**BANA 7031(001) Probability Models:**

**Project**

**Abstract**: The data under analysis consists of weights of stock-picking concepts and performances of portfolios gathered from historical US stock market data. The data covers 20 years, from September 1990 till June 2010, and is further divided into four 5-year periods: 09/1990-06/1995, ..., 09/2005-06/2010. The portfolio under observation consists of 63 S&P500 stocks.

The project involves application of classroom concepts in analyzing the portfolio performance as measured by the annual returns, excess returns and absolute winning rates across time periods and as compared with industry standards.

**Authors**:

Anupam Shukla M13469377)

**Date:** December 2019

**Executive Summary:**

The share prices in the stock market fluctuate daily due to continuous buying and selling. Stock prices move in trends and cycles and are never stable. A trader in the stock market is interested in winning most trades. To be able to do that successfully, an analysis of the past performance of the stock is imperative.

The purpose of this study is to analyze the performance of a portfolio of 63 S&P500 stocks in 5-year periods between 1990 and 2010. The study uses the concepts of empirical distributions, parametric distributions, bootstrapping, hypothesis testing and Bayesian analysis to gather insights into the given stock data. A key metric that has been studied throughout for analysis is the annual returns of stocks in different time periods. The analysis also involves studying the average excess returns and the mean absolute winning rates for the same portfolio across different time periods.

Interesting observations arise upon the completion of the study. A relative comparison of the empirical CDFs indicates that the median of annual returns was the least in the period 2005-2010 and maximum in the period 1990-1995. Further, the period 2000- 2005 saw wide fluctuations in the annual returns varying from -0.08% to 30%. The median annual return for the portfolio over the combined 20-year period (1990-2010) was 15%, which is a very good return.

Conducting hypothesis testing using Wilcoxon test for the annual returns in period 2000-2005, it was concluded that the median return in this period was higher than in the entire period. Using paired t-tests, it was concluded that average annual return is greater in period 1990-1995 as compared to the entire period. Other hypothesis tests revealed that the average excess returns were greater than 0 in all periods except in the period 1995-2000. It was also observed that the mean absolute winning rates were greater than 0.7 only in the 5-year period from 1990 to 1995. (Absolute winning rate greater than 0.7 is considered a good winning percentage for a trader).

Thus, based on the hypothesis testing, the period 1990-1995 saw average excess returns greater than 0, mean absolute winning rate greater than 0.7 and the average annual return greater than the combined 20- year duration.

Bayesian analysis was applied to estimate the posterior density for μ (mean of the normally distributed annual returns) with different priors on μ. It was observed that while a uniform prior, and a normal prior led to a well-defined normally distributed posterior distribution, taking a non-uniform discrete prior did not produce a well-defined distribution.

**Empirical CDF**

The analysis focusses on the performance of the stocks across time. In order to understand the data better, it was required to first estimate the distribution of the random variable. Thus, the following plot illustrates the Empirical CDF derived to estimate and compare the cumulative distribution of Annual Returns for the following periods (the below definition of periods is used in the entire report) –

