FOR UNKNOWN–BUT–BOUNDED ERRORS, INTERVAL ESTIMATES ARE OFTEN BETTER THAN AVERAGING

G. William Walster[1], Vladik Kreinovich[2]
[1]SunPro, 2550 Garcia Avenue, Mountain View, CA 94043-1100, USA
[2]Computer Science Dept., University of Texas at El Paso, El Paso, TX 79968, USA

Abstract. For many measuring devices, the only information that we have about them is their biggest possible error $\varepsilon > 0$. In other words, we know that the error $\Delta x = \tilde{x} - x$ (i.e., the difference between the measured value $\tilde{x}$ and the actual values x) is random, that this error can sometimes become as big as ε or $-\varepsilon$, but we do not have any information about the probabilities of different values of error.

Methods of statistics enable us to generate a better estimate for x by making several measurements $\tilde{x}_1, ..., \tilde{x}_n$. For example, if the average error is 0 ($E(\Delta x) = 0$), then after n measurements, we can take an average $\bar{x} = (\tilde{x}_1 + ... + \tilde{x}_n)/n$, and get an estimate whose standard deviation (and the corresponding confidence intervals) are $\sqrt{n}$ times smaller.

Another estimate comes from interval analysis: for every measurement $\tilde{x}_i$, we know that the actual value x belongs to an interval $[\tilde{x}_i - \varepsilon, \tilde{x}_i + \varepsilon]$. So, x belongs to the intersection of all these intervals. In one sense, this estimate is better than the one based on traditional engineering statistics (i.e., averaging): interval estimation is *guaranteed*. In this paper, we show that for many cases, this intersection is also better in the sense that it gives a more *accurate* estimate for x than averaging: namely, under certain reasonable conditions, the *error of this interval estimate decreases faster (as* $1/n$*) than the error of the average (that only decreases as* $1/\sqrt{n}$*).*

A similar result is proved for a multi-dimensional case, when we measure several auxiliary quantities, and use the measurement results to estimate the value of the desired quantity y.

1. FORMULATION OF A REAL-LIFE PROBLEM

In many cases, we know only the maximum possible error (i.e., an interval). For many measuring devices, the only information that we have about them is their biggest possible error $\varepsilon > 0$ (see, e.g., a survey monograph [NZ91]). In other words, we know that the error $\Delta x = \tilde{x} - x$ (i.e., the difference between the measured and the actual values) is random, that this error can sometimes become as big as ε or $-\varepsilon$, but we do not have any information about the probabilities of different values of error.

Since an error is random, we can improve the accuracy by measuring one and the same quantity several times. Usually, we can safely assume that the errors of different consequent measurements are independent random variables. In this case, methods of mathematical statistics enable us to improve the estimate for x by repeating the measurement several times and combining the results into an improved estimate for x.

The main problem is: *to choose a method that results in a better accuracy.*

What we are going to do. In this paper, we will describe two methods: the simplest statistical method (based on averaging) that is most widely used in engineering and science

applications, and a method based on interval analysis. We will show that, in the general case, interval estimates are more accurate.

Comment. Our intention is to make this paper understandable for all readers of the journal, including those readers who are not experts in mathematical statistics. For these readers, we repeat several definitions (well known in mathematical statistics) that we will need, and describe our proofs in all necessary detail.

2. STATISTICAL METHOD BASED ON AVERAGING

Description of a method. Let us consider a frequent case when the average error $E(\Delta x)$ equals 0 (as usual, $E(\alpha)$ denotes an expected value of α). Then, after n measurements, we can take $\bar{x} = (\tilde{x}_1 + ... + \tilde{x}_n)/n$ as a estimate for x.

This method is well known, and its error estimates are also known (see, e.g., [W90], [B91], and references therein). However, for the readers who are not very familiar with mathematical statistics, we decided to include these estimates and their derivation into this paper (especially since these derivations are very simple and do not occupy lots of space).

So, let us describe how to estimate the accuracy of this estimate $\bar{x}$.

The error of this estimate is (for large n) normally distributed. The error of this estimate is equal to $\Delta x = (\Delta x_1 + ... + \Delta x_n)/n$. Here, the variables Δx_i are independent, and have the same probability distribution with 0 average. Therefore, according to the central limit theorem, for large n, the distribution of Δx is close to Gaussian (=normal).

It is thus sufficient to estimate the average and the standard deviation. Gaussian distribution is uniquely determined by two parameters: average $E(\Delta x)$ and standard deviation $\sigma(\Delta x)$. Therefore, for large n, it is sufficient to estimate these two parameters, and then we will be able to use the estimates based on the normal distribution. Namely, statistics recommends us to choose some k (= 2, 3, 4, ...), and conclude that with high reliability, the actual value of the error Δx belongs to an interval

$$[E(\Delta) - k\sigma(\Delta), E(\Delta) + k\sigma(\Delta)].$$

For example, if we take $k = 2$, then this conclusion is true with probability $\approx 95\%$; for $k = 3$, it is true with probability $\approx 99.9\%$, etc.

So, to get the desired estimate, we must estimate $E(\Delta)$ and $\sigma(\Delta)$.

Estimating the average. Since all the component errors Δx_i have 0 average, the resulting error Δx is also of 0 average.

