

9 Estensione del concetto di valore atteso

Vogliamo ora estendere il concetto di valore atteso di una variabile casuale 1-D al caso n -dimensionale. Ci limitiamo ad analizzare il caso discreto.

DEFINIZIONE: Sia $X = (X_1, \dots, X_n)$ una variabile casuale n -dimensionale con funzione di densità congiunta $f_{X_1, \dots, X_n}(x_1, \dots, x_n)$ e sia $g : \mathbb{R}^n \rightarrow \mathbb{R}$ una funzione definita su X_1, \dots, X_n , cioè $g = g(X_1, \dots, X_n)$ (a sua volta è una variabile casuale).

Definiamo **VALORE ATTESO** di g la quantità:

$$E[g(X_1, \dots, X_n)] = \sum g(x_1, \dots, x_n) f_{X_1, \dots, X_n}(x_1, \dots, x_n)$$

CASI PARTICOLARI

- $g(X_1, \dots, X_n) = X_i \Rightarrow E[g] = E[X_i] = \mu_{X_i}$
- $g(X_1, \dots, X_n) = (X_i - \mu_{X_i})^2$
 $\Rightarrow E[g] = \text{var}[X_i] = \sigma_{X_i}^2$

Limitiamoci al caso bidimensionale ($n = 2$)

- $g(X, Y) = (X - \mu_X)(Y - \mu_Y)$
 $\Rightarrow E[g] = E[(X - \mu_X)(Y - \mu_Y)]$

DEFINIZIONE: Date le variabili casuali X e Y , definiamo **COVARIANZA** di X e Y la quantità:

$$\sigma_{X,Y} = \text{cov}[X, Y] = E[(X - \mu_X)(Y - \mu_Y)]$$

PROPRIETÀ DELLA COVARIANZA

1. $\text{cov}[X, Y] = E[XY] - \mu_X\mu_Y$

Dalla definizione:

$$\begin{aligned} \text{cov}[X, Y] &= E[(X - \mu_X)(Y - \mu_Y)] = \\ &= E[XY - \mu_Y X - \mu_X Y + \mu_X \mu_Y] = \\ &= E[XY] - \mu_Y E[X] - \mu_X E[Y] + \mu_X \mu_Y = \\ &= E[XY] - \mu_X \mu_Y \end{aligned}$$

2. $\text{cov}[aX, bY] = ab \text{cov}[X, Y] \quad \forall a, b \in \mathbb{R}$

$$\begin{aligned} \text{cov}[aX, bY] &= E[(aX - a\mu_X)(bY - b\mu_Y)] = \\ &= ab \text{cov}[X, Y] \end{aligned}$$

3. $\text{cov}[X, Y] = \text{cov}[Y, X]$

4. $\text{cov}[X, X] = \text{var}[X]$

$$\text{cov}[X, X] = E[(X - \mu_X)^2] = \text{var}[X]$$

5. $\text{cov}[X + Y, Z] = \text{cov}[X, Z] + \text{cov}[Y, Z]$

Per provare l'ultima uguaglianza notiamo che:

$$\begin{aligned}
 \text{cov}[X + Y, Z] &= E[(X + Y)Z] - E[X + Y]E[Z] \\
 &\quad \uparrow \\
 &\quad \text{per 1)} \\
 &= E[XZ + YZ] - (E[X] + E[Y])E[Z] \\
 &= E[XZ] + E[YZ] - E[X]E[Z] - E[Y]E[Z] \\
 &= \text{cov}[X, Z] + \text{cov}[Y, Z]
 \end{aligned}$$

La covarianza tra due variabili casuali descrive la possibilità che tra le due variabili possa esistere una **relazione di tipo lineare**.

Se la covarianza è nulla, tra X ed Y è possibile che esista una relazione di tipo non lineare.

Esempio (numerico)

Tabella 1

X	-3	-2	-1	0	1	2	3
Y	9	4	1	0	1	4	9
XY	-27	-8	-1	0	1	8	27

$E[X] = 0$, $E[Y] = 28/7$, $E[XY] = 0$;
 allora $\text{cov}[X, Y] = E[XY] - E[X]E[Y] = 0$,
 ma tra X e Y sussiste la relazione $Y = X^2$ (non lineare).

La covarianza dipende dall'unità di misura utilizzata per misurare X e Y ; perciò si preferisce usare un altro indice, indipendente dall'unità di misura.

