Calcolo delle Probabilità e Statistica 2021/2022

Variabili aleatorie discrete

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

1 Introduzione

In questo capitolo studiamo una particolare classe di variabili aleatorie, le *variabili aleatorie discrete*. In breve, una variabile aleatoria si dice discreta se assume un numero finito (o al più un'infinità numerabile) di valori. Prima di dare la definizione vera e propria di variabile aleatoria discreta, è necessario introdurre la nozione di *densità discreta*.

Definizione 1.1. *Sia $(\Omega, \mathbb{P})$ uno spazio di probabilità e X una variabile aleatoria. La funzione $p_X : \mathbb{R} \to [0,1]$, data da*

$$p_X(x) \;=\; \mathbb{P}(X = x), \qquad \forall\, x \in \mathbb{R},$$

si chiama ***densità discreta*** *o* ***funzione di massa di probabilità*** *o* ***PMF***[a] *di X.*

[a] Dall'inglese *probability mass function*.

Si noti che $p_X(x)$ è la *probabilità* che la variabile aleatoria X assuma il valore x. Per tale ragione, $p_X(x)$ verifica necessariamente le disuguaglianze

$$0 \;\le\; p_X(x) \;\le\; 1, \qquad \forall\, x \in \mathbb{R}.$$

La definizione di *variabile aleatoria discreta* fa intervenire la funzione p_X.

Definizione 1.2. *Sia $(\Omega, \mathbb{P})$ uno spazio di probabilità e X una variabile aleatoria. Si dice che X è una* ***variabile aleatoria discreta*** *(in breve* ***v.a.d.****) se esiste un sottoinsieme $\mathcal{S}_X$ di $\mathbb{R}$, finito o al più infinito numerabile, quindi*

$$\mathcal{S}_X \;=\; \{x_1, \dots, x_n\} \qquad \textit{oppure} \qquad \mathcal{S}_X \;=\; \{x_1, \dots, x_i, \dots\},$$

tale che[a]

$$p_X(x_i) \;>\; 0 \qquad e \qquad \sum_i p_X(x_i) \;=\; 1. \tag{1.1}$$

L'insieme $\mathcal{S}_X$ si chiama ***supporto*** *di X.*

[a] $\sum_i p_X(x_i) = 1$ è una scrittura abbreviata per

$$\sum_{i=1}^{n} p_X(x_i) \;=\; 1 \qquad \text{(caso in cui } \mathcal{S}_X \text{ è finito)}$$
$$\sum_{i=1}^{+\infty} p_X(x_i) \;=\; 1 \qquad \text{(caso in cui } \mathcal{S}_X \text{ è infinito numerabile)}$$

Osservazione 1. *La (1.1) equivale a dire che la variabile aleatoria X assume con probabilità positiva tutti e soli i valori in $\mathcal{S}_X$. In particolare, X assume il valore x_i con probabilità $p_X(x_i) > 0$.*

Tabella della densità discreta. Nel caso in cui $\mathcal{S}_X$ sia un insieme finito, quindi

$$\mathcal{S}_X \;=\; \{x_1, \dots, x_n\},$$

si riportano i valori di p_X in una *tabella*[1]*:*

$$\begin{array}{c|c|c|c|c} X & x_1 & x_2 & \cdots & x_n \\ \hline p_X & p_X(x_1) & p_X(x_2) & \cdots & p_X(x_n) \end{array} \tag{1.2}$$

Esercizio 1.1. *Si lanciano due dadi. Sia*

$$X \;=\; \text{“}\textit{minimo tra i due risultati}\text{”}$$

Mostrare che X è discreta e determinare p_X.

Osservazione. *Quando si chiede di determinare la densità discreta p_X, è sufficiente fornire i valori di p_X per $x \in \mathcal{S}_X$. In particolare, se $\mathcal{S}_X$ è finito, è sufficiente fornire la tabella* (1.2) *della densità discreta.*

Soluzione dell'Esercizio 1.1. Sappiamo che uno spazio di probabilità $(\Omega, \mathbb{P})$ che descrive questo esperimento aleatorio è

$$\Omega \;=\; \big\{(1,1),(1,2),(1,3),\dots,(5,6),(6,6)\big\} \;=\; \{1,2,3,4,5,6\}^2$$

e $\mathbb{P}$ probabilità uniforme, quindi

$$\mathbb{P}(\{(\omega_1,\omega_2)\}) \;=\; \frac{1}{36}, \qquad \forall\,(\omega_1,\omega_2) \in \Omega.$$

La variabile aleatoria X è dunque rappresentata dalla funzione $X\colon \Omega \to \mathbb{R}$ data da

$$X(\omega_1,\omega_2) \;=\; \min(\omega_1,\omega_2), \qquad \forall\,(\omega_1,\omega_2) \in \Omega,$$

dove $\min(\omega_1,\omega_2)$ è il valore minimo tra ω_1 e ω_2. Ad esempio

$$\begin{aligned} X(1,1) &= \min(1,1) \;=\; 1, \\ X(1,2) &= \min(1,2) \;=\; 1, \\ X(1,3) &= \min(1,3) \;=\; 1, \\ &\;\vdots \\ X(5,6) &= \min(5,6) \;=\; 5, \\ X(6,6) &= \min(6,6) \;=\; 6. \end{aligned}$$

Determiniamo la densità discreta di X. Dalla definizione di p_X, abbiamo che

$$p_X(x) \;=\; \mathbb{P}(X = x).$$

[1] In genere nella tabella si riportano solo i valori di p_X per $x \in \mathcal{S}_X$. Tuttavia, come vedremo, ci sono alcuni casi in cui risulta naturale riportare anche il valore di p_X in corrispondenza di qualche $x \notin \mathcal{S}_X$ (per tali x risulta chiaramente $p_X(x) = 0$).

