CBE 20258 Numerical and Statistical Analysis

Algorithm Problem Set 5

due in class Thursday, 3/31
(or in a box in the CBE office by 4pm 4/1)

Problem 1. Systems of Equations and Scaling: The solubility of CO_2 in water is a strong function of the pH, due to the conversion of dissolved CO_2 to carbonic acid (H_2CO_3) and dissociation into HCO_3^- and CO_3^{2-}. We are tasked with determining the equilibrium concentration of all dissolved species in a 0.001 M NaOH solution. Assuming that the NaOH completely dissociates (it is a very strong base), we want to calculate the pH and species concentrations for a solution in equilibrium with a 0.5 atm CO_2 gas (e.g., pCO_2 = 0.5 atm).

a. We have the following equilibrium relationships:

$$\frac{[CO_2]}{pCO_2} = K_{abs} = 3.36x10^{-2}\ M/atm \qquad \frac{[H_2CO_3]}{[CO_2]} = K_{eq} = 1.7x10^{-3}$$

$$\frac{[HCO_3^-][H^+]}{[H_2CO_3]} = Ka_1 = 2.5x10^{-4}\ M \qquad \frac{[CO_3^{2-}][H^+]}{[HCO_3^-]} = Ka_2 = 4.8x10^{-11}\ M$$

$$[H^+][OH^-] = K_w = 10^{-14}\ M^2$$

Using this, write out the relevant equations which must be solved to obtain the concentrations, defining all variables. Don't forget the electroneutrality condition! The net charge of any aqueous solution must be zero – all charges must balance.

b. How should you recast these equations to get a set of equations that can be easily solved in Matlab? (Hint: in addition to the tricks used in the homework earlier this term, two of the variables can be eliminated right off analytically. Remember that it is always a good idea to reduce the dimensionality of a system!)

Problem 2. Optimization: We examine the planning of an orchard of cherry trees. Suppose we have $20,000 to invest in planting cherry trees. We can plant a mix of sour (dessert) cherries and sweet cherries (the ones you like to eat). The sour cherry trees cost $20 each, and the sweet cherry trees cost $10 each. The sweet cherries, however, can be sold for $100 per tree per year when mature if fully pollinated, while the sour cherries can be sold for only $25 per tree when mature. Finally, the yield per tree of the sweet cherry trees depends on their proximity to the sour cherry trees, since they are not self-pollinating. If the number of sour trees is given by x_1 and the number of sweet trees is x_2, we shall take the yield (the fraction of the fully pollinated yield) of the sweet trees to be $x_1 / (x_2 + x_1)$.

Show how you would calculate the optimum number of sweet and sour cherry trees. Be as specific as possible, defining all functions and techniques employed.

Problem 3. Optimization and Error: A farmer friend of yours is trying to figure out how to reduce the uncertainty in his farming income using futures contracts and has come to you for advice. A futures contract is an agreement to sell some quantity of crops at a preset price, with the crops and money actually being exchanged at harvest. This has the advantage of "locking in" current (usually average) prices, but has the disadvantage that if he doesn't actually produce as much as he has contracted to sell, he then has to purchase the difference on the open market at harvest time. Now for the problem:

The farmer produces x_1 bushels of corn, a random variable with a standard deviation of σ_{x_1}. The price he will get per bushel for this corn at harvest time is another random variable x_2 with a standard deviation of σ_{x_2}. Because prices tend to be high when crop yields are low, there is a negative covariance between these variables:

$$\sigma^2_{x_1 x_2} = -\lambda\, \sigma_{x_1}\, \sigma_{x_2}$$

where λ is a positive number between zero and one. The price he will get per bushel via the futures contract is a fixed value p, and the number of bushels sold in this contract is z, the variable you wish to determine. For the purposes of this problem we shall take the expectation value of the price at harvest time (e.g., $\mu_{x_2} = E(x_2)$) to be the same as the futures contract price p. You may use the following values:

$$E(x_2) = p = \$2.50 \ / \ \text{bushel} \qquad \sigma_{x_2} = \$1.00 \ / \ \text{bushel}$$
$$E(x_1) = 10{,}000 \text{ bushels} \qquad \sigma_{x_1} = 2{,}000 \text{ bushels}$$
$$\lambda = 0.5 \ \ (\text{e.g., } \sigma^2_{x_1 x_2} = -\$1{,}000)$$

a. Determine the farmer's expected revenue and standard deviation as a function of z. It is actually *much* easier to do this with the variables first rather than substituting in the numbers right off.

b. Determine the value of z which minimizes the uncertainty in the farmer's income. By how much is the uncertainty reduced over the case z = 0 (no futures trading)?

c. How might your recommendations change if the futures price p is less than the farmer expects to get at harvest (e.g., $p < E(x_2)$)? Please be brief.

Problem 4. Regression and Error: In Senior Lab students measure the natural convection due to a heated wire. Basically, the wire heats the fluid, the fluid expands, and the fluid in the vicinity of the wire rises due to buoyancy forces - just like the draft rolling off a cold window. Theory predicts that the velocity directly over the wire should increase very slowly as we move higher according to the power law relationship:

$$u = C\, x^{s}$$

where x is the height and u is the velocity. The exponent s has a theoretical value of 0.2. A student has made the following set of measurements:

height	*velocity*
1	1.006
2	1.160
3	1.305
4	1.424

Does the student's power law exponent match theory to within experimental error at the 95% confidence level? I want you to carefully define all matrices, assumptions, etc. You can get the final number by hand, however a simple matlab program is much easier!

Problem 5. Model Linearization: In the second project you used three values to determine the time at which a maximum correlation occurred. Here we try seeing what using five values an a bit of regression can do! Suppose we have five measurements of y_i at times t_i:

We are fitting this to the model $y = y_m - a\,(t - t_s)^2$, where the goal is to determine t_s.

a. Set up this problem as a linear regression problem (e.g., linearize it!), clearly identifying all variables and matrices used.

b. Show how you can calculate the matrix of covariance of the fitting parameters, and how these can be used to get the standard deviation of t_s, the quantity of interest.

c. Show how you can calculate the 95% confidence interval for t_s.

Extra Credit: 0 to 4 points added to your mid-term exam grade "above the line" to make it worth your extra effort!

We have stated that the t-distribution is appropriate for connecting the probability distribution to the sample standard deviation for small sample sizes. Here we use a Monte Carlo simulation to prove it! Consider the problem above where we have exact values of $y_m = 1$, $a = 2$, and $t_s = 0.5$. We can generate "data" for t = [-2:2]' (e.g., five values as a column vector) by plugging into the formula above with these values for y_m, a, and t_s. We can also add in a random vector of noise characterized by randn(5,1)*0.1 (e.g., not too big, not too small). Code up the solution you sketched out to the above problem to get the confidence interval for t_s from this vector of generated data and determine whether this contains the "true" value ("if" statements are useful here!). Now put the thing inside a loop (doing it about 10,000 times for different random noise vectors) and determine the fraction of time the true value of t_s falls outside the calculated intervals. If you do it right, it should fail about 5% of the time – largely due to the occasional underestimation of the standard deviation due to the finite sample sizes used.