详细解答

STEP I 1988，第 14 题

概率与统计 · 删失正态分布的期望值

STEP I 1988，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/88-s1-q14

关键思路

变量 X_M 其实就是 $\min(X, M)$：把正态变量（normal variable）中所有高于 M 的部分一律压低到 M。它的期望值（expectation）可拆成两部分：一个在 $x < M$ 上的积分，以及一个常数乘以尾部概率（tail probability）；而这个积分很容易计算，因为 $x\phi(x)$ 除了正负号之外，正是 $\phi(x)$ 的导数。游客的故事只是同一个想法换了个包装。一队巴士所接载的乘客人数是 $\min(\text{需求}, \text{载客量})$，经正态近似（normal approximation）之后，这正好就是一个 X_M。

解答

X_M 的期望值

由于在 $\{X < M\}$ 上 $X_M = X$，而在 $\{X \geq M\}$ 上 $X_M = M$，

$$\mathrm{E}(X_M) = \int_{-\infty}^{M} x\,\phi(x)\,\mathrm{d}x \;+\; M\,\mathrm{P}(X \geq M) = \int_{-\infty}^{M} x\,\phi(x)\,\mathrm{d}x \;+\; M(1-\Phi(M)).$$

? **想一想** 有没有一个函数，其导数是 $x\phi(x)$?

有。由 $\phi(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}$，按链式法则（chain rule）可得 $\phi'(x) = -x\phi(x)$。因此

$$\int_{-\infty}^{M} x\,\phi(x)\,\mathrm{d}x = \Big[-\phi(x)\Big]_{-\infty}^{M} = -\phi(M) + 0,$$

因为当 $x \to -\infty$ 时 $\phi(x) \to 0$。故此

$$\mathrm{E}(X_M) = -\phi(M) + M(1-\Phi(M)),$$

这正是所求的结果。

作为合理性检查，令 $M \to \infty$ 可得 $\mathrm{E}(X_M) \to 0 = \mathrm{E}(X)$，因为 $M(1-\Phi(M)) \to 0$。令 $M \to -\infty$ 则得 $\mathrm{E}(X_M) \approx M$，与预期相符。

巴士

设 N 为某一次到访中被指往巴士站的游客人数。我们假设 1024 名游客各自独立地被指路，往两边的概率各为 $\frac{1}{2}$。于是

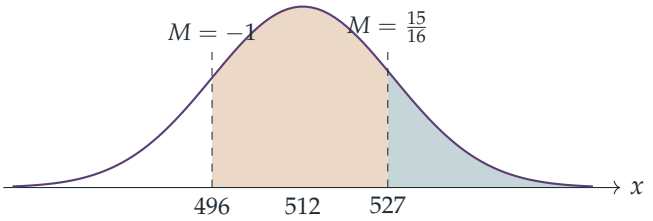
$$N \sim \mathrm{B}(1024, \tfrac{1}{2}), \qquad \mathrm{E}(N) = 512, \qquad \mathrm{Var}(N) = 1024 \cdot \tfrac{1}{2} \cdot \tfrac{1}{2} = 256 = 16^2.$$

由于 $n = 1024$ 这么大，N 可以非常准确地用 $512 + 16X$ 来近似，其中 X 服从标准正态分布（standard normal distribution）。

有 16 辆巴士时，载客量是 $16 \times 31 = 496$。有 17 辆时则是 527。载客量为 C 的车队接载 $\min(N, C)$ 名乘客，所以它每次到访的利润（以 vloska 计）是

$$\begin{aligned}\min(N,C) &\approx \min(512 + 16X,\ C)\\ &= 512 + 16\min\Big(X, \frac{C-512}{16}\Big) = 512 + 16\,X_M, \qquad \text{其中} M = \frac{C-512}{16}.\end{aligned}$$

因此 $C = 496$ 对应 $M = -1$，而 $C = 527$ 对应 $M = \frac{15}{16}$。



当需求落在琥珀色的区域时，第 17 辆巴士接载的是超出 496 的那部分乘客。当需求落在蓝绿色的尾部时，它载满 31 人。

利用 $\phi(1) \approx 0.24197$ 和 $\Phi(1) \approx 0.84134$，并留意 $1 - \Phi(-1) = \Phi(1)$，

$$\mathrm{E}(X_{-1}) = -\phi(1) - \big(1 - \Phi(-1)\big) = -\phi(1) - \Phi(1) \approx -0.24197 - 0.84134 = -1.08332.$$

利用 $\phi(0.9375) \approx 0.25707$ 和 $1 - \Phi(0.9375) \approx 0.17425$，

$$\mathrm{E}(X_{15/16}) \approx -0.25707 + 0.9375 \times 0.17425 \approx -0.25707 + 0.16336 = -0.09371.$$

因此，多一辆巴士令每次到访的期望利润增加

$$16\big(\mathrm{E}(X_{15/16}) - \mathrm{E}(X_{-1})\big) \approx 16 \times 0.98961 \approx 15.83 \text{ vloska}.$$

每年有 50 次到访，所以多一辆巴士的价值大约是

$$50 \times 15.83 \approx 792 \text{ vloska（每年）}.$$

这就是值得考虑的每年贿款上限。付出的金额一旦超过期望的额外利润，合作社平均而言就会得不偿失。

> **想一想** 既然 N 是离散的，我们不是应该使用连续性校正（continuity correction）吗？

这里并不需要。连续性校正的用处，在于近似 $\mathrm{P}(N \le c)$ 这类概率，因为这时被取平均的函数在整数处有跳跃。函数 $k \mapsto \min(k, C)$ 没有跳跃，不作校正的正态近似对它已经极为准确。以二项分布（binomial distribution）作精确计算，结果是 791.68，与上面的数字吻合至小数点后两位。把载客量移动 $\frac{1}{2}$ 反而会令结果变差，得出 775。

验算

- 以中点法则作数值积分（midpoint quadrature）计算 $\int_{-\infty}^{M} x\phi(x)\,\mathrm{d}x + M(1-\Phi(M))$，对多个 M 值所得的结果都与公式吻合至小数点后八位。以 4×10^5 个样本对 $\min(X, M)$ 进行蒙地卡罗模拟（Monte Carlo simulation），结果亦吻合。
- 对 $\mathrm{B}(1024, \frac{1}{2})$ 作精确求和，得出每次到访的额外期望利润为 15.8337，即每年 791.68，与正态近似相符。
- 粗略的合理性检查：只要 $N > 496$，多出的那辆巴士就有收入，这种情况出现的概率约为 0.84，而且它往往满座。因此每次到访的数字比 31 的一半略多，是合理的。

常见错误

- 忘记换算比例：$\min(N, C)$ 是 $512 + 16X_M$，所以 $\mathrm{E}(X_M)$ 之差必须先乘以标准差（standard deviation）16，再乘以 50。
- $M = -1$ 时的正负号错误：$M(1-\Phi(M))$ 一项是 $-(1-\Phi(-1)) = -\Phi(1)$，而不是 $-(1-\Phi(1))$。

STEP 之外

$\mathrm{E}\min(X, M)$ 这个量是一个删失（censored）正态变量的期望值。与它密切相关的表达式，例如 $\mathrm{E}(X-M)^+ = \phi(M) - M(1-\Phi(M))$，是存货理论（inventory theory，即“报贩问题”（newsvendor problem））以及布莱克–舒尔斯（Black–Scholes）期权定价的骨

干。恒等式 $\mathrm{E}\min(N, C+1) - \mathrm{E}\min(N, C) = \mathrm{P}(N > C)$ 提供了一个简洁的方法，让人看清每多一个单位载客量的价值。想进一步了解这些概念，可参阅 Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门教材）。

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。