Lab Worksheet #3: MATH 495

Materials: This worksheet accompanies the m-files found on the web site, which you should download into a directory on your computer. You will also need to download Professor Haario's `mcmcstat` toolbox into the same directory, then use the `addpath` command to point MATLAB towards the mcmcstat directory.

1. (a) First, let's have a look at the function `bodex.m`, which you have seen before. This code is just a stripped version of code from Professor Haario's lectures.

   (b) Next, run `aidsrun_temp.m`, and have a look at the least squares solution for both the first 12, and the last 12, years.

   (c) **Your task** is to add MATLAB commands to `aidsrun_temp.m`—using `bodex.m` as a template—in order to implement Professor Haario's tools for MCMC sampling for the aids data problem. Implement the sampling both with and without sampling the noise variance σ^2. In predicting future cases of aids, which methodology do you think gives the most useful results?

   (d) Using the finished code (`aidsrun.m`) posted on the web (or those you wrote yourself for the previous part) compute confidence intervals for the parameters from the MCMC chains for both 1981-92 and 1993-2004.

2. **Homework Problem:** Next, consider the following data from the 1949 U.S. polio epidemic, which was the second worst in history. The data consist of the cumulative number, or incidence, of polio cases diagnosed on a monthly basis.

month	0	1	2	3	4	5	6	7	8	9	10	11
cases	494	759	1016	1215	1619	2964	8489	22377	32618	38153	41462	42375

   (a) Fit the data with the standard SIR model,

$$\begin{aligned} \frac{dS}{dt} &= -aIS, \quad S(0) = S_0, \\ \frac{dI}{dt} &= aIS - bI, \quad I(0) = I_0, \\ \frac{dR}{dt} &= bI, \quad R(0) = R_0. \end{aligned}$$

   where this time R corresponds to the cumulative number of cases at time t. Assume that there is a baseline (or steady state) level of polio that is well-represented by the initial number of cases 494 (R(0)=494). Also assume that the epidemic is caused by a particularly virulent strain of polio in a single individual (I(0)=1) and is limited to a susceptible subpopulation of the U.S. I will show you how to estimate the size of this group (S(0)). Use initial values of $a = 0.0002$ and $b = 0.1$ in your optimization routine. Next, perform MCMC sampling to sample from the posterior distribution of the parameters. Implement the sampling both with and without sampling the noise variance σ^2. Compute confidence intervals for the parameters from the MCMC chains.

   (b) Let's modify the standard SIR model so that it takes into account the fact that once a patient is diagnosed with polio, he/she remains highly infectious for several days. Test your own modification of the model. Does it fit the data better? Use an MCMC analysis to determine parameter correlations in your new model. Do you think that the additional parameter is worth it?