

## Capitolo 3

### Variabili casuali discrete

#### 3.1 Variabili casuali e loro distribuzioni

In molti degli esempi dei capitoli precedenti, abbiamo calcolato delle probabilità di eventi che si potevano esprimere in termini di una funzione dell'esito di un esperimento aleatorio.

*Esempio 3.1.* Riprendiamo qui l'esempio 1.17, in cui lo spazio campionario è

$$\Omega = \{x = (x_1, x_2, \dots, x_N) : x_i \in \{0, 1\}\} = \{0, 1\}^N,$$

e la probabilità  $P$  è data da

$$P(\{y\}) = p^{\sum_{i=1}^N y_i} (1-p)^{N-\sum_{i=1}^N y_i}.$$

Abbiamo calcolato la probabilità dell'evento

$$A = \{x \in \Omega : \sum_{i=1}^N x_i = n\},$$

dove  $0 \leq n \leq N$ , e dell'evento

$$B = \{x \in \Omega : x_n = 1, x_k = 0 \forall k < n\}.$$

Posto  $X(x) = \sum_{i=1}^N x_i$ , e

$$Y(x) = \begin{cases} \min\{k : x_k = 1\} & \text{se } \{k : x_k = 1\} \neq \emptyset \\ N+1 & \text{altrimenti,} \end{cases}$$

possiamo riscrivere

$$A = \{x \in \Omega : X(x) = n\}$$

e

$$B = \{x \in \Omega : Y(x) = n\}.$$

**Definizione 3.1.** Sia  $(\Omega, P)$  uno spazio di probabilità discreto, e  $E$  un insieme. Una funzione  $X : \Omega \rightarrow E$  si dice *variabile casuale* o *variabile aleatoria* a valori in  $E$ .

Introduciamo un po' di terminologia e di notazioni. Con riferimento alla definizione precedente, se  $E = \mathbb{R}$  diciamo che  $X$  è una variabile casuale *scalare*. Se  $E = \mathbb{R}^n$  diciamo che  $X$  è una variabile casuale *vettoriale* o *vettore aleatorio* di dimensione  $n$ . Se  $E = \mathbb{C}$ , diciamo che  $X$  è una variabile casuale *complessa*.

Se  $A \subseteq E$ , allora  $X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\}$  è un sottoinsieme di  $\Omega$ . Scriviamo  $\{X \in A\}$  e  $P(X \in A)$  in luogo di  $X^{-1}(A)$  e  $P(X^{-1}(A))$  rispettivamente. Nel caso in cui  $A = \{x\}$ , con  $x \in E$ , scriveremo  $\{X = x\}$  e  $P(X = x)$  invece di  $X^{-1}(\{x\})$  e  $P(X^{-1}(\{x\}))$ . Se  $(A_n)_{n \geq 0}$  è una successione di sottoinsiemi disgiunti di  $E$ , allora gli eventi  $\{X \in A_n\}$  sono disgiunti (verificarlo!). Ne segue che

$$P\left(X \in \bigcup_n A_n\right) = \sum_n P(X \in A_n). \quad (3.1)$$

Si noti infine che, essendo  $\Omega$  un insieme finito o numerabile, anche l'immagine di  $X$ ,  $X(\Omega) = \{X(\omega) : \omega \in \Omega\}$ , è un insieme finito o numerabile (in generale, la cardinalità dell'immagine di una funzione è minore o uguale a quella del dominio della stessa). Perciò, se  $A \subseteq E$ , allora  $A \cap X(\Omega)$  è finito o numerabile. Essendo (perché?)  $P(X \in A) = P(X \in A \cap X(\Omega))$ , usando (3.1) si ha

$$\begin{aligned} P(X \in A) &= P\left(X \in \bigcup_{x \in A \cap X(\Omega)} \{x\}\right) \\ &= \sum_{x \in A \cap X(\Omega)} P(X = x). \end{aligned}$$

Poichè, evidentemente,  $P(X = x) = 0$  se  $x \notin X(\Omega)$ , possiamo sinteticamente scrivere l'identità precedente come

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

Possiamo riassumere le precedenti osservazioni come segue.

**Definizione 3.2.** Sia  $X$  una variabile casuale discreta a valori in  $E$ . La funzione

$$\begin{aligned} \mu_X : \mathcal{P}(E) &\rightarrow [0, 1] \\ A &\mapsto P(X \in A) \end{aligned}$$

si dice *distribuzione* o *legge* della variabile casuale  $X$ .

**Proposizione 3.1.** La distribuzione  $\mu_X$  di una variabile casuale discreta  $X$  a valori in  $E$  gode delle seguenti proprietà:

- (i) se  $(A_n)$  è una successione di sottoinsiemi disgiunti di  $E$ , allora

$$\mu_X \left( \bigcup_n A_n \right) = \sum_n \mu_X(A_n);$$

(ii) per ogni sottoinsieme  $A$  di  $E$

$$\mu_X(A) = \sum_{x \in A} P(X = x) = \sum_{x \in A} \mu_X(\{x\}).$$

*Osservazione 3.1.* Nel caso in cui  $E$  è un insieme finito o numerabile, segue dalla Proposizione 3.1 che la coppia  $(E, \mu_X)$  è uno spazio di probabilità discreto. Va tuttavia notato che, in generale, non facciamo alcuna assunzione sulla cardinalità di  $E$ . Abbiamo però osservato che  $X(\Omega)$  è sicuramente un insieme finito o numerabile. Denotando con  $\tilde{\mu}_X$  la restrizione di  $\mu_X$  ai sottoinsiemi di  $X(\Omega)$ , si ha che  $(X(\Omega), \tilde{\mu}_X)$  è uno spazio di probabilità discreto.

## 3.2 Densità discreta

Sia  $X$  una variabile casuale discreta a valori in un insieme  $E$ .

**Definizione 3.3.** La funzione

$$\begin{aligned} p_X : E &\rightarrow [0, 1] \\ x &\mapsto P(X = x) \end{aligned}$$

si dice *densità discreta* della variabile casuale  $X$ .

Nel seguito l'aggettivo “discreta” verrà omissso, fino a quando non introdurremo, nel capitolo 5, la nozione “continua” di densità.

*Osservazione 3.2.* La densità  $p_X$  e la distribuzione  $\mu_X$  di una variabile casuale discreta, si determinano l'un l'altra tramite le relazioni, per  $x \in E, A \subseteq E$

$$p_X(x) = \mu_X(\{x\}), \quad \mu_X(A) = \sum_{x \in A} p_X(x). \quad (3.2)$$

Consideriamo ora la situazione in cui una variabile casuale  $X$  prende valori in un insieme della forma  $E = E_1 \times E_2 \times \cdots \times E_n$ . In questo caso possiamo scrivere, per ogni  $\omega \in \Omega$ ,

$$X(\omega) = (X_1(\omega), X_2(\omega), \dots, X_n(\omega)),$$

dove  $X_i$ , per  $i = 1, \dots, n$ , è una variabile casuale a valori in  $E_i$ . La densità  $p_X : E \rightarrow [0, 1]$  di  $X$  è spesso denotata con  $p_{X_1, \dots, X_n}$ , e viene chiamata *densità congiunta* delle variabili casuali  $X_1, X_2, \dots, X_n$ . Consideriamo ora un sottoinsieme  $I = \{i_1, \dots, i_m\} \subseteq \{1, 2, \dots, n\}$  e conveniamo  $i_1 < i_2 < \cdots < i_m$ . Possiamo dunque considerare la variabile casuale  $(X_{i_1}, X_{i_2}, \dots, X_{i_m})$  a valori in  $E_{i_1} \times E_{i_2} \times \cdots \times E_{i_m}$ . La seguente proposizione stabilisce che la densità di tale variabile casuale è determinabile a partire dalla densità della variabile casuale  $X$ .

**Proposizione 3.2.** *La densità  $p_{X_{i_1}, X_{i_2}, \dots, X_{i_m}}$  della variabile casuale  $(X_{i_1}, X_{i_2}, \dots, X_{i_m})$  è data dalla relazione*

$$p_{X_{i_1}, X_{i_2}, \dots, X_{i_m}}(x_{i_1}, x_{i_2}, \dots, x_{i_m}) = \sum_{j \notin I} \sum_{x_j \in E_j} p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n). \quad (3.3)$$

*Dimostrazione.* Dimostriamo il caso  $n = 2, m = 1, I = \{1\}$ . Il caso generale segue o usando lo stesso argomento, con notazioni più pesanti, oppure osservando che è sempre possibile ridursi a questo caso osservando che, posto  $E_I = \times_{i \in I} E_i$  e  $X_I = (X_{i_1}, X_{i_2}, \dots, X_{i_m})$ , a meno di una permutazione di coordinate, l'insieme  $E$  si può identificare con  $E_I \times E_{I^c}$ , e la variabile  $X$  con la coppia di variabili  $(X_I, X_{I^c})$  a valori rispettivamente in  $E_I$  e  $E_{I^c}$ .

Dunque, si tratta di dimostrare che, per ogni  $x_1 \in E_1$ ,

$$p_{X_1}(x_1) = \sum_{x_2 \in E_2} p_{X_1, X_2}(x_1, x_2). \quad (3.4)$$

Si osservi che l'evento  $\{X_1 = x_1\}$  può essere espresso come unione al più numerabile di eventi disgiunti come segue

$$\{X_1 = x_1\} = \bigcup_{x_2 \in X_2(\Omega)} \{X_1 = x_2, X_2 = x_2\}.$$

Pertanto, per la  $\sigma$ -additività,

$$\begin{aligned} p_{X_1}(x_1) &= P(X_1 = x_1) = \sum_{x_2 \in X_2(\Omega)} P(X_1 = x_2, X_2 = x_2) = \sum_{x_2 \in X_2(\Omega)} P((X_1, X_2) = (x_1, x_2)) \\ &= \sum_{x_2 \in X_2(\Omega)} p_{X_1, X_2}(x_1, x_2) = \sum_{x_2 \in E_2} p_{X_1, X_2}(x_1, x_2) \end{aligned}$$

dove, per l'ultima uguaglianza, abbiamo usato il fatto che se  $x_2 \notin X_2(\Omega)$ , allora l'evento  $\{(X_1, X_2) = (x_1, x_2)\}$  è vuoto e ha dunque probabilità zero.

Nel caso in cui  $k = 1$ , la Proposizione 3.2 consente di esprimere la densità di una componente di un vettore aleatorio in termini della densità congiunta. Le densità delle componenti vengono chiamate *densità marginali*. Come mostriamo nei due seguenti esempi, non è possibile ricostruire la densità congiunta a partire dalla sola conoscenza delle densità marginali.

*Esempio 3.2.* Ad un esame partecipano  $n$  studenti, ed  $m < n$  di essi portano con loro dei testi il cui uso non è consentito, nella speranza di non venire controllati. I due docenti addetti alla sorveglianza, Aldo e Giovanni, decidono di eseguire dei controlli casuali. Aldo controlla  $h$  studenti, mentre Giovanni ne controlla altri  $k$ , distinti da quelli controllati da Aldo. Supponiamo che se uno studente che ha con sé testi non consentiti viene controllato, venga senz'altro espulso dall'aula. Sia  $X_A$  (risp.  $X_G$ ) il numero di studenti espulsi da Aldo (risp. Giovanni). Si determinino le densità congiunte e marginali di  $X_A$  e  $X_G$ .

Supponiamo di “etichettare” gli studenti con numeri tra 1 e  $n$ , assegnando etichetta tra 1 e  $m$  a quelli che hanno portato testi non permessi. Come spazio campionario, possiamo scegliere

$$\Omega = \{(A, G) : A, G \subseteq \{1, 2, \dots, n\}, |A| = h, |G| = k, A \cap G = \emptyset\},$$

munito della probabilità uniforme  $P$ . Le variabili casuali  $X_A$  e  $X_G$  si possono allora definire come segue:

$$X_A(A, G) = |A \cap \{1, 2, \dots, m\}|, \quad X_G(A, G) = |G \cap \{1, 2, \dots, m\}|.$$

Siano  $x_A, x_G$  due interi fissati, con  $1 \leq x_A \leq h, 1 \leq x_G \leq k$ . Anzitutto osserviamo che

$$\begin{aligned} p_{X_A, X_G}(x_A, x_G) &= P(X_A = x_A, X_G = x_G) \\ &= P(X_G = x_G | X_A = x_A) P(X_A = x_A). \end{aligned}$$

Da quanto visto nell'esempio 1.9

$$P(X_A = x_A) = p_{X_A}(x_A) = \frac{\binom{m}{x_A} \binom{n-m}{h-x_A}}{\binom{n}{h}}, \quad (3.5)$$

con la solita convenzione che  $\binom{a}{b} = 0$  se  $b > a$ . Inoltre, condizionatamente al verificarsi dell'evento  $\{X_A = x_A\}$ , Giovanni sceglie  $k$  studenti tra gli  $n - h$  non controllati da Aldo, dei quali  $m - x_A$  hanno testi non ammessi. Perciò

$$P(X_G = x_G | X_A = x_A) = \frac{\binom{m-x_A}{x_G} \binom{n-h-m+x_A}{k-x_G}}{\binom{n-h}{k}}.$$

Mettendo tutto assieme:

$$p_{X_A, X_G}(x_A, x_G) = \frac{\binom{m}{x_A} \binom{n-m}{h-x_A}}{\binom{n}{h}} \frac{\binom{m-x_A}{x_G} \binom{n-h-m+x_A}{k-x_G}}{\binom{n-h}{k}}. \quad (3.6)$$

La densità marginale  $p_{X_A}$  è già stata calcolata in (3.5). Poichè il problema è completamente simmetrico nel ruolo di Aldo e Giovanni, esattamente come in (3.5) troviamo che

$$p_{X_G}(x_G) = \frac{\binom{m}{x_G} \binom{n-m}{k-x_G}}{\binom{n}{k}}. \quad (3.7)$$

Si noti infine che, dalla proposizione 3.2, sappiamo che

$$p_{X_G}(x_G) = \sum_{x_A=1}^h p_{X_A, X_G}(x_A, x_G).$$

Sostituendo i valori ottenuti per le densità in quest'ultima relazione, si ottiene una relazione combinatoria assolutamente non banale!

*Esempio 3.3.* Nello stesso contesto dell'esempio 3.2, supponiamo che Aldo e Giovanni effettuino i controlli in momenti successivi, senza comunicare tra loro, e che soltanto alla fine dei controlli di entrambi venga comunicata agli studenti "imbroglianti" la loro espulsione. In questo caso, dunque, non si esclude la possibilità che uno stesso studente sia controllato sia da Aldo che da Giovanni. Lo spazio campionario che descrive questa procedura è

$$\Omega = \{(A, G) : A, G \subseteq \{1, 2, \dots, n\}, |A| = h, |G| = k\},$$

munito della probabilità uniforme. Le variabili casuali  $X_A$  e  $X_G$  sono definite come nell'esempio 3.2. Inoltre, la diversa procedura di controllo è del tutto ininfluente per il calcolo delle densità marginali, che risultano identiche a quelle calcolate nell'esempio precedente.

