解答例

STEP II 1988、第 15 問

確率・統計 ・ 離散一様分布の和

> **STEP II 1988、問 15**
>
> 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/88-s2-q15

鍵となる考え方

各試験の得点は $\{0, 1, \ldots, m\}$ 上の一様分布（uniform distribution）に従うので、得点の組はどれも等しい確率で現れる。したがって確率は、条件を満たす得点の組の個数を組の総数で割ったものにすぎない。2 科目の場合の個数は単純に数え上げれば求まり、3 科目の場合の個数は 3 科目目の得点で条件付け（conditioning）をして 2 科目の場合の個数を足し合わせれば求まる（これを**畳み込み**（*convolution*）という）。さらに、各試験での対称性（symmetry）$k \mapsto m - k$ を使えば、最も上の範囲はほとんど労せずに得られる。

始める前に一つ注意しておく。問題文に印刷されている 3 科目の公式には誤植（misprint）がある。以下では正しい公式を導き、印刷されたものとどこが異なるかを指摘する。

解答

各試験の得点を X_1, X_2, X_3 とする。試験は互いに独立に採点されるので、これらは独立（independent）であり、それぞれ $\{0, 1, \ldots, m\}$ 上の一様分布に従う。よって順序対 (X_1, X_2) はどれも確率 $(m+1)^{-2}$ をもち、順序付きの 3 つ組はどれも確率 $(m+1)^{-3}$ をもつ。$\{0, \ldots, m\}$ の得点からなる順序対（3 つ組）のうち和が n となるものの個数を $N_2(n)$（$N_3(n)$）と書く。

2 科目の場合

$X_1 = k$ かつ $X_2 = n-k$ で、どちらも $[0, m]$ に入っていなければならないから、

$$\max(0,\, n-m) \le k \le \min(m,\, n).$$

- $0 \le n \le m$ のとき：k は 0 から n まで動くので、$N_2(n) = n+1$ である。
- $m < n \le 2m$ のとき：k は $n-m$ から m まで動くので、$N_2(n) = m-(n-m)+1 = 2m-n+1$ である。

したがって

$$P(X_1+X_2=n) = \begin{cases} \dfrac{n+1}{(m+1)^2}, & 0 \le n \le m, \\ \dfrac{2m-n+1}{(m+1)^2}, & m < n \le 2m, \end{cases}$$

となる。これが $m < n \le 2m$ に対して求める結果であり、あわせて $0 \le n \le m$ に対する結果も得られた。この分布は「テント」形であり、$n = m$ まで 1 段ごとに 1 ずつ増え、そのあと減少していく。

3 科目の場合

3 科目目の得点で条件付けをする。$X_3 = k$ ならば $X_1 + X_2 = n-k$ でなければならないので、

$$N_3(n) = \sum_{k=0}^{m} N_2(n-k) = \sum_{j=n-m}^{n} N_2(j),$$

となる。ここで $j = n-k$ であり、$j < 0$ または $j > 2m$ のときは $N_2(j) = 0$ とする。

$0 \le n \le m$ の場合。 寄与するのは $j = 0, 1, \ldots, n$ のみであり（負の j は 0 を与える）、これらはすべてテントの上り坂の部分にある：

$$N_3(n) = \sum_{j=0}^{n}(j+1) = \frac{(n+1)(n+2)}{2}, \qquad P(X_1+X_2+X_3=n) = \frac{(n+1)(n+2)}{2(m+1)^3}.$$

$m < n \le 2m$ の場合。 このとき $1 \le n-m \le m < n \le 2m$ なので、範囲 $j = n-m, \ldots, n$ はテントの頂上をまたいでいる。そこで頂上で分割する：

$$N_3(n) = \sum_{j=n-m}^{m}(j+1) + \sum_{j=m+1}^{n}(2m-j+1).$$

第 1 の和は $n-m+1$ から $m+1$ までの整数の和である：

$$\sum_{j=n-m}^{m}(j+1) = \frac{(m+1)(m+2)-(n-m)(n-m+1)}{2}.$$

第 2 の和では $i = 2m - j + 1$ とおくと、i は $2m - n + 1$ から m まで動く：

$$\sum_{j=m+1}^{n} (2m - j + 1) = \frac{m(m+1) - (2m-n)(2m-n+1)}{2}.$$

4 つの積を展開すると、

$$\begin{aligned} 2N_3(n) &= (m^2 + 3m + 2) - (n^2 - 2mn + m^2 + n - m) \\ &\quad + (m^2 + m) - (4m^2 - 4mn + n^2 + 2m - n) \\ &= 6mn - 3m^2 - 2n^2 + 3m + 2. \end{aligned}$$

