IE 522 HW02

1. (1.5 points) $\{X_1, \cdots, X_n\}$ is a random sample from Laplace(μ, b), a Laplace distribution with parameters μ and b.

   (1). (1 point) Using the method of moments, find the point estimators for μ and b. Show details. Use the following box or upload a file.

   (2). (0.5 point) To show whether a dataset is from Laplace(μ, b) for some μ, b, why is it sufficient to construct a probability plot using quantiles of Laplace$(0, 1)$?

2. (8.5 points)

Install the ISLR library in R. Smarket is a data frame from this library that contains returns of S&P 500 in the five year period from 1/10/2001 to 12/30/2005. The following shows the first and last three rows of the data frame:

```
library(ISLR)
n=nrow(Smarket)
Smarket[c(1:3,(n-2):n),]
```

```
##      Year   Lag1   Lag2   Lag3   Lag4   Lag5  Volume  Today Direction
## 1    2001  0.381 -0.192 -2.624 -1.055  5.010 1.19130  0.959        Up
## 2    2001  0.959  0.381 -0.192 -2.624 -1.055 1.29650  1.032        Up
## 3    2001  1.032  0.959  0.381 -0.192 -2.624 1.41120 -0.623      Down
## 1248 2005 -0.955  0.043  0.422  0.252 -0.024 1.54047  0.130        Up
## 1249 2005  0.130 -0.955  0.043  0.422  0.252 1.42236 -0.298      Down
## 1250 2005 -0.298  0.130 -0.955  0.043  0.422 1.38254 -0.489      Down
```

For each date, *Today* is the percentage return of the day. *Direction* indicates whether S&P 500 was going up or down during the day. *Volume* is the trading volume on the previous day (in billions). *Lag1* to *Lag5* are the percentage returns in the previous 5 days. From canvas, download ISLRSmarketDates.csv. It contains the corresponding dates. Replace the first column of Smarket by these dates.

```
dates=read.csv("ISLRSmarketDates.csv",header=TRUE)
sp=data.frame(dates,Smarket[,-1])
n=nrow(sp)
sp[c(1:3,(n-2):n),]
```

```
##            Date   Lag1   Lag2   Lag3   Lag4   Lag5  Volume  Today Direction
## 1     1/10/2001  0.381 -0.192 -2.624 -1.055  5.010 1.19130  0.959        Up
## 2     1/11/2001  0.959  0.381 -0.192 -2.624 -1.055 1.29650  1.032        Up
## 3     1/12/2001  1.032  0.959  0.381 -0.192 -2.624 1.41120 -0.623      Down
## 1248 12/28/2005 -0.955  0.043  0.422  0.252 -0.024 1.54047  0.130        Up
## 1249 12/29/2005  0.130 -0.955  0.043  0.422  0.252 1.42236 -0.298      Down
## 1250 12/30/2005 -0.298  0.130 -0.955  0.043  0.422 1.38254 -0.489      Down
```

2.1 (1.5 point) Construct a histogram for *Today* - the percentage return of S&P 500. Make the vertical axis density instead of frequency. Set the number of bins to 50. Add a normal fit to the histogram you obtain. Is the normal distribution fitting the peak well? Add "xlim=c(2,6),ylim=c(0,0.1)" to the hist function to zoom in on the right tail. Make sure you have two plots side by side. Is the normal distribution fitting the tails well?

2.2 (1 point) Suppose a Laplace distribution with parameters μ and b is used to model *Today*. Estimate μ and b. Using the dlaplace(x,μ,b) function in the VGAM library (dlaplace(x,μ,b) is the pdf of the Laplace distribution), add a Laplace fit to the histogram for *Today*. Then, as in the previous question, zoom in on the right tail. Again, you should have two plots side by side. Is the Laplace distribution fitting the peak and tails better than the normal distribution?

2.3 (1.5 points) Add KDEs to the histogram for *Today* using 1/4 of the default bandwidth and four times the default bandwidth (on a single plot). Which fits the data better? Which is less oscillating and smoother? What does these results say about the choice of the bandwidth? (Search to find out how to adjust the bandwidth when you call the *density* function in R.)

2.4 (1 point) Construct a boxplot for *Today*. Investigate the most negative outlier (you could use min and subset functions in R to identify this outlier). Is this outlier there by error or is it a legitimate data point? If it is legitimate, what caused this extreme negative return on that particular date?

2.5 (1 point) Report the correlation matrix for all the variables excluding Date and Direction. Do you see any strong linear relatioship? Construct a scatterplot matrix for these variables. Does it show any significant linear or nonlinear relationship?

2.6 (1 point) Construct a time series plot for Volume. What do you observe near the end of each year? What was the reason?

2.7 (1.5 points) Construct a normal plot and a Laplace plot for *Today*, side by side. Make sure that the ranges of the x-axis are the same for both plots. Between the normal distribution and the Laplace distribution, which is fitting the data better? Are the tails of the distribution for percentage return of S&P 500 fatter or thinner than those of a normal distribution? What are the sample skewness and kurtosis of *Today*?