解答例

STEP III 2004、第 14 問

確率・統計 ・ 仮説検定の誤り

> **STEP III 2004、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/04-s3-q14

鍵となる考え方

どちらの仮説（hypothesis）のもとでも $\overline{X}$ の中心は同じ μ であり、両者の違いは $\overline{X}$ の散らばり具合だけである。したがって、H_0 に反する証拠となるのは、$\overline{X}$ が μ から左右どちらかの側に**大きく離れた**値をとることであり、これはまさに対称な両側の棄却域（rejection region）が検出するものである。その後は、それぞれの仮説のもとで一度ずつ $\overline{X}$ を標準化（standardise）するだけで、すべてが片づく。

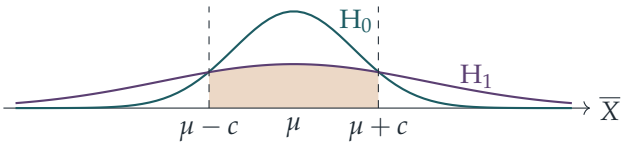
解答

なぜ両側検定なのか

$X \sim \mathrm{N}(\mu, \sigma^2)$ であり標本の値は互いに独立であるから、標本平均（sample mean）は

$$\overline{X} \sim \mathrm{N}\left(\mu,\ \frac{\sigma^2}{n}\right).$$

を満たす。どちらの仮説のもとでも $\overline{X}$ の中心は既知の値 μ であり、唯一の違いは、H_1 のもとではその標準偏差（standard deviation）$\sigma_1/\sqrt{n}$ が $\sigma_0/\sqrt{n}$ より大きいという点である。μ に近い $\overline{X}$ の値は相対的に H_0 のもとで起こりやすく、μ から遠い値は相対的に H_1 のもとで起こりやすい（幅の広い曲線のほうが裾が厚い。図を参照）。どちらの分布も μ に関して対称であるから、$\overline{X}$ が大きな値をとることは、同じだけ小さな値をとることと比べて H_1 を支持する証拠として強いわけではない。よって自然な検定は、$|\overline{X} - \mu|$ がどちらの向きにであれ大きいときに H_0 を棄却するものであり、対称性から受容域（acceptance interval）は μ を中心とすべきである。すなわち、$\mu - c < \overline{X} < \mu + c$ のとき H_0 を受容する。ここで片側検定（one-sided test）を用いるのは誤りである。散らばりが大きいことを示す証拠の半分を無視することになるからである。



網掛けの部分の面積（H_1 が真であるにもかかわらず H_0 を受容する確率）が β であり、破線の外側にあるティール色の曲線の下の面積が α である。

c を求める

H_0 のもとで、$Z = \dfrac{\sqrt{n}\,(\overline{X} - \mu)}{\sigma_0}$ は標準正規分布（standard normal distribution）に従う。$|\overline{X} - \mu| \geq c$、すなわち $|Z| \geq c\sqrt{n}/\sigma_0$ のとき H_1 を採択する。標準正規分布の対称性により、

$$\alpha = \mathrm{P}\left(|Z| \geq \frac{c\sqrt{n}}{\sigma_0}\right) = 2\left(1 - \Phi\left(\frac{c\sqrt{n}}{\sigma_0}\right)\right), \qquad \text{よって} \qquad \Phi\left(\frac{c\sqrt{n}}{\sigma_0}\right) = 1 - \tfrac{1}{2}\alpha.$$

Φ は狭義単調増加であるから、これは $c\sqrt{n}/\sigma_0 = z_\alpha$ を意味し、したがって

$$c = \frac{z_\alpha\,\sigma_0}{\sqrt{n}}.$$

である。

β は n によらない

H_1 のもとで、$W = \dfrac{\sqrt{n}\,(\overline{X} - \mu)}{\sigma_1}$ は標準正規分布に従い、

$$\beta = \mathrm{P}(|\overline{X} - \mu| < c) = \mathrm{P}\left(|W| < \frac{c\sqrt{n}}{\sigma_1}\right) = \mathrm{P}\left(|W| < \frac{z_\alpha\sigma_0}{\sigma_1}\right)$$
$$= 2\,\Phi\left(\frac{z_\alpha\sigma_0}{\sigma_1}\right) - 1,$$

