풀이

STEP I 2006, 문제 12

확률과 통계 • 조건부 확률

> **STEP I 2006, 12번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/06-s1-q12

핵심 아이디어

도로는 그 위의 **모든** 구간이 열려 있을 때에만 열려 있고, 두 마을은 **모든** 도로가 막혔을 때에만 고립된다. 따라서 독립성(independence)을 이용하여 도로별로 따지고, 항상 여사건(complement)을 통해 계산한다. 둘째 부분은 어느 도로를 선택했는지가 중요한 조건부 확률(conditional probability) 문제이다. 구간이 하나인 도로는 구간이 둘인 도로보다 열려 있을 가능성이 더 크므로, 내 도로가 열려 있다는 사실은 내가 어느 도로를 선택했는지에 대한 정보를 준다. 이 점은 정의 $\mathrm{P}(B \mid A) = \mathrm{P}(A \cap B)/\mathrm{P}(A)$를 이용하고, 선택한 도로에 따라 경우를 나누어 처리한다.

풀이

고립될 확률

구간이 둘인 도로를 R_1, R_2라 하고 구간이 하나인 도로를 R_3이라 하자. 한 구간이 열려 있을 확률을 $q = 1 - p$라 하자.

구간이 둘인 도로는 두 구간이 모두 열려 있을 때 정확히 열려 있으며, 독립성에 의해 그 확률은 $q^2 = (1-p)^2$이다. 따라서

$$\mathrm{P}(R_1 \text{ 막힘}) = \mathrm{P}(R_2 \text{ 막힘}) = 1 - (1-p)^2 = p(2-p),$$
$$\mathrm{P}(R_3 \text{ 막힘}) = p.$$

세 도로는 서로 다른 다섯 개의 구간을 사용하므로, 사건 "R_1 막힘", "R_2 막힘", "R_3 막힘"은 서로소인 독립 구간 집합에 의존하고, 따라서 서로 독립이다. 옥스타운(Oxtown)이 캠

빌(Camville)로부터 고립되는 것은 세 도로가 모두 막혔을 때와 정확히 같으므로

$$\mathrm{P}(\text{고립}) = p(2-p)\cdot p(2-p)\cdot p = p^3(2-p)^2,$$

이고, 이것이 구하려던 결과이다.

조건부 확률

내가 도로 R_k를 선택하는 사건을 C_k라 하면 각 k에 대해 $\mathrm{P}(C_k) = \frac{1}{3}$이고, 나의 선택은 침수와 독립이다. 내 도로가 열려 있는 사건을 O, 나머지 두 도로가 모두 막혀 있는 사건을 B라 하자. 구하려는 것은 $\mathrm{P}(B \mid O) = \mathrm{P}(B \cap O)/\mathrm{P}(O)$이다.

? **생각해 보기** 독립성에 의해 답은 그냥 P(나머지 두 도로가 막힘)이 아닐까?

어느 도로를 선택했는지 알고 있다면 그렇다. 그러나 "나머지 두 도로"는 우리의 선택에 따라 다른 쌍이 되며, 내 도로가 열려 있다는 관찰은 짧은 도로 R_3을 선택했을 가능성을 이전보다 높인다. 따라서 선택한 도로에 따라 경우를 나누고 올바르게 가중치를 주어야 하며, 아래의 식이 바로 그 일을 한다.

분모. 도로의 선택에 대해 전확률 법칙(law of total probability)을 쓰면

$$\begin{aligned}\mathrm{P}(O) &= \tfrac{1}{3}(1-p)^2 + \tfrac{1}{3}(1-p)^2 + \tfrac{1}{3}(1-p)\\ &= \tfrac{1}{3}(1-p)[2(1-p)+1] = \tfrac{1}{3}(1-p)(3-2p).\end{aligned}$$

