解答例

STEP II 2014、第 12 問

確率・統計 ・ ハザード関数

STEP II 2014、問 12

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/14-s2-q12

鍵となる考え方

ハザード関数（hazard function）とは、まだ生きているものについての死亡率である。後半の小問はすべて、次のただ一つの観察から導かれる。$F'(t) = f(t)$ であるから、

$$h(t) = \frac{f(t)}{1 - F(t)} = -\frac{\mathrm{d}}{\mathrm{d}t} \ln(1 - F(t)).$$

つまり、h を知ることは $\ln(1 - F)$ の導関数を知ることにほかならない。これを積分し、寿命の始まる点で $F = 0$ となることを用い、さらに F を微分すれば f が得られる。

解答

小問 (i)

事象 $T \geq t$（時刻 t にハエが生きている）を A、事象 $t \leq T \leq t + \delta t$（その後の δt の間に死ぬ）を B とする。$B \cap A = B$ であるから（連続型確率変数では $\mathrm{P}(T = t) = 0$ なので、端点を含めるかどうかは問題にならない）、

$$\mathrm{P}(B \mid A) = \frac{\mathrm{P}(B)}{\mathrm{P}(A)} = \frac{F(t + \delta t) - F(t)}{1 - F(t)}.$$

δt が小さいとき $F(t + \delta t) - F(t) \approx F'(t)\,\delta t = f(t)\,\delta t$ であるから、

$$\mathrm{P}(B \mid A) \approx \frac{f(t)\,\delta t}{1 - F(t)} = h(t)\,\delta t,$$

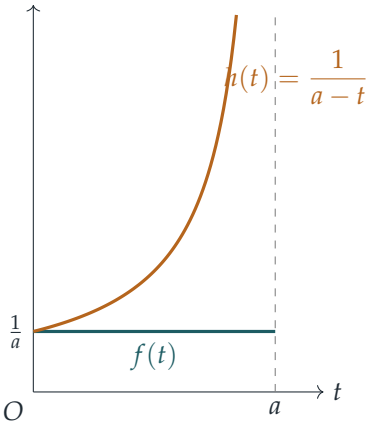
となり、示すべきことが示された。（起こりえない事象を条件にしないために $F(t) < 1$ が必要である。これはちょうど h が定義される範囲である。）

小問 (ii)

ここでは $0 < t < a$ において $f(t) = F'(t) = \frac{1}{a}$ であり、T は $(0, a)$ 上の一様分布（uniform distribution）に従う。このとき

$$h(t) = \frac{1/a}{1 - t/a} = \frac{1}{a - t}, \qquad 0 < t < a.$$

確率密度関数（probability density function）は平坦であるが、ハザードは $h(0) = \frac{1}{a}$ から始まって増加し、$t \to a^-$ のとき $h(t) \to \infty$ となる。これは納得のいく結果である。時刻 a の直前まで生き延びたハエは、残されたごくわずかな時間のうちに確実に死ぬからである。



（図は $a = 2$ として描いてある。）

小問 (iii)

鍵となる考え方により、$t > a$ において $-\dfrac{\mathrm{d}}{\mathrm{d}t}\ln(1 - F(t)) = \dfrac{1}{t}$ である。a から t まで積分し、T は a より大きい値しかとらないので $F(a) = 0$ であることを用いると、

$$-\ln(1 - F(t)) = \ln t - \ln a = \ln\frac{t}{a}, \qquad \text{よって} \qquad 1 - F(t) = \frac{a}{t}.$$

したがって $F(t) = 1 - \dfrac{a}{t}$ であり、

$$f(t) = \frac{a}{t^2}, \qquad t > a$$

となる（$t \le a$ では $f(t) = 0$）。確認として、$\int_a^\infty at^{-2}\,\mathrm{d}t = 1$ であり、$f/(1 - F) = (a/t^2)/(a/t) = 1/t$ である。

小問 (iv)

十分性（「**if**」）。$k > 0$ として、$t > b$ において $f(t) = k\mathrm{e}^{-k(t-b)}$ であるとする。$\int_b^\infty k\mathrm{e}^{-k(t-b)}\,\mathrm{d}t = 1$ であるから、確率はすべて $t > b$ の範囲にあり、$t \le b$ では $f(t) = 0$ かつ $F(t) = 0$ である。よって $t \le b$ では $h(t) = 0/1 = 0$ である。$t > b$ では

$$F(t) = \int_b^t k\mathrm{e}^{-k(s-b)}\,\mathrm{d}s = 1 - \mathrm{e}^{-k(t-b)}, \qquad h(t) = \frac{k\mathrm{e}^{-k(t-b)}}{\mathrm{e}^{-k(t-b)}} = k,$$