Per il calcolo della densità congiunta, si osservi che  $\Omega = \Omega_A \times \Omega_G$ , dove

$$\Omega_A = \{A \subseteq \{1, 2, \dots, n\} : |A| = h\}$$

$$\Omega_G = \{G \subseteq \{1, 2, \dots, n\} : |G| = k\}.$$

Se  $H \subseteq \Omega_A$  e  $K \subseteq \Omega_G$ , gli eventi in  $\Omega$   $H \times \Omega_G$  e  $\Omega_A \times K$  sono indipendenti. Infatti:

$$\begin{aligned} P[(H \times \Omega_G) \cap (\Omega_A \times K)] &= P(H \times K) \\ &= \frac{|H||K|}{|\Omega_A||\Omega_G|} \\ &= \frac{|H||\Omega_G|}{|\Omega_A||\Omega_G|} \frac{|\Omega_A||K|}{|\Omega_A||\Omega_G|} \\ &= P(H \times \Omega_G)P(\Omega_A \times K). \end{aligned}$$

Si noti ora che l'evento  $\{X_A = x_A\}$  è della forma  $H \times \Omega_G$  con  $H = \{A : |A \cap \{1, \dots, m\}| = x_A\}$ , e similmente  $\{X_G = x_G\} = \Omega_A \times K$  con  $K = \{G : |G \cap \{1, \dots, m\}| = x_G\}$ . Dunque i due eventi sono indipendenti. Ma allora

$$\begin{aligned} p_{X_A, X_G}(x_A, x_G) &= P(X_A = x_A, X_G = x_G) \\ &= P(X_A = x_A)P(X_G = x_G) \\ &= p_{X_A}(x_A)p_{X_G}(x_G), \end{aligned}$$

che è diversa dalla densità congiunta trovata nell'esempio precedente, pur essendo le stesse le densità marginali

### 3.3 Indipendenza di variabili casuali

Cominciamo col definire l'indipendenza per una famiglia finita di variabili casuali.

**Definizione 3.4.** Siano  $X_1, X_2, \dots, X_n$  variabili casuali definite nello stesso spazio di probabilità  $(\Omega, P)$ , a valori rispettivamente negli insiemi  $E_1, E_2, \dots, E_n$ . Esse si dicono indipendenti se per ogni scelta di  $A_1 \subseteq E_1, A_2 \subseteq E_2, \dots, A_n \subseteq E_n$  si ha

$$P(X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n) = \prod_{i=1}^n P(X_i \in A_i). \quad (3.8)$$

Osserviamo che se  $X_1, X_2, \dots, X_n$  sono variabili casuali indipendenti e  $\{i_1, i_2, \dots, i_k\} \subset \{1, 2, \dots, n\}$ , allora  $X_{i_1}, X_{i_2}, \dots, X_{i_k}$  sono variabili casuali indipendenti. Infatti se, in (3.8) si sceglie  $A_j = E_j$  per  $j \notin \{i_1, i_2, \dots, i_k\}$  si ottiene

$$\begin{aligned} P(X_{i_1} \in A_{i_1}, X_{i_2} \in A_{i_2}, \dots, X_{i_k} \in A_{i_k}) &= P(X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n) \\ &= \prod_{i=1}^n P(X_i \in A_i) \\ &= \prod_{j=1}^k P(X_{i_j} \in A_{i_j}). \end{aligned}$$

Più in generale, per famiglie non necessariamente finite di variabili casuali, si dà la seguente definizione che, grazie all'osservazione appena fatta, è consistente con la Definizione 3.4

**Definizione 3.5.** Sia  $I$  un insieme qualsiasi di indici, e  $\{X_i : i \in I\}$  una famiglia di variabili casuali a valori negli insiemi  $E_i, i \in I$ . Si dice che le variabili casuali di tale famiglia sono indipendenti se, per ogni  $J \subset I$  finito e per ogni scelta di  $A_j \subseteq E_j, j \in J$ , si ha

$$P\left(\bigcap_{j \in J} \{X_j \in A_j\}\right) = \prod_{j \in J} P(X_j \in A_j).$$

Il semplice confronto tra la Definizione 3.5 e quella di indipendenza tra eventi, ci fornisce la seguente proprietà.

**Proposizione 3.3.** Siano  $\{X_i : i \in I\}$  variabili casuali come nella definizione 3.5. Le seguenti affermazioni sono equivalenti:

- (i) La variabili casuali  $\{X_i : i \in I\}$  sono indipendenti;
- (ii) per ogni scelta di  $A_i \subseteq E_i$ , gli eventi  $\{X_i \in A_i\}, i \in I$  sono indipendenti.

Poichè la distribuzione congiunta di  $n$  variabili casuali è completamente determinata dalla loro densità congiunta, non è sorprendente che l'indipendenza si possa caratterizzare in termini della densità congiunta.

**Proposizione 3.4.** Siano  $X_1, X_2, \dots, X_n$  variabili casuali definite nello stesso spazio di probabilità  $(\Omega, P)$ , a valori rispettivamente negli insiemi  $E_1, E_2, \dots, E_n$ . Denotiamo con  $p_{X_1, \dots, X_n}$  la loro densità congiunta, e con  $p_{X_i}$  le densità marginali. Allora  $X_1, X_2, \dots, X_n$  sono indipendenti se e solo se

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n p_{X_i}(x_i) \quad (3.9)$$

per ogni  $(x_1, \dots, x_n) \in E_1 \times \dots \times E_n$ .

*Dimostrazione.* Assumiamo che  $X_1, X_2, \dots, X_n$  siano indipendenti. Allora

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = P(X_1 \in \{x_1\}, \dots, X_n \in \{x_n\}) = \prod_{i=1}^n P(X_i \in \{x_i\}) = \prod_{i=1}^n p_{X_i}(x_i).$$

Viceversa, assumiamo che valga (3.9). Per semplificare le espressioni che seguono, assumiamo  $n = 2$ ; l'argomento che usiamo è però valido in generale. Siano  $A_1 \in E_1$ ,  $A_2 \in E_2$  fissati ma arbitrari. Si ha

$$\begin{aligned} P(X_1 \in A_1, X_2 \in A_2) &= P((X_1, X_2) \in A_1 \times A_2) = \sum_{x_1 \in A_1, x_2 \in A_2} p_{X_1, X_2}(x_1, x_2) \\ &= \sum_{x_1 \in A_1, x_2 \in A_2} p_{X_1}(x_1) p_{X_2}(x_2) = \sum_{x_1 \in A_1} p_{X_1}(x_1) \sum_{x_2 \in A_2} p_{X_2}(x_2) = P(X_1 \in A_1) P(X_2 \in A_2). \end{aligned}$$

*Osservazione 3.3.* Siano  $X, Y$  due variabili casuali (ma l'argomento è generalizzabile a  $n$  variabili casuali) la cui densità congiunta si fattorizza nella forma

$$p_{X,Y}(x, y) = \alpha(x)\beta(y).$$

Allora

$$p_X(x) = \sum_y p_{X,Y}(x, y) = b\alpha(x),$$

dove  $b = \sum_y \beta(y)$ . Analogamente

$$p_Y(y) = a\beta(y),$$

con  $a = \sum_x \alpha(x)$ . Inoltre

$$1 = \sum_{x,y} p_{X,Y}(x, y) = ab.$$

Di conseguenza

$$p_{X,Y}(x, y) = \alpha(x)\beta(y) = p_X(x)p_Y(y),$$

e dunque  $X$  e  $Y$  sono indipendenti.

*Esempio 3.4.* Si considerino gli esempi 3.2 e 3.3. Le variabili casuali  $X_A, X_G$  dell'Esempio 3.3 sono indipendenti, mentre quelle dell'Esempio 3.2 non sono indipendenti.

*Osservazione 3.4.* Siano  $X_1, X_2, \dots, X_n$  variabili casuali indipendenti, e  $I = \{i_1, \dots, i_h\}$ ,  $J = \{j_1, \dots, j_k\}$  due sottoinsiemi non vuoti e *disgiunti* di  $\{1, 2, \dots, n\}$ . Denotiamo con  $X_I$  la variabile casuale

$$X_I = (X_{i_1}, \dots, X_{i_h}),$$

e, analogamente,  $X_J$ . Mostriamo che  $X_I, X_J$  sono variabili casuali indipendenti. Scriviamo  $x_I$  in luogo di  $(x_{i_1}, \dots, x_{i_h})$ , e analogamente  $x_J$ . Allora, tenendo conto che le variabili casuali  $\{X_l : l \in I \cup J\}$  sono indipendenti, si ha

$$\begin{aligned} p_{X_I, X_J}(x_I, x_J) &= p_{X_{i_1}, \dots, X_{i_h}, X_{j_1}, \dots, X_{j_k}}(x_{i_1}, \dots, x_{i_h}, x_{j_1}, \dots, x_{j_k}) \\ &= \prod_{r=1}^h p_{X_{i_r}}(x_{i_r}) \prod_{s=1}^k p_{X_{j_s}}(x_{j_s}) = p_{X_I}(x_I) p_{X_J}(x_J), \end{aligned}$$

da cui segue l'indipendenza.

Come vedremo, il risultato nell'Osservazione 3.4 viene usato congiuntamente a quello della seguente proposizione, che stabilisce che l'indipendenza si conserva per trasformazioni. Il risultato che segue è enunciato e dimostrato per due variabili casuali, ma è facilmente estendibile a ogni  $n \geq 2$ .

**Proposizione 3.5.** *Siano  $X, Y$  due variabili casuali definite nello stesso spazio di probabilità  $(\Omega, P)$ , e a valori negli insiemi  $E$  e  $F$  rispettivamente. Siano inoltre  $H, K$  due insiemi, e  $f : E \rightarrow H$ ,  $g : F \rightarrow K$  funzioni arbitrarie. Se le variabili casuali  $X$  e  $Y$  sono indipendenti, allora anche le variabili casuali  $f(X)$  e  $g(Y)$  sono indipendenti.*

*Dimostrazione.* Basta osservare che, se  $A \subseteq H$ ,  $B \subseteq K$ ,

$$\begin{aligned} P(f(X) \in A, g(Y) \in B) &= P(X \in f^{-1}(A), Y \in g^{-1}(B)) \\ &= P(X \in f^{-1}(A)) P(Y \in g^{-1}(B)) = P(f(X) \in A) P(g(Y) \in B). \end{aligned}$$

**Esempio 3.5.** Siano  $X_1, \dots, X_n, X_{n+1}, \dots, X_{n+m}$  variabili casuali scalari indipendenti. Allora  $X_1 + \dots + X_n$  e  $X_{n+1} + \dots + X_{n+m}$  sono indipendenti. Basta applicare la proposizione precedente con  $I = \{1, 2, \dots, n\}$ ,  $J = \{n+1, \dots, n+m\}$ ,  $f(x_1, \dots, x_n) = x_1 + \dots + x_n$ ,  $g(x_{n+1}, \dots, x_{n+m}) = x_{n+1} + \dots + x_{n+m}$ .

### 3.4 Il valor medio

La nozione di *media aritmetica* di un insieme finito di numeri reali  $\{x_1, x_2, \dots, x_n\}$  è nota e molto naturale. Una delle sue possibili interpretazioni è quella che si ottiene associando ad ogni  $x_i$  un punto materiale; tali punti materiali vengono posizionati su una retta, ognuno nel punto corrispondente alla coordinata  $x_i$ . Se tali punti materiali hanno tutti la stessa massa, il punto di coordinate  $\frac{1}{n} \sum_{i=1}^n x_i$  è il baricentro di tale sistema di punti materiali. Nel caso in cui i punti non abbiano tutti la stessa massa, il baricentro si ottiene attraverso una media "pesata": se  $m_i$  è la massa del punto materiale in  $x_i$ , scegliendo l'unità di misura per la massa in modo tale che  $\sum_{i=1}^n m_i = 1$ , il baricentro del sistema ha coordinata

$$\mu = \sum_{i=1}^n m_i x_i.$$

In ambito probabilistico, la nozione di *valor medio* corrisponde alla nozione di bari-centro, una volta interpretate le  $x_i$  come i valori assunti da una variabile casuale, e le  $m_i$  come le rispettive probabilità che i corrispondenti valori vengano assunti. Questo ci conduce alla seguente definizione.

**Definizione 3.6.** Sia  $X$  una variabile casuale a valori in  $\mathbb{R}$  o in  $\mathbb{C}$ , definita in uno spazio di probabilità discreto  $(\Omega, P)$ . Si consideri la somma

$$\sum_{\omega \in \Omega} |X(\omega)| P(\{\omega\}). \quad (3.10)$$

Se tale somma ha un valore finito allora diremo che la variabile casuale  $X$  *ammette valor medio*. In tal caso, la quantità

$$E(X) \equiv \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}) \quad (3.11)$$

si dice *valor medio*, o *media*, o *valor atteso* della variabile casuale  $X$ .

*Osservazione 3.5.* Nel caso in cui  $\Omega$  sia un insieme finito, la somma (3.10) ha, ovviamente, valore finito. Nel caso in cui  $\Omega$  sia numerabile, la condizione di finitezza della serie in (3.10) corrisponde alla sommabilità di (3.11).

*Osservazione 3.6.* Se  $X$  è una variabile casuale che assume solo valori reali positivi, la somma (3.11) è sempre definita, anche se può assumere il valore  $+\infty$ . In questo caso denoteremo con  $E(X)$  il valore della somma, anche quando questo è  $+\infty$ . Con questa convenzione, per ogni variabile casuale a valori reali o complessi, la somma in (3.10) è  $E(|X|)$ . Dunque, talvolta scriveremo “ $E(|X|) < +\infty$ ” in luogo di “la variabile  $X$  ammette valor medio”. Si noti anche che, dalla definizione,  $X$  ammette valor medio se e solo se  $|X|$  ammette valor medio.

*Osservazione 3.7.* Data una costante  $c \in \mathbb{C}$ , essa si può identificare con la variabile casuale  $X_c$  che assume solo il valore  $c$ :  $X_c(\omega) = c \forall \omega$ . È ovvio dalla definizione che  $E(X_c) = c$ . D’ora innanzi indicheremo con  $c$  sia il numero complesso che la variabile casuale  $X_c$ .

Le seguenti proprietà formali del valor medio derivano immediatamente dalla precedente definizione.

**Proposizione 3.6.** Siano  $X, Y$  due variabili casuali discrete a valori in  $\mathbb{C}$  o  $\mathbb{R}$ , definite nello stesso spazio di probabilità  $(\Omega, P)$ . Allora valgono le seguenti proprietà:

- (i) (*Monotonia*) Se  $X, Y$  sono a valori reali, entrambe ammettono valor medio e  $X(\omega) \leq Y(\omega)$  per ogni  $\omega \in \Omega$ , allora  $E(X) \leq E(Y)$ .
- (ii) Se  $X$  ammette valor medio, allora

$$|E(X)| \leq E(|X|).$$

- (iii) (*Linearità*) Se  $X$  e  $Y$  ammettono valor medio e  $a, b \in \mathbb{C}$ , allora la variabile casuale  $aX + bY$  definita da

$$(aX + bY)(\omega) = aX(\omega) + bY(\omega),$$

*ammette valor medio e*

$$E(aX + bY) = aE(X) + bE(Y).$$

Le dimostrazioni sono del tutto elementari, e lasciate al lettore. Tali proprietà sono però assolutamente essenziali sia sul piano teorico che su quello computazionale. La prima conseguenza è il fatto che l'insieme delle variabili casuali che ammettono valor medio ha una struttura algebrica e topologica di grande utilità. Fissato uno spazio di probabilità discreto  $(\Omega, P)$ , sia

$$L^1(\Omega, P) := \{X : \Omega \rightarrow \mathbb{R} \text{ tali che } X \text{ ammetta valor medio}\}$$

(trattiamo qui solo il caso di variabili casuali a valori reali, ma lo stesso si può fare per variabili casuali a valori complessi). La Proposizione 3.6 (iii) garantisce che  $L^1(\Omega, P)$  sia uno spazio vettoriale su  $\mathbb{R}$  (dove lo “zero” dello spazio è la funzione costantemente uguale a 0). Per  $X \in L^1(\Omega, P)$ , definiamo

$$\|X\|_1 := E(|X|). \quad (3.12)$$

**Teorema 3.1.** *Supponiamo che lo spazio di probabilità discreto  $(\Omega, P)$  sia tale che*

$$P(\{\omega\}) > 0 \text{ per ogni } \omega \in \Omega. \quad (3.13)$$

*Allora la funzione  $\|\cdot\|_1 : L^1(\Omega, P) \rightarrow \mathbb{R}^+$  è una norma. Inoltre lo spazio normato  $(L^1(\Omega, P), \|\cdot\|_1)$  è completo, cioè è uno spazio di Banach.*

*Dimostrazione.* Ricordiamo che dimostrare che  $\|\cdot\|_1$  è una norma, significa mostrare:

$$\|X\|_1 = 0 \iff X \equiv 0; \quad (3.14)$$

$$X \in L^1(\Omega, P), \lambda \in \mathbb{R} \Rightarrow \|\lambda X\|_1 = |\lambda| \|X\|_1; \quad (3.15)$$

$$X, Y \in L^1(\Omega, P) \Rightarrow \|X + Y\|_1 \leq \|X\|_1 + \|Y\|_1. \quad (3.16)$$

Le proprietà (3.15) e (3.16) (quest'ultima è detta *disuguaglianza triangolare*) sono conseguenze abbastanza semplici della definizione di valor medio. La proprietà (3.14) deriva invece facilmente dall'ipotesi (3.13). Le dimostrazioni sono lasciate per esercizio.

