 480) = 1 - P(Y \le 479) \approx \phi\left(\frac{21}{5\sqrt{10}}\right) = \phi(1.33)$$

$$\approx 0.9082$$

Corr. di Continuità: $\phi\left(\frac{20.5}{5\sqrt{10}}\right) = \phi(1.3) = 0.9032$

2) VARIABILE DI POISSON.

Se $X \sim Po(\lambda)$ $\quad E(X) = \lambda,\quad Var(X) = \lambda \quad (\lambda > 0)$

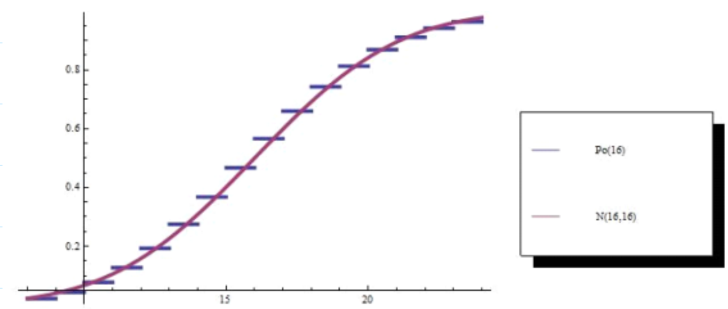
$$\boxed{\forall x:\ P(X \le x) \approx P(\lambda + \sqrt{\lambda}\, Z \le x)}$$

Infatti $\forall n \ge 1$ siano $X_1, \dots, X_n \sim Po\left(\frac{\lambda}{n}\right)$ indip.

$X_1 + \dots + X_n \sim Po(\lambda)$

$$\Rightarrow P(X \le x) = P(X_1 + \dots + X_n \le x) \approx P(\lambda + \sqrt{\lambda}\, Z \le x) \quad \text{T.C.L.}$$

#



ESEMPIO. $X \sim Po(20)$ $\quad P(X>15)$

Calcolo preciso: $P(X>15) = 1-P(X\leq 15) = 1-\sum_{k=0}^{15} e^{-20}\frac{20^k}{k!}$
≈ 0.8435

Approssimazione con la normale:

$P(X>15) = 1-P(X\leq 15)$;

$P(X\leq 15) \approx P(20+\sqrt{20}\,Z \leq 15) = \phi\left(\frac{15-20}{\sqrt{20}}\right) = \phi\left(\frac{-5}{\sqrt{20}}\right)$
$= 1-\phi\left(5/\sqrt{20}\right)$

($X \uparrow$ $Po(20)$)

$\Rightarrow P(X>15) \approx \phi\left(5/\sqrt{20}\right) \simeq \phi(1.12) \approx \boxed{0.8696}$

Con la correzione di continuità: $P(X\leq 15) = P(X\leq 15.5)$

$\rightsquigarrow \phi\left(\frac{4.5}{\sqrt{20}}\right) = \phi(1.01) \approx 0.8438$

3) Somme di variabili i.i.d.

ESEMPIO. Dado 1, 2, 3, 4, 5, 6.

a) Y = esito lancio equilibrato

Calcolare $E(Y)$ $\quad Var(Y)$

$E(Y) = 1\cdot\frac{1}{6} + 2\cdot\frac{1}{6} + \dots + 6\cdot\frac{1}{6}$

$= \frac{1}{6}(1+2+\dots+6) = \frac{1}{6}\cdot\frac{6\cdot 7}{2} = \boxed{\frac{7}{2} = E(Y)}$

$Var(Y) = E(Y^2) - E(Y)^2 = E(Y^2) - \left(\frac{7}{2}\right)^2$

$E(Y^2) = 1^2 P(Y=1) + 2^2 P(Y=2) + \dots + 6^2 P(Y=6)$

$= \frac{1}{6}(1^2+2^2+\dots+6^2) \rightsquigarrow \boxed{Var\,Y = \frac{35}{12}}$

b) Qual è la probabilità che servano almeno 80 lanci affinché la SOMMA dei punteggi ottenuti sia ≥ 300? (Lanci indipendenti, dado non truccato).

X_i = esito del lancio i

$$P(X_1 + \dots + X_{79} < 300)$$

Es $P(X_1 + \dots + X_{79} = 200) = \sum P(X_1 = n_1, X_2 = n_2, \dots, X_{79} = n_{79},\ n_1 + \dots + n_{79} = 200)$

MOLTO COMPLICATO !!

