풀이

STEP II 2024, 문제 12

확률과 통계 • 이산 순서통계량

> **STEP II 2024, 12번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 OCR에서 내려받을 수 있는 2024년 STEP 시험지 묶음(zip 파일) 가운데 STEP II 시험지에 있다:
>
> www.ocr.org.uk/Images/725753-2024-past-papers.zip

핵심 아이디어

첫 번째 게임의 승자는 **유일한** 최솟값을 가진 사람이고, 그런 사람은 많아야 한 명이다. 따라서 "승자가 있다"는 사건은 누가 어떤 수로 이기는지에 따라 서로 배반인(disjoint) 경우들로 나뉘고, 각 경우는 독립(independent)인 확률들의 단순한 곱이다. $\sum i^2$과 $\sum i^3$에 대한 공식이 주어진 것은, 그 결과로 나오는 합이 다항식의 합이라는 힌트이다. 두 번째 게임도 같은 방식으로 풀리는데, 이번에는 사람의 순서쌍(ordered pair)(최솟값을 가진 사람과 최댓값을 가진 사람)과 합할 두 개의 수가 있다.

풀이

풀이 전체에서 k개의 수는 서로 독립이고 각각 $\{1, 2, \ldots, n\}$ 위의 균등분포(uniform distribution)를 따른다. 따라서 특정한 사람이 특정한 수를 받을 확률은 $\frac{1}{n}$이고, 주어진 s개의 값으로 이루어진 집합 안의 수를 받을 확률은 $\frac{s}{n}$이다.

소문항 (i)

$1 \le a \le n-1$인 a를 고정한다. 에이다는 a를 받아야 하고(확률 $\frac{1}{n}$), 나머지 $k-1$명은 각각 $n-a$개의 수 $a+1, \ldots, n$ 중 하나를 받아야 한다(각각 확률 $\frac{n-a}{n}$). 독립성에 의해

$$\mathrm{P}(\text{에이다가 받는 수가 } a,\ \text{나머지 모두 더 큼}) = \frac{1}{n}\left(\frac{n-a}{n}\right)^{k-1} = \frac{(n-a)^{k-1}}{n^k}.$$

($a = n$이면 이 값은 0이 되므로 a는 $n-1$까지만 간다.)

승자가 있는 사건을 W라 하자. 승자는 다른 모든 사람의 수보다 엄격하게 작은 수를 가지므로 승자가 두 명일 수는 없다. 두 사람이 각각 상대보다 작아야 하기 때문이다. 따라서 W는

k명의 사람과 $a = 1, \ldots, n-1$에 걸친 사건 "이 사람이 a를 받고 나머지는 모두 더 큰 수를 받는다"들의 배반 합집합이다. 대칭성(symmetry)에 의해 각 사람의 기여는 에이다의 기여와 같으므로

$$\mathrm{P}(W) = k \sum_{a=1}^{n-1} \frac{(n-a)^{k-1}}{n^k} = \frac{k}{n^k} \sum_{j=1}^{n-1} j^{k-1},$$

이다. 여기서 $j = n - a$로 치환하였고, j도 1부터 $n-1$까지 움직인다.

$k = 4$일 때 $\sum i^3$에 대한 공식을 n 대신 $n-1$을 넣어 쓰면

$$\mathrm{P}(W) = \frac{4}{n^4} \cdot \frac{1}{4}(n-1)^2 n^2 = \frac{(n-1)^2}{n^2},$$

이므로 원하는 결과가 나온다.

소문항 (ii)

$1 \le d \le n-2, 1 \le a \le n-d-1$인 d와 a를 고정한다. 이 조건은 $a + d + 1 \le n$이고 a와 $a+d+1$ 사이에 엄격하게 들어가는 수가 적어도 하나 있음을 보장한다. 에이다는 a를, 밥은 $a+d+1$을 받아야 하고(각각 확률 $\frac{1}{n}$), 나머지 $k-2$명은 각각 d개의 수 $a+1, \ldots, a+d$ 중 하나를 받아야 한다. 독립성에 의해

$$\mathrm{P}(\text{에이다가 받는 수}a\text{, 밥이 받는 수}a+d+1\text{, 나머지는 엄격하게 그 사이}) = \frac{1}{n^2}\left(\frac{d}{n}\right)^{k-2} = \frac{d^{k-2}}{n^k}.$$

> **생각해 보기** 승자가 두 명이 되는 모든 경우가 사람들과 a, d에 대한 합에서 왜 정확히 한 번씩 나타나는가?

승자가 두 명이라는 것은 한 사람이 유일한 최솟값을, 다른 한 사람이 유일한 최댓값을 가진다는 뜻이다. 소문항 (i)에서처럼 낮은 쪽 승자는 유일하고 높은 쪽 승자도 유일하므로, 결과는 서로 다른 두 사람의 **순서쌍**(낮은 쪽 승자, 높은 쪽 승자)과 함께 낮은 수 a, 그리고 높은 수까지의 간격 $d \ge 1$을 결정한다. 나머지 $k-2 \ge 2$명의 수가 엄격하게 그 사이에 있으므로 간격은 적어도 1이어야 한다. 이들을 다르게 고르면 서로 배반인 사건이 된다. 사람의 순서쌍은 $k(k-1)$개이고 각각의 기여는 (에이다, 밥)의 기여와 같으므로

$$\mathrm{P}(\text{승자 두 명}) = k(k-1) \sum_{d=1}^{n-2} \sum_{a=1}^{n-d-1} \frac{d^{k-2}}{n^k} = \frac{k(k-1)}{n^k} \sum_{d=1}^{n-2} (n-1-d)\, d^{k-2},$$

이다. 안쪽 합은 $n-d-1$개의 같은 항으로 이루어져 있기 때문이다.

