

# Estimating Mean under Interval Uncertainty and Variance Constraint

Ali Jalal Kamali and Luc Longpré  
 Department of Computer Science  
 University of Texas at El Paso  
 500 W. University  
 El Paso, TX 79968, USA  
 ajalalkamali@miners.utep.edu  
 longpre@utep.edu

Misha Koshelev  
 Human Neuroimaging Lab  
 Division of Neuroscience  
 Baylor College of Medicine  
 1 Baylor Plaza  
 Houston, TX 77030, USA  
 misha680@gmail.com

**Abstract**—In many practical situations, we have a sample of objects of a given type. When we measure the values of a certain quantity for these objects, we get a sequence of value  $x_1, \dots, x_n$ . When the sample is large enough, then the arithmetic mean  $E$  of the values  $x_i$  is a good approximation for the average value of this quantity for all the objects from this class.

The values  $x_i$  come from measurements, and measurement is never absolutely accurate. Often, the only information that we have about the measurement error is the upper bound  $\Delta_i$  on this error. In this case, once we have the measurement result  $\tilde{x}_i$ , the condition that  $|\tilde{x}_i - x_i| \leq \Delta_i$  implies that the actual (unknown) value  $x_i$  belongs to the interval  $[\tilde{x}_i - \Delta_i, \tilde{x}_i + \Delta_i]$ .

In addition, we often know the upper bound  $V_0$  on the variance  $V$  of the actual values – e.g., we know that the objects belong to the same species, and we know that within-species differences cannot be too high. In such cases, to estimate the average over the class, we need to find the range of possible value of the mean under the constraints that each  $x_i$  belongs to the given interval  $[\underline{x}_i, \bar{x}_i]$  and that the variance  $V(x_1, \dots, x_n)$  is bounded by a given value  $V_0$ . In this paper, we provide efficient algorithms for computing this range.

## I. FORMULATION OF THE PROBLEM

**A standard way to analyze a sample.** In many practical situations, we have a sample of values  $x_1, \dots, x_n$  corresponding to objects of a certain type.

For example,  $x_i$  may represent the height of the  $i$ -th person in a group, or his or her weight, or the toxicity of the  $i$ -th snake of a certain species.

In this case, a standard way to describe the corresponding population is to estimate its mean

$$E = \frac{1}{n} \cdot \sum_{i=1}^n x_i \quad (1)$$

and variance

$$V = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - E)^2. \quad (2)$$

**Case of interval uncertainty.** The above formulas assume that we know the exact values of the characteristics  $x_1, \dots, x_n$ . In practice, these values usually come from measurements, and measurements are never absolutely exact (see, e.g., [7]):

the measurement results  $\tilde{x}_i$  are, in general, different from the actual (unknown) values  $x_i$ :  $\tilde{x}_i \neq x_i$ .

Traditionally, it is assumed that we know the probability distribution of the measurement errors  $\Delta x_i \stackrel{\text{def}}{=} \tilde{x}_i - x_i$ . However, often, the only information we have is the upper bound  $\Delta_i$  on the (absolute value of the) measurement error:  $|\Delta x_i| \leq \Delta_i$ .

In this case, based on the measurement result  $\tilde{x}_i$ , the only information that we have about the actual (unknown) value  $x_i$  is that  $x_i$  belongs to the interval  $\mathbf{x}_i = [\underline{x}_i, \bar{x}_i]$ , where  $\underline{x}_i = \tilde{x}_i - \Delta_i$  and  $\bar{x}_i = \tilde{x}_i + \Delta_i$ .

**Need to estimate mean and variance under interval uncertainty.** In general, different values  $x_i$  from the corresponding intervals  $\mathbf{x}_i$  lead to different values of the mean  $E$  and variance  $V$ . It is therefore desirable to describe the range of possible values of mean and variance when  $x_i$  belong to the corresponding intervals:

$$\mathbf{E} = [\underline{E}, \bar{E}] \stackrel{\text{def}}{=}$$

$$\{E(x_1, \dots, x_n) \mid x_1 \in \mathbf{x}_1, \dots, x_n \in \mathbf{x}_n\}; \quad (3)$$

$$\mathbf{V} = [\underline{V}, \bar{V}] \stackrel{\text{def}}{=}$$

$$\{V(x_1, \dots, x_n) \mid x_1 \in \mathbf{x}_1, \dots, x_n \in \mathbf{x}_n\}. \quad (4)$$

*Comment.* The problem of computing the corresponding ranges is a particular case of a general problem of computing the range

$$\mathbf{y} = [\underline{y}, \bar{y}] \stackrel{\text{def}}{=} \{f(x_1, \dots, x_n) \mid x_1 \in \mathbf{x}_1, \dots, x_n \in \mathbf{x}_n\} \quad (5)$$

of a given function  $f(x_1, \dots, x_n)$  when  $x_i$  are in known intervals. Computing such a range is called *interval computations*; see, e.g., [3], [5].

