详细解答

STEP I 1989，第 15 题

概率与统计 • 分布的比较

STEP I 1989，第 15 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/89-s1-q15

关键思路

以 $12\,\mathrm{km\,h^{-1}}$ 的速率行驶，每公里刚好需时 5 分钟，所以每条路线的行程时间都是一个固定的分钟数，再加上某个随机变量（random variable）的简单倍数：这个变量分别服从均匀分布（uniform distribution）、泊松分布（Poisson distribution）或二项分布（binomial distribution）。于是，每一个关于时间的问题，都变成关于这一个变量的问题。中间的一问则是贝叶斯定理（Bayes′ theorem）的应用，其中三条路线被选中的可能性相等。

解答

行程时间

设 T_A、T_B、T_D 为三条路线的行程时间，以分钟计。由于 $12\,\mathrm{km\,h^{-1}}$ 即每 5 分钟行 1 km：

- Angle Avenue：$T_A = 25 + A$，其中 $A \sim \mathrm{U}(0, 10)$。
- Bend Boulevard：$T_B = 20 + 3B$，其中 $B \sim \mathrm{Po}(4)$。
- Detour Drive：设 X 为绕道的次数。假设五处道路工程互相独立（independent），每处导致绕道的概率均为 $\frac{4}{5}$，则 $X \sim \mathrm{B}(5, \frac{4}{5})$，路程为 $2 + X$ km，而 $T_D = 10 + 5X$。

至少 27 分钟

Angle Avenue。$T_A \geq 27 \iff A \geq 2$，所以

$$\mathrm{P}(T_A \geq 27) = \frac{10 - 2}{10} = \frac{4}{5}.$$

Bend Boulevard。$T_B \geq 27 \iff 3B \geq 7 \iff B \geq 3$，因为 B 是整数。所以

$$\mathrm{P}(T_B \geq 27) = 1 - \mathrm{e}^{-4}\left(1 + 4 + \frac{4^2}{2!}\right) = 1 - 13\mathrm{e}^{-4} \approx 0.7619.$$

Detour Drive。$T_D \geq 27 \iff 5X \geq 17 \iff X \geq 4$。所以

$$\mathrm{P}(T_D \geq 27) = 5\left(\frac{4}{5}\right)^4\frac{1}{5} + \left(\frac{4}{5}\right)^5 = \left(\frac{4}{5}\right)^4\left(1 + \frac{4}{5}\right) = \frac{256}{625}\cdot\frac{9}{5} = \frac{2304}{3125} = 0.73728.$$

已知行程较长，走的是哪条路线?

设 R 为所选的路线，其中 $\mathrm{P}(R = A) = \mathrm{P}(R = B) = \mathrm{P}(R = D) = \frac{1}{3}$，并设 L 为行程时间至少 27 分钟这一事件。根据贝叶斯定理，各个 $\frac{1}{3}$ 因子互相约去：

$$\begin{aligned}\mathrm{P}(R = B \mid L) &= \frac{\mathrm{P}(L \mid B)}{\mathrm{P}(L \mid A) + \mathrm{P}(L \mid B) + \mathrm{P}(L \mid D)} = \frac{1 - 13\mathrm{e}^{-4}}{\frac{4}{5} + 1 - 13\mathrm{e}^{-4} + \frac{2304}{3125}} \\ &= \frac{1 - 13\mathrm{e}^{-4}}{\frac{7929}{3125} - 13\mathrm{e}^{-4}} \approx \frac{0.7619}{2.2992} \approx 0.331.\end{aligned}$$

这个数值只比 $\frac{1}{3}$ 略高，十分合理：三条路线令行程达 27 分钟或以上的可能性几乎相同。

第 (i) 部分

利用 $\mathrm{E}(A) = 5$、$\mathrm{E}(B) = 4$ 和 $\mathrm{E}(X) = 5 \times \frac{4}{5} = 4$，以及期望值的线性性质（linearity of expectation），

$$\mathrm{E}(T_A) = 25 + 5 = 30, \qquad \mathrm{E}(T_B) = 20 + 3 \times 4 = 32, \qquad \mathrm{E}(T_D) = 10 + 5 \times 4 = 30.$$

? **想一想** 有两条路线打成平手。这是计错了吗?

不是。Angle Avenue 和 Detour Drive 的平均时间都刚好是 30 分钟，而 Bend Boulevard 的平均时间是 32 分钟。所以要令平均时间尽量短，我应该选用 Angle Avenue 或 Detour Drive 其中任何一条：两者同样好，而且都胜过 Bend Boulevard。

第 (ii) 部分

现在比较各条路线的 $\mathrm{P}(T < 32)$。

- $T_A < 32 \iff A < 7$，所以 $\mathrm{P}(T_A < 32) = \frac{7}{10} = 0.7$。
- $T_B < 32 \iff 3B < 12 \iff B \leq 3$，所以

$$\mathrm{P}(T_B < 32) = \mathrm{e}^{-4}\left(1 + 4 + 8 + \frac{32}{3}\right) = \frac{71}{3}\mathrm{e}^{-4} \approx 0.4335.$$

- $T_D < 32 \iff 5X < 22 \iff X \leq 4$，所以

$$\mathrm{P}(T_D < 32) = 1 - \left(\frac{4}{5}\right)^5 = 1 - \frac{1024}{3125} = \frac{2101}{3125} \approx 0.6723.$$

由于 $0.7 > 0.6723 > 0.4335$，我应该每次都选用 **Angle Avenue**，这样有 70% 的日子可以在 32 分钟内到达。

验算

- 模拟 400 000 次行程，得出 $\mathrm{P}(T \geq 27) \approx 0.801,\ 0.761,\ 0.740$，平均时间 $\approx 30.0,\ 32.0,\ 30.0$ 分钟，$\mathrm{P}(T < 32) \approx 0.700,\ 0.435,\ 0.670$，以及 $\mathrm{P}(R = B \mid L) \approx 0.331$。全部与精确值吻合。
- 检查边界情况是否合理：在 Detour Drive 上，绕道 3 次需时 25 分钟，绕道 4 次则需时 30 分钟，所以“至少 27 分钟”确实等同于“至少绕道 4 次”。

常见错误

- 处理整数条件时不够严谨：$3B \geq 7$ 表示 $B \geq 3$，而不是 $B \geq 2$；$3B < 12$ 表示 $B \leq 3$，而不是 $B \leq 4$。
- 忘记 Detour Drive 本身有 2 km 的基本路程，或者忘记其他路线上每公里五分钟的骑车时间。
- 在 (i) 中声称只有一条路线胜出。其实两个平均值完全相等。

STEP 之外

第 (i) 和第 (ii) 部分说明，“最好”取决于我们要优化的是甚么：平均值只是概括一个分布的其中一个数字，两条平均值相同的路线，其离散程度和尾部概率可以相差很远。在决策理论（decision theory）中，这就是如何选择*损失函数*（loss function）的问题。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率入门教科书）涵盖了本题用到的各种分布和贝叶斯定理。Blitzstein 与 Hwang 的 *Introduction to Probability*（概率导论）则收录了许多这一类的题目。

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。