CAPITOLO 4

VARIABILI ALEATORIE FINITE

4.1. Variabili aleatorie

In alcuni problemi si è portati a considerare funzioni definite su uno spazio di probabilità .

Esempio 25. *Il numero di teste in* 100 *lanci di una moneta può essere visto come una funzione* $T_{100} : S \to \mathbb{R}$ *con* $S = \{0,1\}^{100}$ *che indica l'insieme dei possibili risultati del lancio di* 100 *monete, indicando testa con* 1 *e croce con* 0, *definita da*

$$T_{100}(a_1, \ldots, a_{100}) = \sum_{i=1}^{100} a_i;$$

ove su S *si considera la probabilità uniforme.*

Per distinguerla dalla teoria delle funzioni e per ricordare che stiamo parlando di fenomeni casuali, queste funzioni vengono chiamate in modo diverso:

Definizione 5. *Dato uno spazio di probabilità finito* (S, P) *ogni funzione* $X : S \to \mathbb{R}$ *è detta* **variabile aleatoria**.

Le variabili aleatorie costituiscono una generalizzazione del concetto di evento, infatti per ogni evento A in uno spazio di probabilità (S, P) la funzione indicatrice di $\mathbb{I}_A$ di A definita da

$$\mathbb{I}_A(s) = \begin{cases} 1 & \text{se } s \in A \\ 0 & \text{se } s \notin A \end{cases}$$

è una variabile aleatoria.

Spesso delle variabili aleatorie ci interessano i valori che esse possono assumere e le relative probabilità : denotiamo con $S_X = X(S) \subseteq \mathbb{R}$ il codominio di una variabile aleatoria X e per ogni $B \subseteq S_X$ con

$$\begin{aligned} P_X(B) &:= P(s \in S : X(s) \in B) \\ &= P(X^{-1}(B)) \\ &=: P(X \in B) \end{aligned}$$

la probabilità che X assuma valori in B. Conseguentemente, $P_X(x) = P_X(\{x\})$ indica la probabilità che X assuma il valore $x \in S_X$.

Lemma 5. *Per ogni variabile aleatoria finita* X, *la coppia* (S_X, P_X) *è uno spazio di probabilità finito detto* **distribuzione di** X.

Dimostrazione. Si ha:
(i) $P_X(S_X) = P(X^{-1}S_X) = P(S) = 1$
(ii) Per ogni $B \subseteq S_X$, $0 \leq P_X(B) \leq P(S) = 1$
(iii) Per ogni $B, C \subseteq S_X$, $B \cap C = \emptyset$,

$$\begin{aligned} P_X(B \cup C) &= P(X^{-1}(B \cup C)) \\ &= P(X^{-1}(B) \cup X^{-1}(C)) \\ &= P(X^{-1}(B)) + P(X^{-1}(C)) \\ &= P(B) + P(C) \end{aligned}$$

in cui la terza uguaglianza deriva, come sempre, dal fatto che gli eventi sono disgiunti e che quindi sono disgiunte le immagini inverse.

Infine, $|S_X| \leq |S| < \infty$. □

Osservazione 6. *Si noti che P_X risulta essere una probabilità proprio perchè utilizziamo X^{-1} per trasportare P in P_X. Se $Z : T \to S$ fosse una fuzione a valori in uno spazio di probabilità (S, P) il trasporto nella direzione opposta dato da $P \circ Z$ non dà in generale uno spazio di probabilità perchè le immagini di insiemi disgiunti non è detto che siano disgiunte.*

Poichè (S_X, P_X) è uno spazio di probabilità gran parte delle informazioni su X possono essere dedotte direttamente dalla sua distribuzione, tant'è che se due variabili aleatorie hanno la stessa distribuzione condividono gran parte delle proprietà . Per cui si pone:

Definizione 6. *Si dice che due variabili aleatorie X definita su (S_1, P_1) ed Y definita su (S_2, P_2) sono uguali in distribuzione, e si denota con $X =^d Y$, se $(S_X, P_X) = (S_Y, P_Y)$*

Da questo punto di vista sembra perfino che non sia necessario introdurre le variabili aleatorie, in quanto il loro studio si riconduce a quello di spazi di probabilità . Tuttavia, solo 'gran parte' delle proprietà dipendono dalla distributione, non tutte. In particolare, in qualunque modello di fenomeni reali si è interessati al valore assunto da una variabile aleatoria e non solo alla sua distribuzione; in questa direzione si osservi che variabili aleatorie uguali in distribuzione non sono necessariamente uguali, anzi possono essere sempre diverse:

Esempio 26. *Nel lancio di una moneta, la funzione indicatrice di testa $\mathbb{I}_T$ e la funzione indicatrice di croce $\mathbb{I}_C$ soddisfano $\mathbb{I}_T =^d \mathbb{I}_C$ ma $\mathbb{I}_T(s) \neq \mathbb{I}_C(s)$ per ogni $s \in S$; in altre parole, i guadagni di chi punta su testa sono opposti a quelli di chi punta su croce*

Come accaduto nell'ultimo esempio, capita di considerare due o più variabili aleatorie $X_1, X_2 \ldots, X_n$ definite su uno stesso spazio di probabilità . Anche ogni combinazione $\phi(X_1, X_2, \ldots, X_n)$ di queste determinata da una funzione $\phi : B \to \mathbb{R}$, con $B \subseteq \mathbb{R}^n$ tale che $S_{X_1} \times$

$S_{X_2} \times \cdots \times S_{X_n} \subseteq B$, è una variabile aleatoria definita da

$$\phi(X_1, X_2, \ldots, X_n)(s) = \phi(X_1(s), X_2(s), \ldots, X_n(s)).$$

Così $X_1 + X_2$, $4X_1$ e X_1^2 sono variabili aleatorie sullo stesso spazio di X_1 e X_2.

La determinazione della distribuzione di

$$Y = \phi(X_1(s), X_2(s), \ldots, X_n(s))$$

può essere laboriosa, come si intravede dall'esempio seguente:

ESEMPIO 27. *Se X_1 e X_2 sono i risultati del lancio di due dadi e $X_1 + X_2$ è la loro somma, con una certa pazienza si puo' ricavare la distribuzione di $X_1 + X_2$ dalle probabilità dei risultati dei lanci dei due dadi. In particolare, $S_{X_1+X_2} = \{2, 3, \ldots, 12\}$ e*

$$P_{X_1+X_2}(s) = \begin{cases} \frac{s-1}{36} & \textit{se } s \leq 6 \\ \frac{12-s+1}{36} & \textit{se } s \geq 7 \end{cases}$$

4.2. Valore atteso

Introduciamo ora una quantità , utilizzabile per l'analisi di variabili alatorie, che può essere giustificata in vari modi.

ESEMPIO 28. *Storicamente fu descritta per la prima volta relativamente al gioco dei punti, in cui due giocatori A e B si disputano 24 gettoni a testa o croce: vince il primo che arriva a 2 vittorie. Naturalmente esiste un modo equo di dividersi i gettoni senza giocare: 12 ognuno. Se però A, che scommette su testa, ha vinto la prima partita, come ci si può dividere la posta senza giocare ulteriormente? Se si decidesse di fermarsi dopo un'ulteriore partita allora se in questa esce ancora testa A prende 24 gettoni, e se esce croce ci si puo' dividere 12 gettoni a testa; quindi dopo una vittoria di A è ragionevole dare i 12 ad A e dividersi gli altri 6 ciascuno, per cui A ha 18 gettoni. Si nota però che $18 = \frac{1}{2}24 + \frac{1}{2}12$. Se invece si decidesse di finire il gioco la probabilità di A di vincere tutta la posta dato che è uscita una testa sarebbe $P(T \cup CT) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$ e si nota che di nuovo $18 = \frac{3}{4} \times 24$.*

ESEMPIO 29. *Supponiamo di scommettere sul risultato di un dado, vincendo 3 se esce il 6 ed altrimenti perdendo 1. In ogni partita la nostra vincita sarà quindi una variabile aleatoria X tale che $X(6) = 3$ e $X(i) = -1$ per ogni $i = 1, \ldots, 5$. Dopo 60 partite ci aspettiamo di aver vinto circa 10 partite e perse 50, con una vincita totale di -20. Si noti che*

$$-20 = 30 - 50 = 3 \times 10 - 1 \times 50 = 3\frac{60}{6} - 1\frac{60 \times 5}{6} = 60(3\frac{1}{6} - 1\frac{5}{6})$$

.

