详细解答

STEP I 1996，第 13 题

概率与统计 · 离散分布与期望值

STEP I 1996，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/96-s1-q13

本题在 *Beyond STEP* 第 12.11 节邮票、杯子与 e 中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch12/stamps-and-e.html

关键思路

游戏在第一次猜错时结束，所以 $N = k$ 的意思是“用 $2, 3, \ldots, k-1$ 个杯时都猜中，然后用 k 个杯时猜错”。把这些概率（probability）相乘，分母便会出现阶乘（factorial）；而求和式中出现阶乘，正强烈暗示指数级数（exponential series）$\mathrm{e}^x = \sum x^n/n!$ 即将登场。计算期望值的窍门，是把 k 的因子与 $k!$ 约去，使每个和都变成 $\sum 1/n!$ 平移后的版本。

解答

N 的分布

把 c 个杯打乱之后，Jim 实际上是随机乱猜，所以他猜中的概率是 $1/c$，猜错的概率是 $1 - 1/c$。每一回合都重新打乱一次，所以各回合互相独立（independent）。

由于杯的数目至少有两个，故 $N \geq 2$。对于 $k \geq 2$，事件 $N = k$ 发生，当且仅当 Jim 用 $2, 3, \ldots, k-1$ 个杯时都猜中，然后用 k 个杯时猜错。由独立性，

$$\begin{aligned} \mathrm{P}(N = k) &= \frac{1}{2} \cdot \frac{1}{3} \cdots \frac{1}{k-1} \cdot \left(1 - \frac{1}{k}\right) = \frac{1}{(k-1)!} \cdot \frac{k-1}{k} \\ &= \frac{k-1}{k!}, \qquad k = 2, 3, 4, \ldots \end{aligned}$$

（当 $k = 2$ 时，猜中部分的乘积是空积，我们得到 $\frac{1}{2}$，正如所料。）

值得留意的是，这个结果有一个简洁的等价形式：$N > k$ 表示 Jim 用 $2, \ldots, k$ 个杯时都猜

中，所以

$$\mathrm{P}(N > k) = \frac{1}{k!} \qquad (k \geq 1).$$

事实上 $\frac{1}{(k-1)!} - \frac{1}{k!} = \frac{k-1}{k!}$，与上面的公式吻合。当 $k \to \infty$ 时，这个概率趋向 0，所以游戏以概率 1 结束，而各概率之和为 1。

期望值

把 k 与 $k!$ 约去，

$$\begin{aligned}\mathrm{E}(N) &= \sum_{k=2}^{\infty} k \cdot \frac{k-1}{k!} = \sum_{k=2}^{\infty} \frac{k(k-1)}{k!} = \sum_{k=2}^{\infty} \frac{1}{(k-2)!} \\ &= \sum_{j=0}^{\infty} \frac{1}{j!} = \mathrm{e},\end{aligned}$$

其中设 $j = k-2$，并在 $x = 1$ 处使用 $\mathrm{e}^x = \sum_{j\geq 0} x^j/j!$。这就是所要求的结果。

方差

> **想一想** 这里哪一个二阶矩（second moment）最容易求和：$\mathrm{E}(N^2)$ 还是 $\mathrm{E}\big(N(N-1)\big)$？

两者皆可，因为 $k^2\,\mathrm{P}(N=k) = k \cdot \dfrac{k(k-1)}{k!}$ 中的因子 $k(k-1)$ 本身已能干净利落地约去。我们求得

$$\begin{aligned}\mathrm{E}(N^2) &= \sum_{k=2}^{\infty} \frac{k^2(k-1)}{k!} = \sum_{k=2}^{\infty} \frac{k}{(k-2)!} = \sum_{j=0}^{\infty} \frac{j+2}{j!} \\ &= \sum_{j=1}^{\infty} \frac{1}{(j-1)!} + 2\sum_{j=0}^{\infty} \frac{1}{j!} = \mathrm{e} + 2\mathrm{e} = 3\mathrm{e},\end{aligned}$$

其中在中间一步，$\sum j/j!$ 的 $j=0$ 项为零，而当 $j \geq 1$ 时 $j/j! = 1/(j-1)!$。因此

$$\mathrm{Var}(N) = \mathrm{E}(N^2) - \big(\mathrm{E}(N)\big)^2 = 3\mathrm{e} - \mathrm{e}^2 = \mathrm{e}(3-\mathrm{e}) \approx 0.766.$$

验算

- 各概率 $\frac{1}{2}, \frac{1}{3}, \frac{1}{8}, \frac{1}{30}, \ldots$ 之和为 1，这可由尾部形式 $\mathrm{P}(N > k) = 1/k!$ 看出。
- 利用尾部公式，$\mathrm{E}(N) = \sum_{k\geq 0} \mathrm{P}(N > k) = 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots = \mathrm{e}$，这是得出同一答案的另一条独立途径。
- 方差（variance）为正数，这是必然的，因为 $\mathrm{e} < 3$。精确求和得 $\mathrm{E}(N) = 2.71828\ldots$ 及 $\mathrm{Var}(N) = 0.76579\ldots$，而模拟 400 000 局游戏所得的平均值为 2.720，方差为 0.767。

常见错误

- 把分布由 $k = 1$ 开始：游戏一开始已有两个杯，所以 $N \geq 2$。
- 漏掉代表猜错的最后一个因子 $1 - \frac{1}{k}$，结果得出 $1/(k-1)!$，而这些“概率”之和是 $\mathrm{e} - 1$，而不是 1。
- 写成 $\mathrm{Var}(N) = \mathrm{E}(N^2) - \mathrm{E}(N)$，而没有减去 $\big(\mathrm{E}(N)\big)^2$。

STEP 之外

尾和公式（tail-sum formula）$\mathrm{E}(N) = \sum_{k \geq 0} \mathrm{P}(N > k)$ 对任何取值于 $\{0, 1, 2, \dots\}$ 的随机变量（random variable）都成立，是大学一年级的标准工具，往往能把繁复的期望值计算化为一行算式，正如本题所见。要把所有矩一并整理起来，更简洁的方法是使用概率生成函数（probability generating function）$G(s) = \mathrm{E}(s^N)$。在本题中可算出 $G(s) = 1 + (s-1)\mathrm{e}^s$，由此 $G'(1) = \mathrm{e}$ 及 $G''(1) = 2\mathrm{e}$ 立即给出平均值和方差。Grimmett 与 Welsh 所著的 *Probability: An Introduction*（概率论入门）正是以这种精神阐述生成函数。

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。