Bayesian growth curve modeling with non-ignorable missing data

Zhiyong Zhang
University of Notre Dame

Model and estimation method

Let $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})'$ be a $T \times 1$ random vector and y_{it} be the observation of individual i at time t $(i = 1, \ldots, N; t = 1, \ldots, T)$. For a growth curve model, we have

$$\mathbf{y}_i = \mathbf{\Lambda}\boldsymbol{\eta}_i + \mathbf{e}_i$$

where $\mathbf{\Lambda}$ is a $T \times q$ factor loading marix determining the growth trajectory, $\boldsymbol{\eta}_i$ is a $q \times 1$ random vector, and $\mathbf{e}_i$ is a vector of residuals or measurement erros. $\mathbf{e}_i$ is often assumed to be normally distributed as $\mathbf{e}_i \sim MN_T(0, \mathbf{\Theta})$. Many times, it is assumed that $\mathbf{\Theta} = \mathbf{I}\phi$ where ϕ is a scalar and $\mathbf{I}$ is an identity matrix. $\boldsymbol{\eta}_i$ are often called random effects because they are different for each individual. The means of $\boldsymbol{\eta}_i$ are fixed effects so that

$$\boldsymbol{\eta}_i = \boldsymbol{\beta} + \boldsymbol{\epsilon}_i$$

where $\boldsymbol{\epsilon}_i \sim MN_q(\mathbf{0}, \mathbf{\Psi})$.

Let $\mathbf{m}_i = (m_{i1}, \ldots, m_{iT})'$ be a vector indicating the missingness in $\mathbf{y}_i$. If $m_{it} = 1$, then y_{it} is missing. Otherwise, $m_{it} = 0$. Assume that missingnesss in $\mathbf{y}_i$ is related to the random effects $\boldsymbol{\eta}_i$ such that

$$\text{logit}(m_{it}) = \boldsymbol{\lambda}_t(1, \boldsymbol{\eta}_i^t)^t = \boldsymbol{\lambda}_t\boldsymbol{\eta}_i^*$$

where $\boldsymbol{\lambda}_t = (\lambda_0, \ldots, \lambda_q)$. For convience, we do not distinguish $\boldsymbol{\eta}_i$ and $\boldsymbol{\eta}_i^*$. The difference between them should be clear in the context. For a study with missing data, the information should include $\mathbf{y}_{i,obs}$ and $\mathbf{m}_i$.

Data augmentation

By augmenting the latent variables $\boldsymbol{\eta}_i$ with $\mathbf{y}_i$ and $\mathbf{m}_i$, the joint distribution is

$$p(\mathbf{y}_i, \mathbf{m}_i, \boldsymbol{\eta}_i|\phi, \mathbf{\Lambda}, \mathbf{\Psi}, \beta, \boldsymbol{\lambda}_t) = p(\mathbf{y}_i|\boldsymbol{\eta}_i, \mathbf{\Lambda}, \phi)p(\boldsymbol{\eta}_i|\boldsymbol{\beta}, \mathbf{\Psi})p(\mathbf{m}_i|\boldsymbol{\eta}_i, \boldsymbol{\lambda}_t).$$

We further assume that there is an underlying normal variable $z_{it} \sim N(\boldsymbol{\lambda}_t\boldsymbol{\eta}_i^*, 1)$ for each m_{it}. If $z_{it} > 0$, then $m_{it} = 1$. Otherwise, $m_{it} = 0$. By augmenting z_{it} with m_{it}, the joint distribution of m_{it} and z_{it} is

$$p(m_{it}, z_{it}|\boldsymbol{\lambda}_t) = p(m_{it}|z_{it})p(z_{it}|\boldsymbol{\eta}_i, \boldsymbol{\lambda_t}).$$

The distribution of z_{it} - $p(z_{it}|\boldsymbol{\eta}_i, \boldsymbol{\lambda_t})$ -is known as a normal distribution with mean $\boldsymbol{\lambda}_t\boldsymbol{\eta}_i^*$ and variance 1 and we need to get the distribution for m_{it} conditional on z_{it}. Note that

$$\begin{aligned} p(m_{it} &= 1|z_{it} > 0) = 1, \;\; p(m_{it} = 1|z_{it} \leq 0) = 0, \\ p(m_{it} &= 0|z_{it} > 0) = 0, \;\; p(m_{it} = 0|z_{it} \leq 0) = 1. \end{aligned}$$

Thus, the distribution for $m_{it}|z_{it}$ can be expressed as

$$p(m_{it}|z_{it}) = \mathcal{I}(m_{it} = 1)\mathcal{I}(z_{it} > 0) + \mathcal{I}(m_{it} = 0)\mathcal{I}(z_{it} \leq 0)$$

where $\mathcal{I}(A)$ is an indicator function which takes 1 if the expression A is true and otherwise 0.

