Statistics 5444: Homework 3

For each homework assignment, turn in at the beginning of class on the indicated due date. Late assignments will only be accepted with special permission. Write each problem up *very* neatly ($\LaTeX$ is preferred). Show all of your work.

Problem 1

In this problem, you will construct a sampler for fitting a line to data, which has Cauchy innovations.

Part 1

Simulate 1,000 points (x,y), where $(x, y) \sim \text{Cauchy}(0, 1)$, with covariance structure

$$\begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}.$$

Plot the realizations of your simulation.

Part 2

Recall that under the model: $y_i = \beta_0 + \beta_1 x_i + \epsilon$, where $\epsilon \sim N(0, \sigma^2)$, we can derive the posterior estimate

$$\beta \sim N((X^T X)^{-1} X^T Y, \sigma^2 (X^T X)^{-1}). \tag{1}$$

Fit a standard regression line of the form $y = \beta_0 + \beta_1 x$ to the data. Plot the residuals and make a QQ plot to illustrate how poorly the "Least Squares" fit performs.

Part 3

Under the gamma-normal (scale) mixture model, we have

$$\beta \sim N((X^T X)^{-1} X^T Y, \frac{\sigma^2}{\gamma} (X^T X)^{-1}),$$

where $\gamma \sim \text{Gamma}(a, b)$. Find a and b so that β has a Cauchy distribution with shift $(X^T X)^{-1} X^T Y$ and scale $\sigma^2 (X^T X)^{-1}$.

Problem 3

In problem 2 we obtained some insight on how to sample from a Cauchy regression model. We will further the insight here. Consider the model:

$$\begin{aligned} y_i &\sim Normal(x_i^T\beta, \sigma^2/\gamma_i), \\ \gamma_i &\sim Gamma(a, b), \end{aligned}$$

for $i = 1, \ldots, N$.

Part 1a

Write out the full conditional distributions for β and $\phi = 1/\sigma^2$, under the reference priors. (Note: for a given value of γ, the full conditional distribution for β should be obvious.)

Part 1b

Write out the full conditional sampling distribution for γ_i, $i = 1, \ldots, N$. Notice that for each sample draw, you used a random γ, so there is a posterior distribution on γ_i for each sample draw.

Part 2

Write out a Gibbs sampling procedure for sampling from $(\beta, \phi = 1/\sigma^2, \gamma_i)$. You do not need to implement this, just write out the pseudo code.

Problem 4

Recall that the *trace* of a matrix A is defined to be the sum of the diagonal elements of A, or equivalently it is the sum of its eigenvalues. Let $tr(A)$ denote the trace of the matrix A.;

Part 1

Show that

$$tr(A + B) = tr(A) + tr(B).$$

Part 2

Show that

$$tr(AB) = tr(BA).$$

Problem 4

Let us consider the example in class where measurements of rats weights were measured through time. Letting x_{ij} denote the weight of rate i in week j. For this we assumed the model

$$x_{ij} \sim N(\alpha_i + \beta_i j, \sigma^2 = 1/\phi),$$

which simply specifies a regression model for each individual rat. We further assumed that each rats regression coefficients were modeled through

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} \sim N(\begin{pmatrix} \alpha_0 \\ \beta_0 \end{pmatrix}, \Sigma).$$

You might interpret (α_0, β_0) as the underlying average population regression coefficients for rats weights. The model also specifies that individuals regression coefficients may not be independent, hence the arbitrary covariance structure Σ. This model is a referred to as a *random effects* model, where the regression coefficients for individual rats are the random effects. Suppose we want to perform a Bayesian analysis, and for convenience we choose to do a conjugate analysis. The conjugate priors are

$$\begin{aligned} \phi &\sim \text{Gamma}(a, b) \\ (\alpha_0, \beta_0)^T &\sim N(\eta, \Psi) \\ \Sigma^{-1} &\sim \text{Wishart}((\rho R)^{-1}, \rho). \end{aligned}$$

Hint: recall $p(\Sigma^{-1}) \propto |\Sigma^{-1}|^{(\rho-2-1)/2} e^{-\frac{1}{2}tr(\rho R \Sigma^{-1})}$.
Derive the full conditional distributions for: $\phi, (\alpha_0, \beta_0)^T$, and Σ^{-1}.

Problem 5

Let $X = (x_1, \ldots x_n)$ and let $x_i \sim N(\mu = 200, \phi = \frac{1}{2})$, where $\phi = 1/\sigma^2$

Part 1

Under reference priors, write down the full conditional distribution for μ and ϕ. You don't need to derive these again, just state what they are.

Part 2

Implement a Metropolis-Hastings sampler for sampling $(\mu, \phi | X)$ in a "block". Discuss your burn-in time, and the proposal you used for this problem.

Part 2

Implement a Gibbs sampler for sampling from the distribution for $(\mu, \phi | X)$, where X is a 100 simulated data points from the above model. Initialize the sampler at $\mu_0 = 0$ and $\phi = 5$. Show the trace plots for both μ and ϕ. Report the burn-in time and illustrate histograms for both of the marginal posteriors (after burn-in).

Problem 6

Let $\mathbf{X} = \{X_1, \ldots X_n\}$, where $X_i = (x_1, \ldots x_k)^T$, and $X_i \sim N(\mu, \Sigma)$. Under $p(\mu) \propto 1$, we found previously that $p(\mu | \mathbf{X}, \Sigma)$ has a multivariate normal distribution with mean $\sum_i X_i / N$ and variance matrix Σ / N.

Under the prior $p(\Sigma) \propto |\Sigma|^{-(k+1)/2}$, find $p(\mu | \mathbf{X})$.

Problem 7

Let $x = (x_1, \ldots, x_n)$, were $x_i \sim \text{Bin}(N, p)$. Find the Jeffreys prior for p.