CAPITOLO 5
DISTRIBUZIONI DI PROBABILTA'
Introduzione

Si ha una *distribuzione di probabilità* quando al valore assunto dalla variabile aleatoria, viene associata la relativa probabilità. Indicata al solito con *X* la variabile aleatoria, si definisce **funzione di ripartizione** $F(x)$ la probabilità che la variabile aleatoria assuma un valore non superiore a *x*:

$$F(x) := p(X \le x).$$

In generale,

$$F(b) - F(a) := p(a \le X \le b).$$

Ricordiamo le definizioni di *media* e di *varianza* di una variabile aleatoria *discreta*:

$$M(X) = \sum x_k p_k$$
$$\sigma^2(X) = \sum (x_k - M(X))^2 p_k .$$

Decisamente interessante è il caso rappresentato dalle variabili aleatorie *continue*; in tal caso, se la funzione di ripartizione è derivabile, la sua derivata si dice **funzione di densità**

$$F'(x) = f(x),$$

ed il termine $f(x)dx$ rappresenta approssimativamente la probabilità che la variabile aleatoria assuma un valore compreso in un intervallo infinitesimo di ampiezza dx contenente il valore $X = x$.
Riepiloghiamo quanto detto fino ad ora nel seguente schema comparativo tra le variabili aleatorie discrete e continue.

v.a. discrete	v.a. continue
$M(X) = \sum x_k p_k$	$M(X) = \int_{-\infty}^{+\infty} x f(x) dx$
$\sigma^2(X) = \sum (x_k - M(X))^2 p_k$	$\sigma^2(X) = \int_{-\infty}^{+\infty} (x - M(X))^2 f(x) dx$
$\sum p_k = 1$	$\int_{-\infty}^{+\infty} f(x) dx = 1$

5.1 Distribuzione *binomiale o di Bernoulli*

Consideriamo esperimenti, riguardanti variabili aleatorie discrete, che possono dare due esiti, come ad esempio il lancio di una moneta, o l'estrazione di una pallina da un'urna contenente palline di due soli colori. Se p è la probabilità che un evento si verifichi, allora $q = 1 - p$ è la probabilità che l'evento non si verifichi.

Esempio. Un'urna contiene 25 palline, di cui 10 bianche e 15 nere. Calcolare la probabilità che, su 5 estrazioni con reinserimento nell'urna della pallina estratta:

1. Le prime tre palline siano bianche;
   - Gli esiti favorevoli all'evento sono: BBBBB-BBBBN-BBBNB-BBBNN. La probabilità si ottiene sommando la probabilità delle cinquine individuate, in virtù dell'incompatibilità degli eventi: $\left(\frac{2}{5}\right)^5 + 2\left(\frac{2}{5}\right)^4\left(\frac{3}{5}\right) + \left(\frac{2}{5}\right)^3\left(\frac{3}{5}\right)^2 = 0,064 = 6,4\%$.
2. Soltanto le prime tre siano bianche
   - Gli esiti favorevoli all'evento sono: BBBNN . La probabilità è $\left(\frac{2}{5}\right)^3\left(\frac{3}{5}\right)^2 = 0,02304 = 2,304\%$

3. Escano tre palline bianche;
- Il numero di esiti favorevoli all'evento è dato dalle combinazioni di 5 elementi presi a 3 per volta. La probabilità è quindi $\binom{5}{3}\left(\frac{2}{5}\right)^3\left(\frac{3}{5}\right)^2 = 0,2304 = 23,04\%$.

