풀이

STEP III 2014, 문제 13

확률과 통계 • 확률생성함수

> **STEP III 2014, 13번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/14-s3-q13

핵심 아이디어

첫 번째 라운드에서 일어나는 일로 조건부(conditioning)를 걸어 생각한다. 이 게임에는 기억이 없다. 게임이 계속되는 라운드가 끝난 뒤에 이어지는 것은 같은 게임의 완전히 새로운 복사본이므로, 그 시점부터 **추가로** 얻는 점수는 N 자체와 같은 분포(distribution)를 가진다. 이로부터 G를 자기 자신으로 나타내는 방정식을 얻고, 이를 풀어 기하급수(geometric series)로 전개한다.

풀이

$p_n = P(N = n)$이라 하면 $G(t) = \sum_{n \geq 0} p_n t^n$이다. 첫 번째 라운드의 결과가 각각 1, 2, 3인 사건을 E_1, E_2, E_3이라 하자. 그러면 $P(E_1) = a$, $P(E_2) = b$, $P(E_3) = c$이다.

소문항 (i)

E_1이 주어지면 게임은 점수 0으로 끝나므로 $P(N = 0 \mid E_1) = 1$이고, $n \geq 1$에 대하여 $P(N = n \mid E_1) = 0$이다. 따라서 조건부 생성함수(generating function)는

$$\sum_{n \geq 0} P(N = n \mid E_1)\, t^n = 1 \cdot t^0 = 1.$$

소문항 (ii)

E_2가 주어지면 첫 라운드가 끝난 뒤에도 점수는 여전히 0이고 게임은 계속된다. 각 라운드의 확률은 모두 같고 이전에 일어난 일과 무관하므로, 두 번째 라운드부터의 라운드들은 점수 0에서 시작하는 새로운 게임을 이룬다. 이 새 게임에서 얻는 점수를 N'이라 하면, N'은

N과 같은 분포를 가진다. 최종 점수는 $0 + N' = N'$이므로

$$P(N = n \mid E_2) = P(N' = n) = p_n \quad \text{모든} n \geq 0,$$

이고, 조건부 생성함수는 $\sum p_n t^n = G(t)$이다.

소문항 (iii)

같은 방식으로, E_3이 주어지면 첫 라운드가 끝난 뒤 점수는 1이고 나머지는 점수 N'을 내는 새로운 게임이므로 최종 점수는 $1 + N'$이다. 따라서 $P(N = 0 \mid E_3) = 0$이고 $n \geq 1$에 대하여 $P(N = n \mid E_3) = p_{n-1}$이다. 조건부 생성함수는 $\sum_{n\geq 1} p_{n-1} t^n = tG(t)$이다.

전확률 법칙(law of total probability)에 의하여

$$p_n = a\,P(N = n \mid E_1) + b\,P(N = n \mid E_2) + c\,P(N = n \mid E_3).$$

$G(t)$에서 t^n의 계수는 p_n이다. 우변에서 1, $G(t)$, $tG(t)$의 t^n의 계수는 정확히 소문항 (i)–(iii)에서 구한 $P(N = n \mid E_1)$, $P(N = n \mid E_2)$, $P(N = n \mid E_3)$이다. 그러므로 모든 $n \geq 0$에 대하여 $G(t)$와 $a + bG(t) + ctG(t)$의 t^n의 계수가 일치하고, 따라서 다음이 증명된다.

$$G(t) = a + bG(t) + ctG(t).$$

> **생각해 보기** $1 - b - ct$로 나눌 수 있는가? 그리고 기하급수로 전개하는 것이 정당한가?

