解答例

STEP I 1988、第 16 問

確率・統計 ・ 連続分布と条件付け

STEP I 1988、問 16

先に進む前に、まず問題を読んでおこう。問題は STEP Support Programme の問題データベースにある：

step.maths.org/questions/88-s1-q16

鍵となる考え方

最初の部分は正規化（normalisation）と積分の練習である。密度（density）は定数倍を除いてしか与えられていないので、全確率が 1 となることを要請してその定数を求めなければならない。また $t^2/(1+t^2)^2$ の積分は、これを $1/(1+t^2)$ と $t/(1+t^2)$ の導関数とに分けることで計算できる。その後の検査の仕組みは、場合分けの短い樹形図にすぎない。意味をもつのは 4 歳と 5 歳での検査だけである（3 歳より前に錆びている車はない）。そこで、4 回目の検査の時点ですでに錆びていたかどうかで場合を分ける。最後の部分は、同じ樹形図の上でのベイズの定理（Bayes' theorem）である。

解答

錆び始める時期の分布

車が年齢 $3+T$ で錆び始めるとする。ここで $T \geq 0$ は、ある定数 k を用いて密度 $f(t) = kt^2/(1+t^2)^2$ をもつ。鍵となる積分は、次のことに気づけば得られる。

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

これは括弧の中を共通の分母 $(1+t^2)^2$ で通分すれば確かめられる。分子は $(1+t^2)-(1-t^2) = 2t^2$ となる。したがって

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

$t \to \infty$ のとき $\arctan t \to \frac{\pi}{2}$、$t/(1+t^2) \to 0$ であるから、$[0,\infty)$ 上の積分は $\frac{\pi}{4}$ である。f が密度であるためには $k \cdot \frac{\pi}{4} = 1$ が必要なので、$k = 4/\pi$ である。ゆえに

$$\boxed{F(t) = \mathrm{P}(\text{錆び始めの年齢が次より小さい：}3+t) = \mathrm{P}(T < t) = \frac{2}{\pi}\left(\arctan t - \frac{t}{1+t^2}\right), \quad t \geq 0.}$$

（$F(t) \to 1$ に注意せよ。どの車もいずれは錆びる。）

必要となる二つの値は次のとおりである。

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

5 年目の検査で不合格になる確率

3 歳より前に錆びている車はないので、1 歳、2 歳、3 歳での検査には必ず合格する（ちょうど 3 回目の検査の瞬間に錆び始める確率は 0 である）。5 歳で不合格になるには、4 歳で合格し、かつ 5 歳で不合格にならなければならない。そして不合格になるには、5 歳の時点で錆びていることが必要である。錆は消えることがないので、これが起こるのは互いに排反（disjoint）な次の二通りである。

- **4 歳の時点ですでに錆びている**（確率 $F(1)$）。4 回目の検査に確率 $\frac{1}{2}$ で合格し、その後も錆びたままなので、確率 $\frac{1}{2}$ で 5 回目の検査に不合格となる。寄与は $\frac{1}{4}F(1)$。
- **4 歳から 5 歳の間に錆び始める**（確率 $F(2) - F(1)$）。4 回目の検査には必ず合格し、5 回目の検査には確率 $\frac{1}{2}$ で不合格となる。寄与は $\frac{1}{2}\big(F(2) - F(1)\big)$。

ここでは問題の意図どおり、錆びた車に対する異なる検査の結果は互いに独立（independent）であり、錆び始めた時刻とも独立であると仮定している。

> **考えてみよう** なぜ 1 歳、2 歳、3 歳での検査に合格する確率を掛けないのだろうか？

それらの検査は、どの車もまだ錆びうる前に行われるので、すべての車が合格するからである。その因子は 1 である。

二つの場合を足し合わせると

$$\begin{aligned}\mathrm{P}(\text{不合格となる年齢が5}) &= \tfrac{1}{4}F(1) + \tfrac{1}{2}\big(F(2) - F(1)\big) = \tfrac{1}{2}F(2) - \tfrac{1}{4}F(1) \\ &= \frac{1}{\pi}\left(\arctan 2 - \frac{2}{5}\right) - \frac{1}{8} + \frac{1}{4\pi} \\ &= \boxed{\frac{1}{\pi}\left(\arctan 2 - \frac{3}{20}\right) - \frac{1}{8} \approx 0.1797.}\end{aligned}$$

5 回目の検査で廃車になったとき、4 回目の検査の時点で錆びていた確率

5 回目の検査の直後に廃車となる車とは、ちょうど（それ以前の検査にはすべて合格したうえで）その検査に不合格となった車のことであり、これはいま求めた事象である。4 回目の検査の時点で錆びていたという事象を A、5 回目の検査に不合格となる事象を D とする。上の一つ目の項目はまさに $A \cap D$ であるから

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

条件付き確率（conditional probability）の定義により

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

分子と分母に 8π を掛けて

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

つまり、5 年で廃車となった車のうち、すでに錆びていながら 4 回目の検査をすり抜けていたものは約 4 分の 1 にすぎず、大部分は 5 年目の間に錆び始めたのである。

検算

- 分布関数（distribution function）であるからには当然そうあるべきとおり、$F(0) = 0$ であり、$t \to \infty$ のとき $F(t) \to 1$ である。$\frac{4}{\pi}\int_0^2 t^2/(1+t^2)^2\,\mathrm{d}t$ にシンプソンの公式（Simpson's rule）を適用すると 0.45018 となり、$F(2)$ と一致する。
- 400 000 台の車のシミュレーション（錆び始める時刻を抽出し、毎年の検査を実行する）では、5 年で不合格となる確率が 0.1789、条件付き確率が 0.255 となり、厳密値 0.1797 および 0.2528 とよく合う。
- 妥当性の確認：$\mathrm{P}(D)$ は $\frac{1}{2}F(2) \approx 0.225$（それ以前の廃車を無視して、5 歳の時点で錆びていてかつ不合格となる確率）より小さくなければならないが、実際そうなっている。

よくある間違い

- 正規化を忘れること。密度は $t^2/(1+t^2)^2$ に比例するだけであり、因子 $4/\pi$ を落とすとその後の答えがすべて狂う。

- 「5 歳の時点で錆びている」を確率 $F(2)$ の一つの場合として扱い、$\frac{1}{2}$ を掛けてしまうこと。これでは、4 歳の時点で錆びていてその検査で除かれた車を無視することになる。
- 最後の部分で、「5 歳で不合格になった」ではなく「5 歳の時点で錆びている」を条件としてしまうこと。

STEP のその先へ

ここでの関数 F は分布関数であり、比 $f(t)/(1-F(t))$ は**ハザード率**（hazard rate）、すなわちそれまで錆びずに残った車が次の瞬間に錆び始める確率の割合である。これは生存解析（survival analysis）と信頼性理論（reliability theory）の中心的な対象である。最後の部分はフィルタリング（filtering）の小さな例であり、不完全な観測（検査）から隠れた状態（錆びているかどうか）を推測している。Grimmett and Welsh, *Probability: An Introduction*（確率論入門）は、このような考え方で連続分布と条件付けを扱っている。

Beyond STEP 解答集：https://beyondstep.keithmcnulty.org/ja/。解答本文は CC0 1.0 である。問題そのものは © Cambridge University Press & Assessment に帰属し、ここには掲載していない。英語版から機械翻訳したものであり、まだ校閲を経ていない。