APPROFONDIMENTO

La distribuzione ipergeometrica

Supponiamo di avere un insieme di N oggetti di due tipi diversi che chiameremo popolazione e supponiamo inoltre che ce ne siano K del primo tipo e $N-K$ del secondo.
Ad esempio, possiamo considerare un'urna contenente 100 palline di cui 60 bianche e 40 nere; oppure una produzione aziendale di 1000 pezzi di cui si sa che 12 sono difettosi; oppure una popolazione di 2000 individui di cui 600 presentano forme di allergia.
Supponiamo che nella popolazione venga scelto a caso un campione di n oggetti: ad esempio, supponiamo che vengano estratte 20 palline dall'urna, che vengano scelti 50 pezzi a caso dall'intera produzione, che vengano selezionate a caso 100 persone fra le 2000.

Consideriamo la variabile aleatoria X che indichi il numero di oggetti fra quelli del campione che siano del primo tipo (oppure del secondo): ad esempio, X conta il numero di palline bianche estratte dall'urna (oppure il numero di quelle nere), il numero di pezzi difettosi trovati (oppure il numero di quelli che non hanno difetti), il numero di persone che risultano allergiche (o che non lo sono).
La prima cosa che dobbiamo determinare è quali valori può assumere X. Osserviamo allora che, se la variabile aleatoria conta il numero di oggetti del primo tipo presenti nel campione, allora non può superare il più piccolo fra n (ampiezza del campione scelto) e K (numero di oggetti del primo tipo); dunque

$$x \in \{0, 1, 2, ..., \min(n, K)\}$$

Di conseguenza il numero di oggetti del campione che sono del secondo tipo, e che sono in numero di $n-x$, non può superare il più piccolo fra n e $N-K$, cioè $n-x \leq N-K$.
Se risolviamo quest'ultima relazione rispetto a x otteniamo $x \geq K-N+n$.
In definitiva allora, la variabile X può assumere valori interi tali che

$$\max(0, K-N+n) \leq x \leq \min(n, K)$$

Indicheremo con S tale insieme.

Ad esempio, se la popolazione è di 2000 individui, di cui 1200 del primo tipo e 800 del secondo, e se si sceglie a caso un campione di 50 individui, la variabile X che conta gli individui del primo tipo presenti nel campione può assumere i valori 0,1,2,....,50.
Se invece il campione scelto è di 950 individui sicuramente almeno 150 saranno del primo tipo ed al massimo quelli del secondo potranno essere 800; ciò è confermato dalla relazione che abbiamo trovato: essendo

$\max(0,K-N+n) = \max(0,1200-2000+950) = \max(0,150) = 150$

$\min(950,1200) = 950$

possiamo dire che la variabile X può assumere valori interi compresi fra 150 e 950:

$S = \{x \text{ interi} \mid 150 \leq x \leq 950\}$.

Determiniamo adesso la funzione distribuzione di probabilità per X; x è il numero di oggetti del primo tipo presenti nel campione, allora il numero di modi di scegliere x oggetti del primo tipo sui K totali è $\binom{K}{x}$; il numero di modi di scegliere i rimanenti $n-x$ oggetti fra quelli del secondo tipo sugli $N-K$ totali è $\binom{N-K}{n-x}$. Osserviamo poi che la scelta del campione può essere fatta in $\binom{N}{n}$ modi diversi, ciascuno ovviamente con la stessa probabilità.

© ISTITUTO ITALIANO EDIZIONI ATLAS

Allora la funzione distribuzione di probabilità di X ha la seguente espressione

$$\boxed{\boldsymbol{f(x) = \begin{cases} \dfrac{\dbinom{K}{x} \cdot \dbinom{N-K}{n-x}}{\dbinom{N}{n}} & x \in S \\ 0 & \textbf{altrove} \end{cases}}}$$

Una distribuzione di probabilità di questo tipo si dice **ipergeometrica**.

Si dimostra che il valore atteso e la varianza di una distribuzione ipergeometrica sono

$$\boxed{\boldsymbol{E(X) = \frac{nK}{N} \qquad V(X) = \frac{nK(N-K)(N-n)}{N^2(N-1)}}}$$

Esempio.

