解答例

STEP III 2016、第 12 問

確率・統計 ・ チェビシェフの不等式

> **STEP III 2016、問 12**
>
> 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/16-s3-q12

鍵となる考え方

チェビシェフ（Chebyshev）の不等式（Chebyshev's inequality）は平均から遠く離れる確率を上から評価するものであるから、何かが平均に近いという主張は余事象（complement）から得られる。各小問でなすべきことは、適切な確率変数（random variable）を見抜き、その平均と標準偏差（standard deviation）を求め、$k\sigma$ がちょうど注目している距離になるように k を選ぶことである。小問 (ii) では、e^n と並んで現れる和 $\sum n^r/r!$ が、ポアソン分布（Poisson distribution）の「指紋」である。

解答

小問 (i)

$100n$ 回の投げのうち表が出る回数を X とする。各回の投げは独立（independent）で、それぞれ確率 0.2 で表が出ると仮定すると、$X \sim \mathrm{B}(100n, 0.2)$ であるから、

$$\mu = 100n \times 0.2 = 20n, \qquad \sigma^2 = 100n \times 0.2 \times 0.8 = 16n, \qquad \sigma = 4\sqrt{n}.$$

事象 $16n \le X \le 24n$ はちょうど $|X - 20n| \le 4n$ であるから、

$$\alpha = \mathrm{P}(|X - \mu| \le 4n) = 1 - \mathrm{P}(|X - \mu| > 4n).$$

$k\sigma = 4n$、すなわち $k \cdot 4\sqrt{n} = 4n$ としたいので、$k = \sqrt{n}$ であり、これは正である。チェビシェフの不等式により

$$\mathrm{P}(|X - \mu| > 4n) = \mathrm{P}(|X - \mu| > \sqrt{n}\,\sigma) \le \frac{1}{(\sqrt{n})^2} = \frac{1}{n},$$

となり、したがって

$$\alpha \geq 1 - \frac{1}{n},$$

が示された。（条件 $n > 1$ は、この評価が意味のある内容をもつようにするためのものにすぎない。$n = 1$ のときは $\alpha \geq 0$ となる。）

小問 (ii)

考えてみよう 確率が $n^r/r!$ のような形をしている確率変数は何か。

$Y \sim \mathrm{Po}(n)$ ならば $r = 0, 1, 2, \ldots$ に対して $\mathrm{P}(Y = r) = \mathrm{e}^{-n} n^r / r!$ であるから、

$$\mathrm{e}^{-n}\left(1 + n + \frac{n^2}{2!} + \cdots + \frac{n^{2n}}{(2n)!}\right) = \mathrm{P}(Y \leq 2n).$$

したがって、示すべき不等式は、$\mathrm{e}^n > 0$ で割ると

$$\mathrm{P}(Y \leq 2n) \geq 1 - \frac{1}{n}.$$

と同値である。ポアソン分布では平均と分散（variance）はともに n であるから、$\mu = n$、$\sigma = \sqrt{n}$ である。

考えてみよう チェビシェフの不等式は両側の評価であるが、$Y \leq 2n$ は片側の事象である。両者をどう対応させればよいか。

$Y \geq 0$ であるから常に $Y - n \geq -n$ が成り立ち、$Y - n < -n$ は起こりえない。よって

$$|Y - n| > n \iff Y - n > n \iff Y > 2n.$$

$k\sigma = n$ となるように $k = \sqrt{n}$ とすると、チェビシェフの不等式により

$$\mathrm{P}(Y > 2n) = \mathrm{P}(|Y - n| > \sqrt{n}\,\sigma) \leq \frac{1}{n},$$

となるので、$\mathrm{P}(Y \leq 2n) = 1 - \mathrm{P}(Y > 2n) \geq 1 - \frac{1}{n}$ である。両辺に e^n を掛けると

$$1 + n + \frac{n^2}{2!} + \cdots + \frac{n^{2n}}{(2n)!} \geq \left(1 - \frac{1}{n}\right)\mathrm{e}^n,$$

となり、これが求める結果である。

検算

これらの評価は正しいが、かなり甘い。$n = 2$ のとき、(i) の二項分布による正確な確率は $\alpha \approx 0.868$ であり、評価の値 0.5 を大きく上回る。$n = 10$ のときは $\alpha \approx 0.9986$ で、評価の値は 0.9 である。(ii) で $n = 2$ のとき、左辺は $1 + 2 + 2 + \frac{4}{3} + \frac{2}{3} = 7$ であり、一方 $\frac{1}{2}\mathrm{e}^2 \approx 3.69$ である。両方の不等式を、(i) では 2 から 30 までのすべての n について、(ii) では 1 から 59 までのすべての n について厳密に確かめた。

よくある間違い

- k を選ぶときに、標準偏差 $4\sqrt{n}$ の代わりに分散 $16n$ を使ってしまう。
- チェビシェフの不等式が評価するのは区間の**外側**にある確率であり、α を得るには余事象を考える必要があることを忘れる。
- (ii) で、下側の裾 $Y < 0$ が何も寄与しない理由を説明しない。

STEP のその先へ

チェビシェフの不等式は**集中不等式**（concentration inequality）の最も簡単な例であり、これを m 回の独立な試行の平均に適用すると大数の弱法則（weak law of large numbers）が得られる。すなわち、ばらつきは $1/\sqrt{m}$ のように縮む。小問 (ii) はまた、指数関数の級数のうち $n^{2n}/(2n)!$ より先の部分（裾）が高々 e^n/n であることも述べており、これは確率によって証明された解析的な事実である。チェルノフ限界（Chernoff bounds）のようなより鋭い道具を使うと、真の確率は 1 に指数関数的に近いことがわかる。さらに詳しくは、Grimmett and Welsh, *Probability: An Introduction*（確率論入門）を参照されたい。

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。