12-1-2022

Epistemic vs. Aleatory: Case of Interval Uncertainty

Marina Tuyako Mizukoshi
Federal University of Goias, tuyako@ufg.br

Weldon Lodwick
University of Colorado Denver, weldon.lodwick@ucdenver.edu

Martine Ceberio
The University of Texas at El Paso, mceberio@utep.edu

Vladik Kreinovich
The University of Texas at El Paso, vladik@utep.edu

Comments:

Technical Report: UTEP-CS-22-127

Epistemic vs. Aleatory: Case of Interval Uncertainty

Marina Tuyako Mizukoshi, Weldon Lodwick, Martine Ceberio, Olga Kosheleva, and Vladik Kreinovich

Abstract Interval computations usually deal with the case of epistemic uncertainty, when the only information that we have about a value of a quantity is that this value is contained in a given interval. However, intervals can also represent aleatory uncertainty – when we know that each value from this interval is actually attained for some object at some moment of time. In this paper, we analyze how to take such aleatory uncertainty into account when processing data. We show that in case when different quantities are independent, we can use the same formulas for dealing with aleatory uncertainty as we use for epistemic one. We also provide formulas for processing aleatory intervals in situations when we have no information about the dependence between the inputs quantities.

1 Outline

There exist many algorithm for dealing with *epistemic* uncertainty, when we do not have full information about the actual value of a physical quantity x. An important case of such uncertainty is the case of *interval* uncertainty when all we know is the lower and upper bounds $\underline{x}$ and $\overline{x}$ on the actual (unknown) value x. This case is called

Marina Tuyako Mizukoshi
Federal University of Goias, Brazil, e-mail: tuyako@ufg.br

Weldon Lodwick
Department of Mathematical and Statistical Sciences, University of Colorado Denver,
1201 Larimer Street, Denver, Colorado 80204, USA, e-mail: weldon.lodwick@ucdenver.edu

Martine Ceberio and Vladik Kreinovich
Department of Computer Science, University of Texas at El Paso, 500 W. University
El Paso, Texas 79968, USA, e-mail: mceberio@utep.edu, vladik@utep.edu

Olga Kosheleva
Department of Computer Science, University of Texas at El Paso, 500 W. University
El Paso, Texas 79968, USA, e-mail: olgak@utep.edu

interval uncertainty, because in this case the set of possible values of x forms an interval $[\underline{x},\overline{x}]$.

In the case of epistemic uncertainty, we have a single (unknown) actual value, and the only thing we know about this value is that it belongs to the given interval. However, intervals emerge also in other situations, e.g., when we actually have several different values of the physical quantity, and the interval represents the range of these actual values. This situation is known as *aleatory uncertainty*. In this paper, we analyze how to take aleatory uncertainty into account during data processing.

2 Epistemic Interval Uncertainty: A Brief Reminder

What is epistemic interval uncertainty: a reminder. In many practical situations, we do not know the exact value of a physical quantity x, the only thing we know about this value x is an interval $[\underline{x},\overline{x}]$ that contains this value; see, e.g., [13]. In this case, the interval represents our lack of knowledge and is, thus, a particular case of *epistemic* uncertainty.

How to process data under epistemic interval uncertainty: general case. In general, data processing means that we apply an algorithm $f(x_1,\ldots,x_n)$ to some values $x_1,\ldots,x_n$, and get the result $y=f(x_1,\ldots,x_n)$. This procedure allows us, given the values x_i, to find the value of a quantity y whose values are related to x_i by the algorithm $y=f(x_1,\ldots,x_n)$. For example, if we know the resistance x_1 and the current x_2, then we can use the known dependence $y=x_1\cdot x_2$ (Ohm's law) to compute the voltage y.

In case of epistemic interval uncertainty, we only know the intervals $[\underline{x}_i,\overline{x}_i]$ that contain the actual (unknown) values x_i. Thus, the only thing that we can conclude about the value $y=f(x_1,\ldots,x_n)$ is that this value is contained in the set

$$\{f(x_1,\ldots,x_n) : x_1\in[\underline{x}_1,\overline{x}_1],\ldots,x_n\in[\underline{x}_n,\overline{x}_n]\}. \quad (1)$$

This set is usually called the *y-range*.

Definition 1. *By a* y-range *of a function* $y=f(x_1,\ldots,x_n)$ *on the intervals* $[\underline{x}_1,\overline{x}_1]$, $\ldots$, $[\underline{x}_n,\overline{x}_n]$*, we mean the set*

$$f([\underline{x}_1,\overline{x}_1],\ldots,x_n\in[\underline{x}_n,\overline{x}_n]) \stackrel{\text{def}}{=} \{f(x_1,\ldots,x_n) : x_1\in[\underline{x}_1,\overline{x}_1],\ldots,x_n\in[\underline{x}_n,\overline{x}_n]\}. \quad (2)$$

Data processing functions $y=f(x_1,\ldots,x_n)$ are usually continuous. For such functions, the following fact is known from calculus:

Proposition 1. *For each continuous function* $f(x_1,\ldots,x_n)$ *and for all intervals* $[\underline{x}_1,\overline{x}_1]$, $\ldots$, $[\underline{x}_n,\overline{x}_n]$*, the y-range is an interval.*

In the following text, we will denote the endpoints of the y-range by $\underline{y}$ and $\overline{y}$. Then, the y-range is equal to the interval $[\underline{y},\overline{y}]$. Computing this interval y-range is knows as *interval computations* [4, 8, 9, 10].

In general, the interval computation problem is NP-hard already for quadratic functions $y = f(x_1,\ldots,x_n)$; see, e.g., [6]. This means that unless P = NP (which most computer scientists believe not to be true), no feasible algorithm can compute all the y-ranges. However, there are many efficient algorithms that help to solve many practical cases of this general problem; see, e.g., [4, 8, 9, 10].

Linearized case. In the important and frequent case when we know the values x_i with a good accuracy, there exist feasible algorithms – actually, straightforward formulas – for interval computations. This case is known as the *linearized case*, and the corresponding techniques is known as *linearization*; see, e.g., [5, 6, 11].

To explain these formulas, we need to take into account that each interval $[\underline{x}_i,\overline{x}_i]$ can be presented as $[\widetilde{x}_i-\Delta_i,\widetilde{x}_i+\Delta_i]$, where

$$\widetilde{x}_i \stackrel{\rm def}{=} \frac{\underline{x}_i+\overline{x}_i}{2} \quad (3)$$

is the interval's midpoint and

$$\Delta_i \stackrel{\rm def}{=} \frac{\overline{x}_i-\underline{x}_i}{2} \quad (4)$$

is the interval's half-width. Each value x_i from this interval can be represented as $\widetilde{x}_i+\Delta x_i$, where the difference

$$\Delta x_i \stackrel{\rm def}{=} x_i-\widetilde{x}_i \quad (5)$$

satisfies the condition $|\Delta x_i| \le \Delta_i$. In these terms, each value $y = f(x_1,\ldots,x_n)$ takes the form

$$y = f(\widetilde{x}_1+\Delta x_1,\ldots,\widetilde{x}_n+\Delta x_n).$$

Since the values Δx_i are small, we can expand this expression in Taylor series and keep only linear terms in this expansion. Thus, we get

$$y = f(\widetilde{x}_1+\Delta x_1,\ldots,\widetilde{x}_n+\Delta x_n) = \widetilde{y}+\sum_{i=1}^n c_i\cdot\Delta x_i,$$

where we denoted

$$\widetilde{y} \stackrel{\rm def}{=} f(\widetilde{x}_1,\ldots,\widetilde{x}_n) \quad (6)$$

and

$$c_i \stackrel{\rm def}{=} \frac{\partial f}{\partial x_i}_{|x_1=\widetilde{x}_1,\ldots,x_n=\widetilde{x}_n}. \quad (7)$$