**Case of fuzzy uncertainty.** A similar problem occurs when, instead of measurement results  $\tilde{x}_i$ , we have expert estimates of the corresponding values  $x_i$ . Expert estimates are often formulated in terms of words from natural language, like “somewhat”, “close to 1.0”, etc.

A natural way to describe these estimates is by using fuzzy values  $X_i$  (i.e., membership functions  $\mu_i(x_i)$ ); see, e.g., [4], [6]. Instead of using membership functions, we can alternatively use  $\alpha$ -cuts

$$X_i(\alpha) \stackrel{\text{def}}{=} \{x_i : \mu(x_i) \geq \alpha\} \quad (6)$$

corresponding to different values  $\alpha$ .

In this case, it is desirable to find the fuzzy numbers  $\mu_f(y)$  corresponding to the mean  $f = E(x_1, \dots, x_n)$  and to the variance  $f = V(x_1, \dots, x_n)$ .

A natural way is to use Zadeh's extension principle

$$\mu_f(y) = \max\{\min(\mu_1(x_1), \dots, \mu_n(x_n)) : f(x_1, \dots, x_n) = y\}, \quad (7)$$

**Processing fuzzy uncertainty can be reduced to processing interval uncertainty.** It is known (see, e.g., [4], [6]) that for continuous functions  $f(x_1, \dots, x_n)$ , Zadeh's extension principle is equivalent to requiring that for every  $\alpha$ , the  $\alpha$ -cut  $Y(\alpha)$  for  $y = f(x_1, \dots, x_n)$  is equal to the range of values of  $f(x_1, \dots, x_n)$  when each  $x_i$  belongs to the corresponding  $\alpha$ -cut  $X_i(\alpha)$ :

$$Y(\alpha) = \{f(x_1, \dots, x_n) \mid x_1 \in X_1(\alpha), \dots, x_n \in X_n(\alpha)\}. \quad (8)$$

Thus, computing the corresponding fuzzy set is equivalent to solving several interval computation problems corresponding to different values  $\alpha$ : e.g., to  $\alpha = 0.1, 0.2, \dots, 1.0$ .

In view of this equivalence, in the following text, we will concentrate on the problem of computing mean and variance under interval uncertainty.

**Computing the range of the mean.** When we pick any of the variables  $x_i$  and increase it to some value  $x'_i > x_i$  (while leaving others intact, i.e.,  $x'_j = x_j$  for all  $j \neq i$ ), the value  $E$  would increase as well. Thus, the smallest value  $\underline{E}$  is attained when each of the variables  $x_i$  attains its smallest possible value  $x_i = \underline{x}_i$ , and its largest value  $\overline{E}$  is attained when each of the variables  $x_i$  attains its largest possible value  $x_i = \overline{x}_i$ :

$$\underline{E} = \frac{1}{n} \cdot \sum_{i=1}^n \underline{x}_i; \quad \overline{E} = \frac{1}{n} \cdot \sum_{i=1}^n \overline{x}_i. \quad (9)$$

**Computing the range of the variance.** The variance (2) is, in general, not monotonic; so, for the variance, the problem of computing the range  $[V, \overline{V}]$  under interval uncertainty is more complex.

Specifically, it turns out that while the lower endpoint  $\underline{V}$  can be computed in linear time [8], the problem of computing  $\overline{V}$  is, in general, NP-hard [1], [2].

**Variance constraints.** In the above expressions, we assume that there is no a priori information about the values of  $E$  and  $V$ .

In some cases, we have *a priori* constraints on the variance:  $V \leq V_0$  for a given  $V_0$ . For example, we know that within a species, there can be no more than 0.1 variation of a certain characteristic.

**Estimating mean under interval uncertainty and variance constraint: a problem.** In the presence of variance constraints, the problem of finding possible values of the mean  $E$  takes the following form:

- *given*:  $n$  intervals  $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$  and a number  $V_0 \geq 0$ ;
- *compute*: the range

$$[\underline{E}, \overline{E}] = \{E(x_1, \dots, x_n) \mid x_i \in \mathbf{x}_i \ \& \ V(x_1, \dots, x_n) \leq V_0\}. \quad (10)$$

This is a problem that we will solve in this paper.

**Case when this problem is (relatively) easy to solve.** Let us first consider the case when  $V_0$  is larger than (or equal to) the largest possible value  $\overline{V}$  of the variance corresponding to the given sample.