In questi ed in molti altri esempi compare quindi la quantità somma dei valori vinti moltiplicati per la probabilità di vincerli e questa quantità può essere quindi un modo di valutare l'esito di una variabile aleatoria. Per cui si pone:

DEFINIZIONE 7. *Data una variabile aleatoria X definita su uno spazio di probabilità (S, P) si dice* **valore atteso** *o* **speranza matematica** *o aspettazione o valor medio di X il valore*

$$E(X) = M(X) = \sum_{s \in S} X(s)P(s). \tag{4.12}$$

Il valore atteso dipende solo dalla distribuzione di X:

LEMMA 6.

$$E(X) = \sum_{x \in S_X} xP_X(x)$$

DIMOSTRAZIONE.

$$\begin{aligned} E(X) &= \sum_{s \in S} X(s)P(s) \\ &= \sum_{x \in S_X} \sum_{s \in S: X(s)=x} xP(s) \\ &= \sum_{x \in S_X} xP_X(x) \end{aligned}$$

□

ESEMPIO 30. *Il valore atteso del risultato X del lancio di un dado è :*

$$E(X) = \sum_{i=1}^{6} i\frac{1}{6} = 3,5.$$

L'esempio mostra che non sempre il valore atteso è un valore che può essere assunto dalla variabile aleatoria e mostra quindi che l'idea che rappresenti il valore 'che ci aspettiamo' è soltanto approssimativa e va specificata nei sensi che abbiamo indicato in precedenza.

ESEMPIO 31. *Il valore atteso di una variabile aleatoria $T_n \approx B(n, p)$ è :*

$$E(T_n) = \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} = np.$$

Alcune proprietà elementari del valore atteso sono:

LEMMA 7. *Se $X_1, X_2, \dots, X_n$ sono variabili aleatorie su uno spazio di probabilità (S, P) e $a_1, a_2, \dots, a_n \in \mathbb{R}$ sono costanti si ha che:*
(i) $E(a) = a$
(ii) $E(a_1X_1 + a_2X_2) = a_1E(X_1) + a_2E(X_2)$ *e pertanto*
(iii) $E(\sum_{i=1}^{n} a_iX_i) = \sum_{i=1}^{n} a_iE(X_i)$

Dimostrazione. (i) $E(a) = \sum_{s\in S} aP(s) = a$.
(ii)

$$\begin{aligned} E(a_1X_1 + a_2X_2) &= \sum_{s\in S}(a_1X_1 + a_2X_2)(s)P(s) \\ &= \sum_{s\in S}(a_1X_1(s) + a_2X_2(s))P(s) \\ &= \sum_{s\in S} a_1X_1(s)P(s) + \sum_{s\in S} a_2X_2(s)P(s) \\ &= a_1E(X_1) + a_2E(X_2). \end{aligned}$$

(iii) segue per induzione su n □

Esprimere una variabile aleatoria come somma di altre variabili può quindi condurre ad una drastica semplificazione del calcolo del valore atteso, come si vede dall'esempio seguente.

Esempio 32. *Se $T_{1000} \approx B(1000, 1/2)$ è il numero di teste in 1000 lanci di una moneta allora $T_{1000} = \sum_{i=1}^{1000} X_i$ con X_i la funzione indicatrice di testa all'i-simo lancio. Si ha che $E(X_i) = 1/2$ per ogni i e*

$$E(T_n) = E(\sum_{i=1}^{1000} X_i) = \sum_{i=1}^{1000} E(X_i) = 500.$$

Pertanto il valore atteso riproduce in questo caso correttamente l'idea di quello che 'ci aspettiamo', così come era stato assunto nelle discussioni iniziali ed anche utilizzato nel teorema di De Moivre-Laplace. Generalizzando si ottiene che se $T_n \approx B(n,p)$ allora $T_n = \sum_{i=1}^{n} X_i$ con $X_i \approx B(1,p)$. Per cui $E(X_i) = 1p - 0(1-p) = p$ e $E(T_n) = np$.

In questo calcolo la proprietà di indipendenza della distribuzione di Bernoulli non è stata utilizzata ed infatti il Lemma vale in generale per tutte le variabili aleatorie.

