1 Relabeling Algorithm in High Dimension

Since the algorithm has been well explained in previous sections and in our previous project, it would not be repeated here. I would like to describe the output of relabeling algorithm for high dimensional data. And explaination will be given.

1.1 Simulated Data

This simulated data set is of 4 components of 3 dimension. MCMC samples are generated by Gibbs sampling algorithm.

The dataset are simulated by randomly choosing from 4 normal distributions whose first dimensional values are -2,0,2,and -1 with weights of 0.125,0.125,0.25,and 0.5 respectively. And the true mean values of first dimension are -2.21,0.174,1.78,and -0.933 with true weights of 0.12,0.18,0.22,and 0.48 respectively.

The plot shows that label switching problem happens here. But it is not evenly or totally,i.e. the label switching is unbalanced. The effect of the label switching is obvious. In this example, two modes of MCMC is well mixed up here, while the other two stands alone. But this cannot be guarranteed. The rest two modes might be coming into mixing when we run enough many iterations. Hence for prediction questions, we might end up with 3 components for this simulated dataset, which is not correct.

An extreme case of unbalanced label switching is that every component remains its own mode for quite a long time, i.e. when MCMC sample size is not big enough, the label switching cannot be observed.

The unbalanced label switching can also be shown from the MCMC sample of estimates for weights of each component. We can see that weight estimates for each components are quite different from their real values.

To simulate the label-switching, or to solve unbalanced label switching problem, we apply Fruhwirth Random Permutation Algorithm.

We can see that for each component switches from different modes.

Estimates of weights (after burn-in) by MCMC samples is 0.21, 0.266,0.261,and 0.264, which are roughtly close 0.25. When we change the burn size, we can see that the weights change and difference is getting smaller.

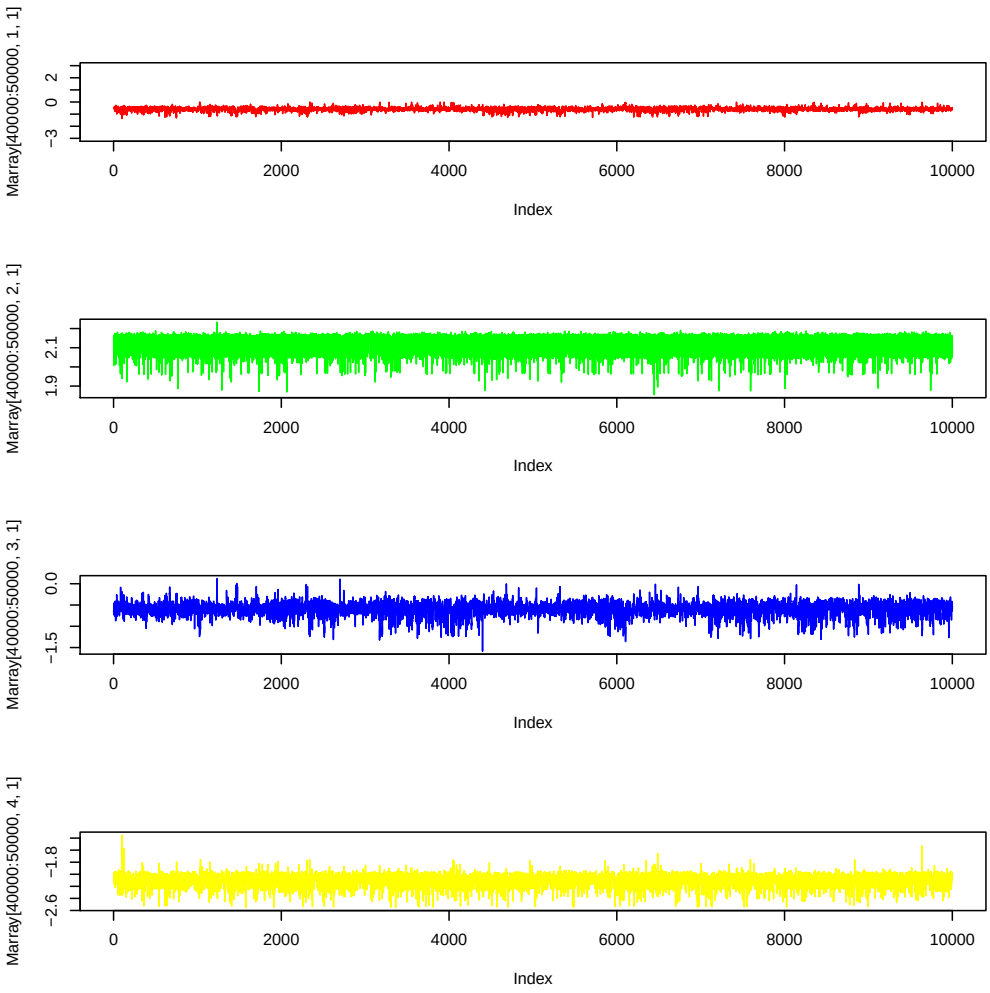

Figure 1: Mean of 1st dimension for 4 components of 3-dimensional data. True means are -2, 0, 2, and -1 for component 1-4 respectiely. The label switching is unbalanced.

