Measurement Errors and Error Analysis

By Eric W. Frew and Kristine Larson, modified by D.J. Scheeres and P. Axelrad.

Describing Measurement Errors

When carrying out experimental measurements one obtains a sequence of measured values, denoted by x_i, $i = 1, 2, \ldots, N$. It is of fundamental interest to describe this entire set of measurement values using only a few numbers. We can do this by defining the mean, standard deviation, and median of this data set.

Mean: The mean of a data set is the average value, and represents the simplest description of a sequence of observations. We often consider this to be the estimate of the value of x.

$$\bar{x} = E[x_i] = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Standard Deviation: The standard deviation of a data set is the square root of the average value of the difference between the mean and the data set.

$$\sigma_x = \sqrt{E[(x_i - \bar{x})^2]} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

For dealing with measured quantities, it's actually more proper to define the "Sample Standard Deviation."

$$\sigma_x = \sqrt{E[(x_i - \bar{x})^2]} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

For a large number of observations, the two will essentially be equal.

The square of the standard deviation is called the variance of x, σ_x^2. The standard deviation provides an estimate of the variation from measurement to measurement in your data. If the standard deviation is "0" (which is unrealistic in general) then all the measurements are the same and there is no apparent error, usually the standard deviation is non-zero and can be used as an approximate estimate of the likely statistical error made in each measurement.

Median: The median of a data set is the value, x_j, such that half the values of x_i are larger than this value and half are smaller. If the data set has an even number of points, the mean of the two values that cross the middle value is often taken. The median provides a different, independent estimate of the central value of the data set and should be contrasted with the mean. A large difference between the mean and median can arise due to an "outlier" measurement that may have large, non-random error associated with it.

In summary, for a given sequence of observed values, or values based on other observations, we denote the estimate of this value as its mean value and represent the expected error in our estimate by the variance. We can use the median as a check on the mean: if their difference is comparable or larger than the standard deviation then it is likely that we have some “bad” observations with a large, non-random error.

Error Analysis

When designing an experiment to measure some quantity of interest it is important to both understand and predict how errors in the measurements translate into errors in the final result. This note will briefly discuss two different methods to predict and quantify the error that results in an experimental measurement.

Consider a scalar *output* quantity x that is related to multiple *input* variables y_i through the *measurement function*

$$x = h(y_1, \cdots, y_n). \tag{1}$$

The variables y_i could be of a similar type (e.g. time taken at two different instances) or of different types (e.g. time and height or distance), and x could be the quantity we are computing, such as coefficient of restitution or distance.

In practice the *measurements* z_i taken of the variables y_i are corrupted by some uncertainty or noise which we will denote v_i ($z_i = y_i + v_i$). Thus the quantity $\bar{x}$ we derive from the measurements is

$$\bar{x} = h(z_1, \cdots, z_n). \tag{2}$$

We are interesting in understanding the error of the output ($\tilde{x} = \bar{x} - x$) as a function of the measurements and the measurement error.

For our experiments, x represents the “true” value of the parameter given the actual values of the variables y_i. Here the quantity $\bar{x}$ represents the value of the parameter computed for one set of measurements.

The previous section describes how to combine a sequence of these values of x into an estimate and its standard deviation. This section describes how to evaluate the sensitivity of this parameter x to measurement errors.

Taylor Series Expansion

In order to analyze error propagation in the measurement function we will use the Taylor series expansion of the measurement function about some nominal values of the input variable. Taking the measurement variables as the nominal values we can rewrite Equation (1) as follows:

$$x = h(y_1, \cdots, y_n) = h(z_1 - v_1, \cdots, z_n - v_n) = h(z_1, \cdots, z_n) - \frac{\partial h}{\partial y_1}(z_1, \cdots, z_n) \cdot v_1 - \cdots - \frac{\partial h}{\partial y_n}(z_1, \cdots, z_n) \cdot v_n + H.O.T \tag{3}$$

Next, notice that the first term on the right-hand side of Equation (3) is equivalent to $\bar{x}$ and therefore

$$\tilde{x} = \bar{x} - x \approx \frac{\partial h}{\partial y_1}(z_1, \cdots, z_n) \cdot v_1 + \cdots + \frac{\partial h}{\partial y_n}(z_1, \cdots, z_n) \cdot v_n \tag{4}$$

where we have dropped the higher order terms.

Random Error

In this section we consider the case when the measurement error v_i is a random variable. In particular, we assume the measurement error has a normal distribution with:

- Zero mean: $E[v_i] = 0$ for all i
- Standard deviation: $\sqrt{E\left[v_i^2\right]} = \sigma_i$ possible different for each i
- No cross-correlation: $E[v_i \cdot v_j] = 0 \quad i \neq j$

With these assumptions we can see that the error of the output has a mean of zero.

$$\begin{aligned} E[\tilde{x}] &= E\left[\frac{\partial h}{\partial y_1}(z_1, \cdots, z_n) \cdot v_1 + \cdots + \frac{\partial h}{\partial y_n}(z_1, \cdots, z_n) \cdot v_n\right] \\ &= \frac{\partial h}{\partial y_1}(z_1, \cdots, z_n) \cdot E[v_1] + \cdots + \frac{\partial h}{\partial y_n}(z_1, \cdots, z_n) \cdot E[v_n] = 0 \end{aligned} . \tag{5}$$

