CAPITOLO TERZO – VARIABILI CASUALI

§1. Le variabili casuali e la loro distribuzione di probabilità

In molte situazioni, dato uno spazio di probabilità S, si è interessati non tanto agli eventi elementari (o composti) relativi ad S, quanto ad un valore numerico associato a ciascun evento elementare. In un precedente esempio relativo a famiglie con due figli abbiamo posto l'attenzione sul numero dei figli maschi. Generalmente, lanciando un dado, si considera il valore numerico della faccia uscita e, lanciando una coppia di dadi, interessa il punteggio totale realizzato oppure il valore massimo fra i due ottenuti. Spesso si considera il numero di teste ottenute lanciando un certo numero di volte una moneta. Questa corrispondenza tra esiti di un esperimento e numeri non è confinata ai giochi. Ad esempio, se si deve esaminare uno stock di pezzi meccanici e si controlla un campione, si vuole determinare il numero di pezzi difettosi presenti nel campione stesso. L'obiettivo dello studio di una popolazione può essere la distribuzione delle altezze o dei pesi: di ciascun individuo scelto a caso interessa allora il valore numerico che esprime, in una prefissata unità, la misura della grandezza da esaminare.

In generale, ai possibili eventi elementari $e_1, e_2,..., e_n$ di un dato spazio S sono associati dei valori numerici: più tecnicamente è data una *funzione* avente come dominio lo spazio S e che a ciascun evento elementare di S associa un numero reale. Funzioni di tal genere sono dette **variabili casuali** o **aleatorie**: variabili in quanto suscettibili di assumere valori diversi, casuali poiché il valore da esse assunto dipende dall'esito di un esperimento casuale, ossia da quale evento elementare si è realizzato in una data prova. Una variabile casuale sarà indicata con X. Data una variabile casuale X indichiamo con $x_1, x_2,..., x_n$ l'insieme dei suoi possibili valori. Se x_i è un valore della variabile casuale X, indichiamo con p_i la probabilità che assuma il valore x_i, in formula: $p_i = p(X = x_i)$. La determinazione di p_i è subordinata alla scelta del modello probabilistico relativo allo spazio di eventi relativamente al quale la variabile casuale è definita. Dato lo spazio di probabilità S, p_i si calcola semplicemente sommando le probabilità degli eventi elementari ai quali è associato il valore x_i di X e si perviene ad uno schema di questo tipo:

X	x_1	x_2	...	x_n
	p_1	p_2	...	p_n

nella quale si riportano i valori della variabile casuale e le rispettive probabilità, ossia la **distribuzione di probabilità** della variabile casuale.

***Esempio* 1**. Data la variabile casuale X = "numero di figli maschi" nelle famiglie con due figli, determinare la distribuzione di probabilità.
Evidentemente X può assumere soltanto i valori 0, 1 e 2. Assumendo l'equiprobabilità e l'indipendenza tra le nascite maschili e femminili (e quindi l'equiprobabilità degli eventi elementari MM, MF, FM, FF) si ha immediatamente la seguente distribuzione di probabilità:

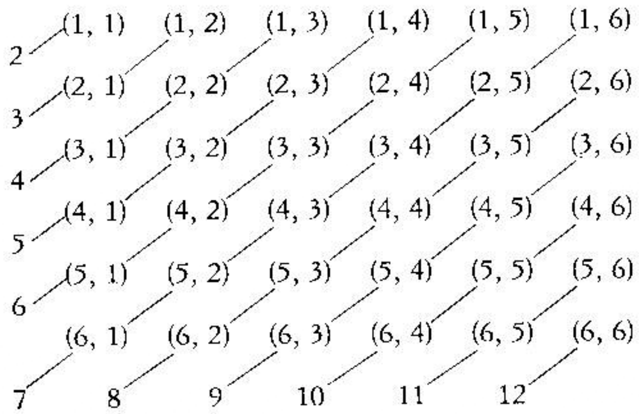
X	0	1	2
	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

***Esempio* 2**. Si determini la distribuzione di probabilità della variabile casuale X = "punteggio ottenuto lanciando un dado".
Si ha immediatamente:

X	1	2	3	4	5	6
	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

***Esempio* 3**. Si lanciano due dadi. Determinare la distribuzione di probabilità della variabile casuale X = “somma dei punteggi dei due dadi”.

