Frontiers In Global Optimization, pp. 1-2
C. A. Floudas and P. M. Pardalos, Editors
©2003 Kluwer Academic Publishers

Probabilities, Intervals, What Next? Optimization Problems Related to Extension of Interval Computations to Situations with Partial Information about Probabilities

V. Kreinovich
Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA
vladik@cs.utep.edu

Abstract

When we have only interval ranges $[\underline{x}_i, \overline{x}_i]$ of sample values $x_1, \ldots, x_n$, what is the interval $[\underline{V}, \overline{V}]$ of possible values for the variance V of these values? We show that the problem of computing the upper bound $\overline{V}$ is NP-hard. We provide a feasible (quadratic time) algorithm for computing the exact lower bound $\underline{V}$ on the variance of interval data. We also provide feasible algorithms that computes $\overline{V}$ under reasonable easily verifiable conditions, in particular, in case interval uncertainty is introduced to maintain privacy in a statistical database.

We also extend the main formulas of interval arithmetic for different arithmetic operations x_1 op x_2 to the case when, for each input x_i, in addition to the interval $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$ of possible values, we also know its mean E_i (or an interval $\mathbf{E}_i$ of possible values of the mean), and we want to find the corresponding bounds for $y = x_1$ op x_2 and its mean. In this case, we are interested not only in the bounds for y, but also in the bounds for the mean of y. We formulate and solve the corresponding optimization problems, and describe remaining open problems.

Keywords: interval computations, robust statistics, optimization.

1 Introduction: Data Processing–From Computing to Probabilities to Intervals

Why data processing? In many real-life situations, we are interested in the value of a physical quantity y that is difficult or impossible to measure directly. Examples of such

quantities are the distance to a star and the amount of oil in a given well. Since we cannot measure y directly, a natural idea is to measure y *indirectly*. Specifically, we find some easier-to-measure quantities $x_1, \ldots, x_n$ which are related to y by a known relation $y = f(x_1, \ldots, x_n)$; this relation may be a simple functional transformation, or complex algorithm (e.g., for the amount of oil, numerical solution to an inverse problem). Then, to estimate y, we first measure the values of the quantities $x_1, \ldots, x_n$, and then we use the results $\widetilde{x}_1, \ldots, \widetilde{x}_n$ of these measurements to to compute an estimate $\widetilde{y}$ for y as $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$.

For example, to find the resistance R, we measure current I and voltage V, and then use the known relation $R = V/I$ to estimate resistance as $\widetilde{R} = \widetilde{V}/\widetilde{I}$.

Computing an estimate for y based on the results of direct measurements is called *data processing*; data processing is the main reason why computers were invented in the first place, and data processing is still one of the main uses of computers as number crunching devices.

Comment. In this paper, for simplicity, we consider the case when the relation between x_i and y is known exactly; in some practical situations, we only known an approximate relation between x_i and y.

Why interval computations? From computing to probabilities to intervals. Measurement are never 100% accurate, so in reality, the actual value x_i of i-th measured quantity can differ from the measurement result $\widetilde{x}_i$. Because of these *measurement errors* $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i$, the result $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ of data processing is, in general, different from the actual value $y = f(x_1, \ldots, x_n)$ of the desired quantity y [16].

It is desirable to describe the error $\Delta y \stackrel{\text{def}}{=} \widetilde{y} - y$ of the result of data processing. To do that, we must have some information about the errors of direct measurements.

What do we know about the errors Δx_i of direct measurements? First, the manufacturer of the measuring instrument must supply us with an upper bound Δ_i on the measurement error. If no such upper bound is supplied, this means that no accuracy is guaranteed, and the corresponding "measuring instrument" is practically useless. In this case, once we performed a measurement and got a measurement result $\widetilde{x}_i$, we know that the actual (unknown) value x_i of the measured quantity belongs to the interval $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$, where $\underline{x}_i = \widetilde{x}_i - \Delta_i$ and $\overline{x}_i = \widetilde{x}_i + \Delta_i$.

In many practical situations, we not only know the interval $[-\Delta_i, \Delta_i]$ of possible values of the measurement error; we also know the probability of different values Δx_i within this interval. This knowledge underlies the traditional engineering approach to estimating the error of indirect measurement, in which we assume that we know the probability distributions for measurement errors Δx_i.

In practice, we can determine the desired probabilities of different values of Δx_i by comparing the results of measuring with this instrument with the results of measuring the same quantity by a standard (much more accurate) measuring instrument. Since the standard measuring instrument is much more accurate than the one use, the difference between these two measurement results is practically equal to the measurement error; thus, the empirical distribution of this difference is close to the desired probability distribution

for measurement error. There are two cases, however, when this determination is not done:

- First is the case of cutting-edge measurements, e.g., measurements in fundamental science. When a Hubble telescope detects the light from a distant galaxy, there is no "standard" (much more accurate) telescope floating nearby that we can use to calibrate the Hubble: the Hubble telescope is the best we have.
- The second case is the case of measurements on the shop floor. In this case, in principle, every sensor can be thoroughly calibrated, but sensor calibration is so costly – usually costing ten times more than the sensor itself – that manufacturers rarely do it.

In both cases, we have no information about the probabilities of Δx_i; the only information we have is the upper bound on the measurement error.

In this case, after we performed a measurement and got a measurement result $\widetilde{x}_i$, the only information that we have about the actual value x_i of the measured quantity is that it belongs to the interval $\mathbf{x}_i = [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$. In such situations, the only information that we have about the (unknown) actual value of $y = f(x_1, \ldots, x_n)$ is that y belongs to the range $\mathbf{y} = [\underline{y}, \overline{y}]$ of the function f over the box $\mathbf{x}_1 \times \ldots \times \mathbf{x}_n$:

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

The process of computing this interval range based on the input intervals $\mathbf{x}_i$ is called *interval computations*; see, e.g., [5, 6, 7, 12].

Interval computations as an optimization problem. The main problem of interval computations can be naturally reformulated as an optimization problem. Indeed, $\underline{y}$ is the solution to the following problem: $f(x_1, \ldots, x_n) \to \min$, under the conditions

$$\underline{x}_1 \le x_1 \le \overline{x}_1; \quad \ldots \quad \underline{x}_n \le x_n \le \overline{x}_n,$$

and $\overline{y}$ is the solution to the maximization problem $f(x_1, \ldots, x_n) \to \max$ under the same conditions.

Interval computations techniques: brief reminder. Historically the first method for computing the enclosure for the range is the method which is sometimes called "straightforward" interval computations. This method is based on the fact that inside the computer, every algorithm consists of elementary operations (arithmetic operations, min, max, etc.). For each elementary operation $f(a, b)$, if we know the intervals $\mathbf{a}$ and $\mathbf{b}$ for a and b, we can compute the exact range $f(\mathbf{a}, \mathbf{b})$. The corresponding formulas form the so-called *interval arithmetic*. For example,

$$[\underline{a}, \overline{a}] + [\underline{b}, \overline{b}] = [\underline{a} + \underline{b}, \overline{a} + \overline{b}]; \quad [\underline{a}, \overline{a}] - [\underline{b}, \overline{b}] = [\underline{a} - \overline{b}, \overline{a} - \underline{b}];$$

$$[\underline{a}, \overline{a}] \cdot [\underline{b}, \overline{b}] = [\min(\underline{a} \cdot \underline{b}, \underline{a} \cdot \overline{b}, \overline{a} \cdot \underline{b}, \overline{a} \cdot \overline{b}), \max(\underline{a} \cdot \underline{b}, \underline{a} \cdot \overline{b}, \overline{a} \cdot \underline{b}, \overline{a} \cdot \overline{b})].$$

In straightforward interval computations, we repeat the computations forming the program f step-by-step, replacing each operation with real numbers by the corresponding operation

of interval arithmetic. It is known that, as a result, we get an enclosure $\mathbf{Y} \supseteq \mathbf{y}$ for the desired range.

In some cases, this enclosure is exact. In more complex cases (see examples below), the enclosure has excess width.

There exist more sophisticated techniques for producing a narrower enclosure, e.g., a centered form method. However, for each of these techniques, there are cases when we get an excess width. Reason: as shown in [9, 18], the problem of computing the exact range is known to be NP-hard even for polynomial functions $f(x_1, \ldots, x_n)$ (actually, even for quadratic functions f).

What we are planning to do? First, we analyze a specific interval computations problem – when we use traditional statistical data processing algorithms $f(x_1, \ldots, x_n)$ to process the results of direct measurements.

Then, we extend our analysis to the case when for each input x_i, in addition to the interval $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$ of possible values, we have partial information about the probabilities: specifically, we know its mean E_i (or an interval $\mathbf{E}_i$ of possible values of the mean).

We formulate and solve the corresponding optimization problems, and describe remaining open problems.

2 First Step Beyond Intervals: Error Estimation for Traditional Statistical Data Processing Algorithms under Interval Uncertainty

When we have n results $x_1, \ldots, x_n$ of repeated measurement of the same quantity (at different points, or at different moments of time), traditional statistical approach usually starts with computing their sample average $E = (x_1 + \ldots + x_n)/n$ and their (sample) variance

$$V = \frac{(x_1 - E)^2 + \ldots + (x_n - E)^2}{n} \tag{1}$$

(or, equivalently, the sample standard deviation $\sigma = \sqrt{V}$); see, e.g., [16].

