详细解答

STEP III 1991，第 16 题

概率与统计 • 联合分布与边际分布

STEP III 1991，第 16 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/91-s3-q16

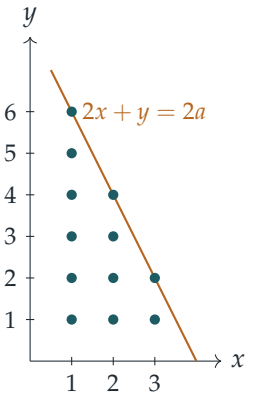
关键思路

这里的一切都是在一个由格点（lattice points）组成的三角形上的有限和。先把区域画出来：对每个 x，我们要知道哪些 y 是容许的；对每个 y，又有哪些 x 是容许的。一旦求和的上下限定对了，每个边际分布（marginal distribution）都是一个等差数列（arithmetic progression）之和。唯一要留神的地方，是对给定的 y，容许的最大 x 是 $\lfloor (2a-y)/2 \rfloor$，而这取决于 y 是偶数还是奇数。

解答

区域

我们需要 $x \geq 1$、$y \geq 1$ 和 $2x + y \leq 2a$。由于 $y \geq 1$，必须有 $2x \leq 2a - 1$，所以 $x \leq a - 1$。对于固定的 x（$1 \leq x \leq a - 1$），容许的值是 $y = 1, 2, \ldots, 2a - 2x$。同样地，因为 $x \geq 1$，所以 $y \leq 2a - 2$。



图中所示为 $a = 4$ 的情况：第 x 列有 $2a - 2x$ 个点。

X 的边际分布与 c 的值

固定 x，并记 $n = 2a - 2x$。把等差数列 $2x + 1, \ldots, 2x + n$ 求和，

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

而在其他情况下 $\mathrm{P}(X = x) = 0$。

所有概率之和必须等于 1。设 $k = a - x$，它由 1 取到 $a - 1$；于是 $2a + 2x + 1 = 4a + 1 - 2k$，并且

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

令 c 乘以此和等于 1，便得

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

正如所求。

Y 的边际分布

固定 y，其中 $1 \le y \le 2a - 2$。容许的 x 是 $1, 2, \ldots, N$，其中 N 是满足 $2N \le 2a - y$ 的最大整数。对这 N 项求和，

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

y **为偶数**。此时 $2a-y$ 是偶数，$N=\frac{1}{2}(2a-y)$，所以 $N+1+y=\frac{1}{2}(2a+2+y)$，并且

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

其中用了 c 的值。这就是所要求的结果。

想一想 当 y 是奇数时，为甚么不能用 $N=\frac{1}{2}(2a-y)$?

当 y 是奇数时，$\frac{1}{2}(2a-y)$ 不是整数。满足 $2x\le 2a-y$ 的最大整数 x 是 $N=\frac{1}{2}(2a-y-1)$：向下取整把那半个舍去了。

y **为奇数**。取 $N=\frac{1}{2}(2a-1-y)$，便得 $N+1+y=\frac{1}{2}(2a+1+y)$，所以

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

（这两种情况可以写成 $3\big((2a+1)^2-(y+1)^2\big)$ 和 $3\big(4a^2-(y+1)^2\big)$ 除以同一个分母。）

Y 的期望值

把 $y\,\mathrm{P}(Y=y)$ 求和，就要把偶数和奇数的 y 分开处理。较简洁的做法是回到联合分布（joint distribution），逐列求和。对于固定的 x，同前面一样取 $n=2a-2x$，

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

再次设 $k=a-x$，于是 $n=2k$，$6x+2n+1=6a-2k+1$。每列之和变成

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

展开得 $k(2k+1)(6a+1-2k)=(6a+1)(2k^2+k)-4k^3-2k^2$。利用求和至 $a-1$ 的标准公式，

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

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

所以

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

因此

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

验算

- 当 $a = 2$ 时只有两个点，$(1,1)$ 和 $(1,2)$，权重分别为 $3c$ 和 $4c$，所以 $c = \frac{1}{7}$。这与 $6/(2\cdot 1\cdot 21)$ 一致。于是 $\mathrm{E}(Y) = \frac{3+8}{7} = \frac{11}{7}$，而公式给出 $\frac{33}{21} = \frac{11}{7}$。
- X 和 Y 的边际分布公式（奇偶两种情况）以及 $\mathrm{E}(Y)$，已对 a 由 2 至 29 的每个值用分数作精确枚举核对。例如，$a = 3$ 给出 $\mathrm{E}(Y) = \frac{67}{29}$。
- 当 a 很大时，$\mathrm{E}(Y) \approx \frac{3}{4}a$。权重 $2x + y$ 偏重靠近斜边的点，所以 Y 明显高于在三角形上取相等权重时所得的 $\frac{2}{3}a$。

常见错误

- 让 x 一直取到 a：$x = a$ 这一点会迫使 $y \le 0$。
- 对奇数 y 使用 $N = \frac{1}{2}(2a - y)$，而它并不是整数。
- 试图从分两种情况的边际分布求 $\mathrm{E}(Y)$。这样做是行得通的，但沿联合分布的各列求和，就可以完全避开奇偶之分。

STEP 之外

从联合分布求边际分布，就是把另一个变量求和消去，而难处总在于区域的上下限。对于连续变量，同样的工作变成在二重积分（double integral）中选取积分限，而交换求和次序（正如我们求 $\mathrm{E}(Y)$ 时所做的）则变成交换积分次序。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门）涵盖了离散和连续变量的联合分布与边际分布。

Beyond STEP 详细解答：https://beyondstep.keithmcnulty.org/zh-hans/。解答文字采用 CC0 1.0；题目本身版权属 Cambridge University Press & Assessment 所有（©），本文件不予转载。由机器从英文翻译而成；尚未经审阅。