13.6 積み重なる誤差

ワークシート

問題

STEP II 1993、問 15

先に進む前に、まず問題を読んでおこう。問題は STEP Support Programme の問題データベースにある：

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

ヒント

まずは自力で問題に取り組んでみよう。行き詰まったら、ヒントを**一つずつ**読み、そのたびに問題に戻ること。**手を止めて、やってみよう**と記したヒントが重要な段階である。各ヒントの後には計算用の余白がある。

? **考えてみよう**　ある数の小数部分（fractional part）U が $[0,1)$ 上の一様分布（uniform distribution）に従うとき、その数を切り捨てたときの誤差はどうなるか。四捨五入したときはどうか。それぞれの平均（mean）と分散（variance）は何か。

段階 1　誤差の合計

扱う内容：1 つの数の誤差と、1500 個の数の和

? **考えてみよう**　ある数の小数部分（fractional part）が $[0,1)$ 上の一様分布（uniform distribution）に従うとき、その数を切り捨て（truncation）たときの誤差はどうなるか。四捨五入（rounding）したときはどうか。それぞれの平均（mean）と分散（variance）を求めよ。

手を止めて、やってみよう 誤差の合計は 1500 個の独立（independent）な誤差の和である。それぞれのコンピュータについて、その平均と分散はいくらか。また、どの定理を使えばこれを正規分布（normal distribution）として扱えるか。

段階 2 大きな誤差が生じる確率

扱う内容：1500 個の数の和

考えてみよう 四捨五入するコンピュータについては、15 を標準化（standardise）して正規分布表を用いよ。切り捨てるコンピュータについては、誤差の合計はほぼ確実にどの範囲にあるか。

段階 3 加算は何回か

扱う内容：LEP に必要な加算の回数

> **考えてみよう** n 回の加算の後、LEP の誤差の標準偏差（standard deviation）はいくらか。$\mathrm{P}(|Z| < 1.645) = 0.9$ を用いて n について解け。

背後にある数学

偏りと分散

切り捨ては**偏りがある**（biased）。すなわち、その誤差の平均は 0 でなく、偏りは線形に蓄積する。四捨五入は**不偏**（unbiased）であり、その誤差はランダムウォークのように振る舞って $\sqrt{n}$ 程度で増える。100 万回の演算のあとでは、その差は 500 000 程度の誤差と 300 程度の誤差の違いになる。これは、実験における系統誤差（systematic error、たとえば校正の狂った測定器）と偶然誤差（random error）の区別と同じである。多数の測定値を平均すれば偶然誤差は小さくなるが、系統誤差にはまったく効果がない。

浮動小数点

あらゆるコンピュータは実数を有限の精度で保存し、あらゆる算術演算はその結果を丸める。現代のコンピュータは IEEE 754 規格に従っており、表現可能な最も近い数に丸める。ちょうど中間の場合には最後の桁が偶数になる方へ丸めるが、これはまさに丸めが平均的に不偏となるようにするためである。それでも、多数の数を不用意に足し合わせると精度が失われることがある。カハン（William Kahan）の**補償付き総和**（compensated summation）アルゴリズム（1965 年）は、失われた下位の桁を記録しておき、それを足し戻す。1991 年の湾岸戦争では、パトリオットミサイルの砲台が飛来するミサイルの迎撃に失敗した。その時計は時間を 0.1 秒単位で数えていたが、0.1 は 2 進法では正確に表せない数である。100 時間の稼働のあと、切り捨て誤差は約 3 分の

1 秒にまで増え、標的の位置を 0.5 キロメートル以上見誤るのに十分な大きさになっていた。

バンクーバー指数

バンクーバー証券取引所の指数は、1982 年 1 月に 1000.000 という値で創設された。1983 年 11 月までに、市場全体はおおむね上昇していたにもかかわらず、指数は約 524 にまで下がっていた。コンサルタントたちは、指数が 1 日に何千回も再計算され、そのたびに小数第 3 位まで（四捨五入ではなく）切り捨てられていたことを突き止めた。週末をかけて正しく計算し直した結果、指数は約 1098 で取引を再開した。この出来事は、数値計算の講義で定番の戒めとなっている。

次の一歩

1. **VOZ の補正。** VOZ が最後に答えに $\frac{n}{2}$ を加えるとすると、その誤差はどのような分布に従うか。これで LEP と同じくらい良くなったといえるか。

2. **小数第 1 位への四捨五入。** 数を小数第 1 位に四捨五入し、捨てられる桁が一様分布に従うとき、10 000 個の数を足したときの誤差の総和の標準偏差を求めよ。

3. **バンクーバー指数。** 指数が 1 日に 3000 回再計算され、1 回の切り捨てで平均 0.0005 が失われたとすると、取引日 250 日の 1 年間でどれだけ失われるか。

4. **厳密な分布。** 2 つの四捨五入誤差の和について、厳密な確率密度関数（三角形になる）を求めよ。3 つの場合については、その概形を描け。どのくらいの速さで正規分布らしく見えてくるか。

さらに学ぶために

- **Nicholas Higham, *Accuracy and Stability of Numerical Algorithms***（SIAM）。丸め誤差に関する定評ある本であり、第 1 章は読みやすい。
- **David Goldberg, “What every computer scientist should know about floating-point arithmetic”**, *ACM Computing Surveys* 23 (1991)。古典的な論文で、オンラインで無料で読める。
- **Tim Harford, *How to Make the World Add Up***（Bridge Street）。データにおける系統誤差を扱っている。

詳しい解説付きの解答は第 13.6 節にある：

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

毎週 1 問の STEP の問題を丁寧に解き明かす The STEP Up。メールで無料配信中：keithmcnulty.substack.com