1. Let $\boldsymbol{Y}$ be a $n \times p$ data matrix. Let O_{ij} denote if Y_{ij} is missing, i.e., $Y_{ij} = \texttt{NA}$ if $O_{ij} = 0$. Consider the model for the complete data $\boldsymbol{Y}$ that

$$\boldsymbol{Y}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

with priors $\boldsymbol{\mu} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0)$ and $\boldsymbol{\Sigma} \sim InvWishart(v, \boldsymbol{I}_p)$. Suppose you can only sample univariate truncated normal distribution but not multivariate truncated normal distribution, describe how to implement a Gibbs sampler to make inference about the parameters of the model.

Complete-data Likelihood:

$$P(Y|\mu,\Sigma) = P(Y|\mu,\Sigma)\cdot P(O|Y,\mu,\Sigma)$$

$$\propto \prod_{i=1}^{n} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{(y_i-\mu)^T\Sigma^{-1}(y_i-\mu)}{2}\right)$$

Including Ymiss, and Yobserved.

$$P(Y,\mu,\Sigma) \propto P(Y|\mu,\Sigma)\cdot P(\mu)\cdot P(\Sigma)$$

$$\propto |\Sigma|^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^{n}(y_i-\mu)^T\Sigma^{-1}(y_i-\mu)}{2}\right)\cdot \exp\left(-\frac{(\mu-\mu_0)^T\Lambda_0^{-1}(\mu-\mu_0)}{2}\right)$$

$$\cdot |\Sigma|^{-\frac{\nu+p+1}{2}} \exp\left(-\frac{1}{2}tr(\Sigma^{-1})\right)$$

Actually, Y_{mis} is condisered parameters in this model:

For each i, we have $Y_{obs,i}$, $Y_{mis,i}$.

And $Y_{mis,i} \mid Y_{obs,i}, \mu, \Sigma$ is a conditional distribution of a MVN distribution, so WLOG, let $Y_{i1}, \dots, Y_{ik_i}$ be observed Y and $Y_{ik_i+1} \dots Y_{ip}$ be missing data. Then, $\mu = \begin{pmatrix} \mu_{obs,i} \\ \mu_{mis,i} \end{pmatrix}$ $\Sigma = \begin{pmatrix} \Sigma_{i,11} & \Sigma_{i,12} \\ \Sigma_{i,21} & \Sigma_{i,22} \end{pmatrix}$

in which $\begin{cases} \Sigma_{i,11} \text{ describes the } Cov(Y_{obs,i}) \\ \Sigma_{i,12} \text{ describes the } Cov(Y_{obs,i}, Y_{mis,i}) \\ \Sigma_{i,21} = \Sigma_{i,12}^{T} \\ \Sigma_{i,22} \text{ describes the } Cov(Y_{mis,i}) \end{cases}$

Matrix partition.

Therefore:

$$(Y_{mis,i} \mid Y_{obs,i}, \mu, \Sigma) \sim MVN(\mu_i, \Sigma_i)$$

in which: $\mu_i = \mu_{mis,i} + \Sigma_{i,21}\Sigma_{i,11}^{-1}(Y_{obs,i} - \mu_{obs,i})$

$\Sigma_i = \Sigma_{i,22} - \Sigma_{i,21}\Sigma_{i,11}^{-1}\Sigma_{i,12}$

Since $Y_{mis,i}$ has been generated, we got a "pseudo complete data" Y_i, then we can continue the Gibbs Sampler for μ and Σ as usual:

$$p(\mu \mid Y, \Sigma) \propto \exp\left(-\frac{1}{2}\left[\left(n \cdot \mu^T\Sigma^{-1}\mu - \Sigma_{i=1}^{n} y_i^T \Sigma^{-1}\mu - \mu^T\Sigma^{-1}\Sigma_{i=1}^{n} y_i\right) + \left(\mu^T\Lambda_0^{-1}\mu - \mu_0^T\Lambda_0^{-1}\mu - \mu^T\Lambda_0^{-1}\mu_0\right)\right]\right)$$

$$= \exp\left(-\frac{1}{2}\left(\mu^T(n\Sigma^{-1} + \Lambda_0^{-1})\mu - (\Sigma_{i=1}^{n} y^T\Sigma^{-1} + \mu_0^T\Lambda_0^{-1})\mu - \mu^T(\Sigma^{-1}\Sigma_{i=1}^{n} y_i + \Lambda_0^{-1}\mu_0)\right)\right)$$

$$\propto \exp\left(-\frac{1}{2}\left[\mu - (n\Sigma^{-1} + \Lambda_0^{-1})^{-1}(\Sigma^{-1}\Sigma_{i=1}^{n} y_i + \Lambda_0^{-1}\mu_0)\right]^T (n\Sigma^{-1} + \Lambda_0^{-1}) \cdot \left[\mu - (n\Sigma^{-1} + \Lambda_0^{-1})^{-1}(\Sigma_0^{-1}\Sigma_{i=1}^{n} y_i + \Lambda_0^{-1}\mu_0)\right]\right)$$

$$\sim MVN\left((n\Sigma^{-1}+\Lambda_0^{-1})^{-1}(\Sigma^{-1}\textstyle\sum_{i=1}^{n} y_i+\Lambda_0^{-1}\mu_0),\ (n\Sigma^{-1}+\Lambda_0^{-1})^{-1} \right)$$

$$p(\Sigma \mid Y, \mu) \propto |\Sigma|^{-\frac{n+\nu+p+1}{2}} \exp\left(-\frac{1}{2} tr\left[\left(\frac{\sum_{i=1}^{n}(y_i-\mu)^T(y_i-\mu)\cdot I}{2}+I\right)\Sigma^{-1}\right]\right)$$

$$\sim Inv\text{-}Wishart\left(n+\nu,\ \left(\frac{\sum_{i=1}^{n}(y_i-\mu)^T(y_i-\mu)}{2}+1\right)I\right)$$

2. Simulate data from the following model with $n = 20, m = 10, p = 5$,

$$
\begin{aligned}
y_{ij} &= \beta_{0i} + \sum_{k=1}^{p} \beta_{ki} x_{ijk} + \epsilon_{ij}, \quad i = 1, ..., n; j = 1, ..., m \\
\beta_{0i} &\sim N(\mu_0, \tau_0^2), \quad i = 1, ..., n \\
\beta_{ki} &\sim N(\mu_k, \tau_k^2), \quad i = 1, ..., n; k = 1, ..., p \\
\sigma_{ijk} &\sim N(0, \sigma^2)
\end{aligned}
$$

with your choice of $\boldsymbol{\mu}$, $\boldsymbol{\tau}$, and σ^2. Fit (1) a linear regression with common β_0 and $\{\beta_k\}_{k=1:p}$ for all n groups, (2) a random intercept model, and (3) a random slope model. Summarize your posterior distributions of the regression coefficients.

