详细解答

STEP III 2016，第 12 题

概率与统计 · 切比雪夫不等式

> **STEP III 2016，第 12 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/16-s3-q12

关键思路

切比雪夫不等式（Chebyshev's inequality）给出的是远离均值（mean）的概率的上界，所以要证明某个量接近均值，就要从补事件（complement）入手。每一部分的工作都是：找出合适的随机变量（random variable），求出它的均值和标准差（standard deviation），再选取 k，使 $k\sigma$ 恰好等于我们关心的距离。在第 (ii) 部分，和式 $\sum n^r/r!$ 与 e^n 并列出现，正是泊松分布（Poisson distribution）的标志。

解答

第 (i) 部分

设 X 为 $100n$ 次抛掷中得到正面的次数。假设各次抛掷互相独立（independent），而每次得到正面的概率为 0.2，则 $X \sim \mathrm{B}(100n, 0.2)$，因此

$$\mu = 100n \times 0.2 = 20n, \qquad \sigma^2 = 100n \times 0.2 \times 0.8 = 16n, \qquad \sigma = 4\sqrt{n}.$$

事件 $16n \le X \le 24n$ 正好就是 $|X - 20n| \le 4n$，所以

$$\alpha = \mathrm{P}(|X - \mu| \le 4n) = 1 - \mathrm{P}(|X - \mu| > 4n).$$

我们希望 $k\sigma = 4n$，即 $k \cdot 4\sqrt{n} = 4n$，故 $k = \sqrt{n}$，这是正数。由切比雪夫不等式得

$$\mathrm{P}(|X - \mu| > 4n) = \mathrm{P}(|X - \mu| > \sqrt{n}\,\sigma) \le \frac{1}{(\sqrt{n})^2} = \frac{1}{n},$$

因此

$$\alpha \ge 1 - \frac{1}{n},$$

得证。（条件 $n > 1$ 只是为了让这个界有实质内容：当 $n = 1$ 时，它只说 $\alpha \ge 0$。）

第 (ii) 部分

想一想 哪一个随机变量的概率形如 $n^r/r!$?

若 $Y \sim \mathrm{Po}(n)$，则对 $r = 0, 1, 2, \ldots$ 有 $\mathrm{P}(Y = r) = \mathrm{e}^{-n} n^r / r!$，所以

$$\mathrm{e}^{-n}\left(1 + n + \frac{n^2}{2!} + \cdots + \frac{n^{2n}}{(2n)!}\right) = \mathrm{P}(Y \le 2n).$$

所求不等式两边除以 $\mathrm{e}^n > 0$ 后，便等价于

$$\mathrm{P}(Y \le 2n) \ge 1 - \frac{1}{n}.$$

泊松分布的均值和方差（variance）都是 n，所以 $\mu = n$，$\sigma = \sqrt{n}$。

想一想 切比雪夫不等式是双侧的，但 $Y \le 2n$ 是单侧事件。两者如何对应起来?

由于 $Y \ge 0$，必有 $Y - n \ge -n$，所以 $Y - n < -n$ 不可能发生。因此

$$|Y - n| > n \iff Y - n > n \iff Y > 2n.$$

取 $k = \sqrt{n}$，使 $k\sigma = n$，由切比雪夫不等式得

$$\mathrm{P}(Y > 2n) = \mathrm{P}(|Y - n| > \sqrt{n}\,\sigma) \le \frac{1}{n},$$

所以 $\mathrm{P}(Y \le 2n) = 1 - \mathrm{P}(Y > 2n) \ge 1 - \frac{1}{n}$。两边乘以 e^n，

$$1 + n + \frac{n^2}{2!} + \cdots + \frac{n^{2n}}{(2n)!} \ge \left(1 - \frac{1}{n}\right)\mathrm{e}^n,$$

这正是所求结果。

验算

这些界是正确的，但相当宽松。当 $n = 2$ 时，第 (i) 部分的准确二项分布（binomial distribution）概率为 $\alpha \approx 0.868$，而界只是 0.5；当 $n = 10$ 时，$\alpha \approx 0.9986$，而界是 0.9。在第 (ii) 部分取 $n = 2$，左边为 $1 + 2 + 2 + \frac{4}{3} + \frac{2}{3} = 7$，而 $\frac{1}{2}\mathrm{e}^2 \approx 3.69$。我们已就第 (i) 部分中 n 由 2 至 30、第 (ii) 部分中 n 由 1 至 59 的每一个值，准确验证了两个不等式。

常见错误

- 选取 k 时，误用方差 $16n$ 代替标准差 $4\sqrt{n}$。
- 忘记切比雪夫不等式给出的是落在区间*以外*的概率的上界，因此要用补事件才能得到 α。
- 在第 (ii) 部分没有解释为何下尾 $Y < 0$ 没有任何贡献。

STEP 之外

切比雪夫不等式是集中不等式（concentration inequality）最简单的例子；把它应用于 m 次独立试验的平均值，便得到弱大数定律（weak law of large numbers）：离散程度以 $1/\sqrt{m}$ 的速度收窄。第 (ii) 部分还告诉我们，指数级数（exponential series）在 $n^{2n}/(2n)!$ 之后的尾项之和至多为 e^n/n——这是一个用概率证明的分析学事实。更精细的工具，例如切尔诺夫界（Chernoff bounds），可以证明真正的概率与 1 的差距是指数级地小。如欲深入了解，可参阅 Grimmett 与 Welsh 的 *Probability: An Introduction*（《概率论导引》）。

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。