12-1-2010

Computing the Range of Variance-to-Mean Ratio under Interval and Fuzzy Uncertainty

Sio-Long Lo
Macau University of Science and Technology

Gang Xiang
University of Texas at El Paso

Comments:
Technical Report: UTEP-CS-10-28a
Published in *Proceedings of the 30th Annual Conference of the North American Fuzzy Information Processing Society NAFIPS'2011*, El Paso, Texas, March 18-20, 2011.

Computing the Range of Variance-to-Mean Ratio under Interval and Fuzzy Uncertainty

Sio-Long Lo
Faculty of Information Technology
Macau University of Science and Technology (MUST)
Avenida Wai Long, Taipa, MACAU SAR, China
Email: akennetha@gmail.com

Gang Xiang
Philips Healthcare Informatics - EII
4050 Rio Bravo, Suite 200
El Paso, TX 79902
Email: gxiang@sigmaxi.net

***Abstract*—In many practical problems such as radar imaging, it is useful to compute the variance-to-mean ratio. The need is important because for the sum of k identical independent signal components, both the variance and the mean are multiplied by k, so this ratio is independent on k and thus, provides useful information about the components. In practice, we only know the samples values with uncertainty. It is therefore necessary to compute the variance-to-mean ratio under this uncertainty. In this paper, we present efficient algorithms for computing this ratio under interval and fuzzy uncertainty.**

I. Formulation of the Problem

Need for variance-to-mean ratio. In engineering and scientific practice, the usual way to process the measurement results or other estimates $x_1, \ldots, x_n$ of the same quantity is to compute their mean $E = \dfrac{1}{n} \cdot \sum\limits_{i=1}^{n} x_i$ and their variance $V = \dfrac{1}{n} \cdot \sum\limits_{i=1}^{n} (x_i - E)^2$; see, e.g., [7].

In many practical problems, it is useful to compute the variance-to-mean ratio $R \stackrel{\text{def}}{=} V/E$. This ratio – also known as *index of dispersion*, *dispersion index*, *coefficient of dispersion*, or *Fano factor* – is used to quantify whether a set of observed occurrences are clustered or dispersed compared to a standard statistical model; see, e.g., [1], [9].

This ratio is also useful, e.g., in radar imaging, to detect the presence of signal in noisy environments. This radar use is based on the assumption that, in the absence of the signal, the noise s consists of several independent components $s = s_1 + \ldots + s_k$,

- the mean $E[s]$ of the noise is equal to the sum of the means $E[s] = E[s_1] + \ldots + E[s_k]$ and
- the variance $V[s]$ of the signal is equal to the sum of the variances $V[s] = V[s_1] + \ldots + V[s_k]$; see, e.g. [7], [8].

In particular, when the noise consists of k identical independent signal components, both the variance and the mean are multiplied by k: $E[s] = k \cdot E[s_i]$ and $V[s] = k \cdot V[s_i]$. Thus, we get $V[s]/E[s] = V[s_i]/E[s_i]$.

When the components are not identical, we get a similar formula with average ratios. Indeed, we have $E[s] = k \cdot E_{av}$ and $V[s] = k \cdot V_{av}$, where E_{av} and V_{av} are the average mean and variance of different components:

$$E_{av} \stackrel{\text{def}}{=} \frac{E[s_1] + \ldots + E[s_k]}{k}; \quad V_{av} \stackrel{\text{def}}{=} \frac{V[s_1] + \ldots + V[s_k]}{k}.$$

Thus,

$$\frac{V[s]}{E[s]} = \frac{V_{av}}{E_{av}}.$$

When the signal *is* present, the variance-to-mean ratio changes. Thus, the difference between the observed ratio $\dfrac{V[s]}{E[s]}$ to the value $\dfrac{V_{av}}{E_{av}}$ corresponding to pure noise is an indication that there is a signal.

Need to take uncertainty into account. Traditional statistical estimates – like the above estimates for E and V – are based on the simplifying assumption that we know the exact values of the observations $x_1, \ldots, x_n$. In practice, the sample values $\widetilde{x}_1, \ldots, \widetilde{x}_n$ come from measurement or from expert estimation; in both cases, these values are only approximately equal to the actual (unknown) values x_i.

