# STATISTICA

GIUSEPPE DE NICOLAO
Dipartimento di Informatica e Sistemistica
Università di Pavia

# SOMMARIO

- V.C. vettoriali
- Media e varianza campionarie
- Proprietà degli stimatori
- Intervalli di confidenza

# V.C. VETTORIALI

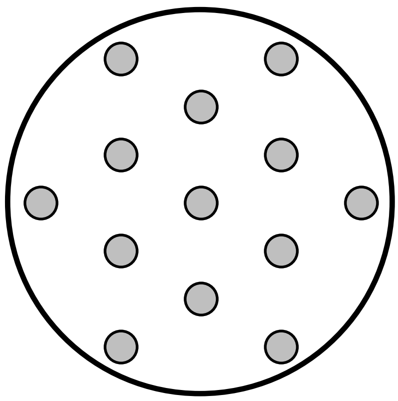
*Variabile Casuale Vettoriale X:* un esperimento casuale il cui esito è un vettore $X = [X_1\ X_2 \ldots X_n]'$ di numeri reali

*Esempio 1:* Lancio di due dadi. $X_1$ = *risultato del primo dado,* $X_2$ = *risultato del secondo dado.*

*Esempio 2:* Fermo per la strada una persona a caso e ne misuro la statura $X_1$ ed il peso $X_2$.

*Esempio 3:* Gli errori di misura $X_1, X_2, \ldots, X_n$ compiuti in $n$ misurazioni.

*Esempio 4:* Campionamento di un Wf: misure $X_1, X_2, \ldots, X_{13}$ nei diversi siti.



*Funzione di distribuzione:*

$$F_X(x) = F_X(x_1,x_2,\ldots,x_n) = P(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n)$$

*Densità di probabilità:* $f_X(x) = \dfrac{\partial^n F_X(x)}{\partial x_1 \partial x_2 \ldots \partial x_n}$

*Indipendenza:* $X_1$ e $X_2$ sono indipendenti $\Leftrightarrow$ $F_X(x_1,x_2) = F_{X_1}(x_1)F_{X_2}(x_2) \Leftrightarrow f_X(x_1,x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$

*Interpretazione:* Conoscere il valore di $X_1$ non mi è di nessun aiuto nel prevedere il valore assunto da $X_2$.

*Esempio 1 (2 dadi):* $X_1$ e $X_2$ indipendenti

*Esempio 2 (statura e peso):* $X_1$ e $X_2$ non indipendenti

*Esempio 3 (errori di misura):* in un buon sistema di misura, $X_1$, $X_2$ ,…, $X_n$ dovrebbero essere indipendenti

*Esempio 4 (campionamento Wf):* è ragionevole aspettarsi correlazione

In pratica, si assume che $X_1$ e $X_2$ siano indipendenti tutte le volte che sono generate da fenomeni fisici "non interagenti".

*Media:* $E[X] = [\, E[X_1] \;\; E[X_2] \; \dots \; E[X_n] \,]'$

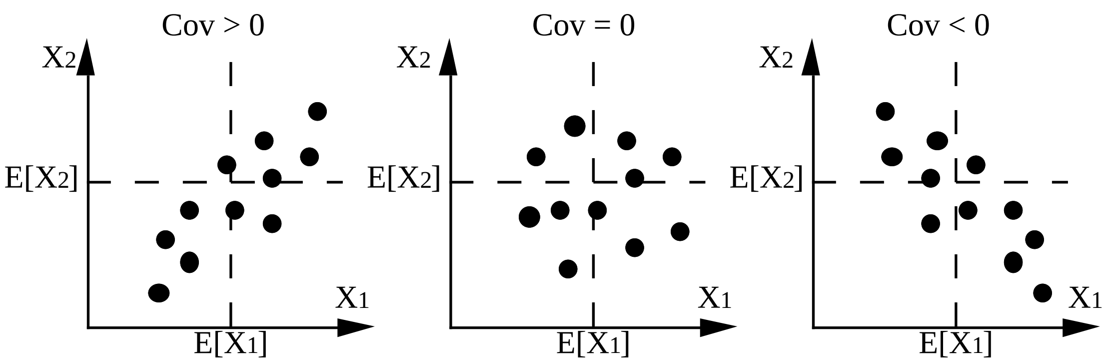
*Matrice varianza:* $Var[X] = E[(X-E[X])(X-E[X])']$

$$Var[X] = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n} \\ \dots & \dots & \dots & \dots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{bmatrix} \quad \text{(è simmetrica)}$$

$$\sigma_i^2 = Var[X_i]$$
$$\sigma_{ij} = \mathrm{E}[(X_i-E[X_i])(X_j-E[X_j])] = Cov[X_i,X_j]$$

*Incorrelazione:* $X_1$ e $X_2$ si dicono incorrelate se $Cov[X_1,X_2] = 0$

- $Cov[X_1,X_2] > 0$: se $X_1 > E[X_1]$ mi aspetto $X_2 > E[X_2]$
- $Cov[X_1,X_2] < 0$: se $X_1 > E[X_1]$ mi aspetto $X_2 < E[X_2]$



*Proprietà di media e varianza:*

- $E[AX] = AE[X]$ ($A$ matrice reale $m \times n$)

$$\Downarrow$$

$$E[X_1 + X_2] = E[X_2] + E[X_2]$$

- $Var[AX] = AVar[X]A'$

$$\Downarrow$$

$$Var[X_1 + X_2] = Var[X_1] + 2Cov[X_1, X_2] + Var[X_2]$$

- $X_1$ e $X_2$ incorrelate $\Leftrightarrow$ $Var[X_1 + X_2] = Var[X_1] + Var[X_2]$

- $X_1$ e $X_2$ indipendenti $\Rightarrow$ $X_1$ e $X_2$ incorrelate
  (in generale, $X_1$ e $X_2$ incorrelate $\not\Rightarrow$ $X_1$ e $X_2$ indipendenti)

