Azzolini Riccardo 2020-03-26

# Variabili aleatorie discrete

## 1 Parametri di una variabile aleatoria discreta

Esistono alcuni parametri numerici che permettono di quantificare in modo riassuntivo i valori assunti da una variabile aleatoria. Nel caso discreto, essi sono definiti nel seguente modo:

- Data una variabile aleatoria discreta $X$ che assume i valori $x_1, \dots, x_n$, si chiama **speranza matematica** o **valore medio** (o ancora *media*, *attesa*, *valore atteso*), e si indica con $\mu$ o $E(X)$, la quantità

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

  Siccome la speranza matematica è la somma dei possibili valori assunti da $X$, ciascuno pesato in base alla probabilità con la quale viene assunto, essa ha intuitivamente il significato di media dei valori assunti da $X$ (ed è appunto per questo che viene indicata con i termini *valore medio* e *media*).

- Si definisce **varianza** della variabile aleatoria discreta $X$, avente valore medio $E(X)$, la quantità

  $$\begin{aligned}\sigma^2 = \mathrm{Var}(x) &= \sum_{i=1}^{n} (x_i - E(X))^2\, p(x_i) \\ &= E[(X - E(X))^2]\end{aligned}$$

  Essa esprime la dispersione dei valori assunti da $X$ attorno al valore medio $E(X)$: più tali valori sono "concentrati" vicino alla media, più è piccola la varianza, e viceversa.

  L'elevamento al quadrato della differenza $x_i - E(X)$ è necessario[1] perché, altrimenti (per la definizione del valore medio) il risultato della sommatoria sarebbe sempre 0.

[1] Un'alternativa sarebbe considerare il valore assoluto di tale differenza, ma la varianza è convenzionalmente definita usando il quadrato perché così risulta più "comoda" da gestire nei calcoli.

- Si definisce **deviazione standard** o **scarto quadratico medio** della variabile aleatoria discreta $X$ la quantità

$$\sigma = \sqrt{\sigma^2} = \sqrt{\mathrm{Var}(X)} = \sqrt{E[(X - E(X))^2]}$$

  Rispetto alla varianza, la deviazione standard ha il vantaggio avere la stessa unità di misura dei valori di $X$ (ad esempio, se i valori di $X$ sono misurati in metri, allora la varianza sarà in metri quadri, mentre la deviazione standard sarà ancora in metri).

## 2 Problema: lotteria

*Problema*: Un giocatore acquista un biglietto di una lotteria: può vincere il primo premio di €5000 con probabilità 0.001, e il secondo premio di €2000 con probabilità 0.003. Quale dovrebbe essere il giusto prezzo del biglietto?

Si può considerare "giusto" (equo) un prezzo corrispondente alla vincita media che si può aspettare chi compra il biglietto. Allora, si considera una variabile aleatoria discreta $X$, che assume valori corrispondenti alle vincite, ovvero tale che

$$\begin{aligned} P\{X = 5000\} &= p(5000) = 0.001 \\ P\{X = 2000\} &= p(2000) = 0.003 \end{aligned}$$

(e quindi anche $P\{X = 0\} = p(0) = 1 - (0.001 + 0.003) = 0.996$, ma il valore 0 può non essere considerato, perché non contribuisce alla media), e si calcola:

$$\begin{aligned} E(X) &= 5000\,p(5000) + 2000\,p(2000) \\ &= 5000 \cdot 0.001 + 2000 \cdot 0.003 \\ &= 11 \end{aligned}$$

Il prezzo giusto per il biglietto dovrebbe perciò essere €11.

## 3 Problema: lotteria nazionale

*Problema*: In una lotteria nazionale vengono messi in palio i seguenti premi:

| | |
|---|---|
| 1° premio | € 3 000 000 |
| 2° premio | € 2 000 000 |
| 3° premio | € 1 000 000 |
| 5 premi da | € 100 000 |
| 20 premi da | € 10 000 |
| 100 premi da | € 1000 |

Si vendono 2 milioni di biglietti.

