Bayesian Analysis of Structural Equation Models

Outline:

• Definition of SEMs
  1. Normal items
  2. Underlying normal
  3. Exponential family items

• Prior distributions - proper, conditionally conjugate

• Posterior computation via Gibbs sampling

• Issue with slow mixing & weak identifiability
Normal Structural Equation Models (SEMs)
(Bollen, 1989)

• Data for subject $i$ ($i = 1, \ldots, n$) consist of two multivariate vectors, $\mathbf{y}_i = (y_{i1}, \ldots, y_{ip})'$ and $\mathbf{x}_i = (x_{i1}, \ldots, x_{iq})'$.

• Each of these observed data vectors are assumed to consist of measurements of one or more latent variables.

• These continuous latent variables (or “traits”) typically have a substantive interpretation.
• For example, they may be quantities which can only be measured indirectly through a battery of *indicators* or *items*, e.g.,
  
  – poverty,
  – neighborhood quality,
  – intelligence,
  – neurological or reproductive health,

• Latent variables may also be introduced purely for convenience in modeling some feature of the data:

  – Parsimonious characterization of covariance in multivariate data (*computational advantages over alternatives*)
  – Covariance structure modeling - e.g., gene networks
  – Informative censoring - e.g., dropout time depends on latent health status in longitudinal study
  – Informative missingness - individuals not showing up to clinic have different health status
Components of SEM

1. Measurement Model

\[ y_i = \alpha + \Lambda_y \eta_i + \epsilon_i \]
\[ x_i = \beta + \Lambda_x \xi_i + \delta_i \]

\( \alpha, \beta = \) vectors of intercepts
\( \Lambda_y = \) coefficients (factor loadings) on latent variables \( \eta_i \)
\( \Lambda_x = \) coefficients on latent varaibles \( \xi_i \)
\( \epsilon_i \sim N_p(0, \Sigma_y), \Sigma_y = \) diagonal
\( \delta_i \sim N_q(0, \Sigma_x), \Sigma_x = \) diagonal

Conditional Independence Assumption: all the elements of \( y_i \)
and \( x_i \) are independent given the latent variables \( \eta_i \) and \( \xi_i \).

In other words - dependency on the shared latent variables explains all of the correlation in the observed variables.
2. Latent Variable Model:

$$\eta_i = B\eta_i + \Gamma\xi_i + \zeta_i,$$

$\eta_i = r \times 1$ vector of endogenous latent variables
$\xi_i = s \times 1$ vector of latent exogenous variables

$B = r \times r$ matrix characterizing dependency among $\eta_i$
- constrained so that $(I - B)$ non-singular

$\Gamma = r \times s$ matrix characterizing dependency of $\eta_i$ on $\xi_i$

$\zeta_i \sim N(0, I) = r \times 1$ disturbance vector - uncorrelated with
$\xi_i \sim N(0, I)$
Some Comments

• Clearly a very broad & flexible modeling framework for multivariate data.

• Identifiability will be an issue & it is appealing to choose a model (or models) which is identified by the observed data - even for Bayesians!

• In other words (at least for large samples), it is appealing to have information in the data about all of the parameters in the model.
Likelihood

The complete data likelihood of the observed \{\mathbf{x}, \mathbf{y}\} and latent \{\boldsymbol{\eta}, \boldsymbol{\xi}\} can be expressed as follows:

\[
L(\mathbf{y}, \mathbf{x}, \boldsymbol{\eta}, \boldsymbol{\xi}; \Theta) = \prod_{i=1}^{n} \left\{ \prod_{j=1}^{p} N(y_{ij}; \alpha_j + \lambda_j^{(y)' \eta_i}; \tau_{y,j}) \right\} \left\{ \prod_{h=1}^{q} N(x_{ih}; \beta_h + \lambda_h^{(x)' \xi_i}; \tau_{x,h}) \right\} \left\{ \prod_{l=1}^{r} N(\eta_{il}; \mathbf{b}_l^{'} \eta_{(l)} + \gamma_l^{'} \xi_i, 1) \right\} \left\{ \prod_{m=1}^{s} N(\xi_{im}; 0, 1) \right\}
\]

\(\lambda_j^{(y)} = j^{\text{th}}\) row of \(\mathbf{\Lambda}_y\)

\(\tau_{y,j} = j^{\text{th}}\) diagonal element of \(\mathbf{\Sigma}_y\)

\(\lambda_h^{(x)} = h^{\text{th}}\) row of \(\mathbf{\Lambda}_x\)

\(\tau_{x,h} = h^{\text{th}}\) diagonal element of \(\mathbf{\Sigma}_x\)

\(\mathbf{b}_l = l^{\text{th}}\) row of \(\mathbf{B}\) (excluding \(l^{\text{th}}\) column)

\(\eta_{(l)} = \text{subvector of } \boldsymbol{\eta}_i \text{ excluding } l^{\text{th}}\) element

\(\gamma_l = l^{\text{th}}\) row vector of \(\mathbf{\Gamma}\)

