풀이

STEP III 2003, 문제 13

확률과 통계 • 흡수 마르코프 연쇄

STEP III 2003, 13번 문제

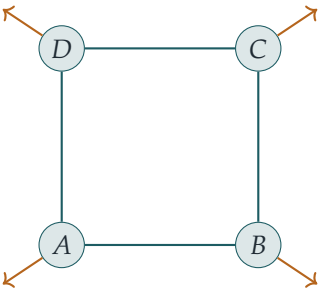
더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:

step.maths.org/questions/03-s3-q13

핵심 아이디어

이것은 "첫걸음 분석(first-step analysis)"이다. 토끼가 다음에 무엇을 하는지에 대해 조건을 걸면, 한 방에서 출발했을 때의 확률을 그 방에서 갈 수 있는 방들에서 출발했을 때의 확률로 나타낼 수 있다. 그러면 네 개의 연립일차방정식(linear equations)을 얻고, 정사각형의 대칭성(symmetry)(대각선 AC에 대한 반사는 B와 D를 맞바꾼다)을 이용하면 계산이 줄어든다. 마지막 문항에서는 C에 들어가는 것을 실패로 치는 변형된 토끼굴에 같은 방법을 적용한 다음, 조건부 확률(conditional probability)의 정의를 이용한다.

풀이



지상으로 나가는 굴

풀이 전체에서 각 움직임은 토끼의 과거와 독립적으로 각각 확률 $\frac{1}{3}$로 선택된다고 본다(어느 방에 "있게 된" 토끼에게 각 선택이 "똑같이 가능하다"는 말을 이렇게 해석한다). 토끼가 결국 A의 굴을 통해 밖으로 나가는 사건을 E_A라 하자.

소문항 (i)

토끼가 A에 있다고 하고, 다음 움직임에 대해 조건을 건다.

- 확률 $\frac{1}{3}$로 토끼는 곧바로 A의 굴로 올라가고, 사건 E_A가 일어난다.
- 확률 $\frac{1}{3}$로 토끼는 B로 달려간다. 이제 토끼는 B에 있고, 앞으로의 움직임은 어떻게 그곳에 왔는지와 무관하므로 결국 A로 나갈 확률은 p_B이다.
- 확률 $\frac{1}{3}$로 토끼는 D로 달려가고, 마찬가지로 이때의 확률은 p_D이다.

전확률 법칙(law of total probability)에 의해

$$p_A = \tfrac{1}{3}\cdot 1 + \tfrac{1}{3}p_B + \tfrac{1}{3}p_D,$$

이므로 증명이 끝난다.

소문항 (ii)

다른 방들에서 같은 논법을 쓰면 방정식이 세 개 더 나온다. 유일한 차이는 B, C, D의 굴로 나가면 E_A가 **실패**한 것이므로 그 갈래의 기여가 $\frac{1}{3}\cdot 0$이라는 점이다.

$$\begin{aligned} p_B &= \tfrac{1}{3}p_A + \tfrac{1}{3}p_C, \\ p_C &= \tfrac{1}{3}p_B + \tfrac{1}{3}p_D, \\ p_D &= \tfrac{1}{3}p_A + \tfrac{1}{3}p_C. \end{aligned}$$

p_B와 p_D에 대한 방정식의 우변이 같으므로 $p_B = p_D$이다. 이 공통값을 b라 하자. (이것이 AC에 대한 반사 대칭이다.) 그러면

$$p_A = \tfrac{1}{3} + \tfrac{2}{3}b, \qquad p_C = \tfrac{2}{3}b,$$

이고, 이를 p_B에 대한 방정식에 대입하면

$$b = \tfrac{1}{3}\left(\tfrac{1}{3} + \tfrac{2}{3}b + \tfrac{2}{3}b\right) = \tfrac{1}{9} + \tfrac{4}{9}b \quad\Longrightarrow\quad \tfrac{5}{9}b = \tfrac{1}{9} \quad\Longrightarrow\quad b = \tfrac{1}{5}.$$

따라서

$$p_A = \tfrac{1}{3} + \tfrac{2}{15} = \frac{7}{15}, \qquad \text{이고 또한} \qquad p_B = p_D = \tfrac{1}{5},\ p_C = \tfrac{2}{15}.$$

생각해 보기 토끼가 애초에 밖으로 나간다는 것을, 그래서 이 방정식들이 정말로 확률을 결정한다는 것을 어떻게 알 수 있는가?

