MATH 8820 (Fall 2018)
Homework 3

Shirong Zhao

Before to answer the following questions, first we do a small trick just as that in class. if $x \sim MVN(\mu, \Sigma)$, then

$$f(X) \propto \exp\{-\frac{(X-\mu)'\Sigma^{-1}(X-\mu)}{2}\}$$
$$\propto \exp\{-\frac{(X'\Sigma^{-1}X - 2X'\Sigma^{-1}\mu)}{2}\}$$

That is to say, if

$$f(X) \propto \exp\{-\frac{(X'aX - 2X'b)}{2}\}, \ then$$
$$X \sim MVN(a^{-1}b, a^{-1})$$

1. (a) **Solution.**

The basic idea is that The prior specification for the regression coefficients focuses on observable quantities in that the elicitation is based on the availability of historical data D_0 and a scalar quantity α_0 quantifying the uncertainty in D_0. Then D_0 and α_0 are used to specify a prior for the regression coefficients in a semiautomatic fashion.

To be specific, the power prior distribution of β for the current study can be defined as

$\pi(\theta|D_0, \alpha_0) \propto L(\theta|D_0)^{\alpha_0}\pi_0(\theta|c_0)$,

where c_o is a specified hyperparameter for the initial prior, and α_0 is a scalar prior parameter that weights the historical data relative to the likelihood of the current study.

For our linear regression model, we can construct the posterior distrbution based on the power prior.

Data Model: $Y|X\beta \sim MVN(X\beta, \phi^{-1}I)$ with n observations

History Data Model: $Y_0|X_0\beta \sim MVN(X_0\beta, \phi^{-1}I)$ with n_0 observations

Here we assume β and ϕ is Normal-Gamma Prior Distribution.

initial prior for $\beta|\phi \sim MVN(a, \phi^{-1}R)$

initial prior for $\phi \sim Gamma(a_0, b_0)$

Suppose $\alpha_0 \in [0,1]$ is a scalar. Then the posterior distribution for β, ϕ based on the power prior is

$$\begin{aligned}
P(\beta,\phi|X,Y,X_0,Y_0,\alpha_0) &\propto L(\beta,\phi|X,Y)L(\beta,\phi|X_0,Y_0)^{\alpha_0}\pi_0(\beta|\phi)\pi_0(\phi)\\
&\propto \phi^{\frac{n}{2}}\exp\{-\frac{\phi(Y-X\beta)'(Y-X\beta)}{2}\}\\
&\times (\phi^{\frac{n_0}{2}}\exp\{-\frac{\phi(Y_0-X_0\beta)'(Y_0-X_0\beta)}{2}\})^{\alpha_0}\\
&\times \phi^{\frac{p}{2}}\exp\{-\frac{\phi(\beta-a)'R^{-1}(\beta-a)}{2}\}\\
&\times \phi^{a_0-1}\exp\{-\phi b_0\}
\end{aligned}$$

The conditional posterior distribution of β can be derived as

$$P(\beta|else) \propto \exp\{-\phi(\frac{\alpha_0(Y_0-X_0\beta)'(Y_0-X_0\beta)+(Y-X\beta)'(Y-X\beta)+(\beta-a)'R^{-1}(\beta-a)}{2})\}$$

$$\propto \exp\{-\phi\frac{\beta'(\alpha_0X_0'X_0+R^{-1}+X'X)\beta-2\beta'(\alpha_0X_0'Y_0+X'Y+R^{-1}a)}{2}\}$$

After some algebra, we can find the conditional posterior distribution of β

$\beta \sim MVN((\alpha_0X_0'X_0+X'X+R^{-1})^{-1}(\alpha_0X_0'Y_0+X'Y+R^{-1}a), \phi^{-1}(\alpha_0X_0'X_0+X'X+R^{-1})^{-1})$

And the conditional posterior distribution of ϕ

$\phi \sim Gamma(\frac{n_0\alpha_0+n+p}{2}+a_0, \frac{\alpha_0(Y_0-X_0\beta)'(Y_0-X_0\beta)+(Y-X\beta)'(Y-X\beta)+(\beta-a)'R^{-1}(\beta-a)}{2}+b_0)$

Clearly, when $\alpha_0=0$, we have

$\beta \sim MVN((X'X+R^{-1})^{-1}(X'Y+R^{-1}a), \phi^{-1}(X'X+R^{-1})^{-1})$

$\phi \sim Gamma(\frac{n+p}{2}+a_0, \frac{(Y-X\beta)'(Y-X\beta)+(\beta-a)'R^{-1}(\beta-a)}{2}+b_0)$

we recover our regular posterior distribution for multiple linear regression model discussed in class. when $\alpha_0=1$, it means the history data and current data are weighted equally. Hence α_0 measures the importance of historical data on current estimation of parameters.

□

(b) **Solution.**

The information about G-prior is from https://en.wikipedia.org/wiki/G-prior

Assume the ε_i are iid normal with zero mean and variance ϕ^{-1}. Then the g-prior for β is the multivariate normal distribution with prior mean a, and covariance matrix proportional to $\phi^{-1}(X'X)^{-1}$ i.e., $\beta|\phi \sim MVN(a, g\phi^{-1}(X'X)^{-1})$, where g is a positive scalar parameter.

