解答例

STEP I 1988、第 15 問

確率・統計 ・ 二項検定

> **STEP I 1988、問 15**
>
> 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/88-s1-q15

鍵となる考え方

「得点の半分より多くをファウルから得る」という条件を、ファウルの数とゴールの数の間の簡単な不等式（inequality）に書き直す。そのうえでゴールの数で条件付け（conditioning）をすると、あとは等確率なファウルの値を数えるだけになる。後半では、**以前の**分布のもとで 300 回の試行のうち 9 回のファウルが起こる回数がどれくらいと期待されるか、そして何も変わっていないとしたら観測された回数がどれほど意外なものかを考える。

解答

確率

ある試合におけるある特定のチームについて、そのチームが挙げるゴールの数を G、そのチームに対して犯されるファウルの数を F とする。このとき G と F は独立（independent）であり、

$$\mathrm{P}(G=0),\ldots,\mathrm{P}(G=4) = \tfrac{1}{9},\tfrac{2}{9},\tfrac{3}{9},\tfrac{2}{9},\tfrac{1}{9}, \qquad \mathrm{P}(F=f)=\tfrac{1}{10} \quad (f=0,1,\ldots,9).$$

である。チームはゴールから $2G$ 点、ファウルから F 点を得るので、合計は $2G+F$ 点である。得点の半分より多くをファウルから得るのは、ちょうど

$$F > \tfrac{1}{2}(2G+F) \iff F > 2G.$$

のときである。（チームがまったく得点しない場合は $F=0=2G$ となり、これは条件を満たさない。厳密な不等号を用いればこの場合も正しく扱える。）

$G = g$ のもとで条件を満たす F の値は $2g+1, \ldots, 9$ であり、これは等確率な 10 個の値のうち $9-2g$ 個である。よって F と G の独立性を用いると、

$$\mathrm{P}(F > 2G \mid G = g) = \frac{9-2g}{10}, \qquad g = 0, 1, 2, 3, 4,$$

となり、値は $\frac{9}{10}, \frac{7}{10}, \frac{5}{10}, \frac{3}{10}, \frac{1}{10}$ である。全確率の公式（law of total probability）により、

$$\mathrm{P}(F > 2G) = \frac{1\cdot 9 + 2\cdot 7 + 3\cdot 5 + 2\cdot 3 + 1\cdot 1}{9 \times 10} = \frac{45}{90} = \frac{1}{2}.$$

この簡潔な答えには理由がある。ゴールの分布は 2 に関して対称であり、$(9-2g)/10$ は g の一次式なので、その平均は $g = 2$ における値、すなわち $\frac{1}{2}$ に等しいのである。

ファウルの分布は変化したか？

あるチームが犯すファウルの数は、相手チームに対して犯されるファウルの数であるから、以前のルールのもとでは、各チームが 1 試合で 9 回のファウルを犯す確率は $\frac{1}{10}$ である。300 のチーム試合のうち、チームが 9 回のファウルを犯したものの数を X とする。次の検定を行う。

$$H_0 : \ p = \tfrac{1}{10} \ \text{（変化なし）} \qquad \text{対} \qquad H_1 : \ p < \tfrac{1}{10},$$

ここで p はチームが 1 試合で 9 回のファウルを犯す確率である。ルール変更は多数のファウルが起こる可能性を**下げる**ことを意図したものなので、対立仮説（alternative hypothesis）は片側（one-sided）である。

> **考えてみよう** ここで二項分布（binomial distribution）のモデルを用いることは正当化されるか？

300 のチーム試合が独立で、それぞれ同じ確率 p をもつ必要がある。問題文は 1 試合の中での独立性を与えており、異なる試合どうしも独立であると仮定するのは妥当である。よって H_0 のもとで $X \sim \mathrm{B}(300, \frac{1}{10})$ であり、

$$\mathrm{E}(X) = 30, \qquad \mathrm{Var}(X) = 300 \times \tfrac{1}{10} \times \tfrac{9}{10} = 27, \qquad \text{標準偏差（standard deviation）} \sqrt{27} \approx 5.20.$$

である。観測値は $X = 12$ であり、期待値 30 より 18 小さく、これは標準偏差の約 3.5 倍にあたる。正規近似（normal approximation）と連続性補正（continuity correction）を用いると、

$$\mathrm{P}(X \le 12) \approx \Phi\left(\frac{12.5 - 30}{\sqrt{27}}\right) = \Phi(-3.37) \approx 0.0004.$$

二項分布による正確な確率はさらに小さく、約 0.0001 である。いずれにせよ、何も変わっていなければ、これほど極端な結果が起こるのは 1000 シーズンに 1 回未満である。これは通常の 5% や 1% の有意水準（significance level）さえはるかに超えている

ので、H_0 を棄却する。これはファウルの数の分布が変化したことの十分な証拠である。しかも期待された方向、すなわち 9 回のファウルが起こる試合が少なくなる方向に変化している。

注意すべき点が二つある。この検定が示すのは P(9 回のファウル) が減少したことだけであり、分布の残りの部分がどう変化したかは示さない（ただし、それが自然な説明ではある）。また、シーズンの間に他の要因が変わった可能性もあるので、ルール変更がその減少を引き起こしたことを統計学で証明することはできない。

検算

- ゴールの確率 $\frac{1}{9}(1+2+3+2+1)$ の和は、当然そうあるべきとおり 1 である。
- 5×10 通りのすべての組 (g, f) を正確な分数で列挙すると $\mathrm{P}(F > 2G) = \frac{1}{2}$ となり、200,000 試合のシミュレーションでは 0.4997 となった。
- 二項確率を直接足し合わせると、$X \sim \mathrm{B}(300, 0.1)$ に対して $\mathrm{P}(X \leq 12) = 9.2 \times 10^{-5}$ となる。正規近似（3.8×10^{-4}）と平均 30 のポアソン近似（Poisson approximation）（1.7×10^{-4}）も、この確率が非常に小さいことで一致している。

よくある間違い

- F を $2G$ ではなく G と比較してしまう。1 ゴールは 2 点の価値がある。
- $F > 2G$ ではなく $F \geq 2G$ を数えてしまう。「半分より多く」は厳密な不等号であり、得点のないチームはファウルから得点を得ていない。
- 標本の大きさとして 300 のチーム試合ではなく 150 試合を用いてしまう。あるいは、計算の前に帰無仮説（null hypothesis）と二項分布のモデルを述べ忘れる。

STEP のその先へ

後半は二項母比率に関する仮説検定（hypothesis test）であり、求めた裾の確率は p 値（p-value）である。大学では、尤度比検定（likelihood ratio test）やカイ二乗検定（chi-squared test）を学ぶ。これらは一つの値だけでなく、観測されたファウル数の分布全体を以前の分布と比較するものである。Blitzstein and Hwang, *Introduction to Probability*（確率論入門）は、確率の側面（条件付け、二項分布、正規近似）を非常にわかりやすく扱っている。

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。