解答例

STEP II 2000、第 14 問

確率・統計 ・ 中央値の分布

STEP II 2000、問 14

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-s2-q14

この問題は、*Beyond STEP* の第 13.4 節「平均か中央値か？」で詳しく解説されている：
https://beyondstep.keithmcnulty.org/ja/ch13/mean-or-median.html

鍵となる考え方

中央値（median）の確率密度関数（density）は $y^r(1-y)^s$ という形の多項式なので、与えられた積分ですべてが片づく。まず正規化定数（normalising constant）を求めるのに一度使い、次に y の指数を 1 だけ、さらに 2 だけ上げて、1 次と 2 次のモーメント（moment）を求めるのに使う。最後の小問では、Y と $\bar{X}$ はともに $\frac{1}{2}$ を中心としており、正規分布（normal distribution）の場合、中心から一定の距離以内に入る確率は標準偏差（standard deviation）だけで決まる。散らばりの小さい確率変数のほうが勝つのである。

解答

定数 k

確率密度関数は積分すると 1 にならなければならない。与えられた結果を $r = s = n$ として用いると、

$$1 = \int_0^1 ky^n(1-y)^n \,\mathrm{d}y = k\,\frac{n!\,n!}{(2n+1)!}, \qquad \text{よって} \qquad k = \frac{(2n+1)!}{(n!)^2},$$

となり、示すべき式が得られる。

平均

確率密度関数は $y = \frac{1}{2}$ に関して対称である（y を $1-y$ に置き換えても変わらない）から、$\mathrm{E}(Y) = \frac{1}{2}$ である。これを積分で確かめるには、$r = n+1$、$s = n$ とすればよい：

$$\mathrm{E}(Y) = k\int_0^1 y^{n+1}(1-y)^n \,\mathrm{d}y = \frac{(2n+1)!}{(n!)^2}\cdot\frac{(n+1)!\,n!}{(2n+2)!} = \frac{n+1}{2n+2} = \frac{1}{2}.$$

分散

次に $r = n+2$、$s = n$ とする：

$$\mathrm{E}(Y^2) = \frac{(2n+1)!}{(n!)^2}\cdot\frac{(n+2)!\,n!}{(2n+3)!} = \frac{(n+2)(n+1)}{(2n+3)(2n+2)} = \frac{n+2}{2(2n+3)}.$$

したがって

$$\mathrm{Var}(Y) = \frac{n+2}{2(2n+3)} - \frac{1}{4} = \frac{2(n+2)-(2n+3)}{4(2n+3)} = \frac{1}{4(2n+3)}.$$

中央値と平均の比較

各 X_i は $[0,1]$ 上の一様分布（uniform distribution）に従うので、$\mathrm{E}(X_i) = \frac{1}{2}$、$\mathrm{Var}(X_i) = \frac{1}{12}$ である。X_i は互いに独立（independent）であるから、

$$\mathrm{E}(\bar{X}) = \tfrac{1}{2}, \qquad \mathrm{Var}(\bar{X}) = \frac{1}{(2n+1)^2}\cdot(2n+1)\cdot\frac{1}{12} = \frac{1}{12(2n+1)}.$$

考えてみよう どちらの確率変数も $\frac{1}{2}$ を中心としている。中心から $d/\sqrt{n}$ 以内に入る確率がどちらのほうが大きいかを、ただ一つの数で決めるとしたら、それは何か？

標準偏差である。n が大きいとき、$Y \sim \mathrm{N}(\frac{1}{2}, \sigma_Y^2)$、$\bar{X} \sim \mathrm{N}(\frac{1}{2}, \sigma_{\bar{X}}^2)$ とみなしてよい。$W \sim \mathrm{N}(\frac{1}{2}, \sigma^2)$ かつ $c > 0$ ならば、標準正規分布の分布関数（distribution function）を Φ と書くと、

$$\mathrm{P}\left(|W - \tfrac{1}{2}| < c\right) = \mathrm{P}\left(|Z| < \frac{c}{\sigma}\right) = 2\Phi\left(\frac{c}{\sigma}\right) - 1,$$

