풀이

STEP I 1992, 문제 14

확률과 통계 • 통계적 주장의 해석

> **STEP I 1992, 14번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/92-s1-q14

이 문제는 *Beyond STEP*의 13.1 절 **신문 머리기사 읽기**에서 단계별 안내와 함께 자세히 다룬다:
https://beyondstep.keithmcnulty.org/ko/ch13/headlines.html

핵심 아이디어

모든 헤드라인은 숫자 하나를 가지고 곧바로 결론으로 비약한다. 각 헤드라인에 대해 두 가지를 물어야 한다. 첫째, 같은 것끼리 비교하고 있는가? 즉 같은 대상을 같은 방식으로 측정하고, 위험에 노출된 정도로 보정했는가? 둘째, 그 차이가 단지 우연일 수는 없는가? 둘째 질문에 대해서는 드물고 서로 독립인 사건의 횟수를 나타내는 푸아송 모형(Poisson model)이 자연스러운 기준이 되며, 이 모형의 표준편차(standard deviation)는 평균의 제곱근이다. 이 모형이 언제 들어맞고 언제 들어맞지 않는지 아는 것이 (ii)부터 (iv)까지의 핵심이다.

풀이

이 풀이 전체에서 "평균적으로"는 1974–1989년의 연평균, 즉 한 해에 보행자 사망 1080명, 항공 사망 180명을 뜻한다.

소문항 (i)

"여섯 배"라는 말은 비 $1080/180 = 6$에서 나왔지만, 이것은 위험이 아니라 단순 총계를 비교한 것이다. 위험이란 노출 단위당 사망자 수, 즉 이동 한 번당, 한 시간당, 또는 이동 거리 1킬로미터당 사망자 수이다. 폴다비아의 거의 모든 사람은 매일 도로 근처를 걷지만, 비행기를 타는 사람은 상대적으로 적고 그마저도 가끔이다. 안전성을 비교하려면 다음과 같은

것이 필요하다.

$$\frac{\text{보행자 사망자 수}}{\text{총 보행 시간(또는 거리)}} \quad \text{대} \quad \frac{\text{항공 사망자 수}}{\text{총 비행 시간(또는 거리)}},$$

그런데 헤드라인은 두 분모 중 어느 것도 알려 주지 않는다. 어떤 척도를 쓰느냐에 따라 결론이 반대로 나올 수도 있다. 킬로미터당으로 보면 비행은 매우 안전해 보이지만, 시간당이나 이동 한 번당으로 보면 그렇지 않을 수도 있다.

다른 문제도 있다. 여기서 "걷기"는 사실 걸어가다가 차량에 치이는 것을 뜻하며, 이 수치는 한적한 곳의 보행자에 대해서는 아무것도 말해 주지 않는다. 항공 사망자에는 승객뿐 아니라 승무원도 포함된다. 그리고 어떤 기간의 평균으로 1991년의 안전성에 대해 무언가를 "증명"할 수는 없다. 기껏해야 1974–1989년에 대한 증거일 뿐이다.

소문항 (ii)

얼핏 보면 125는 평균 180보다 한참 낮다. 이 감소가 우연 이상의 것인가?

사망이 한 번에 한 명씩 서로 독립적으로 일어난다면 평균이 180인 푸아송 모형이 적절할 것이다. 그 표준편차는 $\sqrt{180} \approx 13.4$이므로 125는 평균보다 표준편차의 약 4.1배만큼 낮고, 우연히 일어날 가능성은 매우 작다($P(X \leq 125) \approx 10^{-5}$).

그러나 항공 사망은 한 번에 한 명씩 일어나지 않는다. 사망은 무리 지어 일어난다. 추락 사고 한 번으로 수십 명이나 수백 명이 목숨을 잃을 수 있다. 자연스러운 모형은 **치명적 추락 사고**의 횟수가 푸아송 분포를 따르고, 각 사고의 사망자 수는 확률변수라는 것이다. 이렇게 하면 변동성이 크게 늘어난다. 예를 들어 치명적 사고가 한 해 평균 6번 일어나고 한 번에 30명씩 사망한다고 하자. 그러면 연간 사망자 수는 $N \sim \text{Po}(6)$일 때 $30N$이고, 그 표준편차는 $30\sqrt{6} \approx 73$이며,

$$P(\text{사망자 수} \leq 125) = P(N \leq 4) \approx 0.29.$$

이 (예시용) 수치에 따르면 사망자가 125명인 해는 전혀 이상하지 않다. 평소보다 사고가 한두 번만 적으면 된다. 사고마다 규모까지 달라지게 하면 퍼짐은 더 커진다.