1. Qual è il valore medio della vincita per chi acquista un biglietto?
2. Se il biglietto costa €5, il gioco è equo?

La probabilità di acquistare un dato biglietto è

$$\frac{1}{2\,000\,000} = \frac{1}{2 \cdot 10^6} = \frac{1}{2} \cdot 10^{-6} = 0.5 \cdot 10^{-6} = 5 \cdot 10^{-7}$$

Allora, la variabile aleatoria $X$, che assume valori corrispondenti alle vincite, ha questa densità:

$$\begin{aligned}
p(3\,000\,000) &= 5 \cdot 10^{-7} \\
p(2\,000\,000) &= 5 \cdot 10^{-7} \\
p(1\,000\,000) &= 5 \cdot 10^{-7} \\
p(100\,000) &= 5 \cdot (5 \cdot 10^{-7}) = 25 \cdot 10^{-7} = 2.5 \cdot 10^{-6} \\
p(10\,000) &= 20 \cdot (5 \cdot 10^{-7}) = 100 \cdot 10^{-7} = 1 \cdot 10^{-5} \\
p(1000) &= 100 \cdot (5 \cdot 10^{-7}) = 500 \cdot 10^{-7} = 5 \cdot 10^{-5}
\end{aligned}$$

Il suo valore medio è quindi:

$$\begin{aligned}
E(X) &= 3\,000\,000\,p(3\,000\,000) + 2\,000\,000\,p(2\,000\,000) + 1\,000\,000\,p(1\,000\,000) \\
&\quad + 100\,000\,p(100\,000) + 10\,000\,p(10\,000) + 1000\,p(1000) \\
&= 3 \cdot 10^6 \cdot 5 \cdot 10^{-7} + 2 \cdot 10^6 \cdot 5 \cdot 10^{-7} + 1 \cdot 10^6 \cdot 5 \cdot 10^{-7} \\
&\quad + 1 \cdot 10^5 \cdot 2.5 \cdot 10^{-6} + 1 \cdot 10^4 \cdot 1 \cdot 10^{-5} + 1 \cdot 10^3 \cdot 5 \cdot 10^{-5} \\
&= 1.5 + 1 + 0.5 + 0.25 + 0.1 + 0.05 \\
&= 3.4
\end{aligned}$$

Le risposte alle domande poste dal problema sono quindi le seguenti:

1. Il valore medio della vincita è €3.40.
2. Siccome il prezzo del biglietto, €5, è maggiore della vincita media, il gioco *non* è equo: invece, è a sfavore di chi compra i biglietti.

# 4 Alcune proprietà del valore medio

Siano $X, Y$ delle variabili aleatorie discrete. Si può dimostrare che il valore medio è lineare, cioè che:

$$\begin{aligned}
E(kX) &= kE(X) \quad \forall k \in \mathbb{R} \\
E(X + Y) &= E(X) + E(Y)
\end{aligned}$$

Inoltre, se $X$ e $Y$ sono indipendenti, allora vale anche:

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

# 5 Problema: somma di lanci di un dado

*Problema*: La variabile aleatoria $X$ indica la somma dei punti ottenuti con il lancio di due dadi. Calcolarne il valore medio.

La soluzione più immediata consiste nell'enumerare tutte le coppie di lanci che danno ciascuna delle possibili somme; Siccome i due lanci sono indipendenti, ogni coppia di lanci ha probabilità

$$\frac{1}{6} \cdot 16 = \frac{1}{36}$$

quindi mediante l'enumerazione delle coppie è possibile ricavare da densità di $X$:

$$\begin{aligned}
x_1 &= 2 \leftarrow (1,1) & p(x_1) &= \frac{1}{36} \\
x_2 &= 3 \leftarrow (1,2),\ (2,1) & p(x_2) &= \frac{2}{36} \\
x_3 &= 4 \leftarrow (1,3),\ (2,2),\ (3,1) & p(x_3) &= \frac{3}{36} \\
x_4 &= 5 \leftarrow (1,4),\ (2,3),\ (3,2),\ (4,1) & p(x_4) &= \frac{4}{36} \\
x_5 &= 6 \leftarrow (1,5),\ (2,4),\ (3,3),\ (4,2),\ (5,1) & p(x_5) &= \frac{5}{36} \\
x_6 &= 7 \leftarrow (1,6),\ (2,5),\ (3,4),\ (4,3),\ (5,2),\ (6,1) & p(x_6) &= \frac{6}{36} \\
x_7 &= 8 \leftarrow (2,6),\ (3,5),\ (4,4),\ (5,3),\ (6,2) & p(x_7) &= \frac{5}{36} \\
x_8 &= 9 \leftarrow (3,6),\ (4,5),\ (5,4),\ (6,3) & p(x_8) &= \frac{4}{36} \\
x_9 &= 10 \leftarrow (4,6),\ (5,5),\ (6,4) & p(x_9) &= \frac{3}{36} \\
x_{10} &= 11 \leftarrow (5,6),\ (6,5) & p(x_{10}) &= \frac{2}{36} \\
x_{11} &= 12 \leftarrow (6,6) & p(x_{11}) &= \frac{1}{36}
\end{aligned}$$

