풀이

STEP II 1988, 문제 16

확률과 통계 • 정규분포

> **STEP II 1988, 16번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/88-s2-q16

핵심 아이디어

근에 관한 모든 것은 하나의 확률변수(random variable) B에 의해 결정되므로, X_1과 X_2에 관한 각 질문을 B가 어느 구간에 놓이는가에 관한 질문으로 바꾸어야 한다. 판별식(discriminant)은 근이 언제 실수인지 알려 주고, 근과 계수의 관계(Vieta's formulas)(합 $-2B$, 곱 1)를 이용하면 (i)과 (ii)의 조건을 쉽게 옮길 수 있다. 그 뒤에는 각 답이 표준정규(standard normal) 확률변수에 대한 조건부 확률(conditional probability) 또는 조건부 기댓값(conditional expectation)이 되며, 기댓값은 짧은 적분 하나만 계산하면 된다.

풀이

이하에서 $B \sim \mathrm{N}(0,1)$이고, 확률밀도함수(density)는 $\phi(z) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-z^2/2}$, 분포함수(distribution function)는 Φ이다. 근이 실수라는 사건(event)을 R이라 하자.

실근

$X^2 + 2BX + 1$의 판별식은 $4B^2 - 4$이므로, 근이 실수일 필요충분조건은 $B^2 \geq 1$, 즉 $|B| \geq 1$이다. 정규분포의 대칭성(symmetry)에 의해

$$\mathrm{P}(R) = \mathrm{P}(|B| \geq 1) = 2(1 - \Phi(1)) = 2(1 - 0.84134) = 0.31731\ldots \approx 0.317.$$

(중근(equal root)인 경우 $|B| = 1$을 포함하든 안 하든 차이가 없다. 그 확률이 0이기 때문이다.)

근이 실수일 때 근은 다음과 같다.

$$X_{1,2} = -B \pm \sqrt{B^2 - 1}, \qquad X_1 + X_2 = -2B, \qquad X_1 X_2 = 1.$$

소문항 (i)

> **생각해 보기** "두 근이 모두 $\frac{1}{5}$보다 크다"를 어떻게 B에 관한 조건으로 바꿀 수 있을까?

먼저 두 근은 모두 양수여야 한다. 두 근의 곱이 $1 > 0$이므로 두 근의 부호는 같고, 그 부호는 두 근의 합 $-2B$의 부호와 같다. 따라서 $B < 0$이어야 하고, R과 함께 생각하면 이는 $B \le -1$을 뜻한다. $b = -B \ge 1$이라 두면 근은 $b \pm \sqrt{b^2 - 1}$이고, 둘 다 양수이다.

곱이 1이므로 작은 근은 큰 근의 역수(reciprocal)이다.

$$b - \sqrt{b^2 - 1} = \frac{1}{b + \sqrt{b^2 - 1}}.$$

따라서 작은 근이 $\frac{1}{5}$보다 클 필요충분조건은 큰 근이 5보다 작은 것이며, 이때 큰 근도 자동으로 $\frac{1}{5}$보다 크다. 그러므로 조건은

$$\sqrt{b^2 - 1} < 5 - b.$$

이를 위해서는 $b < 5$이어야 하고, $1 \le b < 5$에서는 양변이 모두 음이 아니므로 부등식을 바꾸지 않고 제곱할 수 있다.

$$b^2 - 1 < 25 - 10b + b^2 \iff 10b < 26 \iff b < 2.6.$$

($2.6 < 5$이므로 $b \ge 5$인 값은 어차피 조건을 만족하지 않는다.) 따라서 근이 실수일 때 두 근이 모두 $\frac{1}{5}$보다 클 필요충분조건은 $-2.6 < B \le -1$이고,

$$\begin{aligned} \mathrm{P}(X_1, X_2 > \tfrac{1}{5} \mid R) &= \frac{\mathrm{P}(-2.6 < B \le -1)}{\mathrm{P}(|B| \ge 1)} = \frac{\Phi(2.6) - \Phi(1)}{2(1 - \Phi(1))} \\ &= \frac{0.99534 - 0.84134}{0.31731} = 0.4853\ldots \approx 0.485. \end{aligned}$$

