풀이

STEP I 2001, 문제 12

확률과 통계 • 수형도

> **STEP I 2001, 12번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/01-s1-q12

핵심 아이디어

모든 것은 **누가 무엇을 아는지**를 꼼꼼히 추적하는 데 달려 있다. 어떤 학생이 누군가가 소문을 들었다는 사실을 아는 것은 그 사람이 자신에게 말해 주었거나 자신이 그 사람에게 말해 주었을 때뿐이다. 누구도 소문을 어디서 들었는지는 전하지 않는다. 각 단계에서 각 학생이 아는 것을 기록해 두면 수형도(tree diagram)는 작아진다. 모든 갈림길은 두 사람 중 하나를 고르는 선택이므로 경로는 여덟 개뿐이고, 결국 이 경로들은 모두 같은 확률(probability)을 가진다는 것이 밝혀진다.

풀이

설정하기

네 학생을 A, B, C, D라 하자. 어느 단계에서든 학생 X가 소문을 들었다고 **아는** 사람은 X 자신, X에게 소문을 말해 준 사람, 그리고 X가 소문을 말해 준 사람이다. X가 소문을 들으면 X는 그 집합 밖의 사람들 가운데에서 균등하게 무작위로(uniformly at random) 한 명을 고른다. 그 집합에 이미 네 학생이 모두 들어 있으면 소문은 X에서 멈춘다.

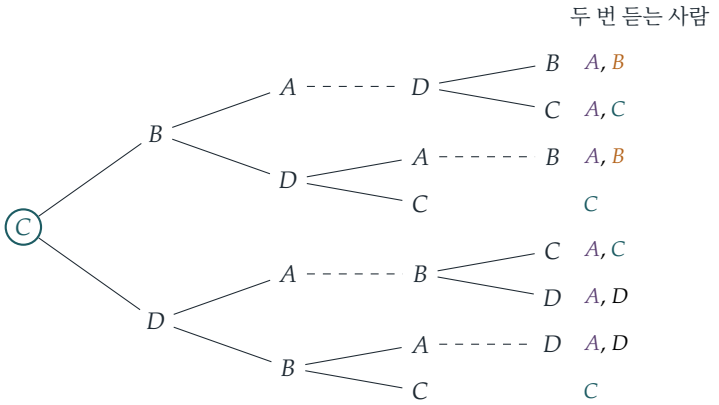
처음에 A가 C에게 말했으므로 A는 $\{A, C\}$를 알고 C도 $\{A, C\}$를 안다.

수형도

C는 A를 빼고 B 또는 D에게 각각 확률 $\frac{1}{2}$로 말한다. 먼저 $C \to B$ 갈래를 살펴보자. 이제 C는 $\{A, B, C\}$를 알고 B는 $\{B, C\}$를 안다 (C가 소문의 출처를 말하지 않았으므로 A는 모른다). 따라서 B는 A 또는 D에게 각각 확률 $\frac{1}{2}$로 말한다.

- $B \to A$. 이제 아서는 $\{A, B, C\}$를 알므로 D에게 말할 수밖에 없고, 그러면 아서는 모두가 소문을 들었음을 알게 된다. D는 $\{A, D\}$만 알므로 D는 B 또는 C에게 각각 확률 $\frac{1}{2}$로 말한다. 어느 쪽이든 듣는 사람의 집합이 완성된다. B는 이미 $\{A, B, C\}$를 알고 있었고 C도 마찬가지였으며, 이제 각자 D에 대해 알게 된다. 따라서 소문은 멈춘다.
- $B \to D$. 이제 B는 $\{B, C, D\}$를 알고 D는 $\{B, D\}$를 알므로 D는 A 또는 C에게 말한다.
  - $D \to C$: C는 이미 $\{A, B, C\}$를 알고 있었고 이제 D에 대해 알게 되므로 소문은 멈춘다. 아서는 소문을 다시 듣지 못한다.
  - $D \to A$: 아서는 $\{A, C, D\}$를 알므로 B에게 말하고, 그러면 이 학생은 네 명을 모두 알게 된다. 소문은 멈춘다.

$C \to D$ 갈래는 B와 D의 역할을 서로 바꾼 것과 같다. 전체 수형도는 다음과 같다. 강제된 이동(확률 1)은 점선으로 나타냈고, 그 밖의 모든 가지에는 확률 $\frac{1}{2}$가 붙는다.



