풀이

STEP III 2019, 문제 12

확률과 통계 • 무작위 부분집합

> **STEP III 2019, 12번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 OCR에서 내려받을 수 있는 2019년 STEP 시험지 묶음(zip 파일) 가운데 STEP III 시험지에 있다:
>
> www.ocr.org.uk/Images/696590-2019-past-papers.zip

이 문제는 *Beyond STEP*의 9.12 절 **무작위 부분집합**에서 단계별 안내와 함께 자세히 다룬다: https://beyondstep.keithmcnulty.org/ko/ch09/random-subsets.html

핵심 아이디어

S의 부분집합(subset)을 균등하게 무작위로(uniformly at random) 고르는 것은 정수 $1, 2, \ldots, n$을 하나씩 차례로 살피면서 각 정수마다 공정한 동전을 던져 그 정수를 넣을지 말지 정하는 것과 같다. 이 점을 알아차리면 문제의 모든 사건(event)은 각 정수에 대한 개별적인 진술이 되고, 정수들은 서로 독립(independent)으로 행동하며, 답은 하나의 작은 확률을 n제곱한 값이 된다.

풀이

부분집합의 개수 세기

S의 부분집합은 n개의 정수 각각에 대해 넣을지 뺄지를 정하면 완전히 결정된다. 각 정수마다 독립적으로 2가지 선택이 있고, 선택의 패턴이 다르면 서로 다른 부분집합이 나온다. 따라서 $\varnothing$(모두 뺀 경우)과 S(모두 넣은 경우)를 포함한 부분집합의 개수는

$$2 \times 2 \times \cdots \times 2 = 2^n.$$

이다.

> ? **생각해 보기** 왜 정수들을 독립적인 동전 던지기로 취급해도 되는가?

A가 T 위에서 균등분포를 따른다고 하자. 2^n개의 넣기/빼기 패턴 각각이 하나의 부분집합

이므로, 각 패턴의 확률은 $2^{-n} = (\frac{1}{2})^n$이다. 이는 독립적인 공정한 동전 n개를 던질 때 특정한 패턴이 나올 확률과 정확히 같다. 따라서 사건 "$1 \in A$", "$2 \in A$", ..., "$n \in A$"는 서로 독립이고, 각각의 확률은 $\frac{1}{2}$이다. 또 $A_1, \ldots, A_m$도 서로 독립적으로 선택되므로, mn개의 사건 "$i \in A_k$"는 모두 독립이고 각각의 확률은 $\frac{1}{2}$이다.

소문항 (i)

2^n개의 부분집합 중 정확히 절반이 1을 포함한다(1을 포함하는 각 부분집합을, 그 부분집합에서 1을 제거한 부분집합과 짝지으면 된다). 따라서

$$\mathrm{P}(1 \in A_1) = \tfrac{1}{2}.$$

이다.

소문항 (ii)

정수 i에 대하여, i가 A_1과 A_2에 동시에 속하지는 **않는** 사건을 E_i라 하자. 그러면 $A_1 \cap A_2 = \varnothing$인 것은 모든 E_i가 일어날 때, 그리고 그때에만 성립한다. 독립성에 의해

$$\mathrm{P}(i \in A_1 \text{ 그리고} i \in A_2) = \tfrac{1}{2} \cdot \tfrac{1}{2} = \tfrac{1}{4}, \qquad \text{이므로} \qquad \mathrm{P}(E_i) = \tfrac{3}{4}.$$

이다. 사건 E_i는 정수 i에 대한 동전에만 의존하므로 $E_1, \ldots, E_n$은 독립이고,

$$\mathrm{P}(A_1 \cap A_2 = \varnothing) = \mathrm{P}(E_1) \cdots \mathrm{P}(E_n) = \left(\tfrac{3}{4}\right)^n,$$

이 되어 원하는 결과를 얻는다.

같은 논증은 집합이 몇 개이든 성립한다. 정수 i가 $A_1 \cap A_2 \cap A_3$에 속할 확률은 $(\frac{1}{2})^3 = \frac{1}{8}$이므로, 교집합(intersection)을 피할 확률은 $\frac{7}{8}$이고,

$$\mathrm{P}(A_1 \cap A_2 \cap A_3 = \varnothing) = \left(\tfrac{7}{8}\right)^n.$$

이다. 일반적으로 i가 m개의 집합 모두에 속할 확률은 2^{-m}이므로

$$\mathrm{P}(A_1 \cap A_2 \cap \cdots \cap A_m = \varnothing) = \left(1 - \frac{1}{2^m}\right)^n.$$

이다.

