ESTENSIONE DEL CONCETTO DI VALORE ATTESO

Vogliamo ora estendere il concetto di valore atteso di una v.c.(1-D) al caso n-dimensionale.
Ci limitiamo ad analizzare il solo caso discreto.

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

- $g(X_1,\dots,X_n)=X_i$

$\Rightarrow \quad E[g]=E[X_i]=\mu_{X_i}$

- $g(X_1,\dots,X_n)=(X_i-\mu_{X_i})^2$

$\Rightarrow \quad E[g]=E[(X_i-\mu_{X_i})^2]=var[X_i]=\sigma^2_{X_i}$

Limitiamoci al caso bidimensionale ($n=2$).

- $g(X,Y)=(X-\mu_X)(Y-\mu_Y)$

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

DEFINIZIONE: Date le v.c. X, Y, definiamo **COVARIANZA** di X e Y la quantità:

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

PROPRIETÀ DELLA COVARIANZA

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

dalla definizione

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

$$= E[XY] - \mu_Y \underbrace{E[X]}_{\mu_X} - \mu_X \underbrace{E[Y]}_{\mu_Y} + \mu_X \cdot \mu_Y$$

$$= E[XY] - \mu_X \cdot \mu_Y$$

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

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

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

banale

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

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

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

$$\text{cov}[X+Y, Z] \underset{\text{per 1)}}{=} E[(X+Y)Z] - E[X+Y]\cdot E[Z] =$$

$$= E[XZ + YZ] - (E[X] + E[Y])\cdot E[Z]$$

$$= E[XZ] + E[YZ] - E[X]\cdot E[Z] - E[Y]E[Z]$$

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

La covarianza tra due v.c. descrive la possibilità che tra le due variabili possa esserci una relazione di tipo lineare.

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

Esempio (numerico)

X	-3	-2	-1	0	1	2	3
Y	9	4	1	0	1	4	9

XY -27 -8 -1 0 1 8 27

$E[X] = \mu_X = \frac{\sum x_i}{7} \equiv 0$; $E[Y] = \frac{28}{7}$; $E[XY] \equiv 0$

allora $cov[X,Y] = E[X \cdot Y] - E[X] \cdot E[Y] \equiv 0$

ma tra X e Y sussiste la relazione $Y = X^2$ (non lineare).

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

DEFINIZIONE: Date le v.c. X, Y con covarianza $cov[X,Y]$ e deviazioni standard σ_X, σ_Y rispettivamente, definiamo COEFFICIENTE DI CORRELAZIONE di X e Y il numero:

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

PROPRIETÀ DI $\rho_{X,Y}$

1) $|\rho_{X,Y}| \leq 1$

2) $\rho_{X,Y} = 0$ se $cov[X,Y] = 0$

3) se $Y = mX + q$ (dipendenza lineare)

$$\rho_{X,Y} = 1 \quad \text{se } m > 0$$

$$\rho_{X,Y} = -1 \quad \text{se } m < 0$$

Infatti:

$$\mathrm{cov}[X,Y] = \mathrm{cov}[X, mX+q] = \mathrm{cov}[X, mX] + \underset{0}{\underset{\parallel}{\mathrm{cov}[X,q]}}$$

$$= m\,\mathrm{cov}[X,X] = m\,\mathrm{var}[X] = m\sigma_X^2 \gtrless 0 \quad m \gtrless 0$$

$$\mathrm{var}[Y] = \mathrm{var}[mX+q] = \mathrm{var}[mX] = m^2\mathrm{var}[X]$$

$\sigma_Y = m\,\sigma_X > 0$ per def. di deviazione standard

$$\Rightarrow \rho_{X,Y} = \begin{cases} \dfrac{m\sigma_X^2}{\sigma_X \cdot m\sigma_X} = 1 & \text{se } m > 0 \\[2ex] \dfrac{-m\sigma_X^2}{\sigma_X \cdot m\sigma_X} = -1 & \text{se } m < 0. \end{cases}$$

4) $\forall a, b \in \mathbb{R}^+$

$$\rho_{aX,bY} = \rho_{X,Y}$$

Infatti:

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

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

$$\mathrm{var}[bY] = b^2\,\mathrm{var}[Y] \Rightarrow \sigma_{bY} = b\,\sigma_Y > 0$$

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

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

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

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

Esempio : Calcolare le coefficiente di correlazione tra le v.c. X, Y dell'esempio del lancio dei due tetraedri.

$$\rho_{X,Y} = \frac{\operatorname{cov}[X,Y]}{\sigma_X \cdot \sigma_Y}$$

1) dobbiamo calcolare la cov[X,Y] cioè $E[XY]$; $E[X]$; $E[Y]$.

