Interval Computations and Interval-Related Statistical Techniques: Tools for Estimating Uncertainty of the Results of Data Processing and Indirect Measurements

Vladik Kreinovich

Dept. Computer Science, University of Texas, El Paso, TX 79968-0518, USA
vladik@utep.edu

Summary. In many practical situations, we only know the upper bound Δ on the (absolute value of the) measurement error Δx, i.e., we only know that the measurement error is located on the interval $[-\Delta, \Delta]$. The traditional engineering approach to such situations is to assume that Δx is uniformly distributed on $[-\Delta, \Delta]$, and to use the corresponding statistical techniques. In some situations, however, this approach underestimates the error of indirect measurements. It is therefore desirable to directly process this interval uncertainty. Such "interval computations" methods have been developed since the 1950s. In this chapter, we provide a brief overview of related algorithms, results, and remaining open problems.

Table 1. List of Symbols

y	actual (unknown) value of the desired quantity
x_i	actual (unknown) value of the i-th auxiliary quantity
n	number of auxiliary quantities
$f(x_1, \ldots, x_n)$	relation between $x_1, \ldots, x_n$ and y: $y = f(x_1, \ldots, x_n)$
$\widetilde{x}_i$	result of measuring x_i
$\widetilde{y}$	estimate for y: the result of indirect measurement
$\Delta x_i = \widetilde{x}_i - x_i$	the i-th measurement error
$\Delta y = \widetilde{y} - y$	inaccuracy of indirect measurement
Δ_i	upper bound on the absolute value $\lvert\Delta x_i\rvert$ of Δx_i
Δ	resulting upper bound on $\lvert\Delta y\rvert$
$\boldsymbol{x}_i = [\underline{x}_i, \overline{x}_i]$	interval of possible values of x_i
$\boldsymbol{y} = [\underline{y}, \overline{y}]$	interval of possible values of the desired quantity y
$\boldsymbol{Y}$	enclosure for $\boldsymbol{y}$, i.e., an interval such that $\boldsymbol{y} \subseteq \boldsymbol{Y}$
$\rho(x)$	probability density function (pdf)
$F(x)$	cumulative distribution function (cdf)
δx_i	simulated value of the i-th measurement error
$a^{(k)}$	value of a quantity a on the k-th iteration

1 Importance of Data Processing and 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 $x_1, \ldots, x_i, \ldots, x_n$ which are related to y by a known relation $y = f(x_1, \ldots, x_i, \ldots, x_n)$; 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 $x_1, \ldots, x_i, \ldots, x_n$, and then we use the results $\widetilde{x}_1, \ldots, \widetilde{x}_i, \ldots, \widetilde{x}_n$ of these measurements to compute an estimate $\widetilde{y}$ for y as $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_i, \ldots, \widetilde{x}_n)$; see Fig. 1.

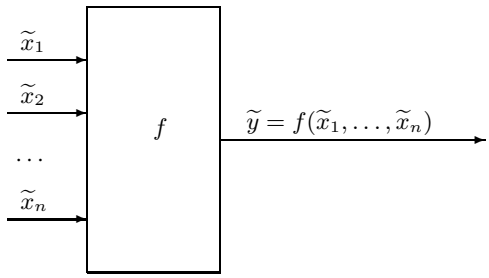

Fig. 1. Indirect measurement

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 chapter, 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.

2 Estimating Uncertainty for the Results of Data Processing and Indirect Measurements: An Important Metrological Problem

Measurements 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 $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_i, \ldots, \widetilde{x}_n)$ of data

processing is, in general, different from the actual value $y = f(x_1, \ldots, x_i, \ldots, x_n)$ of the desired quantity y.

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.

3 Uncertainty of Direct Measurements: Brief Description, Limitations, Need for Overall Error Bounds (i.e., Interval Uncertainty)

Upper bounds on measurement errors. What do we know about the errors Δx_i of direct measurements? First, the manufacturers of a measuring device usually provide us with an upper bound Δ_i for the (absolute value of) possible measurement errors, i.e., with the bound Δ_i for which we are guaranteed that $|\Delta x_i| \le \Delta_i$.

The need for such a bound comes from the very nature of a measurement process. Indeed, if no such bound is provided, this means that the actual value x_i can be arbitrarily different from the "measurement result" $\widetilde{x}_i$ as possible. Such a value $\widetilde{x}_i$ is not a measurement, it is a wild guess.

Since the (absolute value of the) measurement error $\Delta x_i = \tilde{x}_i - x_i$ is bounded by the given bound Δ_i, we can therefore guarantee that the actual (unknown) value of the desired quantity belongs to the interval $\boldsymbol{x}_i \stackrel{\text{def}}{=} [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$.

Example. If the measured value of a quantity is $\widetilde{x}_i = 1.0$, and the upper bound Δ_i on the measurement error is 0.1, this means that the (unknown) actual value of the measured quantity can be anywhere between $1 - 0.1 = 0.9$ and $1 + 0.1 = 1.1$, i.e., that it can take any value from the interval $[0.9, 1.1]$.

Probabilities. 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 *probabilities* of different values Δx_i within this interval; see, e.g., Rabinovich [1].

In most practical applications, it is assumed that the corresponding measurement errors are normally distributed with 0 mean and known standard deviations.

Numerous engineering techniques are known (and widely used) for processing this uncertainty; see, e.g., Rabinovich [1].

In practice, we can determine the desired probabilities of different values of Δx_i by comparing the results $\widetilde{x}_i$ of measuring with this instrument with the results $\widetilde{x}_i^{\,\text{st}}$ 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 $\widetilde{\Delta} x_i = \widetilde{x}_i - \widetilde{x}_i^{\,\text{st}}$ is close to the desired probability distribution for the measurement error $\Delta x_i = \widetilde{x}_i - x_i$.

Sometimes, we do not know probabilities. There are two cases 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.

4 Data Processing and Indirect Measurements Under Interval Uncertainty: The Main Problem of Interval Computations

In the case when the only information we have is the upper bound on the measurement error, 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 $\boldsymbol{x}_i = [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$. In such situations, the only information that we have about the (unknown) actual value of $y = f(x_1, \ldots, x_i, \ldots, x_n)$ is that y belongs to the range $\boldsymbol{y} = [\underline{y}, \overline{y}]$ of the function f over the box $\boldsymbol{x}_1 \times \ldots \times \boldsymbol{x}_i \times \ldots \times \boldsymbol{x}_n$:

$$\boldsymbol{y} = [\underline{y}, \overline{y}] = \{f(x_1, \ldots, x_i, \ldots, x_n) \,|\, x_1 \in \boldsymbol{x}_1, \ldots, x_i \in \boldsymbol{x}_i, \ldots, x_n \in \boldsymbol{x}_n\}. \quad (1)$$

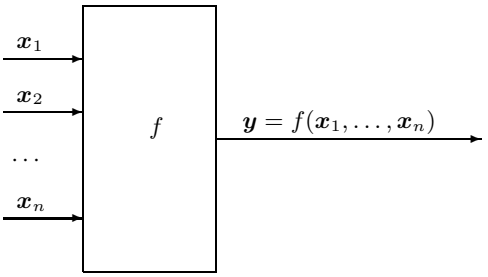

Fig. 2. Interval computations

The process of computing this interval range based on the input intervals $\boldsymbol{x}_i$ is called *interval computations*; see, e.g., Jaulin et al. [2] and Kearfott et al. [3].

5 Uniform Distributions: Traditional Engineering Approach to Interval Uncertainty

Brief description. In the case of interval uncertainty, we only know the intervals, we do not know the probability distributions on these intervals. A traditional statistical approach to the situation when several probability distributions are possible

is to select the "most uncertain" distribution, i.e., the distribution which has the largest possible value of the entropy $S \stackrel{\text{def}}{=} -\int \rho(x) \cdot \ln(\rho(x))\,\mathrm{d}x$ (here $\rho(x)$ denotes the probability density). For details on this Maximum Entropy approach and its relation to interval uncertainty (and Laplace's principle of indifference), see, e.g., Jaynes at al. [4].

One can easily check that for a single variable x_1, among all distributions located on a given interval, the entropy is the largest when this distribution is *uniform* on this interval. Indeed, a function $\rho(x) \geq 0$ is a probability density function on the given interval if $\int \rho(x)\,\mathrm{d}x = 1$. Thus, to find the probability density function that maximizes entropy, we must maximize entropy $-\int \rho(x) \cdot \ln(\rho(x))\,\mathrm{d}x$ under the constraint $\int \rho(x)\,\mathrm{d}x = 1$. According to the Lagrange multiplier method, for some value λ (Lagrange multiplier), the desired constraint optimization problem is equivalent to an unconstraint optimization problem of maximizing the expression $-\int \rho(x) \cdot \ln(\rho(x))\,\mathrm{d}x + \lambda \cdot (\int \rho(x)\,\mathrm{d}x - 1)$. Differentiating this expression with respect to each of the variables $\rho(x)$ and equating the derivative to 0, we conclude that $-\ln(\rho(x)) - 1 + \lambda = 0$, hence $\rho(x) = \exp(\lambda - 1)$. The probability density has the same value for all x from the given interval, hence we indeed have a uniform distribution.

In the case of several variables, we can similarly conclude that the distribution with the largest value of the entropy is the one which is uniformly distributed in the corresponding box $\boldsymbol{x}_1 \times \ldots \times \boldsymbol{x}_i \times \ldots \times \boldsymbol{x}_n$, i.e., a distribution in which:

- each variable Δx_i is uniformly distributed on the corresponding interval $[-\Delta_i, \Delta_i]$, and
- variables corresponding to different inputs are statistically independent.

