풀이

STEP III 1994, 문제 12

확률과 통계 • 첫걸음 분석

> **STEP III 1994, 12번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/94-s3-q12

핵심 아이디어

게임을 두 개씩 묶어서 본다. 2점 차의 리드는 짝수 번의 게임이 끝난 뒤에만 나타날 수 있고, 게임 두 개가 끝날 때마다 두 선수는 각각 한 번씩 서브(serve)를 넣었으므로 서브 차례가 처음 상태로 돌아온다. 그러면 경기는 단순한 반복 시행이 된다. 각 묶음은 A가 둘 다 이기거나, B가 둘 다 이기거나, 하나씩 나누어 갖는데, 마지막 경우에는 다시 동점이 된다. 표준 테니스도 처음 두 묶음만 따로 다룬다는 점을 빼면 같은 구조이다.

풀이

이 풀이 전체에서

$$a = p_A p_B, \qquad b = q_A q_B, \qquad s = p_A q_B + q_A p_B.$$

라고 쓴다. 두 선수가 한 번씩 서브를 넣는 연속된 두 게임에서, a는 A가 두 게임을 모두 이길 확률, b는 B가 두 게임을 모두 이길 확률, s는 두 사람이 한 게임씩 이길 확률이다(마지막 경우는 A와 B가 모두 자기 서브 게임을 지키거나, 둘 다 상대 서브 게임을 브레이크하는 경우이다). 문제의 의도대로 서로 다른 게임의 결과는 독립(independent)이라고 가정한다. 가능한 결과는 이 세 가지뿐이므로

$$a + b + s = (p_A + q_A)(p_B + q_B) = 1.$$

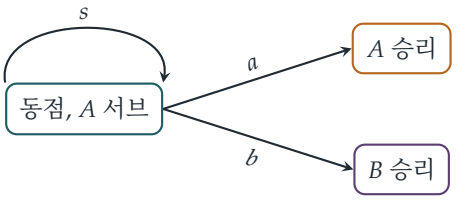
이다.

단축 테니스

점수 차는 게임마다 ± 1씩 변하므로, 짝수 번의 게임 뒤에는 짝수이고 홀수 번의 게임 뒤에는 홀수이다. 따라서 정확히 2점 차의 리드는 $2, 4, 6, \ldots$ 번째 게임 뒤에만 처음 생길 수 있고,

짝수 번의 게임 뒤에도 경기가 계속되고 있다면 점수는 동점이다.

그러므로 경기를 두 게임씩 묶는다(1–2번째 게임, 3–4번째 게임, ...). A가 서브할 차례인 동점 상태에서 시작하면, 다음 묶음에서 확률 a로 A의 승리로 경기가 끝나고, 확률 b로 B의 승리로 끝나며, 확률 s로 다시 A가 서브할 차례인 동점 상태로 돌아온다. 이때의 상황은 처음과 완전히 같으므로



$$P_{\text{short}} = a + sP_{\text{short}} \quad \Longrightarrow \quad P_{\text{short}} = \frac{a}{1-s} = \frac{a}{a+b} = \frac{p_A p_B}{p_A p_B + q_A q_B},$$

이다. 여기서 $1 - s = a + b$를 이용하였다. ($1 - s$로 나누는 것은 $s = 1$이 아닌 한 문제없다. $s = 1$은 모든 묶음이 나누어져 경기가 영원히 끝나지 않는 퇴화된 경우이다.)

B가 먼저 서브하면, 각 묶음은 B의 서브 게임 뒤에 A의 서브 게임이 오는 형태가 된다. A가 둘 다 이길 확률, B가 둘 다 이길 확률, 하나씩 나누어 가질 확률은 각각 $p_B p_A$, $q_B q_A$, $q_B p_A + p_B q_A$이다. 묶음 안에서 두 게임의 순서는 상관없으므로 이는 같은 a, b, s이다. 같은 논증으로 같은 답이 나오므로, P_{short}는 누가 먼저 서브하는지에 의존하지 않는다.

4게임 만에 승부가 나는 경우

4게임이 끝나기 전에는 승부가 날 수 없다. 4게임(두 묶음, A가 먼저 서브) 뒤의 점수 차는 짝수이고, 점수가 4–0, 3–1, 1–3, 0–4일 때에만 정확히 승부가 난다. 두 묶음으로 나타내면 다음과 같다.

- A의 4–0: A가 두 묶음을 모두 이기는 경우로, 확률은 $a^2 = p_A^2 p_B^2$이다.
- 0–4: B가 두 묶음을 모두 이기는 경우로, 확률은 $b^2 = q_A^2 q_B^2$이다.
- A의 3–1: A가 한 묶음을 이기고 다른 묶음은 나누어지는 경우로, 순서는 어느 쪽이든 되며 확률은 $2as$이다.
- 1–3: 마찬가지로 $2bs$이다.

이들을 더하면

$$\Pr(\text{4게임 만에 승부가 남}) = p_A^2 p_B^2 + q_A^2 q_B^2 + 2(p_A p_B + q_A q_B)(p_A q_B + q_A p_B),$$

이므로 요구된 결과를 얻는다.