In G-prior case,

Data Model: $Y|X\beta \sim MVN(X\beta, \phi^{-1}I)$ with n observations

Here we assume β and ϕ is Normal-Gamma Prior Distribution.
prior for $\beta|\phi \sim MVN(a, g\phi^{-1}(X'X)^{-1})$
prior for $\phi \sim Gamma(a_0, b_0)$
Then the posterior distribution for β, ϕ based on G-prior is

$$\begin{aligned} P(\beta,\phi|X,Y) &\propto L(\beta,\phi|X,Y)\pi(\beta|\phi)\pi(\phi) \\ &\propto \phi^{\frac{n}{2}}\exp\{-\frac{\phi(Y-X\beta)'(Y-X\beta)}{2}\} \\ &\times \phi^{\frac{p}{2}}\exp\{-\frac{(\beta-a)'(g\phi^{-1}(X'X)^{-1})^{-1}(\beta-a)}{2}\} \\ &\times \phi^{a_0-1}\exp\{-\phi b_0\} \end{aligned}$$

The conditional posterior distribution for β is

$$\begin{aligned} P(\beta|else) &\propto \exp\{-\frac{\phi(Y-X\beta)'(Y-X\beta)}{2}\}\exp\{-\frac{(\beta-a)'(g\phi^{-1}(X'X)^{-1})^{-1}(\beta-a)}{2}\} \\ &\propto \exp\{-\phi(\frac{(Y-X\beta)'(Y-X\beta)+(\beta-a)'(X'Xg^{-1})(\beta-a)}{2})\} \\ &\propto \exp\{-\phi(\frac{\beta'(X'X+X'Xg^{-1})\beta-2\beta'(X'Y+X'Xg^{-1}a)}{2})\} \end{aligned}$$

After some algebra, we can find the conditional posterior distribution of β
$\beta \sim MVN((X'X+X'Xg^{-1})^{-1}(X'Y+X'Xg^{-1}a), \phi^{-1}(X'X+X'Xg^{-1})^{-1})$
And also the conditional posterior distribution of ϕ
$\phi \sim Gamma(\frac{n+p}{2}+a_0, \frac{(Y-X\beta)'(Y-X\beta)+(\beta-a)'(X'Xg^{-1})(\beta-a)}{2}+b_0)$
Clearly, if $g(X'X)^{-1} = R$, then we have
$\beta \sim MVN((X'X+R^{-1})^{-1}(X'Y+R^{-1}a), \phi^{-1}(X'X+R^{-1})^{-1})$
$\phi \sim Gamma(\frac{n+p}{2}+a_0, \frac{(Y-X\beta)'(Y-X\beta)+(\beta-a)'R^{-1}(\beta-a)}{2}+b_0)$
we recover our regular posterior distribution for multiple linear regression model discussed in class.

□

(c) **Solution.** For Power Prior Distributions, please see Function Gibbs.MLR.Powerprior in the attached file.
for G-prior, please see Gibbs.MLR.Gprior in the attached file

□

(d) **Solution.**

We simulate the data the same as that in "Lecture6.R" from class. However, we simulate a noise history data set. For noise data, we use following parameters:

$\beta_0 = c(-2, -1, -3)$; $\phi_0 = 2$

You can also check the codes for details. Here we try 7 different α_0 values and also 7 different g values, please see the outcomes from the following Table 1

Table 1: Comparison between Different Models

Model	α_0 or g	β_0	β_1	β_2	ϕ
TRUE	NA	2.0000	1.0000	3.0000	1.0000
MLR	NA	1.8741	1.0197	3.2204	0.8988
Power Prior	0.0000	1.8776	1.0186	3.2123	0.8970
Power Prior	0.1000	1.5004	0.8739	2.7059	0.1621
Power Prior	0.2000	1.2427	0.7425	2.2029	0.1045
Power Prior	0.4000	0.7477	0.5398	1.5258	0.0734
Power Prior	0.6000	0.3846	0.3797	1.0010	0.0646
Power Prior	0.8000	0.1313	0.2359	0.5533	0.0614
Power Prior	1.0000	-0.0994	0.1289	0.2109	0.0607
G-Prior	0.1000	0.1731	0.0932	0.2915	0.0639
G-Prior	0.5000	0.6280	0.3420	1.0715	0.0848
G-Prior	1.0000	0.9410	0.5139	1.6025	0.1090
G-Prior	5.0000	1.5671	0.8546	2.6796	0.2636
G-Prior	10.0000	1.7126	0.9223	2.9222	0.3913
G-Prior	50.0000	1.8420	1.0026	3.1583	0.7060
G-Prior	100.0000	1.8588	1.0109	3.1865	0.7943

It shows clearly that, as α_0 increases, the estimates of using power prior becomes worse, that is because as α_0 increases, the noise history data becomes more and more important. However, if the history data is not noise or the history data is very similar to current data, then as α_0 increases, the estimates of power prior should become better.

