STAT5020 Assignment2

WU Zhihao

March 22, 2022

Question 1

Consider the following linear SEM with dichotomous (y_{i1}, y_{i2}, y_{i3}), continuous $(y_{i4}, y_{i5}, y_{i6}, y_{i7})$, and binary variables (y_{i8}, y_{i9}):

$$\begin{cases} y_{ik}^* = \mu_k + \boldsymbol{\lambda}_k \boldsymbol{\omega}_i + \epsilon_{ik}, & \text{for } k = 1, 2, 3, \\ y_{ik} = \mu_k + \boldsymbol{\lambda}_k \boldsymbol{\omega}_i + \epsilon_{ik}, & \text{for } k = 4, 5, 6, 7, \\ \vartheta_{ik} = \mu_k + \boldsymbol{\lambda}_k \boldsymbol{\omega}_i, \quad \text{for } k = 8, 9, & \\ \eta_i = b \times d_i + \gamma_1 \times \xi_{1i} + \gamma_2 \times \xi_{2i} + \delta_i, & \\ \boldsymbol{\xi}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Phi}),\ \delta_i \sim \mathcal{N}(0, \psi_\delta),\ \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}_{\boldsymbol{\epsilon}}), & \end{cases} \quad \text{for } i = 1, \ldots, 500, \tag{1}$$

where

- $\boldsymbol{\omega}_i = (\eta_i, \xi_{1i}, \xi_{2i})^\top$ is a 3×1 vector of latent variable,
- y_{ik}^* is the latent continuous measurement for dichotomous observed variable y_{ik}, for $k = 1, 2, 3$,
- ϑ_{ik} is the canonical parameter for binary y_{ik}, for $k = 8, 9$,
- the first three diagonal elements in $\boldsymbol{\Psi}_{\boldsymbol{\epsilon}}$ are fixed and equal to 1, due to identifiability.

For the dichotomous variables, the conditional distributions of the related latent continuous variables:

$$y_{ik}^* \sim \begin{cases} \mathcal{N}(\mu_k + \boldsymbol{\lambda}_k \boldsymbol{\omega}_i, 1)\mathrm{I}_{(-\infty,0)}(y_{ik}^*) & \text{for } y_{ik} = 0, \\ \mathcal{N}(\mu_k + \boldsymbol{\lambda}_k \boldsymbol{\omega}_i, 1)\mathrm{I}_{(0,\infty)}(y_{ik}^*) & \text{for } y_{ik} = 1, \end{cases} \quad \text{for } k = 1, 2, 3, \tag{2}$$

where $\mathrm{I}_A(y)$ is an indicator function that takes the value 1 if y is in A, and 0 otherwise.

For the binary variables, the distributions are

$$\begin{cases} y_{ik} \sim \text{Bernoulli}(p_{ik}), \\ p_{ik} = \dfrac{\exp\{\vartheta_{ik}\}}{1 + \exp\{\vartheta_{ik}\}}, \end{cases} \quad \text{for } k = 8, 9. \tag{3}$$

The following prior inputs of the hyperparameter values in the conjugate prior distributions of the parameters are considered:

- location hyperparameters for normal distributions are set equal to the true values,
- scale hyperparameters for normal distributions (covariance matrix) are taken to be diagonal matrix with diagonal elements equal to 0.16,
- for inverse gamma distribution, the first and second hyperparameters are equal to 9 and 3,
- for inverse Wishart distribution, the two hyperparameters are $\rho_0 = 10, \mathbf{R}_0^{-1} = 7\mathbf{I}$.

To summarise the result of Bayesian analysis based on 10 replications, bias and RMS are calculated in the following table.

