STA 601: Linear Mixed Effects Models

STA 601 Fall 2021

Merlise Clyde

Nov 4, 2021

Linear Mixed Effects Models

$$y_{ij} = \beta^T x_{ij} + \gamma^T z_{ij} + \epsilon_{ij}, \qquad \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$$
$$\gamma_j \overset{iid}{\sim} N_p(0, \Sigma)$$

Linear Mixed Effects Models

$$\begin{aligned} y_{ij} &= \beta^T x_{ij} + \gamma^T z_{ij} + \epsilon_{ij}, \qquad \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2) \\ \gamma_j &\overset{iid}{\sim} N_p(0, \Sigma) \end{aligned}$$

- Fixed effects contribution $\beta^T x_{ij}$, x_{ij} is a $d \times 1$ vector with β is constant across groups j, $j = 1, \ldots, J$

Linear Mixed Effects Models

$$
\begin{aligned}
&y_{ij} = \beta^T x_{ij} + \gamma^T z_{ij} + \epsilon_{ij}, \qquad \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2) \\
&\gamma_j \overset{iid}{\sim} N_p(0, \Sigma)
\end{aligned}
$$

- Fixed effects contribution $\beta^T x_{ij}$, x_{ij} is a $d \times 1$ vector with β is constant across groups j, $j = 1, \ldots, J$
- Random effects $\gamma_j^T z_{ij}$, z_{ij} is a $p \times 1$ vector with $\gamma_j \overset{iid}{\sim} N_p(0, \Sigma)$ for groups $j = 1, \ldots, J$

Linear Mixed Effects Models

$$y_{ij} = \beta^T x_{ij} + \gamma^T z_{ij} + \epsilon_{ij}, \qquad \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$$
$$\gamma_j \overset{iid}{\sim} N_p(0, \Sigma)$$

- Fixed effects contribution $\beta^T x_{ij}$, x_{ij} is a $d \times 1$ vector with β is constant across groups j, $j = 1, \dots, J$
- Random effects $\gamma_j^T z_{ij}$, z_{ij} is a $p \times 1$ vector with $\gamma_j \overset{iid}{\sim} N_p(0, \Sigma)$ for groups $j = 1, \dots, J$
- Designed to accomodate correlated data due to nested/hierarchical structure/repeated measurements

Linear Mixed Effects Models

$$y_{ij} = \beta^T x_{ij} + \gamma^T z_{ij} + \epsilon_{ij}, \qquad \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$$
$$\gamma_j \overset{iid}{\sim} N_p(0, \Sigma)$$

- Fixed effects contribution $\beta^T x_{ij}$, x_{ij} is a $d \times 1$ vector with β is constant across groups j, $j = 1, \dots, J$
- Random effects $\gamma_j^T z_{ij}$, z_{ij} is a $p \times 1$ vector with $\gamma_j \overset{iid}{\sim} N_p(0, \Sigma)$ for groups $j = 1, \dots, J$
- Designed to accomodate correlated data due to nested/hierarchical structure/repeated measurements
- students w/in schools; patients w/in hospitals; additional covariates

Linear Mixed Effects Models

$$
\begin{aligned}
y_{ij} &= \beta^T x_{ij} + \gamma^T z_{ij} + \epsilon_{ij}, \qquad \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2) \\
\gamma_j &\overset{iid}{\sim} N_p(0, \Sigma)
\end{aligned}
$$

- Fixed effects contribution $\beta^T x_{ij}$, x_{ij} is a $d \times 1$ vector with β is constant across groups j, $j = 1, \ldots, J$
- Random effects $\gamma_j^T z_{ij}$, z_{ij} is a $p \times 1$ vector with $\gamma_j \overset{iid}{\sim} N_p(0, \Sigma)$ for groups $j = 1, \ldots, J$
- Designed to accomodate correlated data due to nested/hierarchical structure/repeated measurements
- students w/in schools; patients w/in hospitals; additional covariates
- As before not inherently Bayesian! It's just a model/likelihood specification! Population parameters, $\theta = (\beta, \Sigma, \sigma^2)$

Likelihoods

- Complete Data Likelihood $(\{\gamma_i\}, \theta)$

$$L(\{\beta_i\}, \theta)) \propto \prod_j N(\gamma_j; 0, \Sigma) \prod_i N(y_{ij}; \beta^T x_{ij} + \gamma_j^T z_{ij}, \sigma^2)$$

