

How to Separate Absolute and Relative Error Components: Interval Case

C. Servin

*Computer Science and Information Technology Systems Department
El Paso Community College (EPCC)
919 Hunter Dr. Room A-1720
El Paso, TX 79915, USA
E-mail: cservin1@epcc.edu*

O. Kosheleva

*Department of Teacher Education
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
E-mail: olgak@utep.edu*

V. Kreinovich

*Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
E-mail: vladik@utep.edu*

Usually, measurement errors contain both absolute and relative components. To correctly gauge the amount of measurement error for all possible values of the measured quantity, it is important to separate these two error components. For probabilistic uncertainty, this separation can be obtained by using traditional probabilistic techniques. The problem is that in many practical situations, we do not know the probability distribution, we only know the upper bound on the measurement error. In such situations of interval uncertainty, separation of absolute and relative error components is not easy. In this paper, we propose a technique for such a separation based on the maximum entropy approach, and we provide feasible algorithms – both sequential and parallel – for the resulting separation.

Keywords: absolute error, relative error, interval uncertainty, Maximum Entropy approach

1. Need to Separate Absolute and Relative Error Components: General Case

Measurement uncertainty: a brief reminder. Most information about the world comes from measurements. Computers that process this information treat all the numerical inputs as exact numbers, but, of course, measurements are never 100% accurate. The measurement results \tilde{x} is, in general, different from the actual (unknown) value x of the measured quantity: in other words, the *measurement error* $\Delta x \stackrel{\text{def}}{=} \tilde{x} - x$ is, in general, different from 0; see, e.g.,⁹.

Because of this measurement uncertainty, the results of data processing are also, in general, different from what we would get if we knew the exact values of the corresponding quantities. To make adequate decisions based on the results of data processing, we need to gauge this uncertainty. For this, we need to know how accurate were the corresponding measurements.

How measurement uncertainty is usually described. Situations change. As a result, in slightly different situations, when we use the same measuring instrument several times to measure the same quantity, in general, we get slightly different results. We cannot predict which result we will get for each measurement. At best, what we can do is find the frequency with which different measurement errors appear in such a measurement. In other words, at best, we can determine, for each measured value x , the probability distribution $\rho_x(\Delta x)$ that describes the probability of different values of the measurement error Δx .

In many cases, the measurement error is the result of a joint action of the large number of small independent factors. In such cases, the measurement error is the sum of several independent variables of about the same size. It is known – this is called the Central Limit Theorem – that the distribution of such a sum is close to Gaussian; see, e.g.,¹⁰. Gaussian distributions are indeed frequently observed in metrology; see, e.g.,⁹.

In general, a Gaussian distribution is described by two parameters:

- its mean and
- its standard deviation σ .

If the mean is different from 0 – i.e., if we have a *bias* – then we can simply re-calibrate this measurement instrument by subtracting this bias from all the measurement results. Thus, we can safely assume that the bias is 0, and that the only characteristic describing the measurement error of measuring

the quantity x is the standard deviation $\sigma(x)$.

Case of interval uncertainty. To find (and compensate for) the bias and to find the standard deviation $\sigma(x)$, we need several times to compare:

- the results of a measurement performed by the analyzed measuring instrument and
- the results of measuring the same quantity with a much more accurate (“standard”) measuring instrument.

Such standard instruments are expensive. Thus, determining the mean and standard deviation is usually a very expensive and time-consuming procedure. Because of this, in many practical situations, this procedure is not performed. In such situations, the only information that we have about the measurement error Δx is the upper bound $\Delta(x)$ on its absolute value:

$$|\Delta x| = |\tilde{x} - x| \leq \Delta(x).$$

This upper bound has to be provided by the manufacturer of the measuring instrument. (Indeed, if no such upper bound is provided, then, based on the measurement results, we cannot say anything at all about the actual value, it can be arbitrarily large or it can be arbitrarily small. In other words, this would be a wild guess, not a measurement.)

