12.9 치우친 동전으로 공정하게 정하기

학습지

문제

STEP I 2019, 11번 문제

더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 OCR에서 내려받을 수 있는 2019년 STEP 시험지 묶음(zip 파일) 가운데 STEP I 시험지에 있다:

www.ocr.org.uk/Images/696590-2019-past-papers.zip

힌트

먼저 혼자 힘으로 문제를 풀어 보자. 막히면 힌트를 **하나씩** 읽고, 힌트를 하나 읽을 때마다 다시 문제로 돌아가자. **잠깐, 직접 해 보기**라고 표시된 힌트가 핵심 단계이다. 각 힌트 뒤에는 풀이를 적을 공간이 있다.

? **생각해 보기** 한 라운드에서 결정이 내려질 확률은 얼마인가? 결정이 내려지지 않을 확률은? $2pq$가 가질 수 있는 가장 큰 값은 얼마인가?

1단계 n번째 라운드에서의 결정

다루는 내용: 소문항 (i), 라운드 n

? **생각해 보기** 한 라운드에서 동전들의 결과가 일치할 확률은 얼마인가? n번째 라운드에서 결정되려면 $n-1$번의 불일치 다음에 한 번의 일치가 일어나야 한다.

2단계 n번째 라운드까지의 결정

다루는 내용: 소문항 (i), 라운드 n까지

> **잠깐, 직접 해 보기** 처음 n번의 라운드에서 아무것도 결정되지 않을 확률은 얼마인가? $pq = p(1-p)$는 최대 얼마까지 커질 수 있는가?

3단계 세 사람

다루는 내용: 소문항 (ii)

생각해 보기 첫 번째 라운드는 세 동전이 모두 일치할 때 결정된다. 그렇지 않은 경우, 다수가 앞면(heads)이었다면 혼자 다른 동전을 앞면으로 뒤집는다. 이제 두 번째 라운드에서는 무엇이 필요한가? 다수가 뒷면(tails)이었다면 어떠한가?

잠깐, 직접 해 보기 $p+q=1$과 $p^3+q^3=(p+q)^3-3pq(p+q)$를 이용하여 모든 것을 $s=pq$로 나타내어라.

숨은 수학

폰 노이만의 기법

문제의 절차는 빠르지만 치우쳐 있다. 폰 노이만의 절차는 공정하다. 두 번 던져서 HT나 TH이면 결정하고, HH나 TT이면 반복한다. 각 시도는 확률 $2pq$로 성공하므로 평균 $\frac{1}{2pq}$번의 시도가 필요하며, 동전이 매우 치우쳐 있으면 느리다(그림 12.9.1). 문제의 절차는 라운드마다 확률 $p^2+q^2\geq\frac{1}{2}$로 결정을 내리지만, $p>\frac{1}{2}$이면 그 답은 호텔 쪽으로 기운다.

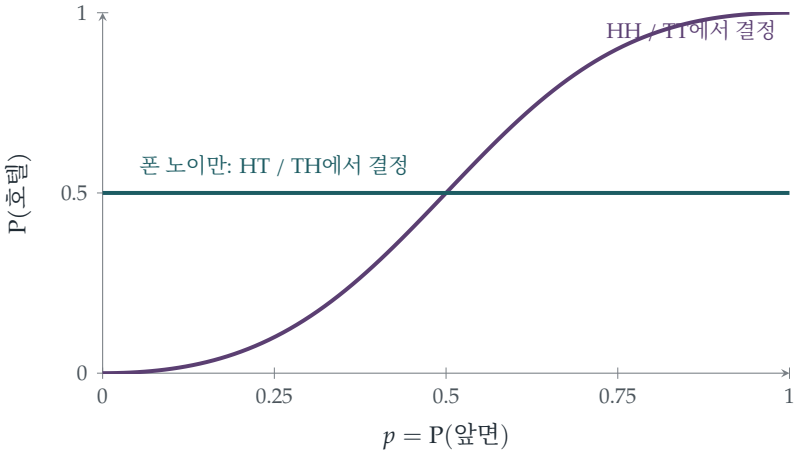

그림 12.9.1 동전의 치우침에 따른 호텔을 고를 확률. 동전들이 일치할 때 결정하는 방법(자주색)은 빠르지만 더 잘 나오는 면 쪽으로 기운다. 폰 노이만의 절차(청록색)는 모든 p에 대하여 정확히 공정하지만, 동전이 매우 치우쳐 있으면 시간이 더 오래 걸린다.