Likelihoods

- Complete Data Likelihood $(\{\gamma_i\}, \theta)$

$$L(\{\beta_i\}, \theta)) \propto \prod_j N(\gamma_j; 0, \Sigma) \prod_i N(y_{ij}; \beta^T x_{ij} + \gamma_j^T z_{ij}, \sigma^2)$$

- Marginal likelihood based on just observed data $(\{y_{ij}\}, \{x_{ij}\}, \{z_{ij}\})$

Likelihoods

- Complete Data Likelihood $(\{\gamma_i\}, \theta)$

$$L(\{\beta_i\}, \theta)) \propto \prod_j N(\gamma_j; 0, \Sigma) \prod_i N(y_{ij}; \beta^T x_{ij} + \gamma_j^T z_{ij}, \sigma^2)$$

- Marginal likelihood based on just observed data $(\{y_{ij}\}, \{x_{ij}\}, \{z_{ij}\})$

$$L(\{\beta_i\}, \theta)) \propto \prod_j \int N(\gamma_j; 0, \Sigma) \prod_i N(y_{ij}; \beta^T x_{ij} + \gamma_j^T z_{ij}, \sigma^2)\, d\gamma_j$$

Likelihoods

- Complete Data Likelihood $(\{\gamma_i\}, \theta)$

$$L(\{\beta_i\}, \theta)) \propto \prod_j N(\gamma_j; 0, \Sigma) \prod_i N(y_{ij}; \beta^T x_{ij} + \gamma_j^T z_{ij}, \sigma^2)$$

- Marginal likelihood based on just observed data $(\{y_{ij}\}, \{x_{ij}\}, \{z_{ij}\})$

$$L(\{\beta_i\}, \theta)) \propto \prod_j \int N(\gamma_j; 0, \Sigma) \prod_i N(y_{ij}; \beta^T x_{ij} + \gamma_j^T z_{ij}, \sigma^2)\, d\gamma_j$$

- Option A: we can calculate this integral by brute force algebraically

Likelihoods

- Complete Data Likelihood $(\{\gamma_i\}, \theta)$

$$L(\{\beta_i\}, \theta)) \propto \prod_j N(\gamma_j; 0, \Sigma) \prod_i N(y_{ij}; \beta^T x_{ij} + \gamma_j^T z_{ij}, \sigma^2)$$

- Marginal likelihood based on just observed data $(\{y_{ij}\}, \{x_{ij}\}, \{z_{ij}\})$

$$L(\{\beta_i\}, \theta)) \propto \prod_j \int N(\gamma_j; 0, \Sigma) \prod_i N(y_{ij}; \beta^T x_{ij} + \gamma_j^T z_{ij}, \sigma^2) \, d\gamma_j$$

- Option A: we can calculate this integral by brute force algebraically
- Option B: (lazy option) We can calculate marginal exploiting properties of Gaussians as sums will be normal - just read off the first two moments!

Marginal Distribution

- Express observed data as vectors for each group j: (Y_j, X_j, Z_j) where Y_j is $n_j \times 1$, X_j is $n_j \times d$ and Z_j is $n_j \times p$;

Marginal Distribution

- Express observed data as vectors for each group j: (Y_j, X_j, Z_j) where Y_j is $n_j \times 1$, X_j is $n_j \times d$ and Z_j is $n_j \times p$;

- Group Specific Model (1):

$$
\begin{aligned}
&Y_j = X_j\beta + Z_j\gamma + \epsilon_j, \qquad \epsilon_j \sim N(0, \sigma^2 I_{n_j}) \\
&\gamma_j \overset{iid}{\sim} N(0, \Sigma)
\end{aligned}
$$

Marginal Distribution

- Express observed data as vectors for each group j: (Y_j, X_j, Z_j) where Y_j is $n_j \times 1$, X_j is $n_j \times d$ and Z_j is $n_j \times p$;

- Group Specific Model (1):

$$
\begin{aligned}
Y_j &= X_j\beta + Z_j\gamma + \epsilon_j, \qquad \epsilon_j \sim N(0, \sigma^2 I_{n_j}) \\
\gamma_j &\overset{iid}{\sim} N(0, \Sigma)
\end{aligned}
$$

