解答例

STEP III 1989、第 15 問

確率・統計 ・ 中央値の分布

STEP III 1989、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/89-s3-q15

鍵となる考え方

$2n+1$ 個の値の中央値（median）が z の近くにあるのは、ちょうど 1 個の値が z にある微小な区間に入り、n 個の値がそれより下に、n 個の値がそれより上にあるときである。小問 (i) と (ii) は「すべて下にある」部分と「すべて上にある」部分の確率を与え、どの値がどの役割を担うかの選び方を数え上げれば確率密度関数（density）が得られる。その後は、確率密度関数の積分が 1 になるという事実から積分の値がただちに求まり、同じ技法を n の代わりに $n+1$ として用いれば分散（variance）が得られる。

解答

以下を通して、X は $[-c,c]$ 上で確率密度 $\frac{1}{2c}$ をもつので、$-c \le k \le c$ に対して

$$\mathrm{P}(X > k) = \frac{c-k}{2c}, \qquad \mathrm{P}(X < k) = \frac{c+k}{2c}.$$

である。

小問 (i)

n 個の値は独立（independent）であるから、確率を掛け合わせて

$$\mathrm{P}(n \text{ 個の値がすべて次より大きい：} k) = \left(\frac{c-k}{2c}\right)^n.$$

小問 (ii)

同様にして

$$\mathrm{P}(n \text{ 個の値がすべて次より小さい：} k) = \left(\frac{c+k}{2c}\right)^n.$$

中央値の確率密度関数

$X_1, \ldots, X_{2n+1}$ を標本とし、Z をその中央値、すなわち $(n+1)$ 番目に小さい値とする。X は連続型であるから、値が一致する確率は 0 であり、そのような場合は無視してよい。各 X_i は $[-c, c]$ に属するので、Z も $[-c, c]$ に属する。

z を $(-c, c)$ にとり、小さな $\delta z > 0$ をとる。$(\delta z)^2$ のオーダーの項を除けば、事象 $z < Z \le z + \delta z$ が起こるのはちょうど次の場合である。

- 値のうち 1 個が $(z, z+\delta z]$ に入る。
- 残りの値のうち n 個が z より小さい。
- 残りの n 個の値が $z+\delta z$ より大きい。

（微小な区間に 2 個以上の値が入る配置の確率は $(\delta z)^2$ のオーダーであり、極限では影響しない。）

> **考えてみよう** $2n+1$ 個の値をこれら 3 つの役割に割り当てる方法は何通りあるか。

微小な区間に入る値の選び方が $2n+1$ 通りあり、その後、残りのうちどの n 個を z より下に置くかの選び方が $\binom{2n}{n}$ 通りある。したがって割り当ては

$$(2n+1)\binom{2n}{n} = \frac{(2n+1)!}{(n!)^2}$$

通りである。これらは互いに排反であり、どれも同じ確率をもつ。ある 1 つの割り当てを固定すると、独立性と、$k = z$ とした小問 (i), (ii) により、その確率は

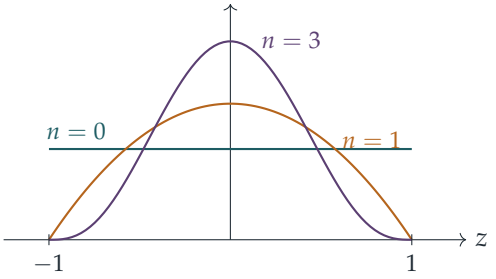
$$\frac{\delta z}{2c}\left(\frac{c+z}{2c}\right)^n\left(\frac{c-z-\delta z}{2c}\right)^n.$$

となる。すべての割り当てについて足し合わせ、δz で割って $\delta z \to 0$ とすると、Z の確率密度関数は

$$\begin{aligned} g(z) &= \frac{(2n+1)!}{(n!)^2}\cdot\frac{1}{2c}\cdot\frac{(c+z)^n(c-z)^n}{(2c)^{2n}} \\ &= \frac{(2n+1)!}{(n!)^2(2c)^{2n+1}}\,(c^2-z^2)^n, \qquad -c \le z \le c, \end{aligned}$$

である。ここで $(c+z)(c-z) = c^2 - z^2$ を用いた。これが求める結果である。

下の図は $c = 1$ のときの確率密度関数を示している。$n = 0$ のとき「中央値」は 1 個の値そのものであり、一様分布（uniform distribution）に従う。n が大きくなるにつれて、中央値は 0 の近くに集中する。