となり、これは定数である。

必要性（「**only if**」）。$t \le b$ で $h(t) = 0$、$t > b$ で $h(t) = k$（定数）であるとする。

? **考えてみよう** $t \le b$ で $h(t) = 0$ であることから、$F(b)$ について何がわかるか。

$h(t) = 0$ となるところでは $f(t) = h(t)(1 - F(t)) = 0$ である。よって f はすべての $t \le b$ で 0 となり、$t \le b$ において $F(t) = \int_{-\infty}^t f = 0$ が得られる。特に $F(b) = 0$ である。（これにより、そこで $F < 1$ であることも確かめられるので、h は実際に $t \le b$ で定義されている。）

$t > b$ では $-\dfrac{\mathrm{d}}{\mathrm{d}t}\ln(1 - F(t)) = k$ である。b から t まで積分し、$F(b) = 0$ を用いると、

$$-\ln(1 - F(t)) = k(t - b), \qquad F(t) = 1 - \mathrm{e}^{-k(t-b)},$$

となり、これを微分して $t > b$ において $f(t) = k\mathrm{e}^{-k(t-b)}$ を得る。最後に、k は正でなければならない。$f \ge 0$ であるから $k \ge 0$ であり、もし $k = 0$ ならば F はいつまでも 0 のままで $F(t) \not\to 1$ となり、T はまともな確率変数にならないからである。以上で両方向が示された。

小問 (v)

ここでも $t > 0$ において $-\dfrac{\mathrm{d}}{\mathrm{d}t}\ln(1 - F(t)) = \dfrac{\lambda}{\theta^\lambda}t^{\lambda-1}$ である。$T > 0$ より $F(0) = 0$ であることを用いて 0 から t まで積分する。$\lambda > 0$ であるから $t \to 0$ のとき $t^\lambda \to 0$ となることに注意すると、

$$-\ln(1 - F(t)) = \frac{\lambda}{\theta^\lambda} \cdot \frac{t^\lambda}{\lambda} = \left(\frac{t}{\theta}\right)^\lambda.$$

よって $F(t) = 1 - \exp(-(t/\theta)^\lambda)$ であり、微分すると

$$f(t) = \frac{\lambda}{\theta^\lambda}\,t^{\lambda-1}\exp\left(-\left(\frac{t}{\theta}\right)^\lambda\right), \qquad t > 0.$$

これはまともな確率密度関数である。実際、正の値をとり、$t \to \infty$ のとき $F(t) \to 1$ となる。

検算

- どの答えも、$f/(1-F)$ を直接計算すれば確かめられる。たとえば (v) では、指数関数の因子が打ち消し合って $\lambda\theta^{-\lambda}t^{\lambda-1}$ が残る。
- 小問 (v) で $\lambda = 1$ とすると、ハザードは定数 $1/\theta$ となり $f(t) = \frac{1}{\theta}\mathrm{e}^{-t/\theta}$ が得られる。これは小問 (iv) で $b = 0, k = 1/\theta$ としたものである。
- 数値的に確かめると、$\lambda = \frac{1}{2}, \theta = 2$ のとき f を $(0,\infty)$ で積分すると 1.0000 となり、$F(1) = 1 - \mathrm{e}^{-1/\sqrt{2}} \approx 0.5069$ は f を 0 から 1 まで直接積分した値と一致する。

よくある間違い

- 寿命の始まる点で $F = 0$ という条件を忘れ、F に任意の積分定数（constant of integration）を残したままにしてしまう。
- (iv) で「必要十分」（if and only if）の一方向しか証明しない、あるいは $t \le b$ で $h = 0$ であることを用いて $F(b) = 0$ を導くことをしない。
- マイナス符号を落とす。正しくは $\frac{\mathrm{d}}{\mathrm{d}t}\ln(1-F) = -f/(1-F)$ である。

STEP のその先へ

関係式 $1 - F(t) = \exp\left(-\int_0^t h(s)\,\mathrm{d}s\right)$ は、**生存解析**（survival analysis）と信頼性理論（reliability theory）の基礎である。(v) の分布はワイブル分布（Weibull distribution）である。$\lambda < 1$ ならハザードは減少し（初期故障）、$\lambda = 1$ なら無記憶性（memoryless）をもつ指数分布（exponential distribution）となり、$\lambda > 1$ ならハザードは増加する（摩耗故障）。生存解析は、医学の臨床試験から工学まで、あらゆるところで用いられている。基礎となる確率論については、Blitzstein and Hwang, *Introduction to Probability*（確率論入門）を参照されたい。

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。