A question in Pólya model

史灵生*

During my course of Stochastics, a question in Pólya model is asked by students from time to time. The question is why a fault assumption may lead to a seemingly 'correct' solution to the following problem (Exercise 1.23(d) [1]).

Pólya model. Suppose there are r red and b black balls in a box. Each time pick one randomly from the box and put back $d+1 \geq 0$ balls with the same color to the picked one. Show that the probability of picking a black ball any time is always $b/(b+r)$.

I first give a correct proof and then discuss the question by students.

Proof. Let B_i denote the event that the picked ball is black at the ith time, and assume that there are in total X_n black balls picked within the first n times. Then $X_n = \sum_{i=1}^n I_{B_i}$ and for $\{n_1, n_2, ..., n_k\} \cup \{m_1, m_2, ..., m_{n-k}\} = \{1, 2, ..., n\}$, we have

$$P(B_{n_1} ... B_{n_k} B_{m_1}^c ... B_{m_{n-k}}^c) = \prod_{i=0}^{k-1}(b+di) \prod_{j=0}^{n-k+1}(r+dj) \prod_{l=0}^{n-1}(b+r+dl)^{-1}.$$

By the finite additivity, we obtain

$$P(X_n = k) = \binom{n}{k} \prod_{i=0}^{k-1}(b+di) \prod_{j=0}^{n-k+1}(r+dj) \prod_{l=0}^{n-1}(b+r+dl)^{-1}.$$

Note that $P(B_{n+1} \mid X_n = k) = (b+kd)/(b+r+nd)$. Setting $b' = b+d$ and by the total probability law, we have

$$\begin{aligned}
P(B_{n+1}) &= \sum_{k=0}^{n} P(X_n = k) P(B_{n+1} \mid X_n = k) \\
&= \sum_{k=0}^{n} \binom{n}{k} \prod_{i=0}^{k}(b+di) \prod_{j=0}^{n-k+1}(r+dj) \prod_{l=0}^{n}(b+r+dl)^{-1} \\
&= \frac{b}{b+r} \sum_{k=0}^{n} \binom{n}{k} \prod_{i=0}^{k-1}(b'+di) \prod_{j=0}^{n-k+1}(r+dj) \prod_{l=0}^{n-1}(b'+r+dl)^{-1} \\
&= b/(b+r). \quad \square
\end{aligned}$$

Some students used an induction on n with a fault assumption of the distribution of X_n but resulting in the same value for $P(B_{n+1})$. Let us see how they did. By induction, $P(B_i) = b/(b+r)$

*清华数学系计算数学与运筹学研究所副教授，E-mail: lshi@math.tsinghua.edu.cn

for $i \le n$. Then they thought $X_n \sim B(n, b/(b+r))$ (binomial distribution) by a mistake and went on with

$$\begin{aligned} P(B_{n+1}) &= \sum_{k=0}^{n} P(X_n = k) P(B_{n+1} \mid X_n = k) \\ &= \sum_{k=0}^{n} \binom{n}{k} \left(\frac{b}{b+r}\right)^k \left(\frac{r}{b+r}\right)^{n-k} \frac{b+kd}{b+r+nd} \\ &= \frac{b}{b+r+nd} + \frac{d}{b+r+nd} \sum_{k=0}^{n} k \binom{n}{k} \left(\frac{b}{b+r}\right)^k \left(\frac{r}{b+r}\right)^{n-k} \\ &= \frac{b}{b+r+nd} + \frac{d}{b+r+nd} \frac{nb}{b+r} = \frac{b}{b+r}. \end{aligned}$$

Why may this happen? In fact, we can compute $P(B_{n+1})$ as follows,

$$\begin{aligned} P(B_{n+1}) &= \sum_{k=0}^{n} P(X_n = k) P(B_{n+1} \mid X_n = k) \\ &= \sum_{k=0}^{n} P(X_n = k)(b+kd)/(b+r+nd) \\ &= \frac{b}{b+r+nd} + \frac{d}{b+r+nd} \sum_{k=0}^{n} k P(X_n = k) \\ &= \frac{b}{b+r+nd} + \frac{dEX_n}{b+r+nd}. \end{aligned}$$

Thus it is easily seen that one would get the same value for $P(B_{n+1})$ whenever $EX_n = nb/(b+r)$ regardless of the distribution of X_n at all!

参考文献

[1] 钱敏平和叶俊,《随机数学》, 大学数学第二版, 萧树铁主编, 高等教育出版社, 2004。

****

数学趣闻

Napoleon: You have written this huge book on the system of the world without once mentioning the author of the universe.

Laplace: Sire, I had no need of that hypothesis.

Later when told by Napoleon about the incident, Lagrange commented: Ah, but that is a fine hypothesis. It explains so many things.