풀이

STEP I 1990, 문제 14

확률과 통계 • 기댓값과 최적 정지

> **STEP I 1990, 14번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/90-s1-q14

핵심 아이디어

공을 다시 넣으므로 각 시행은 독립(independent)이고, 따라서 매번 흰 공, 빨간 공, 검은 공이 나올 확률은 항상 같은 값으로 고정되어 있다. Juggins의 경우에는 세 번의 시행에서 살아남을 확률과 살아남은 시행 한 번의 평균 점수만 알면 된다. Muggins의 경우에는 맞바꿈(trade-off)이 문제가 된다. 한 번 더 뽑으면 평균적으로 점수가 조금 늘지만 이미 쌓아 둔 점수를 모두 잃을 위험이 있고, 그 위험의 크기는 N과 함께 커진다. "뽑을 때의 기대 점수"와 "멈출 때의 점수"를 비교하면 문턱값(threshold)이 나온다.

풀이

한 번의 시행에서 공이 흰 공, 빨간 공, 검은 공인 사건을 각각 W, R, B라 하자. 공을 다시 넣으므로 각 시행은 다른 시행과 독립이고,

$$\mathrm{P}(W) = \frac{5}{10} = \frac{1}{2}, \qquad \mathrm{P}(R) = \frac{3}{10}, \qquad \mathrm{P}(B) = \frac{2}{10} = \frac{1}{5}.$$

이다.

Juggins의 전략

Juggins의 최종 점수가 0이 되는 것은 처음 세 번 안에 검은 공이 나올 때이고, 오직 그때뿐이다. 그렇지 않으면 세 번 모두 검은 공이 아니고 그는 멈춘다. 따라서

$$\mathrm{P}(\text{최종 점수}0) = 1 - \mathrm{P}(\text{세 번 모두 검은 공이 아님}) = 1 - \left(\frac{4}{5}\right)^3 = 1 - \frac{64}{125} = \frac{61}{125}.$$

이다.

기댓값(expected value)을 구하기 위해 그의 최종 점수를 S라 하자. 처음 세 번이 모두 검은

공이 아니면(확률 $\frac{64}{125}$) $S = X_1 + X_2 + X_3$이다. 여기서 X_i는 i번째 시행이 검은 공이 아니라는 조건 아래에서 그 시행의 점수이다. 검은 공이 아니라는 조건 아래에서 그 공이 흰 공일 확률은 $\frac{1/2}{4/5} = \frac{5}{8}$, 빨간 공일 확률은 $\frac{3/10}{4/5} = \frac{3}{8}$이다. 그러므로

$$\mathrm{E}(X_i) = 1 \cdot \tfrac{5}{8} + 2 \cdot \tfrac{3}{8} = \tfrac{11}{8}.$$

이다. 나머지 모든 경우에는 $S = 0$이다. 따라서

$$\mathrm{E}(S) = \tfrac{64}{125} \cdot 3 \cdot \tfrac{11}{8} = \tfrac{264}{125} = 2.112.$$

이다.

? **생각해 보기** 여기서 왜 $\frac{1}{2}$와 $\frac{3}{10}$이 아니라 $\frac{5}{8}$과 $\frac{3}{8}$을 쓰는가?

이미 살아남았다는 조건을 걸었으므로 세 번의 시행은 모두 검은 공이 아님이 알려져 있고, 따라서 확률의 합이 1이 되도록 다시 비율을 맞춰야 한다. 조건부 확률(conditional probability)을 쓰지 않고 같은 계산을 할 수도 있다. $\mathrm{E}(S)$는 흰 공과 빨간 공으로 이루어진 $2^3 = 8$가지 순서 각각에 대해 (확률)×(점수)를 더한 값이다. 각 시행은 기댓값으로 $1 \cdot \frac{1}{2} + 2 \cdot \frac{3}{10} = \frac{11}{10}$만큼 기여하고, 그 점수는 나머지 두 시행도 살아남을 때(확률 $\frac{16}{25}$)에만 유지된다. 이렇게 해도 다시 $3 \cdot \frac{11}{10} \cdot \frac{16}{25} = \frac{264}{125}$를 얻는다.

Muggins의 결정

Muggins가 N점을 가진 상태에서 한 번 더 뽑고 (그다음 멈춘다고) 하자. 그의 점수는 $N+1$, $N+2$ 또는 0이 되므로

$$\mathrm{E}(\text{한 번 더 뽑은 뒤의 점수}) = \tfrac{1}{2}(N+1) + \tfrac{3}{10}(N+2) + \tfrac{1}{5} \cdot 0 = \tfrac{4}{5}N + \tfrac{11}{10}.$$

이다. 평균적으로 뽑는 것이 멈추는 것보다 나은 것은 정확히

$$\tfrac{4}{5}N + \tfrac{11}{10} > N \iff \tfrac{11}{10} > \tfrac{1}{5}N \iff N < 5.5.$$

일 때이다. 한 번 뽑을 때의 평균 이득은 $\frac{11}{10}$이고, 평균 손실은 검은 공이 나올 확률에 걸려 있는 점수를 곱한 $\frac{1}{5}N$이다. 점수는 정수이므로, 이로부터 다음 전략을 생각할 수 있다.