잎에 이르는 각 경로에는 확률 $\frac{1}{2}$인 선택이 정확히 세 번 들어 있으므로 (강제된 이동은 인수 1을 곱할 뿐이다) 여덟 가지 결과 각각의 확률은 $\frac{1}{8}$이다. 오른쪽의 표시는 그 경로에서 소문을 두 번 듣는 사람을 나타낸다. (어느 경로에서든 한 학생이 소문을 듣는 횟수는 많아야 두 번이다.)

생각해 보기 C에게서 막 소문을 들은 B는 왜 아서가 소문을 들었다는 것을 모를까?

C가 누구에게서 들었는지 말하지 않기 때문이다. B가 추론할 수 있는 것은 C가 소문을 들었다는 사실뿐이다. 바로 이 점이 소문을 계속 이어지게 한다. $C \to B \to A$ 갈래에서 아서는 이미 알고 있던 것을 다시 듣게 된다.

아서가 다시 듣는 경우

아서는 두 개의 짧은 경로 $C \to B \to D \to C$와 $C \to D \to B \to C$를 제외한 모든 경로에서 소문을 다시 듣는다. 따라서

$$\mathrm{P}(\text{아서가 다시 들음}) = 1 - 2 \times \tfrac{1}{8} = \tfrac{3}{4},$$

이고, 이것이 보이려던 결과이다.

버사와 찬드라

버사가 소문을 두 번 듣는 것은 경로 $C \to B \to A \to D \to B$와 $C \to B \to D \to A \to B$에서뿐이므로

$$\mathrm{P}(\text{버사가 두 번 들음}) = \tfrac{2}{8} = \tfrac{1}{4}.$$

찬드라는 네 경로에서 소문을 두 번 듣는다. 수형도의 각 절반에 두 개씩이다 ($C \to B \to A \to D \to C$, $C \to B \to D \to C$, $C \to D \to A \to B \to C$, $C \to D \to B \to C$). 따라서

$$\mathrm{P}(\text{찬드라가 두 번 들음}) = \tfrac{4}{8} = \tfrac{1}{2}.$$

찬드라가 버사보다 소문을 다시 들을 가능성이 더 큰 것은 찬드라가 소문을 일찍 들었기 때문이다. 사슬에서 나중에 소문을 듣는 사람들은 찬드라가 소문을 들었다는 사실을 모르므로, 찬드라는 끝까지 소문을 전해 받을 수 있는 대상으로 남는다.

검산

B와 D 사이의 대칭성(symmetry)에 의해 딜라일라가 소문을 두 번 들을 확률도 $\frac{1}{4}$여야 하며, 실제로 수형도에는 그런 경로가 두 개 있다. 결과의 합도 맞는다. 같은 확률을 가진 잎이 여덟 개이고 전체 확률은 1이다. 컴퓨터로 이 과정을 정확히 열거(enumeration)하면 $\frac{3}{4}$, $\frac{1}{4}$, $\frac{1}{2}$를 얻고, 소문 200 000번을 시뮬레이션(simulation)한 결과는 0.750, 0.250, 0.501이었다.

흔한 실수

- 듣는 사람이 사슬에서 더 앞에 있는 모든 사람을 안다고 생각하는 것. 전해지는 것은 말해 준 사람이 누구인지뿐이며, 모든 것을 다 안다면 소문은 결코 아서에게 돌아오지 않을 것이다.
- **말해 준 사람**도 무언가를 알게 된다는 사실을 잊는 것. X는 Y에게 말해 준 뒤 Y가 소문을 들었음을 알게 되며, 바로 이 점이 아서의 이동을 강제한다.
- 아서가 소문을 다시 듣자마자 수형도 그리기를 멈추는 것. 버사와 찬드라를 다루려면 그 뒤의 갈래들이 필요하다.

STEP 너머

이 소문은 네 학생 위에서 움직이는 무작위 행보(random walk)인데, 그 규칙이 지금까지의 이력에 의존하며, 여기서는 그 이력을 각 학생의 지식으로 기록했다. 다음 단계가 현재 상태에만 의존하도록 상태를 충분히 풍부하게 잡으면 이 과정은 **마르코프 연쇄**(Markov chain)가 된다. 마르코프 연쇄는 확률과정(random processes) 첫 강좌의 중심 대상이며, 소문 모형과 전염병 모형은 그 고전적인 응용이다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 마르코프 연쇄를 딱 알맞은 수준으로 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.