A questo punto, è possibile calcolare il valore medio:

$$\begin{aligned} E(X) &= \sum_{i=1}^{11} x_i p(x_i) \\ &= \frac{1}{36}(2 \cdot 1 + 3 \cdot 2 + 4 \cdot 3 + 5 \cdot 4 + 6 \cdot 5 + 7 \cdot 6 \\ &\qquad + 8 \cdot 5 + 9 \cdot 4 + 10 \cdot 3 + 11 \cdot 2 + 12 \cdot 1) \\ &= \frac{252}{36} = 7 \end{aligned}$$

Invece, per risolvere il problema con meno calcoli si può sfruttare la linearità del valore medio. A tale scopo, si definiscono le variabili aleatorie $X_1$ e $X_2$, che rappresentano i risultati dei due lanci del dato: entrambe assumono i valori $x_1 = 1, \dots, x_6 = 6$ in modo equiprobabile (le loro distribuzioni marginali sono $p_1(x_i) = p_2(x_i) = \frac{1}{6}\ \forall i = 1, \dots, 6$), quindi hanno valore medio

$$\begin{aligned} E(X_1) = E(X_2) &= \sum_{i=1}^{6} x_i \overbrace{p_1(x_i)}^{=\frac{1}{6}} = \frac{1}{6}\sum_{i=1}^{6} x_i = \frac{1}{6}\sum_{i=1}^{6} i \\ &= \frac{1}{6} \cdot \underbrace{\frac{6(6+1)}{2}}_{\text{formula di Gauss}} = \frac{42}{12} = \frac{7}{2} \end{aligned}$$

Infine, siccome $X = X_1 + X_2$, il calcolo del valore medio diventa:

$$E(X) = E(X_1 + X_2) = E(X_1) + E(X_2) = \frac{7}{2} + \frac{7}{2} = 7$$

# 6 Problema: numero di teste

*Problema*: Trovare la varianza della variabile aleatoria $X$, definita come il numero di teste ottenute con tre lanci successivi di una moneta.

Per prima cosa, si enumerano le possibili sequenze di lanci per determinare la densità di $X$, considerando che ciascuna delle sequenze ha probabilità $(\frac{1}{2})^3 = \frac{1}{8}$:

$$\begin{array}{lll} x_1 = 0 \ \leftarrow \ (C, C, C) & & p(x_1) = \dfrac{1}{8} \\ x_2 = 1 \ \leftarrow \ (T, C, C),\ (C, T, C),\ (C, C, T) & & p(x_2) = \dfrac{3}{8} \\ x_3 = 2 \ \leftarrow \ (T, T, C),\ (T, C, T),\ (C, T, T) & & p(x_3) = \dfrac{3}{8} \\ x_4 = 3 \ \leftarrow \ (T, T, T) & & p(x_4) = \dfrac{1}{8} \end{array}$$

Si calcola poi il valore medio:

$$
\begin{aligned}
E(X) &= \sum_{i=1}^{4} x_i p(x_i) \\
&= 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} \\
&= \frac{3+6+3}{8} = \frac{12}{8} = \frac{3}{2}
\end{aligned}
$$

A questo punto, ci sono vari modi di calcolare la varianza:

- Il modo più immediato consiste nell'applicare direttamente la definizione di varianza come sommatoria:

$$
\begin{aligned}
\mathrm{Var}(X) &= \sum_{i=1}^{4} (x_i - E(X))^2 p(x_i) \\
&= \frac{1}{8}\left(0 - \frac{3}{2}\right)^2 + \frac{3}{8}\left(1 - \frac{3}{2}\right)^2 + \frac{3}{8}\left(2 - \frac{3}{2}\right)^2 + \frac{1}{8}\left(3 - \frac{3}{2}\right)^2 \\
&= \frac{1}{8}\left(-\frac{3}{2}\right)^2 + \frac{3}{8}\left(-\frac{1}{2}\right)^2 + \frac{3}{8}\left(\frac{1}{2}\right)^2 + \frac{1}{8}\left(\frac{3}{2}\right)^2 \\
&= \frac{1}{8}\left(\frac{9}{4} + 3 \cdot \frac{1}{4} + 3 \cdot \frac{1}{4} + \frac{9}{4}\right) = \frac{24}{32} = \frac{3}{4}
\end{aligned}
$$

- La varianza non è altro che il valore medio di $(X - E(X))^2$. Allora, si può sfruttare la linearità del valore medio per semplificare la sommatoria da calcolare:

$$
\begin{aligned}
\mathrm{Var}(X) &= E[(X - E(X))^2] \\
&= E[X^2 - 2X\underbrace{E(X)}_{\text{costante}} + \underbrace{(E(X))^2}_{\text{costante}}] \\
&= E(X^2) - 2E(X)E(X) + (E(X))^2 \\
&= E(X^2) - (E(X))^2 \\
&= \sum_{i=1}^{4} x_i^2 p(x_i) - \left(\frac{3}{2}\right)^2 \\
&= 0^2 p(0) + 1^2 p(1) + 2^2 p(2) + 3^2 p(3) - \frac{9}{4} \\
&= \frac{3}{8} + 4 \cdot \frac{3}{8} + 9 \cdot \frac{1}{8} - \frac{9}{4} \\
&= \frac{3 + 12 + 9 - 18}{8} \\
&= \frac{6}{8} = \frac{3}{4}
\end{aligned}
$$

- Per variabili aleatorie indipendenti, sfruttando la formula

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2$$

(ricavata nella soluzione precedente) e le proprietà del valore medio, si dimostra che la varianza della somma è uguale alla somma delle varianze:

$$\begin{aligned}
\mathrm{Var}(X+Y) &= E((X+Y)^2) - (E(X+Y))^2 \\
&= E(X^2 + 2XY + Y^2) - (E(X) + E(Y))^2 \\
&= E(X^2) + 2\underbrace{E(XY)}_{\substack{=E(X)E(Y) \\ \text{perché } X,Y \text{ indipendenti}}} + E(Y^2) - (E(X))^2 - 2E(X)E(Y) - (E(Y))^2 \\
&= E(X^2) - (E(X))^2 + E(Y^2) - (E(Y))^2 \\
&\quad + 2E(X)E(Y) - 2E(X)E(Y) \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y)
\end{aligned}$$

Per utilizzare questa proprietà, si definiscono innanzitutto tre variabili aleatorie $X_1, X_2, X_3$, corrispondenti agli esiti dei singoli lanci, ciascuna delle quali assume con uguale probabilità i valori 1 (testa) e 0 (croce):

$$p_i(1) = p_i(0) = \frac{1}{2} \quad i = 1, 2, 3$$

Si calcolano poi il valore medio e la varianza di queste variabili:

$$E(X_i) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}$$

$$\begin{aligned}
\mathrm{Var}(X_i) &= E(X_i^2) - (E(X_i))^2 \\
&= 0^2 \cdot \frac{1}{2} + 1^2 \cdot \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}
\end{aligned}$$

Infine, siccome il numero di teste ottenute nei tre lanci è $X = X_1 + X_2 + X_3$, e $X_1, X_2, X_3$ sono indipendenti (ciascun lancio non dà informazioni sugli altri), la varianza di $X$ è data da:

$$\mathrm{Var}(X) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \mathrm{Var}(X_3) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$$

# 7 Proprietà delle densità discrete

In generale, una funzione $p(x)$ è una densità discreta se e solo se:

$$\begin{gathered}
p(x_i) \geq 0 \quad \forall x_i \\
\sum_{x_i} p(x_i) = 1
\end{gathered}$$

Entrambe queste proprietà derivano dagli assiomi della probabilità ($p(x_i)$ è appunto la probabilità di un evento, $\{X = x_i\}$):

- una mappa di probabilità assume valori in $[0, 1]$, quindi $\geq 0$;
- $\sum_{x_i} p(x_i)$ somma le probabilità di tutti gli eventi elementari $\{X = x_i\}$, perciò è la probabilità dello spazio totale, e deve allora valere 1.

