MATH 8820 (Fall 2018)
Homework 4

Shirong Zhao

Problem 1

(1) **Solution.**

In the GLM setting, the data set consists of n observations with univariate response y_i and a p-dimensional vectors of covariates x_i, $i = 1, 2, \ldots, n$. The observations are assumed to be independent with exponential family density

$$f(y_i|\theta_i) = \exp\{(y_i\theta_i - b(\theta_i))/\phi_i\}c(y_i, \theta_i)$$

For exponential family density, we know that the mean is $\mu_i = E(y_i|\theta_i) = b'(\theta_i)$. Moreover, we have the link relation $g(\mu_i) = \eta_i = \boldsymbol{x_i}'\boldsymbol{\beta}$, $i = 1, 2, \ldots, n$.

If we specify the prior distribution for $\boldsymbol{\beta}$ as $N(\boldsymbol{a}, \boldsymbol{R})$, then the posterior distribution for $\boldsymbol{\beta}$ is

$$\pi(\boldsymbol{\beta}) \propto \exp\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{a})'\boldsymbol{R}^{-1}(\boldsymbol{\beta} - \boldsymbol{a}) + \sum_{i=1}^{n} \frac{y_i\theta_i - b(\theta_i)}{\phi_i}\}$$

However, the posterior distribution is not easy to get. Hence we need M-H algorithm. The author uses the idea from Iterative Weighted Least Squares (IWLS) and proposes the following proposal distribution.

$$J(\beta^*|\beta^{(t-1)}) = N(\boldsymbol{m}^{(t-1)}, \boldsymbol{C^{(t-1)}})$$

$$\boldsymbol{m}^{(t-1)} = (\boldsymbol{R}^{-1} + \boldsymbol{X}'\boldsymbol{W(\beta^{(t-1)})}\boldsymbol{X})^{-1} \times (\boldsymbol{R}^{-1}\boldsymbol{a} + \boldsymbol{X}'\boldsymbol{W(\beta^{(t-1)})}\boldsymbol{\tilde{y}(\beta^{(t-1)}})$$

$$\boldsymbol{C}^{(t-1)} = (\boldsymbol{R}^{-1} + \boldsymbol{X}'\boldsymbol{W(\beta^{(t-1)})}\boldsymbol{X})^{-1}$$

And $\boldsymbol{\tilde{y}(\beta^{(t-1)}}$ is a vector of transformed observations, where $\tilde{y}_i(\boldsymbol{\beta}) = \eta_i + (y_i - \mu_i)g'(\mu_i)$, and $\boldsymbol{W(\beta^{(t-1)})}$ is an associated diagonal matrix with respective components, where $W_i(\boldsymbol{\beta}) = b''(\theta_i)g'(\mu_i)^2$

The M-H ratio is calculated as usual as following:

$$r = \min\{1, \frac{p(\beta^*|Y)J(\beta^{(t-1)}|\beta^*)}{p(\beta^{(t-1)}|Y)J(\beta^*|\beta^{(t-1)})}\}$$

□

(2) **Solution.**

For the logistic regression, we have the following setup as the same as that in Lecture Notes.

$$y_i \sim Bin(\pi_i, 1)$$

$$\log(\frac{\pi_i}{1-\pi_i}) = \beta_0 + \beta_1 x_i = \boldsymbol{x}'\boldsymbol{\beta}$$

The density for Bernoulli distribution is

$$f(y_i) = \pi_i^{y_i}(1-\pi_i)^{1-y_i} = (\frac{\pi_i}{1-\pi_i})^{y_i}(1-\pi_i) = \exp(y_i \log(\frac{\pi_i}{1-\pi_i}) + \log(1-\pi_i))$$

