풀이

STEP I 1996, 문제 14

확률과 통계 • 연속된 결과와 조건부확률

> **STEP I 1996, 14번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/96-s1-q14

핵심 아이디어

현재의 연속(run)이 어떻게 끝나는지에 따라 조건화(conditioning)한다. 동전의 면이 바뀌는 순간, 그 이전에 일어난 일은 모두 무관해진다. 이전의 연속은 끊어지고, 새 연속의 길이는 1이 된다. 따라서 앞면이 나온 뒤에는 뒷면이 나오는 순간 게임이 "다시 시작"되며, 그 상태는 맨 처음 던진 결과가 뒷면이었을 때와 정확히 같다. 그 반대도 마찬가지이다. 이로부터 미지수 두 개에 대한 일차방정식(linear equation) 두 개를 얻는다.

풀이

$x = \mathrm{P}(A \mid H)$, $y = \mathrm{P}(A \mid T)$라 하자. 모든 던지기는 서로 독립(independent)이다. $r, s \geq 1$이라 한다.

$\mathrm{P}(A \mid H)$에 대한 방정식

첫 번째 던지기에서 앞면이 나왔다고 하자. 이제 앞면 하나로 이루어진 연속이 있고, 뒷면의 연속은 아직 시작되지 않았다. 그다음 $r-1$번의 던지기(즉 2번째부터 r번째까지)를 살펴보자.

- 이들이 모두 앞면이면(확률 p^{r-1}), r번째 던지기에서 앞면 r개의 연속이 완성된다. 뒷면은 한 번도 나오지 않았으므로 뒷면 s개의 연속이 먼저 나왔을 리 없고, A가 일어난다.
- 그렇지 않으면(확률 $1-p^{r-1}$), 앞면의 연속이 길이 r에 이르기 전에 $2, \ldots, r$번째 던지기 중에 뒷면이 나온다. 그러한 첫 뒷면에서 앞면의 연속은 r에 이르지 못한 채 끊어지고, 그 이전에는 뒷면이 나온 적이 없으므로 뒷면 s개의 연속도 아직 일어나지 않았다. 이제 정확히 뒷면 하나로 이루어진 연속을 가지고 있으며, 그 이전의 어떤 것도 앞으로의 연속에

영향을 줄 수 없다.

생각해 보기 그 뒷면 이후로 A가 일어날 확률이 정확히 $\mathrm{P}(A \mid T)$인 이유는 무엇인가?

첫 번째 던지기에서 뒷면이 나온 뒤의 상황은 다음과 같다. 뒷면 하나로 이루어진 연속이 있고, 진행 중인 앞면의 연속은 없으며, 앞으로의 던지기는 과거와 독립이다. 이는 위의 두 번째 경우에서 첫 뒷면이 나온 시점의 상황과 정확히 같다. 이후의 던지기는 이전의 던지기와 독립이고, 두 연속 중 어느 것이 먼저 나오는지는 현재의 연속과 앞으로의 던지기에만 의존하므로, 그 시점부터 A가 일어날 조건부 확률은 y이다. 따라서 전확률 법칙(law of total probability)에 의해

$$\mathrm{P}(A \mid H) = p^{r-1} + \left(1 - p^{r-1}\right)\mathrm{P}(A \mid T),$$

이므로 $\alpha = r - 1$이다.

$\mathrm{P}(A \mid T)$에 대한 방정식

첫 번째 던지기에서 뒷면이 나왔다고 하자. 2번째부터 s번째까지의 던지기를 살펴보자.

- 모두 뒷면이면(확률 q^{s-1}), 앞면이 나오기 전에 뒷면 s개의 연속이 완성되므로 A는 일어나지 않는다.
- 그렇지 않으면(확률 $1 - q^{s-1}$), 뒷면이 연달아 s개에 이르기 전에 앞면이 나온다. 그 앞면에서 앞면 하나로 이루어진 연속이 있고 진행 중인 뒷면의 연속은 없는데, 이는 첫 번째 던지기에서 앞면이 나온 뒤의 상황과 정확히 같다. 따라서 그 시점부터 A는 확률 x로 일어난다.

