10^{th} IMEKO TC7 International Symposium
June 30–July 2, 2004, Saint-Petersburg, Russia

PROBABILITIES, INTERVALS, WHAT NEXT? EXTENSION OF INTERVAL COMPUTATIONS TO SITUATIONS WITH PARTIAL INFORMATION ABOUT PROBABILITIES

Vladik Kreinovich[1], *Gennady N. Solopchenko*[2], *Scott A. Ferson*[3], *Lev Ginzburg*[3], *Richard Aló*[4]

[1]Computer Science, University of Texas, El Paso, USA
[2]State Technical Univ. of St. Petersburg, Russia
[3]Applied Biomathematics, 100 North Country Road, Setauket, NY 11733, USA
[4]University of Houston-Downtown, One Main Street, Houston, TX 77002, USA

Abstract. In many real-life situations, we are interested in the value of a physical quantity y that is difficult or impossible to measure directly. To estimate y, we find some easier-to-measure quantities $x_1, \ldots, x_n$ which are related to y by a known relation $y = f(x_1, \ldots, x_n)$. Measurements are never 100% accurate; hence, the measured values $\widetilde{x}_i$ are different from x_i, and the resulting estimate $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ is different from the desired value $y = f(x_1, \ldots, x_n)$. How different?

Traditional engineering to error estimation in data processing assumes that we know the probabilities of different measurement error $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i$.

In many practical situations, we only know the upper bound Δ_i for this error; hence, after the measurement, the only information that we have about x_i is that it belongs to the interval $\mathbf{x}_i \stackrel{\text{def}}{=} [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$. In this case, it is important to find the range $\mathbf{y}$ of all possible values of $y = f(x_1, \ldots, x_n)$ when $x_i \in \mathbf{x}_i$.

We start with a brief overview of the corresponding *interval computation* problems. We then discuss what to do when, in addition to the upper bounds Δ_i, we have some partial information about the probabilities of different values of Δx_i.

Keywords: indirect measurements, total error, interval computations

1. FORMULATION OF THE PROBLEM

1.1. Why indirect measurements?

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.

1.2. 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 [11].

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., [3,4].

1.3. 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,

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

$$[\underline{a}, \overline{a}] \cdot [\underline{b}, \overline{b}] =$$

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

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 [8], 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).

1.4. 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., [11]), 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: $|\Delta_s x| \le \Delta_s$.

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

$$\bar{x} = \frac{\widetilde{x}^{(1)} + \ldots + \widetilde{x}^{(N)}}{N}$$

of the corresponding measurement results

$$\widetilde{x}^{(1)} = x + \Delta x^{(1)}, \ldots, \widetilde{x}^{(N)} = x + \Delta x^{(N)}.$$

- 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:

1.5. New problem in precise terms

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.

2. HOW WE SOLVE THIS PROBLEM

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.

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 1. *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

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

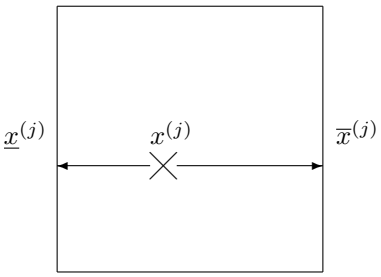
where $p_i \stackrel{\text{def}}{=} (E_i - \underline{x}_i)/(\overline{x}_i - \underline{x}_i)$.

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

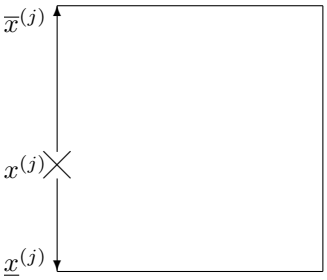
$p^{(j)} \cdot \underline{p}^{(j)}$, where $\overline{p}^{(j)} \stackrel{\text{def}}{=} (x_1^{(j)} - \underline{x}_1)/(\overline{x}_1 - \underline{x}_1)$ and $\underline{p}^{(j)} \stackrel{\text{def}}{=} 1 - \overline{p}^{(j)}$:



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 $\le N - 1$ 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 2. *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 3. *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 3; the proof of Theorem 2 is similar. The formula for $\overline{E}$ given in Theorem 1 can be simplified if we consider two cases: $p_1 \le p_2$ and $p_1 \ge p_2$. 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 [12] lead to the following bound:

Theorem 4. *For the inverse $y = 1/x_1$, the range of possible values of E is*

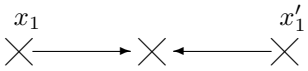
$$\mathbf{E} = [1/E_1, p_1/\overline{x}_1 + (1-p_1)/\underline{x}_1].$$

(Here p_1 denotes the same value as in Theorem 1).

