풀이

# STEP II 2019, 문제 12

확률과 통계 • 연속 분포와 분위수

> **STEP II 2019, 12번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 OCR에서 내려받을 수 있는 2019년 STEP 시험지 묶음(zip 파일) 가운데 STEP II 시험지에 있다:
>
> www.ocr.org.uk/Images/696590-2019-past-papers.zip

## 핵심 아이디어

여기서 누적분포함수(cumulative distribution function)는 $[0,1]$에서 단순히 $x^n$이므로, 모든 분위수(quantile)를 정확히 적을 수 있다. 분포의 비율 $p$가 그 아래에 놓이는 값은 $p^{1/n}$이다. 그러면 평균(mean)과 분위수의 비교는 거듭제곱의 비교가 되고, 양변을 $n$제곱하면 이는 $\left(1+\frac{1}{n}\right)^n$에 관한 명제로 바뀐다. 이것이 바로 이항전개(binomial expansion)로 어림하기에 알맞은 양이다.

## 풀이

### 소문항 (i)

확률밀도함수(density)를 곱하여 적분하면

$$\mu = \mathrm{E}(X) = \int_0^1 x \cdot nx^{n-1}\,\mathrm{d}x = \int_0^1 nx^n\,\mathrm{d}x = \frac{n}{n+1},$$
$$\mathrm{E}(X^2) = \int_0^1 nx^{n+1}\,\mathrm{d}x = \frac{n}{n+2}.$$

따라서

$$\sigma^2 = \mathrm{E}(X^2) - \mu^2 = \frac{n}{n+2} - \frac{n^2}{(n+1)^2} = \frac{n\left[(n+1)^2 - n(n+2)\right]}{(n+1)^2(n+2)}$$
$$= \frac{n\left[n^2+2n+1-n^2-2n\right]}{(n+1)^2(n+2)} = \frac{n}{(n+1)^2(n+2)},$$

이므로 원하는 결과를 얻는다.

**소문항 (ii)**

$0 \le x \le 1$에서 분포함수(distribution function)는

$$\mathrm{F}(x) = \mathrm{P}(X \le x) = \int_0^x nt^{n-1}\,\mathrm{d}t = x^n.$$

$n = 2$이면 $\mathrm{F}(x) = x^2$이다. 하위 사분위수(lower quartile) $q_1$은 $q_1^2 = \frac{1}{4}$를 만족하고, 상위 사분위수(upper quartile) $q_3$은 $q_3^2 = \frac{3}{4}$를 만족한다. 둘 다 $[0,1]$에 속하므로 양의 제곱근을 택한다.

$$q_1 = \tfrac{1}{2}, \qquad q_3 = \tfrac{\sqrt{3}}{2}, \qquad q_3 - q_1 = \frac{\sqrt{3}-1}{2}.$$

소문항 (i)에 의해 $\sigma^2 = \dfrac{2}{9\cdot 4} = \dfrac{1}{18}$이므로 $\sigma = \dfrac{1}{3\sqrt{2}} = \dfrac{\sqrt{2}}{6}$이고 $2\sigma = \dfrac{\sqrt{2}}{3}$이다.

$\dfrac{\sqrt{3}-1}{2} < \dfrac{\sqrt{2}}{3}$임을 보이고자 한다. 6을 곱하면 이는 다음과 동치(equivalent)이다.

$$3\sqrt{3} - 3 < 2\sqrt{2} \iff 3\sqrt{3} < 3 + 2\sqrt{2}.$$

마지막 부등식의 양변은 모두 양수이므로, 이는 양변의 제곱 사이의 부등식과 동치이다.

$$27 < 9 + 12\sqrt{2} + 8 = 17 + 12\sqrt{2} \iff 10 < 12\sqrt{2},$$

그런데 $\sqrt{2} > 1$이므로 $12\sqrt{2} > 12 > 10$이 되어 이 부등식은 참이다. 모든 단계가 동치였으므로 이 사슬을 거슬러 올라가면 사분위범위(interquartile range)는 $2\sigma$보다 작다.

> **생각해 보기** 여기서는 양변을 제곱해도 안전한데, 일반적으로는 왜 그렇지 않은가?

제곱이 부등식 $a < b$를 보존하는 것은 $a$와 $b$가 음이 아닐 때뿐이다(예를 들어 $-3 < 2$이지만 $9 > 4$이다). 그래서 빨셈이 남아 있는 $3\sqrt{3} - 3 < 2\sqrt{2}$를 바로 제곱하지 않고, 먼저 양변이 분명히 양수인 $3\sqrt{3} < 3 + 2\sqrt{2}$ 꼴로 정리한 것이다.