It also shows clearly that, as g increases, the estimates of using G-prior becomes better, that is because as g increases, we are more close to the outcome of our regular MLR. Indeed, the mean of posterior conditional distribution for *beta* of using G-prior is actually weighted average of maximum likelihood estimator and prior of *beta* (in this case, it is just a)

Ideally, one should also need to specify different priors for α_0 or g, and then also get the posterior distribution of these two.

□

2. (a) **Solution.** Since $b = (b_1, ..., b_n)' \sim N(0, \tau^2(D - \rho W)^{-1})$, the the density for b is

$$P(b) \propto \exp(-\frac{b'(D - \rho W)b}{2\tau^2})$$

Let's focus the numerator,

$$
\begin{aligned}
b'(D-\rho W)b &= (b_1,...,b_n)\begin{bmatrix} D_{11} & -\rho W_{12} & -\rho W_{13} & \dots & -\rho W_{1n} \\ -\rho W_{21} & D_{22} & -\rho W_{23} & \dots & -\rho W_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\rho W_{n1} & -\rho W_{n2} & -\rho W_{n3} & \dots & D_{nn} \end{bmatrix}(b_1,...,b_n)' \\
&= D_{ii}b_i^2 + 2\sum_{k\neq i}^{n}(-\rho W_{ki})b_k b_i + something\ unrelevant\ to\ b_i
\end{aligned}
$$

Hence

$$P(b_i|b_{(-i)}) = \exp(-\frac{D_{ii}b_i^2 + 2\sum_{k\neq i}^{n}(-\rho W_{ki})b_k b_i}{2\tau^2})$$

Recall that $D_{ii} = \sum_{k=1}^{n} W_{ki}$ and $W_{ii} = 0$. Therefore, the conditional distribution of b_i is

$$b_i|b_{(-i)} \sim N(\frac{\sum_{k=1}^{n}\rho W_{ki}b_k}{\sum_{k=1}^{n} W_{ki}}, \frac{\tau^2}{\sum_{k=1}^{n} W_{ki}})$$

□

(b) **Solution.** Since Data Model

$$Y \sim N(\beta_0 + b, \sigma^2)$$

Prior for b is

$$b = (b_1,...,b_n)' \sim N(0, \tau^2(D-\rho W)^{-1})$$

Prior for σ^{-2} is

$$\sigma^{-2} \sim Gamma(a_\sigma, b_\sigma)$$

Prior for τ^{-2} is

$$\tau^{-2} \sim Gamma(a_\tau, b_\tau)$$

Then the posterior distribution is

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

Let's first work on the conditional distribution for β_0

$$\begin{aligned}
\beta_0|else &\propto \exp\{-\frac{(Y-\beta_0-b)'(Y-\beta_0-b)}{2\sigma^2}\} \\
&\propto \exp\{-\frac{n\beta_0^2 - 2\beta_0\sum_{i=1}^n (Y_i - b_i)}{2\sigma^2}\} \\
\beta_0|Y,b,\sigma^{-2} &\sim N(\frac{\sum_{i=1}^n (Y_i-b_i)}{n}, \frac{\sigma^2}{n})
\end{aligned}$$

The conditional distribution for σ^{-2} is

$$\begin{aligned}
\sigma^{-2}|else &\propto (\tau^{-2})^{\frac{n}{2}} \exp\{-\frac{(Y-\beta_0-b)'(Y-\beta_0-b)}{2\sigma^2}\}(\sigma^{-2})^{a_\sigma - 1}\exp\{-\sigma^{-2}b_\sigma\} \\
&\propto (\sigma^{-2})^{\frac{n}{2}+a_\sigma-1}\exp\{-\sigma^{-2}(\frac{(Y-\beta_0-b)'(Y-\beta_0-b)}{2}+b_\sigma)\} \\
\sigma^{-2}|Y,b,\beta_0 &\sim Gamma(\frac{n}{2}+a_\sigma, \frac{(Y-\beta_0-b)'(Y-\beta_0-b)}{2}+b_\sigma)
\end{aligned}$$

The conditional distribution for τ^{-2} is

$$\begin{aligned}
\tau^{-2}|else &\propto (\tau^{-2})^{\frac{n}{2}} \exp\{-\frac{b'(D-\rho W)b}{2\tau^2}\}(\tau^{-2})^{a_\tau-1}\exp\{-\tau^{-2}b_\tau\} \\
&\propto (\tau^{-2})^{\frac{n}{2}+a_\tau-1}\exp\{-\tau^{-2}(\frac{b'(D-\rho W)b}{2}+b_\tau)\} \\
\tau^{-2}|b &\sim Gamma(\frac{n}{2}+a_\tau, \frac{b'(D-\rho W)b}{2}+b_\tau)
\end{aligned}$$

The conditional distribution for b is

$$\begin{aligned}
b|else &\propto \exp\{-\frac{(Y-\beta_0-b)'(Y-\beta_0-b)}{2\sigma^2}\}\exp\{-\frac{b'(D-\rho W)b}{2\tau^2}\} \\
&\propto \exp\{-(\frac{(Y-\beta_0-b)'(Y-\beta_0-b)}{2\sigma^2}+\frac{b'(D-\rho W)b}{2\tau^2})\} \\
&\propto \exp\{-(\frac{(Y^*-b)'(Y^*-b)}{2\sigma^2}+\frac{b'R^*b}{2\sigma^2})\}, where\ Y^* = Y-\beta_0, R^* = \frac{(D-\rho W)\sigma^2}{\tau^2} \\
&\propto \exp\{-(\frac{b'(I+R^*)b - 2b'Y^*}{2\sigma^2})\},
\end{aligned}$$

$$b|Y,\beta_0,\tau^{-2},\sigma^{-2} \sim MVN((I+R^*)^{-1}Y^*, \sigma^2(I+R^*)^{-1}), Hence$$

$$b|Y,\beta_0,\tau^{-2},\sigma^{-2} \sim MVN((I+\frac{(D-\rho W)\sigma^2}{\tau^2})^{-1}(Y-\beta_0), \sigma^2(I+\frac{(D-\rho W)\sigma^2}{\tau^2})^{-1})$$

The conditional distribution for b_i is

$$\begin{aligned}
b_i|else &\propto \exp\{-\frac{(Y_i-\beta_0-b_i)^2}{2\sigma^2}\}\exp\{-\frac{(b_i-\mu_i)^2}{2\tau_i^2}\}, \\
where\ \mu_i &= \frac{\sum_{k=1}^n \rho W_{ki} b_k}{\sum_{k=1}^n W_{ki}}, \tau_i^2 = \frac{\tau^2}{\sum_{k=1}^n W_{ki}} \\
&\propto \exp\{-\frac{(Y_i-\beta_0-b_i)^2 + A(b_i-\mu_i)^2}{2\sigma^2}\}, where\ A = \frac{\sigma^2\sum_{k=1}^n W_{ki}}{\tau^2} \\
&\propto \exp\{-\frac{b_i^2(1+A) - 2(Y_i-\beta_0+A\mu_i)b_i}{2\sigma^2}\} \\
b_i|Y_i,\beta_0,b_{(-i)},\sigma^{-2},\tau^{-2} &\sim N(\frac{Y_i-\beta_0+A\mu_i}{1+A}, \frac{\sigma^2}{1+A}), \\
where\ A &= \frac{\sigma^2\sum_{k=1}^n W_{ki}}{\tau^2}, \mu_i = \frac{\sum_{k=1}^n \rho W_{ki} b_k}{\sum_{k=1}^n W_{ki}}
\end{aligned}$$

□

(c) **Solution.**

We follows the sample order of parameters as the codes in the Lecture Notes

First Algorithms (Updating all of the spacial random effects)

Step 1: Initialize $b^{(0)}$, $\beta_0^{(0)}$, $\sigma^{2(0)}$, $\tau^{2(0)}$. Consider $i = 1$.

Step 2: First Sample for $\beta_0^{(i)}$ from full posterior conditional distribution for β_0

$\beta_0^{(i)} \sim \beta_0^{(i)}|Y, b^{(i-1)}, \sigma^{2(i-1)}$

Step 3: Then sample for $b^{(i)}$ from full posterior conditional distribution for b

$b^{(i)} \sim b^{(i)}|Y, \beta_0^{(i)}, \tau^{2(i-1)}, \sigma^{2(i-1)}$

Step 4: Then sample for $\sigma^{2(i)}$ from full posterior conditional distribution for σ^2

$\sigma^{2(i)} \sim \sigma^{2(i)}|Y, \beta_0^{(i)}, b^{(i)}$

Step 5: Then sample for $\tau^{2(i)}$ from full posterior conditional distribution for τ^2

$\tau^{2(i)} \sim \tau^{2(i)}|Y, b^{(i)}$

Step 6: Increase i by 1 and then repeat Step2-Step5 for enough large iterations

Second Algorithms (Updating the spacial random effects one at a time)

Step 1: Initialize $b^{(0)} = (b_1^{(0)}, b_2^{(0)}, ..., b_n^{(0)})$, $\beta_0^{(0)}$, $\sigma^{2(0)}$, $\tau^{2(0)}$. Consider $i = 1$.

Step 2: First Sample for $\beta_0^{(i)}$ from full posterior conditional distribution for β_0

$\beta_0^{(i)} \sim \beta_0^{(i)}|Y, b^{(i-1)}, \sigma^{2(i-1)}$

Step 3: Then for $j = 1, 2, .., n$, sample for $b_j^{(i)}$ from full posterior conditional distribution for b_j

$b_j^{(i)} \sim b_j^{(i)}|Y, \beta_0^{(i)}, \tau^{2(i-1)}, \sigma^{2(i-1)}, b_1^{(i)}, b_2^{(i)}, ..., b_{j-1}^{(i)}, b_{j+1}^{(i-1)}, ..., b_n^{(i-1)}$

Step 4: Then sample for $\sigma^{2(i)}$ from full posterior conditional distribution for σ^2

$\sigma^{2(i)} \sim \sigma^{2(i)}|Y, \beta_0^{(i)}, b^{(i)}$

Step 5: Then sample for $\tau^{2(i)}$ from full posterior conditional distribution for τ^2

$\tau^{2(i)} \sim \tau^{2(i)}|Y, b^{(i)}$

Step 6: Increase i by 1 and then repeat Step2-Step5 for enough large iterations

$\square$

(d) **Solution.**

For the First Algorithm, we define a function named "fbg1'; for the Second Algorithm, we define a function named "ig1". Moreover, we know the true data and also we simulate a data set with noise.

The outcome using our two algorithms are shown below. In terms of graph, our estimation using either First Algorithm or Second Algorithm, performs very well.

Figure 1: True Graph

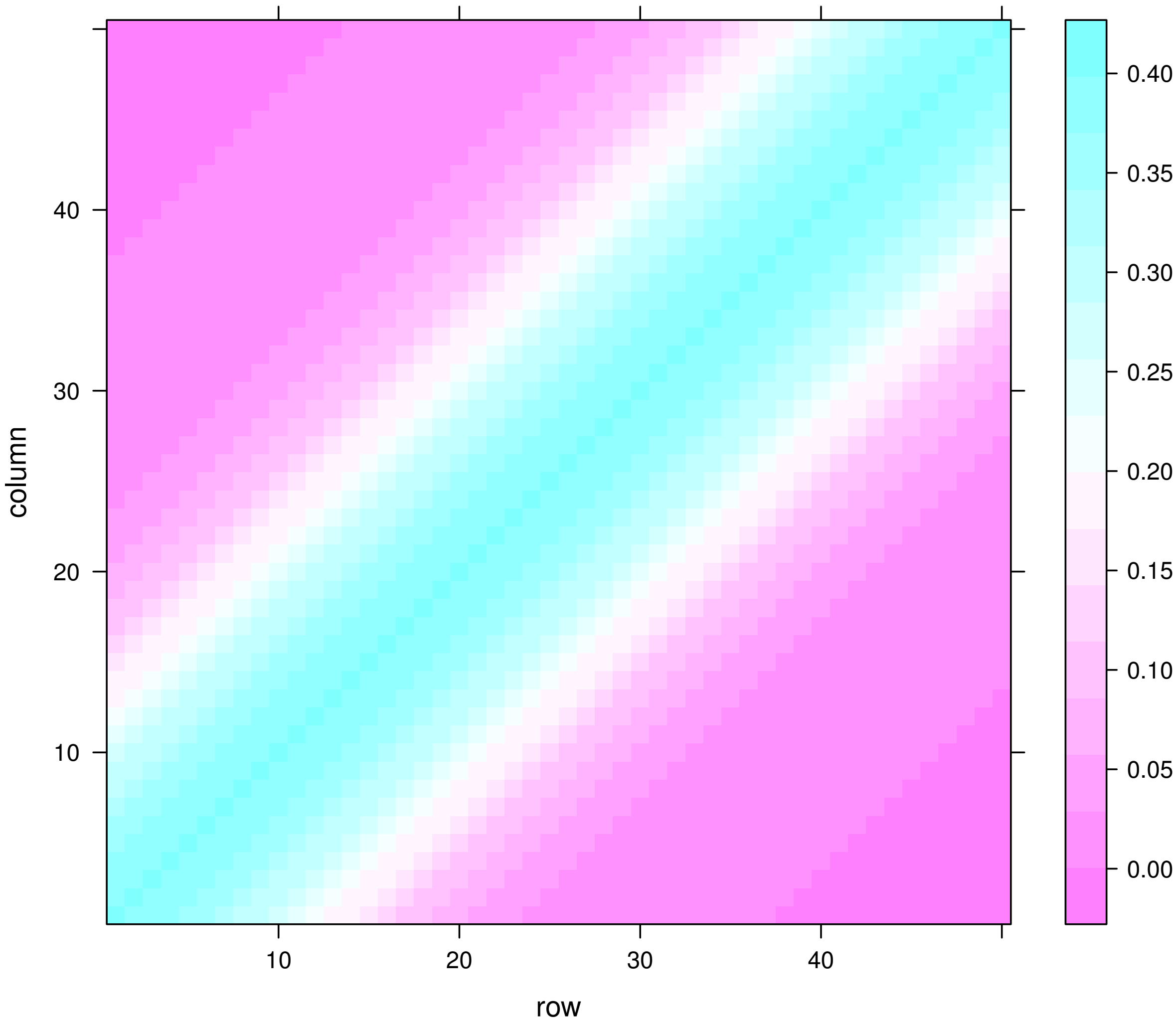


Figure 2: Graph of Our Simulated Data)

