풀이

STEP II 2010, 문제 12

확률과 통계 • 평균과 중앙값

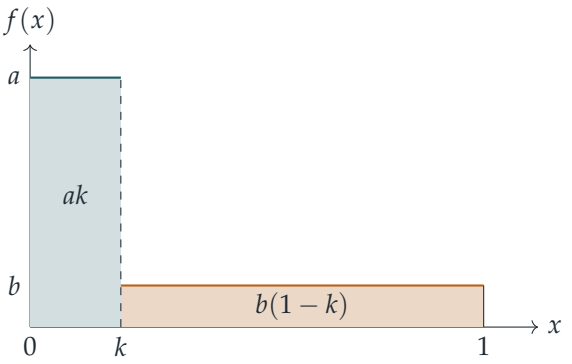
STEP II 2010, 12번 문제

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

step.maths.org/questions/10-s2-q12

핵심 아이디어

이 밀도함수(density function)는 두 단계로 된 계단함수(step function)이므로 모든 확률은 직사각형의 넓이이고, 실질적인 조건은 전체 넓이가 1이라는 것 하나뿐이다. 이 조건이 k를 a와 b로 결정하며, k를 소거하고 나면 나머지는 모두 대수 계산이다. 중앙값(median)에 대해서는 높은 첫 번째 직사각형이 이미 확률의 절반을 담고 있는지만 따지면 된다.



풀이

그림의 두 직사각형의 넓이는 ak와 $b(1-k)$이므로, 밀도함수의 적분값이 1이 될 필요충분조건은

$$ak + b(1-k) = 1, \qquad \text{즉,} \qquad k = \frac{1-b}{a-b}.$$

이다. ($a > b$이므로 $a-b$로 나눌 수 있다.) 이후 줄곧 이 k의 값을 사용한다.

$a > 1$이고 $b < 1$인 이유. 등식 $ak + b(1-k) = 1$은 1이 a와 b의 가중평균(weighted average)이고, 그 가중치 k와 $1-k$가 모두 0과 1 사이에 엄격하게 놓인다는 뜻이다. 양의 가중치를 갖는 가중평균은 두 값 사이에 엄격하게 놓이므로 $b < 1 < a$이다. 말없이 식으로 나타내면 $1 - b = k(a-b) > 0$이고 $a - 1 = (1-k)(a-b) > 0$이다.

소문항 (i)

각 구간에서 $xf(x)$를 적분하면

$$\mathrm{E}(X) = \int_0^k ax\,\mathrm{d}x + \int_k^1 bx\,\mathrm{d}x = \tfrac{1}{2}ak^2 + \tfrac{1}{2}b(1-k^2) = \tfrac{1}{2}(b + (a-b)k^2).$$

이다. 이제 $(a-b)k^2 = \dfrac{(1-b)^2}{a-b}$이므로

$$\mathrm{E}(X) = \frac{b(a-b) + (1-b)^2}{2(a-b)} = \frac{ab - b^2 + 1 - 2b + b^2}{2(a-b)} = \frac{1 - 2b + ab}{2(a-b)},$$

이고, 이것이 구하는 결과이다.

소문항 (ii)

누적분포함수(cumulative distribution function)는

$$\mathrm{F}(x) = \begin{cases} ax & 0 \le x < k, \\ ak + b(x-k) & k \le x \le 1. \end{cases}$$

이다. 이 함수는 $[0,1]$에서 연속이고 순증가하므로, 중앙값 M은 $\mathrm{F}(M) = \frac{1}{2}$의 유일한 해이다. 어느 식을 쓸지는 첫 번째 직사각형이 확률의 절반 이상을 담고 있는지에 따라 달라진다.

> ? **생각해 보기** "첫 번째 직사각형의 넓이가 $\frac{1}{2}$ 이상이다"를 어떻게 a와 b에 대한 조건으로 바꿀까?

k를 대입하고 양수인 분모 $a-b$를 없애면 다음을 얻는다.

$$ak \ge \tfrac{1}{2} \iff 2a(1-b) \ge a - b \iff 2a - 2ab \ge a - b \iff a + b \ge 2ab.$$

$a + b \ge 2ab$인 경우. 이때 $\mathrm{F}(k) = ak \ge \frac{1}{2}$이므로 중앙값은 $[0,k]$에 있고 $aM = \frac{1}{2}$를 만족한다. 따라서

$$M = \frac{1}{2a}.$$

$a + b \le 2ab$인 경우. 이때 $ak \le \frac{1}{2}$이므로 중앙값은 $[k,1]$에 있다. 오른쪽 끝에서부터 계산하는 것이 가장 쉽다. M의 오른쪽에 있는 확률은 $b(1-M)$이고, 이것이 $\frac{1}{2}$와 같아야 한다. 따라서

$$M = 1 - \frac{1}{2b} = \frac{2b-1}{2b}.$$

$a+b=2ab$일 때는 두 식이 모두 적용되며 값이 일치한다. 이때 $1-\frac{1}{2b}=\frac{2ab-a}{2ab}=\frac{b}{2ab}=\frac{1}{2a}$이기 때문이다.

소문항 (iii)

각 경우에 $\mathrm{E}(X)-M$을 계산하여 양수임을 보인다.

$a+b\geq 2ab$인 경우. 공통분모 $2a(a-b)$로 통분하면

$$\mathrm{E}(X)-\frac{1}{2a}=\frac{a(1-2b+ab)-(a-b)}{2a(a-b)}=\frac{b-2ab+a^2b}{2a(a-b)}=\frac{b(a-1)^2}{2a(a-b)}.$$

$a+b\leq 2ab$인 경우. 공통분모 $2b(a-b)$로 통분하면

$$\begin{aligned}\mathrm{E}(X)-\frac{2b-1}{2b}&=\frac{b(1-2b+ab)-(2b-1)(a-b)}{2b(a-b)}\\&=\frac{(b-2b^2+ab^2)-(2ab-2b^2-a+b)}{2b(a-b)}=\frac{a(1-b)^2}{2b(a-b)}.\end{aligned}$$

> **생각해 보기** 이 값들이 단지 음이 아닌 것이 아니라 엄격하게 양수인 이유는 무엇인가?

두 경우 모두 a, b, $a-b$는 양수이고, 처음에 $a>1$이고 $b<1$임을 보였으므로 제곱 항도 엄격하게 양수이다. 따라서 모든 경우에 $\mathrm{E}(X)-M>0$이고, $M<\mathrm{E}(X)$이다.

검산

- $a=3$, $b=\frac{1}{2}$로 두면 $k=\frac{1}{5}$, $a+b=\frac{7}{2}\geq 3=2ab$이고, $\mathrm{E}(X)=\frac{1-1+\frac{3}{2}}{5}=0.3$인 반면 $M=\frac{1}{6}$이다. 그 차 $0.1\overline{3}$은 $\frac{b(a-1)^2}{2a(a-b)}=\frac{2}{15}$와 일치한다.
- $a=\frac{5}{4}$, $b=\frac{9}{10}$로 두면 이번에는 $a+b=2.15\leq 2.25=2ab$이므로 $M=1-\frac{5}{9}=\frac{4}{9}\approx 0.444$이고, $\mathrm{E}(X)=\frac{13}{28}\approx 0.464$이다.
- $a\to 1^+$이면 (따라서 $b\to 1^-$이면) 밀도함수는 균등분포(uniform distribution)의 밀도함수로 수렴하고, 두 차는 모두 0으로 수렴한다. 대칭성에서 예상할 수 있듯이 평균과 중앙값은 $\frac{1}{2}$에서 일치한다.

임의로 고른 300개의 쌍 (a,b)에 대하여, F에 이분법(bisection)을 적용해 구한 중앙값이 경우별 공식과 일치하는지, 중점 수치적분(midpoint quadrature)으로 구한 평균이 소문항 (i)의 결과와 일치하는지, 그리고 $\mathrm{E}(X)-M$이 위의 인수분해된 식과 같은지를 수치적으로 확인하였다.

흔한 실수

- k를 소거하는 것을 잊는 실수. 답은 a와 b만으로 나타내야 한다.
- 두 번째 경우의 중앙값을 $b(M-k)=\frac{1}{2}$의 해로 써서, $[0,k)$에 이미 쌓인 확률 ak를 빠뜨리는 실수.
- $\mathrm{E}(X)\geq M$까지만 결론짓는 실수. 엄격한 부등식을 얻으려면 $a\neq 1$이고 $b\neq 1$이어야 하며, 바로 이것을 첫 부분에서 보였다.

STEP 너머

이 밀도함수는 감소하므로 "오른쪽으로 치우쳐(skewed to the right)" 있다. 즉 긴 꼬리가 평균을 중앙값보다 오른쪽으로 끌어당긴다. 이런 양상은 흔하지만 언제나 성립하는 것은 아니며, 대학의 첫 확률론 강의에서는 단봉분포(unimodal distribution)에서 평균, 중앙값, 최빈값(mode)이 어떤 관계에 있는지, 그리고 익숙한 경험 법칙이 어디에서 성립하지 않는지를 배운다. 이와 같은 계단함수는 히스토그램(histogram)의 기본 구성 요소이기도 하고, 측도론(measure theory)에서 적분을 정의할 때 쓰는 단순함수(simple function)의 기본 구성 요소이기도 하다. 연속분포에 대해 더 알고 싶다면 Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문서)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.