CAPITOLO UNDICESIMO

VARIABILI CASUALI

1. INTRODUZIONE

Una **variabile casuale** (v.c.) detta anche **variabile aleatoria**, è una funzione misurabile e a valori reali definita sullo spazio campione Ω. Ad ogni elemento dello spazio campione corrisponde uno e un solo valore della v.c.; si definisce una variabile casuale quando si crea una corrispondenza tra l'insieme dei risultati di una prova e l'insieme dei numeri reali. Tuttavia, la corrispondenza tra eventi e numeri non deve essere necessariamente biunivoca.

Si deve sottolineare, inoltre, che una v.c., se è vero che può assumere, prima di una data prova, un valore qualsiasi, dopo la prova assume uno ed un solo valore numerico, detto **determinazione della v.c.**

Generalmente, come sinonimo di v.c. si usa l'espressione **variabile stocastica**, anche se nel 1987 Dall'Aglio ha operato una distinzione definendo *stocastico* ciò che concerne il calcolo delle probabilità, e *casuale* ciò che deriva da situazioni in cui esiste indifferenza probabilistica nelle preferenze, nei risultati, etc.

Le v.c. possono essere *discrete* o *continue*. Nello schema seguente si distinguono le due tipologie di v.c. di cui daremo una definizione precisa nel corso del capitolo:

Variabili casuali discrete	Variabili casuali continue
Bernoulli	Normale
Binomiale	Lognormale
Poisson	Gamma
Ipergeometrica	Chi-quadrato
Uniforme	*t* di Student
	F di Fisher-Snedecor
	Esponenziale negativa
	Uniforme

2. VARIABILI CASUALI DISCRETE

Una v.c. è **discreta** se i valori che assume sono in corrispondenza di un insieme numerabile.

Ad una v.c. discreta è associata la **funzione di probabilità**, che esprime la probabilità:

$$P(X = x) = p_x$$

di ogni valore assunto dalla v.c.; essa è definita se e solo se:

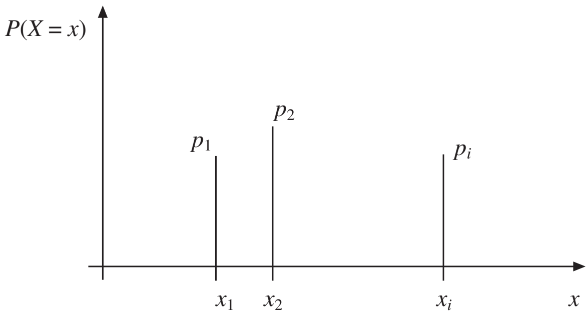
— $p(x_i) \geq 0 \ \forall \ i = 1, 2, \ldots;$

— $\sum_{i=1}^{\infty} p(x_i) = 1 \cdot$

Una v.c. ha una configurazione analoga a quella di una variabile statistica, ne differisce perché, in corrispondenza delle singole modalità, non si trovano le frequenze (assolute o relative), ma le probabilità.

Ne consegue che l'insieme delle probabilità dei diversi valori possibili di una v.c. costituisce una **distribuzione di probabilità** (simile, in sostanza, ad una distribuzione di frequenza) di tale v.c.

Volendo indicare una v.c. graficamente si pongono in ascissa di un sistema di riferimento cartesiano i valori reali che essa assume e in ordinata le corrispondenti probabilità:



Variabile casuale discreta

Nell'inferenza statistica fondamentale importanza ha la **funzione di ripartizione** o **funzione di distribuzione cumulativa** che esprime la probabilità che la v.c. X assuma valori inferiori o uguali ad un valore prefissato, cioè per ogni x reale:

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p(x_i)$$

Essa è non decrescente, compresa sempre tra 0 e 1, ed è tale per cui:

$$\lim_{x \to -\infty} F(x) = 0; \qquad \lim_{x \to +\infty} F(x) = 1$$

Inoltre, è continua da destra.

