

MATH 559: BAYESIAN THEORY AND METHODS

MCMC FOR HIERARCHICAL NON-LINEAR REGRESSION MODELS

Non-linear regression: We consider non-linear regression data from m individuals. Suppose we have the mean model for individual i as

$$\mu(x; \beta_{i0}, \beta_{i1}, \beta_{i2}) = e^{\beta_{i2}} \text{expit}\{\beta_{i0} + \beta_{i1}x\}$$

where $\text{expit}(x) = e^x / (1 + e^x)$, and where it is considered that the parameters $(\beta_{i0}, \beta_{i1}, \beta_{i2})$ for each individual are different, thereby necessitating a hierarchical model: if $i = 1, \dots, m$ indexes individuals in the study, and $j = 1, \dots, n$ indexes the observations made on each individual, then

- **STAGE 1: Observed data:**

$$Y_{ij} \sim \text{Normal}(\mu(x_{ij}; \beta_{i0}, \beta_{i1}, \beta_{i2}), \sigma_i^2) \quad i = 1, \dots, m, j = 1, \dots, n \text{ independently}$$

- **STAGE 2: Population model:** for $i = 1, \dots, m$ independently, assume that

$$(\beta_{i0}, \beta_{i1}, \beta_{i2}) \sim \text{Normal}_3(\mu, \Sigma) \quad \sigma_i^2 \sim \text{InverseGamma}(a/2, b/2).$$

where $\mu \in \mathbb{R}^3$ and Σ is a 3×3 positive definite matrix, and $a, b > 0$ are fixed scalars.

- **STAGE 3: Prior model:** the prior model is the **Normal-Inverse Wishart** model

$$\Sigma \sim \text{InverseWishart}(\nu, \Psi) \quad \mu | \Sigma \sim \text{Normal}_3(\eta, \Sigma/\lambda)$$

where $\eta \in \mathbb{R}^3$, $\nu > 1$ is a scalar and Ψ is a 3×3 positive definite matrix.

MCMC: We now formulate a Gibbs sampler strategy:

1. For the **first stage** parameters conditional on the first stage parameters, we have a standard linear model, and the full conditional posteriors exhibit a conditional independent structure for $i = 1, \dots, m$. However, the analysis is not quite the same as in the non-hierarchical case, and so we again use a Gibbs sampler structure. If $\mathbf{X}_i$ is the $n \times 2$ design matrix for the i th simple linear regression, $\mathbf{y}_i$ is the vector of response data, we have

- $\beta_i | \sigma_i^2, \mu, \Sigma$: we have that the full conditional posterior is proportional to

$$\exp \left\{ -\frac{1}{2\sigma_i^2} (\mathbf{y}_i - \mu(x_i; \beta_{i0}, \beta_{i1}, \beta_{i2}))^\top (\mathbf{y}_i - \mu(x_i; \beta_{i0}, \beta_{i1}, \beta_{i2})) \right\} \exp \left\{ -\frac{1}{2} (\beta_i - \mu)^\top \Sigma^{-1} (\beta_i - \mu) \right\}.$$

where $\mu(x_i; \beta_{i0}, \beta_{i1}, \beta_{i2})$ is the $n \times 1$ vector of means for individual i . In this case, the full conditional posterior is not available in closed form, so we must use the Metropolis-Hastings algorithm: this can be achieved using a block-update of all the parameters for an individual together, or using a Metropolis-within-Gibbs approach where each parameters is updated separately conditional on the others.

- $\sigma_i^2 | \beta_i, \mu, \Sigma$: we have that the full conditional posterior is proportional to

$$\left(\frac{1}{\sigma_i^2} \right)^{n/2} \exp \left\{ -\frac{1}{2\sigma_i^2} (\mathbf{y}_i - \mu(x_i; \beta_{i0}, \beta_{i1}, \beta_{i2}))^\top (\mathbf{y}_i - \mu(x_i; \beta_{i0}, \beta_{i1}, \beta_{i2})) \right\} \left(\frac{1}{\sigma_i^2} \right)^{a/2+1} \exp \left\{ -\frac{b}{2} \right\}$$

and then by inspection, we see that

$$\pi_n(\sigma_i^2 | \beta_i, \mu, \Sigma) \equiv \text{InverseGamma}(a_n/2, b_{ni}/2)$$

and

$$a_n = n + a \quad b_{ni} = (\mathbf{y}_i - \mu(x_{ij}; \beta_{i0}, \beta_{i1}, \beta_{i2}))^\top (\mathbf{y}_i - \mu(x_{ij}; \beta_{i0}, \beta_{i1}, \beta_{i2})) + b$$

These two full conditional distributions can be sampled directly using `rmvn` from the `mvnfast` library, and the `rgamma` function (after reciprocation) respectively.

