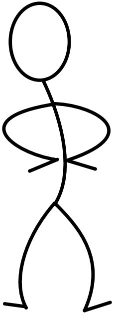
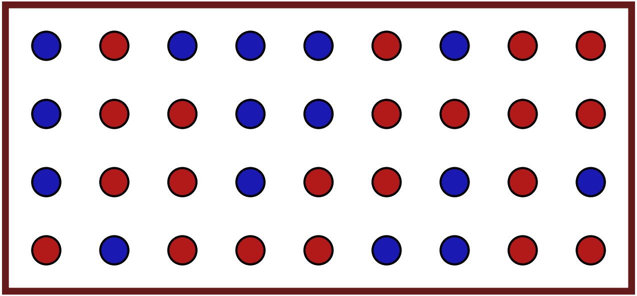
Distribuzione geometrica



Esempio (a)

Consideriamo la seguente variabile aleatoria T.

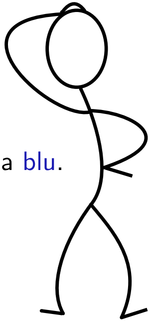


Estraiamo ripetutamente una biglia (con reimbussolamento).

T conta il numero di biglie rosse estratte prima della prima biglia blu.

Se blu esce alla prima estrazione 0, se esce alla seconda 1, ecc.

In quale spazio di probabilità “vive” questa v.a.?



Esempio (b)

Lo spazio campionario è $\boldsymbol{U}^{\mathbb{N}}$.

Con $\boldsymbol{U}^{\mathbb{N}}$ denotiamo un insieme delle sequenze (infinite) di elementi di U:

$$\boldsymbol{U}^{\mathbb{N}} = \Big\{\langle x_0, x_1, \ldots, x_i, \ldots\rangle \ : \ x_i \in \boldsymbol{U} \text{ per ogni } i \in \mathbb{N}\Big\}$$

Come eventi prendiamo gli insiemi della forma (esempio):

$$R \times B \times B \times R \times \boldsymbol{U}^{\mathbb{N} \setminus 4}$$

$$= \{4, 5, \ldots\}$$

Questo è l'insieme delle sequenze che cominciano con rossa,blu,blu,rossa

A questo evento ha probabilità:

$$P\Big(R \times B \times B \times R \times \boldsymbol{U}^{\mathbb{N} \setminus 4}\Big) = P(R) \cdot P(B) \cdot P(B) \cdot P(R)$$

Esempio (c)

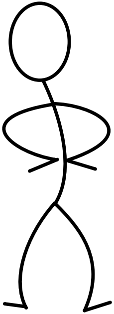
Possiamo ora calcolare $P(T{=}3)$.

Ovvero, la probabilità di ottenere rossa,rossa,rossa,blu:

$$P\Big(R \times R \times R \times B \times \boldsymbol{U}^{\mathbb{N}\smallsetminus 4}\Big) \;=\; \frac{21}{36}\cdot\frac{21}{36}\cdot\frac{21}{36}\cdot\frac{15}{36}$$

In generale:

$$P(T{=}k) \;=\; \left(\frac{21}{36}\right)^k \cdot \frac{15}{36}$$



Terminologia.

Sia $\langle\Omega, \mathcal{E}, P\rangle$ uno spazio di probabilità con due eventi non banali, cioè:

$$\mathcal{E} = \big\{\varnothing,\ \Omega,\ R,\ B\big\} \qquad\qquad \text{dove } R = \neg B.$$

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

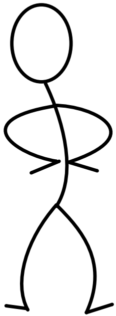
La variabile aleatoria $T : \Omega^{\mathbb{N}} \to \mathbb{N}$ così definita:

$$T(\langle x_0, x_1, \ldots\rangle) = \min\{i : x_i \in B\}$$

Si dice **v.a. geometrica** di parametro p.