**Note:** We have a conditional linear regression structure
Let $\theta_1 = (\alpha', \lambda'_y)'$, where $\lambda_y = \text{vector of free elements of } \Lambda_y$
Let $\theta_2 = (\beta', \lambda'_x)'$, where $\lambda_x = \text{vector of free elements of } \Lambda_x$
Let $\theta_3 = (b, \gamma')$, where $b$ & $\gamma$ are free elements of $B, \Gamma$

Then, we can reexpress the likelihood as follows

$$L(y, x, \eta, \xi; \Theta) = \prod_{i=1}^{n} \left\{ \prod_{j=1}^{p} \mathcal{N}(y_{ij}; u'_{1ij} \theta_1, \tau_{y,j}) \right\}$$
$$\left\{ \prod_{h=1}^{q} \mathcal{N}(x_{ih}; u'_{2ih} \theta_2, \tau_{x,h}) \right\}$$
$$\left\{ \prod_{l=1}^{r} \mathcal{N}(\eta_{il}; u'_{3il} \theta_3, 1) \right\}$$
$$\left\{ \prod_{m=1}^{s} \mathcal{N}(\xi_{im}; 0, 1) \right\}$$

where the elements of $u_{1ij}$, $u_{2ih}$, $u_{3il}$ consist of indicator variables & latent variables

**Hence:** given the latent variables, we have 3 independent univariate linear regression models
Bayesian Specification

• To complete a Bayesian specification of the model, we choose prior distributions for the unknown coefficients, $\theta_1, \theta_2, \theta_3$, and residual variances, $\tau$.

• A convenient choice is:

$$
\pi(\theta_1, \theta_2, \theta_3) = \text{N}(\theta_{1,0}, \Sigma_{\theta_1}) \text{N}(\theta_{2,0}, \Sigma_{\theta_2}) \text{N}(\theta_{3,0}, \Sigma_{\theta_3})
$$

$$
\pi(\tau) = \left\{ \prod_{j=1}^{p} \text{IG}(\tau_{y,j}; a_{y,j}, b_{y,j}) \right\} \left\{ \prod_{h=1}^{q} \text{IG}(\tau_{x,h}; a_{x,h}, b_{x,h}) \right\}.
$$
Due to the conditionally linear regression structure, this prior is conditionally-conjugate, so that the full conditional posterior distributions of $\theta_1$, $\theta_2$, $\theta_3$ are multivariate normal:

$$
\pi(\theta_1 | \theta_2, \theta_3, \tau, \eta, \xi, y, x) = N(\hat{\theta}_1, \hat{\Sigma}_{\theta_1})
$$
$$
\pi(\theta_2 | \theta_1, \theta_3, \tau, \eta, \xi, y, x) = N(\hat{\theta}_2, \hat{\Sigma}_{\theta_2})
$$
$$
\pi(\theta_3 | \theta_1, \theta_2, \tau, \eta, \xi, y, x) = N(\hat{\theta}_3, \hat{\Sigma}_{\theta_3})
$$

\begin{align*}
\hat{\Sigma}_{\theta_1} &= \left( \Sigma_{\theta_1}^{-1} + \Sigma_{\eta}^{p} \sum_{j=1}^{p} \tau_{y,j}^{-1} \mathbf{u}_{1ij} \mathbf{u}_{1ij}' \right)^{-1} \\
\hat{\theta}_1 &= \hat{\Sigma}_{\theta_1} \left( \Sigma_{\theta_1}^{-1} \theta_{1,0} + \Sigma_{\eta}^{p} \sum_{j=1}^{p} \tau_{y,j}^{-1} \mathbf{u}_{1ij} \mathbf{y}_{ij} \right) \\
\hat{\Sigma}_{\theta_2} &= \left( \Sigma_{\theta_2}^{-1} + \Sigma_{\eta}^{q} \sum_{h=1}^{q} \tau_{x,h}^{-1} \mathbf{u}_{2ih} \mathbf{u}_{2ih}' \right)^{-1} \\
\hat{\theta}_2 &= \hat{\Sigma}_{\theta_2} \left( \Sigma_{\theta_2}^{-1} \theta_{2,0} + \Sigma_{\eta}^{q} \sum_{h=1}^{q} \tau_{x,h}^{-1} \mathbf{u}_{2ih} \mathbf{x}_{ih} \right) \\
\hat{\Sigma}_{\theta_3} &= \left( \Sigma_{\theta_3}^{-1} + \Sigma_{\eta}^{r} \sum_{l=1}^{r} \mathbf{u}_{3il} \mathbf{u}_{3il}' \right)^{-1} \\
\hat{\theta}_3 &= \hat{\Sigma}_{\theta_3} \left( \Sigma_{\theta_3}^{-1} \theta_{3,0} + \Sigma_{\eta}^{r} \sum_{l=1}^{r} \mathbf{u}_{3il} \mathbf{\eta}_{il} \right)
\end{align*}
• The full conditional posterior distribution for the residual variances are also conditionally-conjugate

\[
\pi(\tau_{y,j} \mid \theta, \eta, \xi, y, x) = IG(a_{y,j} + \frac{n}{2}, b_{y,j} + \frac{1}{2} \sum_{i=1}^{n} (y_{ij} - \alpha_j - \lambda_j(y')^{2})^2),
\]

\[
\pi(\tau_{x,h} \mid \theta, \eta, \xi, y, x) = IG(a_{x,h} + \frac{n}{2}, b_{x,h} + \frac{1}{2} \sum_{i=1}^{n} (x_{ih} - \beta_h - \lambda_h(x')^{2})^2)
\]

