PDA 4010 Assignment 6 赵[illegible] 120090346

Problem 1:

(a) $Y_i \sim \text{Bernoulli}\left(\frac{\exp(\beta_0+\beta^T x_i)}{1+\exp(\beta_0+\beta^T x_i)}\right)$

① $P(\beta_0 \mid \beta, \gamma, y, X) \propto P(y \mid \beta_0, \beta, \gamma, y, X)\, P(\beta_0)$

$$= \left(\prod_{i=1}^{n_1} \frac{e^{\beta_0+(\beta\gamma)^T x_i}}{1+e^{\beta_0+(\beta\gamma)^T x_i}} \prod_{j=1}^{n_2} \frac{1}{1+e^{\beta_0+(\beta\gamma)^T x_j}}\right) \text{dnorm}(\beta_0; 0, 16)$$

$$\log(r_1) = \frac{\log[P(\beta_0^* \mid \beta, \gamma, y, X)]}{\log[P(\beta_0 \mid \beta, \gamma, y, X)]} = \frac{\left[\sum_{i=1}^{n_1}\log\left(\frac{e^{\beta_0^*+(\beta\gamma)^T x_i}}{1+e^{\beta_0^*+(\beta\gamma)^T x_i}}\right) + \sum_{j=1}^{n_2}\log\left(\frac{1}{1+e^{\beta_0^*+(\beta\gamma)^T x_j}}\right)\right] + \log(\text{dnorm}(\beta_0^*; 0, 16))}{\left[\sum_{i=1}^{n}\log\left(\frac{e^{\beta_0+(\beta\gamma)^T x_i}}{1+e^{\beta_0+(\beta\gamma)^T x_i}}\right) + \sum_{j=1}^{n_2}\log\left(\frac{1}{1+e^{\beta_0+(\beta\gamma)^T x_j}}\right)\right] + \log(\text{dnorm}(\beta_0; 0, 16))}$$

② if $\gamma_j = 1$:

$P(\beta_j \mid \beta_0, \gamma, y, X) \propto P(y \mid \beta_0, \beta, \gamma, y, X)\, P(\beta_j)$

$$= \left(\prod_{i=1}^{n_1} \frac{e^{\beta_0+(\beta\gamma)^T x_i}}{1+e^{\beta_0+(\beta\gamma)^T x_i}} \prod_{j=1}^{n_2} \frac{1}{1+e^{\beta_0+(\beta\gamma)^T x_j}}\right) \prod_{j=1}^{5} \text{dnorm}(\beta_j; 0, 4)$$

if $\gamma_j = 0$, $P(\beta_j \mid \beta_0, \gamma, y, X) = P(\beta_j)$

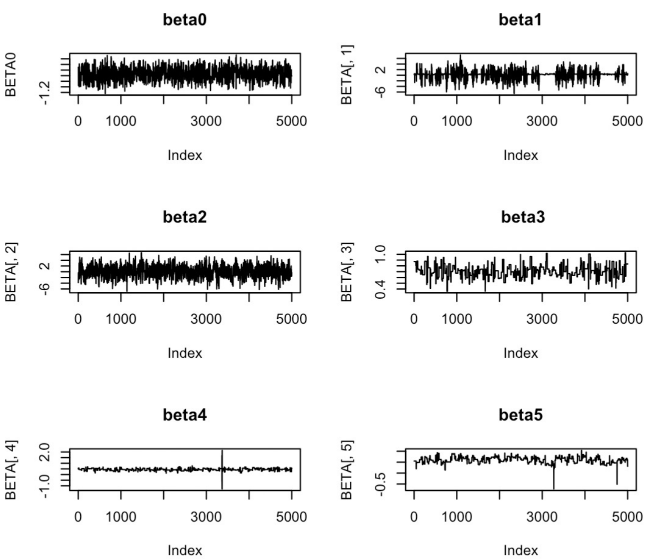
③ $$P(\gamma_j = 1 \mid \gamma_{-j}, \beta_0, \beta, y, X) = \frac{P(y \mid \gamma_j = 1, \gamma_{-j}, \beta_0, \beta, X)\, P(\gamma_j = 1)}{P(y \mid \gamma_j = 1, \gamma_{-j}, \beta_0, \beta, X)\, P(\gamma_j = 1) + P(y \mid \gamma_j = 0, \gamma_{-j}, \beta_0, \beta, X)\, P(\gamma_j = 0)}$$

