풀이

STEP III 2023, 문제 12

확률과 통계 • 조합론적 확률

> **STEP III 2023, 12번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 OCR에서 내려받을 수 있는 2023년 STEP 시험지 묶음(zip 파일) 가운데 STEP III 시험지에 있다:
>
> www.ocr.org.uk/Images/707102-2023-past-papers.zip

핵심 아이디어

한 켤레를 이루는 두 양말은 똑같아 보이지만, 선택은 $2n$개의 **실제** 양말 가운데 $2k$개로 이루어진 집합을 균등하게 무작위로(uniformly at random) 고르는 것이다. 따라서 모든 양말에 이름표를 붙이고 집합의 개수를 센다. 그러면 모든 것은 선택을 단계별로 구성하는 문제가 된다. 먼저 어느 켤레가 온전히 나타나는지 고르고, 다음으로 다른 어느 켤레가 양말을 한 짝만 내놓는지 고르고, 마지막으로 그런 켤레 각각에서 어느 짝을 가져올지 고른다. 소문항 (iii)은 익숙한 기법 $r\binom{n}{r} = n\binom{n-1}{r-1}$을 쓰는데, 이 등식은 가중합을 확률들의 합으로 바꾸어 준다.

풀이

$2n$개의 양말이 모두 서로 구별되도록 이름표를 붙인다. 똑같은 가능성을 갖는 선택은 $\binom{2n}{2k}$가지이고, 아래의 모든 확률은 (유리한 선택의 수)/$\binom{2n}{2k}$이다.

소문항 (i)

온전한 켤레가 하나도 없는 선택은 각 켤레에서 많아야 한 짝만 포함하므로, 그 $2k$개의 양말은 서로 다른 $2k$개의 켤레에서 나온다. 이런 선택을 만들려면 양말을 내놓는 $2k$개의 켤레를 $\binom{n}{2k}$가지 방법으로 고르고($2k \le n$이므로 가능하다), 각 켤레에서 두 짝 중 하나를 2^{2k}가지 방법으로 고르면 된다. 이렇게 고른 각각은 서로 다른 집합을 주고, 켤레가 없는 모든 집합은 이런 방식으로 생기므로

$$\mathrm{P}(\text{짝 없음}) = \frac{\binom{n}{2k}\,2^{2k}}{\binom{2n}{2k}}.$$

소문항 (ii)

$0 \le r \le k$인 r을 고정한다. 온전한 켤레가 정확히 r개인 선택은 다음으로 이루어진다.

- 통째로 뽑힌 r개의 켤레($2r$개의 양말),
- 그리고 나머지 $2k - 2r = 2(k-r)$개의 양말. 이들은 모두 서로 다른 켤레에서 나오며(그렇지 않으면 켤레가 r개보다 많아진다), 이미 쓰인 r개의 켤레에서는 하나도 나오지 않는다.

따라서 온전한 켤레 r개를 $\binom{n}{r}$가지 방법으로 고르고, 남은 $n-r$개의 켤레에서 한 짝씩 내놓는 $2(k-r)$개의 켤레를 $\binom{n-r}{2(k-r)}$가지 방법으로 고르고, 이들 각각에서 한 짝을 $2^{2(k-r)}$가지 방법으로 고른다. 이번에도 켤레가 정확히 r개인 집합은 각각 정확히 한 번씩 만들어지므로

$$\mathrm{P}(X_{n,k} = r) = \frac{\binom{n}{r}\binom{n-r}{2(k-r)}\, 2^{2(k-r)}}{\binom{2n}{2k}},$$

이고, 이것이 구하려던 결과이다. ($r = 0$일 때 이것은 다시 소문항 (i)이 된다.)

소문항 (iii)

$1 \le r \le k$라 하자.

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

이므로 소문항 (ii)에서

$$r\,\mathrm{P}(X_{n,k} = r) = \frac{n\binom{n-1}{r-1}\binom{n-r}{2(k-r)}\, 2^{2(k-r)}}{\binom{2n}{2k}}.$$

이제 켤레가 $n-1$개이고, 뽑는 양말이 $2(k-1)$개이고, 찾은 켤레가 $r-1$개인 경우에 소문항 (ii)를 적용한다. $2k \le n$에서 $2(k-1) \le n-1$이 따라 나오고 $0 \le r-1 \le k-1$이므로 이렇게 해도 된다. $(n-1)-(r-1) = n-r$이고 $(k-1)-(r-1) = k-r$이므로

$$\mathrm{P}(X_{n-1,k-1} = r-1) = \frac{\binom{n-1}{r-1}\binom{n-r}{2(k-r)}\, 2^{2(k-r)}}{\binom{2n-2}{2k-2}}.$$

