詳細解答

STEP III 2019，第 11 題

概率與統計 • 泊松稀疏化

> **STEP III 2019，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 2019 年 STEP 試卷（zip 檔案）內的 STEP III 試卷：
>
> www.ocr.org.uk/Images/696590-2019-past-papers.zip

關鍵思路

一切都建基於以到店顧客人數或取沙人數作條件（conditioning），然後對一個泊松（Poisson）級數求和。第 (i) 部分是經典的泊松分佈「稀疏化」（thinning）：把二項概率與泊松概率相乘之後，對那個看不見的總數求和，結果會收縮成一個指數函數。一旦知道取沙人數服從泊松分佈，第 (ii) 及 (iii) 部分便只需要求 $(1-k)^Y$ 的期望值（expectation），而那又是另一個指數級數。

解答

設 N 為一天內的顧客人數，故 $N \sim \mathrm{Po}(\lambda)$；設 Y 為取沙的人數。我們假設顧客之間的決定互相獨立（independent），並且與 N 獨立，因此在 $N = n$ 的條件下，有 $Y \sim \mathrm{B}(n, p)$。

第 (i) 部分

對 $y = 0, 1, 2, \ldots$，以 N 作條件（只有 $n \geq y$ 的項有貢獻）：

$$\begin{aligned}
\mathrm{P}(Y = y) &= \sum_{n=y}^{\infty} \mathrm{e}^{-\lambda} \frac{\lambda^n}{n!} \binom{n}{y} p^y (1-p)^{n-y} \\
&= \mathrm{e}^{-\lambda} \frac{(p\lambda)^y}{y!} \sum_{n=y}^{\infty} \frac{(\lambda(1-p))^{n-y}}{(n-y)!} \qquad \text{由於} \frac{\lambda^n}{n!} \cdot \frac{n!}{y!\,(n-y)!} = \frac{\lambda^y \lambda^{n-y}}{y!\,(n-y)!} \\
&= \mathrm{e}^{-\lambda} \frac{(p\lambda)^y}{y!} \mathrm{e}^{\lambda(1-p)} = \mathrm{e}^{-p\lambda} \frac{(p\lambda)^y}{y!},
\end{aligned}$$

其中我們令 $m = n - y$，並用了 $\sum_{m \geq 0} x^m / m! = \mathrm{e}^x$。這正是 $\mathrm{Po}(p\lambda)$ 的概率，所以 $Y \sim \mathrm{Po}(p\lambda)$，得證。

第 (ii) 部分

每位取沙的顧客都會留下原有沙量的 $1-k$ 比例。所以經過 Y 位取沙顧客後，剩下的質量是 $S(1-k)^Y$，而被取走的質量是

$$T = S(1-(1-k)^Y).$$

由第 (i) 部分，

$$\mathrm{E}[(1-k)^Y] = \sum_{y=0}^{\infty}(1-k)^y\,\mathrm{e}^{-p\lambda}\frac{(p\lambda)^y}{y!} = \mathrm{e}^{-p\lambda}\,\mathrm{e}^{p\lambda(1-k)} = \mathrm{e}^{-kp\lambda}.$$

因此，由期望值的線性性質（linearity of expectation），

$$\mathrm{E}(T) = S(1-\mathrm{E}[(1-k)^Y]) = (1-\mathrm{e}^{-kp\lambda})S,$$

這正是所求的結果。

第 (iii) 部分

想一想 怎樣把「取走沙的 k 比例」轉化為關於某一粒特定沙粒的概率？

由於沙粒極多，而且沒有任何一粒被特別挑出，所以每當有人取走剩餘沙的 k 比例時，一粒仍在袋中的沙粒被取走的概率是 k，並且與之前發生的事獨立。（等價地說，金沙粒落在某一份沙中的機會，等於該份沙佔總質量的比例。）

設 G 為店員取得金沙粒這一事件。在 $Y=y$ 的條件下，金沙粒在全部 y 位顧客手下倖存的概率是 $(1-k)^y$，之後店員取得它的概率是 k。所以

$$\mathrm{P}(G \mid Y=y) = k(1-k)^y,$$

再利用第 (ii) 部分求得的期望值，

$$\mathrm{P}(G) = k\,\mathrm{E}[(1-k)^Y] = k\,\mathrm{e}^{-kp\lambda}.$$

這與第 (ii) 部分吻合：金沙粒在當天倖存的概率是 $\mathrm{e}^{-kp\lambda}$，恰好就是剩餘沙量的期望比例。

$k=0$ 的情況。此時 $\mathrm{P}(G)=0$。這是合理的：當 $k=0$ 時，沒有人（包括店員）會取走任何沙，所以金沙粒一直留在袋中。

$k\to 1$ 的極限。此時 $\mathrm{P}(G)\to \mathrm{e}^{-p\lambda}$，即 $\mathrm{P}(Y=0)$。當 k 接近 1 時，第一位取沙的顧客幾乎會把沙全部取走，金沙粒幾乎必然也在其中。店員基本上只有在沒有任何顧客取沙時才能得到金沙粒，而那時店員會取走（幾乎）整袋沙。