Esempio 1

Si consideri l'esperimento consistente nel lancio di un dado non truccato, determinare:

a) *i valori che assume la funzione di ripartizione relativa alla v.c. associata al lancio del dado;*
b) *la rappresentazione grafica della funzione di ripartizione.*

La v.c. X associata al lancio del dado assume i soli valori 1, 2, 3, …, 6, i quali sono ugualmente probabili e con probabilità pari a 1/6.

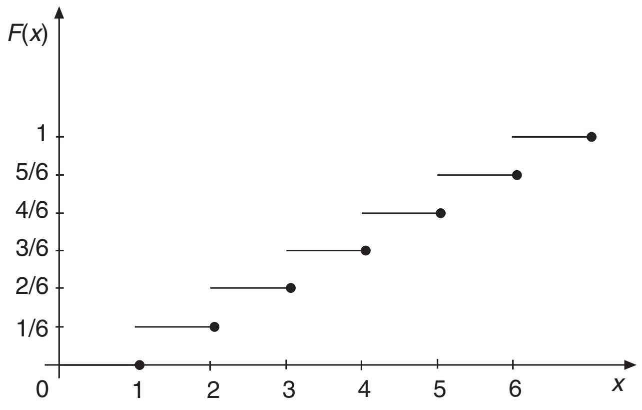
a) La funzione di ripartizione assume, quindi, i seguenti valori:

x	$F(x)$
< 1	0
$i \le x \le i + 1$	$i/6$ per $i = 1, 2, \ldots, 5$
$x \ge 6$	1

Tabella 1

In effetti, anche se i valori della x inferiori ad 1 o superiori a 6 sono impossibili, ha senso calcolare i possibili valori della funzione di ripartizione.

b) La rappresentazione grafica della funzione di ripartizione è una funzione a gradini:



I valori caratteristici di una v.c. sono: il *valore medio* e la *varianza*.

2.1 Valore medio

Data una **v.c. discreta** X che assume i valori $x_1, x_2, \ldots$, con probabilità $p_1, p_2, \ldots$, il **valore atteso** o **speranza matematica** corrisponde al valore medio della distribuzione di probabilità della v.c., in simboli:

$$E(X) = \mu = x_1 p_1 + x_2 p_2 + \ldots = \sum_{i=1}^{\infty} x_i p_i$$

L'espressione *speranza matematica* è tipicamente usata nei giochi d'azzardo.

Esempio 2

Un giocatore d'azzardo vince 20€ *se una carta estratta da un mazzo è di cuori, ne vince* 40 *se la carta è di quadri, ne perde* 30 *se la carta è di fiori e non vince né perde se la carta è di picche. Determinare il suo guadagno atteso.*

La v.c. X è data dalla vincita (o dalla perdita) e la sua funzione di probabilità è rappresentata nella tabella seguente:

x_i	20	40	–30	0
p_i	0,25	0,25	0,25	0,25

Tabella 2

La speranza matematica è data 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 guadagno atteso dal giocatore è di 7,5€. A questo punto sarebbe anche possibile stabilire una misura dell'equità del gioco in base alle poste stabilite per giocare.

2.2 Varianza

La **varianza** di una v.c. è quel valore caratteristico che indica la dispersione dei valori intorno al valore medio. Essa è definita come la speranza matematica del quadrato della differenza tra la variabile X e la sua speranza matematica; sia σ^2 la varianza della v.c. X, in simboli:

$$\sigma^2 = Var(X) = E\left[(X-\mu)^2\right]$$

Se la v.c. X è **discreta** la varianza è pari a:

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

Esempio 3

Determinare la varianza della v.c. definita nella tabella 2.

Per determinare la varianza della v.c. presa in esame si consideri il seguente schema di calcolo:

x_i	p_i	$x_i - \mu$	$(x_i - \mu)^2$	$(x_i - \mu)^2 p_i$
20	0,25	12,5	156,25	39,0625
40	0,25	32,5	1.056,25	264,0625
–30	0,25	–37,5	1.406,25	351,5625
0	0,25	–7,5	56,25	14,0625

Schema 1

Applicando la formula, la varianza è:

$$\sigma^2 = 39{,}0625 + 264{,}0625 + 351{,}5625 + 14{,}0625 = 668{,}75$$

Al solito si definisce lo **scarto quadratico medio** (o **deviazione standard**) σ la radice quadrata della varianza.

