详细解答

STEP II 2021，第 11 题

概率与统计 • 条件概率的递推关系

STEP II 2021，第 11 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2021 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696588-2021-past-papers.zip

本题在 *Beyond STEP* 第 12.5 节最后一位乘客的座位中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch12/airline-seat.html

关键思路

以 T_1 坐在哪里为条件（condition on）。若 T_1 坐了自己的座位 S_1，其他人都不受影响；若 T_1 坐了 S_k，那么 $T_2, \ldots, T_{k-1}$ 都能安然入座，而 T_k 就成了新的“随机选座者”。关键的观察是：此时 T_k 的处境与当初的 T_1 完全相同，只是列车变小了，而空着的座位 S_1 则扮演 T_k 自己座位的角色。这样问题便转化为一条递推关系（recurrence）。

解答

全文中，“T_r 被挤走”指 T_r 到达时发现 S_r 已被人坐了。只有被挤走的乘客（或 T_1）才会随机选座。

第 (i) 部分

当 $n = 2$ 时，T_1 各以概率（probability）$\frac{1}{2}$ 坐 S_1 或 S_2，而 T_2 能坐到 S_2 恰好当 T_1 坐了 S_1。所以

$$P_2 = \frac{1}{2}.$$

当 $n = 3$ 时，以 T_1 的座位为条件，每种情况的概率均为 $\frac{1}{3}$：

- S_1：T_2 和 T_3 都发现自己的座位空着，所以 T_3 坐到 S_3。
- S_3：T_3 必定坐不到 S_3。

▶ S_2：T_2 被挤走，在 S_1 和 S_3 之间随机选择；T_3 坐到 S_3 恰好当 T_2 选了 S_1，概率为 $\frac{1}{2}$。

因此

$$P_3 = \tfrac{1}{3}\cdot 1 + \tfrac{1}{3}\cdot 0 + \tfrac{1}{3}\cdot\tfrac{1}{2} = \tfrac{1}{2}.$$

第 (ii) 部分

假设 T_1 坐在 S_k，其中 $2 \le k \le n-1$。那么 $T_2, \ldots, T_{k-1}$ 都发现自己的座位空着并坐下。当 T_k 到达时，空座位是

$$S_1,\ S_{k+1},\ S_{k+2},\ \ldots,\ S_n,$$

共 $n-k+1$ 个座位，而 T_k 在其中均匀随机地选择。尚未到达的乘客是 $T_{k+1}, \ldots, T_n$，各人都想坐自己的座位，而尚未到达的人之中没有人被分配到 S_1。

? **想一想** 为甚么这恰好就是有 $n-k+1$ 个座位的原问题?

重新标记：把 T_k 称为“第一位乘客”，并当作 S_1 是 T_k 获分配的座位。在有 m 个座位的原问题中，第一位乘客从全部 m 个座位中均匀地选择，之后每位乘客若自己的座位空着便坐下，否则从空座位中均匀地选择，而第一位乘客获分配的座位没有其他人想要。这里取 $m = n-k+1$，乘客 $T_k, T_{k+1}, \ldots, T_n$ 在座位 $S_1, S_{k+1}, \ldots, S_n$ 上正好遵循这些规则，而 S_1 没有其他人想要。这些规则除了用来判断乘客自己的座位是否空着之外，从不用到座位的标签，所以两个过程的概率完全相同。在较小的问题中，T_n 仍是最后一位乘客，所以

$$\mathrm{P}(T_n \text{ 坐在} S_n \mid T_1 \text{ 坐在} S_k) = P_{n-k+1}.$$

现在，对 $n \ge 3$，以 T_1 的座位为条件，$S_1, \ldots, S_n$ 每个的概率均为 $\frac{1}{n}$。若是 S_1，T_n 必定坐到 S_n；若是 S_n，T_n 必定坐不到。所以由全概率公式（law of total probability），

$$P_n = \frac{1}{n}\cdot 1 + \frac{1}{n}\sum_{k=2}^{n-1} P_{n-k+1} + \frac{1}{n}\cdot 0 = \frac{1}{n}\left(1 + \sum_{r=2}^{n-1} P_r\right),$$

其中我们代入了 $r = n-k+1$，当 k 由 2 变到 $n-1$ 时，它由 $n-1$ 递减至 2。这就是所求的结果。

第 (iii) 部分

我们断言对每个 $n \ge 2$ 都有 $P_n = \frac{1}{2}$。我们使用强归纳法（strong induction）。

基本情况。由第 (i) 部分得 $P_2 = \frac{1}{2}$。

归纳步骤。设 $n \ge 3$，并假设对 $r = 2, 3, \ldots, n-1$ 有 $P_r = \frac{1}{2}$。第 (ii) 部分中的和有 $n-2$ 项，每项都等于 $\frac{1}{2}$，所以

$$P_n = \frac{1}{n}\left(1 + \frac{n-2}{2}\right) = \frac{1}{n}\cdot\frac{n}{2} = \frac{1}{2}.$$

