Computational Statistics

Generalized Linear Mixed Model

Chenyin Gao

Dept. Statistics, Sun Yat-sen University

Contents

1 Progect Summary

1.1 Model Notation

In this project, we consider a clustering problem. Suppose we have observed n observations, each observation is a binary process, i.e. the response $Y_{ij} = 0$ or 1, $i = 1, \cdots, n$, $j = 1, \cdots, T$. Here n is the number of subjects and T is the length of observation. In general, T might vary across subjects, time points may also be different. In this project, however, we simply assume that all subjects have common time length and time points. We also assume that these subjects belong to two clusters. For each cluster, the conditional expectation of response variable is:

$$\begin{aligned} P_{ij} &= \mathbb{E}(Y_{ij}|U_i = 1, X_{1,ij}, Z_{1,i}) = g^{-1}(\beta_1 X_{1,ij} + Z_{1,i}) \\ P_{ij} &= \mathbb{E}(Y_{ij}|U_i = 2, X_{2,ij}, Z_{2,i}) = g^{-1}(\beta_2 X_{2,ij} + Z_{2,i}) \end{aligned} \tag{1}$$

where U is cluster membership, $X_{c,ij}$ and $Z_c, i(c = 1,2)$ are fixed and random effects, respectively. The link function $g^{-1}(x) = \frac{\exp(x)}{1+\exp(x)}$ is given. In a typical clustering problem, U is usually unknown, and hence we treat U as another random effect.

For random effects, we assume that $Z_{c,i} \sim N(0, \sigma_c^2)$ and $\mathbb{P}(U = 1) = \pi_1$ (then $\pi_2 = 1 - \pi_1$). Then the parameter to be estimated is $\Omega = \{\beta_1, \beta_2, \sigma_1, \sigma_2, \pi_1\}$. Treating random effects as missing data, one can write the complete data likelihood function as

$$L(\Omega|Y_{ij}, U_i, Z_{U_i,i}) = \prod_{i=1}^{n}\prod_{c=1}^{2}\{\pi_c f_c(Z_{c,i})[\prod_{j=1}^{T} f_c(Y_{ij}|Z_{c,i})]\}^{w_{ic}} \tag{2}$$

where $f_c(Z_{c,i})$ is the density function of Normal distribution, $f_c(Y_{ij}|Z_{c,i}) = \mathbb{P}^{Y_{ij}}(1-\mathbb{P}_{ij})^{1-Y_{ij}}$. w_{ic} is the dummy variable of U_i , i.e.

$$w_{ic} = \begin{cases} 1 & , \quad if\ subject \text{ i } belongs\ to\ cluster\ c \\ 0 & , \quad otherwise \end{cases}$$

1.2 Simulation Setup and Requirement

Generate 100 simulations. In each simulation, set n = 100 and T = 10. The true values of parameter are: $\beta_1 = 1, \beta_2 = 1, \pi_1 = 0.6, \sigma_1 = 2$ and $\sigma_2 = 10$

Use N(0,1) to generate the fixed effect X, and use them for all 100 simulations and use MCEM to evaluate the loglikelihood function. In the E-step, perform K = 500 Gibbs sampling incorporated with a Metropolis-Hastings step, and drop the first 100 as a burn-in procedure.

2 Generalized Linear Mixed Model(GLMM)

Given the simulation parameters: $n = 100$,$T = 10$,$\beta_1 = \beta_2 = 1$,$\pi_1 = 0.6$, $\sigma_1 = 2$,$\sigma_2 = 10$, we could obtain,

Observed variables

$$\mathbf{Y} = \begin{bmatrix} Y_{11} & Y_{12} & \cdots Y_{1T} & \\ Y_{21} & Y_{22} & \cdots Y_{2T} & \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n1} & Y_{n2} & \cdots Y_{nT}) & \end{bmatrix} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_n \end{bmatrix}$$

Additional unobserved or unobservable variables

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \\ \vdots \\ \mathbf{U}_n \end{bmatrix}, \mathbf{Z} = \begin{bmatrix} \mathbf{Z}_{U_1,1} \\ \mathbf{Z}_{U_2,2} \\ \vdots \\ \mathbf{Z}_{U_n,n} \end{bmatrix}$$