This is indeed one of the main ways how interval uncertainty is treated in engineering practice: if we only know that the value of some variable is in the interval $[\underline{x}_i, \overline{x}_i]$, and we have no information about the probabilities, then we assume that the variable x_i is uniformly distributed on this interval.

Limitations. To explain the limitations of this engineering approach, let us consider the simplest possible algorithm $y = f(x_1, \ldots, x_i, \ldots, x_n) = x_1 + \ldots + x_i + \ldots + x_n$. For simplicity, let us assume that the measured values of all n quantities are 0s $\widetilde{x}_1 = \ldots = \widetilde{x}_i = \ldots = \widetilde{x}_n = 0$, and that all n measurements have the same error bound Δ_x; $\Delta_1 = \ldots = \Delta x_i = \ldots = \Delta_n = \Delta_x$.

In this case, $\Delta y = \Delta x_1 + \ldots + \Delta x_i + \ldots + \Delta x_n$. Each of n component measurement errors can take any value from $-\Delta_x$ to Δ_x, so the largest possible value of Δy is attained when all of the component errors attain the largest possible value $\Delta x_i = \Delta_x$. In this case, the largest possible value Δ of Δy is equal to $\Delta = n \cdot \Delta_x$.

Let us see what the maximum entropy approach will predict in this case. According to this approach, we assume that Δx_i are independent random variables, each of which is uniformly distributed on the interval $[-\Delta, \Delta]$. According to the Central Limit theorem (see, e.g., Sheskin [5]), when $n \to \infty$, the distribution of the sum of n independent identically distributed bounded random variables tends to Gaussian. This means that for large values n, the distribution of Δy is approximately normal.

A normal distribution is uniquely determined by its mean and variance. When we add several independent variables, their means and variances add up. For each uniform distribution Δx_i on the interval $[-\Delta_x, \Delta_x]$ of width $2\Delta_x$, the probability density is equal to $\rho(x) = \dfrac{1}{2\Delta_x}$, so the mean is 0 and the variance is

$$V = \int_{-\Delta_x}^{\Delta_x} x^2 \cdot \rho(x)\,\mathrm{d}x = \frac{1}{2\Delta_x} \cdot \int_{-\Delta_x}^{\Delta_x} x^2\,\mathrm{d}x = \frac{1}{2\Delta_x} \cdot \frac{1}{3} \cdot x^3\big|_{-\Delta_x}^{\Delta_x} = \frac{1}{3} \cdot \Delta_x^2. \quad (2)$$

Thus, for the sum Δy of n such variables, the mean is 0, and the variance is equal to $(n/3) \cdot \Delta_x^2$. Thus, the standard deviation is equal to $\sigma = \sqrt{V} = \Delta_x \cdot \frac{\sqrt{n}}{\sqrt{3}}$.

It is known that in a normal distribution, with probability close to 1, all the values are located within the $k \cdot \sigma$ vicinity of the mean: for $k = 3$, it is true with probability 99.9%, for $k = 6$, it is true with probability $1-10^{-6}$%, etc. So, practically with certainty, Δy is located within an interval $k \cdot \sigma$ which grows with n as $\sqrt{n}$.

For large n, we have $k \cdot \Delta_x \cdot \frac{\sqrt{n}}{\sqrt{3}} \ll \Delta_x \cdot n$, so we get a serious underestimation of the resulting measurement error. This example shows that estimates obtained by selecting a single distribution can be very misleading.

6 Techniques for Estimating the Uncertainty of the Results of Indirect Measurements in Situations when the Measurement Errors of Direct Measurements are Relatively Small

Linearization: main idea. When the measurement errors Δx_i are relatively small, we can use linearization.

By definition of the measurement error $\Delta x_i = \widetilde{x}_i - x_i$, hence $x_i = \widetilde{x}_i - \Delta x_i$. When the measurement errors Δx_i of direct measurements are relatively small, we can expand the expression

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

$$f(\widetilde{x}_1, \ldots, \widetilde{x}_i, \ldots, \widetilde{x}_n) - f(\widetilde{x}_1 - \Delta x_1, \ldots, \widetilde{x}_i - \Delta x_i, \ldots, \widetilde{x}_n - \Delta x_n) \quad (3)$$

in Taylor series and only keep linear terms in the resulting expansion. Since

$$y = f(\widetilde{x}_1 - \Delta x_1, \ldots, \widetilde{x}_i - \Delta x_i, \ldots, \widetilde{x}_n - \Delta x_n) \approx$$

$$f(\widetilde{x}_1, \ldots, \widetilde{x}_i, \ldots, \widetilde{x}_n) - \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \cdot \Delta x_i, \quad (4)$$

we conclude that $\Delta y = \widetilde{y} - y = \sum_{i=1}^{n} c_i \cdot \Delta x_i$, where $c_i = \frac{\partial f}{\partial x_i}$.

The dependence of Δy on Δx_i is linear: it is increasing relative to x_i if $c_i \geq 0$ and decreasing if $c_i < 0$. So, to find the largest possible value Δ of Δy, we must take:

- the largest possible value $\Delta x_i = \Delta_i$ when $c_i \geq 0$, and
- the smallest possible value $\Delta x_i = -\Delta_i$ when $c_i < 0$.

In both cases, the corresponding term in the sum has the form $|c_i| \cdot \Delta_i$, so we can conclude that

$$\Delta = \sum_{i=1}^{n} |c_i| \cdot \Delta_i. \quad (5)$$

Similarly, the smallest possible value of Δy is equal to $-\Delta$. Thus, the range of possible values of y is equal to $[\underline{y}, \overline{y}] = [\widetilde{y} - \Delta, \widetilde{y} + \Delta]$. So, to compute Δ, it is sufficient to know the partial derivatives c_i.

Case of analytical formulas. In the simplest case when the algorithm $f(x_1, \ldots, x_i, \ldots, x_n)$ consists of a simple analytical expression, we can find explicit analytical formulas for the partial derivatives and thus compute the desired bound Δ.

Techniques based on sensitivity analysis (automatic differentiation). In the general case, a natural way to compute partial derivatives comes directly from the definition. By definition, a partial derivative is defined as a limit

$$\frac{\partial f}{\partial x_i} = \lim_{h_i \to 0} \frac{f(\widetilde{x}_1, \ldots, \widetilde{x}_{i-1}, \widetilde{x}_i + h_i, \widetilde{x}_{i+1}, \ldots, \widetilde{x}_n) - f(\widetilde{x}_1, \ldots, \widetilde{x}_i, \ldots, \widetilde{x}_n)}{h_i}. \tag{6}$$

In turn, a limit, by its definition, means that when the values of h_i is small, the corresponding ratio is very close to the partial derivative. Thus, we can estimate the partial derivative as the ratio

$$c_i = \frac{\partial f}{\partial x_i} \approx \frac{f(\widetilde{x}_1, \ldots, \widetilde{x}_{i-1}, \widetilde{x}_i + h_i, \widetilde{x}_{i+1}, \ldots, \widetilde{x}_n) - f(\widetilde{x}_1, \ldots, \widetilde{x}_i, \ldots, \widetilde{x}_n)}{h_i} \tag{7}$$

for some small value h_i.

After we have computed n such ratios, we can then compute the desired bound Δ on $|\Delta y|$ as $\Delta = \sum_{i=1}^{n} |c_i| \cdot \Delta_i$.

In general, this procedure requires n divisions by h_i and n multiplications by Δ_i. The procedure can be made faster if we select $h_i = \Delta_i$. In this case, we get

$$\Delta = \sum_{i=1}^{n} |f(\widetilde{x}_1, \ldots, \widetilde{x}_{i-1}, \widetilde{x}_i + \Delta_i, \widetilde{x}_{i+1}, \ldots, \widetilde{x}_n) - \widetilde{y}|. \tag{8}$$

Advanced Monte-Carlo simulation techniques. The above algorithm requires that we call the data processing algorithm $n + 1$ times: first to compute the value $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_i, \ldots, \widetilde{x}_n)$, and then n more times to compute the values

$$f(\widetilde{x}_1, \ldots, \widetilde{x}_{i-1}, \widetilde{x}_i + \Delta_i, \widetilde{x}_{i+1}, \ldots, \widetilde{x}_n) \tag{9}$$

and thus, the corresponding partial derivatives.

In many practical situations, the data processing algorithms are time-consuming, and we process large amounts of data, with the number n of data points in thousands. In this case, the use of the above linearization algorithm would require thousands of times longer than data processing itself – which itself is already time consuming. Is it possible to estimate Δ faster?

The answer is "yes", it is possible to have a Monte-Carlo-type algorithm which estimates Δ by using only a constant number of calls to the data processing algorithm f; for details, see, e.g. Kreinovich et al. [6] and Kreinovich Ferson [7].

At first glance, since we know that the measurement errors are located within the intervals $[-\Delta_i, \Delta_i]$, it sounds reasonable to select distributions located on these intervals. However, it can be shown that this does not lead to the desired estimates.

It turns out that it is possible to estimate the interval uncertainty if we use a distribution d which is *not* located on the interval $[-\Delta_i, \Delta_i]$ – namely, a distribution related to the *basic Cauchy* distribution with the probability density function $\rho(x) = \dfrac{1}{\pi \cdot (x^2+1)}$. The resulting Cauchy deviate method works in the linearized case – when the function $f(x_1, \ldots, x_i, \ldots, x_n)$ is reasonably smooth and the box $[\underline{x}_1, \overline{x}_1] \times \ldots \times [\underline{x}_i, \overline{x}_i] \times \ldots \times [\underline{x}_n, \overline{x}_n]$ is small enough, so that on this box, we can reasonably approximate the function f by its linear terms.

