詳細解答

STEP II 1989，第 16 題

概率與統計 • 二項分佈

STEP II 1989，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/89-s2-q16

關鍵思路

把整批產品分成表現不同的兩部分：樣本中的 n 件小配件，以及從未被抽查的 N 件。每一部分的次品數目都服從二項分佈（binomial distribution），而且兩個數目互相獨立（independent）。一切取決於樣本：如果樣本中有兩件或以上次品，整批最終會完美無瑕，否則一切維持不變。因此，若一批產品最後有 $k \geq 1$ 件次品，樣本中必定只有 0 或 1 件次品，其餘的次品則在另外 N 件之中。

解答

記 $q = 1 - p$。設 X 為樣本中次品的數目，Y 為該批其餘 N 件中次品的數目。每件小配件各自獨立地以概率（probability）p 為次品，所以

$$X \sim \mathrm{B}(n, p), \qquad Y \sim \mathrm{B}(N, p),$$

而且 X 與 Y 獨立，因為它們所數算的是兩組互不相交的小配件中的次品。

設 K 為經過品質管制後該批產品中次品的數目。若 $X \geq 2$，所有次品都會被修正，於是 $K = 0$。若 $X \leq 1$，則甚麼都不會修正，於是 $K = X + Y$。所以

$$K = \begin{cases} X + Y, & X \leq 1, \\ 0, & X \geq 2. \end{cases}$$

K 的分佈

取 $1 \le k \le N$。由於 $k \ge 1$，事件 $K = k$ 不可能來自 $X \ge 2$ 的情況，所以它恰好在 $X = 0,\ Y = k$ 或 $X = 1,\ Y = k-1$ 時發生。這兩個事件互斥，而由獨立性可得

$$\begin{aligned}\mathrm{P}(K=k) &= \mathrm{P}(X=0)\,\mathrm{P}(Y=k) + \mathrm{P}(X=1)\,\mathrm{P}(Y=k-1)\\ &= q^n\binom{N}{k}p^k q^{N-k} + npq^{n-1}\binom{N}{k-1}p^{k-1}q^{N-k+1}\\ &= \left[\binom{N}{k} + n\binom{N}{k-1}\right]p^k q^{N+n-k}.\end{aligned}$$

兩項中 p 的冪次相同，q 的冪次也相同，所以才能合併。現在把二項式係數（binomial coefficient）通分，公分母為 $k!\,(N-k+1)!$：

$$\binom{N}{k} = \frac{N!\,(N-k+1)}{k!\,(N-k+1)!}, \qquad n\binom{N}{k-1} = \frac{n\,N!\,k}{k!\,(N-k+1)!}.$$

相加後，分子為 $N!\,(N-k+1+nk) = N!\,[N+1+k(n-1)]$，所以

$$\mathrm{P}(K=k) = \frac{[N+1+k(n-1)]\,N!}{(N-k+1)!\,k!}\,p^k(1-p)^{N+n-k},$$

正如所求。

> **想一想** 當 $k = 0$ 時會怎樣？

這條公式只在 $k \ge 1$ 時成立。當 $k = 0$ 時，它給出 $\frac{(N+1)N!}{(N+1)!}q^{N+n} = q^{N+n}$，這是完全沒有次品的概率。但只要 $X \ge 2$，$K = 0$ 同樣會發生，所以實際上

$$\mathrm{P}(K=0) = q^{N+n} + \mathrm{P}(X \ge 2) = q^{N+n} + 1 - q^n - npq^{n-1}.$$

