풀이

STEP II 2014, 문제 12

확률과 통계 • 위험 함수

> **STEP II 2014, 12번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/14-s2-q12

핵심 아이디어

위험 함수(hazard function)는 아직 살아 있는 개체에 대한 사망률이다. 뒤따르는 소문항은 모두 한 가지 관찰에서 나온다. $F'(t) = f(t)$이므로

$$h(t) = \frac{f(t)}{1 - F(t)} = -\frac{\mathrm{d}}{\mathrm{d}t} \ln(1 - F(t)).$$

따라서 h를 안다는 것은 $\ln(1 - F)$의 도함수를 안다는 뜻이다. 이를 적분하고, 수명이 시작되는 곳에서 $F = 0$이라는 사실을 이용한 다음, F를 미분하여 f를 얻는다.

풀이

소문항 (i)

A를 사건 $T \geq t$(파리가 시각 t에 살아 있음), B를 사건 $t \leq T \leq t + \delta t$(파리가 다음 δt 동안 죽음)라 하자. $B \cap A = B$이므로(연속확률변수(continuous random variable)에서는 $\mathrm{P}(T = t) = 0$이므로 끝점의 포함 여부는 상관없다)

$$\mathrm{P}(B \mid A) = \frac{\mathrm{P}(B)}{\mathrm{P}(A)} = \frac{F(t + \delta t) - F(t)}{1 - F(t)}.$$

δt가 작을 때 $F(t + \delta t) - F(t) \approx F'(t)\,\delta t = f(t)\,\delta t$이므로

$$\mathrm{P}(B \mid A) \approx \frac{f(t)\,\delta t}{1 - F(t)} = h(t)\,\delta t,$$

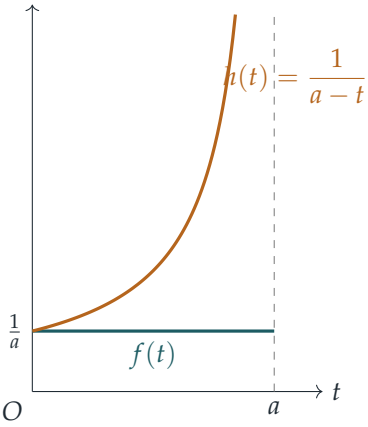
이고, 이것이 보이려던 것이다. (불가능한 사건을 조건으로 삼지 않으려면 $F(t) < 1$이어야 하며, 이것이 바로 h가 정의되는 범위이다.)

소문항 (ii)

여기서는 $0 < t < a$에서 $f(t) = F'(t) = \frac{1}{a}$이다. 즉 T는 $(0, a)$ 위의 균등분포(uniform distribution)를 따른다. 그러면

$$h(t) = \frac{1/a}{1 - t/a} = \frac{1}{a - t}, \qquad 0 < t < a.$$

밀도함수는 평평하지만, 위험 함수는 $h(0) = \frac{1}{a}$에서 시작하여 증가하고, $t \to a^-$일 때 $h(t) \to \infty$이다. 이는 자연스럽다. 거의 시각 a까지 살아남은 파리는 남은 아주 짧은 시간 안에 반드시 죽는다.



(그림은 $a = 2$로 그린 것이다.)

소문항 (iii)

핵심 아이디어에 따라 $t > a$에서 $-\frac{\mathrm{d}}{\mathrm{d}t}\ln(1 - F(t)) = \frac{1}{t}$이다. a부터 t까지 적분하고, T는 a보다 큰 값만 가지므로 $F(a) = 0$임을 이용하면

$$-\ln(1 - F(t)) = \ln t - \ln a = \ln\frac{t}{a}, \qquad \text{따라서} \qquad 1 - F(t) = \frac{a}{t}.$$

그러므로 $F(t) = 1 - \frac{a}{t}$이고

$$f(t) = \frac{a}{t^2}, \qquad t > a$$

이다($t \le a$에서는 $f(t) = 0$). 검산하면 $\int_a^\infty at^{-2}\,\mathrm{d}t = 1$이고, $f/(1 - F) = (a/t^2)/(a/t) = 1/t$이다.

소문항 (iv)

"충분성". $k > 0$에 대하여 $t > b$에서 $f(t) = k\mathrm{e}^{-k(t-b)}$라 하자. $\int_b^\infty k\mathrm{e}^{-k(t-b)}\,\mathrm{d}t = 1$이므로 모든 확률이 $t > b$에 놓여 있고, 따라서 $t \le b$에서 $f(t) = 0$, $F(t) = 0$이다. 그러므로

$t \le b$에서 $h(t) = 0/1 = 0$이다. $t > b$에서는

$$F(t) = \int_b^t k\mathrm{e}^{-k(s-b)}\,\mathrm{d}s = 1 - \mathrm{e}^{-k(t-b)}, \qquad h(t) = \frac{k\mathrm{e}^{-k(t-b)}}{\mathrm{e}^{-k(t-b)}} = k,$$

이고, 이는 상수이다.

