풀이

STEP II 1999, 문제 14

확률과 통계 • 최적 정지

> **STEP II 1999, 14번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/99-s2-q14

핵심 아이디어

주사위에는 기억이 없다. 다시 던지기로 결정할 때마다 같은 게임의 맨 처음으로 되돌아가는 것이므로, 그 시점부터 기대할 수 있는 최선은 똑같은 수 V, 곧 게임 전체의 최대 기대 점수(expected score)이다. 그러면 매번 던진 뒤의 결정은 단순한 비교가 된다. 지금 가진 수가 V보다 크면 그 수를 갖고, 그렇지 않으면 다시 던진다. 이것을 식으로 적으면 V에 대한 방정식 하나가 나오고, 그 방정식을 풀면 값과 전략(strategy)을 모두 알 수 있다.

풀이

설정하기

1이 나온 던지기를 **강제 종료**(*forced*)라고 부르자. 게임은 그 자리에서 점수 1로 끝난다. 그 밖의 던지기 결과 $x \in \{2,3,4,5,6\}$이 나온 뒤에는 멈출지(점수 x) 계속할지를 선택한다. 모든 점수는 1과 6 사이에 있으므로, 모든 전략의 기대 점수는 $[1,6]$에 속한다.

$$V = \text{어떤 전략으로든 얻을 수 있는 가장 큰 기대 점수}$$

라 하자(엄밀히는 모든 전략에 대한 최소 상계(least upper bound)이며, 이 값이 실제로 달성됨을 곧 보게 된다).

점수 4를 얻는 전략

다음 전략을 생각하자. **4, 5, 6 중 하나가 나오는 즉시 멈춘다**(물론 1이 나와도 멈춘다). 이 전략의 기대 점수를 E라 하자. 첫 번째 던지기에 대해 조건화(conditioning)한다. 각각 확률(probability) $\frac{1}{6}$로 $1,4,5,6$이 나오고 게임은 그 점수로 끝난다. 확률 $\frac{2}{6}$로 2 또는 3이 나오는데, 이때는 다시 던지고, 주사위에는 기억이 없으므로 거기서부터의 기대 점수는 다시

E이다. 따라서

$$E = \tfrac{1}{6}(1+4+5+6) + \tfrac{2}{6}E \quad \Longrightarrow \quad \tfrac{4}{6}E = \tfrac{16}{6} \quad \Longrightarrow \quad E = 4.$$

(멈추지 않고 n번 던질 확률이 $(\tfrac{1}{3})^n \to 0$이므로 게임은 확률 1로 끝난다.) 그러므로 $V \geq 4$이다.

어떤 전략도 4를 넘을 수 없다

아무 전략이나 하나 택하자. 이 전략은 지금까지 던진 결과 전체에 의존해도 된다. 첫 번째 던지기 결과 x를 보자.

- $x = 1$이면 점수는 1이다.
- $x \geq 2$이고 전략이 멈추면 점수는 x이다.
- $x \geq 2$이고 전략이 계속하면, 이제부터 하는 일은 새로 시작한 같은 게임에 대한 어떤 전략일 뿐이다(앞선 던지기는 나중 던지기에 영향을 줄 수 없다). 따라서 여기서부터의 기대 점수는 많아야 V이다.

어느 경우든 $x \geq 2$일 때의 기대 점수는 많아야 $\max(x, V)$이다. 그러므로 모든 전략의 기대 점수는 많아야

$$g(V) = \tfrac{1}{6}\Big(1 + \sum_{x=2}^{6} \max(x, V)\Big),$$

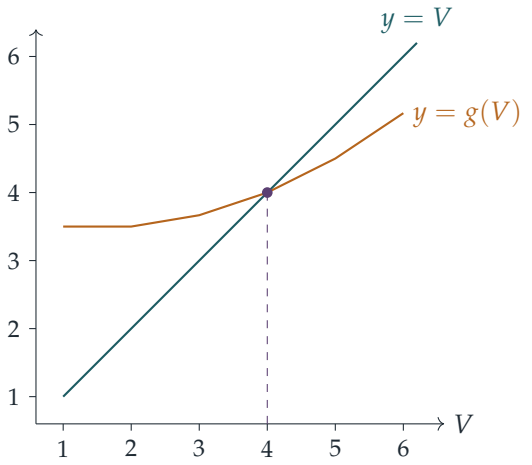
이고, 전략들에 대한 최소 상계를 취하면 $V \leq g(V)$를 얻는다.

