13.6 쌓이는 오차

학습지

문제

STEP II 1993, 15번 문제

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

step.maths.org/questions/93-s2-q15

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 어떤 수의 소수 부분 U가 $[0,1)$에서 균등분포(uniform distribution)를 따른다면, 그 수를 버림할 때 오차는 무엇인가? 반올림할 때는 어떠한가? 각각의 평균과 분산(variance)은 무엇인가?

1단계 전체 오차

다루는 내용: 수 하나의 오차와 1500개의 수 더하기

생각해 보기 어떤 수의 소수 부분(fractional part)이 $[0,1)$에서 균등분포(uniform distribution)를 따른다면, 그 수를 버림(truncation)할 때의 오차는 무엇인가? 반올림(rounding)할 때는 어떠한가? 각각의 평균(mean)과 분산(variance)을 구하여라.

잠깐, 직접 해 보기 전체 오차는 서로 독립(independent)인 1500개의 오차의 합이다. 각 컴퓨터에 대하여 그 평균과 분산은 무엇인가? 어떤 정리를 쓰면 이것을 정

규분포(normal distribution)로 다룰 수 있는가?

2단계 큰 오차가 생길 확률

다루는 내용: 1500개의 수 더하기

생각해 보기 반올림하는 컴퓨터에 대해서는 15를 표준화(standardise)하고 정규분포표를 이용하여라. 버림하는 컴퓨터에 대해서는, 전체 오차가 거의 확실히 어디에 있는가?

3단계 덧셈을 몇 번 해야 하는가?

다루는 내용: LEP의 덧셈 횟수

> **생각해 보기** 덧셈을 n번 한 뒤 LEP의 오차의 표준편차(standard deviation)는 얼마인가? $\mathrm{P}(|Z| < 1.645) = 0.9$를 이용하여 n에 대하여 풀어라.

숨은 수학

편향과 분산

버림은 **편향**(biased)되어 있다. 즉 버림 오차의 평균은 영이 아니며, 편향은 선형으로 쌓인다. 반올림은 **비편향**(unbiased)이며, 그 오차는 확률보행처럼 행동하여 $\sqrt{n}$에 비례하여 커진다. 연산을 백만 번 하고 나면 그 차이는 500 000 정도의 오차와 300 정도의 오차 사이의 차이가 된다. 이는 실험에서 계통 오차(systematic error, 예를 들어 잘못 보정된 측정 기구)와 무작위 측정 오차(random error)를 구별하는 것과 같다. 측정을 여러 번 하여 평균을 내면 무작위 오차는 줄어들지만 계통 오차는 전혀 줄어들지 않는다.

부동소수점

모든 컴퓨터는 실수를 제한된 정밀도로 저장하며, 모든 산술 연산은 그 결과를 반올림한다. 현대 컴퓨터는 IEEE 754 표준을 따르는데, 이 표준은 표현 가능한 가장 가까운 수로 반올림하고, 두 수의 한가운데인 경우에는 마지막 자릿수가 짝수가 되는 쪽으로 정한다. 바로 이렇게 해야 반올림이 평균적으로 비편향이 되기 때문이다. 그렇더라도 많은 수를 부주의하게 더하면 정확도를 잃을 수 있다. William Kahan의 **보정 합산**(compensated summation) 알고리즘(1965년)은 잃어버린 하위 자릿수를 추적하여 다시 더해 준다. 1991년 걸프 전쟁에서 한 패트리엇 미사일 포대는 날아오는 미사일을 요격하는 데 실패했는데, 그 포대의 시계가 시간을 10분의 1초 단위로 셌기 때문이다. 10분의 1은 이진법으로 정확하게 저장할 수 없는 수이다. 100시간 동안 작동한 뒤 버림 오차는 약 3분의 1초로 커져 있었고, 이는 목표물의 위치를 반 킬로미터 넘게 잘못 판단하기에 충분했다.

밴쿠버 지수

밴쿠버 증권거래소 지수는 1982년 1월 1000.000의 값으로 만들어졌다. 1983년 11월이 되자 시장이 전반적으로 상승했는데도 지수는 약 524까지 떨어져 있었다. 자문가들은 지수가 하루에 수천 번 다시 계산되었고 그때마다 소수점 아래 셋째 자리까지 (반올림이 아니라) 버림되었다는 사실을 알아냈다. 주말 동안 올바르게 다시 계산한 지수는 약 1098에서 다시 시작했다. 이 일화는 수치 계산 강의에서 빠지지 않는 경고 사례이다.

더 도전하기

1 **보정된 VOZ.** VOZ가 마지막에 답에 $\frac{n}{2}$를 더한다면 그 오차는 어떤 분포를 따르는가? 이제 LEP만큼 좋아졌는가?

2 **소수점 아래 첫째 자리로 반올림하기.** 수를 소수점 아래 첫째 자리로 반올림하고, 버려지는 자릿수가 균등분포를 따른다고 하자. 10 000개의 수를 더할 때 전체 오차의 표준편차를 구하라.

3 **밴쿠버 지수.** 지수를 하루에 3000번 다시 계산하고 한 번 버릴 때마다 평균 0.0005를 잃었다면, 거래일이 250일인 1년 동안 지수는 얼마나 줄어들겠는가?

4 **정확한 분포.** 두 반올림 오차의 합의 정확한 확률밀도함수(삼각형 모양)를 구하라. 세 개의 합에 대해서는 그 개형을 그려 보라. 얼마나 빨리 정규분포처럼 보이기 시작하는가?

더 읽을거리

- **Nicholas Higham, *Accuracy and Stability of Numerical Algorithms*** (SIAM). 반올림 오차에 관한 권위 있는 책으로, 첫 장은 읽기 쉽다.
- **David Goldberg, "What every computer scientist should know about floating-point arithmetic"**, *ACM Computing Surveys* 23 (1991). 고전적인 글로, 온라인에서 무료로 볼 수 있다.
- **Tim Harford, *How to Make the World Add Up*** (Bridge Street). 데이터 속 계통 오차를 다룬다.

안내에 따른 전체 풀이는 제 13.6절에 있다:

https://beyondstep.keithmcnulty.org/ko/ch13/errors-add-up.html

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