In this section, we consider situations when we do not know the exact values of the quantities $x_1, \ldots, x_n$, we only know the intervals $\mathbf{x}_1, \ldots, \mathbf{x}_n$ of possible values of x_i. In such situations, for different possible values $x_i \in \mathbf{x}_i$, we get different values of E and V. The question is: what are the intervals $\mathbf{E}$ and $\mathbf{V}$ of possible values of E and V?

The practical importance of this question was emphasized, e.g., in [13, 14] on the example of processing geophysical data.

For E, the straightforward interval computations leads to the exact range:

$$\mathbf{E} = \frac{\mathbf{x}_1 + \ldots + \mathbf{x}_n}{n}, \text{ i.e., } \underline{E} = \frac{\underline{x}_1 + \ldots + \underline{x}_n}{n}, \text{ and } \overline{E} = \frac{\overline{x}_1 + \ldots + \overline{x}_n}{n}.$$

For V, straightforward interval computations lead to an excess width. For example, for $\mathbf{x}_1 = \mathbf{x}_2 = [0, 1]$, the variance is $V = (x_1 - x_2)^2/4$ and hence, the actual range $\mathbf{V} = [0, 0.25]$. On the other hand, $\mathbf{E} = [0, 1]$, hence

$$\frac{(\mathbf{x}_1 - \mathbf{E})^2 + (\mathbf{x}_2 - \mathbf{E})^2}{2} = [0, 1] \supset [0, 0.25].$$

More sophisticated methods of interval computations also sometimes lead to an excess width.

Reason: in the formula for the average E, each variable only occurs once, and it is known that for such formulas, straightforward interval computations lead to the exact range (see, e.g., [4]). In the expression for variance, each variable x_i occurs several times: explicitly, in $(x_i - E)^2$, and explicitly, in the expression for E. In such cases, often, dependence between intermediate computation results leads to excess width of the results of straightforward interval computations. Not surprisingly, we do get excess width when applying straightforward interval computations to the formula (1).

For variance, we can actually prove that the corresponding optimization problem is difficult:

Theorem 1. *Computing $\overline{V}$ is NP-hard.*

Proof. By definition, a problem is NP-hard if any problem from the class NP can be reduced to it. Therefore, to prove that a problem $\mathcal{P}$ is NP-hard, it is sufficient to reduce one of the known NP-hard problems $\mathcal{P}_0$ to $\mathcal{P}$.

In this case, since $\mathcal{P}_0$ is known to be NP-hard, this means that every problem from the class NP can be reduced to $\mathcal{P}_0$, and since $\mathcal{P}_0$ can be reduced to $\mathcal{P}$, thus, the original problem from the class NP is reducible to $\mathcal{P}$.

For our proof, as the known NP-hard problem $\mathcal{P}_0$, we take a *subset* problem: given n positive integers $s_1, \ldots, s_n$, to check whether there exist signs $\eta_i \in \{-1, +1\}$ for which the signed sum $\sum\limits_{i=1}^{n} \eta_i \cdot s_i$ equals 0.

We will show that this problem can be reduced to the problem of computing $\overline{V}$, i.e., that to every instance $(s_1, \ldots, s_n)$ of the problem $\mathcal{P}_0$, we can put into correspondence such an instance of the $\overline{V}$-computing problem that based on its solution, we can easily check whether the desired signs exist.

As this instance, we take the instance corresponding to the intervals $[\underline{x}_i, \overline{x}_i] = [-s_i, s_i]$. We want to show that for the corresponding problem, $\overline{V} = C_0$, where we denoted $C_0 \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum\limits_{i=1}^{n} s_i^2$, if and only if there exist signs η_i for which $\sum \eta_i \cdot s_i = 0$.

1°. Let us first show that in all cases, $\overline{V} \le C_0$.

Indeed, it is known that the formula for the finite population variance can be reformulated in the following equivalent form:

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

Since $x_i \in [-s_i, s_i]$, we can conclude that $x_i^2 \le s_i^2$ hence $\sum x_i^2 \le \sum s_i^2$. Since $E^2 \ge 0$, we thus conclude that $V \le \frac{1}{n} \cdot \sum\limits_{i=1}^{n} s_i^2 = C_0$. In other words, every possible value V of the sample variance is smaller than or equal to C_0. Thus, the largest of these possible values, i.e., $\overline{V}$, also cannot exceed C_0, i.e., $\overline{V} \le C_0$.

2°. Let us now prove that if the desired signs η_i exist, then $\overline{V} = C_0$.

Indeed, in this case, for $x_i = \eta_i \cdot s_i$, we have $E = 0$ and $x_i^2 = s_i^2$, hence

$$V = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - E)^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} s_i^2 = C_0.$$

So, the variance V is always $\leq C_0$, and it attains the value C_0 for some x_i. Therefore, $\overline{V} = C_0$.

3°. To complete the proof of Theorem 1, we must show that, vice versa, if $\overline{V} = C_0$, then the desired signs exist.

Indeed, let $\overline{V} = C_0$. The variance is a continuous function on a compact set $\mathbf{x}_1 \times \ldots \times \mathbf{x}_n$, hence its maximum on this compact set is attained for some values $x_1 \in \mathbf{x}_1 = [-s_1, s_1], \ldots, x_n \in \mathbf{x}_n = [-s_n, s_n]$. In other words, for the corresponding values of x_i, the variance V is equal to C_0.

Since $x_i \in [-s_i, s_i]$, we can conclude that $x_i^2 \leq s_i^2$; since $E^2 \geq 0$, we get $V \leq C_0$. If $|x_i|^2 < s_i^2$ or $E^2 > 0$, then we would have $\sigma^2 < C_0$. Thus, the only way to have $V = C_0$ is to have $x_i^2 = s_i^2$ and $E = 0$. The first equality leads to $x_i = \pm s_i$, i.e., to $x_i = \eta_i \cdot s_i$ for some $\eta_i \in \{-1, +1\}$. Since E is, by definition, the (arithmetic) average of the values x_i, the equality $E = 0$ then leads to $\sum_{i=1}^{n} \eta_i \cdot s_i = 0$. So, if $\overline{V} = C_0$, then the desired signs do exist. The theorem is proven.

The very fact that computing the range of a quadratic function is NP-hard was first proven by Vavasis [18] (see also [9]). We have shown that this difficulty happens even for very simple quadratic functions frequently used in data processing.

A natural question is: maybe the difficulty comes from the requirement that the range be computed exactly? In practice, it is often sufficient to compute, in a reasonable amount of time, a usefully accurate estimate $\widetilde{\overline{V}}$ for $\overline{V}$, i.e., an estimate $\widetilde{\overline{V}}$ which is accurate with a given accuracy $\varepsilon > 0$: $\left|\widetilde{\overline{V}} - \overline{V}\right| \leq \varepsilon$. Alas, a simple modification of the above proof shows that for any ε, such computations are also NP-hard:

Theorem 2. *For every $\varepsilon > 0$, the problem of computing $\overline{V}$ with accuracy ε is NP-hard.*

It is worth mentioning that $\overline{V}$ can be computed exactly in exponential time $O(2^n)$:

Theorem 3. *There exists an algorithm that computes $\overline{V}$ in exponential time.*

Proof. Let $x_1^{(0)} \in \mathbf{x}_1, \ldots, x_n^{(0)} \in \mathbf{x}_n$ be the values for which the variance V attains maximum on the box $\mathbf{x}_1 \times \ldots \times \mathbf{x}_n$.

Let us pick one of the n variables x_i, and let us fix the values of all the other variables x_j $(j \neq i)$ at $x_j = x_j^{(0)}$. When we substitute $x_j = x_j^{(0)}$ for all $j \neq i$ into the expression for finite population variance, V becomes a quadratic function of x_i. This function of one variable should attain its maximum on the interval $\mathbf{x}_i$ at the value $x_i^{(0)}$.

By definition, the variance V is a sum of non-negative terms; thus, its value is always non-negative. Therefore, the corresponding quadratic function of one variable always has a global minimum. This function is decreasing before this global minimum, and increasing after it. Thus, its maximum on the interval $\mathbf{x}_i$ is attained at one of the endpoints of this interval.

In other words, for each variable x_i, the maximum is attained either for $x_i = \underline{x}_i$, or for $x_i = \overline{x}_i$. Thus, to find $\overline{V}$, it is sufficient to compute V for 2^n possible combinations $(x_1^{\pm}, \ldots, x_n^{\pm})$, where $x_i^- \stackrel{\text{def}}{=} \underline{x}_i$ and $x_i^+ \stackrel{\text{def}}{=} \overline{x}_i$, and find the largest of the resulting 2^n numbers. The theorem is proven.

For computing $\underline{V}$, there a feasible algorithm: specifically, our algorithm is *quadratic-time*, i.e., it requires $O(n^2)$ computational steps (arithmetic operations or comparisons) for n interval data points $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$.