L'evento $\{X = x\}$ è dato da

$$\{X = x\} \;=\; \{(\omega_1, \omega_2) \in \Omega\colon X(\omega_1, \omega_2) = x\} \;=\; \{(\omega_1, \omega_2) \in \Omega\colon \min(\omega_1, \omega_2) = x\}.$$

Quindi, se $x \notin \{1, 2, 3, 4, 5, 6\}$ è chiaro che

$$\{X = x\} \;=\; \emptyset \qquad \Longrightarrow \qquad p_X(x) \;=\; \mathbb{P}(\emptyset) \;=\; 0.$$

Al contrario, se $x \in \{1, 2, 3, 4, 5, 6\}$ si ha:

$$\begin{aligned}
\{X = 1\} &= \big\{(1,1),(1,2),(2,1),(1,3),(3,1),(1,4),(4,1),(1,5),(5,1),(1,6),(6,1)\big\},\\
\{X = 2\} &= \big\{(2,2),(2,3),(3,2),(2,4),(4,2),(2,5),(5,2),(2,6),(6,2)\big\},\\
\{X = 3\} &= \big\{(3,3),(3,4),(4,3),(3,5),(5,3),(3,6),(6,3)\big\},\\
\{X = 4\} &= \big\{(4,4),(4,5),(5,4),(4,6),(6,4)\big\},\\
\{X = 5\} &= \big\{(5,5),(5,6),(6,5)\big\},\\
\{X = 6\} &= \big\{(6,6)\big\}
\end{aligned}$$

Quindi

$$\begin{aligned}
p_X(1) &= \frac{11}{36},\\
p_X(2) &= \frac{9}{36},\\
p_X(3) &= \frac{7}{36},\\
p_X(4) &= \frac{5}{36},\\
p_X(5) &= \frac{3}{36},\\
p_X(6) &= \frac{1}{36}.
\end{aligned}$$

In conclusione, X è una variabile aleatoria discreta[2] con supporto $\mathcal{S}_X = \{1, 2, 3, 4, 5, 6\}$ e densità discreta data da

X	1	2	3	4	5	6
p_X	$\frac{11}{36}$	$\frac{9}{36}$	$\frac{7}{36}$	$\frac{5}{36}$	$\frac{3}{38}$	$\frac{1}{36}$

$\square$

Variabili aleatorie costanti. Sia a un numero reale e X la variabile aleatoria costante uguale ad a. Allora X è una variabile aleatoria discreta con supporto $\mathcal{S}_X = \{a\}$ e densità discreta

X	a
p_X	1

[2]Infatti $p_X(1)$ >0, …, $p_X(6) > 0$ e $\sum_{i=1}^{6} p_X(i) = 1$.

Variabili aleatorie indicatrici. Sia A un evento e $X = 1_A$ la variabile aleatoria indicatrice relativa all'evento A. Allora X è una variabile aleatoria discreta con supporto $\mathcal{S}_X = \{0,1\}$ e densità discreta

X	0	1
p_X	$1 - \mathbb{P}(A)$	$\mathbb{P}(A)$

2 Caratterizzazione delle variabili aleatorie discrete

Sia X una variabile aleatoria discreta. Che relazione esiste tra p_X e la distribuzione di X, che abbiamo indicato con $\mathbb{P}_X$? Che relazione c'è invece tra p_X e F_X, la funzione di ripartizione di X? Le risposte a queste domande sono fornite dal seguente teorema.

Teorema 2.1. *Sia $(\Omega, \mathbb{P})$ uno spazio di probabilità e X una variabile aleatoria. Le seguenti affermazioni sono equivalenti tra loro:*

1) *X è una **variabile aleatoria discreta** (con densità discreta p_X e supporto $\mathcal{S}_X = \{x_1, x_2, \ldots\}$, finito o al più infinito numerabile).*

2) *F_X è una funzione **costante a tratti**: F_X è una funzione costante tranne nei punti $x_1, x_2, \ldots$ di $\mathcal{S}_X$, in cui F_X salta (verso l'alto) con ampiezza del salto pari a*

$$F_X(x_i) - F_X(x_i-) \;=\; p_X(x_i).$$

Quindi F_X è data dalla seguente formula:

$$F_X(x) \;=\; \sum_{x_i \le x} p_X(x_i), \qquad \forall\, x \in \mathbb{R}. \tag{2.1}$$

3) *$\mathbb{P}_X$, la distribuzione di X, è **concentrata nei punti** $x_1, x_2, \ldots$ **di** $\mathcal{S}_X$:*

$$\mathbb{P}_X \;=\; \sum_i p_X(x_i)\, \delta_{x_i},$$

dove δ_{x_i} è la delta di Dirac in x_i.

Infine, vale la formula

$$\mathbb{P}(X \in B) \;=\; \sum_{x_i \in B} p_X(x_i), \qquad \forall\, B \subset \mathbb{R}. \tag{2.2}$$

Non riportiamo la dimostrazione del Teorema 2.1. Notiamo solamente che la formula (2.1), la quale fornisce il valore della funzione di ripartizione in $x \in \mathbb{R}$, è un caso particolare della formula (2.2), ricordando che per definizione di F_X si ha

$$F_X(x) \;=\; \mathbb{P}(X \le x) \;=\; \mathbb{P}(X \in (-\infty, x]), \qquad \forall\, x \in \mathbb{R}.$$

Esercizio 2.1. *Sia* $G\colon \mathbb{R} \to [0,1]$ *una funzione data da*

$$G(x) \;=\; \begin{cases} 0, & x < 0, \\ 1/2, & 0 \le x < 1, \\ 2/3, & 1 \le x < 2, \\ 11/12, & 2 \le x < 3, \\ 1, & x \ge 3. \end{cases}$$

(a) *Mostrare che* G *è una funzione di ripartizione.*

Sia dunque X *una variabile aleatoria con funzione di ripartizione* $F_X = G$.

(b) *Mostrare che* X *è discreta. Determinare supporto e densità discreta di* X.

(c) *Trovare* $\mathbb{P}_X$, *la distribuzione di* X.

(d) *Calcolare* $\mathbb{P}(X > 1/2)$, $\mathbb{P}(2 < X \le 4)$, $\mathbb{P}(1 < X < 2)$, $\mathbb{P}(X < 3)$.

(e) *Mostrare che* $Y = (X-2)^2$ *è una variabile aleatoria discreta. Determinare* $\mathcal{S}_Y$ *e* p_Y.

Soluzione.