DEFINIZIONE: Date le variabili casuali X e Y con covarianza $\text{cov}[X, Y]$ e deviazioni standard rispettivamente σ_X e σ_Y , definiamo **COEFFICIENTE DI CORRELAZIONE** di X e Y il numero:

$$\rho_{X,Y} = \frac{\text{cov}[X, Y]}{\sigma_X \sigma_Y} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}$$

Proprietà di $\rho_{X,Y}$

1. $|\rho_{X,Y}| \leq 1$
2. $\rho_{X,Y} = 0$ se $\text{cov}[X, Y] = 0$
3. Se $Y = mX + q$ (dipendenza lineare), allora

$$\begin{cases} \rho_{X,Y} = 1 & \text{se } m > 0, \\ \rho_{X,Y} = -1 & \text{se } m < 0. \end{cases}$$

Infatti:

$$\begin{aligned} \text{cov}[X, Y] &= \text{cov}[X, mX + q] = \text{cov}[X, mX] + \text{cov}[X, q] \\ &= m \text{cov}[X, X] = m \text{var}[X] = m\sigma_X^2 \geq 0 \text{ se } m \geq 0 \end{aligned}$$

\uparrow
 $=0$

e

$\text{var}[Y] = \text{var}[mX + q] = m^2 \text{var}[X] \Rightarrow$
 $\Rightarrow \sigma_Y = m\sigma_X > 0$ per definizione di deviazione standard.

Quindi:

$$\rho_{X,Y} = \begin{cases} \frac{m\sigma_X^2}{\sigma_X \cdot m\sigma_X} = 1 & \text{se } m > 0, \\ \frac{-|m|\sigma_X^2}{\sigma_X \cdot |m|\sigma_X} = -1 & \text{se } m < 0. \end{cases}$$

4. $\forall a, b \in \mathbb{R}^+ \quad \rho_{aX, bY} = \rho_{X,Y}$

Infatti:

$$\rho_{aX, bY} = \frac{\text{cov}[aX, bY]}{\sigma_{aX}\sigma_{bY}} = \frac{ab \text{cov}[X, Y]}{ab\sigma_X\sigma_Y} = \rho_{X,Y}$$

$$\text{var}[aX] = a^2 \text{var}[X] \Rightarrow \sigma_{aX} = a\sigma_X > 0$$

$$\text{var}[bX] = b^2 \text{var}[X] \Rightarrow \sigma_{bX} = b\sigma_X > 0$$

DEFINIZIONE: Date le variabili casuali X e Y e la funzione $g(X, Y)$, definiamo **VALORE ATTESO CONDIZIONATO** di g , dato $X = x$, la quantità:

$$E[g(X, Y)|X = x] = \sum_y g(x, y) f_{Y|X}(y|x)$$

- se $g(X, Y) = Y$

$$E[Y|X = x] = E[Y|X] = \sum y f_{Y|X}(y|x)$$

Esempio. Calcolare il coefficiente di correlazione tra le variabili casuali X, Y dell'esempio del lancio dei due tetraedri.

Ricordiamo che $\rho_{X,Y} = \frac{\text{cov}[X,Y]}{\sigma_X \sigma_Y}$.

Dobbiamo calcolare la $\text{cov}[X, Y]$ e quindi $E[XY]$, $E[X]$ ed $E[Y]$.

- $g(X, Y) = XY$

$$\begin{aligned} E[XY] &= \sum xy f_{X,Y}(x, y) = \\ &= 1 \cdot 1 f_{X,Y}(1, 1) + 1 \cdot 2 f_{X,Y}(1, 2) + 1 \cdot 3 f_{X,Y}(1, 3) + \\ &+ 1 \cdot 4 f_{X,Y}(1, 4) + 2 \cdot 2 f_{X,Y}(2, 2) + 2 \cdot 3 f_{X,Y}(2, 3) + \\ &+ 2 \cdot 4 f_{X,Y}(2, 4) + 3 \cdot 3 f_{X,Y}(3, 3) + 3 \cdot 4 f_{X,Y}(3, 4) + \\ &+ 4 \cdot 4 f_{X,Y}(4, 4) = \\ &= \frac{1}{16} + 2 \cdot \frac{1}{16} + 3 \cdot \frac{1}{16} + 4 \cdot \frac{1}{16} + 4 \cdot \frac{2}{16} + 6 \cdot \frac{1}{16} + \\ &+ 8 \cdot \frac{1}{16} + 9 \cdot \frac{3}{16} + 12 \cdot \frac{1}{16} + 16 \cdot \frac{4}{16} = \frac{135}{16} \end{aligned}$$

