

How to Gauge Accuracy of Measurements and of Expert Estimates: Beyond Normal Distributions

Christian Servin^{1,2}, Aline Jaimes², Craig Tweedie², Aaron Velasco²,
Omar Ochoa², and Vladik Kreinovich²

¹Information Technology Department, El Paso Community College,
El Paso, TX 79915, USA

²Cyber-ShARE Center, University of Texas at El Paso, El Paso, TX 79968, USA,
christians@miners.utep.edu, ajaimes@mines.utep.edu, ctweedie@utep.edu,
aavelasco@utep.edu, omar@miners.utep.edu, vladik@utep.edu

Abstract. To properly process data, we need to know the accuracy of different data points, i.e., accuracy of different measurement results and expert estimates. Often, this accuracy is not given. For such situations, we describe how this accuracy can be estimated based on the available data.

1 Formulation of the Problem

Need to gauge accuracy. To properly process data, it is important to know the accuracy of different data values, i.e., the accuracy of different measurement results and expert estimates; see, e.g., [3–5]. In many cases, this accuracy information is available, but in many other practical situations, we do not have this information. In such situations, it is necessary to extract this accuracy information from the data itself.

Extracting uncertainty from data: traditional approach. The usual way to gauge of the uncertainty of a measuring instrument is to compare the result \tilde{x} produced by this measuring instruments with the result \tilde{x}_s of measuring the same quantity x by a much more accurate (“standard”) measuring instrument.

Since the “standard” measuring instrument is much more accurate than the instrument that we are trying to calibrate, we can safely ignore the inaccuracy of its measurements and take \tilde{x}_s as a good approximation to the actual value x . In this case, the difference $\tilde{x} - \tilde{x}_s$ between the measurement results can serve as a good approximation to the desired measurement accuracy $\Delta x = \tilde{x} - x$.

Traditional approach cannot be applied for calibrating state-of-the-art measuring instruments. The above traditional approach works well for many measuring instruments. However, we cannot apply this approach for calibrating state-of-the-art instrument, because these instruments are the best we have. There are no other instruments which are much more accurate than these ones – and which can therefore serve as standard measuring instruments for our calibration.

Such situations are ubiquitous; for example:

- in the environmental sciences, we want to gauge the accuracy with which the Eddy covariance tower measure the Carbon and heat fluxes; see, e.g., [1];
- in the geosciences, we want to gauge how accurately seismic [2], gravity, and other techniques reconstruct the density at different depths and different locations.

How state-of-the-art measuring instruments are calibrated: case of normally distributed measurement errors. Calibration of state-of-the-art measuring instruments is possible if we make a usual assumption that the measurement errors are normally distributed with mean 0. Under this assumption, to fully describe the distribution of the measurement errors, it is sufficient to estimate the standard deviation σ of this distribution.

There are two possible approaches for estimating this standard deviation. The first approach is applicable when we have several similar measuring instruments. For example, we can have two nearby towers, or we can bring additional sensors to the existing tower. In such a situation, instead of a single measurement result \tilde{x} , we have two different results $\tilde{x}^{(1)}$ and $\tilde{x}^{(2)}$ of measuring the same quantity x . Here, by definition of the measurement error, $\tilde{x}^{(1)} = x + \Delta x^{(1)}$ and $\tilde{x}^{(2)} = x + \Delta x^{(2)}$ and therefore, $\tilde{x}^{(1)} - \tilde{x}^{(2)} = \Delta x^{(1)} - \Delta x^{(2)}$.

Each of the random variables $\Delta x^{(1)}$ and $\Delta x^{(2)}$ is normally distributed with mean 0 and (unknown) standard deviation σ (i.e., variance σ^2). Since the two measuring instruments are independence, the corresponding random variables $\Delta x^{(1)}$ and $\Delta x^{(2)}$ are also independent, and so, the variance of their difference is equal to the sum of their variances $\sigma^2 + \sigma^2 = 2\sigma^2$. Thus, the standard deviation σ' of this difference is equal to $\sqrt{2} \cdot \sigma$. We can estimate this standard deviation σ' based on the observed differences $\tilde{x}^{(1)} - \tilde{x}^{(2)}$ and therefore, we can estimate σ as $\frac{\sigma'}{\sqrt{2}}$.