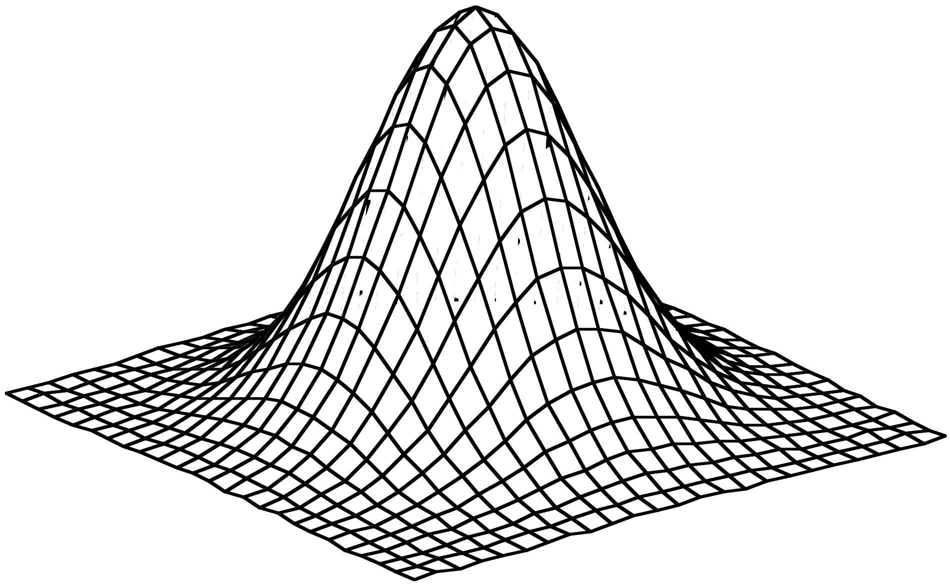
*Definizione:* Coefficiente di correlazione

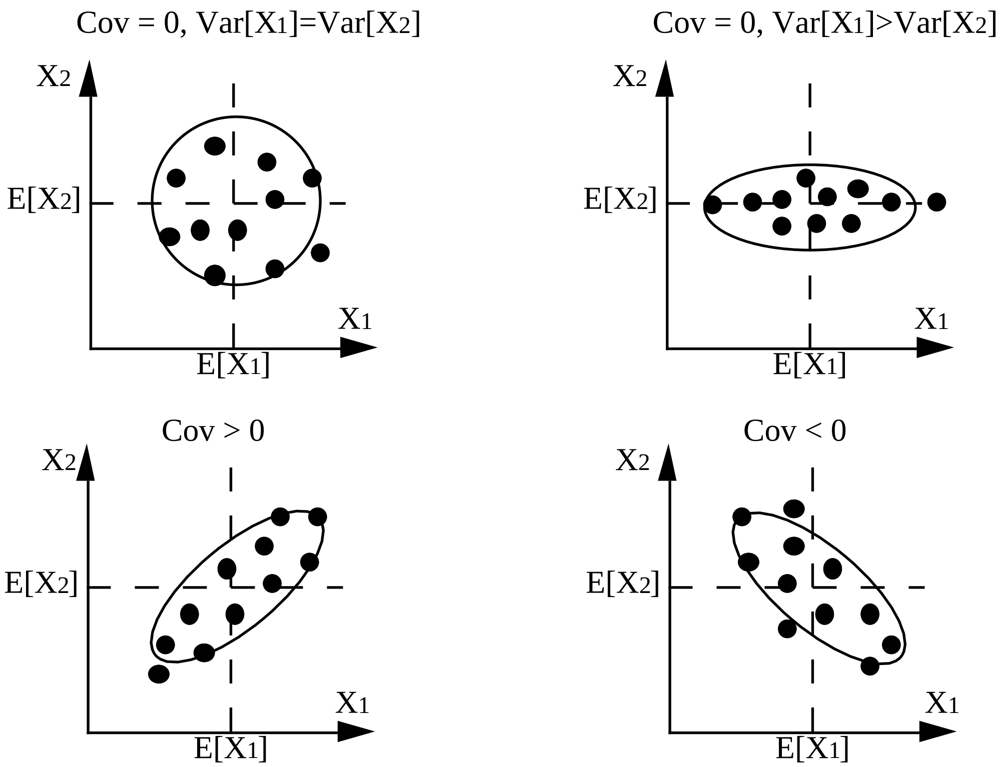
$$r_{ij} := \frac{Cov[X_i, X_j]}{\sqrt{Var[X_i]\, Var[X_j]}} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$$

*Proprietà:* $0 \leq |r_{ij}| \leq 1$

*V.C. gaussiana vettoriale:*

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}}\, e^{-\frac{(x-m)'\Sigma^{-1}(x-m)}{2}}$$





*Proprietà:*

- $E[X] = m$
- $Var[X] = \Sigma$
- E' completamente caratterizzata da media e matrice varianza
- Le singole V.C. $X_1, X_2, \ldots, X_n$ sono gaussiane
- $X_1, X_2$ incorrelate $\Rightarrow$ $X_1, X_2$ indipendenti
  (per V.C. gaussiane indipendenza e incorrelazione sono nozioni equivalenti)
- Data una matrice $A$ ed un vettore $B$, se definisco $Y = AX + B$ la nuova V.C. vettoriale $Y$ è ancora gaussiana con $E[Y] = AE[X]+B$, $Var[Y] = AVar[X]A'$

$$\Downarrow$$

$X_i$ gaussiane $\Rightarrow$ $Y=a_0+a_1X_1+\ldots+a_nX_n$ *è gaussiana*

*(la combinazione lineare di V.C. gaussiane è gaussiana)*

*Teorema centrale del limite:* Siano $X_i$ V.C. indip.. Allora, sotto larghe ipotesi, la V.C. $Y = X_1+X_2+\ldots+X_i$ tende, per $i \rightarrow \infty$, ad una V.C. con f.d.d. gaussiana.

# MEDIA E VARIANZA CAMPIONARIE

*Ipotesi:* Sono disponibili $n$ valori $x_i$ indipendenti e identicamente distribuiti con media $\mu$ e varianza $\sigma^2$

*Problema (classico):* Stimare $\mu$ e $\sigma^2$

*Media campionaria:*

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

*Varianza campionaria:*

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{X})^2}{n-1}$$

La media campionaria e la varianza campionaria (da non confondere con la media e la varianza "teoriche") sono *stimatori* che dipendono dai dati

