풀이

STEP II 2004, 문제 14

확률과 통계 • 포함배제의 원리

STEP II 2004, 14번 문제

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

step.maths.org/questions/04-s2-q14

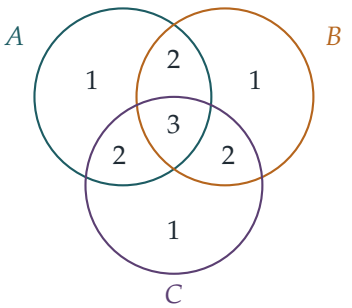
핵심 아이디어

포함배제 원리(inclusion–exclusion)는 중복으로 센 것을 바로잡는 셈법이다. $\mathrm{P}(A) + \mathrm{P}(B) + \mathrm{P}(C)$를 더하면 겹치는 부분을 너무 여러 번 세게 되는데, 두 사건씩의 항과 세 사건 모두의 항이 이 셈을 바로잡는다. 푸딩 문제에서는 **여사건**(complement)을 다루는 편이 훨씬 쉽다. E_i를 푸딩 i가 비어 있는 사건이라 하자. **어떤** 푸딩이 비어 있을 확률은 합사건(union)의 확률이고, 이것이 바로 포함배제 원리가 다루는 대상이다. 게다가 E_i들의 각 교사건(intersection)의 확률은 매우 간단하다.

풀이

포함배제 공식

$A \cup B \cup C$를 벤 다이어그램(Venn diagram)의 서로소(disjoint)인 일곱 영역으로 나눈다. 영역들을 통째로 모아 만든 사건의 확률은 그 영역들의 확률의 합이므로, 우변에서 각 영역이 정확히 한 번씩 세어짐을 확인하면 충분하다.



그림의 숫자는 각 영역이 A, B, C 가운데 몇 개에 속하는지를 나타낸다. 세 사건 가운데 정확히 k개에 속하는 영역을 하나 잡자. 여기서 $k = 1, 2$ 또는 3이다.

- 이 영역은 항 $\mathrm{P}(A), \mathrm{P}(B), \mathrm{P}(C)$ 가운데 k개에 나타난다.
- 이 영역은 두 사건씩의 교사건 가운데 $\binom{k}{2}$개에 나타난다. 예를 들어 이 영역이 $A \cap B$에 속하는 것은 A와 B 모두에 속할 때와 정확히 같기 때문이다.
- 이 영역은 $k = 3$일 때만 세 사건의 교사건에 나타난다. 즉 $\binom{k}{3}$번 나타난다.

따라서 이 영역이 최종적으로 세어지는 횟수는

$$k - \binom{k}{2} + \binom{k}{3} = \begin{cases} 1 - 0 + 0 = 1 & (k = 1), \\ 2 - 1 + 0 = 1 & (k = 2), \\ 3 - 3 + 1 = 1 & (k = 3). \end{cases}$$

이다. $A \cup B \cup C$의 모든 영역은 정확히 한 번씩 세어지고, 합사건 바깥의 영역은 전혀 세어지지 않는다. 따라서 우변은 $\mathrm{P}(A \cup B \cup C)$와 같다.

(같은 방법으로, 두 사건에 대한 법칙 $\mathrm{P}(X \cup Y) = \mathrm{P}(X) + \mathrm{P}(Y) - \mathrm{P}(X \cap Y)$를 $X = A \cup B$, $Y = C$에 적용한 뒤 다시 $(A \cap C) \cup (B \cap C)$에 적용해도 된다.)

각 푸딩에 6펜스 동전이 적어도 하나씩 들어갈 확률

모형. 골고루 저은 뒤 똑같은 양으로 나눈다는 것은, 각 6펜스 동전이 다른 동전과 독립적(independent)으로 푸딩 1, 2, 3에 각각 확률 $\frac{1}{3}$로 들어간다는 뜻이다. $r \geq 1$이라 하자.

E_i를 푸딩 i에 6펜스 동전이 하나도 없는 사건이라 하자.

- E_1은 r개의 동전이 모두 푸딩 1을 피할 때 일어난다. 각 동전이 그럴 확률은 $\frac{2}{3}$이므로 독립성에 의해 $\mathrm{P}(E_1) = \left(\frac{2}{3}\right)^r$이고, E_2, E_3도 마찬가지이다.
- $E_1 \cap E_2$는 모든 동전이 푸딩 3에 들어갈 때 일어나므로 $\mathrm{P}(E_1 \cap E_2) = \left(\frac{1}{3}\right)^r$이고, 나머지 쌍도 마찬가지이다.
- $E_1 \cap E_2 \cap E_3$이 일어나려면 세 푸딩이 모두 비어 있어야 하는데, $r \geq 1$이므로 이는 불가능하다. 따라서 그 확률은 0이다.