In this case, the constraint  $V \leq V_0$  is always satisfied. Thus, in this case, the desired range simply coincides with the range of all possible values of  $E$ , i.e., with the *arithmetic average* (9) of the corresponding intervals.

**Another case when this problem is (relatively) easy to solve.** Another such case is when  $V_0 = 0$ .

In this case, the constraint  $V \leq V_0$  means that the variance  $V$  should be equal to 0. In this case, all non-negative values  $(x_i - E)^2$  should also be equal to 0 – otherwise, the average  $V$  of these values  $(x_i - E)^2$  would be positive. So, we have  $x_i = E$  for all  $i$  and thus, all the actual (unknown) values should coincide:  $x_1 = \dots = x_n$ . In this case, we know that this common value  $x_i$  belongs to each of  $n$  intervals  $\mathbf{x}_i$ , so it belongs to their *intersection*.

$$\mathbf{x}_1 \cap \dots \cap \mathbf{x}_n. \quad (11)$$

A value  $E$  belongs to the interval  $[\underline{x}_i, \overline{x}_i]$  if it is larger than or equal to its lower endpoint  $\underline{x}_i$  and smaller than or equal to its upper endpoint  $\overline{x}_i$ . Thus, for a value  $E$  to belong to all  $n$  intervals, it has to be larger than or equal to all  $n$  lower endpoints  $\underline{x}_1, \dots, \underline{x}_n$ , and it has to be smaller than or equal to all  $n$  upper endpoints  $\overline{x}_1, \dots, \overline{x}_n$ .

A number  $E$  is larger than or equal to  $n$  given numbers  $\underline{x}_1, \dots, \underline{x}_n$  if and only if it is larger than or equal to the largest of these  $n$  numbers, i.e., if  $\max(\underline{x}_1, \dots, \underline{x}_n) \leq E$ . Similarly, a number  $E$  is smaller than or equal to  $n$  given numbers  $\overline{x}_1, \dots, \overline{x}_n$  if and only if it is smaller than or equal to the smallest of these  $n$  numbers, i.e., if  $E \leq \min(\overline{x}_1, \dots, \overline{x}_n)$ . So, the intersection consists of all the numbers which are located between these two bounds, i.e., the intersection coincides with the interval

$$[\underline{E}, \overline{E}] = [\max(\underline{x}_1, \dots, \underline{x}_n), \min(\overline{x}_1, \dots, \overline{x}_n)]. \quad (12)$$

**General case.** In the general case, when  $V_0$  is larger than 0 but smaller than the upper endpoint  $\bar{V}$ , we should get intervals intermediate between intersection and arithmetic average. In this paper, we show how to compute the corresponding interval for  $E$ .

## II. MAIN RESULT

**Algorithm.** The following feasible algorithm solves the problem of computing the range  $[E, \bar{E}]$  of the range of the mean under interval uncertainty and variance constraint:

- First, we compute the values

$$E^- \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^n \underline{x}_i \text{ and } V^- \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^n (\underline{x}_i - E^-)^2;$$

$$E^+ \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^n \bar{x}_i \text{ and } V^+ \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^n (\bar{x}_i - E^+)^2.$$

- If  $V^- \leq V_0$ , then we return  $\underline{E} = E^-$ .
- If  $V^+ \leq V_0$ , then we return  $\bar{E} = E^+$ .
- If at least one these inequalities does not hold, i.e., if  $V_0 < V^-$  or  $V_0 < V^+$ , then we sort the all  $2n$  endpoints  $\underline{x}_i$  and  $\bar{x}_i$  into a non-decreasing sequence

$$z_1 \leq z_2 \leq \dots \leq z_{2n}$$

and consider  $2n - 1$  zones  $[z_k, z_{k+1}]$ .

- For each zone  $[z_k, z_{k+1}]$ , we take:
  - for every  $i$  for which  $\bar{x}_i \leq z_k$ , we take  $x_i = \bar{x}_i$ ;
  - for every  $i$  for which  $z_{k+1} \leq \underline{x}_i$ , we take  $x_i = \underline{x}_i$ ;
  - for every other  $i$ , we take  $x_i = \alpha$ ; let us denote the number of such  $i$ 's by  $n_k$ .