	γ_1	γ_2	b	λ_2	λ_3	λ_5	λ_6	λ_7	λ_9
BIAS	-0.00009	0.00468	-0.00458	-0.0019	-0.01069	0.00814	0.00580	0.00246	-0.03997
RMS	0.01679	0.01975	0.01420	0.02788	0.0334	0.02567	0.03410	0.02462	0.13967
	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	μ_9
BIAS	0.00467	0.00095	-0.00512	-0.01290	-0.00285	-0.01337	-0.01402	0.02343	0.03102
RMS	0.02386	0.02107	0.01922	0.04127	0.03779	0.02338	0.04288	0.13771	0.09301
	$\psi_{\epsilon 4}$	$\psi_{\epsilon 5}$	$\psi_{\epsilon 6}$	$\psi_{\epsilon 7}$	ψ_{δ}	ϕ_{11}	ϕ_{21}	ϕ_{12}	ϕ_{22}
BIAS	0.00082	-0.00932	-0.00734	-0.00949	0.01944	-0.03261	0.00319	0.00319	0.07710
RMS	0.01120	0.02182	0.04059	0.03541	0.02083	0.10400	0.04350	0.04350	0.18017

Table 1: Bias and RMS of 10 replications.

Question 2

a) The measurement equation model in matrix form is

$$
\begin{bmatrix} y_{i1}^{(g)} \\ y_{i2}^{(g)} \\ y_{i3}^{(g)} \\ \tilde{z}_{i1}^{(g)} \\ \tilde{z}_{i2}^{(g)} \\ \tilde{z}_{i3}^{(g)} \\ \tilde{z}_{i4}^{(g)} \\ \tilde{z}_{i5}^{(g)} \\ \vartheta_{i4}^{(g)} \\ \vartheta_{i5}^{(g)} \\ \vartheta_{i6}^{(g)} \\ \vartheta_{i7}^{(g)} \end{bmatrix}
=
\begin{bmatrix} \mu_1^{(g)} \\ \mu_2^{(g)} \\ \mu_3^{(g)} \\ \mu_4^{(g)} \\ \mu_5^{(g)} \\ \mu_6^{(g)} \\ \mu_7^{(g)} \\ \mu_8^{(g)} \\ \mu_9^{(g)} \\ \mu_{10}^{(g)} \\ \mu_{11}^{(g)} \\ \mu_{12}^{(g)} \end{bmatrix}
+
\begin{bmatrix} 1 & 0 & 0 & 0 \\ \lambda_{21}^{(g)} & 0 & 0 & 0 \\ \lambda_{31}^{(g)} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & \lambda_{52}^{(g)} & 0 & 0 \\ 0 & \lambda_{62}^{(g)} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \lambda_{83}^{(g)} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & \lambda_{10,4}^{(g)} \\ 0 & 0 & 0 & \lambda_{11,4}^{(g)} \\ 0 & 0 & 0 & \lambda_{12,4}^{(g)} \end{bmatrix}
\begin{bmatrix} \eta_i^{(g)} \\ \xi_{i1}^{(g)} \\ \xi_{i2}^{(g)} \\ \xi_{i3}^{(g)} \end{bmatrix}
+
\begin{bmatrix} \epsilon_{i1}^{(g)} \\ \epsilon_{i2}^{(g)} \\ \epsilon_{i3}^{(g)} \\ \epsilon_{i4}^{(g)} \\ \epsilon_{i5}^{(g)} \\ \epsilon_{i6}^{(g)} \\ \epsilon_{i7}^{(g)} \\ \epsilon_{i8}^{(g)} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad i = 1, \cdots, N_g, \quad g = 1, 2. \tag{4}
$$

where

- observed variables $y_{ik}^{(g)}$, $k = 1, 2, 3$, are continuous,
- $\tilde{z}_{ik}^{(g)}$ is the latent continuous measurement for corresponding ordered categorical variable $z_{ik}^{(g)}$,
- $\vartheta_{ik}^{(g)}$ is the canonical parameter for variable $y_{ik}^{(g)}$, $k = 4, 5, 6, 7$, from exponential family distributions,

and the structural equation model is

$$
\eta_i^{(g)} = \gamma_1^{(g)} \xi_{i1}^{(g)} + \gamma_2^{(g)} \xi_{i2}^{(g)} + \gamma_3^{(g)} \xi_{i3}^{(g)} + \delta_i^{(g)}. \tag{5}
$$

The conditions needed for model identification:

1. for the loading matrix, fix the first non-zero element in each column equal to 1,
2. for the random measurement (residual) errors, assume that the ones related to the ordered categorical variables in the measurement equation model follow standard normal distribution, i.e.