분자. R_1을 선택했다면 R_1은 열려 있고 R_2, R_3은 막혀 있어야 한다. 도로들의 독립성에 의해 이 확률은 $(1-p)^2\cdot p(2-p)\cdot p$이다. R_2를 선택한 경우도 같다. R_3을 선택했다면 R_3은 열려 있고 구간이 둘인 두 도로는 모두 막혀 있어야 하므로 확률은 $(1-p)\cdot p^2(2-p)^2$이다. 따라서

$$\begin{aligned}\mathrm{P}(B\cap O) &= \tfrac{1}{3}\left[2p^2(2-p)(1-p)^2 + p^2(2-p)^2(1-p)\right]\\ &= \tfrac{1}{3}\,p^2(1-p)(2-p)[2(1-p)+(2-p)]\\ &= \tfrac{1}{3}\,p^2(1-p)(2-p)(4-3p).\end{aligned}$$

답. 나누고, $p<1$일 때 0이 아닌 인수 $\frac{1}{3}(1-p)$를 약분하면

$$\mathrm{P}(B \mid O) = \frac{p^2(2-p)(4-3p)}{3-2p}.$$

인수 p, $2-p$, $4-3p$, $3-2p$는 모두 일차식이므로, 이것은 완전히 인수분해된 형태이다.

극한 $p \to 1$

$p \to 1$일 때 모든 인수가 1로 수렴하므로

$$\mathrm{P}(B \mid O) \to \frac{1 \cdot 1 \cdot 1}{1} = 1.$$

이는 납득할 만하다. 침수가 거의 확실할 때 내 도로가 열려 있다는 것은 운 좋은 우연이며, 내 도로와 독립적으로 침수되는 다른 도로들은 거의 확실히 나와 같은 행운을 누리지 못했다. 내 도로가 열려 있다는 사실은 다른 도로들이 열려 있는지에 대해 아무것도 알려 주지 않는다. 따라서 다른 어떤 선택을 했더라도 실패했으리라고 거의 확신할 수 있다. ($p = 1$ 자체에서는 조건 사건의 확률이 0이므로 조건부 확률이 정의되지 않으며, 극한만이 의미를 가진다.)

검산

- $p = 0$이면 아무것도 막히지 않으며, 식은 마땅히 0을 준다.
- $p = \frac{1}{2}$이면 식은 $\frac{\frac{1}{4} \cdot \frac{3}{2} \cdot \frac{5}{2}}{2} = \frac{15}{32}$를 준다. 막힌 구간의 2^5가지 모든 패턴과 세 가지 도로 선택을 모두 정확히 열거해도 같은 값이 나오며, $p = \frac{2}{3}$일 때 $\frac{32}{45}$가 나오는 것처럼 다른 여러 p값에 대해서도 마찬가지이다.
- $0 \le p < 1$에서 답은 항상 0과 1 사이에 있다. 예를 들어 $p = 0.99$일 때 약 0.9996이다.

흔한 실수

- 구간이 둘인 도로가 막힐 확률을 $1 - (1 - p)^2$(적어도 하나가 막힘) 대신 p^2(두 구간이 모두 막힘)으로 잡는 것.
- 조건을 거는 것을 잊는 것. 나머지 두 도로가 막혀 있을 확률은, 무작위로 선택한 내 도로가 열려 있다는 조건 아래에서의 조건부 확률과 같지 않다.
- 관찰 **이후에** 세 가지 선택에 같은 가중치를 주는 것. 같은 것은 사전 확률(prior probability) $\frac{1}{3}$이지 사후 확률(posterior probability)이 아니다.

STEP 너머

직렬로 연결된 구간들로 이루어진 도로들이 병렬로 놓인 구조는 **신뢰성 네트워크**(*reliability network*)의 가장 간단한 예이다. 네트워크가 두 점을 연결할 확률은 일반적으로 스미기 이론(percolation theory)에서 연구되며, 이 이론에서는 p가 변함에 따라 행동이 급격하게 바뀌는 현상이 핵심 주제이다. 둘째 부분은 모습을 바꾼 베이즈 정리(Bayes' theorem)이

다. "내 도로가 열려 있다"는 관찰이 어느 도로를 선택했는지에 대한 확률을 갱신한다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 바로 이러한 관점에서 조건부 확률과 독립성을 다룬다.

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.