풀이

STEP III 1995, 문제 12

확률과 통계 • 방사 대칭 확률밀도

STEP III 1995, 12번 문제

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

step.maths.org/questions/95-s3-q12

이 문제는 *Beyond STEP*의 12.22 절 **평균이 없는 분포**에서 단계별 안내와 함께 자세히 다룬다:
https://beyondstep.keithmcnulty.org/ko/ch12/cauchy.html

핵심 아이디어

방사 대칭(radially symmetrical)인 밀도함수(density)는 원점으로부터의 거리 $r = \sqrt{u^2+v^2}$를 통해서만 (u,v)에 의존하므로, 원점을 중심으로 어떻게 회전해도 똑같아 보인다. 특히 꼭짓점이 원점인 부채꼴(sector)에 떨어질 확률은 그 부채꼴의 중심각에만 의존한다. 사건 $U/V < k$는 직선 $u = kv$와 u축이 잘라 내는 서로 마주 보는 한 쌍의 부채꼴이므로, 그 확률은 단지 어떤 각을 2π로 나눈 값이다.

풀이

결합 밀도함수와 방사 대칭

X와 Y는 각각 밀도함수 $\phi(t) = \frac{1}{\sqrt{2\pi}}e^{-t^2/2}$를 가진다. X와 Y는 독립(independent)이므로, 결합 밀도함수(joint density)는 주변 밀도함수(marginal density)의 곱이다.

$$f(x,y) = \phi(x)\,\phi(y) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}\cdot\frac{1}{\sqrt{2\pi}}e^{-y^2/2} = \frac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2)},$$

이 식은 모든 실수 x와 y에 대하여 성립하며, 이것이 보이려던 것이다.

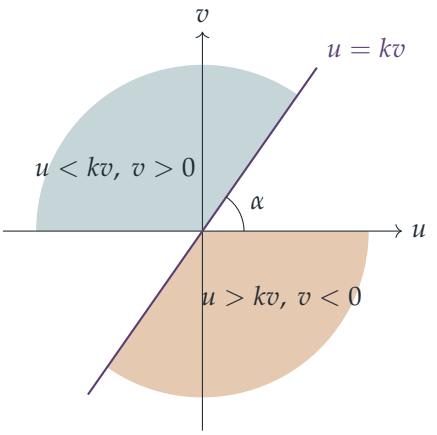
이 밀도함수가 **방사 대칭**이라는 것은, 점 (x,y)에서의 값이 그 점의 방향과는 관계없이 원점으로부터의 거리 $r = \sqrt{x^2+y^2}$에만 의존한다는 뜻이다. 여기서는 $f(x,y) = \frac{1}{2\pi}e^{-r^2/2}$이다. 같은 말로, 이 밀도함수는 원점을 중심으로 하는 평면의 어떤 회전에 의해서도 변하지 않으며, 그 등고선(contours)은 원점을 중심으로 하는 원이다.

$\mathrm{P}(U/V < k)$에 대한 항등식

U와 V가 방사 대칭인 결합 밀도함수 $h(u,v) = \rho\left(\sqrt{u^2+v^2}\right)$를 가진다고 하자. 직선 $v=0$의 확률은 0이므로(넓이가 0이다), $V=0$인 경우는 무시해도 된다. 부등식 $U/V < k$에 V를 곱하되, $V<0$일 때는 부등호의 방향을 바꾸면

$$\left\{\frac{U}{V} < k\right\} = \{U < kV,\ V > 0\} \cup \{U > kV,\ V < 0\},$$

이고, 우변의 두 사건은 서로 배반이다.



그림($k>0$인 경우)에서 보듯이, 원점을 지나는 직선 $u=kv$와 u축은 평면을 네 개의 부채꼴로 나누고, 색칠된 두 부채꼴이 사건 $U/V<k$를 이룬다. 사상 $(u,v)\mapsto(-u,-v)$는 원점을 중심으로 하는 π만큼의 회전이며, 영역 $\{u<kv,\ v>0\}$를 정확히 $\{u>kv,\ v<0\}$로 옮긴다(두 부등식에 모두 -1을 곱하면 된다). 방사 대칭인 밀도함수는 $h(-u,-v)=h(u,v)$를 만족하므로, 이 두 영역의 확률은 같다. 따라서

$$\mathrm{P}\left(\frac{U}{V} < k\right) = 2\,\mathrm{P}(U < kV,\ V > 0),$$

이고, 이것이 보여야 할 것이었다.

U/V의 밀도함수

극좌표(polar coordinates) $u = r\cos\theta$, $v = r\sin\theta$를 사용하자. 영역 $\{u<kv,\ v>0\}$는 부채꼴 $\alpha<\theta<\pi$이다. 여기서 $\alpha\in(0,\pi)$는 반직선 $u=kv, v>0$이 양의 u축과 이루는 각이므로 $\cot\alpha = k$이다. (실제로 $0<\theta<\pi$이면 $\sin\theta>0$이고, $u<kv$는 $\cot\theta<k$가 된다. cot은 $(0,\pi)$에서 감소하므로, 이는 $\theta>\alpha$와 같다.)

