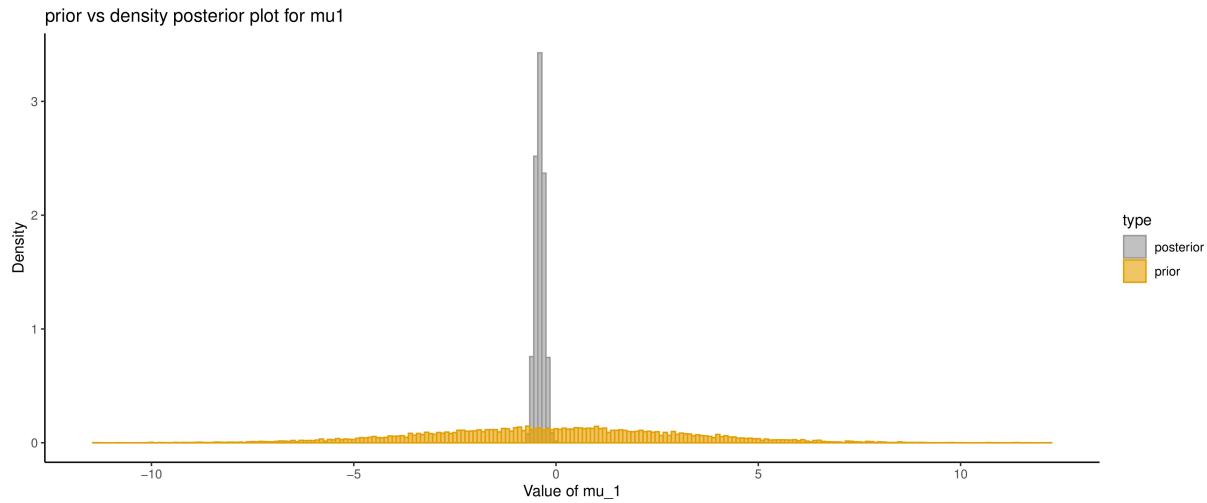


```

p1_1 <- ggplot(data1_1, aes(x=value, fill=type, color=type)) +
  geom_histogram(aes(y=after_stat(density)), alpha=0.6,
                 position = 'identity', binwidth = .1) +
  scale_color_manual(values=c("#999999", "#E69F00"))+
  scale_fill_manual(values=c("#999999", "#E69F00"))+
  labs(title="prior vs density posterior plot for mu1",x="Value of mu_1",
       y = "Density")+
  theme_classic()
p1_1

```



plot the posterior vs prior of  $\mu_2$ ,

```

# Build dataset with different distributions
data1_2 <- data.frame(
  type = c( rep("posterior", 10001), rep("prior", 10000) ),
  value = c(l1[,3], y)
)

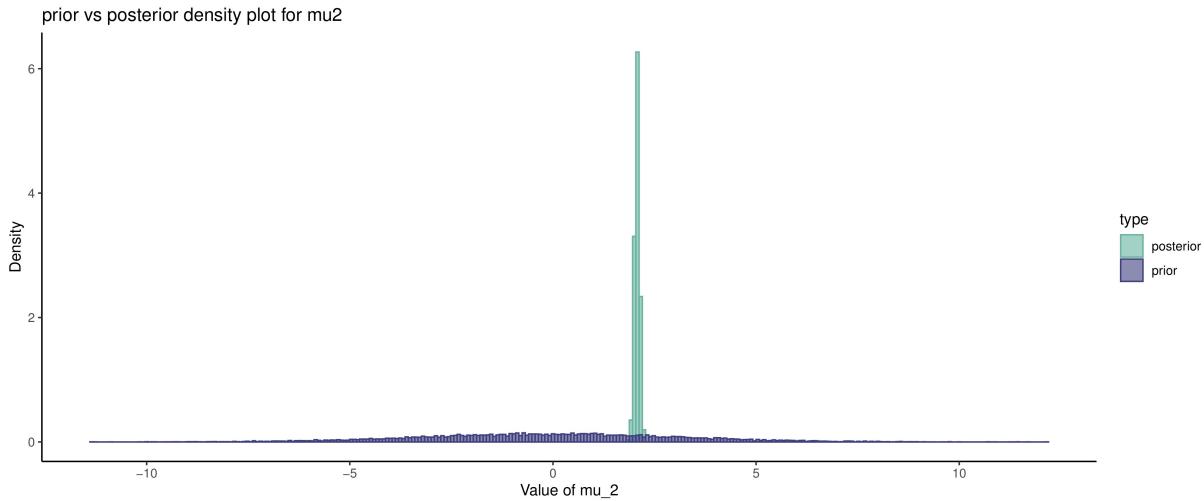
p1_2 <- ggplot(data1_2, aes(x=value, fill=type, color=type)) +
  geom_histogram(aes(y=..density..), alpha=0.6, position = 'identity',
                 binwidth = .08) +
  scale_color_manual(values=c("#69b3a2", "#404080"))+
  scale_fill_manual(values=c("#69b3a2", "#404080"))+
  labs(title="prior vs posterior density plot for mu2",x="Value of mu_2",
       y = "Density")+
  theme_classic()
p1_2

```

```

## Warning: The dot-dot notation ('..density..') was deprecated in ggplot2 3.4.0.
## i Please use 'after_stat(density)' instead.

```



From these plots, we could clearly see that all posteriors of these three parameters of interest have some differences with their corresponding prior distribution. This means that definitely there is some information in the data and through gibbs sampler we learn more information to update our beliefs.

The posterior distribution of three parameters of interest are very concentrated so we can think of that the change point location is around 11 or 12.

posterior mean and variance of each of the parameters of interest,

```
#posterior mean
mk=mean(l1[,1])
m1=mean(l1[,2])
m2=mean(l1[,3])

#posterior variance
vk=var(l1[,1])
v1=var(l1[,2])
v2=var(l1[,3])

mean_int = c(mk,m1,m2)
var_int = c(vk,v1,v2)
param=c('k','mu_1','mu_2')

param_1 <- data.frame('parameter'=param, "posterior mean"=mean_int,
                      "posterior variance" = var_int)
param_1

##   parameter posterior.mean posterior.variance
## 1          k     11.3003700    0.247368863
## 2        mu_1     -0.4006117    0.012155493
## 3        mu_2      2.0708639    0.003800127
```

(e)

Repeat (d) on the second data series, use same question index with (d), and comment the differences at last.

i.

