풀이

STEP III 2007, 문제 13

확률과 통계 • 무작위 행보의 기댓값

STEP III 2007, 13번 문제

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

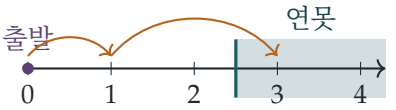
step.maths.org/questions/07-s3-q13

핵심 아이디어

반 미터가 붙어 있는 것은 "가장자리에 정확히 착지하는" 경우에 대한 의심을 없애기 위해서이다. $(n-\frac{1}{2})\,\mathrm{m}$ 떨어진 곳에서 출발하면, 개구리가 뛴 거리의 합이 n 이상일 때 그리고 그때에만 개구리는 연못 안에 있다. 따라서 문제의 핵심은 1과 2로 이루어진 누적합이 처음으로 n에 도달하는 시점이다. 작은 n에 대해서는 가능한 경우를 그냥 나열하면 되고, 일반적인 공식에 대해서는 u_n의 세 값으로부터 세 상수에 대한 세 방정식을 얻는다.

풀이

처음 j번의 점프로 이동한 총거리(미터 단위)를 S_j라 하고, 연못에 도달하는 데 필요한 점프 횟수를 T라 하자. $(n-\frac{1}{2})\,\mathrm{m}$ 떨어진 곳에서 출발하면, 개구리가 j번 점프한 뒤 연못 안에 있을 필요충분조건은 $S_j > n-\frac{1}{2}$이고, S_j가 정수이므로 이는 $S_j \geq n$과 같다. 따라서 $p_n(j) = \mathrm{P}(T=j)$이며, 여기서 T는 $S_j \geq n$을 만족하는 최초의 j이다.



그림은 $n=3$(가장자리가 2.5 m 떨어져 있음)인 경우와 "짧게, 길게"라는 점프 순서를 보여주며, 이 경우 두 번의 점프로 연못에 도달한다.

소문항 (i)

$n = 2$이면 가장자리는 1.5 m 떨어져 있다. 첫 점프가 길면(확률 q) 개구리는 곧바로 연못에 들어가므로 $p_2(1) = q$이다. 그렇지 않으면 첫 점프는 짧고(확률 p), 개구리는 가장자리에서 0.5 m 떨어진 곳에 있게 되어 두 번째 점프가 **어떤 것이든** 연못에 들어간다. 점프들은 독립(independent)이므로

$$p_2(2) = p \times 1 = p,$$

이고, 이것이 구하려던 결과이다. (또한 $j \geq 3$이면 $p_2(j) = 0$이고, $p + q = 1$이므로 확률분포(probability distribution)가 빠짐없이 완성됨을 확인할 수 있다.)

소문항 (ii)

$\frac{1}{2}$ m 떨어진 곳에서는 어떤 점프든 연못에 착지하므로 $u_1 = 1$이다.

소문항 (i)을 이용하면

$$u_2 = 1 \cdot p_2(1) + 2 \cdot p_2(2) = q + 2p = 2 - q.$$

$n = 3$이면 $S_j \geq 3$이어야 한다.

- $S_1 \leq 2$이므로 한 번의 점프로는 결코 충분하지 않다. 즉 $p_3(1) = 0$이다.
- 두 번의 점프가 충분하지 않은 것은 두 점프가 모두 짧은 경우($S_2 = 2$)뿐이고, 나머지 모든 경우에는 $S_2 \geq 3$이다. 따라서 $p_3(2) = 1 - p^2$이다.
- 두 번 짧게 점프한 뒤에는 개구리가 0.5 m 떨어진 곳에 있으므로 세 번째 점프는 반드시 연못에 착지한다. 즉 $p_3(3) = p^2$이다.

그러므로

$$\begin{aligned} u_3 &= 2(1 - p^2) + 3p^2 = 2 + p^2 \\ &= 2 + (1 - q)^2 = 3 - 2q + q^2, \end{aligned}$$

이고, 이것이 구하려던 결과이다.

생각해 보기 이 값들을 더 빨리 알아내는 방법이 있을까? 그리고 그 방법은 모든 n으로 확장될까?

있다. 첫 번째 점프에 대해 조건부로 생각하면 된다. 첫 점프 후 개구리는 n이 1 또는 2만큼 줄어든 같은 상황에 놓이므로, $n \geq 2$일 때

$$u_n = 1 + p\,u_{n-1} + q\,u_{n-2}, \qquad u_0 = 0,$$

이 성립한다. 여기서 $u_0 = 0$은 1.5 m에서 첫 점프를 길게 하는 경우를 처리한다. 이로부터 다시 $u_2 = 1 + p$, $u_3 = 1 + p(1 + p) + q = 2 + p^2$을 얻는다. 소문항 (iii)의 형태

가 옳은 이유도 바로 여기에 있다. 특성방정식(auxiliary equation) $\lambda^2 - p\lambda - q = 0$은 $(\lambda - 1)(\lambda + q) = 0$으로 인수분해되어 $(-q)^{n-1}$ 항을 준다. ("+1"에서 비롯된) 중근 1이 $B + Cn$을 설명한다.