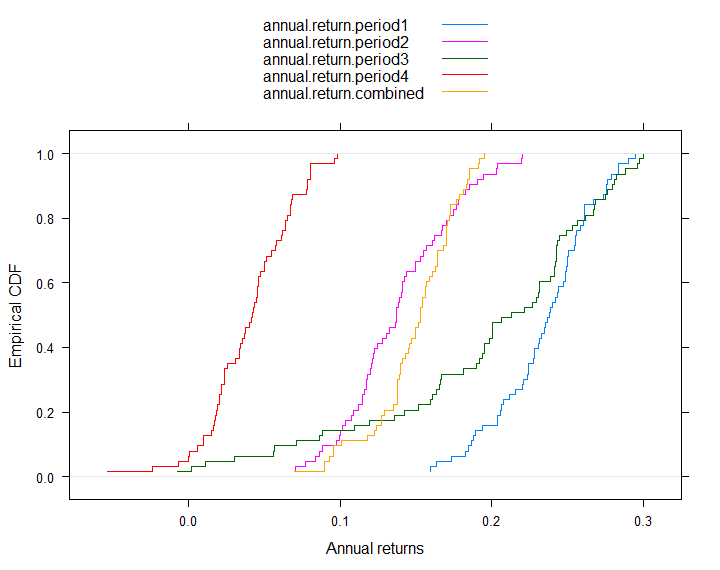
Period1 = 09/1990-06/1995

Period2 = 09/1995-06/2000

Period3 = 09/2000-06/2005

Period4 = 09/2005-06/2010

Combined = 09/1990-06/2010



**Figure 1: ECDF plots for annual returns in different time periods**

From the above ecdf plot, it can be inferred that -

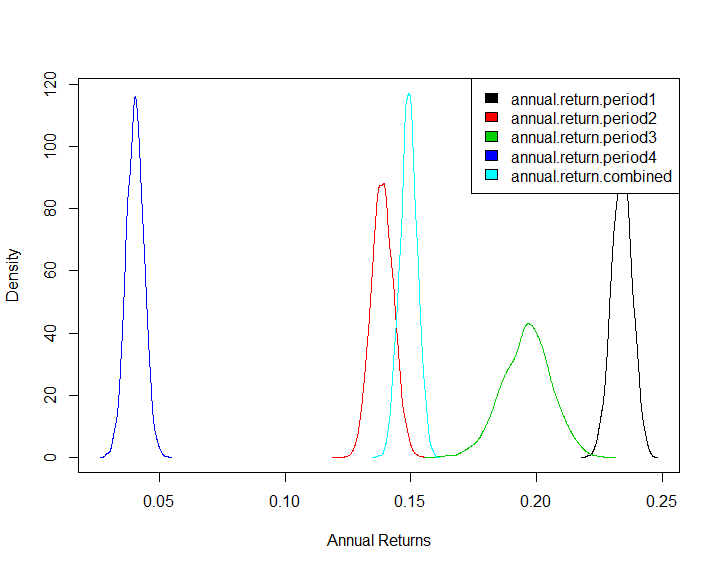
a. The median return was maximum in period 1. It dropped in period 2 and then rose in period3. It again drops in period 4 and is the least in this period.

b. Almost 4% stocks in period 4 gave negative returns.

c. Stocks in period 3 have a wide range of returns from -0.08% to almost 30%.

**Non-Parametric Bootstrapping**

The next step in the analysis was to calculate standard errors of sampling distributions and the confidence interval for the estimated parameters in order to make inferences. Let the statistic be the expected annual return of the stocks. Non-parametric bootstrapping was conducted to estimate the standard error and confidence interval of the sampling distribution of the statistic.



**Figure 2: Sampling distribution of Expected Annual Returns in different time periods**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Period** | **Expected Return** | **Standard error** | **Normal CI** | **Pivotal CI** | **Percentile CI** |
| Period1 | 0.234 | 0.004 | (0.2259,0.242) | (0.2262,0.2421) | (0.2258,0.2417) |
| Period2 | 0.1386 | 0.0044 | (0.1298,0.1474) | (0.1296,0.1496) | (0.1303,0.1475) |
| Period3 | 0.196 | 0.0097 | (0.1767,0.2154) | (0.1774,0.2159) | (0.1762,0.2147) |
| Period4 | 0.0404 | 0.0035 | (0.0334,0.0474) | (0.0337,0.0473) | (0.0335,0.0470) |
| Combined period | 0.1492 | 0.0034 | (0.1423,0.1561) | (0.1423,0.1556) | (0.1428,0.1561) |

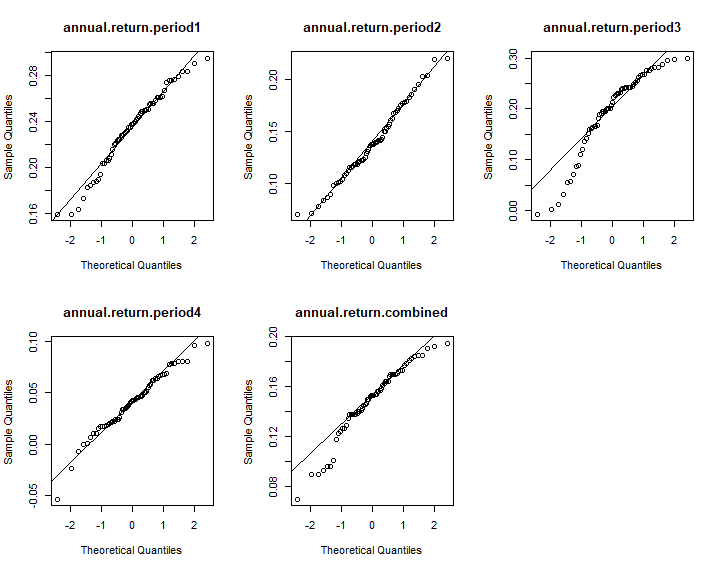
**Table 1: Non-parametric bootstrap results on annual returns**

**Observations**:

1. The expected returns are the highest in period 1 while lowest in period 4. The expected return over the 20-year (combined) period is almost 15%.
2. The confidence intervals in each period are similar for normal, pivotal and percentile methods. This indicates that the distributions can be assumed to be normal (which will also be established in later sections of this report). Figure 2 supports this assumption.
3. The maximum variation in the expected returns is observed in period 3.

