Stat238: Problem Set 3
Due Wednesday Oct. 26

October 8, 2016

Comments:

- Please note my comments in the syllabus about academic integrity. You can work together on the problems, but your final writeup must be your own and you should wrestle with the problems on your own first.
- It's due at the start of class on paper. The syllabus discusses the penalty for turning it in late.
- Please give the names of anyone you worked with on the problem set on what you hand in.
- Please include your code in your solution. For extensive code that doesn't naturally fit within your response to a given problem, please include the code in an Appendix.

Problems

1. This problem will consider the influence of prior distributions and what happens when the posterior is improper.

   (a) For the rats data from Lab 6, derive the full conditionals for a "Gibbs" sampler that samples from the closed-form conditional distributions. Use the model from Lab 6, but with inverse gamma priors for the various σ^2 terms. You can use results from the book or class to avoid any derivations in which you recognize what the conditional distribution must be.

   (b) In PS2, we saw that $p(\sigma) \propto 1/\sigma$ (equivalent to $p(\sigma^2) \propto 1/\sigma^2$, which can be written as an improper $IG(0,0)$ prior) results in an improper posterior. Code up your own Gibbs sampler or use NIMBLE with a user-defined distribution (see class demo code) so as to be able to use an improper prior. Run the sampler on the rats data and assess convergence based on the trace plots; do they give any indication of the impropriety?

   (c) Consider the solution to BDA 5.10a from PS2. This is not exactly the same model, but the reasoning about why the posterior is not integrable should still hold. In light of where the infinite mass in the posterior is, can you explain your results from part (b)? Under what situations do you think the sampler would give an indication of the impropriety?

2. This problem will explore the effect of using different MCMC algorithms on MCMC performance. Consider the following model, known as the *litters* example in the BUGS examples (and with data in *litters.csv*). There are $J = 2$ groups of rat litters, with $N = 16$ litters (i.e., mothers) in each group, and the number of pups in litter i in group j is n_{ij}. The number of pups that survive in litter i, j is y_{ij}. Survival of the pups in a litter is governed by a survival probability for each litter, p_{ij}, and the

probabilities for the litters within a group are considered to come from a common distribution, $p_{ij} \sim \text{Beta}(a_j, b_j)$,thereby borrowing strength across the litters in a group. We'll use the prior distribution from the original BUGS example, $a_j \sim \text{Ga}(1, 0.001)$ and $b_j \sim \text{Ga}(1, 0.001)$ for $j \in \{1, 2\}$, though in a real analysis we'd want to be more careful if we wanted to specify a non-informative prior. This problem is known to show high posterior dependence between $a_1$and b_1 as well as between a_2 and b_2.

You should do one of the two following variations on the problem, but you do not need to do both. Choose whichever software is of most interest for you to explore. In either case assess convergence using the graphical and quantitative measures we've discussed and seen in BDA.

(a) Fit the model using MCMC with NIMBLE using NIMBLE's default MCMC configuration and assess convergence. Now explore various samplers that you might use in place of the default samplers. Some things you might consider are slice sampling, blocking parameters and the approach discussed in Problem 3 of this problem set, which in NIMBLE we call a 'crossLevel' sampler. Details on various samplers can be found via *help(samplers)*.

(b) Alternatively, explore the use of MCMC using Hamiltonian Monte Carlo in Stan or using the MCMC tools in PyMC and assess convergence. You can also qualitatively compare the MCMC to that obtained from a default MCMC in NIMBLE based on the traceplots for the hyperparameters found in *nimble_litters_traces.pdf*.

3. *Context*: Consider the following strategy for building a joint sampler for one or more hyperparameters, ϕ, and a set of random effects, θ, that depend on ϕ. Suppose that one can derive the closed-form conditional for $\theta|\phi, y, \cdot$ where $\cdot$ stands for any other parameters in the model. In this context, one could integrate over θ analytically, but often one doesn't do this, either because one doesn't want to do the derivation or because integrating over θ would induce a complicated covariance structure for any dependencies of θ. For example, we've seen that if we have

$$\begin{aligned} y_{ij} &\sim N(\theta_j, \sigma^2) \\ \theta_j &\sim N(\mu, \tau^2) \end{aligned}$$

and one integrates over θ, one gets correlation in the covariance matrix for $y = \{\{y_{ij}\}\}$ (although in that case the dependence is a simple blocked structure).

Problem: Show that the following joint sampler for $\{\theta, \phi\}$ is equivalent to using Metropolis sampling for $\phi|y$.

(a) Sample $\phi^*|\phi$ from some symmetric distribution

(b) Sample $\theta^*|\phi^*, y, \cdot$ from its conditional distribution, $p(\theta|\phi, y, \cdot)$

(c) Accept or reject $\{\phi^*, \theta^*\}$ as a single joint Metropolis-Hastings proposal.

Basically you'll need to work out the r quantity in BDA equation 11.2 for the above algorithm and show it is the same as the r for a Metropolis sampling approach for ϕ in a model in which θ has been integrated out. This should be straightforward using the relationships of the posterior joint, conditional, and marginal distributions.

4. (Extra credit) Often practicioners will subsample (or "thin") their MCMC chain to reduce the autocorrelation in the MCMC samples. This is often done under the (mistaken) belief that it improves one's inference, but in fact, thinning *always* reduces the amount of information you have (i.e., it increases the Monte Carlo variance of averages of posterior samples). Prove that systematic subsampling from

a Markov chain increases the variance of sample means relative to using the entire chain. Assume that the chain is stationary (i.e., in the MCMC context that we have achieved burn-in and are sampling from the posterior) and consider what this implies about the distribution of the samples. One way to start the proof is by systematically dividing the samples from the chain into k subsamples of size n each and defining the subsample means, $\bar{\theta}_k$:

$$\begin{aligned}
\bar{\theta}_1 &= \frac{1}{n}(\theta_1 + \theta_{k+1} + \cdots + \theta_{nk-k+1}) \\
\bar{\theta}_2 &= \frac{1}{n}(\theta_2 + \theta_{k+2} + \cdots + \theta_{nk-k+2}) \\
&\cdots \\
\bar{\theta}_k &= \frac{1}{n}(\theta_k + \theta_{2k} + \cdots + \theta_{nk})
\end{aligned}$$

Then consider how the variance of the mean for the entire sample, $\bar{\theta}$, relates to the variance of a subsample mean. Note that here the indexing refers to the iterations from the chain, not the components of the parameter vector.