풀이

STEP I 2005, 문제 12

확률과 통계 • 벤 다이어그램과 범위

> **STEP I 2005, 12번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/05-s1-q12

핵심 아이디어

"어떤 값의 범위가 모순 없이 가능한가?"라는 꼴의 문제는, 각 영역의 값은 모르지만 모두 음이 아니어야(*non-negative*) 하는 벤 다이어그램(Venn diagram)에 관한 문제이다. 모든 영역을 되도록 적은 수의 자유 미지수로 나타낸다. 그러면 "이 영역 ≥ 0"이라는 각 조건이 하나의 경계가 되고, 이 경계들을 모두 모으면 범위가 나온다. 소문항 (ii)에서는 세 개의 단일 확률을 더할 때 각 마법사가 몇 번씩 세어지는지를 따지는 세기 논법(counting argument)으로, 개별 영역을 구하지 않고도 "정확히 두 개"의 확률을 얻는다.

풀이

소문항 (i)

호빗이 파이프 담배를 피우는 사건을 S, 모자를 쓰는 사건을 H라 하자. $t = \mathrm{P}(S \cap H)$로 놓는다. 벤 다이어그램의 네 영역은 다음과 같다.

$$\mathrm{P}(S \cap H) = t, \qquad \mathrm{P}(S \cap H') = 0.7 - t,$$
$$\mathrm{P}(S' \cap H) = 0.4 - t, \qquad \mathrm{P}(S' \cap H') = 1 - (0.7 + 0.4 - t) = t - 0.1.$$

네 값이 모두 음이 아니게 하는 t를 고르면 실제로 성립하는 확률분포(probability distribution)가 되고, 역도 성립한다. 실제로 제약이 되는 조건은 $t \geq 0.1$(마지막 영역에서)과 $t \leq 0.4$(세 번째 영역에서)이며, 나머지 조건은 이로부터 따라 나온다. 따라서 $0.1 \leq t \leq 0.4$이고, $p = 0.7 - t$이므로

$$0.3 \leq p \leq 0.6.$$

양 끝값은 모두 실제로 나온다. $p = 0.6$은 파이프를 피우는 호빗 중 모자를 쓰는 호빗이 가능한 한 적을 때(모든 호빗이 두 가지 중 적어도 하나를 할 때)이고, $p = 0.3$은 모자를 쓰는 호빗이 모두 파이프를 피울 때이다.

소문항 (ii): 정확히 두 개

마법사가 모자, 망토, 반지를 착용하는 사건을 각각 H, C, R라 하자. 마법사가 세 물건 중 정확히 한 개, 정확히 두 개를 착용할 확률을 각각 a, b라 하자. 정확히 하나도 착용하지 않을 확률은 0.05, 정확히 세 개를 착용할 확률은 0.1이므로

$$0.05 + a + b + 0.1 = 1 \quad \Longrightarrow \quad a + b = 0.85.$$

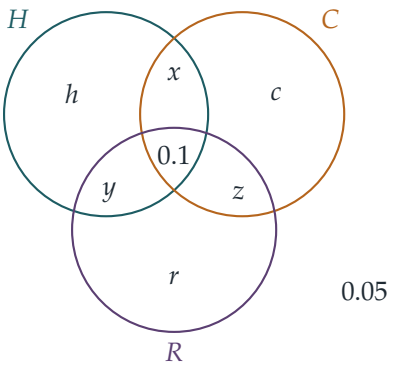
이제 주어진 세 확률을 더하면 $\mathrm{P}(H) + \mathrm{P}(C) + \mathrm{P}(R) = 0.7 + 0.8 + 0.4 = 1.9$이다. 이 합에서 정확히 k개의 물건을 착용한 마법사는 k번 세어지므로

$$a + 2b + 3(0.1) = 1.9 \quad \Longrightarrow \quad a + 2b = 1.6.$$

두 식을 빼면 $b = 0.75$(그리고 $a = 0.1$)이다. 마법사가 물건 중 정확히 두 개를 착용할 확률은 $\frac{3}{4}$이다.

소문항 (ii): q의 범위

그림과 같이 안쪽의 일곱 영역에 이름을 붙인다. 정확히 한 개인 영역은 h, c, r, 정확히 두 개인 영역은 x, y, z이고, 가운데는 0.1, 바깥은 0.05이다.



