풀이

STEP III 2007, 문제 14

확률과 통계 • 기하학적 확률

> **STEP III 2007, 14번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/07-s3-q14

핵심 아이디어

모든 구멍을 둘러싸는 가장 작은 중심원(또는 정사각형)은 **가장 먼** 구멍에 의해 정해진다. 따라서 구해야 하는 것은 최댓값(maximum)의 분포이며, $P(\max \le r)$는 한 번 던졌을 때의 확률(probability)을 n제곱한 것에 지나지 않는다. 원판에서는 한 구멍이 O로부터 떨어진 거리의 제곱 R^2이 $[0,1]$에서 균등분포(uniform distribution)를 따른다는 사실이 드러나고, 덕분에 모든 계산이 깔끔해진다. 정사각형에서는 알맞은 "거리"가 $\max(|x|,|y|)$이며, 이것은 원판에서의 반지름과 정확히 똑같이 행동한다. 이 우연의 일치 덕분에 소문항 (iii)도 거의 공짜로 해결된다.

풀이

소문항 (i)

n개의 구멍이 O로부터 각각 거리 $R_1, \dots, R_n$에 있다고 하자. 던지기가 서로 독립이므로 이 값들도 독립(independent)이다. 한 번 던질 때 $0 \le r \le 1$이면, 구멍이 O로부터 거리 r 이내에 있을 필요충분조건은 반지름 r인 원판 안에 떨어지는 것이므로

$$P(R_i \le r) = \frac{\pi r^2}{\pi \cdot 1^2} = r^2.$$

$U_i = R_i^2$이라 하자. 그러면 $0 \le u \le 1$에 대하여 $P(U_i \le u) = P(R_i \le \sqrt{u}) = u$이므로 각 U_i는 $[0,1]$**에서 균등분포를 따르고**, U_i들은 서로 독립이다.

O를 중심으로 하고 모든 구멍을 둘러싸는 가장 작은 원의 반지름은 $M = \max_i R_i$이고 넓이는 $\pi M^2 = \pi \max_i U_i$이다. $V = \max_i U_i$라 쓰자. 독립성에 의하여

$$P(V \le v) = P(U_1 \le v) \cdots P(U_n \le v) = v^n \qquad (0 \le v \le 1),$$

이므로 V의 확률밀도함수(probability density function)는 nv^{n-1}이고

$$E(V) = \int_0^1 v \cdot nv^{n-1}\,\mathrm{d}v = \frac{n}{n+1}.$$

따라서 n개의 구멍을 모두 둘러싸는 가장 작은 원의 넓이의 기댓값(expected value)은

$$\boxed{\frac{n\pi}{n+1}}.$$

O에 가장 가까운 $n-1$개의 구멍을 둘러싸는 원의 반지름은 **두 번째로 큰** 거리이므로, 그 넓이는 πW이다. 여기서 W는 $U_1, \ldots, U_n$ 가운데 두 번째로 큰 값이다.

? 생각해 보기 두 번째로 큰 값의 분포는 어떻게 구할까?

$W \le w$일 필요충분조건은 U_i 가운데 w를 넘는 것이 많아야 하나라는 것이다. 하나도 넘지 않거나(확률 w^n) 정확히 하나가 넘는다. 후자의 경우 어느 것이 넘는지 고르는 방법이 n가지이므로 확률은 $n(1-w)w^{n-1}$이다. 따라서

$$P(W \le w) = w^n + n(1-w)w^{n-1} = nw^{n-1} - (n-1)w^n.$$

미분하는 대신 $[0,1]$에 값을 갖는 모든 확률변수(random variable)에 대해 성립하는 $E(W) = \int_0^1 P(W > w)\,\mathrm{d}w$를 이용할 수 있다.

$$E(W) = \int_0^1 (1 - nw^{n-1} + (n-1)w^n)\,\mathrm{d}w = 1 - 1 + \frac{n-1}{n+1} = \frac{n-1}{n+1}.$$

따라서 가장 가까운 $n-1$개의 구멍을 둘러싸는 원의 넓이의 기댓값은

$$\boxed{\frac{(n-1)\pi}{n+1}}.$$

(이 결과는 보기 좋은 규칙에 들어맞는다. 서로 독립인 n개의 균등분포 점은 $[0,1]$을 기댓값이 모두 $\frac{1}{n+1}$인 $n+1$개의 간격으로 나누므로, k번째로 작은 값의 평균은 $\frac{k}{n+1}$이다.)

소문항 (ii)

Q를 지나고 판의 변에 평행한 좌표축을 잡으면 판은 $-1 \le x, y \le 1$이 된다. 둘러싸는 정사각형의 변도 판의 변에 평행하다고 하자. Q를 중심으로 하고 한 변의 절반이 s인 정사각형은 $\{|x| \le s,\ |y| \le s\}$이므로, 이 정사각형이 구멍 (x, y)를 포함할 필요충분조건은 $\max(|x|, |y|) \le s$이다.