となる。ここで $Z \sim \mathrm{N}(0,1)$ である。Φ は狭義単調増加であるから、これは σ の狭義**単調減少**関数である。よって $c = d/\sqrt{n}$ とすると、示すべき不等式が成り立つのは $\sigma_Y > \sigma_{\bar{X}}$、すなわち $\mathrm{Var}(Y) > \mathrm{Var}(\bar{X})$ のときであり、かつそのときに限る：

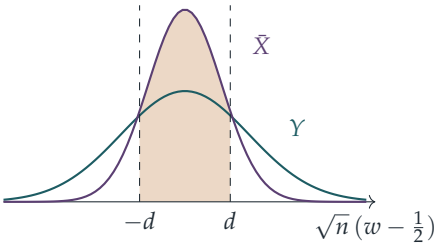
$$\begin{aligned}\frac{1}{4(2n+3)} > \frac{1}{12(2n+1)} &\iff 12(2n+1) > 4(2n+3)\\ &\iff 24n + 12 > 8n + 12\\ &\iff n > 0.\end{aligned}$$

これはすべての $n \geq 1$ で成り立つ。したがって、正規近似（normal approximation）が妥当となるほど n が大きければ、正の数 d が何であっても、Y についての確率は $\bar{X}$ についての確率より真に小さい。以上で証明が完了する。

その差がどれほどかを見るために、標準化（standardise）する：

$$\frac{d/\sqrt{n}}{\sigma_Y} = 2d\sqrt{\frac{2n+3}{n}} \to 2\sqrt{2}\,d, \qquad \frac{d/\sqrt{n}}{\sigma_X} = d\sqrt{\frac{12(2n+1)}{n}} \to 2\sqrt{6}\,d.$$

漸近的には、平均の標準偏差は $\sqrt{3}$ 倍だけ小さい。



図は、二つの極限の確率密度関数を $\sqrt{n}\,(w-\frac{1}{2})$ のスケールで示したものである。幅の狭いほうの曲線（平均）は、中心付近のどの帯においても、より多くの確率を含む。

検算

- $n = 0$ のとき：一つの一様確率変数の「中央値」は X_1 そのものであり、実際に公式から $k = 1$、$\mathrm{Var}(Y) = \frac{1}{12}$ という一様分布の分散が得られ、あるべき結果と一致する。
- 二つの分散は $n = 0$ のときともに $\frac{1}{12}$ に等しく、$n \geq 1$ では異なる値になる。これは上の不等式 $n > 0$ と合っている。
- $n = 50$（101 個の一様確率変数）でシミュレーションを行うと、$\mathrm{Var}(Y) \approx 0.00243$ となり、$1/412 \approx 0.00243$ と一致した。$d = 0.3$ とすると $\mathrm{P}(|Y - \frac{1}{2}| < d/\sqrt{n}) \approx 0.607$、$\mathrm{P}(|\bar{X} - \frac{1}{2}| < d/\sqrt{n}) \approx 0.859$ となり、正規近似による値 0.611 と 0.860 に近かった。

よくある間違い

- $\mathrm{Var}(\bar{X}) = \frac{1}{12}$ と書いてしまう。あるいは、分母に $(2n+1)^2$ が現れてから $2n+1$ 倍するのではなく、$2n+1$ で一度だけ割ってしまう。
- 向きを取り違える：分散が**大きい**ほど、中心付近に入る確率は**小さく**なる。
- $\mathrm{E}(Y^2)$ には $r = 2n$ ではなく $r = n+2$ を使うことを忘れる。

STEP のその先へ

中央値の確率密度関数はベータ分布（Beta distribution）Beta$(n+1, n+1)$ であり、与えられた積分はベータ関数（Beta function）である。いま行った比較は**漸近相対効率**（asymptotic relative efficiency）という考え方そのものである。一様分布のデータでは、標本平均（sample mean）は中央値の 3 分の 1 の分散で中心を推定するが、裾の重い（heavy-tailed）分布のデータでは中央値のほうが勝ることもある。順序統計量（order statistics）とベータ関数については、Blitzstein and Hwang, *Introduction to Probability*（確率論入門）によくまとめられている。

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。