풀이

STEP II 1999, 문제 13

확률과 통계 • 연속확률변수

> **STEP II 1999, 13번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/99-s2-q13

핵심 아이디어

역학 단원으로 분류되어 있지만, 이 문제는 사실 연속확률변수(continuous random variable)에 관한 문제이다. **짧은** 조각이 간단한 분포를 따른다는 점만 알아차리면 모든 것이 쉬워진다. 막대가 부러지는 지점이 막대 위에서 균등하게 분포하면, 짧은 조각의 길이는 전체 길이의 $[0, \frac{1}{2}]$에서 균등분포(uniform distribution)를 따른다. 비율 R는 짧은 조각의 길이에 대한 증가함수이므로, "$R < r$"는 짧은 조각만에 관한 명제로 바뀐다. 이로부터 분포함수(distribution function)를 얻고, 이를 미분하면 확률밀도함수를 얻는다.

풀이

R의 분포

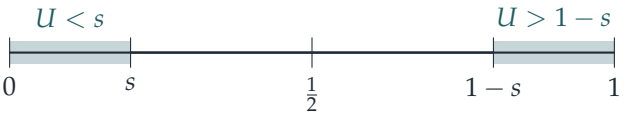
막대의 길이를 1로 놓자. 비율은 단위에 의존하지 않는다. 부러지는 지점이 한쪽 끝에서 거리 U만큼 떨어져 있다고 하고, U는 $[0,1]$에서 균등분포를 따른다고 하자. 두 조각의 길이는 U와 $1-U$이므로, 짧은 조각의 길이는

$$S = \min(U, 1-U),$$

이고 긴 조각의 길이는 $1-S$이다. 그러면 $0 \le S \le \frac{1}{2}$이고

$$R = \frac{S}{1-S}.$$

$0 \le s \le \frac{1}{2}$에 대하여, 사건 $S < s$는 정확히 $U < s$ 또는 $U > 1-s$일 때 일어난다. 이 둘은 서로소인 두 구간이고 길이의 합이 $2s$이므로 $\mathrm{P}(S < s) = 2s$이다. 다시 말해 S는 $[0, \frac{1}{2}]$에서 균등분포를 따른다.



이제 $0 \le r \le 1$이라 하자. 함수 $s \mapsto \dfrac{s}{1-s}$는 $[0, \frac{1}{2}]$에서 순증가한다(도함수가 $1/(1-s)^2 > 0$이다). 따라서 부등식을 다음과 같이 바꿀 수 있다.

$$R < r \iff \frac{S}{1-S} < r \iff S < r(1-S) \iff S < \frac{r}{1+r},$$

여기서 양변에 $1-S$를 곱해도 되는 것은 $1 - S \ge \frac{1}{2} > 0$이기 때문이다. $0 \le r \le 1$이면 $\dfrac{r}{1+r}$의 값은 $[0, \frac{1}{2}]$에 속하므로

$$\boxed{\mathrm{P}(R < r) = \frac{2r}{1+r} \qquad (0 \le r \le 1).}$$

짧은 조각이 긴 조각보다 길 수는 없으므로 R는 항상 $[0,1]$에 속한다. 따라서 $r \le 0$이면 $\mathrm{P}(R < r) = 0$이고 $r \ge 1$이면 $\mathrm{P}(R < r) = 1$이며, 이는 $r = 1$에서 공식의 값이 1인 것과 맞아떨어진다.

? **생각해 보기** 왜 U를 직접 써서 $R = U/(1-U)$로 계산하지 않는가?

그 식은 $U \le \frac{1}{2}$일 때만 옳다. $U > \frac{1}{2}$이면 짧은 조각은 $1-U$이고 $R = (1-U)/U$이다. 이 두 경우로 나누어 계산할 수도 있는데, 대칭성에 의해 각 경우가 $\frac{r}{1+r}$씩 기여한다. S로 넘어가면 이런 경우 나누기를 한 번에 처리할 수 있다.

확률밀도함수, 평균, 분산

분포함수 $F(r) = \dfrac{2r}{1+r} = 2 - \dfrac{2}{1+r}$를 $(0,1)$에서 미분하면 확률밀도함수(probability density function)

$$f(r) = \frac{2}{(1+r)^2} \quad (0 \le r \le 1), \qquad f(r) = 0 \text{ 그 외.}$$

를 얻는다. 검산으로, $\int_0^1 f(r)\,\mathrm{d}r = F(1) - F(0) = 1$이다.

