Penn State STAT 540
Homework #1, due Thursday, March 15, 2018 at midnight

What you have to submit in a Canvas submission folder: (i) Your `R` code, (ii) a pdf file that contains a clear writeup for the questions below. Note that your code should be readily usable, without any modifications. For example, to read in the data make sure the code accesses the website rather than a directory on your computer.

1. (Return to the previous homework problem.) Let $X_1, \dots, X_n$ iid from Gamma(α, β) distribution. Consider Bayesian inference for α, β with prior distributions $\alpha \sim N(0, 3)$, $\beta \sim N(0, 3)$. The data for this problem are available here: `http://personal.psu.edu/muh10/540/Rcode/hw1.dat`

   (a) Use importance sampling to approximate the expectations. Provide pseudocode, including relevant distributions. Is the variability of your Monte Carlo approximation guaranteed to be finite? Explain your answer. Provide Monte Carlo standard errors for your estimates.

   (b) Construct an all-at-once Metropolis-Hastings (AMH) algorithm to approximate the posterior expectation of α, β. Provide pseudocode for your algorithm. This should include any distributions you had to derive. How did you determine the Monte Carlo approximations were good enough? That is, discuss stopping criteria (how you determined chain length), the number of starting values you tried, how you obtained initial values etc.

   (c) Construct a variable-at-a-time Metropolis-Hastings (VMH) algorithm. You need to provide the same level of detail here as you did for the previous algorithm.

   (d) Provide a table with all approximations along with any error approximations, as well as the computational time taken by the algorithms. (To make this easy to read, the rows of your table should correspond to algorithms.)

   (e) Compare the efficiency of the four algorithms – importance sampling, AMH, VMH, and the Laplace approximation from the previous homework – using the methodology discussed in class, e.g. effective sample size, effective samples per second etc. The Monte Carlo algorithms are easier to compare than the comparison between them and the Laplace approximation; you may have to think carefully about this. Be sure to clearly explain why your approaches for comparing the algorithms are reasonable.

   (f) Which algorithm would you recommend for this problem? How would you order the algorithms in terms of ease of implemention?

2. Lightbulbs: Assume that lightbulb lifetimes for a lightbulb made by a particular company are independent and exponentially distributed with

expectation θ. Suppose in an experiment, m bulbs are switched on at the same time, but are only completely observed up to time τ. Let the lifetimes of these bulbs be $A_1, A_2, \ldots, A_m$. However, since the bulbs are only observed till time τ, not all these lifetimes will be observed. Now suppose that at time τ, the experimenter observes the number of lightbulbs, W, still working at time τ, and the lifetimes of all lightbulbs that stopped working by τ. For convenience, denote these bulbs that stopped working by time τ as $A_1, A_2, \ldots, A_{m-W}$. Hence, the missing information consists of the lifetimes of the lightbulbs still working at time τ, $A_{m-W+1}, \ldots, A_m$. For a particular experiment, let τ be 184 days and $m = 100$. The observations for this experiment are as follows. The data on the lightbulb lifetimes for the bulbs that stopped working by τ are here: `http://personal.psu.edu/muh10/540/hwdir/bulbs.dat` (Assume that the remaining bulbs were still working at time τ.) Let the prior for θ be Gamma(1,100) (with parameterization so it has mean 100 and variance 100^2).

(a) Using auxiliary variables for the missing lightbulb data, construct an MCMC algorithm to approximate the posterior distribution of θ. Provide the same level of detail about your MCMC algorithm as you did in the previous problem.

(b) Construct a different MCMC algorithm, this time without using auxiliary variables/data augmentation. Again, provide details.

(c) Overlay posterior density plot approximations for the two algorithms. Provide a table that shows the posterior mean approximations for θ along with MCMC standard errors.

(d) For the auxiliary variable method, plot the approximate posterior pdf for one of the "missing" lightbulbs, then overlay the approximate posterior pdfs for a lightbulb made by the company. You should notice that they are different. Report the posterior mean estimates for each of them.

(e) Looking at just the ability to approximate the posterior per iteration of the algorithm, which of the two MCMC algorithms is more efficient? Now accounting for computing costs, which of the two MCMC algorithms is more efficient? Which algorithm would you recommend?

(f) This course is focused on computing but it is worth noting some basics about inference. Compare your results above with what would happen to inference if you ignored the missing data by overlaying the density plots.