풀이

STEP I 2013, 문제 13

확률과 통계 • 경우의 수와 기댓값

> **STEP I 2013, 13번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/13-s1-q13

핵심 아이디어

수 $1, \ldots, 52$는 서로소(disjoint)인 26개의 "짝 쌍" $\{k, 53-k\}$로 나뉘고, 버려지는 쌍이란 두 원소를 모두 선택한 짝 쌍일 뿐이다. 따라서 완전한 쌍이 정확히 k개인 선택의 수를 세려면, 어느 쌍이 완전한지를 먼저 고른 다음 나머지 자리를 **서로 다른** 쌍에서 온 수로 채우되, 매번 짝 중 어느 쪽을 택할지는 자유롭게 고르면 된다. 기댓값(expectation)에 대해서는 26개의 쌍에 대한 선형성(linearity)을 이용하면 빠르게 확인할 수 있다.

풀이

선택의 수 세기

정수들을 26개의 쌍 $\{1,52\}, \{2,51\}, \ldots, \{26,27\}$로 나누어 쓰자. 1부터 52까지의 두 정수의 합이 53이 되는 것은 정확히 두 수가 이 쌍들 중 하나를 이룰 때이고, 각 정수는 정확히 한 쌍에 속한다. 따라서 X는 두 원소가 모두 선택된 일곱 개의 수 안에 들어 있는 쌍의 개수이다. $2X \le 7$이므로 X는 $0, 1, 2, 3$의 값을 가진다.

$\binom{52}{7}$가지 선택은 모두 같은 확률로 일어나므로, $X = k$인 선택의 수를 센다.

- 완전한 쌍 k개를 고른다: $\binom{26}{k}$가지. 이들은 일곱 개의 수 중 $2k$개를 차지한다.
- 나머지 $7-2k$개의 수는 남은 $26-k$개의 쌍 가운데 **서로 다른** $7-2k$개의 쌍에서 와야 한다(같은 쌍에서 두 수를 고르면 완전한 쌍이 하나 더 생긴다): $\binom{26-k}{7-2k}$가지.
- 그 각각의 쌍에서 두 원소 중 하나를 택한다: 2^{7-2k}가지.

$X = k$인 각 선택은 이 방법으로 정확히 한 번씩 나타나므로

$$\mathrm{P}(X = k) = \frac{\binom{26}{k}\binom{26-k}{7-2k}\,2^{7-2k}}{\binom{52}{7}}, \qquad k = 0, 1, 2, 3.$$

수치 계산

$$\begin{aligned}
X = 0: &\quad \binom{26}{7}2^7 = 657\,800 \times 128 = 84\,198\,400,\\
X = 1: &\quad 26\binom{25}{5}2^5 = 26 \times 53\,130 \times 32 = 44\,204\,160,\\
X = 2: &\quad \binom{26}{2}\binom{24}{3}2^3 = 325 \times 2024 \times 8 = 5\,262\,400,\\
X = 3: &\quad \binom{26}{3}\binom{23}{1}2 = 2600 \times 23 \times 2 = 119\,600,
\end{aligned}$$

이고 $\binom{52}{7} = 133\,784\,560$이다. 이 네 개의 수를 더하면 실제로 $\binom{52}{7}$가 되는데, 이는 빠뜨리거나 중복해서 센 것이 없다는 좋은 검산이다.

? **생각해 보기** 아홉 자리 수가 들어 있는 분수는 어떻게 약분할까?

문제에 있는 주석이 힌트이다. $5593 = 7 \times 17 \times 47$이 분모가 되어야 한다. 실제로 $\binom{52}{7} = 23\,920 \times 5593$이고, 네 개의 수는 모두 23 920의 배수이다.

$$\begin{aligned}
84\,198\,400 &= 3520 \times 23\,920, & 44\,204\,160 &= 1848 \times 23\,920,\\
5\,262\,400 &= 220 \times 23\,920, & 119\,600 &= 5 \times 23\,920.
\end{aligned}$$

따라서 X의 분포(distribution)는 다음과 같다.

k	0	1	2	3
$\mathrm{P}(X = k)$	$\frac{3520}{5593}$	$\frac{1848}{5593}$	$\frac{220}{5593}$	$\frac{5}{5593}$

(두 번째 값을 기약분수로 나타내면 $\frac{264}{799}$이다.) 그리고 마땅히 그래야 하듯이 $3520 + 1848 + 220 + 5 = 5593$이다.

기댓값

$$\mathrm{E}(X) = \frac{0 \times 3520 + 1 \times 1848 + 2 \times 220 + 3 \times 5}{5593} = \frac{2303}{5593}.$$

이제 $2303 = 7^2 \times 47$이고 $5593 = 7 \times 17 \times 47$이므로

$$\mathrm{E}(X) = \frac{7^2 \times 47}{7 \times 17 \times 47} = \frac{7}{17},$$

이 되어 요구한 결과를 얻는다.

기댓값을 구하는 더 빠른 방법

$j = 1, \ldots, 26$에 대하여 j번째 쌍의 두 원소가 모두 선택되면 $I_j = 1$, 그렇지 않으면 0이라 하자. 그러면 $X = I_1 + \cdots + I_{26}$이다. 주어진 한 쌍의 두 원소를 모두 고르면 나머지 50개에서 5개를 더 고르게 되므로

$$\mathrm{P}(I_j = 1) = \frac{\binom{50}{5}}{\binom{52}{7}} = \frac{7 \times 6}{52 \times 51} = \frac{7}{442}.$$

기댓값의 선형성에 의하여(이는 독립성을 필요로 하지 않는다. 실제로 I_j들은 서로 종속이다)

$$\mathrm{E}(X) = 26 \times \frac{7}{442} = \frac{7}{17}.$$

검산

- 네 확률의 합은 1이고, 두 방법 모두 같은 값 $\mathrm{E}(X) = \frac{7}{17} \approx 0.412$를 준다.
- 300 000번의 무작위 선택을 컴퓨터로 모의실험한 결과, $X = 0, 1, 2, 3$의 상대도수는 각각 약 0.630, 0.329, 0.039, 0.001로, 정확한 값 0.6294, 0.3304, 0.0393, 0.0009와 잘 맞았다.
- 상식적인 확인: 일곱 개의 수만으로 완전한 쌍을 세 개 만드는 일은 드물어야 하므로, $\mathrm{P}(X = 3)$이 매우 작은 것은 당연하다.

흔한 실수

- $X = 0$인 선택의 수를 $\binom{52}{7}$에서 대충 무언가를 뺀 값으로 세거나, 각 쌍에서 어느 원소를 택하는지에 대한 인수 2^7을 잊고 $\binom{26}{7}$만으로 세는 것.
- $X = 1$일 때 나머지 다섯 수를 고르는 방법을 $\binom{50}{5}$로 세는 것: 이렇게 하면 두 번째 완전한 쌍이 생길 수 있어 과다하게 세게 된다.

STEP 너머

개수를 지시확률변수(indicator variable)의 합으로 쓰고 기댓값의 선형성을 이용하는 것은 대학 1학년 확률론에서 가장 유용한 기법 중 하나이다. 이 방법을 쓰면 분포를 구하는 것이 절망적일 때에도 기댓값을 얻을 수 있다. 이 문제에서는 전체 분포를 구할 수 있었지만, 예를 들어 무작위 순열(permutation)의 고정점(fixed point) 개수의 기댓값을 구할 때에는 지시확률변수가 자연스러운 출발점이다. Blitzstein과 Hwang의 *Introduction to Probability*(확률론 입문)에서는 이 아이디어를 거듭해서 사용한다.

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.