Explanatory variables(fixed effect)

$$X = \begin{vmatrix} X_{U_1,11} & X_{U_1,12} & \cdots & X_{U_1,1T} \\ X_{U_2,21} & X_{U_2,22} & \cdots & X_{U_2,2T} \\ \vdots & \vdots & \ddots & \vdots \\ X_{U_n,n1} & X_{U_n,n2} & \cdots & X_{U_n,nT} \end{vmatrix}$$

2.1 Complete Log-Likelihood

Given the necesssary parameters for each component of Ω, we could write the augmented logged liklihood as

$$\begin{aligned}
l(\Omega|\mathbf{Y},\mathbf{U},\mathbf{Z}) &= \ln L(\Omega|\mathbf{Y},\mathbf{U},\mathbf{Z}) \\
&= \ln \prod_{i=1}^{n}\prod_{c=1}^{2}\{\pi_c f_c(Z_{c,i})[\prod_{j=1}^{T} f_c(Y_{ij}|Z_{c,i})]\}^{w_{ic}} \\
&= \sum_{i=1}^{n}\sum_{c=1}^{2} w_{ic}[\ln \pi_c + \ln f_c(Z_{c,i}) + \sum_{j=1}^{T} \ln f_c(Y_{ij}|Z_{c,i})] \\
&= \sum_{i=1}^{n}\sum_{c=1}^{2} w_{ic}\{\ln \pi_c - \frac{Z_{c,i}^2}{2\sigma_c^2} - \frac{1}{2}\ln(2\pi\sigma_c^2) + \sum_{j=1}^{T}[Y_{ij}\ln P_{ij} + (1-Y_{ij})\ln(1-P_{ij})]\} \\
&= n\sum_{c=1}^{2} w_{ic}[\ln \pi_c - \frac{1}{2}\ln(2\pi\sigma_c^2)] - \sum_{i=1}^{n}\sum_{c=1}^{2} w_c \frac{Z_{c,i}^2}{2\sigma_c^2} \\
&+ \sum_{i=1}^{n}\sum_{c=1}^{2} w_{ic}\sum_{j=1}^{T}[Y_{ij}(\beta_c X_{c,ij} + Z_{c,i}) - Y_{ij}\ln(1+\exp(\beta_c X_{c,ij} + Z_{c,i})) - (1-Y_{ij})\ln(1+\exp(\beta_c X_{c,ij} + Z_{c,i}))] \\
&= n\sum_{c=1}^{2} w_{ic}[\ln \pi_c - \frac{1}{2}\ln(2\pi\sigma_c^2)] \\
&+ \sum_{i=1}^{n}\sum_{c=1}^{2} w_c\{-\frac{Z_{c,i}^2}{2\sigma_c^2} + \sum_{j=1}^{T}[Y_{ij}(\beta_c X_{c,ij} + Z_{c,i}) - \ln(1+\exp(\beta_c X_{c,ij} + Z_{c,i}))]\}
\end{aligned}$$

which is equal to

$$\begin{aligned}
l(\Omega|\mathbf{Y},\mathbf{U},\mathbf{Z}) &= \sum_{i=1}^{n} \ln f_{(U_i,Z_{U_i,i})}(U_i, Z_{U_i,i}|\pi_c,\sigma_1,\sigma_2) + \sum_{i=1}^{n}\sum_{j=1}^{T} \ln f_{Y_{ij}|(U_i,Z_{U_i,i})}(Y_{ij}|(U_i,Z_{U_i,i}),\beta_1,\beta_2) \\
&= \triangleq \ln f_{(\mathbf{U},\mathbf{Z})}(\mathbf{U},\mathbf{Z}|\pi_c,\sigma_1,\sigma_2) + \ln f_{\mathbf{Y}|(\mathbf{U},\mathbf{Z})}(\mathbf{Y}|\mathbf{U},\mathbf{Z},\beta_1,\beta_2)
\end{aligned}$$