Futher augmenting z_{it}, the joint distribution becomes

$$p(\mathbf{y}_i, \mathbf{m}_i, \boldsymbol{\eta}_i, \mathbf{z}_i|\phi, \boldsymbol{\Lambda}, \boldsymbol{\Psi}, \beta, \boldsymbol{\lambda}_t) = p(\mathbf{y}_i|\boldsymbol{\eta}_i, \boldsymbol{\Lambda}, \phi)p(\boldsymbol{\eta}_i|\boldsymbol{\beta}, \boldsymbol{\Psi})\prod_{t=1}^{T} p(m_{it}|z_{it})p(z_{it}|\boldsymbol{\eta}_i, \boldsymbol{\lambda_t}).$$

Then the likelihood function is

$$\begin{aligned} L &= \prod_{i=1}^{n} p(\mathbf{y}_i, \mathbf{m}_i, \boldsymbol{\eta}_i, \mathbf{z}_i|\phi, \boldsymbol{\Lambda}, \boldsymbol{\Psi}, \beta, \boldsymbol{\lambda}_t) \\ &= \prod_{i=1}^{n}\left[p(\mathbf{y}_i|\boldsymbol{\eta}_i, \boldsymbol{\Lambda}, \phi)p(\boldsymbol{\eta}_i|\boldsymbol{\beta}, \boldsymbol{\Psi})\prod_{t=1}^{T} p(m_{it}|z_{it})p(z_{it}|\boldsymbol{\eta}_i, \boldsymbol{\lambda_t})\right] \\ &\propto \prod_{i=1}^{n}\Bigg\{|\mathbf{I}\phi|^{-1/2}\exp\left[-\frac{1}{2}(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i)'(\mathbf{I}\phi)^{-1}(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i)\right] \\ &\quad \times|\boldsymbol{\Psi}|^{-1/2}\exp\left[-\frac{1}{2}(\boldsymbol{\eta}_i - \boldsymbol{\beta})'\boldsymbol{\Psi}^{-1}(\boldsymbol{\eta}_i - \boldsymbol{\beta})\right] \\ &\quad \times\prod_{t=1}^{T}\mathcal{I}(m_{it}=1)\mathcal{I}(z_{it}>0) + \mathcal{I}(m_{it}=0)\mathcal{I}(z_{it}\leq 0) \\ &\quad \times\exp\left[-\frac{1}{2}(z_{it} - \boldsymbol{\lambda}_t\boldsymbol{\eta}_i^*)^2\right]\Bigg\} \end{aligned} \tag{1}$$

Priors

For $\boldsymbol{\beta}$, a multivariate normal prior is used

$$p(\boldsymbol{\beta}) = MN_q(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0) = (2\pi)^{-q/2}|\boldsymbol{\Sigma}_0|^{-1/2}\exp\left[-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right].$$

For $\boldsymbol{\Psi}$, the inverse Wishart prior is used

$$p(\boldsymbol{\Psi}) = IW(m_0, V_0) = \frac{|V_0|^{m_0/2}|\boldsymbol{\Psi}|^{-\frac{m_0+q+1}{2}}\exp\left[-\mathrm{tr}(V_0\boldsymbol{\Psi}^{-1})/2\right]}{2^{m_0 q/2}\Gamma(m_0/2)}.$$

For ϕ, the inverse gamma distribution prior is used

$$p(\phi) = IG(\alpha_0, \gamma_0) = \frac{\gamma_0^{\alpha_0}}{\Gamma(\alpha_0)}\phi^{-(\alpha_0+1)}\exp\left(-\frac{\gamma}{\phi_0}\right).$$

For $\boldsymbol{\lambda}_t$, a multivariate normal prior is also used

$$p(\boldsymbol{\lambda}_t) = MN_{q+1}(\boldsymbol{\lambda}_{t0}, \mathbf{S}_{t0}) = (2\pi)^{-q/2}|\mathbf{S}_{t0}|^{-1/2}\exp\left[-\frac{1}{2}(\boldsymbol{\lambda}_t - \boldsymbol{\lambda}_{t0})'\mathbf{S}_{t0}^{-1}(\boldsymbol{\lambda}_t - \boldsymbol{\lambda}_{t0})\right].$$

Posteriors

With the likelihood function and the priors, the joint posterior distribution is readily available,

$$p(\phi, \boldsymbol{\Psi}, \beta, \boldsymbol{\lambda}_t, t = 1, \ldots, T|\mathbf{y}_i, \mathbf{m}_i, \boldsymbol{\eta}_i, \mathbf{z}_i) = L \times p(\boldsymbol{\beta}) \times p(\boldsymbol{\Psi}) \times p(\phi)\prod_{t=1}^{T} p(\boldsymbol{\lambda}_t).$$

Alghough the Markov chains can be generated directly from the joint posterior distribution, to improve efficiency, we first obtain the conditional posterior distributions for model parameters and then use Gibbs sampling procedure to generate Markov chains.