Per calcolare $E[X]$ ed $E[Y]$ possiamo utilizzare

• $g(X, Y) = X$

$$\begin{aligned}
 E[X] &= \sum x f_{X,Y}(x, y) = 1 \cdot \sum_y f_{X,Y}(1, y) + \\
 &+ 2 \cdot \sum_y f_{X,Y}(2, y) + 3 \cdot \sum_y f_{X,Y}(3, y) + \\
 &+ 4 \cdot \sum_y f_{X,Y}(4, y) = \\
 &= 1 \cdot \left(\frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} \right) + 2 \cdot \left(\frac{2}{16} + \frac{1}{16} + \frac{1}{16} \right) + \\
 &+ 3 \cdot \left(\frac{3}{16} + \frac{1}{16} \right) + 4 \cdot \frac{4}{16} = \frac{40}{16}
 \end{aligned}$$

• $g(X, Y) = Y$

$$\begin{aligned}
 E[Y] &= \sum y f_{X,Y}(x, y) = 1 \cdot \sum_x f_{X,Y}(x, 1) + \\
 &+ 2 \cdot \sum_x f_{X,Y}(x, 2) + 3 \cdot \sum_x f_{X,Y}(x, 3) + \\
 &+ 4 \cdot \sum_x f_{X,Y}(x, 4) = 1 \cdot \frac{1}{16} + 2 \cdot \left(\frac{1}{16} + \frac{2}{16} \right) + \\
 &+ 3 \cdot \left(\frac{1}{16} + \frac{1}{16} + \frac{3}{16} \right) + 4 \cdot \left(\frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} \right) = \\
 &= \frac{50}{16}
 \end{aligned}$$

oppure molto più **velocemente** attraverso le funzioni di densità marginali: $E[X] = \sum x f_X(x)$ ed $E[Y] = \sum y f_Y(y)$, da cui

$$\begin{aligned} E[X] &= 1 \cdot f_X(1) + 2 \cdot f_X(2) + 3 \cdot f_X(3) + 4 \cdot f_X(4) \\ &= 1 \cdot \frac{4}{16} + 2 \cdot \frac{4}{16} + 3 \cdot \frac{4}{16} + 4 \cdot \frac{4}{16} = \frac{40}{16} \end{aligned}$$

ed

$$\begin{aligned} E[Y] &= 1 \cdot f_Y(1) + 2 \cdot f_Y(2) + 3 \cdot f_Y(3) + 4 \cdot f_Y(4) \\ &= 1 \cdot \frac{1}{16} + 2 \cdot \frac{3}{16} + 3 \cdot \frac{5}{16} + 4 \cdot \frac{7}{16} = \frac{50}{16} \end{aligned}$$

cosicchè:

$$\text{cov}[X, Y] = E[XY] - E[X]E[Y] = \frac{135}{16} - \frac{40}{16} \cdot \frac{50}{16} = \frac{5}{8}$$

Calcoliamo ora σ_X e σ_Y .

$$\sigma_X^2 = \text{var}[X] = E[X^2] - E[X]^2$$

ed $E[X^2] = \sum x^2 f_X(x) \Rightarrow$

$$\begin{aligned} E[X^2] &= 1 \cdot f_X(1) + 4 \cdot f_X(2) + 9 \cdot f_X(3) + 16 \cdot f_X(4) \\ &= 1 \cdot \frac{4}{16} + 4 \cdot \frac{4}{16} + 9 \cdot \frac{4}{16} + 16 \cdot \frac{4}{16} = \frac{120}{16} = \frac{15}{2} \end{aligned}$$

perciò

$$\text{var}[X] = \frac{15}{2} - \left(\frac{5}{2}\right)^2 = \frac{5}{4} \Rightarrow \sigma_X = \frac{\sqrt{5}}{2}$$

