풀이

# STEP II 1995, 문제 14

확률과 통계 • 가설 검정

> **STEP II 1995, 14번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/95-s2-q14

## 핵심 아이디어

밀도함수(density)를 $x^2\phi(x)$로 쓴다. 여기서 $\phi$는 표준정규분포(standard normal distribution)의 밀도함수이다. 그러면 필요한 모든 적분이 표준정규확률변수 $Z$의 적률(moment)이 되는데, 바로 여기서 $\mathrm{E}(Z^4)$에 관한 힌트가 쓰인다. 데이터가 "믿음을 흔드는지" 판단하려면, 가설(hypothesis) $(\mu = 0)$이 참일 때 적어도 이만큼 극단적인 결과가 나올 확률이 얼마인지 묻는다. 그 확률이 아주 작고, $\mu$를 옮기면 데이터가 평범해진다면, 그 데이터는 가설에 불리한 증거가 된다.

## 풀이

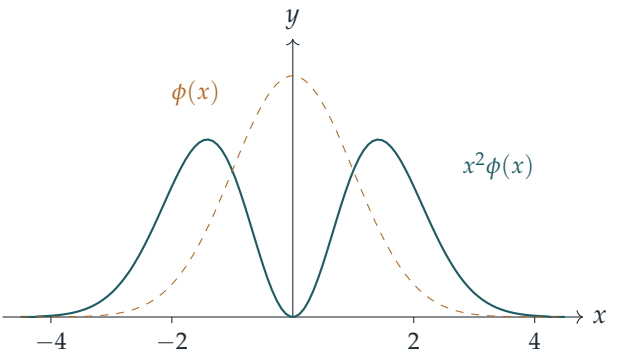
### $A$ 구하기

$u = x$, $\mathrm{d}v = x\mathrm{e}^{-x^2/2}\,\mathrm{d}x$로 두고 부분적분(integration by parts)하면 $v = -\mathrm{e}^{-x^2/2}$이므로

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

이다. 여기서 표준정규분포의 적분을 이용하였다. 전체 확률은 1이어야 하므로

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




이 밀도함수는 대칭이고 $x = \pm\sqrt{2}$에서 두 개의 봉우리를 가지며 $x = 0$에서 0이다. 대칭성에 의해 $\mathrm{E}(X) = 0$이고,

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

**소문항 (i)**

가설 아래에서 $a = 87.3$만큼 멀리 떨어진 결과가 나올 가능성은 얼마일까? 앞에서처럼 부분적분하면 (이번에는 $\phi'(x) = -x\phi(x)$이므로 $v = -\phi(x)$로 둔다)

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

또한 $a > 0$일 때 적분 구간에서 $x/a \geq 1$이므로 $\mathrm{P}(Z \geq a) = \int_a^{\infty} \phi(x)\,\mathrm{d}x \leq \int_a^{\infty} \frac{x}{a}\phi(x)\,\mathrm{d}x = \frac{\phi(a)}{a}$이다. 따라서 대칭성에 의해

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

이는 터무니없이 작은 값이다. 가설이 참이라면 이런 값은 사실상 절대 관찰되지 않아야 한다. 반면 $\mu$가 85나 89 근처라면 (87.3이 $\mathrm{f}(x - \mu)$의 봉우리 가운데 하나에 놓이게 되므로) 87.3이라는 결과는 지극히 평범하다. 따라서 87.3이라는 결과 하나만으로도 믿음은 완전히 흔들려야 한다. 유일한 탈출구는 결과 자체를 의심하는 것(예를 들어 기록 오류)이며, 분별 있는 과학자라면 가설을 버리기 전에 이를 먼저 확인할 것이다.

> **생각해 보기** 모든 특정한 값의 확률이 0인데, 관찰값 하나로 판단하는 것은 불공평하지 않은가?

요점은 정확히 87.3이라는 값의 확률이 아니라, 가설이 확률을 몰아 두는 곳에서 **적어도** 이만큼 멀리 떨어진 값의 확률이다. 마찬가지로, 가설 아래에서 87.3에서의 밀도는 적절한 대립가설 아래에서의 밀도의 대략 $10^{-1650}$배이다. 어느 쪽으로 보든, 이렇게 극단적인 경우에는 관찰값 하나로 충분하다.