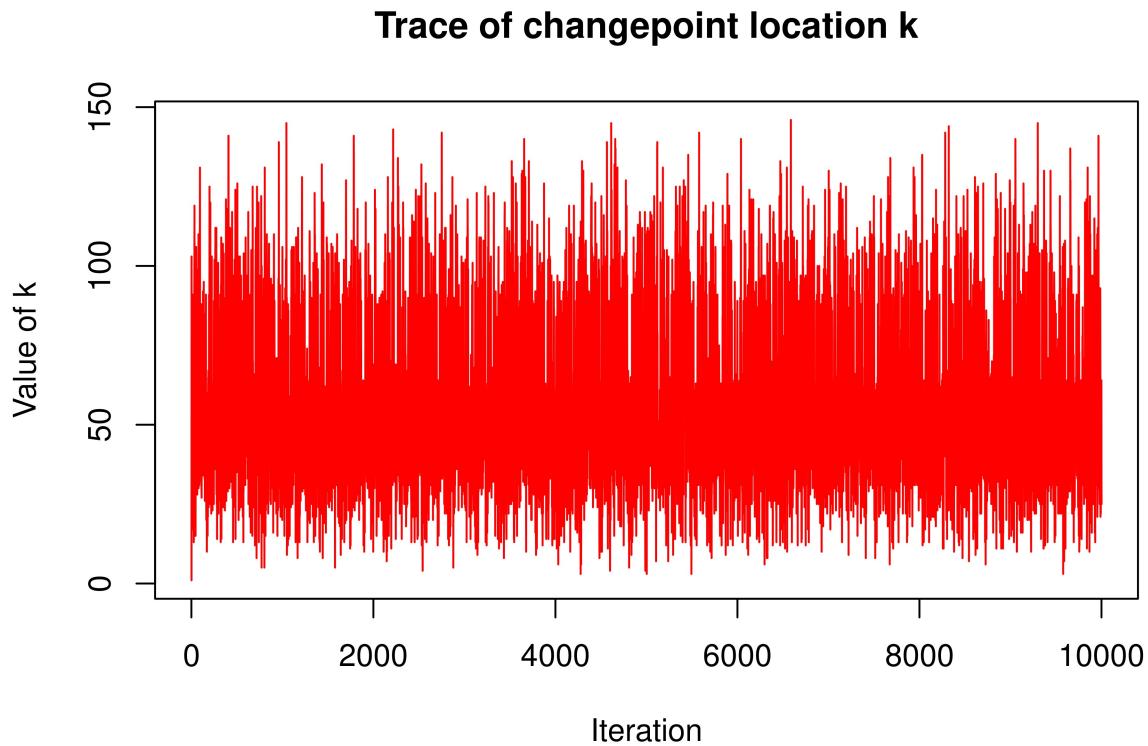
Run Gibbs sampler on the second data series with initial values  $X^{(0)} = (k^{(0)}, \mu_1^{(0)}, \mu_2^{(0)}) = (1, 1, 1)$  and generate 10000 samples,

```
l2=pmix.gibbs(1,1,1,dataset2,10000)
```

ii.

First, we look at the trace of changepoint location for  $k$ ,

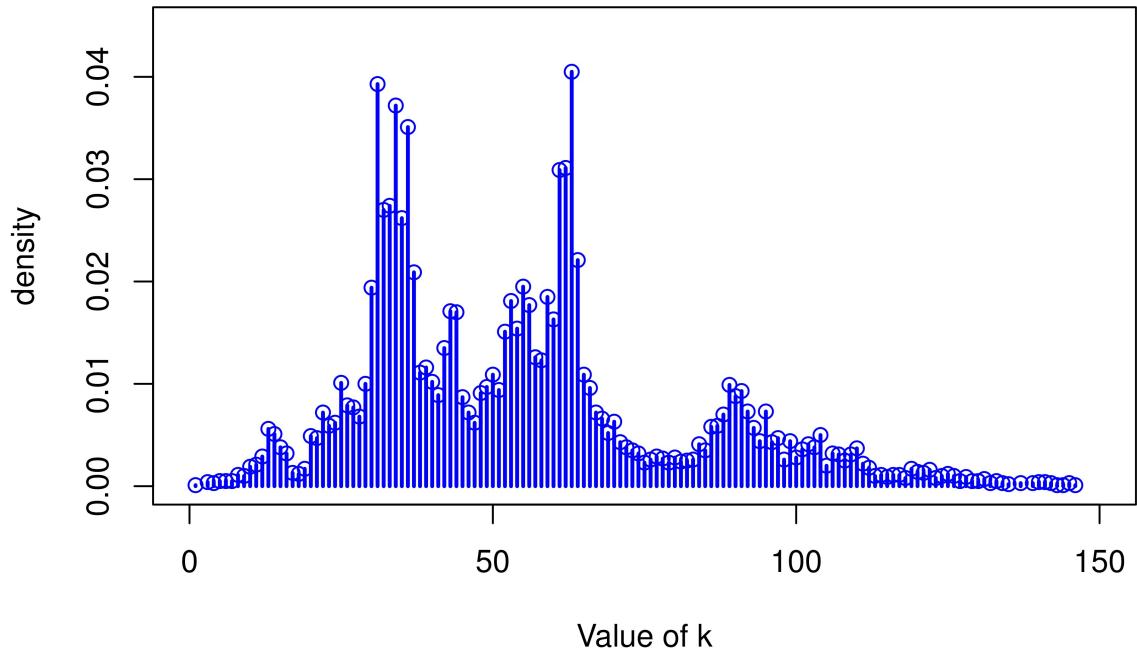
```
plot(l2[,1],type = 'l',col='red',main = 'Trace of changepoint location k',
     xlab = 'Iteration',ylab = 'Value of k')
```



Also, we can plot the estimated posterior distribution of  $k$ ,

```
dk2=as.data.frame(table(l2[,1]))
plot(as.vector(dk2$Var1),as.vector(dk2$Freq/sum(dk2$Freq)),type='h',
     xlim=c(0,150),ylim = c(0,0.045),lwd=2,col='blue',
     main = 'Estimated posterior distribution of k',xlab = 'Value of k',
     ylab = 'density')
points(as.vector(dk2$Var1),as.vector(dk2$Freq/sum(dk2$Freq)),col='blue',lwd=1)
```

