SPERANZA MATEMATICA DI V.A. DISCRETE CON DISTRIBUZIONE NOTEVOLE

1) Distribuzione Bernoulliana: $X \sim B(p)$.

$S_X = \{0, 1\}$ insieme finito: (*) ok

$$\begin{cases} p_X(0) = 1-p \\ p_X(1) = p \end{cases}$$

$$\mathbb{E}[X] = 0 \cdot p_X(0) + 1 \cdot p_X(1) = 0 \cdot (1-p) + 1 \cdot p = p$$

Quindi, se $X = 1_A$ (funzione indicatrice dell'evento $A \in \mathcal{F}$), $\mathbb{E}[1_A] = P(A)$.

2) Distribuzione Binomiale: $X \sim Bin(n, p)$.

$S_X = \{0, 1, \dots, n\}$ insieme finito: (*) ok

$$\left\{ p_X(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{per } k \in S_X \right.$$

$$\mathbb{E}[X] = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} =$$

$h = k-1$

$$= \sum_{k=1}^{n} \frac{n(n-1)!}{(k-1)!(n-1-(k-1))!} p^{k-1+1} (1-p)^{n-1-(k-1)} =$$

BINOMIO NEWTON

$$= np \sum_{h=0}^{n-1} \binom{n-1}{h} p^h (1-p)^{n-1-h} = np(p+1-p)^{n-1} = np$$

Il risultato $X \sim Bin(n,p) \Rightarrow \mathbb{E}[X] = np$ si può dimostrare in maniera alternativa come segue.

Ricordiamo che una maniera canonica per ottenere una $X \sim Bin(n,p)$ è la seguente:

$$\Omega = \underbrace{\{0,1\} \times \cdots \times \{0,1\}}_{n \text{ volte}} \qquad P(\{\omega\}) = p^{X(\omega)}(1-p)^{n-X(\omega)} \quad \text{dove } X(\omega) = \sum_{i=1}^{n} \omega_i$$

$$\forall \omega = (\omega_1, \dots, \omega_n) \in \Omega$$

Allora possiamo considerare le v.a. più semplici $X_1, \dots, X_n : \Omega \to \mathbb{R}$ tali che $X_i(\omega) = \omega_i \quad \forall \omega \in \Omega. \quad i = 1, \dots, n.$

In corrispondenza si ha:
$$\begin{cases} X = X_1 + \cdots + X_n & \text{(per costruzione)} \\ \mathbb{E}[X_i] = P(X_i = 1) = p & \text{(per quanto detto nella Bernoulliana)} \\ & \forall i = 1, \dots, n \end{cases}$$

Allora

$$\mathbb{E}[X] = \mathbb{E}[X_1 + \cdots + X_n] \underset{\text{linearità}}{=} \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n] = \underbrace{p + \cdots + p}_{n \text{ volte}} = np.$$

3) Distribuzione Ipergeometrica

$S_X = \{0, 1, \dots, n\}$ insieme finito: (*) ok.

$$\begin{cases} p_X(k) = \dfrac{\binom{m_1}{k}\binom{m_2}{n-k}}{\binom{m_1+m_2}{n}} \end{cases}$$ per $k \in S_X$

con $n < m_1 + m_2$

$$\mathbb{E}[X] = \sum_{k=0}^{n} k \frac{\binom{m_1}{k}\binom{m_2}{n-k}}{\binom{m_1+m_2}{n}}$$

anche qui bisogna gestire l'espressione con i fattoriali ...

NON LO FACCIAMO

Consideriamo invece il procedimento alternativo visto per la Binomiale. Anche in questo caso (pensando alle estrazioni senza reinserimento) si ha

$$X = X_1 + \dots + X_n \quad \text{dove} \quad X_i = \begin{cases} 1 & \text{estratta pallina di tipo 1} \\ 0 & \text{estratta pallina di tipo 2} \end{cases} \sim B\left(\frac{m_1}{m_1+m_2}\right) \quad \forall i \in \{1, \dots, n\}.$$

Allora

$$\mathbb{E}[X] = \mathbb{E}[X_1 + \dots + X_n] \underset{\text{linearità}}{=} \mathbb{E}[X_1] + \dots + \mathbb{E}[X_n] = \underbrace{\frac{m_1}{m_1+m_2} + \dots + \frac{m_1}{m_1+m_2}}_{n \text{ volte}} = n \frac{m_1}{m_1+m_2}$$

COMMENTI.

