Distribuzione ipergeometrica.

Prologo

Sia $\Omega = C_{b+r,n}$ con $b, r, n \in \mathbb{N}$ $n \leq b+r$

(l'insieme delle combinazioni di n elementi in $b+r$ elementi)

Per semplicità si pensi che r indichi il numero delle palline rosse e b quello delle palline bianche contenute in un'urna. Si supponga inoltre sempre per semplicità che le palline rosse siano numerate da 1 a r, mentre le bianche sono numerate da $r+1$ a $b+r$.

Gli eventi elementari $\omega \in \Omega$ descrivono quindi una estrazione di n elementi fra $b+r$; pertanto in assenza di ulteriori indicazioni la probabilità del singolo evento elementare è data dalla distribuzione uniforme, cioè $\mathbb{P}(\{\omega\}) = \frac{1}{\#\Omega} = \frac{1}{\# C_{b+r,n}}$ $= \frac{1}{\binom{b+r}{n}}$. Cioè con la distrib. uniforme non è importante considerare le palline di due colori diversi ed inoltre non è neppure importante l'ordine con cui vengono estratte.

In questo caso lo spazio di probabilità costruito descrive l'estrazione di n numeri (o palline numerate) senza reimbussolamento e senza tener conto dell'ordine. Infatti se $n=5$ e $\omega = \{1, 2, 3\}$, allora la probabilità che nella prima estrazione esca 1 è $\frac{1}{5}$, che nella seconda esca 2 è $\frac{1}{4}$ e che nella 3^a esca 3 è $\frac{1}{3}$. Tuttavia, se non ci curiamo dell'ordine allora dovremo moltiplicare $\frac{1}{3} \cdot \frac{1}{4} \cdot \frac{1}{5}$ per il numero di permutazioni dei tre numeri cioè $3!$

Quindi $\frac{3!}{3 \cdot 4 \cdot 5} = \frac{1}{10} = \frac{1}{\binom{5}{3}} = \frac{1}{\frac{5!}{3!(5-3)!}} = \frac{3!\,2!}{5!} = \frac{2!}{4 \cdot 5}$

$= \frac{1}{10}$

Consideriamo ora degli eventi più complessi, ovvero per $k \leq r$

$$A_k = \{ \omega \in C_{b+r,n} : \text{ in } \omega \text{ vi sono esattamente } k \text{ elementi con indice minore o uguale di } r \}$$

A_k indica cioè l'estrazione, senza rimpiazzo di n palline di cui esattamente k sono di colore rosso.

Allora $$\underline{P}(A_k) = P\left(\bigcup_{\omega \in A_k} \{\omega\}\right) = \sum_{\omega \in A_k} \underline{P}(\{\omega\}) = \frac{1}{\# C_{b+r,n}} \cdot \# A_k$$

$$= \frac{1}{\binom{b+r}{n}} \sum_{\omega \in A_k} 1 .$$

D'altra parte se $\omega \in A_k$ vuol dire che in ω abbiamo esattamente k numeri minori o uguali di r (cioè k palline rosse). Però queste palline rosse sono state selezionate in un insieme di r palline complessive, quindi avremo $\binom{r}{k}$ possibilità. In questo modo sono stati fissati k degli n elementi di ω; ne rimangono allora $n-k$ da considerare fra le palline bianche, per cui ci sono esattamente $\binom{b}{n-k}$ opzioni. Quindi $\binom{r}{k} \cdot \binom{b}{n-k}$ è la cardinalità di A_k, vale a dire $\# A_k = \binom{r}{k}\binom{b}{n-k}$.

Pertanto

$$\underline{P}(A_k) = \frac{1}{\binom{b+r}{n}} \cdot \binom{r}{k}\binom{b}{n-k} .$$

Se in $\Omega = C_{b+r,\,n}$ con distribuzione uniforme, consideriamo la v.a. $X : \Omega \longrightarrow \mathbb{R}$ t.c. $X(\omega)$ = numero delle palline rosse fra n estratte (senza rimpiazzo), allora $X(\Omega) = \{0, 1, 2, 3, \dots, r\}$ e

$$P(X = k) = P(A_k) = \frac{\binom{r}{k}\binom{b}{n-k}}{\binom{b+r}{n}}, \qquad k = 0, 1, 2, 3, \dots, r$$

Provare che $\{\eta : \eta = A_k,\ k = 0, 1, 2, \dots, r\}$ $(r \leq n,\ b+r \leq n)$ con $P(\eta) = \dfrac{\binom{r}{k}\binom{b}{n-k}}{\binom{b+r}{n}}$

è uno spazio di probabilità (con distribuzione di probabilità non uniforme). Questa distribuzione di probabilità è chiamata ipergeometrica.