Nei successivi paragrafi ci occuperemo delle seguenti v.c. discrete: la *v.c. di Bernoulli*, la *v.c. binomiale*, la *v.c. di Poisson* e la *v.c. ipergeometrica*; mentre della *v.c. uniforme discreta* ci occuperemo trattando le v.c. continue.

3. LA VARIABILE CASUALE DI BERNOULLI

La **v.c. di Bernoulli**, dal nome del matematico svizzero che diede grande contributo allo sviluppo della probabilità, è la distribuzione di probablità di una **variabile dicotomica** (o **binaria**) che assume valori 0 e 1 con probabilità $1 - p$ e p, rispettivamente. Generalmente si attribuisce il valore 1 al successo e il valore 0 all'insuccesso, quindi p è la probabilità del successo e $q = 1 - p$ è la probabilità dell'insuccesso. Essa è il caso particolare ($n = 1$) di una v.c. binomiale di cui ci occuperemo nel prossimo paragrafo. La *media* e la *varianza* di una v.c. di Bernoulli, indicata con $X \sim Ber\ (1, p)$, sono rispettivamente:

$$E(X) = p;\ Var(X) = p\ (1 - p)$$

Tale modello probabilistico si ritrova nel piano degli esperimenti, nel controllo di qualità, nelle analisi cliniche, nei test psicologici etc., in quanto spesso, nelle più svariate situazioni sperimentali, si è interessati ad accertare se un evento E (detto *successo*) si verifica oppure no: vivo-morto; utile-guasto; funziona-non funziona; superato-respinto etc.

4. LA VARIABILE CASUALE BINOMIALE

Si supponga di avere n eventi indipendenti di tipo bernoulliano, tutti della stessa specie, ciascuno dei quali ha una probabilità p di accadere e una probabilità $q = 1 - p$ di non accadere. Se la probabilità del successo di uno di questi eventi è p, allora la probabilità che accadano tutti gli n eventi è data da p^n, allo stesso modo si può dire che la probabilità che tutti gli n eventi non accadano è data da $(1 - p)^n$.

Per calcolare le probabilità che la **v.c. binomiale** assuma, rispettivamente, i valori $x = 0, 1, 2, \ldots, n$ si deve considerare, ogni volta, il numero di modi in cui x oggetti si possono prendere da n, in altri termini, si deve considerare il numero $\binom{n}{x}$ di combinazioni di n oggetti a x a x.

Pertanto, al fenomeno **numero di successi in *n* prove indipendenti** è associata la seguente distribuzione di probabilità:

$$\begin{cases} x_i & 0 & 1 & 2 & \ldots & n \\ p_i & q^n & \binom{n}{1} pq^{n-1} & \binom{n}{2} p^2 q^{n-2} & \ldots & p^n \end{cases}$$

La v.c. binomiale ha funzione di distribuzione:

$$P\left(X=x\right)=\binom{n}{x}p^{x}q^{n-x}=\frac{n!}{x!\left(n-x\right)!}p^{x}q^{n-x} \tag{4.1}$$

media e varianza, rispettivamente, pari a:

$$E(X) = np;\ Var(X) = np(1-p)$$

e si indica in questo modo $X \sim Bin\ (n, p)$.

Il grafico rappresentativo della distribuzione binomiale presenta un andamento diverso a seconda del valore assunto da p, infatti se esso è uguale, minore, maggiore a 0,5 la distribuzione presenta, rispettivamente, simmetria, asimmetria negativa, asimmetria positiva.

Essendo utilizzabile in tutti i casi in cui gli esiti di una prova possono essere ridotti a due (successo - insuccesso), la distribuzione binomiale è tipicamente adoperata nel caso dei controlli di qualità di un dato pezzo che può essere classificato come difettoso oppure non difettoso.

