풀이

STEP III 2020, 문제 11

확률과 통계 • 확률변수의 함수

> **STEP III 2020, 11번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 OCR에서 내려받을 수 있는 2020년 STEP 시험지 묶음(zip 파일) 가운데 STEP III 시험지에 있다:
>
> www.ocr.org.uk/Images/696589-2020-past-papers.zip

핵심 아이디어

순감소함수(strictly decreasing function)는 부등식의 방향을 뒤집는다. 또한 f는 자기 자신의 역함수이므로 $\mathrm{f}(X) \leq y$의 양변에 이 함수를 적용하면 X에만 관한 사건을 얻는다. 그다음 기댓값(expectation)은 부분적분(integration by parts)으로 구하는데, 이 과정에서 f'가 f로 바뀐다. 소문항 (ii)에서는 Z와 X 사이의 관계가 두 변수에 대해 대칭이므로, 자기역함수(self-inverse) f에 대해 $Z = \mathrm{f}(X)$이고 소문항 (i)을 적용할 수 있다. 그러면 부등식은 결국 분산(variance)이 양수라는 사실을 말하는 것일 뿐이다.

풀이

소문항 (i)

f^{-1}가 $[a, b]$ 전체에서 정의되므로 f는 $[a, b]$에서 자기 자신으로의 전단사(bijection)이고, 따라서 Y는 $[a, b]$ 안의 값을 가진다. $y \in [a, b]$를 고정하자. f는 순감소이므로 부등식의 양변에 이 함수를 적용하면 부등식의 방향이 바뀌고, 또 $\mathrm{f}(\mathrm{f}(x)) = x$이다. 따라서 $x \in [a, b]$에 대해

$$\mathrm{f}(x) \leq y \iff \mathrm{f}(\mathrm{f}(x)) \geq \mathrm{f}(y) \iff x \geq \mathrm{f}(y).$$

(역방향 함의도 $x \geq \mathrm{f}(y)$에 f를 적용하면 같은 방식으로 성립한다.) X는 $[a, b]$에서 균등분포(uniform distribution)를 따르고 $\mathrm{f}(y) \in [a, b]$이므로

$$\mathrm{P}(Y \leq y) = \mathrm{P}(X \geq \mathrm{f}(y)) = \frac{b - \mathrm{f}(y)}{b - a},$$

가 되어 요구된 식을 얻는다.

미분하면 Y의 확률밀도함수(probability density function)는

$$g(y) = -\frac{\mathrm{f}'(y)}{b-a} \quad \text{단, } a \le y \le b,$$

이고, 그 밖의 범위에서는 0이다. f가 감소함수이므로 이 값은 음이 아니다. 이제 $u = y^2$, $\mathrm{d}v = -\mathrm{f}'(y)\,\mathrm{d}y$로 두고 부분적분하면

$$\begin{aligned}\mathrm{E}(Y^2) &= \frac{1}{b-a}\int_a^b y^2(-\mathrm{f}'(y))\,\mathrm{d}y \\ &= \frac{1}{b-a}\Big[-y^2\mathrm{f}(y)\Big]_a^b + \int_a^b \frac{2y\,\mathrm{f}(y)}{b-a}\,\mathrm{d}y \\ &= \frac{-b^2a + a^2b}{b-a} + \int_a^b \frac{2y\,\mathrm{f}(y)}{b-a}\,\mathrm{d}y,\end{aligned}$$