Concentriamoci invece sulla questione della completezza. Ricordiamo che una successione  $(X_n)_{n \geq 1}$  di elementi di  $L^1(\Omega, P)$  si dice *successione di Cauchy* se

$$\lim_{n \rightarrow +\infty} \sup_{l, m \geq n} \|X_m - X_l\|_1 = 0. \quad (3.17)$$

Mostrare che uno spazio normato è completo significa dimostrare che ogni successione di Cauchy ammette un limite nella convergenza *indotta* dalla norma, cioè

esiste (unico)  $X \in L^1(P)$  tale che

$$\lim_{n \rightarrow +\infty} \|X_n - X\|_1 = 0 \quad (3.18)$$

(l'unicità del limite è una semplice conseguenza di (3.14)). Fissato  $\omega \in \Omega$  è chiaro che

$$\|X_m - X_l\|_1 = \sum_{\omega' \in \Omega} |X_m(\omega') - X_l(\omega')| P(\{\omega'\}) \geq |X_m(\omega) - X_l(\omega)| P(\{\omega\}). \quad (3.19)$$

Usando (3.19), (3.17) e il fatto che  $P(\{\omega\}) > 0$ , ne segue che la successione di numeri reali  $(X_n(\omega))_{n \geq 1}$  è una successione di Cauchy, e quindi converge, essendo  $\mathbb{R}$  uno spazio completo. È dunque lecito definire, per ogni  $\omega \in \Omega$ ,

$$X(\omega) := \lim_{n \rightarrow +\infty} X_n(\omega).$$

Dimostriamo ora che  $X \in L^1(\Omega, P)$ , e che vale (3.18). Cominciamo con ricordare che una semplice conseguenza della disuguaglianza triangolare è il fatto che

$$|\|X_m\|_1 - \|X_l\|_1| \leq \|X_m - X_l\|_1,$$

da cui segue che la successione di numeri reali  $(\|X_n\|_1)_{n \geq 1}$  è di Cauchy. Perciò converge e, in particolare, è limitata. Quindi

$$M := \sup_{n \geq 1} \|X_n\|_1 < +\infty.$$

Sia ora  $A$  un sottoinsieme finito di  $\Omega$ . Poichè il limite conserva le somme finite:

$$\sum_{\omega \in A} |X(\omega)| P(\{\omega\}) = \lim_{n \rightarrow +\infty} \sum_{\omega \in A} |X_n(\omega)| P(\{\omega\}) \leq \lim_{n \rightarrow +\infty} \|X_n\|_1 \leq M.$$

Ricordando la definizione di somma infinita:

$$\sum_{\omega \in \Omega} |X(\omega)| P(\{\omega\}) = \sup_{A \subseteq \Omega: |A| < +\infty} \sum_{\omega \in A} |X(\omega)| P(\{\omega\}) \leq M,$$

segue che  $X \in L^1(\Omega, P)$ . Resta da dimostrare (3.18). Usando nuovamente il fatto che il limite conserva le somme finite, per  $A \subseteq \Omega$  finito si ha

$$\begin{aligned} \sum_{\omega \in A} |X(\omega) - X_n(\omega)| P(\{\omega\}) &= \lim_{m \rightarrow +\infty} \sum_{\omega \in A} |X_m(\omega) - X_n(\omega)| P(\{\omega\}) \\ &\leq \limsup_{m \rightarrow +\infty} \|X_m - X_n\|_1 \leq \sup_{l, m \geq n} \|X_m - X_l\|_1. \end{aligned}$$

Prendendo l'estremo superiore sui sottoinsiemi finiti di  $\Omega$  nella precedente, si trova

$$\|X - X_n\|_1 \leq \sup_{l, m \geq n} \|X_m - X_l\|_1$$

che, assieme a (3.17), implica (3.18).

La definizione di valor medio che abbiamo data risulta poco operativa, in quanto in genere di una variabile casuale è nota solo la distribuzione. Per questo motivo è utile esprimere il valor medio di una variabile casuale in termini della sua densità. Nella seguente proposizione faremo, in realtà, qualcosa in più, ovvero esprimeremo in termini della densità di  $X$  la media di ogni “funzione” della variabile  $X$ . Useremo, di qui in avanti, la seguente notazione. Sia  $X$  una variabile casuale discreta a valori in un insieme generico  $E$ , e sia  $f : E \rightarrow K$ , con  $K = \mathbb{R}$  o  $\mathbb{C}$ . Allora  $f \circ X$  è una variabile casuale a valori in  $K$ , che denoteremo con  $f(X)$ .

**Proposizione 3.7.** *Sia  $X$  una variabile casuale discreta a valori in un insieme generico  $E$ , sia  $p_X$  la sua densità e sia  $f : E \rightarrow K$ , con  $K = \mathbb{R}$  o  $\mathbb{C}$ .  $f(X)$  ammette valor medio se e solo se*

$$\sum_{x \in E} |f(x)| p_X(x) < +\infty, \quad (3.20)$$

*In questo caso*

$$E(f(X)) = \sum_{x \in E} f(x) p_X(x). \quad (3.21)$$

*In particolare, se  $E = K$ ,  $X$  ammette valor medio se e solo se*

$$\sum_{x \in K} |x| p_X(x) < +\infty,$$

*e in questo caso*

$$E(X) = \sum_{x \in K} x p_X(x).$$

*Dimostrazione.* Dalla definizione di valor medio, abbiamo che  $f(X)$  ammette valor medio se e solo se

$$\sum_{\omega \in \Omega} |f(X(\omega))| P(\{\omega\}) < +\infty. \quad (3.22)$$

Cominciamo col mostrare che la somma in (3.20) coincide con quella in (3.22). Si noti anzitutto, che la famiglia di eventi  $\{\omega \in \Omega : X(\omega) = x\}$ , al variare di  $x \in X(\Omega)$ , costituisce una partizione di  $\Omega$ . Ma allora

$$\begin{aligned} \sum_{\omega \in \Omega} |f(X(\omega))| P(\{\omega\}) &= \sum_{x \in X(\Omega)} \sum_{\omega : X(\omega) = x} |f(X(\omega))| P(\{\omega\}) \\ &= \sum_{x \in X(\Omega)} |f(x)| P(X = x) \\ &= \sum_{x \in E} |f(x)| p_X(x). \end{aligned}$$

Ciò mostra che (3.20) equivale al fatto che  $f(X)$  ammetta valor medio. Per concludere la dimostrazione, occorre mostrare che, se  $f(X)$  ammette valor medio, allora la somma in (3.21) coincide con

$$E(f(X)) = \sum_{\omega \in \Omega} f(X(\omega)) P(\{\omega\}).$$

Ma per questo basta ripetere l'argomento appena usato per mostrare che la somma in (3.20) coincide con quella in (3.22).

### 3.5 Spazi $L^p$ , momenti, varianza

Sia  $p \geq 1$  un numero reale, e  $(\Omega, P)$  uno spazio di probabilità discreto. Definiamo

$$L^p(\Omega, P) := \{X : \Omega \rightarrow \mathbb{R} \text{ tali che } E(|X|^p) < +\infty\}.$$

È chiaro che per  $p = 1$  ritroviamo la definizione data nel paragrafo precedente. Un primo, semplice, risultato è il seguente.

**Proposizione 3.8.** *L'insieme  $L^p(\Omega, P)$  è uno spazio vettoriale su  $\mathbb{R}$ .*

*Dimostrazione.* Il fatto che  $X \in L^p(\Omega, P)$ ,  $\lambda \in \mathbb{R}$  implichi  $\lambda X \in L^p(\Omega, P)$  è immediato, e lo lasciamo verificare al lettore. Resta da dimostrare che se  $X, Y \in L^p(\Omega, P)$  allora  $X + Y \in L^p(\Omega, P)$ , cioè

$$E(|X + Y|^p) < +\infty.$$

Essendo  $E(|X + Y|^p) \leq E[(|X| + |Y|)^p]$ , basta dimostrare la finitezza di quest'ultimo. Consideriamo la funzione  $\varphi : [0, +\infty) \rightarrow \mathbb{R}$  data da  $\varphi(x) = x^p$ . Poiché  $p \geq 1$ , la funzione  $\varphi$  è convessa. Pertanto, se  $x, y \geq 0$

$$\left(\frac{x+y}{2}\right)^p = \varphi\left(\frac{x+y}{2}\right) \leq \frac{\varphi(x) + \varphi(y)}{2} = \frac{x^p + y^p}{2},$$

o, equivalentemente,

$$(x+y)^p \leq 2^{p-1}(x^p + y^p).$$

Usando quest'ultima disuguaglianza:

$$E[(|X| + |Y|)^p] \leq 2^{p-1}[E(|X|^p) + E(|Y|^p)] < +\infty.$$

Per  $X \in L^p(\Omega, P)$ , definiamo

$$\|X\|_p := [E(|X|^p)]^{1/p}. \quad (3.23)$$

Il teorema seguente generalizza il Teorema 3.1. Per  $p > 1$  la dimostrazione è un po' più delicata, e la omettiamo.

**Teorema 3.2.** *Supponiamo che lo spazio di probabilità discreto  $(\Omega, P)$  sia tale che*

$$P(\{\omega\}) > 0 \text{ per ogni } \omega \in \Omega. \quad (3.24)$$

*Allora la funzione  $\|\cdot\|_p : L^p(\Omega, P) \rightarrow \mathbb{R}^+$  è una norma. Inoltre lo spazio normato  $(L^p(\Omega, P), \|\cdot\|_p)$  è completo, cioè è uno spazio di Banach.*

**Proposizione 3.9.** *Se  $1 \leq p < q$ , allora  $L^q(\Omega, P) \subseteq L^p(\Omega, P)$ .*

*Dimostrazione.* Si osservi che, per ogni  $x \in \mathbb{R}$ , vale la disuguaglianza

$$|x|^p \leq 1 + |x|^q$$

(per verificarlo è sufficiente osservare che  $|x|^p \leq 1$  se  $|x| \leq 1$  e  $|x|^p \leq |x|^q$  se  $|x| > 1$ ). Pertanto, se  $X \in L^q(\Omega, P)$ ,

$$E(|X|^p) \leq 1 + E(|X|^q) < +\infty,$$

da cui segue la tesi.

*Osservazione 3.8.* Dimostreremo nel prossimo paragrafo una versione più forte della proposizione precedente, cioè che se  $X \in L^q(\Omega, P)$  allora

$$\|X\|_p \leq \|X\|_q.$$

Quest'ultima disuguaglianza implica, in particolare, che l'inclusione  $L^q(\Omega, P) \rightarrow L^p(\Omega, P)$  è continua rispetto alle corrispondenti norme.

**Definizione 3.7.** Sia  $X \in L^k(\Omega, P)$ , dove  $k \geq 1$  è un intero. Diremo allora che  $X$  ammette *momento di ordine  $k$*  e la quantità  $E(X^k)$  si dice *momento di ordine  $k$* .

Poiché le costanti ammettono momento di ogni ordine,  $X$  ammette momento di ordine  $k$  se e solo se  $X - c$  ammette momento di ordine  $k$  per ogni  $c \in \mathbb{R}$ . Se  $X$  ammette momento di ordine  $k \geq 1$ , per la Proposizione 3.9,  $X$  ammette valor medio. Allora ha senso considerare il momento di ordine  $k$  di  $X - E(X)$ , cioè  $E[(X - E(X))^k]$ , che viene detto *momento centrato di ordine  $k$* . Si vede subito che il momento centrato di ordine 1 vale zero. Il momento centrato di ordine 2 si dice *varianza*, ed è denotato con  $Var(\cdot)$ :

$$Var(X) = E[(X - E(X))^2].$$

Si osservi che la Varianza non è un operatore lineare. Infatti, per  $a, b \in \mathbb{R}$

$$Var(aX) = a^2 Var(X),$$

e

$$Var(X + b) = Var(X).$$

La verifica di tali identità è semplice, ed è lasciata al lettore.

Notiamo ora che  $Var(X) = \|X - E(X)\|_2^2$ , ossia la Varianza è il quadrato della distanza in  $L^2(\Omega, P)$  di  $X$  dalla costante  $E(X)$ : si può dunque interpretare come una misura della “dispersione” dei valori della variabile  $X$  attorno alla sua media. In particolare valgono i seguenti risultati.

**Proposizione 3.10.** *Sia  $X \in L^2(\Omega, P)$ .*

- (i)  *$Var(X) = 0$  se e solo se  $X$  è quasi certamente costante, cioè se esiste una costante  $c \in \mathbb{R}$  tale che  $P(X = c) = 1$ .*

(ii) Per ogni  $c \in \mathbb{R}$ ,  $\text{Var}(X) \leq \|X - c\|_2^2$ .

*Dimostrazione.* (i) La dimostrazione è semplice, ed è basata sull'osservazione che se  $Y \in L^1(\Omega, P)$  e  $Y \geq 0$ , allora  $E(Y) = 0$  se e solo se  $P(Y = 0) = 1$ . I dettagli sono lasciati come esercizio.

(ii) Si consideri il polinomio in  $c$

$$p(c) = \|X - c\|_2^2 = E[(X - c)^2] = c^2 - 2E(X)c + E(X^2).$$

Si vede subito che  $p(c)$  ha minimo assoluto per  $c = E(X)$ , e  $p(E(X)) = \text{Var}(X)$ .

Nella precedente proposizione, la (i) afferma che le variabili casuali con varianza zero sono costanti a meno di insiemi di probabilità zero, mentre la (ii) fornisce una caratterizzazione “variazionale” del valor medio, affermando che esso è la costante che realizza la distanza minima da  $X$  nel senso di  $L^2$ . Va notato che, se assumiamo l'ipotesi (3.24), allora la Proposizione 3.10 (i) si potrebbe enunciare come segue:  $\text{Var}(X) = 0$  se e solo se  $X$  è costante. L'ipotesi (3.24), sostanzialmente non restrittiva per spazi di probabilità discreti, è invece non ragionevole, come vedremo, in spazi di probabilità generali. Scegliamo pertanto di non fare tale assunzione, in modo da avere enunciati che resteranno validi nell'ambito della teoria generale.

