풀이

STEP II 2017, 문제 13

확률과 통계 • 기댓값과 무한급수

> **STEP II 2017, 13번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/17-s2-q13

핵심 아이디어

모든 경우에 문이 k번째 시도에서 열린다는 것은 처음 $k-1$번의 시도가 실패하고 k번째 시도가 성공한다는 것과 정확히 같으므로, 그 확률은 "실패" 인수들의 곱에 "성공" 인수 하나를 곱한 것이다. 경우마다 달라지는 것은 매번 뽑을 때 주머니 속에 열쇠가 몇 개 있느냐뿐이다. 확률을 적고 나면 기댓값(expectation)은 합으로 주어진다. (i)에서는 미분한 기하급수(geometric series), (ii)에서는 등차급수(arithmetic series), 그리고 (iii)에서는 조화급수(harmonic series)의 일부를 남기는 망원합(telescoping sum)이다.

풀이

이 풀이 전체에서 X를 문이 열리는 시도의 번호라 하고, $q = 1 - \frac{1}{n}$이라 하자.

소문항 (i)

열쇠를 다시 넣으므로 매번 n개의 열쇠가 모두 든 주머니에서 이전과 독립적으로 뽑는다. 따라서 각 시도는 확률 $\frac{1}{n}$으로 성공하고,

$$\mathrm{P}(X = k) = q^{k-1} \cdot \frac{1}{n}, \qquad k = 1, 2, 3, \ldots$$

이다. $0 < q < 1$이므로 이항전개(binomial expansion)에 의해

$$(1-q)^{-2} = 1 + 2q + 3q^2 + \cdots = \sum_{k=1}^{\infty} k q^{k-1}.$$

따라서

$$\mathrm{E}(X) = \sum_{k=1}^{\infty} k\, q^{k-1} \frac{1}{n} = \frac{1}{n}(1-q)^{-2} = \frac{1}{n} \cdot n^2 = n.$$

시도 횟수의 기댓값은 n이다.

소문항 (ii)

이번에는 실패할 때마다 주머니에서 틀린 열쇠가 하나씩 빠지므로, j번째 시도에서는 열쇠가 $n-j+1$개 있다. 문이 k번째 시도에서 열리려면 $k-1$번 실패한 다음 성공해야 한다.

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

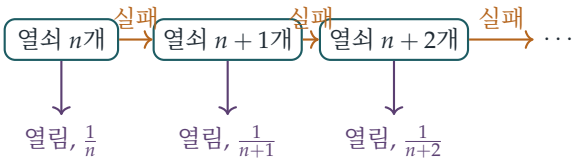
곱이 연쇄적으로 약분되기 때문이다. (이는 직접 보아도 분명하다. 열쇠를 다시 넣지 않고 하나씩 뽑으면 맞는 열쇠는 n개의 자리 가운데 균등하게(uniformly) 무작위인 자리에 놓인다.) 따라서 X는 $\{1,2,\ldots,n\}$ 위의 균등분포를 따르고

$$\mathrm{E}(X)=\frac{1}{n}\sum_{k=1}^{n}k=\frac{1}{n}\cdot\frac{n(n+1)}{2}=\frac{n+1}{2}.$$

소문항 (iii)

확률. 주머니는 n개의 열쇠로 시작하여 실패할 때마다 틀린 열쇠가 하나씩 늘어나며, 맞는 열쇠는 언제나 정확히 하나 들어 있다. 따라서 j번째 시도에서는 열쇠가 $n+j-1$개 있고,

$$\mathrm{P}(\text{시도}j\text{가 실패}\mid\text{이전 시도가 모두 실패})=\frac{n+j-2}{n+j-1}.$$



$k-1$번의 실패와 그 뒤의 성공에 대한 조건부확률(conditional probability)을 곱하면

$$\begin{aligned}\mathrm{P}(X=k)&=\frac{n-1}{n}\cdot\frac{n}{n+1}\cdots\frac{n+k-3}{n+k-2}\cdot\frac{1}{n+k-1}\\&=\frac{n-1}{n+k-2}\cdot\frac{1}{n+k-1}=\frac{n-1}{(n+k-1)(n+k-2)},\end{aligned}$$

이다. 각 분자가 바로 앞의 분모와 약분되어 첫 번째 분자 $n-1$과 마지막 실패 인수의 분모 $n+k-2$만 남기 때문이다. $k=1$일 때는 실패 인수가 없으며, 공식은 마땅히 그래야 하듯 $\frac{n-1}{n(n-1)}=\frac{1}{n}$을 준다. 이것이 구하려던 결과이다.

