Probabilità e Processi Stocastici (455AA)

Lezione 8

Dario Trevisan – *https://web.dm.unipi.it/trevisan*

15/10/2025

Varianza e deviazione standard

Varianza

La moda, la mediana ed il valore medio sono tutti indicatori *puntuali*, riassumono tutta la legge con un singolo valore.

Per descrivere in modo più efficace una variabile X si affianca un indicatore della dispersione, ossia della "concentrazione" della sua legge intorno ad un indicatore puntuale.

- Un indicatore di dispersione molto usato è la *deviazione standard* (o scarto quadratico medio), definita come

$$\sigma_X = \sqrt{\mathrm{Var}(X)} \geq 0$$

dove la *varianza* è

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] =: K_{XX}$$

$$\underset{0}{\rotatebox{90}{$\geq$}}$$

L'unità di misura di σ_X è la stessa di X, mentre $\mathrm{Var}(X)$ ha come unità di misura il quadrato dell'unità di X.

Operativamente:

1. si calcola il valor medio di X,

L'unità di misura di σ_X è la stessa di X, mentre $\mathrm{Var}(X)$ ha come unità di misura il quadrato dell'unità di X.

Operativamente:

1. si calcola il valor medio di X,
2. si considera lo scarto (variabile aleatoria centrata)

$$X - \mathbb{E}[X],$$

$$\mathbb{E}[(X - \mathbb{E}[X])] = \mathbb{E}[X] - \mathbb{E}[X]$$
$$= 0$$

L'unità di misura di σ_X è la stessa di X, mentre $\mathrm{Var}(X)$ ha come unità di misura il quadrato dell'unità di X.

Operativamente:

1. si calcola il valor medio di X,
2. si considera lo scarto (variabile aleatoria centrata)

$$X - \mathbb{E}[X],$$

3. se ne prende il quadrato e il valor medio

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right].$$

L'unità di misura di σ_X è la stessa di X, mentre $\operatorname{Var}(X)$ ha come unità di misura il quadrato dell'unità di X.

Operativamente:

1. si calcola il valor medio di X,
2. si considera lo scarto (variabile aleatoria centrata)
$$X - \mathbb{E}[X],$$
3. se ne prende il quadrato e il valor medio
$$\operatorname{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right].$$
4. Per la deviazione standard si prende la radice quadrata
$$\sigma_X = \sqrt{\mathbb{E}\left[(X - \mathbb{E}[X])^2\right]}.$$

$\mathbb{E}[g(X)] = \int g(x)\, p(X=x)\, dx$

$$\text{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right].$$

Partendo dalla densità di X e usando $g(x) = (x - \mathbb{E}[X])^2$, si trova la formula

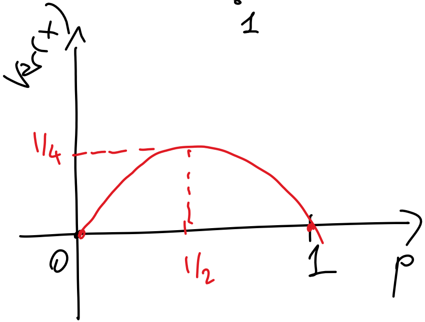
$$\text{Var}(X) = \begin{cases} \sum_{x \in \mathbb{R}} (x - \mathbb{E}[X])^2 P(X = x) & \text{se } X \text{ ha densità discreta} \\ \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 p(X = x) dx & \text{se } X \text{ ha densità continua.} \end{cases}$$

Esempi

$X \in \{0,1\}$ $\qquad \underline{\text{Bernoulli}}\,(p)$ $\qquad p = P(X=1) = \mathbb{E}[X]$

$$\mathrm{Var}(X) = \sum_{x=0}^{1} (x-p)^2 P(X=x) = p^2 \cdot (1-p) + (1-p)^2 p$$

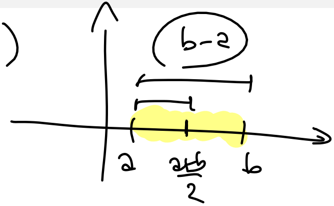
$$= p(1-p)\,[\underbrace{p + (1-p)}_{1}] = p(1-p) \geq 0$$



$$\sigma_X = \sqrt{p(1-p)}$$

Esempi

X uniforme continua su (a, b)



$a = 100$

$b = 101$

$$\frac{a+b}{2} = \frac{201}{2} = 100{,}5$$

$$\mathrm{Var}(X) = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \frac{1}{b-a}\, dx = \frac{1}{b-a} \int_a^b \underbrace{\left(x - \frac{a+b}{2}\right)^2}_{\downarrow} dx$$

$$= \frac{1}{b-a} \left. \frac{\left(x - \frac{a+b}{2}\right)^3}{3} \right|_a^b = \frac{\left(\frac{b-a}{2}\right)^3 - \left(\frac{a-b}{2}\right)^3}{3(b-a)}$$

$$= \frac{\cancel{2}(b-a)^{\cancel{3}\,2}}{3 \cdot \cancel{8}\,4} \frac{1}{\cancel{b-a}} = \frac{1}{12}(b-a)^2$$

