풀이

STEP I 2011, 문제 13

확률과 통계 • 연속확률변수

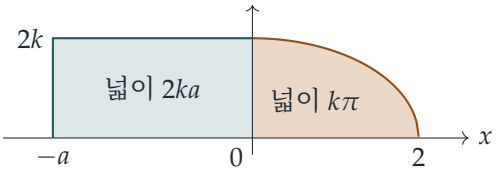
STEP I 2011, 13번 문제

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

step.maths.org/questions/11-s1-q13

핵심 아이디어

밀도함수(density)의 그래프를 그려 보자. 이 그래프는 $[-a, 0)$ 위의 높이 $2k$인 직사각형과 $[0, 2]$ 위의 반지름 2인 사분원(quarter-circle)을 세로로 k배 한 것을 이어 붙인 모양이다. 그래프 아래의 넓이가 곧 확률이므로, 적분을 하지 않아도 직사각형과 사분원만으로 대부분의 계산을 할 수 있다. 전체 넓이는 1이어야 하고, 이 조건에서 k가 정해진다. 그다음부터 d에 관한 모든 물음은 어떤 점의 오른쪽에 넓이가 얼마나 있는지를 묻는 문제로 귀결된다.



풀이

k 구하기

$[0, 2]$에서 $\sqrt{4 - x^2}$의 그래프는 원 $x^2 + y^2 = 4$의 4분의 1이므로

$$\int_0^2 \sqrt{4 - x^2}\,\mathrm{d}x = \tfrac{1}{4}\pi \cdot 2^2 = \pi.$$

(주어진 원시함수(antiderivative)로 계산해도 $2\arcsin 1 + 0 - 0 = \pi$로 일치한다.) 따라서 전체 확률은 $2ka + k\pi = 1$이고,

$$k = \frac{1}{2a + \pi}.$$

소문항 (i)

기댓값(expectation) $\mathrm{E}(X)$의 적분을 0에서 나눈다.

$$\int_{-a}^{0} 2kx\,\mathrm{d}x = k[x^2]_{-a}^{0} = -ka^2.$$

두 번째 부분에서는 $(4-x^2)^{3/2}$의 도함수(derivative)가 $-3x(4-x^2)^{1/2}$이므로

$$\int_{0}^{2} kx\sqrt{4-x^2}\,\mathrm{d}x = k\Big[-\tfrac{1}{3}(4-x^2)^{3/2}\Big]_{0}^{2} = k\cdot\tfrac{1}{3}\cdot 8 = \tfrac{8}{3}k.$$

두 부분을 더하면

$$\mathrm{E}(X) = k\Big(\frac{8}{3} - a^2\Big) = \frac{8-3a^2}{3(2a+\pi)}.$$

소문항 (ii)

0의 오른쪽에 있는 확률은 사분원의 넓이이다.

$$\mathrm{P}(X>0) = k\pi = \frac{\pi}{2a+\pi}.$$

생각해 보기 $\mathrm{P}(X>0)$를 알면 어떻게 d의 부호를 알 수 있을까?

밀도함수가 $(-a,2)$에서 양수이므로, 함수 $t \mapsto \mathrm{P}(X>t)$는 $-a \le t \le 2$에서 연속(continuous)이고 순감소(strictly decreasing)한다. 이 함수는 1에서 0까지 줄어들므로 $\mathrm{P}(X>d) = \frac{1}{10}$을 만족하는 d는 정확히 하나 있다. 함수가 감소하므로 $d<0$일 필요충분조건은 $\mathrm{P}(X>0) < \frac{1}{10}$이다. 다시 말해, 0의 오른쪽에 있는 확률이 10분의 1보다 작을 때 d는 음수이다.

이제 $2a > 9\pi$라고 하자. 그러면 $2a+\pi > 10\pi$이므로

$$\mathrm{P}(X>0) = \frac{\pi}{2a+\pi} < \frac{1}{10},$$

따라서 구하는 대로 $d<0$이다. (이 논증은 거꾸로도 성립한다. 즉 $d<0$이면 $2a>9\pi$이다.)

