풀이

STEP II 2010, 문제 13

확률과 통계 • 독립사건

STEP II 2010, 13번 문제

더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:

step.maths.org/questions/10-s2-q13

핵심 아이디어

로절린드는 파딥을 상대로 적어도 한 판을 이기고, **동시에** 퀜틴을 상대로도 적어도 한 판을 이길 때, 그리고 그때에만 합격한다. "적어도 하나"는 언제나 여사건(complement)으로 다루는 것이 가장 쉽다. n판 가운데 적어도 한 판을 이길 확률은 1에서 모든 판을 질 확률을 뺀 것이다. 각 판이 독립(independent)이면 각 전략의 성공 확률은 이런 인수 두 개의 곱이 되고, 전략을 비교하는 일은 차를 인수분해(factorise)하는 일이 된다. 질 확률을 $a = 1 - p$, $b = 1 - q$로 쓰면 계산이 깔끔해진다.

풀이

서로 다른 판의 결과는 독립이고, 파딥(또는 퀜틴)과의 각 판은 이전에 무슨 일이 있었든 확률 p(또는 q)로 이긴다고 가정한다. 순서는 상관없다. 자격을 얻은 뒤에 경기를 멈추든 멈추지 않든, 그녀가 자격을 얻는 결과는 정확히 똑같다.

로절린드가 파딥과 m번, 퀜틴과 n번 경기한다고 하자. 파딥을 상대로 적어도 한 판을 이기는 사건을 A, 퀜틴을 상대로 적어도 한 판을 이기는 사건을 B라 하자. A가 일어나지 않는 것은 m판을 모두 질 때뿐이므로 $\mathrm{P}(A) = 1 - (1-p)^m$이고, 마찬가지로 $\mathrm{P}(B) = 1 - (1-q)^n$이다. 두 사건은 서로 다른 판에 관한 것이므로 독립이고,

$$\mathrm{P}(\text{합격}) = (1 - (1-p)^m)(1 - (1-q)^n).$$

이제부터 $a = 1 - p$, $b = 1 - q$로 쓴다. $0 < p < q < 1$이므로 $0 < b < a < 1$이다.

소문항 (i)

파딥과 두 번, 퀜틴과 한 번 경기하면 성공 확률은 $(1-a^2)(1-b)$이고, 퀜틴과 두 번, 파딥과 한 번 경기하면 $(1-a)(1-b^2)$이다. 각 제곱의 차를 인수분해하면

$$\begin{aligned}(1-a^2)(1-b)-(1-a)(1-b^2) &= (1-a)(1-b)[(1+a)-(1+b)]\\ &= (1-a)(1-b)(a-b)\\ &= pq(q-p).\end{aligned}$$

$p, q > 0$이고 $q > p$이므로 이 값은 양수이다. 따라서 파딥과 두 번 경기하는 쪽이 합격할 확률이 더 크고, 그녀는 그 방법을 택해야 한다.

소문항 (ii)

세 전략의 성공 확률은 다음과 같다.

$$S_1 = (1-a^3)(1-b), \qquad S_2 = (1-a^2)(1-b^2), \qquad S_3 = (1-a)(1-b^3).$$

$1-x^k = (1-x)(1+x+\cdots+x^{k-1})$이므로 각각은 인수 $(1-a)(1-b) = pq$를 가진다.

전략 1은 언제나 전략 3보다 낫다. 다음이 성립한다.

$$\begin{aligned}S_1 - S_3 &= pq[(1+a+a^2)-(1+b+b^2)]\\ &= pq\,(a-b)(1+a+b) > 0,\end{aligned}$$

$a > b$이기 때문이다. 따라서 전략 3은 결코 최선의 선택이 아니며, 전략 1과 전략 2만 비교하면 된다.

전략 1과 전략 2의 비교. 다음이 성립한다.

$$S_1 - S_2 = pq[(1+a+a^2)-(1+a)(1+b)] = pq[a^2 - b(1+a)].$$

다시 p와 q로 나타내면

$$S_1 - S_2 = pq\,D, \qquad \text{여기서} \qquad D = (1-p)^2 - (1-q)(2-p).$$

$pq > 0$이므로 전략 1이 전략 2보다 나은 것은 정확히 $D > 0$일 때이다.

생각해 보기 조건 $q - p > \frac{1}{2}$는 D를 양수로 만드는 데 어떻게 도움이 되는가?

D에서 q를 포함하는 항은 $-(1-q)(2-p)$뿐이고, $2-p > 0$이므로 이 항은 q가 커질수록 커진다. 따라서 q를 조건이 허용하는 가장 작은 값으로 바꾸어 본다. $q - p > \frac{1}{2}$이면 $1-q < \frac{1}{2} - p$이고, $2-p > 0$이므로

$$(1-q)(2-p) < (\tfrac{1}{2}-p)(2-p) = 1 - \tfrac{5}{2}p + p^2.$$

그러므로

$$D > (1 - 2p + p^2) - (1 - \tfrac{5}{2}p + p^2) = \tfrac{1}{2}p > 0.$$

따라서 $S_1 > S_2$이고, $S_1 > S_3$과 함께 이는 $q - p > \frac{1}{2}$일 때 로절린드가 전략 1을 택해야 함을 보여 준다.

