풀이

STEP I 2007, 문제 12

확률과 통계 • 조건부 확률

STEP I 2007, 12번 문제

더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:

step.maths.org/questions/07-s1-q12

핵심 아이디어

두 번째 사탕을 뽑을 때 주머니들이 어떤 상태인지는 앞에서 무슨 일이 일어났는지에 따라 달라지므로 미리 알 수 없다. 그래서 조건화(conditioning)를 한다. 즉 두 번째로 뽑기 전에 일어난 모든 경우로 나누고, 각 경우에 빨간 사탕이 나올 확률(probability)을 구한 다음, 전확률 법칙(law of total probability)을 이용해 더한다. 소문항 (ii)에는 유용한 지름길이 있다. 방금 무작위로 사탕 하나를 받은 주머니에 들어 있는 빨간 사탕 개수의 기댓값(expected value)은 쉽게 적을 수 있으므로, 옮겨진 사탕의 색에 따라 경우를 여러 번 나눌 필요가 없다.

풀이

소문항 (i)

첫 번째 사탕이 빨간색인 사건(event)을 R_1, 두 번째 사탕이 빨간색인 사건을 R_2라 하자. 모든 사탕이 첫 번째로 뽑힐 가능성은 같으므로

$$\mathrm{P}(R_1) = \frac{a}{N}.$$

R_2에 대해서는 첫 번째 사탕의 색으로 조건화한다. 첫 번째 사탕이 빨간색이면 남은 $N-1$개 중 $a-1$개가 빨간색이고, 빨간색이 아니면 남은 $N-1$개 중 a개가 빨간색이다. (여기서

$N \geq 2$이므로 두 번째 사탕이 반드시 존재한다.) 전확률 법칙에 의해

$$\begin{aligned} \mathrm{P}(R_2) &= \mathrm{P}(R_1)\,\mathrm{P}(R_2 \mid R_1) + \mathrm{P}(R_1')\,\mathrm{P}(R_2 \mid R_1') \\ &= \frac{a}{N} \cdot \frac{a-1}{N-1} + \frac{N-a}{N} \cdot \frac{a}{N-1} \\ &= \frac{a[(a-1)+(N-a)]}{N(N-1)} = \frac{a(N-1)}{N(N-1)} = \frac{a}{N}. \end{aligned}$$

따라서 요구한 대로 $\mathrm{P}(R_2) = \mathrm{P}(R_1)$이다. (대칭성(symmetry)으로도 같은 결론을 얻는다. 서로 다른 두 사탕의 순서쌍은 모두 같은 확률로 나오므로, 두 번째 사탕이 특정한 사탕일 가능성은 첫 번째 사탕이 그 사탕일 가능성과 같다.)

소문항 (ii)

여기서 "첫 번째 사탕"은 첫 번째 동전 던지기 후에 옮겨지는 사탕이고, "두 번째 사탕"은 두 번째 동전 던지기 후에 고르는 사탕이다. 이 사탕들이 빨간색인 사건을 각각 R_1, R_2라 하자. 두 번의 동전 던지기는 서로 독립(independent)이고, 각각은 어떤 사탕이 뽑히는지와도 독립이다.

첫 번째 사탕. 첫 번째 동전 던지기로 조건화하면

$$\mathrm{P}(R_1) = p \cdot \frac{a}{N} + q \cdot \frac{b}{N} = \frac{pa+qb}{N}.$$

두 번째 사탕. 두 번의 동전 던지기에 대해 네 가지 경우가 있다.

앞면, 앞면. 주머니 1에서 사탕 하나를 꺼내고, 이어서 주머니 1에서 또 하나를 꺼낸다. 이는 정확히 소문항 (i)의 상황을 주머니 1(사탕 N개, 빨간색 a개)에 적용한 것이므로, 두 번째 사탕이 빨간색일 확률은 a/N이다.

앞면, 뒷면. 사탕 하나가 주머니 1에서 주머니 2로 옮겨져 주머니 2에는 이제 사탕이 $N+1$개 있고, 그다음 주머니 2에서 뽑는다. 주머니 2에는 확률 a/N으로 빨간 사탕이 $b+1$개, 확률 $1-a/N$으로 b개 있으므로

$$\mathrm{P}(R_2 \mid \mathrm{HT}) = \frac{a}{N} \cdot \frac{b+1}{N+1} + \left(1 - \frac{a}{N}\right)\frac{b}{N+1} = \frac{b+\frac{a}{N}}{N+1} = \frac{Nb+a}{N(N+1)}.$$

? 생각해 보기 두 번째 줄은 왜 이렇게 깔끔하게 정리되는가?