となる。ここで、前の部分で得た $c\sqrt{n} = z_\alpha\sigma_0$ を代入した。c に含まれる $\sqrt{n}$ が標準化で現れる $\sqrt{n}$ と打ち消し合うので、β は α と σ_0/σ_1 のみに依存し、n にはよらない。

考えてみよう 標本が大きいほど、検定はよくなるはずではないのか？

通常はそのとおりである。n が大きくなるにつれて、二つの標本分布が互いに離れていくからである。ところがここでは決して離れない。どちらも中心は μ であり、n を大きくすると**両方の**標準偏差が同じ因子 $\sqrt{n}$ で縮むだけである。図全体が単に拡大縮小されるだけなので、面積 α と β は変わらない。標本平均は分散（variance）を検定するための統計量（statistic）としては不適切であり、個々の値がもつ散らばりについての情報を捨ててしまうのである。

比 σ_1/σ_0 の最小値

$r = \sigma_1/\sigma_0 > 1$ とおく。

条件 $\alpha < 0.05$。これは $\Phi(z_\alpha) = 1 - \frac{1}{2}\alpha > 0.975 \approx \Phi(1.960)$ を意味するので、Φ が増加関数であることから $z_\alpha > 1.960$ である。

条件 $\beta < 0.05$。これは $2\Phi(z_\alpha/r) - 1 < 0.05$、すなわち $\Phi(z_\alpha/r) < 0.525 \approx \Phi(0.063)$ を意味するので、$z_\alpha/r < 0.063$、つまり

$$r > \frac{z_\alpha}{0.063}.$$

である。両者を合わせると、$r > z_\alpha/0.063 > 1.960/0.063$ が必要である。逆に、$r > 1.960/0.063$ ならば、z_α を 1.960 よりわずかに大きく、かつ $z_\alpha < 0.063\,r$ となるように選べるので、両方の条件が成り立つ。よって比は

$$\frac{1.960}{0.063} \approx 31.1,$$

を超えなければならず、σ_1/σ_0 の最小値はおよそ 31 である。標本の大きさにかかわらず、標本平均がこれほど確実に二つを見分けるには、標準偏差がおよそ 30 倍も異なっていなければならないのである。

検算

- $\mu = 3$、$\sigma_0 = 1$、$\sigma_1 = 2.5$、$\alpha = 0.1$、$n = 1, 4, 9, 25$ としてシミュレーションを行ったところ、H_0 のもとでの棄却率はいずれも約 0.10 となり、H_1 のもとでの受容率はどの n でも約 0.489 となって、$2\Phi(1.645/2.5) - 1 \approx 0.489$ と一致した。
- $r = 1.960/0.063$ かつちょうど $\alpha = 0.05$ のとき、$\beta = 2\Phi(0.063) - 1 \approx 0.050$ となり、予想どおり境界上にある。Φ の逆関数の正確な値を用いると $r \approx 31.3$ となるので、答え 31 は与えられた数表の値の丸めの影響を受けない。
- 妥当性の確認：$\sigma_1 = \sigma_0$ ならば $\beta = 2\Phi(z_\alpha) - 1 = 1 - \alpha$ となる。二つの仮説が一致するときには、当然こうならなければならない。

よくある間違い

- $\overline{X} > \mu + c$ のような片側の棄却域を用いること。H_1 のもとでも平均は動かないので、一方の裾をひいきする理由はない。
- 棄却域が両側に裾をもつことを忘れて、$\alpha = 1 - \Phi(c\sqrt{n}/\sigma_0)$ と書いてしまうこと。
- r についての不等式の向きを逆にしてしまうこと。β を小さくするには z_α/r が小さくなければならず、したがって r は大きくなければならない。

STEP のその先へ

この問題は、二種類の誤りの間のトレードオフと、検定統計量の選び方がなぜ重要なのかを示している。ネイマン–ピアソンの補題（Neyman–Pearson lemma）によれば、最も検出力の高い検定は二つの仮説のもとでの尤度（likelihood）を比較するものである。ここではそれにより統計量 $\sum(X_i - \mu)^2$ が導かれ、その χ^2_n 分布は、$\overline{X}$ とは異なり、n が大きくなるにつれて実際に二つの仮説を分離する。これらの考え方への読みやすい入門書としては、Blitzstein と Hwang の *Introduction to Probability*（確率論入門）、あるいは Grimmett と Welsh の *Probability: An Introduction*（確率論への招待）を参照されたい。

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。