Per quanto visto sopra

$$P(T{=}k) \;=\; (1-p)^k \cdot p$$



Esempio (d)

Qual'è la probabilità che una biglia blu non venga estratta prima della quinta estrazione? Ovvero quanto vale $P(T \geq 5)$?

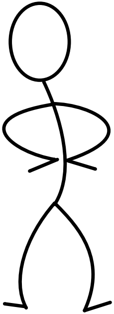
$$\sum_{k=5}^{\infty} P(T=k) \;=\; \sum_{k=5}^{\infty}(1-p)^k \cdot p$$

$$1 - \sum_{k=0}^{4} P(T=k) \;=\; \sum_{k=0}^{4}(1-p)^k \cdot p$$

```
sum ( dgeom ( 0:4, 15/36 ) ) = 0.9324564

pgeom ( 4, 15/36 ) = 0.9324564
```

Distribuzione di Poisson



Definizione

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

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

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

Infatti:

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

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

k	Binomiale $n = 100$ $p = 0.01$	Poisson $\lambda = 1$
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

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

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

Infatti:

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

L'approssimazione è meno buona per p grande

k	Binomiale $n = 100$ $p = 0.33$	Poisson $\lambda = 33$
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

Calcoliamo la probabilità che in un anno si verifichino più di 10 incidenti tra un ciclista e un mezzo a motore supponendo che la probabilità che si verifichi un tale incidente in un giorno sia 0.024.

Posto

$$\lambda = 365(0.024) = 8.76$$

Allora

$$\begin{aligned} P(X = k) &= \frac{e^{-8.76}(8.76)^k}{k!} \\ P(X > 10) &= \sum_{k=11}^{+\infty} \frac{e^{-8.76}(8.76)^k}{k!} \\ P(X > 10) &= 1 - \sum_{k=0}^{10} \frac{e^{-8.76}(8.76)^k}{k!} \end{aligned}$$

Otteniamo $P(X > 10) = 27\%$

```
1 - ppois( 10, 8.76 ) = 0.265962
```

Ogni anno in Piemonte transitano (o si formano) 127 temporali. Qual'è la probabilità che in un anno se ne verifichino più di 150?.

Posto

$$\lambda = 127$$

Allora

$$\begin{aligned} P(X = k) &= \frac{e^{-127}(127)^k}{k!} \\ P(X > 150) &= \sum_{k=151}^{+\infty} \frac{e^{-127}(127)^k}{k!} \\ P(X > 150) &= 1 - \sum_{k=0}^{150} \frac{e^{-127}(127)^k}{k!} \end{aligned}$$

Otteniamo $P(X > 150) = 2.1\%$ `1 - ppois( 150, 127 ) = 0.02068217`

Esempio

Ogni anno in Piemonte ci sono mediamente 63 giorni piovosi. Qual'è la probabilità che in un anno se ne verifichino più di 75?

$$P(X=k) = \binom{365}{k}\left(\frac{63}{365}\right)^k\left(1-\frac{63}{365}\right)^{365-k}$$

$$P(X>75) = \sum_{k=76}^{365} P(X=k) = 4.4\%$$

```
1 - pbinom(75,365,63/365) = 0.04417319
```

N.B. Non sarebbe corretto usare la distribuzione di Poisson

$$P(X=k) = \frac{e^{-63}(63)^k}{k!}$$

$$P(X>75) = \sum_{k=76}^{+\infty} \frac{e^{-63}(63)^k}{k!} = 1-\sum_{k=0}^{75} \frac{e^{-63}(63)^k}{k!} = 6.1\%$$

```
1 - ppois( 75, 63 ) = 0.06092581
```

Somma di v.a.i. di Poisson.