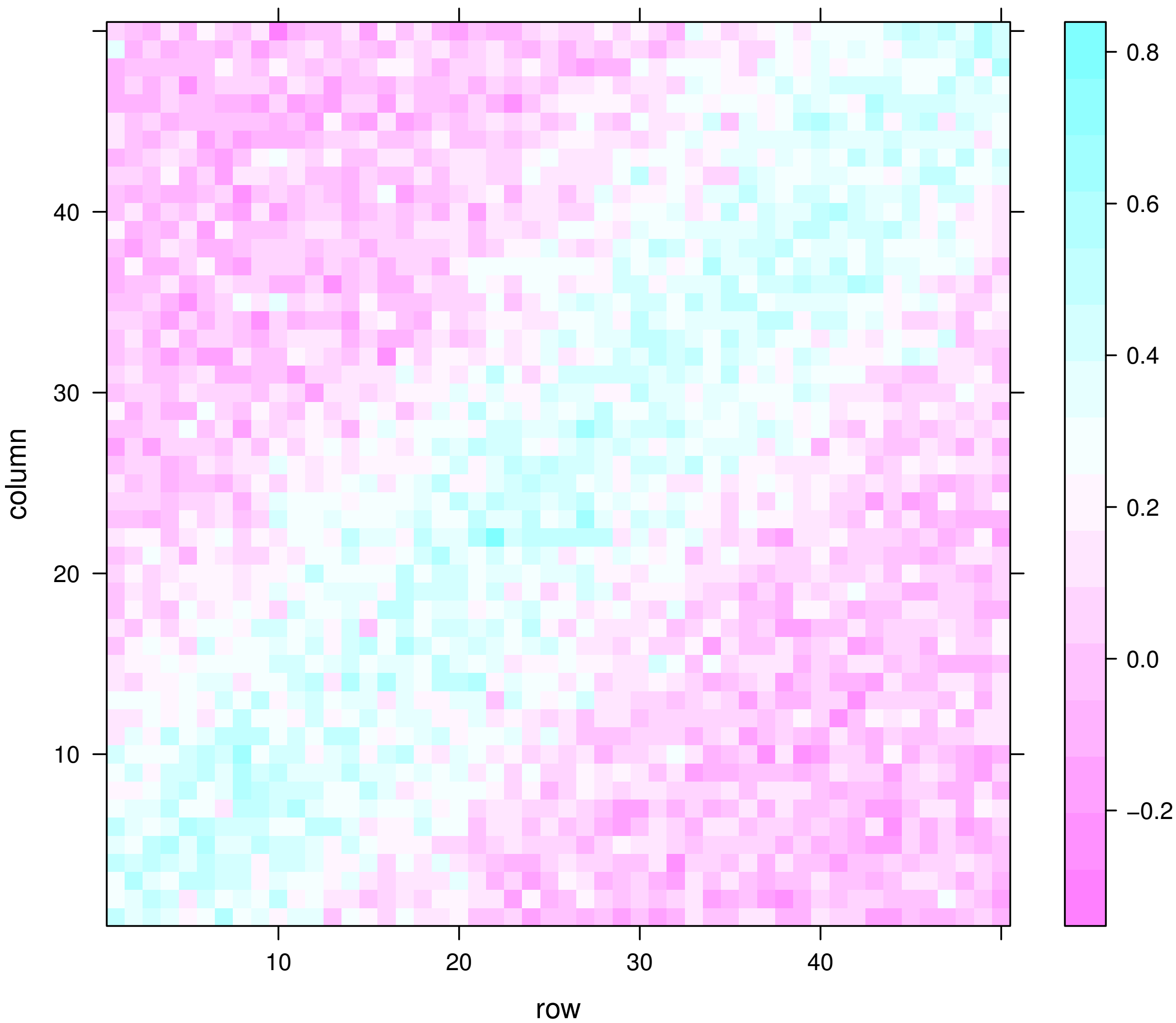


Figure 3: Estimated Graph using Function fbg1 (or the First Algorithm)