Espressione alternativa per la varianza

Vale l'indentità

momento secondo

$$\mathrm{Var}\,(X) = \mathbb{E}\left[X^2\right] - (\mathbb{E}\,[X])^2.$$

Dimostrazione

$$\mathbb{E}[(X-\mathbb{E}[X])^2] = \mathbb{E}[X^2 + (\mathbb{E}[X])^2 - 2X\mathbb{E}[X]]$$

$$= \mathbb{E}[X^2] + (\mathbb{E}[X])^2 - 2\mathbb{E}[X]\mathbb{E}[X]$$

$$= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \geq 0$$

Un altro esempio

Esponenziale $\quad \mathcal{E}xp(\lambda) \quad \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$

$$\mathbb{E}[X] = \int_0^{+\infty} x \lambda e^{-\lambda x}\, dx = [y = \lambda x] = \int_0^{+\infty} y e^{-y} \frac{dy}{\lambda}$$

$$dy = \lambda dx$$

$$= \frac{1}{\lambda} \underbrace{\int_0^{+\infty} y e^{-y} dy}_{\text{Integro per parti}} = \cancel{\frac{1}{\lambda}\, y \cdot (-e^{-y}) \Big|_0^{+\infty}} - \frac{1}{\lambda} \int_0^{+\infty} 1 \cdot (-e^{-y})\, dy$$

$$= \frac{1}{\lambda} \int_0^{+\infty} e^{-y} dy = \frac{1}{\lambda}$$

mediana $= \dfrac{\log(2)}{\lambda}$

Un altro esempio

$$\mathbb{E}[X^2] = \int_0^{+\infty} x^2 \lambda e^{-\lambda x}\, dx = [y = \lambda x]$$

$$= \int_0^{+\infty} \frac{y^2}{\lambda^2} e^{-y}\, dy = \frac{1}{\lambda^2}\int_0^{+\infty} y^2 \underbrace{e^{-y}}_{(-e^{-y})'}\, dy$$

$$= \frac{1}{\lambda^2}\, \cancel{y^2(-e^{-y})\Big|_0^{+\infty}} + \frac{2}{\lambda^2} \underbrace{\int_0^{+\infty} y\, e^{-y}\, dy}_{= 1}$$

$$= \frac{2}{\lambda^2}$$

$$\boxed{\mathbb{E}[X] = \frac{1}{\lambda} \qquad \sigma_X = \frac{1}{\lambda}}$$

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

Diseguaglianza di Chebyshev

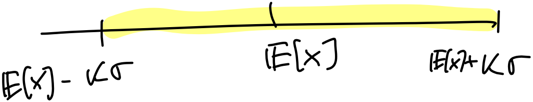
Sia $X \in \mathbb{R}$ una variabile aleatoria. Allora per ogni costante $k > 0$, si ha

$$P(|X - \mathbb{E}[X]| > k) \leq \frac{\mathrm{Var}(X)}{k^2},$$

- o, equivalentemente, per ogni $k \geq 1$,

$$P(\mathbb{E}[X] - k\sigma_X \leq X \leq \mathbb{E}[X] + k\sigma_X) \geq 1 - \frac{1}{k^2}.$$

$k=2 \to \frac{3}{4}$



$k=3 \to \frac{8}{9}$

Diseguaglianza di Chebyshev

Sia $X \in \mathbb{R}$ una variabile aleatoria. Allora per ogni costante $k > 0$, si ha

$$\boxed{P(|X - \mathbb{E}[X]| > k) \leq \frac{\mathrm{Var}(X)}{k^2},}$$

- o, equivalentemente, per ogni $k \geq 1$,

$$P(\mathbb{E}[X] - k\sigma_X \leq X \leq \mathbb{E}[X] + k\sigma_X) \geq 1 - \frac{1}{k^2}.$$

- Una conseguenza è che, se $\sigma_X = 0$ la variabile X è constante con probabilità 1

Diseguaglianza di Chebyshev

Sia $X \in \mathbb{R}$ una variabile aleatoria. Allora per ogni costante $k > 0$, si ha

$$P(|X - \mathbb{E}[X]| > k) \leq \frac{\mathrm{Var}(X)}{k^2},$$

- o, equivalentemente, per ogni $k \geq 1$,

$$P(\mathbb{E}[X] - k\sigma_X \leq X \leq \mathbb{E}[X] + k\sigma_X) \geq 1 - \frac{1}{k^2}.$$

