詳細解答

STEP I 1988，第 14 題

概率與統計 • 刪失正態分佈的期望值

STEP I 1988，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/88-s1-q14

關鍵思路

變量 X_M 其實就是 $\min(X, M)$：把正態變量（normal variable）中所有高於 M 的部分一律壓低到 M。它的期望值（expectation）可拆成兩部分：一個在 $x < M$ 上的積分，以及一個常數乘以尾部概率（tail probability）；而這個積分很容易計算，因為 $x\phi(x)$ 除了正負號之外，正是 $\phi(x)$ 的導數。遊客的故事只是同一個想法換了個包裝。一隊巴士所接載的乘客人數是 $\min(\text{需求}, \text{載客量})$，經正態近似（normal approximation）之後，這正好就是一個 X_M。

解答

X_M 的期望值

由於在 $\{X < M\}$ 上 $X_M = X$，而在 $\{X \geq M\}$ 上 $X_M = M$，

$$\mathrm{E}(X_M) = \int_{-\infty}^{M} x\,\phi(x)\,\mathrm{d}x \;+\; M\,\mathrm{P}(X \geq M) = \int_{-\infty}^{M} x\,\phi(x)\,\mathrm{d}x \;+\; M(1-\Phi(M)).$$

? **想一想** 有沒有一個函數，其導數是 $x\phi(x)$？

有。由 $\phi(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}$，按鏈式法則（chain rule）可得 $\phi'(x) = -x\phi(x)$。因此

$$\int_{-\infty}^{M} x\,\phi(x)\,\mathrm{d}x = \Big[-\phi(x)\Big]_{-\infty}^{M} = -\phi(M) + 0,$$

因為當 $x \to -\infty$ 時 $\phi(x) \to 0$。故此

$$\mathrm{E}(X_M) = -\phi(M) + M(1-\Phi(M)),$$

這正是所求的結果。

作為合理性檢查，令 $M \to \infty$ 可得 $\mathrm{E}(X_M) \to 0 = \mathrm{E}(X)$，因為 $M(1-\Phi(M)) \to 0$。令 $M \to -\infty$ 則得 $\mathrm{E}(X_M) \approx M$，與預期相符。

巴士

設 N 為某一次到訪中被指往巴士站的遊客人數。我們假設 1024 名遊客各自獨立地被指路，往兩邊的概率各為 $\frac{1}{2}$。於是

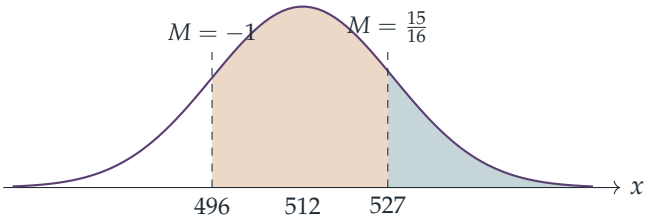
$$N \sim \mathrm{B}(1024, \tfrac{1}{2}), \qquad \mathrm{E}(N) = 512, \qquad \mathrm{Var}(N) = 1024 \cdot \tfrac{1}{2} \cdot \tfrac{1}{2} = 256 = 16^2.$$

由於 $n = 1024$ 這麼大，N 可以非常準確地用 $512 + 16X$ 來近似，其中 X 服從標準正態分佈（standard normal distribution）。

有 16 輛巴士時，載客量是 $16 \times 31 = 496$。有 17 輛時則是 527。載客量為 C 的車隊接載 $\min(N, C)$ 名乘客，所以它每次到訪的利潤（以 vloska 計）是

$$\begin{aligned}\min(N,C) &\approx \min(512 + 16X,\ C) \\ &= 512 + 16\min\Big(X, \frac{C-512}{16}\Big) = 512 + 16\,X_M, \qquad \text{其中} M = \frac{C-512}{16}.\end{aligned}$$

