解答例

STEP II 2008、第 13 問

確率・統計 ・ 非復元抽出

> **STEP II 2008、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/08-s2-q13

鍵となる考え方

すべての駒（counter）を追いかける必要はない。黒い駒だけを追えばよい。黒い駒それぞれの行き先は 2 回の取り出しで決まり、各取り出しにおいて特定の駒が取り出される確率は単に（取り出す個数）/（袋の中の個数）である。1 回目の取り出しで何が起こるかによって場合分けすると、どちらの小問も短い和になり、しかもどちらも予想以上にきれいに簡約される。

解答

以下を通して、「無作為に取り出す」とは、袋の中の k 個の駒の組がどれも等しい確率で取り出されることを意味する。繰り返し用いる事実は次のとおりである：m 個の駒が入った袋から k 個を無作為に取り出すとき、**特定の**駒がその中に含まれる確率は、対称性（symmetry）により k/m である（m 個の駒はどれも等しい確率で含まれ、平均して k 個が含まれるからである）。

1 回目の移動の後、袋 P には $n-k$ 個、袋 Q には $n+k$ 個の駒が入っている。2 回目の移動では、その $n+k$ 個のうち k 個が取り出されて P に移される。

小問 (i)

P から出発する黒い駒を B とする。1 回目の移動で B が動くかどうかで場合分けする。

- B が P から取り出されない場合：確率は $\dfrac{n-k}{n}$ である。このとき B は最後まで P にとどまる。2 回目の移動では Q からしか駒を取り出さないからである。

- B が P から取り出される場合：確率は $\dfrac{k}{n}$ である。このとき B は Q の中の $n+k$ 個の駒の 1 つであり、B が P に戻るのは、戻すために取り出される k 個の中に B が含まれるときであり、そのときに限る。その確率は $\dfrac{k}{n+k}$ である。

これらの場合は互いに排反（disjoint）であり、すべての可能性を尽くしているので、

$$\begin{aligned}\mathrm{P}(B\text{ が最後に入っている袋が}P) &= \frac{n-k}{n} + \frac{k}{n}\cdot\frac{k}{n+k} = \frac{(n-k)(n+k)+k^2}{n(n+k)} \\ &= \frac{n^2}{n(n+k)} = \frac{n}{n+k}.\end{aligned}$$

これは k について狭義単調減少であるから、$0 \le k \le n$ の範囲では $k=0$ のときに限り最大となり、そのとき値は 1 である。（驚くことではない。何も移さなければ、黒い駒は確実にその場にとどまる。）

小問 (ii)

今度は、P から出発する黒い駒を B_1、Q から出発する黒い駒を B_2 とする。2 つが最終的に同じ袋に入っているという事象を S とする。再び、1 回目の移動で B_1 が動くかどうかで場合分けする。

場合 1：B_1 が P にとどまる（確率 $\frac{n-k}{n}$）。このとき B_1 は何が起こっても最終的に P にあるので、S が起こるのは、Q の中の $n+k$ 個の駒から取り出される k 個の中に B_2 が含まれるときであり、そのときに限る：

$$\mathrm{P}(S \mid \text{場合 1}) = \frac{k}{n+k}.$$

場合 2：B_1 が Q に移る（確率 $\frac{k}{n}$）。このとき 2 つの黒い駒はどちらも Q の中の $n+k$ 個に含まれており、S が起こるのは、2 回目の取り出しでその**両方**を取り出すか、**どちらも**取り出さないときである。

> **考えてみよう** $n+k$ 個の駒から無作為に選んだ k 個の組が、与えられた 2 つの駒を両方とも含む確率、あるいはどちらも含まない確率はいくらか。

部分集合を数える。等しい確率で起こる選び方は $\binom{n+k}{k}$ 通りある。黒い駒を両方含む選び方では、残りの $k-2$ 個を残りの $n+k-2$ 個から選ぶ。どちらも含まない選び方では、k 個すべてを $n+k-2$ 個の白い駒から選ぶ。割り算をして、

$$\begin{aligned}\mathrm{P}(\text{両方}) &= \binom{n+k-2}{k-2}\Big/\binom{n+k}{k} = \frac{k(k-1)}{(n+k)(n+k-1)}, \\ \mathrm{P}(\text{どちらも含まない}) &= \binom{n+k-2}{k}\Big/\binom{n+k}{k} = \frac{n(n-1)}{(n+k)(n+k-1)}.\end{aligned}$$

