HW 7

Problem 1

(1)

The posterior is

$$
\begin{aligned}
p(\theta|y) &\propto p(\theta)p(y|\theta) \\
&\propto \prod_{i=1}^{2} \frac{1}{\pi\left\{1+(y_i-\theta)^2\right\}}
\end{aligned}
$$

So the integral is

$$
\begin{aligned}
&\int_{-\infty}^{\infty} p(\theta|y)\ d\theta \\
&\propto \int_{-\infty}^{\infty} \prod_{i=1}^{2} \frac{1}{\pi\left\{1+(y_i-\theta)^2\right\}}\ d\theta \\
&\leq \int_{-\infty}^{\infty} \frac{1}{\pi\left\{1+(y_1-\theta)^2\right\}}\ d\theta \\
&= 1.
\end{aligned}
$$

The last line is because we form a density of Cauchy distribution in the integral.

(2)

Take derivative w.r.t. θ, we have

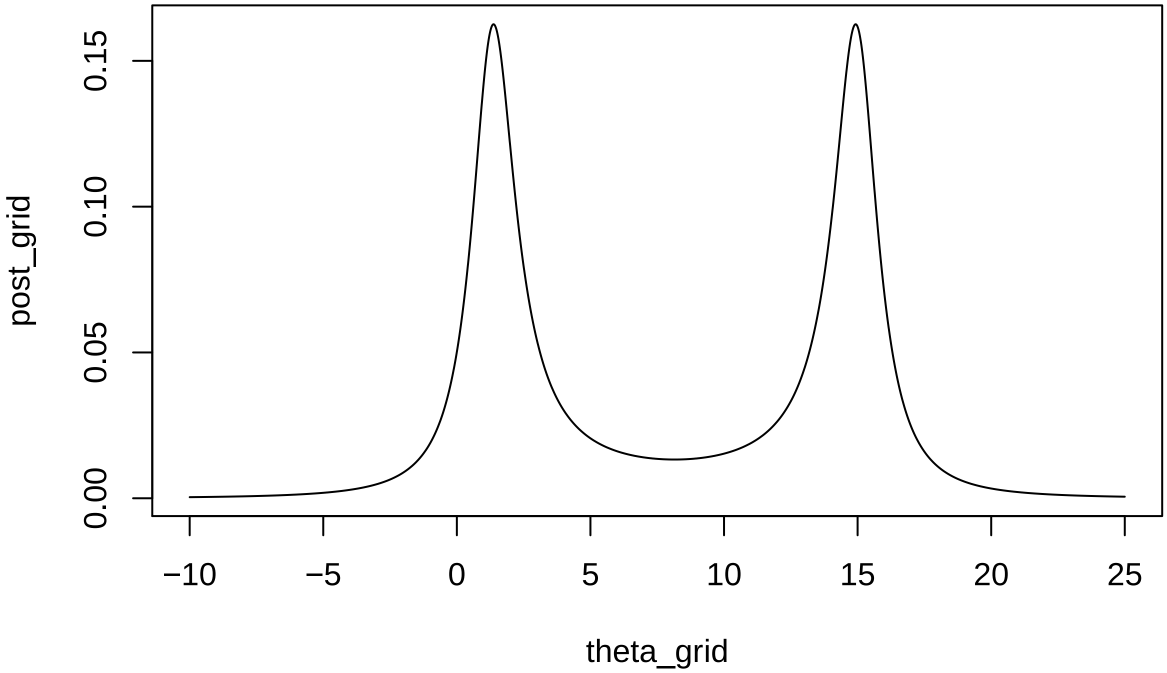
$$
\frac{dp(\theta|y)}{d\theta} = -\frac{2}{\pi^2} \frac{\left[1+(y_1-\theta)^2\right](\theta-y_2)+\left[1+(y_2-\theta)^2\right](\theta-y_1)}{\left[1+(y_1-\theta)^2\right]^2\left[1+(y_2-\theta)^2\right]^2}
$$

When $y_1 = y_2$ we have that $\frac{dp(\theta|y)}{d\theta} = 0$ only at $\theta = y_1 = y_2$ so the density is unimodal.

(3)

```
set.seed(0)
y1 <- 1.3
y2 <- 15.0
post <- function(theta) {
  1/pi^2/(1+(y1-theta)^2)/(1+(y2-theta)^2)
}

normalize_const <- 1/integrate(post, -Inf, Inf)$value

post <- function(theta) {
  normalize_const * 1/pi^2/(1+(y1-theta)^2)/(1+(y2-theta)^2)
}
```

```r
theta_grid <- seq(-10, 25, .01)
post_grid <- post(theta_grid)

plot(theta_grid, post_grid, type='l')
```



