Supplement for Bayesian Inference for Finite Mixture Regression Model Based on Non-iterative Algorithm

1 Conditional distributions in IBF with non-informative prior

2. Conditional distribution $\pi(\boldsymbol{\beta}_j|\sigma_j^2, \mathbf{y}, \mathbf{G})$

The conditional density of $\boldsymbol{\beta}_j$ is given by

$$\pi(\boldsymbol{\beta}_j|\sigma_j^2, \mathbf{y}, \mathbf{G}) \propto \exp\bigg\{ -\frac{\sum_{i=1}^n G_{ij}(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j)^2}{2\sigma_j^2} \bigg\}.$$

Let

$$\mathbf{G}_{(j)} = (G_{1j}, \cdots, G_{nj}), \quad \mathbf{W}_j = \mathrm{diag}(\mathbf{G}_{(j)}), \quad \mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_{p-1})^{\mathrm{T}},$$

and

$$\mathbf{B}_j^{-1} = \mathbf{X}^{\mathrm{T}}\mathbf{W}_j\mathbf{X}, \quad \mathbf{b}_j = \mathbf{B}_j\mathbf{X}^{\mathrm{T}}\mathbf{W}_j\mathbf{y},$$

then

$$\begin{aligned}\sum_{i=1}^n G_{ij}(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j)^2 &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_j)^{\mathrm{T}}\mathbf{W}_j(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_j) \\ &= \boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{W}_j\mathbf{X}\boldsymbol{\beta}_j - 2\boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{W}_j\mathbf{y} + \mathbf{y}^{\mathrm{T}}\mathbf{W}_j\mathbf{y} \\ &= (\boldsymbol{\beta}_j - \mathbf{b}_j)^{\mathrm{T}}\mathbf{B}_j^{-1}(\boldsymbol{\beta}_j - \mathbf{b}_j) + \mathbf{y}^{\mathrm{T}}\mathbf{W}_j\mathbf{y} - \mathbf{b}_j^{\mathrm{T}}\mathbf{B}_j^{-1}\mathbf{b}_j\end{aligned}$$

therefore

$$\pi(\boldsymbol{\beta}_j|\sigma_j^2, \mathbf{y}, \mathbf{G}_{(j)}) \propto \exp\bigg\{ -\frac{(\boldsymbol{\beta}_j - \mathbf{b}_j)^{\mathrm{T}}\mathbf{B}_j^{-1}(\boldsymbol{\beta}_j - \mathbf{b}_j)}{2\sigma_j^2} \bigg\}, \tag{1}$$

the right side of Equation (1) is the kernel of Mutivariate Normal Disitrbution with mean vector $\mathbf{b}_j$ and covariance matrix $\sigma_j^2\mathbf{B}_j$, so

$$\boldsymbol{\beta}_j|\sigma_j^2, \mathbf{y}, \mathbf{G}_{(j)} \sim \mathbb{N}(\mathbf{b}_j, \sigma_j^2\mathbf{B}_j).$$

3. Conditional distribution $\pi(\sigma_j^2|\mathbf{y}, \mathbf{G}_{(j)})$

The joint distribution of $\boldsymbol{\beta}_j$ and σ_j^2 conditional on $(\mathbf{y}, \mathbf{G}_{(j)})$ is that

$$\begin{aligned}\pi(\boldsymbol{\beta}_j, \sigma_j^2 | \mathbf{y}, \mathbf{G}_{(j)}) &\propto \prod_{i=1}^{n} \left(\lambda_j \frac{1}{\sqrt{2\pi}\sigma_j} \exp\{-\frac{(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j)^2}{2\sigma_j^2}\} \right)^{G_{ij}} \frac{1}{\sigma_j^2} \\ &\propto \left(\frac{1}{\sigma_j^2}\right)^{\frac{\sum_{i=1}^n G_{ij}}{2}+1} \exp\left\{ -\frac{\sum_{i=1}^n G_{ij}(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j)^2}{2\sigma_j^2} \right\} \\ &= \left(\frac{1}{\sigma_j^2}\right)^{\frac{\sum_{i=1}^n G_{ij}}{2}+1} \exp\left\{ -\frac{(\boldsymbol{\beta}_j - \mathbf{b}_j)^{\mathrm{T}}\mathbf{B}_j^{-1}(\boldsymbol{\beta}_j - \mathbf{b}_j) + \mathbf{y}^{\mathrm{T}}\mathbf{W}_j\mathbf{y} - \mathbf{b}_j^{\mathrm{T}}\mathbf{B}_j^{-1}\mathbf{b}_j}{2\sigma_j^2} \right\}.\end{aligned}$$

We integrate $\boldsymbol{\beta}_j$ from the about equation and get

$$\begin{aligned}\pi(\sigma_j^2 | \mathbf{y}, \mathbf{G}_{(j)}) &\propto \left(\frac{1}{\sigma_j^2}\right)^{\frac{\sum_{i=1}^n G_{ij}-p}{2}+1} \exp\left\{ -\frac{\mathbf{y}^{\mathrm{T}}\mathbf{W}_j\mathbf{y} - \mathbf{b}_j^{\mathrm{T}}\mathbf{B}_j^{-1}\mathbf{b}_j}{2\sigma_j^2} \right\} \\ &= \left(\frac{1}{\sigma_j^2}\right)^{\frac{\sum_{i=1}^n G_{ij}-p}{2}+1} \exp\left\{ -\frac{\sum_{i=1}^n G_{ij}(y_i - \mathbf{x}_i^{\mathrm{T}}\mathbf{b}_j)^2}{2\sigma_j^2} \right\},\end{aligned} \tag{2}$$

due to the fact

$$\begin{aligned}\sum_{i=1}^{n} G_{ij}(y_i - \mathbf{x}_i^{\mathrm{T}}\mathbf{b}_j)^2 &= (\mathbf{y} - \mathbf{X}\mathbf{b}_j)^{\mathrm{T}}\mathbf{W}_j(\mathbf{y} - \mathbf{X}\mathbf{b}_j) \\ &= \mathbf{b}_j^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{W}_j\mathbf{X}\mathbf{b}_j - 2\mathbf{b}_j^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{W}_j\mathbf{y} + \mathbf{y}^{\mathrm{T}}\mathbf{W}_j\mathbf{y} \\ &= \mathbf{y}^{\mathrm{T}}\mathbf{W}_j\mathbf{y} - \mathbf{b}_j^{\mathrm{T}}\mathbf{B}_j^{-1}\mathbf{b}_j.\end{aligned}$$