題目中的「$k \le N$」應理解為 $1 \le k \le N$。

$k = N+1$ 的情況

要使 $K = N+1$，需要 $X \le 1$ 且 $X + Y = N+1$。由於 $Y \le N$，這就迫使 $X = 1$ 且 $Y = N$，所以

$$\mathrm{P}(K=N+1) = npq^{n-1}\cdot p^N = np^{N+1}q^{n-1}.$$

把 $k = N+1$ 代入公式，得到係數

$$\frac{[N+1+(N+1)(n-1)]\,N!}{0!\,(N+1)!} = \frac{(N+1)\,n\,N!}{(N+1)!} = n,$$

而 q 的冪為 $q^{N+n-(N+1)} = q^{n-1}$。所以公式給出 $np^{N+1}q^{n-1}$，結果正確。（當 $k \ge N+2$ 時概率為零，因為 $K \le N+1$。）

次品數目的期望值

我們可以對 $k\,\mathrm{P}(K=k)$ 求和，但利用以 X 和 Y 表示 K 的寫法會快得多。以 $\mathbf{1}_{\{X\le 1\}}$ 表示事件 $X\le 1$ 的指示變量（indicator），則

$$K=(X+Y)\,\mathbf{1}_{\{X\le 1\}}=X\,\mathbf{1}_{\{X\le 1\}}+Y\,\mathbf{1}_{\{X\le 1\}}.$$

對於第一項，$X\,\mathbf{1}_{\{X\le 1\}}$ 在 $X=1$ 時等於 1，其餘情況等於 0，所以它的期望值（expectation）是 $\mathrm{P}(X=1)=npq^{n-1}$。對於第二項，$Y$ 與 X 獨立，所以

$$\mathrm{E}[Y\,\mathbf{1}_{\{X\le 1\}}]=\mathrm{E}[Y]\,\mathrm{P}(X\le 1)=Np\left(q^n+npq^{n-1}\right).$$

相加，並提出因子 pq^{n-1}，

$$\begin{aligned}\mathrm{E}[K]&=npq^{n-1}+Np\,q^{n-1}(q+np)\\&=pq^{n-1}\left[n+N(1-p)+Nnp\right]\\&=\left[N+n+pN(n-1)\right]p(1-p)^{n-1},\end{aligned}$$

正如所求。

驗算

- 各概率之和應為 1。取 $n=2$、$N=1$：$\mathrm{P}(K=1)=(2+1)pq^2=3pq^2$，$\mathrm{P}(K=2)=2p^2q$，以及 $\mathrm{P}(K=0)=q^3+p^2$，三者相加得 $q^3+3pq^2+3p^2q+p^3=1$（用到 $p^2=p^2(p+q)$）。
- 若 n 極大，樣本中有次品的批次幾乎全部都會被修正，期望值趨向 0，這與因子 $(1-p)^{n-1}$ 吻合。
- 我們以精確分數驗算了每一條公式，範圍是 $2\le n\le 6$、$0\le N\le 6$ 以及若干個 p 值，並且模擬了整個程序：取 $n=4$、$N=6$、$p=0.3$ 時，公式給出 $\mathrm{E}[K]\approx 1.585$，而模擬 200 000 批所得的結果是 1.581。

常見錯誤

- 把「兩件或以上」的規則套用在整批的次品總數上。次品會否被修正，只由樣本中的數目 X 決定。
- 忘記 $K=0$ 還有來自已修正批次的額外貢獻，所以公式在 $k=0$ 時不成立。
- 以硬算求和的方法計算 $\mathrm{E}[K]$。這樣做是可行的，但用指示變量把 K 拆開，再利用獨立性，會簡短得多。

STEP 之外

把一個隨機變量（random variable）寫成若干部分各自乘以指示變量之和，再運用期望值的線性性質（linearity of expectation），是大學一年級概率論中最有用的技巧之一。當只需要平均值時，這個方法可以免卻求出整個分佈的工夫。這道題目也是品質管制中驗收抽樣（acceptance sampling）的一個小型模型，當中我們要選擇 n，在檢測成本與漏網的次品之間取得平衡。想進一步了解，可參閱 Grimmett 與 Welsh 的 *Probability: An Introduction*（概率論入門），或 Blitzstein 與 Hwang 的 *Introduction to Probability*（概率論導論），後者全書貫徹使用指示變量。

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。