SDS 383C: Statistical Modeling I, Fall 2021
Homework 5, Due Nov 25, 12:00 Noon

Instructor: Abhra Sarkar (abhra.sarkar@utexas.edu)
Teaching Assistant: Angela Ting (angelating@utexas.edu)
Department of Statistics and Data Sciences
The University of Texas at Austin
2317 Speedway D9800, Austin, TX 78712-1823, USA

All homework must be submitted typed-in as a single pdf file. Name the file "firstname-lastname-SDS383C-HW-5.pdf" Submit this file without compression such as zip or rar. Figures accompanying the solutions must be presented close to the actual solution. Computer codes will be rarely evaluated but must still be submitted separately from the main file. Codes must be commented properly and should run easily on other machines. Precise, concise, clear, innovative solutions may be rewarded with bonus points. Explain your answer with logic reasoning and/or mathematical proofs. Organize your solutions in the same order as they were presented. If you can solve a problem using multiple techniques, present only your best solution.

1. (20 points) Let the true posterior of a parameter θ be $p(\theta \mid data) \equiv t_3(\theta)$.
   (a) Using importance resampling, with a normal distribution as the reference density, draw $n = 100$ samples from the posterior. Provide a histogram of the sampled values, superimposed over the actual posterior. Repeat it for $n = 10,000$.
   (b) Using importance sampling with $n = 100$ samples again (but not using the samples drawn in part(a)), estimate $\mathbb{E}(\theta \mid data)$ and $\mathrm{var}(\theta \mid data)$. Compare with the corresponding true values. Repeat it for $n = 10,000$.

2. (10 points) Using Gibbs algorithm, draw 1000 appropriately burned-in and thinned samples from a bivariate Normal distribution with means $(\mu_1, \mu_2) = (0, 2)$, variances $(\sigma_1^2, \sigma_2^2) = (1, 1)$ and correlation $\rho = 0.75$. Provide a scatter plot of the sampled values, superimposed over a contour of the actual density. Also, provide trace and autocorrelation plots of the sampled values.

3. (30 points) For an AR(1) process with mean μ, correlation $|\rho| < 1$ and error variance σ^2

$$y_t = \mu + \rho(y_{t-1} - \mu) + \epsilon_t, \qquad \epsilon_t \text{ iid with } \mathbb{E}(\epsilon_t) = 0 \text{ and } \mathrm{var}(\epsilon_t) = \sigma^2,$$

show that (a) $\mathrm{var}(y_t) = \frac{\sigma^2}{(1-\rho^2)}$,

(b) $\mathrm{corr}(y_t, y_{t-k}) = \rho^k$,

(c) $\mathbb{E}(\overline{y}_n) = \mu$,

(d) $\mathrm{var}(\overline{y}_n) = \frac{\sigma^2}{n^2(1-\rho^2)} \left\{ n + 2\sum_{j=1}^{n-1}(n-j)\rho^j \right\}$,

(e) $n\mathrm{var}(\overline{y}_n) \to \frac{\sigma^2}{(1-\rho)^2}$ as $n \to \infty$,

(f) $\mathbb{E}(s_n^2) \to \mathrm{var}(y_t)$ as $n \to \infty$.

[Hints: work with transformed variables $z_t = y_t - \mu$ and assume that the time index t can be extended to include $t = 0, -1, \ldots, -\infty$.]

4. (20 points) Simulate 500 points from an AR(1) process with mean $\mu = 0$, correlation $\rho = 0.75$ and Normally distributed errors with variance $\sigma^2 = 1$. Treating these sampled values as your data points $y_1, \ldots, y_{500}$ and assuming $\mu = 0$ to be known, fit a Bayesian model following the guidelines below.

   (a) Find out conjugate priors for ρ and σ^2.

   (b) List the full conditionals under your conjugate priors.

   (c) Using a Gibbs sampler based on these full conditionals with suitably chosen prior hyper-parameters, draw $2,000$ burned-in and thinned samples from the posterior.

   (d) Using these sampled values, construct 90% credible intervals for ρ and σ^2.

   [Hints: ignore the likelihood contribution of the first data point if that helps.]

5. (20 points) The 'galaxies' dataset from package 'MASS' in R gives the velocities in km-s/sec of 82 galaxies (export this dataset from R if you are using a different programming language). Divide all these values by 1000. Use these scaled values as your data points. Using Gibbs sampling, fit a Bayesian location-scale mixture of normals

$$f(y) = \textstyle\sum_{k=1}^{K} \pi_k \text{Normal}(y \mid \mu_k, \sigma_k^2)$$

   with $K = 3, 4, 5, 6, 7, 8$ components. Use priors

$$\boldsymbol{\pi} \sim \text{Dir}(1/K, \ldots, 1/K), \quad \mu_k \sim \text{Normal}(\mu_0, \sigma_0^2), \quad \sigma_k^2 \sim \text{Inv-Ga}(a_0, b_0)$$

   with suitable choices of the prior hyper-parameters. Provide a brief general description of your algorithm, explicitly listing the full conditionals. Summarize your results by showing the posterior mean densities and their point-wise 90% credible intervals superimposed over a histogram of the data points in a single figure with 3×2 panels.

6. (20 points) The 'faithful' dataset from package 'datasets' in R gives eruption and waiting times of the old faithful geyser in Yellowstone national park (export this dataset from R if you are using a different programming language).

   Using Gibbs sampling, fit a Bayesian location-scale mixture of bivariate normals

$$f(\mathbf{y}) = \textstyle\sum_{k=1}^{K} \pi_k \ \text{MVN}_2(\mathbf{y} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

   to this dataset with $K = 2, 3, 4, 5$ components. Use the priors

$$\boldsymbol{\pi} \sim \text{Dir}(1/K, \ldots, 1/K), \quad \boldsymbol{\mu}_k \sim \text{MVN}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0), \quad \boldsymbol{\Sigma}_k \sim \text{IW}\left(\nu_0, \boldsymbol{\Psi}_0\right)$$

   with suitable choices of the prior hyper-parameters. Provide a brief general description of your algorithm, explicitly listing the full conditionals. Summarize your results by showing contours of the fitted posterior mean densities superimposed over a scatterplot of the data points in a single figure with 2×2 panels.