Error Analysis (First Draft)

Chris Clark

June 22, 2006

1 Types of Error

Most of the time when we try to measure some quantity in the physical world we have the potential for being off by a bit. Under different circumstances, the error that we are off by may vary greatly, so it is very useful to have a measurement of the quality of the measurement, which we call the **error analysis**. First we choose a physical **quantity** X to determine and then we perform an **experiment** which is a set of measurements of this quantity. Each **measurement** gives a **value** x_i. The **outcome** O_X of the experiment is the set of all results. Finally, we present the **result** $\bar{x}$, which is the average of the results in the outcome. The **true value** X_o is the value that would be measured if our experiment had no error. A good experiment will produce a result very close to the true value, $\bar{x} \cong X_o$.

Sources of error are commonly classified as either *systematic error* or *random error*. A **systematic error** is any error that skews the measured values by the same amount in the same direction every time.[1] For example, if you were making measurements with a stopwatch that always started at 0.01s instead of 0.00s, then you would have a systematic error. A **random error** is any error that is not a systematic error. For example, if you were measuring an object with a ruler and you were off by a fraction of a millimeter, then you would have a random error.

Random error usually skews measured values higher and lower with equal probability. If this is the case, then it possible to eliminate it by averaging, which we will denote averaging by enclosing the expression in angle brackets or covering it with a bar. First, we represent the value of a measurement by breaking it down into three components.

$$\text{Measurement} = \text{True Value} + \text{Systematic Error} + \text{Random Error}$$

Now averaging is distributive so we have

$$\langle \text{Measurement} \rangle = \langle \text{True Value} \rangle + \langle \text{Systematic Error} \rangle + \langle \text{Random Error} \rangle$$

[1] If an error source skews the measured values by a roughly constant amount in one direction, but has a slightly different effect each time, then it is probably an error source that has both a systematic and a random component.

The first two terms on the right hand side are just constants, so we can drop the averaging symbols on them. The last term has an average of approximately zero in most cases yielding

$$\langle \text{Measurement} \rangle = \text{True Value} + \text{Systematic Error}$$

We can't get rid of the systematic error with statistical methods, but by using quality equipment and improving our experimental technique we can usually reduce systematic error to a negligible level. Therefore, if you average your measured values and take care to minimize systematic error then you will obtain a result that is approximately equal to the true value.

It is not always the case that random error cancels with averaging. For example, if you are measuring the speed of a rolling ball and there is an intermittent breeze that is always in the same direction, then there will be a random error in the speed that biases it in one direction. Averaging will not be able to eliminate this type of random error.

2 Measuring Error

A good way to visualize the outcome of an experiment is with a *histogram*. A **histogram** is a chart made by stacking up squares in columns based on the values of measurements, with one square added per measurement. This makes the more frequent results form taller columns.

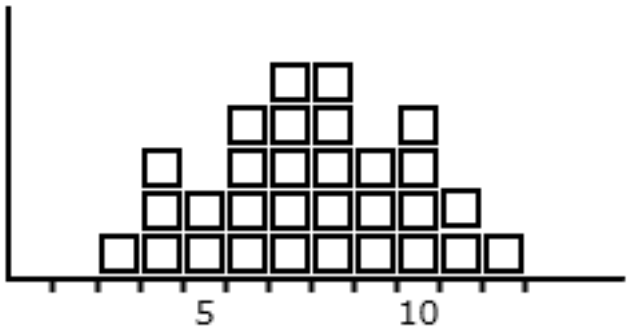

Figure 1: Sample Histogram

In this figure, each square represents one measurement of a quantity X. The value of the measurement, which we call x, is the number of the column that it lies in. We can write the value of the i^{th} measurement as x_i. Most quantities that we measure are not always integers, but we can still use histograms by making each column a **bin** that collects any value falling within a specific range. We will refer to the shape that the histogram forms as a **distribution** of measurements.