• $g(X,Y) = XY$

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

$$= 1\cdot 1 \, f_{XY}(1,1) + 1\cdot 2 \cdot f_{X,Y}(1,2) + 1\cdot 3 \, f_{X,Y}(1,3) + 1\cdot 4 \cdot f_{XY}(1,4)$$
$$2\cdot 2 \, f_{X,Y}(2,2) + 2\cdot 3 \, f_{XY}(2,3) + 2\cdot 4 \cdot f_{XY}(2,4) +$$
$$3\cdot 3 \, f_{X,Y}(3,3) + 3\cdot 4 \, f_{X,Y}(3,4) + 4\cdot 4 \cdot f_{X,Y}(4,4)$$

$$= \frac{1}{16} + 2\cdot\frac{1}{16} + 3\cdot\frac{1}{16} + 4\cdot\frac{1}{16} + 4\cdot\frac{2}{16} + 6\cdot\frac{1}{16} + 8\cdot\frac{1}{16} +$$
$$9\cdot\frac{3}{16} + 12\cdot\frac{1}{16} + 16\cdot\frac{4}{16} = \frac{135}{16}$$

per calcolare $E[X]$ e $E[Y]$ possiamo utilizzare

• $g(X,Y) = X$

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

• $g(X,Y) = Y$

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

oppure molto più velocemente attraverso le funzioni di densità marginali.

$$E[X] = \Sigma\, x f_X(x) = 1 \cdot f_X(1) + 2 \cdot f_X(2) + 3 \cdot f_X(3) + 4 \cdot f_X(4)$$

$$= 1 \cdot \frac{4}{16} + 2 \cdot \frac{4}{16} + 3 \cdot \frac{4}{16} + 4 \cdot \frac{4}{16} = \frac{40}{16}$$

$$E[Y] = \Sigma\, y f_Y(y) = 1 \cdot f_Y(1) + 2 \cdot f_Y(2) + 3 \cdot f_Y(3) + 4 \cdot f_Y(4)$$

$$= 1 \cdot \frac{1}{16} + 2 \cdot \frac{3}{16} + 3 \cdot \frac{5}{16} + 4 \cdot \frac{7}{16} = \frac{50}{16}$$

cosicchè:

$$cov[X,Y] = E[XY] - E[X] \cdot E[Y]$$

$$= \frac{135}{16} \cdot \frac{5}{2} \cdot \frac{25}{8} = \frac{10}{16} = \frac{5}{8}$$

Calcoliamo ora σ_X, σ_Y.

$$\sigma_X = \sqrt{var[X]} \; ; \; \sigma_Y = \sqrt{var[Y]}$$

$$var[X] = E[X^2] - E[X]^2$$

e

$$E[X^2] = \Sigma\, x^2 f_X(x) = 1 \cdot f_X(1) + 2^2 \cdot f_X(2) + 3^2 \cdot f_X(3) + 4^2 \cdot f_X(4)$$

$$= 1 \cdot \frac{4}{16} + 4 \cdot \frac{4}{16} + 9 \cdot \frac{4}{16} + 16 \cdot \frac{4}{16}$$

$$= \frac{120}{16} = \frac{15}{2}$$

perciò

$$var[X] = \frac{15}{2} - \frac{25}{4} = \frac{5}{4} \quad \Rightarrow \sigma_X = \frac{\sqrt{5}}{2}$$

Analogamente

$$E[Y^2] = \Sigma\, y^2 f_Y(y) = 1 \cdot \frac{1}{16} + 4 \cdot \frac{3}{16} + 9 \cdot \frac{5}{16} + 16 \cdot \frac{7}{16} = \frac{170}{16} = \frac{85}{8}$$

$$var[Y] = \frac{85}{5} - \frac{625}{64} = \frac{55}{64} \quad \Rightarrow \sigma_Y = \frac{\sqrt{55}}{8}$$

quindi

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

Ricordando che:

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

- calcoliamo $E[Y|X=2]$.

