풀이

STEP I 2015, 문제 12

확률과 통계 • 푸아송 분포의 솎아내기

> **STEP I 2015, 12번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/15-s1-q12

핵심 아이디어

도착한 인원수를 조건으로 두면 수술이 필요한 사람의 수는 이항분포(binomial distribution)를 따르는 개수가 된다. 그다음 가능한 모든 도착 인원수에 대하여 더하면(전확률 법칙(law of total probability)) 수술이 필요한 사람 수의 분포를 얻고, 이 합은 지수함수의 급수(exponential series)로 정리된다. 마지막 소문항에서 핵심이 되는 사실은 평균이 같은 두 독립인 푸아송(Poisson) 개수를 그 합을 조건으로 두면, 둘 사이의 나뉨이 이항분포를 따른다는 것이다.

풀이

어느 날 도착하는 부상자의 수를 X라 하고, 그중 수술이 필요한 사람의 수를 Y라 하자.

소문항 (i)

$X = n$일 때, n명의 부상자 각각은 다른 사람과 독립적으로 확률 $\frac{1}{4}$로 수술이 필요하다. 따라서 조건부로 $Y \sim \mathrm{B}(n, \frac{1}{4})$이고, $0 \le r \le n$에 대하여

$$\mathrm{P}(Y = r \mid X = n) = \binom{n}{r}\left(\frac{1}{4}\right)^r\left(\frac{3}{4}\right)^{n-r} = \binom{n}{r}\frac{3^{n-r}}{4^n}.$$

이다. ($r > n$이면 확률은 0이다.)

소문항 (ii)

$r \geq 0$을 고정하자. $Y = r$는 $X \geq r$일 때만 일어날 수 있으므로, 전확률 법칙에 의해

$$\begin{aligned} \mathrm{P}(Y = r) &= \sum_{n=r}^{\infty} \mathrm{P}(X = n)\,\mathrm{P}(Y = r \mid X = n) \\ &= \sum_{n=r}^{\infty} \frac{\mathrm{e}^{-8} 8^n}{n!} \cdot \frac{n!}{r!\,(n-r)!} \left(\frac{1}{4}\right)^r \left(\frac{3}{4}\right)^{n-r}. \end{aligned}$$

이다. $n!$은 약분된다. 이제 $8^n = 8^r \cdot 8^{n-r}$로 나누면 $8^r (\frac{1}{4})^r = 2^r$이고 $8^{n-r} (\frac{3}{4})^{n-r} = 6^{n-r}$이므로

$$\begin{aligned} \mathrm{P}(Y = r) &= \frac{\mathrm{e}^{-8} 2^r}{r!} \sum_{n=r}^{\infty} \frac{6^{n-r}}{(n-r)!} \\ &= \frac{\mathrm{e}^{-8} 2^r}{r!} \sum_{k=0}^{\infty} \frac{6^k}{k!} \qquad (k = n - r) \\ &= \frac{\mathrm{e}^{-8} 2^r}{r!}\, \mathrm{e}^6 = \frac{\mathrm{e}^{-2} 2^r}{r!}. \end{aligned}$$

이다. 이는 모든 $r = 0, 1, 2, \ldots$에 대하여 성립하므로, 하루에 수술이 필요한 사람의 수는 평균이 $8 \times \frac{1}{4} = 2$인 푸아송 분포를 따른다.

생각해 보기 e^6은 어디에서 나왔는가?

r에만 의존하는 것을 모두 합 밖으로 꺼내고 나면, 남는 것은 $\sum 6^k / k!$, 곧 e^6의 지수함수 급수이다. 8^n을 $8^r \cdot 8^{n-r}$로 나누는 것이 바로 남은 합이 $n - r$에만 의존하도록 만드는 단계이다.

소문항 (iii)

월요일과 화요일에 수술이 필요한 사람의 수를 각각 M, T라 하자. 소문항 (ii)에 의해 각각은 $\mathrm{Po}(2)$를 따른다. 서로 다른 날의 도착은 독립이고 부상자들은 서로 독립적으로 수술이 필요하므로, M과 T는 독립이다. 그러면

$$\mathrm{P}(M = 8 \mid M + T = 12) = \frac{\mathrm{P}(M = 8,\ T = 4)}{\mathrm{P}(M + T = 12)}.$$

이다. 독립성에 의해 분자는

$$\mathrm{P}(M = 8)\,\mathrm{P}(T = 4) = \frac{\mathrm{e}^{-2} 2^8}{8!} \cdot \frac{\mathrm{e}^{-2} 2^4}{4!} = \frac{\mathrm{e}^{-4}\, 2^{12}}{8!\, 4!}.$$