The value  $\alpha$  is determined from the condition that for the selected vector  $x$ , we have  $V(x) = V_0$ , i.e., from solving the following quadratic equation:

$$\frac{1}{n} \cdot \left( \sum_{i: \bar{x}_i \leq z_k} (\bar{x}_i)^2 + \sum_{i: z_{k+1} \leq \underline{x}_i} \underline{x}_i^2 + n_k \cdot \alpha^2 \right) - \frac{1}{n^2} \cdot \left( \sum_{i: \bar{x}_i \leq z_k} \bar{x}_i + \sum_{i: z_{k+1} \leq \underline{x}_i} \underline{x}_i + n_k \cdot \alpha \right)^2 = V_0. \quad (13)$$

Then:

- if none of the two roots of the above quadratic equation belongs to the zone, this zone is dismissed;
- if one or more roots belong to the zone, then for each of these roots, based on this  $\alpha$ , we compute the value

$$E_k(\alpha) = \frac{1}{n} \cdot \left( \sum_{i: \bar{x}_i \leq z_k} \bar{x}_i + \sum_{i: z_{k+1} \leq \underline{x}_i} \underline{x}_i + n_k \cdot \alpha \right). \quad (14)$$

- After that:
  - if  $V_0 < V^-$ , we return the smallest of the values  $E_k(\alpha)$  as  $\underline{E}$ :  $\underline{E} = \min k, \alpha E_k(\alpha)$ ;

- if  $V_0 < V^+$ , we return the largest of the values  $E_k(\alpha)$  as  $\bar{E}$ :  $\bar{E} = \max k, \alpha E_k(\alpha)$ .

*Comment.* The correctness of this algorithm is proven in the special Proof section.

**Computation time of this algorithm.** Sorting  $2n$  numbers requires time  $O(n \cdot \log(n))$ .

Once the values are sorted, we can then go zone-by-zone, and perform the corresponding computations. A straightforward implementation of the above algorithm would require time  $O(n^2)$ : for each of  $2n$  zones, we need linear time to compute several sums of  $n$  numbers.

However, in reality, only the sum for the first zone requires linear time. Once we have the sums for each zone, computing the sum for the next zone requires changing a few terms – values  $x_j$  which changed status. Each value  $x_j$  changes once, so overall, to compute all these sums, we still need linear time.

Thus, after sorting, the algorithm requires only linear computations time  $O(n)$ . So, if the endpoints are already given to us as sorted, we only take linear time.

If we still need to sort, then we need time

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

**Toy example.** Let us illustrate the above example of a simple example in which we have two intervals  $\mathbf{x}_1 = [-1, 0]$  and  $\mathbf{x}_2 = [0, 1]$ , and the bound  $V_0$  is equal to 0.16.

In this case, according to the above algorithm, we compute the values

$$E^- = \frac{1}{2} \cdot (-1 + 0) = -0.5;$$

$$V^- = \frac{1}{2} \cdot (((-1) - (-0.5))^2 + (0 - (-0.5))^2) = 0.25;$$

$$E^+ = \frac{1}{2} \cdot (0 + 1) = 0.5;$$

$$V^+ = \frac{1}{2} \cdot ((0 - 0.5)^2 + (1 - 0.5)^2) = 0.25.$$

Here,  $V_0 < V^-$  and  $V_0 < V^+$ , so for computing both bounds  $\underline{E}$  and  $\bar{E}$ , we need to consider different zones.

By sorting the 4 endpoints  $-1, 0, 0, \text{ and } 1$ , we get  $z_1 = -1 \leq z_2 = 0 \leq z_3 = 0 \leq z_4 = 1$ . Thus, here, we have 3 zones  $[z_1, z_2] = [-1, 0]$ ,  $[z_2, z_3] = [0, 0]$ , and  $[z_3, z_4] = [0, 1]$ .

1) For the first zone  $[z_1, z_2] = [-1, 0]$ , according to the above algorithm, we select  $x_2 = 0$  and  $x_1 = \alpha$ . To determine the value  $\alpha$ , we form the quadratic equation (13):

$$\frac{1}{2} \cdot (0^2 + \alpha^2) - \frac{1}{4} \cdot (0 + \alpha)^2 = V_0 = 0.16.$$

This equation is equivalent to

$$\frac{1}{2} \cdot \alpha^2 - \frac{1}{4} \cdot \alpha^2 = \frac{1}{4} \cdot \alpha^2 = 0.16,$$

hence  $\alpha^2 = 0.64$  and  $\alpha = \pm 0.8$ . Of the two roots  $\alpha = -0.8$  and  $\alpha = 0.8$ , only the first root belongs to the zone  $[-1, 0]$ . For this root, we compute the value (14):

$$E_1 = \frac{1}{2} \cdot (0 + \alpha) = \frac{1}{2} \cdot (0 + (-0.8)) = -0.4.$$

2) For the second zone  $[z_2, z_3] = [0, 0]$ , according to the above algorithm, we select  $x_1 = x_2 = 0$ . In this case, there is no need to compute  $\alpha$ , so we directly compute

$$E_2 = \frac{1}{2} \cdot (0 + 0) = 0.$$

3) For the third zone  $[z_3, z_4] = [0, 1]$ , according to the above algorithm, we select  $x_1 = 0$  and  $x_2 = \alpha$ . To determine the value  $\alpha$ , we form the quadratic equation (13):

$$\frac{1}{2} \cdot (0^2 + \alpha^2) - \frac{1}{4} \cdot (0 + \alpha)^2 = V_0 = 0.16.$$