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

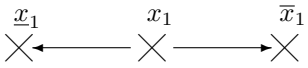
$$f(\alpha \cdot x_1 + (1-\alpha) \cdot x_1') \le \alpha \cdot f(x_1) + (1-\alpha) \cdot f(x_1').$$

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 1, 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 5. *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 6. *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 5; the proof of Theorem 6 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 4), 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 1, 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 5 is proven.

Theorem 7. *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 8. *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 7; the proof of Theorem 8 is similar. Similarly to the proof of Theorem 5, 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 1, 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.

3. ADDITIONAL RESULTS

The above techniques assume that we already know the moments etc., but how can we compute them based on the measurement results? For example, 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?

It turns out that most such problems are computationally difficult (to be more precise, NP-hard), and we provide feasible algorithms that compute these bounds under reasonable easily verifiable conditions [1,5,9].

ACKNOWLEDGMENTS

This work was supported in part by NASA under cooperative agreement NCC5-209, by the Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant F49620-00-1-0365, by NSF grants EAR-0112968, EAR-0225670, and EIA-0321328, by the Army Research Laboratories grant DATM-05-02-C-0046, by the Hewlett-Packard equipment grants 89955.1 and 89955.2, and by the Personal Interface AccessGrid awarded by the Education, Outreach and Training Partnership for Advanced Computational Infrastructure EOT-PACI.

This work was partly performed during V.K.'s visit to Brazil. This visit was sponsored by the Brazilian funding agency CTINFO/CNPq.

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

REFERENCES

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

[2] S.A. Ferson, V. Kreinovich, L. Ginzburg, D.S. Myers, K. Sentz, *Constructing Probability Boxes and Dempster-Shafer Structures*, Sandia National Laboratories, Report SAND2002-4015, January 2003.

[2a] L. Jaulin, M. Keiffer, O. Didrit, E. Walter, *Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control and Robotics*, Springer-Verlag, London, 2001.

[4] R.B. Kearfott, V. Kreinovich (eds.), *Applications of Interval Computations*, Kluwer, Dordrecht, 1996.

[5] V. Kreinovich, "Probabilities, Intervals, What Next? Optimization Problems Related to Extension of Interval Computations to Situations with Partial Information about Probabilities", *Journal of Global Optimization* (to appear).

[6] V. Kreinovich, S.A. Ferson, "A New Cauchy-Based Black-Box Technique for Uncertainty in Risk Analysis", *Reliable Engineering and Systems Safety* (to appear).

[7] V. Kreinovich, S.A. Ferson, and L. Ginzburg, "Exact Upper Bound on the Mean of the Product of Many Random Variables With Known Expectations", *Reliable Computing*, **9**(6), 441–463, 2003.

[8] V. Kreinovich, A. Lakeyev, J. Rohn, P. Kahl, *Computational complexity and feasibility of data processing and interval computations*, Kluwer, Dordrecht, 1997.

[9] V. Kreinovich and L. Longpré, "Computational complexity and feasibility of data processing and interval computations, with extension to cases when we have partial information about probabilities", In: V. Brattka, M. Schroeder, K. Weihrauch, N. Zhong, *Proc. of the Conf. on Computability and Complexity in Analysis CCA'2003*, Cincinnati, Ohio, USA, August 28–30, 2003, pp. 19–54.

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

[11] S. Rabinovich, *Measurement Errors: Theory and Practice*, Amer. Inst. Phys., N.Y., 1993.

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

[13] R. Trejo, V. Kreinovich, "Error Estimations for Indirect Measurements: Randomized vs. Deterministic Algorithms For 'Black-Box' Programs", In: S. Rajasekaran et al. (eds.), *Handbook on Randomized Computing*, Kluwer, 2001, pp. 673–729.

[14] P. Walley, *Statistical Reasoning with Imprecise Probabilities*, Chapman & Hall, N.Y., 1991.

Authors: Corresponding author: Vladik Kreinovich, Department of Computer Science, University of Texas at El Paso, 500 W. University, El Paso, TX 79968, USA, phone 1-915-747-6951, fax 1-915-747-5030, email vladik@cs.utep.edu.

Gennady N. Solopchenko, State Technical Univ. of St. Petersburg, Information-Measuring Department, Politechnicheskaya 29, 195251 St. Petersburg, Russia.

Scott A. Ferson, Lev Ginzburg, Applied Biomathematics, 100 North Country Road, Setauket, NY 11733, USA, email scott@ramas.com.

Richard Aló, Center for Computational Sciences and Advanced Distributed Simulation, University of Houston-Downtown, One Main Street, Houston, TX 77002, USA, email ralo@uh.edu.