因此对 $r = 2, \ldots, n$ 有 $P_r = \frac{1}{2}$，由归纳法，对所有 $n \geq 2$ 都有 $P_n = \frac{1}{2}$。（第 (i) 部分的 $P_3 = \frac{1}{2}$ 与此吻合。）

第 (iv) 部分

当 $n = 2$ 时，$T_{n-1} = T_1$，他以概率 $\frac{1}{2}$ 坐在 S_1。所以 $Q_2 = \frac{1}{2}$。

现在设 $n \geq 3$，于是 T_{n-1} 不是 T_1，再次以 T_1 的座位 S_k 为条件：

- $k = 1$：每个人都坐到自己的座位，所以 T_{n-1} 坐到 S_{n-1}。
- $k = n$：$T_2, \ldots, T_{n-1}$ 都发现自己的座位空着，所以 T_{n-1} 同样坐到 S_{n-1}。
- $k = n-1$：T_{n-1} 必定坐不到 S_{n-1}。
- $2 \leq k \leq n-2$：与第 (ii) 部分完全一样，余下的旅程就是以 T_k 为首、有 $m = n-k+1$ 个座位的原问题。在这列较小的列车中，T_{n-1} 是倒数第二位乘客，所以概率为 Q_{n-k+1}。这里 m 取遍 $3, 4, \ldots, n-1$。

因此，对 $n \geq 3$，

$$Q_n = \frac{1}{n}\left(2 + \sum_{m=3}^{n-1} Q_m\right).$$

当 $n = 3$ 时和为空，故 $Q_3 = \frac{2}{3}$。若对 $3 \leq m \leq n-1$ 有 $Q_m = \frac{2}{3}$，则

$$Q_n = \frac{1}{n}\left(2 + \frac{2(n-3)}{3}\right) = \frac{1}{n} \cdot \frac{2n}{3} = \frac{2}{3},$$

所以由强归纳法，对每个 $n \geq 3$ 都有 $Q_n = \frac{2}{3}$。总结：

$$Q_2 = \tfrac{1}{2}, \qquad Q_n = \tfrac{2}{3} \quad (n \geq 3).$$

? **想一想** 有没有方法不用任何递推关系就看出 $\frac{1}{2}$ 和 $\frac{2}{3}$?

有的。每当有人随机选座时，S_1 与每个后来乘客仍然空着的座位都受到同等对待，所以在过程结束之前，S_1 和 S_n 两者中先被坐的那个是哪一个，可能性相等；T_n 成功恰好当 S_1 先被坐，得 $\frac{1}{2}$。对于 T_{n-1}（其中 $n \geq 3$），考虑 S_1, S_{n-1}, S_n 之中哪一个最先被随机选座者坐了；每个的可能性相等，而 T_{n-1} 只有在那个是 S_{n-1} 时才会被挤走，得 $1 - \frac{1}{3} = \frac{2}{3}$。

验算

- $n = 3$ 的情况可以用手列举：(T_1, T_2, T_3) 的结果是 (S_1, S_2, S_3)，概率为 $\frac{1}{3}$；(S_3, S_2, S_1)，概率为 $\frac{1}{3}$；以及 (S_2, S_1, S_3)、(S_2, S_3, S_1)，概率各为 $\frac{1}{6}$。所以 T_3 坐到 S_3 的概率为 $\frac{1}{2}$，T_2 坐到 S_2 的概率为 $\frac{2}{3}$，与所求得的一致。

- 用电脑对 $n = 2, \ldots, 8$ 精确列举每一种可能的旅程，得出 $P_n = \frac{1}{2}$ 全部成立、对 $n \geq 3$ 有 $Q_n = \frac{2}{3}$，并证实第 (ii) 部分的条件概率（conditional probability）对每个 k 都成立。以 $n = 20$ 模拟 10^5 次旅程，得到 0.500 和 0.668。

常见错误

- 忘记 T_1 本身可以选 S_1，这正是递推关系中“$1+$”的来源。
- 在第 (iv) 部分漏掉 T_1 坐了 S_n 的情况：这对 T_{n-1} 是好消息，所以常数是 2 而不是 1。
- 在 $n = 2$ 时套用 Q 的递推关系，但此时 T_{n-1} 就是 T_1，答案是 $\frac{1}{2}$ 而不是 $\frac{2}{3}$。

STEP 之外

这里的递推关系是一种首步分析（first-step analysis）：以第一次随机选择为条件，并认出同一问题的较小版本。同样的技巧可以求出马尔可夫链（Markov chain）的命中概率（hitting probability）和期望命中时间（expected hitting time），这是大学一年级概率论的核心课题。上面的对称性论证让人初尝一个选得好的“哪件事先发生?”问题如何取代冗长的计算。Grimmett 与 Welsh 的 *Probability: An Introduction*（《概率导论》）对这两种思路都有阐述。

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。