详细解答

STEP III 2023，第 12 题

概率与统计 · 组合概率

STEP III 2023，第 12 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2023 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/707102-2023-past-papers.zip

关键思路

虽然同一对袜子的两只看起来一样，但抽出的是从 $2n$ 只*实际存在*的袜子中均匀随机（uniformly random）选取的 $2k$ 只所组成的集合，所以我们为每只袜子编上标号，然后数集合的个数。这样一来，一切都归结为分阶段构造一个选法：先选出哪些对是完整出现的，再选出哪些其他的对只贡献一只袜子，最后选出每一个这样的对中取哪一只。第 (iii) 部分用的是熟悉的技巧 $r\binom{n}{r} = n\binom{n-1}{r-1}$，它把一个加权和转化为一个概率（probability）之和。

解答

为袜子编上标号，使全部 $2n$ 只互不相同。共有 $\binom{2n}{2k}$ 种机会均等的选法，而下面每个概率都等于（有利选法的数目）$/\binom{2n}{2k}$。

第 (i) 部分

一个没有完整一对的选法，从每一对中最多只含一只袜子，所以它的 $2k$ 只袜子来自 $2k$ 个不同的对。要构造这样的选法，先选出有代表出现的 $2k$ 对，有 $\binom{n}{2k}$ 种方法（因为 $2k \le n$，这是可行的），然后从每一对的两只中选一只，有 2^{2k} 种方法。每一种这样的选择都给出不同的集合，而每个没有成对的集合都以这种方式产生，所以

$$\mathrm{P}(\text{没有成对}) = \frac{\binom{n}{2k}\,2^{2k}}{\binom{2n}{2k}}.$$

第 (ii) 部分

固定 r，其中 $0 \le r \le k$。一个恰好有 r 对完整袜子的选法由以下部分组成：

- 整对取出的 r 对（共 $2r$ 只袜子），以及
- 另外 $2k-2r = 2(k-r)$ 只袜子，全部来自不同的对（否则成对的数目会多于 r），而且都不来自已用的 r 对。

因此，我们以 $\binom{n}{r}$ 种方法选出 r 个完整的对；从余下的 $n-r$ 对中，以 $\binom{n-r}{2(k-r)}$ 种方法选出贡献一只袜子的 $2(k-r)$ 对；再从这些对中各取一只，有 $2^{2(k-r)}$ 种方法。同样地，每个恰好有 r 对的集合都恰好被产生一次，所以

$$\mathrm{P}(X_{n,k} = r) = \frac{\binom{n}{r}\binom{n-r}{2(k-r)}\,2^{2(k-r)}}{\binom{2n}{2k}},$$

此即所求。（当 $r = 0$ 时，这就是第 (i) 部分。）

第 (iii) 部分

设 $1 \le r \le k$。由于

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

由第 (ii) 部分得

$$r\,\mathrm{P}(X_{n,k} = r) = \frac{n\binom{n-1}{r-1}\binom{n-r}{2(k-r)}\,2^{2(k-r)}}{\binom{2n}{2k}}.$$

现在把第 (ii) 部分应用于 $n-1$ 对袜子、抽出 $2(k-1)$ 只、找到 $r-1$ 对的情况。这是容许的，因为由 $2k \le n$ 可推出 $2(k-1) \le n-1$，而且 $0 \le r-1 \le k-1$。由于 $(n-1)-(r-1) = n-r$ 及 $(k-1)-(r-1) = k-r$，

$$\mathrm{P}(X_{n-1,k-1} = r-1) = \frac{\binom{n-1}{r-1}\binom{n-r}{2(k-r)}\,2^{2(k-r)}}{\binom{2n-2}{2k-2}}.$$

两者的分子除了因子 n 之外完全相同，所以

$$\begin{aligned}\frac{r\,\mathrm{P}(X_{n,k} = r)}{\mathrm{P}(X_{n-1,k-1} = r-1)} &= \frac{n\binom{2n-2}{2k-2}}{\binom{2n}{2k}} = n \cdot \frac{(2n-2)!}{(2k-2)!\,(2n-2k)!} \cdot \frac{(2k)!\,(2n-2k)!}{(2n)!} \\ &= n \cdot \frac{2k(2k-1)}{2n(2n-1)} = \frac{k(2k-1)}{2n-1}.\end{aligned}$$

（上面的除法是安全的：$n-r \ge 2k-r \ge 2(k-r)$，所以这里每个二项式系数（binomial coefficient）都是正数。）两边相乘即得所求结果。

想一想 这个恒等式对求 $\mathrm{E}(X_{n,k})$ 有甚么帮助?

期望值（expectation）正好是左边各项之和，而右边各项加起来是一个完整的概率分布（probability distribution）。期望值中 $r=0$ 的一项为零，所以

$$\begin{aligned}\mathrm{E}(X_{n,k}) = \sum_{r=1}^{k} r\,\mathrm{P}(X_{n,k}=r) &= \frac{k(2k-1)}{2n-1}\sum_{r=1}^{k}\mathrm{P}(X_{n-1,k-1}=r-1)\\ &= \frac{k(2k-1)}{2n-1}\sum_{s=0}^{k-1}\mathrm{P}(X_{n-1,k-1}=s).\end{aligned}$$

变量 $X_{n-1,k-1}$ 数的是 $2(k-1)$ 只袜子中的对数，所以它只取值 $0,1,\ldots,k-1$，因此最后的和等于 1。故此

$$\mathrm{E}(X_{n,k}) = \frac{k(2k-1)}{2n-1}.$$

验算

- 当 $k=1$（两只袜子）时，答案是 $\frac{1}{2n-1}$：无论先抽到哪一只，余下的 $2n-1$ 只中恰好有一只是它的配对。
- 用指示变量（indicator variables）可以直接求出平均值：第 i 对完整出现的概率是 $\binom{2n-2}{2k-2}/\binom{2n}{2k} = \frac{2k(2k-1)}{2n(2n-1)}$，乘以 n 对，同样得到 $\frac{k(2k-1)}{2n-1}$。
- 当 $n=6$、$k=3$ 时：第 (i) 部分给出 $\frac{64}{924} = \frac{16}{231}$，而平均值是 $\frac{15}{11}$。以电脑对所有 $n \le 8$ 枚举全部选法，证实了 (ii) 中的分布、(iii) 中的恒等式以及平均值。

常见错误

- 把同一对的两只袜子当作真正相同，然后去数“颜色组合”：这些组合并非机会均等，所以必须对有标号的袜子计数。
- 忘记了因子 $2^{2(k-r)}$，它对应每个单只出现的对中取哪一只。
- 在 (iii) 中没有留意到 $\sum_s \mathrm{P}(X_{n-1,k-1}=s)=1$，而尝试硬算这个和。

STEP 之外

第二项验算用的是期望值的线性性（linearity of expectation）配合指示变量，它无需知道分布就能求出平均值；即使“第 i 对完整出现”这些事件并不独立，这方法依然成立。(iii) 中的恒等式是一个尺寸偏倚（size-biasing）关系：以 r 为分布加权，会得到一个较小问题的平移版本；这个技巧同样可以求出二项分布（binomial distribution）和超几何

分布（hypergeometric distribution）的平均值。Blitzstein 与 Hwang 的 *Introduction to Probability*（《概率导论》）全书都采用这类指示变量和故事式证明。

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。