- Population Mean $E[Y_j] = E[X_j\beta + Z_j\gamma_j + \epsilon_j] = X_j\beta$

Marginal Distribution

- Express observed data as vectors for each group j: (Y_j, X_j, Z_j) where Y_j is $n_j \times 1$, X_j is $n_j \times d$ and Z_j is $n_j \times p$;
- Group Specific Model (1):

$$
\begin{aligned}
&Y_j = X_j\beta + Z_j\gamma + \epsilon_j, \qquad \epsilon_j \sim N(0, \sigma^2 I_{n_j}) \\
&\gamma_j \overset{iid}{\sim} N(0, \Sigma)
\end{aligned}
$$

- Population Mean $E[Y_j] = E[X_j\beta + Z_j\gamma_j + \epsilon_j] = X_j\beta$
- Covariance $V[Y_j] = V[X_j\beta + Z_j\gamma_j + \epsilon_j] = Z_j\Sigma Z_j^T + \sigma^2 I_{n_j}$

Marginal Distribution

- Express observed data as vectors for each group j: (Y_j, X_j, Z_j) where Y_j is $n_j \times 1$, X_j is $n_j \times d$ and Z_j is $n_j \times p$;
- Group Specific Model (1):

$$
\begin{aligned}
Y_j &= X_j\beta + Z_j\gamma + \epsilon_j, \qquad \epsilon_j \sim N(0, \sigma^2 I_{n_j}) \\
\gamma_j &\overset{iid}{\sim} N(0, \Sigma)
\end{aligned}
$$

- Population Mean $E[Y_j] = E[X_j\beta + Z_j\gamma_j + \epsilon_j] = X_j\beta$
- Covariance $V[Y_j] = V[X_j\beta + Z_j\gamma_j + \epsilon_j] = Z_j\Sigma Z_j^T + \sigma^2 I_{n_j}$
- Group Specific Model (2)

$$
Y_j \mid \beta, \Sigma, \sigma^2 \overset{ind}{\sim} N(X_j\beta, Z_j\Sigma Z_j^T + \sigma^2 I_{n_j})
$$

Priors

- Model (1) leads to a simple Gibbs sampler if we use conditional (semi-) conjugate priors on $\theta = (\beta, \Sigma, \phi = 1/\sigma^2)$

$$
\begin{aligned}
\beta &\sim N(\mu_0, \Psi_0^{-1}) \\
\phi &\sim \mathsf{Gamma}(v_0/2, v_o \sigma_0^2/2)
\end{aligned}
$$

Priors

- Model (1) leads to a simple Gibbs sampler if we use conditional (semi-) conjugate priors on $\theta = (\beta, \Sigma, \phi = 1/\sigma^2)$

$$\beta \sim N(\mu_0, \Psi_0^{-1})$$
$$\phi \sim \mathsf{Gamma}(v_0/2, v_o\sigma_0^2/2)$$

- One complication is that the covariance matrix Σ must be **positive definite and symmetric**.

Priors

- Model (1) leads to a simple Gibbs sampler if we use conditional (semi-) conjugate priors on $\theta = (\beta, \Sigma, \phi = 1/\sigma^2)$

$$\beta \sim N(\mu_0, \Psi_0^{-1})$$
$$\phi \sim \text{Gamma}(v_0/2, v_o\sigma_0^2/2)$$

- One complication is that the covariance matrix Σ must be **positive definite and symmetric**.

- "Positive definite" means that for all $x \in \mathcal{R}^p$, $x^T\Sigma x > 0$.

Priors

- Model (1) leads to a simple Gibbs sampler if we use conditional (semi-) conjugate priors on $\theta = (\beta, \Sigma, \phi = 1/\sigma^2)$

$$\begin{aligned} \beta &\sim N(\mu_0, \Psi_0^{-1}) \\ \phi &\sim \text{Gamma}(v_0/2, v_o\sigma_0^2/2) \end{aligned}$$

- One complication is that the covariance matrix Σ must be **positive definite and symmetric**.

- "Positive definite" means that for all $x \in \mathcal{R}^p$, $x^T\Sigma x > 0$.

- Basically ensures that the diagonal elements of Σ (corresponding to the marginal variances) are positive.

Priors

- Model (1) leads to a simple Gibbs sampler if we use conditional (semi-) conjugate priors on $\theta = (\beta, \Sigma, \phi = 1/\sigma^2)$

$$\beta \sim N(\mu_0, \Psi_0^{-1})$$
$$\phi \sim \text{Gamma}(v_0/2, v_o\sigma_0^2/2)$$

- One complication is that the covariance matrix Σ must be **positive definite and symmetric**.

