Capitolo 2

Variabili Aleatorie Discrete

2.1 Variabili aleatorie

Uno dei concetti fondamentali della teoria della probabilità è quello di *variabile aleatoria*. Prima di darne una definizione formale, illustriamo questo concetto mediante alcuni esempi:

- il numero di chiamate effettuate ad un centralino telefonico in un certo intervallo di tempo è una quantità aleatoria ed assume valori diversi in intervalli di tempo diversi (della stessa ampiezza) in dipendenza di circostanze accidentali;

- il tempo percorso da un furgone che, da un magazzino, rifornisce una sede periferica è una quantità aleatoria ed assume valori diversi (pur effettuando lo stesso percorso) in dipendenza delle condizioni di traffico e/o di altri fattori accidentali;

- il tempo che intercorre fra due successivi guasti di una stessa apparecchiatura elettronica è una quantità aleatoria che dipende da vari fattori accidentali.

Si pone il problema di costruire una teoria matematica che consenta lo studio di fenomeni di questo tipo. Si noti che, nonostante la diversità del contenuto reale degli esempi precedenti, essi presentano nella sostanza lo stesso aspetto dal punto di vista matematico. Infatti, in ciascun caso interviene una quantità che, in un modo o nell'altro, descrive i fenomeni che si stanno studiando; sotto l'effetto di circostanze casuali, ciascuna di queste quantità può prendere una varietà di valori diversi, e non si può stabilire in anticipo quale valore la quantità in esame assumerà, perchè essa varia in modo casuale da prova a prova.

Sinonimi di variabile aleatoria sono *numero aleatorio* e *variabile casuale*. Da un punto di vista formale, diamo la seguente definizione di variabile casuale.

Definizione 2.1 Sia $(\Omega, \mathcal{A}, P)$ uno spazio di probabilità. Si dice *variabile aleatoria* un'applicazione $X : \Omega \to \mathbb{R}$ tale che, per ogni $t \in \mathbb{R}$, l'insieme $\{\omega \in \Omega \,:\, X(\omega) \leq t\}$ sia in $\mathcal{A}$

Una variabile aleatoria (v.a.) è dunque una funzione di ω tale che si possa calcolare $P\{\omega \,:\, X(\omega) \leq t\}$, cioè tale che abbia senso calcolare la probabilità che X assuma valori al più uguali a t, per ogni $t \in \mathbb{R}$. Più in generale è fondamentale per le v.a. il calcolo di probabilità del tipo

$P\{\omega \,:\, X(\omega) \in A\}$, dove A è un sottoinsieme di $\mathbb{R}$. Il problema è pertanto quello di studiare l'applicazione:

$$A \to P\{\omega \,:\, X(\omega) \in A\} \tag{2.1}$$

che ad ogni sottoinsieme $A \subseteq \mathbb{R}$ associa la probabilità che X assuma valori appartenenti ad A. Questa applicazione viene chiamata *legge* o *distribuzione* di X.

Nota 2.2 In generale l'applicazione (2.1) potrebbe non essere definita per ogni sottoinsieme $A \subset \mathbb{R}$, cioè può accadere che $\{\omega \,:\, X(\omega) \in A\}$ non sia un evento (in altre parole non sia un elemento della σ-algebra $\mathcal{A}$). Per gli scopi del nostro corso, supporremo nel seguito che condizione richiesta sia sempre verificata.

Se $\{\omega \,:\, X(\omega) \leq a\}$ è un evento, cioè appartiene ad $\mathcal{A}$, per ogni $a \in R$, allora segue:

1. $\{\omega \,:\, X(\omega) > a\}$ è un evento : infatti, in base alla seconda proprietà della definizione di σ-algebra, si ha $\{\omega \,:\, X(\omega) > a\} = \{\omega \,:\, X(\omega) \leq a\}^c$.

2. $\{\omega \,:\, a < X(\omega) \leq b\}$ è un evento per ogni $a, b \in R$, con $a < b$, infatti:
$$\{\omega \,:\, a < X(\omega) \leq b\} = \{\omega \,:\, X(\omega) \leq b\} \cap \{\omega \,:\, X(\omega) > a\}$$
appartiene ad $\mathcal{A}$ in quanto intersezione di due eventi appartenenti ad $\mathcal{A}$.

3. $\{\omega \,:\, X(\omega) = x\}$ è un evento in quanto possiamo scrivere tale insieme come l'intersezione di un numero infinito di eventi
$$\{\omega \,:\, X(\omega) = x\} := \bigcap_n \left\{\omega \,:\, x - \frac{1}{n} < X(\omega) \leq x\right\} \tag{2.2}$$
e quindi appartiene ad $\mathcal{A}$ per la terza proprietà della definizione di σ-algebra.

Nel seguito, per semplicità, scriveremo $\{X \leq t\}$, $\{X \in A\}$ e $\{X \leq x\}$ anzichè rispettivamente $\{\omega \,:\, X(\omega) \leq t\}$, $\{\omega \,:\, X(\omega) \in A\}$ e $\{\omega \,:\, X(\omega) = x\}$. Inoltre indicheremo con Ω_X il codominio di X.

2.2 Variabili aleatorie discrete

Sia $X : \Omega \to \mathbb{R}$ una v.a. ed assumiamo che il codominio Ω_X di X contenga al più un'infinità numerabile di valori, cioè $\Omega_X = \{x_1, \ldots, x_k, \ldots\}$; supponiamo inoltre che l'insieme $\{x_1, \ldots, x_k, \ldots\}$ non abbia punti di accumulazione.

Sia $t \in \mathbb{R}$ e consideriamo l'insieme dei valori di Ω_X al più uguali a t, siano essi $x_1 < \cdots < x_h \leq t$. Poichè gli insiemi $\{X = x_i\}$, $(i = 1, \ldots, h)$, sono eventi in base alla (2.2), essendo il codominio un insieme discreto, possiamo scrivere:

$$\begin{aligned}\{\omega \in \Omega \,:\, X \leq t\} &= \{\omega \in \Omega \,:\, X = x_1\} \cup \cdots \cup \{\omega \in \Omega \,:\, X = x_h\} \\ &= \bigcup_{x_i \leq t} \{\omega \in \Omega \,:\, X = x_i\}\,. \end{aligned} \tag{2.3}$$

Assegnata un v.a. discreta X, consideriamo la funzione $p : \mathbb{R} \to \mathbb{R}^+$ definita da

$$p(x) = P\{X = x\} \ .$$

Ovviamente tale funzione gode delle seguenti proprietà:

$$1. \quad p(x) \geq 0 \qquad \text{per ogni } x \in \mathbb{R} \tag{2.4}$$

$$2. \quad p(x) > 0 \qquad \text{per } x \in \Omega_X \tag{2.5}$$

$$3. \quad \sum_{x \in \Omega_X} p(x) = 1 \ . \tag{2.6}$$

La (2.5) è ovvia; per quanto riguarda la (2.6), osserviamo che gli eventi $\{X = x_i\}$ sono a due a due disgiunti in quanto X è una funzione che associa ad ogni $\omega \in \Omega$ uno ed un solo valore $X(\omega) = x$; inoltre poichè X assume valori solo in Ω_X, ne segue che

$$\bigcup_{i=1}^{\infty} \{X = x_i\} = \{X = x_1\} \cup \cdots \cup \{X = x_i\} \cup \cdots = \{X \in \Omega_X\} = \Omega$$

e pertanto:

$$\sum_{i=1}^{\infty} p(x_i) = \sum_{i=1}^{\infty} P\{X = x_i\} \ = \ P\left(\bigcup_{i=1}^{\infty} \{X = x_i\} \right) = P(\Omega) = 1 \ .$$

La funzione p sopra definita viene chiamata *funzione di probabilità* o *densità discreta*.

La conoscenza della densità consente di determinare la legge di X; infatti assegnato $A \subset \mathbb{R}$, la probabilità di tale evento è data da:

$$P\{X \in A\} = \sum_{x_i \in A} P\{X = x_i\} = \sum_{x_i \in A} p(x_i) \ . \tag{2.7}$$

Un'importante variabile aleatoria è la funzione indicatrice che abbiamo introdotto all'inizio. Sia $(\Omega, \mathcal{A}, P)$ uno spazio di probabilità e sia $A \in \mathcal{A}$, con $P(A) = p$. Allora

$$P(|A| = 1) = P(A) = p$$
$$P(|A| = 0) = P(A^c) = 1 - p \ .$$

In questo caso, pertanto, la funzione di densità assume valori positivi su 0 e 1, e zero altrove. Sotto questo punto di vista, il concetto di variabile aleatoria può essere considerato una generalizzazione del concetto di evento: a ciascun evento A si associa un numero 0 (se l'evento è falso) oppure 1 (se l'evento è vero); più in generale ad un evento A si può associare un certo numero reale.

