详细解答

STEP I 2007，第 13 题

概率与统计 • 条件概率

STEP I 2007，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/07-s1-q13

本题在 *Beyond STEP* 第 12.1 节精确地取条件中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch12/conditioning.html

关键思路

虽然圆片是逐个取出的，但每一组四个圆片出现的可能性都相同，所以大部分工夫都花在数组合（combination）上。真正值得留意的是各个条件。知道第一个取出的是哪一个圆片，或者只知道某个指定的圆片确实被取出，余下三个圆片都仍然是从剩下十个之中随机选出的。只知道有某个有号码的圆片被取出，则是性质完全不同的信息，而最后两部分正是为了显示这个分别而设。

解答

设 N 为取出的四个圆片之中有号码的圆片数目。由于袋子经过摇匀，每一个由四个不同圆片组成的有序序列出现的可能性都相同，因此每一个由四个圆片组成的无序集合出现的可能性也相同。这样的集合共有

$$\binom{11}{4} = 330$$

个。要组成一个恰好有 k 个有号码圆片的集合，须从 5 个有号码的圆片中选 k 个，再从 6 个空白圆片中选 $4-k$ 个，所以

$$\mathrm{P}(N=k) = \frac{\binom{5}{k}\binom{6}{4-k}}{330}.$$

第 (i) 部分

四个全部有号码，即 $k = 4$：

$$\mathrm{P}(N = 4) = \frac{\binom{5}{4}\binom{6}{0}}{330} = \frac{5}{330} = \frac{1}{66}.$$

换一个算法，按次序逐个抽取，$\frac{5}{11} \cdot \frac{4}{10} \cdot \frac{3}{9} \cdot \frac{2}{8} = \frac{1}{66}$。

第 (ii) 部分

设 T_1 为“3”号圆片最先被取出这个事件（event）。已知 T_1 发生，摇匀的袋子里现在有 10 个圆片，其中 4 个有号码、6 个空白，而接着取出的三个圆片是从这些圆片中随机选出的。四个圆片全部有号码，当且仅当这三个全部有号码，所以

$$\mathrm{P}(N = 4 \mid T_1) = \frac{\binom{4}{3}}{\binom{10}{3}} = \frac{4}{120} = \frac{1}{30}.$$

第 (iii) 部分

仍然已知 T_1 发生，我们需要从余下十个圆片抽出的三个之中，恰好再有一个有号码的圆片：

$$\mathrm{P}(N = 2 \mid T_1) = \frac{\binom{4}{1}\binom{6}{2}}{\binom{10}{3}} = \frac{4 \times 15}{120} = \frac{1}{2}.$$

第 (iv) 部分

现在设 T 为“3”号圆片在某一次被取出这个事件。

> ? **想一想** “3”号圆片是最先取出，还是在其他某一次取出，有没有分别?

没有分别。已知 T 发生，取出的四个圆片就是“3”号圆片连同另外三个，而根据对称性（symmetry），从其余十个圆片中选出三个的每一个集合，出现的可能性都相同。这正是第 (iii) 部分的情况，当中抽取的次序从来都无关重要。所以答案同样是 $\frac{1}{2}$。

直接验证如下：$\mathrm{P}(T) = \binom{10}{3}/\binom{11}{4} = \frac{120}{330} = \frac{4}{11}$，而包含“3”号圆片、另外一个有号码圆片和两个空白圆片的集合共有 $\binom{4}{1}\binom{6}{2} = 60$ 个，所以

$$\mathrm{P}(N = 2 \mid T) = \frac{60/330}{120/330} = \frac{1}{2}.$$

第 (v) 部分

设 F 为第一个圆片有号码这个事件。按第一个圆片上的号码把 F 分拆：F 是“j 号圆片最先被取出”这五个事件（$j = 1, \ldots, 5$）的不相交并集（disjoint union），每个事件的概率