$$\Downarrow$$

*$\bar{X}$ e $S^2$ sono variabili casuali!*

## Proprietà della media campionaria (osservazioni i.i.d.)

*Media della media campionaria:* $E[\bar{X}] = \mu$

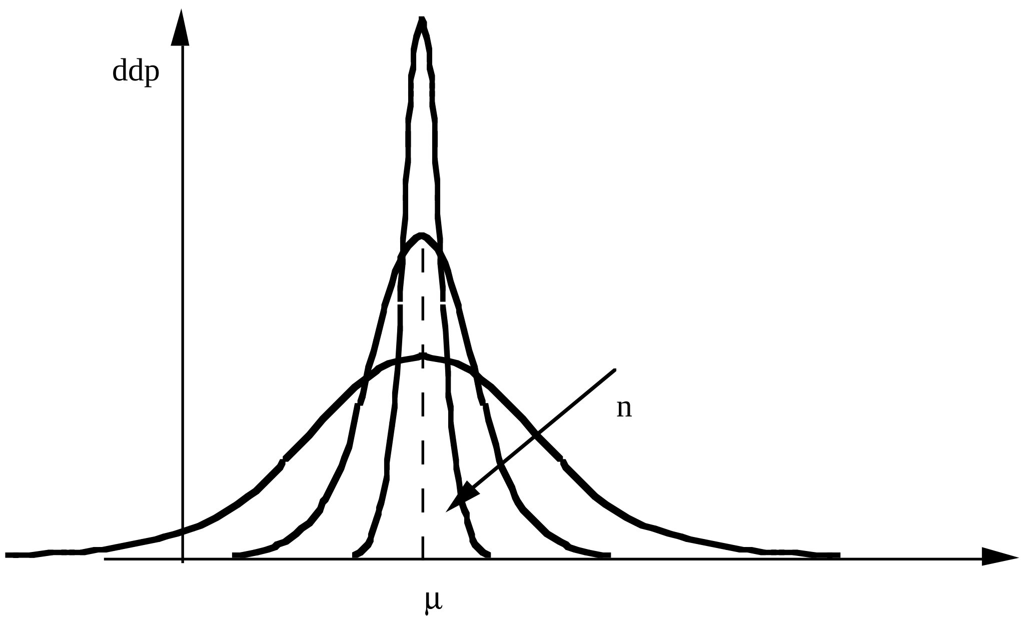
Infatti:

$$E\left[\frac{\sum_{i=1}^{n} x_i}{n}\right] = \frac{1}{n}\sum_{i=1}^{n} E[x_i] = \frac{n\mu}{n} = \mu$$

*Varianza della media campionaria:* $Var[\bar{X}] = \frac{\sigma^2}{n}$

Infatti:

$$Var\left[\frac{\sum_{i=1}^{n} x_i}{n}\right] = \frac{1}{n^2}\sum_{i=1}^{n} Var[x_i] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$




*Distribuzione:*

- $x_i$ gaussiani $\Rightarrow \bar{X}$ gaussiana
  (somma di V.C. gaussiane)
- $x_i$ non gaussiani $\Rightarrow \bar{X}$ gaussiana per $n \to \infty$
  (per il teorema centrale del limite)

*Standardizzazione:*

$$Z = \frac{\bar{X} - E[\bar{X}]}{\sqrt{Var[\bar{X}]}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$\Downarrow$$

$$\bar{X} = \mu + \frac{\sigma}{\sqrt{n}} Z \,, \quad Z \sim N(0,1)$$

*La media campionaria fornisce il valore della media a meno di un termine di errore che diventa piccolo per $n \to \infty$*

*Definizione: Standard Error*

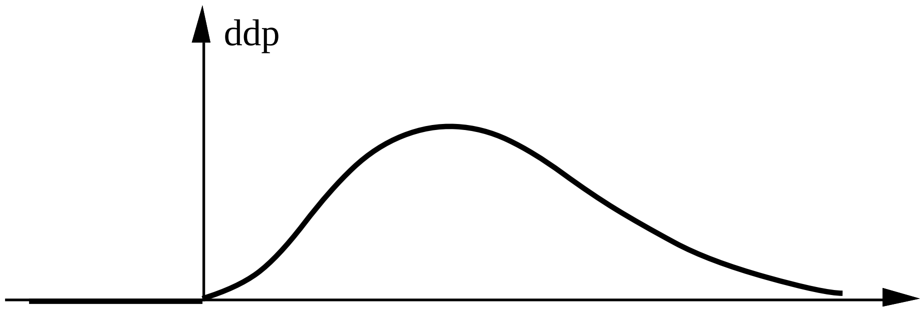
$$SE := \frac{\sigma}{\sqrt{n}}$$

## Parentesi: la V.C. *"chi-quadrato"*

*Definizione:* Siano $Z_i$, $i=1,\ldots n$, delle V.C. gaussiane standard tra loro indipendenti. Allora,

$$\chi_n^2 := \sum_{i=1}^{n} Z_i^2$$

è una V.C. che prende il nome di "chi quadrato" ad $n$ gradi di libertà.




*Proprietà:*

- $E[\chi_n^2] = n$
- $Var[\chi_n^2] = 2n$
- $\chi_n^2$ tende ad una d.d.p. gaussiana per $n \to \infty$
  (per il teorema centrale del limite)

# Proprietà della varianza campionaria (osservazioni i.i.d.)

*Media della varianza campionaria:* $E[S^2] = \sigma^2$

*Distribuzione:*

- $x_i$ gaussiani $\Rightarrow$ $S^2$ è una $\chi_{n-1}^2$ ed è indipendente da $\bar{X}$
  Precisamente:

$$\frac{(n-1)S^2}{\sigma^2} = \chi_{n-1}^2$$

- La distribuzione precedente non è robusta rispetto all'ipotesi di normalità delle osservazioni.

# PROPRIETA' DEGLI STIMATORI

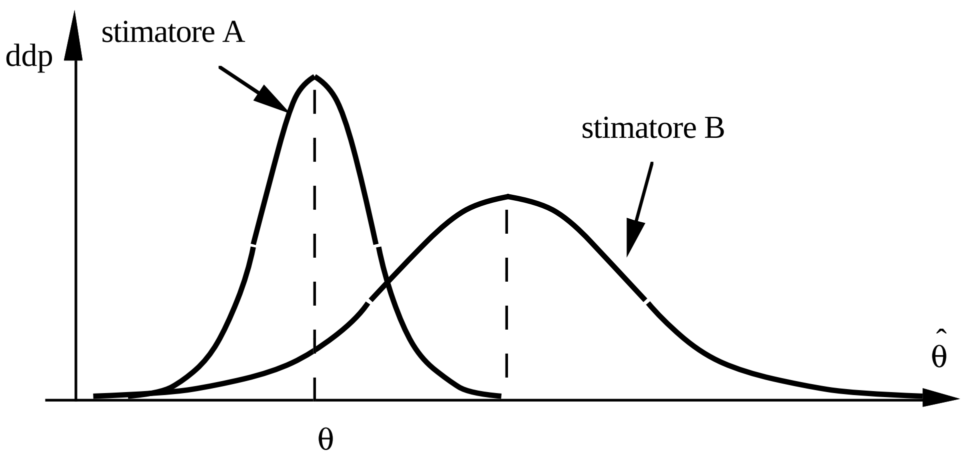
D'ora in poi, $\hat{\theta}$ indica uno stimatore di un parametro $\theta$.

