7.1 *Glossary of Notation*

- $\mathbf{Y}$: A $n \times p$ matrix containing all multivariate skew-normal regression outcomes such that y_{ij} is the j^{th} outcome observed for subject i, where $i = 1, ..., n$ and $j = 1, ...p$.
- $\mathbf{X}$: A $n \times m$ matrix containing all multivariate skew-normal regression covariates such that x_{ij} is the j^{th} covariate value for subject i.
- $\mathbf{B}$: A $m \times p$ matrix containing all multivariate skew-normal regression coefficients such that $\mathbf{B} = \left[\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_p\right]$, where β_{ij} is interpreted as the effect of covariate i on outcome j for $i = 1, ..., m$ and $j = 1, ..., p$.
- $\mathbf{E}$: A $n \times p$ matrix of error terms in the multivariate skew-normal regression model component. $\mathbf{E}$ is made up of row vectors $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, ..., \epsilon_{ip})$, where $\boldsymbol{\epsilon}_i \overset{iid}{\sim} N_p(0, \boldsymbol{\Sigma})$ for $i = 1, ..., n$.
- $\boldsymbol{\Sigma}$: A $p \times p$ covariance matrix that defines the correlation between the p multivariate normal outcomes.
- $\boldsymbol{\Omega}$: A $p \times p$ covariance scale matrix that defines the correlation between the p multivariate skew-normal outcomes.
- $\boldsymbol{\psi}$: A $p \times 1$ vector containing the skewness parameter for each outcome.
- $\boldsymbol{\alpha}$: A $p \times 1$ vector containing the skewness parameter for each outcome.
- $\mathbf{t}$: A $n \times 1$ vector of truncated normal random effects used in the stochastic representation of the multivariate skew-normal distribution. For $i = 1, ..., n$, $t_i \overset{iid}{\sim} T_{[0,\infty)(0,1)}$
- $\mathbf{X}^*$: A $n \times (m+1)$ matrix constructed by column binding $\mathbf{t}$ to $\mathbf{X}$
- $\mathbf{B}^*$: A $(m+1) \times p$ matrix constructed by row binding $\boldsymbol{\psi}^T$ to $\mathbf{B}$.

7.2 *Derivation of Full Conditional Distributions*

7.2.1 *Multivariate Skew-Normal Regression.* Without loss of generality, we derive the full conditional distributions for the multivariate skew-normal regression model component under the assumption that all observations belong to a single cluster. To make the extension to the case where more than one cluster is specified, simply apply these distributional forms to cluster specific parameters and data.

The multivariate skew-normal regression model can be written as follows.

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{t}\boldsymbol{\psi}^T + \mathbf{E} = \mathbf{X}^*\mathbf{B}^* + \mathbf{E}$$

For convenience, we define $\mathbf{X}^*$ as a $n \times (m+1)$ matrix constructed by column binding $\mathbf{t}$ to $\mathbf{X}$, and $\mathbf{B}^*$ as a $(m+1) \times p$ matrix constructed by row binding $\boldsymbol{\psi}^T$ to $\mathbf{B}$. We assume that $t_i \overset{iid}{\sim} T_{[0,\infty)}(0,1)$ and that $\mathbf{E}$ is made of row vectors $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, ..., \epsilon_{ip})$ for $i = 1, ..., n$, where $\boldsymbol{\epsilon}_i \overset{iid}{\sim} N_p(0, \boldsymbol{\Sigma})$.

The conditional likelihood for this model is given below.

$$p(\mathbf{Y}|\mathbf{X}^*, \mathbf{B}^*, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-n/2} \mathrm{exp}\left\{-\frac{1}{2}\mathrm{tr}(\mathbf{Y} - \mathbf{X}^*\mathbf{B}^*)^T(\mathbf{Y} - \mathbf{X}^*\mathbf{B}^*)\boldsymbol{\Sigma}^{-1}\right\}$$

We choose conjugate priors for $\mathbf{B}^*$ and $\boldsymbol{\Sigma}$ as follows.

$$\boldsymbol{\Sigma} \sim \text{inverse-Wishart}(\mathbf{V}_0, \nu_0)$$

$$\mathbf{B}^*|\boldsymbol{\Sigma} \sim MN_{(m+1)\times p}(\mathbf{B}_0^*, \mathbf{L}_0^{-1}, \boldsymbol{\Sigma})$$

We now dervive the joint posterior distribution of the parameters $\mathbf{B}^*$ and $\boldsymbol{\Sigma}$.