### 3.6 Funzione generatrice dei momenti

**Definizione 3.8.** Sia  $X$  una variabile casuale scalare. La funzione  $\gamma_X : \mathbb{R} \rightarrow (0, +\infty]$  definita da

$$\gamma_X(t) = E(e^{tX})$$

è detta *funzione generatrice dei momenti* della variabile casuale  $X$ .

*Osservazione 3.9.* Notare che la funzione generatrice dei momenti di una variabile casuale può assumere il valore  $+\infty$  per qualche valore di  $t$ , nel caso in cui la variabile casuale  $e^{tX}$  non ammetta valor medio. Si osservi, inoltre, che necessariamente  $\gamma_X(0) = 1$

La funzione generatrice dei momenti gioca un ruolo importante nel calcolo delle probabilità, tuttavia in applicazioni che vanno al di là del contenuto di questo libro. Essa è utile, in alcuni casi, anche per il calcolo dei momenti di una variabile casuale, ove sia possibile far uso del seguente risultato.

**Teorema 3.3.** Sia  $X$  una variabile casuale scalare e  $\gamma_X$  la sua funzione generatrice. Supponiamo esista  $a > 0$  tale che  $\gamma_X(t) < +\infty$  per ogni  $t \in (-a, a)$ . Allora

- (i) La variabile casuale  $X$  ammette momenti di ogni ordine.
- (ii)  $\gamma_X$  è infinitamente derivabile in  $(-a, a)$ , e

$$\gamma_X^{(n)}(0) = E(X^n).$$

(iii) Esiste  $\varepsilon > 0$  tale che, per  $t \in (-\varepsilon, \varepsilon)$ ,  $\gamma_X(t)$  è dato dalla serie di Taylor

$$\gamma_X(t) = \sum_{n=0}^{+\infty} E(X^n) \frac{t^n}{n!}. \quad (3.25)$$

*Dimostrazione.* (i) Cominciamo con l'osservare che anche  $\gamma_{|X|}(t) < +\infty$  per  $t \in (-a, a)$ . Per  $t < 0$  la cosa è ovvia, visto che

$$e^{t|X|} \leq 1 \Rightarrow \gamma_{|X|}(t) \leq 1.$$

Per  $0 < t < a$ , poiché per ogni  $x \in \mathbb{R}$  si ha  $e^{|x|} \leq e^x + e^{-x}$ , abbiamo che

$$\gamma_{|X|}(t) = E(e^{t|X|}) \leq E(e^{tX}) + E(e^{-tX}) = \gamma_X(t) + \gamma_X(-t).$$

Ora si noti che, per  $x \in \mathbb{R}$ ,  $n \geq 0$

$$e^{|x|} = \sum_{k=0}^{+\infty} \frac{|x|^k}{k!} \geq \frac{|x|^n}{n!}. \quad (3.26)$$

Sia ora  $0 < b < a$ . Usando (3.26) abbiamo

$$|X(\omega)|^n \leq \frac{n!}{b^n} e^{b|X(\omega)|}.$$

Perciò

$$E(|X|^n) \leq \frac{n!}{b^n} \gamma_{|X|}(b) < +\infty,$$

e quindi  $X$  ammette momenti di ogni ordine.

(ii) Osserviamo anzitutto che, per ogni  $t \in (-a, a)$  e  $n \geq 0$ , la variabile casuale  $|X|^n e^{t|X|}$  ammette valor medio (e perciò lo ammette anche  $X^n e^{tX}$ , essendo  $|X|^n e^{tX} \leq |X|^n e^{t|X|}$ ). Scegliamo  $h > 0$  tale che  $|t| + h < a$ . Usando (3.26) come al punto precedente, abbiamo che

$$|X|^n \leq \frac{n!}{h^n} e^{h|X|}.$$

Perciò,

$$|X^n| e^{tX} \leq \frac{n!}{h^n} e^{(|t|+h)|X|}. \quad (3.27)$$

La variabile casuale al membro destro della precedente uguaglianza ammette valor medio, per quanto mostrato all'inizio del punto precedente. Pertanto anche  $|X|^n e^{t|X|}$  ammette valor medio.

A questo punto, mostriamo per induzione su  $n \geq 0$  che

$$\gamma_X^{(n)}(t) = E(X^n e^{tX}) \quad (3.28)$$

per  $t \in (-a, a)$ , dove si noti che, per quanto appena dimostrato, il valor medio in (3.28) esiste. Per  $n = 0$  non c'è nulla da dimostrare. Per il passo induttivo, si noti che, se  $t \pm h \in (-a, a)$ , usando l'ipotesi induttiva (3.28)

$$\frac{\gamma_X^{(n)}(t+h) - \gamma_X^{(n)}(t)}{h} = E \left( X^n e^{tX} \left[ \frac{e^{hX} - 1}{h} \right] \right).$$

Pertanto

$$\left| \frac{\gamma_X^{(n)}(t+h) - \gamma_X^{(n)}(t)}{h} - E(X^{n+1} e^{tX}) \right| \leq E \left( |X|^n e^{tX} \left| \frac{e^{hX} - 1}{h} - X \right| \right). \quad (3.29)$$

A questo punto usiamo il fatto che, per ogni  $x \in \mathbb{R}$ ,

$$|e^x - 1 - x| \leq \frac{x^2}{2} e^{|x|} \quad (3.30)$$

(verificarlo!). Ma allora, usando (3.30) in (3.29) si ha

$$\begin{aligned} \left| \frac{\gamma_X^{(n)}(t+h) - \gamma_X^{(n)}(t)}{h} - E(X^{n+1} e^{tX}) \right| &\leq h E \left( |X|^{n+2} e^{(|t|+|h|)|X|} \right) \\ &\leq h E \left( |X|^{n+2} e^{b|X|} \right), \end{aligned} \quad (3.31)$$

dove  $b$  è scelto in modo tale che  $|t| + |h| < b < a$ . Per quanto visto sopra,  $E(|X|^{n+2} e^{b|X|}) < +\infty$ , e quindi

$$\lim_{h \rightarrow 0} \left| \frac{\gamma_X^{(n)}(t+h) - \gamma_X^{(n)}(t)}{h} - E(X^{n+1} e^{tX}) \right| = 0,$$

che conclude la dimostrazione del passo induttivo.

- (iii) Usiamo una generalizzazione della disuguaglianza in (3.30), che consiste nello stimare il resto dell'espansione di Taylor per la funzione  $e^x$ :

$$\left| e^x - \sum_{k=0}^{n-1} \frac{x^k}{k!} \right| \leq \frac{|x|^n}{n!} e^{|x|}. \quad (3.32)$$

Poniamo ora  $\varepsilon = \frac{a}{2}$ . Per  $t \in (-\varepsilon, \varepsilon)$ , usando (3.32), abbiamo che

$$\left| e^{tX} - \sum_{k=0}^{n-1} X^k \frac{t^k}{k!} \right| \leq \frac{|t|^n}{n!} |X|^n e^{|tX|}. \quad (3.33)$$

Perciò, usando la disuguaglianza  $|E(Y)| \leq E(|Y|)$ , otteniamo

$$\left| E(e^{tX}) - \sum_{k=0}^{n-1} E(X^k) \frac{t^k}{k!} \right| \leq E \left( \left| e^{tX} - \sum_{k=0}^{n-1} X^k \frac{t^k}{k!} \right| \right) \leq \frac{|t|^n}{n!} E(|X|^n e^{|tX|}). \quad (3.34)$$

A questo punto, usando (3.27) con  $h = a/2$  abbiamo

$$\left| E(e^{tX}) - \sum_{k=0}^{n-1} E(X^k) \frac{t^k}{k!} \right| \leq \frac{|t|^n}{n!} \frac{n!}{(a/2)^n} E(e^{(|t|+\frac{a}{2})|X|}) = \left( \frac{t}{a/2} \right)^n E(e^{(|t|+\frac{a}{2})|X|}). \quad (3.35)$$

Poiché quest'ultimo valor medio è finito e  $\left(\frac{t}{a/2}\right)^n$  tende a zero per  $n \rightarrow +\infty$ , si conclude che

$$\lim_{n \rightarrow +\infty} \left| E(e^{tX}) - \sum_{k=0}^{n-1} E(X^k) \frac{t^k}{k!} \right| = 0,$$

che implica

$$E(e^{tX}) = \lim_{n \rightarrow +\infty} \sum_{k=0}^{n-1} E(X^k) \frac{t^k}{k!} = \sum_{n=0}^{+\infty} E(X^n) \frac{t^n}{n!},$$

e la dimostrazione è conclusa.

*Osservazione 3.10.* Con un po' di fatica in più su può dimostrare che la serie in (3.25) converge a  $\gamma_X(t)$  per ogni  $t \in (-a, a)$ .

*Osservazione 3.11.* Nella dimostrazione del Teorema 3.3 non abbiamo mai usato la seguente espressione per la funzione generatrice dei momenti, che è quella che più useremo nei calcoli espliciti:

$$\gamma_X(t) = \sum_{x \in \mathbb{R}} e^{tx} p_X(x). \quad (3.36)$$

Nel caso in cui la variabile casuale  $X$  assuma un numero finito di valori, cioè  $|X(\Omega)| < +\infty$ , la somma in (3.36) è una somma finita. In questo caso, tutte le affermazioni del Teorema 3.3 si dimostrano facilmente da (3.36). Se invece  $|X(\Omega)|$  è numerabile, non c'è alcun vantaggio ad usare (3.36) per dimostrare il Teorema 3.3. La dimostrazione qui data ha il vantaggio di rimanere inalterata nel contesto più generale, che vedremo più avanti, di variabili casuali definite in spazi di probabilità non discreti.

Più avanti in questo capitolo vedremo numerosi esempi di calcolo della funzione generatrice.

### 3.7 Disuguaglianze

Come in molti altri settori della matematica, in probabilità le disuguaglianze giocano un ruolo fondamentale. La prima che vediamo chiarisce il significato della varianza

come indice della dispersione dei valori di una variabile casuale: la probabilità di una deviazione dalla media di una variabile casuale  $X$  si può stimare dall'alto con la sua varianza.

**Proposizione 3.11.** (*Disuguaglianza di Chebyshev*)

- (i) Sia  $X$  una variabile casuale a valori reali positivi, che ammette valor medio. Allora, per ogni  $\varepsilon > 0$

$$P(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon}.$$

- (ii) Sia  $X \in L^2(\Omega, P)$ . Allora, per ogni  $\varepsilon > 0$

$$P(|X - E(X)| > \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}.$$

*Dimostrazione.* (i) Tenendo conto che  $X(\omega) \geq 0$  per ogni  $\omega$ , si ha la disuguaglianza

$$X \geq \varepsilon 1_{X \geq \varepsilon}.$$

Per monotonia del valor medio:

$$E(X) \geq \varepsilon E(1_{X \geq \varepsilon}) = \varepsilon P(X \geq \varepsilon),$$

da cui la tesi segue subito.

- (ii) Poichè

$$P(|X - E(X)| > \varepsilon) = P((X - E(X))^2 > \varepsilon^2),$$

è sufficiente applicare quanto dimostrato in (i) a  $(X - E(X))^2$  e  $\varepsilon^2$ .

La seguente disuguaglianza è utile quando si voglia confrontare il valor medio di una variabile casuale con quelli di sue opportune funzioni.

**Proposizione 3.12.** (*Disuguaglianza di Jensen*) Sia  $X$  una variabile casuale scalare e  $\varphi: \mathbb{R} \rightarrow \mathbb{R}$  una funzione convessa. Si assuma che le variabili casuali  $X$  e  $\varphi(X)$  ammettano valor medio. Allora

$$\varphi(E(X)) \leq E(\varphi(X)).$$

*Dimostrazione.* Il fatto che  $\varphi$  sia convessa è equivalente ad affermare che, per ogni  $x_0 \in \mathbb{R}$ , esiste  $\lambda(x_0) \in \mathbb{R}$  tale che per ogni  $x \in \mathbb{R}$

$$\varphi(x) \geq \varphi(x_0) + \lambda(x_0)(x - x_0) \quad (3.37)$$

(il lettore verifichi quest'ultima affermazione). Posto  $x = X(\omega)$  e  $x_0 = E(X)$  in (3.37) otteniamo, per tutti gli  $\omega \in \Omega$

$$\varphi(X(\omega)) \geq \varphi(E(X)) + \lambda(E(X))(X - E(X)). \quad (3.38)$$

Prendendo il valor medio dei due membri di (3.38) la conclusione segue subito.

Un esempio di applicazione della disuguaglianza di Jensen è il seguente confronto tra norme  $L^p$  di una variabile casuale.

**Corollario 3.1.** *Sia  $X \in L^q(\Omega, P)$  e sia  $1 \leq p \leq q$ . Allora*

$$\|X\|_p \leq \|X\|_q.$$

*Dimostrazione.* È sufficiente applicare la disuguaglianza di Jensen alla variabile casuale  $|X|$  e alla funzione convessa  $\varphi(x) = |x|^{q/p}$ .

L'ultima disuguaglianza di questo paragrafo è utile ogni qual volta si voglia stimare la media di un prodotto di variabili casuali. Un'applicazione importante la vedremo nel prossimo paragrafo.

**Proposizione 3.13.** *(Disuguaglianza di Cauchy-Schwartz)*

*Siano  $X, Y \in L^2(\Omega, P)$ . Allora  $XY$  ammette valor medio, e*

$$|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}. \quad (3.39)$$

*L'uguaglianza in (3.39) vale se e solo se esiste  $c \in \mathbb{R}$  tale che  $P(X = cY) = 1$ .*

*Dimostrazione.* Per mostrare che  $XY$  ammette valor medio, si osservi che per ogni  $x, y \in \mathbb{R}$

$$|xy| \leq \frac{1}{2}x^2 + \frac{1}{2}y^2,$$

come si vede dal fatto che

$$\frac{1}{2}x^2 + \frac{1}{2}y^2 - |xy| = \frac{1}{2}(|x| - |y|)^2 \geq 0.$$

Ma allora

$$|XY| \leq \frac{1}{2}X^2 + \frac{1}{2}Y^2,$$

da cui

$$E(|XY|) \leq \frac{1}{2}E(X^2) + \frac{1}{2}E(Y^2) < +\infty.$$

Per mostrare la disuguaglianza (3.39) è necessario un argomento più fine. La disuguaglianza (3.39) fa parte della classe di disuguaglianze di Cauchy-Schwartz, che si dimostrano tutte in modo analogo.

Anzitutto, non è restrittivo supporre  $P(X = 0) < 1$ , e  $P(Y = 0) < 1$ . Equivalentemente  $E(X^2) > 0$ ,  $E(Y^2) > 0$ . In caso contrario la (3.39) è banalmente verificata ( $0 = 0!$ ). Inoltre possiamo assumere  $E(XY) \geq 0$ . In caso contrario è sufficiente rimpiazzare  $X$  con  $-X$ , e notare che l'intero enunciato non viene modificato da tale rimpiazzamento. Poniamo allora

$$X_* = \frac{X}{\sqrt{E(X^2)}}, \quad Y_* = \frac{Y}{\sqrt{E(Y^2)}}.$$

Si noti che

$$\begin{aligned} 0 &\leq E[(X_* - Y_*)^2] = E(X_*^2) + E(Y_*^2) - 2E(X_*Y_*) \\ &= 2 - 2 \frac{E(XY)}{\sqrt{E(X^2)E(Y^2)}}, \end{aligned}$$

da cui (3.39) segue.