무작위성 추출

폰 노이만의 기법은 최초의 **무작위성 추출기**(*randomness extractor*), 곧 불완전한 무작위성을 완전한 무작위성으로 바꾸는 절차이다. 이 기법은 낭비가 많다(두 번 던질 때마다 공정한 비트를 많아야 하나 얻고, 하나도 얻지 못할 때도 많다). 더 나은 방법들은 버려진 HH와 TT 쌍을 재활용하여 동전에 담긴 무작위성을 거의 전부 추출하는데, 그 양은 던질 때마다 **엔트로피**(*entropy*) $-p \log_2 p - q \log_2 q$비트로 측정된다. 추출기는 컴퓨터의 하드웨어 난수 생성기에 쓰인다. 이 생성기는 예측할 수 없지만 치우친 물리적 잡음을 모아서, 암호에 쓸 수 있도록 치우치지 않은 비트로 바꾸어야 한다.

폰 노이만과 난수

존 폰 노이만은 1951년 논문 "Various techniques used in connection with random digits"(난수와 관련하여 쓰이는 여러 기법)에서 이 기법을 설명했다. 이 논문은 그와 동료들이 핵무기 설계를 위해 최초의 컴퓨터 시뮬레이션 몇 가지("몬테카를로 방법(Monte Carlo method)")를 수행하던 시기에 쓰였다. 같은 논문에서 그는 산술로 난수를 만들어 내는 일에 대해 유명한 말을 남겼다. "산술적 방법으로 난수를 만들어 내려는 사람은 누구든, 당연히, 죄를 짓고 있는 상태이다."

더 도전하기

1 **라운드 수의 기댓값.** 소문항 (i)의 절차와 폰 노이만의 절차에 대하여 라운드 수의 기댓값(expected value)을 구하여라. 어떤 p에 대하여 폰 노이만의 절차가 더 느린가?

2 **공정한 셋 중 하나 고르기.** 치우친 동전을 사용하여 세 가지 선택지 가운데 하나를 각각 정확히 확률 $\frac{1}{3}$로 고르는 절차를 고안하여라.

3 **공정한 동전으로 만드는 치우친 동전.** 공정한 동전을 사용하여 앞면이 나올 확률이 정확히 $\frac{1}{3}$인 동전을 흉내 내는 절차를 고안하여라. 던지는 횟수의 기댓값은 얼마인가?

4 **다시 세 사람.** 소문항 (ii)에서 그들이 호텔에 갈 확률을 p의 함수로 나타내면 무엇인가? $p = \frac{1}{2}$일 때 이 절차는 공정한가?

더 읽을거리

- **John von Neumann, "Various techniques used in connection with random digits"**, *National Bureau of Standards Applied Mathematics Series* 12 (1951). 짧고 유명한 논문이다.
- **Donald Knuth, *The Art of Computer Programming*, Vol. 2** (Addison-Wesley), 난수를 다룬 3장.
- **Michael Mitzenmacher & Eli Upfal, *Probability and Computing*** (Cambridge University Press), 무작위성 추출과 엔트로피를 다룬다.

안내에 따른 전체 풀이는 제 12.9절에 있다:

https://beyondstep.keithmcnulty.org/ko/ch12/von-neumann.html

매주 STEP 문제 하나를 끝까지 풀어 본다. The STEP Up은 이메일로 무료로 받아 볼 수 있다: keithmcnulty.substack.com