식을 정리하면 $G(t)(1 - b - ct) = a$이다. $1 - b = a + c > c > 0$이므로 $|t| < (1 - b)/c$일 때마다 $1 - b - ct \neq 0$이고, $(1 - b)/c = 1 + a/c > 1$이므로 이 범위는 $|t| \leq 1$을 포함한다. 이러한 t에 대하여 $|ct/(1 - b)| < 1$이므로

$$G(t) = \frac{a}{1 - b - ct} = \frac{a}{1 - b} \cdot \frac{1}{1 - \dfrac{ct}{1 - b}} = \frac{a}{1 - b} \sum_{n\geq 0} \left(\frac{c}{1 - b}\right)^n t^n,$$

이다. t^n의 계수를 읽으면

$$P(N = n) = \frac{ac^n}{(1 - b)^{n+1}}, \qquad n \geq 0.$$

(같은 이야기로, 계수를 비교하면 $p_0 = a + bp_0$이고 $n \geq 1$에 대하여 $p_n = bp_n + cp_{n-1}$이다. 즉 $p_0 = a/(1 - b)$이고 $p_n = \frac{c}{1-b} p_{n-1}$이며, 이로부터 같은 공식을 얻는다.)

소문항 (iv)

평균(mean)은 $\mu = G'(1)$이다. $G(t) = a(1 - b - ct)^{-1}$을 미분하고 $1 - b - c = a$를 이용하면

$$G'(t) = \frac{ac}{(1 - b - ct)^2}, \qquad \text{따라서} \qquad \mu = G'(1) = \frac{ac}{(1 - b - c)^2} = \frac{ac}{a^2} = \frac{c}{a},$$

이다. 그러면 $1+\mu = \dfrac{a+c}{a} = \dfrac{1-b}{a}$이므로

$$\frac{\mu^n}{(1+\mu)^{n+1}} = \frac{c^n}{a^n}\cdot\frac{a^{n+1}}{(1-b)^{n+1}} = \frac{ac^n}{(1-b)^{n+1}} = P(N=n),$$

이 되어 원하는 결과를 얻는다.

검산

- 확률의 합은 1이다: $\sum_n p_n = G(1) = a/(1-b-c) = a/a = 1$.
- 직접적인 논증: 결과가 2인 "낭비된" 라운드는 무시하자. 남은 각 라운드는 확률 $a/(a+c)$로 멈춤이고 확률 $c/(a+c)$로 1점이다. 따라서 N은 처음 멈추기 전까지 얻은 점수의 개수이고, $P(N=n) = \left(\frac{c}{a+c}\right)^n \frac{a}{a+c}$이다. $a+c = 1-b$이므로 이는 같은 공식이다.
- $a = \frac{1}{5}$, $b = \frac{3}{10}$, $c = \frac{1}{2}$이면 $\mu = \frac{5}{2}$이고 $P(N=0) = \frac{2}{7} \approx 0.2857$, $P(N=1) = \frac{10}{49} \approx 0.2041$이다. 2×10^5번의 게임을 컴퓨터로 시뮬레이션한 결과 0.2855, 0.2038, 그리고 평균 점수 2.498을 얻었다.

흔한 실수

- 경우 3의 조건부 생성함수를 $tG(t)$가 아니라 $G(t)$로 쓰는 실수. 이미 얻은 1점 때문에 t의 모든 거듭제곱이 하나씩 올라간다.
- 첫 라운드 이후의 게임이 원래 게임의 복사본인 **이유**를 말하지 않는 실수. 이는 각 라운드가 과거와 독립(independent)으로 같은 확률을 가진다는 사실에 근거한다.

STEP 너머

첫 단계로 조건부를 거는 것은 마르코프 연쇄(Markov chain) 이론의 기본 도구로, 그곳에서는 **첫걸음 분석**(*first-step analysis*)이라 부른다. 그리고 여기서 얻은 답은 모습을 바꾼 기하분포(geometric distribution)이다. 소문항 (iv)는 $\{0,1,2,\dots\}$ 위의 기하분포가 평균만으로 결정된다는 것을 보여 준다. G에 대한 방정식을 세우는 같은 기법으로 분기 과정(branching process)도 다룰 수 있으며, 이는 가문의 성(姓)이 사라질 확률에 관한 유명한 결과로 이어진다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 생성함수와 분기 과정을 매우 읽기 쉽게 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.