Esempio 2.3 Consideriamo una sequenza di tre lanci di una moneta regolare. Sia X la quantità:

X = "numero di teste in una successione di tre lanci di una moneta"

Poichè $\Omega = \{$(TTT), (TTC), (TCT), (TCC), (CTT), (CTC), (CCT), (CCC)$\}$, allora X può assumere solo i valori 0,1,2,3, cioè $\Omega_X = \{0, 1, 2, 3\}$. Ciascuna delle sequenze di risultati ha probabilità uguale a 1/8. Se indichiamo con ω la generica sequenza di risultati, allora si ha:

ω	$X(\omega) = x$	$P(\omega)$
T T T	3	1/8
T T C	2	1/8
T C T	2	1/8
T C C	1	1/8
C T T	2	1/8
C T C	1	1/8
C C T	1	1/8
C C C	0	1/8

La funzione di probabilità è data da:

$$\begin{aligned}
p(0) &= P(X = 0) = P(\text{CCC}) = 1/8 \\
p(1) &= P(X = 1) = P(\text{CCT} \cup \text{CTC} \cup \text{TCC}) = P(\text{CCT}) + P(\text{CTC}) + P(\text{TCC}) = 3/8 \\
p(2) &= P(X = 2) = P(\text{CTT} \cup \text{TTC} \cup \text{TCT}) = P(\text{CTT}) + P(\text{TTC}) + P(\text{TCT}) = 3/8 \\
p(3) &= P(X = 3) = P(\text{TTT}) = 1/8 \ .
\end{aligned}$$

In Figura 2.1 viene data la rappresentazione grafica della funzione di probabilità della v.a. $X = |T_1| + |T_2| + |T_3|$. E' immediato verificare la proprietà (2.6):

$$p(0) + p(1) + p(2) + p(3) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1 \ .$$

Sullo stesso spazio campionario Ω si può considerare la variabile aleatoria Y = "numero di coppie consecutive di teste":

$$Z = |\text{T}_1\text{T}_2| + |\text{T}_2\text{T}_3| \ .$$

In questo caso, come si vede dalla tabella precedente, la v.a. Z può assumere i valori $0, 1, 2$, cioè $\Omega_Z = \{0, 1, 2\}$. La funzione di probabilità è data da:

$$\begin{aligned}
p(0) &= P(Z = 0) = P(\text{CCC} \cup \text{CCT} \cup \text{CTC} \cup \text{TCC} \cup \text{TCT}) \\
& = P(\text{CCC}) + P(\text{CCT}) + P(\text{CTC}) + P(\text{TCC}) + P(\text{TCT}) = 5/8 \\
p(1) &= P(Z = 1) = P(\text{TTC} \cup \text{CTT}) = P(\text{TTC}) + P(\text{CTT}) = 2/8 \\
p(2) &= P(Z = 2) = P(\text{TTT}) = 1/8 \ .
\end{aligned}$$

Verifichiamo anche in questo caso la proprietà (2.6):

$$p(0) + p(1) + p(2) = \frac{5}{8} + \frac{2}{8} + \frac{1}{8} = 1 \ .$$

In Figura 2.2 viene data la rappresentazione grafica della funzione di probabilità della v.a. Z. ♣

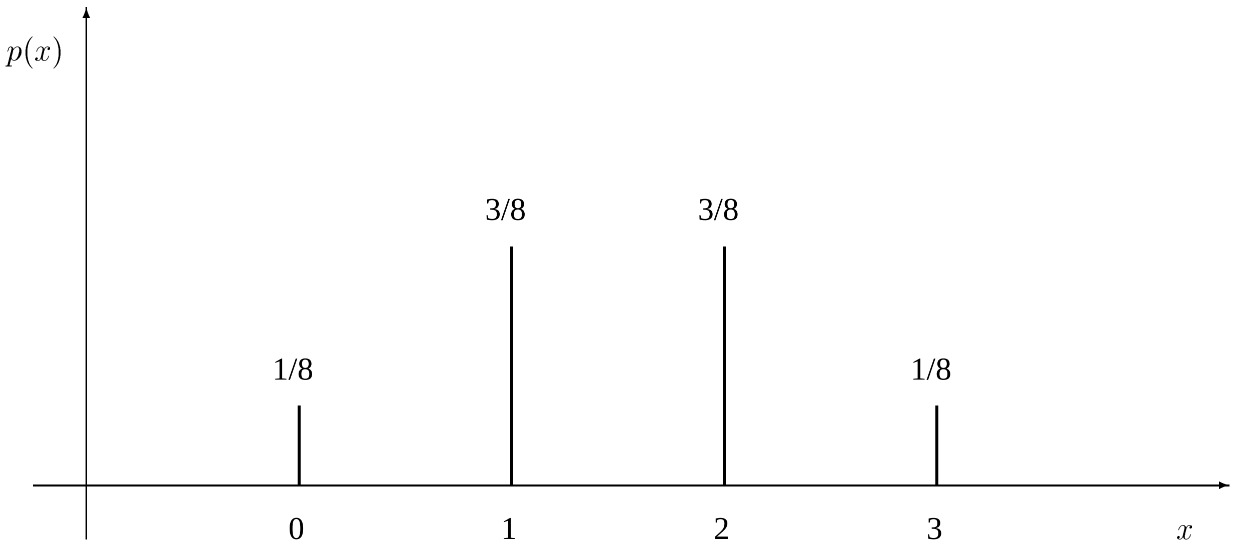


Figura 2.1: Rappresentazione grafica della distribuzione di probabilità del numero di teste uscite nel lancio di tre monete.

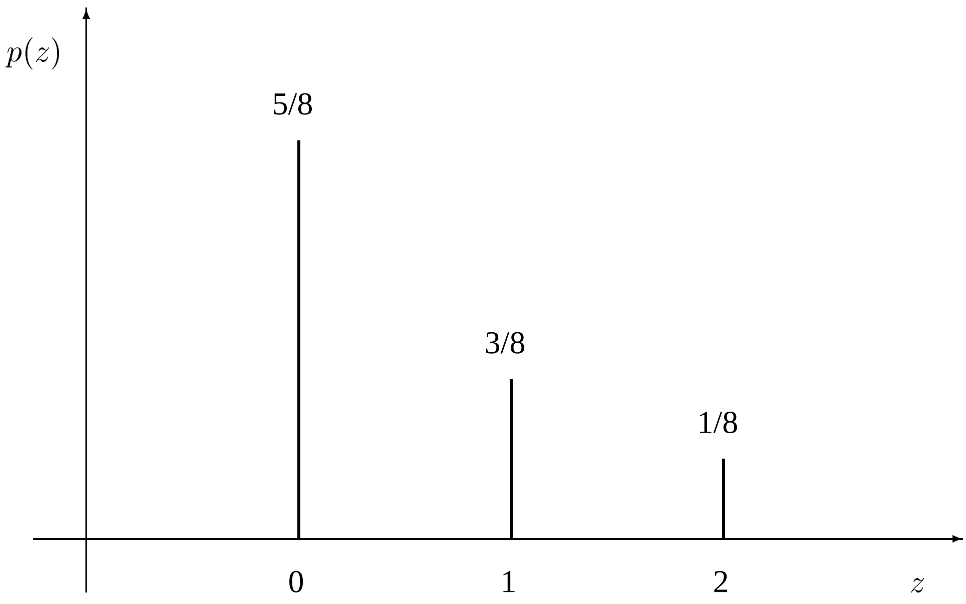


Figura 2.2: Rappresentazione grafica della distribuzione di probabilità della v.a. Z.

Insieme alla funzione di probabilità di una variabile aleatoria si definisce un'altra funzione importante ai fini della descrizione delle principali proprietà di una variabile aleatoria: la *funzione di ripartizione* o *funzione di distribuzione*.

Definizione 2.4 Sia X una variabile aleatoria discreta definita su $(\Omega, \mathcal{A}, P)$ ed a valori in $\Omega_X = \{x_1, x_2, \ldots\}$. Si definisce *funzione di ripartizione* di X la funzione $F : \mathbb{R} \to [0, 1]$ definita da:

$$F(x) := P(X \leq x) \; . \tag{2.8}$$

Sia $x_h \in \Omega_X$ allora per la (2.3) si ha:

$$F(x_h) = p(x_1) + p(x_2) + \cdots + p(x_h) \tag{2.9}$$

Più in generale, per $t \in \mathbb{R}$, si ha:

$$F(x) = P(X \leq x) = \sum_{t \leq x} p(t) \; . \tag{2.10}$$

L'ordinata della funzione di ripartizione calcolata nel punto $x \in \mathbb{R}$ indica pertanto la probabilità con cui la variabile aleatoria X assume valori al più uguali a x. Dalla definizione di funzione di ripartizione seguono le seguenti proprietà:

1. $\underline{F(x) \text{ è non decrescente}}$: infatti per $x_1 \leq x_2$ si ha $\{X \leq x_1\} \subseteq \{X \leq x_2\}$ e quindi per la proprietà di monotonia si ha:

$$P\{X \leq x_1\} \leq P\{X \leq x_2\} \; ,$$

cioè $F(x_1) \leq F(x_2)$.

2. $\underline{F(x) \text{ è continua da destra, cioè } \lim_{\varepsilon \to 0^+} F(x + \varepsilon) = F(x)}$: infatti considerati gli eventi $\{X \leq x\}$ e $\{X \leq x + \varepsilon\}$, risulta:

$$\lim_{\varepsilon \to 0^+} \{X \leq x + \varepsilon\} = \{X \leq x\}$$

e quindi:

$$\lim_{\varepsilon \to 0^+} P\{X \leq x + \varepsilon\} = P\{X \leq x\} \; .$$

3. $\underline{F(x_i) - F(x_{i-1}) = p(x_i) \text{ per } x_i \in \Omega_X}$, infatti:

$$\begin{aligned} F(x_i) - F(x_{i-1}) &= P\{X \leq x_i\} - P\{X \leq x_{i-1}\} = P\{x_{i-1} < X \leq x_i\} \\ &= P\{X = x_i\} = p(x_i) \; . \end{aligned}$$

4. $\underline{F(x) = 0 \text{ per } x < x_1 \text{ e } F(x) = 1 \text{ per } x \geq x_k}$, infatti per $x < x_1$ si ha $\{X \leq x\} = \emptyset$ e per $x \geq x_k$ si ha $\{X \leq x\} = \Omega$.

In Figura 2.3 diamo la funzione di ripartizione della v.a. $X = |T_1| + |T_2| + |T_3|$.

Esercizio 2.5 Sia X la v.a. "numero di teste ottenute in due lanci consecutivi di una moneta equilibrata".

