详细解答

STEP II 2003，第 14 题

概率与统计 · 假设检验的错误

STEP II 2003，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/03-s2-q14

关键思路

掷出第一个 6 所需的投掷次数服从几何分布（geometric distribution），而它的尾部概率形式极为简洁：事件“需要多于 M 次投掷”其实就是“首 M 次投掷全部失败”。因此两个错误概率都纯粹是 M 次幂，而要在两者之间消去 M，只需取对数（logarithm）即可。

解答

两个错误概率

把各次投掷视为互相独立（independent），每次掷出 6 的概率同为 p。设 X 为直至首次掷出 6 为止（包括该次）的投掷次数，则

$$X > M \iff \text{首 } M \text{ 次投掷中没有一次是 6}, \qquad \text{所以} \qquad \mathrm{P}(X > M) = (1-p)^M,$$

这里用到了各次投掷的独立性。（毋须求几何级数（geometric series）$\sum p(1-p)^{k-1}$ 的和：直接考虑事件本身更快捷。）

这个检验恰好在 $X > M$ 时接受 H_1。因此 α，即实际上 $p = B$ 时接受 H_1 的概率，是

$$\alpha = (1-B)^M,$$

而 β，即实际上 $p = A$ 时接受 H_0（亦即 $X \le M$）的概率，是

$$\beta = 1 - \mathrm{P}(X > M \mid p = A) = 1 - (1-A)^M.$$

消去 M

我们想把 $(1-A)^M$ 写成 $\alpha = (1-B)^M$ 的幂。由于 $0 < B < 1$，$\ln(1-B)$ 有定义且不等于零，所以对 α 取对数可得

$$M = \frac{\ln \alpha}{\ln(1-B)}.$$

于是

$$(1-A)^M = \mathrm{e}^{M\ln(1-A)} = \exp\left(\frac{\ln(1-A)}{\ln(1-B)}\ln\alpha\right) = \alpha^K, \qquad \text{其中} \quad K = \frac{\ln(1-A)}{\ln(1-B)}.$$

因此

$$\beta = 1-\alpha^K, \qquad K = \frac{\ln(1-A)}{\ln(1-B)},$$

而 K 只取决于 A 和 B，与 M 无关，正如题目所求。

? **想一想** 关于 K 的大小，我们可以说些甚么？

由于 $0 < A < B < 1$，我们有 $1 > 1-A > 1-B > 0$，所以两个对数都是负数，且 $\ln(1-B) < \ln(1-A) < 0$。两者相除，可知 K 是正数，并且

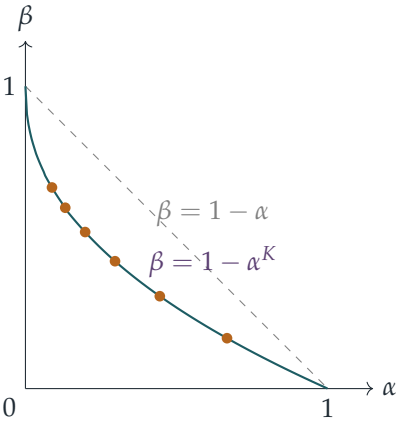
$$0 < K < 1.$$

图像的形状正是由此决定。

β 对 α 的图像

当 M 依次取 $1, 2, 3, \ldots$ 时，$\alpha = (1-B)^M$ 在 $(0,1)$ 内取值，并递减趋向 0；严格来说，图像是一列点，全部位于曲线 $\beta = 1-\alpha^K$，$0 \le \alpha \le 1$ 之上。对于这条曲线：

- 它通过 $(0,1)$ 和 $(1,0)$；
- $\dfrac{\mathrm{d}\beta}{\mathrm{d}\alpha} = -K\alpha^{K-1} < 0$，所以 β 随 α 增加而减少：令其中一种错误较难发生，另一种错误就会较易发生；
- 由于 $K - 1 < 0$，当 $\alpha \to 0^+$ 时斜率（gradient）趋向 $-\infty$（曲线从 $(0,1)$ 垂直出发），而在 $\alpha = 1$ 处斜率等于 $-K$，介乎 -1 与 0 之间；
- $\dfrac{\mathrm{d}^2\beta}{\mathrm{d}\alpha^2} = K(1-K)\alpha^{K-2} > 0$，所以曲线是凸的（convex）；
- 对于 $0 < \alpha < 1$ 和 $0 < K < 1$，我们有 $\alpha^K > \alpha$，所以 $\beta < 1-\alpha$：曲线位于连接 $(0,1)$ 与 $(1,0)$ 的直线下方。



图中的曲线是按 $A=\frac{1}{6}$、$B=\frac{1}{3}$（所以 $K\approx 0.45$）绘画的，并标出了 $M=1,\ldots,6$ 所对应的点；M 增加时，点会向左上方移动。

验算

- 取 $A=\frac{1}{6}$、$B=\frac{1}{3}$、$M=4$。直接计算，$\alpha=(\frac{2}{3})^4\approx 0.1975$，$\beta=1-(\frac{5}{6})^4\approx 0.5177$。以 $K=\ln\frac{5}{6}/\ln\frac{2}{3}\approx 0.4497$ 计算，$1-\alpha^K\approx 0.5177$，结果相同。把这个检验模拟 2×10^5 次，所得结果为 0.198 和 0.518。
- 极端情况的表现亦合乎常理：若 A 接近 B，则 $K\approx 1$ 而 $\beta\approx 1-\alpha$，这个检验比起不理会投掷结果、随机作决定好不了多少；若 A 极小，则 $K\approx 0$，对于任何合理的 α，都可以令 β 变得很小。

常见错误

- 混淆两个假设（hypothesis）：这里 H_0 是较大的值 B，所以等了很久才掷出 6，是不利于 H_0 的证据。
- 把 K 上下颠倒，写成 $\ln(1-B)/\ln(1-A)$，其值大于 1；检查 $\beta<1-\alpha$（任何合理的检验都胜过乱猜）便可发现这个错误。

STEP 之外

β 对 α 的曲线，就是两类错误之间的取舍曲线（trade-off curve），与统计学和机器学习中广泛使用的 ROC 曲线（ROC curve）关系密切。奈曼–皮尔逊引理（Neyman–Pearson lemma）表明，对于像这样的两个简单假设（simple hypotheses），基于似然比（likelihood

ratio）的检验能就每个 α 给出最小可能的 β；这里似然比对 X 是单调的（monotonic），所以“当 X 很大时拒绝 H_0”这个检验其实是最优的。想进一步了解这些概念，可先阅读 Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门）打好基础，然后再读任何一门数理统计（mathematical statistics）的入门课程。

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。