Assignment2

Yi Ren

December 2020

1 Question 2.1

1. No
2. Yes
3. $\{\mu_{r-1,c}, \mu_{r,c-1}, \mu_{r,c}, \mu_{r,c+1}, \mu_{r+1,c+1}\}$
4. No
5. No
6. $\{Z_m^n : n \in [N], m \in [M]\} \cup \{C^n : n \in [N]\}$

2 Question 2.2

2.1 2.2.7

The code can be found on `https://github.com/Yi-Ren1999/DD2434-A2`

2.2 2.2.8

Size	1	2	3	4	5
Small	0.016	0.015	0.011	0.008	0.04
Medium	$4.33e^{-18}$	$3.09e^{-20}$	$1.05e^{-16}$	$6.58e^{-16}$	$1.48e^{-18}$
Large	$3.28e^{-69}$	$1.1e^{-66}$	$2.52e^{-68}$	$1.24e^{-66}$	$3.53e^{-69}$

Table 1: Caption

3 Question 2.3

3.1 2.3.9

The code can be found on `https://github.com/Yi-Ren1999/DD2434-A2`

3.2 2.3.10

We have the following equation known:

$$
\begin{aligned}
P(D \mid \mu, r) &= (\frac{\gamma}{2\pi})^{N/2} exp\{-\frac{\gamma}{2}\sum_{n=1} N(x_n-\mu)^2\} \\
P(\mu \mid \gamma) &= N(\mu \mid \mu_0, (\lambda_0\gamma)^{-1}) \\
P(\gamma) &= Gam(\gamma \mid a, b)
\end{aligned}
$$

while the equation for exact posterior is :

$$
\begin{aligned}
&p(\mu, \gamma \mid D) \\
=&p(u, r, D)(D) \\
=&p(D \mid \mu, \gamma)p(\mu \mid \gamma)p(\gamma)/p(D)
\end{aligned}
$$

where P(D) is a constant and all the other terms can be calculated by the above equation

3.3 2.3.11

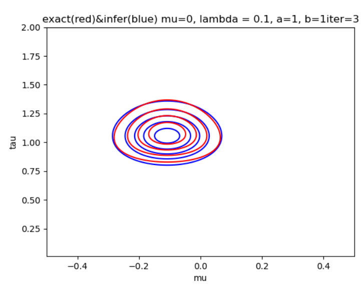


Figure 1: $\mu = 0\lambda = 0.1a = 1b = 1$

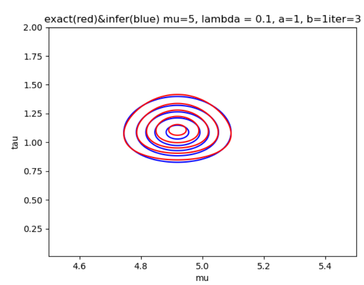


Figure 2: $\mu = 5\lambda = 0.1a = 1b = 1$

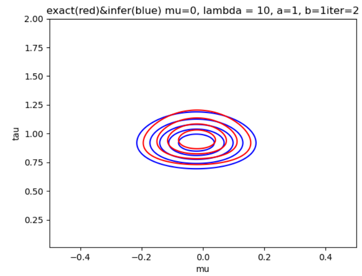


Figure 3: $\mu = 0 \lambda = 10 a = 1 b = 1$

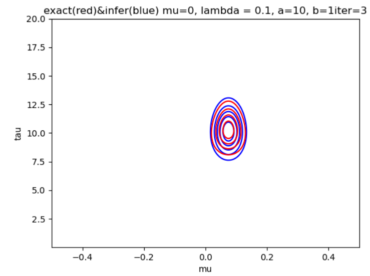


Figure 4: $\mu = 0 \lambda = 0.1 a = 10 b = 1$

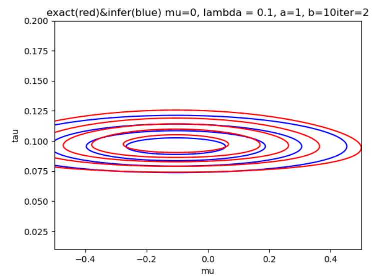


Figure 5: $\mu = 0 \lambda = 0.1 a = 1 b = 10$

From the above 5 examples we can see that the VI algorithm performs well in inferring exact posterior generally, but at the same time different coefficients of the input data can influence the accuracy. The selection of μ doesn't change the accuracy.When the λ gets bigger, the accuracy drop. When the ratio of a/b is too large or small, the accuracy of the algorithm decreases.