1. Descrivere lo spazio degli eventi.
2. Calcolare la funzione di densità di X e rappresentarla graficamente;
3. Calcolare la funzione di ripartizione di X e rappresentarla graficamente;

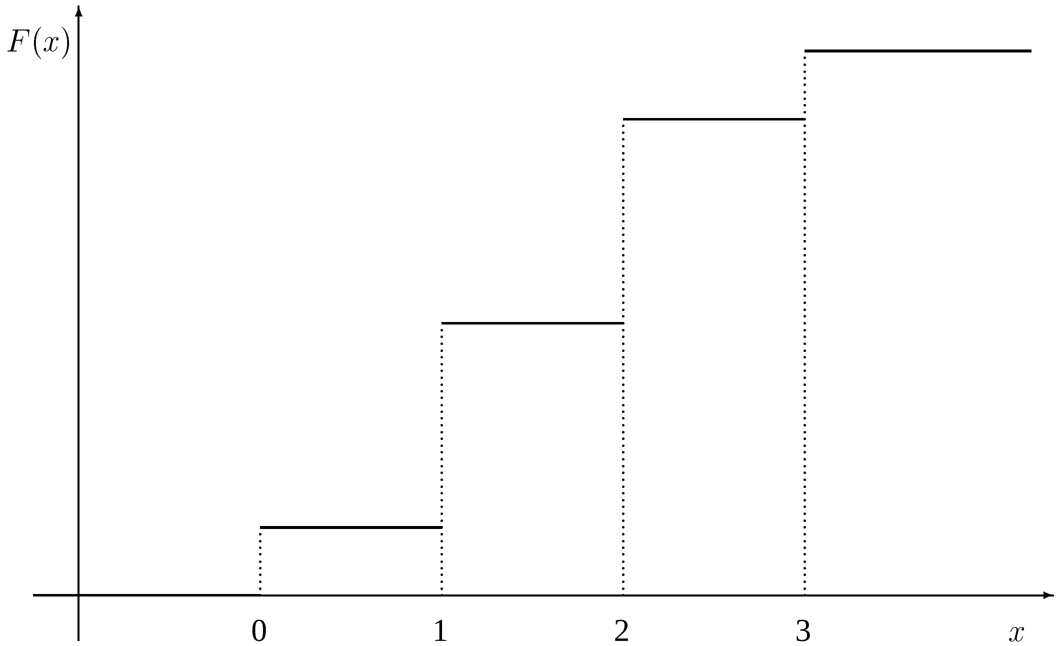


Figura 2.3: Rappresentazione grafica della funzione di ripartizione della v.a. $X = |T_1| + |T_2| + |T_3|$ concernente il numero di teste uscite nel lancio di tre monete.

Soluzione. Lo spazio degli eventi Ω è il seguente:

$$\Omega = \{TT, CT, TC, CC\} .$$

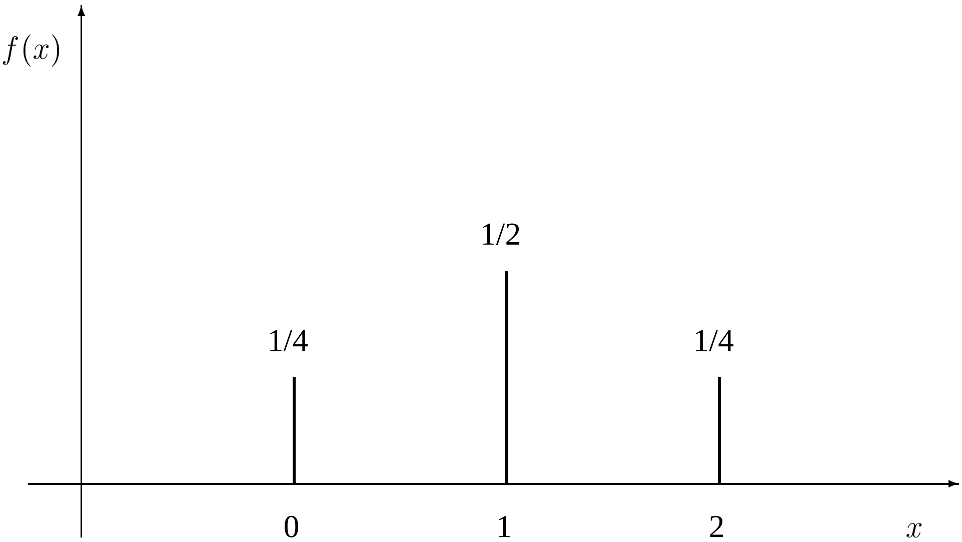
ciascuno con probabilità $\frac{1}{4}$ in quanto assumiamo l'indipendenza degli eventi elementari $\{C, T\}$ associati a ciascun lancio, cioè ad esempio $P(TT) = P(T) \cdot P(T) = \frac{1}{2}\frac{1}{2} = \frac{1}{4}$.

Considerata la v.a. X="numero di teste ottenute in due lanci consecutivi di una moneta equilibrata", ne segue che la funzione di densità assume valori strettamente positivi nell'insieme $\mathcal{X} = \{0, 1, 2\}$:

$$\begin{aligned} p(0) &= P(X = 0) = P(CC) = \frac{1}{4} = 0.25 \\ p(1) &= P(X = 1) = P(CT) + P(TC) = \frac{1}{2} = 0.5 \\ p(2) &= P(X = 2) = P(TT) = \frac{1}{4} = 0.25 . \end{aligned}$$

Si ha pertanto:

$$p(x) = \begin{cases} 0.25 & \text{per } x = 0 \\ 0.50 & \text{per } x = 1 \\ 0.25 & \text{per } x = 2 \\ 0 & \text{altrove.} \end{cases}$$



La funzione di ripartizione $F(x)$ viene costruita a partire dalla funzione di densità in base alla sua definizione:

$$F(t) = P(X \leq t).$$

Si ha pertanto:

$$F(x) = \begin{cases} 0 & \text{per } x < 0 \\ 0.25 & \text{per } 0 \leq x < 1 \\ 0.75 & \text{per } 1 \leq x < 2 \\ 1.00 & \text{per } 2 \leq x \end{cases}$$

e tale funzione ha la seguente rappresentazione grafica: ♣

2.3 Distribuzioni congiunte

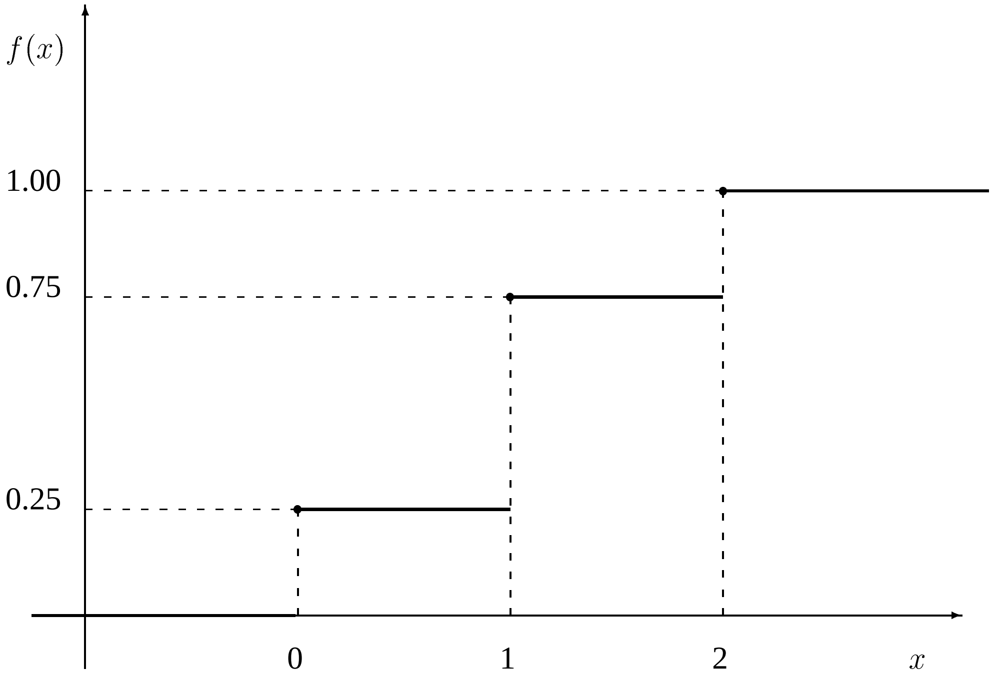
Definizione 2.6 Una v.a. m-dimensionale discreta (o *vettore aleatorio* discreto) è un'applicazione $\mathbf{X} = (X_1, \ldots, X_m) : \Omega \to \mathbb{R}$ tale che le applicazioni $X_1, \ldots, X_m$ siano delle v.a. discrete.

Ovviamente, se $\mathbf{X}$ è una v.a. discreta, allora essa può assumere al più un'infinità numerabile di valori. Infatti se $\mathbf{x} = (x_1, \ldots, x_m) \in \mathbb{R}$, allora

$$\{\omega \in \Omega \, : \, \mathbf{X} = \mathbf{x}\} := \{\omega \in \Omega \, : \, X_1 = x_1\} \cap \cdots \cap \{\omega \in \Omega \, : \, X_m = x_m\} \qquad (2.11)$$

Tale relazione mostra che il v.a. $\mathbf{X}$ risulta uguale a $\mathbf{x}$ se e solo se X_1 assume valore x_1, X_2 assume valore x_2, ..., X_m assume valore x_m; ne segue che $\{\mathbf{X} = \mathbf{x}\}$ è un evento in quanto intersezione di eventi. Nel seguito, anzichè la (2.11), utilizzeremo sempre la seguente scrittura sintetica:

$$\{\mathbf{X} = \mathbf{x}\} := \{X_1 = x_1\} \cap \cdots \cap \{X_m = x_m\} \, .$$



Al solito, denoteremo con Ω_X il codominio di $\mathbf{X}$.

