풀이

STEP I 1988, 문제 16

확률과 통계 • 연속 분포와 조건부

STEP I 1988, 16번 문제

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

step.maths.org/questions/88-s1-q16

핵심 아이디어

첫 부분은 정규화(normalisation)와 적분 연습이다. 밀도함수(density)가 상수배를 제외하고만 주어져 있으므로, 전체 확률이 1이 되어야 한다는 조건으로 그 상수를 구해야 한다. 또 $t^2/(1+t^2)^2$의 적분은 이를 $1/(1+t^2)$와 $t/(1+t^2)$의 도함수로 나누면 풀린다. 그다음 검사 방식은 짧은 경우의 수 나무로 정리된다. 나이 3 이전에는 어떤 차도 녹슬지 않으므로 의미가 있는 것은 나이 4와 5의 검사뿐이다. 그래서 네 번째 검사 때 차가 이미 녹슬어 있었는지에 따라 경우를 나누고, 마지막 부분은 같은 나무에 베이즈 정리(Bayes' theorem)를 적용한다.

풀이

녹이 생기기 시작하는 시점의 분포

차가 나이 $3+T$에 녹슬기 시작한다고 하자. 여기서 $T \geq 0$는 어떤 상수 k에 대하여 밀도함수 $f(t) = k\,t^2/(1+t^2)^2$를 가진다. 핵심 적분은 다음을 알아차리는 데서 나온다.

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{t}{1+t^2}\right) = \frac{1-t^2}{(1+t^2)^2}, \qquad \text{따라서} \qquad \frac{t^2}{(1+t^2)^2} = \frac{1}{2}\left(\frac{1}{1+t^2} - \frac{1-t^2}{(1+t^2)^2}\right),$$

이는 괄호 안을 공통분모 $(1+t^2)^2$로 통분하여 확인할 수 있다. 분자는 $(1+t^2)-(1-t^2) = 2t^2$이다. 그러므로

$$\int_0^t \frac{s^2}{(1+s^2)^2}\,\mathrm{d}s = \frac{1}{2}\left(\arctan t - \frac{t}{1+t^2}\right).$$

$t \to \infty$일 때 $\arctan t \to \frac{\pi}{2}$이고 $t/(1+t^2) \to 0$이므로, $[0,\infty)$에서의 적분값은 $\frac{\pi}{4}$이다. f가 밀도함수가 되려면 $k \cdot \frac{\pi}{4} = 1$이어야 하므로 $k = 4/\pi$이다. 따라서

$$\boxed{F(t) = \mathrm{P}(\text{다음 나이 이전에 녹슮: } 3+t) = \mathrm{P}(T < t) = \frac{2}{\pi}\left(\arctan t - \frac{t}{1+t^2}\right), \quad t \geq 0.}$$

($F(t) \to 1$임에 주목하자. 모든 차는 결국 녹슨다.)

필요한 두 값은 다음과 같다.

$$F(1) = \frac{2}{\pi}\left(\frac{\pi}{4} - \frac{1}{2}\right) = \frac{1}{2} - \frac{1}{\pi} \approx 0.1817,$$
$$F(2) = \frac{2}{\pi}\left(\arctan 2 - \frac{2}{5}\right) \approx 0.4502.$$

5주년 검사에서 불합격할 확률

나이 3 이전에는 어떤 차도 녹슬지 않으므로, 나이 1, 2, 3의 검사는 반드시 통과한다(세 번째 검사 순간에 정확히 녹이 생기기 시작할 확률은 0이다). 나이 5에 불합격하려면 차는 나이 4의 검사를 통과하고 나이 5의 검사에서 불합격해야 하며, 불합격하려면 나이 5에 녹슬어 있어야 한다. 녹은 결코 사라지지 않으므로, 이런 일이 일어나는 방식은 서로 배반(disjoint)인 두 가지이다.

- **나이 4에 이미 녹슬어 있는 경우** (확률 $F(1)$). 확률 $\frac{1}{2}$로 네 번째 검사를 통과하고, 여전히 녹슨 채로 확률 $\frac{1}{2}$로 다섯 번째 검사에서 불합격한다. 기여분: $\frac{1}{4}F(1)$.
- **나이 4와 5 사이에 녹이 생기기 시작하는 경우** (확률 $F(2) - F(1)$). 네 번째 검사는 반드시 통과하고, 확률 $\frac{1}{2}$로 다섯 번째 검사에서 불합격한다. 기여분: $\frac{1}{2}\big(F(2) - F(1)\big)$.

여기서는 문제가 의도한 대로, 녹슨 차에 대한 서로 다른 검사의 결과가 서로 독립(independent)이고 녹이 생기기 시작한 시점과도 독립이라고 가정한다.

