Maximize standard deviation given range

Thursday, April 29, 2021 15:29

Question:

n is a natural number at least 3. $a_0, a_1, \dots, a_{n-1}$ are real numbers. Mean (and sum) of a_i is 0. Range of a_i is r, $r \ge 0$.

(The cases of $n = 1,2$ are easy. There is nothing to be maximized, (sum of squares) having only one possible value to take.)

Maximize $\sigma = \sqrt{\frac{\sum_{i=0}^{n-1} a_i^2}{n}}$.

Solution 1 (Complete):

Assume there is such an arrangement with range equals to r. Without loss of generality, let us assume that $a_0 \le a_1 \le \dots \le a_{n-1}$.
Our goal is to get an arrangement with sum of squares at least that of the original arrangement, while having only at most 3 distinct values among a_i.
To achieve this, we establish the following lemma:

(Inspired by statistics - How to maximize Std Dev given a range of possible values, a number of values, and a specific mean? - Mathematics Stack Exchange and **reminded by *Fu Jia Cheng Charles*.**)

If $a \le b$ and $\delta \ge 0$,

$$(a-\delta)^2 + (b+\delta)^2 \ge a^2 + b^2.$$

Proof:

$$a^2 - 2a\delta + \delta^2 + b^2 + 2b\delta + \delta^2 = a^2 + b^2 + 2(b-a)\delta + \delta^2 \ge a^2 + b^2$$

Now, we use an *C++-like* algorithm (language not important), fixing a_0 and a_{n-1} but trying to push middle a_i's to the bounds a_0 and a_{n-1}, with the help from the above lemma.

$k_0 = 1, k_1 = n-2$
`while` $(k_0 < k_1)$
  $\delta = \min\{x_{k_0} - x_0, x_{n-1} - x_{k_1}\}$
  x_{k_0} `-=` δ, x_{k_1} `+=` δ
  `if` $(\delta == x_{k_0} - x_0)$
    k_0`++`
  `else` `// This line can be "if (`δ `==` $x_{n-1} - x_{k_1}$`)", but the risk is having` $k_0 > k_1$ `after the loop ends.`
    k_1`--`

- Each time we try to push x_{k_0} or x_{k_1} to reach lower bound a_0 and upper bound a_{n-1}, by $(x_{k_0} - \delta)^2 + (x_{k_1} + \delta)^2 \ge x_{k_0}^2 + x_{k_1}^2$. Within the whole procedure, $x_0 \le \dots \le x_{n-1}$ is maintained, so as the conditions $x_0 = \dots = x_{k_0 - 1}$ and $x_{k_1+1} = \dots = x_{n-1}$.
- Originally and during the whole course of the procedure, $k_0 \le k_1$; after the loop, (we must have) $k_0 = k_1$. Hence we may simply use a new variable (for the ease of presentation) $p = k_0 = k_1$. Let $q = n - 1 - p$ to further simplify expressions (actually I mean making the expressions more symmetrically).
  - Note that $0 < p, q < n-1$.
- Therefore, we have $0 = \sum_{i=0}^{n-1} x_i = \underbrace{x_0 + \dots + x_{p-1}}_{p} + x_p + \underbrace{x_{p+1} + \dots + x_{n-1}}_{n-1-p} = px_0 + x_p + qx_{n-1}$, and

$$\sum_{i=0}^{n-1} x_i^2 = px_0^2 + x_p^2 + qx_{n-1}^2.$$

Now we maximize $\sum_i x_i^2$ keeping these constraints.

$$x_0 = \dots = x_{p-1} \le x_p \le x_{p+1} = \dots = x_{n-1} = x_0 + r$$

$$x_0 \le 0 - qx_{n-1} - px_0 = -(p+q)x_0 - qr \le x_0 + r$$

$$-\frac{(q+1)r}{n} = -\frac{(q+1)r}{p+q+1} \le x_0 \le -\frac{qr}{p+q+1} = -\frac{qr}{n}$$

$$-\frac{pr}{n(n-1)} = \frac{-(n-1)(q+1)r + nqr}{n(n-1)} = -\frac{(q+1)r}{n} + \frac{qr}{n-1} \le x_0 + \frac{qr}{n-1} \le -\frac{qr}{n} + \frac{qr}{n-1} = \frac{qr}{n(n-1)}$$

$$0 \le \left(x_0 + \frac{qr}{n-1}\right)^2 \le \left(\frac{\max\{p,q\}\,r}{n(n-1)}\right)^2$$

