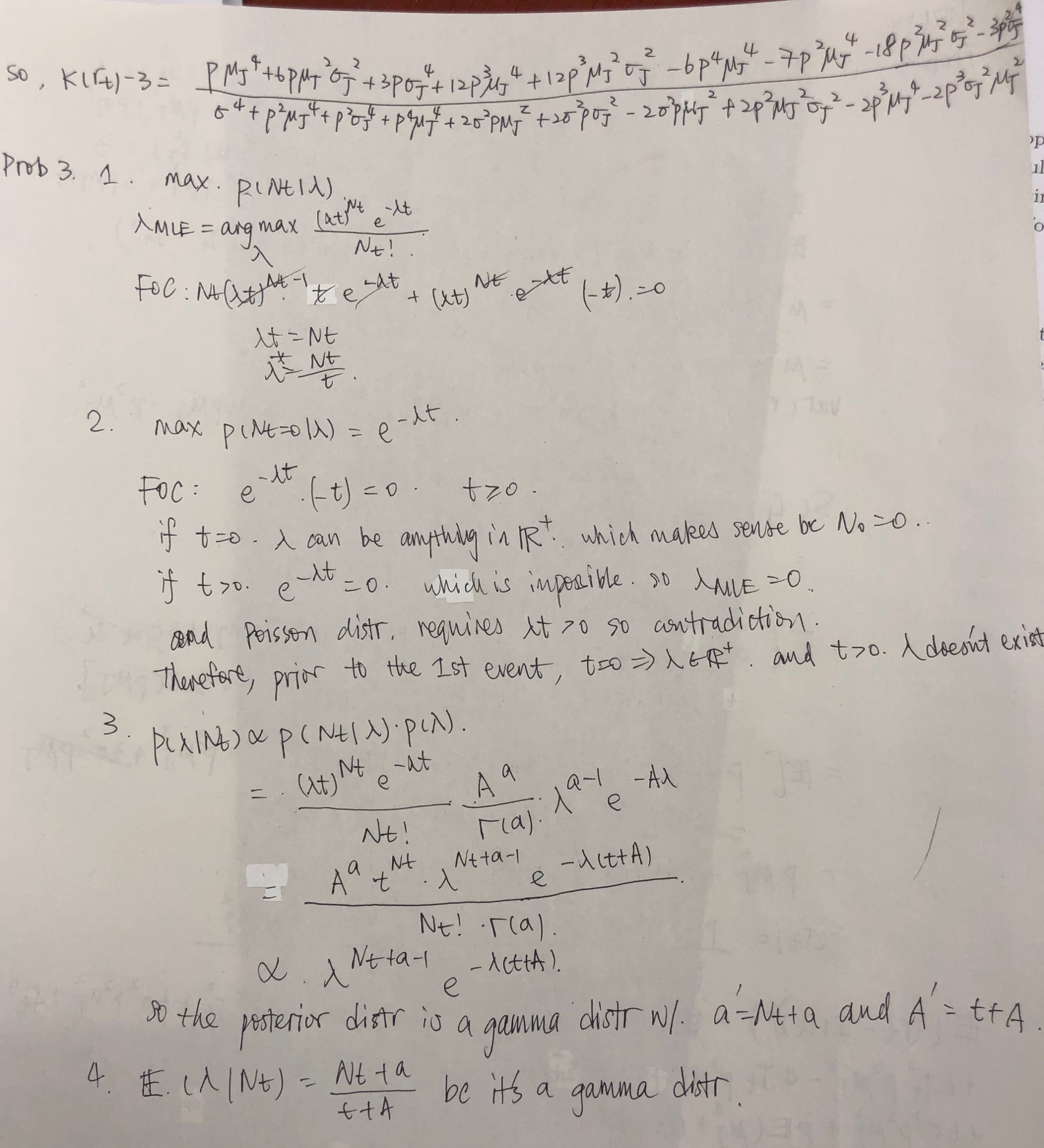
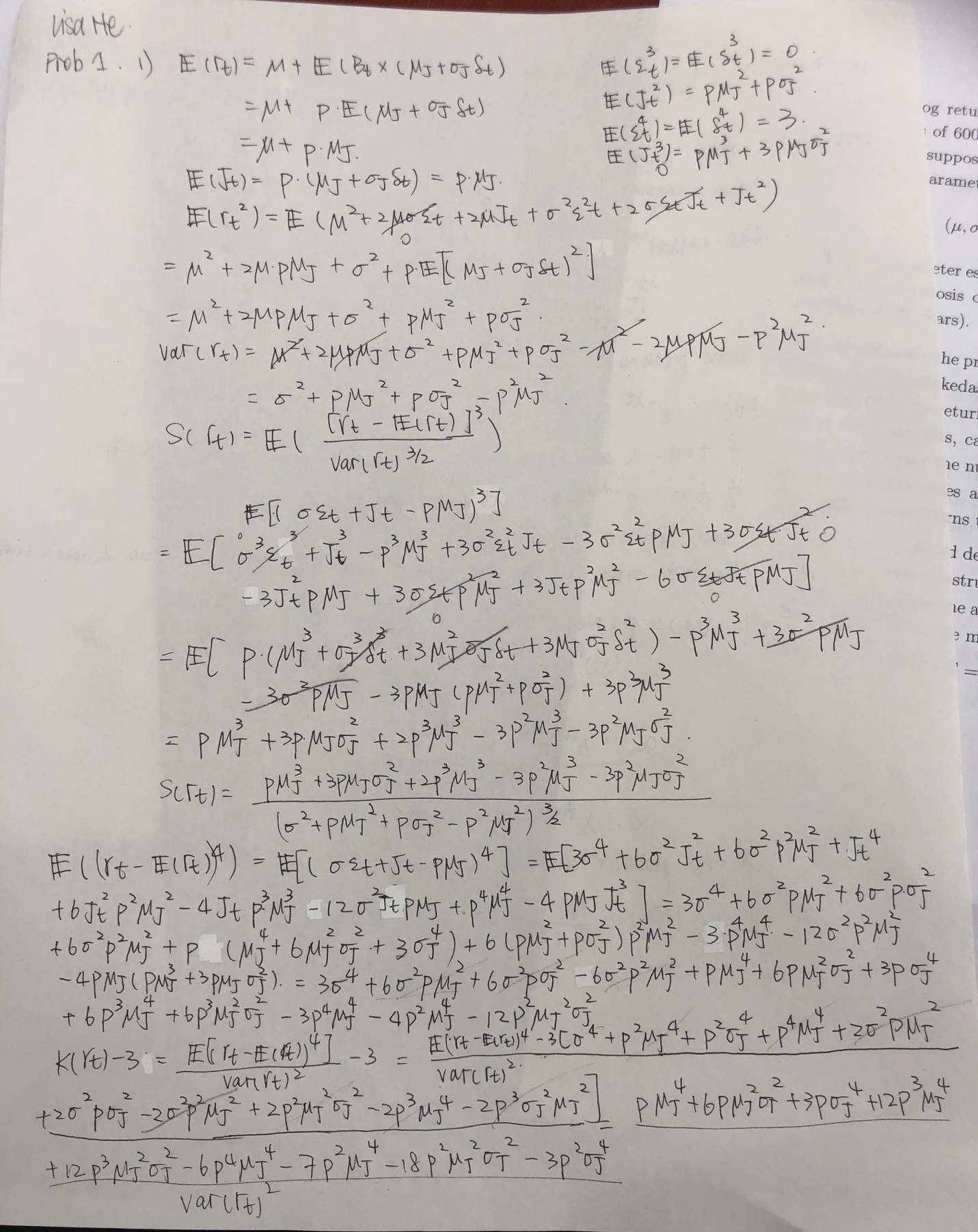
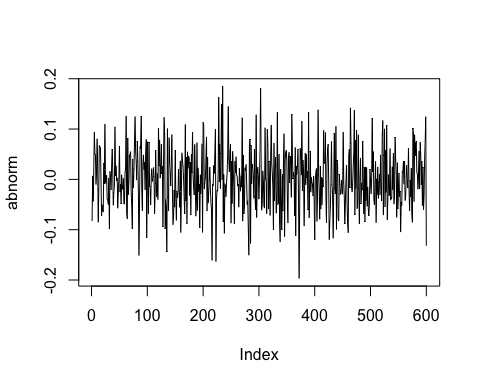
prob1

