CBE 20258 Algorithm Problem Set 3

Due in class 2/24/17

Problem 1). Radiocarbon dating is a classic technique for determining the age of once living samples such as wood, pottery (the organic material in it) or shells. The idea is that an organism sequesters carbon from the atmosphere, and this process stops after death. Interestingly, due to a balance between the solar neutron flux and the rate of radioactive decay, the atmospheric isotopic concentration of C14 is reasonably constant (or it was until the first atmospheric A bomb test, anyway). Thus, by measuring the ratio of the radioactivity of a sample to that of a standard you can calculate when it stopped sequestering carbon. The formula for the age is given by:

$$t = \frac{t_{1/2}}{\ln(2)} \ln\left(\frac{r_0 - r_b}{r_s - r_b}\right)$$

where $t_{1/2}$ is the *half-life* of the C14 isotope, and has a value of 5730 years. The three count rates r_s, r_b, and r_0 are the observed rates of the sample, background, and the standard. These are given by:

$$r_s = \frac{N_s}{T_s} \quad ; \quad r_b = \frac{N_b}{T_b} \quad ; \quad r_0 = \frac{N_0}{T_0}$$

where the total number of counts observed in time T_s, T_b, T_0 were N_s, N_b, and N_0. In my high school science project I worked out a way to do this with sample processing in my basement and counting using borrowed time on various liquid scintillation spectrometers in the DC area. Typical count times were about 200 minutes each (e.g., things were set up to run overnight), and the average count rates of background and standard were 32cpm and 46cpm, respectively. The sample rate varied according to the age of the sample.

A curious feature of radioactive decay is that it is characterized by the Poisson distribution, where the variance in the number of counts equals the number of counts (e.g., $\sigma_N = N^{1/2}$). So now we come to the problem:

a. A researcher reports an age for a sample of 8,000 years, but doesn't give the confidence interval (a cardinal sin in this field). Using the error propagation method described in class, calculate the 95% error bounds on this reported age. Use 200 minutes as the count times for all three needed rates. You can do the derivatives analytically, or you can use the finite difference approach in matlab described in class. It is particularly simple if you use the matlab anonymous function syntax.

b. Repeat the confidence interval calculation for a sample which is reported to be 30,000 years old. Why are these error bounds weird?

c. Redo the calculation for parts a & b using a matlab simulation: using the random number generator, simulate the experiment for 10000 trials, sort the calculated ages (a vector), and then determine the age of the 250th and 9750th trial (e.g., the error bounds). Compare these to the values calculated by the formula in a and b and comment. Note:

if, in a trial, the sample rate is less than the background rate, the age is indistinguishable from infinite age.

Problem 2). On the first algorithm assignment you determined the optimum depth of the second probe in a two-probe system for calculating the temperature gradient at the surface. Suppose your boss says that this isn't good enough, and suggests adding in a third probe (e.g., one at a depth h and the second at a depth 2h). Take the expected temperature profile to be about T_0 + A*exp(b*x) where A = 2°C and b = 0.5 cm^{-1}. The probes have a random error of ±0.1°C.

a. Redo the calculations for the two probe system, this time calculating the random experimental error using the correct error propagation formula used in class and the expected temperature distribution given above. What is the optimum value of h, and what is the best we can do calculating the derivative?

b. For your three probe system you will not want to weight the probes evenly. Determine the optimum weighting of the three probes for getting the surface derivative (Hint: you need to keep another term in your Taylor series expansion for the algorithm error, and try to get the quadratic term exactly!)

c. Determine the random experimental error for this new three probe formula as a function of h, and determine both the optimum h and the minimum derivative measurement error.

Problem 3). You are using a falling-ball rheometer to measure the viscosity of a liquid. The fluid is sufficiently viscous that the Stokes sedimentation equation described in class is assumed to apply (you will learn all about that next term). The experiment consists of measuring how long (t) it takes a sphere of radius (a) to fall a distance (L). The sphere density is (ρ_s) and the fluid density is (ρ_f). The velocity U_s = L/t. We have the measurements (regarded as independent):

$$U_s = \frac{2}{9}\frac{\left(\rho_s - \rho_f\right) g\, a^2}{\mu}$$

	value	1σ
a	$100\,\mu m$	$1\,\mu m$
ρ_s	$1.18\,g/cm^3$	$0.01\,g/cm^3$
ρ_f	$0.98\,g/cm^3$	$0.002\,g/cm^3$
g	$980.6\,cm/s^2$	0
L	$2\,cm$	$0.01\,cm$
t	$230\,s$	$0.5\,s$

a. What is the 2σ error bound for the measured viscosity (μ)? Report this in units of g/cm-s.

b. Which measurement had the largest contribution to the error in the final result (e.g., which measurement would you want to spend some time and money improving)?

Hint: It is much easier if you do this one with variables using the computer. I would write a short function (anonymous functions work fine here, and are very convenient – if you haven't used them yet, try googling "matlab anonymous function" and read up on it!) which takes in all the parameters as a vector and spits out the viscosity. You can then take the derivative using the finite difference method as discussed in class and calculate the error. You can also see which term has the largest contribution to the overall error (e.g., part b). You can do it all using pencil and paper, of course: your choice.

Problem 4). In the second project last year (with solution given in last week's tutorial), students found that the average Notre Dame high temperature excursion in 2015 was 0.937 degrees above the historic averages, and the standard deviation was 9.938 degrees.

a. If we assume that all the days are independent, what is the probability of the average excursion being this large or higher?

b. For an exponential decay in the correlation of the residual (evenly spaced observations), such as was observed for all the cities, the correct number of independent data points in an average is reduced (roughly) by the factor $1/(1 + 4*\text{correlation length}^2)^{1/2}$. In this formula the correlation length is the dimensionless value (e.g., units of "indices", or correlation length in days divided by the time between observations), and is the inverse of the exponential decay rate. Based on this, what should be the correct answer to part a?

What you should get from these questions:

This week's algorithm assignment has a mixture of straight up pencil & paper questions as well as those which are most easily addressed using the computer. All of them deal with error propagation and analysis in one way or another.

1. This problem has lots of parts to it: it introduces you to the Poisson distribution found in radioactive decay (as well as many other examples, such as counting tracers to get concentrations, etc.). It demonstrates the application and limitations of the standard error propagation formula. It introduces you to the idea of a Monte Carlo simulation: the use of a random number generator to perform numerical experiments. As computers get faster, this is an increasingly valuable technique. Finally, it gets you to use the "anonymous function" utility for matlab, which is a very convenient tool.

2. This a more sophisticated version of the problem you studied in the first algorithm assignment, requiring a bit more mathematics (extra term in the Taylor series).

3. Here we are both doing error propagation as well as using it to do a sensitivity analysis. A lot of times you will need to answer the question of "where to I spend my time/money to get the greatest improvement in my answer?" For error minimization it all boils down to determining which variable has the largest contribution to the overall error, which is the one you want to dial down first. Error propagation is a great way to determine this, as you can get the contribution of each measurement to overall error as a vector.

4. Stats packages tend to assume that error is independent, often leading to an underestimation of the standard deviation of the final result. Time series data is almost never independent – particularly when the variation is much slower than the sampling time. This can have profound effects on conclusions drawn from the data. Here we use our modeled covariance to determine the effect on probabilities.