Estimating the standard deviation. Since the errors Δx_i are all mutually independent, and distributed according to the same probability distribution, the standard deviation $\sigma[\bar{x}]$ of $\bar{x}$ is equal to $\sigma/\sqrt{n}$ (where by σ, we denoted the standard deviation of Δx_i).

So, to get the desired estimate for $\sigma[\Delta x]$, it is necessary to estimate $\sigma[\Delta x_i]$. A standard deviation is defined as a square root of the average value $E[(\Delta x_i - E(\Delta x_i))^2]$.

Since $E(\Delta x_i) = 0$, we conclude that σ^2 equals to the average $E(\Delta x_i^2)$ of Δx_i^2. Due to the fact that measurement errors cannot exceed ε, we have $\Delta x_i^2 \le \varepsilon^2$. Therefore, the average $E(\Delta x_i^2)$ also does not exceed ε^2. So, $\sigma^2 \le \varepsilon^2$, and $\sigma \le \varepsilon$.

Therefore, $\sigma(\Delta x) \le \varepsilon/\sqrt{n}$.

Final error estimate. Since $E(\Delta x) = 0$, we conclude that the error Δx belongs to an interval $[-k\sigma(\Delta x), k\sigma(\Delta x)]$. In other words, we conclude that $|\Delta x| \le k\sigma(\Delta x)$. Due to the fact that $\sigma(\Delta x) \le \varepsilon/\sqrt{n}$, we conclude that for large n,

$$|\Delta x| \le k\varepsilon/\sqrt{n}.$$

In other words, after n measurements, the standard deviation and the resulting confidence intervals can be made $\approx \sqrt{n}$ times smaller.

Let us describe this estimate formally.

Definition 1. *Assume that a number $p_0 \in (0,1)$ is given. This number is called a confidence level. Assume also that some probability density $\rho(x)$ is defined on the interval $[-\varepsilon, \varepsilon]$, for which the average is 0 (i.e., $\int x\rho(x)\,dx = 0$). Let $\tilde{x}_1, ..., \tilde{x}_n$ be independent random variables with the probability density $\rho(x-a)$ for some (unknown) a. By an accuracy of the estimate $\bar{x} = (\tilde{x}_1 + ... + \tilde{x}_n)/n$, we mean the smallest $\delta > 0$, for which the inequality $|\bar{x} - a| \le \delta$ holds with probability $\ge p_0$. This accuracy will be denoted by $\delta_n^{\rm av}$.*

Comment. In other words, $\delta_n^{\rm av}$ is the smallest real number, for which

$$P(|\bar{x} - a| \le \delta) \ge p_0,$$

where $P(A)$ denotes the probability of an event A.

To describe how $\delta_n^{\rm av}$ decreases with n, we must recall the standard notations describing the asymptotic behavior of sequences:

Denotations. *We say that two sequences a_n and b_n are asymptotically equivalent and denote it by $a_n \sim b_n$ if $\lim a_n/b_n = 1$. Notation $a_n = O(b_n)$ means that there exists a $C > 0$ such that for all n, $a_n \le Cb_n$. Notation $a_n = o(b_n)$ means that $a_n/b_n \to 0$ as $n \to \infty$.*

PROPOSITION. *Assume that on an interval $[-\varepsilon, \varepsilon]$, a probability density $\rho(x)$ is defined for which the average is 0 (i.e., $\int x\rho(x)\,dx = 0$). Then, for every confidence level p_0, there exists a constant C such that $\delta_n^{\rm av} \sim C/\sqrt{n}$.*

Comment. This is a well-known (and reasonably easy-to-prove) result (see, e.g., [I81]).

3. INTERVAL ESTIMATES

The origin of these estimates. In the above estimates, we applied traditional statistical methods; however, this was not a typical application of these methods, because we knew nothing about the probability distribution. It would be more natural to use the fact that

the only thing we know about each measurement error Δx_i is that it belongs to an interval $[-\varepsilon, \varepsilon]$.

In general, there is an area of statistics that designs statistical estimation techniques for the situations in which we do not know the exact distribution or its shape, and we only know the class of possible distributions; this area is called *robust statistics* [H81], [W90]. In particular, estimates for the case when we know only the intervals of possible values of input data have been developed (since the pioneer work [M64]) in *interval mathematics* (see, e.g., [M79], [A88], [K96]). In this sense, interval mathematics can be viewed as a particular case of robust statistics.

The estimate itself. For every measurement $\tilde{x}_i$, we know that the error $\Delta x_i = \tilde{x}_i - x$ belongs to an interval $[-\varepsilon, \varepsilon]$. Therefore, the actual value x belongs to an interval $\mathbf{x}_i = [\tilde{x}_i - \varepsilon, \tilde{x}_i + \varepsilon]$. So, after n measurements, we can conclude that x belongs to the intersection of n corresponding intervals

$$\bigcap_{i=1}^{n} \mathbf{x}_i = \bigcap_{i=1}^{n} [\tilde{x}_i - \varepsilon, \tilde{x}_i + \varepsilon].$$

In other words, we conclude that $x \in [x^-, x^+]$, where

$$x^- = \max_{1 \le i \le n} (\tilde{x}_i - \varepsilon) = \max_{1 \le i \le n} \tilde{x}_i - \varepsilon,$$

$$x^+ = \min_{1 \le i \le n} (\tilde{x}_i + \varepsilon) = \min_{1 \le i \le n} \tilde{x}_i + \varepsilon.$$

This estimate was proposed in [W88] (see also [VS89], [D91]).

