详细解答

STEP II 2005，第 14 题

概率与统计 · 混合分布

STEP II 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-s2-q14

关键思路

这个密度是两个我们早已熟悉的密度的加权混合：一个是标准正态分布（standard normal distribution），另一个是 $[0,\lambda]$ 上的均匀分布（uniform distribution）。因此，f 的每个积分都可拆成正态部分和均匀部分，而每一部分都可以直接由标准事实（总概率为 1、均值、$\mathrm{E}(X^2)$）读出，完全不必做任何艰深的积分。方差（variance）则由 $\mathrm{E}(X^2)-\mu^2$ 求得。

解答

常数 k、均值与方差

由于 ϕ 和 g 都是概率密度函数（probability density function），两者在整条实数线上的积分都是 1。所以

$$1=\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = k\,[1+\lambda\cdot 1] \quad\Longrightarrow\quad k=\frac{1}{1+\lambda}.$$

至于均值（mean），我们利用两个成分各自的均值：标准正态分布的均值为 0，而 $[0,\lambda]$ 上的均匀密度 g 的均值为 $\frac{1}{2}\lambda$。因此

$$\mu=\int_{-\infty}^{\infty} xf(x)\,\mathrm{d}x = k\left[\int x\phi(x)\,\mathrm{d}x+\lambda\int xg(x)\,\mathrm{d}x\right]=\frac{1}{1+\lambda}\left[0+\lambda\cdot\frac{\lambda}{2}\right]=\frac{\lambda^2}{2(1+\lambda)}.$$

同样地，标准正态分布满足 $\int x^2\phi(x)\,\mathrm{d}x=1$（由于其均值为 0，这就是它的方差），而

$$\int x^2 g(x)\,\mathrm{d}x=\frac{1}{\lambda}\int_0^{\lambda} x^2\,\mathrm{d}x=\frac{\lambda^2}{3}.$$

所以

$$\mathrm{E}(X^2)=\frac{1}{1+\lambda}\left[1+\lambda\cdot\frac{\lambda^2}{3}\right]=\frac{3+\lambda^3}{3(1+\lambda)}.$$

想一想 为甚么 $\mathrm{E}(X^2)$ 可以这样拆开，方差却不可以?

期望值（expectation）是积分，而积分是线性的，所以 $\mathrm{E}(X^2)$ 是各成分二阶矩（second moment）的加权组合，权重与 f 由各成分密度组合而成时所用的完全相同。方差并非线性（它涉及 μ^2），所以我们必须先把二阶矩组合起来，然后才减去 μ^2。这样做便得

$$\begin{aligned}\sigma^2 = \mathrm{E}(X^2) - \mu^2 &= \frac{3+\lambda^3}{3(1+\lambda)} - \frac{\lambda^4}{4(1+\lambda)^2} \\ &= \frac{4(3+\lambda^3)(1+\lambda) - 3\lambda^4}{12(1+\lambda)^2} \\ &= \frac{12+12\lambda+4\lambda^3+4\lambda^4-3\lambda^4}{12(1+\lambda)^2} = \frac{\lambda^4+4\lambda^3+12\lambda+12}{12(1+\lambda)^2},\end{aligned}$$

这正是所要求的结果。

$\lambda = 2$ 的情况

此时 $k = \frac{1}{3}$，而在 $[0,2]$ 上 $\lambda g(x) = 1$，所以

$$f(x) = \begin{cases} \frac{1}{3}\phi(x) & x < 0 \text{ 或} x > 2, \\ \frac{1}{3}(\phi(x)+1) & 0 \le x \le 2. \end{cases}$$

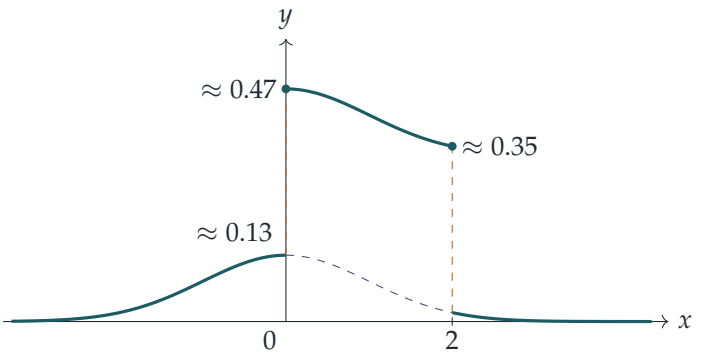
此外

$$\mu = \frac{4}{6} = \frac{2}{3}, \qquad \sigma^2 = \frac{16+32+24+12}{12\cdot 9} = \frac{84}{108} = \frac{7}{9}, \qquad \sigma = \frac{\sqrt{7}}{3}.$$

第 (i) 部分

这条曲线是标准正态钟形曲线的三分之一，并在区间 $[0,2]$ 上向上提升了 $\frac{1}{3}$。所以它在 $x = 0$ 处向上跳跃，在 $x = 2$ 处向下跳跃，跳跃的高度都是 $\frac{1}{3}$。利用 $\phi(0) = 1/\sqrt{2\pi} \approx 0.399$ 和 $\phi(2) \approx 0.054$：

