Siano X e Y due v.a. discrete Abbiamo definito la densità congiunta: $p_{X,Y}(x,y) = P(X=x, Y=y)$

Se $g: \mathbb{R}^2 \to \mathbb{R}$ allora $E[g(X,Y)] = \sum_{x,y} g(x,y)\, p_{X,Y}(x,y)$

<u>Osservazione</u> tutto questo si estende al caso di n v.a.:

$$E[g(X_1,\dots,X_n)] = \sum_{x_1,\dots,x_n} g(x_1,x_2,\dots,x_n)\, p_{X_1,\dots,X_n}(x_1,\dots,x_n)$$

Proposizione Siano X e Y due variabili aleatorie (definite nello stesso spazio campionario).
Allora $E(X+Y) = E(X) + E(Y)$

Dim usiamo la formula precedente nel caso

$g(x,y) = x+y$, cioè $g(X,Y) = X+Y$

$$E(X+Y) = E(g(X,Y)) = \sum_{x,y} g(x,y)\, p_{X,Y}(x,y) = \sum_{x,y} (x+y)\, p_{X,Y}(x,y)$$

$$= \sum_{x,y} \left[x\, p_{X,Y}(x,y) + y\, p_{X,Y}(x,y) \right] = \sum_{x,y} x\, p_{X,Y}(x,y) + \sum_{x,y} y\, p_{X,Y}(x,y)$$

$$= \sum_{x} x \underbrace{\sum_{y} p_{X,Y}(x,y)}_{p_X(x)} + \sum_{y} y \underbrace{\sum_{x} p_{X,Y}(x,y)}_{p_Y(y)}$$

$$= E(X) + E(Y)$$

Ricordiamo che due variabili aleatorie sono indipendenti se e solo se

$$p_{X,Y}(x,y) = p_X(x)\, p_Y(y)$$

<u>Teorema</u> Siano X e Y due v.a. indipendenti. Allora

$$E(XY) = E(X)\,E(Y)$$

<u>Dim</u>. Sia $g(x,y) = xy$. $E(XY) = E[g(X,Y)] = \sum_{x,y} xy\, p_{X,Y}(x,y)$

$$= \sum_{x,y} xy\, p_X(x)\, p_Y(y) = \sum_x x\, p_X(x) \sum_y y\, p_Y(y) = E(X)\,E(Y)$$

Il problema della varianza di una somma di v.a.

Siano X e Y due v.a.

$$Var(X+Y) = E\left[(X+Y-E(X)-E(Y))^2\right] =$$

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

$$= Var(X) + Var(Y) + \_\_\_\_$$

Definizione Si dice covarianza fra due v.a. la quantità

$$Cov(X,Y) = E\left[(X-E(X))(Y-E(Y))\right]$$

Dunque: $Var(X+Y) = Var(X) + Var(Y) + 2\,Cov(X,Y)$

dove $Cov(X,Y) = E\left[(X-E(X))(Y-E(Y))\right]$

Proprietà della covarianza

1. $Cov(X,X) = Var(X)$ 2. $Cov(X,Y) = Cov(Y,X)$

3. $Cov(X,Y) = E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y)$

$$\overset{!}{=} E(XY) - E(X)E(Y)$$

4. $Cov(X, aY) = a\, Cov(X,Y)$

$Cov(X, Y+Z) = Cov(X,Y) + Cov(X,Z)$

<u>Corollario</u> Se X e Y sono indipendenti allora

<u>$Cov(X,Y) = 0$ e quindi $Var(X+Y) = Var(X) + Var(Y)$</u>

Due variabili al. X e Y si dicono <u>scorrelate</u> se $Cov(X,Y) = 0$. Abbiamo visto che due variabili a. indipendenti sono scorrelate. Il viceversa in generale è falso.

Esempio Siano X e $Y \sim Be(\frac{1}{2})$ indipendenti.

Siano $Z = X+Y$, $W = X-Y$

$$Cov(Z,W) = Cov(X+Y, X-Y) \overset{(4)}{=} Cov(X+Y, X) - Cov(X+Y, Y)$$

$$\overset{(2)}{=} Cov(X, X+Y) - Cov(Y, X+Y) \overset{(4)}{=} Cov(X,X) + Cov(X,Y) - Cov(Y,X) - Cov(Y,Y) = 0$$

Quindi Z e W sono scorrelate.

$P(Z=2) = \frac{1}{4}$. $P(W=1) = \frac{1}{4}$ $\quad P(Z=2, W=1) = 0$

cioè Z e W non sono indipendenti

Sia X e Y due v.a. definite nello stesso spazio campionario.

Supponiamo di voler "stimare" la variabile Y a partire da trasformazioni lineari della variabile X.

" $aX+b \simeq Y$ ". Quello che si fa è trovare a e b in modo tale che $aX+b$ e Y siano il più vicine possibile in "media quadratica", cioè che rendano minimo l'"errore"

$$e(a,b) = E\left[\left(Y-(aX+b)\right)^2\right]$$

Il problema è quindi di trovare i valori di a e b che minimizzino $e(a,b)$

Sia (a^*,b^*) il punto di minimo. Si ha

$$e(a^*,b^*) = Var(Y)\left[1 - \rho^2(X,Y)\right]$$

dove $\rho(X,Y) = \dfrac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$ è detto <u>coefficiente di correlazione</u>