What is the accuracy of this estimate? Numerical experiments [W88] showed that this estimate is often much more accurate than the averaging. In [D91], it is proved that under some reasonable assumptions, the *standard deviation* error of the interval estimate decreases as $1/n$ and is thus much better than the accuracy of the averaging (for which the error is $\sim 1/\sqrt{n}$).

In the following sections, we will show that not only the interval estimate is better "in the average" (in the sense of standard deviations), but it is better in the *vast majority* of cases: namely, whatever probability $p_0 < 1$ we choose (be it 90%, or 99.999%), starting from some n, interval estimates are better in $\ge p_0$ of all the cases.

4. WHAT DISTRIBUTIONS ARE POSSIBLE?

To estimate the accuracy of an interval estimate, we must know the original error distribution. The accuracy of the interval estimate depends on the unknown probability density $\rho(x)$. The only thing we know about $\rho(x)$ is that this function is different from 0 only for $x \in [-\varepsilon, \varepsilon]$.

Our main assumption about the distribution. We know also that both values ε and $-\varepsilon$ are possible values of the error. How can we formulate that additional knowledge in terms of the function $\rho(x)$?

A typical way to determine a probability density from the experimental data is as follows: Suppose that we have a sample $\alpha_1, ..., \alpha_N$. Then, we divide the interval $[\min \alpha_i, \max \alpha_i]$ into several subintervals, and for every subinterval I, estimate the frequency f_I with which the values α_i fall into this interval as $f_I = N(I)/N$ (where $N(I)$ is a total number of those values α_i that are in I). Then, we can define an empirical density function $\tilde{\rho}(x)$ (that approximates the actual density function $\rho(x)$) as follows: for $x \in I$, we take $\tilde{\rho}(x) = f_I$.

If we fix a finite set of subintervals, and apply this procedure to our case, then due to the fact that both values ε and $-\varepsilon$ are possible, the probability that $N(I) > 0$ for subintervals I that contain ε or $-\varepsilon$ will tend to 1 as N increases. Therefore, for sufficiently large N, we will have $f_I > 0$ for these intervals (with probability close to 1), and, as a result, $\tilde{\rho}(\varepsilon) > 0$ and $\tilde{\rho}(-\varepsilon) > 0$ for the empirical approximation $\tilde{\rho}(x)$ to the probability density function.

Also, the above-described approximation function $\tilde{\rho}(x)$ (defined on the interval $[-\varepsilon, \varepsilon]$) is continuous in some neighborhoods of ε and $-\varepsilon$.

Therefore, it seems reasonable to assume that the same properties hold for the actual density function $\rho(x)$: namely, that $\rho(-\varepsilon) > 0$, $\rho(\varepsilon) > 0$, and that $\rho(x)$ is continuous in some vicinity of ε and $-\varepsilon$.

Comment. Our motivations of the assumption that the probability density is positive (and continuous) at the endpoints are not 100% convincing, and we therefore *do not claim* that all real-life distributions should satisfy this property. However, this assumption (as we will see later) is necessary to prove that interval estimates are better than averaging. So, what we *do claim* is that there is a reasonable class of distributions for which this assumption holds, and for which interval estimates are better than averaging.

This assumption describes a generic case. From the mathematical viewpoint, the above restrictions mean that out of all possible values of $\rho(-\varepsilon)$ and $\rho(\varepsilon)$ (i.e., out of all non-negative numbers), we exclude only one value 0. So, our case is really the generic case.

5. MAIN RESULT: ONE-DIMENSIONAL CASE

Definition 2. *Assume that a number $p_0 \in (0, 1)$ is given. This number will be called a confidence level. Assume also that some probability density $\rho(x)$ is defined on the interval $[-\varepsilon, \varepsilon]$, such that $\rho(\varepsilon) > 0$, $\rho(-\varepsilon) > 0$, and that is continuous in some neighborhoods of ε and $-\varepsilon$. Let $\tilde{x}_1, ..., \tilde{x}_n$ be independent random variables with the probability density $\rho(x - a)$ for some (unknown) a. By an accuracy of the interval estimate $[x^-, x^+]$ (where $x^- = \max \tilde{x}_i - \varepsilon$, and $x^+ = \min \tilde{x}_i + \varepsilon$), we mean the smallest $\delta > 0$, for which both inequalities $|x^- - a| \le \delta$ and $|x^+ - a| \le \delta$ hold with probability $\ge p_0$. This accuracy will be denoted by $\delta_n^{\rm int}$.*

Comment. In other words, $\delta_n^{\rm int}$ is the smallest real number, for which

$$P(|x^- - a| \le \delta \ \&\ |x^+ - a| \le \delta) \ge p_0.$$

THEOREM 1. *Assume that on an interval* $[-\varepsilon, \varepsilon]$, *a probability density* $\rho(x)$ *is defined, for which* $\rho(\varepsilon) > 0$, $\rho(-\varepsilon) > 0$, *and* ρ *is continuous in some neighborhoods of* ε *and* $-\varepsilon$. *Then,* $\delta_n^{\rm int} = O(1/n)$.

So, **the accuracy of the interval estimate** is $\leq C/n$, and thus, for large n, it **is much better than the accuracy of averaging** $\delta_n^{\rm av} \sim C_{\rm av}/\sqrt{n}$. Besides, the averaging is applicable only to the case when $E(\Delta x) = 0$, while an interval estimate is also applicable to distributions with non-zero average.