---

(1) for linear regression, we have:

$$y_{ij} = \beta_0 + \sum_{k=1}^{p} x_{ijk} \cdot \beta_k + \varepsilon_{ij}$$

Set $\beta_l \sim N(\mu_l, \tau_l^2)$, $l = 0, 1, \cdots, p$. and a conjugate prior for $\sigma^2 \sim IG\left(\frac{\nu_0}{2}, \frac{\nu_0^2 \sigma_0^2}{2}\right)$.

$$p(y_{ij} \mid \beta, \sigma^2) \propto (\sigma^2)^{-\frac{1}{2}} \exp\left(-\frac{y_{ij} - \sum_{k=1}^{p} x_{ijk} \cdot \beta_k - \beta_0}{2\sigma^2}\right)$$

$$p(Y, \beta, \sigma^2) \propto (\sigma^2)^{-\frac{mn}{2}} \exp\left(-\frac{\sum_{i=1}^{n} \sum_{j=1}^{m} \left(y_{ij} - \sum_{k=1}^{p} x_{ijk} \cdot \beta_k - \beta_0\right)^2}{2\sigma^2}\right) \cdot 1$$

$$\cdot (\sigma^2)^{-\frac{\nu_0}{2} - 1} \exp\left(-\frac{1}{\sigma^2} \cdot \frac{\nu_0^2 \sigma_0^2}{2}\right) \cdot \prod_{l=1}^{p} \exp\left(-\frac{(\beta_l - \mu_l)^2}{2\tau_l^2}\right)$$

$$\sum_{k=1}^{p} x_{ijk} \cdot \beta_k = X_{ij} \beta$$

$$p(\beta_0 \mid \text{others}) \propto \exp\left(-\frac{1}{2}\left(\left(\frac{mn}{\sigma^2} + \frac{1}{\tau_0^2}\right)\beta_0^2 - 2\beta_0\left(\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij} - X_{ij}\beta)}{\sigma^2} + \frac{\mu_l}{\tau^2}\right)\right)$$

$$\sim N\left(\frac{\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij} - X_{ij}\beta)}{\sigma^2} + \frac{\mu_l}{\tau_0^2}}{mn/\sigma^2 + 1/\tau_0^2}, \left(mn/\sigma^2 + 1/\tau_0^2\right)^{-1}\right)$$

$$p(\beta_k \mid \text{others}) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(y_{ij}-\beta_0-x_{ij}\beta\right)^2\right)\cdot\exp\left(-\frac{(\beta_k-\mu_k)^2}{2\tau_k^2}\right)$$

$$\propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{m}\left((x_{ijk}\beta_k)^2-2x_{ijk}\beta_k\left(y_{ij}-\beta_0-\sum_{\ell\neq k}x_{ij\ell}\beta_\ell\right)\right]-\frac{(\beta_k-\mu_k)^2}{2\tau_k^2}\right)\right)$$

$$=\exp\left(-\frac{1}{2}\left(\beta_k^2\left(\sum_{i=1}^{n}\sum_{j=1}^{m}x_{ijk}^2/\sigma^2+1/\tau_k^2\right)-2\beta_k\left(\sum_{i=1}^{n}\sum_{j=1}^{m}\left(y_{ij}-\beta_0-\sum_{\ell\neq k}x_{ij\ell}\beta_\ell\right)/\sigma^2+\mu_k/\tau_k^2\right)\right)\right)$$

$$\sim N\left(\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\left(y_{ij}-\beta_0-\sum_{\ell\neq k}x_{ij\ell}\beta_\ell\right)/\sigma^2+\mu_k/\tau_k^2}{\sum_{i=1}^{n}\sum_{j=1}^{m}x_{ijk}^2/\sigma^2+1/\tau_k^2},\ \left(\sum_{i=1}^{n}\sum_{j=1}^{m}x_{ijk}^2/\sigma^2+1/\tau_k^2\right)^{-1}\right)$$

$$p(\sigma^2\mid\text{others})\propto(\sigma^2)^{-\frac{mn+\nu_0}{2}-1}\exp\left(-\frac{1}{\sigma^2}\cdot\left(\frac{\nu_0^2\sigma_0^2}{2}+\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-x_{ij}\beta-\beta_0)^2}{2}\right)\right)$$

$$\sim IG\left(\frac{mn+\nu_0}{2},\ \frac{\nu_0^2\sigma_0^2+\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-x_{ij}\beta-\beta_0)^2}{2}\right)$$

(2) Based on (1), we have β_{0i} instead of β_0, also,

we have $\beta_{0i}\sim N(\mu_0,\tau_0^2)$. and $\mu_0\sim N(0,\tau_{\mu_0}^2)$

$\tau_0^2\sim IG\left(\frac{\nu_{\tau_0}}{2},\frac{\nu_\tau\sigma_{\tau_0}^2}{2}\right)$

Joint distribution:

$$p(Y,\beta,\sigma^2,\mu_0,\tau_0)\propto(\sigma^2)^{-\frac{mn}{2}}\exp\left(-\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\left(y_{ij}-\sum_{k=1}^{p}x_{ijk}\beta_k-\beta_{0i}\right)^2}{2\sigma^2}\right)$$

$$\cdot(\sigma^2)^{-\frac{\nu_0}{2}-1}\exp\left(-\frac{1}{\sigma^2}\cdot\frac{\nu_0^2\sigma_0^2}{2}\right)$$

$$\prod_{i=1}^{m}(\tau_0)^{-\frac{1}{2}}\exp\left(-\frac{(\beta_{0i}-\mu_0)^2}{2\tau_0^2}\right)\cdot\prod_{i=1}^{m}\exp\left(-\frac{\mu_0^2}{2\tau_\mu^2}\right)\cdot$$

$$\prod_{i=1}^{m}(\tau_0^2)^{-\frac{\nu_\tau}{2}-1}\exp\left(-\frac{\nu_\tau\sigma_\tau^2}{2\tau_0^2}\right)$$

$$p(\beta_{0i}\mid\text{others})\propto\exp\left(-\frac{\sum_{j=1}^{m}\left(y_{ij}-\sum_{k=1}^{p}x_{ijk}\cdot\beta_k-\beta_{0i}\right)^2}{2\sigma^2}-\frac{(\beta_{0i}-\mu_0)^2}{2\tau_0^2}\right)$$

$$\propto\exp\left(-\frac{1}{2}\left(\beta_{0i}^2\left(m/\sigma^2+1/\tau_{0i}^2\right)-2\beta_{0i}\left(\sum_{j=1}^{m}\left(y_{ij}-\sum_{k=1}^{p}x_{ijk}\beta_k\right)/\sigma^2+\mu_0/\tau_0^2\right)\right)\right)$$

$$\sim N\left(\frac{\sum_{j=1}^{m}(y_{ij}-\sum_{k=1}^{p}x_{ijk}-\beta_k)/\sigma^2+\mu_0/\tau_{0i}^2}{m/\sigma^2+1/\tau_{0i}^2},\ (m/\sigma^2+1/\tau_{0i}^2)^{-1}\right)$$

