详细解答

STEP II 2022，第 12 题

概率与统计 • 平均值、中位数与众数

STEP II 2022，第 12 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2022 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696587-2022-past-papers.zip

关键思路

一切都源自累积分布函数（cumulative distribution function）F，而这里它只是一个简单的多项式。由于 F 在 $[0,1]$ 上递增，数 a 位于中位数（median）之下，当且仅当 $\mathrm{F}(a) < \frac{1}{2}$。因此，要比较平均值（mean）或众数（mode）与中位数的大小，我们根本不需要求出中位数本身：只需计算 F 在平均值或众数处的值，再与 $\frac{1}{2}$ 比较即可。

解答

第 (i) 部分

由于总概率为 1，

$$1 = k\int_0^1 \left(x^n - x^{n+1}\right)\mathrm{d}x = k\left(\frac{1}{n+1} - \frac{1}{n+2}\right) = \frac{k}{(n+1)(n+2)},$$

所以 $k = (n+1)(n+2)$，正如所求。于是

$$\mu = k\int_0^1 \left(x^{n+1} - x^{n+2}\right)\mathrm{d}x = (n+1)(n+2)\cdot\frac{1}{(n+2)(n+3)} = \frac{n+1}{n+3}.$$

第 (ii) 部分

当 $0 \le x \le 1$ 时，

$$\mathrm{F}(x) = (n+1)(n+2)\left(\frac{x^{n+1}}{n+1} - \frac{x^{n+2}}{n+2}\right) = x^{n+1}((n+2) - (n+1)x).$$

由于在 $(0,1)$ 上 $\mathrm{f}>0$，F 在该区间上严格递增（strictly increasing），而中位数 m 是唯一满足 $\mathrm{F}(m)=\frac{1}{2}$ 的点。因此

$$\mu < m \iff \mathrm{F}(\mu) < \tfrac{1}{2}.$$

现在代入 $\mu=\frac{n+1}{n+3}$，

$$(n+2)-(n+1)\mu = \frac{(n+2)(n+3)-(n+1)^2}{n+3} = \frac{3n+5}{n+3},$$

所以

$$\mathrm{F}(\mu) = \left(\frac{n+1}{n+3}\right)^{n+1}\frac{3n+5}{n+3}.$$

因此，条件 $\mathrm{F}(\mu)<\frac{1}{2}$ 两边乘以正数 $2\left(\frac{n+3}{n+1}\right)^{n+1}$ 后，变为

$$\frac{2(3n+5)}{n+3} < \left(\frac{n+3}{n+1}\right)^{n+1}.$$

由于 $\frac{2(3n+5)}{n+3}=\frac{6(n+3)-8}{n+3}=6-\frac{8}{n+3}$ 而且 $\frac{n+3}{n+1}=1+\frac{2}{n+1}$，这正是题目给出的不等式。所以只要该不等式成立，$\mu$ 就小于中位数。

要证明这个不等式，设 $N=n+1$，于是 $N\geq 3$ 是整数。$\left(1+\frac{2}{N}\right)^N$ 的二项展开（binomial expansion）中每一项都是正数，而且共有 $N+1\geq 4$ 项，所以总和至少等于首四项之和：

$$\begin{aligned}\left(1+\frac{2}{N}\right)^N &\geq 1+N\cdot\frac{2}{N}+\frac{N(N-1)}{2}\cdot\frac{4}{N^2}+\frac{N(N-1)(N-2)}{6}\cdot\frac{8}{N^3}\\ &= 3+2\left(1-\frac{1}{N}\right)+\frac{4}{3}\left(1-\frac{3}{N}+\frac{2}{N^2}\right)\\ &= \frac{19}{3}-\frac{6}{N}+\frac{8}{3N^2}.\end{aligned}$$

想一想 这个四项之和只是一个下界。我们怎样证明它大于 $6-\frac{8}{N+2}$？

我们把两者相减，再乘以正的分母 $3N^2(N+2)$ 以消去分母：

$$\begin{aligned}&3N^2(N+2)\left[\frac{19}{3}-\frac{6}{N}+\frac{8}{3N^2}-6+\frac{8}{N+2}\right]\\ &\quad= N^2(N+2)-18N(N+2)+8(N+2)+24N^2\\ &\quad= N^3+8N^2-28N+16.\end{aligned}$$

