詳細解答

STEP I 2017，第 12 題

概率與統計 • 期望值與條件化

> **STEP I 2017，第 12 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/17-s1-q12

關鍵思路

主辦者收取的款項永遠相同，都是 Nc；唯一的不確定之處，在於究竟需不需要派發獎金。因此期望利潤（expected profit）等於收入減去 J 乘以「有人選中中獎號碼」的概率，而整個問題最終歸結為其互補事件（complementary event），即沒有人選中。這個概率是一個冪 $(1-p)^N$，而近似式 $(*)$ 正是為處理這類式子而設的。

解答

第 (i) 部分

設 W 為中獎號碼，並設 E 為沒有參加者選中 W 的事件。無論 W 結果是甚麼，每位參加者都獨立（independently）地以概率 $1-\frac{1}{N}$ 避開它，所以

$$\mathrm{P}(E) = \left(1-\frac{1}{N}\right)^N.$$

在任何情況下，主辦者都收到 Nc。若 E 發生，則無人中獎，利潤為 Nc；否則獎金由中獎者瓜分，但不論中獎者有多少人，派出的總額都是 J，所以利潤為 $Nc-J$。因此

$$\text{期望利潤} = Nc - J\,\mathrm{P}(E^c) = Nc - J\left[1-\left(1-\frac{1}{N}\right)^N\right].$$

在 $(*)$ 中取 $a=1$，得 $(1-\frac{1}{N})^N \approx \mathrm{e}^{-1}$，所以期望利潤約為 $Nc - J(1-\mathrm{e}^{-1})$。若 $J = 2Nc$，則變為

$$Nc - 2Nc\left(1-\mathrm{e}^{-1}\right) = Nc\left(\frac{2}{\mathrm{e}}-1\right).$$

由於 $\mathrm{e} > 2$，我們有 $2/\mathrm{e} < 1$，所以此值為負：主辦者預期會虧蝕。

第 (ii) 部分

共有 γN 個熱門號碼和 $(1-\gamma)N$ 個冷門號碼。每位參加者恰好選一個號碼，所以全部 N 個號碼的概率之和為 1：

$$\gamma N \cdot \frac{a}{N} + (1-\gamma)N \cdot \frac{b}{N} = 1 \quad \Longrightarrow \quad \gamma a + (1-\gamma)b = 1.$$

主辦者仍然均勻地（uniformly）選取 W，所以 W 是熱門號碼的概率為 γ，是冷門號碼的概率為 $1-\gamma$。

? **想一想** 參加者不再均勻地選號。這會否改變我們求 $\mathrm{P}(E)$ 的方法？

只會透過避開 W 的概率產生影響，而這個概率現在取決於 W 屬於哪一類號碼。因此我們以此作條件（condition）。若 W 是熱門號碼，每位參加者獨立地以概率 $1-\frac{a}{N}$ 避開它；若是冷門號碼，則以概率 $1-\frac{b}{N}$ 避開。由全概率定律（law of total probability）及 $(*)$，

$$\mathrm{P}(E) = \gamma\left(1-\frac{a}{N}\right)^N + (1-\gamma)\left(1-\frac{b}{N}\right)^N \approx \gamma\,\mathrm{e}^{-a} + (1-\gamma)\,\mathrm{e}^{-b}.$$

與第 (i) 部分一樣，期望利潤為 $Nc - J(1-\mathrm{P}(E))$，所以

$$\text{期望利潤} \approx \gamma J\,\mathrm{e}^{-a} + (1-\gamma)J\,\mathrm{e}^{-b} + (Nc - J),$$

這正是所要求的形式，其中

$$A = \gamma J, \qquad B = (1-\gamma)J, \qquad C = Nc - J.$$

情況 $\gamma = \frac{1}{8}$，$a = 9b$。 由上述關係得 $\frac{1}{8}\cdot 9b + \frac{7}{8}b = 1$，即 $2b = 1$，所以

$$b = \tfrac{1}{2}, \qquad a = \tfrac{9}{2}.$$

當 $J = 2Nc$ 時，我們有 $A = \frac{1}{4}Nc$、$B = \frac{7}{4}Nc$ 及 $C = -Nc$，所以期望利潤約為

$$Nc\left(\tfrac{1}{4}\mathrm{e}^{-9/2} + \tfrac{7}{4}\mathrm{e}^{-1/2} - 1\right).$$

括號內第一項為正，所以只需證明 $\frac{7}{4}\mathrm{e}^{-1/2} > 1$，即 $\sqrt{\mathrm{e}} < \frac{7}{4}$。由於 $\mathrm{e} < 3$，

$$\sqrt{\mathrm{e}} < \sqrt{3} < \tfrac{7}{4},$$

其中最後一步成立，是因為兩邊皆為正，且 $3 < \frac{49}{16}$（即 $48 < 49$）。因此括號內的值為正，主辦者預期會獲利。獎金與第 (i) 部分相同，但由於中獎號碼往往是冷門號碼，獎金無人認領的情況更為常見。

驗算

- 當 $J = 2Nc$ 時，第 (i) 部分給出的期望利潤約為 $(2/\mathrm{e} - 1)Nc \approx -0.264\,Nc$，而最後一部分給出 $\left(\frac{1}{4}\mathrm{e}^{-4.5} + \frac{7}{4}\mathrm{e}^{-0.5} - 1\right)Nc \approx 0.064\,Nc$。正如所述，是一筆小額利潤。
- 當 $N = 800$ 時，用精確的冪 $(1 - a/N)^N$ 計算得 $0.0640\,Nc$，可見近似式 $(*)$ 在此非常準確。
- 若 $\gamma = 1$（每個號碼都是「熱門」的），上述關係迫使 $a = 1$，而第 (ii) 部分的公式便化為第 (i) 部分的 $J\mathrm{e}^{-1} + Nc - J$。
- 以電腦對 $N = 4$ 作完全枚舉（參加者選號的全部 4^4 種方式），結果與兩條未經近似的公式皆吻合。

常見錯誤

- 糾結於獎金如何分配。只要至少有一人中獎，不論中獎人數多少，主辦者派出的總額都是 J。
- 誤以為中獎號碼是熱門號碼的概率取決於 a 或 b。主辦者是均勻選號的，所以它是熱門號碼的概率為 γ。
- 把關係式寫成 $a + b = 1$ 或 $\gamma a + b = 1$，忘記計算每類號碼各有多少個。

STEP 之外

選中中獎號碼的參加者人數服從參數為 N 和 p 的二項分佈（binomial distribution），而 $(*)$ 正是理解以下事實的第一步：當 N 很大而 Np 固定時，它近似於泊松分佈（Poisson distribution）——$(1-p)^N \to \mathrm{e}^{-Np}$ 不過就是泊松分佈取值為零的概率。現實中設有獎金累積（roll-over）的彩票正是以這種方式分析的，而選擇冷門號碼的玩家確實有較高的期望回報。Grimmett 與 Welsh 的 *Probability: An Introduction*（《概率導論》）一書，詳細發展了泊松極限，以及類似本題所用的條件論證。

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。