详细解答

STEP III 2013，第 13 题

概率与统计 · 矩与分布函数

STEP III 2013，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/13-s3-q13

关键思路

由于 $\mathrm{F}' = \mathrm{f}$，表达式 $2\mathrm{F}(x)\mathrm{f}(x)$ 正是 $(\mathrm{F}(x))^2$ 的导数（derivative）。因此第 (i)(b) 部分只需一次分部积分（integration by parts）；而只要适当选取 g，关于 Y 的每个问题都能化为关于 X 的问题。至于各个界（bound），则来自把 f 或 F 换成第 (i)(a) 部分中较大的量 M 或 Mx，而每次替换都朝着令积分变大的方向进行。

解答

全文中，$\mathrm{F}(x) = \int_0^x \mathrm{f}(t)\,\mathrm{d}t$，故 $\mathrm{F}'(x) = \mathrm{f}(x)$；又由于 X 取值于 $[0,1]$，有 $\mathrm{F}(0) = 0$ 及 $\mathrm{F}(1) = 1$。

第 (i) 部分

(a) 对 $0 \le x \le 1$，由于对每个 t 都有 $0 \le \mathrm{f}(t) \le M$，

$$0 \le \int_0^x \mathrm{f}(t)\,\mathrm{d}t \le \int_0^x M\,\mathrm{d}t = Mx,$$

即 $0 \le \mathrm{F}(x) \le Mx$。

(b) 由链式法则(chain rule)，$\dfrac{\mathrm{d}}{\mathrm{d}x}(\mathrm{F}(x))^2 = 2\mathrm{F}(x)\mathrm{f}(x)$。取 $u = \mathrm{g}(x)$ 及 $\mathrm{d}v = 2\mathrm{F}(x)\mathrm{f}(x)\,\mathrm{d}x$ 进行分部积分，于是 $v = (\mathrm{F}(x))^2$，

$$\begin{aligned}\int_0^1 2\mathrm{g}(x)\mathrm{F}(x)\mathrm{f}(x)\,\mathrm{d}x &= \Big[\mathrm{g}(x)(\mathrm{F}(x))^2\Big]_0^1 - \int_0^1 \mathrm{g}'(x)(\mathrm{F}(x))^2\,\mathrm{d}x \\ &= \mathrm{g}(1) - \int_0^1 \mathrm{g}'(x)(\mathrm{F}(x))^2\,\mathrm{d}x,\end{aligned}$$

因为 $\mathrm{F}(1) = 1$ 且 $\mathrm{F}(0) = 0$。（这里需要 g 可微（differentiable），题目已隐含此假设。）

第 (ii) 部分

(a) Y 的概率密度函数（probability density function）的积分必须等于 1。在 (i)(b) 中取 $\mathrm{g}(x) = 1$，使 $\mathrm{g}' = 0$，得

$$\int_0^1 2\mathrm{F}(y)\mathrm{f}(y)\,\mathrm{d}y = 1,$$

故 $k = 2$。（亦可直接计算：$\int_0^1 k\mathrm{Ff}\,\mathrm{d}y = \frac{k}{2}[\mathrm{F}^2]_0^1 = \frac{k}{2}$。）另请注意 $k\mathrm{Ff} \geq 0$，所以这确实是一个概率密度函数。

(b) 我们有

$$\mathrm{E}(Y^n) = \int_0^1 2y^n\,\mathrm{F}(y)\mathrm{f}(y)\,\mathrm{d}y.$$

上界。由 (i)(a)，$\mathrm{F}(y) \leq My$，而其余因子 $2y^n\mathrm{f}(y)$ 均非负，所以

$$\mathrm{E}(Y^n) \leq \int_0^1 2y^n \cdot My \cdot \mathrm{f}(y)\,\mathrm{d}y = 2M\int_0^1 y^{n+1}\mathrm{f}(y)\,\mathrm{d}y = 2M\mu_{n+1}.$$

下界。现在在 (i)(b) 中取 $\mathrm{g}(y) = y^n$，于是 $\mathrm{g}(1) = 1$ 及 $\mathrm{g}'(y) = ny^{n-1}$：

$$\mathrm{E}(Y^n) = 1 - \int_0^1 ny^{n-1}(\mathrm{F}(y))^2\,\mathrm{d}y.$$

（当 $n = 0$ 时，该积分就是 0。）要求 $\mathrm{E}(Y^n)$ 的下界，我们需要求这个积分的上界。

> **想一想** $(\mathrm{F}(y))^2$ 的两个因子中，应该把哪一个换成 My?

