详细解答

STEP III 2018，第 12 题

概率与统计 · 顺序统计量

STEP III 2018，第 12 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/18-s3-q12

关键思路

事件“第 k 小的数不大于 y”与事件“n 个数之中至少有 k 个不大于 y”是同一回事，而不大于 y 的数的个数服从二项分布（binomial distribution）。由此可以立即得出分布函数（distribution function）。将它求导会得到两个和式，而 (ii) 中的两条二项式恒等式正好令它们相消（telescope）。一旦知道了密度函数（density），由于密度函数的积分等于 1，便可以顺手求出一个积分；而同一个积分（把 n 和 k 各自平移）就给出期望值。

解答

把这 n 个数记为 $X_1, \ldots, X_n$。它们互相独立（independent），每个都在 $[0,1]$ 上服从均匀分布（uniform distribution）。

第 (i) 部分

固定 y，其中 $0 \le y \le 1$，并设 N 为满足 $X_i \le y$ 的 X_i 的个数。每个 X_i 满足 $X_i \le y$ 的概率为 y，且与其他各数独立，所以 $N \sim \mathrm{B}(n, y)$。

$Y_k \le y$ 成立，当且仅当这些数之中至少有 k 个不大于 y。若 $Y_k \le y$，则 Y_k 以及比它小的 $k-1$ 个数都不大于 y，所以 $N \ge k$。反过来，若 $N \ge k$，则最小的 k 个数都不大于 y，特别地 $Y_k \le y$。因此

$$\mathrm{P}(Y_k \le y) = \mathrm{P}(N \ge k) = \sum_{m=k}^{n} \binom{n}{m} y^m (1-y)^{n-m},$$

这就是 $(*)$。

第 (ii) 部分

把阶乘写出来，对 $1 \le m \le n$，

$$m\binom{n}{m} = \frac{n!}{(m-1)!\,(n-m)!} = n \cdot \frac{(n-1)!}{(m-1)!\,((n-1)-(m-1))!} = n\binom{n-1}{m-1},$$

同理，对 $0 \le m \le n-1$，

$$(n-m)\binom{n}{m} = \frac{n!}{m!\,(n-m-1)!} = n \cdot \frac{(n-1)!}{m!\,((n-1)-m)!} = n\binom{n-1}{m}.$$

（当 $m = n$ 时，左边为 0；若把 $\binom{n-1}{n}$ 理解为 0，右边亦为 0。）

密度函数。 Y_k 的密度函数 $f(y)$ 是 $(*)$ 的导数。用乘积法则（product rule）逐项求导，

$$f(y) = \sum_{m=k}^{n} \binom{n}{m} \Big[m\,y^{m-1}(1-y)^{n-m} - (n-m)\,y^{m}(1-y)^{n-m-1} \Big].$$

在第二部分中，由于因子 $n-m$，$m = n$ 的项等于零，所以该和式实际上只到 $m = n-1$。利用上述两条恒等式，

$$f(y) = n \sum_{m=k}^{n} \binom{n-1}{m-1} y^{m-1}(1-y)^{n-m} \;-\; n \sum_{m=k}^{n-1} \binom{n-1}{m} y^{m}(1-y)^{n-1-m}.$$

? **想一想** 这两个和式看起来几乎一样。怎样才能把它们对齐?