• Normal conditional posterior distributions can also be derived for the latent variables for each subject:

\[
\pi(\eta_i \mid \theta_1, \theta_2, \theta_3, \xi, y, x) = N(\hat{\eta}_i, \hat{V}_{\eta_i})
\]

\[
\pi(\xi_i \mid \theta_1, \theta_2, \theta_3, \eta, y, x) = N(\hat{\xi}_i, \hat{V}_{\xi_i})
\]

derivation of \(\hat{\eta}_i, \hat{V}_{\eta_i}, \hat{\xi}_i, \hat{V}_{\xi_i}\) is left as an exercise

- proceeds via standard algebraic routes
• Gibbs Sampling Algorithm:

1. Choose hyperparameters in the priors for \( \theta_1, \theta_2, \theta_3, \tau \) & initial values for these parameters & for the latent variables.

2. Update the latent variables \( \eta_i, \xi_i \) by sampling from their multivariate normal full conditional posterior distributions for \( i = 1, \ldots, n \).

3. Update the parameters, \( \theta_1, \theta_2, \theta_3 \), by sampling from their multivariate normal full conditional posterior distribution.

4. Update the residual variances, \( \tau \), by sampling each element from its inverse-gamma full conditional.

5. Repeat steps 2-3 a large number of times; discard a burn-in; and base inferences on posterior summaries calculated from the collected draws.

Again - underlying normal/probit models can be used to allow categorical \( y \)'s and \( x \)'s
Application - Spermatotoxicity Data

- **Interest**: effect of agent known to cause infertility at high dosages after short term exposure.

- Study designed to detect toxicant-induced changes in sperm motility at low doses

- Adult male rats randomized to one of four dose groups (0, 8, 24, 72), 10 animals/group.

- Rats dosed daily for 14 days, sperm were obtained from the proximal cauda epididymidis of each animal on day 15.

- Sperm had been exposed to the toxicant during their final differentiation in the testis & maturation in the caput and corpus epididymidis. It is during this time that the capability for progressive motion develops in sperm.
• Using computer aided sperm analysis (CASA) (reviewed by Boyers, Davis, and Katz, 1989), the proportion of motile (surviving) sperm was recorded for a sample of 100-200 sperm from each rat, as were the \((x, y)\) coordinates of points (60 per sec) along the path travelled by each surviving sperm.

• We use three kinematic measures (LDV: linear displacement velocity, LNR: linearity, and PRD: predictability), to quantify the progressive motility of each surviving sperm.
Summary of sperm motility data (LDV = linear displacement velocity, LNR = linearity, PRD = predictability).

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Dose</th>
<th>Mean†</th>
<th>SD†</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDV</td>
<td>0</td>
<td>88.4</td>
<td>9.21</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>76.1</td>
<td>7.54</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>82.1</td>
<td>15.6</td>
</tr>
<tr>
<td></td>
<td>72</td>
<td>77.2</td>
<td>13.3</td>
</tr>
<tr>
<td>LNR</td>
<td>0</td>
<td>0.219</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.216</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>0.207</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>72</td>
<td>0.206</td>
<td>0.020</td>
</tr>
<tr>
<td>PRD</td>
<td>0</td>
<td>25.5</td>
<td>2.70</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>22.0</td>
<td>2.52</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>24.3</td>
<td>5.05</td>
</tr>
<tr>
<td></td>
<td>72</td>
<td>23.6</td>
<td>3.38</td>
</tr>
</tbody>
</table>

† Statistics are based on average values for each animal
• Let’s initially focus on the case in which data are collapsed into animal-specific means, $y_i = (y_{i1}, y_{i2}, y_{i3})'$.

• We then consider the model:

$$y_{ij} = \alpha_j + \lambda_j \eta_i + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \tau_j),$$
$$\eta_i = \beta x_i + \delta_i, \quad \delta_i \sim N(0, 1)$$

$\alpha = (\alpha_1, \alpha_2, \alpha_3)'$ are intercepts
$\lambda = (\lambda_1, \lambda_2, \lambda_3)'$ are factor loadings
$\beta =$ dose effect on latent progressive motility

• Note this is a special case of the SEM in which $x_i = \xi_i$ (latent predictor is observed)
Homework Exercise

• Write down the complete data likelihood for the sperm example.
• Choose prior distributions for the parameters (identifiability).
• Derive conditional posterior distributions for each of the unknowns.
• Outline a Gibbs sampling algorithm for posterior computation.