DDA 4010 Assignment 5 赵松霖 12009034b

Problem 1 :

(a) ∵ $\mathrm{Var}(y_{ij}|u,\tau^2)$ is bigger, because it includes both within and between group variability, while $\mathrm{Var}(y_{ij}|\theta_j,\sigma^2)$ only includes within group variability.

(b) if θ_j and σ^2 is known,

since $y_{i_1 j}, y_{i_2 j}|\theta_j,\sigma^2 \overset{iid}{\sim} N(\theta_j,\sigma^2)$

∴ $\mathrm{cov}(y_{i_1 j}, y_{i_2 j}|\theta_j,\sigma^2) = 0$, i.e., knowing $y_{i_1 j}$ gives no information about $y_{i_2 j}$ if θ_j is known.

if θ_j is unknown,

knowing about $y_{i_1 j}$ gives information about θ_j, thus gives information about $y_{i_2 j}$.

if $y_{i_1 j}$ is large, we expect θ_j is large, and $y_{i_2 j}$ is large.

∴ $\mathrm{cov}(y_{i_1 j}, y_{i_2 j}|u,\tau^2) > 0$

(c) ∵ ① $\mathrm{Var}(y_{ij}|\theta_j,\sigma^2) = \sigma^2$

② $\mathrm{Var}(\bar{y}_{\cdot j}|\theta_j,\sigma^2) = \mathrm{Var}\left(\frac{\sum_{i=1}^{n_j} y_{ij}}{n_j}|\theta_j,\sigma^2\right) = \frac{\sigma^2}{n_j}$

③ $\mathrm{Cov}(y_{i_1 j}, y_{i_2 j}|\theta_j,\sigma^2) = E(y_{i_1 j}y_{i_2 j}|\theta_j,\sigma^2) - E(y_{i_1 j}|\theta_j,\sigma^2)E(y_{i_2 j}|\theta_j,\sigma^2)$

$= 0$

④ $\mathrm{Var}(y_{ij}|u,\tau^2) = \mathrm{Var}\left(E(y_{ij}|\theta_j,\sigma^2)|u,\tau^2\right) + E\left(\mathrm{Var}(y_{ij}|\theta_j,\sigma^2)|u,\tau^2\right)$

$= \mathrm{Var}(\theta_j|u,\tau^2) + E(\sigma^2|u,\tau^2)$

$= \tau^2 + \sigma^2$

⑤ $\mathrm{Var}(\bar{y}_{\cdot j}|u,\tau^2) = \mathrm{Var}\left(E(y_{\cdot j}|\theta_j,\sigma^2)|u,\tau^2\right) + E\left(\mathrm{Var}(y_{\cdot j}|\theta_j,\sigma^2)|u,\tau^2\right)$

$= \mathrm{Var}(\theta_j|u,\tau^2) + E\left(\frac{\sigma^2}{n_j}|u,\tau^2\right)$

$= \tau^2 + \frac{\sigma^2}{n_j}$

⑥ $\mathrm{cov}(y_{i_1 j}, y_{i_2 j}|u,\tau) = E(\mathrm{cov}(y_{i_1 j}, y_{i_2 j}|\theta_j,\sigma^2)|u,\tau^2) + \mathrm{Cov}\left[E(y_{i_1 j}|\theta_j,\sigma^2), E(y_{i_2 j}|\theta_j,\sigma^2)|u,\tau^2\right]$

$= \mathrm{cov}(\theta_j,\theta_j|u,\tau^2)$

$= \mathrm{Var}(\theta_j|u,\tau^2)$

$= \tau^2$

$\mathrm{Var}(y_{ij}|\theta_i,\sigma^2) < \mathrm{Var}(y_{ij}|u,\tau^2)$

$\mathrm{cov}(y_{i_1 j}, y_{i_2 j}|\theta_j,\sigma^2) = 0$

$\mathrm{cov}(y_{i_1 j}, y_{i_2 j}|u,\tau^2) > 0$

which align with the results in (a) and (b)

(d) $p(u|\theta_1,\cdots,\theta_m,\sigma^2,\tau^2,y_1,\cdots,y_m)$

$= \frac{p(u,\theta_1,\cdots,\theta_m,\sigma^2,\tau^2,y_1,\cdots,y_m)}{\int p(u,\theta_1,\cdots,\theta_m,\sigma^2,\tau^2,y_1,\cdots,y_m)\,du}$