The conditional posterior distribution for ϕ is an inverse gamma distribution

$$p(\phi|\boldsymbol{\beta}, \boldsymbol{\eta}_i, \mathbf{y}_i, i = 1, \ldots, n) = IG(\alpha_1, \gamma_1)$$

where

$$\alpha_1 = \alpha + \frac{nT}{2}$$

and

$$\gamma_1 = \gamma + \frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i)'(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i).$$

The conditional posterior distribution for $\boldsymbol{\beta}$ is a multivariate normal distribution

$$p(\boldsymbol{\beta}|\boldsymbol{\Psi}, \boldsymbol{\eta}_i, i = 1, \ldots, n) = MN_q(\boldsymbol{\beta}_1, \boldsymbol{\Sigma}_1)$$

where

$$\boldsymbol{\beta}_1 = \boldsymbol{\Sigma}_1(\boldsymbol{\Psi}^{-1}\sum_{i=1}^{n}\boldsymbol{\eta}_i + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\beta}_0)$$

and

$$\boldsymbol{\Sigma}_1 = (n\boldsymbol{\Psi}^{-1} + \boldsymbol{\Sigma}_0^{-1})^{-1}.$$

The conditional posterior distribution for $\boldsymbol{\Psi}$ is an inverse Wishart distribution

$$p(\boldsymbol{\Psi}|\boldsymbol{\beta}, \boldsymbol{\eta}_i, i = 1, \ldots, n) = IW(m_1, V_1)$$

where

$$m_1 = m_0 + n$$

and

$$V_1 = V_0 + \sum_{i=1}^{n}(\boldsymbol{\eta}_i - \boldsymbol{\beta})(\boldsymbol{\eta}_i - \boldsymbol{\beta})'.$$

The conditional posterior distribution for $\boldsymbol{\lambda}_t, t = 1, \ldots, T$ is a multivaraite normal distribution

$$p(\boldsymbol{\lambda}_t|\boldsymbol{\eta}_i, z_{it}, i = 1, \ldots, n) = MN_{q+1}(\boldsymbol{\lambda}_{t1}, \mathbf{S}_{t1})$$

where

$$\boldsymbol{\lambda}_{t1} = \mathbf{S}_{t1}(\mathbf{S}_{t0}^{-1}\lambda_{t0} + \boldsymbol{\eta}_*'\boldsymbol{z}_{*t})$$

and

$$\mathbf{S}_{t1} = (\mathbf{S}_{t0}^{-1} + \boldsymbol{\eta}_*'\boldsymbol{\eta}_*)^{-1}$$

with $\boldsymbol{\eta}_* = (\boldsymbol{\eta}_1^*, \ldots, \boldsymbol{\eta}_n^*)'$ and $\boldsymbol{z}_{*t} = (z_{i1}, \ldots, z_{in})'$.

The conditional posterior distribution for $\boldsymbol{\eta}_i$ is a multivariate normal distribution with

$$p(\boldsymbol{\eta}_i|\boldsymbol{\beta}, \boldsymbol{\Psi}, \phi, \boldsymbol{\lambda}_t, \mathbf{y}_i, t = 1, \ldots, T) = MN_q(\boldsymbol{\eta}_{i1}, U_{i1})$$

where

$$\boldsymbol{\eta}_{i1} = U_{i1}(\boldsymbol{\Lambda}'\phi\mathbf{y}_i + \boldsymbol{\Psi}^{-1}\boldsymbol{\beta} + \boldsymbol{\lambda}_*'\boldsymbol{a}_i)$$

and

$$U_{i1} = (\mathbf{\Lambda}'\phi\mathbf{\Lambda} + \mathbf{\Psi}^{-1} + \boldsymbol{\lambda}_*'\boldsymbol{\lambda}_*)^{-1}$$

with $\boldsymbol{a}_i = (z_{i1} - \lambda_{10}, \ldots, z_{it} - \lambda_{t0})'$ and

$$\boldsymbol{\lambda}_* = \begin{pmatrix} \lambda_{11} & \ldots & \lambda_{1q} \\ \vdots & \ddots & \vdots \\ \lambda_{T1} & \ldots & \lambda_{Tq} \end{pmatrix}.$$

The conditional posterior distribution for the underlying variable z_{it} is a truncated normal distribution,

$$z_{it}|\boldsymbol{\lambda}_t, \boldsymbol{\eta}_i, m_{it} \sim \begin{cases} N(\boldsymbol{\lambda}_t\boldsymbol{\eta}_i^*, 1)I(0, +\infty) & m_{it} = 1 \\ N(\boldsymbol{\lambda}_t\boldsymbol{\eta}_i^*, 1)I(-\infty, 0] & m_{it} = 0 \end{cases}.$$

Finally, the conditional posterior distribution for the missing data $y_{it.}^{miss}$ is a normal distribution

$$y_{it}^{miss} = N(\Lambda_t\boldsymbol{\eta}_i, \phi)$$

where Λ_t is the tth row of $\mathbf{\Lambda}$.

