풀이

STEP I 2003, 문제 13

확률과 통계 • 기하분포의 기댓값

> **STEP I 2003, 13번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/03-s1-q13

핵심 아이디어

첫 번째 부분에서는 각 팀의 골 대신 **실축**(miss) 횟수를 센다. 각 팀이 다섯 번씩 찬 뒤 두 팀이 동점인 것은 정확히 두 팀의 실축 횟수가 같을 때이다. 기댓값(expectation)을 구할 때는 서든데스(sudden death) 단계가 두 번의 킥으로 이루어진 라운드 단위로 진행되고, 각 라운드가 같은 확률로 승부차기(shoot-out)를 끝내는 독립(independent) 시행이라는 점에 주목한다. 따라서 라운드 수는 기하분포(geometric distribution)를 따르고, 킥 횟수의 기댓값은 "항상 차는 열 번"과 "필요할 확률로 가중된 추가 킥"의 합으로 나뉜다.

풀이

모형이 의도하는 대로, 모든 킥은 다른 모든 킥과 독립이고, A팀은 항상 확률 p_A로, B팀은 확률 p_B로 득점한다고 가정한다. $q_A = 1 - p_A$, $q_B = 1 - p_B$로 둔다.

열 번의 킥 후 동점일 확률 α

A팀이 처음 다섯 번의 킥 중 실축한 횟수를 X, B팀이 실축한 횟수를 Y라 하자. 각 팀의 다섯 번의 킥은 실축 확률이 같은 독립 시행이므로

$$X \sim \mathrm{B}(5, q_A), \qquad Y \sim \mathrm{B}(5, q_B),$$

즉 이항분포(binomial distribution)를 따르고, 두 팀의 킥이 독립이므로 X와 Y도 독립이다.

A팀은 $5 - X$골, B팀은 $5 - Y$골을 넣으므로, 점수가 같은 것은 정확히 $X = Y$일 때이다.

$i = 0, 1, \ldots, 5$에 대한 사건 $\{X = Y = i\}$는 서로 배반이므로

$$\begin{aligned}\alpha = \mathrm{P}(X = Y) &= \sum_{i=0}^{5} \mathrm{P}(X = i)\,\mathrm{P}(Y = i) \\ &= \sum_{i=0}^{5} \binom{5}{i} q_A^i p_A^{5-i} \binom{5}{i} q_B^i p_B^{5-i} \\ &= \sum_{i=0}^{5} \binom{5}{i}^2 (1-p_A)^i (1-p_B)^i p_A^{5-i} p_B^{5-i},\end{aligned}$$

이고, 이것이 구하는 식이다. (대신 골 수를 세어 각 팀이 $5-i$골씩 넣는 경우를 생각해도 $\binom{5}{i} = \binom{5}{5-i}$이므로 같은 합을 얻는다.)

킥 횟수의 기댓값

열 번의 킥 후에도 동점이면 서든데스 단계는 **라운드**(round) 단위로 진행된다. 각 라운드에서 A가 한 번, B가 한 번 찬다. 한 라운드에서 두 킥 중 정확히 하나만 들어가면 승부가 결정되고, 그 확률은

$$p_A q_B + q_A p_B = p_A(1-p_B) + p_B(1-p_A) = p_A + p_B - 2p_A p_B = \beta.$$

이다. 그렇지 않으면(둘 다 득점하거나 둘 다 실축하면) 다음 라운드를 진행한다. $\beta > 0$이어야 하므로, 어떤 라운드도 승부를 결정할 수 없는 자명한 경우 $p_A = p_B = 0$과 $p_A = p_B = 1$은 제외한다.

> ? **생각해 보기** 승부차기가 라운드 도중에 끝날 수 없는 이유는 무엇인가?

규칙에 따르면 한 팀은 득점하고 다른 팀은 득점하지 못했을 때 두 팀이 찬 횟수가 같아야 한다. A가 실축하더라도 B는 여전히 자기 킥을 차야 한다(B도 실축할 수 있기 때문이다). 따라서 서든데스의 킥은 항상 짝을 이루어 나오고, 전체 킥 횟수는 $10 + 2R$이다. 여기서 R은 진행된 라운드 수이다(처음 열 번의 킥에서 승부가 결정되면 $R = 0$).

서든데스에 도달했다는 조건 아래에서 각 라운드는 확률 β로 승부가 결정되는 독립 시행이므로, R은 기하분포를 따른다.

$$\mathrm{P}(R = k \mid \text{열 번 후 동점}) = (1-\beta)^{k-1}\beta, \qquad k = 1, 2, 3, \ldots$$

그 평균은

$$\mathrm{E}(R \mid \text{열 번 후 동점}) = \sum_{k=1}^{\infty} k\beta(1-\beta)^{k-1} = \beta \cdot \frac{1}{(1-(1-\beta))^2} = \frac{1}{\beta},$$

이다. 여기서 $|x| < 1$일 때 성립하는 $\sum_{k\geq 1} kx^{k-1} = (1-x)^{-2}$를 이용했으며, $0 < \beta \leq 1$이므로 이를 적용할 수 있다.

처음 열 번의 킥에서 승부가 결정되면 $R = 0$이다. 열 번의 킥 후 동점인지 아닌지로 조건을 나누면

$$\mathrm{E}(R) = \alpha \cdot \frac{1}{\beta} + (1-\alpha)\cdot 0 = \frac{\alpha}{\beta}.$$

따라서 킥 횟수의 기댓값은

$$\mathrm{E}(10+2R) = 10 + \frac{2\alpha}{\beta},$$

이고, 이것이 증명하려던 결과이다.

검산

- 두 팀의 실력이 같아 $p_A = p_B = \frac{1}{2}$인 경우: 방데르몽드(Vandermonde) 항등식(Vandermonde's identity)에 의해 $\sum_i \binom{5}{i}^2 = \binom{10}{5} = 252$이므로 $\alpha = 252/2^{10} = \frac{63}{256}$이고, $\beta = \frac{1}{2}$이다. 킥 횟수의 기댓값은 $10 + \frac{63}{64} = \frac{703}{64} \approx 10.98$이다.
- 한 팀은 완벽하고 다른 팀은 전혀 못 넣는 경우($p_A = 1$, $p_B = 0$), $\alpha = 0$이다($i = 0$인 항만 남을 수 있는데, 그 항에는 $p_B^5 = 0$이 들어 있다). 따라서 상식대로 정확히 10번의 킥만 이루어진다.
- α에 대한 식은 처음 열 번의 킥에서 나올 수 있는 2^{10}가지 결과를 모두 직접 나열하여 확인했고, $10 + 2\alpha/\beta$는 2×10^5번의 승부차기 시뮬레이션으로 확인했다. $p_A = \frac{3}{4}$, $p_B = \frac{2}{3}$일 때 식은 11.264, 시뮬레이션은 11.25를 준다.

흔한 실수

- β를 A가 한 라운드를 이길 확률 $p_A(1-p_B)$로 잡는 것. 승부차기는 어느 팀이든 한 라운드를 이기면 끝난다.
- 서든데스가 확률 α로만 일어난다는 것을 잊고 답을 $10 + 2/\beta$로 쓰는 것.
- 서든데스의 각 단계에서 킥을 두 번이 아니라 한 번으로 세는 것.

STEP 너머

$\mathrm{E}(R) = \alpha\,\mathrm{E}(R \mid \text{동점}) + (1-\alpha)\,\mathrm{E}(R \mid \text{동점 아님})$으로 나누는 단계는 전체 기댓값의 법칙(law of total expectation)으로, 대학 1학년 확률론에서 가장 유용한 도구 중 하나이다. 서든데스 단계는 흡수 상태(absorbing state)를 가진 간단한 마르코프 연쇄(Markov chain)이기도 하며, 여기서 한 "흡수까지 걸리는 시간의 기댓값" 계산은 랜덤워크(random walk)와 도박꾼의 파산(gambler's ruin) 문제로 일반화된다. Grimmett와 Welsh의 *Probability: An*

Introduction(확률론 입문)은 조건부 기댓값(conditional expectation)과 이러한 연쇄를 모두 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.