La variabile aleatoria ipergeometrica

Nell'estrarre n biglie senza reinserimento da un'urna che contiene N biglie, di cui m sono bianche e $N-m$ nere, sia X il numero di biglie bianche presenti tra le n estratte. Allora

$$P(X=k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}} \qquad k = 0, 1, \ldots, n.$$

Una variabile aleatoria dotata di tale densità discreta per opportuni valori n, m, N è detta variabile aleatoria *ipergeometrica*.

Ricordando che $\binom{r}{k} > 0$ quando $0 \leq k \leq r$, risulta $\binom{m}{k}\binom{N-m}{n-k} > 0$ per $\begin{cases} 0 \leq k \leq m \\ 0 \leq n-k \leq N-m \end{cases}$ ossia $\begin{cases} 0 \leq k \leq m \\ n+m-N \leq k \leq n \end{cases}$ da cui segue che

$$\begin{aligned} &P(X=k) > 0 \quad \text{se } \max(0, n+m-N) \leq k \leq \min(n,m), \\ &P(X=k) = 0 \quad \text{altrimenti.} \end{aligned}$$

Notiamo che, in virtù della formula di Vandermonde, risulta:

$$\sum_{k=0}^{n} P(X=k) = \sum_{k=0}^{n} \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}} = \frac{1}{\binom{N}{n}} \sum_{k=0}^{n} \binom{m}{k}\binom{N-m}{n-k} = \frac{\binom{N}{n}}{\binom{N}{n}} = 1.$$

Esempio. In un sistema multiutente vi sono 8 utenti collegati, di cui 5 richiedono l'accesso ad Internet ed i rimanenti non lo richiedono. Se si scelgono a caso 4 utenti collegati, qual è la probabilità che al più 3 di essi richiedano l'accesso ad Internet?

Soluzione. Sia X il numero di utenti che richiede l'accesso ad Internet tra i 4 utenti scelti. Poiché X ha distribuzione ipergeometrica di parametri $n=4$, $m=5$, $N=8$ è:

$$p(k) = P(X=k) = \frac{\binom{5}{k}\binom{3}{4-k}}{\binom{8}{4}}.$$

Notiamo che: $p(0)=0$, $p(1)=\frac{1}{14}$, $p(2)=\frac{6}{14}$, $p(3)=\frac{6}{14}$, $p(4)=\frac{1}{14}$. Pertanto, la probabilità richiesta è $P(X \le 3) = 1 - P(X=4) = 1 - \frac{1}{14} = \frac{13}{14}$.

Esempio. Un rivenditore acquista componenti elettriche a lotti di 10. Controlla a caso 3 componenti in ogni lotto e lo accetta solo se nessuno dei 3 pezzi controllati è difettoso. Se il 30% dei lotti ha 4 pezzi difettosi e il 70% ha 1 pezzo difettoso, qual è la percentuale dei lotti che il rivenditore rifiuterà?

Soluzione. Sia X il numero di pezzi difettosi tra i 3 controllati, e sia $B = \{$il lotto ha 4 pezzi difettosi$\}$, cosicché $\overline{B} = \{$il lotto ha 1 pezzo difettoso$\}$; si ha

$$
\begin{aligned}
P(X=0) &= P(X=0|B)\,P(B) + P(X=0|\overline{B})\,P(\overline{B}) \\
&= \frac{\binom{4}{0}\binom{6}{3}}{\binom{10}{3}} \cdot \frac{3}{10} + \frac{\binom{1}{0}\binom{9}{3}}{\binom{10}{3}} \cdot \frac{7}{10} = \frac{6 \cdot 5 \cdot 4}{10 \cdot 9 \cdot 8} \cdot \frac{3}{10} + \frac{9 \cdot 8 \cdot 7}{10 \cdot 9 \cdot 8} \cdot \frac{7}{10} \\
&= 0{,}05 + 0{,}49 = 0{,}54
\end{aligned}
$$

pertanto il rivenditore rifiuterà il 46% dei lotti. Notiamo inoltre che

$$
P(B|X=0) = \frac{P(X=0|B)\,P(B)}{P(X=0)} = \frac{0{,}05}{0{,}54} \approx 0{,}0926.
$$

