풀이

STEP II 1994, 문제 12

확률과 통계 • 조건부 확률

> **STEP II 1994, 12번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/94-s2-q12

핵심 아이디어

첫 번째 던지기로 "포인트(point)" s가 정해지고 나면, 그 뒤의 던지기는 모두 결과가 세 가지인 똑같은 독립(independent) 시행이다. s가 나오면 이기고, 7이 나오면 지며, 그 밖의 값이 나오면 계속한다. 따라서 포인트가 주어졌을 때 게임이 끝나는 던지기의 차례는 기하분포(geometric distribution)를 따르고, 문제 전체는 포인트에 따라 경우를 나눈 뒤 전확률 법칙(law of total probability)으로 더하는 일로 귀결된다. 마지막 부분은 조건으로 삼는 사건만 다를 뿐 같은 계산이다.

풀이

S를 공정한 주사위 두 개의 눈의 합이라 하자. 일어날 가능성이 같은 36가지 결과를 세면

$$\mathrm{P}(S = s) = \frac{6 - |s - 7|}{36}, \qquad s = 2, 3, \ldots, 12,$$

이므로 합 $2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$의 확률은 $\frac{1}{36} \times (1, 2, 3, 4, 5, 6, 5, 4, 3, 2, 1)$이다.

첫 번째 던지기에서 이길 확률

첫 번째 합이 7 또는 11이면 그녀는 곧바로 이긴다.

$$\mathrm{P}(\text{승리, 던지기1}) = \frac{6}{36} + \frac{2}{36} = \frac{2}{9}.$$

두 번 이상 던질 때, n번째 던지기에서 이길 확률

그녀가 두 번 이상 던지는 것은 첫 번째 합이 포인트 $4, 5, 6, 8, 9, 10$ 가운데 하나일 때이고, 오직 그때뿐이다. 그 확률은

$$\mathrm{P}(M) = \frac{3+4+5+5+4+3}{36} = \frac{2}{3}$$

이다. 여기서 M은 사건 "그녀가 두 번 이상 던진다"이다.

포인트 s를 고정하고 $p_s = \mathrm{P}(S = s)$라 쓰자. 그 뒤의 각 던지기는 앞선 던지기들과 독립이고, 각 던지기에서

$$\mathrm{P}(\text{합}s) = p_s, \qquad \mathrm{P}(\text{합}7) = \tfrac{1}{6}, \qquad \mathrm{P}(\text{둘 다 아님}) = r_s = 1 - p_s - \tfrac{1}{6}.$$

이다. ($s \neq 7$이므로 이 두 사건은 서로 배반(disjoint)이다.) 그녀가 던지기 $n \geq 2$에서 이기는 것은 던지기 $2, \ldots, n-1$이 모두 "둘 다 아님"이고 던지기 n에서 s가 나올 때이며, 오직 그때뿐이다. 따라서 독립성에 의해

$$\mathrm{P}(\text{승리, 던지기}n \mid \text{포인트}s) = r_s^{n-2}\, p_s.$$

이다. 포인트는 확률이 같은 것끼리 짝을 이룬다(4와 10, 5와 9, 6과 8). $k = 3, 4, 5$에 대하여 $p_s = k/36$이라 쓰면 $r_s = (30-k)/36$이고, 이는 각각 $\frac{3}{4}$, $\frac{13}{18}$, $\frac{25}{36}$이다. 전확률 법칙에 의해 $n \geq 2$일 때

$$\mathrm{P}(\text{승리, 던지기}n) = \sum_s p_s \cdot r_s^{n-2} p_s = 2\sum_{k=3}^{5} \left(\frac{k}{36}\right)^2 \left(\frac{30-k}{36}\right)^{n-2}.$$

이다. 던지기 $n \geq 2$에서 이기면 M이 일어나므로, $\mathrm{P}(M) = \frac{2}{3}$으로 나누면 각 계수에 $\frac{3}{2}$가 곱해진다.

$$2 \cdot \frac{k^2}{1296} \cdot \frac{3}{2} = \frac{k^2}{432} \quad \text{이로부터} \quad \frac{9}{432} = \frac{1}{48}, \quad \frac{16}{432} = \frac{1}{27}, \quad \frac{25}{432}.$$

따라서

$$\mathrm{P}(\text{승리, 던지기}n \mid M) = \frac{1}{48}\left(\frac{3}{4}\right)^{n-2} + \frac{1}{27}\left(\frac{13}{18}\right)^{n-2} + \frac{25}{432}\left(\frac{25}{36}\right)^{n-2},$$

이고, 이것이 구하려던 결과이다.

? 생각해 보기 하나의 "평균" 포인트와 하나의 기하분포로 계산하면 왜 틀릴까?