- "Positive definite" means that for all $x \in \mathcal{R}^p$, $x^T \Sigma x > 0$.

- Basically ensures that the diagonal elements of Σ (corresponding to the marginal variances) are positive.

- Also, ensures that the correlation coefficients for each pair of variables are between -1 and 1.

Priors

- Model (1) leads to a simple Gibbs sampler if we use conditional (semi-) conjugate priors on $\theta = (\beta, \Sigma, \phi = 1/\sigma^2)$

$$\beta \sim N(\mu_0, \Psi_0^{-1})$$
$$\phi \sim \text{Gamma}(v_0/2, v_o\sigma_0^2/2)$$

- One complication is that the covariance matrix Σ must be **positive definite and symmetric**.
- "Positive definite" means that for all $x \in \mathcal{R}^p$, $x^T\Sigma x > 0$.
- Basically ensures that the diagonal elements of Σ (corresponding to the marginal variances) are positive.
- Also, ensures that the correlation coefficients for each pair of variables are between -1 and 1.
- Our prior for Σ should thus assign probability one to set of positive definite matrices.

Inverse-Wishart distribution

- Analogous to the univariate case, the **inverse-Wishart distribution** is the corresponding conditionally conjugate prior for Σ (multivariate generalization of the inverse-gamma).

Inverse-Wishart distribution

- Analogous to the univariate case, the **inverse-Wishart distribution** is the corresponding conditionally conjugate prior for Σ (multivariate generalization of the inverse-gamma).

- Hoff covers the construction of Wishart and inverse-Wishart random variables in Chapter 7.

Inverse-Wishart distribution

- Analogous to the univariate case, the **inverse-Wishart distribution** is the corresponding conditionally conjugate prior for Σ (multivariate generalization of the inverse-gamma).

- Hoff covers the construction of Wishart and inverse-Wishart random variables in Chapter 7.

- A random variable $\Sigma \sim \text{IW}_p(\eta_0, \boldsymbol{S}_0^{-1})$, where Σ is positive definite and $p \times p$, has pdf

$$p(\Sigma) \propto |\Sigma|^{\frac{-(\eta_0+p+1)}{2}} \exp\left\{-\frac{1}{2}\text{tr}(\boldsymbol{S}_0\Sigma^{-1})\right\}$$

Inverse-Wishart distribution

- Analogous to the univariate case, the **inverse-Wishart distribution** is the corresponding conditionally conjugate prior for Σ (multivariate generalization of the inverse-gamma).

- Hoff covers the construction of Wishart and inverse-Wishart random variables in Chapter 7.

- A random variable $\Sigma \sim \text{IW}_p(\eta_0, \boldsymbol{S}_0^{-1})$, where Σ is positive definite and $p \times p$, has pdf

$$p(\Sigma) \propto |\Sigma|^{\frac{-(\eta_0+p+1)}{2}} \exp\left\{-\frac{1}{2}\text{tr}(\boldsymbol{S}_0\Sigma^{-1})\right\}$$

where

- $\eta_0 > p - 1$ is the "degrees of freedom", and
- $\boldsymbol{S}_0$ is a $p \times p$ positive definite matrix.

Mean

- For this distribution, $E[\Sigma] = \frac{1}{\eta_0 - p - 1}\boldsymbol{S}_0$, for $\eta_0 > p + 1$.

Mean

- For this distribution, $E[\Sigma] = \frac{1}{\eta_0 - p - 1}\boldsymbol{S}_0$, for $\eta_0 > p + 1$.
- If we are very confident in a prior guess Σ_0, for Σ, then we might set
  - η_0, the degrees of freedom to be very large, and
  - $\boldsymbol{S}_0 = (\eta_0 - p - 1)\Sigma_0$.

  In this case, $E[\Sigma] = \frac{1}{\eta_0 - p - 1}\boldsymbol{S}_0 = \frac{1}{\eta_0 - p - 1}(\eta_0 - p - 1)\Sigma_0 = \Sigma_0$, and Σ is tightly (depending on the value of η_0) centered around Σ_0.

Mean

- For this distribution, $E[\Sigma] = \frac{1}{\eta_0 - p - 1} \boldsymbol{S}_0$, for $\eta_0 > p + 1$.
- If we are very confident in a prior guess Σ_0, for Σ, then we might set
  - η_0, the degrees of freedom to be very large, and
  - $\boldsymbol{S}_0 = (\eta_0 - p - 1)\Sigma_0$.