If we multiply a random variable distributed according to the above basic Cauchy distribution d by a value Δ, then we get a Cauchy distribution with a parameter Δ, i.e., a distribution described by the following density function: $\rho(x) = \dfrac{\Delta}{\pi \cdot (x^2 + \Delta^2)}$. It is known that if $\xi_1, \ldots, \xi_i, \ldots, \xi_n$ are independent variables distributed according to Cauchy distributions with parameters Δ_i, then, for every n real numbers $c_1, \ldots, c_i, \ldots, c_n$, the corresponding linear combination $c_1 \cdot \xi_1 + \ldots + c_i \cdot \xi_i + \ldots + c_n \cdot \xi_n$ is also Cauchy distributed, with the parameter Δ equal to the desired value $\Delta = |c_1| \cdot \Delta_1 + \ldots + |c_i| \cdot \Delta_i + |c_n| \cdot \Delta_n$.

Thus, if for some number of iterations N, we simulate $\delta x_i^{(k)}$ $(1 \le k \le N)$ as Cauchy distributed with parameter Δ_i, then, in the linear approximation, the corresponding differences

$$\delta y^{(k)} \stackrel{\text{def}}{=} f(\widetilde{x}_1 + \delta x_1^{(k)}, \ldots, \widetilde{x}_i + \delta x_i^{(k)}, \ldots, \widetilde{x}_n + \delta x_n^{(k)}) - \widetilde{y} \tag{10}$$

are distributed according to the Cauchy distribution with the parameter Δ. The resulting values $\delta y^{(1)}, \ldots, \delta y^{(k)}, \ldots, \delta y^{(N)}$ are therefore a sample from the Cauchy distribution with the unknown parameter Δ. Based on this sample, we can estimate the value Δ.

In order to estimate Δ, we can apply the Maximum Likelihood Method which leads to the following equation:

$$\frac{1}{1 + \left(\dfrac{\delta y^{(1)}}{\Delta}\right)^2} + \ldots + \frac{1}{1 + \left(\dfrac{\delta y^{(k)}}{\Delta}\right)^2} + \ldots + \frac{1}{1 + \left(\dfrac{\delta y^{(N)}}{\Delta}\right)^2} = \frac{N}{2}. \tag{11}$$

The left-hand side of this equation is an increasing function that is equal to 0 (hence smaller than $N/2$) for $\Delta = 0$ and larger than $N/2$ for $\Delta = \max \left|\delta y^{(k)}\right|$; therefore the solution to this equation can be found by applying a bisection method to the interval $\left[0, \max \left|\delta y^{(k)}\right|\right]$.

Simulation of Cauchy distribution with parameter Δ_i can be based on the functional transformation of uniformly distributed sample values:

$$\delta x_i^{(k)} = \Delta_i \cdot \tan(\pi \cdot (r_i - 0.5)), \tag{12}$$

where r_i is uniformly distributed on the interval $[0, 1]$.

As a result, we arrive at the following algorithm (see, e.g., Kreinovich Ferson [7] and Trejo et al. [8]):

- Apply f to the midpoints: $\widetilde{y} := f(\widetilde{x}_1, \ldots, \widetilde{x}_i, \ldots, \widetilde{x}_n)$;
- For $k = 1, 2, \ldots, N$, repeat the following:
  - use the standard random number generator to compute n numbers $r_i^{(k)}$, $i = 1, 2, \ldots, n$, that are uniformly distributed on the interval $[0, 1]$;

- compute Cauchy distributed values $c_i^{(k)} := \tan(\pi \cdot (r_i^{(k)} - 0.5))$;
- compute the largest value of $|c_i^{(k)}|$ so that we will be able to normalize the simulated approximation errors and apply f to the values that are within the box of possible values: $K := \max_i |c_i^{(k)}|$;
- compute the simulated approximation errors $\delta x_i^{(k)} := \Delta_i \cdot c_i^{(k)}/K$;
- compute the simulated "actual values" $x_i^{(k)} := \widetilde{x}_i + \delta x_i^{(k)}$;
- apply the program f to the simulated measurement results and compute the simulated approximation error for y:

$$\Delta y^{(k)} := K \cdot \left(f\left(x_1^{(k)}, \ldots, x_i^{(k)}, \ldots, x_n^{(k)}\right) - \widetilde{y} \right); \tag{13}$$

- Compute Δ by applying the bisection method to solve the equation

$$\frac{1}{1 + \left(\dfrac{\Delta y^{(1)}}{\Delta}\right)^2} + \ldots + \frac{1}{1 + \left(\dfrac{\Delta y^{(k)}}{\Delta}\right)^2} + \ldots + \frac{1}{1 + \left(\dfrac{\Delta y^{(N)}}{\Delta}\right)^2} = \frac{N}{2}. \tag{14}$$

In Kreinovich Ferson [7] and Trejo et al. [8], we found the number of iterations N that would provide the desired (relative) accuracy ε in estimating Δ, i.e., the number of iterations that are needed to guarantee that

$$(1 - \varepsilon) \cdot \widetilde{\Delta} \le \Delta \le (1 + \varepsilon) \cdot \widetilde{\Delta} \tag{15}$$

with a given certainty p_0.

In practice, it is reasonable to get a certainty $p_0 = 95\%$ and accuracy $\varepsilon = 0.2$ (20%).

To get an accuracy ε with 95% certainty, we must pick $N = 8/\varepsilon^2$. In particular, to get a 20% accuracy ($0.2 \cdot \Delta$) with 95% certainty, i.e., to guarantee that

$$0.8 \cdot \widetilde{\Delta} \le \Delta \le 1.2 \cdot \widetilde{\Delta} \tag{16}$$

with certainty $\ge 95\%$, we need $N = 8/(0.2)^2 = 200$ runs.

In general, the required number of calls to a model depends only on the desired accuracy ε and not on n – so for large n, these methods are much faster.

Comment. It is important to mention that we assumed that the function f is reasonably linear within the box

$$[\widetilde{x}_1 - \Delta_1, \widetilde{x}_1 + \Delta_1] \times \ldots \times [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i] \times \ldots \times [\widetilde{x}_n - \Delta_n, \widetilde{x}_n + \Delta_n]. \tag{17}$$

However, the simulated values δ_i may be outside the box. When we get such values, we do not use the function f for them, we use a linearized function that is equal to f within the box, and that is extended linearly for all other values.

7 Techniques for Error Estimation in the General Case of Interval Uncertainty

Need for interval computations. In many application areas, it is sufficient to have an approximate estimate of y. However, in some applications, it is important to

guarantee that the (unknown) actual value y of a certain quantity does not exceed a certain threshold y_0. The only way to guarantee this is to have an interval $\boldsymbol{Y} = [\underline{Y}, \overline{Y}]$ which is guaranteed to contain y (i.e., for which $\boldsymbol{y} \subseteq \boldsymbol{Y}$) and for which $\overline{Y} \leq y_0$.

For example, in nuclear engineering, we must make sure that the temperatures and the neutron flows do not exceed the critical values; when planning a space flight, we want to guarantee that the space ship lands on the planet and does not fly past it, etc.

The interval $\boldsymbol{Y}$ which is guaranteed to contain the actual range $\boldsymbol{y}$ is usually called an *enclosure* for this range. So, in such situations, we need to compute either the original range or at least an enclosure for this range. Computing such an enclosure is also one of the main tasks of interval computations.

Traditional numerical methods are often not sufficient. The main limitations of the traditional numerical mathematics approach to error estimation was that often, no clear distinction was made between approximate (non-guaranteed) and guaranteed (= interval) error bounds.

For example, for iterative methods, many papers on numerical mathematics consider the rate of convergence as an appropriate measure of approximation error. Clearly, if we know that the error decreases as $O(1/n)$ or as $O(a^{-n})$, we gain some information about the corresponding algorithms – and we also gain a knowledge that for large n, the second method is more accurate. However, in real life, we make a fixed number n of iterations. If the only information we have about the approximation error is the above asymptotics, then we still have no idea how close the result of n-th iteration is to the actual (desired) value.

It is therefore important to emphasize the need for guaranteed methods, and to develop techniques for producing *guaranteed* estimates. Such guaranteed estimates is what interval computations are about.

Interval computations: a brief history. The notion of interval computations is reasonably recent, it dates back to the 1950s, but the main problem is known since Archimedes who used guaranteed two-sided bounds to compute π; see, e.g., Achimedes [9].

Since then, many useful guaranteed bounds have been developed for different numerical methods. There have also been several general descriptions of such bounds, often formulated in terms similar to what we described above. For example, in the early 20th century, the concept of a function having values which are bounded within limits was discussed by W. H. Young in [10]. The concept of operations with a set of multi-valued numbers was introduced by R. C. Young, who developed a formal algebra of multi-valued numbers [11]. The special case of closed intervals was further developed by P. S. Dwyer in [12].

Interval computations in their current form were independently invented by three researchers in three different parts of the world: by M. Warmus in Poland [13], by T. Sunaga in Japan [14], and by R. Moore in the USA [15].

The active interest in interval computations started with Moore's 1966 monograph [16]. This interest was enhanced by the fact that in addition to estimates for general numerical algorithms, Moore's monograph also described *practical* applications which have already been developed in his earlier papers and technical reports: in particular, interval computations were used to make sure that even when we take all the uncertainties into account, the trajectory of a space flight is guaranteed to reach the Moon.