- Una conseguenza è che, se $\sigma_X = 0$ la variabile X è constante con probabilità 1
- Informalmente scriviamo

$$X \approx \mathbb{E}[X] \pm \sigma_X$$

Dimostrazione di Chebychev

Markov $\Rightarrow P(Z > c) \leq \frac{\mathbb{E}[Z]}{c}$ $Z \geq 0$

$Z = (X - \mathbb{E}[X])^2$ $\mathbb{E}[Z] = Var(X)$

$c = k^2$

$$P\left(\sqrt{(X-\mathbb{E}[X])^2 > k^2}\right) \leq \frac{Var(X)}{k^2}$$

$$\parallel$$

$$P(|X - \mathbb{E}[X]| > k)$$

Standardizzazione

Data X, la variabile $Y = X - \mathbb{E}[X]$ è **centrata**, ossia $\mathbb{E}[Y] = 0$:

- La **standardizzazione** X è la variabile

$$\hat{X} = \frac{X - \mathbb{E}[X]}{\sigma_X},$$

che è centrata e ha deviazione standard $\sigma_{\hat{X}} = 1$.

~~Esempi~~

OSS

$$\mathrm{Var}(aX+b) =$$

$$= \mathbb{E}\left[\left(aX+b-\mathbb{E}[aX+b]\right)^2\right]$$

$$= \mathbb{E}\left[\left(aX+b-a\,\mathbb{E}[X]-b\right)^2\right]$$

$$= \mathbb{E}\left[a^2\left(X-\mathbb{E}[X]\right)^2\right] = a^2\,\mathrm{Var}(X)$$

$$\mathrm{Var}\left(\frac{X-\mathbb{E}[X]}{\sigma_X}\right) = \frac{1}{\sigma_X^2}\,\mathrm{Var}(X) = \frac{\sigma_X^2}{\sigma_X^2} = 1$$

Binomiale $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$

X = "numero di successi in n esperimenti indipendenti"

$X_1, X_2, \dots, X_n \in \{0,1\}$ Bernoulli(p) indipendenti

$$X = X_1 + X_2 + \dots + X_n$$

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \dots + \mathbb{E}[X_n] = \underbrace{p + p + \dots + p}_{n \text{ volte}} = np$$

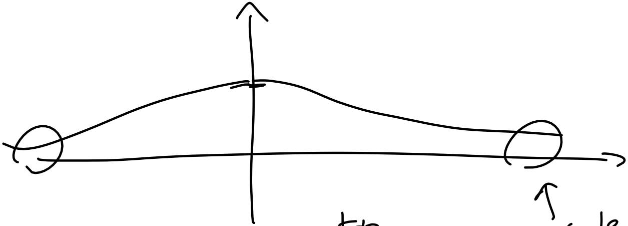
$$\mathrm{Var}(X) = \mathrm{Var}(X_1 + X_2 + \dots + X_n) =$$ OK

$$= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \dots + \mathrm{Var}(X_n) = np(1-p)$$

$$\sigma_X = \sigma_{X_1 + X_2 + \dots + X_n} = n\,\sigma_{X_1} = n\sqrt{p(1-p)}$$ No

Esempio X Cauchy v.a. continua con densità

$$p(X=x) = \frac{1}{\pi}\frac{1}{(1+x^2)} \quad x \in \mathbb{R}$$



code sono "pesanti"

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} \frac{1}{\pi} \frac{x}{1+x^2}\,dx \underset{?}{=} 0$$

$$\mathbb{E}[|X|] = \int_{-\infty}^{+\infty} \frac{1}{\pi} \frac{|x|}{1+x^2}\,dx = +\infty \quad \text{NON integrabile}$$

Esempio $\quad p(Z=z) \propto \frac{1}{z^3} \quad \underline{z>1}$

Calcolare $\frac{1}{\int_1^{+\infty} \frac{1}{z^3} dz}$

· CDF, quantile

$$\mathbb{E}[Z] = \int_1^{+\infty} \cancel{z}\, c \frac{1}{z^{\cancel{3}2}} dz = c \int_1^{+\infty} \frac{1}{z^2} dz < +\infty \quad \text{ok}$$

$$\mathbb{E}[Z^2] = \int_1^{+\infty} \cancel{z^2}\, c \cdot \frac{1}{z^{\cancel{3}}} dz = c \int_1^{+\infty} \frac{1}{z} dz = +\infty$$

$(\mathrm{Var}(Z) = +\infty)$

Covarianza

La dispersione nel caso vettoriale

- L'estensione del valor medio al caso vettoriale $X \in \mathbb{R}^d$ è immediata:

$$\mathbb{E}[X] = (\mathbb{E}[X_1], \mathbb{E}[X_2], \ldots, \mathbb{E}[X_d]).$$

La dispersione nel caso vettoriale

- L'estensione del valor medio al caso vettoriale $X \in \mathbb{R}^d$ è immediata:

$$\mathbb{E}[X] = (\mathbb{E}[X_1], \mathbb{E}[X_2], \ldots, \mathbb{E}[X_d]).$$

- Estensione della varianza? Le varianze delle componenti non sono "sufficienti" a descrivere la dispersione della legge.