Supponiamo ora che (3.39) valga come uguaglianza. Allora, per quanto appena visto,  $E[(X_* - Y_*)^2] = 0$  che, come osservato nella dimostrazione della Proposizione 3.10, equivale a

$$P(X_* = Y_*) = 1 \iff P\left(X = \frac{\sqrt{E(X^2)}}{\sqrt{E(Y^2)}}Y\right) = 1,$$

cioè  $P(X = cY) = 1$  con  $c = \frac{\sqrt{E(X^2)}}{\sqrt{E(Y^2)}}$ .

Viceversa, se  $P(X = cY) = 1$ , allora (perché?)  $E(XY) = cE(Y^2)$ ,  $E(X^2) = c^2E(Y^2)$ , da cui si verifica che la (3.39) vale come uguaglianza.

### 3.8 Covarianza e coefficiente di correlazione

Media, varianza e momenti sono quantità che si riferiscono ad una singola variabile casuale. L'oggetto che ora definiamo è un indice della relazione che sussiste tra due variabili casuali.

**Definizione 3.9.** Siano  $X, Y$  due variabili casuali scalari definite sullo stesso spazio di probabilità  $(\Omega, P)$ , che ammettono valor medio. Se la variabile casuale  $(X - E(X))(Y - E(Y))$  ammette media, la quantità

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

si dice *covarianza* tra  $X$  e  $Y$ .

Si noti che ogni qual volta  $X, Y$  e  $XY$  ammettono valor medio, allora la covarianza è ben definita, e

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

**Proposizione 3.14.** Siano  $X, Y \in L^2(\Omega, P)$ . Allora  $\text{Cov}(X, Y)$  è ben definita. Inoltre

$$|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X)\text{Var}(Y)}. \quad (3.40)$$

Infine, assumendo  $\text{Var}(X) > 0$ ,  $\text{Var}(Y) > 0$ ,  $\text{Cov}(X, Y) = \sqrt{\text{Var}(X)\text{Var}(Y)}$  (risp.  $\text{Cov}(X, Y) = -\sqrt{\text{Var}(X)\text{Var}(Y)}$ ) se e solo se esistono costanti  $a > 0$  (risp.  $a < 0$ ) e  $b \in \mathbb{R}$  tali che

$$P(X = aY + b) = 1. \quad (3.41)$$

*Dimostrazione.* Si applica la Proposizione 3.13 alle variabili casuali  $\bar{X} = X - E(X)$  e  $\bar{Y} = Y - E(Y)$ . Si ottiene immediatamente l'esistenza della covarianza, e la disuguaglianza (3.40) segue da (3.39). Inoltre, se (3.40) vale come uguaglianza, allora (3.39) vale come uguaglianza per  $\bar{X}$  e  $\bar{Y}$ , e quindi esiste  $c \in \mathbb{R}$  per cui  $P(\bar{X} = c\bar{Y}) = 1$ . Notare che dev'essere  $c \neq 0$ , essendo  $E(\bar{X}^2) = \text{Var}(X) > 0$ . Ma allora

$$P(X = cY + E(X) - cE(Y)) = 1,$$

cioè (3.41) vale con  $a = c$  e  $b = E(X) - cE(Y)$ . Infine, un calcolo diretto simile a quello fatto nella Proposizione 3.13, mostra che se vale (3.41) con  $a \neq 0$  allora

$$\text{Cov}(X, Y) = \frac{a}{|a|} \sqrt{\text{Var}(X)\text{Var}Y},$$

e la dimostrazione è conclusa.

**Definizione 3.10.** Siano  $X, Y \in L^2(\Omega, P)$ , tali che  $\text{Var}(X) > 0$ ,  $\text{Var}(Y) > 0$ . La quantità

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}Y}}$$

Si dice *coefficiente di correlazione* tra  $X$  e  $Y$ . Se  $\rho_{X,Y} = 0$  (equivalentemente  $\text{Cov}(X, Y) = 0$ ) diremo che le variabili casuali  $X$  e  $Y$  sono *scorrelate*

Le seguenti proprietà del coefficiente di correlazione discendono immediatamente dalla Proposizione 3.14

**Proposizione 3.15.** Siano  $X, Y \in L^2(\Omega, P)$ , tali che  $\text{Var}(X) > 0$ ,  $\text{Var}(Y) > 0$ . Allora  $|\rho(X, Y)| \leq 1$ . Inoltre  $\rho(X, Y) = 1$  (risp  $\rho(X, Y) = -1$ ) se e solo se esistono  $a > 0$  (risp.  $a < 0$ ) e  $b \in \mathbb{R}$  tali che

$$P(X = aY + b) = 1.$$

Il coefficiente di correlazione è un'indice del grado di correlazione *lineare* tra le variabili  $X$  e  $Y$ , ossia di quanto “bene”  $X$  possa essere approssimata da funzioni affini di  $Y$ . Tale affermazione è di ovvio significato quando  $|\rho(X, Y)| = 1$ , vista la Proposizione 3.15. Per capire il caso generale, formuliamo il seguente problema. Siano  $X, Y \in L^2(\Omega, P)$ , con  $\text{Var}(X) > 0$ ,  $\text{Var}(Y) > 0$ , due variabili casuali fissate, e cerchiamo di determinare la funzione affine di  $Y$  che meglio approssima  $X$  del senso della distanza  $\|\cdot - \cdot\|_2$ . In altre parole, cerchiamo le costanti  $a, b \in \mathbb{R}$  tali che la distanza

$$\|X - (aY + b)\|_2$$

sia la minima possibile. Questo equivale a minimizzare la funzione, nelle variabili  $a, b$

$$\varphi(a, b) = \|X - (aY + b)\|_2^2 = E[(X - aY - b)^2] = E(Y^2 a^2 + b^2 + 2E(Y)ab - 2E(XY)a - 2E(X)b + E(X^2)).$$

Con le tecniche standard del calcolo differenziale in più variabili, si vede che  $\varphi(a, b)$  ammette minimo assoluto in corrispondenza delle soluzioni del sistema

$$\begin{aligned}\frac{\partial}{\partial a}\varphi(a,b) &= 0 \\ \frac{\partial}{\partial b}\varphi(a,b) &= 0\end{aligned}$$

la cui unica soluzione è

$$\begin{aligned}a^* &= \frac{\text{Cov}(X,Y)}{\text{Var}(Y)} \\ b^* &= E(X) - \frac{E(Y)\text{Cov}(X,Y)}{\text{Var}(Y)}.\end{aligned}$$

Il valore del minimo assoluto è

$$\varphi(a^*, b^*) = \text{Var}(X)[1 - \rho^2(X,Y)].$$

Dunque  $X$  è tanto meglio approssimabile da funzioni affini di  $Y$  quanto più vicino ad uno è  $\rho^2(X,Y)$ . Se, viceversa, le variabili  $X$  e  $Y$  sono scorrelate, allora  $a^* = 0$ , cioè la migliore approssimazione di  $X$  con funzioni affini di  $Y$  *non* dipende da  $Y$ , ossia è una costante.

### 3.9 Valor medio e indipendenza

Nella proposizione che segue, mostriamo che l'indipendenza implica una proprietà moltiplicativa del valor medio.

**Proposizione 3.16.** *Siano  $X, Y$  variabili casuali scalari indipendenti, definite nello stesso spazio di probabilità e che ammettono valor medio. Allora  $XY$  ammette valor medio e*

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

*Dimostrazione.* Cominciamo col mostrare che  $XY$  ammette valor medio. Usando il fatto che  $p_{X,Y}(x,y) = p_X(x)p_Y(y)$  si ha:

$$\begin{aligned}E(|XY|) &= \sum_{x,y} |x||y|p_{X,Y}(x,y) = \sum_x |x|p_X(x) \sum_y |y|p_Y(y) \\ &= E(|X|)E(|Y|) < +\infty.\end{aligned}$$

La dimostrazione che  $E(XY) = E(X)E(Y)$  è del tutto analoga.

**Corollario 3.2.** *Siano  $X, Y$  due variabili casuali scalari indipendenti, che ammettono valor medio. Allora  $\text{Cov}(X,Y)$  è ben definita, e vale zero.*

Il Corollario 3.2 segue immediatamente dalla Proposizione 3.16. In particolare, ne segue che due variabili casuali scalari indipendenti che ammettono media, so-

no scorrelate. Il viceversa non è necessariamente vero, come mostra l'esempio che segue.

*Esempio 3.6.* Sia  $Z$  una variabile casuale tale che

$$p_Z(0) = p_Z(\pi/2) = p_Z(\pi) = \frac{1}{3}.$$

Posto  $X = \sin(Z)$  e  $Y = \cos(Z)$ , si vede subito che  $XY \equiv 0$  e  $E(Y) = 0$ , da cui

$$\text{Cov}(X, Y) = 0$$

segue immediatamente. Inoltre

$$P(X = 1, Y = 1) = 0 \neq \frac{1}{9} = P(X = 1)P(Y = 1).$$

Chiudiamo questo paragrafo con un risultato semplice ma utile. È facile vedere (provare!) in generale che se due variabili casuali  $X, Y$  ammettono momento secondo, allora

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

Una semplice dimostrazione per induzione, dimostra allora quanto segue.

**Corollario 3.3.** *Siano  $X_1, X_2, \dots, X_n$  variabili casuali che ammettono momento secondo, tali che  $\text{Cov}(X_i, X_j) = 0$  se  $i \neq j$ . Allora*

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i).$$

Per il Corollario 3.2, il risultato del Corollario 3.3 vale, in particolare, se le variabili  $X_1, X_2, \dots, X_n$  sono indipendenti.

**Corollario 3.4.** *Siano  $X_1, X_2, \dots, X_n$  variabili casuali scalari indipendenti, e siano  $\gamma_{X_i}$  le relative funzioni generatrici. Allora, per ogni  $t \in \mathbb{R}$  per cui  $\gamma_{X_i}(t) < +\infty$  per ogni  $i = 1, 2, \dots, n$ , si ha*

$$\gamma_{X_1 + X_2 + \dots + X_n}(t) = \prod_{i=1}^n \gamma_{X_i}(t).$$

*Dimostrazione.* Per la Proposizione 3.5, fissato  $t \in \mathbb{R}$ , le variabili casuali  $e^{tX_1}, e^{tX_2}, \dots, e^{tX_n}$  sono indipendenti. È perciò sufficiente applicare ricorsivamente la Proposizione 3.16.

### 3.10 Classi notevoli di variabili casuali discrete

#### 3.10.1 Variabili casuali di Bernoulli

Una variabile casuale  $X$  è detta *variabile casuale di Bernoulli* se prende valori nell'insieme  $\{0, 1\}$ . Posto

$$p = p_X(1),$$

ne segue che  $p_X(0) = 1 - p$ . La distribuzione di  $X$  è dunque completamente determinata dal parametro  $p \in [0, 1]$ . Diremo allora che  $X$  è una variabile casuale di Bernoulli di parametro  $p$ , e scriveremo

$$X \sim Be(p).$$

È immediato vedere che

$$E(X) = E(X^2) = p,$$

e dunque

$$Var(X) = E(X^2) - E^2(X) = p(1 - p).$$

#### 3.10.2 Variabili casuali binomiali

Una variabile casuale  $X$  si dice *variabile casuale Binomiale* di parametri  $n \geq 1$  e  $p \in [0, 1]$ , e si scrive

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

se  $X$  ha la medesima distribuzione di una variabile casuale del tipo  $X_1 + \dots + X_n$ , dove  $X_i \sim Be(p)$ , e le  $X_i$  sono tra loro indipendenti. È facile calcolare la densità di  $X$ . Se le  $X_i$  rappresentano gli esiti di  $n$  prove ripetute indipendenti per ognuna delle quali l'esito positivo ( $X_i = 1$ ) ha probabilità  $p$ , l'evento  $\{X_1 + \dots + X_n = k\}$  rappresenta la probabilità di avere ottenuto  $k$  esiti positivi. Tale probabilità è stata calcolata nell'Esempio 1.17:

$$p_X(k) = P(X_1 + \dots + X_n = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Inoltre

$$E(X) = E(X_1 + \dots + X_n) = np,$$

e, usando il Corollario 3.3

$$Var(X) = \sum_{i=1}^n Var(X_i) = np(1 - p).$$

Un'ultima, rilevante proprietà delle variabili casuali binomiali è la seguente.

**Proposizione 3.17.** *Siano  $X \sim B(n, p)$  e  $Y \sim B(m, p)$  due variabili casuali indipendenti. Allora  $X + Y \sim B(n + m, p)$ .*

*Dimostrazione.* Siano  $X_1, X_2, \dots, X_n, X_{n+1}, \dots, X_{n+m}$  variabili casuali di Bernoulli di parametro  $p$  indipendenti. La distribuzione congiunta di  $(X, Y)$  è uguale (perché?) alla distribuzione congiunta di  $(X_1 + \dots + X_n, X_{n+1} + \dots + X_{n+m})$ . Ne segue che  $X + Y$  ha la stessa distribuzione di  $X_1 + \dots + X_n + X_{n+1} + \dots + X_{n+m}$  che, per definizione, è una variabile casuale binomiale di parametri  $n + m$  e  $p$ .

### 3.10.3 Variabili casuali Geometriche

Abbiamo visto nell'Esempio 1.17 che, se si eseguono  $N$  prove ripetute indipendenti per ognuna delle quali l'esito positivo ha probabilità  $p$ , allora la probabilità di ottenere il primo esito positivo al tentativo  $n + 1$ -esimo, con  $n < N$ , è  $p(1 - p)^n$ . Per liberarsi della restrizione  $n < N$ , immaginiamo di eseguire una successione infinita di prove ripetute  $N = +\infty$ . C'è un problema, però: l'insieme che descrive gli esiti di una successione infinita di prove,  $\Omega = \{0, 1\}^{\mathbb{N} \setminus \{0\}}$ , non è numerabile, e dunque non può essere lo spazio campionario di un spazio di probabilità discreto. Come vedremo nel prossimo capitolo, introducendo una nozione più generale di spazio di probabilità, tale problema può essere superato. Tralasciando i dettagli tecnici è possibile definire una variabile casuale  $X$  tale che l'evento  $\{X = n\}$  corrisponde a "il primo esito positivo è stato ottenuto al tentativo  $n + 1$ -esimo", ed ha probabilità  $p(1 - p)^n$ . In altre parole  $X$  è il *numero di tentativi falliti prima del primo successo*.

In generale, diremo che una variabile casuale  $X$  a valori in  $\mathbb{N}$  è una variabile casuale Geometrica di parametro  $p \in (0, 1)$ , e scriveremo

$$X \sim Ge(p),$$

se

$$p_X(n) = p(1 - p)^n,$$

per ogni  $n \geq 0$ . Le variabili casuali Geometriche godono di una proprietà rilevante, detta *perdita di memoria*, che si può esprimere con l'affermazione: *se nei primi  $n$  tentativi non è stato ottenuto alcun successo, la probabilità di dover attendere altri  $m$  tentativi prima del primo successo non dipende da  $n$* . La conseguente inutilità di puntare, ad esempio, sui numeri ritardatari nel gioco del Lotto è, spero, evidente!