*Esempio:* $\theta = \mu$, $\hat{\theta} = \bar{X}$

*Osservazione fondamentale:* $\hat{\theta} = \hat{\theta}(x_1,\ldots,x_n)$ è una funzione dei dati $x_1,\ldots,x_n$ che sono V.C. $\Rightarrow$ *anche* $\hat{\theta}$ *è una V.C.*

$$\Downarrow$$

*Per giudicare uno stimatore non basta vedere se "ci azzecca" in uno o due casi (sono risultati casuali!) ma bisogna analizzare la sua d.d.p.*

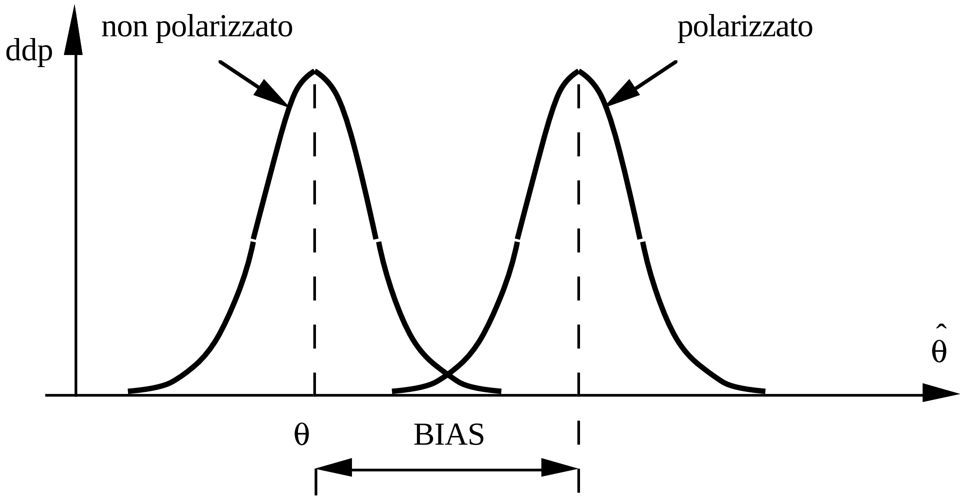



*lo stimatore A è migliore di B*

(una volta ogni tanto può però capitare che A commetta un errore maggiore di quello commesso da B)

## Proprietà degli stimatori

*Non polarizzazione:* Uno stimatore si dice *non-polarizzato* (unbiased) quando $E[\hat{\theta}] = \theta$. La differenza $E[\hat{\theta}] - \theta$ è detta BIAS (errore sistematico).




*Esempio 1:* Media campionaria ($\theta = \mu$, $\hat{\theta} = \overline{X}$).

$$E[\overline{X}] = \mu \Rightarrow \text{stimatore non polarizzato}$$

*Esempio 2:* Varianza campionaria ($\theta = \sigma^2$, $\hat{\theta} = S^2$).

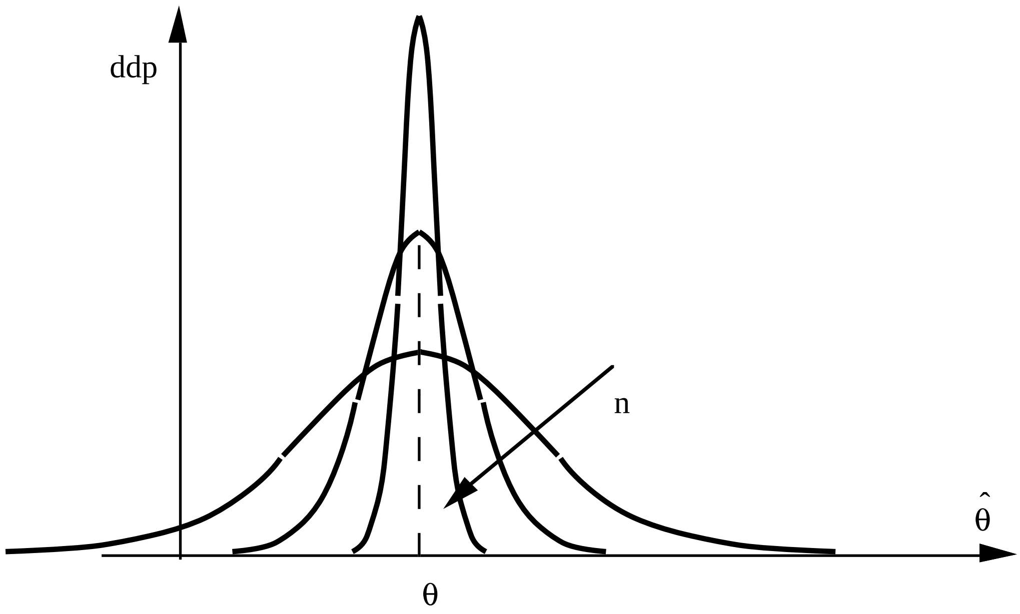
$$E[S^2] = \sigma^2 \Rightarrow \text{stimatore non polarizzato}$$

*Osservazione:* Se nel denominatore di $S^2$ usassi $n$ invece di $n$-$1$, lo stimatore sarebbe polarizzato.

*Consistenza:* Uno stimatore si dice *consistente* quando

$$\lim_{n\to\infty} \hat{\theta} = \theta$$

(nel senso che $\lim_{n\to\infty} E[\hat{\theta}] = \theta$, $\lim_{n\to\infty} Var[\hat{\theta}-\theta]^2 = 0$, per es.)




*Interpretazione:* al crescere di *n*, l'errore di stima tende a zero.