$P(\beta_k|\text{others}) \sim$

$$N\left(\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-\beta_{0i}-\sum_{\ell\neq k}x_{ij\ell}\beta_\ell)/\sigma^2+\mu_k/\tau_k^2}{\sum_{i=1}^{n}\sum_{j=1}^{m}x_{ijk}^2/\sigma^2+1/\tau_k^2},\ \left(\sum_{i=1}^{n}\sum_{j=1}^{m}x_{ijk}^2/\sigma^2+1/\tau_k^2\right)^{-1}\right)$$

the only difference between random effect model and linear regression is using β_{0i} insted of β_0

$$P(\sigma^2|\text{others}) \sim IG\left(\frac{mn+\nu_0}{2},\ \frac{\nu_0\sigma_0^2+\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-x_{ij}\beta-\beta_{0i})^2}{2}\right)$$

$$P(\mu_0|\text{others}) \sim N\left(\frac{\sum_i\beta_{0i}}{m/\tau_0^2+1/\tau_{\mu_0}^2},\ (m/\tau_0^2+1/\tau_{\mu_0}^2)^{-1}\right)$$

$$P(\tau_0|\text{others}) \sim IG\left(\frac{m+\nu_{\tau_0}}{2},\ \frac{\sum_i(\beta_{0i}-\mu_0)^2+\nu_{\tau_0}\sigma_{\tau_0}^2}{2}\right)$$

(3) Based on (2) we have other assumptions:

$\beta_{ki} \sim N(\mu_k, \tau_k)$ $\mu_k \sim N(0, \tau_{\mu_k})$ $\tau_k \sim IG\left(\nu_{\tau_k}, \frac{\nu_{\tau_k}\sigma_{\tau_k}^2}{2}\right)$

$\beta_{0i}|\text{others} \sim$ the difference is β_{ki} insted of β_k

$$N\left(\frac{\sum_{j=1}^{m}(y_{ij}-\sum_{k=1}^{p}x_{ijk}-\beta_{ki})/\sigma^2+\mu_0/\tau_{0i}^2}{m/\sigma^2+1/\tau_{0i}^2},\ (m/\sigma^2+1/\tau_{0i}^2)^{-1}\right)$$

$$P(\beta_{ki} \mid \text{others}) \propto \exp\left(-\frac{1}{2}\left(\frac{\sum_{j=1}^{m}(y_{ij} - \sum_{k=1}^{p} x_{ijk} \cdot \beta_{ki} - \beta_{0i})^2}{\sigma^2} + \frac{(\beta_{ki} - \mu_k)^2}{\tau_k^2}\right)\right)$$

$$\propto \exp\left(-\frac{1}{2}\left(\beta_{ki}^2\left(\sum_{j=1}^{m} x_{ijk}^2/\sigma^2 + 1/\tau_k^2\right) - 2\beta_{ki}\left(\sum_{j=1}^{m}(y_{ij} - \sum_{\ell \neq k} x_{ij\ell}\beta_{\ell i} - \beta_{0i})^2/\sigma^2 + \mu_k/\tau_k^2\right)\right)\right)$$

$$\sim N\left(\frac{\sum_{j=1}^{m}(y_{ij} - \sum_{\ell \neq k} x_{ij\ell}\beta_{\ell i} - \beta_{0i})^2/\sigma^2 + \mu_k/\tau_k^2}{\sum_{j=1}^{m} x_{ijk}^2/\sigma^2 + 1/\tau_k^2}, \left(\sum_{j=1}^{m} x_{ijk}^2/\sigma^2 + 1/\tau_k^2\right)^{-1}\right)$$

$$P(\sigma^2 \mid \text{others}) \sim IG\left(\frac{mn + \nu_0}{2}, \frac{\nu_0 \sigma_0^2 + \sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij} - \sum_{k=1}^{p} x_{ijk}\beta_{ki} - \beta_{0i})^2}{2}\right)$$

$$P(\mu_0 \mid \text{others}) \sim N\left(\frac{\sum_i \beta_{0i}}{m/\tau_0^2 + 1/\tau_{\mu_0}^2}, \left(m/\tau_0^2 + 1/\tau_{\mu_0}^2\right)^{-1}\right)$$

$$P(\tau_0 \mid \text{others}) \sim IG\left(m + \nu_{\tau_0}, \frac{\sum_i(\beta_{0i} - \mu_0)^2 + \nu_{\tau_0}\sigma_{\tau_0}^2}{2}\right)$$

$$P(\mu_k \mid \text{others}) \sim N\left(\frac{\sum_i \beta_{ki}}{m/\tau_k^2 + 1/\tau_{\mu_k}^2}, \left(m/\tau_k^2 + 1/\tau_{\mu_k}^2\right)^{-1}\right)$$

$$P(\tau_k^2 \mid \text{others}) \sim IG\left(m + \nu_{\tau_k}, \frac{\sum_i(\beta_{ki} - \mu_k)^2 + \nu_{\tau_k}\sigma_{\tau_k}^2}{2}\right)$$

