풀이

STEP I 2003, 문제 14

확률과 통계 • 푸아송 분포

> **STEP I 2003, 14번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/03-s1-q14

핵심 아이디어

제인(Jane)이 만나는 친구의 수는 전화한 친구의 수 자체가 아니라, 그 수를 2에서 잘라 낸(capped) 값이다. 이렇게 상한이 있는 확률변수(random variable)는 세 가지 경우만 구별한다. 아무도 전화하지 않는 경우, 정확히 한 명이 전화하는 경우, 두 명 이상이 전화하는 경우이다. 따라서 푸아송(Poisson) 급수(series) 전체를 더할 필요가 전혀 없다. $P(0)$과 $P(1)$만 구하면 되고, "두 명 이상"의 확률은 나머지 확률 전부이다. 새 친구에 대해서는 그 친구가 전화하는지 여부를 조건으로 한다.

풀이

원래의 친구들

어느 날 전화하는 친구의 수를 N이라 하면 $N \sim \mathrm{Po}(2)$이고

$$P(N=k) = \frac{\mathrm{e}^{-2}2^k}{k!}, \qquad k = 0, 1, 2, \ldots$$

이다. 제인은 $S = \min(N, 2)$명의 친구를 만난다. 즉 $N = 0$이면 $S = 0$, $N = 1$이면 $S = 1$, $N \geq 2$이면 $S = 2$이다. 이제

$$P(N=0) = \mathrm{e}^{-2}, \qquad P(N=1) = 2\mathrm{e}^{-2}, \qquad P(N \geq 2) = 1 - 3\mathrm{e}^{-2},$$

이므로

$$\mathrm{E}(S) = 0 \cdot \mathrm{e}^{-2} + 1 \cdot 2\mathrm{e}^{-2} + 2\left(1 - 3\mathrm{e}^{-2}\right) = 2 - 4\mathrm{e}^{-2},$$

이고, 이것이 구하려던 결과이다.

같은 계산을 더 깔끔하게 쓰는 방법이 있으며, 아래에서 다시 사용한다. $\{0,1,2,\dots\}$에 속하는 값을 갖는 임의의 확률변수 X에 대하여

$$\mathrm{E}(\min(X,2)) = 2 - 2P(X=0) - P(X=1),$$

이 성립한다. $\min(X,2)$는 $X=0$일 때 2보다 2만큼, $X=1$일 때 1만큼 작고, 그 밖의 경우에는 2와 같기 때문이다. $X=N$으로 놓으면 다시 $2-2\mathrm{e}^{-2}-2\mathrm{e}^{-2}$를 얻는다.

새 친구가 생긴 경우

새 친구가 전화하면 $B=1$, 그렇지 않으면 $B=0$으로 두면 $P(B=1)=p$이다. 문제에서 B는 N과 독립(independent)이라고 하였다. 이제 전화하는 사람의 총수는 $M=N+B$이고, 제인은 $\min(M,2)$명의 친구를 만난다.

> **?** **생각해 보기** M의 분포 전체가 필요한가?

필요하지 않다. 위의 항등식에 따르면 $P(M=0)$과 $P(M=1)$만 있으면 된다. 독립성을 이용하여 확률을 곱하면

$$\begin{aligned} P(M=0) &= P(N=0)\,P(B=0) = (1-p)\mathrm{e}^{-2}, \\ P(M=1) &= P(N=1)\,P(B=0) + P(N=0)\,P(B=1) = 2(1-p)\mathrm{e}^{-2} + p\,\mathrm{e}^{-2}. \end{aligned}$$

이다. 따라서

$$\begin{aligned} \mathrm{E}(\min(M,2)) &= 2 - 2(1-p)\mathrm{e}^{-2} - 2(1-p)\mathrm{e}^{-2} - p\,\mathrm{e}^{-2} \\ &= 2 - (4-3p)\mathrm{e}^{-2}. \end{aligned}$$

다른 방법으로, B를 조건으로 할 수도 있다. 새 친구가 전화하지 않으면(확률 $1-p$) 평균은 앞에서와 같이 $2-4\mathrm{e}^{-2}$이다. 새 친구가 전화하면(확률 p) 제인은 $\min(N+1,2)$명을 만나는데, 이 값은 $N=0$이면 1, 그 밖에는 2이므로 평균은 $\mathrm{e}^{-2}+2(1-\mathrm{e}^{-2}) = 2-\mathrm{e}^{-2}$이다. 그러므로 평균은

$$(1-p)(2-4\mathrm{e}^{-2}) + p(2-\mathrm{e}^{-2}) = 2-(4-3p)\mathrm{e}^{-2}.$$

이다. 따라서 이제 제인은 하루에 평균 $2-(4-3p)\mathrm{e}^{-2}$명의 친구를 만나며, 이는 $3p\,\mathrm{e}^{-2}$만큼 늘어난 것이다.

검산

- $p=0$이면 새 친구는 전혀 전화하지 않으므로 $2-4\mathrm{e}^{-2} \approx 1.4587$을 다시 얻는다.
- $p=1$이면 $2-\mathrm{e}^{-2} \approx 1.8647$을 얻는데, 이는 당연히 $\min(N+1,2)$의 평균이다.

- 답은 p에 대하여 증가하고 항상 2보다 작으므로 타당하다.
- 정확한 분포에 대하여 $\min(k+b,2)$를 더한 값과, $p=0.3$으로 4×10^5일을 모의실험한 결과 모두 약 1.580으로, $2-3.1\mathrm{e}^{-2}$와 일치한다.

흔한 실수

- 친구 두 명이라는 상한을 잊고 $\mathrm{E}(N)=2$ 또는 $2+p$라고 답하는 것.
- $N+B$를 푸아송 분포로 취급하는 것. 베르누이(Bernoulli) 확률변수는 푸아송 확률변수가 아니므로, 그 합도 푸아송 분포를 따르지 않는다.
- 전화하는 사람이 정확히 한 명이 되는 두 가지 경우(원래 친구 한 명이 전화하고 새 친구는 전화하지 않는 경우, 원래 친구는 아무도 전화하지 않고 새 친구가 전화하는 경우) 중 하나를 빠뜨리는 것.

STEP 너머

항등식 $\mathrm{E}(\min(X,2))=P(X\geq1)+P(X\geq2)$는 음이 아닌 정수 값을 갖는 확률변수에 대한 꼬리 합 공식(tail-sum formula) $\mathrm{E}(X)=\sum_{k\geq1}P(X\geq k)$의 특수한 경우이다. 이는 생존 함수(survival function)를 적분하는 것의 이산 버전이다. 이와 같이 상한이 있는 평균은 대기행렬 모형(queueing model)과 재고 모형(inventory model)에 나타나며, 수용 능력이 제한되어 있을 때 서비스를 받는 고객 수의 기댓값을 준다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 여기서 사용한 두 도구인 꼬리 합 공식과 조건부 기댓값(conditional expectation)을 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.