This equation is equivalent to

$$\frac{1}{2} \cdot \alpha^2 - \frac{1}{4} \cdot \alpha^2 = \frac{1}{4} \cdot \alpha^2 = 0.16,$$

hence  $\alpha^2 = 0.64$  and  $\alpha = \pm 0.8$ . Of the two roots  $\alpha = -0.8$  and  $\alpha = 0.8$ , only the second root belongs to the zone  $[0, 1]$ . For this root, we compute the value (14):

$$E_3 = \frac{1}{2} \cdot (0 + \alpha) = \frac{1}{2} \cdot (0 + 0.8) = 0.4.$$

Here, we have a value  $E_k$  for all three zones, so we return

$$\underline{E} = \min(E_1, E_2, E_3) = -0.4;$$

$$\overline{E} = \max(E_1, E_2, E_3) = 0.4.$$

### III. PROOF OF THE ALGORITHM'S CORRECTNESS

1°. Let us first show that it is sufficient to prove correctness for the case of the upper endpoint  $\overline{E}$ .

Indeed, one can easily see that if we replace the original values  $x_i$  with the new values  $x'_i = -x_i$ , then the means changes sign  $E' = -E$  while the variance remains the same  $V' = V$ .

When each  $x_i$  is known with interval uncertainty  $x_i \in \mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$ , the corresponding interval for  $x'_i = -x_i$  is equal to  $\mathbf{x}'_i = [-\overline{x}_i, -\underline{x}_i]$ . The resulting interval  $\mathbf{E}' = [\underline{E}', \overline{E}']$  for  $E'$  is similarly equal to  $[-\overline{E}, -\underline{E}]$ , so  $\overline{E}' = -\underline{E}$  and thus,  $\underline{E} = -\overline{E}'$ .

Thus, if we know how to compute the upper endpoint  $\overline{E}$  for an arbitrary set of intervals  $\mathbf{x}_1, \dots, \mathbf{x}_n$ , we can compute  $\underline{E}$  for a given set of intervals  $\mathbf{x}_1 = [\underline{x}_1, \overline{x}_1], \dots, \mathbf{x}_n = [\underline{x}_n, \overline{x}_n]$  as follows:

- we compute  $n$  auxiliary intervals  $\mathbf{x}'_i = [-\overline{x}_i, -\underline{x}_i]$ ,  $i = 1, \dots, n$ ;
- we use the known algorithm to find the upper endpoint  $\overline{E}'$  for the range of the mean when  $x'_i \in \mathbf{x}'_i$  and  $V(x') \leq V_0$ ;
- we take  $\underline{E} = -\overline{E}'$ .

2°. Let us prove that the largest possible values  $\overline{E}$  is attained for some values  $x_i \in [\underline{x}_i, \overline{x}_i]$  for which  $V(x) \leq V_0$ .

Indeed, the variance function  $V(x_1, \dots, x_n)$  is continuous; thus, the set of all the values  $x = (x_1, \dots, x_n)$  for which  $V(x_1, \dots, x_n) \leq V_0$  is closed.

The box  $\mathbf{x}_1 \times \dots \times \mathbf{x}_n$  is closed and bounded and thus, compact. The set  $S$  of all the values  $x \in \mathbf{x}_1 \times \dots \times \mathbf{x}_n$  for which  $V(x) \leq V_0$  is a closed subset of a compact set and therefore, compact itself. A continuous function attains its maximum on a compact set at some point. In particular, this means that the function  $E(x)$  attains its maximum  $\overline{E}$  at some point  $x$ , i.e., that there exist values  $x = (x_1, \dots, x_n)$  for which  $E(x_1, \dots, x_n) = \overline{E}$ .

In the following text, we will consider these optimizing values.

3°. Let us prove that for the optimizing vector  $x$ , for all  $i$  for which we have  $x_i < E$ , we have  $x_i = \overline{x}_i$ .