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

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	" " " " "	2	
2	" " " " "	3	
4	" " " " "	4	.

Viene estratta una pallina.

X v.c. che indica il n° inciso sulla pallina estratta

Y v.c.: $Y = \frac{1}{2}(X-2)^2$.

Calcolare:

- la funzione di densità congiunta $f_{X,Y}$ e le marginali
- covarianza
- $P[X>2 \mid Y=\frac{1}{2}]$

$p(1) = \frac{2}{12} = \frac{1}{6}$; $p(2) = \frac{4}{12} = \frac{1}{3}$; $p(3) = \frac{1}{6}$; $p(4) = \frac{1}{3}$

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

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

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

$X = 4 \rightarrow Y = 2$

	$X=1$	$X=2$	$X=3$	$X=4$	f_Y
$Y=0$	0	•	0	0	$= 1/3$
$Y=\frac{1}{2}$	•	0	•	0	$= 1/6$; $= 1/6$
$Y=2$	0	0	0	•	$= 1/3$
f_X	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{3}$	1

Dai valori $p(1); p(2); p(3); p(4)$ ottengo subito i valori delle f. di densità marginale f_X.
Dai valori congiunti di X e Y individuo subito nella tabella a doppia entrata le coppie (x,y) per le quali la f. di densità congiunta è $\neq 0$. (•)
Quindi

$$f_{X,Y}(1,\tfrac{1}{2}) = \tfrac{1}{6}$$

$$f_{X,Y}(2,0) = \tfrac{1}{3}$$

$$f_{X,Y}(3,\tfrac{1}{2}) = \tfrac{1}{6}$$

$$f_{X,Y}(4,2) = \tfrac{1}{3}$$

da cui ricavo i valori della marginale f_Y.

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

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

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

- ovviamente X, Y non sono indipendenti. Infatti presa la coppia $(x,y) = (1,0)$

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

$$f_{XY} \neq f_X \cdot f_Y$$

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

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

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

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

$$P[X>2 \mid Y=\tfrac{1}{2}] = P[X=3 \mid Y=\tfrac{1}{2}] + P[X=4 \mid Y=\tfrac{1}{2}]$$

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

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

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

Esempio (prova scritta del 08.01.2017)

Sia X v.c. che assume i valori $\{0,1\}$ ed Y v.c. che assume i valori $\{2,3\}$. Sapendo che

$$P[Y=2] = \frac{2}{5} \quad , \quad P[X=0 \mid Y=2] = P[Y=2 \mid X=0] = \frac{2}{3}$$

calcolare

$$P[X=1 \mid Y=3] \; .$$

Costruisco la tabella a doppia entrata

	$X=0$	$X=1$	f_Y
$Y=2$			$2/5$
$Y=3$			
f_X			1

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

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

Dalla tabella ricavo subito:

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

mentre da

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

ricavo

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

e quindi

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

e da

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

ricavo

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

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

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

TEOREMA 1

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

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

Nel caso discreto

$$E[g_1(X)\, g_2(Y)] = \sum_i \sum_j g_1(x_i)\, g_2(y_j)\, f_{X,Y}(x_i, y_j)$$

$$\Vert \leftarrow \text{INDIPENDENZA}$$

$$= \sum_i \sum_j g_1(x_i)\, g_2(y_j)\, f_X(x_i) \cdot f_Y(y_j)$$

$$= \sum_i g_1(x_i) f_X(x_i) \cdot \sum_j g_2(y_j) f_Y(y_j)$$

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

TEOREMA 2

Se X, Y v.c. indipendenti allora $\mathrm{cov}[X,Y] \equiv 0$

Basta scegliere:

$g_1(X) = X - \mu_X$

$g_2(Y) = Y - \mu_Y$

$$\mathrm{cov}[X,Y] = E[(X-\mu_X)\cdot(Y-\mu_Y)] = E[g_1 \cdot g_2] \overset{\text{TH. 1}}{=} E[g_1] \cdot E[g_2]$$

$$= E[(X-\mu_X)] \cdot E[(Y-\mu_Y)]$$

$$= \{E[X]-\mu_X\} \cdot \{E[Y]-\mu_Y\}$$

$$= (\mu_X - \mu_X) \cdot (\mu_Y - \mu_Y) \equiv 0$$

oppure

$g_1(X) = X$; $g_2(Y) = Y$, dal TH. 1 .

