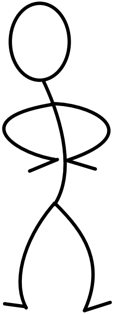
Distribuzione binomiale



Definizione.

Sia $\langle \Omega, \mathcal{E}, P \rangle$ uno spazio di probabilità.

Siano B e R due eventi tali che $B = \neg R$

Sia $p = P(B)$ e quindi $P(R) = P(\neg B) = 1 - p$.

spesso di scrive q per $1 - p$

La variabile aleatoria $X : \Omega^n \to \mathbb{R}$ così definita:

$$X(\langle x_1, \ldots, x_n \rangle) = \Big|\big\{ i \ : \ x_i \in B \big\}\Big|$$

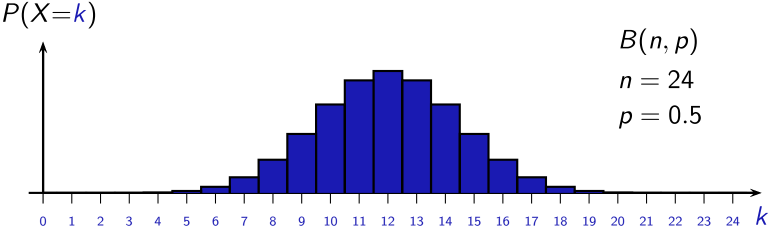
Si dice **v.a. binomiale** di parametri n e p.

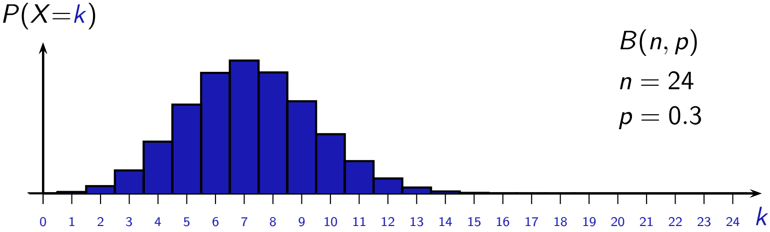
Si indica con $B(n, p)$.

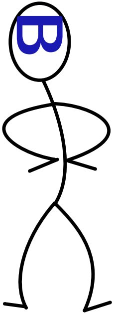
Per quanto visto $P(X{=}k) \ = \ \binom{n}{k} p^k (1-p)^{n-k}$.

$B(1, p)$ si dice **v.a. di Bernoulli**.

Esempi.


$P(X=k)$
$B(n,p)$
$n = 24$
$p = 0.5$
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
k


$P(X=k)$
$B(n,p)$
$n = 24$
$p = 0.3$
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
k


B

Esempio.

La prevalenza del daltonismo nella popolazione maschile è $p = 6\%$.

Qual è la probabilità di avere almeno 2 daltonici in un campione di 25?

Il numero di daltonici in una popolazione di 25 è una v.a. $B(25, p)$

La probabilità di avere esattamente k daltonici è $\binom{25}{k} p^k (1-p)^{25-k}$.

Quindi la risposta è $\displaystyle\sum_{k=2}^{25} \binom{25}{k} p^k (1-p)^{25-k} = 45\%$

Un'altra risposta è $1 - p'$ dove p' è la probabilità di avere < 2 daltonici.

Poiché $p' = \binom{25}{0} p^0 (1-p)^{25} + \binom{25}{1} p^1 (1-p)^{25-1}$

Otteniamo $1 - (1-p)^{25} - 25\, p\, (1-p)^{24}$. (Più semplice da calcolare)

Somma di v.a. indipendenti.

Per $i = 1, 2$ siano $X_i : \Omega \to \mathbb{R}$ variabili aleatorie indipendenti $B(1, p)$.

Esempio: $\Omega = \mathbf{U}^2$, ovvero due estrazioni dall'urna di biglie rosse e blu.

X_1 è 0 se la 1^{a} biglia estratta è rossa, 1 se è blu.

