详细解答

STEP III 2022，第 11 题

概率与统计 · 二项分布的平均偏差

STEP III 2022，第 11 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2022 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696587-2022-past-papers.zip

关键思路

硬币公平，所以 X 的分布（distribution）关于其均值（mean）对称（symmetric），即低于均值的概率与高于均值的概率一一对应。利用这种对称性，可以把绝对值（absolute value）化为单侧的和。接着，(i)(c) 中的恒等式（identity）把 $\sum r\binom{2n}{r}$ 化为一个较小的二项分布（binomial distribution）的尾部和（tail sum），而这个尾部和同样可以用对称性求出。

解答

全题中 $X \sim \mathrm{B}(N, \frac{1}{2})$，所以 $\mathrm{P}(X = r) = \binom{N}{r}2^{-N}$，且 $\mu = \frac{1}{2}N$。由于 $\binom{N}{r} = \binom{N}{N-r}$，我们有 $\mathrm{P}(X = r) = \mathrm{P}(X = N - r)$：分布关于 $\frac{1}{2}N$ 对称。

第 (i)(a) 部分

当 $N = 2n$ 时，映射 $r \mapsto 2n - r$ 把 $\{0, \ldots, n-1\}$ 映到 $\{n+1, \ldots, 2n\}$，所以由对称性得

$$\mathrm{P}(X \leq n-1) = \mathrm{P}(X \geq n+1).$$

三个事件 $X \leq n-1$、$X = n$、$X \geq n+1$ 涵盖了所有结果，所以 $2\,\mathrm{P}(X \leq n-1) + \mathrm{P}(X = n) = 1$，由此得

$$\mathrm{P}(X \leq n-1) = \tfrac{1}{2}(1 - \mathrm{P}(X = n)).$$

第 (i)(b) 部分

此时 $\mu = n$，所以

$$\delta = \sum_{r=0}^{2n} |r-n| \binom{2n}{r} \frac{1}{2^{2n}}.$$

$r = n$ 的项为零。在 $r \geq n+1$ 的各项中代入 $r = 2n-s$，得 $|r-n| = n-s$ 及 $\binom{2n}{r} = \binom{2n}{s}$，其中 $0 \leq s \leq n-1$。因此，高于均值的各项恰好重复了低于均值的各项：

$$\delta = 2\sum_{r=0}^{n-1} (n-r)\binom{2n}{r}\frac{1}{2^{2n}} = \sum_{r=0}^{n-1}(n-r)\binom{2n}{r}\frac{1}{2^{2n-1}},$$

得证。

第 (i)(c) 部分

对于 $1 \leq r \leq 2n$，

$$r\binom{2n}{r} = \frac{r\,(2n)!}{r!\,(2n-r)!} = 2n \cdot \frac{(2n-1)!}{(r-1)!\,(2n-r)!} = 2n\binom{2n-1}{r-1}.$$

现在把 (b) 中的和拆成 $2^{2n-1}\delta = A - B$，其中

$$A = n\sum_{r=0}^{n-1}\binom{2n}{r}, \qquad B = \sum_{r=1}^{n-1} r\binom{2n}{r}$$

（B 中 $r=0$ 的项为零，所以略去）。

和 A。由 (a)，$\sum_{r=0}^{n-1}\binom{2n}{r} = 2^{2n}\,\mathrm{P}(X \leq n-1) = 2^{2n-1} - \frac{1}{2}\binom{2n}{n}$，所以

$$A = n\,2^{2n-1} - \tfrac{n}{2}\binom{2n}{n}.$$

? **想一想** 当 s 由 0 到 $n-2$ 时，怎样求 $\binom{2n-1}{s}$ 的总和?

和 B。由上述恒等式，令 $s = r-1$，

$$B = 2n\sum_{s=0}^{n-2}\binom{2n-1}{s}.$$

这一行 $\binom{2n-1}{s}$ $(0 \leq s \leq 2n-1)$ 共有偶数个项，总和为 2^{2n-1}。它在 $s \mapsto 2n-1-s$ 下对称，而这个映射把 $\{0,\ldots,n-1\}$ 与 $\{n,\ldots,2n-1\}$ 互换，所以前一半的和是 2^{2n-2}。我们的和比这一半少了最后一项，所以

$$\sum_{s=0}^{n-2}\binom{2n-1}{s} = 2^{2n-2} - \binom{2n-1}{n-1} = 2^{2n-2} - \tfrac{1}{2}\binom{2n}{n},$$

