풀이

STEP I 2010, 문제 13

확률과 통계 • 푸아송 대기 시간

> **STEP I 2010, 13번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/10-s1-q13

핵심 아이디어

"첫 문자가 1시간에서 2시간 사이에 도착한다"는 것은 "처음 1시간 동안에는 문자가 없지만, 처음 2시간 동안 문자가 하나도 없는 것은 아니다"와 같다. 이 두 사건은 각각 사건이 한 번도 일어나지 않을 푸아송(Poisson) 확률이고, 이는 지수함수(exponential) 하나로 나타난다. 여기서 $x = \mathrm{e}^{\lambda}$로 치환하면 모든 것이 이차방정식(quadratic)으로 바뀌고, 나머지 일은 비에트 정리(Vieta's formulas, 근과 계수의 관계: 두 근의 합과 곱)가 해 준다. 밀드러드의 경우도 마찬가지이다.

풀이

일반적인 푸아송 모형을 사용한다. 즉 길이가 t시간인 임의의 구간에서 조지가 받는 문자의 개수는 평균이 λt인 푸아송 분포를 따르고, 겹치지 않는 구간들에서의 개수는 서로 독립(independent)이다. 시간은 아침이 시작되는 순간부터 잰다. 시간은 연속적이므로 첫 문자가 정확히 1시간 또는 정확히 2시간에 도착할 확률은 0이다. 따라서 구간의 끝점을 포함하든 포함하지 않든 상관없다.

조지의 방정식

조지가 첫 문자를 받을 때까지 걸리는 시간을 시간 단위로 T라 하자. 그러면 $T > t$인 것은 정확히 $[0, t]$에서 문자가 하나도 오지 않는 경우이므로

$$\mathrm{P}(T > t) = \mathrm{P}(\mathrm{Poisson}(\lambda t) = 0) = \mathrm{e}^{-\lambda t}.$$

사건 $\{1 < T \leq 2\}$는 $\{T > 1\}$에서 $\{T > 2\}$를 뺀 것이고, $\{T > 2\} \subseteq \{T > 1\}$이므로

$$p = \mathrm{P}(1 < T \leq 2) = \mathrm{P}(T > 1) - \mathrm{P}(T > 2) = \mathrm{e}^{-\lambda} - \mathrm{e}^{-2\lambda}.$$

양변에 $e^{2\lambda}$를 곱하면 $pe^{2\lambda} = e^{\lambda} - 1$, 즉

$$pe^{2\lambda} - e^{\lambda} + 1 = 0,$$

이 되어 요구된 식을 얻는다.

λ의 두 양수 값

$x = e^{\lambda}$로 놓자. $\lambda \mapsto e^{\lambda}$는 $(0, \infty)$에서 $(1, \infty)$로 가는 순증가 전단사함수(bijection)이므로, 양수 해 λ는

$$px^2 - x + 1 = 0.$$

의 해 중 $x > 1$인 것과 정확히 일대일로 대응한다. 여기서 $p > 0$이다(예를 들어 $[1, 2]$에서 문자가 정확히 하나 오고 그 전에는 하나도 오지 않으면 항상 일어나는 사건의 확률이기 때문이다). 따라서 이 식은 실제로 이차방정식이다.

> **생각해 보기** 두 근이 실수임을 보이는 것은 쉽다. 두 근이 모두 1보다 크다는 것은 어떻게 알 수 있을까?

서로 다른 두 실근. $4p < 1$이므로 판별식(discriminant)은 $1 - 4p > 0$이다. 따라서 서로 다른 두 실근 $x_1 < x_2$가 존재한다.

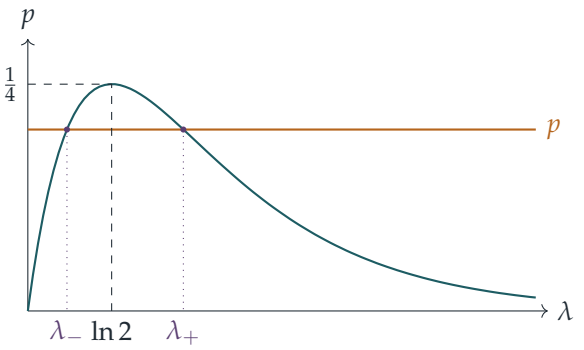
두 근 모두 1보다 크다. 비에트 정리에 의해 $x_1 + x_2 = \frac{1}{p}$이고 $x_1 x_2 = \frac{1}{p}$이다. 그러면

$$(x_1 - 1)(x_2 - 1) = x_1 x_2 - (x_1 + x_2) + 1 = \frac{1}{p} - \frac{1}{p} + 1 = 1 > 0,$$

이므로 $x_1 - 1$과 $x_2 - 1$은 부호가 같다. 또한 $p < \frac{1}{4}$에서 $\frac{1}{p} > 4$이므로 두 수의 합은 $x_1 + x_2 - 2 = \frac{1}{p} - 2 > 0$이다. 따라서 둘 다 양수이다. 즉 $x_1 > 1$이고 $x_2 > 1$이다.

그러므로 $\lambda_{\pm} = \ln\left(\dfrac{1 \pm \sqrt{1 - 4p}}{2p}\right)$는 방정식을 만족하는 서로 다른 두 양수 λ 값이다.

