Valore atteso e varianza di una variabile casuale

Paolo Sciattella

Valore atteso

Valore atteso

- Le proprietà di una variabile casuale possono essere completamente dedotte dalla sua distribuzione di probabilità.
- D'altra parte sappiamo che spesso è necessaria una descrizione più sintetica, che, tramite pochi valori, ci permetta di cogliere le caratteristiche essenziali della distribuzione.
- Spesso siamo interessati a conoscere il **valore medio** che una variabile casuale può assumere in un gran numero di prove.
- Tale valore viene chiamato anche **valore atteso** o **speranza matematica**.

Valore atteso

Il **valore medio** o **atteso** di una variabile casuale X, indicato con $E(X)$, è definito come:

$$E(X) = \sum_i x_i P(x_i)$$ se la v. c. è **discreta**

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$ se la v. c. è **continua**

Valore atteso

Esempio: considerando la prova *lancio di due dadi* e la variabile casuale ***discreta*** X definita come somma dei punteggi ottenuti per ciascun dado, la distribuzione di probabilità di X è:

X	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
***P*(*X*)**	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

$$E(X) = \sum_i x_i P(x_i) =$$

$$= 2\frac{1}{36} + 3\frac{2}{36} + 4\frac{3}{36} + 5\frac{4}{36} + 5\frac{4}{36} + 6\frac{5}{36} + 7\frac{6}{36} + 8\frac{5}{36} + 9\frac{4}{36} + 10\frac{3}{36}$$
$$+ 11\frac{2}{36} + 12\frac{1}{36} = \mathbf{7}$$

Valore atteso

Esempio: *Tempo di attesa a una fermata dell'autobus*
Supponiamo che il tempo di attesa di una persona a una fermata dell'autobus sia una variabile aleatoria continua X, misurata in minuti, con distribuzione uniforme tra 0 e 10 minuti. Questo significa che ogni valore tra 0 e 10 è ugualmente probabile.

Valore atteso

Esempio: *Tempo di attesa a una fermata dell'autobus*

La funzione di densità di X sarà:

$$f(x) = \begin{cases} \dfrac{1}{10} \; se \; x \in [0; 10] \\ 0 \; altrimenti \end{cases}$$

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx = \int_{0}^{10} x \frac{1}{10} dx$$

Valore atteso

Esempio: *Tempo di attesa a una fermata dell'autobus*

Ricordando che:

$$\int x^n dx = \frac{x^{n+1}}{n+1}$$

Avremo che

$$E(X) = \int_0^{10} x \frac{1}{10} dx = \frac{1}{10} \left[\frac{x^2}{2}\right]_0^{10} = \frac{1}{10}\frac{10^2}{2} - 0 = \mathbf{5}$$

Valore atteso

Esempio: *Tempo di attesa a una fermata dell'autobus*

Ricordando che:

$$\int x^n dx = \frac{x^{n+1}}{n+1}$$

Il tempo di attesa medio è pari a 5 minuti

Avremo che

$$E(X) = \int_0^{10} x \frac{1}{10} dx = \frac{1}{10} \left[\frac{x^2}{2}\right]_0^{10} = \frac{1}{10}\frac{10^2}{2} - 0 = \mathbf{5}$$

Varianza

Varianza

- Sappiamo che la media evidenzia solamente la dimensione del fenomeno descritto dalla variabile casuale, ma non fornisce alcuna informazione sulla sua variabilità, che sappiamo essere una caratteristica di grande importanza.
- Una misura della variabilità della variabile casuale è la sua **varianza**.

Varianza

Per definire la varianza di una variabile casuale X, dobbiamo innanzitutto spiegare come calcolare il valore atteso di una **funzione** della v. c. X:

$$Y = g(X)$$

ossia come calcolare il valore atteso della v. c. Y, cioè $E(Y)$

Varianza

Il calcolo del valore atteso è per definizione dato da:

$$E(Y) = \sum_i y_i P(y_i)$$

se la v. c. è **discreta**

$$E(Y) = \int_{-\infty}^{+\infty} y f(y) dy$$

se la v. c. è **continua**

Varianza

- Queste formule presumono la previa conoscenza della $P(y_i)$ o della $f(y)$;
- tuttavia ciò diventa superfluo poiché per il calcolo del valore atteso si può utilizzare direttamente la funzione di probabilità o di densità della v.c. X.