$= \frac{p(y_1,\cdots,y_m|\theta_1,\cdots,\theta_m,\sigma^2)\,p(\theta_1,\cdots,\theta_m|u,\tau^2)\,p(u)\,p(\sigma^2,\tau^2)}{p(y_1,\cdots,y_m|\theta_1,\cdots,\theta_m,\sigma^2)\,p(\sigma^2,\tau^2)\int p(\theta_1,\cdots,\theta_m|u,\tau^2)\,p(u)\,du}$

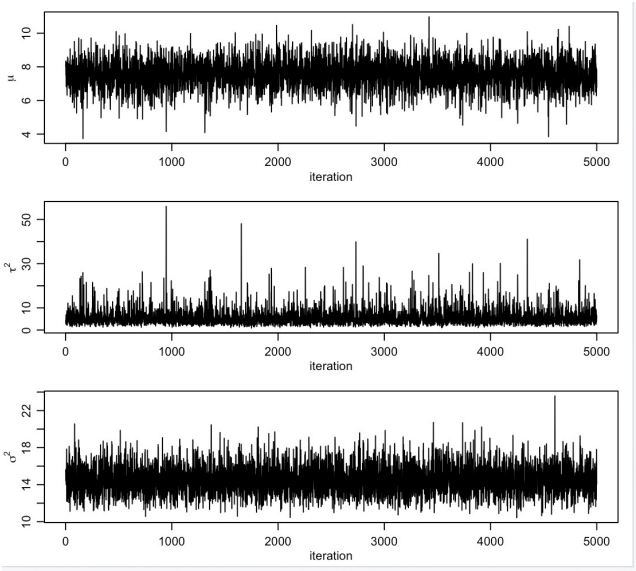
$= \frac{p(\theta_1,\cdots,\theta_m|u,\tau^2)\,p(u)}{\int p(\theta_1,\cdots,\theta_m|u,\tau^2)\,p(u)\,du}$

$= p(u|\theta_1,\cdots,\theta_m,\tau^2)$

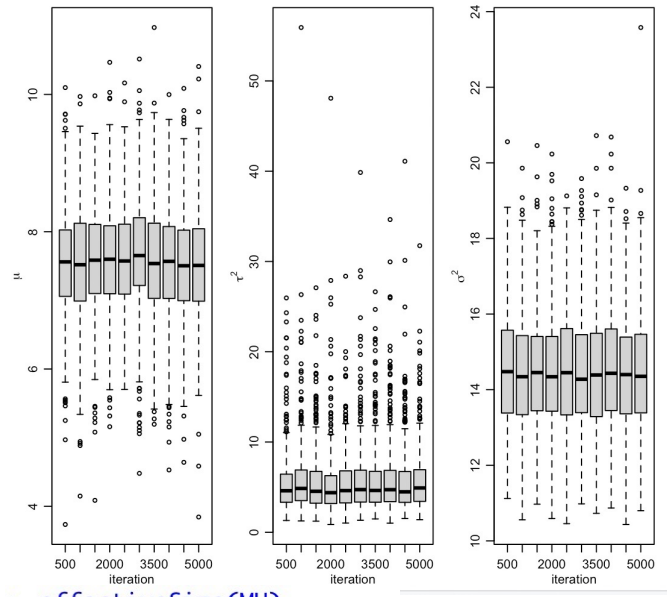
This means u does not depend on $y_1, \cdots, y_n$ and σ^2, i.e., u does not depend on the observed data, if $\theta_1, \cdots, \theta_m$ is known.

Problem 2:

(a) trace plot:



Stationary plot:




```
> effectiveSize(MU)
    var1
4147.251
> effectiveSize(TAU_SQ)
    var1
3729.928
> effectiveSize(SIGMA_SQ)
    var1
4713.354
```

effective sample size of u

effective sample size of τ^2

effective sample size of σ^2

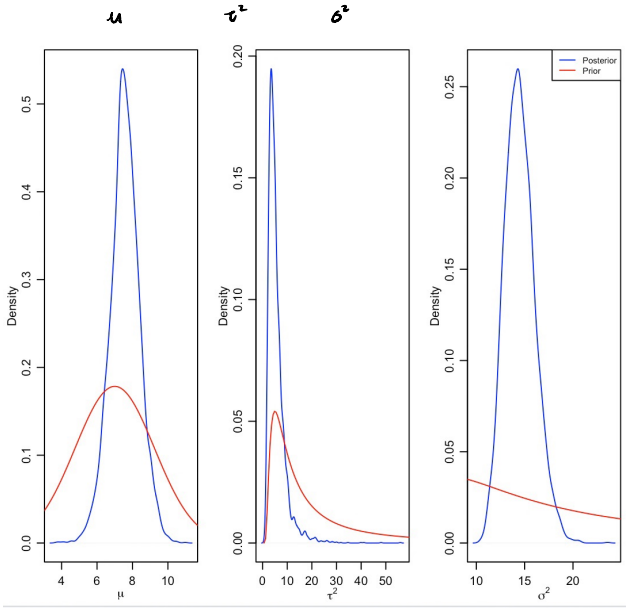
From the stationary plot and trace plot, we can see our MCMC samples reaches stationary. effective sample sizes are all > 1000, so the number of iterations is large enough.