$$= \frac{a_j}{a_j + b_j}$$

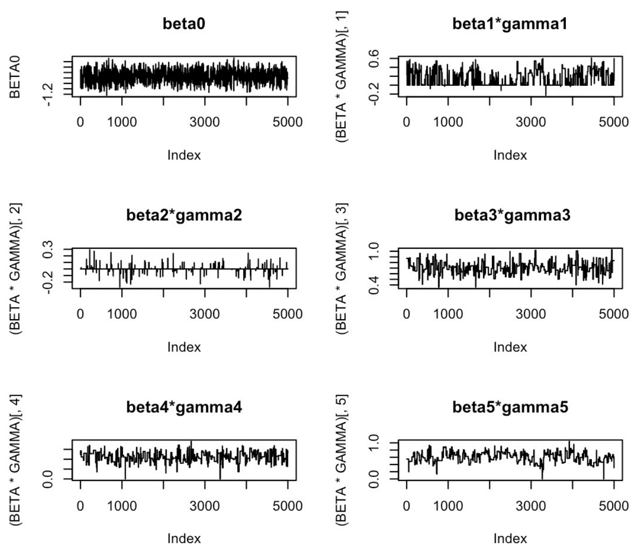
$\therefore\ \gamma_j \sim \text{Bernoulli}\left(\frac{a_j}{a_j+b_j}\right)$

$$r = \frac{P(y \mid \gamma_j = 1, \gamma_{-j}, \beta_0, \beta, X)\, P(\gamma_j = 1)}{P(y \mid \gamma_j = 0, \gamma_{-j}, \beta_0, \beta, X)\, P(\gamma_j = 0)}$$

trace plot of β_0, β



trace plot of $\beta_0, \beta \cdot \gamma$



effective sample size:

```
> apply(BETA, 2, effectiveSize)
[1] 707.32124 611.79844 221.01823 341.66034  93.47357
> apply(BETA*GAMMA, 2, effectiveSize)
[1]   93.66448 2261.98663  221.01823  260.83283   88.36948
```

I use $S = 10000$ and truncate the first 5000 iterations, the effective sample size of $\beta_1, \dots, \beta_5$ and $\beta_1\gamma_1, \dots, \beta_5\gamma_5$ are shown above, showing a generally acceptable mixing.

(b)

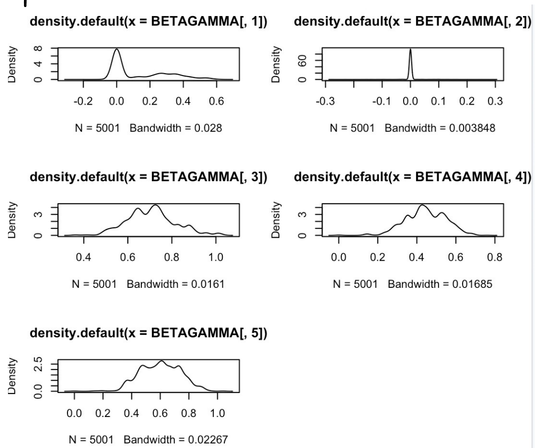
```
> table(as.data.frame(GAMMA))
, , V3 = 1, V4 = 0, V5 = 0

   V2
V1    0    1
  0   0    0
  1   0    0

, , V3 = 1, V4 = 1, V5 = 0

   V2
V1    0    1
  0   0    0
  1   5    0

, , V3 = 1, V4 = 0, V5 = 1

   V2
V1    0    1
  0   9    0
  1   0    0

, , V3 = 1, V4 = 1, V5 = 1

   V2
V1    0    1
  0 2557  126
  1 2165  139
```

```
> sort(table(as.data.frame(GAMMA)), decreasing = TRUE)[1:5]
[1] 2557 2165  139  126    9
> sort(table(as.data.frame(GAMMA)), decreasing = TRUE)[1:5] / 5000
[1] 0.5114 0.4330 0.0278 0.0252 0.0018
```

$\downarrow$ 00111 $\quad\downarrow$ 10111 $\quad\downarrow$ 11111 $\quad\downarrow$ 01111 $\quad\downarrow$ 10110

The approximation is good, since the top 5 have similar variables that is selected.