Esempio

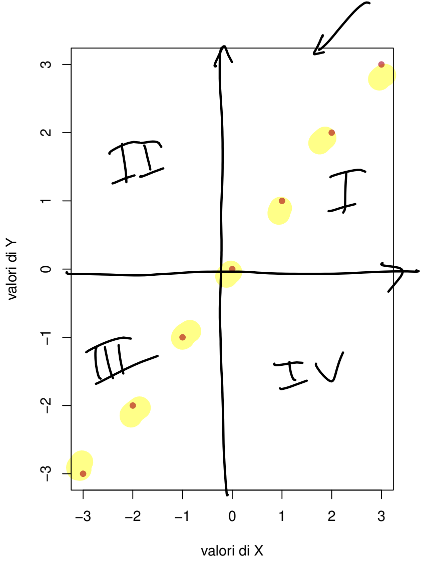
Si considerino due variabili X, Y uniformi discrete sui valori $\{-3, -2, -1, 0, 1, 2, 3\}$ (in modo che siano già centrate). La deviazione standard risulta $\sigma_X = \sigma_Y = 2$.

Ma non abbiamo indicazioni sulla la “dispersione” della variabile congiunta (X, Y):

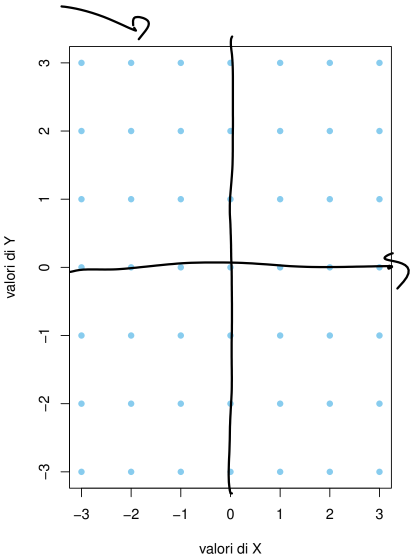
- potrebbe essere noto che $X = Y$, e quindi la densità discreta della variabile congiunta è “concentrata” sulla diagonale principale;

Ma non abbiamo indicazioni sulla la “dispersione” della variabile congiunta (X, Y):

- potrebbe essere noto che $X = Y$, e quindi la densità discreta della variabile congiunta è “concentrata” sulla diagonale principale;
- oppure le due variabili potrebbero essere indipendenti, e quindi la densità è “diffusa” su tutte le possibili coppie di valori.



X = Y



X indip da Y

La covarianza: definizione e proprietà

Date due variabili aleatorie reali X, Y, si definisce la covarianza come la quantità (reale)

$$\mathrm{Cov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right].$$

- La covarianza è una estensione della varianza

$$\mathrm{Cov}(X, X) = \mathbb{E}\left[(X - \mathbb{E}[X])(X - \mathbb{E}[X])\right]$$
$$= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathrm{Var}(X)$$

La covarianza: definizione e proprietà

Date due variabili aleatorie reali X, Y, si definisce la covarianza come la quantità (reale)

$$\mathrm{Cov}(X,Y) = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right].$$

- La covarianza è una estensione della varianza
- è una funzione *bilineare* (ossia separatamente lineare) di X e Y

$$\mathrm{Cov}(X, aY + Z) = a\,\mathrm{Cov}(X,Y) + \mathrm{Cov}(X,Z)$$

La covarianza: definizione e proprietà

Date due variabili aleatorie reali X, Y, si definisce la covarianza come la quantità (reale)

$$\operatorname{Cov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right].$$

- La covarianza è una estensione della varianza
- è una funzione *bilineare* (ossia separatamente lineare) di X e Y
- A volte si indica anche con K_{XY} c_{XY}

Date variabili aleatorie reali X, Y, Z e una costante $a > 0$, valgono le seguenti proprietà:

1. $\mathrm{Cov}\,(X, X) = \mathrm{Var}\,(X)$

Date variabili aleatorie reali X, Y, Z e una costante $a > 0$, valgono le seguenti proprietà:

1. $\text{Cov}\,(X, X) = \text{Var}\,(X)$
2. (simmetria) $\text{Cov}\,(X, Y) = \text{Cov}\,(Y, X)$

Date variabili aleatorie reali X, Y, Z e una costante $a > 0$, valgono le seguenti proprietà:

1. $\mathrm{Cov}\,(X, X) = \mathrm{Var}\,(X)$
2. (simmetria) $\mathrm{Cov}\,(X, Y) = \mathrm{Cov}\,(Y, X)$
3. (bilinearità) $\mathrm{Cov}\,(X + Z, Y) = \mathrm{Cov}\,(X, Y) + \mathrm{Cov}\,(Z, Y)$ e similmente $\mathrm{Cov}\,(X, Y + Z) = \mathrm{Cov}\,(X, Y) + \mathrm{Cov}\,(X, Z)$. Inoltre $\mathrm{Cov}\,(aX, Y) = a\,\mathrm{Cov}\,(X, Y) = \mathrm{Cov}\,(X, aY)$.