T.C.L $\quad X_1 + \dots + X_{79} \approx 79\mu + \sqrt{79}\,\sigma Z \qquad Z \sim N(0,1)$

$$\sigma^2 = \frac{35}{12} \qquad \mu = \frac{7}{2}$$

$$P(X_1 + \dots + X_{79} < 300) = P(X_1 + \dots + X_{79} \leq 299)$$

$$\approx P\left(79 \times \frac{7}{2} + \sqrt{79}\sqrt{\frac{35}{12}}\, Z \leq 299\right)$$

$$= \phi\left(\frac{299 - 79 \times 7/2}{\sqrt{79} \times \sqrt{\frac{35}{12}}}\right) \approx \phi(1.55) \approx 0.94$$

PROBLEMI INVERSI

ESERCIZIO. Moneta equilibrata, n lanci indipendenti. Determinare $n \in \mathbb{N}$ affinché la probabilità che le Teste su n lanci appaiano con FREQUENZA $\geq 48\%$ superi il 99%.

$$P(\text{Frequenza nei primi } n \text{ lanci} \geq 0.48)$$

$$P\left(\frac{X_1 + \dots + X_n}{n} \geq 0.48\right) \qquad X_i \sim B\left(\tfrac{1}{2}\right)$$

Trovare n affinché $P\left(\frac{X_1 + \dots + X_n}{n} \geq 0.48\right) \geq 99\%$

$$P(X_1 + \dots + X_n \geq 0.48n) = 1 - P(X_1 + \dots + X_n < 0.48n)$$

$$P(X_1 + \dots + X_n < 0.48n) \underset{TCL}{\approx} P\left(\frac{n}{2} + \sqrt{n \tfrac{1}{2} \cdot \tfrac{1}{2}}\, Z < 0.48n\right)$$

$$P(X_1+\dots+X_n < 0.48n) \underset{TCL}{\simeq} P\left(\frac{n}{2}+\sqrt{n}\,\frac{1}{2}\cdot\frac{1}{2}\,Z < 0.48n\right)$$

$$= P\left(\frac{\sqrt{n}}{2} Z < 0.48n - 0.5n\right) = P\left(Z < \frac{-0.04n}{\sqrt{n}}\right)$$

$$= P(Z < -0.04\sqrt{n}) = \underline{\Phi(-0.04\sqrt{n})}$$

Trovare n affinché $1-\Phi(-0.04\sqrt{n}) > 0.99$

$$\iff \quad \Phi(0.04\sqrt{n}) > 0.99$$

Dato che $\Phi(2.33) = 0.9901$ basta che

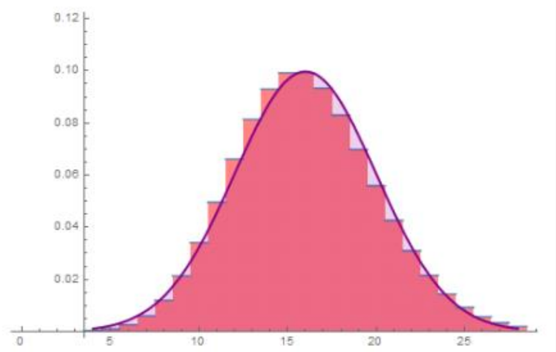
$0.04\sqrt{n} > 2.33$ cioè $\sqrt{n} > \dfrac{2.33}{0.04}$,

$$n > \left(\frac{2.33}{0.04}\right)^2 \simeq 3393$$

APPROSSIMAZIONE DI DENSITÀ DISCRETE

1) OSSERVAZIONE. Il teorema Centrale del limite mostra che sotto adeguate ipotesi la funzione di distribuzione di $X_1+\dots+X_n$ "assomiglia" a quella di una variabile normale, per questo si parla di "convergenza in distribuzione".

In alcuni casi anche la densità (discreta se $X_1+\dots+X_n$ è discreta, o continua se $X_1+\dots+X_n$ è continua) "assomiglia" a quella di una variabile normale: ciò spiega come mai i grafici di svariate densità discrete assomiglino ad una gaussiana. Questo fatto non è una conseguenza diretta del Teorema Centrale del limite.