→ C'è una differenza tra i due approcci alternativi visti per Binomiale e Ipergeometrica: $\begin{cases} X_1, \dots, X_n \text{ indipendenti nel } 1^\circ \text{ caso,} \\ X_1, \dots, X_n \text{ non indipendenti nel } 2^\circ \text{ caso.} \end{cases}$

Questo non ha influenza sui risultati che coincidono se poniamo $p = \frac{n_1}{n_1 + n_2}$.
Al contrario ci sarà una differenza nel caso della <u>varianza</u> di cui parleremo prossimamente.

→ I risultati ottenuti ci dicono che, nel caso di un numero finito di estrazioni casuali, la media della v.a. che conta il numero di oggetti di un certo tipo estratti non cambia se le estrazioni casuali sono con o senza reinserimento.

→ In generale $\mathbb{E}[X]$ dipende dalla distribuzione di X, e non da come è fatta $X : \Omega \to \mathbb{R}$.
In effetti, nel caso discreto (l'unico che abbiamo visto finora) $\mathbb{E}[X]$ dipende solo dalla densità discreta p_X. Quindi quel che abbiamo visto per l'ipergeometrica, con riferimento alle estrazioni casuali senza reinserimento, vale anche per le estrazioni casuali in blocco.

4) DISTRIBUZIONE DI POISSON : $X \sim \text{Poisson}(\lambda)$.

$S_X = \{0,1,2,\dots\}$ insieme non finito.

$$\begin{cases} p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{per } k \in S_X \end{cases}$$

La condizione (*) è

$$\sum_{k\geq 0} |k|\, p_X(k) < \infty.$$

Se è vera

$$\mathbb{E}[X] = \sum_{k \geq 0} k\, p_X(k).$$

ATTENZIONE: Qui il valore assoluto è superfluo

Allora calcoliamo la serie senza il valore assoluto.

$$\sum_{k\geq 0} k \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k \geq 1} k \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k\geq 1} \cancel{k} \frac{\lambda^{k-1+1}}{\underbrace{\cancel{k}\cdot(k-1)!}} e^{-\lambda} = \lambda \sum_{k \geq 1} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} =$$

$$\underset{h=k-1}{=} \lambda \underbrace{\sum_{h \geq 0} \frac{\lambda^h}{h!} e^{-\lambda}}_{=1} = \lambda.$$

Quindi vale (*) perché $\lambda < \infty$, ed inoltre $\mathbb{E}[X] = \lambda$.

5) DISTRIBUZIONE GEOMETRICA ; $X \sim Geo(p)$

$S_X = \{0, 1, 2, \dots\}$ insieme non finito.

$\begin{cases} p_X(k) = (1-p)^k p \quad \text{per } k \in S_X \end{cases}$

La condizione (*) è

$$\sum_{k \geq 0} |k| p_X(k) < \infty$$

Se è vera

$$\mathbb{E}[X] = \sum_{k\geq 0} k \, p_X(k).$$

ATTENZIONE:

Qui il valore assoluto è superfluo

Allora calcoliamo la serie senza il valore assoluto.

Partiamo da $\sum_{k\geq 0} (1-p)^k p = 1 \Longrightarrow p \sum_{k\geq 0} (1-p)^k = 1 \Longrightarrow \sum_{k \geq 0} (1-p)^k = \frac{1}{p}$.

Allora, derivando rispetto a p (questo è un caso in cui derivata della serie coincide con la serie delle derivate), si ha:

$$\sum_{k\geq 0} k (1-p)^{k-1} (-1) = -p^{-2}$$

; i segni meno si cancellano

$$\sum_{k\geq 0} k (1-p)^{k-1} = \frac{1}{p^2} ;$$

moltiplicando per $p(1-p)$ membro a membro

$$\sum_{k\geq 0} k (1-p)^k p = \underbrace{\frac{(1-p)p}{p^2}}_{} .$$

$$= \frac{1-p}{p} = \frac{1}{p} - 1.$$

Quindi vale (*) perché $\frac{1}{p} - 1 < \infty$ (oss. $p \neq 0$) e $\boxed{\mathbb{E}[X] = \frac{1}{p} - 1}$.