(a) G è una funzione di ripartizione, infatti G verifica le seguenti proprietà:

1) G è monotona crescente.
2) G è continua a destra.
3) $\lim_{x\to-\infty} G(x) = 0$.
4) $\lim_{x\to+\infty} G(x) = 1$.

(b) Dato che G è costante a tratti, segue direttamente dal Teorema 2.1 che X è una variabile aleatoria discreta. Inoltre, dal Teorema 2.1 sappiamo che i punti di salto di G sono gli elementi del supporto $\mathcal{S}_X$ di X, mentre l'ampiezza di ogni salto è la probabilità che X assuma quel valore. Perciò, $\mathcal{S}_X = \{0,1,2,3\}$ e

X	0	1	2	3
p_X	1/2	1/6	1/4	1/12

(c) Dal Teorema 2.1 si ha che

$$\mathbb{P}_X \;=\; \sum_i p_X(x_i)\,\delta_{x_i} \;=\; \frac{1}{2}\,\delta_0 + \frac{1}{6}\,\delta_1 + \frac{1}{4}\,\delta_2 + \frac{1}{12}\,\delta_3.$$

(d) Per la formula (2.2) si ha che:

$$\mathbb{P}(X > 1/2) \;=\; \sum_{x_i > 1/2} p_X(x_i) \;=\; p_X(1) + p_X(2) + p_X(3) \;=\; 1 - p_X(0) \;=\; \frac{1}{2}$$

$$\begin{aligned}
\mathbb{P}(2 < X \leq 4) &= \sum_{2<x_i\leq 4} p_X(x_i) = p_X(3) = \frac{1}{12} \\
\mathbb{P}(1 < X < 2) &= \sum_{1<x_i<2} p_X(x_i) = 0, \\
\mathbb{P}(X < 3) &= \sum_{x_i<3} p_X(x_i) = p_X(0) + p_X(1) + p_X(2) = 1 - p_X(3) = \frac{11}{12}.
\end{aligned}$$

Un modo alternativo per calcolare queste probabilità è tramite la funzione di ripartizione:

$$\begin{aligned}
\mathbb{P}(X > 1/2) &= 1 - \mathbb{P}(X \leq 1/2) = 1 - F_X(1/2) = \frac{1}{2}, \\
\mathbb{P}(2 < X \leq 4) &= F_X(4) - F_X(2) = \frac{1}{12}, \\
\mathbb{P}(1 < X < 2) &= F_X(2-) - F_X(1) = 0, \\
\mathbb{P}(X < 3) &= F_X(3-) = \frac{11}{12}.
\end{aligned}$$

(e) Determiniamo p_Y. Dato che $p_Y(y) = \mathbb{P}(Y = y)$, iniziamo col determinare l'evento $\{Y = y\}$ al variare di $y \in \mathbb{R}$:

$$\{Y = y\} = \{(X-2)^2 = y\}.$$

Dato che $S_X = \{0, 1, 2, 3\}$, è chiaro che $(X-2)^2$ può essere uguale solo a $0, 1, 4$. In particolare, si ha che

$$\begin{aligned}
\{Y = 0\} &= \{X = 2\}, \\
\{Y = 1\} &= \{X = 1\} \cup \{X = 3\}, \\
\{Y = 4\} &= \{X = 0\}.
\end{aligned}$$

Perciò

$$\begin{aligned}
p_Y(0) &= p_X(2) = \frac{1}{4}, \\
p_Y(1) &= p_X(1) + p_X(3) = \frac{1}{4}, \\
p_Y(4) &= p_X(0) = \frac{1}{2}
\end{aligned}$$

In conclusione, Y è una variabile aleatoria discreta con supporto $\mathcal{S}_Y = \{0, 1, 4\}$ e densità discreta data da

Y	0	1	4
p_Y	1/4	1/4	1/2

□

3 Indici di sintesi di una distribuzione: μ e σ^2

La distribuzione o legge di una variabile aleatoria può essere descritta in maniera sintetica tramite due quantità numeriche, la *media* (o *valore atteso*) e la *varianza*.

La ***media*** è un **indice di posizione**, ovvero indica qual è il valore "centrale" della distribuzione. Essa è una generalizzazione della *media aritmetica* di n numeri reali $x_1, \dots, x_n$:

$$\mu_{\text{Aritm}} \;=\; \frac{x_1 + \dots + x_n}{n}.$$

In questa formula tutti i numeri x_i hanno lo stesso "peso" pari a $\frac{1}{n}$, mentre la media che andremo a definire sarà una *media pesata* (con le probabilità degli x_i).

La ***varianza*** è un **indice di dispersione**, ossia dice quanto la distribuzione si concentra attorno alla media. È una generalizzazione della media aritmetica delle distanze al quadrato degli x_i da μ:

$$\frac{(x_1 - \mu)^2 + \dots + (x_n - \mu)^2}{n}.$$

3.1 Media o valore atteso

Definizione 3.1. *Sia X una variabile aleatoria discreta con supporto $\mathcal{S}_X = \{x_1, x_2, \dots\}$. La **media** (o **valore atteso**) di X è data da*[a]

$$\mathbb{E}[X] \;=\; \sum_i x_i \, p_X(x_i).$$

La media si indica anche con μ oppure μ_X.

[a] Il simbolo $\mathbb{E}$ deriva dall'inglese *expected value* (valore atteso). Segnaliamo inoltre che è ormai di uso comune l'utilizzo delle parentesi quadre anziché tonde nell'espressione $\mathbb{E}[X]$.

Esercizio 3.1.

1) *Sia $a \in \mathbb{R}$ una costante. Mostrare che $\mathbb{E}[a] = a$.*

2) *Sia A un evento. Mostrare che $\mathbb{E}[1_A] = \mathbb{P}(A)$.*

Soluzione.

