풀이

STEP II 2003, 문제 12

확률과 통계 • 분기 과정

STEP II 2003, 12번 문제

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

step.maths.org/questions/03-s2-q12

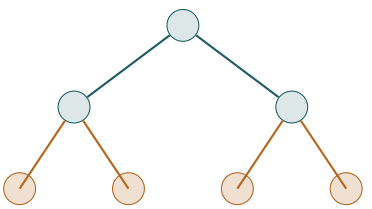
이 문제는 *Beyond STEP*의 12.15 절 **멸종인가 폭발인가**에서 단계별 안내와 함께 자세히 다룬다:
https://beyondstep.keithmcnulty.org/ko/ch12/branching.html

핵심 아이디어

이것은 분지 과정(branching process)이다. 각 세대는 바로 앞 세대로부터 만들어지며, 모든 입자는 서로 독립적으로 분열한다. 모든 계산을 떠받치는 단 하나의 사실은, m개의 입자가 각각 독립적으로 $\mathrm{B}(2, p)$를 따르는 수의 자손을 낳으면 자손의 총수는 $\mathrm{B}(2m, p)$를 따른다는 것이다. 그래서 확률을 구할 때도, 기댓값을 구할 때도 앞 세대의 크기를 조건으로 둔다.

풀이

모형의 의도대로, 서로 다른 입자들은 서로에 대해서도, 과거에 대해서도 독립적으로 분열한다고 가정한다. $0 < t < T$ 동안 살아 있는 하나의 입자에 대하여 $X_0 = 1$이라 쓴다. $t = nT$에서 살아 있는 X_{n-1}개의 입자는 각각 자기 자손으로 바뀌므로, X_n은 제$n-1$세대가 낳은 자손의 총수이다.



$X_0 = 1$

$X_1 = 2$

$X_2 \leq 4$

그림은 $X_1 = 2$인 경우를 보여 준다. 이때 두 입자는 합쳐서 최대 네 개의 자손을 낳고, 가능한 네 자손은 각각 독립적으로 확률 p로 태어난다.

결합 확률

$X_1 \sim \mathrm{B}(2, p)$이므로 $\mathrm{P}(X_1 = 2) = p^2$이다. $X_1 = 2$일 때 두 입자는 각각 독립적으로 $\mathrm{B}(2, p)$를 따르는 수의 자손을 낳으므로, X_2는 성공 확률이 p인 독립 시행 4번에서의 성공 횟수이다. 즉 $X_2 \mid X_1 = 2 \sim \mathrm{B}(4, p)$이다. 따라서

$$\mathrm{P}(X_1 = 2 \text{ 그리고} X_2 = 2) = \mathrm{P}(X_1 = 2)\,\mathrm{P}(X_2 = 2 \mid X_1 = 2) = p^2 \cdot \binom{4}{2} p^2 q^2 = 6p^4 q^2,$$

이고, 이것이 구하려던 결과이다.

조건부 확률

조건부 확률(conditional probability)의 정의를 쓰려면 $\mathrm{P}(X_2 = 2)$가 필요하다. 이것은 X_1을 조건으로 두어 구한다.

- $X_1 = 0$이면 개체군이 소멸했으므로 $X_2 = 0$이다.
- $X_1 = 1$이면(확률 $2pq$) $X_2 \sim \mathrm{B}(2, p)$이고 $\mathrm{P}(X_2 = 2 \mid X_1 = 1) = p^2$이다.
- $X_1 = 2$이면 그 기여는 방금 구한 $6p^4q^2$이다.

따라서

$$\mathrm{P}(X_2 = 2) = 2pq \cdot p^2 + 6p^4 q^2 = 2p^3 q\,(1 + 3pq),$$

이고, $0 < p < 1$일 때

$$\mathrm{P}(X_1 = 2 \mid X_2 = 2) = \frac{6p^4 q^2}{2p^3 q(1 + 3pq)} = \frac{3pq}{1 + 3pq}.$$

이다. 이 값을 $\frac{9}{25}$로 두면 $75pq = 9 + 27pq$이므로 $48pq = 9$, 즉 $pq = \frac{3}{16}$이다. $q = 1 - p$를 대입하면

$$p^2 - p + \tfrac{3}{16} = 0 \quad \Longleftrightarrow \quad (p - \tfrac{1}{4})(p - \tfrac{3}{4}) = 0,$$

이므로 $p = \frac{1}{4}$ 또는 $p = \frac{3}{4}$이다. ($p = 0$과 $p = 1$인 경우는 제외된다. $p = 0$이면 사건 $X_2 = 2$가 일어날 수 없고, $p = 1$이면 조건부 확률이 0이다.)