Indeed, since  $V = M - E^2$ , where  $M \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^n x_i^2$ , we conclude that

$$\frac{\partial V}{\partial x_i} = \frac{\partial M}{\partial x_i} - \frac{\partial E^2}{\partial x_i} = \frac{\partial M}{\partial x_i} - 2 \cdot E \cdot \frac{\partial E}{\partial x_i}.$$

Here,  $\frac{\partial E}{\partial x_i} = \frac{1}{n}$ ,  $\frac{\partial M}{\partial x_i} = \frac{2x_i}{n}$ , and therefore,

$$\frac{\partial V}{\partial x_i} = \frac{2 \cdot (x_i - E)}{n}. \quad (15)$$

If we change only one value  $x_i$ , by replacing it with  $x_i + \Delta x_i$ , with a small  $\Delta x_i$ , the value of  $V$  changes by

$$\Delta V = \frac{\partial V}{\partial x_i} \cdot \Delta x_i + o(\Delta x_i) = \frac{2}{n} \cdot (x_i - E) \cdot \Delta x_i + o(\Delta x_i). \quad (16)$$

When  $x_i < E$ , i.e., when  $x_i - E < 0$ , then for small  $\Delta x_i > 0$ , we have a negative  $\Delta V$ , i.e., the variance decreases, while the mean  $E$  increases by  $\frac{1}{n} \cdot \Delta x_i > 0$ . Thus, if we had  $x_i < E$  and  $x_i \neq \overline{x}_i$  for some  $i$ , then we could, by slightly increasing  $x_i$ , further increase  $E$  while decreasing  $V$  (and thus, keeping the constraint  $V \leq V_0$ ). So, in this case, the vector  $x$  cannot be the one that maximizes  $E$  under the constraint  $V \leq V_0$ .

This conclusion proves that for the optimizing vector, when  $x_i < E$ , we have  $x_i = \overline{x}_i$ .

4°. Let us assume that an optimizing vector has a component  $x_i$  which is strictly inside the corresponding interval  $[\underline{x}_i, \overline{x}_i]$ , i.e., for which  $\underline{x}_i < x_i < \overline{x}_i$ . Due to Part 3 of this proof, we cannot have  $x_i < E$ , so we must have  $x_i \geq E$ . Let us prove that in this case,

- for every  $j$  for which  $E \leq x_j < x_i$ , we have  $x_j = \overline{x}_j$ , and
- for every  $k$  for which  $x_k > x_i$ , we have  $x_k = \underline{x}_k$ .

4.1°. Let us first prove that if  $x_i \in (\underline{x}_i, \overline{x}_i)$ , and  $E \leq x_j < x_i$ , then  $x_j = \overline{x}_j$ .

We will prove this by contradiction. Indeed, let us assume that we have  $E \leq x_j < x_i$  and  $x_j < \bar{x}_j$ . In this case, we can, in principle, slightly increase  $x_j$ , to  $x_j + \Delta x_j$  and slightly decrease  $x_i$ , to  $x_i - \Delta x_i$ , and still stay within the corresponding intervals  $\mathbf{x}_i$  and  $\mathbf{x}_j$ . We select  $\Delta x_j$  and  $\Delta x_i$  in such a way that the resulting change  $\Delta V$  in the variance  $V$  is non-negative. Here,

$$\Delta V = \frac{\partial V}{\partial x_j} \cdot \Delta x_j - \frac{\partial V}{\partial x_i} \cdot \Delta x_i + o(\Delta x_i) + o(\Delta x_j). \quad (17)$$

Substituting the formula (15) for the derivative  $\frac{\partial V}{\partial x_j}$  into this formula, we conclude that

$$\Delta V = \frac{2}{n} \cdot ((x_j - E)\Delta x_j - (x_i - E) \cdot \Delta x_i) + o(\Delta x_i) + o(\Delta x_j). \quad (18)$$

Thus, for every  $\Delta x_j$ , to get  $\Delta V = 0$ , we select

$$\Delta x_i = \frac{x_j - E}{x_i - E} \cdot \Delta x_j + o(\Delta x_j). \quad (19)$$

For this selection, the variance does not change, but the mean  $E$  is changed by

$$\Delta E = \frac{1}{n} \cdot (\Delta x_j - \Delta x_i) = \left(1 - \frac{x_j - E}{x_i - E}\right) \cdot \Delta x_j + o(\Delta x_j) = \frac{x_i - x_j}{x_i - E} \cdot \Delta x_j + o(\Delta x_j). \quad (20)$$

Since  $x_j < x_i$ , for small  $\Delta x_j$ , we have  $\Delta E > 0$ . Thus, we can further increase the mean without violating the constraint  $V \leq V_0$ . This contradicts to our assumption that  $x$  is the optimizing vector. Thus, when  $E < x_j < x_i$ , we cannot have  $x_j < \bar{x}_j$  – so we must have  $x_j = \bar{x}_j$ .

4.2°. Let us first prove that if  $x_i \in (\underline{x}_i, \bar{x}_i)$ ,  $E \leq x_i$ , and  $x_k > x_i$ , then  $x_k = \underline{x}_k$ .

