STAT 515
Homework #11, due Monday, Apr. 23 at 2:30pm
This homework must be submitted electronically to ANGEL. I strongly encourage the use of LaTeX.

Please make every assignment easier to grade by neatly organizing your writeup and clearly labeling your final answers when appropriate. Try using LaTeX!

1. Suppose that $Y_1, \ldots, Y_n$ is a simple random sample from $N(\theta, 1)$, where we assume that $(\log \theta - \mu)/\sigma$ has a t distribution on r degrees of freedom (this is the example we discussed briefly in class on April 11). Assume that μ, σ, and r are known (parameters on the prior distribution like this are called *hyperparameters* in a Bayesian context).

   (a) Describe a Metropolis-Hastings algorithm for sampling from the posterior distribution of $\theta \mid \mathbf{Y}$, using a normal proposal distribution centered at the current value of the Markov chain and with variance τ^2.

   (b) Take $\mu = 0$, $\sigma = 5$, and $r = 4$. For the dataset $Y_1, \ldots, Y_{100}$ at `http://sites.stat.psu.edu/~dhunter/515/hw/hw11prob1b.txt`, implement your M-H algorithm starting the chain at $\theta_0 = 1$ and running for 50,000 steps. Use $\tau^2 = 1$.

   (c) Record the acceptance rate of your MH algorithm. Then create a trace plot in which you plot the values of θ_i against i. Comment on it: Does it appear that your Markov chain is effectively "mixing"?

   (d) Create a histogram of the θ_i values. Also, report a point estimate (the posterior mean) along with a 95% credible interval for θ based on your posterior sample. (For the latter, just use the 0.025 and 0.975 sample quantiles of your sample.)

   (e) To create a 95% confidence interval for the true posterior mean, a naive idea would be to try

   $$\hat{\mu} \pm \frac{1.96}{\sqrt{n}} * \sqrt{\frac{1}{n-1} \sum_{i=1}^{50,000} (\theta_i - \hat{\mu})^2},$$

   where $\hat{\mu}$ is your point estimate from part (d). Explain why this interval has a totally different interpretation than your interval from part (d). (Hint: Only one of these intervals tries to capture the MCMC error.) Also, explain why this idea is likely to produce an interval that is narrower than it should be—and hence, it is not a very good idea from a statistical point of view.

   (f) Replicate part (c) using a value of τ^2 that appears "too small". Then do the same thing for a value of τ^2 that appears "too big". In each case, try to explain why the chain does not appear to be as effective as in part (c).

2. The goal of this problem is to analyze a dataset in which the observations are assumed to come from a piecewise-homogeneous Poisson process in which the Poisson rate starts at one value and then changes at some unknown time to a different value.

   The data are at `http://sites.stat.psu.edu/~dhunter/515/hw/hw11prob2.txt`, where the events have been binned into 50 time periods of equal length. A model, adapted by Murali Haran from Chapter 5 of "Bayes and Empirical Bayes Methods for Data Analysis" by Carlin and Louis (2000), for the binned counts $Y_1, \ldots, Y_{50}$ is as follows:

   $$Y_i \mid k, \theta, \lambda \sim \begin{cases} \text{Poisson}(\theta) & \text{for } i = 1, \ldots, k; \\ \text{Poisson}(\lambda) & \text{for } i = k+1, \ldots, 50. \end{cases}$$

The prior distributions on the k, θ, and λ parameters, along with two hyperparameters b_1 and b_2, are as follows:

$$\begin{aligned}
\theta \mid b_1 &\sim \text{Gamma}(0.5, b_1) \\
\lambda \mid b_2 &\sim \text{Gamma}(0.5, b_2) \\
k &\sim \text{Unif}\{1, \ldots, 50\} \\
b_1 &\sim \text{Inverse Gamma}(0, 1) \\
b_2 &\sim \text{Inverse Gamma}(0, 1),
\end{aligned}$$

where b_1 and b_2 are independent and k, θ, and λ are conditionally independent given b_1 and b_2. The density functions of the $\text{Gamma}(\alpha, \beta)$ and $\text{Inverse Gamma}(\alpha, \beta)$ distributions are, respectively,

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} \quad \text{and} \quad f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{-\alpha-1} e^{-1/(x\beta)}.$$

Please take note: The stated priors for b_1 and b_2 are not proper probability distributions: Integrating $\exp\{-1/b\}/b$ from 0 to ∞ does not converge to a finite value. However, if you simply use the improper prior density $\exp\{-1/b\}/b$, you will obtain proper full conditional distributions for all of the parameters.

Specific instructions are as follows:

(a) Derive the full conditional densities or mass functions (up to a constant) for each of the five parameters conditional on the other four and the data. For θ, λ, b_1, and b_2, describe these full conditionals as coming from some named parametric family, and give the parameters.

(b) Implement a one-variable-at-a-time Metropolis-Hastings algorithm to sample from the posterior distribution. For θ, λ, b_1, and b_2, use Gibbs sampling from the full conditionals; for k, use a Metropolis-Hastings update where the proposal distribution is uniform on $\{2, \ldots, 49\}$. Run at least one million iterations of the algorithm.

(c) Give a 95% credible interval for k (use the 0.025 and 0.975 quantiles of the posterior distribution of k). Then, produce a plot of the data (Time vs. Count) and overlay the mean of the Poisson process on the same plot, where this mean is determined by the posterior means of k, θ, and λ.

If you are interested in seeing an application of the model in this problem to a real dataset, or if you simply get stuck on this problem, you might find the writeup by Murali Haran at `http://sites.stat.psu.edu/~mharan/MCMCtut/MCMC.html` to be helpful.