Pag. BETA 14,15,16

4. Fra probabilità e statistica

Le variabili aleatorie discrete e le distribuzioni di probabilità

Consideriamo i punteggi che si possono ottenere nel lancio di due dadi a sei facce e calcoliamo la probabilità che si verifichi ognuno di essi.
Esaminiamo la seguente tabella a doppia entrata.

▶ Tabella 1

SECONDO DADO \ PRIMO DADO	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

I casi possibili sono 36. La probabilità che il risultato sia 2 è $p(2) = \frac{1}{36}$, che sia 3 è $p(3) = \frac{2}{36} = \frac{1}{18}$, che sia 4 è $p(4) = \frac{3}{36} = \frac{1}{12}$ e così via.
Se indichiamo con X un generico valore che può essere assunto dal risultato, possiamo riassumere le probabilità che si ottengono con una tabella.

▶ Tabella 2

X	2	3	4	5	6	7	8	9	10	11	12
$p(X)$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

Notiamo che gli eventi che abbiamo considerato sono tutti quelli possibili; inoltre, essi si escludono a vicenda, quindi costituiscono una partizione dell'insieme universo U di tutti gli eventi.

X viene detta *variabile aleatoria* (o *casuale*) e, poiché può assumere un numero finito di valori, è chiamata *discreta*.

▶ La definizione si può estendere al caso di una *infinità numerabile* di valori $x_1, x_2, \ldots, x_n, \ldots$ Per esempio, se lanciamo più volte un dado e ci chiediamo qual è la probabilità che esca 4 per n volte consecutive, la variabile aleatoria ha gli infiniti valori 1, 2, 3, 4, 5, …

DEFINIZIONE

Variabile aleatoria discreta

Una variabile aleatoria discreta X è una variabile che può assumere i valori $x_1, x_2, \ldots, x_n$ corrispondenti a eventi aleatori $E_1, E_2, \ldots, E_n$, non impossibili, che si escludono a vicenda e tali che sicuramente uno di essi si verifichi.

Diciamo inoltre che la tabella appena compilata descrive la *distribuzione di probabilità* relativa alla variabile X.

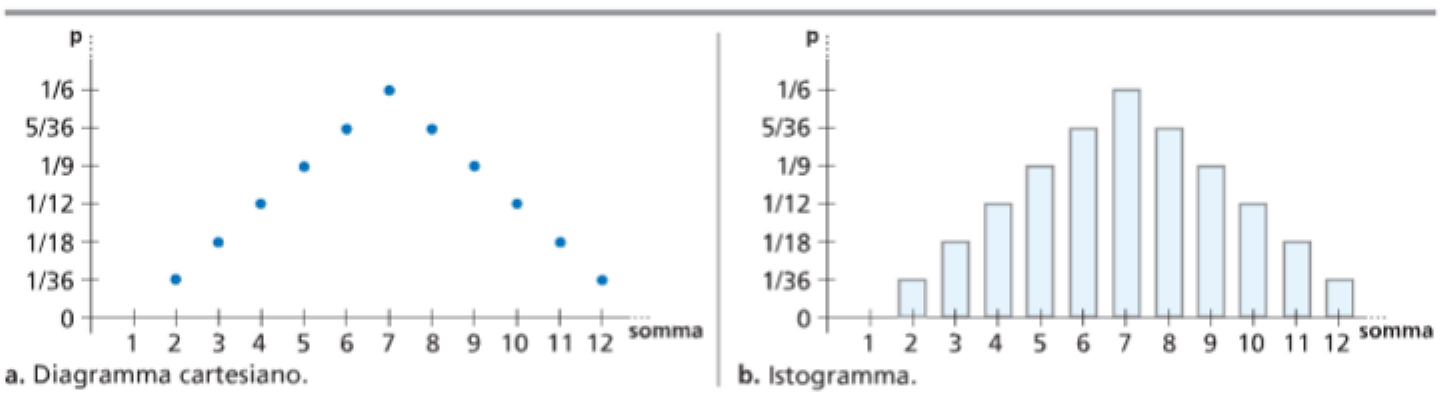
▶ Poiché gli eventi di una variabile aleatoria sono incompatibili e la loro unione coincide con l'insieme universo U, si ha:
$$p_1 + p_2 + \ldots + p_n = 1.$$

DEFINIZIONE

Distribuzione di probabilità

Data una variabile aleatoria discreta X, con valori $x_1, x_2, \ldots, x_n$, la successione delle probabilità $p_1, p_2, \ldots, p_n$ a essi associate si chiama distribuzione di probabilità della variabile X.