2. For the **second stage** parameters conditional on the first stage parameters, if $\beta_i = (\beta_{i0}, \beta_{i1})$ are regarded as known quantities, then independently for $i = 1, \dots, n$

$$\beta_i | \mu, \Sigma \sim \text{Normal}_d(\mu, \Sigma).$$

We treat the β_i vectors as pseudo-data, and attempt to perform inference for the second stage “population” parameters. Here, the $NIW(\eta, \lambda, \nu, \Psi)$ prior is conjugate to the corresponding second-stage “likelihood”, and it follows that the posterior distribution for μ and Σ is $NIW(\mu_n, \lambda_n, \nu_n, \Psi_n)$ where

$$\eta_n = \frac{m\bar{\beta} + \lambda\eta}{m + \lambda} \quad \lambda_n = m + \lambda \quad \nu_n = m + \nu$$

and

$$\Psi_n = \sum_{i=1}^m (\beta_i - \bar{\beta})(\beta_i - \bar{\beta})^\top + \frac{m\lambda}{m + \lambda} (\bar{\beta} - \eta)(\bar{\beta} - \eta)^\top + \Psi \quad \bar{\beta} = \frac{1}{m} \sum_{i=1}^m \beta_i.$$

and η, ν and Ψ are hyperparameters. Then we have the results

$$\pi_n(\mu | \Sigma, -) \equiv \text{Normal}(\eta_n, \Sigma / \lambda_n) \quad \pi_n(\Sigma | -) \equiv \text{InverseWishart}(\nu_n, \Psi_n).$$

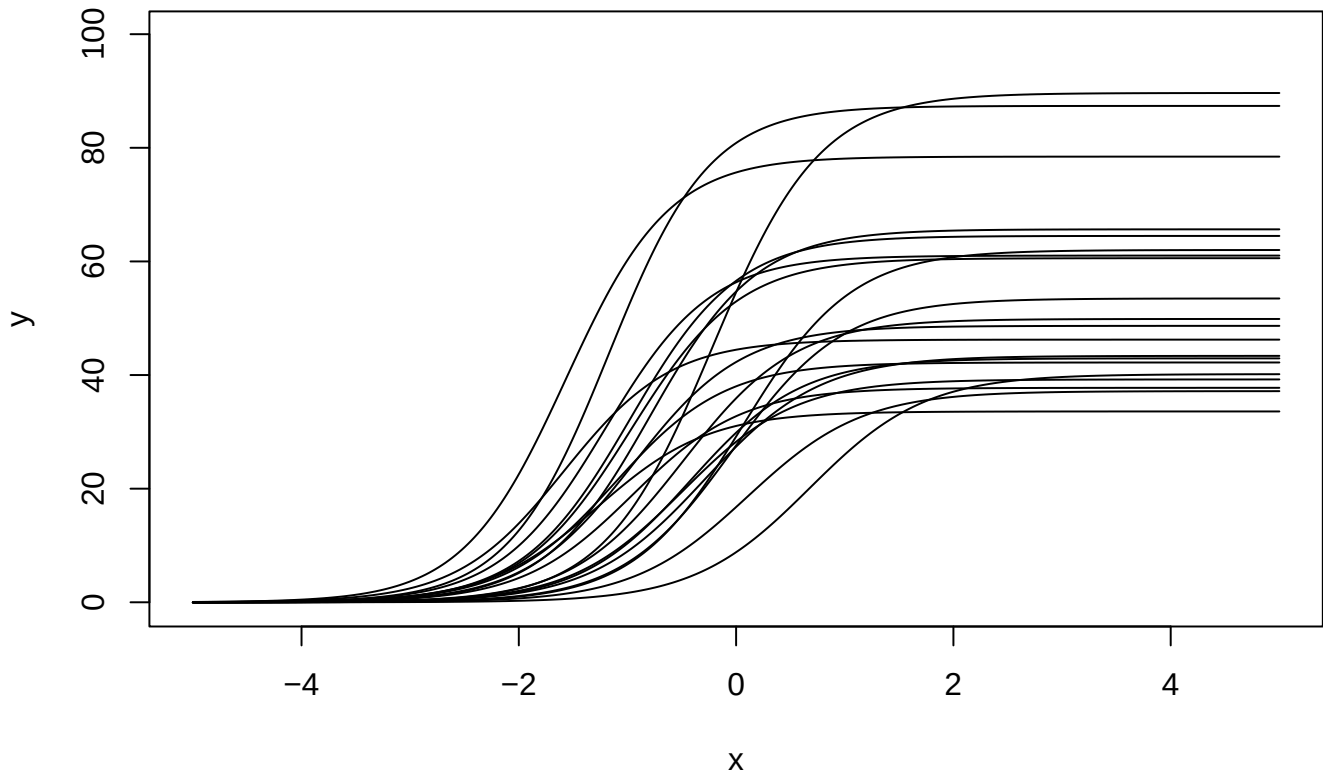
These two full conditional distributions can be sampled directly using `rmvn` from the `mvnfast` library, and the `riwish` function from the `MCMCpack` library.

