풀이

STEP II 1998, 문제 14

확률과 통계 • 이항분포와 푸아송 분포의 기댓값

> **STEP II 1998, 14번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/98-s2-q14

핵심 아이디어

데이터베이스의 분류와 달리 이 문제는 확률 문제이다. 새 제도에서 받을 **기댓값**(*expected value*) 급여가 고정 급여 A보다 크면 둠 박사는 새 제도를 택해야 한다. 낙제하는 학생 수는 이항분포(binomial distribution)를 따르므로 $\mathrm{E}[1/(1+X)]$를 구해야 한다. 다루기 불편한 인수 $\frac{1}{k+1}$은 항등식 $\frac{1}{k+1}\binom{N}{k} = \frac{1}{N+1}\binom{N+1}{k+1}$로 흡수되고, 이로써 합은 $(p+q)^{N+1}$의 거의 완전한 이항전개가 된다. 같은 기법을 $\frac{\lambda^k}{(k+1)!}$에 적용하면 푸아송(Poisson) 모형도 처리할 수 있다.

풀이

새 제도를 받아들여야 할 때

둠 박사의 학생 가운데 낙제하는 학생 수를 X라 하고, $q = 1 - p$로 둔다. 학생들은 각각 확률 p로 서로 독립적으로 낙제하므로 $X \sim \mathrm{B}(N, p)$이다.

$$\mathrm{P}(X = k) = \binom{N}{k} p^k q^{N-k}, \qquad k = 0, 1, \ldots, N.$$

둠 박사는 기대 급여로 두 제도를 판단하는 데 만족한다고 가정한다. 새 제도에서 급여는 확률변수(random variable) $S = B/(1+X)$이므로

$$\mathrm{E}(S) = B \sum_{k=0}^{N} \frac{1}{k+1} \binom{N}{k} p^k q^{N-k}.$$

생각해 보기 합 안에 있는 $\frac{1}{k+1}$을 어떻게 없앨 수 있을까?

계승(factorial)을 이용한다.

$$\frac{1}{k+1}\binom{N}{k} = \frac{N!}{(k+1)!\,(N-k)!} = \frac{1}{N+1}\cdot\frac{(N+1)!}{(k+1)!\,((N+1)-(k+1))!}$$
$$= \frac{1}{N+1}\binom{N+1}{k+1}.$$

이를 대입하고, p의 거듭제곱이 새 이항계수와 맞도록 p를 곱하고 나누면

$$\mathrm{E}(S) = \frac{B}{(N+1)p}\sum_{k=0}^{N}\binom{N+1}{k+1}p^{k+1}q^{N-k}$$
$$= \frac{B}{(N+1)p}\sum_{j=1}^{N+1}\binom{N+1}{j}p^{j}q^{N+1-j}$$
$$= \frac{B}{(N+1)p}\left[(p+q)^{N+1} - q^{N+1}\right]$$

이다. 여기서 $j = k+1$로 두고 이항정리(binomial theorem)를 썼는데, 이 합은 $(p+q)^{N+1}$의 전개에서 $j=0$인 항 q^{N+1} 하나만 빠진 것이기 때문이다. $p+q=1$이므로

$$\mathrm{E}(S) = \frac{B(1-(1-p)^{N+1})}{(N+1)p}.$$

$\mathrm{E}(S) > A$이면 받아들여야 한다. 양변에 양수인 $(N+1)p$를 곱하면(이 공식이 의미를 가지려면 $p>0$이어야 한다. $p=0$이면 단순히 B와 A를 비교하면 된다) 이 조건은

$$A(N+1)p < B(1-(1-p)^{N+1}),$$

이 되고, 이것이 구하는 조건이다.