Case of interval uncertainty. Traditional methods for taking the measurement uncertainty into account are based on the assumption that we know the probabilities of different values of the measurement error $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i$. Often, however, we do not know the probabilities, the only thing we know is the upper bound Δ_i on the measurement errors: $|\Delta x_i| \le \Delta_i$; see, e.g., [7]. In this case, based on the measurement result $\widetilde{x}_i$, the only information that we have about the actual (unknown) value x_i is that x_i belongs to the interval

$$\mathbf{x}_i = [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i].$$

For different values x_i within the corresponding intervals, in general, we get different values of the variance-to-mean ratio R. It is therefore desirable to find the range $\mathbf{R} = [\underline{R}, \overline{R}]$ of this ratio when $x_i \in \mathbf{x}_i$:

$$[\underline{R}, \overline{R}] = \{R(x_1, \ldots, x_n) : x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n\}.$$

Comment. The problem of computing this range is a particular case of a general problem of *interval computations*, where we need to compute the range

$$[\underline{y}, \overline{y}] = \{f(x_1, \ldots, x_n) : x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n\}$$

of a given function $f(x_1, \ldots, x_n)$ on given intervals $\mathbf{x}_1$, ..., $\mathbf{x}_n$; see, e.g., [3], [5].

Case of fuzzy uncertainty. For expert estimates, we rarely have the upper bounds on the estimation errors. Instead, we have "fuzzy" estimates of the approximation error Δx_i, e.g., saying that "usually, the approximation error is about 0.1, and it is rarely larger than 0.2". Fuzzy logic is a natural way to formalize such natural-language statements; see, e.g., [4], [6]. Thus, for each i, we have a membership function $\mu_i(x_i)$ which describe the degree to which different values x_i are possible.

Based on these membership functions, we must find the degree $\mu(R)$ to which different values of the ratio R are possible. A value R is possible if it is equal to $R(x_1, \ldots, x_n)$ for some possible values $x_1, \ldots, x_n$:

$$R \text{ is possible } \Leftrightarrow$$

$$\exists x_1 \ldots \exists x_n \, (x_1 \text{ is possible } \& \ldots \& \, x_n \text{ is possible } \&$$

$$R(x_1, \ldots, R_n) = R).$$

We know the degrees $\mu_i(x_i)$ to which different values x_i are possible. Thus, if we use $\min$ to describe &, and $\max$ to describe $\vee$ (and thus $\exists$), we arrive at *Zadeh's extension principle*, according to which

$$\mu(R) = \max_{x_i : R(x_1, \ldots, R_n) = R} \min(\mu_1(x_1), \ldots, \mu_n(x_n)).$$

From the computational viewpoint, the case of fuzzy uncertainty can be reduced to the case of interval uncertainty. An alternative way to describe a membership function $\mu_i(x_i)$ is to describe, for each possible values $\alpha \in [0, 1]$, the set of all values x_i for which the degree of possibility is at least α. This set $\{x_i : \mu_i(x_i) \geq \alpha\}$ is called an *alpha-cut* and is denoted by $X_i(\alpha)$.

It is known (see, e.g., [4], [6]), that the for alpha-cuts, Zadeh's extension principle takes the following form: for every α, we have

$$R(\alpha) = \{R(x_1, \ldots, x_n) : x_i \in X_i(\alpha)\}.$$

Thus, for every α, finding the alpha-cut of the resulting membership function $\mu(R)$ is equivalent to applying interval computations to the corresponding intervals $X_1(\alpha)$, ..., $X_n(\alpha)$.

Because of this reduction, in the following text, we will only consider the case of interval uncertainty. Thus, we are arrive at the following problem:

Problem. Given the intervals $[\underline{x}_1, \overline{x}_1], \ldots, [\underline{x}_n, \overline{x}_n]$ with $\underline{x}_i > 0$, find the range $[\underline{R}, \overline{R}]$ of possible values of the ratio $R = \dfrac{V}{E}$, where $E = \dfrac{1}{n} \cdot \sum_{i=1}^{n} x_i$ and $V = \dfrac{1}{n} \cdot \sum_{i=1}^{n} (x_i - E)^2$.