1) Ricordiamo che a denota ovviamente una costante, ma anche una variabile aleatoria (la variabile aleatoria costante uguale ad a stessa). Come variabile aleatoria, sappiamo che è una variabile aleatoria discreta con supporto $\mathcal{S}_a = \{a\}$ e densità discreta p_a che verifica

$$p_a(a) \;=\; 1.$$

Quindi, dalla definizione di valore atteso, otteniamo

$$\mathbb{E}[a] \;=\; a \, p_a(a) \;=\; a.$$

2) Ricordiamo che la variabile aleatoria 1_A è una variabile aleatoria discreta con supporto $\mathcal{S}_{1_A} = \{0, 1\}$ e densità discreta

1_A	0	1
p_{1_A}	$1 - \mathbb{P}(A)$	$\mathbb{P}(A)$

Dalla definizione di valore atteso, si ha dunque che

$$\mathbb{E}[1_A] \;=\; 0 \cdot p_{1_A}(0) + 1 \cdot p_{1_A}(1) \;=\; p_{1_A}(1) \;=\; \mathbb{P}(A).$$

$\square$

Nel seguito capiterà spesso di dover calcolare il valore atteso di una funzione di X:

$$Y \;=\; h(X).$$

Risulta dunque particolarmente utile il seguente risultato.

Teorema 3.1. *Sia X una variabile aleatoria discreta con densità discreta p_X e supporto $\mathcal{S}_X = \{x_1, x_2, \dots\}$. Inoltre, siano $h\colon \mathbb{R} \to \mathbb{R}$ e $Y = h(X)$. Allora*

$$\mathbb{E}[Y] \;=\; \mathbb{E}[h(X)] \;=\; \sum_i h(x_i)\, p_X(x_i).$$

Dimostrazione. Supponiamo per semplicità che $\mathcal{S}_X$ sia finito, quindi $\mathcal{S}_X = \{x_1, \dots, x_n\}$. Segue allora che anche il supporto di Y, indicato con $\mathcal{S}_Y$, è finito. Si noti infatti che

$$\mathcal{S}_Y \;= \big\{y \in \mathbb{R}\colon y = h(x_i) \text{ per qualche } x_i \in \mathcal{S}_X\big\}.$$

Perciò $\mathcal{S}_Y = \{y_1, \dots, y_m\}$ e, necessariamente, $m \le n$. Per ogni $j = 1, \dots, m$, poniamo

$$\mathcal{S}_{y_j} \;:=\; \big\{x \in \mathcal{S}_X\colon h(x) = y_j\big\}.$$

Si noti che gli insiemi $\mathcal{S}_{y_1}, \dots, \mathcal{S}_{y_m}$ sono disgiunti e la loro unione è uguale a $\mathcal{S}_X$, ossia $\mathcal{S}_{y_1}, \dots, \mathcal{S}_{y_m}$ sono una partizione di $\mathcal{S}_X$.

Possiamo dunque scrivere $\mathbb{E}[Y]$ come segue, partendo dalla sua definizione, (nelle tre uguaglianze intermedie sono evidenziate in blu le differenze rispetto alla formula precedente)

$$\begin{aligned}
\mathbb{E}[Y] \;&=\; \sum_{j=1}^m y_j\, p_Y(y_j) \;=\; \sum_{j=1}^m y_j \Bigg(\sum_{\substack{i=1,\dots,n\\ h(x_i)=y_j}} p_X(x_i)\Bigg) \;=\; \sum_{j=1}^m y_j \Bigg(\sum_{\substack{i=1,\dots,n\\ x_i\in\mathcal{S}_{y_j}}} p_X(x_i)\Bigg) \\
&=\; \sum_{j=1}^m \Bigg(\sum_{\substack{i=1,\dots,n\\ x_i\in\mathcal{S}_{y_j}}} y_j\, p_X(x_i)\Bigg) \;=\; \sum_{j=1}^m \Bigg(\sum_{\substack{i=1,\dots,n\\ x_i\in\mathcal{S}_{y_j}}} h(x_i)\, p_X(x_i)\Bigg) \;=\; \sum_{i=1}^n h(x_i)\, p_X(x_i),
\end{aligned}$$

dove l'ultima uguaglianza segue dal fatto che $\mathcal{S}_{y_1}, \dots, \mathcal{S}_{y_m}$ sono una partizione di $\mathcal{S}_X$, quindi ogni x_i compare in una e una sola sommatoria interna. $\square$

Un'importante proprietà del valore atteso è la *linearità*.

Teorema 3.2 (Linearità del valore atteso). *Sia X una variabile aleatoria discreta. Inoltre, siano $h\colon \mathbb{R} \to \mathbb{R}$, $g\colon \mathbb{R} \to \mathbb{R}$ e $a, b \in \mathbb{R}$ costanti. Allora*

$$\mathbb{E}[a\,X + b] \;=\; a\,\mathbb{E}[X] + b \tag{3.1}$$

e, più in generale,

$$\mathbb{E}[a\,h(X) + b\,g(X)] \;=\; a\,\mathbb{E}[h(X)] + b\,\mathbb{E}[g(X)]. \tag{3.2}$$

Dimostrazione. Dimostriamo la formula (3.1), dato che la formula (3.2) si dimostra in modo analogo. Tale formula si dimostra applicando il Teorema 3.1 con $Y = h(X)$, dove

$$h(x) \;=\; a\,x + b, \qquad \forall\, x \in \mathbb{R}.$$

Dal Teorema 3.1 si ha che

$$\mathbb{E}[a\,X + b] \;=\; \sum_i (a\,x_i + b)\,p_X(x_i) \;=\; a \sum_i x_i\,p_X(x_i) + b \sum_i p_X(x_i).$$

Dato che $\sum_i x_i\,p_X(x_i) = \mathbb{E}[X]$ e $\sum_i p_X(x_i) = 1$, si ottiene la formula (3.1).

3.2 Varianza

Definizione 3.2. *Sia X una variabile aleatoria discreta con densità discreta p_X e supporto $\mathcal{S}_X = \{x_1, x_2, \ldots\}$. La **varianza** di X è data da*

$$\mathrm{Var}(X) \;=\; \mathbb{E}[(X - \mathbb{E}[X])^2] \;=\; \sum_i (x_i - \mathbb{E}[X])^2\,p_X(x_i).$$

La varianza si indica anche con σ^2 oppure σ_X^2.

