BIOS 731 Advanced Statistical Computing
Fall 2014
Homework 3 (Bayesian Computation)
Due 10/31/2014 midnight

Instruction: Please submit both write-ups and programs. The programs need to be written in a high-level language (no compilation required), and `R` is highly recommended. The codes for all problems need to be saved in a single file named NAME hw3.EXT. Replace NAME by your name, and EXT by proper extension name, e.g., R, sas, etc. Provide adequate comments in the codes to clearly mark the section for different questions. The codes should generate all results and figures in the homework. Please make sure the codes are "self-contained", e.g., does not depend on platform, can be run at any other machine in any subdirectory, and does not require user input.

Question 1: Denote by $N_+(\mu, \mu^-, \sigma^2)$ the truncated normal distribution with left truncation point μ^-, i.e. the distribution with density

$$f(x \mid \mu, \mu^-, \sigma^2) = \frac{\exp(-(x-\mu)^2/2\sigma^2)}{\sqrt{2\pi}\sigma[1-\Phi((\mu^- - \mu)/\sigma)]} I[x \geq \mu^-]$$

1. Using the classical CDF inversion technique, design and implement an algorithm to simulate the truncated normal distribution

2. Let
$$g(x \mid \alpha, \mu^-) = \alpha \exp(-\alpha(x-\mu^-)) I[x \geq \mu^-].$$
   (a) Show that there is a constant $M(\alpha, \mu^-)$, such that
   $$f(x \mid \mu, \mu^-, \sigma^2) \leq M(\alpha, \mu^-) g(x \mid \alpha, \mu^-).$$
   (b) Using the accept-reject method, design and implement an algorithm to simulate the truncated normal distribution.
   (c) Derive the closed form of the acceptance probability in your designed algorithm and provide the numerical justification of your results.
   (d) Find the best choice of α by maximizing the acceptance probability. Verify your results by numerical experiments.

3. Perform a simulation study to compare the two algorithms that you developed in Part 1 and Part 2. For the comparison, we mainly focus on the computational time and the accuracy of the density estimations.

Question 2: Consider a finite Normal mixture model with K components, where K is fixed and pre-specified. Suppose the data are $y_1, \ldots, y_n$ then for $i = 1, 2, \ldots, n$,

$$y_i \mid \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\sigma}^{-2} \sim \sum_{k=1}^{K} p_k \text{Normal}(\mu_k, \sigma_k^2) \tag{1}$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)^T$, $\boldsymbol{\sigma}^{-2} = (\sigma_1^{-2}, \ldots, \sigma_K^{-2})^T$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_K)^T$. For priors and hyperpriors, we assume that

$$\begin{aligned} \mu_k &\overset{\text{i.i.d.}}{\sim} \text{Uniform}[y_{\min}, y_{\max}] \\ \sigma_k^{-2} \mid \beta &\overset{\text{i.i.d.}}{\sim} \text{Gamma}(\alpha, \beta) \\ \beta &\sim \text{Gamma}(a, b) \\ \boldsymbol{\lambda} &\sim \text{Dirichlet}(\theta_1, \ldots, \theta_K) \end{aligned}$$

where $y_{\min} = \min y_1, \ldots, y_n$ and $y_{\max} = \max y_1, \ldots, y_n$; and α, a, b and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)^T$ are fixed and pre-specified hyperparameters

1. Suppose $K = 2$. Using the Laplace method, design and implement an algorithm to estimate the marginal posterior distribution of μ_1, μ_2, σ_1^{-2}, σ_2^{-2} and λ_1.

2. Suppose $K \geq 2$. Without introducing any auxiliary variables, design and implement an MCMC algorithm to simulate the joint posterior distribution of $\boldsymbol{\mu}$, $\boldsymbol{\sigma}^{-2}$ and $\boldsymbol{\lambda}$.

3. Suppose $K \geq 2$. For each y_i, we introduce a latent indicator $z_i \in \{1, \ldots, K\}$; and assume that

$$\begin{aligned} [y_i \mid z_i = k, \mu_k, \sigma_k^2\ ] &\overset{\text{i.i.d.}}{\sim} \text{Normal}(\mu_k, \sigma_k^2) \\ \Pr(z_i = k) &= \lambda_k \end{aligned} \tag{2}$$

   (a) Show that (1) and (2) are equivalent.

   (b) Design and implement an MCMC algorithm to simulate the joint posterior distribution of $\boldsymbol{\mu}$, $\boldsymbol{\sigma}^{-2}$ and $\boldsymbol{\lambda}$.

4. Perform a simulation study to compare the three algorithms that you developed in Part 1, Part 2 and Part 3, when $K = 2$; and compare the two MCMC algorithms when $K = 10$. For the comparison, we mainly focus on the computational time and the accuracy of parameter estimations.

Question 3: Consider a linear regression model.

$$\mathbf{y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2 \sim \text{Normal}(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$$

where $\mathbf{y} = (y_1, \ldots, y_n)^T$ is an $n \times 1$ vector and $\mathbf{X} = (x_{ij})$ is an $n \times p$ matrix. The parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ and σ^2 are of our primarary interests. We consider the following prior specifications

$$\begin{aligned} [\beta_j \mid \sigma^2, z_j = k\ ] &\overset{\text{i.i.d.}}{\sim} \text{Normal}(0, \sigma^2\tau_k^2) \quad \text{for } k = 0, 1 \\ \Pr(z_j = 1) &= \pi \\ \sigma^{-2} &\sim \text{Gamma}(\alpha_1, \alpha_2) \end{aligned}$$

where τ_0^2, τ_1^2, α_1, α_2 and π are pre-determined

1. Design and implement a Gibbs sampler to simulate marginal posterior distribution of $\boldsymbol{\beta}$ and σ^2.

2. Perform simulation studies to evaluate the performance of the proposed algorithm. In particular, simulate data with the following settings: $\mathbf{X}$ are generated from the multivariate normal distribution with zero mean and compound symmetric covariance variance $0.75\mathbf{I}_p + 0.25\mathbf{1}_p\mathbf{1}_p^T$. The true parameters are set as follows

$$\boldsymbol{\beta} = (0.6, 1.2, 1.8, 2.4, 3.0, \underbrace{0, 0, \ldots, 0}_{p-5})^T, \qquad \sigma^2 = 1$$

   The suggested hyper parameter specifications are

$$\tau_0^2 = \frac{1}{10n}, \quad \tau_1^2 = \log(n), \qquad \pi = 0.5$$

   Please consider three different cases $(n, p) = (100, 100), (100, 200)$ and $(200, 500)$.