6) DISTRIBUZIONE GEOMETRICA TRASLATA : $Y \sim GeoTraslata(p)$

Qui, invece di procedere con la definizione (si dovrebbe fare riferimento alla serie $\sum_{k \geq 1} k(1-p)^{k-1} p$ sia per la condizione (*), sia per il calcolo di $\mathbb{E}[X]$), osserviamo che

$$Y = X + 1 \quad \text{con} \quad X \sim Geo(p).$$

Allora

$$\mathbb{E}[Y] = \mathbb{E}[X+1] \overset{\text{linearità}}{=} \underbrace{\mathbb{E}[X]}_{\substack{= \frac{1}{p} - 1 \\ \text{(calcolato prima)}}} + \underbrace{\mathbb{E}[1]}_{=1} = \frac{1}{p} - 1 + 1 = \frac{1}{p}.$$

7) DISTRIBUZIONI BINOMIALE NEGATIVA e BINOMIALE NEGATIVA TRASLATA.

Anche in questo caso non faremo riferimento alla definizione

(si dovrebbero considerare le serie

$$\sum_{k\geq 0} k \binom{k+r-1}{r-1} p^r (1-p)^k$$

$$\sum_{h\geq r} h \binom{h-1}{r-1} p^r (1-p)^{h-r}$$

).

Al contrario consideriamo le somme di Geometriche (e Geometriche traslate) opportune (viste in passato) e usiamo la linearità:

$$X = \underbrace{X_1 + \dots + X_r}_{Geo(p)\ \text{indip.}} \Rightarrow \mathbb{E}[X] \overset{\text{LIN.}}{=} \mathbb{E}[X_1] + \dots + \mathbb{E}[X_r] = \underbrace{\frac{1}{p} - 1 + \dots + \frac{1}{p} - 1}_{r\ \text{volte}} = r\left(\frac{1}{p} - 1\right)$$

$$Y = \underbrace{Y_1 + \dots + Y_r}_{GeoTraslate(p)\ \text{indip.}} \Rightarrow \mathbb{E}[Y] \overset{\text{LIN.}}{=} \mathbb{E}[Y_1] + \dots + \mathbb{E}[Y_r] = \underbrace{\frac{1}{p} + \dots + \frac{1}{p}}_{r\ \text{volte}} = \frac{r}{p}$$

I valori ottenuti $\mathbb{E}[Y] = \frac{r}{p}$ e $\mathbb{E}[X] = r\left(\frac{1}{p} - 1\right)$ sono in accordo con altre formule.

Infatti si ha $Y = X + r$, da cui segue

$$\mathbb{E}[Y] = \mathbb{E}[X+r] \overset{\text{linearità}}{=} \mathbb{E}[X] + r.$$

In effetti si ha

$$\underbrace{\mathbb{E}[Y]}_{= \frac{r}{p}} = \underbrace{\underbrace{\mathbb{E}[X]}_{r\left(\frac{1}{p}-1\right)} + r}_{= \frac{r}{p} - \cancel{r} + \cancel{r}} .$$

ok

MOMENTI, VARIANZA e COVARIANZA.

Le definizioni di seguito possono essere date anche se X non è discreta.
Noi ora penseremo solo al caso discreto (l'unico che abbiamo visto finora).

Def. Il MOMENTO k-esimo di X è (se esiste finito) $\mathbb{E}[X^k]$

Def. Il MOMENTO CENTRATO k-esimo di X è (se esiste finito) $\mathbb{E}[(X-\mathbb{E}[X])^k]$

Def. La VARIANZA di X è (se esiste finita) $\mathbb{E}[(X-\mathbb{E}[X])^2]$,
cioè il momento centrato di ordine 2.

Useremo il simbolo $Var[X]$.