2 Relabeling Algorithm

We mainly apply Steven's algorithm to solve label switching problem. We also apply some other algorithm such as identifiability constraint and comparison between these methods will be given.

Due to time limit, I would like to give the result on 3 components of a

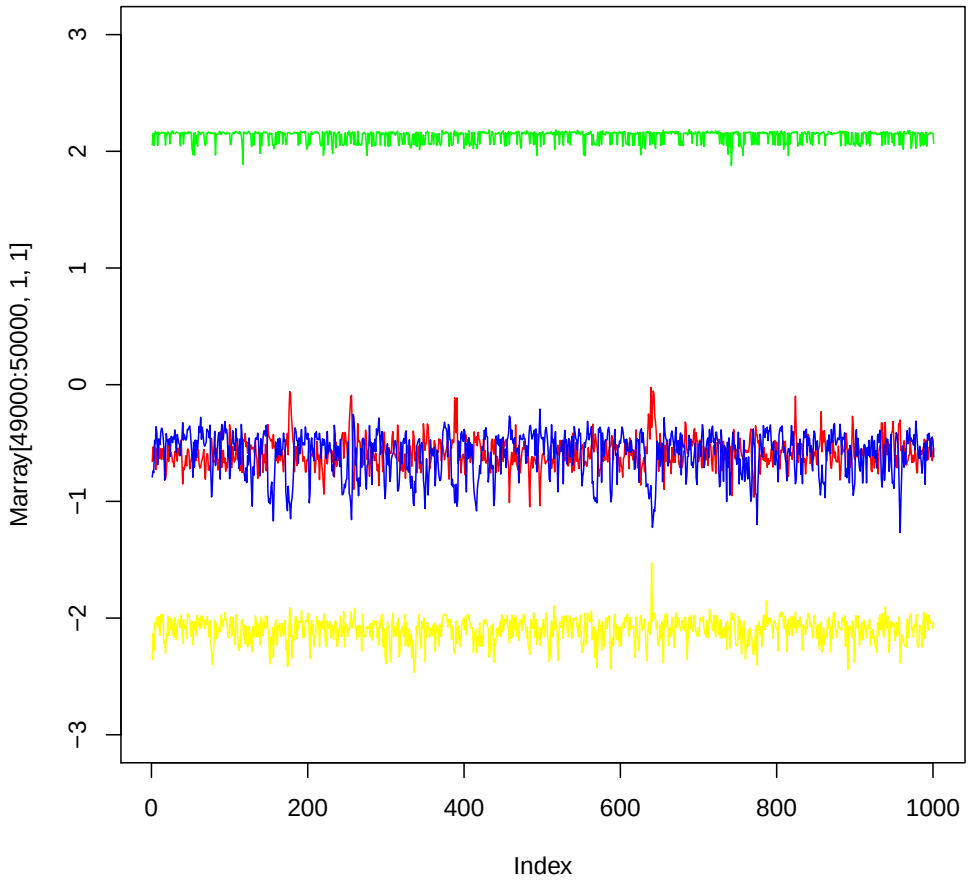

Figure 2: Mean of 1st dimension for 4 components of 3-dimensional data in one plot shows an unbalanced label switching problem.

three dimensional dataset. MCMC samples are generated by Gibbs sampling and Fruhwirth Random Permutation Algorithm.

The dataset are simulated by randomly choosing from 3 normal distributions whose first dimensional values are -2,0,and 2 with weights of 0.50,0.25,and 0.25 respectively. And the true mean values of first dimension are -1.83,-0.306,and 1.70 with true weights of 0.52,0.24,and 0.24 respectively.