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

$\Rightarrow \mathrm{cov}[X,Y] = E[XY] - E[X] \cdot E[Y] \equiv 0$

OSSERVAZIONE

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

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

DEFINIZIONE: Due v.c. X, Y si dicono non correlate sse $cov[X,Y]=0$.

quindi

INDIPENDENZA $\Rightarrow$ NON CORRELAZIONE

$\not\Leftarrow$

vedi esempio in cui $cov(X,Y)=0$ ma $Y=X^2$.

N.B. Esiste un solo caso in cui $cov[X,Y]=0 \Rightarrow$ indipendenza.
Ciò avviene quando la funzione di densità congiunta di X e Y è GAUSSIANA.

COMBINAZIONI LINEARI DI VARIABILI CASUALI

Date n v.c. $X_1, \dots X_n$ si consideri:

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

si ha che:

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

  banale

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

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

si spezza nella somma in cui $i = J$

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

e nella somma in cui $i \neq J$

$$
\begin{aligned}
&2 \sum_{\substack{i=1 \\ i \neq J}}^{n} \sum_{J=1}^{n} E\left[(X_i - E[X_i]) \cdot (X_J - E[X_J])\right] \\
&= 2 \sum_{\substack{i \\ i \neq J}} \sum_{J} cov[X_i, X_J]
\end{aligned}
$$

per semplicità scegliamo

$g = X_1 + X_2$

$$var[X_1+X_2] = E[(X_1+X_2)^2] - (E[X_1+X_2])^2$$

$$= E[X_1^2 + 2X_1X_2 + X_2^2] - (E[X_1]+E[X_2])^2$$

$$= E[X_1^2] + 2E[X_1X_2] + E[X_2^2] -$$

$$(E[X_1]^2 + 2E[X_1]\cdot E[X_2] + E[X_2]^2)$$

$$= var[X_1] + var[X_2] + 2\,cov[X_1,X_2]$$

Quanti sono i termini $\sum_i \sum_{\substack{J \\ i \neq J}} cov[X_i, X_J]$?

$m = 2 \rightarrow cov[X_1,X_2] \rightarrow 1$

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

$m = 4 \rightarrow cov[X_1,X_2],\ cov[X_1,X_3],\ cov[X_1,X_4],$
$cov[X_2,X_3],\ cov[X_2,X_4],$
$cov[X_3,X_4] \rightarrow 6$

$\vdots$

$m = k \rightarrow \ \dots\dots \ \rightarrow \dfrac{k(k-1)}{2}$

quindi nel caso generale in cui $g = X_1 + \dots + X_m$

i termini sono $\dfrac{m(m-1)}{2}$.

(2) $g(X_1,\dots,X_m) = a_1X_1 + \dots + a_mX_m \qquad a_i \in \mathbb{R} \ \forall i=1,\dots,m$

$$E[g] = E\left[\sum_{i=1}^{m} a_i X_i\right] = \sum_{i=1}^{m} E[a_i X_i] = \sum_{i=1}^{m} a_i E[X_i]$$

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

CASO PARTICOLARE

$m = 2$

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

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

$$var[g] = var[X \pm Y] = var[X] + var[Y] \pm 2\cdot cov[X,Y]$$

(3) $g(X_1,\dots,X_m) = X_1 + \dots + X_m$ con X_i a due a due NON CORRELATE cioè $\forall i,j \ cov[X_i,X_j] \equiv 0 \quad i\neq j$

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

(4) $g(X_1,\dots,X_m) = X_1 + \dots + X_m$ con X_i a due a due NON CORRELATE, identicamente distribuite.