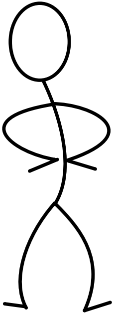
Siano X, Y v.a.i. di Poisson con parametro λ e μ. Allora

$$
\begin{aligned}
P(X+Y=k) &= \sum_{i=0}^{k} P(X=i) \cdot P(Y=k-i) \\
&= \sum_{i=0}^{k} e^{-\lambda} \frac{\lambda^i}{i!} e^{-\mu} \frac{\mu^{k-i}}{(k-i)!} \\
&= e^{-(\lambda+\mu)} \frac{1}{k!} \sum_{i=0}^{k} \binom{k}{i} \lambda^i \mu^{k-i} \\
&= e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^k}{k!}
\end{aligned}
$$

Binomio di Newton

Quindi $X+Y$ è una v.a. di Poisson con parametro $\lambda+\mu$.

Il valore atteso



Definizione

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

Sia $X : \Omega \to \{x_0, x_1 \dots\} \subseteq \mathbb{R}$ una variabile aleatoria discreta.

Il **valore atteso**
(o **media della popolazione**) di X è

In inglese **expected value**
oppure **population mean**

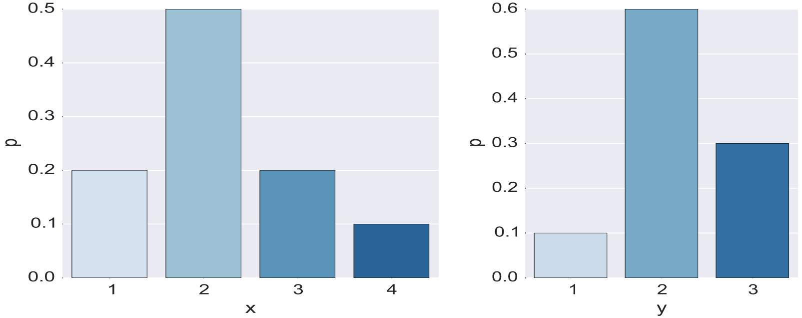
$$\boldsymbol{E(X)} = \sum_{k \in \mathbb{N}} x_k \cdot P(X{=}x_k)$$

È come scrivere $\sum_{k=0}^{\infty}$

Si indica anche con $\boldsymbol{\mu_X}$
oppure semplicemente con $\boldsymbol{\mu}$

Esempio

Le variabili aleatorie X ed Y hanno le seguenti distribuzioni.



$$E(X) = 1 \cdot 0.2 + 2 \cdot 0.5 + 3 \cdot 0.2 + 4 \cdot 0.1 = 2.2$$

$$E(Y) = 1 \cdot 0.1 + 2 \cdot 0.6 + 3 \cdot 0.3 = 2.2$$

Il valore atteso di $B(n, p)$

Se X è una v.a. con distribuzione $B(1, p)$ allora

$$\begin{aligned} E(X) &= \sum_{k\in\mathbb{N}} x_k \cdot P(X{=}x_k) \\ &= \sum_{k=0}^{1} x_k \cdot P(X{=}x_k) \\ &= 0 \cdot P(X{=}0) + 1 \cdot P(X{=}1) \;=\; 0 \cdot (1-p) + 1 \cdot p \;=\; p \end{aligned}$$

Se X è una v.a. con distribuzione $B(n, p)$ allora

$$E(X) \;=\; \sum_{k=0}^{n} k \cdot P(X{=}k) \;=\; \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \;=\, ?$$

Laborioso da calcolare (più avanti useremo un metodo più veloce).

Il valore atteso della somma di v.a.

Siano $X, Y : \Omega \to \mathbb{N}$ v.a. discrete a valori in $\mathbb{N}$ (per semplicità).

$$
\begin{aligned}
E(X+Y) &= \sum_{k\in\mathbb{N}} k \cdot P(X{+}Y{=}k) \\
&= \sum_{h,i\in\mathbb{N}} (h+i)\cdot P(X,Y=h,i) \\
&= \sum_{h,i\in\mathbb{N}} h\cdot P(X,Y=h,i) + \sum_{h,i\in\mathbb{N}} i\cdot P(X,Y=h,i) \\
&= \sum_{h\in\mathbb{N}} h\cdot P(X{=}h) + \sum_{i\in\mathbb{N}} i\cdot P(Y{=}i) \\
&= E(X)+E(Y)
\end{aligned}
$$

Si noti che: $\displaystyle\sum_{i\in\mathbb{N}} P(X,Y=h,i) \; = \; P(X{=}h)$

Il valore atteso della somma di v.a.

