

VARIABILI ALEATORIE

Spesso in un esperimento aleatorio non ci interessano tanto i dettagli del risultato quanto piuttosto un valore numerico determinato dal risultato. Queste quantità numeriche vengono chiamate **variabili aleatorie** e le indicheremo con **X, Y, ...**

Esempio 1. Una famiglia ha tre figli: lo spazio campione è dunque

$$\mathbf{S} = \{(m, m, m), (m, m, f), (m, f, m), (f, m, m), \\ (m, f, f), (f, m, f), (f, f, m), (f, f, f)\}$$

Per esempio (m,m,f) significa che il primogenito è un maschio, il secondo un maschio e la più piccola è una femmina.

Assumiamo che la **probabilità sia quella uniforme**, quindi ogni esito ha probabilità **$1/|\mathbf{S}| = 1/8$** .

Siamo interessati al numero di bambine della famiglia:

X = numero di bambine della famiglia, assume valori 0,1,2,3.

Determiniamo la probabilità che X assuma ciascuno dei quattro valori:

$$\{X = 0\} = \{(m, m, m)\}, \quad P(X = 0) = \frac{1}{8}$$

$$\{X = 1\} = \{((m, m, f), (m, f, m), (f, m, m))\} \quad P(X = 1) = \frac{3}{8}$$

$$\{X = 2\} = \{((m, f, f), (f, f, m), (f, m, f))\} \quad P(X = 2) = \frac{3}{8}$$

$$\{X = 3\} = \{(f, f, f)\}, \quad P(X = 3) = \frac{1}{8}$$

Osserva che

$$P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 1$$

Esempio 2. Una punto viene collocato sul segmento $[0,1]$ con probabilità uniforme. Siamo interessati a conoscere il valore **X** dell'ordinata corrispondente alla posizione del punto. Questo numero assume tutti i valori reali in $[0,1]$.

Determiniamo la probabilità che X assuma un valore minore o uguale a x:

Assumendo che $0 \leq x \leq 1$

$$P(X \leq x) = x$$

Osserva inoltre che per $x \leq 0$

$$P(X \leq x) = 0$$

mentre per $x \geq 1$

$$P(X \leq x) = 1$$

VARIABILE ALEATORIA

Definizione. Una variabile aleatoria è una funzione X che ad ogni evento elementare dello spazio campione S associa un numero reale.

$\{X = x\} \subset S$ è un evento e $P(X = x)$ è la sua probabilità .

Anche $\{X \leq x\} \subset S$ è un evento e $P(X \leq x)$ è la sua probabilità .

Una **variabile aleatoria** (v.a. nel seguito) si dice **discreta** se i valori che assume sono **finiti o numerabili** (non necessariamente interi), si dice **continua** se i valori che assume sono **non numerabili** (numeri reali).

Nell'esempio 1 abbiamo una variabile discreta:

$$\{X \leq 2\} = \{X = 0\} \cup \{X = 1\} \cup \{X = 2\},$$

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8}$$

Nell'esempio 2 abbiamo una variabile continua:

$$P(X \leq -0,2) = 0$$

$$P(X \leq 0,4) = 0,4$$

$$P(X \leq 0,8) = 0,8$$

$$P(X \leq 2,8) = 1$$

VARIABILI ALEATORIE DISCRETE

Consideriamo dapprima le v.a. **discrete** e per fissare le idee le successive definizioni saranno date per il caso in cui **X** assume **n** valori **x_1, \dots, x_n** .

Definizione. La funzione

$$p(x_i) = P(X = x_i), \quad i = 1, 2, \dots, n$$

è detta **distribuzione di probabilità** della variabile aleatoria X.

Osserva che la funzione p verifica

$$p(x_i) \geq 0 \text{ per ogni } i = 1, 2, \dots, n \text{ e } \sum_{i=1}^n p(x_i) = 1$$

Nell'esempio 1: $X \in \{0, 1, 2, 3\}$,

$$\begin{aligned} p(0) &= P(X = 0) = \frac{1}{8}, & p(1) &= P(X = 1) = \frac{3}{8} \\ p(2) &= P(X = 2) = \frac{3}{8}, & p(3) &= P(X = 3) = \frac{1}{8} \end{aligned}$$

Nota che questa variabile discreta assume valori interi, questo non succede sempre.

Esempio 3. Un dado viene lanciato, prima del lancio viene puntato un euro, la vincita è uguale a due volte il numero uscito diviso 6. L'importo X della vincita assume i valori: $x_i = 2/6, 4/6, 6/6, 8/6, 10/6, 12/6$ ognuno con **probabilità** $1/6$, quindi:
 $p(x_i) = P(X = x_i) = 1/6$. Si osservi che inoltre che:

$$p(x_i) \geq 0 \text{ per ogni } i = 1, 2, \dots, 6 \text{ e } \sum_{i=1}^6 p(x_i) = \frac{6}{6} = 1.$$

FUNZIONE DISTRIBUZIONE

Definizione. La funzione

$$F(x) = P(X \leq x), \quad x \in \mathbb{R}$$

è detta **funzione distribuzione (o di ripartizione)** della v.a. X .