평균(mean)을 구하기 위해 $r = (1+r) - 1$로 쓰면

$$\begin{aligned}\mathrm{E}(R) &= \int_0^1 \frac{2r}{(1+r)^2}\,\mathrm{d}r = 2\int_0^1 \left(\frac{1}{1+r} - \frac{1}{(1+r)^2}\right)\mathrm{d}r \\ &= 2\Big[\ln(1+r) + \frac{1}{1+r}\Big]_0^1 = 2\left(\ln 2 + \tfrac{1}{2} - 1\right) \\ &= 2\ln 2 - 1 \approx 0.386.\end{aligned}$$

이차 적률(second moment)을 구하기 위해 $r^2 = (1+r)^2 - 2(1+r) + 1$로 쓰면

$$\frac{r^2}{(1+r)^2} = 1 - \frac{2}{1+r} + \frac{1}{(1+r)^2}.$$

이다. 그러면

$$\begin{aligned}\mathrm{E}(R^2) &= 2\int_0^1 \left(1 - \frac{2}{1+r} + \frac{1}{(1+r)^2}\right) \mathrm{d}r = 2\Big[r - 2\ln(1+r) - \frac{1}{1+r}\Big]_0^1 \\ &= 2\left(1 - 2\ln 2 - \tfrac{1}{2} + 1\right) = 3 - 4\ln 2.\end{aligned}$$

따라서 분산(variance)은

$$\begin{aligned}\mathrm{Var}(R) &= \mathrm{E}(R^2) - (\mathrm{E}(R))^2 = 3 - 4\ln 2 - (4(\ln 2)^2 - 4\ln 2 + 1) \\ &= 2 - 4(\ln 2)^2 \approx 0.0782.\end{aligned}$$

이다. $\ln 2$ 항들이 상쇄되어 이렇게 깔끔한 꼴이 남는다.

검산

- $r = 1$에서 확률은 당연히 1이다. $r = \frac{1}{2}$에서는 $\frac{2}{3}$이다. 긴 조각이 짧은 조각의 두 배보다 길어지는 것은 정확히 부러지는 지점이 막대의 양 끝 3분의 1 구간에 떨어질 때이고, 이 구간들의 길이의 합은 $\frac{2}{3}$이다.
- 평균 0.386은 0과 1 사이에 있고 $\frac{1}{2}$보다 작다. 이는 확률밀도함수가 $r = 0$에서 2로부터 $r = 1$에서 $\frac{1}{2}$로 감소한다는 사실과 잘 맞는다.
- 무작위로 10^6번 부러뜨린 시뮬레이션에서 표본평균은 0.3862, 표본분산은 0.0783이었다. 이론값은 $2\ln 2 - 1 = 0.38629$와 $2 - 4(\ln 2)^2 = 0.07819$이다.

흔한 실수

- 모든 U에 대하여 $R = U/(1-U)$로 놓는 것. 이 값은 1보다 클 수 있으므로 짧은 조각과 긴 조각의 비율이 아니다.
- 부러지는 지점이 양쪽 끝 어느 쪽 근처에도 있을 수 있다는 것을 잊어서 $\mathrm{P}(R < r) = \frac{r}{1+r}$를 얻는 것.
- 짧은 조각의 평균 $\frac{1}{4}$를 써서 $\mathrm{E}(R) = \frac{1/4}{3/4} = \frac{1}{3}$로 계산하는 것. 일반적으로 $\mathrm{E}\big(g(S)\big) \neq g\big(\mathrm{E}(S)\big)$이다.

STEP 너머

여기서 쓴 방법은 확률변수를 변환할 때 쓰는 표준적인 "분포함수 기법(distribution function technique)"이다. g가 증가함수이고 $Y = g(X)$이면 $F_Y(y) = F_X\left(g^{-1}(y)\right)$이고, 이를 미분하면 변수변환 공식(change-of-variables formula) $f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\mathrm{d}g^{-1}/\mathrm{d}y\right|$를 얻는다. 같은 아이디어는 순서통계량(order statistics)과, 세 조각으로 삼각형을 만들 수 있을 확률과 같은 고전적인 막대 부러뜨리기 문제(broken-stick problem)로 이어진다. 연속확률변수의 변환에 대해서는 Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)을 읽어 보기 좋다.

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.