Estrazioni da un'urna

Consideriamo un'urna contenente N palline di cui K sono arancioni e le altre $N-K$ di altro colore; studiamo un paio di problemi relativi ad estrazioni successive di palline dall'urna, considerando le due diverse procedure di estrazione:

- **con ripetizione:** dopo aver estratto ed osservato ogni pallina, rimetto la pallina nell'urna prima di procedere all'estrazione successiva.

  Le varie estrazioni sono quindi condotte sempre nelle stesse condizioni (la composizione dell'urna non varia), quindi *i risultati relativi ad estrazioni diverse sono indipendenti tra loro*. Si parla quindi di *prove ripetute* o anche di *campionamento bernoulliano.*

- **senza ripetizione:** dopo aver estratto ed osservato ogni pallina, *non* rimetto la pallina nell'urna prima di procedere all'estrazione successiva.

  Qui ogni estrazione modifica la composizione dell'urna e quindi incide sulle probabilità degli eventi relativi alle estrazioni successive.

Parleremo alternativamente di "singola estrazione" o "prova"; in ogni prova indichiamo come "successo" l'estrazione di una pallina arancione e come "insuccesso" l'estrazione di una pallina di un altro colore.

Problema 1

Se estraggo dall'urna n palline, con che probabilità k di queste saranno arancioni?

Cerchiamo quindi la probabilità dell'evento

$$(E_n = k) = \{k \text{ successi in } n \text{ prove}\}$$

che ha senso solo se $k \in \{0, 1, \cdots, n\}$.

Banalmente, se $n = 1$ e $k = 1$, la probabilità cercata è

$$P(E_1 = 1) = p = \frac{K}{N}.$$

Se si effettuano più prove, bisogna distinguere due casi, a seconda delle diverse procedure di estrazione.

(a) *Estrazioni con ripetizione:*

Qui gli eventi elementari sono tutte le possibili n-uple in cui in ogni posizione c'è una delle N palline. Poiché si considera possibile l'estrazione della stessa pallina più volte, per preservare l'equiprobabilità dei risultati conviene distinguere i diversi ordinamenti con cui possono presentarsi i risultati delle prove.

Il numero dei casi possibili è dato quindi dalle disposizioni con ripetizione di N elementi di classe n

$$\text{numero casi possibili} = D^R_{N,n} = N^n.$$

Per quanto riguarda il conteggio dei casi favorevoli all'evento $(E_n = k)$, la cosa è più complicata.

Cominciamo quindi a calcolare la probabilità dell'evento “prima k successi e poi $n-k$ insuccessi”, che è contenuto in $(E_n = k)$.

Poiché in ogni estrazione abbiamo S in K modi possibili (le palline arancioni sono K) e F in $N-K$,

$$\text{numero casi favorevoli a } (\underbrace{S, \cdots S}_{k \text{ volte}}, \underbrace{F, \cdots F}_{n-k \text{ volte}}) = K^k (N-K)^{n-k}$$

e quindi

$$P(\underbrace{S, \cdots S}_{k \text{ volte}}, \underbrace{F, \cdots F}_{n-k \text{ volte}}) = \frac{K^k (N-K)^{n-k}}{N^n} = p^k (1-p)^{n-k}. \tag{1}$$

Questi però non esauriscono l'evento $(E_n = k)$, perché bisogna considerare tutte le n-uple con k successi e $n-k$ insuccessi, indipendentemente dall'ordine in cui si presentano. Ogni n-upla di questo tipo ha probabilità data da (1).

Per ottenere la probabilità $P(E_n = k)$ bisogna quindi moltiplicare (1) per il numero di possibili ordinamenti degli n simboli, k dei quali sono delle S e $n-k$ sono delle F, cioè per $\frac{n!}{k!(n-k)!} = \binom{n}{k}$.

