Homework 4

This is the final homework and will be due on the final day of class. I shall keep adding some more problems in this on the later topics.

Problem 1: Suppose (u, v, w) have a trivariate mean zero normal distribution with $\text{var}(u) = \text{var}(v) = \text{w} = 1$ and $\text{cov}(u, v) = 0.99, \text{cov}(u, w) = 0.1$ and $\text{cov}(v, w) = 0$. The purpose of this problem is to study the effect of blocking in Gibbs samplers. Compare two sampling schemes: (a) sample iteratively from $[u \mid v, w], [v \mid u, w], [w \mid u, v]$ and (b) $[v \mid w], [u \mid v, w], [w \mid u, v]$. Justify why (b) is a valid Gibbs sampler. Summarize your findings. You should report effective sample sizes in addition to showing trace plots.

Problem 2: Consider the logistic regression model

$$\text{pr}(y_i = 1 \mid \beta) = F(x_i'\beta), \quad i = 1, \ldots, n,$$

with $y_i \in \{0, 1\}$ the binary response and $x_i \in \mathbb{R}^d$ the covariates for observation i, and $F(t) = e^t/(1 + e^t)$ the cumulative distribution function of a logistic distribution. The logistic density has heavier tails than Gaussian. O'Brien and Dunson (2004) observed that the logistic density can be accurately approximated by a $t_\nu(0, \sigma^2)$ density, with $\nu = 7.3$ and $\sigma^2 = \pi^2(\nu - 2)/(3\nu)$.

(i) Overlay the logistic density with a $t_\nu(0, \sigma^2)$ density with the above choices of ν and σ^2. How do you think they obtained the expression for (ν, σ^2)? [Hint: for a given ν, what would be a natural way to obtain a candidate for σ^2?]

(ii) Motivated by the above, suppose we want to use the cdf of the $t_\nu(0, \sigma^2)$ density (let's call it F^*) as a surrogate for the logistic cdf, and fit the model

$$\text{pr}(y_i = 1 \mid \beta) = F^*(x_i'\beta), \quad i = 1, \ldots, n,$$

with a $N(0, \xi^{-1}I_d)$ prior on β. Describe the steps of an appropriate data-augmentation Gibbs sampler to sample from the posterior distribution of β.

Problem 3: Consider the linear regression model with t-error and a ridge prior on β,

$$\begin{aligned} y_i \mid \beta, \sigma^2, \xi &\sim t_\nu(x_i'\beta, \sigma^2), \quad i = 1, \ldots, n, \\ \beta \mid \xi &\sim N(0, \xi^{-1}\sigma^2 I_d), \\ \sigma^2 &\sim IG(a/2, b/2) \\ \xi &\sim \pi_\xi, \end{aligned}$$

where π_ξ is as in the pdf file on ridge regression posted in ecampus.
(i) Describe the steps of an MCMC algorithm to sample from the joint posterior of $(\beta, \sigma^2, \xi) \mid y$.
(ii) Run this Gibbs sampler on data that is simulated as in the second example in the ridge regression file ($n = 100, d = 7$ and the first, third, fourth columns of X correlated), with the last step replaced by `y = mustar + sigstar*rt(n,df=4)`. You may assume the degrees of freedom to be known when you fit the model.

Run the MCMC for 20,000 iterations after discarding 5000 iterations as burn-iin. Provide trace plots and density plots as in the example file, and also report effective sample sizes for seven coordinates of β, σ^2, and ξ. **Your code should not include any inner loop inside the MCMC loop; in other words, all operations should be vectorized/matricized.**