This approach is not applicable in the geosciences applications, when we usually have only one seismic map, only one gravity map, etc. In such situations, we have several measurement results $\tilde{x}^{(i)}$ with, in general, different standard deviations $\sigma^{(i)}$. For every two measuring instruments i and j , the difference $\tilde{x}^{(i)} - \tilde{x}^{(j)}$ is normally distributed with the variance $(\sigma^{(i)})^2 + (\sigma^{(j)})^2$. By comparing actual measurement results, we can estimate this variance and thus, get an estimate e_{ij} for the sum. As a result, e.g., for the case when we have three different measuring instruments, we get three values e_{ij} for which:

$$e_{12} = (\sigma^{(1)})^2 + (\sigma^{(2)})^2; \quad e_{13} = (\sigma^{(1)})^2 + (\sigma^{(3)})^2;$$

$$e_{23} = (\sigma^{(2)})^2 + (\sigma^{(3)})^2.$$

Here, we have a system of three linear equations with three unknowns, from which we can uniquely determined all three desired variances $(\sigma^{(i)})^2$:

$$(\sigma^{(1)})^2 = \frac{e_{12} + e_{13} - e_{23}}{2}; \quad (\sigma^{(2)})^2 = \frac{e_{12} + e_{23} - e_{13}}{2};$$

$$\left(\sigma^{(3)}\right)^2 = \frac{e_{13} + e_{23} - e_{12}}{2}.$$

Need to go beyond normal distributions, and resulting problem. In practice, the distribution of measurement errors is often different from normal; this is the case, e.g., in measuring fluxes [1]. In such cases, we can still use the same techniques to find the standard deviation of the measurement error. However, in general, it is not enough to know the standard deviation to uniquely determine the distribution: e.g., we may have (and we sometimes do have) an asymmetric distribution, for which the skewness is different from 0 (i.e., equivalently, the expected value of $(\Delta x)^3$ is different from 0).

It is known that in this case, in contrast to the case of the normal distribution, we cannot uniquely reconstruct the distribution of Δx from the known distribution of the difference $\Delta x^{(1)} - \Delta x^{(2)}$. Indeed, if we have an asymmetric distribution for Δx , i.e., a distribution which is not invariant under the transformation $\Delta x \rightarrow -\Delta x$, this means that the distribution for $\Delta y \stackrel{\text{def}}{=} -\Delta x$ is different from the distribution for Δx . However, since $\Delta y^{(1)} - \Delta y^{(2)} = \Delta x^{(2)} - \Delta x^{(1)}$, the y -difference is also equal to the difference between two independent variables with the distribution Δx and thus, distribution for the difference $\Delta y^{(1)} - \Delta y^{(2)}$ is exactly the same as for the difference $\Delta x^{(1)} - \Delta x^{(2)}$. In other words, if we know the distribution for the difference $\Delta x^{(1)} - \Delta x^{(2)}$, we cannot uniquely reconstruct the distribution for Δx , because, in addition to the original distribution for Δx , all the observations are also consistent with the distribution for $\Delta y = -\Delta x$.

This known non-uniqueness naturally leads to the following questions:

- first, a theoretical question: since we cannot uniquely reconstruct the distribution for Δx , what information about this distribution can we reconstruct?
- second, a practical question: for those characteristics of Δx which can be theoretically reconstructed, we need to design computationally efficient algorithms for reconstructing these characteristics.

2 Technique for Solving the Problem

Technique to use. To solve these questions, let us use the Fourier analysis technique.