*La radice quadrata della varianza si chiama **deviazione standard** (o **scarto quadratico medio** o anche **scostamento quadratico medio**) e si indica con σ oppure σ_X.*

Osservazione 1. *Se X è una grandezza fisica espressa in una certa unità di misura, allora la deviazione standard ha il vantaggio, a differenza della varianza, di essere espressa nella stessa unità di misura di X (la varianza ha invece come unità di misura il quadrato dell'unità di misura di X).*

Per calcolare la varianza di una variabile aleatoria è utile la seguente formula.

Teorema 3.3. *Sia X una variabile aleatoria discreta con densità discreta p_X e supporto $\mathcal{S}_X = \{x_1, x_2, \dots\}$. Vale che*

$$\mathrm{Var}(X) \;=\; \mathbb{E}[X^2] - \mathbb{E}[X]^2 \;=\; \sum_i x_i^2\, p_X(x_i) - \mathbb{E}[X]^2. \tag{3.3}$$

Dimostrazione. Dimostriamo la prima uguaglianza, dato che la seconda è una conseguenza del Teorema 3.1.
Si ha che

$$\mathrm{Var}(X) \;=\; \mathbb{E}\big[(X - \mathbb{E}[X])^2\big] \;=\; \mathbb{E}\big[X^2 - 2\,X\,\mathbb{E}[X] + \mathbb{E}[X]^2\big].$$

Dalla linearità del valore atteso, si ottiene (si noti che $\mathbb{E}[X]$ è una costante)

$$\mathrm{Var}(X) \;=\; \mathbb{E}[X^2] - 2\,\mathbb{E}[X]^2 + \mathbb{E}[X]^2 \;=\; \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

□

A differenza del valore atteso, la varianza non è lineare. Più precisamente, la varianza possiede le seguenti proprietà.

Teorema 3.4 (Proprietà della varianza). *Siano X una variabile aleatoria discreta e $a, b \in \mathbb{R}$ costanti. Allora*

1) $\mathrm{Var}(X) \geq 0$.

2) $\mathrm{Var}(b) = 0$ *e viceversa: se* $\mathrm{Var}(X) = 0$ *allora X è una variabile aleatoria costante.*

3) $\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$.

Dimostrazione.

1) Per definizione di varianza, si ha che

$$\mathrm{Var}(X) \;=\; \sum_i (x_i - \mathbb{E}[X])^2\, p_X(x_i).$$

Dato che ogni addendo è maggiore o uguale di zero, segue che la somma (ovvero la varianza di X) è anch'essa maggiore o uguale di zero.

2) Dalla formula (3.3) si ha che

$$\mathrm{Var}(b) \;=\; \mathbb{E}[b^2] - \mathbb{E}[b]^2.$$

Ricordando che il valore atteso di una costante è pari alla costante stessa, otteniamo

$$\mathbb{E}[b^2] - \mathbb{E}[b]^2 \;=\; b^2 - b^2 \;=\; 0.$$

Quindi $\mathrm{Var}(b) = 0$.
Viceversa, sia X una generica variabile aleatoria discreta, di cui sappiamo che

$$\mathrm{Var}(X) \;=\; \sum_i (x_i - \mathbb{E}[X])^2\, p_X(x_i) \;=\; 0.$$

Dato che ogni addendo è maggiore o uguale di zero, la somma è nulla se e solo se ciascun addendo è nullo. Quindi

$$(x_i - \mathbb{E}[X])^2\, p_X(x_i) \;=\; 0, \qquad \forall\, i.$$

Tale prodotto è nullo se e solo se $p_X(x_i) = 0$ oppure $(x_i - \mathbb{E}[X])^2 = 0$ (che significa $x_i = \mathbb{E}[X]$). Non è possibile che $p_X(x_i) = 0$ per ogni i, altrimenti non sarebbe vero che $\sum_i p_X(x_i) = 1$. D'altra parte, essendo i valori x_i distinti tra loro, esiste solo un i per cui vale che $x_i = \mathbb{E}[X]$ (si noti che $\mathbb{E}[X]$ è una costante). Quindi il supporto della variabile aleatoria X è costituito da un unico valore, $\mathcal{S}_X = \{\mathbb{E}[X]\}$, da cui segue che X è la variabile aleatoria costante uguale a $\mathbb{E}[X]$.

3) Dalla definizione di varianza si ha che

$$\mathrm{Var}(aX + b) \;=\; \mathbb{E}\big[\big(aX + b - \mathbb{E}[aX + b]\big)\big].$$

Ricordiamo che $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$, quindi

$$\begin{aligned}\mathrm{Var}(aX + b) \;&=\; \mathbb{E}\big[\big(aX + b - a\mathbb{E}[X] - b\big)\big] \\ &=\; \mathbb{E}\big[\big(aX - a\mathbb{E}[X]\big)\big] \;=\; a^2\, \mathbb{E}\big[\big(X - \mathbb{E}[X]\big)\big] \;=\; a^2\, \mathrm{Var}(X).\end{aligned}$$

$\square$

4 Distribuzioni discrete notevoli

In questa sezione vediamo le principali distribuzioni discrete.

Distribuzione uniforme discreta. Sia $\{x_1, \ldots, x_n\}$ un sottoinsieme finito di $\mathbb{R}$. Diciamo che X ha **distribuzione uniforme discreta sull'insieme** $\{x_1, \ldots, x_n\}$ se X è una variabile aleatoria discreta con $\mathcal{S}_X = \{x_1, \ldots, x_n\}$ e densità discreta data da

X	x_1	x_2	$\cdots$	x_n
p_X	$\frac{1}{n}$	$\frac{1}{n}$	$\cdots$	$\frac{1}{n}$

In tal caso scriviamo

$$X \;\sim\; \mathrm{Unif}\,(\{x_1, \ldots, x_n\}).$$

Si noti che

$$\begin{aligned}\mathbb{E}[X] \;&=\; \mu_{\mathrm{Aritm}} \;=\; \frac{x_1 + \cdots + x_n}{n}, \\ \mathrm{Var}(X) \;&=\; \frac{(x_1 - \mathbb{E}[X])^2 + \cdots + (x_n - \mathbb{E}[X])^2}{n}.\end{aligned}$$

Distribuzione di Bernoulli. Sia $0 \leq p \leq 1$. Diciamo che X ha **distribuzione di Bernoulli** (o **bernoulliana**) **di parametro** p se X è una variabile aleatoria discreta con $\mathcal{S}_X = \{0, 1\}$ e densità discreta data da

