详细解答

STEP II 2019，第 12 题

概率与统计 · 连续分布与分位数

STEP II 2019，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 2019 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

这里的累积分布函数（cumulative distribution function）在 $[0,1]$ 上就是 x^n，所以每个分位数（quantile）都可以准确写出：分布中有比例 p 落在其下方的值是 $p^{1/n}$。于是，比较平均值与某个分位数就变成比较幂，而把两边同时取 n 次方后，问题便化为关于 $\left(1+\frac{1}{n}\right)^n$ 的命题，这正是二项式展开（binomial expansion）最擅长估计的量。

解答

第 (i) 部分

对概率密度函数（probability density function）积分，

$$\mu = \mathrm{E}(X) = \int_0^1 x \cdot nx^{n-1}\,\mathrm{d}x = \int_0^1 nx^n\,\mathrm{d}x = \frac{n}{n+1},$$
$$\mathrm{E}(X^2) = \int_0^1 nx^{n+1}\,\mathrm{d}x = \frac{n}{n+2}.$$

因此

$$\sigma^2 = \mathrm{E}(X^2) - \mu^2 = \frac{n}{n+2} - \frac{n^2}{(n+1)^2} = \frac{n\left[(n+1)^2 - n(n+2)\right]}{(n+1)^2(n+2)}$$
$$= \frac{n\left[n^2+2n+1-n^2-2n\right]}{(n+1)^2(n+2)} = \frac{n}{(n+1)^2(n+2)},$$

得证。

第 (ii) 部分

当 $0 \le x \le 1$ 时，分布函数为

$$\mathrm{F}(x) = \mathrm{P}(X \le x) = \int_0^x nt^{n-1}\,\mathrm{d}t = x^n.$$

取 $n = 2$, 则 $\mathrm{F}(x) = x^2$。下四分位数 (lower quartile) q_1 满足 $q_1^2 = \frac{1}{4}$，上四分位数（upper quartile）q_3 满足 $q_3^2 = \frac{3}{4}$；两者都在 $[0,1]$ 内，所以取正根：

$$q_1 = \tfrac{1}{2}, \qquad q_3 = \tfrac{\sqrt{3}}{2}, \qquad q_3 - q_1 = \frac{\sqrt{3}-1}{2}.$$

由第 (i) 部分，$\sigma^2 = \dfrac{2}{9\cdot 4} = \dfrac{1}{18}$，所以 $\sigma = \dfrac{1}{3\sqrt{2}} = \dfrac{\sqrt{2}}{6}$，而 $2\sigma = \dfrac{\sqrt{2}}{3}$。

我们要证明 $\dfrac{\sqrt{3}-1}{2} < \dfrac{\sqrt{2}}{3}$。两边乘以 6，这等价于

$$3\sqrt{3} - 3 < 2\sqrt{2} \iff 3\sqrt{3} < 3 + 2\sqrt{2}.$$

最后这个不等式两边都是正数，所以它等价于两边平方后的不等式：

$$27 < 9 + 12\sqrt{2} + 8 = 17 + 12\sqrt{2} \iff 10 < 12\sqrt{2},$$

这是成立的，因为 $\sqrt{2} > 1$ 给出 $12\sqrt{2} > 12 > 10$。每一步都是等价的，所以沿着这条推理链倒推回去，可知四分位数间距（interquartile range）小于 2σ。

> ? **想一想** 为甚么在这里两边平方是安全的，但一般情况下却不是?

只有当 a 和 b 都非负时，平方才能保持不等式 $a < b$（例如 $-3 < 2$，但 $9 > 4$）。正因如此，我们先把不等式整理成 $3\sqrt{3} < 3 + 2\sqrt{2}$，使两边明显都是正数，而不是直接把仍含有减法的 $3\sqrt{3} - 3 < 2\sqrt{2}$ 平方。

第 (iii) 部分

由二项式定理,

$$(1+x)^n = 1 + nx + \frac{n(n-1)}{2}\,x^2 + \cdots,$$

而第 $(k+1)$ 项是

$$\binom{n}{k} x^k = \frac{n(n-1)\cdots(n-k+1)}{k!}\,x^k.$$

中位数与下四分位数。 由于 $\mathrm{F}(x) = x^n$，中位数（median）m 满足 $m^n = \frac{1}{2}$，下四分位数 q 满足 $q^n = \frac{1}{4}$，所以

$$m = \left(\tfrac{1}{2}\right)^{1/n}, \qquad q = \left(\tfrac{1}{4}\right)^{1/n}.$$