Comments.

- For readers' convenience, all the proofs are placed in Section 7.
- The proof of Theorem 1 is rather simple and straightforward. However, we decided to place it here, because it is new (as far as we know), of importance to the interval community, and not completely trivial.
- Our Theorem 1 is in good accordance with the known statistical results: e.g., it is known that for the *uniform distribution* on the interval $[-\varepsilon, \varepsilon]$, the estimator $(1/2) \cdot (\min \tilde{x}_i + \max \tilde{x}_i)$ is optimal in some reasonable sense (to be more precise, it is a minimax estimate [I81], Ch. 1, Section 3.4).
- If the condition that $\rho(\varepsilon) > 0$ and $\rho(-\varepsilon) > 0$ is not satisfied, that the conclusion of Theorem 1 is not necessarily true: indeed, e.g., if ρ is a uniform distribution on an interval $[-\varepsilon/2, \varepsilon/2]$, then $-\varepsilon/2 \leq \tilde{x}_i \leq \varepsilon/2$, and, therefore, $x^- = \max \tilde{x}_i - \varepsilon \leq \varepsilon/2 - \varepsilon = -\varepsilon/2$ and similarly, $x^+ \geq \varepsilon/2$. In this case, for all n, the interval accuracy $\delta_n^{\rm int} \geq \varepsilon/2$, and $\delta_n^{\rm int}$ does not tend to 0 at all. In other cases (e.g., if $\rho(\pm\varepsilon) = 0$ and the derivatives $\rho'(\pm\varepsilon)$ are different from 0), we have $\delta_n^{\rm int} \to 0$, but much slower than in the case described by Theorem 1.

Empirical confirmation. Numerical simulations, performed by one of the authors (G.W.W.), confirmed the above theoretical conclusion that interval estimates are more accurate [W88].

An example when the interval estimate is even more accurate (see, e.g., [I81]). The fact that we have only an upper estimate for $\delta_n^{\rm int}$ is not accidental: in some cases, the accuracy is even better. For example, let us consider the case when the error takes only two possible values ε and $-\varepsilon$ with equal probability 1/2. In this case, when we measure one and the same value x several times, we can get only two possible results: $x - \varepsilon$ and $x + \varepsilon$. So, if we encountered two different results, then they are equal to $x - \varepsilon$ and $x + \varepsilon$.

Since the probability to have an error ε is equal to 1/2, and the errors are independent, the probability that all n errors are equal to ε is equal to 2^{-n}. Likewise, the probability that all n errors are equal to $-\varepsilon$, is equal to 2^{-n}. In all remaining cases (i.e., with probability $1 - 2 \times 2^{-n}$), we have two different measurement results among n. Therefore, one of them is equal to $x+\varepsilon$, and one of them is equal to $x+\varepsilon$. Hence, $x^- = \max \tilde{x}_i - \varepsilon = x + \varepsilon - \varepsilon = x$ and likewise, $x^+ = x^- = x$. So, in this situation, the interval estimate is *absolutely accurate* ($\delta_n^{\rm int} = 0$).

This happens, when $1 - 2 \times 2^{-n} = 1 - 2^{-(n-1)} \geq p_0$, i.e., when $2^{-(n-1)} \leq 1 - p_0$, and $n \geq 1 - \log_2(1 - p_0) = \log_2(2/(1 - p_0))$.

These values of n are quite reasonable: e.g., for $p_0 = 95\%$, this is true for $n \geq 6$; for $p = 99.9\%$, it is true for $n \geq 11$, etc.

6. ADDITIONAL RESULT: MULTI-DIMENSIONAL CASE

Formulation of a physical problem. Suppose that we are interested in the values of some physical quantity y, and it is either impossible, or too difficult to measure them directly. So, we apply an *indirect measurement* procedure, i.e., we measure other quantities $x_1, ..., x_n$ that are in known relationship with the desired quantity y (and maybe some auxiliary quantities $y_2, ..., y_p$ that are also not directly measurable), and then reconstruct y from the results $\tilde{x}_i$ of these measurements. In some cases, we have an explicit formula that expresses y in terms of the measurement results. However, in the general case, the relation between x_i and y is *implicit*, i.e., it is given in terms of the equations $F_j(x_1, ..., x_n, y, y_2, .., y_p) = 0$ $(j = 1, 2, ...)$.

In general, for a system of equations to have a unique solution, we need as many equations as there are unknowns. In our case, we have p unknowns $y, y_2, ..., y_p$, therefore, we need p equations $(1 \leq j \leq p)$. If we perform exactly p measurements, then we can compute the estimate $\tilde{y}$ for y by solving a system of p equations $F_j(\tilde{x}_1, ..., \tilde{x}_n, \tilde{y}, \tilde{y}_2, ..., \tilde{y}_p) = 0$ $(1 \leq j \leq p)$ with p unknowns $\tilde{y}, \tilde{y}_2, ..., \tilde{y}_p$.

To improve the accuracy of this estimate, it is reasonable to repeat each measurement of $x_1, ..., x_n$ several times. After $k-$th measurement $(1 \leq k \leq N)$, we get n measurement results $\tilde{x}_1^{(k)}, ..., \tilde{x}_n^{(k)}$. Totally, we have $N \cdot n$ measurement results. What accuracy in y can we achieve?