Regression by Stan: $\nearrow$ intercept $\nearrow$ slope

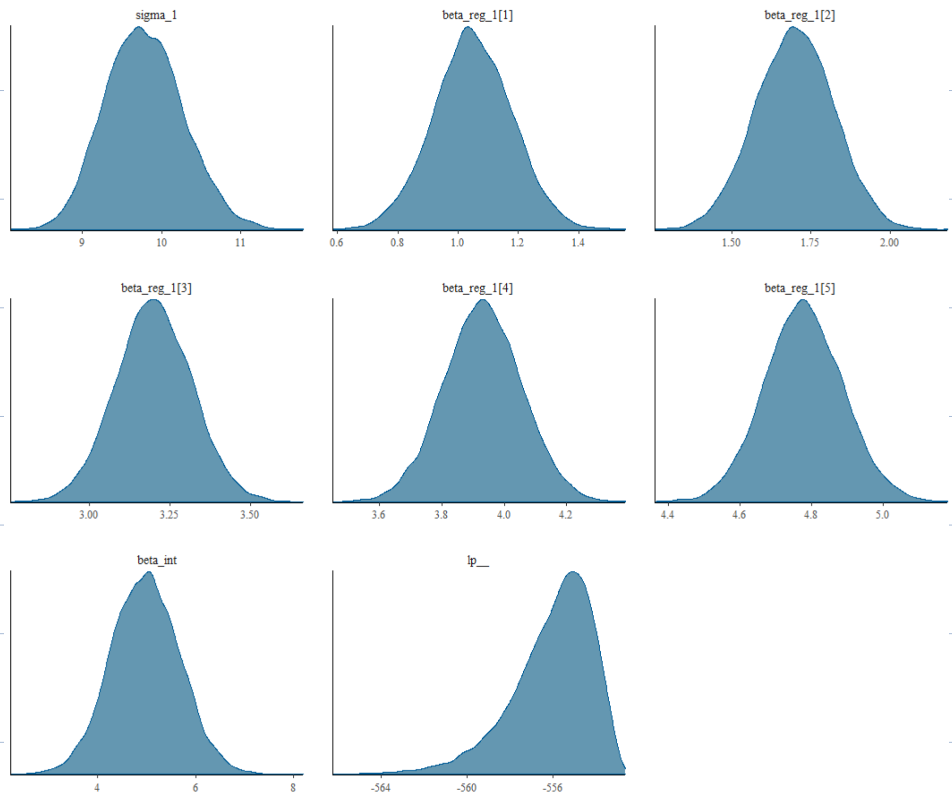
1 set $\mu_0 = 5$ $\mu_{others} = 1:5$

$2_0 = 1$ $2_{others} = 1:5$

$\sigma = 1.5$

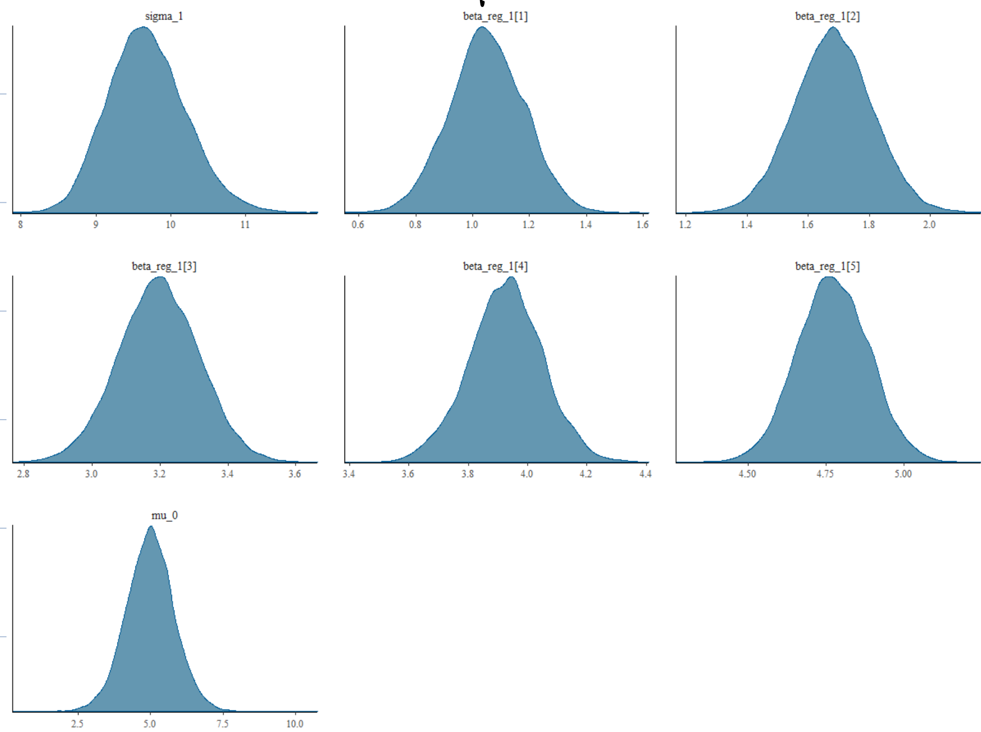
in data generating process.

Case 1: Common Bayesian Regression



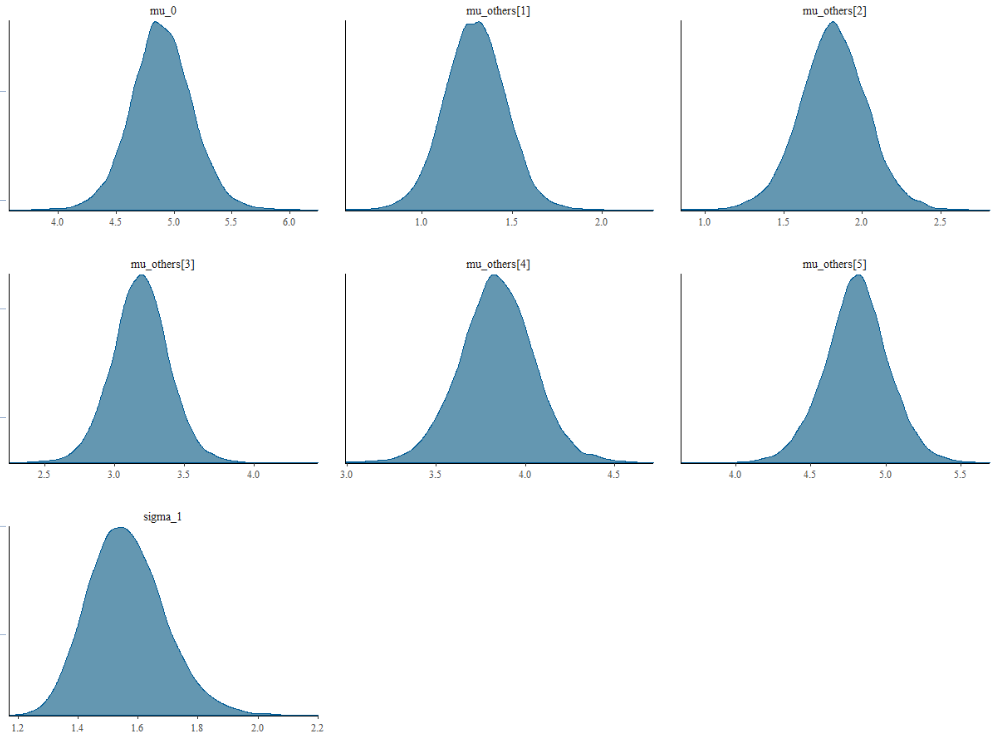
σ^2 is not accurate, but other slopes remain around the true value. It may be because the randomness of slopes and intercepts are added into the residuals.

Case 2: Random Intercept



it's close to the last model, but σ^2 decreases a little. but not so obvious.

Case 3:



We can see the distribution of σ^2 changes to our simulated parameter. And others remain close to the setting.

Conclusion: As long as there exists a group-wise difference in the intercept or slope, the random effect model will have smaller residuals than merging all of them together.

Using the `broadway` dataset, fit a linear mixed effect model in the following form

$$y_{ijk} = \alpha_i + \beta_j + \epsilon_{ijk} \qquad \varepsilon_{ijk} \sim (0, \sigma^2)$$

where y_{ijk} is the weekly gross for show i at year j and week k, with the following priors

set $\pi(\sigma^2) \propto 1 : 1/\sigma^2$

$$\alpha_i \sim N(\mu_\alpha, \tau_\alpha)$$

$$\beta_j \sim N(\mu_\beta, \tau_\beta)$$

$$f(y \mid \alpha, \beta, \sigma^2) \propto (\sigma^2)^{-\frac{\sum_{i=1}^{2}\sum_{j=1}^{J_j} n_{ij}}{2}} \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{2}\sum_{j=1}^{J_i}\sum_{k=1}^{k_{ij}}(y_{ijk}-\alpha_i-\beta_j)^2\right)\right)$$

$$\text{Joint}: f(y, \alpha, \beta, \sigma^2) \propto (\sigma^2)^{-\frac{\sum_{i=1}^{2}\sum_{j=1}^{J_j} n_{ij}}{2}} \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{2}\sum_{j=1}^{J_i}\sum_{k=1}^{k_{ij}}(y_{ijk}-\alpha_i-\beta_j)^2\right)\right)$$

$$\cdot \prod_{i=1}^{n} \exp\left(-\frac{(\alpha_i-\mu_\alpha)^2}{2\tau_\alpha^2}\right) \cdot \prod_{j=1}^{J}\left(-\frac{(\beta_i-\mu_\beta)^2}{2\tau_{\beta^2}}\right) \cdot 1/\sigma^2$$

$$f(\alpha_i \mid \text{others}) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_{j=1}^{J_i}\sum_{k=1}^{k_{ij}}(\alpha_i-(y_{ijk}-\beta_j))^2 - \frac{1}{2\tau_\alpha^2}(\alpha_i-\mu_\alpha)^2\right)$$

$$\propto \exp\left(-\frac{1}{2}\left(\alpha_i^2\left(n_i/\sigma^2 + 1/\tau_\alpha^2\right) - 2\alpha_i\left(\frac{\sum_{j=1}^{J_i}\sum_{k=1}^{k_{ij}}(y_{ijk}-\beta_j)}{\sigma^2} + \frac{\mu_\alpha}{\tau_\alpha^2}\right)\right)\right)$$

