详细解答

STEP III 2021，第 12 题

概率与统计 · 和的方差

STEP III 2021，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 2021 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696588-2021-past-papers.zip

本题在 *Beyond STEP* 第 12.4 节两两独立，却非相互独立中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch12/pairwise-independence.html

关键思路

把全队的得分写成一些指示型变量（indicator variable）之和，每对球员对应一个变量，然后展开这个和的方差（variance）。一切都取决于两个不同的球员对之间如何相互影响。如果两对没有共同的球员，它们显然是独立（independent）的，所以真正要处理的是共有一名球员的两对。在 (i) 中，这样的两对原来是独立的，所以交叉项消失。在 (iii) 中则不然，而 (ii) 中的恒等式正是计算它们贡献的工具。

解答

第 (i) 部分

每颗骰子都是公正的，各球员独立掷骰。不论两名指定球员中的第一人掷出甚么，二人掷出相同点数的概率都是 $\frac{1}{6}$。所以

$$\mathrm{P}(X_{12}=1)=\mathrm{P}(X_{23}=1)=\frac{1}{6}.$$

$X_{12}=1$ 和 $X_{23}=1$ 同时发生，当且仅当球员 1、2、3 都掷出相同点数，所以

$$\mathrm{P}(X_{12}=1,\ X_{23}=1)=\frac{1}{6}\cdot\frac{1}{6}=\frac{1}{36}=\mathrm{P}(X_{12}=1)\,\mathrm{P}(X_{23}=1).$$

从边际概率（marginal probability）中减去这个值，便得到其余三格。例如 $\mathrm{P}(X_{12}=1,X_{23}=0)=\frac{1}{6}-\frac{1}{36}=\frac{1}{6}\cdot\frac{5}{6}$；同理 $\mathrm{P}(X_{12}=0,X_{23}=1)=\frac{5}{6}\cdot\frac{1}{6}$，以及 $\mathrm{P}(X_{12}=0,X_{23}=$

$0) = 1 - \frac{1}{6} - \frac{1}{6} + \frac{1}{36} = \frac{5}{6} \cdot \frac{5}{6}$。所以联合分布（joint distribution）等于边际分布的乘积，X_{12} 与 X_{23} 是独立的。

> ? **想一想** 为甚么知道 $X_{12} = 1$ 对 X_{23} 毫无信息?

已知球员 1 与 2 点数相同，球员 2 的点数在 $\{1, \ldots, 6\}$ 上仍然是均匀分布（uniform）的，而球员 3 与之相同的概率是 $\frac{1}{6}$，跟之前完全一样。由于每一面出现的机会均等，共有的那名球员并不带来任何信息。

设 $T = \sum_{i<j} X_{ij}$ 为总得分，它是 $N = \binom{n}{2}$ 项之和。每个 X_{ij} 都是参数为 $\frac{1}{6}$ 的伯努利（Bernoulli）变量，所以 $\mathrm{E}(X_{ij}) = \frac{1}{6}$，$\mathrm{Var}(X_{ij}) = \frac{1}{6} \cdot \frac{5}{6} = \frac{5}{36}$。因此

$$\mathrm{E}(T) = \frac{N}{6} = \frac{n(n-1)}{12}.$$

至于方差，任取两个不同的球员对。如果它们没有共同球员，它们取决于互不相交的两组骰子，所以是独立的。如果它们共有一名球员，那么把球员重新编号后，由上面的论证可知它们也是独立的。所以各 X_{ij} 是成对独立（pairwise independent）的，每个协方差（covariance）都是零，和的方差等于方差之和：

$$\mathrm{Var}(T) = N \cdot \frac{5}{36} = \frac{5n(n-1)}{72}.$$

第 (ii) 部分

设 $S = Y_1 + \cdots + Y_m$。由线性（linearity）得 $\mathrm{E}(S) = 0$，所以 $\mathrm{Var}(S) = \mathrm{E}(S^2) - 0^2 = \mathrm{E}(S^2)$。展开平方，每个 Y_i^2 出现一次，而每个 $i < j$ 的乘积 $Y_i Y_j$ 出现两次：

$$S^2 = \sum_{i=1}^{m} Y_i^2 + 2 \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} Y_i Y_j.$$

取期望值（expectation）并利用线性，便得到所求的公式。

第 (iii) 部分

对两名球员而言，$\mathrm{P}(\text{掷出相同的偶数}) = \frac{3}{36} = \frac{1}{12}$，同样 $\mathrm{P}(\text{掷出相同的奇数}) = \frac{1}{12}$。所以

$$\mathrm{P}(Z_{ij} = 1) = \mathrm{P}(Z_{ij} = -1) = \tfrac{1}{12}, \qquad \mathrm{P}(Z_{ij} = 0) = \tfrac{5}{6}.$$