For this problem, we will also consider two classes of estimates: estimates based on averaging, and interval estimates.

Estimates based on averaging. There are two possible estimates based on averaging:

- first, we can take, for every i, an average $\bar{x}_i$ of N values $\tilde{x}_i^{(k)}$, $1 \leq k \leq N$, and use these averages to estimate y.
- second, we can use each sequence of n measurement results to compute an estimate $\tilde{y}^{(k)}$ for y, and then take an average of the resulting N values $\tilde{y}^{(1)}, ..., \tilde{y}^{(N)}$.

The first estimate is computationally preferable, because in this case we must solve a non-linear system of equations only once, while in the second case, we must repeat this time-consuming procedure N times. Therefore, in the following text, we will consider only the first estimate.

Comment. By doing so, we do not lose any accuracy, because one can prove that the accuracy of the second estimate is (asymptotically) the same.

Interval estimate. For every measurement, we usually know the biggest possible value of a measurement error. In other words, we know that for every i, the measurement error $\Delta x_i = \tilde{x}_i - x_i$ belongs to an interval $[-\varepsilon_i, \varepsilon_i]$. Therefore, the actual value of x_i belongs to an interval $\mathbf{x}_i^{(k)} = [\tilde{x}_i^{(k)} - \varepsilon_i, \tilde{x}_i^{(k)} + \varepsilon_i]$. Similar to averaging, we can consider two possible interval estimates:

- first, we can take for every i, the intersection $\mathbf{x}_i$ of the corresponding N intervals $\mathbf{x}_i^{(1)}, ..., \mathbf{x}_i^{(N)}$, and then use these intervals $\mathbf{x}_i$ to find an interval estimate for y.
- or, alternatively, we can use each of N measurement results to find the corresponding interval estimate for y, and then take the intersection of the resulting N intervals for y.

Similar to averaging, the first estimate is computationally preferable, so we will consider only the first estimate. Again, similar to averaging, the second estimate has the same accuracy.

Let us now formulate the problem in mathematical terms.

Mathematical formulation of the problem and the main results.

Definition 3. *By an indirect measurement situation, we mean a tuple consisting of two positive integers n and p, an open region $U \subseteq R^n$, and p differentiable functions*

$$f_j : U \times R^p \to R,\ 1 \le j \le p,$$

such that for every $\vec{x} \in U$, there exists a unique value $\vec{y} \in R^p$ for which $\vec{F}(\vec{x}, \vec{y}) = \vec{0}$. We will say that these values $\vec{y}$ correspond to $\vec{x}$. In particular, we will say that the corresponding value of $y = y_1$ is an estimate that corresponds to $\vec{x}$. We say that an indirect measurement situation is non-degenerate if for every $\vec{x}$ and for the corresponding $\vec{y}$, $\det |\partial F_j / \partial y_l| \neq 0$.

Comment. In the following text, we will assume that our indirect measurement situation is non-degenerate.

Definition 4. *Assume that n values $x_1, ..., x_n$ are given. They will be called actual (or true) values of x_i and denoted by x_i. The value $y = y_1$ that corresponds to these actual values of x_i, will be called an actual, or true value of y, and will be denoted by y.*

Definition 5. *For every set of Nn real numbers $\tilde{x}_1^{(k)}, ..., \tilde{x}_n^{(k)}$, $1 \le k \le N$, by an estimate based on averaging $\bar{x}_i$ for x_i, we mean an average $\bar{x}_i = (\tilde{x}_i^{(1)} + ... + \tilde{x}_i^{(N)})/N$. By a estimate based on averaging $\bar{y}$ for y, we mean an estimate that corresponds to $\bar{x}_1, ..., \bar{x}_n$.*

Definition 1′. *Assume that a number $p_0 \in (0, 1)$ is given. This number will be called a confidence level. Assume also that for every i from 1 to n, a probability density $\rho_i(x)$ with 0 average is defined on some interval $[-\varepsilon_i, \varepsilon_i]$. Let $\tilde{x}_i^{(k)}$, $1 \le i \le n$, $1 \le k \le N$, be Nn independent random variables, and the probability density for $\tilde{x}_i^{(k)}$ is $\rho_i(x - x_i)$ (where x_i denotes an actual value). By an accuracy of an estimate $\bar{y}$ based on averaging, we mean the smallest $\delta > 0$ for which the inequality $|\bar{y} - y| \le \delta$ holds with probability $\ge p_0$. This accuracy will be denoted by $\delta_n^{\rm av}$.*

THEOREM 2. *For every set of probability densities ρ_i, and for any confidence level p_0, $\delta_n^{\rm av} = O(1/\sqrt{n})$.*

Comment. Unlike the 1-dimensional case, we cannot state that in all cases, $\delta_n^{\rm av} \sim C/\sqrt{n}$ for some $C > 0$, because our definitions do not exclude the case when y is identically 0 (e.g., when $F_j = y_j$), and in this case, $\Delta y = 0$ and $\delta_n^{\rm av} = 0$. These cases, when y is actually

not depending on x_i at all, are physically meaningless: why should we measure x_i if y does not depend on these values? As we will see from the proof, if there is a dependency of y on x_i (i.e., if measuring x_i makes sense), then indeed $\delta_n^{\rm av} \sim C/\sqrt{n}$ for some $C > 0$.