계속할 확률 r_s가 포인트에 따라 달라지기 때문이다. 세 기하분포의 혼합(mixture)은 그 자체로는 기하분포가 아니며, 그래서 답에 거듭제곱이 하나가 아니라 세 개 따로 나타난다.

m번보다 많이 던질 때

$m > 1$에 대하여 M_m을 사건 "그녀가 m번보다 많이 던진다"라 하자. 이 사건은 첫 번째 던지기에서 포인트 s가 나오고 던지기 $2, \ldots, m$이 모두 "둘 다 아님"일 때이며, 오직 그때뿐이

다. 따라서

$$\mathrm{P}(M_m) = \sum_s p_s\, r_s^{m-1} = \frac{2}{36}\sum_{k=3}^{5} k\Big(\frac{30-k}{36}\Big)^{m-1}.$$

이다. $n \le m$이면 던지기 n에서 이기는 사건은 M_m과 함께 일어날 수 없으므로 조건부 확률(conditional probability)은 0이다. $n > m$이면 던지기 n에서 이길 때 M_m이 일어나므로

$$\mathrm{P}(\text{승리, 던지기}n \mid M_m) = \frac{\mathrm{P}(\text{승리, 던지기}n)}{\mathrm{P}(M_m)} = \frac{9(\frac{3}{4})^{n-2} + 16(\frac{13}{18})^{n-2} + 25(\frac{25}{36})^{n-2}}{36\Big[3(\frac{3}{4})^{m-1} + 4(\frac{13}{18})^{m-1} + 5(\frac{25}{36})^{m-1}\Big]},$$

이다. 여기서 인수 $\frac{2}{1296}$와 $\frac{2}{36}$는 약분하였다. 분자와 분모에 $\frac{1}{432}$을 곱하면 다음과 같이 쓸 수도 있다.

$$\frac{\frac{1}{48}(\frac{3}{4})^{n-2} + \frac{1}{27}(\frac{13}{18})^{n-2} + \frac{25}{432}(\frac{25}{36})^{n-2}}{\frac{1}{4}(\frac{3}{4})^{m-1} + \frac{1}{3}(\frac{13}{18})^{m-1} + \frac{5}{12}(\frac{25}{36})^{m-1}}, \qquad n > m.$$

분모가 $\mathrm{P}(M_m \mid M)$이라는 점에 주목하자. 마땅히 그래야 한다. 분자는 M을 조건으로 한 앞의 답이고, $M_m \subseteq M$이기 때문이다.

검산

- $m = 1$이면 분모는 $\frac{1}{4} + \frac{1}{3} + \frac{5}{12} = 1$이 되어 앞 부분의 결과로 돌아간다.
- $n = 2$일 때 조건부 확률 공식은 $\frac{1}{48} + \frac{1}{27} + \frac{25}{432} = \frac{50}{432} = \frac{25}{216} \approx 0.1157$을 준다. 직접 계산한 $\frac{3}{2} \cdot \frac{2(9+16+25)}{1296}$도 같은 값이다.
- $m = 2, n = 3$일 때 답은 $\frac{107}{930} \approx 0.1151$이다. 모든 공식을 분수를 이용해 게임을 한 단계씩 정확히 전개한 결과와 비교하여 확인하였고, 4×10^5번의 게임을 모의실험한 결과도 표본 오차 범위 안에서 일치하였다.

흔한 실수

- $\mathrm{P}(M) = \frac{2}{3}$으로 나누는 것을 잊는 실수. 그러면 계수가 $\frac{1}{72}$, $\frac{2}{81}$, $\frac{25}{648}$이 된다.
- 하나의 비율로 $\mathrm{P}(M_m) = (\text{어떤 값})^{m-1}$이라 쓰는 실수. 조건으로 삼는 사건 역시 포인트들에 대한 혼합이다.
- "둘 다 아님"인 던지기의 횟수를 $n-2$가 아니라 $n-1$로 세는 실수. 던지기 1은 포인트를 정하고, 던지기 n은 이기는 던지기이다.

STEP 너머

이 게임은 흡수 상태(absorbing state) "승리"와 "패배"를 가진 작은 마르코프 연쇄(Markov chain)이고, 여기서 흡수될 때까지 걸리는 시간은 기하분포들의 혼합이다. m번의 던지기

를 넘어 살아남았다는 조건을 주면 혼합 가중치가 가장 오래 버티는 포인트(여기서는 다시 나올 가능성이 가장 작은 4와 10) 쪽으로 옮겨 간다. 이는 무기억성(memoryless) 분포들의 혼합이 그 자체로는 무기억성을 갖지 않는다는 것을 보여 주는 간단한 예이다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 이 수준에서 조건부 확률과 마르코프 연쇄를 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.