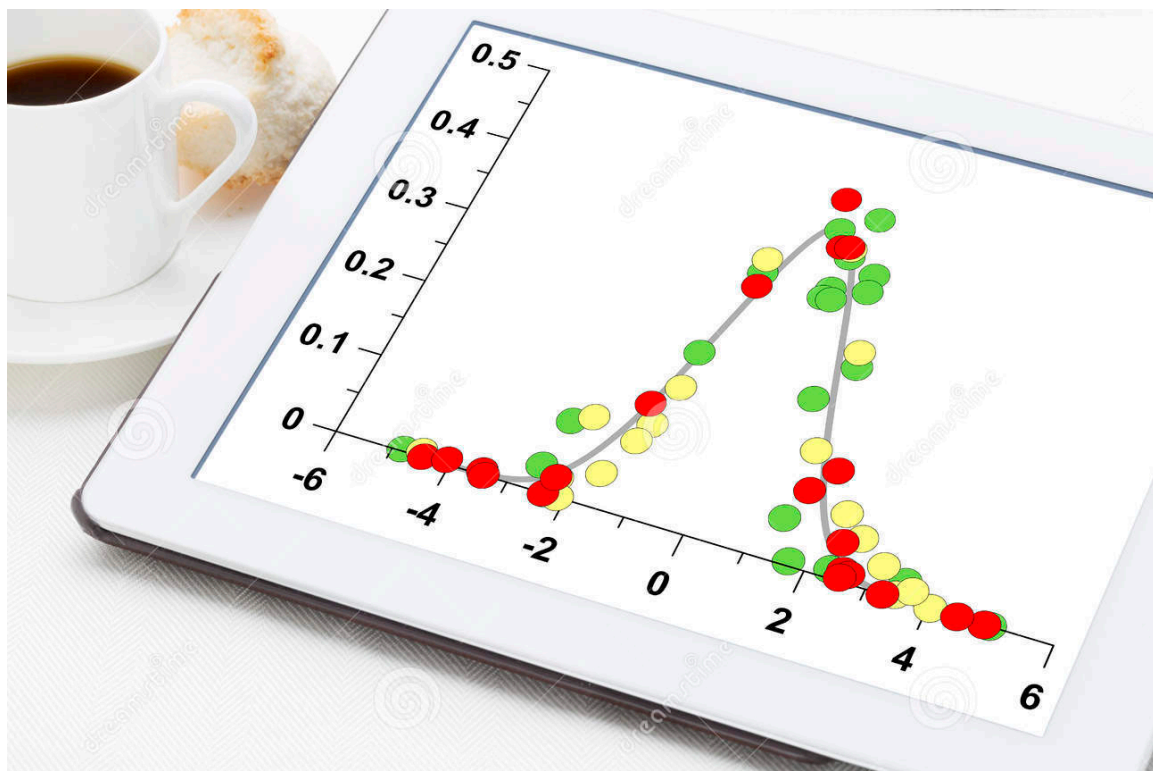


DOCENTE: Vincenzo Pappalardo
MATERIA: Matematica

Distribuzioni di probabilità



Variabili casuali discrete e distribuzioni di probabilità

Definizioni

Nel corso di un mese si è rilevato che il prezzo di un certo prodotto al kilogrammo ha assunto i seguenti valori per il numero dei giorni indicati:

Prezzo al kg in euro	4,70	5,25	4,98	4,80
Numero giorni	9	6	12	3

Il prezzo è una *grandezza* che ha assunto più valori nel corso del mese, risultando *variabile* in seguito a eventi *casuali* avvenuti nel mercato del prodotto. Diciamo allora che il prezzo è descritto da una **variabile casuale**.

Possiamo associare a ogni valore del prezzo la sua frequenza relativa:

Prezzo	4,70	5,25	4,98	4,80
Frequenza relativa	$\frac{3}{10}$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{1}{10}$

$$f_{\text{relativa}} = \frac{n}{N}$$

n= numero di volte in cui un valore si presenta
N= numero totale delle rilevazioni effettuate

Abbiamo descritto una variabile casuale empirica che deriva da una rilevazione statistica. Essa assume valori, dovuti al caso, che hanno una determinata frequenza relativa. È **empirica** in quanto abbiamo rilevato i valori *dopo* che si sono verificati. Infatti non avremmo potuto sapere in precedenza il valore che il prezzo avrebbe assunto, né per quanti giorni si sarebbe mantenuto.

Altro esempio

Abbiamo un'urna contenente 6 palline verdi e 4 gialle. Estraiamo senza reimmissione 3 palline e valutiamo quante volte potrebbe uscire la pallina verde. Essa può non uscire mai, oppure una volta, o due, o tre volte. Il numero delle volte è una variabile casuale **teorica**, perché a ogni suo valore possiamo associare un valore di probabilità calcolato teoricamente.

Indichiamo con X la variabile e con $P(X)$ la successione delle probabilità corrispondenti ai valori di X :

X	0	1	2	3
$P(X)$	$\frac{1}{30}$	$\frac{3}{10}$	$\frac{1}{2}$	$\frac{1}{6}$

Gli eventi si escludono a vicenda e, fra i valori considerati, uno si deve sicuramente verificare, quindi la somma delle probabilità è 1. La successione delle probabilità viene detta *distribuzione di probabilità* della variabile X .

DEFINIZIONE

Variabile casuale (o aleatoria) discreta

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

DEFINIZIONE

Distribuzione di probabilità

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

Possiamo rappresentare la distribuzione di probabilità di una variabile casuale discreta per mezzo di un istogramma o di un diagramma cartesiano.

ESEMPIO

Abbiamo un'urna con tre palline numerate da 1 a 3. Estraiamo consecutivamente due palline, rimettendo ogni volta la pallina estratta nell'urna. Costruiamo la distribuzione di probabilità della variabile casuale $X = \text{«somma dei due numeri estratti»}$.

Le possibili coppie di palline che si possono avere dall'estrazione sono:

somma dei numeri estratti	
(1, 1)	2
(1, 2); (2, 1)	3
(1, 3); (3, 1); (2, 2)	4
(2, 3); (3, 2)	5
(3, 3)	6

dove ogni coppia ha probabilità $\frac{1}{9}$.

Quindi la variabile casuale X può assumere i valori 2, 3, 4, 5, 6, a cui corrispondono le probabilità:

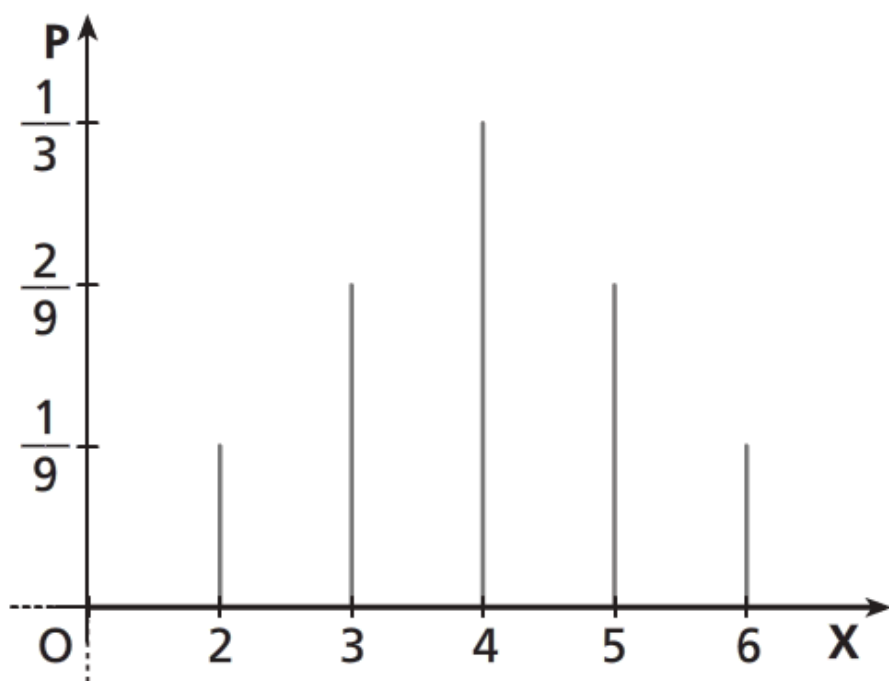
$$P(X = 2) = P(X = 6) = \frac{1}{9};$$

$$P(X = 3) = P(X = 5) = \frac{2}{9};$$

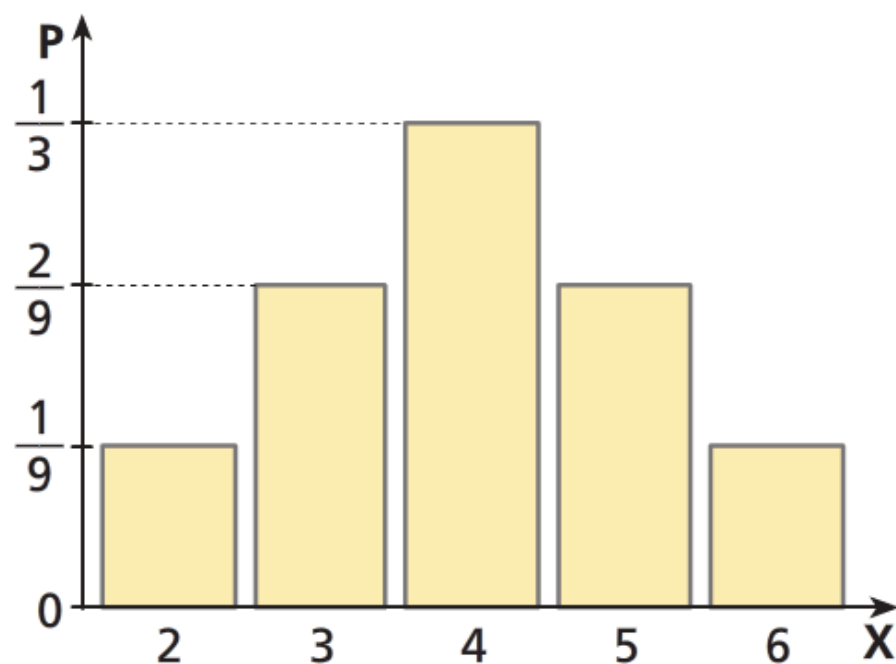
$$P(X = 4) = \frac{3}{9} = \frac{1}{3}.$$

Abbiamo una variabile casuale X con la seguente distribuzione di probabilità:

X	2	3	4	5	6
$P(X)$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{3}$	$\frac{2}{9}$	$\frac{1}{9}$



a. Diagramma cartesiano della distribuzione di probabilità di X .



b. Istogramma della distribuzione di probabilità di X .

La funzione di ripartizione

Riprendiamo il problema in cui estraiamo senza reimmissione 3 palline da un'urna contenente 6 palline verdi e 4 gialle e consideriamo la distribuzione di probabilità relativa all'uscita di palline verdi:

X	0	1	2	3
$P(X)$	$\frac{1}{30}$	$\frac{3}{10}$	$\frac{1}{2}$	$\frac{1}{6}$

Ci domandiamo qual è la probabilità che le palline verdi escano al massimo una volta, oppure al massimo due volte ecc.

Applichiamo il seguente teorema (vedere modulo calcolo delle probabilità):

TEOREMA

Teorema della probabilità totale

Dati n eventi a due a due incompatibili $E_1, E_2, \dots, E_n$, la probabilità della loro unione è uguale alla somma delle loro singole probabilità:

$$p(E_1 \cup E_2 \cup \dots \cup E_n) = p(E_1) + p(E_2) + \dots + p(E_n).$$

Per il teorema della probabilità totale, a questa domanda possiamo rispondere associando a ciascun valore della variabile casuale la somma cumulativa delle probabilità, ossia la somma delle probabilità dei valori che lo precedono e della probabilità del valore stesso.

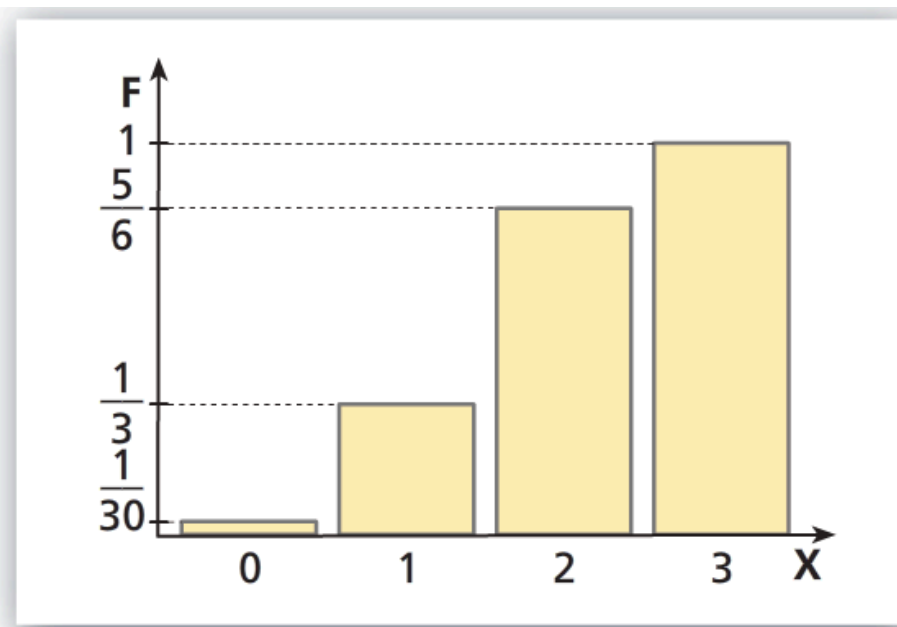
Per esempio, nella tabella a fianco,

$$\frac{1}{3} = \frac{1}{30} + \frac{3}{10}, \quad \frac{5}{6} = \frac{1}{3} + \frac{1}{2}.$$

X	0	1	2	3
$P(X)$	$\frac{1}{30}$	$\frac{3}{10}$	$\frac{1}{2}$	$\frac{1}{6}$
$F(X)$	$\frac{1}{30}$	$\frac{1}{3}$	$\frac{5}{6}$	1

La funzione che a ogni valore di X fa corrispondere la relativa somma cumulativa di probabilità è chiamata *funzione di ripartizione*.

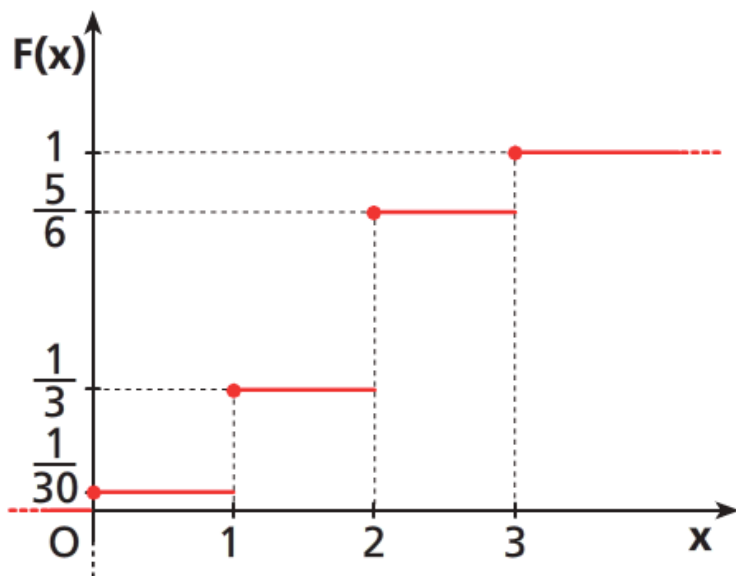
A essa corrisponde la rappresentazione qui a fianco.



● Invece dell'insieme dei valori di X , come dominio della funzione di ripartizione si può prendere tutto $\mathbb{R}$. In questo caso:

$$F(x) = \begin{cases} 0 & \text{se } x < x_1 \\ \sum_{k=1}^i p_k & \text{se } x_i \leq x < x_{i+1} \\ 1 & \text{se } x \geq x_n \end{cases}$$

$$\begin{aligned} &\bullet F(x) = F(x_1) \\ &\text{per } x_1 \leq x < x_2, \\ &F(x) = F(x_2) \text{ per } \\ &x_2 \leq x < x_3, \dots \end{aligned}$$



◀ **Figura 4** L'andamento della funzione di ripartizione, con dominio $\mathbb{R}$, nel caso dell'esempio precedente.

La funzione è definita per casi:

$$F(x) = \begin{cases} 0 & \text{se } x < 0 \\ \frac{1}{30} & \text{se } 0 \leq x < 1 \\ \frac{1}{3} & \text{se } 1 \leq x < 2 \\ \frac{5}{6} & \text{se } 2 \leq x < 3 \\ 1 & \text{se } x \geq 3 \end{cases}$$

DEFINIZIONE

Funzione di ripartizione

Si chiama funzione di ripartizione di una variabile casuale X la funzione $F(x)$ che fornisce la probabilità che X assuma un valore non superiore a un valore prefissato x :

$$F(x_i) = P(X \leq x_i) = p_1 + p_2 + \dots + p_i = \sum_{k=1}^i p_k.$$

X	x_1	x_2	$x_{\dots}$	x_n
$P(X)$	p_1	p_2	$p_{\dots}$	p_n
$F(X)$	p_1	$p_1 + p_2$	$p_1 + p_2 + \dots$	$p_1 + p_2 + \dots + p_n$

esempio

Possiamo utilizzare la funzione di ripartizione per calcolare, per esempio, la probabilità che le palline verdi si presentino più di una volta: la probabilità che sia estratta al più una pallina verde è $F(1) = P(X \leq 1) = \frac{1}{3}$; allora la probabilità cercata è quella dell'evento contrario, cioè:

$$1 - F(1) = 1 - \frac{1}{3} = \frac{2}{3}.$$

Operazioni sulle variabili casuali

Operazioni tra una variabile e delle costanti

Un'urna contiene 12 palline, di cui una pallina con il numero 3, quattro palline con il 4, due con il 7 e cinque con il 9. Estraiamo una pallina.

