1. (30 points)

(a) (10 points) $y \mid \mu \sim N(\mu, \sigma^2 = 25)$, $\mu \sim N(\mu_0 = -5, \tau_0^2 = 100)$.
The marginal likelihood function of $\boldsymbol{y}$ is

$$
\begin{aligned}
m(\boldsymbol{y}) &= \int_{-\infty}^{\infty} p(\boldsymbol{y} \mid \mu)\pi(\mu)d\mu \\
&= \int_{-\infty}^{\infty} \frac{1}{(\sigma\sqrt{2\pi})^n} \exp\left\{-\frac{\sum(y_i-\mu)^2}{2\sigma^2}\right\} \cdot \frac{1}{\tau_0\sqrt{2\pi}} \exp\left\{-\frac{(\mu-\mu_0)^2}{2\tau_0^2}\right\} d\mu \\
&= \frac{1}{(\sigma\sqrt{2\pi})^n}\frac{1}{\tau_0\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\left(\frac{n}{2\sigma^2}+\frac{1}{2\tau_0^2}\right)\mu^2 + \left(\frac{\sum y_i}{\sigma^2}+\frac{\mu_0}{\tau_0^2}\right)\mu - \frac{\sum y_i^2}{2\sigma^2} - \frac{\mu_0^2}{2\tau_0^2}\right\} d\mu \\
&= \frac{1}{(\sigma\sqrt{2\pi})^n}\frac{1}{\tau_0\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\left(\frac{n}{2\sigma^2}+\frac{1}{2\tau_0^2}\right)\left(\mu - \frac{\frac{\sum y_i}{\sigma^2}+\frac{\mu_0}{\tau_0^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau_0^2}}\right)^2 - \frac{\sum y_i^2}{2\sigma^2} - \frac{\mu_0^2}{2\tau_0^2} + \frac{\left(\frac{\sum y_i}{\sigma^2}+\frac{\mu_0}{\tau_0^2}\right)^2}{\frac{2n}{\sigma^2}+\frac{2}{\tau_0^2}}\right\} d\mu \\
&= \frac{1}{(\sigma\sqrt{2\pi})^n}\frac{1}{\tau_0\sqrt{2\pi}} \exp\left\{-\frac{\sum y_i^2}{2\sigma^2} - \frac{\mu_0^2}{2\tau_0^2} + \frac{\left(\frac{\sum y_i}{\sigma^2}+\frac{\mu_0}{\tau_0^2}\right)^2}{\frac{2n}{\sigma^2}+\frac{2}{\tau_0^2}}\right\} \int_{-\infty}^{\infty} -\left(\frac{n}{2\sigma^2}+\frac{1}{2\tau_0^2}\right)\left(\mu - \frac{\frac{\sum y_i}{\sigma^2}+\frac{\mu_0}{\tau_0^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau_0^2}}\right)^2 d\mu \\
&= \frac{1}{(\sigma\sqrt{2\pi})^n}\frac{1}{\tau_0\sqrt{2\pi}} \exp\left\{-\frac{\sum y_i^2}{2\sigma^2} - \frac{\mu_0^2}{2\tau_0^2} + \frac{\left(\frac{\sum y_i}{\sigma^2}+\frac{\mu_0}{\tau_0^2}\right)^2}{\frac{2n}{\sigma^2}+\frac{2}{\tau_0^2}}\right\} \sqrt{2\pi\frac{1}{\frac{n}{\sigma^2}+\frac{1}{\tau_0^2}}} \\
&= \frac{1}{(\sigma\sqrt{2\pi})^n}\frac{1}{\sqrt{\frac{n\tau_0^2}{\sigma^2}+1}} \exp\left\{-\frac{\sum y_i^2}{2\sigma^2} - \frac{\mu_0^2}{2\tau_0^2} + \frac{\left(\frac{\sum y_i}{\sigma^2}+\frac{\mu_0}{\tau_0^2}\right)^2}{\frac{2n}{\sigma^2}+\frac{2}{\tau_0^2}}\right\}
\end{aligned}
$$

Plug in the values we get $m(\boldsymbol{y}) = 5.641 \times 10^{-18}$.