Esempio 1

Un'urna contiene 4 *palline bianche e* 8 *rosse.*

Determinare la probabilità che, estraendo a caso 4 palline dall'urna e rimettendo ogni volta la pallina estratta nell'urna, sortiscano a) 0*; b)* 1*; c)* 2*; d)* 3*; e)* 4 *palline bianche. Rappresentare graficamente la distribuzione di probabilità ottenuta.*

Le probabilità richieste si ottengono applicando la formula (4.1) in cui $p=\frac{4}{12}=\frac{1}{3}$ è la probabilità di estrarre una pallina bianca dall'urna e $q=\frac{8}{12}=\frac{2}{3}$ è la probabilità di estrarre una pallina rossa dall'urna.

Esse sono:

a) $P\left(X{=}0\right)=\binom{4}{0}\left(\frac{1}{3}\right)^{0}\left(\frac{2}{3}\right)^{4}=0{,}1975;$

b) $P\left(X{=}1\right)=\binom{4}{1}\left(\frac{1}{3}\right)^{1}\left(\frac{2}{3}\right)^{3}=0{,}3951;$

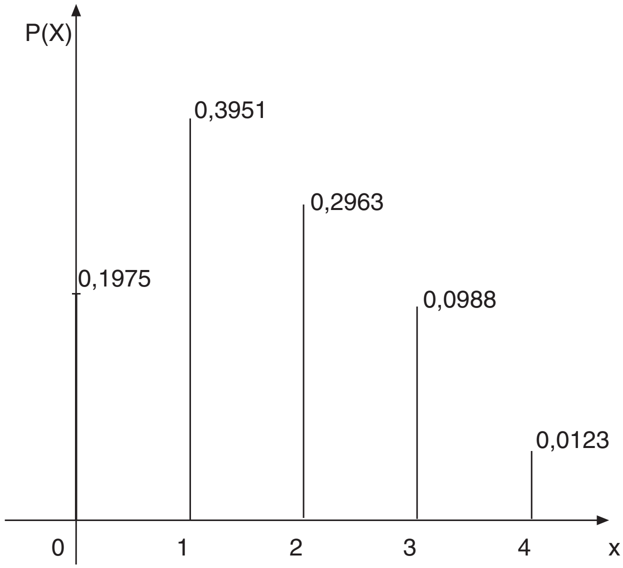
c) $P(X=2)=\binom{4}{2}\left(\frac{1}{3}\right)^{2}\left(\frac{2}{3}\right)^{2}=0{,}2963;$

d) $P(X=3)=\binom{4}{3}\left(\frac{1}{3}\right)^{3}\left(\frac{2}{3}\right)^{1}=0{,}0988;$

e) $P(X=4)=\binom{4}{4}\left(\frac{1}{3}\right)^{4}\left(\frac{2}{3}\right)^{0}=0{,}0123.$

È ovvio che la somma di tutte queste probabilità è uguale a 1.

La rappresentazione grafica della distribuzione di probabilità è la seguente:



Il grafico è asimmetrico.

Esempio 2

Si supponga che la probabilità che nasca un maschio sia esattamente uguale alla probabilità che nasca una femmina $\left(\frac{1}{2}\right)$.

Determinare la probabilità che in una famiglia di 5 *figli:*

a) *almeno uno sia maschio;*
b) 4 *siano femmine.*

Si tratta di un problema di prove ripetute e quindi può essere utilizzata la distribuzione binomiale:

$$P(X = x) = \binom{n}{x} p^x q^{n-x}$$

Applicando la formula abbiamo:

n = numero delle prove = 5 (numero dei figli);

$p = \frac{1}{2}$ probabilità di nascita di un maschio;

$q = \frac{1}{2}$ probabilità di nascita di una femmina.

