풀이

STEP III 2006, 문제 13

확률과 통계 • 기하학적 확률

STEP III 2006, 13번 문제

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

step.maths.org/questions/06-s3-q13

핵심 아이디어

둘레(perimeter)는 성격이 매우 다른 두 조각, 곧 길이가 π인 호(arc)와 길이가 2인 지름(diameter)으로 이루어져 있으므로, 각 점이 어느 조각 위에 놓이는지를 조건으로 삼는다. 세 경우 각각에서 삼각형의 넓이는 꼭짓점 C를 중심에 둔 $\frac{1}{2}ab\sin C$라는 간단한 공식으로 주어지고, 조건부 기댓값(conditional expectation)은 짧은 적분으로 구해진다. 그런 다음 전체 기댓값의 법칙(law of total expectation)으로 조각들을 다시 합친다.

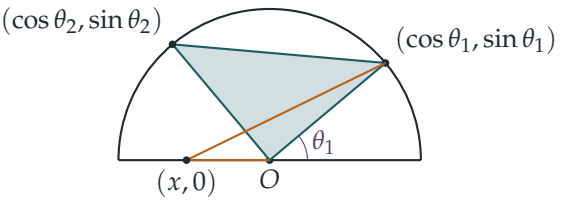
풀이

설정

지름의 중심 O를 원점에 두고, 지름은 x축을 따라 $(-1,0)$에서 $(1,0)$까지, 호는 위쪽 반평면에 놓는다. "둘레 위에서 무작위로"라는 말은 전체 길이가 $\pi+2$인 경계 전체를 따라 길이에 대해 균등하게 분포한다(uniformly distributed)는 뜻이다. 따라서 각 점은 독립적으로(independently)

- 확률 $\dfrac{\pi}{\pi+2}$로 호 위에 놓이며, 이때 점은 θ가 $[0,\pi]$에서 균등분포를 따르는 $(\cos\theta,\sin\theta)$이다.
- 확률 $\dfrac{2}{\pi+2}$로 지름 위에 놓이며, 이때 점은 x가 $[-1,1]$에서 균등분포를 따르는 $(x,0)$이다.

(단위원에서 호의 길이는 각과 같으므로, 호 위에서 길이에 대해 균등한 것은 각에 대해 균등한 것과 같다.)



두 점 가운데 지름 위에 놓인 점의 개수를 D라 하자. 그러면 $D = 2, 1, 0$일 확률은 각각

$$\frac{4}{(\pi+2)^2}, \qquad \frac{2\cdot 2\pi}{(\pi+2)^2} = \frac{4\pi}{(\pi+2)^2}, \qquad \frac{\pi^2}{(\pi+2)^2},$$

이다. 여기서 두 점의 독립성을 이용했다(가운데 경우의 인수 2는 어느 점이 지름 위에 있는지를 센 것이다). 모든 경우에 삼각형은 O를 꼭짓점으로 가지므로, 그 넓이는 $\frac{1}{2}|OP|\,|OQ|\sin\angle POQ$이다.

두 점이 모두 지름 위에 있는 경우

세 꼭짓점이 모두 한 직선 위에 있으므로 삼각형은 퇴화(degenerate)하고 $A = 0$이다. 따라서 $\mathrm{E}(A \mid D = 2) = 0$이다.

한 점은 호 위에, 한 점은 지름 위에 있는 경우

두 점을 $P = (\cos\theta, \sin\theta)$와 $R = (x, 0)$이라 하자. 그러면 $|OP| = 1$, $|OR| = |x|$이고, 밑변 OR는 지름 위에 놓이며 P는 그 위로 높이 $\sin\theta \geq 0$인 곳에 있으므로

$$A = \tfrac{1}{2}\,|x|\sin\theta.$$

이다. x와 θ는 독립이므로

$$\mathrm{E}(A \mid D = 1) = \tfrac{1}{2}\,\mathrm{E}|x|\,\mathrm{E}(\sin\theta) = \tfrac{1}{2}\cdot\frac{1}{2}\int_{-1}^{1}|x|\,\mathrm{d}x\cdot\frac{1}{\pi}\int_0^{\pi}\sin\theta\,\mathrm{d}\theta = \tfrac{1}{2}\cdot\tfrac{1}{2}\cdot\frac{2}{\pi} = \frac{1}{2\pi}.$$

두 점이 모두 호 위에 있는 경우

이제 $P = (\cos\theta_1, \sin\theta_1)$, $Q = (\cos\theta_2, \sin\theta_2)$이고, θ_1, θ_2는 서로 독립이며 $[0, \pi]$에서 균등분포를 따른다. 두 반지름의 길이는 모두 1이고 각 POQ는 $|\theta_1 - \theta_2|$이므로

$$A = \tfrac{1}{2}\,|\sin(\theta_1 - \theta_2)|.$$

이다.

