VARIABILI ALEATORIE DISCRETE

Lezione 26/05/05

CASO 1 X è una variabile aleatoria discreta (scalare)

$\Rightarrow X$ può assumere i valori $x_1, x_2, \ldots, x_i, \ldots \in \mathbb{R}$
(quantità numerabile)

$p(x_i) = P(\{X = x_i\})$ - densità di probabilità discreta -

- $\boxed{\sum_i p(x_i) = 1}$
- $\boxed{p(x_i) \in [0,1]}$

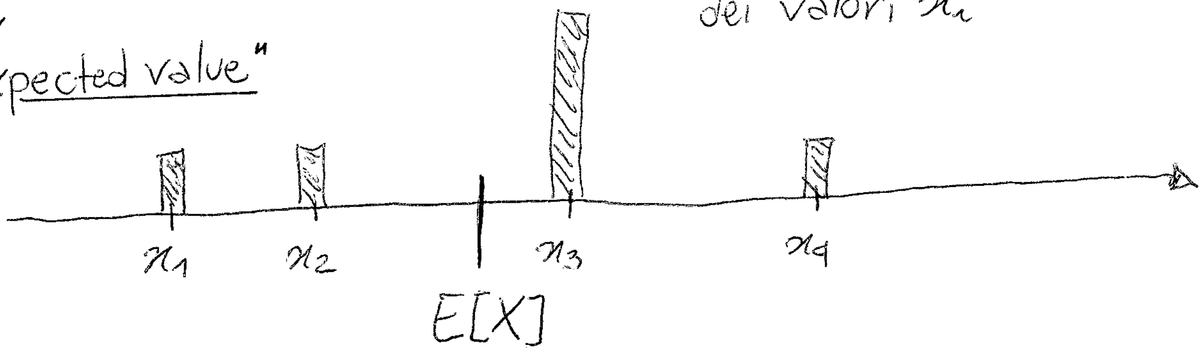
VALORE ATTESO (MEDIA, SPERANZA MATEMATICA)

$$E[X] \triangleq \sum_i x_i \, p(x_i)$$

combinazione convessa dei valori x_i

$\downarrow$

"expected value"



- $E[X+Y] = E[X] + E[Y]$
- $E[aX] = aE[X]$

$\Big\}$ l'operatore VALORE ATTESO è lineare

$\Rightarrow E[X]$ è una misura della locazione della densità di probabilità di X

CASO 2

X è una variabile aleatoria (discreta) vettoriale

$\Rightarrow$ X può assumere i valori $\boldsymbol{x}^{(1)}, x^{(2)}, \ldots, \boldsymbol{x}^{(i)}, \ldots \in \mathbb{R}^n$

$$p(x^{(i)}) = P(\{X = x^{(i)}\})$$ - densità di probabilità discreta -

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$ dove X_i sono variabili aleatorie scalari

$i = 1, \ldots, n$

$\rightarrow$ $p(x)$ è la densità di probabilità congiunta di $X_1, \ldots, X_n$.

$$x^{(i)} = \begin{bmatrix} x_1^{(i)} \\ x_2^{(i)} \\ \vdots \\ x_n^{(i)} \end{bmatrix}$$

$$p(x^{(i)}) = P\left(\{X_1 = x_1^{(i)}\} \cap \{X_2 = x_2^{(i)}\} \cap \ldots \cap \{X_n = x_n^{(i)}\}\right)$$

- $\sum_i p(x^{(i)}) = 1$
- $p(x^{(i)}) \in [0,1]$ per ogni i

$$E[X] = \sum_i x^{(i)} p(x^{(i)})$$ è un vettore!!!

$$\parallel$$

$$E\left[\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}\right] \overset{?}{=} \begin{pmatrix} E[X_1] \\ \vdots \\ E[X_n] \end{pmatrix}$$ da dimostrare...

probabilità marginali

Consideriamo una v.a. X_j, $j=1,\dots,n$.

$$p(z_k) = P(\{X_j = z_k\}) =$$

$$= P(X \in A_k) = \sum_{x^{(i)} \in A_k} P(\{X = x^{(i)}\})$$

$$A_k = \{x^{(i)} : x_j^{(i)} = z_k\}$$

densità di probabilità congiunta di X

densità di probabilità marginale di X_j

$\Rightarrow$ NOTA LA DENSITÀ DI PROBABILITÀ CONGIUNTA DI X È POSSIBILE CALCOLARE LE DENSITÀ DI PROBABILITÀ MARGINALI DELLE X_j, $j=1,\dots,n$.