Date variabili aleatorie reali X, Y, Z e una costante $a > 0$, valgono le seguenti proprietà:

1. $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$
2. (simmetria) $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
3. (bilinearità) $\mathrm{Cov}(X + Z, Y) = \mathrm{Cov}(X, Y) + \mathrm{Cov}(Z, Y)$ e similmente $\mathrm{Cov}(X, Y + Z) = \mathrm{Cov}(X, Y) + \mathrm{Cov}(X, Z)$. Inoltre $\mathrm{Cov}(aX, Y) = a\,\mathrm{Cov}(X, Y) = \mathrm{Cov}(X, aY)$.
4. (varianza della somma) $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y)$.

$$\mathrm{Var}(X+Y) = \mathrm{Cov}(X+Y, X+Y) =$$
$$= \mathrm{Cov}(X+Y, X) + \mathrm{Cov}(X+Y, Y)$$
$$= \mathrm{Cov}(X, X) + \mathrm{Cov}(Y, X) + \mathrm{Cov}(X, Y) + \mathrm{Cov}(Y, Y)$$

Date variabili aleatorie reali X, Y, Z e una costante $a > 0$, valgono le seguenti proprietà:

1. $\operatorname{Cov}(X, X) = \operatorname{Var}(X)$
2. (simmetria) $\operatorname{Cov}(X, Y) = \operatorname{Cov}(Y, X)$
3. (bilinearità) $\operatorname{Cov}(X + Z, Y) = \operatorname{Cov}(X, Y) + \operatorname{Cov}(Z, Y)$ e similmente $\operatorname{Cov}(X, Y + Z) = \operatorname{Cov}(X, Y) + \operatorname{Cov}(X, Z)$. Inoltre $\operatorname{Cov}(aX, Y) = a \operatorname{Cov}(X, Y) = \operatorname{Cov}(X, aY)$.
4. (varianza della somma) $\operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2 \operatorname{Cov}(X, Y)$.
5. (formula alternativa) $\operatorname{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$.

Dimostrazione

Indipendenza e covarianza

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{cov}(X,Y)$$

Se due variabili reali X, Y sono indipendenti (rispetto ad una informazione I), allora sono *non correlate*, ossia

$$\mathrm{Cov}\,(X, Y) = 0.$$

- In particolare,

$$\mathrm{Var}\,(X+Y) = \mathrm{Var}\,(X) + \mathrm{Var}\,(Y)\,.$$

$$\mathrm{Cov}(X,Y) = \mathbb{E}\left[(X-\mathbb{E}[X])\,(Y-\mathbb{E}[Y])\right]$$

$$= \mathbb{E}[g(X) \cdot f(Y)] = \mathbb{E}[g(X)]\,\mathbb{E}[f(Y)]$$

$$= 0 \cdot 0 = 0$$

Dimostrazione

Questo fatto segue dalla formula alternativa per la covarianza

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$$

- e la proprietà vista del valor medio del prodotto di variabili indipendenti,

$$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y].$$

X, Y indip $\Longrightarrow$ $\mathrm{Cov}(X, Y) = 0$

$\not\Longleftarrow$ NON in generale

Segno della covarianza

- Se X, Y sono indipendenti, allora sono *non correlate*, ossia $\operatorname{Cov}(X, Y) = 0$

Segno della covarianza

- Se X, Y sono indipendenti, allora sono *non correlate*, ossia $\operatorname{Cov}(X, Y) = 0$
- Il viceversa non è vero in generale: se $\operatorname{Cov}(X, Y) = 0$, non necessariamente X, Y sono indipedenti.

Segno della covarianza

- Se X, Y sono indipendenti, allora sono *non correlate*, ossia $\mathrm{Cov}\,(X, Y) = 0$
- Il viceversa non è vero in generale: se $\mathrm{Cov}\,(X, Y) = 0$, non necessariamente X, Y sono indipedenti.
- Se $\mathrm{Cov}\,(X, Y) > 0$, X, Y sono *positivamente* correlate

Segno della covarianza

- Se X, Y sono indipendenti, allora sono *non correlate*, ossia $\mathrm{Cov}\,(X, Y) = 0$
- Il viceversa non è vero in generale: se $\mathrm{Cov}\,(X, Y) = 0$, non necessariamente X, Y sono indipedenti.
- Se $\mathrm{Cov}\,(X, Y) > 0$, X, Y sono *positivamente* correlate
- Se $\mathrm{Cov}\,(X, Y) < 0$, sono *negativamente* correlate.

