LE VARIABILI CASUALI

Sommario: 1. Generalità. - 2. La variabile casuale binomiale. - 3. La variabile casuale di Poisson. - 4. La variabile casuale normale. 5. La variabile casuale chi-quadrato. - 6. La variabile casuale *t* di Student. - 7. La variabile casuale *F* di Fisher-Snedecor. - 8. Teorema del limite centrale.

1. GENERALITÀ

Dato uno spazio campione S è possibile definire una funzione che associ ad ogni elemento di S un valore scelto in un opportuno insieme numerico A. Tale funzione viene detta **funzione aleatoria** (o **stocastica**) o anche **variabile casuale** (v.c.) o **stocastica**.

Nel seguito del capitolo preferiremo quest'ultima espressione, del resto molto diffusa. Posto $T = \{esce\ Testa\}$, $C = \{esce\ Croce\}$, se consideriamo l'evento $E = \{lancio\ di\ due\ monete\} = \{TT, CT, TC, CC\}$ possiamo definire la v.c. X nel modo seguente:

Tabella 1

Punti campione	*TT*	*CT*	*TC*	*CC*
X	2	1	1	0

È evidente che la v.c. è stata definita associando ad ogni punto campione il numero di teste (T) contenuto in esso, e che l'insieme A in questo caso è dato da $A = \{0, 1, 2\}$.

Se la v.c. può variare con continuità all'interno dell'insieme A viene detta v.c. **continua**, in caso contrario (come nell'esempio sopraesposto) è una v.c. **discreta**. Per tutti e due i tipi di v.c. è interessante definire una **funzione di probabilità**, ossia una regola che consenta di definire con quale probabilità una v.c. assuma un determinato valore.

Nel caso di v.c. discrete, detti:

$$x_1, x_2, x_3, \ldots, x_n$$

i valori che può assumere la v.c. X e conoscendo le probabilità:

$$P(X = x_k) \qquad k=1, 2, \ldots, n$$

è possibile definire la funzione:

$$P(X = x) = f\,(x)$$

che gode delle seguenti proprietà:

1. $f\,(x) \geq 0$
2. $\sum_{k=1}^{n} f\left(x_k\right) = 1$

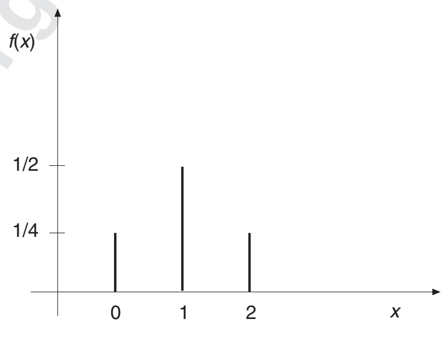
La rappresentazione grafica della funzione di probabilità è detta **grafico di probabilità**.

Esempio

Di seguito vengono illustrate la funzione e il grafico di probabilità relativi alla v.c. definita nella tabella seguente:

Tabella 2

x	0	1	2
***f*(*x*)**	1/4	1/2	1/4



La distribuzione di probabilità di una v.c. discreta resta pienamente determinata una volta definita la sua **funzione di distribuzione cumulativa**, detta anche **funzione di distribuzione**, ossia la funzione che permette di calcolare con quale probabilità la v.c. assume un valore minore o uguale a un determinato valore critico:

$$P(X \le x) = F(x)$$

La funzione di distribuzione può essere calcolata dalla funzione di probabilità eseguendo la somma dei singoli valori di probabilità:

$$F(x) = P(X \le x) = \sum_{u \le x} f(u)$$

Una trattazione analoga si ha per le v.c. continue per le quali si considera la probabilità che X cada in un ragionevole intervallo (a, b).

Possiamo allora definire, in analogia con quanto fatto per le v.c. discrete, due proprietà per la funzione di probabilità $f(x)$:

1. $f(x) \ge 0$
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$

Qual è la probabilità che x cada nell'intervallo (a, b)?

La probabilità che X cada nell'intervallo (a, b) è data da:

$$P(a < X < b) = \int_a^b f(x) dx$$

Data una v.c. X che può assumere n valori, il **valore medio** di X può essere definito come la somma dei prodotti tra i singoli valori assunti e la probabilità che la X ha di assumere tali valori.