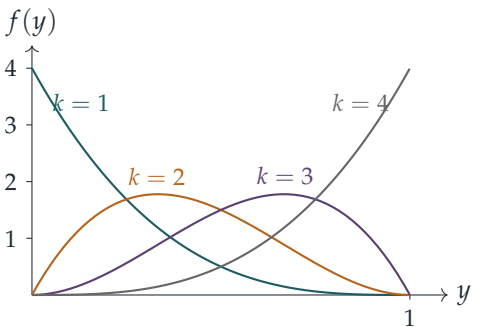
在第二个和式中令 $j = m+1$。它变成

$$n \sum_{j=k+1}^{n} \binom{n-1}{j-1} y^{j-1}(1-y)^{n-j},$$

这与第一个和式逐项相同，只是从 $k+1$ 而不是从 k 开始。除了第一个和式中 $m = k$ 的项之外，其余全部抵消：

$$f(y) = n\binom{n-1}{k-1} y^{k-1}(1-y)^{n-k},$$

正如所求。（若 $k = n$，第二个和式是空和，第一个和式只有 $m = n$ 一项，所以结果依然成立。）



$n = 4$ 时的密度函数：四个数之中第 k 小的数集中在 $k/5$ 附近。

积分。 密度函数在其取值范围上的积分等于 1，这里的范围是 $[0,1]$。因此

$$n\binom{n-1}{k-1}\int_0^1 y^{k-1}(1-y)^{n-k}\,\mathrm{d}y = 1,$$

所以

$$\int_0^1 y^{k-1}(1-y)^{n-k}\,\mathrm{d}y = \frac{1}{n\binom{n-1}{k-1}} = \frac{(k-1)!\,(n-k)!}{n!}.$$

此式对所有整数 $1 \le k \le n$ 都成立。

第 (iii) 部分

按定义，

$$\mathrm{E}(Y_k) = \int_0^1 y\,f(y)\,\mathrm{d}y = n\binom{n-1}{k-1}\int_0^1 y^{k}(1-y)^{n-k}\,\mathrm{d}y.$$

> **想一想** 这个积分与刚才求得的并不完全一样。取哪些 n 和 k 的值，才能把它化成那个形式？

被积函数（integrand）是 $y^{K-1}(1-y)^{N-K}$，其中 $K = k+1$、$N = n+1$，且 $1 \le K \le N$，所以 (ii) 的公式适用：

$$\int_0^1 y^{k}(1-y)^{n-k}\,\mathrm{d}y = \frac{k!\,(n-k)!}{(n+1)!}.$$

由于 $n\binom{n-1}{k-1} = \dfrac{n!}{(k-1)!\,(n-k)!}$，

$$\mathrm{E}(Y_k) = \frac{n!}{(k-1)!\,(n-k)!}\cdot\frac{k!\,(n-k)!}{(n+1)!} = \frac{k}{n+1}.$$

验算

- 当 $n = 1$、$k = 1$ 时，只有一个均匀分布的数，而 $\mathrm{E}(Y_1) = \frac{1}{2}$，理应如此。
- 对称性（symmetry）：把每个 X_i 换成 $1 - X_i$，第 k 小的数就变成第 k 大的数，所以 $\mathrm{E}(Y_k) + \mathrm{E}(Y_{n+1-k}) = 1$。确实 $\frac{k}{n+1} + \frac{n+1-k}{n+1} = 1$。
- 这 n 个数把 $[0,1]$ 分成 $n+1$ 段间隔，而答案表明每段间隔的平均长度都是 $\frac{1}{n+1}$，从对称性看来非常合理。
- 当 $n = 5$、$k = 2$ 时：$(*)$ 在 $y = 0.3$ 处的值为 0.47178；用电脑模拟 200 000 个样本，得到 0.4726，而 Y_2 的样本平均值为 0.33333，理论值为 $\frac{1}{3}$。

常见错误

- 写成 $\mathrm{P}(Y_k \leq y) = \binom{n}{k} y^k (1-y)^{n-k}$，即恰好有 k 个值小于 y 的概率，而不是至少有 k 个。
- 忘记第二个和式中 $m = n$ 的项等于零，以致相消看似差了一项而不成立。
- 在 (iii) 中重用积分时只平移 k（而没有平移 n）。

STEP 之外

这些 Y_k 是样本的顺序统计量（order statistics），而 Y_k 服从参数为 k 和 $n-k+1$ 的贝塔分布（Beta distribution）。(ii) 中的积分就是贝塔函数 $B(a,b) = \int_0^1 y^{a-1}(1-y)^{b-1}\,\mathrm{d}y = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$，你日后会与伽玛函数（Gamma function）一并再遇见它。顺序统计量是样本中位数和无母数统计（non-parametric statistics）的基础。Blitzstein 和 Hwang 的 *Introduction to Probability*（《概率导论》）对顺序统计量和贝塔分布有十分精彩的论述，其中包括贝塔积分的一个“故事式”证明，与这里的论证非常相似。

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。