都是 $\frac{1}{11}$。第 (iii) 部分的论证并没有用到号码是 3 这一点，所以对每个 j 都有

$$\mathrm{P}(N=2 \mid j \text{ 号圆片最先}) = \tfrac{1}{2}.$$

因此，根据全概率公式（law of total probability），

$$\mathrm{P}(N=2 \mid F) = \sum_{j=1}^{5} \frac{1/11}{5/11} \cdot \frac{1}{2} = \frac{1}{2}.$$

（直接的做法：已知 F 发生，其余三个圆片是从 4 个有号码和 6 个空白的圆片中随机抽出的三个，这又回到第 (iii) 部分。）

第 (vi) 部分

设 A 为最少有一个有号码的圆片被取出这个事件，即 A 就是事件 $N \geq 1$。这与之前的信息并不属于同一类：它没有指明某一个圆片，再让其余的保持随机。我们改为使用条件概率（conditional probability）的定义。由于 $N=2$ 蕴含 A，

$$\mathrm{P}(N=2 \mid A) = \frac{\mathrm{P}(N=2)}{\mathrm{P}(N \geq 1)}.$$

> **想一想** 为甚么不可以直接说“已知有一个有号码的圆片，所以其余三个是从剩下十个之中随机抽出的”？

因为我们不知道已知的是哪一个圆片。一个有两个有号码圆片的集合，与一个有四个的集合，同样满足 A；并没有哪一个圆片可以让我们单独抽出来放在一旁。所以这个条件是按 N 本身的值为各个集合重新加权，我们必须老老实实地数。

我们有

$$\mathrm{P}(N=2) = \frac{\binom{5}{2}\binom{6}{2}}{330} = \frac{10 \times 15}{330} = \frac{150}{330}, \qquad \mathrm{P}(N=0) = \frac{\binom{6}{4}}{330} = \frac{15}{330},$$

所以 $\mathrm{P}(N \geq 1) = \frac{315}{330}$，而

$$\mathrm{P}(N=2 \mid A) = \frac{150}{315} = \frac{10}{21}.$$

验算

N 的分布（distribution）加起来应该正确：当 $k = 0, \ldots, 4$ 时，$\binom{5}{k}\binom{6}{4-k}$ 的数目分别是 $15, 100, 150, 60, 5$，总和为 330。在第 (ii) 和第 (iii) 部分，其余有号码圆片数目的条件分布，在 120 个之中的数目分别是 $20, 60, 36, 4$，总和同样正确。我们亦以精确分数列出全部 $11 \times 10 \times 9 \times 8 = 7920$ 个有序抽取结果，核对了全部六个答案（$\frac{1}{66}$、$\frac{1}{30}$、$\frac{1}{2}$、$\frac{1}{2}$、$\frac{1}{2}$、$\frac{10}{21}$）。留意 $\frac{10}{21} \approx 0.476$ 略小于 $\frac{1}{2}$，这是合理的：以 $N \geq 1$ 为条件只剔除了 $N=0$ 的情况，所以与知道某个指定圆片被取出相比，它在 $N=1$ 上留下了较多的权重。

常见错误

- 把第 (iv) 和第 (vi) 部分当作同一类条件。“3 号圆片被取出”指明了一个圆片；“有一个有号码的圆片被取出”则没有，两者的答案并不相同。
- 在第 (vi) 部分，忘记了条件是 $N \geq 1$，反而除以 $\mathrm{P}(N = 1)$ 或 $\frac{5}{11}$。
- 分子数的是有序抽取，分母数的却是无序集合。两者选定其一，然后贯彻始终。

STEP 之外

第 (v) 和第 (vi) 部分之间的对比，与著名的“男孩还是女孩”悖论（“boy or girl” paradox）是同一个陷阱：得知某个指定的孩子是男孩，与得知最少有一个是男孩，会得出不同的答案。在大学里，这一点会有严格的表述：以一个事件为条件，就是把概率测度（probability measure）限制在该事件上再重新归一化（renormalise），而这里的数目正是超几何分布（hypergeometric distribution）。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率入门），以及 Blitzstein 与 Hwang 的 *Introduction to Probability*（概率导论），都仔细讨论了这些悖论。

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。