풀이

STEP I 2014, 문제 13

확률과 통계 • 삼각분포

STEP I 2014, 13번 문제

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

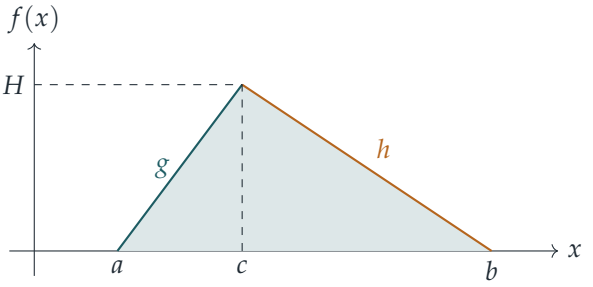
step.maths.org/questions/14-s1-q13

핵심 아이디어

밀도함수(density function)의 그래프는 밑변이 a에서 b까지이고 꼭짓점(apex)이 c 위에 있는 삼각형이며, 그 넓이는 1이어야 한다. 이 조건으로 꼭짓점의 높이가 정해지고, 그러면 각 일차함수 조각은 두 점으로 완전히 결정된다. 평균(mean)과 중앙값(median)을 구할 때에는 각 일차함수 조각이 0에서 시작하도록 변수를 평행이동하면 적분이 깔끔해진다. 중앙값의 경우 가장 먼저 물어야 할 것은 그것이 c의 어느 쪽에 있느냐이다.

풀이

풀이 내내 $a < c < b$로 두어 두 조각이 모두 실제로 존재하도록 한다(문제의 공식들이 의미를 가지려면 $c - a$와 $b - c$가 0이 아니어야 한다).



두 일차함수 조각

꼭짓점의 높이를 $H = g(c) = h(c)$라 하자. f의 그래프 아래 영역은 밑변이 $b-a$이고 높이가 H인 삼각형이며, 전체 확률(probability)은 1이므로

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

g는 일차함수이고 $g(a) = 0$이므로 어떤 상수 k에 대하여 $g(x) = k(x-a)$이며, $g(c) = H$로부터 $k(c-a) = \dfrac{2}{b-a}$를 얻는다. 따라서

$$g(x) = \frac{2(x-a)}{(b-a)(c-a)},$$

이고, 이것이 구하는 식이다. 같은 방법으로 $h(c) = H$인 $h(x) = m(b-x)$로 두면

$$h(x) = \frac{2(b-x)}{(b-a)(b-c)}.$$

$b-a$, $c-a$, $b-c$가 모두 양수이므로, 요구된 대로 g는 증가함수이고 h는 감소함수이다.

소문항 (i)

$\mathrm{E}(X) = I_1 + I_2$로 쓰자. 여기서 $I_1 = \int_a^c x\,g(x)\,\mathrm{d}x$이고 $I_2 = \int_c^b x\,h(x)\,\mathrm{d}x$이다.

I_1에서 $x = a + t$로 치환(substitution)하면 t는 0부터 $c-a$까지 변한다.

$$\begin{aligned} I_1 &= \frac{2}{(b-a)(c-a)} \int_0^{c-a} (a+t)\,t\,\mathrm{d}t = \frac{2}{(b-a)(c-a)} \left[\frac{a(c-a)^2}{2} + \frac{(c-a)^3}{3}\right] \\ &= \frac{(c-a)(3a+2(c-a))}{3(b-a)} = \frac{(c-a)(a+2c)}{3(b-a)}. \end{aligned}$$

I_2에서 $x = b - s$로 치환하면 s는 0부터 $b-c$까지 변한다.

$$\begin{aligned} I_2 &= \frac{2}{(b-a)(b-c)} \int_0^{b-c} (b-s)\,s\,\mathrm{d}s = \frac{2}{(b-a)(b-c)} \left[\frac{b(b-c)^2}{2} - \frac{(b-c)^3}{3}\right] \\ &= \frac{(b-c)(3b-2(b-c))}{3(b-a)} = \frac{(b-c)(b+2c)}{3(b-a)}. \end{aligned}$$

두 식을 더하면 분자는

$$\begin{aligned} (c-a)(a+2c) + (b-c)(b+2c) &= (ac + 2c^2 - a^2 - 2ac) + (b^2 + 2bc - bc - 2c^2) \\ &= b^2 - a^2 + bc - ac \\ &= (b-a)(b+a+c). \end{aligned}$$

이를 $3(b-a)$로 나누면

$$\mathrm{E}(X) = \tfrac{1}{3}(a+b+c),$$

이고, 이것이 구하는 결과이다. (이것은 삼각형의 무게중심(centroid)의 x좌표인데, 우연이 아니다. 평균은 밀도함수 아래 영역의 균형점이기 때문이다.)

