解答例

STEP I 2006、第 14 問

確率・統計 ・ 非復元抽出

> **STEP I 2006、問 14**
>
> 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/06-s1-q14

鍵となる考え方

赤い菓子を r 回目に取り出すということは、青い菓子を $r-1$ 回続けて取り出した後に赤い菓子を取り出すということである。復元抽出（with replacement）では毎回の取り出しが同じ条件で行われるので、確率は累乗の単純な積になる。そのうえで n を連続変数とみなし、対数を微分して n について最大化する。非復元抽出（without replacement）の場合は、菓子を**すべて**取り出すと想像するのがもっとも見通しがよい。そうすると赤い菓子はどの位置にも等しい確率で現れる。

解答

以下を通して $n \geq 0$ は整数とし、R を赤い菓子が初めて取り出される回の番号とする。

小問 (i)

取り出した菓子は袋に戻され、袋はよく振られるので、各回の取り出しは互いに独立（independent）であり、どの回でも赤い菓子が取り出される確率は $\frac{1}{n+1}$、青い菓子が取り出される確率は $\frac{n}{n+1}$ である。事象 $R=r$ は、最初の $r-1$ 回が青で r 回目が赤であることを意味するので、独立性より

$$P(R=r) = \left(\frac{n}{n+1}\right)^{r-1} \frac{1}{n+1} = \frac{n^{r-1}}{(n+1)^r}, \qquad r = 1, 2, 3, \ldots$$

となる（これは成功確率 $\frac{1}{n+1}$ の幾何分布（geometric distribution）である）。

ここで r を固定し、これを n の関数（function）とみなす：

$$f(n) = \frac{n^{r-1}}{(n+1)^r}.$$

$r=1$ ならば $f(n)=\frac{1}{n+1}$ であり、これは $n=0$ で最大となる（青い菓子がなければ、最初の菓子は必ず赤である）。そこで $r\geq 2$ とする。このとき $f(0)=0$ であり、$n>0$ では $f(n)>0$ であるから、n を実数変数とみなして $n>0$ の範囲で対数（logarithm）をとることができる：

$$\ln f(n)=(r-1)\ln n-r\ln(n+1),\qquad \frac{\mathrm{d}}{\mathrm{d}n}\ln f(n)=\frac{r-1}{n}-\frac{r}{n+1}=\frac{(r-1)-n}{n(n+1)}.$$

考えてみよう n は整数でなければならない。微分してよいのだろうか？

よい。導関数（derivative）は f が増加・減少する範囲を調べるためだけに使うからである。$n>0$ では分母 $n(n+1)$ は正なので、導関数は $0<n<r-1$ で正、$n>r-1$ で負である。したがって f は $[0,r-1]$ で狭義単調増加、$[r-1,\infty)$ で狭義単調減少である。$r-1$ 自体が整数なので、これは他のどの整数よりも大きな値を与え、隣り合う整数どうしを比較する必要はない。

よって、どの場合も（$r=1$ を含めて）確率を最大にするのは

$$n=r-1,$$

であり、最大値は $\frac{(r-1)^{r-1}}{r^r}$ である（$r=1$ のときは 1 と読む）。たとえば $r=3$ のとき、最良の袋は青い菓子が $n=2$ 個のもので、確率は $\frac{4}{27}$ となる。

小問 (ii)

今度は各回の取り出しは独立ではないが、簡単な方法が二つある。

直接積をとる方法。 $R=r$ となるには、最初の $r-1$ 回それぞれで青を取り出し、その後に赤を取り出す必要がある。これは青い菓子が十分にあるとき、すなわち $r-1\leq n$ のときにのみ可能である。それまでの取り出しがすべて青であったという条件のもとで、k 回目の取り出しは $n+2-k$ 個の菓子（そのうち 1 個が赤）の入った袋から行われる。よって

$$P(R=r)=\frac{n}{n+1}\cdot\frac{n-1}{n}\cdot\frac{n-2}{n-1}\cdots\frac{n-r+2}{n-r+3}\cdot\frac{1}{n-r+2}.$$

各分子は次の分母と約分され、積は次のように打ち消し合って簡単になる：

$$P(R=r)=\begin{cases}\dfrac{1}{n+1}, & 1\leq r\leq n+1,\\[2ex] 0, & r>n+1.\end{cases}$$

対称性を使う方法。 赤い菓子が出たところで止めても、赤い菓子がどの位置にあるかには影響しない。そこで、袋が空になるまで取り出し続けると想像する。（互いに区別

できる）菓子の並び方 $(n+1)!$ 通りはすべて同様に確からしいので、赤い菓子は $n+1$ 個の位置のどこにも等しい確率で現れる。これで同じ答えが得られる。

最大化のために r を固定する。$n < r-1$ ならば確率は 0 である。$n \geq r-1$ ならば確率は $\frac{1}{n+1}$ であり、これは n が増えると狭義単調に減少する。よって確率を最大にするのは条件を満たす最小の袋、すなわち

$$n = r-1,$$

であり、最大確率は $\frac{1}{r}$ である（$n = r-1$ のとき、赤い菓子は r 個の菓子のどれであることも同様に確からしく、それを最後に取り出すのは r 通りの同様に確からしい結果のうちの一つである）。

検算

- (i) では $\sum_{r\geq 1} P(R=r) = \frac{1}{n+1} \cdot \dfrac{1}{1-n/(n+1)} = 1$ であり、(ii) では $n+1$ 個の等しい値 $\frac{1}{n+1}$ の和は 1 である。
- (i) で $r=3$ のとき、$n=1,2,3$ はそれぞれ $\frac{1}{8}$、$\frac{4}{27} \approx 0.148$、$\frac{9}{64} \approx 0.141$ を与えるので、$n=2$ が最大である。
- (ii) については、$n \leq 6$ に対して菓子の並び方をすべて列挙して確かめた。(i) を $n=3$、$r=2$ でシミュレーションしたところ 0.188 となり、厳密値 $\frac{3}{16} = 0.1875$ と一致した。

よくある間違い

- (i) で $\left(\frac{n}{n+1}\right)^r \frac{1}{n+1}$ と書いてしまう：赤の前に来る青は $r-1$ 個だけである。
- (ii) で $r > n+1$ のとき答えが 0 になることを忘れる。まさにこのことが、「n をできるだけ小さくする」という方針が $n=0$ まで行ってしまうのを防いでいる。

STEP のその先へ

どちらの最大化も、実は**最尤推定**（maximum likelihood estimation）である。赤い菓子が初めて r 回目に現れたことは観察したが n はわからないとき、その観察結果をもっとも起こりやすくする n の値は、どちらの実験でも $n = r-1$ である。これは大学 1 年次の統計学の礎となる考え方である。(ii) の対称性による議論は**交換可能性**（exchangeability）の一例であり、これを使うと多くの非復元抽出の問題が簡単になる。

どちらの考え方も Blitzstein and Hwang, *Introduction to Probability*（確率論入門）で展開されている。

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。