详细解答

STEP I 2005，第 14 题

概率与统计 · 混合型随机变量

STEP I 2005，第 14 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/05-s1-q14

关键思路

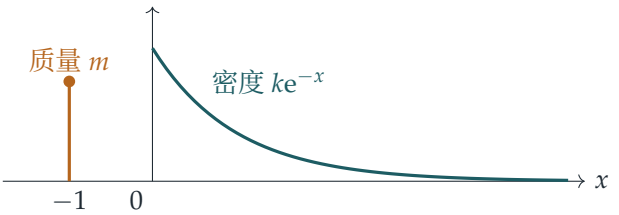
X 是一个混合型随机变量（mixed random variable）：有一团概率集中在 -1 这一点上，其余则连续地分布在 $[0,\infty)$ 上。处理的方法是把每一项计算都拆成两部分：来自 $X=-1$ 的离散项，加上在 $[0,\infty)$ 上对密度所作的积分。只要留意到连续部分不过是一个乘了常数的指数分布（exponential distribution），大部分积分便都是我们早已认识的了。

解答

把题目给出的累积概率(cumulative probability)求导,可知 X 的连续部分的密度(density)为

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

所以 X 的样子是这样的：在 -1 处有一个大小为 m 的点质量（point mass），在 $[0,\infty)$ 上则是曲线 $k\mathrm{e}^{-x}$。



第 (i) 部分

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

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

（由于 $m \geq 0$ 是一个概率，而且 $m < \frac{1}{2}$，我们有 $k > \frac{1}{2} > 0$，所以密度确实是正的。）

第 (ii) 部分

拆成两部分，

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

其中用到 $\int_0^\infty x\mathrm{e}^{-x}\,\mathrm{d}x = 1$（可用分部积分法（integration by parts），或认出这是参数为 1 的指数分布的平均值）。这正是所要求的结果。

第 (iii) 部分

方差（*variance*）。同样地，利用 $\int_0^\infty x^2\mathrm{e}^{-x}\,\mathrm{d}x = 2$（分部积分两次），

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

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 中位数（median）可以是 -1，或者落在 -1 与 0 之间吗?

中位数。不可以：X 从不在 $(-1,0)$ 内取值，而且 $\mathrm{P}(X \leq -1) = m < \frac{1}{2}$，所以单靠点质量并不足以累积到一半的概率。条件 $m < \frac{1}{2}$ 正正是用在这里。因此中位数 M 满足 $M \geq 0$ 及 $\mathrm{P}(X \leq M) = \frac{1}{2}$：

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

于是

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

作为一致性检查，由 $0 \le m < \frac{1}{2}$ 得 $1 < 2 - 2m \le 2$，所以 $M > 0$，与应有的情况相符。由于分布在 $[0, \infty)$ 上是连续的，这个 M 同时满足 $\mathrm{P}(X \ge M) = \frac{1}{2}$，所以中位数是毫不含糊的。

第 (iv) 部分

当 $X = -1$ 时，$|X|^{1/2} = 1$；当 $X \ge 0$ 时，$|X|^{1/2} = \sqrt{X}$。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 怎样把 $\int_0^\infty \sqrt{x}\, \mathrm{e}^{-x}\, \mathrm{d}x$ 化成题目给出的那个积分?

题目给出的积分含有 e^{-y^2}，这提示我们代入 $x = y^2$（其中 $y \ge 0$），于是 $\sqrt{x} = y$ 及 $\mathrm{d}x = 2y\, \mathrm{d}y$；积分上下限 0 与 ∞ 保持不变：

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

因此

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

验算

- 当 $m = 0$ 时，X 恰好是参数为 1 的指数分布：我们的公式给出平均值 1、方差 1、中位数 $\ln 2$ 及 $\mathrm{E}(\sqrt{X}) = \frac{1}{2}\sqrt{\pi} = \Gamma(\frac{3}{2})$，全部都是标准数值。
- 当 $m \to \frac{1}{2}$ 时，中位数趋向 $\ln 1 = 0$，这是合理的：一半的概率位于 -1，所以中位数会滑落到连续部分的最底端。
- 取 $m = 0.3$ 并抽取 4×10^5 个样本作模拟，得出平均值 0.402（公式为 0.4）、方差 1.551（公式为 1.54）、中位数 0.336（公式为 $\ln 1.4 \approx 0.336$）及 $\mathrm{E}(|X|^{1/2}) \approx 0.922$（公式为 0.920）。

常见错误

- 在 $\mathrm{E}(X^2)$ 或 $\mathrm{E}(|X|^{1/2})$ 中处理点质量时出错：它的贡献是 $+m$ 而不是 $-m$，因为 $(-1)^2 = |-1|^{1/2} = 1$；漏掉这一项同样是错的。
- 解 $(1 - m)(1 - \mathrm{e}^{-M}) = \frac{1}{2}$ 来求中位数，这样便忽略了在 -1 处已经累积的概率 m。
- 换元时没有把 $\mathrm{d}x$ 换成 $2y\, \mathrm{d}y$，结果得出 $\frac{1}{4}\sqrt{\pi}$ 而不是 $\frac{1}{2}\sqrt{\pi}$。

STEP 之外

同时包含点质量和密度的分布在实际应用中十分常见（一宗保险索偿有某个概率为零，否则便是连续的）。在大学里，这类分布会用统一的方法处理：把期望值（expectation）写成对分布函数（distribution function）的积分，$\mathrm{E}(g(X)) = \int g(x)\,\mathrm{d}F(x)$，一次过涵盖离散、连续和混合三种情况。第 (iv) 部分的积分就是伽玛函数（Gamma function）的值 $\Gamma(\frac{3}{2}) = \frac{1}{2}\sqrt{\pi}$。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门）是很好的下一步读物。

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。