Si usa il termine SCARTO QUADRATICO MEDIO di una v.a. X per $\sqrt{Var[X]}$.

La covarianza riguarda coppie di v.a. definite su uno stesso spazio di probabilità e la vedremo dopo.

PROPRIETÀ DELLA VARIANZA

Per la proprietà di monotonia del valor medio, ed essendo $(X-\mathbb{E}[X])^2 \geq 0$, allora

$\mathbb{E}[(X-\mathbb{E}[X])^2] \geq \mathbb{E}[0]$, e quindi $Var[X] \geq 0$.

In realtà si può dire di più; si ha

$Var[X] = 0 \iff X$ è una v.a. costante
(si intende il caso in cui esiste $x_0 \in \mathbb{R}$ tale che si ha $p_X(x_0)=1$; in corrispondenza si ha $x_0 = \mathbb{E}[X]$).

$\iff X = \mathbb{E}[X]$.

Quindi la varianza ha un valore minimo; al contrario non ammette un valore massimo.

In ogni modo possiamo dire che:

varianza piccola, distribuzione concentrata vicino alla media;
varianza grande, distribuzione non concentrata vicino alla media.

Per capire meglio questo presentiamo il seguente risultato:

DISUGUAGLIANZA DI CHEBYSHEV*

$$\forall a>0 \qquad P(|X-\mathbb{E}[X]|\geq a) \leq \frac{Var[X]}{a^2}$$

(per ogni v.a. X con media e varianza finite; vale anche per v.a. non discrete)

* Si tratta di un matematico russo per il quale ci sono diverse traslitterazioni.

Pafnutij L'vovič Čebyšëv

Da Wikipedia, l'enciclopedia libera.

Pafnutij L'vovič Čebyšëv (in russo: Пафнутий Львович Чебышёв?; Akatovo, 16 maggio 1821 – San Pietroburgo, 8 dicembre 1894) è stato un matematico e statistico russo.

Egli è considerato uno dei padri fondatori della grande scuola matematica russa. Tra i suoi allievi presso l'Università di San Pietroburgo vanno menzionati Dmitrij Grave, Aleksandr Korkin, Aleksandr Ljapunov, Egor Zolotarëv, Andrej Markov padre e Konstantin Posse.

I polinomi di Čebyšëv gli devono il nome, così come esiste una famiglia di filtri elettronici analogici chiamati filtri di Čebyšëv. Egli è altresì noto per i suoi risultati nell'ambito della probabilità e della statistica, dove tra l'altro riscoprì, indipendentemente da Bienaymé (di cui però divenne amico), quella che ora è chiamata disuguaglianza di Čebyšëv.

Il suo nome si trova traslitterato in vari altri modi: **Chebychev** e **Chebyshov** in inglese; **Cebisceff** e **Chebycheff** in italiano; **Tchebycheff** e **Tschebyscheff** in francese; **Tchebychev**, **Tschebyschow** e **Tschebyscheff** in tedesco.

Pafnutij L'vovič Čebyšëv

Indice [nascondi]

Quindi si ha

$$\boxed{\forall a>0 \qquad P\left(|X-\mathbb{E}[X]|\geq a\right)\leq \frac{\mathrm{Var}[X]}{a^2}}$$

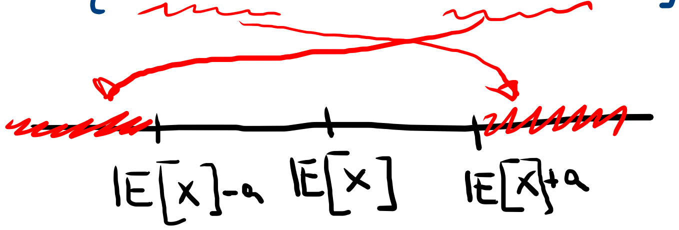
e lo dimostreremo nel caso di X v.a. discreta.

Se $\dfrac{\mathrm{Var}[X]}{a^2}\geq 1$ (questo può accadere se a è abbastanza vicino a zero)
si ha una disuguaglianza banale; il 1° membro è in $[0,1]$.

