**Lezione** 15/11/2022

# Variabile geometrica

**Definizione** (Variabile Geometrica)**.** *In uno schema* successo/insuccesso con rimpiazzo*, la variabile geometrica $X$ di parametro $p$, rappresenta il numero di prove necessarie affinchè si verifichi il primo successo, dove $p$ è la probabilità che si verifichi un successo in una singola prova. Si indica con $X \sim Geo(p)$.*

**Esempio.** *Lanciamo una moneta regolare finchè non otteniamo testa. La variabile $X$, che indica il numero di lanci necessari affinchè esca per la prima volta testa, è una variabile geometrica di parametro $\frac{1}{2}$ ($X \sim Geo(\frac{1}{2})$). Dunque,*

- *$X = 1$ vuol dire che è uscito testa al primo lancio (T);*
- *$X = 2$ vuol dire che è uscito testa al secondo lancio (CT);*
- *$X = 3$ vuol dire che è uscito testa al terzo lancio (CCT);*
- *$X = i$ vuol dire che è uscito testa all'i-esimo lancio.*

Troviamo la distribuzione di probabilità di una v.a. geometrica:

**Esercizio 1.** *Supponiamo di lanciare un dado tante volte fino a quando non otteniamo 6. Qual è la probabilità che mi fermi dopo 5 lanci?*

**Svolgimento.** *Sia $X$ la variabile aleatoria che conta il numero necessario di prove per ottenere per la prima volta 6. $X$ è una variabile geometrica di parametro $\frac{1}{6}$ ($X \sim Geo(\frac{1}{6})$). La probabilità che mi fermi dopo 5 lanci è $P(X = 5)$. Osservo che $X = 5$ equivale all' intersezione degli eventi $X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 0$ e $X_5 = 1$, dove $X_i$ è la variabile di Bernuilli che descrive ogni singolo lancio del dado ($X_i = 1$ se esce 6 e $X_i = 0$ altrimenti). Dunque,*

$$P(X = 5) = P(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 1) =$$

$$P(X_1 = 0) \cdot P(X_2 = 0) \cdot P(X_3 = 0) \cdot P(X_4 = 0) \cdot P(X_5 = 1)$$

$$(\textit{dato che gli eventi sono indipendenti}) = \left(\frac{5}{6}\right) \cdot \left(\frac{5}{6}\right) \cdot \left(\frac{5}{6}\right) \cdot \left(\frac{5}{6}\right) \cdot \left(\frac{1}{6}\right) =$$

$$\left(\frac{5}{6}\right)^4 \cdot \left(\frac{1}{6}\right)$$

In generale, la distribuzione di probabilità di $X \sim Geo(p)$ è data dal seguente teorema:

**Teorema** (Distribuzione di probabilità di una variabile geometrica)**.** *Sia $X$ una variabile geometrica di parametro $p$, allora*

$$P(X = k) = p(1-p)^{k-1}.$$

**Teorema.** *Si dimostra che $P(X > k) = (1-p)^k$.*

**Teorema.** *Si dimostra che $P(X \geq k) = (1-p)^{k-1}$.*

**Dimostrazione.** $P(X \geq k) = P(X = k) + P(X > k) = p \cdot (1-p)^{k-1} + (1-p)^k = (1-p)^k \cdot \left[\frac{p}{1-p} + 1\right] = \frac{(1-p)^k}{(1-p)} = (1-p)^{k-1}$.

**Esercizio 2.** *Da un'urna con 10 palline bianche e 15 palline nere, si eseguono estrazioni con rimpiazzo fino all'estrazione di una pallina nera.*

*(a) Calcola la probabilità che servano 20 estrazioni (per l'estrazione della pallina nera);*

*(b) calcola la probabilità che servano almeno 10 estrazioni.*

**Svolgimento.** *(a) L'estrazione con rimpiazzo è uno schema successo/insuccesso con rimpiazzo, dove il "successo" è l'evento "esce una pallina bianca" (perchè è l'evento che ci interessa). In ogni singola prova, la probabilità di ottenere successo è $\frac{3}{5}$. Se $X \sim Geo(\frac{3}{5})$, allora dobbiamo calcolare $P(X = 20)$.*

$$P(X = 20) = \frac{3}{5} \cdot \left(\frac{2}{5}\right)^{19}.$$

*(b)* $P(X \geq 10) = \left(\frac{2}{5}\right)^{9}.$

## Proprietà di mancanza di memoria della variabile geometrica

**Teorema** (Proprietà di mancanza di memoria della variabile geometrica)**.** *Se $X$ è una variabile geometrica, allora*

$$P(X = k + m \mid X > k) = P(X = m).$$

**Svolgimento.** *Sia* $X \sim Geo(p)$, *allora*
$P(X = k+m \mid X > k) = \dfrac{P(X = k+m, X > k)}{P(X > k)}$ *(per la formula della probabilità condizionata)* $= \dfrac{P(X = k+m)}{P(X > k)}$ *(l'intersezione di* $X = k+m$ *e* $X > k$ *è* $X = k+m$*)*$=\dfrac{p \cdot (1-p)^{k+m-1}}{(1-p)^k}$ *(applicando la formula della probabilità della variabile geometrica)*$=p \cdot (1-p)^{m-1}$ $= P(X = m)$ *(applicando la formula della probabilità della variabile geometrica).*