$\Downarrow$

$\mu_{X_i} \equiv \mu \quad \forall i=1,\dots,m$

$\sigma_{X_i}^2 \equiv \sigma^2 \quad$ " "

$$E[g] = m\mu$$

$$var[g] = m\sigma^2$$

Infatti

$$E[g] = E[X_1 + \dots + X_m] = \sum_{i=1}^{m} E[X_i] = \sum_{i=1}^{m} \mu_{x_i} = m\mu$$

$$var[g] = var[X_1 + \dots + X_m] = var\left[\sum_{i=1}^{m} X_i\right] = \sum_{i=1}^{m} var[X_i]$$

$$= \sum_{i=1}^{m} \sigma^2_{x_i} = m\sigma^2.$$

(5) $g(X_1, \dots, X_m) = \frac{1}{m}(X_1 + \dots + X_m) = \bar{X}_m$ MEDIA CAMPIONARIA

con X_i a due a due non correlate e identicamente distribuite con media μ e varianza σ^2.

$$E[g] = E[\bar{X}_m] = \mu$$

$$var[g] = var[\bar{X}_m] = \frac{\sigma^2}{m}$$

Infatti:

$$E[g] = E\left[\frac{1}{m}(X_1 + \dots + X_m)\right] = \frac{1}{m} E\left[\sum_{i=1}^{m} X_i\right] = \frac{1}{m} \sum_{i=1}^{m} E[X_i]$$

$$= \frac{1}{m} \sum_{i=1}^{m} \mu_{x_i} = \frac{1}{m} \cdot m\mu = \mu$$

$$var[g] = var\left[\frac{1}{m}(X_1 + \dots + X_m)\right] = \frac{1}{m^2} var\left[\sum_{i=1}^{m} X_i\right]$$

$$= \frac{1}{m^2} \sum_{i=1}^{m} var[X_i] = \frac{1}{m^2} \cdot \sum_{i=1}^{m} \sigma^2_{x_i} = \frac{1}{m^2} \cdot m\sigma^2 = \frac{\sigma^2}{m}$$

OSSERVAZIONE

In (4) e in (5) l'ipotesi di non correlazione a due a due può essere sostituita con l'ipotesi più forte di INDIPENDENZA (indip $\underset{\not\Leftarrow}{\Rightarrow}$ non correl.)

In tale caso le v.c. $X_1, \dots, X_m$ si dicono indipendenti e identicamente distribuite cioè *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

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

Esempio

Le precipitazioni annuali a BS hanno distribuzione normale $N(\mu = 12,08\,mm;\ \sigma = 3,1\,mm)$. Le precipitazioni di anni successivi sono indipendenti. Calcolare:

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

$$Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

- X_1 per il 1° anno; X_2 per il 2° anno. X_1, X_2 sono indip.

$$Y = X_1 + X_2 \sim N(\mu_Y, \sigma_Y^2) \quad \text{dove } \mu_Y = E[Y] = m\mu = 2 \cdot 12,08 = 24,16$$

$$\sigma_Y^2 = var[Y] = m\sigma^2 = 2 \cdot (3,1)^2 = 19,22$$

$$P[Y > 25] = P\left[Z > \frac{25-\mu_Y}{\sigma_Y}\right] = P\left[Z > \frac{0,84}{\sqrt{19,22}}\right] = P[Z > 0.1916]$$

$$= 1 - P[Z < 0.1916] \cong 1 - 0.576 = 0.424$$

- se $Y = \alpha X + \beta$ con $\alpha, \beta \in \mathbb{R}$ $\alpha \neq 0$ e $X \sim N(\mu, \sigma^2)$ allora

$$Y \sim N(\mu_Y, \sigma_Y^2) \quad \text{dove } \mu_Y = \alpha\mu_X + \beta$$

$$\sigma_Y^2 = \alpha^2 \sigma_X^2$$

scelto $\alpha = -1$ e $\beta = 0$

$$Y = -X \sim N(-\mu_X, \sigma_X^2)$$

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

Chiamo $T = X_1 - X_2$ allora $T \sim N(\mu_T, \sigma_T^2)$ dove:

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

$$\sigma_T^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2 = 2\,\sigma^2 = 19{,}22$$

Quindi:

$$P[T > 3] = P\left[\frac{T - \mu_T}{\sigma_T} > \frac{3 - \mu_T}{\sigma_T}\right] = P\left[Z > \frac{3}{\sqrt{19{,}22}}\right]$$

$$= P[Z > 0.6843] = 1 - P[Z \leq 0.6843]$$

$$\simeq 1 - 0.754 = 0.246.$$