Lisa He



1. Becuase we have more flexibility with bernoulli normal misture model and that we can see more negative jumps than positive jumps which is consistent with the data. the variance is not consistent over time which is also consistent with market data. and the stock return can have fat tails under the bernoulli normal misture model

N <- 600  
set.seed(124)  
norm <- rnorm(N)  
abnorm <- norm\*0.063 + 0.0045  
plot(abnorm,type = "l")

 It does not look like the data, because the data is non-stationary. there are times where there's a lot of variance in the returns and times where there isnt, however, the simulation from normal distribution's variance is stationary aross time.

mu <- 0.012  
sig <- 0.05   
p<- 0.15  
muj<- -0.05  
sigj <- 0.17  
#mean 0.0045  
mu + p\*muj

## [1] 0.0045

#variance =0.00715375  
sig\*\*2+p\*muj\*\*2+p\*sigj\*\*2-p\*\*2\*muj\*\*2

## [1] 0.00715375

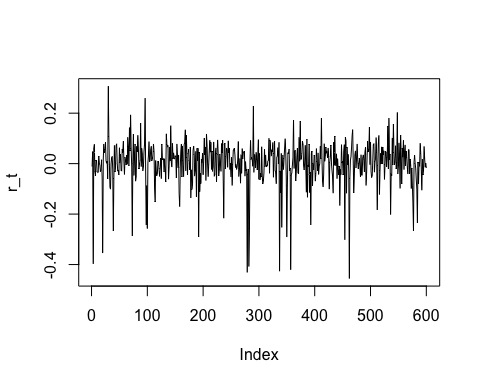
#skewness = -0.9319173  
(p\*muj\*\*3+3\*p\*muj\*sigj\*\*2+2\*p\*\*3\*muj\*\*3-3\*p\*p\*muj\*\*3-3\*p\*p\*muj\*sigj\*\*2)/(sig\*\*2+p\*muj\*\*2+p\*sigj\*\*2-p\*\*2\*muj\*\*2)\*\*(3/2)

## [1] -0.9319173

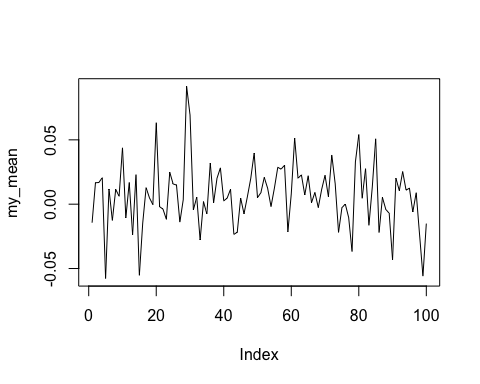
#kurtosis =7.002188  
(3\*sig\*\*4+6\*sig\*sig\*p\*muj\*\*2+6\*sig\*sig\*p\*sigj\*sigj-6\*sig\*sig\*p\*p\*muj\*muj+p\*muj\*\*4+6\*p\*muj\*muj\*sigj\*sigj+3\*p\*sigj\*\*4+6\*p\*\*3\*muj\*\*4+6\*p\*\*3\*muj\*\*2\*sigj\*\*2-3\*p\*\*4\*muj\*\*4-4\*p\*p\*muj\*\*4-12\*p\*p\*muj\*\*2\*sigj\*\*2)/(sig\*\*2+p\*muj\*\*2+p\*sigj\*\*2-p\*\*2\*muj\*\*2)\*\*2-3

## [1] 7.002188

delta\_t <- rnorm(N)  
epsilon\_t <- rnorm(N)  
B\_t <- rbinom(N, size = 1, prob = p)  
J\_t <- B\_t\*(muj+sigj\*delta\_t)  
r\_t <- mu + sig\*epsilon\_t+J\_t  
plot(r\_t,type = "l")