3 Monte Carlo Expectation Maximization

3.1 EM Algorithm

By taking expectation of **U** and **Z** given **Y** under the current estimate of the parameters $\Omega^{(t)}$, we could write the expected augmented logged likelihood as

$$
\begin{aligned}
Q(\Omega,\Omega^{(t)}) &= \mathbb{E}_{(\mathbf{U},\mathbf{Z})|(\mathbf{Y},\Omega^{(t)})} \ln L(\Omega|\mathbf{Y},\mathbf{U},\mathbf{Z}) \\
&= \mathbb{E}_{(\mathbf{U},\mathbf{Z})|(\mathbf{Y},\Omega^{(t)})} \\
&= \sum_{i=1} \sum_{c=1}^{2} \mathbb{E}_{\mathbf{U}|(\mathbf{Y},\Omega^{(t)})}(w_{ic})[\ln \pi_c - \frac{1}{2}\ln(2\pi\sigma_c^2)] \\
&+ \sum_{i=1}^{n} \sum_{c=1}^{2} \mathbb{E}_{(\mathbf{U},\mathbf{Z})|(\mathbf{Y},\Omega^{(t)})} w_c\{-\frac{Z_{c,i}^2}{2\sigma_c^2} + \sum_{j=1}^{T}[Y_{ij}(\beta_c X_{c,ij}+Z_{c,i}) - \ln(1+\exp(\beta_c X_{c,ij}+Z_{c,i}))]\}
\end{aligned}
$$

Notice that in the expected log-likelihood, $\Omega^{(t)}$ could be decomposed into seperate component

$$
\begin{aligned}
Q(\Omega,\Omega^{(t)}) &= \mathbb{E}_{(\mathbf{U},\mathbf{Z})|(\mathbf{Y},\Omega^{(t)})} \ln f_{(\mathbf{U},\mathbf{Z})}(\mathbf{U},\mathbf{Z}|\pi_c,\sigma_1,\sigma_2) + \mathbb{E}_{(\mathbf{U},\mathbf{Z})|(\mathbf{Y},\Omega^{(t)})} \ln f_{\mathbf{Y}|(\mathbf{U},\mathbf{Z})}(\mathbf{Y}|\mathbf{U},\mathbf{Z},\beta_1,\beta_2) \\
&=\triangleq P(\Omega,\Omega^{(t)}) + R(\Omega,\Omega^{(t)})
\end{aligned}
$$

3.2 Monte Carlo Integrating

In order to compute the integral above, we use Monte Carlo Integrating to approximate it. Suppose that $\{(\mathbf{U}_{(k)},\mathbf{Z}_{(k)}, k=1,2,\cdots,K)\} \overset{i.i.d}{\sim} f_{(\mathbf{U},\mathbf{Z}|\mathbf{Y})}(\mathbf{U},\mathbf{Z}|\mathbf{Y}),\Omega)$ and we sample m times to approximate.

Based on Mean Value Method

$$
Q(\Omega,\Omega^{(t)}) \approx \frac{1}{m}\sum_{k=1}^{m} \sum_{i=1,c=U_{(k),i}} \left[\ln \pi_c - \frac{1}{2}\ln(2\pi\sigma_c^2) - \frac{Z_{c,i}^2}{2\sigma_c^2} + \sum_{j=1}^{T}[Y_{ij}(\beta_c X_{c,ij}+Z_{c,i}) - \ln(1+\exp(\beta_c X_{c,ij}+Z_{c,i}))\right] \tag{3}
$$

3.3 MLE

The partial derivatives of the parameters are given by

$$
\begin{aligned}
\frac{\partial Q(\Omega,\Omega^{(t)})}{\partial \pi_1} &= \frac{1}{m}\sum_{k=1}^{m}\sum_{i=1}^{n} \mathbb{I}_{\{U_{(k),i},i=1\}} \frac{1}{\pi_1} - \frac{1}{m}\sum_{k=1}^{m}\sum_{i=1}^{n} \mathbb{I}_{\{U_{(k),i},i=2\}} \frac{1}{1-\pi_1} \\
\frac{\partial Q(\Omega,\Omega^{(t)})}{\partial \sigma_c^2} &= \frac{1}{m}\sum_{k=1}^{m}\sum_{i=1}^{n} \mathbb{I}_{\{U_{(k),i=c}\}} \left(-\frac{1}{2\sigma_c^2} + \frac{Z_{(k),c,i}^2}{2\sigma_c^4}\right) \\
\frac{\partial Q(\Omega,\Omega^{(t)})}{\partial \beta_c} &= \frac{1}{m}\sum_{k=1}^{m}\sum_{i=1}^{n} \mathbb{I}_{\{U_{(k),i=c}\}} \sum_{j=1}^{T}\left[Y_i j X_{c,ij} - \frac{X_{c,ij}\exp(\beta_c X_{c,ij}+Z_{(k),c,i})}{1+\exp(\beta_c X_{c,ij}+Z_{(k),c,i})}\right]
\end{aligned}
$$

