詳細解答

# STEP III 樣本試卷，第 16 題

概率與統計 • 負二項分佈

> **STEP III 樣本試卷，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/spec-s3-q16

## 關鍵思路

等到 $r$ 個兒子出生，等同於先等一個兒子出生，然後重新開始等下一個，如此重複 $r$ 次。因此 $C$ 可以拆成 $r$ 個分佈相同的等候時間之和，餘下的工作交給期望值的線性（linearity of expectation）便可。至於第 (ii) 部分，竅門是想像總統停止生育之後，嬰兒仍然繼續出生：這樣一來，他在某個時刻之前是否已經停下，就變成一個關於固定數目的孩子的問題，再利用男孩與女孩之間的對稱性（symmetry）便可解決。

## 解答

以下一律假設每次出生的嬰兒是男孩的概率為 $\frac{1}{2}$，並且與其他各次出生互相獨立（independent）。

### 第 (i) 部分

對於 $i = 1, 2, \ldots, r$，設 $X_i$ 為第 $(i-1)$ 個兒子出生之後，直至第 $i$ 個兒子出生為止（包括這個兒子在內）所生的孩子數目。（當 $i = 1$ 時，由最初開始數起。）於是

$$C = X_1 + X_2 + \cdots + X_r.$$

每個 $X_i$ 所數的，都是在一輪全新的、獨立而公平的出生之中，直至第一個男孩出現為止的出生次數，所以每個 $X_i$ 的分佈（distribution）都與 $X$ 相同，其中

$$\mathrm{P}(X = k) = \left(\tfrac{1}{2}\right)^{k-1} \cdot \tfrac{1}{2} = \left(\tfrac{1}{2}\right)^{k}, \qquad k = 1, 2, 3, \ldots,$$

因為 $X = k$ 表示先有 $k-1$ 個女孩，然後才有一個男孩。

要求出 $\mathrm{E}(X)$，可把幾何級數（geometric series）$\sum_{k\geq 0} x^k = \dfrac{1}{1-x}$（適用於 $|x|<1$）求導，得到 $\sum_{k\geq 1} kx^{k-1} = \dfrac{1}{(1-x)^2}$。乘以 $x$ 並代入 $x=\frac{1}{2}$，得

$$\mathrm{E}(X) = \sum_{k=1}^{\infty} k\left(\tfrac{1}{2}\right)^k = \frac{\frac{1}{2}}{\left(1-\frac{1}{2}\right)^2} = 2.$$

這個結果合乎常理：如果所有嬰兒之中有一半是男孩，那麼平均要等兩次出生才會有一個男孩。

根據期望值的線性（無論各個 $X_i$ 是否獨立，這個性質都成立，不過這裏它們確實是獨立的），

$$\mathrm{E}(C) = \mathrm{E}(X_1) + \cdots + \mathrm{E}(X_r) = 2r.$$

**第 (ii) 部分**

想像有一個無窮無盡的嬰兒序列，各個嬰兒的性別早已預先決定，彼此獨立而且公平；總統按着這個序列生孩子，直至第 $r$ 個兒子出生為止。這樣做完全不會改變 $C$，卻讓我們即使在他提早停下的情況下，仍然可以談論「首 $2r-1$ 個嬰兒」。

**想一想** 怎樣只用序列中的首 $2r-1$ 個嬰兒，來描述事件 $C<2r$？

由於 $C$ 是整數，$C<2r$ 即是 $C\leq 2r-1$，也就是說，第 $r$ 個兒子在第 $2r-1$ 個位置或之前出現。這種情況發生，當且僅當至少有 $r$ 個男孩出現在首 $2r-1$ 個嬰兒之中。所以

$$\mathrm{P}(C<2r) = \mathrm{P}(\text{至少有}r\text{ 個男孩出現在首}2r-1\text{ 個嬰兒之中}).$$

現在設 $B$ 和 $G$ 分別為這 $2r-1$ 個嬰兒之中男孩和女孩的數目，於是 $B+G=2r-1$。因為 $2r-1$ 是奇數，$B$ 與 $G$ 不可能相等，而且兩者之中恰好有一個不小於 $r$：

- 若 $B\geq r$，則 $G=2r-1-B\leq r-1$；
- 若 $B\leq r-1$，則 $G\geq r$。

所以事件 $\{B\geq r\}$ 與 $\{G\geq r\}$ 互斥（mutually exclusive），而且兩者合起來涵蓋了所有可能情況，由此得到

$$\mathrm{P}(B\geq r) + \mathrm{P}(G\geq r) = 1.$$

但每次出生時，男孩和女孩出現的機會均等，所以把「男孩」和「女孩」這兩個標籤對調，並不會改變任何概率，因此 $\mathrm{P}(B\geq r) = \mathrm{P}(G\geq r)$。故此兩者各為 $\frac{1}{2}$，而

$$\mathrm{P}(C<2r) = \tfrac{1}{2},$$

這正是所要證明的。

由此可見，家庭人數的期望值是 $2r$，但家庭人數少於 $2r$ 的情況卻恰好佔一半。家庭人數的分佈是偏斜（skewed）的：略小於 $2r$ 的數值十分常見，而偶爾出現的一長串女孩則把平均值拉高。

## 驗算

- 取 $r = 1$：$\mathrm{E}(C) = 2$，而 $C < 2$ 表示第一個嬰兒是男孩，其概率為 $\frac{1}{2}$。
- 取 $r = 2$：$C < 4$ 表示 $C = 2$ 或 3，概率分別為 $\frac{1}{4}$ 和 $2 \cdot \frac{1}{8} = \frac{1}{4}$，總和為 $\frac{1}{2}$。
- 對 $r = 1, \ldots, 8$ 把 $c\binom{c-1}{r-1}2^{-c}$ 精確地求和，每次都得到 $\mathrm{E}(C) = 2r$ 和 $\mathrm{P}(C < 2r) = \frac{1}{2}$；以 $r = 3$ 和 200 000 個家庭進行模擬，得出平均值為 6.00，以及 $\mathrm{P}(C < 6) \approx 0.499$。

## 常見錯誤

- 以「孩子之中有一半是男孩，所以 $C = 2r$」作為論證。這句話就平均而言是對的，但 $C$ 本身是隨機的；第 (i) 部分需要把它分解成各段等候時間，而不能只靠猜測。
- 在第 (ii) 部分只考慮總統實際所生的孩子。當 $C < 2r - 1$ 時，根本就沒有 $2r - 1$ 個孩子；正是把出生序列延伸至停止點之後，計數的論證才得以成立。
- 忘記了 $2r - 1$ 是奇數。正是這一點排除了男孩與女孩數目相同的情況，並使兩個事件互補（complementary）。

## STEP 之外

$C$ 的分佈就是負二項（negative binomial）分佈：$\mathrm{P}(C = c) = \binom{c-1}{r-1}2^{-c}$，其中 $c \geq r$。把一個隨機變量（random variable）寫成若干個較簡單部分之和，再運用期望值的線性，是大學一年級概率論中最有力的工具之一；而第 (ii) 部分把「等候時間」與「固定試驗次數中的計數」聯繫起來的竅門，正是連結負二項分佈與二項（binomial）分佈的那種對偶性（duality）（在連續時間的情況下，它連結的是伽瑪（gamma）分佈與泊松（Poisson）分佈）。Blitzstein 與 Hwang 合著的 *Introduction to Probability* 以大量例子講解了這兩個想法。

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。