详细解答

STEP I 1988，第 16 题

概率与统计 · 连续分布与条件化

STEP I 1988，第 16 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-q16

关键思路

第一部分是一道归一化（normalisation）与积分的练习：题目只给出密度函数（density）到相差一个常数的程度，所以我们必须要求总概率为 1，从而求出这个常数；而 $t^2/(1+t^2)^2$ 的积分，只要把它拆成 $1/(1+t^2)$ 与 $t/(1+t^2)$ 的导数，便迎刃而解。其后的检验制度只是一棵简短的分情况树形图。只有车龄 4 和 5 时的检验才有影响（车龄 3 之前没有汽车会生锈），所以我们按汽车在第四次检验时是否已经生锈来分情况，而最后一部分就是在同一棵树形图上运用贝叶斯定理（Bayes' theorem）。

解答

开始生锈时间的分布

设汽车在车龄 $3+T$ 时开始生锈，其中 $T \geq 0$ 的密度函数为 $f(t) = k\,t^2/(1+t^2)^2$，k 为某常数。关键的积分来自以下观察：

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{t}{1+t^2}\right) = \frac{1-t^2}{(1+t^2)^2}, \qquad \text{故} \qquad \frac{t^2}{(1+t^2)^2} = \frac{1}{2}\left(\frac{1}{1+t^2} - \frac{1-t^2}{(1+t^2)^2}\right),$$

只要把括号内的式子通分，以 $(1+t^2)^2$ 为公分母，便可验证：分子为 $(1+t^2)-(1-t^2) = 2t^2$。因此

$$\int_0^t \frac{s^2}{(1+s^2)^2}\,\mathrm{d}s = \frac{1}{2}\left(\arctan t - \frac{t}{1+t^2}\right).$$

当 $t \to \infty$ 时，$\arctan t \to \frac{\pi}{2}$ 而 $t/(1+t^2) \to 0$，所以在 $[0,\infty)$ 上的积分为 $\frac{\pi}{4}$。要使 f 成为密度函数，我们需要 $k \cdot \frac{\pi}{4} = 1$，故 $k = 4/\pi$。于是

$$F(t) = \mathrm{P}(\text{开始生锈时车龄小于} 3+t) = \mathrm{P}(T < t) = \frac{2}{\pi}\left(\arctan t - \frac{t}{1+t^2}\right), \quad t \geq 0.$$

（留意 $F(t) \to 1$：每辆汽车终究都会生锈。）

我们需要以下两个数值：

$$F(1) = \frac{2}{\pi}\left(\frac{\pi}{4} - \frac{1}{2}\right) = \frac{1}{2} - \frac{1}{\pi} \approx 0.1817,$$
$$F(2) = \frac{2}{\pi}\left(\arctan 2 - \frac{2}{5}\right) \approx 0.4502.$$

第五周年检验不合格

车龄 3 之前没有汽车生锈，所以车龄 1、2 和 3 时的检验必定合格（恰好在第三次检验那一刻开始生锈的概率为零）。汽车要在车龄 5 时不合格，必须在车龄 4 时合格，然后在车龄 5 时不合格，而不合格的前提是车龄 5 时已经生锈。锈蚀永远不会消失，所以这件事有两种互斥（disjoint）的发生方式。

- *车龄 4 时已经生锈*（概率为 $F(1)$）。它以概率 $\frac{1}{2}$ 通过第四次检验，然后在仍然生锈的情况下，以概率 $\frac{1}{2}$ 在第五次检验不合格。贡献：$\frac{1}{4}F(1)$。
- *在车龄 4 与 5 之间开始生锈*（概率为 $F(2) - F(1)$）。它必定通过第四次检验，并以概率 $\frac{1}{2}$ 在第五次检验不合格。贡献：$\frac{1}{2}\big(F(2) - F(1)\big)$。

这里我们按题目的原意假设：一辆生锈汽车在各次检验的结果彼此独立（independent），亦与开始生锈的时间独立。

> ? **想一想** 为甚么我们不用乘上通过车龄 1、2 和 3 时各次检验的概率?

