解答例

STEP III 1991、第 16 問

確率・統計 ・ 同時分布と周辺分布

STEP III 1991、問 16

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/91-s3-q16

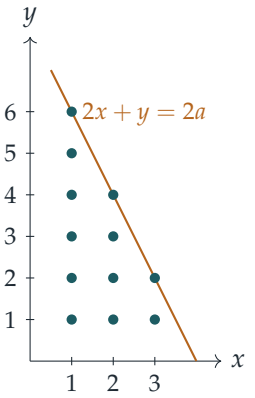
鍵となる考え方

ここで扱うものはすべて、三角形をなす格子点（lattice points）の上の有限和である。まず領域を図に描くこと。各 x に対してどの y が許されるか、また各 y に対してどの x が許されるかを把握する必要がある。和の範囲さえ正しく定まれば、各周辺分布（marginal distribution）は等差数列（arithmetic progression）の和になる。唯一の微妙な点は、与えられた y に対して許される最大の x が $\lfloor (2a-y)/2 \rfloor$ であり、これが y の偶奇によって変わることである。

解答

領域

条件は $x \geq 1$、$y \geq 1$ および $2x + y \leq 2a$ である。$y \geq 1$ であるから $2x \leq 2a - 1$ が必要であり、したがって $x \leq a - 1$ となる。$1 \leq x \leq a - 1$ を満たす x を固定すると、許される値は $y = 1, 2, \ldots, 2a - 2x$ である。同様に、$x \geq 1$ であるから $y \leq 2a - 2$ となる。



図は $a = 4$ の場合を示している。列 x には $2a - 2x$ 個の点がある。

X の周辺分布と c の値

x を固定し、$n = 2a - 2x$ とおく。等差数列 $2x + 1, \ldots, 2x + n$ の和をとると

$$\begin{aligned} \mathrm{P}(X = x) &= c \sum_{y=1}^{n} (2x + y) = c \left(2xn + \frac{n(n+1)}{2} \right) = c \frac{n}{2} (4x + n + 1) \\ &= c\,(a - x)(2a + 2x + 1), \qquad x = 1, 2, \ldots, a - 1, \end{aligned}$$

であり、それ以外では $\mathrm{P}(X = x) = 0$ である。

確率の総和は 1 でなければならない。$k = a - x$ とおくと、k は 1 から $a - 1$ までを動く。このとき $2a + 2x + 1 = 4a + 1 - 2k$ であり、

$$\begin{aligned} \sum_{k=1}^{a-1} k(4a + 1 - 2k) &= (4a + 1) \frac{(a-1)a}{2} - 2 \cdot \frac{(a-1)a(2a-1)}{6} \\ &= a(a-1) \left(\frac{4a+1}{2} - \frac{2a-1}{3} \right) = \frac{a(a-1)(8a+5)}{6}. \end{aligned}$$

これの c 倍を 1 とおくと

$$c = \frac{6}{a(a-1)(8a+5)},$$

となり、求める結果が得られる。

Y の周辺分布

$1 \le y \le 2a - 2$ を満たす y を固定する。許される x は $1, 2, \ldots, N$ であり、ここで N は $2N \le 2a - y$ を満たす最大の整数である。これら N 個の項について

$$\sum_{x=1}^{N} (2x + y) = N(N + 1) + yN = N(N + 1 + y).$$

y が偶数の場合。このとき $2a-y$ は偶数で $N=\frac{1}{2}(2a-y)$ であるから、$N+1+y=\frac{1}{2}(2a+2+y)$ となり、c の値を用いると

$$\mathrm{P}(Y=y)=c\cdot\frac{(2a-y)(2a+2+y)}{4}=\frac{3(2a-y)(2a+2+y)}{2a(a-1)(8a+5)},$$

を得る。これが求める結果である。

考えてみよう y が奇数のとき、なぜ $N=\frac{1}{2}(2a-y)$ を使えないのだろうか。

y が奇数のとき、$\frac{1}{2}(2a-y)$ は整数ではない。$2x\le 2a-y$ を満たす最大の整数 x は $N=\frac{1}{2}(2a-y-1)$ である。切り捨てによって半端な分が落ちるのである。

y が奇数の場合。$N=\frac{1}{2}(2a-1-y)$ とすると $N+1+y=\frac{1}{2}(2a+1+y)$ となるから

$$\mathrm{P}(Y=y)=c\cdot\frac{(2a-1-y)(2a+1+y)}{4}=\frac{3(2a-1-y)(2a+1+y)}{2a(a-1)(8a+5)}.$$

