풀이

STEP I 2004, 문제 14

확률과 통계 • 배열의 개수 세기

STEP I 2004, 14번 문제

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

step.maths.org/questions/04-s1-q14

핵심 아이디어

상자 안의 동전을 한 개씩 꺼내 모두 한 줄로 늘어놓는다고 생각하자. 해적들은 이 줄을 두 개의 납 동전에서 끊을 뿐이다. 첫째 해적은 첫 번째 납 동전 앞의 금화를, 둘째 해적은 두 납 동전 사이의 금화를, 셋째 해적은 두 번째 납 동전 뒤의 금화를 가진다. 동전의 모든 순서가 일어날 가능성이 같으므로, 결국 문제는 두 납 동전이 놓일 수 있는 자리를 세는 것으로 귀결된다.

풀이

실제로는 셋째 해적이 남은 동전을 그냥 쓸어 담지만, $n+2$개의 동전을 모두 무작위 순서로 꺼낸다고 상상하자. 그렇게 해도 누가 어떤 동전을 가지는지는 달라지지 않는다. 꺼내는 순서가 무작위이므로, ($n+2$개의 자리 중에서) 납 동전 두 개의 자리를 고르는 $\binom{n+2}{2}$가지 방법은 모두 일어날 가능성이 같다.

첫째, 둘째, 셋째 해적이 가져간 금화의 개수를 각각 A, B, C라 하자. 그러면 $A+B+C=n$이고, 동전의 줄은 다음과 같다.



여기서 금화는 호박색, 납 동전은 자주색으로 나타내었다. 납 동전의 자리가 (A, B, C)를 결정하고, $A+B+C=n$을 만족하는 음이 아닌 정수의 순서쌍 각각은 정확히 한 가지 자리 선택에서 나온다. 따라서 $\binom{n+2}{2} = \frac{1}{2}(n+1)(n+2)$개의 순서쌍 (A, B, C)는 모두 일어날 가능성이 같다.

소문항 (i)

첫째 해적이 금화를 하나도 얻지 못하는 것은 정확히 $A = 0$일 때, 즉 맨 처음 꺼낸 동전이 납 동전일 때이다. $n+2$개의 동전 각각이 처음 나올 가능성이 같으므로 그 확률(probability)은 $\frac{2}{n+2}$이다. 따라서

$$\mathrm{P}(A \geq 1) = 1 - \frac{2}{n+2} = \frac{n}{n+2}.$$

(경우의 수를 세어도 같은 답을 얻는다. $A = 0$이면 $B + C = n$이 남고, 이 방정식의 해는 $n+1$개이며, $(n+1)/\frac{1}{2}(n+1)(n+2) = \frac{2}{n+2}$이다.)

소문항 (ii)

> **생각해 보기** 둘째 해적은 무작위한 시점에 동전을 꺼내기 시작한다. 그녀가 정말 첫째 해적보다 불리하지 않을까?

그렇다. 불리하지 않다. $B = 0$인 순서쌍을 세어 보자. 이때 $A + C = n$이고, 이 방정식의 해는 소문항 (i)과 똑같이 $n+1$개이다. 따라서

$$\mathrm{P}(B \geq 1) = 1 - \frac{n+1}{\frac{1}{2}(n+1)(n+2)} = \frac{n}{n+2}.$$

사실 모든 순서쌍 (A, B, C)의 가능성이 같으므로 세 몫은 **교환 가능**(*exchangeable*)하다. 즉 셋째 해적까지 포함해 모든 해적이 금화를 얻을 확률은 똑같이 $\frac{n}{n+2}$이다.

소문항 (iii)

A, B, C가 모두 1 이상이어야 한다. $A = A' + 1, B = B' + 1, C = C' + 1$로 놓자. 그러면 구하는 순서쌍은 $A' + B' + C' = n - 3$의 음이 아닌 정수해와 대응한다. 별과 막대(stars and bars) 방법에 따르면 이러한 해는 $\binom{n-1}{2}$개인데, 이는 직접 확인할 수도 있다. 이웃한 금화 사이의 $n-1$개의 틈 중에서 납 동전이 들어갈 두 곳을 고르면 된다. 따라서 $n \geq 1$일 때

$$\mathrm{P}(A \geq 1,\ B \geq 1,\ C \geq 1) = \frac{\binom{n-1}{2}}{\binom{n+2}{2}} = \frac{(n-1)(n-2)}{(n+1)(n+2)}.$$

$n = 1$과 $n = 2$일 때 이 값은 0인데, 금화가 세 개보다 적으면 세 해적이 모두 금화를 가질 수 없으므로 당연히 그래야 한다. ($n = 0$일 때 확률은 자명하게 0이다. 이때는 $\binom{n-1}{2}$가 존재하지 않는 틈을 세게 되므로 공식이 적용되지 않는다.)

검산

- $n = 1$일 때 동전의 배열은 GLL, LGL, LLG이고 모두 가능성이 같다. 첫째 해적은 GLL에서만 금화를 얻으며 그 확률은 $\frac{1}{3} = \frac{1}{1+2}$이고, 둘째 해적은 LGL에서만 금화를 얻으며 그

확률도 $\frac{1}{3}$이다.

- $n = 3$일 때 조건을 만족하는 배열은 $\binom{5}{2} = 10$가지 배열 중 GLGLG 하나뿐이고, 공식은 $\frac{2 \cdot 1}{4 \cdot 5} = \frac{1}{10}$을 준다.
- $n = 1, \ldots, 14$에 대해 정확히 열거한 결과는 세 공식과 모두 일치한다. $n = 5$인 상자 2×10^5개를 시뮬레이션한 결과는 0.715, 0.714, 0.286으로, 이론값 $\frac{5}{7} \approx 0.714$, $\frac{5}{7}$, $\frac{12}{42} = \frac{2}{7} \approx 0.286$과 잘 맞는다.

흔한 실수

- 둘째 해적이 가득 찬 상자에서 새로 꺼내기 시작하는 것처럼 다루는 것. 그녀에게 남은 금화의 개수는 첫째 해적의 결과에 따라 달라진다.
- 소문항 (iii)에서 $\mathrm{P}(A \geq 1), \mathrm{P}(B \geq 1), \mathrm{P}(C \geq 1)$을 곱하는 것. 세 몫은 합이 n이 되어야 하므로 독립(independent)이 아니다.

STEP 너머

납 동전의 자리와 $A + B + C = n$의 해 사이의 대응은 "별과 막대" 전단사(bijection)이다. 이는 무작위로 놓은 표지들 사이의 간격이 교환 가능하다는 것을 보여 준다. 막대 위에 점을 균등하게 떨어뜨리는 연속 버전은 균등 간격(uniform spacings) 이론과 디리클레 분포(Dirichlet distribution)로 이어지며, 이런 종류의 순서통계량(order statistics)은 대학 1학년 확률론 곳곳에서 등장한다. Blitzstein과 Hwang의 *Introduction to Probability*(확률론 입문)는 이러한 경우의 수 세기와 그에 따르는 대칭성 논증을 매우 잘 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.