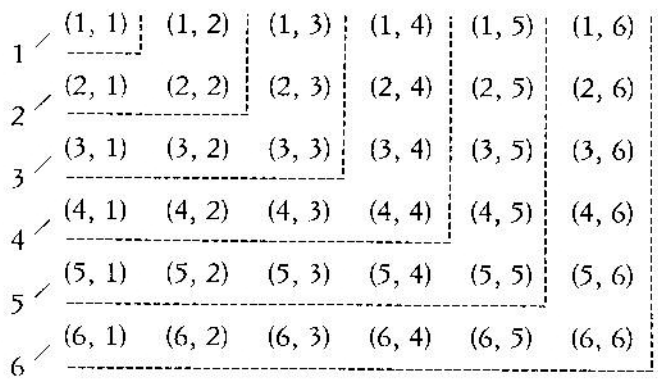
Gli eventi elementari sono 36, ciascuno con probabilità $\frac{1}{36}$:



I valori della variabile casuale X vanno da 2 a 12. Per determinare le rispettive probabilità basta moltiplicare per $\frac{1}{36}$ il numero di coppie cui corrisponde il dato valore di X:

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

***Esempio* 4**. Si lanciano due dadi. Determinare la distribuzione di probabilità della variabile casuale X = “massimo punteggio ottenuto”.
I valori di X vanno da 1 a 6:



Le probabilità si determinano come nell’esempio precedente:

X	1	2	3	4	5	6
	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

***Esempio* 5**. Si lancia quattro volte una moneta. Determinare la distribuzione di probabilità della variabile casuale X = “numero di teste ottenute”.
I valori che X può assumere sono 0, 1, 2, 3, 4 e gli eventi elementari sono 16:

TTTT TTTC TTCT TTCC TCTT TCTC TCCT TCCC
CTTT CTTC CTCT CTCC CCTT CCTC CCCT CCCC

ciascuno con probabilità $\frac{1}{16}$.

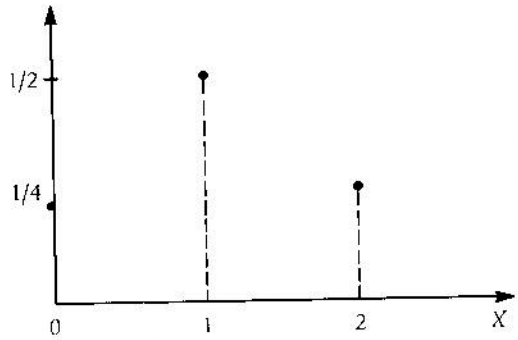
Dato che gli eventi elementari favorevoli a $X = 0$, $X = 1$, $X = 2$, $X = 3$, $X = 4$ sono rispettivamente 1, 4, 6, 4, 1, si ottiene:

X	0	1	2	3	4
	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

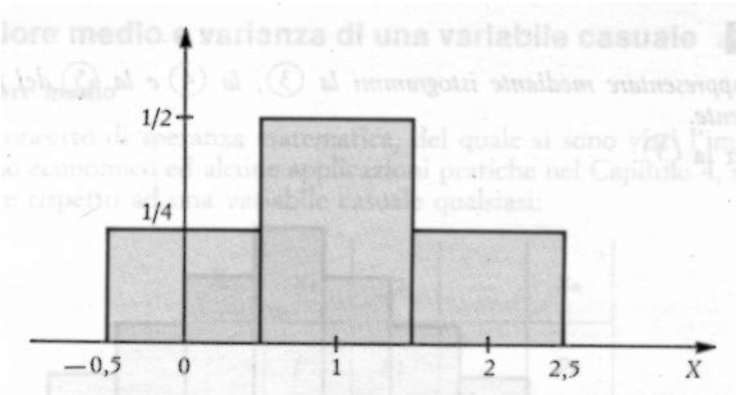
§2. Rappresentazioni grafiche

Si rivela conveniente, data una variabile casuale X, rappresentarne graficamente la distribuzione di probabilità in un diagramma cartesiano nel quale si riportano in ascissa i valori di X e in ordinata le rispettive probabilità. È utile inoltre considerare il relativo *istogramma* come illustriamo nei seguenti grafici che corrispondono ai cinque esempi del paragrafo precedente.