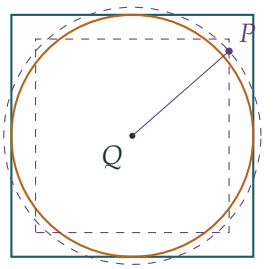
i번째 구멍에 대하여 $S_i = \max(|x_i|, |y_i|)$라 하자. $0 \le s \le 1$이면

$$P(S_i \le s) = \frac{(2s)^2}{2^2} = s^2,$$

이고, 이는 소문항 (i)의 $P(R_i \le r) = r^2$과 같다. 둘러싸는 가장 작은 정사각형은 한 변의 절반이 $\max_i S_i$이고 넓이는 $(2\max_i S_i)^2 = 4\max_i S_i^2$이다. 소문항 (i)와 똑같이 S_i^2들은 서로 독립이고 $[0,1]$에서 균등분포를 따르므로, 그 최댓값의 기댓값은 $\frac{n}{n+1}$이다. 둘러싸는 가장 작은 정사각형의 넓이의 기댓값은

$$\boxed{\frac{4n}{n+1}}.$$

소문항 (iii)



이제 판은 정사각형이고, Q를 중심으로 하여 모든 구멍을 둘러싸는 가장 작은 원을 구하려 한다. 구멍 $P = (x, y)$는 Q로부터 거리 $\sqrt{x^2+y^2}$에 있고

$$\sqrt{x^2+y^2} \;\ge\; \max(|x|, |y|) = S,$$

이며, 등호는 P가 좌표축 위에 있을 때만 성립한다. (그림에서 P를 지나는 점선 원은 P를 지나는 점선 정사각형의 내접원보다 결코 작지 않다.) 따라서 D가 구멍들이 Q로부터 떨어진 거리 가운데 가장 큰 값이면 $D \ge \max_i S_i$이고, 둘러싸는 원의 넓이는

$$\pi D^2 \;\ge\; \pi \max_i S_i^2.$$

을 만족한다. 소문항 (ii)에서 $\max_i S_i^2$은 원형 판에서의 $\max_i R_i^2$과 정확히 같은 분포를 따르므로

$$E(\pi D^2) \;\ge\; \pi\, E(\max_i S_i^2) = \frac{n\pi}{n+1},$$

이고, 이것은 소문항 (i)에서 구한 원형 판의 답이다. 이 부등식(inequality)은 등호가 성립하지 않는다. 양의 확률로 어떤 구멍도 좌표축 위에 있지 않으며, 그때는 $D > \max_i S_i$이기 때문이다. 따라서 정사각형 판에서의 원의 넓이의 기댓값이 **더 크다**.

직관적으로 설명하면, 정사각형 판은 원형 판을 포함하고(원판은 정사각형의 내접원이다) 다트가 모서리 쪽까지 퍼지므로, 가장 먼 구멍이 더 멀리 있는 경향이 있다.

검산

- $n = 1$이면 공식은 $\frac{\pi}{2}$와 2를 준다. 균등분포를 따르는 R^2의 평균은 $\frac{1}{2}$이므로 맞다. $n \to \infty$이면 넓이는 각각 π와 4, 즉 판 전체의 넓이로 다가가는데, 당연히 그래야 한다.
- $n = 2, 3, 5$에 대하여 2×10^5회 시행한 몬테카를로 시뮬레이션(Monte Carlo simulation)은 $\frac{n\pi}{n+1}$, $\frac{(n-1)\pi}{n+1}$, $\frac{4n}{n+1}$과 유효숫자 세 자리 정도까지 일치했다. 예를 들어 $n = 3$일 때 $\frac{3\pi}{4} \approx 2.3562$에 대하여 2.3561을 얻었다.
- 소문항 (iii)에 대해서는 정확한 분포를 수치 적분(quadrature)한 결과, $n = 3$일 때 넓이의 기댓값이 원판에서는 2.356인 데 비해 정사각형 판에서는 약 3.257이었다. 주장한 대로 더 크다.

흔한 실수

- $E(M)$을 구한 다음 제곱하는 것. 넓이의 기댓값은 $\pi\big(E(M)\big)^2$이 아니라 $\pi E(M^2)$이다.
- 중심으로부터의 거리가 $[0, 1]$에서 균등분포를 따른다고 여기는 것. 넓이에 비례하여 맞는다면 균등분포를 따르는 것은 R이 아니라 R^2이다. 바깥쪽 고리가 더 넓기 때문이다.
- 두 번째로 큰 값을 구할 때, 어느 구멍이 원 밖에 있는지에 해당하는 인수 n을 빠뜨리는 것.

STEP 너머

표본의 최댓값과 두 번째로 큰 값은 **순서통계량**(order statistics)이다. $[0, 1]$에서 균등분포를 따르는 서로 독립인 n개의 확률변수에 대하여 k번째로 작은 값은 평균이 $\frac{k}{n+1}$인 베타 분포(Beta distribution)를 따른다. 소문항 (iii)은 **확률적 지배**(stochastic dominance)(결합(coupling))의 예이다. 한 확률변수가 다른 확률변수보다 항상 크거나 같으면 그 기댓값도 그러하므로, 자세한 계산 없이 비교할 수 있다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문), 그리고 Blitzstein과 Hwang의 *Introduction to Probability*(확률론 입문)는 모두 순서통계량과 확률변수의 변환을 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.