이제 $k = 4$로 두고 $m = n-1$이라 쓰면 d는 1부터 $m-1$까지 움직인다. 주어진 두 공식을 n

대신 $m-1$을 넣어 쓰면

$$\begin{aligned}\sum_{d=1}^{m-1}(m-d)\,d^2 &= m\cdot\tfrac{1}{6}(m-1)m(2m-1)-\tfrac{1}{4}(m-1)^2m^2\\ &= \tfrac{1}{12}m^2(m-1)[2(2m-1)-3(m-1)]\\ &= \tfrac{1}{12}m^2(m-1)(m+1).\end{aligned}$$

따라서 $m-1=n-2$, $m+1=n$이므로

$$\mathrm{P}(\text{승자 두 명}) = \frac{12}{n^4}\cdot\frac{1}{12}(n-1)^2(n-2)n = \frac{(n-2)(n-1)^2}{n^3},$$

이고, 이것이 원하는 결과이다.

$k=4$일 때 승자가 한 명인 경우와 두 명인 경우의 비교

어떤 사람이 유일한 최솟값을 가지는 사건을 L, 어떤 사람이 유일한 최댓값을 가지는 사건을 H라 하자. 소문항 (i)에서 $\mathrm{P}(L)=\frac{(n-1)^2}{n^2}$이다. 모든 수 x를 $n+1-x$로 바꾸면 균등분포를 따르는 수는 여전히 균등분포를 따르고 "최소"와 "최대"가 서로 바뀌므로 $\mathrm{P}(H)=\mathrm{P}(L)$이기도 하다.

승자가 정확히 두 명인 사건은 $L\cap H$이므로 $T=\mathrm{P}(L\cap H)=\frac{(n-2)(n-1)^2}{n^3}$이라 쓰자. 승자가 정확히 한 명이라는 것은 L 또는 H가 일어나되 둘 다 일어나지는 않는다는 뜻이므로

$$\mathrm{P}(\text{정확히 한 명}) = \mathrm{P}(L)+\mathrm{P}(H)-2\,\mathrm{P}(L\cap H) = \frac{2(n-1)^2}{n^2}-2T.$$

승자가 두 명일 확률이 한 명일 확률보다 큰 것은 다음과 같을 때, 그리고 그때에만이다.

$$\begin{aligned}T>\frac{2(n-1)^2}{n^2}-2T &\iff 3\cdot\frac{(n-2)(n-1)^2}{n^3}>\frac{2(n-1)^2}{n^2}\\ &\iff 3(n-2)>2n \iff n>6,\end{aligned}$$

여기서 양수 $\frac{(n-1)^2}{n^3}$으로 나누었다($n\geq 3$이므로 양수이다). $n=6$일 때는 두 확률이 같으므로 최솟값은

$$n=7.$$

검산

- $n=3, k=4$일 때 결과는 81가지이다. 1로 이기는 승자가 있으려면 나머지 세 수가 $\{2,3\}$에 속해야 하므로 $4\cdot 2^3=32$가지이다. 2로 이기는 승자가 있으려면 나머지 세 수가 모두 3이어야 하므로 4가지이다. 따라서 $\mathrm{P}(W)=\frac{36}{81}=\frac{4}{9}=\frac{(3-1)^2}{3^2}$이다.
- $n=6$일 때 "정확히 한 명"과 "정확히 두 명"의 확률은 모두 $\frac{25}{54}\approx 0.463$이고, $n=7$일 때는 각각 $\frac{144}{343}\approx 0.420$, $\frac{180}{343}\approx 0.525$이다. 이 값들과 $n\leq 8$, $k=2,3,4$에 대한 두 소문항의 공식은 모든 결과를 컴퓨터로 열거하여 확인하였다.

- $n \to \infty$이면 같은 수가 나오는 경우가 드물어지므로, 네 명이 있으면 거의 항상 유일한 최솟값과 유일한 최댓값이 모두 있어야 한다. 실제로 $\frac{(n-2)(n-1)^2}{n^3} \to 1$이다.

흔한 실수

- 소문항 (i)에서 인수 k를, 또는 소문항 (ii)에서 인수 $k(k-1)$을 빠뜨리는 것. 소문항 (ii)에서 두 승자는 한 명은 낮은 쪽, 한 명은 높은 쪽이므로 순서가 있는 쌍이다.
- "승자가 정확히 한 명"을 $\mathrm{P}(L) + \mathrm{P}(H) - \mathrm{P}(L \cap H)$로 계산하는 것. 이것은 "승자가 적어도 한 명"이며, 겹치는 부분은 두 번 빼야 한다.
- $n = 6$이라고 답하는 것. 그때는 두 확률이 같을 뿐 엄격하게 크지 않다.

STEP 너머

사건을 "누가"와 "어디서"로 이름 붙은 배반인 조각들로 나누는 것은 표본의 최솟값과 최댓값, 즉 순서 통계량(order statistics)의 분포를 구하는 방법의 이산 버전이다. 연속 표본에서는 같은 논증으로 최솟값과 최댓값의 결합 밀도함수(joint density)를 얻고, 이 문제를 흥미롭게 만드는 같은 수가 나오는 경우는 사라진다. Grimmett과 Welsh의 *Probability: An Introduction*(확률론 입문)은 이런 아이디어를 접하기에 좋은 책이다.

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.