***Esempio* 1**. Il diagramma cartesiano è:

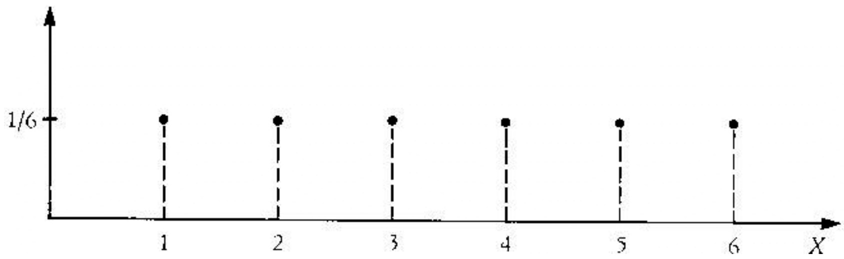


Per ottenere la rappresentazione mediante un istogramma si construiscono tanti rettangoli aventi basi di lunghezza unitaria centrate nei valori di X e come altezze le relative probabilità.

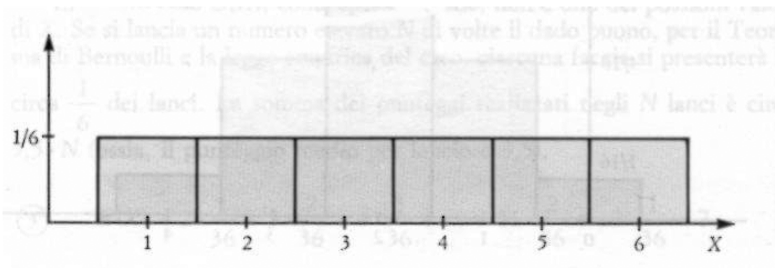


A proposito della rappresentazione con istogrammi è bene osservare che $p_i = p(X = x_i)$ può essere visualizzata mediante l'area del rettangolo la cui base ha x_i come punto medio. Così $p_0 = p(X = 0) = \frac{1}{4}$ è l'area del rettangolo a sinistra, $p_1 = p(X = 1) = \frac{1}{2}$ è l'area del rettangolo centrale e $p_2 = p(X = 2) = \frac{1}{4}$ è l'area del rettangolo a destra. Si noti anche che la somma delle aree dei tre rettangoli è 1.

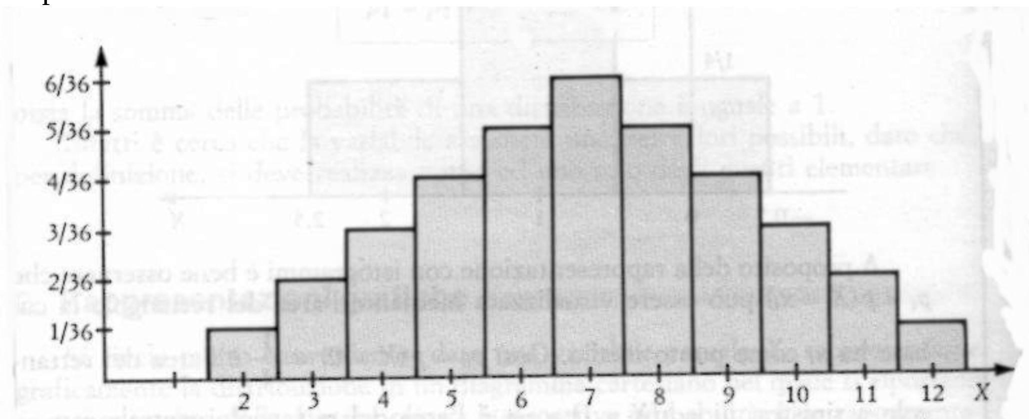
***Esempio* 2.** La rappresentazione cartesiana è:

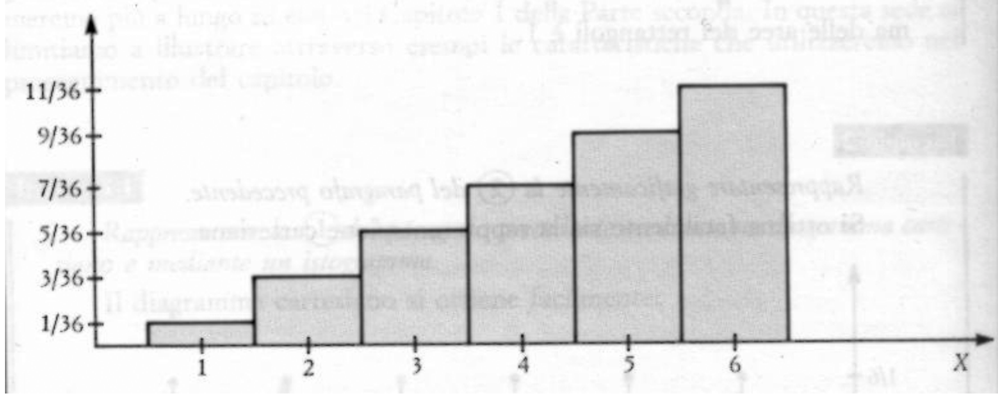


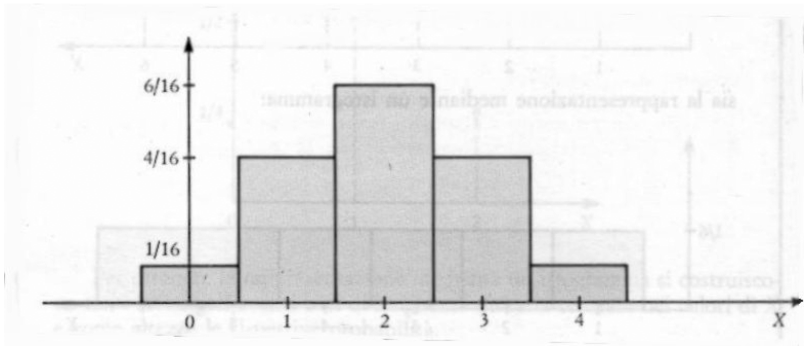
e il corrispondente istogramma risulta:



Di seguito riportiamo gli istogrammi relativi alle variabili casuali degli Esempi 3, 4 e 5 del paragrafo precedente:







§3. Valore medio e varianza di una variabile casuale

A) Valore medio

Il concetto di speranza matematica si può generalizzare rispetto ad una variabile casuale X qualsiasi:

X	x_1	x_2	...	x_n
	p_1	p_2	...	p_n

Definizione. Si dice **valore medio** (o **speranza matematica**) di una variabile casuale X, e lo indichiamo con $E(X)$, la quantità: $E(X) = p_1 \cdot x_1 + p_2 \cdot x_2 + \ldots + p_n \cdot x_n$.
Intuitivamente, $E(X)$ rappresenta il valore che *in media* ci si aspetta di ottenere in una ripetizione di prove dai cui esiti dipendono i valori di X.
***Esempio* 1**. Determinare il valore medio delle variabili casuali degli Esempi 1-5 del §1.

1) $E(X) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1$

Tale valore indica, come del resto è intuitivamente ovvio, che, assumendo l'equiprobabilità degli eventi elementari MM, MF, FM, FF, il numero medio dei maschi per famiglia è 1 (su N famiglie con due figli il numero dei maschi è N).

2) $E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2} = 3{,}5$

In questo caso $E(X)$, come spesso accade, non è uno dei possibili valori di X. Se si lancia un numero elevato N di volte il dado buono, per la legge empirica del caso ciascuna faccia si presenterà in circa $\frac{1}{6}$ dei lanci. La somma dei punteggi realizzati negli N lanci è circa $3{,}5N$, ossia il punteggio medio per lancio è 3,5.

3) $E(X) = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + \ldots + 11 \cdot \frac{2}{36} + 12 \cdot \frac{1}{36} = 7$

Come è intuitivamente plausibile, essendo 3,5 il punteggio medio nel lancio di un dado, il valore medio nel lancio di due dadi è il doppio, cioè 7.

4) $E(X) = 1 \cdot \frac{1}{36} + 2 \cdot \frac{3}{36} + 3 \cdot \frac{5}{36} + 4 \cdot \frac{7}{36} + 5 \cdot \frac{9}{36} + 6 \cdot \frac{9}{36} = \frac{161}{36} \cong 4{,}47$

Il fatto che i valori più alti di X abbiano maggiore probabilità di quelli più bassi giustifica il valore relativamente elevato di $E(X)$.

5) $E(X) = 0 \cdot \frac{1}{16} + 1 \cdot \frac{4}{16} + 2 \cdot \frac{6}{16} + 3 \cdot \frac{4}{16} + 4 \cdot \frac{1}{16} = 2$

Anche qui il risultato era largamente prevedibile: in 4 lanci di una moneta buona il valore medio del numero di teste ottenute è 2.

