풀이

STEP II 2006, 문제 14

확률과 통계 • 연속확률변수

> **STEP II 2006, 14번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/06-s2-q14

핵심 아이디어

이 문제 전체는 하나의 역도함수(antiderivative)에 달려 있다. 분모 $x\ln x$는 $\ln x$에 x를 곱한 것이고, $1/x$는 $\ln x$의 도함수(derivative)이므로, 피적분함수(integrand)는 $g(x) = \ln x$에 대하여 $g'(x)/g(x)$ 꼴이다. 따라서

$$\int \frac{1}{x\ln x}\,\mathrm{d}x = \ln|\ln x| + C.$$

그 다음부터 각 소문항은 "전체 확률은 1이다"를 적용하고, 부호와 함께 밀도함수(density)가 실제로 0이 아닌 구간이 어디인지를 꼼꼼히 따지는 문제가 된다.

풀이

그래프 그리기

$x > 0, x \neq 1$에 대하여 $y = \dfrac{1}{x\ln x}$라 하자.

부호. $x > 0$이므로 y의 부호는 $\ln x$의 부호와 같다. 즉 $0 < x < 1$에서 $y < 0$이고 $x > 1$에서 $y > 0$이다. 곡선은 좌표축과 만나지 않는다.

양 끝에서의 행동.

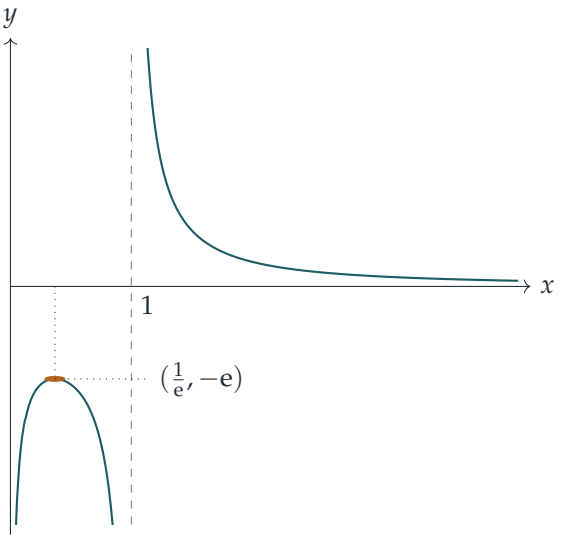
- $x \to 0^+$일 때: 문제에서 $x\ln x \to 0$임이 주어져 있고, 이 근처에서 $x\ln x < 0$이므로 $y \to -\infty$이다.
- $x \to 1$일 때: $x\ln x \to 0$이며, 왼쪽에서는 음수, 오른쪽에서는 양수이므로 $x \to 1^-$일 때 $y \to -\infty, x \to 1^+$일 때 $y \to +\infty$이다. 직선 $x = 1$은 수직 점근선(vertical asymptote)이다.

▶ $x \to \infty$일 때: $x \ln x \to \infty$이므로 $y \to 0^+$이고, x축이 점근선이다.

극점. 몫의 미분법(quotient rule)(또는 연쇄법칙(chain rule))에 의해

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{\ln x + 1}{(x \ln x)^2},$$

이고, 이 값은 $\ln x = -1$, 즉 $x = 1/\mathrm{e}$일 때에만 0이며, 이때 $y = \dfrac{1}{(1/\mathrm{e})(-1)} = -\mathrm{e}$이다. $0 < x < 1/\mathrm{e}$에서는 $\ln x + 1 < 0$이므로 y는 증가하고, $x > 1/\mathrm{e}$ $(x \neq 1)$에서는 감소한다. 따라서 $(1/\mathrm{e}, -\mathrm{e})$는 왼쪽 가지의 극대점(local maximum)이고, 오른쪽 가지는 $+\infty$에서 0까지 꾸준히 감소한다.



소문항 (i)

여기서 $a = \frac{1}{4}$, $b = \frac{1}{2}$이고, 둘 다 $x \ln x < 0$인 구간 $(0, 1)$에 속한다. 위의 역도함수를 이용하면

$$\begin{aligned}\int_{1/4}^{1/2} \frac{\mathrm{d}x}{x \ln x} &= \Big[\ln|\ln x|\Big]_{1/4}^{1/2} = \ln|\ln \tfrac{1}{2}| - \ln|\ln \tfrac{1}{4}| \\ &= \ln(\ln 2) - \ln(2 \ln 2) = -\ln 2.\end{aligned}$$

전체 확률은 1이어야 하므로 $\lambda \cdot (-\ln 2) = 1$이고, 따라서

$$\lambda = -\frac{1}{\ln 2}.$$

음의 부호는 바로 우리에게 필요한 것이다. $[\frac{1}{4}, \frac{1}{2}]$에서 $x \ln x < 0$이므로, 그 구간에서 f가 음이 아니려면 λ가 음수여야 한다.

소문항 (ii)

$\lambda = 1$이면 밀도함수가 음이 아니어야 하므로 구간 $[a, b]$는 $x \ln x > 0$인 곳에 놓여야 하는데, $b > a > 1$이므로 실제로 그렇다. 그러면 $\ln a, \ln b > 0$이고

$$1 = \int_a^b \frac{\mathrm{d}x}{x \ln x} = \ln(\ln b) - \ln(\ln a) = \ln \frac{\ln b}{\ln a}.$$

따라서 $\dfrac{\ln b}{\ln a} = \mathrm{e}$, 즉 $\ln b = \mathrm{e} \ln a = \ln\left(a^{\mathrm{e}}\right)$이고, ln은 일대일(one-to-one)이므로 요구한 대로 $b = a^{\mathrm{e}}$이다.

