풀이

STEP I 1999, 문제 14

확률과 통계 • 조건부 확률

> **STEP I 1999, 14번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/99-s1-q14

핵심 아이디어

연속확률변수(continuous random variable)는 이산(discrete) 확률변수를 만들기 위한 장치일 뿐이다. 밀도함수(density)를 4, 3, 2, 1점을 얻을 네 확률로 바꾸고 나면, 소문항 (i)은 한 줄짜리 기댓값(expectation) 계산이다. 소문항 (ii)는 조건부확률(conditional probability) 문제인데, 합계가 정확히 3이 되어야 하므로 이기는 경우가 매우 적어서 모두 나열할 수 있다.

풀이

소문항 (i)

$0 \le a \le b \le 1$에 대하여

$$\mathrm{P}(a \le X < b) = \int_a^b 2x \,\mathrm{d}x = b^2 - a^2.$$

(연속확률변수에서는 끝점을 포함하든 제외하든 차이가 없다.) 다트 한 번으로 얻는 점수를 S라 하자. 그러면

$$\mathrm{P}(S=4) = \left(\tfrac{1}{4}\right)^2 - 0 = \tfrac{1}{16}, \qquad \mathrm{P}(S=3) = \left(\tfrac{1}{2}\right)^2 - \left(\tfrac{1}{4}\right)^2 = \tfrac{3}{16},$$

$$\mathrm{P}(S=2) = \left(\tfrac{3}{4}\right)^2 - \left(\tfrac{1}{2}\right)^2 = \tfrac{5}{16}, \qquad \mathrm{P}(S=1) = 1 - \left(\tfrac{3}{4}\right)^2 = \tfrac{7}{16}.$$

이 확률들의 합은 마땅히 그래야 하듯 1이다. 바깥쪽 고리일수록 확률이 더 크다는 점에 주목하자. 중심에서 먼 곳일수록 넓이가 더 넓기 때문에 밀도 $2x$가 x와 함께 커진다.

따라서

$$\mathrm{E}(S) = 4 \cdot \tfrac{1}{16} + 3 \cdot \tfrac{3}{16} + 2 \cdot \tfrac{5}{16} + 1 \cdot \tfrac{7}{16} = \frac{4+9+10+7}{16} = \frac{30}{16} = \frac{15}{8},$$

이므로 원하는 결과를 얻는다.

소문항 (ii)

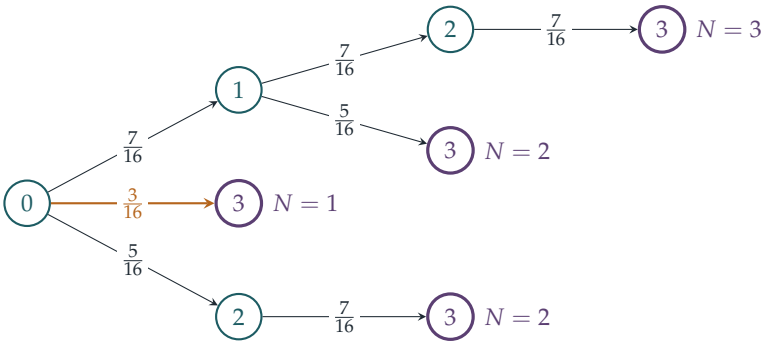
각 던지기는 독립(independent)이고, 모두 위의 S와 같은 분포를 따른다고 가정한다. 내가 이기는 사건을 W, 게임에서 던지는 횟수를 N이라 하자.

? **생각해 보기** 어떤 점수의 나열이 승리로 이어지는가?

누적 합계가 먼저 넘어서는 일 없이 3에 도달할 때, 그리고 오직 그때만 내가 이긴다. 모든 점수는 적어도 1이므로 합계는 매번 적어도 1씩 늘어나고, 게임은 많아야 세 번 던지면 끝난다. 이기는 게임은 합이 정확히 3이 되는 양의 점수들의 나열이다. (합계가 3보다 작을 때만 다시 던지므로, 이러한 나열은 모두 실제로 하나의 완결된 게임이다.) 그런 나열은 다음 네 가지뿐이다.

$$(3), \qquad (1,2), \qquad (2,1), \qquad (1,1,1).$$

점수 4는 단번에 합계가 3을 넘어서게 하므로 결코 나타나지 않는다.