There are two distinct measures of the error of a result, *accuracy* and *precision*. The **accuracy** tells you how close your result is to the accepted result. Accuracy is almost always expressed in terms of the **percent error** [2], which is given by

$$\frac{\text{Experimental Result} - \text{Accepted Result}}{\text{Accepted Result}} \times 100 \tag{1}$$

The **precision** tells you how consistent your measured values are, with higher precision indicating that all of your data points are clustered around one value. Effectively, precision is a measure of the width of the distribution for the experiment. In order to quantify the width of the distribution we need a formula that accounts for the contribution to the spread by each measured value. The first thing that comes to mind is the **deviation** of each value, which is given by

$$d_i \equiv x_i - \bar{x} \tag{2}$$

but this won't help to much because if we average the deviation over all the values, we will usually get an answer close to zero. This is because random error usually is equally likely to skew the values in either direction from the average and this will cause the deviations to cancel positive with negative. But there is a quick fix for this—we can just take the absolute value of the deviation and average that. This defines the **mean deviation** [3]

$$\alpha_{O_X} = \langle |x_i - \bar{x}| \rangle \tag{3}$$

where the subscript O_X indicates that the expression is computed using the outcome of the experiment measuring the quantity X.

The mean deviation is not used vary often. Instead of taking the absolute value, the square of the deviation is used. The square is again always positive, but it is more convenient than the absolute values because squares are easier to work with analytically and there are other advantages that we will see soon. Averaging the square of the deviation and taking the square root yields the **standard deviation** [4]

$$\sigma_{O_X} = \sqrt{\langle (x_i - \bar{x})^2 \rangle} \tag{4}$$

There is a useful identity involving the standard deviation that can be derived easily.

$$\begin{aligned}\sigma_{O_X}^2 &= \left\langle (x_i - \langle x_i \rangle)^2 \right\rangle = \left\langle x_i^2 - 2x_i\langle x_i \rangle + \langle x_i \rangle^2 \right\rangle \\ &= \langle x_i^2 \rangle - \langle 2x_i \langle x_i \rangle \rangle + \langle x_i \rangle^2 \\ &= \langle x_i^2 \rangle - 2\langle x_i \rangle^2 + \langle x_i \rangle^2 = \langle x_i^2 \rangle - \langle x_i \rangle^2\end{aligned}$$

Now we want to consider the shape that a histogram of the outcome would take if we had performed an infinite number of measurements. This limiting

[2] Percent error is also called percent deviation

[3] The mean deviation is also called the average deviation

[4] This is the biased or population standard deviation. There are arguments that it is better to divide by N-1 instead of N in the average, but we will not pursue that here. In Microsoft Excel, the population standard deviation is stdevp.

distribution would give us a sense for what the error itself really looks like instead of just seeing a few pieces of the puzzle. Without further information about the measurement process, we can't be sure what the histogram would look like. But the **central limit theorem** proves that if the measurement process is affected by many independent random errors, then the limiting distribution of an infinite number of measurements will be approximately a *Gaussian distribution*. Whether or not this is the reason, it is true that very often random error in measurements produces a **Gaussian distribution**, which is a bell-shaped curve with the functional form

$$G(x) = Ae^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{5}$$

where A is the height parameter, σ is the width parameter, and μ is the center of the peak.

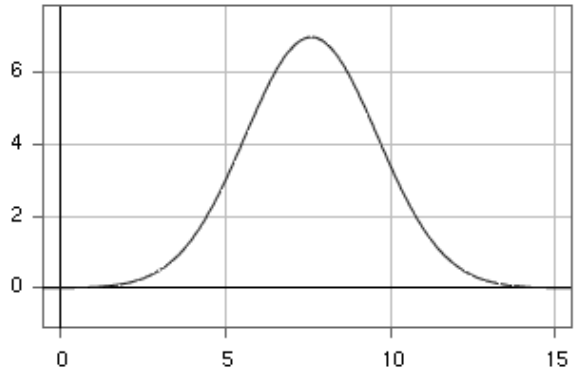

Figure 2: Gaussian Distribution

We can now view this continuous distribution as the real form of the random error and we can perform statistical operations on the function rather than on the discrete data points. In this way, we treat the quantity X that we are measuring as a **random variable**. The actual value of a random variable is not just a single number, it can take on many specific values, just as many repeated measurements of the same quantity can take on many values. But a random variable's value is not completely random, it takes on values with a certain probability distribution. Every random variable X has a corresponding **probability distribution** ρ_X which determines the probability of each specific value of X. The probability distribution is a function such that the probability that X takes on the specific value x is proportional to $\rho_X(x)$. As we just realized, the probability distribution will look like a Gaussian distribution. When dealing with continuous random variables that can take the value of any real number, the probability of getting a specific value x is always zero, but the probability

of x falling in some range can be non-zero. To be precise, we can write the probability that X takes on a value between x_1 and x_2 as

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} \rho_X(x)\ dx \tag{6}$$

Notice that X must always take on a value between $-\infty$ and ∞ so $P(-\infty \leq X \leq \infty) = 1$ and therefore

$$1 = \int_{-\infty}^{\infty} \rho_X(x)\ dx \tag{7}$$

This equation indicates that the probability distribution is **normalized**, which just means that its integral over all values of x is equal to 1.