Since then, interval computations have been actively used in many areas of science and engineering; see, e.g., interval website [17] and Jaulin et al. [2].

Comment. Early papers on interval computations can be found on the interval computations website [17].

First step: interval arithmetic. Our goal is to find the range of a given function $f(x_1, \ldots, x_i, \ldots, x_n)$ on the given intervals $\boldsymbol{x}_1 = [\underline{x}_1, \overline{x}_1], \ldots, \boldsymbol{x}_i = [\underline{x}_i, \overline{x}_i], \ldots, \boldsymbol{x}_n = [\underline{x}_n, \overline{x}_n]$.

This function $f(x_1, \ldots, x_i, \ldots, x_n)$ is given as an algorithm. In particular, we may have an explicit analytical expression for f, in which case this algorithm consists of simply computing this expression.

When we talk about algorithms, we usually mean an algorithm (program) written in a high-level programming language like Java or C. Such programming languages allows us to use arithmetic expressions and many other complex constructions. Most of these constructions, however, are not directly hardware supported inside a computer. Usually, only simple arithmetic operations are implemented: addition, subtraction, multiplication, and $1/x$ (plus branching). Even division a/b is usually not directly supported, it is performed as a sequence of two elementary arithmetic operations:

- first, we compute $1/b$;
- then, we multiply a by $1/b$.

When we input a general program into a computer, the computer *parses* it, i.e., represents it a sequence of elementary arithmetic operations.

Since a computer performs this parsing anyway, we can safely assume that the original algorithm $f(x_1, \ldots, x_i, \ldots, x_n)$ is already represented as a sequence of such elementary arithmetic operations.

Let us start our analysis of the interval computation techniques with the simplest possible case when the algorithm $f(x_1, \ldots, x_i, \ldots, x_n)$ simply consists of a single arithmetic operation: addition, subtraction, multiplication, or computing $1/x$.

Let us start by estimating the range of the addition function $f(x_1, x_2) = x_1 + x_2$ on the intervals $[\underline{x}_1, \overline{x}_1]$ and $[\underline{x}_2, \overline{x}_2]$. This function is increasing with respect to both its variables. We already know how to compute the range $[\underline{y}, \overline{y}]$ of a monotonic function. So, the range of addition is equal to $[\underline{x}_1 + \underline{x}_2, \overline{x}_1 + \overline{x}_2]$.

The desired range is usually denoted as $f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_i, \ldots, \boldsymbol{x}_n)$; in particular, for addition, this notation takes the form $\boldsymbol{x}_1 + \boldsymbol{x}_2$. Thus, we can define "addition" of two intervals as follows:

$$[\underline{x}_1, \overline{x}_1] + [\underline{x}_2, \overline{x}_2] = [\underline{x}_1 + \underline{x}_2, \overline{x}_2 + \overline{x}_2]. \tag{18}$$

This formula makes perfect intuitive sense: if one town has between 700 and 800 thousand people, and it merges with a nearby town whose population is between 100 and 200 thousand, then:

- the smallest possible value of the total population of the new big town is when both populations are the smallest possible, $700 + 100 = 800$, and
- the largest possible value is when both populations are the largest possible, i.e., $800 + 200 = 1000$.

The subtraction function $f(x_1, x_2) = x_1 - x_2$ is increasing with respect to x_1 and decreasing with respect to x_2, so we have

$$[\underline{x}_1, \overline{x}_1] - [\underline{x}_2, \overline{x}_2] = [\underline{x}_1 - \overline{x}_2, \overline{x}_1 - \underline{x}_2]. \tag{19}$$

These operations are also in full agreement with common sense. For example, if a warehouse originally had between 6.0 and 8.0 tons, and we moved between 1.0 and 2.0 tons to another location, then the smallest amount left is when we start with the smallest possible value 6.0 and move the largest possible value 2.0, resulting in $6.0 - 2.0 = 4.0$. The largest amount left is when we start with the largest possible value 8.0 and move the smallest possible value 1.0, resulting in $8.0 - 1.0 = 7.0$.

For multiplication $f(x_1, x_2) = x_1 \cdot x_2$, the direction of monotonicity depends on the actual values of x_1 and x_2: e.g., when $x_2 > 0$, the product increases with x_1, otherwise it decreases with x_1. So, unless we know the signs of the product beforehand, we cannot tell whether the maximum is attained at $x_1 = \underline{x}_1$ or at $x_1 = \overline{x}_1$. However, we know that it is always attained at one of these endpoints. So, to find the range of the product, it is sufficient to try all $2 \cdot 2 = 4$ combinations of these endpoints:

$$[\underline{x}_1, \overline{x}_1] \cdot [\underline{x}_2, \overline{x}_2] =$$
$$[\min(\underline{x}_1 \cdot \underline{x}_2, \underline{x}_1 \cdot \overline{x}_2, \overline{x}_1 \cdot \underline{x}_2, \overline{x}_1 \cdot \overline{x}_2), \max(\underline{x}_1 \cdot \underline{x}_2, \underline{x}_1 \cdot \overline{x}_2, \overline{x}_1 \cdot \underline{x}_2, \overline{x}_1 \cdot \overline{x}_2)]. \tag{20}$$

Finally, the function $f(x_1) = 1/x_1$ is decreasing wherever it is defined (when $x_1 \neq 0$), so if $0 \notin [\underline{x}_1, \overline{x}_1]$, then

$$\frac{1}{[\underline{x}_1, \overline{x}_1]} = \left[\frac{1}{\overline{x}_1}, \frac{1}{\underline{x}_1}\right]. \tag{21}$$

The formulas for addition, subtraction, multiplication, and reciprocal of intervals are called formulas of *interval arithmetic*.

Comment. Alternative faster-to-compute formulas which lead to slightly wider intervals are described, e.g., in Cerimele et al. [18].

Straightforward ("naive") interval computations. Historically the first method for computing the enclosure for the general case is the method which is sometimes called "straightforward" interval computations. In this method, 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 $\boldsymbol{Y} \supseteq \boldsymbol{y}$ for the desired range.

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

Example. Let us illustrate the above idea on the example of estimating the range of the function $f(x_1) = x_1 - x_1^2$ on the interval $x_1 \in [0, 0.8]$.

We start with parsing the expression for the function, i.e., describing how a computer will compute this expression; it will implement the following sequence of elementary operations:

$$r_1 := x_1 \cdot x_1; \quad r_2 := x_1 - r_1. \tag{22}$$

According to straightforward interval computations, we perform the same operations, but with *intervals* instead of *numbers*:

$$\boldsymbol{r}_1 := [0, 0.8] \cdot [0, 0.8] = [0, 0.64]; \quad \boldsymbol{r}_2 := [0, 0.8] - [0, 0.64] = [-0.64, 0.8]. \tag{23}$$

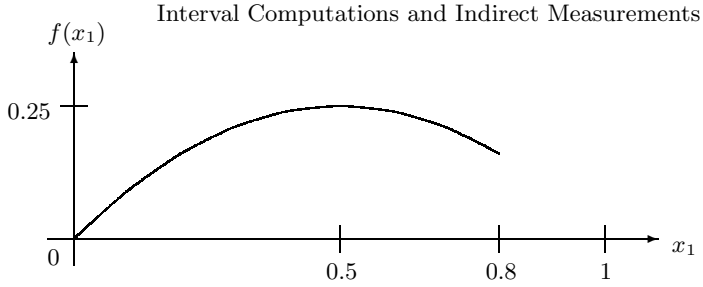

Fig. 3. Range of the function $f(x_1) = x_1 - x_1^2$ on the interval $[0, 0.8]$

For this function, the actual range is $f(\boldsymbol{x}_1) = [0, 0.25]$; see Fig. 3.

Interval computations go beyond straightforward technique. People who are vaguely familiar with interval computations sometimes erroneously assume that the above straightforward ("naive") techniques is all there is in interval computations. In conference presentations (and even in published papers), one often encounters a statement: "I tried interval computations, and it did not work". What this statement usually means is that they tried the above straightforward approach and – not surprisingly – it did not work well.

In reality, interval computations is *not a single algorithm*, it is a *problem* for which many different techniques exist. Let us now describe some of such techniques.

Comment. For each of the known techniques, there are cases when we get an excess width. The reason is that the problem of computing the exact range is NP-hard even for polynomial functions $f(x_1, \ldots, x_i, \ldots, x_n)$ – actually, even for quadratic functions f (see, e.g., Kreinovich et al. [19]).

Centered form. One of such techniques is the centered form; see, e.g., Jaulin et al. [2]. This technique is based on the same Taylor series expansion ideas as linearization. We start by representing each interval $\boldsymbol{x}_i = [\underline{x}_i, \overline{x}_i]$ in the form $[\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$, where $\widetilde{x}_i = (\underline{x}_i + \overline{x}_i)/2$ is the midpoint of the interval $\boldsymbol{x}_i$ and $\Delta_i = (\overline{x}_i - \underline{x}_i)/2$ is the half-width of this interval.

After that, we use the Taylor expansion. In linearization, we simply ignored quadratic and higher order terms. Here, instead, we use the Taylor form with a remainder term. Specifically, the centered form is based on the formula

$$f(x_1, \ldots, x_i, \ldots, x_n) = f(\widetilde{x}_1, \ldots, \widetilde{x}_i, \ldots, \widetilde{x}_n) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\eta_1, \ldots, \eta_i, \ldots, \eta_n) \cdot (x_i - \widetilde{x}_i), \tag{24}$$

where each η_i is some value from the interval $\boldsymbol{x}_i$.

