풀이

STEP I 1990, 문제 15

확률과 통계 • 음이항분포

> **STEP I 1990, 15번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/90-s1-q15

핵심 아이디어

첫 번째 부분에서는 사건을 서로 독립(independent)인 두 조각, 즉 마지막 던지기에서 일어나는 일과 그 이전의 모든 던지기에서 일어나는 일로 나눈다. 나머지 부분에서 중요한 사실은 동전에 기억이 없다는 것이다. 앞면이 막 나왔다면 다음 앞면을 기다리는 일은 처음부터 새로 시작되므로, 기하분포(geometric distribution)를 따르는 대기 시간이 짝수일 확률과 홀수일 확률만 알면 된다.

풀이

이 풀이 전체에서 $q = 1 - p$로 둔다. 각 던지기는 독립이고, 각각 확률 p로 앞면이 나온다.

k번째 앞면의 분포

k번째 앞면이 나오는 던지기의 번호를 T_k라 하자. 사건 $\{T_k = n\}$은 정확히 다음 두 조건이 성립할 때 일어난다.

- n번째 던지기가 앞면이고,
- $1, 2, \ldots, n-1$번째 던지기 중 앞면이 정확히 $k-1$번 (따라서 뒷면이 $n-k$번) 나온다.

두 조건은 모두 필요하다. n번째 던지기가 뒷면이라면 k번째 앞면은 n번째 던지기 이전이나 이후에 나오고, n번째에 나오지는 않는다. 처음 $n-1$번의 던지기에서 앞면의 개수가 다르다면 n번째 던지기는 k번째 앞면이 되지 않는다. 또한 두 조건을 함께 만족하면 충분하다.

이 두 사건은 서로 겹치지 않는 던지기들에 관한 것이므로 독립이다. 첫 번째 사건의 확률은 p이다. 두 번째 사건에서 $k-1$개의 앞면은 처음 $n-1$번의 던지기 중 $\binom{n-1}{k-1}$가지 위치 집합

어디에든 놓일 수 있다. 이러한 배열 각각은 $k-1$개의 앞면과 $n-k$개의 뒷면으로 이루어진 특정한 수열이며, 독립성에 의해 그 확률은 $p^{k-1}q^{n-k}$이다. 또 서로 다른 배열은 서로 배반(mutually exclusive)이다. 따라서

$$\mathrm{P}(T_k = n) = p \cdot \binom{n-1}{k-1} p^{k-1} q^{n-k} = \binom{n-1}{k-1} p^k (1-p)^{n-k}, \qquad n \geq k,$$

이고, 이것이 구하려던 결과이다.

짝수에서 짝수로

$(k-1)$번째 앞면 이후에 추가로 필요한 던지기 횟수를 $G = T_k - T_{k-1}$라 하자 (단, $T_0 = 0$). T_{k-1}번째 던지기 이후의 던지기들은 그 시점까지 일어난 모든 일과 독립이다. 따라서 G는 T_{k-1}과 독립이고, 다음과 같은 기하분포를 따른다.

$$\mathrm{P}(G = g) = q^{g-1}p, \qquad g = 1, 2, \ldots$$

(이는 첫 번째 부분에서 $k = 1$인 경우이다.)

? **생각해 보기** T_{k-1}이 짝수일 때, T_k는 언제 짝수인가?

$T_k = T_{k-1} + G$이므로, T_{k-1}이 짝수이면 T_k는 정확히 G가 짝수일 때 짝수이다. G는 T_{k-1}과 독립이므로, T_{k-1}이 짝수라는 조건을 걸어도(conditioning) G의 분포는 변하지 않는다. 그러므로

$$\mathrm{P}(T_k \text{ 짝수} \mid T_{k-1} \text{ 짝수}) = \mathrm{P}(G \text{ 짝수}).$$

등비급수(geometric series)의 합을 구하면

$$\mathrm{P}(G \text{ 짝수}) = p\left(q + q^3 + q^5 + \cdots\right) = \frac{pq}{1-q^2} = \frac{pq}{(1-q)(1+q)} = \frac{q}{1+q},$$
$$\mathrm{P}(G \text{ 홀수}) = p\left(1 + q^2 + q^4 + \cdots\right) = \frac{p}{1-q^2} = \frac{1}{1+q}.$$

따라서 구하는 확률은

$$\frac{q}{1+q} = \frac{1-p}{2-p}.$$

이 답은 k에 의존하지 않는다.