We examine the performance of the algorithm for the following simulated data.

```
set.seed(2300)
#Simulation
library(mvnfast)
m<-20
mu<-c(1.5,2.0,4)
Cor.mat<-matrix(c(1,0.75,0.25,0.75,1,0.25,0.25,0.75,1),3,3)
Sigma<-2*diag(c(1,0.05,0.2))%*% (Cor.mat%*%diag(c(1,0.05,0.2)))
be<-rmvn(m,mu,Sigma)
ysd<-sqrt(1/rgamma(m,4,4))

expit<-function(x){return(exp(x)/(1+exp(x)))}
mu.func<-function(x,b0,b1,b2){
  return(exp(b2)*expit(b0+b1*x))
}

xv<-seq(-5,5,by=0.01)
x<-seq(-5,5.0,by=0.5)
n<-length(x)
par(mar=c(4,4,1,0))
plot(xv,xv,type='n',ylim=range(-0.25,100),xlab='x',ylab='y')
y<-matrix(0,nrow=m,ncol=n)
for(i in 1:m){
  muv<-mu.func(x,be[i,1],be[i,2],be[i,3])
  lines(xv,mu.func(xv,be[i,1],be[i,2],be[i,3]))
  y[i,]<-muv+rnorm(n)*ysd[i]
}
```