$$\begin{aligned}
p(\mathbf{B}^*, \boldsymbol{\Sigma}|\mathbf{X}^*, \mathbf{Y}) &\propto p(\mathbf{Y}|\mathbf{X}^*, \mathbf{B}^*, \boldsymbol{\Sigma})p(\mathbf{B}^*|\boldsymbol{\Sigma})p(\boldsymbol{\Sigma}) \\
&\propto |\boldsymbol{\Sigma}|^{-n/2}\mathrm{exp}\left\{-\frac{1}{2}\mathrm{tr}((\mathbf{Y} - \mathbf{X}^*\mathbf{B}^*)^T(\mathbf{Y} - \mathbf{X}^*\mathbf{B}^*)\boldsymbol{\Sigma}^{-1})\right\} \\
&\times |\boldsymbol{\Sigma}|^{-m/2}\mathrm{exp}\left\{-\frac{1}{2}\mathrm{tr}((\mathbf{B}^* - \mathbf{B}_0^*)^T\mathbf{L}_0(\mathbf{B}^* - \mathbf{B}_0^*)\boldsymbol{\Sigma}^{-1})\right\} \\
&\times |\boldsymbol{\Sigma}|^{(\nu_0+p+1)/2}\mathrm{exp}\left\{-\frac{1}{2}\mathrm{tr}(\mathbf{V}_0\boldsymbol{\Sigma}^{-1})\right\}
\end{aligned}$$

7.2.2 *Multinomial Logit Regression.*

7.2.3 *Multivariate Normal Conditional Imputation.* The multivariate normal conditional imputation derivations are given for a single cluster without loss of generality. In practive, the data and parameters in this section would be replaced by cluster specific estimates in the case of clustering.

For a given observation vector $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, we allow for missingness in at most $p-1$ of the multivariate outcomes through the use of a conditional imputation step embedded within our Gibbs sampler. Suppose $\mathbf{y}$ contains q missing observations and can be partitioned into two vectors $\mathbf{y_1}$ and $\mathbf{y_2}$ such that $\mathbf{y_1}$ is a $q \times 1$ vector of missing observations and $\mathbf{y_2}$ is a $(p-q) \times 1$ vector of complete observations. Similarly, partition $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as follows.

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

We will use these quantities to derive the conditional distribution $f(\mathbf{y_1}|\mathbf{y_2}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$.

$$\begin{aligned}
f(\mathbf{y_1}|\mathbf{y_2}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) &\propto f(\mathbf{y_1}, \mathbf{y_2}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
&\propto \exp\left\{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\} \\
&= \exp\left\{-\frac{1}{2}\begin{bmatrix} \mathbf{y}_1-\boldsymbol{\mu}_1 \\ \mathbf{y}_2-\boldsymbol{\mu}_2 \end{bmatrix}^T \boldsymbol{\Sigma}^{-1} \begin{bmatrix} \mathbf{y}_1-\boldsymbol{\mu}_1 \\ \mathbf{y}_2-\boldsymbol{\mu}_2 \end{bmatrix}\right\} \\
&= \exp\left\{-\frac{1}{2}\begin{bmatrix} \mathbf{y}_1-\boldsymbol{\mu}_1 \\ \mathbf{y}_2-\boldsymbol{\mu}_2 \end{bmatrix}^T \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y}_1-\boldsymbol{\mu}_1 \\ \mathbf{y}_2-\boldsymbol{\mu}_2 \end{bmatrix}\right\} \\
&= \exp\left\{-\frac{1}{2}\begin{bmatrix} \mathbf{y}_1-\boldsymbol{\mu}_1 \\ \mathbf{y}_2-\boldsymbol{\mu}_2 \end{bmatrix}^T \begin{bmatrix} \boldsymbol{\Sigma}^*_{11} & \boldsymbol{\Sigma}^*_{12} \\ \boldsymbol{\Sigma}^*_{21} & \boldsymbol{\Sigma}^*_{22} \end{bmatrix} \begin{bmatrix} \mathbf{y}_1-\boldsymbol{\mu}_1 \\ \mathbf{y}_2-\boldsymbol{\mu}_2 \end{bmatrix}\right\} \\
&= \exp\left\{-\frac{1}{2}\left[(\mathbf{y}_1-\boldsymbol{\mu}_{cond})^T\boldsymbol{\Sigma}_{cond}^{-1}(\mathbf{y}_1-\boldsymbol{\mu}_{cond})\right]\right\}
\end{aligned}$$

$$\Rightarrow \mathbf{y_1}|\mathbf{y_2}, \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim N_q(\boldsymbol{\mu}_{cond}, \boldsymbol{\Sigma}_{cond})$$

$$\boldsymbol{\mu}_{cond} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{y}_2-\boldsymbol{\mu}_2), \qquad \boldsymbol{\Sigma}_{cond} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$$

The blockwise inversion formula was used to invert $\mathbf{\Sigma}$ according to the following reparameterizations.

$$\mathbf{\Sigma}_{11}^{*} = \mathbf{\Sigma}_{11}^{-1} + \mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}$$

$$\mathbf{\Sigma}_{12}^{*} = -\mathbf{\Sigma}_{11}\mathbf{\Sigma}_{12}(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{-1}$$

$$\mathbf{\Sigma}_{21}^{*} = -(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}$$

$$\mathbf{\Sigma}_{22}^{*} = (\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{-1}$$