La disuguaglianza ci dice che l'evento $\{|X-\mathbb{E}[X]|\geq a\}=\{X-\mathbb{E}[X]\geq a\}\cup\{X-\mathbb{E}[X]\leq -a\}$

$$=\{X\geq \mathbb{E}[X]+a\}\cup\{X\leq \mathbb{E}[X]-a\}$$

ha probabilità che non può essere troppo grande
e diventa piccola se $\mathrm{Var}[X]$ è piccola.



DIMOSTRAZIONE (per il caso discreto; per il caso generale il procedimento è simile).

Si ha

FORMULA per trasformazione di X

$$\mathrm{Var}[X] = \mathbb{E}\left[(X-\mathbb{E}[X])^2\right] = \sum_{x_k \in S_X} (x_k - \mathbb{E}[X])^2 p_X(x_k) =$$

OSS: in generale $z^2 = |z|^2 \ \forall z \in \mathbb{R}$

$$= \sum_{\substack{x_k \in S_X: \\ |x_k - \mathbb{E}[X]| \geq a}} |x_k - \mathbb{E}[X]|^2 p_X(x_k) + \sum_{\substack{x_k \in S_X: \\ |x_k - \mathbb{E}[X]| < a}} |x_k - \mathbb{E}[X]|^2 p_X(x_k).$$

Essendo le due sommatorie non negative (somme di quadrati), si ottiene una minorazione togliendo una delle due e quindi la seconda:

$$\mathrm{Var}[X] \geq \sum_{\substack{x_k \in S_X: \\ |x_k - \mathbb{E}[X]| \geq a}} |x_k - \mathbb{E}[X]|^2 p_X(x_k).$$

Allora

$$\mathrm{Var}[X] \geq \sum_{\substack{x_k \in S_X \\ |x_k - \mathbb{E}[X]| \geq a}} a^2 p_X(x_k) = a^2 \cdot \sum_{\substack{x_k \in S_X \\ |x_k - \mathbb{E}[X]| \geq a}} p_X(x_k) = a^2 P(|X - \mathbb{E}[X]| \geq a)$$

e, dividendo membro a membro per a^2, si ottiene

$$\boxed{\frac{\mathrm{Var}[X]}{a^2} \geq P(|X - \mathbb{E}[X]| \geq a)}.$$

ALCUNE FORMULE PER LA VARIANZA (non solo per il caso discreto)

1) Formula alternativa: $\mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

(Nel seguito scriverò $\mathbb{E}^2[X]$ anziché $(\mathbb{E}[X])^2$).

<u>DIM</u> $\mathrm{Var}[X] = \mathbb{E}[(X-\mathbb{E}[X])^2] = \mathbb{E}[X^2 - 2X\cdot\mathbb{E}[X] + \mathbb{E}^2[X]] \overset{\text{linearità}}{=}$

$= \mathbb{E}[X^2] - 2\underbrace{\mathbb{E}[X]\cdot\mathbb{E}[X]}_{=\mathbb{E}^2[X]} + \mathbb{E}^2[X] = \mathbb{E}[X^2] - \mathbb{E}^2[X].$

2) Sia $a \in \mathbb{R}$. Allora $\mathrm{Var}[aX] = a^2\,\mathrm{Var}[X]$.

<u>DIM</u> $\mathrm{Var}[aX] = \mathbb{E}[(aX-\mathbb{E}[aX])^2] = \mathbb{E}[(aX - a\mathbb{E}[X])^2] = \mathbb{E}[a^2(X-\mathbb{E}[X])^2] \overset{\text{linearità}}{=} a^2\underbrace{\mathbb{E}[(X-\mathbb{E}[X])^2]}_{=\mathrm{Var}[X]}$

oppure

$\mathrm{Var}[aX] \underset{\text{formula alternativa}}{=} \mathbb{E}[(aX)^2] - \mathbb{E}^2[aX] = \mathbb{E}[a^2X^2] - (a\mathbb{E}[X])^2 = a^2\,\mathbb{E}[X^2] - a^2\,\mathbb{E}^2[X]$

$= a^2(\mathbb{E}[X^2] - \mathbb{E}^2[X]) \underset{\text{formula alternativa}}{=} a^2\,\mathrm{Var}[X].$