Esempio 33. *Se T_n è il totale di n estrazioni dalla tombola senza reinserimento allora $T_n = \sum_{i=1}^{n} X_i$, con X_i i-simo numero estratto. Poichè*

$$E(X_i) = \sum_{j=1}^{90} \frac{j}{90} = \frac{91}{2}$$

per ogni $i \leq 90$, e $E(X_i) = 0$ per ogni $i > 90$, si ha

$$E(T_n) = \begin{cases} n\frac{91}{2} & se\ n \leq 90 \\ 90\frac{91}{2} & se\ n \geq 90. \end{cases}$$

Si noti il caso $n = 90$, in cui tutte le palline sono estratte e $T_{90} \equiv 90\frac{91}{2}$ con probabilità uno.

Ci sono altri modi per indicare un 'valore medio' di una variabile aleatoria, utili in particolari contesti. Tra questi vi è il centro o i centri delle probabilità :

DEFINIZIONE 8. *Si dice* **mediana** *di una variabile aleatoria* X *definita su* (S, P) *ogni valore* $m(X)$ *tale che*

$$P(X \leq m(X)) \geq \frac{1}{2} \text{ e } P(X \geq m(X)) \geq \frac{1}{2}.$$

ESEMPIO 34. *Tutti i valori in* $[3, 4]$ *sono mediane del risultato del lancio di un dado.*

ESEMPIO 35. *La mediana del numero* T_n *di teste in* 100 *lanci di una moneta è* 50*, come si vede per simmetria, mentre per* 101 *lanci è un qualunque numero in* $[50, 51]$

Come si è visto dagli esempi, la mediana non è necessariamente unica e può coincidere o meno con il valore atteso.

DEFINIZIONE 9. *Si dice* **moda** *di una variabile aleatoria* X *ogni valore* $mo(X)$ *tale che per ogni* $x \in S_X$

$$P(X = mo(X)) \geq P(X = x).$$

E' moda di una variabile aleatoria $T_n \approx B(n, p)$ ogni valore in $[\lfloor p(n+1) \rfloor - 1, \lfloor p(n+1) \rfloor] \cap \mathbb{N}$ mentre tutti i valori possibili sono moda per le variabili aleatorie uniformi. La moda di una variabile aleatoria è utile solo in alcune situazioni: ad esempio, se si è obbligati a scommettere su uno solo dei risultati possibili di un esperimento casuale la moda permette di massimizzare le probabilità di vittoria.

Quando si osservano dei dati $\{x_1, \ldots, x_n\}$ dopo aver condotto un esperimento casuale è possibile definire lo spazio di probabilità $(S_{\mathbb{O}}, P_{\mathbb{O}})$, che chiameremo empirico o campionario, in cui $S_{\mathbb{O}}$ è costituito dai valori osservati e $P_{\mathbb{O}}$ è determinato dalla frequenze, ossia

$$P_{\mathbb{O}}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{x_i = x}.$$

In tale spazio il valore atteso della variabile aleatoria identità I è la media empirica dei dati:

$$\begin{aligned} \mu = E(I) &= \sum_{x \in \{x_1, \ldots, x_n\}} x P_{\mathbb{O}}(x) \\ &= \sum_{x \in \{x_1, \ldots, x_n\}} x \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{x_i = x} \\ &= \frac{1}{n} \sum_{i=1}^{n} x_i. \end{aligned}$$

La mediana di I è la mediana empirica $m = m(I)$ tale che $|\{i : x_i \leq m\}| \geq n/2$ e $|\{i : x_i \geq m\}| \geq n/2$. Si noti che le mediane empiriche si possono determinare ordinando (debolmente) le osservazioni x_i, che vengono in questo caso denotate con $\{x_{(1)}, \ldots, x_{(n)}\}$ con $x_{(1)} \leq \cdots \leq x_{(n)}$, e scegliendo poi un valore nell'intervallo (banale se n è dispari) $[x_{(\lfloor \frac{n+1}{2} \rfloor)}, x_{(\lceil \frac{n+1}{2} \rceil)}]$. La moda di I è la moda dei dati, ossia ogni valore $mo(I)$ che massimizza la frequenza di osservazione.

Esempio 36. *Con dati* $5.0, 3.5, 2.3, 5.0, 1.2, 1.6$ *risultano:* $\mu = 3.1, m \in [2.3, 3.5], mo = 5.0$.

Prima di condurre un esperimento è possibile valutare a priori le quantità empiriche appena descritte considerandole come funzioni di variabili aleatorie $X_1, \ldots, X_n$. In particolare la media empirica delle variabili aleatorie diviene

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

di cui si discuteranno varie proprietà nel seguito.