이다. 여기서 $\mathrm{f}(b) = a$와 $\mathrm{f}(a) = b$를 이용했다. 첫째 항은 $-ab(b-a)/(b-a) = -ab$이다. 적분의 더미 변수(dummy variable) y를 x로 바꿔 쓰면

$$\mathrm{E}(Y^2) = -ab + \int_a^b \frac{2x\,\mathrm{f}(x)}{b-a}\,\mathrm{d}x.$$

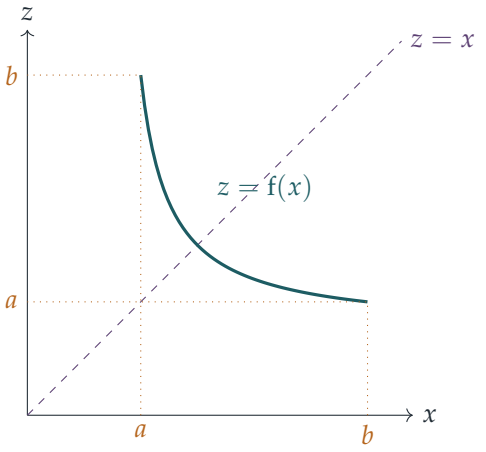
소문항 (ii)

$\frac{1}{c} = \frac{1}{a} + \frac{1}{b}$에서 $c = \dfrac{ab}{a+b}$이고, $\frac{1}{c} > \frac{1}{a}$이므로 $c < a$임에 주의하자. Z에 대해 풀면

$$\frac{1}{Z} = \frac{1}{c} - \frac{1}{X} = \frac{X-c}{cX}, \qquad \text{따라서} \qquad Z = \mathrm{f}(X) \ \text{여기서}\ \mathrm{f}(x) = \frac{cx}{x-c}.$$

$x \ge a > c$이므로 분모는 $[a,b]$에서 양수이다. f가 소문항 (i)의 세 조건을 만족하는지 확인하자.

- $\frac{1}{\mathrm{f}(a)} = \frac{1}{c} - \frac{1}{a} = \frac{1}{b}$이므로 $\mathrm{f}(a) = b$이다. 마찬가지로 $\frac{1}{\mathrm{f}(b)} = \frac{1}{c} - \frac{1}{b} = \frac{1}{a}$이므로 $\mathrm{f}(b) = a$이다.
- $\frac{1}{\mathrm{f}(x)} = \frac{1}{c} - \frac{1}{x}$는 $[a,b]$에서 양수이고 순증가하므로, $\mathrm{f}(x)$는 그 구간에서 순감소한다.
- 관계식 $\frac{1}{z} + \frac{1}{x} = \frac{1}{c}$는 x와 z에 대해 대칭이다. 따라서 $z = \mathrm{f}(x)$일 필요충분조건은 $x = \mathrm{f}(z)$이고, 이는 $\mathrm{f} = \mathrm{f}^{-1}$임을 뜻한다.



f의 그래프는 (a,b)에서 (b,a)까지 이어지며, 자기역함수라면 마땅히 그래야 하듯 직선 $z=x$에 대해 대칭이다.

생각해 보기 왜 $\mathrm{E}(Z^2)$뿐 아니라 $\mathrm{E}(Z)$까지 구해야 할까?

부등식은 $\mathrm{Var}(Z)>0$에서 나와야 하고, 분산을 구하려면 두 적률(moment)이 모두 필요하다. 로그는 각 적률에 나타나는데, 그 일차 항은 소거되고 제곱 항만 남는다.

두 적분 모두 먼저 나눗셈을 해 두자.

$$\mathrm{f}(x)=c+\frac{c^2}{x-c},\qquad x\,\mathrm{f}(x)=\frac{cx^2}{x-c}=c\left(x+c+\frac{c^2}{x-c}\right).$$

$L=\ln\left(\dfrac{b-c}{a-c}\right)$라 두면, $b-c>a-c>0$이므로 이 값은 양수이다. 그러면

$$\mathrm{E}(Z)=\frac{1}{b-a}\int_a^b\left(c+\frac{c^2}{x-c}\right)\mathrm{d}x=c+\frac{c^2L}{b-a}.$$

소문항 (i)에 의해

$$\begin{aligned}\mathrm{E}(Z^2)&=-ab+\frac{2c}{b-a}\int_a^b\left(x+c+\frac{c^2}{x-c}\right)\mathrm{d}x\\&=-ab+\frac{2c}{b-a}\left(\frac{b^2-a^2}{2}+c(b-a)+c^2L\right)\\&=-ab+c(a+b)+2c^2+\frac{2c^3L}{b-a}.\end{aligned}$$

