풀이

STEP III 2021, 문제 11

확률과 통계 • 지수분포

> **STEP III 2021, 11번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 OCR에서 내려받을 수 있는 2021년 STEP 시험지 묶음(zip 파일) 가운데 STEP III 시험지에 있다:
>
> www.ocr.org.uk/Images/696588-2021-past-papers.zip

이 문제는 *Beyond STEP*의 12.18 절 **무기억성**에서 단계별 안내와 함께 자세히 다룬다: https://beyondstep.keithmcnulty.org/ko/ch12/memoryless.html

핵심 아이디어

Y와 Z에 관한 모든 질문은 사실 X가 어느 구간에 떨어지는가에 관한 질문이다. 사건 $Y = n$은 $n \le X < n+1$이고, 사건 $Z < z$는 각 n마다 하나씩, 서로 배반인(disjoint) 조각 $n \le X < n+z$들로 나뉜다. 지수분포(exponential distribution)의 꼬리 확률(tail) $\mathrm{P}(X \ge x) = \mathrm{e}^{-\lambda x}$ 덕분에 각 조각은 등비급수(geometric series)의 한 항이 된다.

풀이

이 풀이 전체에서 $q = \mathrm{e}^{-\lambda}$로 쓴다. $\lambda > 0$이므로 $0 < q < 1$이다. $x \ge 0$에 대하여

$$\mathrm{P}(X \ge x) = \int_x^\infty \lambda \mathrm{e}^{-\lambda t}\,\mathrm{d}t = \mathrm{e}^{-\lambda x},$$

이고, X는 연속(continuous) 확률변수이므로 $0 \le a < b$에 대하여 $\mathrm{P}(a \le X < b) = \mathrm{P}(a < X < b) = \mathrm{e}^{-\lambda a} - \mathrm{e}^{-\lambda b}$이다.

소문항 (i)

Y는 X의 정수 부분(integer part)이므로, $Y = n$인 것은 정확히 $n \le X < n+1$일 때이다. 따라서

$$\mathrm{P}(Y = n) = \mathrm{e}^{-n\lambda} - \mathrm{e}^{-(n+1)\lambda} = (1 - \mathrm{e}^{-\lambda})\mathrm{e}^{-n\lambda},$$

이고, 이것이 구하는 결과이다. (즉 Y는 0에서 시작하고 공비가 q인 기하분포(geometric distribution)를 따른다.)

소문항 (ii)

$Z = X - Y$는 X의 소수 부분(fractional part)이므로 $0 \le Z < 1$이다. $0 \le z \le 1$이라 하자. 사건 $Z < z$는 정확히 정수 $n = Y$에 대하여 $n \le X < n + z$일 때 일어난다. $n = 0, 1, 2, \ldots$에 대한 이 사건들은 서로 배반이므로

$$\begin{aligned} \mathrm{P}(Z < z) &= \sum_{n=0}^{\infty} \mathrm{P}(n \le X < n + z) = \sum_{n=0}^{\infty} \left(\mathrm{e}^{-n\lambda} - \mathrm{e}^{-n\lambda - \lambda z}\right) \\ &= (1 - \mathrm{e}^{-\lambda z}) \sum_{n=0}^{\infty} q^n = \frac{1 - \mathrm{e}^{-\lambda z}}{1 - \mathrm{e}^{-\lambda}}, \end{aligned}$$

이다. 여기서 $0 < q < 1$이므로 수렴(converge)하는 등비급수의 합을 구하였다. 이것이 구하는 결과이다.

소문항 (iii)

(ii)의 분포함수(distribution function)를 미분하면 Z의 확률밀도함수(density)는

$$g(z) = \frac{\lambda \mathrm{e}^{-\lambda z}}{1 - \mathrm{e}^{-\lambda}} \qquad (0 \le z \le 1),$$

이고, 그 밖의 경우에는 $g(z) = 0$이다. $u = z$, $\mathrm{d}v = \lambda \mathrm{e}^{-\lambda z}\,\mathrm{d}z$로 놓고 부분적분(integration by parts)하면

$$\int_0^1 z\lambda \mathrm{e}^{-\lambda z}\,\mathrm{d}z = \left[-z\mathrm{e}^{-\lambda z}\right]_0^1 + \int_0^1 \mathrm{e}^{-\lambda z}\,\mathrm{d}z = -\mathrm{e}^{-\lambda} + \frac{1 - \mathrm{e}^{-\lambda}}{\lambda}.$$

이것을 $1 - \mathrm{e}^{-\lambda}$로 나누면

$$\mathrm{E}(Z) = \frac{1}{\lambda} - \frac{\mathrm{e}^{-\lambda}}{1 - \mathrm{e}^{-\lambda}} = \frac{1}{\lambda} - \frac{1}{\mathrm{e}^{\lambda} - 1}.$$