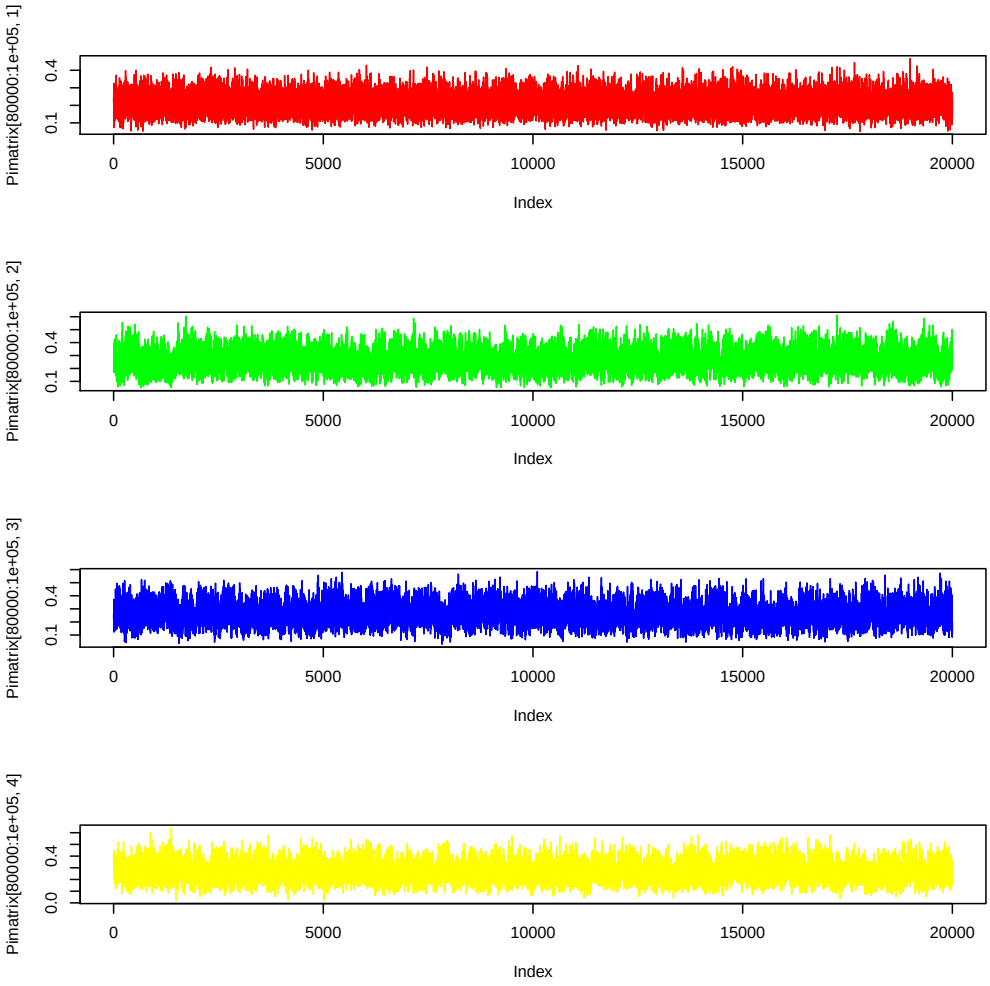

Figure 3: Weights of 4 components. True weights are 0.12, 0.18, 0.22, and 0.48. But weights of simulated MCMC are 0.2101204, 0.2651040, 0.2628557, and 0.2619199 which shows label switching happens here

The plot shows that label switching problem happens here. And it is almost evenly and totally,i.e. the label switching is roughly balanced since we only run for 10000 iterations. The effect of the label switching is obvious. In this example, Three modes of MCMC are well mixed up here.

Then we apply Steven's relabeling algorithm on the MCMC samples (burn-in size is 6000). The plots shows that the mix-up almost complet-

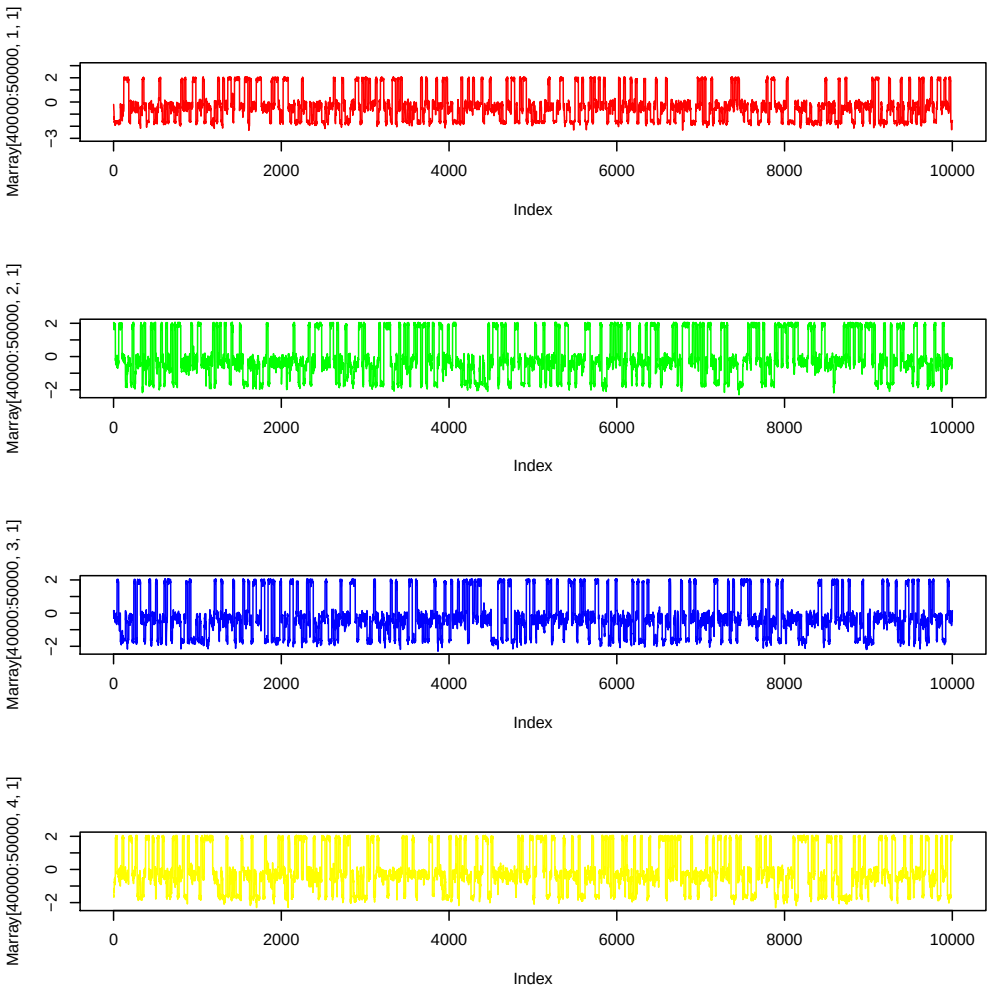

Figure 4: MCMC estimates for 1st dimension value for 4 components shows an obvious label switching between 4 components.

edly separated. The means of the first dimension are -2.02,1.51,and -0.614.

