풀이

STEP III 2022, 문제 11

확률과 통계 • 이항분포의 평균편차

> **STEP III 2022, 11번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 OCR에서 내려받을 수 있는 2022년 STEP 시험지 묶음(zip 파일) 가운데 STEP III 시험지에 있다:
>
> www.ocr.org.uk/Images/696587-2022-past-papers.zip

핵심 아이디어

공정한 동전(fair coin)을 던지므로 X의 분포는 평균(mean)에 대하여 대칭(symmetric)이고, 따라서 평균보다 작은 쪽의 확률은 평균보다 큰 쪽의 확률과 같다. 이 대칭성 덕분에 절댓값(absolute value)이 들어간 합을 한쪽만의 합으로 바꿀 수 있다. 그다음 (i)(c)의 항등식(identity)이 $\sum r\binom{2n}{r}$를 더 작은 이항분포(binomial distribution)의 꼬리합(tail sum)으로 바꾸고, 이 꼬리합도 다시 대칭성으로 계산한다.

풀이

이 풀이 전체에서 $X \sim \mathrm{B}(N, \frac{1}{2})$이므로 $\mathrm{P}(X = r) = \binom{N}{r}2^{-N}$이고 $\mu = \frac{1}{2}N$이다. $\binom{N}{r} = \binom{N}{N-r}$이므로 $\mathrm{P}(X = r) = \mathrm{P}(X = N - r)$이다. 즉 분포는 $\frac{1}{2}N$에 대하여 대칭이다.

소문항 (i)(a)

$N = 2n$일 때 사상 $r \mapsto 2n - r$은 $\{0, \ldots, n-1\}$을 $\{n+1, \ldots, 2n\}$으로 보내므로, 대칭성에 의하여

$$\mathrm{P}(X \leq n-1) = \mathrm{P}(X \geq n+1).$$

세 사건 $X \leq n-1$, $X = n$, $X \geq n+1$이 모든 결과를 덮으므로 $2\,\mathrm{P}(X \leq n-1) + \mathrm{P}(X = n) = 1$이고, 이로부터

$$\mathrm{P}(X \leq n-1) = \tfrac{1}{2}(1 - \mathrm{P}(X = n)).$$

소문항 (i)(b)

이때 $\mu = n$이므로

$$\delta = \sum_{r=0}^{2n} |r-n| \binom{2n}{r} \frac{1}{2^{2n}}.$$

$r = n$인 항은 0이다. $r \geq n+1$인 항들에 $r = 2n - s$를 대입하면 $0 \leq s \leq n-1$에 대하여 $|r-n| = n-s$이고 $\binom{2n}{r} = \binom{2n}{s}$이다. 따라서 평균보다 큰 쪽의 항들은 작은 쪽의 항들을 정확히 되풀이한다.

$$\delta = 2\sum_{r=0}^{n-1} (n-r)\binom{2n}{r}\frac{1}{2^{2n}} = \sum_{r=0}^{n-1}(n-r)\binom{2n}{r}\frac{1}{2^{2n-1}},$$

이것이 보이려던 식이다.

소문항 (i)(c)

$1 \leq r \leq 2n$일 때

$$r\binom{2n}{r} = \frac{r\,(2n)!}{r!\,(2n-r)!} = 2n \cdot \frac{(2n-1)!}{(r-1)!\,(2n-r)!} = 2n\binom{2n-1}{r-1}.$$

이제 (b)의 합을 $2^{2n-1}\delta = A - B$로 나눈다. 여기서

$$A = n\sum_{r=0}^{n-1}\binom{2n}{r}, \qquad B = \sum_{r=1}^{n-1} r\binom{2n}{r}$$

이다(B의 $r=0$인 항은 0이므로 뺀다).

합 A. (a)에 의하여 $\sum_{r=0}^{n-1}\binom{2n}{r} = 2^{2n}\,\mathrm{P}(X \leq n-1) = 2^{2n-1} - \frac{1}{2}\binom{2n}{n}$이므로

$$A = n\,2^{2n-1} - \tfrac{n}{2}\binom{2n}{n}.$$

> **생각해 보기** s가 0부터 $n-2$까지일 때 $\binom{2n-1}{s}$를 어떻게 모두 더할까?

합 B. 위의 항등식과 $s = r-1$에 의하여

$$B = 2n\sum_{s=0}^{n-2}\binom{2n-1}{s}.$$

$0 \leq s \leq 2n-1$인 행 $\binom{2n-1}{s}$는 항의 개수가 짝수이고 합이 2^{2n-1}이다. 이 행은 $\{0,\ldots,n-1\}$과 $\{n,\ldots,2n-1\}$을 서로 바꾸는 $s \mapsto 2n-1-s$에 대하여 대칭이므로, 앞쪽 절반의 합은 2^{2n-2}이다. 우리의 합은 그 절반에서 한 항 모자란 곳에서 멈추므로

$$\sum_{s=0}^{n-2}\binom{2n-1}{s} = 2^{2n-2} - \binom{2n-1}{n-1} = 2^{2n-2} - \tfrac{1}{2}\binom{2n}{n},$$