In other words, the data processing function $y = f(x_1,\ldots,x_n)$ becomes linear:

$$y = \widetilde{y}+\sum_{i=1}^n c_i\cdot(x_i-\widetilde{x}_i), \quad (8)$$

For such linear functions, there exists a known explicit formula for their y-ranges:

Proposition 2. *The y-range of a linear function (8) on intervals* $[\widetilde{x}_1-\Delta_1,\widetilde{x}_1+\Delta_1]$, $\ldots$, $[\widetilde{x}_n-\Delta_n,\widetilde{x}_n+\Delta_n]$ *is equal to* $[\widetilde{y}-\Delta,\widetilde{y}+\Delta]$, *where* $\widetilde{y}=f(\widetilde{x}_1,\ldots,\widetilde{x}_n)$ *and*

$$\Delta=\sum_{i=1}^{n}|c_i|\cdot\Delta_i. \tag{9}$$

Comment. This proposition is known (see, e.g., [5, 6, 11]), but for completeness, we include its proof. For readers' convenience, all the proofs are placed in the special (last) Proofs section.

3 Aleatory Interval Uncertainty: Towards the Precise Formulation of the Problem

Why interval uncertainty in the aleatory case. Aleatory uncertainty means that in different situations, we have different actual values of the corresponding quantity. What can we say about the set of these actual values?

For each physical quantity, there are usually bounds; see, e.g., [3, 14]: for example, all speeds are limited by the speed of light, all distances are bounded by the size of the Universe, etc. So, for each physical quantity, its set of actual values is bounded.

For each physical quantity x, its values change with time $x(t)$, and in physics, these changes are usually continuous; see, e.g., [3, 14]. As we have mentioned, the range of a continuous function $x(t)$ on any time interval is an interval. So, we can conclude that the set of all actual values of a quantity is an interval.

Comment. In this section, we consider the situation when all the values from this interval actually occur for some objects at some moments of time. Of course, in quantum physics, changes may be discrete, and we may have a discrete set of values. The following text analyzes what happens if we take into account that not all the values from the interval are physically possible.

Notations. To distinguish an aleatory interval from the epistemic one, we will denote aleatory intervals with capital letters. For example, while the epistemic interval for a quantity x is denoted by $[\underline{x},\overline{x}]$, its aleatory interval will be denoted by $[\underline{X},\overline{X}]$.

Resulting problem: informal description. How can we take aleatory interval uncertainty into account in data processing, when:

- we have aleatory information about the quantities $x_1,\ldots,x_n$,
- we have a data processing algorithm $y=f(x_1,\ldots,x_n)$, and
- we want to find out what aleatory knowledge we have about the quantity y.

To be more precise: for each of n quantities $x_1,\ldots,x_n$, we know an aleatory interval $[\underline{X}_i,\overline{X}_i]$, i.e., we know that all the values x_i from this interval $[\underline{X}_i,\overline{X}_i]$ are actually occurring. We want to known which values y are actually occurring.

Strictly speaking, to find out which values y occur, we need to know which n-tuples $(x_1,\ldots,x_n)$ occur, i.e., we need to know the set S of all the actually occurring n-tuples $(x_1,\ldots,x_n)$. Once we know this set, we can conclude that the set of actual values of y is equal to

$$f(S) \stackrel{\text{def}}{=} \{f(x_1,\ldots,x_n) : (x_1,\ldots,x_n) \in S\}.$$

We do not know this set S, we only have three pieces of information about this set:

- first, since we only know that the values $x_i \in [\underline{x}_i,\overline{x}_i]$ actually occur, the set S is contained in the "box"
$$[\underline{X}_1,\overline{X}_1] \times \ldots \times [\underline{X}_n,\overline{X}_n]$$
of all the n-tuples for which $x_i \in [\underline{X}_i,\overline{X}_i]$ for all i;
- second, since each values x_i from each aleatory interval $[\underline{X}_i,\overline{X}_i]$ is actually occurring, for each i and for each value $x_i \in [\underline{X}_i,\overline{X}_i]$, there must exist an n-tuple in the set S for which the i-th component is equal exactly to this value;
- third, since we consider the case when all dependencies are continuous, we can conclude that the set S is connected – as the set of values of a continuous function over a time interval.

We will call sets S satisfying these three properties *conceivable sets*.

A value y is definitely attained if y is contained in $f(S)$ for all conceivable sets. Thus, we arrive at the following definition.

Definition 2. *Let $[\underline{X}_1,\overline{X}_1]$, ..., $[\underline{X}_n,\overline{X}_n]$ be intervals, and let $f(x_1,\ldots,x_n)$ be a continuous real-valued function of n real variables.*

- *We say that a set $S \subseteq [\underline{X}_1,\overline{X}_1] \times \ldots \times [\underline{X}_n,\overline{X}_n]$ is* conceivable *if this set is connected and for every i and for every value $x_i \in [\underline{X}_i,\overline{X}_i]$, there exists an n-tuple $(x_1,\ldots,x_i,\ldots,x_n) \in S$ whose i-th component is equal to this value x_i.*
- *By the* aleatory y-set*, we mean the set of all the real numbers y for which, for every conceivable set S, there exists an n-tuple $(x_1,\ldots,x_n) \in S$ for which*
$$f(x_1,\ldots,x_n) = y.$$

Notation. In the following text, we will denote the aleatory y-set by

$$f_a([\underline{X}_1,\overline{X}_1],\ldots,[\underline{X}_n,\overline{X}_n]).$$

Proposition 3. *The aleatory y-set is either an interval, or an empty set.*

Comment. An example when the aleatory y-set is empty will be given in the following section.

Notation. Because of this result, in the following text, we will also call the aleatory y-set the aleatory *y-interval*. We denote this interval by $[\underline{Y},\overline{Y}]$.

4 How to Compute the Aleatory y-Interval: Linearized Case

Let us denote the midpoint of the i-th aleatory interval by $\widetilde{X}_i$ and its half-width by $\delta_i > 0$. Let us consider the linearized case, when the data processing algorithm takes the form

$$y = \widetilde{Y} + \sum_{i=1}^{n} c_i \cdot (x_i - \widetilde{X}_i), \tag{10}$$

where $\widetilde{X}_i$ is the midpoint of the aleatory interval $[\underline{X}_i, \overline{X}_i]$,

$$c_i \stackrel{\text{def}}{=} \frac{\partial f}{\partial x_i}_{|x_1=\widetilde{X}_1,\ldots,x_n=\widetilde{X}_n},$$

and

$$\widetilde{Y} \stackrel{\text{def}}{=} f(\widetilde{X}_1, \ldots, \widetilde{X}_n). \tag{11}$$

In this case, we can efficiently compute the aleatory y-interval by using the following result.

