풀이

STEP I 1987, 문제 15

확률과 통계 • 연속확률변수

> **STEP I 1987, 15번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/87-s1-q15

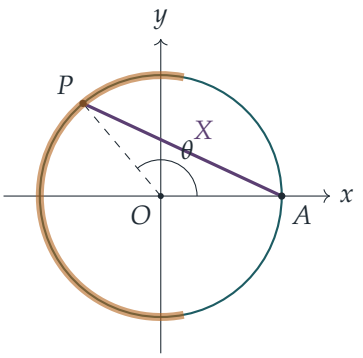
핵심 아이디어

"원 위에서 균등하다(uniform)"는 말은 P가 중심에서 이루는 각이 균등분포(uniform distribution)를 따른다는 뜻이다. 따라서 자연스러운 첫걸음은 θ가 $[0, 2\pi)$에서 균등분포를 따른다고 하고 $P = (\cos\theta, \sin\theta)$로 쓰는 것이다. 그러면 $(1, 0)$까지의 거리는 단위원의 현(chord)이 되고, 반각 공식(half-angle formula)을 쓰면 이것이 사인 하나로 정리된다. 그 다음부터는 평균(mean)은 적분 한 번으로 구해지고, 분산(variance)은 적분이 거의 필요 없으며, 확률은 각의 구간에 관한 문제가 된다.

풀이

설정하기

원점을 O라 하고, 양의 x축에서 OP까지의 각을 θ라 하자. 원 위에서 P를 균등하게 고른다는 것은 호의 길이(arc length)를 균등하게 고른다는 뜻이므로, θ는 $[0, 2\pi)$에서 확률밀도(density) $\frac{1}{2\pi}$인 균등분포를 따른다. $A = (1, 0)$, $P = (\cos\theta, \sin\theta)$로 놓는다.



그러면 $1-\cos\theta = 2\sin^2\frac{\theta}{2}$를 이용하여

$$X^2 = (\cos\theta - 1)^2 + \sin^2\theta = 2 - 2\cos\theta = 4\sin^2\tfrac{\theta}{2},$$

를 얻는다. $0 \le \frac{\theta}{2} < \pi$이므로 $\sin\frac{\theta}{2} \ge 0$이고, 따라서 양의 제곱근을 취해도 안전하다.

$$X = 2\sin\tfrac{\theta}{2}.$$

평균

$$\mathrm{E}(X) = \frac{1}{2\pi}\int_0^{2\pi} 2\sin\tfrac{\theta}{2}\,\mathrm{d}\theta = \frac{1}{2\pi}\Big[-4\cos\tfrac{\theta}{2}\Big]_0^{2\pi} = \frac{1}{2\pi}(4+4) = \frac{4}{\pi}.$$

따라서 $\mathrm{E}(X) = \frac{4}{\pi} \approx 1.273$이다.

분산

> **생각해 보기** $\mathrm{E}(X^2)$를 구하려면 $4\sin^2\frac{\theta}{2}$를 적분해야 할까?

그렇지 않다. 제곱을 풀어 쓴 꼴이 더 쉽다. $X^2 = 2 - 2\cos\theta$이고, $\int_0^{2\pi}\cos\theta\,\mathrm{d}\theta = 0$이므로 한 바퀴 전체에서 $\cos\theta$의 평균은 0이다. 따라서 $\mathrm{E}(X^2) = 2$이고,

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = 2 - \frac{16}{\pi^2} \approx 0.379.$$

X가 평균보다 클 확률

$2\sin\frac{\theta}{2} > \frac{4}{\pi}$, 즉 $\sin\frac{\theta}{2} > \frac{2}{\pi}$이어야 한다. $0 < \frac{2}{\pi} < 1$이므로 이 부등식은 해를 가진다. $\alpha = \arcsin\frac{2}{\pi}$라 하면 α는 $(0, \frac{\pi}{2})$에 있다. $\frac{\theta}{2}$가 $[0,\pi)$를 지나는 동안 사인값은 0에서 1까지 증가했다가 다시 감소하며, 사인값이 $\sin\alpha$보다 큰 것은 정확히

$$\alpha < \tfrac{\theta}{2} < \pi - \alpha, \qquad \text{즉} \qquad 2\alpha < \theta < 2\pi - 2\alpha.$$

일 때이다. 이것이 그림에서 굵게 표시한 호, 곧 원에서 A의 반대편에 있는 점들이다. θ가 균등분포를 따르므로, 구하는 확률은 이 구간의 길이를 2π로 나눈 값이다.

$$\mathrm{P}(X > \mathrm{E}(X)) = \frac{2\pi - 4\alpha}{2\pi} = \boxed{1 - \frac{2}{\pi}\arcsin\frac{2}{\pi}} \approx 0.561.$$

검산

- X의 범위는 $[0,2]$이고, 평균 $4/\pi \approx 1.27$은 중점 1보다 크다. 이는 그림과 잘 맞는다. 원의 상당 부분에서 현의 길이가 최댓값에 가깝고(P가 반대편을 지날 때 X는 천천히 변한다), 따라서 분포가 큰 값 쪽으로 치우쳐 있다.
- 같은 이유로 평균보다 클 확률은 $\frac{1}{2}$보다 조금 커야 하는데, 실제로 0.561이 그러하다.
- 수치적으로, 등간격으로 놓인 2×10^5개의 각에 대해 중점법(midpoint rule)으로 평균을 내면 평균 1.273240, 분산 0.378861, $\mathrm{P} = 0.56066$을 얻으며, 10^6개의 점으로 한 모의실험(simulation)도 표본오차 범위 안에서 일치한다.

흔한 실수

- $\theta \in (-\pi, \pi]$로 잡고 $X = 2\sin\frac{\theta}{2}$로 두는 것. 이 경우 $\sin\frac{\theta}{2}$가 음수일 수 있으므로 $X = 2\left|\sin\frac{\theta}{2}\right|$로 써야 한다. $[0, 2\pi)$를 쓰면 절댓값을 피할 수 있다.
- P의 x좌표가 $[-1,1]$에서 균등분포를 따른다고 보는 것. 원 위에서 균등하다는 것은 x에 대해서가 아니라 각(호의 길이)에 대해 균등하다는 뜻이다.

STEP 너머

이 문제는 확률변수의 변환(transformation of a random variable)의 첫 번째 예이다. X의 확률밀도함수는 $[0,2]$에서 $f(x) = \dfrac{2}{\pi\sqrt{4-x^2}}$로, 아크사인형 분포(arcsine-type law)이며, 위의 모든 답을 이것으로부터 다시 구할 수 있다. 이 문제는 또한 "무작위로"라는 말이 얼마나 미끄러운지도 암시한다. 원의 현을 무작위로 고르는 자연스러운 방법이 여러 가지 있는데, 방법에 따라 서로 다른 답이 나온다(베르트랑의 역설(Bertrand's paradox)). 두 아이디어 모두 Blitzstein과 Hwang의 *Introduction to Probability*(확률론 입문)에서 자세히 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.