**Proposizione 3.18.** *Sia  $X \sim Ge(p)$ . Allora, per ogni  $n, m \geq 0$ ,*

$$P(X \geq n + m | X \geq n) = P(X \geq m). \quad (3.42)$$

*Dimostrazione.* Anzitutto si osservi che, poichè  $\{X \geq n + m\} \subseteq \{X \geq n\}$

$$\begin{aligned}
 P(X \geq n+m | X \geq n) &= \frac{P(\{X \geq n+m\} \cap \{X \geq n\})}{P(\{X \geq n\})} \\
 &= \frac{P(\{X \geq n+m\})}{P(\{X \geq n\})}.
 \end{aligned}$$

Dobbiamo dunque calcolare probabilità del tipo  $P(X \geq n)$ . Si ottiene:

$$P(X \geq n) = \sum_{k=n}^{+\infty} p(1-p)^k = p(1-p)^n \sum_{h=0}^{+\infty} (1-p)^h = (1-p)^n,$$

dove si è usato il fatto di saper calcolare la somma di una serie geometrica:

$$\sum_{h=0}^{+\infty} (1-p)^h = \frac{1}{1-(1-p)} = \frac{1}{p}.$$

Ma allora:

$$\begin{aligned}
 P(X \geq n+m | X \geq n) &= \frac{P(\{X \geq n+m\})}{P(\{X \geq n\})} \\
 &= \frac{(1-p)^{n+m}}{(1-p)^n} \\
 &= (1-p)^m \\
 &= P(X \geq m).
 \end{aligned}$$

La proprietà di perdita di memoria, caratterizza le variabili casuali Geometriche nella classe delle variabili casuali a valori naturali, come mostriamo nella seguente Proposizione.

**Proposizione 3.19.** *Sia  $X$  una variabile casuale a valori in  $\mathbb{N}$ , tale che (3.42) è verificata per ogni  $n, m \geq 0$ . Allora  $X$  è una variabile casuale Geometrica oppure  $P(X=0) = 1$ .*

*Dimostrazione.* Posto  $m = 1$  in (3.42), si ottiene

$$P(X \geq n+1) = P(X \geq n)P(X \geq 1)$$

che, per induzione, implica

$$P(X \geq n) = P(X \geq 1)^n,$$

per ogni  $n \geq 1$ . Ma allora, osservando che  $\{X = n\} = \{X \geq n\} \setminus \{X \geq n+1\}$ ,

$$p_X(n) = P(X = n) = P(X \geq n) - P(X \geq n+1) = P(X \geq 1)^n - P(X \geq 1)^{n+1} = (1 - P(X \geq 1))P(X \geq 1)^n,$$

per ogni  $n \geq 1$ . Inoltre

$$p_X(0) = P(X = 0) = 1 - P(X \geq 1).$$

Dunque, posto  $p = 1 - P(X \geq 1)$ ,

$$p_X(n) = p(1-p)^n$$

per ogni  $n \geq 0$ , che conclude la dimostrazione.

Calcoliamo ora media e varianza di una variabile casuale Geometrica. A questo scopo calcoliamo la funzione generatrice dei momenti di una variabile casuale geometrica.

$$\gamma_X(t) = \sum_{n=0}^{+\infty} e^{tn} p(1-p)^n = p \sum_{n=0}^{+\infty} ((1-p)e^t)^n.$$

quest'ultima è una serie geometrica di ragione  $(1-p)e^t$  cioè  $t < -\log(1-p)$ . Perciò

$$\gamma_X(t) = \begin{cases} \frac{p}{1-(1-p)e^t} & \text{se } t < -\log(1-p) \\ +\infty & \text{altrimenti.} \end{cases}$$

Per il Teorema 3.3,  $E(X^n) = \gamma^{(n)}(0)$  per ogni  $n \geq 1$ . Essendo

$$\begin{aligned} \gamma'_X(t) &= \frac{p(1-p)e^t}{(1-(1-p)e^t)^2}, \\ \gamma''_X(t) &= \frac{p(1-p)e^t \left[ (1-(1-p)e^t)^2 + 2(1-(1-p)e^t)(1-p)e^t \right]}{(1-(1-p)e^t)^4}, \end{aligned}$$

si trova

$$E(X) = \gamma'(0) = \frac{1-p}{p}, \quad E(X^2) = \gamma''(0) = \frac{(1-p)(2-p)}{p^2},$$

da cui segue

$$\text{Var}(X) = E(X^2) - E^2(X) = \frac{1-p}{p^2}.$$

### 3.10.4 Variabili casuali di Poisson

In numerose situazioni concrete, ci si trova a considerare delle variabili casuali binomiali i cui parametri,  $n$  e  $p$ , sono tali che  $n$  è molto grande e  $p$  è molto piccolo. In altre parole, si eseguono numerose ripetizioni di una prova che ha esito positivo con probabilità piccola. Supponiamo, ad esempio, di considerare il numero di accessi, in un fissato intervallo di tempo, ad un certo servizio (lo sportello di un ufficio pubblico, un pagina web, un centralino...). Vi sarà un numero grande di utenti, diciamo  $n$ , che ha accesso a tale servizio. Si osserva però che, tipicamente, un numero di utenti molto minore di  $n$  accede effettivamente al servizio in un intervallo di tempo della lunghezza di quello fissato. Un ragionevole modello matematico per tale situazione,

è di supporre che ogni utente abbia una probabilità piccola,  $p \ll 1$ , di accedere al servizio. In una approssimazione che è assai buona in molti casi, assumiamo che il valore di  $p$  sia uguale per tutti gli utenti, e che ogni utente si comporti in modo indipendente dagli altri. Sotto queste ipotesi, il numero  $X$  di utenti che effettivamente accede al servizio è una variabile casuale binomiale di parametri  $n$  e  $p$ , cioè

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Se  $n$  è molto grande, calcoli espliciti con questa densità risultano estremamente pesanti, se non impraticabili, data la difficoltà di calcolare i coefficienti binomiali per grandi valori di  $n$ . È allora interessante analizzare il comportamento asintotico di  $p_X(k)$  quando  $n \rightarrow +\infty$  e  $p \rightarrow 0$ . Per ottenere un comportamento limite non banale, è necessario che  $p$  vada a zero “proporzionalmente” a  $\frac{1}{n}$ . Per semplicità poniamo semplicemente  $p = \frac{\lambda}{n}$  con  $\lambda > 0$ , ma il calcolo che segue si può facilmente modificare per coprire il caso  $p = p_n$  con  $\lim_{n \rightarrow +\infty} n p_n = \lambda > 0$ . Si ha

$$\begin{aligned} p_X(k) &= \frac{n!}{k!(n-k)!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \frac{n(n-1) \cdots (n-k+1)}{n^k} \frac{1}{\left(1 - \frac{\lambda}{n}\right)^k} \left(1 - \frac{\lambda}{n}\right)^n. \end{aligned}$$

D'altra parte

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{n(n-1) \cdots (n-k+1)}{n^k} &= 1, \\ \lim_{n \rightarrow +\infty} \left(1 - \frac{\lambda}{n}\right)^k &= 1 \end{aligned}$$

e

$$\lim_{n \rightarrow +\infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}.$$

Ne segue che

$$\lim_{n \rightarrow +\infty} p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Si noti che l'espressione appena ottenuta è la densità di una variabile casuale a valori in  $\mathbb{N}$ , in quanto

$$\sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = 1.$$

Tale procedura di approssimazione giustifica la definizione di una nuova classe di variabili casuali. Una variabile casuale  $X$  a valori in  $\mathbb{N}$  si dice variabile casuale di Poisson di parametro  $\lambda > 0$ , e si scrive

$$X \sim Po(\lambda),$$

se

$$p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

per ogni  $k \in \mathbb{N}$ . Tali variabili casuali vengono comunemente usate per modellare, tra le altre, quantità del tipo “numero di accessi ad un servizio”, come abbiamo sopra giustificato.

È agevole calcolare la funzione generatrice dei momenti di  $X \sim Po(\lambda)$ :

$$\gamma_X(t) = e^{-\lambda} \sum_{n=0}^{+\infty} \frac{(e^t n)^n}{n!} = e^{\lambda(e^t - 1)}.$$

Ne segue che

$$\gamma'_X(t) = \lambda e^t \gamma_X(t), \quad \gamma''_X(t) = \lambda e^t [\gamma_X(t) + \gamma'_X(t)],$$

da cui  $E(X) = \gamma'_X(0) = \lambda$ ,  $E(X^2) = \gamma''_X(0) = \lambda + \lambda^2$  e perciò

$$\text{Var}(X) = \lambda.$$

Il risultato che ora illustriamo afferma che la somma di due variabili casuali di Poisson indipendenti è ancora una variabile casuale di Poisson.

**Proposizione 3.20.** *Siano  $X \sim Po(\lambda)$  e  $Y \sim Po(\mu)$  variabili casuali indipendenti. Allora  $X + Y \sim Po(\lambda + \mu)$ .*

Prima di dimostrare la Proposizione 3.20, mostriamo un risultato del tutto generale.

**Proposizione 3.21.** *Siano  $X$  e  $Y$  due variabili casuali discrete scalari, e sia  $p_{X,Y}$  la loro densità congiunta. Allora*

$$p_{X+Y}(z) = P(X + Y = z) = \sum_x p_{X,Y}(x, z - x) = \sum_y p_{X,Y}(z - y, y). \quad (3.43)$$

*Dimostrazione.* Usando la  $\sigma$ -additività della probabilità, si ha

$$\begin{aligned} P(X + Y = z) &= P\left(\bigcup_{x \in X(\Omega)} \{X = x, Y = z - x\}\right) \\ &= \sum_{x \in X(\Omega)} P(X = x, Y = z - x) \\ &= \sum_x p_{X,Y}(x, z - x), \end{aligned}$$

ove si è osservato che  $P(X = x, Y = z - x) = 0$  se  $x \notin X(\Omega)$ . L'ultima uguaglianza in (3.43) si ottiene scambiando i ruoli di  $X$  e  $Y$ .

Nel caso in cui le variabili casuali  $X$  e  $Y$  nella Proposizione 3.21 siano indipendenti, la (3.43) assume la forma

$$p_{X+Y}(z) = \sum_x p_X(x) p_Y(z-x) = \sum_y p_X(z-y) p_Y(y), \quad (3.44)$$

che ora useremo.

*Dimostrazione (Dimostrazione della Proposizione 3.20).* Per la (3.44) si ha

$$\begin{aligned} p_{X+Y}(n) &= \sum_{k=0}^n e^{-\lambda} e^{-\mu} \frac{\lambda^k}{k!} \frac{\mu^{n-k}}{(n-k)!} \\ &= e^{-(\lambda+\mu)} \frac{1}{n!} \sum_{k=0}^n \binom{n}{k} \lambda^k \mu^{n-k} \\ &= e^{-(\lambda+\mu)} \frac{(\lambda + \mu)^n}{n!}, \end{aligned}$$

dove nell'ultima uguaglianza abbiamo applicato la formula di Newton per la potenza di un binomio.

### 3.10.5 Approssimazione di Poisson

Nell'introdurre le variabili casuali di Poisson abbiamo dimostrato il seguente risultato.

**Proposizione 3.22.** Per  $n \geq 1$  sia  $X_n \sim B(n, p_n)$ , e supponiamo che esista il limite

$$\lim_{n \rightarrow +\infty} np_n =: \lambda.$$

Allora, per ogni  $k \geq 0$

$$\lim_{n \rightarrow +\infty} p_{X_n}(k) = e^{-\lambda} \frac{\lambda^k}{k!} = p_X(k),$$

dove  $X \sim Po(\lambda)$

In questo paragrafo ci proponiamo di migliorare il precedente risultato in due direzioni.

- Mostriamo che sotto opportune condizioni, la densità di una opportuna variabile di Poisson è una buona approssimazione della densità di una somma di variabili di Bernoulli indipendenti; nel caso in cui i parametri delle variabili di Bernoulli siano tutti uguali, la loro somma è una variabile casuale Binomiale, ma non richiederemo tale omogeneità.
- Forniremo un risultato di approssimazione non solo nella forma di limite, ma fornendo una stima esplicita dell'errore per ogni fissato valore del numero di variabili di Bernoulli e dei loro rispettivi parametri.

Nel seguito, se  $E$  è un insieme e  $p : E \rightarrow [0, 1]$  è una funzione per cui

$$\sum_{x \in E} p(x) = 1,$$

diremo che  $p$  è una *densità discreta* su  $E$ . Cominciamo con il fornire una nozione di *distanza* fra due densità discrete.

**Definizione 3.11.** Siano  $p$  e  $q$  due densità discrete sullo stesso insieme  $E$ . Chiameremo *distanza in variazione totale* fra  $p$  e  $q$  la quantità

$$d(p, q) := \frac{1}{2} \sum_{x \in E} |p(x) - q(x)|.$$

Non è difficile mostrare che  $d(\cdot, \cdot)$  è effettivamente una distanza nell'insieme delle densità discrete su  $E$ .

Illustreremo ora un metodo molto utile per fornire stime sulla distanza in variazione totale tra due densità discrete, noto come il *metodo dell'accoppiamento*. Accoppiare due densità  $p$  e  $q$  significa realizzare nello stesso spazio di probabilità due variabili casuali  $X$  e  $Y$  le cui densità siano rispettivamente  $p$  e  $q$ . Ciò è equivalente a definire una densità congiunta di cui  $p$  e  $q$  siano le densità marginali. Chiaramente questo può essere fatto in molti modi diversi. Ad esempio, date due densità discrete  $p$  e  $q$ , un modo di "accoppiarle" è di considerare due variabili casuali *indipendenti* di densità, rispettivamente,  $p$  e  $q$ . Questo modo semplice risulta, nella gran parte di casi, scarsamente efficace.

Il risultato che segue costituisce lo strumento fondamentale del metodo dell'accoppiamento.

**Proposizione 3.23.** Siano  $X$  e  $Y$  due variabili casuali, definite nello stesso spazio di probabilità  $(\Omega, P)$ , e a valori in  $E$ . Allora

$$d(p_X, p_Y) \leq P(X \neq Y).$$

*Dimostrazione.* Osserviamo che

$$\begin{aligned} p_X(x) - p_Y(x) &= P(X = x) - P(Y = x) \leq P(X = x) - P(X = x, Y = x) \\ &= P(\{X = x\} \setminus \{X = x, Y = x\}) = P(X = x, Y \neq x). \end{aligned}$$

Analogamente,

$$p_Y(x) - p_X(x) \leq P(Y = x, X \neq x).$$

Ne segue che

$$|p_X(x) - p_Y(x)| \leq P(X = x, Y \neq x) + P(Y = x, X \neq x). \quad (3.45)$$

Inoltre, considerando il fatto che  $\{\{X = x\} : x \in X(\Omega)\}$  è una partizione di  $\Omega$ , abbiamo che

$$\begin{aligned}
 P(X \neq Y) &= P\left(\bigcup_{x \in X(\Omega)} \{X \neq Y\} \cap \{X = x\}\right) \\
 &= \sum_x P(X \neq Y, X = x) = \sum_x P(X = x, Y \neq x).
 \end{aligned}$$

Per simmetria, si ha anche  $P(X \neq Y) = \sum_x P(Y = x, X \neq x)$ , e quindi

$$2P(X \neq Y) = \sum_x [P(X = x, Y \neq x) + P(Y = x, X \neq x)]. \quad (3.46)$$

La conclusione segue dal confronto di (3.45) e (3.46).

Vediamo ora l'applicazione del precedente risultato a due casi speciali.