생각해 보기 함수 ρ를 모를 때, (U,V)가 꼭짓점이 원점이고 중심각이 β인 부채꼴 안에 있을 확률은 얼마인가?

극좌표에서 넓이 요소(area element)는 $r\,\mathrm{d}r\,\mathrm{d}\theta$이므로, 부채꼴 $\theta_1 < \theta < \theta_1 + \beta$의 확률은

$$\int_{\theta_1}^{\theta_1+\beta}\int_0^\infty \rho(r)\,r\,\mathrm{d}r\,\mathrm{d}\theta = \beta\int_0^\infty \rho(r)\,r\,\mathrm{d}r = \frac{\beta}{2\pi},$$

이다. $\beta = 2\pi$로 두면 평면 전체가 되어 $2\pi\int_0^\infty \rho(r)\,r\,\mathrm{d}r = 1$이기 때문이다. 다시 말해, ρ가 무엇이든 (U, V)의 방향은 균등분포(uniformly distributed)를 따른다.

우리의 부채꼴은 중심각이 $\pi - \alpha$이므로

$$\mathrm{G}(k) = \mathrm{P}\Big(\frac{U}{V} < k\Big) = 2\cdot\frac{\pi-\alpha}{2\pi} = 1 - \frac{\alpha}{\pi}.$$

$\alpha \in (0,\pi)$에서 $\cot\alpha = k$이므로 $\alpha = \frac{\pi}{2} - \arctan k$이다($\cot(\frac{\pi}{2} - \psi) = \tan\psi$이므로 양변 모두 $(0,\pi)$에 속하고 코탄젠트 값이 k이다). 그러므로

$$\mathrm{G}(k) = \frac{1}{2} + \frac{1}{\pi}\arctan k, \qquad -\infty < k < \infty.$$

이 함수는 분포함수(distribution function)가 그래야 하듯이 0에서 1까지 증가한다. 미분하면 U/V의 밀도함수를 얻는다.

$$\mathrm{g}(k) = \mathrm{G}'(k) = \frac{1}{\pi(1+k^2)}, \qquad -\infty < k < \infty,$$

이것이 보이려던 것이다. 특히 이 결과는 첫 부분의 서로 독립인 표준정규분포(standard normal) 확률변수에 대한 X/Y에도 적용된다.

검산

- $\mathrm{G}(0) = \frac{1}{2}$: 대칭성에 의해 U/V가 양수일 가능성과 음수일 가능성은 같다.
- $\int_{-\infty}^{\infty}\frac{\mathrm{d}k}{\pi(1+k^2)} = \frac{1}{\pi}\big[\arctan k\big]_{-\infty}^{\infty} = 1$.
- 서로 다른 세 가지 방사 대칭 분포(서로 독립인 두 표준정규분포, 단위원판 위의 균등분포, e^{-r}에 비례하는 밀도함수)에서 각각 2×10^5개의 점을 뽑는 시뮬레이션(simulation)을 해 보면, 매번 $\mathrm{P}(U/V < 0.7) \approx 0.694$를 얻었고, 이는 $\frac{1}{2} + \frac{1}{\pi}\arctan 0.7 \approx 0.6944$와 일치한다. 여러 k 값에 대해서도 항등식의 양변이 표집 오차(sampling error) 범위 안에서 일치하였다.

흔한 실수

- $\{U/V < k\} = \{U < kV\}$로 쓰는 것: $V < 0$일 때 V를 곱하면 부등호의 방향이 바뀐다.
- 주의 없이 $\alpha = \arctan(1/k)$로 두는 것: $k < 0$이면 이 값은 $(-\frac{\pi}{2}, 0)$에 속하지만, 필요한 각은 $(0,\pi)$에 속한다. $\alpha = \frac{\pi}{2} - \arctan k$로 쓰면 모든 k에 대해 성립한다.

STEP 너머

밀도함수 $\frac{1}{\pi(1+k^2)}$은 표준 **코시 분포**(Cauchy distribution)이다. 이 분포는 평균(mean)이 존재하지 않는 것으로 유명하며($|k|\,\mathrm{g}(k)$의 적분이 발산한다), 그래서 큰 수의 법칙(law of large numbers)이 성립하지 않는다. 서로 독립인 코시 확률변수 n개의 평균은 그중 하나와 똑같은 코시 분포를 따른다. 여기서 쓴 극좌표 논증은 야코비안(Jacobian)을 이용해 결합 밀도함수를 변환하는 방법, 즉 대학 1학년에서 배우는 표준적인 기법을 처음 맛보는 것이다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)이 두 가지를 모두 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.