풀이

STEP III 2015, 문제 13

확률과 통계 • 균등 확률변수의 함수

STEP III 2015, 13번 문제

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

step.maths.org/questions/15-s3-q13

핵심 아이디어

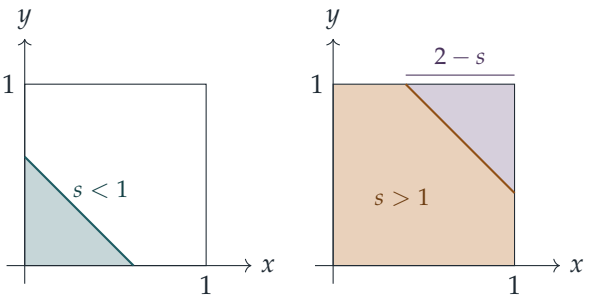
X와 Y가 서로 독립(independent)이고 $[0,1]$에서 균등분포(uniform distribution)를 따르므로, 순서쌍 (X,Y)는 단위정사각형(unit square) 위에서 균등하게 분포한다. 따라서 X와 Y에 관한 모든 확률은 그저 넓이이다. $X+Y \le s$나 $Y \le rX$ 같은 사건은 정사각형을 직선을 따라 자르며, 한쪽 부분의 넓이는 삼각형의 넓이이거나 정사각형에서 삼각형을 뺀 넓이이다. 그다음에는 $1/(X+Y)$와 $X/(X+Y)$에 관한 사건을 이미 넓이를 구한 사건들로 다시 써서 그 분포를 얻는다.

풀이

풀이 전체에서 (X,Y)는 넓이가 1인 단위정사각형 $[0,1]^2$ 위에서 균등분포를 따른다. 따라서 정사각형의 임의의 영역 A에 대하여 $\mathrm{P}((X,Y)\in A)$는 A의 넓이와 같다.

소문항 (i)

$[0,2]$의 값을 가지는 $S=X+Y$를 생각하고, $F(s)=\mathrm{P}(S\le s)$라 하자. 사건 $S\le s$는 정사각형에서 직선 $x+y=s$ 위 또는 그 아래에 있는 부분이다.



$0 \le s \le 1$이면 직선은 두 축과 $(s,0)$, $(0,s)$에서 만나고, 그 아래 영역은 직각을 낀 두 변의 길이가 s인 직각이등변삼각형이다. $1 \le s \le 2$이면 직선 위의 영역이 꼭짓점 $(1,1)$ 쪽 모서리에 있는 삼각형이 된다. 이 삼각형의 직각을 낀 두 변은 직선에서 모서리까지 이어지며, 그 길이는 $1-(s-1)=2-s$이다. 따라서

$$F(s) = \mathrm{P}(X+Y \le s) = \begin{cases} 0 & s < 0, \\ \frac{1}{2}s^2 & 0 \le s \le 1, \\ 1 - \frac{1}{2}(2-s)^2 & 1 \le s \le 2, \\ 1 & s > 2. \end{cases}$$

가운데 두 식은 $s=1$에서 일치하며(둘 다 $\frac{1}{2}$이다), 당연히 그래야 한다.

소문항 (ii)

$T = 1/S$라 하자. $0 < S \le 2$이므로(값 $S=0$의 확률은 0이다), T는 $[\frac{1}{2}, \infty)$의 값을 가진다. $t > 0$에 대하여, $s \mapsto 1/s$가 $s > 0$에서 감소하므로

$$\mathrm{P}(T \le t) = \mathrm{P}\left(S \ge \tfrac{1}{t}\right) = 1 - F\left(\tfrac{1}{t}\right),$$

이다. 여기서 S가 연속확률변수(continuous random variable)이므로 $\mathrm{P}(S = 1/t) = 0$이라는 사실을 이용하였다.

> **생각해 보기** F의 어느 식을 사용해야 할까? 그것은 t가 어디에 있는지가 아니라 $1/t$가 어디에 있는지에 따라 정해진다.

$t \ge 1$이면 $0 < 1/t \le 1$이므로 첫 번째 식을 사용한다: $\mathrm{P}(T \le t) = 1 - \frac{1}{2}t^{-2}$. $\frac{1}{2} \le t \le 1$이면 $1 \le 1/t \le 2$이므로 두 번째 식을 사용한다: $\mathrm{P}(T \le t) = \frac{1}{2}\left(2 - t^{-1}\right)^2$. 그리고 $t < \frac{1}{2}$이면 $\mathrm{P}(T \le t) = 0$이다.

미분하면 밀도함수(density)를 얻는다. $t \ge 1$일 때

$$f(t) = \frac{\mathrm{d}}{\mathrm{d}t}\left(1 - \tfrac{1}{2}t^{-2}\right) = t^{-3},$$

이고, $\frac{1}{2} \le t \le 1$일 때는 연쇄법칙(chain rule)에 의해

$$f(t) = \left(2 - t^{-1}\right) \cdot t^{-2} = 2t^{-2} - t^{-3},$$

이며, 그 밖에서는 $f(t) = 0$이다. 이것이 구하는 밀도함수이다.

기댓값(expectation)은 다음과 같다.

$$\begin{aligned} \mathrm{E}\Big(\frac{1}{X+Y}\Big) = \int_{1/2}^{\infty} t f(t)\,\mathrm{d}t &= \int_{1/2}^{1} \left(2t^{-1} - t^{-2}\right)\mathrm{d}t + \int_{1}^{\infty} t^{-2}\,\mathrm{d}t \\ &= \left[2\ln t + t^{-1}\right]_{1/2}^{1} + \left[-t^{-1}\right]_{1}^{\infty} \\ &= (0+1) - (-2\ln 2 + 2) + 1 \\ &= 2\ln 2. \end{aligned}$$

t^{-2}가 $[1,\infty)$에서 적분 가능하므로 이 이상적분(improper integral)은 수렴한다.

소문항 (iii)

$[0,\infty)$의 값을 가지는 $R = Y/X$를 생각하자(사건 $X = 0$의 확률은 0이다). $r \ge 0$에 대하여 사건 $R \le r$는 $Y \le rX$, 즉 정사각형에서 원점을 지나는 직선 $y = rx$ 위 또는 그 아래에 있는 부분이다.

$0 \le r \le 1$이면 이 직선은 오른쪽 변의 점 $(1, r)$에서 정사각형을 빠져나가고, 그 아래 영역은 밑변이 1, 높이가 r인 삼각형으로 넓이는 $\frac{1}{2}r$이다. $r \ge 1$이면 직선은 윗변의 점 $(1/r, 1)$에서 빠져나가고, 직선 위의 영역은 윗변을 따라 놓인 밑변이 $1/r$이고 높이가 1인 삼각형으로 넓이는 $\frac{1}{2r}$이다. 따라서

$$H(r) = \mathrm{P}\Big(\frac{Y}{X} \le r\Big) = \begin{cases} 0 & r < 0, \\ \frac{1}{2}r & 0 \le r \le 1, \\ 1 - \dfrac{1}{2r} & r \ge 1. \end{cases}$$

이제 $[0,1]$의 값을 가지는 $Z = \dfrac{X}{X+Y} = \dfrac{1}{1+R}$를 생각하자. $0 < z \le 1$에 대하여, $r \mapsto 1/(1+r)$가 $r \ge 0$에서 감소하므로

$$\mathrm{P}(Z \le z) = \mathrm{P}\Big(1 + R \ge \frac{1}{z}\Big) = \mathrm{P}\Big(R \ge \frac{1-z}{z}\Big) = 1 - H\Big(\frac{1-z}{z}\Big).$$

이번에도 어느 식을 쓸지는 $(1-z)/z$가 어디에 있는지에 따라 정해진다. 이 값이 1 이하인 것은 정확히 $z \ge \frac{1}{2}$일 때이다.

- $\frac{1}{2} \le z \le 1$일 때: $\mathrm{P}(Z \le z) = 1 - \dfrac{1-z}{2z} = \dfrac{3}{2} - \dfrac{1}{2z}$.

- $0 < z \le \frac{1}{2}$일 때: $\mathrm{P}(Z \le z) = 1 - \left(1 - \dfrac{z}{2(1-z)}\right) = \dfrac{z}{2(1-z)}$.

두 식 모두 $z = \frac{1}{2}$에서 $\frac{1}{2}$이다. 미분하면(두 번째 식은 $\frac{z}{2(1-z)} = \frac{1}{2}\left(\frac{1}{1-z} - 1\right)$로 고쳐 쓴다) Z의 밀도함수는

$$g(z) = \begin{cases} \dfrac{1}{2(1-z)^2} & 0 \le z \le \frac{1}{2}, \\ \dfrac{1}{2z^2} & \frac{1}{2} \le z \le 1, \\ 0 & \text{그 밖.} \end{cases}$$

$g(1-z) = g(z)$임에 주목하자. 즉 밀도함수는 $\frac{1}{2}$에 대하여 대칭이다.

기댓값. 적분하지 않고도 바로 구할 수 있다. X와 Y는 서로 독립이고 같은 분포를 따르므로, 순서쌍 (X, Y)와 (Y, X)는 같은 분포를 가진다. 따라서 $\dfrac{X}{X+Y}$와 $\dfrac{Y}{X+Y}$는 같은 기댓값을 가진다. 둘의 합은 정확히 1이므로

$$\mathrm{E}\left(\frac{X}{X+Y}\right) = \frac{1}{2}.$$

적분으로 확인하기.

$$\mathrm{E}(Z) = \int_0^{1/2} \frac{z}{2(1-z)^2}\,\mathrm{d}z + \int_{1/2}^1 \frac{1}{2z}\,\mathrm{d}z.$$

두 번째 적분은 $\frac{1}{2}\left[\ln z\right]_{1/2}^1 = \frac{1}{2}\ln 2$이다. 첫 번째 적분에서 $u = 1 - z$로 치환하면 $\mathrm{d}z = -\mathrm{d}u$이고, $z = 0, \frac{1}{2}$는 $u = 1, \frac{1}{2}$에 대응한다:

$$\begin{aligned} \int_0^{1/2} \frac{z}{2(1-z)^2}\,\mathrm{d}z &= \frac{1}{2}\int_{1/2}^1 \frac{1-u}{u^2}\,\mathrm{d}u = \frac{1}{2}\left[-\frac{1}{u} - \ln u\right]_{1/2}^1 \\ &= \frac{1}{2}\left(-1 - (-2 + \ln 2)\right) = \frac{1}{2}(1 - \ln 2). \end{aligned}$$

더하면 $\mathrm{E}(Z) = \frac{1}{2}(1 - \ln 2) + \frac{1}{2}\ln 2 = \frac{1}{2}$이 되어 위의 값이 확인된다.

검산

- 각 분포함수(distribution function)는 0에서 1까지 증가하고, 식이 바뀌는 지점에서 연속이다: $F(1) = \frac{1}{2}$, $\mathrm{P}(T \le 1) = \frac{1}{2}$, $H(1) = \frac{1}{2}$, $\mathrm{P}(Z \le \frac{1}{2}) = \frac{1}{2}$. 이 값들이 모두 $\frac{1}{2}$인 것은 정사각형이 대각선에 대하여 대칭이기 때문이다.
- 밀도함수 f를 적분하면 1이 된다: $\int_{1/2}^1 (2t^{-2} - t^{-3})\,\mathrm{d}t = 2 - \frac{3}{2} = \frac{1}{2}$이고 $\int_1^\infty t^{-3}\,\mathrm{d}t = \frac{1}{2}$이다.
- $2\ln 2 \approx 1.386$은 당연히 1과 2 사이에 있고, 볼록함수 $1/s$에 대한 옌센 부등식(Jensen's inequality)이 예측하듯이 $1/\mathrm{E}(X+Y) = 1$보다 크다. 정사각형 위에서 $\iint (x +$

$y)^{-1}\,\mathrm{d}x\,\mathrm{d}y$를 직접 수치 계산하면 1.3863을 얻는다. 분포함수들과 $\mathrm{E}(Z) = \frac{1}{2}$도 컴퓨터 시뮬레이션으로 확인하였다.

흔한 실수

- F나 H의 식을 $1/t$나 $(1-z)/z$가 아니라 t나 z에 따라 고르는 것. 그러면 두 식이 뒤바뀐다.
- $1/s$와 $1/(1+r)$가 감소함수여서 부등호의 방향을 바꿔야 한다는 사실을 잊는 것: $T \le t$는 $S \ge 1/t$를 뜻한다.
- 소문항 (i)에서 모서리 삼각형의 변의 길이를 잘못 구하는 것: 그 길이는 $s-1$이 아니라 $2-s$이다.

STEP 너머

(X, Y)에 관한 명제를 넓이로 바꾼 다음 분포함수를 미분하는 것은, 확률변수(random variable)의 함수의 분포를 구하는 일반적인 **분포함수법**(distribution function method)이다. 대학에서는 여기에 결합밀도함수(joint density)에 대한 변수변환(야코비안, Jacobian) 공식이 더해지는데, 이 공식을 쓰면 $X+Y$와 $X/(X+Y)$를 함께 다룰 수 있다. 특히 여기서 $X+Y$의 밀도함수는 두 균등분포 밀도함수의 **합성곱**(convolution)이다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 두 방법을 모두 다루며, Blitzstein과 Hwang의 *Introduction to Probability*(확률 입문)에는 같은 맥락의 예제가 많이 실려 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.