  In this case, $E[\Sigma] = \frac{1}{\eta_0 - p - 1} \boldsymbol{S}_0 = \frac{1}{\eta_0 - p - 1}(\eta_0 - p - 1)\Sigma_0 = \Sigma_0$, and Σ is tightly (depending on the value of η_0) centered around Σ_0.
- If we are not at all confident but we still have a prior guess Σ_0, we might set
  - $\eta_0 = p + 2$, so that the $E[\Sigma] = \frac{1}{\eta_0 - p - 1} \boldsymbol{S}_0$ is finite.
  - $\boldsymbol{S}_0 = \Sigma_0$

Wishart distribution

- Just as we had with the gamma and inverse-gamma relationship in the univariate case, we can also work in terms of the **Wishart distribution** (multivariate generalization of the gamma) instead.

Wishart distribution

- Just as we had with the gamma and inverse-gamma relationship in the univariate case, we can also work in terms of the **Wishart distribution** (multivariate generalization of the gamma) instead.

- The **Wishart distribution** provides a conditionally-conjugate prior for the precision matrix Σ^{-1} in a multivariate normal model.

Wishart distribution

- Just as we had with the gamma and inverse-gamma relationship in the univariate case, we can also work in terms of the **Wishart distribution** (multivariate generalization of the gamma) instead.
- The **Wishart distribution** provides a conditionally-conjugate prior for the precision matrix Σ^{-1} in a multivariate normal model.
- Specifically, if $\Sigma \sim \mathrm{IW}_p(\eta_0, \boldsymbol{S}_0)$, then $\Phi = \Sigma^{-1} \sim \mathrm{W}_p(\eta_0, \boldsymbol{S}_0^{-1})$.

Wishart distribution

- Just as we had with the gamma and inverse-gamma relationship in the univariate case, we can also work in terms of the **Wishart distribution** (multivariate generalization of the gamma) instead.
- The **Wishart distribution** provides a conditionally-conjugate prior for the precision matrix Σ^{-1} in a multivariate normal model.
- Specifically, if $\Sigma \sim \mathrm{IW}_p(\eta_0, \boldsymbol{S}_0)$, then $\Phi = \Sigma^{-1} \sim \mathrm{W}_p(\eta_0, \boldsymbol{S}_0^{-1})$.
- A random variable $\Phi \sim \mathrm{W}_p(\eta_0, \boldsymbol{S}_0^{-1})$, where Φ has dimension $(p \times p)$, has pdf

$$f(\Phi) \ \propto \ |\Phi|^{\frac{\eta_0 - p - 1}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{S}_0 \Phi)\right\}.$$

Wishart distribution

- Just as we had with the gamma and inverse-gamma relationship in the univariate case, we can also work in terms of the **Wishart distribution** (multivariate generalization of the gamma) instead.
- The **Wishart distribution** provides a conditionally-conjugate prior for the precision matrix Σ^{-1} in a multivariate normal model.
- Specifically, if $\Sigma \sim \text{IW}_p(\eta_0, \boldsymbol{S}_0)$, then $\Phi = \Sigma^{-1} \sim \text{W}_p(\eta_0, \boldsymbol{S}_0^{-1})$.
- A random variable $\Phi \sim \text{W}_p(\eta_0, \boldsymbol{S}_0^{-1})$, where Φ has dimension $(p \times p)$, has pdf

$$f(\Phi) \ \propto \ |\Phi|^{\frac{\eta_0 - p - 1}{2}} \exp\left\{-\frac{1}{2}\text{tr}(\boldsymbol{S}_0 \Phi)\right\}.$$

- Here, $E[\Phi] = \eta_0 \boldsymbol{S}_0$.

Conditional posterior for Σ

$$
\begin{aligned}
Y_j \mid \beta, \gamma_j, \sigma^2 &\overset{ind}{\sim} N(X_j\beta + Z_j\gamma_j, \sigma^2 I_{n_j}) \\
\gamma_j \mid \Sigma &\overset{iid}{\sim} N(0, \Sigma) \\
\Sigma &\sim \mathrm{IW}_p(\eta_0, \boldsymbol{S}_0^{-1}) \\
\beta &\sim N(\mu_0, \Psi_0^{-1}) \\
\phi &\sim \mathsf{Gamma}(v_0/2, v_o\sigma_0^2/2)
\end{aligned}
$$

