풀이

STEP III 2005, 문제 14

확률과 통계 • 연속확률변수

> **STEP III 2005, 14번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/05-s3-q14

핵심 아이디어

이 문제의 모든 계산은 인용해도 좋다고 주어진 단 하나의 적분에 기대고 있으므로, 핵심은 각 적분에 어떤 m과 n이 필요한지 알아내는 것이다. 치환(substitution) $x = k^2/u$는 피적분함수(integrand)의 모양은 그대로 둔 채 구간 $[0,k]$와 $[k,\infty)$를 서로 맞바꾸며, 이 대칭성 덕분에 중앙값(median)이 곧바로 정해진다. 시간 $T = s/V$에 대해서도 같은 역수(reciprocal) 기법을 쓰면, T의 밀도함수(density)가 V와 정확히 같은 族이 아니라 같은 분포족(family of distributions)에 속하고 k만 s/k로 바뀐 꼴임을 알 수 있다. 그다음에는 V에 대한 결과로부터 중앙값과 평균을 바로 읽어 낼 수 있다.

풀이

상수 C

주어진 결과에서 $m = a, n = 2a$로 놓는다. a는 양의 정수이고 $2a \geq a$이므로 이렇게 놓을 수 있다. 그러면 $n + 2 = 2a + 2$, $n - m + 1 = a + 1$이므로

$$\int_0^\infty \frac{x^a}{(x+k)^{2a+2}}\,\mathrm{d}x = \frac{a!\,a!}{(2a+1)!\,k^{a+1}}.$$

밀도함수를 그 범위 전체에서 적분하면 1이므로

$$1 = Ck^{a+1} \cdot \frac{a!\,a!}{(2a+1)!\,k^{a+1}} \qquad \Longrightarrow \qquad C = \frac{(2a+1)!}{a!\,a!},$$

이고, 이것이 구하려던 결과이다.

치환과 중앙값

좌변의 적분에서 $x = k^2/u$로 놓으면 $\mathrm{d}x = -\dfrac{k^2}{u^2}\,\mathrm{d}u$이다. x가 0에서 v까지 움직일 때 u는 ∞에서 k^2/v까지 내려간다. 또

$$x + k = \frac{k^2}{u} + k = \frac{k(u+k)}{u},$$

이므로

$$\frac{x^a}{(x+k)^{2a+2}} = \frac{k^{2a}/u^a}{k^{2a+2}(u+k)^{2a+2}/u^{2a+2}} = \frac{u^{a+2}}{k^2\,(u+k)^{2a+2}}.$$

여기에 $\mathrm{d}x$를 곱하면 k^2/u^2이 약분되고, 음의 부호는 적분 구간의 위끝과 아래끝을 뒤바꾼다.

$$\int_0^v \frac{x^a}{(x+k)^{2a+2}}\,\mathrm{d}x = \int_\infty^{k^2/v} \frac{u^a}{(u+k)^{2a+2}}\,(-\mathrm{d}u) = \int_{k^2/v}^\infty \frac{u^a}{(u+k)^{2a+2}}\,\mathrm{d}u,$$

이것이 구하려던 항등식이다.

생각해 보기 v를 어떤 값으로 잡으면 이 항등식이 중앙값에 대해 무언가를 말해 주는가?

$v = k$로 잡으면 $k^2/v = k$도 된다. 양변에 Ck^{a+1}을 곱하면 항등식은 다음과 같이 된다.

$$\begin{aligned}\mathrm{P}(V < k) &= \int_0^k \mathrm{f}(x)\,\mathrm{d}x \\ &= \int_k^\infty \mathrm{f}(u)\,\mathrm{d}u = \mathrm{P}(V > k).\end{aligned}$$

V는 연속(continuous)확률변수이므로 $\mathrm{P}(V = k) = 0$이고, 위의 두 확률의 합은 1이다. 따라서 각각은 $\frac{1}{2}$이고, V의 중앙값은 k이다.

V의 기댓값

$$\mathrm{E}(V) = Ck^{a+1}\int_0^\infty \frac{x^{a+1}}{(x+k)^{2a+2}}\,\mathrm{d}x.$$

이번에는 $m = a+1$, $n = 2a$로 놓는다. 이를 위해서는 $2a \geq a+1$이어야 하는데, $a \geq 1$이므로 성립한다. 그러면 $n - m + 1 = a$이고,

$$\mathrm{E}(V) = \frac{(2a+1)!}{a!\,a!}\,k^{a+1}\cdot\frac{(a+1)!\,(a-1)!}{(2a+1)!\,k^a} = k\cdot\frac{(a+1)!}{a!}\cdot\frac{(a-1)!}{a!} = \frac{(a+1)k}{a}.$$

T의 분포

속력은 양수이므로, $t > 0$일 때 양변에 $V/t > 0$을 곱해도 부등호의 방향이 바뀌지 않는다.

$$T < t \iff \frac{s}{V} < t \iff V > \frac{s}{t}.$$

따라서

$$\mathrm{P}(T<t)=\mathrm{P}\Big(V>\frac{s}{t}\Big)=\int_{s/t}^{\infty}\frac{Ck^{a+1}x^{a}}{(x+k)^{2a+2}}\,\mathrm{d}x, \qquad (*)$$

이고, 이것이 구하려던 결과이다.