(b) (10 points) Bayes factor

$$
\begin{aligned}
BF_{01} &= \frac{p(\boldsymbol{y} \mid \mu = -4)}{\int_{-\infty}^{\infty} p(\boldsymbol{y} \mid \mu)\pi(\mu)d\mu} = \frac{\frac{1}{(\sigma\sqrt{2\pi})^n}\exp\left\{-\frac{\sum(y_i-\mu)^2}{2\sigma^2}\right\}}{m(\boldsymbol{y})} \\
&= \frac{\frac{1}{(5\sqrt{2\pi})^{12}}\exp\left\{-\frac{\sum(y_i+4)^2}{50}\right\}}{m(\boldsymbol{y})} \\
&= \frac{3.168 \times 10^{-18}}{5.641 \times 10^{-18}} = 0.562.
\end{aligned}
$$

(c) (10 points) The posterior probability of H_0:

$$
\begin{aligned}
p(H_0 \mid D) &= \frac{p(D \mid H_0)p(H_0)}{p(D \mid H_0)p(H_0) + p(D \mid H_1)p(H_1)} \\
&= \frac{\frac{p(D|H_0)}{p(D|H_1)}p(H_0)}{\frac{p(D|H_0)}{p(D|H_1)}p(H_0) + p(H_1)} = \frac{BF_{01}p(H_0)}{BF_{01}p(H_0) + p(H_1)} \\
&= \frac{(0.562)(0.4)}{(0.562)(0.4) + 0.6} = 0.273.
\end{aligned}
$$

2. (30 points)

(a) (10 points) Implement the Metropolis algorithm.
The posterior,

$$\begin{aligned} p(\theta \mid \boldsymbol{y}) &\propto p(\boldsymbol{y} \mid \theta)p(\theta) \\ &\propto \theta^{y_1}(\frac{1}{4}-\theta)^{y_2}(\frac{1}{4}+\theta)^{y_3}(\frac{1}{2}-\theta)^{y_4} \end{aligned}$$

```
y <- c(16,12,41,31)
pf <- function(theta){
  theta^y[1]*(1/4-theta)^y[2]*(1/4+theta)^y[3]*(1/2-theta)^y[4]
}
l <- 500
m <- 2000
theta <- rep(NA,m+l)
theta[1]=0.1
n  <- 1
for(k in 2:(l+m)){
  thetanew <- runif(1,0,1/4)
  theta[k] <- theta[k-1]
  r <- min(pf(thetanew)/pf(theta[k-1]),1)
  if(runif(1)<r){
    theta[k] <- thetanew
    n <- n+1
  }
}
thetas <- theta[(l+1):(l+m)]
n/(l+m)
```

The moving rate is 25.88%.

(b) (10 points) Implement Gibbs sample.
Use two latent variables z_1 and z_2 as follows: split the third cell into two with probabilities $\frac{1}{4}$ and θ and split the fourth cell into two with probabilities $\frac{1}{4}$ and $\frac{1}{4}-\theta$.

y_1	y_2	z_1	$y_3 - z_1$	z_2	$y_4 - z_2$
θ	$\frac{1}{4}$ - θ	$\frac{1}{4}$	θ	$\frac{1}{4}$	$\frac{1}{4}$ - θ

Then,

$$\begin{aligned} p(\theta \mid \boldsymbol{y}, \boldsymbol{z}) &\propto \theta^{y_1}(\frac{1}{4}-\theta)^{y_2}(\frac{1}{4})^{z_1}\theta^{y_3-z_1}(\frac{1}{4})^{z_2}(\frac{1}{4}-\theta)^{y_4-z_2} \\ &\propto \theta^{y_1+y_3-z_1}(\frac{1}{4}-\theta)^{y_2+y_4-z_2} \\ &\propto (4\theta)^{y_1+y_3-z_1}(1-4\theta)^{y_2+y_4-z_2} \end{aligned}$$