- The conditional posterior (full conditional) $\Sigma \mid \boldsymbol{\gamma}, \boldsymbol{Y}$, is then

$$
\begin{aligned}
\pi(\Sigma \mid \boldsymbol{\gamma}, \boldsymbol{Y}) &\propto \pi(\Sigma) \cdot \pi(\boldsymbol{\gamma} \mid \Sigma) \\
&\propto \underbrace{|\Sigma|^{\frac{-(\eta_0+p+1)}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{S}_0\Sigma^{-1})\right\}}_{\pi(\Sigma)} \cdot \underbrace{\prod_{j=1}^{J} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\boldsymbol{\gamma}_j^T \Sigma^{-1} \gamma_j\right]\right\}}_{\pi(\boldsymbol{\gamma}|\Sigma)}
\end{aligned}
$$

Conditional posterior for Σ

$$
\begin{aligned}
Y_j \mid \beta, \gamma_j, \sigma^2 &\overset{ind}{\sim} N(X_j\beta + Z_j\gamma_j, \sigma^2 I_{n_j}) \\
\gamma_j \mid \Sigma &\overset{iid}{\sim} N(0, \Sigma) \\
\Sigma &\sim \mathrm{IW}_p(\eta_0, \boldsymbol{S}_0^{-1}) \\
\beta &\sim N(\mu_0, \Psi_0^{-1}) \\
\phi &\sim \textsf{Gamma}(v_0/2, v_o\sigma_0^2/2)
\end{aligned}
$$

- The conditional posterior (full conditional) $\Sigma \mid \boldsymbol{\gamma}, \boldsymbol{Y}$, is then

$$
\begin{aligned}
\pi(\Sigma \mid \boldsymbol{\gamma}, \boldsymbol{Y}) &\propto \pi(\Sigma) \cdot \pi(\boldsymbol{\gamma} \mid \Sigma) \\
&\propto \underbrace{|\Sigma|^{\frac{-(\eta_0+p+1)}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{S}_0\Sigma^{-1})\right\}}_{\pi(\Sigma)} \cdot \underbrace{\prod_{j=1}^{J} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\boldsymbol{\gamma}_j^T\Sigma^{-1}\gamma_j\right]\right\}}_{\pi(\boldsymbol{\gamma}|\Sigma)}
\end{aligned}
$$

- $\Sigma \mid \{\gamma_j\}, \boldsymbol{Y} \sim \mathrm{IW}_p\left(\eta_0 + J, (\boldsymbol{S}_0 + \sum_{j=1}^{J}\gamma_j\gamma_j^T)^{-1}\right)$

Posterior Continued

$$
\begin{aligned}
\pi(\Sigma \mid \boldsymbol{\gamma}, \boldsymbol{Y}) &\propto \pi(\Sigma) \cdot \pi(\boldsymbol{\gamma} \mid \Sigma) \\
&\propto |\Sigma|^{\frac{-(\eta_0+p+1)}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{S}_0\Sigma^{-1})\right\} \cdot \prod_{j=1}^{J} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\boldsymbol{\gamma}_j^T\Sigma^{-1}\gamma_j\right]\right\} \\
&\propto |\Sigma|^{\frac{-(\eta_0+p+J+1)}{2}} \exp\left\{-\frac{1}{2}\left[\mathrm{tr}\left[\boldsymbol{S}_0\Sigma^{-1}\right] + \sum_{j=1}^{J}\gamma_j^T\Sigma^{-1}\gamma_j\right]\right\}, \\
&\propto |\Sigma|^{\frac{-(\eta_0+p+J+1)}{2}} \exp\left\{-\frac{1}{2}\left[\mathrm{tr}\left[\boldsymbol{S}_0\Sigma^{-1}\right] + \sum_{j=1}^{J}\mathrm{tr}\left[\gamma_j\gamma_j^T\Sigma^{-1}\right]\right]\right\}, \\
&\propto |\Sigma|^{\frac{-(\eta_0+p+J+1)}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{S}_0\Sigma^{-1} + \sum_{j=1}^{J}\gamma_j\gamma_j^T\Sigma^{-1}\right]\right\}, \\
&\propto |\Sigma|^{\frac{-(\eta_0+p+J+1)}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\left(\boldsymbol{S}_0 + \sum_{j=1}^{J}\gamma_j\gamma_j^T\right)\Sigma^{-1}\right]\right\},
\end{aligned}
$$

