详细解答

STEP III 2003，第 12 题

概率与统计 · 指数分布的等候时间

STEP III 2003，第 12 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-s3-q12

关键思路

如果任何一段时间内的打扰次数都服从泊松分布（Poisson distribution），那么“下一次打扰在 t 分钟之后才出现”与“在接下来的 t 分钟内完全没有打扰”是同一个事件，这样就把离散分布（discrete distribution）变成连续分布（continuous distribution）：等候时间服从指数分布（exponential distribution）。其后，相继两次打扰之间的各个间隔互相独立（independent），所以寻找一个长间隔的过程，就是一连串独立试验，每次试验不是“成功”（间隔达半小时或以上）就是“失败”。余下的工作交给几何分布（geometric distribution）和条件期望（conditional expectation）便可。

解答

T 的分布

打扰平均每十分钟出现一次，所以发生率是每分钟 $\frac{1}{10}$ 次，而在长 t 分钟的时段内，打扰次数服从平均值为 $\frac{t}{10}$ 的泊松分布。

从刚刚发生的那次打扰开始计时。到下一次打扰的时间 T 不少于 t，当且仅当在接下来的 t 分钟内没有打扰，所以

$$\mathrm{P}(T \geq t) = \mathrm{P}(\text{没有任何打扰，为时}t) = \mathrm{e}^{-t/10}.$$

因此，对于 $t \geq 0$，工作时间少于 t 分钟的概率是

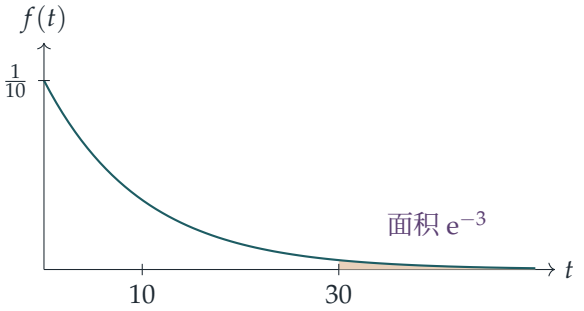
$$\mathrm{P}(T < t) = 1 - \mathrm{e}^{-t/10}.$$

这就是 T 的累积分布函数（cumulative distribution function）$F(t)$（当 $t < 0$ 时 $F(t) = 0$）。

求导，得

$$f(t) = F'(t) = \tfrac{1}{10}\,\mathrm{e}^{-t/10} \quad (t \geq 0), \qquad f(t) = 0 \quad (t < 0).$$

所以 T 服从平均值为 10 分钟的指数分布。



打扰次数的期望值

由于这个过程没有记忆（互不重叠的时段内的泊松计数互相独立），每当一次打扰发生，到下一次打扰的时间都重新开始计算，其分布与 T 相同，并且独立于之前发生的一切。所以相继的间隔 $T_1, T_2, \ldots$ 互相独立，各自的分布都与 T 相同。

如果一个间隔不少于 30 分钟，就称它为长间隔。每个间隔是长间隔的概率为

$$p = \mathrm{P}(T \geq 30) = \mathrm{e}^{-30/10} = \mathrm{e}^{-3},$$

而且各个间隔互相独立。设 N 为第一个长间隔出现之前的打扰次数（刚刚发生的那一次不计算在内）。每个短间隔都以一次打扰告终，而第一个长间隔就是我那不受打扰的半小时，所以 N 就是第一个长间隔之前的短间隔数目。因此

$$\mathrm{P}(N = n) = (1-p)^n p, \qquad n = 0, 1, 2, \ldots$$

? **想一想** 这个几何分布由 0 而不是由 1 开始，我们怎样求它的平均值?

对几何级数（geometric series）$\sum_{n\geq 0} x^n = \frac{1}{1-x}$ 求导，可得当 $|x| < 1$ 时 $\sum_{n\geq 1} n x^{n-1} = \frac{1}{(1-x)^2}$。取 $x = 1-p$，

$$\mathrm{E}(N) = \sum_{n=1}^{\infty} n(1-p)^n p = p(1-p)\cdot\frac{1}{p^2} = \frac{1-p}{p}.$$

（另一个做法：$N+1$ 是直至首次成功为止（包括该次）的试验次数，其平均值为 $\frac{1}{p}$。）代入 $p = \mathrm{e}^{-3}$，

$$\mathrm{E}(N) = \frac{1-\mathrm{e}^{-3}}{\mathrm{e}^{-3}} = \mathrm{e}^3 - 1,$$

正是所要求的结果。

短间隔长度的期望值

我们要求 $\mathrm{E}(T \mid T < 30)$。已知 $T < 30$ 时，T 的条件概率密度函数（conditional density）在 $0 \le t < 30$ 上是 $f(t)/\mathrm{P}(T < 30)$，所以

$$\mathrm{E}(T \mid T < 30) = \frac{1}{1 - \mathrm{e}^{-3}} \int_0^{30} \frac{t}{10}\, \mathrm{e}^{-t/10}\, \mathrm{d}t.$$

用分部积分法（integration by parts），取 $u = t$ 和 $\mathrm{d}v = \frac{1}{10}\mathrm{e}^{-t/10}\,\mathrm{d}t$，于是 $v = -\mathrm{e}^{-t/10}$，

$$\begin{aligned}\int_0^{30} \frac{t}{10}\, \mathrm{e}^{-t/10}\, \mathrm{d}t &= \Big[-t\, \mathrm{e}^{-t/10} \Big]_0^{30} + \int_0^{30} \mathrm{e}^{-t/10}\, \mathrm{d}t \\ &= -30\, \mathrm{e}^{-3} + 10(1 - \mathrm{e}^{-3}) \\ &= 10 - 40\, \mathrm{e}^{-3}.\end{aligned}$$

因此，把分子和分母同乘以 e^3，

$$\mathrm{E}(T \mid T < 30) = \frac{10 - 40\, \mathrm{e}^{-3}}{1 - \mathrm{e}^{-3}} = \frac{10\, \mathrm{e}^3 - 40}{\mathrm{e}^3 - 1} \text{ 分钟} \approx 8.43 \text{ 分钟}.$$

这个数值略小于 10，是合理的：我们已经把长间隔剔除了。

等候时间的期望值

在不受打扰的半小时开始之前，等候时间就是 N 个短间隔之和。已知 $N = n$ 时，这是 n 个独立的间隔，每个都以短于 30 分钟为条件，所以已知 $N = n$ 时等候时间的期望值是 $n\,\mathrm{E}(T \mid T < 30)$。再对 N 取平均，

$$\mathrm{E}(\text{等候时间}) = \mathrm{E}(N)\, \mathrm{E}(T \mid T < 30) = (\mathrm{e}^3 - 1) \cdot \frac{10\, \mathrm{e}^3 - 40}{\mathrm{e}^3 - 1} = 10\, \mathrm{e}^3 - 40 \text{ 分钟},$$

即大约 161 分钟，或 2 小时 41 分钟。（如果想求直至试卷完成为止的时间，就要加上那半小时本身：$10\, \mathrm{e}^3 - 10 \approx 191$ 分钟。）

验算

- $\int_0^{\infty} \frac{1}{10}\mathrm{e}^{-t/10}\,\mathrm{d}t = 1$，所以 f 确实是一个概率密度函数，而它的平均值是 10，与“平均每十分钟一次”吻合。
- 在没有条件的情况下，$\mathrm{E}(T) = 10$ 应该可以分拆为 $\mathrm{P}(T < 30)\,\mathrm{E}(T \mid T < 30) + \mathrm{P}(T \ge 30)\,\mathrm{E}(T \mid T \ge 30)$。由无记忆性（memorylessness）可知 $\mathrm{E}(T \mid T \ge 30) = 40$，而果然有 $(10 - 40\mathrm{e}^{-3}) + 40\mathrm{e}^{-3} = 10$。
- 把这个过程模拟运行 200 000 次，得出平均打扰次数为 19.1（对照 $\mathrm{e}^3 - 1 \approx 19.09$），平均等候时间为 161.1 分钟（对照 $10\mathrm{e}^3 - 40 \approx 160.86$）。

常见错误

- 使用 $\mathrm{E}(N) = \frac{1}{p} = \mathrm{e}^3$：这样连长间隔也计算在内，但长间隔结束时，那半小时之前并没有再出现打扰。
- 把 $\mathrm{E}(N)$ 乘以无条件平均值 10 来求等候时间。组成等候时间的间隔已知是短间隔，所以它们的平均值是 $\mathrm{E}(T \mid T < 30)$，而不是 10。

STEP 之外

这些打扰构成一个泊松过程（*Poisson process*），而各个间隔是互相独立的指数分布变量这个结果，正是泊松过程的基本性质之一；最后一步 $\mathrm{E}(\text{共}N\text{ 项之和}) = \mathrm{E}(N) \times \mathrm{E}(\text{每项})$ 是一个取条件的论证，与沃尔德恒等式（*Wald's identity*）关系密切。两者合起来，通向更新理论（renewal theory）和排队模型（queueing models），这些都是大学一、二年级概率课程的内容。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门）以合适的程度讲解泊松过程和条件期望。

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。