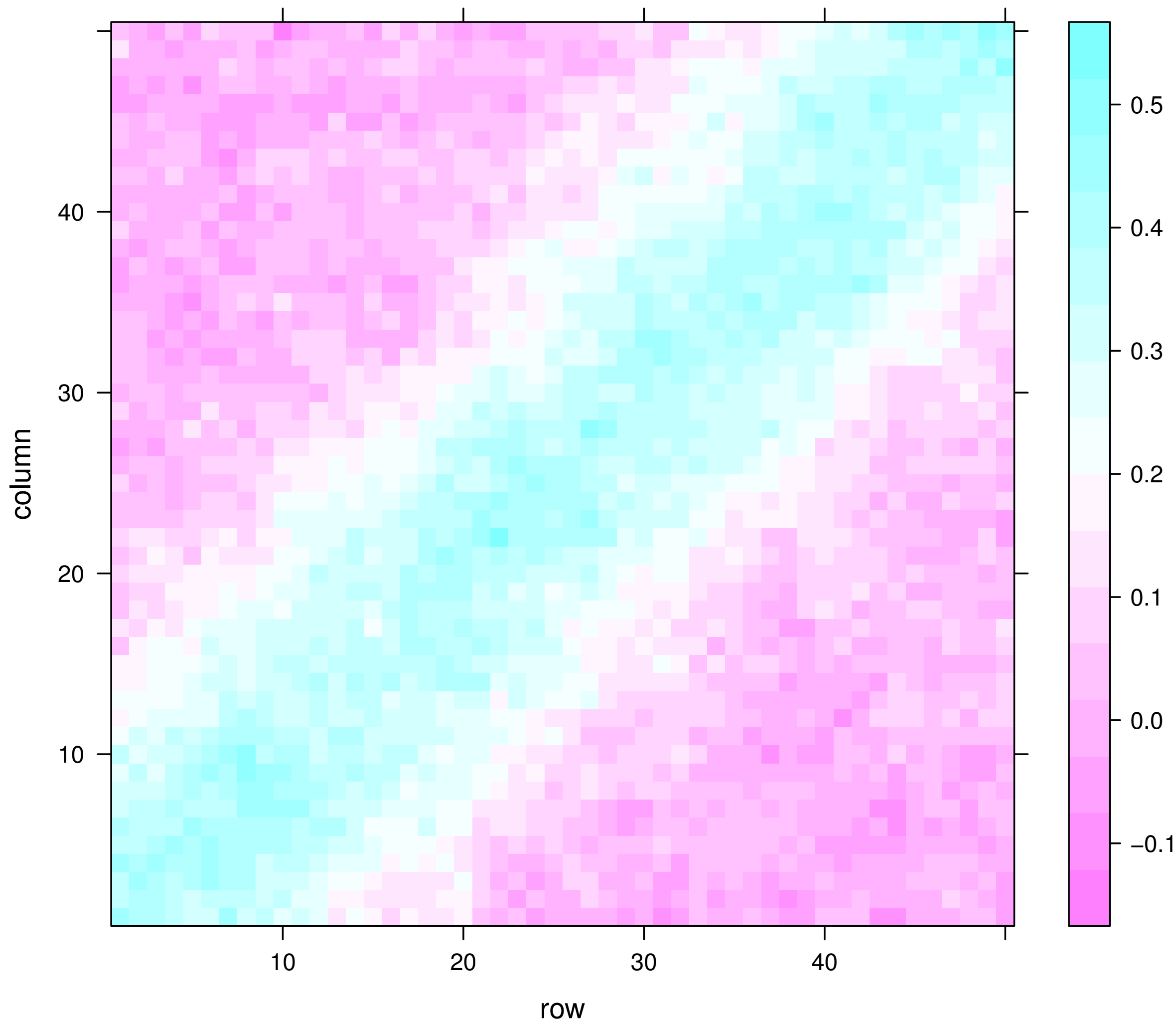


Figure 4: Estimated Graph using Function ig1 (or the Second Algorithm)