소문항 (iii)

$u_n = A(-q)^{n-1} + B + Cn$에 $n = 1, 2, 3$을 대입하고 소문항 (ii)를 이용하면 다음을 얻는다.

$$A + B + C = 1, \tag{1}$$
$$-qA + B + 2C = 2 - q, \tag{2}$$
$$q^2A + B + 3C = 3 - 2q + q^2. \tag{3}$$

빼서 B를 소거하면

$$(2) - (1): \quad -(1+q)A + C = 1 - q, \tag{4}$$
$$(3) - (2): \quad q(1+q)A + C = 1 - q + q^2. \tag{5}$$

이제 $q \times (4) + (5)$로 A를 소거하면

$$(1+q)C = q - q^2 + 1 - q + q^2 = 1, \qquad \text{따라서} \qquad C = \frac{1}{1+q},$$

이고, 이것이 구하려던 결과이다. 그러면 (4)로부터

$$(1+q)A = \frac{1}{1+q} - (1-q) = \frac{1-(1-q^2)}{1+q} = \frac{q^2}{1+q}, \qquad \text{따라서} \qquad A = \frac{q^2}{(1+q)^2},$$

이고, (1)로부터

$$B = 1 - \frac{q^2}{(1+q)^2} - \frac{1}{1+q} = \frac{(1+q)^2 - q^2 - (1+q)}{(1+q)^2} = \frac{q}{(1+q)^2}.$$

따라서

$$u_n = \frac{q^2(-q)^{n-1} + q + (1+q)n}{(1+q)^2}.$$

큰 n. $0 \le q \le 1$이므로 항 $A(-q)^{n-1}$과 B는 유계(bounded)이고(각각의 크기는 많아야 $\frac{1}{4}$이다), Cn은 한없이 커진다. 그러므로 큰 n에 대해

$$u_n \approx Cn = \frac{n}{1+q} = \frac{n}{p+2q},$$

이다. 여기서 $1 = p + q$이므로 $1 + q = p + 2q$임을 이용하였다.

이것이 예상되는 이유. 점프 한 번의 평균 길이는 $1 \cdot p + 2 \cdot q = p + 2q$미터이다. 개구리는 약 n미터를 이동해야 하고, 큰수의 법칙(law of large numbers)에 의해 N이 클 때 N번 점프한 총길이는 $N(p+2q)$에 가깝다. 따라서 필요한 점프 횟수는 $n/(p+2q)$에 가깝다. 출발점의 반 미터 차이와 마지막에 넘어서는 거리(개구리는 가장자리를 넘어 많아야 1.5 m 지점에 착지한다)가 유계이므로 이 근사는 좋은 근사이다. 이들은 점프 횟수를 유계인 양만큼만 바꾸며, 이는 n에 비해 무시할 수 있다. 그 유계인 보정이 바로 $A(-q)^{n-1} + B$ 부분이다.

검산

- 공식에 $n = 0$을 대입하면 마땅히 그래야 하듯이 $\left(q^2 \cdot (-1/q) + q\right)/(1+q)^2 = 0 = u_0$이고, 공식은 위의 점화식(recurrence)을 만족한다.
- $q = 0$(모든 점프가 짧음)이면 $u_n = n$인데, 개구리는 정확히 n번의 단위 점프가 필요하므로 옳다.
- 모든 점프 순서를 (분수로) 정확히 열거한 결과가 $n = 1, \ldots, 12$와 여러 p 값에 대해 공식과 일치하였다. $q = 0.7$, $n = 200$으로 한 시뮬레이션에서는 평균 117.90번의 점프가 나왔으며, 공식으로는 117.89, $n/(p+2q) \approx 117.65$이다.

흔한 실수

- "연못에 착지한다"를 정확히 $S_j = n$인 것으로 다루는 실수. 반 미터 때문에 개구리는 넘어설 수 있으며, $S_j \geq n$인 모든 순서가 해당된다.
- $p_3(2) = 2pq + q^2$이라 쓰면서 긴 점프가 먼저 오는 경우를 잊는 실수. 실제로 $pq + qp + q^2 = 1 - p^2$은 두 순서를 모두 포함한다.

STEP 너머

u_n에 대한 점화식은 재생 방정식(renewal equation)이다. μ가 평균 보폭일 때 $u_n \sim n/\mu$라는 결과는 기본 재생 정리(elementary renewal theorem)이다. 왈드의 항등식(Wald's identity) $\mathrm{E}(S_T) = \mu\,\mathrm{E}(T)$로부터 정확한 형태를 얻는다. 여기서 S_T는 n 또는 $n+1$이므로 $(1+q)u_n = n + \mathrm{P}$(넘어섬)이다. Grimmett and Welsh, *Probability: An Introduction*(확률론 입문)은 이런 관점에서 무작위 행보(random walk)와 생성함수(generating function)를 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.