1. (20 points) Consider the linear model

\[ y \mid \beta, \sigma^2 \sim N(X\beta, \sigma^2 I) \]
\[ p(\beta, \sigma^2) \propto 1/\sigma^2, \]

where \( y \) is an \( n \times 1 \) vector and \( X \) is a known \( n \times p \) matrix of predictors with rank \( p \).

Let \( s^2 = (y - X\hat{\beta})'(y - X\hat{\beta}) \), where \( \hat{\beta} = (X'X)^{-1}X'y \). Show that

\[ \frac{(\beta - \hat{\beta})'X'X(\beta - \hat{\beta})}{ps^2} \mid y \sim F_{p,n-p}, \]

where \( F_{p,n-p} \) is the \( F \) distribution with \((p, n-p)\) degrees of freedom.
2. (20 points) Let \( y \) be an \( n \times 1 \) vector of outcomes. Consider the following hierarchical model:

\[
\begin{align*}
y | \mu, \Sigma & \sim N(\mu, \Sigma) \\
\mu | \Sigma & \sim N(\theta, \alpha \Sigma) \\
\Sigma & \sim IW(\nu, S),
\end{align*}
\]

where \( \mu \) is an unknown \( n \times 1 \) vector and \( \Sigma \) is an unknown \( n \times n \) variance-covariance matrix. Assume that \( \theta \) (an \( n \times 1 \) vector), \( \alpha \) (a scalar), \( \nu \) (a scalar), and \( S \) (an \( n \times n \) symmetric and positive definite matrix) are known.

(a) Find the joint posterior distribution for \( \mu \) and \( \Sigma \), i.e. \( p(\mu, \Sigma | y) \).

(b) Find the marginal posterior for \( \Sigma \), i.e., \( p(\Sigma | y) \).

(c) Find \( p(\mu | \Sigma, y) \).

(d) Describe how you will draw samples from the marginal posterior distribution of \( \mu \), i.e. from \( p(\mu | y) \), using Normal and Inverse Wishart generators only.
3. (20 points) Consider the Bayesian Seemingly Unrelated Regression (SUR) model, where $y_i$ is an $n \times 1$ vector of outcomes from the $i$-th subject such that:

$$y_i = X_i' \beta_i + \epsilon_i \quad \epsilon_i \overset{iid}{\sim} N(0, \Sigma) \quad \text{for} \quad i = 1, 2, \ldots, N,$$

where $X_i'$ is a known $n \times p$ matrix of predictors, $\beta_i$ is the $p \times 1$ slope vector for the $i$-th subject and $\Sigma$ is an unknown $n \times n$ symmetric positive definite matrix. Consider the priors: $\beta_i \overset{iid}{\sim} N(\mu_\beta, V_\beta)$ and $\Sigma \sim IW(\nu, S)$, where $\nu \geq n + 2$. Find the full conditional distribution of the $\beta_i$’s and $\Sigma$ and describe a Gibbs sampler to estimate this model.
4. (20 points) Let \( y_i \) be an \( m \times 1 \) vector of outcomes such that
\[
y_i \sim N(0, a_{ii} \Lambda) \quad \text{and} \quad \text{cov}(y_i, y_j) = a_{ij} \Lambda \quad \text{for} \quad i, j = 1, 2, \ldots, n,
\]
where \( \Lambda \) is an unknown \( m \times m \) symmetric positive definite matrix and \( a_{ij} \)'s are unknown scalars such that the \( n \times n \) matrix with \( a_{ij} \) as its \((i, j)\)-th element, say \( \mathbf{A} = \{a_{ij}\} \), is an \( n \times n \) symmetric positive definite matrix. Let \( \mathbf{y} = (y_1', \ldots, y_n')' \) be the \( mn \times 1 \) vector formed by stacking up the \( y_i \)'s in a single column.

(a) Show that the variance-covariance matrix of \( \mathbf{y} \), conditional upon the parameters, is \( \text{var} \{\mathbf{y}\} = \mathbf{A} \otimes \Lambda \), where \( \otimes \) is the Kronecker product for matrices. You can find the basic properties of \( \otimes \) that you will need for this problem in Wikipedia (http://en.wikipedia.org/wiki/Kronecker_product).

(b) Show that there exists an \( m \times m \) matrix \( \mathbf{S}_1 \) and an \( n \times n \) matrix \( \mathbf{S}_2 \) such that
\[
\mathbf{y}'(\mathbf{A} \otimes \Lambda)^{-1}\mathbf{y} = \text{tr}(\mathbf{S}_1 \Lambda^{-1}) = \text{tr}(\mathbf{S}_2 \mathbf{A}^{-1}),
\]
where \( \text{tr}(\cdot) \) denotes the trace function for a matrix.

(c) Let \( \mathbf{A} \sim IW(\nu_A, \mathbf{S}_A) \) and \( \Lambda \sim IW(\nu_\Lambda, \mathbf{S}_\Lambda) \), where the hyperparameters \( \nu_A, \nu_\Lambda, \mathbf{S}_A \) and \( \mathbf{S}_\Lambda \) are known. Find \( p(\mathbf{A} | \mathbf{y}, \Lambda) \) and \( p(\Lambda | \mathbf{A}, \mathbf{y}) \).
5. (20 points) Consider the following multivariate random effects model:

\[ y_i = X'_i \beta_i + w_i + \epsilon_i, \quad i = 1, 2, \ldots, n, \]

where \( y_i \) is an \( m \times 1 \) vector of outcomes, \( X'_i \) is a known \( m \times p \) matrix of predictors, \( \beta_i \) is a \( p \times 1 \) vector of slopes specific to subject \( i \), \( w_i \) is an \( m \times 1 \) unknown random effect and \( \epsilon_i \) iid \( \sim N(0, \Psi) \). Assume that the \( w_i \)'s are normally distributed with mean 0 and \( \text{cov}(w_i, w_j) = a_{ij} \Lambda \), where \( a_{ij} \)'s are unknown constants such that \( A = \{a_{ij}\} \) is a symmetric positive definite matrix and \( \Lambda \) is an unknown symmetric positive definite matrix as well. Suppose \( \beta_i \) iid \( \sim N(\mu_\beta, V_\beta) \), \( \Lambda \sim IW(\nu_\Lambda, S_\Lambda) \), \( A \sim IW(\nu_A, S_A) \) and \( \Psi \sim IW(\nu_\Psi, S_\Psi) \). Assume that all the hyperparameters are unknown.

Find the full conditional distributions for the \( \beta_i \)'s, the \( w_i \)'s, \( \Lambda \), \( A \) and \( \Psi \) and describe a Gibbs sampler to draw posterior samples for these parameters.