So, if $\eta = 4\theta$, then the full conditional distributions of (z_1, z_2, η) is

$$\begin{aligned}
(\eta \mid \boldsymbol{y}, \boldsymbol{z}) &\propto \eta^{y_1+y_3-z_1}(1-\eta)^{y_2+y_4-z_2} \sim Beta(y_1+y_3-z_1+1, y_2+y_4-z_2+1), \\
(z_1 \mid \eta, \boldsymbol{y}) &\propto Bin(y_3, \frac{1/4}{1/4+\theta}) = Bin(y_3, \frac{1}{1+\eta}), \\
(z_2 \mid \eta, \boldsymbol{y}) &\propto Bin(y_4, \frac{1/4}{1/2-\theta}) = Bin(y_4, \frac{1}{2-\eta}).
\end{aligned}$$

```
l <- 500
m <- 2000
eta <- rep(NA,m+l)
z1 <- rep(NA,m+l)
z2 <- rep(NA,m+l)
z1[1] = z2[1] = 0
eta[1] <- 0.1
for(k in 2:(l+m)){
  eta[k]<-rbeta(1,y[1]+y[3]-z1[k-1]+1,y[2]+y[4]-z2[k-1]+1)
  z1[k]<-rbinom(1,y[3],1/(1+eta[k]))
  z2[k]<-rbinom(1,y[4],1/(2-eta[k]))
}
thetas<-eta[(l+1):(l+m)]/4
z1s<-z1[(l+1):(l+m)]
z2s<-z2[(l+1):(l+m)]
hist(thetas,prob=T,nclass=30)

t<-seq(0,0.25,0.001)
n0<-length(t)
p<-rep(NA,n0)
for(i in 1:n0){
  p[i]<-mean(dbeta(4*t[i],y[1]+y[3]-z1s+1,y[2]+y[4]-z2s+1))*4
}
lines(t,p,lty=2)
```

(c) (10 points)

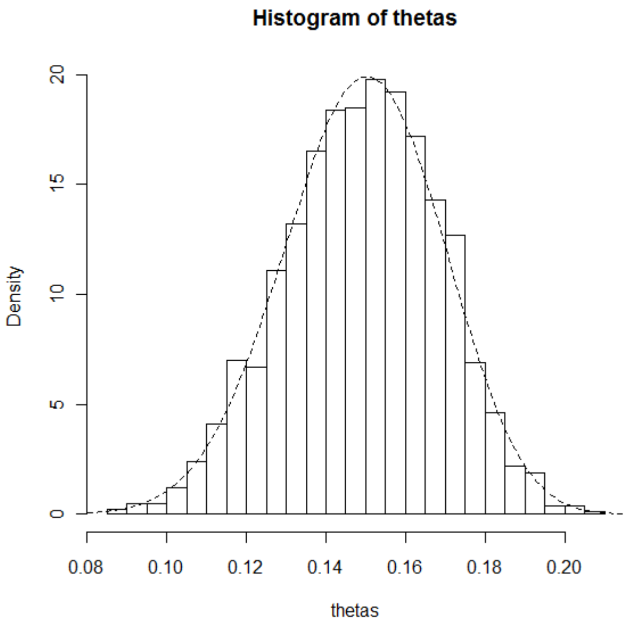
```
mean(thetas)
var(thetas)
```

$$\begin{aligned}
E(\theta \mid \boldsymbol{y}) &= 0.1491752 \\
Var(\theta \mid \boldsymbol{y}) &= 0.0003777287.
\end{aligned}$$

3. (40 points)

(a)(c) (20 points)

```
model;
{
inv_delta0 ~ dgamma(a0,b0)
```

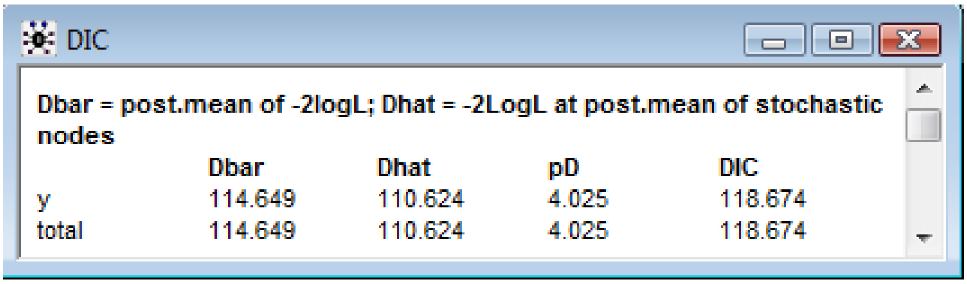



```
delta0 <- 1 / inv_delta0
inv_tau2 <- 1 / tau2
theta ~ dnorm(20,inv_tau2)

for( i in 1 : 3 ) {
mu[i] ~ dnorm(theta,inv_delta1)
for( j in 1 : N ) {
y[j , i] ~ dnorm(mu[i],inv_delta0)
}
}
inv_delta1 ~ dgamma(a1,b1)
delta1 <- 1 / inv_delta1
}
#data
list(N=10,tau2=100,a1=1,a0=1,b0=1,b1=1,
y=structure(
.Data=c(23,28,23,
25,27,20,
21,27,25,
22,29,21,
21,26,22,
22,29,23,
20,27,21,
23,30,20,
19,28,19,
22,27,20),
.Dim=c(10,3))
)
```

```
#initial
list(theta=30,inv_delta0=0.5,inv_delta1=0.5,mu=c(0,0,0))
```