Therefore, $\theta_i = \log(\frac{\pi_i}{1-\pi_i})$, and hence $\pi_i = \frac{\exp(\theta_i)}{1+\exp(\theta_i)}$ and $b(\theta_i) = -\log(1-\pi_i) = \log(1+\exp(\theta_i))$. Also $g(\mu_i) = g(\pi_i) = log(\frac{\pi_i}{1-\pi_i}) = \boldsymbol{x_i}'\boldsymbol{\beta}$. After some algebra, it can be shown that

$$\tilde{y}_i(\boldsymbol{\beta}) = \boldsymbol{x_i}'\boldsymbol{\beta} + (y_i - \pi_i)\frac{1}{\pi_i(1-\pi_i)}$$

$$W_i(\boldsymbol{\beta}) = \pi_i(1-\pi_i)$$

where $\pi_i = \frac{\exp(\boldsymbol{x_i}'\boldsymbol{\beta})}{1+\exp(\boldsymbol{x_i}'\boldsymbol{\beta})}$. Then after specifying initial value for $\boldsymbol{\beta}$, we can get $\boldsymbol{m}^{(t-1)}$ and $\boldsymbol{C}^{(t-1)}$, and follow the steps in (1) to get the posterior distribution samples for $\boldsymbol{\beta}$

□

(3) **Solution.**

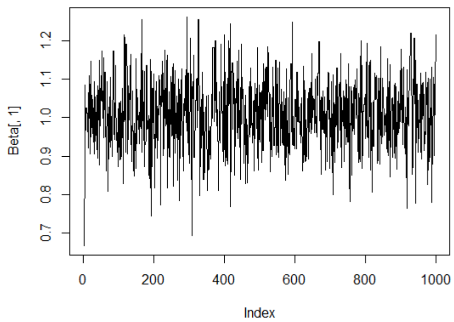
Please check the function DG(y, X, beta=c(0,0), a=rep(0,2), R=100*diag(2), iter=1e3)

To show our model actually works. We simulate 1000 data. And we let $\boldsymbol{\beta} = c(1,2)$. And we use the methods developed in the paper to run 1×10^3 iterations. Our estimates and plots are shown below.[1]

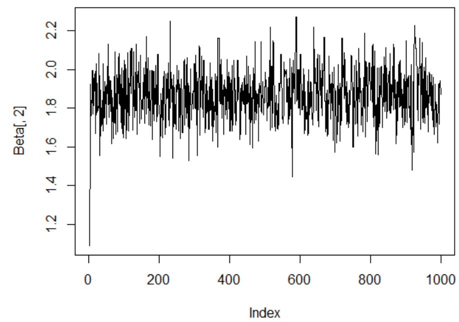
Table 1: Outcome of Simulations

Parameter	True	Estimate	HPD
β_1	1	1.0099	[0.8393, 1.1996]
β_2	2	1.8695	[1.6493, 2.1658]

[1]For the details of our simulation, please check the attached codes.



