풀이

STEP I 2008, 문제 12

확률과 통계 • 이산 균등분포의 최댓값

> **STEP I 2008, 12번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/08-s1-q12

핵심 아이디어

X는 두 점수 중 더 큰 값일 뿐이다. "큰 값이 r과 같다"를 직접 구하려면 여러 경우를 따져야 하지만, "큰 값이 r 이하이다"는 쉽다. 이는 정확히 두 점수가 모두 r 이하일 때 일어나기 때문이다. 그래서 먼저 누적분포(cumulative distribution) $\mathrm{P}(X \le r)$를 구한 뒤 차를 취한다. 같은 누적분포를 이용하면 중앙값(median)도 거의 바로 구할 수 있다.

풀이

X_1과 X_2를 서로 독립(independent)인 두 점수라 하고, $X = \max(X_1, X_2)$라 하자 ($X_1 = X_2$일 때 정의의 두 경우가 일치하므로 X는 잘 정의된다(well defined)).

소문항 (i)

$r = 0, 1, \ldots, N$에 대하여 사건 $\{X \le r\}$는 사건 $\{X_1 \le r$ 그리고 $X_2 \le r\}$와 같다. 두 사건의 확률은 각각 r/N이고, X_1과 X_2가 독립이므로 두 사건도 독립이다. 따라서

$$\mathrm{P}(X \le r) = \frac{r}{N} \cdot \frac{r}{N} = \frac{r^2}{N^2}.$$

이 식은 $r = 0$일 때도 성립하며, 이때 값은 0이다. 그러므로 $r = 1, 2, \ldots, N$에 대하여

$$\mathrm{P}(X = r) = \mathrm{P}(X \le r) - \mathrm{P}(X \le r - 1) = \frac{r^2 - (r-1)^2}{N^2} = \frac{2r - 1}{N^2},$$

이고, 이것이 구하려던 결과이다. (확인 삼아 보면, $\sum_{r=1}^{N}(2r-1) = N^2$이므로 이 확률들의 합은 1이다.)

소문항 (ii)

주어진 두 합 공식을 이용하면

$$\begin{aligned}\mu = \sum_{r=1}^{N} r\,\frac{2r-1}{N^2} &= \frac{1}{N^2}\left(2\sum_{r=1}^{N} r^2 - \sum_{r=1}^{N} r\right)\\ &= \frac{1}{N^2}\left(\frac{N(N+1)(2N+1)}{3} - \frac{N(N+1)}{2}\right)\\ &= \frac{N+1}{N}\cdot\frac{2(2N+1)-3}{6} = \frac{(N+1)(4N-1)}{6N}.\end{aligned}$$

$N = 100$일 때

$$\mu = \frac{101\times 399}{600} = \frac{40\,299}{600} = 67.165,$$

이고, 이것이 구하려던 값이다.

소문항 (iii)

소문항 (i)에서 $\mathrm{P}(X \le m) = m^2/N^2$이고 $\mathrm{P}(X \ge m) = 1 - \mathrm{P}(X \le m-1) = 1 - (m-1)^2/N^2$이다. 중앙값에 대한 두 조건은 다음과 같이 바뀐다.

$$\mathrm{P}(X \le m) \ge \tfrac{1}{2} \iff m^2 \ge \tfrac{1}{2}N^2 \iff m \ge \frac{N}{\sqrt{2}},$$
$$\mathrm{P}(X \ge m) \ge \tfrac{1}{2} \iff (m-1)^2 \le \tfrac{1}{2}N^2 \iff m-1 \le \frac{N}{\sqrt{2}},$$

여기서 양의 제곱근을 취했는데, $m \ge 1$이어서 m과 $m-1$이 음이 아니므로 이는 타당하다. 따라서 m은 다음 부등식을 만족하는 정수여야 한다.

$$\frac{N}{\sqrt{2}} \le m \le \frac{N}{\sqrt{2}} + 1.$$

생각해 보기 길이가 1인 구간에는 정수가 두 개 들어갈 수 있다. 중앙값이 두 개일 수도 있을까?

