Esercitazione 1 - Statistica (parte II)

Davide Passaretti

9/2/2017

Contents

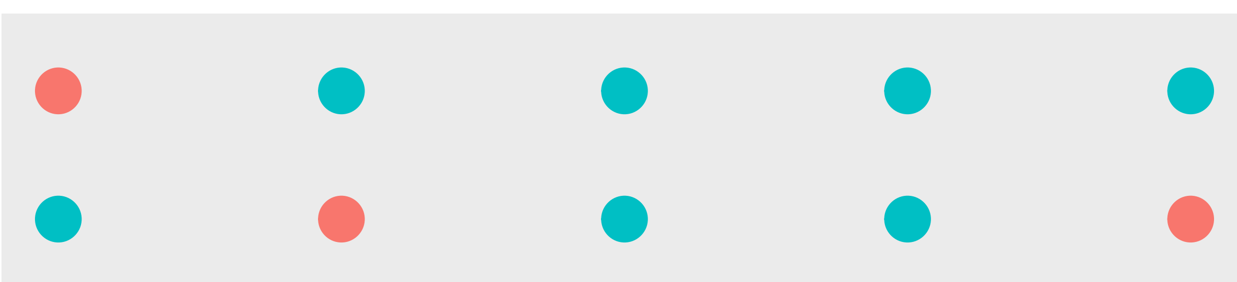
1 Variabili casuali discrete

Le variabili casuali discrete assumono valori finiti (es: successo = **1** ed insuccesso = **0**) oppure un'infinità numerabile di valori (es: numero di clienti in attesa = $\{\mathbf{0, 1, 2, \ldots}\}$).

Le variabili trattate saranno:

- Variabili che descrivono prove indipendenti:
  - Bernoulli, Binomiale e sua approssimazione alla Poisson
  - Geometrica e Binomiale Negativa
- Variabili che descrivono prove dipendenti:
  - Ipergeometrica
  - *Facoltativo*: come ricavare una v.c. Ipergeometrica Negativa

L'urna
Supponiamo di avere un'urna contenente 3 palline rosse e 7 palline azzurre.



2 Variabili che descrivono prove indipendenti

Le seguenti variabili partono dall'assunzione che la probabilità nelle varie prove sia costante ed uguale a π. Tale assunzione identifica uno schema di **campionamento con ripetizione** (o con rimessa): una volta estratta una pallina dall'urna, essa viene reinserita nell'urna prima dell'estrazione successiva. Per questo motivo, ciascuna estrazione (o prova) è indipendente dalle altre.

2.1 Variabile casuale di Bernoulli

Ciascuna estrazione indipendente è una v.c. di Bernoulli di parametro π:

$$\mathcal{X} \sim \mathbf{Ber}(\pi)$$

Nel nostro caso, π è la probabilità di estrarre una pallina rossa. Essendo presenti 3 palline rosse su 10 totali, $\pi = \frac{3}{10} = 0.3$.

Dimostrare che il valore atteso di una estrazione è uguale alla probabilità di successo.

Identificando con 1 i successi e con 0 gli insuccessi (palline azzurre):

$$E(\mathcal{X}) = 0.3 \times 1 + (1 - 0.3) \times 0 = 0.3$$

Dimostrare che la varianza di una estrazione è uguale al prodotto tra la probabilità di successo e quella di insuccesso.

La varianza è determinabile come media dei quadrati meno quadrato della media. Abbiamo appena visto che la media di una estrazione è uguale a $\pi = 0.3$.

$$Var(\mathcal{X}) = \underbrace{0.3 \times 1^2 + (1 - 0.3) \times 0^2}_{E(\mathcal{X}^2)} - \underbrace{0.3^2}_{[E(\mathcal{X})]^2} = 0.3 - 0.3^2 = 0.3(1 - 0.3) = 0.21$$

