PHYS 209, 2008/09 Assignment 2

Due 5pm October 31, 2008

Turn in to the box in the hallway outside the lab.

Late assignments not accepted.

Combination of Errors

In this problem set, we will explore why adding errors "in quadrature" is usually the right thing to do.

The first thing you need to know is that **it isn't always the right thing to do.** If the errors in your measurements are correlated (meaning that the error in one value is related in some semi-predictable way to the errors in other values, then simply adding errors in quadrature will almost certainly underestimate the real errors (although it is possible this will overestimate the real errors).

1. Assume that you've measured three quantities, x, y, and z and that the errors in those measurements are independent. The real quantity you are after is given by $p = \sqrt{x/y}\ln z$. Find the formula for the fractional error in p (i.e. $\delta p/p$), in terms of x, y, z and the errors δx, δy, and δz. The answer should be expressed so that you have three terms that each contain only one of x, y, or z.

2. We're going to digress to a very simple case here to see that adding errors in quadrature "does the right thing." Imagine that you have measured f_1 and f_2. You need to add these two quantities: $f = f_1 + f_2$. Adding the errors in quadrature, for this case we find: $\delta f^2 = \delta f_1^2 + \delta f_2^2$.

   To see that this gets the error right, we need to consider what "getting it right" means. Let's start with the meaning of δf_1. If you were to do many measurements of the quantity f_1, then δf_1 should be the standard deviation of the distribution of f_1 measurements. Same for δf_2. So δf should be the standard deviation of the distribution of f values obtained from many measurements of f_1 and f_2.

   What we want to do below is to simulate many measurements of f_1 and f_2, look at the distributions of our simulated measurements, and then look at the distribution of f values we get from adding $f = f_1 + f_2$. At the end we hope to find that the distribution width of f values is given by $\delta f^2 = \delta f_1^2 + \delta f_2^2$.

   (a) Take $f_1 = 98.0000$ and $f_2 = 102.000$ as "the truth." The experimental uncertainty in our measurements is $\delta f_1 = \delta f_2 = 5$. Now, let's imagine we make a large number of independent measurements of these two quantities. Do this by generating simulated data on the computer. This exercise is probably most easily done in a spreadsheet (or in a small computer program if you are able to write one). You need to generate random numbers that are distributed according to a Gaussian distribution. Most computer random number generators provide what are called "uniform deviates," which are uniformly distributed between 0 and 1.

You can generate Gaussian distributed random numbers with the following formula:

$$g = \sigma\sqrt{-2\ln x_1}\cos(2\pi x_2),$$

in which σ is the standard deviation of the distribution, and x_1 and x_2 are uniform deviates between 0 and 1 (see Numerical Recipes for a proof).

Generate at least 2000 values for each of f_1 and f_2 and plot histograms of these simulated data to ensure we actually did generate Gaussian distributed data.

You can create the histogram data from your simulated measurements many different ways, but one way is right in the spreadsheet. An example formula to use is:

`=SUM((A$1:A$2000>=$D3)*(A$1:A$2000<$D4))`

This formula will find the number of cells in the range A1:A2000 that contain values lying between the values specified in D3 and D4. **This formula needs to be entered in a special way!!!**. If you type in the formula and press enter, it won't work the way we want it to. Type in the formula, then hold down shift and control and then press enter. This tells the spreadsheet that your formula is an "array formula." Feel free to search on Google if you'd like to know more about array formulae. Our formula uses the `SUM` function because the result of each comparison is either true (1) or false (0), and we simply add up the 1's to get a count of those cells that satisfy both criteria. Plot separate histograms for each of f_1 and f_2 using a bin width of 1. Don't print them out just yet - we'll need to add to them.

Fit each of the two distributions to Gaussian curves:

$$f(x) = A\exp(-(x - x_0)^2/2\sigma^2).$$

If you calculated your simulated data correctly, you should have a good fit, and the x_0 and σ values should reflect our earlier choices for "the truth" and the measurement error.

Turn in histograms, along with the best fits to the Gaussian curves. Do your fit parameters make sense?

(b) Now go back to the spreadsheet and calculate $f = f_1 + f_2$, and generate a histogram from these values. Again fit the histogram to a Gaussian. What is the standard deviation of this distribution? Does it match what you expect?