By setting the above partial derivative to 0, we get the maximum likelihood estimators

$$\begin{aligned}\hat{\pi}_1 &= \frac{1}{mn}\sum_{k=1}^{m}\sum_{i=1}^{n}\mathbb{I}_{\{U_{(k)},i=1\}} \\ \hat{\sigma}_c &= \sqrt{\frac{\sum_{k=1}^{m}\sum_{i=1}^{n}\mathbb{I}_{\{U_{(k)},i=c\}}Z^2_{(k),c,i}}{\sum_{k=1}^{m}\sum_{i=1}^{n}\mathbb{I}_{\{U_{(k),i=c}\}}}}\end{aligned} \tag{4}$$

To compute the MLE of β_c, we use direct numerical maximization proposed by Newton-Raphson Method. The second order partial derivative of β_c is denoted as

$$\frac{\partial^2 Q(\Omega,\Omega^{(t)})}{\partial\beta_c^2} = -\frac{1}{m}\sum_{k=1}^{m}\sum_{i=1}^{n}\mathbb{I}_{\{U_{(k),i=c}\}}\sum_{j=1}^{T}\frac{X^2_{c,ij}\exp(\beta_c X_{c,ij}+Z_{(k),c,i})}{(1+\exp(\beta_c X_{c,ij}+Z_{(k),c,i}))^2} \tag{5}$$

Algorithm 1 Newton-Raphson Method

1: Initialize $\hat{\beta}_c^{(0)}$

2: t$\leftarrow$ 0

3: $\hat{\beta}_c^{(t+1)} \leftarrow \hat{\beta}_c^{(t)} - \frac{\frac{\partial Q(\Omega,\Omega^{(t)})}{\partial\beta_c}|_{\hat{\beta}_c^{(t)}}}{\frac{\partial^2 Q(\Omega,\Omega^{(t)})}{\partial\beta_c^2}|_{\hat{\beta}_c^{(t)}}}$

4: Repeat step 2-3 until convergence

4 Markov Chain Sampler

Since it difficult to sample directly form multivariate distribution $f_{(\mathbf{U},\mathbf{Z}|\mathbf{Y})}(\mathbf{U},\mathbf{Z}|\mathbf{Y}),\Omega)$.We can use Gibbs Sampling, a Markov chain Monte Carlo (MCMC) algorithm to obtain a sequence of observations which are approximated from the multivariate distribution.

First, we need to calculate the conditional distributions

$$\frac{f_{(U_i,Z_{(U_i,i)}|\mathbf{Y}_i)}(U_i,Z_{U_i,i}|\mathbf{Y}_i,\Omega)}{f_{Z_{(U_i,i)}|\mathbf{Y}_i}(Z_{U_i,i}|\mathbf{Y}_i,\Omega)} = f_{U_i|(Z_{U_i},i,\mathbf{Y}_i)}(U_i|Z_{U_i,i},\mathbf{Y}_i) \tag{6}$$

and

$$\frac{f_{(U_i,Z_{(U_i,i)}|\mathbf{Y}_i)}(U_i,Z_{U_i,i}|\mathbf{Y}_i,\Omega)}{f_{U_i|\mathbf{Y}_i}(U_i|\mathbf{Y}_i,\Omega)} = f_{Z_{U_i,i}|(U_i,\mathbf{Y}_i)}(Z_{U_i,i}|(U_i,\mathbf{Y}_i)) \tag{7}$$