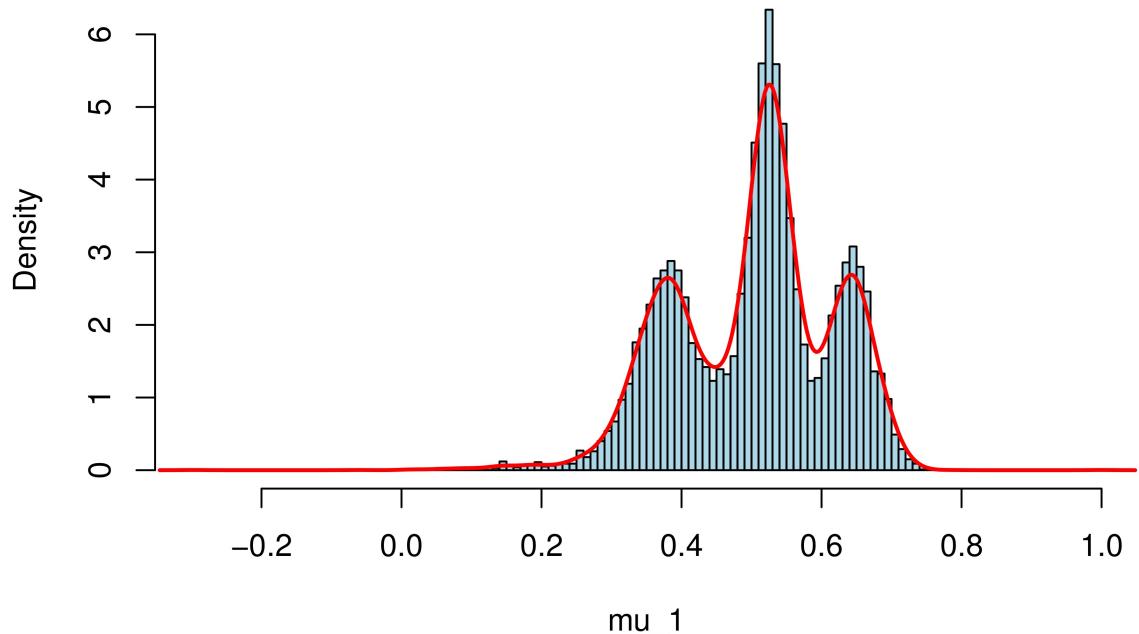
## Estimated posterior distribution of k



Next, let us visualize the density plot and trace plot of  $\mu_1$  and  $\mu_2$ ,

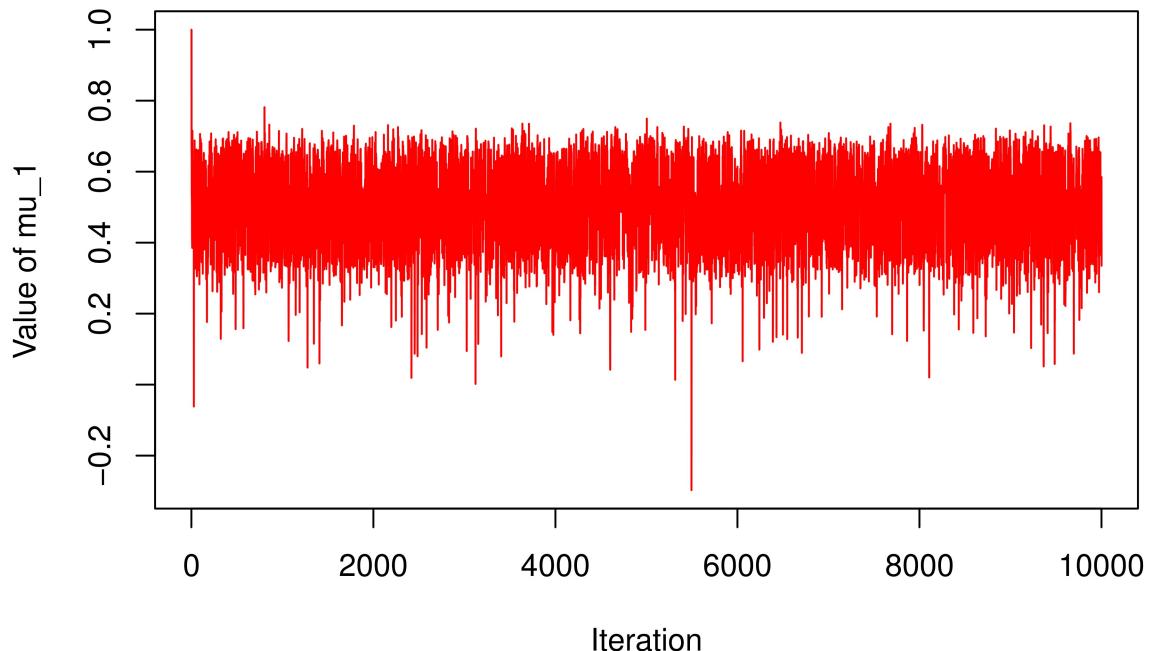
```
hist(12[,2],freq = FALSE, ,breaks=100, col = 'lightblue',
     main = 'Estimated posterior distribution of mu_1',xlab = 'mu_1')
lines(density(12[,2]),col='red',lwd=2)
```

## Estimated posterior distribution of mu\_1



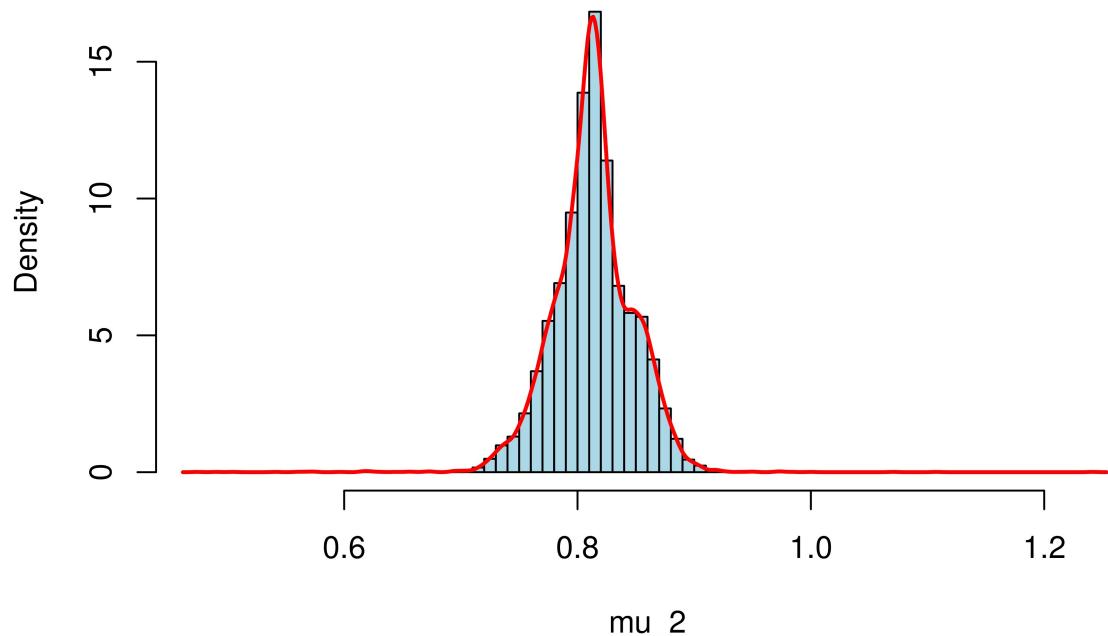
```
plot(l2[,2], col = 'red',main = 'Trace of mu_1',xlab = 'Iteration',
     ylab = 'Value of mu_1',type='l')
```

