5. Variabili casuali e distribuzioni di probabilità

Notazioni
X = variabile aleatoria casuale o aleatoria (v.a.)
p_i = probabilità dell'evento "$X = x_i$"
$F(x)$ = funzione di ripartizione
$E(X)$ (o μ) = valore medio di una v.a.
X' = variabile casuale scarto dalla media μ
$Var(X)$ (o σ^2) = varianza di una v.a.
σ = deviazione standard o scarto quadratico medio (s.q.m.)
Z = v.a. standardizzata
p_{ij} = funzione di probabilità congiunta
$f(x_i), g(y_j)$ = funzioni di probabilità marginali
$Cov(X, Y)$ = covarianza di X e Y
$\rho(X, Y)$ = indice di correlazione di X e Y
$\bar{X}_n$ = media campionaria

5.1 Variabili casuali

5.1.1 Introduzione

Nel precedente capitolo sono stati oggetto di studio gli eventi aleatori e, in base alla loro natura, si è visto quale impostazione è più idonea a esprimere una valutazione di probabilità degli eventi stessi e quali regole di calcolo applicare per calcolare le probabilità di eventi complessi.

Lo studio dei fenomeni aleatori può essere affrontato in modo completo mediante le *variabili casuali* o *aleatorie*.

Per ogni fenomeno aleatorio si considerano tutti gli esiti che si possono avere e ad ognuno di essi si associa la probabilità che ha di verificarsi, così come, in una distribuzione di frequenza, a ogni modalità è associata la sua frequenza. A seconda che si manifestino modalità discrete o continue, si hanno *distribuzioni di probabilità discrete o distribuzioni di probabilità continue*.

5.1.2 Variabili casuali discrete

Nel calcolo delle probabilità è fondamentale il concetto di variabile *aleatoria* o *casuale*. In molti fenomeni aleatori l'esito di un esperimento è una grandezza che assume valori variabili in modo casuale, cioè valori che si presentano con una determinata probabilità.

Lo studio di una variabile aleatoria (v.a.) si svolge come segue: anzitutto si rappresenta graficamente la v.a. per esaminarne l'andamento, successivamente si sintetizza la distribuzione con valori sintetici quali il valor medio, il valore modale, la varianza, lo scarto quadratico medio.

Esempi di variabili aleatorie sono i seguenti.

a) Il numero di persone che arrivano a una cassa di un supermercato.
b) Il numero di particelle emesse da una sostanza radioattiva.
c) Il valore di un titolo azionario fra una settimana.
d) Il guadagno o la perdita di un giocatore in n partite.
e) Il primo numero estratto nel gioco della tombola.
f) Il numero di autovetture che arrivano a un casello autostradale.
g) Il numero di lanci di una moneta, finché non si presenta testa.
h) La durata della vita di una data persona.

Per alcune variabili aleatorie i valori sono in numero finito, come negli esempi e), d), per altre sono un'infinità numerabile, come negli esempi a), b), f), g), per altre ancora sono tutti i valori reali appartenenti a un intervallo, come negli esempi c), h). Quest'ultimi due casi sono esempi di *variabili casuali continue*, mentre i precedenti – casi e), d), a), b), f), g) – sono esempi di *variabili casuali discrete*.

Si esaminino i seguenti ulteriori esempi.

1) Si lancia due volte una moneta regolare; "testa" si può presentare o nessuna volta (0), o una sola volta (1), o due volte (2).
2) Si lanciano due dadi; la somma dei punti è uno degli 11 valori: 2, 3, 4, … 12.

3) Un giocatore lancia un dado. Se si presenta un numero primo, egli vince quel numero di euro, ma se si presenta un numero non primo egli perde quel numero di euro.