점수가 5 이하인 동안은 계속 뽑고, 6 이상이 되는 즉시 멈춘다.

? **생각해 보기** 한 번 더 뽑는 경우와 멈추는 경우만 비교하는 것은 근시안적으로 보인다. $N \geq 6$에서 뽑고 나중에 또 뽑을 계획을 세우는 것이 이득일 수는 없을까?

그렇지 않다. 점수가 일단 6 이상이 되면, 이후에 몇 번을 더 뽑든 점수는 6 이상으로 남거나 (아니면 0으로 게임이 끝나므로), 이후의 모든 결정 시점에서 한 번 뽑기 비교는 다시 "멈춰

라"라고 말한다. 그런 점수에서 한 번 더 뽑을 때마다 최종 점수의 기댓값이 줄어드므로, 그 상태에서 가장 좋은 계획은 즉시 멈추는 것이다. 6 미만에서는 한 번 더 뽑는 것만으로도 이미 멈추는 것보다 낫고, 계속할 수 있는 선택지는 그 이득을 더 늘려 줄 뿐이다. 따라서 이 문턱값 규칙은 단지 합리적인 전략이 아니라 가능한 최선의 전략이다.

이 전략을 쓸 때 N에서 시작한 최종 점수의 기댓값을 $V(N)$이라 하자. 그러면 $N \geq 6$에 대해 $V(N) = N$이고, $N \leq 5$에 대해 $V(N) = \frac{1}{2}V(N+1) + \frac{3}{10}V(N+2)$이다. 위에서부터 내려가며 계산하면 다음과 같다.

N	6	5	4	3	2	1	0
$V(N)$	6	5.1	4.35	3.705	3.1575	2.690	2.292

따라서 Muggins는 한 게임에 평균 약 2.29점을 얻고, 이는 Juggins의 2.112점보다 높다.

검산

- 컴퓨터로 최대 세 번까지의 모든 시행 순서를 정확히 나열하였다. 그 결과 $\mathrm{P}(0) = \frac{61}{125} = 0.488$과 $\mathrm{E}(S) = \frac{264}{125}$가 확인되었다. 2×10^5번의 게임으로 몬테카를로(Monte Carlo) 시뮬레이션을 하였더니 0.488과 2.111이 나왔다.
- $N = 5$에서 뽑으면 $\frac{4}{5} \cdot 5 + 1.1 = 5.1 > 5$이다. $N = 6$에서는 $5.9 < 6$이다. 따라서 전환점은 5와 6 사이에 있다.
- 모든 전략에 대해 가치 반복(value iteration)을 수행하였더니 같은 규칙이 나왔다. 대신 ≥ 5나 ≥ 7에서 멈추면 평균이 각각 2.261과 2.265로, 6에서 멈출 때의 2.292보다 낮다.

흔한 실수

- $\mathrm{E}(S)$를 $3 \times \frac{11}{10}$으로 계산하는 것. 이는 나중에 검은 공이 나오면 앞의 점수가 모두 사라진다는 사실을 잊은 것이다.
- 조건부 평균 $\frac{11}{8}$에 조건 없는 확률까지 곱하는 것. 이렇게 하면 생존 인자를 두 번 세게 된다.
- 문턱값을 $N \leq 5.5$로 쓰고 N이 정수라는 점을 언급하지 않는 것. 규칙은 "5에서는 뽑고, 6에서는 멈춘다"이다.

STEP 너머

Muggins의 문제는 **최적 정지**(*optimal stopping*) 문제이다. "지금 멈추는 것이 한 단계 더 가는 것보다 나으면 멈춘다"는 규칙을 **한 단계 앞보기**(*one-step look-ahead*) 규칙이라 하며, 이

문제에서처럼 그 규칙이 멈추라고 하는 상태들의 집합을 결코 벗어날 수 없을 때에는 언제나 최적이다. 같은 아이디어는 마르코프(Markov) 결정 과정(Markov decision process)과 동적 계획법(dynamic programming)으로 이어진다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 기댓값과 마르코프 연쇄(Markov chain)의 배경 지식을 다루고, Blitzstein과 Hwang의 *Introduction to Probability*(확률론 입문)에는 이와 비슷한 첫 단계 분석(first-step analysis)이 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.