### Trace of mu\_1



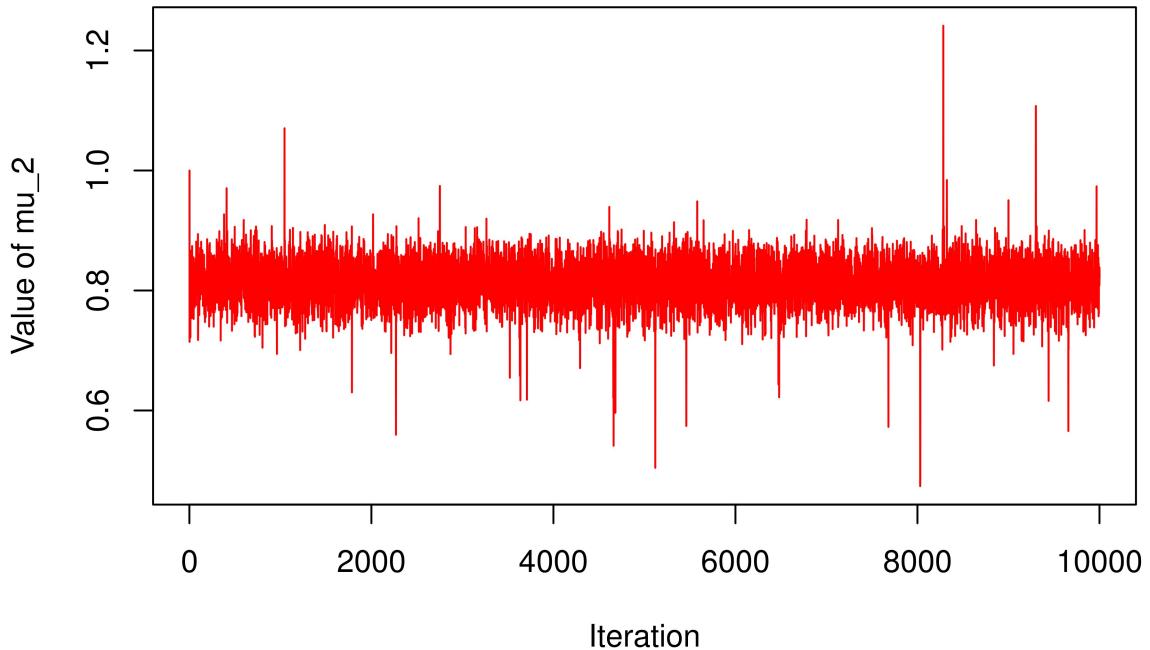
```
hist(l2[,3],freq = FALSE, ,breaks=100, col = 'lightblue',
     main = 'Estimated posterior distribution of mu_2',xlab = 'mu_2')
lines(density(l2[,3]),col='red',lwd=2)
```

## Estimated posterior distribution of mu\_2



```
plot(l2[,3], col = 'red',main = 'Trace of mu_2',xlab = 'Iteration',
     ylab = 'Value of mu_2',type='l')
```

## Trace of mu\_2



Then, we calculate the effective sample size of our output using the autocorrelation of the changepoint location,

```
#get the autocorrelation of changepoint location k
auto_cf_12_k = acf(l2[,1],plot = FALSE)
#get the first order sample autocorrelation of changepoint location k
acf2 = auto_cf_12_k$acf[2]
acf2
```

```
## [1] 0.4392691
```

so we can obtain that  $\rho(X_1^{(t-1)}, X_1^{(t)}) = 0.05868507$ , the by the formula of ESS

$$T_{ESS} = \frac{1 - \rho}{1 + \rho} \cdot T$$

```
T_ESS = round(10000*((1-acf2)/(1+acf2)))
print(paste("thus the effective sample size is", T_ESS))
```

```
## [1] "thus the effective sample size is 3896"
```

This ESS size is big enough to converge. Therefore, we don't need to run the sampler again for a longer period because the effective sample size is high. This means comparing to i.i.d. sampling, there are not much information loss in the sample from the Markov Chain.

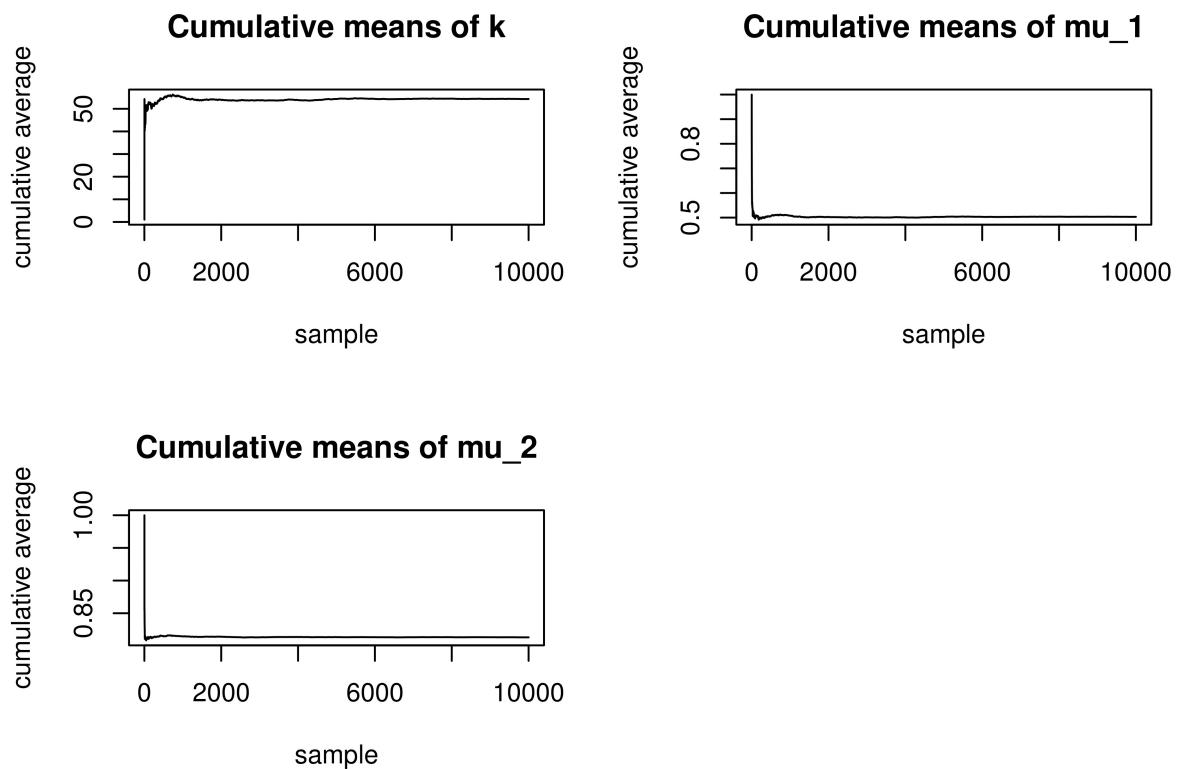
Alternately, we can assess the convergence by cumulative mean average, and the plots showed that the average converges very fast.

```

par(mfrow = c(2, 2))
cum_avg1 <- cummean(l2[,1])
cum_avg2 <- cummean(l2[,2])
cum_avg3 <- cummean(l2[,3])

plot(cum_avg1,type='l',xlab = 'sample',ylab = 'cumulative average',
      main = 'Cumulative means of k')
plot(cum_avg2,type='l',xlab = 'sample',ylab = 'cumulative average',
      main = 'Cumulative means of mu_1')
plot(cum_avg3,type='l',xlab = 'sample',ylab = 'cumulative average',
      main = 'Cumulative means of mu_2')

```



iii.