**MLE and its asymptotic distribution**

Maximum Likelihood Estimation can be applied for parametric distributions. To apply this method, it was first required to understand the underlying distribution. QQ plots and Shapiro-Wilk test results were assessed to determine the normality of the annual returns data in all the periods.



**Figure 3: QQ plots for annual returns**

Using Shapiro-Wilk test, the following p-value for the respective periods were obtained:

annual.period.1 = 0.1654

annual.period.2 = 0.5709

annual.period.3 = 0.0001

annual.period.4 = 0.2809

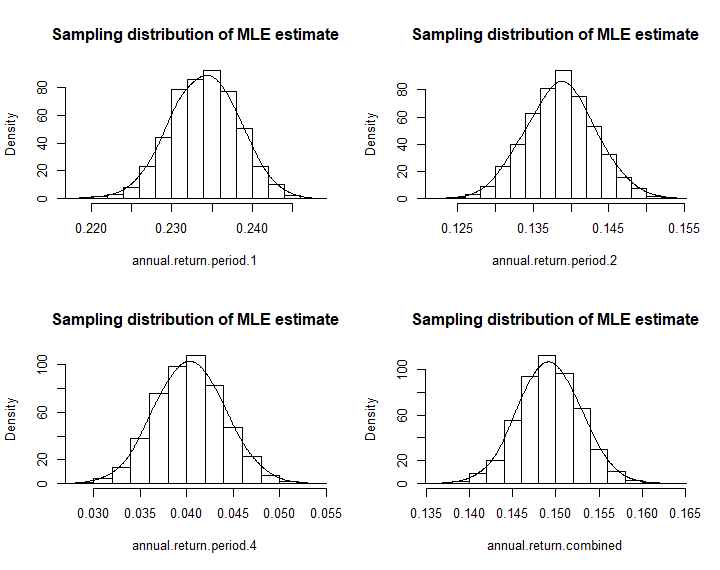
annual.period.5 = 0.01

Both the QQ plots and Shapiro-Wilk tests establish that for annual returns in all periods follow normal distributions except in period 3.

MLE is calculated as per the formula used below:

and

In order to establish the property of asymptotic normality, the sampling distribution of MLE estimates using parametric bootstrapping were plotted to obtain the below graph:



**Figure 4: Sampling Distribution of MLE estimate for expected Annual Return**

Further, the MLE estimates and the corresponding confidence intervals obtained using parametric bootstrapping are givenasbelow:

|  |  |  |  |
| --- | --- | --- | --- |
| **Period** | **MLE estimate** | **Standard error** | **Bootstrap CI** |
| Period1 | 0.2340 | 0.004 | (0.2258,0.2422) |
| Period2 | 0.1386 | 0.0044 | (0.1297,0.1475) |
| Period4 | 0.0404 | 0.0036 | (0.0331,0.0476) |
| Combined period | 0.1492 | 0.0035 | (0.1422,0.1563) |

**Table 2: Parametric bootstrap results on annual returns**

**Hypothesis Testing**

1. Comparison of median annual returns across period 3 and combined period using **Wilcoxon** signed rank test:

H0 = Median annual return is greater in period 3 compared to combined period

Ha = Median annual return is less in period 3 compared to combined period

Wilcoxon test was used to test the null hypothesis as period 3 data was observed to be non-normal.

The test returned V = 1753 with p-value = 1 which indicates insufficient evidence to reject the null hypothesis that the median annual returns in period 3 is greater than median annual returns in the combined period.

2. Comparison of average annual returns across periods all periods with combined period using paired **t-test:**

H0 = Average annual return is less in period 1 compared to combined period

Ha = Average annual return is greater in period 1 compared to combined period

This test was repeated for periods 2, 3 and 4.

