详细解答

STEP I 2003，第 12 题

概率与统计 • 乘积的期望值

> **STEP I 2003，第 12 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/03-s1-q12

关键思路

期望值（expected value）就是对所有等可能的结果，把（数值 × 概率）加起来，所以两部分都归结为列出所有有序数对（ordered pairs），再把它们的乘积相加。二重和 $\sum_r \sum_s rs$ 可以分解为 $(\sum r)^2$，第 (i) 部分要做的就只有这一步。第 (ii) 部分唯一的分别，是 $r = s$ 的数对不再可能出现，所以我们取同一个二重和，再除去它的对角线项。

解答

设 X 为第一个球上的数字，Y 为第二个球上的数字。

第 (i) 部分

由于第一个球会放回袋中，第二次抽取时袋中仍是全部的球，结果与第一次无关。所以对任意 $r, s \in \{1, \ldots, n\}$，

$$\mathrm{P}(X = r,\ Y = s) = \mathrm{P}(X = r)\,\mathrm{P}(Y = s) = \frac{1}{n} \cdot \frac{1}{n} = \frac{1}{n^2}.$$

这样的有序数对 (r, s) 共有 n^2 个，每个给出乘积 rs，所以按期望值的定义，

$$\mathrm{E}(XY) = \sum_{r=1}^{n} \sum_{s=1}^{n} rs\,\mathrm{P}(X = r,\ Y = s) = \frac{1}{n^2} \sum_{r=1}^{n} \sum_{s=1}^{n} rs,$$

这正是题目要求我们解释的式子。

要化简这个式子，留意 r 与 s 无关，所以可以把它提到内层求和之外：

$$\sum_{r=1}^{n} \sum_{s=1}^{n} rs = \sum_{r=1}^{n} r\left(\sum_{s=1}^{n} s\right) = \left(\sum_{r=1}^{n} r\right)^2 = \tfrac{1}{4} n^2 (n+1)^2.$$

因此

$$\mathrm{E}(XY) = \frac{(n+1)^2}{4}.$$

（等价地说，X 与 Y 独立（independent），所以 $\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y) = \left(\frac{n+1}{2}\right)^2$。）

第 (ii) 部分

要抽出两个球，须有 $n \geq 2$。现在第一个球不放回，所以可能的结果是满足 $r \neq s$ 的有序数对 (r, s)。对这样的一个数对，

$$\mathrm{P}(X = r,\ Y = s) = \mathrm{P}(X = r)\,\mathrm{P}(Y = s \mid X = r) = \frac{1}{n} \cdot \frac{1}{n-1},$$

因为取走 r 号球之后，第二个球是余下 $n-1$ 个球中任何一个的可能性均等。这样的数对共有 $n(n-1)$ 个，全部等可能，而 $r = s$ 的数对概率为 0。因此

$$\mathrm{E}(XY) = \frac{1}{n(n-1)} \sum_{r \neq s} rs.$$

想一想 怎样只对 $r \neq s$ 的数对把 rs 加起来?

先取对所有数对的和（已在第 (i) 部分求出），再减去 $r = s$ 的对角线项，这些项各自贡献 $r \cdot r = r^2$：

$$\sum_{r \neq s} rs = \sum_{r=1}^{n} \sum_{s=1}^{n} rs - \sum_{r=1}^{n} r^2 = \tfrac{1}{4} n^2 (n+1)^2 - \tfrac{1}{6} n(n+1)(2n+1).$$

提出公因式 $\frac{1}{12} n(n+1)$，

$$\begin{aligned} \sum_{r \neq s} rs &= \tfrac{1}{12} n(n+1)[3n(n+1) - 2(2n+1)] \\ &= \tfrac{1}{12} n(n+1)(3n^2 - n - 2) \\ &= \tfrac{1}{12} n(n+1)(n-1)(3n+2). \end{aligned}$$

除以 $n(n-1)$（因为 $n \geq 2$，它不等于零），

$$\mathrm{E}(XY) = \frac{(n+1)(3n+2)}{12},$$

这就是所求的结果。

验算

取 $n = 3$。有放回时，九个乘积的平均值是 $\frac{1}{9}(1+2+3)^2 = 4$，而 $\frac{(3+1)^2}{4} = 4$。无放回时，无序数对 $\{1,2\}, \{1,3\}, \{2,3\}$ 给出乘积 $2, 3, 6$，平均值为 $\frac{11}{3}$，而 $\frac{4 \cdot 11}{12} = \frac{11}{3}$。当 $n = 2$ 时，

无放回的乘积只有 2 一个，而公式给出 $\frac{3\cdot 8}{12} = 2$。我们亦以逐一列举的方式，对 12 或以下的每个 n 准确验证了两条公式。

比较两个答案也是一个很好的合理性检查：

$$\frac{(n+1)(3n+2)}{12} - \frac{(n+1)^2}{4} = -\frac{n+1}{12} < 0.$$

无放回时答案较小，这是合理的：我们失去了 $n \cdot n$ 这类数值大的对角线乘积。

常见错误

- 在第 (ii) 部分除以 n^2。可能的有序数对只有 $n(n-1)$ 个，每个的概率是 $\frac{1}{n(n-1)}$。
- 求和时用无序数对，计数时却用有序数对（或者反过来）。对无序数对求和，会令总和与个数同时减半，所以比值不变；但把两者混用，结果就会相差一个因子 2。
- 在第 (ii) 部分写 $\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y)$。这需要独立性，而球不放回时 X 与 Y 并不独立。

STEP 之外

我们求得的差正好就是协方差 (covariance)：$\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y) = -\frac{n+1}{12}$，它是负数，因为第一个数字大，第二个数字也大的可能性就会稍低。在无放回抽样（sampling without replacement）中求和的方差（variance）时，需要的正是这类协方差，由此可以得出超几何分布（hypergeometric distribution）的方差，以及抽样调查中使用的“有限总体校正”（finite population correction）。Blitzstein 与 Hwang 的 *Introduction to Probability*（概率论入门教材）正是以这种方式发展协方差与指示变量（indicator）方法。

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。