**Lemma 3.1.** A) Siano  $X \sim Be(p)$  e  $Y \sim Po(p)$ . Allora

$$d(p_X, p_Y) \leq p^2.$$

B) Siano  $X \sim Po(\lambda)$  e  $Y \sim Po(\mu)$ . Allora

$$d(p_X, p_Y) \leq |\lambda - \mu|.$$

*Dimostrazione.*

A) Per capire l'idea della dimostrazione, supponiamo che  $p$  sia “piccolo”, e  $Y \sim Po(p)$ . Poiché  $P(Y = 0) = e^{-p} \simeq 1$ , è abbastanza facile vedere che, posto  $X := \mathbf{1}_{\{Y \geq 1\}}$ , le densità di  $X$  e  $Y$  sono “vicine” in variazione totale; infatti le variabili casuali  $X$  e  $Y$  sono definite sullo stesso spazio di probabilità, e quindi è possibile applicare la Proposizione 3.23. Tuttavia  $X \sim Be(1 - e^{-p})$  mentre, nell'enunciato del Teorema, compare  $X \sim Be(p)$ . Ma per  $p$  piccolo,  $1 - e^{-p} \simeq p$ . È quindi necessaria una “leggera” correzione della costruzione precedente. Sia  $Z$  una variabile casuale avente la seguente distribuzione:

$$P(Z = k) = \begin{cases} e^{-p} \frac{p^k}{k!} & \text{per } k \geq 1 \\ 1 - p & \text{per } k = 0 \\ e^{-p} - (1 - p) & \text{per } k = -1 \end{cases}$$

Essendo  $e^{-p} \geq 1 - p$ , si tratta di una buona definizione. Definiamo poi

$$X := \mathbf{1}_{\{Z \neq 0\}}$$

e

$$Y := Z \mathbf{1}_{\{Z \neq -1\}}.$$

Lasciamo al lettore la semplice verifica del fatto che  $X \sim Be(p)$  e  $Y \sim Po(p)$ . Inoltre, osservando che  $\{X \neq Y\} = \{Z \neq 0, 1\}$ , per la Proposizione 3.23 e la disuguaglianza  $e^{-p} \geq 1 - p$ , abbiamo:

$$d(p_X, p_Y) \leq 1 - P(Z=0) - P(Z=1) = 1 - e^{-p} - pe^{-p} \leq 1 - (1-p) - p(1-p) = p^2.$$

B) Senza perdita di generalità, possiamo supporre  $\lambda > \mu$ . Consideriamo, nello stesso spazio di probabilità,  $X \sim Po(\lambda)$  e  $Z_n \sim B(n, \frac{\mu}{\lambda})$ ,  $n \geq 0$  (si intende  $Z_0 \equiv 0$ ), in modo tale che, per ogni  $n \geq 1$ ,  $X$  e  $Z_n$  siano indipendenti. Definiamo

$$Y(\omega) := Z_{X(\omega)}.$$

Cominciamo con il mostrare che  $Y \sim Po(\mu)$ .

$$\begin{aligned} P(Y=k) &= P\left(\bigcup_{n \geq k} \{X=n, Z_n=k\}\right) = \sum_{n \geq k} P(X=n, Z_n=k) \\ &= \sum_{n \geq k} P(X=n)P(Z_n=k) = \sum_{n \geq 1} e^{-\lambda} \frac{\lambda^n}{n!} \binom{n}{k} \left(\frac{\mu}{\lambda}\right)^k \left(1 - \frac{\mu}{\lambda}\right)^{n-k} \\ &= e^{-\lambda} \frac{\lambda^k}{k!} \left(\frac{\mu}{\lambda}\right)^k \sum_{n \geq k} \frac{\lambda^{n-k}}{(n-k)!} \left(1 - \frac{\mu}{\lambda}\right)^{n-k} = e^{-\lambda} \frac{\mu^k}{k!} e^{\lambda-\mu} = e^{-\mu} \frac{\mu^k}{k!}. \end{aligned}$$

Usando ora la Proposizione 3.23

$$\begin{aligned} d(p_X, p_Y) &\leq 1 - P(X=Y) = 1 - P\left(\bigcup_{n \geq 0} \{X=n, Z_n=n\}\right) \\ &= 1 - \sum_{n \geq 0} P(X_n=n)P(Z_n=n) = 1 - \sum_{n \geq 0} e^{-\lambda} \frac{\lambda^n}{n!} \left(\frac{\mu}{\lambda}\right)^n = 1 - e^{-\lambda} \sum_{n \geq 0} \frac{\mu^n}{n!} \\ &= 1 - e^{-(\lambda-\mu)} \leq \lambda - \mu, \end{aligned}$$

dove, nell'ultima disuguaglianza abbiamo usato il fatto che  $1 - e^{-x} \leq x$  per ogni  $x \in \mathbb{R}$ .

Siamo ora in condizione di enunciare il principale risultato di questo paragrafo, noto come *Legge dei Piccoli Numeri*.

**Teorema 3.4.** *Siano  $X_1, X_2, \dots, X_n$  variabili aleatorie indipendenti tali che  $X_i \sim Be(q_i)$ . Posto  $S_n := X_1 + X_2 + \dots + X_n$  e  $W_n \sim Po(q_1 + q_2 + \dots + q_n)$ , abbiamo*

$$d(p_{S_n}, p_{W_n}) \leq \sum_{i=1}^n q_i^2.$$

Prima di dimostrare il Teorema 3.4, notiamo come da esso segua la Proposizione 3.22. Se le  $X_i$  sono indipendenti e  $q_i \equiv \frac{p_i}{n}$  per ogni  $i$ , allora  $S_n \sim B(n, p_n)$ , e  $q_1 + q_2 + \dots + q_n = np_n$ . Allora, per  $k \geq 0$ , se  $np_n \rightarrow \lambda$ , posto  $Y \sim Po(\lambda)$  e  $W_n \sim Po(np_n)$ , usando anche il Lemma 3.1 B),

$$d(p_{S_n}, p_Y) \leq d(p_{S_n}, p_{W_n}) + d(p_{W_n}, p_Y) \leq np_n^2 + |\lambda - np_n|$$

da cui

$$\lim_{n \rightarrow +\infty} d(p_{S_n}, p_Y) = 0. \quad (3.47)$$

Ovviamente la Proposizione 3.22 segue da (3.47), essendo

$$|p_{S_n}(k) - p_Y(k)| \leq 2d(p_{S_n}, p_Y).$$

**Dimostrazione del Teorema 3.4.** Per  $i = 1, 2, \dots, n$ , siano  $Z_i$  variabili casuali indipendenti con distribuzione

$$P(Z_i = k) = \begin{cases} e^{-p_i} \frac{p_i^k}{k!} & \text{per } k \geq 1 \\ 1 - p_i & \text{per } k = 0 \\ e^{-p_i} - (1 - p_i) & \text{per } k = -1 \end{cases}$$

Posto  $X_i := \mathbf{1}_{\{Z_i \neq 0\}}$  e  $Y_i := Z_i \mathbf{1}_{\{Z_i \neq -1\}}$  abbiamo che

- $X_i \sim Be(q_i)$  e  $Y_i \sim Po(q_i)$ , e  $P(X_i \neq Y_i) \leq q_i^2$ , come visto nella dimostrazione del Lemma 3.1 A).
- Per la Proposizione 3.5, le  $X_i$  sono indipendenti.
- Analogamente, le  $Y_i$  sono indipendenti, e quindi  $W_n := Y_1 + Y_2 + \dots + Y_n \sim Po(q_1 + q_2 + \dots + q_n)$ .

Pertanto, se  $S_n := X_1 + X_2 + \dots + X_n$ ,

$$d(p_{S_n}, p_{W_n}) \leq P(S_n \neq W_n) \leq P\left(\bigcup_{i=1}^n \{X_i \neq Y_i\}\right) \leq \sum_{i=1}^n P(X_i \neq Y_i) \leq \sum_{i=1}^n q_i^2.$$

□

### 3.10.6 Funzione di ripartizione

**Definizione 3.12.** Sia  $X$  una variabile casuale discreta scalare. La funzione  $F_X : \mathbb{R} \rightarrow [0, 1]$  definita da

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

si dice *funzione di ripartizione* della variabile casuale  $X$ .

Alcune proprietà elementari delle funzione di ripartizione sono mostrate nella seguente proposizione.

**Proposizione 3.24.** Sia  $F_X$  la funzione di ripartizione di una variabile casuale discreta  $X$ . Allora

- (i)  $F_X$  è non decrescente
- (ii)  $F_X$  è continua da destra.

(iii)

$$\lim_{x \rightarrow -\infty} F_X(x) = 0.$$

(iv)

$$\lim_{x \rightarrow +\infty} F_X(x) = 1.$$

*Dimostrazione.* (i) Se  $x < y$  allora  $\{X \leq x\} \subseteq \{X \leq y\}$ . Perciò

$$F_X(x) = P(X \leq x) \leq P(X \leq y) = F_X(y).$$

(ii) Sia  $x \in \mathbb{R}$ . Basta dimostrare che se  $(x_n)$  è una successione decrescente tale che  $x_n \downarrow x$  allora

$$\lim_{n \rightarrow +\infty} F_X(x_n) = F(x). \quad (3.48)$$

Si osservi che

$$\{X \leq x\} = \bigcap_n \{X \leq x_n\},$$

e che la successione di eventi  $\{X \leq x_n\}$  è decrescente. Per la Proposizione 1.3 si ha

$$F_X(x) = P(X \leq x) = \lim_{n \rightarrow +\infty} P(X \leq x_n) = \lim_{n \rightarrow +\infty} F_X(x_n).$$

(iii) È sufficiente applicare la Proposizione 1.3 alla famiglia decrescente di eventi  $\{X \leq x_n\}$  dove  $x_n \downarrow -\infty$ , osservando che

$$\bigcap_n \{X \leq x_n\} = \emptyset.$$

(iv) È sufficiente applicare la Proposizione 1.3 alla famiglia crescente di eventi  $\{X \leq x_n\}$  dove  $x_n \uparrow +\infty$ , osservando che

$$\bigcup_n \{X \leq x_n\} = \Omega.$$

Per quanto visto nell'Osservazione 3.2, la funzione di ripartizione si può esprimere in termini della densità come segue:

$$F_X(x) = \sum_{y \leq x} p_X(y).$$

Viceversa, mostriamo che la densità è esprimibile in termini della funzione di ripartizione. In particolare, la funzione di ripartizione determina completamente la distribuzione di una variabile casuale.

**Proposizione 3.25.** *Se  $X$  è una variabile casuale discreta scalare, per ogni  $x \in \mathbb{R}$  vale la relazione*

$$p_X(x) = F_X(x) - F_X(x^-),$$

dove

$$F_X(x^-) = \lim_{y \rightarrow x^-} F_X(y).$$

*Dimostrazione.* Se  $y < x$ , si ha che

$$P(X \in (y, x]) = P(X \leq x) - P(X \leq y) = F_X(x) - F_X(y).$$

Allora, se  $(y_n)$  è una successione di elementi di  $(-\infty, x)$  tali che  $y_n \uparrow x$ , poichè gli eventi  $\{X \in (y_n, x]\}$  formano una successione decrescente per cui  $\bigcap_n \{X \in (y_n, x]\} = \{X = x\}$ , si ha

$$\begin{aligned} p_X(x) = P(X = x) &= \lim_{n \rightarrow +\infty} P(X \in (y_n, x]) = \lim_{n \rightarrow +\infty} [F_X(x) - F_X(y_n)] = F_X(x) - \lim_{n \rightarrow +\infty} F_X(y_n) \\ &= F_X(x) - F_X(x^-). \end{aligned}$$

*Osservazione 3.12.* Dalla proposizione precedente segue che  $F_X$  è discontinua in  $x$  se e solo se  $p_X(x) > 0$ , e il valore di  $p_X(x)$  rappresenta l'entità del "salto" di  $F_X$  in  $x$ . Nel caso in cui  $X(\Omega)$  non abbia punti di accumulazione, dall'identità

$$F_X(x) = \sum_{y \leq x} p_X(y).$$

segue che  $F_X$  è una funzione costante a tratti.

La funzione di ripartizione è spesso utile nei calcoli con le distribuzioni di variabili casuali. In particolare, come ora vedremo, è usata quando si tratta di calcolare la distribuzione del massimo o del minimo di variabili casuali indipendenti.

**Proposizione 3.26.** *Siano  $X_1, X_2, \dots, X_n$  variabili casuali scalari indipendenti. Definiamo:*

$$Z = \max(X_1, \dots, X_n), \quad W = \min(X_1, \dots, X_n).$$

Allora

$$F_Z(x) = \prod_{k=1}^n F_{X_k}(x), \quad (3.49)$$

e

$$F_W(x) = 1 - \prod_{k=1}^n [1 - F_{X_k}(x)]. \quad (3.50)$$

*Dimostrazione.* Cominciamo col dimostrare (3.49). Si osservi che

$$\{Z \leq x\} = \bigcap_{k=1}^n \{X_k \leq x\},$$

usando l'indipendenza delle  $X_k$  si ha

$$F_Z(x) = P(Z \leq x) = \prod_{k=1}^n P(X_k \leq x) = \prod_{k=1}^n F_{X_k}(x).$$

Per quanto riguarda (3.50), si vede che

$$\{W > x\} = \bigcap_{k=1}^n \{X_k > x\}.$$

e perciò

$$F_W(x) = 1 - P(W > x) = 1 - \prod_{k=1}^n P(X_k > x) = 1 - \prod_{k=1}^n [1 - P(X_k \leq x)] = 1 - \prod_{k=1}^n [1 - F_{X_k}(x)].$$

*Esempio 3.7.* Siano  $X_1, \dots, X_n \sim Ge(p)$  indipendenti. Determiniamo la densità di  $Z = \max(X_1, \dots, X_n)$  e  $W = \min(X_1, \dots, X_n)$ .

Come si vede dai calcoli fatti nella Proposizione 3.18,

$$F_{X_i}(k) = 1 - (1 - p)^{k+1}.$$

Dunque

$$F_Z(k) = \left[1 - (1 - p)^{k+1}\right]^n.$$

Essendo, per  $k \in \mathbb{N}$ ,  $F_Z(k^-) = F_Z(k - 1)$ , si ha, usando la Proposizione 3.25

$$p_Z(k) = F_Z(k) - F_Z(k - 1) = \left[1 - (1 - p)^{k+1}\right]^n - \left[1 - (1 - p)^k\right]^n.$$

Analogamente,

$$F_W(k) = 1 - \left[(1 - p)^{k+1}\right]^n = 1 - [(1 - p)^n]^{k+1},$$

che coincide con la funzione di ripartizione di una variabile casuale Geometrica di parametro  $1 - (1 - p)^n$ . Poichè la funzione di ripartizione individua completamente la distribuzione, possiamo concludere che

$$W \sim Ge(1 - (1 - p)^n).$$

### 3.11 Un'applicazione alla finanza: il modello di Black-Scholes discreto

Consideriamo un mercato finanziario molto semplificato, in cui sia presente un unico titolo rischioso, ad esempio un'azione, e in cui non vi sia inflazione e non vi siano costi di transazione. Sia  $X_0 \equiv x$  il valore odierno del titolo, che viene aggiornato una volta al giorno. Le variabili casuali  $X_1, X_2, \dots$  rappresentano i valori del titolo nei giorni successivi.

Gli istituti finanziari, ad esempio le banche, offrono ai loro clienti le cosiddette *opzioni*. Un esempio di opzione è l'*opzione call europea*: il cliente, che ha investito

nel titolo, acquista il diritto di vendere al giorno  $N$  all'istituto finanziario una (o più) unità del titolo ad un prezzo  $\pi = \pi(X_N)$  dato da

$$\pi(X_N) = \begin{cases} X_N & \text{se } X_N \geq a \\ a & \text{se } X_N < a \end{cases}$$

dove  $a$  è un prezzo minimo prefissato. In questo modo il cliente limita il rischio di perdite.