(c) posterior density of $\beta_j \gamma_j$



posterior mean of $\beta_j \gamma_j$

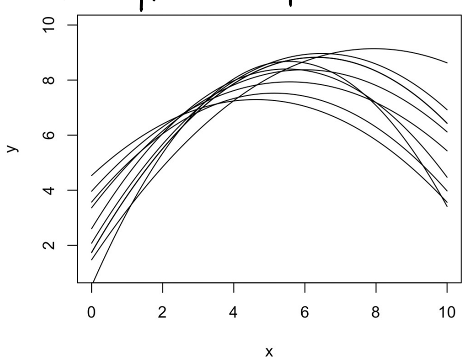
```
> apply(BETAGAMMA, 2, mean)
[1] 0.1364959951 0.0004558604 0.7059484018 0.4447755610 0.6005650203
```

$Pr(\gamma_j = 1 | x, y)$

```
> apply(GAMMA, 2, mean)
[1] 0.4617077 0.0529894 1.0000000 0.9982004 0.9990002
```

Problem 2:

(a) OLS for 10 plots



$\beta_j \sim MVN(\theta, \Sigma) \quad j = 1, 2, \cdots, 10$

$Y_j \sim MVN(X_j \beta_j, \sigma^2 I) \quad j = 1, 2, \cdots, 10$

$\therefore \hat{\theta}$ = sample mean of $\beta_{j,OLS}$

$\hat{\Sigma}$ = sample covariance of $\beta_{j,OLS}$

$\hat{\sigma}^2$ = mean of sample variance in each group

$\hat{\theta}$

```
> apply(BETA, 2, mean)
[1]  2.86875  1.85485 -0.15925
> cov(BETA)
            [,1]        [,2]         [,3]
[1,]  2.00120764 -0.69321312  0.044309549
[2,] -0.69321312  0.27555421 -0.020742679
[3,]  0.04430955 -0.02074268  0.001968451
```

$\hat{\Sigma}$

$\hat{\sigma}^2$

```
> apply(SIGMA_SQ.OLS, 2, mean)
[1] 1.748102
```

(b)

priors: $\theta \sim MVN(\mu_0, \Lambda_0)$

$\Sigma \sim$ inverse - wishart(η_0, S_0^{-1}) $\quad\quad$ $\Sigma^{-1} \sim W(4, \hat{\Sigma}^{-1})$

$\Sigma \sim$ inv-wishart$(4, \hat{\Sigma}^{-1})$

$\sigma^2 \sim$ inverse-gamma$(\frac{\nu_0}{2}, \frac{\nu_0}{2}\sigma_0^2)$

posterior: $\beta_j | y_j, X_j, \sigma^2, \theta, \Sigma \sim MVN\left((\Sigma^{-1} + X_j^T X_j/\sigma^2)^{-1} (\Sigma^{-1}\theta + X_j^T y_j/\sigma^2),\ (\Sigma^{-1} + X_j^T X_j/\sigma^2)^{-1} \right)$

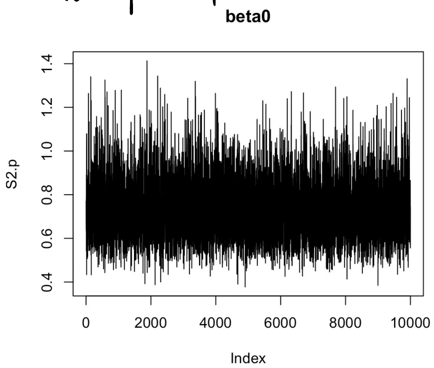
$\theta | \Sigma, \beta_1, \dots, \beta_m \sim MVN\left((\Lambda_0^{-1} + m\Sigma^{-1})^{-1} (\Lambda_0^{-1}\mu_0 + m\Sigma^{-1}\bar{\beta}),\ (\Lambda_0^{-1} + m\Sigma^{-1})^{-1} \right)$