La densità discreta di una variabile di Poisson confrontata con la densità di una normale di stessa media e varianza

In particolare si ha:

Se X è binomiale o Poisson di media μ e varianza σ

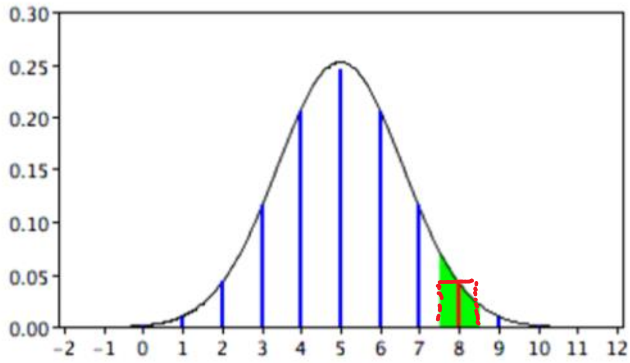
$$P(X=k) \approx f_{\mu+\sigma Z}(k)$$

2) È interessante, dal punto di vista pratico, sapere che se X è binomiale o Poisson, per k valore assunto si ha

$$P(X=k) = P\left(k-\tfrac{1}{2} \le X \le k+\tfrac{1}{2}\right)$$

$$= P\left(X \le k+\tfrac{1}{2}\right) - P\left(X \le k-\tfrac{1}{2}\right)$$

$$\approx P\left(\mu+\sigma Z \le k+\tfrac{1}{2}\right) - P\left(\mu+\sigma Z \le k-\tfrac{1}{2}\right)$$



Tale valore si trova usando le tabelle della Φ, ed è più facile da calcolare rispetto a $f_{\mu+\sigma Z}(k)$

ESEMPIO $X \sim B(12, \frac{1}{2})$ Si vuole calcolare $P(X=7)$

a) Valore preciso b) Usando la Φ c) Approssimaz. con la densità normale

OSSERVAZIONI FINALI

Il Th. Centrale del limite vale sotto condizioni meno restrittive:

- le $(X_i)_i$ possono non essere indipendenti (ma "quasi")
- le $(X_i)_i$ possono non essere ident. distribuite (ma "quasi").

Di fatto, sovrapposizioni di fenomeni casuali danno luogo a una variabile (quasi) normale.
Ciò spiega come la distribuzione delle altezze ad esempio sia distribuita in modo normale. In genere l'approssimazione è tanto più buona quanto si è vicini alla media.

ESEMPIO. Altezza delle donne

ALTEZZE

Le altezze degli studenti del corso FAMP 2016 / 17

```
data1 := {174.00, 180.00, 175.00, 160.00, 174.00, 172.00, 179.00, 170.00,
   165.00, 176.00, 178.00, 179.00, 162.00, 187.00, 178.00, 166.00, 185.00,
   172.00, 162.00, 180.00, 180.00, 183.00, 165.00, 165.00, 158.00, 177.00,
   172.00, 174.00, 170.00, 175.00, 182.00, 187.00, 165.00, 185.00, 170.00,
   173.00, 163.00, 183.00, 184.00, 162.00, 175.00, 185.00, 158.00, 179.00,
   175.00, 172.00, 169.00, 168.00, 180.00, 162.00, 179.00, 181.00, 170.00,
   180.00, 167.00, 175.00, 187.00, 175.00, 167.00, 169.00, 184.00, 177.00,
   186.00, 177.00, 178.00, 183.00, 170.00, 180.00, 183.00, 187.00, 168.00,
   178.00, 165.00, 173.00, 183.00, 167.00, 170.00, 169.00, 163.00, 175.00,
   165.00, 186.00, 168.00, 185.00, 176.00, 176.00, 185.00, 175.00, 157.00,
   177.00, 183.00, 167.00, 173.00, 180.00, 174.00, 181.00, 169.00, 183.00,
   163.00, 187.00, 157.00, 173.00, 172.00, 168.00, 170.00, 180.00, 167.00,
   186.00, 172.00, 171.00, 176.00, 154.00, 173.00, 156.00, 170.00, 178.00,
   165.00, 163.00, 173.00, 174.00, 187.00, 198.00, 157.00, 179.00, 171.00,
   183.00, 186.00, 175.00, 165.00, 176.00, 172.00, 183.00, 175.00, 180.00}
```

```
Max[{data1}]
```

198.

