Error analysis for labs: some background

PHY 252, Spring 2007

As you probably have learned by now, physics gives particular meanings to certain terms associated with error analyis:

Error or uncertainty refers to the best estimate of the quantitative range within which one can trust his or her results. The error can be calculated even if there is no "official" accepted value for what is being measured. To John Q. Public, "error" refers to how badly he screwed up, or how far off his answer is from the "official" value.

Systematic error refers to an error that is consistent from measurement to measurement. For example, if you always round down to the nearest tic mark on a meter stick when measuring length, you will make a systematic error of measuring a slightly shorter length.

Random error refers to an error which fluctuates in an unpredictable fashion from measurement to measurement. For example, if you were asked to determine the number of births per day at Stony Brook University Hospital, you would likely get a slightly different number each day you inquired even though the total number of births per year might remain fairly steady from year to year.

Accuracy refers to the degree to which our value is correct within uncertainty. This is largely a matter of having the correct calibration of all of our reference measurements. That is, if someone gave us a miscalibrated meter stick that was shorter than the official length of a meter, we might measure the length of a table with great precision (lots of decimal places) but poor accuracy (what we think is a meter is not really a meter).

Precision can be thought of as the number of meaningful digits to a measurement. A measurement of a length as being 1.03424 meters is more precise than a measurement of 1.03 meters; however, if the measurement was made with an incorrectly calibrated meterstick, the measurement will have high precision but low accuracy!

Reference standard is the thing that you measure against. For example, if we're measuring the length of various tables, we might use a meter stick for the measurement. The meter stick is then our reference standard for the measurements of the lengths of the tables, and if our reference standard is inaccurate, then all measurements made using it will also be inaccurate no matter how precise they are. The international system of units, or SI units (for *Le Syst‘eme International d’Unités*), tries as much as possible to use reference standards based on natural phenomena that anyone can replicate in a properly-equipped lab (mass is the glaring exception to this principle; mass is defined relative to a particular platinum-iridium cylinder sitting in Paris). The base units are a choice of seven well-defined units which by convention are regarded as dimensionally independent: the metre, the kilogram, the second, the ampere, the kelvin, the mole, and the candela. As an example, the SI unit definition of the second is as follows:

> The second is the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium 133 atom.

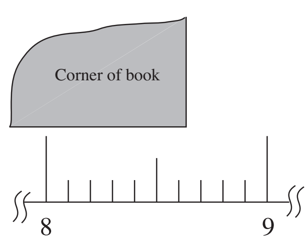
The name of the game in our lab activities is to provide estimates of the error of all of our measurements.

In this course, we will use only SI units. The U.S. is one of the only countries left on the planet that does *not* use SI units in everyday transactions; it may be that Liberia is the only other country that shares this characteristic. Even within our own country, SI units are becoming more and more prevalent within industry because of international trade. If you dare use inches, or pounds, or gallons in this course, you will be publicly humiliated and 0.01 points will be taken off of your grade.

SI units grew out of the metric system that was created during the French Revolution, and originally the unit of length was derived from a measurement of the diameter of the earth. There’s lots of good information out there on SI units; good places to start are the *Bureau International des Poids et Mesures* or BIPM web site, Wikipedia on `SI`, and the National Institute of Standards and Technology or NIST web site.

1 Measuring a length

Let’s say that you wanted to measure the width of your textbook, and when you held it up with a ruler aligned to one edge you saw the following at the other edge:



Everyone can agree that the number is between 8.6 and 8.7 centimeters or cm, but it’s tough to decide whether it’s 8.63 or 8.64, for example. A reasonable way to report this might be to say that it is 8.63 ± 0.03 cm, or 8.65 ± 0.05 cm. In writing these numbers, notice that we did not go beyond the precision of the error in specifying the value; that is, it makes no sense to write 8.63247 ± 0.1

cm. Also note that a number without units is meaningless in this context; if you were to just say that the length was about 8.6 it would be impossible to be certain if the length was 8.6 cm or 8.6 yards, for example. Also, the exact measure of the length of the book is only as good as A) the calibration of your meter stick and B) how accurately you were able to line up the zero point of the meter stick with the other end of the book. Also, if the book has a very rough edge, it may be difficult to say exactly where the edge is; you should account for that in reporting the error of your measurement (if the edge were bumpy or wavy at the level of 0.2 cm, you would say that the length was 8.6 ± 0.2 cm). Generally speaking, your estimate of the error should include all factors that you think provide a reasonable estimate of how well others can trust your measurement.

