풀이

STEP III 1991, 문제 15

확률과 통계 • 표본 속 쌍의 개수 세기

STEP III 1991, 15번 문제

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

step.maths.org/questions/91-s3-q15

핵심 아이디어

k장의 카드로 이루어진 모든 집합이 뽑힐 가능성은 똑같으므로, 각 확률은 조건을 만족하는 패(hand)의 개수를 $\binom{2n}{k}$로 나눈 값이다. 주어진 쌍 구조를 갖는 패의 개수를 세려면 단계별로 패를 만들어 나가면 된다. 먼저 어떤 종류가 완전한 쌍으로 나타나는지 고른다. 다음으로 다른 어떤 종류가 한 장씩만 나타나는지 고른다. 마지막으로 한 장만 있는 각 종류에 대해 두 장 가운데 어느 카드를 가지고 있는지 고른다. 이 개수를 구해 두면 문제의 모든 부분이 하나의 공식의 특수한 경우가 된다.

풀이

$2n$장의 카드를 모두 서로 구별할 수 있다고 생각하자(한 종류의 두 카드에 각각 1과 2라는 번호를 붙였다고 보면 된다). 그러면 k장의 카드를 비복원 추출(without replacement)로 뽑을 때, 크기가 k인 $\binom{2n}{k}$개의 부분집합이 모두 같은 확률로 뽑힌다. 어떤 종류의 두 카드가 모두 뽑혔을 때 패에 그 종류의 쌍(pair)이 있다고 한다.

네 장, 쌍 없음

쌍이 없는 패는 서로 다른 네 종류의 카드를 한 장씩 포함한다. 종류를 고르는 방법은 $\binom{n}{4}$가지이고, 각 종류에서 카드를 고르는 방법은 2가지이므로 이런 패는 $2^4\binom{n}{4}$개이다.

$$\binom{2n}{4} = \frac{2n(2n-1)(2n-2)(2n-3)}{24} = \frac{4n(n-1)(2n-1)(2n-3)}{24},$$

이므로

$$\mathrm{P}(\text{쌍 없음}) = \frac{16 \cdot n(n-1)(n-2)(n-3)/24}{4n(n-1)(2n-1)(2n-3)/24} = \frac{4(n-2)(n-3)}{(2n-1)(2n-3)},$$

이고, 이것이 구하려던 결과이다. (여기서 $n \geq 4$이므로 서로 다른 네 종류가 반드시 존재한다.)

네 장, 정확히 한 쌍

쌍으로 나타나는 종류를 고른다(n가지). 나머지 두 장은 남은 $n-1$개의 종류 가운데 서로 다른 두 종류에서 나와야 하는데, 이 두 종류를 고르는 방법은 $\binom{n-1}{2}$가지이고 각 종류마다 카드를 고르는 방법이 2가지이다. 따라서 패의 개수는

$$n \cdot \frac{(n-1)(n-2)}{2} \cdot 4 = 2n(n-1)(n-2),$$

이고, 그러므로

$$\mathrm{P}(\text{정확히 한 쌍}) = \frac{24 \cdot 2n(n-1)(n-2)}{4n(n-1)(2n-1)(2n-3)} = \frac{12(n-2)}{(2n-1)(2n-3)}.$$

k장, 정확히 r개의 쌍

패에 쌍이 정확히 r개 있다고 하자. 이 쌍들이 $2r$장의 카드를 차지한다. 나머지 $k-2r$장은 모두 서로 다른 종류에 속하며, 그 가운데 어느 것도 쌍을 이룬 r개의 종류에 속하지 않는다. 그렇지 않으면 쌍이 더 생기기 때문이다. 따라서 다음을 센다.

- 쌍을 이루는 종류: $\binom{n}{r}$가지;
- 한 장씩 나타나는 $k-2r$개의 종류를 나머지 $n-r$개의 종류에서 고르는 방법: $\binom{n-r}{k-2r}$가지;
- 한 장만 있는 각 종류에서 두 장 가운데 어느 카드를 가지는지: 2^{k-2r}가지.

쌍이 정확히 r개인 각 패는 이러한 선택의 순서 가운데 정확히 하나로부터 얻어진다. 그러므로

$$\boxed{\mathrm{P}(\text{정확히}r\text{개의 쌍}) = \frac{2^{k-2r}\binom{n}{r}\binom{n-r}{k-2r}}{\binom{2n}{k}}}$$

생각해 보기 이 값이 0이 되는 경우가 있는가? 그리고 네 장일 때의 답과 일치하는가?

