詳細解答

STEP III 2001，第 12 題

概率與統計 • 以指示變量求期望值

STEP III 2001，第 12 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

step.maths.org/questions/01-s3-q12

關鍵思路

在第 (i) 部分，不要嘗試求出黑球數目的分佈：把這個數目拆開，每個黑球對應一個指示變數（indicator），然後運用期望值的線性性質（linearity of expectation）。不放回地抽球直至袋空，其實就是把所有球隨機排成一列，所以每個指示變數都歸結為一個對稱性論證（symmetry argument）。在第 (ii) 部分，兩次抽到白球之間袋中的球並無改變，所以等候每個白球出現的次數服從幾何分佈（geometric distribution）；而題目中的「hence」，則來自這個觀察：取走一個白球之後，我們回到同樣的處境，只是白球少了一個。

解答

第 (i) 部分

設想一直抽下去，直至袋空為止。由於每次都是隨機抽取，而且抽出的球不放回，這 $b+w$ 個球被抽出的次序，就是它們的一個均勻隨機排列（uniformly random ordering）。到最後一個白球被取走時已抽出的黑球，正好就是在這個排列中位於最後一個白球之前的黑球。

把黑球編號為 $1, 2, \ldots, b$；若黑球 j 在最後一個白球之前被抽出，令 $I_j = 1$，否則令 $I_j = 0$。設 N 為最後一個白球被取走時已抽出的黑球數目，則

$$N = I_1 + I_2 + \cdots + I_b, \qquad \text{所以} \qquad \mathrm{E}(N) = \sum_{j=1}^{b} \mathrm{P}(I_j = 1)$$

這是根據期望值的線性性質（它並不需要各個 I_j 之間互相獨立）。

想一想 黑球 j 排在所有白球之後的概率是多少？

只看由黑球 j 和 w 個白球組成的這 $w+1$ 個球。在所有球的均勻隨機排列中，這 $w+1$ 個球的相對次序同樣是均勻隨機的，所以黑球 j 處於它們當中 $w+1$ 個位置的任何一個，機會均等。它排在這組球最後的概率是 $\frac{1}{w+1}$，而這正是事件 $I_j = 0$。因此

$$\mathrm{P}(I_j = 1) = 1 - \frac{1}{w+1} = \frac{w}{w+1},$$

從而

$$\boxed{\mathrm{E}(N) = \frac{bw}{w+1}.}$$

第 (ii) 部分

現在黑球會放回袋中，所以在抽到白球之前，袋中始終有 b 個黑球和 w 個白球。因此每次抽到黑球的概率是 $p = \frac{b}{b+w}$，抽到白球的概率是 $q = \frac{w}{b+w}$，而且與之前各次抽取獨立，因為每次袋中的球都相同。

設 X 為抽到第一個白球之前抽到黑球的次數。那麼 $X = n$ 表示先抽到 n 次黑球，然後抽到一個白球，所以

$$\mathrm{P}(X = n) = p^n q, \qquad n = 0, 1, 2, \ldots$$

若 $b = 0$，則 $X = 0$，結果顯然成立，故取 $b \geq 1$；這時 $0 < p < 1$，利用當 $|x| < 1$ 時 $\sum_{n\geq 1} nx^{n-1} = \frac{1}{(1-x)^2}$（把等比級數（geometric series）求導即得），

$$\mathrm{E}(X) = \sum_{n=0}^{\infty} np^n q = pq \sum_{n=1}^{\infty} np^{n-1} = \frac{pq}{(1-p)^2} = \frac{p}{q} = \frac{b/(b+w)}{w/(b+w)} = \frac{b}{w},$$

因為 $1 - p = q$。這就是所要證明的結果。

至於最後一個白球，把抽到黑球的總次數寫成 $X_w + X_{w-1} + \cdots + X_1$，其中 X_k 是袋中恰好有 k 個白球期間抽到黑球的次數。當袋中尚餘 k 個白球時，袋中有 b 個黑球和 k 個白球，所以這個階段正是把 w 換成 k 的「第一個白球」問題，而剛才證明的結果給出 $\mathrm{E}(X_k) = \frac{b}{k}$。根據期望值的線性性質，最後一個白球被取走時抽到黑球的期望次數是

$$\boxed{\sum_{k=1}^{w} \frac{b}{k} = b\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{w}\right).}$$

驗算

當 $w = 1$ 時，兩部分的結果都很容易直接看出。在 (i) 中，每一個黑球排在白球之前的概率都是 $\frac{1}{2}$，得出 $\frac{b}{2}$，與 $\frac{bw}{w+1}$ 相符。在 (ii) 中，只有一個階段，而總和就是 $\frac{b}{1} = b$。在 (i) 中當 $w \to \infty$ 時，幾乎每個黑球都在最後一個白球之前被抽出，而確實有 $\frac{bw}{w+1} \to b$。

我們對所有 $b \leq 4$、$w \leq 4$ 的情況逐一列出每個排列來驗算 (i)；當 $b = 3$、$w = 4$ 時，結果恰好是 $\frac{12}{5}$。至於 (ii)，以 $b = 3$、$w = 4$ 進行 2×10^5 次模擬，所得平均值約為 6.245，而 $3(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4}) = \frac{25}{4} = 6.25$。

常見錯誤

- 在 (i) 中，假設各個黑球是否被抽出是互相獨立的。期望值的線性性質並不需要獨立性，這正是指示變數如此好用的原因。
- 在 (ii) 中，採用了把抽到白球那一次也計算在內的幾何分佈，結果得出 $\frac{1}{q} = \frac{b+w}{w}$，比答案多了一。
- 在最後一部分的每個階段都以 $b + w$ 作分母：袋中白球的數目每過一個階段就減少一個。

STEP 之外

第二部分其實是喬裝了的優惠券收集問題（coupon collector's problem）：把各個階段的幾何等候時間加起來，便得到一個調和和（harmonic sum），其增長速度與 $\ln w$ 相若。像第 (i) 部分那樣把一個計數寫成指示變數之和，是概率論中最強而有力的技巧之一；日後計算隨機置換（random permutation）中不動點（fixed point）、紀錄（record）和循環（cycle）的期望數目時，你會再次遇上它。Blitzstein 與 Hwang 合著的 *Introduction to Probability*（概率論入門）中，有許多關於這兩個想法的例子。

Beyond STEP 詳細解答：https://beyondstep.keithmcnulty.org/zh-hant/。解答文字採用 CC0 1.0；題目本身版權屬 Cambridge University Press & Assessment 所有（©），本文件不予轉載。由機器從英文翻譯而成；尚未經審閱。