詳細解答

STEP II 2024，第 11 題

概率與統計 • 分組檢測的期望值

STEP II 2024，第 11 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 OCR 提供下載的 2024 年 STEP 試卷（zip 檔案）內的 STEP II 試卷：

www.ocr.org.uk/Images/725753-2024-past-papers.zip

關鍵思路

只有當混合樣本的檢測通常呈陰性時，合併檢測（pooling）才划算；而每人的期望（expected）檢測次數原來是 $1+\frac{1}{k}-(1-p)^k$。要問這個數何時不超過 1，就變成比較 $(1-p)$ 與 $k^{-1/k}$ 的大小，這正是第 (i) 部分要討論 $x^{1/x}$ 的原因：當 p 較大時，最佳的組別人數就是使 $k^{1/k}$ 最大的那個整數。

解答

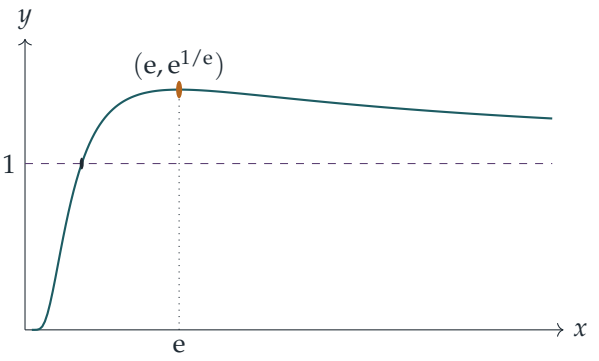
第 (i) 部分

對 $x>0$，記 $y=x^{1/x}=\mathrm{e}^{(\ln x)/x}$，於是 $\ln y=\dfrac{\ln x}{x}$。求導（differentiating）得

$$\frac{1}{y}\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{1-\ln x}{x^2}.$$

由於 $y>0$，$\frac{\mathrm{d}y}{\mathrm{d}x}$ 與 $1-\ln x$ 同號：當 $0<x<\mathrm{e}$ 時為正，在 $x=\mathrm{e}$ 時為零，當 $x>\mathrm{e}$ 時為負。因此曲線只有一個轉折點（turning point），即位於 $(\mathrm{e},\mathrm{e}^{1/\mathrm{e}})\approx(2.72,\,1.44)$ 的極大值（maximum）點。

至於圖像兩端：當 $x\to 0^+$ 時，$(\ln x)/x\to-\infty$，所以 $y\to 0$；當 $x\to\infty$ 時，$(\ln x)/x\to 0$，所以 $y\to 1$，且從上方趨近（因為對 $x>1$ 有 $\ln y>0$）。曲線經過點 $(1,1)$。



對正整數 n 而言，此函數在 $(0, \mathrm{e}]$ 上遞增，在 $[\mathrm{e}, \infty)$ 上遞減，所以 $1^{1/1} < 2^{1/2}$ 且 $3^{1/3} > 4^{1/4} > 5^{1/5} > \cdots$。因此最大值是 $2^{1/2}$ 或 $3^{1/3}$。把兩者都升到六次方，得 $2^3 = 8$ 和 $3^2 = 9$，所以 $3^{1/3} > 2^{1/2}$，且

$$\max_n n^{1/n} = 3^{1/3}.$$

第 (ii) 部分

記 $q = 1 - p$。考慮一組 k 人。由獨立性（independence），組內無人帶有該酶的概率是 q^k，而只有在這種情況下混合樣本才呈陰性。所以該組所用的檢測次數為

$$\begin{cases} 1 \text{ 次檢測} & \text{概率為} q^k, \\ 1 + k \text{ 次檢測} & \text{概率為} 1 - q^k, \end{cases}$$

