详细解答

STEP II 2002，第 13 题

概率与统计 · 变换后的分布函数

STEP II 2002，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/02-s2-q13

关键思路

这一题虽然放在力学部分，其实是一道关于变换分布函数（distribution function）的概率题。一切都建基于一条公式：对 G 求导得 $G'(y) = f(y) \cdot \dfrac{2}{(2-F(y))^2}$，其中 $f = F'$ 是 X 的概率密度函数（density）。因此，Y 的密度就是 X 的密度乘以一个权重，而这个权重被夹在 $\frac{1}{2}$ 与 2 之间。有了这一点，Y 的每一个期望值（expectation）都是 X 的密度乘以一个在 $[\frac{1}{2}, 2]$ 内的权重的积分，只要比较被积函数（integrand），便可得出各个不等式。条件 $a > 0$ 的作用，是确保被积函数非负。

解答

G 的性质

我们按题目的原意，假设 X 在 $[a,b]$ 上有密度 $f(y) = F'(y) \geq 0$，因此 X 的全部概率都落在 $[a,b]$ 内。

由于 $F(a) = 0$ 及 $F(b) = 1$，

$$G(a) = \frac{0}{2-0} = 0, \qquad G(b) = \frac{1}{2-1} = 1.$$

由商法则（quotient rule），

$$G'(y) = \frac{F'(y)(2-F(y)) - F(y)(-F'(y))}{(2-F(y))^2} = \frac{2F'(y)}{(2-F(y))^2}.$$

分子非负，因为分布函数是非递减的，所以 $F'(y) \geq 0$。分母是 $2 - F(y) \geq 1$ 的平方，故为正数。因此 $G'(y) \geq 0$。

至于那个界，由于 F 是分布函数，故 $0 \le F(y) \le 1$。于是

$$1 \le 2 - F(y) \le 2 \quad \Longrightarrow \quad 1 \le (2 - F(y))^2 \le 4 \quad \Longrightarrow \quad \tfrac{1}{2} \le \frac{2}{(2 - F(y))^2} \le 2,$$

其中各项皆为正数，所以取平方后不等号方向不变；取倒数会令不等号反向，再乘以 2 则只是把各项按比例放大。这就是所求的双重不等式。

综合而言，$G(a) = 0$、$G(b) = 1$ 以及 G 非递减，说明 G 确实是 $[a, b]$ 上的一个分布函数，所以随机变量（random variable）Y 是有意义的。它的密度是

$$g(y) = G'(y) = w(y)\, f(y), \qquad \text{其中} w(y) = \frac{2}{(2 - F(y))^2} \in [\tfrac{1}{2}, 2].$$

$\mathrm{E}(Y)$ 的界

> **想一想** 为甚么我们可以把不等式 $\frac{1}{2} \le w(y) \le 2$ 乘以 $yf(y)$，而不等号方向保持不变?

因为在 $[a, b]$ 上 $yf(y) \ge 0$：我们有 $f(y) \ge 0$，而且 $y \ge a > 0$。假设 $a > 0$ 正是用在这里。因此

$$\tfrac{1}{2}\, yf(y) \le y\, w(y) f(y) \le 2\, yf(y) \qquad \text{对于} a \le y \le b,$$

而在 $[a, b]$ 上积分后，不等式依然成立：

$$\tfrac{1}{2} \int_a^b yf(y)\, \mathrm{d}y \;\le\; \int_a^b y\, g(y)\, \mathrm{d}y \;\le\; 2 \int_a^b yf(y)\, \mathrm{d}y,$$

即

$$\tfrac{1}{2}\, \mathrm{E}(X) \le \mathrm{E}(Y) \le 2\, \mathrm{E}(X),$$

这正是所要证明的。

$\mathrm{Var}(Y)$ 的界

同样地，$y^2 f(y) \ge 0$（这只需要 $f \ge 0$），所以把 $w(y) \le 2$ 乘以它再积分，得

$$\mathrm{E}(Y^2) = \int_a^b y^2\, w(y) f(y)\, \mathrm{d}y \le 2 \int_a^b y^2 f(y)\, \mathrm{d}y = 2\, \mathrm{E}(X^2).$$