B) **Varianza**

Dopo il valore medio, il parametro più usato per caratterizzare le distribuzioni di probabilità delle variabili casuali è la varianza, la quale indica quanto sono "dispersi" i valori della variabile aleatoria relativamente al suo valore medio. Data una variabile casuale X qualsiasi:

X	x_1	x_2	...	x_n
	p_1	p_2	...	p_n

sia $E(X)$ il suo valore medio.

Consideriamo la variabile casuale $X - E(X)$:

X	$x_1 - E(X)$	$x_2 - E(X)$	...	$x_n - E(X)$
	p_1	p_2	...	p_n

i cui valori sono le "distanze" tra i valori di X e il valore medio $E(X)$. Per ottenere una misura della dispersione dei valori di X occorre rendere sempre positivi i valori di $X - E(X)$ (si verifica facilmente che il valore medio di $X - E(X)$ vale zero), ad esempio considerando la variabile casuale $(X - E(X))^2$:

X	$(x_1 - E(X))^2$	$(x_2 - E(X))^2$	...	$(x_n - E(X))^2$
	p_1	p_2	...	p_n

Definizione. Si dice **varianza** della variabile casuale X, e la indichiamo con $var(X)$, il valore medio di $(X - E(X))^2$, ossia:

$$var(X) = (x_1 - E(X))^2 \cdot p_1 + (x_2 - E(X))^2 \cdot p_2 + \ldots + (x_n - E(X))^2 \cdot p_n$$

Il procedimento con cui abbiamo definito la varianza illustra il motivo per cui tale parametro indica la "dispersione" dei valori di X: se i valori di una variabile casuale sono concentrati vicino al valore medio $E(X)$, il valore medio dei valori di $(X - E(X))^2$ sarà basso; in caso contrario vi saranno quadrati delle differenze elevati, e elevato sarà anche il loro valore medio.

***Esempio* 2**. Calcolare la varianza delle variabili casuali degli Esempi 1-3 del §1.

Per il calcolo delle varianze delle variabili casuali è necessario aver prima determinato il loro valore medio (vedi Esempio 1).

1) Valori di X: 0 1 2

Dato che $E(X) = 1$, si ha:

Valori di $X - E(X)$:	–1	0	1
Valori di $(X - E(X))^2$:	1	0	1
Rispettive probabilità:	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Quindi: $var(X) = 1 \cdot \frac{1}{4} + 0 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = \frac{1}{2}$

2) Valori di X: 1 2 3 4 5 6

Dato che $E(X) = \frac{7}{2}$, si ha:

Valori di $X - E(X)$:	$-\frac{5}{2}$	$-\frac{3}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{2}$	$\frac{5}{2}$
Valori di $(X - E(X))^2$:	$\frac{25}{4}$	$\frac{9}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{9}{4}$	$\frac{25}{4}$
Rispettive probabilità:	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Quindi: $$var(X) = \frac{25}{4}\cdot\frac{1}{6} + \frac{9}{4}\cdot\frac{1}{6} + \frac{1}{4}\cdot\frac{1}{6} + \frac{1}{4}\cdot\frac{1}{6} + \frac{9}{4}\cdot\frac{1}{6} + \frac{25}{4}\cdot\frac{1}{6} = \frac{35}{12}$$

3) Valori di X: 2 3 4 5 6 7 8 9 10 11 12

Dato che $E(X) = 7$, si ha:

Valori di $X - E(X)$:	–5	–4	–3	–2	–1	0	1	2	3	4	5
Valori di $(X - E(X))^2$:	25	16	9	4	1	0	1	4	9	16	25
Rispettive probabilità:	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Quindi, osservando che i primi cinque addendi coincidono con gli ultimi cinque e che il sesto è nullo:

$$var(X) = \left(25\cdot\frac{1}{36} + 16\cdot\frac{2}{36} + 9\cdot\frac{3}{36} + 4\frac{4}{36} + 5\cdot\frac{1}{36}\right)\cdot 2 = \frac{105}{18} = \frac{35}{6}$$

La varianza, come parametro che indica la dispersione, ha il difetto di essere di secondo grado rispetto ai valori di X (così, ad esempio, se i valori di X sono metri, la varianza è espressa in metri quadrati).

Si introduce allora, accanto alla varianza, la **deviazione standard**, che solitamente si indica con σ (sigma), che è definita come la radice quadrata della varianza: $\sigma = \sqrt{var(X)}$.