풀이

STEP II 2016, 문제 12

확률과 통계 • 포함배제와 교란순열

STEP II 2016, 12번 문제

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

step.maths.org/questions/16-s2-q12

이 문제는 *Beyond STEP*의 9.8 절 **아무도 자기 셔츠를 받지 못한다**에서 단계별 안내와 함께 자세히 다룬다: https://beyondstep.keithmcnulty.org/ko/ch09/derangements.html

핵심 아이디어

세 사건(event)의 경우에는 $A \cup B$를 하나의 사건으로 보고 두 사건에 대한 규칙을 두 번 적용한다. 같은 교대 패턴(한 사건씩 더하고, 두 사건의 교집합을 빼고, 세 사건의 교집합을 더하고, ...)은 사건이 몇 개이든 성립한다. 이를 "카드 i가 i번째 자리에 있다"는 사건들에 적용하면, r개의 사건으로 이루어진 항은 모두 확률이 같으므로 합의 각 층이 하나의 분수로 정리된다. "정확히 하나"의 경우에는 어느 카드가 일치하는지를 고정하고, 나머지 카드는 하나도 일치하지 않을 확률을 구한다.

풀이

세 사건

$A \cup B \cup C = (A \cup B) \cup C$로 쓰고, A 대신 $A \cup B$, B 대신 C를 넣어 두 사건에 대한 규칙을 적용한다.

$$\mathrm{P}(A \cup B \cup C) = \mathrm{P}(A \cup B) + \mathrm{P}(C) - \mathrm{P}\big((A \cup B) \cap C\big).$$

첫째 항은 $\mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$이다. 마지막 항은 분배법칙(distributive law)에 의해 $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$이고, 두 사건에 대한 규칙을 한 번 더 적용하면

$$\begin{aligned}\mathrm{P}\big((A \cap C) \cup (B \cap C)\big) &= \mathrm{P}(A \cap C) + \mathrm{P}(B \cap C) - \mathrm{P}\big((A \cap C) \cap (B \cap C)\big) \\ &= \mathrm{P}(A \cap C) + \mathrm{P}(B \cap C) - \mathrm{P}(A \cap B \cap C).\end{aligned}$$

두 식을 모두 대입하면

$$\mathrm{P}(A\cup B\cup C)=\mathrm{P}(A)+\mathrm{P}(B)+\mathrm{P}(C)-\mathrm{P}(A\cap B)-\mathrm{P}(B\cap C)-\mathrm{P}(C\cap A)+\mathrm{P}(A\cap B\cap C),$$

이 되어 요구된 결과를 얻는다.

네 사건

네 개의 단일 사건의 확률을 더하고, 여섯 개의 두 사건 교집합(intersection)의 확률을 빼고, 네 개의 세 사건 교집합의 확률을 더한 뒤, 네 사건 모두의 교집합의 확률을 뺀다.

$$\begin{aligned}\mathrm{P}(A\cup B\cup C\cup D)=\ &\mathrm{P}(A)+\mathrm{P}(B)+\mathrm{P}(C)+\mathrm{P}(D)\\&-\mathrm{P}(A\cap B)-\mathrm{P}(A\cap C)-\mathrm{P}(A\cap D)\\&-\mathrm{P}(B\cap C)-\mathrm{P}(B\cap D)-\mathrm{P}(C\cap D)\\&+\mathrm{P}(A\cap B\cap C)+\mathrm{P}(A\cap B\cap D)+\mathrm{P}(A\cap C\cap D)+\mathrm{P}(B\cap C\cap D)\\&-\mathrm{P}(A\cap B\cap C\cap D).\end{aligned}$$

소문항 (i)–(iii)

$n!$가지 배열은 모두 같은 정도로 일어날 수 있다. 특정한 r장의 카드가 각자 제자리에 있어야 한다면 나머지 $n-r$장의 카드는 $(n-r)!$가지 방법으로 배열할 수 있으므로, 그 확률은 $(n-r)!/n!$이다. 따라서

$$\text{(i) } \mathrm{P}(E_i)=\frac{1}{n},\qquad \text{(ii) } \mathrm{P}(E_i\cap E_j)=\frac{1}{n(n-1)},\qquad \text{(iii) } \mathrm{P}(E_i\cap E_j\cap E_k)=\frac{1}{n(n-1)(n-2)}.$$

((ii)는 다음과 같이 보아도 같다. 카드 i가 i번째 자리에 놓일 확률은 $\frac{1}{n}$이고, 그다음 카드 j가 남은 $n-1$개의 자리 중 j번째 자리에 놓일 확률은 $\frac{1}{n-1}$이다.)

적어도 하나가 일치할 확률

적어도 한 장의 카드가 일치하는 것은 정확히 $E_1\cup E_2\cup\cdots\cup E_n$이 일어날 때이다. 위의 패턴은 n개의 사건으로 확장된다. 즉 r번째 층은 서로 다른 r개의 사건의 모든 교집합의 확률을 합한 것에 $(-1)^{r+1}$을 곱한 것이다.

> **생각해 보기** 어떤 사건들인지 모르는 채로, r개의 사건을 고르는 모든 경우에 대한 합을 어떻게 계산할 수 있을까?

