풀이

STEP I 2000, 문제 12

확률과 통계 • 기하분포

STEP I 2000, 12번 문제

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

step.maths.org/questions/00-s1-q12

핵심 아이디어

각 경우마다 **지금까지의 시도가 모두 실패했다는 조건에서, 다음 시도가 성공할 확률은 얼마인가?**를 묻는다. 그런 다음 처음 $n-1$번 시도의 실패 확률들과 n번째 시도의 성공 확률을 곱한다. 세 가지 규칙은 서로 매우 다른 모양의 답을 준다. 하나는 조건부 확률(conditional probability)이 점점 커지는 경우이고, 하나는 처음부터 일정한 경우이며, 나머지 하나는 첫 시도 이후에만 일정한 경우이다.

이 풀이 전체에서 필요한 시도 횟수를 N이라 하자. 그러면 구하려는 것은 $n = 1, 2, 3, \ldots$에 대한 $\mathrm{P}(N = n)$이다. 문에 맞는 열쇠를 **맞는** 열쇠라고 부르자.

풀이

소문항 (i)

이 경우에는 열쇠를 한 번도 반복하지 않으므로, 열쇠들을 균등하게 무작위인(uniformly random) 순서로 시도하는 것과 같다. 대칭성(symmetry)에 의해 맞는 열쇠는 그 순서의 k개 자리 중 어느 자리에나 같은 확률로 놓인다. 따라서

$$\mathrm{P}(N = n) = \frac{1}{k} \quad \text{단, } n = 1, 2, \ldots, k, \qquad \mathrm{P}(N = n) = 0 \text{ 단, } n > k.$$

대칭성 논증이 너무 매끄럽게 느껴진다면, 대신 조건부 확률들을 곱해 보자. j번 실패한 뒤에는 아직 시도하지 않은 열쇠가 $k-j$개 남아 있고 그중 하나가 맞는 열쇠이므로, $1 \le n \le k$일 때

$$\mathrm{P}(N = n) = \frac{k-1}{k} \cdot \frac{k-2}{k-1} \cdots \frac{k-n+1}{k-n+2} \cdot \frac{1}{k-n+1} = \frac{1}{k},$$

이다. 이 곱은 앞뒤 항이 차례로 약분되는 망원 곱(telescoping product)이기 때문이다.

소문항 (ii)

이번에는 매 시도가 k개의 열쇠 전체에서 새로, 독립적으로(independent) 고르는 것이므로, 이전에 무슨 일이 있었든 각 시도는 확률 $\frac{1}{k}$로 성공한다. 정확히 n번의 시도가 필요하려면 처음 $n-1$번은 실패하고 n번째에 성공해야 한다.

$$\mathrm{P}(N=n) = \left(\frac{k-1}{k}\right)^{n-1}\frac{1}{k}, \qquad n = 1, 2, 3, \ldots$$

이것은 모수가 $\frac{1}{k}$인 기하분포(geometric distribution)이며, n에는 상한이 없다.

소문항 (iii)

첫 번째 시도는 확률 $\frac{1}{k}$로 성공하므로 $\mathrm{P}(N=1) = \frac{1}{k}$이다.

이제 ($m \geq 1$인 어떤 m에 대하여) m번째 시도가 실패했다고 하자. 그러면 m번째 시도에서 쓴 열쇠는 틀린 열쇠이다. $m+1$번째 시도에서는 나머지 $k-1$개의 열쇠 중에서 균등하게 고른다.

> **생각해 보기** m번째 시도 이전의 이력이 $m+1$번째 시도의 성공 확률에 영향을 주는가?

그렇지 않다. 중요한 것은 어떤 열쇠가 제외되었는가뿐이며, 제외된 열쇠는 언제나 **틀린** 열쇠이다. 따라서 고를 수 있는 $k-1$개의 열쇠에는 항상 맞는 열쇠가 포함되어 있고, 이전에 무슨 일이 있었든 $m+1$번째 시도는 확률 $\frac{1}{k-1}$로 성공하고 확률 $\frac{k-2}{k-1}$로 실패한다.

그러므로 $n \geq 2$일 때, 첫 번째 시도가 실패하고(확률 $\frac{k-1}{k}$), $2, \ldots, n-1$번째 시도가 실패하며(각각 확률 $\frac{k-2}{k-1}$), n번째 시도가 성공한다(확률 $\frac{1}{k-1}$). 따라서

$$\mathrm{P}(N=n) = \frac{k-1}{k}\left(\frac{k-2}{k-1}\right)^{n-2}\frac{1}{k-1} = \frac{1}{k}\left(\frac{k-2}{k-1}\right)^{n-2}, \qquad n \geq 2,$$

이고, 여기에 $\mathrm{P}(N=1) = \frac{1}{k}$가 더해진다.

흥미롭게도 $\mathrm{P}(N=2) = \frac{1}{k}$ 역시 소문항 (i)과 같다. ($k=2$이면 이 규칙에 따라 열쇠를 번갈아 써야 하므로 N은 각각 확률 $\frac{1}{2}$로 1 또는 2이다. 0^0을 1로 읽으면 공식도 이 결과를 준다.)

검산

- 각 분포의 합은 1이다. (iii)에서는 $\frac{1}{k} + \frac{1}{k}\sum_{j\geq 0}\left(\frac{k-2}{k-1}\right)^j = \frac{1}{k} + \frac{1}{k}(k-1) = 1$이다.
- 평균(mean)도 합리적으로 나온다. (i)에서는 $\frac{k+1}{2}$, (ii)에서는 k, (iii)에서는 $1 + \frac{(k-1)^2}{k}$이다. 마지막 열쇠를 반복하지 않는 것은 (ii)에 비해 조금 도움이 되지만, 어떤 열쇠도 반복하지 않는 것보다는 훨씬 덜 도움이 된다.

- $k = 2, \ldots, 5$와 $n \leq 7$에 대하여 모든 선택의 순서를 분수로 열거하여 세 공식을 정확히 확인하였다. 또한 $k = 5$인 (iii)을 2×10^5번 시뮬레이션(simulation)한 결과 $\mathrm{P}(N = 3) \approx 0.1496$을 얻었는데, 정확한 값은 $\frac{1}{5} \cdot \frac{3}{4} = 0.15$이다.

흔한 실수

- (iii)에서 첫 번째 시도에도 $\frac{k-2}{k-1}$을 쓰는 것. 첫 선택은 k개의 열쇠 전체에서 하므로 첫 번째 인수는 $\frac{k-1}{k}$이다.
- (i)에서 $n > k$일 때 $\mathrm{P}(N = n) = 0$임을 밝히지 않는 것.
- (iii)에서 제외된 열쇠가 맞는 열쇠일 수도 있다고 생각하는 것. 제외되는 열쇠는 언제나 방금 실패한 열쇠뿐이다.

STEP 너머

소문항 (iii)은 작은 **마르코프 연쇄**(Markov chain)이다. 다음에 일어날 일은 전체 이력이 아니라 현재 상태(여기서는 마지막으로 시도한 열쇠가 무엇인가)에만 의존한다. 과거를 무관하게 만드는 상태를 찾아내는 것이 마르코프 연쇄 이론의 핵심 기법이며, 이를 통해 복잡한 조건부 확률이 간단한 곱으로 바뀐다. Grimmett과 Welsh의 *Probability: An Introduction*(확률론 입문)은 정확히 이 수준에서 마르코프 연쇄를 소개한다.

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.