详细解答

STEP II 2025，第 12 题

概率与统计 · 泊松分布

STEP II 2025，第 12 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 2025 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/750486-2025-past-papers.zip

关键思路

对于泊松分布（Poisson distribution），相邻两项概率的比有简单的形式 $p_r/p_{r-1} = \lambda/r$，因此每一个差（difference）d_r 都可以写成某个已知概率乘以一个关于 r 的简单有理因子。d_r 的正负号告诉我们 p_r 在哪里上升、在哪里下降；再深入一层，二阶差分（second difference）$d_{r+1} - d_r$ 的正负号告诉我们 d_r 在哪里上升、在哪里下降，而这个正负号由一个关于 r 的二次式（quadratic）决定。那个看似古怪的数 $\lambda + \frac{1}{2} + \sqrt{\lambda + \frac{1}{4}}$，其实只不过是该二次式的一个根（再加一）。

解答

全文中 $p_r = \mathrm{e}^{-\lambda}\lambda^r/r!$，所以当 $r \geq 1$ 时

$$p_r = \frac{\lambda}{r}\,p_{r-1}, \qquad p_{r-1} = \frac{r}{\lambda}\,p_r,$$

而且每个 p_r 都严格为正。

第 (i) 部分

当 $r \geq 1$ 时，

$$d_r = p_r - p_{r-1} = p_{r-1}\left(\frac{\lambda}{r} - 1\right) = \frac{p_{r-1}}{r}\,(\lambda - r).$$

由于 $p_{r-1}/r > 0$，d_r 的正负号与 $\lambda - r$ 的正负号相同；又因为 λ 不是整数，d_r 永不为零。设 $m = \lfloor \lambda \rfloor$，即小于 λ 的最大整数，于是 $\lambda - 1 < m < \lambda$（由于 $\lambda \notin \mathbb{Z}$，两个不等式（inequality）都是严格的）。那么

- 当 $1 \leq r \leq m$ 时，$r < \lambda$，所以 $d_r > 0$，且 $p_0 < p_1 < \cdots < p_m$；

▶ 当 $r \geq m+1$ 时，$r > \lambda$，所以 $d_r < 0$，且 $p_m > p_{m+1} > p_{m+2} > \cdots$。

（若 $\lambda < 1$，则 $m = 0$，第一组为空。）因此对每个 $r \neq m$ 都有 $p_r < p_m$：m 是唯一的最可能取值，而且 $\lambda - 1 < m < \lambda$。

第 (ii) 部分

现在看看 d_r 本身如何变化。当 $r \geq 1$ 时，设 $e_r = d_{r+1} - d_r = p_{r+1} - 2p_r + p_{r-1}$。把所有项都用 p_r 表示，

$$e_r = p_r\left(\frac{\lambda}{r+1} - 2 + \frac{r}{\lambda}\right) = \frac{p_r}{\lambda(r+1)}\Big(\lambda^2 - 2\lambda(r+1) + r(r+1)\Big).$$

前面的因子为正，所以 e_r 的正负号与下式相同：

$$Q(r) = r^2 + (1-2\lambda)r + \lambda^2 - 2\lambda.$$

这个二次式的判别式（discriminant）是 $(1-2\lambda)^2 - 4(\lambda^2 - 2\lambda) = 1 + 4\lambda$，所以它的根是

$$\alpha = \lambda - \tfrac{1}{2} - \sqrt{\lambda + \tfrac{1}{4}}, \qquad \beta = \lambda - \tfrac{1}{2} + \sqrt{\lambda + \tfrac{1}{4}},$$

而 $Q(r) < 0$ 恰好当 $\alpha < r < \beta$。因此

$$d_{r+1} < d_r \iff \alpha < r < \beta, \qquad d_{r+1} > d_r \iff r < \alpha \text{ 或 } r > \beta.$$

注意 $\beta + 1 = \lambda + \frac{1}{2} + \sqrt{\lambda + \frac{1}{4}}$ 不是整数，所以 β 也不是整数，而 $k = \lfloor \beta + 1 \rfloor = \lfloor \beta \rfloor + 1$ 正是满足下式的整数：

$$k < \lambda + \tfrac{1}{2} + \sqrt{\lambda + \tfrac{1}{4}} < k + 1.$$

? **想一想** 既然 $r < \alpha$ 时 d_r 也是递增的，我们到底应该在哪里寻找 d_r 的最小值?