X	0	1
p_X	$1-p$	p

In tal caso scriviamo

$$X \;\sim\; B(p).$$

Le variabili aleatorie di Bernoulli sono tutte e sole variabili aleatorie indicatrici. Infatti

$$X \;=\; 1_A, \qquad\qquad \text{con } A \;=\; \{X = 1\}.$$

Si noti che

$$\begin{aligned} \mathbb{E}[X] \;&=\; p, \\ \mathrm{Var}(X) \;&=\; p(1-p). \end{aligned}$$

Distribuzione binomiale. Consideriamo ora una generalizzazione della distribuzione bernoulliana: la distribuzione binomiale. Come abbiamo visto, una v.a. X ha distribuzione di Bernoulli se è una v.a. indicatrice di un qualche evento A. In altre parole, se X vale solamente 1 o 0 a seconda che l'evento A si verifichi oppure no. Immaginiamo ora di ripetere n volte l'esperimento aleatorio a cui l'evento A si riferisce, in modo tale che i vari esperimenti siano tra loro "indipendenti". Per ciascun esperimento consideriamo la corrispondente v.a. bernoulliana. Abbiamo dunque n variabili aleatorie bernoulliane:

$$X_1 \;\sim\; B(p), \qquad X_2 \;\sim\; B(p), \qquad \ldots \qquad X_n \;\sim\; B(p).$$

Consideriamo la seguente variabile aleatoria:

$$X \;=\; \text{"n}^{\text{o}} \text{ di successi negli } n \text{ esperimenti"},$$

dove *successo* significa che l'evento A si è verificato, quindi X è il numero di volte che A si è verificato negli n esperimenti. Notiamo che

$$X \;=\; X_1 + \cdots + X_n.$$

Vediamo un esempio in cui si verifica questa situazione. Come si vedrà nell'esempio, X ha una distribuzione particolare che si chiama *distribuzione binomiale di parametri n e p*.

Esempio 4.1. *Si consideri un'urna contenente b palline bianche ed r palline rosse. Si effettuano n estrazioni* con reimmissione. *Sia*[a]

$$X \;=\; \text{“}n^o \text{ di palline bianche estratte”}.$$

Mostrare che X è una variabile aleatoria discreta e determinarne supporto e densità discreta.

[a]Si noti che $X = X_1 + \cdots + X_n$, dove

$$X_i \;=\; \text{“vale 1 se si estrae una pallina bianca all'}i\text{-esima estrazione, 0 altrimenti”},$$

per ogni $i = 1, \ldots, n$. Quindi $X_i \sim B(p)$, dove $p = \frac{b}{b+r}$.

Soluzione. Chiaramente $\mathcal{S}_X = \{0, 1, 2, \ldots, n\}$. Resta dunque da calcolare $p_X(k)$ per $k = 0, 1, 2, \ldots, n$. Sia

$$A_k \;=\; \{X = k\}, \qquad \text{per ogni } k = 0, 1, 2, \ldots, n.$$

Si noti che l'evento A_k ha la seguente interpretazione:

$$A_k \;=\; \text{“si estraggono } k \text{ palline bianche ed } n-k \text{ palline rosse”}.$$

Abbiamo già calcolato la probabilità di A_k nel capitolo riguardante il *calcolo combinatorio*. Ricordiamo che

$$\mathbb{P}(A_k) \;=\; \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n,$$

dove $p = \frac{b}{b+r}$ è la probabilità di estrarre una pallina bianca in una singola estrazione. Quindi

$$p_X(k) \;=\; \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n.$$

Come seguirà dalla definizione, X ha distribuzione binomiale di parametri n e p, ovvero $X \sim B(n, p)$. $\square$

Siano $n \in \mathbb{N}$ e $0 \leq p \leq 1$. Diciamo che X ha **distribuzione binomiale di parametri n e p** se X è una variabile aleatoria discreta con $\mathcal{S}_X = \{0, 1, 2, \ldots, n\}$ e densità discreta data da

$$p_X(k) = \mathbb{P}(X = k) = \binom{n}{k} p^k \, (1-p)^{n-k}, \quad \text{per ogni } k = 0, \ldots, n,$$

cioè

X	0	1	$\cdots$	$n-1$	n
p_X	$\binom{n}{0}(1-p)^n$	$\binom{n}{1} p\,(1-p)^{n-1}$	$\cdots$	$\binom{n}{n-1} p^{n-1}(1-p)$	$\binom{n}{n} p^n$

In tal caso scriviamo

$$X \;\sim\; B(n, p).$$

Si noti che quando $n = 1$, X ha distribuzione di Bernoulli, ovvero $B(1, p) = B(p)$.

Notiamo inoltre che, per la formula del binomio di Newton, vale che

$$\sum_{k=0}^{n} p_X(k) \;=\; \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \;=\; 1, \qquad \forall\, 0 \le p \le 1,\; n \in \mathbb{N}.$$

Questo dimostra che p_X è effettivamente una densità discreta.

Proposizione 4.1. *Siano $0 \le p \le 1$, $n \in \mathbb{N}$ e $X \sim B(n,p)$. Allora*

$$\begin{aligned} \mathbb{E}[X] &= np, \\ \mathrm{Var}(X) &= np(1-p). \end{aligned}$$

Dimostrazione.