The algorithm $\underline{\mathcal{A}}$ is as follows:

- First, we sort all $2n$ values $\underline{x}_i$, $\overline{x}_i$ into a sequence $x_{(1)} \le x_{(2)} \le \ldots \le x_{(2n)}$.
- Second, we compute $\underline{E}$ and $\overline{E}$ and select all "small intervals" $[x_{(k)}, x_{(k+1)}]$ that intersect with $[\underline{E}, \overline{E}]$.
- For each of the selected small intervals $[x_{(k)}, x_{(k+1)}]$, we compute the ratio $r_k = S_k/N_k$, where
$$S_k \stackrel{\text{def}}{=} \sum_{i:\underline{x}_i \ge x_{(k+1)}} \underline{x}_i + \sum_{j:\overline{x}_j \le x_{(k)}} \overline{x}_j,$$
and N_k is the total number of such i's and j's If $r_k \in [x_{(k)}, x_{(k+1)}]$, then we compute
$$V_k \stackrel{\text{def}}{=} \frac{1}{n} \cdot \left(\sum_{i:\underline{x}_i \ge x_{(k+1)}} (\underline{x}_i - r_k)^2 + \sum_{j:\overline{x}_j \le x_{(k)}} (\overline{x}_j - r_k)^2 \right).$$
If $N_k = 0$, we take $V_k \stackrel{\text{def}}{=} 0$.
- Finally, we return the smallest of the values V_k as $\underline{V}$.

Theorem 4. *The algorithm $\underline{\mathcal{A}}$ always compute $\underline{V}$ is quadratic time.*

Proof. Let us first show that this algorithm is indeed correct.

1°. Indeed, let $x_1^{(0)} \in \mathbf{x}_1, \ldots, x_n^{(0)} \in \mathbf{x}_n$ be the values for which the variance V attains minimum on the box $\mathbf{x}_1 \times \ldots \times \mathbf{x}_n$.

Let us pick one of the n variables x_i, and let us fix the values of all the other variables x_j $(j \ne i)$ at $x_j = x_j^{(0)}$. When we substitute $x_j = x_j^{(0)}$ for all $j \ne i$ into the expression for finite population variance, V becomes a quadratic function of x_i. This function of one variable should attain its minimum on the interval $\mathbf{x}_i$ at the value $x_i^{(0)}$.

As we have shown in the proof of Theorem 3, this function is decreasing before this global minimum, and increasing after it. This global minimum is attained when $\partial V/\partial x_i = 0$. Differentiating the formula (1) with respect to x_i, we conclude that
$$\frac{\partial V}{\partial x_i} = \frac{1}{n} \cdot \left(2(x_i - E) + \sum_{j=1}^{n} 2(E - x_j) \cdot \frac{\partial E}{\partial x_j} \right).$$
Since $\partial E/\partial x_i = 1/n$, we conclude that
$$\frac{\partial V}{\partial x_i} = \frac{2}{n} \cdot \left((x_i - E) + \sum_{j=1}^{n} (E - x_j) \cdot \frac{1}{n} \right).$$

Here, $\sum(E - x_j) = n \cdot E - \sum x_j$. By definition of the average E, this difference is 0, hence the above formula takes the form $\partial V/\partial x_i = (2/n) \cdot (x_i - E)$. So, this function attains the minimum when $x_i - E = 0$, i.e., when $x_i = E$.

Since $E = (1/n) \cdot (x_i + \sum_i' x_j)$, where $\sum_i'$ means the sum over all $j \neq i$, the equality $x_i = E$ means that $x_i = x_i/n + (1/n) \cdot \sum_i' x_j^{(0)}$. Moving terms containing x_i into the left-hand side and dividing by the coefficient at x_i, we conclude that the minimum is attained when $x_i = E_i \stackrel{\text{def}}{=} \dfrac{1}{n-1} \cdot \sum_i' x_j^{(0)}$, i.e., when x_i is equal to the arithmetic average E_i of all other elements.

2°. Let us now use the knowledge of a *global* minimum to describe where the desired function attains its minimum on the interval $\mathbf{x}_i$.

In our general description of non-negative quadratic functions of one variable, we mentioned that each such function is decreasing before the global minimum and increasing after it. Thus, for $x_i < E_i$, the function V is decreasing; for $x_i > E_i$, this function in increasing. Therefore:

- If $E_i \in \mathbf{x}_i$, the global minimum of the function V of one variable is attained within the interval $\mathbf{x}_i$, hence the minimum on the interval $\mathbf{x}_i$ is attained for $x_i = E_i$.
- If $E_i < \underline{x}_i$, the function V is increasing on the interval $\mathbf{x}_i$ and therefore, its minimum on this interval is attained when $x_i = \underline{x}_i$.
- Finally, if $E_i > \overline{x}_i$, the function V is decreasing on the interval $\mathbf{x}_i$ and therefore, its minimum on this interval is attained when $x_i = \overline{x}_i$.

3°. Let us reformulate the above conditions in terms of the average

$$E = \frac{1}{n} \cdot x_i + \frac{n-1}{n} \cdot E_i.$$

- In the first case, when $x_i = E_i$, we have $x_i = E = E_i$, so $E \in \mathbf{x}_i$.
- In the second case, we have $E_i < \underline{x}_i$ and $x_i = \underline{x}_i$. Therefore, in this case, $E < \underline{x}_i$.
- In the third case, we have $E_i > \overline{x}_i$ and $x_i = \overline{x}_i$. Therefore, in this case, $E > \underline{x}_i$.

Thus:

- If $E \in \mathbf{x}_i$, then we cannot be in the second or third cases. Thus, we are in the first case, hence $x_i = E$.
- If $E < \underline{x}_i$, then we cannot be in the first or the third cases. Thus, we are the second case, hence $x_i = \underline{x}_i$.
- If $E > \overline{x}_i$, then we cannot be in the first or the second cases. Thus, we are in the third case, hence $x_i = \overline{x}_i$.

4°. So, as soon as we determine the position of E with respect to all the bounds $\underline{x}_i$ and $\overline{x}_i$, we will have a pretty good understanding of all the values x_i at which the minimum is

attained. Hence, to find the minimum, we will analyze how the endpoints $\underline{x}_i$ and $\overline{x}_i$ divide the real line, and consider all the resulting sub-intervals.

Let the corresponding subinterval $[x_{(k)}, x_{(k+1)}]$ by fixed. For the i's for which $E \notin \mathbf{x}_i$, the values x_i that correspond to the minimal finite population variance are uniquely determined by the above formulas.

For the i's for which $E \in \mathbf{x}_i$ the selected value x_i should be equal to E. To determine this E, we can use the fact that E is equal to the average of all thus selected values x_i, in other words, that we should have

$$E = \frac{1}{n} \cdot \left(\sum_{i: \underline{x}_i \geq x_{(k+1)}} \underline{x}_i + (n - N_k) \cdot E + \sum_{j: \overline{x}_j \leq x_{(k)}} \overline{x}_j \right),$$

where $(n - N_k) \cdot E$ combines all the points for which $E \in \mathbf{x}_i$. Multiplying both sides of this equality by n and subtracting $n \cdot E$ from both sides, we conclude that $E = S_k/N_k$ – what we denoted, in the algorithm's description, by r_k. If thus defined r_k does not belong to the subinterval $[x_{(k)}, x_{(k+1)}]$, this contradiction with our initial assumption shows that there cannot be any minimum in this subinterval, so this subinterval can be easily dismissed.

The corresponding variance is denoted by V_k. If $N_k = 0$, this means that E belongs to all the intervals $\mathbf{x}_i$ and therefore, that the lower endpoint $\underline{V}$ is exactly 0 – so we assign $V_k = 0$. So, the algorithm is indeed correct.

5°. To complete the proof of the theorem, we must show that this algorithm indeed requires quadratic time. Indeed, sorting requires $O(n \cdot \log(n))$ steps (see, e.g., [1]), and the rest of the algorithm requires linear time ($O(n)$) for each of $2n$ subintervals, i.e., the total quadratic time. The theorem is proven.

NP-hardness of computing $\overline{V}$ means, crudely speaking, that there are no general ways for solving all particular cases of this problem (i.e., computing $\overline{V}$) in reasonable time.

However, we show that there are algorithms for computing $\overline{V}$ for many reasonable situations. Namely, we propose an efficient algorithm that computes $\overline{V}$ for the case when all the interval midpoints ("measured values") $\widetilde{x}_i = (\underline{x}_i + \overline{x}_i)/2$ are definitely different from each other, in the sense that the "narrowed" intervals $[\widetilde{x}_i - \Delta_i/n, \widetilde{x}_i + \Delta_i/n]$ – where $\Delta_i = (\underline{x}_i - \overline{x}_i)/2$ is the interval's half-width – do not intersect with each other.

