Proposal for Advanced Applied Statistics

Yu Han

1. Introduction

It is common for researchers to use high dimensional matrices to store large collections of data and extract information from the covariance matrices. Therefore, research on high dimensional covariance matrices means a lot to analyzing big data and putting it to application. Researchers from home and abroad have studied a lot and enjoyed several mature results on the estimation of static matrices. However, in some real cases, covariance may vary with other variables, for instance, time. In medical imaging, if we statistically store MRI signals of a patient at every particular time, then its covariance matrix will be time-dependent. So it will be of great value if we can detect the dynamic property of covariance matrices and apply appropriate methods to analysis the covariances through data.

2. Data Description

I am gonna analyze the fMRI data of the attention deficit hyperactivity disorder (ADHD) patients. The data set was made available by Neuro Bureau and the ADHD-200 consortium and can be downloaded from https://www.nitrc.org/plugins/mwiki/index.php/neurobureau:AthenaPipeline. I wil focus on the fMRI behavior of Individual 010002 and try to capture its fMRI signals at different time.

First, I read the data and look into its dimension.

```
MyData <- read.csv(file="/Users/HanY/Desktop/0010002.csv", header=TRUE, sep=",")
nrow(MyData)
```

```
## [1] 172
```

```
ncol(MyData)
```

```
## [1] 353
```

Then I get some infomation of the data. The data consists 172 scanned data of 351 different regions of interest (ROI).

```
MyData[1:5,1:5]
```

```
##                                       File Sub_brick    Mean_1    Mean_3
## 1 sfnwmrda0010002_session_1_rest_1.nii.gz         1 -0.010454 -0.089631
## 2 sfnwmrda0010002_session_1_rest_1.nii.gz         2  0.006257 -0.094698
## 3 sfnwmrda0010002_session_1_rest_1.nii.gz         3  0.016927 -0.092042
## 4 sfnwmrda0010002_session_1_rest_1.nii.gz         4  0.006936 -0.082161
## 5 sfnwmrda0010002_session_1_rest_1.nii.gz         5 -0.029020 -0.060132
##      Mean_4
## 1 -0.064186
## 2 -0.072936
## 3 -0.052112
## 4 -0.011082
## 5  0.028198
```

To check the dynamic characteristics of the data, I split the 172 observations into two sets, each consisting 86 observations. Then, I plot the heatmaps of the first 20 ROIs for them respectively.

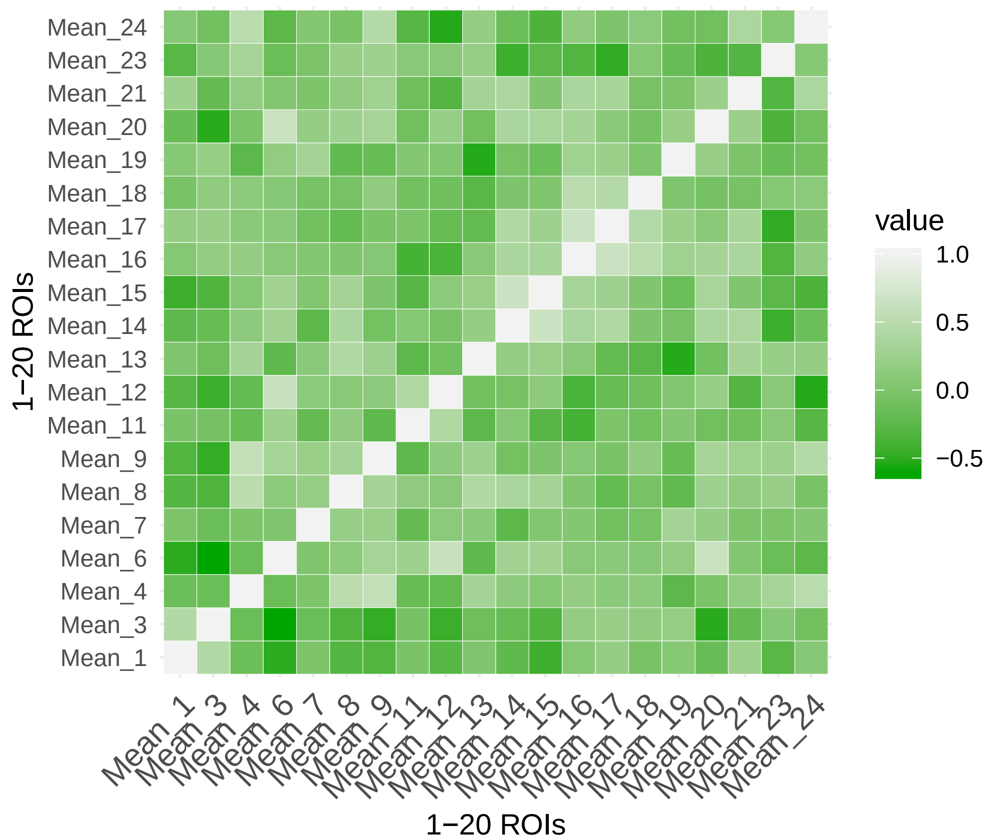
```
MyData <- read.csv(file="/Users/HanY/Desktop/0010002.csv", header=TRUE, sep=",")
dim(MyData)
```

```
## [1] 172 353
```