Then, suppose that $(U_{(k),i},Z_{(k),U_{(k),i},i})$ is the ith component of the kth sample, we want to draw the ith component of the $(k+1)$th sample. We draw

$$\begin{aligned}U_{(k+1),i} &\sim f_{U_i|Z_{U_i,i},\mathbf{Y}_i}(u|Z_{U_i,i},\mathbf{Y}_i,\Omega) \\ Z_{(k+1),U_{(k+1),i},i} &\sim f_{Z_{U_i,i}|U_i,\mathbf{Y}_i}(z|U_i,\mathbf{Y}_i,\Omega)\end{aligned}$$

4.1 Metropolis-Hastings Algorithm

To sample $Z_{(k+1),U_{(k+1),i},i}$ from $f_{Z_{U_i,i}|U_i,\mathbf{Y}_i}(z|U_i,\mathbf{Y}_i,\Omega)$, let $h_{Z_{U_{(k),i},i}}(z)$ be the candidate distribution. Since the candidate distribution should be similar to $f_{Z_{U_i,i}|U_i,\mathbf{Y}_i}(z|U_i,\mathbf{Y}_i,\Omega)$, we can choose $h_{Z_{U_{(k),i},i}}(z) = f_{U_i}(z|\Omega)$ and the acceptance function is

$$A_{k,\mathbf{Y}_i}(z,z^*) = \min\left[1, \frac{f_{Z_{U_i,i}|U_i,\mathbf{Y}_i}(z^*|U_i,\mathbf{Y}_i,\Omega)f_{U_i}(z|\Omega)}{f_{Z_{U_i,i}|U_i,\mathbf{Y}_i}(z|U_i,\mathbf{Y}_i,\Omega)f_{U_i}(z^*|\Omega)}\right]$$

where $\frac{f_{Z_{U_i,i}|U_i,\mathbf{Y}_i}(z^*|U_i,\mathbf{Y}_i,\Omega)f_{U_i}(z|\Omega)}{f_{Z_{U_i,i}|U_i,\mathbf{Y}_i}(z|U_i,\mathbf{Y}_i,\Omega)f_{U_i}(z^*|\Omega)}$ can be written as,

$$\frac{f_{Z_{U_i,i}|U_i,\mathbf{Y}_i}(z^*|U_i,\mathbf{Y}_i,\Omega)f_{U_i}(z|\Omega)}{f_{Z_{U_i,i}|U_i,\mathbf{Y}_i}(z|U_i,\mathbf{Y}_i,\Omega)f_{U_i}(z^*|\Omega)} = \exp\left[\sum_{j=1}^{T} Y_{ij}(z^*-z)\right]\prod_{j=1}^{T}\frac{1+\exp(\beta_i X_{ij}+z)}{1+\exp(\beta_i X_{ij}+z^*)}$$

We begin our Gibbs sampler incorporated a Metropolis-Hastings step as follow

Algorithm 2 MCMC incorporated Metropolis-Hastings

```
for i=1:n do
    Initialize(U_(0),i, Z_(0),1,i, Z_(0),2,i)
    for c do=1:2
        k ← 0
        for k do=1:K_2
            Draw z* ~ f_c(z|Ω)
            Accept z* as Z_(k+1),c,i with probability A_k,Y_i(z, z*); otherwise, retain the original Z_(k),c,i
        Burn-in procedure and let the last K+1 samples be the final samples {Z_(k),c,i, k = 0,1,···,K}
    k ← 0
    for k do=1:K Draw U_(k+1),i ~ f_{U_i|Z_{U_i},i,Y_i}(u|Z_{U_i},i,Y_i,Ω)
    Let the last m samples be the final samples {U_(k),i, k = 0,1,···,K}
Burn-in procedure and return the m samples {(U_(i),i, Z_(i),1,i, Z_(i),2,i), k = 0,1,···,m}
```

4.2 MCEM

Unfortunately, we don't known $f_{(\mathbf{U},\mathbf{Z})|\mathbf{Y}}(\mathbf{U},\mathbf{Z}|\mathbf{Y},\Omega)$, so we use $f_{(\mathbf{U},\mathbf{Z})|\mathbf{Y}}(\mathbf{U},\mathbf{Z}|\mathbf{Y},\Omega^{(t)})$ in the $(t+1)$th step to estimate the distribution. To generate $\{(U_{(k)},Z_{(k)}),k=1,2,\cdots,m\} \overset{i.i.d}{\sim} f_{(\mathbf{U},\mathbf{Z})|\mathbf{Y}}(\mathbf{U},\mathbf{Z}|\mathbf{Y},\Omega^{(t)})$.