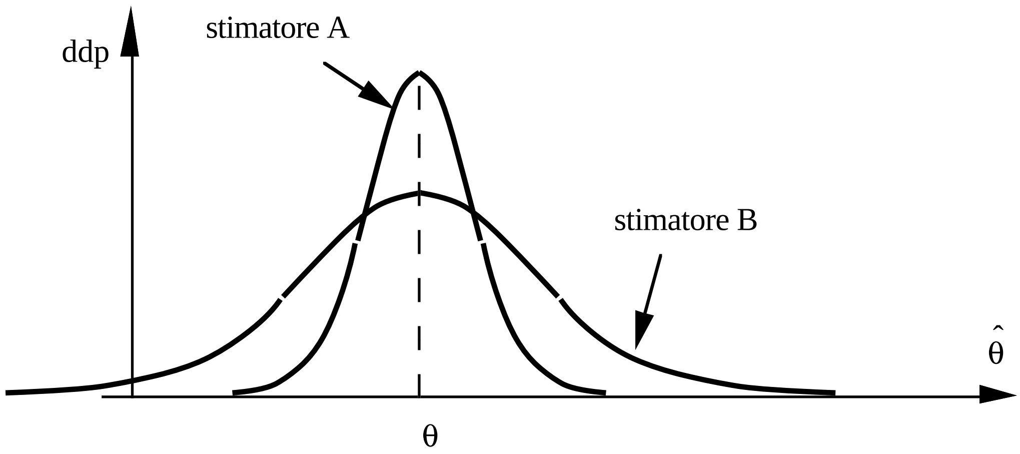
*Legge dei grandi numeri:* La media campionaria è uno stimatore consistente della media.

*Proprietà:* $S^2$ è uno stimatore consistente della varianza.

*Asintotica normalità:* Per $n \to \infty$, la f.d.d. di $\hat{\theta}$ tende a diventare gaussiana.

*Esempio:* $\overline{X}$ è asintoticamente normale (per il teorema centrale del limite).

*Osservazione:* Per gli stimatori *non polarizzati*, $Var[\hat{\theta}]$ costituisce un indice della precisione dello stimatore.




*Stimatore a minima varianza:* Lo stimatore $\hat{\theta}^m$ è detto *a minima varianza* se $\mathrm{Var}[\hat{\theta}^m] < \mathrm{Var}[\hat{\theta}]$, dove $\hat{\theta}$ è un qualsiasi altro stimatore ($\hat{\theta}^m$ è il miglior stimatore possibile).

*Esempio:* Se i dati $x_i$ sono i.i.d. gaussiani, la media campionaria è lo stimatore a minima varianza.

*Osservazione:* Per stimatori polarizzati, la varianza da sola è poco indicativa.

*Esempio:* $\hat{\theta} = 10$, indipendentemente dai dati $\Rightarrow \mathrm{Var}[\hat{\theta}] = 0$ (ma è un pessimo stimatore).

*Indice di accuratezza per stimatori polarizzati:*
Errore quadratico medio (*Mean Square Error*)

$$MSE := E[(\hat{\theta}-\theta)^2]$$

(mi dice quanto è concentrata attorno allo zero la d.d.p. dell'errore di stima)

*Osservazione: MSE* pesa sia *BIAS* che variabilità; infatti:

$$MSE := E[(\hat{\theta}-E[\hat{\theta}] + E[\hat{\theta}]-\theta)^2] =$$

$$=E[(\hat{\theta}-E[\hat{\theta}])^2 + 2(\hat{\theta}-E[\hat{\theta}])(E[\hat{\theta}]-\theta) + (E[\hat{\theta}]-\theta)^2)] =$$

$$= Var[\hat{\theta}] + 2E[\hat{\theta}-E[\hat{\theta}]](E[\hat{\theta}]-\theta) + (E[\hat{\theta}]-\theta)^2 =$$

$$= Var[\hat{\theta}] + BIAS^2$$

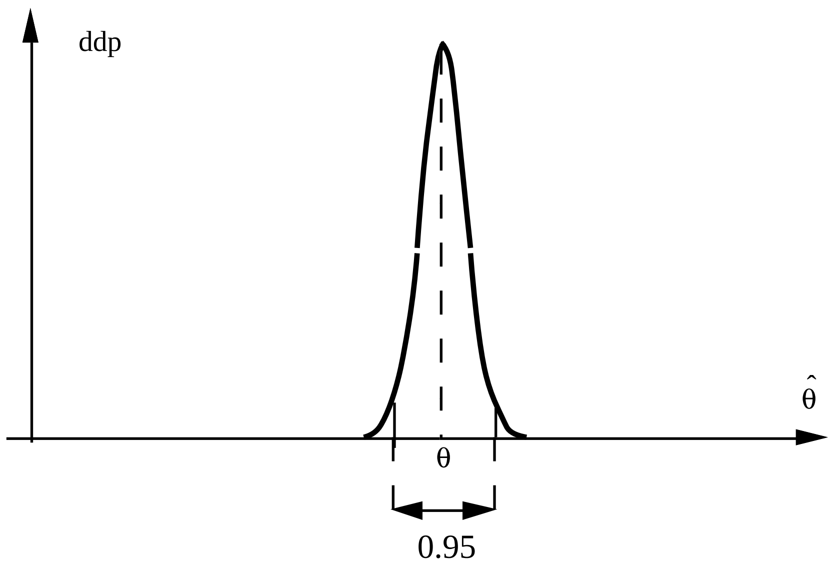
# INTERVALLI DI CONFIDENZA

Dato che $\hat{\theta}(x_1,\dots,x_n)$ è una V.C. conoscere $\hat{\theta}$ non mi dà garanzie sul valore effettivo di θ.

$$\Downarrow$$

*E' privo di senso fornire stime $\hat{\theta}$ senza fornire indicazioni sull'ampiezza dell'errore di stima $\theta$-$\hat{\theta}$*

Se $\hat{\theta}$ è non polarizzato, può essere sufficiente conoscere $Var[\hat{\theta}]$ (se è piccola, la stima è accurata)




Se conosco la d.d.p. di $\hat{\theta}$ posso dare indicazioni più precise

*Intervallo di confidenza al 95%:* $I_{95} = [L_{95}, U_{95}]$ tale che:

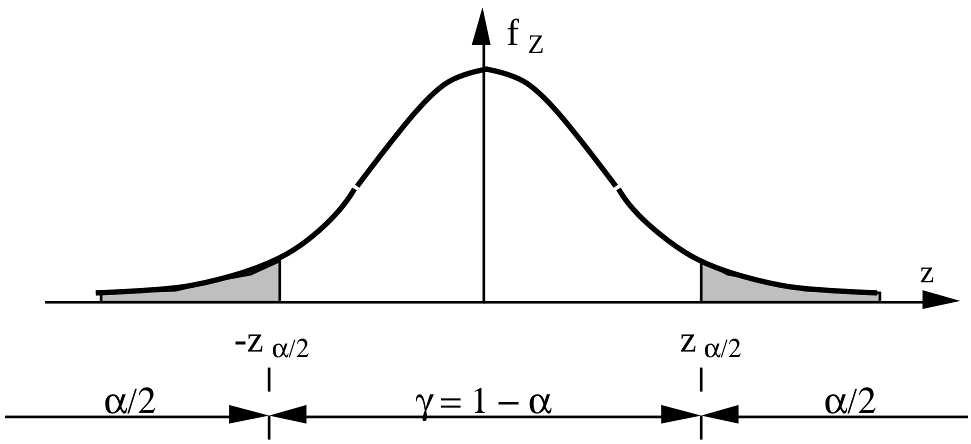
$$P(\theta \in I_{95}) = 0.95 = \gamma \text{ (coeff. di confidenza)}$$

$$\alpha := 1 - \gamma$$

## Intervalli di confidenza per la media di V.C. gaussiane i.i.d., $\sigma^2$ nota

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Cerco $z_{\alpha/2}$ tale che $P(Z \geq z_{\alpha/2}) = \alpha/2$




| $\alpha$ | 0.10 | 0.05 | 0.01 | 0.002 | 0.001 |
|---|---|---|---|---|---|
| $z_\alpha$ | 1.28 | 1.645 | 2.33 | 2.88 | 3.090 |
| $z_{\alpha/2}$ | 1.645 | 1.96 | 2.576 | 3.090 | 3.291 |

$$P(|Z| \leq z_{\alpha/2}) = 1 - \alpha = \gamma$$

$$\Downarrow$$

$$P\left(\left|\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}\right| \leq z_{\alpha/2}\right) = \gamma$$

$$\Downarrow$$

$$P\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = \gamma$$

Scegliendo $\gamma = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow z_{\alpha/2} = 1.96$

$$\Downarrow$$

$$I_{95} = [\bar{X} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}]$$

In pratica, se risulta $I_{95} = [52.26 - 0.03, 52.26 + 0.03]$, scriveremo $\mu = 52.26 \pm 0.03$

*Osservazione:* per dimezzare $I_\gamma$ bisogna quadruplicare il numero $n$ dei dati (la precisione costa!).

Scegliendo $\gamma = 0.9973 \Rightarrow \alpha = 0.0027 \Rightarrow z_{\alpha/2} = 3$

$$\Downarrow$$

$$I_\gamma = [\bar{X} - 3\frac{\sigma}{\sqrt{n}}, \bar{X} + 3\frac{\sigma}{\sqrt{n}}]$$

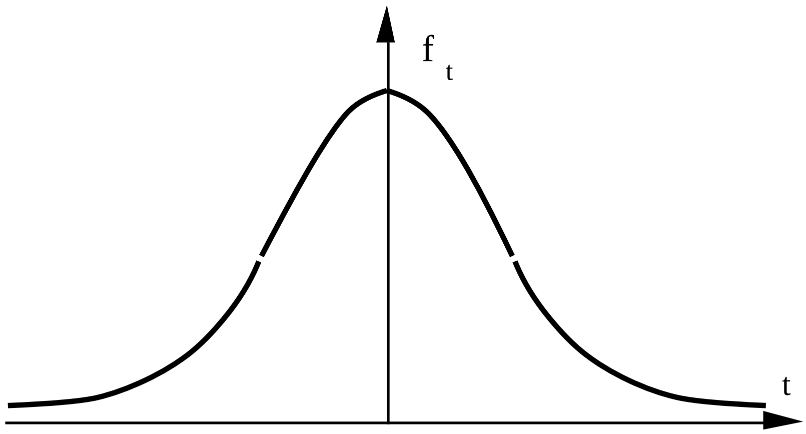
**Parentesi: la V.C. *"t di Student"***

*"Student": pseudonimo di W.S. Gosset (1876-1937)*

*Definizione:* Siano $Z_i$, $i=0,\dots n$, delle V.C. gaussiane standard i.i.d.. Allora,

$$t_n := \frac{Z_0 \sqrt{n}}{\sqrt{\sum_{i=1}^{n} Z_i^2}} = \frac{Z_0 \sqrt{n}}{\chi_n}$$

è una V.C. che prende il nome di "t di Student" ad $n$ gradi di libertà.




*Proprietà:*

- Simile ad una gaussiana ma più "panciuta"
- $E[t_n] = 0$
- $Var[t_n] = \frac{n}{n-2}$
- $t_n$ tende ad una d.d.p. gaussiana standard per $n \rightarrow \infty$

## Intervalli di confidenza per la media di V.C. gaussiane i.i.d., $\sigma^2$ non nota

Si vede che

$$\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} = t_{n-1}$$

Cerco $t_{\alpha/2}$ tale che $P(t_{n-1} \geq t_{\alpha/2}) = \alpha/2$

*(risulta sempre* $t_{\alpha/2} > z_{\alpha/2}$*)*

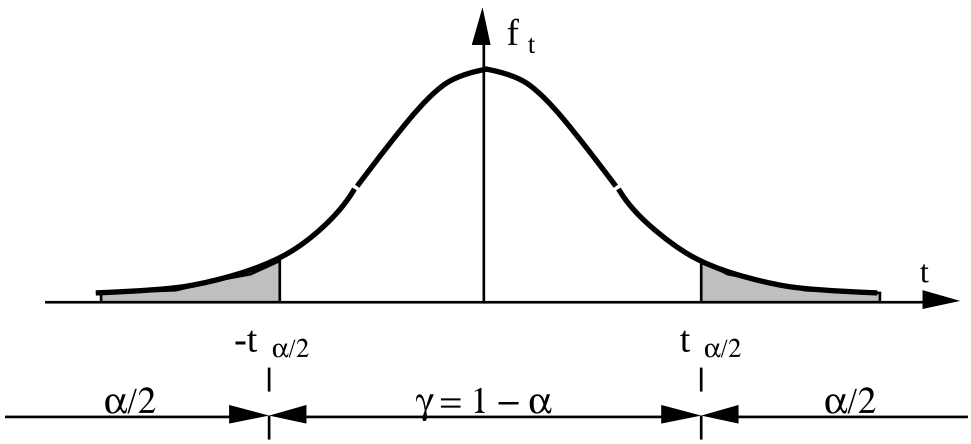



*Tabella: Valori di* $t_{\alpha/2}$

| n | $\alpha = 0.1$ | $\alpha=0.05$ |
|---|---|---|
| **1** | 6.314 | 12.706 |
| **2** | 2.920 | 4.303 |
| **3** | 2.353 | 3.182 |
| **4** | 2.132 | 2.776 |
| **5** | 2.015 | 2.571 |
| **6** | 1.943 | 2.447 |
| **7** | 1.895 | 2.365 |
| **8** | 1.860 | 2.306 |
| **9** | 1.833 | 2.262 |
| **10** | 1.812 | 2.228 |
| **11** | 1.796 | 2.201 |
| **12** | 1.782 | 2.179 |
| **13** | 1.771 | 2.160 |
| **14** | 1.761 | 2.145 |
| **15** | 1.753 | 2.131 |

