풀이

STEP II 2022, 문제 12

확률과 통계 • 평균, 중앙값, 최빈값

STEP II 2022, 12번 문제

더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 OCR에서 내려받을 수 있는 2022년 STEP 시험지 묶음(zip 파일) 가운데 STEP II 시험지에 있다:

www.ocr.org.uk/Images/696587-2022-past-papers.zip

핵심 아이디어

모든 것은 누적분포함수(cumulative distribution function) F에서 나오며, 여기서 이 함수는 간단한 다항식이다. F는 $[0,1]$에서 증가하므로, 수 a가 중앙값(median)보다 작을 필요충분조건은 $\mathrm{F}(a) < \frac{1}{2}$이다. 따라서 평균(mean)이나 최빈값(mode)을 중앙값과 비교할 때 중앙값 자체를 구할 필요가 전혀 없다. 평균이나 최빈값에서 F의 값을 계산하여 $\frac{1}{2}$과 비교하기만 하면 된다.

풀이

소문항 (i)

전체 확률이 1이므로

$$1 = k\int_0^1 \left(x^n - x^{n+1}\right)\mathrm{d}x = k\left(\frac{1}{n+1} - \frac{1}{n+2}\right) = \frac{k}{(n+1)(n+2)},$$

이고, 따라서 요구한 대로 $k = (n+1)(n+2)$이다. 그러면

$$\mu = k\int_0^1 \left(x^{n+1} - x^{n+2}\right)\mathrm{d}x = (n+1)(n+2)\cdot\frac{1}{(n+2)(n+3)} = \frac{n+1}{n+3}.$$

소문항 (ii)

$0 \le x \le 1$일 때

$$\mathrm{F}(x) = (n+1)(n+2)\left(\frac{x^{n+1}}{n+1} - \frac{x^{n+2}}{n+2}\right) = x^{n+1}((n+2) - (n+1)x).$$

$(0,1)$에서 $\mathrm{f} > 0$이므로 F는 그 구간에서 순증가하고, 중앙값 m은 $\mathrm{F}(m) = \frac{1}{2}$을 만족하는 유일한 점이다. 따라서

$$\mu < m \iff \mathrm{F}(\mu) < \tfrac{1}{2}.$$

이제 $\mu = \frac{n+1}{n+3}$이면

$$(n+2) - (n+1)\mu = \frac{(n+2)(n+3) - (n+1)^2}{n+3} = \frac{3n+5}{n+3},$$

이므로

$$\mathrm{F}(\mu) = \left(\frac{n+1}{n+3}\right)^{n+1} \frac{3n+5}{n+3}.$$

그러므로 조건 $\mathrm{F}(\mu) < \frac{1}{2}$은 양수 $2\left(\frac{n+3}{n+1}\right)^{n+1}$을 곱하면 다음과 같이 된다.

$$\frac{2(3n+5)}{n+3} < \left(\frac{n+3}{n+1}\right)^{n+1}.$$

$\frac{2(3n+5)}{n+3} = \frac{6(n+3)-8}{n+3} = 6 - \frac{8}{n+3}$이고 $\frac{n+3}{n+1} = 1 + \frac{2}{n+1}$이므로, 이것은 바로 주어진 부등식(inequality)이다. 따라서 그 부등식이 성립하면 언제나 μ는 중앙값보다 작다.

이 부등식을 증명하기 위해 $N = n+1$로 놓으면 $N \geq 3$인 정수이다. $\left(1 + \frac{2}{N}\right)^N$의 이항전개(binomial expansion)의 모든 항은 양수이고 항은 $N+1 \geq 4$개이므로, 그 합은 처음 네 항의 합 이상이다.

$$\begin{aligned}\left(1+\frac{2}{N}\right)^N &\geq 1 + N\cdot\frac{2}{N} + \frac{N(N-1)}{2}\cdot\frac{4}{N^2} + \frac{N(N-1)(N-2)}{6}\cdot\frac{8}{N^3} \\ &= 3 + 2\left(1 - \frac{1}{N}\right) + \frac{4}{3}\left(1 - \frac{3}{N} + \frac{2}{N^2}\right) \\ &= \frac{19}{3} - \frac{6}{N} + \frac{8}{3N^2}.\end{aligned}$$

생각해 보기 네 항의 합은 하계(lower bound)일 뿐이다. 이것이 $6 - \frac{8}{N+2}$보다 크다는 것을 어떻게 보일까?