Analogamente al caso unidimensionale, anche per le v.a. m-dimensionali si può definire una *densità discreta*, ponendo:

$$p(\mathbf{x}) := P\{\mathbf{X} = \mathbf{x}\} \ . \tag{2.12}$$

dove, come in precedenza, la funzione $p(\mathbf{x})$ verifica le tre proprietà:

$$1. \quad p(\mathbf{x}) \ \geq \ 0 \qquad \text{per ogni } \mathbf{x} \in \mathbb{R}^m \tag{2.13}$$

$$2. \quad p(\mathbf{x}) \ > \ 0 \qquad \text{per } \mathbf{x} \in \Omega_X \tag{2.14}$$

$$3. \quad \sum_{\mathbf{x} \in \Omega_X} p(\mathbf{x}) = 1 \ . \tag{2.15}$$

Analogamente al caso di variabili aleatorie, chiameremo *funzione di densità* (discreta) una funzione p che soddisfi le tre condizioni viste sopra e, se p è una funzione di densità, allora esistono certamente uno spazio di probabilità $(\Omega, \mathcal{A}, P)$ ed una v.a. m-dimensionale $\mathbf{X}$ di cui p sia la densità.

I ragionamenti che hanno portato alla (2.7) si possono ripetere esattamente per v.a. discrete m-dimensionali $\mathbf{X}$ e $A \subset \mathbb{R}^m$, e si ottiene:

$$P(\mathbf{x} \in A) = \sum_{\mathbf{x} \in A} p(\mathbf{x}) \ .$$

Se $X_1, \ldots, X_m$ sono v.a. reali, la densità della v.a. m-dimensionale $\mathbf{X} = (X_1, \ldots, X_m)$ si chiama *densità congiunta* delle v.a. $X_1, \ldots, X_m$; al contrario, se $\mathbf{X} = (X_1, \ldots, X_m)$ è una v.a. m-dimensionale, le densità delle v.a. reali $X_1, \ldots, X_m$ – denotate rispettivamente con $p_1, \ldots, p_m$ – vengono chiamate *densità marginali*.

Un importante caso è quello $m = 2$, cioè di v.a. bidimensionali, in cui le densità congiunte e marginali possono essere assegnate mediante una *tabella a doppia entrata*. Per semplicità di notazione, in questo caso la v.a. bidimensionale viene denotata con (X, Y) anzichè con (X_1, X_2);

inoltre assumiamo che il codominio delle v.a. X e Y sia costituito rispettivamente da k e h valori distinti, cioè $\Omega_X = \{x_1, \dots, x_k\}$ e $\Omega_Y = \{y_1, \dots, y_h\}$ (nel caso generale, dovremmo considerare tabelle con un numero infinito di righe e/o colonne). Posto $p_{ij} = p(x_i, y_j)$, $p_{i0} = p_X(x_i)$ e $p_{0j} = p_Y(y_j)$, si ottiene la seguente tabella a doppia entrata:

	y_1	$\dots$	y_j	$\dots$	y_h	
x_1	p_{11}	$\dots$	p_{1j}	$\dots$	p_{1h}	p_{10}
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_i	p_{i1}	$\dots$	p_{ij}	$\dots$	p_{ih}	p_{i0}
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_k	p_{k1}	$\dots$	p_{kj}	$\dots$	p_{kh}	p_{k0}
	p_{01}	$\dots$	p_{0j}	$\dots$	p_{0h}	1

dove $p_{10}, \dots, p_{k0}$ costituisce la distribuzione marginale di X e $p_{01}, \dots, p_{0h}$ costituisce la densità marginale di Y, cioè

$$\begin{aligned} p_{i0} &= p_X(x_i) := P\{X = x_i\} \qquad i = 1, \dots, k\,, \\ p_{0j} &= p_Y(y_j) := P\{Y = y_j\} \qquad j = 1, \dots, h\,. \end{aligned}$$

Ovviamente risulta:

$$\begin{aligned} p_{i0} &:= p_X(x_i) = P\{X = x_i\} = P\left(\bigcup_{j=1}^{h}\{X = x_i, Y = y_j\}\right) = \sum_{j=1}^{h} P\{X = x_i, Y = y_j\} \\ &= \sum_{j=1}^{h} p(x_i, y_j) = \sum_{j=1}^{h} p_{ij} \\ p_{0j} &:= p_Y(y_j) = P\{Y = y_j\} = P\left(\bigcup_{i=1}^{k}\{X = x_i, Y = y_j\}\right) = \sum_{i=1}^{k} P\{X = x_i, Y = y_j\} \\ &= \sum_{i=1}^{k} p(x_i, y_j) = \sum_{i=1}^{k} p_{ij}\ . \end{aligned}$$

Si noti infine che non è possibile, in generale, conoscendo le sole densità marginali, ricostruire la densità congiunta in quanto – come si può anche vedere dalla tabella precedente – densità congiunte diverse possono avere le stesse densità marginali.

Definizione 2.7 Le m v.a. discrete $X_1, \dots, X_m$, assegnate su uno stesso spazio di probabilità $(\Omega, \mathcal{A}.P)$, si dicono *indipendenti* se e solo se per ogni scelta di $A_1, \dots, A_m \subset \mathbb{R}$ si ha:

$$P\{X_1 \in A_1, \dots, X_m \in A_m\} = P\{X_1 \in A_1\} \cdots P\{X_m \in A_m\}\ . \tag{2.16}$$

Diremo che le v.a. $X_1, \dots, X_n, \dots$ (in numero infinito) sono indipendenti se e solo se per ogni scelta di m indici $i_1, \dots, i_m$, le variabili $X_{i_1}, \dots, X_{i_m}$ risultano indipendenti.

Il significato intuitivo della nozione di indipendenza di v.a. è analogo a quello di indipendenza di eventi: m v.a. sono indipendenti se la conoscenza dei valori assunti da alcune di esse non dà informazioni che modifichino la previsione di quelli assunti dalle altre. Del resto, la (2.16) scritta nel caso di due v.a., per ogni $A_1 \subseteq \Omega_{X_1}, A_2 \subseteq \Omega_{X_2}$ diviene:

$$P\{X_1 \in A_1, X_2 \in A_2\} = P\{X_1 \in A_1\}P\{X_2 \in A_2\} \tag{2.17}$$

che non significa altro che il fatto che i due eventi $\{X_1 \in A_1\}$ e $\{X_2 \in A_2\}$ sono fra loro indipendenti *per ogni scelta* degli insiemi A_1, A_2.

Nella (2.16), in particolare scelti $A_1 = \{x_1\}, \ldots, A_m = \{x_m\}$, si ottiene:

$$P\{X_1 = x_1, \ldots, X_m = x_m\} = P\{X_1 = x_1\} \cdots P\{X_m = x_m\} \; . \tag{2.18}$$

Indicato con p la densità congiunta di $X_1, \ldots, X_m$ e con $p_1, \ldots, p_m$ le rispettive densità di $X_1, \ldots, X_m$, posto $\mathbf{x} = (x_1, \ldots, x_m)$, la (2.18) si può scrivere:

$$p(\mathbf{x}) = p_1(x_1) \cdots p_m(x_m) \; . \tag{2.19}$$

Al contrario, supponiamo che valga la (2.18); se $A_1, \ldots, A_m \subset \mathbb{R}$, posto $A = A_1 \times A_2 \times \cdots \times A_m$, si ha:

$$\begin{aligned} P\{X_1 \in A_1, \ldots, X_m \in A_m\} &= \sum_{x_1 \in A_1, \ldots, x_m \in A_m} p(x_1, \ldots, x_m) \\ &= \sum_{x_1 \in A_1, \ldots, x_m \in A_m} p_1(x_1) \cdots p_m(x_m) \\ &= \sum_{x_1 \in A_1} p_1(x_1) \cdots \sum_{x_m \in A_m} p_m(x_m) \\ &= P\{X_1 \in A_1\} \cdots P\{X_m \in A_m\} \end{aligned}$$

Ciò dimostra che la (2.19) è una condizione equivalente all'indipendenza di $X_1, \ldots, X_m$. In particolare, nel caso di v.a. indipendenti, tramite la (2.19) è possibile calcolare la densità congiunta a partire dalle marginali, cosa che non è vera in generale.

Esercizio 2.8 Completare la seguente distribuzione di probabilità congiunta nell'ipotesi di indipendenza stocastica delle v.a. X e Y:

	x_1	x_2	x_3	
y_1	*	0.2	*	0.4
y_2	*	*	0.105	*
y_3	*	*	*	*
	*	*	0.3	1

Soluzione.

	x_1	x_2	x_3	
y_1	0.08	0.2	0.12	0.40
y_2	0.07	0.175	0.105	0.35
y_3	0.05	0.125	0.075	0.25
	0.2	0.5	0.3	1

♣

2.3.1 Densità condizionali

Siano assegnato due v.a. discrete (X, Y) avente densità congiunta $p(x, y)$. Assumendo $p_X(x) > 0$, in accordo alla definizione di probabilità condizionale, si ha:

$$P(Y = y|X = x) = \frac{P(Y = y, X = x)}{P(X = x)} = \frac{p(x, y)}{p_X(x)} \qquad p_X(x) > 0 \ .$$

Analogamente per $P(X = x|Y = y)$ assumendo stavolta $p_Y(y) > 0$.