분자는 옮긴 뒤 주머니 2에 있는 빨간 사탕 개수의 기댓값이다. 처음부터 있던 b개에, 들어온 사탕이 빨간색일 확률 a/N을 더한 것이다. 이 경우 주머니 2의 크기는 $N+1$로 고정되어 있으므로, 빨간 사탕을 뽑을 확률은 이 기댓값을 $N+1$로 나눈 것이다.

뒷면, 앞면. 주머니의 역할을 바꾸어 같은 논증을 하면, 주머니 1에는 이제 사탕이 $N+1$개 있고 빨간 사탕 개수의 기댓값은 $a+b/N$이므로

$$\mathrm{P}(R_2 \mid \mathrm{TH}) = \frac{Na+b}{N(N+1)}.$$

뒷면, 뒷면. 소문항 (i)을 주머니 2에 적용하면 확률은 b/N이다.

이제 네 경우를 각각의 확률 p^2, pq, qp, q^2와 함께 합친다.

$$\begin{aligned}\mathrm{P}(R_2) &= p^2\frac{a}{N} + pq\,\frac{(Nb+a)+(Na+b)}{N(N+1)} + q^2\frac{b}{N}\\ &= p^2\frac{a}{N} + pq\,\frac{(N+1)(a+b)}{N(N+1)} + q^2\frac{b}{N}\\ &= \frac{p^2a + pqa + pqb + q^2b}{N}\\ &= \frac{pa(p+q)+qb(p+q)}{N} = \frac{pa+qb}{N},\end{aligned}$$

마지막 단계에서 $p+q=1$을 이용했다. 이는 $\mathrm{P}(R_1)$과 같으며, 이것이 보여야 할 내용이었다.

검산

- $p=1$이면 모든 일이 주머니 1에서 일어나고, 두 사탕 모두에 대한 답 a/N은 정확히 소문항 (i)이다. $p=0$이면 주머니 2에 대한 소문항 (i)을 얻는다.
- $a=b=0$이면 빨간 사탕이 없으므로 두 확률 모두 0이고, $a=b=N$이면 모든 사탕이 빨간색이므로 둘 다 1이다.
- $N=5$, $a=2$, $b=4$, $p=0.3$일 때 공식은 $(0.6+2.8)/5=0.68$을 준다. 전체 과정을 200 000번 시뮬레이션(simulation)한 결과 첫 번째 사탕은 0.679, 두 번째 사탕은 0.681이었고, 작은 N, a, b 전체에 대해 정확히 모든 경우를 따져 본 결과도 일치한다.

흔한 실수

- 옮긴 뒤 받는 주머니에는 사탕이 N개가 아니라 $N+1$개 있고, 주는 주머니에는 $N-1$개 있다는 것을 잊는다.
- HH와 TT 경우에 두 번째 뽑기를 첫 번째 뽑기와 독립인 것으로 취급한다. 독립은 아니지만, 어쨌든 소문항 (i)에서 답이 바뀌지 않는다는 것을 알고 있으므로 계산을 다시 하지 말고 그 결과를 이용하면 된다.

STEP 너머

두 소문항 모두 깊은 아이디어의 예이다. 항아리에서 무엇을 꺼냈는지 보지 않고 뽑으면 항아리에 대한 우리의 지식은 바뀌지 않으므로, 다음 뽑기도 같은 분포(distribution)를 가진다. 엄밀하게 말하면 뽑기의 수열은 **교환가능**(*exchangeable*)하며, 소문항 (ii)에서 다음에 뽑는 사탕이 빨간색일 확률인 양 $p \times$ (주머니 1의 빨간 사탕 비율) $+ q \times$ (주머니 2의 빨간 사탕 비율)은 **마팅게일**(*martingale*)이다. 마팅게일이란 미래 값의 기댓값이 현재 값과 같은 양을 말한다. 마팅게일은 도박에서 금융에 이르기까지 현대 확률론의 핵심이다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 조건부 기댓값(conditional expectation)과 이러한 아이디어를 처음 접하기에 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.