풀이

STEP III 1993, 문제 15

확률과 통계 • 대기 시간의 분포

> **STEP III 1993, 15번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/93-s3-q15

핵심 아이디어

마지막 던지기에 주목한다. 실험이 n번째 던지기에서 멈추는 것은 바로 그 던지기가 아직 모자랐던 면의 두 번째를 채워 줄 때이다. 따라서 처음 $n-1$번의 던지기에는 그 면이 정확히 한 번 나오고 나머지는 모두 다른 면이 나와야 한다. 그러면 분포는 간단한 두 항의 합이 된다. 이제 기댓값(expectation)은 $\sum n(n-1)x^{n-2}$ 꼴의 급수 두 개로 나타나며, 이는 등비급수(geometric series)를 두 번 미분하여 얻는다.

풀이

$0 < p < 1$이라고 가정하고(그렇지 않으면 실험이 끝나지 않는다), 각 던지기는 독립(independent)이라고 가정한다.

N의 분포(distribution)

앞면(head)이 적어도 두 번, 뒷면(tail)이 적어도 두 번 나와야 하므로 $N \geq 4$이다. $n \geq 4$를 고정한다. 사건 $N = n$에서 n번째 던지기는 앞면 또는 뒷면이고, 이 두 경우는 서로소(disjoint)이다.

n번째 던지기가 앞면인 경우. 그 직전까지는 목표에 도달하지 못했고, n번째 던지기에서 앞면이 더해져 목표에 도달했다. 따라서 처음 $n-1$번의 던지기에는 이미 뒷면이 적어도 두 번 있었지만 앞면은 한 번뿐이었다. 즉 앞면이 정확히 한 번, 뒷면이 $n-2$번이다(그리고 $n-2 \geq 2$인데, 이는 $n \geq 4$이기 때문이다). 역으로, 이러한 수열 뒤에 앞면이 나오면 정확히 n번째에서 멈춘다. 한 번 나온 앞면은 $n-1$개의 자리 중 어디에나 올 수 있으므로, 이 경우의 확률은

$$(n-1)\,p\,q^{n-2} \cdot p = (n-1)p^2q^{n-2}.$$

n번째 던지기가 뒷면인 경우. 앞면과 뒷면을 바꾸어 같은 논증을 하면, 확률은 $(n-1)q^2p^{n-2}$이다.

두 경우를 더하면

$$\boxed{\mathrm{P}(N=n)=(n-1)(p^2q^{n-2}+q^2p^{n-2}),\qquad n=4,5,6,\ldots}$$

이며, $\mathrm{P}(N=n)=0$이 되는 것은 $n\le 3$일 때이다.

생각해 보기 마지막 던지기가 앞면일 때, 처음 $n-1$번의 던지기에 앞면이 두 번 있을 수 없는 이유는 무엇인가?

만약 그렇다면, 그 던지기들에는 이미 앞면 두 번과 (멈추는 데 필요한) 뒷면 두 번이 들어 있게 되므로 실험은 n번째 던지기 이전에 끝났을 것이다. 바로 이 때문에 앞면의 개수가 정확히 하나로 정해진다.

기댓값

$|x|<1$일 때 $\sum_{n\ge 0}x^n=(1-x)^{-1}$을 두 번 미분하면 다음을 얻는다.

$$\sum_{n=2}^{\infty}n(n-1)x^{n-2}=\frac{2}{(1-x)^3}.$$

우리의 합은 $n=4$부터 시작하므로, $n=2$ 항과 $n=3$ 항, 즉 2와 $6x$를 빼면

$$\sum_{n=4}^{\infty}n(n-1)x^{n-2}=\frac{2}{(1-x)^3}-2-6x.$$

이제

$$\mathrm{E}(N)=\sum_{n=4}^{\infty}n\,\mathrm{P}(N=n)=p^2\sum_{n=4}^{\infty}n(n-1)q^{n-2}+q^2\sum_{n=4}^{\infty}n(n-1)p^{n-2}.$$

이 급수에 $x=q$(따라서 $1-x=p$)를 대입하고, 그다음 $x=p$(따라서 $1-x=q$)를 대입한다. $0<p,q<1$이므로 둘 다 성립하며, 다음을 얻는다.

$$\begin{aligned}\mathrm{E}(N)&=p^2\Big(\frac{2}{p^3}-2-6q\Big)+q^2\Big(\frac{2}{q^3}-2-6p\Big)\\&=\frac{2}{p}+\frac{2}{q}-2(p^2+q^2)-6pq(p+q).\end{aligned}$$

$p+q=1$이므로 $\frac{1}{p}+\frac{1}{q}=\frac{p+q}{pq}=\frac{1}{pq}$이고 $p^2+q^2=(p+q)^2-2pq=1-2pq$이다. 따라서

$$\begin{aligned}\mathrm{E}(N)&=\frac{2}{pq}-2(1-2pq)-6pq\\&=\frac{2}{pq}-2-2pq=2\Big(\frac{1}{pq}-1-pq\Big),\end{aligned}$$

이 되어 요구한 식을 얻는다.

검산

- 확률의 합은 1이다. 인수가 하나 적은 같은 급수로부터 $\sum_{n\geq 4}(n-1)p^2q^{n-2} = 1 - p^2 - 2p^2q$를 얻고, 앞면과 뒷면을 바꾼 식을 더하면 $1-(p^2+q^2)-2pq(p+q) = 1-1+2pq-2pq = 1$이 남는다. 여기서 $p^2+q^2 = 1-2pq$를 이용하였다.
- 공정한 동전이면 $pq = \frac{1}{4}$이고 $\mathrm{E}(N) = 2(4-1-\frac{1}{4}) = \frac{11}{2}$이다. 직접 계산하면 $\mathrm{P}(N=4) = \frac{3}{8}$, $\mathrm{P}(N=5) = \frac{1}{4}$, ...이고, $n\,\mathrm{P}(N=n)$의 부분합은 5.5에 가까워진다.
- $p \to 0$일 때 $\mathrm{E}(N) \approx 2/p$이다. 앞면이 두 번 나오기를 기다려야 하고, 각각 약 $1/p$번의 던지기가 걸리기 때문이다.
- 길이 10까지의 모든 수열을 나열하면 분포가 정확히 재현되고, $p = 0.3$으로 한 시뮬레이션에서는 평균이 7.10으로, 공식에서 얻은 7.1038과 맞았다.

흔한 실수

- $\binom{n}{2}$라고 쓰는 실수(올바른 값은 $n-1$이다). 마지막 던지기의 위치는 고정되어 있으므로, 처음 $n-1$번의 던지기 중 단 한 번 나온 앞면(또는 뒷면)의 위치만 자유롭다.
- 표준 급수를 사용할 때 $n=2$ 항과 $n=3$ 항을 빼는 것을 잊는 실수.

STEP 너머

N은 평행이동된 두 **음이항**(*negative binomial*) 대기 시간의 혼합이고, 등비급수를 미분하는 요령은 **확률생성함수**(*probability generating function*)를 손으로 만든 형태이다. $G(s) = \mathrm{E}(s^N)$으로 두면 평균은 $G'(1)$로, 분산은 $G''(1)$로부터 똑같이 빠르게 얻을 수 있다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 생성함수와 이와 같은 대기 시간 문제를 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.