풀이

STEP III 2025, 문제 12

확률과 통계 • 재귀적으로 정의된 확률변수

> **STEP III 2025, 12번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 OCR에서 내려받을 수 있는 2025년 STEP 시험지 묶음(zip 파일) 가운데 STEP III 시험지에 있다:
>
> www.ocr.org.uk/Images/750486-2025-past-papers.zip

핵심 아이디어

소문항 (i)은 하나의 이중합(double sum)을 두 가지 방식으로 쓴 것에 불과하다. 양변 모두 같은 순서쌍 (r,s)들의 집합에 대하여 $\mathrm{f}(r)\,\mathrm{g}(s)$를 더한 것인데, 한쪽은 행별로, 다른 쪽은 열별로 센 것이다. 소문항 (ii)에서는 바로 앞의 값 X_{n-1}에 대하여 조건을 걸면(conditioning) 각 확률 $\mathrm{P}(X_n = r)$이 정확히 (i)의 좌변에 나타나는 꼴의 합으로 주어진다. 그런 다음 합의 순서를 바꾸면(swapping the order of summation) $\mathrm{E}(X_n)$을 $\mathrm{E}(X_{n-1})$로 알아볼 수 있는 꼴로 바꿀 수 있다.

풀이

소문항 (i)

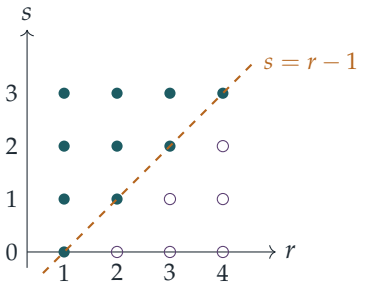
괄호를 전개하면 좌변은

$$\sum_{r=1}^{m+1}\ \sum_{s=r-1}^{m} \mathrm{f}(r)\,\mathrm{g}(s),$$

즉

$$1 \le r \le m+1 \quad \text{이고} \quad r-1 \le s \le m.$$

을 만족하는 모든 정수 순서쌍 (r,s)에 대하여 $\mathrm{f}(r)\,\mathrm{g}(s)$를 더한 합이다. 같은 순서쌍들의 집합을 이번에는 s를 먼저 고정하여 기술해 보자. 위의 조건에서 $0 \le s \le m$이어야 하고($s \ge r-1 \ge 0$이므로), 이러한 s가 주어졌을 때 가능한 r는 $1 \le r$이고 $r \le s+1$인 것들이다. 이때 $s+1 \le m+1$이므로 $r \le m+1$이라는 조건은 저절로 성립한다. 역으로 $0 \le s \le m$이고 $1 \le r \le s+1$인 순서쌍은 모두 원래 조건을 만족한다. 아래 그림은 $m=3$일 때 순서쌍들이 이루는 삼각형을 보여 준다.



열별로(r를 고정하여) 더하면 좌변이 되고, 행별로(s를 고정하여) 더하면

$$\sum_{s=0}^{m}\sum_{r=1}^{s+1} \mathrm{f}(r)\,\mathrm{g}(s) = \sum_{s=0}^{m}\Big(\mathrm{g}(s)\sum_{r=1}^{s+1}\mathrm{f}(r)\Big),$$

가 되는데, 이것이 바로 우변이므로 원하는 결과를 얻는다. 합이 유한하므로 순서를 바꾸는 것은 언제나 허용된다.

소문항 (ii)(a)

X_1은 0 또는 1일 확률이 같으므로 $\mathrm{E}(X_1) = \frac{1}{2}$이다.

X_2에 대해서는 X_1에 대하여 조건을 건다. $X_1 = 0$이면 X_2는 $\{0,1\}$ 위의 균등분포(uniform distribution)를 따르고, $X_1 = 1$이면 X_2는 $\{0,1,2\}$ 위의 균등분포를 따른다. 전확률 법칙(law of total probability)에 의하여

$$\begin{aligned}
\mathrm{P}(X_2 = 0) &= \tfrac{1}{2}\cdot\tfrac{1}{2} + \tfrac{1}{2}\cdot\tfrac{1}{3} = \tfrac{5}{12},\\
\mathrm{P}(X_2 = 1) &= \tfrac{1}{2}\cdot\tfrac{1}{2} + \tfrac{1}{2}\cdot\tfrac{1}{3} = \tfrac{5}{12},\\
\mathrm{P}(X_2 = 2) &= \tfrac{1}{2}\cdot 0 + \tfrac{1}{2}\cdot\tfrac{1}{3} = \tfrac{1}{6},
\end{aligned}$$

이므로 원하는 결과를 얻는다(세 값의 합이 1이라는 것도 유용한 검산이다). 따라서

$$\mathrm{E}(X_2) = 0\cdot\tfrac{5}{12} + 1\cdot\tfrac{5}{12} + 2\cdot\tfrac{1}{6} = \tfrac{3}{4}.$$

소문항 (ii)(b)

수학적 귀납법(induction)에 의하여 X_{n-1}은 $\{0,1,\ldots,n-1\}$의 값을 가진다(한 단계마다 가능한 최댓값이 하나씩 커진다). $X_{n-1} = s$가 주어지면 X_n은 $s+2$개의 값 $0,1,\ldots,s+1$ 위의 균등분포를 따르므로

$$\mathrm{P}(X_n = r \mid X_{n-1} = s) = \begin{cases} \dfrac{1}{s+2}, & 0 \le r \le s+1,\\[2mm] 0, & r > s+1. \end{cases}$$

이다. 값 $r = 0$은 언제나 가능하므로 전확률 법칙에 의하여

$$\mathrm{P}(X_n = 0) = \sum_{s=0}^{n-1} \frac{\mathrm{P}(X_{n-1} = s)}{s+2}.$$

이다. $1 \le r \le n$인 경우에는 $s+1 \ge r$, 즉 $s \ge r-1$인 s만 기여하므로

$$\mathrm{P}(X_n = r) = \sum_{s=r-1}^{n-1} \frac{\mathrm{P}(X_{n-1} = s)}{s+2}, \qquad r = 1, 2, \ldots, n.$$

이다. ($r = 1$일 때 이 합은 $r = 0$일 때의 합과 같으므로, 모든 $n \ge 1$에 대하여 $\mathrm{P}(X_n = 0) = \mathrm{P}(X_n = 1)$이며, 이는 (a)에서 나온 두 값 $\frac{5}{12}$와 들어맞는다.)