MCMC Sampling

- Model (1) leads to a simple Gibbs sampler if we use conditional (semi-) conjugate priors on $\theta = (\beta, \Sigma, \phi = 1/\sigma^2)$

MCMC Sampling

- Model (1) leads to a simple Gibbs sampler if we use conditional (semi-) conjugate priors on $\theta = (\beta, \Sigma, \phi = 1/\sigma^2)$
- Model (2) can be challenging to update the variaance components! no conjugacy and need to ensure that MH updates maintain the positive-definiteness of Σ (can reparameterize)

MCMC Sampling

- Model (1) leads to a simple Gibbs sampler if we use conditional (semi-) conjugate priors on $\theta = (\beta, \Sigma, \phi = 1/\sigma^2)$
- Model (2) can be challenging to update the variaance components! no conjugacy and need to ensure that MH updates maintain the positive-definiteness of Σ (can reparameterize)
- Is Gibbs always more efficient?

MCMC Sampling

- Model (1) leads to a simple Gibbs sampler if we use conditional (semi-) conjugate priors on $\theta = (\beta, \Sigma, \phi = 1/\sigma^2)$
- Model (2) can be challenging to update the variaance components! no conjugacy and need to ensure that MH updates maintain the positive-definiteness of Σ (can reparameterize)
- Is Gibbs always more efficient?
- No - because the Gibbs sampler can have high autocorrelation in updating the $\{\gamma_j\}$ from their full conditional and then updating θ from their full full conditionals given the $\{\gamma_j\}$

MCMC Sampling

- Model (1) leads to a simple Gibbs sampler if we use conditional (semi-) conjugate priors on $\theta = (\beta, \Sigma, \phi = 1/\sigma^2)$
- Model (2) can be challenging to update the variaance components! no conjugacy and need to ensure that MH updates maintain the positive-definiteness of Σ (can reparameterize)
- Is Gibbs always more efficient?
- No - because the Gibbs sampler can have high autocorrelation in updating the $\{\gamma_j\}$ from their full conditional and then updating θ from their full full conditionals given the $\{\gamma_j\}$
- slow mixing

MCMC Sampling

- Model (1) leads to a simple Gibbs sampler if we use conditional (semi-) conjugate priors on $\theta = (\beta, \Sigma, \phi = 1/\sigma^2)$

- Model (2) can be challenging to update the variaance components! no conjugacy and need to ensure that MH updates maintain the positive-definiteness of Σ (can reparameterize)

- Is Gibbs always more efficient?

- No - because the Gibbs sampler can have high autocorrelation in updating the $\{\gamma_j\}$ from their full conditional and then updating θ from their full full conditionals given the $\{\gamma_j\}$

- slow mixing

- update β using (2) instead of (1) (marginalization so is independent of γ_j's

Marginal update for β

$$Y_j \mid \beta, \Sigma, \sigma^2 \overset{ind}{\sim} N(X_j\beta, Z_j \Sigma Z_j^T + \sigma^2 I_{n_j})$$
$$\beta \sim N(\mu_0, \Psi_0^{-1})$$

Marginal update for β

$$Y_j \mid \beta, \Sigma, \sigma^2 \overset{ind}{\sim} N(X_j\beta, Z_j \Sigma Z_j^T + \sigma^2 I_{n_j})$$
$$\beta \sim N(\mu_0, \Psi_0^{-1})$$

- Let $\Phi_j = (Z_j \Sigma Z_j^T + \sigma^2 I_{n_j})^{-1}$ (precision in model 2)

Marginal update for β

$$Y_j \mid \beta, \Sigma, \sigma^2 \overset{ind}{\sim} N(X_j\beta, Z_j \Sigma Z_j^T + \sigma^2 I_{n_j})$$
$$\beta \sim N(\mu_0, \Psi_0^{-1})$$

- Let $\Phi_j = (Z_j \Sigma Z_j^T + \sigma^2 I_{n_j})^{-1}$ (precision in model 2)

$$\pi(\beta \mid \Sigma, \sigma^2, \boldsymbol{Y}) \propto |\Psi_0|^{1/2} \exp\left\{-\frac{1}{2}(\beta - \mu_0)^T \Psi_0 (\beta - \mu_0)\right\}\cdot$$
$$\prod_{j=1}^{J} |\Phi_j|^{1/2} \exp\left\{-\frac{1}{2}(Y_j - X_j\beta)^T \Phi_j (Y_j - X_j\beta)\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left((\beta - \mu_0)^T \Psi_0 (\beta - \mu_0) + \sum_j (Y_j - X_j\beta)^T \Phi_j (Y_j - X_j\beta)\right)\right\}$$