Definition 6. *Assume that n positive real numbers ε_i are given. For every set of Nn real numbers $\tilde{x}_i^{(k)}$, $1 \le i \le n$, $1 \le k \le N$, by an interval estimate for x_i, we mean an interval $\mathbf{x}_i = [x_i^-, x_i^+]$, where $x_i^- = \max_k \tilde{x}_i^{(k)} - \varepsilon_i$ and $x_i^+ = \min_k \tilde{x}_i^{(k)} + \varepsilon_i]$. By an interval estimate $\mathbf{y} = [y^-, y^+]$, we mean an interval consisting of the values of y_1 for all vectors $\vec{y}$, for which $\vec{F}(\vec{x}, \vec{y}) = \vec{0}$ for some $x_i \in \mathbf{x}_i$.*

Definition 7. *Assume that a number $p_0 \in (0,1)$ is given. This number will be called a confidence level. Assume also that for every i from 1 to n, a probability density $\rho_i(x)$ is defined on the interval $[-\varepsilon_i, \varepsilon_i]$ such that $\rho_i(-\varepsilon_i) > 0$, $\rho(\varepsilon_i) > 0$, and ρ_i is continuous in some neighborhoods of ε_i and ε_i. Let $\tilde{x}_i^{(k)}$, $1 \le i \le n$, $1 \le k \le N$, be Nn independent random variables, and the probability density for $\tilde{x}_i^{(k)}$ is $\rho_i(x - x_i)$, where x_i are actual values. By an accuracy of the interval estimate $[y^-, y^+]$, we mean the smallest $\delta > 0$ for which both inequalities $|y^- - y| \le \delta$ and $|y^+ - y| \le \delta$, hold with probability $\ge p_0$. This accuracy will be denoted by $\delta_n^{\rm int}$.*

THEOREM 3. $\delta_n^{\rm int} = O(1/n)$.

Comments.

- So, in multi-dimensional case, the accuracy of the interval estimate is also much better than the accuracy of the estimate based on averaging.
- Theorems 2 and 3 are easy to prove (see Section 7).

7. PROOFS

Proof of Theorem 1.
Comment. The main ideas of this proof are similar to the ideas from [D91].

1. To estimate the accuracy $\delta_n^{\rm int}$, let us first estimate for a given δ, the probability P that $|x^- - a| \le \delta$ and $|x^+ - a| \le \delta$.

This probability is close to 1, so it is more convenient to estimate the probability of the opposite event, i.e., estimate P as $1 - p$, where p is the probability that either $|x^- - a| > \delta$, or $|x^+ - a| > \delta$.

Since $P(A \vee B) \le P(A) + P(B)$, we can estimate p as $p \le p^- + p^+$, where $p^- = P(|x^- - a| > \delta)$ and $p^+ = P(|x^+ - a| > \delta)$. Therefore, if we find δ, for which $p^- \le (1-p_0)/2$ and $p^+ \le (1-p_0)/2$, we will have $p \le p^- + p^+ \le 1 - p_0$, hence $P = 1 - p \ge p_0$. So, for this δ, $\delta_n^{\rm int} \le \delta$.

2. Let us estimate p^-.

2.1. Since $x^- = \max \tilde{x}_i - \varepsilon$ and $\tilde{x}_i \le a + \varepsilon$ for all i, we can conclude that $\max \tilde{x}_i \le a + \varepsilon$ and $x^- \le a$. Therefore, the difference $x^- - a$ is always non-positive, and hence the inequality $|x^- - a| > \delta$ is equivalent to $x^- - a < -\delta$, i.e., to $x^- \ge a - \delta$. So, $p^- = P(x^- < a - \delta)$.

2.2. Since $x^- = \max \tilde{x}_i - \varepsilon$, the condition $x^- < a - \delta$ is equivalent to $\max \tilde{x}_i < a + \varepsilon - \delta$. The biggest of n numbers $\tilde{x}_i$ is smaller than $a + \varepsilon - \delta$ if and only if all n numbers are smaller than that, i.e.,

$$p^- = P((\tilde{x}_1 < a + \varepsilon - \delta)\ \&\ ...\ \&\ (\tilde{x}_n < a + \varepsilon - \delta)).$$

Since the variables $\tilde{x}_i$ are independent, we conclude that

$$p^- = P(\tilde{x}_1 < a + \varepsilon - \delta) \cdot\ ...\ \cdot\ P(\tilde{x}_n < a + \varepsilon - \delta).$$

Since $\tilde{x}_i$ are equally distributed, we have $p^- = Q^n$, where by Q, we denoted

$$P(\tilde{x}_1 < a + \varepsilon - \delta).$$

So, in order to estimate p^-, it is necessary to estimate Q.

3. Let us estimate Q.

3.1. For small δ, $\tilde{x}_1 \le a + \varepsilon$; so, this probability Q is close to 1. Therefore, it is convenient to estimate it by estimating the probability q of the opposite event as $Q = 1 - q$, where $q = P(\tilde{x}_1 \ge a + \varepsilon - \delta) = P(a + \varepsilon - \delta \le \tilde{x}_1 \le a + \varepsilon)$.