| n | $\alpha = 0.1$ | $\alpha=0.05$ |
|---|---|---|
| **16** | 1.746 | 2.120 |
| **17** | 1.740 | 2.110 |
| **18** | 1.734 | 2.101 |
| **19** | 1.729 | 2.093 |
| **20** | 1.725 | 2.086 |
| **21** | 1.721 | 2.080 |
| **22** | 1.717 | 2.074 |
| **23** | 1.714 | 2.069 |
| **24** | 1.711 | 2.064 |
| **25** | 1.708 | 2.060 |
| **26** | 1.706 | 2.056 |
| **27** | 1.703 | 2.052 |
| **28** | 1.701 | 2.048 |
| **29** | 1.699 | 2.045 |
| **30** | 1.697 | 2.042 |

| n | $\alpha = 0.1$ | $\alpha=0.05$ |
|---|---|---|
| **40** | 1.684 | 2.021 |
| **60** | 1.671 | 2.000 |
| **120** | 1.658 | 1.980 |
| **∞** | 1.645 | 1.960 |

$$P(|t_{n-1}| \leq t_{\alpha/2}) = 1 - \alpha = \gamma$$

$$\Downarrow$$

$$P\left(\left|\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}\right| \leq t_{\alpha/2}\right) = \gamma$$

$$\Downarrow$$

$$P\left(\bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right) = \gamma$$

*Osservazioni:*

- Unica differenza rispetto al caso con varianza nota:
  uso $S$ e $t_{\alpha/2}$ al posto di $\sigma$ e $z_{\alpha/2}$

- Non conoscere $\sigma^2$ allarga l'intervallo

- Per $n \to \infty$ , $t_{\alpha/2}$ diventa gaussiana $\Rightarrow$ $z_{\alpha/2}$ al posto di $t_{\alpha/2}$

## Intervalli di confidenza per la varianza di V.C. gaussiane i.i.d.

$$\frac{(n-1)S^2}{\sigma^2} = \chi^2_{n-1}$$

Cerco $\chi^2_{\alpha/2,n-1}$ e $\chi^2_{1-\alpha/2,n-1}$

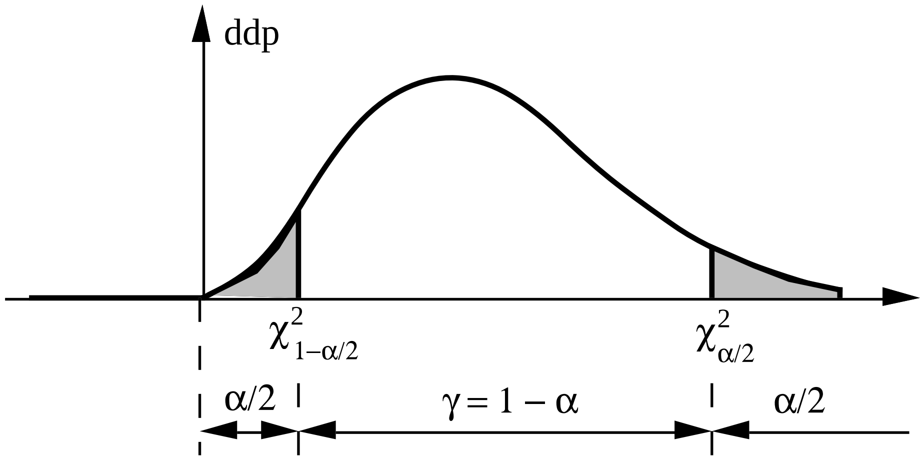



*Tabella: Valori di $\chi^2_{\alpha,n}$*

| *n* | *α=0.975* | *α=0.95* | *α=0.05* | *α=0.025* |
|---|---|---|---|---|
| **1** | 0.00 | 0.00 | 3.84 | 5.02 |
| **2** | 0.05 | 0.10 | 5.99 | 7.38 |
| **3** | 0.22 | 0.35 | 7.81 | 9.35 |
| **4** | 0.48 | 0.71 | 9.49 | 11.14 |
| **5** | 0.83 | 1.15 | 11.07 | 12.38 |
| **6** | 1.24 | 1.64 | 12.59 | 14.45 |
| **7** | 1.69 | 2.17 | 14.07 | 16.01 |
| **8** | 2.18 | 2.73 | 15.51 | 17.53 |
| **9** | 2.70 | 3.33 | 16.92 | 19.02 |
| **10** | 3.25 | 3.94 | 18.31 | 20.48 |
| **11** | 3.82 | 4.57 | 19.68 | 21.92 |
| **12** | 4.40 | 5.23 | 21.03 | 23.34 |
| **13** | 5.01 | 5.89 | 22.36 | 24.74 |
| **14** | 5.63 | 6.57 | 23.68 | 26.12 |
| **15** | 6.27 | 7.26 | 25.00 | 27.49 |

| *n* | *α=0.975* | *α=0.95* | *α=0.05* | *α=0.025* |
|---|---|---|---|---|
| **16** | 6.91 | 7.96 | 26.30 | 28.85 |
| **17** | 7.56 | 8.67 | 27.59 | 30.19 |
| **18** | 8.23 | 9.39 | 28.87 | 31.53 |
| **19** | 8.91 | 10.12 | 30.14 | 32.85 |
| **20** | 9.59 | 10.85 | 31.41 | 34.17 |
| **25** | 13.12 | 14.61 | 37.65 | 40.65 |
| **30** | 16.79 | 18.49 | 43.77 | 46.98 |
| **40** | 24.43 | 26.51 | 55.76 | 59.34 |
| **50** | 32.36 | 34.76 | 67.50 | 71.42 |
| **60** | 40.48 | 43.19 | 79.08 | 83.30 |
| **70** | 48.76 | 51.74 | 90.53 | 95.02 |
| **80** | 57.15 | 60.39 | 101.88 | 106.63 |
| **90** | 65.65 | 69.13 | 113.14 | 118.14 |
| **100** | 74.22 | 77.93 | 124.34 | 129.56 |

$$P(\chi^2_{1-\alpha/2,n-1} \le \chi^2_{n-1} \le \chi^2_{\alpha/2,n-1}) = 1 - \alpha = \gamma$$

$$\Downarrow$$

$$P\left(\chi^2_{1-\alpha/2,n-1} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{\alpha/2,n-1}\right) = \gamma$$

$$\Downarrow$$

$$P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right) = \gamma$$

*Attenzione:* Intervalli poco robusti rispetto all'ipotesi di gaussianità dei dati

# CONCLUSIONI

- Essendo basato su dati casuali, lo stimatore è una V.C.

