解答例

STEP II 1995、第 14 問

確率・統計 ・ 仮説検定

> **STEP II 1995、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/95-s2-q14

鍵となる考え方

密度関数（density function）を $x^2\phi(x)$ の形に書く。ここで ϕ は標準正規分布（standard normal distribution）の密度関数である。すると、必要な積分はすべて標準正規確率変数 Z のモーメント（moment）になる。$\mathrm{E}(Z^4)$ に関するヒントが効いてくるのはここである。データが「確信を揺るがす」かどうかを判断するには、仮説（hypothesis）（$\mu = 0$）が正しいとしたとき、少なくともこれほど極端な結果がどの程度の確率で得られるかを問う。その確率がきわめて小さく、しかも μ をずらせばデータが何の変哲もないものになるならば、そのデータは仮説に不利な証拠となる。

解答

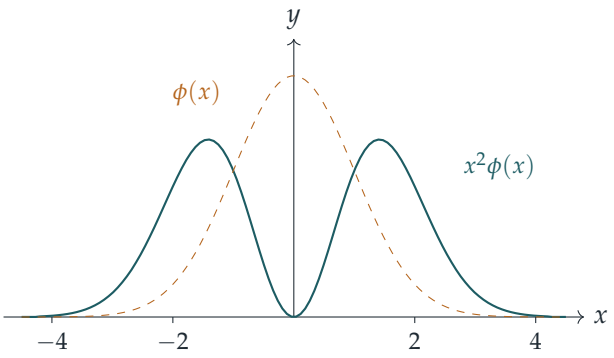
A を求める

$u = x$、$\mathrm{d}v = x\mathrm{e}^{-x^2/2}\,\mathrm{d}x$ とおいて部分積分（integration by parts）を行う。このとき $v = -\mathrm{e}^{-x^2/2}$ である。

$$\int_{-\infty}^{\infty} x^2\mathrm{e}^{-x^2/2}\,\mathrm{d}x = \left[-x\mathrm{e}^{-x^2/2}\right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \mathrm{e}^{-x^2/2}\,\mathrm{d}x = 0 + \sqrt{2\pi},$$

ここで標準正規分布の積分を用いた。確率の総和は 1 でなければならないので、

$$A = \frac{1}{\sqrt{2\pi}}, \qquad \text{すなわち} \qquad \mathrm{f}(x) = x^2\phi(x), \quad \phi(x) = \tfrac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}.$$



この密度関数は対称で、$x = \pm\sqrt{2}$ に二つの山をもち、$x = 0$ で 0 となる。対称性から $\mathrm{E}(X) = 0$ であり、

$$\mathrm{E}(X^2) = \int_{-\infty}^{\infty} x^4\phi(x)\,\mathrm{d}x = \mathrm{E}(Z^4) = 3, \qquad \text{よって} \qquad \mathrm{Var}(X) = 3.$$

すなわち分散（variance）は 3 である。

小問 (i)

仮説のもとで、$a = 87.3$ ほど遠く離れた結果はどの程度起こりやすいだろうか。先ほどと同様に部分積分すると（今度は $\phi'(x) = -x\phi(x)$ より $v = -\phi(x)$ とする）、

$$\mathrm{P}(X \geq a) = \int_a^{\infty} x \cdot x\phi(x)\,\mathrm{d}x = a\phi(a) + \mathrm{P}(Z \geq a).$$

また $a > 0$ のとき、積分範囲では $x/a \geq 1$ であるから、$\mathrm{P}(Z \geq a) = \int_a^{\infty} \phi(x)\,\mathrm{d}x \leq \int_a^{\infty} \frac{x}{a}\phi(x)\,\mathrm{d}x = \frac{\phi(a)}{a}$ が成り立つ。したがって対称性より

$$\mathrm{P}(|X| \geq 87.3) \leq 2\left(a + \frac{1}{a}\right)\phi(a) \approx 70\,\mathrm{e}^{-3810.6} \approx 10^{-1653}.$$

これはばかげているほど小さい。仮説が正しければ、このような値を目にすることは事実上ありえない。一方、μ が 85 や 89 程度であれば（つまり 87.3 が $\mathrm{f}(x - \mu)$ の山の一つに位置するならば）、87.3 という結果はまったくありふれたものである。したがって、87.3 という結果がただ一つ得られただけで、確信は完全に揺らぐはずである。唯一の逃げ道は結果そのもの（たとえば記録の誤り）を疑うことであり、分別のある科学者なら、仮説を捨てる前にまずそれを確認するだろう。