We also applied identifiability restriction on this MCMC sample. By restriction on the order of weights, the means of the first dimension are -0.377,-0.372,and -0.380.

By restriction on the order of first dimension value, the means of the first dimension vlaues are -2.02,-0.611, and 1.50.

By comparing of above results, we can see that Steven's relabeling algo-

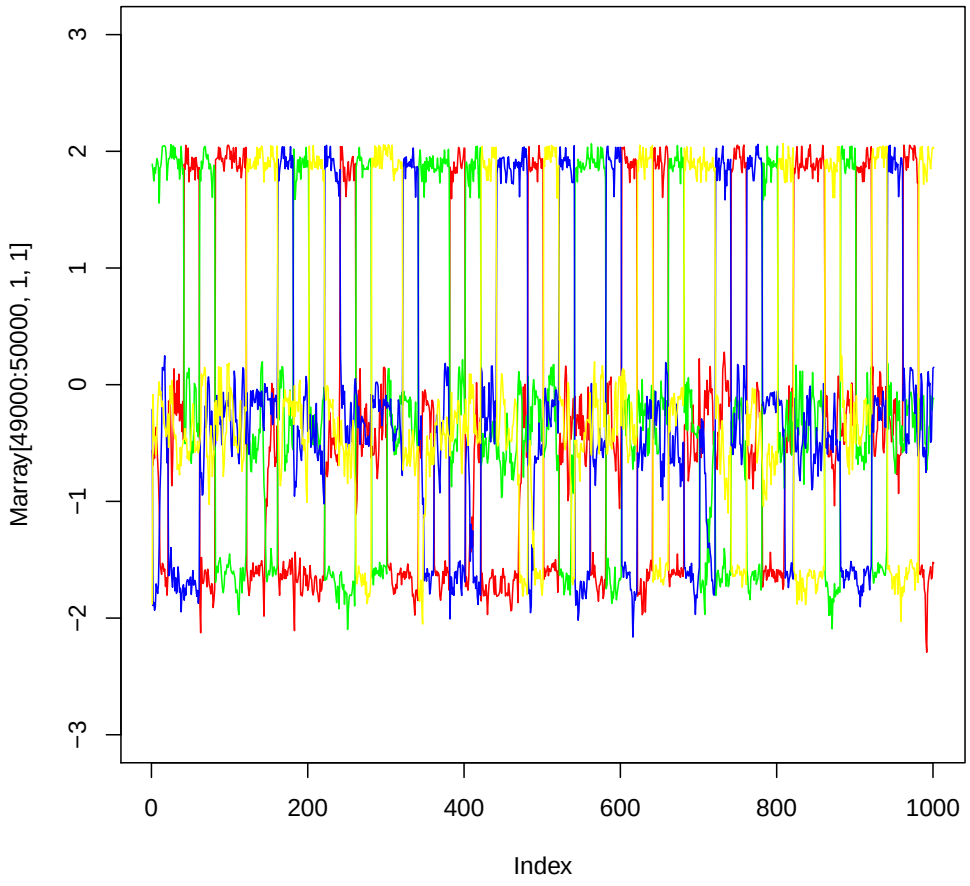

Figure 5: MCMC estimates for 1st dimension value for 4 components in one plot.All modes are totally mixed up.

rithm gives the best result. By the order of the value dimensional data value also gives a fairly good value. By the order of weights does not come up with a good result at all since the weight itself is totally mixed up.

Steven's relabeling algorithm is preferred since identifiability restriction for hight dimensional data is usually not easy to give. Also, as Steven argued that many of the choice will be ineffective and label-switching problem will remain.

Table 1: Component Weights Change with Sample Size

Sample Size(Start:End)	π_i	π_2	π_3	π_4
40000:50000	0.2089794	0.2658891	0.2607384	0.2643931
40000:60000	0.2092240	0.2631768	0.2630629	0.2645363
40000:70000	0.2097820	0.2644694	0.2626630	0.2630856
40000:80000	0.2097905	0.2657061	0.2633637	0.2611397
40000:90000	0.2096614	0.2644653	0.2638334	0.2620399
40000:100000	0.2098225	0.2644618	0.2629233	0.2627923