IL VICEVERSA NON È IN GENERALE POSSIBILE.

$$E[X_j] = \sum_k z_k\, p(z_k) = \sum_k z_k \left(\sum_{x^{(i)} \in A_k} P(\{X = x^{(i)}\}) \right)$$

$$= \underbrace{\sum_i x_j^{(i)} P(\{X = x^{(i)}\})}_{} = \sum_i x_j^{(i)} p(x^{(i)})$$

questa è proprio la j-esima componente del vettore $E[X] = \sum_i x^{(i)} p(x^{(i)})$

OSSERVAZIONE - Consideriamo la v.a. $Z = \begin{bmatrix} X \\ Y \end{bmatrix}$.
Non conosciamo la densità di probabilità congiunta di Z, ma conosciamo $E[X]$ e $E[Y]$. Allora posso calcolare

$$E[Z] = \begin{bmatrix} E[X] \\ E[Y] \end{bmatrix}.$$

$$\underbrace{E[X+Y]}_{\substack{\text{Calcolato usando} \\ \text{la densità congiunta} \\ \text{di } X \text{ e } Y}} = \underbrace{E[X]}_{\substack{\text{Calcolato usando} \\ \text{la densità marginale} \\ \text{di } X}} + \underbrace{E[Y]}_{\substack{\text{Calcolato usando} \\ \text{la densità marginale} \\ \text{di } Y}}$$

PROPOSIZIONE - Se X e Y sono v.a. indipendenti, allora $E[XY] = E[X]E[Y]$.

$\Rightarrow$ NOTA - Se X e Y sono v.a. indipendenti, allora

$$\underbrace{p(x,y)}_{\text{congiunta}} = \underbrace{p(x)p(y)}_{\text{marginal}}$$

$$\underbrace{E[XY]}_{\substack{\text{richiede la} \\ \text{densità congiunta} \\ \text{di } X \text{ e } Y}} = \underbrace{E[X]}_{\substack{\text{richiede la} \\ \text{densità marginale} \\ \text{di } X}} \underbrace{E[Y]}_{\substack{\text{richiede la densità} \\ \text{marginale di } Y}}$$

VARIANZA

Si consideri la v.a. X (scalare)

Si definisce <u>VARIANZA</u> di X la quantità:

$$\boxed{Var(X) \triangleq E\left[(X - E[X])^2\right]}$$

$$\Rightarrow Var(X) = \sum_i \underbrace{\left(x_i - E[X]\right)^2}_{\text{scarto quadratico}} p(x_i)$$

La <u>varianza</u> di una variabile aleatoria X è una misura della <u>dispersione</u> della densità di probabilità di X rispetto al valor medio $E[X]$.

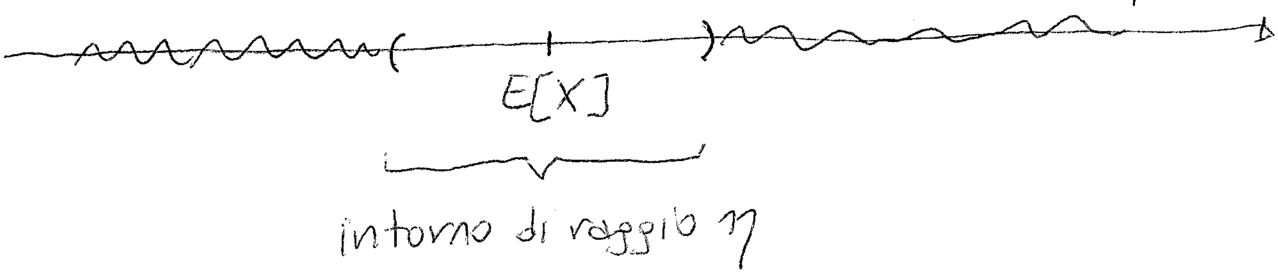
Infatti vale la seguente disuguaglianza:

<u>DISUGUAGLIANZA DI CHEBYSHEV</u>

Fissato $\eta > 0$, allora:

$$P\left\{|X - E[X]| > \eta\right\} \leq \frac{Var(X)}{\eta^2}$$

la probabilità di assumere valori al di fuori dell'intorno di $E[X]$ di raggio η è tanto più piccola quanto più $Var(X)$ è piccola.



