详细解答

STEP III 2017，第 13 题

概率与统计 · 期望值与方差

STEP III 2017，第 13 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/17-s3-q13

关键思路

把 $X - x$ 写成 $(X - \mu) + (\mu - x)$。第一部分的平均值（mean）为零，所以平方后取期望值（expectation）时，交叉项（cross term）会消失。剩下的就是方差（variance）加上一个常数，这与“$\mathrm{E}(X^2) = \mathrm{Var}(X) + \mu^2$”是同一条恒等式。至于均匀分布（uniform distribution）的情况，Y 是 X 的函数，但并非一对一（one-to-one），所以求它的概率密度函数（probability density function）最稳妥的方法，是经由它的累积分布函数（cumulative distribution function）入手。

解答

函数 V

对任何固定的数 x，

$$\begin{aligned}\mathrm{V}(x) = \mathrm{E}((X-x)^2) &= \mathrm{E}(((X-\mu)+(\mu-x))^2)\\ &= \mathrm{E}((X-\mu)^2) + 2(\mu-x)\,\mathrm{E}(X-\mu) + (\mu-x)^2\\ &= \sigma^2 + (x-\mu)^2,\end{aligned}$$

这里用了期望值的线性性质（linearity of expectation）（此处 $\mu - x$ 是常数）以及 $\mathrm{E}(X - \mu) = 0$。

Y 的平均值

由于 $\mathrm{V}(x) = \sigma^2 + (x - \mu)^2$ 对每个实数 x 都成立，我们可以代入 $x = X$：

$$Y = \mathrm{V}(X) = \sigma^2 + (X-\mu)^2.$$

想一想 既然 V 是由期望值定义的，$\mathrm{V}(X)$ 岂不是只是一个数?

不是：V 是一个普通的实变量函数（一个关于 x 的二次式）。取期望值时 x 是固定不变的。把随机变量（random variable）X 代入这个二次式，便得到一个新的随机变量。所以

$$\mathrm{E}(Y) = \sigma^2 + \mathrm{E}((X-\mu)^2) = \sigma^2 + \sigma^2 = 2\sigma^2,$$

这就是 $(*)$。

均匀分布的情况

若 X 在 $[0,1]$ 上均匀分布，则 $\mu = \frac{1}{2}$，而 $\sigma^2 = \int_0^1 x^2\,\mathrm{d}x - \frac{1}{4} = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$。因此

$$\mathrm{V}(x) = \tfrac{1}{12} + (x-\tfrac{1}{2})^2 = x^2 - x + \tfrac{1}{3}.$$

Y 的分布。 现在 $Y = \frac{1}{12} + (X-\frac{1}{2})^2$。当 X 取遍 $[0,1]$ 时，$|X-\frac{1}{2}|$ 取遍 $[0,\frac{1}{2}]$，所以 Y 的取值范围是 $[\frac{1}{12}, \frac{1}{12}+\frac{1}{4}] = [\frac{1}{12}, \frac{1}{3}]$。

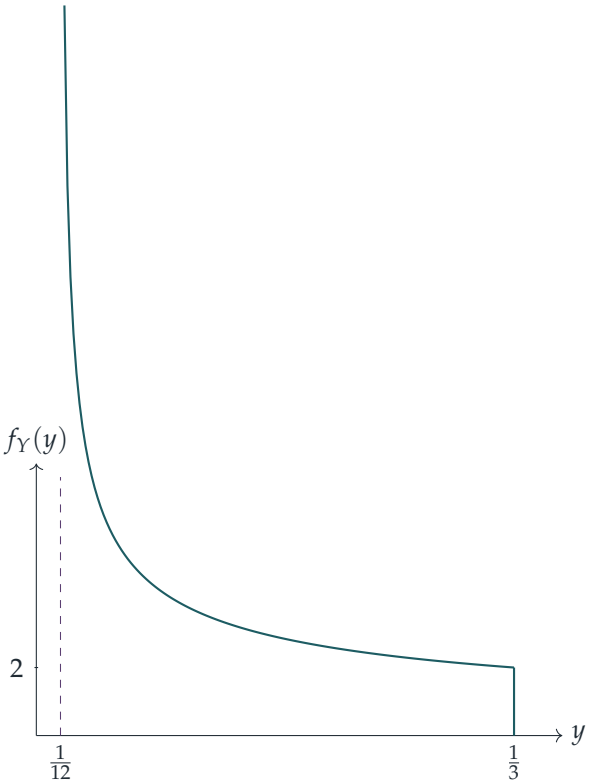
映射 $X \mapsto Y$ 是二对一的（点 $\frac{1}{2} \pm h$ 给出相同的 Y），所以我们改用分布函数处理。对于 $\frac{1}{12} \le y \le \frac{1}{3}$，