What we want to find is the probability density $\rho(z)$ describing the distribution of the measurement error $z \stackrel{\text{def}}{=} \Delta x$. In order to find the unknown probability density, we will first find its Fourier transform $F(\omega) = \int \rho(z) \cdot e^{i\omega \cdot z} dz$. By definition, this Fourier transform is equal to the mathematical expectation of the function $e^{i\omega \cdot z}$: $F(\omega) = E[e^{i\omega \cdot z}]$. Such a mathematical expectation is also known as a *characteristic function* of the random variable z .

Based on the observed values of the difference $z^{(1)} - z^{(2)}$, we can estimate the characteristic function $D(\omega)$ of this difference: $D(\omega) = E[e^{i\omega \cdot (z^{(1)} - z^{(2)})}]$.

Here,

$$e^{i\omega \cdot (z^{(1)} - z^{(2)})} = e^{(i\omega \cdot z^{(1)}) + (-i\omega \cdot z^{(2)})} = e^{i\omega \cdot z^{(1)}} \cdot e^{-i\omega \cdot z^{(2)}}.$$

Measurement errors $z^{(1)}$ and $z^{(2)}$ corresponding to two measuring instruments are usually assumed to be independent. Thus, the variables $e^{i\omega \cdot z^{(1)}}$ and $e^{-i\omega \cdot z^{(2)}}$ are also independent. It is known that the expected value of the product of two independent variables is equal to the product of their expected values, thus,

$$D(\omega) = E \left[e^{i\omega \cdot z^{(1)}} \right] \cdot E \left[e^{-i\omega \cdot z^{(2)}} \right],$$

i.e., $D(\omega) = F(\omega) \cdot F(-\omega)$. Here, $F(-\omega) = E \left[e^{-i\omega \cdot z} \right] = E \left[\left(e^{i\omega \cdot z} \right)^* \right]$, where t^* means complex conjugation, i.e., an operation that transforms $t = a + b \cdot i$ into $t^* = a - b \cdot i$. Thus, $F(-\omega) = F^*(\omega)$, and the above formula takes the form

$$D(\omega) = F(\omega) \cdot F^*(\omega) = |F(\omega)|^2.$$

In other words, the fact that we know $D(\omega)$ means that we know the absolute value (modulus) of the complex-valued function $F(\omega)$.

In these terms, the problems becomes: how can we reconstruct the complex-valued function $F(\omega)$ if we only know its absolute value?

3 Is It Possible to Estimate Accuracy?

How to use Fourier techniques to solve the theoretical question. First, let us address the theoretical question: since, in general, we cannot reconstruct $\rho(z)$ (or, equivalently, $F(\omega)$) uniquely, what information about $\rho(z)$ (and, correspondingly, about $F(\omega)$) can we reconstruct?

To solve this theoretical question, let us take into account the practical features of this problem. First, it needs to be mentioned that, from the practical viewpoint, we need to take into account that the situation in, e.g., Eddy covariance tower measurements is more complex than we described, because the tower does not measure *one* single quantity, it simultaneously measuring *several* quantities: carbon flux, heat flux, etc. Since these different measurements are based on data from the same sensors, it is reasonable to expect that the resulting measurement errors are correlated. Thus, to fully describe the measurement uncertainty, it is not enough to describe the distribution of each 1-D measurement error, we need to describe a joint distribution of all the measurement errors $z = (z_1, z_2, \dots)$. In this multi-D case, we can use the multi-D Fourier transforms and characteristic functions, where for $\omega = (\omega_1, \omega_2, \dots)$, we define $F(\omega) = E \left[e^{i\omega \cdot z} \right]$, with $\omega \cdot z \stackrel{\text{def}}{=} \omega_1 \cdot z_1 + \omega_2 \cdot z_2 + \dots$.