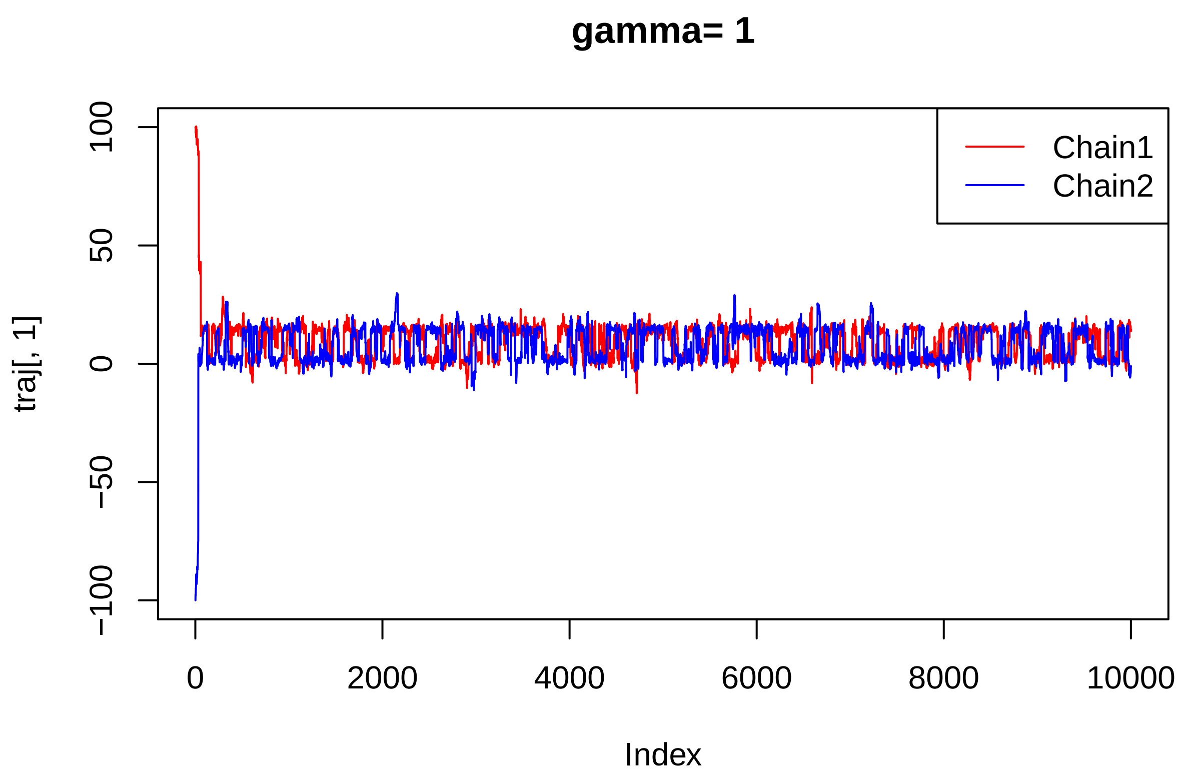
(4)

Let the jump distribution be $J_t(a|b) = \frac{1}{\pi\gamma}\frac{1}{1+(a-b)^2/\gamma^2}$. Obviously $J_t(a|b) = J_t(b|a)$ for all a, b since $(a-b)^2 = (b-a)^2$. The scale parameter γ is of our choice.

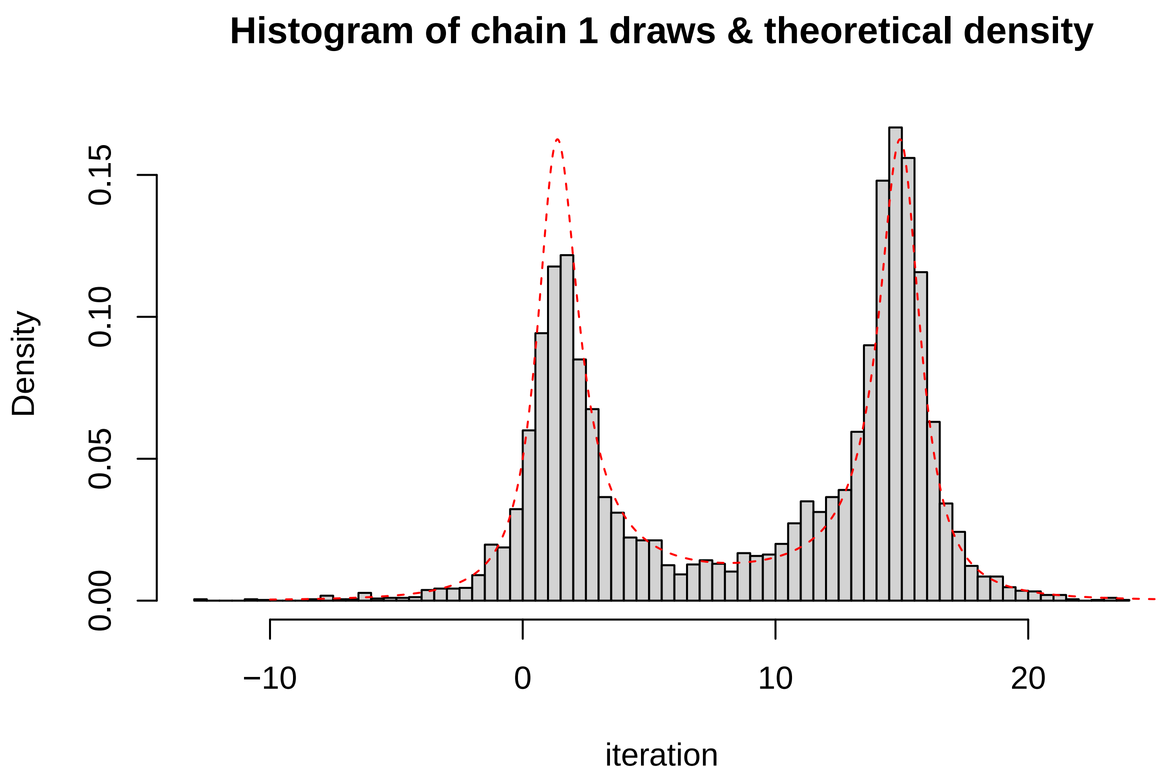
```r
problem1 <- function(gamma, n, init) {
  traj <- matrix(nrow=n+1, ncol=2)
  traj[1, ] <- init

  accept <- matrix(0, nrow=n, ncol=2)
  for (i in 1:n) {
    temp_theta <- rcauchy(1, location=traj[i, 1], scale=gamma)
    log_ratio <- min(0, log(post(temp_theta)) - log(post(traj[i, 1])))
    log_U <- log(runif(1))
    if (log_U <= log_ratio) {
      traj[i+1, 1] <- temp_theta
      accept[i, 1] <- 1
    } else {
      traj[i+1, 1] <- traj[i, 1]
    }
    temp_theta <- rcauchy(1, location=traj[i, 2], scale=gamma)
    log_ratio <- min(0, log(post(temp_theta)) - log(post(traj[i, 2])))
    log_U <- log(runif(1))
    if (log_U <= log_ratio) {
      traj[i+1, 2] <- temp_theta
      accept[i, 2] <- 1
    } else {
```

```r
      traj[i+1, 2] <- traj[i, 2]
    }
  }
  plot(traj[, 1], type='l', col='red', ylim=c(-100, 100), main=paste('gamma=', gamma))
  lines(traj[, 2], col='blue')
  legend('topright', col=c('red', 'blue'), lty=c(1, 1), legend=c('Chain1', 'Chain2'))
  print(paste("Chain1 acceptance rate:", mean(accept[, 1])))
  print(paste("Chain2 acceptance rate:", mean(accept[, 2])))
  hist(tail(traj[, 1], -.2*n), freq=F, breaks=60, main='Histogram of chain 1 draws & theoretical densit
  lines(theta_grid, post_grid, lty=2, col='red')
}
```

```r
problem1(1, 1e4, c(100, -100))
```

