解答例

STEP II 1993、第 15 問

確率・統計 ・ 中心極限定理

> **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

この問題は、*Beyond STEP* の第 13.6 節「**積み重なる誤差**」で詳しく解説されている：
https://beyondstep.keithmcnulty.org/ja/ch13/errors-add-up.html

鍵となる考え方

誤差の合計は、各数に 1 つずつ生じる多数の小さな独立な誤差の和である。そこで、*1 つの*数に生じる誤差の平均と分散を求め、それをすべての数について足し合わせ、中心極限定理（Central Limit Theorem）を用いて合計を近似的に正規分布に従うものとして扱う。2 台のコンピュータの違いは、すべて平均から生じる。一方の機械の誤差は平均すると打ち消し合うが、もう一方の誤差はすべて同じ向きに偏るのである。

解答

1 つの数に生じる誤差

数を $a = m + U$ とおく。ここで $m = \lfloor a \rfloor$ は整数、U はその小数部分であり、$U \sim \mathrm{U}[0,1)$ である。誤差を（近似値）−（真の値）と定義する。符号の取り方は誤差の大きさには影響しない。

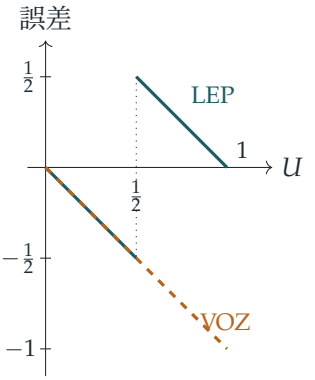
LEP（**四捨五入**（*rounding*））。$U < \frac{1}{2}$ ならば数は m に置き換えられ、誤差は $-U$ である。$U \geq \frac{1}{2}$ ならば $m+1$ に置き換えられ、誤差は $1-U$ である。よって誤差は

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

であり、$(-\frac{1}{2}, \frac{1}{2}]$ の各値を一定の密度でとる。すなわち X は $[-\frac{1}{2}, \frac{1}{2}]$ 上の一様分布に従

う（図ののこぎり波）。

VOZ（**切り捨て**（*truncation*））。数は m に置き換えられるので、誤差は $Y = -U$ であり、$(-1, 0]$ 上の一様分布に従う。



長さ 1 の区間上の一様分布（uniform distribution）の分散（variance）は、区間がどこにあっても $\frac{1}{12}$ である。よって

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

（念のため記しておくと、$\mathrm{Var}(U) = \int_0^1 u^2 \,\mathrm{d}u - \left(\frac{1}{2}\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$ である。）

1500 個の数を足す

コンピュータは近似値どうしを正確に足し合わせるので、答えに含まれる誤差は個々の数に生じる誤差の和である。小数部分は互いに独立であるから、個々の誤差も独立（independent）であり、それらの分散は足し合わせることができる。$n = 1500$ 個の数に対して

$$\mathrm{Var}(\text{誤差の合計}) = \frac{1500}{12} = 125, \qquad \text{標準偏差} \sqrt{125} = 5\sqrt{5} \approx 11.18.$$

中心極限定理により、独立で同一の分布に従う 1500 個の誤差の和は、正規分布（normal distribution）に非常に近い。

LEP。誤差の合計 E_L は近似的に $\mathrm{N}(0, 125)$ に従うので、

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

連続性補正（continuity correction）は必要ない。誤差は整数ではなく連続型確率変数だからである。

VOZ。誤差の合計 E_V は近似的に $\mathrm{N}(-750, 125)$ に従う。誤差の大きさが 15 以下となるのは $E_V \geq -15$ のときに限られ、これは平均より

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

標準偏差も上にある。この確率はまったく無視できるほど小さいので、

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

切り捨ては常に下向きに、平均して $\frac{1}{2}$ ずつずれるので、1500 個の数を足すと VOZ はおよそ 750 だけずれてしまう。

> **考えてみよう**　どちらの機械も同じ分散の誤差を生じる。それなのに、なぜ答えがこれほど違うのだろうか。

誤差の合計のばらつきは同じで、およそ ±11 程度であるが、その**中心**は同じではない。LEP の誤差は 0 を中心としているので、合計は 0 の周りをふらつく。VOZ の誤差はすべて平均 $-\frac{1}{2}$ をもつので、偏り（bias）は n に比例して増大する一方、ランダムなばらつきは $\sqrt{n}$ 程度でしか増えない。偏りのほうが圧倒的に勝るのである。

LEP は何回足し算できるか

LEP が n 個の数を足すとする。その誤差の合計は近似的に $\mathrm{N}\left(0, \frac{n}{12}\right)$ に従い、次が必要である。

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

Φ は増加関数で $\Phi(1.6449) = 0.95$ であるから、これは

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

と同値である。確率は n が増えると減少するので、$n \leq 443$ では 0.9 以上であり、$n = 444$ で初めて 0.9 を下回る。よって LEP は **443** 個の数を足すことができる。（「足し算」を $+$ 記号の使用 1 回と数えるならば、443 個の数を足すには 442 回の足し算が必要である。どちらの解釈を採るかを明記すれば、どちらでも構わない。）

検算

- **1.342 は妥当か。**誤差の合計 15 は 1 標準偏差（11.18）を少し超える程度であり、正規分布では平均から 1.34 標準偏差より遠くにある部分がおよそ 18% である。おおむね妥当である。

- **数値的な確認。** 1500 個の四捨五入された数の組を 10^5 回シミュレーションしたところ、$\mathrm{P}(|E_L| > 15) \approx 0.179$ となった。一様分布の和の正確な分布からは 0.1797 が得られ、正規近似による値 0.1797 と一致する。後半については、正確な確率は $n = 443$ で 0.90019、$n = 444$ で 0.89981 であり、境界が正しいことが確かめられる。

よくある間違い

- LEP の誤差を $[0, \frac{1}{2}]$ 上の一様分布とすること。四捨五入では上にも下にもずれうるので、誤差は $[-\frac{1}{2}, \frac{1}{2}]$ 上の一様分布である。
- VOZ の 1 つの数あたりの平均 $-\frac{1}{2}$ を忘れ、LEP と同じ答えを出してしまうこと。
- 最後の小問で 0.95 の代わりに $\Phi(z) = 0.9$ を用いること。事象 $|E| < 10$ は両側なので、それぞれの裾に 0.05 ずつが割り当てられる。

STEP のその先へ

これは浮動小数点演算（floating-point arithmetic）における丸め誤差（rounding error）の古典的な解析である。偏りのない丸めでは累積誤差は $\sqrt{n}$ 程度で増大するが、偏りのある規則では n 程度で増大する。IEEE 754 規格が、ちょうど半分のときに常に切り上げるのではなく「偶数への丸め（round half to even）」を採用しているのはこのためである。n 個の一様分布の和には正確な分布（アーウィン・ホール分布（Irwin–Hall distribution））があり、それがどれほど速く正規分布に近づくかを調べることは、中心極限定理への良い入り口となる。Grimmett and Welsh, *Probability: An Introduction*（確率論入門）では、この定理が証明され、詳しく論じられている。

Beyond STEP 解答集：https://beyondstep.keithmcnulty.org/ja/。解答本文は CC0 1.0 である。問題そのものは © Cambridge University Press & Assessment に帰属し、ここには掲載していない。英語版から機械翻訳したものであり、まだ校閲を経ていない。