詳細解答

STEP II 2003，第 14 題

概率與統計 • 假設檢驗的錯誤

> **STEP II 2003，第 14 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：
>
> step.maths.org/questions/03-s2-q14

關鍵思路

擲出第一個 6 所需的投擲次數服從幾何分佈（geometric distribution），而它的尾部概率形式極為簡潔：事件「需要多於 M 次投擲」其實就是「首 M 次投擲全部失敗」。因此兩個錯誤概率都純粹是 M 次冪，而要在兩者之間消去 M，只需取對數（logarithm）即可。

解答

兩個錯誤概率

把各次投擲視為互相獨立（independent），每次擲出 6 的概率同為 p。設 X 為直至首次擲出 6 為止（包括該次）的投擲次數，則

$$X > M \iff \text{首 } M \text{ 次投擲中沒有一次是 6}, \qquad \text{所以} \qquad \mathrm{P}(X > M) = (1-p)^M,$$

這裏用到了各次投擲的獨立性。（毋須求幾何級數（geometric series）$\sum p(1-p)^{k-1}$ 的和：直接考慮事件本身更快捷。）

這個檢驗恰好在 $X > M$ 時接受 H_1。因此 α，即實際上 $p = B$ 時接受 H_1 的概率，是

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

而 β，即實際上 $p = A$ 時接受 H_0（亦即 $X \leq M$）的概率，是

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

消去 M

我們想把 $(1-A)^M$ 寫成 $\alpha = (1-B)^M$ 的冪。由於 $0 < B < 1$，$\ln(1-B)$ 有定義且不等於零，所以對 α 取對數可得

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

於是

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

因此

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

而 K 只取決於 A 和 B，與 M 無關，正如題目所求。

? **想一想** 關於 K 的大小，我們可以說些甚麼？

由於 $0 < A < B < 1$，我們有 $1 > 1-A > 1-B > 0$，所以兩個對數都是負數，且 $\ln(1-B) < \ln(1-A) < 0$。兩者相除，可知 K 是正數，並且

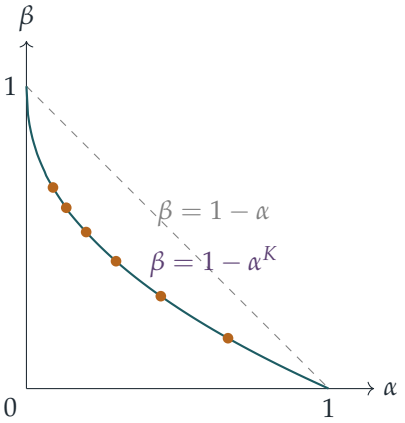
$$0 < K < 1.$$

圖像的形狀正是由此決定。

β 對 α 的圖像

當 M 依次取 $1, 2, 3, \ldots$ 時，$\alpha = (1-B)^M$ 在 $(0,1)$ 內取值，並遞減趨向 0；嚴格來說，圖像是一列點，全部位於曲線 $\beta = 1-\alpha^K$，$0 \le \alpha \le 1$ 之上。對於這條曲線：

- 它通過 $(0,1)$ 和 $(1,0)$；
- $\dfrac{\mathrm{d}\beta}{\mathrm{d}\alpha} = -K\alpha^{K-1} < 0$，所以 β 隨 α 增加而減少：令其中一種錯誤較難發生，另一種錯誤就會較易發生；
- 由於 $K - 1 < 0$，當 $\alpha \to 0^+$ 時斜率（gradient）趨向 $-\infty$（曲線從 $(0,1)$ 垂直出發），而在 $\alpha = 1$ 處斜率等於 $-K$，介乎 -1 與 0 之間；
- $\dfrac{\mathrm{d}^2\beta}{\mathrm{d}\alpha^2} = K(1-K)\alpha^{K-2} > 0$，所以曲線是凸的（convex）；
- 對於 $0 < \alpha < 1$ 和 $0 < K < 1$，我們有 $\alpha^K > \alpha$，所以 $\beta < 1-\alpha$：曲線位於連接 $(0,1)$ 與 $(1,0)$ 的直線下方。



圖中的曲線是按 $A = \frac{1}{6}$、$B = \frac{1}{3}$（所以 $K \approx 0.45$）繪畫的，並標出了 $M = 1, \ldots, 6$ 所對應的點；M 增加時，點會向左上方移動。

驗算

- 取 $A = \frac{1}{6}$、$B = \frac{1}{3}$、$M = 4$。直接計算，$\alpha = (\frac{2}{3})^4 \approx 0.1975$，$\beta = 1 - (\frac{5}{6})^4 \approx 0.5177$。以 $K = \ln \frac{5}{6} / \ln \frac{2}{3} \approx 0.4497$ 計算，$1 - \alpha^K \approx 0.5177$，結果相同。把這個檢驗模擬 2×10^5 次，所得結果為 0.198 和 0.518。
- 極端情況的表現亦合乎常理：若 A 接近 B，則 $K \approx 1$ 而 $\beta \approx 1 - \alpha$，這個檢驗比起不理會投擲結果、隨機作決定好不了多少；若 A 極小，則 $K \approx 0$，對於任何合理的 α，都可以令 β 變得很小。

常見錯誤

- 混淆兩個假設（hypothesis）：這裏 H_0 是較大的值 B，所以等了很久才擲出 6，是不利於 H_0 的證據。
- 把 K 上下顛倒，寫成 $\ln(1 - B) / \ln(1 - A)$，其值大於 1；檢查 $\beta < 1 - \alpha$（任何合理的檢驗都勝過亂猜）便可發現這個錯誤。

STEP 之外

β 對 α 的曲線，就是兩類錯誤之間的取捨曲線（trade-off curve），與統計學和機器學習中廣泛使用的 ROC 曲線（ROC curve）關係密切。奈曼–皮爾遜引理（Neyman–Pearson lemma）表明，對於像這樣的兩個簡單假設（simple hypotheses），基於似然比（likelihood

ratio）的檢驗能就每個 α 給出最小可能的 β；這裏似然比對 X 是單調的（monotonic），所以「當 X 很大時拒絕 H_0」這個檢驗其實是最優的。想進一步了解這些概念，可先閱讀 Grimmett 與 Welsh 的 *Probability: An Introduction*（概率論入門）打好基礎，然後再讀任何一門數理統計（mathematical statistics）的入門課程。

Beyond STEP 詳細解答：https://beyondstep.keithmcnulty.org/zh-hant/。解答文字採用 CC0 1.0；題目本身版權屬 Cambridge University Press & Assessment 所有（©），本文件不予轉載。由機器從英文翻譯而成；尚未經審閱。