II. Efficient Algorithm for Computing $\underline{R}$

Main idea. To compute $\underline{R}$, we apply the following algorithm. First, we sort all $2n$ endpoints $\underline{x}_i$ of the original intervals into a sorted sequence

$$z_1 \leq z_2 \leq \ldots \leq z_{2n}.$$

Thus, we divide the real line into $2n + 1$ *zones* $(-\infty, z_1]$, $[z_1, z_2]$, ..., $[z_{n-1}, z_n]$, and $[z_n, \infty)$. If we denote $z_0 \stackrel{\text{def}}{=} -\infty$ and $z_{n+1} = \infty$, then we can describe all these zones as $[z_k, z_{k+1}]$, for $k = 0, 1, \ldots, n$.

For each of these zones $[z_k, z_{k+1}]$, for each i, we take the following value $x_i \in [\underline{x}_i, \overline{x}_i]$:

- if $\overline{x}_i \leq z_k$, we take $x_i = \overline{x}_i$;
- if $z_{k+1} \leq \underline{x}_i$, we take $x_i = \underline{x}_i$;
- in all other cases, we take $x_i = z$.

The value z is determined from the condition that for the selected sequence x_i, we have

$$E + \frac{V}{2E} = z,$$

i.e., equivalently, $E^2 + \dfrac{1}{2} \cdot V = z \cdot E$. Both E^2 and V are quadratic functions of x_i, so we get a quadratic equation to determine z. Of all the roots of this quadratic equation, we only consider the values $z \in [z_k, z_{k+1}]$.

Algorithm in detail. First, we sort $2n$ values $\underline{x}_i, \overline{x}_i$ in an increasing order,

$$z_1 \leq z_2 \leq \ldots \leq z_{2n},$$

and define $z_0 = -\infty$ and $z_{2n+1} = +\infty$. For each zone $[z_k, z_{k+1}]$, $k = 0, \ldots, 2n$, we then do the following:

- For every i, we take:
  - if $\overline{x}_i \leq z_k$, we take $x_i = \overline{x}_i$;
  - if $\underline{x}_i \geq z_{k+1}$, we take $x_i = \underline{x}_i$.

  We count the number n_k of all the indices i for which $\overline{x}_i \leq z_k$ or $\underline{x}_i \geq z_{k+1}$.
- If $n_k = n$, then we compute the ratio R based on the selected values x_i.
- If $n_k \neq n$, then, based on the above assignments, we calculate the values

$$e_k = \sum_{i : \overline{x}_i \leq z_k} \underline{x}_i + \sum_{j : \underline{x}_j \geq z_{k+1}} \overline{x}_j, \tag{1}$$

$$m_k = \sum_{i : \overline{x}_i \leq z_k} (\underline{x}_i)^2 + \sum_{j : \underline{x}_j \geq z_{k+1}} (\overline{x}_j)^2, \tag{2}$$

$$A_k = n_k \cdot (n_k - n); \quad B_k = -2n_k \cdot e_k \cdot \mu_k, \tag{3}$$

$$C_k = e_k^2 + n \cdot m_k, \tag{4}$$

and solve the quadratic equation

$$A_k \cdot \mu_k^2 + B_k \cdot \mu_k + C_k = 0. \tag{5}$$

For each solution μ_k which is within the zone $[z_k, z_{k+1}]$, we compute

$$E_k = \frac{e_k}{n} + \frac{n - n_k}{n} \cdot \mu_k, \tag{6}$$

$$M_k = \frac{m_k}{n} + \frac{n - n_k}{n} \cdot \mu_k^2, \quad (7)$$

and

$$R = \frac{M_k - E_k^2}{E_k}. \quad (8)$$

The smallest of all the computed values R is the desired lower endpoint $\underline{R}$.

Mathematical comment. For reader's convenience, the justification of this algorithm is given in a special Justifications section.