figureTrace Plot for β_1



figureTrace Plot for β_2

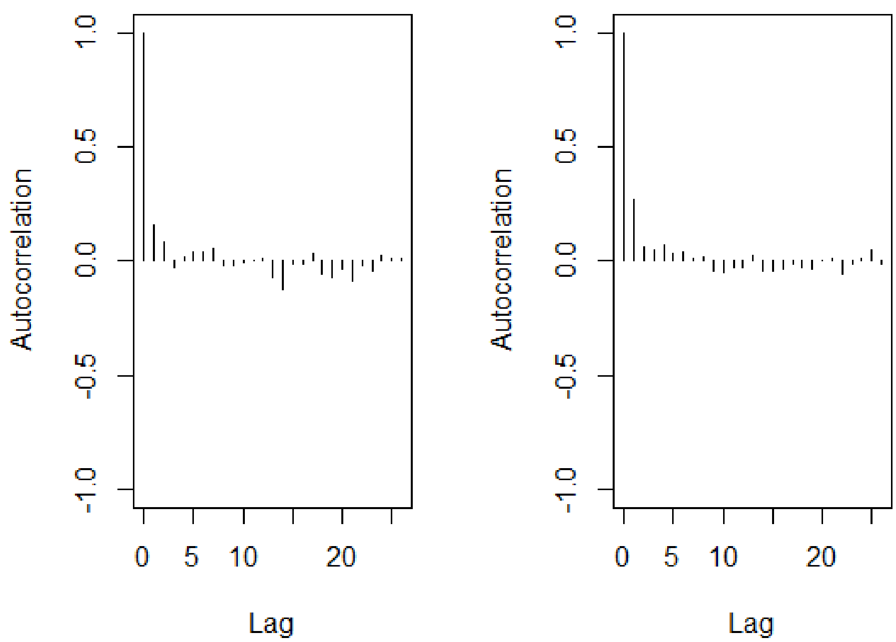


Figure 1: Autocorrelation for β_1 and β_2

And the effective sample size is 1087 for β_1 and is 572 for β_2. Overall, it appears that the chain has converged for all parameters and the HPD covers the true value of parameters.

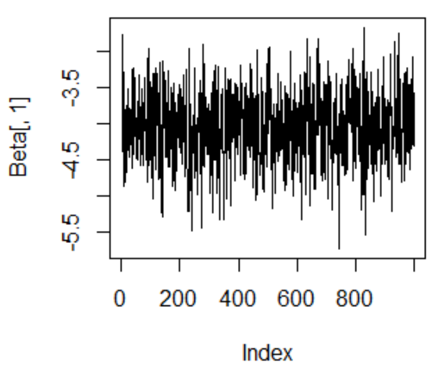
□

(4) **Solution.**

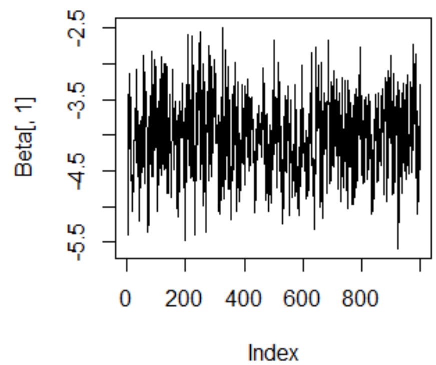
We use the methods developed in the paper to run 1×10^3 iterations. Our estimates and plots (denoted as DG, HPD1, ESS1(effective sample size), ACC1(accepting rates)) compared with the outcomes using your MH codes (denoted as MH, HPD2, ESS2(effective sample size), ACC2(accepting rates)) are shown below.

Table 2: Outcome of Simulations

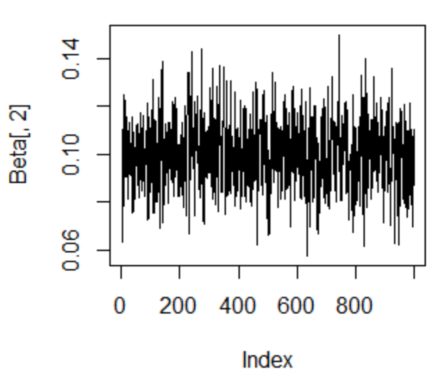
Param	DG	MH	HPD1	HPD2	ESS1	ESS2	ACC1	ACC2
β_1	-4.0538	-4.0111	[-4.9903,-3.0660]	[-5.0323, -2.9847]	740	395	0.95	0.42
β_2	0.1004	0.0990	[0.0714, 0.1270]	[0.0698, 0.1289]	698	390	0.95	0.44