그러므로

$$\mathrm{P}(A \mid T) = \left(1 - q^{s-1}\right)\mathrm{P}(A \mid H),$$

이고 $\beta = s - 1$이다.

(극단적인 경우도 잘 들어맞는다. $r = 1$이면 첫 번째 방정식은 $x = 1$이 되는데, 앞면 하나가 이미 길이 r인 연속이므로 옳다. $s = 1$이면 두 번째 방정식은 $y = 0$이 된다.)

방정식 풀기

두 번째 방정식을 첫 번째 방정식에 대입하면

$$x = p^{r-1} + \left(1 - p^{r-1}\right)\left(1 - q^{s-1}\right)x.$$

이제

$$1 - \left(1 - p^{r-1}\right)\left(1 - q^{s-1}\right) = p^{r-1} + q^{s-1} - p^{r-1}q^{s-1},$$

이고, 이 값은 양수이다($q^{s-1}(1-p^{r-1}) \geq 0$이므로 $0 < p < 1$일 때 이 값은 $p^{r-1} > 0$ 이상이다). 따라서 이 값으로 나눌 수 있다. $D = p^{r-1} + q^{s-1} - p^{r-1}q^{s-1}$이라 쓰면

$$\mathrm{P}(A \mid H) = \frac{p^{r-1}}{D}, \qquad \mathrm{P}(A \mid T) = \frac{p^{r-1}\left(1-q^{s-1}\right)}{D}.$$

마지막으로 $\mathrm{P}(H) = p, \mathrm{P}(T) = q$와 전확률 법칙에 의해

$$\begin{aligned}\mathrm{P}(A) &= p\,\mathrm{P}(A \mid H) + q\,\mathrm{P}(A \mid T)\\ &= \frac{p^{r-1}\left(p+q-q^{s}\right)}{D}\\ &= \frac{p^{r-1}\left(1-q^{s}\right)}{p^{r-1}+q^{s-1}-p^{r-1}q^{s-1}},\end{aligned}$$

이다. 여기서 $p+q=1$을 이용하였다.

검산

- 대칭성: $r = s$인 공정한 동전이라면 답은 $\frac{1}{2}$이어야 한다. $p = q = \frac{1}{2}, r = s = n$이면 분자는 $2^{1-n}(1-2^{-n})$이고 분모는 $2^{2-n} - 2^{2-2n} = 2 \cdot 2^{1-n}(1-2^{-n})$이므로 $\frac{1}{2}$을 얻는다.
- 작은 경우: $r = 1$이면 $\mathrm{P}(A) = 1 - q^s$인데, 이는 처음 s번의 던지기가 모두 뒷면은 아닐 확률이므로 옳다.
- $p = \frac{1}{2}, r = 3, s = 2$일 때 공식은 $\mathrm{P}(A \mid H) = \frac{2}{5}, \mathrm{P}(A \mid T) = \frac{1}{5}, \mathrm{P}(A) = \frac{3}{10}$을 준다. 전체 마르코프 연쇄(Markov chain)를 정확히 풀어 이를 확인하였고, 200 000번의 게임을 시뮬레이션한 결과는 0.298이었다.

흔한 실수

- p^{r-1} 대신 p^r을 쓰는 실수: 첫 앞면은 이미 주어졌으므로 앞면이 $r-1$개만 더 필요하다.
- $\mathrm{P}(A)$를 구하려면 마지막에 p와 q로 가중해야 한다는 것을 잊는 실수. 조건부 확률만으로는 답이 되지 않는다.

STEP 너머

이 논증은 마르코프 연쇄의 **첫 단계 분석**(*first-step analysis*)이다. 상태는 "현재의 면과 연속의 길이"이고, 다시 시작되는 성질이 바로 마르코프 성질(Markov property)이다. 같은 방법으로 앞면과 뒷면으로 이루어진 임의의 패턴에 대한 흡수 확률(absorption probability)과 기대 대기 시간(expected waiting time)을 구할 수 있으며, 이는 어떤 패턴이 놀랍게도 다른

패턴을 이기는 페니 게임(Penney's game)으로 이어진다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 마르코프 연쇄와 첫 단계 분석을 알맞은 수준에서 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.