풀이

STEP II 1990, 문제 15

확률과 통계 • 순서통계량

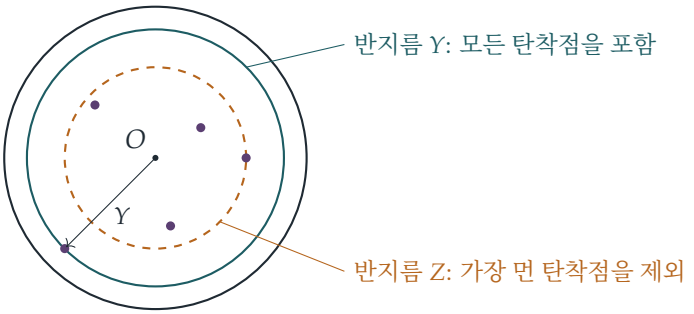
STEP II 1990, 15번 문제

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

step.maths.org/questions/90-s2-q15

핵심 아이디어

탄착점들을 모두 둘러싸는 O 중심의 가장 작은 원의 반지름은 **가장 먼** 탄착점까지의 거리와 같으므로, Y는 서로 독립(independent)인 거리들의 최댓값이다. 최댓값이 y 이하인 것은 **모든** 탄착점이 거리 y 이내에 있을 때와 정확히 같으므로, 누적분포함수(cumulative distribution function)는 곱으로 바뀐다. 가장 먼 탄착점을 제외하고 나면 구하려는 반지름은 **두 번째로** 큰 거리이며, 그 분포함수는 주어진 원 바깥에 놓일 수 있는 탄착점이 몇 개인지를 세어서 얻는다.



풀이

n개의 탄착점이 O로부터 떨어진 거리를 $R_1, \ldots, R_n$이라 하자. 사격은 독립적으로 이루어지므로 이들은 서로 독립이고, 각각의 분포함수는 다음과 같다.

$$P(R_i \le t) = t^2, \qquad 0 \le t \le 1.$$

Y의 분포

모든 탄착점을 포함하는 O 중심의 가장 작은 원의 반지름은 $Y = \max(R_1, \ldots, R_n)$이다. $0 \le y \le 1$에 대하여

$$P(Y \le y) = P(R_1 \le y, \ldots, R_n \le y) = \prod_{i=1}^{n} P(R_i \le y) = (y^2)^n = y^{2n},$$

이며, 가운데 단계에서 독립성을 이용하였다. 이를 미분하면 Y의 확률밀도함수(probability density function)는

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} y^{2n} = 2ny^{2n-1}, \qquad 0 \le y \le 1,$$

이고, 이것이 구하려던 결과이다. ($[0,1]$ 바깥에서는 0이며, 적분하면 당연히 1이 된다.)

원의 넓이는 πY^2이므로

$$E(\pi Y^2) = \pi \int_0^1 y^2 \cdot 2ny^{2n-1}\,\mathrm{d}y = 2n\pi \int_0^1 y^{2n+1}\,\mathrm{d}y = \frac{2n\pi}{2n+2} = \frac{n\pi}{n+1}.$$

가장 먼 탄착점 제외하기

탄착점이 남아 있으려면 $n \ge 2$이어야 한다. 가장 먼 탄착점을 제거하면, 나머지를 둘러싸는 O 중심의 가장 작은 원의 반지름은 $R_1, \ldots, R_n$ 중 두 번째로 큰 값인 Z이다.

> **생각해 보기** 사건 $Z \le z$를 반지름 z인 원 바깥에 떨어지는 탄착점의 개수로 어떻게 나타낼 수 있을까?

두 번째로 큰 거리가 z 이하인 것은 O로부터 z보다 멀리 떨어진 탄착점이 **많아야 하나**일 때와 정확히 같다. 각 탄착점은 독립적으로 확률 z^2으로 거리 z 이내에 놓이므로, 반지름 z인 원 바깥에 있는 탄착점의 개수는 이항분포(binomial distribution) $\mathrm{B}(n, 1-z^2)$를 따른다. 따라서 $0 \le z \le 1$에 대하여

$$\begin{aligned} P(Z \le z) &= P(\text{바깥에 탄착점이 없음}) + P(\text{바깥에 탄착점이 정확히 하나}) \\ &= z^{2n} + n(1-z^2)z^{2(n-1)} \\ &= nz^{2n-2} - (n-1)z^{2n}. \end{aligned}$$

이를 미분하면 Z의 확률밀도함수를 얻는다.

$$f_Z(z) = 2n(n-1)z^{2n-3} - 2n(n-1)z^{2n-1} = 2n(n-1)z^{2n-3}(1-z^2), \qquad 0 \le z \le 1.$$

그러므로 새 원의 넓이의 기댓값은

$$\begin{aligned} E(\pi Z^2) &= 2n(n-1)\pi \int_0^1 \left(z^{2n-1} - z^{2n+1}\right) \mathrm{d}z \\ &= 2n(n-1)\pi \left(\frac{1}{2n} - \frac{1}{2n+2}\right) \\ &= 2n(n-1)\pi \cdot \frac{2}{2n(2n+2)} = \left(\frac{n-1}{n+1}\right)\pi, \end{aligned}$$

이고, 이것이 구하려던 결과이다.

검산

- $n = 1$이면 첫 번째 답은 $\pi/2$이다. 탄착점이 하나일 때 $P(R^2 \le u) = u$이므로 R^2은 $[0,1]$에서 균등분포(uniform distribution)를 따르고, 따라서 넓이 πR^2의 기댓값은 $\pi/2$이다.
- $n = 2$이면 두 번째 답은 $\pi/3$으로, 두 탄착점 중 **더 가까운** 것에 대한 넓이의 기댓값이다. 첫 번째 답 $2\pi/3$과 함께 평균을 내면 탄착점이 하나일 때의 값인 $\pi/2$가 되며, 반드시 그래야 한다.
- $n = 5$로 2×10^5회 시행한 시뮬레이션에서 넓이의 기댓값은 2.618과 2.095로 나왔으며, 이론값은 $5\pi/6 \approx 2.618$과 $4\pi/6 \approx 2.094$이다.

흔한 실수

- 남은 탄착점들이 마치 $n-1$개짜리 새로운 표본인 것처럼 $P(Z \le z) = z^{2(n-1)}$이라고 쓰는 것. 제외된 탄착점은 무작위로 고른 것이 아니므로, 남은 탄착점들은 O 쪽으로 치우쳐 있다.
- 원 바깥에 있는 탄착점이 어느 것인지에 해당하는 인수 n을 빠뜨리는 것.
- $E(Y)$를 구한 뒤 제곱하는 것. 넓이의 기댓값은 $\pi E(Y^2)$이지 $\pi\big(E(Y)\big)^2$이 아니다.

STEP 너머

이것들은 **순서통계량**(*order statistics*)이다. $U_i = R_i^2$이 $[0,1]$에서 균등분포를 따르므로 넓이들은 바로 n개의 균등분포 확률변수의 순서통계량이며, n개의 독립인 균등분포 확률변수 중 k번째로 작은 값은 평균이 $k/(n+1)$인 베타분포(beta distribution)를 따른다. 여기서 구한 두 답은 각각 $k = n$과 $k = n-1$인 경우이다. 깔끔하게 그려 보면, n개의 무작위 점이 $[0,1]$을 기대 길이가 모두 같은 $n+1$개의 구간으로 나눈다는 것이다. Grimmett and Welsh, *Probability: An Introduction*(확률론 입문서)에서 이런 종류의 변환과 순서통계량을 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.