The **variance** of a distribution is simply the square of its standard deviation. We can translate the definition of standard deviation to get an expression for variance that utilizes the probability distribution.

$$\sigma_X^2 = \langle (x_i - \bar{x})^2 \rangle$$

$$= \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2$$

for a distribution containing N measured values. Next we switch the variable of summation from the index i to the values x, where we must assume that the values are all integers. But there can be many values of i that all have the same value, so we must introduce a factor $n(x)$ which is the number of values in the distribution with value x.

$$= \frac{1}{N} \sum_{x=-\infty}^{\infty} (x - \bar{x})^2 n(x)$$

If we drop the requirement that x must be an integer, then we have to convert the sum to an integral.

$$= \int_{-\infty}^{\infty} (x - \bar{x})^2 \frac{n(x)}{N}\ dx$$

In the limit $N \to \infty$, $n(x)$ increases along with N so that $n(x)/N \to \rho(x)$. Therefore

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - \bar{x})^2 \rho_X(x)\ dx \tag{8}$$

One of the other reasons why the standard deviation is used as a measure of precision is that the standard deviation of a Gaussian is exactly half of the width of the peak at half the height of its maximum. It turns out that if you integrate a Gaussian probability distribution from $\bar{x} - \sigma_X$ to $\bar{x} + \sigma_X$ you get about 0.6827.[5] This means that about 68.27% of measurements will yield a value that lies within one standard deviation of the average value.

[5]The integral of the Gaussian distribution is not analytical, it is called the error function.

When presenting results it is desirable to include an estimate of the error right with the result itself. This is accomplished by appending a distance that the true value should lie within. By convention, the distance used is just the standard deviation. It is true that there is a reasonable chance that the true value will not fall within one standard deviation of the result, but under normal circumstances the only distance that we could be sure the true value lies within would be far too large to be of use. This is because there is usually a slight probability that any given measurement will give a value that is very far from the true value. Therefore it is convenient to just state results as

$$X_o = \bar{x} \pm \sigma_X \tag{9}$$

After this form is obtained, it can be compared with other experiments for *consistency*. Two results $X_o = \bar{x} \pm \sigma_X$ and $X'_o = \bar{x}' \pm \sigma'_X$ are **consistent** iff $|\bar{x} - \bar{x}'| \leq \sigma_X + \sigma'_X$.[6] In other words, two results are consistent iff the half width at half maximum of their probability distributions overlap.

3 Worst-Case Error Propagation

So far we have seen the standard deviation used on the outcome of an experiment to get a sense of the spread of the data. It is also possible to estimate the expected standard deviation of the outcome by accounting for each of the random error sources, which is called **error propagation**. It will take some mathematical work to derive the error propagation formula, but first we need to know how to account for the random error sources. First of all we must assume that we are aware of all the random error sources. This is usually not a problem because we just treat each measured quantity as having one combined random error source, which may in fact include many independent error sources. If we are directly measuring the quantity of interest, then that is all there is to it—the expected standard deviation is the standard deviation of the distribution of measurements. The only problem is that we do not know the distribution of measurements before we make them, so how can we know it's standard deviation? This is where we can make an estimate. The **least count** is the size of the smallest division on the measuring device. A competent experimenter using working equipment should rarely make errors larger than the least count. In fact, you should always be able to round to the correct division, so you will probably always be within half of the least count. For example, on a normal ruler the least count is 1mm and you should never be off by more than 0.5mm. But clearly there is a good chance that your measured value will differ from the true value by 0.1mm, so there is some nonzero spread to the distribution of measurements. A good rule to use is to assume the standard deviation of your distribution of measurements is half of the least count. This means that about 68% of your measurements lie within half a least count from the true value.