? **考えてみよう** どの特定の値も確率は 0 なのに、一つの観測値から判断するのは不公平ではないだろうか。

問題は 87.3 というちょうどその値の確率ではなく、仮説が確率の重みを置く場所から**少なくともこれだけ離れた**値の確率である。同様に、仮説のもとでの 87.3 における密

度は、適当な対立仮説のもとでのそこでの密度のおよそ 10^{-1650} 倍である。いずれにしても、これほど極端であれば観測値は一つで十分である。

小問 (ii)

$X_1, \ldots, X_{1000}$ を互いに独立な結果とし、$\bar{X}$ をその平均とする。仮説のもとで各 X_i は平均 0、分散 3 をもつので、

$$\mathrm{E}(\bar{X}) = 0, \qquad \mathrm{Var}(\bar{X}) = \frac{3}{1000}, \qquad \mathrm{sd}(\bar{X}) = \sqrt{0.003} \approx 0.0548.$$

分散は有限で、1000 は大きいから、中心極限定理（central limit theorem）により、f 自体は正規分布の密度とは似ても似つかないにもかかわらず、$\bar{X}$ は近似的に $\mathrm{N}(0, 0.003)$ に従う。観測された平均は 0 から

$$\frac{0.23}{0.0548} \approx 4.2$$

標準偏差（standard deviation）分だけ離れており、

$$\mathrm{P}(|\bar{X}| \geq 0.23) \approx \mathrm{P}(|Z| \geq 4.2) \approx 3 \times 10^{-5}.$$

つまり、仮説が正しければ、標本平均が 0 からこれほど離れるのは、実験全体を十万回繰り返して三回程度である。$\mu \approx 0.23$ という対立仮説のもとでは、同じデータはまったく典型的なものになる。したがってこの場合も、(i) ほど圧倒的ではないにせよ、確信は深刻に揺らぐはずである。慣例的などの有意水準（significance level）、例えば 1% や 0.1% でも、$\mu = 0$ は棄却される。

> **考えてみよう** チェビシェフの不等式を使えば、中心極限定理を避けられないだろうか。

チェビシェフ（Chebyshev）の不等式を用いると、近似なしに $\mathrm{P}(|\bar{X}| \geq 0.23) \leq 0.003/0.23^2 \approx 0.057$ が得られる。これは示唆的ではあるが決定的ではない。これほど極端な結果が排除されるのは約 6% の水準までにすぎないからである。中心極限定理を使えばはるかに鋭い値が得られ、裾が $\mathrm{e}^{-x^2/2}$ のように減衰する分布の観測値が 1000 個あるならば、この近似は信頼してよい。

検算

- シンプソンの公式（Simpson's rule）により $\int x^2 \mathrm{e}^{-x^2/2}\,\mathrm{d}x = 2.506628\ldots = \sqrt{2\pi}$、また $\int x^2 \mathrm{f}(x)\,\mathrm{d}x = 3.000$ が得られる。
- 裾の確率の公式 $\mathrm{P}(|X| \geq a) = 2(a\phi(a) + \mathrm{P}(Z \geq a))$ は、$a = 3$ において数値積分の結果と一致する（いずれも 0.029291）。

- $|X|$ は、独立な三つの $\mathrm{N}(0,1)$ 変数からなるベクトルの長さと同じ密度をもつ。この方法で X をシミュレーションし、1000 個の値の平均を 4000 回求めたところ、平均が $2\,\mathrm{sd}$ を超えたのは約 4.9% で、正規分布による予測 4.6% に近く、0.23 を超えたものは一つもなかった。

よくある間違い

- 指数関数の部分が標準正規分布に見えるからといって $\mathrm{Var}(X) = 1$ とすること。余分な x^2 が確率の重みを外側へ押し出すので、分散は 3 になる。
- (ii) で、$\bar{X}$ ではなく単一の X に正規分布による検定を適用すること、あるいは分散を 1000 で割り忘れること。
- (i) で「観測値一つでは何も証明できない」と論じること。これほど極端な結果であれば、証明できるのである。

STEP のその先へ

これは、仮説検定（hypothesis testing）を第一原理から眺めたものである。大学一年次の統計学の講義では、これを p 値（p-value）、尤度比（likelihood ratio）（ネイマン–ピアソンの補題（Neyman–Pearson lemma））、中心極限定理を用いて定式化する。ちなみに $|X|$ は自由度 3 のカイ分布（chi distribution）に従い、これは気体分子の速さのマクスウェル–ボルツマン分布（Maxwell–Boltzmann distribution）にほかならない。Grimmett and Welsh, *Probability: An Introduction*（確率論入門）は、ここで用いた極限定理を扱っている。

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。