### 7.1 Esercizio di verifica di queste proprietà

Data la funzione

$$p(x) = \frac{x+3}{15} \quad x = 1, 2, 3$$

verificare se essa è la densità di una qualche variabile aleatoria discreta $X$.

- La funzione assume valori $\geq 0$:

$$p(1) = \frac{4}{15} \geq 0 \qquad p(2) = \frac{5}{15} = \frac{1}{3} \geq 0 \qquad p(3) = \frac{6}{15} = \frac{2}{5} \geq 0$$

- La somma dei valori assunti è 1:

$$\sum_{i=1}^{3} f(x_i) = \frac{4+5+6}{15} = \frac{15}{15} = 1$$

  (dove $x_1 = 1,\ x_2 = 2,\ x_3 = 3$).

Quindi $p(x)$ è una densità di probabilità discreta.

## 8 Continuità della funzione di ripartizione

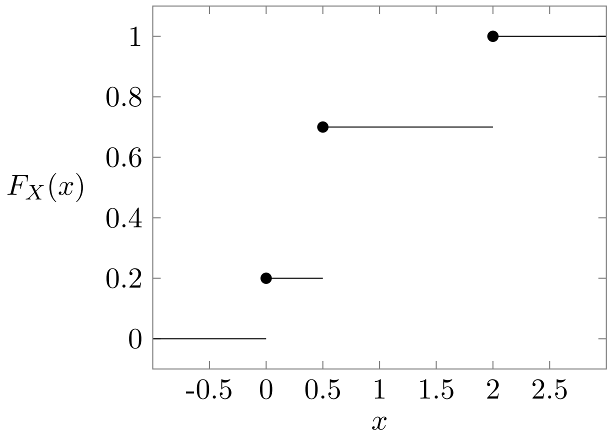
La funzione di ripartizione $F_X(x)$ di una variabile aleatoria discreta $X$ è *continua da destra.*

Infatti, essa è definita da

$$F_X(x) = P\{X \leq x\} = \sum_{x_i \leq x} p(x_i)$$

e allora:

- per gli $x$ nei quali $p(x) = 0$, $F_x(x)$ è costante, e quindi continua;
- gli $x$ dove $p(x) > 0$ provocano invece un salto, e, siccome la funzione di ripartizione è la somma dei $p(x_i)$ per $x_i$ minore *o uguale* a $x$, $F_X(x)$ assume già il valore "successivo" al salto, e in seguito rimarrà costante (fino eventualmente al prossimo salto), quindi è continua da destra (ma non da sinistra) in $x$.




# 9 Esercizio: calcolo della densità dalla funzione di ripartizione

Data la funzione di ripartizione

$$F_X(x) = \begin{cases} 0 & x < -2 \\ 0.2 & -2 \leq x < 0 \\ 0.7 & 0 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

della variabile aleatoria discreta $X$, determinare la densità discreta $p(x)$.

Per definizione, gli unici punti in cui $p(x) > 0$ saranno quelli in cui la funzione di ripartizione ha dei salti, e i valori di $p(x)$ saranno appunto le ampiezze di tali salti:

$$p(x) = \begin{cases} 0.2 - 0 & x = -2 \\ 0.7 - 0.2 & x = 0 \\ 1 - 0.7 & x = 2 \\ 0 & \text{altrimenti} \end{cases} = \begin{cases} 0.2 & x = -2 \\ 0.5 & x = 0 \\ 0.3 & x = 2 \\ 0 & \text{altrimenti} \end{cases}$$

Per confermare la soluzione, si osserva che la $p(x)$ calcolata soddisfa le proprietà della densità discreta:

- $p(x_i) \geq 0 \quad \forall x_i$
- $\sum_{x_i} p(x_i) = 0.2 + 0.5 + 0.3 = 1$