詳細解答

STEP II 1988，第 15 題

概率與統計 • 離散均勻分佈變量之和

> **STEP II 1988，第 15 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/88-s2-q15

關鍵思路

每份試卷的分數在 $\{0, 1, \ldots, m\}$ 上均勻分佈（uniform distribution），所以每一組可能的分數出現的機會均等，而概率（probability）就是符合條件的分數組合數目除以所有組合的總數。兩份試卷的數目只須簡單列舉便可得到；三份試卷的數目，則可先以第三份試卷的分數為條件，再把兩份試卷的數目相加而得（這就是卷積（convolution））。之後利用對稱性（symmetry）（把每份試卷的分數作 $k \mapsto m - k$ 的變換），最高的那個範圍幾乎不費工夫便可得出。

動筆之前先提醒一句：試卷上印出的三份試卷公式有一處印刷錯誤。下文會推導出正確的公式，並指出印刷版本錯在哪裏。

解答

設 X_1, X_2, X_3 為各份試卷的分數。由於各份試卷獨立評分，它們是獨立（independent）的，而且各自在 $\{0, 1, \ldots, m\}$ 上均勻分佈。因此每個有序對 (X_1, X_2) 的概率是 $(m+1)^{-2}$，每個有序三元組的概率是 $(m+1)^{-3}$。記 $N_2(n)$ 和 $N_3(n)$ 分別為分數取自 $\{0, \ldots, m\}$ 而總和為 n 的有序對和有序三元組的數目。

兩份試卷

我們需要 $X_1 = k$ 及 $X_2 = n - k$，而且兩者都在 $[0, m]$ 之內，所以

$$\max(0,\, n - m) \le k \le \min(m,\, n).$$

- 若 $0 \le n \le m$：k 由 0 取到 n，所以 $N_2(n) = n + 1$。

▶ 若 $m < n \le 2m$：k 由 $n-m$ 取到 m，所以 $N_2(n) = m - (n-m) + 1 = 2m - n + 1$。

因此

$$P(X_1 + X_2 = n) = \begin{cases} \dfrac{n+1}{(m+1)^2}, & 0 \le n \le m, \\ \dfrac{2m-n+1}{(m+1)^2}, & m < n \le 2m, \end{cases}$$

這就是題目要求的 $m < n \le 2m$ 的結果，連同 $0 \le n \le m$ 的結果。這個分佈形如「帳篷」：每一步上升一，直到 $n = m$，然後逐步下降。

三份試卷

以第三份試卷的分數為條件：若 $X_3 = k$，則需要 $X_1 + X_2 = n - k$，所以

$$N_3(n) = \sum_{k=0}^{m} N_2(n-k) = \sum_{j=n-m}^{n} N_2(j),$$

其中 $j = n - k$，而當 $j < 0$ 或 $j > 2m$ 時 $N_2(j) = 0$。

情況 $0 \le n \le m$。 只有 $j = 0, 1, \ldots, n$ 有貢獻（負的 j 給出零），而且它們全都落在帳篷上升的一段：

$$N_3(n) = \sum_{j=0}^{n} (j+1) = \frac{(n+1)(n+2)}{2}, \qquad P(X_1 + X_2 + X_3 = n) = \frac{(n+1)(n+2)}{2(m+1)^3}.$$

情況 $m < n \le 2m$。 這時 $1 \le n - m \le m < n \le 2m$，所以 $j = n-m, \ldots, n$ 這個範圍橫跨帳篷的頂點。我們在頂點處把它分成兩段：

$$N_3(n) = \sum_{j=n-m}^{m} (j+1) + \sum_{j=m+1}^{n} (2m - j + 1).$$

第一個和是把由 $n - m + 1$ 到 $m + 1$ 的整數相加：

$$\sum_{j=n-m}^{m} (j+1) = \frac{(m+1)(m+2) - (n-m)(n-m+1)}{2}.$$

在第二個和中，設 $i = 2m - j + 1$，它由 $2m - n + 1$ 取到 m：

$$\sum_{j=m+1}^{n} (2m - j + 1) = \frac{m(m+1) - (2m-n)(2m-n+1)}{2}.$$

把四個乘積展開，

$$\begin{aligned} 2N_3(n) &= (m^2 + 3m + 2) - (n^2 - 2mn + m^2 + n - m) \\ &\quad + (m^2 + m) - (4m^2 - 4mn + n^2 + 2m - n) \\ &= 6mn - 3m^2 - 2n^2 + 3m + 2. \end{aligned}$$