 It doesn't look like the data, but there is a lot more negative jumps than positive jumps. what;s missing is that in the real data, there's periods of times with a lot of negative jumps and positive jumps and some periods with neither. whereas the simulation doesn't achieve that

numN <- 100  
N<- 12  
my\_mean <- c()  
my\_error <-c()  
for (i in 1:numN) {  
delta\_t <- rnorm(N)  
epsilon\_t <- rnorm(N)  
B\_t <- rbinom(N, size = 1, prob = p)  
J\_t <- B\_t\*(muj+sigj\*delta\_t)  
r\_t <- mu + sig\*epsilon\_t+J\_t  
my\_mean[i] <- mean(r\_t)  
my\_error[i] <- std.error(r\_t)  
}  
t\_mean <- mean(my\_mean)  
t\_sd <- sqrt(var(my\_mean))  
#var is var(my\_mean) = 0.0006082747  
t\_stat <- t\_mean/(t\_sd/sqrt(numN))  
plot(my\_mean,type= "l")



# it looks like a distribution from the plot  
skewness(my\_mean) #0.1260242

## [1] 0.1707318

kurtosis(my\_mean)#1.373097

## [1] 1.406444

jarque.bera.test(my\_mean)

##   
## Jarque Bera Test  
##   
## data: my\_mean  
## X-squared = 9.8247, df = 2, p-value = 0.007355

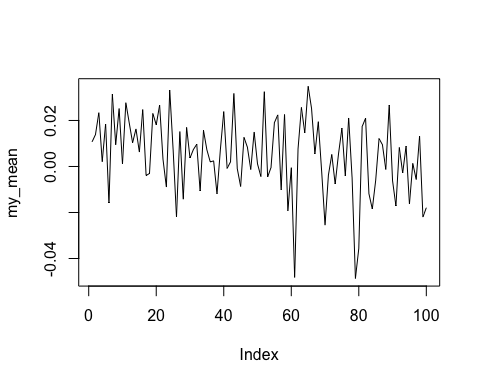
#p-value = 0.01017 so it couldn't be a normal distribution  
mean(my\_error)

## [1] 0.02123394

#the mean of the standard errors are 0.02217715 which is slightly higher than the standard deviation

t\_stat is 3.044169 which is greater than 1.96 so we reject the null hypothesis

numN <- 100  
N<- 24  
my\_mean <- c()  
my\_error <-c()  
for (i in 1:numN) {  
delta\_t <- rnorm(N)  
epsilon\_t <- rnorm(N)  
B\_t <- rbinom(N, size = 1, prob = p)  
J\_t <- B\_t\*(muj+sigj\*delta\_t)  
r\_t <- mu + sig\*epsilon\_t+J\_t  
my\_mean[i] <- mean(r\_t)  
my\_error[i] <- std.error(r\_t)  
}  
t\_mean <- mean(my\_mean)  
t\_sd <- sqrt(var(my\_mean))  
#var is var(my\_mean) = 0.0003389792  
t\_stat <- t\_mean/(t\_sd/sqrt(numN))  
plot(my\_mean,type= "l")



# it doesnt look like a distribution from the plot  
skewness(my\_mean) #-0.5748082

## [1] -0.647919

kurtosis(my\_mean)#0.5307763

## [1] 0.6570971

jarque.bera.test(my\_mean)

##   
## Jarque Bera Test  
##   
## data: my\_mean  
## X-squared = 9.4395, df = 2, p-value = 0.008918

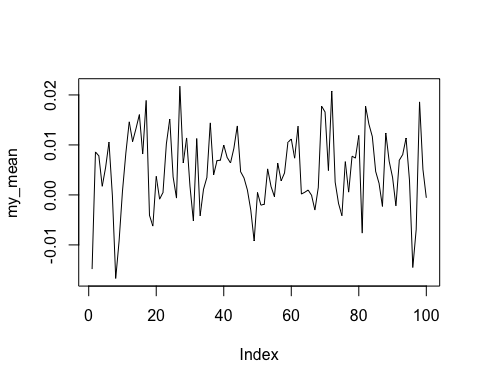
#p-value = 0.02749 so it couldn't be a normal distribution  
mean(my\_error)

