详细解答

STEP II 2014，第 12 题

概率与统计 · 风险函数

STEP II 2014，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/14-s2-q12

关键思路

风险函数（hazard function）是对仍然存活者而言的死亡率。后面各部分全都源于一个观察：由于 $F'(t) = f(t)$，

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

所以知道 h，就等于知道 $\ln(1 - F)$ 的导数（derivative）。先积分，再利用寿命开始时 $F = 0$ 这一事实，最后对 F 求导便得到 f。

解答

第 (i) 部分

设 A 为事件 $T \geq t$（苍蝇在时间 t 仍然存活），B 为事件 $t \leq T \leq t + \delta t$（苍蝇在接下来的 δt 内死亡）。由于 $B \cap A = B$（对连续随机变量（continuous random variable）而言 $\mathrm{P}(T = t) = 0$，所以端点是否包括在内并无影响），

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

当 δt 很小时，$F(t + \delta t) - F(t) \approx F'(t)\,\delta t = f(t)\,\delta t$，因此

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

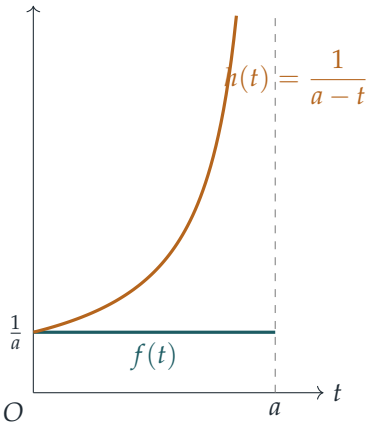
此即所求。（我们需要 $F(t) < 1$，以免以一个不可能事件作为条件；而这正是 h 有定义的范围。）

第 (ii) 部分

这里当 $0 < t < a$ 时 $f(t) = F'(t) = \frac{1}{a}$：$T$ 在 $(0, a)$ 上服从均匀分布（uniform distribution）。于是

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

概率密度是平的，但风险函数由 $h(0) = \frac{1}{a}$ 开始递增，且当 $t \to a^-$ 时 $h(t) \to \infty$。这是合理的：一只存活至接近时间 a 的苍蝇，必定会在余下极短的时间内死亡。



（图中取 $a = 2$。）

第 (iii) 部分

利用关键思路，当 $t > a$ 时 $-\dfrac{\mathrm{d}}{\mathrm{d}t}\ln(1 - F(t)) = \dfrac{1}{t}$。由 a 积分至 t，并因 T 只取大于 a 的值而有 $F(a) = 0$，得

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

因此 $F(t) = 1 - \dfrac{a}{t}$，且

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

（而当 $t \le a$ 时 $f(t) = 0$）。验算一下：$\int_a^\infty at^{-2}\,\mathrm{d}t = 1$，且 $f/(1 - F) = (a/t^2)/(a/t) = 1/t$。

第 (iv) 部分

“若”的方向。设当 $t > b$ 时 $f(t) = k\mathrm{e}^{-k(t-b)}$，其中 $k > 0$。由于 $\int_b^\infty k\mathrm{e}^{-k(t-b)}\,\mathrm{d}t = 1$，全部概率都落在 $t > b$，所以当 $t \le b$ 时 $f(t) = 0$ 及 $F(t) = 0$。因此当 $t \le b$ 时 $h(t) = 0/1 = 0$。当 $t > b$ 时，

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

这是一个常数。

“仅当”的方向。设当 $t \le b$ 时 $h(t) = 0$，而当 $t > b$ 时 $h(t) = k$（常数）。

> **想一想** 当 $t \le b$ 时 $h(t) = 0$，这告诉我们 $F(b)$ 是甚么?

在 $h(t) = 0$ 之处，有 $f(t) = h(t)(1 - F(t)) = 0$。所以对所有 $t \le b$，f 都等于零，从而当 $t \le b$ 时 $F(t) = \int_{-\infty}^{t} f = 0$；特别地 $F(b) = 0$。（这亦确认了该处 $F < 1$，所以 h 在 $t \le b$ 上确实有定义。）

当 $t > b$ 时，$-\dfrac{\mathrm{d}}{\mathrm{d}t}\ln(1 - F(t)) = k$。由 b 积分至 t，并利用 $F(b) = 0$，

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

求导后得当 $t > b$ 时 $f(t) = k\mathrm{e}^{-k(t-b)}$。最后，$k$ 必须是正数：因为 $f \ge 0$，所以 $k \ge 0$；而若 $k = 0$，则 F 会永远停留在 0，于是 $F(t) \not\to 1$，T 便不是一个真正的随机变量。至此两个方向都已证明。

第 (v) 部分

同样地，当 $t > 0$ 时 $-\dfrac{\mathrm{d}}{\mathrm{d}t}\ln(1 - F(t)) = \dfrac{\lambda}{\theta^{\lambda}}t^{\lambda-1}$。由 0 积分至 t，因 $T > 0$ 而有 $F(0) = 0$，并注意 $\lambda > 0$ 使得当 $t \to 0$ 时 $t^{\lambda} \to 0$，

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

所以 $F(t) = 1 - \exp(-(t/\theta)^{\lambda})$，求导得

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

这是一个真正的概率密度函数（probability density function）：它是正的，而且当 $t \to \infty$ 时 $F(t) \to 1$。

验算

- 每个答案都可以直接计算 $f/(1 - F)$ 来验证。例如在 (v) 中，指数因子互相抵消，剩下 $\lambda\theta^{-\lambda}t^{\lambda-1}$。
- 在第 (v) 部分取 $\lambda = 1$，得常数风险 $1/\theta$ 及 $f(t) = \frac{1}{\theta}\mathrm{e}^{-t/\theta}$，这正是第 (iv) 部分取 $b = 0$、$k = 1/\theta$ 的情况。
- 数值上，取 $\lambda = \frac{1}{2}$、$\theta = 2$，将 f 在 $(0, \infty)$ 上积分得 1.0000，而 $F(1) = 1 - \mathrm{e}^{-1/\sqrt{2}} \approx 0.5069$ 与直接将 f 由 0 积分至 1 的结果吻合。

常见错误

- 忘记寿命开始时 $F = 0$ 这个条件，以致 F 中留下一个任意的积分常数。
- 在 (iv) 中只证明了“当且仅当”（if and only if）的其中一个方向，或者没有利用当 $t \leq b$ 时 $h = 0$ 来推出 $F(b) = 0$。
- 漏掉负号：$\frac{\mathrm{d}}{\mathrm{d}t} \ln(1 - F) = -f/(1 - F)$。

STEP 之外

关系式 $1 - F(t) = \exp\left(-\int_0^t h(s)\,\mathrm{d}s\right)$ 是*存活分析*（survival analysis）及可靠性理论（reliability theory）的基础。(v) 中的分布是韦伯分布（Weibull distribution）：$\lambda < 1$ 给出递减的风险（早期故障），$\lambda = 1$ 给出具无记忆性（memoryless）的指数分布（exponential distribution），而 $\lambda > 1$ 则给出递增的风险（损耗老化）。存活分析的应用范围极广，由医学临床试验到工程学都有。关于其背后的概率论，可参阅 Blitzstein 与 Hwang 的 *Introduction to Probability*（《概率导论》）。

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。