Covarianza di variabili Bernoulli

Siano $X \in \{0,1\}$, indicatrice dell'evento A, $Y \in \{0,1\}$ indicatrice dell'evento B.

- Allora $XY \in \{0,1\}$ è indicatrice di "A e B" e

$$\mathrm{Cov}\,(X,Y) = \mathbb{E}\,[XY] - \mathbb{E}\,[X]\,\mathbb{E}\,[Y] = P(A \text{ e } B) - P(A)P(B).$$

marginali

Covarianza di variabili Bernoulli

Siano $X \in \{0,1\}$, indicatrice dell'evento A, $Y \in \{0,1\}$ indicatrice dell'evento B.

- Allora $XY \in \{0,1\}$ è indicatrice di "A e B" e

$$\mathrm{Cov}\,(X,Y) = \mathbb{E}\,[XY] - \mathbb{E}\,[X]\,\mathbb{E}\,[Y] = P(A \text{ e } B) - P(A)P(B).$$

- X e Y sono positivamente correlate se e solo se

$$P(A \text{ e } B) - P(A)P(B) > 0, \quad \text{ossia} \quad \frac{P(A \text{ e } B)}{P(A)P(B)} > 1,$$

Covarianza di variabili Bernoulli

Siano $X \in \{0, 1\}$, indicatrice dell'evento A, $Y \in \{0, 1\}$ indicatrice dell'evento B.

- Allora $XY \in \{0, 1\}$ è indicatrice di "A e B" e

$$\mathrm{Cov}\,(X, Y) = \mathbb{E}\,[XY] - \mathbb{E}\,[X]\,\mathbb{E}\,[Y] = P(A \text{ e } B) - P(A)P(B).$$

- X e Y sono positivamente correlate se e solo se

$$P(A \text{ e } B) - P(A)P(B) > 0, \quad \text{ossia} \quad \frac{P(A \text{ e } B)}{P(A)P(B)} > 1,$$

- ossia $P(A|B) > P(A)$ (B è una informazione a favore di A)

Covarianza di variabili Bernoulli

Siano $X \in \{0,1\}$, indicatrice dell'evento A, $Y \in \{0,1\}$ indicatrice dell'evento B.

- Allora $XY \in \{0,1\}$ è indicatrice di "A e B" e

$$\mathrm{Cov}\,(X,Y) = \mathbb{E}\,[XY] - \mathbb{E}\,[X]\,\mathbb{E}\,[Y] = P(A \text{ e } B) - P(A)P(B).$$

- X e Y sono positivamente correlate se e solo se

$$P(A \text{ e } B) - P(A)P(B) > 0, \quad \text{ossia} \quad \frac{P(A \text{ e } B)}{P(A)P(B)} > 1,$$

- ossia $P(A|B) > P(A)$ (B è una informazione a favore di A)
- sono negativamente correate se e solo se il rapporto è < 1,

Covarianza di variabili Bernoulli

Siano $X \in \{0, 1\}$, indicatrice dell'evento A, $Y \in \{0, 1\}$ indicatrice dell'evento B.

- Allora $XY \in \{0, 1\}$ è indicatrice di "A e B" e

$$\mathrm{Cov}\,(X, Y) = \mathbb{E}\,[XY] - \mathbb{E}\,[X]\,\mathbb{E}\,[Y] = P(A \text{ e } B) - P(A)P(B).$$

- X e Y sono positivamente correlate se e solo se

$$P(A \text{ e } B) - P(A)P(B) > 0, \quad \text{ossia} \quad \frac{P(A \text{ e } B)}{P(A)P(B)} > 1,$$

- ossia $P(A|B) > P(A)$ (B è una informazione a favore di A)
- sono negativamente correate se e solo se il rapporto è < 1,
- sono non correlate se e solo se il rapporto vale 1 **ossia** A, B sono indipendenti.

Valor medio e varianza di variabili Binomiali

$$X = X_1 + \dots + X_n$$

$$\mathbb{E}[X] = np$$

$$\mathrm{Var}(X) = np(1-p) \qquad \sigma_X = \sqrt{np(1-p)}$$

es $p = \frac{1}{2}$ $\quad \boxed{X \simeq \frac{n}{2} \pm \frac{\sqrt{n}}{2}} \quad \sqrt{n} \ll n$

Intepretazione del segno della covarianza

- Come intepretare il segno della covarianza nel caso di variabili generali?

Intepretazione del segno della covarianza

- Come intepretare il segno della covarianza nel caso di variabili generali?
- Vedremo una spiegazione precisa trattando la regressione lineare.