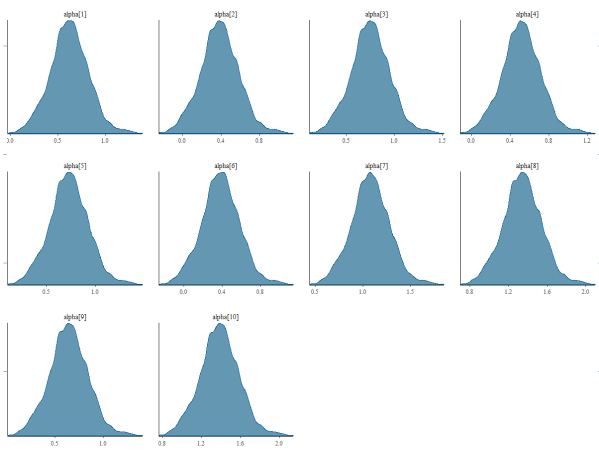
$$\sim N\left(\frac{\frac{\sum_{j=1}^{J_i}\sum_{k=1}^{k_{ij}}(y_{ijk}-\beta_j)}{\sigma^2} + \frac{\mu_\alpha}{\tau_\alpha^2}}{n_i/\sigma^2 + 1/\tau_\alpha^2}, \left(n_i/\sigma^2 + 1/\tau_\alpha^2\right)^{-1}\right)$$

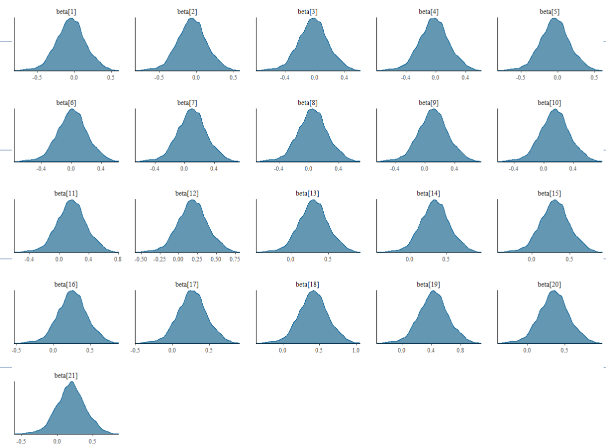
$$f(\beta_j \mid \text{others}) \propto \exp\left(-\frac{1}{2}\left(\beta_j^2\left(n_j/\sigma^2 + 1/\tau_\beta^2\right) - 2\beta_j\left(\frac{\sum_{i=1}^{I_j}\sum_{k=1}^{k_{ij}}(y_{ijk}-\alpha_i)}{\sigma^2} + \frac{\mu_\beta}{\tau_\beta^2}\right)\right)\right)$$

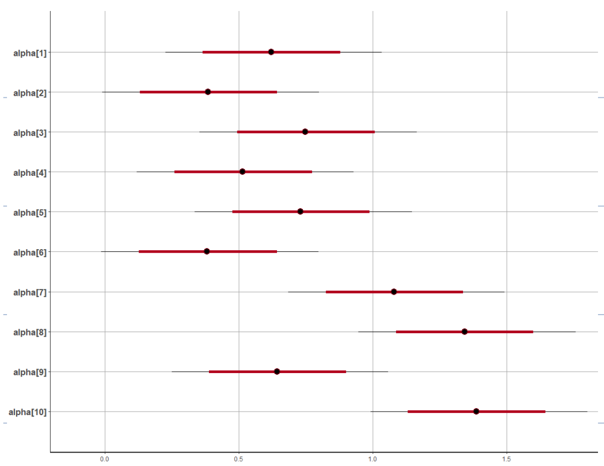
$$\sim N\left(\frac{\frac{\sum_{i=1}^{I_j}\sum_{k=1}^{k_{ij}}(y_{ijk}-\alpha_i)}{\sigma^2} + \frac{\mu_\beta}{\tau_\beta^2}}{n_j/\sigma^2 + 1/\tau_\beta^2}, \left(n_j/\sigma^2 + 1/\tau_\beta^2\right)^{-1}\right)$$

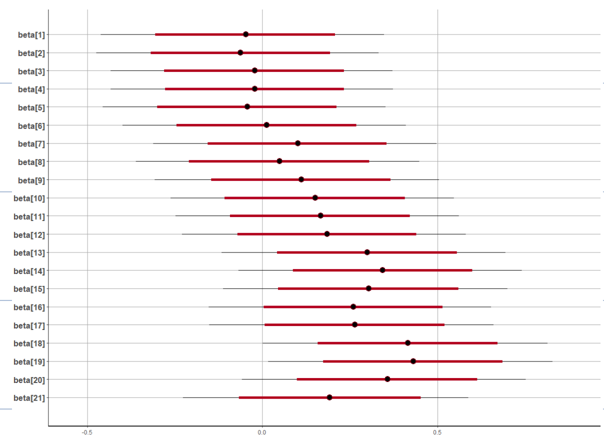
$$\sigma^2 \mid \text{others} \propto IG\left(\frac{N}{2}, \frac{\sum_{i=1}^{2}\sum_{j=1}^{J_i}\sum_{k=1}^{k_{ij}}(y_{ijk}-\alpha_i-\beta_j)^2}{2}\right)$$

After using stan, here is the posterior distribution for all 10 α's and 21 β's

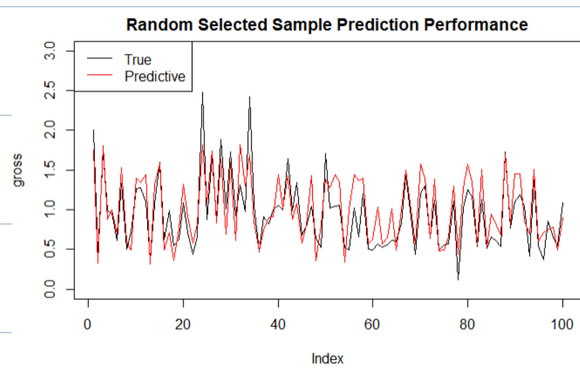








Here is the prediction performance if I randomly pick 100 samples in the dataset and make a prediction.