2.2 Variabile casuale Binomiale

La variabile casuale Binomiale è la somma di n prove Bernoulliane di parametro π:

$$\mathcal{Y} \sim \mathbf{Bin}(n, \pi)$$

Essendo una somma di n variabili indipendenti, il suo valore atteso è uguale a:

$$E(\mathcal{Y}) = n \times E(\mathcal{X}) = n\pi$$

e la sua varianza è uguale a:

$$Var(\mathcal{Y}) = n \times Var(\mathcal{X}) = n\pi(1 - \pi)$$

La sua `pmf` (*probability mass function*), ovvero la sua funzione di probabilità, descrive la probabilità di ottenere k successi in n prove indipendenti:

$$\mathbf{P}(\mathcal{Y} = k) = \binom{n}{k} \pi^k \, (1 - \pi)^{n-k}$$

Si noti che il numero di successi k ha senso solo se minore o uguale al numero delle prove n. Inoltre, se $n = 1$, allora $k \leq 1$ e $\mathcal{Y}$ risulta uguale ad $\mathcal{X}$. Infatti, per $n = k = 1$, ci si sta chiedendo qual è la probabilità di ottenere 1 successo in 1 prova, che è esattamente ciò che la v.c. di Bernoulli descrive.

a) Qual è la probabilità di pescare 3 palline rosse in 5 estrazioni indipendenti?

In questo caso, $n = 5$, $\pi = 0.3$, $k = 3$:

$$\mathbf{P}(\mathcal{Y} = 3) = \binom{5}{3} 0.3^3 \, (1 - 0.3)^{5-3} = \frac{5!}{3! \, (5-3)!} \, 0.3^3 \times 0.7^2 = 0.1323$$

b) Qual è la probabilità di pescare almeno una pallina rossa in 6 estrazioni indipendenti?

In questo caso, $n = 6$, $\pi = 0.3$, $k > 0$. Dunque, è più facile calcolare il complemento ad 1 della probabilità di non estrarre alcuna pallina rossa nelle prima 6 estrazioni:

$$\mathbf{P}(\mathcal{Y} > 0) = \mathbf{P}(\mathcal{Y} \neq 0) = 1 - \mathbf{P}(\mathcal{Y} = 0) = 1 - \left[\binom{6}{0} 0.3^0 \, (1 - 0.3)^{6-0} \right] =$$

$$= 1 - [1 \times 1 \times 0.7^6] = 1 - 0.117649 = 0.882351$$

c) Qual è la probabilità di pescare almeno due palline rosse in 6 estrazioni indipendenti?

In questo caso, $n = 6$, $\pi = 0.3$, $k > 1$. Dunque, è più facile calcolare il complemento ad 1 della probabilità di estrarre al più una pallina rossa nelle prima 6 estrazioni:

$$\mathbf{P}(\mathcal{Y} > 1) = 1 - \overbrace{[\mathbf{P}(\mathcal{Y} = 0) + \mathbf{P}(\mathcal{Y} = 1)]}^{\mathbf{P}(\mathcal{Y} \leq 1)} = 1 - \left[0.117649 + \binom{6}{1} 0.3^1 \, (1 - 0.3)^{6-1} \right] =$$

$$= 1 - [0.117649 + 6 \times 0.3 \times 0.7^5] = 1 - 0.420175 = 0.579825$$

d) Qual è la probabilità di pescare al più una pallina rossa in 9 estrazioni indipendenti?

In questo caso, $n = 9$, $\pi = 0.3$, $k = \{0, 1\}$. Dunque, bisogna sommare i valori della `pmf` corrispondenti ai due diversi k.

$$\mathbf{P}(\mathcal{Y} \leq 1) = \mathbf{P}(\mathcal{Y} = 0) + \mathbf{P}(\mathcal{Y} = 1) = 0.7^9 + \binom{9}{1} 0.3^1 \, (1 - 0.3)^{9-1} =$$

$$= 0.7^9 + 9 \times 0.3 \times 0.7^8 = 0.1960032$$

2.3 Approssimazione della v.c. Binomiale alla v.c. di Poisson

La v.c. di Poisson viene usata essenzialmente in due contesti:

- come approssimazione della v.c. Binomiale per n grande e π piccolo
- come probabilità di accadimento di eventi rari tra loro indipendenti in intervalli spaziali o temporali.