Da un'urna contenente 10 palline bianche e 7 palline rosse vengono estratte 2 palline senza rimpiazzo.
Indichiamo con $i = 1, 2$ le variabili aleatorie

$$X_i := \begin{cases} 1 & \text{se la } i\text{-esima pallina estratta è rossa} \\ 0 & \text{altrimenti} \end{cases}$$

Calcolare $P(X_1 = 1)$, $P(X_2 = 1)$ e $P(X_1 = 1, X_2 = 1)$
e verificare che $P(X_1 = 1)P(X_2 = 1) \neq P(X_1 = 1, X_2 = 1)$

Calcolare la probabilità di indovinare l'estrazione di 5, 7, 10, 11 su 7 numeri (senza rimpiazzo) tra 90 numeri (da 1 a 90)

Soluzioni alla pag. successiva

$\underline{P}(X_1=1)=\underline{P}(X_1=1, X_2\in\{0,1\}) = \underline{P}(\{X_1=1\}\cap\{X_2\in\{0,1\}\})$

$=\underline{P}(\{X_1=1\}\cap\Omega) = \underline{P}(\{X_1=1\}) = \frac{n}{b+n} = \frac{7}{17}$

$\underline{P}(X_2=1)=\underline{P}(X_1\in\{0,1\}, X_2=1) = \underline{P}(\Omega\cap\{X_2=1\}) = \underline{P}(X_2=1) = \underline{P}(X_1$

mentre $\underline{P}(X_1=1, X_2=1) = \dfrac{\binom{7}{2}\binom{10}{0}}{\binom{17}{2}} = \frac{7}{17}\cdot\frac{6}{16} \neq \left(\frac{7}{17}\right)^2$

Per comprendere meglio il fatto che $\underline{P}(X_2=1)=\underline{P}(X_1=1)$ si consideri il seguente caso generale in cui $\Omega = D_K^{b+r}$, quindi $\omega=(\omega_1,..,\omega_K)$, con ω_i che può essere 1 o 0 (bianca o rossa). Si osservi che si tiene in considerazione qual è l'ordine. Ora se $A_i=\{\omega_i=1\}$ cioè la i-esima estrazione è 1 cioè la i-esima pallina estratta è bianca, abbiamo che $\underline{P}(A_1) = \frac{b}{b+r}$.

D'altra parte, sapendo che la funzione $\phi_{1,K}: D_K^{b+n} \to D_K^{b+n}$ è iniettiva e suriettiva.

$\phi_{1,K}(\omega_1,...,\omega_K) = (\omega_K,\omega_2,..,\omega_{K-1},\omega_1)$

sono stati scambiati

perché $\underbrace{\phi_{1,K}\circ\phi_{1,K}}_{\text{composto}} = id_{D_K^{b+r}}$, allora $\#A_1=\#A_K$

(lo stesso discorso si applica per A_J e A_m con $J\neq m$ usando $\phi_{J,m}$)

Quindi $\underline{P}(A_1)=\underline{P}(A_K) = \frac{b}{b+r}$, come vale $P(A_J)=\underline{P}(A_m$

Per quanto riguarda la probabilità di indovinare 4 numeri su 7 estrazioni possiamo utilizzare il modello ipergeometrico. Infatti, per semplicità consideriamo i 4 numeri selezionati $\{5,7,10,11\}$ come l'insieme delle palline rosse, mentre i rimanenti $\{k \in \mathbb{N} : 1 \le k \le 90\} \setminus \{5,7,10,11\}$ le palline bianche.

In tal caso, $n=7$, $r=4$, $b=86$, per cui

$$\mathbb{P}(A_4) = \frac{\binom{4}{4}\binom{86}{3}}{\binom{90}{7}} = \frac{86!}{3!\cancel{83!}} \cdot \frac{7!\cancel{83!}}{90!} = \frac{4 \cdot 5 \cdot 6 \cdot 7}{87 \cdot 88 \cdot 89 \cdot 90}$$

Esercizio

Un'urna contiene 4 palline verdi e 6 palline gialle. Da essa se ne estraggono 4 senza reimpiazzo.

(a) Calcolare la probabilità che tra le palline estratte ce ne siano due verdi e due gialle

(b) Calcolare la probabilità che le prime due estratte siano verdi e le seconde due siano gialle

(c) Calcolare la probabilità che le prime due estratte siano gialle e le seconde due gialle

Esercizio

Un'urna contiene $2n$ palline di cui n rosse e n gialle. Da essa vengono estratte n palline. Calcolare la probabilità che tra le palline estratte ve ne siano esattamente k rosse.

Si provi inoltre come conseguenza del precedente risultato che

$$\sum_{k=0}^{n} \binom{n}{k}^2 = \binom{2n}{n}$$