Computational comment. If we take the above algorithm literally, then for each of the $2n+1 = O(n)$ zones, we need to compute the sums e_k and m_k, each of which takes linear time $O(n)$ to compute – which would take $O(n) \times O(n) = O(n^2)$ time. Indeed, the initial values e_0 and m_0 take linear time. However, once we have computed the sums e_k and m_k, to find the next values e_{k+1} and m_{k+1}, we only need to take into account the values $\underline{x}_i$ and $\overline{x}_j$ which start satisfying the inequality $\overline{x}_i \le z_k$ or which stop satisfying the inequality $\underline{x}_j \ge z_{k+1}$. Each value i an j passes through this change only once, so totally, we need to update $O(n)$ terms in computing all the sums $e_1, m_1, \ldots$

Thus, after sorting, the total computation time is $O(n) + O(n) = O(n)$. Since sorting take times $O(n \cdot \log(n))$, the total computation time of this algorithm is

$$O(n \cdot \log(n)) + O(n) = O(n \cdot \log(n)).$$

III. Efficient Algorithm for Computing $\overline{R}$ When No More Than C Intervals Have a Common Point

Formulation of the case. We consider the case when, for some fixed integer C, at most C intervals $[\underline{x}_i, \overline{x}_i]$ can have a common interior point.

For example, for $C = 1$, this means that no two intervals can have a common point. For $C = 2$, this means that while it is possible than a pair of intervals has a common point, no three intervals have a common point, etc.

Algorithm. In this case, to compute $\overline{R}$, we use the following algorithm. First, we sort $2n$ values $\underline{x}_i, \overline{x}_i$ in an increasing order,

$$z_1 \le z_2 \le \ldots \le z_{2n},$$

and define $z_0 = -\infty$ and $z_{2n+1} = +\infty$. For each zone $[z_k, z_{k+1}]$, $k = 0, \ldots, 2n$, we then do the following:

- For every i for which $\overline{x}_i < z_k$, we take $x_i = \underline{x}_i$.
- For every i for which $z_{k+1} < \underline{x}_i$, we take $x_i = \overline{x}_i$.

For all other i, we take all possible combinations of $\underline{x}_i$ and $\overline{x}_i$. For each zone and for each such combination, we compute the ratio R.

The largest of the resulting ratios is returned as $\overline{R}$.

Computational complexity. Sorting requires time

$$O(n \cdot \log(n)).$$

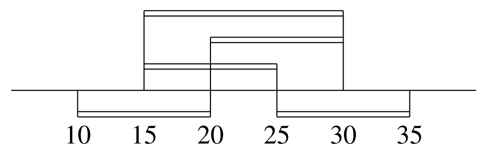
After sorting, for each zone, we have no more than C intervals with two possible values x_i (see Justifications section). So, for a fixed C, we have $2^C = O(1)$ possible combinations $x = (x_1, \ldots, x_n)$. For each combination, we need linear time to compute R – but, similarly to the case of $\underline{R}$, we can update the values computed for the previous zone, and this requires a total linear time.

Thus, similar to the case of $\underline{R}$, we have an algorithm that takes time $O(n)$ after sorting and the total time

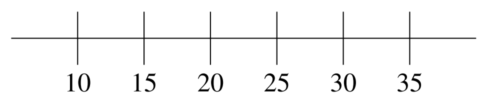
$$O(n \cdot \log(n)) + O(n) = O(n \cdot \log(n)).$$

IV. Computing $\overline{R}$: Numerical Example

Description of the example. To illustrate our algorithm for computing the upper endpoint $\overline{R}$, let us consider the following 5 intervals: $x_1 = [10, 20]$, $x_2 = [15, 25]$, $x_3 = [20, 30]$, $x_4 = [25, 35]$, and $x_5 = [15, 30]$. In this case, $C = 3$. These intervals are shown by double lines.



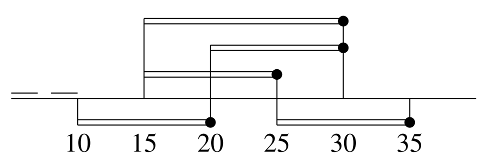
First step: sorting and computing zones. In this case, sorting the endpoints of the given intervals leads to $x_{(0)} = -\infty$, $x_{(1)} = 10$, $x_{(2)} = 15$, $x_{(3)} = 20$, $x_{(4)} = 25$, $x_{(5)} = 30$, $x_{(6)} = 35$, $x_{(7)} = +\infty$. Here, we have 7 zones:



Let us analyze these zones one by one.