Analogamente

$$\begin{aligned} E[Y^2] &= 1 \cdot f_Y(1) + 4 \cdot f_Y(2) + 9 \cdot f_Y(3) + 16 \cdot f_Y(4) \\ &= 1 \cdot \frac{1}{16} + 4 \cdot \frac{3}{16} + 9 \cdot \frac{5}{16} + 16 \cdot \frac{7}{16} = \frac{170}{16} = \frac{85}{8} \end{aligned}$$

da cui

$$\text{var}[Y] = \frac{85}{8} - \left(\frac{25}{8}\right)^2 = \frac{55}{64} \Rightarrow \sigma_Y = \frac{\sqrt{55}}{8}$$

quindi

$$\rho_{X,Y} = \frac{5/8}{\sqrt{5}/2 \cdot \sqrt{55}/8} = \frac{2}{\sqrt{11}}$$

Adesso calcoliamo $E[Y|X = 2]$

Ricordiamo che

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

e

$$E[Y|X = 2] = \sum_i y_i f_{Y|X}(y_i|x = 2)$$

se $x = 2 \Rightarrow y \geq 2$

$$f_{Y|X}(2|2) = \frac{f_{X,Y}(2, 2)}{f_X(2)} = \frac{1}{2}$$

$$f_{Y|X}(3|2) = \frac{f_{X,Y}(2, 3)}{f_X(2)} = \frac{1}{4}$$

$$f_{Y|X}(4|2) = \frac{f_{X,Y}(2, 4)}{f_X(2)} = \frac{1}{4}$$

perciò

$$E[Y|X = 2] = 2 \cdot \frac{1}{2} + 3 \cdot \frac{1}{4} + 4 \cdot \frac{1}{4} = \frac{11}{4}$$

Esempio. (prova scritta del 17.07.2007)

Si consideri un'urna contenente 12 palline numerate.

2 palline hanno inciso il numero 1.

4 palline hanno inciso il numero 2.

2 palline hanno inciso il numero 3.

4 palline hanno inciso il numero 4.

Viene estratta una pallina. Sia X la variabile casuale che indica il numero inciso sulla pallina estratta ed Y la variabile casuale data da $Y = \frac{1}{2}(X - 2)^2$.

Calcolare:

1. la funzione di densità congiunta $f_{X,Y}$ e le relative marginali;

2. la covarianza;

3. $P[X > 2|Y = 1/2]$.

1)

$$P(X = 1) = P(X = 3) = \frac{2}{12} = \frac{1}{6};$$

$$P(X = 2) = P(X = 4) = \frac{4}{12} = \frac{1}{3}.$$

Inoltre

$$X = 1 \Rightarrow Y = \frac{1}{2}$$

$$X = 2 \Rightarrow Y = 0 \quad \Rightarrow X = \{1, 2, 3, 4\}$$

$$X = 3 \Rightarrow Y = \frac{1}{2} \quad \Rightarrow Y = \{0, \frac{1}{2}, 2\}$$

$$X = 4 \Rightarrow Y = 2$$

	X=1	X=2	X=3	X=4	f_x
Y=0	○	●	○	○	
Y=1/2	●	○	●	○	
Y=2	○	○	○	●	
f_x	1/6	1/3	1/6	1/3	1

Handwritten annotations in the table:
 - Row Y=0: Green arrow from X=2 to X=4 labeled "= 1/3".
 - Row Y=1/2: Green arrow from X=1 to X=3 labeled "= 1/6".
 - Row Y=1/2: Green arrow from X=3 to X=4 labeled "= 1/6".
 - Row Y=2: Green arrow from X=4 to the right labeled "= 1/3".
 - Bottom right cell: "1" circled in red.

Figura 1: Tabella dei valori della funzione di densità congiunta $f_{X,Y}$

Dai valori di $P(X = i)$, $i = 1, \dots, 4$ otteniamo subito i valori della funzione di densità marginale f_X .

Dai valori congiunti di X e Y individuiamo subito nella tabella a doppia entrata le coppie (x, y) per le quali la funzione di densità congiunta è diversa da zero (●)