Se si scelgono a caso n biglie senza reinserimento da un insieme di N biglie, delle quali la frazione $p = m/N$ è bianca, allora il numero di biglie bianche selezionate X ha distribuzione ipergeometrica. È ragionevole supporre che se m e N sono grandi rispetto a n, allora il fatto che si effettui o meno reinserimento ad ogni estrazione possa essere trascurabile. Non tenendo conto delle biglie già estratte, ogni altra estrazione darà una biglia bianca con probabilità approssimativamente pari a p, se m e N sono grandi rispetto a n. In tal caso la legge di X è approssimata da una legge binomiale:

$$P(X = k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}} = \frac{\frac{(m)_k}{k!} \cdot \frac{(N-m)_{n-k}}{(n-k)!}}{\frac{(N)_n}{n!}} = \binom{n}{k} \frac{(m)_k \cdot (N-m)_{n-k}}{(N)_n}$$

Per m grande risulta $(m)_k = m(m-1)\cdots(m-k-1) \approx m^k$ e pertanto

$$P(X = k) \approx \binom{n}{k} \frac{m^k\,(N-m)^{n-k}}{N^n} = \binom{n}{k} \left(\frac{m}{N}\right)^k \left(\frac{N-m}{N}\right)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k}$$

con $p = \dfrac{m}{N}$, e con m e N grandi rispetto a n e k.

Proposizione. Se X è una variabile aleatoria ipergeometrica di parametri n, N e m, allora per $p = \dfrac{m}{N}$ si ha

$$E(X) = n\,p, \qquad \mathrm{Var}(X) = n\,p\,(1-p)\left(1 - \frac{n-1}{N-1}\right).$$

Dimostrazione. Il momento di ordine k di X è dato da:

$$E(X^k) = \sum_{i=0}^{n} i^k\, P(X=i) = \sum_{i=1}^{n} i^k \binom{m}{i}\binom{N-m}{n-i}\Big/\binom{N}{n}.$$

Utilizzando le identità

$$i\binom{m}{i} = \frac{i\,m!}{i!\,(m-i)!} = \frac{m\,(m-1)!}{(i-1)!\,(m-i)!} = m\binom{m-1}{i-1}, \qquad n\binom{N}{n} = N\binom{N-1}{n-1}$$

si ha

$$E(X^k) = \frac{n\,m}{N}\sum_{i=1}^{n} i^{k-1}\binom{m-1}{i-1}\binom{N-m}{n-i}\Big/\binom{N-1}{n-1}.$$

Ponendo $j = i - 1$ in $E(X^k) = \dfrac{n\,m}{N} \displaystyle\sum_{i=1}^{n} i^{k-1} \binom{m-1}{i-1}\binom{N-m}{n-i} \Big/ \binom{N-1}{n-1}$ si ha

$$E(X^k) = \frac{n\,m}{N} \sum_{j=0}^{n-1} (j+1)^{k-1} \binom{m-1}{j}\binom{N-m}{n-1-j} \Big/ \binom{N-1}{n-1} = \frac{n\,m}{N}\, E[(Y+1)^{k-1}]$$

con Y variabile aleatoria ipergeometrica di parametri $n-1$, $N-1$ e $m-1$. Ponendo $k=1$ e $k=2$ si ha rispettivamente:

$$E(X) = n\,\frac{m}{N} = n\,p, \qquad E(X^2) = \frac{n\,m}{N}\, E(Y+1) = \frac{n\,m}{N}\left[\frac{(n-1)\,(m-1)}{N-1} + 1\right].$$

Da ciò, ricordando che $p = m/N$, segue

$$\begin{aligned}
\mathrm{Var}(X) &= \frac{n\,m}{N}\left[\frac{(n-1)\,(m-1)}{N-1} + 1 - \frac{n\,m}{N}\right] = n\,p\left[\frac{(n-1)\,(Np-1)}{N-1} + 1 - n\,p\right] \\
&= n\,p\left[(n-1)\left(p - \frac{1-p}{N-1}\right) + 1 - n\,p\right] \\
&= n\,p\left[(n-1)\,p - (n-1)\,\frac{1-p}{N-1} + 1 - n\,p\right] = n\,p\,(1-p)\left(1 - \frac{n-1}{N-1}\right).
\end{aligned}$$

Poiché risulta

$$\mathrm{Var}(X) = n\,p\,(1-p)\left(1 - \frac{n-1}{N-1}\right),$$

è facile vedere che quando $N \to \infty$ la varianza di una variabile aleatoria ipergeometrica tende alla varianza di una variabile aleatoria binomiale, data da $n\,p\,(1-p)$.