Zone corresponding to $k = 0$. In the zone $(-\infty, 10]$ corresponding to $k = 0$, the algorithm leads to the following choice of the values x_i; the selected values are marked by a star in the table and by a black circle in the picture; the dashed line represents the zone:

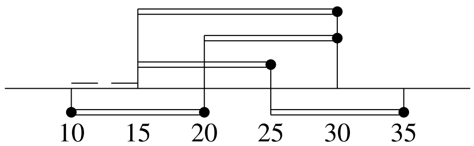
i	$\underline{x}_i$	$\overline{x}_i$
1	10	20*
2	15	25*
3	20	30*
4	25	35*
5	15	30*



For this zone, $E_0 = 28$, $M_0 = 810$, $V_0 = M_0 - E_0^2 = 26$, and $R_0 = V_0/E_0 = 0.928571$.

Zone corresponding to $k = 1$. In the zone $[10, 15]$ corresponding to $k = 1$, the algorithm leads to the following choice of the values x_i:

i	$\underline{x}_i$	$\overline{x}_i$
1	10*	20*
2	15	25*
3	20	30*
4	25	35*
5	15	30*



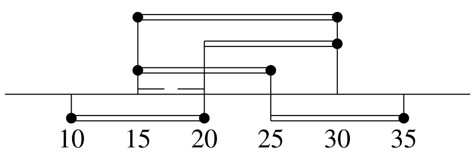
Here, for the element x_1, we have two different options: $x_1 = 10$ and $x_1 = 20$.

For the first choice $x_1 = 10$, we get $E_{1,1} = 26$, $M_{1,1} = 750$, $V_{1,1} = 74$, and $R_{1,1} = 2.84615$.

For the second choice $x_1 = 20$, we get $E_{1,2} = 28$, $M_{1,2} = 810$, $V_{1,2} = 76$, and $R_{1,2} = 0.928571$.

Zone corresponding to $k = 2$. In the zone $[15, 20]$ corresponding to $k = 2$, the algorithm leads to the following choice of the values x_i:

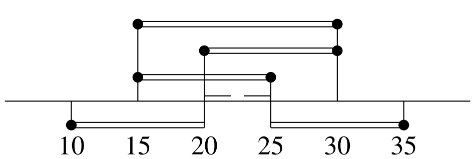
i	$\underline{x}_i$	$\overline{x}_i$
1	10*	20*
2	15*	25*
3	20	30*
4	25	35*
5	15*	30*



Here, we have 8 possible combinations of the values x_1, x_2, and x_5. For these combinations, we get $R_{2,1} = 4.47619$, $R_{2,2} = 3.91667$, $R_{2,3} = 3.73913$, $R_{2,4} = 2.84615$, $R_{2,5} = 2.86957$, $R_{2,6} = 2.07692$, $R_{2,7} = 2$, and $R_{2,8} = 0.928571$.

Zone corresponding to $k = 3$. In the zone $[20, 25]$ corresponding to $k = 3$, the algorithm leads to the following choice of the values x_i:

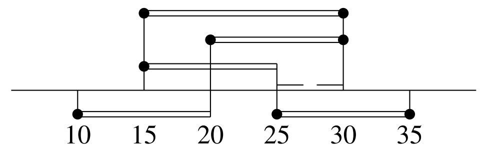
i	$\underline{x}_i$	$\overline{x}_i$
1	10*	20
2	15*	25*
3	20*	30*
4	25	35*
5	15*	30*



Here, we also have 8 different values R: $R_{3,1} = 3.89474$, $R_{3,2} = 3.90909$, $R_{3,3} = 4.47619$, $R_{3,4} = 3.91667$, $R_{3,5} = 3.52381$, $R_{3,6} = 3.08333$, $R_{3,7} = 3.73913$, and $R_{3,8} = 2.84615$.

Zone corresponding to $k = 4$. In the zone $[25, 30]$ corresponding to $k = 4$, the algorithm leads to the following choice of the values x_i:

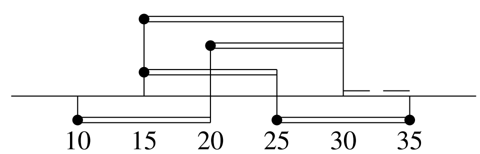
i	$\underline{x}_i$	$\overline{x}_i$
1	10*	20
2	15*	25
3	20*	30*
4	25*	35*
5	15*	30*



Here, we get the values $R_{4,1} = 1.52941$, $R_{4,2} = 2.5$, $R_{4,3} = 3.89474$, $R_{4,4} = 3.90909$, $R_{4,5} = 2.84211$, $R_{4,6} = 3$, $R_{4,7} = 4.47619$, and $R_{4,8} = 3.91667$.

