詳細解答

STEP III 2011，第 13 題

概率與統計 • 離散分佈的眾數

STEP III 2011，第 13 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/11-s3-q13

關鍵思路

要找出離散隨機變量（discrete random variable）最可能的取值，我們不用微分，而是比較相鄰的概率（probability）。比值 $\mathrm{P}(X = r+1)/\mathrm{P}(X = r)$ 是 r 的一個簡單有理函數（rational function），而問它何時大於 1，就變成一個關於 r 的線性不等式（linear inequality）。比值大於 1 時概率上升，一旦跌到 1 以下概率便下降，所以峯值就在轉折前的最後一個 r。同一方法無論放回抽樣（with replacement）還是不放回抽樣（without replacement）都適用，改變的只是那個比值。

解答

第 (i) 部分

在放回抽樣下，各次抽取互相獨立，每次抽到黑球的概率都是 $p = b/n$，所以 $X \sim \mathrm{B}(k, p)$：

$$\mathrm{P}(X = r) = \binom{k}{r} p^r (1-p)^{k-r}, \qquad r = 0, 1, \ldots, k.$$

先假設 $0 < b < n$，使得這些概率全部為正。對於 $0 \le r \le k-1$，

$$\frac{\mathrm{P}(X = r+1)}{\mathrm{P}(X = r)} = \frac{k-r}{r+1} \cdot \frac{p}{1-p} = \frac{(k-r)\,b}{(r+1)(n-b)}.$$

由於分母為正，這個比值大於 1 當且僅當

$$\begin{aligned} & (k-r)b > (r+1)(n-b) \\ \iff \quad & kb - rb > rn - rb + n - b \\ \iff \quad & (k+1)b > (r+1)n \\ \iff \quad & r+1 < \frac{(k+1)b}{n}. \end{aligned}$$

同理，比值等於 1 當且僅當 $r+1 = m$，小於 1 當且僅當 $r+1 > m$，其中

$$m = \frac{(k+1)b}{n}.$$

因此，隨着 r 增大，每一步若終點 $r+1 < m$，$\mathrm{P}(X = r)$ 便嚴格遞增；若終點 $r+1 > m$，便嚴格遞減。注意由於 $0 < b < n$，有 $0 < m < k+1$，所以轉折發生在取值範圍 $0, \ldots, k$ 之內。

想一想 如果 m 不是整數，最大的概率究竟在哪裏？

若 m 不是整數，記 $M = \lfloor m \rfloor$，則 $M < m < M+1$ 且 $0 \leq M \leq k$。概率嚴格遞增直至 $\mathrm{P}(X = M)$（每一步的終點都 $\leq M < m$），之後嚴格遞減（之後每一步的終點都 $\geq M+1 > m$）。因此黑球數目唯一的最可能取值是

$$\left\lfloor \frac{(k+1)b}{n} \right\rfloor,$$

正如所求。

不唯一的情況。若 m 是整數，由 $r = m-1$ 到 $r = m$ 這一步的比值恰好是 1，所以 $\mathrm{P}(X = m-1) = \mathrm{P}(X = m)$，而這兩個相等的值比其餘所有值都大（概率嚴格上升至 $m-1$，並在 m 之後嚴格下降）。由於 $1 \leq m \leq k$，兩個取值都是可能的。所以答案不唯一，當且僅當 $(k+1)b/n$ 是整數（且 $0 < b < n$），此時 $m-1$ 和 m 同樣是最可能的取值。例如 $n = 3$、$b = 1$、$k = 5$ 給出 $m = 2$，而確實 $\mathrm{P}(X = 1) = \mathrm{P}(X = 2) = \frac{80}{243}$。

平凡情況不在上述討論之內：若 $b = 0$，則必然 $X = 0$，與公式相符；若 $b = n$，則必然 $X = k$，而公式（會給出 $k+1$）並不適用，因為上面的比值涉及除以 $n-b$。

第 (ii) 部分

在不放回抽樣下，X 服從超幾何分佈（hypergeometric distribution）：把樣本視為 k 個球組成的無序集合，所有 $\binom{n}{k}$ 個集合機會均等，

$$\mathrm{P}(X = r) = \frac{\binom{b}{r}\binom{n-b}{k-r}}{\binom{n}{k}}.$$

由於 $k \le b$，最大可能取值是 $r = k$；最小的是 $r_0 = \max(0, k-(n-b))$，因為非黑球最多只能抽出 $n-b$ 個。對於 $r_0 \le r \le k$，概率為正；而對於 $r_0 \le r \le k-1$，利用 $\binom{a}{r+1} = \binom{a}{r}\frac{a-r}{r+1}$ 和 $\binom{a}{s-1} = \binom{a}{s}\frac{s}{a-s+1}$，

$$\frac{\mathrm{P}(X=r+1)}{\mathrm{P}(X=r)} = \frac{b-r}{r+1}\cdot\frac{k-r}{n-b-k+r+1}.$$

