Urna di Pólya

Si consideri un'urna contenente b palline bianche e r palline rosse. Si estrae a caso una pallina, si osserva il colore, e la si reintroduce nell'urna insieme ad **una pallina aggiuntiva** dello stesso colore. Questo procedimento viene ripetuto per k estrazioni complessive.

1. Distribuzione del numero totale di palline bianche estratte

Sia X la variabile aleatoria che conta il numero di palline bianche tra le k estrazioni.

Allora:

$$\mathbb{P}(X = h) = \binom{k}{h} \cdot \frac{(b)_h (r)_{k-h}}{(b+r)_k} \quad \text{per } h = 0, 1, \ldots, k$$

dove $(a)_n$ denota il simbolo di Pochhammer (o fattoriale crescente):

$$(a)_n = a(a+1)(a+2)\cdots(a+n-1)$$

Questa distribuzione è detta **distribuzione beta-binomiale**.

2. Probabilità marginale che la i-esima estrazione sia bianca

Vogliamo dimostrare che, per ogni $i = 1, \ldots, k$, vale:

$$\mathbb{P}(\text{la } i\text{-esima è bianca}) = \frac{b}{b+r}$$

Dimostrazione (per induzione o simmetria di scambiabilità)

Grazie alla simmetria e scambiabilità delle estrazioni nel processo di Pólya, ogni estrazione ha la stessa distribuzione marginale.

Supponiamo che l'urna parta con b palline bianche e r rosse. Il processo incrementa il totale di palline, ma ogni estrazione dipende solo dal rapporto attuale, che si evolve in modo *rinforzato*.

Questo comportamento implica che ogni variabile indicatrice $I_j = 1$ se la j-esima estrazione è bianca, ha valore atteso:

$$\mathbb{E}[I_j] = \frac{b}{b+r}, \quad \forall j$$

e quindi anche la probabilità marginale:

$$\mathbb{P}(\text{la } i\text{-esima è bianca}) = \frac{b}{b+r}$$

Esempio concreto: $b = 2$, $r = 1$, $k = 3$

Obiettivo: calcolare $\mathbb{P}(B_2)$ e $\mathbb{P}(B_3)$

Passo 1: Calcolo di $\mathbb{P}(B_2)$ Ci sono due possibilità per la prima estrazione:

- Prima estrazione bianca: $\mathbb{P}(B_1) = \frac{2}{3}$
  - Dopo rinforzo: 3 bianche, 1 rossa
  - $\mathbb{P}(B_2 \mid B_1) = \frac{3}{4}$
- Prima estrazione rossa: $\mathbb{P}(R_1) = \frac{1}{3}$
  - Dopo rinforzo: 2 bianche, 2 rosse
  - $\mathbb{P}(B_2 \mid R_1) = \frac{2}{4} = \frac{1}{2}$

Applicando la formula della probabilità totale:

$$\mathbb{P}(B_2) = \mathbb{P}(B_2 \mid B_1) \cdot \mathbb{P}(B_1) + \mathbb{P}(B_2 \mid R_1) \cdot \mathbb{P}(R_1) = \frac{3}{4} \cdot \frac{2}{3} + \frac{1}{2} \cdot \frac{1}{3} = \frac{8}{12} = \frac{2}{3}$$

Passo 2: Calcolo di $\mathbb{P}(B_3)$ Ora analizziamo tutte le sequenze possibili delle prime due estrazioni:

- $B_1 B_2$:
  - Probabilità: $\frac{2}{3} \cdot \frac{3}{4} = \frac{1}{2}$
  - Urna: 4 bianche, 1 rossa $\Rightarrow \mathbb{P}(B_3 \mid B_1 B_2) = \frac{4}{5}$
  - Contributo: $\frac{1}{2} \cdot \frac{4}{5} = \frac{4}{10}$
- $B_1 R_2$:

- Probabilità: $\frac{2}{3} \cdot \frac{1}{4} = \frac{1}{6}$
- Urna: 3 bianche, 2 rosse $\Rightarrow \mathbb{P}(B_3 \mid B_1 R_2) = \frac{3}{5}$
- Contributo: $\frac{1}{6} \cdot \frac{3}{5} = \frac{3}{30}$

- $R_1 B_2$:
  - Probabilità: $\frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}$
  - Urna: 3 bianche, 2 rosse $\Rightarrow \mathbb{P}(B_3 \mid R_1 B_2) = \frac{3}{5}$
  - Contributo: $\frac{1}{6} \cdot \frac{3}{5} = \frac{3}{30}$
- $R_1 R_2$:
  - Probabilità: $\frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}$
  - Urna: 2 bianche, 3 rosse $\Rightarrow \mathbb{P}(B_3 \mid R_1 R_2) = \frac{2}{5}$
  - Contributo: $\frac{1}{6} \cdot \frac{2}{5} = \frac{2}{30}$

Somma dei contributi:

$$\mathbb{P}(B_3) = \frac{4}{10} + \frac{3}{30} + \frac{3}{30} + \frac{2}{30} = \frac{20}{30} = \frac{2}{3}$$

Conclusione

Per ogni $i = 1, \ldots, k$:

$$\mathbb{P}(\text{la } i\text{-esima estrazione è bianca}) = \frac{b}{b+r}$$

come atteso. La struttura del processo di Pólya conserva la distribuzione marginale invariata ad ogni passo, pur con dipendenza tra le variabili.