풀이

STEP III 2004, 문제 14

확률과 통계 • 가설 검정의 오류

STEP III 2004, 14번 문제

더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:

step.maths.org/questions/04-s3-q14

핵심 아이디어

두 가설(hypothesis) 모두에서 $\overline{X}$의 중심은 같은 값 μ이고, 다른 것은 $\overline{X}$가 얼마나 퍼져 있는가뿐이다. 따라서 H_0에 반하는 증거는 $\overline{X}$가 어느 쪽으로든 μ에서 **멀리 떨어진** 곳에 놓이는 것이며, 대칭인 양측(two-sided) 기각역(rejection region)이 바로 이것을 잡아낸다. 그 다음에는 모든 것이 각 가설 아래에서 한 번씩 $\overline{X}$를 표준화(standardising)하는 일로 귀결된다.

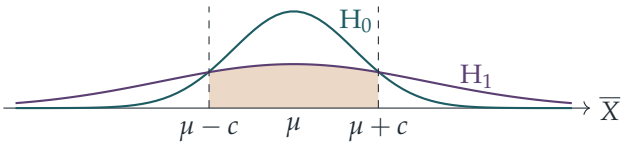
풀이

왜 양측검정인가

$X \sim \mathrm{N}(\mu, \sigma^2)$이고 표본값들이 서로 독립이므로, 표본평균(sample mean)은

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

을 만족한다. 어느 가설 아래에서든 $\overline{X}$의 중심은 알려진 값 μ이며, 유일한 차이는 H_1 아래에서 표준편차(standard deviation) $\sigma_1/\sqrt{n}$이 $\sigma_0/\sqrt{n}$보다 크다는 것이다. μ에 가까운 $\overline{X}$의 값은 상대적으로 H_0 아래에서 더 나오기 쉽고, μ에서 먼 값은 상대적으로 H_1 아래에서 더 나오기 쉽다(더 넓은 곡선의 꼬리가 더 두껍다. 그림을 보라). 두 분포가 모두 μ에 대하여 대칭이므로, $\overline{X}$의 큰 값이 같은 정도로 작은 값보다 H_1에 대한 더 강한 증거가 되지는 않는다. 그러므로 자연스러운 검정은 $|\overline{X} - \mu|$가 어느 방향으로든 클 때 H_0을 기각하는 것이고, 대칭성에 의해 채택구간(acceptance interval)은 μ를 중심으로 해야 한다. 즉 $\mu - c < \overline{X} < \mu + c$일 때 H_0을 채택한다. 단측검정(one-sided test)은 여기서 틀린 방법인데, 더 큰 퍼짐에 대한 증거의 절반을 무시하게 되기 때문이다.



색칠한 부분의 넓이(H_1이 참인데도 H_0을 채택하는 경우)가 β이고, 점선 바깥쪽 청록색 곡선 아래의 넓이가 α이다.

c 구하기

H_0 아래에서 $Z = \dfrac{\sqrt{n}\,(\overline{X} - \mu)}{\sigma_0}$는 표준정규분포(standard normal distribution)를 따른다. $|\overline{X} - \mu| \geq c$일 때, 즉 $|Z| \geq c\sqrt{n}/\sigma_0$일 때 H_1을 채택한다. 표준정규분포의 대칭성에 의해

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

Φ는 순증가함수(strictly increasing function)이므로 이는 $c\sqrt{n}/\sigma_0 = z_\alpha$를 뜻하고, 따라서

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

이다.

β는 n에 의존하지 않는다

H_1 아래에서 $W = \dfrac{\sqrt{n}\,(\overline{X} - \mu)}{\sigma_1}$는 표준정규분포를 따르고,

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

이다. 여기서 앞 단계에서 얻은 $c\sqrt{n} = z_\alpha\sigma_0$를 대입하였다. c에 들어 있는 $\sqrt{n}$이 표준화에서 나온 $\sqrt{n}$과 약분되므로, β는 n이 아니라 α와 σ_0/σ_1에만 의존한다.

? 생각해 보기 표본이 클수록 검정이 더 좋아져야 하지 않을까?

