풀이

STEP III 2013, 문제 13

확률과 통계 • 적률과 분포함수

> **STEP III 2013, 13번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/13-s3-q13

핵심 아이디어

$\mathrm{F}' = \mathrm{f}$이므로 식 $2\mathrm{F}(x)\mathrm{f}(x)$는 $(\mathrm{F}(x))^2$의 도함수(derivative)이다. 그래서 소문항 (i)(b)는 한 번의 부분적분(integration by parts)으로 끝나며, g를 알맞게 고르면 Y에 관한 모든 질문을 X에 관한 질문으로 바꿀 수 있다. 부등식들은 f나 F를 소문항 (i)(a)에서 얻은 더 큰 양 M 또는 Mx로 바꾸어 얻는데, 매번 적분이 커지는 방향으로 바꾼다.

풀이

이 풀이 전체에서 $\mathrm{F}(x) = \int_0^x \mathrm{f}(t)\,\mathrm{d}t$이므로 $\mathrm{F}'(x) = \mathrm{f}(x)$이고, X가 $[0,1]$에 놓이므로 $\mathrm{F}(0) = 0, \mathrm{F}(1) = 1$이다.

소문항 (i)

(a) $0 \le x \le 1$일 때, 모든 t에 대하여 $0 \le \mathrm{f}(t) \le M$이므로

$$0 \le \int_0^x \mathrm{f}(t)\,\mathrm{d}t \le \int_0^x M\,\mathrm{d}t = Mx,$$

이고, 이것이 곧 $0 \le \mathrm{F}(x) \le Mx$이다.

(b) 연쇄법칙(chain rule)에 의해 $\dfrac{\mathrm{d}}{\mathrm{d}x}(\mathrm{F}(x))^2 = 2\mathrm{F}(x)\mathrm{f}(x)$이다. $u = \mathrm{g}(x)$, $\mathrm{d}v = 2\mathrm{F}(x)\mathrm{f}(x)\,\mathrm{d}x$, 즉 $v = (\mathrm{F}(x))^2$으로 두고 부분적분하면

$$\begin{aligned}\int_0^1 2\mathrm{g}(x)\mathrm{F}(x)\mathrm{f}(x)\,\mathrm{d}x &= \Big[\mathrm{g}(x)(\mathrm{F}(x))^2\Big]_0^1 - \int_0^1 \mathrm{g}'(x)(\mathrm{F}(x))^2\,\mathrm{d}x \\ &= \mathrm{g}(1) - \int_0^1 \mathrm{g}'(x)(\mathrm{F}(x))^2\,\mathrm{d}x,\end{aligned}$$

이다. 여기서 $F(1) = 1, F(0) = 0$을 썼다. (g가 미분가능(differentiable)해야 하는데, 문제에서 이를 암묵적으로 가정하고 있다.)

소문항 (ii)

(a) Y의 밀도함수(density)를 적분하면 1이 되어야 한다. (i)(b)에서 $g(x) = 1$로 두면 $g' = 0$이므로

$$\int_0^1 2F(y)f(y)\,dy = 1,$$

이고, 따라서 $k = 2$이다. (같은 방법으로 $\int_0^1 kFf\,dy = \frac{k}{2}[F^2]_0^1 = \frac{k}{2}$로 보아도 된다.) 또한 $kFf \geq 0$이므로 이것은 실제로 밀도함수이다.

(b) 다음이 성립한다.

$$E(Y^n) = \int_0^1 2y^n F(y)f(y)\,dy.$$

상한(*upper bound*). (i)(a)에 의해 $F(y) \leq My$이고, 나머지 인수 $2y^n f(y)$는 음이 아니므로

$$E(Y^n) \leq \int_0^1 2y^n \cdot My \cdot f(y)\,dy = 2M\int_0^1 y^{n+1}f(y)\,dy = 2M\mu_{n+1}.$$

하한(*lower bound*). 이번에는 (i)(b)를 $g(y) = y^n$으로 두고 쓴다. 그러면 $g(1) = 1$, $g'(y) = ny^{n-1}$이므로

$$E(Y^n) = 1 - \int_0^1 ny^{n-1}(F(y))^2\,dy.$$

($n = 0$이면 이 적분은 그냥 0이다.) $E(Y^n)$의 하한을 구하려면 이 적분의 **상한**을 구해야 한다.

> **생각해 보기** $(F(y))^2$의 두 인수 중 어느 것을 My로 바꾸어야 할까?