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=0}^{n} k\, p_X(k) = \sum_{k=1}^{n} k\, p_X(k) = \sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k} = \sum_{k=1}^{n} \frac{n!}{(k-1)!\,(n-k)!} p^k (1-p)^{n-k} \\
&= n\,p \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!\,(n-k)!} p^{k-1} (1-p)^{n-k} \\
&\underset{h=k-1}{\overset{\uparrow}{=}} n\,p \sum_{h=0}^{n-1} \frac{(n-1)!}{h!\,(n-1-h)!} p^h (1-p)^{n-1-h} \underset{\substack{\text{form. binom. Newton:} \\ \sum_{h=0}^{n-1}(\cdots)=(p+(1-p))^{n-1}=1}}{\overset{\uparrow}{=}} n\,p.
\end{aligned}$$

Per quanto riguarda la varianza, dato che $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X^2] - n^2p^2$, resta da calcolare $\mathbb{E}[X^2]$. Inoltre, $\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] = \mathbb{E}[X(X-1)] + np$, quindi dobbiamo calcolare $\mathbb{E}[X(X-1)]$. Si ha che

$$\begin{aligned}
\mathbb{E}[X(X-1)] &= \sum_{k=0}^{n} k\,(k-1)\, p_X(k) = \sum_{k=2}^{n} k\,(k-1)\, p_X(k) \\
&= \sum_{k=2}^{n} k\,(k-1) \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=2}^{n} k\,(k-1) \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=2}^{n} \frac{n!}{(k-2)!\,(n-k)!} p^k (1-p)^{n-k} \\
&= n\,(n-1)\,p^2 \sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!\,(n-k)!} p^{k-2} (1-p)^{n-k} \\
&\underset{h=k-2}{\overset{\uparrow}{=}} n\,(n-1)\,p^2 \sum_{h=0}^{n-2} \frac{(n-2)!}{h!\,(n-2-h)!} p^h (1-p)^{n-2-h} \\
&\underset{\substack{\text{form. binom. Newton:} \\ \sum_{h=0}^{n-2}(\cdots)=(p+(1-p))^{n-2}=1}}{\overset{\uparrow}{=}} n\,(n-1)\,p^2.
\end{aligned}$$

Quindi $\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + np = n(n-1)p^2 + np$, perciò $\mathrm{Var}(X) = \mathbb{E}[X^2] - n^2p^2 = np(1-p)$. $\square$

Distribuzione di Poisson. La distribuzione di Poisson che ora introduciamo è un "caso limite" della distribuzione binomiale, che si ottiene a partire dalla distribuzione binomiale quando

$$n \to +\infty, \qquad p \to 0, \qquad np \to \lambda,$$

dove $\lambda > 0$ è una costante fissata. Anche se in modo impreciso, possiamo dire più semplicemente che se $X \sim B(n,p)$, *n è "molto grande"* e *p è "molto piccolo"*, allora X ha approssimativamente distribuzione di Poisson di parametro $\lambda = np$. Questa osservazione può essere d'aiuto in certi casi, dato che risulta più facile fare i conti con la distribuzione di Poisson rispetto alla distribuzione binomiale.
Diciamo che X ha **distribuzione di Poisson di parametro** λ se X è una variabile aleatoria discreta con $\mathcal{S}_X = \{0,1,2,\dots\}$ e densità discreta data da

$$p_X(k) = \mathbb{P}(X = k) = \mathrm{e}^{-\lambda}\,\frac{\lambda^k}{k!}, \qquad \text{per ogni } k = 0,1,2,\dots$$

In tal caso scriviamo

$$X \sim \mathrm{Poisson}(\lambda).$$

Si ricordi che

$$\sum_{k=0}^{+\infty} \frac{x^k}{k!} = \mathrm{e}^x, \qquad \forall\, x \in \mathbb{R}. \tag{4.1}$$

Quindi si verifica che

$$\sum_{k=0}^{+\infty} p_X(k) = \sum_{k=0}^{+\infty} \mathrm{e}^{-\lambda}\,\frac{\lambda^k}{k!} = 1.$$

Proposizione 4.2. *Siano $\lambda > 0$ e $X \sim \mathrm{Poisson}(\lambda)$. Allora*

$$\begin{aligned} \mathbb{E}[X] &= \lambda, \\ \mathrm{Var}(X) &= \lambda. \end{aligned}$$

Dimostrazione.

$$\begin{aligned} \mathbb{E}[X] &= \sum_{k=0}^{+\infty} k\,\mathrm{e}^{-\lambda}\,\frac{\lambda^k}{k!} = \sum_{k=1}^{+\infty} k\,\mathrm{e}^{-\lambda}\,\frac{\lambda^k}{k!} = \mathrm{e}^{-\lambda}\sum_{k=1}^{+\infty} k\,\frac{\lambda^k}{k!} = \mathrm{e}^{-\lambda}\sum_{k=1}^{+\infty} \frac{\lambda^k}{(k-1)!} \\ &= \lambda\,\mathrm{e}^{-\lambda}\sum_{k=1}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!} \\ &\underset{\substack{\uparrow \\ h=k-1}}{=} \lambda\,\mathrm{e}^{-\lambda}\sum_{h=0}^{+\infty} \frac{\lambda^h}{h!} \underset{\substack{\uparrow \\ (4.1)}}{=} \lambda. \end{aligned}$$

Per quanto riguarda la varianza, ricordiamo la formula

$$\mathrm{Var}(X) \;=\; \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X^2] - \lambda^2.$$

Resta dunque da calcolare $\mathbb{E}[X^2]$. Per la linearità del valore atteso (si usa in particolare la formula (3.2) con $h(X) = X(X-1)$, $g(X) = X$, $a = 1$ e $b = 1$), abbiamo che

$$\mathbb{E}[X^2] \;=\; \mathbb{E}[X(X-1) + X] \;=\; \mathbb{E}[X(X-1)] + \mathbb{E}[X] \;=\; \mathbb{E}[X(X-1)] + \lambda,$$

Perciò, è sufficiente calcolare $\mathbb{E}[X(X-1)]$. Si ha che

$$\begin{aligned}\mathbb{E}[X(X-1)] \;&=\; \sum_{k=0}^{+\infty} k\,(k-1)\,\mathrm{e}^{-\lambda}\,\frac{\lambda^k}{k!} \;=\; \sum_{k=2}^{+\infty} k\,(k-1)\,\mathrm{e}^{-\lambda}\,\frac{\lambda^k}{k!} \;=\; \mathrm{e}^{-\lambda}\sum_{k=2}^{+\infty} k\,(k-1)\,\frac{\lambda^k}{k!} \\ &=\; \mathrm{e}^{-\lambda}\sum_{k=2}^{+\infty}\frac{\lambda^k}{(k-2)!} \;=\; \lambda^2\,\mathrm{e}^{-\lambda}\sum_{k=2}^{+\infty}\frac{\lambda^{k-2}}{(k-2)!} \;\underset{\substack{\uparrow\\ h=k-2}}{=}\; \lambda^2\,\mathrm{e}^{-\lambda}\sum_{h=0}^{+\infty}\frac{\lambda^h}{h!} \;\underset{\substack{\uparrow\\ (4.1)}}{=}\; \lambda^2.\end{aligned}$$

Quindi $\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \lambda = \lambda^2 + \lambda$, perciò $\mathrm{Var}(X) = \mathbb{E}[X^2] - \lambda^2 = \lambda$. $\square$

5 Il problema del giornalaio

Consideriamo il seguente problema del giornalaio, che è un problema di economia riguardante la gestione delle scorte.