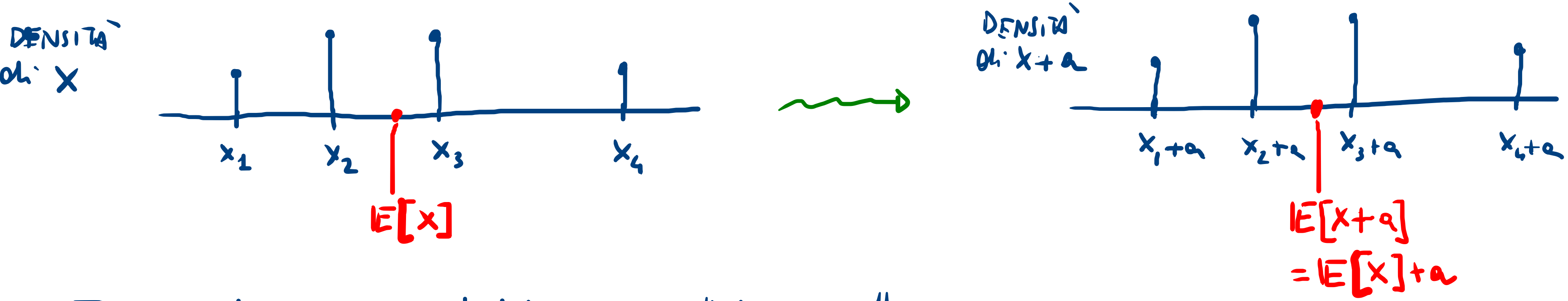
3) Sia $a \in \mathbb{R}$. Allora $Var[X+a] = Var[X]$

<u>DIM</u> $Var[X+a] = \mathbb{E}\left[\left(X+a-\underbrace{\mathbb{E}[X+a]}_{=\mathbb{E}[X]+a}\right)^2\right] = \mathbb{E}\left[\left(X+\cancel{a}-\mathbb{E}[X]-\cancel{a}\right)^2\right] = \mathbb{E}\left[\left(X-\mathbb{E}[X]\right)^2\right] = Var[X]$.

linearità

<u>OSS</u>. Anche in questo caso si può fare una dimostrazione con la formula alternativa.

<u>OSS</u>. Il risultato appena dimostrato ha la seguente interpretazione.
Facciamo riferimento al caso discreto per fissare le idee.



In entrambi i casi le distribuzioni si "disperdono" rispetto alle alle loro medie nello stesso modo; quindi non sorprende che si abbia la stessa varianza.

VARIANZA DI UNA SOMMA DI V.A. e INTRODUZIONE ALLA COVARIANZA

Si vuole dare una formula per $Var[X_1+X_2]$, dove entrambe le v.a. sono non costanti (se ad esempio si avesse X_2 costante, per quanto visto si avrebbe $Var[X_1]$).

Si ha

$$Var[X_1+X_2] = \mathbb{E}\Big[\big(X_1+X_2-\underbrace{\mathbb{E}[X_1+X_2]}_{=\mathbb{E}[X_1]+\mathbb{E}[X_2]}\big)^2\Big] = \mathbb{E}\Big[\big(X_1+X_2-\mathbb{E}[X_1]-\mathbb{E}[X_2]\big)^2\Big] =$$

linearità

$$= \mathbb{E}\Big[\big(X_1-\mathbb{E}[X_1]+X_2-\mathbb{E}[X_2]\big)^2\Big] = \mathbb{E}\Big[\big(X_1-\mathbb{E}[X_1]\big)^2+\big(X_2-\mathbb{E}[X_2]\big)^2+2\big(X_1-\mathbb{E}[X_1]\big)\big(X_2-\mathbb{E}[X_2]\big)\Big] =$$

quadrato di binomio

$$= \underbrace{\mathbb{E}\Big[\big(X_1-\mathbb{E}[X_1]\big)^2\Big]}_{=Var[X_1]} + \underbrace{\mathbb{E}\Big[\big(X_2-\mathbb{E}[X_2]\big)^2\Big]}_{=Var[X_2]} + 2\,\underbrace{\mathbb{E}\Big[\big(X_1-\mathbb{E}[X_1]\big)\big(X_2-\mathbb{E}[X_2]\big)\Big]}_{\overset{def.}{=} Cov(X_1,X_2)}$$

questa è la covarianza tra X_1 e X_2

Quindi:

$$\mathrm{Var}[X_1+X_2] = \mathrm{Var}[X_1]+\mathrm{Var}[X_2] + 2\,\mathrm{Cov}(X_1,X_2).$$

$\Leftrightarrow$ OSS Analogia con il quadrato di binomio.