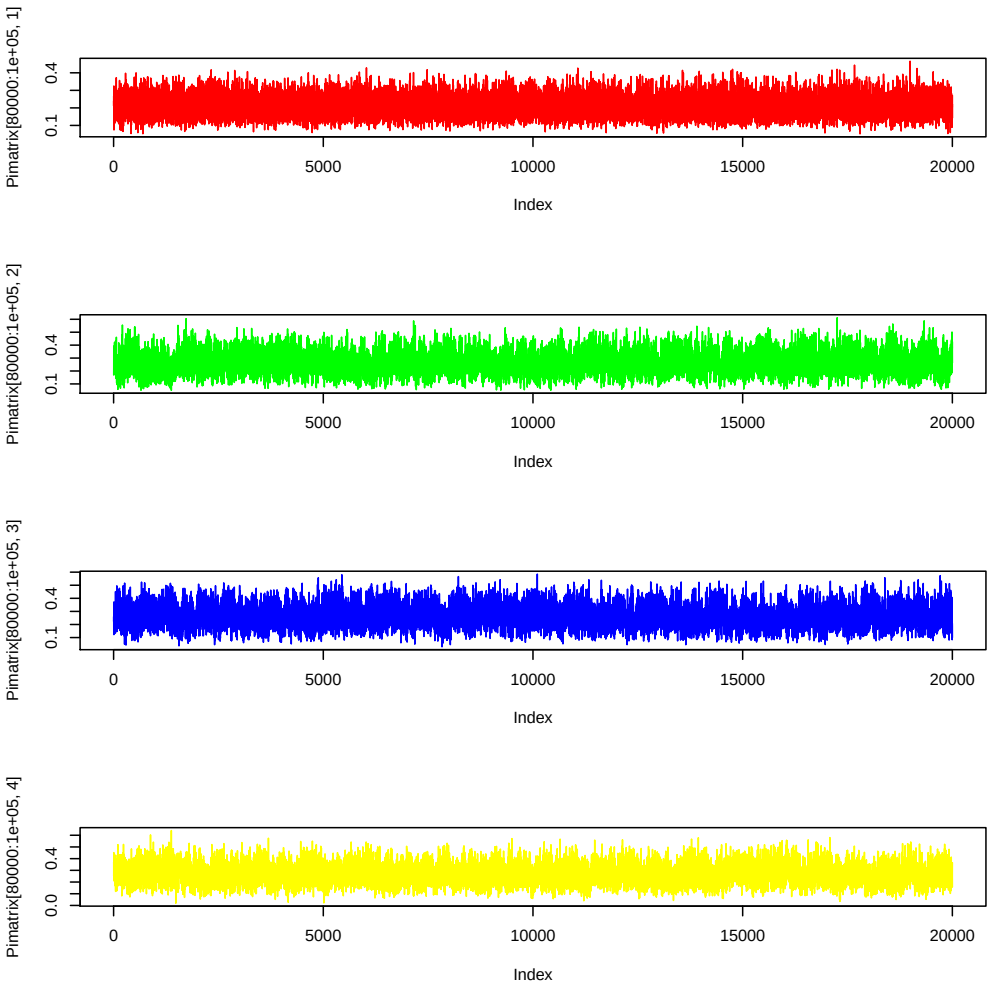

Figure 6: Weights of 4 components. True weights are 0.12, 0.18, 0.22, and 0.48. But weights of simulated MCMC are 0.2101204 0.2651040 0.2628557 0.2619199 which shows label switching happens here.

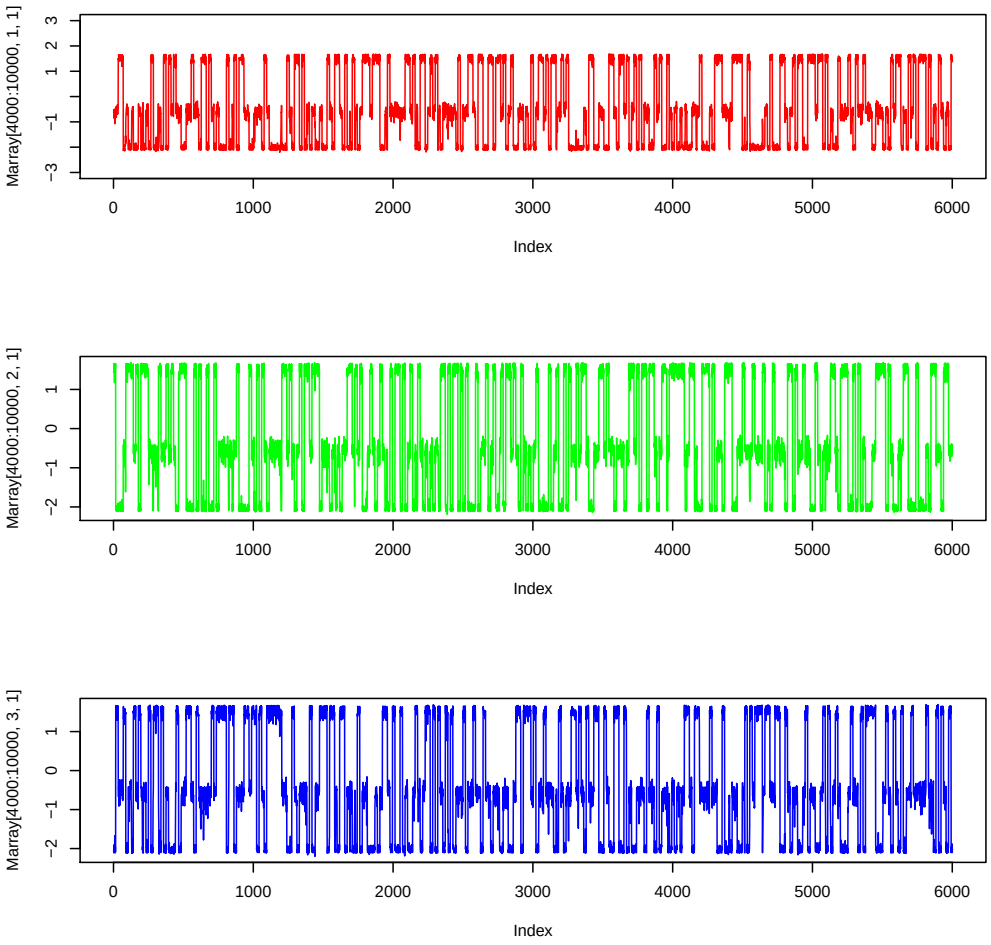

Figure 7: MCMC estimates for 1st dimension value for 4 components shows an obvious label switching between 4 components.

