STA 410/2102, Fall 2015 — Assignment #3

Due at the start of class on December 8. Please hand it in on 8 1/2 by 11 inch paper, stapled in the upper left, with no other packaging.

This assignment is to be done by each student individually. You may discuss it in general terms with other students, but the work you hand in should be your own. In particular, you should not leave any discussion with someone else with any written notes (either paper or electronic).

In this assignment, you will use Gibbs sampling to find the posterior distribution of the parameters for the same model for data on beetles as was used for Assignment #2.

To simplify the assignment, the α parameters, giving the abundances of the ten species of beetle, should be considered known, with the following values:

s:	1	2	3	4	5	6	7	8	9	10
α_s:	0.05	0.20	0.05	0.06	0.04	0.15	0.05	0.15	0.20	0.05

The set of model parameters therefore consists of μ_s, ν_s, and ρ_s for $s = 1, \ldots, 10$, as described in Assignment #2. (Note that in a real problem the species abundances, and also the standard deviations for the normal distributions used, would be parameters rather than being assumed known, and when they are unknown it is still possible to sample for the posterior distribution using Gibbs sampling, if they are given prior distributions that are conjugate.)

You should use a prior distribution for the model parameters in which all parameters are independent, the μ_s and ν_s have the $N(1, 2^2)$ distribution, and the ρ_s have the uniform distribution on $(0, 1)$, which is also the Beta$(1, 1)$ distribution.

You should write an R function that performs Gibbs sampling for this model, taking as arguments the data, an initial set of parameter values, and the number of Gibbs sampling iterations to do. Each iteration of Gibbs sampling should first update the unknown species indicators, and then each of the parameters in turn. Updating the species indicators first avoids the need to provide initial values for them. The order in which the parameters are updated is arbitrary, since they are independent conditional on the species indicators. You may find that defining other functions makes writing the Gibbs sampling function easier, and if you like you may reuse any functions you wrote for Assignment #2, or the functions from the solution to Assignment #2 from the course web page, or modified versions of them. You should put all these function definitions in a source file separate from the R script file. This file of function definitions should not refer to the particular data set you are analysing.

In a separate R script file, you should run your Gibbs sampling function to obtain a sample from the posterior distribution of the parameters, given the data from the course web page. You should produce trace plots showing the parameter values along the chain, and the values of the species indicators for a few of the 500 beetles that you choose as representative. You should also output estimates for the posterior means and standard deviations of the parameters, found using the values from the chain after discarding an appropriate number of initial "burn-in" iterations. You may produce additional output (of reasonable size) if you think it is helpful.

You should hand in your derivations of the conditional distributions needed for Gibbs sampling (which may be hand-written), along with the source for your functions, the source and output (text and plots) of your script (which may be one file, if you use `knitr::spin`), and a discussion of how well Gibbs sampling seems to work for this problem (eg, how fast it converges, and how dependent the subsequent points from the Markov chain are).