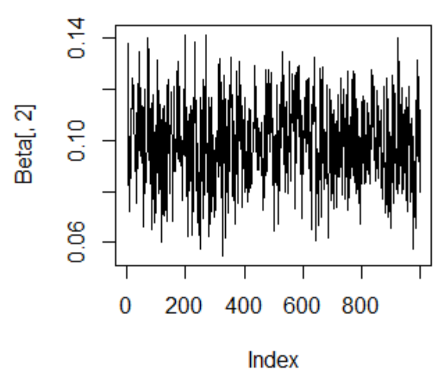
figureTrace Plot for β_1 in DG



figureTrace Plot for β_1 in MH



figureTrace Plot for β_2 in DG



figureTrace Plot for β_2 in MH

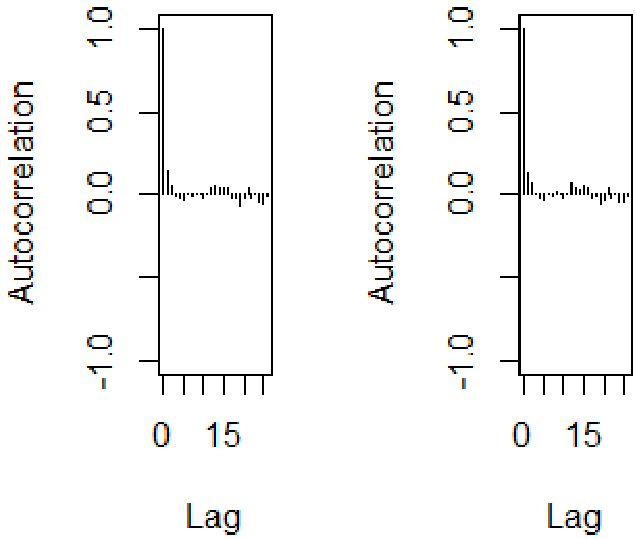


Figure 2: Autocorrelation for β_1 and β_2 in DG

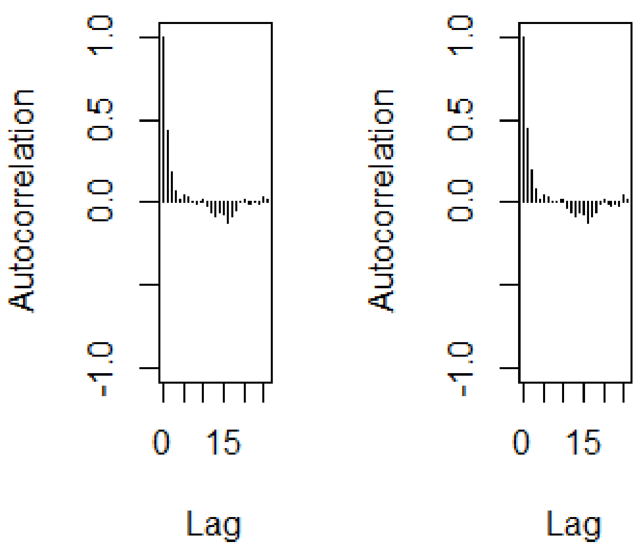


Figure 3: Autocorrelation for β_1 and β_2 in MH

As we can see that the estimates, HPD, trace plots are very similar. The difference is that DG method has a larger effective sample size, lower correlation and higher accepting rate than MH method. And therefore DG method works very well.

This is not surprising, as mentioned by the author,"The Markov chain formed by the above scheme leads to an algorithm with high acceptance rates due to the good approximation of the proposal to the true posterior distribution and the chain moves are dictated by the structure of the model."

□

Problem 2

(1) **Solution.**

The idea about CMP prior is that with p predictor variables, we evaluate our prior information at p locations in the predictor space. For each location, we specify a prior distribution for the mean of potential observations at the location. These means are conditioned on their locations in predictor space. By assuming independent priors for the various locations, the prior on the regression coefficients $\boldsymbol{\beta}$ is then induced from the CMP.

Consider a random observation y that has the density or mass function of the form

$$f(y|\theta,\phi;w) = h(\phi,y,w)\exp\{\frac{w}{\phi}(\theta y - r(\theta))\}$$

This is a one parameter exponential family density.

Consider a GLM model. Assume n independent observations y_i with corresponding covariate p vectors $\boldsymbol{x_i}$ such that

$$y_i \sim f(y_i|\theta_i,\phi;w_i), \quad E(y_i) = m_i, \quad g(m_i) = \boldsymbol{x_i'\beta}$$