Consideriamo la variabile casuale $X = \text{«numero della pallina estratta»}$ e la sua distribuzione di probabilità:

X	3	4	7	9
$P(X)$	$\frac{1}{12}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{5}{12}$

Da questa variabile casuale possiamo ottenerne altre, i cui valori sono distribuiti con *le stesse probabilità* dei valori di X , moltiplicando e/o sommando delle costanti ai suoi valori.

Per esempio, se moltiplichiamo i valori di X per 5, otteniamo la variabile casuale, indicata con $5X$, avente distribuzione di probabilità:

$5X$	15	20	35	45
$P(5X)$	$\frac{1}{12}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{5}{12}$

Se invece ai valori della variabile X aggiungiamo 3, otteniamo la variabile casuale, indicata con $X + 3$, avente distribuzione di probabilità:

$X + 3$	6	7	10	12
$P(X + 3)$	$\frac{1}{12}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{5}{12}$

$$p_{classica} = \frac{n}{N}$$

n= numero casi favorevoli
N= numero casi possibili

In generale, date una **variabile casuale** X e una **costante** h , si denota con:

- hX la variabile casuale i cui valori sono dati dal prodotto di h per i valori della variabile X ;
- $X + h$ la variabile casuale i cui valori sono dati dalla somma dei valori della variabile X con h .

Per entrambe queste variabili, i valori assunti hanno le **stesse probabilità dei corrispondenti valori di X** .

Somma di due variabili

I) Prendiamo ora anche una seconda urna. Questa contiene due palline con il numero 1 e tre con il 2. Estraiamo una pallina.

Consideriamo la variabile casuale $Y =$ «numero della pallina estratta dalla seconda urna» e la sua distribuzione di probabilità:

Y	1	2
$P(Y)$	$\frac{2}{5}$	$\frac{3}{5}$

Estraiamo una pallina dalla prima urna e una pallina dalla seconda. Consideriamo la variabile casuale «somma dei due numeri estratti». Indichiamo questa variabile con la notazione $X + Y$ e diciamo che è la **somma delle variabili casuali** X e Y .

Per determinare la distribuzione di probabilità di $X + Y$, costruiamo la tabella in cui le righe corrispondono ai valori della X e le colonne corrispondono a quelli della Y . All'interno della tabella, ogni casella contiene la probabilità di estrazione di due palline aventi rispettivamente il valore x_i e y_j , corrispondenti alla riga i e alla colonna j che individuano la casella. Tale probabilità è la **probabilità congiunta** del prodotto di due eventi **indipendenti**:

$X \backslash Y$	1	2
3	$\frac{1}{30}$	$\frac{1}{20}$
4	$\frac{2}{15}$	$\frac{1}{5}$
7	$\frac{1}{15}$	$\frac{1}{10}$
9	$\frac{1}{6}$	$\frac{1}{4}$

A_i = «estrazione del numero x_i dalla prima urna» $i = 1, 2, 3, 4$;

B_j = «estrazione del numero y_j dalla seconda urna» $j = 1, 2$.

● Dati due eventi indipendenti A e B , si ha:

$$p(A \cap B) = p(A) \cdot p(B).$$

Per esempio, nella tabella, il valore $\frac{1}{20}$, relativo

alla probabilità che escano 3 e 2, si ottiene con:

$$\frac{1}{12} \cdot \frac{3}{5} = \frac{1}{20}.$$

La variabile casuale $X + Y$ ha per valori le possibili somme tra i valori delle due variabili X e Y . La probabilità di ogni valore z_i di $Z = X + Y$ è uguale alla somma delle probabilità (contenute nella tabella) delle possibili coppie di numeri la cui somma è uguale a z_i ; infatti tali coppie corrispondono a eventi incompatibili e z_i corrisponde al loro evento unione. Per esempio, la somma 5 si ottiene con (4 e 1) o (3 e 2), quindi:

$$P(Z = 5) = \frac{2}{15} + \frac{1}{20} = \frac{11}{60}.$$

$X + Y$	4	5	6	8	9	10	11
$P(X + Y)$	$\frac{1}{30}$	$\frac{11}{60}$	$\frac{1}{5}$	$\frac{1}{15}$	$\frac{1}{10}$	$\frac{1}{6}$	$\frac{1}{4}$

● Dati due eventi incompatibili A e B , si ha:

$$p(A \cup B) = p(A) + p(B).$$

II) Consideriamo ancora l'urna contenente una pallina con il numero 3, quattro con il 4, due con il 7 e cinque palline con il 9 e la variabile casuale $X = \text{«numero della pallina estratta»}$.

X	3	4	7	9
$P(X)$	$\frac{1}{12}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{5}{12}$

Estraiano consecutivamente due numeri senza reimmissione.

Indichiamo con Y la variabile casuale relativa alla seconda estrazione.

Vogliamo determinare la distribuzione di probabilità della variabile casuale $X + Y = \text{«somma dei due numeri estratti»}$. I due eventi

$A = \text{«estrazione di un numero dall'urna»}$,

$B|A = \text{«estrazione di un numero dall'urna essendosi verificato } A\text{»}$

sono **dipendenti**, quindi le *probabilità congiunte* sono calcolate moltiplicando la probabilità dell'estrazione del primo numero per quella del secondo nell'ipotesi che il primo numero sia uscito:

$X \backslash Y$	3	4	7	9
3	0	$\frac{1}{33}$	$\frac{1}{66}$	$\frac{5}{132}$
4	$\frac{1}{33}$	$\frac{1}{11}$	$\frac{2}{33}$	$\frac{5}{33}$
7	$\frac{1}{66}$	$\frac{2}{33}$	$\frac{1}{66}$	$\frac{5}{66}$
9	$\frac{5}{132}$	$\frac{5}{33}$	$\frac{5}{66}$	$\frac{5}{33}$

La distribuzione di probabilità della variabile casuale $X + Y$ si ottiene procedendo in modo analogo al caso precedente:

$X + Y$	7	8	10	11	12	13	14	16	18
$P(X + Y)$	$\frac{2}{33}$	$\frac{1}{11}$	$\frac{1}{33}$	$\frac{4}{33}$	$\frac{5}{66}$	$\frac{10}{33}$	$\frac{1}{66}$	$\frac{5}{33}$	$\frac{5}{33}$

Giochi aleatori

La speranza matematica

Una persona decide di partecipare al seguente gioco. Viene lanciato un dado. Se esce il numero 1 essa riceve 7 euro, se esce un numero pari 4 euro, mentre in caso contrario deve pagare 2 euro.

Vengono effettuati 8 lanci e alla fine si rileva che il numero 1 è uscito una volta, un numero pari 2 volte e per 5 volte la persona ha dovuto effettuare il pagamento pattuito.

Calcoliamo se a questa persona è convenuto o meno partecipare al gioco:

$$€ (7 \cdot 1 + 4 \cdot 2 - 2 \cdot 5) = € 5.$$

Il guadagno totale è stato positivo e il guadagno medio è stato di $\frac{5}{8} = 0,625$ euro.

Possiamo ottenere lo stesso valore scrivendo i calcoli nella forma

$$7 \cdot \frac{1}{8} + 4 \cdot \frac{2}{8} - 2 \cdot \frac{5}{8} = 0,625,$$

cioè moltiplicando i guadagni positivi e negativi per le frequenze degli esiti favorevoli e sfavorevoli dei lanci del dado.

Se al posto delle frequenze empiriche (valori a posteriori) sostituiamo le probabilità (valori a priori), abbiamo:

$$7 \cdot \frac{1}{6} + 4 \cdot \frac{1}{2} - 2 \cdot \frac{1}{3} = 2,5.$$

Il valore ottenuto rappresenta non la vincita media riscontrata, ma quella che il giocatore avrebbe potuto sperare di realizzare in media. Questo valore si chiama proprio *speranza matematica*.

I vari guadagni (positivi e negativi) rappresentano i valori di una variabile casuale avente, come distribuzione di probabilità, le probabilità di vincita o di perdita:

<i>X</i>	7	4	−2
<i>P(X)</i>	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$

Consideriamo la distribuzione di probabilità di una variabile casuale X , i cui valori sono i guadagni (positivi e negativi) di un gioco d'azzardo:

X	S_1	S_2	$S_{\dots}$	S_n
$P(X)$	p_1	p_2	$p_{\dots}$	p_n

DEFINIZIONE

Speranza matematica

Si chiama speranza matematica $M(X)$ di X la somma dei prodotti di ogni valore della variabile X per la rispettiva probabilità:

$$M(X) = S_1 \cdot p_1 + S_2 \cdot p_2 + \dots + S_n \cdot p_n.$$

Se $M(X) = 0$, si ha un **gioco equo**.

Se $M(X) > 0$, il **gioco è favorevole** (per il giocatore).

Se $M(X) < 0$, il **gioco è sfavorevole**.

Riprendiamo l'esempio precedente. La speranza matematica del giocatore è 2,5, quindi possiamo dire che il gioco gli è favorevole.

Calcoliamo quanto dovrebbe essere la somma P (cioè la posta) che il giocatore deve corrispondere affinché il gioco sia equo.

Impostiamo la relazione della speranza matematica come equazione con incognita P :

$$7 \cdot \frac{1}{6} + 4 \cdot \frac{1}{2} - P \cdot \frac{1}{3} = 0 \rightarrow P = 9,5.$$

Consideriamo ora lo stesso gioco con un'altra «organizzazione».

Al giocatore viene chiesto un *prezzo* per partecipare al gioco gestito da un *banco*.

Questo prezzo non è altro che la posta da pagarsi prima che il gioco inizi.

In caso di esito favorevole il giocatore incasserà una vincita lorda di:

- 16,5 euro se esce la faccia 1;
- 13,5 euro se esce un numero pari.

La **posta** P che il giocatore deve pagare, affinché il gioco sia equo, è **la somma dei prodotti delle vincite lorde per le rispettive probabilità di vincita**:

$$P = 16,5 \cdot \frac{1}{6} + 13,5 \cdot \frac{1}{2} = 9,5.$$

Parliamo di vincita *lorda* perché, per ottenere il guadagno (ossia la vincita netta), alla somma vinta dobbiamo togliere la posta che viene pagata ogni volta che si gioca:

$$16,5 - 9,5 = 7,$$

$$13,5 - 9,5 = 4.$$

DEFINIZIONE

Speranza matematica di una somma

Si chiama speranza matematica di una somma S il prodotto dell'importo della somma per la probabilità p di ottenerla: $S \cdot p$.

In un gioco dove si conseguono le vincite lorde

$$V_1, V_2, \dots, V_n,$$

con probabilità

$$p_1, p_2, \dots, p_n,$$

si ha equità se la posta P è la somma delle speranze matematiche delle vincite lorde:

$$P = V_1 \cdot p_1 + V_2 \cdot p_2 + \dots + V_n \cdot p_n.$$

ESEMPIO

In un gioco con un mazzo di 40 carte si estraggono contemporaneamente 2 carte. Se sono due figure, si vincono 52 euro, e se è una sola figura, si vincono 26 euro. Per partecipare al gioco viene richiesta una posta di 20 euro.

Le probabilità di vincita sono $\frac{11}{130}$ e $\frac{28}{65}$.

Affiché il gioco sia equo, la posta dovrebbe essere la somma delle speranze matematiche delle vincite lorde:

$$P = 52 \cdot \frac{11}{130} + 26 \cdot \frac{28}{65} = 15,6.$$

Il gioco è sfavorevole essendo la posta richiesta superiore.

Possiamo anche verificare che la speranza matematica dei guadagni, in base alla posta richiesta di 20 euro, è negativa,

$$M(X) = (52 - 20) \cdot \frac{11}{130} + (26 - 20) \cdot \frac{28}{65} - 20 \cdot \frac{63}{130} = -4,4,$$

e il risultato è proprio la differenza fra la posta in caso di equità e la posta richiesta.

Valori caratterizzanti una
variabile casuale discreta

Il valore medio

Consideriamo un'urna che contiene quattro palline con il numero 50, cinque palline con il 54, otto con il 58 e tre con il 62.

Effettuiamo 50 estrazioni di una pallina nelle stesse condizioni, cioè rimettendo ogni volta la pallina estratta nell'urna. Otteniamo la seguente tabella che mostra per ogni valore delle palline il numero delle volte che è uscito:

Valore	Frequenza assoluta
50	10
54	12
58	24
62	4

Se calcoliamo la media aritmetica, otteniamo:

$$M = \frac{50 \cdot 10 + 54 \cdot 12 + 58 \cdot 24 + 62 \cdot 4}{50} = 55,76.$$

Avremmo anche potuto scrivere la media nella forma:

$$M = 50 \cdot \frac{10}{50} + 54 \cdot \frac{12}{50} + 58 \cdot \frac{24}{50} + 62 \cdot \frac{4}{50} = 55,76,$$

dove compaiono le frequenze relative.

Se accompagniamo ogni valore con la relativa probabilità, otteniamo la seguente distribuzione:

X	50	54	58	62
$P(X)$	$\frac{1}{5}$	$\frac{1}{4}$	$\frac{2}{5}$	$\frac{3}{20}$

La somma dei prodotti di ogni valore per la sua probabilità dà un **valore medio** che ha carattere di previsione:

$$M(X) = 50 \cdot \frac{1}{5} + 54 \cdot \frac{1}{4} + 58 \cdot \frac{2}{5} + 62 \cdot \frac{3}{20} = 56.$$

Il valore ottenuto, $\mu = 56$, esprime in modo sintetico la variabile casuale ed è un **valore teorico** che, fra l'altro, nel nostro esempio non può mai essere ottenuto.

DEFINIZIONE

Valore medio

Data la distribuzione di probabilità di una variabile casuale discreta X :

X	x_1	x_2	$x_{\dots}$	x_n
$P(X)$	p_1	p_2	$p_{\dots}$	p_n

il valore medio $M(X)$ di X è la somma dei prodotti di ogni valore assunto dalla variabile casuale per la corrispondente probabilità:

$$M(X) = x_1 \cdot p_1 + x_2 \cdot p_2 + \dots + x_n \cdot p_n = \sum_{i=1}^n x_i \cdot p_i.$$

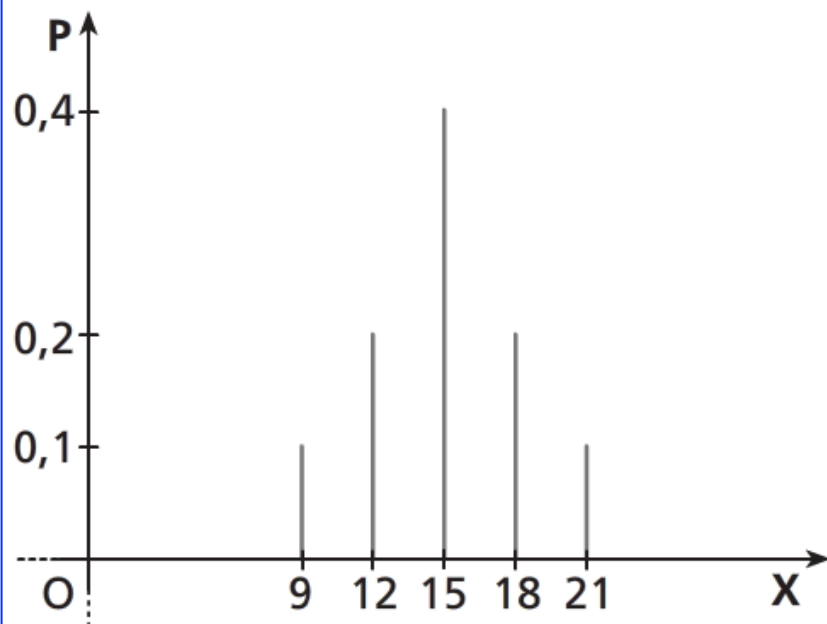
La varianza e la deviazione standard

Sono date le seguenti variabili discrete e le relative distribuzioni di probabilità:

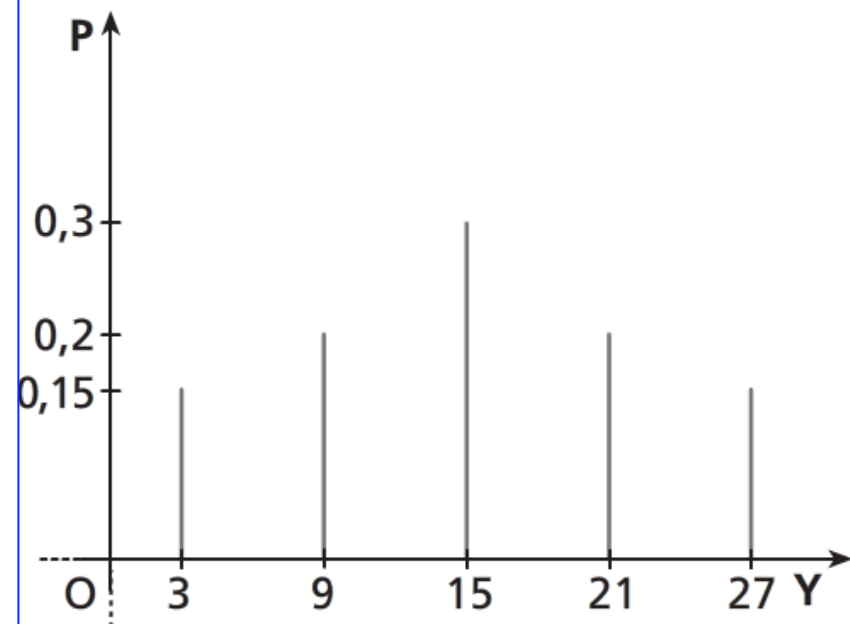
X	9	12	15	18	21
$P(X)$	0,1	0,2	0,4	0,2	0,1

Y	3	9	15	21	27
$P(Y)$	0,15	0,2	0,3	0,2	0,15

Esse hanno la stessa media $M(X) = M(Y) = \mu = 15$. Tuttavia, osserviamo le rappresentazioni grafiche delle distribuzioni di probabilità.



a. Grafico della distribuzione di probabilità di X.



b. Grafico della distribuzione di probabilità di Y.