In generale:

date X_i, per $i = 1, \ldots, n$ variabili aleatorie a valori in $\mathbb{R}$

e posto $X = c_1 X_1 + \cdots + c_n X_n$, dove $c_1, \ldots, c_n \in \mathbb{R}$ sono costanti.

Ovvero X è **combinazione lineare** di $X_1, \ldots, X_n$

Allora

$$E(X) = c_1 E(X_1) + \cdots + c_n E(X_n)$$

Il valore atteso di $B(n, p)$

Sia X è una v.a. con distribuzione $B(n, p)$.

Possiamo immaginare che $X = X_1 + \cdots + X_n$ dove X_i sono $B(1, p)$.

Allora

$$
\begin{aligned}
E(X) &= E(X_1) + \ \ldots \ + E(X_n) \\
&= \quad p \quad + \ldots + \quad p \quad = \quad np
\end{aligned}
$$

Il valore atteso di una v.a. di Poisson

Sia X è una v.a. con distribuzione di Poisson con parametro λ.

$$
\begin{aligned}
E(X) &= \sum_{k=0}^{\infty} k\; e^{-\lambda} \frac{\lambda^k}{k!} &&= e^{-\lambda} \sum_{k=0}^{\infty} k\; \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=1}^{\infty} k\; \frac{\lambda^k}{k!} &&= e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} \\
&= \lambda\; e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} &&= \lambda\; e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \\
& &&= \lambda
\end{aligned}
$$

Il valore atteso di una v.a. geometrica

Sia X è una v.a. con distribuzione geometrica con parametro p.

$$
\begin{aligned}
E(X) &= \sum_{k=0}^{\infty} k\,(1-p)^k p &&= p\sum_{k=0}^{\infty} k\,(1-p)^k \\
&= p\,(1-p)\sum_{k=0}^{\infty} k\,(1-p)^{k-1} &&= p\,(p-1)\sum_{k=0}^{\infty} -\frac{d(1-p)^k}{dp} \\
&= p\,(1-p)\frac{d}{dp}-\sum_{k=0}^{\infty}(1-p)^k &&= p\,(1-p)\frac{d}{dp}-\frac{1}{p} \\
& &&= \frac{1-p}{p}
\end{aligned}
$$

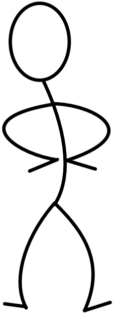
Il valore atteso del prodotto di v.a.i.

Siano X, Y v.a. indipendenti a valori in $\mathbb{N}$ (solo per per semplicità)

$$
\begin{aligned}
E(X \cdot Y) &= \sum_{k \in \mathbb{N}} k \cdot P(X \cdot Y = k) \\
&= \sum_{h,i \in \mathbb{N}} h \cdot i \cdot P(X, Y = h, i) \\
&= \sum_{h,i \in \mathbb{N}} h \cdot i \cdot P(X{=}h) \cdot P(Y{=}i) \\
&= \sum_{h \in \mathbb{N}} h \cdot P(X{=}h) \cdot \sum_{i \in \mathbb{N}} i \cdot P(Y{=}i) \\
&= E(X) \cdot E(Y)
\end{aligned}
$$

Perché indipendenti

La varianza



Definizione

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

Sia $X : \Omega \to \{x_0, x_1, \ldots, \} \subseteq \mathbb{R}$ una variabile aleatoria discreta.