这里用了 $\binom{2n}{n} = \binom{2n-1}{n-1} + \binom{2n-1}{n} = 2\binom{2n-1}{n-1}$。因此

$$B = n\,2^{2n-1} - n\binom{2n}{n}.$$

合并起来。

$$2^{2n-1}\delta = A - B = \left(n - \frac{n}{2}\right)\binom{2n}{n} = \frac{n}{2}\binom{2n}{n}, \qquad \text{所以} \qquad \delta = \frac{n}{2^{2n}}\binom{2n}{n}.$$

第 (ii) 部分

现在 $N = 2n+1$，$\mu = n + \frac{1}{2}$，这并不是 X 可取的值。对称映射 $r \mapsto 2n+1-r$ 把 $\{0,\ldots,n\}$ 与 $\{n+1,\ldots,2n+1\}$ 互换，并且保持 $|r-\mu|$ 和 $\binom{2n+1}{r}$ 不变。所以，与 (i)(b) 一样，

$$\delta = 2\sum_{r=0}^{n}\left(n+\tfrac{1}{2}-r\right)\binom{2n+1}{r}\frac{1}{2^{2n+1}} = \frac{1}{2^{2n+1}}\sum_{r=0}^{n}(2n+1-2r)\binom{2n+1}{r}.$$

- 由同样的对称性，这一行 $\binom{2n+1}{r}$ 前一半的和是 2^{2n+1} 的一半：$\sum_{r=0}^{n}\binom{2n+1}{r} = 2^{2n}$。
- 由恒等式 $r\binom{2n+1}{r} = (2n+1)\binom{2n}{r-1}$（证明与 (i)(c) 完全相同），再用 (i)(a)，

$$\sum_{r=1}^{n} r\binom{2n+1}{r} = (2n+1)\sum_{s=0}^{n-1}\binom{2n}{s} = (2n+1)\left(2^{2n-1} - \tfrac{1}{2}\binom{2n}{n}\right).$$

因此

$$\sum_{r=0}^{n}(2n+1-2r)\binom{2n+1}{r} = (2n+1)2^{2n} - (2n+1)\left(2^{2n} - \binom{2n}{n}\right) = (2n+1)\binom{2n}{n},$$

而当 $N = 2n+1$ 时，

$$\boxed{\delta = \frac{2n+1}{2^{2n+1}}\binom{2n}{n}}$$

验算

- $N = 3$（$n = 1$）：$|X - \frac{3}{2}|$ 以概率 $\frac{1}{4}$ 取值 $\frac{3}{2}$，以概率 $\frac{3}{4}$ 取值 $\frac{1}{2}$，所以 $\delta = \frac{3}{8} + \frac{3}{8} = \frac{3}{4}$。公式给出 $\frac{3\cdot 2}{8} = \frac{3}{4}$。当 $n = 0$（$N = 1$）时，公式给出 $\frac{1}{2}$，同样正确。
- $N = 4$（$n = 2$）：(i)(c) 给出 $\frac{2\cdot 6}{16} = \frac{3}{4}$，与 $N = 3$ 的结果相同。事实上，$N = 2n+1$ 时的 δ 总是等于 $N = 2n+2$ 时的 δ：由于 $\binom{2n+2}{n+1} = \frac{2(2n+1)}{n+1}\binom{2n}{n}$，把偶数情况公式中的 n 换成 $n+1$，便化简为奇数情况的公式。
- 经电脑验算，对所有 $N \le 60$，两条公式都与 δ 的定义完全吻合；例如 $N = 10$ 给出 $\delta = \frac{315}{256} \approx 1.230$。

常见错误

- 忘记对称性带来的因子 2；(i)(b) 中出现 2^{2n-1} 而不是 2^{2n}，正是因为这个因子。
- 作平移 $s = r - 1$ 后弄错 s 的范围：和止于 $n-2$，比半行少一项，而正是缺少的那一项 $\binom{2n-1}{n-1}$ 令答案不为零。
- 在 (ii) 中把 μ 当作整数：其实没有任何一项为零，而且这一行的两半各有 $n+1$ 项。

STEP 之外

斯特林公式（Stirling's formula）给出 $\binom{2n}{n} \approx 4^n/\sqrt{\pi n}$，所以当 N 很大时，$\delta \approx \sqrt{N/(2\pi)}$。由于 $\sigma = \frac{1}{2}\sqrt{N}$，这就是 $\delta \approx \sigma\sqrt{2/\pi}$，恰好是正态分布（normal distribution）的平均绝对偏差（mean absolute deviation），正如中心极限定理（central limit theorem）所预测。Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率导论》）既讲解了二项分布的正态近似，也讲解了其背后的中心极限定理。

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。