Vediamo che nel primo caso i valori risultano più vicini al valore medio, pertanto c'è una minore *dispersione*.

Per misurare questa dispersione possiamo considerare gli scarti di ogni valore dal valore medio, creando così la variabile casuale $X - M(X)$, che chiamiamo *variabile casuale scarto*, con la seguente distribuzione di probabilità:

$X - M(X)$	-6	-3	0	3	6
$P[X - M(X)]$	0,1	0,2	0,4	0,2	0,1

Osserviamo che il valore medio della variabile casuale scarto è nullo:

$$M[X - M(X)] = -6 \cdot 0,1 - 3 \cdot 0,2 + 3 \cdot 0,2 + 6 \cdot 0,1 = 0.$$

Allo stesso risultato si giunge se si opera in modo analogo con la variabile Y . Il valore medio dello scarto non è quindi adatto per esprimere la dispersione.

DEFINIZIONE

Variabile casuale scarto

Data una variabile casuale discreta X , si chiama variabile casuale scarto di X la variabile casuale che ha per valori le differenze fra i valori di X e il valore medio $M(X)$, cioè è la variabile $X - M(X)$.

Questa variabile non è molto significativa perché **il suo valore medio è sempre nullo**. Per avere informazioni riguardo alla **dispersione** dei valori di una variabile X , si considera allora la variabile casuale quadrato dello scarto di X .

Consideriamo allora i *quadrati degli scarti dal valore medio*, cioè la variabile $[X - M(X)]^2$:

$[X - M(X)]^2$	36	9	0	9	36
$P([X - M(X)]^2)$	0,1	0,2	0,4	0,2	0,1

Assumiamo come indice della dispersione il valore medio della variabile casuale quadrato degli scarti, che chiamiamo *varianza* e indichiamo con $var(X)$:

$$var(X) = 36 \cdot 0,1 + 9 \cdot 0,2 + 9 \cdot 0,2 + 36 \cdot 0,1 = 10,8.$$

DEFINIZIONE

Varianza

Data la distribuzione di probabilità di una variabile casuale discreta X :

X	x_1	x_2	$x_{\dots}$	x_n
P	p_1	p_2	$p_{\dots}$	p_n

con valore medio $M(X) = \mu$, si chiama varianza $var(X)$ di X il valore medio della variabile casuale quadrato dello scarto di X :

$$\begin{aligned} var(X) &= (x_1 - \mu)^2 \cdot p_1 + (x_2 - \mu)^2 \cdot p_2 + \dots + (x_n - \mu)^2 \cdot p_n = \\ &= \sum_{i=1}^n (x_i - \mu)^2 \cdot p_i. \end{aligned}$$

La varianza ci dice quanto sono concentrati i valori della variabile X attorno al valore medio μ . La varianza si può anche indicare con $\sigma^2(X)$.

Possiamo anche considerare, come indice della dispersione, la radice quadrata della varianza, detta *deviazione standard* $\sigma(X)$:

$$\sigma(X) = \sqrt{\text{var}(X)} = \sqrt{10,8} \simeq 3,29.$$

Per la seconda variabile casuale Y abbiamo i seguenti valori:

$Y - M(Y)$	-12	-6	0	6	12
$[Y - M(Y)]^2$	144	36	0	36	144
P	0,15	0,2	0,3	0,2	0,15

$$\text{var}(Y) = 144 \cdot 0,15 + 36 \cdot 0,2 + 36 \cdot 0,2 + 144 \cdot 0,15 = 57,6;$$

$$\sigma(Y) = \sqrt{57,6} \simeq 7,59.$$

DEFINIZIONE

Deviazione standard

Si chiama deviazione standard $\sigma(X)$ di una variabile casuale X la radice quadrata della varianza di X :

$$\sigma(X) = \sqrt{\text{var}(X)}.$$

Noi possiamo, attraverso il **valore di sintesi** μ e il **valore di dispersione** σ^2 , individuare le **caratteristiche di una variabile casuale**.

TEOREMA

Sia X una variabile casuale, allora la varianza di X è uguale alla differenza tra il valore medio della variabile X^2 e il quadrato del valore medio di X :

$$\text{var}(X) = M(X^2) - [M(X)]^2.$$

Le proprietà del valore medio e della varianza

PROPRIETÀ

Siano X e Y due variabili casuali discrete e k, h due costanti numeriche.

1. Se X è *costante*, cioè assume sempre uno stesso valore a , allora $M(X) = a$;
2. $M(k \cdot X) = k \cdot M(X)$;
3. $M(X + h) = M(X) + h$;
4. $M(X + Y) = M(X) + M(Y)$.

PROPRIETÀ

Siano X e Y due variabili casuali discrete e k, h due costanti numeriche:

1. $\text{var}(k \cdot X) = k^2 \cdot \text{var}(X)$;
2. $\text{var}(X + h) = \text{var}(X)$;
3. se le variabili sono *indipendenti*, $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$.

La semidifferenza tra $\text{var}(X + Y)$ e $[\text{var}(X) + \text{var}(Y)]$ viene detta **covarianza** di X e Y e si indica con **$\text{cov}(X, Y)$** :

$$\text{cov}(X, Y) = \frac{\text{var}(X + Y) - [\text{var}(X) + \text{var}(Y)]}{2}.$$

Ha senso parlare di covarianza solo per le variabili dipendenti, poiché *la covarianza di variabili indipendenti è nulla*.

Distribuzioni di probabilità di variabili casuali discrete

La distribuzione uniforme discreta

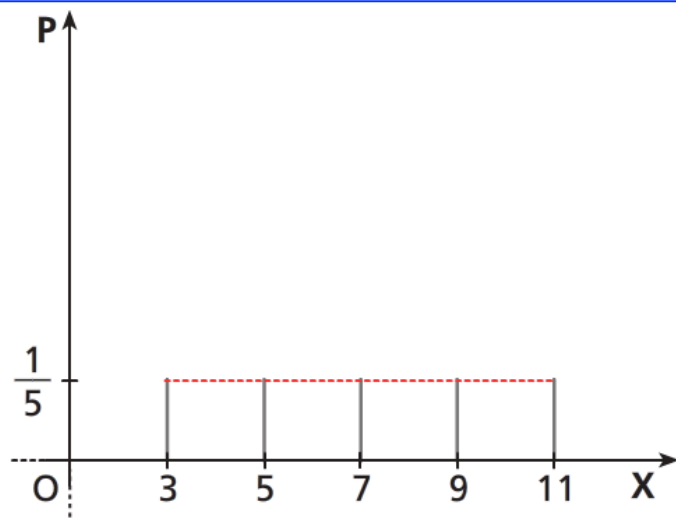
Consideriamo la seguente distribuzione di probabilità relativa a una variabile casuale X_1 e la corrispondente funzione di ripartizione.

X_1	3	5	7	9	11
P	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$
F	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{3}{5}$	$\frac{4}{5}$	1

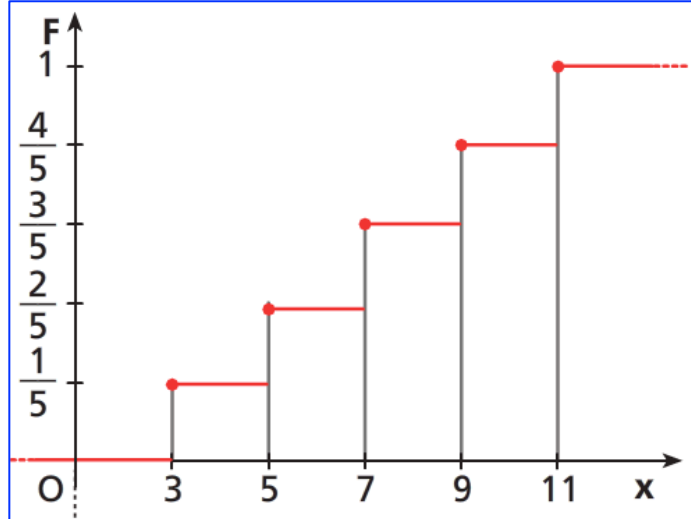
● Per ogni fenomeno specifico si può costruire una sua variabile casuale. Lo statistico ha il compito di esaminare i fenomeni e di classificarli, andando oltre l'apparente diversità, per ricondurli a quelle «regolarità» che li descrivono.

Osserviamo che tutti i valori della variabile hanno la stessa probabilità $\frac{1}{5}$.

In questo caso la distribuzione di probabilità è detta **uniforme** e il suo grafico ha un andamento rettilineo (figura 6a).



a. Grafico della distribuzione di probabilità di X_1 .



b. Grafico della funzione di ripartizione di X_1 .

DEFINIZIONE

Distribuzione uniforme discreta

Si dice che una variabile casuale discreta ha una distribuzione di probabilità uniforme se tutti i suoi valori hanno la stessa probabilità.

Se i valori di una variabile casuale X sono $1, 2, 3, \dots, n$ e tutti hanno probabilità $p = \frac{1}{n}$, si può dimostrare che:

$$M(X) = \frac{n+1}{2} \quad \text{e} \quad \text{var}(X) = \frac{n^2-1}{12}.$$

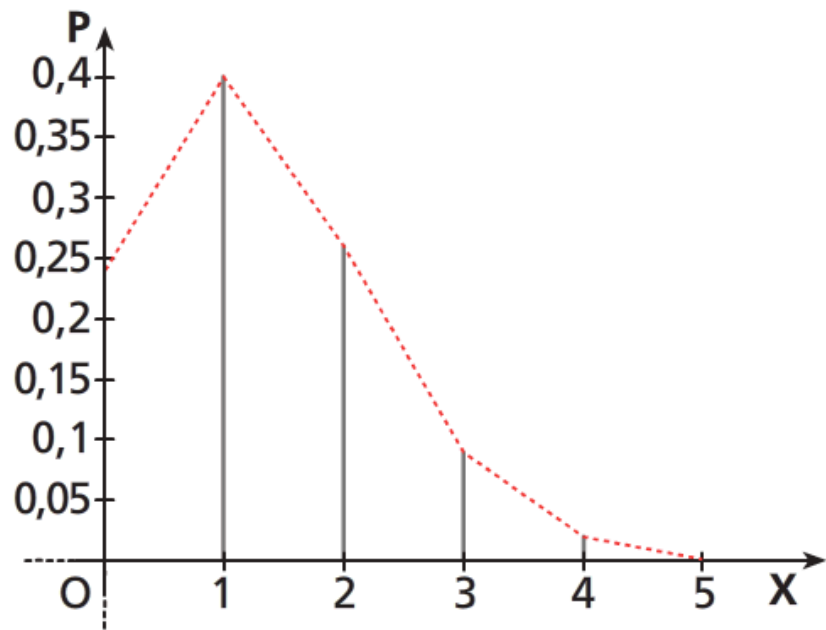
La distribuzione binomiale

Un'urna contiene 8 palline nere e 24 bianche. Si estraggono consecutivamente cinque palline rimettendo ogni volta la pallina estratta nell'urna. Consideriamo l'evento «uscita della pallina nera».

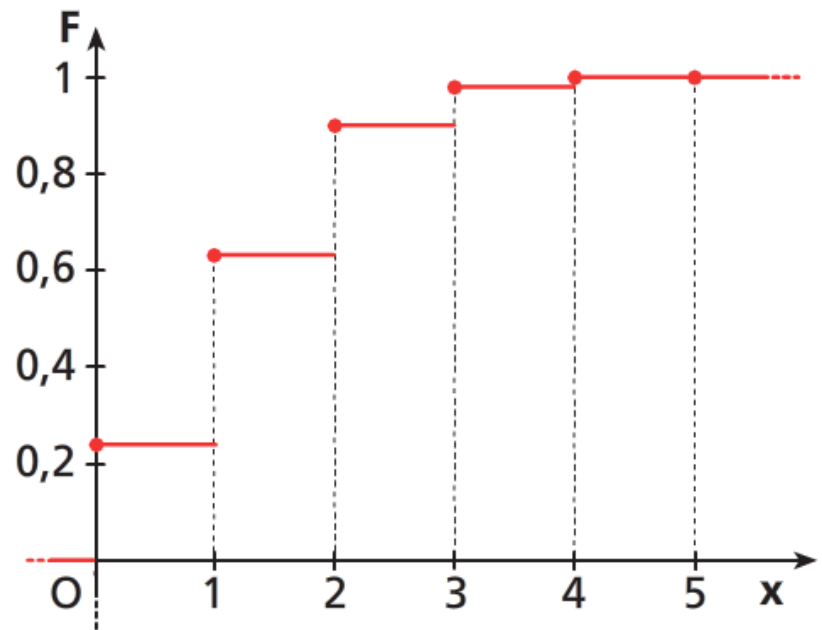
Abbiamo un evento, con probabilità costante $p = \frac{1}{4}$, sottoposto a 5 prove tutte nelle stesse condizioni.

Consideriamo la variabile casuale X corrispondente al numero delle volte in cui l'evento può verificarsi.

X	0	1	2	3	4	5
P	0,237305	0,395508	0,263672	0,087891	0,014648	0,000977
F	0,237305	0,632813	0,896484	0,984375	0,999023	1



a. Grafico della distribuzione di probabilità di X .



b. Grafico della funzione di ripartizione di X .

La distribuzione di probabilità della variabile casuale X corrispondente al numero delle volte in cui l'evento può verificarsi, è data dalla seguente legge (derivante dal modello probabilistico delle prove ripetute):

DEFINIZIONE

Distribuzione binomiale (o di Bernoulli)

Si dice che una variabile casuale discreta X , con valori $x = 0, 1, 2, \dots, n$, ha una distribuzione di probabilità binomiale di *parametri* n e p se:

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}.$$

Se consideriamo **n prove**, esse corrispondono alla *somma di n variabili casuali indipendenti* e pertanto si hanno i seguenti valori:

$$M(X) = np \quad (\text{valore medio})$$

$$\text{var}(X) = npq = np(1 - p) \quad (\text{varianza})$$

$$\sigma(X) = \sqrt{npq} = \sqrt{np(1 - p)} \quad (\text{deviazione standard})$$

Una variabile casuale con distribuzione binomiale descrive il numero di volte che si può verificare un evento aleatorio di probabilità p su n prove.

ESEMPIO

Una macchina di precisione produce pezzi di ricambio per elettrodomestici con una percentuale del 2% di pezzi difettosi.

Calcoliamo il numero medio di pezzi difettosi che si possono prevedere su una produzione giornaliera di 400 pezzi e la deviazione standard.

Essendo $n = 400$, $p = 0,02$ e $q = 0,98$, otteniamo:

$$\mu = 400 \cdot 0,02 = 8 \quad \text{e} \quad \sigma = \sqrt{400 \cdot 0,02 \cdot 0,98} = 2,8.$$

■ La distribuzione di Poisson

Un evento ha probabilità $p = 0,005$. Calcoliamo la probabilità che su 800 prove, *effettuate nelle stesse condizioni*, l'evento si verifichi 6 volte.

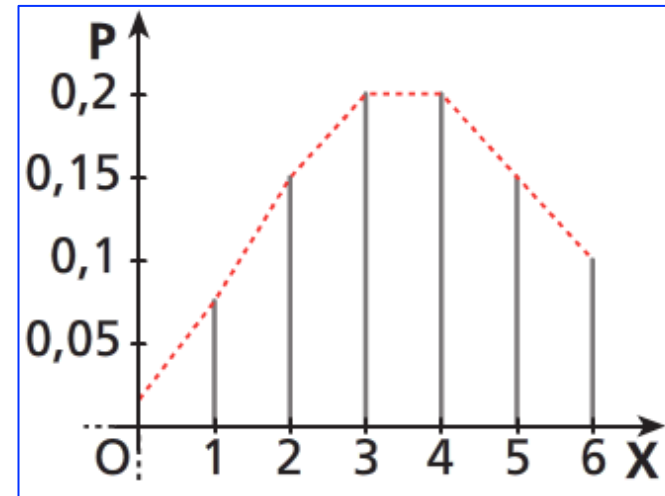
Per calcolare la probabilità richiesta dobbiamo applicare lo schema delle prove ripetute:

$$p_{(6,800)} = \binom{800}{6} \cdot 0,005^6 \cdot 0,995^{794} \simeq 0,10433.$$

Al variare di $x = 0, 1, 2, 3, 4, 5, 6$, otteniamo la seguente distribuzione di probabilità:

X	0	1	2	3	4	5	6
P	0,018316	0,073263	0,146525	0,195367	0,195367	0,156293	0,104196

Grafico della distribuzione



Quando, come in questo caso, n è elevato e p piccolo, per ottenere un valore approssimato del precedente si può utilizzare la seguente **distribuzione di probabilità** dovuta a **Poisson**:

DEFINIZIONE

Distribuzione di Poisson

Si dice che una variabile casuale discreta X , con valori $x = 0, 1, 2, \dots, n$, ha una distribuzione di probabilità di Poisson di *parametro* λ se:

$$P(X = x) = \frac{\lambda^x}{x!} \cdot e^{-\lambda}, \quad \text{con } \lambda = np$$

dove λ è il numero medio di eventi per intervallo di tempo.