当 $N\geq 3$ 时，我们有 $N^2+8N-28\geq 9+24-28=5>0$，所以 $N^3+8N^2-28N+16=N(N^2+8N-28)+16>0$。因此

$$6-\frac{8}{n+3} < \frac{19}{3}-\frac{6}{N}+\frac{8}{3N^2} \leq \left(1+\frac{2}{n+1}\right)^{n+1},$$

即题目给出的不等式对每个整数 $n>1$ 都成立，所以 X 的中位数大于 μ。

第 (iii) 部分

在 $(0,1)$ 上,

$$f'(x) = k\left(nx^{n-1} - (n+1)x^n\right) = kx^{n-1}(n-(n+1)x),$$

当 $x < \frac{n}{n+1}$ 时它是正数，当 $x > \frac{n}{n+1}$ 时它是负数。所以 f 先升后降，众数为 $M = \frac{n}{n+1}$。

与第 (ii) 部分一样,$M > m$ 当且仅当 $F(M) > \frac{1}{2}$。由于 $(n+2) - (n+1)M = (n+2) - n = 2$,

$$F(M) = 2\left(\frac{n}{n+1}\right)^{n+1},$$

所以我们需要 $\left(\frac{n}{n+1}\right)^{n+1} > \frac{1}{4}$，即

$$\left(1+\frac{1}{n}\right)^{n+1} < 4.$$

> ? **想一想** 既然 n 可以是 2 或以上的任何整数，题目给出关于 $\left(1+\frac{1}{x}\right)^{x+1}$ 的事实怎样帮到我们?

由于该函数在 x 为正数时递减，它在任何整数 $n \geq 2$ 处的值都不大于它在 $x = 2$ 处的值:

$$\left(1+\frac{1}{n}\right)^{n+1} \leq \left(\frac{3}{2}\right)^3 = \frac{27}{8} < 4.$$

因此 $F(M) > \frac{1}{2}$，即 X 的众数大于其中位数。

综合各部分，对每个 $n > 1$ 都有 平均值 < 中位数 < 众数。

验算

- 取 $n = 2$: 平均值为 $\frac{3}{5} = 0.6$, 众数为 $\frac{2}{3} \approx 0.667$。以数值方法解 $4x^3 - 3x^4 = \frac{1}{2}$, 得中位数 ≈ 0.6143, 恰好介乎两者之间。此外, $F(0.6) = 0.4752 < \frac{1}{2}$, 而 $F(\frac{2}{3}) = \frac{16}{27} \approx 0.593 > \frac{1}{2}$。
- 当 $n = 2$（$N = 3$）时，展开式刚好有四项，所以下界取等号：两边都等于 $\frac{125}{27} \approx 4.63$, 稳稳大于 $6 - \frac{8}{5} = 4.4$。
- 平均值 < 中位数 < 众数这个大小次序，已用电脑对 2 至 79 的每个 n 验证过。

常见错误

- 试图明确求出中位数：方程 $x^{n+1}(n+2-(n+1)x) = \frac{1}{2}$ 没有简洁的解，而你也不需要求出它。

- 弄错不等号方向：由于 F 递增，$\mu < m$ 对应的是 $\mathrm{F}(\mu) < \frac{1}{2}$。
- 忘记说明为何可以舍去后面的二项式项（因为它们全是正数）。

STEP 之外

这个密度函数是一个贝塔分布（Beta distribution）Beta$(n+1, 2)$，它是 $[0,1]$ 上最重要的分布族之一，在贝叶斯统计（Bayesian statistics）中处于核心地位。平均值 < 中位数 < 众数这个次序，是左边有长尾（负偏态，negative skew）的分布的典型特征，但它并非普遍定律，所以这里必须加以证明。第 (iii) 部分中的递减函数在 $x \to \infty$ 时趋向 e，这是一个经典极限，你将会在分析（analysis）课程中正式学到。Blitzstein 与 Hwang 的 *Introduction to Probability*（《概率导论》）对贝塔分布有出色的论述。

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。