積分

確率密度関数をその値域全体で積分すると 1 になるので

$$\frac{(2n+1)!}{(n!)^2(2c)^{2n+1}} \int_{-c}^{c} (c^2 - z^2)^n \, \mathrm{d}z = 1,$$

であり、したがって

$$\int_{-c}^{c} (c^2 - z^2)^n \, \mathrm{d}z = \frac{(n!)^2(2c)^{2n+1}}{(2n+1)!}.$$

である。これはすべての整数 $n \geq 0$ について成り立ち、このことをすぐ後で用いる。

平均と分散

確率密度関数 g は z の偶関数（even function）であるから、$zg(z)$ は奇関数（odd function）であり、対称な区間 $[-c, c]$ 上でのその積分は 0 である。よって

$$\mathrm{E}(Z) = 0.$$

ゆえに $\mathrm{var}(Z) = \mathrm{E}(Z^2)$ である。$K_n = \dfrac{(2n+1)!}{(n!)^2(2c)^{2n+1}}$ および $I_n = \displaystyle\int_{-c}^{c} (c^2 - z^2)^n \, \mathrm{d}z = \frac{1}{K_n}$ とおく。

考えてみよう $z^2(c^2 - z^2)^n$ を展開せずに積分するにはどうすればよいか。

$z^2 = c^2 - (c^2 - z^2)$ と書く。すると

$$\begin{aligned} \mathrm{E}(Z^2) &= K_n \int_{-c}^{c} \big(c^2 - (c^2 - z^2)\big)(c^2 - z^2)^n \, \mathrm{d}z \\ &= K_n \big(c^2 I_n - I_{n+1}\big) = c^2 - \frac{K_n}{K_{n+1}}, \end{aligned}$$

となる。ここで $K_n I_n = 1$ と、積分の結果で n の代わりに $n+1$ とした $I_{n+1} = 1/K_{n+1}$ を用いた。さて

$$\frac{K_n}{K_{n+1}} = \frac{(2n+1)!}{(2n+3)!} \cdot \frac{((n+1)!)^2}{(n!)^2} \cdot \frac{(2c)^{2n+3}}{(2c)^{2n+1}} = \frac{(n+1)^2 \cdot 4c^2}{(2n+2)(2n+3)} = \frac{2(n+1)c^2}{2n+3}.$$

したがって

$$\operatorname{var}(Z) = c^2 \left(1 - \frac{2n+2}{2n+3}\right) = \frac{c^2}{2n+3}.$$

である。

検算

- $n = 0$ のとき、中央値は X の 1 個の値にすぎない。確率密度関数は $\frac{1}{2c}$ となり、積分は $2c$、分散は $\frac{c^2}{3}$ となる。これは $[-c, c]$ 上の一様分布の分散である。
- $n = 1$ かつ $c = 1$ のとき、積分の公式から $\int_{-1}^{1} (1 - z^2)\,\mathrm{d}z = \frac{1 \cdot 8}{6} = \frac{4}{3}$ が得られ、これは正しい。
- $c = 1.7$ として大きさ 5（$n = 2$）の標本を 10^5 個シミュレーションしたところ、中央値の分散は 0.415 となり、$c^2/7 \approx 0.413$ とよく一致した。

よくある間違い

- 中央の値の選び方に対応する因子 $2n+1$ を忘れ、$\binom{2n}{n}$ だけを用いてしまう。
- 中央の値自身の確率密度に対応する因子 $\frac{1}{2c}$ を落とし、累乗が $(2c)^{2n+1}$ ではなく $(2c)^{2n}$ のままになってしまう。
- $\mathrm{E}(Z^2)$ を求めるために $(c^2 - z^2)^n$ を二項展開してしまう。それでも計算はできるが、和を簡単にするのは難しい。

STEP のその先へ

中央値は標本の**順序統計量**（order statistics）の 1 つである。同じ数え上げの議論により、任意の連続型分布からとった m 個の値のうち r 番目に小さい値の確率密度関数が $\frac{m!}{(r-1)!\,(m-r)!} F^{r-1}(1-F)^{m-r} f$ として得られる。一様分布の場合、これらは**ベータ分布**（beta distributions）であり、導いた積分はベータ関数（beta function）$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ の特別な場合である。分散 $c^2/(2n+3)$ は $1/n$ のように小さくなり、これは標本中央値の漸近正規性（asymptotic normality）への第一歩である。Blitzstein と Hwang によ

る *Introduction to Probability*（確率論入門）は、順序統計量とベータ分布をよく扱っている。

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。