소문항 (iii)

$a = \mathrm{e}$이면 소문항 (ii)에 의해 $b = \mathrm{e}^{\mathrm{e}}$이다. $1 < \frac{3}{2} < 2 < \mathrm{e}$이므로 구간 $[\mathrm{e}^{3/2}, \mathrm{e}^2]$ 전체가 $[a, b] = [\mathrm{e}, \mathrm{e}^{\mathrm{e}}]$ 안에 들어 있고, 따라서

$$\mathrm{P}(\mathrm{e}^{3/2} \le X \le \mathrm{e}^2) = \Big[\ln(\ln x)\Big]_{\mathrm{e}^{3/2}}^{\mathrm{e}^2} = \ln 2 - \ln \tfrac{3}{2} = \ln \tfrac{4}{3}.$$

이 값을 분수로 나타내기 위해 매클로린 급수(Maclaurin series) $\ln(1+t) = t - \frac{t^2}{2} + \frac{t^3}{3} - \frac{t^4}{4} + \cdots$에 $t = \frac{1}{3}$을 대입하고 네 항까지 취하면

$$\ln \tfrac{4}{3} \approx \frac{1}{3} - \frac{1}{18} + \frac{1}{81} - \frac{1}{324} = \frac{108 - 18 + 4 - 1}{324} = \frac{93}{324} = \frac{31}{108}.$$

> **생각해 보기** 네 항에서 멈추는 것이 적절하다는 것을 어떻게 알 수 있을까?

이 급수는 항의 크기가 감소하는 교대급수(alternating series)이므로, 네 항까지 취했을 때의 오차는 다섯째 항 $\frac{1}{5 \cdot 3^5} = \frac{1}{1215} < 0.001$보다 작다. 따라서 요구한 대로 $\mathrm{P} = \ln \frac{4}{3} \approx \frac{31}{108}$이다.

소문항 (iv)

이제 $a = \mathrm{e}^{1/2}$이므로 소문항 (ii)에 의해

$$b = a^{\mathrm{e}} = \mathrm{e}^{\mathrm{e}/2}.$$

> **생각해 보기** 적분하기 전에, 밀도함수가 실제로 0이 아닌 곳은 어디인가?

$\mathrm{e} < 3$이므로 $\mathrm{e}/2 < \frac{3}{2}$이고, 따라서 $b = \mathrm{e}^{\mathrm{e}/2} < \mathrm{e}^{3/2}$이다. 구간 $[\mathrm{e}^{3/2}, \mathrm{e}^2]$는 완전히 b의 오른쪽에 있고, 그곳에서 $\mathrm{f}(x) = 0$이다. 따라서

$$\mathrm{P}(\mathrm{e}^{3/2} \le X \le \mathrm{e}^2) = 0.$$

($1/(x \ln x)$를 무작정 $\mathrm{e}^{3/2}$부터 e^2까지 적분하면 다시 $\ln \frac{4}{3}$이 나오는데, 이는 틀린 답이다.)

검산

- 수치적으로 $e^{-1} \approx 0.368$이고, $(0,1)$에서 y를 훑어보면 그 점에서 최댓값 $\approx -2.718 = -e$를 얻는다.
- 심프슨 공식(Simpson's rule)으로 계산하면 $\int_{1/4}^{1/2} \frac{dx}{x \ln x} \approx -0.6931 = -\ln 2$이므로 $\lambda \approx -1.4427$이다.
- 여러 $a > 1$ 값(예를 들어 $a = 2$, $a = 7.3$)에 대하여 $\int_a^{a^e} \frac{dx}{x \ln x}$를 계산하면 1.0000이 나온다.
- $\ln \frac{4}{3} \approx 0.28768$이고 $\frac{31}{108} \approx 0.28704$이므로, 두 값은 0.0007 이내에서 일치한다.
- (iv)에서 $b = e^{e/2} \approx 3.89$인 반면 $e^{3/2} \approx 4.48$이다.

흔한 실수

- $\ln x < 0$인 소문항 (i)에서 $\ln(\ln x)$라고 쓰는 것: 역도함수에는 절댓값이 필요하므로 $\ln|\ln x|$이다.
- $\lambda = \frac{1}{\ln 2}$라고 답하는 것: 양수 λ는 "밀도함수"를 음수로 만든다.
- 소문항 (iv)에서 $[e^{3/2}, e^2]$가 받침(support) $[a,b]$ 안에 있는지 확인하지 않고 그 구간에서 적분하는 것.

STEP 너머

치환 $u = \ln x$를 하면 이 밀도함수는 $[\ln a, \ln b]$에서의 $1/u$가 되고, 조건 $b = a^e$는 단지 $\int_{\ln a}^{\ln b} \frac{du}{u} = 1$일 뿐이다. 이처럼 확률변수(random variable)를 변환하는 것(여기서는 $U = \ln X$)은 대학 1학년 확률론의 핵심 도구로, 확률변수의 함수의 밀도함수를 연쇄법칙으로부터 구한다. 같은 함수 $1/(x \ln x)$는 해석학에서도 유명하다. $\sum 1/(n \ln n)$은 항이 0으로 수렴하는데도, 이 적분과의 비교에 의해 발산(diverge)한다. 확률변수의 변환에 대해서는 Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)을 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.