풀이

STEP I 2007, 문제 13

확률과 통계 • 조건부 확률

> **STEP I 2007, 13번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/07-s1-q13

이 문제는 *Beyond STEP*의 12.1 절 **조건을 정확하게 걸기**에서 단계별 안내와 함께 자세히 다룬다: https://beyondstep.keithmcnulty.org/ko/ch12/conditioning.html

핵심 아이디어

원판은 한 번에 하나씩 나오지만 네 개의 원판으로 이루어진 모든 집합이 나올 가능성은 같으므로, 작업의 대부분은 조합(combination)의 개수를 세는 일이다. 흥미로운 부분은 조건에 있다. **어느** 원판이 먼저 나왔는지 알거나, 단순히 특정한 원판이 뽑혔다는 **사실**만 알면, 나머지 세 원판은 남은 열 개 중에서 무작위로 고른 것이 된다. 그러나 **어떤** 번호 원판이 뽑혔다는 것만 아는 경우는 전혀 다른 종류의 정보이며, 마지막 두 소문항은 바로 그 차이를 보여 주기 위한 것이다.

풀이

뽑은 네 원판 가운데 번호 원판의 개수를 N이라 하자. 주머니를 흔들었으므로 서로 다른 네 원판의 모든 순서 있는 나열은 가능성이 같고, 따라서 순서 없는 네 원판의 모든 집합도 가능성이 같다. 이러한 집합은

$$\binom{11}{4} = 330$$

개 있다. 번호 원판이 정확히 k개인 집합은 번호 원판 5개 중 k개, 빈 원판 6개 중 $4-k$개를 골라서 만들어지므로

$$\mathrm{P}(N=k) = \frac{\binom{5}{k}\binom{6}{4-k}}{330}.$$

소문항 (i)

네 개 모두 번호 원판이라는 것은 $k = 4$라는 뜻이다.

$$\mathrm{P}(N = 4) = \frac{\binom{5}{4}\binom{6}{0}}{330} = \frac{5}{330} = \frac{1}{66}.$$

같은 계산을 순서대로 뽑는 방식으로 하면 $\frac{5}{11} \cdot \frac{4}{10} \cdot \frac{3}{9} \cdot \frac{2}{8} = \frac{1}{66}$이다.

소문항 (ii)

원판 "3"이 처음에 뽑히는 사건을 T_1이라 하자. T_1이 주어지면 흔든 주머니에는 이제 번호 원판 4개와 빈 원판 6개, 모두 10개의 원판이 들어 있고, 다음 세 원판은 이들 중에서 무작위로 고른 것이다. 네 개가 모두 번호 원판인 것은 이 세 개가 모두 번호 원판일 때이고, 그때뿐이므로

$$\mathrm{P}(N = 4 \mid T_1) = \frac{\binom{4}{3}}{\binom{10}{3}} = \frac{4}{120} = \frac{1}{30}.$$

소문항 (iii)

여전히 T_1이 주어졌을 때, 남은 열 개에서 뽑은 세 원판 중 번호 원판이 정확히 하나 더 있어야 한다.

$$\mathrm{P}(N = 2 \mid T_1) = \frac{\binom{4}{1}\binom{6}{2}}{\binom{10}{3}} = \frac{4 \times 15}{120} = \frac{1}{2}.$$

소문항 (iv)

이제 원판 "3"이 어느 시점에든 뽑히는 사건을 T라 하자.

> ? **생각해 보기** 원판 "3"이 처음에 나왔는지, 아니면 다른 차례에 나왔는지가 중요한가?

중요하지 않다. T가 주어지면 뽑은 네 원판은 원판 "3"과 다른 세 원판으로 이루어지고, 대칭성(symmetry)에 의해 나머지 열 원판에서 고른 세 원판의 모든 집합은 가능성이 같다. 이는 정확히 소문항 (iii)의 상황이며, 거기서도 뽑는 순서는 전혀 중요하지 않았다. 따라서 답은 다시 $\frac{1}{2}$이다.

직접 확인해 보자. $\mathrm{P}(T) = \binom{10}{3}/\binom{11}{4} = \frac{120}{330} = \frac{4}{11}$이고, 원판 "3", 다른 번호 원판 하나, 빈 원판 두 개를 포함하는 집합의 개수는 $\binom{4}{1}\binom{6}{2} = 60$이므로

$$\mathrm{P}(N = 2 \mid T) = \frac{60/330}{120/330} = \frac{1}{2}.$$

소문항 (v)

첫 번째 원판이 번호 원판인 사건을 F라 하자. 첫 번째 원판에 적힌 번호에 따라 F를 나누자. F는 다섯 사건 "원판 j가 처음에 뽑힌다"($j = 1, \dots, 5$)의 서로소 합집합(disjoint union)이고, 각 사건의 확률은 $\frac{1}{11}$이다. 소문항 (iii)의 논증은 번호가 3이라는 사실을 사용하지 않았으므로, 각 j에 대하여

$$\mathrm{P}(N = 2 \mid \text{원판 } j\text{가 처음}) = \tfrac{1}{2}.$$

따라서 전확률 법칙(law of total probability)에 의해

$$\mathrm{P}(N = 2 \mid F) = \sum_{j=1}^{5} \frac{1/11}{5/11} \cdot \frac{1}{2} = \frac{1}{2}.$$

(직접 보면, F가 주어졌을 때 나머지 세 원판은 번호 원판 4개와 빈 원판 6개 중에서 무작위로 고른 세 개이므로 다시 소문항 (iii)과 같다.)