等確率の $(m+1)^3$ 通りの 3 つ組の個数で割って、

$$P(X_1 + X_2 + X_3 = n) = \frac{6mn - 3m^2 - 2n^2 + 3m + 2}{2(m+1)^3}, \qquad m < n \le 2m.$$

考えてみよう これは問題文に印刷された公式とは異なる。どちらが正しいのだろうか？

印刷された公式では $-3m^2$ の代わりに $-4m^2$、$(m+1)^3$ の代わりに $(m+1)^2$ となっているが、どちらも誤植である。具体例を一つ調べれば決着がつく。$m = 1$、$n = 2$ のとき、$\{0, 1\}$ から選んだ 3 つの得点の和が 2 となるのは、等確率の 8 通りのうちちょうど 3 通りなので、確率は $\frac{3}{8}$ である。我々の公式は $\frac{12-3-8+3+2}{16} = \frac{3}{8}$ を与えるが、印刷された公式は $\frac{12-4-8+3+2}{8} = \frac{5}{8}$ を与える。そもそも分母は等確率の結果の個数なので、いずれにせよ $(m+1)^3$ でなければならない。さらなる確認として、$n = m$ のとき中央の公式は $\frac{(m+1)(m+2)}{2(m+1)^3}$ を与え、第 1 の場合と一致する。このとき 2 つの和がつながるので、これは当然そうなるべきである。

$2m < n \le 3m$ の場合。 各得点 X_i を $m - X_i$ で置き換える。これもまた $\{0, \ldots, m\}$ 上の一様分布に従い、新しい 3 つの得点も依然として独立であるから、

$$\begin{aligned} P(X_1 + X_2 + X_3 = n) &= P\big((m - X_1) + (m - X_2) + (m - X_3) = n\big) \\ &= P(X_1 + X_2 + X_3 = 3m - n). \end{aligned}$$

$2m < n \le 3m$ のとき $0 \le 3m - n < m$ なので、n を $3m - n$ に置き換えて第 1 の場合を適用できる：

$$P(X_1 + X_2 + X_3 = n) = \frac{(3m - n + 1)(3m - n + 2)}{2(m+1)^3}, \qquad 2m < n \le 3m.$$

（同じ対称性を、$3m - n$ の代わりに $2m - n$ として用いれば、$n \le m$ に対する 2 科目の結果から $n > m$ に対する結果が得られる。）

検算

- 小さい場合：$m = 1$ のとき、3 つの公式は $n = 0, 1, 2, 3$ に対してそれぞれ $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}$ を与える。これらの和は 1 であり、公平な硬貨を 3 回投げたときの二項分布（binomial distribution）と一致する。
- 中央の公式は $n \mapsto 3m - n$ に関して対称である：$6mn - 3m^2 - 2n^2$ において n を $3m - n$ に置き換えても値は変わらない。対称性の議論から、そうでなければならない。
- 1 から 8 までのすべての m について全数列挙を行うと、5 つの公式はすべて個数と完全に一致する。たとえば $m = 10$、$n = 15$ ではどちらの方法でも $\frac{91}{1331}$ となる。

よくある間違い

- $N_2(j)$ に 2 通りの公式があることを忘れ、中央の範囲で和の最初から最後まで $j + 1$ を使ってしまう。
- 印刷された中央の公式にたどり着こうとして、計算を「無理やり」合わせてしまう。丁寧な計算が与えられた答えと食い違ったら、自分を疑う前にごく小さな場合で試してみるとよい。

STEP のその先へ

独立な確率変数の和の分布は、それぞれの分布の**畳み込み**であり、我々の和 $\sum_k N_2(n-k)$ はまさにそれである。確率母関数（probability generating function）を使えば、この計算は機械的になる：個数 $N_3(n)$ は $(1 + t + \cdots + t^m)^3$ における t^n の係数である。試験の科目数を増やしていくと、分布の形はテントから滑らかな山へと変わり、正規曲線（normal curve）に近づいていく（中心極限定理（central limit theorem））。Grimmett と Welsh による *Probability: An Introduction*（確率論入門）は、母関数と中心極限定理をちょうどよい水準で扱っている。

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。