詳細解答

STEP II 1993，第 16 題

概率與統計 • 泊松過程

> **STEP II 1993，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/93-s2-q16

關鍵思路

這裏有兩個互相獨立的隨機來源：巴士帶來一個幾何分佈（geometric distribution）的等候時間，而乘客則構成一個泊松過程（Poisson process）。先對第一班巴士到站的時間取條件（condition）。一旦知道這個時間，候車人數就只是一段已知長度時間內的泊松計數。第 (iii) 部分接着運用全概率定理（law of total probability），而真正的問題是：只憑統計表和計算機，怎樣把所得的無窮和計算至兩位有效數字。

解答

以 8:oo 之後的分鐘數來量度時間。設 T 為第一班巴士到站的時間，並設 X_t 為在 $(0,t]$ 內到達的乘客人數。乘客過程是速率為每分鐘 4 人的泊松過程，所以 $X_t \sim \mathrm{Po}(4t)$。巴士只可能在 $t = 1,2,3,\ldots$ 到站（「剛過八時」表示可能在 8:oo 到站的那班巴士已經來過並開走了），而且巴士與乘客互相獨立。

第 (i) 部分

第一班巴士在 8 時 n 分到站，當且僅當第 $1,2,\ldots,n-1$ 分鐘都沒有巴士到站，而第 n 分鐘有巴士到站。這些事件互相獨立，所以

$$\mathrm{P}(T = n) = \left(\tfrac{3}{4}\right)^{n-1}\tfrac{1}{4}, \qquad n = 1,2,3,\ldots$$

因此 T 服從參數為 $\frac{1}{4}$ 的幾何分佈。（驗算一下：這些概率之和為 $\frac{1}{4}/(1-\frac{3}{4}) = 1$。）

第 (ii) 部分

在 8:05 候車的人，正是在 $(0,5]$ 內到達的那些人。沒有人離開過，因為 8:05 之前沒有巴士到站。所以我們要求的是在 $T=5$ 的條件下 X_5 的分佈。

> **想一想** 知道第一班巴士在 8:05 到站，會否改變乘客人數的分佈？

不會。事件 $T=5$ 只取決於巴士，而巴士與乘客互相獨立。所以對 $T=5$ 取條件並不改變 X_5 的分佈，即 $X_5 \sim \mathrm{Po}(4\times 5)=\mathrm{Po}(20)$。因此

$$\mathrm{P}(m \text{ 人候車} \mid T=5)=\frac{\mathrm{e}^{-20}\,20^m}{m!}, \qquad m=0,1,2,\ldots$$

第 (iii) 部分

巴士未能接載所有人，當且僅當候車人數超過 25，即 $X_T \geq 26$。設 F 為這個事件。對 T 取條件，並如 (ii) 般運用獨立性，

$$\mathrm{P}(F)=\sum_{n=1}^{\infty}\mathrm{P}(T=n)\,\mathrm{P}(X_n\geq 26)=\sum_{n=1}^{\infty}\left(\tfrac{3}{4}\right)^{n-1}\tfrac{1}{4}\,\mathrm{P}(\mathrm{Po}(4n)\geq 26).$$

現在我們需要一個方案，把這個無窮和化為有限的工作量，而誤差要遠低於 0.005。我們要的是兩位有效數字，而答案最後會是 0.1 這個數量級。

第 1 步：改為處理補事件，令尾項容易處理。 記 $w_n=\left(\tfrac{3}{4}\right)^{n-1}\tfrac{1}{4}$ 及 $q_n=\mathrm{P}(\mathrm{Po}(4n)\leq 25)$。由於 $\sum w_n=1$，

$$\mathrm{P}(F)=1-\sum_{n=1}^{\infty}w_n\,q_n.$$

這個形式較好，因為當 n 很大時 q_n 極小。當 $n\geq 11$ 時，平均值 $4n\geq 44$ 遠高於 25。由於 q_n 隨 n 增加而遞減（時間愈長，乘客愈多），

$$\sum_{n\geq 11}w_n q_n \leq q_{11}\sum_{n\geq 11}w_n = q_{11}\left(\tfrac{3}{4}\right)^{10}.$$

用正態近似（normal approximation）可得 $q_{11}\approx\Phi((25.5-44)/\sqrt{44})=\Phi(-2.79)\approx 0.003$，所以這個尾項小於 0.0002，可以略去。這樣便只剩下十項。

第 2 步：最初幾項。 當 $n=1$ 和 $n=2$ 時（平均值為 4 和 8），有 26 名或以上乘客的可能性極低，在遠超所需的準確度之內 $q_1=q_2=1$，所以這兩項的貢獻是 $w_1+w_2=\tfrac{1}{4}+\tfrac{3}{16}$。