이 경우 $-a \le d < 0$이므로, d의 오른쪽 넓이는 직사각형 중 d부터 0까지의 폭 $-d$인 부분과 사분원 전체를 합한 것이다.

$$\mathrm{P}(X>d) = 2k(-d) + k\pi = \frac{\pi-2d}{2a+\pi}.$$

이것을 $\frac{1}{10}$과 같다고 놓으면 $10\pi - 20d = 2a+\pi$이므로

$$d = \frac{9\pi - 2a}{20}.$$

이 값은 실제로 $[-a,0)$에 속한다. $2a>9\pi$이므로 음수이고, $d \ge -a$는 $9\pi+18a \ge 0$과 동치인데 이는 참이다.

소문항 (iii)

이제 $d = \sqrt{2} > 0$이므로 d는 원 모양 부분 아래에 있고

$$\mathrm{P}(X > \sqrt{2}) = k \int_{\sqrt{2}}^{2} \sqrt{4 - x^2}\,\mathrm{d}x.$$

주어진 결과를 이용하면

$$\begin{aligned}\int_{\sqrt{2}}^{2} \sqrt{4 - x^2}\,\mathrm{d}x &= \left[2\arcsin(\tfrac{1}{2}x) + \tfrac{1}{2}x\sqrt{4 - x^2}\right]_{\sqrt{2}}^{2} \\ &= (2 \cdot \tfrac{\pi}{2} + 0) - (2 \cdot \tfrac{\pi}{4} + \tfrac{\sqrt{2}}{2} \cdot \sqrt{2}) \\ &= \frac{\pi}{2} - 1.\end{aligned}$$

(기하적으로 이것은 $x = \sqrt{2}$에서의 현(chord)이 잘라 낸 활꼴(circular segment)이다. 그 넓이는 원판의 8분의 1인 $\frac{\pi}{2}$에서 넓이가 1인 직각삼각형을 뺀 것과 같다.) 따라서

$$\frac{\frac{\pi}{2} - 1}{2a + \pi} = \frac{1}{10} \quad \Longrightarrow \quad 5\pi - 10 = 2a + \pi \quad \Longrightarrow \quad a = 2\pi - 5.$$

이 값은 양수이고($a \approx 1.283$), $2a \leq 9\pi$를 만족하므로 $d > 0$이라는 점에서 소문항 (ii)와도 일치한다.

검산

- 소문항 (i)에서 $a = 1$이면 평균은 $\frac{5}{3(2+\pi)} \approx 0.3242$이다. 표본 2×10^5개로 몬테카를로 시뮬레이션(Monte Carlo simulation)을 하면 0.3236이 나왔다.
- 소문항 (ii)의 공식은 경계인 $2a = 9\pi$에서 $d = 0$을 주므로 조건과 맞는다. $a = 20$이면 $d = \frac{9\pi - 40}{20} \approx -0.586$이고, $\mathrm{P}(X > d) = \frac{1}{10}$을 수치적으로 풀어도 같은 값이 나온다.
- 소문항 (iii)에서 $a = 2\pi - 5$일 때 심프슨 공식(Simpson's rule)으로 계산하면 $\mathrm{P}(X > \sqrt{2}) = 0.1000$이다.

흔한 실수

- k를 모르는 채로 두는 것. 정규화(normalisation) 조건 $2ka + k\pi = 1$이 숨겨진 첫 단계이며, 모든 답은 a만으로 나타내어진다.
- 소문항 (ii)에서 사분원을 빠뜨리고 $\mathrm{P}(X > d) = -2kd$라고만 쓰는 것. $[0, 2]$ 위의 확률도 모두 음수인 d의 오른쪽에 있다.
- $\int_{-a}^{0} 2kx\,\mathrm{d}x$의 부호를 틀리는 것. 구간 전체에서 $x < 0$이므로 이 값은 음수이다.

STEP 너머

값 d는 X의 상위 10% 점, 곧 0.9-분위수(quantile)이다. 이를 구하는 것은 누적분포함수(cumulative distribution function) F의 역함수를 구하는 것이며, 바로 이 역함수가 균등분포를 따르는 확률변수로부터 다른 확률변수를 시뮬레이션할 때 쓰인다(역변환 샘플링(inverse transform sampling)). 한쪽은 평평하고 다른 쪽은 휘어진 이 밀도함수의 혼합된 모양은 분위수를 왜 구간별로 나누어 구해야 하는 경우가 많은지 보여 준다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 분포함수와 분위수를 꼼꼼히 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.