**Test Results**:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Period** | **t-statistic** | **p-value** | **Reject H0** | **Conclusion** |
| Period1 | 36.713 | < 2.2e-16 | Yes | Average annual return is greater in period 1 compared to combined period |
| Period2 | -1.7844 | 0.9604 | No | Average annual return is lesser in period 2 compared to combined period |
| Period3 | 6.7355 | 3.078e-09 | Yes | Average annual return is greater in period 1 compared to combined period |
| Period4 | -39.362 | 1 | No | Average annual return is lesser in period 4 compared to combined period |

3. Test whether average excess returns are greater than 0 in any period using **t-test**

H0 = Average excess return is less than 0 in a period

Ha = Average excess return is greater than 0 in a period

**Test results:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Period** | **t-statistic** | **p-value** | **Reject H0** | **Conclusion** |
| Period1 | 9.9866 | 7.859e-15 | Yes | Average excess return is greater than 0 in period1 |
| Period2 | -17.25 | 1 | No | Average excess return is less than 0 in period2 |
| Period3 | 23.454 | < 2.2e-16 | Yes | Average excess return is greater than 0 in period3 |
| Period4 | 10.152 | 4.156e-15 | Yes | Average excess return is greater than 0 in period4 |
| Combined | 16.484 | < 2.2e-16 | Yes | Average excess return is greater than 0 in combined period |

4. Test whether absolute win rate is more than 0.7 in any period using **t-test** (an absolute winning rate of more than 0.7 is considered a good winning percentage for a trader)

H0 = Average absolute winning rate is less than 0.7 in a period

Ha = Average absolute winning rate is greater than 0.7 in a period

**Test results:**

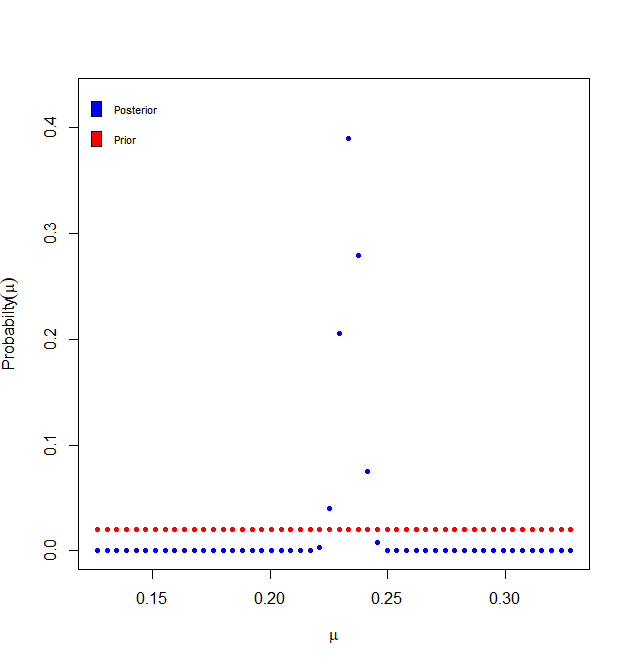
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Period** | **t-statistic** | **p-value** | **Reject H0** | **Conclusion** |
| Period1 | 13.726 | < 2.2e-16 | Yes | Average absolute winning rate is greater than 0.7 in period1 |
| Period2 | -7.5286 | 1 | No | Average absolute winning rate is less than 0.7 in period2 |
| Period3 | -0.2099 | 0.5828 | No | Average absolute winning rate is less than 0.7 in period3 |
| Period4 | -22.039 | 1 | No | Average absolute winning rate is less than 0.7 in period 4 |
| Combined | -5.0974 | 1 | No | Average absolute winning rate is less than 0.7 in combined period |

**Bayesian Analysis**

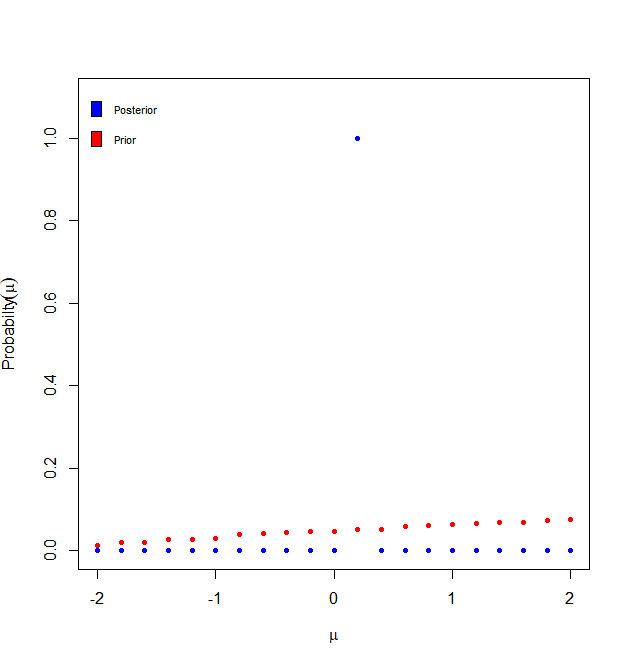
Let the unknown parameter be the Mean of Annual Return in period 1 such that Annual Returns are normally distributed.