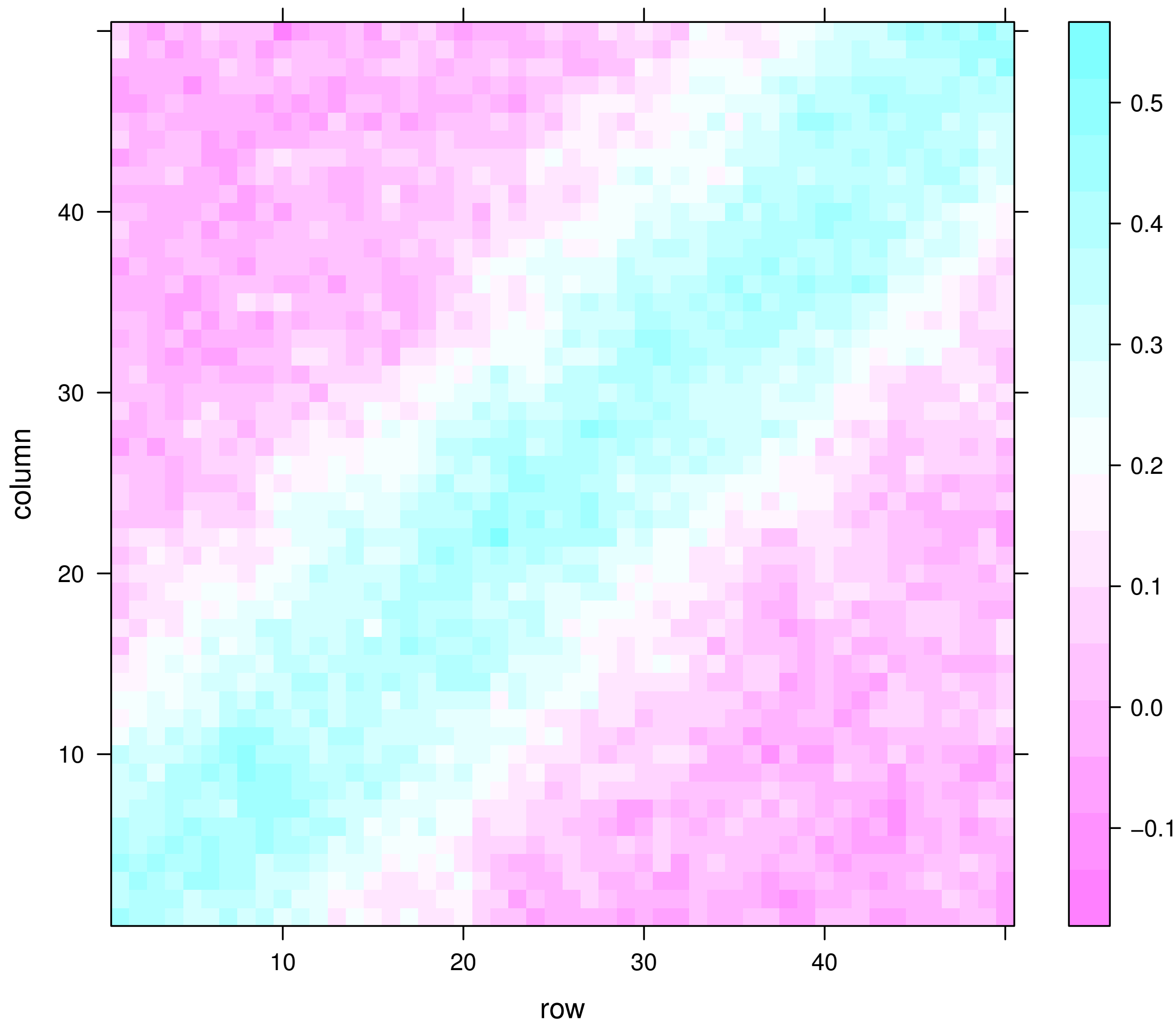
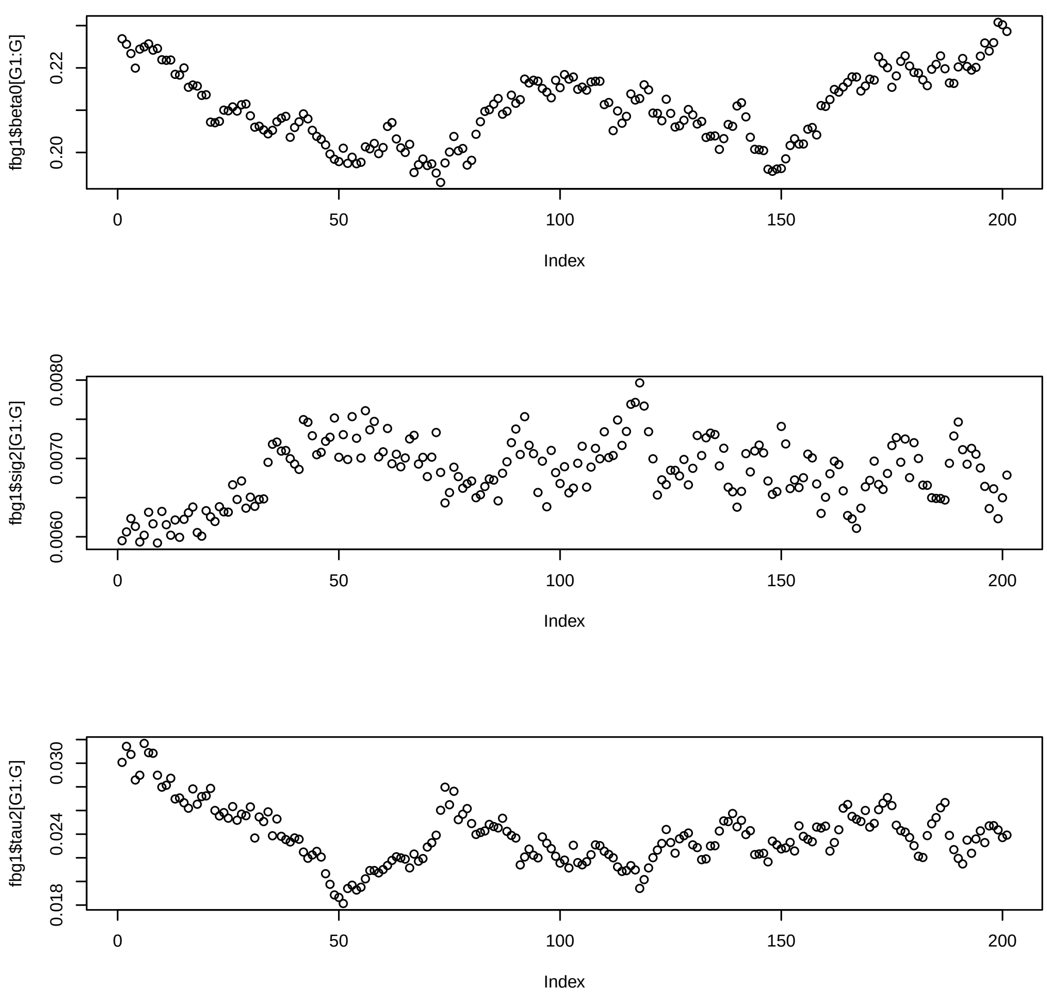


Figure 5: Estimated Chain of Parameters using Function fbg1 (or the First Algorithm)



□

Figure 6: Estimated Chain of Parameters using Function ig1 (or the Second Algorithm)