解答例

STEP II 1988、第 16 問

確率・統計 ・ 正規分布

STEP II 1988、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/88-s2-q16

鍵となる考え方

根に関するすべては、ただ一つの確率変数（random variable）B によって決まる。したがって、X_1 と X_2 についての問いはどれも、B がどの区間にあるかという問いに置き換えればよい。判別式（discriminant）を見れば根が実数になる条件がわかり、解と係数の関係（Vieta's formulas）（和 $-2B$、積 1）を使えば (i) と (ii) の条件は容易に言い換えられる。そのあとは、どの答えも標準正規（standard normal）確率変数についての条件付き確率（conditional probability）または条件付き期待値（conditional expectation）であり、期待値の計算には短い積分が一つ必要なだけである。

解答

以下では一貫して $B \sim \mathrm{N}(0,1)$ とし、その確率密度関数（density）を $\phi(z) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-z^2/2}$、分布関数（distribution function）を Φ とする。根が実数であるという事象を R とおく。

実数解

$X^2 + 2BX + 1$ の判別式は $4B^2 - 4$ であるから、根が実数になるのはちょうど $B^2 \geq 1$、すなわち $|B| \geq 1$ のときである。正規分布の対称性により、

$$\mathrm{P}(R) = \mathrm{P}(|B| \geq 1) = 2(1 - \Phi(1)) = 2(1 - 0.84134) = 0.31731\ldots \approx 0.317.$$

（重解の場合 $|B| = 1$ を含めるかどうかは結果に影響しない。その確率は 0 だからである。）

根が実数のとき、根は

$$X_{1,2} = -B \pm \sqrt{B^2 - 1}, \qquad X_1 + X_2 = -2B, \qquad X_1 X_2 = 1.$$

小問 (i)

考えてみよう 「両方の根が $\frac{1}{5}$ より大きい」を B についての条件にどう言い換えればよいか。

まず、両方の根が正でなければならない。積は $1 > 0$ なので二つの根は同符号であり、その符号は和 $-2B$ の符号に等しい。よって $B < 0$ が必要であり、R と合わせると $B \le -1$ となる。$b = -B \ge 1$ とおくと、根は $b \pm \sqrt{b^2 - 1}$ であり、どちらも正である。

積が 1 であるから、小さい方の根は大きい方の根の逆数である：

$$b - \sqrt{b^2 - 1} = \frac{1}{b + \sqrt{b^2 - 1}}.$$

したがって、小さい方の根が $\frac{1}{5}$ より大きいことは、大きい方の根が 5 より小さいことと同値であり、このとき大きい方の根も自動的に $\frac{1}{5}$ より大きい。よって条件は

$$\sqrt{b^2 - 1} < 5 - b.$$

これには $b < 5$ が必要であり、$1 \le b < 5$ では両辺とも非負なので、不等式の向きを変えずに両辺を 2 乗できる：

$$b^2 - 1 < 25 - 10b + b^2 \iff 10b < 26 \iff b < 2.6.$$

（$2.6 < 5$ なので、$b \ge 5$ の値はいずれにせよ条件を満たさない。）以上より、根が実数であるという条件のもとで、両方の根が $\frac{1}{5}$ より大きいのはちょうど $-2.6 < B \le -1$ のときであり、

$$\begin{aligned} \mathrm{P}(X_1, X_2 > \tfrac{1}{5} \mid R) &= \frac{\mathrm{P}(-2.6 < B \le -1)}{\mathrm{P}(|B| \ge 1)} = \frac{\Phi(2.6) - \Phi(1)}{2(1 - \Phi(1))} \\ &= \frac{0.99534 - 0.84134}{0.31731} = 0.4853\ldots \approx 0.485. \end{aligned}$$

小問 (ii)

$|X_1 + X_2| = 2|B|$ であるから、求めるものは $\mathrm{E}(2|B| \mid |B| \ge 1)$ である。事象を条件とする条件付き期待値は、その事象の上で積分し、その事象の確率で割ればよい：

$$\mathrm{E}(|B| \mid |B| \ge 1) = \frac{1}{\mathrm{P}(|B| \ge 1)} \int_{|z| \ge 1} |z|\, \phi(z)\, \mathrm{d}z = \frac{2\int_1^\infty z\, \phi(z)\, \mathrm{d}z}{2(1 - \Phi(1))},$$

ここで ϕ の対称性を用いた。$z\phi(z)$ は $\phi(z)$ の導関数にマイナスをつけたものなので、この積分は容易である：

$$\int_1^\infty z\phi(z)\,\mathrm{d}z = \Big[-\phi(z)\Big]_1^\infty = \phi(1) = \frac{\mathrm{e}^{-1/2}}{\sqrt{2\pi}} = 0.24197\ldots$$

したがって

$$\mathrm{E}(|X_1+X_2|\,|\,R) = \frac{2\phi(1)}{1-\Phi(1)} = \frac{2\times 0.24197}{0.15866} = 3.050\ldots \approx 3.05.$$

検算

- $B=-2.6$ を試すと、方程式は $X^2-5.2X+1=0$ となり、根はちょうど 5 と $\frac{1}{5}$ である。これは (i) の境界の場合である。
- R の上では $2|B|\geq 2$ なので、(ii) の答えは 2 より大きくなければならない。実際そうなっており、値ももっともらしい。1 より先の正規分布の裾は薄いので、$|B|$ はたいてい 1 をわずかに上回る程度である（その条件付き平均は約 1.53）。
- B として標準正規分布に従う値を 2×10^6 個生成したシミュレーション（simulation）では 0.3177、0.4855、3.049 が得られ、0.317、0.485、3.05 と一致した。

よくある間違い

- 条件付けを忘れること。(i) と (ii) では $\mathrm{P}(R)=0.317$ で割らなければならない。条件なしの確率 $\Phi(2.6)-\Phi(1)=0.154$ は答えではない。
- (i) で $B\geq 1$ を含めてしまうこと。このとき両方の根は負なので、どちらも $\frac{1}{5}$ より大きくはない。
- $\mathrm{E}|X_1+X_2|$ を $|\mathrm{E}(X_1+X_2)|$ として計算してしまうこと。後者は対称性により 0 である。

STEP のその先へ

量 $\phi(1)/(1-\Phi(1))$ は**切断**正規分布（truncated normal distribution）の平均であり、比 $\phi(a)/(1-\Phi(a))$ は逆ミルズ比（inverse Mills ratio）として知られている。これは統計学（たとえば打ち切りデータ（censored data）のモデル）や、正規分布のハザード率（hazard rate）として信頼性理論（reliability theory）に現れる。係数がランダムな多項式は平均していくつの実数解をもつか、といったランダム多項式（random polynomials）に関する問いは、確率論の中で意外なほど豊かな分野をなしている。条件付き期待値

と連続分布については、Grimmett と Welsh による *Probability: An Introduction*（確率論入門）、または Blitzstein と Hwang による *Introduction to Probability*（確率論入門）を参照するとよい。

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。