$$\sum_{i=0}^{n-1} x_i^2 = px_0^2 + (p+q)^2x_0^2 + 2qr(p+q)x_0 + q^2r^2 + qx_0^2 + 2qrx_0 + qr^2 = (p+q)(p+q+1)x_0^2 + 2qr(p+q+1)x_0 + (q^2+q)r^2 = n(n-1)x_0^2 + 2nqrx_0 + (q^2+q)r^2 = n(n-1)\left(\left(x_0 + \frac{qr}{n-1}\right)^2 - \left(\frac{qr}{n-1}\right)^2 + \frac{(q^2+q)r^2}{n(n-1)}\right)$$

$$= n(n-1)\left(\left(x_0 + \frac{qr}{n-1}\right)^2 + \frac{r^2}{n-1}\left(\frac{q^2+q}{n} - \frac{q^2}{n-1}\right)\right) \le \frac{\max\{p,q\}^2 r^2}{n(n-1)} + \frac{r^2}{n-1}\left((n-1)(q^2+q) - nq^2\right) = r^2\left(\frac{\max\{p,q\}^2}{n(n-1)} + \left(\frac{q(n-1-q)}{n-1}\right)\right) = \frac{r^2}{n(n-1)}\left(\max\{p,q\}^2 + pqn\right) = \frac{r^2}{n(n-1)} \cdot \max\{p^2 + pqn, q^2 + qpn\}$$

$$= \frac{r^2}{n(n-1)} \cdot \max\{p^2 + p(n-1-p)n, \dots_{\text{by symmetry}}\} = \frac{r^2}{n(n-1)} \cdot \max\{p^2(1-n) + np(n-1), \dots_{\text{by symmetry}}\} = \frac{r^2}{n} \cdot \max\{-p^2 + np, -q^2 + nq\} = \frac{r^2}{n}\left(\max\left\{-\left(p - \frac{n}{2}\right)^2, -\left(q - \frac{n}{2}\right)^2\right\} + \frac{n^2}{4}\right) = \frac{r^2}{n}\left(-\min\left\{\left(p - \frac{n}{2}\right)^2, \left(n - 1 - p - \frac{n}{2}\right)^2\right\} + \frac{n^2}{4}\right)$$

$$= \frac{r^2}{n}\left(-\min\left\{\left(p - \frac{n}{2}\right)^2, \left(p - \left(\frac{n}{2} - 1\right)\right)^2\right\} + \frac{n^2}{4}\right)$$

- When n is even, $\min\{\text{(see above)}\}$ is at least 0. When n is odd, it is at lesat $\left(\frac{1}{2}\right)^2$, or you can refer to below for an **unnecessary** alternative manipulation / proof.

$$= \frac{r^2}{n}\left(-\frac{\left(p - \frac{n}{2}\right)^2 + \left(p - \left(\frac{n}{2} - 1\right)\right)^2}{2} + \frac{\left|\left(p - \frac{n}{2} + p - \frac{n}{2} + 1\right)(-1)\right|}{2} + \frac{n^2}{4}\right) = \frac{r^2}{n}\left(-\frac{p^2 + p^2 - (n+n-2)p + \frac{n^2}{4} + \frac{n^2 - 4n + 4}{4}}{2} + \frac{n^2}{4} + \left|p - \frac{n-1}{2}\right|\right)$$

$\because \min\{a,b\} = \frac{a+b}{2} - \frac{|a-b|}{2}$

$$= \frac{r^2}{n}\left(-p^2 + (n-1)p + \frac{2n-2}{4} + \left|p - \frac{n-1}{2}\right|\right) = \frac{r^2}{n}\left(-\left(p - \frac{n-1}{2}\right)^2 + \frac{n^2 - 2n + 1 + 2n - 2}{4} + \left|p - \frac{n-1}{2}\right|\right) = \frac{r^2}{n}\left(-\left(p - \frac{n-1}{2}\right)^2 + \left|p - \frac{n-1}{2}\right| + \frac{n^2-1}{4}\right)$$

- When n is odd, $p - \frac{n-1}{2}$ is integer and we can prove that $-x^2 + |x| \le 0, x = p - \frac{n-1}{2}$. May refer to this graph, and notice that the extremum is $\frac{1}{4}$, but it is the case of n is even, which is already handled above (no need to make it so complicated to do).

$$\therefore \sum_{i=0}^{n-1} x_i^2 \le \begin{cases} \frac{r^2}{n}\left(\frac{n^2}{4}\right) = \frac{r^2 n}{4}, n \text{ is even} \\ \frac{r^2}{n}\left(\frac{n^2-1}{4}\right) = \frac{r^2(n-1)(n+1)}{4n}, n \text{ is odd} \end{cases}$$

(It is also true when $n = 1$ or 2.)