이 나무 그림은 승리로 끝나는(보라색 원으로 표시) 각 경로를 따라 누적 합계를 보여 주며, 각 던지기의 확률을 함께 적어 두었다.

독립성에 의해 각 나열의 확률은 그 점수들의 확률의 곱이다.

$$\begin{aligned}
\mathrm{P}(W \cap N = 1) &= \mathrm{P}(3) = \tfrac{3}{16} = \tfrac{768}{4096}, \\
\mathrm{P}(W \cap N = 2) &= \mathrm{P}(1,2) + \mathrm{P}(2,1) = 2 \cdot \tfrac{7}{16} \cdot \tfrac{5}{16} = \tfrac{70}{256} = \tfrac{1120}{4096}, \\
\mathrm{P}(W \cap N = 3) &= \mathrm{P}(1,1,1) = \left(\tfrac{7}{16}\right)^3 = \tfrac{343}{4096}.
\end{aligned}$$

이 세 사건은 서로 배반(disjoint)이고 합치면 W가 되므로

$$\mathrm{P}(W) = \frac{768 + 1120 + 343}{4096} = \frac{2231}{4096}.$$

그러므로

$$\mathrm{P}(N = 3 \mid W) = \frac{\mathrm{P}(W \cap N = 3)}{\mathrm{P}(W)} = \frac{343/4096}{2231/4096} = \frac{343}{2231},$$

이므로 원하는 결과를 얻는다.

검산

- 네 점수의 확률 $\frac{1}{16}$, $\frac{3}{16}$, $\frac{5}{16}$, $\frac{7}{16}$의 합은 1이다. 이들이 분모가 16인 홀수 분자의 분수인 것은 $(k+1)^2 - k^2 = 2k+1$이기 때문이다.
- 목표값 $343 = 7^3$을 보면 세 번 던져 이기는 경우가 $(1,1,1)$이어야 함을 곧바로 알 수 있으며, 이는 설정이 맞음을 확인해 준다.
- 10^6번의 게임을 컴퓨터 모의실험(simulation)으로 돌린 결과 $\mathrm{P}(W) \approx 0.5452$(정확한 값 0.5447), $\mathrm{P}(N = 3 \mid W) \approx 0.1535$(정확한 값 $343/2231 \approx 0.1537$)를 얻었다.

흔한 실수

- 점수 구간들의 확률이 모두 같다고 보는 것. 확률은 구간의 폭이 아니라 밀도함수를 적분해서 얻는다.
- $(1,2)$는 세고 $(2,1)$을 빠뜨리는 것. 이렇게 하면 $\mathrm{P}(W) = \frac{1671}{4096}$이 나온다.
- $\mathrm{P}(W)$로 나누지 않고 $\mathrm{P}(N = 3 \text{ 이고} W) = \frac{343}{4096}$을 답으로 쓰는 것.

STEP 너머

소문항 (ii)는 작은 마르코프 연쇄(Markov chain)이다. 상태는 누적 합계이고, 3과 "초과(bust)"는 흡수 상태(absorbing state)이다. 목표값이 더 커지면 경로를 일일이 나열하는 것은 더 이상 실용적이지 않다. 대신 합계 t에서 출발하여 이길 확률을 w_t라 쓰면 점화식(recurrence) $w_t = \sum_s \mathrm{P}(S = s)\, w_{t+s}$를 얻는다. 이러한 첫걸음 분석(first-step analysis)은 대학 1학년 확률론의 핵심 도구이다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 무작위 행보(random walk)와 마르코프 연쇄를 다루는 부분에서 이를 설명한다.

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.