Varianza

Infatti, data la v. c. X, e la sua trasformazione $Y = g(X)$, valgono le seguenti relazioni:

$$E(Y) = \sum_i y_i P(y_i) = \sum_i g(x_i) P(x_i)$$ se la v. c. è **discreta**

$$E(Y) = \int_{-\infty}^{+\infty} y f(y) dy = \int_{-\infty}^{+\infty} g(x) f(x) dx$$ se la v. c. è **continua**

Varianza

Per esempio, consideriamo la seguente variabile casuale:

X	**–2**	**–1**	**0**	**+1**	**+2**
***P*(*X*)**	0,1	0,2	0,4	0,2	0,1

e la sua funzione, $Y = X^2$. In corrispondenza di tale funzione possiamo definire la distribuzione di probabilità della v. c. Y:

Y	**0**	**+1**	**+4**
***P*(*Y*)**	0,4	0,4	0,2

Varianza

Il valore atteso di $Y = X^2$:

Y	0	+1	+4
$P(Y)$	0,4	0,4	0,2

si può calcolare come:

$$E(Y) = \sum_i y_i P(y_i) = 0 \cdot 0{,}4 + 1 \cdot 0{,}4 + 2 \cdot 0{,}4 = \mathbf{1,2}$$

Varianza

In alternativa, il valore atteso di $Y = X^2$:

X	−2	−1	0	+1	+2
$P(X)$	0,1	0,2	0,4	0,2	0,1

si può calcolare come:

$$E(Y) = \sum_i g(x_i)P(x_i) =$$

$$(-2)^2 \cdot 0{,}1 + (-1)^2 \cdot 0{,}2 + (0)^2 \cdot 0{,}4 + (1)^2 \cdot 0{,}2 + (2)^2 \cdot 0{,}1 = \mathbf{1,2}$$

Varianza

Detto ciò, poiché la varianza di una variabile casuale è la media degli scostamenti al quadrato dalla media, ossia il valore atteso della funzione $Y = g(X) = [X - E(X)]^2$, possiamo dare la seguente definizione:

La **varianza** $\boldsymbol{V(X)}$ di una variabile casuale X **è definita da**:

$$V(X) = \sum_i [x_i - E(X)]^2 P(x_i)$$ se la v. c. è **discreta**

$$V(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2 f(x) dx$$ se la v. c. è **continua**

Varianza

La varianza misura la differenza quadratica tra i possibili valori della v. c. e il suo valore atteso, con pesi dati dalle probabilità di osservare tali valori.

La varianza è in effetti il valore atteso della v. c. $[X - E(X)]^2$ e si può scrivere nei due modi seguenti:

$$V(X) = E\{[X - E(X)]^2\}$$

$$V(X) = E(X^2) - [E(X)]^2$$

Varianza

La varianza risulta nulla se X assume probabilità 1 in corrispondenza di un solo valore e probabilità zero altrove, mentre è tanto più elevata quanto più alta è la dispersione intorno al valore atteso.

La radice quadrata della varianza prende il nome di scostamento quadratico medio o **deviazione standard**

$$SD(X) = \sqrt{V(X)}$$

Variabili casuali standardizzate

Variabili casuali standardizzate

- In generale, la distribuzione di probabilità di una v. c. (o la funzione di densità di una v. c. continua) è una funzione matematica che può dipendere da uno o più parametri, ossia da costanti numeriche che ne determinano la forma. Di conseguenza, il valore atteso e la varianza di molte v. c. dipendono da tali parametri.

Variabili casuali standardizzate

Sia Y una v. c. funzione lineare di una v. c. X, ossia $Y = a + bX$, in cui a e b sono delle costanti, si hanno le seguenti proprietà:

$$E(Y) = E(a + bX) = a + bE(X)$$

e

$$V(Y) = V(a + bX) = b^2V(X)$$

Variabili casuali standardizzate

Se X è una variabile casuale con valore atteso $E(X)$ e deviazione standard $SD(X)$, allora:

$$Y = \frac{X - E(X)}{SD(X)}$$

Y **è una variabile casuale standardizzata,** caratterizzata da avere valore atteso nullo e varianza unitaria.
$E(Y) = 0; V(Y) = 1$

Variabili casuali standardizzate

Esempio: si consideri la distribuzione di probabilità connessa all'estrazione di una famiglia da un collettivo di 100 famiglie e all'osservazione del numero di figli presenti.

N. di figli	**0**	**1**	**2**	**3**	**4**	**5**
Probabilità	0,20	0,35	0,25	0,12	0,06	0,02

Si possono facilmente calcolare:

$$E(X) = 0 \cdot 0{,}20 + 1 \cdot 0{,}35 + 2 \cdot 0{,}25 + 3 \cdot 0{,}12 + 4 \cdot 0{,}06 + 5 \cdot 0{,}02 = 1{,}55$$

$$E(X^2) = 0^2 \cdot 0{,}2 + 1^2 \cdot 0{,}35 + 2^2 \cdot 0{,}25 + 3^2 \cdot 0{,}12 + 4^2 \cdot 0{,}06 + 5^2 \cdot 0{,}02 = 3{,}89$$

Variabili casuali standardizzate

Da cui derivano:

$$V(X) = E(X^2) - [E(X)]^2 = 3{,}89 - 1{,}55^2 = 1{,}488$$

e

$$\mathrm{SD}(X) = \sqrt{V(X)} = \sqrt{1{,}488} = 1{,}220$$

Ne deriva che la v. c. X può essere standardizzata come:

$$Y = \frac{X - 1{,}55}{1{,}220}$$