$$
\epsilon_k^{(g)} \sim \mathcal{N}(0, 1), \quad k = 4, \cdots, 8, \tag{6}
$$

3. for the thresholds related to the ordered categorical variables $z_{ik}^{(g)}$s, select the first group as the reference group, impose the following restrictions to have the same scale among groups:

$$
\alpha_{k,m}^{(1)} = \alpha_{k,m}^{(2)}, \quad k = 1, \cdots, 5, \ m = 1, \cdots, b_k, \tag{7}
$$

and fix the both end thresholds of the first group:

$$
\alpha_{k,1}^{(1)} = \Phi^{-1}(f_{k,1}^{*(1)}), \ \alpha_{k,b_k}^{(1)} = \Phi^{-1}(f_{k,b_k}^{*(1)}), \tag{8}
$$

where $\Phi(\cdot)$ is the distribution function of standard normal distribution, $f_{k,1}^{*(1)}$ is the first-group frequency of the first category, while $f_{k,b_k}^{*(1)}$ is the first-group cumulative frequency of the first $(b_k - 1)$ categories.

b) Since in the scenario of multisample data, the joint prior distribution for the factor loading matrix and the the unique variance of the error measurement may cause problem under the constrained situation

$$\mathbf{\Lambda}^{(1)} = \mathbf{\Lambda}^{(2)} = \mathbf{\Lambda}, \quad \mathbf{\Psi_\epsilon}^{(1)} \neq \mathbf{\Psi_\epsilon}^{(2)}, \tag{9}$$

then we consider the normal prior for each $\mathbf{\Lambda}^{(g)}$, and independent inverse Gamma prior for each $\mathbf{\Psi_\epsilon}^{(g)}$, while in the scenario of uni-sample data, we usual consider the inverse Gamma prior for $\mathbf{\Psi_\epsilon}$ and the conditional normal prior distribution for $\mathbf{\Lambda} \,|\, \mathbf{\Psi_\epsilon}$.

c) <u>**Bayes Factor**</u>

Define

$$\mathbf{\Lambda}^{(g)} = \begin{bmatrix} \vdots \\ \mathbf{\Lambda}_k^{(g)} \\ \vdots \end{bmatrix}, \quad \boldsymbol{\omega}_i^{(g)} = \begin{bmatrix} \eta_i^{(g)} \\ \xi_{i1}^{(g)} \\ \xi_{i2}^{(g)} \\ \xi_{i3}^{(g)} \end{bmatrix} \tag{10}$$

where row-vector $\mathbf{\Lambda}_k^{(g)}$ is the k-th row of matrix $\mathbf{\Lambda}^{(g)}$.

Rewrite the measurement equation model with vector and matrix notation

$$M_1: \quad \boldsymbol{y}_i^{(g)} = \boldsymbol{\mu}^{(g)} + \mathbf{\Lambda}^{(g)}\boldsymbol{\omega}_i^{(g)} + \boldsymbol{\epsilon}_i^{(g)}. \tag{11}$$

For the hypothesis testing problem

$$H_0: \ \mathbf{\Lambda}^{(1)} = \mathbf{\Lambda}^{(2)} = \mathbf{\Lambda} \text{ v.s. } \ H_1: \ \mathbf{\Lambda}^{(1)} \neq \mathbf{\Lambda}^{(2)}, \tag{12}$$

consider another measurement equation model

$$M_2: \quad \boldsymbol{y}_i^{(g)} = \boldsymbol{\mu}^{(g)} + \mathbf{\Lambda}\boldsymbol{\omega}_i^{(g)} + \boldsymbol{\epsilon}_i^{(g)}, \tag{13}$$

and convert this hypothesis testing problem to the model comparison problem between M_1 and M_2.

Consider the link model M_t for linking M_1 and M_2,

$$M_t: \quad \boldsymbol{y}_i^{(g)} = \boldsymbol{\mu}^{(g)} + t\mathbf{\Lambda}^{(g)}\boldsymbol{\omega}_i^{(g)} + (1-t)\mathbf{\Lambda}\boldsymbol{\omega}_i^{(g)} + \boldsymbol{\epsilon}_i^{(g)}. \tag{14}$$

where for $t = 1$, model M_t reduces to M_1, while for $t = 0$, model M_t reduces to M_2.