Si ha una formula più generale nel caso di n addendi:

$$\mathrm{Var}[X_1+\dots+X_n] = \sum_{i=1}^{n}\mathrm{Var}[X_i] + 2\sum_{\substack{i,j=1\\ i<j}}^{n}\mathrm{Cov}(X_i,X_j).$$

$\leftarrow$ anche perché $\mathrm{Cov}(\cdot,\cdot)$ è simmetrica (vedere qui di seguito)

FORMULE PER LA COVARIANZA (non solo per il caso discreto).

1) Formula alternativa $\quad \mathrm{Cov}(X_1,X_2) = \mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\,\mathbb{E}[X_2]$

DIM

$$\mathrm{Cov}(X_1,X_2) = \mathbb{E}[(X_1-\mathbb{E}[X_1])(X_2-\mathbb{E}[X_2])] = \mathbb{E}[X_1X_2 - X_1\mathbb{E}[X_2] - X_2\mathbb{E}[X_1] + \mathbb{E}[X_1]\mathbb{E}[X_2]] \overset{\text{linearità}}{=}$$

$$= \mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2] + \mathbb{E}[X_1]\mathbb{E}[X_2] = \mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\,\mathbb{E}[X_2].$$

2) $\mathrm{Cov}(X_1,X_2) = \mathrm{Cov}(X_2,X_1)$ (SIMMETRIA)

IL PRODOTTO TRA NUMERI REALI È COMMUTATIVO

DIM

$$\mathrm{Cov}(X_1,X_2) = \mathbb{E}[(X_1-\mathbb{E}[X_1])(X_2-\mathbb{E}[X_2])] = \mathbb{E}[(X_2-\mathbb{E}[X_2])(X_1-\mathbb{E}[X_1])] = \mathrm{Cov}(X_2,X_1)$$

oppure $\quad \mathrm{Cov}(X_1,X_2) = \mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2] = \mathbb{E}[X_2X_1] - \mathbb{E}[X_2]\mathbb{E}[X_1] = \mathrm{Cov}(X_2,X_1).$

3) $\mathrm{Cov}(X,X) = \mathrm{Var}[X]$

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

oppure

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

OSSERVAZIONE

A differenza della varianza, che è sempre non negativa (anzi è positiva se la v.a. non è costante) la covarianza può assumere anche un valore negativo.

Si potrebbero dare altre formule del tipo

$\mathrm{Cov}(a_1 X_1, a_2 X_2) = \ldots$

$\mathrm{Cov}(X_1 + a_1, X_2 + a_2) = \ldots$

ESERCIZIO

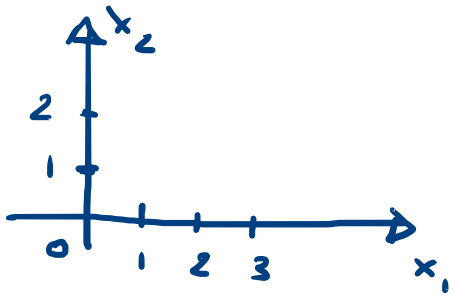
Sia $\underline{X} = (X_1, X_2)$ una v.a. discreta con la seguente densità congiunta:

$$P_{\underline{X}}(0,1) = P_{\underline{X}}(0,2) = P_{\underline{X}}(1,0) = P_{\underline{X}}(2,0) = P_{\underline{X}}(3,0) = \frac{1}{5}.$$

Calcolare $\mathbb{E}[X_1]$, $\mathbb{E}[X_2]$, $Var[X_1]$, $Var[X_2]$, $Cov(X_1, X_2)$.

