11-1-2022

Need for Optimal Distributed Measurement of Cumulative Quantities Explains the Ubiquity of Absolute and Relative Error Components

Hector A. Reyes
The University of Texas at El Paso, hareyes2@miners.utep.edu

Aaron D. Brown
The University of Texas at El Paso, adbrown3@miners.utep.edu

Jeffrey Escamilla
The University of Texas at El Paso, jescamilla2@miners.utep.edu

Ethan D. Kish
The University of Texas at El Paso, edkish@miners.utep.edu

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

Comments:
Technical Report: UTEP-CS-22-111

Need for Optimal Distributed Measurement of Cumulative Quantities Explains the Ubiquity of Absolute and Relative Error Components

Hector A. Reyes, Aaron D. Brown, Jeffrey Escamilla, Ethan D. Kish, and Vladik Kreinovich

Abstract In many practical situations, we need to measure the value of a cumulative quantity, i.e., a quantity that is obtained by adding measurement results corresponding to different spatial locations. How can we select the measuring instruments so that the resulting cumulative quantity can be determined with known accuracy – and, to avoid unnecessary expenses, not more accurately than needed? It turns out that the only case where such an optimal arrangement is possible is when the required accuracy means selecting the upper bounds on absolute and relative error components. This results provides a possible explanation for the ubiquity of such two-component accuracy requirements.

1 Formulation of the Problem

Need for distributed measurements. In many practical situations, we are interested in estimating the value x of a cumulative quantity: e.g., we want to estimate the overall amount of oil in a given area, the overall amount of CO_2 emissions, etc.

How to perform distributed measurements. Measuring instruments usually measure quantities in a given location, i.e., they measure local values $x_1, \ldots, x_n$ that together form the desired value

$$x = x_1 + \ldots + x_n.$$

So, a natural way to produce an estimate $\widetilde{x}$ for the cumulative value x is:

- to place measuring instruments at several locations within a given area,

Hector A. Reyes, Aaron D. Brown, Jeffrey Escamilla, Ethan D. Kish, and Vladik Kreinovich
Department of Computer Science, University of Texas at El Paso, 500 W. University
El Paso, Texas 79968, USA, e-mail: hareyes2@miners.utep.edu,
jescamilla2@miners.utep.edu, adbrown3@miners.utep.edu,
edkish@miners.utep.edu, vladik@utep.edu

- to measure the values x_i of the desired quantity in these locations, and
- to add up the results $\widetilde{x}_1 + \ldots + \widetilde{x}_n$ of these measurement:

$$\widetilde{x} = \widetilde{x}_1 + \ldots + \widetilde{x}_n.$$

Need for optimal planning. Usually, we want to reach a certain estimation accuracy. To achieve this accuracy, we need to plan how accurate the deployed measurement instruments should be. Use of accurate measuring instruments is often very expensive, while budgets are usually limited. It is therefore desirable to come up with the deployment plan that would achieve the desired overall accuracy within the minimal cost. This implies, in particular, that the resulting estimate should not be more accurate than needed – this would mean that we could use less accurate (and thus, cheaper) measuring instruments.

What we do in this paper. In this paper, we provide a condition under which such optimal planning is possible – and the corresponding optimal planning algorithm. The resulting condition will explain why usually, measuring instruments are characterized by their absolute and relative accuracy.

2 Let Us Formulate the Problem in Precise Terms

How we can describe measurement accuracy. Measurements are never absolutely accurate, the measurement result $\widetilde{x}_i$ is, in general, different from the actual (unknown) value x_i of the corresponding quantity; see, e.g., [5]. In other words, the difference $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i$ is, in general, different from 0. This difference is known as the *measurement error*.

For each measuring instrument, we should know how large the measurement error can be. In precise terms, we need to know an upper bound Δ on the absolute value $|\Delta x_i|$ of the measurement error. This upper bound should be provided by the manufacturer of the measuring instrument. Indeed, if no such upper is known, this means that whatever the reading of the measuring instrument, the actual value can be as far away from this reading as possible, so we get no information whatsoever about the actual value – in this case, this is not a measuring instrument, it is a wild guess.

Ideally, in addition to knowing that the measurement error Δx_i is somewhere in the interval $[-\Delta, \Delta]$, it is desirable to know how probable are different values from this interval, i.e., what is the probability distribution on the measurement error. Sometimes, we know this probability distribution, but in many practical situations, we don't know it, and the upper bound is all we know. So, in this section, we will consider this value as the measure of the instrument's accuracy; see, e.g., [1, 2, 3, 4].

This upper bound Δ may depend on the measured value. For example, if we are measuring current in the range from 1 mA to 1 A, then it is relatively easy to maintain accuracy of 0.1 mA when the actual current is 1 mA – this means measuring with one correct decimal digit. We can get values 0.813. . . , 0.825. . . , but since the measurement accuracy is 0.1, this means that these measurement results may correspond to the same actual value. In other words, whatever the measuring instrument shows, only one digit is meaningful and significant – all the other digits may be caused by measurement errors. On the other hand, to maintain the same accuracy of 0.1 mA when we measure currents close to 1 A would mean that we need to distinguish between values 0.94651 A = 946.51 mA and 0.94637 A = 946.37 mA, since the difference between these two values is larger than 0.1 mA. This would mean that we require that in the measurement result, we should have not one, but four significant digits – and this would require much more accurate measurements.

Because of this, we will explicitly take into account that the accuracy Δ depends on the measured value: $\Delta = \Delta(x)$. Usually, small changes in x lead to only small changes in the accuracy, so we can safely assume that the dependence $\Delta(x)$ is smooth.

What we want. We want to estimate the desired cumulative value x with some accuracy δ. In other words, we want to make sure that the difference between our estimate $\widetilde{x}$ and the actual value x does not exceed δ: $|\widetilde{x} - x| \leq \delta$.

The cumulative value is estimated based on n measurement results. As we have mentioned, the accuracy that we can achieve in each measurement, in general, depends on the measured value: the larger the value of the measured quantity, the more difficult it is to maintain the corresponding accuracy. It is therefore reasonable to conclude that, whatever measuring instruments we use to measure each value x_i, it will be more difficult to estimate the larger cumulative value x with the same accuracy. Thus, it makes sense to require that the desired accuracy δ should also depend on the value that we want to estimate $\delta = \delta(x)$: the larger the value x, the larger the uncertainty $\delta(x)$ that we can achieve.

So, our problem takes the following form:

- we want to be able to estimate the cumulative value x with given accuracy $\delta(x)$ – i.e., we are given a function $\delta(x)$ and we want to estimate the cumulative value with this accuracy;
- we want to find the measuring instruments that would guarantee this estimation accuracy – and that would be optimal for this task, i.e., that would not provide better accuracy than needed.

Let us describe what we want in precise terms. To formulate this problem in precise terms, let us analyze what estimation accuracy we can achieve if we use, for each of n measurements, the measuring instrument characterized by the accuracy $\Delta(x)$.

Based on each measurement result $\widetilde{x}_i$, we can conclude that the actual value x_i of the corresponding quantity is located somewhere in the interval

$$[\widetilde{x}_i - \Delta(x_i), \widetilde{x}_i + \Delta(x_i)] :$$

the smallest possible value is $\widetilde{x}_i - \Delta(x_i)$, the largest possible value is $\widetilde{x}_i + \Delta(x_i)$.

When we add the measurement results, we get the estimate $\widetilde{x} = \widetilde{x}_1 + \ldots + \widetilde{x}_n$ for the desired quantity x. What are the possible values of this quantity? The sum

$x = x_1 + \ldots + x_n$ attains its smallest value if all values x_i are the smallest, i.e., when

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

i.e., when

$$x = \widetilde{x} - (\Delta(x_1) + \ldots + \Delta(x_n)).$$

Similarly, the sum $x = x_1 + \ldots + x_n$ attains its largest value if all values x_i are the largest, i.e., when

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

i.e., when

$$x = \widetilde{x} + (\Delta(x_1) + \ldots + \Delta(x_n)).$$

Thus, all we can conclude about the value x is that this value belongs to the interval

$$[\widetilde{x} - (\Delta(x_1) + \ldots + \Delta(x_n)), \widetilde{x} + (\Delta(x_1) + \ldots + \Delta(x_n))].$$

This means that we get an estimate of x with the accuracy $\Delta(x_1) + \ldots + \Delta(x_n)$.

Our objective is to make sure that this is exactly the desired accuracy $\delta(x)$. In other words, we want to make sure that whenever $x = x_1 + \ldots + x_n$, we should have

$$\delta(x) = \Delta(x_1) + \ldots + \Delta(x_n).$$

Substituting $x = x_1 + \ldots + x_n$ into this formula, we get

$$\delta(x_1 + \ldots + x_n) = \Delta(x_1) + \ldots + \Delta(x_n). \quad (1)$$

We do not know a priori what will be the values x_i, so if we want to maintain the desired accuracy $\delta(x)$ – and make sure that we do not get more accuracy – we should make sure that the equality (1) be satisfied for all possible values $x_1, \ldots, x_n$.