소문항 (iv)

$0 \le z_1 < z_2 \le 1$이라 하자. 사건 $\{Y = n$ 이고$z_1 < Z < z_2\}$는 정확히 $n + z_1 < X < n + z_2$이므로

$$\mathrm{P}(Y = n \text{ 이고} z_1 < Z < z_2) = \mathrm{e}^{-\lambda(n+z_1)} - \mathrm{e}^{-\lambda(n+z_2)} = \mathrm{e}^{-n\lambda}\left(\mathrm{e}^{-\lambda z_1} - \mathrm{e}^{-\lambda z_2}\right).$$

? **생각해 보기** 독립성을 판정하려면 이것을 무엇과 비교해야 할까?

곱 $\mathrm{P}(Y=n)\,\mathrm{P}(z_1<Z<z_2)$와 비교해야 한다. Z는 연속 확률변수이므로 $\mathrm{P}(Z=z_1)=0$이고, 따라서 (ii)에 의하여

$$\mathrm{P}(z_1<Z<z_2)=\mathrm{P}(Z<z_2)-\mathrm{P}(Z<z_1)=\frac{\mathrm{e}^{-\lambda z_1}-\mathrm{e}^{-\lambda z_2}}{1-\mathrm{e}^{-\lambda}}.$$

여기에 (i)의 $\mathrm{P}(Y=n)=(1-\mathrm{e}^{-\lambda})\mathrm{e}^{-n\lambda}$를 곱하면 인수 $1-\mathrm{e}^{-\lambda}$가 약분되어 위의 결합확률(joint probability)과 정확히 같아진다. 따라서 모든 음이 아닌 정수 n과 모든 구간 $(z_1,z_2)\subseteq[0,1]$에 대하여

$$\mathrm{P}(Y=n \text{ 이고} z_1<Z<z_2)=\mathrm{P}(Y=n)\,\mathrm{P}(z_1<Z<z_2)$$

이다. Y는 음이 아닌 정수값만 갖고 Z는 $[0,1)$의 값만 가지므로, 이러한 꼴의 사건들이 결합분포(joint distribution) 전체를 결정한다. 그러므로 Y와 Z는 **독립**(*independent*)**이다**. 즉 X가 어느 단위 구간에 떨어졌는지 안다고 해서 그 구간 안의 어디에 떨어졌는지에 대해서는 아무것도 알 수 없다.

검산

- (ii)에 $z=1$을 대입하면 당연히 그래야 하듯 $\mathrm{P}(Z<1)=1$이 되고, $z=0$을 대입하면 0이 된다.
- $\lambda\to 0$일 때 X의 밀도는 각 단위 구간 위에서 거의 평평하므로, Z는 $[0,1]$ 위의 균등분포(uniform distribution)에 가까워야 한다. 실제로 $\mathrm{e}^{\lambda}-1=\lambda+\frac{1}{2}\lambda^2+\cdots$로부터 $\frac{1}{\mathrm{e}^{\lambda}-1}=\frac{1}{\lambda}-\frac{1}{2}+O(\lambda)$이므로 $\mathrm{E}(Z)\to\frac{1}{2}$이다. $\lambda\to\infty$일 때는 X가 거의 항상 아주 작으므로 $\mathrm{E}(Z)\to 0$이다.
- $\lambda=1$일 때 $\mathrm{E}(Z)=1-\frac{1}{\mathrm{e}-1}\approx 0.4180$이며, 이는 X의 값 200 000개를 컴퓨터로 모의실험한 결과와 일치한다. (iv)의 모의실험 결합 빈도 역시 주변분포(marginal distribution)들의 곱과 일치하였다.

흔한 실수

- $Z<z$가 첫 번째 단위 구간뿐 아니라 모든 단위 구간에서 일어날 수 있다는 점을 잊고 $\mathrm{P}(Z<z)=\mathrm{P}(X<z)$로 쓰는 것.
- n, z_1, z_2를 한 가지만 골라 확인하고 독립이라고 결론 내리는 것. 인수분해(factorisation)는 이들 모두에 대하여 성립해야 한다.

STEP 너머

(iv)의 독립성은 지수분포의 **무기억성**(*memoryless property*)이 드리운 그림자이다. 즉 $X \geq n$이라는 조건 아래에서 초과량 $X - n$은 n이 무엇이든 다시 같은 비율의 지수분포를 따른다. 이 때문에 지수분포는 포아송 과정(Poisson process)의 대기 시간(waiting time)을 모형화하며, 그 정수 부분은 이산 무기억 분포인 기하분포가 된다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 두 분포와 포아송 과정을 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.