Noticing that the right side of Equation (2) is the kernel of inverse-Gamma density, thus

$$\sigma_j^2 | \mathbf{y}, \mathbf{G}_{(j)} \sim \mathrm{IG}\left(\frac{\sum_{i=1}^n G_{ij} - p}{2}, \frac{\sum_{i=1}^n G_{ij}(y_i - \mathbf{x}_i^{\mathrm{T}}\mathbf{b}_j)^2}{2} \right).$$

2 IBF algorithm with conjugate prior

2.1 The conjugate prior

The prior of $\boldsymbol{\beta}_j$ and σ_j^2 are given as the Normal-Inverse Gamma prior,

$$\boldsymbol{\beta}_j | \sigma_j^2 \sim N_p(\mathbf{b}, \sigma_j^2\mathbf{B}), \quad \sigma_j^2 \sim \mathrm{IG}(c/2, d/2),$$

where $\mathbf{b}_{p\times 1}$ is prior mean vector, $\mathbf{B}_{p\times p}$ is a positive definite matrix, and $\mathrm{IG}(c/2, d/2)$ denotes the Inverse-Gamma distribution with parameters $c > 0,\ d > 0$. Note that when $c \to 0,\ d \to 0$, $\pi(\sigma_j^2) \propto 1/\sigma_j^2$, thus we get a non-informative prior for σ_j^2. In the following derivation, we will show that the conditional joint posterior of $\boldsymbol{\beta}_j$ and σ_j^2 also follows the Normal-Inverse Gamma distribution, therefore Normal-Inverse Gamma distribution is the conjugate prior for $\boldsymbol{\beta}_j$ and σ_j^2.

The mixing probabilities $\boldsymbol{\lambda}$ are given a Dirichlet prior with concentration α:

$$(\lambda_1, \cdots, \lambda_g) \sim \mathcal{D}(\alpha, \cdots, \alpha)$$

where $\mathcal{D}(\alpha, \cdots, \alpha)$ stands for the Dirichlet distribution on the simplex $\{(\lambda_1, \cdots, \lambda_g) : \sum_{j=1}^{g} \lambda_g \leq 1\}$ with a density function proportional to

$$\lambda_1^{\alpha-1} \cdots \lambda_{g-1}^{\alpha-1}(1 - \lambda_1 - \cdots \lambda_{g-1})^{\alpha-1}.$$

Note that the Dirichlet distribution is the natural conjugate distribution for the mixture proportions. Notice that the choice of $\alpha = 1$ gives a uniform prior over the space $\lambda_1 + \cdots + \lambda_g = 1$.

We suppose all hyper-parameters are known, then given complete data $(\mathbf{y}, \mathbf{G})$, the complete posterior density of $\boldsymbol{\theta}$ is shown as

$$\pi(\boldsymbol{\theta}|\mathbf{y}, \mathbf{G}) \propto L(\boldsymbol{\theta}|\mathbf{y}, \mathbf{G}) \prod_{j=1}^{g} \pi(\boldsymbol{\beta}_j|\sigma_j^2)\pi(\sigma_j^2)\pi(\boldsymbol{\lambda}). \tag{3}$$

2.2 Conditional distributions

In order to implement the IBF sampling in fitting FMNR model, we now present some required conditional posterior distributions, whose densities are all proportional to the joint posterior density $\pi(\boldsymbol{\theta}|\mathbf{y}, \mathbf{G})$ in Equation (3).

1. Conditional distribution $\pi(\mathbf{G}_i|\boldsymbol{\theta}, \mathbf{y})$

For $i = 1, \cdots, n$, the conditional posterior distribution of $\mathbf{G}_i = (G_{i1}, \cdots, G_{ig})^{\mathrm{T}}$ is the same as in the non-informative prior situation, that is

$$\pi(G_{i1}, \cdots, G_{ig}|\boldsymbol{\theta}, \mathbf{y}) \propto \prod_{j=1}^{g} (\lambda_j \mathrm{N}(y_i; \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j, \sigma_j^2))^{G_{ij}},$$

thus we obtain that

$$\mathbf{G}_i|\boldsymbol{\theta}, \mathbf{y} \sim \mathrm{Mult}(1, r_{i1}, \cdots, r_{ig}), \tag{4}$$

where

$$r_{ij} = \frac{\lambda_j \mathrm{N}(y_i; \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j, \sigma_j^2)}{\sum_{k=1}^{g} \lambda_k \mathrm{N}(y_i; \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_k, \sigma_k^2)}, \tag{5}$$

2. Conditional distribution $\pi(\boldsymbol{\beta}_j|\sigma_j^2, \mathbf{y}, \mathbf{G})$

The conditional density of $\boldsymbol{\beta}_j$ is given by

$$\pi(\boldsymbol{\beta}_j|\sigma_j^2, \mathbf{y}, \mathbf{G}) \propto \exp\{-\frac{\sum_{i=1}^n G_{ij}(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j)^2}{2\sigma_j^2}\}\exp\{-\frac{(\boldsymbol{\beta}_j - \mathbf{b})^{\mathrm{T}}\mathbf{B}^{-1}(\boldsymbol{\beta}_j - \mathbf{b})}{2\sigma_j^2}\}.$$