Definizione 2.9 Sia (X, Y) un vettore aleatorio discreto con funzione di probabilità $p(x, y)$ e densità di probabilità marginali $p_X(x)$ e $p_Y(y)$. Si dice *densità di probabilità condizionale* di Y dato $X = x$ la funzione

$$p_{Y|X}(y|x) := P(Y = y|X = x) = \frac{p(x, y)}{p_X(x)} \qquad p_X(x) > 0 \ ;$$

analogamente si dice *densità di probabilità condizionale* di X dato $Y = y$ la funzione

$$p_{X|Y}(x|y) := P(X = x|Y = y) = \frac{p(x, y)}{p_Y(y)} \qquad p_Y(y) > 0 \ .$$

Con particolare riferimento ad una distribuzione doppia di probabilità, è possibile individuare $2 + h + k$ distribuzioni semplici:

- 2 <u>distribuzioni marginali</u>

$$X = \begin{cases} x_1 & x_2 & \cdots & x_k \\ p_{10} & p_{20} & \cdots & p_{k0} \end{cases} \qquad Y = \begin{cases} y_1 & y_2 & \cdots & y_h \\ p_{01} & p_{02} & \cdots & p_{0h} \end{cases}$$

- h <u>distribuzioni condizionate di X dato $Y = y_j$</u>, la cui distribuzione generica è:

$$\begin{cases} x_1 & x_2 & \cdots & x_k \\ p_{X|Y}(x_1|y_j) & p_{X|Y}(x_2|y_j) & \cdots & p_{X|Y}(x_k|y_j) \end{cases}$$

- k distribuzioni condizionate di Y dato $X = x_i$ la cui distribuzione generica è:

$$\begin{cases} y_1 & y_2 & \cdots & y_h \\ p_{Y|X}(y_1|x_i) & p_{Y|X}(y_2|x_i) & \cdots & p_{Y|X}(y_h|x_i) \end{cases}$$

Esercizio 2.10 Il responsabile di un giornale finanziario desidera conoscere quali sono le aree di investimento di maggiore interesse fra gli abbonati al giornale. Vengono scelti a caso 330 (investitori) professionisti abbonati a cui viene inviato un questionario. I risultati sono sintetizzati nella seguente tabella:

	Area di investimento				
professione	Fondi azionari	Fondi Obbligazionari	Azioni	Titoli di stato	Totale
Medico	30	25	15	0	70
Avvocato	29	34	12	6	81
Commercialista	50	35	29	15	129
Altro	21	14	10	5	50
Totale	130	108	66	26	330

1. Valutare la probabilità che un investitore non sia né un avvocato né un commercialista;

2. Valutare la probabilità che l'investitore sia un commercialista avente maggiore interesse nei fondi obbligazionari;

3. Valutare la probabilità che il professionista sia un avvocato, nel caso in cui si abbia maggiore interesse nell'acquisto di azioni;

4. Valutare la probabilità che l'investimento principale del professionista non sia nei titoli di stato;

5. Sia A l'evento che l'investitore sia un avvocato e B l'evento che l'area di maggiore interesse siano i titoli di stato. Verificare se gli venti A e B sono mutuamente esclusivi.

Esercizio 2.11 Siano X e Y due variabili aleatorie discrete, a valori rispettivamente in $\Omega_X = \{0, 1\}$ e $\Omega_Y = \{1, 2\}$. Considerate le seguenti probabilità congiunte in funzione del parametro α:

$$P(X = 0, Y = 1) = \alpha$$
$$P(X = 0, Y = 2) = 1 - 4\alpha$$
$$P(X = 1, Y = 1) = \alpha\ .$$

Calcolare l'insieme dei valori ammissibili per α. Verificare se esistono valori di α tali che X e Y risultano stocasticamente indipendenti ed in tal caso calcolarli.

Soluzione. Trattandosi di una distribuzione di probabilità deve ovviamente risultare:

$$\begin{aligned} P(X=0,Y=1)+P(X=0,Y=2)+P(X=1,Y=1)+P(X=1,Y=2) = \\ = \alpha + 1 - 4\alpha + \alpha + P(X=1,Y=2) = 1 \end{aligned}$$

da cui si ottiene $P(X=1,Y=2)=2\alpha$ e quindi la seguente tabella:

		X		
		0	1	
Y	1	α	α	2α
	2	$1-4\alpha$	2α	$1-2\alpha$
		$1-3\alpha$	3α	1

Affinchè i valori in tabella siano delle probabilità devono essere verificate preliminarmente le condizioni:

$$0 \le \alpha \le 1/2\,; \qquad 0 \le 1-2\alpha \le 1\,; \qquad 0 \le 1-3\alpha \le 1\,; \qquad 0 \le 1-4\alpha \le 1$$

da cui segue $0 \le \alpha \le 1/4$.

Le v.a. X e Y risultano stocasticamente indipendenti se e solo se è soddisfatto il seguente sistema ($p_{ij} = p_{i0}p_{0j}$ per ogni i,j):

$$\begin{aligned} \alpha &= 2\alpha(1-3\alpha) \\ \alpha &= 6\alpha^2 \\ 1-4\alpha &= (1-2\alpha)(1-3\alpha) \\ 2\alpha &= 3\alpha(1-2\alpha) \end{aligned}$$

Le quattro equazioni conducono alla equazione seguente:

$$6\alpha^2 - \alpha = 0$$

soddisfatto per $\alpha_1 = 0$ e $\alpha_2 = 1/6$, valori entrambi compatibili con l'insieme dei valori ammissibili. In particolare, per $\alpha = 0$, la distribuzione è concentrata sulla coppia di valori $(X=0, Y=2)$. ♣

2.4 Speranza matematica

Sia X una v.a. definita su uno spazio di probabilità $(\Omega, \mathcal{A}, P)$.

Definizione 2.12 Sia X una v.a. (discreta) che assuma i valori $x_1, x_2, \ldots$ ed indichiamo con $p(x)$ la sua densità. Assumiamo che:

$$\sum_i |x_i| p(x_i) < +\infty\ .$$

In questo caso si definisce *speranza matematica* di X, la quantità

$$\mathbb{E}[X] := \sum_i x_i\, p(x_i) = \sum_i x_i P\{X = x_i\} \tag{2.20}$$

In altre parole la speranza matematica di X è data dalla (2.20) a condizione che la serie dei valori assoluti converga. Sinonimi di speranza matematica sono i termini *media*, *valore medio*, *attesa*, *valore atteso*, *previsione*. In alcuni testi, più raramente, la speranza matematica viene denotata col simbolo $\mathbb{P}[X]$. Se consideriamo la funzione indicatrice $|A|$ di un evento A che risulta vero con probabilità p e falso con probabilità $1-p$, si ha:

$$\mathbb{E}[A] = p \cdot 1 + (1-p) \cdot 0 = p = P(A)$$

e pertanto la speranza matematica della funzione indicatrice di un evento A coincide con la probabilità dell'evento stesso. Sotto questo punto di vista, così come il concetto di variabile aleatoria generalizza il concetto di evento, allora il concetto di speranza matematica generalizza il concetto di probabilità.

Il concetto di speranza matematica generalizza il concetto di probabilità.

In effetti, osservando che i termini che intervengono nella (2.20) non sono altro che i possibili valori di X moltiplicati per la probabilità con cui essi vengono assunti, il significato intuitivo della speranza matematica è quello della media dei possibili valori di X. Nel caso in cui $\mathbb{E}[X] = 0$, si dice che X è *centrata*.

Si noti che la speranza matematica di una v.a. X dipende unicamente dalla sua funzione di densità: se due v.a. X e Y hanno la stessa densità allora avranno anche la stessa speranza matematica, cioè $\mathbb{E}[X] = \mathbb{E}[Y]$.

In generale, se $\mathbf{X} = (X_1, \ldots, X_m)$ è una v.a. m-dimensionale, chiameremo *speranza matematica* di $\mathbf{X}$ il vettore avente per componenti le speranze matematiche delle componenti di $\mathbf{X}$, cioè:

$$\mathbb{E}[\mathbf{X}] := (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_m])^t \, .$$

Posto $\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}]$ e $\mu_i = \mathbb{E}[X_i]$, per $i = 1, \ldots, m$, la relazione precedente possiamo scriverla:

$$\boldsymbol{\mu} \;\; = \;\; (\mu_1, \ldots, \mu_m)^t \, .$$

dove ^{t} indica vettore trasposto.

Esempio 2.13 Una *slot machine* dispone di due quadranti: in ogni quadrante possono comparire 3 diversi tipi di figure: mele, fragole, ciliege. La macchina è strutturata in modo che i due quadranti girino in modo indipendente l'uno dall'altro, e quando si arrestano compaia una delle tre figure in ognuno dei due quadranti. Indichiamo con M, F e C rispettivamente gli eventi "mela", "fragola" e "ciliegia"

Dopo aver osservato attentamente il gioco per un lungo periodo, si stabilisce che le probabilità di uscita di ogni figura sono diverse, e supponiamo che la distribuzione di probabilità sui tre eventi $\Omega = \{M, F, C\}$ sia la seguente:

$$P(M) = 0.1 \qquad P(F) = 0.4 \qquad P(C) = 0.5 \, .$$

Il gioco consiste nell'abbassare la leva che permette di far girare le figure; il risultato che comparirà sarà una delle 9 possibili coppie di figure: $\{(M,M),(M,F),(M,C),(F,M),(F,F),(F,C),(C,M),(C,F),(C,C)\}$. Il giocatore inserisce un euro nella macchina e, in corrispondenza di uno dei possibili eventi, riceve le seguenti somme:

Risultato	Somma corrisposta
(M,M)	10 euro
(F,F)	2 euro
(C,C)	1 euro
figure diverse	0 euro

In base a quanto detto sopra, i risultati non sono equiprobabili; inoltre poichè gli eventi sono indipendenti si ha:

Risultato	Probabilità
(M,M)	$P(M,M) = P(M)\cdot P(M) = 0.1\cdot 0.1 = 0.01$
(F,F)	$P(F,F) = P(F)\cdot P(F) = 0.4\cdot 0.4 = 0.16$
(C,C)	$P(C,C) = P(C)\cdot P(C) = 0.5\cdot 0.5 = 0.25$
figure diverse	$1-\{P(M,M)+P(F,F)+P(C,C\} = 1-(0.01+0.16+0.25) = 0.58$

Considerata la v.a. $X =$"guadagno in una giocata", essa è data da:

$$X \;=\; 10\cdot|(M,M)| + 2\cdot|(F,F)| + 1\cdot|(C,C)| - 0.10\ ,$$

da cui si ottiene la seguente distribuzione di probabilità di X:

$$\begin{array}{lclclcl} x_1 & = & 9 & \quad & p_1 & = & P(M,M) = 0.01 \\ x_2 & = & 1 & & p_2 & = & P(F,F) = 0.16 \\ x_3 & = & 0 & & p_3 & = & P(C,C) = 0.25 \\ x_4 & = & -1 & & p_4 & = & 1-\{P(M,M)+P(F,F)+P(C,C)\} = 0.58\ . \end{array}$$