Second, we need to take into account that while theoretically, we can consider all possible values of the difference $z^{(1)} - z^{(2)}$, in practice, we can only get values which are proportional to the smallest measuring unit h . For example, if we measure distance and the smallest distance we can measure is centimeters, then the measuring instrument can only return values 0 cm, 1 cm, 2 cm, etc. In other words, in reality, the value z can only take discrete values. If we take the smallest value of z as the new starting point (i.e., as 0), then the possible values of z take the form $z = 0, z = h, z = 2h, \dots$, until we reach the upper bound $z = N \cdot h$ for

some integer N . For these values, in the 1-D case, the Fourier transform takes the form

$$F(\omega) = E [e^{i\omega \cdot z}] = \sum_{k=0}^N p_k \cdot e^{i\omega \cdot k \cdot h},$$

where p_k is the probability of the value $z = k \cdot h$. This formula can be equivalently rewritten as $F(\omega) = \sum_{k=0}^N p_k \cdot s^k$, where $s \stackrel{\text{def}}{=} e^{i\omega \cdot h}$. Similarly, in the multi-D case, we have $z = (k_1 \cdot h_1, k_2 \cdot h_2, \dots)$, and thus,

$$e^{i\omega \cdot k \cdot h} = e^{i\omega \cdot (k_1 \cdot h_1 + k_2 \cdot h_2 + \dots)} = e^{i\omega_1 \cdot k_1 \cdot h_1} \cdot e^{i\omega_2 \cdot k_2 \cdot h_2} \cdot \dots,$$

so we have

$$F(\omega) = \sum_{k_1=0}^{N_1} \sum_{k_2=0}^{N_2} \dots p_k \cdot s_1^{k_1} \cdot s_2^{k_2} \cdot \dots,$$

where $s_k \stackrel{\text{def}}{=} e^{i\omega_k \cdot h_k}$. In other words, we have a polynomial of the variables s_1, s_2, \dots :

$$P(s_1, s_2, \dots) = \sum_{k_1=0}^{N_1} \sum_{k_2=0}^{N_2} \dots p_k \cdot s_1^{k_1} \cdot s_2^{k_2} \cdot \dots$$

Different values of ω correspond to different values of $s = (s_1, s_2, \dots)$. Thus, the fact that we know the values of $|F(\omega)|^2$ for different ω is equivalent to knowing the values of $|P(s)|^2$ for all possible values $s = (s_1, s_2, \dots)$.

In these terms, the theoretical question takes the following form: we know the values $D(s) = |P(s)|^2 = P(s) \cdot P^*(s)$ for some polynomial $P(s)$, we need to reconstruct this polynomial. In the 1-D case, each complex-valued polynomial of degree N has, in general, N complex roots $s^{(1)}, s^{(2)}, \dots$, and can, therefore, be represented as $|P(s)|^2 = \text{const} \cdot (s - s^{(1)}) \cdot (s - s^{(2)}) \cdot \dots$. In this case, there are many factors, so there are many ways to represent it as a product – which explains the above-described non-uniqueness of representing $D(s)$ as the product of two polynomials $P(s)$ and $P^*(s)$.

Interestingly, in contrast to the 1-D case, in which each polynomial can be represented as a product of polynomials of 1st order, in the multi-D case, a generic polynomial *cannot* be represented as a product of polynomials of smaller degrees. This fact can be easily illustrated on the example of polynomials of two variables.

To describe a general polynomial of two variables $\sum_{k=0}^n \sum_{l=1}^n c_{kl} \cdot s_1^k \cdot s_2^l$ in which each of the variables has a degree $\leq n$, we need to describe all possible coefficients c_{kl} . Each of the indices k and l can take $n + 1$ possible values $0, 1, \dots, n$, so overall, we need to describe $(n + 1)^2$ coefficients.

When two polynomials multiply, the degrees add: $s^m \cdot s^{m'} = s^{m+m'}$. Thus, if we represent $P(s)$ as a product of two polynomials, one of them must have a degree $m < n$, and the other one degree $n - m$. In general:

- we need $(m + 1)^2$ coefficients to describe a polynomial of degree m and

- we need $(n - m + 1)^2$ coefficients to describe a polynomial of degree $n - m$,
- so to describe arbitrary products of such polynomials, we need $(m + 1)^2 + (n - m + 1)^2$ coefficients.