이제 $(*)$에서 $u=s/x$로 치환하면 $x=s/u$, $\mathrm{d}x=-\dfrac{s}{u^2}\,\mathrm{d}u$이다. x가 s/t에서 ∞까지 움직일 때 u는 t에서 0까지 내려간다. $x+k=(s+ku)/u$이므로

$$\mathrm{P}(T<t)=\int_0^t Ck^{a+1}\,\frac{s^a/u^a}{(s+ku)^{2a+2}/u^{2a+2}}\cdot\frac{s}{u^2}\,\mathrm{d}u=\int_0^t\frac{Ck^{a+1}s^{a+1}\,u^a}{(s+ku)^{2a+2}}\,\mathrm{d}u.$$

이 식은 모든 $t>0$에 대해 성립하고, $t\le 0$이면 $\mathrm{P}(T<t)=0$이다. 그러므로 t에 대해 미분하면 T의 밀도함수를 얻는다.

$$\mathrm{g}(t)=\frac{Ck^{a+1}s^{a+1}\,t^a}{(s+kt)^{2a+2}}\qquad(t>0).$$

이를 f와 비교하기 위해 분모에서 k^{2a+2}을 묶어 내고 $K=s/k$로 쓰면

$$\mathrm{g}(t)=\frac{C\,(s/k)^{a+1}\,t^a}{(t+s/k)^{2a+2}}=\frac{CK^{a+1}t^a}{(t+K)^{2a+2}}.$$

이것은 밀도함수 f에서 k를 $K=s/k$로 바꾼 것과 정확히 같다. 따라서 T는 V와 같은 분포족에 속한다.

중앙값과 평균

T는 모수가 $K=s/k$인 같은 꼴의 분포를 따르므로, 앞의 결과를 그대로 적용할 수 있다.

$$\text{중앙값}T=\frac{s}{k},\qquad \mathrm{E}(T)=\frac{(a+1)}{a}\cdot\frac{s}{k}.$$

그러므로 시간의 중앙값과 속력의 중앙값의 곱은

$$\frac{s}{k}\cdot k=s,$$

이고, 주장한 대로 이것은 거리와 같다. 시간의 기댓값과 속력의 기댓값의 곱은

$$\mathrm{E}(T)\,\mathrm{E}(V)=\frac{(a+1)s}{ak}\cdot\frac{(a+1)k}{a}=s\Big(1+\frac{1}{a}\Big)^2>s,$$

이다. a는 양의 정수이므로 $1+1/a>1$이기 때문이다. 이로써 문제의 풀이가 끝난다.

검산

- $a=1, k=1$이면 $C=3!/1=6$이고, $6x/(x+1)^4$을 $[0,\infty)$에서 수치적분(numerical integration)하면 1.000000이 나온다. $[0,1]$에서는 0.500000이 나와 중앙값이 맞음을 확인할 수 있다.

- $a = 2, k = 0.7$ 등 여러 쌍에 대해 수치적분을 하면 $\mathrm{E}(V) = 1.05000 = \frac{3}{2}(0.7)$이 나온다. $s = 3.2$로 두면 $\mathrm{E}(T)\mathrm{E}(V) = 7.20000 = 3.2 \times \frac{9}{4}$이 나온다.
- $a = 2, k = s = 1$로 두고 속력 2×10^5개를 모의실험(simulation)한 결과, 속력의 중앙값 $\times$ 시간의 중앙값 ≈ 1.000001이었지만 속력의 평균은 ≈ 1.504(이론값: 1.5)였다.
- 중앙값의 곱이 s와 같다는 사실은 공식 없이도 알 수 있다. $T = s/V$는 V에 대한 감소함수(decreasing function)이므로, T의 중앙값은 s를 V의 중앙값으로 나눈 값이다.

흔한 실수

- 주어진 적분을 인용할 때 $n \geq m$인지 확인하지 않는 것. $\mathrm{E}(V)$의 경우에는 $a \geq 1$이 필요한데, 바로 이 때문에 a가 양의 정수여야 한다. 더 높은 차수의 적률(moment)에서는 결국 공식이 성립하지 않게 되며, 실제로 $\mathrm{E}(V^{a+1})$은 무한대이다.
- T의 밀도함수를 구하려고 $(*)$를 미분할 때 인수 s/t^2(연쇄법칙(chain rule))을 빠뜨리는 것.
- $\mathrm{E}(s/V) = s/\mathrm{E}(V)$라고 가정하는 것. 마지막 소문항의 요점이 바로 이것이 거짓이라는 것이다.

STEP 너머

마지막 부등식은 옌센 부등식(Jensen's inequality)의 한 예이다. $v > 0$에서 $1/v$는 볼록(convex)하므로 $\mathrm{E}(1/V) \geq 1/\mathrm{E}(V)$이고, 등호는 V가 상수일 때만 성립한다. 반면 중앙값은 어떤 단조함수(monotonic function)를 거쳐도 그대로 옮겨 간다. 치환 $V/(V+k)$는 V를 $\frac{1}{2}$에 대해 대칭인 베타(Beta) 분포 $\mathrm{Beta}(a+1, a+1)$를 따르는 확률변수로 바꾸는데, 이것도 중앙값이 k인 이유를 보는 또 다른 방법이다. Blitzstein과 Hwang의 *Introduction to Probability*(확률론 입문)는 옌센 부등식, 변수변환(change of variables), 베타 분포족을 매우 읽기 쉽게 다룬다.

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.