매 단계마다 토끼는 확률 $\frac{1}{3}$로 밖으로 나가므로, n단계 뒤에도 여전히 땅속에 있을 확률은 $(\frac{2}{3})^n \to 0$이다. 즉 토끼는 확률 1로 밖으로 나간다. 또한 네 개의 연립일차방정식은 해가 하나뿐이므로(우리는 아무런 선택 없이 풀었다), 위의 값들이 바로 구하는 확률이다.

소문항 (iii)

토끼가 C를 한 번도 방문하지 않는 사건을 N이라 하자. 구하는 것은

$$\mathrm{P}(N \mid E_A) = \frac{\mathrm{P}(N \cap E_A)}{\mathrm{P}(E_A)},$$

이고, 토끼는 A에서 출발하므로 분모는 $p_A = \frac{7}{15}$이다.

분자를 구하기 위해, 토끼가 현재 각각 A, B, D에 있을 때(그리고 아직 C를 방문하지 않았을 때) C에 **한 번도 들어가지 않고** A의 굴로 나갈 확률을 q_A, q_B, q_D라 하자. 첫걸음 분석은 앞과 똑같이 적용되는데, 다만 이제는 C로 들어가는 움직임을 실패로 친다.

$$q_A = \tfrac{1}{3} + \tfrac{1}{3}q_B + \tfrac{1}{3}q_D,$$
$$q_B = \tfrac{1}{3}q_A + \tfrac{1}{3}\cdot 0 + \tfrac{1}{3}\cdot 0,$$

왜냐하면 B에서 토끼는 A로 되돌아가거나, C로 달려 들어가거나(실패), 잘못된 굴로 나가기(실패) 때문이다. 대칭성에 의해 $q_D = q_B = \frac{1}{3}q_A$이므로

$$q_A = \tfrac{1}{3} + \tfrac{2}{9}q_A \quad\Longrightarrow\quad \tfrac{7}{9}q_A = \tfrac{1}{3} \quad\Longrightarrow\quad q_A = \tfrac{3}{7}.$$

그러므로

$$\mathrm{P}(N \mid E_A) = \frac{3/7}{7/15} = \frac{45}{49}.$$

검산

- 네 값 p_A, p_B, p_C, p_D는 각 방에서 출발하여 A로 나갈 확률이다. 대칭성에 의해 이들은 또한 A에서 출발하여 각각 A, B, C, D로 나갈 확률이기도 하므로, 합이 1이어야 한다. 실제로 $\frac{7}{15} + \frac{3}{15} + \frac{2}{15} + \frac{3}{15} = 1$이다.
- 마땅히 그래야 하듯이 $p_A > p_B = p_D > p_C$이다. 토끼가 A에서 먼 곳에서 출발할수록 A로 나갈 가능성은 작아진다.
- (iii)의 답은 1에 가깝다. A로 나가는 토끼는 대개 짧게만 돌아다닌 뒤에 나간다. 토끼 4×10^5마리를 모의실험한 결과 p_A는 0.469, 조건부 확률은 0.918이었으며, 이론값은 $\frac{7}{15} \approx 0.467$과 $\frac{45}{49} \approx 0.918$이다.

흔한 실수

- p_B, p_C, p_D에 대한 방정식에 $\frac{1}{3}$의 "지금 나가기" 항을 포함하는 실수. A가 아닌 다른 굴로 나가는 것은 0을 기여한다.
- (iii)의 답으로 $q_A = \frac{3}{7}$을 제시하는 실수. 이것은 $\mathrm{P}(N \cap E_A)$이다. 문제는 조건부 확률을 묻고 있으므로 여전히 p_A로 나누어야 한다.

STEP 너머

토끼는 **마르코프 연쇄**(*Markov chain*)를 따라 움직이고 있다. 즉 다음 움직임은 지금 어디에 있는지에만 의존한다. 지상은 **흡수 상태**(*absorbing state*)이고, p_A, p_B, p_C, p_D는 **흡수 확률**(*absorption probabilities*)이며, 첫걸음 분석은 이를 언제나 연립일차방정식으로 바꾸어 준다. 같은 아이디어로 흡수될 때까지 걸리는 시간의 기댓값, 도박꾼의 파산(gambler's ruin) 확률 등 훨씬 많은 것을 구할 수 있다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 마르코프 연쇄를 처음 공부하기에 좋은 설명을 담고 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.