詳細解答

# STEP I 2005，第 14 題

概率與統計 • 混合型隨機變量

**STEP I 2005，第 14 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/05-s1-q14

## 關鍵思路

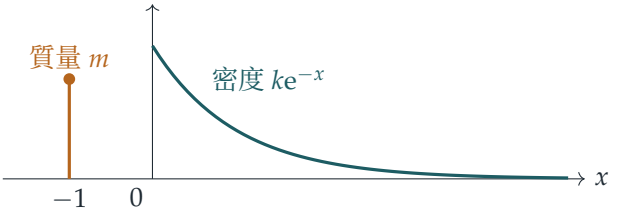
$X$ 是一個混合型隨機變量（mixed random variable）：有一團概率集中在 $-1$ 這一點上，其餘則連續地分佈在 $[0,\infty)$ 上。處理的方法是把每一項計算都拆成兩部分：來自 $X=-1$ 的離散項，加上在 $[0,\infty)$ 上對密度所作的積分。只要留意到連續部分不過是一個乘了常數的指數分佈（exponential distribution），大部分積分便都是我們早已認識的了。

## 解答

把題目給出的累積概率（cumulative probability）求導，可知 $X$ 的連續部分的密度（density）為

$$f(x)=\frac{\mathrm{d}}{\mathrm{d}x}k(1-\mathrm{e}^{-x})=k\mathrm{e}^{-x},\qquad x\geq 0.$$

所以 $X$ 的樣子是這樣的：在 $-1$ 處有一個大小為 $m$ 的點質量（point mass），在 $[0,\infty)$ 上則是曲線 $k\mathrm{e}^{-x}$。



**第 (i) 部分**

總概率為 1。令 $x \to \infty$，得 $\mathrm{P}(X \geq 0) = k(1-0) = k$，所以

$$m + k = 1, \qquad \text{即} \qquad k = 1 - m.$$

（由於 $m \geq 0$ 是一個概率，而且 $m < \frac{1}{2}$，我們有 $k > \frac{1}{2} > 0$，所以密度確實是正的。）

**第 (ii) 部分**

拆成兩部分，

$$\begin{aligned}\mathrm{E}(X) &= (-1)\cdot m + \int_0^\infty x\,(1-m)\mathrm{e}^{-x}\,\mathrm{d}x \\ &= -m + (1-m)\cdot 1 = 1 - 2m,\end{aligned}$$

其中用到 $\int_0^\infty x\mathrm{e}^{-x}\,\mathrm{d}x = 1$（可用分部積分法（integration by parts），或認出這是參數為 1 的指數分佈的平均值）。這正是所要求的結果。

**第 (iii) 部分**

*方差*（*variance*）。同樣地，利用 $\int_0^\infty x^2\mathrm{e}^{-x}\,\mathrm{d}x = 2$（分部積分兩次），

$$\mathrm{E}(X^2) = (-1)^2\cdot m + (1-m)\cdot 2 = 2 - m.$$

因此

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = 2 - m - (1-2m)^2 \\ &= 1 + 3m - 4m^2 = (1-m)(1+4m).\end{aligned}$$

**想一想** 中位數（median）可以是 $-1$，或者落在 $-1$ 與 0 之間嗎？

*中位數*。不可以：$X$ 從不在 $(-1,0)$ 內取值，而且 $\mathrm{P}(X \leq -1) = m < \frac{1}{2}$，所以單靠點質量並不足以累積到一半的概率。條件 $m < \frac{1}{2}$ 正正是用在這裏。因此中位數 $M$ 滿足 $M \geq 0$ 及 $\mathrm{P}(X \leq M) = \frac{1}{2}$：

$$\begin{aligned} m + (1-m)(1-\mathrm{e}^{-M}) &= \tfrac{1}{2} \\ 1 - (1-m)\mathrm{e}^{-M} &= \tfrac{1}{2} \\ \mathrm{e}^{-M} &= \frac{1}{2(1-m)}.\end{aligned}$$