? **생각해 보기** $\mathrm{E}|\sin(\theta_1 - \theta_2)|$를 구하려면 $\theta_1 - \theta_2$의 분포가 필요할까?

그렇지 않다. 먼저 θ_2를 조건으로 삼는다. θ_2를 고정하면, θ_1이 $[0,\pi]$ 위를 움직일 때 각 $\theta_1-\theta_2$는 길이가 π인 구간 위를 움직인다. 함수 $|\sin u|$는 주기(period)가 π이므로, 길이가 π인 어떤 구간 위에서 적분하더라도 $[0,\pi]$ 위에서 적분한 값과 같다.

$$\int_0^{\pi} |\sin(\theta_1-\theta_2)|\,\mathrm{d}\theta_1 = \int_0^{\pi} \sin u\,\mathrm{d}u = 2.$$

따라서 θ_2가 무엇이든 $\mathrm{E}\big(|\sin(\theta_1-\theta_2)| \,\big|\, \theta_2\big) = 2/\pi$이고, 그러므로

$$\mathrm{E}(A \mid D=0) = \tfrac{1}{2}\cdot\frac{2}{\pi} = \frac{1}{\pi}.$$

이다.

종합하기

전체 기댓값의 법칙에 의해

$$\begin{aligned}\mathrm{E}(A) &= \frac{4}{(\pi+2)^2}\cdot 0 + \frac{4\pi}{(\pi+2)^2}\cdot\frac{1}{2\pi} + \frac{\pi^2}{(\pi+2)^2}\cdot\frac{1}{\pi}\\ &= \frac{2+\pi}{(\pi+2)^2} = \frac{1}{2+\pi},\end{aligned}$$

이고, 이는 요구한 대로 $(2+\pi)^{-1}$이다.

검산

- 세 경우의 확률의 합 $\dfrac{4+4\pi+\pi^2}{(\pi+2)^2}$는 1이다.
- 각 조건부 기댓값은 그 경우에 가능한 최대 넓이 이하이다. $\frac{1}{\pi}\approx 0.318 \le \frac{1}{2}$(두 반지름이 직각을 이룰 때)이고 $\frac{1}{2\pi}\approx 0.159 \le \frac{1}{2}$이다.
- 점의 쌍 10^6개로 몬테카를로 시뮬레이션(Monte Carlo simulation)을 하면 $\mathrm{E}(A)\approx 0.1943$을 얻으며, 이는 $1/(2+\pi)\approx 0.1945$와 잘 맞는다. 수치적분(quadrature)으로도 조건부 값 $1/\pi$와 $1/(2\pi)$를 확인했다.

흔한 실수

- 각 점이 호 또는 지름 위에 놓일 확률을 각각 $\frac{1}{2}$로 잡는 것. "둘레 위에서 무작위로"는 각 조각에 그 길이만큼의 가중치를 준다.
- $\mathrm{P}(D=1)$에서 인수 2를 빠뜨리는 것. 두 점 중 어느 것이든 지름 위의 점이 될 수 있다.
- 절댓값 없이 $A=\frac{1}{2}\sin(\theta_1-\theta_2)$로 쓰는 것. 그러면 대칭성 때문에 기댓값이 0이 된다.

STEP 너머

이 문제는 **기하학적 확률**(geometric probability)의 작은 한 조각이며, 그 고전적인 예로는 뷔퐁의 바늘(Buffon's needle)과 베르트랑의 역설(Bertrand's paradox)이 있다. 베르트랑의 역설은 "무작위로"라는 말을 왜 분명하게 정해야 하는지를 보여 주는데, 이는 바로 우리가 둘레를 따른 길이를 측도로 택해서 한 일이다. 여기서 핵심 역할을 한 도구, 곧 분할(partition)을 조건으로 삼아 평균을 내는 것(전체 기댓값의 법칙)은 대학 1학년 확률론의 중심 내용이다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)이 이를 잘 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.