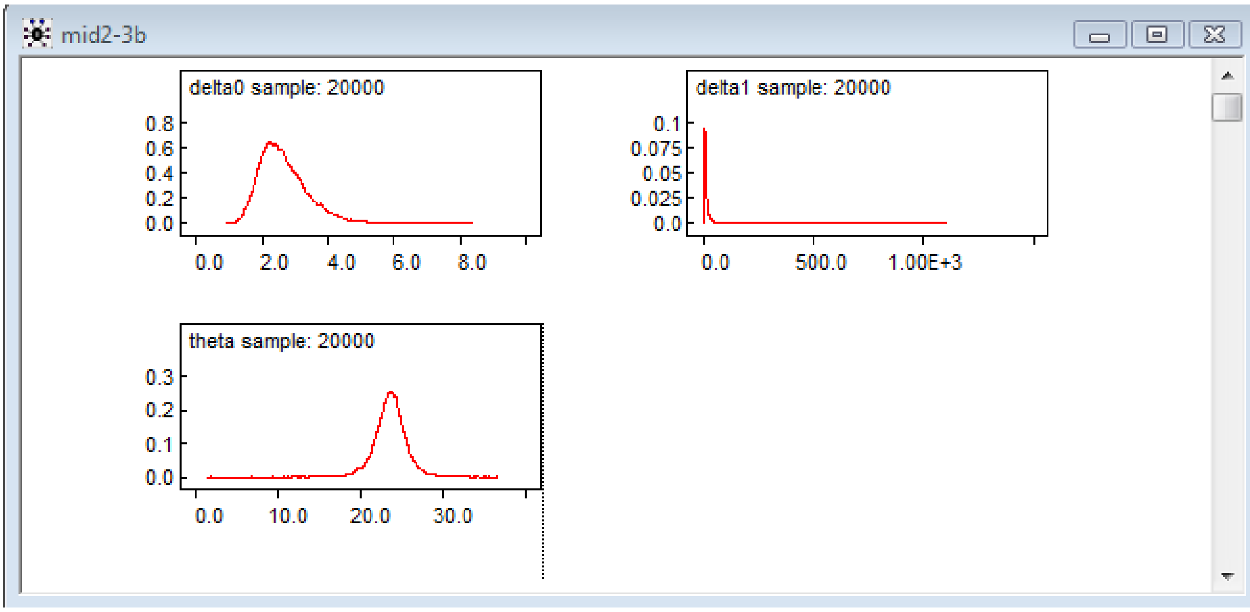
node	mean	sd	error	5.0%	95.0%	start	sample
delta0	2.668	0.7591	0.005535	1.688	4.077	1001	20000
delta1	12.54	20.89	0.1711	2.543	35.8	1001	20000
theta	23.51	1.956	0.01454	20.38	26.5	1001	20000


DIC

Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes

	Dbar	Dhat	pD	DIC
y	114.649	110.624	4.025	118.674
total	114.649	110.624	4.025	118.674

(b) (10 points)



(d) (10 points)
Consider the model under $H_0 : \delta_1 = 0$:

```
model;
{
inv_delta0 ~ dgamma(a0,b0)
delta0 <- 1 / inv_delta0
inv_tau2 <- 1 / tau2
theta ~ dnorm(20,inv_tau2)
```

```
for( i in 1 : 3 ) {
for( j in 1 : N ) {
y[j , i] ~ dnorm(theta,inv_delta0)
}
}
}
```

DIC

Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes

	Dbar	**Dhat**	**pD**	**DIC**
y	158.744	156.783	1.961	160.705
total	158.744	156.783	1.961	160.705

$DIC = 160.705$ for $H_0 > DIC = 118.674$ for H_a. Therefore, $H_a : \delta_1 \neq 0$ is better.