푸아송 모형

N이 크고 p가 작으며 Np가 적당한 크기일 때, 이항분포 $\mathrm{B}(N,p)$는 평균이 $\lambda = Np$인 푸아송 분포로 잘 근사된다. 따라서 둠 박사의 학급이 크고, 각 학생이 낙제할 가능성은 낮으며, 낙제가 여전히 독립적으로 일어난다면 낙제자 수를 푸아송 분포로 모형화할 수 있다. (더 일반적으로, 푸아송 모형은 낙제가 하나씩, 독립적으로, 일정한 평균 비율로 일어날 때 알맞다.)

$X \sim \mathrm{Po}(\lambda)$, $\lambda = Np$일 때도 같은 아이디어가 통하는데, 이번에는 $\frac{1}{k+1}\cdot\frac{1}{k!} = \frac{1}{(k+1)!}$을 쓴다.

$$\mathrm{E}\Big(\frac{B}{1+X}\Big) = B\sum_{k=0}^{\infty}\frac{1}{k+1}\cdot\frac{\mathrm{e}^{-\lambda}\lambda^{k}}{k!} = \frac{B\mathrm{e}^{-\lambda}}{\lambda}\sum_{k=0}^{\infty}\frac{\lambda^{k+1}}{(k+1)!}$$
$$= \frac{B\mathrm{e}^{-\lambda}}{\lambda}(\mathrm{e}^{\lambda}-1) = \frac{B(1-\mathrm{e}^{-\lambda})}{\lambda},$$

마지막 합은 지수함수의 급수(exponential series)에서 첫째 항 1이 빠진 것이기 때문이다. 따라서 이 모형에서 둠 박사의 기대 급여는

$$\frac{B(1-\mathrm{e}^{-Np})}{Np} \text{ 파운드.}$$

검산

- $N = 1$: 직접 계산하면 $\mathrm{E}(S) = B\left(q + \frac{1}{2}p\right) = B\left(1 - \frac{1}{2}p\right)$이고, 공식은 $B\,\frac{1-q^2}{2p} = B\,\frac{1+q}{2} = B\left(1 - \frac{1}{2}p\right)$를 준다.
- $p \to 0$: 이항분포로 구한 식은 B로 수렴하고(아무도 낙제하지 않는다), $\lambda \to 0$일 때 푸아송 분포로 구한 식도 마찬가지이다.
- 근사가 잘 맞아야 할 때 두 답은 일치한다. $N = 1000$, $p = 0.003$이면 $\mathrm{E}(S)/B$는 이항분포 공식으로 0.31655, 푸아송 공식으로 0.31674이다. $N = 20$, $p = 0.3$일 때 정확한 전수 계산과 모의실험은 모두 $\mathrm{E}(S)/B \approx 0.1586$을 준다.

흔한 실수

- $\mathrm{E}\left(\frac{1}{1+X}\right) = \frac{1}{1+\mathrm{E}(X)} = \frac{1}{1+Np}$로 쓰는 것. 기댓값은 비선형 함수를 그대로 통과하지 않는다. 실제로 $\frac{1}{1+x}$은 볼록함수(convex function)이므로 옌센 부등식(Jensen's inequality)에 의해 참값이 더 크다.
- 첨자를 바꾼 합에서 $j = 0$인 항이 빠진다는 것을 잊고 $\mathrm{E}(S) = B/((N+1)p)$를 얻는 것.

STEP 너머

여기서 쓴 기법은 $\frac{1}{1+x} = \int_0^1 t^x \,\mathrm{d}t$를 이용해 $\mathrm{E}\left(1/(1+X)\right)$를 계산하는 방법의 특수한 경우이며, 이를 통해 확률생성함수(probability generating function) $G(t) = \mathrm{E}(t^X)$와 연결된다. 답은 $\int_0^1 G(t)\,\mathrm{d}t$이다. $G(t) = (q+pt)^N$과 $G(t) = \mathrm{e}^{\lambda(t-1)}$에 대해 직접 확인해 보아라. 생성함수와 이항분포가 푸아송 분포로 수렴하는 현상은 Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)에 잘 설명되어 있다.

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.