풀이

STEP II 1997, 문제 12

확률과 통계 • 이항분포의 홀짝 확률

STEP II 1997, 12번 문제

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

step.maths.org/questions/97-s2-q12

핵심 아이디어

메시지가 지금 **맞는지** **틀린지**만 추적한다. 한 번 잘못 들을 때마다 이 상태가 뒤집히므로, Albert에게 돌아오는 메시지는 잘못 전달된 귓속말의 횟수가 짝수일 때, 그리고 그 때에만 맞다. 이항 전개(binomial expansion)에서 짝수 번째 항만 골라내는 것이 바로 $(a+b)^n + (a-b)^n$이 하는 일이다. 홀수 번째 항은 서로 상쇄되기 때문이다. 변형된 게임에서는, 한 번 실패한 뒤 Albert가 **틀린** 메시지를 전달한다는 점에 주목한다. 따라서 그 뒤의 각 회차에서는 잘못 듣는 횟수가 **홀수**여야 한다.

풀이

세 명의 참가자

참가자가 n명이면 귓속말은 n번 전달된다. Albert가 Bertha에게, Bertha가 다음 참가자에게, 이런 식으로 원을 따라 Zebedee까지 전달되고, 마지막으로 Zebedee가 다시 Albert에게 전달한다. 각 귓속말은 확률 p로 올바르게 들리며, 서로 독립(independent)이다. Albert가 참인 결과를 듣는지는 잘못 들린 귓속말의 횟수에만 달려 있고, 동전이 어느 면을 보였는지와는 무관하다.

참가자가 세 명이면 귓속말은 세 번이고, 각각 확률 $\frac{2}{3}$으로 올바르게 들린다. Albert에게 도달하는 메시지는 잘못 들린 귓속말이 하나도 없거나, 정확히 두 번일 때(두 번째 오류가 첫 번째 오류를 되돌린다) 맞다. 따라서

$$\Pr(\text{승리}) = \left(\tfrac{2}{3}\right)^3 + 3\left(\tfrac{1}{3}\right)^2\left(\tfrac{2}{3}\right) = \frac{8}{27} + \frac{6}{27} = \frac{14}{27}.$$

n명의 참가자

$q = 1 - p$라 하자. 잘못 들린 귓속말의 횟수는 $\mathrm{B}(n, q)$ 분포(distribution)를 따르고, 이 횟수가 짝수일 때 게임에서 이긴다.

$$P = \sum_{k\text{ 짝수}} \binom{n}{k} q^k p^{n-k}.$$

이제

$$(p+q)^n + (p-q)^n = \sum_{k=0}^{n} \binom{n}{k} p^{n-k} q^k (1 + (-1)^k) = 2 \sum_{k\text{ 짝수}} \binom{n}{k} p^{n-k} q^k,$$

이다. k가 홀수인 항은 상쇄되고 k가 짝수인 항은 두 번씩 나타나기 때문이다. $p + q = 1$이고 $p - q = 2p - 1$이므로

$$\boxed{P = \tfrac{1}{2}(1 + (2p-1)^n).}$$

$n = 3$, $p = \frac{2}{3}$이면 이 식은 $\frac{1}{2}(1 + \frac{1}{27}) = \frac{14}{27}$이 되어 앞의 결과와 일치한다.

변형된 게임

Albert가 Bertha에게 귓속말하는 횟수를 N이라 하자. 메시지가 원을 한 바퀴 완전히 도는 것을 한 **회차**라 부르면, 모든 회차는 Albert의 귓속말로 시작한다.

> **생각해 보기** 한 회차가 실패한 뒤, 다음 회차는 단지 첫 회차를 새로 반복하는 것일까?

아니다. 첫 회차에서 Albert는 올바른 결과를 보낸다. 만약 Zebedee에게서 그것을 잘못 듣는다면, Albert는 자신이 **들은** 것, 즉 틀린 메시지를 전달한다. 틀린 메시지에서 출발하면, 그 회차는 잘못 들린 귓속말의 횟수가 **홀수**일 때, 그리고 그때에만 Albert가 참인 결과를 듣는 것으로 끝난다. 그 확률은 $1 - P$이다. 그 회차가 또 실패하면 Albert는 다시 틀린 메시지를 듣고(그리고 전달하고), 따라서 그 뒤의 모든 회차는 이 회차와 같다. 각 회차는 서로 독립인데, 각각의 듣기가 독립이기 때문이다.