This algorithm $\overline{\mathcal{A}}$ is as follows:

- First, we sort all $2n$ endpoints of the narrowed intervals $\widetilde{x}_i - \Delta_i/n$ and $\widetilde{x}_i + \Delta_i/n$ into a sequence $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(2n)}$. This enables us to divide the real line into $2n + 1$ segments ("small intervals") $[x_{(k)}, x_{(k+1)}]$, where we denoted $x_{(0)} \stackrel{\text{def}}{=} -\infty$ and $x_{(2n+1)} \stackrel{\text{def}}{=} +\infty$.
- Second, we compute $\underline{E}$ and $\overline{E}$ and pick all "small intervals" $[x_{(k)}, x_{(k+1)}]$ that intersect with $[\underline{E}, \overline{E}]$.
- For each of remaining small intervals $[x_{(k)}, x_{(k+1)}]$, for each i from 1 to n, we pick the following value of x_i:
  - if $x_{(k+1)} < \widetilde{x}_i - \Delta_i/n$, then we pick $x_i = \overline{x}_i$;

- if $x_{(k)} > \widetilde{x}_i + \Delta_i/n$, then we pick $x_i = \underline{x}_i$;
- for all other i, we consider both possible values $x_i = \overline{x}_i$ and $x_i = \underline{x}_i$.

As a result, we get one or several sequences of x_i. For each of these sequences, we check whether the average E of the selected values $x_1, \ldots, x_n$ is indeed within this small interval, and if it is, compute the variance by using the formula (1).

- Finally, we return the largest of the computed variances as $\overline{V}$.

Theorem 5. *The algorithm $\overline{\mathcal{A}}$ computes $\overline{V}$ is quadratic time for all the cases in which the "narrowed" intervals do not intersect with each other.*

This algorithm also works when, for some fixed k, no more than k "narrowed" intervals can have a common point:

Theorem 6. *For every positive integer k, the algorithm $\overline{\mathcal{A}}$ computes $\overline{V}$ is quadratic time for all the cases in which no more than k "narrowed" intervals can have a common point.*

Proof. Let us first show that this algorithm is indeed correct.

1°. Similarly to the proof of Theorem 4, let $x_1, \ldots, x_n$ be the values at which the finite population variance attain its maximum on the box $\mathbf{x}_1 \times \ldots \times \mathbf{x}_n$. If we fix the values of all the variables but one x_i, then V becomes a quadratic function of x_i. When the function V attains maximum over $x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n$, then this quadratic function of one variable will attain its maximum on the interval $\mathbf{x}_i$ at the point x_i.

We have already shown, in the proof of Theorem 4, that this quadratic function has a (global) minimum at $x_i = E_i$. Since this quadratic function of one variable is always non-negative, it cannot have a global maximum. Therefore, its maximum on the interval $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$ is attained at one of the endpoints of this interval.

An arbitrary quadratic function of one variable is symmetric with respect to the location of its global minimum, so its maximum on any interval is attained at the point which is the farthest from the minimum. There is exactly one point which is equally close to both endpoints of the interval $\mathbf{x}_i$: its midpoint $\widetilde{x}_i$. Depending on whether the global minimum is to the left, to the right, or exactly at the midpoint, we get the following three possible cases:

1. If the global minimum E_i is to the left of the midpoint $\widetilde{x}_i$, i.e., if $E_i < \widetilde{x}_i$, then the upper endpoint is the farthest from E_i. In this case, the maximum of the quadratic function is attained at its upper endpoint, i.e., $x_i = \overline{x}_i$.

2. Similarly, if the global minimum E_i is to the right of the midpoint $\widetilde{x}_i$, i.e., if $E_i > \widetilde{x}_i$, then the lower endpoint is the farthest from E_i. In this case, the maximum of the quadratic function is attained at its lower endpoint, i.e., $x_i = \underline{x}_i$.

3. If $E_i = \widetilde{x}_i$, then the maximum of V is attained at both endpoints of the interval $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$.

2°. In the third case, we have either $x_i = \underline{x}_i$ or $x_i = \overline{x}_i$. Depending on whether x_i is equal to the lower or to the upper endpoints, we can "combine" the corresponding situations with Cases 1 and 2. As a result, we arrive at the conclusion that one of the following two situations happen:

1. either $E_i \le \widetilde{x}_i$ and $x_i = \overline{x}_i$;

2. either $E_i \ge \widetilde{x}_i$ and $x_i = \underline{x}_i$.

3°. Similarly to the proof of Theorem 4, let us reformulate these conclusions in terms of the average E of the maximizing values $x_1, \ldots, x_n$.

By definition, $E_i = \frac{1}{n-1} \cdot \sum_i' x_j$, and $\sum_j' x_j = \sum_j x_j - x_i$. By definition of E, we have $\sum\limits_j x_j = n \cdot E$, therefore, $E_i = \frac{n \cdot E - x_i}{n-1}$. Let us apply this formula to the above three cases.

In the first case, we have $\widetilde{x}_i \ge E_i$. So, in terms of E, we get the inequality $\widetilde{x}_i \ge \frac{n \cdot E - x_i}{n-1}$. Multiplying both sides of this inequality by $n-1$, and using the fact that in this case, $x_i = \overline{x}_i = \widetilde{x}_i + \Delta_i$, we conclude that $(n-1) \cdot \widetilde{x}_i \ge n \cdot E - \widetilde{x}_i - \Delta_i$. Moving all the terms but $n \cdot E$ to the left-hand side and dividing by E, we get the following inequality: $E \le \widetilde{x}_i + \Delta_i/n$.

In the second case, we similarly get the inequality $E \le \widetilde{\Delta}_i/n$. So:

- In Case 1, we have $E \le \widetilde{x}_i + \Delta_i/n$ and $x_i = \overline{x}_i$.

- In Case 2, we have $E \ge \widetilde{x}_i - \Delta_i/n$ and $x_i = \underline{x}_i$.

Therefore:

- If $E < \widetilde{x}_i - \Delta_i/n$, this means that we cannot be in Case 2. So we must be in Case 1 and therefore, we must have $x_i = \overline{x}_i$.

- If $E > \widetilde{x}_i + \Delta_i/n$, this means that we cannot be in Case 1. So, we must be in Case 2 and therefore, we must have $x_i = \underline{x}_i$.

The only case when we do not know which endpoint for x_i we should choose is the case when E belongs to the narrowed interval $[\widetilde{x}_i - \Delta/n, \widetilde{x}_i + \Delta_i]$.

4°. Hence, once we know where E is with respect to the endpoints of all narrowed intervals, we can determine the values of all optimal x_i – except for those that are within this narrowed interval. Since we consider the case when no more than k narrowed intervals can have a common point, we have no more than k undecided values x_i. Trying all possible combinations of lower and upper endpoints for these $\le k$ values requires $\le 2^k$ steps.

Thus, the overall number of steps is $O(2^k \cdot n^2)$. Since k is a constant, the overall number of steps is thus $O(n^2)$. The theorem is proven.

3 Important Example: Interval Computations Related to Privacy in Statistical Databases

Need for privacy. Privacy is an important issue in the statistical analysis of human-related data. For example, to check whether in a certain geographic area, there is a gender-based discrimination, we can use the census data to check, e.g., whether for all people from this area who have the same same level of education, there is a correlation between salary and gender. One can think of numerous possible questions of this type related to different

sociological, political, medical, economic, and other questions. From this viewpoint, it is desirable to give researches *ability to perform* whatever *statistical analysis* of this data that is reasonable for their specific research.

On the other hand, we do not want to give them direct access to the raw census data, because a large part of the census data is *confidential*. For example, for most people (those who work in private sector) salary information is confidential. Suppose that a corporation is deciding where to built a new plant and has not yet decided between two possible areas. This corporation would benefit from knowing the average salary of people of needed education level in these two areas, because this information would help them estimate how much it will cost to bring local people on board. However, since salary information is confidential, the company should not be able to know the exact salaries of different potential workers.

The need for privacy is also extremely important for *medical* experiments, where we should be able to make statistical conclusions about, e.g., the efficiency of a new medicine without disclosing any potentially embarrassing details from the individual medical records.

Such databases in which the outside users cannot access individual records but can solicit statistical information are often called *statistical databases*.

How privacy is protected now and why it is not always sufficient. At present, one of the main (and most efficient) methods of protecting privacy in databases is the disaggregation of the data: instead of keeping a record with all the information about a person, we divide this record into several subrecords. For example, instead of keeping a single census record about a female professor leaving in New Jersey with three cats, we split this record into several subrecords: a subrecord about a person living in New Jersey with three cats (this subrecord will be useful for pet statistics), a subrecord about a female professor living in New Jersey (this subrecord will be useful for gender-based employment statistics), etc.

Such disaggregation helps to protect privacy. Indeed, if we keep the original full records, then we can narrow down a request in such a way that only one person will qualify: e.g., a request about the average salary of all female professors living in New Jersey with three cats etc. will eventually lead to an actual salary of that person. On the other hand, when records are disaggregated, whatever query we ask, be it an average salary of all New Jersey residents with exactly three cats or an average salary of all female professors from New Jersey, we will most likely not narrow down to a single person.

Disaggregation is very useful for protecting privacy, but it is not sufficient. Indeed, suppose that we keep a university salary database; for privacy protection, we keep all the records anonymous so the only information in the database is the actual salary values. What happens if we allow all possible statistical queries, including queries like "How many people have salary 83.6K or smaller"? By asking appropriate queries, we can find the salary values close to which the answer changes – and these values are exactly the actual salaries from the database. Thus, if we know that, e.g., the university president is the highest-paying professor, we will be able to get her salary as the largest of these actual salaries. How can we avoid this privacy violation?