Abbiamo finalmente che

$$P(E_n = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

In effetti la probabilità (1) può essere calcolata molto più rapidamente sfruttando il fatto che le prove sono indipendenti: poiché in ogni prova si ha $P(S) = p$ e $P(F) = 1-p$ e le prove sono indipendenti

$$P(\underbrace{S, \cdots S}_{k \text{ volte}}, \underbrace{F, \cdots F}_{n-k \text{ volte}}) = \underbrace{P(S) \cdots P(S)}_{k \text{ volte}} \underbrace{P(F) \cdots P(F)}_{n-k \text{ volte}} = p^k (1-p)^{n-k}.$$

(b) *Estrazioni senza ripetizione:*

Poiché le palline estratte non vengono rimpiazzate, onde evitare che le palline finiscano bisogna considerare dei vincoli aggiuntivi su k ed N, in particolare:

$n \in \{0, 1, \cdots, N\}$ (non posso estrarre più palline di quelle presenti nell'urna);

$k \in \{0, 1, \cdots, K\}$ e $n-k \in \{0, 1, \cdots, N-K\}$ (non posso estrarre più palline arancioni di quelle presenti nell'urna e lo stesso per quelle non arancioni).

Qui gli eventi elementari sono tutte le possibili n-uple in cui in ogni posizione c'è una delle N palline, ma poiché non sono possibili ripetizioni, non è necessario distinguere gli ordinamenti dei risultati, per cui il numero dei casi possibili è dato dalle combinazioni senza ripetizione di N elementi di classe n

$$\text{numero casi possibili} = C_{N,n} = \binom{N}{n}.$$

Per quanto riguarda i casi favorevoli, basta richiedere che k palline siano scelte tra le K bianche e $n-k$ tra le $N-K$ non bianche, per cui

$$P(E_n = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}.$$

Problema 2

Eseguo prove una dopo l'altra e mi fermo appena si verifica il k-esimo succeso. Con che probabilità effettuo esattamente n prove?

Cerchiamo quindi la probabilità dell'evento

$$(T_k = n) = \{\text{il } k\text{-esimo successo all'}n\text{-ma prova}\}$$

che ha senso solo se $n \in \{k, k+1, \cdots\}$.

Si noti che anche qui, come nel problema precedente, abbiamo k successi in n prove, ma qui il problema è molto diverso perché cambia la struttura dell'esperimento: mentre prima era certo il numero di prove da effettuare (n) e l'incertezza era sul numero di successi (k poteva assumere diversi valori), adesso sappiamo quanti successi si verificano (k) e l'incertezza è sul numero di prove da effettuare.

L'evento $(T_k = n)$ si verifica se e solo se ho un successo all'n-ma prova $k-1$ successi nelle prime $n-1$ prove quindi

$$(T_n = k) = \{k-1 \text{ successi nelle prime } n-1 \text{ prove}\} \cap \{\text{successo all'ultima prova}\}$$

(a) *Estrazioni con ripetizione:*

In questo caso gli eventi relativi a prove diverse sono indipendenti tra loro, quindi

$$\begin{aligned} P(T_n = k) &= P(\{k-1 \text{ successi nelle prime } n-1 \text{ prove}\} \cap \{\text{successo all'ultima prova}\}) \\ &= P(\{k-1 \text{ successi nelle prime } n-1 \text{ prove}\})P(\{\text{successo all'ultima prova}\}). \end{aligned}$$

Dal problema 1(a) sappiamo che

$$P(\{k-1 \text{ successi nelle prime } n-1 \text{ prove}\}) = \binom{n-1}{k-1} p^{k-1}(1-p)^{n-k}$$

e la probabilità di successo in una prova è p, quindi

$$\begin{aligned} P(T_n = k) &= P(\{k-1 \text{ successi nelle prime } n-1 \text{ prove}\})P(\{\text{successo all'ultima prova}\}) \\ &= \binom{n-1}{k-1} p^{k-1}(1-p)^{n-k} p = \binom{n-1}{k-1} p^{k}(1-p)^{n-k}. \end{aligned}$$

(b) *Estrazioni senza ripetizione:*

Anche qui, se non vogliamo finire le palline troppo presto, bisogna introdurre i vincoli aggiuntivi su n e k già visti nel problema 1(b).