(b) posterior means of u, τ^2, σ^2

```
> apply(M, MARGIN=2, FUN=mean)
[1]  7.563386  5.581711 14.493790
```

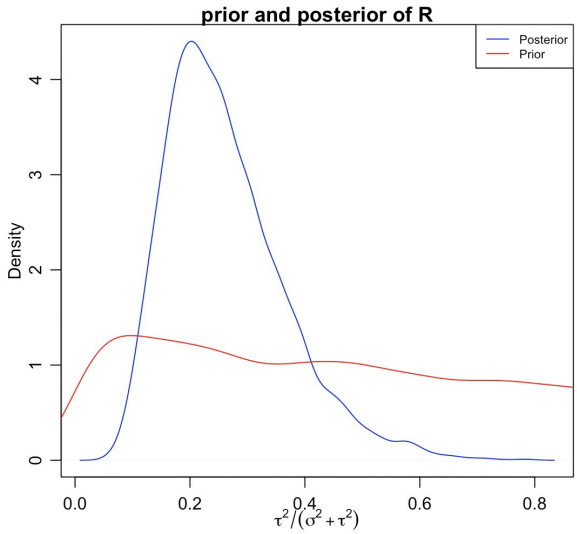
posterior confidence regions of μ, τ^2, σ^2 respectively.

```
> apply(M, MARGIN = 2, FUN = quantile, probs = c(0.025, 0.975))
          [,1]     [,2]     [,3]
2.5%  5.877648  1.83903 11.76387
97.5% 9.200664 15.28747 17.82563
```



remark: the posteriors are more centered than prior distribution. So after observing the data, we are more confident about the true value of μ, τ^2 and σ^2. μ is around 7.5, τ^2 is around 5.5, and σ^2 is around 14.5.

(c)



posterior mean of R

```
> mean(R.posterior)
[1] 0.2618374
```

$R = \frac{\tau^2}{\sigma^2+\tau^2}$ represents the proportion of the between school variability among total variability of y_{ij}.

After observing the data, the posterior mean of R is 0.26, which means the within group variability contributes most of the total variability.

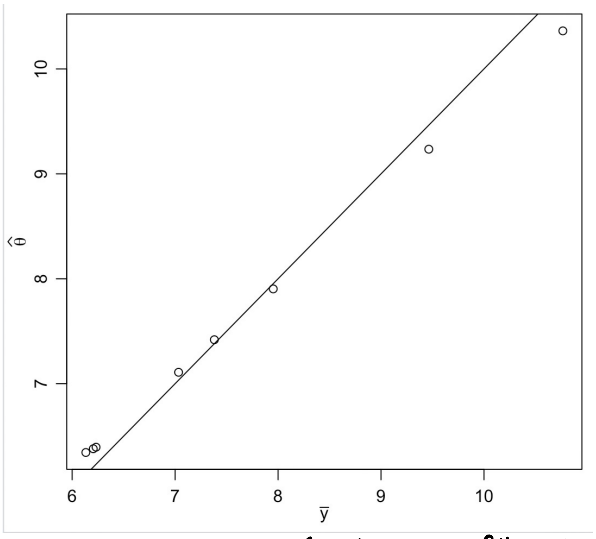
(d) $p(\theta_1 < \theta_6 | y)$

```
> mean(THETA[,7]<THETA[,6])
[1] 0.5208
```

$p(\theta_1 < \theta_{-1} | y)$

```
> mean(minimum==7)
[1] 0.3208
```

(e)



relationship of $\bar{y}_j$ and $\hat{\theta}_j$ does not follow $y=x$.

① when $\bar{y}_j$ is away from u, there will be shrinkage effect:
$\hat{\theta}_j$ moves close to u, difference between $\bar{y}_j$ and $\hat{\theta}_j$ will be larger if $|\bar{y}_j - u|$ is larger.