소문항 (vi)

번호 원판이 적어도 하나 뽑히는 사건을 A라 하자. 즉 A는 사건 $N \geq 1$이다. 이것은 앞에서와 같은 종류의 정보가 **아니다**. 어떤 원판 하나를 지목하고 나머지를 무작위로 남겨 두지 않기 때문이다. 대신 조건부 확률(conditional probability)의 정의를 사용한다. $N = 2$이면 A가 성립하므로

$$\mathrm{P}(N = 2 \mid A) = \frac{\mathrm{P}(N = 2)}{\mathrm{P}(N \geq 1)}.$$

> ? **생각해 보기** 왜 "번호 원판 하나가 알려져 있으니 나머지 세 개는 남은 열 개에서 무작위로 고른 것"이라고 그냥 말할 수 없는가?

어느 원판이 알려진 원판인지 모르기 때문이다. 번호 원판이 두 개인 집합도 번호 원판이 네 개인 집합과 똑같은 방식으로 A를 만족하며, 따로 떼어 놓을 수 있는 단 하나의 원판이 없다. 그래서 이 조건은 N 자체에 따라 집합들의 가중치를 다시 매기게 되고, 우리는 개수를 세어야 한다.

계산하면

$$\mathrm{P}(N = 2) = \frac{\binom{5}{2}\binom{6}{2}}{330} = \frac{10 \times 15}{330} = \frac{150}{330}, \qquad \mathrm{P}(N = 0) = \frac{\binom{6}{4}}{330} = \frac{15}{330},$$

이므로 $\mathrm{P}(N \geq 1) = \frac{315}{330}$이고

$$\mathrm{P}(N = 2 \mid A) = \frac{150}{315} = \frac{10}{21}.$$

검산

N의 분포(distribution)는 합이 맞아야 한다. $k = 0, \ldots, 4$에 대한 개수 $\binom{5}{k}\binom{6}{4-k}$는 $15, 100, 150, 60, 5$이고, 합은 330이다. 소문항 (ii)와 (iii)에서 추가로 나오는 번호 원판 개수의 조건부 분포는 120개 중 개수가 $20, 60, 36, 4$이며, 역시 합이 올바르다. 또한 $11 \times 10 \times 9 \times 8 = 7920$가지의 순서 있는 뽑기를 모두 나열하여 정확한 분수로 여섯 개의 답($\frac{1}{66}$, $\frac{1}{30}$, $\frac{1}{2}$, $\frac{1}{2}$, $\frac{1}{2}$, $\frac{10}{21}$)을 모두 확인하였다. $\frac{10}{21} \approx 0.476$이 $\frac{1}{2}$보다 조금 작다는 점에 주목하자. 이는 당연한 결과이다. $N \geq 1$이라는 조건은 $N = 0$인 경우만 제거하므로, 특정한 원판을 아는 경우보다 $N = 1$에 더 많은 가중치를 남긴다.

흔한 실수

- 소문항 (iv)와 (vi)를 같은 종류의 조건으로 취급하는 것. "원판 3이 뽑힌다"는 원판을 지목하지만 "번호 원판이 뽑힌다"는 그렇지 않으며, 답도 다르다.
- 소문항 (vi)에서 조건이 $N \geq 1$이라는 것을 잊고 $\mathrm{P}(N = 1)$이나 $\frac{5}{11}$로 나누는 것.
- 분자에서는 순서 있는 뽑기를, 분모에서는 순서 없는 집합을 세는 것. 하나를 골라 끝까지 그것을 유지해야 한다.

STEP 너머

소문항 (v)와 (vi)의 대비는 유명한 "아들 또는 딸" 역설과 같은 함정이다. **특정한** 아이가 아들이라는 말을 듣는 것과 **적어도 한** 아이가 아들이라는 말을 듣는 것은 서로 다른 답을 준다. 대학에서는 사건에 대한 조건부를 확률측도(probability measure)를 제한하고 다시 정규화하는 것으로 다루어 이를 엄밀하게 하며, 여기서 센 개수들은 초기하분포(hypergeometric distribution)를 이룬다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문), 그리고 Blitzstein과 Hwang의 *Introduction to Probability*(확률론 입문) 모두 이러한 역설을 자세히 논의한다.

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.