풀이

STEP II 1990, 문제 16

확률과 통계 • 베이즈 정리

> **STEP II 1990, 16번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/90-s2-q16

핵심 아이디어

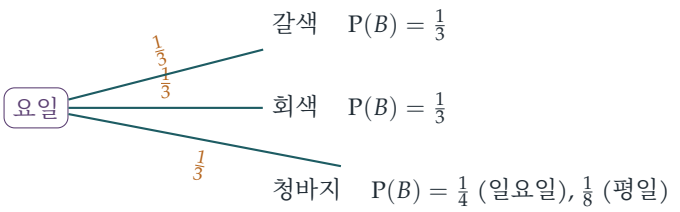
이 문제는 베이즈 정리(Bayes' theorem)를 다른 모습으로 바꾸어 놓은 것이다. 사진을 보면 검은 넥타이를 맸다는 것을 알 수 있고, 우리가 구하려는 것은 그 사실이 **주어졌을** 때 그날이 무슨 요일인지에 대한 확률이다. 그래서 먼저 순방향으로 계산한다. 각 종류의 날마다 바지에 따라 갈라지는 수형도를 이용하여 검은 넥타이를 맬 확률을 구한다. 그다음 조건부 확률(conditional probability)의 방향을 뒤집는다. 월요일 규칙은 수형도의 가지 하나만 바꾸므로, 두 번째 부분은 추가로 할 일이 거의 없다.

풀이

사진이 일요일에 찍힌 사건을 S, 내가 검은 넥타이를 맨 사건을 B라 하자. 사진은 일곱 요일 중 어느 날에 찍혔을 가능성이 모두 같으므로, 사진을 보기 전에는 $\mathrm{P}(S) = \frac{1}{7}$이다. (나중에 다룰) 월요일을 제외한 모든 날에 각 바지를 입을 확률은 $\frac{1}{3}$이다.

각 종류의 날에 검은 넥타이를 맬 확률

- **갈색 바지.** 크라바트는 제외되므로, 넥타이는 검은 넥타이, 물방울무늬 넥타이, 레지멘탈 넥타이 중 하나를 임의로 고른다: $\mathrm{P}(B) = \frac{1}{3}$.
- **회색 바지.** 물방울무늬 넥타이는 제외되므로, 검은 넥타이, 레지멘탈 넥타이, 크라바트 중에서 고른다: $\mathrm{P}(B) = \frac{1}{3}$.
- **일요일의 청바지.** 넥타이는 항상 매며, 네 가지 모두 중에서 고른다: $\mathrm{P}(B) = \frac{1}{4}$.
- **평일의 청바지.** 확률 $\frac{1}{2}$로 스웨트셔츠를 입고 넥타이를 매지 않는다. 그렇지 않으면 네 넥타이 중 하나를 맨다: $\mathrm{P}(B) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$.



가지를 따라 더하면(바지에 대한 전확률 법칙(law of total probability))

$$\mathrm{P}(B \mid \text{일요일}) = \tfrac{1}{3}\left(\tfrac{1}{3}+\tfrac{1}{3}+\tfrac{1}{4}\right) = \tfrac{1}{3}\cdot\tfrac{11}{12} = \tfrac{11}{36} = \tfrac{22}{72},$$
$$\mathrm{P}(B \mid \text{평일}) = \tfrac{1}{3}\left(\tfrac{1}{3}+\tfrac{1}{3}+\tfrac{1}{8}\right) = \tfrac{1}{3}\cdot\tfrac{19}{24} = \tfrac{19}{72}.$$

첫 번째 답

일요일의 사전 확률(prior probability)을 $\frac{1}{7}$, 여섯 평일 전체의 사전 확률을 $\frac{6}{7}$로 하여 베이즈 정리를 적용하면

$$\mathrm{P}(S \mid B) = \frac{\frac{1}{7}\cdot\frac{22}{72}}{\frac{1}{7}\cdot\frac{22}{72}+\frac{6}{7}\cdot\frac{19}{72}} = \frac{22}{22+6\times 19} = \frac{22}{136} = \frac{11}{68},$$

이므로 원하는 결과를 얻는다. (공통 인수 $\frac{1}{7\times 72}$가 약분되는데, 모든 것을 분모 72로 맞춰 두면 좋은 이유가 바로 이것이다.)