? 생각해 보기 왜 나이 1, 2, 3의 검사를 통과할 확률을 곱하지 않는가?

그 검사들은 어떤 차도 녹슬 수 없는 시기에 이루어지므로 모든 차가 통과하기 때문이다. 그 인수는 1이다.

두 경우를 더하면

$$\begin{aligned}\mathrm{P}(\text{불합격 시점: 나이5}) &= \tfrac{1}{4}F(1) + \tfrac{1}{2}\big(F(2) - F(1)\big) = \tfrac{1}{2}F(2) - \tfrac{1}{4}F(1) \\ &= \frac{1}{\pi}\left(\arctan 2 - \frac{2}{5}\right) - \frac{1}{8} + \frac{1}{4\pi} \\ &= \boxed{\frac{1}{\pi}\left(\arctan 2 - \frac{3}{20}\right) - \frac{1}{8} \approx 0.1797.}\end{aligned}$$

다섯 번째 검사 후 폐차되었을 때, 네 번째 검사 때 녹슬어 있었을 확률

다섯 번째 검사 직후 폐차되는 차는 정확히 (이전 검사를 모두 통과하고) 그 검사에서 불합격한 차이며, 이는 방금 구한 사건이다. 네 번째 검사 때 녹슬어 있었다는 사건을 A, 다섯 번째 검사에서 불합격했다는 사건을 D라 하자. 위의 첫 번째 항목이 바로 $A \cap D$이므로

$$\mathrm{P}(A \cap D) = \tfrac{1}{4}F(1) = \frac{1}{8} - \frac{1}{4\pi} = \frac{\pi - 2}{8\pi}.$$

조건부 확률(conditional probability)의 정의에 의해

$$\mathrm{P}(A \mid D) = \frac{\mathrm{P}(A \cap D)}{\mathrm{P}(D)} = \frac{\dfrac{\pi - 2}{8\pi}}{\dfrac{1}{\pi}\left(\arctan 2 - \dfrac{3}{20}\right) - \dfrac{1}{8}}.$$

분자와 분모에 8π를 곱하면

$$\boxed{\mathrm{P}(A \mid D) = \frac{\pi - 2}{8\arctan 2 - \frac{6}{5} - \pi} = \frac{5(\pi - 2)}{40\arctan 2 - 6 - 5\pi} \approx 0.253.}$$

따라서 5년째에 폐차된 차 가운데 이미 녹슨 상태로 네 번째 검사를 빠져나간 차는 약 4분의 1에 불과하고, 대부분은 다섯 번째 해 동안 녹슨 것이다.

검산

- 분포함수(distribution function)라면 그래야 하듯이 $F(0) = 0$이고 $t \to \infty$일 때 $F(t) \to 1$이다. $\frac{4}{\pi}\int_0^2 t^2/(1+t^2)^2\,\mathrm{d}t$에 심프슨 공식(Simpson's rule)을 적용하면 0.45018이 나와 $F(2)$와 일치한다.
- 차 400 000대에 대한 시뮬레이션(simulation)(녹이 생기기 시작하는 시점을 표본추출하고 매년 검사를 실행)에서 5년째에 불합격할 확률은 0.1789, 조건부 확률은 0.255로 나왔으며, 정확한 값은 각각 0.1797과 0.2528이다.
- 타당성 점검: $\mathrm{P}(D)$는 $\frac{1}{2}F(2) \approx 0.225$(이전의 폐차를 무시했을 때 5년째에 녹슬어 있고 불합격할 확률)보다 작아야 하며, 실제로 그렇다.

흔한 실수

- 정규화를 잊는 것. 밀도함수는 $t^2/(1+t^2)^2$에 **비례**할 뿐이며, 인수 $4/\pi$를 빠뜨리면 이후의 모든 답이 틀어진다.
- "나이 5에 녹슬어 있음"을 확률 $F(2)$인 하나의 경우로 보고 $\frac{1}{2}$을 곱하는 것. 이는 나이 4에 녹슬어 있었고 그 검사에서 폐차된 차들을 무시한다.
- 마지막 부분에서 "나이 5에 불합격함" 대신 "나이 5에 녹슬어 있음"을 조건으로 삼는 것.

STEP 너머

여기서 함수 F는 분포함수이고, 비 $f(t)/(1-F(t))$는 **위험률**(hazard rate), 즉 지금까지 차가 녹슬지 않고 버텼다는 조건에서 순간적으로 녹슬 가능성이다. 이는 생존 분석(survival analysis)과 신뢰성 이론(reliability theory)의 중심 대상이다. 마지막 부분은 필터링(filtering)의 작은 예로, 불완전한 관측(검사)으로부터 숨겨진 상태(녹슬었는지 아닌지)를 추론하는 것이다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 이러한 관점에서 연속분포와 조건부 확률을 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.