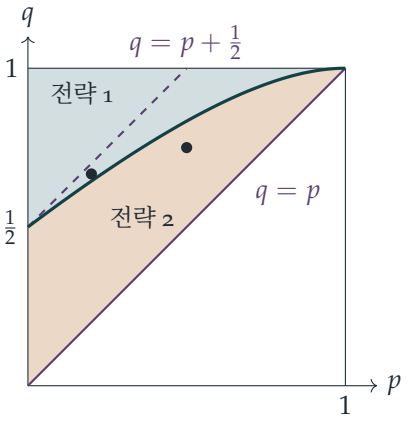
$q - p < \frac{1}{2}$이면 어느 쪽이든 나을 수 있다. 두 가지 예를 보자.

- $p = \frac{1}{2}, q = \frac{3}{4}$ (따라서 $q - p = \frac{1}{4}$). 이때 $D = \frac{1}{4} - \frac{1}{4} \cdot \frac{3}{2} = -\frac{1}{8} < 0$이다. 실제로 $S_1 = \frac{7}{8} \cdot \frac{3}{4} = \frac{42}{64}$이고 $S_2 = \frac{3}{4} \cdot \frac{15}{16} = \frac{45}{64}$이므로 **전략 2가 더 낫다**.
- $p = \frac{1}{5}, q = \frac{2}{3}$ (따라서 $q - p = \frac{7}{15} < \frac{1}{2}$). 이때 $D = \frac{16}{25} - \frac{1}{3} \cdot \frac{9}{5} = \frac{16}{25} - \frac{15}{25} = \frac{1}{25} > 0$이다. 실제로 $S_1 = \frac{61}{125} \cdot \frac{2}{3} = \frac{122}{375}$이고 $S_2 = \frac{9}{25} \cdot \frac{8}{9} = \frac{120}{375}$이므로 **전략 1이 더 낫다**.

(p, q) 평면에서 보면 전체 그림이 분명해진다. $D = 0$으로 놓으면 경계 곡선 $q = 1 - \dfrac{(1-p)^2}{2-p}$를 얻는다. 다음을 확인할 수 있다.

$$\left(p + \tfrac{1}{2}\right) - \left(1 - \frac{(1-p)^2}{2-p}\right) = \frac{p}{2(2-p)} > 0,$$

따라서 이 곡선은 직선 $q = p + \frac{1}{2}$의 바로 아래에 놓인다. 곡선 위쪽에서는 전략 1이 최선이고, 곡선과 대각선 $q = p$ 사이에서는 전략 2가 최선이다.



두 점은 위의 예이다. $(\frac{1}{2}, \frac{3}{4})$는 전략 2 영역에 있고, $(\frac{1}{5}, \frac{2}{3})$는 곡선 바로 위에 있다.

검산

소문항 (i)에서 $q \to 1$로 보내 보자. 퀜틴과의 경기는 반드시 이기므로 파딥과의 경기만 중요하고, 그 경기가 많을수록 분명히 유리하다. 이는 차 $pq(q-p) \to p(1-p) > 0$과 일치한다. 소문항 (ii)에서는 무작위로 뽑은 $p < q$ 쌍 2000개에 대해 가능한 2^4가지 결과를 모두

분수로 정확히 나열하여 S_1, S_2, S_3의 식과 인수분해를 확인하였다. $p = \frac{1}{2}, q = \frac{3}{4}$로 한 시뮬레이션(simulation)에서는 $S_1 \approx 0.657$, $S_2 \approx 0.703$을 얻었고, 정확한 값은 $\frac{21}{32} = 0.65625$와 $\frac{45}{64} \approx 0.7031$이다.

흔한 실수

- "적어도 한 번 이길 확률"을 $1 - (1-p)^m$이 아니라 mp로 계산하는 실수. 서로 겹치는 사건의 확률을 더하면 중복해서 세게 된다.
- 전략 3을 잊는 실수. 전략 1이 최선임을 보이려면 다른 **두** 전략 모두보다 나음을 보여야 한다.
- $q - p < \frac{1}{2}$인 예를 들면서 조건 $0 < p < q < 1$을 확인하지 않는 실수.

STEP 너머

직관적으로는 "가장 약한 곳에 힘을 쏟아라"라고 생각하기 쉽지만, 두 번째 소문항은 그것이 언제나 옳지는 않음을 보여 준다. 파딥과의 추가 경기는 수확 체감(diminishing returns)을 겪으며, p가 너무 작지 않으면 퀜틴과의 두 번째 경기가 더 가치 있다. 정해진 시행 횟수를 배분하여 "적어도 한 번 성공할" 확률들의 곱을 최대화하는 문제는 신뢰성 이론(reliability theory)과 실험 계획법(design of experiments)의 바탕이 되는 최적화(optimisation) 문제의 작은 사례이다. Grimmett과 Welsh의 *Probability: An Introduction*(확률론 입문)은 독립과 여사건을 꼼꼼하게 다루며, 이런 성격의 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.