The prior on $\boldsymbol{\beta}$ is induced from a CMP on $\tilde{\boldsymbol{m}} = (\tilde{m}_1, \tilde{m}_2, , \tilde{m}_p)$, where $\tilde{m}_i = E(y_i \tilde{|} \tilde{x}_i)$, i=1, 2, , p, is the mean response for a potentially observable response $\tilde{y}_i$ at covariate vector $\tilde{x}_i$. Assume that the p covariate vectors $\tilde{x}_i$ are linearly independent. And if we specify an arbitrary proper prior on $\tilde{\boldsymbol{m}}$, $\pi_0(\tilde{\boldsymbol{m}})$, and if the $\tilde{m}_i$'s can be assessed independently, then the CMP is

$$\pi(\tilde{\boldsymbol{m}}) \propto \prod_{i=1}^{p} \pi_{0i}(\tilde{m}_i)$$

In this instance, we can obtain the induced prior for $\boldsymbol{\beta}$

$$\pi(\boldsymbol{\beta}) \propto \prod_{i=1}^{p} \pi_{0i}(g^{-1}(\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\boldsymbol{\beta}))/\dot{g}(g^{-1}(\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\boldsymbol{\beta}))$$

where $\dot{g}(x) = dg(x)/dx$

□

(2) **Solution.**

For the logistic regression, we have the following setup as the same as that in Lecture Notes.

$$y_i \sim Bin(\pi_i, 1)$$

$$\log(\frac{\pi_i}{1-\pi_i}) = \beta_0 + \beta_1 x_i = \boldsymbol{x}'\boldsymbol{\beta}$$

The density for Bernoulli distribution is

$$f(y_i) = \pi_i^{y_i}(1-\pi_i)^{1-y_i} = (\frac{\pi_i}{1-\pi_i})^{y_i}(1-\pi_i) = \exp(y_i \log(\frac{\pi_i}{1-\pi_i}) + \log(1-\pi_i))$$

Therefore $\pi_i = \frac{\exp(\theta_i)}{1+\exp(\theta_i)}$ and $g(\mu_i) = g(\pi_i) = \log(\frac{\pi_i}{1-\pi_i}) = \boldsymbol{x_i}'\boldsymbol{\beta}$.

The prior on $\boldsymbol{\beta}$ is induced from a CMP on $\tilde{\boldsymbol{m}} = (\tilde{m}_1, \tilde{m}_2, , \tilde{m}_p)$, where $\tilde{m}_i = E(\widetilde{y_i|}\tilde{x}_i)$, i=1, 2, , p, is the success probability for a potentially observable response $\tilde{y}_i$ at covariate vector $\tilde{x}_i$. Assume that the p covariate vectors $\tilde{x}_i$ are linearly independent. In particular, one could specify that independently, $\tilde{m}_i \sim beta(a_{1i}, a_{2i})$. that is

$$\pi(\tilde{\boldsymbol{m}}) \propto \prod_{i=1}^{p} \tilde{m}_i^{a_{1i}-1}(1-\tilde{m}_i)^{a_{2i}-1}$$

This independence CMP induces a prior on $\boldsymbol{\beta}$. The induced prior on $\boldsymbol{\beta}$ for the model $g^{-1}(\boldsymbol{x'\beta}) = F(\boldsymbol{x'\beta})$ is

$$\pi(\boldsymbol{\beta}) \propto \prod_{i=1}^{p}(F(\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\boldsymbol{\beta}))^{a_{1i}-1}(1-F(\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\boldsymbol{\beta}))^{a_{2i}-1}f(\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\boldsymbol{\beta})$$

For the logistic regression, we have $F(\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\boldsymbol{\beta}) = \frac{\exp(\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\boldsymbol{\beta})}{1+\exp(\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\boldsymbol{\beta})}$, and $f(\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\boldsymbol{\beta}) = \frac{\exp(\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\boldsymbol{\beta})}{(1+\exp(\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\boldsymbol{\beta}))^2}$
Therefore the CMP prior for the logistic regression is

$$\begin{aligned}
\pi(\boldsymbol{\beta}) &\propto \prod_{i=1}^{p}(\frac{\exp(\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\boldsymbol{\beta})}{1+\exp(\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\boldsymbol{\beta})})^{a_{1i}-1}(\frac{1}{1+\exp(\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\boldsymbol{\beta})})^{a_{2i}-1}\frac{\exp(\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\boldsymbol{\beta})}{(1+\exp(\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\boldsymbol{\beta}))^2} \\
&\propto \prod_{i=1}^{p}(\frac{\exp(\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\boldsymbol{\beta})}{1+\exp(\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\boldsymbol{\beta})})^{a_{1i}}(\frac{1}{1+\exp(\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\boldsymbol{\beta})})^{a_{2i}}
\end{aligned}$$

