Stat 591 – Homework 06 **Due: Wednesday 12/11**

Your group should submit a write-up that includes solutions the problems stated below, along with any <u>*relevant*</u> *pictures/graphs or computer code/output.*

1. Problem 9.11 in [GDS].

2. Look at Example 9.3 in Section 9.1 of [GDS]; see, also, the last paragraph in Section 9.1.1. The problem is formulated as follows: $X_{j1}, \ldots, X_{jn} \overset{\text{iid}}{\sim} \mathsf{N}(\mu_j, \sigma^2)$, independent across $j = 1, \ldots, p$. Here $(\mu_1, \ldots, \mu_p, \sigma^2)$ are unknown. Take an exchangeable prior for $(\mu_1, \ldots, \mu_p)$:

$$\begin{aligned}(\mu_1, \ldots, \mu_p) \mid (m, v, \sigma^2) &\overset{\text{iid}}{\sim} \mathsf{N}(m, v) \\ (m, v, \sigma^2) &\sim \pi(m, v, \sigma^2) \propto 1/\sigma^2.\end{aligned}$$

   Modify your Gibbs sampler for the ANOVA problem in Homework 05 to this setting. In particular, explain how you would approximate $\mathsf{E}(\mu_j \mid X)$ using the Gibbs sampler output. There are several ways this can be done, some better than others. Look at Section 7.4.5 in [GDS] on Rao–Blackwellization, and Equation (9.1).

3. Consider the multiple testing problem in Scott & Berger (*JSPI*, 2006). In particular, note that the goal there is to evaluate $p_j := \mathsf{P}(\mu_j = 0 \mid x)$, $j = 1, \ldots, M$.

   (a) Suppose one can sample from the posterior distribution of $(\mu_1, \ldots, \mu_M, p, \sigma^2, V)$. How could you use that posterior sample to evaluate $p_1, \ldots, p_M$?

   (b) Why do they opt for an importance sampling strategy to evaluate $p_1, \ldots, p_M$ instead of a method like you described in part (a)? In other words, what is the advantage of importance sampling over MCMC in this case?

4. *Students' choice.* Complete one (or more) of the following computational exercises.

   (a) For the multiple testing problem, implement the Scott & Berger (*JSPI*, 2006) importance sampling procedure to compute $p_j := \mathsf{P}(\mu_j = 0 \mid x)$, $j = 1, \ldots, M$. Using all the same settings, reproduce the results in their Table 1 for the prior $\pi(p) = 11p^{10}$ and $n = 25, 100$.

   (b) Park & Casella (*JASA*, 2008) propose a Bayesian lasso method for variable selection in regression.[1] Reproduce the results displayed in their Figure 2 for the Bayesian lasso and least squares estimates only. For handling the Bayesian lasso parameter λ, you may use either the empirical Bayes (Section 3.1) or the full Bayes (Section 3.2) procedure.

   (c) Consider the simple many-normal-means problem discussed in Castillo & van der Vaart (*Annals*, 2012).[1] The theory presented there is relatively sophisticated (you don't have to read it) but, ultimately, they give some guidelines for

[1]Links to references available on course website; data available there too.

choosing good priors in this problem. A relatively simple prior, which is not covered by their theory, is the following:

$$(\theta_1, \ldots, \theta_n) \mid \omega \overset{\text{iid}}{\sim} \omega\delta_0 + (1-\omega)\mathsf{Unif}(-\infty, \infty)$$
$$\omega \sim \mathsf{Beta}(an, 1),$$

where "$\mathsf{Unif}(-\infty, \infty)$" denotes a flat prior, and $a > 0$ is a fixed constant.

i. Derive a Gibbs sampler to simulate from the posterior distribution of the mean vector $\theta = (\theta_1, \ldots, \theta_n)$.

ii. Consider using the coordinate-wise posterior mean $\hat{\theta}$ as an estimate of θ. Perform the simulation described in Section 3.4 of the paper and compute the mean square errors, $\mathsf{E}_\theta \|\hat{\theta} - \theta\|^2$, for the Bayes estimate above. Compare your results with those given in Table 1 of the paper.