풀이

STEP II 2008, 문제 12

확률과 통계 • 독립사건

> **STEP II 2008, 12번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/08-s2-q12

핵심 아이디어

개수의 기댓값(expectation)은 그 개수를 이루는 개별 사건들의 확률의 합이므로, 첫 번째 부분은 p에 대한 이차방정식 하나로 귀결된다. 두 번째 부분에서는 판사들이 피고인을 보지도 않고 마음속으로 동전을 던지므로, 왕의 선고는 진실과 독립(independent)이다. 올바른 선고가 내려질 확률을 t로 나타낸 뒤 $\frac{1}{2}$을 중심으로 정리하면 답이 한 줄로 나온다.

풀이

p 구하기

당나귀, 비버, 켄타우로스가 각각 유죄 판결을 내리는 사건을 A, B, C라 하고, 이들의 지시확률변수(indicator variable)를 I_A, I_B, I_C라 하자(사건이 일어나면 1, 그렇지 않으면 0이다). 그러면 $X = I_A + I_B + I_C$이고, $\mathrm{E}(I_A) = \mathrm{P}(A)$ 등이 성립하므로 기댓값의 선형성(linearity of expectation)에 의해

$$\mathrm{E}(X) = p + \frac{p}{3} + p^2.$$

이다. (선형성에는 독립성이 필요하지 않다. 독립성은 나중에만 사용한다.) 이 값을 $\frac{4}{3}$으로 놓고 양변에 3을 곱하면

$$3p^2 + 4p - 4 = 0 \quad \Longrightarrow \quad (3p-2)(p+2) = 0.$$

이다. 확률은 음수가 될 수 없으므로 $p + 2 \neq 0$이고, 따라서

$$p = \frac{2}{3}.$$

이다. 그러므로 세 판사가 "유죄"라고 말할 확률은

$$\mathrm{P}(A) = \frac{2}{3}, \qquad \mathrm{P}(B) = \frac{2}{9}, \qquad \mathrm{P}(C) = \frac{4}{9},$$

이며, 당연히 모두 $[0,1]$에 속한다.

왕이 "유죄"라고 말할 확률

적어도 두 판사가 유죄 판결을 내리는 사건을 K라 하자. 그러면 K는 왕이 피고인에게 유죄를 선고하는 사건이다. 위의 세 확률을 a, b, c로 쓰자. 정확히 두 명이 유죄 판결을 내리는 경우는 세 가지, 세 명 모두 유죄 판결을 내리는 경우는 한 가지이다. 판사들의 독립성을 이용하면

$$\begin{aligned}\mathrm{P}(K) &= ab(1-c) + ac(1-b) + bc(1-a) + abc \\ &= ab + ac + bc - 2abc.\end{aligned}$$

이다. 여기서 $ab = \frac{4}{27}$, $ac = \frac{8}{27}$, $bc = \frac{8}{81}$, $abc = \frac{16}{243}$이므로

$$\mathrm{P}(K) = \frac{36}{243} + \frac{72}{243} + \frac{24}{243} - \frac{32}{243} = \frac{100}{243}.$$

이다. 간단히 이 값을 g라 하자.

t 구하기

피고인이 실제로 유죄인 사건을 D라 하면 $\mathrm{P}(D) = t$이다. 왕의 선고가 옳은 것은 D와 K가 둘 다 일어나거나 둘 다 일어나지 않을 때이다.

> **생각해 보기** 왜 $\mathrm{P}(D)$와 $\mathrm{P}(K)$를 곱할 수 있는가?

판사들은 "완전히 무작위로" 결정한다. 즉 그들의 판결은 증거에 의존하지 않으므로 K는 D와 독립이다. 따라서

$$\mathrm{P}(\text{옳은 선고}) = \mathrm{P}(D)\,\mathrm{P}(K) + \mathrm{P}(D')\,\mathrm{P}(K') = tg + (1-t)(1-g).$$

이다. 이제 $\frac{1}{2}$을 빼고 인수분해하면

$$\begin{aligned}tg + (1-t)(1-g) - \tfrac{1}{2} &= 2tg - t - g + \tfrac{1}{2} \\ &= (2t-1)(g - \tfrac{1}{2}).\end{aligned}$$

이다. 그러므로 왕의 선고가 옳을 확률이 $\frac{1}{2}$인 것은 $(2t-1)(g-\frac{1}{2}) = 0$일 때이고, 그때뿐이다. $g = \frac{100}{243} \neq \frac{1}{2}$이므로 두 번째 인수는 0이 아니고, 따라서

$$t = \frac{1}{2}.$$

이어야 한다. 인수분해를 쓰지 않고 같은 결과를 얻을 수도 있다. $\frac{100}{243}t + \frac{143}{243}(1-t) = \frac{1}{2}$에서 $143 - 43t = \frac{243}{2}$이므로 $43t = \frac{43}{2}$, 즉 $t = \frac{1}{2}$이다.

여기서 p의 값은 거의 상관이 없다는 점에 주목하자. 판사들의 확률이 무엇이든 $\mathrm{P}(K) \neq \frac{1}{2}$이기만 하면, 증거를 무시하는 법정이 절반의 경우에 옳은 판결을 내리는 것은 유죄인 피고인과 무죄인 피고인이 똑같이 흔할 때이고, 그때뿐이다.

검산

- $p = \frac{2}{3}$일 때 $\frac{2}{3} + \frac{2}{9} + \frac{4}{9} = \frac{12}{9} = \frac{4}{3}$으로, 요구 조건을 만족한다.
- 여덟 가지 판결 패턴을 모두 분수로 정확히 나열하면 $\mathrm{P}(K) = \frac{100}{243} \approx 0.412$, $\mathrm{E}(X) = \frac{4}{3}$을 얻는다.
- $t = \frac{1}{2}$로 200 000회 시행한 시뮬레이션에서 왕의 선고가 옳은 비율은 0.5007이었다.
- 공식의 타당성 점검: $t = 1$(모든 피고인이 유죄)이면 왕이 옳을 확률은 g이고, $t = 0$이면 $1 - g$이다. 이 확률은 t에 대한 일차식이므로 중점에서 $\frac{1}{2}$을 지난다.

흔한 실수

- "적어도 두 명"을 $ab + ac + bc$로 계산하는 것. 이렇게 하면 세 명 모두인 경우가 세 번 세어진다. 올바른 식은 $2abc$를 빼야 한다(또는 서로소인 네 가지 경우를 직접 더한다).
- $p = -2$를 버리는 것을 잊거나, p^2과 $p/3$이 올바른 확률인지 확인하지 않는 것.
- "왕이 옳게 선고한다"를 $\mathrm{P}(K)$만으로 생각하여, 올바르게 무죄 방면된 무고한 피고인들을 빠뜨리는 것.

STEP 너머

첫 번째 부분은 강력한 기법의 작은 예이다. 개수를 지시확률변수들의 합으로 나타내면, 사건들이 서로 독립이 아니더라도 그 기댓값은 단순히 확률들의 합이 된다. 대학 확률론에서는 이 기법을 끊임없이 사용하는데, 예를 들어 무작위 순열(permutation)의 고정점(fixed point) 개수의 기댓값을 구할 때 쓴다. 두 번째 부분은 출력이 진실과 독립인 분류기(classifier)에 관한 것으로, 검사에서의 민감도(sensitivity), 특이도(specificity), 베이즈 정리(Bayes' theorem)를 생각하는 출발점이 된다. Blitzstein과 Hwang의 *Introduction to Probability*(확률론 입문)는 지시확률변수를 많이 활용하며, 다음 단계로 읽기에 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.