짝수 번째 던지기에서 정확히 두 번 앞면

이제 세 번째 앞면이 나올 때까지 던진다. 세 대기 시간을 $G_1 = T_1$, $G_2 = T_2 - T_1$, $G_3 = T_3 - T_2$라 하자. 같은 무기억성(memoryless) 논증에 의해 이들은 독립이고, 각각 위와 같은 기하분포를 따른다. 다음과 같이 쓰자.

$$e = \mathrm{P}(G_i \text{ 짝수}) = \frac{q}{1+q}, \qquad o = \mathrm{P}(G_i \text{ 홀수}) = \frac{1}{1+q}.$$

시작점은 0번째 던지기이고, 이는 짝수이다. 짝수 간격은 현재 위치의 홀짝성(parity)을 유지하고, 홀수 간격은 홀짝성을 바꾼다. 다음 표는 (G_1, G_2, G_3)의 가능한 홀짝 패턴 여덟 가지와 그에 따른 (T_1, T_2, T_3)의 홀짝성을 나열한 것이다.

G_1	G_2	G_3	T_1	T_2	T_3	짝수 번째 앞면
짝	짝	짝	짝	짝	짝	3
짝	짝	홀	짝	짝	홀	**2**
짝	홀	짝	짝	홀	홀	1
짝	홀	홀	짝	홀	짝	**2**
홀	짝	짝	홀	홀	홀	0
홀	짝	홀	홀	홀	짝	1
홀	홀	짝	홀	짝	짝	**2**
홀	홀	홀	홀	짝	홀	1

짝수 번째 던지기에서 앞면이 정확히 두 번 나오는 경우는 세 가지이고, 그 확률은 각각 e^2o, eo^2, o^2e이다. 간격들의 독립성에 의해 구하는 확률은

$$\begin{aligned} e^2o + 2eo^2 = eo(e+2o) &= \frac{q}{(1+q)^2} \cdot \frac{q+2}{1+q} \\ &= \frac{q(2+q)}{(1+q)^3} = \frac{(1-p)(3-p)}{(2-p)^3}. \end{aligned}$$

검산

- 첫 번째 부분의 검산으로, $n \geq k$에 대해 $\mathrm{P}(T_k = n)$을 모두 더하면 1이 되어야 한다. 이는 음이항 급수 $\sum_{n \geq k} \binom{n-1}{k-1} q^{n-k} = (1-q)^{-k}$이다.
- 두 홀짝 확률 $\frac{q}{1+q}$와 $\frac{1}{1+q}$를 더하면 1이다. p가 1에 가까우면 첫 앞면은 거의 항상 곧바로, 즉 홀수 번째 던지기에서 나오므로 $\mathrm{P}(G \text{ 짝수}) \to 0$이고, 공식도 그렇게 준다.
- 공정한 동전($p = \frac{1}{2}$)에 대해 최종 답은 $\frac{(1/2)(5/2)}{(3/2)^3} = \frac{10}{27} \approx 0.370$이다. 2×10^5회의 시뮬레이션 결과는 0.369였다. 세 대기 시간에 대해 직접 합을 계산한 결과는 $p = 0.3, 0.5, 0.8$에서 공식과 10^{-9} 이내로 일치했다.

흔한 실수

- 첫 번째 부분에서 $\binom{n}{k}$를 사용하는 것. 마지막 던지기는 반드시 앞면이어야 하므로, 자유롭게 정해지는 것은 처음 $n-1$번의 던지기뿐이다.
- 두 번째 부분에서 음이항분포를 짝수 n에 대해 더하는 것. 무기억성을 가진 간격을 이용하면 이 계산을 완전히 피할 수 있다.

- 마지막 부분에서 짝수 위치의 앞면을 셀 때 위치들이 독립인 것처럼 다루는 것. 각 위치는 바로 앞의 위치로부터 만들어지므로 독립이 아니다.

STEP 너머

k번째 앞면이 나오는 시각은 **음이항분포**(negative binomial distribution)를 따른다. 이를 독립인 기하분포 간격들의 합으로 쓰는 것은, 포아송 과정(Poisson process)의 도착 시각이 독립인 지수분포 대기 시간들로 만들어지는 방식의 이산 버전이다. 마지막 부분의 홀짝 기록은 두 상태 마르코프 연쇄(Markov chain)이며, 이는 대학 1학년 확률론 강의에서 일찍 만나는 주제이다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)이 이 모든 내용을 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.