In tutti e tre gli esempi considerati, prima di compiere l'esperimento non è possibile dire quale valore assumeranno le variabili aleatorie *numero di "teste"*, *somma dei punti* e *vincita* (positiva o negativa), però si sa che certamente ne assumeranno *uno e uno solo* di quelli possibili indicati. Ogni valore assunto dalla variabile *dipende* dal verificarsi di un evento aleatorio, sottoinsieme di uno spazio campionario S, ossia si associa uno specifico numero a ciascun esito dell'esperimento. Gli eventi costituiscono una *partizione* di S, perché *uno e uno solo* degli eventi si verificherà con una sua propria probabilità. Una siffatta associazione è detta *variabile aleatoria*; per la precisione si pone la seguente definizione.

Una ***variabile aleatoria*** *X, su uno spazio campionario S, è una trasformazione da S nell'insieme dei numeri reali ¬ tali che l'immagine inversa (o preimmagine) di ogni intervallo di ¬ sia un evento di S.*

Occorre sottolineare che se S è uno spazio discreto in cui ogni sottoinsieme è un evento, allora ogni funzione a valori reali definita su S è una variabile aleatoria. D'altra parte, si può mostrare che se S è non numerabile, allora certe funzioni a valori reali definite su S non sono variabili casuali.

Se X e Y sono v.a. definite sullo stesso spazio campionario S, allora $X + Y$, $X + k$, kX e XY (con $k \in$ ¬) sono funzioni su S definite da:

$$(X + Y)(s) = X(s) + Y(s) \qquad (kX)(s) = kX(s)$$
$$(X + k)(s) = X(s) + k \qquad (XY)(s) = X(s)\, Y(s)$$

per ogni $s \in S$.

Si può mostrare che anche queste sono variabili aleatorie. (Il che è banale nel caso in cui ogni sottoinsieme di S sia un evento.)

Per designare la probabilità degli eventi "X si trasforma in x_i" e "X si trasforma nell'intervallo $[a, b]$" impiegheremo la notazione abbreviata $P(X = x_i) = p_i$ e $P(a \leq X \leq b)$, ossia:

$$p_i = P(X = x_i) = P(\{s \in S \mid X(s) = x_i\})$$
$$P(a \leq X \leq b) = P(\{s \in S \mid a \leq X(s) \leq b\})$$

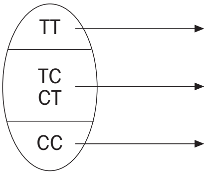
Esempio 1

Prendiamo in esame il precedente esempio 1.

Lo spazio campionario S è costituito da 4 eventi:

$$S = \{TT, TC, CT, CC\}$$

Al numero di teste 0 è associato l'evento {CC}, al numero di teste 1 è associato l'evento {TC, CT}, al numero di teste 2 è, infine, associato l'evento {TT}: questi eventi sono sottoinsiemi di S e ne costituiscono una partizione. Rappresentiamo la partizione determinata dalla v.a. "numero di teste" e le corrispondenti probabilità:



Valori	Probabilità
0	$\frac{1}{4}$
1	$\frac{2}{4}$
2	$\frac{1}{4}$

Per conoscere una variabile aleatoria non basta conoscere i valori che essa assume, ma occorre anche conoscere la probabilità con la quale ciascun valore si verifica.

Sia allora X una v.a. *discreta* su uno spazio campionario S con un insieme immagine finito:

$$X(S) = \{x_1, x_2, \ldots, x_n\}$$

$X(S)$ diventa uno spazio di probabilità definendo la probabilità di x_1 come $P(X = x_i)$, che scriviamo nella forma $p_i = f(x_i)$. Questa funzione f su $X(S)$, che cioè è definita da:

$$p_i = f(x_i) = P(X = x_i)$$

è detta **funzione di probabilità** o **distribuzione di probabilità di *X*** e viene di solito data nella forma di una tabella:

X	P
x_1	p_1
x_2	p_2
…	…
x_n	p_n

con $p_i \geq 0$ e $\sum_{i=1}^{n} p_i = 1$

o anche, graficamente, per mezzo di diagrammi e di istogrammi.

Indice

Cap.1 Dati e frequenze

Cap.2 Indici di posizione

Tabelle

Esercizi

Risposte agli esercizi