## [1] 0.01653043

#the mean of the standard errors are 0.01663182 which is slightly lower than the standard deviation

t\_stat is 0.15277 which is less than 1.96 so we keep the null hypothesis

numN <- 100  
N<- 120  
my\_mean <- c()  
my\_error <-c()  
for (i in 1:numN) {  
delta\_t <- rnorm(N)  
epsilon\_t <- rnorm(N)  
B\_t <- rbinom(N, size = 1, prob = p)  
J\_t <- B\_t\*(muj+sigj\*delta\_t)  
r\_t <- mu + sig\*epsilon\_t+J\_t  
my\_mean[i] <- mean(r\_t)  
my\_error[i] <- std.error(r\_t)  
}  
t\_mean <- mean(my\_mean)  
t\_sd <- sqrt(var(my\_mean))  
#var is var(my\_mean) = 6.445296e-05  
t\_stat <- t\_mean/(t\_sd/sqrt(numN))  
plot(my\_mean,type= "l")



# it looks like a distribution from the plot  
skewness(my\_mean) #0.02659691

## [1] -0.2177322

kurtosis(my\_mean)#-0.3752769

## [1] 0.0683155

jarque.bera.test(my\_mean)

##   
## Jarque Bera Test  
##   
## data: my\_mean  
## X-squared = 0.88539, df = 2, p-value = 0.6423

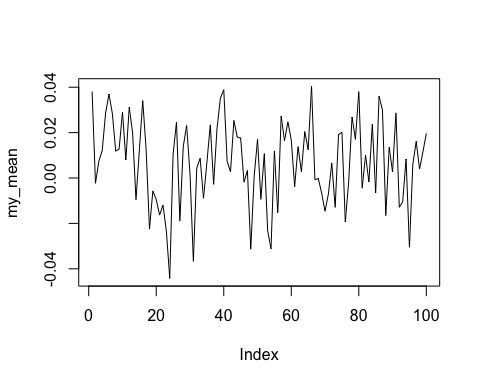
#p-value = 0.8009 so it could be a normal distribution  
mean(my\_error)

## [1] 0.007674494

#the mean of the standard errors are 0.007614783 which is slightly lower than the standard deviation

t\_stat is 6.69 which is greater than 1.96 so we reject the null hypothesis

numN <- 100  
N<- 24  
my\_mean <- c()  
my\_error <-c()  
for (i in 1:numN) {  
delta\_t <- rnorm(N)  
epsilon\_t <- rnorm(N)  
B\_t <- rbinom(N, size = 1, prob = p)  
J\_t <- B\_t\*(muj+sigj\*delta\_t)  
r\_t <- mu + sig\*epsilon\_t+J\_t  
my\_mean[i] <- mean(r\_t)  
my\_error[i] <- std.error(r\_t)  
}  
t\_mean <- mean(my\_mean)  
t\_sd <- sqrt(var(my\_mean))  
#var is var(my\_mean) = 0.0002818953  
t\_stat <- t\_mean/(t\_sd/sqrt(numN))  
plot(my\_mean,type= "l")



# it looks like a distribution from the plot  
skewness(my\_mean) #-0.1605593

## [1] -0.3708221

kurtosis(my\_mean)#-0.5593615

## [1] -0.3546094

jarque.bera.test(my\_mean)

##   
## Jarque Bera Test  
##   
## data: my\_mean  
## X-squared = 2.7392, df = 2, p-value = 0.2542

#p-value = 0.4663 so it could be a normal distribution  
mean(my\_error)

## [1] 0.01673394

#the mean of the standard errors are 0.01727482 which is slightly higherer than the standard deviation

t\_stat is 3.8055 which is less than 1.96 so we reject the null hypothesis

prob 2:

DBV <- read\_excel("/Users/Lisa/Documents/Econometrics/ass1/DBV.xlsx")  
names(DBV)= c("Date", "Open","High"," Low","Close","Volume","Adj")  
attach(DBV)  
dbv\_ts<-ts(DBV,start = 1, frequency = 1)  
# Growth rates:  
log.dbv <- log(dbv\_ts)  
r\_dbv <- diff(log.dbv)  
head(log.dbv)

## Time Series:  
## Start = 1   
## End = 6   
## Frequency = 1   
## Date Open High \tLow Close Volume Adj  
## 1 20.87095 3.216874 3.218876 3.214868 3.218876 9.517825 3.174079  
## 2 20.87102 3.222071 3.223266 3.218876 3.220075 11.203679 3.175278  
## 3 20.87110 3.216874 3.217675 3.214466 3.216874 9.775654 3.172077  
## 4 20.87117 3.217675 3.218476 3.216072 3.218076 12.333146 3.173279  
## 5 20.87124 3.219276 3.220075 3.218476 3.218876 12.899220 3.174079  
## 6 20.87147 3.218876 3.220075 3.216874 3.219276 9.622450 3.174479

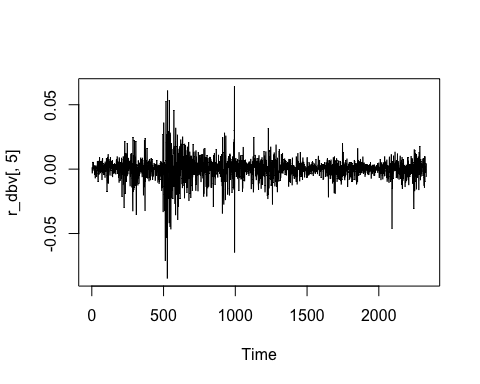
head(r\_dbv)

## Time Series:  
## Start = 2   
## End = 7   
## Frequency = 1   
## Date Open High \tLow Close  
## 2 7.453509e-05 0.005196853 0.0043903881 0.004008021 0.0011993205  
## 3 7.452953e-05 -0.005196853 -0.0055911488 -0.004409749 -0.0032012831  
## 4 7.452398e-05 0.000801202 0.0008006806 0.001605821 0.0012016424  
## 5 7.451842e-05 0.001600681 0.0015994005 0.002403847 0.0008003202  
## 6 2.235220e-04 -0.000399920 0.0000000000 -0.001601883 0.0003999200  
## 7 7.449622e-05 0.001598761 0.0003994408 0.001601883 -0.0003999200  
## Volume Adj  
## 2 1.6858541 0.0011993327  
## 3 -1.4280250 -0.0032012688  
## 4 2.5574915 0.0012016090  
## 5 0.5660741 0.0008003271  
## 6 -3.2767698 0.0003999653  
## 7 0.2709872 -0.0003999653

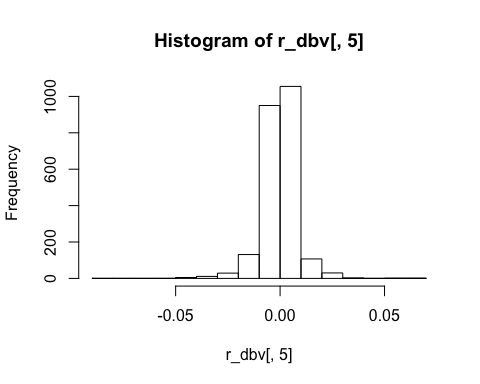
dim(r\_dbv)

## [1] 2330 7

plot(r\_dbv[,5])



hist(r\_dbv[,5])



my\_T<- 2330  
my\_skew<-skewness(r\_dbv[,5])  
skew\_t <- my\_skew/sqrt(6/my\_T)  
#skew\_t is -16.9339 whose absolute value is greater than 1.96, so we reject the null hypothesis  
  
my\_kurt <-kurtosis(r\_dbv[,5])  
kurt\_t <-(my\_kurt)/sqrt(24/my\_T)  
#kurt\_t is 131.615 whose absolute value is greater than 1.96, so we reject the null hypothesis  
jarque.bera.test(r\_dbv[,5])

