解答例

STEP II 1989、第 16 問

確率・統計 ・ 二項分布

> **STEP II 1989、問 16**
>
> 先に進む前に、まず問題を読んでおこう。問題は STEP Support Programme の問題データベースにある：
>
> step.maths.org/questions/89-s2-q16

鍵となる考え方

バッチを、振る舞いの異なる二つの部分に分ける。標本に含まれる n 個のウィジェットと、一度も抜き取られない N 個のウィジェットである。それぞれの部分に含まれる不良品の個数は二項分布（binomial distribution）に従い、二つの個数は互いに独立（independent）である。すべては標本で決まる。標本に不良品が 2 個以上あればバッチは最終的に完全なものになり、そうでなければ何も変わらない。したがって、最終的に $k \geq 1$ 個の不良品を含むバッチでは、標本中の不良品は 0 個または 1 個であり、残りは他の N 個の中にあったはずである。

解答

$q = 1 - p$ とおく。標本に含まれる不良ウィジェットの個数を X、バッチ内の他の N 個に含まれる不良ウィジェットの個数を Y とする。各ウィジェットは独立に確率 p で不良品であるから、

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

であり、X と Y は互いに素（disjoint）なウィジェットの集合における不良品を数えているので、独立である。

品質管理（quality control）を経た後のバッチに含まれる不良ウィジェットの個数を K とする。$X \geq 2$ ならばすべての不良が修正され、$K = 0$ となる。$X \leq 1$ ならば何も修正されず、$K = X + Y$ となる。よって

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

K の分布

$1 \le k \le N$ とする。$k \ge 1$ であるから、事象 $K = k$ は $X \ge 2$ の場合からは生じ得ず、ちょうど $X = 0,\ Y = k$ または $X = 1,\ Y = k-1$ のときに起こる。これらは互いに排反であり、独立性より

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

両方の項で p の累乗も q の累乗も同じであり、だからこそ一つにまとめられる。次に、二項係数（binomial coefficient）を共通の分母 $k!\,(N-k+1)!$ で表す。

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

足し合わせると分子は $N!\,(N-k+1+nk) = N!\,[N+1+k(n-1)]$ となるので、

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

となり、示すべき式が得られた。

考えてみよう $k = 0$ のときはどうなるか？

この公式は $k \ge 1$ のときにしか成り立たない。$k = 0$ のとき公式は $\frac{(N+1)N!}{(N+1)!}q^{N+n} = q^{N+n}$ を与えるが、これは不良品がまったくなかった確率である。しかし $K = 0$ は $X \ge 2$ のときにも必ず起こるので、実際には

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

である。問題文の「$k \le N$」は $1 \le k \le N$ と読むべきである。

$k = N+1$ の場合

$K = N+1$ となるには、$X \le 1$ かつ $X + Y = N+1$ が必要である。$Y \le N$ であるから、これより $X = 1$ かつ $Y = N$ でなければならず、

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

公式に $k = N+1$ を代入すると、係数は

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

となり、q の累乗は $q^{N+n-(N+1)} = q^{n-1}$ である。よって公式は $np^{N+1}q^{n-1}$ を与え、これは正しい。（$K \le N+1$ であるから、$k \ge N+2$ に対する確率は 0 である。）

不良品の個数の期待値

$k\,\mathrm{P}(K=k)$ を総和してもよいが、X と Y を用いた K の表し方を使うほうがはるかに速い。事象 $X \le 1$ の指示関数（indicator function）を $\mathbf{1}_{\{X\le 1\}}$ と書くと、

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

第 1 項について、$X\mathbf{1}_{\{X\le 1\}}$ は $X=1$ のとき 1、それ以外のとき 0 に等しいので、その期待値（expectation）は $\mathrm{P}(X=1) = npq^{n-1}$ である。第 2 項について、Y は X と独立であるから

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

これらを足し合わせ、因数 pq^{n-1} をくくり出すと、

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

となり、示すべき式が得られた。

検算

- 確率の総和は 1 になるはずである。$n=2$、$N=1$ とすると、$\mathrm{P}(K=1) = (2+1)pq^2 = 3pq^2$、$\mathrm{P}(K=2) = 2p^2q$、$\mathrm{P}(K=0) = q^3 + p^2$ であり、これらを足すと（$p^2 = p^2(p+q)$ を用いて）$q^3 + 3pq^2 + 3p^2q + p^3 = 1$ となる。
- n が非常に大きければ、標本に不良品を含むバッチはほぼすべて修正され、期待値は 0 に近づく。これは因数 $(1-p)^{n-1}$ と整合する。
- $2 \le n \le 6$、$0 \le N \le 6$ およびいくつかの p の値について、すべての公式を厳密な分数で確かめ、さらにこの手順をシミュレーションした。$n=4$、$N=6$、$p=0.3$ のとき公式は $\mathrm{E}[K] \approx 1.585$ を与え、200 000 個のバッチのシミュレーションでは 1.581 となった。

よくある間違い

- 「2 個以上」という規則をバッチ全体の不良品の個数に適用してしまう。不良が修正されるかどうかを決めるのは、標本中の個数 X だけである。
- 修正されたバッチから $K=0$ への追加の寄与があることを忘れる。そのため公式は $k=0$ では成り立たない。

- E[K] を力ずくの総和で計算する。それでも求まるが、指示関数を使って K を分解し、独立性を用いるほうがはるかに短い。

STEP のその先へ

確率変数を、指示関数を掛けた部分の和として表し、期待値の線形性（linearity of expectation）を用いる方法は、大学 1 年の確率論で最も役に立つ技法の一つである。平均だけが必要なときに、分布を求めずに済む。また、この問題は品質管理における抜き取り検査（acceptance sampling）の小さなモデルにもなっている。そこでは検査の費用と見逃される不良品とのバランスをとるように n を選ぶ。さらに学ぶには、Grimmett and Welsh, *Probability: An Introduction*（確率論入門）や、全編を通じて指示関数を活用している Blitzstein and Hwang, *Introduction to Probability*（確率入門）を参照するとよい。

Beyond STEP 解答集：https://beyondstep.keithmcnulty.org/ja/。解答本文は CC0 1.0 である。問題そのものは © Cambridge University Press & Assessment に帰属し、ここには掲載していない。英語版から機械翻訳したものであり、まだ校閲を経ていない。