Let $p(\boldsymbol{Y}, \boldsymbol{Z}, \tilde{\boldsymbol{Z}}, \mathbf{\Omega} \,|\, \boldsymbol{\theta}, t)$ be the complete-data likelihood of the link model M_t, and

$$\begin{aligned} U(\boldsymbol{\theta}, \tilde{\boldsymbol{Z}}, \mathbf{\Omega}, \boldsymbol{Y}, \boldsymbol{Z}, t) &= \frac{d}{dt}\log p(\boldsymbol{Y}, \boldsymbol{Z}, \tilde{\boldsymbol{Z}}, \mathbf{\Omega} \,|\, \boldsymbol{\theta}, t) \\ &= \sum_{g=1}^{2}\sum_{k=1}^{3}\frac{1}{\psi_{\epsilon k}^{(g)}}\left[y_{ik}^{(g)} - \mu_k^{(g)} - t\mathbf{\Lambda}_k^{(g)}\boldsymbol{\omega}_i^{(g)} - (1-t)\mathbf{\Lambda}_k\boldsymbol{\omega}_i^{(g)}\right]\left(\mathbf{\Lambda}_k^{(g)} - \mathbf{\Lambda}_k\right)\boldsymbol{\omega}_i^{(g)} \\ &\quad + \sum_{g=1}^{2}\sum_{k=4}^{8}\frac{1}{\psi_{\epsilon k}^{(g)}}\left[\tilde{z}_{i(k-3)}^{(g)} - \mu_k^{(g)} - t\mathbf{\Lambda}_k^{(g)}\boldsymbol{\omega}_i^{(g)} - (1-t)\mathbf{\Lambda}_k\boldsymbol{\omega}_i^{(g)}\right]\left(\mathbf{\Lambda}_k^{(g)} - \mathbf{\Lambda}_k\right)\boldsymbol{\omega}_i^{(g)} \\ &\quad + \sum_{g=1}^{2}\sum_{k=9}^{12}\frac{1}{\psi_{\epsilon k}^{(g)}}\left[y_{ik} - \dot{b}\left(\mu_k^{(g)} + t\mathbf{\Lambda}_k^{(g)}\boldsymbol{\omega}_i^{(g)} + (1-t)\mathbf{\Lambda}_k\boldsymbol{\omega}_i^{(g)}\right)\right]\left(\mathbf{\Lambda}_k^{(g)} - \mathbf{\Lambda}_k\right)\boldsymbol{\omega}_i^{(g)}. \end{aligned} \tag{15}$$

On the fixed grid $t = t_{(s)}$, generate observations $(\boldsymbol{\theta}^{(j)}, \tilde{\boldsymbol{Z}}^{(j)}, \mathbf{\Omega}^{(j)})$ from the joint posterior distribution

$\left[\boldsymbol{\theta}, \tilde{\boldsymbol{Z}}, \boldsymbol{\Omega} \mid \boldsymbol{Y}, \boldsymbol{Z}, t_{(s)}\right]$ and calculate

$$\bar{U}_{(s)} = \frac{1}{J} U(\boldsymbol{\theta}^{(j)}, \tilde{\boldsymbol{Z}}^{(j)}, \boldsymbol{\Omega}^{(j)}, \boldsymbol{Y}, \boldsymbol{Z}, t_{(s)}) \tag{16}$$

then the estimated log-transformed Bayes Factor is

$$\widehat{\log B_{12}} = \frac{1}{2} \sum_{s=0}^{S} (t_{(s+1)} - t_{(s)}) \left(\bar{U}_{(s+1)} + \bar{U}_{(s)}\right). \tag{17}$$

DIC

For each model M_k, (k = 1, 2), the DIC_k is estimated by

$$\widehat{\mathrm{DIC}_k} = -\frac{4}{J} \sum_{j=1}^{J} \log p(\boldsymbol{Y} \mid \boldsymbol{\theta}_k^{(j)}, M_k) + 2 \log p(\boldsymbol{Y} \mid \tilde{\boldsymbol{\theta}}_k, M_k) \tag{18}$$

where $\left\{\boldsymbol{\theta}_k^{(j)}, j = 1, \cdots J\right\}$ is a sample of observations from the posterior distribution under model M_k, and the $\tilde{\boldsymbol{\theta}}_k$ is the Bayesian estimate of $\boldsymbol{\theta}_k$.