풀이

STEP II 2013, 문제 13

확률과 통계 • 동전 던지기에서의 연속된 결과

> **STEP II 2013, 13번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/13-s2-q13

핵심 아이디어

사건(event) $S = n$은 처음 $n+1$번의 던지기 결과가 무엇인지 정확히 말해 준다. 즉 같은 결과가 n번 나온 뒤, 그 연속(run)을 끊는 결과가 한 번 나온다. $A = n$도 마찬가지이다. 번갈아 나오는 결과가 n번 이어진 뒤, 같은 결과가 반복된다. 따라서 각 확률(probability)은 가능한 첫 번째 결과마다 하나씩, 두 곱의 합이다. 그러면 S와 A를 비교하는 일은 네 항으로 된 차를 인수분해(factorisation)하는 문제로 귀결되며, q를 $1-p$로 바꾸지 않고 p와 q를 독립된 문자로 그대로 두면 인수분해가 더 잘 보인다.

풀이

전체에 걸쳐 각 던지기는 독립(independent)이고, 앞면이 나올 확률은 p, 뒷면이 나올 확률은 q이며, $p, q > 0$이고 $p \neq q$이다.

일반 공식

$S = n$은 정확히 $1, \ldots, n$번째 결과가 모두 첫 번째 결과와 같고 $n+1$번째 결과가 다를 때 일어난다. 첫 번째가 앞면이면 이는 $\mathrm{H}^n\mathrm{T}$이고 확률은 $p^n q$이다. 첫 번째가 뒷면이면 $\mathrm{T}^n\mathrm{H}$이고 확률은 $q^n p$이다. 두 경우는 서로 배반(disjoint)이므로

$$\mathrm{P}(S = n) = p^n q + q^n p = pq\left(p^{n-1} + q^{n-1}\right).$$

$A = n$은 정확히 $1, \ldots, n$번째 결과가 번갈아 나오고 $n+1$번째 결과가 n번째 결과와 **같아서** 교대가 끝날 때 일어난다.

- $n = 2m$(짝수)일 때: 연속 $\mathrm{HTHT}\cdots\mathrm{HT}$에는 앞면이 m번, 뒷면이 m번 있고 T로 끝나므로, 다음 결과는 반드시 T여야 하며 확률은 $p^m q^m \cdot q$이다. 뒷면으로 시작하면 $q^m p^m \cdot p$이

다. 따라서

$$\mathrm{P}(A = 2m) = p^m q^m (p + q).$$

- $n = 2m + 1$(홀수)일 때: 연속 HTH$\cdots$H에는 앞면이 $m + 1$번, 뒷면이 m번 있고 H로 끝나므로, 다음 결과는 반드시 H여야 하며 확률은 $p^{m+2}q^m$이다. 뒷면으로 시작하면 $q^{m+2}p^m$이다. 따라서

$$\mathrm{P}(A = 2m + 1) = p^m q^m \left(p^2 + q^2\right).$$

소문항 (i)

$A = 1$은 두 번째 결과가 첫 번째 결과를 반복한다는 뜻, 즉 HH 또는 TT이다. 따라서 $\mathrm{P}(A = 1) = p^2 + q^2$이다(위에서 $m = 0$인 경우). $S = 1$은 두 번째 결과가 첫 번째 결과와 다르다는 뜻, 즉 HT 또는 TH이므로

$$\mathrm{P}(S = 1) = pq + qp = 2pq.$$

그러면

$$\mathrm{P}(A = 1) - \mathrm{P}(S = 1) = p^2 - 2pq + q^2 = (p - q)^2 > 0,$$

이며, $p \neq q$이므로 부등호는 엄격하다. 따라서 $\mathrm{P}(S = 1) < \mathrm{P}(A = 1)$이다.

소문항 (ii)

공식으로부터

$$\mathrm{P}(S = 2) = p^2 q + q^2 p = pq(p + q) = \mathrm{P}(A = 2),$$

이므로 요구된 결과가 성립한다. 길이가 3일 때는

$$\mathrm{P}(S = 3) = p^3 q + q^3 p = pq\left(p^2 + q^2\right) = \mathrm{P}(A = 3),$$

이므로 이 두 확률도 같다.

소문항 (iii)

짝수 길이. $n > 1$이라 하자. 두 확률을 모두 $pq > 0$으로 나누면

$$\frac{\mathrm{P}(S = 2n)}{pq} = p^{2n-1} + q^{2n-1}, \qquad \frac{\mathrm{P}(A = 2n)}{pq} = p^{n-1}q^{n-1}(p + q) = p^n q^{n-1} + p^{n-1}q^n.$$

? **생각해 보기** $p^{2n-1} + q^{2n-1} - p^n q^{n-1} - p^{n-1}q^n$의 부호를 어떻게 알 수 있을까?