**Objective**: To estimate the posterior density for μ (mean of a normal distribution) with a discrete prior on μ

Case 1: find the posterior density with a uniform prior on mu



Case 2: find the posterior density with a non- uniform prior on mu

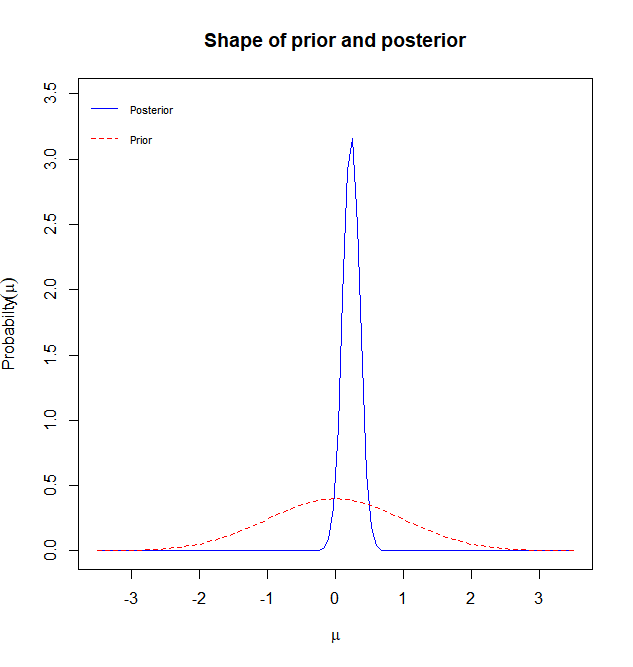


When a uniform or a flat prior is assumed, the resulting posterior is approximately normal. The posterior distribution is equivalent to the frequentist approach in this instance and is given as ~N (.

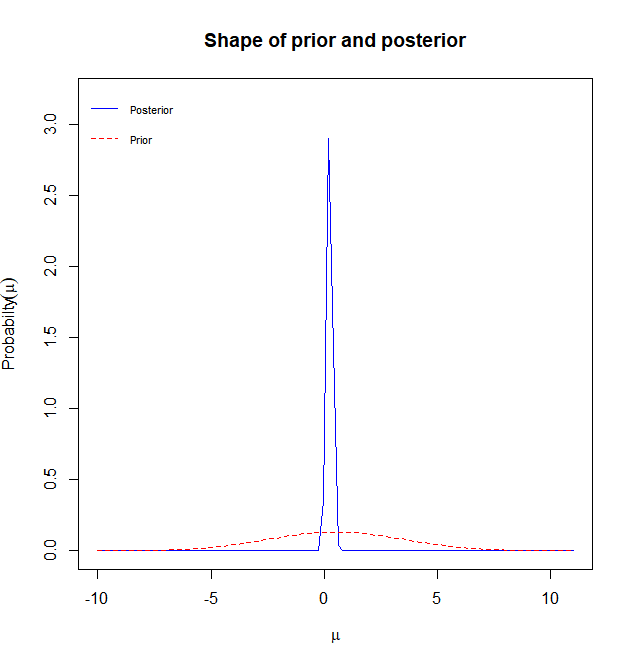
When a non-uniform prior is considered, the posterior distribution is not well defined. Hence, it is not desirable to move ahead with this priori belief.