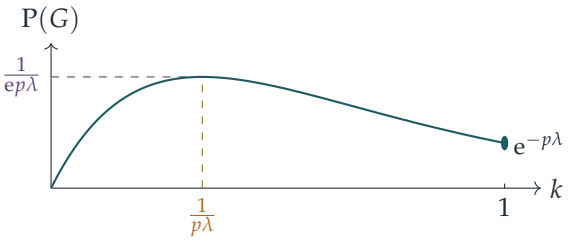
求最大值。記 $\mu = p\lambda$，並對 $0 \le k < 1$ 設 $f(k) = k\,\mathrm{e}^{-\mu k}$。則

$$f'(k) = \mathrm{e}^{-\mu k}(1 - \mu k),$$

它在 $k < 1/\mu$ 時為正，在 $k > 1/\mu$ 時為負。由於 $\mu > 1$，點 $k = 1/\mu$ 位於 $(0,1)$ 內，所以 f 在該點之前遞增、之後遞減，該點給出 $[0,1)$ 上的最大值。因此概率在

$$k = \frac{1}{p\lambda},$$

時取得最大值，而最大概率是 $\dfrac{1}{\mathrm{e}\,p\lambda}$。（若 $p\lambda \le 1$，則 f 在整個 $[0,1)$ 上遞增，沒有最大值，只有當 $k \to 1$ 時的上確界（supremum）$\mathrm{e}^{-p\lambda}$。這就是題目限定 $p\lambda > 1$ 的原因。）



上圖顯示 $p\lambda = 3$ 時的 $\mathrm{P}(G) = k\,\mathrm{e}^{-kp\lambda}$。

驗算

- 第 (i) 部分通過明顯的檢驗：$p = 1$ 時還原為 $\mathrm{Po}(\lambda)$，而 $p = 0$ 時必然有 $Y = 0$。
- 在第 (ii) 部分中，$k = 0$ 給出取走的沙量為零；而當 $\lambda \to \infty$ 時，取走質量的期望值趨向 S，理應如此。
- 取 $\lambda = 3$、$p = 0.6$、$k = 0.4$，公式給出 $\mathrm{P}(G) = 0.4\,\mathrm{e}^{-0.72} \approx 0.1947$。以電腦模擬 200 000 天，得到 0.1953。
- 取 $p\lambda = 3$，在細密網格上 $k\,\mathrm{e}^{-3k}$ 的最大值出現在 $k \approx 0.3333$，其值為 $0.1226 \approx 1/(3\mathrm{e})$。

常見錯誤

- 把取走的沙量寫成 kSY，彷彿每位顧客都取走原來沙量的 k 比例。其實每位顧客取走的是剩下沙量的一部分，所以各次取走的量構成一個等比數列（geometric sequence）。

- 把 $\mathrm{E}[(1-k)^Y]$ 算成 $(1-k)^{\mathrm{E}(Y)} = (1-k)^{p\lambda}$。非線性函數的期望值並不等於期望值的函數。
- 忘記檢查 $1/(p\lambda)$ 是否真的落在容許範圍 $0 \leq k < 1$ 內。條件 $p\lambda > 1$ 正是用在這裏。

STEP 之外

第 (i) 部分就是泊松分佈的稀疏化性質。第 (ii) 及 (iii) 部分中出現的期望值 $\mathrm{E}[s^Y] = \mathrm{e}^{\mu(s-1)}$，是 $\mathrm{Po}(\mu)$ 的*概率母函數*（probability generating function）在 $s = 1-k$ 處的值。利用母函數，稀疏化幾乎一行就能證明：若 $\mathrm{E}[s^N] = \mathrm{e}^{\lambda(s-1)}$，則 $\mathrm{E}[s^Y] = \mathrm{E}[(1-p+ps)^N] = \mathrm{e}^{\lambda p(s-1)}$。在泊松過程（Poisson process）的理論中，這就成為以下命題：對泊松過程的點作獨立標記，所得的仍然是泊松過程。Grimmett 與 Welsh 合著的 *Probability: An Introduction*（《概率導論》）對母函数及這些概念有清晰的論述。

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。