$$\begin{aligned}\mathrm{P}(Y \le y) &= \mathrm{P}((X-\tfrac{1}{2})^2 \le y - \tfrac{1}{12}) = \mathrm{P}(\tfrac{1}{2} - s \le X \le \tfrac{1}{2} + s), \quad \text{其中} s = \sqrt{y - \tfrac{1}{12}},\\ &= 2s = 2\sqrt{y - \tfrac{1}{12}},\end{aligned}$$

因为 $0 \le s \le \frac{1}{2}$，所以该区间位于 $[0,1]$ 之内，长度为 $2s$。（在 $y = \frac{1}{3}$ 时，这给出 $2 \cdot \frac{1}{2} = 1$，理应如此。）求导，得

$$f_Y(y) = \frac{1}{\sqrt{y - \frac{1}{12}}} \quad \text{当} \tfrac{1}{12} < y \le \tfrac{1}{3}, \qquad f_Y(y) = 0 \text{ 其他情况。}$$

这个概率密度函数在 $y = \frac{1}{12}$ 附近无界，但仍然可积（integrable）：Y 最有可能接近它的最小值，因为 $(X-\frac{1}{2})^2$ 在 $X = \frac{1}{2}$ 附近很平坦。



验证 $(*)$。 利用概率密度函数，并代换 $t = y - \frac{1}{12}$（于是 t 由 0 变到 $\frac{1}{4}$），

$$\mathrm{E}(Y) = \int_{1/12}^{1/3} \frac{y}{\sqrt{y - \frac{1}{12}}}\,\mathrm{d}y = \int_0^{1/4} (t^{1/2} + \tfrac{1}{12}t^{-1/2})\,\mathrm{d}t = \left[\tfrac{2}{3}t^{3/2} + \tfrac{1}{6}t^{1/2}\right]_0^{1/4}$$
$$= \tfrac{2}{3}\cdot\tfrac{1}{8} + \tfrac{1}{6}\cdot\tfrac{1}{2} = \tfrac{1}{12} + \tfrac{1}{12} = \tfrac{1}{6}.$$

由于 $2\sigma^2 = 2\cdot\frac{1}{12} = \frac{1}{6}$，结果与 $(*)$ 吻合。

验算

- 概率密度函数的积分等于 1：$\int_0^{1/4} t^{-1/2}\,\mathrm{d}t = 2\cdot\frac{1}{2} = 1$。
- $\mathrm{V}(x)$ 在 $x = \mu$ 处取最小值，最小值为 σ^2：平均值正是使期望平方距离最小的那一点。在均匀分布的情况下，$\mathrm{V}(\frac{1}{2}) = \frac{1}{12}$，即 Y 取值范围的下端。
- 以 400 000 个均匀分布的数值进行电脑模拟，得到 Y 的样本平均值约为 0.1667，与 $\frac{1}{6}$ 吻合。

常见错误

- 把 $\mathrm{V}(X)$ 当作常数 σ^2，结果得出 $\mathrm{E}(Y) = \sigma^2$。
- 使用一对一的变量变换公式 $f_Y = f_X \left|\mathrm{d}x/\mathrm{d}y\right|$ 时只计算了一个分支，结果只得到正确概率密度函数的一半。其实 $X = \frac{1}{2} + s$ 和 $X = \frac{1}{2} - s$ 两者都有贡献。

STEP 之外

$\mathrm{V}(x) = \sigma^2 + (x - \mu)^2$ 是最简单的偏差–方差分解（bias–variance decomposition）。如果你用 x 去估计某个量，均方误差（mean squared error）就会分拆为数据本身的离散程度加上偏差的平方。结果 $\mathrm{E}(Y) = 2\sigma^2$ 说明，X 的两个独立副本之间的期望平方距离是 $2\sigma^2$。统计学家一再用到这个事实，例如在解释为何样本方差（sample variance）要除以 $n - 1$ 时。Blitzstein 与 Hwang 合著的 *Introduction to Probability*（《概率导论》）涵盖了这两个概念，也讲解了如何求随机变量的非单调函数的概率密度函数。

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。