□

(3) **Solution.**

Notice that the likelihood function for logistic regression is

$$\pi(\boldsymbol{\beta}|\boldsymbol{y}) \propto \prod_{i=1}^{n}(\frac{\exp(\boldsymbol{x_i}'\boldsymbol{\beta})}{1+\exp(\boldsymbol{x_i}'\boldsymbol{\beta})})^{y_i}(\frac{1}{1+\exp(\boldsymbol{x_i}'\boldsymbol{\beta})})^{1-y_i}$$

Therefore the posterior distribution for $\boldsymbol{\beta}$ is

$$\begin{aligned}
\pi(\boldsymbol{\beta}|\boldsymbol{y}) \propto &\prod_{i=1}^{n}(\frac{\exp(\boldsymbol{x_i}'\boldsymbol{\beta})}{1+\exp(\boldsymbol{x_i}'\boldsymbol{\beta})})^{y_i}(\frac{1}{1+\exp(\boldsymbol{x_i}'\boldsymbol{\beta})})^{1-y_i} \\
&\times \prod_{i=1}^{p}(\frac{\exp(\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\boldsymbol{\beta})}{1+\exp(\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\boldsymbol{\beta})})^{a_{1i}}(\frac{1}{1+\exp(\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\boldsymbol{\beta})})^{a_{2i}}
\end{aligned}$$

If we let $a_{1i} = \tilde{y}_i$ and $a_{2i} = 1 - \tilde{y}_i$, and also let $\boldsymbol{x}_{n+i} = \tilde{\boldsymbol{x}}_{\boldsymbol{i}}$, $y_{n+i} = \tilde{y}_i$, i=1, 2, , p, then we have

$$\pi(\boldsymbol{\beta}|\boldsymbol{y}) \propto \prod_{i=1}^{n}(\frac{\exp(\boldsymbol{x_i}'\boldsymbol{\beta})}{1+\exp(\boldsymbol{x_i}'\boldsymbol{\beta})})^{y_i}(\frac{1}{1+\exp(\boldsymbol{x_i}'\boldsymbol{\beta})})^{1-y_i}$$
$$\times \prod_{i=1}^{p}(\frac{\exp(\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\boldsymbol{\beta})}{1+\exp(\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\boldsymbol{\beta})})^{\tilde{y}_i}(\frac{1}{1+\exp(\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\boldsymbol{\beta})})^{1-\tilde{y}_i}$$
$$\propto \prod_{i=1}^{n+p}(\frac{\exp(\boldsymbol{x_i}'\boldsymbol{\beta})}{1+\exp(\boldsymbol{x_i}'\boldsymbol{\beta})})^{y_i}(\frac{1}{1+\exp(\boldsymbol{x_i}'\boldsymbol{\beta})})^{1-y_i}$$

Therefore it is now easy to use the method in Problem 1. The difference is that we have $n+p$ observations rather than just n observations, and consequently, we do not have the normal prior $N(\boldsymbol{a}, \boldsymbol{R})$ for $\boldsymbol{\beta}$. Now consider $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n, \boldsymbol{x}n+1, ..., \boldsymbol{x}_{n+p})'$, then we can implement the method in Problem 1 with the following proposal distribution

$$J(\beta^*|\beta^{(t-1)}) = N(\boldsymbol{m}^{(t-1)}, \boldsymbol{C}^{\boldsymbol{(t-1)}})$$