Per rappresentare una distribuzione di probabilità possiamo utilizzare, oltre che una tabella, un diagramma cartesiano o un istogramma. Per l'esempio precedente abbiamo le rappresentazioni della figura 2.



▲ Figura 2 Rappresentazioni grafiche della distribuzione di probabilità p della variabile X = «somma dei punti realizzati nel lancio di due dadi».

La legge empirica del caso

Consideriamo ancora il lancio, ripetuto più volte, di due dadi. Possiamo studiare i risultati ottenuti da un punto di vista statistico, costruendo una tabella delle frequenze. La tabella che riportiamo è riferita a 10, 100 e 1000 lanci. Abbiamo indicato anche le frequenze relative espresse mediante numeri decimali e le probabilità della tabella 2, approssimate alla terza cifra decimale.

La legge empirica del caso si chiama anche:
LEGGE DEI GRANDI NUMERI

LANCI DI DUE DADI							
SOMMA DEI PUNTI	10 LANCI F	f	100 LANCI F	f	1000 LANCI F	f	PROBABILITÀ
2	0	0	1	0,01	27	0,027	0,028
3	1	0,1	4	0,04	51	0,051	0,056
4	0	0	10	0,10	81	0,081	0,083
5	1	0,1	7	0,07	110	0,110	0,111
6	3	0,3	16	0,16	140	0,140	0,139
7	2	0,2	19	0,19	171	0,171	0,167
8	1	0,1	5	0,05	139	0,139	0,139
9	1	0,1	12	0,12	111	0,111	0,111
10	0	0	15	0,15	89	0,089	0,083
11	1	0,1	7	0,07	53	0,053	0,056
12	0	0	4	0,04	28	0,028	0,028

▶ Nella tabella, F rappresenta la frequenza, f la frequenza relativa, ossia $\frac{F}{N}$, dove N è il numero dei lanci. Abbiamo scritto le probabilità della tabella 2 in forma decimale, per poter fare meglio il confronto con le frequenze relative.

Osserviamo che, aumentando il numero dei lanci, le frequenze relative tendono ad avvicinarsi ai valori delle probabilità.

In generale si può affermare che, in un grande numero di prove, la frequenza relativa di un evento aleatorio è molto vicina alla probabilità dell'evento. La differenza fra i due valori tende a diminuire all'aumentare del numero di prove che si eseguono.

Questa affermazione prende il nome di **legge empirica del caso**.

La probabilità statistica

Abbiamo definito la probabilità come quoziente fra numero di casi favorevoli e numero di casi possibili. La probabilità così definita viene anche detta probabilità **«a priori»**, perché è calcolata senza che vengano eseguite prove concrete.
Tuttavia ci sono eventi aleatori per i quali non è possibile calcolare la probabilità in questo modo. Per esempio, non è possibile calcolare a priori la probabilità che esca 2 in un dado truccato.
In casi come questi viene in aiuto la legge empirica del caso. Accettiamo infatti come probabilità, che chiamiamo **probabilità statistica**, la frequenza relativa di un evento che si ottiene da un numero abbastanza elevato di prove, tutte ripetute nelle stesse condizioni. Il valore della probabilità statistica è un valore **a posteriori**.

ESEMPIO Il metodo statistico viene utilizzato nel campo delle assicurazioni per calcolare la probabilità che ha una persona di essere in vita o di morire entro un certo periodo.
La tabella a lato, basata su un'ipotetica popolazione di 100 000 nati vivi, riporta il numero di coloro che sono vivi alle diverse età. Dalla tabella si possono calcolare sia le probabilità di morte sia quelle di vita.