따라서 한 해의 수치만으로는 알 수 있는 것이 거의 없다. 1990년에 비행이 얼마나 이루어졌는지(항공 교통량이 적었다면 어차피 사망자도 적을 것으로 예상된다), 그리고 실제 추세가 있는지 보기 위해 여러 해의 수치도 알아야 한다. 헤드라인의 주장은 뒷받침되지 않는다.

소문항 (iii)

여기에는 별개의 잘못이 두 가지 있다.

하루로는 증가를 보여 줄 수 없다. "증가하고 있다"는 것은 시간에 따른 추세에 대한 주장인데, 하루의 수치로는 아무런 비교도 할 수 없다. 오랜 기간에 걸친 일별(또는 연별) 수치가 필요하다.

일곱 명은 특이한 수치조차 아니다. 우리가 가진 도로 관련 수치는 보행자에 대한 것뿐이다. 한 해에 1080명, 즉

$$\lambda = \frac{1080}{365} \approx 2.96 \text{ 명/일}.$$

보행자 사망은 푸아송 분포로 꽤 잘 모형화된다. 드물고, 대부분 사고 한 건당 한 명이며, 대략 서로 독립이기 때문이다. 어느 하루에 사망한 보행자 수를 $X \sim \mathrm{Po}(2.96)$라 하자. 그러면

$$P(X \geq 7) = 1 - \sum_{k=0}^{6} \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!} \approx 0.031,$$

이므로 보행자가 7명 이상 사망하는 날은 한 해에 약 $365 \times 0.031 \approx 11$일 있을 것으로 예상된다. 게다가 헤드라인은 "사망한 사람들"이라고 했으므로 운전자, 승객, 자전거 이용자도 포함된다. 하루 전체 도로 사망률은 2.96보다 높을 수밖에 없고, 그러면 7은 더욱 대수롭지 않은 수치가 된다.

> ? **생각해 보기** 평균의 두 배가 넘는 사망자가 나온 날이라면 걱정할 만하지 않은가?

작은 수에서는 그렇지 않다. 평균이 약 3일 때 푸아송 표준편차는 $\sqrt{3} \approx 1.7$로 평균에 비해 크므로, 날마다 비율상 큰 변동이 있는 것이 정상이다. 최악의 날을 골라내는 것은 선택 효과이기도 하다. 한 해에는 365일이 있으니 신문은 언제든 나쁜 날 하나쯤은 찾아낼 수 있다. 요일도 중요하다. 교통량과 걸어 다니는 사람 수는 요일과 계절에 따라 달라지기 때문이다. 비율이 일정한 푸아송 모형은 대략적인 일차 근사일 뿐이다.

소문항 (iv)

이것은 실제로 따져 볼 만한 근거가 있는 헤드라인이다. 먼저 통계를 보고, 그다음 해석을 보자.

1350은 우연 이상인가? 여기서는 (iii)에서 말한 이유로 푸아송 모형이 좋은 출발점이다. 평균이 1080이면 표준편차는 $\sqrt{1080} \approx 32.9$이고,

$$\frac{1350 - 1080}{32.9} \approx 8.2.$$

평균보다 표준편차의 8배만큼 높은 값이 나올 확률은 이 모형에서 10^{-15} 정도이다. 실제 연도별 퍼짐이 (예컨대 혹독한 겨울 때문에) 푸아송 값의 두세 배라 하더라도, 1350은 꾸준한 비율 1080에 비하면 여전히 매우 이례적이다. 그러니 무언가가 바뀌었을 가능성이 매우 높다. 문제는 그것이 무엇이냐이다.

무엇과 비교하는가? 헤드라인은 1990년을 1989년이 아니라 16년 평균과 비교한다. 보행자 사망이 이 기간 동안 꾸준히 늘어 왔다면 평균 1080은 기간의 중간(1981–82년 무렵)을 나타낼 것이고, 1990년은 그보다 훨씬 높아야 한다. 예를 들어 1974년의 약 855명에서 시작해 해마다 30명씩 꾸준히 늘었다면 평균은 1080이고 1990년에는 약 1335명이 된다. 이 경우

1990년은 전혀 "증가"가 아니라 오랜 추세가 이어진 것일 뿐이다. 그 추세는 헤드라인감일 수 있지만, 다른 헤드라인이어야 한다. 두 이야기를 구별하려면 연도별 수치가 필요하다.