$$\boldsymbol{m}^{(t-1)} = (\boldsymbol{R}^{-1} + \boldsymbol{X}'\boldsymbol{W}\boldsymbol{(}\boldsymbol{\beta}^{\boldsymbol{(t-1)}}\boldsymbol{)}\boldsymbol{X})^{-1} \times (\boldsymbol{R}^{-1}\boldsymbol{a} + \boldsymbol{X}'\boldsymbol{W}(\boldsymbol{\beta}^{\boldsymbol{(t-1)}})\tilde{\boldsymbol{y}}\boldsymbol{(}\boldsymbol{\beta}^{\boldsymbol{(t-1)}}))$$

$$\boldsymbol{C}^{(t-1)} = (\boldsymbol{R}^{-1} + \boldsymbol{X}'\boldsymbol{W}\boldsymbol{(}\boldsymbol{\beta}^{\boldsymbol{(t-1)}}\boldsymbol{)}\boldsymbol{X})^{-1}$$

$$\tilde{y}_i(\boldsymbol{\beta}) = \boldsymbol{x_i}'\boldsymbol{\beta} + (y_i - \pi_i)\frac{1}{\pi_i(1-\pi_i)}, \quad i = 1, 2, ..., n+p$$

$$W_i(\boldsymbol{\beta}) = \pi_i(1-\pi_i)$$

where $\pi_i = \frac{\exp(\boldsymbol{x_i}'\boldsymbol{\beta})}{1+\exp(\boldsymbol{x_i}'\boldsymbol{\beta})}$.

□

Problem 3

For the CAR model described in Problem 2 of Homework 3, the posterior distribution is

$$P(\beta_0, b, \sigma^{-2}, \tau^{-2}|X, Y) \propto (\sigma^{-2})^{\frac{n}{2}} \exp\{-\frac{(Y-\beta_0-b)'(Y-\beta_0-b)}{2\sigma^2}\}$$
$$\times (\tau^{-2})^{\frac{n}{2}} \exp\{-\frac{b'(D-\rho W)b}{2\tau^2}\}$$
$$\times (\sigma^{-2})^{a_\sigma - 1} \exp\{-\sigma^{-2}b_\sigma\}$$
$$\times (\tau^{-2})^{a_\tau - 1} \exp\{-\tau^{-2}b_\tau\}$$

ρ only appears one time, however, we could not know the exact posterior distribution of ρ. Therefore we develop the following Metropolis-Hastings algorithm that can be used to sample ρ. I develop two methods to sample ρ.

(1)

Since $\rho \in [-1, 1]$, the proposal distribution for ρ is reflected random walk, i.e. $J(\rho|\rho^{(s)}) = Unif(\rho^{(s)} - \delta, \rho^{(s)} + \delta)$. If $\rho < -1$, we use $-2 - \rho$. If $\rho > 1$, we use $2 - \rho$. However we need to tune δ so that the accepting rate is roughly 35%.

(2)

Since $\rho \in [-1, 1]$, then $\rho+1 \in [0, 2]$, and hence $(\rho+1)/2 \in [0, 1]$, and therefore $\log(\frac{(\rho+1)/2}{1-(\rho+1)/2}) = \log(\frac{1+\rho}{1-\rho}) \in (-\infty, +\infty)$. The proposal distribution is $J(\rho|\rho^{(s)}) : \log(\frac{1+\rho}{1-\rho}) \sim N(\log(\frac{1+\rho^{(s)}}{1-\rho^{(s)}}), c)$. However we need to tune c so that the accepting rate is roughly 35%.

The M-H ratio is calculated as usual as following:

$$r = \min\{1, \frac{p(\rho^*|Y)J(\rho^{(s)}|\rho^*)}{p(\rho^{(s)}|Y)J(\rho^*|\rho^{(s)})}\}$$

Where ρ^* is the proposed ρ given $\rho^{(s)}$ from the proposal distribution. If we assume the prior for $\rho \sim Unif(-1, 1)$, then $p(\rho^*|Y) = (\tau^{-2})^{\frac{n}{2}} \exp\{-\frac{b'(D-\rho^* W)b}{2\tau^2}\}$.

Both M-H should work.