풀이

STEP II 2021, 문제 11

확률과 통계 • 조건부 확률의 점화식

> **STEP II 2021, 11번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 OCR에서 내려받을 수 있는 2021년 STEP 시험지 묶음(zip 파일) 가운데 STEP II 시험지에 있다:
>
> www.ocr.org.uk/Images/696588-2021-past-papers.zip

이 문제는 *Beyond STEP*의 12.5 절 **마지막 승객의 좌석**에서 단계별 안내와 함께 자세히 다룬다:
https://beyondstep.keithmcnulty.org/ko/ch12/airline-seat.html

핵심 아이디어

T_1이 어디에 앉는지를 조건으로 놓는다(conditioning). T_1이 자기 좌석 S_1에 앉으면 나머지 모두는 문제없다. T_1이 S_k에 앉으면 $T_2, \ldots, T_{k-1}$은 무사히 자기 자리에 앉고, T_k가 새로운 "무작위 선택자"가 된다. 결정적인 관찰은, 이때 T_k가 처음의 T_1과 정확히 같은 처지에 놓인다는 것이다. 다만 열차가 더 작아졌고, 비어 있는 좌석 S_1이 T_k 자신의 좌석 역할을 한다. 이로써 문제는 점화식(recurrence)으로 바뀐다.

풀이

이 풀이 전체에서 "T_r이 밀려난다"는 것은 T_r이 도착했을 때 S_r이 이미 차 있다는 뜻이다. 무작위로 좌석을 고르는 사람은 밀려난 승객(또는 T_1)뿐이다.

소문항 (i)

$n = 2$일 때, T_1은 각각 확률(probability) $\frac{1}{2}$로 S_1 또는 S_2에 앉고, T_2가 S_2에 앉는 것은 정확히 T_1이 S_1에 앉았을 때이다. 따라서

$$P_2 = \frac{1}{2}.$$

$n = 3$일 때, 각각 확률 $\frac{1}{3}$인 T_1의 좌석을 조건으로 놓는다.

- S_1: T_2와 T_3 모두 자기 좌석이 비어 있으므로, T_3은 S_3에 앉는다.

- S_3: T_3은 확실히 S_3에 앉지 못한다.
- S_2: T_2가 밀려나서 S_1과 S_3 중 하나를 무작위로 고른다. T_3이 S_3에 앉는 것은 정확히 T_2가 S_1을 고를 때이며, 그 확률은 $\frac{1}{2}$이다.

그러므로

$$P_3 = \tfrac{1}{3}\cdot 1 + \tfrac{1}{3}\cdot 0 + \tfrac{1}{3}\cdot\tfrac{1}{2} = \tfrac{1}{2}.$$

소문항 (ii)

T_1이 $2 \le k \le n-1$인 S_k에 앉는다고 하자. 그러면 $T_2, \ldots, T_{k-1}$은 모두 자기 좌석이 비어 있으므로 그 자리에 앉는다. T_k가 도착했을 때 빈 좌석은

$$S_1,\ S_{k+1},\ S_{k+2},\ \ldots,\ S_n,$$

으로 모두 $n-k+1$개이며, T_k는 이 중에서 균등하게 무작위로(uniformly at random) 하나를 고른다. 아직 오지 않은 승객은 $T_{k+1}, \ldots, T_n$이고, 이들은 각자 자기 좌석을 원하며, 아직 오지 않은 승객 중 누구에게도 S_1이 배정되어 있지 않다.

> ? **생각해 보기** 왜 이것이 정확히 좌석이 $n-k+1$개인 원래 문제와 같은가?