Denote $\mathbf{G}_{(j)} = (G_{1j}, \cdots, G_{nj})$, $\mathbf{W}_j = \mathrm{diag}(\mathbf{G}_{(j)})$, and

$$\mathbf{B}_j^{-1} = \mathbf{X}^{\mathrm{T}}\mathbf{W}_j\mathbf{X} + \mathbf{B}^{-1}, \mathbf{b}_j = \mathbf{B}_j(\mathbf{X}^{\mathrm{T}}\mathbf{W}_j\mathbf{y} + \mathbf{B}^{-1}\mathbf{b}), \tag{6}$$

then we have

$$\begin{aligned}
&\sum_{i=1}^n G_{ij}(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j)^2 + (\boldsymbol{\beta}_j - \mathbf{b})^{\mathrm{T}}\mathbf{B}^{-1}(\boldsymbol{\beta}_j - \mathbf{b}) \\
&= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_j)^{\mathrm{T}}\mathbf{W}_j(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_j) + (\boldsymbol{\beta}_j - \mathbf{b})^{\mathrm{T}}\mathbf{B}^{-1}(\boldsymbol{\beta}_j - \mathbf{b}) \\
&= \boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{W}_j\mathbf{X}\boldsymbol{\beta}_j - 2\boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{W}_j\mathbf{y} + \mathbf{y}^{\mathrm{T}}\mathbf{W}_j\mathbf{y} + \boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{B}^{-1}\boldsymbol{\beta}_j - 2\boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{b} + \mathbf{b}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{b} \\
&= \boldsymbol{\beta}_j^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{W}_j\mathbf{X} + \mathbf{B}^{-1})\boldsymbol{\beta}_j - 2\boldsymbol{\beta}_j^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{W}_j\mathbf{y} + \mathbf{B}^{-1}\mathbf{b}) + \mathbf{y}^{\mathrm{T}}\mathbf{W}_j\mathbf{y} + \mathbf{b}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{b} \\
&= (\boldsymbol{\beta}_j - \mathbf{b}_j)^{\mathrm{T}}\mathbf{B}_j^{-1}(\boldsymbol{\beta}_j - \mathbf{b}_j) + \mathbf{y}^{\mathrm{T}}\mathbf{W}_j\mathbf{y} + \mathbf{b}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{b} - \mathbf{b}_j^{\mathrm{T}}\mathbf{B}_j^{-1}\mathbf{b}_j
\end{aligned}$$

consequently,

$$\pi(\boldsymbol{\beta}_j|\sigma_j^2, \mathbf{y}, \mathbf{G}_{(j)}) \propto \exp\Big\{-\frac{(\boldsymbol{\beta}_j - \mathbf{b}_j)^{\mathrm{T}}\mathbf{B}_j^{-1}(\boldsymbol{\beta}_j - \mathbf{b}_j)}{2\sigma_j^2}\Big\}, \tag{7}$$

thus

$$\boldsymbol{\beta}_j|\sigma_j^2, \mathbf{y}, \mathbf{G}_{(j)} \sim \mathbb{N}(\mathbf{b}_j, \sigma_j^2\mathbf{B}_j). \tag{8}$$

3. Conditional distribution $\pi(\sigma_j^2|\mathbf{y}, \mathbf{G}_{(j)})$ The joint distribution of $\boldsymbol{\beta}_j$ and σ_j^2 conditional on $(\mathbf{y}, \mathbf{G}_{(j)})$ is that

$$\begin{aligned}
&\pi(\boldsymbol{\beta}_j, \sigma_j^2|\mathbf{y}, \mathbf{G}_{(j)}) \\
&\propto \prod_{i=1}^n \left(\lambda_j \frac{1}{\sqrt{2\pi}\sigma_j}\exp\{-\frac{(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j)^2}{2\sigma_j^2}\}\right)^{G_{ij}} \pi(\boldsymbol{\beta}_j|\sigma_j^2)\pi(\sigma_j^2) \\
&\propto \left(\frac{1}{\sigma_j^2}\right)^{\frac{\sum_{i=1}^n G_{ij}}{2}+1} \exp\Big\{-\frac{\sum_{i=1}^n G_{ij}(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j)^2}{2\sigma_j^2}\Big\}\times \\
&\left(\frac{1}{\sigma_j^2}\right)^{\frac{p}{2}} \exp\Big\{-\frac{(\boldsymbol{\beta}_j - \mathbf{b})^{\mathrm{T}}\mathbf{B}^{-1}(\boldsymbol{\beta}_j - \mathbf{b})}{2\sigma_j^2}\Big\}\left(\frac{1}{\sigma_j^2}\right)^{\frac{c}{2}} \exp\Big\{-\frac{d}{2\sigma_j^2}\Big\} \\
&= \left(\frac{1}{\sigma_j^2}\right)^{\frac{\sum_{i=1}^n G_{ij}+p+c}{2}+1} \exp\Big\{-\frac{\sum_{i=1}^n G_{ij}(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j)^2 + (\boldsymbol{\beta}_j - \mathbf{b})^{\mathrm{T}}\mathbf{B}^{-1}(\boldsymbol{\beta}_j - \mathbf{b}) + d}{2\sigma_j^2}\Big\} \\
&= \left(\frac{1}{\sigma_j^2}\right)^{\frac{\sum_{i=1}^n G_{ij}+p+c}{2}+1} \exp\Big\{-\frac{(\boldsymbol{\beta}_j - \mathbf{b}_j)^{\mathrm{T}}\mathbf{B}_j^{-1}(\boldsymbol{\beta}_j - \mathbf{b}_j) + \mathbf{y}^{\mathrm{T}}\mathbf{W}_j\mathbf{y} + \mathbf{b}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{b} - \mathbf{b}_j^{\mathrm{T}}\mathbf{B}_j^{-1}\mathbf{b}_j + d}{2\sigma_j^2}\Big\}.
\end{aligned}$$