由第 (i) 部分，$r < \lambda$ 时 $d_r > 0$，$r > \lambda$ 时 $d_r < 0$，所以最小值必定在负项之中，即 $r \geq m+1$。此外 $\beta > \lambda$（因为 $\sqrt{\lambda + \frac{1}{4}} > \frac{1}{2}$）而 $\alpha < \lambda$，所以每个 $r \geq m+1$ 都满足 $r > \alpha$；而由 $k > \beta > \lambda$ 可得 $k \geq m+1$。因此在整数 $r \geq m+1$ 上：

▶ 当 $m+1 \leq r \leq k-1$ 时：$r < \beta$，所以 $e_r < 0$，得 $d_{m+1} > d_{m+2} > \cdots > d_k$；

▶ 当 $r \geq k$ 时：$r > \beta$，所以 $e_r > 0$，得 $d_k < d_{k+1} < d_{k+2} < \cdots$。

（当 $k = m+1$ 时，第一组为空。）所以 d_k 严格小于其他每一个负的 d_r，当然也小于所有正项。d_r 的最小值出现在 $r = k$，正如所求。换言之：p_r 在 $r = k-1$ 与 $r = k$ 之间下降得最陡。

第 (iii) 部分

首先, 若 $\lambda < 1$, 则由第 (i) 部分, 每个 d_r 都是负数。但当 $r \to \infty$ 时 $d_r \to 0$ (因为 $p_r \to 0$), 所以对任何 r，总有后面的某一项比 d_r 大：最大值根本无法取得，更不会在 $r = 1$ 取得。因此我们需要 $\lambda > 1$（由于 λ 不是整数，$\lambda = 1$ 已被排除)。

现设 $\lambda > 1$，于是 $d_1 > 0$。正项为 $d_1, \ldots, d_m$，任何最大值都在它们之中。

- 若最大值在 $r = 1$，则 $d_2 < d_1$，即 $e_1 < 0$。
- 反之，若 $e_1 < 0$，则 $\alpha < 1 < \beta$，所以每个满足 $1 \leq r < \beta$ 的整数 r 都有 $e_r < 0$。由于 $m < \lambda < \beta$，得 $d_1 > d_2 > \cdots > d_m$，而其后各项都是负数。所以最大值在 $r = 1$。

因此，当 $\lambda > 1$ 时，最大值在 $r = 1$ 恰好当 $e_1 < 0$，即当 $Q(1) < 0$：

$$Q(1) = 1 + 1 - 2\lambda + \lambda^2 - 2\lambda = \lambda^2 - 4\lambda + 2 < 0 \iff 2 - \sqrt{2} < \lambda < 2 + \sqrt{2}.$$

结合 $\lambda > 1$（以及 $2 - \sqrt{2} < 1$)，条件为

$$1 < \lambda < 2 + \sqrt{2}.$$

第 (iv) 部分

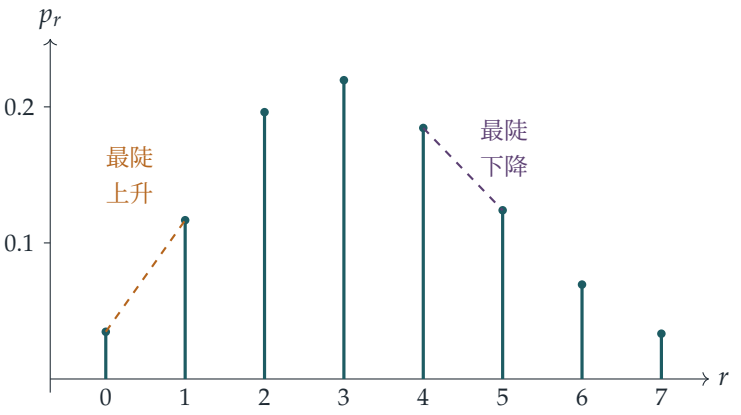
当 $\lambda = 3.36$ 时，$\lambda + \frac{1}{4} = 3.61 = 1.9^2$，所以：