이름을 바꿔 붙여 보자. T_k를 "첫 번째 승객"이라 부르고, S_1을 T_k에게 배정된 좌석이라고 간주한다. 좌석이 m개인 원래 문제에서는 첫 번째 승객이 m개의 좌석 전체에서 균등하게 하나를 고르고, 이후의 각 승객은 자기 좌석이 비어 있으면 거기에 앉고 그렇지 않으면 빈 좌석 중에서 균등하게 하나를 고르며, 첫 번째 승객에게 배정된 좌석을 원하는 사람은 아무도 없다. 여기서 $m = n-k+1$이라 하면, 승객 $T_k, T_{k+1}, \ldots, T_n$은 좌석 $S_1, S_{k+1}, \ldots, S_n$을 가지고 정확히 이 규칙을 따르며, S_1을 원하는 사람은 아무도 없다. 규칙은 승객 자신의 좌석이 비어 있는지 판단할 때 말고는 좌석 이름을 전혀 쓰지 않으므로, 두 과정의 확률은 똑같다. 작아진 문제에서도 T_n은 여전히 마지막 승객이므로

$$\mathrm{P}(T_n\text{의 좌석이}S_n \mid T_1\text{의 좌석이}S_k) = P_{n-k+1}.$$

이제 $n \ge 3$일 때, $S_1, \ldots, S_n$ 각각의 확률이 $\frac{1}{n}$인 T_1의 좌석을 조건으로 놓는다. 그것이 S_1이면 T_n은 확실히 S_n에 앉고, S_n이면 T_n은 확실히 앉지 못한다. 따라서 전확률 법칙(law of total probability)에 의해

$$P_n = \frac{1}{n}\cdot 1 + \frac{1}{n}\sum_{k=2}^{n-1} P_{n-k+1} + \frac{1}{n}\cdot 0 = \frac{1}{n}\left(1 + \sum_{r=2}^{n-1} P_r\right),$$

이다. 여기서 $r = n-k+1$로 치환하였으며, k가 2부터 $n-1$까지 변할 때 이 값은 $n-1$부터 2까지 내려간다. 이것이 구하는 결과이다.

소문항 (iii)

모든 $n \geq 2$에 대하여 $P_n = \frac{1}{2}$임을 보이자. 강한 수학적 귀납법(strong induction)을 쓴다.

기본 단계. 소문항 (i)에서 $P_2 = \frac{1}{2}$이다.

귀납 단계. $n \geq 3$이라 하고, $r = 2, 3, \ldots, n-1$에 대하여 $P_r = \frac{1}{2}$라고 가정하자. 소문항 (ii)의 합은 $n-2$개의 항으로 이루어지고 각 항이 $\frac{1}{2}$이므로

$$P_n = \frac{1}{n}\left(1 + \frac{n-2}{2}\right) = \frac{1}{n} \cdot \frac{n}{2} = \frac{1}{2}.$$

따라서 $r = 2, \ldots, n$에 대하여 $P_r = \frac{1}{2}$이고, 귀납법에 의해 모든 $n \geq 2$에 대하여 $P_n = \frac{1}{2}$이다. (소문항 (i)의 $P_3 = \frac{1}{2}$와도 일치한다.)

소문항 (iv)

$n = 2$이면 $T_{n-1} = T_1$이고, 이 승객은 확률 $\frac{1}{2}$로 S_1에 앉는다. 따라서 $Q_2 = \frac{1}{2}$이다.

이제 $n \geq 3$이라 하자. 그러면 T_{n-1}은 T_1이 아니며, 다시 T_1의 좌석 S_k를 조건으로 놓는다.

- $k = 1$: 모두가 자기 좌석에 앉으므로 T_{n-1}은 S_{n-1}에 앉는다.
- $k = n$: $T_2, \ldots, T_{n-1}$ 모두 자기 좌석이 비어 있으므로, 이번에도 T_{n-1}은 S_{n-1}에 앉는다.
- $k = n-1$: T_{n-1}은 확실히 S_{n-1}에 앉지 못한다.
- $2 \leq k \leq n-2$: 소문항 (ii)와 똑같이, 나머지 여정은 T_k가 이끄는 좌석 $m = n - k + 1$개짜리 원래 문제이다. 그 작아진 열차에서 T_{n-1}은 끝에서 두 번째 승객이므로 확률은 Q_{n-k+1}이다. 여기서 m은 $3, 4, \ldots, n-1$에 걸쳐 변한다.