Privacy leads to intervals. A natural way to fully describe a single real-valued random variable η is to provide the values of its cumulative density function (CDF)

$$F(x) = \text{Prob}(\eta \leq x)$$

for all possible real numbers x. Once we know $F(x)$, we can determine the values of all possible statistical characteristics of this random variable – e.g., its first moment, second moment, variance, etc. Thus, it is natural to allow the users to solicit the values of $F(x)$ for different x; from this information, the users will be able to reconstruct all other statistical characteristics.

For discrete data $x_1, \ldots, x_n$, the corresponding sample distribution – in which each value x_i occurs with probability $1/n$ – is described by the CDF $F(x)$ for which

$$F(x) = (1/n) \cdot \#\{i : x_i \leq x\}.$$

To get the full information about the data, we should allow the user to ask for the values $F(x)$ for all possible real numbers x. However, as we have mentioned, once we know the values $F(x)$ for all x, we can determine all the values x_i. Thus, if we want to keep privacy, we must only allow the users to know $F(x)$ for some fixed values $x^{(1)} \leq \ldots \leq x^{(m)}$. This way, instead of the actual values x_i, all we know is an *interval* $[x^{(k)}, x^{(k+1)}]$ that contains x_i. Intervals corresponding to different values are *almost disjoint*, i.e., either disjoint (intersect in at most one point) or identical. How can we compute statistical characteristics based on this information?

Theorem 7. *There exists a quadratic-time algorithm that computes the exact range* $\mathbf{V}$ *of the variance* V *for the case when intervals* $\mathbf{x}_i$ *of possible values of* x_i *are pairwise almost disjoint.*

Proof. Since there exists an algorithm that computes $\underline{V}$ in feasible time, it is sufficient to produce a feasible algorithm for computing $\overline{V}$.

According to the proof of Theorems 3, 5, and 6, the values $x_i \in \mathbf{x}_i$ that lead to the largest possible value of V satisfy the following property:

- if $E \leq \underline{x}_i$, then $x_i = \overline{x}_i$;
- if $E \geq \overline{x}_i$, then $x_i = \underline{x}_i$;
- if $E \in (\underline{x}_i, \overline{x}_i)$, then $x_i = \underline{x}_i$ or $x_i = \overline{x}_i$.

In order to use this property to compute $\overline{V}$, we test all possible locations of E in relation to the intervals $\mathbf{x}_i$: $E = \underline{x}_i$, $E = \overline{x}_i$, and $E \in (\underline{x}_i, \overline{x}_i)$ for different $i = 1, 2, \ldots, n$.

Let us first consider the cases when $E = \underline{x}_i$ (the case when $E = \overline{x}_i$ is treated similarly). In these cases, since the intervals $\mathbf{x}_i$ are almost disjoint, the above property uniquely determines the values x_i; thus, we can compute E, check whether it indeed satisfies the corresponding condition, and if yes, compute the corresponding value V.

Let us now consider the cases when $E \in (\underline{x}_i, \overline{x}_i)$. Let k denote the number of different intervals of such type, and let n_j, $j = 1, \ldots, k$ denote the number of intervals $\mathbf{x}_i$ that coincide with j-th interval. Then, $n = n_1 + \ldots + n_k$. For each of these k intervals $\mathbf{x}_j$, the values of x_i are uniquely determined when $\overline{x}_j \leq \underline{x}_i$ or $\overline{x}_i \leq \underline{x}_j$; for the remaining n_j values x_i, we

have $x_i = \underline{x}_i$ or $x_i = \overline{x}_i$. Modulo transposition, the resulting set of values $\{x_1, \ldots, x_n\}$ is uniquely determined by how many of these n_j x_i's are equal to $\overline{x}_i$. The number of such x_i's can be 0, 1, 2,..., $n_j + 1$. Thus, the total number of such combinations is equal to $n_j + 1$. Overall, for all j from 1 to k, we have $\sum_{j=1}^{k}(n_j + 1) = \sum_{j=1}^{k} n_j + k = n + k \leq 2n$ resulting sets $\{x_1, \ldots, x_n\}$. For each of these sets, we compute E, check that the resulting E is indeed inside the corresponding interval $\mathbf{x}_i$, and if it is, we compute V.

Thus, we have $\leq 2n + n = 3n$ cases, for each of which we need $O(n)$ computations to compute V. The largest of these V is the desired $\overline{V}$, and we compute it in time $\leq 3n \cdot O(n) = O(n^2)$. The proposition is proven.

Comment. Similar algorithms can be provided for computing the exact range of covariance between two interval-valued data sequences; in general, the problem of computing the range for covariance is NP-hard [15].

4 Second Step Beyond Intervals: Extension of Interval Arithmetic to Situations with Partial Information about Probabilities

Practical problem. In some practical situations, in addition to the lower and upper bounds on each random variable x_i, we know the bounds $\mathbf{E}_i = [\underline{E}_i, \overline{E}_i]$ on its mean E_i.

Indeed, in measurement practice (see, e.g., [16]), the overall measurement error Δx is usually represented as a sum of two components:

- a *systematic* error component $\Delta_s x$ which is defined as the expected value $E[\Delta x]$, and
- a *random* error component $\Delta_r x$ which is defined as the difference between the overall measurement error and the systematic error component: $\Delta_r x \stackrel{\text{def}}{=} \Delta x - \Delta_s x$.

In addition to the bound Δ on the overall measurement error, the manufacturers of the measuring instrument often provide an upper bound Δ_s on the systematic error component: $|\Delta_s x| \leq \Delta_s$.

This additional information is provided because, with this additional information, we not only get a bound on the accuracy of a single measurement, but we also get an idea of what accuracy we can attain if we use repeated measurements to increase the measurement accuracy. Indeed, the very idea that repeated measurements can improve the measurement accuracy is natural: we measure the same quantity by using the same measurement instrument several (N) times, and then take, e.g., an arithmetic average $\bar{x} = (\widetilde{x}^{(1)} + \ldots + \widetilde{x}^{(N)})/N$ of the corresponding measurement results $\widetilde{x}^{(1)} = x + \Delta x^{(1)}$, ..., $\widetilde{x}^{(N)} = x + \Delta x^{(N)}$.

- If systematic error is the only error component, then all the measurements lead to exactly the same value $\widetilde{x}^{(1)} = \ldots = \widetilde{x}^{(N)}$, and averaging does not change the value – hence does not improve the accuracy.

- On the other hand, if we know that the systematic error component is 0, i.e., $E[\Delta x] = 0$ and $E[\widetilde{x}] = x$, then, as $N \to \infty$, the arithmetic average tends to the actual value x. In this case, by repeating the measurements sufficiently many times, we can determine the actual value of x with an arbitrary given accuracy.

In general, by repeating measurements sufficiently many times, we can arbitrarily decrease the random error component and thus attain accuracy as close to Δ_s as we want.

When this additional information is given, then, after we performed a measurement and got a measurement result $\widetilde{x}$, then not only we get the information that the actual value x of the measured quantity belongs to the interval $\mathbf{x} = [\widetilde{x} - \Delta, \widetilde{x} + \Delta]$, but we can also conclude that the expected value of $x = \widetilde{x} - \Delta x$ (which is equal to $E[x] = \widetilde{x} - E[\Delta x] = \widetilde{x} - \Delta_s x$) belongs to the interval $\mathbf{E} = [\widetilde{x} - \Delta_s, \widetilde{x} + \Delta_s]$.

If we have this information for every x_i, then, in addition to the interval $\mathbf{y}$ of possible value of y, we would also like to know the interval of possible values of $E[y]$. This additional interval will hopefully provide us with the information on how repeated measurements can improve the accuracy of this indirect measurement. Thus, we arrive at the following problem.

Resulting optimization problem. In more optimization terms, we want to solve the following problem: given an algorithm computing a function $f(x_1, \ldots, x_n)$ from R^n to R; and values $\underline{x}_1$, $\overline{x}_1$, $\ldots$, $\underline{x}_n$, $\overline{x}_n$, $\underline{E}_1$, $\overline{E}_1$, $\ldots$, $\underline{E}_n$, $\overline{E}_n$, we want to find

$$\underline{E} \stackrel{\text{def}}{=} \min\{E[f(x_1, \ldots, x_n)] \mid \text{ all distributions of } (x_1, \ldots, x_n) \text{ for which}$$

$$x_1 \in [\underline{x}_1, \overline{x}_1], \ldots, x_n \in [\underline{x}_n, \overline{x}_n], E[x_1] \in [\underline{E}_1, \overline{E}_1], \ldots E[x_n] \in [\underline{E}_n, \overline{E}_n]\};$$

and $\overline{E}$ which is the maximum of $E[f(x_1, \ldots, x_n)]$ for all such distributions.

In addition to considering all possible distributions, we can also consider the case when all the variables x_i are independent.