The upper bound is obtained by the manufacturer of the measuring instrument by analyzing the whole bunch of similar mass-produced measuring instruments, with different biases and different values of other characteristics. This is a much cheaper and much less time-consuming process than individual calibration of each individual measuring instrument – which is needed for the probability uncertainty description. The use of this process allows the manufacturer to produce relatively cheap and affordable sensors and other measuring instruments.

In this case, once we know the measurement result \tilde{x} , what can we conclude about the measurement error? The dependence of $\Delta(x)$ on x is usually smooth: relatively small changes of x lead to relatively small changes in $\Delta(x)$. Measurements are usually reasonably accurate, so the measurement result \tilde{x} is very close to the actual value x . Thus, the value $\Delta(x)$ is very close to the value $\Delta(\tilde{x})$. The relative difference $\frac{\Delta(\tilde{x}) - \Delta(x)}{\Delta(x)}$ between these two bounds is proportional to the relative difference $\frac{\tilde{x} - x}{x} = \frac{\Delta x}{x}$. Thus, the absolute difference between these two bounds is of the order

$(\Delta(x))^2$ – quadratic in terms of $\Delta(x)$:

$$\Delta(\tilde{x}) - \Delta(x) = \frac{\Delta(\tilde{x}) - \Delta(x)}{\Delta(x)} \cdot \Delta(x) \sim \frac{\Delta x}{x} \cdot \Delta(x) \sim$$

$$\frac{\Delta(x)}{x} \cdot \Delta(x) = \frac{1}{x} \cdot (\Delta(x))^2.$$

Thus, this difference is much smaller than $\Delta(x)$. So, from the practical viewpoint, this difference can be safely ignored. For example, if the measurement error is 5%, then its square is 0.25%, and is indeed negligible in comparison: e.g., no one says that an instrument has accuracy 5.25%.

Since, from the practical viewpoint, we have $\Delta(\tilde{x}) \approx \Delta(x)$, the above inequality takes the form $|\tilde{x} - x| \leq \Delta(\tilde{x})$. Thus, once we know the measurement result \tilde{x} , the only information that we gain about the actual (unknown) value x of the measured quantity is that this value must satisfy the inequality

$$\tilde{x} - \Delta(\tilde{x}) \leq x \leq \tilde{x} + \Delta(\tilde{x}), \quad (1)$$

i.e., that this value x is in the interval $[\tilde{x} - \Delta(\tilde{x}), \tilde{x} + \Delta(\tilde{x})]$. Because of this, situations when we only know the upper bounds on the measurement errors are known as situations of *interval uncertainty*; see, e.g.,^{2,5,6}.

Need for extrapolation and interpolation. In practice, during any time interval, we can only perform a finite number of tests. In these tests, we estimate the values $\sigma(x_i)$ or $\Delta(x_i)$ for finitely many different values

$$x_1 < x_2 < \dots < x_n.$$

In practice, we will encounter other values of the quantity x . We therefore need to extrapolate and/or interpolate the estimates $\sigma(x_i)$ or $\Delta(x_i)$ to other values of the quantity x .

How to extrapolate and interpolate. The dependence of $\sigma(x)$ and $\Delta(x)$ on the corresponding value x is usually smooth. Thus, a natural idea is to expand this dependence in Taylor series and keep only a few first terms in this expansion. As a result, the dependence $\sigma(x)$ or $\Delta(x)$ becomes a polynomial. So:

- we can estimate the coefficients of this polynomial based on the known values $\sigma(x_i)$ or $\Delta(x_i)$, and then
- we can use the resulting polynomial formula to estimate the corresponding metrological characteristic $\sigma(x)$ or $\Delta(x)$ for all x .

Absolute and relative error components. The simplest polynomial is a linear function. In the vicinity of each value, each smooth function can be reasonably well approximated by a linear function; see, e.g.,¹. So, for some reasonable-size range, can safely use linear expressions $\sigma(x) = \sigma_0 + \sigma_1 \cdot x$ and $\Delta(x) = \Delta_0 + \Delta_1 \cdot x$.

What is meaning of the two terms in this expansion?