그러므로 $n \geq 3$일 때

$$Q_n = \frac{1}{n}\left(2 + \sum_{m=3}^{n-1} Q_m\right).$$

$n = 3$이면 합은 빈 합이므로 $Q_3 = \frac{2}{3}$이다. $3 \leq m \leq n-1$에 대하여 $Q_m = \frac{2}{3}$이면

$$Q_n = \frac{1}{n}\left(2 + \frac{2(n-3)}{3}\right) = \frac{1}{n} \cdot \frac{2n}{3} = \frac{2}{3},$$

이므로 강한 수학적 귀납법에 의해 모든 $n \geq 3$에 대하여 $Q_n = \frac{2}{3}$이다. 정리하면

$$Q_2 = \tfrac{1}{2}, \qquad Q_n = \tfrac{2}{3} \quad (n \geq 3).$$

생각해 보기 점화식 없이 $\frac{1}{2}$와 $\frac{2}{3}$을 알아낼 방법이 있는가?

있다. 누군가 무작위로 고를 때마다 S_1과, 나중에 올 승객의 아직 비어 있는 각 좌석은 똑같이 취급된다. 따라서 과정이 끝날 때까지 좌석 S_1과 S_n 중 어느 쪽이 먼저 차게 될지는 같은

정도로 가능하다. T_n이 성공하는 것은 정확히 S_1이 먼저 찰 때이므로 $\frac{1}{2}$를 얻는다. T_{n-1}의 경우($n \geq 3$), S_1, S_{n-1}, S_n 중 어느 것이 무작위 선택자에 의해 가장 먼저 차는지 살펴보자. 각각이 같은 정도로 가능하고, T_{n-1}이 밀려나는 것은 그것이 S_{n-1}일 때뿐이므로 $1 - \frac{1}{3} = \frac{2}{3}$을 얻는다.

검산

- $n = 3$인 경우는 손으로 나열할 수 있다. (T_1, T_2, T_3)의 결과는 확률 $\frac{1}{3}$인 (S_1, S_2, S_3), 확률 $\frac{1}{3}$인 (S_3, S_2, S_1), 그리고 각각 확률 $\frac{1}{6}$인 (S_2, S_1, S_3), (S_2, S_3, S_1)이다. 따라서 T_3이 S_3에 앉을 확률은 $\frac{1}{2}$, T_2가 S_2에 앉을 확률은 $\frac{2}{3}$으로, 앞에서 구한 것과 같다.
- $n = 2, \ldots, 8$에 대해 가능한 모든 여정을 컴퓨터로 정확히 열거하면, 모든 경우에 $P_n = \frac{1}{2}$이고 $n \geq 3$에서 $Q_n = \frac{2}{3}$이며, 모든 k에 대해 소문항 (ii)의 조건부 확률(conditional probability)도 확인된다. $n = 20$으로 10^5번의 여정을 모의실험한 결과는 0.500과 0.668이었다.

흔한 실수

- T_1이 S_1 자체를 고를 수 있다는 것을 잊는 실수. 이 경우가 점화식의 "$1 +$"를 준다.
- 소문항 (iv)에서 T_1이 S_n에 앉는 경우를 빠뜨리는 실수. 이 경우는 T_{n-1}에게 좋은 소식이므로 상수는 1이 아니라 2가 된다.
- $n = 2$에서 Q에 대한 점화식을 적용하는 실수. 이때 T_{n-1}은 T_1이고 답은 $\frac{2}{3}$이 아니라 $\frac{1}{2}$이다.

STEP 너머

여기서 쓴 점화식은 첫 단계 분석(first-step analysis)이다. 즉 첫 번째 무작위 선택을 조건으로 놓고, 같은 문제의 더 작은 사본을 알아보는 것이다. 같은 기법으로 마르코프 연쇄(Markov chain)의 도달 확률(hitting probability)과 기대 도달 시간(expected hitting time)을 구할 수 있으며, 이는 대학 1학년 확률론의 핵심 주제이다. 위의 대칭성(symmetry) 논증은 잘 고른 "어느 것이 먼저 일어나는가?"라는 질문이 긴 계산을 대신할 수 있음을 맛보여 준다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)이 두 아이디어를 모두 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.