소문항 (ii)

먼저 꼭짓점의 왼쪽에 확률이 얼마나 있는지 구한다. 이것은 왼쪽 삼각형의 넓이이다.

$$\mathrm{P}(X \le c) = \tfrac{1}{2}(c-a)H = \frac{c-a}{b-a}.$$

생각해 보기 중앙값이 c의 왼쪽에 있는지 오른쪽에 있는지 어떻게 알 수 있을까?

중앙값 M은 $\mathrm{P}(X \le M) = \frac{1}{2}$를 만족한다. $\mathrm{P}(X \le c) \ge \frac{1}{2}$이면 중앙값은 c이거나 c의 왼쪽에 있고, 그렇지 않으면 오른쪽에 있다. 그런데 $\frac{c-a}{b-a} \ge \frac{1}{2}$인 것은 정확히 $c \ge \frac{1}{2}(a+b)$일 때, 즉 꼭짓점이 $[a,b]$의 중점이거나 그 오른쪽에 있을 때이다. 따라서 경우는 c와 $\frac{1}{2}(a+b)$를 비교하여 나뉜다.

경우 $c \ge \frac{1}{2}(a+b)$. 이때 $a < M \le c$이고, $a \le m \le c$에 대하여 a부터 m까지의 작은 삼각형의 넓이는

$$\mathrm{P}(X \le m) = \int_a^m g(x)\,\mathrm{d}x = \frac{(m-a)^2}{(b-a)(c-a)}.$$

이 값을 $\frac{1}{2}$로 놓고 양의 제곱근을 취하면($M > a$이므로)

$$M = a + \sqrt{\tfrac{1}{2}(b-a)(c-a)}.$$

경우 $c < \frac{1}{2}(a+b)$. 이번에는 중앙값이 c의 오른쪽에 있으므로, 오른쪽 끝에서부터 계산하는 편이 쉽다. $c \le m \le b$에 대하여

$$\mathrm{P}(X \ge m) = \int_m^b h(x)\,\mathrm{d}x = \frac{(b-m)^2}{(b-a)(b-c)}.$$

이 값을 $\frac{1}{2}$로 놓고 $M < b$를 이용하면

$$M = b - \sqrt{\tfrac{1}{2}(b-a)(b-c)}.$$

$c = \frac{1}{2}(a+b)$일 때에는 분포가 대칭이며, 두 공식 모두 $M = \frac{1}{2}(a+b)$를 준다(예를 들어 첫 번째 공식은 $a + \sqrt{\frac{1}{4}(b-a)^2} = a + \frac{1}{2}(b-a)$를 준다). 따라서 두 경우는 매끄럽게 이어진다.

검산

- $a = 0, b = 4, c = 1$일 때: $\mathrm{P}(X \le c) = \frac{1}{4} < \frac{1}{2}$이므로 두 번째 경우에 해당하고 $M = 4 - \sqrt{6} \approx 1.5505$이다. 평균은 $\frac{5}{3}$이다. 컴퓨터로 200 000개의 값을 모의실험(simulation)한 결과 표본 중앙값은 1.5505, 표본 평균은 1.668이었다.
- $a = 0, b = 4, c = 3$일 때: 첫 번째 경우에 해당하고 $M = \sqrt{6} \approx 2.449$이며, 이는 마땅히 그래야 하듯 앞의 예를 거울에 비춘 모습이다.

- 중앙값은 항상 최빈값(mode) c와 중점 $\frac{1}{2}(a+b)$ 사이에 있다. 예를 들어 첫 번째 경우 $M \leq c$이고, $c \geq \frac{1}{2}(a+b)$일 때 $(b-a)(c-a) \geq \frac{1}{2}(b-a)^2$이므로 $M \geq \frac{1}{2}(a+b)$이다.

흔한 실수

- 중앙값을 구할 때 밀도함수의 엉뚱한 조각을 사용하는 것: 먼저 중앙값이 c의 어느 쪽에 있는지 확인하라.
- 음의 제곱근을 취하는 것: 그러면 M이 $[a, b]$ 밖에 놓이게 된다.
- 평균의 적분을 평행이동 없이 x에 대하여 바로 전개하는 것: 답은 나오지만 계산이 훨씬 무거워지고 실수가 끼어든다.

STEP 너머

삼각분포(triangular distribution)는 c가 중점일 때 서로 독립(independent)인 두 균등(uniform) 확률변수의 합의 분포이며, 최솟값, 최댓값, 가장 가능성이 높은 값만 알려져 있을 때(프로젝트 계획에서처럼) 대략적인 모형으로 널리 쓰인다. 독립인 확률변수들의 합의 분포를 계산하다 보면 합성곱(convolution)에 이르게 되는데, 이것은 대학 1학년 확률론의 핵심 도구이다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 이에 관해 읽기 좋은 책이다.

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.