Quindi

$$f_{X,Y}(1, \frac{1}{2}) = \frac{1}{6}, \quad f_{X,Y}(2, 0) = \frac{1}{3},$$

$$f_{X,Y}(3, \frac{1}{2}) = \frac{1}{6}, \quad f_{X,Y}(4, 2) = \frac{1}{3},$$

da cui si ricavano i valori della marginale f_Y :

$$f_Y(0) = \frac{1}{3}$$

$$f_Y(\frac{1}{2}) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

$$f_Y(2) = \frac{1}{3}$$

2) Ovviamente X ed Y non sono indipendenti. Infatti presa la coppia $(X, Y) = (1, 0)$

$$f_{X,Y}(1, 0) = 0, \text{ ma } f_X(1) = \frac{1}{6} \text{ e } f_Y(0) = \frac{1}{3}$$

$$\Rightarrow f_{X,Y} \neq f_X \cdot f_Y$$

$$E[XY] = 1 \cdot \frac{1}{2} \cdot \frac{1}{6} + 2 \cdot 0 \cdot \frac{1}{3} + 3 \cdot \frac{1}{2} \cdot \frac{1}{3} + 4 \cdot 2 \cdot \frac{1}{3} = 3$$

$$E[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{3} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{3} = \frac{8}{3}$$

$$E[Y] = 0 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + 2 \cdot \frac{1}{3} = \frac{5}{6}$$

$$\Rightarrow \text{cov}[X, Y] = 3 - \frac{20}{9} = \frac{7}{9}$$

$$3) P[X > 2|Y = \frac{1}{2}] = P[X = 3|Y = \frac{1}{2}] + \\ + P[X = 4|Y = \frac{1}{2}]$$

$$P[X = 3|Y = \frac{1}{2}] = f_{X|Y}(3|\frac{1}{2}) = \frac{f_{X,Y}(3, \frac{1}{2})}{f_Y(\frac{1}{2})} = \frac{1/6}{1/3} = \frac{1}{2}$$

$$P[X = 4|Y = \frac{1}{2}] = f_{X|Y}(4|\frac{1}{2}) = \frac{f_{X,Y}(4, \frac{1}{2})}{f_Y(\frac{1}{2})} = \frac{0}{1/3} = 0$$

$$\Rightarrow P[X > 2|Y = \frac{1}{2}] = \frac{1}{2}$$

Esempio. (prova scritta del 09.01.2017)

Sia X la variabile casuale che assume i valori $\{0, 1\}$
 ed Y la variabile casuale che assume i valori $\{2, 3\}$.

Sapendo che

$$P[Y = 2] = \frac{2}{5},$$

$$P[X = 0|Y = 2] = P[Y = 2|X = 0] = \frac{2}{3},$$

calcolare

$$P[X = 1|Y = 3].$$

Costruisco la tabella a doppia entrata

	$X = 0$	$X = 1$	f_Y
$Y = 2$			$\frac{2}{5}$
$Y = 3$			
f_X			①

Ricordo che $P[X = x|Y = y] = \frac{P[X=x,Y=y]}{P[Y=y]}$,

$$P[Y = y|X = x] = \frac{P[X=x,Y=y]}{P[X=x]}.$$

Dalla tabella ricavo subito $P[Y = 3] = 1 - \frac{2}{5} = \textcircled{\frac{3}{5}}$

mentre da

$$P[X = 0|Y = 2] = \frac{P[X=0,Y=2]}{P[Y=2]}$$

ottengo

$$P[X = 0, Y = 2] = P[X = 0|Y = 2] \cdot P[Y = 2] = \\ = \frac{2}{3} \cdot \frac{2}{5} = \frac{4}{15}$$

e quindi

$$P[X = 1, Y = 2] = \frac{2}{5} - \frac{4}{15} = \frac{2}{15}$$

e da

$$P[Y = 2|X = 0] = \frac{P[X=0,Y=2]}{P[X=0]}$$

ricavo

$$P[X = 0] = \frac{P[X=0,Y=2]}{P[Y=2|X=0]} = \frac{4/15}{2/3} = \frac{2}{5}$$

Di conseguenza

$$P[X = 1] = 1 - \frac{2}{5} = \frac{3}{5}$$

e infine

$$P[X = 0|Y = 3] = \frac{3}{5} - \frac{4}{15} = \frac{2}{15}$$

$$P[X = 1|Y = 3] = \frac{3}{5} - \frac{2}{15} = \frac{7}{15}$$

Pertanto

$$P[X = 1|Y = 3] = \frac{P[X=1,Y=3]}{P[Y=3]} = \frac{7/15}{3/5} = \frac{7}{9}$$

TEOREMA 1. Siano X, Y variabili casuali indipendenti e g_1, g_2 funzioni tali che $g_1 = g_1(X)$ e $g_2 = g_2(Y)$ siano variabili casuali. Allora vale

$$E[g_1 \cdot g_2] = E[g_1] \cdot E[g_2]$$

Nel caso discreto

$$\begin{aligned} E[g_1(X)g_2(Y)] &= \sum_i \sum_j g_1(x_i)g_2(y_j) \underbrace{f_{X,Y}(x_i, y_j)}_{=f_X(x_i)f_Y(y_j)} \\ &= \sum_i \sum_j g_1(x_i)g_2(y_j) f_X(x_i) f_Y(y_j) \\ &= \sum_i g_1(x_i) f_X(x_i) \sum_j g_2(y_j) f_Y(y_j) = E[g_1]E[g_2] \end{aligned}$$