4. Using the `broadway` dataset from the previous question, fit an alternative linear mixed effect model in the following form

$$y_{ijk} = \alpha_i + \beta_k + \gamma_i x_{ijk} + \epsilon_{ijk}$$

where y_{ijk} is the weekly gross for show i at year j and week k and $x_{ijk} = j$, i.e., x_{ijk} is the year index from 1 to 21. *Notice that in this question, we assume the outcome is linear in year, instead of treating year as an indicator as in the previous question.* Consider the following priors

$$\begin{aligned}\alpha_i &\sim N(\mu_\alpha, \tau_\alpha) & \mu_\alpha = 0 \quad \tau_\alpha = 1\\ \beta_k &\sim N(\mu_\beta, \tau_\beta) & \mu_\beta = 0 \quad \tau_\beta = 1\\ \gamma_i &\sim N(\mu_\gamma, \tau_\gamma) & \mu_\gamma = 0 \quad \tau_\gamma = 1\end{aligned}$$

This question is pretty similar to Question 3.

$$\alpha_i \mid \text{others} \sim N\left(\frac{\frac{\sum_{j=1}^{J_i}\sum_{k=1}^{k_{ij}}(y_{ijk} - \beta_j - \gamma_i x_{ijk})}{\sigma^2} + \frac{\mu_\alpha}{\tau_\alpha^2}}{n_i/\sigma^2 + 1/\tau_\alpha^2},\ \left(n_i/\sigma^2 + 1/\tau_\alpha^2\right)^{-1} \right)$$

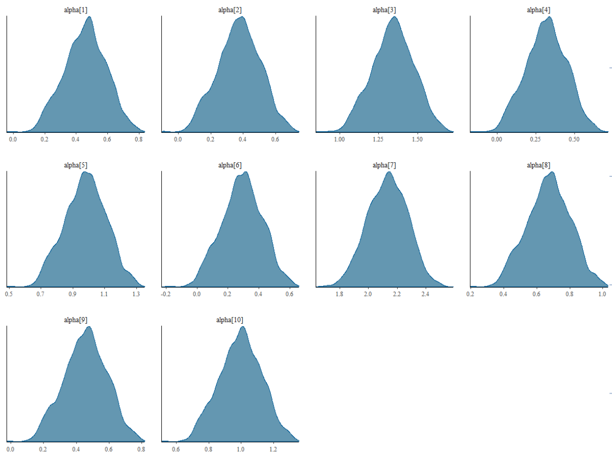
$$\beta_k \mid \text{others} \sim N\left(\frac{\frac{\sum_{i=1}^{I_k}\sum_{j=1}^{J_{ki}}(y_{ijk} - \alpha_i - \gamma_i x_{ijk})}{\sigma^2} + \frac{\mu_\beta}{\tau_\beta^2}}{n_k/\sigma^2 + 1/\tau_\beta^2},\ \left(n_k/\sigma^2 + 1/\tau_\beta^2\right)^{-1} \right)$$

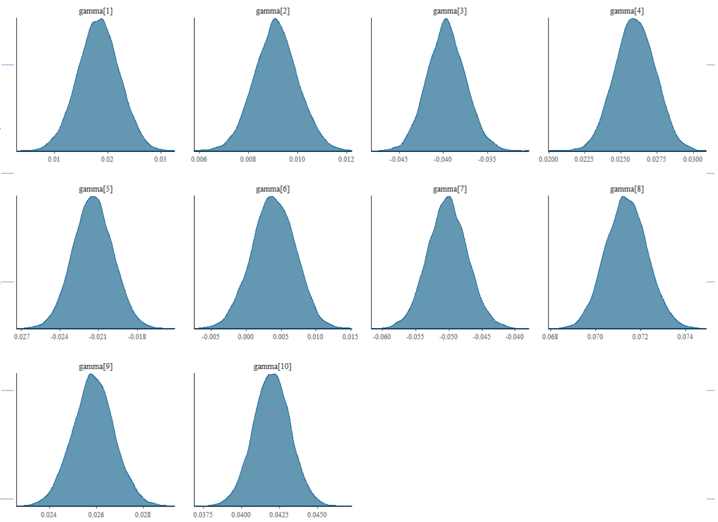
$$p(\gamma_i \mid \text{others}) \propto \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{k=1}^{k_i}\sum_{j=1}^{J_{ik}}(y_{ijk} - \alpha_i - \beta_k - \gamma_i x_{ijk})^2\right) - \frac{(\gamma_i - \mu_\gamma)^2}{2\tau_\gamma^2}\right)$$

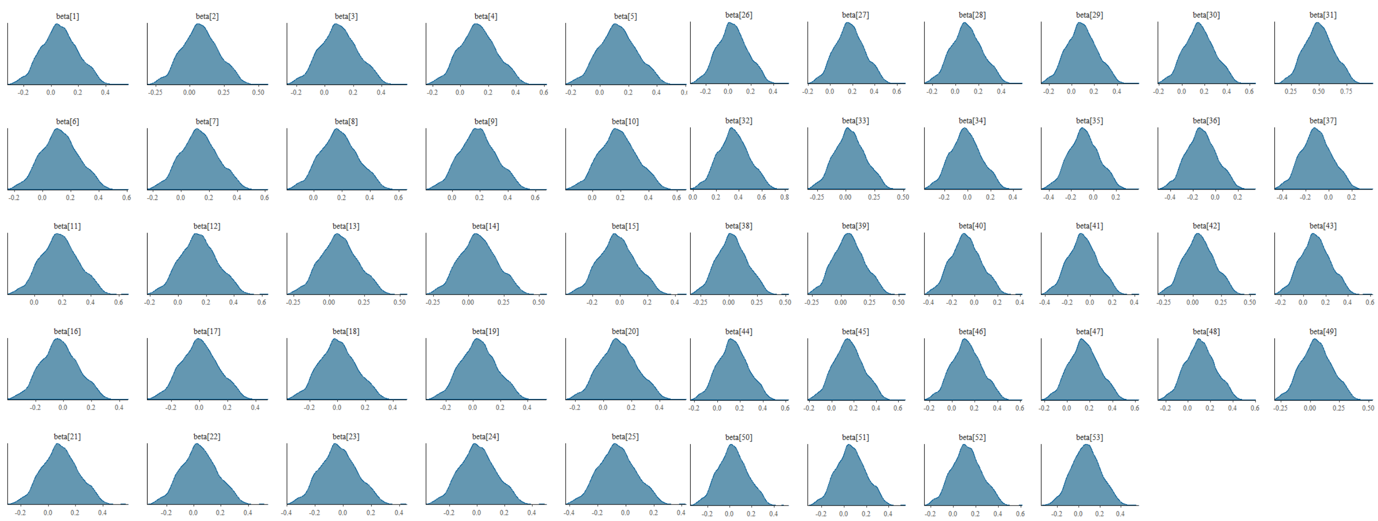
$$\propto \exp\left(-\frac{1}{2}\left(\gamma_i^2\left(\sum_{k=1}^{k_i}\sum_{j=1}^{J_{ik}} x_{ijk}^2/\sigma^2 + 1/\tau_\gamma^2\right) - 2\gamma_i\left(\sum_{k=1}^{k_i}\sum_{j=1}^{J_{ik}}(y_{ijk} - \alpha_i - \beta_k)\, x_{ijk}/\sigma^2 + \mu_\gamma/\tau_\gamma^2\right)\right)\right)$$