In terms of probability density function, this probability q is equal to

$$q = \int_{\varepsilon - \delta}^{\varepsilon} \rho(x)\, dx.$$

This expression q is equal to the difference $I(\varepsilon) - I(\varepsilon - \delta)$, where $I(\alpha) = \int^{\alpha} \rho(x)\, dx$. It is well known that the derivative of an integral is equal to the integrated function. So, $dI(\alpha)/d\alpha = \rho(\alpha)$. In particular, for $\alpha = \varepsilon$, it means that

$$\lim_{\delta \to 0} \frac{I(\varepsilon) - I(\varepsilon - \delta)}{\delta} = \rho(\varepsilon),$$

i.e., $\lim(q/\delta) = \rho(\varepsilon)$. Using $o(.)$ notation, we can rewrite this equality as $q = \delta\rho(\varepsilon) + \Delta q$, where $\Delta q = o(\delta)$.

The denotation $\Delta q = o(\delta)$ means that $\Delta q/\delta \to 0$ as $\delta \to 0$, i.e., that for every $k > 0$, there exists a δ_0^- such that for all $\delta \le \delta_0^-$, $|\Delta q| \le k\delta$. In particular, if we take $k = \rho(\varepsilon)/2$, we conclude that for all $\delta \le \delta_0$, $q \ge \rho(\varepsilon)\delta - k\delta = \delta\rho(\varepsilon)/2$.

3.2. Therefore, $Q = 1 - q \le 1 - \delta\rho(\varepsilon)/2$.

4. Now, we can use this estimate for Q to estimate p^-: $p^- = Q^n \le (1 - \delta\rho(\varepsilon)/2)^n$.

5. Based on these estimates, how to choose δ for which $p^- \le (1 - p_0)/2$?

In view of the above estimate, to guarantee that $p^- \le (1-p_0)/2$, we must take δ for which $\delta \le \delta_0^-$ and $(1-\delta\rho(\varepsilon)/2)^n \le (1-p_0)/2$. Applying equivalent transformations to the second of these inequalities, we consequently get

$$1 - \frac{\delta\rho(\varepsilon)}{2} \le \big(\frac{1-p_0}{2}\big)^{1/n},$$

$$\frac{\delta\rho(\varepsilon)}{2} \ge 1 - \big(\frac{1-p_0}{2}\big)^{1/n},$$

and finally, $\delta \ge \delta_n^-$, where we denoted

$$\delta_n^- = \frac{2}{\rho(\varepsilon)}\big(1 - \big(\frac{1-p_0}{2}\big)^{1/n}\big).$$

So, if $\delta \le \delta_0^-$ and $\delta \ge \delta_n^-$, then we can guarantee that $p^- \le (1-p_0)/2$.

6. Let us estimate the asymptotic behavior of δ_n^- when $n \to \infty$.

First,

$$\big(\frac{1-p_0}{2}\big)^{1/n} = \big[\exp\big(\ln\frac{1-p_0}{2}\big)\big]^{1/n} = \exp\big(\frac{1}{n}\ln\frac{1-p_0}{2}\big).$$

The expression under exp tends to 0 as $n \to \infty$, and $\exp(\alpha) = 1 + \alpha + o(\alpha)$. Therefore,

$$\big(\frac{1-p_0}{2}\big)^{1/n} = 1 + \frac{1}{n}\ln\frac{1-p_0}{2} + o\big(\frac{1}{n}\big).$$

Hence,

$$1 - \big(\frac{1-p_0}{2}\big)^{1/n} = -\frac{1}{n}\ln\frac{1-p_0}{2} + o\big(\frac{1}{n}\big),$$

and

$$\delta_n^- = \frac{2}{\rho(\varepsilon)}\{1 - \big(\frac{1-p_0}{2}\big)^{1/n}\} = -\frac{2}{\rho(\varepsilon)}\frac{1}{n}\ln\frac{1-p_0}{2} + o\big(\frac{1}{n}\big).$$

So,

$$\delta_n^- = \frac{A^-}{n} + o\big(\frac{1}{n}\big),$$

where we denoted

$$A^- = -\frac{2}{\rho(\varepsilon)}\ln\frac{1-p_0}{2}.$$

7. Likewise, one can define δ_n^+ and δ_0^+ such that if $\delta \le \delta_0^+$ and $\delta \ge \delta_n^+$, then $p^+ \le (1-p_0)/2$, and prove that

$$\delta_n^+ = \frac{A^+}{n} + o\big(\frac{1}{n}\big)$$

for some $A^+ > 0$.

8. Finally, we are ready to estimate $\delta_n^{\rm int}$.

Let us take $\delta_n = \max(\delta_n^+, \delta_n^-)$. Then,

$$\delta_n = \frac{A}{n} + o(\frac{1}{n}),$$

where $A = \max(A^-, A^+)$. When $n \to \infty$, this expression tends to 0, hence for sufficiently large n, $\delta_n \le \min(\delta_0^-, \delta_0^+)$. Therefore, for such n, for $\delta = \delta_n$, we get $p^- \le (1-p_0)/2$ and $p^+ \le (1-p_0)/2$ and hence, $P \ge p_0$. Since we defined $\delta_n^{\rm int}$ as the smallest δ with this property, we can conclude that $\delta_n^{\rm int} \le \delta_n = (A/n) + o(1/n)$ and $\delta_n^{\rm int} = O(1/n)$. Q.E.D.