Per concludere la discussione sul valore atteso osserviamo che data una variabile aleatoria X ed una funzione $\Phi : S_X \to \mathbb{R}$, ed il risultato seguente sul cambiamento di variabili permette di semplificare il calcolo del valore atteso della variabile aleatoria $\Phi(X)$.

Teorema 6. *Data una variabile aleatoria X su (S, P) e $\Phi : B \to \mathbb{R}$ con $S_X \subseteq B$ si ha*

$$\begin{aligned} E(\Phi(X)) &= \sum_{\phi \in S_{\Phi(X)}} \phi P(\Phi(X) = \phi) \\ &= \sum_{x \in S_X} \Phi(x) P_X(x) \end{aligned}$$

Dimostrazione. la prima uguaglianza discende dal Lemma 6. Per la seconda si ha

$$\begin{aligned} E(\Phi(X)) &= \sum_{s \in S} \Phi(X(s)) P(s) \\ &= \sum_{x \in S_X} \sum_{s \in S : X(s) = x} \Phi(x) P(s) \\ &= \sum_{x \in S_X} \Phi(x) P_X(x) \end{aligned}$$

□

La prima espressione di $E(\Phi(X))$ data nel teorema richiede la determinazione di $(S_{\Phi(X)}, P_{\Phi(X)})$, mentre per la seconda espressione basta la distribuzione di X che può essere utilizzata per tutte le Φ.

4.3. Indipendenza tra variabili aleatorie

E' stata sollevata nell'ultimo capitolo la questione della dipendenza tra variabili aleatorie. Per definire la mancanza di dipendenza estendiamo il concetto di indipendenza tra eventi. Per comprendere come realizzare questa estensione partiamo dall'esempio seguente.

ESEMPIO 37. *Se* $X_i = \mathbb{I}_{A_i}$, *con* A_i *eventi indipendenti in uno spazio di probabilità* (S, P), *allora*

$$\begin{aligned} P(\mathbb{I}_{A_1} = 1)P(\mathbb{I}_{A_1} = 1) &= P(A_1)P(A_2) \\ &= P(A_1 \cap A_2) \\ &= P(\mathbb{I}_{A_1}\mathbb{I}_{A_2} = 1) \\ &= P(\mathbb{I}_{A_1} = 1, \mathbb{I}_{A_2} = 1) \end{aligned} \tag{4.13}$$

ma abbiamo anche visto che

$$\begin{aligned} P(\mathbb{I}_{A_1} = 1)P(\mathbb{I}_{A_1} = 0) &= P(A_1)P(A_2^c) \\ &= P(A_1 \cap A_2^c) \\ &= P(\mathbb{I}_{A_1} = 1, \mathbb{I}_{A_2} = 0) \end{aligned} \tag{4.14}$$

e così via, di modo che

$$P(\mathbb{I}_{A_1} = i)P(\mathbb{I}_{A_1} = j) = P(\mathbb{I}_{A_1} = i, \mathbb{I}_{A_2} = j)$$

per ogni $i, j = 0, 1$.

Per definire l'indipendenza si può quindi generalizzare questa proprietà :

DEFINIZIONE 10. n *variabili aleatorie* $X_1, , \ldots, X_n$ *definite su uno spazio di probabilità* (S, P) *sono indipendenti se*

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i) \tag{4.15}$$

Per $n = 2$, quindi, X_1, X_2 sono indipendenti se $P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1)P(X_2 = x_2)$ per ogni possibile valore x_1, x_2. Si noti che nella definizione di indipendenza per variabili aleatorie non è stata richiesta la fattorizzazione delle probabilità per sottofamiglie: mostreremo tra poco che questa proprietà è conseguenza della definizione data.

ESEMPIO 38. *Due estrazioni successive* X_1 *e* X_2 *dalla tombola sono indipendenti se effettuate con reinserimento e dipendenti se non c'è reinserimento; infatti, nel primo caso* $P(X_1 = k, X_2 = m) = \frac{1}{90^2} = P(X_1 = k)P(X_2 = m)$ *per ogni* $k, m = 1, \ldots, 90$, *mentre nel secondo caso* $P(X_1 = 1, X_2 = 1) = 0 \neq \frac{1}{90^2} = P(X_1 = 1)P(X_2 = 1)$.