La **varianza** di X è

$$\boldsymbol{Var(X)} = E\Big(\big(X - E(X)\big)^2\Big) = \sum_{k \in \mathbb{N}} \big(x_k - E(X)\big)^2 \cdot P(X{=}x_k)$$

$$= E\big(X^2\big) - E(X)^2$$

Si indica anche con $\boldsymbol{\sigma_X^2}$

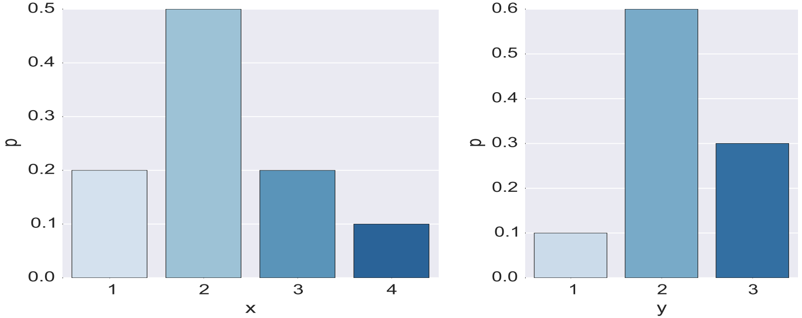
oppure semplicemente con $\boldsymbol{\sigma^2}$

$\sigma = \sqrt{Var(X)}$ si chiama **deviazione standard**

oppure **scarto quadratico medio**

Esempio

Le variabili aleatorie X ed Y hanno le seguenti distribuzioni.



$$Var(X) = (1 - 2.2)^2 \cdot 0.2 + (2 - 2.2)^2 \cdot 0.5 + (3 - 2.2)^2 \cdot 0.3 + (4 - 2.2)^2 \cdot 0.1$$

$$Var(Y) = (1 - 2.2)^2 \cdot 0.1 + (2 - 2.2)^2 \cdot 0.6 + (3 - 2.2)^2 \cdot 0.3$$

$$Var(X) - Var(Y) = (1.2)^2 \cdot 0.1 - (0.8)^2 \cdot 0.1 + (1.8)^2 \cdot 0.1 > 0$$

La covarianza.

Siano X, Y variabili aleatorie.

$$
\begin{aligned}
Var(X+Y) &= E\big((X+Y)^2\big) - E(X+Y)^2 \\
&= E(X^2+Y^2+2XY) - \big[E(X)+E(Y)\big]^2 \\
&= E(X^2)+E(Y^2)+E(2XY)-E(X)^2-E(Y)^2-2E(X)E(Y) \\
&= Var(X)+Var(Y) \; + \; 2\left[\boxed{E(XY)-E(X)E(Y)}\right]
\end{aligned}
$$

Si chiama **covarianza**

Si indica con $\boldsymbol{Cov(XY)}$

E vale $\boxed{\textbf{0 se } \boldsymbol{X, Y} \textbf{ indipendenti}}$

La varianza di $B(n, p)$

Sia X è una v.a. con distribuzione $B(1, p)$.

Allora

Nel nostro caso $X^2 = X$.

$$Var(X) \;=\; E(X^2) - E(X)^2 \;=\; p - p^2 \;=\; p\,(1-p).$$

Sia X è una v.a. con distribuzione $B(n, p)$.

Possiamo immaginare che $X = X_1 + \cdots + X_n$ dove X_i sono $B(1, p)$.

Allora

$$Var(X) = Var(X_1) + \cdots + Var(X_n) = n\,p\,(1-p)$$

La varianza di una v.a. di Poisson

Sia X è una v.a. con distribuzione di Poisson con parametro λ.

$$\begin{aligned}
E(X^2) &= \sum_{k=0}^{\infty} k^2\, e^{-\lambda} \frac{\lambda^k}{k!} &&= \sum_{k=1}^{\infty} k^2\, e^{-\lambda} \frac{\lambda^k}{k!} \\
&= \sum_{k=1}^{\infty} k\, e^{-\lambda} \frac{\lambda^k}{(k-1)!} &&= \lambda\, e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda\, e^{-\lambda} \sum_{k=0}^{\infty} (k+1) \frac{\lambda^k}{k!} &&= \lambda\, E(X+1) \\
&= \lambda\, (E(X)+1) &&= \lambda\, (\lambda+1)
\end{aligned}$$

Quindi $Var(X) \;=\; E(X^2) - \lambda^2 = \lambda$