Pertanto il valore atteso di X è dato da:

$$\begin{aligned} \mathbb{E}[X] &= p_1x_1 + p_2x_2 + p_3x_3 + p_4x_4 \\ &= 0.01\cdot 9 + 0.16\cdot 1 + 0.25\cdot 0 - 0.58\cdot 1 = -0.33\ . \end{aligned}$$

Il valore ottenuto può essere interpretato dicendo che, disponendo di un capitale iniziale di 10 euro ed effettuando 100 partite, il guadagno finale sarà di -33 euro, cioè il capitale finale sarà all'incirca di 67 euro, cioè la nostra somma si sarà ridotta di 1/3. ♣

2.4.1 Proprietà della speranza matematica.

La speranza matematica possiede le seguenti proprietà:

1. Per ogni $\beta \in \mathbb{R}$, la speranza della variabile aleatoria costante $X = \beta$ coincide con la costante stessa:
$$\mathbb{E}(\beta) = \beta \, .$$

2. Se X è una v.a. avente speranza matematica finita e $\alpha, \beta \in \mathbb{R}$, allora anche la v.a. $\alpha X + \beta$ ha speranza matematica finita e risulta:
$$\mathbb{E}[\alpha X + \beta] = \alpha \, \mathbb{E}[X] + \beta \, . \tag{2.21}$$
Infatti, dalla definizione di valore atteso segue immediatamente:
$$\mathbb{E}[\alpha X + \beta] = \sum_i (\alpha x_i + \beta) p_i \; = \; \alpha \sum_i x_i p_i + \beta \sum_i p_i = \alpha \mathbb{E}[X] + \beta$$

3. Se X e Y sono due v.a. aventi speranza matematica finita, allora anche la v.a. $X \pm Y$ ha speranza matematica finita e si ha:
$$\mathbb{E}[X \pm Y] = \mathbb{E}[X] \pm \mathbb{E}[Y] \, . \tag{2.22}$$
Infatti, in base alla definizione di distribuzione marginali, nel caso discreto si ha:
$$\begin{aligned}\mathbb{E}[X \pm Y] &= \sum_i \sum_j (x_i + y_j) p_{ij} = \sum_i \sum_j x_i p_{ij} + \sum_i \sum_j y_j p_{ij} \\ &= \sum_i x_i \sum_j p_{ij} + \sum_j y_j \sum_i p_{ij} = \sum_i x_i p_i + \sum_j y_j p_j \\ &= \mathbb{E}[X] + \mathbb{E}[Y] \, .\end{aligned}$$

4. Le due precedenti proprietà possono essere sintetizzate affermando che se due v.a. X e Y hanno speranza matematica finita, allora anche la v.a. $\alpha X + \beta Y$ ha speranza matematica finita e si ha:
$$\mathbb{E}[\alpha X + \beta Y] = \alpha \, \mathbb{E}[X] + \beta \, \mathbb{E}[Y] \, . \tag{2.23}$$
dove α, β sono numeri reali arbitrari.

5. Se X e Y sono v.a. indipendenti aventi speranza matematica finita, allora si ha:
$$\mathbb{E}[XY] = \mathbb{E}[X] \, \mathbb{E}[Y] \tag{2.24}$$
Infatti, nel caso di v.a. indipendenti si ha $p_{ij} = p_i p_j$ e quindi
$$\begin{aligned}\mathbb{E}[XY] &= \sum_i \sum_j x_i y_j p_{ij} = \sum_i \sum_j x_i y_j p_i p_j \\ &= \sum_i x_i p_i \sum_j y_j p_j = \mathbb{E}[X] \, \mathbb{E}[Y]\end{aligned}$$

La media non è l'unico parametro che caratterizza una distribuzione; nella maggior parte dei casi, infatti, essa è insufficiente a descrivere la distribuzione ed a rappresentare le caratteristiche peculiari.

2.5 Momenti e varianza

In modo generale si può affermare che tutte le caratteristiche di una distribuzione di probabilità sono rappresentate da una famiglia di parametri chiamati *momenti*.

Definizione 2.14 Sia k un intero positivo, sia c una costante ed assumiamo che v.a. $(X-c)^k$ abbia speranza matematica finita. Si definisce *momento di ordine k* intorno al punto c la quantità:

$$\mathbb{E}[(X-c)^k] := \sum_i (x_i - c)^k p(x_i) \tag{2.25}$$

Due casi notevoli. Il primo si ottiene per $c = 0$, in tal caso si ha il *momento di ordine k*, indicato con $\mathbb{E}[X^k]$; il secondo si ottiene per $c = \mathbb{E}[X] = \mu$, in tal caso si ottiene il *momento centrale di ordine k* intorno alla media, denotato con $\mathbb{E}[(X-\mu)^k]$, sempre nell'ipotesi in cui esso esista finito.

Il momento centrale di ordine 2 costituisce un caso di particolare importanza.

Definizione 2.15 Sia X una variabile aleatoria definita su un dato spazio di probabilità $(\Omega, \mathcal{A}, P)$ avente momento secondo $\mathbb{E}[X^2]$ finito. Si definisce *varianza* di X la quantità $\text{Var}(X) := \sigma^2 = \mathbb{E}[(X-\mu)^2]$. La quantità $\sigma := \sqrt{\text{Var}(X)}$ viene chiamata *scarto quadratico medio* o *deviazione standard* della v.a. X.

Si noti che in particolare risulta:

$$\begin{aligned}\text{Var}(X) &= \mathbb{E}[(X-\mu)^2] = \mathbb{E}[X^2 - 2\mu X + \mu^2] = \mathbb{E}[X^2] - 2\mu\mathbb{E}[X] + \mu^2 \\ &= \mathbb{E}[X^2] - \mu^2 ,\end{aligned} \tag{2.26}$$

essendo $\mathbb{E}[X] = \mu$.
Sia $\mathbb{E}[X^2] < \infty$. Allora si definisce la *v.a. standardizzata* Z

$$Z := \frac{X - \mathbb{E}[X]}{\sqrt{\text{Var}(X)}} = \frac{X-\mu}{\sigma} . \tag{2.27}$$

Si noti che risulta $\mathbb{E}[Z] = 0$ e $\text{Var}(Z) = 1$.

2.5.1 Proprietà della varianza.

1. Se α è una costante, allora:
$$\text{Var}(\alpha) = 0 .$$

2. Se X è una v.a. e α, β sono delle costanti, allora la v.a. $\alpha X + \beta$ ha momento secondo finito e si ha:
$$\text{Var}(\alpha X + \beta) = \alpha^2 \text{Var}(X) . \tag{2.28}$$

   Infatti, posto $\mathbb{E}[X] = \mu$, risulta:

$$\begin{aligned}\text{Var}(\alpha X + \beta) &= \mathbb{E}\{[(\alpha X + \beta) - \mathbb{E}(\alpha X + \beta)]^2\} = \mathbb{E}\{[(\alpha X + \beta) - \alpha\mu - \beta]^2\} \\ &= \mathbb{E}\{[\alpha(X-\mu)]^2\} = \mathbb{E}\{\alpha^2 (X-\mu)^2\} = \alpha^2 \text{Var}(X) .\end{aligned}$$

3. Se X è una v.a. avente momento secondo finito e valore atteso μ e c è una costante qualunque, allora si dimostra la seguente relazione:

$$\mathbb{E}[(X-\mu)^2] \le \mathbb{E}[(X-c)^2] \qquad \text{per ogni } c \in \mathbb{R} \ .$$

4. Se X e Y sono due v.a. indipendenti aventi momento secondo finito, allora la v.a. $X+Y$ ha momento secondo finito e risulta:

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) \tag{2.29}$$

2.6 Disuguaglianze di Markov e Cebicev

La varianza, oltre a dare un'idea della dispersione intorno al valor medio, consente di valutare quantitativamente, mediante una maggiorazione certe probabilità, anche in mancanza di ulteriori informazioni sulla distribuzione della v.a. X.

Teorema 2.16 (Disuguaglianza di Markov) Sia X una v.a. a valori non negativi. Posto $\mu = \mathbb{E}[X]$ si ha per ogni $\alpha > \mu$:

$$P(X \ge \alpha) \le \frac{\mu}{\alpha} \ . \tag{2.30}$$

Dimostrazione. Per semplicità diamo la dimostrazione nel caso in cui X sia una v.a. discreta che assume valori $\{x_1, x_2, \dots, x_n\}$. Se $\alpha \le \mu$ l'asserto è banale, per cui possiamo considerare il caso $\alpha > \mu$. La tesi segue dalla definizione di valore atteso tenendo conto che

$$\mu = \sum_{i=1}^{n} p_i x_i \ge \sum_{i\,:\,x_i > \alpha} p_i x_i \ge \alpha \sum_{i\,:\,x_i > \alpha} p_i = \alpha P(X \ge \alpha) \ .$$

□

Corollario 2.17 (Disuguaglianza di Cebicev) Sia X una v.a. avente valore atteso $\mu = \mathbb{E}[X]$ e varianza $\sigma^2 = \text{Var}(X)$. Allora si ha per ogni $k \ge 1$:

$$P(|X-\mu| \ge k\sigma) \le \frac{1}{k^2} \ .$$

Dimostrazione. La dimostrazione segue applicando la disuguaglianza di Markov (2.30) alla v.a. $Y = (X-\mu)^2$ con $\alpha = (k\sigma)^2$. Tenendo conto che i due eventi $\{|X-\mu| \ge k\sigma\}$ e $\{(X-\mu)^2 \ge (k\sigma)^2\}$ coincidono e che $\mathbb{E}[Y] = \mathbb{E}[(X-\mu)^2] = \sigma^2$, segue:

$$P(|X-\mu| \ge k\sigma) \equiv P((X-\mu)^2 \ge (k\sigma)^2) \le \frac{\mathbb{E}[Y]}{k^2\sigma^2} = \frac{1}{k^2} \ .$$

□

Esempio 2.18 In un grande supermercato vi sono 30 casse. In un qualunque giorno di sabato, l'incasso globale di una di tali casse può essere assimilato ad una variabile aleatoria avente distribuzione non nota con valor medio $\mu = 2000$ euro e scarto quadratico medio $\sigma = 250$ euro (stimati in base a dati ricavati in precedenza). Valutare la probabilità che l'incasso globale della cassa n. 6 sia compreso fra 1500 e 2500 euro.

