解答例

STEP I 1996、第 12 問

確率・統計 ・ 無作為標本の最大値

STEP I 1996、問 12

先に進む前に、まず問題を読んでおこう。問題は STEP Support Programme の問題データベースにある：

step.maths.org/questions/96-s1-q12

鍵となる考え方

採点者の選び方は、$\{1,2,\ldots,m\}$ から相異なる n 個の点数からなる集合（set）を選ぶことに帰着し、そのような集合はどれも同様に確からしい。したがって、確率は集合の個数を $\binom{m}{n}$ で割ったものになる。最大の点数が k であるということは、点数を一つ確定させ、残りの $n-1$ 個をそれより小さいものの中から選ぶということである。すると、確率の総和が 1 であることから、恒等式（identity）はおのずと得られる。期待値（expectation）については、$k\binom{k-1}{n-1}$ を一つの二項係数（binomial coefficient）に書き直し、同じ恒等式をもう一度使えるようにするのが要点である。

解答

K の分布

採点者は相異なる n 個の点数をつけるので、対称性により、$\{1,\ldots,m\}$ の要素数 n の部分集合はどれも、使われる点数の集合として同様に確からしい。（どの答案にどの点数がつくかは K に影響しない。）そのような部分集合は $\binom{m}{n}$ 個ある。

事象 $K=k$ が起こるのは、集合が k を含み、かつ残りの $n-1$ 個の点数がすべて $\{1,\ldots,k-1\}$ に属するとき、かつそのときに限る。その $n-1$ 個の点数の選び方は $\binom{k-1}{n-1}$ 通りである。これには $k-1\geq n-1$、すなわち $k\geq n$ が必要である。相異なる n 個の正の整数の最大値は n 以上だからである。また、当然 $k\leq m$ である。



（図は $m=10$、$n=4$、$k=7$ の場合である。）よって

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

恒等式

K は $n, n+1, \ldots, m$ のうちちょうど一つの値をとるから、これらの確率の総和は 1 である：

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

これは $1\le n\le m$ を満たす**すべての**整数について成り立ち、次の計算で必要になる。

期待値

定義により

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

> ? **考えてみよう** 恒等式が再び使えるように、因数 k を消すにはどうすればよいか。

二項係数の中に吸収させればよい：

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

ここで $\binom{k}{n}=\binom{(k+1)-1}{(n+1)-1}$ である。k が n から m まで動くとき、$j=k+1$ は $n+1$ から $m+1$ まで動く。そこで、恒等式を n、m の代わりに $n+1$、$m+1$ として用いると（$1\le n+1\le m+1$ なので許される）、

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

したがって

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

検算

- $n = m$ のとき：すべての点数が使われるので、確実に $K = m$ であり、実際 $\frac{m(m+1)}{m+1} = m$ となる。
- $n = 1$ のとき：K はただ一つの点数で、$\{1, \ldots, m\}$ 上の一様分布に従い、平均は $\frac{m+1}{2}$ となって一致する。
- $m = 6, n = 2$ のとき：$\{1, \ldots, 6\}$ から選ぶ 15 組をすべて書き出すと $\mathrm{E}(K) = \frac{70}{15} = \frac{14}{3}$ となり、公式からも $\frac{2 \cdot 7}{3} = \frac{14}{3}$ が得られる。また、$1 \leq n \leq m \leq 9$ のすべての場合について、すべての部分集合を書き出して分布、恒等式、平均を確かめた。
- 答えのうまい見方：選ばれた n 個の点数は、残りの $m - n$ 個の点数を $n + 1$ 個のすき間に分け、それぞれのすき間の大きさの期待値は等しく $\frac{m-n}{n+1}$ である。よって $\mathrm{E}(K) = m - \frac{m-n}{n+1} = \frac{n(m+1)}{n+1}$ となる。

よくある間違い

- 分子では答案への点数の割り当てを順序つきで数え、分母では順序なしの集合を数えてしまう（あるいはその逆）。一貫していればどちらの数え方でもよい。因数 $n!$ が約分されるからである。
- $\binom{k-1}{n-1}$ の代わりに $\binom{k}{n}$ を使ってしまう。これは点数がすべて k **以下**である集合の個数であり、最大の点数が k に**等しい**集合の個数ではない。

STEP のその先へ

この恒等式はパスカルの三角形（Pascal's triangle）の「ホッケースティック恒等式」（hockey-stick identity）であり、ここでは同じものを二通りに数えるという、組合せ論の標準的な手法で証明した。K は非復元抽出による無作為標本の最大値であり、**順序統計量**（order statistic）の一例である。「ドイツ戦車問題」（German tank problem）はこの問題を逆にして、観測された K から m を推定するものである。$\mathrm{E}(K) = \frac{n(m+1)}{n+1}$ を m について解くと、不偏推定量（unbiased estimator）$\frac{n+1}{n}K - 1$ が得られる。順序統計量や数え上げによる議論についてさらに知りたければ、Blitzstein and Hwang, *Introduction to Probability*（確率論入門）を参照されたい。

Beyond STEP 解答集：https://beyondstep.keithmcnulty.org/ja/。解答本文は CC0 1.0 である。問題そのものは © Cambridge University Press & Assessment に帰属し、ここには掲載していない。英語版から機械翻訳したものであり、まだ校閲を経ていない。