그러려면 양 끝점이 모두 정수여야 하고, 이는 $N/\sqrt{2}$가 정수여야 함을 뜻한다. 그런데 $N/\sqrt{2} = k$이면 $\sqrt{2} = N/k$가 유리수(rational number)가 되어 모순이다. 따라서 두 부등식은 모두 등호 없이 성립하고, 길이가 1인 이 열린구간에는 정수가 정확히 하나 들어 있으며,

$$m = \left\lceil \frac{N}{\sqrt{2}} \right\rceil,$$

즉 $N/\sqrt{2}$보다 큰 가장 작은 정수이다.

$N = 100$일 때 $N/\sqrt{2} = 50\sqrt{2} \approx 70.71$이므로 $m = 71$이다. 직접 확인하면 $71^2 = 5041 \ge 5000 = \frac{1}{2}\times 100^2$이고 $70^2 = 4900 \le 5000$이므로 두 조건이 모두 성립한다.

소문항 (iv)

소문항 (iii)에서 어떤 $0 < \theta < 1$에 대하여 $m = N/\sqrt{2} + \theta$이므로

$$\frac{m}{N} = \frac{1}{\sqrt{2}} + \frac{\theta}{N} \to \frac{1}{\sqrt{2}} \quad \text{단, } N \to \infty.$$

소문항 (ii)에서

$$\frac{\mu}{N} = \frac{(N+1)(4N-1)}{6N^2} = \frac{1}{6}\left(1+\frac{1}{N}\right)\left(4-\frac{1}{N}\right) \to \frac{4}{6} = \frac{2}{3}.$$

그러므로 N이 클 때

$$\frac{\mu}{m} = \frac{\mu/N}{m/N} \approx \frac{2/3}{1/\sqrt{2}} = \frac{2\sqrt{2}}{3}.$$

검산

- $N = 2$일 때: 확률이 같은 네 쌍 가운데 $X = 1$인 경우가 한 번, $X = 2$인 경우가 세 번이므로 소문항 (i)에서 구한 $\frac{1}{4}$, $\frac{3}{4}$와 일치하고, $\mu = \frac{7}{4}$도 $\frac{3\times 7}{12}$와 일치한다.
- $N \le 40$에 대하여 컴퓨터로 N^2개의 쌍을 모두 나열하면 소문항 (i)과 (ii)가 정확히 확인된다. $N \le 2000$에 대하여 모든 정수를 중앙값의 정의에 대입해 보면 항상 중앙값이 정확히 하나 있고, 그 값은 $\lceil N/\sqrt{2} \rceil$이다.
- $N = 10^5$일 때 $\mu/m \approx 0.942812$이고, 이는 $2\sqrt{2}/3 \approx 0.942809$와 거의 같다.

흔한 실수

- $\mathrm{P}(X = r)$를 $2 \times \frac{r}{N} \cdot \frac{1}{N}$로 계산하는 실수. 이렇게 하면 $X_1 = X_2 = r$인 경우를 두 번 세게 된다. 올바르게 세면 $2(r-1)+1 = 2r-1$이다.
- $\mathrm{P}(X \ge m) = 1 - m^2/N^2$로 쓰는 실수. 이것은 $\mathrm{P}(X > m)$이지 $\mathrm{P}(X \ge m)$가 아니다.
- $N = 100$일 때 70.71을 올림하지 않고 가장 가까운 정수로 반올림하여 $m = 70$이라고 답하는 실수.

STEP 너머

X는 **순서통계량**(order statistic)의 가장 간단한 예, 즉 표본의 최댓값이다. $N \to \infty$이면 X/N은 $[0,1]$ 위에서 확률밀도(density)가 $2x$인 연속확률변수(continuous random variable)에 가까워지는데, 이 확률변수의 평균은 $\frac{2}{3}$이고 중앙값은 $1/\sqrt{2}$이다. 바로 이 때문에

소문항 (iv)의 비가 나타난다. 더 큰 표본의 순서통계량을 다루면 베타분포(beta distribution)로 이어진다. Blitzstein과 Hwang의 *Introduction to Probability*(확률론 입문)는 순서통계량을 이해하기 쉽게 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.