풀이

STEP I 2002, 문제 14

확률과 통계 • 확률에서의 등비급수

> **STEP I 2002, 14번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/02-s1-q14

핵심 아이디어

내가 모두 몇 개의 번호를 입력하는지를 조건(conditioning)으로 삼는다. 그 개수가 정해지면 경보가 울린 순간부터 내가 돈을 손에 넣는 순간까지의 시간도 정해지므로, 경찰이 먼저 도착할 확률은 그 시간 t에 대하여 $1-\mathrm{e}^{-\lambda t}$일 뿐이다. 이것을 시도 횟수에 대하여 평균하면 답이 나온다. 두 방법 모두에서 이 평균은 $q=\mathrm{e}^{-\lambda/r}$에 대한 등비급수(geometric series)가 되므로, q에 미리 이름을 붙여 두는 것이 좋다.

풀이

순서대로 입력하는 경우

목록 $1,2,\ldots,n$에서 내 번호의 위치를 K라 하자. 그러면 K는 내가 시도하는 횟수이고, $k=1,\ldots,n$에 대하여 $P(K=k)=\frac{1}{n}$이다.

$K=k$라 하자. 나는 1분에 r개의 속도로 번호를 입력하므로, 연속한 두 번호 사이의 간격은 $1/r$분이다. 시간은 내가 첫 번호를 입력한 순간부터 잰다.

- $k=1$이면 첫 번호가 맞으므로, 틀린 번호는 한 번도 입력되지 않고 경찰도 결코 호출되지 않는다. 경찰이 먼저 도착할 확률은 $0=1-\mathrm{e}^{0}$이다.
- $k\geq 2$이면 첫 번호가 틀리므로 시각 0에 경보가 울리고, 맞는 번호는 k번째로 시각 $(k-1)/r$에 입력된다. 경찰이 $t=(k-1)/r$분 이내에 도착하면 내가 돈을 손에 넣기 전에 도착하는 것이고, 그 확률은 $1-\mathrm{e}^{-\lambda(k-1)/r}$이다.

따라서 식 $1-\mathrm{e}^{-\lambda(k-1)/r}$은 $k=1$을 포함한 모든 k에 대하여 성립한다. 내가 돈을 손에 넣기

전에 경찰이 도착하는 사건을 A라 하자. 전확률 법칙(law of total probability)에 의하여

$$P(A) = \sum_{k=1}^{n} P(K=k)\,P(A \mid K=k) = \sum_{k=1}^{n} \frac{1-\mathrm{e}^{-\lambda(k-1)/r}}{n},$$

이고, 이것이 구하려던 식이다.

간단히 하기 위해 $q = \mathrm{e}^{-\lambda/r}$로 놓는다. $\lambda, r > 0$이므로 $0 < q < 1$이고, $\mathrm{e}^{-\lambda(k-1)/r} = q^{k-1}$이다. 따라서

$$\begin{aligned} P(A) &= 1 - \frac{1}{n}\sum_{k=1}^{n} q^{k-1} = 1 - \frac{1-q^n}{n(1-q)} \\ &= 1 - \frac{1-\mathrm{e}^{-\lambda n/r}}{n\left(1-\mathrm{e}^{-\lambda/r}\right)}, \end{aligned}$$

이다. 여기서 등비급수의 합을 구했는데, $q \neq 1$이므로 이는 타당하다.

무작위로 입력하는 경우

이제 각 시도는 이전 시도와 독립적으로 $1, \ldots, n$에서 균등하게 새로 고른 번호이다(나는 무엇을 시도했는지 기억하지 못한다). 따라서 각 시도는 서로 독립적으로 확률 $\frac{1}{n}$로 맞는다. 맞는 번호를 포함하여 그때까지의 시도 횟수를 N이라 하면, N은 기하분포(geometric distribution)를 따른다.

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

> **?** **생각해 보기** 번호가 반복될 수 있을 때에도 시간에 대한 논증이 여전히 성립하는가?

그렇다. 경찰은 첫 번째 틀린 번호에서 호출되며, $N = k \geq 2$이면 그것은 바로 첫 시도, 즉 시각 0이다. 돈은 k번째 시도, 즉 시각 $(k-1)/r$에 나온다. 반복은 N의 분포만 바꿀 뿐 시간 관계는 바꾸지 않는다. 그러므로 앞에서와 똑같이 $P(A \mid N=k) = 1 - q^{k-1}$이고, 이 식은 $k = 1$일 때에도 맞다.

$p = \frac{n-1}{n}$이라 쓰자. 그러면

$$\begin{aligned} P(A) &= \sum_{k=1}^{\infty} \frac{1}{n}\,p^{k-1}\left(1-q^{k-1}\right) = \frac{1}{n}\sum_{k=1}^{\infty} p^{k-1} - \frac{1}{n}\sum_{k=1}^{\infty}(pq)^{k-1} \\ &= \frac{1}{n}\cdot\frac{1}{1-p} - \frac{1}{n}\cdot\frac{1}{1-pq}. \end{aligned}$$

이다. $0 \leq p < 1$이고 $0 < pq < 1$이므로 두 급수가 모두 수렴(converge)하고, 따라서 합을 둘로 나누어도 된다. 첫째 항은 $\frac{1}{n}\cdot n = 1$인데, 이는 N이 확실히 유한하다는 사실을 말해 줄 뿐이다. 둘째 항에 대해서는

$$n(1-pq) = n - (n-1)q,$$

이므로

$$P(A) = 1 - \frac{1}{n - (n-1)\mathrm{e}^{-\lambda/r}},$$

이고, 이것이 구하려던 식이다.

검산

- $n = 1$이면 틀린 번호가 결코 없으며, 두 식 모두 0을 준다. 첫째 식은 $1 - q^1 = 1 - q$이기 때문이고, 둘째 식은 분모가 1이기 때문이다.
- $\lambda \to \infty$일 때(경찰이 즉시 도착할 때) $q \to 0$이고, 두 답 모두 $1 - \frac{1}{n}$에 가까워진다. 이는 첫 시도가 틀릴 확률이므로, 마땅히 그래야 한다.
- 무작위로 입력하는 쪽이 더 불리해야 한다. $n = 5$, $\lambda = 0.7$, $r = 3$이면 순서대로 입력할 때 0.3382, 무작위로 입력할 때 0.4543을 얻는다. 각각 2×10^5번 시행한 시뮬레이션에서는 0.339와 0.456이 나왔다.

흔한 실수

- $(k-1)/r$ 대신 시각 k/r을 쓰는 것. 시계는 첫 번째 (틀린) 번호에서 시작하므로, k번째 번호는 간격 $k-1$개만큼 뒤에 온다.
- 무작위인 경우에 각 k에 확률 $\frac{1}{n}$을 주거나, 합을 n에서 멈추는 것. 반복이 허용되면 N은 기하분포를 따르며 상한이 없다.

STEP 너머

여기서 경찰의 도착 시간은 지수분포(exponential distribution)를 따르는 확률변수(random variable)이고, 계산하는 것은 확률변수 X에 대한 $P(T < X)$로, X를 조건으로 삼아 구한다. X가 기하분포를 따르면 $E\left(q^X\right)$가 X의 확률생성함수(probability generating function)의 q에서의 값이므로 답이 깔끔한 닫힌 꼴을 갖는다. 실제로 두 답은 모두 $1 - E\left(q^{N-1}\right)$이다. 생성함수와 지수분포는 대학 1학년 확률론의 핵심 내용이다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 두 주제를 모두 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.