Robust Bayesian growth curve modeling using Student's t distribution

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

If y_{it} is not normally distributed, for example, with a long tail or outliers, the normality assumption may not hold any more. A possible solution is to use the Student's t distribution such that

$$e_{it} \sim t(k, 0, \phi)$$

where k is the degree of freedom and ϕ can be viewed as variance.

Likelihood function

Conditonally on $\boldsymbol{\eta}_i$, the distribution of $\mathbf{y}_i$ is a multivariate student distribtion with density function

$$p(\mathbf{y}_i|\boldsymbol{\eta}_i, k, \phi, \mathbf{\Lambda}, \beta) = \frac{\Gamma(\frac{k+T}{2})}{\Gamma(\frac{k}{2})k^{T/2}\pi^{T/2}} |\mathbf{I}\phi|^{-1/2} \left[1 + \frac{1}{k}(\mathbf{y}_i - \mathbf{\Lambda}\boldsymbol{\eta}_i)'(\mathbf{I}\phi)^{-1}(\mathbf{y}_i - \mathbf{\Lambda}\boldsymbol{\eta}_i)\right]^{-(k+T)/2}.$$

Thus, the joint distribution of $\mathbf{y}_i$ and $\boldsymbol{\eta}_i$ is

$$\begin{aligned}
p(\mathbf{y}_i, \boldsymbol{\eta}_i|k, \phi, \mathbf{\Lambda}, \mathbf{\Psi}, \beta) &= p(\boldsymbol{\eta}_i|\mathbf{\Psi}, \beta)p(\mathbf{y}_i|\boldsymbol{\eta}_i, k, \phi, \mathbf{\Lambda}) \\
&= (2\pi)^{-q/2}|\mathbf{\Psi}|^{-1/2} \exp\left[-\frac{1}{2}(\boldsymbol{\eta}_i - \boldsymbol{\beta})'\mathbf{\Psi}^{-1}(\boldsymbol{\eta}_i - \boldsymbol{\beta})\right] \\
&\times \frac{\Gamma(\frac{k+T}{2})}{\Gamma(\frac{k}{2})k^{T/2}\pi^{T/2}} |\mathbf{I}\phi|^{-1/2} \left[1 + \frac{1}{k}(\mathbf{y}_i - \mathbf{\Lambda}\boldsymbol{\eta}_i)'(\mathbf{I}\phi)^{-1}(\mathbf{y}_i - \mathbf{\Lambda}\boldsymbol{\eta}_i)\right]^{-(k+T)/2}.
\end{aligned}$$

Thus, the likelihood function for the model is

$$\begin{aligned}
L &\propto \prod_{i=1}^{n} p(\mathbf{y}_i, \boldsymbol{\eta}_i | k, \phi, \boldsymbol{\Lambda}, \boldsymbol{\Psi}, \boldsymbol{\beta}) \\
&\propto |\boldsymbol{\Psi}|^{-n/2} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{\eta}_i - \boldsymbol{\beta})'\boldsymbol{\Psi}^{-1}(\boldsymbol{\eta}_i - \boldsymbol{\beta})\right] \\
&\times \left[\frac{\Gamma(\frac{k+T}{2})}{\Gamma(\frac{k}{2})}\right]^n k^{-\frac{nT}{2}} \phi^{-\frac{nT}{2}} \left\{\prod_{i=1}^{n}\left[1 + \frac{1}{k}\phi^{-1}(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i)'(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i)\right]\right\}^{-(k+T)/2}.
\end{aligned}$$

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

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

For k, we first consider an exponential distribution prior,

$$p(k) = \mathrm{Exp}(\lambda) = \lambda \exp(-\lambda k).$$

Posteriors

The joint posterior distribution for $(\boldsymbol{\beta}, \boldsymbol{\Psi}, \phi, k)$ is the product of the likelihood function and the priors

$$p(\boldsymbol{\beta}, \boldsymbol{\Psi}, \phi, k | \mathbf{Y}, \boldsymbol{\eta}) = L \times p(\boldsymbol{\beta}) \times p(\boldsymbol{\Psi}) \times p(\phi) \times p(k).$$

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 kernal for the conditional posterior distribution of k is

$$p(k|\phi, \eta_i, \mathbf{y}_i, i = 1, \ldots, n) \propto \left[\frac{\Gamma(\frac{k+T}{2})}{\Gamma(\frac{k}{2})}\right]^n k^{-\frac{nT}{2}} \left[\exp(-\lambda k)\right] \left\{\prod_{i=1}^{n} \left[1 + \frac{1}{k}\phi^{-1}(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i)'(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i)\right]\right\}^{-(k+T)/2}.$$

The kernal for the conditional posterior distribution of ϕ is

$$\begin{aligned} p(\phi|k, \eta_i, \mathbf{y}_i, i = 1, \ldots, n) &= \phi^{-(\alpha_0+1)-nT/2} \exp\left(-\frac{\gamma_0}{\phi}\right) \\ &\times \left\{\prod_{i=1}^{n} \left[1 + \frac{1}{k}\phi^{-1}(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i)'(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i)\right]\right\}^{-(k+T)/2}. \end{aligned}$$

Finally, the conditional posterior distribution for $\boldsymbol{\eta}_i$ is

$$\begin{aligned} p(\boldsymbol{\eta}_i|\phi, \boldsymbol{\Psi}, \boldsymbol{\beta}, k, \mathbf{y}_i) &\propto \exp\left[-\frac{1}{2}(\boldsymbol{\eta}_i - \boldsymbol{\beta})'\boldsymbol{\Psi}^{-1}(\boldsymbol{\eta}_i - \boldsymbol{\beta})\right] \\ &\times \left[1 + \frac{1}{k}(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i)'(\mathbf{I}\phi)^{-1}(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i)\right]^{-(k+T)/2}. \end{aligned}$$

With the full set of conditional posterior distributions, Gibbs sampling can be used to generate Markov chains for model parameters and latent variables. Because the conditional posterior distributions for k, ϕ, and $\boldsymbol{\eta}_i$ do not follow an existing distribution, the Metropolis-Hastings sampling method is used within each itereation of Gibbs sampling. Specifically, in using M-H algorithm, the exponential proposal distribution, the inverse gamma proposal distribution, and the multivariate normal proposal distribution are used for k, ϕ, and $\boldsymbol{\eta}_i$, respectively.

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 $e_{it} \sim N(0, \phi)$ or $e_{it} \sim t(k, 0, \phi)$ and $Cov\left[(v_{1i}, v_{2i})'\right] = \boldsymbol{\Psi} = \begin{pmatrix} \sigma_L^2 & \sigma_{LS} \\ \sigma_{LS} & \sigma_S^2 \end{pmatrix}$. We set the population parameters at $\beta_1 = 50$, $\beta_2 = 5$, $\phi = 25$, and

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

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 degrees of freedom will be set at 3, 10, and 50.

The following simulation studies should be conducted.

1. With generated student's t data, whether the parameters can be recovered?
2. What the influence of analyzing student's t data as normal data?
3. What the influence of analyzing normal data as student's t data?

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] ~ dt(muY[i,t], Inv_Sig_e2, k)
      muY[i,t]<-LS[i,1]+LS[i,2]*t
    }
  }

  k~dunif(3, 30)

  #Priors for model parameter
  for (i in 1:1){
    bL[i] ~ dnorm(0, 1.0E-6)
    bS[i] ~ dnorm(0, 1.0E-6)
  }
  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])
}
```