Intepretazione del segno della covarianza

- Come intepretare il segno della covarianza nel caso di variabili generali?
- Vedremo una spiegazione precisa trattando la regressione lineare.
- X e Y sono positivamente correlate se, sapendo che $X > \mathbb{E}[X]$ allora è più probabile che sia anche $Y > \mathbb{E}[Y]$ (e similmente, sapendo $X \leq \mathbb{E}[X]$, è più probabile che sia $Y \leq \mathbb{E}[Y]$).

Intepretazione del segno della covarianza

- Come intepretare il segno della covarianza nel caso di variabili generali?
- Vedremo una spiegazione precisa trattando la regressione lineare.
- X e Y sono positivamente correlate se, sapendo che $X > \mathbb{E}[X]$ allora è più probabile che sia anche $Y > \mathbb{E}[Y]$ (e similmente, sapendo $X \leq \mathbb{E}[X]$, è più probabile che sia $Y \leq \mathbb{E}[Y]$).
- Graficamente, la densità congiunta tra (X, Y) è più concentrata nel primo e terzo quadrante cartesiano, avendo posto l'origine nel vettore dei valor medi

Intepretazione del segno della covarianza

- Come intepretare il segno della covarianza nel caso di variabili generali?
- Vedremo una spiegazione precisa trattando la regressione lineare.
- X e Y sono positivamente correlate se, sapendo che $X > \mathbb{E}[X]$ allora è più probabile che sia anche $Y > \mathbb{E}[Y]$ (e similmente, sapendo $X \leq \mathbb{E}[X]$, è più probabile che sia $Y \leq \mathbb{E}[Y]$).
- Graficamente, la densità congiunta tra (X, Y) è più concentrata nel primo e terzo quadrante cartesiano, avendo posto l'origine nel vettore dei valor medi
- La correlazione negativa indica che è più concentrata nel secondo e quarto quadrante.

Matrice delle covarianze

$$\begin{array}{cc} & d \\ d & \begin{pmatrix} Var(X_1) & & & \\ & Var(X_2) & & \\ & & \ddots & \\ & & & Var(X_d) \end{pmatrix} \end{array}$$

Dato un vettore aleatorio $X = (X_1, \ldots, X_d) \in \mathbb{R}^d$, si definisce la matrice delle covarianze di X la matrice di numeri reali $\Sigma_X \in \mathbb{R}^{d \times d}$

$$(\Sigma_X)_{i,j} = \mathrm{Cov}\,(X_i, X_j) = \mathbb{E}\,[X_i X_j] - \mathbb{E}\,[X_i]\,\mathbb{E}\,[X_j] \quad \text{per } i, j \in \{1, \ldots, d\}.$$

- Notazioni alternative: $\mathrm{Var}\,(X)$, K_{XX} o Q_X.

Proprietà

- La matrice delle covarianze è simmetrica $\Sigma_X = \Sigma_X^T$, dove T indica l'operazione di trasposizione.

$$Cov(x_i, x_j) = Cov(x_j, x_i) \Rightarrow \Sigma_X = \Sigma_X^T$$

Proprietà

- La matrice delle covarianze è simmetrica $\Sigma_X = \Sigma_X^T$, dove T indica l'operazione di trasposizione.
- (trasformazioni affini) Sia $X \in \mathbb{R}^d$ e

$$Y = AX + b \quad \text{ossia} \quad Y_i = \sum_{j=1}^{d} A_{ij} X_j + b_i,$$

dove $A \in \mathbb{R}^{k \times d}$ è una matrice e $b \in \mathbb{R}^k$ è un vettore (costanti). Vale

$$\Sigma_{AX+b} = A \Sigma_X A^T.$$

Proprietà

- La matrice delle covarianze è simmetrica $\Sigma_X = \Sigma_X^T$, dove T indica l'operazione di trasposizione.
- (trasformazioni affini) Sia $X \in \mathbb{R}^d$ e

$$. \quad Y = AX + b \quad \text{ossia} \quad Y_i = \sum_{j=1}^{d} A_{ij} X_j + b_i,$$

dove $A \in \mathbb{R}^{k \times d}$ è una matrice e $b \in \mathbb{R}^k$ è un vettore (costanti). Vale

$$\Sigma_{AX+b} = A \Sigma_X A^T.$$

- In particolare, se $k = 1$ e $A = v^T$, con $v \in \mathbb{R}^d$, si ottiene che

$$\operatorname{Var}(v \cdot X) = \Sigma_{v \cdot X} = v^T \Sigma_X v,$$

Alunedì

ossia Σ_X è (semi-)definita positiva.