이다. 분모의 경우, 독립인 푸아송 확률변수들의 합은 평균을 더한 푸아송 분포를 따르므로 $M + T \sim \mathrm{Po}(4)$이다. 이 사실을 인용하고 싶지 않다면 직접 유도할 수 있다.

$$\mathrm{P}(M + T = 12) = \sum_{m=0}^{12} \frac{\mathrm{e}^{-2} 2^m}{m!} \cdot \frac{\mathrm{e}^{-2} 2^{12-m}}{(12-m)!} = \frac{\mathrm{e}^{-4} 2^{12}}{12!} \sum_{m=0}^{12} \binom{12}{m} = \frac{\mathrm{e}^{-4}\, 4^{12}}{12!},$$

여기서 $\sum_m \binom{12}{m} = 2^{12}$을 이용하였다. 나누면

$$\begin{aligned} \mathrm{P}(M = 8 \mid M + T = 12) &= \frac{\mathrm{e}^{-4}\,2^{12}}{8!\,4!} \cdot \frac{12!}{\mathrm{e}^{-4}\,4^{12}} \\ &= \binom{12}{8}\left(\frac{1}{2}\right)^{12} = \frac{495}{4096}. \end{aligned}$$

이다. $495 = 3^2 \cdot 5 \cdot 11$은 홀수이고 $4096 = 2^{12}$이므로 이 분수는 기약분수(lowest terms)이다. 따라서 구하는 확률은 $\dfrac{495}{4096}$이다.

? 생각해 보기 답이 왜 이항분포의 확률처럼 생겼는가?

합이 12라는 조건 아래에서, 두 날의 평균이 같으므로 12건의 수술 사례 각각은 월요일에서 왔을 가능성과 화요일에서 왔을 가능성이 같다. 따라서 월요일의 개수는 $\mathrm{B}(12, \frac{1}{2})$를 따른다. 원래의 평균 8이 완전히 사라졌다는 점에 주목하자.

검산

- 소문항 (ii)의 평균은 $\mathrm{E}(Y) = \mathrm{E}\big(\mathrm{E}(Y \mid X)\big) = \mathrm{E}(X/4) = 2$와 일치한다.
- $r = 0, \ldots, 14$에 대하여 소문항 (ii)의 급수를 수치적으로 더하면 반올림 오차 범위 안에서 $\mathrm{e}^{-2}2^r/r!$이 다시 나온다. 예를 들어 $\mathrm{P}(Y = 0) = \sum_n \mathrm{e}^{-8}8^n(\frac{3}{4})^n/n! = \mathrm{e}^{-8}\mathrm{e}^{6}$이다.
- $m = 0, \ldots, 12$에 대한 조건부 확률 $\binom{12}{m}/2^{12}$의 합은 1이고, $495/4096 \approx 0.121$이다. 전체 모형(푸아송 도착 후 각자 확률 $\frac{1}{4}$로 수술)을 컴퓨터로 시뮬레이션한 결과도 이 값과 부합하였다.

흔한 실수

- 소문항 (ii)의 합을 $n = r$이 아니라 $n = 0$부터 시작하는 것: r명보다 적게 도착하면 r명이 수술을 받아야 할 수는 없다.
- 소문항 (iii)에서 각 날에 대해 $\mathrm{Po}(8)$을 사용하는 것. 조건은 수술 사례에 대한 것이므로 관련된 평균은 2이다. (여기서는 어차피 평균이 약분되지만, 풀이 과정은 올바라야 한다.)
- 기약분수로 나타내라고 했는데 답을 $\binom{12}{8}/2^{12}$이나 $1980/16384$로 남겨 두는 것.

STEP 너머

소문항 (ii)는 **솎아내기**(*thinning*)라고 불린다. 푸아송 과정(Poisson process)의 각 점을 독립적으로 확률 p로 남기면, 남은 점들은 비율이 p배가 된 푸아송 과정을 이루며, 놀랍게도

버려진 점들과 독립이다. 소문항 (iii)은 이것의 역이다. 합을 조건으로 두면 독립인 푸아송 개수들은 다항분포(multinomial distribution)에 따라 나뉜다. 이 두 사실은 푸아송 과정 이론의 토대가 되며, 이는 대학의 첫 확률론 강의에서 만나게 된다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문서)이 둘 다 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.