詳細解答

STEP II 1993，第 15 題

概率與統計 • 中心極限定理

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

本題在 *Beyond STEP* 第 13.6 節累積起來的誤差中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch13/errors-add-up.html

關鍵思路

總誤差是許多微小而獨立的誤差之和，每個數各有一個。因此，先求出單獨一個數的誤差的平均值（mean）和方差（variance），再把所有數的結果加起來，然後利用中心極限定理（Central Limit Theorem）把總誤差視為近似服從正態分佈（normal distribution）。兩部電腦之間的全部差別都來自平均值：一部機器的誤差平均而言互相抵銷，另一部的誤差卻全都偏向同一方。

解答

一個數的誤差

設某數為 $a = m + U$，其中 $m = \lfloor a \rfloor$ 是整數，U 是它的小數部分（fractional part），且 $U \sim \mathrm{U}[0,1)$。我們把誤差定義為（近似值）−（真實值）；正負號的約定並不影響誤差的大小。

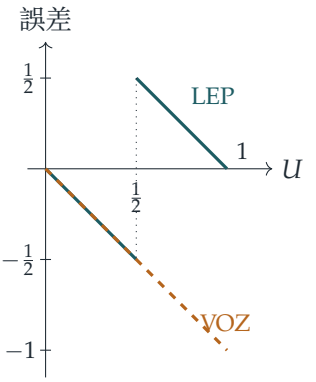
LEP（四捨五入，*rounding*）。若 $U < \frac{1}{2}$，該數會被換成 m，誤差為 $-U$；若 $U \geq \frac{1}{2}$，該數會被換成 $m+1$，誤差為 $1-U$。所以誤差為

$$X = \begin{cases} -U, & 0 \leq U < \frac{1}{2}, \\ 1-U, & \frac{1}{2} \leq U < 1, \end{cases}$$

它以恆定的密度取 $\left(-\frac{1}{2}, \frac{1}{2}\right]$ 內的每一個值：X 在 $\left[-\frac{1}{2}, \frac{1}{2}\right]$ 上服從均勻分佈（uniform dis-

tribution）（即圖中的鋸齒形）。

VOZ（截尾，*truncation*）。該數會被換成 m，所以誤差為 $Y = -U$，在 $(-1, 0]$ 上服從均勻分佈。



在長度為 1 的區間上的均勻分佈，不論區間位於何處，方差都是 $\frac{1}{12}$。所以

$$\mathrm{E}(X) = 0, \quad \mathrm{Var}(X) = \tfrac{1}{12}; \qquad \mathrm{E}(Y) = -\tfrac{1}{2}, \quad \mathrm{Var}(Y) = \tfrac{1}{12}.$$

（順帶記下：$\mathrm{Var}(U) = \int_0^1 u^2 \,\mathrm{d}u - \left(\frac{1}{2}\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$。）

把 1500 個數相加

答案的誤差等於各個數的誤差之和，因為電腦把近似值相加時是完全準確的。各個小數部分互相獨立（independent），所以各個誤差亦互相獨立，它們的方差可以相加。對於 $n = 1500$ 個數，

$$\mathrm{Var}(\text{總誤差}) = \frac{1500}{12} = 125, \qquad \text{標準差} \sqrt{125} = 5\sqrt{5} \approx 11.18.$$

根據中心極限定理，1500 個獨立同分佈（independent, identically distributed）的誤差之和非常接近正態分佈。

LEP。總誤差 E_L 近似服從 $\mathrm{N}(0, 125)$，所以

$$\begin{aligned}\mathrm{P}(|E_L| > 15) &\approx \mathrm{P}\left(|Z| > \frac{15}{\sqrt{125}}\right) = \mathrm{P}(|Z| > 1.342) \\ &= 2\big(1 - \Phi(1.342)\big) \approx 2(1 - 0.9102) \approx 0.18.\end{aligned}$$

