13.6 累積起來的誤差

工作紙

題目

STEP II 1993，第 15 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

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

提示

請先自行嘗試解題。如果遇到困難，請逐一閱讀提示，每讀完一個便回到題目再試。標有停一停，試一試的提示是關鍵步驟。每個提示之後均留有空位供你演算。

> **想一想** 若某數的小數部分 U 在 $[0,1)$ 上均勻分佈（uniform），把該數截斷時誤差是多少？捨入時又是多少？兩者的平均值（mean）和方差（variance）各是多少？

階段 1 總誤差

涵蓋：單一數字的誤差，以及 1500 個數字相加

想一想 若某數的小數部分（fractional part）在 $[0, 1)$ 上服從均勻分佈（uniform distribution），把它截斷（truncate）時誤差是多少？把它捨入（round）時又是多少？分別求出兩種誤差的平均值（mean）和方差（variance）。

停一停，試一試 總誤差是 1500 個獨立（independent）誤差之和。對每部電腦而言，總誤差的平均值和方差分別是多少？哪一條定理容許你把它當作正態分佈（normal distribution）來處理？

階段 2　出現大誤差的概率

涵蓋：1500 個數字相加

> **想一想**　對於採用捨入的電腦，把 15 標準化（standardise），然後查正態分佈表。至於採用截斷的電腦，總誤差幾乎必定落在哪裏？

階段 3 要做多少次加法？

涵蓋：LEP 要做多少次加法

想一想 做了 n 次加法之後，LEP 的誤差的標準差（standard deviation）是多少？利用 $\mathrm{P}(|Z| < 1.645) = 0.9$，解出 n。

背後的數學

偏差與方差

截斷是有偏（biased）的：其誤差的平均值不是零，而偏差（bias）會線性累積。捨入是無偏（unbiased）的，其誤差的表現有如隨機漫步，增長速度相當於 $\sqrt{n}$。經過一百萬次運算後，兩者的分別就是數量級為 500 000 的誤差與數量級為 300 的誤差之別。這與實驗中系統誤差（systematic error，例如儀器校準失準）和隨機測量誤差（random error）之間的區別如出一轍：把多次測量結果取平均，可以減少隨機誤差，對系統誤差卻毫無作用。

浮點數

所有電腦都只能以有限的精確度儲存實數，每次算術運算都要把結果捨入。現代電腦遵從 IEEE 754 標準，把結果捨入至最接近的可表示數；若剛好位於兩數正中，則取末位為偶數的一個，目的正是令捨入平均而言無偏。即使如此，把大量數字相加時若不小心處理，仍會損失準確度；William Kahan 的補償求和（compensated summation）算法（1965 年）會記錄失去的低位數字，再把它們加回去。1991 年波斯灣戰爭期間，一個愛國者導彈連未能攔截一枚來襲導彈，原因是它的時鐘以十分之一秒為單位計時，而這個數無法以二進制準確儲存；連續運作 100 小時後，截斷誤差已增至約三分之一秒，足以令目標位置的估算偏差超過半公里。

溫哥華指數

溫哥華證券交易所指數於 1982 年 1 月設立，起始值為 1000.000。到 1983 年 11 月，儘管大市普遍上升，指數卻已跌至約 524。顧問發現，指數每天重新計算數千次，而每次都截斷（而非捨入）至小數點後三位。經過一個周末的正確重算，指數重開時約為 1098。這件事已成為數值計算課程中的經典警示。

延伸練習

1 **修正 VOZ。**若 VOZ 在最後把 $\frac{n}{2}$ 加到答案上，其誤差的分佈是甚麼？它現在是否與 LEP 一樣好？

2 **捨入至小數點後一位。**若把各數捨入至小數點後一位，而被捨去的數字均勻分佈，求把 10 000 個數相加時總誤差的標準差。

3 **溫哥華指數。**若指數每天重新計算 3000 次，每次截斷平均損失 0.0005，那麼在一年 250 個交易日內，指數會損失多少？

4 **精確分佈。**求兩個捨入誤差之和的精確密度函數（density，呈三角形）。若是三個誤差之和，試繪出其草圖。它多快便開始貌似正態分佈？

延伸閱讀

- **Nicholas Higham, *Accuracy and Stability of Numerical Algorithms***（SIAM）。關於捨入誤差的權威著作；第一章淺白易讀。
- **David Goldberg, “What every computer scientist should know about floating-point arithmetic”**, *ACM Computing Surveys* 23 (1991)。經典之作，可在網上免費閱讀。
- **Tim Harford, *How to Make the World Add Up***（Bridge Street），探討數據中的系統誤差。

完整的引導式解答見第 13.6 節：

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

每星期一道 STEP 題目，完整講解：The STEP Up，以電郵免費發送，訂閱網址：keithmcnulty.substack.com