In simboli, il valore medio di una v.c. discreta è:

$$E(X) = x_1 P(X = x_1) + \ldots + x_n P(X = x_n) = \sum_{i=1}^{n} x_i P(X = x_i)$$

Per una v.c. continua, il valore medio è:

$$E(X) = \mu = \int_{-\infty}^{+\infty} x f(x) dx$$

Esempio

Un giocatore d'azzardo vince 20 € e se una carta estratta da un mazzo è di cuori, ne vince 40 € e se la carta è di quadri, ne perde 30 € e se la carta è di fiori e non vince né perde se la carta è di picche. La variabile casuale *X* è data dalla vincita (o dalla perdita) e la sua funzione di probabilità è rappresentata nella tabella seguente:

Tabella 3

Carta	cuori	quadri	fiori	picche
x	20	40	–30	0
***f*(*x*)**	0,25	0,25	0,25	0,25

Il valore medio è dato da:

$$E(X) = 20 \cdot 0{,}25 + 40 \cdot 0{,}25 - 30 \cdot 0{,}25 + 0 \cdot 0{,}25 = 7{,}5$$

Ne consegue che il probabile guadagno del giocatore sarà di 7,5 €; in base a questo calcolo sarebbe anche possibile stabilire una misura dell'equità del gioco in base alle poste stabilite per giocare.

Detta μ la media di una v.c. è possibile definire una misura della variabilità detta **varianza**.

Tale quantità può essere espressa, per una v.c. discreta, nel modo seguente:

$$Var(X) = \sum_{i=1}^{n} (x_i - \mu)^2 p_i$$

Per una v.c. continua la varianza è:

$$Var(X) = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx$$

2. LA VARIABILE CASUALE BINOMIALE

La v.c. **binomiale** è caratterizzata da una distribuzione di probabilità detta anch'essa binomiale*.

I parametri che la descrivono sono la probabilità p di ottenere un successo in una prova, la probabilità $q = 1 - p$ di ottenere un insuccesso e n il numero di prove effettuate. Con questi parametri la probabilità che su n prove si abbiano esattamente x successi è data da:

$$f(x) = P(X = x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

dove si indica con X la v.c. {numero di successi} e con $x = 1, 2, \ldots, n$ le prove.

Il simbolo $\boldsymbol{n}$! viene detto ***n* fattoriale**, e rappresenta il prodotto di n numeri interi consecutivi a partire da 1.

Esempio

La probabilità di ottenere 4 teste in 10 lanci di una moneta non truccata è data da:

$$f(4) = P(X = 4) = \frac{10!}{4!6!} \cdot 0{,}5^4 \cdot 0{,}5^6 =$$

$$= \frac{3.628.800}{24 \cdot 720} \cdot 0{,}0625 \cdot 0{,}015625 \cong 0{,}205$$

La distribuzione binomiale è caratterizzata dai seguenti indici:

$$E(X) = np \qquad Var(X) = npq$$

3. LA VARIABILE CASUALE DI POISSON

La v.c. di Poisson, o **v.c. degli eventi rari**, assume rilievo quando si tratta di determinare il numero di volte in cui si verifica un evento casuale **in un dato intervallo di tempo/spazio**. Essa è la più adatta per descrivere i fenomeni in cui, su un grande numero n di prove, per ciascuna delle quali la probabilità di successo è piccola, si verificano mediamente λ successi. È molto utilizzata per studiare il numero di guasti, di clienti in arrivo, di auto in coda etc.

La distribuzione di Poisson si ottiene come **limite** della **distribuzione binomiale** assumendo $np = \lambda$ e $n \to \infty$.

La funzione di distribuzione di Poisson è data da:

$$P(X = x) = \lim_{\substack{n\to\infty \\ p\to 0}} \binom{n}{x} p^x (1-p)^{n-x} = \exp(-\lambda)\frac{\lambda^x}{x!}$$

Il valore medio e la varianza di una variabile di Poisson sono dati da:

$$E(X) = \lambda \qquad\qquad Var(X) = \lambda$$

4. LA VARIABILE CASUALE NORMALE

Tra le distribuzioni di probabilità continue un posto di preminenza spetta senz'altro alla v.c. **normale**. Di espressione matematica piuttosto complessa, la funzione di distribuzione dipende dai parametri σ e μ:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

con π =3,141592… ed exp è la funzione esponenziale.

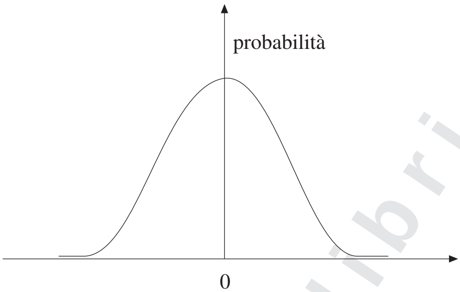
Se standardizziamo la v.c., sottraendo la media μ e dividendo la differenza per lo scarto quadratico medio σ l'espressione si semplifica, in quanto la variabile standardizzata:

$$z = \frac{x-\mu}{\sigma}$$

ha media nulla e scarto quadratico medio unitario. La funzione di probabilità diventa quindi:

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \qquad \text{con } z = \frac{x-\mu}{\sigma}$$

Il grafico della funzione di probabilità della variabile standardizzata è una curva molto conosciuta, detta **curva normale** o **Gaussiana** o **a campana**, presentata nella figura seguente:



Caratteristica di tale curva è che l'area compresa tra la stessa e l'asse delle ascisse, individua percentuali di probabilità ben definite. Nell'intervallo (–1, +1) si trova infatti circa il 68,27% della probabilità totale, nell'intervallo (–2, +2) circa il 95,45% e nell'intervallo (–3, +3) circa il 99,73%. Altri valori, riportati in apposite tabelle, risultano molto utili nella teoria della stima.