Una grande azienda ha 150 impiegati, 12 dei quali hanno un alto tasso di assenteismo. All'avvicinarsi delle feste natalizie, l'azienda decide di assegnare 40 premi ad altrettanti impiegati estratti a sorte. Calcoliamo la probabilità che i premi vengono assegnati

a. a nessun impiegato assenteista;

b. a 3 impiegati assenteisti;

c. a meno di 5 assenteisti.

Se X è la variabile aleatoria che indica il numero di impiegati assenteisti che ricevono un premio, la funzione distribuzione di probabilità ha equazione

$$f(x) = \begin{cases} \dfrac{\dbinom{12}{x} \cdot \dbinom{138}{40-x}}{\dbinom{150}{40}} & x = 0,1,2,\dots\dots\dots,12 \\ 0 & \text{altrove} \end{cases}$$

Si ha che: **a.** $f(0) = \dfrac{\dbinom{12}{0} \cdot \dbinom{138}{40}}{\dbinom{150}{40}} = 0{,}020$ **b.** $f(3) = \dfrac{\dbinom{12}{3} \cdot \dbinom{138}{37}}{\dbinom{150}{40}} = 0{,}266$

c. La probabilità che meno di 5 assenteisti vincano dei premi è data da $f(0) + f(1) + f(2) + f(3) + f(4)$

ed essendo

$$f(1) = \frac{\dbinom{12}{1} \cdot \dbinom{138}{39}}{\dbinom{150}{40}} = 0{,}099 \qquad f(2) = \frac{\dbinom{12}{2} \cdot \dbinom{138}{38}}{\dbinom{150}{40}} = 0{,}212 \qquad f(4) = \frac{\dbinom{12}{4} \cdot \dbinom{138}{36}}{\dbinom{150}{40}} = 0{,}217$$

si ottiene che $p(X < 5) = 0{,}020 + 0{,}099 + 0{,}212 + 0{,}266 + 0{,}217 = 0{,}814$

© ISTITUTO ITALIANO EDIZIONI ATLAS

Esercizi

1 In un vivaio ci sono 30 trote Fario e 70 trote salmonate. Se si pescano 20 trote, qual è la probabilità di trovare 8 Fario? [0,1162]

2 Una commissione d'esame in un concorso pubblico è formata da 10 persone delle quali 4 sono uomini. Qual è la probabilità che, estraendo una sottocommissione formata da 5 persone, non ci sia nemmeno un uomo? $\left[\frac{1}{42}\right]$

3 In un controllo di qualità si estrae un campione di 15 pezzi da un lotto di 100 che si sa che contiene 5 pezzi difettosi. Calcola la probabilità che il lotto venga accettato se la condizione di accettabilità è aver trovato al massimo un pezzo difettoso nel campione.
(Suggerimento: si tratta di una distribuzione ipergeometrica in cui $N = 100$, $K = 5$, $n = 15$; la probabilità di accettare il lotto è quindi data da $p(X = 0) + p(X = 1)$) [$\approx 0{,}84$]

4 In una popolazione di 200 cani di un canile municipale ve ne sono 30 che sono malati. Se si sceglie a caso un campione di 50 cani calcola la probabilità che:

a. non vi siano cani malati; [$0{,}786 \cdot 10^{-6}$]

b. ci siano 10 cani malati; [0,091]

c. ci siano meno di 10 cani malati. [0,821]

5 In un liceo scientifico ci sono 250 iscrizioni in prima classe e di queste 130 sono ai corsi sperimentali. Scelto un campione di 62 studenti, calcola la probabilità:

a. che ci siano 20 studenti iscritti ai corsi ordinari; [0,0019]

b. che ci siano meno di 40 studenti dei corsi sperimentali; [0,9837]

c. che ci siano più di 30 studenti dei corsi sperimentali. [0,6951]

6 In un paese dove vivono 1580 nuclei familiari, ve ne sono 875 che sono proprietari della casa in cui abitano e gli altri sono affittuari. Si estrae da tale popolazione un campione di 80 famiglie e si vuole calcolare la probabilità che:

a. ci siano meno di 40 proprietari; [0,1339]

b. più di 30 affittuari; [0,8852]

c. più di 50 proprietari. [0,0755]

© Istituto Italiano Edizioni Atlas