Qui la tratteremo come approssimazione della Binomiale, rimandando l'altra accezione alla prossima esercitazione in cui verrà legata alla v.c. Esponenziale Negativa.

Detto ciò, la v.c. di Poisson rappresenta una buona approssimazione del modello Binomiale se $n > 100$ e $n\pi < 10$ (ma queste sono misure indicative fornite in letteratura). In realtà si cerca un numero di prove sufficientemente ampio ed una probabilità di successo vicina quanto più possibile allo zero. Vedremo come nel nostro esempio l'approssimazione è sconsigliata.

La convergenza della Binomiale alla Poisson è indicata a seguire:

$$\mathcal{Y} \sim \mathbf{Bin}(n, \pi) \rightarrow \mathcal{P} \sim \mathbf{Pois}(\lambda = n\pi)$$

Si noti che la media e la varianza della v.c. di Poisson sono entrambe uguali a λ. La `pmf` è uguale a:

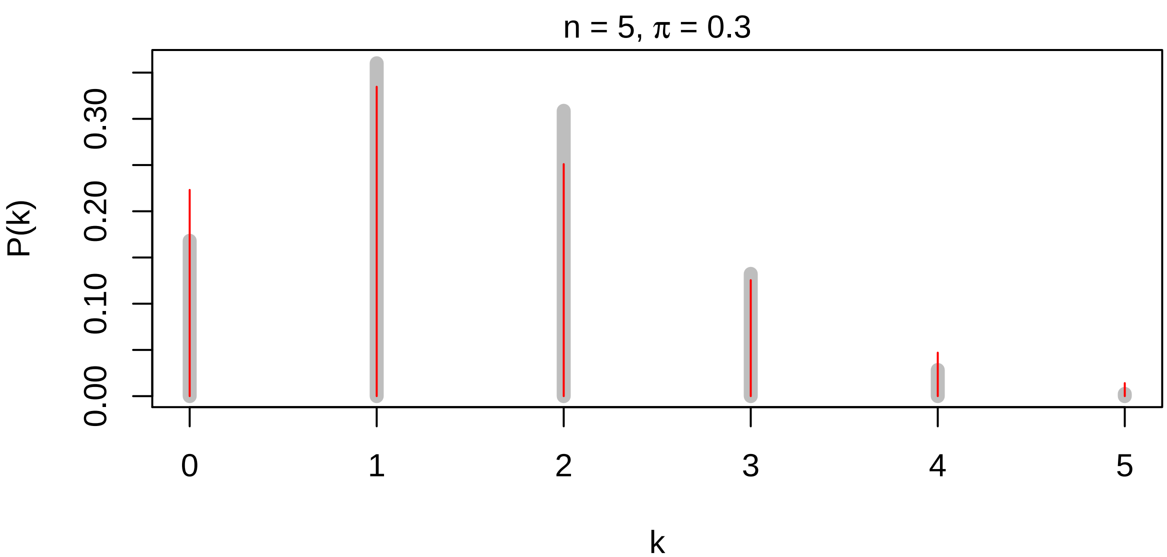
$$\mathbf{P}(\mathcal{P} = k) = \frac{\lambda^k \, e^{-\lambda}}{k!}$$

Ricalcoliamo i punti a), b), c), d) dell'esercizio precedente usando l'approssimazione della v.c. Binomiale alla Poisson e verifichiamo con l'aiuto dei grafici quanto differiscono le `pmf`. Le aste grigie indicano la `pmf` della v.c. Binomiale, mentre quelle rosse (più sottili) indicano l'approssimazione alla Poisson.

a) Qual è la probabilità di pescare 3 palline rosse in 5 estrazioni indipendenti?