- 由第 (i) 部分，众数（mode）是 $m = 3$;
- $\lambda + \frac{1}{2} + \sqrt{\lambda + \frac{1}{4}} = 3.86 + 1.9 = 5.76$，所以 $k = 5$：最陡的下降是由 p_4 到 p_5;
- $1 < 3.36 < 2 + \sqrt{2} \approx 3.41$，所以最陡的上升是由 p_0 到 p_1（只是刚刚好：$\alpha = 0.96$ 很接近 1，所以 d_1 与 d_2 几乎相等)。

利用 $p_r = \frac{3.36}{r} p_{r-1}$ 及 $p_0 = \mathrm{e}^{-3.36} \approx 0.035$，得出近似值

$$p_0, \ldots, p_7 \approx 0.035,\ 0.117,\ 0.196,\ 0.220,\ 0.185,\ 0.124,\ 0.069,\ 0.033.$$

草图应显示：一开始就急升，然后逐渐平缓，在 $r = 3$ 达到峰值；随后下降越来越陡，在 $r = 4$ 与 $r = 5$ 之间最陡，之后再渐趋平缓。



验算

- 对 20,000 个随机的 $\lambda \in (0, 40)$ 计算 p_r 及 d_r，证实众数是 $\lfloor \lambda \rfloor$，d_r 的最小值在 $\lfloor \lambda + \frac{1}{2} + \sqrt{\lambda + \frac{1}{4}} \rfloor$，而且（当 $\lambda > 1$ 时）d_r 的最大值在 $r = 1$ 恰好当 $\lambda < 2 + \sqrt{2} \approx 3.414$。
- 当 $\lambda = 3.36$：$d_1 \approx 0.0820$ 而 $d_2 \approx 0.0794$，所以 d_1（以些微之差）是最大的；$d_4, d_5, d_6 \approx -0.035, -0.061, -0.055$，所以 d_5 是最小的，与预测一致。
- 小的 λ：当 $\lambda \to 0$ 时，$\lambda + \frac{1}{2} + \sqrt{\lambda + \frac{1}{4}} \to 1$，所以 $k = 1$。事实上，此时几乎所有概率都集中在 0，最大的下降就是第一个，即 $p_1 - p_0$。

常见错误

- 忘记 d_r 在 $r < \alpha$ 时也是递增的，因而没有解释为何最小值不可能在该区域（那是 $d_r > 0$ 的区域）。
- 在第 (iii) 部分只检查 $d_2 < d_1$，而漏掉 $\lambda < 1$ 的情况——此时每个 d_r 都是负数，数列根本取不到最大值。
- 写成 $\lambda - 1 \leq m \leq \lambda$：正正因为 λ 不是整数，这些不等式才是严格的。

STEP 之外

差 d_r 与 e_r 是一阶及二阶导数（derivative）的离散版本，而 k 标示了泊松分布的一个离散“拐点”（point of inflection）。当 λ 很大时，泊松分布近似于平均值为 λ、标准差（standard deviation）为 $\sqrt{\lambda}$ 的正态分布（normal distribution），而后者的拐点位于 $\lambda \pm \sqrt{\lambda}$；可与这里的 $\alpha, \beta \approx \lambda \pm \sqrt{\lambda}$ 比较。利用比值 p_r / p_{r-1} 的技巧是求离散分布众数的标准方法，在 Grimmett 与 Welsh 的 *Probability: An Introduction*（《概率导论》），或 Blitzstein 与

Hwang 的 *Introduction to Probability*（《概率导论》）中有更深入的探讨。

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。