H, C, R의 전체 확률과 지금까지 구한 소문항 (ii)의 결과로부터

$$h + x + y = 0.6, \qquad c + x + z = 0.7,$$
$$r + y + z = 0.3, \qquad x + y + z = 0.75.$$

마지막 식을 이용해 앞의 세 식에서 각각 두 미지수를 소거하면

$$h = z - 0.15, \qquad c = y - 0.05, \qquad r = x - 0.45.$$

(검산하면 $h+c+r = 0.75-0.65 = 0.1 = a$이다.) 따라서 전체 분포는 $x+y+z = 0.75$를 만족하는 x, y, z로 결정되며, 여섯 영역이 모두 음이 아닐 때, 그리고 그때에만 실제로 성립하는 분포가 된다.

$$x \geq 0.45, \qquad y \geq 0.05, \qquad z \geq 0.15.$$

? 생각해 보기 x는 얼마나 커질 수 있을까?

$x = 0.75 - y - z$이고 $y+z \geq 0.05+0.15 = 0.2$이므로 $x \leq 0.55$이며, 등호는 $y = 0.05$, $z = 0.15$일 때 성립한다. 가장 작은 값 $x = 0.45$는 예를 들어 $y = 0.1, z = 0.2$일 때 나온다. y와 z를 연속적으로 움직일 수 있으므로 그 사이의 모든 값도 나온다. 따라서

$$0.45 \leq x \leq 0.55.$$

끝으로, 망토를 착용한 상태에서 모자는 쓰고 반지는 끼지 않은 경우가 정확히 영역 x이므로

$$q = \mathrm{P}(H \cap R' \mid C) = \frac{\mathrm{P}(H \cap C \cap R')}{\mathrm{P}(C)} = \frac{x}{0.8}, \qquad \text{이므로} \qquad \frac{9}{16} \leq q \leq \frac{11}{16}.$$

검산

- 극단인 $x = 0.55, y = 0.05, z = 0.15$에서는 $h = 0, c = 0, r = 0.1$이다. 여덟 영역의 합은 1이고, 요구한 대로 $\mathrm{P}(H) = 0.7$, $\mathrm{P}(C) = 0.8$, $\mathrm{P}(R) = 0.4$가 된다.
- 주어진 자료와 모순되지 않는 여덟 영역의 값 배정을 0.01 간격으로 모두 열거해 보았다. 모든 경우에 $x+y+z = 0.75$이고, q는 0.5625부터 0.6875까지 변한다. 마찬가지로 (i)에서 p는 0.3부터 0.6까지 변한다.

흔한 실수

- (i)에서 "둘 다 아님" 영역을 잊어서 $t \geq 0.1$ 대신 하한 $t \geq 0$을 얻고, 그 결과 잘못된 상한 $p \leq 0.7$을 얻는 실수.
- (ii)에서 $\mathrm{P}(C) = 0.8$ 대신 1로 나누는 실수. "망토를 착용했을 때"라는 조건(condition)이 분모를 바꾼다.
- H, C, R를 독립(independent)으로 취급하는 실수. 문제 어디에도 독립이라는 말은 없으며, 실제로 $\mathrm{P}(H \cap C \cap R) = 0.1 \neq 0.7 \times 0.8 \times 0.4$이다.

STEP 너머

선형 방정식과 음이 아닐 조건 아래에서, 미지의 확률들에 대해 선형인 양의 범위를 구하는 것은 작은 **선형 계획법**(*linear programme*) 문제이다. 극값은 언제나 실행 가능 영역(feasible

region)의 꼭짓점에서 나타나며, 그래서 우리의 극값도 영역을 o으로 놓는 데서 나왔다. 같은 아이디어로 소문항 (i)에서 쓴 프레셰 부등식(Fréchet bounds) $\max(0, \mathrm{P}(A) + \mathrm{P}(B) - 1) \leq \mathrm{P}(A \cap B) \leq \min(\mathrm{P}(A), \mathrm{P}(B))$을 얻는다. 포함-배제 원리(inclusion–exclusion)와 조건부 확률(conditional probability)에 대해 더 알고 싶다면 Blitzstein과 Hwang의 *Introduction to Probability*(확률론 입문)를 읽어 보라.

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.