Compare the prior and estimated posterior distributions of the three parameters of interest ( $k, \mu_1, \mu_2$ ).

First, plot the posterior vs prior of  $k$ , the prior of  $k$  is also discrete but we plot it as a line with points in order to visualize more clearly,

```

plot(as.vector(dk2$Var1),as.vector(dk2$Freq/sum(dk2$Freq)),type='h',
     xlim=c(0,150),ylim = c(0,0.045),lwd=2,col='blue',
     main = 'Posterior vs prior of k',xlab = 'Value of k',ylab = 'density')
points(as.vector(dk2$Var1),as.vector(dk2$Freq/sum(dk2$Freq)),col='blue',lwd=1)
lines(prior.k(dataset2),col='green')
points(prior.k(dataset2),col='green')

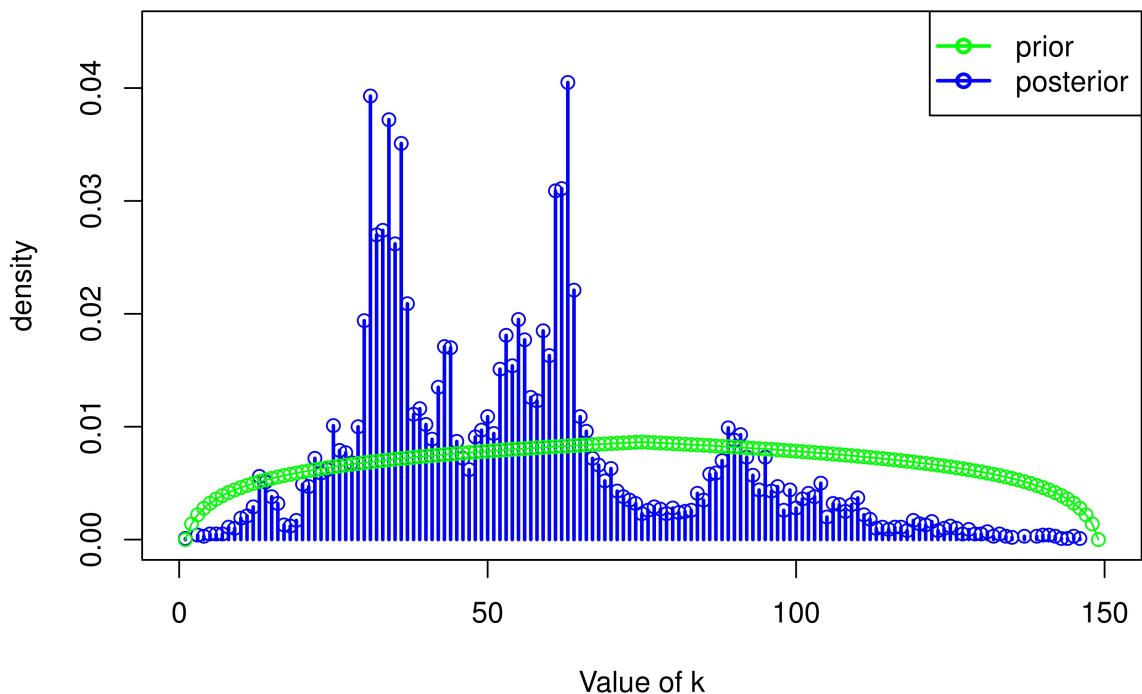
```

```

legend(x = "topright",
       legend = c("prior", "posterior"),
       pch = c(1,1),
       col = c("green", "blue"),
       lwd = 2)

```

**Posterior vs prior of k**



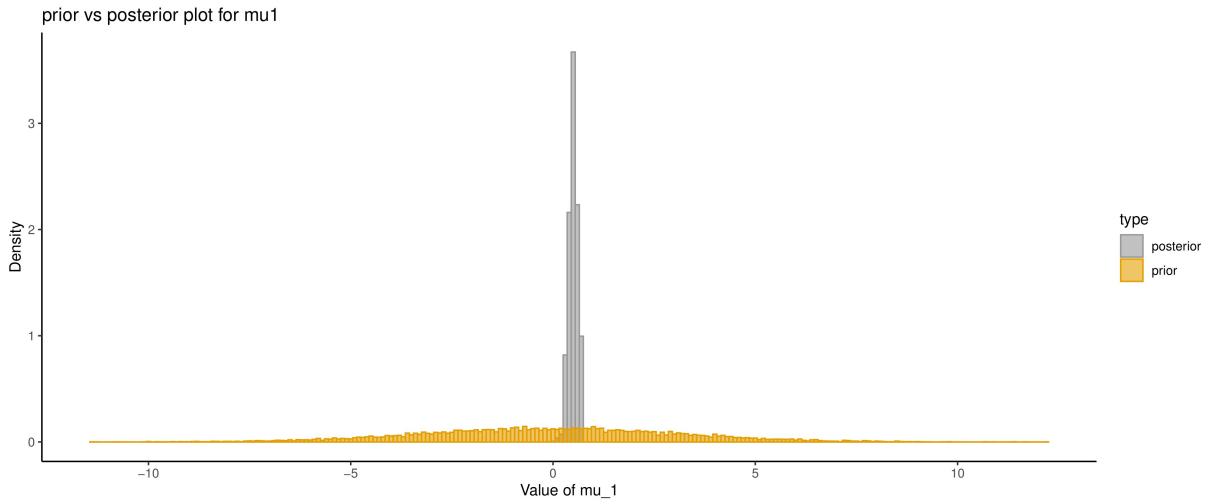
plot the posterior vs prior of  $\mu_1$ ,

```

# Build dataset with different distributions
data2_1 <- data.frame(
  type = c(rep("posterior", 10001), rep("prior", 10000) ),
  value = c(12[,2], y)
)

p2_1 <- ggplot(data2_1, aes(x=value, fill=type, color=type)) +
  geom_histogram(aes(y=..density..), alpha=0.6, position = 'identity',
                 binwidth = .1) +
  scale_color_manual(values=c("#999999", "#E69F00"))+
  scale_fill_manual(values=c("#999999", "#E69F00"))+
  labs(title="prior vs posterior plot for mu1",x="Value of mu_1",
       y = "Density")+
  theme_classic()
p2_1

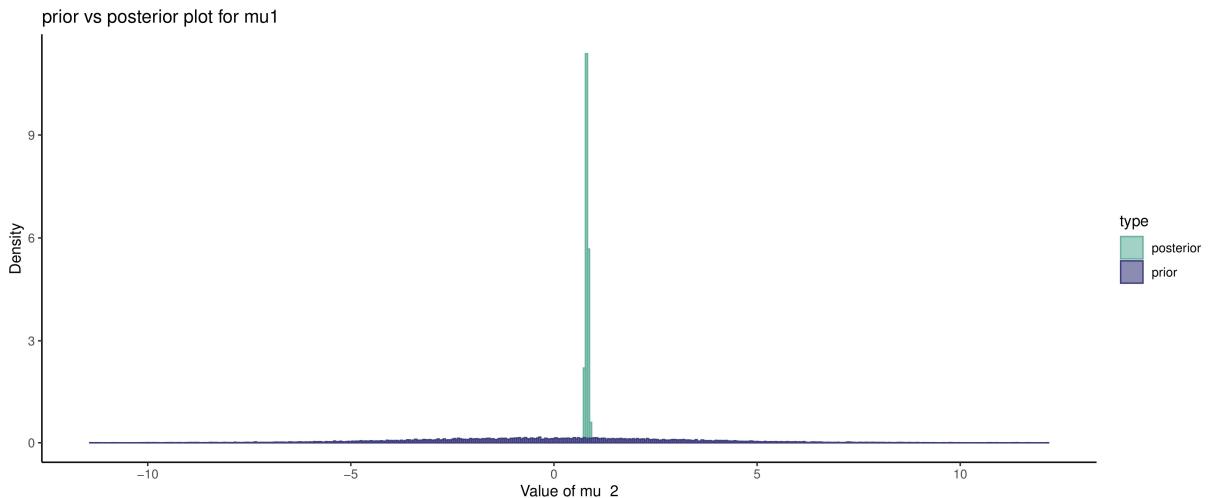
```