不独立。如果 $Z_{12} = 1$ 且 $Z_{23} = -1$，球员 2 的点数便要既是偶数又是奇数。所以

$$\mathrm{P}(Z_{12} = 1,\ Z_{23} = -1) = 0 \neq \tfrac{1}{12} \cdot \tfrac{1}{12} = \mathrm{P}(Z_{12} = 1)\,\mathrm{P}(Z_{23} = -1),$$

因此 Z_{12} 与 Z_{23} 并不独立。

平均值。由对称性，$\mathrm{E}(Z_{ij}) = \frac{1}{12} - \frac{1}{12} = 0$，所以总得分 $U = \sum_{i<j} Z_{ij}$ 的平均值为 0。

方差。每个 Z_{ij} 的平均值都是零，所以可以对这 $N=\binom{n}{2}$ 个变量 Z_{ij} 应用 (ii)。我们需要两类期望值。

- 平方项：Z_{ij}^2 等于 1 当且仅当球员 i 与 j 点数相同，所以 $\mathrm{E}(Z_{ij}^2)=\frac{1}{6}$。
- 两个不同球员对的乘积。如果两对没有共同球员，两个变量独立且平均值为零，所以 $\mathrm{E}(Z_{ij}Z_{kl})=\mathrm{E}(Z_{ij})\,\mathrm{E}(Z_{kl})=0$。如果它们共有一名球员，例如 $Z_{12}Z_{23}$，那么只有当三名球员都掷出相同点数时乘积才非零。这时两个因子同为 $+1$（偶数）或同为 -1（奇数），所以乘积是 1。因此 $\mathrm{E}(Z_{12}Z_{23})=\frac{1}{36}$。

想一想 有多少个无序的“球员对之对”恰好共有一名球员?

先以 n 种方法选出共有的球员，再从其余 $n-1$ 名球员中以 $\binom{n-1}{2}$ 种方法选出他的两个搭档。这样得到 $n\binom{n-1}{2}=\frac{1}{2}n(n-1)(n-2)$ 个这样的“球员对之对”，每个只计算一次，正是 (ii) 中双重求和所需要的。

综合起来，

$$\begin{aligned}\mathrm{Var}(U)&=\binom{n}{2}\cdot\frac{1}{6}+2\cdot\frac{n(n-1)(n-2)}{2}\cdot\frac{1}{36}\\&=\frac{n(n-1)}{36}\big(3+(n-2)\big)\\&=\frac{n(n-1)(n+1)}{36}=\frac{1}{36}n(n^2-1),\end{aligned}$$

得证。

验算

- 以例子中的掷骰结果 $3,3,3,3,6,6,2$ 为例，配对计数得 $6+1=7$，带符号计数得 $-6+1=-5$，与题目一致。
- 当 $n=2$ 时只有一对，所以 $\mathrm{Var}(U)=\mathrm{E}(Z_{12}^2)=\frac{1}{6}$，而公式给出 $\frac{2\cdot3}{36}=\frac{1}{6}$。
- 当 $n=3$ 时，若三颗骰子点数全部相同，则 $U=\pm3$（各自概率为 $\frac{1}{72}$）。否则恰好有一对点数相同，$U=\pm1$，总概率为 $\frac{15}{36}$。所以 $\mathrm{E}(U^2)=9\cdot\frac{1}{36}+\frac{15}{36}=\frac{2}{3}$，与 $\frac{3\cdot8}{36}$ 相符。以电脑精确列举 $n\le6$ 的所有结果，证实了两个平均值和两个方差；例如 $n=6$ 时，(i) 得 $\frac{25}{12}$，(iii) 得 $\frac{35}{6}$。

常见错误

- 只检查了一格便声称 X_{12} 与 X_{23} 独立，却没有说明为何其他格也随之成立。对于只取两个值的变量，其他格确实随之成立，但要把这一点说出来。

- 在 (iii) 中，忘记了共有一名球员的两对不论共同点数是偶数还是奇数，贡献都是 $+\frac{1}{36}$。负号在乘积中互相抵消。
- 数错有重叠的“球员对之对”的数目，例如用了 $n(n-1)(n-2)$（有序）之后再乘以 2。

STEP 之外

第 (i) 部分是一个经典例子：变量成对独立，却并非相互独立（mutually independent）：若 $X_{12} = X_{23} = 1$，则必然 $X_{13} = 1$。计算方差只需要成对独立，这个想法应用于随机化算法（randomised algorithms）和生日问题（birthday problem）之中。第 (iii) 部分展示了当各项相依时，利用协方差计算和的方差的一般方法。两者在 Blitzstein 与 Hwang 的 *Introduction to Probability*（《概率导论》）中均有详细阐述。

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。