解答例

STEP I 1998、第 13 問

確率・統計 ・ 復元抽出

> **STEP I 1998、問 13**
>
> 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/98-s1-q13

鍵となる考え方

これは条件付き確率（conditional probability）の問題である。条件となる事象は「n 回の取り出しでちょうど 1 個のドロップ（pastille）を食べた」であり、これをそのドロップを**いつ**食べたかによって分割する。袋の中身が変わるのは赤を食べた瞬間だけなので、分割した各部分は単純な確率の積になる。それらを足し合わせると等比級数（geometric series）となり、そこから 2 つの n 乗の差が現れるのである。

解答

最初のドロップの個数を $N = r + b$ とおく。$r \geq 1$ かつ $b \geq 1$ と仮定する（$b = 0$ かつ $n \geq 2$ ならば、ちょうど 1 個だけ食べることは不可能であり、条件とする事象が存在しない）。

各回の取り出しにおける袋の状態。 赤でないドロップは袋に戻されるので、袋が変化するのは赤を食べたときだけである。赤を食べる前、袋には N 個のドロップが入っており、そのうち b 個が赤でない。食べた後は $N-1$ 個が入っており、赤でないものは依然として b 個である。毎回取り出す前に袋をよく振るので、各回の取り出しはその時点で袋に入っているものからの新たな一様（uniform）な選択であり、したがって 1 回ごとの確率を掛け合わせることができる。

食べた時点による分割。 $k = 1, 2, \ldots, n$ に対し、n 回の取り出しで食べた唯一のドロップが k 回目に食べられたという事象を E_k とする。このとき、$1, \ldots, k-1$ 回目は満杯の袋から赤でないものを取り出し、k 回目は満杯の袋から赤を取り出し、$k+1, \ldots, n$ 回

目は 1 個減った袋から赤でないものを取り出す：

$$\mathrm{P}(E_k) = \left(\frac{b}{N}\right)^{k-1} \frac{r}{N} \left(\frac{b}{N-1}\right)^{n-k} = \frac{r\,b^{n-1}}{N^k\,(N-1)^{n-k}}.$$

事象 $E_1, \ldots, E_n$ は互いに排反（disjoint）であり、それらの和事象（union）はちょうど 1 個のドロップを食べたという事象 E である。よって

$$\mathrm{P}(E) = r\,b^{n-1} \sum_{k=1}^{n} \frac{1}{N^k (N-1)^{n-k}} = \frac{r\,b^{n-1}}{N^n (N-1)^n} \sum_{k=1}^{n} N^{n-k} (N-1)^k,$$

ここで各項の分子と分母に $N^n(N-1)^n$ を掛けた。

? **考えてみよう** $\sum_{k=1}^{n} N^{n-k}(N-1)^k$ はどのように計算すればよいか。

これは初項 $N^{n-1}(N-1)$、公比 $\frac{N-1}{N}$ の等比級数である。さらに手早く済ませるには、おなじみの因数分解（factorisation）

$$x^n - y^n = (x-y)(x^{n-1} + x^{n-2}y + \cdots + y^{n-1})$$

において $x = N$、$y = N-1$ とすれば $x - y = 1$ となるので

$$\sum_{k=1}^{n} N^{n-k}(N-1)^k = (N-1) \sum_{j=0}^{n-1} N^{n-1-j}(N-1)^j = (N-1)(N^n - (N-1)^n).$$

したがって

$$\mathrm{P}(E) = \frac{r\,b^{n-1}(N^n - (N-1)^n)}{N^n (N-1)^{n-1}}.$$

条件付き確率。 「最後の回に食べ、しかも 1 個だけ食べた」という事象は E_n であり、$\mathrm{P}(E_n) = \dfrac{r\,b^{n-1}}{N^n}$ である。$E_n \subseteq E$ であるから

$$\begin{aligned}\mathrm{P}(E_n \mid E) = \frac{\mathrm{P}(E_n)}{\mathrm{P}(E)} &= \frac{r\,b^{n-1}}{N^n} \cdot \frac{N^n (N-1)^{n-1}}{r\,b^{n-1}(N^n - (N-1)^n)} \\ &= \frac{(N-1)^{n-1}}{N^n - (N-1)^n} = \frac{(r+b-1)^{n-1}}{(r+b)^n - (r+b-1)^n},\end{aligned}$$

となり、求める結果が得られた。

r と b は、その合計を除いて完全に消えることに注意しよう。実は、この計算からはさらに多くのことがわかる。E が与えられたとき、ドロップが k 回目に食べられた確率は $(\frac{N-1}{N})^k$ に比例するので、早い回ほど遅い回より**起こりやすい**のである。これは納得できる。早く食べれば、赤でないものの割合が大きい袋が残り、その後「もう赤が出ない」という状態が続く確率が高くなるからである。

検算

- $n = 1$：公式は $\dfrac{1}{N-(N-1)} = 1$ を与える。取り出しは 1 回だけでそれが最後の回なので、正しい。
- $n = 2$：公式は $\dfrac{N-1}{2N-1}$ を与え、これは $\frac{1}{2}$ よりわずかに小さい。早い回のほうが少し起こりやすいことと整合している。
- $r = 3$、$b = 5$、$n = 4$ のとき、公式は $343/1695 \approx 0.2024$ を与える。4 回の取り出しがすべて同様に確からしいとした場合の値は $\frac{1}{4}$ である。4×10^5 回のシミュレーションでは 0.2028 となった。また、袋の状態を追跡しながらすべての取り出し列を厳密に数え上げたところ、$1 \le r, b \le 4$ および $1 \le n \le 8$ のすべてについて公式が正確に再現された。

よくある間違い

- 赤を食べた後の赤でないものの取り出しに b/N を用いてしまう。袋の中のドロップは $N-1$ 個になっており、これこそがこの問題の要点である。全体を通して b/N を用いると、どの E_k も同じ確率になり、答えは $1/n$ になってしまう。
- 条件付けを忘れる：$\mathrm{P}(E_n) = rb^{n-1}/N^n$ だけでは答えにならない。ちょうど 1 個のドロップを食べたことが与えられているからである。

STEP のその先へ

ここでの袋は単純なマルコフ連鎖（Markov chain）である。その状態は残っている赤の個数であり、状態が 1 つ下がる確率は現在の状態だけに依存する。このような連鎖の最終状態を条件として、それがたどった経路について問うという考え方は、「隠れマルコフモデル（hidden Markov models）」や、過去の履歴についてのベイズ推論（Bayesian inference）の根底にある考え方である。Grimmett と Welsh の *Probability: An Introduction*（確率論入門）は、条件付き確率とマルコフ連鎖をちょうどこの水準で扱っており、Blitzstein と Hwang の *Introduction to Probability*（確率入門）には同じ趣旨の問題が数多く収録されている。

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。