Zone corresponding to $k = 5$. In the zone $[30, 35]$ corresponding to $k = 5$, the algorithm leads to the following choice of the values x_i:

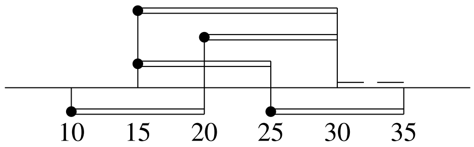
i	$\underline{x}_i$	$\overline{x}_i$
1	10*	20
2	15*	25
3	20*	30
4	25*	35*
5	15*	30



Here, we have two cases corresponding to $x_4 = 25$ and $x_4 = 35$: $R_{5,1} = 1.52941$ and $R_{5,2} = 3.89474$.

Zone corresponding to $k = 6$. Finally, in the zone $[35, \infty)$ corresponding to $k = 6$, the algorithm leads to the following choice of the values x_i:

i	$\underline{x}_i$	$\overline{x}_i$
1	10*	20
2	15*	25
3	20*	30
4	25*	35
5	15*	30



Here, $R_6 = 1.52941$.

Final result. As the desired value $\overline{R}$, we return the largest of the computed ratios, i.e., the value $\overline{R} = 4.47619$.

V. Justifications of the Algorithms

Justification of an algorithm for computing $\underline{R}$. From calculus, we know that a continuous function $R(x_1, \ldots, x_n)$ attains its minimum on a closed interval $[\underline{x}_i, \overline{x}_i]$ when:

- either the minimum is attained in the interior $(\underline{x}_i, \overline{x}_i)$ of the interval and $\dfrac{\partial R}{\partial x_i} = 0$,
- or the minimum is attained at the left endpoint $\underline{x}_i$ of the interval and $\dfrac{\partial R}{\partial x_i} \geq 0$,
- or the minimum is attained at the right endpoint $\overline{x}_i$ of the interval and $\dfrac{\partial R}{\partial x_i} \leq 0$.

Here,

$$\frac{\partial E}{\partial x_i} = \frac{\partial}{\partial x_i}\left(\frac{1}{n} \cdot \sum_{j=1}^{n} x_j\right) = \frac{1}{n},$$

and

$$\frac{\partial V}{\partial x_i} = \frac{\partial}{\partial x_i}\left(\frac{1}{n} \cdot \sum_{j=1}^{n} x_j^2 - E^2\right) = \frac{2}{n} \cdot x_i - 2E \cdot \frac{\partial E}{\partial x_i} = \frac{2 \cdot (x_i - E)}{n}.$$

Thus,

$$\frac{\partial R}{\partial x_i} = \frac{\partial}{\partial x_i}\left(\frac{V}{E}\right) = \frac{\dfrac{\partial V}{\partial x_i} \cdot E - V \cdot \dfrac{\partial E}{\partial x_i}}{E^2} = \frac{2x_i \cdot E - 2E^2 - V}{n \cdot E^2},$$

i.e.,

$$\frac{\partial R}{\partial x_i} = \frac{2}{n \cdot E} \cdot (x_i - z),$$

where $z \stackrel{\text{def}}{=} E + \dfrac{V}{2E}$.

Thus:

- the condition $\dfrac{\partial R}{\partial x_i} = 0$ is equivalent to $x_i = z$,
- the condition $\dfrac{\partial R}{\partial x_i} \geq 0$ is equivalent to $x_i \geq z$, and
- the condition $\dfrac{\partial R}{\partial x_i} \leq 0$ is equivalent to $x_i \leq z$.

So, the above conclusions can be reformulated as follows:

- either the minimum is attained in the interior $(\underline{x}_i, \overline{x}_i)$ of the interval and $x_i = z$,
- or the minimum is attained at the left endpoint $\underline{x}_i$ of the interval and $x_i = \underline{x}_i \geq z$,
- or the minimum is attained at the right endpoint $\overline{x}_i$ of the interval and $x_i = \overline{x}_i \leq z$.