The Gibbs sampler proceeds as follows.

```
library(MCMCpack)
a<-b<-1
eta<-c(0,0,0)
lambda<-0.1
nu<-4
Psi<-nu*diag(rep(1,3))

#Starting values via Non-linear least squares
old.beta<-matrix(0,nrow=m,ncol=3)
old.sigsq<-old.ssq<-old.prior<-rep(0,m)
for(i in 1:m){
  yi<-y[i,]
  fit0<-nls(yi~exp(b2+b0+b1*x)/(1+exp(b0+b1*x)),start=list(b0=0.5,b1=1,b2=4))
  old.beta[i,]<-coef(fit0)
  old.sigsq[i]<-summary(fit0)$sigma^2
}
old.mu<-apply(old.beta,2,mean)
old.Sigma<-var(old.beta)

for(i in 1:m){
  old.ssq[i]<-sum((y[i,]-mu.func(x,old.beta[i,1],old.beta[i,2],old.beta[i,3]))^2)
  old.prior[i]<-dmvnm(old.beta[i,],old.mu,old.Sigma,log=T)
}

nsamp<-2000; nburn<-50; nthin<-1
nits<-nburn+nsamp*ntthin
ico<-0

beta.samp<-array(0,c(nsamp,m,3))
sigsq.samp<-matrix(0,nrow=nsamp,ncol=m)
mu.samp<-matrix(0,nrow=nsamp,ncol=3)
Sigma.samp<-array(0,c(nsamp,3,3))