Soffermiamoci sull'ultimo caso esaminato nell'esempio, e cerchiamo di generalizzarlo. Possiamo affermare che la probabilità di conseguire h successi (ognuno con probabilità p) su n prove è data dalla relazione:

$$P_h = \binom{n}{h} p^h q^{n-h} \qquad q = 1 - p \quad .$$

Si parla di *distribuzione* facendo riferimento allo studio di *tutti* i possibili esiti, ovvero dei possibili valori che può assumere la variabile aleatoria discreta riferita all'evento considerato.
Esempio. Si lancia 4 volte una moneta regolare, sempre nelle stesse condizioni. Studiamo la variabile aleatoria "numero di volte che esce testa".
Riassumiamo nella seguente tabella le 4 possibilità:

X	**0**	**1**	**2**	**3**	**4**
$p(X)$	$\binom{4}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^4 = \frac{1}{16}$	$\binom{4}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^3 = \frac{4}{16}$	$\binom{4}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 = \frac{6}{16}$	$\binom{4}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^1 = \frac{4}{16}$	$\binom{4}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^0 = \frac{1}{16}$
$\sum p(X)$	$= \frac{1}{16}$	$= \frac{5}{16}$	$= \frac{11}{16}$	$= \frac{15}{16}$	$= \frac{16}{16} = 1$

La distribuzione che abbiamo appena visto prende il nome del matematico J. Bernoulli (1654-1705), considerato tra i fondatori del moderno calcolo delle probabilità, autore dell'opera, pubblicata postuma nel 1713, *Ars conjectandi*.
Osserviamo i seguenti fatti, relativi alla distribuzione di Bernoulli, facendo riferimento all'esempio trattato.

1. *Simmetria della distribuzione*: la probabilità che esca h volte testa è uguale a quella che esca $n-h$ volte croce. Questa proprietà vale soltanto nel caso particolare $p = q = \frac{1}{2}$.
2. *Condizione di normalizzazione*: la somma delle probabilità è uguale a 1. Questa probabilità vale sempre, per l'incompatibilità degli eventi.
3. In genere, la probabilità che la variabile aleatoria assuma un qualsiasi valore è abbastanza bassa. Questo fatto è conseguenza di un numero di prove abbastanza elevato.
4. Il valore della variabile aleatoria che presenta la massima probabilità è $\frac{n}{2}$ se n è pari, la parte intera di $\frac{n}{2}$ oppure $\frac{n+1}{2}$ se n è dispari.
5. *La media della distribuzione binomiale* è $M = np$.
6. *Lo scarto quadratico medio della distribuzione binomiale è* $\sigma = \sqrt{npq}$.

5.2 Distribuzione *Normale* o di Gauss

Quando il numero di prove è molto grande, la distribuzione binomiale può essere approssimata da una funzione, la nota curva a *campana* (o, più semplicemente, *campana*) di Gauss:

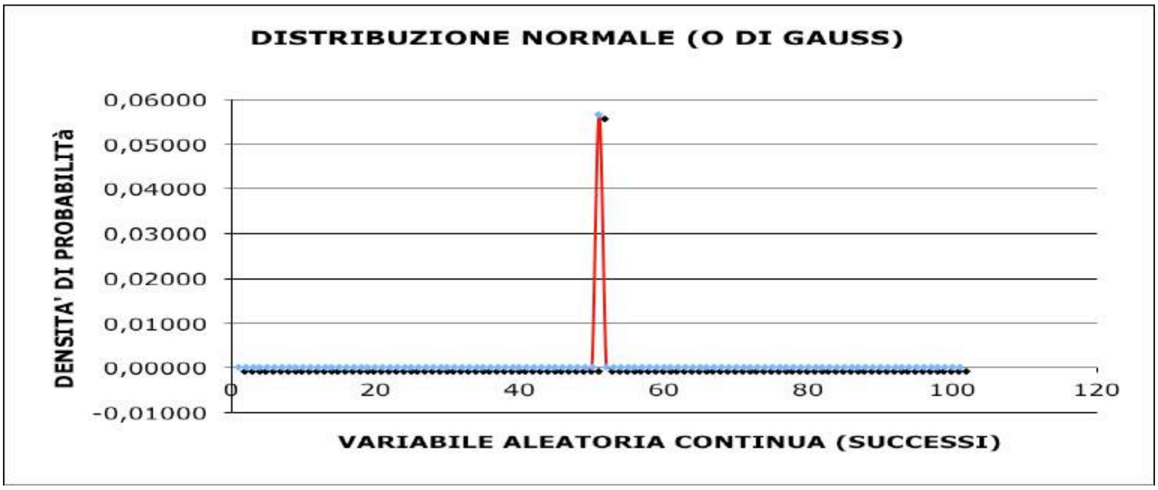
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-M)^2}{2\sigma^2}} ,$$

dove con $M = np$ e $\sigma = \sqrt{npq}$ si sono indicati rispettivamente la media e lo scarto quadratico medio della distribuzione.