"필요성". $t \le b$에서 $h(t) = 0$이고 $t > b$에서 $h(t) = k$(상수)라 하자.

> ? **생각해 보기** $t \le b$에서 $h(t) = 0$이라는 사실은 $F(b)$에 대해 무엇을 알려 주는가?

$h(t) = 0$인 곳에서는 $f(t) = h(t)(1 - F(t)) = 0$이다. 따라서 모든 $t \le b$에서 f가 0이 되고, $t \le b$에서 $F(t) = \int_{-\infty}^t f = 0$이다. 특히 $F(b) = 0$이다. (이로써 그 범위에서 $F < 1$임도 확인되므로, h는 실제로 $t \le b$에서 정의된다.)

$t > b$에서는 $-\dfrac{\mathrm{d}}{\mathrm{d}t}\ln(1 - F(t)) = k$이다. b부터 t까지 적분하고 $F(b) = 0$을 이용하면

$$-\ln(1 - F(t)) = k(t - b), \qquad F(t) = 1 - \mathrm{e}^{-k(t-b)},$$

이고, 미분하면 $t > b$에서 $f(t) = k\mathrm{e}^{-k(t-b)}$이다. 끝으로 k는 양수여야 한다. $f \ge 0$이므로 $k \ge 0$이고, 만약 $k = 0$이면 F가 영원히 0에 머물러 $F(t) \not\to 1$이 되므로 T는 제대로 된 확률변수가 아니다. 이로써 양방향이 모두 증명되었다.

소문항 (v)

이번에도 $t > 0$에서 $-\dfrac{\mathrm{d}}{\mathrm{d}t}\ln(1 - F(t)) = \dfrac{\lambda}{\theta^\lambda}t^{\lambda-1}$이다. $T > 0$이므로 $F(0) = 0$임을 이용하여 0부터 t까지 적분하고, $\lambda > 0$이므로 $t \to 0$일 때 $t^\lambda \to 0$임에 주의하면

$$-\ln(1 - F(t)) = \frac{\lambda}{\theta^\lambda} \cdot \frac{t^\lambda}{\lambda} = \left(\frac{t}{\theta}\right)^\lambda.$$

따라서 $F(t) = 1 - \exp(-(t/\theta)^\lambda)$이고, 미분하면

$$f(t) = \frac{\lambda}{\theta^\lambda}\,t^{\lambda-1}\exp\left(-\left(\frac{t}{\theta}\right)^\lambda\right), \qquad t > 0.$$

이는 제대로 된 밀도함수이다. 양수이고, $t \to \infty$일 때 $F(t) \to 1$이기 때문이다.

검산

- 각 답은 $f/(1 - F)$를 직접 계산하여 확인할 수 있다. 예를 들어 (v)에서는 지수 인자가 약분되어 $\lambda\theta^{-\lambda}t^{\lambda-1}$이 남는다.
- 소문항 (v)에서 $\lambda = 1$로 두면 상수 위험 함수 $1/\theta$와 $f(t) = \frac{1}{\theta}\mathrm{e}^{-t/\theta}$를 얻는데, 이는 소문항 (iv)에서 $b = 0, k = 1/\theta$인 경우이다.
- 수치적으로, $\lambda = \frac{1}{2}$, $\theta = 2$일 때 f를 $(0, \infty)$에서 적분하면 1.0000이 되고, $F(1) = 1 - \mathrm{e}^{-1/\sqrt{2}} \approx 0.5069$는 f를 0부터 1까지 직접 적분한 값과 일치한다.

흔한 실수

- 수명이 시작되는 곳에서 $F = 0$이라는 조건을 잊고, F에 임의의 적분상수(constant of integration)를 남겨 두는 실수.
- (iv)에서 "필요충분조건(if and only if)"의 한 방향만 증명하거나, $t \leq b$에서 $h = 0$임을 이용하여 $F(b) = 0$을 이끌어 내지 않는 실수.
- 음의 부호를 빠뜨리는 실수: $\frac{\mathrm{d}}{\mathrm{d}t} \ln(1 - F) = -f/(1 - F)$.

STEP 너머

관계식 $1 - F(t) = \exp\left(-\int_0^t h(s)\,\mathrm{d}s\right)$는 **생존 분석**(survival analysis)과 신뢰성 이론(reliability theory)의 기초이다. (v)의 분포는 와이블 분포(Weibull distribution)이다. $\lambda < 1$이면 감소하는 위험 함수(초기 고장), $\lambda = 1$이면 무기억성(memoryless)을 가진 지수분포, $\lambda > 1$이면 증가하는 위험 함수(마모)를 얻는다. 생존 분석은 의학 임상시험에서 공학에 이르기까지 어디에서나 쓰인다. 바탕이 되는 확률론에 대해서는 Blitzstein과 Hwang의 *Introduction to Probability*(확률론 입문)를 참고하라.

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.