풀이

STEP II 2012, 문제 13

확률과 통계 • 푸아송 과정과 최근접 거리

> **STEP II 2012, 13번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/12-s2-q13

이 문제는 *Beyond STEP*의 12.19 절 **평면 위의 무작위 점**에서 단계별 안내와 함께 자세히 다룬다: https://beyondstep.keithmcnulty.org/ko/ch12/spatial-poisson.html

핵심 아이디어

"가장 가까운 슈퍼마켓까지의 거리가 y 이상이다"라는 사건은 "우리가 고른 점을 중심으로 하는 반지름 y인 원 안에 슈퍼마켓이 하나도 없다"라는 사건과 정확히 같고, 푸아송(Poisson) 모형은 이 사건의 확률을 바로 알려 준다. 이로부터 Y의 누적분포함수(cumulative distribution function)를 얻고, 이를 미분(differentiation)하면 확률밀도함수(density)를 얻는다. 그러면 적률(moments)은 가우스(Gauss)형 적분이 된다. 부분적분(integration by parts)을 한 번 하면 $\mathrm{E}(Y)$가 가정해도 되는 적분으로 바뀌고, 치환(substitution) $u = y^2$을 하면 $\mathrm{E}(Y^2)$는 초등적으로 계산된다.

풀이

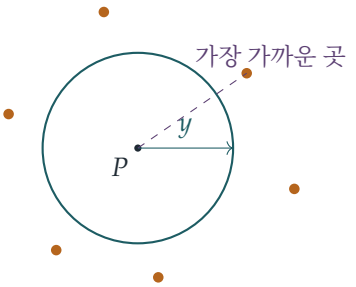
전체에 걸쳐 영역이 충분히 넓어서 경계 효과는 무시할 수 있다고 본다. 따라서 고른 점을 중심으로 하는 반지름 y인 원은 우리가 다루는 모든 y에 대하여 영역 안에 놓인다.

원 안에 슈퍼마켓이 없을 확률

반지름 y인 원 안에 있는 슈퍼마켓의 개수를 N이라 하자. 이 원의 넓이는 πy^2이므로 $N \sim \mathrm{Po}(\pi k y^2)$이고

$$\mathrm{P}(N = 0) = \mathrm{e}^{-\pi k y^2}.$$

Y의 분포



무작위로 고른 점을 P라 하자. 가장 가까운 슈퍼마켓이 P로부터 거리 y 이상 떨어져 있는 것은 P를 중심으로 하는 반지름 y인 열린 원 안에 슈퍼마켓이 하나도 없을 때이고, 오직 그때뿐이다. 따라서 $y \geq 0$에 대하여

$$\mathrm{P}(Y \geq y) = \mathrm{e}^{-\pi k y^2}, \qquad \text{따라서} \qquad \mathrm{P}(Y < y) = 1 - \mathrm{e}^{-\pi k y^2}.$$

(Y는 연속확률변수이므로 $<$로 쓰든 $\leq$로 쓰든 차이가 없다.) 이것이 Y의 누적분포함수 $F(y)$이고, 거리는 음수가 될 수 없으므로 $y < 0$에서는 $F(y) = 0$이다. $y \geq 0$에서 미분하면

$$f(y) = F'(y) = \frac{\mathrm{d}}{\mathrm{d}y}\left(1 - \mathrm{e}^{-\pi k y^2}\right) = 2\pi k y\, \mathrm{e}^{-\pi k y^2},$$

이고, 이것이 구하는 확률밀도함수이다. 간단히 확인해 보면, $y \to \infty$일 때 $F(y) \to 1$이므로 f를 적분하면 1이 된다.

평균

$$\mathrm{E}(Y) = \int_0^\infty y \cdot 2\pi k y\, \mathrm{e}^{-\pi k y^2}\, \mathrm{d}y.$$

생각해 보기 지수함수 앞에 y^2이 붙어 있을 때 주어진 적분을 어떻게 이용할 수 있을까?