##   
## Jarque Bera Test  
##   
## data: r\_dbv[, 5]  
## X-squared = 17646, df = 2, p-value < 2.2e-16

#p-value < 2.2e-16 so reject the null

## Including Plots

You can also embed plots, for example:

GSPC <- read\_excel("/Users/Lisa/Documents/Econometrics/ass1/GSPC.xlsx")  
names(GSPC)= c("Date", "Open","High"," Low","Close","Volume","Adj")  
attach(GSPC)

## The following objects are masked from DBV:  
##   
## Low, Adj, Close, Date, High, Open, Volume

GSPC\_ts<-ts(GSPC,start = 1, frequency = 1)  
# Growth rates:  
log.GSPC <- log(GSPC\_ts)  
r\_GSPC <- diff(log.GSPC)  
head(log.GSPC)

## Time Series:  
## Start = 1   
## End = 6   
## Frequency = 1   
## Date Open High \tLow Close Volume Adj  
## 1 20.87095 7.181425 7.192445 7.178988 7.190201 21.72030 7.190201  
## 2 20.87102 7.190186 7.197884 7.189394 7.197697 21.70660 7.197697  
## 3 20.87110 7.197525 7.200485 7.195592 7.197877 21.73457 7.197877  
## 4 20.87117 7.197854 7.200634 7.195750 7.199589 21.59783 7.199589  
## 5 20.87124 7.199790 7.200335 7.197166 7.197323 21.54456 7.197323  
## 6 20.87147 7.197301 7.199335 7.193145 7.193926 21.49082 7.193926

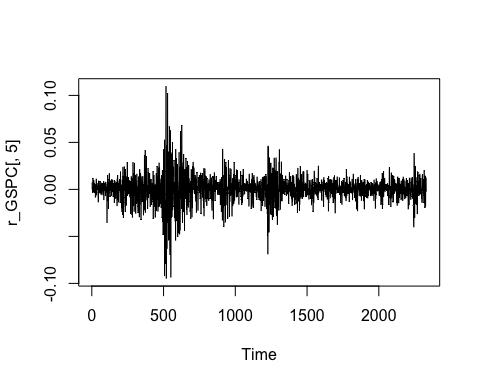
head(r\_GSPC)

## Time Series:  
## Start = 2   
## End = 7   
## Frequency = 1   
## Date Open High \tLow Close  
## 2 7.453509e-05 0.0087614173 0.0054389752 0.0104064021 0.0074961151  
## 3 7.452953e-05 0.0073390971 0.0026002227 0.0061982038 0.0001795701  
## 4 7.452398e-05 0.0003293054 0.0001492882 0.0001574339 0.0017118783  
## 5 7.451842e-05 0.0019359091 -0.0002985076 0.0014160653 -0.0022656719  
## 6 2.235220e-04 -0.0024898072 -0.0010005646 -0.0040211208 -0.0033968847  
## 7 7.449622e-05 -0.0033744044 -0.0001718288 -0.0023933758 0.0020935009  
## Volume Adj  
## 2 -0.01370482 0.0074961151  
## 3 0.02797396 0.0001795701  
## 4 -0.13674633 0.0017118783  
## 5 -0.05327029 -0.0022656719  
## 6 -0.05374031 -0.0033968847  
## 7 0.21927063 0.0020935009

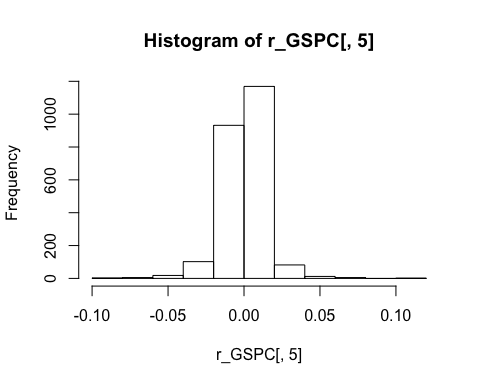
dim(r\_GSPC)