$$a)\ P(X \geq 1) = \sum_{x=1}^{5}\binom{5}{x}\left(\frac{1}{2}\right)^{x}\cdot\left(\frac{1}{2}\right)^{5-x} = \sum_{x=1}^{5}\binom{5}{x}\left(\frac{1}{2}\right)^{5} = \binom{5}{1}\frac{1}{32}+\binom{5}{2}\frac{1}{32}+\binom{5}{3}\frac{1}{32}+\binom{5}{4}\frac{1}{32}+\binom{5}{5}\frac{1}{32} =$$

$$= 5\cdot\frac{1}{32}+10\cdot\frac{1}{32}+10\cdot\frac{1}{32}+5\cdot\frac{1}{32}+1\cdot\frac{1}{32} = 0,96875$$

$$b)\ P(X=4) = \binom{5}{4}\left(\frac{1}{2}\right)^{4}\cdot\frac{1}{2} = \binom{5}{4}\left(\frac{1}{2}\right)^{5} = 5\cdot\frac{1}{32} = 0,15625$$

Esempio 3

È noto che il 38% *dei dipendenti di una multinazionale è di colore. Considerando un campione casuale di* 18 *dipendenti, determinare:*

a) *la probabilità che al massimo* 4 *dipendenti siano di colore;*
b) *la probabilità che sia di colore un numero di dipendenti compreso tra* 6 *e* 9*;*
c) *la media e la varianza della distribuzione.*

La v.c. discreta X rappresenta il numero di persone di colore è:

$$P = (X = x) = \binom{18}{x}(0,38)^{x}(0,62)^{18-x}$$

a) La probabilità che al massimo 4 persone siano di colore è data da:

$$P(X \leq 4) = P(X=0)+P(X=1)+P(X=2)+P(X=3)+P(X=4)$$

Sapendo che n = 18, p = 0,38 e q = 0,62 si ha:

$$P(X=0) = \binom{18}{0}(0,38)^{0}(0,62)^{18} = 0,00018$$

$$P(X=1) = \binom{18}{1}(0,38)^{1}(0,62)^{17} = 0,00202$$

$$P(X=2) = \binom{18}{2}(0,38)^{2}(0,62)^{16} = 0,01053$$

$$P(X=3) = \binom{18}{3}(0,38)^{3}(0,62)^{15} = 0,03443$$

$$P(X=4) = \binom{18}{4}(0,38)^{4}(0,62)^{14} = 0,07913$$

per cui:

$P(X \leq 4)$ = 0,00018 + 0,00202 + 0,01053 + 0,03443 + 0,07913 = 0,12629

b) La probabilità che sia di colore un numero di dipendenti compreso tra 6 e 9 è:

$P(6 \leq X \leq 9) = P(X = 6) + P(X = 7) + P(X = 8) + P(X = 9)$

Le probabilità considerate sono, rispettivamente:

$$P(X = 6) = \binom{18}{6}(0,38)^6(0,62)^{12} = 0,18033$$

$$P(X = 7) = \binom{18}{7}(0,38)^7(0,62)^{11} = 0,18947$$

$$P(X = 8) = \binom{18}{8}(0,38)^8(0,62)^{10} = 0,15968$$

$$P(X = 9) = \binom{18}{9}(0,38)^9(0,62)^{9} = 0,10874$$

per cui la probabilità richiesta è:

$P(6 \leq X \leq 9) = 0,18033 + 0,18947 + 0,15968 + 0,10874 = 0,63822$

c) La media e la varianza della distribuzione sono pari, rispettivamente a:

$E(X) = np = 18 \cdot 0,38 = 6,84$

$Var(X) = npq = 18 \cdot 0,38 \cdot 0,62 = 4,2408$

5. 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 $n \to \infty$. Per ricavare la distribuzione di Poisson osserviamo che se p è molto piccolo, il numero medio di eventi sarà molto più piccolo di n (totalità delle prove), cosicché x, che è il numero di successi, sarà estremamente più piccolo di n. Nell'espressione della distribuzione binomiale possiamo allora fare due approssimazioni.