2 Error propagation

Continuing with the imaginary experiment of measuring your book, let's say that its width was x and its height is y and that we wish to calculate the area $A = x \cdot y$ of the book. If there's some error to the width x of Δx, and some error to the height y of Δy, then it's pretty obvious that there is going to be some error in the calculated area A:

$$\begin{aligned} A &= (x \pm \Delta x)(y \pm \Delta y) = x(1 \pm \frac{\Delta x}{x})y(1 \pm \frac{\Delta y}{y}) \\ &= xy\left(1 \pm \frac{\Delta x}{x} \pm \frac{\Delta y}{y} \pm \frac{\Delta x}{x}\frac{\Delta y}{y}\right) \simeq xy\left(1 \pm \frac{\Delta x}{x} \pm \frac{\Delta y}{y}\right) \\ &\simeq xy \pm (\Delta x + \Delta y) \end{aligned}$$

where we have assumed that $\Delta x/x$ and $\Delta y/y$ are both small so that $(\Delta x/x)(\Delta y/y)$ is really small and can be safely ignored. That is, we simply add the two fractional errors to get the net error. In fact, if the two errors are uncorrelated (that is, there's no reason to expect that an undermeasurement of x is associated with an undermeasurement of y), we should really add the two errors in a root mean square fashion:

$$\Delta A = \Delta(xy) = \sqrt{(\Delta x)^2 + (\Delta y)^2} \qquad \text{for} \quad A = x + y \tag{1}$$

Also, note that if one error is significantly larger than the other, we can get a good idea of the error by just considering the dominant term in such a root mean square or RMS sum.

It turns out that similar rules apply for division errors, addition errors, and power law errors:

$$\Delta A = \sqrt{(\Delta x)^2 + (\Delta y)^2} \qquad \text{for} \qquad A = x + y \tag{2}$$

$$\frac{\Delta A}{A} = \sqrt{(\frac{\Delta x}{x})^2 + (\frac{\Delta y}{y})^2} \qquad \text{for} \qquad A = x/y \tag{3}$$

$$\frac{\Delta A}{A} = \sqrt{2(\frac{\Delta x}{x})^2 + (\frac{\Delta y}{y})^2} \qquad \text{for} \qquad A = x^2 y \tag{4}$$

$$\frac{\Delta A}{A} = \sqrt{\frac{1}{2}(\frac{\Delta x}{x})^2 + (\frac{\Delta y}{y})^2} \qquad \text{for} \qquad A = x^{1/2} y \tag{5}$$

and so on. We can generate rules for other operations (like $\sin()$ and $\cos()$, for example) by using a first-order Taylor series expansion

$$f(x - x_0) \simeq f(x_0) + (x - x_0)\frac{d}{dx}f(x)|_{x=x_0} \tag{6}$$

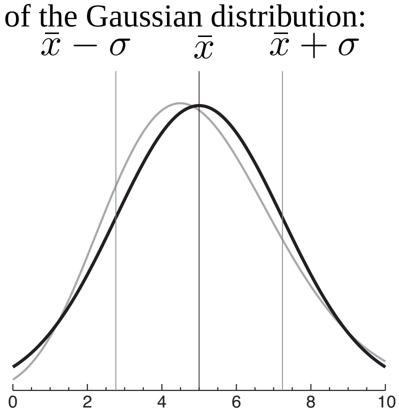
to come up with the right rule for any operation. The assumption here is that the errors are small.

3 Random error

In some measurements, there is a random element involved. Let's say that it is your job to measure the fraction of times that a coin lands face up. You might refuse to make the measurement, saying that you know the answer: it's going to land face up exactly 50% of the time. OK, but what if you make two measurements? If you flip the coin twice, do you expect it to land face up once, and face down once, every time you flip it twice? Of course not! Since each flip of the coin is uncorrelated with the previous flip (the coin has no reason to remember how it landed last time), there is an intrinsic measurement error which for our purposes we can approximate as being equal to the square root of the number of events, or $\sqrt{N}$. That is, if you flip a coin 10 times so that you would expect to have 5 heads, about two-thirds of the time you'll find that the number of heads you get is within the range $5 - \sqrt{5} \simeq 3$ and $5 + \sqrt{5} = \simeq 7$, and that one third of the time you'll get fewer than 3 or more than 7 heads! In other words, you expect to get something like a Gaussian distribution of events about a mean value of $\bar{N}$ of