Similarly, let us assume that we have  $x_k > x_i$  and  $x_k > \underline{x}_k$ . In this case, we can, in principle, slightly increase  $x_i$ , to  $x_i + \Delta x_i$  and slightly decrease  $x_k$ , to  $x_k - \Delta x_k$ , and still stay within the corresponding intervals  $\mathbf{x}_i$  and  $\mathbf{x}_k$ . We select  $\Delta x_i$  and  $\Delta x_k$  in such a way that the resulting change  $\Delta V$  in the variance  $V$  is non-negative. Here,

$$\Delta V = \frac{\partial V}{\partial x_i} \cdot \Delta x_i - \frac{\partial V}{\partial x_k} \cdot \Delta x_k + o(\Delta x_i) + o(\Delta x_k) = \frac{2}{n} \cdot ((x_i - E)\Delta x_i - (x_k - E) \cdot \Delta x_k) + o(\Delta x_i) + o(\Delta x_k). \quad (21)$$

Thus, for every  $\Delta x_i$ , to get  $\Delta V = 0$ , we select

$$\Delta x_k = \frac{x_i - E}{x_k - E} \cdot \Delta x_i + o(\Delta x_i). \quad (22)$$

For this selection, the variance does not change, but the mean  $E$  is changed by

$$\Delta E = \frac{1}{n} \cdot (\Delta x_i - \Delta x_k) = \left(1 - \frac{x_i - E}{x_k - E}\right) \cdot \Delta x_i + o(\Delta x_i) =$$

$$\frac{x_k - x_i}{x_k - E} \cdot \Delta x_i + o(\Delta x_i). \quad (23)$$

Since  $x_k > x_i$ , for small  $\Delta x_i$ , we have  $\Delta E > 0$ . Thus, we can further increase the mean without violating the constraint  $V \leq V_0$ . This contradicts to our assumption that  $x$  is the optimizing vector. Thus, when  $x_i < x_k$ , we cannot have  $x_k > \underline{x}_k$  – so we must have  $x_k = \underline{x}_k$ .

5°. Let us now consider the case when for all the components  $x_i \geq E$  of the optimizing vector  $x$ , we have either  $x_i = \underline{x}_i$  or  $x_i = \bar{x}_i$ . Let us show that in this case, all the values  $x_i$  for which  $x_i = \bar{x}_i$  are smaller than or equal to all the values  $x_j$  for which  $x_j = \underline{x}_j$ .

We will prove this statement by contradiction. Let us assume that there exists  $i$  and  $j$  for which  $E \leq x_j < x_i$ ,  $x_j = \underline{x}_j$  and  $x_i = \bar{x}_i$ . In this case, we can slightly increase the value  $x_j$ , to  $x_j + \Delta x_j$ , and slightly decrease the value  $x_i$ , to  $x_i - \Delta x_i$ , and still stay within the corresponding intervals. Similarly to Part 4 of this proof, for every  $\Delta x_j > 0$ , to get  $\Delta V = 0$ , we must select

$$\Delta x_i = \frac{x_j - E}{x_i - E} \cdot \Delta x_j + o(\Delta x_j). \quad (24)$$

For this selection, the variance does not change, but the mean  $E$  is changed by

$$\Delta E = \frac{1}{n} \cdot (\Delta x_j - \Delta x_i) = \left(1 - \frac{x_j - E}{x_i - E}\right) \cdot \Delta x_j + o(\Delta x_j) = \frac{x_i - x_j}{x_i - E} \cdot \Delta x_j + o(\Delta x_j). \quad (25)$$

Since  $x_j < x_i$ , for small  $\Delta x_j$ , we have  $\Delta E > 0$ . Thus, we can further increase the mean without violating the constraint  $V \leq V_0$ . This contradicts to our assumption that  $x$  is the optimizing vector. So, when  $E \leq x_j < x_i$ , we cannot have  $x_j = \underline{x}_j$  and  $x_i = \bar{x}_i$ .

This contradiction proves that all the values  $x_i$  for which  $x_i = \bar{x}_i$  are indeed smaller than or equal to all the values  $x_j$  for which  $x_j = \underline{x}_j$ .

6°. Due to Parts 3, 4, and 5 of this proof, there exists a threshold value  $\alpha$  such that

- for all  $j$  for which  $x_j < \alpha$ , we have  $x_j = \bar{x}_j$ , and
- for all  $k$  for which  $x_k > \alpha$ , we have  $x_k = \underline{x}_k$ .

Indeed, in the case described in Part 4, as such  $\alpha$ , we can take the value  $x_i$  that is strictly inside the corresponding interval  $\mathbf{x}_i$ . In the case described in Part 5, since all the upper endpoints from the optimizing vector are smaller than or equal to all the lower endpoints, we can take any value  $\alpha$  between the largest of the optimal values  $\bar{x}_j$  and smallest of the optimal values  $\underline{x}_j$ .

