풀이

STEP II 1998, 문제 13

확률과 통계 • 지수분포

> **STEP II 1998, 13번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/98-s2-q13

핵심 아이디어

첫 번째 결과는 지수분포(exponential distribution)의 **무기억성**(*memoryless property*)이다. 어떤 서비스가 이미 시간 t 동안 계속되었다면, 남은 시간은 막 새로 시작한 서비스와 같은 분포를 따른다. 이 성질을 얻고 나면 가게 이야기는 거의 계산할 필요가 없다. 서비스를 받는 도중인 손님은 확률의 관점에서 새로 서비스를 시작하는 손님과 다를 바가 없다. 따라서 중요한 모든 순간에, 아직 서비스를 받고 있는 사람들 사이의 경주는 서로 같은 분포를 따르는 독립(independent)인 지수 시간들 사이의 경주이다.

풀이

무기억성

$x \geq 0$에 대하여

$$\mathrm{P}(X > x) = \int_x^\infty \lambda \mathrm{e}^{-\lambda u}\,\mathrm{d}u = \Big[-\mathrm{e}^{-\lambda u}\Big]_x^\infty = \mathrm{e}^{-\lambda x}.$$

$s, t \geq 0$을 택하자. 사건 $\{X > s+t\}$는 $\{X > t\}$에 포함되므로, 두 사건의 교집합(intersection)은 그냥 $\{X > s+t\}$이다. 조건부 확률(conditional probability)의 정의에 의하여

$$\begin{aligned}\mathrm{P}(X > s+t \mid X > t) &= \frac{\mathrm{P}(X > s+t \text{ 이고} X > t)}{\mathrm{P}(X > t)} = \frac{\mathrm{P}(X > s+t)}{\mathrm{P}(X > t)} \\ &= \frac{\mathrm{e}^{-\lambda(s+t)}}{\mathrm{e}^{-\lambda t}} = \mathrm{e}^{-\lambda s} = \mathrm{P}(X > s),\end{aligned}$$

이고, 이것이 구하려던 결과이다. ($\mathrm{P}(X > t) = \mathrm{e}^{-\lambda t} > 0$이므로 나눌 수 있었다.)

두 손님이 모두 아직 서비스를 받고 있을 확률

모형: 두 서비스 시간은 독립이고 각각 주어진 지수분포를 따른다. 내가 도착했을 때 서비스를 받고 있는 두 손님을 A와 B라 하자.

> **생각해 보기** 손님 A는 이미 알 수 없는 시간 동안 서비스를 받아 왔다. 그렇다면 앞으로 얼마나 더 걸리는지의 분포가 달라지지 않을까?

달라지지 않으며, 바로 이것이 첫 번째 부분이 말해 주는 내용이다. A의 서비스가 지금까지 u 동안 계속되었다고 하자. 그러면 적어도 t만큼 더 계속될 확률은

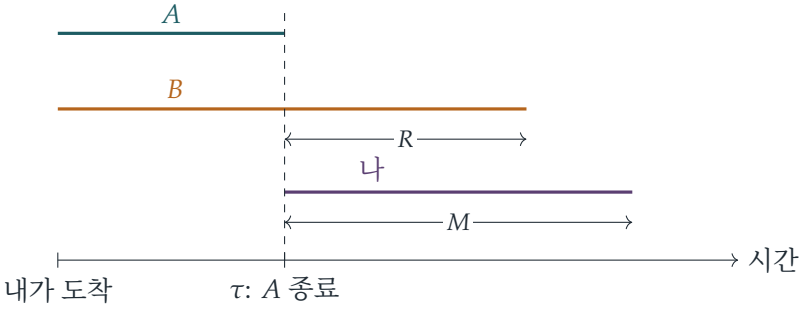
$$\mathrm{P}(X > u + t \mid X > u) = \mathrm{P}(X > t) = \mathrm{e}^{-\lambda t},$$

이고, 이는 u의 값과 관계없다. 따라서 내가 들어온 순간부터 잰 A의 **남은** 서비스 시간은 모수(parameter)가 λ인 같은 지수분포를 따르며, B도 마찬가지이다. 전체 서비스 시간이 독립이므로 두 남은 시간도 독립이다. 그러므로

$$\mathrm{P}(\text{두 손님 모두 아직 서비스 중, 시각 } t) = \mathrm{e}^{-\lambda t} \cdot \mathrm{e}^{-\lambda t} = \mathrm{e}^{-2\lambda t}.$$

다른 손님이 끝날 때 나는 아직 서비스를 받고 있을까?