其期望檢測次數是 $1 + k(1 - q^k)$。共有 $r = N/k$ 組，而期望值具可加性，所以期望總次數為

$$E = \frac{N}{k}(1 + k(1 - q^k)) = N\left(1 + \frac{1}{k} - (1-p)^k\right).$$

第 (iii) 部分

由第 (ii) 部分，$E \le N$ 當且僅當

$$\frac{1}{k} \le (1-p)^k \iff 1 - p \ge k^{-1/k} = \frac{1}{k^{1/k}},$$

這裏取了正的 k 次方根（兩邊均為正數，而 $t \mapsto t^{1/k}$ 是遞增的）。

> ? **想一想** 哪個 k 容許最大的 p？

對給定的 k，條件是 $p \le 1 - k^{-1/k}$，而當 $k^{1/k}$ 最大時，這個上界也最大。由第 (i) 部分，對每個整數 $k > 1$ 都有 $k^{1/k} \le 3^{1/3}$，所以無論選哪個 k，都必須有 $p \le 1 - 3^{-1/3}$。反過來，N 是 3 的倍數，所以 $k = 3$ 是容許的組別人數，而取 $k = 3$ 時條件正好是

$p \le 1 - 3^{-1/3}$（等號成立時 $E = N$）。因此最大值為

$$p = 1 - 3^{-1/3} \approx 0.307.$$

要證明這個值大於 $\frac{1}{4}$：

$$1 - 3^{-1/3} > \tfrac{1}{4} \iff 3^{-1/3} < \tfrac{3}{4} \iff \tfrac{1}{3} < \tfrac{27}{64} \iff 64 < 81,$$

其中中間一步兩邊取立方（對正數而言，取立方保持大小次序）。最後一個式子成立，所以 $1 - 3^{-1/3} > \frac{1}{4}$。

第 (iv) 部分

由二項式展開（binomial expansion），

$$(1-p)^k = 1 - kp + \binom{k}{2}p^2 - \cdots.$$

若 pk 很小，則 $\binom{k}{2}p^2 < \frac{1}{2}(pk)^2$ 更小得多，之後各項亦然，所以 $(1-p)^k \approx 1 - pk$。代入第 (ii) 部分的結果，

$$E \approx N\left(1 + \frac{1}{k} - (1 - pk)\right) = N\left(\frac{1}{k} + pk\right).$$

（事實上，伯努利不等式（Bernoulli's inequality）$(1-p)^k \ge 1 - pk$ 表明這個近似永遠不會低估 E；而把展開式多保留一項，則可證明誤差最多為 $\frac{1}{2}(pk)^2 N$。）

取 $p = 0.01$ 和 $k = 10$，得 $pk = 0.1$，確實很小，而

$$E \approx N\left(\frac{1}{10} + 0.1\right) = 0.2N,$$

所以這個方案只需約 20% 的檢測次數，相比之下逐一檢測每個人需要 N 次。

驗算

- 取 $p = 0.01$ 和 $k = 10$ 作精確計算：$E/N = 1.1 - 0.99^{10} \approx 1.1 - 0.9044 = 0.1956$，約為 19.6%，與近似值之差在誤差上界 $\frac{1}{2}(0.1)^2 = 0.005$ 之內。
- 當 $p = 1 - 3^{-1/3}$ 且 $k = 3$ 時：$(1-p)^3 = \frac{1}{3}$，所以 $E = N(1 + \frac{1}{3} - \frac{1}{3}) = N$，剛好落在邊界上。
- 檢查第 (ii) 部分是否合理：當 $p \to 0$ 時，$E \to N/k$（每組一次檢測）；當 $p \to 1$ 時，$E \to N(1 + \frac{1}{k})$（每組都要重新逐一檢測）。此公式亦已用電腦逐一列舉小組別的所有結果加以驗證。

常見錯誤

- 第 (i) 部分答 $e^{1/e}$：n 必須是整數，所以應比較 $2^{1/2}$ 與 $3^{1/3}$。
- 第 (iii) 部分只驗證了 $k = 3$，卻沒有解釋為何其他組別人數不可能容許更大的 p。
- 在呈陽性的組別中漏計混合檢測本身，得出 k 次而非 $k+1$ 次檢測。

STEP 之外

這就是 Dorfman 的羣組檢測（group testing）方案，它在第二次世界大戰期間為篩查士兵的血液樣本而設計，後來在 COVID-19 檢測中再度被採用。把 $\frac{1}{k} + pk$ 最小化，得 $k \approx 1/\sqrt{p}$（所以這裏 $k = 10$ 是最佳選擇），而所需檢測次數約為 $2\sqrt{p}\,N$；資訊理論（information theory）中研究的更巧妙的自適應（adaptive）方案，效果還要更好。關於背後的概率理論，可參閱 Blitzstein 與 Hwang 的 *Introduction to Probability*（《概率導論》）。

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。