We integrate $\boldsymbol{\beta}_j$ from the about equation and get

$$\pi(\boldsymbol{\beta}_j, \sigma_j^2|\mathbf{y}, \mathbf{G}_{(j)}) \propto \left(\frac{1}{\sigma_j^2}\right)^{\frac{\sum_{i=1}^n G_{ij}+c}{2}+1} \exp\left\{-\frac{\mathbf{y}^{\mathrm{T}}\mathbf{W}_j\mathbf{y} + \mathbf{b}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{b} - \mathbf{b}_j^{\mathrm{T}}\mathbf{B}_j^{-1}\mathbf{b}_j + d}{2\sigma_j^2}\right\}$$
$$= (\frac{1}{\sigma_j^2})^{\frac{\sum_{i=1}^n G_{ij}+c}{2}+1} \exp\left\{-\frac{\sum_{i=1}^n G_{ij}(y_i - \mathbf{x}_i^{\mathrm{T}}\mathbf{b}_j)^2 + (\mathbf{b}_j - \mathbf{b})^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{b}_j - \mathbf{b}) + d}{2\sigma_j^2}\right\},$$

noticing that the right side of Equation (11) is the kernel of inverse-Gamma density, we have

$$\sigma_j^2|\mathbf{y}, \mathbf{G}_{(j)} \sim \mathrm{IG}(\frac{\sum_{i=1}^n G_{ij} + c}{2}, \frac{\sum_{i=1}^n G_{ij}(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{b}_j)^2 + (\mathbf{b}_j - \mathbf{b})^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{b}_j - \mathbf{b}) + d}{2}). \tag{9}$$

4. Conditional distribution $\pi(\boldsymbol{\lambda}|\mathbf{G})$

Finally, we get the conditional density of $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_g)^{\mathrm{T}}$,

$$\pi(\lambda_1, \cdots, \lambda_g|\mathbf{G}) \propto \prod_{j=1}^{g} \lambda_j^{\sum_{i=1}^n G_{ij}+\alpha-1}, \tag{10}$$

the left side of (10) is the kernel of Dirichlet distribution, therefore

$$\boldsymbol{\lambda}|\mathbf{G} \sim \mathcal{D}(\sum_{i=1}^n G_{i1} + \alpha, \cdots, \sum_{i=1}^n G_{ig} + \alpha). \tag{11}$$

2.3 IBF sampler

The IBF algorithm is similar to the one in non-informative case, denoting $\boldsymbol{\theta}_0$ is the posterior mode of $\pi(\boldsymbol{\theta}|\mathbf{y}, \mathbf{G})$ obtained by EM algorithm, the IBF algorithm includes the following four steps.

Step 1. Draw i.i.d. samples $\mathbf{G}_i^{(l)} = (G_{i1}^{(l)}, \cdots, G_{ig}^{(l)})^{\mathrm{T}} \sim \mathrm{M}(r_{i1}^0, \cdots, r_{ig}^0)$, $l = 1, \cdots, L$, for $i = 1, 2, \cdots, n$ based on (4), where r_{ij}^0 is the same as r_{ij} in (5), with $\boldsymbol{\theta}$ replaced by $\boldsymbol{\theta}_0$. Denote $\mathbf{G}^{(l)} = (\mathbf{G}_1^{(l)}, \mathbf{G}_2^{(l)}, \cdots, \mathbf{G}_n^{(l)})$.

Step 2. Calculate the weights

$$\omega_l = \frac{\frac{1}{\pi(\boldsymbol{\theta}_0|\mathbf{y},\mathbf{G}^{(l)})}}{\sum_{l=1}^L \frac{1}{\pi(\boldsymbol{\theta}_0|\mathbf{y},\mathbf{G}^{(l)})}}, \quad l = 1, 2 \cdots, L,$$

where $\pi(\boldsymbol{\theta}_0|\mathbf{y}, \mathbf{G}^{(l)})$ is calculated as $\pi(\boldsymbol{\theta}|\mathbf{y}, \mathbf{G})$ with $\boldsymbol{\theta}$ and $\mathbf{G}$ replaced by $\boldsymbol{\theta}_0$ and $\mathbf{G}^{(l)}$.

Step 3. Choose a subset from $\{\mathbf{G}^{(l)}\}_{l=1}^L$ via resampling without replacement from the discrete distribution on $\{\mathbf{G}^{(l)}\}_{l=1}^L$ with probabilities $\{\omega_l\}_{l=1}^L$ to obtain an i.i.d. sample of size $K(< L)$ approximately from $\pi(\mathbf{G}|\mathbf{y})$, denoted by $\{\mathbf{G}^{(l_k)}\}_{k=1}^K$.