7°. Let us show that because of the property proven in Part 6, once we know to which zone  $\alpha$  belongs, we can uniquely determine all the components  $x_j$  of the corresponding vector  $x$  – a candidate for the optimal vector.

7.1°. Indeed, if  $\bar{x}_j < \alpha$ , then, since we have  $x_j < \bar{x}_j$ , we get  $x_j < \alpha$ . Thus, due to Part 6, we have  $x_j = \bar{x}_j$ .

7.2°. If  $\alpha < \underline{x}_j$ , then, since we have  $\underline{x}_j < x_j$ , we get  $\alpha < x_j$ . Thus, due to Part 6, we have  $x_j = \underline{x}_j$ .

7.3°. Let us now consider the remaining case when neither of the above two conditions is satisfied and thus, we have  $\underline{x}_j \leq \alpha \leq \bar{x}_j$ .

In this case, we cannot have  $x_j < \alpha$ , because then, due to Part 6, we would have  $x_j = \bar{x}_j$  and thus,  $\bar{x}_j < \alpha$ , which contradicts to the inequality  $\alpha \leq \bar{x}_j$ .

Similarly, we cannot have  $\alpha < x_j$ , because then, due to Part 6, we would have  $x_j = \underline{x}_j$  and thus,  $\alpha < \underline{x}_j$ , which contradicts to the inequality  $\underline{x}_j \leq \alpha$ .

Thus, the only possible value here is  $x_j = \alpha$ .

7.3°. Overall, we conclude that for each  $\alpha$ , we get exactly the arrangement formulated in our algorithm.

8°. Let us prove that when  $V_0 < V^+$ , then the maximum is attained when  $V = V_0$ .

Let us prove this by contradiction. Let us assume that  $V_0 < V^+$  and that the maximum of  $E$  is attained for some vector  $x = (x_1, \dots, x_n)$ , with  $x_i \in [\underline{x}_i, \bar{x}_i]$ , for which  $V(x) < V_0$ .

Since  $V < V_0 < V^+$ , we have  $V(x) < V^+ = V(\bar{x}_1, \dots, \bar{x}_n)$ . Thus,  $x = (x_1, \dots, x_n) \neq \bar{x} \stackrel{\text{def}}{=} (\bar{x}_1, \dots, \bar{x}_n)$  – otherwise, we would get  $V(x) = V(\bar{x}) = V^+$ . So, there exists an index  $i$  for which  $x_i \neq \bar{x}_i$ . Since  $x_i \in [\underline{x}_i, \bar{x}_i]$ , this means that  $x_i < \bar{x}_i$ . Thus, we can increase  $x_i$  by a small positive value  $\varepsilon > 0$ , to a new value  $x'_i = x_i + \varepsilon > x_i$ , and still remain inside the interval  $[\underline{x}_i, \bar{x}_i]$ .

The function  $V(x_1, \dots, x_n)$  describing covariance continually depends on  $x_i$ . Since  $V(x) < V_0$ , for sufficiently small  $\varepsilon$ , we will have  $V(x_1, \dots, x_{i-1}, x'_i, x_{i+1}, \dots, x_n) < V_0$ . Thus, the new vector still satisfies the constraint – but for this new vector, the mean is larger (by  $\varepsilon/n > 0$ ) than for the original vector  $x$ .

This contradicts to our assumption that the mean  $E(x)$  of the vector  $x$  is the largest possible under the given constraint  $V \leq V_0$ .

The above contradiction shows that when  $V_0 < V^+$ , then for the optimizing vector  $x$ , we have  $V(x) = V_0$ . This fact enables us to determine  $\alpha$  – as do in the algorithm – by solving the equation  $V(x(\alpha)) = V_0$ , where  $x(\alpha)$  is a vector corresponding to the given  $\alpha$ .

Correctness is proven.

## REFERENCES

- [1] S. Ferson, L. Ginzburg, V. Kreinovich, L. Longpré, and M. Aviles, “Computing Variance for Interval Data is NP-Hard”, *ACM SIGACT News*, 2002, Vol. 33, No. 2, pp. 108–118.
- [2] S. Ferson, L. Ginzburg, V. Kreinovich, L. Longpré, and M. Aviles, “Exact Bounds on Finite Populations of Interval Data”, *Reliable Computing*, 2005, Vol. 11, No. 3, pp. 207–233.
- [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, Pennsylvania, 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] G. Xiang, M. Ceberio, and V. Kreinovich, “Computing Population Variance and Entropy under Interval Uncertainty: Linear-Time Algorithms”, *Reliable Computing*, 2007, Vol. 13, No. 6, pp. 467–488.