- 在 0 的左边紧贴处，$f \approx 0.133$；在 0 的右边紧贴处，$f \approx 0.466$（最大值）；
- 在 2 的左边紧贴处，$f \approx 0.351$；在 2 的右边紧贴处，$f \approx 0.018$；
- 在 $[0,2]$ 上曲线递减，形状与正态曲线一致；
- 当 $x \to \pm\infty$ 时 $f \to 0$，而且处处 $f > 0$。



$[0,2]$ 上的紫红色虚线是 $\frac{1}{3}\phi(x)$，画出来以作比较；实线则是 $y=f(x)$。（纵轴的比例已被拉长。）

第 (ii) 部分

设 $F(x)=\mathrm{P}(X\le x)=\int_{-\infty}^{x} f(t)\,\mathrm{d}t$。对 $\frac{1}{3}\phi$ 积分得 $\frac{1}{3}\Phi$，而 $[0,2]$ 上额外的 $\frac{1}{3}$ 所贡献的，是 $\frac{1}{3}$ 乘以 $[0,2]$ 中位于 x 左边那一段的长度。所以

$$F(x)=\begin{cases}\frac{1}{3}\Phi(x) & x<0,\\ \frac{1}{3}\big(\Phi(x)+x\big) & 0\le x\le 2,\\ \frac{1}{3}\big(\Phi(x)+2\big) & x>2.\end{cases}$$

各段在 $x=0$ 和 $x=2$ 处互相吻合，这是必然的（没有点质量（point mass）时，累积分布函数（CDF）是连续的），而且当 $x\to\infty$ 时 $F(x)\to\frac{1}{3}(1+2)=1$。

第 (iii) 部分

这里 $\mu+2\sigma=\frac{2}{3}+\frac{2}{3}\sqrt{7}\approx 2.43$，大于 2，所以在该处要用 F 的最后一个分支。由于 X 是连续的，

$$\begin{aligned}\mathrm{P}(0<X<\mu+2\sigma)&=F\big(\tfrac{2}{3}+\tfrac{2}{3}\sqrt{7}\big)-F(0)\\&=\tfrac{1}{3}\big(\Phi(\tfrac{2}{3}+\tfrac{2}{3}\sqrt{7})+2\big)-\tfrac{1}{3}\Phi(0)\\&=\tfrac{1}{3}\big(0.9921+2-0.5\big)=\frac{2.4921}{3}=0.8307.\end{aligned}$$

验算

- 当 $\lambda\to 0$ 时，均匀部分消失：公式给出 $\mu\to 0$ 和 $\sigma^2\to\frac{12}{12}=1$，正是标准正态分布的数值，合乎预期。

- 对于 $\lambda = 2$，全方差定律（law of total variance，见下文）以另一种方法给出相同的答案。各成分的权重为 $\frac{1}{3}, \frac{2}{3}$，方差为 $1, \frac{1}{3}$，均值为 $0, 1$。方差的平均值是 $\frac{1}{3} + \frac{2}{9} = \frac{5}{9}$，均值的方差是 $\frac{2}{3} - \frac{4}{9} = \frac{2}{9}$，总和是 $\frac{7}{9}$。
- 以数值方法对 f 积分（中点法则，在 0 和 λ 处分段），取 $\lambda = 0.5, 1, 2, 3.7$，所得的总概率 1、均值和方差公式全部吻合，误差小于 10^{-6}；对于 $\lambda = 2$，结果是 $\mu = 0.6667$ 和 $\sigma^2 = 0.7778$。
- 直接计算 $\Phi(2.4305)$ 得 0.99246，而非题目所引用的 0.9921；概率则是 0.8308，而非 0.8307。这微小的差异源自题目所给的数值，而我们应该采用题目的数值。

常见错误

- 忘记了因子 $k = \frac{1}{3}$，以致总概率变成 3。
- 把方差算成 $k[1 + \lambda \cdot \frac{\lambda^2}{12}]$，即把各成分的方差组合起来。混合分布的方差不能线性地组合；二阶矩才可以。
- 把 f 画成连续曲线：它在 $x = 0$ 和 $x = \lambda$ 处有跳跃。

STEP 之外

形如 $\sum_i w_i f_i(x)$ 的密度，其中权重 $w_i \geq 0$ 且总和为 1，称为混合分布（mixture distribution）：以概率 w_i 选出成分 i，然后从 f_i 中抽样。这里的权重是 $\frac{1}{1+\lambda}$ 和 $\frac{\lambda}{1+\lambda}$。混合分布的方差其实就是全方差定律的化身：各成分方差的平均值，加上各成分均值的方差。混合分布是统计学和机器学习中聚类（clustering）方法的基础。Blitzstein 与 Hwang 的 *Introduction to Probability*（概率论入门教材）对条件期望值（conditional expectation）和全方差定律有清晰的讲解。

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。