② when $\bar{y}_j$ is close to u, $\bar{y}_j$ is also close to $\hat{\theta}_j$.

```
> sum/sum(n)
[1] 7.691278
```
$\Rightarrow \bar{y}$

```
> mean(MU)
[1] 7.563386
```
$\Rightarrow \hat{u}$

Problem 3:

(a) posterior mean of β:

```
> beta.posterior
 [1]  0.277807220  0.002181378  0.532223831  1.438391578 -0.764108880 -0.068558382  0.134361426 -0.065762314
 [9]  0.108240773 -0.266088740  0.358137273  0.229835939  0.708270476 -0.279304157 -0.058860836
```

95% CI of β:

```
> apply(BETA, 2, quantile, probs = c(0.025, 0.975))
            [,1]       [,2]      [,3]        [,4]      [,5]       [,6]       [,7]       [,8]       [,9]
2.5%  0.02763848 -0.3274552 0.2068963 -0.06838007 -2.297244 -0.3407333 -0.1424860 -0.2882754 -0.2198412
97.5% 0.52641303  0.3392396 0.8583802  2.89560840  0.808366  0.2153559  0.4140429  0.1641163  0.4265987
           [,10]      [,11]      [,12]     [,13]      [,14]      [,15]
2.5%  -0.61742527 0.02330779 -0.2339990 0.2991445 -0.5242016 -0.2950214
97.5%  0.09618591 0.67841679  0.6969059 1.1266416 -0.0372050  0.1901738
```

M, Ed, U_2, Ineq, Prob are strongly predictive of crime rates.
because their 95% credible intervals do not contain 0.

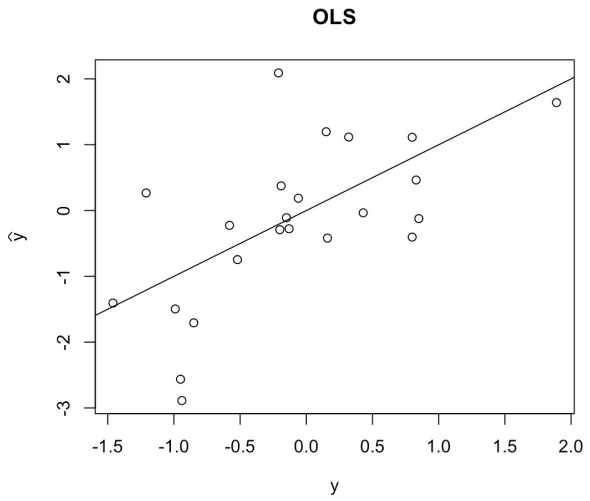
M, Ed, U_2, Ineq have significant positive relationship with crime rates,
and Prob has significant negative relationship with crime rates.
other variables do not have significant relationship with crime rates.

Least square estimates for β:

```
> beta.ols
               [,1]
M      0.2865177028
So    -0.0001179958
Ed     0.5445161778
Po1    1.4716146465
Po2   -0.7817757455
LF    -0.0659672893
M.F    0.1313002714
Pop   -0.0702910179
NW     0.1090590127
U1    -0.2705407273
U2     0.3687335028
GDP    0.2380580097
Ineq   0.7262918200
Prob  -0.2852262729
Time  -0.0615771841
```

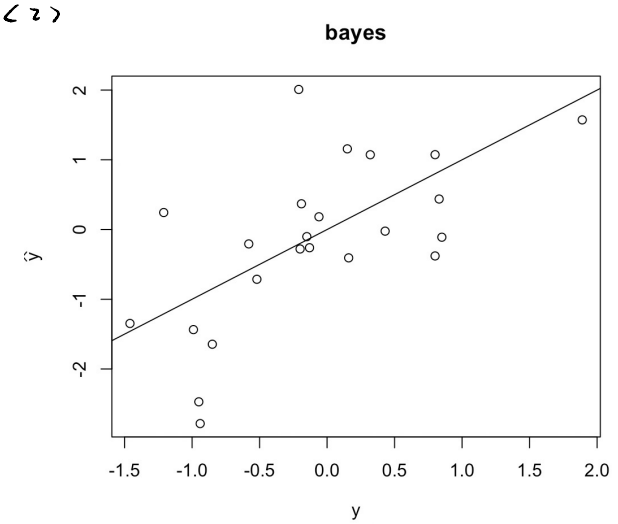
Least square estimates are similar to marginal posterior means of β.

(b) <1>




```
> error
[1] 0.8877496
```

<2>




```
> error
[1] 0.8221626
```

$\hat{\beta}_{Bayes}$ has slightly better accuracy for the test data than $\hat{\beta}_{OLS}$

$\hat{\beta}_{OLS}$ works best for training data, but may not work well for test data.