포함배제 원리에 의해, 적어도 한 푸딩이 비어 있을 확률은

$$p_r = \mathrm{P}(E_1 \cup E_2 \cup E_3) = 3\left(\tfrac{2}{3}\right)^r - 3\left(\tfrac{1}{3}\right)^r.$$

이다. 각 푸딩에 6펜스 동전이 적어도 하나씩 들어 있을 확률은 그 여사건의 확률이다.

$$1 - 3\left(\tfrac{2}{3}\right)^r + 3\left(\tfrac{1}{3}\right)^r = \frac{3^r - 3 \cdot 2^r + 3}{3^r}.$$

적어도 여섯 개의 6펜스 동전

$p_r < \frac{1}{3}$이 되는 r의 값을 구해야 한다. 직접 계산하면

$$p_5 = \frac{3 \cdot 32 - 3}{243} = \frac{93}{243} = \frac{31}{81} > \frac{27}{81} = \frac{1}{3}, \qquad p_6 = \frac{3 \cdot 64 - 3}{729} = \frac{189}{729} = \frac{7}{27} < \frac{9}{27} = \frac{1}{3}.$$

이다. 더 작은 r이 조건을 만족하지 않음을 확실히 하기 위해 p_r이 감소함을 보이자. $r \geq 1$일 때

$$p_{r+1} - p_r = 3\left(\tfrac{2}{3}\right)^r \left(\tfrac{2}{3} - 1\right) - 3\left(\tfrac{1}{3}\right)^r \left(\tfrac{1}{3} - 1\right) = -\left(\tfrac{2}{3}\right)^r + 2\left(\tfrac{1}{3}\right)^r = \left(\tfrac{1}{3}\right)^r (2 - 2^r),$$

이고, 이 값은 $r = 1$일 때 0, $r \geq 2$일 때 음수이다. 따라서 $p_1 = p_2 \geq p_3 \geq p_4 \geq p_5 = \frac{31}{81}$이고, $r \leq 5$인 모든 r에 대해 $p_r > \frac{1}{3}$이다. ($r = 0$이면 모든 푸딩이 반드시 비어 있다.) 그러므로 어떤 푸딩에 6펜스 동전이 없을 확률이 $\frac{1}{3}$보다 작으려면 $r \geq 6$이어야 한다. 또 p_r은 계속 감소하므로 $r \geq 6$인 모든 r에 대해 실제로 $p_r \leq p_6 < \frac{1}{3}$이다. 따라서 요리사는 요구된 대로 6펜스 동전을 6개 이상 사용해야 한다.

각 푸딩에 6펜스 동전이 두 개씩

이제 $r = 6$이다. T를 각 푸딩에 6펜스 동전이 정확히 두 개씩 들어 있는 사건, S를 각 푸딩에 적어도 하나씩 들어 있는 사건이라 하자. 앞 문항에서

$$\mathrm{P}(S) = 1 - p_6 = 1 - \frac{7}{27} = \frac{20}{27}.$$

이다. (서로 구별되는) 여섯 개의 동전을 푸딩에 보내는 방법은 $3^6 = 729$가지이고 모두 같은 정도로 일어날 수 있다. 각 푸딩에 두 개씩 보내는 방법의 수는 여섯 개의 대상을 이름이 붙은 세 쌍으로 나누는 방법의 수인 $\dfrac{6!}{2!\,2!\,2!} = 90$이다. 따라서 $\mathrm{P}(T) = \frac{90}{729} = \frac{10}{81}$이다.

> **?** **생각해 보기** 왜 $\mathrm{P}(T \cap S)$가 그냥 $\mathrm{P}(T)$일까?

모든 푸딩에 동전이 두 개씩 있으면 당연히 모든 푸딩에 적어도 하나씩 있으므로 $T \subseteq S$이고 $T \cap S = T$이다. 따라서 조건부확률(conditional probability)은

$$\mathrm{P}(T \mid S) = \frac{\mathrm{P}(T)}{\mathrm{P}(S)} = \frac{10/81}{20/27} = \frac{10}{81} \cdot \frac{27}{20} = \frac{1}{6}.$$

이다.

검산

$r = 1$과 $r = 2$일 때 공식은 모든 푸딩에 6펜스 동전이 들어 있을 확률을 0으로 주는데, 이는 옳다. 동전 세 개보다 적은 수로 푸딩 세 개를 채울 수는 없기 때문이다. $r = 3$일 때 공식은

$\frac{27-24+3}{27} = \frac{2}{9}$를 주는데, 이는 직접 센 값 $3!/3^3 = 6/27$과 일치한다. $r \leq 9$에 대해 3^r가지 배정을 모두 정확히 나열해 보면 공식과 일치하고, $p_5 = \frac{31}{81} \approx 0.383$과 $p_6 = \frac{7}{27} \approx 0.259$가 확인되며, 최종 답 $\frac{1}{6} \approx 0.167$을 얻는다(푸딩 200 000개를 시뮬레이션하면 0.166이 나왔다).

흔한 실수

- 6펜스 동전을 구별할 수 없다고 보고 $6 = a + b + c$로 나타내는 28가지 방법이 같은 정도로 일어난다고 세는 실수. 실제로는 그렇지 않다. $(2,2,2)$는 90가지 방법으로 일어날 수 있지만 $(6,0,0)$은 한 가지 방법으로만 일어난다.
- $r = 6$과 $r = 5$만 확인하는 실수. "6개 이상 사용해야 한다"를 보이려면 더 작은 모든 r이 조건을 만족하지 않음도 보여야 하며, 그래서 단조성(monotonicity)을 보이는 단계가 중요하다.
- 세 사건의 교사건이 공사건이라는 것을 잊거나, 아무 설명 없이 그 확률을 $(0)^r$로 두는 실수.

STEP 너머

n개의 사건에 대한 일반적인 포함배제 원리는 모든 k중 교사건의 교대합으로 이루어지며, $\sum_{j\geq 1}(-1)^{j-1}\binom{k}{j} = 1$을 이용하여 위와 똑같은 세기 논증으로 증명된다. 이를 n개의 푸딩에 적용하면 **전사**(surjection)의 개수를 세게 되며, 제2종 스털링 수(Stirling numbers of the second kind)와 고전적인 쿠폰 수집가 문제(coupon-collector problem)로 이어진다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)에서 포함배제 원리와 이러한 점유 문제들을 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.