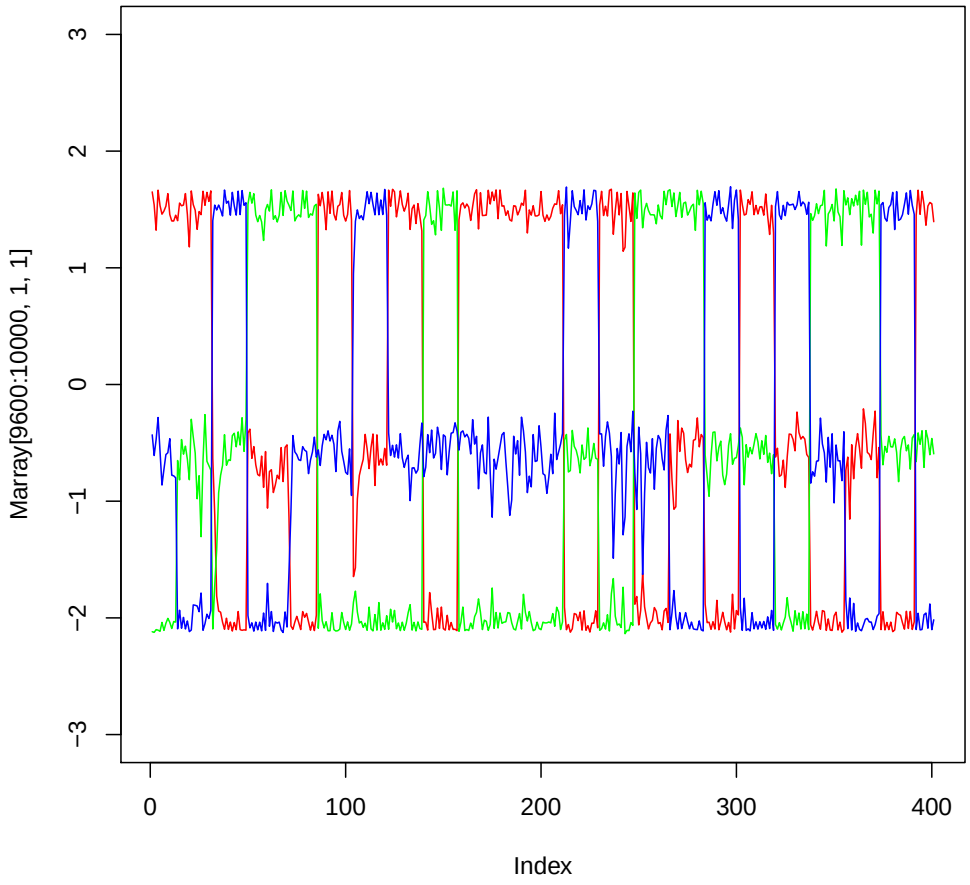

Figure 8: MCMC estimates for 1st dimension value for 3 components in one plot.All modes are totally mixed up.