### 소문항 (ii)

$X_1, \ldots, X_{1000}$을 서로 독립인 결과들이라 하고 $\bar{X}$를 그 평균이라 하자. 가설 아래에서 각 $X_i$의 평균은 0, 분산은 3이므로

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

분산이 유한하고 1000이 충분히 크므로, f 자체는 정규분포의 밀도함수와 전혀 닮지 않았는데도 중심극한정리(central limit theorem)에 의해 $\bar{X}$는 근사적으로 $\mathrm{N}(0, 0.003)$을 따른다. 관찰된 평균은 0에서

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

표준편차만큼 떨어져 있고,

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

따라서 가설이 참이라면, 실험 전체를 십만 번 반복할 때 0에서 이만큼 떨어진 표본평균은 세 번 정도 나타난다. $\mu \approx 0.23$인 대립가설 아래에서는 같은 데이터가 지극히 전형적이다. 이 경우에도 믿음은 심각하게 흔들려야 하지만, (i)만큼 압도적이지는 않다. 통상적인 어떤 유의수준(significance level), 예를 들어 1%나 심지어 0.1%에서도 $\mu = 0$을 기각할 것이다.

> ? **생각해 보기** 체비쇼프 부등식(Chebyshev's inequality)을 써서 중심극한정리를 피할 수는 없을까?

체비쇼프 부등식을 쓰면 아무런 근사 없이 $\mathrm{P}(|\bar{X}| \geq 0.23) \leq 0.003/0.23^2 \approx 0.057$을 얻는다. 이는 시사적이지만 결정적이지는 않은데, 이렇게 극단적인 결과가 약 6% 수준까지만 배제되기 때문이다. 중심극한정리는 훨씬 더 날카로운 수치를 주며, 꼬리가 $\mathrm{e}^{-x^2/2}$처럼 감소하는 분포에서 1000개의 관찰값을 얻었으므로 이 근사는 믿을 만하다.

## 검산

- 심프슨 공식(Simpson's rule)으로 계산하면 $\int x^2 \mathrm{e}^{-x^2/2}\,\mathrm{d}x = 2.506628\ldots = \sqrt{2\pi}$이고, $\int x^2 \mathrm{f}(x)\,\mathrm{d}x = 3.000$이다.
- 꼬리 확률 공식 $\mathrm{P}(|X| \geq a) = 2(a\phi(a) + \mathrm{P}(Z \geq a))$는 $a = 3$에서 수치적분 결과와 일치한다 (둘 다 0.029291).
- $|X|$는 서로 독립인 세 개의 $\mathrm{N}(0,1)$ 변수로 이루어진 벡터의 길이와 같은 밀도함수를 가진다. 이 방법으로 $X$를 모의실험하여 1000개 값의 평균을 내는 일을 4000번 반복했더니, 평균의 약 4.9%가 2 sd 밖에 있었는데 이는 정규분포의 예측값 4.6%에 가깝고, 0.23 밖에 있는 평균은 하나도 없었다.

## 흔한 실수

- 지수함수 부분이 표준정규분포처럼 생겼다고 해서 $\mathrm{Var}(X) = 1$로 두는 것. 추가된 $x^2$이 확률을 바깥쪽으로 밀어내므로 분산은 3이 된다.
- (ii)에서 정규분포 검정을 $\bar{X}$가 아니라 하나의 $X$에 적용하거나, 분산을 1000으로 나누는 것을 잊는 것.
- (i)에서 "관찰값 하나로는 아무것도 증명할 수 없다"고 주장하는 것. 결과가 이렇게 극단적이면 증명할 수 있다.

## STEP 너머

이것은 가설검정(hypothesis testing)을 기본 원리에서부터 바라본 것이다. 대학 1학년 통계학 강의에서는 이를 $p$-값($p$-value), 우도비(likelihood ratio) (네이만–피어슨 보조정리(Neyman–Pearson lemma)), 그리고 중심극한정리로 형식화한다. 덧붙여, $|X|$는 자유도(degrees of freedom)가 3인 카이 분포(chi distribution)를 따르는데, 이는 기체 분자의 속력에 대한 맥스웰–볼츠만 분포(Maxwell–Boltzmann distribution)이다. Grimmett and Welsh, *Probability: An Introduction*(확률론 입문)에서 여기서 사용한 극한 정리들을 다룬다.

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.