Step 4. Based on (11), generate

$$\boldsymbol{\lambda}^{(l_k)}|\mathbf{y}, \mathbf{G}^{(l_k)} \sim \mathcal{D}(\sum_{i=1}^n G_{i1}^{(l_k)} + \alpha, \cdots, \sum_{i=1}^n G_{ig}^{(l_k)} + \alpha), \quad k = 1, \cdots, K.$$

For $j = 1, \cdots, g$, based on (9), generate

$$\sigma_j^{2(l_k)}|\mathbf{y}, \mathbf{G}_{(j)}^{(l_k)} \sim \text{IG}\Big(\frac{\sum_{i=1}^n G_{ij}^{(l_k)} + c}{2}, \frac{\sum_{i=1}^n G_{ij}^{(0)}(y_i - \mathbf{x}_i^{\text{T}}\mathbf{b}_j^{(l_k)})^2 + (\mathbf{b}_j^{(l_k)} - \mathbf{b})^{\text{T}}\mathbf{B}^{-1}(\mathbf{b}_j^{(l_k)} - \mathbf{b}) + d}{2}\Big),$$

and based on (8), generate

$$\boldsymbol{\beta}_j^{(l_k)}|\sigma_j^{2(l_k)}, \mathbf{y}, \mathbf{G}_{(j)}^{(l_k)} \sim \mathbb{N}\big(\mathbf{b}_j^{(l_k)}, \sigma_j^{2(l_k)}\mathbf{B}_j^{(l_k)}\big),$$

where

$$\mathbf{G}_{(j)}^{(l_k)} = (G_{1j}^{(l_k)}, \cdots, G_{nj}^{(l_k)}), \quad \mathbf{W}_j^{(l_k)} = \text{diag}(\mathbf{G}_{(j)}^{(l_k)}),$$

and

$$\mathbf{b}_j^{(l_k)} = \mathbf{B}_j^{(l_k)}(\mathbf{X}^{\text{T}}\mathbf{W}_j^{(l_k)}\mathbf{y} + \mathbf{B}^{-1}\mathbf{b}), \quad \mathbf{B}_j^{(l_k)} = \big(\mathbf{X}^{\text{T}}\mathbf{W}_j^{(l_k)}\mathbf{X} + \mathbf{B}^{-1}\big)^{-1}.$$

Finally, $\{(\boldsymbol{\beta}^{(l_k)}, \boldsymbol{\sigma}^{\mathbf{2}(l_k)}, \boldsymbol{\lambda}^{(l_k)})\}_{k=1}^K$ are i.i.d. samples from $\pi(\boldsymbol{\beta}, \boldsymbol{\sigma}^{\mathbf{2}}, \boldsymbol{\lambda}|\mathbf{y})$.

3 Figures

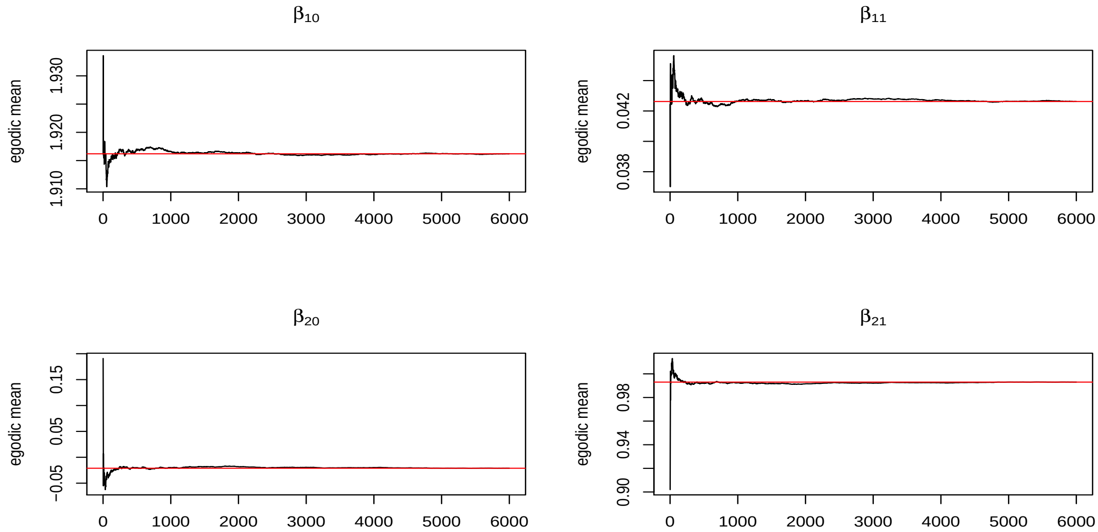

Figure S1: Ergodic mean of 6000 posterior samples of regression coefficients based on IBF, with the red line denoting the posterior mean.