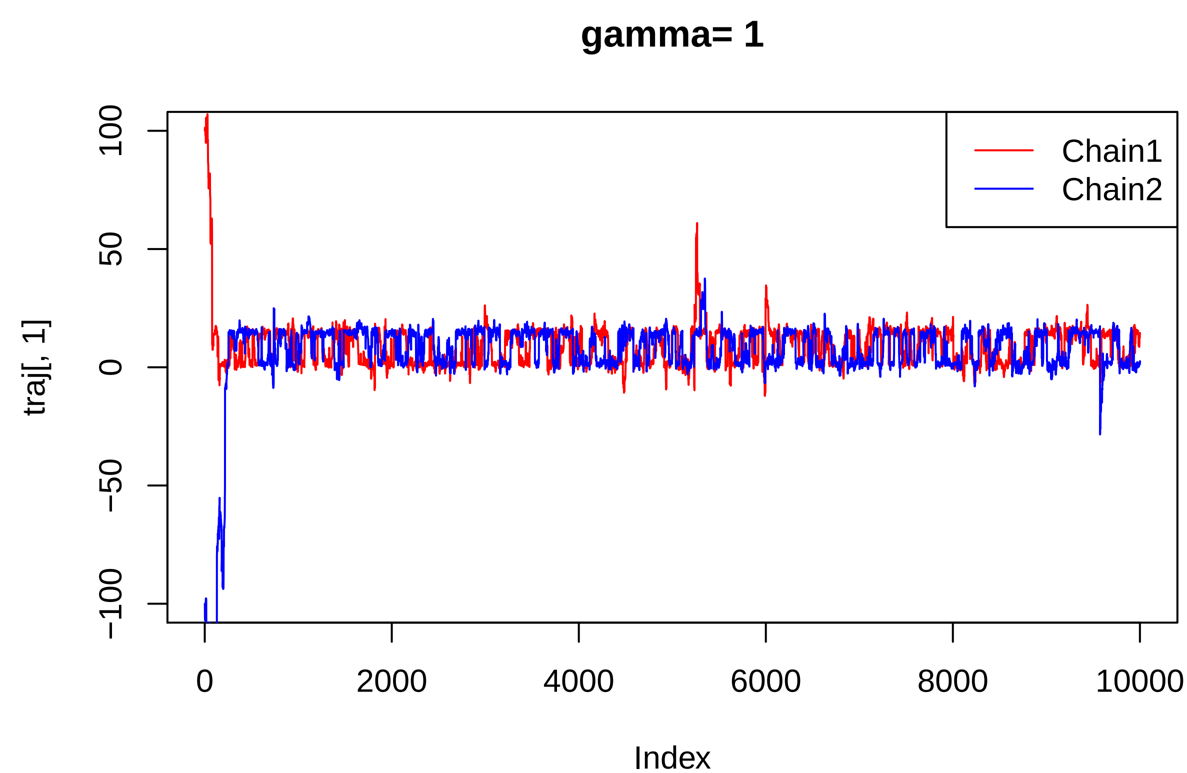



```
## [1] "Chain1 acceptance rate: 0.6648"
## [1] "Chain2 acceptance rate: 0.6612"
```

