解答例

STEP III 2010、第 1 問

純粋数学 • 平均と分散

> **STEP III 2010、問 1**
>
> 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/10-s3-q1

鍵となる考え方

A と B は最初の n 個の数の平均（mean）と分散（variance）であり、C と D は数をもう一つ加えたときの平均と分散である。分散は「2 乗の平均から平均の 2 乗を引いたもの」という形で扱うのが最も楽であり、小問 (ii) がまさにそれを与えてくれる。これを使うと、D は x_{n+1} の二次式（quadratic）として表され、二つの不等式（inequality）はどちらもその二次式についての主張となる。

解答

以下では簡単のため $x = x_{n+1}$ と書く。

小問 (i)

定義より $\sum_{k=1}^{n} x_k = nA$ であるから、

$$C = \frac{1}{n+1}\left(\sum_{k=1}^{n} x_k + x\right) = \frac{nA + x}{n+1}.$$

小問 (ii)

2 乗を展開し、$\sum_{k=1}^{n} x_k = nA$ と $\sum_{k=1}^{n} 1 = n$ を用いると、

$$\begin{aligned} B &= \frac{1}{n}\sum_{k=1}^{n}(x_k^2 - 2Ax_k + A^2) = \frac{1}{n}\sum_{k=1}^{n} x_k^2 - \frac{2A}{n}\cdot nA + \frac{1}{n}\cdot nA^2 \\ &= \frac{1}{n}\sum_{k=1}^{n} x_k^2 - A^2, \end{aligned}$$

となり、示すべき式が得られた。

小問 (iii)

小問 (ii) は任意の数の並びについて成り立つので、これを（平均が C である）$n+1$ 個の数 $x_1, \ldots, x_{n+1}$ に適用すると

$$D = \frac{1}{n+1}\sum_{k=1}^{n+1} x_k^2 - C^2.$$

を得る。再び小問 (ii) より $\sum_{k=1}^{n} x_k^2 = n(B+A^2)$ であるから、$\sum_{k=1}^{n+1} x_k^2 = n(B+A^2)+x^2$ である。C に小問 (i) を用いると、

$$D = \frac{nB+nA^2+x^2}{n+1} - \frac{(nA+x)^2}{(n+1)^2}.$$

両辺に $(n+1)^2$ を掛けて展開する：

$$\begin{aligned}(n+1)^2 D &= (n+1)(nB+nA^2+x^2) - (n^2A^2+2nAx+x^2)\\ &= n(n+1)B + (n^2+n-n^2)A^2 - 2nAx + (n+1-1)x^2\\ &= n(n+1)B + n(x^2-2Ax+A^2)\\ &= n(n+1)B + n(x-A)^2.\end{aligned}$$

したがって

$$\boxed{D = \frac{n}{n+1}B + \frac{n}{(n+1)^2}(x_{n+1}-A)^2.}$$

一つ目の不等式。 $n+1$ を掛けると

$$(n+1)D = nB + \frac{n}{n+1}(x_{n+1}-A)^2 \;\geq\; nB,$$

となる。最後の項は平方の非負の定数倍だからである。これは x_{n+1} のすべての値について成り立ち、等号が成立するのはちょうど $x_{n+1} = A$ のときである。

? **考えてみよう** 負の数の平方根を気にすることなく、$D < B$ を x_{n+1} についての条件に書き換えるにはどうすればよいか？

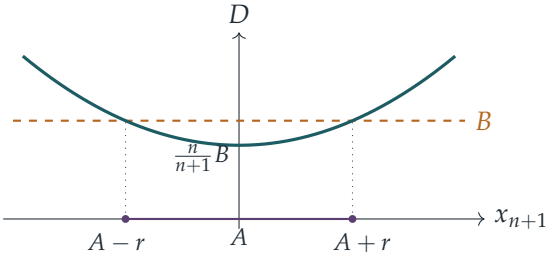
二つ目の不等式。 まず、B は平方の平均なので $B \geq 0$ であり、したがって $\sqrt{(n+1)B/n}$ は実数であることに注意する。枠で囲んだ式を用いると、以下の各変形はいずれも逆にたどることができる（両辺から同じ量を引くことと、正の数 $(n+1)^2/n$ を掛けるこ

としかしていない）：

$$\begin{aligned} D < B &\iff \frac{n}{(n+1)^2}(x_{n+1}-A)^2 < B - \frac{n}{n+1}B = \frac{B}{n+1} \\ &\iff (x_{n+1}-A)^2 < \frac{(n+1)B}{n} \\ &\iff |x_{n+1}-A| < \sqrt{\frac{(n+1)B}{n}} \\ &\iff A - \sqrt{\frac{(n+1)B}{n}} < x_{n+1} < A + \sqrt{\frac{(n+1)B}{n}}. \end{aligned}$$

3 番目の変形では両辺の非負の平方根をとっているが、これは非負の数どうしの狭義の不等式を保つ。以上が求める「必要十分（if and only if）」の主張である。（$B = 0$ のときはこの区間は空であり、実際このとき常に $D \geq 0 = B$ となる。）

下の図は、$n = 3, A = 0, B = 1$ の場合に D を x_{n+1} の関数（function）として描いたものである。これは $x_{n+1} = A$ で最小値 $\frac{n}{n+1}B$ をとる放物線（parabola）であり、上で求めた区間上でちょうど水準 B を下回る。



ここで $r = \sqrt{(n+1)B/n}$ である。

検算

- $n = 1$ とする。このとき $B = 0$ であり、公式は $D = \frac{1}{4}(x_2 - x_1)^2$ を与える。これは二つの数 x_1, x_2 について正しい。実際、その平均は 2 数の中点である。
- $x_1, x_2, x_3 = 0, 0, 3$ とすると $A = 1, B = 2$ であり、ここに $x_4 = 1 = A$ を加える。直接計算すると $D = \frac{1}{4}(1+1+4+0) = \frac{3}{2} = \frac{3}{4}B$ となり、これは最小値である。
- 20 000 個の無作為な数の並びについて、同値性 $D < B \iff (x_{n+1}-A)^2 < (n+1)B/n$ を含む三つの結果すべてを（分数を用いて）厳密に確かめ、さらに区間の両端が有理数となる 216 通りの場合について、両端で $D = B$ となることを確認した。

よくある間違い

- 小問 (ii) で展開する際に、$\sum_{k=1}^{n} A^2 = nA^2$ であって A^2 ではないことを忘れる。
- D を $\sum x_k^2$ を用いて表したところで止めてしまう。問題は D を B, A, x_{n+1}, n で表すことを求めているので、この和は小問 (ii) を用いて置き換えなければならない。
- $(x_{n+1} - A)^2 < k$ の平方根をとって $x_{n+1} - A < \sqrt{k}$ だけを得てしまい、区間の下端を失う。

STEP のその先へ

D の公式は**更新式**（updating formula）である。これを使えば、コンピュータはすべてのデータを保存することなく、データの流れの平均と分散を値を一つずつ受け取りながら保持できる。この考え方は Welford のアルゴリズムで洗練されている。項 $n(x_{n+1} - A)^2/(n+1)^2$ は、統計学における分散分析（analysis of variance）の基礎にある、分散を「群内」と「群間」の部分に分解する考え方の小さな一例である。詳しくは Grimmett and Welsh, *Probability: An Introduction*（確率論入門）を参照されたい。

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。