Sia X_i la quantità aleatoria "incasso di una cassa durante la giornata della i-esima cassa", con $i = 1, \ldots, 30$. In mancanza di ulteriori informazioni, si può ritenere che le variabili X_i siano indipendenti ed identicamente distribuite, anche se la distribuzione non è nota. Il quesito si risolve mediante un'applicazione della disuguaglianza di Cebycev:

$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2} \,.$$

Infatti si richiede la probabilità che l'incasso sia compreso fra 1500 e 2500 euro, cioè $P(1500 \leq X \leq 2500)$. Tale quantità può essere scritta come

$$P(1500 \leq X \leq 2500) = P(|X - 2000| \leq 500) \tag{2.31}$$

e poichè $\sigma = 250$ euro, allora dalla (2.31), segue $500 = 2 \cdot \sigma$. Pertanto risulta

$$P(|X - 2000| \leq 500) = P(|X - 2000| \leq 2 \cdot \sigma) \geq 1 - \frac{1}{2^2} = 0.75 \,.$$

♣

2.7 Covarianza e Correlazione.

Definizione 2.19 Siano X, Y due v.a. aventi varianza finita, ed indichiamo con μ_X e μ_Y le rispettive speranze matematiche. Si definisce *covarianza* di X e Y, e si indica col simbolo $\text{Cov}(X, Y)$, la quantità

$$\text{Cov}(X, Y) := \mathbb{E}\{[X - \mu_X][Y - \mu_Y]\} \,. \tag{2.32}$$

Si nota che la covarianza è uguale alla speranza del prodotto degli scarti delle variabili aleatorie X e Y dalle rispettive speranze, analogamente al caso statistico. In particolare si ha:

$$\text{Cov}(X, X) = \mathbb{E}\{[X - \mu_X][X - \mu_X]\} = \mathbb{E}\{[X - \mu_X]^2\} = \text{Var}(X) \,.$$

Dalla definizione seguono subito due proprietà:

$$\begin{aligned} \text{Cov}(X, \alpha) &= \mathbb{E}\{[X - \mu_X][\alpha - \alpha]\} = 0 && (2.33) \\ \text{Cov}(\alpha X, \beta Y) &= \mathbb{E}\{[\alpha X - \alpha\mu_X][\beta Y - \beta\mu_Y]\} \\ &= \alpha\beta\mathbb{E}\{[X - \mu_X][Y - \mu_Y]\} = \alpha\beta\text{Cov}(X, Y) \,. && (2.34) \end{aligned}$$

Inoltre, svolgendo la (2.32), si ottiene:

$$\begin{aligned}\text{Cov}(X,Y) &= \mathbb{E}\{[X-\mu_X][Y-\mu_Y]\} = \mathbb{E}[XY-\mu_X Y-\mu_Y X+\mu_X\mu_Y]\\ &= \mathbb{E}[XY]-\mu_X\mathbb{E}[Y]-\mu_Y\mathbb{E}[X]+\mu_X\mu_Y\\ &= \mathbb{E}[XY]-\mu_X\mu_Y\,. \end{aligned} \tag{2.35}$$

Dalla (2.24), segue che se X e Y sono stocasticamente indipendenti, allora $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ e quindi si ha Cov$(X,Y)=0$.

La covarianza di X e Y è definita se le due v.a. hanno varianza finita, ed in tal caso si dimostra il seguente risultato.

Proposizione 2.20 Sia (X,Y) una coppia di v.a. aventi varianza finita. Risulta allora:

$$|\text{Cov}(X,Y)| \le \sqrt{\text{Var}(X)\,\text{Var}(Y)}\,. \tag{2.36}$$

In particolare l'uguaglianza vale solo in uno dei due seguenti casi:

1. se almeno una delle due v.a., X o Y, è costante;

2. se X è funzione lineare di Y, cioè sono legati da una relazione del tipo $Y=\gamma X+\beta$.

Dimostrazione. Per $\alpha\in\mathbb{R}$ si consideri la varianza di $\alpha X+Y$:

$$\text{Var}(\alpha X+Y)=\alpha^2\text{Var}(X)+\text{Var}(Y)+2\alpha\text{Cov}(X,Y)$$

e dovendo risultare Var$(\alpha X+Y)\ge 0$ per ogni α, deve risultare:

$$\text{Cov}^2(X,Y)-\text{Var}(X)\text{Var}(Y)\le 0$$

da cui segue la (2.36).

Per quanto concerne l'uguaglianza nella (2.36), essa è banalmente verificata se una delle due v.a., ad esempio Y, è uguale ad una costante c in quanto Var$(c)=0$ e Cov$(X,c)=0$ e quindi si ha l'identità $0=0$. Nell'altro caso in cui $Y=\gamma X+\beta$, risulta:

$$\begin{aligned}\text{Cov}(X,Y) &= \text{Cov}(X,\gamma X+\beta)=\gamma\text{Cov}(X,X)=\gamma\text{Var}(X)\\ \text{Var}(Y) &= \gamma^2\text{Var}(X)\end{aligned}$$

e quindi:

$$\text{Cov}^2(X,Y)=\gamma^2\text{Var}^2(X)=\text{Var}(X)\cdot\text{Var}(Y)\,.$$

□

2.7.1 Varianza della somma di variabili aleatorie

Siano X, Y due v.a. aventi varianza finita e speranze matematiche rispettivamente uguali a μ_X e μ_Y. Si ha allora:

$$\begin{aligned}\text{Var}(X+Y) &= \mathbb{E}[\{(X+Y)-(\mu_X+\mu_Y)\}^2] = \mathbb{E}[\{(X-\mu_X)+(Y-\mu_Y)\}^2] \\ &= \mathbb{E}[(X-\mu_Y X)^2] + \mathbb{E}[Y-\mu_Y)^2] + \mathbb{E}[(X-\mu_X)(Y-\mu_Y)] \\ &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)\ . \end{aligned} \tag{2.37}$$

Più in generale, si può dimostrare la relazione:

$$\text{Var}(\alpha X + \beta Y) = \alpha^2 \text{Var}(X) + \beta^2 \text{Var}(Y) + 2\alpha\beta \text{Cov}(X,Y)\ . \tag{2.38}$$

In particolare, nel caso in cui X e Y sono v.a. indipendenti risulta Cov$(X,Y) = 0$ e quindi si ritrova la (2.29).

La (2.37) si estende al caso n dimensionale come segue. Se $X, \ldots, X_m$ sono v.a. aventi varianza finita, allora si ha

$$\text{Var}(X_1 + \cdots + X_m) = \sum_{i=1}^{m} \text{Var}(X_i) + 2\sum_{i<j} \text{Cov}(X_i, X_j)\ . \tag{2.39}$$

In particolare, se le m v.a. sono indipendenti, allora risulta Cov$(X_i, X_j) = 0$ per $i \neq j$ e quindi dalla (2.39) segue:

$$\text{Var}(X_1 + \cdots + X_m) = \sum_{i=1}^{m} \text{Var}(X_i)\ . \tag{2.40}$$

Tale relazione generalizza la (2.29).

In generale, posto $\sigma_{ij} = \text{Cov}(X_i, X_j)$ e $\sigma_{ii} = \text{Var}(X_i)$, per $i, j = 1, \ldots, m$, si definisce *matrice di varianze e covarianze* la matrice:

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{m1} & \sigma_{m2} & \cdots & \sigma_{mm} \end{pmatrix}\ .$$

Definizione 2.21 Date due v.a. X e Y, aventi varianza non nulla, si definisce *coefficiente di correlazione lineare* di X e Y, e si indica col simbolo ρ, la quantità

$$\rho \ = \ \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}\ . \tag{2.41}$$

dove $\sigma_X = \sqrt{\text{Var}(X)}$ e $\sigma_Y = \sqrt{\text{Var}(Y)}$.

In altre parole, in base alla (2.36), il coefficiente di correlazione è uguale alla covarianza normalizzata, cioè la covarianza diviso il valore massimo teoricamente possibile. Ne segue che si ha $-1 \leq \rho \leq 1$.

Due proprietà del coefficiente di correlazione:

1. $\rho(X,Y)$ è invariante per scala, cioè risulta $\rho(\alpha X, \beta Y) = \rho(X,Y)$ per ogni $\alpha, \beta \in \mathbb{R}$. Infatti, in base alle (2.28) e (2.34), segue:
$$\begin{aligned}\rho(\alpha X, \beta Y) &= \frac{\mathrm{Cov}(\alpha X, \beta Y)}{\sqrt{\mathrm{Var}(\alpha X)\mathrm{Var}(\beta Y)}} = \frac{\alpha\beta \cdot \mathrm{Cov}(X,Y)}{\sqrt{\alpha^2\mathrm{Var}(X)\cdot\beta^2\mathrm{Var}(Y)}} \\ &= \frac{\alpha\beta \cdot \mathrm{Cov}(X,Y)}{\alpha\beta\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \rho(X,Y)\ .\end{aligned}$$

2. Risulta $|\rho(X,Y)| = 1$ se e solo se esistono $\alpha, \beta \in \mathbb{R}$ tali che $Y = \alpha X + \beta$.
Infatti, sia $\rho(X,Y) = 1$, allora in base alla definizione vuol dire che $\mathrm{Cov}^2(X,Y) = \mathrm{Var}(X)\mathrm{Var}(Y)$ e quindi in base a quanto detto prima, segue l'asserto.
Al contrario, se $Y = \alpha X + \beta$, allora $\mathrm{Cov}^2(X,Y) = \alpha^2\mathrm{Var}^2(X)$ e $\mathrm{Var}(Y) = \alpha^2\mathrm{Var}(X)$.