TEOREMA 2. Se X, Y sono variabili casuali indipendenti allora $\text{cov}[X, Y] = 0$

Basta scegliere:

$$g_1(X) = X - \mu_X$$

$$g_2(Y) = Y - \mu_Y$$

cosicchè

$$\begin{aligned} \text{cov}[X, Y] &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= E[g_1 g_2] \underbrace{=}_{\text{Th.1}} E[g_1]E[g_2] = E[X - \mu_X]E[Y - \mu_Y] \\ &= (E[X] - \mu_X)(E[Y] - \mu_Y) = 0 \end{aligned}$$

oppure

$g_1(X) = X, g_2(Y) = Y$ e dal teorema 1)

$$E[XY] = E[X]E[Y] \Rightarrow$$

$$\Rightarrow \text{cov}[X, Y] = E[XY] - E[X]E[Y] = 0.$$

OSSERVAZIONE

$\text{cov}[X, Y] = 0 \not\Rightarrow X, Y$ indipendenti

cioè se $\text{cov}[X, Y] = 0$ non vale $f_{X,Y}(x, y) = f_X(x)f_Y(y)$

DEFINIZIONE: Due variabili casuali X, Y si dicono **non correlate** se e solo se $\text{cov}[X, Y] = 0$.

Quindi:

INDIPENDENZA \Rightarrow NON CORRELAZIONE
 \Leftarrow

Vedi esempio in cui $\text{cov}[X, Y] = 0$, ma $Y = X^2$.

Nota bene: esiste un solo caso in cui $\text{cov}[X, Y] = 0 \Rightarrow$ indipendenza. Ciò avviene quando la funzione di densità congiunta di X ed Y è **GAUSSIANA**.

COMBINAZIONI LINEARI DI VARIABILI CASUALI

Date n variabili casuali X_1, \dots, X_n si consideri:

1) $g(X_1, \dots, X_n) = X_1 + \dots + X_n$

si ha che:

- $E[g] = \sum_{i=1}^n E[X_i]$

La dimostrazione è ovvia.

- $\text{var}[g] = \sum_{i=1}^n \text{var}[X_i] + 2 \sum_{\substack{i, j \\ i \neq j}} \text{cov}[X_i, X_j]$

Infatti:

$$\begin{aligned} \text{var}[g] &= \text{var} \left[\sum_{i=1}^n X_i \right] \\ &= E \left[\left(\sum_{i=1}^n X_i - E \left[\sum_{i=1}^n X_i \right] \right)^2 \right] \\ &= E \left[\left(\sum_{i=1}^n X_i - \sum_{i=1}^n E[X_i] \right)^2 \right] \\ &= E \left[\left(\sum_{i=1}^n (X_i - E[X_i]) \right)^2 \right] \end{aligned}$$

$$\begin{aligned}
 &= E \left[\sum_{i=1}^n \sum_{j=1}^n (X_i - E[X_i]) (X_j - E[X_j]) \right] \\
 &= \sum_{i=1}^n \sum_{j=1}^n E [(X_i - E[X_i]) (X_j - E[X_j])]
 \end{aligned}$$

si spezza nella somma in cui $i = j$

$$\Rightarrow \sum_{i=1}^n E [(X_i - E[X_i])^2] = \sum_{i=1}^n \text{var}[X_i]$$

e nella somma in cui $i \neq j$

$$\begin{aligned}
 &2 \sum_{\substack{i=1 \\ i \neq j}}^n \sum_{j=1}^n E [(X_i - E[X_i]) (X_j - E[X_j])] \\
 &= 2 \sum_{\substack{i=1 \\ i \neq j}}^n \sum_{j=1}^n \text{cov}[X_i, X_j].
 \end{aligned}$$

