详细解答

STEP II 2001，第 14 题

概率与统计 · 最小方差估计量

> **STEP II 2001，第 14 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/01-s2-q14

关键思路

T 是两个独立（independent）伯努利变量（Bernoulli variable）的加权组合，因此它的平均值（mean）和方差（variance）可直接由期望值的线性性质（linearity of expectation），以及适用于独立的 X、Y 的法则 $\operatorname{Var}(aX+cY)=a^2\operatorname{Var}(X)+c^2\operatorname{Var}(Y)$ 得出。权重经过特意选取，使得无论 λ 取何值，平均值都是 p：T 是 p 的*无偏估计量*（unbiased estimator），而题目要找的正是其中最好的一个。至于最后一部分，中心极限定理（central limit theorem）令 $\overline{T}$ 近似服从正态分布（normal distribution），而平均值固定的正态变量，方差最小时最为集中。

解答

T 的平均值与方差

这里 X 和 Y 都是伯努利变量，且

$$\mathrm{E}(X)=p,\quad \operatorname{Var}(X)=p(1-p),\qquad \mathrm{E}(Y)=2p,\quad \operatorname{Var}(Y)=2p(1-2p).$$

（若一个变量以概率 q 取值 1，否则取值 0，则 $\mathrm{E}=q$ 且 $\mathrm{E}(\cdot^2)=q$，所以方差是 $q-q^2=q(1-q)$。）由期望值的线性性质，

$$\mathrm{E}(T)=\lambda p+\tfrac{1}{2}(1-\lambda)\cdot 2p=\lambda p+(1-\lambda)p=p,$$

正如所求。由于两枚硬币互相独立，X 和 Y 亦独立，所以没有协方差（covariance）项，而

$$\begin{aligned}\operatorname{Var}(T)&=\lambda^2\operatorname{Var}(X)+\tfrac{1}{4}(1-\lambda)^2\operatorname{Var}(Y)\\&=\lambda^2p(1-p)+\tfrac{1}{2}(1-\lambda)^2p(1-2p).\end{aligned}$$

求方差的最小值

记 $V(\lambda) = \mathrm{Var}(T)$。这是 λ 的二次函数，首项系数 $p(1-p) + \frac{1}{2}p(1-2p)$ 为正（因为 $0 < p < \frac{1}{2}$，两项皆为正），所以它只有一个最小值，出现在 $V'(\lambda) = 0$ 之处：

$$V'(\lambda) = 2\lambda p(1-p) - (1-\lambda)p(1-2p) = 0.$$

两边除以 $p > 0$，再把含 λ 的项合并，

$$\lambda[2(1-p) + (1-2p)] = 1 - 2p \quad \Longrightarrow \quad \lambda = \frac{1-2p}{3-4p},$$

这正是题目给出的值。（这里 $3 - 4p > 1 > 0$，所以不会出现除以零的情况。）

我们还需要最小值本身。设 $a = p(1-p)$ 及 $c = \frac{1}{2}p(1-2p)$，则 $V = a\lambda^2 + c(1-\lambda)^2$，在 $\lambda = c/(a+c)$（即上面的值）处取得最小值，由此得

$$V_{\min} = a\Big(\frac{c}{a+c}\Big)^2 + c\Big(\frac{a}{a+c}\Big)^2 = \frac{ac(a+c)}{(a+c)^2} = \frac{ac}{a+c}.$$

现在 $a + c = \frac{1}{2}p[2(1-p) + (1-2p)] = \frac{1}{2}p(3-4p)$ 而 $ac = \frac{1}{2}p^2(1-p)(1-2p)$，所以

$$V_{\min} = \frac{p(1-p)(1-2p)}{3-4p}.$$

n 次掷币的平均值

设 $T_1, \ldots, T_n$ 为 n 次掷币中 T 的取值；它们互相独立，各自的平均值为 p，方差为 $V(\lambda)$。于是

$$\mathrm{E}(\overline{T}) = p, \qquad \mathrm{Var}(\overline{T}) = \frac{V(\lambda)}{n}.$$

想一想 T 只取四个值。为甚么我们可以把 $\overline{T}$ 当作正态变量来处理?

因为 $\overline{T}$ 是 n 个独立同分布（independent, identically distributed）且方差有限的变量的平均值，而 $n > 30$ 够大，所以根据中心极限定理，不论单独一个 T 的分布形状如何，$\overline{T}$ 都近似服从 $\mathrm{N}(p, V(\lambda)/n)$。记 $\sigma_\lambda = \sqrt{V(\lambda)/n}$ 及 $Z = (\overline{T} - p)/\sigma_\lambda \approx \mathrm{N}(0,1)$，则

$$\mathrm{P}(|\overline{T} - p| < b) = \mathrm{P}\Big(|Z| < \frac{b}{\sigma_\lambda}\Big) \approx \Phi\Big(\frac{b}{\sigma_\lambda}\Big) - \Phi\Big(-\frac{b}{\sigma_\lambda}\Big) = 2\Phi\Big(\frac{b}{\sigma_\lambda}\Big) - 1,$$

其中用到对称性 $\Phi(-z) = 1 - \Phi(z)$。

由于 $b > 0$ 固定，而 Φ 严格递增（strictly increasing），上式在 b/σ_λ 最大时最大，亦即在 σ_λ 最小时最大，也就是 $V(\lambda)$ 取最小值之时。这发生在 $\lambda = (1-2p)/(3-4p)$，此时

$$\sigma_\lambda^2 = \frac{V_{\min}}{n} = \frac{p(1-p)(1-2p)}{(3-4p)n} = s^2.$$

所以当 λ 变动时，$\mathrm{P}(|\overline{T} - p| < b)$ 的最大值约为 $2\Phi(b/s) - 1$，正如所求。

验算

- 当 $\lambda = 1$ 时，$T = X$，公式给出 $\operatorname{Var}(T) = p(1-p)$；当 $\lambda = 0$ 时，$T = \frac{1}{2}Y$，公式给出 $\frac{1}{4} \cdot 2p(1-2p)$。两者都正确。
- $V_{\min}$ 比这两个值都小：它等于 $p(1-p)$ 乘以 $\frac{1-2p}{3-4p} < 1$。把两枚硬币结合起来，胜过单独使用其中任何一枚。
- 数值上，取 $p = 0.3$、$n = 100$、$b = 0.03$：最佳的 λ 是 $\frac{2}{9}$，$s \approx 0.0216$，而 $2\Phi(b/s) - 1 \approx 0.835$。模拟 40 000 轮、每轮掷币 100 次，得出 0.836；而 $\lambda = 1$（只用硬币 A）则只得约 0.487。

常见错误

- 写成 $\operatorname{Var}(\frac{1}{2}(1-\lambda)Y) = \frac{1}{2}(1-\lambda)\operatorname{Var}(Y)$：常数从方差中提出来时要平方。
- 忘记说明为甚么把方差减至最小就能使概率最大：这需要 Φ 递增，以及对每个 λ 而言 $\overline{T}$ 的平均值都是 p。

STEP 之外

这是估计理论（estimation theory）的初次体验：在 p 的所有无偏估计量 $\lambda X + \frac{1}{2}(1-\lambda)Y$ 之中，我们找出了方差最小的一个。答案是按各个观测值的方差成反比来加权，这与统计学中用来结合精确度不同的量度结果的法则相同（逆方差加权，inverse-variance weighting），而最小值 $ac/(a+c)$ 正是“并联电阻”公式 $1/V_{\min} = 1/a + 1/c$。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门教材）涵盖了最后一步背后的中心极限定理。

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。