노출이 달라졌는가? 인구가 늘거나, 차량이 늘거나, 걷는 일이 늘면 걷기가 이동 한 번당 더 위험해지지 않아도 사망자는 늘어난다. 공정한 비교는 인구 100 000명당 또는 보행 거리 1킬로미터당 사망자 수와 같은 비율로 해야 한다.

측정 방식이 달라졌는가? 사망을 기록하거나 분류하는 방식이 바뀌면 겉보기에 급증한 것처럼 보일 수 있다. 사고 후 얼마 뒤까지의 사망을 도로 사망으로 셀지, 누구를 보행자로 셀지가 바뀌거나, 보고가 더 충실해지는 경우가 그 예이다. 이러한 변화가 없었는지 확인해야 한다.

단어가 맞는가? "사상자"는 보통 사망자 **또는 부상자**를 뜻한다. 인용된 수치는 사망자만의 것이므로 헤드라인은 수치가 보여 주는 것을 과장한다. 부상자는 훨씬 많고, 전혀 다른 양상을 보일 수도 있다.

요약하면, 1350으로의 증가는 일정한 비율 1080 주위의 푸아송 변동으로 설명하기에는 너무 크다. 그러나 이를 "우려스럽다"고 하기 전에 최근의 연간 수치(갑작스러운 증가와 장기 추세를 구별하기 위해)와 인구 및 교통량 수치(횟수를 비율로 바꾸기 위해)가 필요하다. 또한 정의가 바뀌지 않았는지도 확인해야 한다.

검산

확률은 푸아송 급수를 정확히 더하는 방법과 2×10^5회 시행의 시뮬레이션으로 확인했다. (iii)에서는 $\lambda = 1080/365$일 때 $P(X \geq 7) = 0.0315$이다(시뮬레이션: 0.0317). (ii)에서는 한 해에 사고가 6번, 한 번에 사망자가 30명일 때 $P(N \leq 4) = 0.285$이다. 사고 규모가 1부터 59까지 균등하게 변하도록 하면 시뮬레이션에서 표준편차는 84, $P(\text{사망자 수} \leq 125) \approx 0.28$이었다. 푸아송 횟수에 대한 간단한 어림 규칙은, 평균에서 제곱근의 약 두 배 이내의 값은 대수롭지 않다는 것이다. 하루의 경우 $3 + 2 \times 1.7 \approx 6.4$이므로 7은 그 범위를 간신히 벗어날 뿐이며, 365일 가운데에서 고른다면 그런 날은 반드시 생긴다. 한 해의 경우 $1080 + 2 \times 33 = 1146$이고, 이는 1350에 한참 못 미친다.

흔한 실수

- (i)에서 비율 대신 총계를 비교하는 것. 위험에는 언제나 분모가 필요하다.
- (ii)에서 푸아송 표준편차 $\sqrt{180}$을 쓰는 것. 항공 사고의 사망은 무리 지어 일어나므로 퍼짐이 훨씬 크다.
- (iv)에서 16년 평균과의 비교, 인구, 정의가 그 증가를 설명하는지 묻지 않고 "표준편차의 8배이니 실제 증가이다"에서 멈추는 것.

STEP 너머

소문항 (ii)는 **과대산포**(*overdispersion*)에 관한 것이다. 각각 크기가 무작위인 무리가 푸아송 분포를 따르는 개수만큼 모여 이루어진 총합은 **복합 푸아송**(*compound Poisson*) 확률변수이며, 그 분산은 λ가 아니라 $\lambda\,\mathrm{E}[S^2]$이다. 소문항 (iv)는 가설 검정(hypothesis testing)으로, 그리고 통계적으로 유의한 변화가 실제로 무엇을 의미하는가라는 훨씬 어려운 질문으로 이어진다. 두 주제 모두 대학 확률과 통계 과정의 초반에 나온다. Blitzstein과 Hwang의 *Introduction to Probability*(확률론 입문)는 푸아송 모형과 그 한계에 대해 읽기 좋은 책이다.

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.