详细解答

STEP I 2002，第 14 题

概率与统计 · 概率中的等比级数

STEP I 2002，第 14 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/02-s1-q14

关键思路

以我总共按了多少个号码作为条件 (conditioning)。这个数目一旦确定，警钟响起与我取得款项之间相隔的时间也随之确定，于是警察先到的概率就是对应该段时间的 $1-\mathrm{e}^{-\lambda t}$。再对尝试次数取平均，便得出答案。在两种方案中，这个平均都会化为以 $q=\mathrm{e}^{-\lambda/r}$ 为公比的等比级数（geometric series），所以值得及早为 q 命名。

解答

顺序按键

设 K 为我的号码在 $1,2,\ldots,n$ 这列数中的位置。那么 K 就是我尝试的次数，而且对 $k=1,\ldots,n$ 有 $P(K=k)=\frac{1}{n}$。

假设 $K=k$。我每分钟按 r 个号码，所以相邻两个号码相隔 $1/r$ 分钟。时间由我按下第一个号码的一刻起计。

- 若 $k=1$，第一个号码就是正确的，我从未输入过错误号码，警察也不会接到通知。警察先到的概率是 $0=1-\mathrm{e}^{0}$。
- 若 $k\geq 2$，第一个号码是错的，所以警钟在时间 0 响起，而正确号码是第 k 个，在时间 $(k-1)/r$ 输入。如果警察在 $t=(k-1)/r$ 分钟内到达，他们就会在我取得款项之前到达，其概率为 $1-\mathrm{e}^{-\lambda(k-1)/r}$。

因此公式 $1-\mathrm{e}^{-\lambda(k-1)/r}$ 适用于所有 k，包括 $k=1$。设 A 为“警察在我取得款项之前到

达”这个事件。根据全概率公式（law of total probability），

$$P(A) = \sum_{k=1}^{n} P(K=k)\,P(A \mid K=k) = \sum_{k=1}^{n} \frac{1-\mathrm{e}^{-\lambda(k-1)/r}}{n},$$

正如题目所求。

为了化简，设 $q = \mathrm{e}^{-\lambda/r}$。由于 $\lambda, r > 0$，我们有 $0 < q < 1$，而且 $\mathrm{e}^{-\lambda(k-1)/r} = q^{k-1}$。所以

$$\begin{aligned} P(A) &= 1 - \frac{1}{n}\sum_{k=1}^{n} q^{k-1} = 1 - \frac{1-q^n}{n(1-q)} \\ &= 1 - \frac{1-\mathrm{e}^{-\lambda n/r}}{n\left(1-\mathrm{e}^{-\lambda/r}\right)}, \end{aligned}$$

这里用了等比级数求和，因为 $q \neq 1$，所以这一步成立。

随机按键

现在每次尝试都是从 $1, \ldots, n$ 中均匀地重新抽出一个号码，与之前的尝试独立（independent）（我不记得自己试过哪些号码）。所以每次尝试正确的概率都是 $\frac{1}{n}$，而且各次互相独立。若 N 是直至按中正确号码为止（包括该次）的尝试次数，则 N 服从几何分布（geometric distribution）：

$$P(N=k) = \frac{1}{n}\left(\frac{n-1}{n}\right)^{k-1}, \qquad k = 1, 2, 3, \ldots$$

想一想 当号码可以重复时，关于时间的论证是否仍然成立？

成立。警察在第一个错误号码出现时接到通知，而若 $N = k \geq 2$，那正是第一次尝试，发生在时间 0。款项在第 k 次尝试时到手，即时间 $(k-1)/r$。重复号码只会改变 N 的分布，不会改变时间的计算。所以与之前完全一样，$P(A \mid N=k) = 1 - q^{k-1}$，而这对 $k = 1$ 也是对的。

记 $p = \frac{n-1}{n}$。那么

$$\begin{aligned} P(A) &= \sum_{k=1}^{\infty} \frac{1}{n} p^{k-1}\left(1-q^{k-1}\right) = \frac{1}{n}\sum_{k=1}^{\infty} p^{k-1} - \frac{1}{n}\sum_{k=1}^{\infty} (pq)^{k-1} \\ &= \frac{1}{n}\cdot\frac{1}{1-p} - \frac{1}{n}\cdot\frac{1}{1-pq}. \end{aligned}$$

把和式拆开是可以的，因为两个级数都收敛（converge）：$0 \leq p < 1$ 而且 $0 < pq < 1$。第一项是 $\frac{1}{n}\cdot n = 1$，这只不过是说 N 必定是有限的。至于第二项，

$$n(1-pq) = n - (n-1)q,$$

所以

$$P(A) = 1 - \frac{1}{n - (n-1)\mathrm{e}^{-\lambda/r}},$$

正如题目所求。

验算

- 当 $n = 1$ 时，永远不会出现错误号码，而两条公式都给出 0：第一条是因为 $1 - q^1 = 1 - q$，第二条是因为分母等于 1。
- 当 $\lambda \to \infty$（警察即时到达）时，$q \to 0$，两个答案都趋向 $1 - \frac{1}{n}$。这正是第一次尝试出错的概率，理应如此。
- 随机按键应该较差。取 $n = 5$、$\lambda = 0.7$、$r = 3$，顺序按键得 0.3382，随机按键得 0.4543。两种方案各做 2×10^5 次模拟试验，结果分别是 0.339 和 0.456。

常见错误

- 把时间写成 k/r 而不是 $(k-1)/r$。计时由第一个（错误的）号码开始，所以第 k 个号码是在 $k-1$ 个间隔之后出现。
- 在随机的情况下，给每个 k 的概率都是 $\frac{1}{n}$，或者求和只加到 n 为止。既然容许重复，N 服从几何分布，而且没有上界。

STEP 之外

这里警察的到达时间是一个指数随机变量（exponential random variable），而整个计算就是对随机的 X 求 $P(T < X)$，方法是以 X 为条件。当 X 服从几何分布时，答案有简洁的闭合形式（closed form），因为 $E\left(q^X\right)$ 正是 X 的概率生成函数（probability generating function）在 q 处的值。事实上，两个答案都是 $1 - E\left(q^{N-1}\right)$。生成函数和指数分布是大学一年级概率论的核心内容。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门）对两者都有很好的讲解。

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。