풀이

STEP III 1988, 문제 15

확률과 통계 • 기댓값의 점화식

> **STEP III 1988, 15번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/88-s3-q15

핵심 아이디어

그날의 **첫 번째** 책에 대해 조건을 건다(condition). 첫 책이 책꽂이에 꽂히고 나면, 그날의 나머지는 더 짧은 책꽂이로 새로 시작하는 하루와 정확히 같아 보인다. 따라서 길이 n인 책꽂이에 꽂히는 책 수의 기댓값(expected value)은 길이 $n-1$과 $n-2$일 때의 기댓값과 연결된다. 이로부터 선형 차분방정식(linear difference equation)을 얻고, 이를 2계 점화식(second-order recurrence)과 똑같은 방법으로 푼다. 즉 동차(homogeneous) 부분은 특성 이차방정식(auxiliary quadratic)으로 풀고, 우변의 상수 1에 대해서는 특수해(particular solution)를 더한다.

풀이

이 풀이 전체에서 책꽂이의 길이가 n인 날 책꽂이에 꽂히는 책의 수를 N_n이라 하자. 그러면 $M(n) = \mathrm{E}(N_n)$이다. 각 책은 서로 독립적으로 두껍거나 얇으며, 각 책이 두꺼울 확률은 p이다. 책은 항상 책꽂이에 들어갈 수 있는 것보다 많다고 주어졌으므로, 하루는 언제나 어떤 책이 들어가지 못하는 것으로 끝난다.

소문항 (i)

$n = 0$이면 빈 공간이 전혀 없다. 첫 번째 책은 두껍든 얇든 들어가지 않으므로, 그 책과 그 뒤의 모든 책은 수레로 간다. 따라서 항상 $N_0 = 0$이고, $M(0) = 0$이다.

소문항 (ii)

$n = 1$일 때 첫 번째 책을 살펴보자.

- 첫 책이 얇으면(확률 q) 책꽂이에 들어가 책꽂이를 채운다. 다음 책은 적어도 한 단위가 필요하므로 들어가지 않고, 그 뒤로는 모든 책이 수레로 간다. 따라서 $N_1 = 1$이다.
- 첫 책이 두꺼우면(확률 p) 두 단위가 필요하므로 들어가지 않고, $N_1 = 0$이다.

그러므로 $M(1) = q \cdot 1 + p \cdot 0 = q$이다.

소문항 (iii)

$n \geq 2$라 하고 다시 첫 번째 책에 대해 조건을 걸자. 첫 책은 많아야 $2 \leq n$ 단위만 필요하므로, 어떤 책이든 들어간다.

- 첫 책이 얇으면(확률 q) 책꽂이에 꽂히고 $n-1$ 단위가 남는다.
- 첫 책이 두꺼우면(확률 p) 책꽂이에 꽂히고 $n-2$ 단위가 남는다.

> **생각해 보기** 첫 번째 경우에 **추가로** 꽂히는 책 수의 기댓값이 왜 정확히 $M(n-1)$인가?

첫 책 다음의 책들은 차례대로 도착하고, 첫 책과 독립적으로 같은 확률로 두껍거나 얇으며, 같은 규칙에 따라 놓인다. 즉 들어가지 않는 책이 나올 때까지 꽂고, 그러면 멈춘다. 첫 책이 바꾼 것은 빈 공간뿐이며, 이제 그 공간은 $n-1$ 단위이다. 따라서 그날의 나머지는 길이 $n-1$인 책꽂이로 보내는 하루 전체의 복사본이고, 추가로 꽂히는 책 수의 기댓값은 $M(n-1)$이다. 마찬가지로 첫 책이 두꺼우면 그 값은 $M(n-2)$이다. ($n-2=0$일 때도 소문항 (i)에 의해 이는 성립한다.)

첫 번째 책까지 세면, 전체 기댓값의 법칙(law of total expectation)에 의해

$$\begin{aligned} M(n) &= q\big(1 + M(n-1)\big) + p\big(1 + M(n-2)\big) \\ &= (p+q) + qM(n-1) + pM(n-2), \end{aligned}$$

이고, $p+q=1$이므로

$$M(n) - qM(n-1) - pM(n-2) = 1 \qquad (n \geq 2),$$

가 되어 요구된 식을 얻는다.

닫힌 형태

먼저 동차방정식 $M(n) - qM(n-1) - pM(n-2) = 0$을 보자. $M(n) = \lambda^n$을 대입해 보면 특성방정식(auxiliary equation)

$$\lambda^2 - q\lambda - p = 0.$$

을 얻는다. $q = 1 - p$이므로 $\lambda = 1$은 근이고$(1 - q - p = 0)$, 두 근의 곱은 $-p$이므로

$$\lambda^2 - q\lambda - p = (\lambda - 1)(\lambda + p).$$

이다. 따라서 임의의 A와 B에 대하여 $A(-p)^n$과 상수 B는 동차방정식을 만족한다.