Per esempio, se scegliamo $n = 2$

$$g = X_1 + X_2$$

$$\begin{aligned}
 \text{var}[X_1 + X_2] &= E [(X_1 + X_2)^2] - (E[X_1 + X_2])^2 \\
 &= E[X_1^2 + 2X_1X_2 + X_2^2] - (E[X_1] + E[X_2])^2
 \end{aligned}$$

$$\begin{aligned}
 &= E[X_1^2] + 2E[X_1X_2] + E[X_2^2] + \\
 &- \left(E[X_1]^2 + 2E[X_1]E[X_2] + E[X_2]^2 \right) \\
 &= \text{var}[X_1] + \text{var}[X_2] + 2\text{cov}[X_1, X_2]
 \end{aligned}$$

Quanti sono i termini di covarianza?

$$n = 2 \rightarrow \text{cov}[X_1, X_2] \rightarrow 1$$

$$n = 3 \rightarrow \text{cov}[X_1, X_2], \text{cov}[X_1, X_3], \text{cov}[X_2, X_3] \rightarrow 3$$

$$\begin{aligned}
 n = 4 \rightarrow \text{cov}[X_1, X_2], \text{cov}[X_1, X_3], \text{cov}[X_1, X_4], \\
 \text{cov}[X_2, X_3], \text{cov}[X_2, X_4], \text{cov}[X_3, X_4] \rightarrow 6
 \end{aligned}$$

⋮

$$n = k \rightarrow \dots \dots \dots \rightarrow \frac{k(k-1)}{2}$$

quindi nel caso generale, in cui $g = X_1 + \dots + X_n$, i termini di covarianza sono $\frac{n(n-1)}{2}$, mentre i termini di varianza sono n , per un totale di $\frac{n(n+1)}{2}$ termini differenti.

2) $g(X_1, \dots, X_n) = a_1X_1 + \dots + a_nX_n$, con $a_i \in \mathbb{R}$, per $i = 1, \dots, n$.

- $E[g] = \sum_{i=1}^n a_i E[X_i]$

La dimostrazione è ovvia.

$$\bullet \text{ var}[g] = \sum_{i=1}^n a_i^2 \text{ var}[X_i] + 2 \sum_{\substack{i, j \\ i \neq j}} a_i a_j \text{ cov}[X_i, X_j]$$

Infatti:

$$\begin{aligned} \text{var}[g] &= \text{var} \left[\sum_{i=1}^n a_i X_i \right] \\ &= \sum_{i=1}^n \text{var}[a_i X_i] + 2 \sum_{\substack{i=1 \\ i \neq j}}^n \sum_{j=1}^n \text{cov}[a_i X_i, a_j X_j] \\ &= \sum_{i=1}^n a_i^2 \text{var}[X_i] + 2 \sum_{\substack{i=1 \\ i \neq j}}^n \sum_{j=1}^n a_i a_j \text{cov}[X_i, X_j] \end{aligned}$$

CASO PARTICOLARE

$n=2$, $g(X, Y) = X \pm Y$

$$E[g] = E[X \pm Y] = E[X] \pm E[Y]$$

$$\text{var}[g] = \text{var}[X \pm Y] = \text{var}[X] + \text{var}[Y] \pm 2 \text{cov}[X, Y]$$

3) $g(X_1, \dots, X_n) = X_1 + \dots + X_n$, con X_i a due a due NON CORRELATE, cioè $\text{cov}[X_i, X_j] = 0, i \neq j$

- $E[g] = \sum_{i=1}^n E[X_i]$

- $\text{var}[g] = \text{var} \left[\sum_{i=1}^n X_i \right] = \sum_{i=1}^n \text{var}[X_i]$

4) $g(X_1, \dots, X_n) = X_1 + \dots + X_n$, con X_i a due a due NON CORRELATE ed IDENTICAMENTE DISTRIBUITE, cioè $\text{cov}[X_i, X_j] = 0$ per $i \neq j$; $\mu_{X_i} = \mu$ e $\sigma_{X_i}^2 = \sigma^2$, $\forall i = 1, \dots, n$.