In generale la funzione distribuzione può essere scritta come:

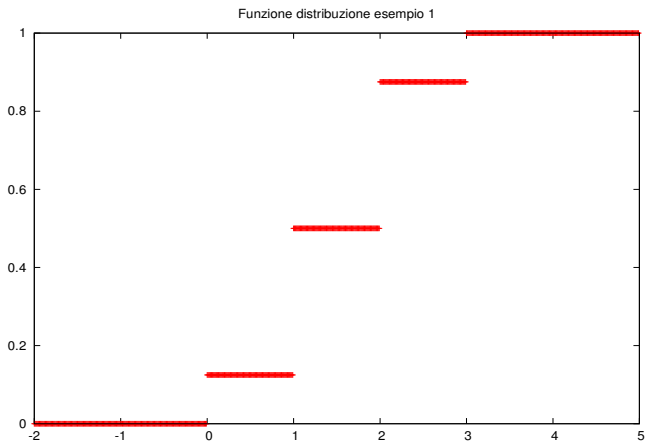
$$F(x) = \sum_{x_i \leq x} p(x_i)$$

dove la somma va su tutte le $x_i \leq x$.

Le probabilità $p(x_i)$ sono non negative per cui la funzione distribuzione è una funzione non decrescente della x .

Inoltre:

$$\lim_{x \rightarrow -\infty} F(x) = 0 \qquad \lim_{x \rightarrow \infty} F(x) = 1$$



Nei punti di discontinuità (corrispondenti ai valori che la variabile discreta può assumere) il valore di $F(x)$ è quello più grande.

VALORE ATTESO

Sia X una v.a. discreta che assume n valori x_1, \dots, x_n .

Definizione. Il **valore atteso** (o valore medio o speranza matematica) della variabile aleatoria X , indicato con $\mathbb{E}(X)$ è il numero

$$\mathbb{E}(X) = \sum_{i=1}^n x_i P(X = x_i) = \sum_{i=1}^n x_i p(x_i)$$

Notare che le definizioni sono le stesse di quelle della statistica descrittiva per la media campionaria, con le frequenze relative sostituite dalle probabilità .

Le proprietà del valore atteso sono le stesse della media campionaria: se a e c sono costanti $\mathbb{E}(aX + c) = a\mathbb{E}(X) + c$

Nell'esempio 1: X assume valori 0,1,2,3 e $p(0) = \frac{1}{8}$, $p(1) = \frac{3}{8}$,
 $p(2) = \frac{3}{8}$ $p(3) = \frac{1}{8}$. Quindi

$$\mathbb{E}(X) = \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{12}{8} = 1.5$$

Nell'esempio 3: X assume valori 2/6, 4/6, 6/6, 8/6, 10/6, 12/6
 quindi $x_i = 2/6, 4/6, 6/6, 8/6, 10/6, 12/6$. Inoltre

$$p(x_i) = \frac{1}{6}$$

per ognuna delle x_i . Quindi

$$\mathbb{E}(X) = \frac{1}{6} \frac{2}{6} + \frac{1}{6} \frac{4}{6} + \frac{1}{6} \frac{6}{6} + \frac{1}{6} \frac{8}{6} + \frac{1}{6} \frac{10}{6} + \frac{1}{6} \frac{12}{6} = 1,166666667$$

Il gioco può essere vantaggioso!!!! Si punta un euro contro un
 valore atteso della vincita di un euro e 16 centesimi.

VALORE ATTESO DI UNA FUNZIONE DI X

Sia g una funzione, allora $g(X)$ è una variabile aleatoria ed il suo valore atteso è dato da:

$$\mathbb{E}(g(X)) = \sum_i g(x_i)P(X = x_i) = \sum_i g(x_i)p(x_i)$$

Esempio: Sia X il valore che si ottiene lanciando un dado e sia $g(x) = \sqrt{x}$, allora:

$$\mathbb{E}(g(X)) = \frac{1}{6}(\sqrt{1} + \sqrt{2} + \sqrt{3} + \sqrt{4} + \sqrt{5} + \sqrt{6}) = 1,805$$

VARIANZA E DEVIAZIONE STANDARD

Definizione. Sia X una variabile aleatoria con valore atteso $\mu = \mathbb{E}(X)$, la **varianza di X** denotata con $\text{var}(X)$ è

$$\text{var}(\mathbf{X}) = \mathbb{E}\left((\mathbf{X} - \mu)^2\right) = \sum_i (\mathbf{x}_i - \mu)^2 \mathbf{p}(\mathbf{x}_i)$$