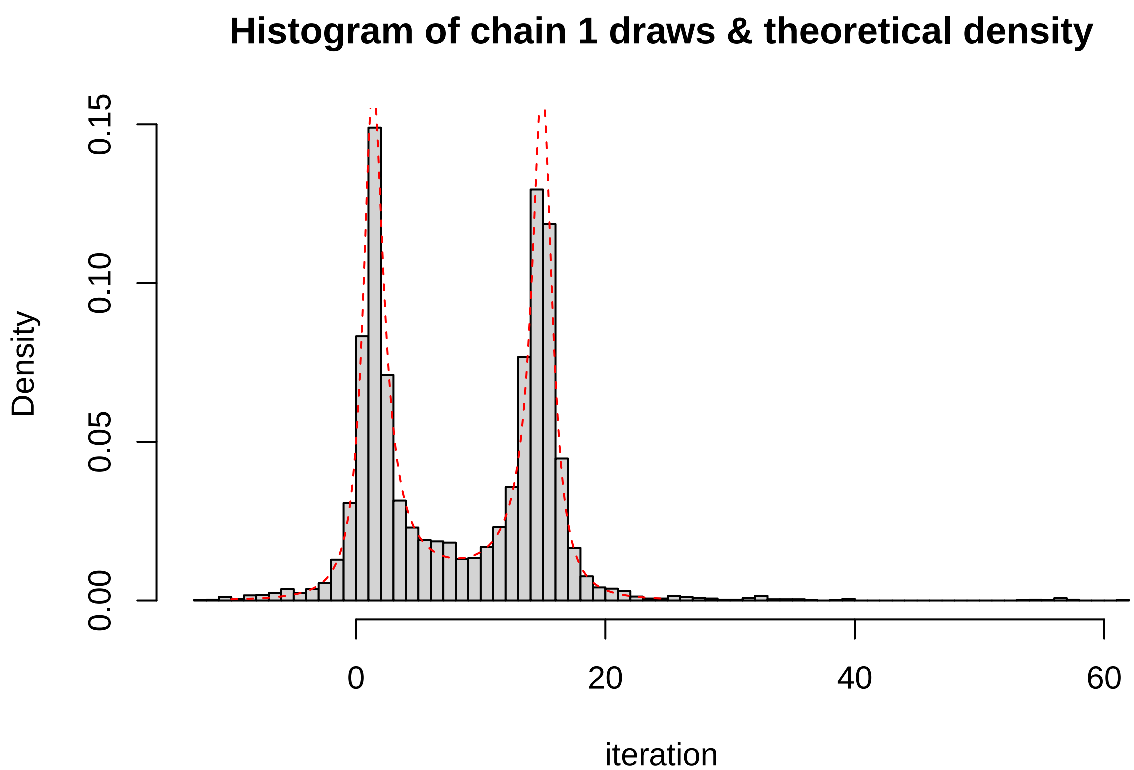


From the traceplot we see that starting from very different points, two chains stably converged. The posterior draws do not perfectly recover the theoretical density, there are fewer sample from the left mode.

```r
problem1(1, 1e4, c(100, -100))
```

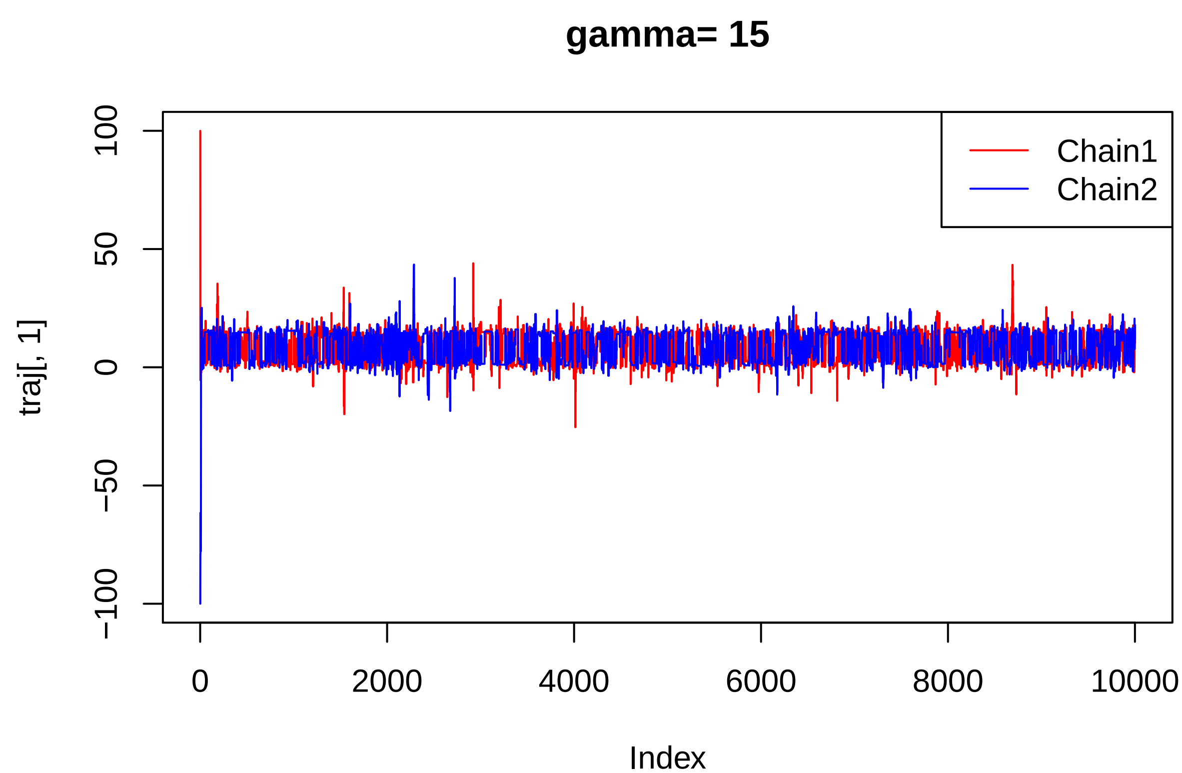



```
## [1] "Chain1 acceptance rate: 0.6645"
## [1] "Chain2 acceptance rate: 0.6641"
```

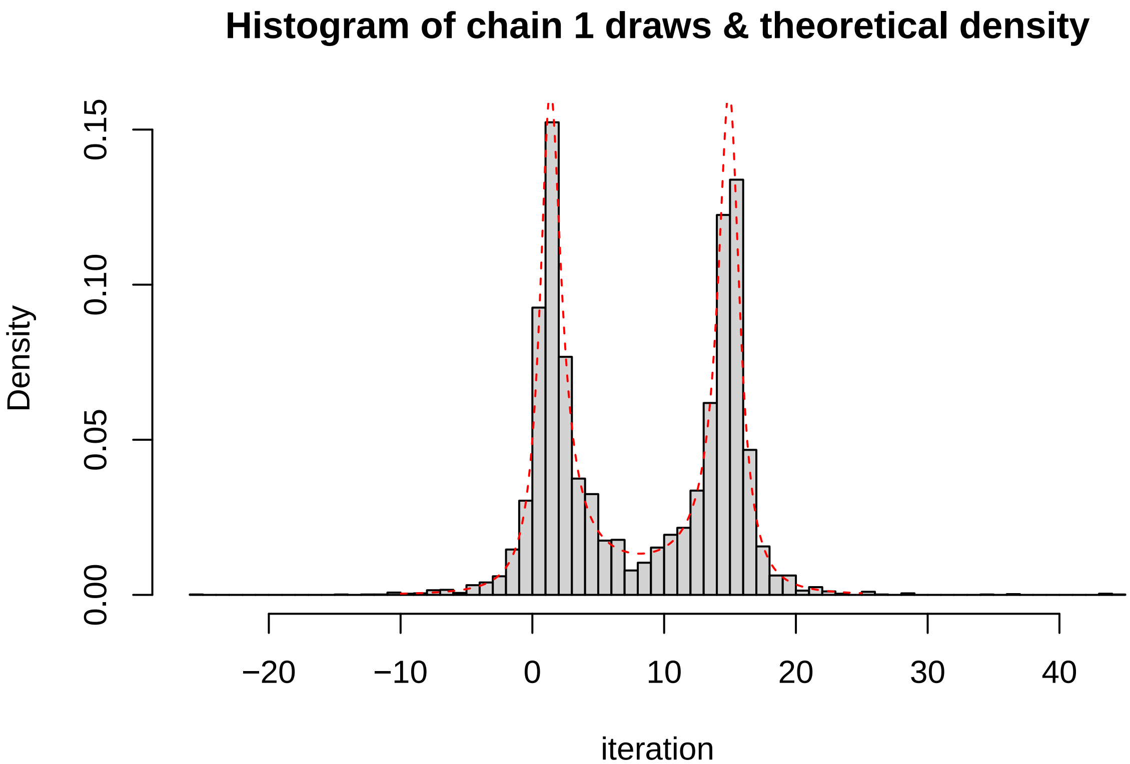


From the traceplot we see that starting from very different points, two chains stably converged. The posterior draws do not perfectly recover the theoretical density, there are fewer sample from the right mode.

```
problem1(15, 1e4, c(100, -100))
```




```
## [1] "Chain1 acceptance rate: 0.2057"
## [1] "Chain2 acceptance rate: 0.1955"
```



From the traceplot we see that starting from very different points, two chains stably converged. With larger scale the draws recovers the target distribution a little bit better.

Problem 2

Since there is no data provided for the eight schools example, I will use the data from table 11.2 (which leads to result of table 11.3) to conduct the simulation. The model remains the same.

(1)

The sampling distribution is

$$p(y|\theta, \sigma^2) = \prod_{j=1}^{J}\prod_{i=1}^{n_j} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_{ij}-\theta_j)^2}{2\sigma^2}\right).$$

The joint prior is

$$p(\theta|\mu, \tau^2) = \prod_{j=1}^{J} \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{(\theta_j-\mu)^2}{2\tau^2}\right).$$

The joint hyperprior is

$$p(\mu, \log\sigma, \log\tau) \propto \tau,$$

for $\sigma > 0$ and $\tau > 0$.

(2)

The unormalized joint posterior is

$$p(\theta, \mu, \log\sigma, \log\tau|y) \propto \tau \prod_{j=1}^{J} N(\theta_j|\mu, \tau^2) \prod_{j=1}^{J}\prod_{i=1}^{n_j} N(y_{ij}|\theta_j, \sigma^2).$$

(3)

In t-th iteration,

- Sample $(\sigma^2)^{(t)}$ from $p(\sigma^2|\theta^{(t-1)},\mu^{(t-1)},\tau^{(t-1)},y)\sim \text{Inv-}\chi^2(n,\hat\sigma^2)$ where $\hat\sigma^2 = n^{-1}\sum_{j=1}^J\sum_{i=1}^{n_j}(y_{ij}-\theta_j^{(t-1)})^2$.
- Sample $(\tau^2)^{(t)}$ from $p(\tau^2|\theta^{(t-1)},\mu^{(t-1)},\sigma^{(t)},y)\sim \text{Inv-}\chi^2(J-1,\hat\tau^2)$ where $\hat\tau^2 = (J-1)^{-1}\sum_{j=1}^J(\theta_j^{(t-1)}-\mu^{(t-1)})^2$.
- Sample $\theta^{(t)}\in\mathbb{R}^J$ from $p(\theta,\mu^{(t-1)},\sigma^{(t)},\tau^{(t)}|y)$, which is $N(\hat\theta,V_\theta)$.

For convenience let's temporarily denote $\mu^{(t-1)},\sigma^{(t)},\tau^{(t)}$as just μ,σ,τ in the formula of conditional posteriors. Then $\hat\theta\in\mathbb{R}^J$ is a vector and on jth coordinate we have

$$\hat\theta_j = \frac{\frac{1}{\tau^2}\mu + \frac{n_j}{\sigma^2}\bar y_{\cdot j}}{\frac{1}{\tau^2}+\frac{n_j}{\sigma^2}}.$$

The covariance matrix V_θ is a diagonal matrix with $1/(\frac{1}{\tau^2}+\frac{n_j}{\sigma^2})$ being jth diagonal element.

- Sample $\mu^{(t)}$ from $p(\mu|\theta^{(t)},\sigma^{(t)},\tau^{(t)},y)\sim N(\hat\mu,\tau^2/J)$ where $\hat\mu = (1/J)\sum_{j=1}^J\theta_j^{(t)}$.

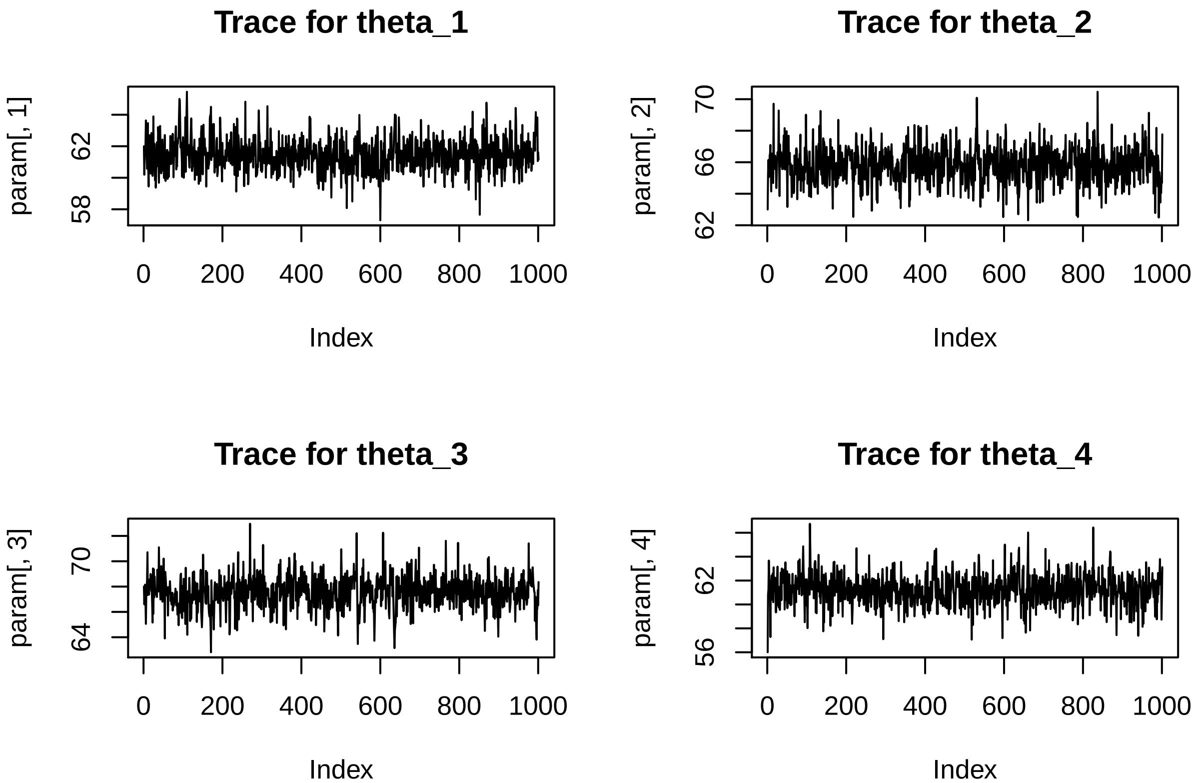
(4)

```r
library(extraDistr)
y_a <- c(62, 60, 63, 59)
y_b <- c(63, 67, 71, 64, 65, 66)
y_c <- c(68, 66, 71, 67, 68, 68)
y_d <- c(56, 62, 60, 61, 63, 64, 63, 59)

J <- 4
n <- c(4, 6, 6, 8)
theta <- c(mean(y_a), mean(y_b), mean(y_c), mean(y_d))

iter <- 1e3
param <- matrix(0, ncol=J+3, nrow=iter+1)
param[1, 1:J] <- c(62, 63, 68, 56)
param[1, J+1] <- mean(param[1, 1:J])
```

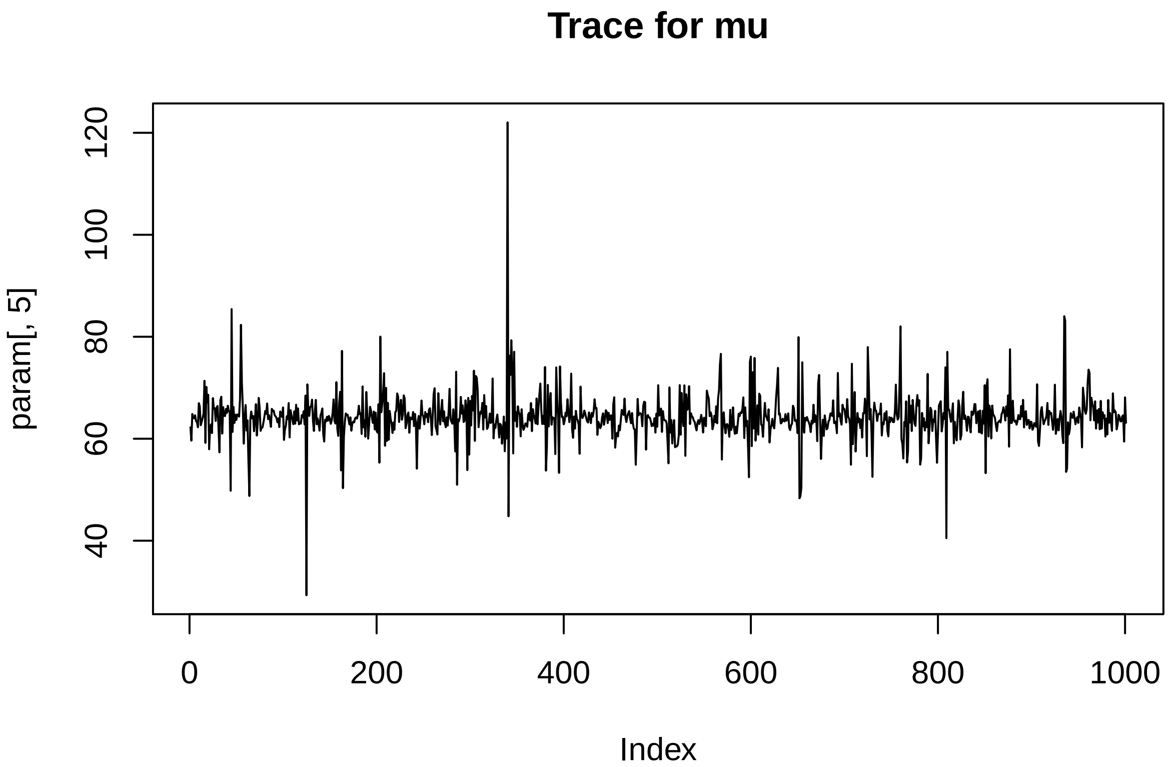
```r
for (i in 1:iter) {
  hat_sigsq <- 1/sum(n)*(sum((y_a-param[i, 1])^2) + sum((y_b-param[i, 2])^2) +     sum((y_c-param[i, 3]
  param[i+1, J+2] <- rinvchisq(1, n, hat_sigsq)

  hat_tausq <- (1/(J-1))*sum((param[i, 1:J] - param[i, J+1])^2)
  param[i+1, J+3] <- rinvchisq(1, J-1, hat_tausq)

  hat_theta <- (1/param[i+1, J+3]*param[i, J+1] + n/param[i+1, J+2]*theta) / (1/param[i+1, J+3] + n/para
  hat_Sig <- diag(1/param[i+1, J+3] + n/param[i+1, J+2])
  param[i+1, 1:J] <- MASS::mvrnorm(1, hat_theta, hat_Sig)

  hat_mu <- 1/J*sum(param[i+1, 1:J])
  param[i+1, J+1] <- rnorm(1, hat_mu, sqrt(param[i+1, J+3]/J))
}
```

```r
par(mfrow=c(2,2))
plot(param[, 1], type='l', main='Trace for theta_1')
```