Proposition 4. *Suppose that we have n intervals*

$$[\widetilde{X}_1 - \delta_1, \widetilde{X}_1 + \delta_1], \ldots, [\widetilde{X}_n - \delta_n, \widetilde{X}_n + \delta_n],$$

and the function $y = f(x_1, \ldots, x_n)$ has the form (10). *Let us denote*

$$\delta \stackrel{\text{def}}{=} 2 \cdot \max_{i=1,\ldots,n} |c_i| \cdot \delta_i - \sum_{i=1}^{n} |c_i| \cdot \delta_i.$$

Then, the aleatory y-interval has the following form:

- *if $\delta < 0$, then $f_a([\widetilde{X}_1 - \delta_1, \widetilde{X}_1 + \delta_1], \ldots, [\widetilde{X}_n - \delta_n, \widetilde{X}_n + \delta_n]) = \emptyset$; and*
- *if $\delta \geq 0$, then $f_a([\widetilde{X}_1 - \delta_1, \widetilde{X}_1 + \delta_1], \ldots, [\widetilde{X}_n - \delta_n, \widetilde{X}_n + \delta_n]) = [\widetilde{Y} - \delta, \widetilde{Y} + \delta]$.*

Corollary 1. *For the case when we have two intervals $[\underline{X}_1, \overline{X}_1]$ and $[\underline{X}_2, \overline{X}_2]$ and $f(x_1, x_2) = x_1 + x_2$, we get the aleatory y-interval*

$$[\underline{Y}, \overline{Y}] = [\underline{X}_1, \overline{X}_1] +_a [\underline{X}_2, \overline{X}_2] = [\min(\underline{X}_1 + \overline{X}_2, \overline{X}_1 + \underline{X}_2), \max(\underline{X}_1 + \overline{X}_2, \overline{X}_1 + \underline{X}_2)].$$

Corollary 2. *For the case when we have two intervals $[\underline{X}_1, \overline{X}_1]$ and $[\underline{X}_2, \overline{X}_2]$ and $f(x_1, x_2) = x_1 - x_2$, we get the aleatory y-interval*

$$[\underline{Y}, \overline{Y}] = [\underline{X}_1, \overline{X}_1] -_a [\underline{X}_2, \overline{X}_2] = [\min(\underline{X}_1 - \underline{X}_2, \overline{X}_1 - \overline{X}_2), \max(\underline{X}_1 - \underline{X}_2, \overline{X}_1 - \overline{X}_2)].$$

Example. For intervals $[0,2]$ and $[1,4]$, we get

$$[0,2] +_a [1,4] = [\min(0+4, 2+1), \max(0+4, 2+1)] =$$

$$[\min(3,4), \max(3,4)] = [3,4].$$

This is clearly different from the usual interval addition

$$[0,2] + [1,4] = [0+1, 2+4] = [1,6]$$

corresponding to epistemic or independent aleatory case.

5 Computing the Aleatory y-Interval Is, in General, NP-Hard

Discussion. In the previous section, we showed that when the function $y = f(x_1, \ldots, x_n)$ is linear, we can effectively compute the resulting aleatory y-interval. It turns out that – similarly to the above-mentioned case of epistemic interval uncertainty, if we consider the next-in-complexity class of functions – quadratic functions – the problem becomes NP-hard.

Proposition 5. *For quadratic functions $f(x_1, \ldots, x_n)$, the problem of computing the aleatory y-interval is NP-hard.*

6 Case of Partial or Full Independence

Discussion. In the above development, we considered the case when all we know is the set of actual values for each quantity x_i, and we do not know whether there is any dependence between these quantities. Because we allow cases of possible dependence, we can have somewhat counter-intuitive conclusions. For example, for the function $f(x_1, x_2) = x_1 \cdot x_2$ and aleatory intervals $[\underline{X}_1, \overline{X}_1] = [0, a_1]$ and $[\underline{X}_2, \overline{X}_2] = [0, a_2]$ for some $a_i > 0$, the aleatory y-interval consists of a single value 0.

Indeed, in this case, each conceivable set S contains a 2-tuple $(0, x_2)$ for which $x_1 \cdot x_2 = 0$. Thus, all y-ranges contain 0 and hence, the aleatory y-interval – which is the intersection of these y-ranges – also contains 0.

On the other hand, the set S consisting of all the points $(0, x_2)$ and $(x_1, 0)$ is conceivable. For this set, $f(S) = \{0\}$. Thus, the intersection of all y-ranges cannot contain any non-zero values and is, thus, indeed equal to $[\underline{Y}, \overline{Y}] = [0,0]$.

In some cases, however, we know that some sets of quantities are independent, i.e., that the fact that one of them has a value x_i should not affect the set of actually occurring values of other quantities x_j from this set. In such cases, all possible combinations $(x_i, x_j, \ldots)$ of actually occurring values are also actually occurring. So, we get the following modification of Definition 2.

Definition 3. *Let $[\underline{X}_1, \overline{X}_1]$, ..., $[\underline{X}_n, \overline{X}_n]$ be intervals, let $f(x_1, \ldots, x_n)$ be a continuous real-valued function of n real variables, and let $\mathscr{F}$ be a class of subsets $F \subseteq \{1, \ldots, n\}$ that contains all 1-element subsets $\{1\}, \ldots, \{n\}$.*

- *We say that a set $S \subseteq [\underline{X}_1, \overline{X}_1] \times \ldots \times [\underline{X}_n, \overline{X}_n]$ is $\mathscr{F}$-*conceivable *if:*

  – *this set is connected, and*
  – *for each set $F = \{i_1, i_2, \ldots\} \in \mathscr{F}$ and for each combination $(x_{i_1}, x_{i_2}, \ldots)$ of values $x_{i_k} \in [\underline{X}_{i_k}, \overline{X}_{i_k}]$, there exists an n-tuple $(x_1, \ldots, x_i, \ldots, x_n) \in S$ with these values $x_{i_1}, x_{i_2}, \ldots$*

- *By the* $\mathscr{F}$-aleatory y-set*, we mean the set of all the real numbers y for which, for every conceivable set S, there exists an n-tuple $(x_1, \ldots, x_n) \in S$ for which*

$$f(x_1, \ldots, x_n) = y.$$

Notation. In the following text, we will denote the $\mathscr{F}$-aleatory y-set by

$$f_a^{\mathscr{F}}([\underline{X}_1, \overline{X}_1], \ldots, [\underline{X}_n, \overline{X}_n]).$$

Proposition 6. *The $\mathscr{F}$-aleatory y-set is either an interval, or an empty set.*

Notation. Because of this result, in the following text, we will also call the $\mathscr{F}$-aleatory y-set the $\mathscr{F}$-aleatory *y-interval*. We denote this interval by $[\underline{Y}, \overline{Y}]$.