**소문항 (iii)**

이항정리(binomial theorem)에 의해

$$(1+x)^n = 1 + nx + \frac{n(n-1)}{2}\,x^2 + \cdots,$$

이고 $(k+1)$번째 항은

$$\binom{n}{k} x^k = \frac{n(n-1)\cdots(n-k+1)}{k!}\,x^k.$$

**중앙값과 하위 사분위수.** $\mathrm{F}(x) = x^n$이므로 중앙값(median) $m$은 $m^n = \frac{1}{2}$를, 하위 사분위수 $q$는 $q^n = \frac{1}{4}$를 만족한다. 따라서

$$m = \left(\tfrac{1}{2}\right)^{1/n}, \qquad q = \left(\tfrac{1}{4}\right)^{1/n}.$$

$\mu$, $m$, $q$는 모두 양수이고 $t \mapsto t^n$은 $t > 0$에서 순증가하므로, 이들의 $n$제곱을 비교하여 대소를 비교할 수 있다. 또한

$$\mu^n = \left(\frac{n}{n+1}\right)^n = \frac{1}{\left(1+\frac{1}{n}\right)^n}.$$

$A = \left(1+\frac{1}{n}\right)^n$이라 하자. 그러면

$$\mu < m \iff \mu^n < \tfrac{1}{2} \iff A > 2, \qquad \mu > q \iff \mu^n > \tfrac{1}{4} \iff A < 4.$$

**$\mu$는 중앙값보다 작다.** 전개식에 $x = \frac{1}{n}$을 대입한다. 모든 항이 음이 아니므로 $A$는 처음 세 항의 합 이상이다.

$$A \geq 1 + n \cdot \frac{1}{n} + \frac{n(n-1)}{2} \cdot \frac{1}{n^2} = 2 + \frac{n-1}{2n} > 2,$$

여기서 마지막 부등식은 $n > 1$이므로 등호 없이 성립한다. ($n \geq 2$이면 셋째 항이 실제로 전개식에 나타난다.) 따라서 $\mu < m$이다.

**$\mu$는 하위 사분위수보다 크다.** $x = \frac{1}{n}$일 때 $(k+1)$번째 항은

$$\binom{n}{k}\frac{1}{n^k} = \frac{1}{k!} \cdot \frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-k+1}{n} \leq \frac{1}{k!},$$

인데, $k$개의 분수 $\frac{n-j}{n}$ $(0 \leq j \leq k-1)$이 각각 0과 1 사이에 있기 때문이다. ($k = 0$이면 항은 $1 = \frac{1}{0!}$이다.) $k = 0, 1, \ldots, n$에 대해 더하면

$$A \leq 1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} < 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots < 4,$$

가운데 부등식은 무한급수(infinite series)에 양의 항이 더 있기 때문에 성립하고, 마지막 부등식은 가정해도 된다고 주어진 결과이다. 따라서 $A < 4$이고, 그러므로 $\mu > q$이다.

두 결과를 합치면, 모든 정수 $n > 1$에 대해 하위 사분위수는 $\mu$보다 작고, $\mu$는 중앙값보다 작다.

**?** **생각해 보기** $n$제곱을 비교하는 것이 왜 $\mu$와 $m$의 대소를 결정해 주는가?

양수 $a$와 $b$에 대해 $t^n$은 $t > 0$에서 순증가하므로, $a < b$일 필요충분조건은 $a^n < b^n$이다. 세 수가 모두 양수이므로 $n$제곱을 해도 잃는 것이 없고, 그렇게 하면 $m$과 $q$에 있는 번거로운 $n$제곱근이 사라진다.

## 검산

- $n = 2$일 때: $\mu = \frac{2}{3}$이고 $\sigma^2 = \frac{2}{9 \cdot 4} = \frac{1}{18}$이며, 이는 $\mathrm{E}(X^2) - \mu^2 = \frac{1}{2} - \frac{4}{9}$와 일치한다.
- 수치로 보면, $n = 2$일 때 사분위범위는 $0.366$이고 $2\sigma = 0.471$이다.
- $n = 2$일 때 세 값은 $q = 0.5$, $\mu \approx 0.667$, $m \approx 0.707$이고, $n = 10$일 때는 $q \approx 0.871$, $\mu \approx 0.909$, $m \approx 0.933$이다. 순서 $q < \mu < m$은 2부터 2000까지의 모든 $n$에 대해 컴퓨터로 확인하였다.
- $n \to \infty$일 때 $A = \left(1 + \frac{1}{n}\right)^n \to \mathrm{e} \approx 2.718$이며, 이는 2와 4 사이에 여유 있게 들어간다.

## 흔한 실수

- 양변이 양수인지 먼저 확인하지 않고 $3\sqrt{3} - 3 < 2\sqrt{2}$를 바로 제곱하는 것.
- 역수(reciprocal)를 취할 때 부등호 방향을 틀리는 것: $\mu^n < \frac{1}{2}$는 $A < 2$가 아니라 $A > 2$에 대응한다.
- $(k+1)$번째 항이 $\frac{1}{k!}$ 이하라고 하지 않고 $\frac{1}{k!}$와 같다고 주장하는 것.

## STEP 너머

양 $\left(1 + \frac{1}{n}\right)^n$은 $n$이 커짐에 따라 증가하며 극한 e에 수렴(converge)한다. 여러분이 $\sum 1/k!$와 항별로 비교하여 증명한 부등식은, 해석학(analysis) 입문 과목의 표준 결과인 $\mathrm{e} = \sum_{k \geq 0} 1/k!$를 보이는 첫 단계이다. 이 분포 자체는 $\mathrm{Beta}(n, 1)$ 분포, 즉 베타 분포(Beta distribution)이다. 또한 $\mathrm{P}(\max \leq x) = x^n$이므로, 이는 $[0, 1]$ 위의 독립인 균등(uniform) 확률변수 $n$개의 최댓값의 분포이기도 하다. 두 갈래에 대해 더 알고 싶다면 Blitzstein과 Hwang의 *Introduction to Probability*(확률론 입문)와 Abbott의 *Understanding Analysis*(해석학 이해)를 참고하라.

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.