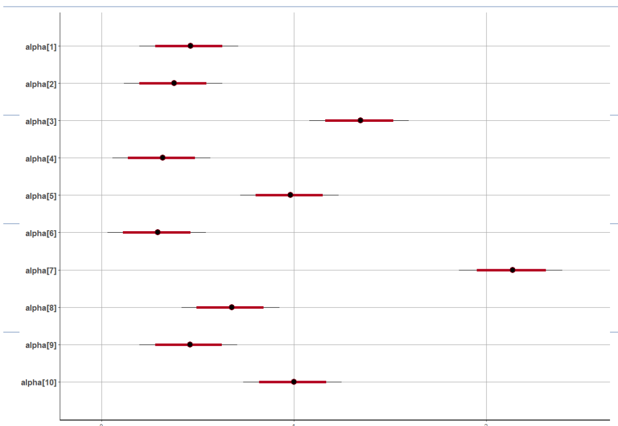
$$\sim N\left(\frac{\sum_{k=1}^{k_i}\sum_{j=1}^{J_{ik}}(y_{ijk} - \alpha_i - \beta_k)\cdot x_{ijk}/\sigma^2 + \mu_\gamma/\tau_\gamma^2}{\frac{\sum_{k=1}^{k_i}\sum_{j=1}^{J_{ik}} x_{ijk}^2}{\sigma^2} + 1/\tau_\gamma^2},\ \left(\frac{\sum_{k=1}^{k_i}\sum_{j=1}^{J_{ik}} x_{ijk}^2}{\sigma^2} + 1/\tau_\gamma^2\right)^{-1} \right)$$

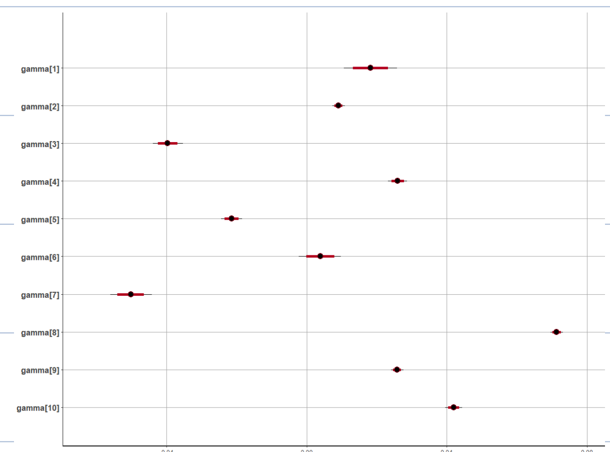
Here is the posterior distribution for all α's, β's and γ's

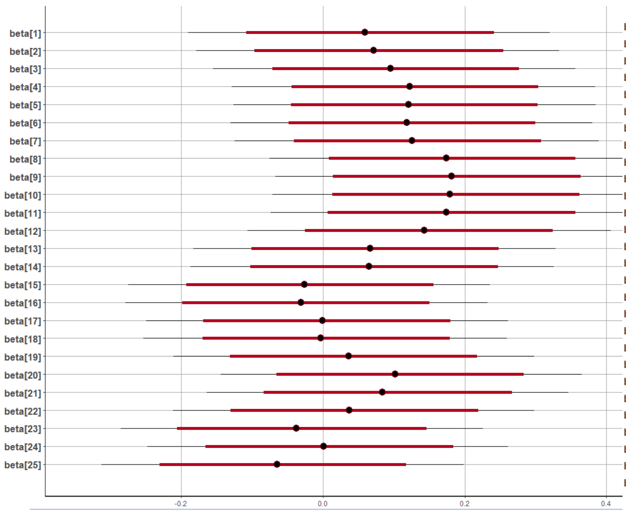


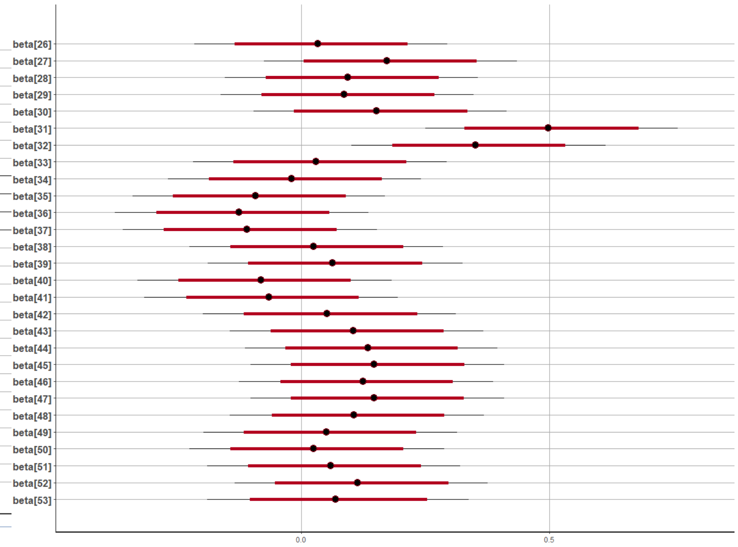




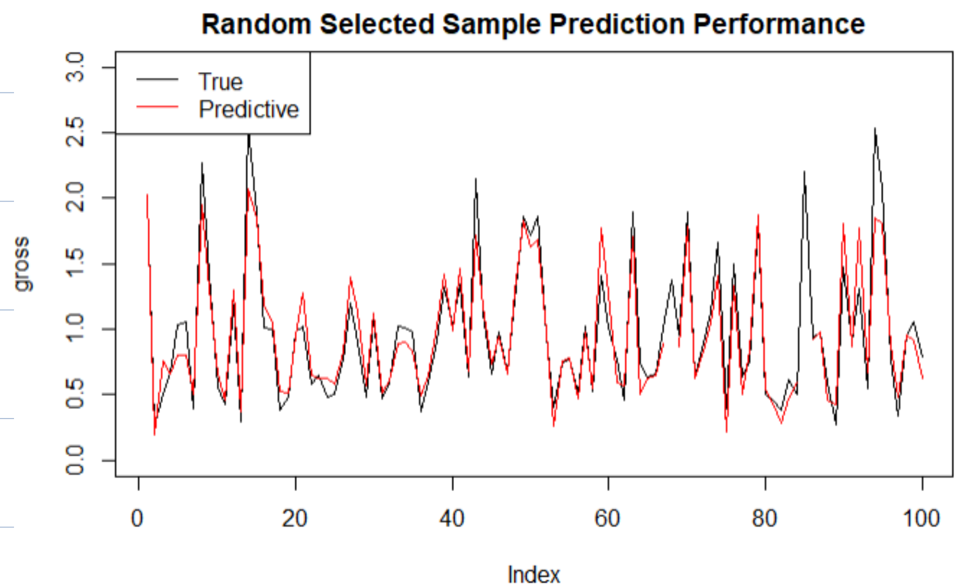








If randon select 100 samples, the predictive performace is shown as follows:



It performs better than the last model!

5. Using the `mtcars` dataset from the `datasets` package in R. Consider the the 'qsec' variable (1/4 mile time) as the outcome and all other variables as predictors. Fit a linear regression model with spike-and-slab prior for the regression coefficients. Summary your regression coefficients, including the posterior mean, sd, 95% credible intervals, and the probability of each predictor being included.