- When we have only the constant term, i.e., when $\sigma_1 = 0$ or $\Delta_1 = 0$, the corresponding characteristic of the measurement error is *absolute* in the sense that it does not change – it is the same for all the values of the measured quantity x . This error component is therefore known as the *absolute* error component.
- When we only have the linear term, i.e., when $\sigma_0 = 0$ or $\Delta_0 = 0$, then the absolute value of the measurement-error characteristic changes, but the *relative* error $\frac{\sigma(x)}{x}$ or $\frac{\Delta(x)}{x}$ remains constant. This component is thus called a *relative* error component.

Need to separate absolute and relative error components. It is therefore desirable to estimate the coefficients of the above linear dependence based on the known values $\sigma(x_i)$ or $\Delta(x_i)$. In other words, we need to *separate* the overall measurement errors into absolute and relative error components.

Comment. In general, the larger the value of the measured quantity, the larger the measurement error. However, we do not want to exclude the opposite situations, when larger values of the quantity can be measured with higher accuracy – and thus, with smaller measurement errors. In such situations, the corresponding coefficient σ_1 or Δ_1 is negative.

2. How to Separate Absolute and Relative Error Components: Probabilistic Case

Towards formal description of the problem. Suppose that, for each of several different values x_i , we perform several (s_i) measurements and get several values \tilde{x}_{ik} , $k = 1, 2, \dots, s_i$. In line with the above description, we assume that all the measurement errors are normally distributed with mean 0 and standard deviation $\sigma_0 + \sigma_1 \cdot x_i$, for some unknown values x_i , σ_0 , and σ_1 , and that measurement errors corresponding to different measurements are independent.

In this case, each value $\Delta x_{ik} = \tilde{x}_{ik} - x_i$ of the measurement error is normally distributed with mean 0 and standard deviation $\sigma_0 + \sigma_1 \cdot x_i$. Thus,

the corresponding probability density has the form

$$\frac{1}{\sqrt{2\pi} \cdot (\sigma_0 + \sigma_1 \cdot x_i)} \cdot \exp\left(-\frac{(\tilde{x}_{ik} - x_i)^2}{2(\sigma_0 + \sigma_1 \cdot x_i)^2}\right).$$

Since all measurement errors are independent, the overall probability density is equal to the product of these expressions:

$$\prod_i \prod_{k=1}^{s_i} \frac{1}{\sqrt{2\pi} \cdot (\sigma_0 + \sigma_1 \cdot x_i)} \cdot \exp\left(-\frac{(\tilde{x}_{ik} - x_i)^2}{2(\sigma_0 + \sigma_1 \cdot x_i)^2}\right). \quad (2)$$

A natural idea to select the values of the corresponding parameters x_i , σ_0 , and σ_1 which is the most probable, i.e., for which the corresponding probability (2) is the largest possible. This is known as the *Maximum Likelihood Method*; see, e.g.,¹⁰. Thus, we arrive at the following solution.

Solution. Based on the measurement results \tilde{x}_{ik} , we find the values x_i , σ_0 , and σ_1 that maximize the expression (2).

3. How to Separate Absolute and Relative Error

Components: Interval Case

Towards a precise formulation of the problem: what we want and what we have. We want to find the expression $\Delta(x) = \Delta_0 + \Delta_1 \cdot x$ that describes the desired upper bounds on the measurement error for all the values x from some range $[\underline{x}, \bar{x}]$.

To find this expression, we can use the values $\Delta(x_i)$ provided for several values $\underline{x} < x_1 < x_2 < \dots < x_n < \bar{x}$ of the measured quantity x .

Important conditions. We want to make sure that we do not underestimate the measurement errors. So, for all i , we must have

$$\Delta(x_i) \leq \Delta_0 + \Delta_1 \cdot x_i. \quad (3)$$

Another constraint is that the value $\Delta(x) = \Delta_0 + \Delta_1 \cdot x$ must be non-negative for all values x from the interval $[\underline{x}, \bar{x}]$. It is easy to see that this condition is equivalent to requiring that this value is non-negative at the two endpoints of this interval, i.e., that we have:

$$0 \leq \Delta_0 + \Delta_1 \cdot \underline{x} \quad (4)$$

and

$$0 \leq \Delta_0 + \Delta_1 \cdot \bar{x}. \quad (5)$$

These conditions are not sufficient. These conditions are necessary, but not sufficient to determine the desired values Δ_0 and Δ_1 . Indeed, if we have some values Δ_0 and Δ_1 that satisfy inequalities (3)–(5), and we increase the value Δ_0 , then all these inequalities will still be preserved.