Since $\eta_i \in \boldsymbol{x}_i$, the value of the i-th derivative belongs to the interval range of this derivative on these intervals. We also know that $x_i - \widetilde{x}_i \in [-\Delta_i, \Delta_i]$. Thus, we can conclude that

$$f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_i, \ldots, \boldsymbol{x}_n) \subseteq f(\widetilde{x}_1, \ldots, \widetilde{x}_i, \ldots, \widetilde{x}_n) +$$

$$\sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_i, \ldots, \boldsymbol{x}_n) \cdot [-\Delta_i, \Delta_i]. \tag{25}$$

To compute the ranges of the partial derivatives, we can use straightforward interval computations.

Example. Let us illustrate this method on the above example of estimating the range of the function $f(x_1) = x_1 - x_1^2$ over the interval $[0, 0.8]$. For this interval, the midpoint is $\widetilde{x}_1 = 0.4$; at this midpoint, $f(\widetilde{x}_1) = 0.24$. The half-width is $\Delta_1 = 0.4$. The only partial derivative here is $\dfrac{\partial f}{\partial x_1} = 1 - 2x_1$, its range on $[0, 0.8]$ is equal to $1 - 2 \cdot [0, 0.8] = [-0.6, 1]$. Thus, we get the following enclosure for the desired range $\boldsymbol{y}$:

$$\boldsymbol{y} \subseteq \boldsymbol{Y} = 0.24 + [-0.6, 1] \cdot [-0.4, 0.4] = 0.24 + [-0.4, 0.4] = [-0.16, 0.64]. \tag{26}$$

This enclosure is narrower than the "naive" estimate $[-0.64, 0.8]$, but it still contains excess width.

How can we get better estimates? In the centered form, we, in effect, ignored quadratic and higher order terms, i.e., terms of the type $\dfrac{\partial^2 f}{\partial x_i \partial x_j} \cdot \Delta x_i \cdot \Delta x_j$. When the estimate is not accurate enough, it means that this ignored term is too large. There are two ways to reduce the size of the ignored term:

- we can try to decrease this quadratic term, or
- we can try to explicitly include higher order terms in the Taylor expansion formula, so that the remainder term will be proportional to say Δx_i^3 and thus, be much smaller.

Let us describe these two ideas in detail.

First idea: bisection. Let us first describe the situation in which we try to minimize the second-order remainder term. In the above expression for this term, we cannot change the second derivative. The only thing we can decrease is the difference $\Delta x_i = x_i - \widetilde{x}_i$ between the actual value and the midpoint. This value is bounded by the half-width Δ_i of the box. So, to decrease this value, we can subdivide the original box into several narrower subboxes. Usually, we divide into two subboxes, so this subdivision is called *bisection*.

The range over the whole box is equal to the union of the ranges over all the subboxes. The widths of each subbox are smaller, so we get smaller Δx_i and hopefully, more accurate estimates for ranges over each of this subbox. Then, we take the union of the ranges over subboxes.

Example. Let us illustrate this idea on the above $x_1 - x_1^2$ example. In this example, we divide the original interval $[0, 0.8]$ into two subintervals $[0, 0.4]$ and $[0.4, 0.8]$. For both intervals, $\Delta_1 = 0.2$.

In the first subinterval, the midpoint is $\widetilde{x}_1 = 0.2$, so $f(\widetilde{x}_1) = 0.2 - 0.04 = 0.16$. The range of the derivative is equal to $1 - 2 \cdot [0, 0.4] = 1 - [0, 0.8] = [0.2, 1]$, hence we get an enclosure $0.16 + [0.2, 1] \cdot [-0.2, 0.2] = [-0.04, 0.36]$.

For the second interval, $\widetilde{x}_1 = 0.6$, $f(0.6) = 0.24$, the range of the derivative is $1 - 2 \cdot [0.4, 0.8] = [-0.6, 0.2]$, hence we get an enclosure

$$0.24 + [-0.6, 0.2] \cdot [-0.2, 0.2] = [0.12, 0.36]. \tag{27}$$

The union of these two enclosures is the interval $[-0.04, 0.36]$. This enclosure is much more accurate than before.

Further bisection leads to even more accurate estimates – the smaller the subintervals, the more accurate the enclosure.

Bisection: general comment. The more subboxes we consider, the smaller Δx_i and thus, the more accurate the corresponding enclosures. However, once we have more boxes, we need to spend more time processing these boxes. Thus, we have a trade-off between computation time and accuracy: the more computation time we allow, the more accurate estimates we will be able to compute.

Additional idea: monotonicity checking. If the function $f(x_1, \ldots, x_i, \ldots, x_n)$ is monotonic over the original box $\boldsymbol{x}_1 \times \ldots \times \boldsymbol{x}_i \ldots \times \boldsymbol{x}_n$, then we can easily compute its exact range. Since we used the centered form for the original box, this probably means that on that box, the function is not monotonic: for example, with respect to x_1, it may be increasing at some points in this box, and decreasing at other points.

However, as we divide the original box into smaller subboxes, it is quite possible that at least some of these subboxes will be outside the areas where the derivatives are 0 and thus, the function $f(x_1, \ldots, x_i, \ldots, x_n)$ will be monotonic. So, after we subdivide the box into subboxes, we should first check monotonicity on each of these subboxes – and if the function is monotonic, we can easily compute its range.

In calculus terms, a function is increasing with respect to x_i if its partial derivative $d_i \stackrel{\text{def}}{=} \dfrac{\partial f}{\partial x_i}$ is non-negative everywhere on this subbox. Thus, to check monotonicity, we should find the range $[\underline{d}_i, \overline{d}_i]$ of this derivative (we need to do it anyway to compute the centered form expression):

- if $\underline{d}_i \geq 0$, this means that the derivative is everywhere non-negative and thus, the function f is increasing in x_i;
- if $\overline{d}_i \leq 0$, this means that the derivative is everywhere non-positive and thus, the function f is decreasing in x_i.

If $\underline{d}_i < 0 < \overline{d}_i$, then we have to use the centered form.

If the function is monotonic (e.g., increasing) only with respect to some of the variables x_i, then

- to compute $\overline{y}$, it is sufficient to consider only the value $x_i = \overline{x}_i$, and
- to compute $\underline{y}$, it is sufficient to consider only the value $x_i = \underline{x}_i$.

For such subboxes, we reduce the original problem to two problems with fewer variables, problems which are thus easier to solve.

Example. For the example $f(x_1) = x_1 - x_1^2$, the partial derivative is equal to $1 - 2 \cdot x_1$.

On the first subbox $[0, 0.4]$, the range of this derivative is $1 - 2 \cdot [0, 0.4] = [0.2, 1]$. Thus, the derivative is always non-negative, the function is increasing on this subbox, and its range on this subbox is equal to $[f(0), f(0.4)] = [0, 0.16]$.

On the second subbox $[0.4, 0.8]$, the range of the derivative is $1 - 2 \cdot [0.4, 0.8] = [-0.6, 0.2]$. Here, we do not have guaranteed monotonicity, so we can use the centered form to get the enclosure $[0.12, 0.36]$ for the range.

The union of these two enclosures is the interval $[0, 0.36]$, which is slightly more accurate than before. Further bisection leads to even more accurate estimates.

Comment. We got the exact range because of the simplicity of our example, in which the extreme point 0.5 of the function $f(x_1) = x_1 - x_1^2$ is exactly in the middle of the interval $[0, 1]$. Thus, when we divided the box in two, both subboxes have the monotonicity property. In the general case, the extremal point will be inside one of the subboxes, so we will have excess width.

General Taylor techniques. As we have mentioned, another way to get more accurate estimates is to use so-called *Taylor techniques*, i.e., to explicitly consider second-order and higher-order terms in the Taylor expansion; see, e.g., Berz et al. [20], Neumaier [21], and references therein.

Let us illustrate the main ideas of Taylor analysis on the case when we allow second order terms. In this case, the formula with a remainder takes the form

$$f(x_1, \ldots, x_i, \ldots, x_n) = f(\widetilde{x}_1, \ldots, \widetilde{x}_i, \ldots, \widetilde{x}_n) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\widetilde{x}_1, \ldots, \widetilde{x}_n) \cdot (x_i - \widetilde{x}_i) +$$

$$\frac{1}{2} \cdot \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{\partial^2 f}{\partial x_i \partial x_j}(\eta_1, \ldots, \eta_n) \cdot (x_i - \widetilde{x}_i) \cdot (x_j - \widetilde{x}_j). \tag{28}$$

Thus, we get the enclosure

$$f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_i, \ldots, \boldsymbol{x}_n) \subseteq \boldsymbol{Y} \stackrel{\text{def}}{=} f(\widetilde{x}_1, \ldots, \widetilde{x}_i, \ldots, \widetilde{x}_n) +$$

$$\sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\widetilde{x}_1, \ldots, \widetilde{x}_i, \ldots, \widetilde{x}_n) \cdot [-\Delta_i, \Delta_i] + \tag{29}$$

$$\frac{1}{2} \cdot \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \cdot [-\Delta_i, \Delta_i] \cdot [-\Delta_j, \Delta_j].$$

Example. Let us illustrate this idea on the above example of $f(x_1) = x_1 - x_1^2$. Here, $\Delta_1 = 0.4$, $\widetilde{x}_1 = 0.4$, so $f(\widetilde{x}_1) = 0.24$ and $\dfrac{\partial f}{\partial x_1}(\widetilde{x}_1) = 1 - 2 \cdot 0.4 = 0.2$. The second derivative is equal to -2, so the Taylor estimate takes the form

$$\boldsymbol{Y} = 0.24 + 0.2 \cdot [-0.4, 0.4] - [-0.4, 0.4]^2.$$

Strictly speaking, if we interpret Δx_1^2 as $\Delta x_1 \cdot \Delta x_1$ and use the formulas of interval multiplication, we get the interval

$$[-0.4, 0.4]^2 = [-0.4, 0.4] \cdot [-0.4, 0.4] = [-0.16, 0.16]$$

and thus, the enclosure

$$\boldsymbol{Y} = 0.24 + [-0.08, 0.08] - [-0.16, 0.16] = [0.16, 0.32] - [-0.16, 0.16] = [0, 0.48]$$

for the desired range. However, we can view x^2 as a special function, for which the range over $[-0.4, 0.4]$ is known to be $[0, 0.16]$. In this case, the above enclosure takes the form

$$\boldsymbol{Y} = 0.24 + [-0.08, 0.08] - [0, 0.16] = [0.16, 0.32] - [0, 0.16] = [0, 0.32]$$

which is much closer to the actual range $[0, 0.25]$.