한 장씩 나타나는 종류에 대해서는 $k-2r \leq n-r$, 즉 $r \geq k-n$이어야 한다. $k > n$이면 어떤 k장의 카드에도 쌍이 적어도 $k-n$개 있어야 하는데, $\binom{n-r}{k-2r} = 0$이 이 경우를 자동으로 처리해 준다. $k = 4$일 때 $r = 0$인 경우는 $2^4\binom{n}{4}/\binom{2n}{4}$, $r = 1$인 경우는 $2^2\, n\binom{n-1}{2}/\binom{2n}{4}$이고, 이는 위에서 구한 두 개수와 정확히 같다.

k가 짝수, 모든 카드가 쌍을 이룸

$k = 2m$이라 하자. 패가 m개의 쌍으로 이루어지는 것은 패가 m개의 완전한 종류의 합집합일 때, 그리고 그때뿐이다. 이런 패는 $\binom{n}{m}$개 있다(일반 공식에서 $r = m$인 경우이다). 따라서

$$\mathrm{P} = \frac{\binom{n}{m}}{\binom{2n}{2m}} = \frac{n!}{m!\,(n-m)!} \cdot \frac{(2m)!\,(2n-2m)!}{(2n)!}.$$

이를 간단히 하기 위해 계승(factorial)의 짝수 인수와 홀수 인수를 나누어 묶는다.

$$(2j)! = \big(2 \cdot 4 \cdots 2j\big)\big(1 \cdot 3 \cdots (2j-1)\big) = 2^j\, j!\,\big(1 \cdot 3 \cdots (2j-1)\big).$$

$O_j = 1 \cdot 3 \cdots (2j-1)$, $O_0 = 1$로 놓자. 그러면 $(2j)!/j! = 2^j O_j$이므로

$$\mathrm{P} = \frac{2^m O_m \cdot 2^{n-m} O_{n-m}}{2^n O_n} = \frac{O_m\, O_{n-m}}{O_n}.$$

O_n을 O_{n-m}으로 나누면 $2n-2m+1 = 2n-k+1$부터 $2n-1$까지의 홀수가 남으므로

$$\mathrm{P}(\tfrac{1}{2}k\text{개의 쌍}) = \frac{1 \times 3 \times 5 \times \cdots \times (k-1)}{(2n-1)(2n-3)\cdots(2n-k+1)},$$

이고, 이것이 구하려던 결과이다.

깔끔한 직접적인 논증도 있다. 카드 한 벌 전체를 섞고 맨 위의 k장을 패로 삼자. 맨 위 카드를 보자. 그 짝은 나머지 $2n-1$개의 위치 어디에나 같은 가능성으로 있을 수 있고, 그 가운데 $k-1$개가 패 안에 있으므로 짝이 패 안에 있을 확률은 $\frac{k-1}{2n-1}$이다. 짝이 패 안에 있다면 그 쌍을 제거한다. 남은 것은 $2n-2$장의 카드를 무작위로 배열한 것이고, 패는 그 가운데 맨 위의 $k-2$장이다. 같은 논증을 반복하면 $\frac{k-3}{2n-3}$을 얻고, 이렇게 계속하여 $\frac{1}{2n-k+1}$까지 내려간다. 이 조건부 확률(conditional probability)들을 곱하면 같은 곱을 얻는다.

검산

- 네 장일 때 가능한 경우는 쌍이 0개, 1개, 2개인 세 경우이고, 쌍이 2개일 확률은 $\binom{n}{2}/\binom{2n}{4} = \frac{3}{(2n-1)(2n-3)}$이다. 실제로

  $$4(n-2)(n-3) + 12(n-2) + 3 = 4n^2 - 8n + 3 = (2n-1)(2n-3),$$

  이므로 세 확률의 합은 1이다.

- $n = 5$이고 네 장일 때 확률은 $\frac{8}{21}$, $\frac{4}{7}$, $\frac{1}{21}$이다. $4 \le n \le 8$이고 k가 3부터 $2n-1$까지인 모든 경우에 대해 컴퓨터로 모든 패를 열거해 본 결과, 일반 공식 및 모두 쌍인 경우의 공식과 일치했다.

- $k = 2$일 때 모두 쌍인 경우의 공식은 $\frac{1}{2n-1}$을 주는데, 이는 두 번째 카드가 첫 번째 카드와 짝이 맞을 확률이다.

흔한 실수

- 한 장씩 있는 카드를 남은 $2n-2r$장 전체에서 $\binom{2n-2r}{k-2r}$로 고르는 것. 이렇게 하면 그중 두 장이 같은 종류일 수 있어서 쌍이 더 생긴다.
- 인수 2^{k-2r}를 빠뜨리는 것. 한 장씩 나타나는 각 종류는 두 카드 가운데 어느 것이든 될 수 있다.

STEP 너머

쌍이 정확히 r개인 패의 개수는 "구조를 먼저 고르고, 그다음에 채워 넣는" 전형적인 세기이고, 곱 $1\cdot 3\cdots(2m-1)$(이중 계승(double factorial) $(2m-1)!!$)은 $2m$개의 대상을 쌍으로 나누는 방법의 수를 센다. 이중 계승은 표준정규분포(standard normal distribution)의 짝수 차 적률(moment)로도 다시 등장한다. 무작위 표본에서 몇 개의 쌍이 살아남는지를 묻는 같은 짝 맞추기 문제는 고전적인 "잃어버린 양말" 문제나 생일 문제 유형의 문제에도 나타난다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문서)은 이런 종류의 세기에 대한 좋은 기초를 제공한다.

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.