Utilizzando questa formula, otteniamo $\lambda = 800 \cdot 0,005 = 4$ e:

$$P(X = 6) = \frac{4^6}{6!} e^{-4} \simeq 0,104196,$$

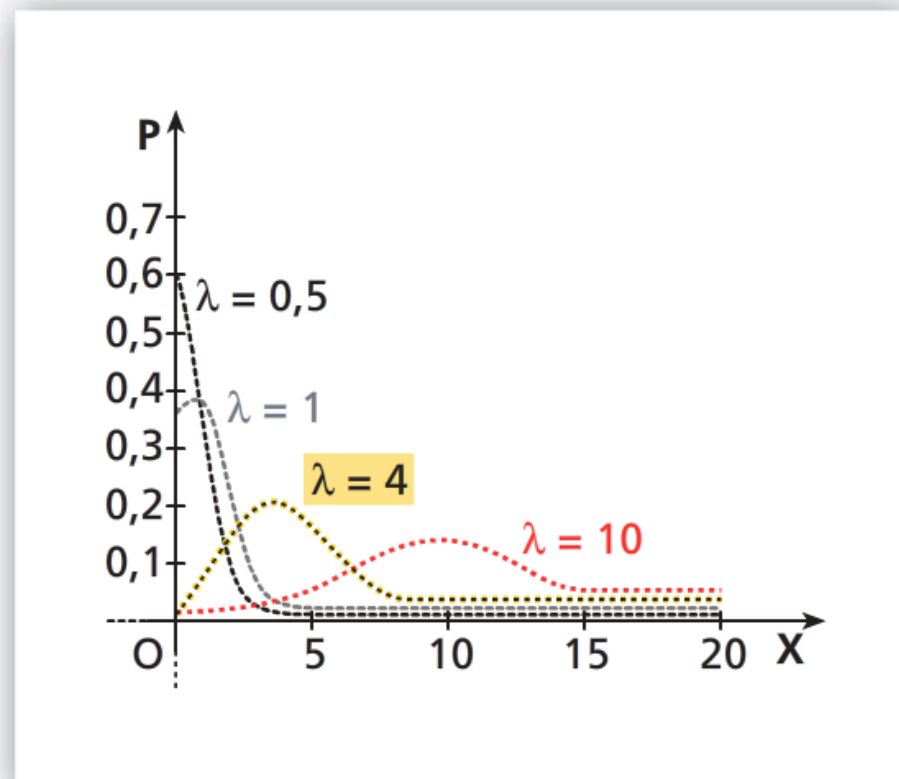
che è una buona approssimazione del valore precedente.

Si può provare che l'approssimazione diventa migliore al crescere di n .

Il parametro λ di una distribuzione di Poisson assume particolare interesse perché si può dimostrare che, per una variabile casuale discreta X con questa distribuzione di probabilità, esso coincide con il valore medio e la varianza della variabile:

$$M(X) = \text{var}(X) = \lambda.$$

Confrontiamo i valori della distribuzione per $\lambda = 0,5, 1, 4, 10$.



Osserviamo che all'aumentare del valore di λ la distribuzione della probabilità tende a diventare simmetrica attorno al valore medio.

La distribuzione di probabilità di Poisson è un modello teorico che rappresenta importanti fenomeni in fisica (decadimento radioattivo), in medicina (malattie rare), in campo economico (file di attesa).

ESEMPIO

Una macchina produce pezzi difettosi con una probabilità $p = 0,006$.
Calcoliamo la probabilità che su 500 pezzi:

- nessuno risulti difettoso;
- risultino difettosi 3 pezzi;
- risultino difettosi più di 5 pezzi.

La variabile casuale relativa al numero di pezzi difettosi ha valore medio e distribuzione di probabilità:

$$\mu = \lambda = 500 \cdot 0,006 = 3, \quad P(X = x) = \frac{3^x}{x!} e^{-3}.$$

La probabilità che nessun pezzo sia difettoso è:

$$P(X = 0) = \frac{3^0}{0!} e^{-3} \simeq 0,04979.$$

La probabilità che i pezzi difettosi siano 3 è $P(X = 3) = \frac{3^3}{3!} e^{-3} \simeq 0,22404$.

Per l'ultimo quesito è necessario calcolare tutti i valori fino a $x = 5$. Nella tabella seguente, indichiamo anche i valori della funzione di ripartizione:

X	0	1	2	3	4	5
P	0,049787	0,149361	0,224042	0,224042	0,168031	0,100819
F	0,049787	0,199148	0,42319	0,647232	0,815263	0,916082

La probabilità che vi siano più di 5 pezzi difettosi è:

$$P(X > 5) = 1 - F(5) \simeq 1 - 0,916082 = 0,083918.$$

ESEMPIO

A uno sportello bancario arrivano in media 30 persone all'ora.
Calcoliamo la probabilità che in 5 minuti:

- arrivino 4 persone;
- arrivino meno di 3 persone.

Calcoliamo in media quante persone arrivano nell'unità di tempo, cioè in 5 minuti:

La distribuzione di probabilità è $P(X = x) = \frac{2,5^x}{x!} e^{-2,5}$: $\lambda = 2,5$

<i>X</i>	0	1	2	3	4
<i>P</i>	0,082085	0,205212	0,256516	0,213763	0,133602
<i>F</i>	0,082085	0,287297	0,543813	0,757576	0,891178

Pertanto $P(X = 4) \simeq 0,133602$.

La probabilità che arrivino meno di 3 persone è $P(X \leq 2) = F(2) \simeq 0,543813$.

Variabili casuali
standardizzate

Definizioni

Le seguenti variabili casuali discrete mostrano il punteggio di due test sottoposti ai 23 alunni di una classe, in momenti successivi. I test si differenziano per il numero dei quesiti, e quindi per il punteggio massimo ottenibile, e per le difficoltà:

X_1	0	1	2	3	4	5	6	7	8
P	$\frac{2}{23}$	$\frac{2}{23}$	$\frac{3}{23}$	$\frac{6}{23}$	$\frac{3}{23}$	$\frac{3}{23}$	$\frac{2}{23}$	$\frac{1}{23}$	$\frac{1}{23}$

X_2	0	1	2	3	4	5
P	$\frac{2}{23}$	$\frac{4}{23}$	$\frac{5}{23}$	$\frac{5}{23}$	$\frac{4}{23}$	$\frac{3}{23}$

Le due variabili casuali hanno rispettivamente valori medi:

$$\mu_1 \simeq 3,478 \quad \text{e} \quad \mu_2 \simeq 2,609.$$

Consideriamo un alunno che ha conseguito 4 punti nel primo test e 3 nel secondo e gli scarti dalle medie dei suoi punteggi:

$$4 - \mu_1 \simeq 0,522; \quad 3 - \mu_2 \simeq 0,391.$$

Potremmo concludere che l'alunno ha avuto un peggioramento, in quanto con il primo test ha uno scarto positivo, rispetto alla media della classe, maggiore di quello del secondo test.

La differenza con la media non è però sufficiente a valutare la situazione, in quanto non tiene conto della diversa difficoltà dei test.

La difficoltà è misurata dalla variabilità dei risultati, e pertanto rapportiamo le differenze dal valore medio con le deviazioni standard, cioè misuriamo gli scarti dai valori medi in unità di σ .

Essendo: $\sigma_1 \simeq 2,061$ e $\sigma_2 \simeq 1,496$,

si hanno: $\frac{4 - \mu_1}{\sigma_1} \simeq 0,253$ e $\frac{3 - \mu_2}{\sigma_2} \simeq 0,261$.

Dobbiamo concludere che, per il secondo test, l'alunno ha conseguito un miglioramento rispetto alla situazione complessiva della classe.

Se operiamo la stessa trasformazione per tutti i valori della variabile casuale X_1 , otteniamo una **nuova variabile casuale Z_1** , detta **standardizzata**, che ha la stessa **distribuzione di probabilità**:

$$Z_1 = \frac{X_1 - \mu_1}{\sigma_1} \simeq \frac{X_1 - 3,478}{2,061}.$$

Z_1	-1,688	-1,202	-0,717	-0,232	0,253	0,738	1,224	1,709	2,194
P	$\frac{2}{23}$	$\frac{2}{23}$	$\frac{3}{23}$	$\frac{6}{23}$	$\frac{3}{23}$	$\frac{3}{23}$	$\frac{2}{23}$	$\frac{1}{23}$	$\frac{1}{23}$

I valori assunti dalla variabile standardizzata Z di una variabile casuale X vengono chiamati **punti zeta** e hanno la stessa distribuzione di probabilità di X .

DEFINIZIONE

Variabile casuale standardizzata

Data una variabile casuale X con valore medio μ e deviazione standard σ , si chiama standardizzata di X la variabile casuale:

$$Z = \frac{X - \mu}{\sigma}.$$

● Effettuando la standardizzazione di una variabile casuale si possono fare confronti tra fenomeni descritti da grandezze diverse, cioè da variabili aventi unità di misura diverse, in quanto la variabile Z è indipendente dall'unità di misura.

PROPRIETÀ

Una variabile standardizzata Z ha valore medio nullo e varianza e deviazione standard unitarie:

$$M(Z) = 0, \quad \text{var}(Z) = \sigma(Z) = 1.$$

Variabili casuali continue

Definizioni

Consideriamo un esperimento in cui abbiamo un cronometro che fermiamo, in modo casuale, premendo un pulsante. Vogliamo determinare la probabilità che venga fermato, per esempio, esattamente dopo 8 secondi. La grandezza legata al fenomeno aleatorio è il tempo, che varia con continuità: a seconda della sensibilità del cronometro, possiamo pensare di considerare, oltre ai secondi, i decimi di secondo, i centesimi, e così via. Se consideriamo un intervallo di tempo, per esempio fra 0 e 10 secondi, a ogni istante possiamo far corrispondere un numero reale e viceversa. In generale, diamo la seguente definizione.

DEFINIZIONE

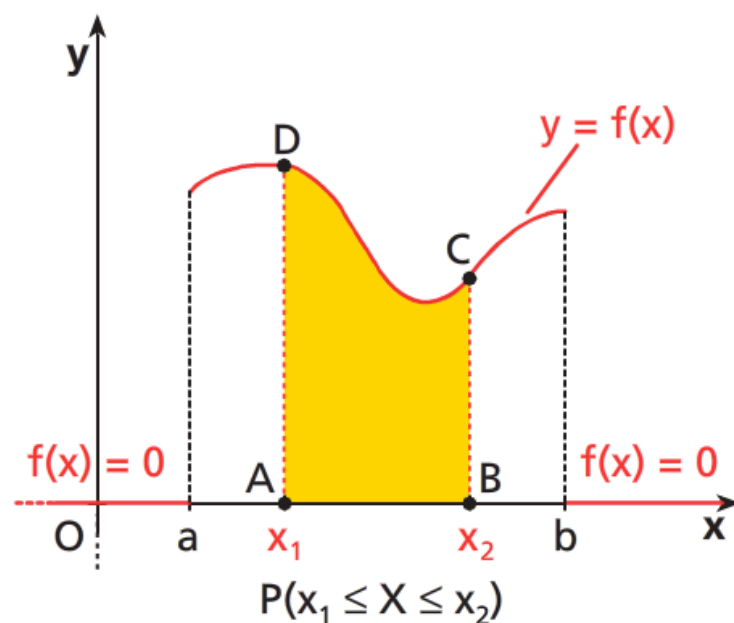
Variabile casuale continua

Una variabile casuale continua X è una grandezza che può assumere tutti i valori reali contenuti in un intervallo.

Il modello matematico che serve per descrivere le probabilità associate a una variabile casuale continua X che varia entro un intervallo $I = [a; b]$, dove $I \subset \mathbb{R}$, non si basa quindi sulle probabilità dei singoli valori di X , bensì sulla probabilità che X assuma valori compresi fra due estremi $x_1, x_2 \in I$.

● Per esempio, nel caso del cronometro, potremmo chiederci qual è la probabilità di fermarlo in un istante compreso fra 7 e 9 secondi.

Per il calcolo utilizziamo una funzione $f(x) \geq 0$, definita punto per punto in $\mathbb{R}$, chiamata **funzione densità di probabilità di X** : il valore della probabilità $P(x_1 \leq X \leq x_2)$ è **uguale all'area compresa fra il grafico di $f(x)$ e l'asse delle ascisse nell'intervallo $[x_1; x_2]$** (figura 10).



► **Figura 10** La probabilità $P(x_1 \leq X \leq x_2)$ è uguale all'area del trapezoide ABCD.

DEFINIZIONE

Funzione densità di probabilità

La funzione densità di probabilità di una variabile casuale continua X è una funzione $f(x)$ tale che:

$$f(x) \geq 0, \quad \forall x \in \mathbb{R}; \quad \int_{-\infty}^{+\infty} f(x) dx = 1.$$

Dalla definizione osserviamo che quando, come nel caso della figura, l'intervallo I in cui varia X ha estremi finiti a e b , la funzione densità di probabilità ha valore:

$$f(x) = 0, \quad \text{per } x < a \text{ e per } x > b.$$

In questo modo l'area complessiva compresa tra il grafico di $f(x)$ e l'asse delle ascisse coincide con quella relativa all'intervallo $I = [a; b]$.

Il fatto che l'area valga 1 indica che la variabile X assume sicuramente un valore compreso fra a e b . Per un intervallo $[x_1; x_2] \subseteq [a; b]$ possiamo scrivere:

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f(x) dx, \quad \text{dove } x_1, x_2 \in I.$$

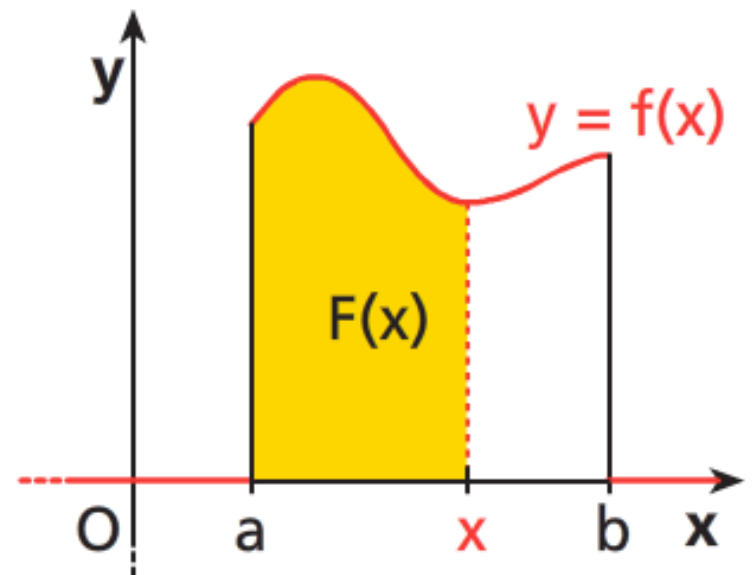
Analogamente al caso discreto, possiamo introdurre il concetto di funzione di ripartizione.

DEFINIZIONE

Funzione di ripartizione

Si chiama funzione di ripartizione di una variabile casuale continua X la funzione $F(x)$ che fornisce la probabilità che il valore della variabile X non superi un determinato valore x :

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt.$$



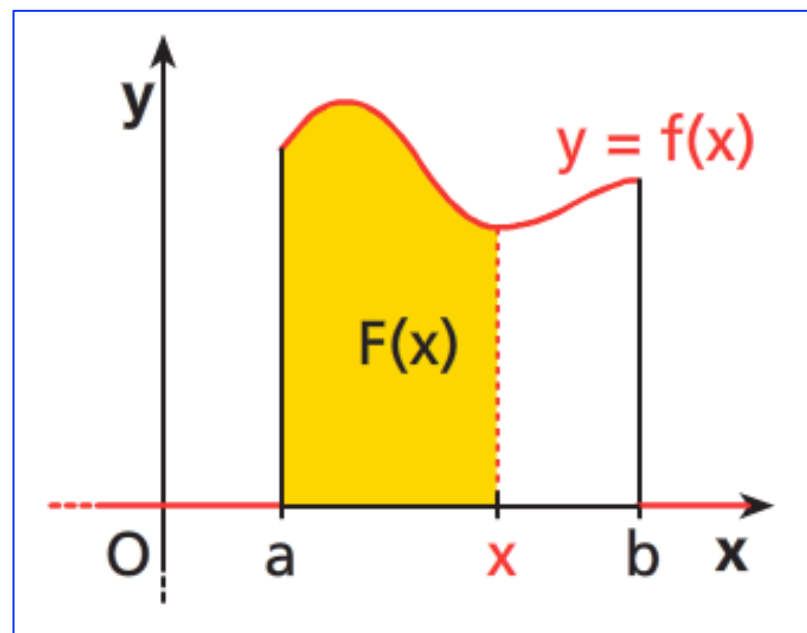
Dalla definizione osserviamo che:

- $F(x)$ è uguale all'area compresa tra il grafico della funzione densità di probabilità $f(x)$ e l'asse delle ascisse nell'intervallo $]-\infty; x]$;
- essa è una primitiva della funzione densità di probabilità $f(x)$, cioè:

$$F'(x) = \frac{dF(x)}{dx} = f(x);$$

- possiamo utilizzare $F(x)$ per calcolare $P(x_1 \leq X \leq x_2)$ applicando la relazione

$$P(x_1 \leq X \leq x_2) = F(x_2) - F(x_1).$$



Anche per le variabili casuali continue possiamo parlare di valore medio, varianza e deviazione standard. Le formule sono l'estensione nel continuo di quelle per le variabili casuali discrete; l'integrale sostituisce la sommatoria.