```

```

mh.sig<-c(0.05,0.05,0.005)
for(iter in 1:nits){

  #Update the betas using Gibbs sampler
  old.Siginv<-solve(old.Sigma)
  new.ssq<-new.prior<-rep(0,m)
  for(i in 1:m){
    for(k in 1:3){
      new.beta<-old.beta[i,]
      new.beta[k]<-old.beta[i,k]+rnorm(1)*mh.sig[k]
      new.ssq[i]<-sum((y[i,]-mu.func(x,new.beta[1],new.beta[2],new.beta[3]))^2)
      new.prior[i]<-dmvn(new.beta,old.mu,old.Sigma,log=T)
      if(log(runif(1)) < -0.5*new.ssq[i]/old.sigsq[i]+new.prior[i]+
        0.5*old.ssq[i]/old.sigsq[i]-old.prior[i]){
        old.beta[i,k]<-new.beta[k]
        old.ssq[i]<-new.ssq[i]
        old.prior[i]<-new.prior[i]
      }
    }
  }

  #Update the sigmas
  a.n<-a+m
  b.n<-old.ssq+b
  old.sigsq<-1/rgamma(m,a.n/2,b.n/2)

  #Update Sigma
  nu.n<-nu+m
  beta.bar<-apply(old.beta,2,mean)
