풀이

STEP III 1994, 문제 13

확률과 통계 • 지수분포를 따르는 수명

STEP III 1994, 13번 문제

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

step.maths.org/questions/94-s3-q13

핵심 아이디어

모든 것은 지수 분포(exponential distribution)를 따르는 수명의 생존 함수(survival function) $\mathrm{P}(T > t) = \mathrm{e}^{-\lambda t}$에서 나온다. 시각 t_0까지 살아남았다는 조건을 거는 조건부 확률(conditional probability)은 이런 지수함수 두 개를 나누는 것일 뿐이며, 유명한 **무기억성**(memoryless property)이 바로 여기서 나온다. 두 밧줄이 모두 끊어질 때까지의 시간은 서로 독립(independent)인 두 수명의 **최댓값**(maximum)이고, 최댓값이 t 이하인 것은 각 수명이 모두 t 이하인 것과 정확히 같다. 따라서 먼저 분포 함수(distribution function)를 구한 뒤 미분한다. 마지막 부분에서 "둘 다 같은 공연 중에 끊어진다"는 것은 두 밧줄이 같은 구간 $((n-1)h,\, nh]$에서 끊어진다는 뜻이고, 이를 n에 대해 모두 더한다. 이 합은 기하급수(geometric series)이다.

풀이

이미 사용 중인 밧줄

밧줄의 전체 수명을 T라 하면, $t \geq 0$에 대하여

$$\mathrm{P}(T > t) = \int_t^\infty \lambda \mathrm{e}^{-\lambda u}\, \mathrm{d}u = \mathrm{e}^{-\lambda t}.$$

밧줄이 t_0시간 동안 버텼다는 것, 즉 $T > t_0$임을 알고 있다. 앞으로 더 버티는 시간을 $S = T - t_0$이라 하자. $s \geq 0$에 대하여 사건 $\{T > t_0 + s\}$는 $\{T > t_0\}$에 포함되므로

$$\mathrm{P}(S > s \mid T > t_0) = \frac{\mathrm{P}(T > t_0 + s)}{\mathrm{P}(T > t_0)} = \frac{\mathrm{e}^{-\lambda(t_0+s)}}{\mathrm{e}^{-\lambda t_0}} = \mathrm{e}^{-\lambda s}.$$

따라서 남은 시간 S의 확률 밀도 함수(probability density function)는 $s \geq 0$일 때 $\lambda \mathrm{e}^{-\lambda s}$이다($s < 0$일 때는 0). 즉 S는 t_0의 값에 관계없이 **같은** 모수 λ를 갖는 지수 분포를 따른다.

해석. 사용한 밧줄은 새 밧줄과 정확히 똑같이 좋다. 밧줄은 과거에 사용된 것을 "기억하지" 않으며, 다음의 짧은 구간 δt 동안 끊어질 확률은 나이에 관계없이 약 $\lambda\,\delta t$이다(일정한 고장률(failure rate)). 이는 갑작스러운 무작위 사고로 생기는 고장에 잘 맞는다. 하지만 실제 밧줄은 해지고 닳으므로 이 모형은 현실적이지 않으며, 오래된 밧줄일수록 곧 끊어질 가능성이 더 커야 한다. 나이가 들수록 고장률이 증가하는 모형이 더 낫다. 이 모형에서는 아무 차이가 없지만, 새 밧줄을 고집하는 곡예사의 태도는 실제로는 현명하다.

두 새 밧줄이 모두 끊어질 때까지의 시간

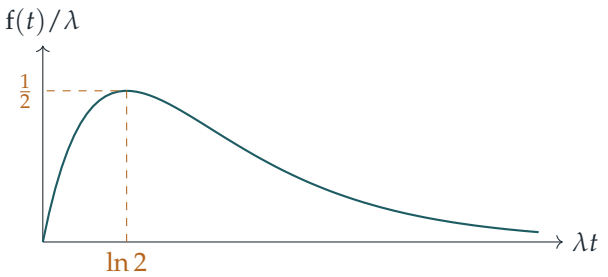
두 새 밧줄의 수명을 T_1, T_2라 하자. 이들은 서로 독립이고, 각각 $t \geq 0$에서 분포 함수 $\mathrm{P}(T_i \leq t) = 1 - \mathrm{e}^{-\lambda t}$를 갖는다. 시각 t까지 둘 다 끊어졌다는 것은 나중에 일어나는 고장이 이미 일어났다는 것과 정확히 같으므로, 구하는 시간은 $M = \max(T_1, T_2)$이다. 이제 $M \leq t$일 필요충분조건은 $T_1 \leq t$ 이고 $T_2 \leq t$인 것이므로, 독립성에 의해

$$\mathrm{P}(M \leq t) = \mathrm{P}(T_1 \leq t)\,\mathrm{P}(T_2 \leq t) = (1 - \mathrm{e}^{-\lambda t})^2 \qquad (t \geq 0),$$

이고 $t < 0$일 때 $\mathrm{P}(M \leq t) = 0$이다. 연쇄 법칙(chain rule)으로 미분하면

$$\mathrm{f}(t) = 2(1 - \mathrm{e}^{-\lambda t}) \cdot \lambda \mathrm{e}^{-\lambda t} = 2\lambda \mathrm{e}^{-\lambda t}(1 - \mathrm{e}^{-\lambda t}) \quad (t \geq 0),$$

이고 $t < 0$일 때 $\mathrm{f}(t) = 0$이므로, 요구된 결과를 얻는다.