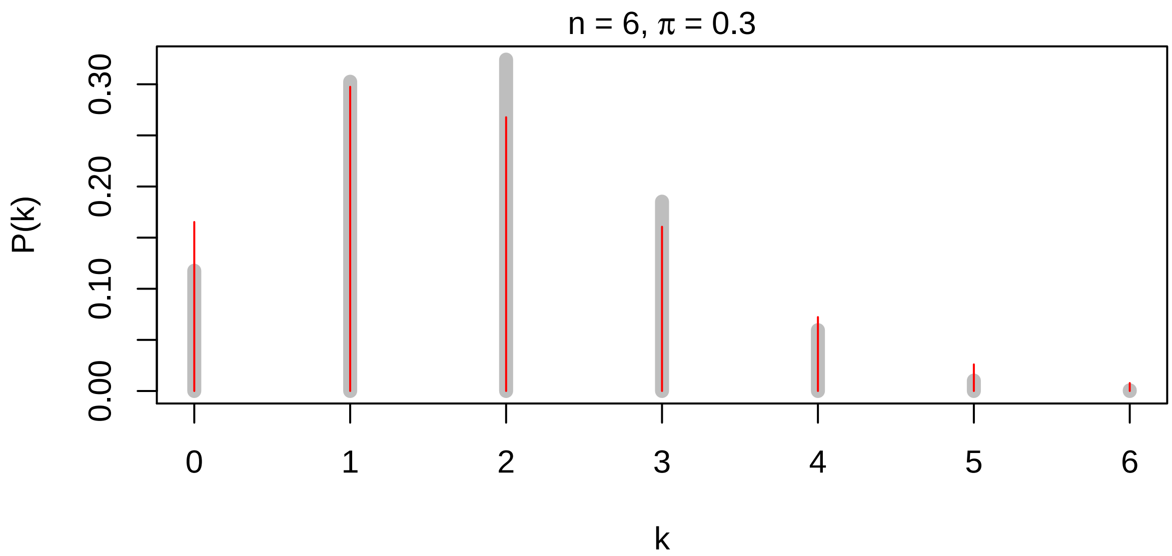
In questo caso $\lambda = 5 \times 0.3 = 1.5$.

$$\mathbf{P}(\mathcal{Y} = 3) \approx \mathbf{P}(\mathcal{P} = 3) = \frac{1.5^3 \, e^{-1.5}}{3!} = 0.1255107$$



b) Qual è la probabilità di pescare almeno una pallina rossa in 6 estrazioni indipendenti?

In questo caso $\lambda = 6 \times 0.3 = 1.8$.



$$\mathbf{P}(\mathcal{Y} > 0) \approx 1 - \mathbf{P}(\mathcal{P} = 0) = 1 - \frac{1.8^0\, e^{-1.8}}{0!} = 1 - e^{-1.8} = 0.8347011$$

c) Qual è la probabilità di pescare almeno due palline rosse in 6 estrazioni indipendenti?

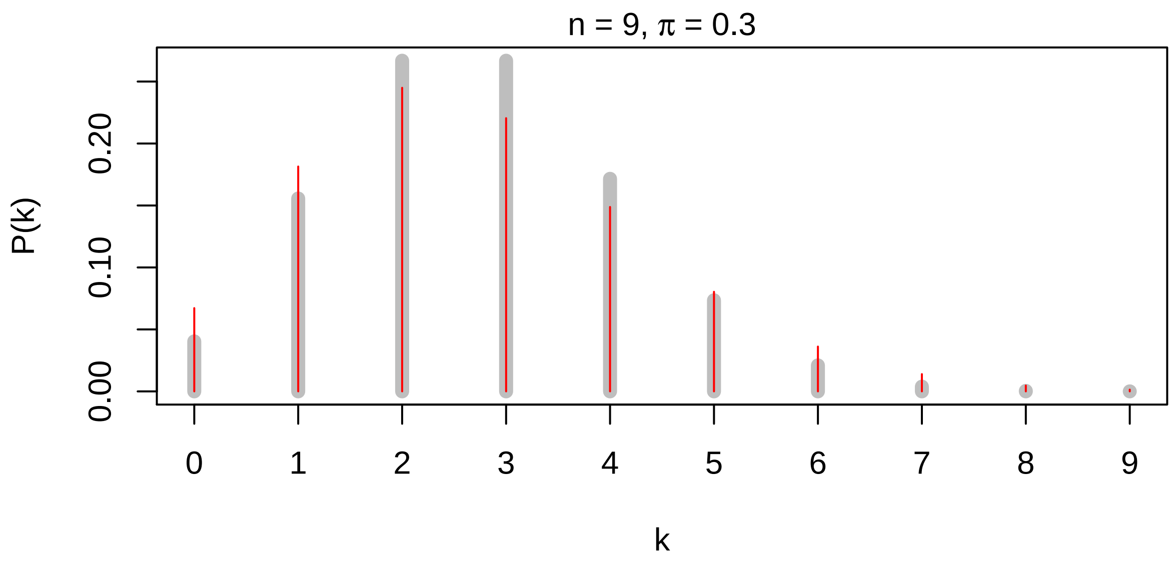
In questo caso $\lambda = 6 \times 0.3 = 1.8$.