보통은 그렇다. n이 커지면서 두 표본분포(sampling distribution)가 서로 멀어지기 때문이다. 그러나 여기서는 결코 멀어지지 않는다. 두 분포 모두 중심이 μ이고, n을 늘리면 두 표준편차가 똑같이 $\sqrt{n}$배만큼 줄어든다. 그림 전체가 단순히 축척만 바뀌므로 넓이 α와 β는 변하지 않는다. 표본평균은 분산을 검정하기에는 나쁜 통계량이다. 개별 값들이 담고 있는 퍼짐에 관한 정보를 버리기 때문이다.

비 σ_1/σ_0의 최솟값

$r = \sigma_1/\sigma_0 > 1$이라 하자.

조건 $\alpha < 0.05$. 이는 $\Phi(z_\alpha) = 1 - \frac{1}{2}\alpha > 0.975 \approx \Phi(1.960)$을 뜻하므로, Φ가 증가함수이므로 $z_\alpha > 1.960$이다.

조건 $\beta < 0.05$. 이는 $2\Phi(z_\alpha/r) - 1 < 0.05$, 즉 $\Phi(z_\alpha/r) < 0.525 \approx \Phi(0.063)$을 뜻하므로 $z_\alpha/r < 0.063$, 곧

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

이다. 둘을 합치면 $r > z_\alpha/0.063 > 1.960/0.063$이어야 한다. 역으로 $r > 1.960/0.063$이면 1.960보다 조금 크면서 $z_\alpha < 0.063\,r$인 z_α를 고를 수 있고, 그러면 두 조건이 모두 성립한다. 따라서 비는

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

보다 커야 하므로, σ_1/σ_0의 최솟값은 약 31이다. 표본 크기가 얼마이든, 표본평균이 이 정도로 믿을 만하게 두 표준편차를 구별하려면 표준편차가 약 서른 배는 차이 나야 한다.

검산

- $\mu = 3, \sigma_0 = 1, \sigma_1 = 2.5, \alpha = 0.1, n = 1, 4, 9, 25$로 모의실험(simulation)을 해 보면, H_0 아래에서의 기각률은 매번 약 0.10이었고, H_1 아래에서의 채택률은 모든 n에 대하여 약 0.489로 $2\Phi(1.645/2.5) - 1 \approx 0.489$와 일치하였다.
- $r = 1.960/0.063$이고 정확히 $\alpha = 0.05$이면 $\beta = 2\Phi(0.063) - 1 \approx 0.050$으로, 예상대로 경계에 놓인다. Φ의 정확한 역함수 값을 쓰면 $r \approx 31.3$이므로, 답 31은 주어진 표 값의 반올림에 영향을 받지 않는다.
- 타당성 확인: $\sigma_1 = \sigma_0$이면 $\beta = 2\Phi(z_\alpha) - 1 = 1 - \alpha$인데, 두 가설이 일치할 때 당연히 그래야 한다.

흔한 실수

- $\overline{X} > \mu + c$와 같은 단측 기각역을 쓰는 것: H_1 아래에서 평균은 움직이지 않으므로 한쪽 꼬리를 더 중시할 이유가 없다.
- 기각역에 꼬리가 두 개라는 것을 잊고 $\alpha = 1 - \Phi(c\sqrt{n}/\sigma_0)$라고 쓰는 것.
- r에 대한 부등식의 방향을 거꾸로 잡는 것: β가 작으려면 z_α/r가 **작아야** 하므로 r가 커야 한다.

STEP 너머

이 문제는 두 종류의 오류(error) 사이의 상충 관계와 검정통계량(test statistic)의 선택이 왜 중요한지를 보여 준다. 네이만-피어슨 보조정리(Neyman–Pearson lemma)에 따르면 가장 강력한 검정은 두 가설 아래에서의 가능도(likelihood)를 비교하는 검정이다. 여기서 그것은 통계량 $\sum(X_i - \mu)^2$로 이어지는데, 이 통계량이 따르는 χ^2_n 분포는 $\overline{X}$와 달리 n이 커질수록 실제로 두 가설을 갈라놓는다. 이러한 개념을 읽기 쉽게 소개한 책으로는 Blitzstein과 Hwang의 *Introduction to Probability*(확률론 입문), 또는 Grimmett과 Welsh의 *Probability: An Introduction*(확률: 입문)을 보라.

Beyond STEP 풀이: https://beyondstep.keithmcnulty.org/ko/. 풀이 글은 CC0 1.0이다. 문제 자체의 저작권은 © Cambridge University Press & Assessment에 있으며 여기에 싣지 않는다. 영어 원문을 기계 번역하였으며 아직 검토되지 않았다.