소문항 (iii)

이번에는 $A_1 \subseteq A_2$가 A_1에는 있지만 A_2에는 없는 정수가 하나도 없다는 뜻이다. 각 정수 i에 대해 (A_1에 속하는가?, A_2에 속하는가?)의 경우는 똑같이 가능한 네 가지이고, 그중 "A_1에 속하고 A_2에 속하지 않음" 하나만 금지된다. 따라서 각 정수는 독립적으로 확률 $\frac{3}{4}$로 조건을 만족하고,

$$\mathrm{P}(A_1 \subseteq A_2) = \left(\tfrac{3}{4}\right)^n.$$

이다. (이는 소문항 (ii)와 일치한다. $A_1 \subseteq A_2$인 것은 A_1과 A_2의 여집합(complement)이 서로소(disjoint)일 때, 그리고 그때에만 성립하며, 균등한 무작위 부분집합의 여집합은 다시 균등한 무작위 부분집합이기 때문이다.)

$A_1 \subseteq A_2 \subseteq A_3$의 경우, 정수 하나 i가 (A_1, A_2, A_3)에 속하는지를 나타내는 패턴을 살펴보자. "속함"은 1, "속하지 않음"은 0으로 쓴다. 사슬(chain) 조건에 따르면 i가 어떤 A_k에 속하면 그 뒤의 모든 집합에도 속해야 하므로, 패턴은 다음 중 하나여야 한다.

$$000, \quad 001, \quad 011, \quad 111.$$

이는 똑같이 가능한 8개의 패턴 중 4개이므로, 각 정수가 조건을 만족할 확률은 $\frac{4}{8} = \frac{1}{2}$이고,

$$\mathrm{P}(A_1 \subseteq A_2 \subseteq A_3) = \left(\frac{1}{2}\right)^n.$$

이다.

m개의 집합으로 이루어진 사슬의 경우, 허용되는 패턴은 0들이 이어진 뒤 1들이 이어지는 형태이다. 이 패턴은 1이 시작되는 위치로 결정되는데, 그 위치는 $1, 2, \ldots, m$ 중 하나이거나 "시작하지 않음"이다. 그러면 2^m개 중 허용되는 패턴이 $m+1$개이므로

$$\mathrm{P}(A_1 \subseteq A_2 \subseteq \cdots \subseteq A_m) = \left(\frac{m+1}{2^m}\right)^n.$$

이다.

검산

- $m = 2$이면 일반 공식은 $(1 - \frac{1}{4})^n$과 $(\frac{3}{4})^n$을 주고, $m = 3$이면 $(\frac{7}{8})^n$과 $(\frac{4}{8})^n$을 주어 특수한 경우와 일치한다. $m = 1$이면 $(\frac{1}{2})^n$($A_1 = \varnothing$인 경우만 가능)과 1(집합 하나는 언제나 사슬)을 주는데, 마땅히 그래야 한다.
- $n = 1$이면 모든 경우를 손으로 나열할 수 있다. 예를 들어 네 가지 선택 중 교집합이 공집합이 아닌 것은 $A_1 = A_2 = \{1\}$ 하나뿐이므로 $\frac{3}{4}$를 얻는다.
- 집합의 모든 선택을 나열하는 컴퓨터 검사로 작은 n과 m에 대해 모든 공식을 확인하였다. 예를 들어 $n = 3$일 때 $\mathrm{P}(A_1 \cap A_2 = \varnothing) = \frac{27}{64}$이고, $\mathrm{P}(A_1 \subseteq A_2 \subseteq A_3) = \frac{1}{8}$이다.

흔한 실수

- A_1의 각 원소가 A_2에 속해야 한다는 논리로 $\mathrm{P}(A_1 \subseteq A_2)$의 답을 $\left(\frac{1}{2}\right)^n$이라고 하는 것. 이는 A_1 자체의 크기도 무작위라는 점을 고려하지 않은 것이다.
- 정수가 어느 집합에도 속하지 않는 패턴을 잊어버려서 사슬 패턴의 개수를 $m+1$이 아니라 m으로 세는 것.

STEP 너머

무작위 부분집합을 독립적인 동전 던지기의 나열(**지시 벡터**(indicator vector))로 나타내는 것은 조합론(combinatorics)의 표준적인 도구이며, 여기서의 사슬 세기는 **부분순서집합**(poset)과 불 격자(Boolean lattice) 이론의 출발점이다. 같은 아이디어를 쓰면 사슬 개수의 기댓값이나 무작위 집합들이 S를 덮을 확률도 별다른 추가 노력 없이 구할 수 있다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 이러한 관점에서 독립성과 지시 변수(indicator variable)를 전개한다.

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.