Il grafico è identico a quello del punto precedente, perciò non è riportato.

$$\mathbf{P}(\mathcal{Y} > 1) \approx 1 - [\mathbf{P}(\mathcal{P} = 0) + \mathbf{P}(\mathcal{P} = 1)] = 1 - \left[e^{-1.8} + \frac{1.8^1\, e^{-1.8}}{1!}\right] = 1 - 2.8\, e^{-1.8} = 0.5371631$$

d) Qual è la probabilità di pescare al più una pallina rossa in 9 estrazioni indipendenti?

In questo caso $\lambda = 9 \times 0.3 = 2.7$.



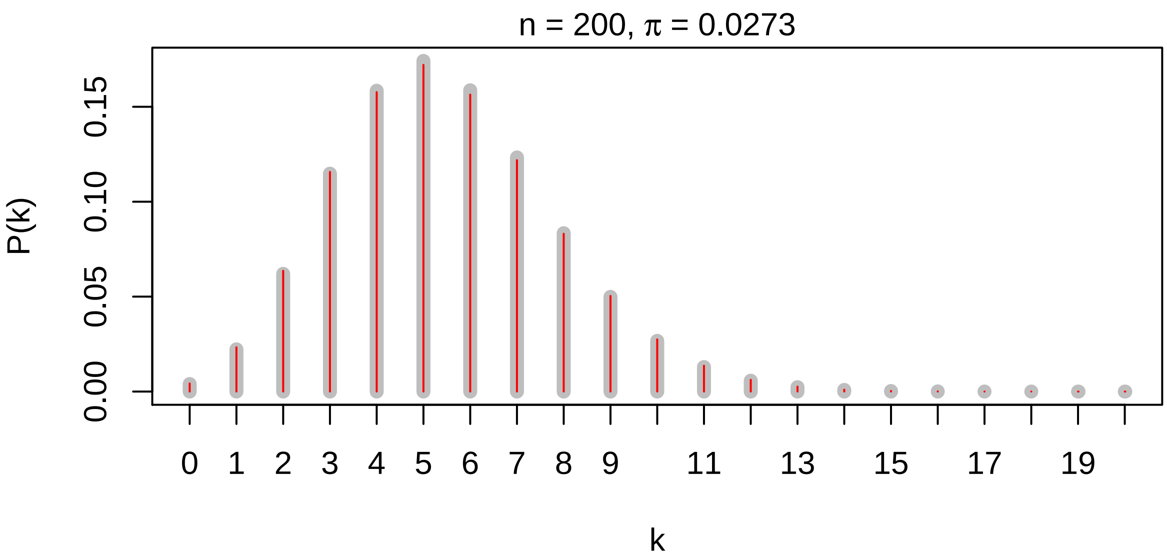
$$\mathbf{P}(\mathcal{Y} \leq 1) \approx \mathbf{P}(\mathcal{P} = 0) + \mathbf{P}(\mathcal{P} = 1) = \frac{2.7^0\, e^{-2.7}}{0!} + \frac{2.7^1\, e^{-2.7}}{1!} = 3.7\, e^{-2.7} = 0.2486604$$

e) Supponiamo di aggiungere altre 100 palline azzurre all'urna. Si calcoli, utilizzando sia la v.c. Binomiale che la Poisson, la probabilità di pescare 8 palline rosse in 200 prove indipendenti.