두 분자는 인수 n을 제외하면 일치하므로

$$\begin{aligned}\frac{r\,\mathrm{P}(X_{n,k} = r)}{\mathrm{P}(X_{n-1,k-1} = r-1)} &= \frac{n\binom{2n-2}{2k-2}}{\binom{2n}{2k}} = n \cdot \frac{(2n-2)!}{(2k-2)!\,(2n-2k)!} \cdot \frac{(2k)!\,(2n-2k)!}{(2n)!} \\ &= n \cdot \frac{2k(2k-1)}{2n(2n-1)} = \frac{k(2k-1)}{2n-1}.\end{aligned}$$

(나누어도 괜찮았다. $n - r \ge 2k - r \ge 2(k-r)$이므로 여기 나오는 이항계수(binomial coefficient)는 모두 양수이다.) 양변에 분모를 곱하면 구하려던 결과를 얻는다.

생각해 보기 이 항등식이 $\mathrm{E}(X_{n,k})$를 구하는 데 어떻게 도움이 되는가?

기댓값(expectation)은 정확히 좌변들의 합이고, 우변들을 더하면 하나의 확률분포(probability distribution) 전체가 된다. 기댓값에서 $r = 0$인 항은 0이므로

$$\mathrm{E}(X_{n,k}) = \sum_{r=1}^{k} r\,\mathrm{P}(X_{n,k} = r) = \frac{k(2k-1)}{2n-1} \sum_{r=1}^{k} \mathrm{P}(X_{n-1,k-1} = r-1)$$
$$= \frac{k(2k-1)}{2n-1} \sum_{s=0}^{k-1} \mathrm{P}(X_{n-1,k-1} = s).$$

확률변수 $X_{n-1,k-1}$은 $2(k-1)$개의 양말 가운데 켤레의 수를 세므로 $0, 1, \ldots, k-1$의 값만 가지고, 따라서 마지막 합은 1이다. 그러므로

$$\mathrm{E}(X_{n,k}) = \frac{k(2k-1)}{2n-1}.$$

검산

- $k = 1$(양말 두 짝)이면 답은 $\frac{1}{2n-1}$이다. 어느 양말이 먼저 나오든, 남은 $2n-1$개의 양말 가운데 정확히 하나가 그 짝이다.
- 지시변수(indicator variable)를 쓰면 평균을 바로 얻는다. 켤레 i가 온전할 확률은 $\binom{2n-2}{2k-2}/\binom{2n}{2k} = \frac{2k(2k-1)}{2n(2n-1)}$이고, 켤레의 수 n을 곱하면 다시 $\frac{k(2k-1)}{2n-1}$이 된다.
- $n = 6, k = 3$일 때 소문항 (i)은 $\frac{64}{924} = \frac{16}{231}$을 주고 평균은 $\frac{15}{11}$이다. $n \le 8$인 모든 경우에 대해 컴퓨터로 모든 선택을 나열하여 (ii)의 분포, (iii)의 항등식, 그리고 평균을 확인하였다.

흔한 실수

- 한 켤레의 양말을 정말로 똑같은 것으로 보고 "색깔 패턴"을 세는 것. 이들은 똑같은 가능성을 갖지 않으므로, 이름표를 붙인 양말을 기준으로 세어야 한다.
- 한 짝만 나온 각 켤레에서 어느 짝을 가져오는지에 대한 인수 $2^{2(k-r)}$을 빠뜨리는 것.
- (iii)에서 $\sum_s \mathrm{P}(X_{n-1,k-1} = s) = 1$임을 알아차리지 못하고 그 합을 무작정 계산하려 드는 것.

STEP 너머

두 번째 검산은 지시변수를 이용한 기댓값의 선형성(linearity of expectation)으로, 분포를 전혀 모르고도 평균을 구한다. 이 방법은 사건 "켤레 i가 온전하다"들이 서로 독

립이 아닌데도 통한다. (iii)의 항등식은 크기 편향(size-biasing) 관계이다. 분포에 r로 가중치를 주면 더 작은 문제의 분포를 평행이동한 것이 나오는데, 이 기법으로 이항분포(binomial distribution)와 초기하분포(hypergeometric distribution)의 평균도 얻을 수 있다. Blitzstein과 Hwang의 *Introduction to Probability*(확률론 입문)는 이와 같은 지시변수와 이야기식 증명(story proof)을 책 전체에 걸쳐 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.