So, we need to have a criterion that would allow us to select one pair of the values Δ_0 and Δ_1 from all the pairs that satisfy the above inequalities.

Idea: use Maximum Entropy Approach. In general, when we make a selection, we decrease the uncertainty of the situation: instead of all possible dependencies $\Delta(x)$ which are consistent with the known values $\Delta(x_i)$, we select one specific dependence. A natural idea is to not invent certainty where there is none, i.e., to select an alternative that maximally preserves the original uncertainty.

In the statistical approach, uncertainty of the probability distribution with the probability density $\rho(x)$ is naturally measured by entropy

$$S = - \int \rho(x) \cdot \ln(\rho(x)) dx,$$

which is proportional to the smallest average number of binary (yes-no) questions that we need to ask to determine the value of the random variable with given accuracy; see, e.g.,^{3,7}. So, the above natural idea leads to maximum entropy approach – when from all possible distributions, we select the one with the largest entropy.

Interval uncertainty corresponds to a *class* of distributions – namely, the class of all the distributions located on the given interval. The average number of binary questions depends on the distribution. In this case, a natural measure of uncertainty is the largest possible average number of questions – i.e., largest possible entropy of all the distributions from this class. We will call this value the *entropy* of the given class of probability distributions.

Thus, a reasonable idea is to select the values Δ_0 and Δ_1 for which thus defined entropy is the closest to the original entropy value.

From the idea to the exact optimization criterion. What is the entropy in the interval case?

We consider n measurements corresponding to n different values x_1, \dots, x_n . To provide a full description of each combination of measurement outcomes, we need to describe n corresponding measurement errors Δx_i . Thus, to get a complete description of measurement accuracy, we

need to provide a probability distribution of the set of the corresponding tuples $(\Delta x_1, \dots, \Delta x_n)$.

Originally, for each i , the only information that we have about each measurement error Δx_i is that it belongs to the interval $[-\Delta(x_i), \Delta(x_i)]$. Thus, the set of all possible values of the tuple is the box

$$B = [-\Delta(x_1), \Delta(x_1)] \times \dots \times [-\Delta(x_i), \Delta(x_i)] \times \dots \times [-\Delta(x_n), \Delta(x_n)].$$

So, we need to consider the class of all possible probability distributions on this box.

The entropy of this class, as we have mentioned earlier, can be described as the largest possible entropy of all the distributions located on this box. It is known (see, e.g.,³) that this largest value is attained when we consider the uniform distribution on this box, with the constant probability density $\rho(x) = \text{const}$. This constant can be determined from the condition that the overall probability be equal to 1: $\int \rho(x) dx = 1$. For the constant, this implies that $\rho \cdot V = 1$, where V denotes the volume of the box. Thus, the probability density is equal to $\rho = \frac{1}{V}$. So, the entropy takes the form

$$\begin{aligned} S &= - \int \rho(x) \cdot \ln(\rho(x)) dx = - \int \frac{1}{V} \cdot \ln\left(\frac{1}{V}\right) dx = \\ &= -V \cdot \frac{1}{V} \cdot \ln\left(\frac{1}{V}\right) = -\ln\left(\frac{1}{V}\right). \end{aligned}$$

Since the logarithm of the ratio is equal to the difference between logarithms, we conclude that

$$\ln\left(\frac{1}{V}\right) = \ln(1) - \ln(V) = -\ln(V),$$

thus $S = \ln(V)$.