**Esempio.** *Sia* $X$ *la variabile aleatoria che rappresenta il numero di prove necessarie per ottenere un numero rosso al gioco della Roulette Francese.*

*Ad ogni partita P(esce un numero nero)=P(esce un numero rosso)=*$\dfrac{18}{37}$ *(la roulette è composta da 37 numeri, 18 rossi, 18 neri e 1 verde).*

*Supponiamo che per cento partite esce sempre un numero nero, alla 101-esima partita la probabilità che esca un numero nero è ancora uguale alla probabilità che esca un numero rosso; il ragionamento intuitivo secondo cui, alla 101-esima partita, è più conveniente puntare sul numero rosso (perchè non è uscito per tante partite) è errato!*

*Per la proprietà di mancanza di memoria di* $X$,

$$P(X = 101 \mid X > 100) = P(X = 1),$$

*dove*

- $X = 101$ *è l'evento "il primo numero rosso esce alla centunesima giocata";*
- $X > 100$ *è l'evento "il primo numero rosso esce dopo cento giocate";*
- $X = 1$ *è l'evento "il primo numero rosso esce alla prima giocata".*

*Dunque, la probabilità che alla giocata 101 esca un numero rosso, sapendo che non è mai uscito in 100 gicate, è uguale alla probabilità che un numero rosso esca alla prima giocata.*

*Per questo motivo si dice che la roulette* non ha memoria, *ovvero non si ricorda delle giocate passate.*

*La proprietà di mancanza di memoria dipende dal fatto che gli eventi* esce un numero rosso alla giocata $i$ *ed* esce un numero rosso alla giocata $i+1$ *sono* indipendenti*: il verificarsi o meno dell'evento* esce un numero rosso alla giocata $i$, *non influisce sulla probabilità dell'evento* esce rosso alla giocata $i+1$. *A prescidere da cosa esce alla giocata* $i$, $P$(*esce un numero rosso alla giocata i+1*) *è sempre uguale a* $\dfrac{18}{37}$.

*Per lo stesso principio, la tecnica di puntare sui numeri ritardatari (i numeri che non escono da tanto tempo) nel gioco del lotto è infondata. In ogni giocata diversa, la probabilità di estrarre il primo numero è sempre uguale a $\frac{1}{90}$, a prescindere da quali siano stati i risultati nelle giocate precedenti.*

Abbiamo dimostrato che se $X$ è una variabile geometrica, allora verifica la proprietà di mancanza di memoria. Vale anche il viceversa: se $X$ è una variabile aleatoria discreta che verifica la proprietà di mancanza di memoria, allora $X$ è una variabile geometrica. Possiamo quindi concludere che le variabili geometriche sono tutte e sole le variabili discrete che soddisfano la proprietà di mancanza di memoria.

### Istogramma di una variabile geometrica

Se $X$ è una variabile geometrica, la probabilità $P(X = k)$ si avvicina sempre più a zero man mano che $k$ diventa più grande. Il seguente è un esempio.

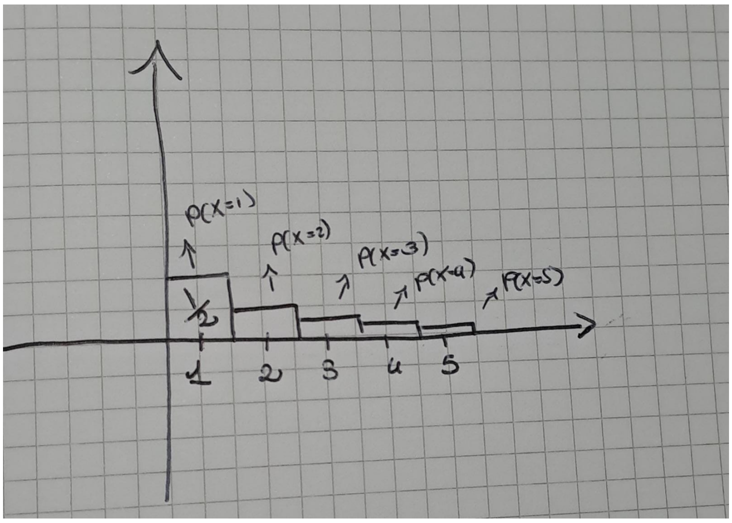
**Esempio.** *Lanciamo una moneta tante volte. Voglio calcolare la probabilità di ottenere testa al $k$-esimo lancio.*

*Sia $X$ la variabile che conta il numero di lanci necessari per ottenere testa per la prima volta, allora $X \sim Geo(\frac{1}{2})$.*

*La distribuzione di probabilità di $X$ è data dalla seguente tabella.*

| $k$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $P(X = k)$ | $\frac{1}{2}$ | $\left(\frac{1}{2}\right)^2$ | $\left(\frac{1}{2}\right)^3$ | $\left(\frac{1}{2}\right)^4$ |

*L'istogramma relativo a $X$ è il seguente.*




*Sia dall'istogramma che dalla tabella, si evince che la probabilità di $X$ decresce al crescere di $k$ e si avvicina sempre più allo zero.*

*Ciò si giustifica matematicamente considerando che*

$$\lim_{k \to +\infty} p \cdot (1-p)^{k-1} = 0.$$

*(nota che $p \leq 1$).*