The Monte Carlo Expectation-Maximization Algorithm we use in every stimulation is given by

Algorithm 3 MCEM

Start with the initial value $\Omega^{(0)}$. Set t=0.

E-STEP:

a. Generate m samples $\{(U_{(i),i}, Z_{(i),1,i}, Z_{(i),2,i}), k = 0, 1, \cdots, m\}$ form $f_{Z_{U_i,i}|U_i,\mathbf{Y}_i}(z|U_i, \mathbf{Y}_i, \Omega)$ through 2

b.Calculate the partial derivatives of $Q(\Omega, \Omega^{(t)})$, the Monte Carlo estimator for every parameters.

M-STEP

$\Omega^{(t+1)} \leftarrow \arg\max_{\Omega} Q(\Omega, \Omega^{(t)})$

$t \leftarrow t+1$

Repeat step 2-6 until convergence and then output the maximum likelihood estimators $\Omega^{(t)}$.

5 Simulation

For $1Y_{ij} = 0$ or 1, $i = 1, \cdots, n, j = 1, \cdots, T$, we begin our simulations as $n = 100$, $T = 10$, $\beta_1 = \beta_2 = 1, \sigma_1 = 5, \sigma_2 = 10, \pi = 0.6$

1. we set the initial value as $\beta_1 = \beta_2 = 0, \sigma_1 = 1, \sigma_2 = 5, \pi = 0.8$.
2. Perform variable step-size[1] Gibbs sampling in each EM iteration
3. Repeat 50 times EM iterate in each experiment
4. Carry out 1000 experiments under different random seeds and display our first 100 experiments results and 1000 experiments MSE

Figure 1. Convergence of β_1

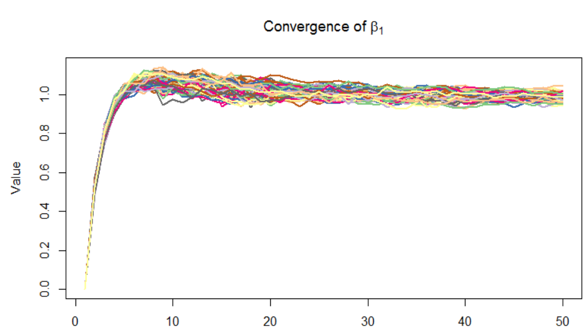
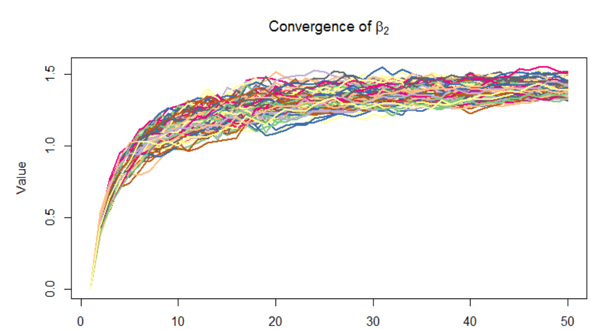


Figure 2. Convergence of β_2



[1]We begin our Monte Carlo Estimation with small sample, as the EM iterate proceeding, we increase our Monte Carlo sampling times accordingly