NOTAZIONE - Spesso si indicano:

$$E[X] = \mu_X$$

$$Var(X) = \sigma_X^2$$

La quantità $\sigma_X = \sqrt{Var(X)}$ si chiama deviazione standard.

Metodo alternativo per calcolare $Var(X)$

$$Var(X) = E[(X - \mu_X)^2] = E[X^2 - 2\mu_X X + \mu_X^2] =$$

$$\underset{\text{linearità}}{=} E[X^2] - 2\mu_X \underbrace{E[X]}_{\mu_X} + \mu_X^2 = E[X^2] - 2\mu_X^2 + \mu_X^2 =$$

$$= E[X^2] - E[X]^2 = E[X^2] - \mu_X^2$$

PROPRIETÀ DELLA VARIANZA

- $Var(aX) = a^2 Var(X)$
- $Var(a+X) = Var(X)$

$a \in \mathbb{R}$ costante

COVARIANZA

Date due variabili aleatorie scalari X e Y, si definisce COVARIANZA di X e Y la quantità:

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

→ richiede la densità congiunta di X e Y

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

$$= E[XY - \mu_Y X - \mu_X Y + \mu_X \mu_Y]$$

$$= E[XY] - \mu_Y E[X] - \mu_X E[Y] + \mu_X \mu_Y$$

linearità

$$= E[XY] - \mu_X\mu_Y - \cancel{\mu_X\mu_Y} + \cancel{\mu_X\mu_Y}$$

$$= E[XY] - E[X]E[Y]$$

espressione alternativa per la covarianza

<u>Definizione</u> - Due v.a. scalari X e Y si dicono <u>SCORRELATE</u> se $Cov(X,Y)=0$.

$\Rightarrow$ Due v.a. scalari X e Y sono <u>SCORRELATE</u> se e solo se $E[XY]=E[X]E[Y]$.

Ricordando che due v.a. scalari X e Y <u>indipendenti</u> sono tali che $E[XY]=E[X]E[Y]$, allora si ottiene:

DUE V. A. SCALARI X e Y INDIPENDENTI SONO SCORRELATE

Il viceversa non è vero in generale, cioè due v.a. scorrelate possono non essere indipendenti

$\Rightarrow$ <u>SCORRELAZIONE</u> proprietà più debole dell'<u>INDIPENDENZA</u>

<u>Definizione</u> - Date due v.a. scalari X e Y, si definisce <u>coefficiente di correlazione</u> di X e Y la quantità:

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$$

Quindi $\rho_{X,Y} = 0$ se e solo se X e Y sono scorrelate.

<u>NOTA</u> - Il coefficiente di correlazione è <u>invariante</u> rispetto ai cambiamenti di scala:

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

---

1 $\mathrm{Var}(X+Y) = E\left[(X+Y-\mu_X-\mu_Y)^2\right]$

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

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

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

linearità

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

$\Rightarrow$ Se X e Y sono <u>scorrelate</u>, allora: $(\mathrm{Cov}(X,Y)=0)$

$$\boxed{\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).}$$

② Consideriamo una variabile vettoriale $Z = \begin{bmatrix} X \\ Y \end{bmatrix}$

Si <u>definisce</u> MATRICE DI COVARIANZA della v.a. Z la matrice:

$$\Sigma_Z = E\left[(Z - E[Z])(Z - E[Z])^T \right]$$

Come è fatta Σ_Z?

$$E\left[\underbrace{\begin{bmatrix} X - \mu_X \\ Y - \mu_Y \end{bmatrix}}_{\Uparrow \; Z - E[Z]} \underbrace{\begin{bmatrix} X - \mu_X & Y - \mu_Y \end{bmatrix}}_{\Uparrow \; (Z - E[Z])^T} \right] =$$

$$= E\left[\begin{pmatrix} (X-\mu_X)^2 & (X-\mu_X)(Y-\mu_Y) \\ (X-\mu_X)(Y-\mu_Y) & (Y-\mu_Y)^2 \end{pmatrix} \right] = \overbrace{\begin{bmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{bmatrix}}^{\Sigma_Z}$$

$\Rightarrow$ Σ_Z è una matrice simmetrica (cioè $\Sigma_Z = \Sigma_Z^T$) ed è tale che gli elementi sulla diagonale principale sono le varianze di X e Y, mentre i valori fuori dalla diagonale sono uguali e coincidono con la covarianza di X e Y.