Taylor methods: general comment. The more terms we consider in the Taylor expansion, the smaller the remainder term and thus, the more accurate the corresponding enclosures. However, once we have more terms, we need to spend more time computing these terms. Thus, for Taylor methods, we also have a trade-off between computation time and accuracy: the more computation time we allow, the more accurate estimates we will be able to compute.

An alternative version of affine and Taylor arithmetic. The main idea of Taylor methods is to approximate the given function $f(x_1, \ldots, x_i, \ldots, x_n)$ by a polynomial of a small order plus an interval remainder term.

In these terms, straightforward interval computations can be viewed as 0-th order Taylor methods in which all we have is the corresponding interval (or, equivalently, the constant term plus the remainder interval). To compute this interval, we repeated the computation of f step by step, replacing operations with numbers by operations with intervals.

We can do the same for higher-order Taylor expansions as well. Let us illustrate how this can be done for the first order Taylor terms. We start with the expressions $x_i = \widetilde{x}_i - \Delta x_i$. Then, at each step, we keep a term of the type $a = \widetilde{a} + \sum\limits_{i=1}^{n} a_i \cdot \Delta x_i + \boldsymbol{a}$. (To be more precise, the keep the coefficients $\widetilde{a}$ and a_i and the interval $\boldsymbol{a}$.)

Addition and subtraction of such terms are straightforward:

$$(\widetilde{a} + \sum_{i=1}^{n} a_i \cdot \Delta x_i + \boldsymbol{a}) + (\widetilde{b} + \sum_{i=1}^{n} b_i \cdot \Delta x_i + \boldsymbol{b}) = (\widetilde{a} + \widetilde{b}) + \sum_{i=1}^{n} (a_i + b_i) \cdot \Delta x_i + (\boldsymbol{a} + \boldsymbol{b}); \quad (30)$$

$$(\widetilde{a} + \sum_{i=1}^{n} a_i \cdot \Delta x_i + \boldsymbol{a}) - (\widetilde{b} + \sum_{i=1}^{n} b_i \cdot \Delta x_i + \boldsymbol{b}) = (\widetilde{a} - \widetilde{b}) + \sum_{i=1}^{n} (a_i - b_i) \cdot \Delta x_i + (\boldsymbol{a} - \boldsymbol{b}). \quad (31)$$

For multiplication, we add terms proportional to $\Delta x_i \cdot \Delta x_j$ to the interval part:

$$(\widetilde{a} + \sum_{i=1}^{n} a_i \cdot \Delta x_i + \boldsymbol{a}) \cdot (\widetilde{b} + \sum_{i=1}^{n} b_i \cdot \Delta x_i + \boldsymbol{b}) = (\widetilde{a} \cdot \widetilde{b}) + \sum_{i=1}^{n} (\widetilde{a} \cdot b_i + \widetilde{b} \cdot a_i) \cdot \Delta x_i +$$

$$(\widetilde{a} \cdot \boldsymbol{b} + \widetilde{b} \cdot \boldsymbol{a} + \sum_{i=1}^{n} a_i \cdot b_i \cdot [0, \Delta_i^2] + \sum_{i=1}^{n} \sum_{j \neq i} a_i \cdot b_j \cdot [-\Delta_i, \Delta_i] \cdot [\Delta_j \cdot \Delta_j]). \quad (32)$$

At the end, we get an expression of the above type for the desired quantity y: $y = \widetilde{y} + \sum\limits_{i=1}^{n} y_i \cdot \Delta x_i + \boldsymbol{y}$. We already know how to compute the range of a linear function, so we get the following enclosure for the final range: $\boldsymbol{Y} = \widetilde{y} + [-\Delta, \Delta] + \boldsymbol{y}$, where $\Delta = \sum\limits_{i=1}^{n} |y_i| \cdot \Delta_i$.

Example. For $f(x_1) = x_1 - x_1^2$, we first compute $x_2 = x_1^2$ and then $y = x_1 - x_2$. We start with the interval $\boldsymbol{x}_1 = \widetilde{x}_1 - \Delta x_1 = 0.4 + (-1) \cdot \Delta_1 + [0, 0]$.

On the next step, we compute the square of this expression. This square is equal to $0.16 + (-0.8) \cdot \Delta x_1 + \Delta x_1^2$. Since $\Delta x_1 \in [-0.4, 0.4]$, we conclude that $\Delta x_1^2 \in [0, 0.16]$ and thus, that $x_2 = 0.16 + (-0.8) \cdot \Delta x_1 + [0, 0.16]$.

For $y = x_1 - x_2$, we now have

$$y = (0.4 - 0.16) + ((-1) - (-0.8)) \cdot \Delta x_1 + ([0,0] - [0,0.16]) =$$

$$0.24 + (-0.2) \cdot \Delta x_1 + [-0.16, 0]. \tag{33}$$

Since $\Delta x_1 \in [-0.4, 0.4]$, we get the enclosure

$$\boldsymbol{Y} = 0.24 + (-0.2) \cdot [-0.4, 0.4] + [-0.16, 0] = [0, 0.32]. \tag{34}$$

Comment. We have described several methods and several ideas. On our simple example, some ideas work better, some lead to wider enclosures. The fact that a method works better on the simple example does not mean that it always works better, it depends on the function. In large-scale practical examples, it is useful to *combine* all these methods and ideas – e.g., bisect and use centered form and monotonicity on subboxes; see, e.g., Jaulin et al. [2].

The interval method – one of the above or their combination – has to be carefully chosen to match the function at hand. There exist several semi-empirical heuristics on which method to choose; see, e.g., Jaulin et al. [2].

8 Situations when, in Addition to the Upper Bounds on the Measurement Error, we Also Have Partial Information About the Probabilities of Different Error Values

Practical problem. In interval computations, we assume that the uncertainty in x_i can be described by the interval of possible values. In real life, in addition to the intervals, we often have some information about the probabilities of different values within this interval. What can we then do?

Which is the best way to describe the corresponding probabilistic uncertainty? One of the main objectives of data processing is to make decisions. A standard way of making a decision is to select the action a for which the expected utility (gain) is the largest possible. This is where probabilities are used: in computing, for every possible action a, the corresponding expected utility. To be more precise, we usually know, for each action a and for each actual value of the (unknown) quantity x, the corresponding value of the utility $u_a(x)$. We must use the probability distribution for x to compute the expected value $\mathrm{e}[u_a(x)]$ of this utility.

In view of this application, the most useful characteristics of a probability distribution would be the ones which would enable us to compute the expected value $\mathrm{e}[u_a(x)]$ of different functions $u_a(x)$.

Which representations are the most useful for this intended usage? General idea. Which characteristics of a probability distribution are the most useful for computing mathematical expectations of different functions $u_a(x)$? The answer to this question depends on the type of the function, i.e., on how the utility value u depends on the value x of the analyzed parameter.

Smooth utility functions naturally lead to moments. One natural case is when the utility function $u_a(x)$ is smooth. We have already mentioned, in Section I, that we usually know a (reasonably narrow) interval of possible values of x. So, to compute the expected value of $u_a(x)$, all we need to know is how the function $u_a(x)$ behaves on this narrow interval. Because the function is smooth, we can expand it into Taylor series. Because the interval is narrow, we can safely consider only linear and quadratic terms in this expansion and ignore higher-order terms: $u_a(x) \approx c_0 + c_1 \cdot (x - x_0) + c_2 \cdot (x - x_0)^2$, where x_0 is a point inside the interval. Thus, we can approximate the expectation of this function by the expectation of the corresponding quadratic expression: $\mathrm{e}[u_a(x)] \approx \mathrm{e}[c_0 + c_1 \cdot (x - x_0) + c_2 \cdot (x - x_0)^2]$, i.e., by the following expression: $\mathrm{e}[u_a(x)] \approx c_0 + c_1 \cdot \mathrm{e}[x - x_0] + c_2 \cdot \mathrm{e}[(x - x_0)^2]$. So, to compute the expectations of such utility functions, it is sufficient to know the first and second moments of the probability distribution.

In particular, if we use, as the point x_0, the average $\mathrm{e}[x]$, the second moment turns into the variance of the original probability distribution. So, instead of the first and the second moments, we can use the mean E and the variance V.

In decision making, non-smooth utility functions are common. In decision making, not all dependencies are smooth. There is often a threshold x_0 after which, say, a concentration of a certain chemical becomes dangerous.