Aggiungendo 100 palline alle 10 già presenti, si ha che la *nuova* probabilità di successo è:

$$\pi = \frac{3}{110} \approx 0.0273$$

Tale valore fa sì che $n\pi = \lambda = 200 \times \frac{3}{110} = \frac{60}{11} \approx 5.45 < 10$. Inoltre $n = 200 > 100$. Quindi l'approssimazione della Binomiale alla Poisson è consigliata.



Rispondiamo ora al quesito:

- **Usando la v.c. Binomiale:**

$$P(\mathcal{Y} = 8) = \binom{200}{8} \left(\frac{3}{110}\right)^8 \left(1 - \frac{3}{110}\right)^{200-8} = 0.08341843$$

- **Usando la v.c. di Poisson:**

$$P(\mathcal{P} = 8) = \frac{\left(\frac{60}{11}\right)^8 \, e^{-\frac{60}{11}}}{8!} = 0.08311301$$

2.4 Variabile casuale Geometrica

La variabile casuale Geometrica descrive la probabilità di dover aspettare la n-esima prova Bernoulliana di parametro π per osservare il primo successo:

$$\mathcal{G} \sim \mathbf{Geo}(\pi)$$

La sua `pmf` è determinabile considerando che nelle prime $n-1$ prove devono essere osservati solo insuccessi, mentre nell'ultima prova bisogna osservare il successo:

$$\mathbf{P}(\mathcal{G} = n) = \underbrace{(1-\pi)^{n-1}}_{\mathbf{Bin}(n-1,\pi),\, 0 \text{ successi}} \times \pi$$

Ritorniamo all'urna originaria contenente 10 palline, con $\pi = 0.3$.

a) Qual è la probabilità che si debba aspettare la 11-esima estrazione indipendente prima di pescare una pallina rossa?

$$\mathbf{P}(\mathcal{G} = 11) = (1 - 0.3)^{11-1} \times 0.3 = 0.7^{10} \times 0.3 = 0.008474257$$

b) Qual è la probabilità che si debba aspettare almeno la terza estrazione indipendente prima di pescare una pallina rossa?

Qui $n \geq 3$. Conviene calcolare il complemento ad 1 della probabilità di osservare il primo successo alla prima prova o alla seconda prova.

$$\mathbf{P}(\mathcal{G} \geq 3) = 1 - \underbrace{[\mathbf{P}(\mathcal{G} = 1) + \mathbf{P}(\mathcal{G} = 2)]}_{\mathbf{P}(\mathcal{G} \leq 2)} = 1 - [(1 - 0.3)^0 \times 0.3 + (1 - 0.3)^1 \times 0.3] = 1 - 0.51 = 0.49$$

2.5 Variabile casuale Binomiale Negativa

La variabile casuale Binomiale Negativa (o *distribuzione di Pascal*) descrive la probabilità di dover aspettare la n-esima prova Bernoulliana di parametro π per osservare i primi k successi:

$$\mathcal{T} \sim \mathbf{BN}(k, \pi)$$

È subito evidente che, per $k = 1$, essa si riduca alla v.c. Geometrica di parametro π.

La sua `pmf` è determinabile considerando che nelle prime $n - 1$ prove devono essere osservati $k - 1$ successi, mentre nell'ultima prova bisogna osservare il k-esimo successo:

$$\mathbf{P}(\mathcal{T} = n) = \underbrace{\binom{n-1}{k-1} \pi^{k-1} (1 - \pi)^{(n-1)-(k-1)}}_{\mathbf{Bin}(n-1,\pi),\, k-1 \text{ successi}} \times \pi = \binom{n-1}{k-1} \pi^{k} (1 - \pi)^{n-k}$$

Si noti che, per $n = k$, $\mathbf{BN}(k, \pi) \equiv \mathbf{Bin}(n, \pi)$, in quanto osservare l'$n$-esimo successo all'$n$-esima prova equivale ad osservare n successi in n prove.

Ritorniamo alla nostra urna.

a) Qual è la probabilità che si debba aspettare la 11-esima estrazione indipendente prima di pescare quattro palline rosse?