## [1] 2330 7

plot(r\_GSPC[,5])



hist(r\_GSPC[,5])



my\_T<- 2330  
my\_skew2<-skewness(r\_GSPC[,5])  
skew\_t2 <- my\_skew2/sqrt(6/my\_T)  
#skew\_t2 is -6.38153 whose absolute value is greater than 1.96, so we reject the null hypothesis  
  
my\_kurt2 <-kurtosis(r\_GSPC[,5])  
kurt\_t2 <-(my\_kurt2)/sqrt(24/my\_T)  
#kurt\_t2 is 96.4317 whose absolute value is greater than 1.96, so we reject the null hypothesis  
  
jarque.bera.test(r\_GSPC[,5])

##   
## Jarque Bera Test  
##   
## data: r\_GSPC[, 5]  
## X-squared = 9360.7, df = 2, p-value < 2.2e-16

#p-value < 2.2e-16 so reject the null

logrDBV <-r\_dbv[,5]  
logrGSPC <-r\_GSPC[,5]  
my\_model<-lm(logrDBV~logrGSPC)  
summary(my\_model)

##   
## Call:  
## lm(formula = logrDBV ~ logrGSPC)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.069545 -0.003080 0.000239 0.003311 0.058668   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.0001153 0.0001367 -0.843 0.399   
## logrGSPC 0.4323363 0.0101522 42.585 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.006598 on 2328 degrees of freedom  
## Multiple R-squared: 0.4379, Adjusted R-squared: 0.4376   
## F-statistic: 1814 on 1 and 2328 DF, p-value: < 2.2e-16

g\_model <-glm(logrDBV~logrGSPC)  
summary(g\_model)

##   
## Call:  
## glm(formula = logrDBV ~ logrGSPC)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.069545 -0.003080 0.000239 0.003311 0.058668   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.0001153 0.0001367 -0.843 0.399   
## logrGSPC 0.4323363 0.0101522 42.585 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 4.3538e-05)  
##   
## Null deviance: 0.18031 on 2329 degrees of freedom  
## Residual deviance: 0.10136 on 2328 degrees of freedom  
## AIC: -16781  
##   
## Number of Fisher Scoring iterations: 2

#rob\_model <- lmrob(logrDBV~logrGSPC)  
#summary(rob\_model)  
my\_model %>%   
 vcovHC() %>%   
 diag() %>%   
 sqrt()

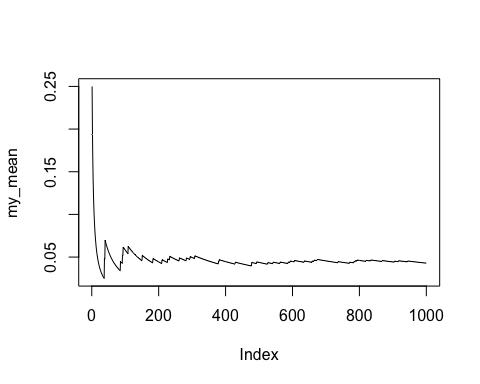
## (Intercept) logrGSPC   
## 0.0001375697 0.0182961682

Under OLS assumptions, the standard error of the intercept is 0.0001367 and that of the slope is 0.0101522. Allowing for non-normalities, the standard error of the intercept is 0.0001367 and that of the slope is 0.0101522. the heteroskedastic standard error of the intercept is 0.0001375697 and that of the slope is 0.0182961682.

The White standard errors are larger than the classic OLS standard errors because the homoskedasticity-only standard errors are only valid of the errors are homoskedastic and that the White standard errors are valid whether or not the errors are heteroskedastic. Therefore, the White standard errors are more general and more conservative.

Prob 3:

N<-1000  
a<-1  
A<-3  
pois <- rpois(N, 0.05)  
sumpois <- cumsum(pois)  
my\_mean <- rep(0, N)  
  
for (i in 1:N){  
 my\_mean[i] <- (a+sumpois[i])/(A+i)  
   
}  
   
  
plot(my\_mean, type = "l")



It doesn't look odd to me. The mean start off close to the mean i set which is 1/A = 1/3 and then it jumps up and slowly goes down and eventually converges to lambda = 0.05