Lo scarto quadratico medio o deviazione standard è invece:

$$\sigma = \sqrt{\text{var}(\mathbf{X})}.$$

Notare che le definizioni sono le stesse della statistica descrittiva, con le frequenze relative sostituite dalle probabilità come anche le proprietà: **$\text{var}(\mathbf{aX} + \mathbf{c}) = \mathbf{a}^2 \text{var}(\mathbf{X})$**

Si noti che in modo analogo alla varianza campionaria, possiamo usare

$$\begin{aligned}\text{var}(\mathbf{X}) &= \mathbb{E}\left((\mathbf{X} - \mu)^2\right) = \sum_i (\mathbf{x}_i - \mu)^2 \mathbf{p}(\mathbf{x}_i) = \sum_i \mathbf{x}_i^2 \mathbf{p}(\mathbf{x}_i) + \\ &+ \mu^2 \sum_i \mathbf{p}(\mathbf{x}_i) - 2\mu \sum_i \mathbf{x}_i \mathbf{p}(\mathbf{x}_i) = \mathbb{E}(\mathbf{X}^2) - \mu^2\end{aligned}$$

ossia:

$$\text{var}(\mathbf{X}) = \mathbb{E}(\mathbf{X}^2) - [\mathbb{E}(\mathbf{X})]^2$$

Esercizio 1. Sia **X** una variabile aleatoria che assume solo i **valori 1 e -1** con **probabilità uniforme**. Sia **Y = 3X**. **Calcolare media e varianza di entrambe.**

$$\mathbb{E}(\mathbf{X}) = 1 \cdot \frac{1}{2} - 1 \cdot \frac{1}{2} = \mathbf{0}$$

$$\text{var}(\mathbf{X}) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \mathbb{E}(X^2) = 1 \cdot \frac{1}{2} + (-1)^2 \cdot \frac{1}{2} = \mathbf{1}$$

La variabile aleatoria **Y** assume i due **valori 3 e -3 con probabilità uniforme**:

$$P(Y = 3) = P(X = 1) = \frac{1}{2}, \quad P(Y = -3) = P(X = -1) = \frac{1}{2}$$

$$\mathbb{E}(\mathbf{Y}) = 3 \cdot \frac{1}{2} - 3 \cdot \frac{1}{2} = \mathbf{0}, \quad \text{var}(\mathbf{Y}) = \mathbb{E}(Y^2) = 3^2 \cdot \frac{1}{2} + (-3)^2 \cdot \frac{1}{2} = \mathbf{9}$$

Si noti, che avremmo potuto usare $\mathbb{E}(Y) = 3 \cdot \mathbb{E}(X) = 0$,
 $\text{var}(Y) = 3^2 \cdot \text{var}(X) = 9$.

Una ulteriore proprietà molto importante è la seguente: se X ed Y sono due variabile aleatorie allora,

- $\mathbb{E}(\mathbf{X} + \mathbf{Y}) = \mathbb{E}(\mathbf{X}) + \mathbb{E}(\mathbf{Y})$.

Inoltre se se X ed Y sono due variabile aleatorie **indipendenti** allora,

- $\mathbf{var}(\mathbf{X} + \mathbf{Y}) = \mathbf{var}(\mathbf{X}) + \mathbf{var}(\mathbf{Y})$.

Definizione. Due variabili aleatorie discrete X e Y si dicono **indipendenti** se gli eventi $\{X = x\}$ e $\{Y = y\}$ sono indipendenti per ogni x , valore di X ed y , valore di Y .

Esercizio 1.

Si lanci un dado e si calcoli media e varianza del valore uscito X .

$$\mathbb{E}(X) = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 3,5$$

$$\text{var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \frac{1}{6}(1+4+9+16+25+36) - (3,5)^2 = 2,916$$

Esercizio 2, metodo 1.

Si lancino due dadi e si calcoli media e varianza del valore uscito Z .
I lanci sono indipendenti e il risultato Z è la somma di X e Y che sono i risultati dei singoli dadi

$$\mathbb{E}(Z) = \mathbb{E}(X) + \mathbb{E}(Y) = 3,5 + 3,5 = 7$$

$$\text{var}(Z) = \text{var}(X) + \text{var}(Y) = 2,916 + 2,916 = 5,83$$

Esercizio 2, metodo 2.

Si lancino due dadi e si calcoli media e varianza del valore uscito Z . La variabile Z assume i valori 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 con probabilità $1/36$, $2/36$, $3/36$, $4/36$, $5/36$, $6/36$, $5/36$, $4/36$, $3/36$, $2/36$, $1/36$.

$$\mathbb{E}(Z) = \frac{1}{36}2 + \frac{2}{36}3 + \dots + \frac{1}{36}12 = 7$$

$$\text{var}(Z) = \mathbb{E}(Z^2) - [\mathbb{E}(Z)]^2 = \frac{1}{36}4 + \frac{2}{36}9 + \dots + \frac{1}{36}144 - 7^2 = 5,83$$