피적분함수를 $y \times \left(2\pi k y\, \mathrm{e}^{-\pi k y^2}\right)$로 나누어 보자. 두 번째 인수는 $-\mathrm{e}^{-\pi k y^2}$의 도함수이다. $u = y$로 두고 부분적분하면

$$\mathrm{E}(Y) = \left[-y\, \mathrm{e}^{-\pi k y^2}\right]_0^\infty + \int_0^\infty \mathrm{e}^{-\pi k y^2}\, \mathrm{d}y = \int_0^\infty \mathrm{e}^{-\pi k y^2}\, \mathrm{d}y,$$

이다. $y \to \infty$일 때 $y\, \mathrm{e}^{-\pi k y^2} \to 0$이고(지수함수가 이긴다) $y = 0$에서도 0이기 때문이다. 이제 지수를 $-\frac{1}{2}x^2$에 맞춘다. $\pi k y^2 = \frac{1}{2}x^2$, 즉 $y = x/\sqrt{2\pi k}$, $\mathrm{d}y = \mathrm{d}x/\sqrt{2\pi k}$로 두자. 적분 구간은 그대로 0에서 ∞까지이므로

$$\mathrm{E}(Y) = \frac{1}{\sqrt{2\pi k}} \int_0^\infty \mathrm{e}^{-x^2/2}\, \mathrm{d}x = \frac{1}{\sqrt{2\pi k}} \sqrt{\frac{\pi}{2}} = \frac{1}{2\sqrt{k}}.$$

분산

$\mathrm{E}(Y^2)$를 구할 때는 주어진 적분이 전혀 필요 없다. y의 거듭제곱이 홀수이므로 $u = \pi k y^2$, $\mathrm{d}u = 2\pi k y\,\mathrm{d}y$로 치환할 수 있기 때문이다.

$$\mathrm{E}(Y^2) = \int_0^\infty y^2 \cdot 2\pi k y\, \mathrm{e}^{-\pi k y^2}\,\mathrm{d}y = \int_0^\infty \frac{u}{\pi k}\,\mathrm{e}^{-u}\,\mathrm{d}u = \frac{1}{\pi k}\int_0^\infty u\,\mathrm{e}^{-u}\,\mathrm{d}u = \frac{1}{\pi k},$$

여기서 $\int_0^\infty u\,\mathrm{e}^{-u}\,\mathrm{d}u = 1$을 이용했다(부분적분을 하거나, 모수가 1인 지수분포(exponential distribution)의 평균임을 알아보면 된다). 따라서

$$\mathrm{Var}(Y) = \mathrm{E}(Y^2) - \big(\mathrm{E}(Y)\big)^2 = \frac{1}{\pi k} - \frac{1}{4k} = \frac{4-\pi}{4\pi k},$$

이므로 원하는 결과를 얻는다.

검산

- **척도 변환.** 슈퍼마켓의 밀도가 k배가 되면 길이는 $\sqrt{k}$배만큼 줄어들어야 하므로 $\mathrm{E}(Y) \propto k^{-1/2}$이고 $\mathrm{Var}(Y) \propto k^{-1}$이어야 한다. 두 답 모두 이 꼴이다.
- **양수 여부.** $4-\pi > 0$이므로 분산은 마땅히 그래야 하듯 양수이다.
- **수치 계산.** $k = 1$일 때 확률밀도함수에 중점 구적법을 적용하면 $\mathrm{E}(Y) = 0.5000$, $\mathrm{Var}(Y) = 0.06831$을 얻고, 이는 $(4-\pi)/(4\pi) \approx 0.06831$과 일치한다. 한 점 주위에 슈퍼마켓을 푸아송 방식으로 무작위로 흩뿌리는 시행을 10^5번 시뮬레이션한 결과 평균은 0.4996, 분산은 0.0681이었다.

흔한 실수

- $\mathrm{P}(Y < y) = \mathrm{e}^{-\pi k y^2}$으로 쓰는 실수. 이것은 거리 y 안에 슈퍼마켓이 **하나도 없을** 확률, 즉 $\mathrm{P}(Y \geq y)$이다.
- $x = \sqrt{2\pi k}\,y$ 대신 $x = \sqrt{\pi k}\,y$로 치환하는 실수. 이렇게 하면 가정한 $\mathrm{e}^{-x^2/2}$가 아니라 e^{-x^2}이 나오고, 답이 $\sqrt{2}$배만큼 어긋난다.

STEP 너머

여기서 슈퍼마켓들은 평면 위의 **푸아송 점과정**(*Poisson point process*)을 이룬다. 이는 숲속의 나무, 하늘의 별, 휴대전화 기지국 등에 쓰이는 "완전한 공간적 무작위성"의 기본 모형이다. Y의 분포는 **레일리 분포**(*Rayleigh distribution*)로, 두 성분이 서로 독립인 정규확률변수인 벡터의 길이로도 나타난다. 또한 우리의 부분적분이 드러낸 깔끔한 꼬리 공식

$\mathrm{E}(Y) = \int_0^\infty \mathrm{P}(Y \geq y)\,\mathrm{d}y$는 음이 아닌 모든 확률변수에 대하여 성립한다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 푸아송 과정과 이러한 꼬리 적분 아이디어를 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.