解答例

STEP III 2000、第 13 問

確率・統計 ・ 確率母関数

STEP III 2000、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/00-s3-q13

鍵となる考え方

どちらの確率変数（random variable）も、「裾」の事象、すなわち「累積」の事象を通して扱うのが最も簡単である。最初の 6 が r 回目より後に出るのは、最初の r 回の投げのすべてにおいて**どの**さいころも 6 を出さないときであり、かつそのときに限る。最後の 6 が r 回目までに出ているのは、**各**さいころがすでに 6 を出しているときであり、かつそのときに限る。これらはいずれも独立な事象の共通部分であるから、その確率は積で表され、連続する値の差をとれば分布が得られる。最初の 6 に関する確率変数が幾何分布（geometric distribution）に従うと分かれば、それに関することはすべてすぐに求まる。

解答

以下では、さいころは互いに独立に振る舞い、また各回の投げも互いに独立であると仮定する。1 個のさいころについて、さいころ i が初めて 6 を出す回の番号を X_i とおくと、

$$\mathrm{P}(X_i > r) = q^r \qquad (r = 0, 1, 2, \dots),$$

である。$X_i > r$ は 6 以外が r 回続くことを意味するからである。このとき $R = \min_i X_i$、$S = \max_i X_i$ である。

最初の 6

事象 $R > r$ は、最初の r 回のうちに n 個のさいころのどれも 6 を出していないこと、すなわち nr 回の個々の投げがすべて 6 以外であったことを表す。独立性により、

$$\mathrm{P}(R > r) = q^{nr}.$$

したがって、$r \geq 1$ に対して

$$\mathrm{P}(R = r) = \mathrm{P}(R > r-1) - \mathrm{P}(R > r) = q^{n(r-1)} - q^{nr} = q^{nr}(q^{-n} - 1),$$

となり、これが求める結果である。

$Q = q^n$ とおくと、これは $\mathrm{P}(R = r) = Q^{r-1}(1-Q)$ と書ける。すなわち R は幾何分布に従い、さいころ全体を 1 回投げることが確率 $1-Q$ で「成功」（少なくとも 1 個が 6）となる。よって確率母関数（probability generating function）は等比級数となる：

$$\begin{aligned} G_R(t) = \sum_{r=1}^{\infty} t^r Q^{r-1}(1-Q) &= (1-Q)\,t \sum_{r=1}^{\infty} (Qt)^{r-1} \\ &= \frac{(1-q^n)\,t}{1-q^n t}, \end{aligned}$$

これは $|t| < q^{-n}$（特に $t = 1$ の近く）で成り立つ。

平均を求めるため、商の微分法（quotient rule）により微分すると、

$$G_R'(t) = (1-q^n)\,\frac{(1-q^n t) + q^n t}{(1-q^n t)^2} = \frac{1-q^n}{(1-q^n t)^2},$$

となるので、

$$\mathrm{E}(R) = G_R'(1) = \frac{1}{1-q^n}.$$

$n = 2$、$p = \frac{1}{6}$ のとき $q^2 = \frac{25}{36}$ であるから、$\mathrm{E}(R) = 1 \big/ \frac{11}{36} = \frac{36}{11}$ となり、求める結果を得る。

最後の 6

事象 $S \leq r$ は、最初の r 回のうちのどこかで、すべてのさいころが 6 を出していることを表す。1 個のさいころについてこの確率は $1-q^r$ であり、さいころは互いに独立であるから、

$$\mathrm{P}(S \leq r) = (1-q^r)^n.$$

連続する値の差をとると、$r \geq 1$ に対して

$$\mathrm{P}(S = r) = \mathrm{P}(S \leq r) - \mathrm{P}(S \leq r-1) = (1-q^r)^n - (1-q^{r-1})^n,$$

となり、求める結果を得る。（$r = 1$ のとき第 2 項は $0^n = 0$ となり、正しい。）

考えてみよう $t^r[(1-q^r)^2-(1-q^{r-1})^2]$ の和をとるのは煩雑に見える。どうすれば等比級数に帰着できるだろうか？

$n=2$ のとき、平方を展開すると

$$\begin{aligned}\mathrm{P}(S=r) &= (1-2q^r+q^{2r})-(1-2q^{r-1}+q^{2r-2})\\ &= 2q^{r-1}(1-q)-q^{2(r-1)}(1-q^2)\\ &= 2p\,q^{r-1}-(1-q^2)\,(q^2)^{r-1}.\end{aligned}$$