Let us analyze what will be the consequences of these conditions in three possible situations:

- when the interval $[\underline{x}_i, \overline{x}_i]$ is to the left of z, i.e., when $\overline{x}_i \leq z$;
- when the interval $[\underline{x}_i, \overline{x}_i]$ is to the right of z, i.e., when $z \leq \underline{x}_i$; and
- when z is strictly inside the interval, i.e., $\underline{x}_i < z < \overline{x}_i$.

In the first situation, we have $x_i \leq z$, thus z cannot be the interior point of the interval. If the minimum is attained for $x_i = \underline{x}_i$, then, according to the above condition, we have $z \leq \underline{x}_i$ but we also have $\overline{x}_i = z$, thus, $\underline{x}_i \leq \overline{x}_i \leq z \leq \underline{x}_i$ hence $\underline{x}_i = \overline{x}_i$ and thus, the minimum is attained for $x_i = \overline{x}_i$. In the remaining case, the minimum is also attained for $x_i = \overline{x}_i$. Thus, in the first situation, the minimum is always attained when $x_i = \overline{x}_i$.

Similarly, in the second situation, when $z \leq \underline{x}_i$, the minimum is attained when $x_i = \underline{x}_i$.

Finally, in the third situation, when $\underline{x}_i < z < \overline{x}_i$, the minimum cannot be attained at $x_i = \underline{x}_i$, because then we would have $z \leq \underline{x}_i < z$ and thus $z < z$ – a contradiction. Similarly, the minimum cannot be attained at $x_i = \overline{x}_i$, because then we would have $z < \overline{x}_i \leq z$ and $z < z$. Thus, in this situation, the minimum has to be attained at an interior point, and we know that in this case, $x_i = z$.

Thus, once we know the location of the unknown value z with respect to the endpoints of all the intervals, we can uniquely determine, for every i, the value x_i at which the ratio R attains its minimum.

This is exactly what we do in the above algorithm: try all possible locations of z with respect to these endpoints; for each possible location, we assign the values x_i according to the above rule and see when we get the smallest possible value of the ratio R.

Justification of an algorithm for computing $\overline{R}$. Similarly to the previous algorithm justification, from calculus, we know that a continuous function $R(x_1, \ldots, x_n)$ attains its maximum on a closed interval $[\underline{x}_i, \overline{x}_i]$ when:

- either the maximum is attained in the interior $(\underline{x}_i, \overline{x}_i)$ of the interval, $\dfrac{\partial R}{\partial x_i} = 0$, and $\dfrac{\partial^2 R}{\partial x_i^2} \leq 0$,
- or the maximum is attained at the left endpoint $\underline{x}_i$ of the interval and $\dfrac{\partial R}{\partial x_i} \leq 0$,
- or the maximum is attained at the right endpoint $\overline{x}_i$ of the interval and $\dfrac{\partial R}{\partial x_i} \geq 0$.

We already have the formula for the first derivative of R. Differentiating the corresponding expression with respect to x_i, we conclude that

$$\frac{\partial^2 R}{\partial x_i^2} = \frac{\partial}{\partial x_i}\left(\frac{\partial R}{\partial x_i}\right) = \frac{2}{n \cdot E^2} \cdot (V - 2x_i \cdot E + (n+1) \cdot E^2).$$

The expression V can be represented as

$$V = \frac{1}{n} \cdot \sum_{j=1}^{n} x_j^2 - E^2 = \frac{1}{n} \cdot \sum_{j \neq i} x_j^2 + \frac{1}{n} \cdot x_i^2 - E^2.$$

Thus, we conclude that

$$V - 2x_i \cdot E + (n+1) \cdot E^2 =$$

$$\frac{1}{n} \cdot \sum_{j \neq i} x_j^2 + \frac{1}{n} \cdot x_i^2 - E^2 - 2x_i \cdot E + (n+1) \cdot E^2,$$

hence,

$$\frac{\partial^2 R}{\partial x_i^2} = \frac{2}{n^2 \cdot E^2} \cdot \left(\sum_{j \neq i} x_j^2 + x_i^2 - 2n \cdot x_i \cdot E + n^2 \cdot E^2 \right).$$

This expression can be represented as

$$\frac{\partial^2 R}{\partial x_i^2} = \frac{2}{n^2 \cdot E^2} \cdot \left(\sum_{j \neq i} x_j^2 + (n \cdot E - x_i)^2 \right).$$

This expression is a sum of squares of positive numbers and is, thus, always positive. Thus, the maximum cannot be attained at an interior point. Therefore, the maximum of the ratio R is always attained at one of the endpoints $x_i = \underline{x}_i$ or $x_i = \overline{x}_i$.