SVOLGIMENTO

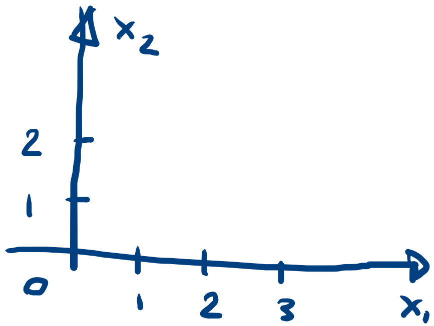
$$\begin{cases} P_{X_1}(0) = P_{\underline{X}}(0,1) + P_{\underline{X}}(0,2) = \frac{1}{5} + \frac{1}{5} = \frac{2}{5} \\ P_{X_1}(1) = P_{\underline{X}}(1,0) = \frac{1}{5} \\ P_{X_1}(2) = P_{\underline{X}}(2,0) = \frac{1}{5} \\ P_{X_1}(3) = P_{\underline{X}}(3,0) = \frac{1}{5} \end{cases}$$



$$\mathbb{E}[X_1] = 0 \cdot \frac{2}{5} + 1 \cdot \frac{1}{5} + 2 \cdot \frac{1}{5} + 3 \cdot \frac{1}{5} = \frac{6}{5}$$

$$Var[X_1] = \mathbb{E}[X_1^2] - \mathbb{E}^2[X_1] = 0^2 \cdot \frac{2}{5} + 1^2 \cdot \frac{1}{5} + 2^2 \cdot \frac{1}{5} + 3^2 \cdot \frac{1}{5} - \left(\frac{6}{5}\right)^2 = \frac{14}{5} - \frac{36}{25} = \frac{70-36}{25} = \frac{34}{25}$$

$$\begin{cases} p_{X_2}(0) = p_{\underline{X}}(1,0) + p_{\underline{X}}(2,0) + p_{\underline{X}}(3,0) = \frac{1}{5} + \frac{1}{5} + \frac{1}{5} = \frac{3}{5} \\ p_{X_2}(1) = p_{\underline{X}}(0,1) = \frac{1}{5} \\ p_{X_2}(2) = p_{\underline{X}}(0,2) = \frac{1}{5} \end{cases}$$



$$\mathbb{E}[X_2] = 0 \cdot \frac{3}{5} + 1 \cdot \frac{1}{5} + 2 \cdot \frac{1}{5} = \frac{3}{5}$$

$$\mathrm{Var}[X_2] = \mathbb{E}[X_2^2] - \mathbb{E}^2[X_2] = 0^2 \cdot \frac{3}{5} + 1^2 \cdot \frac{1}{5} + 2^2 \cdot \frac{1}{5} - \left(\frac{3}{5}\right)^2 = \frac{\overset{1}{\cancel{5}}}{\cancel{5}} - \frac{9}{25} = \frac{25-9}{25} = \frac{16}{25}$$

Oss. $\mathrm{Var}[X_2] < \mathrm{Var}[X_1]$ non sorprende.

Infine

$$\mathrm{Cov}(X_1, X_2) = \mathbb{E}[X_1 \cdot X_2] - \mathbb{E}[X_1] \cdot \mathbb{E}[X_2].$$ Si ha

$$\mathbb{E}[X_1 \cdot X_2] = \sum_{\underline{x} \in S_{\underline{X}}} x_1 \cdot x_2 \, p_{\underline{X}}(x_1, x_2) = 0 \cdot 1 \cdot \frac{1}{5} + 0 \cdot 2 \cdot \frac{1}{5} + 1 \cdot 0 \cdot \frac{1}{5} + 2 \cdot 0 \cdot \frac{1}{5} + 3 \cdot 0 \cdot \frac{1}{5} = 0.$$

Allora $\mathrm{Cov}(X_1, X_2) = 0 - \frac{6}{5} \cdot \frac{3}{5} = -\frac{18}{25}$.

Oss. Quando parleremo di rette di regressione potremo dire che $\mathrm{Cov}(X_1, X_2) < 0$ non sorprende.