그림을 보면 이 결과가 이해된다. λ의 함수로서 $p = e^{-\lambda} - e^{-2\lambda}$는 $\lambda = 0$에서 0이고, $\lambda \to \infty$일 때 0으로 수렴(tends to)하며, ($u = e^{-\lambda}$로 놓으면 $p = u - u^2$이므로) $u = \frac{1}{2}$, 즉 $\lambda = \ln 2$에서 최댓값 $\frac{1}{4}$를 갖는다. 높이가 $p < \frac{1}{4}$인 수평선은 이 그래프와 두 번 만난다. 한산한 전화(문자가 거의 오지 않음)와 바쁜 전화(문자가 대개 첫 1시간 안에 옴)가 같은 p를 줄 수 있는 것이다.



밀드러드: $\lambda_1 + \lambda_2$

각 전화를 따로 보면 조지의 상황과 정확히 같으므로, e^{λ_1}과 e^{λ_2}는 모두 $px^2 - x + 1 = 0$을 만족한다. 이 이차방정식의 근은 두 개뿐이고 $\lambda_1 \neq \lambda_2$이므로, e^{λ_1}과 e^{λ_2}는 바로 그 두 근 x_1과 x_2이다. (이를 위해서는 $p < \frac{1}{4}$가 필요한데, 서로 다른 두 값이 존재하므로 이 조건은 성립한다. $p = \frac{1}{4}$이면 중근이 된다.) 두 근의 곱에 의해

$$e^{\lambda_1+\lambda_2} = e^{\lambda_1}e^{\lambda_2} = x_1x_2 = \frac{1}{p}, \qquad \text{따라서} \qquad \lambda_1 + \lambda_2 = \ln\frac{1}{p} = -\ln p.$$

밀드러드의 첫 문자

? **생각해 보기** 밀드러드가 두 전화로 받는 문자의 총 개수는 어떤 분포를 따를까?

두 개수는 서로 독립인 푸아송 확률변수이므로, 그 합은 평균이 두 평균의 합인 푸아송 분포를 따른다. 즉 t시간 동안 밀드러드가 받는 문자의 총 개수는 $\text{Poisson}((\lambda_1 + \lambda_2)t)$를 따른다. 같은 내용을 다르게 표현하면 다음과 같다. 두 전화의 대기 시간(waiting time)을 T_1, T_2라 하면 이들은 서로 독립이고, 첫 문자는 $T = \min(T_1, T_2)$에 오며

$$\mathrm{P}(T > t) = \mathrm{P}(T_1 > t)\,\mathrm{P}(T_2 > t) = e^{-(\lambda_1+\lambda_2)t}.$$

어느 방법이든 $\mu = \lambda_1 + \lambda_2$로 놓으면 $e^{-\mu} = p$이고, 조지의 경우와 같은 논증에 의해

$$\mathrm{P}(1 < T \leq 2) = e^{-\mu} - e^{-2\mu} = p - p^2 = p(1-p).$$

검산

- $\lambda = 0.8$이면 $p = e^{-0.8} - e^{-1.6} \approx 0.2474$이다. 지수분포를 따르는 대기 시간 2×10^5개를 시뮬레이션(simulation)하면 0.2462를 얻었다.
- $p = 0.2$이면 근은 $\lambda_1 \approx 0.3235$와 $\lambda_2 \approx 1.2859$로 둘 다 양수이고, 그 합은 $1.6094 =$

$\ln 5 = -\ln 0.2$이다. $\min(T_1, T_2)$를 시뮬레이션하면 마지막 확률로 0.1601을 얻었으며, 이는 $p - p^2 = 0.16$과 잘 맞는다.

- 타당성 확인: 예상대로 $p(1-p) < p$이다. 전화가 두 대이면 첫 문자가 더 빨리 오는 경향이 있어서 "1시간 동안 문자가 없음"의 가능성이 더 작아지기 때문이다.

흔한 실수

- x에 대한 근이 실수이고 양수임을 보이고 나서, $\lambda > 0$이려면 $x > 0$이 아니라 $x > 1$이어야 한다는 것을 잊는 경우.
- 마지막 문항에서 두 전화의 확률을 더하거나 곱하는(p^2) 경우. 올바른 사건은 "두 전화 중 어느 쪽이든 첫 문자가 오는 것"이고, 이는 합쳐진 비율이 결정한다.

STEP 너머

푸아송 과정(Poisson process)에서 첫 사건이 일어날 때까지의 대기 시간은 지수분포를 따른다. 여기서 사용한 병합(merging) 성질(비율이 λ_1, λ_2인 서로 독립인 두 푸아송 과정을 합치면 비율이 $\lambda_1 + \lambda_2$인 하나의 푸아송 과정이 된다)은 "솎아내기(thinning)"(각 사건을 독립적으로 확률 q로 남기는 것)와 함께 푸아송 과정의 기본 성질 중 하나이다. 이러한 아이디어는 대기행렬 이론(queueing theory)과 응용 확률론의 상당 부분을 떠받치고 있다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 이 문제에서 자연스럽게 이어지는 방식으로 푸아송 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.