詳細解答

STEP II 2014，第 12 題

概率與統計 • 風險函數

> **STEP II 2014，第 12 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：
>
> step.maths.org/questions/14-s2-q12

關鍵思路

風險函數（hazard function）是*對仍然存活者而言*的死亡率。後面各部分全都源於一個觀察：由於 $F'(t) = f(t)$，

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

所以知道 h，就等於知道 $\ln(1 - F)$ 的導數（derivative）。先積分，再利用壽命開始時 $F = 0$ 這一事實，最後對 F 求導便得到 f。

解答

第 (i) 部分

設 A 為事件 $T \geq t$（蒼蠅在時間 t 仍然存活），B 為事件 $t \leq T \leq t + \delta t$（蒼蠅在接下來的 δt 內死亡）。由於 $B \cap A = B$（對連續隨機變量（continuous random variable）而言 $\mathrm{P}(T = t) = 0$，所以端點是否包括在內並無影響），

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

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

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

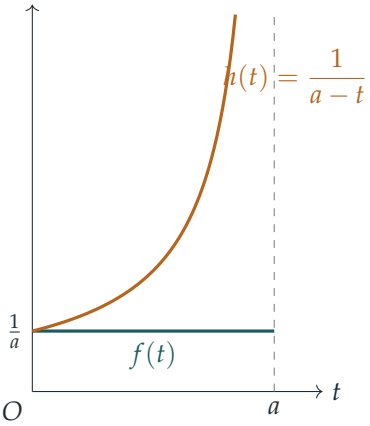
此即所求。（我們需要 $F(t) < 1$，以免以一個不可能事件作為條件；而這正是 h 有定義的範圍。）

第 (ii) 部分

這裏當 $0 < t < a$ 時 $f(t) = F'(t) = \frac{1}{a}$：$T$ 在 $(0, a)$ 上服從均勻分佈（uniform distribution）。於是

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

概率密度是平的，但風險函數由 $h(0) = \frac{1}{a}$ 開始遞增，且當 $t \to a^-$ 時 $h(t) \to \infty$。這是合理的：一隻存活至接近時間 a 的蒼蠅，必定會在餘下極短的時間內死亡。



（圖中取 $a = 2$。）

第 (iii) 部分

利用關鍵思路，當 $t > a$ 時 $-\frac{\mathrm{d}}{\mathrm{d}t}\ln(1 - F(t)) = \frac{1}{t}$。由 a 積分至 t，並因 T 只取大於 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 $F(t) = 1 - \frac{a}{t}$，且

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

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

第 (iv) 部分

「**若**」**的方向**。設當 $t > b$ 時 $f(t) = k\mathrm{e}^{-k(t-b)}$，其中 $k > 0$。由於 $\int_b^\infty k\mathrm{e}^{-k(t-b)}\,\mathrm{d}t = 1$，全部概率都落在 $t > b$，所以當 $t \le b$ 時 $f(t) = 0$ 及 $F(t) = 0$。因此當 $t \le b$ 時

$h(t) = 0/1 = 0$。當 $t > b$ 時，

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

這是一個常數。

「**僅當**」**的方向**。設當 $t \le b$ 時 $h(t) = 0$，而當 $t > b$ 時 $h(t) = k$（常數）。

> ? **想一想** 當 $t \le b$ 時 $h(t) = 0$，這告訴我們 $F(b)$ 是甚麼？

在 $h(t) = 0$ 之處，有 $f(t) = h(t)(1 - F(t)) = 0$。所以對所有 $t \le b$，f 都等於零，從而當 $t \le b$ 時 $F(t) = \int_{-\infty}^t f = 0$；特別地 $F(b) = 0$。（這亦確認了該處 $F < 1$，所以 h 在 $t \le b$ 上確實有定義。）

當 $t > b$ 時，$-\dfrac{\mathrm{d}}{\mathrm{d}t}\ln(1 - F(t)) = k$。由 b 積分至 t，並利用 $F(b) = 0$，

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

求導後得當 $t > b$ 時 $f(t) = k\mathrm{e}^{-k(t-b)}$。最後，$k$ 必須是正數：因為 $f \ge 0$，所以 $k \ge 0$；而若 $k = 0$，則 F 會永遠停留在 0，於是 $F(t) \not\to 1$，T 便不是一個真正的隨機變量。至此兩個方向都已證明。

第 (v) 部分

同樣地，當 $t > 0$ 時 $-\dfrac{\mathrm{d}}{\mathrm{d}t}\ln(1 - F(t)) = \dfrac{\lambda}{\theta^\lambda}t^{\lambda-1}$。由 0 積分至 t，因 $T > 0$ 而有 $F(0) = 0$，並注意 $\lambda > 0$ 使得當 $t \to 0$ 時 $t^\lambda \to 0$，

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

所以 $F(t) = 1 - \exp(-(t/\theta)^\lambda)$，求導得

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

這是一個真正的概率密度函數（probability density function）：它是正的，而且當 $t \to \infty$ 時 $F(t) \to 1$。

驗算

- 每個答案都可以直接計算 $f/(1 - F)$ 來驗證。例如在 (v) 中，指數因子互相抵消，剩下 $\lambda\theta^{-\lambda}t^{\lambda-1}$。
- 在第 (v) 部分取 $\lambda = 1$，得常數風險 $1/\theta$ 及 $f(t) = \frac{1}{\theta}\mathrm{e}^{-t/\theta}$，這正是第 (iv) 部分取 $b = 0$、$k = 1/\theta$ 的情況。

- 數值上，取 $\lambda = \frac{1}{2}$、$\theta = 2$，將 f 在 $(0, \infty)$ 上積分得 1.0000，而 $F(1) = 1 - \mathrm{e}^{-1/\sqrt{2}} \approx 0.5069$ 與直接將 f 由 0 積分至 1 的結果吻合。

常見錯誤

- 忘記壽命開始時 $F = 0$ 這個條件，以致 F 中留下一個任意的積分常數。
- 在 (iv) 中只證明了「當且僅當」（if and only if）的其中一個方向，或者沒有利用當 $t \le b$ 時 $h = 0$ 來推出 $F(b) = 0$。
- 漏掉負號：$\frac{\mathrm{d}}{\mathrm{d}t} \ln(1 - F) = -f/(1 - F)$。

STEP 之外

關係式 $1 - F(t) = \exp\bigl(-\int_0^t h(s)\,\mathrm{d}s\bigr)$ 是*存活分析*（survival analysis）及可靠性理論（reliability theory）的基礎。(v) 中的分佈是韋伯分佈（Weibull distribution）：$\lambda < 1$ 給出遞減的風險（早期故障），$\lambda = 1$ 給出具無記憶性（memoryless）的指數分佈（exponential distribution），而 $\lambda > 1$ 則給出遞增的風險（損耗老化）。存活分析的應用範圍極廣，由醫學臨牀試驗到工程學都有。關於其背後的概率論，可參閱 Blitzstein 與 Hwang 的 *Introduction to Probability*（《概率導論》）。

Beyond STEP 詳細解答：https://beyondstep.keithmcnulty.org/zh-hant/。解答文字採用 CC0 1.0；題目本身版權屬 Cambridge University Press & Assessment 所有（©），本文件不予轉載。由機器從英文翻譯而成；尚未經審閱。