따라서 첫 회차는 확률 P로 성공하고, 그 뒤의 각 회차는 확률 $1 - P$로 성공하고 확률 P로 실패한다. 그러므로 $\Pr(N = 1) = P$이고, $k \geq 2$에 대하여

$$\Pr(N = k) = \underbrace{(1-P)}_{\text{제1회차 실패}} \underbrace{P^{k-2}}_{\text{제2,…,}k-1\text{회차 실패}} \underbrace{(1-P)}_{\text{제}k\text{회차 성공}} = (1-P)^2 P^{k-2}.$$

$0 < p < 1$이라 가정하면 $0 < P < 1$이다. 이때 주어진 합에 $x = P$를 대입하면

$$\sum_{k=2}^{\infty} k P^{k-2} = \frac{1}{P}\left(\sum_{k=1}^{\infty} k P^{k-1} - 1\right) = \frac{1}{P}\left(\frac{1}{(1-P)^2} - 1\right),$$

이고, 따라서

$$\begin{aligned}\mathrm{E}(N) &= P + (1-P)^2 \cdot \frac{1}{P}\left(\frac{1}{(1-P)^2} - 1\right) \\ &= P + \frac{1-(1-P)^2}{P} \\ &= P + (2-P) = 2.\end{aligned}$$

Albert가 Bertha에게 귓속말하는 횟수의 기댓값(expected value)은 참가자의 수와 p의 값에 관계없이 2이다.

이를 더 빨리 확인하는 방법이 있다. 첫 회차가 실패했다는 조건 아래에서, 추가로 필요한 회차의 수는 성공 확률이 $1-P$인 기하분포(geometric distribution)를 따르므로 그 평균은 $1/(1-P)$이다. 그러므로 $\mathrm{E}(N) = 1 + (1-P)\cdot\frac{1}{1-P} = 2$이다.

(경계의 경우: $p = 1$이거나, $p = 0$이고 n이 짝수이면 $P = 1$이고 Albert는 한 번만 귓속말한다. $p = 0$이고 n이 홀수이면 $P = 0$이고 두 번째 회차는 반드시 성공하므로 $N = 2$이다.)

검산

- 맞게 듣거나 틀리게 듣는 2^n가지 경우를 모두 나열하여 정확한 분수로 계산하면, $n \le 8$과 여러 p값에 대하여 $\frac{1}{2}(1+(2p-1)^n)$이 그대로 나온다. 세 명일 때의 $\Pr(\text{승리}) = \frac{14}{27}$도 포함된다.
- $p = \frac{1}{2}$이면 $P = \frac{1}{2}$이다. 메시지가 순전한 잡음이 되므로 당연한 결과이다.
- $(n, p) = (3, \frac{2}{3})$, $(5, 0.9)$, $(4, 0.3)$에 대하여 변형된 게임을 2×10^5번 시뮬레이션했더니, 평균 귓속말 횟수는 각각 2.001, 1.996, 1.994였다.

흔한 실수

- 귓속말을 n번이 아니라 $n-1$번으로 세는 것. Zebedee가 다시 Albert에게 하는 마지막 귓속말도 세어야 한다.
- 변형된 게임의 모든 회차가 확률 P로 성공한다고 보는 것. 그러면 $1/P$가 나오는데, 이는 틀리다. 실패한 뒤에는 Albert가 틀린 메시지를 전달하기 때문이다.

STEP 너머

메시지의 맞음/틀림 상태는 두 상태를 갖는 마르코프 연쇄(Markov chain)이고, $2p-1$은 그 전이 행렬(transition matrix)의 두 번째 고윳값(eigenvalue)이다. 이것이 $(2p-1)^n$이 메시지가 동전의 결과를 얼마나 빨리 잊는지를 결정하는 이유이다. 이와 같은 모형인

이진 대칭 통신로(binary symmetric channel)는 정보 이론(information theory)과 오류 정정 부호(error-correcting codes)의 출발점이다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 이 수준에서 생성함수(generating function)와 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.