[6] This definition can be relaxed a little because of the fact that there is a decent chance that the true value lies outside one standard deviation in both experiments.

Note that this asssumes that the systematic error has been brought down to a negligible level.

So we have handled the case where we are directly measuring the quantity of interest, but that is not usually how experiments are done. Most experiments involve several measurements whose results are plugged into an equation to get the final result. For example, if you are measuring the area of a rectangle, you must measure the width and the height, and then plug these results into the formula $A = w * h$. The error in measuring the width and height is known based on the least count, but what amount of error do these error sources produce in the result for the area? This is where the error propagation formulas are needed.

There is only one error propagation formula that you need to know, which will be derived at the end, but there are other formulas which are either approximations or special cases of the final formula. The simplest type of error propagation formula is the worst case type. Formulas of this type just find the highest and lowest possible values of the result for values of the parameters which lie within their error bounds. The formulas for addition and multiplication of parameters are shown below.

$$F(X,Y) = X + Y \Rightarrow \sigma_F \leq \sigma_X + \sigma_Y$$

$$F(X,Y) = X * Y \Rightarrow \frac{\sigma_F}{|\bar{f}|} \leq \frac{\sigma_X}{|\bar{x}|} + \frac{\sigma_Y}{|\bar{y}|}$$

Error in the form $\sigma_X/\bar{x}$ is called **relative error**. In contrast, error in the regular form σ_X is called **absolute error**. We will not be using this simple form of error propagation because it usually gives an overestimate of the error because it is not so likely that the results of both parameters will be off from the true value in the same direction.

4 Convolution

Consider a random variable F that is the sum of the two random variables X and Y, so $F(X,Y) = X + Y$. We want to find the error propagation formula that gives the standard deviation of F in terms of the standard deviations of X and Y. It turns out to be more convenient to use the variances, but the standard deviation can be obtained simply by taking the square root. For simplicity, we will assume that the probability distributions of X and Y are centered on zero, and hence so is the probability distribution for F. It can be shown that shifting the distributions in this way does not affect the standard deviation of F, but we will not need this to proceed.

$$\sigma_F^2 = \int_{-\infty}^{\infty} f^2 \rho_F(f)\ df$$

The probability distribution $\rho_F(f)$ is determined by adding up all the probabilities for different ways that $x + y = f$. For each given pair x, y that satisfies

this equation, the probability is $\rho_X(x)\rho_Y(y)$. But there are an infinite number of pairs x, y which satisfy this equation, so we must integrate over all of them. This is called the **convolution** of the distributions ρ_X and ρ_Y.

$$\sigma_F^2 = \int_{-\infty}^{\infty} f^2 \left[\int_{-\infty}^{\infty} \rho_X(\alpha)\rho_Y(f-\alpha)\ d\alpha \right]\ df$$

Switching the order of integration,

$$\sigma_F^2 = \int_{-\infty}^{\infty} \rho_X(\alpha) \left[\int_{-\infty}^{\infty} f^2 \rho_Y(f-\alpha)\ df \right]\ d\alpha$$

Let $u = f - \alpha$ so $du = df$ since α is a constant for the f integral.

$$\sigma_F^2 = \int_{-\infty}^{\infty} \rho_X(\alpha) \left[\int_{-\infty}^{\infty} (u+\alpha)^2 \rho_Y(u)\ du \right]\ d\alpha$$

$$\sigma_F^2 = \int_{-\infty}^{\infty} \rho_X(\alpha) \left[\int_{-\infty}^{\infty} (u^2 + 2\alpha u + \alpha^2) \rho_Y(u)\ du \right]\ d\alpha$$

$$\sigma_F^2 = \int_{-\infty}^{\infty} \rho_X(\alpha)(\sigma_Y^2 + \alpha^2)\ d\alpha$$

where the $2\alpha u$ term is zero because the probability distribution is symmetric about zero.

$$\sigma_F^2 = \int_{-\infty}^{\infty} \alpha^2 \rho_X(\alpha)\ d\alpha\ + \sigma_Y^2$$

$$\sigma_F^2 = \sigma_X^2 + \sigma_Y^2 \tag{10}$$

This is the correct formula for error propagation in the special case of addition of two quantities.