두 식을 빼고 양수인 분모 $3N^2(N+2)$를 곱하여 없앤다.

$$\begin{aligned}&3N^2(N+2)\left[\frac{19}{3} - \frac{6}{N} + \frac{8}{3N^2} - 6 + \frac{8}{N+2}\right] \\ &\quad = N^2(N+2) - 18N(N+2) + 8(N+2) + 24N^2 \\ &\quad = N^3 + 8N^2 - 28N + 16.\end{aligned}$$

$N \geq 3$이면 $N^2 + 8N - 28 \geq 9 + 24 - 28 = 5 > 0$이므로 $N^3 + 8N^2 - 28N + 16 = N(N^2 + 8N - 28) + 16 > 0$이다. 따라서

$$6 - \frac{8}{n+3} < \frac{19}{3} - \frac{6}{N} + \frac{8}{3N^2} \leq \left(1 + \frac{2}{n+1}\right)^{n+1},$$

즉 주어진 부등식은 모든 정수 $n > 1$에 대하여 성립하고, 따라서 X의 중앙값은 μ보다 크다.

소문항 (iii)

$(0,1)$에서

$$f'(x) = k\left(nx^{n-1} - (n+1)x^n\right) = kx^{n-1}(n-(n+1)x),$$

이고, 이는 $x < \frac{n}{n+1}$일 때 양수, $x > \frac{n}{n+1}$일 때 음수이다. 따라서 f는 증가하다가 감소하며, 최빈값은 $M = \frac{n}{n+1}$이다.

소문항 (ii)에서와 마찬가지로, $M > m$일 필요충분조건은 $F(M) > \frac{1}{2}$이다. $(n+2) - (n+1)M = (n+2) - n = 2$이므로

$$F(M) = 2\left(\frac{n}{n+1}\right)^{n+1},$$

이고, 따라서 $\left(\frac{n}{n+1}\right)^{n+1} > \frac{1}{4}$, 즉 다음이 필요하다.

$$\left(1+\frac{1}{n}\right)^{n+1} < 4.$$

> ? **생각해 보기** n이 2 이상의 어떤 값이든 될 수 있을 때, $\left(1+\frac{1}{x}\right)^{x+1}$에 대해 주어진 사실이 어떻게 도움이 될까?

이 함수는 양의 x에서 감소하므로, 임의의 정수 $n \geq 2$에서의 값은 $x = 2$에서의 값 이하이다.

$$\left(1+\frac{1}{n}\right)^{n+1} \leq \left(\frac{3}{2}\right)^3 = \frac{27}{8} < 4.$$

따라서 $F(M) > \frac{1}{2}$이고, X의 최빈값은 중앙값보다 크다.

각 소문항의 결과를 종합하면, 모든 $n > 1$에 대하여 평균 < 중앙값 < 최빈값이다.

검산

- $n = 2$로 놓자. 평균은 $\frac{3}{5} = 0.6$이고 최빈값은 $\frac{2}{3} \approx 0.667$이다. $4x^3 - 3x^4 = \frac{1}{2}$을 수치적으로 풀면 중앙값 ≈ 0.6143을 얻는데, 정확히 그 사이에 있다. 또한 $F(0.6) = 0.4752 < \frac{1}{2}$이고 $F(\frac{2}{3}) = \frac{16}{27} \approx 0.593 > \frac{1}{2}$이다.
- $n = 2$ $(N = 3)$일 때 전개식은 정확히 네 항이므로 하계가 등식이 된다. 양변 모두 $\frac{125}{27} \approx 4.63$이며, $6 - \frac{8}{5} = 4.4$보다 넉넉히 크다.
- 평균 < 중앙값 < 최빈값이라는 순서는 2부터 79까지의 모든 n에 대해 컴퓨터로 확인되었다.

흔한 실수

- 중앙값을 직접 구하려고 하는 것: 방정식 $x^{n+1}(n+2-(n+1)x) = \frac{1}{2}$은 깔끔한 해가 없으며, 해를 구할 필요도 없다.
- 부등호 방향을 잘못 잡는 것: F가 증가하므로 $\mu < m$은 $\mathrm{F}(\mu) < \frac{1}{2}$에 대응한다.
- 뒤쪽 이항 항들을 버려도 되는 이유(모두 양수라는 점)를 밝히지 않는 것.

STEP 너머

이 밀도함수(density)는 베타분포(Beta distribution) $\mathrm{Beta}(n+1, 2)$로, $[0,1]$ 위의 가장 중요한 분포족 가운데 하나이며 베이즈 통계학(Bayesian statistics)의 중심에 있다. 평균 < 중앙값 < 최빈값이라는 순서는 왼쪽으로 긴 꼬리를 가진 분포(음의 왜도(negative skew))에서 전형적으로 나타나지만, 보편적인 법칙은 아니다. 그래서 여기서 증명해야 했던 것이다. 소문항 (iii)의 감소함수는 $x \to \infty$일 때 e에 수렴하는데, 이는 해석학(analysis)에서 제대로 만나게 될 고전적인 극한이다. Blitzstein과 Hwang의 *Introduction to Probability*(확률론 입문)에 베타분포에 대한 훌륭한 설명이 있다.

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.