plot the posterior vs prior of  $\mu_2$ ,

```
# Build dataset with different distributions
data2_2 <- data.frame(
  type = c( rep("posterior", 10001), rep("prior", 10000) ),
  value = c(l2[,3], y)
)

p2_2 <- ggplot(data2_2, aes(x=value, fill=type, color=type)) +
  geom_histogram(aes(y=..density..), alpha=0.6, position = 'identity',
                 binwidth = .05) +
  scale_color_manual(values=c("#69b3a2", "#404080"))+
  scale_fill_manual(values=c("#69b3a2", "#404080"))+
  labs(title="prior vs posterior plot for mu1",x="Value of mu_2",
       y = "Density")+
  theme_classic()
p2_2
```



Similarly from these plots, we could clearly see that all posteriors of these three parameters of interest have some differences with their corresponding prior distribution. This means that definitely there is some

information in the data and through gibbs sampler we learn more information to update our beliefs.

But this time the posterior distribution of  $k$  is not concentrated, it emerges with multiple peaks but finally converge to around 53. So we cannot really determine the change point location till now.

posterior mean and variance of each of the parameters of interest,

```
#posterior mean
mk=mean(l2[,1])
m1=mean(l2[,2])
m2=mean(l2[,3])

#posterior variance
vk=var(l2[,1])
v1=var(l2[,2])
v2=var(l2[,3])

mean_int = c(mk,m1,m2)
var_int = c(vk,v1,v2)
param=c('k','mu_1','mu_2')

param_2 <- data.frame('parameter'=param, "posterior mean"=mean_int,
                      "posterior variance" = var_int)
param_2

##   parameter posterior.mean posterior.variance
## 1          k     54.3229677    6.240693e+02
## 2      mu_1     0.5031027   1.261611e-02
## 3      mu_2     0.8131240   1.204371e-03
```

Comments:

Compare to data series 1, the posterior mean and variance of  $\mu_1$  and  $\mu_2$  are both very small. More precisely, the posterior mean of  $\mu_1$  of data series 2 is slightly bigger than data series 1. But the posterior mean of  $\mu_2$  of data series 2 is smaller than data series 1. Their variance of data series 1 and data series 2 are very similar both for  $\mu_1$  and  $\mu_2$ .

But when we look at the change point location  $k$ , the variance of data series 2 is very huge but data series 1's is very small. So the mean of data series 2 is not as convinced as data series 1's mean. Even we can think about in data series 2 there maybe not one change point in it. So the result maybe not right.

(f)

i.

Assume that we consider a set of 5 models  $\{\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3, \mathcal{M}_4, \mathcal{M}_5\}$ .  $\mathcal{M}_m$  means a  $m$  changepoints model.  $m \in \{1, 2, 3, 4, 5\}$  is the index of model and the number of changepoints in this model.

Each model is characterised by a density  $f_m$  and the associated parameter space  $\Theta_m$ , i.e.  $\mathcal{M}_m := \{f_m(\cdot|\theta), \theta \in \Theta_m\}$ , where  $f_m$  is the density and  $\Theta_m$  the parameter space of the  $m$ -changepoints model.

Then we define the parameter space  $\Theta_m$  more precisely,

$$\begin{aligned} \Theta_m := \{&\theta = (\mu_1, \dots, \mu_{m+1}, k_1, \dots, k_m) : \\ &k_i \in \{1, 2, \dots, n-1, n\} \text{ for } i \in \{1, \dots, m\} \text{ and } k_1 < k_2 < \dots < k_m, \\ &\mu_j \in \mathbb{R} \text{ for } j \in \{1, \dots, m+1\}\} \end{aligned}$$

where  $n$  is the length of data series.

Using a hierarchical Bayesian setup, we first place a prior distribution of the set of models,

$$P(\mathcal{M}_m) = p_m = \frac{1}{5} \quad \text{with} \quad \sum_{m=1}^5 P(\mathcal{M}_m) = 1$$

On each parameter space  $\Theta_m$ , we place prior distribution separately for  $\mu_1, \dots, \mu_{m+1}$  and  $k_1, \dots, k_m$ ,

For simplicity, we write  $\theta = (\hat{\mu}, \hat{k})$

$$\hat{\mu} = (\mu_1, \dots, \mu_{i+1}) \quad \text{and} \quad \hat{k}_m = (k_1, \dots, k_i)$$

where  $i$  is conditional on the model.