Problema 5.1. *Un giornalaio vende quotidiani a € 1.50/copia. Il suo guadagno è di € 0.25/copia*[a]. ***Quante copie conviene al giornalaio avere in edicola?***

- *Il numero di copie richieste al giorno non è costante ma è soggetto ad oscillazioni non prevedibili.*
- *Se il numero di copie è insufficiente, il giornalaio ha un mancato guadagno pari a € 0.25 per ogni copia richiesta dopo che ha esaurito quelle ritirate.*
- *Ad ogni copia ritirata ma non venduta corrisponde invece una perdita di € 1.25.*

[a]Le copie non vendute giorno per giorno non possono essere rese, perdono dunque ogni valore.

Per risolvere questo problema il giornalaio dovrà innanzitutto valutare quali probabilità attribuisce al fatto di poter vendere un numero di copie pari a 0, 1, 2, 3, ecc. A tal proposito, potrebbe essere conveniente raccogliere informazioni o effettuare sperimentazioni per avere una migliore base di giudizio. Per semplicità, supponiamo che il giornalaio decida solamente di osservare cosa accade nei primi 50 giorni. Riportiamo nella tabella che segue i dati riguardanti questo periodo di prova:

N$^{\text{O}}$ COPIE RICHIESTE	N$^{\text{O}}$ GIORNI	FREQUENZA RELATIVA
0	1	$\frac{1}{50}$
1	1	$\frac{1}{50}$
2	3	$\frac{3}{50}$
3	6	$\frac{6}{50}$
4	10	$\frac{10}{50}$
5	11	$\frac{11}{50}$
6	9	$\frac{9}{50}$
7	3	$\frac{3}{50}$
8	3	$\frac{3}{50}$
9	2	$\frac{2}{50}$
10	1	$\frac{1}{50}$

La colonna centrale riporta il numero di giorni (su un totale di cinquanta) in cui il numero totale di copie richieste è stato pari a quanto riportato sulla stessa riga della prima colonna. Si noti che in nessun giorno sono state vendute più di 10 copie. L'ultima colonna riporta la *frequenza relativa*, ovvero la frazione (o percentuale) di giorni in cui è stato venduto un numero di copie pari a 0, 1, 2, 3, ecc.
Come abbiamo già detto, supponiamo per semplicità che il giornalaio abbia ragione di credere che questi dati siano significativi, ovvero che in futuro l'andamento delle richieste non se ne scosterà significativamente. In altre parole, supponiamo che le frequenze relative siano delle buone approssimazioni delle probabilità. Ad esempio

$$\mathbb{P}(\text{“n}^{\text{o}}\text{ di copie richieste è uguale a zero”}) \;\approx\; \frac{1}{50}.$$

Introduciamo dunque la variabile aleatoria

$$X \;=\; \text{“n}^{\text{o}}\text{ di copie richieste”}.$$

Allora è ragionevole ritenere che X abbia densità discreta data da

X	0	1	2	3	4	5	6	7	8	9	10
p_X	$\frac{1}{50}$	$\frac{1}{50}$	$\frac{3}{50}$	$\frac{6}{50}$	$\frac{10}{50}$	$\frac{11}{50}$	$\frac{9}{50}$	$\frac{3}{50}$	$\frac{3}{50}$	$\frac{2}{50}$	$\frac{1}{50}$

Supponiamo ora che il giornalaio decida di acquistare k copie, per qualche $k = 1, \dots, 10$. Consideriamo dunque la variabile aleatoria

$$Y_k \;=\; \text{“guadagno avendo acquistato } k \text{ copie”}.$$

L'obiettivo del giornalaio è trovare il numero k che massimizza il “guadagno atteso”, ovvero

trovare $k = 1, \dots, 10$ tale che $\mathbb{E}[Y_k]$, il “guadagno atteso”, sia massimo.

Mostriamo come si calcola $\mathbb{E}[Y_k]$ nel caso $k = 3$. Consideriamo dunque la variabile aleatoria

$$Y_3 \;=\; \text{“guadagno avendo acquistato 3 copie”}.$$

Si ha che

$$Y_3 \;=\; \begin{cases} -3\cdot 1.25 & X \;=\; 0 \\ -3\cdot 1.25 + 1.50 & X \;=\; 1 \\ -3\cdot 1.25 + 2\cdot 1.50 & X \;=\; 2 \\ -3\cdot 1.25 + 3\cdot 1.50 & X \;\geq\; 3 \end{cases}$$

Quindi Y_3 è una variabile aleatoria discreta con densità

Y_3	-3.75	-2.25	-0.75	0.75
p_{Y_3}	$\mathbb{P}(X=0)=\frac{1}{50}$	$\mathbb{P}(X=1)=\frac{1}{50}$	$\mathbb{P}(X=2)=\frac{3}{50}$	$\mathbb{P}(X\geq 3)=\frac{45}{50}$

Perciò

$$\text{guadagno atteso avendo acquistato 3 copie} \;=\; \mathbb{E}[Y_3] \;=\; 0.51.$$

Ripetendo il ragionamento fatto nel caso $k = 3$ anche per gli altri valori di k, si ottiene la seguente tabella:

k	$\mathbb{E}[Y_k]$
1	0.22
2	0.41
3	**0.51**
4	0.43
5	0.05
6	-0.66
7	-1.64
8	-2.71
9	-3.87
10	-5.09

In conclusione, il guadagno atteso è massimo per $k = 3$. Questo è dunque il numero di copie che al giornalaio conviene avere in edicola.