This threshold sometimes comes from the detailed chemical and/or physical analysis. In this case, when we increase the value of this parameter, we see the drastic increase in effect and hence, the drastic change in utility value. Sometimes, this threshold simply comes from regulations. In this case, when we increase the value of this parameter past the threshold, there is no drastic increase in effects, but there is a drastic decrease of utility due to the necessity to pay fines, change technology, etc. In both cases, we have a utility function which experiences an abrupt decrease at a certain threshold value x_0.

Non-smooth utility functions naturally lead to cdfs. We want to be able to compute the expected value $\mathrm{e}[u_a(x)]$ of a function $u_a(x)$ which changes smoothly until a certain value x_0, then drops it value and continues smoothly for $x > x_0$. We usually know the (reasonably narrow) interval which contains all possible values of x. Because the interval is narrow and the dependence before and after the threshold is smooth, the resulting change in $u_a(x)$ before x_0 and after x_0 is much smaller than the change at x_0. Thus, with a reasonable accuracy, we can ignore the small changes before and after x_0, and assume that the function $u_a(x)$ is equal to a constant u^+ for $x < x_0$, and to some other constant $u^- < u^+$ for $x > x_0$.

The simplest case is when $u^+ = 1$ and $u^- = 0$. In this case, the desired expected value $\mathrm{e}[u_a^{(0)}(x)]$ coincides with the probability that $x < x_0$, i.e., with the corresponding value $F(x_0)$ of the cumulative distribution function (cdf). A generic function $u_a(x)$ of this type, with arbitrary values u^- and u^+, can be easily reduced to this simplest case, because, as one can easily check, $u_a(x) = u^- + (u^+ - u^-) \cdot u^{(0)}(x)$ and hence, $\mathrm{e}[u_a(x)] = u^- + (u^+ - u^-) \cdot F(x_0)$.

Thus, to be able to easily compute the expected values of all possible non-smooth utility functions, it is sufficient to know the values of the cdf $F(x_0)$ for all possible x_0.

How to represent partial information about probabilities: general idea. In many cases, we have a complete information about the probability distributions that describe the uncertainty of each of n inputs.

However, a practically interesting case is how to deal with situations when we only have partial information about the probability distributions. How can we represent this partial information?

Case of cdf. If we use cdf $F(x)$ to represent a distribution, then full information corresponds to the case when we know the exact value of $F(x)$ for every x. Partial information means:

- either that we only know approximate values of $F(x)$ for all x, i.e., that for every x, we only know the interval that contains $F(x)$; in this case, we get a *p-box* (Ferson [22]);
- or that we only know the values of $F(x)$ for some x, i.e, that we only know the values $F(x_1), \ldots, F(x_i), \ldots, F(x_n)$ for finitely many values $x = x_1, \ldots, x_i, \ldots, x_n$; in this case, we have a *histogram*.

It is also possible that we know only approximate values of $F(x)$ for some x; in this case, we have an *interval-valued histogram*.

Case of moments. If we use moments to represent a distribution, then partial information means that we either know the exact values of finitely many moments, or that we know intervals of possible values of several moments.

Resulting algorithms. This discussion leads to a natural classification of possible algorithms:

- If we have complete information about the distributions of x_i, then, to get validated estimates on uncertainty of y, we have to use Monte-Carlo-type techniques; see, e.g., Lodwick et al. [23].
- If we have p-boxes, we can use methods from Ferson [22] and Ferson et al. [24].
- If we have histograms, we can use methods from Berleant et al. [25].
- If we have moments, then we can use methods from Granvilliers et al. [26] and Kreinovich [27].

Case study: first moments. In some practical situations, in addition to the lower and upper bounds on each random variable x_i, we know the bounds $\boldsymbol{E}_i = [\underline{E}_i, \overline{E}_i]$ on its mean E_i. Indeed, in measurement practice (see, e.g., Rabinovich [1]), 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 $\mathrm{e}[\Delta x]$, 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 $\bar{x} = \dfrac{\widetilde{x}^{(1)} + \ldots + \widetilde{x}^{(k)} + \ldots + \widetilde{x}^{(N)}}{N}$ of the corresponding measurement results $\widetilde{x}^{(1)} = x + \Delta x^{(1)}, \ldots, \widetilde{x}^{(k)} = x + \Delta x^{(k)}, \ldots, \widetilde{x}^{(N)} = x + \Delta x^{(N)}$.

- If systematic error is the only error component, then all the measurements lead to exactly the same value $\widetilde{x}^{(1)} = \ldots = \widetilde{x}^{(k)} = \ldots = \widetilde{x}^{(N)}$, 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., $\mathrm{e}[\Delta x] = 0$ and $\mathrm{e}[\widetilde{x}] = x$, 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 $\boldsymbol{x} = [\widetilde{x} - \Delta, \widetilde{x} + \Delta]$, but we can also conclude that the expected value of $x = \widetilde{x} - \Delta x$ (which is equal to $\mathrm{e}[x] = \widetilde{x} - \mathrm{e}[\Delta x] = \widetilde{x} - \Delta_s x$) belongs to the interval $\boldsymbol{E} = [\widetilde{x} - \Delta_s, \widetilde{x} + \Delta_s]$.

If we have this information for every x_i, then, in addition to the interval $\boldsymbol{y}$ of possible values of y, we would also like to know the interval of possible values of $\mathrm{e}[y]$. 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:

Precise formulation of the problem. Given an algorithm computing a function $f(x_1, \ldots, x_i, \ldots, x_n)$ from R^n to R, and values $\underline{x}_1$, $\overline{x}_1$, $\ldots$, $\underline{x}_i$, $\overline{x}_i$, $\ldots$, $\underline{x}_n$, $\overline{x}_n$, $\underline{E}_1$, $\overline{E}_1$, $\ldots$, $\underline{E}_i$, $\overline{E}_i$, $\ldots$, $\underline{E}_n$, $\overline{E}_n$, we want to find

$$\underline{E} \stackrel{\text{def}}{=} \min\{\mathrm{e}[f(x_1, \ldots, x_i, \ldots, x_n)] \mid \text{ all distributions of } (x_1, \ldots, x_i, \ldots, x_n)$$
$$\text{for which } x_1 \in [\underline{x}_1, \overline{x}_1], \ldots, x_i \in [\underline{x}_i, \overline{x}_i], \ldots, x_n \in [\underline{x}_n, \overline{x}_n], \tag{35}$$
$$\mathrm{e}[x_1] \in [\underline{E}_1, \overline{E}_1], \ldots, \mathrm{e}[x_i] \in [\underline{E}_i, \overline{E}_i], \ldots, \mathrm{e}[x_n] \in [\underline{E}_n, \overline{E}_n]\};$$

and $\overline{E}$ which is the maximum of $\mathrm{e}[f(x_1, \ldots, x_n)]$ 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.

Algorithms for solving the problem: case of exactly known moments. 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 $\mathrm{e}[x_1 + x_2] = \mathrm{e}[x_1] + \mathrm{e}[x_2]$, if $y = x_1 + x_2$, there is only one possible value for $E = \mathrm{e}[y]$: 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, $\boldsymbol{E} = \boldsymbol{E}_1 + \boldsymbol{E}_2$.

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

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

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 (see, e.g., Granvilliers et al. [26] and Kreinovich [27]): For multiplication $y = x_1 \cdot x_2$, when we have no information about the correlation,

$$\underline{E} = \max(p_1 + p_2 - 1, 0) \cdot \overline{x}_1 \cdot \overline{x}_2 + \min(p_1, 1 - p_2) \cdot \overline{x}_1 \cdot \underline{x}_2 + \min(1 - p_1, p_2) \cdot \underline{x}_1 \cdot \overline{x}_2 + \max(1 - p_1 - p_2, 0) \cdot \underline{x}_1 \cdot \underline{x}_2; \tag{36}$$

$$\overline{E} = \min(p_1, p_2) \cdot \overline{x}_1 \cdot \overline{x}_2 + \max(p_1 - p_2, 0) \cdot \overline{x}_1 \cdot \underline{x}_2 + \max(p_2 - p_1, 0) \cdot \underline{x}_1 \cdot \overline{x}_2 + \min(1 - p_1, 1 - p_2) \cdot \underline{x}_1 \cdot \underline{x}_2, \tag{37}$$

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

For the *inverse* $y = 1/x_1$, the finite range is possible only when $0 \notin \boldsymbol{x}_1$. Without losing generality, we can consider the case when $0 < \underline{x}_1$. In this case, the range of possible values of E is $\boldsymbol{E} = [1/E_1, p_1/\overline{x}_1 + (1 - p_1)/\underline{x}_1]$.

Similar formulas can be produced for max and min, and also 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.

Algorithms for solving the problem: general case. For multiplication (under no assumption about correlation), if we only know the intervals of possible values of E_i, then 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 $1 \in \boldsymbol{p}_1 + \boldsymbol{p}_2$); and
- $p_1 = \min(\overline{p}_1, 1 - \underline{p}_2)$ and $p_2 = 1 - p_1$ (if $1 \in \boldsymbol{p}_1 + \boldsymbol{p}_2$).

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

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

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

Important open problems. What if, in addition to intervals and first moments, we also know second moments? This problem is important for design of computer chips.

What if, in addition to moments, we also know p-boxes?