대칭성(symmetry)에 의해 서로 다른 r개의 사건의 교집합은 모두 같은 확률 $(n-r)!/n!$을 가지며, r개의 사건을 고르는 방법은 $\binom{n}{r}$가지이다. 따라서 r번째 층의 기여는

$$(-1)^{r+1}\binom{n}{r}\frac{(n-r)!}{n!}=(-1)^{r+1}\frac{n!}{r!\,(n-r)!}\cdot\frac{(n-r)!}{n!}=\frac{(-1)^{r+1}}{r!}.$$

$r = 1, 2, \ldots, n$에 대하여 합하면

$$\mathrm{P}(\text{적어도 하나 일치}) = 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n+1}\frac{1}{n!},$$

이 되어 요구된 결과를 얻는다. ($r = 1, 2, 3$일 때 이는 소문항 (i)–(iii)에서 얻은 $n \cdot \frac{1}{n}$, $\binom{n}{2} \cdot \frac{1}{n(n-1)}$, $\binom{n}{3} \cdot \frac{1}{n(n-1)(n-2)}$에 불과하다.)

정확히 하나가 일치할 확률

섞은 m장의 카드 중 어느 것도 제자리에 있지 않을 확률을 p_m이라 하자. 앞의 결과에 의해

$$p_m = 1 - \left(1 - \frac{1}{2!} + \cdots + (-1)^{m+1}\frac{1}{m!}\right) = \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^{m}\frac{1}{m!}.$$

"정확히 하나 일치"라는 사건은 "카드 i는 일치하고 다른 카드는 일치하지 않는다"는 사건들의 i에 대한 서로소인 합집합(disjoint union)이다. E_i가 일어났을 때, 나머지 $n-1$장의 카드는 나머지 $n-1$개의 자리에 어떤 순서로든 같은 정도로 놓일 수 있으므로, 이는 $n-1$장의 카드를 새로 섞는 것과 같고, 그중 어느 것도 일치하지 않을 확률은 p_{n-1}이다. 따라서

$$\mathrm{P}(\text{정확히 하나 일치}) = \sum_{i=1}^{n} \mathrm{P}(E_i)\, p_{n-1} = n \cdot \frac{1}{n} \cdot p_{n-1} = p_{n-1},$$

즉

$$\mathrm{P}(\text{정확히 하나 일치}) = \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^{n-1}\frac{1}{(n-1)!} \qquad (n \geq 2),$$

이고, $n = 1$일 때는 1이다. 같은 말로, 이 확률은 $\sum_{r=0}^{n-1}(-1)^r/r!$이며, 이 식은 $n = 1$인 경우도 포함한다.

검산

$n = 3$일 때 배열은 123, 132, 213, 231, 312, 321이다. 이 중 네 개에 일치하는 카드가 있으므로 확률은 $\frac{4}{6} = \frac{2}{3} = 1 - \frac{1}{2} + \frac{1}{6}$이다. 또 세 개(맞바꾼 배열 132, 213, 321)에서 정확히 하나가 일치하므로 확률은 $\frac{1}{2}$이고, 이는 공식과 일치한다. $n = 2$일 때는 정확히 하나만 일치하는 것이 불가능하며(한 카드가 제자리에 있으면 다른 카드도 제자리에 있다), 공식도 0을 준다. $n \leq 7$에 대하여 컴퓨터로 모든 배열을 나열해 보면 두 공식 모두와 일치한다. 예를 들어 $n = 7$일 때 정확히 하나가 일치할 확률은 $\frac{53}{144} \approx 0.368$이다. $n \to \infty$일 때 두 답은 모두 익숙한 극한(limit)으로 수렴한다. 적어도 하나가 일치할 확률은 $1 - \mathrm{e}^{-1} \approx 0.632$로, 정확히 하나가 일치할 확률은 e^{-1}로 수렴한다.

흔한 실수

- $\mathrm{P}(E_i \cap E_j) = \frac{1}{n^2}$로 쓰는 것. 이 사건들은 독립(independent)이 아니다. 카드 i가 i번째 자리에 놓이면 카드 j가 놓일 수 있는 자리는 $n-1$개만 남기 때문이다.

- "정확히 하나"의 답을 $n \cdot \frac{1}{n} = 1$이나 $\mathrm{P}(E_1)$로 쓰는 것. 카드 i가 제자리에 있고 **동시에** 다른 카드는 하나도 제자리에 있지 않아야 한다.
- "적어도 하나"에 대한 합에서 $r = 1$인 항이 $n \cdot \frac{1}{n} = 1$이라는 것을 잊는 것. 이 때문에 급수(series)가 1로 시작한다.

STEP 너머

이것은 고전적인 **교란순열**(*derangement*) 문제(*problème des rencontres*)이며, 일반적인 포함배제 원리(inclusion–exclusion) 공식은 수학적 귀납법(induction)이나 지시 확률변수(indicator random variables)를 이용하여 깔끔하게 증명된다. 큰 n에 대하여 일치하는 카드의 개수는 근사적으로 평균이 1인 푸아송 분포(Poisson distribution)를 따르는 것으로 밝혀진다. 즉 정확히 k개가 일치할 확률은 $\mathrm{e}^{-1}/k!$로 수렴한다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)과 Blitzstein과 Hwang의 *Introduction to Probability*(확률론 입문)는 모두 포함배제 원리와 일치 문제를 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.