The volume of the box is equal to the product of the lengths $2\Delta(x_i)$ of its sides:

$$V = \prod_{i=1}^n (2\Delta(x_i)) = 2^n \cdot \prod_{i=1}^n \Delta(x_i).$$

The logarithm of the product is equal to the sum of the logarithms, so we conclude that the original entropy is equal to

$$S = \ln(V) = n \cdot \ln(2) + \sum_{i=1}^n \ln(\Delta(x_i)).$$

When we replace each value $\Delta(x_i)$ with a linear expression $\Delta_0 + \Delta_1 \cdot x_i$, we similarly get the values

$$V' = \prod_{i=1}^n (2(\Delta_0 + \Delta_1) \cdot x_i) = 2^n \cdot \prod_{i=1}^n (\Delta_0 + \Delta_1 \cdot x_i), \quad (6)$$

and

$$S' = \ln(V') = n \cdot \ln(2) + \sum_{i=1}^n \ln(\Delta_0 + \Delta_1 \cdot x_i). \quad (7)$$

We want to find the values Δ_0 and Δ_1 for which the entropy S' is the closest to the original entropy S .

For each i , due to inequality (2), we have $0 \leq \Delta(x_i) \leq \Delta_0 + \Delta_1 \cdot x_i$, so the product V of the left-hand sides is always smaller than or equal to the product V' of the right-hand sides. Thus, we always have

$$S = \ln(V) \leq S' = \ln(V').$$

So, selecting the value S' which is the closest to S means selecting the smallest possible value S' .

Adding a constant $n \cdot \ln(2)$ to all the values of the entropy does not change which value is larger. Thus, finding the values Δ_0 and Δ_1 that minimize the entropy S' is equivalent to finding the values Δ_i for which the sum

$$s \stackrel{\text{def}}{=} \sum_{i=1}^n \ln(\Delta_0 + \Delta_1 \cdot x_i) \quad (8)$$

attains the smallest possible value.

So, we arrive at the following exact formulation of the problem.

Formulation of the problem in precise terms. Among all the pairs of values Δ_0 and Δ_1 that satisfy the inequalities (3)–(5), we need to find the values Δ_0 and Δ_1 for which the expression (8) attains the smallest possible value.

Towards the algorithm. First, to somewhat simplify the formulation of the problem, let us notice that the conditions (4) and (5) can be formulated in the form (3).

- For the condition (4), it is sufficient to take $x_0 = \underline{x}$ and $\Delta(x_0) = 0$.
- For the condition (5), it is sufficient to take $x_{n+1} = \bar{x}$ and $\Delta(x_{n+1}) = 0$.

So, instead of the conditions (3)–(5), we can consider the condition (3) for $i = 0, 1, \dots, n, n+1$.

Second, let us take into account that since we have constraints of the non-strict inequality type, then for the minimizing pair of Δ_0 and Δ_1 , some inequalities will be equalities and some strict inequalities.

When at least two inequalities become strict equalities, e.g., equalities j and k , i.e., when $\Delta_0 + \Delta_1 \cdot x_j = \Delta(x_j)$ and $\Delta_0 + \Delta_1 \cdot x_k = \Delta(x_k)$, then we have two linear equations for the two unknowns Δ_0 and Δ_1 . From these two equations, we can find the values of both unknowns:

$$\Delta_1 = \frac{\Delta(x_k) - \Delta(x_j)}{x_k - x_j} \quad (9)$$

and

$$\Delta_0 = \frac{x_j \cdot \Delta(x_k) - x_k \cdot \Delta(x_j)}{x_k - x_j}. \quad (10)$$

The objective function (8) is smooth. Thus, if none of the inequalities are equalities (or if only one inequality is an equality), then the value of the objective function is a (local or global) minimum of the objective function (or a local or global minimum under a single constraint). In both cases, this minimum is minus infinity – corresponding to the case when one of the values $\Delta_0 + \Delta_1 \cdot x_i$ is 0 and thus, its logarithm is equal $s = -\infty$. This cannot be the solution to our problem since, due to inequalities (3), we have $\ln(\Delta(x_i)) \leq \ln(\Delta_0 + \Delta_1 \cdot x_i)$ hence $\sum_{i=1}^n \ln(\Delta(x_i)) \leq s$. Here, $\Delta(x_i) > 0$ for all i , so $-\infty < \ln(\Delta(x_i))$ and $-\infty < s$.