then we place the prior distribution on  $\hat{\mu}_m$  and  $\hat{k}_m$  conditional upon their being  $m$  changepoints,

the locations of the changepoints is uniform over all possible configurations of  $m$  changepoints (conditional upon their being  $m$  changepoints), so

$$P(\hat{k} | \mathcal{M}_m) = f_m^{prior}(\hat{k}) = \frac{1}{C_m^{n-1}}$$

alternatively,

$$\hat{k} | \mathcal{M}_m \sim Unif(C_m^{n-1})$$

for the mean parameter  $\hat{\mu}$ , we use the natural generalisation of the single changepoint model above, then,

$$f_m^{prior}(\hat{\mu}) = \prod_{i=1}^{m+1} \phi_{0,10}(\mu_i)$$

Assume now that we have observed data  $\mathbf{y} = (y_1, \dots, y_n)$ . We define the likelihood function conditional on the model  $\mathcal{M}_m$  and parameter  $\theta = (\hat{\mu}, \hat{k})$  be,

$$l(\mathbf{y} | \theta, \mathcal{M}_m) = l_m(\mathbf{y} | \theta) = \prod_{i=1}^{k_1} \phi_{\mu_1,1}(y_i) \prod_{j=k_1+1}^{k_2} \phi_{\mu_2}(\mathbf{y}) \dots \prod_{q=k_m+1}^n \phi_{\mu_{m+1},1}(\mathbf{y})$$

**ii.**

The full posterior distribution over both model and parameters is, (for simplicity we write  $f^{post}(\mathcal{M}_m, \theta)$  as  $f^{post}(m, \theta)$ )

$$\begin{aligned} f^{post}(m, \theta) &= \frac{p_m f_m^{prior}(\hat{k}) f_m^{prior}(\hat{\mu}) l_m(\mathbf{y} | \hat{k}, \hat{\mu})}{\sum_m^{n-1} p_m \int_{\Theta_m} f_m^{prior}(\hat{k}) f_m^{prior}(\hat{\mu}) l_m(\mathbf{y} | \hat{k}, \hat{\mu}) d\hat{k} d\hat{\mu}} \\ &= \frac{f_m^{prior}(\hat{k}) f_m^{prior}(\hat{\mu}) l_m(\mathbf{y} | \hat{k}, \hat{\mu})}{\sum_m^{n-1} \int_{\Theta_m} f_m^{prior}(\hat{k}) f_m^{prior}(\hat{\mu}) l_m(\mathbf{y} | \hat{k}, \hat{\mu}) d\hat{k} d\hat{\mu}} \\ &= \frac{\frac{1}{C_5^m} \prod_{i=1}^{m+1} \phi_{0,10}(\mu_i) \prod_{i=1}^{k_1} \phi_{\mu_1,1}(y_i) \prod_{j=k_1+1}^{k_2} \phi_{\mu_2}(\mathbf{y}) \dots \prod_{q=k_m+1}^n \phi_{\mu_{m+1},1}(\mathbf{y})}{\sum_m^{n-1} \int_{\Theta_m} \frac{1}{C_5^m} \prod_{i=1}^{m+1} \phi_{0,10}(\mu_i) \prod_{i=1}^{k_1} \phi_{\mu_1,1}(y_i) \prod_{j=k_1+1}^{k_2} \phi_{\mu_2}(\mathbf{y}) \dots \prod_{q=k_m+1}^n \phi_{\mu_{m+1},1}(\mathbf{y}) d\hat{k} d\hat{\mu}} \end{aligned}$$

defined on the disjoint union space

$$\Theta := \bigcup_{m=1}^5 (\{m\} \times \Theta_m)$$

### iii.

Consider the  $1 \rightarrow 2$  move first, we draw the mean  $\mu_3$  and change point location  $k_2$  of the new component from the corresponding prior. And we need to adjust the prior of the change point location a bit for  $k_2$ . So we draw  $k_2$  uniformly from  $\{1, \dots, k_1 - 1, k_1 + 1, \dots, n - 1\}$  which not included  $k_1$ . Therefore, we draw,

$$u_1^{(t-1)} \sim g_1 \quad u_2^{(t-1)} \sim g_2$$

and set

$$\mu_3 = u_1^{(t-1)} \quad k_2 = u_2^{(t-1)}$$

with  $g_1$  being the density of the prior distribution on the  $\mu_2$ ,  $g_2$  being the density of uniform distribution on the discrete set  $\{1, \dots, k_1 - 1, k_1 + 1, \dots, n - 1\}$ .

The corresponding transformations  $T_{1 \rightarrow 2}$  is,

$$T_{1 \rightarrow 2} \begin{pmatrix} \mu_1^{(t-1)} \\ \mu_2^{(t-1)} \\ k_1^{(t-1)} \\ u_1^{(t-1)} \\ u_2^{(t-1)} \end{pmatrix} = \begin{pmatrix} \mu_1^{(t-1)} \\ \mu_2^{(t-1)} \\ u_1^{(t-1)} \\ k_1^{(t-1)} \\ u_2^{(t-1)} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ k_1 \\ k_2 \end{pmatrix}$$

The determinant of the Jacobian of  $T_{1 \rightarrow 2}$  is 1.

In order to include fixed-dimensional moves, we use the setup corresponds to,

$$\rho_{1 \rightarrow 1} = \frac{1}{2}, \rho_{1 \rightarrow 2} = \frac{1}{2}, \rho_{2 \rightarrow 1} = \frac{1}{2}, \rho_{2 \rightarrow 2} = \frac{1}{2}$$

– If the current model is single changepoint model  $\mathcal{M}_1$  (i.e.  $m^{\wedge}\{t-1\}=1$ )

- With probability  $\frac{1}{2}$  perform an update of  $\theta^{(t-1)} = (\mu_1^{(t-1)}, \mu_2^{(t-1)}, k_1^{(t-1)})$  within model  $\mathcal{M}_1$ , i.e.

1. Generate  $\theta$  by its components' full conditional distribution

$$\begin{aligned} \mu_1^{(t)} &\sim f_1^{post}(\mu_1 | \mu_2^{(t-1)}, k_1^{(t-1)}) \\ \mu_2^{(t)} &\sim f_1^{post}(\mu_2 | \mu_1^{(t)}, k_1^{(t-1)}) \\ k_1^{(t)} &\sim f_1^{post}(k | \mu_1^{(t)}, \mu_2^{(t)}) \end{aligned}$$

2. Set  $\theta^{(t)} = \theta = (\mu_1^{(t)}, \mu_2^{(t)}, k_1^{(t)})$  .(Gibbs sampler is a special case of Metropolis-Hastings with acceptance probability 1.)

- Otherwise(with probability  $\frac{1}{2}$ ), attempt a jump to model  $\mathcal{M}_2$ ,i.e.