這裏毋須作連續性校正（continuity correction）：這些誤差是連續隨機變量（continuous random variable），而不是整數。

VOZ。總誤差 E_V 近似服從 $\mathrm{N}(-750, 125)$。只有當 $E_V \geq -15$ 時，誤差的大小才不超過 15，而這個值比平均值高出

$$\frac{-15 - (-750)}{\sqrt{125}} = \frac{735}{11.18} \approx 66$$

個標準差（standard deviation）。這個概率完全可以忽略不計，所以

$$\mathrm{P}(|E_V| > 15) \approx 1.$$

截尾永遠向下偏，平均每次偏 $\frac{1}{2}$，所以加了 1500 個數之後，VOZ 的答案大約相差 750。

想一想 兩部機器所產生的誤差方差相同。為甚麼答案卻相差這麼遠？

總誤差的離散程度是一樣的，大約是 ± 11，但它的中心卻不一樣。LEP 的誤差以 0 為中心，所以總誤差在 0 附近徘徊。VOZ 的誤差平均值全都是 $-\frac{1}{2}$，所以偏差（bias）與 n 成正比地增長，而隨機的離散程度只按 $\sqrt{n}$ 的速度增長。偏差輕易勝出。

LEP 可以做多少次加法？

假設 LEP 把 n 個數相加。它的總誤差近似服從 $\mathrm{N}\left(0, \frac{n}{12}\right)$，而我們需要

$$\mathrm{P}(|E_L| < 10) = 2\Phi\left(\frac{10}{\sqrt{n/12}}\right) - 1 \;\geq\; 0.9 \quad\Longleftrightarrow\quad \Phi\left(\frac{10}{\sqrt{n/12}}\right) \geq 0.95.$$

由於 Φ 是遞增函數，而且 $\Phi(1.6449) = 0.95$，上式即

$$\frac{10}{\sqrt{n/12}} \geq 1.6449 \iff n \leq 12\left(\frac{10}{1.6449}\right)^2 \approx 443.5.$$

這個概率隨 n 增加而下降，所以當 $n \leq 443$ 時它仍然不小於 0.9，到 $n = 444$ 時才首次跌破 0.9。因此 LEP 可以把 **443** 個數相加。（如果你把一次「加法」理解為使用一次 $+$ 號，那麼把 443 個數相加需要用 442 次；只要說明你採用哪一種理解，兩種都可以接受。）

驗算

- 1.342 這個數值合理嗎？總誤差 15 比一個標準差（11.18）略多，而正態分佈中約有 18% 的部分距離平均值超過 1.34 個標準差。這大致吻合。
- 數值檢驗。模擬 10^5 批、每批 1500 個四捨五入後的數，得出 $\mathrm{P}(|E_L| > 15) \approx 0.179$。按均勻分佈之和的精確分佈計算得 0.1797，而正態近似所得的亦是 0.1797。至於第二部分，精確概率在 $n = 443$ 時為 0.90019，在 $n = 444$ 時為 0.89981，印證了這個分界點。

常見錯誤

- 把 LEP 的誤差當作在 $[0, \frac{1}{2}]$ 上均勻分佈：四捨五入可以向上或向下，所以誤差是在 $[-\frac{1}{2}, \frac{1}{2}]$ 上均勻分佈。

- 忘記 VOZ 每個數的誤差平均值是 $-\frac{1}{2}$，因而得出與 LEP 相同的答案。
- 在最後一部分用了 $\Phi(z) = 0.9$ 而不是 0.95：事件 $|E| < 10$ 是雙側的，所以每邊尾部各佔 0.05。

STEP 之外

這是浮點運算（floating-point arithmetic）中捨入誤差（rounding error）的經典分析：無偏的捨入法令累積誤差按 $\sqrt{n}$ 的速度增長，而有偏的規則則令它按 n 的速度增長。這正是 IEEE 754 標準採用「逢半取偶」（round half to even）而不是逢半必進的原因。n 個均勻分佈隨機變量之和有一個精確的分佈（Irwin–Hall 分佈），觀察它變成正態分佈的速度有多快，是認識中心極限定理的一個好途徑；Grimmett 與 Welsh 所著的 *Probability: An Introduction*（概率論入門教材）證明了這條定理，並對它加以討論。

Beyond STEP 詳細解答：https://beyondstep.keithmcnulty.org/zh-hant/。解答文字採用 CC0 1.0；題目本身版權屬 Cambridge University Press & Assessment 所有（©），本文件不予轉載。由機器從英文翻譯而成；尚未經審閱。