第 3 步：中間各項查表。 當 $n = 3, 4, 5$ 時（平均值為 12、16、20），直接從累積泊松分佈表讀出 $q_n = \mathrm{P}(\mathrm{Po}(4n) \le 25)$；這類表通常涵蓋至平均值 20。如果表中也列出平均值 24 和 28，則 $n = 6, 7$ 亦可查表。

第 4 步：較大的平均值用正態近似。 對於餘下直至 10 的 n（平均值最大為 40），使用 $\mathrm{Po}(\mu) \approx \mathrm{N}(\mu, \mu)$ 並作連續性校正（continuity correction）：

$$q_n \approx \Phi\left(\frac{25.5 - 4n}{\sqrt{4n}}\right),$$

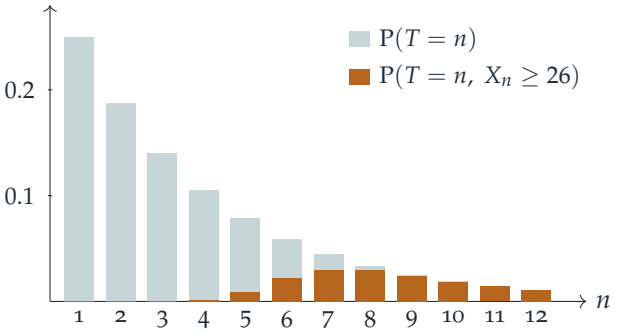
其中 Φ 的值從正態分佈表讀出。這些平均值已經夠大，近似效果良好。更重要的是，權重 w_n 本身已經很小（$w_8 \approx 0.033$，$w_{10} \approx 0.019$），所以 q_n 即使有千分之幾的誤差，對 $\mathrm{P}(F)$ 的影響也小於 0.001。

第 5 步：合併。 用計算機算出 $w_1, \ldots, w_{10}$，並計算

$$\mathrm{P}(F) \approx 1 - \sum_{n=1}^{10} w_n q_n,$$

然後捨入至兩位有效數字。每個被略去或作近似的部分最多約為 0.001，所以除非總和剛好非常接近捨入的分界，否則捨入是穩妥的。萬一真的如此，這個方法也清楚顯示應該改進哪幾項。

題目並沒有要求我們實際計算，但算出來的結果是 $\mathrm{P}(F) \approx 0.19$。下圖顯示這個概率來自何處：大部分來自第一班巴士在 8:06 至 8:10 到站的情況，這時預期的候車人數為 24 至 40，接近或超過載客量。



驗算

- 在 (i) 中，各概率之和為 1，而且 $\mathrm{P}(T = 1) = \frac{1}{4}$，與預期相符。
- 在 (ii) 中，候車人數的平均值為 20，即每分鐘四人、共五分鐘。

- 在 (iii) 中，精確的和（由電腦計算）為 0.1904。以蒙地卡羅模擬（Monte Carlo simulation）模擬 2×10^5 個早上，得 0.189。上述查表加正態近似的方法得 0.1912。三者捨入後都是 0.19。

常見錯誤

- 分界值差了一。只有當候車人數為 26 或以上時，巴士才未能接載所有人，所以我們需要的概率是 $1 - \mathrm{P}(\mathrm{Po}(4n) \leq 25)$，而不是 $1 - \mathrm{P}(\mathrm{Po}(4n) \leq 24)$。
- 把候車人數當作 $\mathrm{Po}(4\,\mathrm{E}[T]) = \mathrm{Po}(16)$。等候時間是隨機的，所以必須把泊松概率對 T 取平均，而不是把平均時間代入。
- 讓和式保持無窮，卻沒有說明為何可以截斷。你必須證明被略去的尾項小得不足以影響第二位有效數字。

STEP 之外

候車人數 X_T 是一個平均值本身隨機的泊松計數，即泊松分佈的一個混合（mixture）。由於 T 服從幾何分佈，X_T 的母函數（generating function）原來有一個簡潔的閉合形式；而像這樣的取條件論證，正是排隊論（queueing theory）的基礎，也是大學第一門概率課程中複合分佈（compound distribution）和混合分佈（mixed distribution）的基礎。怎樣截斷一個和式並估計誤差的上界，這個問題就是數值分析（numerical analysis）的起點。Grimmett 與 Welsh 的 *Probability: An Introduction*（概率論入門教材）以合適的程度講解泊松過程和母函數。

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。