(二つの場合は、同じ分母の上に、それぞれ $3\big((2a+1)^2-(y+1)^2\big)$ と $3\big(4a^2-(y+1)^2\big)$ をのせた形にも書ける。)

Y の期待値

$y\,\mathrm{P}(Y=y)$ を足し上げようとすると、y が偶数の場合と奇数の場合を別々に扱うことになる。同時分布（joint distribution）に立ち返り、各列ごとに縦に和をとるほうがすっきりする。x を固定し、先ほどと同じく $n=2a-2x$ とすると

$$\sum_{y=1}^{n}y(2x+y)=2x\cdot\frac{n(n+1)}{2}+\frac{n(n+1)(2n+1)}{6}=\frac{n(n+1)}{6}\big(6x+2n+1\big).$$

再び $k=a-x$ とおくと、$n=2k$ および $6x+2n+1=6a-2k+1$ である。列の和は

$$\frac{2k(2k+1)(6a+1-2k)}{6}=\frac{k(2k+1)(6a+1-2k)}{3}.$$

となる。展開すると $k(2k+1)(6a+1-2k)=(6a+1)(2k^2+k)-4k^3-2k^2$ である。$a-1$ までの標準的な和の公式を用いると

$$\sum_{k=1}^{a-1}(2k^2+k)=\frac{a(a-1)(2a-1)}{3}+\frac{a(a-1)}{2}=\frac{a(a-1)(4a+1)}{6},$$

$$\sum_{k=1}^{a-1}k^3=\frac{a^2(a-1)^2}{4},\qquad\sum_{k=1}^{a-1}k^2=\frac{a(a-1)(2a-1)}{6}.$$

よって

$$\begin{aligned}
&\sum_{k=1}^{a-1} k(2k+1)(6a+1-2k) \\
&\quad = a(a-1)\left(\frac{(6a+1)(4a+1)}{6} - a(a-1) - \frac{2a-1}{3}\right) \\
&\quad = \frac{a(a-1)}{6}\left(24a^2+10a+1-6a^2+6a-4a+2\right) \\
&\quad = \frac{a(a-1)(6a^2+4a+1)}{2}.
\end{aligned}$$

したがって、期待値（expectation）は

$$\begin{aligned}
\mathrm{E}(Y) &= \frac{c}{3}\cdot\frac{a(a-1)(6a^2+4a+1)}{2} = \frac{6}{a(a-1)(8a+5)}\cdot\frac{a(a-1)(6a^2+4a+1)}{6} \\
&= \frac{6a^2+4a+1}{8a+5}.
\end{aligned}$$

である。

検算

- $a=2$ のとき、点は $(1,1)$ と $(1,2)$ の二つだけであり、重みはそれぞれ $3c$ と $4c$ であるから $c=\frac{1}{7}$ となる。これは $6/(2\cdot 1\cdot 21)$ と一致する。さらに $\mathrm{E}(Y)=\frac{3+8}{7}=\frac{11}{7}$ であり、公式からも $\frac{33}{21}=\frac{11}{7}$ が得られる。
- X の周辺分布の公式、Y の周辺分布の公式（偶奇の両方）、および $\mathrm{E}(Y)$ は、2 から 29 までのすべての a について、分数を用いた厳密な全数列挙で確かめた。たとえば $a=3$ では $\mathrm{E}(Y)=\frac{67}{29}$ となる。
- a が大きいとき、$\mathrm{E}(Y)\approx\frac{3}{4}a$ である。重み $2x+y$ は斜辺に近い点ほど大きいので、Y は三角形上で重みを均等にした場合に得られる $\frac{2}{3}a$ よりもかなり上に位置する。

よくある間違い

- x を a まで動かしてしまうこと。点 $x=a$ では $y\le 0$ とならざるを得ない。
- 奇数の y に対して、整数でない $N=\frac{1}{2}(2a-y)$ を使ってしまうこと。
- 二つの場合に分かれた周辺分布から $\mathrm{E}(Y)$ を求めようとすること。それでも求められるが、同時分布の列ごとに和をとれば、偶奇（parity）による場合分けそのものを避けられる。

STEP のその先へ

同時分布から周辺分布を得るとは、もう一方の変数について和をとって消去することであり、難しいのはいつも領域の範囲である。連続型の確率変数では、同じ作業が二重積分（double integral）の積分範囲を選ぶことになり、（$\mathrm{E}(Y)$ を求めたときのように）和の順序を交換することは積分の順序を交換することになる。Grimmett と Welsh による *Probability: An Introduction*（確率論入門）は、離散型と連続型の両方について同時分布と周辺分布を扱っている。

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。