  Psi.n<-Psi+((m*lambda)/(m+lambda))*((beta.bar-eta) %*% t(beta.bar-eta))+var(old.beta)*(m-1)
  old.Sigma<-riwish(nu.n,Psi.n)

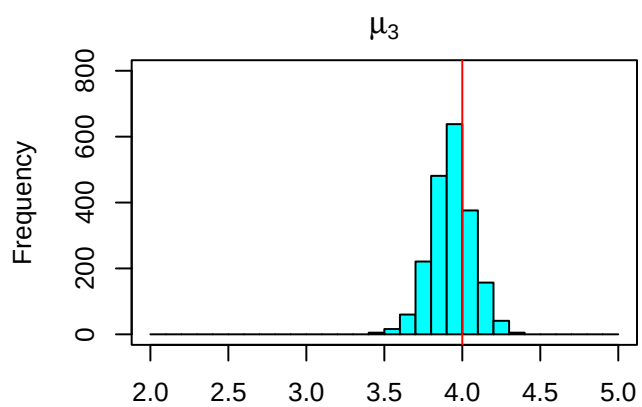
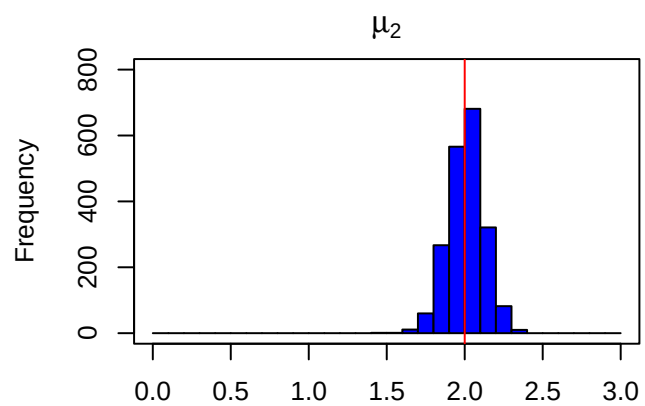
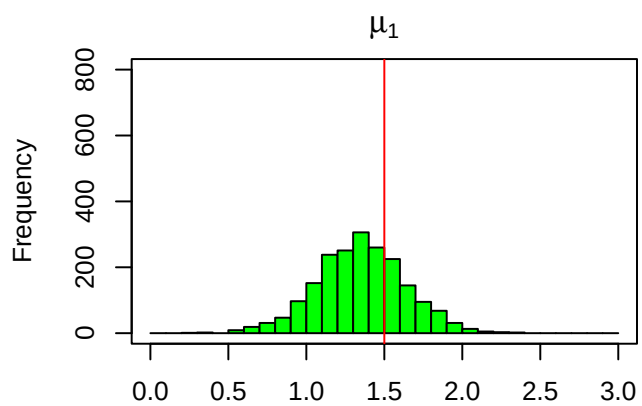
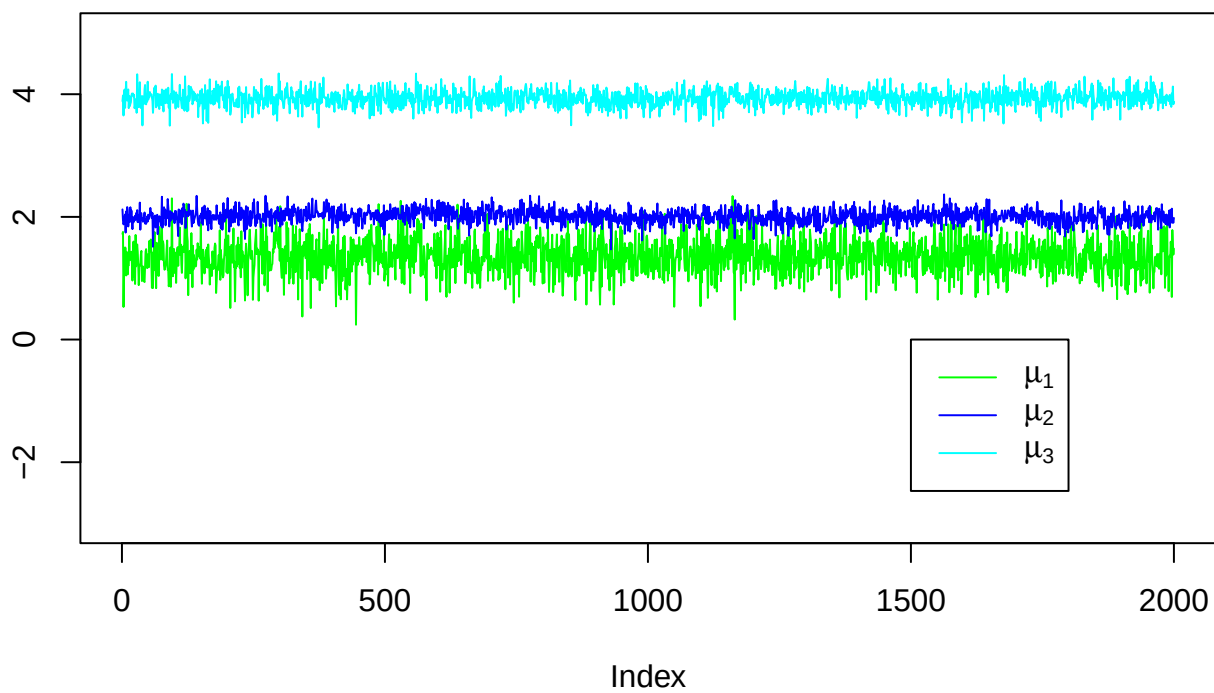
  #Update mu given Sigma
  lambda.n<-lambda+m
  eta.n<-(m*beta.bar+lambda*eta)/(m+lambda)
  old.mu<-matrix(rmvn(1,eta.n,old.Sigma/lambda.n),ncol=1)

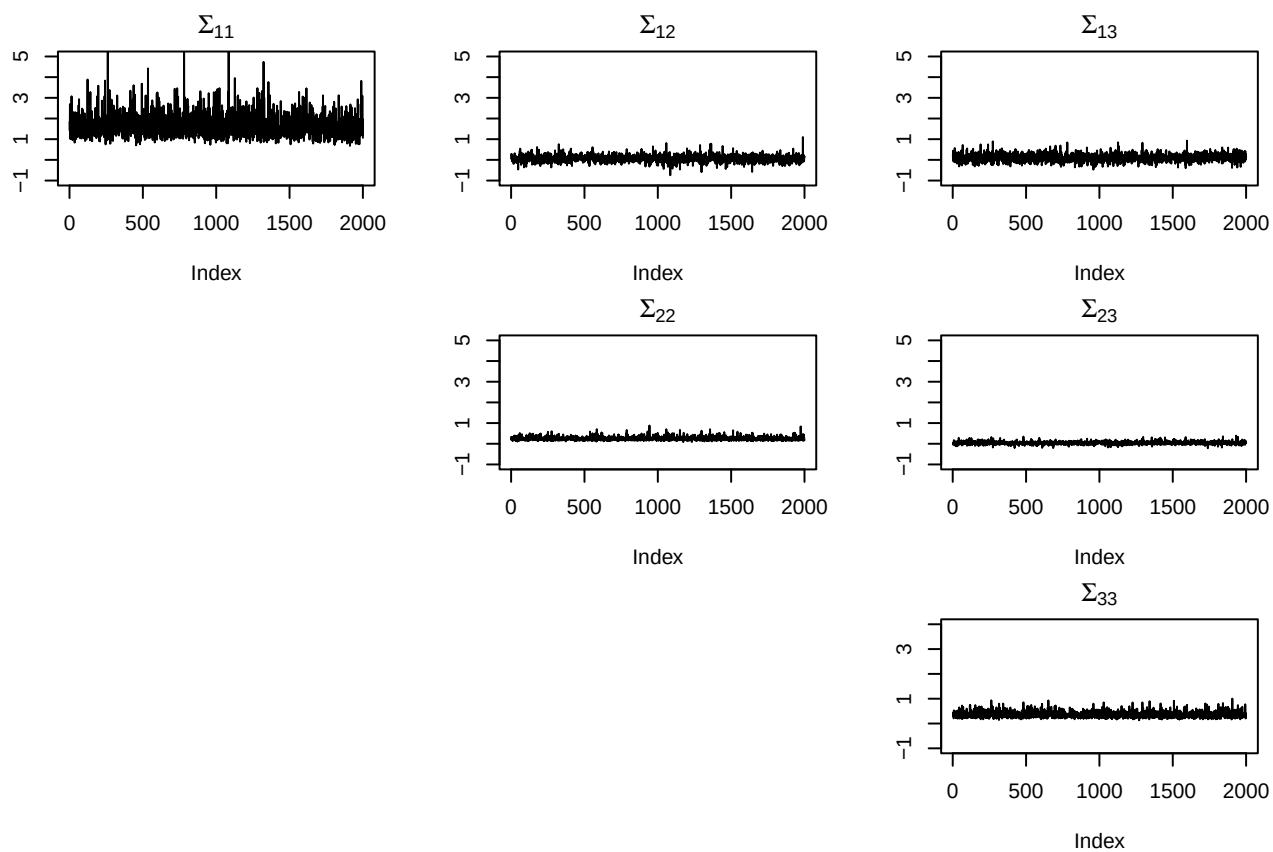
  if(iter > nburn & iter %% nthin == 0){
    ico<-ico+1
    beta.samp[ico,]<-old.beta
    sigsq.samp[ico,]<-old.sigsq
    mu.samp[ico,]<-old.mu
    Sigma.samp[ico,]<-old.Sigma
  }

  for(i in 1:m){
    old.ssq[i]<-sum((y[i,]-mu.func(x,old.beta[i,1],old.beta[i,2],old.beta[i,3]))^2)
    old.prior[i]<-dmvn(old.beta[i,],old.mu,old.Sigma,log=T)
  }
}

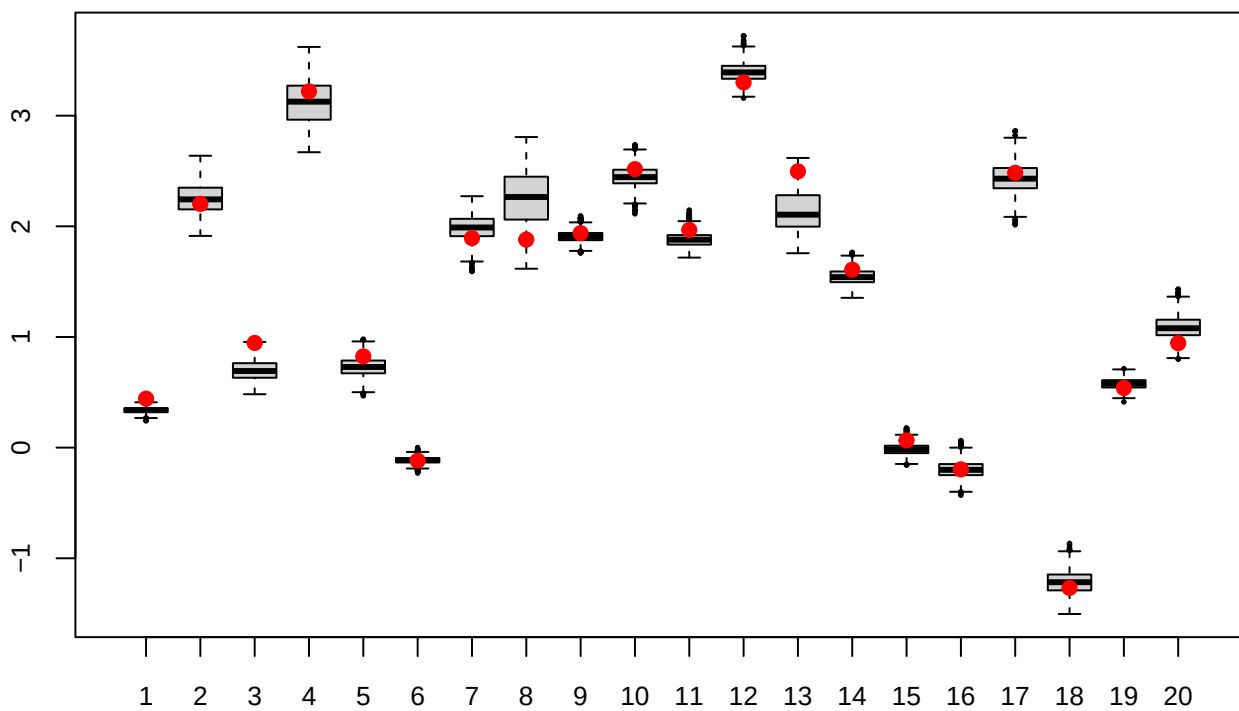
```

