풀이

STEP I 2001, 문제 13

확률과 통계 • 조건부 푸아송 분포

> **STEP I 2001, 13번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/01-s1-q13

핵심 아이디어

깨진 접시의 총 개수를 조건으로 두면(conditioning) 푸아송(Poisson) 문제가 이항(binomial) 문제로 바뀐다. 네 학생이 같은 비율로 서로 독립적으로 접시를 깬다면, 깨진 접시가 모두 몇 개인지 알고 나면 깨진 접시 하나하나는 다른 접시와 독립적으로 네 학생 중 누구에 의해서든 같은 확률로 깨졌다고 볼 수 있다. 따라서 깨진 접시 중 수학자의 몫은 이항분포를 따르고, 미지수 λ는 사라진다.

풀이

설정하기

수학자가 깬 접시의 개수를 M, 나머지 세 학생이 깬 접시의 총 개수를 R이라 하자. 문제에서 각 학생이 깨는 접시의 개수는 다른 학생과 독립적으로 $\mathrm{Po}(\lambda)$를 따른다고 하였다. 따라서 $M \sim \mathrm{Po}(\lambda)$이고, R는 서로 독립인 세 개의 $\mathrm{Po}(\lambda)$ 확률변수의 합이므로 $\mathrm{Po}(3\lambda)$를 따른다. 또한 R는 나머지 세 학생에게만 의존하므로 M과 R는 독립(independent)이다. 총합은 $T = M + R \sim \mathrm{Po}(4\lambda)$이다.

> ? **생각해 보기** 서로 독립인 푸아송 확률변수의 합은 왜 다시 푸아송분포를 따르는가?

서로 독립인 $X \sim \mathrm{Po}(\alpha)$와 $Y \sim \mathrm{Po}(\beta)$에 대하여, n을 나누는 모든 방법에 대해 합하면

$$\mathrm{P}(X+Y=n) = \sum_{k=0}^{n} \frac{\mathrm{e}^{-\alpha}\alpha^k}{k!} \cdot \frac{\mathrm{e}^{-\beta}\beta^{n-k}}{(n-k)!} = \frac{\mathrm{e}^{-(\alpha+\beta)}}{n!} \sum_{k=0}^{n} \binom{n}{k} \alpha^k \beta^{n-k} = \frac{\mathrm{e}^{-(\alpha+\beta)}(\alpha+\beta)^n}{n!}$$

이다. 마지막 등호는 이항정리(binomial theorem)에 의한 것이다. 따라서 $X + Y \sim$

$\mathrm{Po}(\alpha+\beta)$이다. 이를 두 번 적용하면 $R \sim \mathrm{Po}(3\lambda)$를 얻고, 한 번 더 적용하면 $T \sim \mathrm{Po}(4\lambda)$를 얻는다.

M의 조건부 분포

$m = 0, 1, \ldots, 5$에 대하여 사건 $\{M = m,\ T = 5\}$는 $\{M = m,\ R = 5-m\}$과 같다. M과 R가 독립이므로

$$\begin{aligned}
\mathrm{P}(M = m \mid T = 5) &= \frac{\mathrm{P}(M=m)\,\mathrm{P}(R=5-m)}{\mathrm{P}(T=5)} \\
&= \frac{\dfrac{\mathrm{e}^{-\lambda}\lambda^m}{m!} \cdot \dfrac{\mathrm{e}^{-3\lambda}(3\lambda)^{5-m}}{(5-m)!}}{\dfrac{\mathrm{e}^{-4\lambda}(4\lambda)^5}{5!}} \\
&= \frac{5!}{m!\,(5-m)!} \cdot \frac{3^{5-m}}{4^5} = \binom{5}{m}\left(\tfrac{1}{4}\right)^m \left(\tfrac{3}{4}\right)^{5-m}.
\end{aligned}$$

지수함수들이 약분되고, λ의 거듭제곱도 모두 약분된다(분자와 분모에 모두 λ^5이 있다). 따라서 $T = 5$라는 조건 아래에서 λ의 값에 관계없이 $M \sim \mathrm{B}(5, \frac{1}{4})$이다.

답

$m = 3, 4, 5$에 대한 이항확률을 더하면

$$\begin{aligned}
\mathrm{P}(M \geq 3 \mid T = 5) &= \binom{5}{3}\frac{3^2}{4^5} + \binom{5}{4}\frac{3}{4^5} + \binom{5}{5}\frac{1}{4^5} \\
&= \frac{10\times 9 + 5\times 3 + 1}{1024} = \frac{106}{1024} = \frac{53}{512}.
\end{aligned}$$

따라서 깨진 접시 다섯 개 중 수학자가 세 개 이상을 깼을 확률은 $\dfrac{53}{512} \approx 0.104$이다.

검산

- 여섯 개의 조건부 확률 $\binom{5}{m}3^{5-m}/4^5$를 모두 더하면 당연히 그래야 하듯이 $(1+3)^5/4^5 = 1$이 된다.
- 대칭성: 같은 논리로 다섯 개의 접시 중 각 학생 몫의 기댓값은 $5 \times \frac{1}{4} = \frac{5}{4}$이고, 네 학생의 몫을 더하면 5가 된다.
- $\lambda = 0.3, 1, 2.7$ and 5에 대하여 네 개의 푸아송분포로부터 직접(깨진 접시 다섯 개를 네 학생에게 나누는 모든 방법에 대해 합하여) $\mathrm{P}(M \geq 3 \mid T = 5)$를 계산하였더니 매번 $0.103515625 = 53/512$를 얻었다. 설거지 4×10^5회를 모의실험(simulation)한 결과, 정확히 다섯 개가 깨진 회차들 가운데에서 약 0.105를 얻었다.

흔한 실수

- 답에 λ를 남겨 두는 것. 이것은 반드시 약분되어야 하며, 그렇지 않다면 조건부 계산이 어딘가에서 잘못된 것이다.
- $\mathrm{P}(M+R=5)$로 나누면서 R를 $\mathrm{Po}(3\lambda)$가 아니라 $\mathrm{Po}(\lambda)$로 다루는 것. 이렇게 하면 학생이 두 명뿐인 셈이 된다.
- 이항계수를 빠뜨리거나, 사건을 "세 개 이상"이 아니라 "정확히 세 개"로 다루는 것.

STEP 너머

이것은 일반적인 사실의 작은 예이다. $X_1, \ldots, X_k$가 평균이 각각 $\lambda_1, \ldots, \lambda_k$인 서로 독립인 푸아송 확률변수일 때, 그 합이 n이라는 조건 아래에서 벡터 $(X_1, \ldots, X_k)$는 확률이 $\lambda_i / \sum \lambda_j$인 다항분포(multinomial distribution)를 따른다. 거꾸로 말하면, 개수가 $\mathrm{Po}(\mu)$를 따르는 사건들 각각에 k가지 유형 중 하나를 독립적으로 붙이면, 각 유형의 개수는 서로 독립인 푸아송 확률변수가 된다. 이러한 "푸아송 분할(Poisson splitting)"은 푸아송 과정(Poisson process)을 연구하는 데 기본적인 도구이다. Blitzstein과 Hwang의 *Introduction to Probability*(확률론 입문)는 두 방향을 모두 많은 예와 함께 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.