COVARIANZA

Def: La covarianza di due v.a. X ed Y è definita come:

$$\text{Cov}(X,Y) = E[(X - E[X])\cdot(Y - E[Y])] = E[X\cdot Y] - E[X]\cdot E[Y]$$

- Se $\text{Cov}(X,Y) > 0$ allora X ed Y sono positivamente correlate.
- Se $\text{Cov}(X,Y) < 0$ allora X ed Y sono negativamente correlate.
- Se $\text{Cov}(X,Y) = 0$ allora X ed Y sono incorrelate.

Dim: sul quaderno.

Oss: Scopriamo la Covarianza a partire dal calcolo della Varianza di $g(x,y) = X + Y$:

Ricordando che: $\text{Var}(X) = E[(X - E[X])^2]$

- $$\text{Var}(X+Y) = E[(X+Y-E[X+Y])^2]$$

Uso la linearità della $E[\cdot]$ interna.

$$= E[(X+Y-E[X]-E[Y])^2]$$

$$= E[((X-E[X])+(Y-E[Y]))^2]$$

Uso la linearità della $E[\cdot]$ esterna.

$$= E[(X-E[X])^2 + 2(X-E[X])(Y-E[Y]) + (Y-E[Y])^2]$$

$$= E[(X-E[X])^2] + 2E[(X-E[X])(Y-E[Y])] + E[(Y-E[Y])^2]$$

$$= \text{Var}(X) + 2\,\text{Cov}(X,Y) + \text{Var}(Y)$$

Varianza e Correlazione

Corollario: Abbiamo scoperto che la varianza non è lineare:

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\,\text{Cov}(X,Y)$$

A differenza dell'aspettazione che è invece lineare:

$$E[X+Y] = E[X] + E[Y]$$

Ma se X ed Y sono incorrelate, allora

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$$

(X, Y incorrelate $\iff$ $\text{Cov}(X,Y) = 0$)

Studio del Segno della Covarianza

Def: La covarianza si scrive come: $\text{Cov}(X,Y) = E[(X - E[X])\cdot(Y - E[Y])]$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \underbrace{(X - E[X])\cdot(Y - E[Y])}\cdot \overbrace{f_{XY}(x,y)}^{\overset{0}{\wedge\!\shortmid}} \cdot dx\cdot dy$$

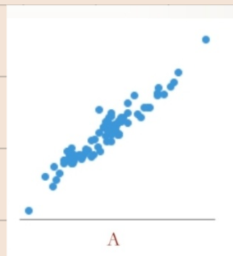
Quando $\text{Cov}(X,Y) \gtreqless 0$? Dipende dal segno degli scarti.

Interpretazione grafici Scatterplot

Se $(X - E[X])$ e $(Y - E[Y])$ hanno lo stesso segno:

$\Rightarrow (X - E[X]) \cdot (Y - E[Y]) \geq 0$.

La $Cov(X,Y) = 0$ quando la funzione si bilancia tra i + e -.



Positiv. Correlato.

Relazione tra Dipendenza e Correlazione

Siano X ed Y v.a., allora:

X,Y INDIPENDENTI ==> X,Y SCORRELATE

se X ed Y sono indipendenti, allora sono incorrelate, quindi:

$\Rightarrow \quad E[X \cdot Y] = E[X] \cdot E[Y]$

$\Rightarrow \quad Var(X+Y) = Var(X) + Var(Y)$

$\Rightarrow \quad Cov(X,Y) = 0$

ma non è vero il contrario: scorrelate $\not\Rightarrow$ indipendenti.

Dim: sul quaderno.

Esempio discreto calcolo Covarianza

X ed Y sono due v.a discrete con funzione di probabilità data dalla tabella di seguito. Calcolare **Cov(X, Y)** e indicare se X e Y *sono correlate.*

Y \ X	0	1	2
-1	1/6	1/6	1/6
1	0	1/2	0

Dati: $g(x,y) = X \cdot Y$

1) Calcolare l'aspettazione del prodotto $E[X \cdot Y]$:

$$E[g(x,y)] = E[X \cdot Y] = \sum_{\substack{x \in \{0,1,2\} \\ y \in \{-1,1\}}} x \cdot y \cdot p_{xy}(x,y)$$

$$= (-1) \cdot 1 \cdot \frac{1}{6} + 1 \cdot 1 \cdot \frac{1}{2} + 2 \cdot (-1) \cdot \frac{1}{6}$$

$$= -\frac{1}{6} - \frac{2}{6} + \frac{1}{2} = 0.$$

Formula conveniente covarianza:

$$\boxed{Cov(X,Y) = E[X \cdot Y] - E[X] \cdot E[Y]}$$

$$\boxed{E[g(x,y)] = \sum_{(x,y) \in S_{x,y}} g(x,y) \cdot p_{xy}(x,y)}$$

2) Calcolare le marginali di $E[X]$ e $E[Y]$:

$$E[X] = \sum_{x \in \{0,1,2\}} x \cdot p_x(x) = 1 \cdot \left(\frac{1}{6} + \frac{1}{2}\right) + 2 \cdot \frac{1}{6} = 1.$$

$$E[Y] = \sum_{y \in \{-1,1\}} y \cdot p(y) = (-1) \cdot \frac{3}{6} + 1 \cdot \frac{1}{2} = 0.$$

$$\boxed{E[X] = \sum_{x \in S_x} x \cdot p_x(x)}$$

3) Calcolare la covarianza $Cov(X,Y)$:

$Cov(X,Y) = 0 - 1 \cdot 0 = 0 \quad \Rightarrow \quad$ X ed Y incorrelate.

Riguardo all'esempio precedemte. Sono X e Y indipendenti?

Con $x = 0$ e $y = 1$ abbiamo:

- $P_{XY}(0,1) = 0 \neq \frac{1}{6} \cdot \frac{1}{2} = P_X(x) \cdot P_Y(y)$

Ricordiamo che X e Y sono indipendenti se:

$P_{X,Y}(x,y) = P_X(a) \cdot P_Y(y) \quad \forall (a,b) \in S_{(a,b)}$

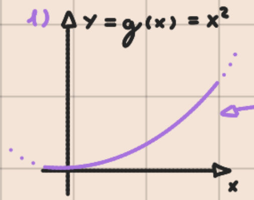
Quindi si tratta di un caso di V.A. **Dipendenti e Incorrelate**.

Nota sulla Dipendenza tra due v.a.

Si può derivare se due V.A. sono dipendenti tra loro semplicemente guardando la funzione ed il grafico:

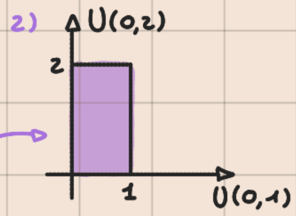
Caso 1) Abbiamo X e $Y = X^2$ si nota come la y è in funzione della x $\Rightarrow$ sono fortemente **dipendenti**.

Caso 2) Abbiamo $X \sim U(0,1)$, $Y \sim U(0,2)$ non sono direttamente in funzione l'una con l'altra $\Rightarrow$ **indipendenti**.



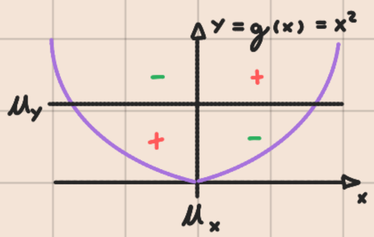
Se conosco la x conosco anche la y

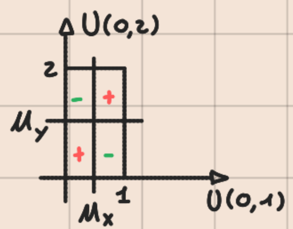
Invece qua no.



Nota sulla Correlazione tra due v.a.

Si denota che in entrambi i casi le variabili X ed Y **sono incorrelate** perché bilanciate tra i + e −.





$\Rightarrow \quad Cov(X,Y) = 0$.

Trasformazioni Lineari

Siano X e Y due V.A. allora:

$$\boxed{Cov(rX + s, tY + v) = r \cdot t \cdot Cov(X,Y)}$$

Dim: sul quaderno.

$$\boxed{Var(aX + bY + c) = a^2 Var(X) + 2ab\, Cov(X,Y) + b^2 \cdot Var(Y)}$$

Dim: sul quaderno.

Correlazione

$$\boxed{\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)} \cdot \sqrt{Var(Y)}}}$$

Il **coefficiente di correlazione** di X e Y V.A. non è altro che la **covarianza normalizzata con gli scarti Q.M.**

Vantaggi rispetto alla covarianza: 1) Se $|\rho(X,Y)| = 1 \Rightarrow X$ e Y si dicono totalmente correlate.

2) $\rho(\alpha X + \beta, \gamma Y + \delta) = \rho(X,Y)$.

Disuguaglianza di Schwarts: $|\rho(X,Y)| \leq 1$.

Oss: $Cov(X,X) = E[(X - E[X]) \cdot (X - E[X])] = E[(X - E[X])^2] = Var(X)$.