因为这些检验都在任何汽车有可能生锈之前进行，所以每辆汽车都会通过；该因子为 1。

把两种情况相加，

$$\begin{aligned}\mathrm{P}(\text{检验不合格时的车龄为5}) &= \tfrac{1}{4}F(1) + \tfrac{1}{2}\big(F(2) - F(1)\big) = \tfrac{1}{2}F(2) - \tfrac{1}{4}F(1)\\ &= \frac{1}{\pi}\left(\arctan 2 - \frac{2}{5}\right) - \frac{1}{8} + \frac{1}{4\pi}\\ &= \boxed{\frac{1}{\pi}\left(\arctan 2 - \frac{3}{20}\right) - \frac{1}{8} \approx 0.1797.}\end{aligned}$$

已知在第五次检验后被销毁，求第四次检验时已生锈的概率

一辆在第五次检验后随即被销毁的汽车，正正就是在该次检验不合格（而之前各次全部合格）的汽车，亦即刚才求得的事件。设 A 为汽车在第四次检验时已经生锈的事件，D 为它在第五次检验不合格的事件。上面第一点所述的正是 $A \cap D$，所以

$$\mathrm{P}(A \cap D) = \tfrac{1}{4}F(1) = \frac{1}{8} - \frac{1}{4\pi} = \frac{\pi - 2}{8\pi}.$$

根据条件概率（conditional probability）的定义，

$$\mathrm{P}(A \mid D) = \frac{\mathrm{P}(A \cap D)}{\mathrm{P}(D)} = \frac{\dfrac{\pi - 2}{8\pi}}{\dfrac{1}{\pi}\left(\arctan 2 - \dfrac{3}{20}\right) - \dfrac{1}{8}}.$$

分子和分母同乘以 8π，

$$\boxed{\mathrm{P}(A \mid D) = \frac{\pi - 2}{8\arctan 2 - \frac{6}{5} - \pi} = \frac{5(\pi - 2)}{40\arctan 2 - 6 - 5\pi} \approx 0.253.}$$

所以在五年时被销毁的汽车当中，只有约四分之一是在已经生锈的情况下侥幸通过第四次检验的；大部分都是在第五年内才生锈。

验算

- $F(0) = 0$，而且当 $t \to \infty$ 时 $F(t) \to 1$，符合分布函数（distribution function）应有的性质。对 $\frac{4}{\pi}\int_0^2 t^2/(1+t^2)^2\,\mathrm{d}t$ 使用辛普森法则（Simpson's rule）得 0.45018，与 $F(2)$ 吻合。
- 模拟 400 000 辆汽车（抽样得出开始生锈的时间，再进行每年的检验），得出五年时不合格的概率为 0.1789，条件概率为 0.255，而精确值分别为 0.1797 和 0.2528。
- 合理性检查：$\mathrm{P}(D)$ 必须小于 $\frac{1}{2}F(2) \approx 0.225$（即车龄五年时已生锈然后不合格的概率，当中不考虑较早被淘汰的情况），而事实的确如此。

常见错误

- 忘记归一化：密度函数只是与 $t^2/(1+t^2)^2$ 成正比，漏掉因子 $4/\pi$ 会令其后每个答案都出错。
- 把“车龄 5 时已生锈”当作概率为 $F(2)$ 的单一情况，然后乘以 $\frac{1}{2}$。这样做忽略了那些在车龄 4 时已生锈、并在该次检验被淘汰的汽车。
- 在最后一部分，以“车龄 5 时已生锈”而非“车龄 5 时不合格”作为条件。

STEP 之外

这里的函数 F 是一个分布函数，而比值 $f(t)/(1-F(t))$ 称为风险率（hazard rate），即已知汽车至今仍未生锈的条件下，瞬间生锈的可能性；它是存活分析（survival analysis）与可靠性理论（reliability theory）的核心对象。最后一部分是滤波（filtering）的一个小

例子：从不完美的观测（各次检验）推断隐藏的状态（是否生锈）。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门）正是以这种精神讲解连续分布与条件概率。

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。