详细解答

STEP I 2013，第 13 题

概率与统计 · 计数与期望值

STEP I 2013，第 13 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/13-s1-q13

关键思路

数字 $1,\ldots,52$ 可分成 26 个互不相交的“伙伴对” $\{k, 53-k\}$，而被弃掉的一对，其实就是我们把两个成员都选中了的伙伴对。因此，要数出恰好有 k 个完整对的选法，我们先选定哪些对是完整的，再用来自不同对的数字填满其余位置，每次可自由选取该对中的哪一个成员。至于期望值（expectation），对这 26 个对运用线性性质（linearity）便可迅速验证结果。

解答

数算选法

把这些整数写成 26 个对 $\{1,52\},\{2,51\},\ldots,\{26,27\}$。由 1 至 52 中的两个整数之和为 53，当且仅当它们组成其中一个对，而每个整数恰好属于一个对。所以 X 就是两个成员都在所选七个数之中的对的数目。由于 $2X \le 7$，X 的取值为 $0,1,2,3$。

所有 $\binom{52}{7}$ 种选法的可能性相等，因此我们数算 $X=k$ 的选法数目。

- 选出 k 个完整对：共 $\binom{26}{k}$ 种方法。这些对用去七个数中的 $2k$ 个。
- 其余 $7-2k$ 个数必须来自余下 $26-k$ 个对中 $7-2k$ 个不同的对（若有两个来自同一对，便会多出一个完整对）：共 $\binom{26-k}{7-2k}$ 种方法。
- 从每个这样的对中取其两个成员之一：共 2^{7-2k} 种方法。

每个 $X=k$ 的选法以这种方式恰好出现一次，所以

$$\mathrm{P}(X=k)=\frac{\binom{26}{k}\binom{26-k}{7-2k}2^{7-2k}}{\binom{52}{7}},\qquad k=0,1,2,3.$$

具体数值

$$\begin{aligned}
X=0:&\quad \tbinom{26}{7}2^7 = 657\,800\times 128 = 84\,198\,400,\\
X=1:&\quad 26\tbinom{25}{5}2^5 = 26\times 53\,130\times 32 = 44\,204\,160,\\
X=2:&\quad \tbinom{26}{2}\tbinom{24}{3}2^3 = 325\times 2024\times 8 = 5\,262\,400,\\
X=3:&\quad \tbinom{26}{3}\tbinom{23}{1}2 = 2600\times 23\times 2 = 119\,600,
\end{aligned}$$

而 $\binom{52}{7} = 133\,784\,560$。这四个数目加起来确实等于 $\binom{52}{7}$，这是一个很好的验证，说明没有遗漏或重复计算。

> **想一想** 分数中有九位数时，我们该如何化简？

题目中的附注就是提示：$5593 = 7\times 17\times 47$ 应该是分母。事实上 $\binom{52}{7} = 23\,920\times 5593$，而这四个数目每一个都是 23 920 的倍数：

$$84\,198\,400 = 3520\times 23\,920, \qquad 44\,204\,160 = 1848\times 23\,920,$$
$$5\,262\,400 = 220\times 23\,920, \qquad 119\,600 = 5\times 23\,920.$$

所以 X 的分布（distribution）为

k	0	1	2	3
$\mathrm{P}(X=k)$	$\frac{3520}{5593}$	$\frac{1848}{5593}$	$\frac{220}{5593}$	$\frac{5}{5593}$

（第二个的最简形式为 $\frac{264}{799}$），而 $3520+1848+220+5 = 5593$，理应如此。

期望值

$$\mathrm{E}(X) = \frac{0\times 3520 + 1\times 1848 + 2\times 220 + 3\times 5}{5593} = \frac{2303}{5593}.$$

由于 $2303 = 7^2\times 47$ 而 $5593 = 7\times 17\times 47$，所以

$$\mathrm{E}(X) = \frac{7^2\times 47}{7\times 17\times 47} = \frac{7}{17},$$

正如所求。

求期望值的捷径

对于 $j = 1,\ldots,26$，若第 j 个对的两个成员都被选中，则设 $I_j = 1$，否则为 0，于是 $X = I_1 + \cdots + I_{26}$。选中某一指定对的两个成员后，须从其余 50 个数中再选 5 个，所以

$$\mathrm{P}(I_j = 1) = \frac{\binom{50}{5}}{\binom{52}{7}} = \frac{7\times 6}{52\times 51} = \frac{7}{442}.$$

由期望值的线性性质（这不需要独立性（independence）；事实上这些 I_j 并不独立），

$$\mathrm{E}(X) = 26 \times \frac{7}{442} = \frac{7}{17}.$$

验算

- 四个概率之和为 1，而两种方法都给出相同的 $\mathrm{E}(X) = \frac{7}{17} \approx 0.412$。
- 以电脑模拟 300 000 次随机选取，得到 $X = 0, 1, 2, 3$ 的相对频率约为 0.630、0.329、0.039、0.001，而准确值为 0.6294、0.3304、0.0393、0.0009。
- 合理性检查：$\mathrm{P}(X = 3)$ 非常小，这合乎预期，因为只有七个数却要有三个完整对。

常见错误

- 把 $X = 0$ 的选法数目算成 $\binom{52}{7}$ 减去某个粗略的数，或只算 $\binom{26}{7}$，忘记了每个对取哪一个成员所带来的因子 2^7。
- 对于 $X = 1$，把其余五个数的选法算成 $\binom{50}{5}$：这容许出现第二个完整对，因而多算了。

STEP 之外

把一个计数写成指示变量（indicator variable）之和，再运用期望值的线性性质，是大学一年级概率论中最有用的技巧之一：即使分布根本无法求出，它仍能给出期望值。本题的完整分布尚可求得，但例如要求随机排列（permutation）的不动点（fixed point）的期望数目，指示变量便是最自然的切入点。Blitzstein 与 Hwang 的 *Introduction to Probability*（《概率导论》）一书反复运用这个想法。

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。