소문항 (ii)(c)

$p_s = \mathrm{P}(X_{n-1} = s)$라 하자. $r = 0$인 항은 기댓값(expectation)에 아무것도 기여하지 않으므로, (b)에 의하여

$$\mathrm{E}(X_n) = \sum_{r=1}^{n} r\,\mathrm{P}(X_n = r) = \sum_{r=1}^{n} \Big(r \sum_{s=r-1}^{n-1} \frac{p_s}{s+2} \Big).$$

이다.

? **생각해 보기** 소문항 (i)을 여기에 어떻게 적용할 수 있을까?

이것은 $m = n-1$($n \ge 1$이므로 허용된다), $\mathrm{f}(r) = r$, $\mathrm{g}(s) = p_s/(s+2)$로 둔 (i)의 좌변이다. 합의 순서를 바꾸면

$$\begin{aligned}
\mathrm{E}(X_n) &= \sum_{s=0}^{n-1} \Big(\frac{p_s}{s+2} \sum_{r=1}^{s+1} r \Big) = \sum_{s=0}^{n-1} \frac{p_s}{s+2} \cdot \frac{(s+1)(s+2)}{2} \\
&= \frac{1}{2} \sum_{s=0}^{n-1} (s+1)\, p_s = \frac{1}{2} \Big(\sum_{s=0}^{n-1} p_s + \sum_{s=0}^{n-1} s\, p_s \Big) \\
&= \tfrac{1}{2} \big(1 + \mathrm{E}(X_{n-1}) \big),
\end{aligned}$$

이다. 여기서 $\sum p_s = 1$을 이용하였다. 이것이 구하는 결과이다.

직관적으로 보면, $X_{n-1} = s$가 주어졌을 때 $\{0, \ldots, s+1\}$ 위의 균등분포를 따르는 확률변수의 평균은 $\frac{1}{2}(s+1)$이므로, 이 결과는 모습을 바꾼 탑 법칙(tower law)이다.

점화식(recurrence)을 풀기 위해 $e_n = \mathrm{E}(X_n)$이라 하자. 양변에서 1을 빼면

$$e_n - 1 = \tfrac{1}{2}\,(e_{n-1} - 1),$$

이므로 $e_n - 1$은 한 단계마다 절반이 된다. $e_0 = 0$이므로 $e_n - 1 = -\big(\frac{1}{2}\big)^n$, 즉

$$\mathrm{E}(X_n) = 1 - \frac{1}{2^n}, \qquad n = 1, 2, \ldots.$$

이다.

검산

이 공식에서 $\mathrm{E}(X_1) = \frac{1}{2}$, $\mathrm{E}(X_2) = \frac{3}{4}$를 얻으며, 이는 (a)와 일치한다. $n = 15$까지 X_n의 정확한 분포를 단계별로 계산해 보면 (b)의 식과 $\mathrm{E}(X_n) = 1 - 2^{-n}$이 모두 확인된다. 예를 들어 $\mathrm{E}(X_5) = \frac{31}{32} \approx 0.969$이고, 200 000회 시행한 컴퓨터 시뮬레이션에서는 0.967을 얻었다. $n \to \infty$일 때 평균은 1에 수렴(converge)하는데, 이는 그럴듯하다. 각 단계에서 평균이 1 쪽으로 남은 거리의 절반만큼 끌려가기 때문이다.

흔한 실수

- $\mathrm{P}(X_n = r)$의 합을 $s = r - 1$이 아니라 $s = r$부터 시작하는 실수. X_n은 $X_{n-1} + 1$까지 올라갈 수 있으므로 값 r는 $s = r - 1$에서도 도달할 수 있다.
- (c)에서 순서를 바꿀 때 합의 범위를 잘못 잡는 실수. 순서를 바꾸고 나면 r는 1부터 $s + 1$까지 움직이므로 $\frac{1}{2}(s+1)(s+2)$가 되고, 인수 $s + 2$가 깔끔하게 약분된다.
- e_1을 초깃값으로 하여 점화식을 풀면서 $1 - 2^{-(n-1)}$로 쓰는 실수. $e_1 = \frac{1}{2}$과 비교하여 확인하라.

STEP 너머

여기서의 과정은 마르코프 연쇄(Markov chain)이다. X_n의 분포는 X_{n-1}에만 의존하며, (b)는 확률벡터(probability vector)에 전이행렬(transition matrix)을 한 번 곱하는 단계에 해당한다. (c)의 지름길은 전기댓값 법칙(law of total expectation) $\mathrm{E}(X_n) = \mathrm{E}\big(\mathrm{E}(X_n \mid X_{n-1})\big)$으로, 이와 같은 많은 문제를 선형 점화식(linear recurrence)으로 바꾸어 준다. 두 아이디어는 모두 Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)에서 전개된다.

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.