Valore medio:

$$M(x) = \int_{-\infty}^{+\infty} x \cdot f(x) dx$$

Varianza:

$$var(X) = \int_{-\infty}^{+\infty} [x - M(x)]^2 f(x) dx$$

Deviazione standard

$$\sigma(X) = \sqrt{var(X)}$$

ESEMPIO

Una variabile casuale continua X varia nell'intervallo $[0; 2]$ e la sua funzione densità di probabilità è:

$$f(x) = \begin{cases} \frac{x^3}{4} & \text{se } 0 \leq x \leq 2 \\ 0 & \text{se } x < 0 \text{ o } x > 2 \end{cases}$$

Controlliamo che si tratti di una funzione densità di probabilità:

$$f(x) = \frac{x^3}{4} \geq 0 \quad \text{nell'intervallo } [0; 2],$$

$$\int_0^2 \frac{x^3}{4} dx = \left[\frac{x^4}{16} \right]_0^2 = 1.$$

Determiniamo la funzione di ripartizione:

$$F(x) = \begin{cases} 0 & \text{se } x < 0 \\ \int_0^x \frac{t^3}{4} dt = \left[\frac{t^4}{16} \right]_0^x = \frac{1}{16} x^4 & \text{se } 0 \leq x \leq 2 \\ 1 & \text{se } x > 2 \end{cases}$$

Calcoliamo la probabilità che X assuma un valore compreso tra 1 e 1,5,

$$P(1 \leq X \leq 1,5) = \int_1^{1,5} \frac{x^3}{4} dx = \left[\frac{x^4}{16} \right]_1^{1,5} = \frac{1}{16} (5,0625 - 1) \simeq 0,2539,$$

oppure, utilizzando la funzione di ripartizione:

$$P(1 \leq X \leq 1,5) = F(1,5) - F(1) = \frac{1}{16} 5,0625 - \frac{1}{16} \simeq 0,2539.$$

Il valore medio è:

$$M(X) = \int_0^2 x \cdot \frac{x^3}{4} dx = \left[\frac{x^5}{20} \right]_0^2 = \frac{32}{20} = \frac{8}{5}.$$

La varianza è:

$$\begin{aligned} var(X) &= \int_0^2 \left(x - \frac{8}{5} \right)^2 \frac{x^3}{4} dx = \\ &= \int_0^2 \left(\frac{1}{4} x^5 - \frac{4}{5} x^4 + \frac{16}{25} x^3 \right) dx = \\ &= \left[\frac{1}{24} x^6 - \frac{4}{25} x^5 + \frac{4}{25} x^4 \right]_0^2 = \frac{8}{75}. \end{aligned}$$

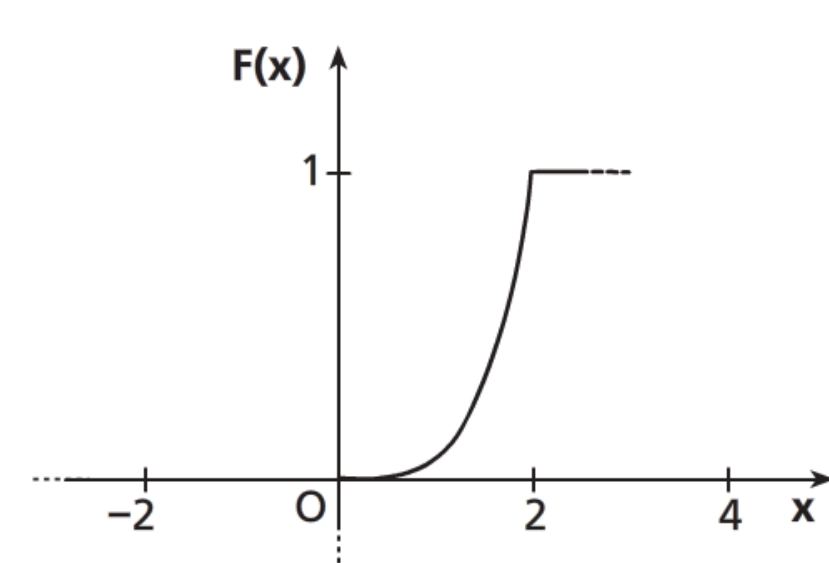
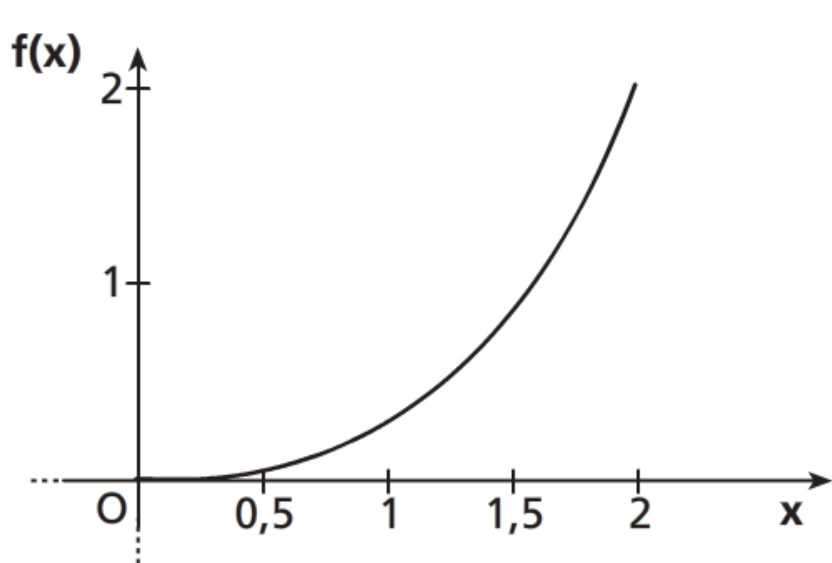
Più semplicemente applicando la proprietà:

$$var(X) = M(X^2) - [M(X)]^2$$

$$\begin{aligned} \text{var}(X) &= \int_0^2 x^2 \cdot \frac{x^3}{4} dx - \left(\frac{8}{5}\right)^2 = \int_0^2 \frac{x^5}{4} dx - \frac{64}{25} = \\ &= \left[\frac{x^6}{24} \right]_0^2 - \frac{64}{25} = \frac{64}{24} - \frac{64}{25} = \frac{8}{75}. \end{aligned}$$

Il grafico della funzione densità di probabilità nell'intervallo $[0; 2]$ è quello di figura 11a.

Il grafico della funzione di ripartizione è quello di figura 11b.



a

b

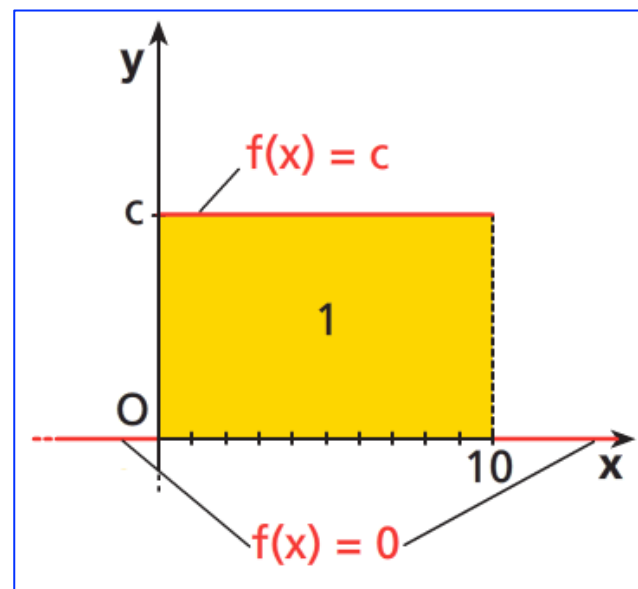
Distribuzioni di probabilità di variabili casuali continue

■ La distribuzione uniforme continua

Consideriamo una variabile casuale continua X definita nell'intervallo $[0; 10]$, che abbia una funzione densità di probabilità del tipo:

$$f(x) = \begin{cases} 0 & \text{se } x < 0 \\ c & \text{se } 0 \leq x \leq 10 \\ 0 & \text{se } x > 10 \end{cases}$$

dove c è un numero reale positivo costante.



Per la definizione di funzione densità di probabilità, l'area nell'intervallo $[0; 10]$ compresa tra questa retta e l'asse delle ascisse, cioè l'area del rettangolo di base 10 e altezza c , deve essere uguale a 1, pertanto:

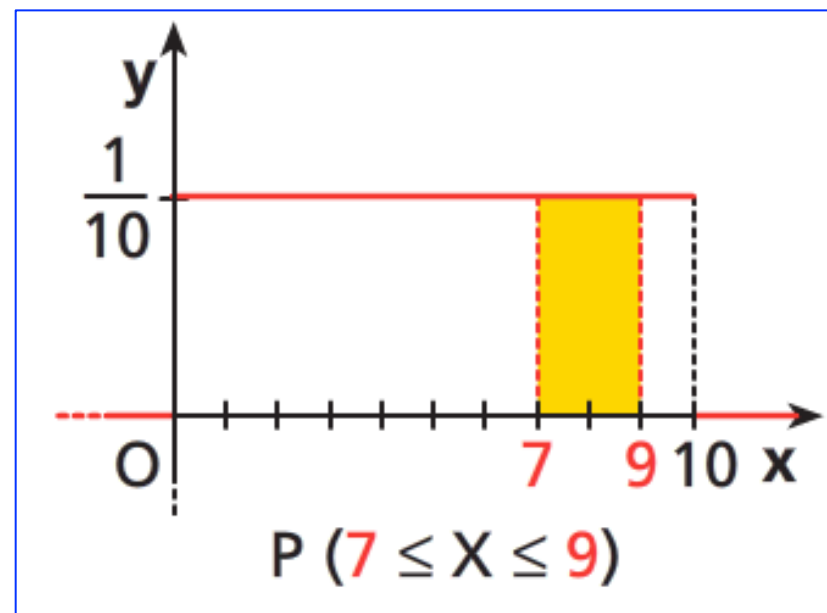
$$10 \cdot c = 1 \rightarrow c = \frac{1}{10}.$$

$$\text{Quindi } f(x) = \frac{1}{10} \text{ per } 0 \leq x \leq 10.$$

Una distribuzione di questo tipo è detta *uniforme*.

La probabilità che la variabile X assuma valori contenuti in un intervallo $[x_1; x_2]$, per esempio $x_1 = 7$ e $x_2 = 9$, è:

$$P(7 \leq X \leq 9) = \int_7^9 \frac{1}{10} dx = \left[\frac{x}{10} \right]_7^9 = \frac{1}{5}.$$

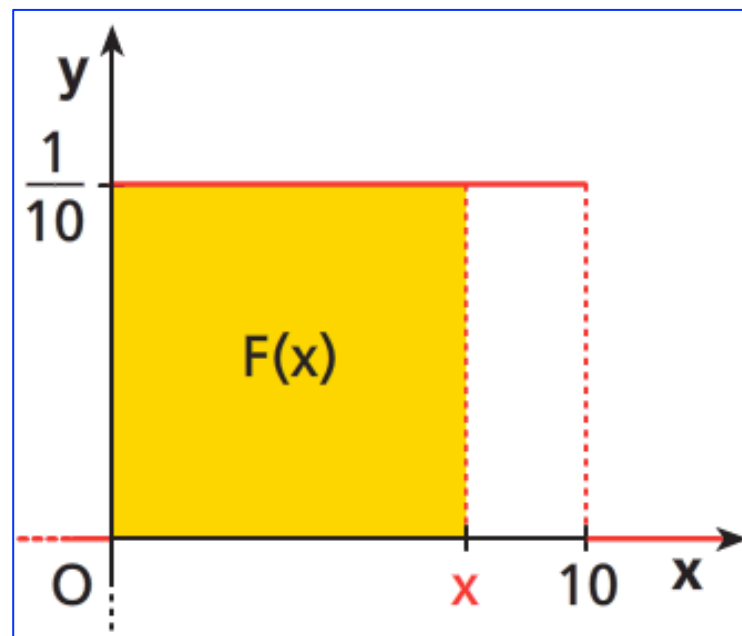


Se consideriamo invece l'intervallo $[0; x]$, otteniamo l'espressione della funzione di ripartizione:

$$F(x) = P(0 \leq X \leq x) = \int_0^x \frac{1}{10} dt = \left[\frac{t}{10} \right]_0^x = \frac{x}{10}, \quad \text{per } 0 \leq x \leq 10.$$

Inoltre si ha:

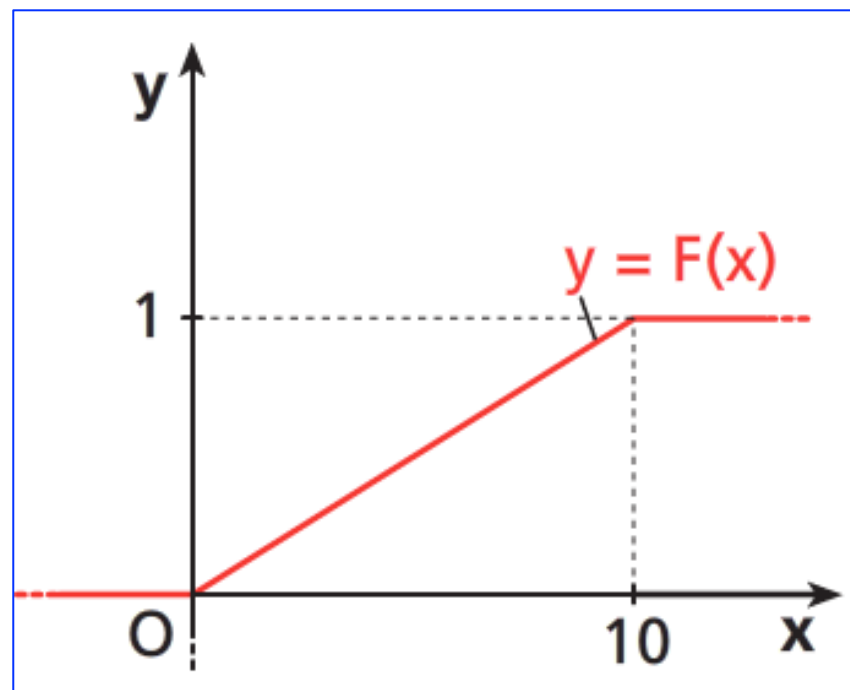
$$F(x) = 0, \quad \text{per } x < 0; \quad F(x) = 1, \quad \text{per } x > 10.$$



Utilizzando la funzione di ripartizione, possiamo calcolare in modo alternativo la probabilità che X assuma un valore compreso fra 7 e 9:

$$P(7 \leq X \leq 9) = F(9) - F(7) = \frac{9}{10} - \frac{7}{10} = \frac{1}{5}.$$

Grafico funzione di ripartizione



In generale abbiamo la seguente definizione.

DEFINIZIONE

Distribuzione uniforme continua

Si dice che una variabile casuale continua X , definita in $[a; b]$, ha una distribuzione uniforme se la sua funzione densità di probabilità è:

$$f(x) = \begin{cases} 0 & \text{se } x < a \\ c & \text{se } a \leq x \leq b \\ 0 & \text{se } x > b \end{cases}$$

$$\text{con } c = \frac{1}{b - a}.$$

Valore medio:

$$M(X) = \frac{a + b}{2}.$$

Varianza:

$$\text{var}(X) = \frac{(b - a)^2}{12}$$

La probabilità che X assuma valori contenuti nell'intervallo $[x_1; x_2]$ (con $a \leq x_1 \leq x_2 \leq b$) è:

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} \frac{1}{b-a} dx = \left[\frac{x}{b-a} \right]_{x_1}^{x_2} = \frac{x_2 - x_1}{b-a}.$$

La funzione di ripartizione di X è:

$$F(x) = \begin{cases} 0 & \text{se } x < a \\ \frac{x-a}{b-a} & \text{se } a \leq x \leq b \\ 1 & \text{se } x > b \end{cases}$$

La distribuzione normale o gaussiana

La distribuzione normale è una funzione densità di probabilità che è stata individuata come il modo con il quale si distribuiscono le misure ripetute, che differiscono fra loro per motivi accidentali, di una stessa grandezza X .

DEFINIZIONE

Distribuzione normale (o di Gauss)

Una funzione densità di probabilità $f(x)$ si dice normale se è definita in $\mathbb{R}$ e ha espressione:

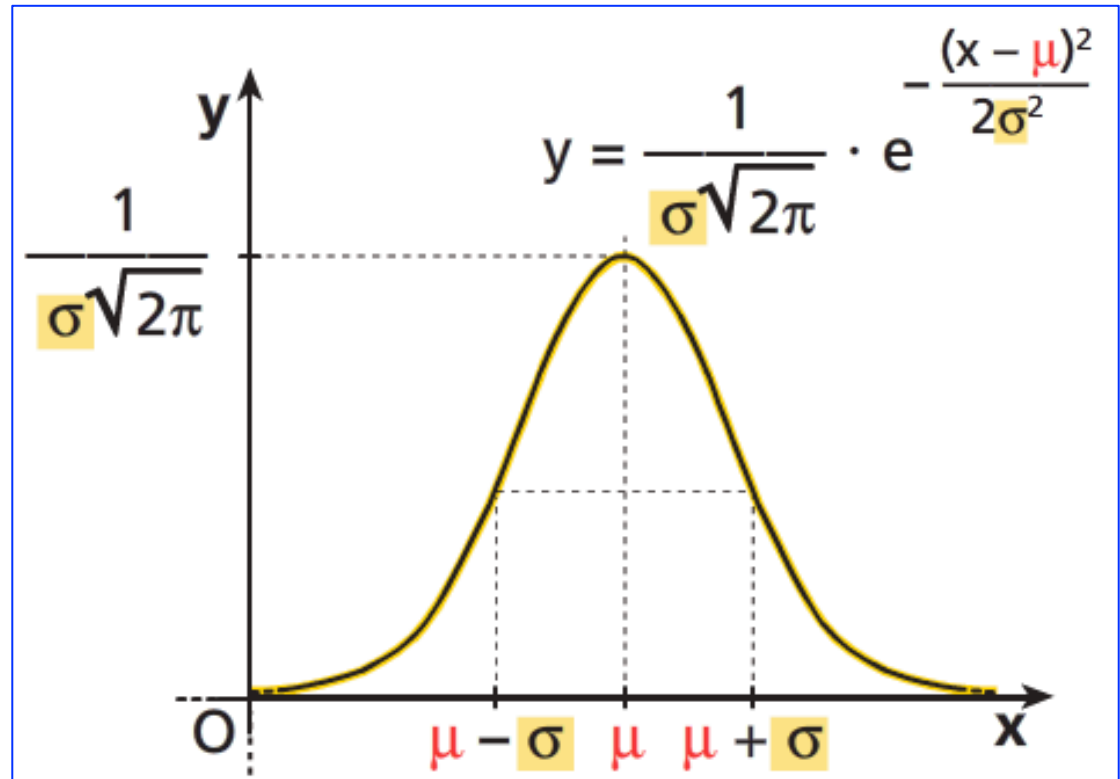
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

dove i parametri μ e σ sono due costanti reali positive.