생각해 보기 부등식(inequality) $V \leq g(V)$는 어떻게 V를 결정하는가?

$h(V) = g(V) - V$를 보자. 각 항 $\max(x, V)$는 V가 증가할 때 변화율 0 또는 1로 증가하고, 그런 항은 다섯 개뿐이므로 g의 증가율은 많아야 $\frac{5}{6}$이다. 따라서 h는 순감소한다. $V = 4$에서

$$g(4) = \tfrac{1}{6}(1+4+4+4+4+5+6) = \tfrac{24}{6} = 4,$$

이므로 $h(4) = 0$이고, 모든 $V > 4$에 대해 $h(V) < 0$이다. $h(V) \geq 0$이므로 $V \leq 4$여야 한다.



결론

두 부분을 합치면 $V = 4$이다. 가능한 최대 기대 점수는 4이고, 이는 4, 5, 6이 나오면 멈추고 2 또는 3이 나오면 다시 던져서 얻어진다. 이것은 바로 "던진 결과가 V 이상이면 멈춘다"라는 규칙이다. 2나 3은 계속해서 기대할 수 있는 4보다 나쁘고, 5나 6은 그보다 좋다.

4는 비기는 경우이다. 4에서 멈추면 4를 얻고, 계속해도 기대 점수는 4이다. 따라서 "5 또는 6에서만 멈춘다"도 똑같이 좋다. 같은 조건화로

$$E' = \tfrac{1}{6}(1+5+6) + \tfrac{3}{6}E' \quad \Longrightarrow \quad E' = 4.$$

다른 규칙은 모두 엄밀히 더 나쁘다. 예를 들어 3 이상에서 멈추면 $\frac{1+3+4+5+6}{5} = \frac{19}{5}$이고, 6에서만 멈추면 $\frac{1+6}{2} = \frac{7}{2}$이다.

검산

"1과 집합 S에 속한 눈에서 멈춘다"라는 규칙에 대해 쓸 수 있는 깔끔한 지름길이 있다. 매번 던질 때 게임이 끝날 확률은 같고, 끝나는 눈은 $\{1\} \cup S$의 어느 것이든 같은 가능성으로 나오므로, 기대 점수는 그 눈들의 평균이다. $S = \{4,5,6\}$이면 $\frac{16}{4} = 4$이고, $S = \{5,6\}$이면 $\frac{12}{3} = 4$이다. 이런 규칙 32가지를 모두 정확히 확인해 보았더니, 4를 주는 것은 이 두 가지뿐이고 그다음으로 좋은 값은 $\frac{19}{5}$였다. 4 이상에서 멈추는 게임을 400 000번 시뮬레이션(simulation)한 결과 평균은 3.998이었다.

흔한 실수

- 한 번 던진 결과의 평균인 3.5를 계속하는 것의 가치로 쓰는 것. 계속하면 다시 던질 선택권이 남아 있으므로 그 가치는 3.5보다 크다. 따라서 3.5와 비교하는 것으로는 "4 이상에

서 멈춘다"를 정당화할 수 없다.

- 1이 강제 종료라는 점을 잊고, 마치 위험 없이 영원히 던질 수 있는 것처럼 계속하는 것의 가치를 적는 것.
- 문턱값 전략(threshold strategy)만 확인하고, 지난 결과에 의존하는 더 교묘한 규칙이 왜 더 나을 수 없는지는 설명하지 않는 것. 위의 무기억성(memorylessness) 논증은 모든 전략을 다룬다.

STEP 너머

이것은 **최적 정지**(*optimal stopping*)의 작은 예이고, 방정식 $V = g(V)$는 동적 계획법(dynamic programming)과 마르코프 결정 과정(Markov decision process)의 기본 도구인 **벨만 방정식**(*Bellman equation*)이다. 같은 추론으로 금융에서 아메리칸 옵션(American option)의 가격을 매기고, 유명한 비서 문제(secretary problem)를 푼다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 여기서 쓴 조건화 논증을 다루고, Blitzstein과 Hwang의 *Introduction to Probability*(확률론 입문)에는 같은 맥락의 문제가 많이 실려 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.