풀이

STEP II 2012, 문제 12

확률과 통계 • 조건부 확률

> **STEP II 2012, 12번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/12-s2-q12

핵심 아이디어

스위치를 누르면 상태가 **뒤바뀐다**(toggle). 따라서 복도에 있는 수영장 스위치가 마지막에 켜져 있는 경우는 처음부터 켜져 있었고 건드리지 않았거나, 처음에 꺼져 있었고 눌렸거나 둘 중 하나이다. 부엌 쪽은 교묘하게 위장한 함정(red herring)이다. 확률 $\frac{1}{2}$로 켜져 있는 스위치는 상태가 뒤바뀌든 아니든 여전히 확률 $\frac{1}{2}$로 켜져 있다. 소문항 (i)에서 하나의 조건부 확률(conditional probability)을 구하고 나면, 일곱 번의 월요일은 이항(binomial) 실험이 되고, "가장 가능성이 큰 값이 3이다"라는 조건은 이웃한 이항 확률 사이의 부등식 두 개가 된다.

풀이

소문항 (i)

다음과 같이 사건(event)을 정의한다.

- H: 손님이 복도에 있는 수영장 스위치를 누른다. 따라서 $\mathrm{P}(H) = \frac{1}{4}$이다.
- A: 처음에 복도의 스위치가 모두 켜져 있다. 따라서 $\mathrm{P}(A) = p$이다.
- L: 마지막에 수영장 조명이 켜져 있다.

손님은 스위치를 무작위로 고르므로 H는 A와 독립(independent)이다.

복도. 스위치를 누른 뒤 복도의 수영장 스위치가 켜져 있는 경우는 정확히 A가 일어나고 H가 일어나지 않은 경우(켜져 있었고 건드리지 않음) 또는 A가 일어나지 않고 H가 일어난

경우(꺼져 있었는데 눌려서 켜짐)이다. 따라서

$$\mathrm{P}(\text{마지막에 복도의 수영장 스위치가 켜져 있음}) = p \cdot \tfrac{3}{4} + (1-p) \cdot \tfrac{1}{4} = \frac{1+2p}{4}.$$

부엌. 부엌의 수영장 스위치는 처음에 확률 $\frac{1}{2}$로 켜져 있다. 눌리지 않으면 그대로이고, 눌리면 상태가 뒤집히는데, 확률 $\frac{1}{2}$로 꺼져 있는 스위치는 뒤집히면 확률 $\frac{1}{2}$로 켜진다. 어느 경우든 손님이 부엌에서 무엇을 하든 이 스위치는 마지막에 확률 $\frac{1}{2}$로 켜져 있다.

부엌 스위치들과 손님의 부엌 선택은 복도에서 일어나는 모든 일과 독립이므로

$$\mathrm{P}(L) = \frac{1+2p}{4} \cdot \frac{1}{2} = \frac{1+2p}{8}.$$

조건부 확률을 구하기 위해 보면, $H \cap L$이 일어나려면 손님이 복도의 수영장 스위치를 누르고, 그 스위치가 처음에 꺼져 있었으며(그래야 눌러서 켜진다), 부엌 스위치가 마지막에 켜져 있어야 한다.

$$\mathrm{P}(H \cap L) = \frac{1}{4} \cdot (1-p) \cdot \frac{1}{2} = \frac{1-p}{8}.$$

그러므로

$$\mathrm{P}(H \mid L) = \frac{(1-p)/8}{(1+2p)/8} = \frac{1-p}{1+2p},$$

이고, 이것이 구하려던 결과이다.

소문항 (ii)

$q = \dfrac{1-p}{1+2p}$라 하자. 각 월요일마다 조명이 켜져 있었다는 사실이 주어지므로, 소문항 (i)에 의해 복도의 수영장 스위치는 확률 q로 눌렸고, 이는 주마다 독립이다(손님들은 서로 독립적으로 행동하고, 매주 월요일 처음 상태가 새로 정해진다). 따라서 스위치가 눌린 날의 수 X는

$$X \sim \mathrm{B}(7, q).$$

를 만족한다. $1-q$도 필요하다.

$$1 - q = \frac{(1+2p) - (1-p)}{1+2p} = \frac{3p}{1+2p}, \qquad \text{따라서} \qquad \frac{q}{1-q} = \frac{1-p}{3p}.$$

(여기서 $p > 0$이다. $p = 0$이면 $q = 1$이고 확실히 $X = 7$이므로 3이 가장 가능성이 큰 값이 될 수 없다.)

> **생각해 보기** "3이 가장 가능성이 큰 값이다"를 어떻게 계산 가능한 형태로 바꿀 수 있을까?

