풀이

STEP II 2002, 문제 13

확률과 통계 • 변환된 분포함수

STEP II 2002, 13번 문제

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

step.maths.org/questions/02-s2-q13

핵심 아이디어

역학 단원에 들어 있지만, 이 문제는 분포함수(distribution function)의 변환을 다루는 확률 문제이다. 모든 것은 하나의 식에 달려 있다. G를 미분하면 $G'(y) = f(y) \cdot \dfrac{2}{(2-F(y))^2}$을 얻는데, 여기서 $f = F'$은 X의 확률밀도함수(density)이다. 따라서 Y의 확률밀도함수는 X의 확률밀도함수에 $\frac{1}{2}$과 2 사이에 갇힌 가중치를 곱한 것이다. 이것만 얻으면 Y의 각 기댓값(expectation)은 X의 확률밀도함수에 $[\frac{1}{2}, 2]$ 안의 가중치를 곱한 것의 적분이 되고, 피적분함수를 비교하면 부등식이 따라 나온다. 조건 $a > 0$은 피적분함수가 음이 아니도록 하기 위한 것이다.

풀이

G의 성질

문제의 의도대로 X가 $[a, b]$에서 확률밀도함수 $f(y) = F'(y) \geq 0$을 가져서, X의 확률이 모두 $[a, b]$ 안에 있다고 가정한다.

$F(a) = 0$이고 $F(b) = 1$이므로

$$G(a) = \frac{0}{2-0} = 0, \qquad G(b) = \frac{1}{2-1} = 1.$$

몫의 미분법(quotient rule)에 의해

$$G'(y) = \frac{F'(y)(2-F(y)) - F(y)(-F'(y))}{(2-F(y))^2} = \frac{2F'(y)}{(2-F(y))^2}.$$

분포함수는 감소하지 않으므로 $F'(y) \geq 0$이고, 따라서 분자는 음이 아니다. 분모는 $2 - F(y) \geq 1$의 제곱이므로 양수이다. 그러므로 $G'(y) \geq 0$이다.

범위를 구하자. F는 분포함수이므로 $0 \le F(y) \le 1$이다. 따라서

$$1 \le 2 - F(y) \le 2 \quad \Longrightarrow \quad 1 \le (2 - F(y))^2 \le 4 \quad \Longrightarrow \quad \tfrac{1}{2} \le \frac{2}{(2 - F(y))^2} \le 2,$$

이다. 여기서 모든 값이 양수이므로 제곱해도 부등호의 방향이 유지되고, 역수를 취하면 방향이 바뀌며, 2를 곱하면 크기만 조정된다. 이것이 구하려던 이중 부등식이다.

$G(a) = 0, G(b) = 1$이고 G가 감소하지 않으므로, G는 $[a, b]$에서 진정한 분포함수이며 확률변수(random variable) Y는 잘 정의된다. 그 확률밀도함수는

$$g(y) = G'(y) = w(y)\, f(y), \qquad \text{단, } w(y) = \frac{2}{(2 - F(y))^2} \in [\tfrac{1}{2}, 2].$$

$\mathrm{E}(Y)$의 범위

생각해 보기 부등식 $\frac{1}{2} \le w(y) \le 2$에 $yf(y)$를 곱해도 부등호의 방향이 유지되는 이유는 무엇인가?

$[a, b]$에서 $yf(y) \ge 0$이기 때문이다. $f(y) \ge 0$이고 $y \ge a > 0$이다. 바로 여기서 가정 $a > 0$이 쓰인다. 따라서

$$\tfrac{1}{2}\, yf(y) \le y\, w(y) f(y) \le 2\, yf(y) \qquad \text{단, } a \le y \le b,$$

이고, $[a, b]$에서 적분해도 부등식은 유지된다.

$$\tfrac{1}{2} \int_a^b yf(y)\, \mathrm{d}y \;\le\; \int_a^b y\, g(y)\, \mathrm{d}y \;\le\; 2 \int_a^b yf(y)\, \mathrm{d}y,$$

즉,

$$\tfrac{1}{2}\, \mathrm{E}(X) \le \mathrm{E}(Y) \le 2\, \mathrm{E}(X),$$

이므로 원하는 결과를 얻는다.