그런데 $c(a+b)=ab$이므로 처음 두 항은 소거된다. 여기서 $\mathrm{E}(Z)^2=c^2+\dfrac{2c^3L}{b-a}+\dfrac{c^4L^2}{(b-a)^2}$를 빼면 L을 포함한 일차 항이 소거되어

$$\mathrm{Var}(Z)=c^2-\frac{c^4L^2}{(b-a)^2}.$$

를 얻는다. Z는 상수가 아니다. 소문항 (i)에 의해 그 분포함수(distribution function) $(b - \mathrm{f}(y))/(b-a)$는 $[a,b]$에서 0부터 1까지 순증가하기 때문이다. 따라서 $\mathrm{Var}(Z) > 0$이고, 이로부터

$$L^2 < \frac{(b-a)^2}{c^2}.$$

L과 $(b-a)/c$는 모두 양수이므로 양변에 제곱근을 취할 수 있다.

$$\ln\left(\frac{b-c}{a-c}\right) < \frac{b-a}{c}.$$

검산

- 소문항 (i)에서 가장 간단한 자기역함수의 예인 $\mathrm{f}(x) = a + b - x$를 시험해 보자. 이때 Y는 $[a,b]$에서 균등분포를 따르고, 공식은 $\mathrm{E}(Y^2) = -ab + \frac{2}{b-a}\int_a^b x(a+b-x)\,\mathrm{d}x = \frac{1}{3}(a^2+ab+b^2)$를 주는데, 이는 균등분포를 따르는 확률변수의 올바른 이차 적률이다.
- $a = 1, b = 3$이면 $c = \frac{3}{4}$, $L = \ln 9 \approx 2.197$이고, $(b-a)/c = \frac{8}{3} \approx 2.667$이다. 부등식이 성립하며, $\mathrm{Var}(Z) = \frac{9}{16} - \frac{81}{256}\cdot\frac{(\ln 9)^2}{4} \approx 0.180$이다. 컴퓨터로 f와 f^2를 수치적분하면 같은 분산을 얻는다.
- $b \to a$일 때 부등식의 양변은 모두 0으로 수렴하고, $\mathrm{Var}(Z) \to 0$이다. 이는 당연한 결과인데, Z가 $[a,b]$ 안에 갇히기 때문이다.

흔한 실수

- 감소함수가 부등식의 방향을 뒤집는다는 사실을 잊고 $\mathrm{P}(Y \le y) = \mathrm{P}(X \le \mathrm{f}(y))$라고 쓰는 실수.
- 밀도 $-\mathrm{f}'(y)/(b-a)$에서 음의 부호를 빠뜨리는 실수. 밀도는 음수일 수 없다.
- $\mathrm{Var}(Z)$가 왜 **엄격하게** 양수인지 설명하지 않거나, 양변이 모두 양수임을 밝히지 않고 제곱근을 취하는 실수.

STEP 너머

소문항 (i)은 밀도의 변수변환 공식(transformation rule)의 특수한 경우이다. 즉, $Y = \mathrm{f}(X)$이고 f가 단조함수이면 $g_Y(y) = g_X\big(\mathrm{f}^{-1}(y)\big)\,\big|(\mathrm{f}^{-1})'(y)\big|$이다. $\mathrm{Var}(Z) > 0$에서 엄격한 부등식을 얻는 것은 강력한 기법의 한 예이다. $\mathrm{E}(Z)^2 \le \mathrm{E}(Z^2)$는 사실 옌센(Jensen) 부등식(또는 코시–슈바르츠(Cauchy–Schwarz) 부등식)이 모습을 바꾼 것이며, 확률변수를 영리하게 고르면 여러 고전적인 부등식을 증명할 수 있다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 두 아이디어를 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.