5 Linear Combinations of Quantities

The formula of the previous section can be generalized to handle an arbitrary linear combination of random variables. First we need to understand the effect of multiplying a random variable by a constant. Without even checking the math, we can see that mutliplying a distribution by a constant c scales it by a factor of c, and since the standard deviation is proportional to the width of the distribution, the standard deviation likewise scales by a factor of C.

$$F(X) = cX \Rightarrow \sigma_F = c\sigma_X$$

Now if we consider a linear combination of two variables, we can consider each term to be a function of the previous form.

$$F(X_1, X_2) = c_1 X_1 + c_2 X_2 = F_1(X_1) + F_2(X_2)$$

$$\Rightarrow \sigma_F^2 = \sigma_{F_1}^2 + \sigma_{F_2}^2 = c_1^2 \sigma_{X_1}^2 + c_2^2 \sigma_{X_2}^2$$

A linear combination of three random variables can be handled by combining two of them into one function.

$$F(X_1, X_2, X_3) = c_1X_1 + c_2X_2 + c_3X_3 = F_{12}(X_1, X_2) + F_3(X_3)$$

$$\Rightarrow \sigma_F^2 = \sigma_{F_{12}}^2 + \sigma_{F_3}^2 = c_1^2\sigma_{X_1}^2 + c_2^2\sigma_{X_2}^2 + c_3^2\sigma {X_3}^2$$

And by repeating this process over and over, we can get the formula for an arbitrary linear combination of n random variables.

$$F(X_1, ..., X_n) = c_1X_1 + \cdots + c_nX_n \Rightarrow \sigma_F^2 = c_1^2\sigma_{X_1}^2 + \cdots + c_n^2\sigma_{X_n}^2 \tag{11}$$

6 Error Propagation Formula

We are finally in a position to derive the general error propagation formula for independent random errors. Let's say the desired quantity F is a function of the n random variables $X_1, ..., X_n$. If we assume that errors in each random variable are small, then we can approximate the function F with the multivariable Taylor series around the centers of each probability distribution.

$$F(X_1, ..., X_n) = F(\bar{x}_1, ..., \bar{x}_n) + \frac{\partial F}{\partial X_1}(X_1 - \bar{x}_1) + \cdots + \frac{\partial F}{\partial X_n}(X_n - \bar{x}_n)$$

where the partial derivatives are evaluated at $X_1 = \bar{x}_1, ..., X_n = \bar{x}_n$. This is just a linear combination of the random variables $(X_i - \bar{x}_i)$ which each have the same standard deviations as X_i since they are just shifted with respect to each other. The partial derivatives are the constant coefficients. The first term has no contribution to the standard deviation because it is just a constant. Therefore by equation (11),

$$\sigma_F^2 = \left(\frac{\partial F}{\partial X_1}\right)^2 \sigma_{X_1}^2 + \cdots + \left(\frac{\partial F}{\partial X_n}\right)^2 \sigma_{X_n}^2 \tag{12}$$

7 Standard Deviation of the Average

This error propagation formula gives us a way to measure how effectively averaging eliminates random error. Suppose we determine that every measurement of quantity X has an error of σ_X. The function for the average of N measured values is

$$F(X_1, ..., X_N) = \frac{X_1 + \cdots + X_N}{N}$$

where each X_i refers to one measurement of the quantity X. Applying the error propagation formula to this F,

$$\sigma_F^2 = \frac{1}{N^2}\sigma_{X_1}^2 + \cdots + \frac{1}{N^2}\sigma_{X_n}^2$$

But all the σ_{X_i} are the same and equal σ_X. So since there are N terms total,

$$\sigma^2_{(avg)} = \frac{\sigma^2_X}{N} \tag{13}$$

Evidently as we increase the number of measurements N, the error in the result falls with a factor of $\sqrt{N}$.

8 References

A nice straightforward introduction to this topic can be found in the first five chapters of Introduction to Error Analysis by John Taylor. A slightly more advanced treatment is given in Data Reduction and Error Analysis for the Physical Sciences by Philip Bevington. The term "least count" came from http://www.rit.edu/ uphysics/uncertainties/Uncertaintiespart1.html. A very interesting discussion of Guassian distributions can be found at http://www-biba.inrialpes.fr/Jaynes/cc07s.pdf.