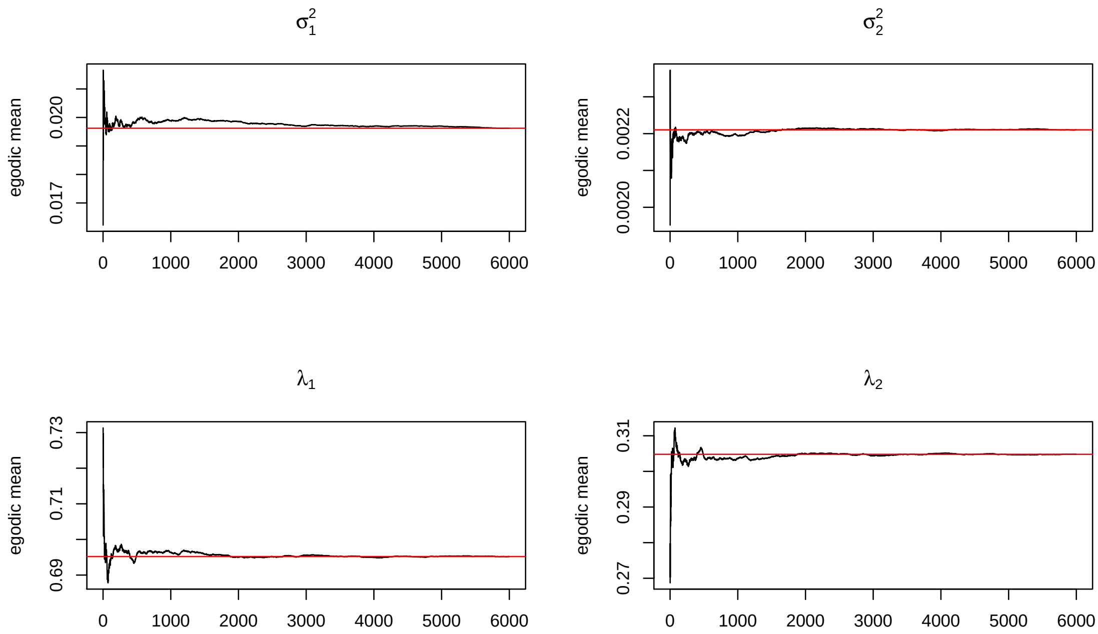

Figure S2: Ergodic mean of 6000 posterior samples of variances and group probabilities based on IBF, with the red line denoting the posterior mean..

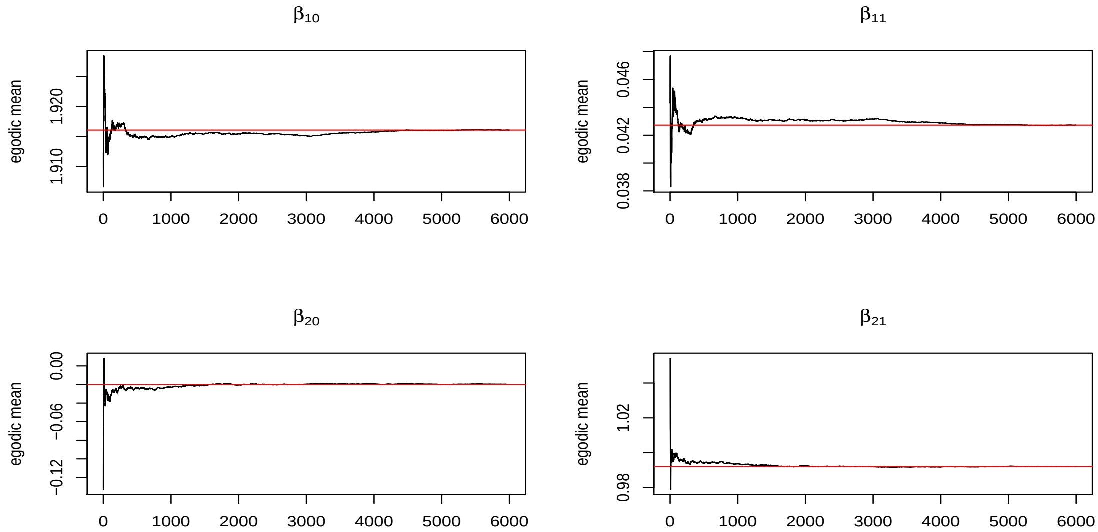

Figure S3: Ergodic mean of 6000 posterior samples of regression coefficients based on Gibbs sampling, with the red line denoting the posterior mean.

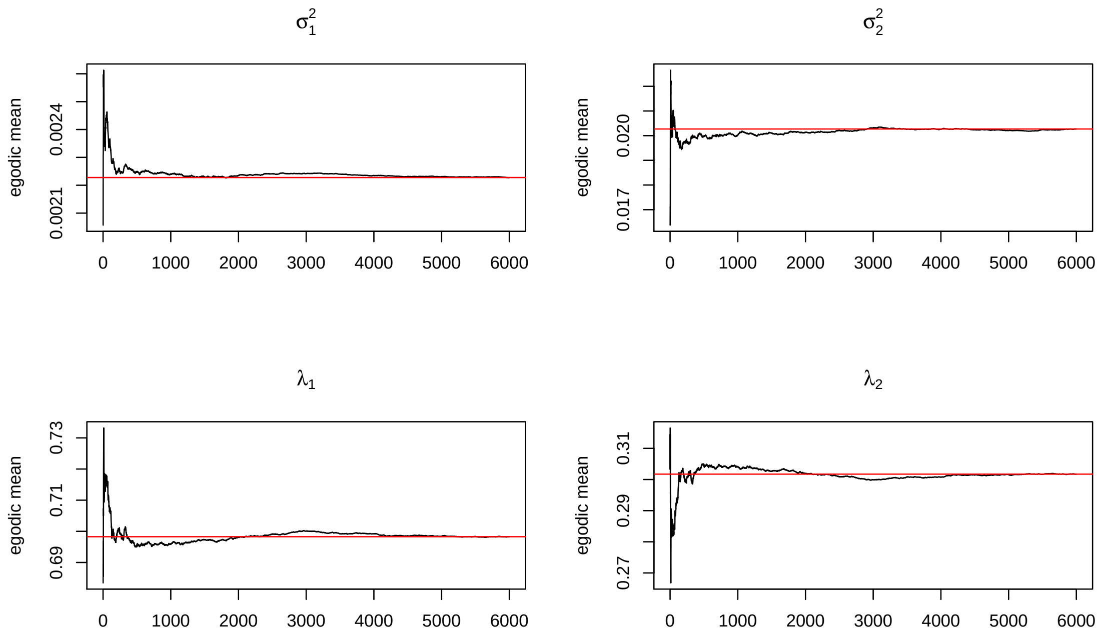

Figure S4: Ergodic mean of 6000 posterior samples of variances and group probabilities based on Gibbs sampling, with the red line denoting the posterior mean.