하나의 지수 분포 수명과 달리, M의 밀도는 $t = 0$에서 0이다. 시작하자마자 두 고장이 한꺼번에 일어날 수는 없기 때문이다. 밀도는 $\mathrm{e}^{-\lambda t} = \frac{1}{2}$인 곳에서 최대가 된다.

두 밧줄이 모두 n번째 공연 중에 끊어질 확률

n번째 공연은 공연 시간의 구간 $((n-1)h,\, nh]$이다. $q = \mathrm{e}^{-\lambda h}$로 놓자. 밧줄 하나에 대하여

$$\mathrm{P}((n-1)h < T_i \leq nh) = \mathrm{e}^{-\lambda(n-1)h} - \mathrm{e}^{-\lambda nh} = q^{n-1}(1 - q).$$

생각해 보기 $\mathrm{f}(t)$를 n번째 공연 구간에서 적분해야 할까?

아니다. $\int_{(n-1)h}^{nh} \mathrm{f}(t)\,\mathrm{d}t$는 **두 번째** 고장이 n번째 공연에서 일어날 확률이다. 첫 번째 밧줄은 그 이전 공연에서 끊어졌을 수도 있다. 우리에게 필요한 것은 **각** 밧줄이 n번째 공연에서 끊어지는 것이고, 두 밧줄은 독립이므로

$$p_n = \left(q^{n-1}(1-q)\right)^2 = \mathrm{e}^{-2\lambda(n-1)h}(1-\mathrm{e}^{-\lambda h})^2.$$

두 밧줄이 같은 공연 중에 끊어질 확률

$n = 1, 2, 3, \ldots$에 대한 사건 "둘 다 n번째 공연에서 끊어진다"는 서로 배반(mutually exclusive)이므로 확률을 더하면 된다. 이는 공비가 $q^2 < 1$인 기하급수이다($\lambda h > 0$이므로):

$$\sum_{n=1}^{\infty} p_n = (1-q)^2 \sum_{n=1}^{\infty} q^{2(n-1)} = \frac{(1-q)^2}{1-q^2} = \frac{1-q}{1+q}.$$

이제 분자와 분모에 $\mathrm{e}^{\lambda h/2}$를 곱하면

$$\frac{1-\mathrm{e}^{-\lambda h}}{1+\mathrm{e}^{-\lambda h}} = \frac{\mathrm{e}^{\lambda h/2} - \mathrm{e}^{-\lambda h/2}}{\mathrm{e}^{\lambda h/2} + \mathrm{e}^{-\lambda h/2}} = \tanh\left(\frac{\lambda h}{2}\right),$$

이므로 요구된 결과를 얻는다.

검산

- $\int_0^{\infty} \mathrm{f}(t)\,\mathrm{d}t = \left[(1-\mathrm{e}^{-\lambda t})^2\right]_0^{\infty} = 1$이므로 f는 제대로 된 확률 밀도 함수이다.
- 극한: $h \to \infty$이면(끝없이 이어지는 하나의 공연) 답 $\tanh(\lambda h/2) \to 1$이다. $h \to 0$이면 답은 0으로 가는데, 두 연속 수명이 아주 작은 공통 구간에 함께 들어가는 일은 거의 없기 때문이다. h가 작을 때 답은 약 $\lambda h/2$이다.
- 수치로 확인하면, $\lambda = 0.7$, $h = 0.8$일 때 급수와 $\tanh(0.28)$ 모두 0.27291을 준다. 밧줄 2×10^5쌍으로 한 모의실험에서는 0.2746을 얻었고, $p_1 \approx 0.1839$는 모의실험 값 0.1848과 일치했다.

흔한 실수

- 최댓값 대신 **최솟값**(minimum)을 사용하는 것. 최솟값은 **첫 번째** 밧줄이 끊어질 때까지의 시간이며, 모수가 2λ인 지수 분포를 따른다.
- p_n을 구하려고 f를 n번째 공연 구간에서 적분하는 것. 그렇게 구한 확률들을 모두 더하면 $\tanh(\lambda h/2)$가 아니라 1이 된다.

STEP 너머

무기억성은 지수 분포를 특징짓는다. 즉 이 성질을 갖는 연속 분포는 지수 분포뿐이다. 신뢰성 이론(reliability theory)에서 "고장률", 즉 위험 함수(hazard function) $\mathrm{f}(t)/\mathrm{P}(T > t)$가 상수인 것은 지수 분포뿐이다. 마모는 증가하는 위험 함수로 모형화하며, 예를 들어 와이블 분포(Weibull distribution)를 쓴다. 독립인 확률변수들의 최댓값과 최솟값은 **순서통계량**(order statistics)의 양 극단에 해당하며, 대학의 첫 확률론 강의에서 만나게 된다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)이 이 내용을 모두 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.