- Caratteristiche fondamentali di uno stimatore: polarizzazione e varianza.

- E' fondamentale fornire insieme alle stime un indice della loro affidabilità

$$\Downarrow$$

*intervalli di confidenza*

- Attenzione alle ipotesi sulle distribuzioni.

# DOMANDE

**V** **F**

1. E[X+Y] = E[X]+E[Y]. ❑ ❑

2. Var[X+Y] = Var[X] + Cov[X,Y] + Var[Y]. ❑ ❑

3. Se due V.C. X e Y sono indipendenti, allora sono anche incorrelate. ❑ ❑

4. La media della varianza campionaria è uguale alla varianza della media campionaria. ❑ ❑

5. La varianza della media campionaria di n osservazioni i.i.d. tende a zero per $n \to \infty$. ❑ ❑

6. La d.d.p. della media campionaria di dati i.i.d. tende a diventare gaussiana al crescere di n. ❑ ❑

7. La d.d.p. della varianza campionaria di dati i.i.d. gaussiani è gaussiana. ❑ ❑

8. Se si raddoppia la numerosità n del campione, l'ampiezza dell'intervallo di confidenza viene dimezzata. ❑ ❑

9. Lo Standard Error fornisce un indice della precisione della media campionaria come stimatore della media. ❑ ❑

10. Se X e Y hanno entrambe media > 0, allora Cov[X,Y] > 0. ❑ ❑

ESERCITAZIONE 1

Date delle V.C. X e Y congiuntamente gaussiane con E[X] = E[Y] = 0, si considerino le seguenti triplette di varianze e covarianze:

1) $\sigma_X^2 = 9, \sigma_Y^2 = 1, \ \sigma_{XY} = 0$

2) $\sigma_X^2 = 1, \sigma_Y^2 = 9, \ \sigma_{XY} = 0$

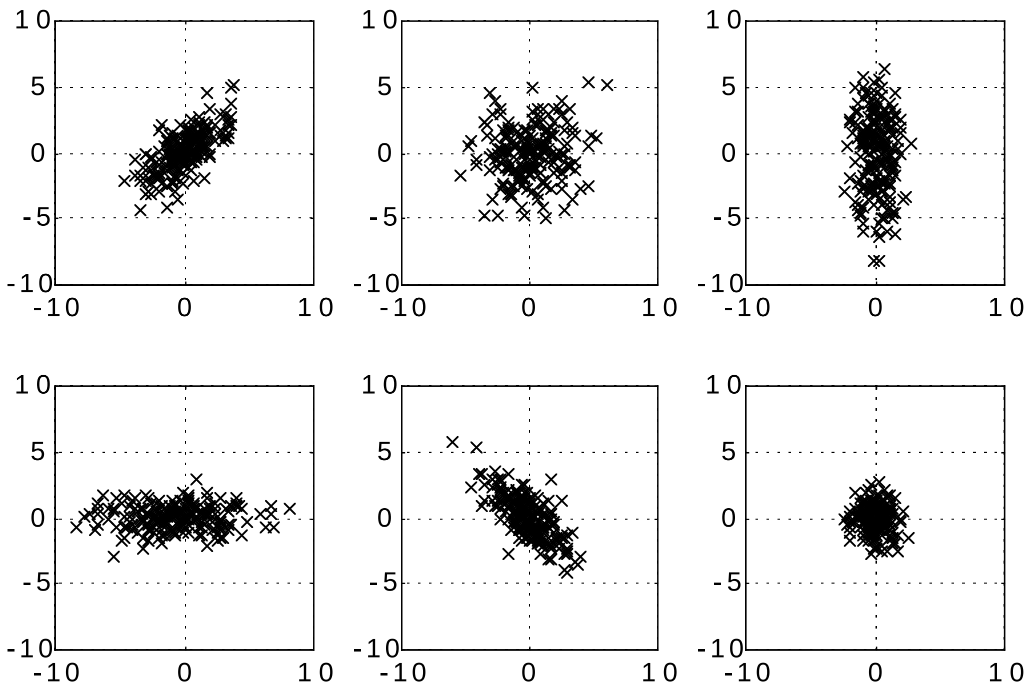
3) $\sigma_X^2 = 3, \sigma_Y^2 = 3, \ \sigma_{XY} = 2.1$

4) $\sigma_X^2 = 3, \sigma_Y^2 = 3, \ \sigma_{XY} = -2.1$

5) $\sigma_X^2 = 1, \sigma_Y^2 = 1, \ \sigma_{XY} = 0$

6) $\sigma_X^2 = 4, \sigma_Y^2 = 4, \ \sigma_{XY} = 0$

Riportare sopra i seguenti scatter plot il numero della corrispondente tripletta $\sigma_X^2$, $\sigma_Y^2$, $\sigma_{XY}$



# RISPOSTE

| | | V | F |
|---|---|---|---|
| 1. | E[X+Y] = E[X]+E[Y]. | ■ | ❑ |
| 2. | Var[X+Y] = Var[X] + Cov[X,Y] + Var[Y]. | ❑ | ■ |
| 3. | Se due V.C. X e Y sono indipendenti, allora sono anche incorrelate. | ■ | ❑ |
| 4. | La media della varianza campionaria è uguale alla varianza della media campionaria. | ❑ | ■ |
| 5. | La varianza della media campionaria di n osservazioni i.i.d. tende a zero per $n \to \infty$. | ■ | ❑ |
| 6. | La d.d.p. della media campionaria di dati i.i.d. tende a diventare gaussiana al crescere di n. | ■ | ❑ |
| 7. | La d.d.p. della varianza campionaria di dati i.i.d. gaussiani è gaussiana. | ❑ | ■ |
| 8. | Se si raddoppia la numerosità n del campione, l'ampiezza dell'intervallo di confidenza viene dimezzata. | ❑ | ■ |
| 9. | Lo Standard Error fornisce un indice della precisione della media campionaria come stimatore della media. | ■ | ❑ |
| 10. | Se X e Y hanno entrambe media > 0, allora Cov[X,Y] > 0. | ❑ | ■ |

# SOLUZIONE ESERCITAZIONE 1