Additional problem: how to estimate bounds on the moments? If we knew the exact values Δx_i of the measurement errors, we could estimate the moments by

using the standard formulas: the mean as $\mathrm{e}(\Delta x) = \frac{1}{n}(\Delta x_1 + \ldots + \Delta x_i + \ldots + \Delta x_n)$, the variance as $\mathrm{var}(\Delta x) = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (\Delta x_i - E(\Delta x))^2$, and the covariance as

$$\mathrm{cov}(\Delta x, \Delta y) = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (\Delta x_i - E(\Delta x)) \cdot (\Delta y_i - E(\Delta y)). \tag{38}$$

In practice, we do not know the actual value of $\Delta x_i = \widetilde{x}_i - x_i$, we only know an approximate value $\widetilde{\Delta} x_i = \widetilde{x}_i - \widetilde{x}_i{}^{\mathrm{st}}$, where $\widetilde{x}_i{}^{\mathrm{st}}$ is the result of measuring the same quantity by a standard (much more accurate) measuring instrument.

For the standard measuring instrument, we often only know the upper bound Δ_i^{st} on its measurement error: $|\widetilde{x}_i{}^{\mathrm{st}} - x_i| \le \Delta_i^{\mathrm{st}}$. In this case, we only know that $\Delta x_i \in [\widetilde{\Delta} x_i - \Delta_i^{\mathrm{st}}, \widetilde{\Delta} x_i + \Delta_i^{\mathrm{st}}]$. So, to find guaranteed bounds for each of the above statistical characteristics $\mathrm{c}(\Delta x_1, \ldots, \Delta x_i, \ldots, \Delta x_n)$, we must find the *range* of possible values of the corresponding characteristics when Δx_i belongs to the corresponding interval $[\underline{\Delta} x_i, \overline{\Delta} x_i]$.

For some characteristics, computing the corresponding range is easy. For example, the mean $E(\Delta x)$ is a monotonic function of all its variables, so its range can be computed as $\mathbf{e}(\Delta) = [\underline{E}, \overline{E}]$, where $\underline{E} = \frac{1}{n}(\underline{\Delta} x_1 + \ldots + \underline{\Delta} x_i + \ldots + \underline{\Delta} x_n)$ and $\overline{E} = \frac{1}{n}(\overline{\Delta} x_1 + \ldots + \overline{\Delta} x_i + \ldots + \overline{\Delta} x_n)$.

For other statistics such as variance $\mathrm{var}(\Delta x)$ or covariance $\mathrm{cov}(\Delta x, \Delta y)$, the problem is, in general, NP-hard; Ferson et al. [28]. In such cases, in general, we have to use approximate techniques. There are, however, practically meaningful situations in which it is possible to efficiently compute the exact range of the variance and of other characteristics; the corresponding algorithms are summarized in Ferson et al. [29], Kreinovich et al. [30], and Kreinovich et al. [31].

Comment. Similar algorithms can be used in the general situation of statistical processing under interval uncertainty. Interval uncertainty can come from measurement errors, but there are also other sources of interval uncertainty:

- A source of interval uncertainty is the existence of detection limits for different sensors: if a sensor, e.g., did not detect any ozone, this means that the ozone concentration is below its detection limit DL, i.e., in the interval $[0, DL]$.
- Yet another source of interval uncertainty is discretized data: if we experiment on the fish and watch it daily, and a fish is alive on Day 5 but dead on Day 6, then all we know about its lifetime is that it is in the interval $[5, 6]$.
- Expert estimates often come as intervals.
- The need to keep privacy in statistical (e.g., medical) databases also often leads to the fact that instead of recording, e.g., exact age, what we only record is the interval $[40, 50]$.

In all these situations, the algorithms from Ferson et al. [29], Kreinovich et al. [30], and Kreinovich et al. [31] can be used.

9 Final Remarks

Traditional statistical approach to processing measurement errors Δx_i is based on the assumption that we have a *full* information about the probability distributions

for these errors. Typically, it is assumed that these errors are independent and normally distributed, with known means and standard deviations.

In practice, however, we often only have a *partial* information about the corresponding probability distributions. For example, sometimes, we only know the upper bound Δ_i on the (absolute values of the) measurement errors, i.e., we only know that Δx_i belongs to the interval $[-\Delta_i, \Delta_i]$. In this case, a usual engineering approach is to select, from several possible distributions, the most "reasonable one" – e.g., the uniform distribution on $[-\Delta_i, \Delta_i]$. We show that this selection sometimes drastically underestimates the error of indirect measurements. To get more adequate estimates, we must use *robust* statistical techniques, i.e., techniques which take into account all the probability distributions which are consistent with our knowledge. For the case of interval uncertainty, such techniques are called *interval computations*. In this chapter, we overviewed interval computations techniques and more general techniques of robust statistics.

Acknowledgement. This work was supported in part by NSF grants HRD-0734825, EAR-0225670, and EIA-0080940, by Texas Department of Transportation grant No. 0-5453, by the Japan Advanced Institute of Science and Technology (JAIST) International Joint Research Grant 2006-08, and by the Max Planck Institut für Mathematik.

The author is greatly thankful to the anonymous referees for valuable suggestions.

References

1. Rabinovich, S.G.: Measurement Errors and Uncertainty. Theory and Practice, Springer Verlag, Berlin (2005)
2. Jaulin, L., et al.: Applied Interval Analysis. Springer Verlag, London (2001)
3. Kearfott, R.B., Kreinovich, V. (eds.): Applications of Interval Computations. Kluwer, Dordrecht (1996)
4. Jaynes, E.T., Bretthorst, G.L.: Probability Theory: the logic of science, Cambridge University Press, Cambridge, UK (2003)
5. Sheskin, D.J.: Handbook of Parametric and Nonparametric Statistical Procedures, Chapman & Hall/CRC, Boca Raton, Florida (2004)
6. Kreinovich, V., et al.: Monte-Carlo-type techniques for processing interval uncertainty, and their potential engineering applications. Reliable Computing **13**(1), 25–69 (2007)
7. Kreinovich, V., Ferson, S.: A new Cauchy-Based black-box technique for uncertainty in risk analysis. Reliability Engineering and Systems Safety **85**(1–3), 267–279 (2004)
8. Trejo, R., Kreinovich, V.: Error estimations for indirect measurements. In: Rajasekaran, S. et al. (eds.), Handbook on Randomized Computing, pp. 673–729, Kluwer, Dordrecht (2001)
9. Archimedes, On the measurement of the circle, In: Heath, T.L. (ed.), The Works of Archimedes. Cambridge University Press, Cambridge (1897); Dover (1953)
10. Young, W.H.: Sull due funzioni a piu valori constituite dai limiti d'una funzione di variable reale a destra ed a sinistra di ciascun punto. Rendiconti Academia di Lincei, Classes di Scienza Fiziche, **17**(5), 582–587 (1908)

11. Young, R.C.: The algebra of multi-valued quantities. Mathematische Annalen **104**, 260–290 (1931)
12. Dwyer, P.S.: Linear Computations, J. Wiley, N.Y. (1951)
13. Warmus, M.: Calculus of approximations. Bulletin de l'Academie Polonaise de Sciences **4**(5), 253–257 (1956)
14. Sunaga, T.: Theory of interval algebra and its application to numerical analysis, RAAG Memoirs, Ggujutsu Bunken Fukuy-kai, Tokyo **2**, 29–46 (547–564) (1958)
15. Moore R.E.: Automatic error analysis in digital computation, Space Div. Report LMSD84821, Lockheed Missiles and Space Co. (1959)
16. Moore, R.E.: Interval Analysis. Prentice-Hall, Englewood Cliffs, N.J. (1966)
17. Interval computations website http://www.cs.utep.edu/interval-comp
18. Cerimele, M.M., Venturini Zilli, M.: Effective numerical approximation by intervals. Freiburger Intervall-Berichte, 85/4, 1–24 (1985)
19. Kreinovich, V., et al.: Computational Complexity and Feasibility of Data Processing and Interval Computations, Kluwer, Dordrecht (1997)
20. Berz, M., Hoffstätter, G.: Computation and Application of Taylor Polynomials with Interval Remainder Bounds. Reliable Computing **4**(1), 83–97 (1998)
21. Neumaier, A.: Taylor forms. Reliable Computing **9**, 43–79 (2002)
22. Ferson, S.: Risk Assessment with Uncertainty Numbers: RiskCalc, CRC Press, Boca Raton, Florida (2002)
23. Lodwick, W.A., Jamison, K.D.: Estimating and validating the cumulative distribution of a function of random variables. Reliable Computing **9**(2), 127–141 (2003)
24. Ferson, S., et al.: Constructing Probability Boxes and Dempster-Shafer Structures. Sandia National Laboratories, Report SAND2002-4015 (2003)
25. Berleant, D., Zhang, J.: Representation and Problem Solving with the Distribution Envelope Determination (DEnv) Method. Reliability Engineering and System Safety **85**(1–3), 153–168 (2004)
26. Granvilliers, L., Kreinovich, V., Mueller, N.: Novel approaches to numerical software with result verification. In: Alt, R. et al. (eds.), Numerical Software with Result Verification, Springer Lectures Notes in Computer Science **2291**, 274–305 (2004)
27. Kreinovich, V.: Probabilities, intervals, what next? Journal of Global Optimization **29**(3) 265–280 (2004)
28. Ferson, S., et al.: Exact bounds on finite populations of interval data. Reliable Computing **11** (3), 207–233 (2005)
29. Ferson, S., et al.: Experimental Uncertainty Estimation and Statistics for Data Having Interval Uncertainty. Sandia National Laboratories, Report SAND2007-0939, May 2007; available as http://www.ramas.com/intstats.pdf and as part of the book DVD
30. Kreinovich, V., et al.: Towards combining probabilistic and interval uncertainty in engineering calculations. Reliable Computing **12**(6), 471–501 (2006)
31. Kreinovich, V., et al.: Interval versions of statistical techniques. Journal of Computational and Applied Mathematics **199**(2), 418–423 (2007)