詳細解答

STEP I 1996，第 12 題

概率與統計 • 隨機樣本的最大值

> **STEP I 1996，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/96-s1-q12

關鍵思路

閱卷員的選擇，歸根究底就是從 $\{1,2,\ldots,m\}$ 中選出一個由 n 個不同分數組成的集合，而每一個這樣的集合出現的機會均等。因此，概率（probability）就是集合的數目除以 $\binom{m}{n}$。說最高分數是 k，即是固定了其中一個分數，餘下的 $n-1$ 個則要從比它小的分數中選出。恆等式（identity）隨之唾手可得，因為各概率之和為 1。至於期望值（expectation），竅門是把 $k\binom{k-1}{n-1}$ 還原成單獨一個二項式係數（binomial coefficient），這樣便可以再次使用同一條恆等式。

解答

K 的分佈

閱卷員給出 n 個互不相同的分數，而根據對稱性，$\{1,\ldots,m\}$ 中每一個大小為 n 的子集（subset），成為所用分數集合的機會均等。（哪一份答卷得到哪一個分數，對 K 並無影響。）這樣的子集共有 $\binom{m}{n}$ 個。

事件 $K=k$ 發生，當且僅當該集合包含 k，而其餘 $n-1$ 個分數全部落在 $\{1,\ldots,k-1\}$ 之內。選出這 $n-1$ 個分數的方法有 $\binom{k-1}{n-1}$ 種。這需要 $k-1\geq n-1$，即 $k\geq n$：n 個互不相同的正整數之中，最大的一個至少是 n。此外，當然還有 $k\leq m$。



（圖中所示為 $m=10$、$n=4$、$k=7$ 的情況。）因此

$$\mathrm{P}(K=k)=\binom{k-1}{n-1}\Big/\binom{m}{n},\qquad n\le k\le m.$$

恆等式

K 必定恰好取 $n, n+1, \ldots, m$ 其中一個值，所以這些概率之和為 1：

$$\sum_{k=n}^{m}\binom{k-1}{n-1}\Big/\binom{m}{n}=1\quad\Longrightarrow\quad\sum_{k=n}^{m}\binom{k-1}{n-1}=\binom{m}{n}.$$

這對所有滿足 $1\le n\le m$ 的整數都成立，下一步正要用到這一點。

期望值

根據定義，

$$\mathrm{E}(K)=\frac{1}{\binom{m}{n}}\sum_{k=n}^{m}k\binom{k-1}{n-1}.$$

? **想一想** 怎樣才能消去因子 k，使恆等式可以再次派上用場？

把它併入二項式係數之中：

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

現在 $\binom{k}{n}=\binom{(k+1)-1}{(n+1)-1}$。當 k 由 n 取到 m 時，$j=k+1$ 由 $n+1$ 取到 $m+1$。於是，以 $n+1$ 和 $m+1$ 代替 n 和 m 來使用恆等式（這是容許的，因為 $1\le n+1\le m+1$），便得

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

所以

$$\begin{aligned}\mathrm{E}(K)&=\frac{n\binom{m+1}{n+1}}{\binom{m}{n}}=n\cdot\frac{(m+1)!}{(n+1)!\,(m-n)!}\cdot\frac{n!\,(m-n)!}{m!}\\&=\frac{n(m+1)}{n+1}.\end{aligned}$$

驗算

- $n=m$：每個分數都用上了，所以必定有 $K=m$，而確實 $\frac{m(m+1)}{m+1}=m$。
- $n=1$：K 是單獨一個分數，在 $\{1,\ldots,m\}$ 上均勻分佈，平均值為 $\frac{m+1}{2}$，與公式吻合。

- $m = 6$、$n = 2$：列出 $\{1,\ldots,6\}$ 中全部 15 對，得 $\mathrm{E}(K) = \frac{70}{15} = \frac{14}{3}$，而公式給出 $\frac{2\cdot 7}{3} = \frac{14}{3}$。我們已就所有 $1 \le n \le m \le 9$ 逐一列出每個子集，驗證了分佈、恆等式和平均值。
- 有一個漂亮的方法可以看出答案：選出的 n 個分數把餘下的 $m-n$ 個分數分成 $n+1$ 個間隔，各間隔的期望大小相等，均為 $\frac{m-n}{n+1}$。於是 $\mathrm{E}(K) = m - \frac{m-n}{n+1} = \frac{n(m+1)}{n+1}$。

常見錯誤

- 分子數的是把分數分配給答卷的有序方式，分母數的卻是無序的集合（或者反過來）。只要前後一致，兩種數法都可行，因為因子 $n!$ 會互相抵消。
- 用了 $\binom{k}{n}$ 而不是 $\binom{k-1}{n-1}$。前者數的是所有分數都*不超過* k 的集合，而不是最高分數*等於* k 的集合。

STEP 之外

這條恆等式就是帕斯卡三角形（Pascal's triangle）中的「曲棍球棒」恆等式（hockey-stick identity），而我們在這裏是用兩種方法數同一樣東西來證明它，這是組合數學中的標準技巧。K 是不放回抽取的隨機樣本中的最大值，屬於*順序統計量*（order statistic）。「德國坦克問題」（German tank problem）把本題倒過來，由觀察所得的 K 去估計 m：從 $\mathrm{E}(K) = \frac{n(m+1)}{n+1}$ 解出 m，便得到無偏估計量（unbiased estimator）$\frac{n+1}{n}K - 1$。想進一步了解順序統計量和計數論證，可參閱 Blitzstein 與 Hwang 的 *Introduction to Probability*（概率論入門教材）。

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。