Proprietà della funzione di Gauss

1. E' simmetrica rispetto alla media. Questo significa che la probabilità di un valore che supera la media di una quantità fissata, è uguale alla probabilità di un valore che è inferiore alla media della stessa quantità.
2. L'area al di sotto della curva è uguale a uno (la curva è *normalizzata*). Sfruttiamo il risultato $\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$. Posto $z = \frac{x-M}{\sigma\sqrt{2}} \Rightarrow dx = \sigma\sqrt{2}dz$. Di conseguenza,
$$\int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-M)^2}{2\sigma^2}} dx = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-z^2} \sigma\sqrt{2}dz = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-z^2} dz = \frac{\sqrt{\pi}}{\sqrt{\pi}} = 1 .$$

Esempio. Lanciamo 100 volte una moneta: la distribuzione binomiale è approssimata dalla distribuzione normale di media $M = np = 100 \cdot \frac{1}{2} = 50$, e scarto quadratico medio $\sigma = \sqrt{npq} = \sqrt{100 \cdot \frac{1}{2} \cdot \frac{1}{2}} = 5$; I grafici che seguono sono stati realizzati con il foglio elettronico excel.



Interpretazione dei risultati

L'equazione della curva normale è. La probabilità che esca testa un numero di volte compreso tra $45 = M - \sigma \le x \le M + \sigma = 55$ è pari al 68,3%, come si evince dalle tavole redatte appositamente per il calcolo di questa probabilità. Formalmente la probabilità richiesta è data dall'integrale $P(45 \leq x \leq 50) = \int_{45}^{55} \frac{1}{5\sqrt{2\pi}} e^{-\frac{(x-50)^2}{50}} dx$ che, come noto, approssima la distribuzione binomiale $P(45 \leq k \leq 50) = \sum_{k=45}^{55} \binom{100}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{100-k}$.

L'elaborazione dei dati e il grafico che seguono, sono stati realizzati con il foglio elettronico excel.

