详细解答

STEP III 2019，第 11 题

概率与统计 · 泊松稀疏化

> **STEP III 2019，第 11 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP III 试卷：
>
> www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

一切都建基于以到店顾客人数或取沙人数作条件（conditioning），然后对一个泊松（Poisson）级数求和。第 (i) 部分是经典的泊松分布“稀疏化”（thinning）：把二项概率与泊松概率相乘之后，对那个看不见的总数求和，结果会收缩成一个指数函数。一旦知道取沙人数服从泊松分布，第 (ii) 及 (iii) 部分便只需要求 $(1-k)^Y$ 的期望值（expectation），而那又是另一个指数级数。

解答

设 N 为一天内的顾客人数，故 $N \sim \mathrm{Po}(\lambda)$；设 Y 为取沙的人数。我们假设顾客之间的决定互相独立（independent），并且与 N 独立，因此在 $N = n$ 的条件下，有 $Y \sim \mathrm{B}(n, p)$。

第 (i) 部分

对 $y = 0, 1, 2, \ldots$，以 N 作条件（只有 $n \geq y$ 的项有贡献）：

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

其中我们令 $m = n - y$，并用了 $\sum_{m \geq 0} x^m / m! = \mathrm{e}^x$。这正是 $\mathrm{Po}(p\lambda)$ 的概率，所以 $Y \sim \mathrm{Po}(p\lambda)$，得证。

第 (ii) 部分

每位取沙的顾客都会留下原有沙量的 $1-k$ 比例。所以经过 Y 位取沙顾客后，剩下的质量是 $S(1-k)^Y$，而被取走的质量是

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

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，由期望值的线性性质（linearity of expectation），

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

这正是所求的结果。

第 (iii) 部分

想一想 怎样把“取走沙的 k 比例”转化为关于某一粒特定沙粒的概率?

由于沙粒极多，而且没有任何一粒被特别挑出，所以每当有人取走剩余沙的 k 比例时，一粒仍在袋中的沙粒被取走的概率是 k，并且与之前发生的事独立。（等价地说，金沙粒落在某一份沙中的机会，等于该份沙占总质量的比例。）

设 G 为店员取得金沙粒这一事件。在 $Y=y$ 的条件下，金沙粒在全部 y 位顾客手下幸存的概率是 $(1-k)^y$，之后店员取得它的概率是 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，

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

这与第 (ii) 部分吻合：金沙粒在当天幸存的概率是 $\mathrm{e}^{-kp\lambda}$，恰好就是剩余沙量的期望比例。

*$k=0$ 的情况。*此时 $\mathrm{P}(G)=0$。这是合理的：当 $k=0$ 时，没有人（包括店员）会取走任何沙，所以金沙粒一直留在袋中。

*$k\to 1$ 的极限。*此时 $\mathrm{P}(G)\to \mathrm{e}^{-p\lambda}$，即 $\mathrm{P}(Y=0)$。当 k 接近 1 时，第一位取沙的顾客几乎会把沙全部取走，金沙粒几乎必然也在其中。店员基本上只有在没有任何顾客取沙时才能得到金沙粒，而那时店员会取走（几乎）整袋沙。

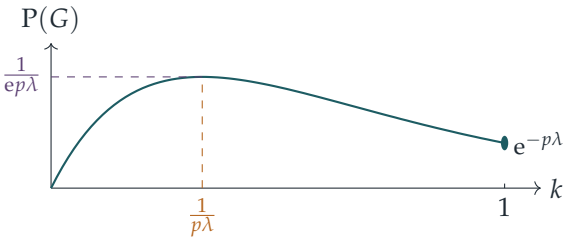
*求最大值。*记 $\mu = p\lambda$，并对 $0\le k<1$ 设 $f(k) = k\,\mathrm{e}^{-\mu k}$。则

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

它在 $k < 1/\mu$ 时为正，在 $k > 1/\mu$ 时为负。由于 $\mu > 1$，点 $k = 1/\mu$ 位于 $(0,1)$ 内，所以 f 在该点之前递增、之后递减，该点给出 $[0,1)$ 上的最大值。因此概率在

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

时取得最大值，而最大概率是 $\dfrac{1}{\mathrm{e}\,p\lambda}$。（若 $p\lambda \le 1$，则 f 在整个 $[0,1)$ 上递增，没有最大值，只有当 $k \to 1$ 时的上确界（supremum）$\mathrm{e}^{-p\lambda}$。这就是题目限定 $p\lambda > 1$ 的原因。）



上图显示 $p\lambda = 3$ 时的 $\mathrm{P}(G) = k\,\mathrm{e}^{-kp\lambda}$。

验算

- 第 (i) 部分通过明显的检验：$p = 1$ 时还原为 $\mathrm{Po}(\lambda)$，而 $p = 0$ 时必然有 $Y = 0$。
- 在第 (ii) 部分中，$k = 0$ 给出取走的沙量为零；而当 $\lambda \to \infty$ 时，取走质量的期望值趋向 S，理应如此。
- 取 $\lambda = 3$、$p = 0.6$、$k = 0.4$，公式给出 $\mathrm{P}(G) = 0.4\,\mathrm{e}^{-0.72} \approx 0.1947$。以电脑模拟 200 000 天，得到 0.1953。
- 取 $p\lambda = 3$，在细密网格上 $k\,\mathrm{e}^{-3k}$ 的最大值出现在 $k \approx 0.3333$，其值为 $0.1226 \approx 1/(3\mathrm{e})$。

常见错误

- 把取走的沙量写成 kSY，仿佛每位顾客都取走*原来*沙量的 k 比例。其实每位顾客取走的是*剩下*沙量的一部分，所以各次取走的量构成一个等比数列（geometric sequence）。
- 把 $\mathrm{E}[(1-k)^Y]$ 算成 $(1-k)^{\mathrm{E}(Y)} = (1-k)^{p\lambda}$。非线性函数的期望值并不等于期望值的函数。
- 忘记检查 $1/(p\lambda)$ 是否真的落在容许范围 $0 \le k < 1$ 内。条件 $p\lambda > 1$ 正是用在这里。

STEP 之外

第 (i) 部分就是泊松分布的稀疏化性质。第 (ii) 及 (iii) 部分中出现的期望值 $\mathrm{E}[s^Y] = \mathrm{e}^{\mu(s-1)}$，是 $\mathrm{Po}(\mu)$ 的概率母函数（probability generating function）在 $s = 1-k$ 处的值。利用母函数，稀疏化几乎一行就能证明：若 $\mathrm{E}[s^N] = \mathrm{e}^{\lambda(s-1)}$，则 $\mathrm{E}[s^Y] = \mathrm{E}[(1-p+ps)^N] = \mathrm{e}^{\lambda p(s-1)}$。在泊松过程（Poisson process）的理论中，这就成为以下命题：对泊松过程的点作独立标记，所得的仍然是泊松过程。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（《概率导论》）对母函数及这些概念有清晰的论述。

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。