LEMMA 8. *Se* $X_1, \ldots, X_n$ *sono variabili aleatorie indipendenti allora*

(I) per ogni $A_1, \ldots, A_n$*,* $A_i \subseteq S_{X_i}$ *si ha*

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i)$$

(II) per ogni $J \subseteq \{1, \ldots, n\}$ *le variabili aleatorie* $X_i, \quad i \in J$*, sono indipendenti,*
(III) per ogni $A_1, \ldots, A_n$*,* $A_i \subseteq S_{X_i}$*, gli eventi* $(X_i \in A_i) = \{s \in S : X_i(s) \in A_i\}$ *sono collettivamente indipendenti.*

Dimostrazione. Indichiamo con $\mathbf{x} = (x_1, \ldots, x_n)$ i vettori n dimensionali:

(I) Fissati $A_1, \ldots, A_n$ come nell'ipotesi si ha

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = P(\cup_{x_1 \in A_1, \ldots, x_n \in A_n} \{\mathbf{X} = \mathbf{x}\}).$$

Gli eventi $\{\mathbf{X} = \mathbf{x}\}$ sono disgiunti per definizione di funzione, quindi per l'indipendenza degli X_i

$$\begin{aligned} P(X_1 \in A_1, \ldots, X_n \in A_n) &= \sum_{x_1 \in A_1, \ldots, x_n \in A_n} P(\mathbf{X} = \mathbf{x}) \\ &= \sum_{x_1 \in A_1, \ldots, x_n \in A_n} \prod_{i=1}^{n} P(X_i = x_i) \\ &= \prod_{i=1}^{n} \sum_{x_1 \in A_1, \ldots, x_n \in A_n} P(X_i = x_i) \\ &= \prod_{i=1}^{n} P(X_i \in A_i) \end{aligned}$$

(II) segue da (I) con

$$A_i = \begin{cases} x_i & \text{se } i \in J \\ S_{X_i} & \text{se } i \notin J \end{cases}$$

essendo

$$\begin{aligned} P(X_i = x_i \text{ se } i \in J) &= P(X_i = x_i \text{ se } i \in J \text{ e } X_j \in S_{X_j} \text{ se } j \notin J) \\ &= \prod_{i \in J} P(X_i = x_i) \prod_{j \notin J} P(X_j \in S_{X_j}) \\ &= \prod_{i \in J} P(X_i = x_i) \end{aligned}$$

(III) segue da (II) prendendo $X_i = \mathbb{I}_{A_i}$. □

Quindi la richiesta di fattorizzazione delle probabilità per tutti i valori del codominio delle variabili aleatorie include già la fattorizzazione delle stesse espressioni per sottinsiemi di funzioni.

Vediamo ora che funzioni di variabili aleatorie indipendenti sono ancora indipendenti, nel senso che

TEOREMA 7. *Date variabili aleatorie indipendenti $X_1, \ldots, X_n$ definite su uno spazio di probabilità (S, P) e due funzioni $\phi : C \to \mathbb{R}$ e $\psi : D \to \mathbb{R}$ tali che $S_{X_1} \times S_{X_2} \times \cdots \times S_{X_k} \subseteq C$ e $S_{X_{k+1}} \times \cdots \times S_{X_n} \subseteq D$ si ha che $T = \phi(X_1, \ldots, X_k)$ e $Z = \psi(X_{k+1}, \ldots, X_n)$ sono indipendenti.*

DIMOSTRAZIONE. Poichè gli eventi nell'unione che segue sono, come al solito, disgiunti, ed essendo le variabili aleatorie indipendenti si ha:

$$
\begin{aligned}
&P(T = t, Z = z)\\
&\quad = P(s : \phi(X_1(s), \ldots, X_k(s)) = t,\\
&\qquad\qquad \psi(X_{k+1}(s), \ldots, X_n(s)) = z)\\
&\quad = P\Big(\bigcup_{\substack{\boldsymbol{x}=(x_1,\ldots,x_n):\phi(x_1,\ldots,x_k)=t,\\ \psi(x_{k+1},\ldots,x_n)=z}} \{s : (X_1(s), \ldots, X_n(s)) = \boldsymbol{x}\}\\
&\quad = \sum_{\substack{\boldsymbol{x}=(x_1,\ldots,x_n):\phi(x_1,\ldots,x_k)=t,\\ \psi(x_{k+1},\ldots,x_n)=z}} P((X_1, \ldots, X_n) = \boldsymbol{x})\\
&\quad = \sum_{\substack{\boldsymbol{x}=(x_1,\ldots,x_n):\phi(x_1,\ldots,x_k)=t,\\ \psi(x_{k+1},\ldots,x_n)=z}} \prod_{i=1}^{n} P(X_i = x_i)\\
&\quad = \sum_{\substack{(x_1,\ldots,x_k):\\ \phi(x_1,\ldots,x_k)=t}} \prod_{i=1}^{k} P(X_i = x_i) \sum_{\substack{(x_{k+1},\ldots,x_n):\\ \psi(x_{k+1},\ldots,x_n)=z}} \prod_{i=k+1}^{n} P(X_i = x_i)\\
&\quad = \sum_{\substack{\boldsymbol{x}=(x_1,\ldots,x_k):\\ \phi(x_1,\ldots,x_k)=t}} P((X_1, \ldots, X_k) = \boldsymbol{x})\\
&\qquad\qquad \times \sum_{\substack{\boldsymbol{x'}=(x_{k+1},\ldots,x_n):\\ \psi(x_{k+1},\ldots,x_n)=z}} P((X_{k+1}, \ldots, X_n) = \boldsymbol{x'})\\
&\quad = P(T = t)P(Z = z).
\end{aligned}
$$

□

Questo risultato ha molte conseguenze interessanti, ad esempio la seguente:

ESEMPIO 39. *Giocando $k+1$ partite alla roulette, se X_i è il numero che esce nell'i-sima partita si ha che le variabili $X_1, \ldots, X_k, X_{k+1}$ sono indipendenti. Ogni strategia che, a k fissato, cerchi di determinare su quale numero scommettere alla $k+1$-sima partita osservando i risultati delle partite precedenti è equivalente ad una funzione $\phi(X_1, \ldots, X_k)$ e, per il teorema appena provato, X_{k+1} ne risulta quindi indipendente.*

Per cui il teorema porta a concludere che non esiste nessuna strategia che dalle prime k partite dia suggerimenti sulla successiva partita.

Si può ovviamente pensare una strategia che non fissi k a priori (tipo aspettare la prima uscita del 3), ma ne rimandiamo l'analisi in quanto richiede un modello con una infinità di possibili risultati, che tratteremo più avanti.

Una delle proprietà fondamentali delle variabili aleatorie indipendenti si riferisce al valore atteso del prodotto di variabili aleatorie.

TEOREMA 8. *Se X_i, $i = 1, \ldots, n$ sono variabili aleatorie indipendenti su (S, P) allora*

$$E(\prod_{i=1}^{n} X_i) = \prod_{i=1}^{n} E(X_i).$$

DIMOSTRAZIONE. Cominciamo da $n = 2$:

$$\begin{aligned} E(X_1 X_2) &= \sum_{t \in S_{X_1 X_2}} t P(X_1 X_2 = t) \\ &= \sum_{t \in S_{X_1 X_2}} \sum_{x_1 \in S_{X_1}, x_2 \in X_2} x_1 x_2 P(X_1 = x_1, X_2 = x_2) \\ &= \sum_{x_1 \in S_{X_1}} x_1 P(X_1 = x_1) \sum_{t \in S_{X_1 X_2}} \sum_{x_2 \in S_{X_2} : x_1 x_2 = t} x_2 P(X_2 = x_2) \\ &= \sum_{x_1 \in S_{X_1}} x_1 P(X_1 = x_1) \sum_{x_2 \in S_{X_2}} x_2 P(X_2 = x_2) \\ &= E(X_1) E(X_2) \end{aligned}$$

poichè la seconda uguaglianza discende dall'additività della probabilità dell'unione disgiunta di insiemi.

Il risultato per n generico discende per induzione su n essendo $\prod_{i=1}^{n-1} X_i$ indipendente da X_n. □

DEFINIZIONE 11. *Variabili aleatorie $X_1, \ldots, X_n$ su uno spazio di probabilità che siano indipendenti e tali che $X_i =^d X_j$ per ogni $i, j = 1, \ldots, n$ si dicono* **indipendenti ed identicamente distribuite** *o* **i.i.d.**.