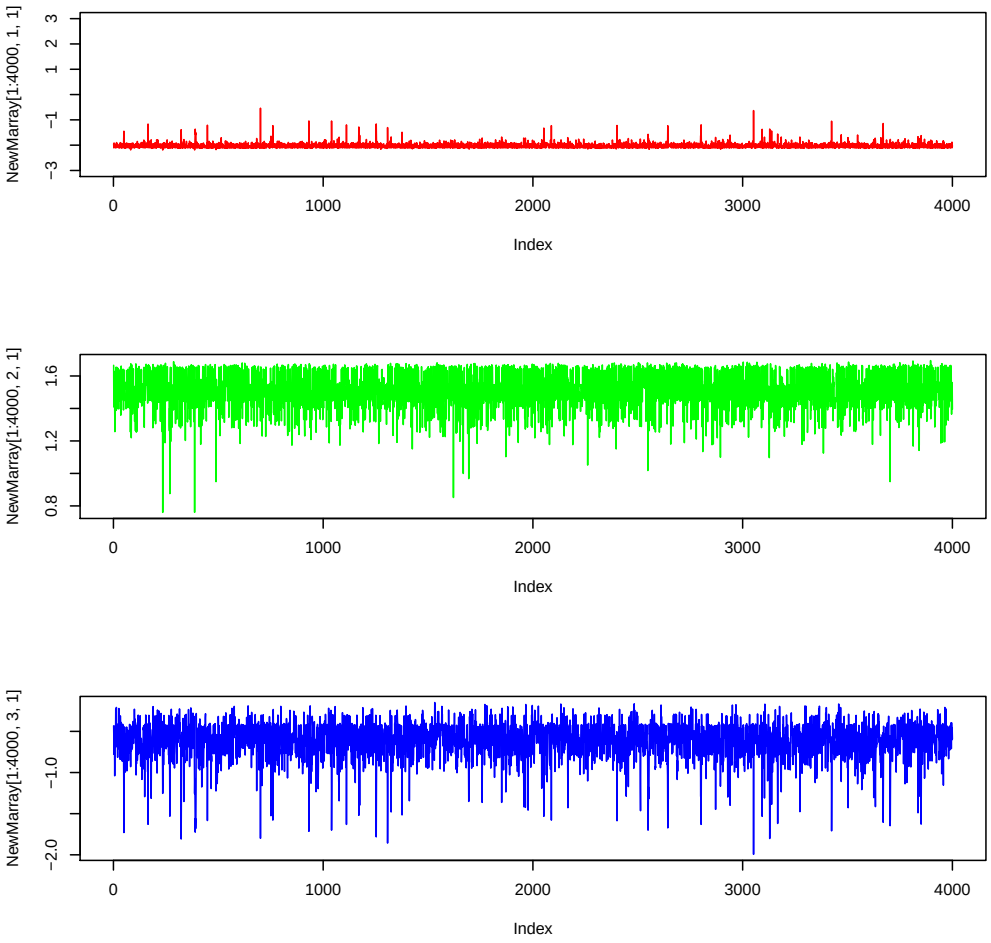

Figure 9: MCMC estimates for 1st dimension value for 3 components after running relabeling algorithm.

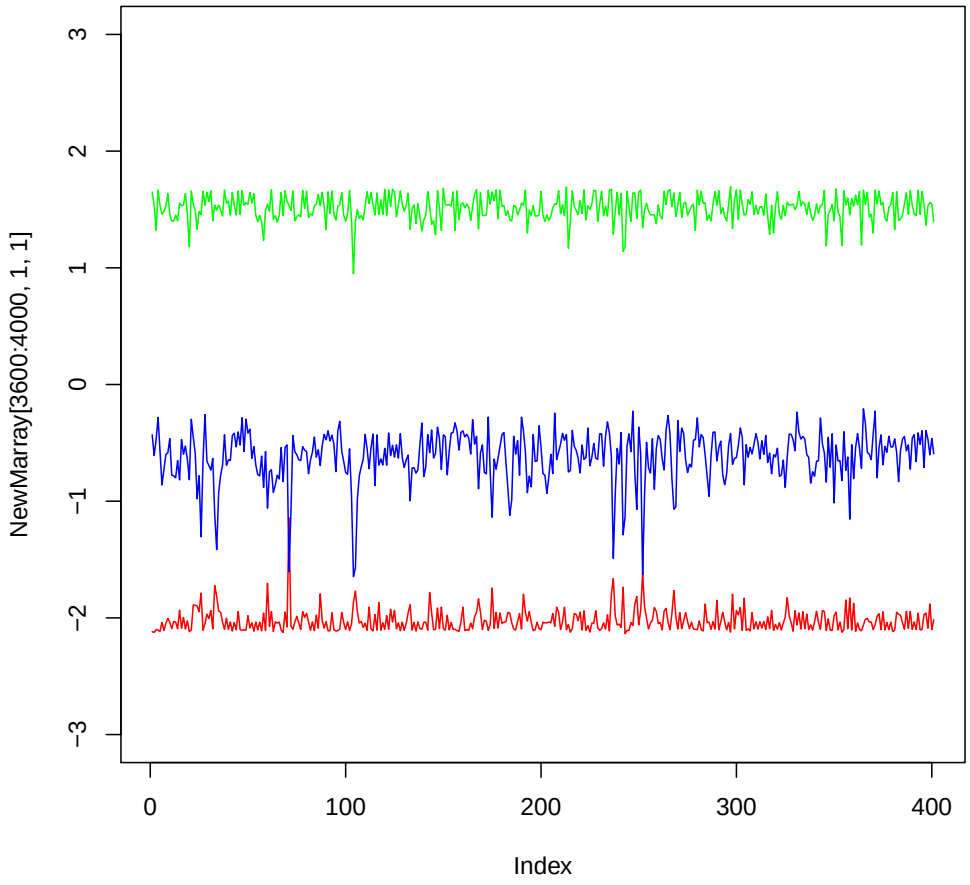

Figure 10: MCMC estimates for 1st dimension value for 3 components in one plot after running relabeling algorithm.