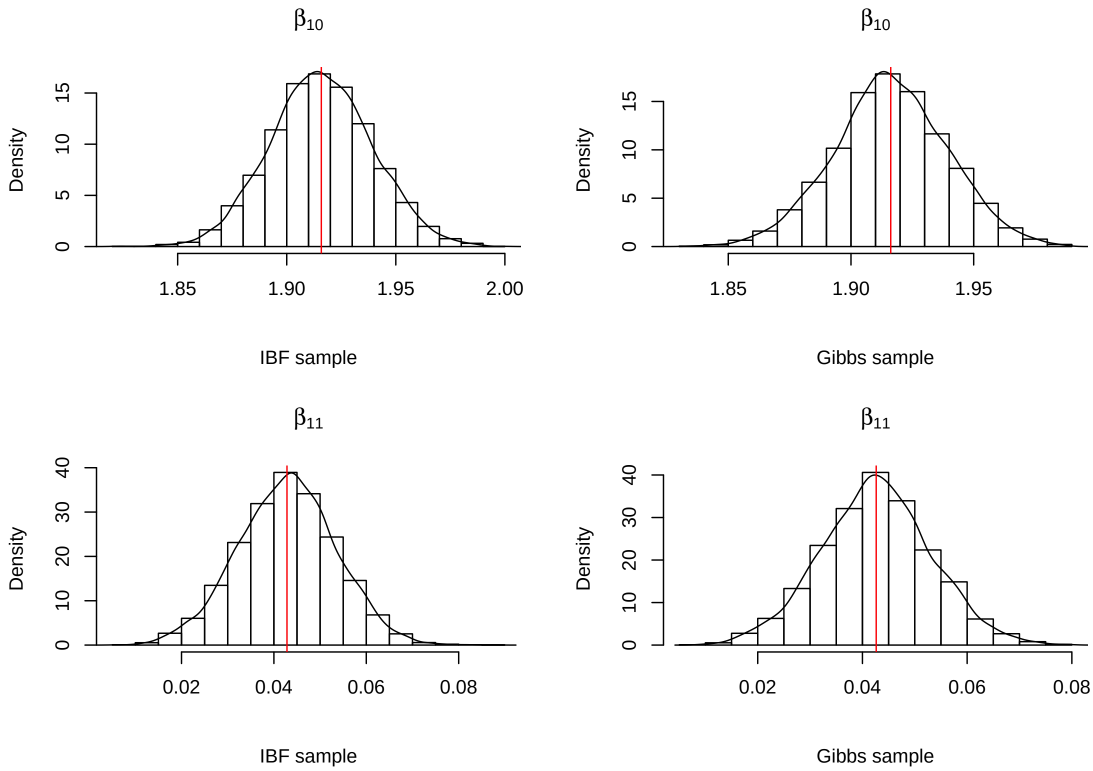

Figure S5: Density histograms of posterior samples for β_{10} and β_{11} based on IBF samples and Gibbs samples.

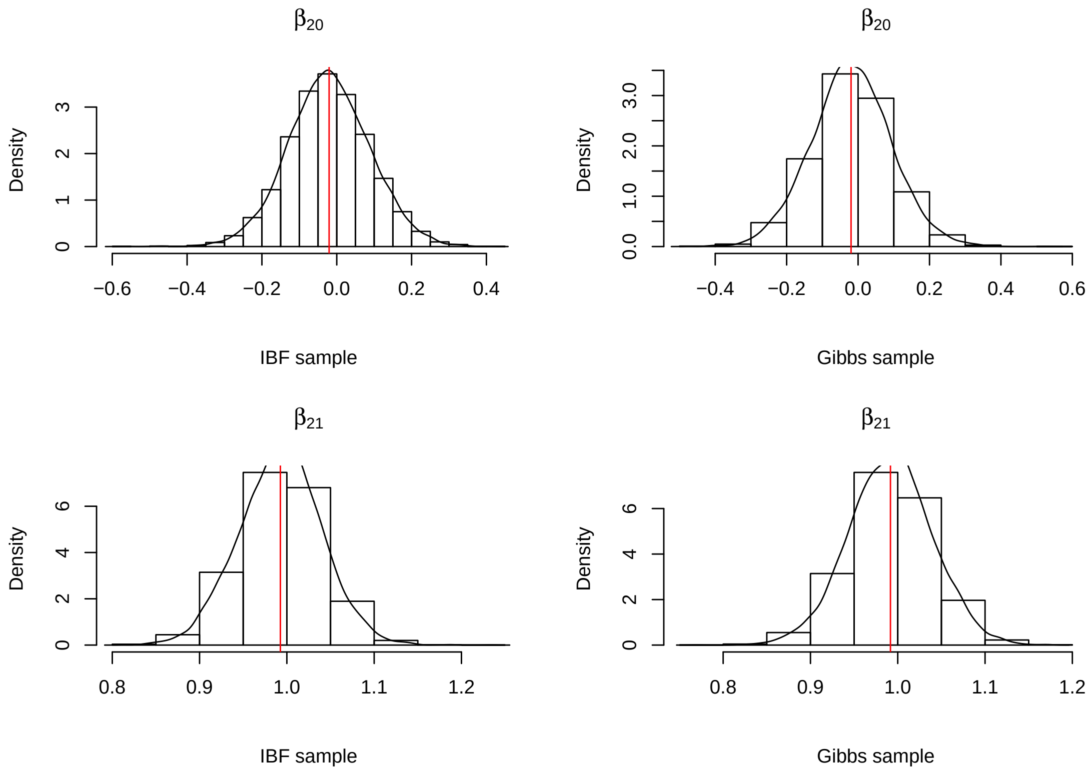

Figure S6: Density histograms of posterior samples for β_{20} and β_{21} based on IBF samples and Gibbs samples.