- $E[g] = n\mu$

- $\text{var}[g] = n\sigma^2$

Infatti:

$$E[g] = E \left[\sum_{i=1}^n X_i \right] = \sum_{i=1}^n E[X_i] = \sum_{i=1}^n \mu_{X_i} = n\mu,$$

$$\text{var}[g] = \text{var} \left[\sum_{i=1}^n X_i \right] = \sum_{i=1}^n \text{var}[X_i] = \sum_{i=1}^n \sigma_{X_i}^2 = n\sigma^2$$

5) $g(X_1, \dots, X_n) = \frac{1}{n} (X_1 + \dots + X_n) = \bar{X}_n$

con X_i a due a due NON CORRELATE ed IDENTICAMENTE DISTRIBUITE con media μ e varianza σ^2 .

\bar{X}_n è detta MEDIA CAMPIONARIA.

- $E[g] = E[\bar{X}_n] = \mu$
- $\text{var}[g] = \text{var}[\bar{X}_n] = \frac{\sigma^2}{n}$

Infatti:

$$E[g] = E\left[\frac{1}{n}\sum_{i=1}^n X_i\right] = \frac{1}{n}\sum_{i=1}^n E[X_i] = \frac{1}{n}\sum_{i=1}^n \mu_{X_i} = \frac{1}{n} \cdot n\mu = \mu,$$

$$\text{var}[g] = \text{var}\left[\frac{1}{n}\sum_{i=1}^n X_i\right] = \frac{1}{n^2}\text{var}\left[\sum_{i=1}^n X_i\right] = \frac{1}{n^2}\sum_{i=1}^n \text{var}[X_i] = \frac{1}{n^2}\sum_{i=1}^n \sigma_{X_i}^2 = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$$

OSSERVAZIONE

Nei punti numero 4) e 5) l'ipotesi di non correlazione a due a due può anche essere sostituita con l'ipotesi **più forte** di **INDIPENDENZA** (più forte perchè **independenza** \Rightarrow **non correlazione**)

In tal caso le variabili casuali X_1, \dots, X_n si dicono **indipendenti ed identicamente distribuite** che nel seguito abbrevieremo con **i.i.d.**

Relativamente al caso 5) è possibile provare che se X_i sono i.i.d. con funzione di densità normale $N(\mu, \sigma^2)$, allora

$$\bar{X}_n \text{ è NORMALE } N\left(\mu, \frac{\sigma^2}{n}\right)$$

Esempio. Le precipitazioni annuali a Brescia hanno distribuzione normale $N(\mu = 12.08 \text{ cm}, \sigma = 3.1 \text{ cm})$. Le precipitazioni di anni successivi sono indipendenti. Calcolare:

- 1) la probabilità che le precipitazioni dei prossimi 2 anni superino i 25 cm;
- 2) la probabilità che le precipitazioni del prossimo anno superino quelle dell'anno successivo per più di 3 cm.

1) Chiamiamo X_1 ed X_2 le precipitazioni (in cm) dei prossimi 2 anni. Introduciamo

$$Y = X_1 + X_2 \sim N(\mu_Y, \sigma_Y^2)$$

dove $\mu_Y = E[Y] = n\mu = 2 \cdot 12.08 \text{ cm} = 24.16 \text{ cm}$
 $\sigma_Y^2 = n\sigma^2 = 2 \cdot (3.1)^2 \text{ cm}^2 = 19.22 \text{ cm}^2$

Quindi

$$\begin{aligned} P[Y > 25] &= P\left[Z > \frac{25 - \mu_Y}{\sigma_Y}\right] = P\left[Z > \frac{0.84}{\sqrt{19.22}}\right] \\ &= P[Z > 0.1916] = 1 - P[Z \leq 0.1916] \cong \\ &\cong 1 - 0.576 = 0.424 \end{aligned}$$

2) La seconda domanda chiede di trovare

$$P[X_1 > X_2 + 3] = P[X_1 - X_2 > 3].$$

Chiamiamo

$$T = X_1 - X_2 \Rightarrow T \sim N(\mu_T, \sigma_T^2) \text{ dove}$$

$$\mu_T = \mu_{X_1} - \mu_{X_2} = \mu - \mu = 0$$

$$\sigma_T^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2 = 2\sigma^2 = 19.22 \text{ cm}^2$$

Quindi:

$$\begin{aligned} P[T > 3] &= P\left[\frac{T - \mu_T}{\sigma_T} > \frac{3 - \mu_T}{\sigma_T}\right] = P\left[Z > \frac{3}{\sqrt{19.22}}\right] \\ &= P[Z > 0.6843] = 1 - P[Z \leq 0.6843] \cong \\ &\cong 1 - 0.754 = 0.246. \end{aligned}$$