Qui $n = 11$, $k = 4$.

$$\mathbf{P}(\mathcal{T} = 11) = \binom{11-1}{4-1} 0.3^4 (1 - 0.3)^{11-4} = 0.220133$$

b) Qual è la probabilità che si debba aspettare almeno la terza estrazione indipendente prima di pescare due palline rosse?

Qui $n \geq 3$, $k = 2$. Conviene calcolare il complemento ad 1 della probabilità di osservare i primi due successi alla prima prova o alla seconda.

$$\mathbf{P}(\mathcal{T} \geq 3) = 1 - [\mathbf{P}(\mathcal{T} = 1) + \mathbf{P}(\mathcal{T} = 2)] = 1 - \left[\underbrace{0}_{\substack{\text{impossibile osservare} \\ \text{2 successi in 1 prova}}} + \underbrace{\binom{2-1}{2-1} 0.3^2 (1 - 0.3)^{2-2}}_{\mathbf{Bin}(2,0.3),\, 2 \text{ successi}} \right] = 1 - 0.3^2 = 0.91$$

c) *Facoltativo e di riepilogo*: qual è la probabilità che si debba aspettare non più di due estrazioni indipendenti prima di pescare almeno 1 pallina rossa?

$$\mathbf{P}(k \geq 1, n \leq 2) = \overbrace{\underbrace{\mathbf{P}(k=1, \text{I prova})}_{\mathbf{Ber}(0.3),\text{ successo}}}^{\text{prob. successo alla I prova}} + \overbrace{\underbrace{\mathbf{P}(k=0, \text{I prova}) \times \mathbf{P}(k=1, \text{II prova})}_{\mathbf{Geo}(0.3),\ \texttt{2 prove}} + \underbrace{\mathbf{P}(k=2, \text{II prova})}_{\mathbf{BN}(2,0.3)\equiv\mathbf{Bin}(2,0.3)}}^{\text{prob. di dover attendere 2 prove per osservare almeno 1 successo}} =$$

$$= 0.3 + (1-0.3) \times 0.3 + 0.3^2 = 0.6$$

3 Variabili che descrivono prove dipendenti

Le seguenti variabili partono dall'assunzione che la probabilità vari nelle varie prove. Tale assunzione identifica uno schema di **campionamento senza ripetizione** (o in blocco): una volta estratta una pallina dall'urna, essa non verrà più reinserita nell'urna. Per questo motivo, ciascuna estrazione (o prova) è dipendente dalle precedenti.

3.1 Variabile casuale Ipergeometrica

La variabile casuale Ipergeometrica descrive la probabilità di osservare k successi in n prove dipendenti. Data un'urna contenente F palline favorevoli su N palline totali, la v.c. Ipergeometrica:

$$\mathcal{H} \sim \mathbf{Iper}(N, F, n)$$

ha `pmf` uguale a:

$$\mathbf{P}(\mathcal{H}=k) = \frac{\overbrace{\binom{F}{k}}^{\substack{\#\text{ modi per ottenere } k \\ \text{successi su } F \text{ disponibili}}} \times \overbrace{\binom{N-F}{n-k}}^{\substack{\#\text{ modi per ottenere } n-k \\ \text{insuccessi su } N-F \text{ disponibili}}}}{\underbrace{\binom{N}{n}}_{\substack{\#\text{ modi in cui estrarre } n \text{ palline} \\ \text{da un'urna che ne contiene N}}}}$$

Torniamo alla nostra urna in cui $N = 10$ e $F = 3$. Risponderemo alle stesse domande poste per la v.c. Binomiale nella Sezione **2.2**, supponendo adesso che le prove siano dipendenti.

a) Qual è la probabilità di pescare 3 palline rosse in 5 estrazioni dipendenti?

Qui $n = 5$, $k = 3$.

$$\mathbf{P}(\mathcal{H}=3) = \frac{\binom{3}{3}\binom{10-3}{5-3}}{\binom{10}{5}} = \frac{1 \times 21}{252} = 0.083333$$

b) Qual è la probabilità di pescare almeno una pallina rossa in 6 estrazioni dipendenti?