Trace plots of μ parameters

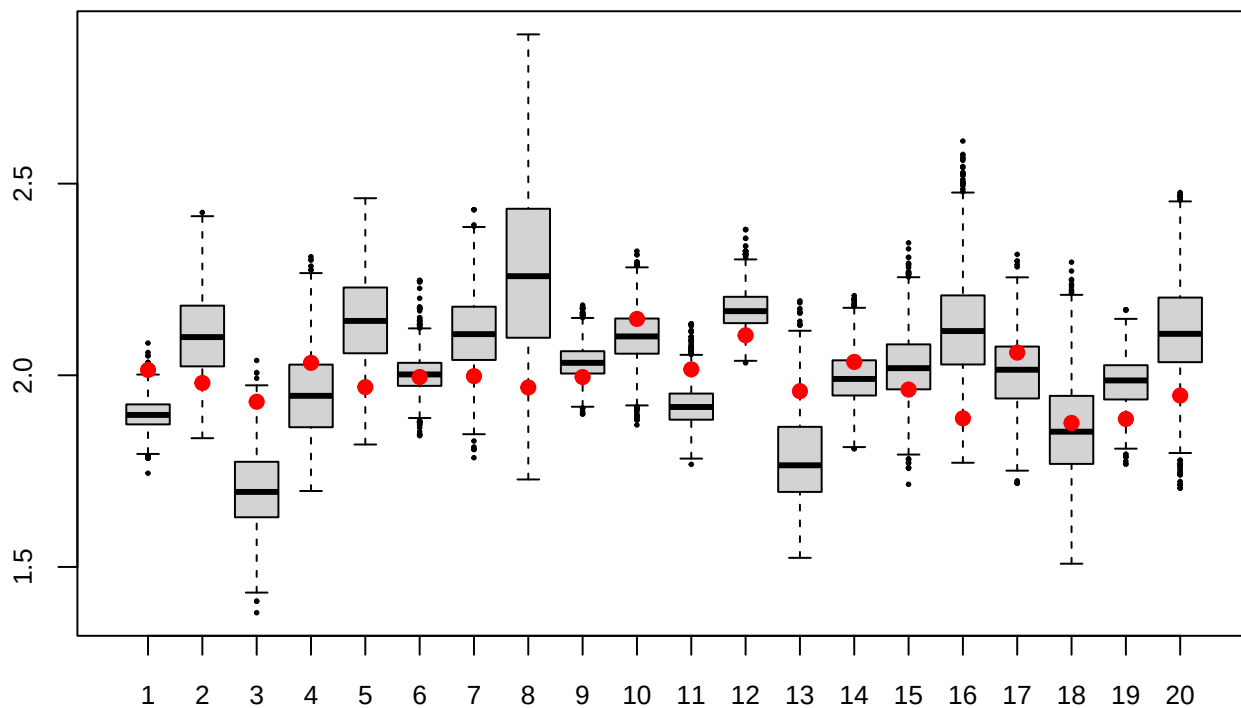




Boxplots of β_{i0} parameters



Boxplots of β_{i1} parameters



Boxplots of β_{i2} parameters