Case of full independence. In the case of *full independence*, when the class $\mathscr{F}$ contains the set $\{1, \ldots, n\}$, all n-tuples are conceivable, and there is only one conceivable set – the set of all combinations of the actual values of x_i:

$$S = [\underline{X}_1, \overline{X}_1] \times \ldots \times [\underline{X}_n, \overline{X}_n].$$

Thus, the $\mathscr{F}$-aleatory y-interval is equal to the y-range of the function $f(x_1, \ldots, x_n)$ on these intervals:

$$f([\underline{X}_1, \overline{X}_1], \ldots, x_n \in [\underline{X}_n, \overline{X}_n]) \stackrel{\text{def}}{=} \{f(x_1, \ldots, x_n) : x_1 \in [\underline{X}_1, \overline{X}_1], \ldots, x_n \in [\underline{X}_n, \overline{X}_n]\}.$$

This is exactly the same formula as for the epistemic uncertainty. Thus, in this case of full independence, to compute the aleatory y-interval, we can use the same interval computations techniques as for epistemic intervals.

Linearized case. In the linearized case, there are explicit formulas for the $\mathscr{F}$-aleatory y-interval:

Proposition 7. *Suppose that we have n intervals*

$$[\widetilde{X}_1 - \delta_1, \widetilde{X}_1 + \delta_1], \ldots, [\widetilde{X}_n - \delta_n, \widetilde{X}_n + \delta_n],$$

and the function $y = f(x_1, \ldots, x_n)$ has the form (10). *Let us denote*

$$\delta \stackrel{\text{def}}{=} 2 \cdot \max_{F \in \mathscr{F}} \sum_{i \in F} |c_i| \cdot \delta_i - \sum_{j=1}^{n} |c_j| \cdot \delta_j.$$

Then, the $\mathscr{F}$-aleatory y-interval has the following form:

- *if* $\delta < 0$, *then* $f_a([\widetilde{X}_1 - \delta_1, \widetilde{X}_1 + \delta_1], \ldots, [\widetilde{X}_n - \delta_n, \widetilde{X}_n + \delta_n]) = \emptyset$; *and*
- *if* $\delta \geq 0$, *then* $f_a([\widetilde{X}_1 - \delta_1, \widetilde{X}_1 + \delta_1], \ldots, [\widetilde{X}_n - \delta_n, \widetilde{X}_n + \delta_n]) = [\widetilde{Y} - \delta, \widetilde{Y} + \delta]$.

7 What If We Take Discreteness Into Account

In the previous text, we assumed that all the changes are continuous and thus, that all the ranges are connected. As we have mentioned, according to quantum physics, there can be discrete transitions. In most cases, however, these transitions are small, so that the distance between the previous state and the new state does not exceed some small number $\varepsilon > 0$. In this case, for the set of actual combinations of values, instead of the original connectedness, we have a similar notion of ε-*connectedness*:

Definition 4. *Let* $\varepsilon > 0$ *be a real number. We say that a set* $S \subseteq \mathbb{R}^n$ *is* ε-connected *if every two points* $x, x' \in S$ *can be connected by a sequence* $x = x^{(1)}, x^{(2)}, \ldots, x^{(m-1)}, x^{(m)} = x'$ *for which* $d(x^{(i)}, x^{(i+1)}) \leq \varepsilon$ *for all* $i = 1, \ldots, m-1$.

It turns out that such sets can be approximated by connected sets:

Definition 5. *We say that a connected set C is* ε-close *to the set S if* $S \subseteq C$ *and every element of C is* ε*-close to some element of the set S.*

Comment. In particular, we say that an interval $[\underline{X}, \overline{X}]$ is an ε-aleatory interval if for every value x from this interval, there is an ε-close actual value.

Proposition 8. *For every* ε*-connected set S, there exists an* ε*-close connected set C.*

One can easily see that for continuous functions $f(x_1, \ldots, x_n)$, the image of an ε-connected set is ε'-connected, for an appropriate ε'.

Proposition 9. *For each box*

$$B = [\underline{X}_1, \overline{X}_1] \times \ldots \times [\underline{X}_n, \overline{X}_n]$$

and for every ε*-connected set* $S \subseteq B$*, the y-range* $f(S)$ *is* ε'*-connected, for*

$$\varepsilon' = c_f(\varepsilon) \stackrel{\text{def}}{=} \sup\{d(f(x), f(x')) : d(x, x') \leq \varepsilon\}. \tag{12}$$

Thus, we arrive at the following definition:

Definition 6. *Let* $\varepsilon > 0$ *be given, let* $[\underline{X}_1, \overline{X}_1]$*, ...,* $[\underline{X}_n, \overline{X}_n]$ *be intervals for which* $\overline{X}_i - \underline{X}_i \geq 2\varepsilon$*, and let* $f(x_1, \ldots, x_n)$ *be a continuous real-valued function of n real variables.*

- *We say that a set $S \subseteq [\underline{X}_1, \overline{X}_1] \times \ldots \times [\underline{X}_n, \overline{X}_n]$ is* ε-conceivable *if this set is connected and for every i and for every value $x_i \in [\underline{X}_i, \overline{X}_i]$, there exists an n-tuple $(x_1, \ldots, x_i, \ldots, x_n) \in S$ whose i-th component is ε-close to this value.*
- *By the* ε-aleatory y-set*, we mean the set of all the real numbers y for which, for every ε-conceivable set S, there exists an n-tuple $(x_1, \ldots, x_n) \in S$ for which*

$$f(x_1, \ldots, x_n) = y.$$

Notation. In the following text, we will denote the ε-aleatory y-set by

$$f_{a,\varepsilon}([\underline{X}_1, \overline{X}_1], \ldots, [\underline{X}_n, \overline{X}_n]).$$

Discussion. From the computational viewpoint, this case can be reduced to the continuous case:

Proposition 10. *For every function $f(x_1, \ldots, x_n)$ and for all intervals $[\underline{X}_1, \overline{X}_1]$, ..., $[\underline{X}_n, \overline{X}_n]$, the ε-aleatory y-set has the form*

$$f_{a,\varepsilon}([\underline{X}_1, \overline{X}_1], \ldots, [\underline{X}_n, \overline{X}_n]) = f_a([\underline{X}_1 + \varepsilon, \overline{X}_1 - \varepsilon], \ldots, [\underline{X}_n + \varepsilon, \overline{X}_n - \varepsilon]).$$

What if we allow discontinuities of arbitrary size? In this case, there is no justification for connectedness or ε-connectedness, so we can have the following new definition, in which $*$ indicates that we no longer require connectedness:

Definition 7. *Let $[\underline{X}_1, \overline{X}_1]$, ..., $[\underline{X}_n, \overline{X}_n]$ be intervals, and let $f(x_1, \ldots, x_n)$ be a continuous real-valued function of n real variables, and let $\varepsilon > 0$.*

- *We say that a set $S \subseteq [\underline{X}_1, \overline{X}_1] \times \ldots \times [\underline{X}_n, \overline{X}_n]$ is* $*$-conceivable *if for every i and for every value $x_i \in [\underline{X}_i, \overline{X}_i]$, there exists an n-tuple $(x_1, \ldots, x_i, \ldots, x_n) \in S$ whose i-th component is ε-close to this value.*
- *By the* $*$-aleatory y-set*, we mean the set of all the real numbers y for which, for every $*$-conceivable set S, there exists an n-tuple $(x_1, \ldots, x_n) \in S$ for which*

$$f(x_1, \ldots, x_n) = y.$$

In some cases, we still have a non-empty $*$-aleatory y-set: e.g., in the above example of multiplication and $[\underline{X}_i, \overline{X}_i] = [0, a_i]$, we still have $[0,0]$ as the $*$-aleatory y-set. However, in the linearized case, this definition does not lead to any meaningful result: this set is always empty.