Analog of straightforward interval computations. The main idea behind straightforward interval computations can be applied here as well. Namely, first, we find out how to solve this problem for the case when $n = 2$ and $f(x_1, x_2)$ is one of the standard arithmetic operations. Then, once we have an arbitrary algorithm $f(x_1, \ldots, x_n)$, we parse it and replace each elementary operation on real numbers with the corresponding operation on quadruples $(\underline{x}, \underline{E}, \overline{E}, \overline{x})$.

To implement this idea, we must therefore know how to, solve the above problem for elementary operations.

Solution. For *addition*, the answer is simple. Since $E[x_1 + x_2] = E[x_1] + E[x_2]$, if $y = x_1 + x_2$, there is only one possible value for $E = E[y]$: the value $E = E_1 + E_2$. This value does not depend on whether we have correlation or nor, and whether we have any information about the correlation. Thus, $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2$.

Similarly, the answer is simple for *subtraction:* if $y = x_1 - x_2$, there is only one possible value for $E = E[y]$: the value $E = E_1 - E_2$. Thus, $\mathbf{E} = \mathbf{E}_1 - \mathbf{E}_2$.

For *multiplication*, if the variables x_1 and x_2 are independent, then $E[x_1 \cdot x_2] = E[x_1] \cdot E[x_2]$. Hence, if $y = x_1 \cdot x_2$ and x_1 and x_2 are independent, there is only one possible value for $E = E[y]$: the value $E = E_1 \cdot E_2$; hence $\mathbf{E} = \mathbf{E}_1 \cdot \mathbf{E}_2$.

The first non-trivial case is the case of multiplication in the presence of possible correlation. When we know the exact values of E_1 and E_2, the solution to the above problem is as follows:

Theorem 8. *For multiplication $y = x_1 \cdot x_2$, when we have no information about the correlation,*

$$\underline{E} = \max(p_1 + p_2 - 1, 0) \cdot \overline{x}_1 \cdot \overline{x}_2 + \min(p_1, 1 - p_2) \cdot \overline{x}_1 \cdot \underline{x}_2 + \min(1 - p_1, p_2) \cdot \underline{x}_1 \cdot \overline{x}_2 +$$

$$\max(1 - p_1 - p_2, 0) \cdot \underline{x}_1 \cdot \underline{x}_2;$$

and

$$\overline{E} = \min(p_1, p_2) \cdot \overline{x}_1 \cdot \overline{x}_2 + \max(p_1 - p_2, 0) \cdot \overline{x}_1 \cdot \underline{x}_2 + \max(p_2 - p_1, 0) \cdot \underline{x}_1 \cdot \overline{x}_2 +$$

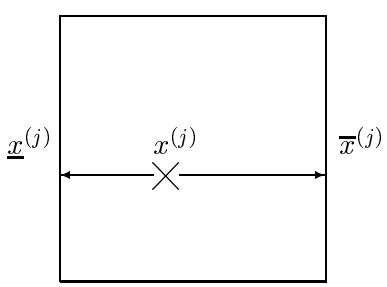
$$\min(1 - p_1, 1 - p_2) \cdot \underline{x}_1 \cdot \underline{x}_2,$$

where $p_i \stackrel{\text{def}}{=} (E_i - \underline{x}_i)/(\overline{x}_i - \underline{x}_i)$.

Proof. Let us show that a general distribution with $E[x_i] = E_i$ can be simplified without changing the values $E[x_i]$ and $E[x_1 \cdot x_2]$. Thus, to describe possible values of $E[x_1 \cdot x_2]$, we do not need to consider all possible distributions, it is sufficient to consider only the simplified ones.

We will describe the simplification for discrete distributions that concentrate on finitely many points $x^{(j)} = (x_1^{(j)}, x_2^{(j)})$, $1 \le j \le N$. An arbitrary probability distribution can be approximated by such distributions, so we do not lose anything by this restriction.

So, we have a probability distribution in which the point $x^{(1)}$ appears with the probability $p^{(1)}$, the point $x^{(2)}$ appears with the probability $p^{(2)}$, etc. Let us modify this distribution as follows: pick a point $x^{(j)} = (x_1^{(j)}, x_2^{(j)})$ that occurs with probability $p^{(j)}$, and replace it with two points: $\overline{x}^{(j)} = (\overline{x}_1, x_2^{(j)})$ with probability $p^{(j)} \cdot \overline{p}^{(j)}$ and $\underline{x}^{(j)} = (\underline{x}_1, x_2^{(j)})$ with probability $p^{(j)} \cdot \underline{p}^{(j)}$, where $\overline{p}^{(j)} \stackrel{\text{def}}{=} (x_1^{(j)} - \underline{x}_1)/(\overline{x}_1 - \underline{x}_1)$ and $\underline{p}^{(j)} \stackrel{\text{def}}{=} 1 - \overline{p}^{(j)}$:

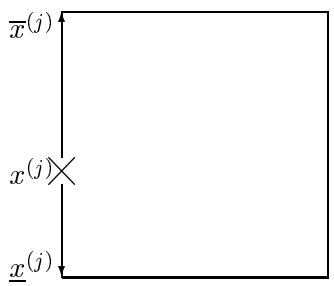


Here, the values $\overline{p}^{(j)}$ and $\underline{p}^{(j)} = 1 - \overline{p}^{(j)}$ are chosen in such a way that $\overline{p}^{(j)} \cdot \overline{x}_1 + \underline{p}^{(j)} \cdot \underline{x}_1 = x_1^{(j)}$. Due to this choice, $p^{(j)} \cdot \overline{p}^{(j)} \cdot \overline{x}_1 + p^{(j)} \cdot \underline{p}^{(j)} \cdot \underline{x}_1 = p^{(j)} \cdot x_1^{(j)}$, hence for the new distribution,

the mathematical expectation $E[x_1]$ is the same as for the old one. Similarly, we can prove that the values $E[x_2]$ and $E[x_1 \cdot x_2]$ do not change.

We started with a general discrete distribution with N points for each of which $x_1^{(j)}$ could be inside the interval $\mathbf{x}_1$, and we have a new distribution for which $\leq N-1$ points have the value x_1 inside this interval. We can perform a similar replacement for all N points and get a distribution with the same values of $E[x_1]$, $E[x_2]$, and $E[x_1 \cdot x_2]$ as the original one but for which, for every point, x_1 is equal either to $\underline{x}_1$, or to $\overline{x}_1$.

For the new distribution, we can perform a similar transformation relative to x_1 and end up – without changing the values x_1 – with the distribution for which always either $x_2 = \underline{x}_1$ or $x_2 = \overline{x}_2$:



Thus, instead of considering all possible distributions, it is sufficient to consider only distributions for which $x_1 \in \{\underline{x}_1, \overline{x}_1\}$ and $x_2 \in \{\underline{x}_2, \overline{x}_2\}$. In other words, it is sufficient to consider only distributions which are located in the four corner points $(\underline{x}_1, \underline{x}_2)$, $(\underline{x}_1, \overline{x}_2)$, $(\overline{x}_1, \underline{x}_2)$, and $(\overline{x}_1, \overline{x}_2)$ of the box $\mathbf{x}_1 \times \mathbf{x}_2$.

Such distribution can be characterized by the probabilities of these four points. These four probabilities must satisfy 3 conditions: that their sum is 1, that $E[x_1]$ is E_1, and that $E[x_2] = E_2$. Thus, we only have one parameter left; optimizing with respect to this parameter, we get the desired formulas for $\underline{E}$ and $\overline{E}$. The theorem is proven.

When we only know the intervals $\mathbf{E}_i$ of possible values of E_i, instead of the values p_i, we have the corresponding intervals $\mathbf{p}_i = (\mathbf{E}_i - \underline{x}_i)/(\overline{E}_i - \underline{x}_i)$. In terms of these intervals, we get the following results:

Theorem 9. *For multiplication under no information about dependence, to find $\underline{E}$, it is sufficient to consider the following combinations of p_1 and p_2:*

- $p_1 = \underline{p}_1$ *and* $p_2 = \underline{p}_2$*;* $p_1 = \underline{p}_1$ *and* $p_2 = \overline{p}_2$*;* $p_1 = \overline{p}_1$ *and* $p_2 = \underline{p}_2$*;* $p_1 = \overline{p}_1$ *and* $p_2 = \overline{p}_2$*;*
- $p_1 = \max(\underline{p}_1, 1 - \overline{p}_2)$ *and* $p_2 = 1 - p_1$ *(if* $1 \in \mathbf{p}_1 + \mathbf{p}_2$*); and*
- $p_1 = \min(\overline{p}_1, 1 - \underline{p}_2)$ *and* $p_2 = 1 - p_1$ *(if* $1 \in \mathbf{p}_1 + \mathbf{p}_2$*).*

The smallest value of $\underline{E}$ for all these cases is the desired lower bound $\underline{E}$.