Marginal Posterior for β

$$\pi(\beta \mid \Sigma, \sigma^2, \boldsymbol{Y}) \propto \exp\left\{-\frac{1}{2}\left((\beta-\mu_0)^T\Psi_0(\beta-\mu_0)+\sum_j (Y_j - X_j\beta)^T\Phi_j(Y_j - X_j\beta)\right)\right\}$$

Marginal Posterior for β

$$\pi(\beta \mid \Sigma, \sigma^2, \boldsymbol{Y}) \\ \propto \exp\left\{-\frac{1}{2}\left((\beta-\mu_0)^T\Psi_0(\beta-\mu_0)+\sum_j (Y_j - X_j\beta)^T\Phi_j(Y_j - X_j\beta)\right)\right\}$$

- Expand, read off precision and fix up mean so that it is a function of MLE's

Marginal Posterior for β

$$\pi(\beta \mid \Sigma, \sigma^2, \boldsymbol{Y}) \\ \propto \exp\left\{-\frac{1}{2}\left((\beta - \mu_0)^T \Psi_0 (\beta - \mu_0) + \sum_j (Y_j - X_j\beta)^T \Phi_j (Y_j - X_j\beta)\right)\right\}$$

- Expand, read off precision and fix up mean so that it is a function of MLE's

- precision

$$\Psi_n = \Psi_0 + \sum_{j=1}^{J} X_j^T \Phi_j X_j$$

- mean

$$\mu_n = \left(\Psi_0 + \sum_{j=1}^{J} X_j^T \Phi_j X_j\right)^{-1} \left(\Psi_0 \mu_0 + \sum_{j=1}^{J} X_j^T \Phi_j X_j \hat{\beta}_j\right)$$

where \hat{\beta}_j = (X_j^T \Phi X_j)^{-1} X_j^T \Phi_j Y_j\right) is the generalized least squares estimate of β for group j

Full conditional for σ^2 or ϕ

$$
\begin{aligned}
Y_j \mid \beta, \gamma_j, \sigma^2 &\overset{ind}{\sim} N(X_j\beta + Z_j\gamma_j, \sigma^2 I_{n_j}) \\
\gamma_j \mid \Sigma &\overset{iid}{\sim} N(0, \Sigma) \\
\Sigma &\sim \mathrm{IW}_p(\eta_0, \boldsymbol{S}_0^{-1}) \\
\beta &\sim N(\mu_0, \Psi_0^{-1}) \\
\phi &\sim \textsf{Gamma}(v_0/2, v_o\sigma_0^2/2)
\end{aligned}
$$

$$
\pi(\phi \mid \beta, \{\gamma_j\}\{Y_j\}) \propto \textsf{Gamma}(\phi; v_0/2, v_o\sigma_0^2/2) \prod_j N(Y_j; X_j\beta + Z_j\gamma_j, \phi^{-1} I_{n_j}))
$$

$$
\phi \mid \{Y_j\}, \beta, \{\gamma_j\} \sim \textsf{Gamma}\left(\frac{v_0 + \sum_j n_j}{2}, \frac{v_o\sigma_0^2 + \sum_j \|Y_j - X_j\beta - Z_j\gamma_j\|^2}{2}\right)
$$

Full conditional for $\{\gamma_j\}$

$$
\begin{aligned}
Y_j \mid \beta, \gamma_j, \sigma^2 &\overset{ind}{\sim} N(X_j\beta + Z_j\gamma_j, \sigma^2 I_{n_j}) \\
\gamma_j \mid \Sigma &\overset{iid}{\sim} N(0, \Sigma) \\
\Sigma &\sim \mathrm{IW}_p(\eta_0, \boldsymbol{S}_0^{-1}) \\
\beta &\sim N(\mu_0, \Psi_0^{-1}) \\
\phi &\sim \mathsf{Gamma}(v_0/2, v_o\sigma_0^2/2)
\end{aligned}
$$

$$
\pi(\gamma \mid \beta, \phi, \Sigma) \propto \mathsf{N}(\gamma_j; 0, \Sigma) \prod_j N(Y_j; X_j\beta + Z_j\gamma_j, \phi^{-1} I_{n_j}))
$$