於是

$$M = \ln\left(2(1-m)\right) = \ln(2-2m).$$

作為一致性檢查，由 $0 \le m < \frac{1}{2}$ 得 $1 < 2 - 2m \le 2$，所以 $M > 0$，與應有的情況相符。由於分佈在 $[0, \infty)$ 上是連續的，這個 $M$ 同時滿足 $\mathrm{P}(X \ge M) = \frac{1}{2}$，所以中位數是毫不含糊的。

**第 (iv) 部分**

當 $X = -1$ 時，$|X|^{1/2} = 1$；當 $X \ge 0$ 時，$|X|^{1/2} = \sqrt{X}$。所以

$$\mathrm{E}(|X|^{1/2}) = 1 \cdot m + (1 - m) \int_0^\infty \sqrt{x}\, \mathrm{e}^{-x}\, \mathrm{d}x.$$

? **想一想** 怎樣把 $\int_0^\infty \sqrt{x}\, \mathrm{e}^{-x}\, \mathrm{d}x$ 化成題目給出的那個積分？

題目給出的積分含有 $\mathrm{e}^{-y^2}$，這提示我們代入 $x = y^2$（其中 $y \ge 0$），於是 $\sqrt{x} = y$ 及 $\mathrm{d}x = 2y\, \mathrm{d}y$；積分上下限 0 與 $\infty$ 保持不變：

$$\int_0^\infty \sqrt{x}\, \mathrm{e}^{-x}\, \mathrm{d}x = \int_0^\infty y\, \mathrm{e}^{-y^2}\, 2y\, \mathrm{d}y = 2 \int_0^\infty y^2 \mathrm{e}^{-y^2}\, \mathrm{d}y = 2 \cdot \tfrac{1}{4}\sqrt{\pi} = \tfrac{1}{2}\sqrt{\pi}.$$

因此

$$\mathrm{E}(|X|^{1/2}) = m + \tfrac{1}{2}\sqrt{\pi}\,(1 - m).$$

## 驗算

- 當 $m = 0$ 時，$X$ 恰好是參數為 1 的指數分佈：我們的公式給出平均值 1、方差 1、中位數 $\ln 2$ 及 $\mathrm{E}(\sqrt{X}) = \frac{1}{2}\sqrt{\pi} = \Gamma(\frac{3}{2})$，全部都是標準數值。
- 當 $m \to \frac{1}{2}$ 時，中位數趨向 $\ln 1 = 0$，這是合理的：一半的概率位於 $-1$，所以中位數會滑落到連續部分的最底端。
- 取 $m = 0.3$ 並抽取 $4 \times 10^5$ 個樣本作模擬，得出平均值 0.402（公式為 0.4）、方差 1.551（公式為 1.54）、中位數 0.336（公式為 $\ln 1.4 \approx 0.336$）及 $\mathrm{E}(|X|^{1/2}) \approx 0.922$（公式為 0.920）。

## 常見錯誤

- 在 $\mathrm{E}(X^2)$ 或 $\mathrm{E}(|X|^{1/2})$ 中處理點質量時出錯：它的貢獻是 $+m$ 而不是 $-m$，因為 $(-1)^2 = |-1|^{1/2} = 1$；漏掉這一項同樣是錯的。
- 解 $(1 - m)(1 - \mathrm{e}^{-M}) = \frac{1}{2}$ 來求中位數，這樣便忽略了在 $-1$ 處已經累積的概率 $m$。
- 換元時沒有把 $\mathrm{d}x$ 換成 $2y\, \mathrm{d}y$，結果得出 $\frac{1}{4}\sqrt{\pi}$ 而不是 $\frac{1}{2}\sqrt{\pi}$。

## STEP 之外

同時包含點質量和密度的分佈在實際應用中十分常見（一宗保險索償有某個概率為零，否則便是連續的）。在大學裏，這類分佈會用統一的方法處理：把期望值（expectation）寫成對分佈函數（distribution function）的積分，$\mathrm{E}(g(X)) = \int g(x)\,\mathrm{d}F(x)$，一次過涵蓋離散、連續和混合三種情況。第 (iv) 部分的積分就是伽瑪函數（Gamma function）的值 $\Gamma(\frac{3}{2}) = \frac{1}{2}\sqrt{\pi}$。Grimmett 與 Welsh 合著的 *Probability: An Introduction*（概率論入門）是很好的下一步讀物。

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。