   To access the dataset and run a simple regression, you can use the following commands in R.

```
library(datasets)
data(mtcars)
lm(qsec ~.,  data = mtcars)
```

$\varepsilon_i \sim N(0, \sigma^2)$

Model: $Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i$

For a spike and slab prior:

$$p(\beta_i) \propto (1-z_i)\cdot(\tau_0^2)^{-\frac{1}{2}} \exp\left(-\frac{\beta_i^2}{2\sigma^2\tau_0^2}\right) + z_i(\tau_1^2)^{-\frac{1}{2}}\exp\left(-\frac{\beta_i^2}{2\sigma^2\tau_1^2}\right)$$

$$z_i \sim Ber(q)$$

Joint $f(y, \beta, z_j, \sigma^2)$

$$\propto (2\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n \left(y_i - \sum_{j=1}^p x_{ij}\cdot\beta_j\right)^2\right) \cdot 1/\sigma^2$$

$$\prod_{j=1}^p \left[(1-z_j)\cdot(\sigma_{\beta_j}^2\tau_0^2)^{-\frac{1}{2}} \exp\left(-\frac{\beta_j^2}{2\sigma_{\beta_j}^2\tau_0^2}\right) + z_j(\sigma_{\beta_j}^2\tau_1^2)^{-\frac{1}{2}}\exp\left(-\frac{\beta_j^2}{2\sigma_{\beta_j}^2\tau_1^2}\right)\right]$$

$$f(\beta_j \mid \text{others}) \propto (1-z_j)(\tau_0^2)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left[\left(\sum_{i=1}^n x_{ij}^2/\sigma^2 + 1/\sigma^2\tau_0^2\right)\beta_j^2 - 2\beta_j\left(\sum_{i=1}^n \sum_{j\neq p}(y_i - x_{ij}\beta_j)/\sigma^2\right) + \frac{\left[\sum_{i=1}^n \sum_{j\neq p}(y_i - x_{ij}\beta_j)/\sigma^2\right]^2}{\sum_{i=1}^n x_{ij}^2/\sigma^2 + 1/\sigma^2\tau_0^2} - \frac{\left[\sum_{i=1}^n \sum_{j\neq p}(y_i - x_{ij}\beta_j)^{x_{ip}}/\sigma^2\right]^2}{\sum_{i=1}^n x_{ij}^2/\sigma^2 + 1/\sigma^2\tau_0^2}\right]\right)$$

$$+ (z_j)(\tau_1^2)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left[\left(\sum_{i=1}^n x_{ij}^2/\sigma^2 + 1/\sigma^2\tau_1^2\right)\beta_j^2 - 2\beta_j\left(\sum_{i=1}^n \sum_{j\neq p}(y_i - x_{ij}\beta_j)/\sigma^2\right) + \right.\right.$$

$$+\frac{\left[\sum_{i=1}^{n}\sum_{j\neq p}(y_i - x_{ij}\beta_j)/\sigma^2\right]}{\sum_{i=1}^{n}x_{ij}^2/\sigma^2 + 1/\sigma^2\tau_1^2} - \frac{\left[\sum_{i=1}^{n}\sum_{j\neq p}(y_i - x_{ij}\beta_j)/\sigma^2\right]^2}{\sum_{i=1}^{n}x_{ij}^2/\sigma^2 + 1/\sigma^2\tau_1^2}\Bigg]\Bigg)\Bigg]$$

$$\sim GMM: w_1 = (1 - z_j)(\tau_0^2)^{-\frac{1}{2}} \cdot \exp\left(\frac{\left[\sum_{i=1}^{n}\sum_{j\neq p}(y_i - x_{ij}\beta_j)/\sigma^2\right]^2}{2\left[\sum_{i=1}^{n}x_{ij}^2/\sigma^2 + 1/\sigma^2\tau_0^2\right]}\right]$$

$$\mu_1 = \frac{\sum_{i=1}^{n}\sum_{j\neq p}(y_i - x_{ij}\beta_j)^{x_{ij}}/\sigma^2}{\sum_{i=1}^{n}x_{ij}^2/\sigma^2 + 1/\sigma_{\beta_j}^2\tau_0^2} \qquad \sigma_1^2 = \left(\sum_{i=1}^{n}x_{ij}^2/\sigma^2 + 1/\sigma_{\beta_j}^2\tau_0^2\right)^{-1}$$

$$w_2 = z_j(\tau_1^2)^{-\frac{1}{2}} \exp\left(\frac{\left[\sum_{i=1}^{n}\sum_{j\neq p}(y_i - x_{ij}\beta_j)/\sigma^2\right]^2}{2\left[\sum_{i=1}^{n}x_{ij}^2/\sigma^2 + 1/\sigma^2\tau_1^2\right]}\right]$$

$$\mu_2 = \frac{\sum_{i=1}^{n}\sum_{j\neq p}(y_i - x_{ij}\beta_j)^{x_{ij}}/\sigma^2}{\sum_{i=1}^{n}x_{ij}^2/\sigma^2 + 1/\sigma_{\beta_j}^2\tau_1^2} \qquad \sigma_2^2 = \left(\sum_{i=1}^{n}x_{ij}^2/\sigma^2 + 1/\sigma_{\beta_j}^2\tau_1^2\right)^{-1}$$

It can be simplified by discussing the case that $z_j = 0$ or $z_j = 1$

$$\sigma^2 \sim IG\left(\frac{n}{2}, \frac{\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p}x_{ij}\cdot\beta_j)^2}{2}\right)$$

$$z_i \mid \text{others} \sim Ber\left(\frac{q N(\beta_i, 0, \sigma_{\beta_i}^2 \tau_1^2)}{q N(\beta_i, 0, \sigma_{\beta_i}^2, \tau_1^2) + (1-q) N(\beta_i, 0, \sigma_{\beta_i}^2, \tau_0)}\right)$$

The probability of being included and the posterior mean are as follows:

↓ this is the case that $\tau_0=1$ $\tau_1=1000$

	1	mpg	cyl	disp	hp	drat	wt	vs	am	gear	carb
Probability Included	0.982	0.014	0.048	0.011	0.012	0.040	0.109	0.033	0.025	0.047	0.039
Estimated Coefficient	9.908	0.075	0.011	-0.011	-0.001	0.774	2.130	1.040	-0.817	0.201	-0.530

	1	mpg	cyl	disp	hp	drat	wt	vs	am	gear	carb
2.5%	-106.760	-0.066	-0.774	-0.093	-0.018	-0.827	0.620	-0.150	-1.855	-0.835	-3.924
50%	17.130	0.071	-0.327	-0.008	-0.002	-0.013	1.529	0.883	-0.823	-0.186	-0.314
97.5%	23.318	0.247	6.577	0.001	0.012	5.956	14.435	1.914	0.251	7.929	0.091

τ_0 isn't small enough, therefore is used another case:

$\tau_0=0.01$. $\tau_1=1000$

	1	mpg	cyl	disp	hp	drat	wt	vs	am	gear	carb
Probability Included	1.000	0.025	0.159	0.014	0.015	0.082	0.945	0.863	0.329	0.196	0.146
Estimated Coefficient	17.513	0.016	-0.327	-0.004	-0.008	0.080	1.375	1.219	-0.534	-0.091	-0.174

	1	mpg	cyl	disp	hp	drat	wt	vs	am	gear	carb
2.5%	11.279	-0.072	-1.139	-0.016	-0.021	-0.514	0.015	-0.114	-2.101	-1.327	-0.659
50%	15.944	0.040	-0.050	-0.005	-0.010	-0.023	1.366	1.540	-0.094	-0.061	-0.099
97.5%	32.134	0.154	0.142	0.005	0.007	0.218	2.361	2.652	0.152	0.189	0.110

This time the model performs well. The intercept. wt, vs are significant not 0 from the first line.