각 순수한 거듭제곱을, 같은 문자의 가장 높은 거듭제곱을 공유하는 교차항과 짝지어 보자.

$$\begin{aligned} p^{2n-1} + q^{2n-1} - p^n q^{n-1} - p^{n-1} q^n &= p^n \left(p^{n-1} - q^{n-1}\right) - q^n \left(p^{n-1} - q^{n-1}\right) \\ &= \left(p^n - q^n\right)\left(p^{n-1} - q^{n-1}\right). \end{aligned}$$

$p, q > 0$이고 $p \neq q$이므로, 임의의 $k \geq 1$에 대하여 수 $p^k - q^k$는 0이 아니고 $p - q$와 부호가 같다($x \mapsto x^k$가 $x > 0$에서 순증가(strictly increasing)하기 때문이다). 여기서 $n \geq 2$이므로 n과 $n-1$이 모두 1 이상이고, 두 괄호는 0이 아니며 부호가 같다. 따라서 그 곱은 양수이고, 다시 $pq > 0$을 곱하면

$$\mathrm{P}(S = 2n) > \mathrm{P}(A = 2n) \qquad (n > 1).$$

($n = 1$이면 두 번째 괄호는 $p^0 - q^0 = 0$이 되며, 이것이 소문항 (ii)의 등식이다.)

홀수 길이. 같은 방법으로, 역시 $n > 1$일 때

$$\frac{\mathrm{P}(S = 2n+1)}{pq} = p^{2n} + q^{2n},$$

$$\frac{\mathrm{P}(A = 2n+1)}{pq} = p^{n-1} q^{n-1}\left(p^2 + q^2\right) = p^{n+1} q^{n-1} + p^{n-1} q^{n+1},$$

이고

$$\begin{aligned} p^{2n} + q^{2n} - p^{n+1} q^{n-1} - p^{n-1} q^{n+1} &= p^{n+1}\left(p^{n-1} - q^{n-1}\right) - q^{n+1}\left(p^{n-1} - q^{n-1}\right) \\ &= \left(p^{n+1} - q^{n+1}\right)\left(p^{n-1} - q^{n-1}\right). \end{aligned}$$

같은 논증에 의해 두 괄호는 모두 0이 아니며 $p - q$와 부호가 같으므로

$$\mathrm{P}(S = 2n+1) > \mathrm{P}(A = 2n+1) \qquad (n > 1).$$

이번에도 인수 $p^{n-1} - q^{n-1}$은 $n = 1$에서 0이 되며, 이는 $\mathrm{P}(S = 3) = \mathrm{P}(A = 3)$과 일치한다.

정리하면, 길이가 1인 같은 면의 연속은 교대 연속보다 일어날 가능성이 작고, 길이 2와 3에서는 가능성이 같으며, 4 이상의 모든 길이에서는 같은 면의 연속이 더 일어나기 쉽다.

검산

- $p = \frac{1}{3}, q = \frac{2}{3}$이라 하자. 그러면 $\mathrm{P}(S = 4) = pq(p^3 + q^3) = \frac{2}{9} \cdot \frac{1}{3} = \frac{2}{27} \approx 0.074$이고 $\mathrm{P}(A = 4) = p^2 q^2 = \frac{4}{81} \approx 0.049$이므로, 실제로 $S = 4$가 더 일어나기 쉽다.
- 모든 $n \geq 1$에 대하여 $\mathrm{P}(S = n)$을 더하면 $q \cdot \frac{p}{1-p} + p \cdot \frac{q}{1-q} = p + q = 1$이 되어, 마땅히 그래야 하는 대로 나온다.
- 각 공식은 여러 p 값에 대해 14번 던지기의 모든 수열을 정확한 분수로 열거하여, 길이 13까지 컴퓨터로 확인하였다.

흔한 실수

- 연속을 **끝내는** 던지기를 잊는 것. $S = n$이 되려면 $n+1$번째 결과가 달라야 한다. 이것이 없으면 $\mathrm{P}(S \geq n)$을 계산하게 된다.
- 교대 연속을 엉뚱한 문자로 끝내는 것. 홀수 길이의 연속은 시작한 문자와 같은 문자로 끝나고, 짝수 길이의 연속은 다른 문자로 끝난다.
- 소문항 (iii)에서 $q = 1 - p$를 대입하는 것. 그러면 식이 p에 대한 다항식이 되어 인수분해를 알아보기가 훨씬 어려워진다.

STEP 너머

S는 기하분포형 확률변수(geometric-type random variable)이다. 첫 번째 결과가 주어지면, 연속의 길이는 첫 번째 변화가 일어날 때까지의 대기 시간(waiting time)이다. 이와 같은 연속은 통계학의 런 검정(runs test)의 바탕이 되는데, 이 검정은 결과의 수열이 무작위로 보이는지 판단하는 데 쓰인다. 또한 생성함수(generating function)를 이용하면 가장 긴 연속의 전체 분포(distribution)를 깔끔하게 구할 수 있다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 생성함수와 대기 시간을 알맞은 수준에서 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.