이웃한 확률을 비교한다. $1 \le k \le 7$에 대하여

$$\frac{\mathrm{P}(X=k)}{\mathrm{P}(X=k-1)} = \frac{\binom{7}{k} q^k (1-q)^{7-k}}{\binom{7}{k-1} q^{k-1} (1-q)^{8-k}} = \frac{8-k}{k} \cdot \frac{q}{1-q}.$$

인수 $\frac{8-k}{k}$는 k가 커질수록 순증가가 아니라 순감소하므로, 확률은 이 비가 1보다 큰 동안 증가하고 1보다 작아지면 감소한다. 따라서 3이 (유일한) 가장 가능성이 큰 값이 될 필요충분조건은 확률이 $k=2$에서 $k=3$으로 갈 때 증가하고 $k=3$에서 $k=4$로 갈 때 감소하는 것이다.

$$\frac{\mathrm{P}(X=3)}{\mathrm{P}(X=2)} = \frac{5}{3}\cdot\frac{q}{1-q} > 1 \qquad \text{그리고} \qquad \frac{\mathrm{P}(X=4)}{\mathrm{P}(X=3)} = \frac{4}{4}\cdot\frac{q}{1-q} < 1.$$

(우리에게 필요한 방향으로는 이 두 조건만으로 충분하다. 3이 가장 가능성이 큰 값이라면 당연히 $\mathrm{P}(X=3) > \mathrm{P}(X=2)$이고 $\mathrm{P}(X=3) > \mathrm{P}(X=4)$이다.)

$\frac{q}{1-q} = \frac{1-p}{3p}$를 대입하고 양의 값 $3p$를 양변에 곱하면

- 두 번째 조건에서 $1-p < 3p$, 즉 $p > \frac{1}{4}$를 얻는다.
- 첫 번째 조건에서 $\frac{5}{3}(1-p) > 3p$, 즉 $5-5p > 9p$이므로 $p < \frac{5}{14}$를 얻는다.

그러므로 $\frac{1}{4} < p < \frac{5}{14}$이고, 이것이 구하려던 결과이다.

검산

- $p=0$이면 복도 스위치가 모두 꺼져 있으므로, 수영장 스위치가 눌린 경우에만 조명이 켜질 수 있다. 소문항 (i)의 결과는 $\mathrm{P}(H \mid L) = 1$로, 마땅히 그래야 하는 값이다. $p=1$이면 그 스위치가 **눌리지 않은** 경우에만 조명이 켜질 수 있는데, 공식도 0을 준다.
- 양 끝점은 $q=\frac{1}{2}$($p=\frac{1}{4}$일 때)와 $q=\frac{3}{8}$($p=\frac{5}{14}$일 때)에 대응한다. $q=\frac{1}{2}$이면 $\mathrm{B}(7,\frac{1}{2})$는 대칭이므로 3과 4의 확률이 같다. $q=\frac{3}{8}$이면 $(n+1)q = 3$이고 2와 3의 확률이 같다. 따라서 등호 없는 부등식이 정확히 맞다.
- 소문항 (i)은 모든 처음 상태와 모든 스위치 선택을 분수로 정확히 나열하여 확인했고, 소문항 (ii)는 간격이 $\frac{1}{2800}$인 격자 위의 p에 대해 $\mathrm{B}(7,q)$ 확률을 계산하여 확인했다. 유일한 최빈값(mode)이 3이 되는 것은 정확히 $\frac{1}{4} < p < \frac{5}{14}$일 때이다.

흔한 실수

- 이미 켜져 있는 스위치를 누르면 **꺼진다**는 사실을 잊는 것. 복도 계산에서는 "켜짐"에 이르는 두 경로를 모두 고려해야 한다.
- 답이 그냥 $\frac{1}{2}$인데도 부엌을 경우로 나누어 계산하는 것. 경우 분석이 틀린 것은 아니지만 계산 실수가 나기 쉬운 곳이다.
- 소문항 (ii)에서 이항분포의 모수로 q 대신 p를 쓰는 것. 조명이 켜져 있다는 조건이 바로 스위치를 누를 확률을 $\frac{1}{4}$에서 q로 바꾸는 것이다.

STEP 너머

소문항 (i)은 베이즈 정리(Bayes' theorem)가 실제로 작동하는 모습이다. 관찰(조명이 켜져 있음)이 손님이 스위치를 눌렀을 확률을 $\frac{1}{4}$에서 $\frac{1-p}{1+2p}$로 갱신한다. 소문항 (ii)는 이 생각을 거꾸로 적용하여, 주어진 가장 가능성 큰 결과와 양립하는 미지의 모수 p의 값이 무엇인지 묻는다. 이것이 통계적 추론(statistical inference)과 최대가능도 추정(maximum likelihood estimation)의 씨앗이다. 연속한 항의 비를 이용해 최빈값을 찾는 방법은 많은 이산 분포(푸아송(Poisson) 분포, 초기하(hypergeometric) 분포)에서도 통한다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 조건부 확률과 이러한 분포들을 꼼꼼하게 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.