이다. 여기서 $\binom{2n}{n} = \binom{2n-1}{n-1} + \binom{2n-1}{n} = 2\binom{2n-1}{n-1}$을 이용하였다. 따라서

$$B = n\,2^{2n-1} - n\binom{2n}{n}.$$

종합하기.

$$2^{2n-1}\delta = A - B = \left(n - \frac{n}{2}\right)\binom{2n}{n} = \frac{n}{2}\binom{2n}{n}, \qquad \text{따라서} \qquad \delta = \frac{n}{2^{2n}}\binom{2n}{n}.$$

소문항 (ii)

이제 $N = 2n+1$이고 $\mu = n + \frac{1}{2}$이며, 이는 X가 가질 수 있는 값이 아니다. 대칭 $r \mapsto 2n+1-r$은 $\{0,\dots,n\}$과 $\{n+1,\dots,2n+1\}$을 서로 바꾸고, $|r-\mu|$와 $\binom{2n+1}{r}$을 모두 보존한다. 따라서 (i)(b)에서와 같이

$$\delta = 2\sum_{r=0}^{n}\left(n+\tfrac{1}{2}-r\right)\binom{2n+1}{r}\frac{1}{2^{2n+1}} = \frac{1}{2^{2n+1}}\sum_{r=0}^{n}(2n+1-2r)\binom{2n+1}{r}.$$

- 같은 대칭성에 의하여 행 $\binom{2n+1}{r}$의 앞쪽 절반의 합은 2^{2n+1}의 절반이다. 즉 $\sum_{r=0}^{n}\binom{2n+1}{r} = 2^{2n}$이다.
- 항등식 $r\binom{2n+1}{r} = (2n+1)\binom{2n}{r-1}$((i)(c)와 똑같이 증명된다)과 (i)(a)에 의하여

$$\sum_{r=1}^{n} r\binom{2n+1}{r} = (2n+1)\sum_{s=0}^{n-1}\binom{2n}{s} = (2n+1)\left(2^{2n-1} - \tfrac{1}{2}\binom{2n}{n}\right).$$

그러므로

$$\sum_{r=0}^{n}(2n+1-2r)\binom{2n+1}{r} = (2n+1)2^{2n} - (2n+1)\left(2^{2n} - \binom{2n}{n}\right) = (2n+1)\binom{2n}{n},$$

이고, $N = 2n+1$일 때

$$\boxed{\delta = \frac{2n+1}{2^{2n+1}}\binom{2n}{n}}$$

이다.

검산

- $N = 3$ $(n=1)$: $|X - \frac{3}{2}|$는 확률 $\frac{1}{4}$로 $\frac{3}{2}$이고 확률 $\frac{3}{4}$로 $\frac{1}{2}$이므로 $\delta = \frac{3}{8} + \frac{3}{8} = \frac{3}{4}$이다. 공식은 $\frac{3\cdot 2}{8} = \frac{3}{4}$를 준다. $n=0$ $(N=1)$이면 공식은 $\frac{1}{2}$를 주는데, 이것도 옳다.
- $N = 4$ $(n=2)$: (i)(c)는 $\frac{2\cdot 6}{16} = \frac{3}{4}$를 주며, 이는 $N=3$일 때와 같다. 실제로 $N = 2n+1$일 때의 δ는 언제나 $N = 2n+2$일 때의 δ와 같다. $\binom{2n+2}{n+1} = \frac{2(2n+1)}{n+1}\binom{2n}{n}$이므로, 짝수 공식에서 n 대신 $n+1$을 넣으면 홀수 공식이 된다.

- 두 공식은 $N \leq 60$인 모든 N에 대하여 δ의 정의와 정확히 일치함을 컴퓨터로 확인하였다. 예를 들어 $N = 10$이면 $\delta = \frac{315}{256} \approx 1.230$이다.

흔한 실수

- 대칭성에서 나오는 인수 2를 잊는 것. (i)(b)에서 2^{2n}이 아니라 2^{2n-1}이 나오는 이유가 바로 이것이다.
- $s = r - 1$로 옮긴 뒤 s의 범위를 잘못 잡는 것. 합은 $n - 2$에서 멈추어 행의 절반에서 한 항이 모자라며, 그 빠진 항 $\binom{2n-1}{n-1}$ 덕분에 답이 0이 아니게 된다.
- (ii)에서 μ를 정수처럼 다루는 것. 이 경우 사라지는 항이 없으며, 행의 두 절반은 각각 $n + 1$개의 항을 가진다.

STEP 너머

스털링 공식(Stirling's formula)에 의하여 $\binom{2n}{n} \approx 4^n/\sqrt{\pi n}$이므로, N이 클 때 $\delta \approx \sqrt{N/(2\pi)}$이다. $\sigma = \frac{1}{2}\sqrt{N}$이므로 이는 $\delta \approx \sigma\sqrt{2/\pi}$이며, 중심극한정리(central limit theorem)가 예측하는 대로 정확히 정규분포(normal distribution)의 평균절대편차(mean absolute deviation)와 같다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 이항분포의 정규근사와 그 바탕에 있는 중심극한정리를 모두 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.