1. Generate  $u_1^{(t-1)} \sim g_1$  and  $u_2^{(t-1)} \sim g_2$  ( $g_1$  and  $g_2$  as we introduced above).

2. Set

$$\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ k_1 \\ k_2 \end{pmatrix} = T_{1 \rightarrow 2} \begin{pmatrix} \mu_1^{(t-1)} \\ \mu_2^{(t-1)} \\ k_1^{(t-1)} \\ u_1^{(t-1)} \\ u_2^{(t-1)} \end{pmatrix} = \begin{pmatrix} \mu_1^{(t-1)} \\ \mu_2^{(t-1)} \\ u_1^{(t-1)} \\ k_1^{(t-1)} \\ u_2^{(t-1)} \end{pmatrix}$$

3. Compute,( $\frac{1}{2} \times \frac{1}{3}$  in the probability of picking one  $\mu$  and one  $k$  from 3  $\mu$  and 2  $k$  in the  $2 \rightarrow 1$  step)

$$\begin{aligned} \alpha &= \min\left\{1, \frac{p_2 f_2^{prior}(k_1, k_2) f_2^{prior}(\mu_1, \mu_2, \mu_3) l_2(y_1, \dots, y_n | \mu_1, \mu_2, \mu_3, k_1, k_2)}{p_1 f_1^{prior}(k_1^{(t-1)}) f_1^{prior}(\mu_1^{(t-1)}, \mu_2^{(t-1)}) l_1(y_1, \dots, y_n | \mu_1^{(t-1)}, \mu_2^{(t-1)}, k_1^{(t-1)})} \cdot \frac{\rho_{2 \rightarrow 1}}{\rho_{1 \rightarrow 2}} \cdot \frac{\frac{1}{2} \times \frac{1}{3}}{g_1(u_1^{(t-1)}) \cdot g_2(u_2^{(t-1)})} \cdot 1\right\} \\ &= \min\left\{1, \frac{(n-1) \prod_{i=1}^3 \phi_{0,10}(\mu_i) l_2(y_1, \dots, y_n | \mu_1, \mu_2, \mu_3, k_1, k_2)}{(\mathbf{C}_2^{n-1}) \prod_{i=1}^2 \phi_{0,10}(\mu_i^{(t-1)}) l_1(y_1, \dots, y_n | \mu_1^{(t-1)}, \mu_2^{(t-1)}, k_1^{(t-1)})} \cdot \frac{1}{6 \cdot g_1(u_1^{(t-1)}) \cdot g_2(u_2^{(t-1)})}\right\} \end{aligned}$$

4. With probability  $\alpha$ , we set  $m^{(t)} = 2$  and  $\theta^{(t)} = (\mu_1^{(t)}, \mu_2^{(t)}, \mu_3^{(t)}, k_1^{(t)}, k_2^{(t)})$ . Otherwise, remain  $m^t = 1$  and  $\theta^{(t)} = \theta^{(t-1)}$ .

– If the current model is two changepoint model  $\mathcal{M}_2$  (i.e.  $m^{(t-1)} = 2$ )

- With probability  $\frac{1}{2}$  perform an update of  $\theta^{(t-1)} = (\mu_1^{(t-1)}, \mu_2^{(t-1)}, \mu_3^{(t-1)}, k_1^{(t-1)}, k_2^{(t-1)})$  within model  $\mathcal{M}_2$ , i.e.

1. Generate  $\theta$  by its components' full conditional distribution

$$\begin{aligned} \mu_1^{(t)} &\sim f_2^{post}(\mu_1 | \mu_2^{(t-1)}, \mu_3^{(t-1)}, k_1^{(t-1)}, k_2^{(t-1)}) \\ \mu_2^{(t)} &\sim f_2^{post}(\mu_2 | \mu_1^{(t)}, \mu_3^{(t-1)}, k_1^{(t-1)}, k_2^{(t-1)}) \\ \mu_3^{(t)} &\sim f_2^{post}(\mu_3 | \mu_1^{(t)}, \mu_2^{(t)}, k_1^{(t-1)}, k_2^{(t-1)}) \\ k_1^{(t)} &\sim f_2^{post}(k | \mu_1^{(t)}, \mu_2^{(t)}, \mu_3^{(t)}, k_2^{(t-1)}) \\ k_2^{(t)} &\sim f_2^{post}(k | \mu_1^{(t)}, \mu_2^{(t)}, \mu_3^{(t)}, k_1^{(t)}) \end{aligned}$$

2. Set  $\theta^{(t)} = \theta = (\mu_1^{(t)}, \mu_2^{(t)}, \mu_3^{(t)}, k_1^{(t)}, k_2^{(t)})$  .(Gibbs sampler is a special case of Metropolis-Hastings with acceptance probability 1.)

- Otherwise(with probability  $\frac{1}{2}$ ), attempt a jump to model  $\mathcal{M}_2$ ,i.e.

1. Drop one  $\mu$  randomly with probability  $\frac{1}{3}$  from  $\{\mu_1^{(t-1)}, \mu_2^{(t-1)}, \mu_3^{(t-1)}\}$ .

2. Drop one  $k$  randomly with probability  $\frac{1}{2}$  from  $\{k_1^{(t-1)}, k_2^{(t-1)}\}$ .

3. Compute the reciprocal

$$\frac{1}{\alpha}$$

4. With probability  $\frac{1}{\alpha}$ , we set  $m^{(t)} = 1$  and  $\theta^{(t)} = (\mu_1^{(t)}, \mu_2^{(t)}, k_1^{(t)})$ . Otherwise, remain  $m^t = 2$  and  $\theta^{(t)} = \theta^{(t-1)}$ .

#### iv.

We remain  $\rho_{2 \rightarrow 1}$  and  $\rho_{1 \rightarrow 2}$  in the acceptance probability,

$$\begin{aligned} \alpha &= \min\left\{1, \frac{p_2 f_2^{prior}(k_1, k_2) f_2^{prior}(\mu_1, \mu_2, \mu_3) l_2(y_1, \dots, y_n | \mu_1, \mu_2, \mu_3, k_1, k_2)}{p_1 f_1^{prior}(k_1^{(t-1)}) f_1^{prior}(\mu_1^{(t-1)}, \mu_2^{(t-1)}) l_1(y_1, \dots, y_n | \mu_1^{(t-1)}, \mu_2^{(t-1)}, k_1^{(t-1)})} \cdot \frac{\rho_{2 \rightarrow 1}}{\rho_{1 \rightarrow 2}} \cdot \frac{\frac{1}{2} \times \frac{1}{3}}{g_1(u_1^{(t-1)}) \cdot g_2(u_2^{(t-1)})} \cdot 1\right\} \\ &= \min\left\{1, \frac{(n-1) \prod_{i=1}^3 \phi_{0,10}(\mu_i) l_2(y_1, \dots, y_n | \mu_1, \mu_2, \mu_3, k_1, k_2)}{(\mathbf{C}_2^{n-1}) \prod_{i=1}^2 \phi_{0,10}(\mu_i^{(t-1)}) l_1(y_1, \dots, y_n | \mu_1^{(t-1)}, \mu_2^{(t-1)}, k_1^{(t-1)})} \cdot \frac{\rho_{2 \rightarrow 1}}{\rho_{1 \rightarrow 2}} \cdot \frac{1}{6 \cdot g_1(u_1^{(t-1)}) \cdot g_2(u_2^{(t-1)})}\right\} \end{aligned}$$