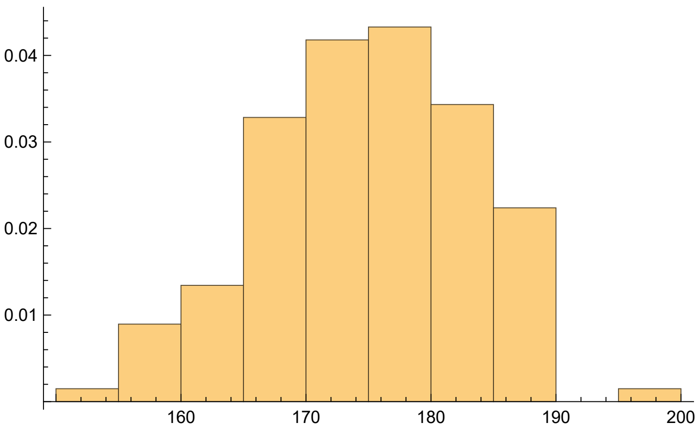
```
Min[{data1}]
```

154.

Distribuzione dei dati (percentuali stud. con data altezza)

```
A1 := Histogram[data1, Automatic, "ProbabilityDensity"]
```

```
A1
```



Media e varianza dei dati

```
m := Mean[data1]
```

```
s := Variance[data1]
```

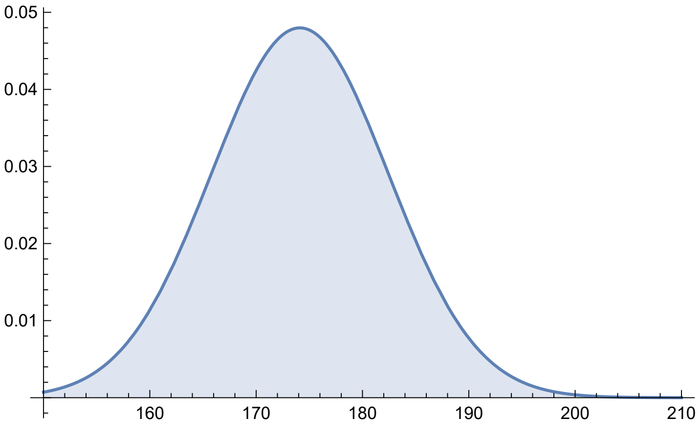
```
{m, s}
```

```
{174.112, 69.1378}
```

Densità continua della normale con data media e varianza

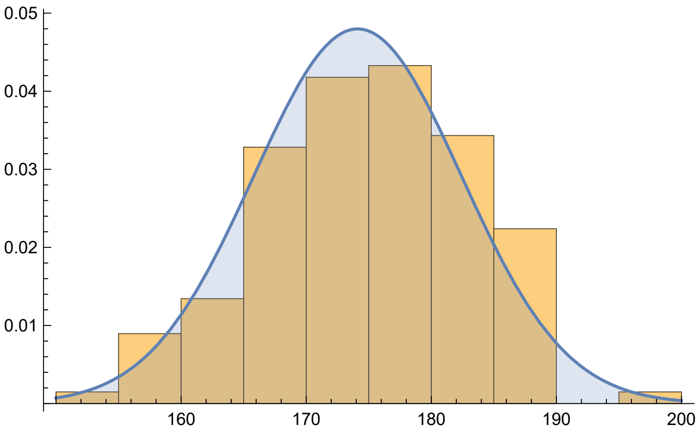
```
B1 := Plot[Table[PDF[NormalDistribution[m, Sqrt[s]], x]],
   {x, 150, 200}, Filling → Axis]
```

```
B1
```