在此範圍內兩個分母都為正，所以比值大於 1 當且僅當

$$(b-r)(k-r) > (r+1)(n-b-k+r+1).$$

展開後，左邊是 $bk-r(b+k)+r^2$，右邊是 $(r+1)(n-b-k+1)+r^2+r$。r^2 項互相抵消，不等式變為

$$\begin{aligned} & bk-r(b+k) > r(n-b-k+1)+(n-b-k+1)+r \\ \iff\quad & bk+b+k+1 > r(n+2)+(n+2) \\ \iff\quad & r+1 < \frac{(b+1)(k+1)}{n+2}. \end{aligned}$$

想一想 轉折點是否真的落在可能取值的範圍 $r_0, \ldots, k$ 之內？

設 $m = \dfrac{(b+1)(k+1)}{n+2}$。由於 $b+1 < n+2$，有 $m < k+1$。至於下端，$m > 0$ 是顯然的；而若 $r_0 = k+b-n > 0$，則需要 $(b+1)(k+1) > (n+2)(k+b-n)$。記 $x = n-b$ 及 $y = n-k$，左邊減右邊等於

$$(n+1-x)(n+1-y) - [(n+1)^2 - 1 - (n+2)(x+y)] = 1+x+y+xy > 0.$$

所以在所有情況下都有 $r_0 < m < k+1$，第 (i) 部分的論證可以原封不動地套用：當 $r+1 < m$ 時概率嚴格上升，一旦 $r+1 > m$ 便嚴格下降。因此，在唯一的情況下，黑球數目最可能的取值是

$$\left\lfloor \frac{(b+1)(k+1)}{n+2} \right\rfloor.$$

它不唯一，當且僅當 $(b+1)(k+1)/(n+2)$ 是整數；此時 $m-1$ 和 m（兩者都是可能的取值，因為 $r_0 \le m-1$ 且 $m \le k$）機會相等，並且比其他任何取值都更可能。例如 $n=10$、$b=7$、$k=5$ 給出 $m=4$，而 $X=3$ 與 $X=4$ 打和，概率各為 $\frac{105}{252} = \frac{5}{12}$。

驗算

- 我們用電腦以分數精確計算了所有 $n \le 24$、$0 \le b \le n$ 及 $1 \le k \le 19$（第 (ii) 部分另要求 $k \le b$）的兩個分佈，以窮舉法找出眾數，並在所有情況下確認了兩條公式及兩個打和條件（第 (ii) 部分共 2565 個情況）。

- 對於 $n = 10$、$b = 6$、$k = 4$：$m = \frac{35}{12}$，公式給出 2，而 $X = 2$ 確實是唯一的眾數。
- 若 n 遠大於 k，取走幾個球幾乎不改變黑球所佔的比例，所以第 (ii) 部分應該與第 (i) 部分相若。事實正是如此：$(b+1)(k+1)/(n+2) \approx (k+1)b/n$。

常見錯誤

- 斷定眾數（mode）是 $\lfloor kb/n \rfloor$，即平均值（mean）的整數部分。眾數接近平均值，但不一定等於其整數部分。
- 忘記當比值恰好等於 1 時，眾數有兩個，而不是一個。
- 在第 (ii) 部分展開 $(r+1)(n-b-k+r+1)$ 時弄錯正負號；r^2 項互相抵消是一個有用的檢查。

STEP 之外

比較相鄰概率是找出任何離散分佈眾數的標準方法，而同一計算作為第 (i) 部分的極限情況，給出泊松分佈（Poisson distribution）的眾數 $\lfloor \lambda \rfloor$。超幾何分佈是抽樣理論的核心，也是統計學中費雪（Fisher）精確檢驗（Fisher's exact test）的基礎。概率像這樣先升後降的分佈稱為單峯分佈（unimodal）；對於二項分佈和超幾何分佈，這源自一個更強的性質——對數凹性（log-concavity）。如欲深入了解，可參閱 Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率導論》），或 Blitzstein 和 Hwang 的 *Introduction to Probability*（《概率導論》）。

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。