Non essendo possibile considerare l'indipendenza tra le prime $n-1$ prove e l'ultima, bisogna procedere utilizzando la probabilità condizionata

$$\begin{aligned} P(T_n = k) &= P(\{k-1 \text{ successi nelle prime } n-1 \text{ prove}\} \cap \{\text{successo all'ultima prova}\}) \\ &= P(\{k-1 \text{ successi nelle prime } n-1 \text{ prove}\}) \\ &\quad P(\{\text{successo all'ultima prova}\}|\{k-1 \text{ successi nelle prime } n-1 \text{ prove}\}). \end{aligned}$$

Dal problema 1(b) sappiamo che

$$P(\{k-1 \text{ successi nelle prime } n-1 \text{ prove}\}) = \frac{\binom{K}{k-1}\binom{N-K}{n-k}}{\binom{N}{n-1}}$$

inoltre, se sappiamo che sono uscite $k-1$ palline arancioni nelle prime $n-1$ prove, per l'n-ma prova l'urna conterrà $N-(n-1)$ palline di cui $K-(k-1)$ arancioni, per cui

$$P(\{\text{successo all'ultima prova}\}|\{k-1 \text{ successi nelle prime } n-1 \text{ prove}\}) = \frac{K-k+1}{N-n+1}$$

e quindi

$$P(T_n = k) = \frac{\binom{K}{k-1}\binom{N-K}{n-k}}{\binom{N}{n-1}} \frac{K-k+1}{N-n+1}.$$

Problema 3

Supponiamo adesso che nell'urna ci siano K_A palline arancioni, K_B bianche e le altre $K_C = N - K_A - K_B$ di altro colore. Con che probabilità in n prove estraggo k_A palline arancioni e k_B bianche?

Il problema ha senso se $k_A + k_B \leq n$. Vista l'analogia col problema 1, procederemo in modo molto simile.

(a) *Estrazioni con ripetizione:*

(Guarda il problema 1(a)) Innanzi tutto indichiamo con

$$p_A = \frac{K_A}{N} \qquad p_B = \frac{K_B}{N} \qquad p_C = \frac{K_C}{N}$$

che rappresentano la probabilità di estrarre, un una singola prova, pallina arancione o bianca o altro rispettivamente.

Anche qui facciamo un passo intermedio trovando una probabilità analoga alla (1): cerchiamo quindi prima la probabilità di osservare palline arancioni nelle prime k_A prove, poi palline bianche nelle successive k_B e poi altro colore nelle restanti $k_C = n - k_A - k_B$. Ricordando che le prove sono tutte indipendenti tra loro, abbiamo

$$\begin{aligned} P(\underbrace{A,\cdots A}_{k_A \text{ volte}},\underbrace{B,\cdots B}_{k_B \text{ volte}}\underbrace{C,\cdots C}_{k_C \text{ volte}}) &= \underbrace{P(A)\cdots P(A)}_{k_A \text{ volte}}\underbrace{P(B)\cdots P(B)}_{k_B \text{ volte}}\underbrace{P(C)\cdots P(C)}_{k_C \text{ volte}} \\ &= p_A^{k_A} p_B^{k_B} p_C^{k_C}. \end{aligned} \tag{2}$$

La probabilità cercata è data dalla (2) moltiplicata per le permutazioni di n elementi di cui k_A uguali tra loro, k_B uguali tra loro e k_C uguali tra loro e quindi

$$P(k_A \text{ arancioni, } k_B \text{ bianche in } n \text{ prove}) = \frac{n!}{k_A!k_B!k_C!} p_A^{k_A} p_B^{k_B} p_C^{k_C}$$

dove, naturalmente, $p_C = 1 - p_A - p_B$ e $k_C = n - k_A - k_B$.

(b) *Estrazioni senza ripetizione:*

Anche qui bisogna introdurre vincoli analoghi a quelli del problema 1(b) e, con un ragionamento del tutto simile, si ottiene

$$P(k_A \text{ arancioni, } k_B \text{ bianche in } n \text{ prove}) = \frac{\binom{K_A}{k_A}\binom{K_B}{k_B}\binom{K_C}{k_c}}{\binom{N}{n}}.$$