$\Sigma | \theta, \beta_1, \dots, \beta_m \sim$ inverse - wishart$(\eta_0 + m, (S_0 + S_\theta)^{-1})$ $\quad\quad$ $S_\theta = \sum_{j=1}^{m} (\beta_j - \theta)(\beta_j - \theta)^T$

$\sigma^2 | \beta_j, y_j, X_j \sim$ inverse-gamma$\left(\frac{\nu_0 + m n_j}{2}, \frac{\nu_0 \sigma_0^2 + SSR}{2} \right)$

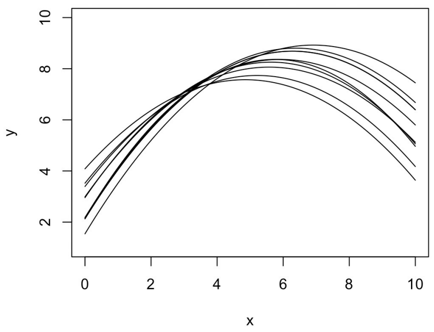
(c) MCMC diagnosis:

trace plot of σ^2:



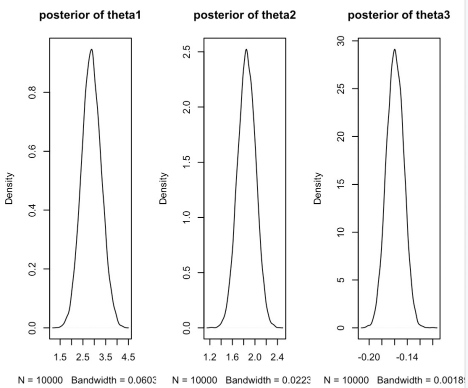
effective sample size =

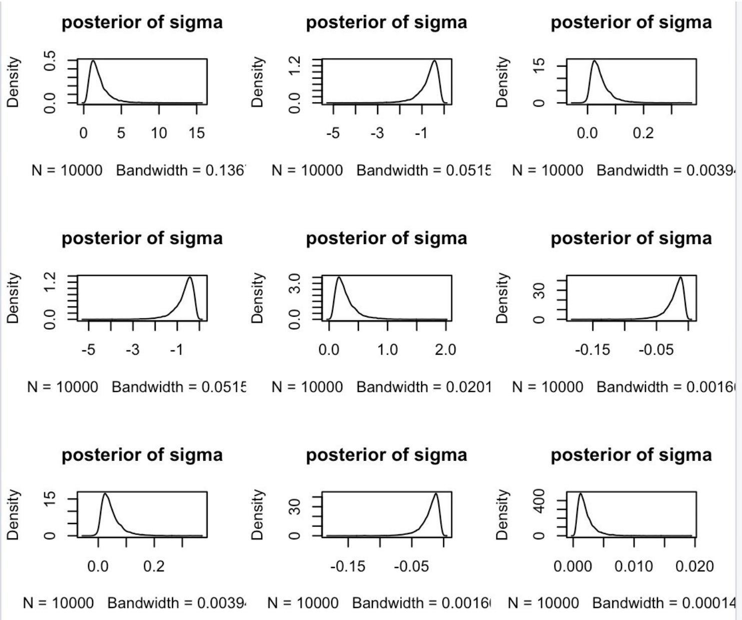
```
> effectiveSize(S2.p)
    var1
6710.077
```



regression lines in a is more spread, and regression lines after GLM is more centered. because GLMM borrows information from other groups, so having shrinkage effect to eliminate outliers.

(d)





$\beta_j \sim MVN(\theta, \Sigma)$

by the plot, $Var(\beta_j)$ is really high, indicating that β_j can be different across groups.

(e) X_{max} that maximizes expected yield:

```
> x_max = - avgtheta[2] / (2*avgtheta[3])
> x_max
[1] 5.819898
```

95% CI for X_{max}:

```
> quantile(X_max, c(0.025, 0.975))
    2.5%    97.5%
5.373661 6.293517
```