> **?** **생각해 보기** 왜 답이 그냥 $\mathrm{P}(B \mid \text{일요일})$이나 $\frac{1}{7}$이 아닐까?

$\mathrm{P}(B \mid \text{일요일}) = \frac{11}{36}$은 엉뚱한 질문에 대한 답이다. 이것은 요일이 주어졌을 때 넥타이를 맬 확률이지, 넥타이가 주어졌을 때 그 요일일 확률이 아니다. 또 $\frac{1}{7}$은 증거를 무시한 값이다. 일요일에는 넥타이를 매지 않는 선택지가 없으므로 검은 넥타이를 맬 가능성이 조금 더 높다($\frac{22}{72}$ 대 $\frac{19}{72}$). 그래서 사진은 일요일일 확률을 $\frac{1}{7} = \frac{11}{77}$에서 $\frac{11}{68}$로 살짝 끌어올린다.

월요일 규칙이 있을 때

이제 월요일에는 반드시 청바지를 입는데, 월요일도 평일이므로 스웨트셔츠 선택지는 여전히 적용된다:

$$\mathrm{P}(B \mid \text{Monday}) = 1\cdot\tfrac{1}{2}\cdot\tfrac{1}{4} = \tfrac{1}{8} = \tfrac{9}{72}.$$

일요일과 나머지 다섯 평일은 변하지 않는다. 각 요일의 사전 확률은 여전히 $\frac{1}{7}$이고 이것이 약분되므로

$$\mathrm{P}(S \mid B) = \frac{\frac{22}{72}}{\frac{22}{72}+\frac{9}{72}+5\cdot\frac{19}{72}} = \frac{22}{22+9+95} = \frac{22}{126} = \frac{11}{63}.$$

이 값은 $\frac{11}{68}$보다 조금 크다. 이는 타당하다. 월요일에는 검은 넥타이를 맬 가능성이 낮아지므로($\frac{19}{72}$가 아니라 $\frac{1}{8}$), 검은 넥타이 사진은 월요일에서 멀어지고 일요일 쪽을 조금 더 가리키게 된다.

검산

- 각 요일에 대해 넥타이별 확률과 "넥타이 없음"의 확률을 더하면 1이 된다. 예를 들어 평일에 넥타이를 매지 않을 확률은 $\frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}$이다.
- 임의의 날 4×10^5개를 모의실험(simulation)하면, 검은 넥타이를 맨 날 중 일요일의 비율은 약 0.161(정확한 값 $\frac{11}{68} \approx 0.1618$)이고, 월요일 규칙이 있을 때는 약 0.174(정확한 값 $\frac{11}{63} \approx 0.1746$)이다.

흔한 실수

- 월요일에도 청바지에는 스웨트셔츠를 입을 확률 절반이 따라온다는 것을 잊는 실수. 월요일도 평일이다. 월요일에 $\frac{1}{8}$ 대신 $\frac{1}{4}$를 쓰면 $\frac{22}{135}$가 나온다.
- 갈색 바지나 회색 바지를 입을 때 네 넥타이 모두 중에서 고르는 실수. 이 두 경우에는 넥타이가 세 개만 남는다.

STEP 너머

"가설이 주어졌을 때 증거의 확률"을 "증거가 주어졌을 때 가설의 확률"로 바꾸는 것이 베이즈 추론(Bayesian inference)의 핵심이다. 깔끔하게 말하면, 사후 오즈(posterior odds)는 사전 오즈(prior odds)에 우도비(likelihood ratio)를 곱한 것과 같다. 여기서 일요일 대 평일의 사전 오즈는 1 : 6이고 우도비는 22 : 19이므로 22 : 114, 즉 11 : 57을 얻는다. 대학 과정에서는 이를 발전시켜 베이즈 통계학(Bayesian statistics)을 배우는데, 여기서는 분포 전체를 같은 방식으로 갱신한다. Blitzstein과 Hwang의 *Introduction to Probability*(확률론 입문)에는 이 문제와 같은 예가 많이 실린, 조건부 확률에 관한 훌륭한 장이 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.