To be more precise, in such a product, we can always multiply one of the polynomials by a constant and divide another one by the same constant, without changing the product. Thus, we can always assume that, e.g., in the first polynomial, the free term c_{00} is equal to 1. As a result, we need one fewer coefficient to describe a general product: $(m + 1)^2 + (n - m + 1)^2 - 1$.

To be able to represent a generic polynomial $P(s)$ of degree n as such a product $P(s) = P_m(s) \cdot P_{n-m}(s)$, we need to make sure that the coefficients at all $(n + 1)^2$ possible degrees $s_1^k \cdot s_2^l$ are the same on both sides of this equation. This requirement leads to $(n + 1)^2$ equations with $(m + 1)^2 + (n - m + 1)^2 - 1$ unknowns.

In general, a system of equations is solvable if the number of equations does not exceed the number of unknowns. Thus, we must have

$$(n + 1)^2 \leq (m + 1)^2 + (n - m + 1)^2 - 1.$$

Opening parentheses, we get

$$n^2 + 2n + 1 \leq m^2 + 2m + 1 + (n - m)^2 + 2 \cdot (n - m) + 1 - 1.$$

The constant terms in both sides cancel each other, as well as the terms $2n$ in the left-hand side and $2m + 2 \cdot (n - m) = 2n$ in the right-hand side, so we get an equivalent inequality $n^2 \leq m^2 + (n - m)^2$. Opening parentheses, we get $n^2 \leq m^2 + n^2 - 2 \cdot n \cdot m + m^2$. Cancelling n^2 in both sides, we get $0 \leq 2m^2 - 2 \cdot n \cdot m$. Dividing both sides by $2m$, we get an equivalent inequality $0 \leq m - n$, which clearly contradicts to our assumption that $m < n$.

Let us go back to our problem. We know the product $D(s) = P(s) \cdot P^*(s)$, and we want to reconstruct the polynomial $P(s)$. We know that this problem is not uniquely solvable, i.e., that there exist other polynomials $Q(s) \neq P(s)$ for which $D(s) = P(s) \cdot P^*(s) = Q(s) \cdot Q^*(s)$. Since, in general, a polynomial $P(s)$ of several variables cannot be represented as a product – i.e., is “prime” in terms of factorization the same way prime numbers are – the fact that the two products coincide means that $Q(s)$ must be equal to one of the two prime factors in the decomposition $D(s) = P(s) \cdot P^*(s)$. Since we know that $Q(s)$ is different from $P(s)$, we thus conclude that $Q(s) = P^*(s)$.

By going back to the definitions, one can see that for the distribution $\rho'(x) = \rho(-x)$, the corresponding polynomial has exactly the form $Q(s) = P^*(s)$. Thus, in general, this is the *only* non-uniqueness that we have: each distribution which is consistent with the observation of differences coincides either with the original distribution $\rho(x)$ or with the distribution $\rho'(x) = \rho(-x)$. In other words, we arrive at the following result.

Answer to the theoretical question. We have proven that, in general, each distribution which is consistent with the observation of differences $\Delta x^{(1)} - \Delta x^{(2)}$ coincides either with the original distribution $\rho(x)$ or with the distribution $\rho'(x) \stackrel{\text{def}}{=} \rho(-x)$.

4 Practical Question: How to Gauge the Accuracy

How to use Fourier techniques to solve the practical question: idea. We want to find a probability distribution $\rho(z)$ which is consistent with the observed characteristic function $D(\omega)$ for the difference. In precise terms, we want to find a function $\rho(z)$ which satisfies the following two conditions:

- the first condition is that $\rho(z) \geq 0$ for all z , and
- the second condition is that $|F(\omega)|^2 = D(\omega)$, where $F(\omega)$ denotes the Fourier transform of the function $\rho(x)$.