4 Question 2.4

5 2.4.12

The code can be found on `https://github.com/Yi-Ren1999/DD2434-A2`

5.1 2.4.13

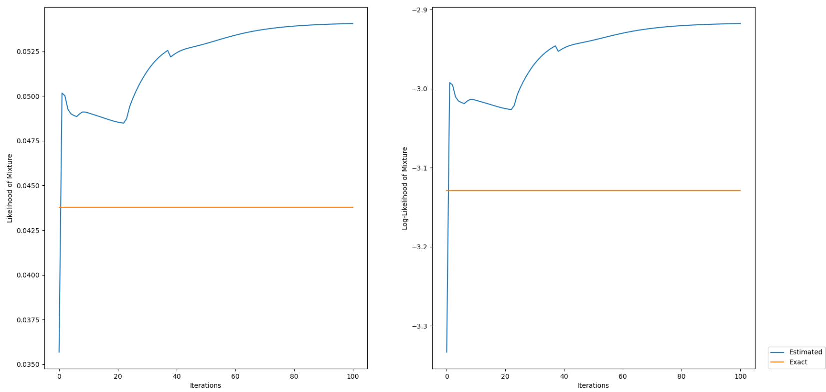


Figure 6: sample=100,n=5,c=3

As shown in the figure, the likelihood of real mixture tree is 0.45 which the Mixture tree created by the EM algorithm rise to about 0.54 at the end of iteration, which shows the inferred mixture tree fit the sample data better than the real tree.
The RF comparison is shown below:

```
4.2 Compare trees and print Robinson-Foulds (RF) distance:

        t0 vs inferred trees
        RF distance:     4
        RF distance:     4
        RF distance:     4
        t1 vs inferred trees
        RF distance:     5
        RF distance:     3
        RF distance:     5
        t2 vs inferred trees
        RF distance:     4
        RF distance:     4
        RF distance:     4
        4.2. Make the likelihood comparison.
```

Figure 7: RF

From RF we know the inferred tree is different from the real tree, which probability means that our Mixture tree over-fits the data. The reason may be the sample size is small, the node is too little or the clusters are not enough, which we will discuss in the next question.

5.2 2.4.14

We want to find out the influence of number of sample, number of nodes and number of clusters on the final result. We have already get the result for sample=100, node=5 and cluster=3.

5.2.1 larger sample

We use sample=1000, nodes=5, cluster=3

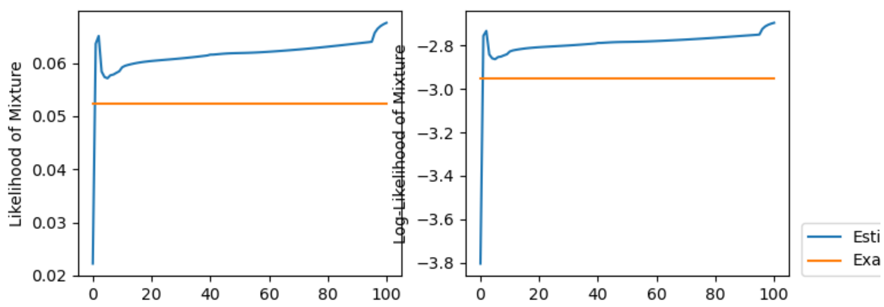


Figure 8: sample=1000,n=5,c=3

As shown in the figure,when the number of sample getting larger,the difference of the inferred model and the real model becomes smaller, this can be recognized both through a closer final probability and a smaller RF matrix

```
4.2 Compare trees and print Robinson-Foulds (RF) distance:

[[4. 4. 4.]
 [3. 3. 3.]
 [4. 4. 4.]]

4.2. Make the likelihood comparison.
```

