详细解答

STEP III 2011，第 13 题

概率与统计 • 离散分布的众数

STEP III 2011，第 13 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/11-s3-q13

关键思路

要找出离散随机变量（discrete random variable）最可能的取值，我们不用微分，而是比较相邻的概率（probability）。比值 $\mathrm{P}(X = r+1)/\mathrm{P}(X = r)$ 是 r 的一个简单有理函数（rational function），而问它何时大于 1，就变成一个关于 r 的线性不等式（linear inequality）。比值大于 1 时概率上升，一旦跌到 1 以下概率便下降，所以峰值就在转折前的最后一个 r。同一方法无论放回抽样（with replacement）还是不放回抽样（without replacement）都适用，改变的只是那个比值。

解答

第 (i) 部分

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

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

先假设 $0 < b < n$，使得这些概率全部为正。对于 $0 \le r \le k-1$，

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

由于分母为正，这个比值大于 1 当且仅当

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

同理，比值等于 1 当且仅当 $r+1=m$，小于 1 当且仅当 $r+1>m$，其中

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

因此，随着 r 增大，每一步若终点 $r+1<m$，$\mathrm{P}(X=r)$ 便严格递增；若终点 $r+1>m$，便严格递减。注意由于 $0<b<n$，有 $0<m<k+1$，所以转折发生在取值范围 $0,\ldots,k$ 之内。

> ? **想一想** 如果 m 不是整数，最大的概率究竟在哪里?

若 m 不是整数，记 $M=\lfloor m\rfloor$，则 $M<m<M+1$ 且 $0\le M\le k$。概率严格递增直至 $\mathrm{P}(X=M)$（每一步的终点都 $\le M<m$），之后严格递减（之后每一步的终点都 $\ge M+1>m$）。因此黑球数目唯一的最可能取值是

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

正如所求。

不唯一的情况。若 m 是整数，由 $r=m-1$ 到 $r=m$ 这一步的比值恰好是 1，所以 $\mathrm{P}(X=m-1)=\mathrm{P}(X=m)$，而这两个相等的值比其余所有值都大（概率严格上升至 $m-1$，并在 m 之后严格下降）。由于 $1\le m\le k$，两个取值都是可能的。所以答案不唯一，当且仅当 $(k+1)b/n$ 是整数（且 $0<b<n$），此时 $m-1$ 和 m 同样是最可能的取值。例如 $n=3$、$b=1$、$k=5$ 给出 $m=2$，而确实 $\mathrm{P}(X=1)=\mathrm{P}(X=2)=\frac{80}{243}$。

平凡情况不在上述讨论之内：若 $b=0$，则必然 $X=0$，与公式相符；若 $b=n$，则必然 $X=k$，而公式（会给出 $k+1$）并不适用，因为上面的比值涉及除以 $n-b$。

第 (ii) 部分

在不放回抽样下，X 服从超几何分布（hypergeometric distribution）：把样本视为 k 个球组成的无序集合，所有 $\binom{n}{k}$ 个集合机会均等，

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

由于 $k \le b$，最大可能取值是 $r = k$；最小的是 $r_0 = \max(0, k-(n-b))$，因为非黑球最多只能抽出 $n-b$ 个。对于 $r_0 \le r \le k$，概率为正；而对于 $r_0 \le r \le k-1$，利用 $\binom{a}{r+1} = \binom{a}{r}\frac{a-r}{r+1}$ 和 $\binom{a}{s-1} = \binom{a}{s}\frac{s}{a-s+1}$，

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

在此范围内两个分母都为正，所以比值大于 1 当且仅当

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

展开后，左边是 $bk - r(b+k) + r^2$，右边是 $(r+1)(n-b-k+1) + r^2 + r$。r^2 项互相抵消，不等式变为

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

想一想 转折点是否真的落在可能取值的范围 $r_0, \ldots, k$ 之内?

设 $m = \dfrac{(b+1)(k+1)}{n+2}$。由于 $b+1 < n+2$，有 $m < k+1$。至于下端，$m > 0$ 是显然的；而若 $r_0 = k+b-n > 0$，则需要 $(b+1)(k+1) > (n+2)(k+b-n)$。记 $x = n-b$ 及 $y = n-k$，左边减右边等于

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

所以在所有情况下都有 $r_0 < m < k+1$，第 (i) 部分的论证可以原封不动地套用：当 $r+1 < m$ 时概率严格上升，一旦 $r+1 > m$ 便严格下降。因此，在唯一的情况下，黑球数目最可能的取值是

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

它不唯一，当且仅当 $(b+1)(k+1)/(n+2)$ 是整数；此时 $m-1$ 和 m（两者都是可能的取值，因为 $r_0 \le m-1$ 且 $m \le k$）机会相等，并且比其他任何取值都更可能。例如 $n = 10$、$b = 7$、$k = 5$ 给出 $m = 4$，而 $X = 3$ 与 $X = 4$ 打和，概率各为 $\frac{105}{252} = \frac{5}{12}$。

验算

- 我们用电脑以分数精确计算了所有 $n \le 24$、$0 \le b \le n$ 及 $1 \le k \le 19$（第 (ii) 部分另要求 $k \le b$）的两个分布，以穷举法找出众数，并在所有情况下确认了两条公式及两个打和条件（第 (ii) 部分共 2565 个情况）。

- 对于 $n = 10$、$b = 6$、$k = 4$：$m = \frac{35}{12}$，公式给出 2，而 $X = 2$ 确实是唯一的众数。
- 若 n 远大于 k，取走几个球几乎不改变黑球所占的比例，所以第 (ii) 部分应该与第 (i) 部分相若。事实正是如此：$(b+1)(k+1)/(n+2) \approx (k+1)b/n$。

常见错误

- 断定众数（mode）是 $\lfloor kb/n \rfloor$，即平均值（mean）的整数部分。众数接近平均值，但不一定等于其整数部分。
- 忘记当比值恰好等于 1 时，众数有两个，而不是一个。
- 在第 (ii) 部分展开 $(r+1)(n-b-k+r+1)$ 时弄错正负号；r^2 项互相抵消是一个有用的检查。

STEP 之外

比较相邻概率是找出任何离散分布众数的标准方法，而同一计算作为第 (i) 部分的极限情况，给出泊松分布（Poisson distribution）的众数 $\lfloor \lambda \rfloor$。超几何分布是抽样理论的核心，也是统计学中费雪（Fisher）精确检验（Fisher's exact test）的基础。概率像这样先升后降的分布称为单峰分布（unimodal）；对于二项分布和超几何分布，这源自一个更强的性质——对数凹性（log-concavity）。如欲深入了解，可参阅 Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率导论》），或 Blitzstein 和 Hwang 的 *Introduction to Probability*（《概率导论》）。

Beyond STEP 详细解答：https://beyondstep.keithmcnulty.org/zh-hans/。解答文字采用 CC0 1.0；题目本身版权属 Cambridge University Press & Assessment 所有（©），本文件不予转载。由机器从英文翻译而成；尚未经审阅。