One way to find the unknown function that satisfies two conditions is to use the method of successive projections. In this method, we start with an arbitrary function $\rho^{(0)}(z)$. On the k -th iteration, we start with the result $\rho^{(k-1)}(z)$ of the previous iteration, and we do the following:

- first, we project this function $\rho^{(k-1)}(z)$ onto the set of all functions which satisfy the first condition; to be more precise, among all the functions which satisfy the first condition, we find the function $\rho'(z)$ which is the closest to $\rho^{(k-1)}(z)$;
- then, we project the function $\rho'(z)$ onto the set of all functions which satisfy the second condition; to be more precise, among all the functions which satisfy the second condition, we find the function $\rho^{(k)}(z)$ which is the closest to $\rho'(z)$.

We continue this process until it converges.

As the distance between the two functions $f(z)$ and $g(z)$ – describing how close they are – it is reasonable to take the natural analog of the Euclidean distance: $d(f, g) \stackrel{\text{def}}{=} \sqrt{\int (f(z) - g(z))^2 dz}$. One can check that for this distance function:

- the closest function in the first part of the iteration is the function $\rho'(z) = \max(0, \rho^{(k-1)}(z))$, and
- on the second part, the function whose Fourier transform is equal to

$$F^{(k)}(\omega) = \frac{\sqrt{|D(\omega)|}}{|F'(\omega)|} \cdot F'(\omega).$$

Thus, we arrive at the following algorithm.

How to use Fourier techniques to solve the practical question: algorithm. We start with an arbitrary function $\rho^{(0)}(z)$. On the k -th iteration, we start with the function $\rho^{(k-1)}(z)$ obtained on the previous iteration, and we do the following:

- first, we compute $\rho'(z) = \max(0, \rho^{(k-1)}(z))$;
- then, we apply Fourier transform to $\rho'(z)$ and get $F'(z)$;
- after that, we compute $F^{(k)}(\omega) = \frac{\sqrt{|D(\omega)|}}{|F'(\omega)|} \cdot F'(\omega)$;

- finally, as the next approximation $\rho^{(k)}(z)$, we take the result of applying the inverse Fourier transform to $F^{(k)}(\omega)$.

We continue this process until it converges.

Acknowledgment

This work was supported in part by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070), by the National Science Foundation grants HRD-0734825, HRD-1242122, DUE-0926721, by Grants 1 T36 GM078000-01 and 1R43TR000173-01 from the National Institutes of Health, and by a grant N62909-12-1-7039 from the Office of Naval Research.

References

1. M. Aubinet, T. Vesala, and D. Papale (eds.), *Eddy Covariance – A Practical Guide to Measurement and Data Analysis*, Springer, Dordrecht, Heidelberg, London, New York, 2012.
2. J. A. Hole, “Nonlinear high-resolution three-dimensional seismic travel time tomography”, *Journal of Geophysical Research*, 1992, Vol. 97, pp. 6553–6562.
3. O. Ochoa, A. Velasco, and C. Servin, “Towards Model Fusion in Geophysics: How to Estimate Accuracy of Different Models”, *Journal of Uncertain Systems*, 2013, Vol. 7, No. 3, pp. 190–197.
4. O. Ochoa, A. A. Velasco, C. Servin, and V. Kreinovich, “Model Fusion under Probabilistic and Interval Uncertainty, with Application to Earth Sciences”, *International Journal of Reliability and Safety*, 2012, Vol. 6, No. 1–3, pp. 167–187.
5. S. Rabinovich, *Measurement Errors and Uncertainties: Theory and Practice*, American Institute of Physics, New York, 2005.
6. C. Servin, O. Ochoa, and A. A. Velasco, “Probabilistic and interval uncertainty of the results of data fusion, with application to geosciences”, *Abstracts of 13th International Symposium on Scientific Computing, Computer Arithmetic, and Verified Numerical Computations SCAN’2008*, El Paso, Texas, September 29 – October 3, 2008, p. 128.