> **생각해 보기** 왜 두 p 값이 같은 답을 주는가?

조건부 확률은 오직 $pq = p(1 - p)$를 통해서만 p에 의존하는데, 이 값은 p를 $1 - p$로 바꾸어도 변하지 않는다. 따라서 근은 $p, 1 - p$ 꼴의 쌍으로 나와야 하며, 여기서는 $\frac{1}{4}$와 $\frac{3}{4}$이다.

기댓값

$X_{n-1} = m$이라 하자. 이 m개의 입자는 각각 평균 $\mathrm{E}(X) = 2p$개의 자손을 낳으므로, 다음 세대 크기의 기댓값은 $2pm$이다.

$$\mathrm{E}(X_n \mid X_{n-1} = m) = 2pm.$$

가능한 m 값들에 대하여 평균을 내면(전체 기댓값의 법칙(law of total expectation))

$$\mathrm{E}(X_n) = \sum_m \mathrm{P}(X_{n-1} = m)\, 2pm = 2p\,\mathrm{E}(X_{n-1}).$$

이다. 말로 하면, 한 구간에 살아 있는 모든 입자는 다음 구간에 평균 $2p$개의 입자를 보탠다. $\mathrm{E}(X_1) = 2p$(또는 $\mathrm{E}(X_0) = 1$)이므로, 이 점화식(recurrence)은 등비수열(geometric sequence)을 준다.

$$\mathrm{E}(X_n) = (2p)^n.$$

극한

$p = \frac{1}{4}$이면 $0 < \frac{1}{2} < 1$이므로 $n \to \infty$일 때 $\mathrm{E}(X_n) = \left(\frac{1}{2}\right)^n \to 0$이다.

$p = \frac{3}{4}$이면 $\frac{3}{2} > 1$이므로 $\mathrm{E}(X_n) = \left(\frac{3}{2}\right)^n \to +\infty$이다.

이처럼 조건부 확률에서는 똑같아 보이던 두 p 값이 장기적으로는 전혀 다른 행동으로 이어진다. 평균적으로 개체군은 $p = \frac{1}{4}$일 때 사라져 가고, $p = \frac{3}{4}$일 때 폭발적으로 늘어난다.

검산

- $p = \frac{3}{4}$일 때 $6p^4q^2 = \frac{243}{2048} \approx 0.1187$이고, 두 세대짜리 개체군 10^6개를 시뮬레이션하면 0.1186이 나왔다.
- $p = \frac{1}{4}$일 때 $pq = \frac{3}{16}$이므로 $3pq = \frac{9}{16}$이고 $\dfrac{9/16}{25/16} = \frac{9}{25}$로, 마땅히 그래야 하는 값이 나온다.
- 여러 p에 대하여 $X_1, \ldots, X_5$의 정확한 분포(각 세대는 앞 세대가 주어질 때 $\mathrm{B}(2m, p)$)를 계산하면 평균이 정확히 $(2p)^n$이 된다.

흔한 실수

- $\mathrm{P}(X_2 = 2)$를 구할 때 $X_1 = 1$, $X_2 = 2$인 경로를 빠뜨리는 것. $X_1 = 0$을 거치는 경로는 올바르게 아무것도 기여하지 않는다.
- $X_2 \mid X_1 = 2$를 서로 독립인 두 복사본의 합, 즉 $\mathrm{B}(4, p)$로 보지 않고 $2 \times \mathrm{B}(2, p)$(한 입자의 자손 수의 두 배)로 보는 것.

STEP 너머

이것은 골턴–왓슨 분지 과정(Galton–Watson branching process)으로, 처음에는 성씨가 왜 사라지는지를 설명하기 위해 연구되었다. 자손 수의 평균 $\mu = 2p$가 모든 것을 결정한다. $\mu \leq 1$이면 개체군은 확률 1로 소멸하고, $\mu > 1$이면 양의 확률로 영원히 살아남는다. $p = \frac{3}{4}$일 때조차 개체군은 여전히 확률 $\frac{1}{9}$로 소멸하는데, 이 값은 확률생성함수(probability generating function)에서 나오는 방정식 $s = (q + ps)^2$의 가장 작은 근이다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 생성함수를 써서 분지 과정을 다루는데, 그 내용이 이 문제에서 곧바로 이어진다.

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.