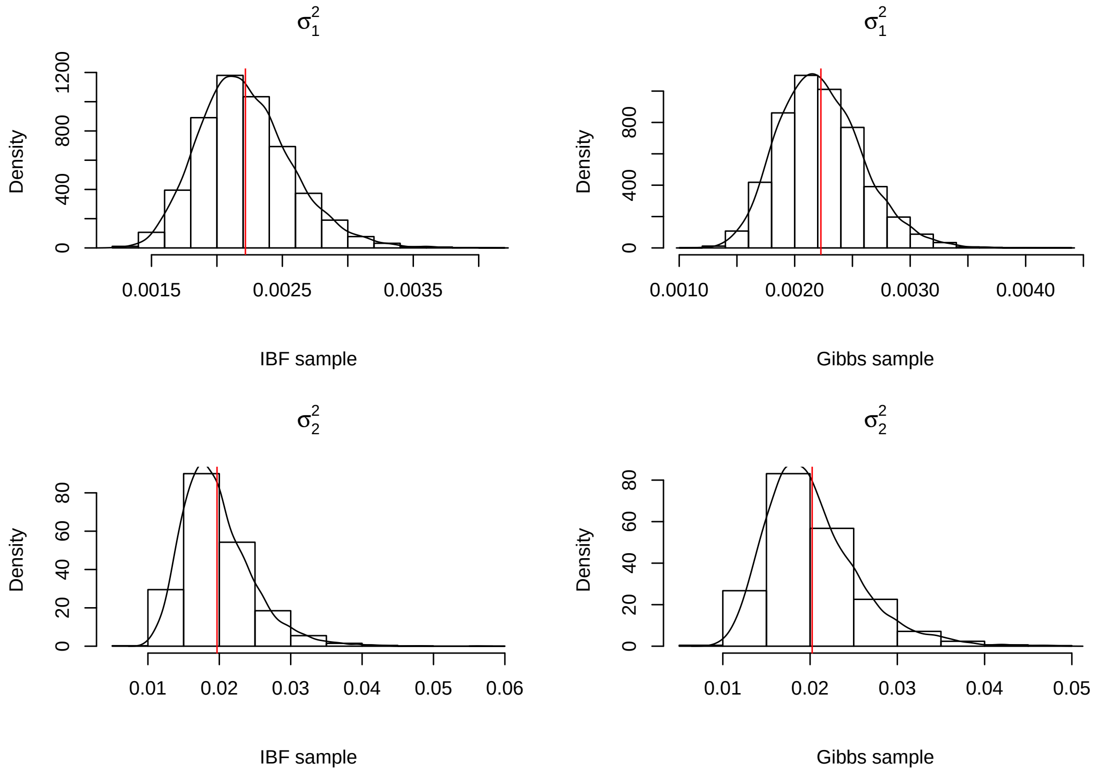

Figure S7: Density histograms of posterior samples for σ_1^2 and σ_2^2 by IBF sampler and Gibbs sampler.

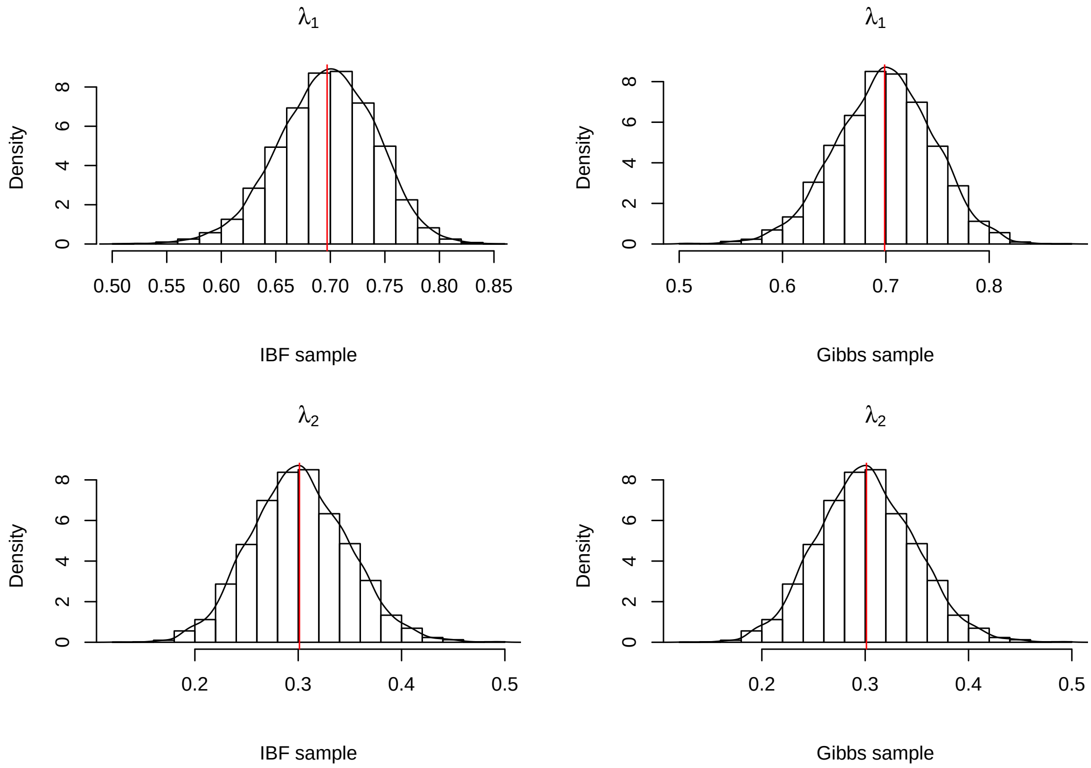

Figure S8: Density histograms of posterior samples for λ_1 and λ_2 by IBF sampler and Gibbs sampler.