하나만 바꾸면 된다. $0 \leq F(y) \leq My$이므로 $(F(y))^2 \leq My\,F(y)$이고, $ny^{n-1} \geq 0$이므로

$$\int_0^1 ny^{n-1}(F(y))^2\,dy \leq nM\int_0^1 y^n F(y)\,dy.$$

두 인수를 모두 바꾸면 $\int y^{n+1}\,dy$만 남는데, 이것은 X에 대해 아무것도 말해 주지 않는다. F 하나를 남겨 두면 한 번 더 부분적분하여 X의 적률(moment)을 만들 수 있다. $u = F(y)$, $dv = y^n\,dy$로 두고 부분적분하면

$$\int_0^1 y^n F(y)\,dy = \left[\frac{y^{n+1}}{n+1}F(y)\right]_0^1 - \int_0^1 \frac{y^{n+1}}{n+1}f(y)\,dy = \frac{1-\mu_{n+1}}{n+1}.$$

이를 종합하면

$$E(Y^n) \geq 1 - \frac{nM}{n+1}(1-\mu_{n+1}) = 1 + \frac{nM}{n+1}\mu_{n+1} - \frac{nM}{n+1},$$

이고, 앞의 상한과 함께 이것이 구하려던 한 쌍의 부등식이다.

(c) (b)의 양 끝을 합치면 $n \geq 0$에 대하여

$$1 + \frac{nM}{n+1}\mu_{n+1} - \frac{nM}{n+1} \leq 2M\mu_{n+1},$$

이고, μ_{n+1}이 들어 있는 항을 오른쪽으로 모으면

$$1 - \frac{nM}{n+1} \leq \left(2M - \frac{nM}{n+1}\right)\mu_{n+1} = \frac{(n+2)M}{n+1}\,\mu_{n+1}.$$

f를 적분하면 1이므로 $M > 0$이고, 따라서 계수 $\frac{(n+2)M}{n+1}$은 양수이다. 그러므로 부등호 방향을 바꾸지 않고 이 계수로 나눌 수 있다.

$$\mu_{n+1} \geq \frac{n+1}{(n+2)M} - \frac{n}{n+2}.$$

이 부등식은 모든 $n \geq 0$에 대하여 성립한다. $m = n+1$로 두면 $m \geq 1$이고, 부등식은

$$\mu_m \geq \frac{m}{(m+1)M} - \frac{m-1}{m+1} \qquad (m \geq 1),$$

이 된다. m을 n으로 다시 쓰면 이것이 구하려던 결과이다.

검산

- 균등분포(uniform distribution)를 따르는 X: $\mathrm{f} = 1$, $M = 1$, $\mu_n = \frac{1}{n+1}$이다. (c)의 하한은 $\frac{n}{n+1} - \frac{n-1}{n+1} = \frac{1}{n+1}$이므로 등호가 성립한다. 이때 Y의 밀도함수는 $2y$이고 $\mathrm{E}(Y^n) = \frac{2}{n+2}$이며, (b)의 두 부등식의 양 끝도 모두 $\frac{2}{n+2}$이다. 이는 자연스러운 결과이다. 우리가 사용한 부등식 $\mathrm{F}(y) \leq My$는 균등분포에서 모두 등식이 되기 때문이다.
- $\mathrm{f}(x) = 3x^2$, $M = 3$, $n = 2$: $\mu_2 = \frac{3}{5}$이고 하한은 $\frac{2}{9} - \frac{1}{3} < 0$이므로 넉넉히 성립한다. 이 밀도함수에 대하여 $\mathrm{E}(Y^2) = \frac{3}{4}$, $\mu_3 = \frac{1}{2}$이므로 (b)의 양 끝은 $1 + 2 \cdot \frac{1}{2} - 2 = 0$과 $2 \cdot 3 \cdot \frac{1}{2} = 3$이고, $\frac{3}{4}$는 그 사이에 있다.
- (i)(b)의 항등식과 모든 부등식은 여러 밀도함수와 여러 가지 g에 대하여 수치적분(numerical integration)으로도 확인하였다.

흔한 실수

- $\left(\mathrm{F}\right)^2$을 M^2y^2으로 누르는 것: 참이기는 하지만 X에 관한 정보를 모두 잃어버리므로 아무 데도 이르지 못한다.
- $\mathrm{E}(Y^n)$의 식에서 적분 앞에 빼기 부호가 붙어 있다는 것을 잊는 것. $\mathrm{E}(Y^n)$의 하한을 얻으려면 적분의 **상한**이 필요하다.
- μ_{n+1}에 대한 부등식에서 멈추고, 첨자를 옮겨 $n \geq 1$에 대한 주어진 꼴로 바꾸지 않는 것.

STEP 너머

밀도함수 $2\mathrm{F}(y)\mathrm{f}(y)$는 우연히 나온 것이 아니다. $\mathrm{P}(\max \leq y) = \big(\mathrm{F}(y)\big)^2$이므로, 이것은 X와 같은 분포를 따르는 서로 독립(independent)인 두 확률변수에 대한 $\max(X_1, X_2)$의 밀도함수이다. 더 일반적으로, n개의 독립인 관측값 중 r번째로 작은 값의 밀도함수는 F와 $1 - \mathrm{F}$의 거듭제곱으로 이루어진다. 이러한 **순서통계량**(order statistics)은 비모수 통계학(non-parametric statistics)의 핵심이다. Blitzstein과 Hwang의 *Introduction to Probability*(확률론 입문)에 순서통계량이 명료하게 설명되어 있다.

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.