Construction of an example (*I am too lazy to prove uniqueness*):

- When n is even,
  - $(a_0, \dots, a_{n-1}) = \left(\underbrace{-\frac{r}{2}, \dots, -\frac{r}{2}}_{\frac{n}{2}}, \underbrace{\frac{r}{2}, \dots, \frac{r}{2}}_{\frac{n}{2}}\right)$.
- When n is odd,
  - $(a_0, \dots, a_{n-1}) = \left(\underbrace{-\frac{n-1}{n} \cdot \frac{r}{2}, \dots, -\frac{n-1}{n} \cdot \frac{r}{2}}_{\frac{n+1}{2}}, \underbrace{\frac{n+1}{n} \cdot \frac{r}{2}, \dots, \frac{n+1}{n} \cdot \frac{r}{2}}_{\frac{n-1}{2}}\right)$
    - $\sum_{i=0}^{n-1} a_i = -\frac{n^2-1}{4n} \cdot r + \frac{n^2-1}{4n} \cdot r = 0$
    - $\sum_{i=0}^{n-1} a_i^2 = \left(\frac{n+1}{2}\left(\frac{n-1}{n}\right)^2 + \frac{n-1}{2}\left(\frac{n+1}{n}\right)^2\right)\frac{r^2}{4} = \frac{r^2}{4} \cdot \frac{(n-1)(n+1)}{n^2}\left(\frac{n+1}{2} + \frac{n-1}{2}\right) = \frac{r^2(n-1)(n+1)}{4n}$
- (Also true when $n = 1$ or 2.)

$$\therefore \sigma = \sqrt{\frac{\sum_{i=0}^{n-1} x_i^2}{n}} \le \begin{cases} \frac{r}{2}, n \text{ is even} \\ \frac{r}{2}\sqrt{1 - \frac{1}{n^2}}, n \text{ is odd} \end{cases}, \forall n \in \mathbb{N}$$

Question restatement:

$n \in \mathbb{N}_{\ge 3}, a_0, \dots, a_{n-1} \in \mathbb{R}, \max_{0 \le i < n} a_i - \min_{0 \le i < n} a_i = r \in \mathbb{R}_{\ge 0}$.

(We do not assume $\bar{a} = 0$ here.)

Maximize

$$\sigma^2 = \frac{1}{n}\sum_{i=0}^{n-1}(a_i - \bar{a})^2 = \frac{\sum_{i=0}^{n-1}\left(a_i - \frac{\sum_{i=0}^{n-1} a_i}{n}\right)^2}{n}.$$

Solution 2 (When n is even):

(By Mr. Cheng Tak Sum)

- Instead of assuming $\bar{a} = 0$ as in the first question version, we try to let $\min_{0 \le i < n} a_i = -\frac{r}{2}$, WLOG. Then, $\max_{0 \le i < n} a_i = \frac{r}{2}$.
- $\sigma^2 = \frac{1}{n}\sum_{i=0}^{n-1}(a_i^2 - 2\bar{a}a_i + \bar{a}^2) = \overline{a^2} - 2\bar{a}\bar{a} + \bar{a}^2 = \overline{a^2} - \bar{a}^2 = \frac{1}{n}\left(\frac{r^2}{4} + \sum_{i=1}^{n-2} a_i^2 + \frac{r^2}{4}\right) - \frac{1}{n^2}\left(\sum_{i=1}^{n-2} a_i\right)^2 \le \frac{1}{n}\left(\frac{r^2}{4} + \sum_{i=1}^{n-2}\frac{r^2}{4} + \frac{r^2}{4}\right) - 0 = \frac{r^2}{4}$
- $\sigma \le \frac{r}{2}$
  - Equality attained only when there are $\frac{n}{2}$ many $-\frac{r}{2}$ and $\frac{n}{2}$ many $\frac{r}{2}$ among a_i's. When $r = 0$ each a_i would have to be 0, but we have n many -0 and n many 0, just for your information.