Dimostrazione

Conseguenze: il coefficiente di correlazione

- Nel caso $d = 2$, scrivendo (X, Y) per la variabile congiunta di due variabili reali X, Y, la matrice delle covarianze è esplicitamente

$$\Sigma_{(X,Y)} = \begin{pmatrix} \mathrm{Var}\,(X) & \mathrm{Cov}\,(X, Y) \\ \mathrm{Cov}\,(X, Y) & \mathrm{Var}\,(Y) \end{pmatrix}.$$

Conseguenze: il coefficiente di correlazione

- Nel caso $d = 2$, scrivendo (X, Y) per la variabile congiunta di due variabili reali X, Y, la matrice delle covarianze è esplicitamente

$$\Sigma_{(X,Y)} = \begin{pmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X, Y) \\ \mathrm{Cov}(X, Y) & \mathrm{Var}(Y) \end{pmatrix}.$$

- Essendo semidefinita positiva, il suo determinante è positivo (o nullo):

$$\det(\Sigma_{(X,Y)}) = \mathrm{Var}(X)\,\mathrm{Var}(Y) - (\mathrm{Cov}(X, Y))^2 \geq 0,$$

Conseguenze: il coefficiente di correlazione

- Nel caso $d = 2$, scrivendo (X, Y) per la variabile congiunta di due variabili reali X, Y, la matrice delle covarianze è esplicitamente

$$\Sigma_{(X,Y)} = \begin{pmatrix} \mathrm{Var}\,(X) & \mathrm{Cov}\,(X, Y) \\ \mathrm{Cov}\,(X, Y) & \mathrm{Var}\,(Y) \end{pmatrix}.$$

- Essendo semidefinita positiva, il suo determinante è positivo (o nullo):

$$\det(\Sigma_{(X,Y)}) = \mathrm{Var}\,(X)\,\mathrm{Var}\,(Y) - (\mathrm{Cov}\,(X, Y))^2 \geq 0,$$

- ossia, dopo alcune operazioni elementari

$$\rho_{XY} := \frac{\mathrm{Cov}\,(X, Y)}{\sigma_X \sigma_Y} \in [-1, 1].$$

Conseguenze: il coefficiente di correlazione

- Nel caso $d = 2$, scrivendo (X, Y) per la variabile congiunta di due variabili reali X, Y, la matrice delle covarianze è esplicitamente
$$\Sigma_{(X,Y)} = \begin{pmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(X,Y) & \mathrm{Var}(Y) \end{pmatrix}.$$
- Essendo semidefinita positiva, il suo determinante è positivo (o nullo):
$$\det(\Sigma_{(X,Y)}) = \mathrm{Var}(X)\,\mathrm{Var}(Y) - (\mathrm{Cov}(X,Y))^2 \geq 0,$$
- ossia, dopo alcune operazioni elementari
$$\rho_{XY} := \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} \in [-1, 1].$$
- ρ_{XY} è il *coefficiente di correlazione* (o indice di correlazione di Pearson).

Standardizzazione nel caso vettoriale

Il teorema spettrale permette di decomporre

$$\Sigma_X = U^T D U,$$

dove $U \in \mathbb{R}^{d \times d}$, è ortogonale $U^T U = Id$ e D è diagonale (gli autovalori di Σ_X).

- La trasformazione UX, corrisponde ad un cambio di coordinate e trasforma la covarianza

  $$\Sigma_{UX} = U \Sigma_X U^T = D$$

  ossia le componenti di UX sono a due a due non correlate.

- Se D è invertibile si può definire una *standardizzazione* di X

$$\hat{X} = \sqrt{D}^{-1} U(X - \mathbb{E}[X])$$

dove $\sqrt{D}$ è la matrice diagonale con entrate date dalla radice quadrata di quelle di D.

- Se D è invertibile si può definire una *standardizzazione* di X

$$\hat{X} = \sqrt{D}^{-1} U(X - \mathbb{E}[X])$$

dove $\sqrt{D}$ è la matrice diagonale con entrate date dalla radice quadrata di quelle di D.

- Usando le proprietà del vettore delle medie e della varianza, si ha

$$\mathbb{E}\left[\hat{X}\right] = 0 \in \mathbb{R}^d \quad \text{e} \quad \Sigma_{\hat{X}} = Id.$$

- Se D è invertibile si può definire una *standardizzazione* di X

$$\hat{X} = \sqrt{D}^{-1} U(X - \mathbb{E}[X])$$

dove $\sqrt{D}$ è la matrice diagonale con entrate date dalla radice quadrata di quelle di D.

- Usando le proprietà del vettore delle medie e della varianza, si ha

$$\mathbb{E}\left[\hat{X}\right] = 0 \in \mathbb{R}^d \quad \text{e} \quad \Sigma_{\hat{X}} = Id.$$

- Attenzione! quando si passa alle osservazioni di un campione, la *standardizzazione* si riferisce all'operazione effettuata sulle marginali (comando `scale()` in R).