μ、m、q 全都是正数，而 $t \mapsto t^n$ 在 $t > 0$ 上严格递增（strictly increasing），所以我们可以通过比较它们的 n 次方来比较它们。另外

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

记 $A = \left(1+\frac{1}{n}\right)^n$。那么

$$\mu < m \iff \mu^n < \tfrac{1}{2} \iff A > 2, \qquad \mu > q \iff \mu^n > \tfrac{1}{4} \iff A < 4.$$

μ 小于中位数。 在展开式中代入 $x = \frac{1}{n}$。每一项都非负，所以 A 至少等于首三项之和：

$$A \geq 1 + n \cdot \frac{1}{n} + \frac{n(n-1)}{2} \cdot \frac{1}{n^2} = 2 + \frac{n-1}{2n} > 2,$$

其中最后一个不等式是严格的，因为 $n > 1$。（当 $n \geq 2$ 时，展开式中确实有第三项。）因此 $\mu < m$。

μ 大于下四分位数。 取 $x = \frac{1}{n}$，第 $(k+1)$ 项是

$$\binom{n}{k}\frac{1}{n^k} = \frac{1}{k!} \cdot \frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-k+1}{n} \leq \frac{1}{k!},$$

因为那 k 个分数 $\frac{n-j}{n}$（$0 \leq j \leq k-1$）每个都在 0 与 1 之间。（当 $k = 0$ 时，该项是 $1 = \frac{1}{0!}$。）对 $k = 0, 1, \ldots, n$ 求和，

$$A \leq 1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} < 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots < 4,$$

中间的不等式成立，是因为无穷级数（infinite series）还包含更多正项；最后一个不等式则是题目容许我们假设的结果。因此 $A < 4$，所以 $\mu > q$。

综合两者，对每个整数 $n > 1$，下四分位数都小于 μ，而 μ 又小于中位数。

> ? **想一想** 为甚么比较 n 次方就能确定 μ 与 m 的大小关系?

对正数 a 和 b，$a < b$ 若且唯若 $a^n < b^n$，因为 t^n 在 $t > 0$ 上严格递增。我们的三个数全是正数，所以取 n 次方不会损失任何信息，而且这样做能消去 m 和 q 中麻烦的 n 次方根。

验算

- 当 $n = 2$：$\mu = \frac{2}{3}$，而 $\sigma^2 = \frac{2}{9 \cdot 4} = \frac{1}{18}$，与 $\mathrm{E}(X^2) - \mu^2 = \frac{1}{2} - \frac{4}{9}$ 吻合。
- 数值上，当 $n = 2$ 时，四分位数间距是 0.366，而 $2\sigma = 0.471$。

- 当 $n = 2$ 时，三个值分别是 $q = 0.5$、$\mu \approx 0.667$、$m \approx 0.707$；当 $n = 10$ 时，它们是 $q \approx 0.871$、$\mu \approx 0.909$、$m \approx 0.933$。对 2 至 2000 的每个 n，顺序 $q < \mu < m$ 都已用电脑验证。
- 当 $n \to \infty$ 时，$A = \left(1 + \frac{1}{n}\right)^n \to \mathrm{e} \approx 2.718$，稳稳地落在 2 与 4 之间。

常见错误

- 没有先确认两边都是正数，就直接把 $3\sqrt{3} - 3 < 2\sqrt{2}$ 平方。
- 取倒数时弄错不等号方向：$\mu^n < \frac{1}{2}$ 对应的是 $A > 2$，而不是 $A < 2$。
- 声称第 $(k+1)$ 项等于 $\frac{1}{k!}$，而不是至多为 $\frac{1}{k!}$。

STEP 之外

$\left(1 + \frac{1}{n}\right)^n$ 随 n 递增并趋向极限 e；而你逐项与 $\sum 1/k!$ 比较所证出的界，正是证明 $\mathrm{e} = \sum_{k \geq 0} 1/k!$ 的第一步，这是大学第一门数学分析（analysis）课程中的标准结果。这个分布本身是 Beta$(n, 1)$ 分布（Beta distribution）；它也是 n 个在 $[0, 1]$ 上独立均匀分布的随机变量之最大值的分布，因为 $\mathrm{P}(\max \leq x) = x^n$。想深入了解这两条线索，可参阅 Blitzstein 与 Hwang 的 *Introduction to Probability*（《概率导论》），以及 Abbott 的 *Understanding Analysis*（《理解分析》）。

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。