$$P(N, \bar{N}) = \frac{1}{\sqrt{2\pi\bar{N}}} \exp\left[-\frac{(N - \bar{N})^2}{2\bar{N}}\right] \tag{7}$$

where $P(N, \bar{N})$ means "the probability of measuring N this particular time, in the case where the average of many measurements is $\bar{N}$." This has the same form as $\exp[-(x - \bar{x})^2/2\sigma^2]$, so that $\sigma = \sqrt{\bar{N}}$ characterizes the width of the Gaussian distribution:



That is, about 2/3 of the points will fit between $\bar{x} - \sigma$ and $\bar{x} + \sigma$ on a Gaussian distribution. Now this is only approximately right:

- The Gaussian plotted above is for a continuous distribution in x, whereas we started out talking about integer numbers of events N. Plotting such events should really give a kind of a staircase distribution, because we have only integer value positions on the x axis.

- For events for which no negative value is possible, the real story is that we should use a Poisson distribution (shown in grey above) which is subtly different from a Gaussian distribution for $\bar{x} \lesssim 30$. This is discussed in more detail in Appendix A.
- The real story is that for a Gaussian distribution one can calculate the number of points that fall outside of some multiple of σ by using the error function erf(x) as described in Appendix B.

Still, the above curve gives you an idea.

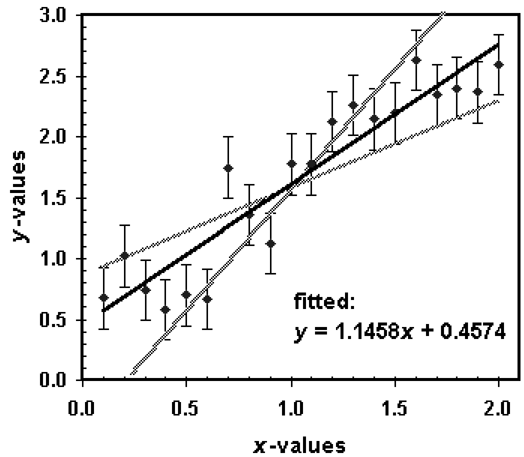
Random error can come in all kinds of measurements beyond simple flipping of coins. For example, when measuring the period of a pendulum, you have to hit a stopwatch with some human reaction time thrown in; this might have a mean (which you can deal with by anticipating the right moment for stopping the watch based on observing the event transpire, or by including it in both the start and the end time), and it might also have a variance σ^2. It's this latter variance that usually factors into measurements you will make.

4 Curve fitting

Frequently in the laboratory you will have the situation that you perform a series of measurements of a quantity y at different values of x. This can often give you a much more accurate determination of a physical parameter than a single measurement will provide, since you are now able to see a correlation of resulting values y over a range of test values x. When you have linear relationship of $y = mx + b$, you can determine the uncertainty in the measured slope m and the intercept b. One way to do this is through a least-squares fit, which is a procedure to minimize the squared error χ^2 of

$$\chi^2 = . \tag{8}$$

Simple least square fits are usually available in spreadsheet programs (it's the `LINEST` function in Excel, for example), and more specialized programs can allow for different uncertainties of the data points and for fits of polynomials, exponentials, Gaussian, and other types of functions to the data. However, if you have no access or experience with spreadsheet programs, you can use a simple, graphical method. Plot the measured points (x, y) and mark for each point (or at least for some representative points) the errors Δx and Δy as bars that extend from the plotted point in the x and y directions. Draw the line that best describes the measured points (***i.e.***, the line that minimizes the sum of the squared distances from the line to the points to be fitted; the least-squares line). This line will give you the best value for slope m and intercept b. Next, draw the steepest and flattest straight lines that are consistent with the error bars of at least two-thirds of the data points. The difference in slope m between the flattest ($m_{\min}$) and steepest ($m_{\max}$) lines should indicate the 2σ range of the slope m, so that the 1σ error Δm in the slope will be given by $\Delta m = (m_{\max} - m_{\min})/2$.