확률이 같을 때의 표준 테니스

4게임 만에 승부가 나지 않으면 점수는 2–2이다. 이는 확률 $s^2 + 2ab$로 일어난다(두 묶음이 모두 나누어지거나, 두 선수가 순서에 관계없이 한 묶음씩 이기는 경우). 2–2에서는 "2점 차로 앞서야 한다"는 규칙이 동점에서 시작하는 단축 테니스 규칙과 정확히 같고, 5번째 게임은 A가 서브한다. 따라서 그 시점부터 A가 이길 확률은 P_{short}이다. 그러므로 일반적으로

$$P_{\text{stan}} = a^2 + 2as + (s^2 + 2ab)P_{\text{short}}.$$

이다. 이제 $p_A = p_B = p$, $q_A = q_B = q$로 두면 $a = p^2$, $b = q^2$, $s = 2pq$이고 $P_{\text{short}} = \dfrac{p^2}{p^2 + q^2}$이다. 그러면

$$P_{\text{stan}} = p^4 + 4p^3q + \frac{6p^4q^2}{p^2 + q^2}.$$

이다.

> **생각해 보기** 모든 것을 p와 q의 거듭제곱으로 전개하지 않고 P_{stan}과 P_{short}를 비교하는 방법이 있을까?

있다. 단축 테니스를 같은 처음 네 게임에 대해 펼쳐 쓰면 된다. $P_{\text{short}} = a + sP_{\text{short}}$를 두 번 적용하면

$$P_{\text{short}} = a + as + s^2P_{\text{short}}.$$

이다. 말로 하면, A가 첫 번째 묶음을 이기거나, 첫 묶음이 나누어지고 A가 두 번째 묶음을 이기거나, 두 묶음이 모두 나누어져 처음부터 다시 시작한다는 뜻이다. P_{stan}의 식에서 이 식을 빼면 s^2P_{short} 항이 소거된다.

$$\begin{aligned} P_{\text{stan}} - P_{\text{short}} &= a^2 + as - a + 2abP_{\text{short}} \\ &= a(a + s - 1) + 2abP_{\text{short}} \\ &= -ab + \frac{2a^2b}{a + b} \qquad (a + s - 1 = -b\text{이므로}) \\ &= \frac{ab(a - b)}{a + b}. \end{aligned}$$

이 식은 임의의 p_A, p_B에 대해 성립한다. $a = p^2$, $b = q^2$를 대입하면

$$P_{\text{stan}} - P_{\text{short}} = \frac{p^2q^2(p^2 - q^2)}{p^2 + q^2} = \frac{p^2q^2(p - q)(p + q)}{p^2 + q^2} = \frac{p^2q^2(p - q)}{p^2 + q^2},$$

이다. 마지막 등호는 $p + q = 1$이기 때문이다. 이것이 요구된 결과이다.

그 이유는 계산 과정에서 드러난다. 두 방식은 A가 첫 번째 묶음을 이기고 B가 두 번째 묶음을 이기는 경우, 또는 그 반대인 경우에만 차이가 난다. 단축 테니스는 이 경우 첫 번째 묶음을 이긴 사람에게 승리를 주고, 표준 테니스는 2–2에서 경기를 계속한다. 두 순서 모두 확률이 ab이므로 차이는 $ab(2P_{\text{short}} - 1)$이고, 이는 더 강한 선수에게 유리하다.

검산

- $p = q = \frac{1}{2}$이면 차이는 0이다. 대칭성(symmetry)에 의해 당연히 그래야 한다. 두 확률이 모두 $\frac{1}{2}$이기 때문이다.
- $p = 0.6$이 면 $P_{\text{short}} = 0.36/0.52 \approx 0.6923$이 고 $P_{\text{stan}} = 0.1296 + 0.3456 + 0.124416/0.52 \approx 0.7145$이며, 그 차이 0.02215는 $0.0576 \times 0.2/0.52$와 일치한다.
- 4게임 만에 승부가 날 확률 $a^2 + b^2 + 2(a+b)s$와 2–2에 도달할 확률 $s^2 + 2ab$를 더하면, 당연히 그래야 하듯이 $(a+b+s)^2 = 1$이다.
- 실제 규칙을 따르는 컴퓨터 시뮬레이션(simulation)에서 $p_A = 0.7$, $p_B = 0.45$로 두면, 누가 먼저 서브하든 A가 단축 테니스에서 이기는 비율은 약 65.6%로, $0.315/0.48 = 0.65625$와 일치한다.

흔한 실수

- 3–1이 두 가지 순서(나눈 뒤 이기기, 또는 이긴 뒤 나누기)로 나올 수 있다는 것을 잊고 인수 2를 빠뜨린다.
- 게임을 이길 확률을 모든 게임에서 같다고 취급한다. 이 확률은 누가 서브하는지에 따라 달라지며, 바로 그 때문에 두 게임씩 묶는 것이 매우 유용하다.
- 무한 등비급수(geometric series)를 쓰면서, 그것이 결국 방정식 $P = a + sP$일 뿐이라는 점을 짚지 않는다.

STEP 너머

묶음 단위로 따지는 이 논증은 흡수 상태(absorbing state)가 있는 **마르코프 연쇄**(Markov chain)의 첫 번째 예이며, 방정식 $P = a + sP$는 흡수 확률(absorption probability)을 구하는 표준적인 방법인 첫 단계 분석(first-step analysis)이다. 같은 아이디어로 테니스 한 세트 전체나 스쿼시 한 경기를 이길 확률도 구할 수 있으며, 더 나아가 도박꾼의 파산(gambler's ruin) 문제와 무작위 행보(random walk)로 이어진다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 바로 이 수준에서 무작위 행보와 마르코프 연쇄를 전개한다.

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.