Theorem 10. *For multiplication under no information about dependence, to find $\overline{E}$, it is sufficient to consider the following combinations of p_1 and p_2:*

- $p_1 = \underline{p}_1$ *and* $p_2 = \underline{p}_2$*;* $p_1 = \underline{p}_1$ *and* $p_2 = \overline{p}_2$*;* $p_1 = \overline{p}_1$ *and* $p_2 = \underline{p}_2$*;* $p_1 = \overline{p}_1$ *and* $p_2 = \overline{p}_2$*;*

- $p_1 = p_2 = \max(\underline{p}_1, \underline{p}_2)$ *(if* $\mathbf{p}_1 \cap \mathbf{p}_2 \neq \emptyset$*); and*
- $p_1 = p_2 = \min(\overline{p}_1, \overline{p}_2)$ *(if* $\mathbf{p}_1 \cap \mathbf{p}_2 \neq \emptyset$*).*

The largest value of $\overline{E}$ *for all these cases is the desired upper bound* $\overline{E}$*.*

Proof. We will prove Theorem 10; the proof of Theorem 9 is similar. The formula for $\overline{E}$ given in Theorem 8 can be simplified if we consider two cases: $p_1 \leq p_2$ and $p_1 \geq p_2$. To find the largest possible value $\overline{E}$ of E, it is sufficient to consider the largest possible values for each of these cases, and then take the largest of the resulting two numbers.

In each case, for a fixed p_2, the formula is linear in p_1. To find the maximum of a linear function on an interval, it is sufficient to consider this interval's endpoints. Thus, the maximum in p_1 is attained when either p_1 attains its smallest possible value $\underline{p}_1$, or when p_1 attains the largest possible value within this case; depending on p_2, this value is either $p_1 = \overline{p}_1$ or $p_1 = p_2$.

Thus, to find the maximum for each cases, it is sufficient to consider only the following cases: $p_1 = \underline{p}_1$, $p_1 = \overline{p}_1$, and $p_1 = p_2$. Similarly, it is sufficient to consider only the following cases for p_2: $p_2 = \underline{p}_2$, $p_2 = \overline{p}_2$, and $p_1 = p_2$.

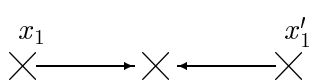
When $p_1 = p_2$, the probability $p_1 = p_2$ can take all possible values from the intersection $\mathbf{p}_1 \cap \mathbf{p}_2$. the formula for $\overline{E}$ is linear in p_1, so to find its maximum, it is sufficient to consider the endpoints of the interval $\mathbf{p}_1 \cap \mathbf{p}_2$, i.e., the values $p_1 = p_2 = \max(\underline{p}_1, \underline{p}_2)$ and $p_1 = p_2 = \min(\overline{p}_1, \overline{p}_2)$. The theorem is proven.

For the *inverse* $y = 1/x_1$, the finite range is possible only when $0 \notin \mathbf{x}_1$. Without losing generality, we can consider the case when $0 < \underline{x}_1$. In this case, methods presented in [17] lead to the following bound:

Theorem 11. *For the inverse* $y = 1/x_1$*, the range of possible values of* E *is* $\mathbf{E} = [1/E_1, p_1/\overline{x}_1 + (1-p_1)/\underline{x}_1]$.

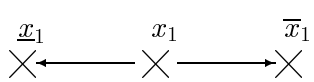
(Here p_1 denotes the same value as in Theorem 8).

Proof. For $x_1 > 0$, the function $f(x_1) \stackrel{\text{def}}{=} 1/x_1$ is convex: for every x_1, x_1', and $\alpha \in [0,1]$, we have $f(\alpha \cdot x_1 + (1-\alpha) \cdot x_1') \leq \alpha \cdot f(x_1) + (1-\alpha) \cdot f(x_1')$. Hence, if we are looking for a minimum of $E[1/x_1]$, we can replace every two points from the probability distribution with their average, and the resulting value of $E[1/x_1]$ will only decrease:



So, the minimum is attained when the probability distribution is concentrated on a single value – which has to be E_1. Thus, the smallest possible value of $E[1/x_1]$ is $1/E_1$.

Due to the same convexity, if we want maximum of $E[1/x_1]$, we should replace every value $x_1 \in [\underline{x}_1, \overline{x}_1]$ by a probabilistic combination of the values $\underline{x}_1, \overline{x}_1$:



So, the maximum is attained when the probability distribution is concentrated on these two endpoints $\underline{x}_1$ and $\overline{x}_1$. Since the average of x_1 should be equal to E_1, we can, similarly to

the proof of Theorem 8, conclude that in this distribution, $\overline{x}_1$ occurs with probability p_1, and $\underline{x}_1$ occurs with probability $1-p_1$. For this distribution, the value $E[1/x_1]$ is exactly the upper bound from the formulation of the theorem. The theorem is proven.

Theorem 12. *For minimum $y = \min(x_1, x_2)$, when x_1 and x_2 are independent, we have $\overline{E} = \min(E_1, E_2)$ and*

$$\underline{E} = p_1 \cdot p_2 \cdot \min(\overline{x}_1, \overline{x}_2) + p_1 \cdot (1-p_2) \cdot \min(\overline{x}_1, \underline{x}_2) + (1-p_1) \cdot p_2 \cdot \min(\underline{x}_1, \overline{x}_2) + (1-p_1) \cdot (1-p_2) \cdot \min(\underline{x}_1, \underline{x}_2).$$

Theorem 13. *For maximum $y = \min(x_1, x_2)$, when x_1 and x_2 are independent, we have $\underline{E} = \max(E_1, E_2)$ and*

$$\overline{E} = p_1 \cdot p_2 \cdot \max(\overline{x}_1, \overline{x}_2) + p_1 \cdot (1-p_2) \cdot \max(\overline{x}_1, \underline{x}_2) + (1-p_1) \cdot p_2 \cdot \max(\underline{x}_1, \overline{x}_2) + (1-p_1) \cdot (1-p_2) \cdot \max(\underline{x}_1, \underline{x}_2).$$

Proof. We will prove Theorem 12; the proof of Theorem 13 is similar. Since $\min(x_1, x_2) \le x_1$, we have $E[\min(x_1, x_2)] \le E[x_1] = E_1$. Similarly, $E[\min(x_1, x_2)] \le E_2$, hence, $E[\min(x_1, x_2)] \le \min(E_1, E_2)$. The value $\min(E_1, E_2)$ is possible when $x_1 = E_1$ with probability 1 and $x_2 = E_2$ with probability 1. Thus, $\min(E_1, E_2)$ is the exact upper bound for $E[\min(x_1, x_2)]$.

For each x_2, the function $x_1 \to \min(x_1, x_2)$ is concave; therefore, if we replace each point $x^{(j)} = (x_1^{(j)}, x_2^{(j)})$ by the corresponding probabilistic combination of the points $(\underline{x}_1, x_2^{(j)})$ and $(\overline{x}_1, x_2^{(j)})$ (as in the proof of Theorem 11), we preserve $E[x_1]$ and $E[x_2]$ and decrease the value $E[\min(x_1, x_2)]$. Thus, when we are looking for the smallest possible value of $E[\min(x_1, x_2)]$, it is sufficient to consider only the distributions for which x_1 is located at one of the endpoints $\underline{x}_1$ or $\overline{x}_1$. Similarly to the proof of Theorem 8, the probability of $\overline{x}_1$ is equal to p_1.

Similarly, we can conclude that to find the largest possible value of $E[\min(x_1, x_2)]$, it is sufficient to consider only distributions in which x_2 can take only two values: $\underline{x}_2$ and $\overline{x}_2$. To get the desired value of E_2, we must have $\overline{x}_2$ with probability p_1 and $\underline{x}_2$ with probability $1-p_2$.

Since we consider the case when x_1 and x_2 are independent, and each of them takes two possible values, we can conclude that $x = (x_1, x_2)$ can take four possible values $(\underline{x}_1, \underline{x}_2)$, $(\underline{x}_1, \overline{x}_2)$, $(\overline{x}_1, \underline{x}_2)$, and $(\overline{x}_1, \overline{x}_2)$, and the probability of each of these values is equal to the product of the probabilities corresponding to x_1 and x_2. For this distribution, $E[\min(x_1, x_2)]$ is exactly the expression from the formulation of the theorem. Theorem 12 is proven.

Theorem 14. *For minimum $y = \min(x_1, x_2)$, when we have no information about the correlation between x_1 and x_2, we have $\overline{E} = \min(E_1, E_2)$,*

$$\underline{E} = \max(p_1 + p_2 - 1, 0) \cdot \min(\overline{x}_1, \overline{x}_2) + \min(p_1, 1-p_2) \cdot \min(\overline{x}_1, \underline{x}_2) + \min(1-p_1, p_2) \cdot \min(\underline{x}_1, \overline{x}_2) + \max(1-p_1-p_2, 0) \cdot \min(\underline{x}_1, \underline{x}_2).$$

Theorem 15. *For maximum $y = \max(x_1, x_2)$, when we have no information about the correlation between x_1 and x_2, we have $\underline{E} = \max(E_1, E_2)$ and*

$$\overline{E} = \min(p_1, p_2) \cdot \max(\overline{x}_1, \overline{x}_2) + \max(p_1 - p_2, 0) \cdot \max(\overline{x}_1, \underline{x}_2) +$$

$$\max(p_2 - p_1, 0) \cdot \max(\underline{x}_1, \overline{x}_2) + \min(1 - p_1, 1 - p_2) \cdot \max(\underline{x}_1, \underline{x}_2).$$

Proof. We will prove Theorem 14; the proof of Theorem 15 is similar. Similarly to the proof of Theorem 12, we can conclude that $\min(E_1, E_2)$ is the attainable upper bound for $E[\min(x_1, x_2)]$. Due to convexity, to find the lower bound for $E[\min(x_1, x_2)]$, it is sufficient to consider distributions located at the four corners of the box $\mathbf{x}_1 \times \mathbf{x}_2$. Similar to the proof of Theorem 8, we conclude that such distribution can be characterized by a single parameter. Optimizing with respect to this parameter, we get the desired formula for $\underline{E}$. The theorem is proven.