Note that you can also handle relationships other than simple linear by appropriate transformation of the data. If you have measurements of position x versus time t under constant acceleration such that $x = (1/2)at^2$, you can plot $\sqrt{x}$ versus t and obtain a fit of $\sqrt{x} = mt$ in which case you can make the association $a = 2m^2$.

A Comparison of Poisson and Gaussian distributions

If the average number of events obtained in many separate measurements is $\bar{x}$, then the probability of counting x events in one particular measurement is given by the Poisson distribution

$$P(x, \bar{x}) = \frac{\bar{x}^x}{x!} \exp(-\bar{x}). \tag{9}$$

The difficulty in using the exact expression for the Poisson distribution is that both x^x and $x!$ increase dramatically for even modest values of x. We wish to find an approximate expression for the Poisson distribution both for analytical and for numerical calculations. This is done using Stirling's approximation, which says

$$x! \simeq \sqrt{2\pi x}\, x^x \exp\left[-x + \frac{1}{12x} + \mathrm{O}(x^{-2})\right]. \tag{10}$$

We can then approximate the Poisson distribution as

$$P(x, \bar{x}) \simeq \frac{1}{\sqrt{2\pi x}} \left(\frac{\bar{x}}{x}\right)^x \exp[x - \bar{x}] \exp\left[-\frac{1}{12x}\right]. \tag{11}$$

In the case where $x \gg \bar{x}$, the term $(\bar{x}/x)^x$ will be small while the term $\exp[x - \bar{x}]$ will become large. To avoid problems in numerical evaluation of the Poisson distribution, we therefore want to combine these terms as follows:

$$\begin{aligned} \ln\Big\{(\frac{\bar{x}}{x})^x \exp[x - \bar{x}]\Big\} &= x \ln(\frac{\bar{x}}{x}) + x - \bar{x} \\ &= x\,[1 + \ln(\frac{\bar{x}}{x})] - \bar{x}. \end{aligned} \tag{12}$$

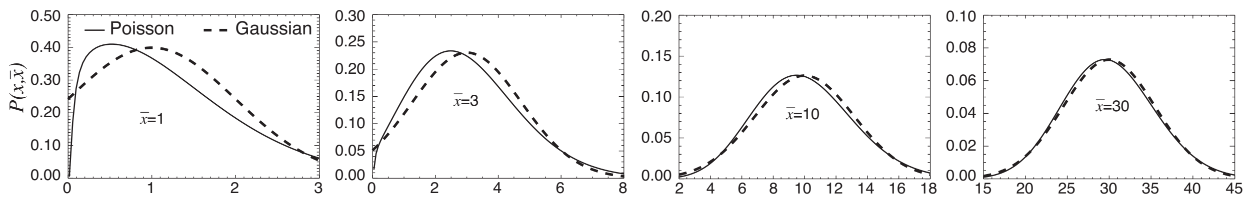
We then write the Poisson distribution as

$$P(x,\bar{x}) \simeq \frac{1}{\sqrt{2\pi x}} \exp\Big\{n[1+\ln(\frac{\bar{x}}{x})] - \bar{x}\Big\} \exp\left[-\frac{1}{12n}\right]. \tag{13}$$

We now have an expression that is more easily evaluated numerically, but it is still too cumbersome to use in simple analytical calculations. For those cases, we can approximate the Poisson distribution with a Gaussian distribution:

$$P(x,\bar{x}) = \frac{1}{\sqrt{2\pi\bar{x}}} \exp\left[-\frac{(x-\bar{x})^2}{2\bar{x}}\right] \tag{14}$$

with a truncation to $P = 0$ for $x < 0$ so as to exclude the incorrect non-zero values of P for negative event numbers x due to the long tails of the Gaussian distribution. Here's a comparison between Poisson and Gaussian distributions for small values of $\bar{x}$:



As you can see, the Gaussian distribution with the limit $P(x < 0) = 0$ is a quite accurate and usable approximation for $\bar{x} \gtrsim 10$.

B The error function erf(x)

The statement that two-thirds of the measurements fall within a range of 1σ of the mean is only approximate. A precise statement is provided by the error function erf(x) of

$$\mathrm{erf}(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x \exp[-t^2]\,dt \tag{15}$$

which has particular values of

σ	x	erf(x)
1	$1/\sqrt{2}$	0.6827
2	$2/\sqrt{2}$	0.9545
3	$3/\sqrt{2}$	0.9973

and values over a larger range as follows: