CBE 20258 Numerical and Statistical Analysis

Algorithm Problem Set 5

due in class Thursday, 4/6
(or in a box in the CBE office by 4pm 4/7)

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 0.01 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 0.2 atm CO_2 gas (e.g., pCO_2 = 0.2 atm). This would roughly be the concentration of CO_2 in a stoichiometric burning of coal in air with all oxygen consumed, for example, so it would be a measure of the use of this base solution for carbon capture in a coal fired power plant.

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. Recast these equations to get a set of equations that can be easily solved in Matlab. If you do it right, you could even get it into the form that you could set it up as an anonymous function (although a function file is probably easier). (Hint: in addition to the tricks used in the homework earlier this term, two or three of the variables can be eliminated right off analytically. Remember that it is always a good idea to reduce the dimensionality of a system! Don't forget how to condition this sort of problem either!)

c. Using fsolve, get the concentrations! Report these concentrations as well as the total carbon concentration in the solution.

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 $x_1 / (x_2 + x_1)$. Develop an appropriate objective function and determine the optimum values of x_1 and x_2. While you can do this analytically, the math gets a bit messy: it is much easier to do it on the computer.

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. In general, the futures price (or at least that received by the farmer) is less than the expected price at harvest (e.g., $p < E(x_2)$). This means that there is a cost to using futures hedging. Calculate the expected return and standard deviation for $\$2<p<\$2.5$ and graphically compare the distributions (you can use the normpdf command to generate the distributions once you know the mean and sd) for selected values in this range. How does this plot affect your recommendations?

Problem 4. Model Linearization: On the mid-term you linearized (or tried to, anyway) the model for the viscosity of a fluid as a function of measured torque with an unknown gap width error given by:

$$\mu = T\left(h + \Delta h_0\right)$$

Where T is the measured torque and h is the measured gap width (thus, there will be a set of each). The unknown parameters are the viscosity μ and the gap width error Δh_0. Here you are going to do it right lots of different ways!

a. Develop four different linearizations that obey the ten commandments of linear regression, clearly identifying A, x, and b for each. One of them may be really weird, but all should obey the rules and all should work. There are actually an infinite number of ways you can do this, but some make much more sense than others.

b. Ordinarily, regression is set up to regress one parameter (which has error) on another (which is assumed not to). The matrix A thus ideally shouldn't involve the variable with error in it (in this case, T). What linearization (likely one of the ones above!) satisfies this criterion and makes the b vector solely a function of T? What do the assumptions of unweighted regression imply about the standard deviation of T for this case?

c. In this problem Δh_0 is typically very small. This means that T is measured to be inversely proportional to h. If the standard deviation of T is roughly proportional to T (finite precision), which of your linearizations are suitable for unweighted linear regression?