Similar formulas can be produced for the cases when there is a strong correlation between x_i: namely, when x_1 is (non-strictly) increasing or decreasing in x_2.

From Elementary Arithmetic Operations to General Algorithms When we have a complex algorithm f, then a step-by-step approach leads to excess width. How can we find the actual range of $E = E[y]$?

At first glance, the exact formulation of this problem requires that we use infinitely many variables, because we must describe all possible probability distributions on the box $\mathbf{x}_1 \times \ldots \times \mathbf{x}_n$ (or, in the independent case, all possible tuples consisting of distributions on all n intervals $\mathbf{x}_1, \ldots, \mathbf{x}_n$). It turns out, however, that we can reformulate these problems in equivalent forms that require only finitely many variables:

Theorem 16. *For a general continuous function $f(x_1, \ldots, x_n)$, $\underline{E}$ is a solution to the following optimization problem: $\sum\limits_{j=0}^{n} p^{(j)} \cdot f(x_1^{(j)}, \ldots, x_n^{(j)}) \to \min$ under the conditions*

$$\sum_{k=0}^{n} p^{(k)} = 1; \quad p^{(j)} \geq 0; \quad \underline{x}_i \leq x_i^{(j)} \leq \overline{x}_i; \quad \underline{E}_i \leq \sum_{j=0}^{n} p^{(j)} \cdot x_i^{(j)} \leq \overline{E}_i \quad (\text{for all } i, j),$$

and $\overline{E}$ is a solution to $\sum\limits_{j=0}^{n} p^{(j)} \cdot f(x_1^{(j)}, \ldots, x_n^{(j)}) \to \max$ under the same constraints.

Proof. In terms of the unknown probabilities $p^{(j)}$, we are minimizing a linear function under linear constraints (equalities and inequalities). Geometrically, the set of all points that satisfy several linear constraints is a polytope. It is well known that to find the minimum of a linear function on a polytope, it is sufficient to consider its vertices (this idea is behind linear programming). In algebraic terms, a vertex can be characterized by the fact that for N variables, N of the original constrains are equalities. Thus, in our case, all but n probabilities $p^{(j)}$ must be equal to 0. The theorem is proven.

5 Open Problems

So far, we have provided explicit formulas for the elementary arithmetic operations $f(x_1, \ldots, x_n)$ for the case when we know the first order moments. What if, in addition to that, we have some information about second order (and/or higher order) moments of x_i? What will we be then able to conclude about the moments of y? Partial answers to this question are given in [11, 17, 19]; it is desirable to find a general answer.

Similarly to Theorem 16, we can reduce the corresponding problems to the constraint optimization problems with finitely many variables. For example, when, in addition to intervals $\mathbf{E}_i$ that contain the first moments $E[x_i]$, we know the intervals $\mathbf{E}_{ik}$ that contain the second moments $E[x_i \cdot x_k]$, then the corresponding bounds $\underline{E}$ and $\overline{E}$ on $E[y]$ can be computed by solving the problems $\sum\limits_{j=0}^{N} p^{(j)} \cdot f(x_1^{(j)}, \ldots, x_n^{(j)}) \to \min(\max)$ under the conditions

$$\sum_{j=0}^{N} p^{(j)} = 1; \quad p^{(j)} \geq 0; \quad \underline{x}_i \leq x_i^{(j)} \leq \overline{x}_i; \quad \underline{E}_i \leq \sum_{j=0}^{n} p^{(j)} \cdot x_i^{(j)} \leq \overline{E}_i;$$

$$\underline{E}_{ik} \leq \sum_{j=0}^{n} p^{(j)} \cdot x_i^{(j)} \cdot x_k^{(j)} \leq \overline{E}_{ik},$$

where $N = n(n+1)/2$.

It is desirable to find explicit analytical expressions for these bounds, at least for the case when $n = 2$ and $f(x_1, \ldots, x_n)$ is an elementary arithmetic operation.

Acknowledgments

This work was supported in part by NASA grant NCC5-209, by the AFOSR grant F49620-00-1-0365, by NSF grants EAR-0112968 and EAR-0225670, by IEEE/ACM SC2001 and SC2002 Minority Serving Institutions Participation Grants, by a research grant from Sandia National Laboratories as part of the Department of Energy Accelerated Strategic Computing Initiative (ASCI), and by Small Business Innovation Research grant 9R44CA81741 to Applied Biomathematics from the National Cancer Institute (NCI), a component of NIH.

This research was partly done when the author was a Visiting Faculty Member at the Fields Institute for Research in Mathematical Sciences (Toronto, Canada). The author is thankful to the anonymous referees for valuable suggestions.

References

[1] Cormen, Th. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2001) "Introduction to Algorithms", MIT Press, Cambridge, MA.

[2] Ferson S., Ginzburg L., Kreinovich V., Longpré L., and Aviles M. (2002), "Computing Variance for Interval Data is NP-Hard", *ACM SIGACT News, vol. 33*, 108–118.

[3] Ferson S., Ginzburg L., Kreinovich V., and Lopez J. (2002), "Absolute Bounds on the Mean of Sum, Product, etc.: A Probabilistic Extension of Interval Arithmetic", *Extended Abstracts of the 2002 SIAM Workshop on Validated Computing*, Toronto, Canada, May 23–25, 70–72.

[4] Hansen, E. (1997), "Sharpness in interval computations", *Reliable Computing, vol. 3*, 7–29.

[5] Jaulin L., Keiffer M., Didrit O., and Walter E. (2001), "Applied Interval Analysis", Springer-Verlag, Berlin.

[6] Kearfott R. B. (1996), "Rigorous Global Search: Continuous Problems", Kluwer, Dordrecht.

[7] Kearfott R. B. and Kreinovich V., eds. (1996), "Applications of Interval Computations" (Pardalos. P. M., Hearn, D., "Applied Optimization", Vol. 3), Kluwer, Dordrecht.

[8] Kreinovich V. (2000), "Beyond Interval Systems: What Is Feasible and What Is Algorithmically Solvable?", In: Pardalos P. M., ed., "Approximation and Complexity in Numerical Optimization: Continuous and Discrete Problems", Kluwer, Dordrecht, 364–379.

[9] Kreinovich V., Lakeyev A., Rohn J., and Kahl P. (1997), "Computational Complexity and Feasibility of Data Processing and Interval Computations" (Pardalos. P. M., Hearn, D., "Applied Optimization", Vol. 10), Kluwer, Dordrecht.

[10] Kreinovich V., Nguyen H. T., Ferson S., and Ginzburg L. (2002), "From Computation with Guaranteed Intervals to Computation with Confidence Intervals", *Proc. 21st Int'l Conf. of North American Fuzzy Information Processing Society NAFIPS'2002*, New Orleans, Louisiana, 418–422.

[11] Kuznetsov V. P. (1991), "Interval Statistical Models", Radio i Svyaz, Moscow (in Russian).

[12] Moore R. E. (1979), "Methods and Applications of Interval Analysis", SIAM, Philadelphia.

[13] Nivlet P., Fournier F., and Royer J. (2001), "A new methodology to account for uncertainties in 4-D seismic interpretation", *Proceedings of the 71st Annual International Meeting of the Society of Exploratory Geophysics SEG'2001*, San Antonio, Texas, September 9–14, 1644–1647.

[14] Nivlet P., Fournier F., and Royer J. (2001), "Propagating interval uncertainties in supervised pattern recognition for reservoir characterization", *Proceedings of the 2001 Society of Petroleum Engineers Annual Conference SPE'2001*, New Orleans, Louisiana, September 30–October 3, paper SPE-71327.

[15] Osegueda, R., Kreinovich, V., Potluri, L., and Aló R. (2002), "Non-Destructive Testing of Aerospace Structures: Granularity and Data Mining Approach", *Proceedings of FUZZ-IEEE'2002*, Honolulu, Hawaii, May 12–17, *Vol. 1*, 685–689.

[16] Rabinovich S. (1993), "Measurement Errors: Theory and Practice", American Institute of Physics, New York.

[17] Rowe, N. C. (1988), "Absolute bounds on the mean and standard deviation of transformed data for constant-sign-derivative transformations", *SIAM Journal of Scientific Statistical Computing, vol. 9*, 1098–1113.

[18] Vavasis S. A. (1991), "Nonlinear Optimization: Complexity Issues", Oxford University Press, N.Y.

[19] Walley, P. (1991), "Statistical Reasoning with Imprecise Probabilities", Chapman and Hall, N.Y.