Indichiamo con $N(\mu; \sigma^2)$ una variabile casuale continua con una funzione densità di probabilità normale e la chiamiamo **variabile casuale normale**. I parametri μ e σ coincidono rispettivamente con il valore medio e la deviazione standard della variabile casuale.

La curva che rappresenta la funzione gaussiana (figura 12) è chiamata *curva degli errori accidentali* o, dalla sua forma dovuta alla simmetria, *curva a campana*.

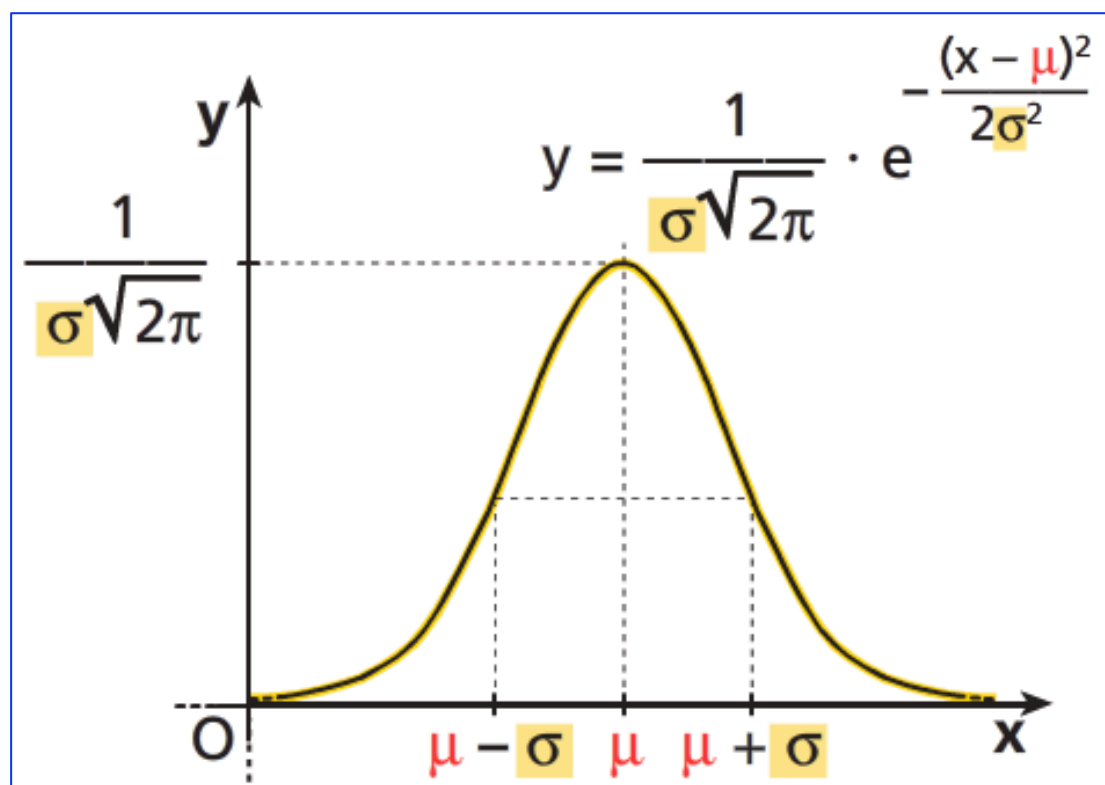
● Sono molto frequenti i fenomeni collettivi naturali, sociali e produttivi con valori che si distribuiscono *normalmente*.



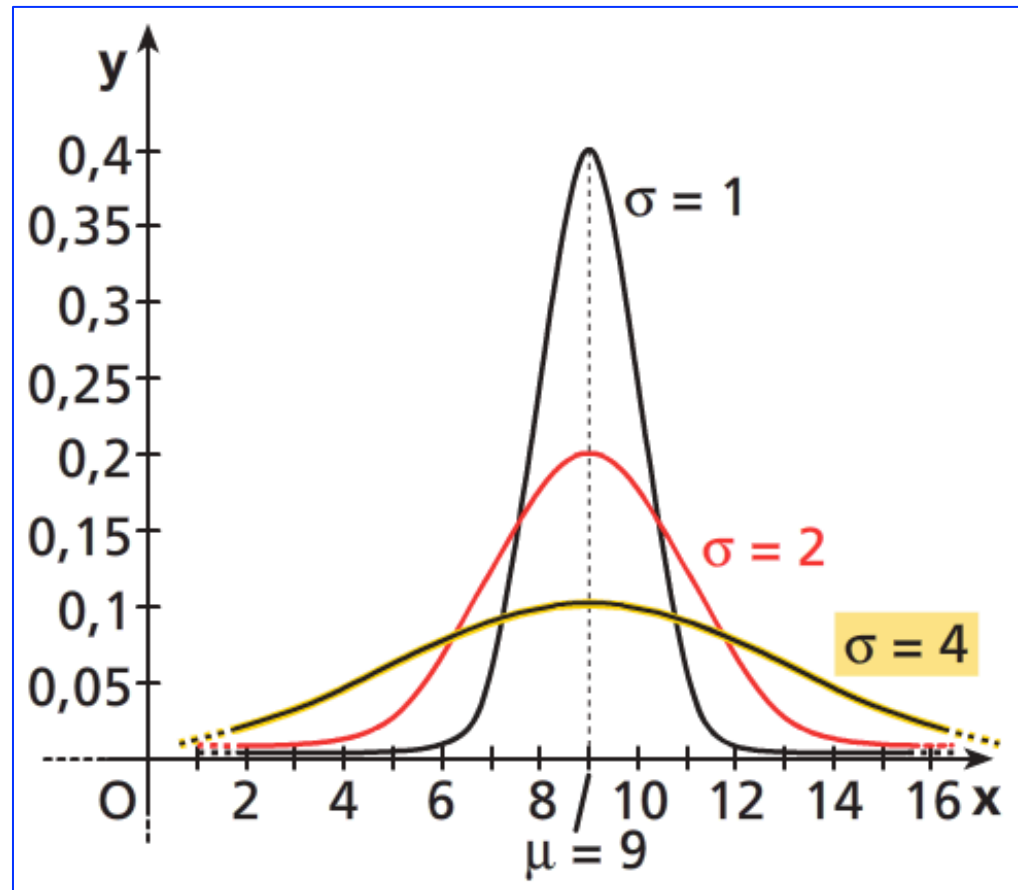
Curva di Gauss (gaussiana)

Sottolineiamo alcune caratteristiche di questa curva:

- è simmetrica rispetto all'asse di equazione $x = \mu$;
- la curva ha un massimo in corrispondenza del punto $\left(\mu; \frac{1}{\sigma\sqrt{2\pi}}\right)$;
- ha l'asse delle ascisse come asintoto orizzontale;
- presenta due punti di flesso in corrispondenza di $\mu - \sigma$ e $\mu + \sigma$.



Se consideriamo più variabili casuali normali con uguale valore medio μ , ma diverso valore di σ , rileviamo che la curva si appiattisce all'aumentare della deviazione standard.



Calcoliamo la probabilità che la variabile casuale normale $X = N(9; 4)$ assuma valori compresi nell'intervallo $[10; 12]$.

Per semplificare il calcolo, *standardizziamo* la funzione densità, cioè trasformiamo la variabile X nella variabile Z :

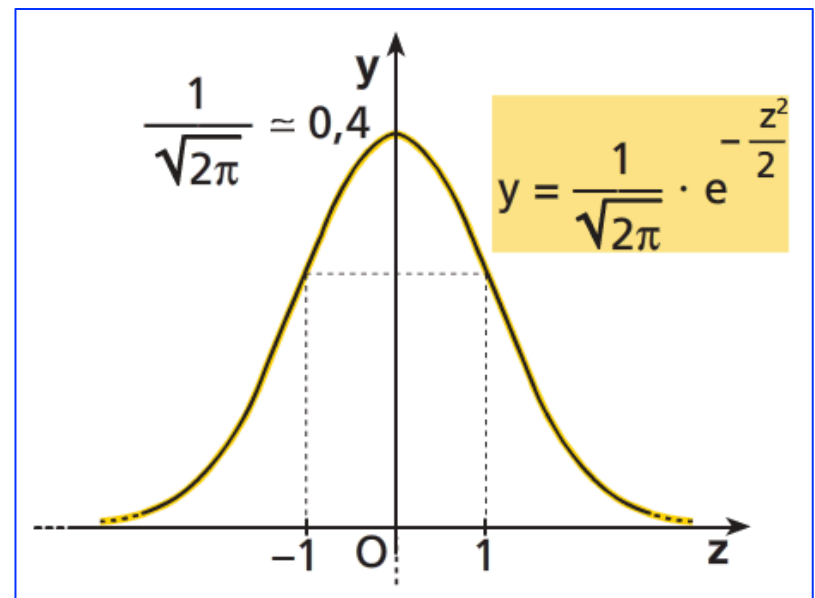
$$Z = \frac{X - \mu}{\sigma}.$$

In questo modo abbiamo ottenuto ancora una *variabile casuale normale*, ma con valore medio $\mu = 0$ e deviazione standard $\sigma = 1$.

Quindi **la variabile casuale standardizzata di una variabile casuale normale è $N(0; 1)$** , a cui corrisponde la funzione gaussiana:

$$f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}}$$

Curva di Gauss standardizzata



Il calcolo della probabilità che la variabile $N(9; 4)$ assuma un valore compreso tra 10 e 12 si riconduce al calcolo della probabilità che la variabile casuale standardizzata $Z = N(0; 1)$ assuma un valore compreso tra 0,5 e 1,5 essendo questi i due valori di Z corrispondenti ai precedenti.

Infatti, applicando la relazione $Z = \frac{X - 9}{2}$, si ha:

$$x_1 = 10 \quad \rightarrow \quad z_1 = \frac{10 - 9}{2} = 0,5;$$

$$x_2 = 12 \quad \rightarrow \quad z_2 = \frac{12 - 9}{2} = 1,5.$$

Esiste una tavola apposita (**tavola di Sheppard**) che fornisce il valore delle aree sottostanti la curva $f(z)$ nell'intervallo $[0; z]$, cioè:

$$F(z) = P(0 < Z < z).$$

Tavola di Sheppard

Valori della funzione $F(z)$. Aree sotto la curva normale standardizzata da 0 a z .

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817

[illegible]

In questa tavola, le righe sono in corrispondenza della parte decimale del valore z e le colonne corrispondono ai centesimi. Per esempio, per trovare il valore di $F(1,35)$ occorre individuare la riga in cui compare il numero 1,3 e scorrerla fino alla colonna corrispondente al numero 0,05; la casella così individuata contiene il valore cercato.

Applichiamo la tavola per risolvere il nostro problema:

$$P(0 < Z < 1,5) = F(1,5) = 0,4332 \quad \text{e} \quad P(0 < Z < 0,5) = F(0,5) = 0,1915.$$

Quindi:

$$\begin{aligned} P(0,5 < Z < 1,5) &= P(0 < Z < 1,5) - P(0 < Z < 0,5) = \\ &= 0,4332 - 0,1915 = 0,2417. \end{aligned}$$

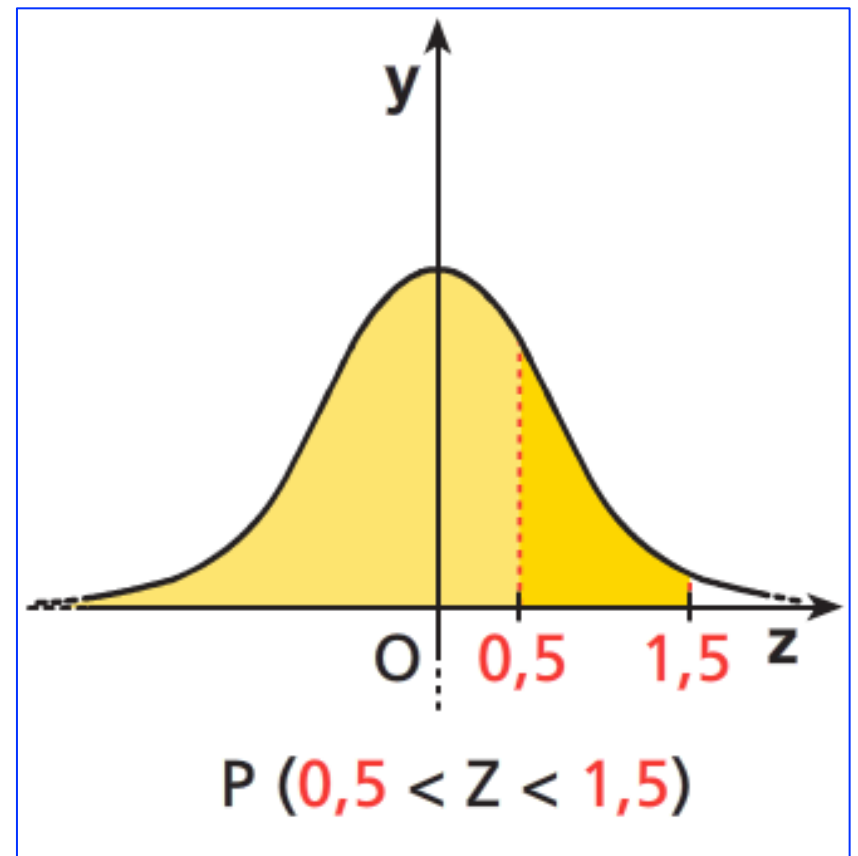
La simmetria della curva gaussiana rispetto all'asse y comporta che *la stessa tavola possa essere utilizzata anche per valori negativi della variabile Z* :

$$P(-z < Z < 0) = P(0 < Z < z),$$

$$\begin{aligned} P(-\infty < Z < -z) &= P(z < Z < +\infty) = \\ &= P(0 < Z < +\infty) - P(0 < Z < z) \\ &= 0,5 - P(0 < Z < z). \end{aligned}$$

$$P(0 < Z < 1,5) = F(1,5) = 0,4332$$

$$P(0 < Z < 0,5) = F(0,5) = 0,1915$$



$$\begin{aligned} P(0,5 < Z < 1,5) &= P(0 < Z < 1,5) - P(0 < Z < 0,5) = \\ &= 0,4332 - 0,1915 = 0,2417. \end{aligned}$$

ESEMPIO

I punteggi assegnati alle prove di un concorso hanno avuto una distribuzione normale e il punteggio medio è stato $\mu = 48$ con una deviazione standard $\sigma = 12$.

Calcoliamo la probabilità:

- a) che il punteggio sia compreso tra 39 e 43;
- b) che il punteggio sia compreso tra 35 e 65;
- c) che il punteggio sia inferiore a 30.

Per la standardizzazione abbiamo la relazione $Z = \frac{X - 48}{12}$.

$$\begin{aligned} \text{a)} \quad x_1 = 39 &\rightarrow z_1 = -0,75; \\ x_2 = 43 &\rightarrow z_2 = -0,42. \end{aligned}$$

$$\begin{aligned} P(-0,75 < Z < -0,42) &= P(0,42 < Z < 0,75) = \\ &= P(0 < Z < 0,75) - P(0 < Z < 0,42) = 0,2734 - 0,1628 = 0,1106. \end{aligned}$$

b) $x_1 = 35 \rightarrow z_1 = -1,08;$
 $x_2 = 65 \rightarrow z_2 = 1,42.$

$$\begin{aligned} P(-1,08 < Z < 1,42) &= P(-1,08 < Z < 0) + P(0 < Z < 1,42) = \\ &= P(0 < Z < 1,08) + P(0 < Z < 1,42) = 0,3599 + 0,4222 = 0,7821. \end{aligned}$$

c) $x = 30 \rightarrow z = -1,5.$

$$\begin{aligned} P(-\infty < Z < -1,5) &= P(1,5 < Z < +\infty) = \\ &= 0,5 - P(0 < Z < 1,5) = 0,5 - 0,4332 = 0,0668. \end{aligned}$$

Vi sono delle aree di probabilità che rivestono importanza in quanto sono frequentemente usate. Sono quelle che in una distribuzione normale corrispondono ai seguenti intervalli:

$$\mu - \sigma < X < \mu + \sigma$$

$$\mu - 2\sigma < X < \mu + 2\sigma$$

$$\mu - 3\sigma < X < \mu + 3\sigma$$

...

$$\mu - n\sigma < X < \mu + n\sigma.$$

Le relative probabilità si determinano ancora considerando i corrispondenti intervalli della variabile standardizzata Z e poi applicando la tavola dei valori di $F(z)$:

$$P(-1 < Z < 1) = P(-1 < Z < 0) + P(0 < Z < 1) = 2 \cdot F(1) = 0,6826$$

$$P(-2 < Z < 2) = P(-2 < Z < 0) + P(0 < Z < 2) = 2 \cdot F(2) = 0,9544$$

$$P(-3 < Z < 3) = P(-3 < Z < 0) + P(0 < Z < 3) = 2 \cdot F(3) = 0,9974$$

...

$$P(-n < Z < n) = P(-n < Z < 0) + P(0 < Z < n) = 2 \cdot F(n).$$

ESEMPIO

Le altezze di un gruppo di 400 persone hanno una distribuzione normale. La media è risultata di 172 cm con una varianza di 25 cm^2 . Determiniamo il numero di persone che hanno un'altezza compresa tra 162 cm e 182 cm.

Essendo $\mu = 172 \text{ cm}$ e $\sigma = 5 \text{ cm}$, gli estremi dell'intervallo considerato non sono altro che $\mu - 2\sigma$ e $\mu + 2\sigma$; esso ha probabilità $2 \cdot F(2)$, cioè corrisponde al 95,44% dei casi. Pertanto le persone che hanno un'altezza compresa tra 162 cm e 182 cm sono:

$$400 \cdot 0,9544 \simeq 382.$$

Possiamo anche procedere in modo inverso. Fissiamo una probabilità e, utilizzando la tavola $F(z)$, determiniamo l'intervallo per il quale si verifica detto valore.

ESEMPIO

Il peso di scatole di detersivo confezionate automaticamente si distribuisce normalmente. Il peso medio è 2,5 kg con una deviazione standard di 0,12 kg. Determiniamo l'intervallo di peso entro il quale si concentra il 99% delle scatole confezionate.

Cerchiamo nella tavola $F(z)$, che fornisce solo i valori per $z > 0$, il valore 0,495 (metà di 0,99). Questo numero corrisponde a un valore compreso tra 2,57 e 2,58. Quindi la probabilità dello 0,99 corrisponde (arrotondando) all'intervallo:

$$-2,57 < Z < 2,57 \quad \text{cioè} \quad \mu - 2,57\sigma < X < \mu + 2,57\sigma.$$

Essendo $\mu = 2,5$ e $\sigma = 0,12$, l'intervallo entro il quale troviamo il 99% della produzione è $2,19 < X < 2,81$.

Esercizi

ESERCIZIO GUIDA

Vengono lanciate contemporaneamente quattro monete non regolari. La probabilità di uscita della faccia testa è $\frac{2}{3}$ e quella di uscita della faccia croce è $\frac{1}{3}$.

- Determiniamo la distribuzione di probabilità e la funzione di ripartizione della variabile casuale X che esprime il numero di croci uscite.
- Calcoliamo la probabilità che il valore della variabile casuale X non superi il valore 2.
- Calcoliamo la probabilità che il valore della variabile casuale X superi il valore 1.

a) Il numero di volte che la faccia croce può comparire varia da 0 a 4. Le probabilità sono:

nessuna croce (e quattro teste)	$P(X = 0) = \binom{4}{0} \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^4 = \frac{16}{81};$
---------------------------------	--

una croce (e tre teste)	$P(X = 1) = \binom{4}{1} \left(\frac{1}{3}\right) \left(\frac{2}{3}\right)^3 = \frac{32}{81};$
-------------------------	--

due croci (e due teste)	$P(X = 2) = \binom{4}{2} \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^2 = \frac{8}{27};$
-------------------------	---

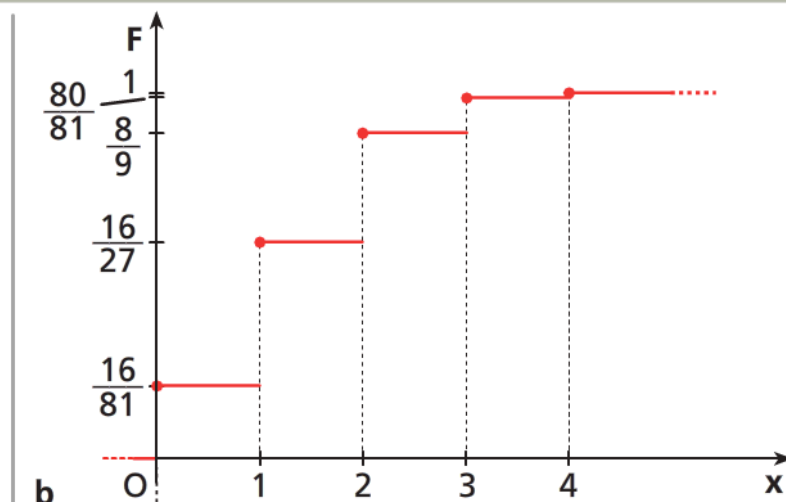
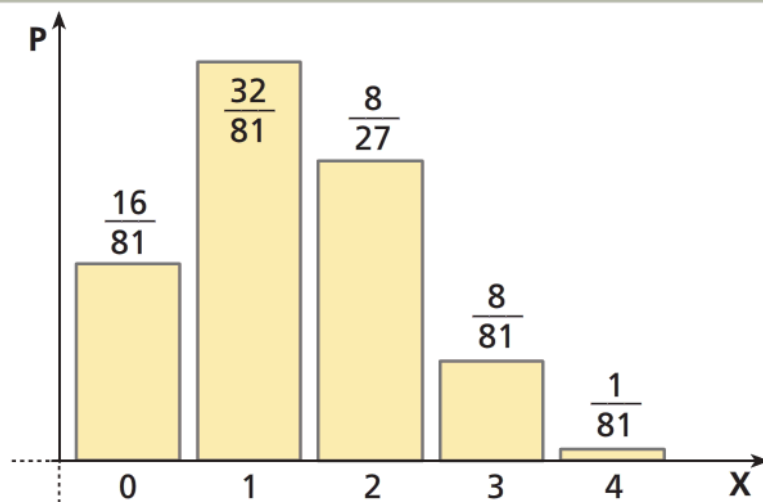
tre croci (e una testa)	$P(X = 3) = \binom{4}{3} \left(\frac{1}{3}\right)^3 \left(\frac{2}{3}\right) = \frac{8}{81};$
-------------------------	---

quattro croci (e nessuna testa)	$P(X = 4) = \binom{4}{4} \left(\frac{1}{3}\right)^4 \left(\frac{2}{3}\right)^0 = \frac{1}{81}.$
---------------------------------	---

La distribuzione di probabilità e la funzione di ripartizione della variabile casuale X sono:

X	0	1	2	3	4
$P(X)$	$\frac{16}{81}$	$\frac{32}{81}$	$\frac{8}{27}$	$\frac{8}{81}$	$\frac{1}{81}$
$F(X)$	$\frac{16}{81}$	$\frac{16}{27}$	$\frac{8}{9}$	$\frac{80}{81}$	1

Le rappresentazioni grafiche della distribuzione di probabilità e della funzione di ripartizione sono le seguenti.



b) $P(X \leq 2) = F(2) = \frac{8}{9}$.

c) $P(X > 1) = 1 - F(1) = 1 - \frac{16}{27} = \frac{11}{27}$.

ESERCIZIO GUIDA

Abbiamo la seguente variabile casuale $X =$ «numero delle giornate di assenza per malattia degli impiegati in una settimana» con la relativa distribuzione di probabilità rilevata statisticamente:

X	0	1	2	3	4	5
$P(X)$	0,49	0,18	0,12	0,1	0,09	0,02

- a) Determina le distribuzioni di probabilità delle variabili $2X$, $X - 3$ e $5X + 1$.
- b) Consideriamo la variabile casuale $Y =$ «numero delle giornate di assenza degli impiegati per motivi diversi dalla malattia» avente distribuzione di probabilità:

Y	0	1	2
$P(Y)$	0,9	0,08	0,02

Determina la distribuzione di probabilità della somma $X + Y$ nel caso in cui le variabili X e Y siano indipendenti.

a) La variabile casuale $2X$ ha i valori di X moltiplicati per 2 e le stesse probabilità:

$2X$	0	2	4	6	8	10
$P(2X)$	0,49	0,18	0,12	0,1	0,09	0,02

La variabile casuale $X - 3$ ha i valori di X sommati alla costante -3 e la stessa distribuzione delle probabilità:

$X - 3$	-3	-2	-1	0	1	2
$P(X - 3)$	0,49	0,18	0,12	0,1	0,09	0,02

La variabile casuale $5X + 1$ ha i valori di X moltiplicati per 5 e sommati a 1; le probabilità non cambiano:

$5X + 1$	1	6	11	16	21	26
$P(5X + 1)$	0,49	0,18	0,12	0,1	0,09	0,02

b) Costruiamo la tabella delle probabilità congiunte. Le probabilità marginali coincidono rispettivamente con le probabilità di X e Y .

Poiché le variabili sono indipendenti, ogni probabilità congiunta è il prodotto delle probabilità marginali:

$X \backslash Y$	0	1	2	$P(X)$
0	0,441	0,0392	0,0098	0,49
1	0,162	0,0144	0,0036	0,18
2	0,108	0,0096	0,0024	0,12
3	0,09	0,008	0,002	0,1
4	0,081	0,0072	0,0018	0,09
5	0,018	0,0016	0,0004	0,02
$P(Y)$	0,9	0,08	0,02	1

e la variabile casuale $X + Y$ ha la seguente distribuzione di probabilità:

$X + Y$	0	1	2	3	4	5	6	7
$P(X + Y)$	0,441	0,2012	0,1322	0,1032	0,0914	0,0272	0,0034	0,0004

ESERCIZIO GUIDA

Un'urna contiene 4 palline nere e 6 verdi. Il giocatore A estrae consecutivamente due palline rimettendo ogni volta la pallina estratta nell'urna. Se esse hanno colore uguale, vince 9 euro, altrimenti paga a B la sua posta. Calcoliamo la posta che A deve pagare, in caso di gioco equo.

La probabilità di vincita di A è $\frac{13}{25}$, quella di perdita è $\frac{12}{25}$ e conosciamo la posta di B che è 9 euro.

Impostiamo la seguente proporzione fra le poste dei due giocatori e le rispettive probabilità di vincita:

$$S_A : 9 = \frac{13}{25} : \frac{12}{25} \quad \rightarrow \quad S_A = \frac{9 \cdot \frac{13}{25}}{\frac{12}{25}} = 9,75.$$

Otteniamo lo stesso risultato impostando la seguente equazione ottenuta uguagliando a 0 la speranza matematica.

$$9 \cdot \frac{13}{25} - S_A \cdot \frac{12}{25} = 0 \quad \rightarrow \quad S_A = 9,75.$$

ESERCIZIO GUIDA

Un giocatore partecipa a un gioco organizzato e paga un prezzo di 4 euro. Il gioco consiste nell'estrarre consecutivamente senza reimmissione due carte da un mazzo di 40: se sono due figure dello stesso seme, il giocatore vince 45 euro, mentre se hanno lo stesso valore vince 17 euro.

- a) Verifichiamo che il gioco non è equo e determiniamo la posta corretta affinché lo sia.
- b) Lasciando immutata la posta modifichiamo la prima vincita affinché il gioco sia equo.

a) I valori delle due probabilità sono: $\frac{2}{130}$ e $\frac{10}{130}$. Calcoliamo il valore della posta in caso di gioco equo applicando la relazione della speranza matematica delle vincite lorde:

$$P = 45 \cdot \frac{2}{130} + 17 \cdot \frac{10}{130} = 2.$$

Essendo il prezzo inferiore alla posta richiesta, il gioco è sfavorevole.

b) Per rispondere al secondo quesito impostiamo la seguente equazione, sempre relativa alla speranza matematica della vincita lorda:

$$x \cdot \frac{2}{130} + 17 \cdot \frac{10}{130} = 4, \quad \text{da cui} \quad x = 175.$$

ESERCIZIO GUIDA

Un'urna contiene 5 palline che portano i numeri 1, 2, 3, 6 e 7. Estraiamo consecutivamente due palline senza rimettere nell'urna la pallina estratta per prima, e consideriamo come variabile casuale X il valore minore dei due numeri estratti. Calcoliamo il valore medio, la varianza e la deviazione standard di X .

La distribuzione di probabilità della variabile casuale X è la seguente:

X	1	2	3	6
P	$\frac{2}{5}$	$\frac{3}{10}$	$\frac{1}{5}$	$\frac{1}{10}$

Il valore medio di X è:

$$M(X) = 1 \cdot \frac{2}{5} + 2 \cdot \frac{3}{10} + 3 \cdot \frac{1}{5} + 6 \cdot \frac{1}{10} = \frac{11}{5} = 2,2.$$

Per determinare la varianza, consideriamo la variabile casuale degli scarti dalla media:

$[X - M(X)]$	-1,2	-0,2	0,8	3,8
P	$\frac{2}{5}$	$\frac{3}{10}$	$\frac{1}{5}$	$\frac{1}{10}$

Il valore medio di $[X - M(X)]$ è nullo:

$$M[X - M(X)] = -1,2 \cdot \frac{2}{5} - 0,2 \cdot \frac{3}{10} + 0,8 \cdot \frac{1}{5} + 3,8 \cdot \frac{1}{10} = 0.$$

Consideriamo poi la variabile casuale dei quadrati degli scarti:

$[X - M(X)]^2$	1,44	0,04	0,64	14,44
P	$\frac{2}{5}$	$\frac{3}{10}$	$\frac{1}{5}$	$\frac{1}{10}$

Il valore medio dei quadrati degli scarti è la varianza di X :

$$var(X) = 1,44 \cdot \frac{2}{5} + 0,04 \cdot \frac{3}{10} + 0,64 \cdot \frac{1}{5} + 14,44 \cdot \frac{1}{10} = 2,16;$$

dalla varianza si ottiene la deviazione standard:

$$\sigma(X) = \sqrt{var(X)} = \sqrt{2,16} \simeq 1,47.$$

La variabile casuale considerata è caratterizzata dal valore di sintesi $\mu = 2,2$ e dal valore di dispersione $\sigma^2 = 2,16$.

I calcoli per determinare i due valori caratteristici della variabile casuale possono essere esposti nella tabella seguente:

x_i	p_i	$x_i p_i$	$x_i - \mu$	$(x_i - \mu)^2$	$(x_i - \mu)^2 p_i$
1	0,4	0,40	-1,20	1,44	0,576
2	0,3	0,60	-0,20	0,04	0,012
3	0,2	0,60	0,80	0,64	0,128
6	0,1	0,60	3,80	14,44	1,444
Σ		2,20			2,16

ESERCIZIO GUIDA

Consideriamo due variabili casuali X e Y aventi rispettivamente distribuzione di probabilità:

X	0	1	2	3	4	5
P	0,45	0,22	0,13	0,1	0,07	0,03

Y	1	2	3
P	0,6	0,21	0,19

Calcoliamo il valore medio e la varianza delle seguenti variabili casuali:

a) $3X$, $Y - 4$ e $3Y + 5$;

b) $X + Y$, sapendo che le due variabili sono indipendenti.

La variabile casuale X ha valore medio $\mu_X = 1,21$ e varianza $\sigma^2_X \simeq 2,05$.
La variabile casuale Y ha valore medio $\mu_Y = 1,59$ e varianza $\sigma^2_Y \simeq 0,62$.

a) $3X \rightarrow$ valore medio: $\mu = 3 \cdot \mu_X = 3 \cdot 1,21 = 3,63$;
varianza: $\sigma^2 = 3^2 \cdot \sigma_X^2 \simeq 9 \cdot 2,05 = 18,45$.

$Y - 4 \rightarrow$ valore medio: $\mu = \mu_Y - 4 = 1,59 - 4 = -2,41$;
varianza: $\sigma^2 = \sigma_Y^2 \simeq 0,62$.

$3Y + 5 \rightarrow$ valore medio: $\mu = 3 \cdot \mu_Y + 5 = 3 \cdot 1,59 + 5 = 9,77$;
varianza: $\sigma^2 = 3^2 \cdot \sigma_Y^2 \simeq 9 \cdot 0,62 = 5,58$.

b) Poiché le variabili casuali sono indipendenti, abbiamo:

$$\begin{aligned}\mu &= \mu_X + \mu_Y = 1,21 + 1,59 = 2,8 && \text{(valore medio);} \\ \sigma^2 &= \sigma_X^2 + \sigma_Y^2 \simeq 2,05 + 0,62 = 2,67 && \text{(varianza).}\end{aligned}$$

ESERCIZIO GUIDA

Date due variabili casuali dipendenti X e Y aventi le probabilità congiunte indicate a fianco, calcoliamo sia il valore medio e la varianza della somma $X + Y$, sia la covarianza di X e Y .

$X \backslash Y$	3	4	7	9
3	0	$\frac{1}{33}$	$\frac{1}{66}$	$\frac{5}{132}$
4	$\frac{1}{33}$	$\frac{1}{11}$	$\frac{2}{33}$	$\frac{5}{33}$
7	$\frac{1}{66}$	$\frac{2}{33}$	$\frac{1}{66}$	$\frac{5}{66}$
9	$\frac{5}{132}$	$\frac{5}{33}$	$\frac{5}{66}$	$\frac{5}{33}$

Le probabilità relative ai valori delle variabili casuali X e Y sono date dalle probabilità marginali. Osserviamo che le due variabili hanno la stessa distribuzione di probabilità.

Il valore medio e la varianza delle due variabili sono: $\mu_X = \mu_Y = 6,5$ e $\sigma^2_X = \sigma^2_Y = 5,75$.

In questo caso il valore medio di $X + Y$ è ancora uguale alla somma dei valori medi di X e Y :

$$\mu = \mu_X + \mu_Y = 6,5 + 6,5 = 13.$$

Per determinare la varianza, invece, dobbiamo applicare la definizione e quindi determinare la distribuzione di probabilità di $X + Y$.

La distribuzione di probabilità che si ottiene è:

$X + Y$	6	7	8	10	11	12	13	14	16	18
P	0	$\frac{2}{33}$	$\frac{1}{11}$	$\frac{1}{33}$	$\frac{4}{33}$	$\frac{5}{66}$	$\frac{10}{33}$	$\frac{1}{66}$	$\frac{5}{33}$	$\frac{5}{33}$

Applicando la definizione, otteniamo che la varianza di $X + Y$ è $\sigma^2 = 10,4\overline{5}$.

La covarianza $cov(X, Y)$ è la semidifferenza tra la varianza di $X + Y$ e la somma delle varianze delle due variabili casuali X e Y :

$$cov(X, Y) = \frac{10,4\overline{5} - (5,75 + 5,75)}{2} = -0,52\overline{27}.$$

Possiamo ottenere questo valore anche direttamente, calcolando:

- gli scarti dei valori della variabile X ;
- gli scarti dei valori della variabile Y ;
- la somma dei prodotti degli scarti per le rispettive probabilità congiunte.