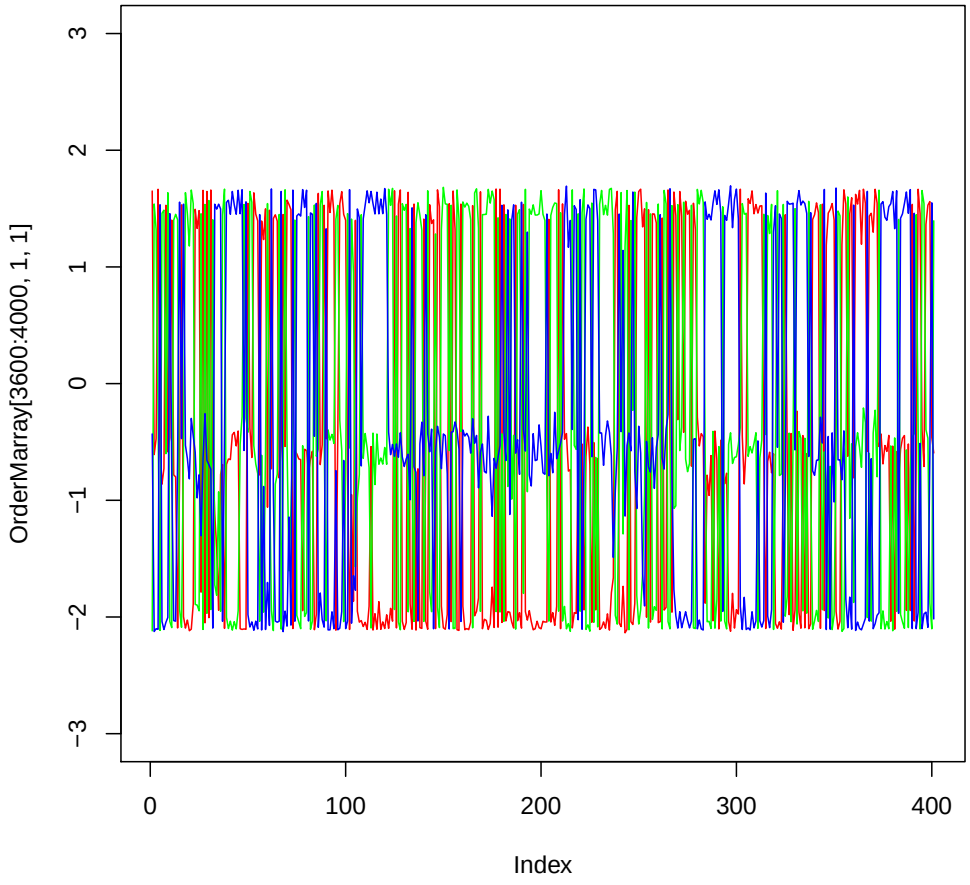

Figure 11: MCMC estimates for 1st dimension value for 3 components in one plot after running identifiability algorithm (by order of weights).

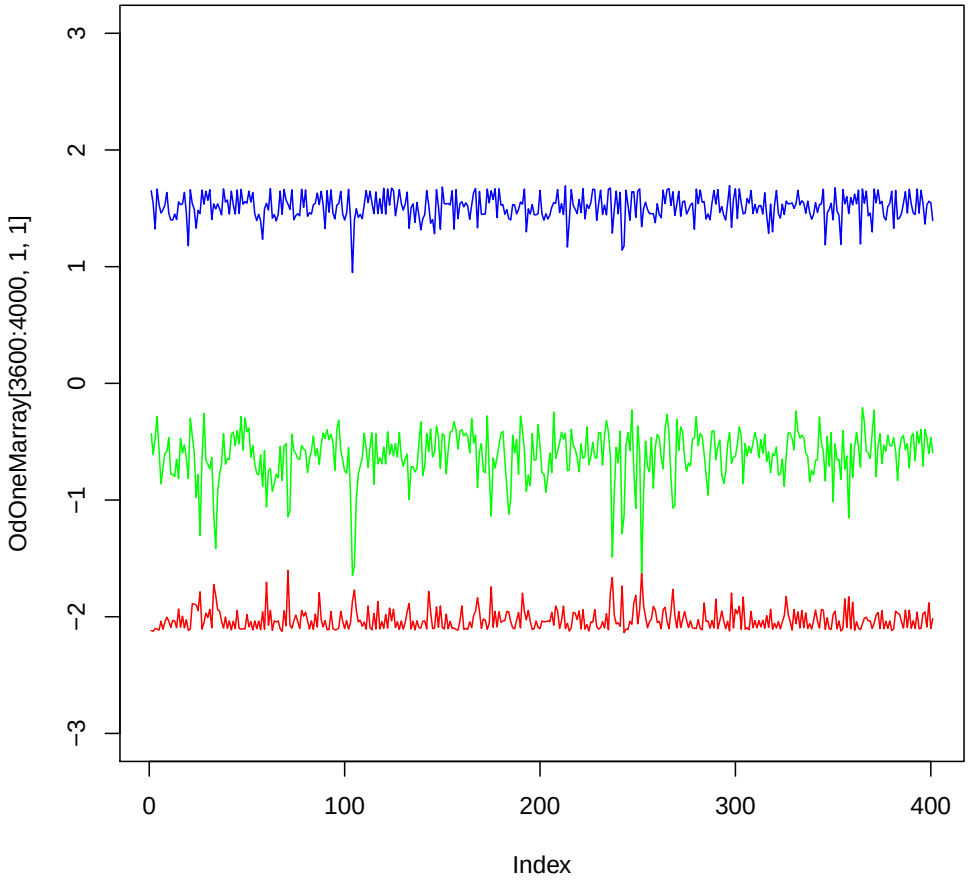

Figure 12: MCMC estimates for 1st dimension value for 3 components in one plot after running identifiability algorithm (by order of first dimension's value).