Per cui, alla fine ponendo $np = \lambda$ avremo:

$$\frac{\lambda^x}{x!}(1-p)^{\lambda/p} = \frac{\lambda^x}{x!}\left[(1-p)^{1/p}\right]^{\lambda}$$

Inoltre, poiché:

$$\lim_{p \to 0}\left[\left(1-p\right)^{1/p}\right]^{\lambda} = \exp\left(-\lambda\right)$$

avremo che la funzione di distribuzione di Poisson è data da:

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

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

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

La v.c. di Poisson si indica con $X \sim Po\ (\lambda)$. Il grafico rappresentativo della distribuzione di probabilità presenta asimmetria positiva, al crescere di λ diviene più simmetrico.

Esempio 1

In una data popolazione il numero delle nascite mensili segue la distribuzione di Poisson, con un tasso di natalità mensile medio $\lambda = 2$.
Determinare la probabilità di avere in un mese un numero di nascite maggiore di 2.

La probabilità di avere un numero di nascite maggiore di 2 in un dato mese si determina in questo modo:

$$P(X>2)=1-\left[P(X=0)+P(X=1)+P(X=2)\right]$$

Con $\lambda = 2$ e $x = 0, 1, 2$ si ha:

$$P(X=0)=\frac{\exp(-2)2^0}{0!}=0{,}135$$
$$P(X=1)=\frac{\exp(-2)2^1}{1!}=0{,}271$$
$$P(X=2)=\frac{\exp(-2)2^2}{2!}=0{,}271$$

quindi:

$$P(X > 2) = 1 - [0{,}135 + 0{,}271 + 0{,}271] = 0{,}323$$

Esempio 2

In un piccolo albergo di provincia ogni 3 *giorni arrivano in media* 4 *clienti.*

Determinare:

a) *la probabilità che in* 3 *giorni arrivi un numero di clienti minore o uguale a* 5*;*
b) *la probabilità che in* 3 *giorni arrivi un numero di clienti maggiore di* 6*.*

a) La probabilità che in 3 giorni arrivi un numero di clienti minore o uguale a 5 è:

$P(X \leq 5) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)$

dove:

$$P(X=0)=\frac{\exp(-4)4^0}{0!}=0{,}0183$$

$$P(X=1)=\frac{\exp(-4)4^1}{1!}=0{,}0733$$

$$P(X=2)=\frac{\exp(-4)4^2}{2!}=0{,}1465$$

$$P(X=3)=\frac{\exp(-4)4^3}{3!}=0{,}1954$$

$$P(X=4)=\frac{\exp(-4)4^4}{4!}=0{,}1954$$

$$P(X=5)=\frac{\exp(-4)4^5}{5!}=0{,}1563$$

da cui:

$P(X \leq 5) = 0{,}0183 + 0{,}0733 + 0{,}1465 + 0{,}1954 + 0{,}1954 + 0{,}1563 = 0{,}7851$

b) La probabilità che in 3 giorni arrivi un numero di clienti maggiore di 6 è data da:

$$\begin{aligned} P(X>6) &= 1-\left[P(X=0)+P(X=1)+P(X=2)+P(X=3)+P(X=4)+P(X=5)\right]= \\ &= 1-P(X\leq 5)=1-0{,}7851=0{,}2149 \end{aligned}$$

6. LA VARIABILE CASUALE IPERGEOMETRICA

La **v.c. ipergeometrica** assume rilievo allorché si fa riferimento alla distribuzione di probabilità associata ad una **estrazione senza ripetizione** da una popolazione di ampiezza finita *H*. Per questo la variabile è detta anche **v.c. dell'estrazione in blocco**.

Se la popolazione è composta da r elementi di un tipo e da $H-r$ elementi di un altro tipo, allora la probabilità di trovare x elementi del primo tipo, da un campione di numerosità n, è data da:

$$P(X=x)=\frac{\binom{r}{x}\binom{H-r}{n-x}}{\binom{H}{n}}$$

Quando N è abbastanza grande ed n è piccolo rispetto a N, tale distribuzione può essere **approssimata** dalla v.c. binomiale.