Taking into account the above conclusions and the known expression for $\dfrac{\partial R}{\partial x_i}$, the above conclusions can be reformulated as follows:

- either the maximum is attained at the left endpoint $\underline{x}_i$ of the interval and $x_i = \underline{x}_i \leq z$,
- or the maximum is attained at the right endpoint $\overline{x}_i$ of the interval and $x_i = \overline{x}_i \geq z$.

Let us analyze what will be the consequences of these conditions in three possible situations:

- when the interval $[\underline{x}_i, \overline{x}_i]$ is completely to the left of z, i.e., when $\overline{x}_i < z$;
- when the interval $[\underline{x}_i, \overline{x}_i]$ is completely to the right of z, i.e., when $z < \underline{x}_i$; and
- when z is inside the interval, i.e., $\underline{x}_i \leq z \leq \overline{x}_i$.

In the first situation $\overline{x}_i < z$, if the maximum is attained for $x_i = \overline{x}_i$, then, according to the above condition, we have $z \leq \overline{x}_i$, thus, $z \leq \overline{x}_i \leq z$ hence $z < z$ – a contradiction. Thus, in the first situation, the maximum is always attained when $x_i = \underline{x}_i$.

Similarly, in the second situation, when $z < \underline{x}_i$, the maximum is attained when $x_i = \overline{x}_i$.

In the third situation, we can have both $x_i = \underline{x}_i$ and $x_i = \overline{x}_i$.

This is exactly what we do in our algorithm: we consider all possible locations of z in relation to the endpoints. For each possible location, we assign unique values x_i to all intervals which are strictly to the left and strictly to the right of the corresponding zone, and try all possible combination of endpoints for intervals that contain this zone.

Since at most C intervals may have a common point, there are no more than C such intervals, so for each zone, we must consider no more than 2^C such assignments. When C is fixed, this is just a constant.

For each of these assignments, we compute R and take the largest of the values R as $\overline{R}$.

Acknowledgment

This project was done during Sio-Long Lo's visit to the University of Texas at El Paso. Sio-Long is thankful to the Macau University of Science and Technology (MUST) and to the University of Texas at El Paso (UTEP) for this research opportunity, and to Professors Liya Ding (MUST) and Vladik Kreinovich (UTEP) for their support.

The authors are thankful to the anonymous referees for valuable suggestions.

References

[1] D. R. Cox and P. A. W. Lewis, *The Statistical Analysis of Series of Events*, Methuen, London, 1966.

[2] P. Debroux, J. Boehm, F. Modave, V. Kreinovich, G. Xiang, J. Beck, K. Tupelly, R. Kandathi, L. Longpre, and K. Villaverde, "Using 1-D radar observations to detect a space explosion core among the explosion fragments: sequential and distributed algorithms", *Proceedings of the 11th IEEE Digital Signal Processing Workshop DSP'04*, Taos Ski Valley, New Mexico, August 1–4, 2004, pp. 273–277.

[3] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter, *Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control and Robotics*, Springer-Verlag, London, 2001.

[4] G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Upper Saddle River, New Jersey: Prentice Hall, 1995.

[5] R. E. Moore, R. B. Kearfott, and M. J. Cloud, *Introduction to Interval Analysis*, SIAM Press, Philadelphia, Pennsylviania, 2009.

[6] H. T. Nguyen and E. A. Walker, *First Course on Fuzzy Logic*, CRC Press, Boca Raton, Florida, 2006.

[7] S. Rabinovich, *Measurement Errors and Uncertainties: Theory and Practice*, Springer Verlag, New York, 2005.

[8] D. G. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman & Hall/CRC Press, Boca Raton, Florida, 2007.

[9] G. Upton and I. Cook, *Oxford Dictionary of Statistics*. Oxford University Press, Oxford, 2006.

[10] G. Xiang, *Fast Algorithms for Computing Statistics under Interval Uncertainty, with Applications to Computer Science and to Electrical and Computer Engineering*, PhD Dissertation, Department of Computer Science, University of Texas at El Paso, 2007.