Ovviamente si può considerare la *matrice di correlazione*. Posto
$$r_{ij} := \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}$$
possiamo scrivere:
$$\mathbf{R} = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ r_{m1} & r_{m2} & \cdots & r_{mm} \end{pmatrix}\ ,$$
dove ovviamente risulta:
$$r_{11} = r_{22} = \cdots = r_{mm} = 1\ .$$

Esempio 2.22 Si prendano in considerazione due diverse possibilità di giocare 100 euro al lotto, supponendo che per ogni euro giocato si abbia un ricavo atteso pari a 80 centesimi con una deviazione standard di 50 centesimi. La prima possibilità, indicata con A, consiste nel giocare rispettivamente 20, 30 e 50 euro su tre ruote s, t, u scelte a caso. La seconda possibilità, indicata con B, consiste nel giocare tutta la somma di 100 euro su una sola ruota v estratta a caso. Qual è, fra le due possibilità A e B, quella meno rischiosa?

Soluzione. Innanzitutto è bene osservare che le due strategie di gioco A e B hanno la stessa speranza matematica. Infatti, denotato con X la v.a. ”ricavo in una giocata per una puntata pari a un euro”, in base ai dati del problema risulta $\mathbb{E}[X] = 0.8$ e $\mathrm{Var}(X) = 0.5$. Denotate poi con A e B i ricavi ottenuti rispettivamente in base alle due strategie A e B, si ha:
$$\begin{aligned}A &:= 20X_s + 30X_t + 50X_u \\ B &:= 100X_v\end{aligned}$$
dove X_s, X_t, X_u e X_v sono le quantità aleatorie ”ricavo da una giocata sulla ruota considerata (s, t, u, v)” aventi tutte la stessa speranza matematica e la stessa varianza. da cui
$$\begin{aligned}\mathbb{E}[A] &= 20\mathbb{E}[X_s] + 30\mathbb{E}[X_t] + 50\mathbb{E}[X_u] = 20\cdot 0.8 + 30\cdot 0.8 + 50\cdot 0.8 = 80 \\ \mathbb{E}[B] &= 100\cdot\mathbb{E}[X_v] = 100\cdot 0.8 = 80\ .\end{aligned}$$

Sono invece diverse invece le varianze. Dalla (2.40) si ha infatti, tenento conto che Var$(X) = 0.5^2 = 0.25$:

$$\begin{aligned}\text{Var}(A) &= \text{Var}(20X_s + 30X_t + 50X_u) = 20^2\text{Var}(X_s) + 30^2\text{Var}(X_t) + 50^2\text{Var}(X_u) \\ &= 400 \cdot 0.25 + 900 \cdot 0.25 + 2500 \cdot 0.25 = 950 \\ \text{Var}(B) &= \text{Var}(100X_v) = 10000\text{Var}(X_v) = 10000 \cdot 0.25 = 2500 \,.\end{aligned}$$

Quindi, sebbene i due tipi di gioco tendano, per un elevato numero di giocate, a far perdere 20 euro per ogni 100 euro giocati, questa perdita si realizza in modo diverso nella strategia A da quella che si realizzarebbe nella strategia B. Il metodo consiste nel dividere la somma che si vuole giocare e giocarla in più volte riduce considerevolmente la varianza del gioco, ed è inferiore per puntate multiple di somme piccole piuttosto che una sola puntata elevata, riducendo quindi il rischio del gioco.

Come caso limite, se si dividono i 100 euro in 100 parti uguali, si avrà una varianza di vincita 100 volte più piccola, cioè una deviazione standard di 10 volte più piccola di quella che si avrebbe giocando tutta la somma in un sol colpo. ♣

Esercizio 2.23 Sia $\Omega = \{a, b, c\}$ e P la distribuzione di probabilità su Ω definita da:

$$P(a) = \frac{1}{2}\,, \qquad P(b) = P(c) = \frac{1}{4}\,.$$

Siano X e Y la variabili aleatorie definite da:

$$X = \mathbb{1}_{\{a\}} - \mathbb{1}_{\{b,c\}}\,, \qquad Y = \mathbb{1}_{\{b\}} - \mathbb{1}_{\{c\}}\,,$$

dove $\mathbb{1}_A(\omega)$ è la funzione indicatrice, cioè $\mathbb{1}_A(\omega) = 1$ se $\omega \in A$ e $\mathbb{1}_A(\omega) = 0$ se $\omega \notin A$. Dimostrare che X e Y non sono stocasticamente indipendenti e che si ha Cov$(X, Y) = 0$.

Soluzione. In base alle ipotesi del problema, si ha $X \in \{-1, 1\}$ e $Y \in \{-1, 0, 1\}$, in particolare:

$$\begin{aligned}\text{evento } a &\;\Rightarrow\; X = 1\,, Y = 0 \\ \text{evento } b &\;\Rightarrow\; X = -1\,, Y = 1 \\ \text{evento } c &\;\Rightarrow\; X = -1\,, Y = -1\,.\end{aligned}$$

Si ha pertanto la seguente distribuzione di probabilità della variabile congiunta (X, Y):

		Y		
X	-1	0	$+1$	
-1	1/4	0	1/4	1/2
$+1$	0	1/2	0	1/2
	1/4	1/2	1/4	1

Ovviamente X e Y non sono stocasticamente indipendenti in quanto

$$P(X=-1, Y=0) = 0 \neq P(X=-1) \cdot P(Y=0) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} .$$

Invece X e Y sono non correlate in quanto risulta Cov$(X,Y)=0$. Infatti:

$$\begin{aligned}
\mathbb{E}[XY] &= \frac{1}{4} \cdot (-1) \cdot (-1) + \frac{1}{2} \cdot 1 \cdot 0 + \frac{1}{4} \cdot (-1) \cdot 1 = 0 \\
\mathbb{E}[X] &= -1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = 0 \\
\mathbb{E}[Y] &= -1 \cdot \frac{1}{4} + 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{4} = 0
\end{aligned}$$

da cui

$$\text{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0 .$$

♣

2.8 Altri esercizi

Esercizio 2.24 Completare la seguente distribuzione di probabilità congiunta nell'ipotesi di indipendenza stocastica delle v.a. X e Y:

	x_1	x_2	x_3	
y_1	*	*	*	*
y_2	0.06	*	*	0.3
y_3	*	0.18	*	*
	*	0.3	*	1

Esercizio 2.25 Completare la seguente distribuzione di probabilità congiunta nell'ipotesi di indipendenza stocastica delle v.a. X e Y:

	x_1	x_2	x_3	
y_1	*	0.2	*	0.4
y_2	*	*	0.105	*
y_3	*	*	*	*
	*	*	0.3	1

Esercizio 2.26 Una moneta regolare (testa/croce) viene lanciata tre volte. Prima di ogni lancio, è possibile fare una scommessa di importo arbitrario sull'evento T (testa). Si considerino le tre strategie:

a. *Lascia o raddoppia* - Si scommette un euro sul primo lancio: in caso di vincita si ”lascia”, in caso di perdita si ”raddoppia” scommettendo due euro sul secondo lancio; se si vince si lascia, se si perde ancora, si scommettono quattro euro sul terzo lancio.

b. Si scommette un euro sul primo lancio: in caso di vincita si sospende il gioco, in caso di perdita si decide si scommettere un euro sia al secondo che al terzo lancio.

c. Si scommette un euro su ciascun lancio.

Calcolare il valore atteso del guadagno su ciascun lancio per ognuna delle tre strategie; valutare successivamente quale delle tre strategie di gioco presenta la maggiore variabilità.

Esercizio 2.27 Un promotore finanziario propone ad uno dei suoi clienti due diverse possibilità di un investimento 5000 euro. La prima possibilità, indicata con A, consiste nell’investire rispettivamente 1000, 1500 e 2500 euro su tre titoli azionari s, t, u; la seconda possibilità, indicata con B, consiste nell’investire l’intera somma di 5000 euro su un solo titolo azionario v. Si assume che i ricavi per tutti i quattro titoli siano stocasticamente indipendenti e seguano le stesse distribuzioni di probabilità tali che – per ogni euro investito in uno qualsiasi dei quattro titoli azionari in esame e nei tempi indicati dal contratto di sottoscrizione – il ricavo abbia una speranza pari a 1.1 euro con una deviazione standard di 0.5 euro. *i*) Qual è, fra le due possibilità A e B, la più redditizia? *ii*) Qual è quella meno rischiosa?

Esercizio 2.28 Sia X la v.a. ”numero di teste ottenute in due lanci consecutivi di una moneta equilibrata”.

1. Descrivere lo spazio degli eventi.

2. Calcolare la funzione di densità di X e rappresentarla graficamente;

3. Calcolare la funzione di ripartizione di X e rappresentarla graficamente;

4. Calcolare la costante c tale che $\mathbb{E}(cX) = 5$.

Esercizio 2.29 Si consideri una moneta regolare e si indichino con C e T rispettivamente gli eventi ”croce” e ”testa” nel generico lancio.

a) Si considerino tre lanci della moneta: si calcoli la probabilità che non si presenti mai la successione CT.

b) Si considerino quattro lanci della moneta: si calcoli la probabilità che non si verifichino le successioni CT o TCC.

Esercizio 2.30 Si consideri la seguente distribuzione concernente una v.a. doppia discreta (X, Y):

		X			
		2	3	5	
	0	0.30	*	0.00	0.40
Y	1	*	0.15	*	0.35
	2	0.15	*	0.10	0.25
		0.45	0.25	0.30	1.00

Calcolare i valori incogniti e successivamente il coefficiente di correlazione di X e Y.

Esercizio 2.31 Si consideri la seguente distribuzione concernente una v.a. doppia discreta (X, Y):

		X			
		2	3	5	
	0	0.30	0.10	0.00	0.40
Y	1	0.00	0.15	0.20	0.35
	2	0.15	0.00	0.10	0.25
		0.45	0.25	0.30	1.00

Calcolare:

1. le distribuzioni marginali di X e di Y;
2. la distribuzione condizionata di $Y|X = 3$;
3. La covarianza di X e Y.