In these terms, the problem takes the following form:

- For which functions $\delta(x)$ is it possible to have a function $\Delta(x)$ for which the equality (1) is satisfied? and
- For the functions $\delta(x)$ for which such function $\Delta(x)$ is possible, how can we find this function $\Delta(x)$ – that describes the corresponding measuring instrument?

This is the problem that we solve in this paper.

3 When Is Optimal Distributive Measurement of Cumulative Quantities Possible?

Let us first analyze when the optimal distributive measurement of a cumulative quantity is possible, i.e., for which functions $\delta(x)$, there exists a function $\Delta(x)$ for which the equality (1) is always satisfied.

We have assumed that the function $\Delta(x)$ is smooth, i.e., differentiable. Thus, the sum $\delta(x)$ of such functions is differentiable too. Since both functions $\Delta(x)$ and $\delta(x)$ are differentiable, we can differentiate both sides of the equality (1) with respect to one of the variables – e.g., with respect to the variable x_1. The terms $\Delta(x_1), \ldots, \Delta(x_n)$ do not depend on x_1 at all, so their derivative with respect to x_1 is 0, and the resulting formula takes the form

$$\delta'(x_1 + \ldots + x_n) = \Delta'(x_1), \tag{2}$$

where, as usual, δ' and Δ' denote the derivatives of the corresponding functions.

The equality (2) holds for all possible values $x_2, \ldots, x_n$. For every real number x_0, we can take, e.g., $x_2 = x_0 - x_1$ and $x_3 = \ldots + x_n = 0$, then we will have $x_1 + \ldots + x_n = x_0$, and the equality (2) takes the form

$$\delta'(x_0) = \Delta'(x_1).$$

The right-hand side does not depend on x_0, which means that the derivative $\delta'(x_0)$ is a constant not depending on x_0 either.

The only functions whose derivative is a constant are linear functions, so we conclude that the dependence $\delta(x)$ is linear:

$$\delta(x) = a + b \cdot x$$

for some constants a and b.

Interestingly, this fits well with the usual description of measurement error [5], as consisting of two components:

- the absolute error component a that does not depend on x at all, and
- the relative error component – according to which, the bound on the measurement error is a certain percentage of the actual value x, i.e., has the form $b \cdot x$ for some constant b (e.g., for 10% accuracy, $b = 0.1$).

Thus, our result explains this usual description.

4 What Measuring Instrument Should We Select to Get the Optimal Distributive Measurement of Cumulative Quantity?

Now that we know for what desired accuracy $\delta(x)$, we can have the optimal distributive measurement of a cumulative quantity, the natural next question is: given one of such functions $\delta(x)$, what measuring instrument – i.e., what function $\Delta(x)$ – should we select for this optimal measurement?

To answer this question, we can take $x_1 = \ldots = x_n$. In this case, $\Delta(x_1) = \ldots = \Delta(x_n)$, so the equality (2) takes the form

$$\delta(n \cdot x_1) = n \cdot \Delta(x_1). \tag{3}$$

We know that $\delta(x) = a + b \cdot x$, so the formula (3) takes the form

$$a + b \cdot n \cdot x_1 = n \cdot \Delta(x_1).$$

If we divide both sides of this equality by x_1, and rename x_1 into x, we get the desired expression for $\Delta(x)$:

$$\Delta(x) = \frac{a}{n} + b \cdot x.$$

In other words:

- the bound on the relative error component of each measuring instrument should be the same as the desired relative accuracy of the cumulative quantity, and
- the bound on the absolute error component should be n times smaller than the desired bound on the absolute accuracy of the cumulative quantity.

Acknowledgments

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

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

The authors are thankful to the participants of the 2022 UTEP/NMSU Workshop on Mathematics, Computer Science, and Computational Science (El Paso, Texas, November 5, 2022) for valuable discussions.

References

1. L. Jaulin, M. Kiefer, O. Didrit, and E. Walter, *Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control, and Robotics*, Springer, London, 2001.
2. B. J. Kubica, *Interval Methods for Solving Nonlinear Contraint Satisfaction, Optimization, and Similar Problems: from Inequalities Systems to Game Solutions*, Springer, Cham, Switzerland, 2019.
3. G. Mayer, *Interval Analysis and Automatic Result Verification*, de Gruyter, Berlin, 2017.
4. R. E. Moore, R. B. Kearfott, and M. J. Cloud, *Introduction to Interval Analysis*, SIAM, Philadelphia, 2009.
5. S. G. Rabinovich, *Measurement Errors and Uncertainty: Theory and Practice*, Springer Verlag, New York, 2005.