对于被减去的那一项，我们需要的是 $(\mathrm{E}(Y))^2$ 的下界。由上一部分，$\mathrm{E}(Y) \ge \frac{1}{2}\mathrm{E}(X)$，而且因为 $X \ge a > 0$，所以 $\mathrm{E}(X) > 0$。两边都是正数，故可以放心取平方：

$$(\mathrm{E}(Y))^2 \ge \tfrac{1}{4}(\mathrm{E}(X))^2.$$

因此

$$\begin{aligned}\operatorname{Var}(Y) &= \mathrm{E}(Y^2) - (\mathrm{E}(Y))^2 \\ &\le 2\,\mathrm{E}(X^2) - \tfrac{1}{4}(\mathrm{E}(X))^2 \\ &= 2(\operatorname{Var}(X) + (\mathrm{E}(X))^2) - \tfrac{1}{4}(\mathrm{E}(X))^2 \\ &= 2\operatorname{Var}(X) + \tfrac{7}{4}(\mathrm{E}(X))^2,\end{aligned}$$

其中第三行用了 $\mathrm{E}(X^2) = \operatorname{Var}(X) + (\mathrm{E}(X))^2$。这就是所求的界。

验算

取 X 为 $[1,3]$ 上的均匀分布（uniform distribution），则 $F(y) = \frac{1}{2}(y-1)$，且 $\mathrm{E}(X) = 2$。由于 $Y \ge 1$，故 $\mathrm{E}(Y) = 1 + \int_1^3 (1 - G(y))\,\mathrm{d}y$；代入 $u = F(y)$ 得

$$\mathrm{E}(Y) = 1 + 2\int_0^1 \frac{2(1-u)}{2-u}\,\mathrm{d}u = 1 + 4(1 - \ln 2) \approx 2.227,$$

稳稳地落在 1 与 4 之间。直观地说，w 在 F 小的地方较小，在 F 接近 1 的地方较大，所以 Y 偏向区间的上端，在这个例子中 $\mathrm{E}(Y) > \mathrm{E}(X)$。我们亦以数值方法验证过：对于随机区间上的 300 个随机幂律分布（power-law distribution），$\frac{1}{2}\mathrm{E}(X) \le \mathrm{E}(Y) \le 2\mathrm{E}(X)$ 与方差（variance）的界都成立；观察所得的比值 $\mathrm{E}(Y)/\mathrm{E}(X)$ 大约介乎 0.92 与 1.31 之间。

常见错误

- 忘记了 $a > 0$ 为何重要：假如 X 可以取负值，$yf(y)$ 便可能是负数，乘以它会令不等号反向。
- 用 $\mathrm{E}(Y) \le 2\mathrm{E}(X)$ 来估计 $\operatorname{Var}(Y)$ 的界：$(\mathrm{E}(Y))^2$ 这一项是被减去的，所以你需要的是它的下界，而不是上界。
- 试图把 $(y-\mu)^2$ 对 g 积分，直接比较 $\operatorname{Var}(Y)$ 与 $\operatorname{Var}(X)$：两个方差所用的平均值并不相同，所以应改用 $\mathrm{E}(Y^2)$ 来处理。

STEP 之外

写成 $g = wf$，是测度变换（change of measure）的一个简单例子：新的分布相对于旧的分布有一个密度（称为似然比（likelihood ratio），或 Radon–Nikodym 导数），而这个比值的界可以直接转化为期望值的界。统计学中的重要性抽样（importance sampling），以及金融数学中所用的概率测度变换，都以这个想法为基础。Grimmett 与 Welsh 的 *Probability:*

An Introduction（概率论入门）讲解分布函数、密度与期望值，程度正好适合由此再深入探究。

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。