Figure 3. Convergence of σ_1

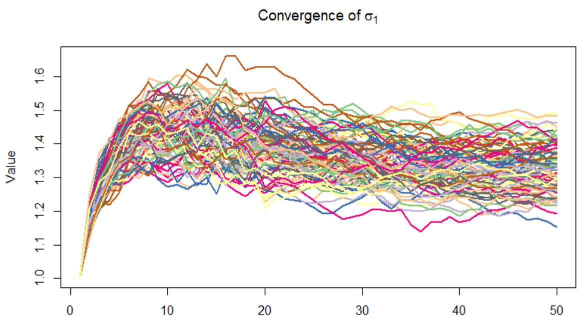


Figure 4. Convergence of σ_2

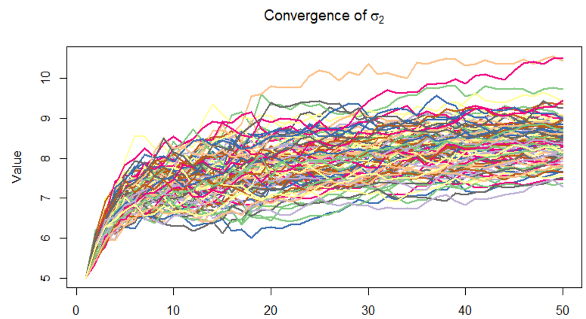


Figure 5. Convergence of π_1

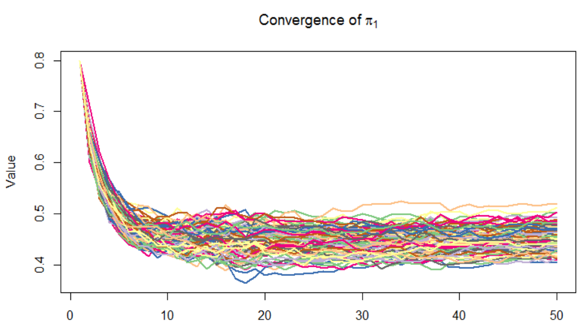


Figure 6. MSE Timeline

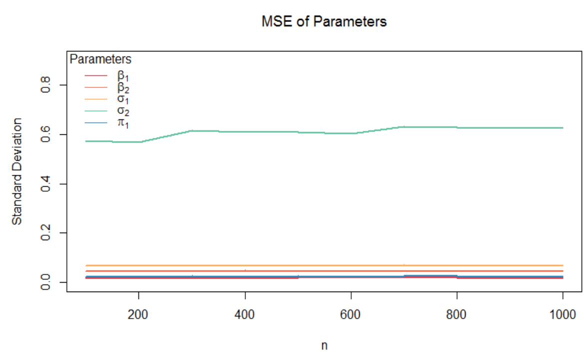


Figure 1, Figure 2, Figure 3, Figure 4 and Figure 5 show the convergence trace for all 100 experiments, the convergence results is displayed in Table 1. After convergence the value continues to show random variation around a line. This line is the approximate MLE estimated by EM algorithm(Figure 6).

Table 1. Compare real and estimated coefficients

	β_1	β_2	σ_1	σ_2	π_1
Real	1.000	1.000	5.000	10.000	0.600
Estimated	0.990	1.404	1.327	8.479	0.455

Also, we can see from Table 5 that the MSE of $\hat{\sigma}_2$ is much bigger than other parameters. This may be the result of the difference of the magnitudes.

Table 2. Simulations for $\beta_1(0) = \beta_2(0) = 0, \sigma_1(0) = 1, \sigma_2(0) = 5, \pi_1(0) = 0.8$, N = 100

	β_1	β_2	σ_1	σ_2	π_1
100	0.019	0.045	0.068	0.572	0.023
200	0.018	0.044	0.068	0.567	0.023
300	0.018	0.046	0.069	0.616	0.025
400	0.018	0.047	0.068	0.609	0.024
500	0.019	0.046	0.069	0.610	0.025
600	0.018	0.046	0.068	0.603	0.025
700	0.019	0.045	0.070	0.633	0.025
800	0.019	0.045	0.068	0.627	0.025
900	0.018	0.045	0.067	0.624	0.025
1000	0.018	0.045	0.067	0.623	0.024

The MSE is calculated as following

$$MSE_\theta = \frac{1}{N} \sum_{n=1}^{N} (\theta^{(n)} - \hat{\theta}^{(n)})^2 \tag{8}$$

where $\theta \in \Omega, \theta^{(n)}$ is the true MLE of θ in the nth experiments and $\hat{\theta}^{(n)}$ is the estimator of $\theta^{(n)}$.