Inoltre, per $n = 1$ la varianza di una variabile aleatoria ipergeometrica coincide con la varianza di una variabile aleatoria binomiale, in quanto entrambe le variabili aleatorie coincidono con una variabile aleatoria di Bernoulli.

La variabile aleatoria uniforme discreta

Nell'estrarre una biglia da un'urna contenente n biglie numerate da 1 a N, denotiamo con X il numero estratto. Allora

$$P(X = k) = \frac{1}{N}, \qquad k = 1, 2, \ldots, N.$$

La variabile aleatoria avente tale densità discreta è detta *uniforme discreta*. Notiamo che la funzione di distribuzione di X è data da: $F(x) = 0$ per $x < 1$, $F(x) = k/N$ per $k \leq x < k + 1$ $(k = 1, 2, \ldots, n - 1)$, $F(x) = 1$ per $x \geq N$.

Proposizione. Se X è una variabile aleatoria uniforme discreta di parametro N, allora

$$E(X) = \frac{N+1}{2}, \qquad \mathrm{Var}(X) = \frac{N^2 - 1}{12}.$$

Dimostrazione. Ricordando che $\displaystyle\sum_{k=1}^{N} k = \frac{N(N+1)}{2}$, si ha

$$E(X) = \sum_{k=1}^{N} k\, P(X = k) = \frac{1}{N} \sum_{k=1}^{N} k = \frac{N+1}{2}.$$

Analogamente, poiché $\displaystyle\sum_{k=1}^{N} k^2 = \frac{N(N+1)(2N+1)}{6}$, si ha

$$E(X^2) = \sum_{k=1}^{N} k^2\, P(X=k) = \frac{1}{N}\sum_{k=1}^{N} k^2 = \frac{(N+1)(2N+1)}{6}.$$

Segue pertanto

$$\begin{aligned}\mathrm{Var}(X) &= \frac{(N+1)(2N+1)}{6} - \frac{(N+1)^2}{4}\\ &= \frac{N+1}{12}\,[2(2N+1)-3(N+1)] = \frac{N+1}{12}\,(N-1)\\ &= \frac{N^2-1}{12}.\end{aligned}$$

Esercizi per casa

4.a) Sia X una variabile aleatoria discreta che assume valori $0, 1, 2, 3$ e tale che

$$p(x) = \frac{1}{2}\,p(x-1) \qquad \text{per } x = 1, 2, 3.$$

(i) Determinare la funzione di probabilità $p(x) = P(X = x)$, per $x = 0, 1, 2, 3$.
(ii) Ricavare $E(X)$.

4.b) Un venditore di automobili ha fissato due appuntamenti. La probabilità di vendere un'automobile nell'i-esimo appuntamento è $p_i = (\frac{1}{2})^i$ $(i = 1, 2)$. Ogni vendita ha la stessa probabilità di riguardare la versione base (del valore di 9000 euro) oppure la versione lusso (del valore di 12000 euro). Indicata con X la variabile aleatoria che descrive il totale dei guadagni del venditore, si determini:
(i) la densità discreta e la funzione di distribuzione di X,
(ii) il valore atteso di X.

4.c) Vi sono due dadi, di cui uno è truccato, nel senso che le facce del dado sono $\{1, 1, 2, 3, 4, 5\}$. Se si sceglie a caso uno dei due dadi e lo si lancia 5 volte, sia X il numero di volte che esce 1. Valutare
(i) $P(X \geq 2)$,
(ii) $E(X)$,
(iii) $Var(X)$.

4.d) Giorgio dispone di due monete truccate: la prima fornisce testa con probabilità $\frac{2}{3}$, la seconda con probabilità $\frac{1}{3}$. Fabrizio sceglie a caso una delle due monete e la lancia finché non esce testa.
(i) Qual è la probabilità che la moneta mostri testa per la prima volta al quarto lancio?
(ii) Qual è la probabilità che siano necessari almeno cinque lanci perché la moneta mostri testa per la prima volta?
(iii) Se Fabrizio sceglie la seconda moneta, qual è il numero medio di lanci che deve effettuare affinché la moneta mostri testa per la prima volta?

4.e) Un esperimento consiste nell'estrarre ripetutamente biglie da un'urna che contiene 4 biglie bianche e 1 nera. Sia Y l'estrazione in cui si estrae la biglia nera per la prima volta. Determinare $P(Y \leq 3)$, $E(Y)$ e $Var(Y)$ nel caso di estrazioni
(i) con reinserimento,
(ii) senza reinserimento.