**Objective**: To estimate the posterior density for μ (mean of a normal distribution) with a normal prior on μ

Case 3: find the posterior density with a N(0,1) prior on mu

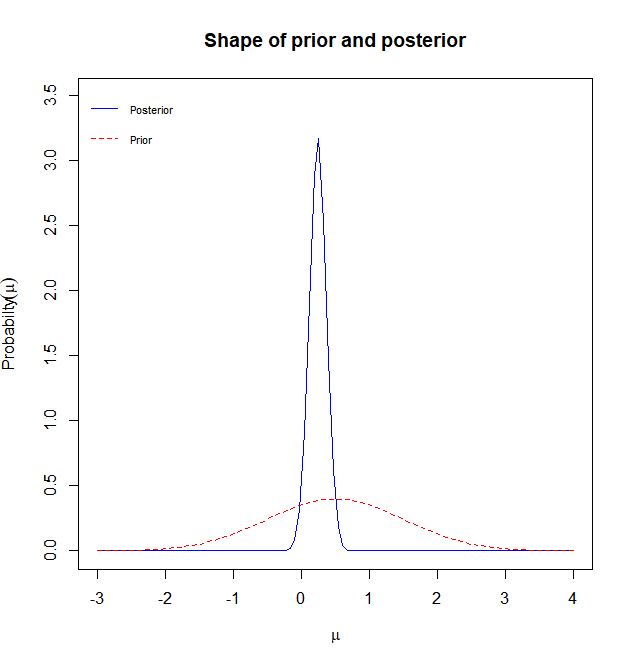
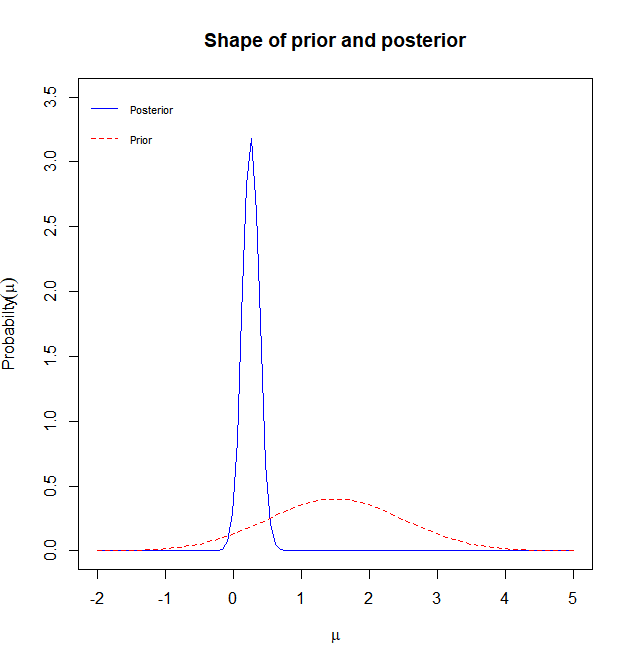


Case 4: find the posterior density with a N(0.5,3) prior on mu

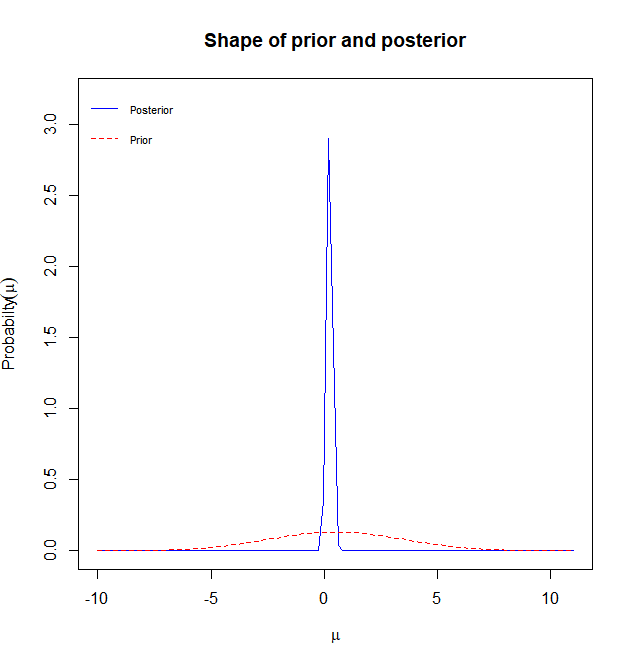
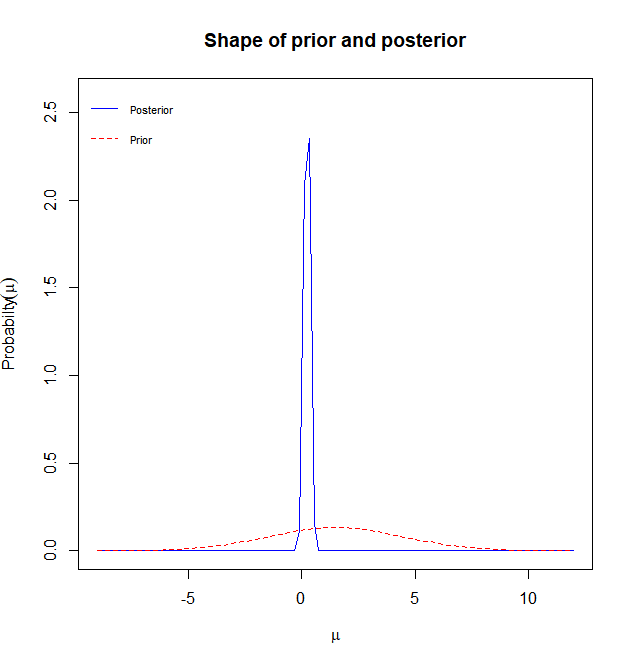