只换其中一个。由于 $0 \leq \mathrm{F}(y) \leq My$，有 $(\mathrm{F}(y))^2 \leq My\,\mathrm{F}(y)$，加上 $ny^{n-1} \geq 0$，所以

$$\int_0^1 ny^{n-1}(\mathrm{F}(y))^2\,\mathrm{d}y \leq nM\int_0^1 y^n\mathrm{F}(y)\,\mathrm{d}y.$$

若把两个因子都换掉，就只剩下 $\int y^{n+1}\,\mathrm{d}y$，对 X 毫无信息；保留一个 F，便可再做一次分部积分，得出 X 的一个矩（moment）。取 $u = \mathrm{F}(y)$ 及 $\mathrm{d}v = y^n\,\mathrm{d}y$ 进行分部积分，

$$\int_0^1 y^n\mathrm{F}(y)\,\mathrm{d}y = \Big[\frac{y^{n+1}}{n+1}\mathrm{F}(y)\Big]_0^1 - \int_0^1 \frac{y^{n+1}}{n+1}\mathrm{f}(y)\,\mathrm{d}y = \frac{1-\mu_{n+1}}{n+1}.$$

综合以上，

$$\mathrm{E}(Y^n) \geq 1 - \frac{nM}{n+1}(1-\mu_{n+1}) = 1 + \frac{nM}{n+1}\mu_{n+1} - \frac{nM}{n+1},$$

连同上界，便是所求的一对不等式（inequality）。

(c) 把 (b) 的两端结合起来，对 $n \geq 0$，

$$1 + \frac{nM}{n+1}\mu_{n+1} - \frac{nM}{n+1} \leq 2M\mu_{n+1},$$

再把含 μ_{n+1} 的项移到右边，

$$1 - \frac{nM}{n+1} \leq \left(2M - \frac{nM}{n+1}\right)\mu_{n+1} = \frac{(n+2)M}{n+1}\,\mu_{n+1}.$$

系数 $\frac{(n+2)M}{n+1}$ 为正（$M > 0$，因为 f 的积分等于 1），所以可以除以它而不改变不等号方向：

$$\mu_{n+1} \geq \frac{n+1}{(n+2)M} - \frac{n}{n+2}.$$

此式对每个 $n \geq 0$ 成立。令 $m = n+1$，则 $m \geq 1$，上式变为

$$\mu_m \geq \frac{m}{(m+1)M} - \frac{m-1}{m+1} \qquad (m \geq 1),$$

这正是所求结果（只需把 m 改名为 n）。

验算

- 均匀分布（uniform distribution）的 X：$\mathrm{f} = 1$，$M = 1$，$\mu_n = \frac{1}{n+1}$。(c) 中的界为 $\frac{n}{n+1} - \frac{n-1}{n+1} = \frac{1}{n+1}$，所以等号成立。此时 Y 的概率密度函数为 $2y$，且 $\mathrm{E}(Y^n) = \frac{2}{n+2}$，而 (b) 中的两个界亦都等于 $\frac{2}{n+2}$。这是合理的：我们用过的每一个不等式 $\mathrm{F}(y) \leq My$，对均匀分布而言都是等式。
- $\mathrm{f}(x) = 3x^2$，$M = 3$，$n = 2$：$\mu_2 = \frac{3}{5}$，而界为 $\frac{2}{9} - \frac{1}{3} < 0$，轻松成立。对这个概率密度函数，$\mathrm{E}(Y^2) = \frac{3}{4}$ 及 $\mu_3 = \frac{1}{2}$，所以 (b) 中的界为 $1 + 2\cdot\frac{1}{2} - 2 = 0$ 及 $2\cdot 3\cdot\frac{1}{2} = 3$，而 $\frac{3}{4}$ 确实介乎两者之间。
- 我们亦以数值积分，对多个概率密度函数及多种 g 的选择，验证了 (i)(b) 中的恒等式及所有不等式。

常见错误

- 以 M^2y^2 作为 $\left(\mathrm{F}\right)^2$ 的上界：这虽然正确，却毫无用处，因为它丢失了关于 X 的所有信息。

- 忘记 $\mathrm{E}(Y^n)$ 的表达式中积分前面带有负号，所以要求 $\mathrm{E}(Y^n)$ 的下界，就需要积分的上界。
- 得出 μ_{n+1} 的界后便停下，没有移动下标，以得出题目所述对 $n \geq 1$ 的形式。

STEP 之外

概率密度函数 $2\mathrm{F}(y)\mathrm{f}(y)$ 并非偶然：它是 X 的两个独立副本的 $\max(X_1, X_2)$ 的概率密度函数，因为 $\mathrm{P}(\max \leq y) = \big(\mathrm{F}(y)\big)^2$。更一般地，$n$ 个独立观测值中第 r 小者的概率密度函数，由 F 和 $1 - \mathrm{F}$ 的幂组成；这些*顺序统计量*（order statistics）是无母数统计（non-parametric statistics）的核心。Blitzstein 与 Hwang 的 *Introduction to Probability*（《概率导论》）对顺序统计量有清晰的讲解。

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。