A가 내가 도착한 뒤 시간 τ에 먼저 끝나고, 내가 곧바로 A를 맡던 점원에게 간다고 하자.



시각 τ에서 B에게 아직 필요한 시간을 R, 나의 서비스 시간을 M이라 하자.

- 나의 서비스는 τ에서 새로 시작하므로 M은 모수가 λ인 지수분포를 따른다. 서로 다른 손님의 서비스 시간은 독립이므로, M은 그 이전의 모든 것과 독립이다.
- R도 모수가 λ인 지수분포를 따른다. 시각 τ에서 B는 이미 어떤 시간(내가 도착하기 전의 시간에 τ를 더한 것) 동안 서비스를 받았고, 다시 무기억성에 의하여 남은 시간은 원래의 분포를 따른다. 또 R은 M과 독립이다.

B가 끝날 때 내가 아직 서비스를 받고 있는 것은 정확히 $M > R$일 때이다. 그런데 M과 R은 독립이고 **같은** 연속분포를 따르므로, 대칭성(symmetry)에 의하여 $\mathrm{P}(M > R) = \mathrm{P}(R >$

M)이다. 또한 연속확률변수(continuous random variable)에 대해서는 $\mathrm{P}(M = R) = 0$이므로, 이 두 확률의 합은 1이다. 그러므로

$$\mathrm{P}(\text{다른 손님이 끝날 때 내가 아직 서비스를 받고 있음}) = \tfrac{1}{2}.$$

대칭성 논증보다 계산이 좋다면 R에 대한 조건을 걸면 된다.

$$\mathrm{P}(M > R) = \int_0^\infty \lambda \mathrm{e}^{-\lambda r} \, \mathrm{P}(M > r) \, \mathrm{d}r = \int_0^\infty \lambda \mathrm{e}^{-2\lambda r} \, \mathrm{d}r = \tfrac{1}{2}.$$

검산

- $t = 0$일 때 $\mathrm{e}^{-2\lambda t} = 1$이다. 내가 도착했을 때 두 손님은 확실히 서비스를 받고 있다. $t \to \infty$이면 확률은 0으로 수렴하는데, 당연히 그래야 한다.
- 답 $\frac{1}{2}$는 λ에 의존하지 않는다. λ를 바꾸는 것은 시간의 단위를 바꾸는 것일 뿐이므로 예상대로이다.
- $\lambda = 0.7$, $t = 0.9$로 한 시뮬레이션(simulation)에서 두 손님이 모두 아직 서비스를 받고 있는 비율은 0.2831로, $\mathrm{e}^{-1.26} \approx 0.2837$과 비교된다. 두 손님이 이미 무작위 시간 동안 서비스를 받아 온 상황의 시뮬레이션에서는, 시행 가운데 0.5000의 비율로 내가 아직 서비스를 받고 있었다.

흔한 실수

- 이미 서비스를 받고 있는 손님들은 "곧 끝날 때가 되었으므로" 남은 시간이 더 짧다고 생각하는 것. 지수분포에서는 이미 일어난 일이 아무런 차이를 만들지 않는다.
- $\mathrm{P}(X > s + t \text{ and } X > t)$를 $\mathrm{P}(X > s + t) \, \mathrm{P}(X > t)$로 쓰는 것. 이 두 사건은 독립이 아니며, 한 사건이 다른 사건을 포함한다.

STEP 너머

무기억성을 가지는 연속분포는 지수분포뿐이다. 이 사실은 **포아송 과정**(*Poisson process*)과 **대기행렬 이론**(*queueing theory*)의 바탕이 되며, 대기행렬 이론에서 점원이 c명이고 서비스 시간이 지수분포를 따르는 가게는 고전적인 M/M/c 대기행렬이다. 두 손님에 대한 계산은 n명으로 확장된다. 모수가 λ인 독립인 지수확률변수 n개의 최솟값은 모수가 $n\lambda$인 지수분포를 따르며(꼬리 확률(tail probability)을 곱하면 된다), 대칭성에 의하여 각 변수가 가장 작을 가능성은 모두 같다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 이 수준에서 지수확률변수와 포아송 과정을 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.