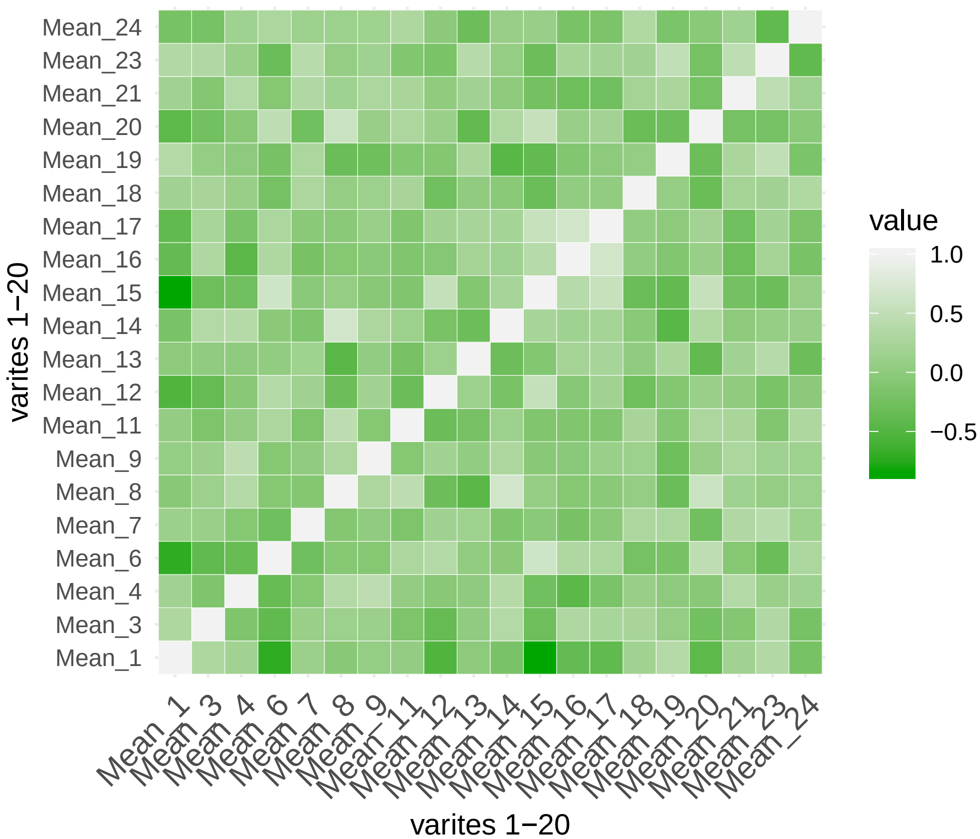
```r
MyData[1:5,1:5]
```

```
##                                     File Sub_brick    Mean_1    Mean_3
## 1 sfnwmrda0010002_session_1_rest_1.nii.gz         1 -0.010454 -0.089631
## 2 sfnwmrda0010002_session_1_rest_1.nii.gz         2  0.006257 -0.094698
## 3 sfnwmrda0010002_session_1_rest_1.nii.gz         3  0.016927 -0.092042
## 4 sfnwmrda0010002_session_1_rest_1.nii.gz         4  0.006936 -0.082161
## 5 sfnwmrda0010002_session_1_rest_1.nii.gz         5 -0.029020 -0.060132
##      Mean_4
## 1 -0.064186
## 2 -0.072936
## 3 -0.052112
## 4 -0.011082
## 5  0.028198
```

```r
half_data1 = MyData[1:86,3:22]
half_data2 = MyData[87:172,3:22]
corr_data1 = cor(half_data1)
corr_data2 = cor(half_data2)
library(reshape2)
library(ggplot2)
melt_data1 = melt(corr_data1)
melt_data2 = melt(corr_data2)
ggplot(data = melt_data1, aes(Var2, Var1, fill = value))+
  xlab('1-20 ROIs')+ylab("1-20 ROIs")+
  geom_tile(color = "white")+
  scale_fill_gradientn(colours = terrain.colors(2)) +
  theme_minimal()+
  theme(axis.text.x = element_text(angle = 45, vjust = 1,
                                   size = 12, hjust = 1))+
  coord_fixed()
```




```
ggplot(data = melt_data2, aes(Var2, Var1, fill = value))+
  xlab('varites 1-20')+ylab("varites 1-20")+
  geom_tile(color = "white")+
  scale_fill_gradientn(colours = terrain.colors(2)) +
  theme_minimal()+
  theme(axis.text.x = element_text(angle = 45, vjust = 1,
                                   size = 12, hjust = 1))+
  coord_fixed()
```



It can be seen clearly in the heatmaps that the correlations among the first 20 ROIs vary a lot with time going. So it may be reasonable for us to capture the dynamic property of the data.

3 Project Goal

I hope to show that the covariance of different brain regions signals of the ADHD patients will vary with time. Thus, it is expected that using dynamic covariance matrices will outperform the static ones in matrix estimation. Once we can verify the dynamic characteristics of the data, we can choose appropriate model and detect the behaviors of different ROI and the correlations between them.

4 Methodology

I plan to use both static and dynamic covariance matrices to model the covariance of different brain regions signals of the ADHD patients at resting state. For the static ones, I will apply adaptive thresholded covariances to model the data. While for the dynamic ones, I will apply kernel smoothing to elementwise estimate the dynamic covariances. Besides, a test may be applied to show the evidence of the the dynamic property of the data.