解答例

STEP II 2009、第 12 問

確率・統計 ・ 半正規分布

> **STEP II 2009、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/09-s2-q12

鍵となる考え方

指数 $-2x^2$ は、正規分布（normal distribution）の密度関数（density）が姿を変えたものである。$-2x^2 = -\dfrac{x^2}{2\sigma^2}$ と書けば $\sigma = \frac{1}{2}$ となるので、f は $Y \sim \mathrm{N}(0, \frac{1}{4})$ の密度関数をちょうど 2 倍し、0 で切り落としたものである。言い換えると、X は $|Y|$ と同じ分布に従う。このことに気づけば、正規化定数（normalising constant）、$\mathrm{E}(X^2)$、中央値（median）はいずれも正規分布に関する事実からただちに求まる。積分が必要なのは $\mathrm{E}(X)$ だけであり、それも（ごく易しい）積分である。

解答

以下を通して、$Y \sim \mathrm{N}(0, \frac{1}{4})$、$Z \sim \mathrm{N}(0,1)$ とする。すると Y は $\frac{1}{2}Z$ と同じ分布に従い、その密度関数は

$$g(y) = \frac{1}{\frac{1}{2}\sqrt{2\pi}}\,\mathrm{e}^{-y^2/(2\cdot\frac{1}{4})} = \frac{2}{\sqrt{2\pi}}\,\mathrm{e}^{-2y^2}.$$

である。

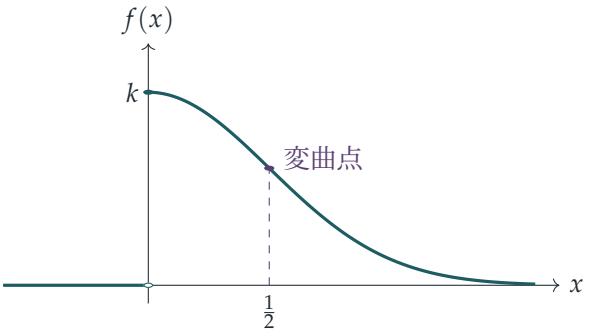
小問 (i)

$x \geq 0$ に対して $f'(x) = -4kx\,\mathrm{e}^{-2x^2} \leq 0$ であるから、f は $x = 0$ における最大値 k から減少し、そこでの傾きは 0 である。また

$$f''(x) = k(16x^2 - 4)\,\mathrm{e}^{-2x^2},$$

は $x = \frac{1}{2}$ で符号が変わるので、$x = \frac{1}{2}$ に変曲点（point of inflection）がある。$x \to \infty$ のとき $f(x) \to 0$ は非常に速く、また $x < 0$ では $f(x) = 0$ である。グラフは釣鐘型曲

線の右半分であり（小問 (ii) より $k \approx 1.60$）、$x = 0$ で 0 から k へ跳ね上がる。



小問 (ii)

全確率は 1 でなければならない。g は 0 に関して対称であり、実数全体での積分が 1 であるから、$[0,\infty)$ での積分は $\frac{1}{2}$ である：

$$\int_0^\infty \frac{2}{\sqrt{2\pi}} \mathrm{e}^{-2x^2}\,\mathrm{d}x = \frac{1}{2} \quad\Longrightarrow\quad \int_0^\infty \mathrm{e}^{-2x^2}\,\mathrm{d}x = \frac{\sqrt{2\pi}}{4}.$$

よって $k \cdot \dfrac{\sqrt{2\pi}}{4} = 1$ であり、

$$k = \frac{4}{\sqrt{2\pi}} = 2\sqrt{\frac{2}{\pi}} \approx 1.596.$$

を得る。$x \geq 0$ に対して $f(x) = 2g(x)$ となることに注意しよう。これにより、X が $|Y|$ と同じ分布に従うことが確かめられる。

小問 (iii)

平均については、因子 x のおかげで積分は初等的に計算できる。$x\,\mathrm{e}^{-2x^2}$ は e^{-2x^2} の導関数の定数倍だからである：

$$\mathrm{E}(X) = k\int_0^\infty x\,\mathrm{e}^{-2x^2}\,\mathrm{d}x = k\Big[-\tfrac{1}{4}\mathrm{e}^{-2x^2}\Big]_0^\infty = \frac{k}{4} = \frac{1}{\sqrt{2\pi}} \approx 0.399.$$

? **考えてみよう** 長い部分積分（integration by parts）をせずに $\displaystyle\int_0^\infty x^2\,\mathrm{e}^{-2x^2}\,\mathrm{d}x$ を求めるにはどうすればよいか？

再び正規分布を使う。関数 $x^2 g(x)$ は偶関数であるから、

$$\mathrm{E}(X^2) = \int_0^\infty x^2 \cdot 2g(x)\,\mathrm{d}x = \int_{-\infty}^\infty x^2 g(x)\,\mathrm{d}x = \mathrm{E}(Y^2) = \mathrm{Var}(Y) = \frac{1}{4},$$

となる。ここで $\mathrm{E}(Y)=0$ を用いた。（$u=x$、$\mathrm{d}v=x\mathrm{e}^{-2x^2}\,\mathrm{d}x$ として部分積分しても同じ結果 $\int_0^\infty x^2\mathrm{e}^{-2x^2}\,\mathrm{d}x=\frac{1}{4}\int_0^\infty \mathrm{e}^{-2x^2}\,\mathrm{d}x=\frac{1}{4k}$ が得られる。）したがって分散（variance）は

$$\mathrm{Var}(X)=\mathrm{E}(X^2)-(\mathrm{E}(X))^2=\frac{1}{4}-\frac{1}{2\pi}=\frac{\pi-2}{4\pi}\approx 0.0908.$$

である。

小問 (iv)

中央値 m は $\mathrm{P}(X\le m)=\frac{1}{2}$ を満たす。X は $|Y|$ と、Y は $\frac{1}{2}Z$ と同じ分布に従うから、標準正規分布の対称性を用いて

$$\mathrm{P}(X\le m)=\mathrm{P}(|Z|\le 2m)=2\Phi(2m)-1,$$

となる。（直接計算すると $\mathrm{P}(X\le m)=\int_0^m 2g(x)\,\mathrm{d}x=2(\mathrm{P}(Y\le m)-\frac{1}{2})$ である。）これを $\frac{1}{2}$ に等しいとおくと

$$\Phi(2m)=\frac{3}{4}.$$

を得る。数表より $\Phi^{-1}(0.75)=0.6745$ であるから、$2m=0.6745$ となり、

$$m\approx 0.337\ \text{（有効数字 3 桁）}.$$

である。

検算

$k=2\sqrt{2/\pi}$ として f を $[0,10]$ 上でシンプソンの公式（Simpson's rule）により数値積分すると、全確率 1.0000、$\mathrm{E}(X)=0.39894$、$\mathrm{Var}(X)=0.090845$ が得られ、$1/\sqrt{2\pi}$ および $(\pi-2)/(4\pi)$ と一致する。$\int_0^m f(x)\,\mathrm{d}x=\frac{1}{2}$ を数値的に解くと $m=0.33724$ となる。標準偏差 $\frac{1}{2}$ の正規乱数 200 000 個について $|Y|$ をシミュレーションしたところ、平均 0.399、分散 0.091 となり、値の 50.0% が 0.337 を下回った。妥当性の確認：中央値 0.337 は平均 0.399 より小さい。右に長い裾をもつ分布ではそうなるはずである。

よくある間違い

- 指数から $\sigma^2=2$ や $\sigma=\frac{1}{4}$ としてしまうこと。$2x^2$ を $x^2/(2\sigma^2)$ と比べれば $\sigma^2=\frac{1}{4}$ となる。
- 因子 2 を忘れること。直線の半分の上では、正規分布の密度関数の積分は $\frac{1}{2}$ にしかならない。
- $\Phi(2m)=\frac{1}{2}$ を解いて $m=0$ としてしまうこと。X の確率のうち m より下にある半分は、Y の確率の中央の半分に対応するので、条件は $\Phi(2m)=\frac{3}{4}$ である。

STEP のその先へ

X は**半正規分布**（*half-normal*）に従う確率変数であり、Y が正規分布に従うときの $|Y|$ は**確率変数の変換**（*transformation of a random variable*）の最も簡単な例である。一般に $X = h(Y)$ ならば、X の密度関数は Y の密度関数をすべての逆像（preimage）にわたって足し合わせることで得られ、ここではそれが因子 2 にあたる。同じ考え方により、Z^2 が自由度（degree of freedom）1 のカイ二乗分布（chi-squared distribution）に従うことがわかり、これが統計学の多くの話題の出発点となる。Grimmett and Welsh, *Probability: An Introduction*（確率論入門）では、確率変数の関数が丁寧に扱われている。

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。