(c) We run 50 times, i.e, sample 50 $\{y_{tr}, X_{tr}\}$'s and $\{y_{te}, X_{te}\}$'s.

```
> mean(Error_OLS)
[1] 0.8174934
> mean(Error_Bayes)
[1] 0.783311
> mean(Error_Bayes < Error_OLS)
[1] 0.9
```

→ prediction error of $\hat{\beta}_{OLS}$

→ prediction error of $\hat{\beta}_{Bayes}$.

$P(\text{error}_{Bayes} < \text{error}_{OLS}) = 0.9$

average prediction error is smaller for Bayes estimator than OLS estimator.

for 90% of the times among 50 iterations, prediction error of $\hat{\beta}_{Bayes}$ is smaller.

Problem 4

(a) $\theta_i = Pr(Y_i = 1 | \alpha, \beta, x_i)$

$$\log\left(\frac{\theta_i}{1-\theta_i}\right) = \alpha + \beta x_i$$

$$\therefore \quad \theta_i(1 + e^{\alpha+\beta x_i}) = e^{\alpha+\beta x_i}$$

$$\theta_i = \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}$$

$$\therefore \prod_{i=1}^{n} P(y_i | \alpha, \beta, x_i) = \prod_{i=1}^{n} \left(\frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}\right)^{y_i} \left(\frac{1}{1+e^{\alpha+\beta x_i}}\right)^{1-y_i}$$

$$= \prod_{i=1}^{n} \frac{e^{(\alpha+\beta x_i) y_i}}{1+e^{\alpha+\beta x_i}}$$

(b) we want $0 \le \theta_i \le 1$ for $10 \le x_i \le 15$

i.e., $0 \le \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}} \le 1$ for $0 \le x_i \le 15$

we want prior be uninformative,

So we let $10^{-2} \le \theta_i \le 1-10^{-2}$ be the prior,

$\therefore$ approximately, $-5 \le \log \frac{\theta_i}{1-\theta_i} \le 5$

$\therefore \quad -5 \le \alpha + \beta x_i \le 5$

when $\alpha = 0$, $\quad -5 \le \beta x_i \le 5$

when $x_i = 10$ be minimum, $\quad -\frac{1}{2} \le \beta \le \frac{1}{2}$

$\therefore$ $\beta \sim N(0, \frac{1}{4})$ be prior of β

when $\beta = 0$, $\quad -5 \le \alpha \le 5$

$\therefore$ $\alpha \sim N(0, 25)$ be prior of α

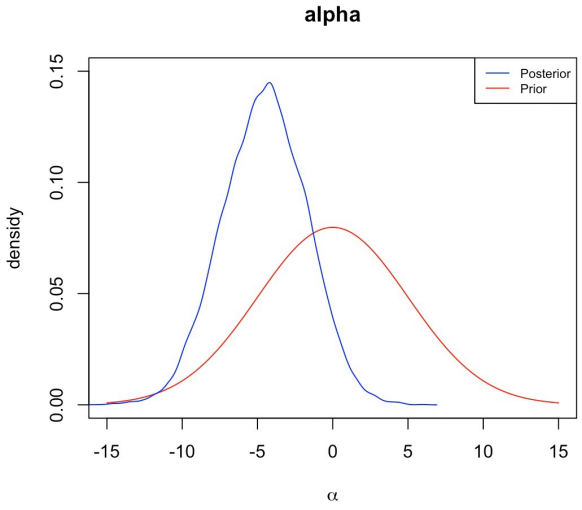
(c) I ran the Metropolis algorithm 50000 times,

the effective sample size is:

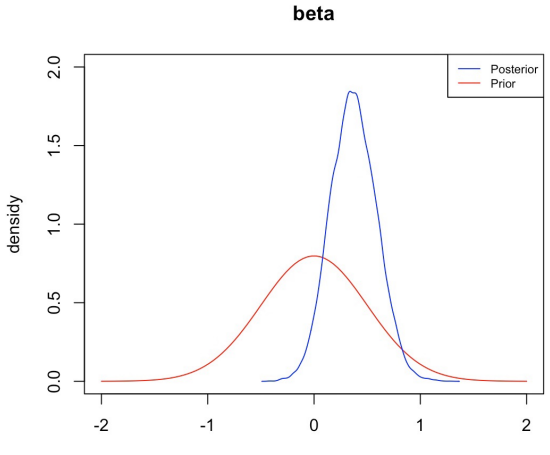
```
> apply(THETA,2,effectiveSize)
[1] 1359.971 1346.402
```

↑ α ↑ β

(d) posterior and prior for α:



posterior and prior for β.



posterior of α and β are both more centered than their own priors.

(e)