详细解答

STEP I 2003，第 14 题

概率与统计 · 泊松分布

STEP I 2003，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/03-s1-q14

关键思路

Jane 见到的朋友数目并不是来访的人数，而是这个人数以 2 为上限后的值。一个设有上限的变量只需考虑三种情况：无人来访、恰好一人来访、两人或以上来访。因此我们根本不必把整个泊松（Poisson）级数加起来，只需求出 $P(0)$ 和 $P(1)$，而“两人或以上”的概率就是余下的部分。至于新朋友，我们按其是否来访作条件分析（conditioning）。

解答

原有的朋友

设 N 为某一天来访的朋友数目，则 $N \sim \mathrm{Po}(2)$，且

$$P(N=k)=\frac{\mathrm{e}^{-2}2^k}{k!}, \qquad k=0,1,2,\ldots$$

Jane 见到 $S=\min(N,2)$ 位朋友。所以当 $N=0$ 时 $S=0$，当 $N=1$ 时 $S=1$，当 $N\geq 2$ 时 $S=2$。现在

$$P(N=0)=\mathrm{e}^{-2}, \qquad P(N=1)=2\mathrm{e}^{-2}, \qquad P(N\geq 2)=1-3\mathrm{e}^{-2},$$

因此

$$\mathrm{E}(S)=0\cdot\mathrm{e}^{-2}+1\cdot 2\mathrm{e}^{-2}+2\left(1-3\mathrm{e}^{-2}\right)=2-4\mathrm{e}^{-2},$$

这正是所要证明的结果。

同一个计算还有一个更简洁的写法，下文会再次用到。对任何取值于 $\{0,1,2,\ldots\}$ 的随机变量（random variable）X，

$$\mathrm{E}(\min(X,2))=2-2P(X=0)-P(X=1),$$

因为 $\min(X,2)$ 与 2 的差距，在 $X=0$ 时是 2，在 $X=1$ 时是 1，其余情况则为零。取 $X=N$，便再次得到 $2-2\mathrm{e}^{-2}-2\mathrm{e}^{-2}$。

加入新朋友之后

若新朋友来访，设 $B=1$，否则设 $B=0$，于是 $P(B=1)=p$。题目指出 B 与 N 独立（independent）。现在来访的总人数是 $M=N+B$，而 Jane 见到 $\min(M,2)$ 位朋友。

> **想一想** 我们需要 M 的完整分布吗?

不需要。根据上面的恒等式，我们只需要 $P(M=0)$ 和 $P(M=1)$。利用独立性把概率相乘：

$$\begin{aligned}P(M=0)&=P(N=0)\,P(B=0)=(1-p)\mathrm{e}^{-2},\\P(M=1)&=P(N=1)\,P(B=0)+P(N=0)\,P(B=1)=2(1-p)\mathrm{e}^{-2}+p\,\mathrm{e}^{-2}.\end{aligned}$$

因此

$$\begin{aligned}\mathrm{E}(\min(M,2))&=2-2(1-p)\mathrm{e}^{-2}-2(1-p)\mathrm{e}^{-2}-p\,\mathrm{e}^{-2}\\&=2-(4-3p)\mathrm{e}^{-2}.\end{aligned}$$

另一个做法是以 B 为条件。若新朋友没有来访（概率为 $1-p$），平均值一如之前，是 $2-4\mathrm{e}^{-2}$。若新朋友来访（概率为 p），Jane 见到 $\min(N+1,2)$ 位朋友，这个数在 $N=0$ 时是 1，其余情况是 2，其平均值为 $\mathrm{e}^{-2}+2(1-\mathrm{e}^{-2})=2-\mathrm{e}^{-2}$。所以平均值是

$$(1-p)(2-4\mathrm{e}^{-2})+p(2-\mathrm{e}^{-2})=2-(4-3p)\mathrm{e}^{-2}.$$

所以 Jane 现在平均每天见到 $2-(4-3p)\mathrm{e}^{-2}$ 位朋友，比之前多了 $3p\,\mathrm{e}^{-2}$。

验算

- 当 $p=0$ 时，新朋友永不来访，我们重新得到 $2-4\mathrm{e}^{-2}\approx 1.4587$。
- 当 $p=1$ 时，我们得到 $2-\mathrm{e}^{-2}\approx 1.8647$，这正是 $\min(N+1,2)$ 的平均值，与预期相符。
- 答案随 p 递增，而且恒小于 2，这是合理的。
- 按精确分布对 $\min(k+b,2)$ 求和，以及取 $p=0.3$ 模拟 4×10^5 天，两者都得出约 1.580，与 $2-3.1\mathrm{e}^{-2}$ 吻合。

常见错误

- 忘记了最多只见两位朋友的上限，答成 $\mathrm{E}(N) = 2$ 或 $2 + p$。
- 把 $N + B$ 当作泊松变量。伯努利（Bernoulli）变量并非泊松变量，所以两者之和也不是泊松变量。
- 恰好一人来访有两种情况（一位旧朋友来访而新朋友没来，或没有旧朋友来访而新朋友来了），却漏掉其中一种。

STEP 之外

恒等式 $\mathrm{E}(\min(X, 2)) = P(X \geq 1) + P(X \geq 2)$ 是尾和公式（tail-sum formula）$\mathrm{E}(X) = \sum_{k \geq 1} P(X \geq k)$ 的特殊情况，该公式适用于取非负整数值的随机变量。它是对生存函数（survival function）积分的离散版本。这类设有上限的平均值常见于排队模型和存货模型，在容量有限的情况下，它们给出获得服务的顾客数目的期望值（expected value）。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门教材）涵盖尾和公式和条件期望（conditional expectation），正是本题所用的两件工具。

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。