Qui $n = 6$, $k > 0$. Dunque, è più facile calcolare il complemento ad 1 della probabilità di non estrarre alcuna pallina rossa nelle prima 6 estrazioni:

$$\mathbf{P}(\mathcal{H}>0) = 1 - \mathbf{P}(\mathcal{H}=0) = 1 - \frac{\binom{3}{0}\binom{10-3}{6-0}}{\binom{10}{6}} = \frac{1 \times 7}{210} = 1 - 0.033333 = 0.966667$$

c) Qual è la probabilità di pescare almeno due palline rosse in 6 estrazioni dipendenti?

In questo caso, $n = 6$, $k > 1$. Dunque, è più facile calcolare il complemento ad 1 della probabilità di estrarre al più una pallina rossa nelle prima 6 estrazioni:

$$\mathbf{P}(\mathcal{H} > 1) = 1 - [\mathbf{P}(\mathcal{H} = 0) + \mathbf{P}(\mathcal{H} = 1)] = 1 - \left[0.033333 + \frac{\binom{3}{1}\binom{10-3}{6-1}}{\binom{10}{6}}\right] =$$

$$= 1 - \left[0.033333 + \frac{3 \times 21}{210}\right] = 1 - 0.333333 = 0.666667$$

d) Qual è la probabilità di pescare al più una pallina rossa in 9 estrazioni dipendenti?

In questo caso, $n = 9$, $k = \{0, 1\}$. Dunque, bisogna sommare i valori della `pmf` corrispondenti ai due diversi k.

$$\mathbf{P}(\mathcal{H} \leq 1) = \mathbf{P}(\mathcal{H} = 0) + \mathbf{P}(\mathcal{H} = 1) = \frac{\binom{3}{0}\overbrace{\binom{10-3}{9-0}}^{=0}}{\binom{10}{9}} + \frac{\binom{3}{1}\overbrace{\binom{10-3}{9-1}}^{=0}}{\binom{10}{9}} = 0$$

È logico, poiché in 9 estrazioni dall'urna senza rimessa, devono essere uscite almeno due palline rosse.

3.2 *Facoltativo*: come ricavare una v.c. Ipergeometrica Negativa

La variabile casuale Ipergeometrica Negativa descrive la probabilità di dover aspettare la n-esima prova dipendente per osservare i primi k successi:

$$\mathcal{H}^- \sim \mathbf{IN}(N, F, k)$$

È l'equivalente della v.c. Binomiale Negativa (e della Geometrica se $k = 1$) in caso di estrazioni senza ripetizione. Per questo motivo, la sua `pmf` è determinabile usando la stessa logica: nelle prime $n - 1$ prove dipendenti devono essere osservati $k - 1$ successi, mentre nell'ultima prova bisogna osservare il k-esimo successo:

$$\mathbf{P}(\mathcal{H}^- = n) = \underbrace{\frac{\binom{F}{k-1}\overbrace{\binom{N-F}{(n-1)-(k-1)}}^{\binom{N-F}{n-k}}}{\binom{N}{n-1}}}_{\mathbf{Iper}(N,F,n-1),\, k-1 \text{ successi}} \times \underbrace{\frac{F-(k-1)}{N-(n-1)}}_{\substack{\text{prob. di successo all'}n\text{-esima prova} \\ \text{dopo aver osservato } k-1 \text{ successi}}}$$

Qual è la probabilità di dover arrivare alla decima (e ultima) estrazione per osservare tutte e 3 le palline rosse?

In questo caso, $N = n = 10$, $F = k = 3$.

$$\mathbf{P}(\mathcal{H}^- = 10) = \underbrace{\frac{\binom{3}{3-1}\binom{10-3}{10-3}}{\binom{10}{10-1}}}_{\substack{\text{prob. di osservare 2 rosse} \\ \text{nelle prime 9 estrazioni}}} \times \underbrace{\frac{3-(3-1)}{10-(10-1)}}_{\substack{\text{prob. di osservare 1 rossa} \\ \text{se l'unica pallina rimasta è rossa}}} = \frac{3}{10} \times 1 = 0.3$$