p= 0,5
q= 0,5
n= 100

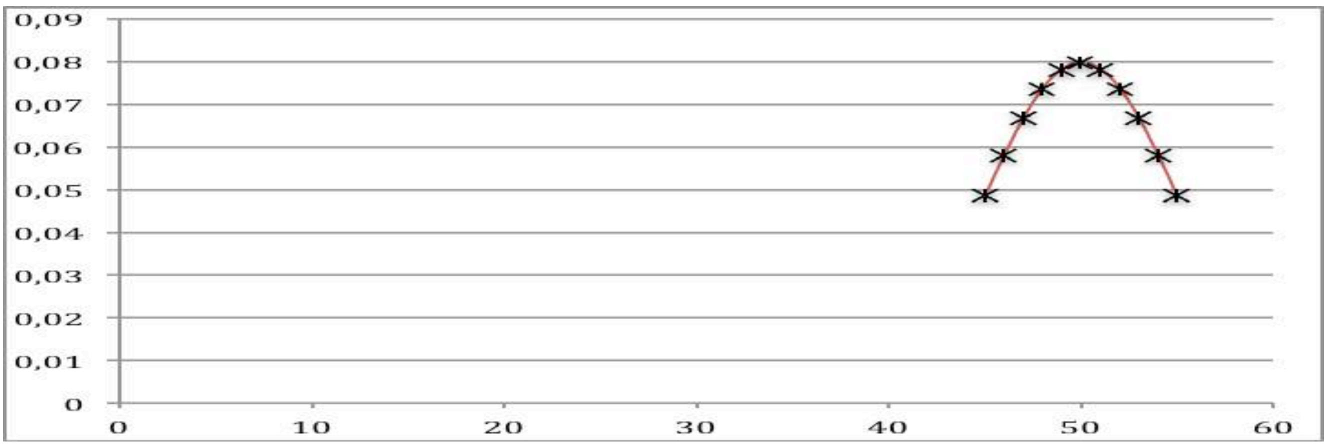
M= 50
$\sigma =$ 5

k min= 45

k max= 55

$\binom{n}{k}$ $\quad P(k) = \binom{n}{k} p^k q^{n-k}$ $\quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-M)^2}{2\sigma^2}}$

k				
45	6,14485E+28	45	0,048474297	0,048394145
46	7,3471E+28	46	0,057958398	0,057938311
47	8,44135E+28	47	0,0665905	0,066644921
48	9,32066E+28	48	0,07352701	0,073654028
49	9,89131E+28	49	0,078028664	0,078208539
50	1,00891E+29	50	0,079589237	0,079788456
51	9,89131E+28	51	0,078028664	0,078208539
52	9,32066E+28	52	0,07352701	0,073654028
53	8,44135E+28	53	0,0665905	0,066644921
54	7,3471E+28	54	0,057958398	0,057938311
55	6,14485E+28	55	0,048474297	0,048394145



Si noti come il grafico della funzione gaussiana, si adatta ai punti della distribuzione di Bernoulli.

Nelle applicazioni si ricorre spesso alla distribuzione normale *standardizzata*, ottenuta mediante la sostituzione formale $z = \frac{x-M}{\sigma}$, da cui segue

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad -\infty < z < +\infty \quad .$$

5.3 Distribuzione *uniforme*

Una variabile aleatoria continua $(a \le X \le b)$ ha una distribuzione uniforme se ha una funzione di ripartizione così definita:

$$F(x) = \begin{cases} 0 & se \quad x \le a \\ (x-a)/(b-a) & se \quad a < x \le b \\ 1 & se \quad x > b \end{cases} .$$

La funzione densità è

$$f(x) = \begin{cases} 0 & se \quad x \le a \\ 1/(b-a) & se \quad a < x \le b \\ 0 & se \quad x > b \end{cases} .$$

Osserviamo, innanzi tutto, che $\int_{-\infty}^{+\infty} f(x)dx = \int_a^b \frac{1}{(b-a)}dx = \frac{b-a}{b-a} = 1$. Inoltre si ha

$M(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_a^b \frac{x}{(b-a)}dx = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}$, e

$$\sigma^2(X) = \int_{-\infty}^{+\infty} (x - M(X))^2 f(x)dx = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \frac{1}{(b-a)}dx =$$

$$\frac{\left(\frac{b-a}{2}\right)^3 - \left(\frac{a-b}{2}\right)^3}{3(b-a)} = \frac{2\left(\frac{b-a}{2}\right)^3}{3(b-a)} = \frac{(b-a)^2}{12} .$$

Esercizio

a) Si dica per quale valore del parametro A, la funzione $g_n(x) = \begin{cases} 0; & x < 0 \\ Ax^n e^{-x}; & x \ge 0 \end{cases}$ *rappresenta una densità di probabilità.*

- Per la condizione di normalizzazione: $1 = \int_0^{+\infty} Ax^n e^{-x}\,dx = An! \Rightarrow A = \frac{1}{n!}$.

b) Si calcolino il valore medio e la deviazione standard della variabile casuale continua la cui densità di probabilità è $g_2(x) = \begin{cases} 0; & x < 0 \\ \frac{x^2e^{-x}}{2}; & x \ge 0 \end{cases}$.

- Il valor medio è $\overline{x} = \int_0^{+\infty} x \frac{(x^2e^{-x})}{2} dx = \lim_{x\to+\infty} F_3(x) = \frac{3!}{2} = 3$, mentre la deviazione standard è data dalla relazione $\sigma_x = \int_0^{+\infty} (x-3)^2 \frac{(x^2e^{-x})}{2} dx = 3$.