1이 근이므로 상수로는 우변의 1을 만들 수 없다. 그래서 Cn을 시도한다. 대입하면

$$\begin{aligned} Cn - qC(n-1) - pC(n-2) &= C\big[(1-q-p)n + q + 2p\big] \\ &= C(q + 2p) = C(1 + p). \end{aligned}$$

이다. 이 값은 $C = \dfrac{1}{1+p}$일 때 1이 되며, $1 + p > 0$이므로 문제없다. 그러면 선형성(linearity)에 의해 A와 B가 무엇이든 $M(n) = A(-p)^n + B + Cn$은 모든 $n \geq 2$에 대하여 (iii)의 방정식을 만족한다.

이제 초깃값을 맞추자. 소문항 (i)에서

$$M(0) = A + B = 0, \qquad \text{따라서} B = -A,$$

이고, 소문항 (ii)에서

$$M(1) = -Ap + B + C = -A(1 + p) + \frac{1}{1+p} = 1 - p.$$

이다. 그러므로

$$A(1 + p) = \frac{1}{1+p} - (1 - p) = \frac{1 - (1 - p^2)}{1 + p} = \frac{p^2}{1+p},$$

이고, 따라서

$$A = \frac{p^2}{(1+p)^2}, \qquad B = -\frac{p^2}{(1+p)^2}, \qquad C = \frac{1}{1+p}.$$

이다. 즉

$$M(n) = \frac{n}{1+p} - \frac{p^2}{(1+p)^2}\big(1 - (-p)^n\big).$$

이다. (i), (ii), (iii)이 $M(0)$, $M(1)$, 그리고 그 뒤의 각 값을 차례로 결정하므로, 이것은 단지 가능한 해 하나가 아니라 유일한 해이다.

검산

- $p = 0$(모두 앓음): $M(n) = n$이며, 당연히 그래야 한다.
- $p = 1$(모두 두꺼움): $M(n) = \frac{n}{2} - \frac{1}{4}\big(1 - (-1)^n\big)$이고, 이는 짝수 n에 대해서는 $\frac{n}{2}$, 홀수 n에 대해서는 $\frac{n-1}{2}$, 즉 $\lfloor n/2 \rfloor$이다. 옳다.

- $p = q = \frac{1}{2}$일 때 점화식에서 직접 구하면 $M(2) = 1 + \frac{1}{4} = \frac{5}{4}$이고, 공식으로는 $\frac{2}{3} - \frac{1}{9} \cdot \frac{3}{4} = \frac{5}{4}$이다.
- n이 클 때 각 책은 평균 $1 + p$ 단위를 차지하므로 약 $n/(1+p)$권의 책을 기대할 수 있다. 공식도 이와 일치하며, 책꽂이 끝에서 낭비되는 공간 때문에 생기는 유계인 보정항이 붙는다.

우리는 $n \leq 12$와 여러 p 값에 대하여 책 꽂기 과정을 정확히 열거한 결과, 그리고 시뮬레이션 결과와 닫힌 형태를 비교해 확인하였다. 예를 들어 $p = \frac{1}{2}$, $n = 5$일 때 두 방법 모두 $M(5) = \frac{103}{32} \approx 3.219$를 준다.

흔한 실수

- 두꺼운 책이 들어가지 못한 뒤에도 얇은 책은 여전히 꽂을 수 있다고 생각하는 것. 문제에서는 그 뒤의 모든 책이 수레로 간다고 했고, 바로 그 때문에 $M(1) = q$이며 더 큰 값이 아니다.
- 특수해로 상수를 시도하는 것. $\lambda = 1$이 특성방정식의 해이므로 상수는 동차 부분에 흡수되어 버리고, Cn이 필요하다.

STEP 너머

첫 단계에 대해 조건을 걸고 남은 부분이 같은 문제의 더 작은 복사본임을 알아차리는 방법을 **첫 단계 분석**(first-step analysis)이라 한다. 이는 마르코프 연쇄(Markov chain)에서 기대 도달 시간(expected hitting time)과 흡수 확률(absorption probability)을 구하는 표준적인 방법이다. 우리의 답은 **재생 정리**(renewal theorem)의 한 예이기도 하다. 크기가 무작위인 물건들을 길이 n에 채워 넣으면, 들어가는 개수는 n을 평균 크기로 나눈 값처럼 증가한다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 무작위 행보(random walk)에 대한 차분방정식과 마르코프 연쇄 입문을 모두 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.