Confronto densità continua / densità discreta

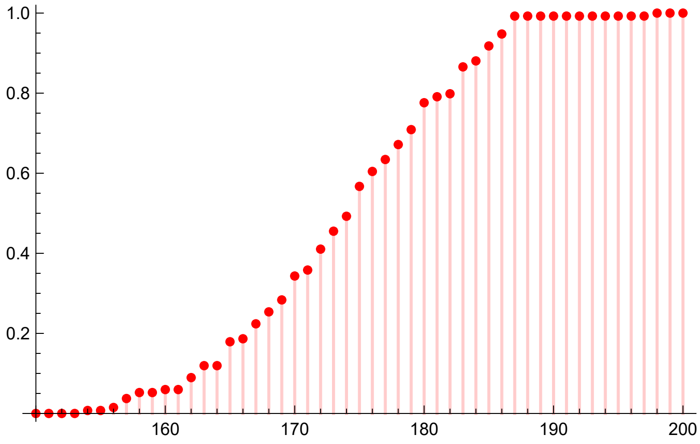
```
Show[A1, B1]
```



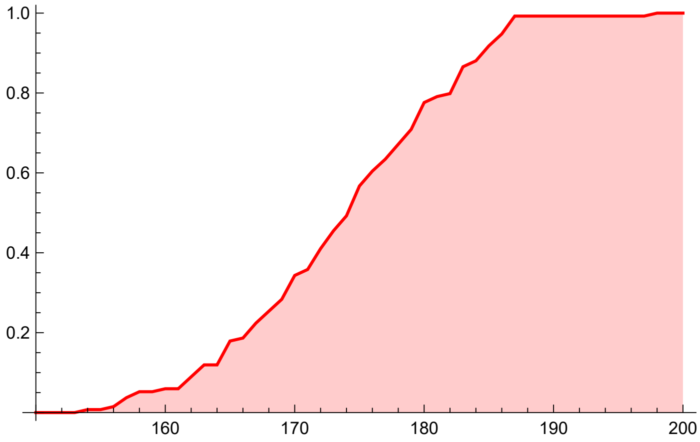

```
DATA = EmpiricalDistribution[data1];
```

Funzione di distribuzione dei dati reali

```
C1 = DiscretePlot[CDF[DATA, x], {x, 150, 200}, PlotStyle → Red]
```



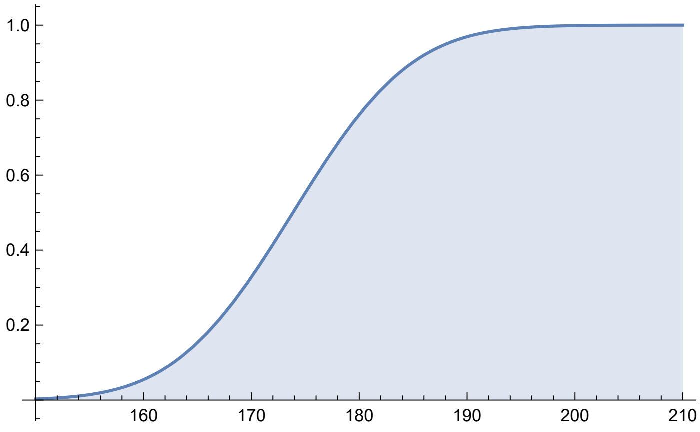

```
C1bis = DiscretePlot[CDF[DATA, x], {x, 150, 200}, PlotStyle → Red, Joined → True]
```



Funzione di distribuizione della normale di data media e varianza

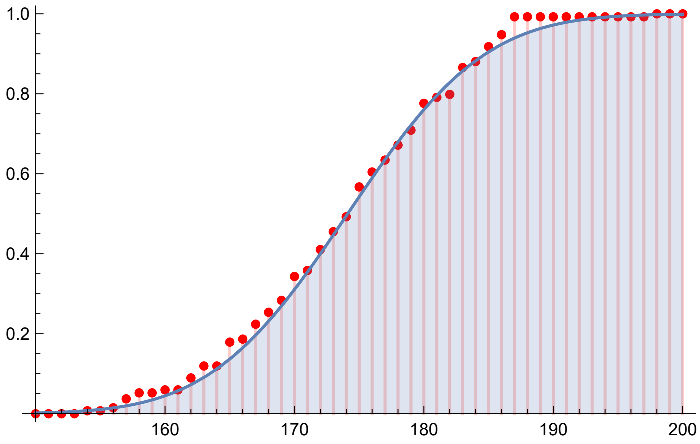
```
D1 := Plot[Table[CDF[NormalDistribution[m, Sqrt[s]], x]],
   {x, 150, 200}, Filling → Axis]
```

```
D1
```



Confronto funzioni di distribuzione normale / dati altezze

```
Show[C1, D1]
```




```
Show[C1bis, D1]
```