Una delle domande fondamentali in finanza è: qualè il *giusto* prezzo per una tale opzione? La nozione di giusto prezzo è legata a quella di *strategia di copertura*. Supponiamo che al momento di vendita dell'opzione, l'istituto finanziario utilizzi un capitale  $V_0 = V$  per attivare un *portafoglio*, di cui una parte investita in  $a_0$  unità del titolo rischioso, la parte restante  $c_0$  investita in un titolo non rischioso, cioè a tasso di rendimento costante. Per semplicità di calcolo assumiamo che tale tasso sia uguale a zero. In altre parole

$$V_0 = a_0 X_0 + c_0.$$

Il giorno successivo, essendo variato da  $X_0$  a  $X_1$  il prezzo del titolo rischioso, il capitale diventa

$$V_1 = V_0 + a_0(X_1 - X_0).$$

A questo punto l'istituto finanziario può modificare la porzione di capitale investita nel titolo rischioso, passando da  $a_0$  ad  $a_1$  unità possedute, con il rimanente  $c_1 = V_1 - a_1 X_1$  mantenuta a rendimento costante. Dopo due giorni il capitale disponibile diventa:

$$V_2 = V_1 + a_1(X_2 - X_1) = V_0 + a_0(X_1 - X_0) + a_1(X_2 - X_1),$$

Iterando il procedimento, si ottiene che il capitale dopo  $n$  giorni è

$$V_n = V_0 + \sum_{i=0}^{n-1} a_i(X_{i+1} - X_i).$$

I coefficienti  $a_0, a_1, \dots, a_{N-1}$  rappresentano la strategia di investimento. Affinché si tratti di una strategia effettivamente implementabile essa dev'essere *non-anticipativa*; in altre parole la scelta di  $a_n$  può dipendere dall'evoluzione del prezzo solo fino all'istante  $n$ , cioè  $a_n = a_n(X_0, X_1, \dots, X_n)$  (ciò significa che la strategia si può basare solo sull'informazione effettivamente disponibile). È chiaro che in questo caso anche  $V_n = V_n(X_0, X_1, \dots, X_n)$ .

**Definizione 3.13.** Per un dato capitale iniziale  $V_0$ , una strategia non-anticipativa  $a_0, a_1, \dots, a_{N-1}$  si dice *strategia di copertura* se, per qualunque evoluzione  $X_1, \dots, X_N$  del prezzo, si ha

$$V_N(X_0, X_1, \dots, X_N) = \pi(X_N).$$

Nel caso in cui vi sia un'unico valore di  $V_0$  per cui esiste una strategia di copertura, diremo che  $V_0$  è il *prezzo* dell'opzione.

Si noti che è anche naturale assumere che una strategia sia tale che, in ogni giorno  $n$ ,  $a_n \geq 0$  e  $a_n X_n \leq V_n$ . Se fosse  $a_n < 0$ , l'istituto avrebbe operato un *vendita allo scoperto*, mentre se  $a_n X_n > V_n$  avrebbe dovuto ricorrere ad ulteriore capitale, cioè ad un prestito. Nel semplice modello di mercato che studieremo tra poco, mostreremo che la strategia di copertura è unica, e soddisfa alle suddette proprietà, che quindi non assumeremo a priori.

Lo stabilire l'esistenza ed, eventualmente, l'unicità del capitale iniziale e della strategia per la copertura dell'opzione, costituisce un problema di estrema rilevanza economica e che si presta ad una trattazione matematica elegante ed efficace. La nozione di copertura si può evidentemente estendere ad una qualunque funzione  $\pi(X_N)$ . I modelli di mercato in cui ad ogni tale funzione è associata una strategia di copertura vengono chiamati mercati *completi*. La completezza di un modello di mercato ammette varie caratterizzazioni matematiche, interessanti e utili. In queste note non affronteremo questo problema; ci limiteremo a semplificare ulteriormente il nostro modello di mercato con un unico titolo, e mostrare che in esso l'opzione europea ammette un prezzo. A tale scopo, definiamo, per  $n = 1, 2, \dots, N$ ,  $R_n = X_n/X_{n-1}$ , e assumiamo che le variabili casuali  $R_n$  possano assumere con probabilità strettamente positiva esattamente due valori  $c$  e  $C$ , con  $0 < c < 1 < C$ . Ciò significa che ogni giorno il prezzo del titolo può soltanto o contrarsi di un fattore  $c$  o dilatarsi di un fattore  $C$ . Osserviamo che  $X_n = R_n R_{n-1} \cdots R_1 x > 0$  se  $x > 0$ .

Possiamo ora analizzare nel dettaglio il problema della copertura. Consideriamo il capitale  $V_{N-1} = V_{N-1}(X_0, \dots, X_{N-1})$  a disposizione dell'istituto al giorno  $N-1$ , e sia  $a_{N-1}$  l'ultimo passo della strategia. Noto il valore  $X_{N-1}$  del prezzo del titolo, il prezzo  $X_N$  del giorno successivo può assumere i due valori  $cX_{N-1}$  e  $CX_{N-1}$ . Se la strategia in esame copre l'opzione, allora dev'essere vero che

$$\pi(X_N) = V_{N-1} + a_{N-1}(X_N - X_{N-1})$$

sia per  $X_N = cX_{N-1}$  che per  $X_N = CX_{N-1}$ . In altre parole

$$\begin{aligned}\pi(cX_{N-1}) &= V_{N-1} - a_{N-1}(1-c)X_{N-1} \\ \pi(CX_{N-1}) &= V_{N-1} + a_{N-1}(C-1)X_{N-1},\end{aligned}$$

dalla quale si ricava

$$a_{N-1} = \frac{\pi(CX_{N-1}) - \pi(cX_{N-1})}{(C-c)X_{N-1}} \quad (3.51)$$

$$V_{N-1} = \frac{1-c}{C-c} \pi(CX_{N-1}) + \frac{C-1}{C-c} \pi(cX_{N-1}). \quad (3.52)$$

Dunque, il capitale  $V_{N-1}$  e la strategia  $a_{N-1}$  sono determinati in modo *univoco* dalla richiesta di copertura dell'opzione. A questo punto il procedimento si può iterare. Per coprire l'opzione è necessario disporre del capitale  $V_{N-1}$  dato da (3.52) al giorno  $N-1$ . Ragionando come prima, il capitale  $V_{N-2}$  e la strategia  $a_{N-2}$  devono soddisfare le relazioni

$$\begin{aligned} V_{N-1}(cX_{N-2}) &= V_{N-2} - a_{N-2}(1-c)X_{N-2} \\ V_{N-1}(CX_{N-2}) &= V_{N-2} + a_{N-2}(C-1)X_{N-2}, \end{aligned}$$

da cui

$$\begin{aligned} a_{N-2} &= \frac{V_{N-1}(CX_{N-2}) - V_{N-1}(cX_{N-2})}{(C-c)X_{N-2}} \\ V_{N-2} &= \frac{1-c}{C-c}V_{N-1}(CX_{N-2}) + \frac{C-1}{C-c}V_{N-1}(cX_{N-2}). \end{aligned}$$

Iterando, per  $n \geq 0$ , troveremo le formule ricorsive

$$a_n = \frac{V_{n+1}(CX_n) - V_{n+1}(cX_n)}{(C-c)X_n} \quad (3.53)$$

$$V_n = \frac{1-c}{C-c}V_{n+1}(CX_n) + \frac{C-1}{C-c}V_{n+1}(cX_n). \quad (3.54)$$

Dovendo essere  $V_N(X_N) = \pi(X_N)$ , questo determina in modo univoco l'intera strategia  $a_0, a_1, \dots, a_{N-1}$ , e il prezzo  $V = V_0(x)$  dell'opzione! È necessario, in realtà, verificare che quest'unica strategia possibile sia effettivamente autofinanziante, cioè che vale la seguente proposizione.

**Proposizione 3.27.** *La strategia  $a_n = a_n(X_n)$  e i corrispondenti valori del capitale  $V_n = V_n(X_n)$  dati da (3.53) e (3.54) soddisfano le relazioni*

$$0 \leq a_n X_n \leq V_n$$

per ogni  $0 \leq n < N$ .

*Dimostrazione.* Si noti che le equazioni (3.53) e (3.54) sono formule ricorsive le cui soluzioni sono funzioni di una variabile  $a_n(x), V_n(x)$ :

$$a_n(x) = \frac{V_{n+1}(Cx) - V_{n+1}(cx)}{(C-c)x} \quad (3.55)$$

$$V_n(x) = \frac{1-c}{C-c}V_{n+1}(Cx) + \frac{C-1}{C-c}V_{n+1}(cx). \quad (3.56)$$

con condizione "terminale"  $V_N(x) = \pi(x)$ . Vogliamo dimostrare che, per ogni  $0 \leq n < N$  e ogni  $x > 0$ , si ha

$$0 \leq xa_n(x) \leq V_n(x). \quad (3.57)$$

Anzitutto notiamo che  $V_N(x) = \pi(x)$  è una funzione non negativa e crescente di  $x > 0$ . È facile vedere usando (3.56), per induzione "all'indietro" su  $n$  che tali proprietà sono trasmesse a tutte le  $V_n(x)$  per  $n \leq N-1$ . In particolare, essendo  $Cx > cx$  per ogni  $x > 0$ , questo implica  $a_n(x) \geq 0$  per ogni  $0 \leq n \leq N-1$ . Resta pertanto da dimostrare la disuguaglianza  $xa_n(x) \leq V_n(x)$  che, per (3.55) e (3.56), equivale a

$$\frac{V_{n+1}(Cx) - V_{n+1}(cx)}{C-c} \leq \frac{1-c}{C-c}V_{n+1}(Cx) + \frac{C-1}{C-c}V_{n+1}(cx) \iff V_{n+1}(Cx) \leq \frac{C}{c}V_{n+1}(cx).$$

Quest'ultima disuguaglianza è verificata se mostriamo che, per ogni  $0 \leq n \leq N$ , ogni  $x > 0$  e  $\rho > 1$

$$V_n(\rho x) \leq \rho V_n(x). \quad (3.58)$$

Tale disuguaglianza è verificata per  $n = N$  (verificarlo distinguendo i tre casi  $x \geq a$ ,  $x < a$ ,  $\rho x \geq a$ ,  $\rho x < a$ ). Inoltre, di nuovo per induzione all'indietro, la proprietà (3.58) si trasmette alle altre  $V_n$ .

Le equazioni ricorsive (3.51) e (3.52) determinano quindi tanto il prezzo dell'opzione, come funzione del valore del titolo al momento in cui l'opzione viene acquistata, quanto la strategia autofinanziante che copre l'opzione. Va notato che tanto il prezzo quanto la strategia *non* dipendono dalla distribuzione di  $(X_1, X_2, \dots, X_N)$ , ma solo dall'ipotesi fatta che il valore  $X_{n+1}$  del titolo al giorno  $n+1$  possa essere o  $cX_n$  oppure  $CX_n$ . Il valore  $V(x)$  del prezzo ammette però un'interpretazione probabilistica che si può estendere a modelli di mercato assai più complessi, e che ha grande rilevanza sia teorica che applicativa. Ricordiamo che, come osservato in precedenza,  $X_n = R_n R_{n-1} \dots R_1 x$ , dove  $R_n = X_n / X_{n-1}$  è la variabile casuale che "aggiorna" il valore del titolo al giorno  $n$ .

**Teorema 3.5.** *Siano  $Q_1, Q_2, \dots, Q_N$  variabili casuali indipendenti, che assumono soltanto i valori  $c$  e  $C$ , e di media  $E(Q_n) = 1$ . Allora la soluzione  $V_n(x)$  dell'equazione ricorsiva (3.56) è, per  $n \leq N-1$ ,*

$$V_n(x) = E[\pi(Q_{n+1} \dots Q_N x)].$$

*In particolare, il prezzo dell'opzione è*

$$V(x) = V_0(x) = E[\pi(Q_1 Q_2 \dots Q_N x)].$$

*Osservazione 3.13.* Posto  $Z_n = Q_1 Q_2 \dots Q_n x$ , il prezzo dell'opzione è dunque  $V = E[\pi(Z_N)]$ . La variabile casuale  $Z_n$  si può interpretare come il valore al giorno  $n$  di un titolo avente lo stesso valore iniziale  $x$  del titolo reale, ma che evolve in un mercato *fittizio* in cui le variabili casuali di "aggiornamento" sono indipendenti e hanno media 1 (*mercato neutrale*). Il prezzo dell'opzione è dato dal valor medio del costo finale a carico dell'istituto finanziario ( $\pi(Z_N)$ ) calcolato non rispetto al mercato "reale" ma rispetto al corrispondente mercato neutrale.

*Dimostrazione (Dimostrazione del Teorema 3.5).*

Si noti, anzitutto, che le richieste di assumere solo i valori  $c$  e  $C$  e di avere media 1 caratterizzano univocamente la distribuzione (comune) dell  $Q_n$ . Infatti dev'essere

$$cP(Q_n = c) + CP(Q_n = C) = 1$$

$$P(Q_n = c) + P(Q_n = C) = 1$$

da cui segue

$$P(Q_n = c) = \frac{C-1}{C-c}, \quad P(Q_n = C) = \frac{1-c}{C-c}.$$

Inoltre, usando l'indipendenza delle  $Q_n$ , posto  $W_n(x) := E[\pi(Q_{n+1} \cdots Q_N x)]$ , per  $n \leq N-2$  si ha

$$\begin{aligned}
W_n(x) &= E[\pi(Q_{n+1} \cdots Q_N x)] \\
&= \sum_{q_{n+1} \in \{c, C\}} \sum_{q_{n+2} \in \{c, C\}} \cdots \sum_{q_N \in \{c, C\}} \pi(q_{n+1} q_{n+2} \cdots q_N x) P(Q_{n+1} = q_{n+1}, Q_{n+2} = q_{n+2}, \dots, Q_N = q_N) \\
&= \sum_{q_{n+1} \in \{c, C\}} \sum_{q_{n+2} \in \{c, C\}} \cdots \sum_{q_N \in \{c, C\}} \pi(q_{n+1} q_{n+2} \cdots q_N x) P(Q_{n+1} = q_{n+1}) P(Q_{n+2} = q_{n+2}) \cdots P(Q_N = q_N) \\
&= \sum_{q_{n+1} \in \{c, C\}} P(Q_{n+1} = q_{n+1}) \sum_{q_{n+2} \in \{c, C\}} \cdots \sum_{q_N \in \{c, C\}} \pi(q_{n+1} q_{n+2} \cdots q_N x) P(Q_{n+2} = q_{n+2}) \cdots P(Q_N = q_N) \\
&= \frac{C-1}{C-c} \sum_{q_{n+2} \in \{c, C\}} \cdots \sum_{q_N \in \{c, C\}} \pi(q_{n+2} \cdots q_N cx) P(Q_{n+2} = q_{n+2}) \cdots P(Q_N = q_N) \\
&\quad + \frac{1-c}{C-c} \sum_{q_{n+2} \in \{c, C\}} \cdots \sum_{q_N \in \{c, C\}} \pi(q_{n+2} \cdots q_N Cx) P(Q_{n+2} = q_{n+2}) \cdots P(Q_N = q_N) \\
&= \frac{C-1}{C-c} W_{n+1}(cx) + \frac{1-c}{C-c} W_{n+1}(Cx).
\end{aligned}$$

Lo stesso argomento per  $n = N-1$  fornisce

$$W_{N-1}(x) = \frac{C-1}{C-c} \pi(cx) + \frac{1-c}{C-c} \pi(Cx).$$

Dunque  $W_n(x)$  risolve l'equazione ricorsiva (3.56), e pertanto  $W_n(x) = V_n(x)$ .