소문항 (ii)

$|X_1 + X_2| = 2|B|$이므로 구하려는 것은 $\mathrm{E}(2|B| \,|\, |B| \ge 1)$이다. 어떤 사건이 주어졌을 때의 조건부 기댓값은 그 사건 위에서 적분한 뒤 사건의 확률로 나누어 구한다.

$$\mathrm{E}(|B| \,|\, |B| \ge 1) = \frac{1}{\mathrm{P}(|B| \ge 1)} \int_{|z| \ge 1} |z|\, \phi(z)\, \mathrm{d}z = \frac{2\int_1^\infty z\, \phi(z)\, \mathrm{d}z}{2(1 - \Phi(1))},$$

여기서 ϕ의 대칭성을 이용하였다. $z\,\phi(z)$가 $\phi(z)$의 도함수에 음의 부호를 붙인 것이므로 적분은 쉽다.

$$\int_1^\infty z\, \phi(z)\, \mathrm{d}z = \Big[-\phi(z)\Big]_1^\infty = \phi(1) = \frac{\mathrm{e}^{-1/2}}{\sqrt{2\pi}} = 0.24197\ldots$$

그러므로

$$\mathrm{E}(|X_1 + X_2| \,|\, R) = \frac{2\phi(1)}{1 - \Phi(1)} = \frac{2 \times 0.24197}{0.15866} = 3.050\ldots \approx 3.05.$$

검산

- $B = -2.6$을 대입해 보자. 방정식은 $X^2 - 5.2X + 1 = 0$이고, 근은 정확히 5와 $\frac{1}{5}$로, (i)의 경계인 경우이다.
- R 위에서 $2|B| \geq 2$이므로 (ii)의 답은 2보다 커야 한다. 실제로 그렇고, 값도 그럴듯하다. 1 너머의 정규분포 꼬리는 얇으므로 $|B|$는 대개 1보다 조금만 크다(조건부 평균은 약 1.53이다).
- B의 표준정규 값 2×10^6개로 모의실험한 결과 0.3177, 0.4855, 3.049를 얻었으며, 이는 0.317, 0.485, 3.05와 일치한다.

흔한 실수

- 조건을 거는 것을 잊는 실수: (i)과 (ii)에서는 반드시 $\mathrm{P}(R) = 0.317$로 나누어야 한다. 무조건부 확률 $\Phi(2.6) - \Phi(1) = 0.154$는 답이 아니다.
- (i)에서 $B \geq 1$을 포함하는 실수. 이때는 두 근이 모두 음수이므로 어느 것도 $\frac{1}{5}$보다 크지 않다.
- $\mathrm{E}|X_1 + X_2|$를 $|\mathrm{E}(X_1 + X_2)|$로 계산하는 실수. 후자는 대칭성에 의해 0이다.

STEP 너머

값 $\phi(1)/(1 - \Phi(1))$은 **절단**(*truncated*) 정규분포의 평균이며, 비 $\phi(a)/(1 - \Phi(a))$는 역밀스 비(inverse Mills ratio)라고 불린다. 이 비는 통계학(예를 들어 중도절단 자료(censored data)의 모형)에 등장하고, 신뢰성 이론(reliability theory)에서는 정규분포의 위험률(hazard rate)로 나타난다. 계수가 무작위인 다항식이 평균적으로 실근을 몇 개 가지는가와 같은 무작위 다항식(random polynomial)에 관한 질문들은 놀라울 만큼 풍부한 확률론의 한 분야를 이룬다. 조건부 기댓값과 연속분포에 대해서는 Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문), 또는 Blitzstein과 Hwang의 *Introduction to Probability*(확률 입문)를 참고하라.

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.