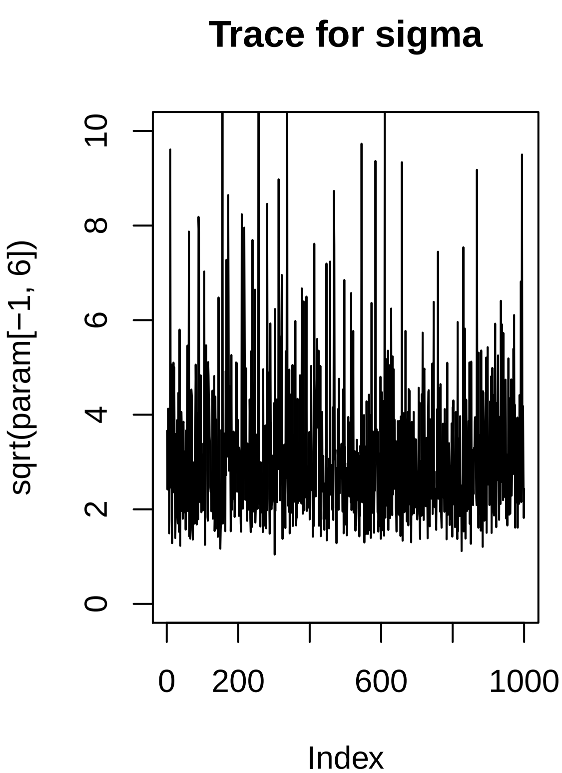
```r
plot(param[, 2], type='l', main='Trace for theta_2')
plot(param[, 3], type='l', main='Trace for theta_3')
plot(param[, 4], type='l', main='Trace for theta_4')
```

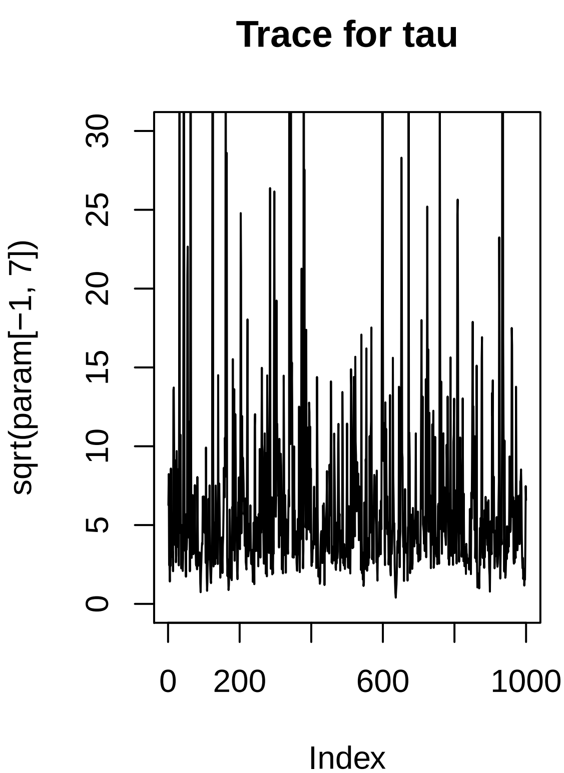



```r
plot(param[, 5], type='l', main='Trace for mu')
```




```
par(mfrow=c(1, 2))
plot(sqrt(param[-1, 6]), type='l', ylim=c(0, 10), main='Trace for sigma')
plot(sqrt(param[-1, 7]), type='l', ylim=c(0, 30), main='Trace for tau')
```






```
param <- tail(param, floor(.8*iter))
param[, 6] <- sqrt(param[, 6])
```

```r
param[, 7] <- sqrt(param[, 7])
res <- t(apply(param, 2, quantile, c(.025, .25, .5, .75, .975)))
rownames(res) <- c('theta_1', 'theta_2', 'theta_3', 'theta_4',
                   'mu', 'sigma', 'tau')
res
```

```
##              2.5%       25%       50%       75%     97.5%
## theta_1 59.496957 60.778436 61.418783 62.005591 63.610911
## theta_2 63.538200 65.104977 65.885347 66.576458 68.041913
## theta_3 65.090785 66.898253 67.622385 68.404140 70.008915
## theta_4 58.664744 60.406291 61.176348 61.959437 63.599529
## mu      55.304893 62.581646 63.950847 65.540477 74.066764
## sigma    1.430931  2.118366  2.731825  3.686346  7.197696
## tau      1.540739  3.227898  4.691281  7.546328 21.038981
```

Problem 3

The acceptance probability $r_B(\theta^*)$ is

$$r_B = \frac{p(\theta^*|y)J_t(\theta^{t-1}|\theta^*)}{p(\theta^*|y)J_t(\theta^{t-1}|\theta^*) + p(\theta^{t-1}|y)J_t(\theta^*|\theta^{t-1})}.$$

Denote the transition kernel by T_t. We have

$$\begin{aligned}
p(\theta^{t-1}|y)T_t(\theta^*|\theta^{t-1}) &= p(\theta^{t-1}|y)J_t(\theta^*|\theta^{t-1})r_B(\theta^*) \\
&= p(\theta^{t-1}|y)J_t(\theta^*|\theta^{t-1})\frac{p(\theta^*|y)J_t(\theta^{t-1}|\theta^*)}{p(\theta^*|y)J_t(\theta^{t-1}|\theta^*) + p(\theta^{t-1}|y)J_t(\theta^*|\theta^{t-1})} \\
&= p(\theta^*|y)J_t(\theta^{t-1}|\theta^*)\frac{p(\theta^{t-1}|y)J_t(\theta^*|\theta^{t-1})}{p(\theta^*|y)J_t(\theta^{t-1}|\theta^*) + p(\theta^{t-1}|y)J_t(\theta^*|\theta^{t-1})} \\
&= p(\theta^*|y)T_t(\theta^{t-1}|\theta^*).
\end{aligned}$$

Thus the detailed balance condition is satisfied and the stationary distribution is our target distribution $p(\theta|y)$.