Proposition 11. *For each linear function $f(x_1, \ldots, x_n)$ of $n \neq 2$ variables that actually depends on all its variables – i.e., for which all the coefficients are different from 0 – the $*$-aleatory y-set is empty.*

8 Proofs

Proof of Proposition 2. The largest value of the expression (8) is attained when each term $c_i \cdot \Delta x_i$ in the sum is the largest.

- When c_i is positive, this largest value is attained when Δx_i is the largest, i.e., when $\Delta x_i = \Delta_i$.
- When c_i is negative, this largest value is attained when Δx_i is the smallest, i.e., when $\Delta x_i = -\Delta_i$.

In both cases, the largest value of this term is equal to $|c_i| \cdot \Delta_i$, so the largest value of y is equal to $\widetilde{y} + \Delta$, where we denoted

$$\Delta = \sum_{i=1}^{n} |c_i| \cdot \Delta_i.$$

Similarly, we can show that the smallest value of y is equal to $\widetilde{y} - \Delta$. Thus, the range of conceivable values of y is indeed equal to $[\widetilde{y} - \Delta, \widetilde{y} + \Delta]$. The proposition is proven.

Proof of Proposition 3. By Definition 2, a real number y belongs to the aleatory y-set if and only if for every conceivable set S, there exists an n-tuple $(x_1, \ldots, x_n) \in S$ for which $f(x_1, \ldots, x_n) = y$, i.e., for which y belongs to the y-range

$$f(S) = \{y = f(x_1, \ldots, x_n) : (x_1, \ldots, x_n) \in S\}.$$

This means that the aleatory y-set is the intersection of y-ranges $f(S)$ corresponding to all conceivable sets S.

Each conceivable set S is bounded – since it is contained in the bounded box $[\underline{X}_1, \overline{X}_1] \times \ldots \times [\underline{X}_n, \overline{X}_n]$ – and connected. The function $f(x_1, \ldots, x_n)$ is continuous, thus the corresponding range $f(S)$ is also bounded and connected, i.e., is an interval.

Intersection of intervals is either an interval or an empty set. The proposition is proven.

Proof of Proposition 4. In this proof, we use ideas from [7].

1°. One can see that to prove the proposition, we need to prove the following two statements:

- that when $\delta \geq 0$ and $|y - \widetilde{Y}| \leq \delta$, then, for each conceivable set S, we have $y = f(x_1, \ldots, x_n)$ for some n-tuple $(x_1, \ldots, x_n) \in S$, and
- that when $|y - \widetilde{Y}| > \delta$, there exists a conceivable set S for which, for all n-tuples $(x_1, \ldots, x_n) \in S$, we have $y \neq f(x_1, \ldots, x_n)$.

Let us prove these two statements one by one.

2°. Let us first prove that if $\delta \geq 0$ and $|y - \widetilde{Y}| \leq \delta$, then the aleatory y-interval contains the corresponding value y.

2.1°. Let is first prove that the aleatory y-interval contains both values $\widetilde{Y} - \delta$ and $\widetilde{Y} + \delta$, i.e., that for every conceivable set S, the range $f(S)$ contains both these values.

To prove this, let us denote by i_0 the index at which the product $|c_i|\cdot\delta_i$ attains its maximum:

$$|c_{i_0}|\cdot\delta_{i_0}=\max_i |c_i|\cdot\delta_i.$$

In these terms, the expression for δ has the following form

$$\delta=2\cdot|c_{i_0}|\cdot\delta_{i_0}-\sum_{i=1}^n |c_i|\cdot\delta_i.$$

The sum in the right-hand side can be represented as

$$\sum_{i=1}^n |c_i|\cdot\delta_i=|c_{i_0}|\cdot\delta_{i_0}+\sum_{i\ne i_0}|c_i|\cdot\delta_i,$$

thus

$$\delta=|c_{i_0}|\cdot\delta_{i_0}-\sum_{i\ne i_0}|c_i|\cdot\delta_i.$$

Let us also denote, by s_i, the sign of the coefficient c_i, i.e.:

- $s_i=1$ if $c_i\ge 0$, and
- $s_i=-1$ if $c_i<0$.

In this case, for all i, we have $c_i\cdot s_i=|c_i|$.

Since the set S is conceivable, there exists an n-tuple $(x_1,\ldots,x_n)$ for which $x_{i_0}=\widetilde{X}_{i_0}+s_{i_0}\cdot\delta_{i_0}$ and $|x_i-\widetilde{X}_i|\le\delta_i$ for all other i. For this n-tuple, we have

$$f(x_1,\ldots,x_n)=\widetilde{Y}+c_{i_0}\cdot(x_{i_0}-\widetilde{X}_{i_0})+\sum_{i\ne i_0}c_i\cdot(x_i-\widetilde{X}_i). \quad (13)$$

Here,

$$c_{i_0}\cdot(x_{i_0}-\widetilde{X}_{i_0})=c_{i_0}\cdot s_{i_0}\cdot\delta_{i_0}=|c_{i_0}|\cdot\delta_{i_0}, \quad (14)$$

and for each $i\ne i_0$, we have

$$c_i\cdot(x_i-\widetilde{X}_i)\ge-|c_i\cdot(x_i-\widetilde{X}_i)|=-|c_i|\cdot|x_i-\widetilde{X}_i|\ge-|c_i|\cdot\delta_i. \quad (15)$$

Due to (14) and (15), we have

$$f(x_1,\ldots,x_n)\ge\widetilde{Y}+|c_{i_0}|\cdot\delta_{i_0}-\sum_{i\ne i_0}|c_i|\cdot\delta_i,$$

i.e., $f(x_1,\ldots,x_n)\ge\widetilde{Y}+\delta$. Thus, for each conceivable set S, the y-range $f(S)$ contains a value which is larger than or equal to $\widetilde{Y}+\delta$. Let us denote this value by y_+.

Since the set S is conceivable, there exists an n-tuple $(x_1,\ldots,x_n)$ for which $x_{i_0}=\widetilde{X}_{i_0}-s_{i_0}\cdot\delta_{i_0}$ and $|x_i-\widetilde{X}_i|\le\delta_i$ for all other i. For this n-tuple, we have the formula (13). Here,

$$c_{i_0}\cdot(x_{i_0}-\widetilde{X}_{i_0})=c_{i_0}\cdot(-s_{i_0}\cdot\delta_{i_0})=-|c_{i_0}|\cdot\delta_{i_0}, \quad (16)$$

and for each $i \neq i_0$, we have

$$c_i \cdot (x_i - \widetilde{X}_i) \le |c_i \cdot (x_i - \widetilde{X}_i)| = |c_i| \cdot |x_i - \widetilde{X}_i| \le |c_i| \cdot \delta_i. \qquad (17)$$

Due to (16) and (17), we have

$$f(x_1,\ldots,x_n) \le \widetilde{Y} - |c_{i_0}| \cdot \delta_{i_0} + \sum_{i \neq i_0} |c_i| \cdot \delta_i,$$

i.e., $f(x_1,\ldots,x_n) \le \widetilde{Y} - \delta$. Thus, for each conceivable set S, the y-range $f(S)$ contains a value which is smaller than or equal to $\widetilde{Y} - \delta$. Let us denote this value by y_-.

The y-range $f(S)$ contains the values y_+ and y_- for which

$$y_- \le \widetilde{Y} - \delta \le \widetilde{Y} + \delta \le y_+.$$

Since the y-range $f(S)$ is an interval, it must also contain both intermediate values $\widetilde{Y} - \delta$ and $\widetilde{Y} + \delta$. So, this statement is proven.

2.2°. Let us now prove that, once we have a value y for which $|y - \widetilde{Y}| \le \delta$, the aleatory y-interval contains the value y.

Indeed, the inequality $|y - \widetilde{Y}| \le \delta$ means that $\widetilde{Y} - \delta \le y \le \widetilde{Y} + \delta$. We have proved that the aleatory y-interval contains both values $\widetilde{Y} - \delta$ and $\widetilde{Y} + \delta$. Thus, since the aleatory y-set is an interval, it should also contain the intermediate value y. The first part of the statement 1° is proven.

3°. Let us now prove that if $|y - \widetilde{Y}| > \delta$, then there exists a conceivable set S for which, for all n-tuples $(x_1,\ldots,x_n) \in S$, we have $f(x_1,\ldots,x_n) \neq y$.

Without loss of generality, we can consider the case when $y - \widetilde{Y} > \delta$. In this case, $y > \widetilde{Y} + \delta$. The case when $y - \widetilde{Y} < -\delta$ can be proven the same way.

Let us take, as S, the set consisting of all n-tuples of the type

$$(\widetilde{X}_1 - s_1 \cdot \delta_1, \ldots, \widetilde{X}_{i-1} - s_{i-1} \cdot \delta_{i-1}, x_i, \widetilde{X}_{i+1} - s_{i+1} \cdot \delta_{i+1}, \ldots, \widetilde{X}_n - s_n \cdot \delta_n), \qquad (18)$$

for all i and for all $x_i \in [\widetilde{X}_i - \delta_i, \widetilde{X}_i + \delta_i]$.

This set consists of n connected components corresponding to different values i. All these components have a common point $(\widetilde{X}_1 - s_1 \cdot \delta_1, \widetilde{X}_2 - s_2 \cdot \delta_2, \ldots, \widetilde{X}_n - s_n \cdot \delta_n)$ through which we can connect points from the two different component sets. Thus, the whole set S is connected.

It is easy to see that for every i, for every point $x_i \in [\widetilde{X}_i - \delta_i, \widetilde{X}_i + \delta_i]$, there exists an n-tuple from the set S with exactly this value of x_i – namely, we can take the corresponding point from the i-th component of the set S. Thus, the set S is conceivable.

Let us show that for all n-tuples $(x_1,\ldots,x_n)$ from this set S, we have $f(x_1,\ldots,x_n) \le \widetilde{Y} + \delta$ and thus, $f(x_1,\ldots,x_n) < y$ and $f(x_1,\ldots,x_n) \neq y$. Since the value y does not belong to the y-range $f(S)$ for one of the conceivable sets, this means that this value y does not belong to the y-aleatory interval $[\underline{Y}, \overline{Y}]$ which is the intersection of the y-ranges corresponding to all conceivable sets.

Indeed, for each n-tuple (18), we have

$$f(x_1,\ldots,x_n)=\widetilde{Y}+\sum_{j=1}^{n}c_j\cdot(x_j-\widetilde{X}_j)=\widetilde{Y}+c_i\cdot(x_i-\widetilde{X}_i)+\sum_{j\neq i}c_j\cdot(x_j-\widetilde{X}_j)=$$

$$\widetilde{y}+c_i\cdot(x_i-\widetilde{X}_i)+\sum_{j\neq i}c_j\cdot s_j\cdot(-\delta_j)=\widetilde{y}+c_i\cdot(x_i-\widetilde{X}_i)-\sum_{j\neq i}|c_j|\cdot\delta_j.$$

Here, $|x_i-\widetilde{X}_i|\le\delta_i$, so $c_i\cdot(x_i-\widetilde{X}_i)\le|c_i|\cdot\delta_i$ and thus,

$$f(x_1,\ldots,x_n)\le\widetilde{y}+|c_i|\cdot\delta_i-\sum_{j\neq i}|c_j|\cdot\delta_j=\widetilde{Y}+\left(2\cdot|c_i|\cdot\delta_i-\sum_{j=1}^{n}|c_j|\cdot\delta_j\right).$$

We have

$$|c_i|\cdot\delta_i\le\max_j|c_j|\cdot\delta_j,$$

therefore

$$f(x_1,\ldots,x_n)\le\widetilde{Y}+\left(2\cdot\max_j|c_j|\cdot\delta_j-\sum_{j=1}^{n}|c_j|\cdot\delta_j\right)=\widetilde{Y}+\delta.$$

The statement is proven, and so is the Proposition.

Proof of Corollary 1.

1°. For the sum function $f(x_1,x_2)=x_1+x_2$, we have $n=2$ and $c_1=c_2=1$. In this case, the expression for δ takes the form

$$\delta=2\cdot\max(\delta_1,\delta_2)-(\delta_1+\delta_2).$$

The expression $\max(\delta_1,\delta_2)$ is equal either to δ_1 or to δ_2 – depending on which of these values is larger. Let us consider both case $\delta_1\ge\delta_2$ and $\delta_1<\delta_2$.

1.1°. Let us first consider the case when $\delta_1\ge\delta_2$.

In this case, $\max(\delta_1,\delta_2)=\delta_1$, and we have $\delta=2\delta_1-(\delta_1+\delta_2)=\delta_1-\delta_2$. Here, $\widetilde{Y}=\widetilde{X}_1+\widetilde{X}_2$, so

$$\underline{Y}=\widetilde{Y}-\delta=\widetilde{X}_1+\widetilde{X}_2-\delta_1+\delta_2=(\widetilde{X}_1-\delta_1)+(\widetilde{X}_2+\delta_2)=\underline{X}_1+\overline{X}_2$$

and similarly,

$$\overline{Y}=\widetilde{Y}+\delta=\widetilde{X}_1+\widetilde{X}_2+\delta_1-\delta_2=(\widetilde{X}_1+\delta_1)+(\widetilde{X}_2-\delta_2)=\overline{X}_1+\underline{X}_2.$$

Since always $\underline{Y}\le\overline{Y}$, we thus have $\underline{X}_1+\overline{X}_2\le\overline{X}_1+\underline{X}_2$.

1.2°. Similarly, when $\delta_1\le\delta_2$, we get $\underline{Y}=\overline{X}_1+\underline{X}_2\le\overline{Y}=\underline{X}_1+\overline{X}_2$.

2°. In both cases,

$$\underline{Y}=\min(\underline{X}_1+\overline{X}_2,\overline{X}_1+\underline{X}_2)\text{ and }\overline{Y}=\max(\underline{X}_1+\overline{X}_2,\overline{X}_1+\underline{X}_2).$$

The corollary is proven.

Proof of Corollary 2. The difference $x_1 - x_2$ can be represented as $x_1 + (-x_2)$, where the set of known actual values of $-x_2$ is equal to

$$\{-x_2 : x_2 \in [\underline{X}_2, \overline{X}_2]\} = [-\overline{X}_2, -\underline{X}_2].$$

If we apply the formula from Corollary 1 to this expression, we get exactly the expression from Corollary 2. The statement is proven.

Proof of Proposition 5.

1°. By definition, a problem is NP-hard if every problem from the class NP can be reduced to this problem; see, e.g., [6, 12]. Thus, to prove that our problem is NP-hard, it is sufficient to show that a known NP-hard problem can be reduced to our problem. Then, for any problem from the class NP, by combining the reduction to the known problem with the reduction of the known problem, we will get the desired reduction to our problem – and thus, prove that our problem is NP-hard.

In this proof, as a known NP-hard problem, we take the following *partition problem*: given m positive integers $s_1, \ldots, s_m$, divide them into two groups with equal sum. If we move all the terms of the desired equality

$$\sum_{i \in G} s_i = \sum_{j \notin G} s_j$$

to the left-hand side, we get an equivalent equality

$$\sum_{i=1}^{n} \eta_i \cdot s_i = 0, \tag{19}$$

where $\eta_i = 1$ if $i \in G$ and $\eta_i = -1$ otherwise. Thus, the partition problem is equivalent to checking whether there exist values $\eta_i \in \{-1, 1\}$ for which the equality (19) is true.

Let us show that each instance of this problem can be reduced to the following instance of the problem of computing the aleatory y-interval: $n = m + 1$,

- $[\underline{X}_i, \overline{X}_i] = [-s_i, s_i]$ for all $i \le m$, and

  $$[\underline{X}_{m+1}, \overline{X}_{m+1}] = [0, s], \tag{20}$$

  where we denoted

  $$s \stackrel{\text{def}}{=} \frac{1}{m} \cdot \sum_{i=1}^{m} s_i^2, \tag{21}$$

  and
-
  $$f(x_1, \ldots, x_m, x_{m+1}) = x_{m+1} - V, \tag{22}$$

  where

  $$V \stackrel{\text{def}}{=} \frac{1}{m} \cdot \sum_{i=1}^{m} x_i^2 - \left(\frac{1}{m} \cdot \sum_{i=1}^{m} x_i \right)^2. \tag{23}$$

Comment. In the following text, we will use the fact that the expression V – which is actually the expression for sample variance – can be equivalently reformulated as

$$\frac{1}{m}\cdot\sum_{i=1}^{m}\left(x_i-\frac{1}{m}\cdot\sum_{j=1}^{m}x_j\right)^2,$$

and is, thus, always non-negative: $V\ge 0$.

2°. Let us recall – see, e.g., [1, 2] – that for the maximum

$$M\stackrel{\text{def}}{=}\max_{x_i\in[-s_i,s_i]}V, \tag{24}$$

we have the following property:

- if the original instance of the partition problem has a solution, then $M=s$, and
- if the original instance of the partition problem does not have a solution, then

$$M<s.$$

Indeed, since $|x_i|\le s_i$, we have $x_i^2\le s_i^2$ and thus,

$$V=\frac{1}{m}\cdot\sum_{i=1}^{m}x_i^2-\left(\frac{1}{m}\cdot\sum_{i=1}^{m}x_i\right)^2\le\frac{1}{m}\cdot\sum_{i=1}^{m}x_i^2\le\frac{1}{m}\cdot\sum_{i=1}^{m}s_i^2=s. \tag{25}$$

Hence, the maximum M of this expression is always smaller than or equal to s.

When the original instance of the partition problem has a solution η_i, then for $x_i=\eta_i\cdot s_i$, we have

$$\sum_{i=1}^{m}x_i=\sum_{i=1}^{m}\eta_i\cdot s_i=0 \tag{26}$$

and $x_i^2=s_i^2$, thus

$$V=\frac{1}{m}\cdot\sum_{i=1}^{m}x_i^2-\left(\frac{1}{m}\cdot\sum_{i=1}^{m}x_i\right)^2=\frac{1}{m}\cdot\sum_{i=1}^{m}x_i^2=\frac{1}{m}\cdot\sum_{i=1}^{m}s_i^2=s.$$

Thus, in this case, the maximum M is indeed equal to s.

Vice versa, if the maximum M is equal to s, this means that for some n-tuples $(x_1,\ldots,x_n)$, we have equality in both inequalities that form the formula (25). The fact that we have equality in the first inequality means that we have $x_i^2=s_i^2$ for all i, i.e., that $x_i=\pm s_i$, i.e., that we have $x_i=\eta_i\cdot s_i$ for some $\eta_i\in\{-1,1\}$. The fact that we have equality in the second inequality means that we have equality (26), i.e., that the values η_i form a solution to the original instance of the partition problem.

Thus, if the original instance of the partition problem does not have a solution, we cannot have $M=s$. Since we always have $M\le s$, this means that in this case, we must have $M<s$. The statement is proven.

Comment. As shown in [1, 2], we can feasibly compute the value $\delta > 0$ such that if the original instance of the partition problem does not have a solution, then we have

$$M \le s - \delta.$$

3°. Let us prove that for the above problem, the aleatory y-interval is equal to

$$[0, s - M].$$

3.1°. First, let us prove that for every conceivable set S, the y-range $f(S)$ contains the interval $[0, s - M]$.

3.1.1°. Let us first prove that the y-range $f(S)$ contains a value which is larger than or equal to $s - M$.

Indeed, by definition of a conceivable set, the set S contains an n-tuple for which $x_{m+1} = s$. The value $y_+ \stackrel{\text{def}}{=} f(x_1, \ldots, x_m, s)$ corresponding to this n-tuple is obtained by subtracting, from $x_{m+1} = s$, the expression V whose maximum is M. Thus, $V \le M$, and therefore, $y_+ = s - V \ge s - M$. So, the y-range $f(S)$ contains a value $y_+ \ge s - M$.

3.1.2°. Let us first prove that the y-range $f(S)$ contains a value which is smaller than or equal to 0.

Indeed, by definition of a conceivable set, the set S contains an n-tuple for which $x_{m+1} = 0$. The value $y_- \stackrel{\text{def}}{=} f(x_1, \ldots, x_m, 0)$ corresponding to this n-tuple is obtained by subtracting, from $x_{m+1} = 0$, a non-negative expression V. Thus, $y_- = 0 - V \le 0$.

3.1.3°. The interval $f(S)$ contains a value $y_- \le 0$ and a value $y_+ \ge s - M$, thus this interval must contain all the values between y_- and y_+, including all the values from the interval $[0, s - M]$. The Statement 3.1 is proven.

3.2°. Let us now prove that there exists a conceivable set S for which the y-range $f(S)$ does not contain any values larger than $s - M$.

Indeed, let $x_1^{\text{opt}}, \ldots, x_m^{\text{opt}}$ be an n-tuple at which the expression V attains its maximum M. Then, we can take the set U consisting of:

- all the $(m+1)$-tuples $(x_1^{\text{opt}}, \ldots, x_m^{\text{opt}}, x_{m+1})$ corresponding to all the values $x_{m+1} \in [0, s]$ and
- all the $(m+1)$-tuples $(x_1, \ldots, x_m, 0)$ corresponding to all m-tuples $(x_1, \ldots, x_m)$.

One can easily check that this set is conceivable.

- For n-tuples from the first component of this set, we have $y = x_{m+1} - M \le s - M$.
- For n-tuples from the second component of this set, we have $y = 0 - V \le 0$ and thus, $y \le s - M$.

Thus, for this set S, all the values from the y-range $f(S)$ are indeed smaller than or equal to $s-M$.

3.3°. Let us prove that there exists a conceivable set S for which the y-range $f(S)$ does not contain any negative values.

Indeed, let us take the set S consisting of:

- all the $(m+1)$-tuples $(0,\ldots,0,x_{m+1})$ corresponding to all the values $x_{m+1} \in [0,s]$ and
- all the $(m+1)$-tuples $(x_1,\ldots,x_m,s)$ corresponding to all m-tuples $(x_1,\ldots,x_m)$.

One can easily check that this set is conceivable.

- For $(m+1)$-tuples from the first component of this set, we have $V=0$, thus

$$y = x_{m+1} - V = x_{m+1} \geq 0.$$

- For $(m+1)$-tuples from the second component of this set, we have $y = s - V$ and, since $V \leq M$, we have $y \geq s - M \geq 0$.

Thus, for this set S, all the values from the y-range $f(S)$ are indeed non-negative.

3.4°. Due to Parts 3.1–3.3 of this proof, the aleatory y-interval – which is equal to the intersection of all the y-ranges $f(S)$ corresponding to conceivable sets S:

- contains the interval $[0,s-M]$ but
- does not contain any values outside this interval.

Thus, indeed, the aleatory y-interval is equal to $[0,s-M]$.

4°. Now we can prove the desired reduction.

If we could compute the aleatory y-range, we would then be able to compute the value $s-M$ and thus, we would be able to check whether $s-M>0$, i.e. whether $M<s$ or $M=s$. Due to Part 2 of this proof, we would then be able to check whether the original instance of the partition problem has a solution.

So, we indeed have a reduction of a known NP-hard problem to our problem. Thus, the problem of computing the aleatory y-interval is NP-hard. The proposition is proven.

Proof of Proposition 6 is similar to the proof of Proposition 3.

Proof of Proposition 7 is similar to the proof of Proposition 4.

Proof of Proposition 8. It is sufficient to take, as the desired set C, the union of all straight line segments that connect all the ε-close pairs of points $x,x' \in S$. One can easily see that this set is connected and ε-close to the original set S.

Proof of Proposition 10. An ε-conceivable set S must contain, for each i:

- n-tuples for which x_i is ε-close to $\underline{X}_i$, i.e., for which $x_i \leq \underline{X}_i + \varepsilon$, and
- n-tuples for which x_i is ε-close to $\overline{X}_i$, i.e., for which $x_i \geq \overline{X}_i - \varepsilon$.

Since the set S is connected, the set S_i of all values x_i corresponding to its n-tuples is also connected, i.e., is an interval containing points $x_i \le \underline{X}_i + \varepsilon$ and $x_i \ge \overline{X}_i - \varepsilon$. Thus, the set S_i contains all the values from the "ε-reduced" interval $[\underline{X}_i + \varepsilon, \overline{X}_i - \varepsilon]$, so it must be conceivable for the reduced intervals.

Vice versa, if we have a set S which is conceivable for the reduced intervals, then, as one can easily check, this set S is ε-conceivable for the original intervals $[\underline{X}_i, \overline{X}_i]$. Thus, the desired equality follows from Definition 2.

Proof of Proposition 11. Connected sets are a particular case of general sets, so the $*$-aleatory y-interval is a subset of the aleatory y-interval $[\underline{Y}, \overline{Y}]$. To prove the proposition, we need to prove that for every $y \in [\underline{Y}, \overline{Y}]$, there exists a $*$-conceivable set S for which $y \notin f(S)$. As such a set, let us take

$$S = \{(x_1, \ldots, x_n) : f(x_1, \ldots, x_n) \ne y\}.$$

The only thing we need to check is that for each i, all the values $x_i \in [\underline{X}_i, \overline{X}_i]$ are represented by n-tuples from this set, Indeed, otherwise, if some value $x_i^{(0)}$ was not represented, this would mean that $f(x_1, \ldots, x_{i-1}, x_i^{(0)}, x_{i+1}, \ldots, x_n) = y$ for all combinations $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ – but for a linear function, this would mean that this function does not depend on the variables $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$ at all – which contradicts to our assumption. The proposition is proven.

Comment. For a non-linear function, as the example of multiplication shows, the $*$-aleatory y-interval can be non-empty – exactly because for multiplication, there is a value $x_1^{(0)} = 0$ for which $f(x_1^{(0)}, x_2) = 0 \cdot x_2 = 0$ for all x_2.

Acknowledgments

This work was supported in part by the National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science), and HRD-1834620 and HRD-2034030 (CAHSI Includes), and by the AT&T Fellowship in Information Technology.

It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478, and by a grant from the Hungarian National Research, Development and Innovation Office (NRDI).

References

1. S. Ferson, L. Ginzburg, V. Kreinovich, and M. Aviles, "Exact bounds on sample variance of interval data", *Extended Abstracts of the 2002 SIAM Workshop on Validated Computing*, Toronto, Canada, May 23–25, 2002, pp. 67–69.

2. 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.
3. R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Addison Wesley, Boston, Massachusetts, 2005.
4. L. Jaulin, M. Kiefer, O. Didrit, and E. Walter, *Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control, and Robotics*, Springer, London, 2001.
5. V. Kreinovich, "Interval computations and interval-related statistical techniques: estimating uncertainty of the results of data processing and indirect measurements", In: F. Pavese, W. Bremser, A. Chunovkina, N. Fisher, and A. B. Forbes (eds.), *Advanced Mathematical and Computational Tools in Metrology and Testing AMTCM'X*, World Scientific, Singapore, 2015, pp. 38–49.
6. V. Kreinovich, A. Lakeyev, J. Rohn, and P. Kahl, *Computational complexity and feasibility of data processing and interval computations*, Kluwer, Dordrecht, 1998.
7. V. Kreinovich, V. M. Nesterov, and N. A. Zheludeva, "Interval methods that are guaranteed to underestimate (and the resulting new justification of Kaucher arithmetic)", *Reliable Computing*, 1996, Vol. 2, No. 2, pp. 119–124.
8. B. J. Kubica, *Interval Methods for Solving Nonlinear Contraint Satisfaction, Optimization, and Similar Problems: from Inequalities Systems to Game Solutions*, Springer, Cham, Switzerland, 2019.
9. G. Mayer, *Interval Analysis and Automatic Result Verification*, de Gruyter, Berlin, 2017.
10. R. E. Moore, R. B. Kearfott, and M. J. Cloud, *Introduction to Interval Analysis*, SIAM, Philadelphia, 2009.
11. H. T. Nguyen, V. Kreinovich, B. Wu, and G. Xiang, *Computing Statistics under Interval and Fuzzy Uncertainty*, Springer Verlag, Berlin, Heidelberg, 2012.
12. C. Papadimitriou, *Computational Complexity*, Addison-Wesley, Reading, Massachusetts, 1994.
13. S. G. Rabinovich, *Measurement Errors and Uncertainty: Theory and Practice*, Springer Verlag, New York, 2005.
14. K. S. Thorne and R. D. Blandford, *Modern Classical Physics: Optics, Fluids, Plasmas, Elasticity, Relativity, and Statistical Physics*, Princeton University Press, Princeton, New Jersey, 2017.