详细解答

STEP II 1989，第 16 题

概率与统计 · 二项分布

> **STEP II 1989，第 16 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/89-s2-q16

关键思路

把整批产品分成表现不同的两部分：样本中的 n 件小配件，以及从未被抽查的 N 件。每一部分的次品数目都服从二项分布（binomial distribution），而且两个数目互相独立（independent）。一切取决于样本：如果样本中有两件或以上次品，整批最终会完美无瑕，否则一切维持不变。因此，若一批产品最后有 $k \geq 1$ 件次品，样本中必定只有 0 或 1 件次品，其余的次品则在另外 N 件之中。

解答

记 $q = 1 - p$。设 X 为样本中次品的数目，Y 为该批其余 N 件中次品的数目。每件小配件各自独立地以概率（probability）p 为次品，所以

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

而且 X 与 Y 独立，因为它们所数算的是两组互不相交的小配件中的次品。

设 K 为经过品质管制后该批产品中次品的数目。若 $X \geq 2$，所有次品都会被修正，于是 $K = 0$。若 $X \leq 1$，则甚么都不会修正，于是 $K = X + Y$。所以

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

K 的分布

取 $1 \le k \le N$。由于 $k \ge 1$，事件 $K = k$ 不可能来自 $X \ge 2$ 的情况，所以它恰好在 $X = 0,\ Y = k$ 或 $X = 1,\ Y = k-1$ 时发生。这两个事件互斥，而由独立性可得

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

两项中 p 的幂次相同，q 的幂次也相同，所以才能合并。现在把二项式系数（binomial coefficient）通分，公分母为 $k!\,(N-k+1)!$：

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

相加后，分子为 $N!\,(N-k+1+nk) = N!\,[N+1+k(n-1)]$，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** 当 $k = 0$ 时会怎样?

这条公式只在 $k \ge 1$ 时成立。当 $k = 0$ 时，它给出 $\frac{(N+1)N!}{(N+1)!}q^{N+n} = q^{N+n}$，这是完全没有次品的概率。但只要 $X \ge 2$，$K = 0$ 同样会发生，所以实际上

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

题目中的“$k \le N$”应理解为 $1 \le k \le N$。

$k = N+1$ 的情况

要使 $K = N+1$，需要 $X \le 1$ 且 $X + Y = N+1$。由于 $Y \le N$，这就迫使 $X = 1$ 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，得到系数

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

而 q 的幂为 $q^{N+n-(N+1)} = q^{n-1}$。所以公式给出 $np^{N+1}q^{n-1}$，结果正确。（当 $k \ge N+2$ 时概率为零，因为 $K \le N+1$。）

次品数目的期望值

我们可以对 $k\,\mathrm{P}(K=k)$ 求和，但利用以 X 和 Y 表示 K 的写法会快得多。以 $\mathbf{1}_{\{X\le 1\}}$ 表示事件 $X\le 1$ 的指示变量（indicator），则

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

对于第一项，$X\,\mathbf{1}_{\{X\le 1\}}$ 在 $X=1$ 时等于 1，其余情况等于 0，所以它的期望值（expectation）是 $\mathrm{P}(X=1)=npq^{n-1}$。对于第二项，$Y$ 与 X 独立，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，并提出因子 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。

验算

- 各概率之和应为 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 极大，样本中有次品的批次几乎全部都会被修正，期望值趋向 0，这与因子 $(1-p)^{n-1}$ 吻合。
- 我们以精确分数验算了每一条公式，范围是 $2\le n\le 6$、$0\le N\le 6$ 以及若干个 p 值，并且模拟了整个程序：取 $n=4$、$N=6$、$p=0.3$ 时，公式给出 $\mathrm{E}[K]\approx 1.585$，而模拟 200 000 批所得的结果是 1.581。

常见错误

- 把“两件或以上”的规则套用在整批的次品总数上。次品会否被修正，只由样本中的数目 X 决定。
- 忘记 $K=0$ 还有来自已修正批次的额外贡献，所以公式在 $k=0$ 时不成立。
- 以硬算求和的方法计算 $\mathrm{E}[K]$。这样做是可行的，但用指示变量把 K 拆开，再利用独立性，会简短得多。

STEP 之外

把一个随机变量（random variable）写成若干部分各自乘以指示变量之和，再运用期望值的线性性质（linearity of expectation），是大学一年级概率论中最有用的技巧之一。当只需要平均值时，这个方法可以免却求出整个分布的工夫。这道题目也是品质管制中验收抽样（acceptance sampling）的一个小型模型，当中我们要选择 n，在检测成本与漏网的次品之间取得平衡。想进一步了解，可参阅 Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门），或 Blitzstein 与 Hwang 的 *Introduction to Probability*（概率论导论），后者全书贯彻使用指示变量。

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。