详细解答

STEP II 1997，第 12 题

概率与统计 · 二项分布的奇偶概率

> **STEP II 1997，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/97-s2-q12

关键思路

只需追踪信息此刻是对还是错。每听错一次，信息便对错互换，所以传回 Albert 的信息正确，当且仅当耳语中出错的次数是偶数。要抽出二项展开式（binomial expansion）中的偶数项，正好可以用 $(a+b)^n+(a-b)^n$，因为奇数项会互相抵消。至于修改后的游戏，要留意一旦失败一次，Albert 传出去的便是错的信息，所以从那时起，每一轮都需要有奇数次听错。

解答

三名玩家

若有 n 名玩家，便有 n 次耳语：Albert 传给 Bertha，Bertha 传给下一位，如此类推绕一圈传到 Zebedee，然后 Zebedee 传回 Albert。每次耳语被正确听到的概率（probability）是 p，而且各次互相独立（independent）。Albert 听到的是否真相，只取决于有多少次耳语被听错，与硬币掷出的结果无关。

三名玩家时有三次耳语，每次正确的概率是 $\frac{2}{3}$。传到 Albert 的信息正确，条件是没有耳语出错，或者恰好有两次出错（第二次错误把第一次抵消了）。所以

$$\Pr(\text{获胜})=\left(\tfrac{2}{3}\right)^3+3\left(\tfrac{1}{3}\right)^2\left(\tfrac{2}{3}\right)=\frac{8}{27}+\frac{6}{27}=\frac{14}{27}.$$

n 名玩家

设 $q = 1 - p$。听错的耳语次数服从 $\mathrm{B}(n, q)$ 分布（distribution），而当这个次数是偶数时游戏便获胜：

$$P = \sum_{k\text{ 为偶数}} \binom{n}{k} q^k p^{n-k}.$$

现在

$$(p+q)^n + (p-q)^n = \sum_{k=0}^{n} \binom{n}{k} p^{n-k} q^k (1 + (-1)^k) = 2 \sum_{k\text{ 为偶数}} \binom{n}{k} p^{n-k} q^k,$$

因为 k 为奇数的项互相抵消，而 k 为偶数的项则出现两次。由于 $p + q = 1$ 及 $p - q = 2p - 1$，

$$\boxed{P = \tfrac{1}{2}(1 + (2p-1)^n).}$$

当 $n = 3$、$p = \frac{2}{3}$ 时，上式给出 $\frac{1}{2}(1 + \frac{1}{27}) = \frac{14}{27}$，与第一部分的结果一致。

修改后的游戏

设 N 为 Albert 向 Bertha 耳语的次数。把信息绕圈传足一周称为一轮，所以每一轮都由 Albert 耳语开始。

> **想一想** 一轮失败之后，下一轮是否只是把第一轮原封不动地重来一次?

不是。在第一轮，Albert 传出的是正确的结果。如果他从 Zebedee 那里听错了，他传下去的是他听到的内容，而那是错的。由错的信息出发，这一轮结束时 Albert 听到真相，当且仅当有奇数次耳语出错，其概率为 $1 - P$。如果这一轮又失败，Albert 便再一次听到（并传出）错的信息，所以之后的每一轮都与这一轮相同。由于各次聆听互相独立，各轮也互相独立。

因此第一轮成功的概率是 P，而之后每一轮成功的概率是 $1 - P$，失败的概率是 P。故此 $\Pr(N = 1) = P$，而对于 $k \geq 2$，

$$\Pr(N = k) = \underbrace{(1-P)}_{\text{第 1 轮失败}} \underbrace{P^{k-2}}_{\text{第 2,…,}k-1\text{ 轮失败}} \underbrace{(1-P)}_{\text{第 }k\text{ 轮成功}} = (1-P)^2 P^{k-2}.$$

假设 $0 < p < 1$，于是 $0 < P < 1$。那么，在题目给出的和式中取 $x = P$，可得

$$\sum_{k=2}^{\infty} kP^{k-2} = \frac{1}{P}\left(\sum_{k=1}^{\infty} kP^{k-1} - 1\right) = \frac{1}{P}\left(\frac{1}{(1-P)^2} - 1\right),$$

因此

$$\begin{aligned}\mathrm{E}(N) &= P + (1-P)^2 \cdot \frac{1}{P}\left(\frac{1}{(1-P)^2} - 1\right)\\ &= P + \frac{1-(1-P)^2}{P}\\ &= P + (2-P) = 2.\end{aligned}$$

Albert 向 Bertha 耳语次数的期望值（expectation）是 2，不论玩家有多少人，也不论 p 是多少。

还有一个更快捷的方法看出这个结果。已知第一轮失败，其后所需的轮数服从几何分布（geometric distribution），成功概率为 $1-P$，所以其平均值是 $1/(1-P)$。故此 $\mathrm{E}(N) = 1 + (1-P)\cdot\frac{1}{1-P} = 2$。

（边界情况：若 $p=1$，或 $p=0$ 而 n 为偶数，则 $P=1$，Albert 只耳语一次。若 $p=0$ 而 n 为奇数，则 $P=0$，第二轮必定成功，所以 $N=2$。）

验算

- 以准确分数列出听对与听错的全部 2^n 种组合，对 $n \le 8$ 及若干个 p 值都得出 $\frac{1}{2}(1+(2p-1)^n)$，当中包括三名玩家时的 $\Pr(\text{获胜}) = \frac{14}{27}$。
- $p = \frac{1}{2}$ 给出 $P = \frac{1}{2}$：信息纯粹是杂讯，理应如此。
- 以 $(n,p) = (3,\frac{2}{3})$、$(5,0.9)$ 及 $(4,0.3)$ 模拟 2×10^5 局修改后的游戏，所得的平均耳语次数分别为 2.001、1.996 及 1.994。

常见错误

- 把耳语次数数成 $n-1$ 而不是 n：最后一次耳语，即由 Zebedee 传回 Albert 的那一次，也要计算在内。
- 把修改后的游戏中每一轮的获胜概率都当作 P。这样会得出 $1/P$，是错的，因为失败之后 Albert 传出的是错的信息。

STEP 之外

信息的对／错状态是一条两状态的马可夫链（Markov chain），而 $2p-1$ 是其转移矩阵（transition matrix）的第二个特征值（eigenvalue）。这就是为甚么 $(2p-1)^n$ 决定了信息“忘记”硬币结果的速度。同一个模型称为二元对称通道（binary symmetric channel），是信息论（information theory）和纠错码（error-correcting codes）的起点。Grimmett 与

Welsh 的 *Probability: An Introduction*（概率论入门）以这个程度讲解生成函数（generating functions）和马可夫链。

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。