9.4 Variabile casuale ipergeometrica

Si supponga di avere un insieme di N unità delle quali F sono classificate come "favorevoli" e $N - F$ come "contrarie", in modo tale che scegliendone una a caso il risultato dell'esperimento possa configurarsi come un "successo" – quando si estrae un'unità favorevole – oppure come un "insuccesso" – se si estrae un'unità contraria. Dall'insieme delle N unità ne vengono estratte n. Se l'estrazione avviene con rimessa la variabile che descrive il numero di successi ha una distribuzione binomiale con probabilità di successo $\pi = F / N$. Quando invece l'estrazione avviene in blocco (senza rimessa) la variabile casuale che descrive il numero di unità favorevoli nelle n estrazioni è ipergeometrica.

La sua distribuzione dipende da tre parametri: il numero complessivo di unità N, il numero di unità favorevoli F e il numero di estrazioni n. Di conseguenza per indicare che X è una variabile casuale ipergeometrica si utilizza la notazione

$$X \sim IP(N, F, n).$$

La variabile casuale ipergeometrica assume valori interi da zero al minimo fra n e F. La probabilità che X assuma valore x si calcola come rapporto fra casi favorevoli e casi possibili. I casi possibili sono dati dal numero delle combinazioni di n elementi scelti fra N

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}.$$

Il numero di modi nei quali è possibile scegliere x elementi fra le F unità favorevoli è

$$\binom{F}{x} = \frac{F!}{x!(F-x)!}.$$

Se fra le n unità estratte ve ne sono x favorevoli, necessariamente le restanti $n-x$ sono contrarie. Il numero di modi nei quali è possibile scegliere $n-x$ elementi fra le $N-F$ unità contrarie è

$$\binom{N-F}{n-x} = \frac{(N-F)!}{(n-x)!(N-F-n+x)!}.$$

I casi favorevoli, ossia il numero di sequenze nelle quali si possono presentare esattamente x unità favorevoli e $n-x$ unità contrarie, risultano

$$\binom{F}{x}\binom{N-F}{n-x} = \frac{F!}{x!(F-x)!}\frac{(N-F)!}{(n-x)!(N-F-n+x)!}.$$

Di conseguenza si ha

$$P(X=x) = \frac{\binom{F}{x}\binom{N-F}{n-x}}{\binom{N}{n}} = \frac{\dfrac{F!}{x!(F-x)!}\dfrac{(N-F)!}{(n-x)!(N-F-n+x)!}}{\dfrac{N!}{n!(N-n)!}},$$

per $x = 0,1,\dots,min\{n,F\}$.

Esempio 9.3. Un'azienda riceve da un fornitore un pacco con 20 pezzi. L'azienda decide di provarne cinque, estraendoli in blocco dal pacco; se nessun pezzo è difettoso la fornitura è accettata altrimenti il pacco è respinto. Si vuole calcolare la probabilità di accettare una fornitura con 6 pezzi difettosi.

La variabile casuale X che descrive il numero di pezzi difettosi ha una distribuzione ipergeometrica. Nella fornitura vi sono $N = 20$ pezzi dei quali $F = 6$ difettosi; se si provano $n = 5$ pezzi si ha $X \sim IP(20,6,5)$. Si vuole la probabilità di $X = 0$; essa risulta

$$P(X=0) = \frac{\dfrac{6!}{0!6!}\dfrac{(20-6)!}{(5-0)!(20-6-5+0)!}}{\dfrac{20!}{5!(20-5)!}} = 0.129,$$

La probabilità che l'azienda accetti un pacco con 6 pezzi difettosi è 0.129.

E' possibile dimostrare che il valore atteso della variabile casuale ipergeometrica è dato da

$$E[X] = n\frac{F}{N},$$

mentre la varianza è

$$Var(X) = n\frac{F}{N}\left(1-\frac{F}{N}\right)\frac{N-n}{N-1}.$$

Se si indica con $\pi = F/N$ la probabilità di estrarre un'unità favorevole in una singola prova, si osserva che il valore atteso della variabile casuale ipergeometrica coincide con quello della variabile casuale binomiale. La circostanza che l'estrazione avvenga senza rimessa non modifica il valore atteso del numero di successi. Infatti l'estrazione in blocco induce dipendenza fra le prove, ma ciò è irrilevante ai fini del calcolo del valore atteso. La varianza invece risulta più piccola poiché la quantità $n\pi(1-\pi)$, corrispondente alla varianza di una variabile casuale binomiale, risulta moltiplicata per un fattore $(N-n)/(N-1)$ minore di uno. Intuitivamente la minore variabilità della variabile casuale ipergeometrica può essere attribuita a una maggiore *efficienza* dell'estrazione in blocco. Se l'estrazione avviene con rimessa la stessa unità può essere estratta più volte. Questa replicazione è invece impossibile quando l'estrazione è in blocco. Nell'esperimento descritto dalla variabile casuale ipergeometrica sono escluse tutte le sequenze nelle quali compare più volte la stessa unità e di conseguenza questa variabile casuale presenta una minore variabilità rispetto alla variabile casuale binomiale.

Tuttavia quando N aumenta, mantenendo costante il rapporto $\pi = F / N$, il rapporto $(N-n)/(N-1)$ tende a 1 e pertanto la varianza della variabile casuale ipergeometrica converge a $n\pi(1-\pi)$. Al crescere di N la probabilità di estrarre più volte la stessa unità, nell'estrazione con rimessa, diventa trascurabile. Di conseguenza, per N sufficientemente grande, la distribuzione di probabilità della variabile casuale ipergeometrica può essere approssimata dalla distribuzione di probabilità della variabile casuale binomiale. La figura 9.2 illustra l'approssimazione della distribuzione di una variabile casuale ipergeometrica con $N = 50, 100, 200$, $F = N/5$ e $n = 10$ mediante la distribuzione di probabilità di una variabile casuale binomiale con parametri $n = 10$ e $\pi = F / N = 0.20$.