Comment. The estimates given in this proof show that increasing $\rho(\varepsilon)$ and $\rho(-\varepsilon)$, i.e., increasing the probability of observing errors close to the endpoints (of the interval assumed to contain errors) increases the accuracy of the interval estimate based on intersection. This result is in good agreement with the experimental data from [W88].

Proof of Theorems 2 and 3. Since $\vec{y}$ is the solution of a system of equations $F_j = 0$ with smooth F_j, we can use the implicit function theorem to compute the derivative of y and y_j with respect to x_i. These derivatives are expressed in terms of the matrix that is inverse to the matrix of partial derivatives $\partial F_j/\partial y_l$. We assumed that the indirect measurement system is non-degenerate, i.e., that this inverse matrix always exists. Therefore, the derivative of y with respect to x_i is always finite, i.e., y is a differentiable function of $x_1, ..., x_n$. Hence, for given actual values x_i, $\Delta y = \sum a_i \Delta x_i + o(\Delta x_i)$ for some a_i.

For averaging, as we have already seen (in 1-dimensional case), $\Delta x_i = O(1/\sqrt{n})$. From this, we would like to conclude that $\Delta y = O(1/\sqrt{n})$. To be able to make this conclusion, we must find a confidence p_0' that will be used to estimate Δx_i, and guarantee confidence level p_0 for y. This p_0' must be thus chosen in such a way that if each of n events (events meaning that the errors Δx_i are smaller than their bounds) happens with probability $\ge p_0'$, then the probability that all these events occur, is greater than p_0. Then, with probability $\ge p_0$, we will have $|\Delta y| \le O(1/\sqrt{n})$.

Finding such p_0' is easy: since all the variables $\tilde{x}_i^{(k)}$ are independent, the probability that each of n events occur is equal to the product of the correspondent probabilities, i.e., it is not smaller than $(p_0')^n$. So, if we choose p_0' in such a way that $(p_0')^n \ge p_0$, we are done. Hence, it is sufficient to take $p_0' = p_0^{1/n}$.

Similarly, in the interval case, $\Delta x_i = O(1/n)$, hence $\Delta y = O(1/n)$. Q.E.D.

CONCLUSIONS

When measurement errors are random, we can improve the accuracy of the measurement result by measuring the same quantity several times and combining the results into a better estimate.

If the only information that we have about the error is the interval of its possible values $[-\varepsilon, \varepsilon]$, then two approaches are possible. The first approach, motivated by traditionally

used statistical methods, prompts us to take an arithmetic mean $\bar{x} = (\tilde{x}_1 + ... + \tilde{x}_n)/n$ of the measurement results. The second approach, motivated by interval computations, is as follows: Each measurement $\tilde{x}_i$ means that the actual value of x belongs to an interval $[\tilde{x}_i - \varepsilon, \tilde{x}_i + \varepsilon]$. So, to get a combined estimate, we can take an intersection of these n intervals.

We prove that for large n, the accuracy of an interval estimate ($O(1/n)$) is much better than the accuracy of the arithmetic mean ($\sim 1/\sqrt{n}$).

A similar result is proved for a multi-dimensional case, when we measure several auxiliary quantities, and use the measurement results to estimate the value of the desired quantity y.

Acknowledgments. This work was supported by an NSF grant No. CDA-9015006, NASA Grants No. 9–482 and NAG 9–757, and a Grant No. PF90–018 from the General Services Administration (GSA), administered by the Materials Research Institute. We are thankful to numerous colleagues, especially to R. Baker Kearfott, for valuable discussions.

REFERENCES

[A88] O. Aberth, *Precise numerical analysis*, Wm. C. Brown Publishers, Dubuque, Iowa, 1988.

[B91] W. H. Beyer, *CRC Standard mathematical tables and formulae*, CRC Press, Boca Raton, FL, 1991.

[D91] N. P. Dyvak, "Interval models of errors and their usage in applied problems". In: *Urgent Problems of Applied Mathematics*, Proceedings of the USSR National Conference, Saratov, May 20–24, 1991, Part 1, pp. 70–75 (in Russian).

[H81] P. J. Huber, *Robust statistics*, Wiley, N.Y., 1981.

[I81] I. A. Ibragimov and R. Z. Has'minskii, *Statistical estimation. Asymptotic theory*, Springer-Verlag, N.Y., 1981.

[K96] B. B. Kearfott and V. Kreinovich, *Applications of Interval Computations*, Kluwer, Dordrecht, 1996.

[M64] R. E. Moore, *The automatic analysis and control of error in digital computation based on the use of interval numbers*, L. B. Rall (ed.), *Proc. of the Advanced Seminar*, Math. Research Center, Wiley, N.Y., 1964, pp. 61–ff.

[M79] R. E. Moore, *Methods and applications of interval analysis*, SIAM, Philadelphia, PA, 1979.

[NZ91] P. V. Novitskii and I. A. Zograph. *Estimating the measurement errors*, Energoatomizdat, Leningrad, 1991 (in Russian).

[VS89] A. P. Voshchinin, G. R. Sotirov, *Optimization in case of uncertainty*, Tehnika, Moscow, Sofia, 1989 (in Russian).

[W90] H. M. Wadsworth, Jr (editor). *Handbook of statistical methods for engineers and scientists*, McGraw-Hill Publishing Co., N.Y., 1990.

[W88] G. W. Walster, "Philosophy and practicalities of interval arithmetic", In: *Reliability in Computing*, Academic Press, N.Y., 1988, pp. 309–323.