Figure 9: RF

5.2.2 more nodes

We use sample=100, nodes=20, cluster=3

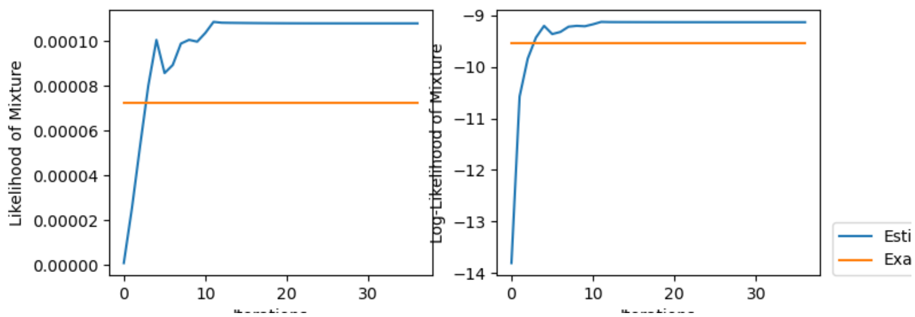
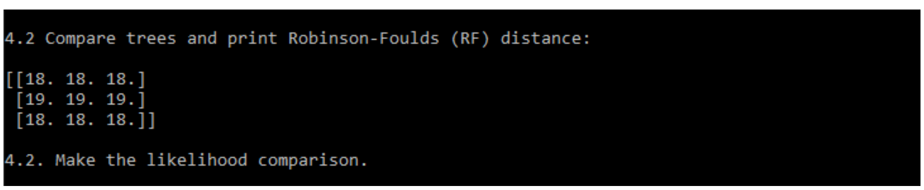


Figure 10: sample=100,n=20,c=3

As shown in the figure,when the number of nodes getting larger,the probability becomes rather small, but the difference of the real and the inferred model gets smaller. This may be because when the number of nodes increase, the uniqueness of getting a certain sample rises and thus the estimated model will be similar to the real one.



```
4.2 Compare trees and print Robinson-Foulds (RF) distance:

[[18. 18. 18.]
 [19. 19. 19.]
 [18. 18. 18.]]

4.2. Make the likelihood comparison.
```

Figure 11: RF

5.2.3 more clusters

We use sample=100, nodes=5, cluster=10

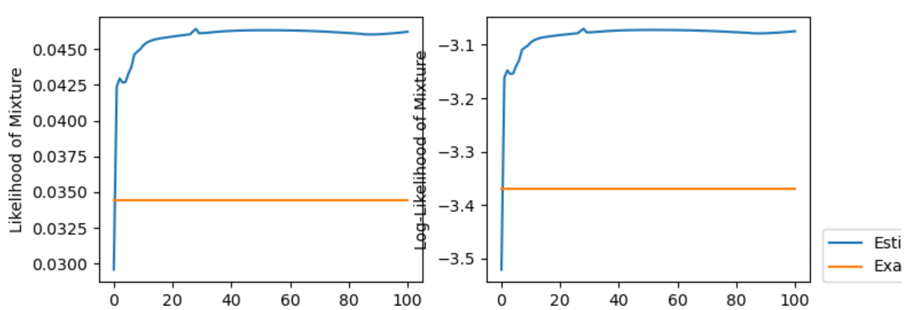


Figure 12: sample=100,n=5,c=10

When the number of cluster increases, the inferred model make a much better outcome than the real model, while the difference between the real and the inferred model is big, this performance is poor as what we want is to estimate the real model not to fit the data.

```
4.2 Compare trees and print Robinson-Foulds (RF) distance:

[[5. 5. 5. 5. 5. 5. 5. 5. 5. 5.]
 [4. 4. 4. 4. 4. 4. 4. 4. 4. 4.]
 [3. 3. 3. 3. 3. 3. 3. 3. 3. 3.]
 [4. 4. 4. 4. 4. 4. 4. 4. 4. 4.]
 [5. 5. 5. 5. 5. 5. 5. 5. 5. 5.]
 [5. 5. 5. 5. 5. 5. 5. 5. 5. 5.]
 [5. 5. 5. 5. 5. 5. 5. 5. 5. 5.]
 [6. 6. 6. 6. 6. 6. 6. 6. 6. 6.]
 [4. 4. 4. 4. 4. 4. 4. 4. 4. 4.]
 [4. 4. 4. 4. 4. 4. 4. 4. 4. 4.]]

4.2. Make the likelihood comparison.
```

Figure 13: RF