除以 $(m+1)^3$ 個機會均等的三元組，

$$P(X_1+X_2+X_3=n)=\frac{6mn-3m^2-2n^2+3m+2}{2(m+1)^3},\qquad m<n\le 2m.$$

想一想 這並不是題目印出的公式。究竟誰對？

印刷版本以 $-4m^2$ 代替了 $-3m^2$，又以 $(m+1)^2$ 代替了 $(m+1)^3$；兩處都是印刷錯誤。只須看一個例子便可定奪：取 $m=1$ 及 $n=2$，由 $\{0,1\}$ 取出的三個分數，在 8 種機會均等的情形中恰好有 3 種總和為 2，所以概率是 $\frac{3}{8}$。我們的公式給出 $\frac{12-3-8+3+2}{16}=\frac{3}{8}$，而印刷的公式卻給出 $\frac{12-4-8+3+2}{8}=\frac{5}{8}$。無論如何，分母必定是 $(m+1)^3$，因為這正是機會均等的結果的總數。再多驗算一次：當 $n=m$ 時，我們中間範圍的公式給出 $\frac{(m+1)(m+2)}{2(m+1)^3}$，與第一種情況吻合；這是理所當然的，因為兩個和在該處正好相接。

情況 $2m<n\le 3m$。 把每個分數 X_i 換成 $m-X_i$。新的分數同樣在 $\{0,\dots,m\}$ 上均勻分佈，而三個新分數仍然獨立，所以

$$\begin{aligned}P(X_1+X_2+X_3=n)&=P\big((m-X_1)+(m-X_2)+(m-X_3)=n\big)\\&=P(X_1+X_2+X_3=3m-n).\end{aligned}$$

當 $2m<n\le 3m$ 時，有 $0\le 3m-n<m$，所以可把第一種情況中的 n 換成 $3m-n$ 來應用：

$$P(X_1+X_2+X_3=n)=\frac{(3m-n+1)(3m-n+2)}{2(m+1)^3},\qquad 2m<n\le 3m.$$

（同樣的對稱性，以 $2m-n$ 代替 $3m-n$，可把兩份試卷在 $n\le m$ 時的結果變成 $n>m$ 時的結果。）

驗算

- 小例子：當 $m=1$ 時，三條公式對 $n=0,1,2,3$ 分別給出 $\frac{1}{8},\frac{3}{8},\frac{3}{8},\frac{1}{8}$，它們的總和是 1，並且與擲三次公平硬幣的二項分佈（binomial distribution）一致。
- 中間範圍的公式在 $n\mapsto 3m-n$ 之下對稱：在 $6mn-3m^2-2n^2$ 中把 n 換成 $3m-n$，式子保持不變，正如對稱性論證所要求的。
- 對由 1 到 8 的每一個 m 作完全列舉，五條公式全部與實際數目完全相符；例如 $m=10$ 及 $n=15$ 時，兩種方法都給出 $\frac{91}{1331}$。

常見錯誤

- 忘記 $N_2(j)$ 有兩條不同的公式，在中間範圍的求和中由頭到尾都用 $j+1$。

► 一心想湊出印刷的中間公式，於是「硬砌」代數運算。如果仔細計算的結果與題目給出的答案不符，先用一個極小的例子檢驗，不要急於懷疑自己。

STEP 之外

獨立隨機變量（random variable）之和的分佈，就是它們各自分佈的*卷積*，而我們的和 $\sum_k N_2(n-k)$ 正是如此。概率母函數（probability generating function）可把這個過程變成機械式的運算：數目 $N_3(n)$ 就是 $(1+t+\cdots+t^m)^3$ 中 t^n 的係數。隨着試卷的數目愈來愈多，分佈的形狀會由帳篷變成平滑的隆起，並趨近正態曲線（這就是中心極限定理（central limit theorem））。Grimmett 與 Welsh 合著的 *Probability: An Introduction*（概率論入門）以合適的程度講解母函數和中心極限定理。

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。