詳細解答

# STEP III 1992，第 16 題

概率與統計 • 泊松稀疏化

> **STEP III 1992，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/92-s3-q16

## 關鍵思路

一旦錯字的總數固定下來，每個錯字都是一次獨立試驗（independent trial），我要麼發現它，要麼看漏，所以第 (i) 部分是一個二項概率（binomial probability）。第 (ii) 部分是條件概率（conditional probability），而分母，即我恰好發現 $m$ 個錯字的概率，需要我們對所有可能的看漏數目逐一相加。這個和原來是一個指數級數（exponential series），正因如此，整個式子才會化簡成簡潔的泊松（Poisson）形式。

## 解答

全篇記 $q = 1 - p$。設 $N$ 為該期報紙的錯字總數，$S$ 為我發現的錯字數目，$U = N - S$ 為我未能發現的數目。題目告訴我們

$$\mathrm{P}(N = k) = \frac{\mathrm{e}^{-\lambda}\lambda^k}{k!}, \qquad k = 0, 1, 2, \ldots$$

### 第 (i) 部分

已知 $N = m + n$ 時，這 $m + n$ 個錯字各自以概率 $p$ 被發現，彼此獨立。因此，在 $N = m + n$ 的條件下，被發現的錯字數目服從二項分佈（binomial distribution）$\mathrm{B}(m + n, p)$。任意指定哪 $m$ 個錯字被發現、哪 $n$ 個被看漏，其概率都是 $p^m q^n$，而這樣的選法共有 $\binom{m+n}{m}$ 種，所以

$$\mathrm{P}(S = m \mid N = m + n) = \binom{m+n}{m} p^m (1-p)^n.$$

### 第 (ii) 部分

我們要求 $\mathrm{P}(U = n \mid S = m)$。事件 $\{S = m,\ U = n\}$ 與 $\{N = m+n,\ S = m\}$ 是同一事件，所以由第 (i) 部分可得

$$\begin{aligned}\mathrm{P}(S = m,\ U = n) &= \mathrm{P}(N = m+n)\,\mathrm{P}(S = m \mid N = m+n)\\ &= \frac{\mathrm{e}^{-\lambda}\lambda^{m+n}}{(m+n)!}\cdot\frac{(m+n)!}{m!\,n!}\,p^m q^n\\ &= \mathrm{e}^{-\lambda}\,\frac{(\lambda p)^m}{m!}\cdot\frac{(\lambda q)^n}{n!}.\end{aligned}$$

因子 $(m+n)!$ 剛好約去，這正是本題得以順利解出的關鍵。

> **想一想** 我們不知道總共有多少個錯字，那麼怎樣求 $\mathrm{P}(S = m)$ 呢？

如果我恰好發現 $m$ 個，看漏的數目可以是 $0, 1, 2, \ldots$ 中的任何一個，而這些情況互斥（mutually exclusive）。因此我們把聯合概率（joint probability）對 $U$ 的所有取值 $j$ 求和：

$$\begin{aligned}\mathrm{P}(S = m) &= \sum_{j=0}^{\infty}\mathrm{e}^{-\lambda}\,\frac{(\lambda p)^m}{m!}\cdot\frac{(\lambda q)^j}{j!} = \mathrm{e}^{-\lambda}\,\frac{(\lambda p)^m}{m!}\,\mathrm{e}^{\lambda q}\\ &= \frac{(\lambda p)^m}{m!}\,\mathrm{e}^{-\lambda p},\end{aligned}$$

這裏用到 $\sum_{j\ge 0} x^j/j! = \mathrm{e}^x$ 及 $-\lambda + \lambda q = -\lambda p$。（由此可見，我發現的錯字數目本身也服從泊松分佈（Poisson distribution），平均值（mean）為 $\lambda p$。）這個概率是正數，所以可以用它作除數。於是

$$\begin{aligned}\mathrm{P}(U = n \mid S = m) &= \frac{\mathrm{P}(S = m,\ U = n)}{\mathrm{P}(S = m)} = \frac{\mathrm{e}^{-\lambda}\,\dfrac{(\lambda p)^m}{m!}\cdot\dfrac{(\lambda q)^n}{n!}}{\mathrm{e}^{-\lambda p}\,\dfrac{(\lambda p)^m}{m!}}\\ &= \frac{(\lambda q)^n}{n!}\,\mathrm{e}^{-\lambda+\lambda p} = \frac{(1-p)^n\lambda^n}{n!}\,\mathrm{e}^{-(1-p)\lambda},\end{aligned}$$

這正是所要證明的結果。

答案完全不依賴 $m$：知道我發現了多少個錯字，並不能告訴我看漏了多少個。無論 $m$ 是多少，看漏的數目都服從平均值為 $(1-p)\lambda$ 的泊松分佈。

## 驗算

- 把最後的答案對 $n = 0, 1, 2, \ldots$ 求和，得 $\mathrm{e}^{(1-p)\lambda}\mathrm{e}^{-(1-p)\lambda} = 1$，條件分佈理應如此。
- 當 $p \to 1$ 時，我會發現所有錯字，而答案在 $n = 0$ 時趨向 1，在 $n \ge 1$ 時趨向 0。吻合。

- 取 $\lambda = 3$、$p = 0.6$、$m = 2$、$n = 1$，公式給出 $1.2\,\mathrm{e}^{-1.2} \approx 0.3614$。模擬 400 000 份報紙，只保留恰好發現兩個錯字的那些，所得結果為 0.360。

## 常見錯誤

- 在第 (i) 部分漏掉二項式係數（binomial coefficient）$\binom{m+n}{m}$：被發現的 $m$ 個錯字可以是 $m+n$ 個之中的任意 $m$ 個。
- 把 $\mathrm{P}(S = m)$ 當成 $\mathrm{P}(N = m)$，或當成第 (i) 部分的答案，而沒有對所有可能的看漏數目求和。

## STEP 之外

這就是泊松分佈的*稀疏化*（thinning，又稱分裂 splitting）性質：如果有 Poisson($\lambda$) 個物件，每個各以概率 $p$ 被保留，那麼被保留的數目與被捨棄的數目是*獨立*的泊松變量，平均值分別為 $\lambda p$ 和 $\lambda(1-p)$。我們求得的聯合概率恰好分解成這兩個部分。稀疏化是泊松過程（Poisson process）理論中的重要工具，它讓你把一串隨機事件分拆成若干串互相獨立的子事件流。Grimmett 與 Welsh 合著的 *Probability: An Introduction*（概率論入門教材）有講述這個性質，也介紹了只需一行便能證明它的母函數（generating function）方法。

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。