詳細解答

STEP I 1987，第 16 題

概率與統計 • 貝葉斯定理與正態近似

STEP I 1987，第 16 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/87-s1-q16

關鍵思路

第 (i) 部分其實是喬裝了的貝葉斯定理（Bayes' theorem）。兩次投票立場相同，是判斷該議員屬於哪個政黨的證據（evidence），因為守成黨（Preservative）議員必定如此，而進步黨（Progressive）議員只是有時如此。所以我們先更新該議員是守成黨議員的概率，然後才預測第三次投票。在第 (ii) 部分，守成黨議員的投票是固定的數目，所以所有隨機性都來自那 40 位進步黨議員。他們投贊成票的人數服從二項分佈（binomial distribution），而我們要用正態分佈（normal distribution）來近似的，正是這個分佈。

解答

第 (i) 部分

設 R 為「被選中的議員是守成黨議員」這事件，故 $P(R) = \frac{60}{100} = \frac{3}{5}$，而 $P(R') = \frac{2}{5}$。設 S_2 為「該議員首兩次投票立場相同」這事件，S_3 為「第三次投票的立場亦相同」這事件。

守成黨議員從不改變主意，所以 $P(S_2 \mid R) = 1$。進步黨議員則隨機投票，每次投票都獨立（independently）地以概率 $\frac{1}{2}$ 投贊成或反對。無論第一次投了甚麼，第二次與之相同的概率都是 $\frac{1}{2}$，所以 $P(S_2 \mid R') = \frac{1}{2}$。因此

$$P(S_2) = \tfrac{3}{5} \cdot 1 + \tfrac{2}{5} \cdot \tfrac{1}{2} = \tfrac{4}{5}, \qquad P(R \mid S_2) = \frac{P(R)\,P(S_2 \mid R)}{P(S_2)} = \frac{3/5}{4/5} = \frac{3}{4}.$$

所以，根據我們觀察所得，該議員是守成黨議員的概率為 $\frac{3}{4}$，是進步黨議員的概率為 $\frac{1}{4}$。

? **想一想** 既然該議員是隨機選出的，為甚麼我們不能直接說 $P(R \mid S_2) = \frac{3}{5}$？

因為資訊 S_2 對守成黨議員而言比對進步黨議員更有可能出現，所以它會改變勝算比

（odds）。觀察之前，勝算比是 3 : 2，偏向守成黨議員。觀察到 S_2 之後，勝算比變成 $3\cdot 1 : 2\cdot \frac{1}{2} = 3:1$。

現在，在已知該議員所屬政黨的條件下，第三次投票與首兩次投票無關，首兩次投票的作用只在於透露該議員屬於哪個政黨。守成黨議員必定重複同一立場，進步黨議員則以概率 $\frac{1}{2}$ 投出相同立場。所以

$$P(S_3 \mid S_2) = \tfrac{3}{4}\cdot 1 + \tfrac{1}{4}\cdot\tfrac{1}{2} = \frac{7}{8}.$$

（等價地，$P(S_3 \mid S_2) = \dfrac{P(S_2 \cap S_3)}{P(S_2)} = \dfrac{\frac{3}{5} + \frac{2}{5}\cdot\frac{1}{4}}{\frac{4}{5}} = \dfrac{7/10}{4/5} = \dfrac{7}{8}$。）

第 (ii) 部分

考慮就一項守成黨政策進行的投票，並假設全部 100 位議員都投票。60 位守成黨議員中有比例 α，即 60α 位議員，總是投反對票，其餘 $60(1-\alpha)$ 位則總是投贊成票。設 X 為投贊成票的進步黨議員人數。每位進步黨議員獨立地以概率 $\frac{1}{2}$ 投贊成票，所以

$$X \sim \mathrm{B}\left(40, \tfrac{1}{2}\right), \qquad E(X) = 20, \qquad \mathrm{Var}(X) = 40\cdot\tfrac{1}{2}\cdot\tfrac{1}{2} = 10.$$

在 100 票中取得簡單多數（simple majority），即至少有 51 票贊成：

$$60(1-\alpha) + X \geq 51 \iff X \geq 60\alpha - 9.$$

由於 $n = 40$ 相當大，而 $p = \frac{1}{2}$ 令二項分佈對稱，我們用 $Y \sim \mathrm{N}(20, 10)$ 來近似 X。這裏 $60\alpha - 9$ 是整數（它是議員人數），所以加上連續性校正（continuity correction）後，

$$P(\text{過半數}) = P(X \geq 60\alpha - 9) \approx P(Y > 60\alpha - 9.5) = P\left(Z > \frac{60\alpha - 29.5}{\sqrt{10}}\right),$$

其中 $Z \sim \mathrm{N}(0,1)$。

這個概率隨 α 增加而遞減，所以恰好當 α 大得足以把它壓到 0.8 以下時，便會舉行選舉。查表得 $P(Z > -0.8416) = 0.8$，所以我們需要

$$\frac{60\alpha - 29.5}{\sqrt{10}} > -0.8416 \iff 60\alpha > 29.5 - 0.8416\sqrt{10} = 26.84$$

$$\iff \alpha > 0.447.$$

由於 60α 必須是整數（議員人數），最小的可能值是 $60\alpha = 27$。所以導致選舉的 α 最小值為

$$\alpha = \frac{27}{60} = 0.45.$$

驗證這個答案：有 26 位倒戈議員時，取得過半數的近似概率為 $P(Z > -1.107) \approx 0.866$；有 27 位倒戈議員時，則為 $P(Z > -0.791) \approx 0.785 < 0.8$。

驗算

- 在 (i) 中，$\frac{7}{8}$ 介乎 $\frac{1}{2}$（已知是進步黨議員）與 1（已知是守成黨議員）之間，理應如此。以 400,000 位隨機議員進行模擬，得出 0.875。
- 在 (ii) 中，精確的二項概率為 $P(X \geq 17) = 0.8659$（26 位倒戈議員）及 $P(X \geq 18) = 0.7852$（27 位倒戈議員）。這些數值與正態近似值在三位小數內一致，並確認分界點位於 26 與 27 位倒戈議員之間，即 $\alpha = 0.45$。

常見錯誤

- 在 (i) 中答 $\frac{3}{5} \cdot 1 + \frac{2}{5} \cdot \frac{1}{2} = \frac{4}{5}$。這忽略了「該議員已經兩次投出相同立場」這項資訊。
- 把簡單多數當作「至少 50 票」。50 對 50 打和並不算過半數，所以需要 51 票。
- 漏了連續性校正。沒有它的話，臨界值會變成 $\alpha > 0.456$，從而得出 28 位倒戈議員而不是 27 位；根據精確的二項計算，這是錯的。

STEP 之外

第 (i) 部分是貝葉斯更新（Bayesian updating）的一個小例子：數據改變我們對某個隱藏「類型」的信念，然後預測就是對更新後的信念取平均（即後驗預測分佈（posterior predictive distribution））。第 (ii) 部分的正態近似是中央極限定理(central limit theorem)的一個例子，而這裏近乎完美的吻合並非偶然，因為當 $p = \frac{1}{2}$ 時二項分佈是對稱的。這兩個概念在 Blitzstein 與 Hwang 的 *Introduction to Probability*（《概率導論》）中都有仔細的闡述。

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。