With the full set of conditional posterior distributions, the Gibbs sampling procedure can be implemented.

Simulation studies

Data will be generated from the following linear growth curve model,

$$\begin{aligned} y_{it} &= L_i + S_i(t-1) + e_{it} \\ L_i &= \beta_1 + v_{1i} \\ S_i &= \beta_2 + v_{2i} \end{aligned},$$

where $e_{it} \sim N(0, \phi)$ and $Cov\left[(v_{1i}, v_{2i})'\right] = \mathbf{\Psi} = \begin{pmatrix} \sigma_L^2 & \sigma_{LS} \\ \sigma_{LS} & \sigma_S^2 \end{pmatrix}$. We set the population parameters at $\beta_1 = 50$, $\beta_2 = 5$, $\phi = 25$, and

$$\mathbf{\Psi} = \begin{pmatrix} 100 & 0 \\ 0 & 25 \end{pmatrix}.$$

Another way to set up the population parameters is to first analyze the empirical example and then use the parameter estiamtes as population values.

The number of measurement occasions can be set at $T = 4$ and 5. For the sample size, $n = 100, 200$ and 500 can be used. The missing data proportions can be set at 10%, 20%, and 40%.

The following simulation studies should be conducted.

1. Investigate parameter estimate bias when the MNAR data are analyze as MAR data.
2. Investigate whether the model parameters can be recovered accurately when missingness depends on $\boldsymbol{\eta}_i$ using the Bayesian proposed.
3. Investigate whether the method can recover the parameters when missingness depends on y.
4. Investigate whether the method can recover the parameters when missingness depends on some unobserved variables.

An empirical example

Pull out a subset of data from the National Longitudianl Study of Youth data base. Conduct the empirical data analysis based on the data.

Appendix
A simple example

```
#WinBUGS codes generated by BAUW: http://bauw.psychstat.org
#USE WITH CAUTION!

Model{
  # Model specification for linear growth curve model
  for (i in 1:N){
    LS[i,1:2]~dmnorm(muLS[i,1:2], Inv_cov[1:2,1:2])
    muLS[i,1]<-bL[1]
    muLS[i,2]<-bS[1]
    for (t in 1:5){
      y[i, t] ~ dnorm(muY[i,t], Inv_Sig_e2)
      muY[i,t]<-LS[i,1]+LS[i,2]*t

      ## missing data mechanism
      m[i,t]~dbern(p[i,t])
      p[i,t]<-phi(gam[t,1]+gam[t,2]*LS[i,1]+gam[t,3]*LS[i,2])
    }
  }

  #Priors for model parameter
  for (i in 1:1){
    bL[i] ~ dnorm(0, 1.0E-6)
    bS[i] ~ dnorm(0, 1.0E-6)
  }
  for (i in 1:3){
     for (t in 1:5){
         gam[t,i]~dnorm(0, .001)
     }
   }
  Inv_cov[1:2,1:2]~dwish(R[1:2,1:2], 2)
  R[1,1]<-1
  R[2,2]<-1
  R[2,1]<-R[1,2]
  R[1,2]<-0
  Inv_Sig_e2 ~ dgamma(.001, .001)
  Sig_e2 <- 1/Inv_Sig_e2
  Cov[1:2,1:2]<-inverse(Inv_cov[1:2,1:2])
  Sig_L2 <- Cov[1,1]
  Sig_S2 <- Cov[2,2]
  Cov_LS <- Cov[1,2]
```

```
  rho_LS <- Cov[1,2]/sqrt(Cov[1,1]*Cov[2,2])
}

# The (naive) starting values for model parameters.
list(bL=c(50), bS = c(1), Inv_Sig_e2 = 1,
Inv_cov= structure(.Data = c(.01,0,0,.04),.Dim=c(2,2)))

# Please put your data here. Remeber you can convert your data by BAUW--Convert Dat
list(N=500, y = structure(.Data = c( 52.04084, NA, NA, NA, NA,  ...),
.Dim=c( 500, 5 )) , m = structure(.Data = c( 1, 0, 0, 0, ...), .Dim=c( 500, 5 )) )
```