详细解答

STEP I 1987，第 16 题

概率与统计 • 贝叶斯定理与正态近似

STEP I 1987，第 16 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/87-s1-q16

关键思路

第 (i) 部分其实是乔装了的贝叶斯定理（Bayes' theorem）。两次投票立场相同，是判断该议员属于哪个政党的证据（evidence），因为守成党（Preservative）议员必定如此，而进步党（Progressive）议员只是有时如此。所以我们先更新该议员是守成党议员的概率，然后才预测第三次投票。在第 (ii) 部分，守成党议员的投票是固定的数目，所以所有随机性都来自那 40 位进步党议员。他们投赞成票的人数服从二项分布（binomial distribution），而我们要用正态分布（normal distribution）来近似的，正是这个分布。

解答

第 (i) 部分

设 R 为“被选中的议员是守成党议员”这事件，故 $P(R) = \frac{60}{100} = \frac{3}{5}$，而 $P(R') = \frac{2}{5}$。设 S_2 为“该议员首两次投票立场相同”这事件，S_3 为“第三次投票的立场亦相同”这事件。

守成党议员从不改变主意，所以 $P(S_2 \mid R) = 1$。进步党议员则随机投票，每次投票都独立（independently）地以概率 $\frac{1}{2}$ 投赞成或反对。无论第一次投了甚么，第二次与之相同的概率都是 $\frac{1}{2}$，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，根据我们观察所得，该议员是守成党议员的概率为 $\frac{3}{4}$，是进步党议员的概率为 $\frac{1}{4}$。

? **想一想** 既然该议员是随机选出的，为甚么我们不能直接说 $P(R \mid S_2) = \frac{3}{5}$?

因为信息 S_2 对守成党议员而言比对进步党议员更有可能出现，所以它会改变胜算比

（odds）。观察之前，胜算比是 $3:2$，偏向守成党议员。观察到 S_2 之后，胜算比变成 $3\cdot 1 : 2\cdot \frac{1}{2} = 3:1$。

现在，在已知该议员所属政党的条件下，第三次投票与首两次投票无关，首两次投票的作用只在于透露该议员属于哪个政党。守成党议员必定重复同一立场，进步党议员则以概率 $\frac{1}{2}$ 投出相同立场。所以

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

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

第 (ii) 部分

考虑就一项守成党政策进行的投票，并假设全部 100 位议员都投票。60 位守成党议员中有比例 α，即 60α 位议员，总是投反对票，其余 $60(1-\alpha)$ 位则总是投赞成票。设 X 为投赞成票的进步党议员人数。每位进步党议员独立地以概率 $\frac{1}{2}$ 投赞成票，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 票中取得简单多数（simple majority），即至少有 51 票赞成：

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

由于 $n = 40$ 相当大，而 $p = \frac{1}{2}$ 令二项分布对称，我们用 $Y \sim \mathrm{N}(20, 10)$ 来近似 X。这里 $60\alpha - 9$ 是整数（它是议员人数），所以加上连续性校正（continuity correction）后，

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

其中 $Z \sim \mathrm{N}(0,1)$。

这个概率随 α 增加而递减，所以恰好当 α 大得足以把它压到 0.8 以下时，便会举行选举。查表得 $P(Z > -0.8416) = 0.8$，所以我们需要

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

$$\iff \alpha > 0.447.$$

由于 60α 必须是整数（议员人数），最小的可能值是 $60\alpha = 27$。所以导致选举的 α 最小值为

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

验证这个答案：有 26 位倒戈议员时，取得过半数的近似概率为 $P(Z > -1.107) \approx 0.866$；有 27 位倒戈议员时，则为 $P(Z > -0.791) \approx 0.785 < 0.8$。

验算

- 在 (i) 中，$\frac{7}{8}$ 介乎 $\frac{1}{2}$（已知是进步党议员）与 1（已知是守成党议员）之间，理应如此。以 400,000 位随机议员进行模拟，得出 0.875。
- 在 (ii) 中，精确的二项概率为 $P(X \geq 17) = 0.8659$（26 位倒戈议员）及 $P(X \geq 18) = 0.7852$（27 位倒戈议员）。这些数值与正态近似值在三位小数内一致，并确认分界点位于 26 与 27 位倒戈议员之间，即 $\alpha = 0.45$。

常见错误

- 在 (i) 中答 $\frac{3}{5} \cdot 1 + \frac{2}{5} \cdot \frac{1}{2} = \frac{4}{5}$。这忽略了“该议员已经两次投出相同立场”这项信息。
- 把简单多数当作“至少 50 票”。50 对 50 打和并不算过半数，所以需要 51 票。
- 漏了连续性校正。没有它的话，临界值会变成 $\alpha > 0.456$，从而得出 28 位倒戈议员而不是 27 位；根据精确的二项计算，这是错的。

STEP 之外

第 (i) 部分是贝叶斯更新（Bayesian updating）的一个小例子：数据改变我们对某个隐藏“类型”的信念，然后预测就是对更新后的信念取平均（即后验预测分布（posterior predictive distribution））。第 (ii) 部分的正态近似是中央极限定理（central limit theorem）的一个例子，而这里近乎完美的吻合并非偶然，因为当 $p = \frac{1}{2}$ 时二项分布是对称的。这两个概念在 Blitzstein 与 Hwang 的 *Introduction to Probability*（《概率导论》）中都有仔细的阐述。

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。