문은 반드시 열린다. 평균을 계산하기 전에 X가 확률 1로 유한함을 확인해 둘 만하다. $\dfrac{n-1}{(n+k-1)(n+k-2)}=(n-1)\left(\dfrac{1}{n+k-2}-\dfrac{1}{n+k-1}\right)$이므로 합이 연쇄적으로 소거

된다.

$$\sum_{k=1}^{N} \mathrm{P}(X=k) = (n-1)\left(\frac{1}{n-1} - \frac{1}{n+N-1}\right) \to 1 \quad \text{단, } N \to \infty.$$

기댓값. $\sum k\,\mathrm{P}(X=k)$가 필요하므로 $\dfrac{k}{(n+k-1)(n+k-2)}$를 부분분수(partial fractions)로 나눈다.

$$k = A(n+k-2) + B(n+k-1)$$

로 놓고 $k = 1-n$을 대입하면 $1-n = -A$, 즉 $A = n-1$이고, $k = 2-n$을 대입하면 $B = 2-n$이다. 따라서

$$k\,\mathrm{P}(X=k) = (n-1)\left(\frac{n-1}{n+k-1} - \frac{n-2}{n+k-2}\right).$$

> **생각해 보기** 계수 $n-1$과 $n-2$가 다르기 때문에 이 합은 깔끔하게 소거되지 않는다. 그렇다면 무엇이 남는가?

S_N을 처음 N개 항의 합이라 하자. 첫 번째 분수에서 분모는 $m = n, \ldots, n+N-1$에 걸쳐 있고, 두 번째 분수에서는 $m = n-1, \ldots, n+N-2$에 걸쳐 있다. 공통 범위 $m = n, \ldots, n+N-2$에서는 각 $\frac{1}{m}$이 계수 $(n-1)-(n-2) = 1$로 나타나고, 양 끝이 남는다.

$$S_N = (n-1)\left(\sum_{m=n}^{n+N-2} \frac{1}{m} + \frac{n-1}{n+N-1} - \frac{n-2}{n-1}\right).$$

마지막 두 항은 유계이다(가운데 항은 양수이고 마지막 항은 고정된 수이다). 반면

$$\sum_{m=n}^{n+N-2} \frac{1}{m} = \sum_{m=1}^{n+N-2} \frac{1}{m} - \sum_{m=1}^{n-1} \frac{1}{m} \to \infty$$

이다. $N \to \infty$일 때 조화급수는 발산(diverge)하고 우리는 고정된 유한한 양만 빼기 때문이다. $n-1 > 0$이므로 $S_N \to \infty$이다. 모든 항 $k\,\mathrm{P}(X=k)$가 양수이므로 $\mathrm{E}(X)$를 정의하는 급수는 무한대로 발산한다. 즉 문이 결국 열린다는 것은 확실하지만, 시도 횟수의 기댓값은 무한대이다.

검산

- $n = 2$일 때 세 답은 $2, \frac{3}{2}, \infty$이다. 두 번째 경우는 같은 열쇠를 다시 시도하는 일이 없으므로 마땅히 가장 효율적이고, 첫 번째 경우는 그 중간이다.
- (iii)에서 $n = 3$이면 공식은 $k = 1, 2, 3$에 대하여 $\frac{1}{3}, \frac{1}{6}, \frac{1}{10}$을 준다. 직접 계산하면 $\frac{1}{3}$, $\frac{2}{3}\cdot\frac{1}{4} = \frac{1}{6}$, $\frac{2}{3}\cdot\frac{3}{4}\cdot\frac{1}{5} = \frac{1}{10}$이다. 2×10^5번의 게임을 컴퓨터로 모의실험한 결과 상대도수는 0.334, 0.168, 0.099였다.

- 역시 $n = 3$일 때 $N = 10^4$인 부분합 S_N은 약 15.6이며, $2\ln N$처럼 계속 커진다. 극한은 존재하지 않는다.

흔한 실수

- (iii)에서 k번째 시도의 주머니 크기를 $n + k - 1$이 아니라 $n + k$로 잡는 것. k번째로 뽑기 전에 추가된 열쇠는 $k - 1$개뿐이다.
- 평균이 무한대이므로 문이 영영 열리지 않을 수도 있다고 결론짓는 것. 확률의 합은 1이다. $1/k^2$처럼 감소하는 꼬리가 너무 두꺼워서 $\sum k\,\mathrm{P}(X = k)$가 수렴하지 않을 뿐이다.
- 두 계수가 같은 것처럼 부분분수를 소거하여 유한한 답을 얻는 것.

STEP 너머

세 번째 경우는 $\mathrm{P}(X = k)$가 $1/k^2$처럼 행동하기 때문에 확률 1로 유한하지만 평균은 무한대인 확률변수(random variable)를 준다. 이와 같은 꼬리가 두꺼운 분포(heavy-tailed distribution)와 유명한 상트페테르부르크 역설(St Petersburg paradox)은 대학 1학년 확률론 강의에서 왜 $\sum k\,\mathrm{P}(X = k)$가 수렴할 때에만 $\mathrm{E}(X)$가 존재한다고 고집하는지를 보여준다. 실패할 때마다 틀린 열쇠가 하나씩 늘어나는 주머니는 "부자는 더 부유해진다" 과정의 표준 모형인 폴리아(Pólya)의 항아리(urn)와 가까운 친척이다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 이산 확률변수의 기댓값과 이러한 미묘한 점들을 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.