Furthermore, we can calculate the standard deviation of the output error $\sigma_{out} = \sqrt{E[\tilde{x} \cdot \tilde{x}]}$. The expected value of the square of the output will have terms consisting of v_i^2 and $v_i v_j$. Since the cross-correlation between error sources is zero, the latter terms will disappear and

$$E[\tilde{x}^2] = \sum_{i=1}^{n} E\left[\left(\frac{\partial h}{\partial y_i}\right)^2 \cdot v_i^2\right] = \sum_{i=1}^{n} \left(\frac{\partial h}{\partial y_i}\right)^2 \cdot E[v_i^2] = \sum_{i=1}^{n} \left(\frac{\partial h}{\partial y_i}\right)^2 \cdot \sigma_i^2 . \tag{6}$$

Thus

$$\sigma_x = \sqrt{E[\tilde{x}^2]} = \sqrt{\sum_{i=1}^{n} \left(\frac{\partial h}{\partial y_i}\right)^2 \cdot \sigma_i^2} . \tag{7}$$

The partial derivatives in Equation (7) are calculated from the specific measurement function being used for each experiment. The standard deviations of the measurement errors are determined based on how the experiment was conducted and what the expected "noise" in the measurements are.

Uncertainty Bounds

Instead of assuming random error on the measurement vectors, error analysis can also be conducted assuming only bounds on the measurement error. In this case we assume the measurement error bounds are symmetric, i.e. $-e_i \leq v_i \leq e_i$ or $|v_i| \leq e_i$ where e_i is a positive scalar representing worst case error.

In this case we can calculate a (positive) bound $\tilde{x}_{bound}$ on the output error such that

$$|\tilde{x}| \leq \tilde{x}_{bound} \tag{8}$$

From Equation (4) we see

$$|\tilde{x}| = \left|\sum_{i=1}^{n} \frac{\partial h}{\partial y_i} \cdot v_i\right| \leq \sum_{i=1}^{n} \left|\frac{\partial h}{\partial y_i} \cdot v_i\right| = \sum_{i=1}^{n} \left|\frac{\partial h}{\partial y_i}\right| \cdot |v_i| \leq \sum_{i=1}^{n} \left|\frac{\partial h}{\partial y_i}\right| \cdot e_i \tag{9}$$

Therefore

$$\tilde{x}_{bound} = \sum_{i=1}^{n} \left|\frac{\partial h}{\partial y_i}\right| \cdot e_i \,. \tag{10}$$

Again, the bound on each measurement are determined based on how the experiment was conducted and what the largest expected measurement errors are.

An example of error propagation from GPS

In GPS, satellites transmit at two frequencies. The ability of a GPS receiver to faithfully record (i.e. precisely) these two signals (called P1 and P2) is about the same, i.e.

$\sigma_{P1} = \sigma_{P2} = 0.5 meters$

The P1 and P2 measurements themselves are contaminated by the effects of the ionosphere. This causes a systematic error in the measurements that will ultimately create navigation errors for GPS users. However, the following linear combination has been shown to remove nearly all the systematic effects of the ionosphere:

$P3 = 2.54 * P1 - 1.54 * P2$

What is the precision of the P3 observable? If we assume P1 and P2 are uncorrelated:

$$\sigma_{P3}^2 = \left(\frac{\partial P3}{\partial P1}\right)^2 \sigma_{P1}^2 + \left(\frac{\partial P3}{\partial P2}\right)^2 \sigma_{P2}^2$$

or

$\sigma_{P3} = 0.5\sqrt{2.54^2 + 1.54^2} \sim 0.5 * 3$

so the price of using P3 is that your observations are three times noisier. The reason people in GPS use it is that it removes a very large systematic error that can create 50 meters of error over very short periods of time.

Another example

The speed of a constant velocity particle can be determined by measuring the time it takes to move a certain distance. If the distance of travel d and the time interval t are two measurements, we have as a measurement function

$$v = \frac{d}{t}.$$

Let's assume we can measure the distance of travel d_0 to within 1 cm (i.e. $e_d = 0.01$m) and the time interval t_0 to within one-tenth of a second ($e_s = 0.1$s). We wish to know the error bound on our calculation of the particle speed.

Since the speed is the output of our experiment we have

$$v = h(d,t) = \frac{d}{t}$$

From Equation (10) we can calculate the bounds on the output from the bounds on the inputs. To do so we need the partial derivatives of the measurement function with respect to each measurement:

$$\begin{aligned} \frac{\partial h}{\partial d} &= \frac{1}{t} \\ \frac{\partial h}{\partial t} &= \frac{d}{t^2} = \frac{v}{t} \end{aligned}.$$

Evaluating these partial derivatives at the measured values yields

$$e_v = \left|\frac{\partial h}{\partial d}(d_0,t_0)\right| \cdot e_d + \left|\frac{\partial h}{\partial t}(d_0,t_0)\right| \cdot e_t = \frac{e_d}{t_0} + \frac{e_t \cdot v_0}{t_0} = \frac{1}{t_0}\left(0.01m + 0.1s \cdot v_0\right).$$

In the above expression the bounds on the measurements and the particle speed v_0 are constants. However, the measured time t_0 is a result of the design of the experiment. The above equation shows that for fixed measurement bounds we can reduce the overall uncertainty bound on our output by increasing the total time t_0 of the experiment. This is equivalent to increasing the distance of travel d_0. Thus, we can use the error analysis to design a better experiment (let the particle move over as long a distance as possible!) Furthermore, we can design this better experiment without have to know the actual result v_0, we only need to know the measurement function that relates it to our measurement variables.