（これらの式は $k=0,1$ のときも成り立ち、そのとき $\mathrm{P}(\text{両方})=0$ である。）したがって

$$\mathrm{P}(S \mid \text{場合 2}) = \frac{k(k-1)+n(n-1)}{(n+k)(n+k-1)}.$$

場合をまとめる。全確率の法則（law of total probability）により、

$$\begin{aligned}\mathrm{P}(S) &= \frac{n-k}{n}\cdot\frac{k}{n+k} + \frac{k}{n}\cdot\frac{k(k-1)+n(n-1)}{(n+k)(n+k-1)} \\ &= \frac{k\big[(n-k)(n+k-1)+k^2-k+n^2-n\big]}{n(n+k)(n+k-1)}.\end{aligned}$$

角括弧の中では $(n-k)(n+k-1) = n^2-k^2-n+k$ であるから、角括弧の中身は $2n^2-2n=2n(n-1)$ となる。よって

$$\boxed{\mathrm{P}(S) = \frac{2k(n-1)}{(n+k)(n+k-1)}}$$

最大化。$k=0$ のとき確率は 0 であるから、$1\le k\le n$ として逆数を最小化する。正の定数 $2(n-1)$ を除けば、

$$\frac{(n+k)(n+k-1)}{k} = k + (2n-1) + \frac{n(n-1)}{k},$$

であるから、整数の範囲で $g(k)=k+\dfrac{n(n-1)}{k}$ を最小化すればよい。隣どうしを比べる：

$$g(k+1)-g(k) = 1-\frac{n(n-1)}{k(k+1)}.$$

これは $k(k+1)<n(n-1)$ のとき、すなわち $k<n-1$ のとき負であり、$k=n-1$ のときちょうど 0 になる。そのとき $k(k+1)=(n-1)n$ となるからである。したがって g は $k=1$ から $k=n-1$ まで狭義単調減少し、$g(n)=g(n-1)$ である。ゆえに確率が最大となるのは

$$k=n-1 \quad \text{および} \quad k=n,$$

のときであり、$n\ge 2$ であるから、どちらも許される範囲にある。$k=n-1$ のとき最大値は

$$\frac{2(n-1)^2}{(2n-1)(2n-2)} = \frac{n-1}{2n-1},$$

であり、$k=n$ のときも同じく $\dfrac{2n(n-1)}{2n(2n-1)} = \dfrac{n-1}{2n-1}$ となる。（g を連続関数とみたときの最小点は $\sqrt{n(n-1)}$ であり、これは $n-1$ と n の間に厳密に位置する。2 つの整数で値が等しくなるのは近似ではなく厳密な事実である。）

検算

- 小問 (i) で $k = n$ のとき：$\frac{1}{2}$。このとき $2n$ 個の駒すべてが Q を通り、そのうち n 個が戻るので、黒い駒はどちらの袋にも等しい確率で入る。正しい。
- 小問 (ii) で $k = 0$ のとき 0 となる。黒い駒どうしは決して出会わないので正しい。$n = 2, k = 1$ のとき $\frac{2}{3 \cdot 2} = \frac{1}{3}$ となり、これは手計算で確かめられる。
- $2 \le n \le 7$ とすべての k について、起こりうる 2 回の取り出しの組をすべて（分数を用いて）厳密に数え上げた。どの場合も両方の式が一致し、(ii) の最大点は常に $\{n-1, n\}$ であった。例えば $n = 4$ のとき最大値は $\frac{3}{7} \approx 0.4286$ である。

よくある間違い

- 2 回目の取り出しで $k/(n+k)$ ではなく k/n を用いてしまう。袋 Q の駒は $n+k$ 個に増えている。
- 小問 (ii) の場合 2 で、「どちらも取り出されない」可能性を忘れる。黒い駒が両方とも Q に残れば、2 つは依然として同じ袋にある。
- 微分を用いて最大化し、$\sqrt{n(n-1)}$ を 1 つの整数に丸めてしまう。離散的な比較により、最大点が 2 つあり値が等しいことがわかる。

STEP のその先へ

両方の小問の背後にある道具は**超幾何**分布（hypergeometric distribution）である。これは、非復元抽出（sampling without replacement）で無作為に取り出した標本に含まれる、印のついた要素の個数の分布である。第 1 段階の結果で条件付け（conditioning）を行い、それを第 2 段階に用いるのが全確率の法則であり、同じ 2 段階の推論は、エーレンフェスト（Ehrenfest）の壺のような、容器間の混合や拡散のモデルの土台となっている。Blitzstein and Hwang, *Introduction to Probability*（確率論入門）は、超幾何分布と条件付けをまさにこのスタイルで扱っている。

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。