因此 $C = 496$ 對應 $M = -1$，而 $C = 527$ 對應 $M = \frac{15}{16}$。



當需求落在琥珀色的區域時，第 17 輛巴士接載的是超出 496 的那部分乘客。當需求落在藍綠色的尾部時，它載滿 31 人。

利用 $\phi(1) \approx 0.24197$ 和 $\Phi(1) \approx 0.84134$，並留意 $1 - \Phi(-1) = \Phi(1)$，

$$\mathrm{E}(X_{-1}) = -\phi(1) - \big(1 - \Phi(-1)\big) = -\phi(1) - \Phi(1) \approx -0.24197 - 0.84134 = -1.08332.$$

利用 $\phi(0.9375) \approx 0.25707$ 和 $1 - \Phi(0.9375) \approx 0.17425$，

$$\mathrm{E}(X_{15/16}) \approx -0.25707 + 0.9375 \times 0.17425 \approx -0.25707 + 0.16336 = -0.09371.$$

因此，多一輛巴士令每次到訪的期望利潤增加

$$16\big(\mathrm{E}(X_{15/16}) - \mathrm{E}(X_{-1})\big) \approx 16 \times 0.98961 \approx 15.83 \text{ vloska}.$$

每年有 50 次到訪，所以多一輛巴士的價值大約是

$$50 \times 15.83 \approx 792 \text{ vloska（每年）}.$$

這就是值得考慮的每年賄款上限。付出的金額一旦超過期望的額外利潤，合作社平均而言就會得不償失。

> **想一想** 既然 N 是離散的，我們不是應該使用連續性校正（continuity correction）嗎？

這裏並不需要。連續性校正的用處，在於近似 $\mathrm{P}(N \le c)$ 這類概率，因為這時被取平均的函數在整數處有跳躍。函數 $k \mapsto \min(k, C)$ 沒有跳躍，不作校正的正態近似對它已經極為準確。以二項分佈（binomial distribution）作精確計算，結果是 791.68，與上面的數字吻合至小數點後兩位。把載客量移動 $\frac{1}{2}$ 反而會令結果變差，得出 775。

驗算

- 以中點法則作數值積分（midpoint quadrature）計算 $\int_{-\infty}^{M} x\phi(x)\,\mathrm{d}x + M(1-\Phi(M))$，對多個 M 值所得的結果都與公式吻合至小數點後八位。以 4×10^5 個樣本對 $\min(X, M)$ 進行蒙地卡羅模擬（Monte Carlo simulation），結果亦吻合。
- 對 $\mathrm{B}(1024, \frac{1}{2})$ 作精確求和，得出每次到訪的額外期望利潤為 15.8337，即每年 791.68，與正態近似相符。
- 粗略的合理性檢查：只要 $N > 496$，多出的那輛巴士就有收入，這種情況出現的概率約為 0.84，而且它往往滿座。因此每次到訪的數字比 31 的一半略多，是合理的。

常見錯誤

- 忘記換算比例：$\min(N, C)$ 是 $512 + 16X_M$，所以 $\mathrm{E}(X_M)$ 之差必須先乘以標準差（standard deviation）16，再乘以 50。
- $M = -1$ 時的正負號錯誤：$M(1-\Phi(M))$ 一項是 $-(1-\Phi(-1)) = -\Phi(1)$，而不是 $-(1-\Phi(1))$。

STEP 之外

$\mathrm{E}\min(X, M)$ 這個量是一個刪失（censored）正態變量的期望值。與它密切相關的表達式，例如 $\mathrm{E}(X-M)^+ = \phi(M) - M(1-\Phi(M))$，是存貨理論（inventory theory，即「報販問題」（newsvendor problem））以及布萊克–舒爾斯（Black–Scholes）期權定價的骨

幹。恆等式 $\mathrm{E}\min(N, C+1) - \mathrm{E}\min(N, C) = \mathrm{P}(N > C)$ 提供了一個簡潔的方法，讓人看清每多一個單位載客量的價值。想進一步了解這些概念，可參閱 Grimmett 與 Welsh 的 *Probability: An Introduction*（概率論入門教材）。

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。