各項は幾何分布（の定数倍）であるから、

$$G_S(t)=\sum_{r=1}^{\infty}t^r\,\mathrm{P}(S=r)=\frac{2pt}{1-qt}-\frac{(1-q^2)\,t}{1-q^2t}.$$

これは理にかなっている。$S=X_1+X_2-R$ であるから、その分布は「パラメータ p の幾何分布の 2 倍から R の分布を引いたもの」になるのである。整理するために、全体を $(1-qt)(1-q^2t)$ で通分し、$1-q^2=p(1+q)$ を用いると、

$$\begin{aligned}2p(1-q^2t)-p(1+q)(1-qt) &= p[\,2-2q^2t-1-q+qt+q^2t\,]\\ &= p[(1-q)+qt(1-q)]=p^2(1+qt).\end{aligned}$$

したがって

$$G_S(t)=\frac{p^2\,t\,(1+qt)}{(1-qt)(1-q^2t)}.$$

確認として、$G_S(1)=\dfrac{p^2(1+q)}{p\cdot p(1+q)}=1$ である。

平均を求めるには、分解した形を微分するのが最も簡単である。各部分は幾何分布の確率母関数であり、その 1 における導関数は、係数倍された平均となるからである。$\dfrac{(1-a)t}{1-at}$ の $t=1$ における微分係数は $\dfrac{1}{1-a}$ である（上の G_R の場合とまったく同じ）から、

$$\mathrm{E}(S)=G_S'(1)=2\cdot\frac{1}{p}-\frac{1}{1-q^2}=\frac{2(1+q)-1}{p(1+q)}=\frac{1+2q}{p(1+q)}.$$

$p=\frac{1}{6}$、$q=\frac{5}{6}$ のとき、

$$\mathrm{E}(S)=12-\frac{36}{11}=\frac{96}{11}\approx 8.73.$$

検算

- $\mathrm{E}(R)+\mathrm{E}(S)=\frac{36}{11}+\frac{96}{11}=12=\mathrm{E}(X_1)+\mathrm{E}(X_2)$ である。$\min+\max=X_1+X_2$ であるから、当然こうならなければならない。

- $n = 1$ のとき最初の 6 と最後の 6 は一致し、G_R は 1 個のさいころの幾何分布の確率母関数 $pt/(1-qt)$ に帰着する。
- さいころの組 2×10^5 組についてモンテカルロ・シミュレーション（Monte Carlo simulation）を行ったところ、標本平均は 3.274 と 8.709 となり、理論値 $\frac{36}{11} \approx 3.273$ と $\frac{96}{11} \approx 8.727$ とよく合った。また、G_S の級数を数値的に和をとって計算すると、いくつかの t の値で閉じた形と一致した。

よくある間違い

- $\mathrm{P}(S = r)$ を「あるさいころが r 回目に初めて 6 を出し、他のさいころはそれより前に出した」として書き、両方のさいころが同じ回に初めて 6 を出す場合を見落とすこと。$\mathrm{P}(S \le r)$ を用いれば、この問題は完全に避けられる。
- 分解した形ではなく、整理した G_S を微分すること。正しくはあるが、計算量がはるかに多く、誤りの機会も多い。

STEP のその先へ

R と S は独立な確率変数の最小値と最大値であり、**順序統計量**（order statistics）の最も簡単な例である。$\mathrm{P}(\min > r)$ と $\mathrm{P}(\max \le r)$ を用いる手法は、どんな分布に対しても有効である。無記憶性（memoryless property）をもつ幾何分布は指数分布（exponential distribution）の離散版であり、連続極限では、n 個の指数分布に従う確率変数の最大値の期待値は調和和 $1 + \frac{1}{2} + \cdots + \frac{1}{n}$（に平均を掛けたもの）となる。これはクーポン収集問題（coupon collector's problem）における重要な一歩である。Grimmett and Welsh, *Probability: An Introduction*（確率論入門）では、母関数とこれらの考え方が詳しく扱われている。

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。