X_2 è 0 se la 2^{a} biglia estratta è rossa, 1 se è blu.

$$X := X_1 + X_2 \ = \begin{cases} 0 & \text{se nessuna biglia estratta è blu} \\ 1 & \text{se una biglia estratta è blu} \\ 2 & \text{se entrambe le biglie estratte sono blu} \end{cases}$$

Inoltre

$P(X{=}0) = (1-p)^2$ $\qquad$ $P(X{=}1) = 2p(p-1)$ $\qquad$ $P(X{=}2) = p^2$

Quindi X è una variabile aleatoria $B(2, p)$.

Somma di v.a.i. $B(n, p)$.

In generale:

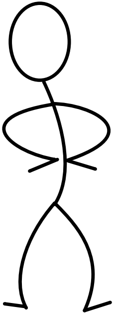
- Se $i = 1, 2$ sono $X_i : \Omega \to \mathbb{R}$ v.a.i. con distribuzione $B(n, p)$ e $B(m, p)$.

  Allora $X := X_1 + X_2$ è una variabile aleatoria $B(n + m, p)$.

- E se $i = 1, \ldots, n$ sono $X_i : \Omega \to \mathbb{R}$ v.a.i. con distribuzione $B(1, p)$.

  Allora $X := X_1 + \cdots + X_n$ è una variabile aleatoria $B(n, p)$.

Distribuzione di Poisson



Definizione

Una variabile aleatoria $X : \Omega \to \mathbb{R}$ si dice **di Poisson** se

$$P(X{=}k) \;=\; e^{-\mu}\frac{\mu^k}{k!}$$

La distrubuzione di Poisson approssima $B(n, \frac{\mu}{n})$ quando $n \to \infty$.

Infatti:

$$\lim_{n\to\infty} \binom{n}{k}\left(\frac{\mu}{n}\right)^k\left(1-\frac{\mu}{n}\right)^{n-k} \;=\; e^{-\mu}\frac{\mu^k}{k!}$$

L'approssimazione è buona per $n \geq 100$ e $p \leq 0.1$

k	Binomiale $n = 100$ $p = 0.01$	Poisson $\mu = 1$
0	0.3660	0.3679
1	0.3697	0.3679
2	0.1849	0.1839
3	0.0610	0.0613
4	0.0146	0.0153
5	0.0028	0.0031

Definizione

Una variabile aleatoria $X : \Omega \to \mathbb{N}$ si dice **di Poisson** se

$$P(X{=}k) \; = \; e^{-\mu}\frac{\mu^k}{k!}$$

La distrubuzione di Poisson approssima $B(n, \frac{\mu}{n})$ quando $n \to \infty$.

Infatti:

$$\lim_{n\to\infty} \binom{n}{k}\left(\frac{\mu}{n}\right)^k\left(1-\frac{\mu}{n}\right)^{n-k} \; = \; e^{-\mu}\frac{\mu^k}{k!}$$

L'approssimazione è meno buona per p grande

k	Binomiale $n = 100$ $p = 0.33$	Poisson $\mu = 33$
20	0.002	0.004
25	0.020	0.028
30	0.071	0.063
35	0.076	0.063
40	0.028	0.031
45	0.004	0.008

Esempio

Problema. Sia B una patologia rara. La probabilità di osservare un caso di B in un anno, è di 0.01 (la probabilità di osservarne 2 è trascurabile). Qual è la probabilità di osservare più di 3 casi in 20 anni?

Primo approccio. Immaginiamo che osservare B in un dato anno sia come estrarre una biglia blu da urna che contiene 99 biglie rosse e 1 biglia blu. La probabilità di estrarre ≥ 3 biglie blu in 20 estrazioni è

$$\sum_{k=3}^{20} \binom{20}{k} (0.1)^k (1-0.1)^{20-k} \quad \sim \quad 0.001$$

Secondo approccio. L'intervallo temporale scelto (un anno) è arbitrario. Avremmo potuto supporre che la probabilità di B in un mese sia $0.01/12$ e chiedere la probabilità di osservare più di 3 casi in 240 mesi? Quindi la distribuzione di Poisson è concettualmente ben giustificata.

$$\sum_{k=3}^{\infty} e^{-0.2} \frac{(0.2)^k}{k!} \quad \sim \quad 0.001$$