Thus, at least two inequalities (3) must be equalities. Hence, the minimizing values Δ_0 and Δ_1 are determined by the formulas (9) and (10) corresponding to some indices j and k . So, we arrive at the following algorithm.

Resulting algorithm. To find the best values of Δ_0 and Δ_1 , we do the following. For each pair of indices $0 \leq j < k \leq n+1$:

- First, we use the formulas (9) and (10) to compute the corresponding values Δ_0 and Δ_1 .
- Then, we check whether the resulting values Δ_0 and Δ_1 , the inequality (3) is satisfied for all $i = 0, 1, \dots, n+1$.
- If the inequality (3) is satisfied for all i , then we compute the value (8).

We then compare all the resulting values of the expression (8) and return the values Δ_0 and Δ_1 for which the expression (8) attains its smallest possible value.

This algorithm is feasible. For each of $O(n^2)$ pairs (j, k) :

- we check $O(n)$ inequalities;
- if inequalities are satisfied, we compute the sum (8) of n numbers;
- then, we compare the sum (8) with largest-sum-so-far, and if the new sum is larger, replace the largest-so-far with this sum.

So, for each of $O(n^2)$ pairs, we need $O(n)$ computational steps, to the total of $O(n^2) \times O(n) = O(n^3)$. Cubic-time algorithms are feasible (see, e.g.,^{4,8}), so we indeed have a feasible algorithm.

This algorithm is parallelizable. If we have an unlimited number of processors, then operations corresponding to each pair (j, k) can be performed on a separate group of processors. For each group, checking all inequalities can be done in parallel by $n + 2$ processors in one step. So, overall, we need $O(n^2) \cdot O(n) = O(n^3)$ processors.

Computing the sum can be done, as usual: in one step we add pairs of values, then add sums of pairs into sums of 4 values, etc. At the end, in $O(\log(n))$ steps, we compute the sum. Thus, the overall time for processing each pair in parallel is $O(\log(n))$.

Now, we have $O(n^2)$ values, we need to find the smallest of these values. This can also be done in $O(\log(n^2)) = O(\log(n))$ steps: first, we find the minimum of pairs, then minima of groups of 4, etc. So overall, we need $O(\log_2(n))$ time on $O(n^3)$ processors. The fact that this problem can be solved in polylog (even log) time on polynomial number of processors means that our problem belongs to the class NC of parallelizable problems; see, e.g.,⁸.

Comment. In this paper, we concentrated on the case when we approximate the dependence $\Delta(x)$ by a linear functions – i.e., by the first two terms in the Taylor expression. Similar ideas and algorithms can be used if we approximate it by a quadratic, cubic, and other fragments of the Taylor series. In this case, instead of pairs (j, k) , we will have to consider triples, quadruples, etc. The algorithm will be more time-consuming but still feasible.

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).

References

1. R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics* (Addison Wesley, Boston, Massachusetts, 2005).
2. 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).
3. E. T. Jaynes and G. L. Bretthorst, *Probability Theory: The Logic of Science* (Cambridge University Press, Cambridge, UK, 2003).
4. V. Kreinovich, A. Lakeyev, J. Rohn, and P. Kahl, *Computational Complexity and Feasibility of Data Processing and Interval Computations* (Kluwer, Dordrecht, 1998).
5. G. Mayer, *Interval Analysis and Automatic Result Verification* (de Gruyter, Berlin, 2017).
6. R. E. Moore, R. B. Kearfott, and M. J. Cloud, *Introduction to Interval Analysis* (SIAM, Philadelphia, 2009).
7. H. T. Nguyen, V. Kreinovich, B. Wu, and G. Xiang, *Computing Statistics under Interval and Fuzzy Uncertainty* (Springer Verlag, Berlin, Heidelberg, 2012).
8. C. Papadimitriou, *Computational Complexity* (Addison-Wesley, Reading, Massachusetts, 1994).
9. S. G. Rabinovich, *Measurement Errors and Uncertainties: Theory and Practice* (Springer, New York, 2005).
10. D. J. Sheskin, *Handbook of Parametric and Non-Parametric Statistical Procedures* (Chapman & Hall/CRC, London, UK, 2011).