详细解答

STEP I 2017，第 12 题

概率与统计 · 期望值与条件化

STEP I 2017，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/17-s1-q12

关键思路

主办者收取的款项永远相同，都是 Nc；唯一的不确定之处，在于究竟需不需要派发奖金。因此期望利润（expected profit）等于收入减去 J 乘以“有人选中中奖号码”的概率，而整个问题最终归结为其互补事件（complementary event），即没有人选中。这个概率是一个幂 $(1-p)^N$，而近似式 $(*)$ 正是为处理这类式子而设的。

解答

第 (i) 部分

设 W 为中奖号码，并设 E 为没有参加者选中 W 的事件。无论 W 结果是甚么，每位参加者都独立（independently）地以概率 $1-\frac{1}{N}$ 避开它，所以

$$\mathrm{P}(E)=\left(1-\frac{1}{N}\right)^N.$$

在任何情况下，主办者都收到 Nc。若 E 发生，则无人中奖，利润为 Nc；否则奖金由中奖者瓜分，但不论中奖者有多少人，派出的总额都是 J，所以利润为 $Nc-J$。因此

$$\text{期望利润}=Nc-J\,\mathrm{P}(E^c)=Nc-J\left[1-\left(1-\frac{1}{N}\right)^N\right].$$

在 $(*)$ 中取 $a=1$，得 $(1-\frac{1}{N})^N\approx \mathrm{e}^{-1}$，所以期望利润约为 $Nc-J(1-\mathrm{e}^{-1})$。若 $J=2Nc$，则变为

$$Nc-2Nc\left(1-\mathrm{e}^{-1}\right)=Nc\left(\frac{2}{\mathrm{e}}-1\right).$$

由于 $\mathrm{e}>2$，我们有 $2/\mathrm{e}<1$，所以此值为负：主办者预期会亏蚀。

第 (ii) 部分

共有 γN 个热门号码和 $(1-\gamma)N$ 个冷门号码。每位参加者恰好选一个号码，所以全部 N 个号码的概率之和为 1：

$$\gamma N \cdot \frac{a}{N} + (1-\gamma)N \cdot \frac{b}{N} = 1 \quad \Longrightarrow \quad \gamma a + (1-\gamma)b = 1.$$

主办者仍然均匀地（uniformly）选取 W，所以 W 是热门号码的概率为 γ，是冷门号码的概率为 $1-\gamma$。

想一想 参加者不再均匀地选号。这会否改变我们求 $\mathrm{P}(E)$ 的方法?

只会通过避开 W 的概率产生影响，而这个概率现在取决于 W 属于哪一类号码。因此我们以此作条件（condition）。若 W 是热门号码，每位参加者独立地以概率 $1-\frac{a}{N}$ 避开它；若是冷门号码，则以概率 $1-\frac{b}{N}$ 避开。由全概率定律（law of total probability）及 $(*)$，

$$\mathrm{P}(E) = \gamma\left(1-\frac{a}{N}\right)^N + (1-\gamma)\left(1-\frac{b}{N}\right)^N \approx \gamma\,\mathrm{e}^{-a} + (1-\gamma)\,\mathrm{e}^{-b}.$$

与第 (i) 部分一样，期望利润为 $Nc - J(1-\mathrm{P}(E))$，所以

$$\text{期望利润} \approx \gamma J\,\mathrm{e}^{-a} + (1-\gamma)J\,\mathrm{e}^{-b} + (Nc - J),$$

这正是所要求的形式，其中

$$A = \gamma J, \qquad B = (1-\gamma)J, \qquad C = Nc - J.$$

情况 $\gamma = \frac{1}{8}$，$a = 9b$。由上述关系得 $\frac{1}{8}\cdot 9b + \frac{7}{8}b = 1$，即 $2b = 1$，所以

$$b = \tfrac{1}{2}, \qquad a = \tfrac{9}{2}.$$

当 $J = 2Nc$ 时，我们有 $A = \frac{1}{4}Nc$、$B = \frac{7}{4}Nc$ 及 $C = -Nc$，所以期望利润约为

$$Nc\left(\tfrac{1}{4}\mathrm{e}^{-9/2} + \tfrac{7}{4}\mathrm{e}^{-1/2} - 1\right).$$

括号内第一项为正，所以只需证明 $\frac{7}{4}\mathrm{e}^{-1/2} > 1$，即 $\sqrt{\mathrm{e}} < \frac{7}{4}$。由于 $\mathrm{e} < 3$，

$$\sqrt{\mathrm{e}} < \sqrt{3} < \tfrac{7}{4},$$

其中最后一步成立，是因为两边皆为正，且 $3 < \frac{49}{16}$（即 $48 < 49$）。因此括号内的值为正，主办者预期会获利。奖金与第 (i) 部分相同，但由于中奖号码往往是冷门号码，奖金无人认领的情况更为常见。

验算

- 当 $J = 2Nc$ 时，第 (i) 部分给出的期望利润约为 $(2/\mathrm{e} - 1)Nc \approx -0.264\,Nc$，而最后一部分给出 $\left(\frac{1}{4}\mathrm{e}^{-4.5} + \frac{7}{4}\mathrm{e}^{-0.5} - 1\right)Nc \approx 0.064\,Nc$。正如所述，是一笔小额利润。
- 当 $N = 800$ 时，用精确的幂 $(1 - a/N)^N$ 计算得 $0.0640\,Nc$，可见近似式 $(*)$ 在此非常准确。
- 若 $\gamma = 1$（每个号码都是“热门”的），上述关系迫使 $a = 1$，而第 (ii) 部分的公式便化为第 (i) 部分的 $J\mathrm{e}^{-1} + Nc - J$。
- 以电脑对 $N = 4$ 作完全枚举（参加者选号的全部 4^4 种方式），结果与两条未经近似的公式皆吻合。

常见错误

- 纠结于奖金如何分配。只要至少有一人中奖，不论中奖人数多少，主办者派出的总额都是 J。
- 误以为中奖号码是热门号码的概率取决于 a 或 b。主办者是均匀选号的，所以它是热门号码的概率为 γ。
- 把关系式写成 $a + b = 1$ 或 $\gamma a + b = 1$，忘记计算每类号码各有多少个。

STEP 之外

选中中奖号码的参加者人数服从参数为 N 和 p 的二项分布（binomial distribution），而 $(*)$ 正是理解以下事实的第一步：当 N 很大而 Np 固定时，它近似于泊松分布（Poisson distribution）——$(1-p)^N \to \mathrm{e}^{-Np}$ 不过就是泊松分布取值为零的概率。现实中设有奖金累积（roll-over）的彩票正是以这种方式分析的，而选择冷门号码的玩家确实有较高的期望回报。Grimmett 与 Welsh 的 *Probability: An Introduction*（《概率导论》）一书，详细发展了泊松极限，以及类似本题所用的条件论证。

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。