詳細解答

STEP III 2018，第 12 題

概率與統計 • 順序統計量

> **STEP III 2018，第 12 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：
>
> step.maths.org/questions/18-s3-q12

關鍵思路

事件「第 k 小的數不大於 y」與事件「n 個數之中至少有 k 個不大於 y」是同一回事，而不大於 y 的數的個數服從二項分佈（binomial distribution）。由此可以立即得出分佈函數（distribution function）。將它求導會得到兩個和式，而 (ii) 中的兩條二項式恆等式正好令它們相消（telescope）。一旦知道了密度函數（density），由於密度函數的積分等於 1，便可以順手求出一個積分；而同一個積分（把 n 和 k 各自平移）就給出期望值。

解答

把這 n 個數記為 $X_1, \ldots, X_n$。它們互相獨立（independent），每個都在 $[0,1]$ 上服從均勻分佈（uniform distribution）。

第 (i) 部分

固定 y，其中 $0 \le y \le 1$，並設 N 為滿足 $X_i \le y$ 的 X_i 的個數。每個 X_i 滿足 $X_i \le y$ 的概率為 y，且與其他各數獨立，所以 $N \sim \mathrm{B}(n, y)$。

$Y_k \le y$ 成立，當且僅當這些數之中至少有 k 個不大於 y。若 $Y_k \le y$，則 Y_k 以及比它小的 $k-1$ 個數都不大於 y，所以 $N \ge k$。反過來，若 $N \ge k$，則最小的 k 個數都不大於 y，特別地 $Y_k \le y$。因此

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

這就是 $(*)$。

第 (ii) 部分

把階乘寫出來，對 $1 \le m \le n$，

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

同理，對 $0 \le m \le n-1$，

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

（當 $m = n$ 時，左邊為 0；若把 $\binom{n-1}{n}$ 理解為 0，右邊亦為 0。）

密度函數。 Y_k 的密度函數 $f(y)$ 是 $(*)$ 的導數。用乘積法則（product rule）逐項求導，

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

在第二部分中，由於因子 $n-m$，$m=n$ 的項等於零，所以該和式實際上只到 $m = n-1$。利用上述兩條恆等式，

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

? **想一想** 這兩個和式看起來幾乎一樣。怎樣才能把它們對齊？

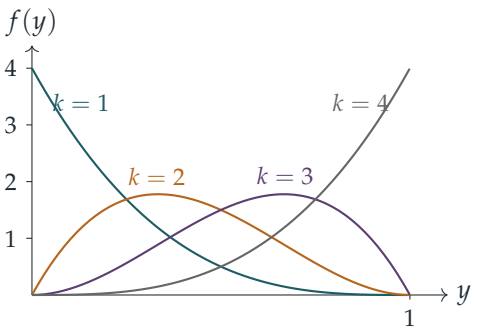
在第二個和式中令 $j = m+1$。它變成

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

這與第一個和式逐項相同，只是從 $k+1$ 而不是從 k 開始。除了第一個和式中 $m = k$ 的項之外，其餘全部抵消：

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

正如所求。（若 $k = n$，第二個和式是空和，第一個和式只有 $m = n$ 一項，所以結果依然成立。）



$n = 4$ 時的密度函數：四個數之中第 k 小的數集中在 $k/5$ 附近。

積分。 密度函數在其取值範圍上的積分等於 1，這裏的範圍是 $[0,1]$。因此

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

所以

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

此式對所有整數 $1 \le k \le n$ 都成立。

第 (iii) 部分

按定義，

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

> **想一想** 這個積分與剛才求得的並不完全一樣。取哪些 n 和 k 的值，才能把它化成那個形式？

被積函數（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) 的公式適用：

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

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

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

驗算

- 當 $n = 1$、$k = 1$ 時，只有一個均勻分佈的數，而 $\mathrm{E}(Y_1) = \frac{1}{2}$，理應如此。
- 對稱性（symmetry）：把每個 X_i 換成 $1 - X_i$，第 k 小的數就變成第 k 大的數，所以 $\mathrm{E}(Y_k) + \mathrm{E}(Y_{n+1-k}) = 1$。確實 $\frac{k}{n+1} + \frac{n+1-k}{n+1} = 1$。
- 這 n 個數把 $[0,1]$ 分成 $n+1$ 段間隔，而答案表明每段間隔的平均長度都是 $\frac{1}{n+1}$，從對稱性看來非常合理。
- 當 $n = 5$、$k = 2$ 時：$(*)$ 在 $y = 0.3$ 處的值為 0.47178；用電腦模擬 200 000 個樣本，得到 0.4726，而 Y_2 的樣本平均值為 0.33333，理論值為 $\frac{1}{3}$。

常見錯誤

- 寫成 $\mathrm{P}(Y_k \leq y) = \binom{n}{k} y^k (1-y)^{n-k}$，即恰好有 k 個值小於 y 的概率，而不是至少有 k 個。
- 忘記第二個和式中 $m = n$ 的項等於零，以致相消看似差了一項而不成立。
- 在 (iii) 中重用積分時只平移 k（而沒有平移 n）。

STEP 之外

這些 Y_k 是樣本的順序統計量（order statistics），而 Y_k 服從參數為 k 和 $n-k+1$ 的貝塔分佈（Beta distribution）。(ii) 中的積分就是貝塔函數 $B(a,b) = \int_0^1 y^{a-1}(1-y)^{b-1}\,\mathrm{d}y = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$，你日後會與伽瑪函數（Gamma function）一併再遇見它。順序統計量是樣本中位數和無母數統計（non-parametric statistics）的基礎。Blitzstein 和 Hwang 的 *Introduction to Probability*（《概率導論》）對順序統計量和貝塔分佈有十分精彩的論述，其中包括貝塔積分的一個「故事式」證明，與這裏的論證非常相似。

Beyond STEP 詳細解答：https://beyondstep.keithmcnulty.org/zh-hant/。解答文字採用 CC0 1.0；題目本身版權屬 Cambridge University Press & Assessment 所有（©），本文件不予轉載。由機器從英文翻譯而成；尚未經審閱。