解答例

STEP III 見本問題、第 16 問

確率・統計 ・ 負の二項分布

> **STEP III 見本問題、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/spec-s3-q16

鍵となる考え方

r 人の息子を待つことは、まず息子を 1 人待ち、そこからまた改めて次の息子を待つ、ということを r 回繰り返すのと同じである。したがって C は r 個の同じ待ち時間の和に分解でき、あとは期待値の線形性（linearity of expectation）が片づけてくれる。小問 (ii) では、大統領が子どもをもうけるのをやめた後も出生が続くと想像するのがこつである。そうすれば、ある時点までにやめているかどうかは**決まった**人数の子どもについての問題となり、男の子と女の子の間の対称性（symmetry）によって決着がつく。

解答

以下では一貫して、各出生は他のすべての出生とは独立に、確率 $\frac{1}{2}$ で男の子であるとする。

小問 (i)

$i = 1, 2, \ldots, r$ に対して、X_i を $(i-1)$ 番目の息子の後から i 番目の息子までに（i 番目の息子を含めて）生まれた子どもの人数とする（$i = 1$ のときは最初から数える）。すると

$$C = X_1 + X_2 + \cdots + X_r.$$

各 X_i は、新たに始まる独立で公平な出生の列において最初の男の子までの出生数を数えたものであるから、どの X_i も X と同じ分布（distribution）をもつ。ここで

$$\mathrm{P}(X = k) = \left(\tfrac{1}{2}\right)^{k-1} \cdot \tfrac{1}{2} = \left(\tfrac{1}{2}\right)^{k}, \qquad k = 1, 2, 3, \ldots,$$

である。$X = k$ とは、$k-1$ 人の女の子の後に男の子が生まれることを意味するからである。

$\mathrm{E}(X)$ を求めるには、$|x| < 1$ に対する等比級数（geometric series）$\sum_{k\geq 0} x^k = \dfrac{1}{1-x}$ を微分して $\sum_{k\geq 1} kx^{k-1} = \dfrac{1}{(1-x)^2}$ を得る。両辺に x を掛けて $x = \frac{1}{2}$ とおくと、

$$\mathrm{E}(X) = \sum_{k=1}^{\infty} k\left(\tfrac{1}{2}\right)^k = \frac{\frac{1}{2}}{\left(1-\frac{1}{2}\right)^2} = 2.$$

これは納得のいく結果である。生まれる赤ちゃんの半数が男の子なら、男の子 1 人を得るまでに平均して 2 回の出生を待つと予想されるからである。

期待値の線形性（これは X_i が独立であってもなくても成り立つが、ここでは実際に独立である）により、

$$\mathrm{E}(C) = \mathrm{E}(X_1) + \cdots + \mathrm{E}(X_r) = 2r.$$

小問 (ii)

限りなく続く赤ちゃんの列の性別があらかじめ独立かつ公平に決められており、大統領はこの列に沿って r 番目の息子に達するまで子どもをもうけると想像する。こう考えても C については何も変わらないが、大統領がそれより早くやめた場合でも「最初の $2r-1$ 人の赤ちゃん」について語ることができるようになる。

考えてみよう 事象 $C < 2r$ を、列の最初の $2r-1$ 人の赤ちゃんだけを用いて表すにはどうすればよいか。

C は整数であるから、$C < 2r$ は $C \leq 2r-1$ を意味する。すなわち、r 番目の息子が $2r-1$ 番目以前に生まれるということである。これが起こるのは、最初の $2r-1$ 人の赤ちゃんの中に男の子が少なくとも r 人いるとき、かつそのときに限る。よって

$$\mathrm{P}(C < 2r) = \mathrm{P}(\text{少なくとも}r\text{ 人の男の子が最初の}2r-1\text{ 人の赤ちゃんの中にいる}).$$

さて、これら $2r-1$ 人の赤ちゃんのうち男の子の人数を B、女の子の人数を G とすると、$B+G = 2r-1$ である。$2r-1$ は奇数なので B と G が等しくなることはなく、ちょうど一方だけが r 以上となる。

- $B \geq r$ ならば $G = 2r-1-B \leq r-1$ である。
- $B \leq r-1$ ならば $G \geq r$ である。

したがって事象 $\{B \geq r\}$ と $\{G \geq r\}$ は互いに排反（mutually exclusive）であり、両者を合わせるとあらゆる場合を尽くす。よって

$$\mathrm{P}(B \geq r) + \mathrm{P}(G \geq r) = 1.$$

ところが、どの出生においても男の子と女の子は等しい確率で生まれるので、「男の子」と「女の子」というラベルを入れ替えてもどの確率も変わらず、$\mathrm{P}(B \geq r) = \mathrm{P}(G \geq r)$ である。ゆえにどちらも $\frac{1}{2}$ であり、

$$\mathrm{P}(C < 2r) = \tfrac{1}{2},$$

となって、示すべきことが示された。

つまり、家族の人数の期待値は $2r$ であるのに、家族の人数が $2r$ より少なくなるのはちょうど半分の場合である。家族の人数の分布は偏っている。$2r$ をわずかに下回る値はよく現れ、ときおり起こる非常に長い女の子の連続が平均を押し上げているのである。

検算

- $r = 1$ のとき：$\mathrm{E}(C) = 2$ であり、$C < 2$ は最初の赤ちゃんが男の子であることを意味し、その確率は $\frac{1}{2}$ である。
- $r = 2$ のとき：$C < 4$ は $C = 2$ または 3 を意味し、その確率はそれぞれ $\frac{1}{4}$ と $2 \cdot \frac{1}{8} = \frac{1}{4}$ で、合計 $\frac{1}{2}$ である。
- $r = 1, \ldots, 8$ について $c\binom{c-1}{r-1}2^{-c}$ を厳密に足し合わせると、いずれの場合も $\mathrm{E}(C) = 2r$ かつ $\mathrm{P}(C < 2r) = \frac{1}{2}$ が得られる。また、$r = 3$ で 200 000 家族のシミュレーションを行うと、平均は 6.00、$\mathrm{P}(C < 6) \approx 0.499$ となった。

よくある間違い

- 「子どもの半分は男の子だから $C = 2r$」と論じること。これは平均としては正しいが、C 自体は確率変数である。小問 (i) では単なる推測ではなく、待ち時間への分解が必要である。
- 小問 (ii) で大統領の実際の子どもだけを考えること。$C < 2r - 1$ のときには $2r - 1$ 人の子どもは**存在しない**。出生の列をやめた時点の先まで延長することによって、はじめてこの数え上げの議論が正当になるのである。
- $2r - 1$ が奇数であることを忘れること。まさにこの事実によって男の子と女の子の同数が排除され、二つの事象が互いに余事象（complementary）となるのである。

STEP のその先へ

C の分布は**負の二項分布**（negative binomial distribution）であり、$c \geq r$ に対して $\mathrm{P}(C = c) = \binom{c-1}{r-1}2^{-c}$ である。確率変数をより単純な部分の和として表し、期待値の線

形性を用いるという手法は、大学 1 年次の確率論において最も強力な道具の一つである。また、小問 (ii) での「待ち時間」を「決まった試行回数の中での個数」に結びつけるこつは、負の二項分布と二項分布を結ぶ双対性（duality）と同じものである（連続時間では、ガンマ分布とポアソン（Poisson）分布を結ぶ双対性にあたる）。Blitzstein と Hwang による *Introduction to Probability*（確率論入門）は、これら両方の考え方を多くの例とともに扱っている。

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。