$\mathrm{Var}(Y)$의 범위

같은 방법으로 $y^2 f(y) \ge 0$이므로(이것은 $f \ge 0$만 있으면 된다), $w(y) \le 2$에 이것을 곱하고 적분하면

$$\mathrm{E}(Y^2) = \int_a^b y^2\, w(y) f(y)\, \mathrm{d}y \le 2 \int_a^b y^2 f(y)\, \mathrm{d}y = 2\, \mathrm{E}(X^2).$$

빼는 항에 대해서는 $(\mathrm{E}(Y))^2$의 **하계**가 필요하다. 앞 부분에서 $\mathrm{E}(Y) \ge \frac{1}{2}\mathrm{E}(X)$이고, $X \ge a > 0$이므로 $\mathrm{E}(X) > 0$이다. 양변이 모두 양수이므로 안심하고 제곱할 수 있다.

$$(\mathrm{E}(Y))^2 \ge \tfrac{1}{4}(\mathrm{E}(X))^2.$$

그러므로

$$\begin{aligned}\operatorname{Var}(Y) &= \mathrm{E}(Y^2) - (\mathrm{E}(Y))^2 \\ &\le 2\,\mathrm{E}(X^2) - \tfrac{1}{4}(\mathrm{E}(X))^2 \\ &= 2(\operatorname{Var}(X) + (\mathrm{E}(X))^2) - \tfrac{1}{4}(\mathrm{E}(X))^2 \\ &= 2\operatorname{Var}(X) + \tfrac{7}{4}(\mathrm{E}(X))^2,\end{aligned}$$

이다. 셋째 줄에서는 $\mathrm{E}(X^2) = \operatorname{Var}(X) + (\mathrm{E}(X))^2$을 사용하였다. 이것이 구하려던 범위이다.

검산

X가 $[1,3]$에서 균등분포(uniform distribution)를 따른다고 하면 $F(y) = \frac{1}{2}(y-1)$이고 $\mathrm{E}(X) = 2$이다. $Y \ge 1$이므로 $\mathrm{E}(Y) = 1 + \int_1^3 (1 - G(y))\,\mathrm{d}y$이고, $u = F(y)$로 치환하면

$$\mathrm{E}(Y) = 1 + 2\int_0^1 \frac{2(1-u)}{2-u}\,\mathrm{d}u = 1 + 4(1 - \ln 2) \approx 2.227,$$

이므로 1과 4 사이에 여유 있게 들어간다. 직관적으로 보면 w는 F가 작은 곳에서 작고 F가 1에 가까운 곳에서 크므로, Y는 구간의 위쪽으로 치우치고 여기서는 $\mathrm{E}(Y) > \mathrm{E}(X)$이다. 또한 임의의 구간 위의 임의의 거듭제곱 법칙 분포 300개에 대해 $\frac{1}{2}\mathrm{E}(X) \le \mathrm{E}(Y) \le 2\mathrm{E}(X)$와 분산 범위가 모두 성립함을 수치적으로 확인하였다. 관찰된 비 $\mathrm{E}(Y)/\mathrm{E}(X)$는 대략 0.92와 1.31 사이에 있었다.

흔한 실수

- $a > 0$이 왜 중요한지 잊는 것: X가 음수일 수 있다면 $yf(y)$가 음수일 수 있고, 그것을 곱하면 부등호의 방향이 바뀐다.
- $\mathrm{E}(Y) \le 2\mathrm{E}(X)$를 사용해 $\operatorname{Var}(Y)$의 범위를 구하는 것: 항 $(\mathrm{E}(Y))^2$은 빼는 항이므로 상계가 아니라 하계가 필요하다.
- $(y-\mu)^2$을 g에 대해 적분하여 $\operatorname{Var}(Y)$를 $\operatorname{Var}(X)$와 직접 비교하려는 것: 두 분산은 서로 다른 평균을 사용하므로, 대신 $\mathrm{E}(Y^2)$을 다루어야 한다.

STEP 너머

$g = wf$로 쓰는 것은 측도 변환(change of measure)의 간단한 예이다. 새 분포는 옛 분포에 대한 밀도(우도비(likelihood ratio), 또는 라돈–니코딤 도함수(Radon–Nikodym derivative))를 가지며, 그 비에 대한 범위는 기댓값에 대한 범위로 곧바로 옮겨 간다. 이 아

이디어는 통계학의 중요도 표집(importance sampling)과 금융수학에서 쓰는 확률측도 변환의 바탕이 된다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 분포함수, 확률밀도함수, 기댓값을 이 내용을 더 발전시키기에 알맞은 수준으로 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.