Si è accennato prima alla centralità della distribuzione normale nella teoria della probabilità. Tale centralità, dimostrata anche in diversi teoremi, è dovuta al fatto che fenomeni naturali manifestatisi in campioni numerosi tendono ad avere, al crescere del numero delle misurazioni, una distribuzione di frequenze che è bene approssimata dalla normale.

5. LA VARIABILE CASUALE CHI-QUADRATO

La **v.c. chi - quadrato** è una v.c. continua generata dalla **somma** di un numero g **di v.c. normali standardizzate e indipendenti al quadrato**, ossia:

$$X = \sum_{i=1}^{g} Z_i^2$$

Il parametro g indica i **gradi di libertà** della distribuzione che rappresentano il numero di unità di informazioni indipendenti in un campione. Essi sono pari alla numerosità campionaria cui vengono sottratti i vincoli noti, cioè i parametri noti della popolazione.

La funzione di densità della v.c. è data da:

$$f(x;g) = \frac{1}{2^{\frac{g}{2}}\Gamma\left(\frac{g}{2}\right)}\exp\left(-\frac{x}{2}\right)x^{\frac{g}{2}-1},\ x \geq 0$$

Al crescere dei gradi di libertà la v.c. chi-quadrato, o v.c. χ^2, tende ad una distribuzione normale.

6. LA VARIABILE CASUALE *t* DI STUDENT

Un'altra v.c. fondamentale in statistica è la **v.c. *t* di Student** data dal **rapporto tra una v.c. normale standardizzata** e la **radice quadrata di una v.c. indipendente** dalla prima **con distribuzione chi-quadrato, rapportata ai propri gradi di libertà**, in simboli:

$$X = \frac{Z}{\sqrt{\frac{Y}{g}}}$$

con $Z \sim N(0, 1)$ e $Y \sim \chi^2_g$.

La media e la varianza sono pari, rispettivamente, a:

$$E(X) = 0 \text{ per } g > 1;\ Var(X) = \frac{g}{g-2} \text{ per } g > 2$$

7. LA VARIABILE CASUALE *F* DI FISHER-SNEDECOR

La **v.c. *F* di Fisher-Snedecor** è una v.c. continua definita come il **rapporto di due v.c. chi - quadrato indipendenti** tra loro, **divise per i rispettivi gradi di libertà**. In simboli, siano $X_1 \sim \chi^2_{g_1}$ e

$X_2 \sim \chi^2_{g_2}$ le due v.c. chi-quadrato, la v.c. F di Fisher-Snedecor, indicata con $F \sim F_{g_1,g_2}$, è:

$$F_{g_1,g_2} = \frac{X_1/g_1}{X_2/g_2}$$

in cui, i due parametri g_1 e g_2 sono i gradi di libertà, rispettivamente, del numeratore e del denominatore.

Il valore medio e la varianza della v.c. sono, rispettivamente:

$$E(X) = \frac{g_2}{g_2 - 2}, \text{ per } g_2 > 2; \quad Var(X) = \frac{2(g_2)^2(g_1 + g_2 - 2)}{g_1(g_1 - 2)^2(g_2 - 4)}, \text{ per } g_2 > 4$$

8. TEOREMA DEL LIMITE CENTRALE

Trattasi di un gruppo di risultati teorici che mostrano, sotto una serie condizioni, la **convergenza alla v.c. normale della somma** (e quindi della media) di **v.c. di qualsiasi tipo**.

In una sua formulazione parziale ma già utilizzabile in molte situazioni l'enunciato del teorema è dato di seguito.

Teorema

La v.c. S_n somma di n v.c. X_i statisticamente indipendenti con la stessa distribuzione, con medie e varianze finite, al tendere di n ad infinito, tende ad assumere una distribuzione normale o gaussiana con media e varianza, rispettivamente:

$$E(S_n) = n\mu \quad Var(S_n) = n\sigma^2$$

Glossario

Distribuzione binomiale: al fenomeno numero di successi in n prove indipendenti è associata la seguente distribuzione di probabilità:

$$\left[\begin{array}{llllll} \boldsymbol{x_i} & 0 & 1 & 2 & \cdots & n \\ \boldsymbol{p_i} & q^n & \binom{n}{1} pq^{n-1} & \binom{n}{2} p^2 q^{n-2} & \cdots & p^n \end{array}\right. \text{con } q = 1 - p$$

dove la probabilità di successo resta costante in ciascuna sottoprova.