Nella tabella seguente sono esposti i calcoli effettuati:

$\begin{matrix} Y - \mu_Y \\ X - \mu_X \end{matrix}$	$-3,5$	$-2,5$	$0,5$	$2,5$	Σ
$-3,5$	0	0,2652	$-0,0265$	$-0,3314$	$-0,0927$
$-2,5$	0,2652	0,5682	$-0,0758$	$-0,9470$	$-0,1894$
$0,5$	$-0,0265$	$-0,0758$	0,0038	0,0947	$-0,0038$
$2,5$	$-0,3314$	$-0,9470$	0,0947	0,9470	$-0,2367$
Σ	$-0,0927$	$-0,1894$	$-0,0038$	$-0,2367$	$-0,5226$

ESERCIZIO GUIDA

Un'urna contiene 30 palline numerate da 1 a 30. Consideriamo le seguenti variabili casuali:

- a) resto della divisione tra il numero presente sulla pallina estratta e il numero 6;
- b) numero di volte in cui può uscire, effettuando 4 estrazioni nelle stesse condizioni, una pallina con un numero divisibile per 3;
- c) numero delle volte in cui può uscire, effettuando 4 estrazioni nelle stesse condizioni, la pallina con il numero 1.

Determiniamo per ogni variabile casuale la distribuzione di probabilità, la funzione di ripartizione, il valore medio, la varianza e la deviazione standard.

Per la variabile casuale b) determiniamo la probabilità di ottenere almeno 2 successi.

- a) I possibili resti della divisione con divisore 6 sono 0, 1, 2, 3, 4, 5 e ognuno di questi valori ha probabilità costante di verificarsi pari a $p = \frac{5}{30} = \frac{1}{6}$. Infatti i numeri sulle palline che divisi per 6 danno resto 1 sono: 1, 7, 13, 19, 25; i numeri che divisi per 6 danno resto 2 sono: 2, 8, 14, 20, 26; e così via.

Abbiamo una **distribuzione uniforme**

X	0	1	2	3	4	5
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
F	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{5}{6}$	1

con i seguenti valori di sintesi per la variabile casuale X :

$$\mu = \frac{0 + 1 + 2 + 3 + 4 + 5}{6} = \frac{15}{6} = 2,5;$$

$$\sigma^2 = \frac{(0 - 2,5)^2 + (1 - 2,5)^2 + (2 - 2,5)^2 + (3 - 2,5)^2 + (4 - 2,5)^2 + (5 - 2,5)^2}{6} = \frac{17,5}{6} = 2,91\bar{6};$$

$$\sigma = \sqrt{2,91\bar{6}} \simeq 1,708.$$

b) È il caso della **distribuzione binomiale** (o di **Bernoulli**).

La probabilità di estrarre un numero divisibile per 3 è $p = \frac{10}{30} = \frac{1}{3}$ e pertanto $q = \frac{2}{3}$. La legge che determina la distribuzione di probabilità è $P(X = x) = \binom{4}{x} \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{4-x}$. Sostituendo a x i valori 0, 1, 2, 3, 4, numero delle possibili volte che in quattro estrazioni può uscire un numero divisibile per 3, otteniamo:

X	0	1	2	3	4
P	$\frac{16}{81}$	$\frac{32}{81}$	$\frac{8}{27}$	$\frac{8}{81}$	$\frac{1}{81}$
F	$\frac{16}{81}$	$\frac{16}{27}$	$\frac{8}{9}$	$\frac{80}{81}$	1

$$\text{Abbiamo: } \mu = 4 \cdot \frac{1}{3} = \frac{4}{3} = 1,3, \quad \sigma^2 = 4 \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{8}{9} = 0,8, \quad \sigma = \sqrt{\frac{8}{9}} \simeq 0,943.$$

$$\text{Inoltre: } P(X \geq 2) = 1 - F(1) = 1 - \frac{16}{27} = \frac{11}{27}.$$

c) Anche in questo caso la variabile casuale assume i valori 0, 1, 2, 3, 4, ma essendo piccolo il valore della probabilità dell'evento $p = \frac{1}{30}$, è opportuno utilizzare la **distribuzione di Poisson**.

Occorre porre $\lambda = 4 \cdot \frac{1}{30} = \frac{2}{15}$; la distribuzione di probabilità è quindi $P(X = x) = \frac{\left(\frac{2}{15}\right)^x}{x!} \cdot e^{-\frac{2}{15}}$.

Otteniamo

<i>X</i>	0	1	2	3	4
<i>P</i>	0,875173	0,116690	0,007779	0,000346	0,000012
<i>F</i>	0,875173	0,991863	0,999642	0,999988	1

$$\text{e: } \mu = \sigma^2 = \lambda = \frac{2}{15}, \quad \sigma = \sqrt{\frac{2}{15}} \simeq 0,365.$$

ESERCIZIO GUIDA

Un concorrente a una prova ha conseguito i seguenti punteggi in tre test diversi: 48, 35, 74. Per questi test conosciamo i punteggi medi 39, 27, 65 e le deviazioni standard 6, 5, 14.

Determiniamo in quale test il concorrente ha conseguito il risultato migliore.

Chiamando X la variabile casuale dei punteggi ottenuti, la variabile standardizzata Z è:

$$Z = \frac{X - \mu}{\sigma}.$$

Trasformiamo i punteggi ottenuti x nei corrispondenti punti zeta:

$$x = 48 \rightarrow z = \frac{48 - 39}{6} = 1,5; \qquad x = 35 \rightarrow z = \frac{35 - 27}{5} = 1,6;$$

$$x = 74 \rightarrow z = \frac{74 - 65}{14} \simeq 0,64.$$

Possiamo concludere che il punteggio migliore si è ottenuto con il secondo test.

ESERCIZIO GUIDA

Dopo aver verificato che la seguente funzione è una funzione densità di probabilità, consideriamo in $[0; 4]$ una variabile casuale continua X avente tale funzione densità.

$$f(x) = \begin{cases} 0 & \text{se } x < 0 \\ \frac{1}{8}x & \text{se } 0 \leq x \leq 4 \\ 0 & \text{se } x > 4 \end{cases}$$

- a) Determiniamo la funzione di ripartizione $F(x)$ di X . b) Rappresentiamo graficamente $f(x)$ e $F(x)$.
c) Calcoliamo $P(1 \leq X \leq 3)$. d) Calcoliamo il valore medio e la varianza di X .

Verifichiamo che la funzione $f(x)$ è una funzione densità di probabilità:

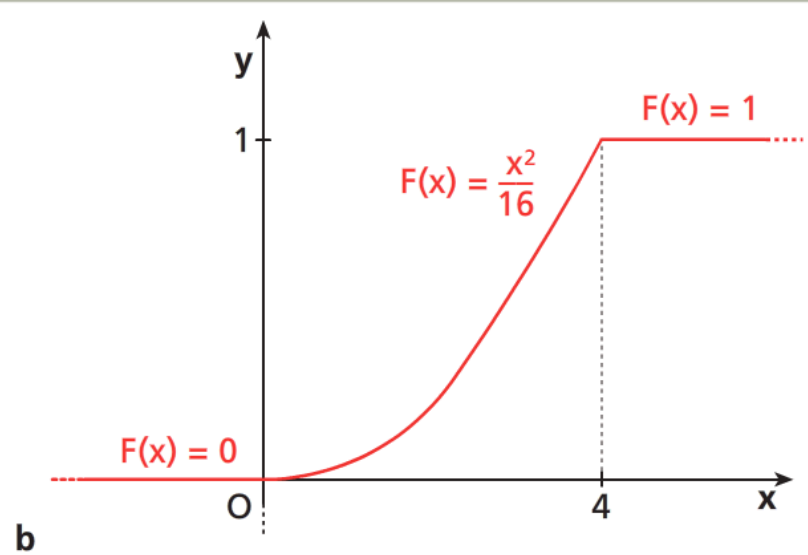
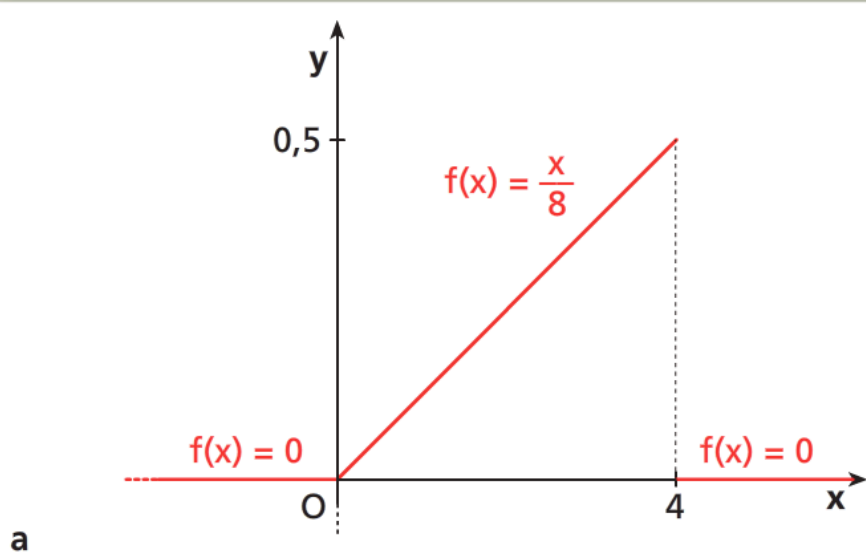
$$f(x) \geq 0 \text{ in } \mathbb{R} \quad \text{e} \quad \int_{-\infty}^{+\infty} f(x) dx = \int_0^4 \frac{1}{8}x dx = \left[\frac{1}{8} \cdot \frac{x^2}{2} \right]_0^4 = 1.$$

- a) Determiniamo la funzione di ripartizione $F(x)$:

$$F(x) = \int_{-\infty}^x f(t) dt = \int_0^x f(t) dt = \int_0^x \frac{1}{8}t dt = \left[\frac{1}{8} \cdot \frac{t^2}{2} \right]_0^x = \frac{x^2}{16}, \quad \text{per } 0 \leq x \leq 4;$$

$$F(x) = 0, \quad \text{per } x < 0; \qquad F(x) = 1, \quad \text{per } x > 4.$$

b) Rappresentiamo graficamente le funzioni $f(x)$ e $F(x)$:



c) Calcoliamo $P(1 \leq X \leq 3)$ in due modi, cioè mediante la funzione densità di probabilità o quella di ripartizione:

$$P(1 \leq X \leq 3) = \int_1^3 \frac{1}{8}x \, dx = \left[\frac{x^2}{16} \right]_1^3 = \frac{9}{16} - \frac{1}{16} = \frac{1}{2};$$

$$P(1 \leq X \leq 3) = F(3) - F(1) = \frac{9}{16} - \frac{1}{16} = \frac{1}{2}.$$

d) Calcoliamo il valore medio di X :

$$M(X) = \int_{-\infty}^{+\infty} x \cdot f(x) \, dx = \int_0^4 x \cdot f(x) \, dx = \int_0^4 x \cdot \frac{1}{8}x \, dx = \left[\frac{1}{8} \cdot \frac{x^3}{3} \right]_0^4 = \frac{8}{3}.$$

Calcoliamo la varianza:

$$\text{var}(X) = M(X^2) - [M(X)]^2 = \int_0^4 x^2 \cdot \frac{1}{8}x \, dx - \left(\frac{8}{3} \right)^2 = \left[\frac{1}{8} \cdot \frac{x^4}{4} \right]_0^4 - \frac{64}{9} = 8 - \frac{64}{9} = \frac{8}{9}.$$

ESERCIZIO GUIDA

Consideriamo una variabile casuale continua X che si distribuisce uniformemente nell'intervallo $[2; 7]$.

- a) Calcoliamo la funzione di ripartizione, il valore medio, la varianza e la deviazione standard di X .
- b) Calcoliamo la probabilità che X assuma valori compresi tra 2,5 e 3.

La funzione densità di probabilità di X , nell'intervallo $[2; 7]$, è $f(x) = \frac{1}{7-2} = \frac{1}{5}$.

- a) In $[2; 7]$ la funzione di ripartizione è:

$$F(x) = \int_2^x \frac{1}{5} dt = \left[\frac{t}{5} \right]_2^x = \frac{x-2}{5}.$$

I valori di sintesi di X valgono:

$$\mu = \frac{7+2}{2} = \frac{9}{2}; \quad \sigma^2 = \frac{(7-2)^2}{12} = \frac{25}{12}; \quad \sigma = \frac{5}{2\sqrt{3}}.$$

- b) $P(2,5 \leq X \leq 3) = \int_{2,5}^3 \frac{1}{5} dx = \left[\frac{x}{5} \right]_{2,5}^3 = \frac{3}{5} - \frac{1}{2} = 0,1$, oppure, con la funzione di ripartizione,

$$P(2,5 \leq X \leq 3) = F(3) - F(2,5) = \frac{1}{5} - \frac{0,5}{5} = \frac{0,5}{5} = 0,1.$$

ESERCIZIO GUIDA

Un'impresa costruisce rasoi elettrici la cui durata è una grandezza che si distribuisce normalmente. La durata media è di 6 anni con una deviazione standard di 1,5 anni.

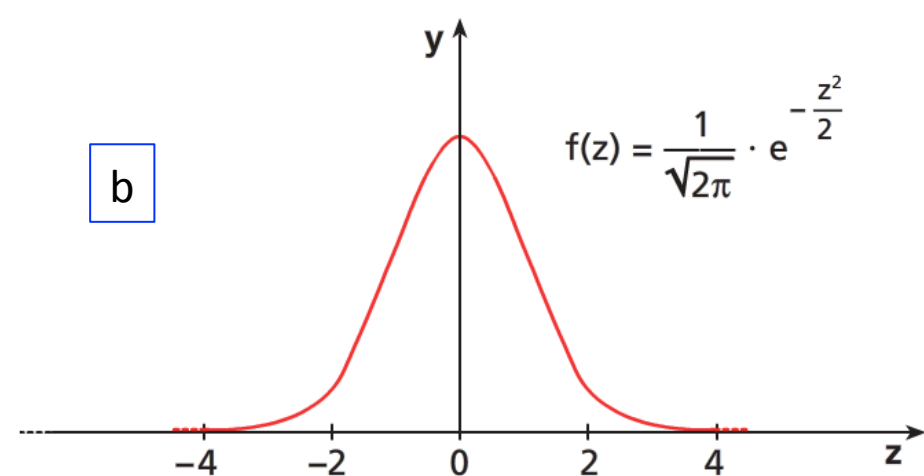
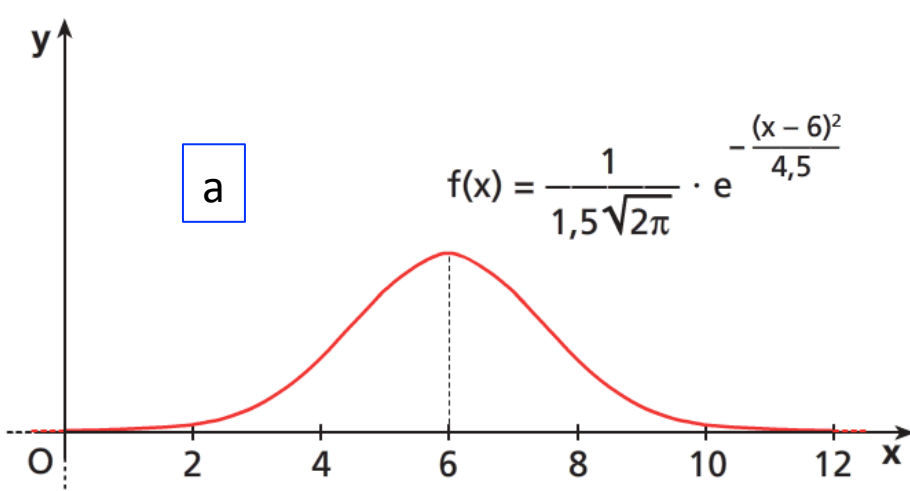
Calcoliamo:

- a) la probabilità che un rasoio abbia una durata fino a 8 anni;
- b) la probabilità che abbia una durata di almeno 7 anni;
- c) il numero di rasoi che, su una produzione di 400, abbia una durata compresa tra 4 e 9 anni.

Consideriamo la variabile casuale normale $N(6; 2,25)$, la cui funzione densità di probabilità ha il grafico riportato in figura *a*.

Utilizziamo la variabile standardizzata $Z = N(0; 1)$ che si ottiene con la trasformazione $Z = \frac{X - 6}{1,5}$. Il grafico è quello in figura *b*.

In entrambi i grafici l'area compresa tra la curva e l'asse delle ascisse è 1, ma per il secondo grafico abbiamo a disposizione la tavola $F(z)$ che fornisce il valore dell'area sotto la curva per valori crescenti di $z > 0$.



a) Probabilità che un rasoio abbia una durata fino a 8 anni.

Il punto zeta corrispondente a $x = 8$ è $z = 1,3\bar{3}$. Sfruttiamo la simmetria di $f(z)$ e utilizziamo la tavola $F(z)$:

$$\begin{aligned} P(X \leq 8) &\simeq P(-\infty < Z \leq 1,33) = P(-\infty < Z < 0) + P(0 \leq Z \leq 1,33) = \\ &= 0,5 + 0,4082 = 0,9082. \end{aligned}$$

b) Probabilità che abbia una durata di almeno 7 anni.

Il punto zeta corrispondente a $x = 7$ è $z = 0,6\bar{7}$:

$$P(X \geq 7) \simeq P(Z \geq 0,67) = 0,5 - P(0 < Z < 0,67) = 0,5 - 0,2486 = 0,2514.$$

c) Calcoliamo la probabilità che un rasoio abbia una durata compresa tra 4 e 9 anni.

Trasformando $x_1 = 4$ e $x_2 = 9$ otteniamo $z_1 = -1,3\bar{3}$ e $z_2 = 2$:

$$\begin{aligned} P(4 \leq X \leq 9) &\simeq P(-1,33 \leq Z \leq 2) = P(0 \leq Z \leq 1,33) + P(0 \leq Z \leq 2) = \\ &= 0,4082 + 0,4772 = 0,8854. \end{aligned}$$

Possiamo valutare il numero di rasoi che, su una produzione di 400, avranno una durata compresa tra 4 e 9 anni. Otteniamo:

$$400 \cdot 0,8854 \simeq 354.$$