Further, it was observed that varying the standard deviation or the mean of the prior distribution did not produce any significant difference in the posterior distribution.

N (0.5,1) N (1.5,1)

N (0.5,3) N (1.5,3)

The above analysis was repeated for other periods and similar results were observed.

**References:** <https://archive.ics.uci.edu/ml/datasets/Stock+portfolio+performance>

**Appendix**

Data



Code

#import excel

library(readxl)

#stock\_performance\_pd1 <- read\_excel("stock\_performance.xlsx", sheet = "1st period")

#stock\_performance\_pd2 <- read\_excel("stock\_performance.xlsx", sheet = "2nd period")

#stock\_performance\_pd3 <- read\_excel("stock\_performance.xlsx", sheet = "3rd period")

#stock\_performance\_pd4 <- read\_excel("stock\_performance.xlsx", sheet = "4th period")

#stock\_performance\_all <- read\_excel("stock\_performance.xlsx", sheet = "all period")

# Remove space from column names

library(stringr)

names(stock\_performance\_pd1)<-str\_replace\_all(names(stock\_performance\_pd1), c(" " = "." , "," = "" ))

names(stock\_performance\_pd2)<-str\_replace\_all(names(stock\_performance\_pd2), c(" " = "." , "," = "" ))

names(stock\_performance\_pd3)<-str\_replace\_all(names(stock\_performance\_pd3), c(" " = "." , "," = "" ))

names(stock\_performance\_pd4)<-str\_replace\_all(names(stock\_performance\_pd4), c(" " = "." , "," = "" ))

names(stock\_performance\_all)<-str\_replace\_all(names(stock\_performance\_all), c(" " = "." , "," = "" ))

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# 1. ECDF of the random variable: 'Annual return' for all periods

library(lattice)

install.packages('latticeExtra')

library(latticeExtra)

set.seed(42)

vals <- data.frame(annual.return.period1 = stock\_performance\_pd1$Annual.Return,

annual.return.period2 = stock\_performance\_pd2$Annual.Return,

annual.return.period3 = stock\_performance\_pd3$Annual.Return,

annual.return.period4 = stock\_performance\_pd4$Annual.Return,

annual.return.combined = stock\_performance\_all$Annual.Return)

ecdfplot(~ annual.return.period1 + annual.return.period2 + annual.return.period3

+ annual.return.period4 + annual.return.combined, data=vals,

auto.key=list(space='top'),xlab = 'Annual returns')

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# 2.a Non-Parametric Bootstrapping for standard errors and CI of statistic: Annual return

install.packages("bootstrap")

library(bootstrap)

# plug-in estimator

mu.hat.pd1<-mean(stock\_performance\_pd1$Annual.Return)

mu.hat.pd2<-mean(stock\_performance\_pd2$Annual.Return)

mu.hat.pd3<-mean(stock\_performance\_pd3$Annual.Return)

mu.hat.pd4<-mean(stock\_performance\_pd4$Annual.Return)

mu.hat.combined<-mean(stock\_performance\_all$Annual.Return)

theta<-function(x){

mean(x)

}

results.pd1 <- bootstrap(stock\_performance\_pd1$Annual.Return,3200,theta)

results.pd2 <- bootstrap(stock\_performance\_pd2$Annual.Return,3200,theta)

results.pd3 <- bootstrap(stock\_performance\_pd3$Annual.Return,3200,theta)

results.pd4 <- bootstrap(stock\_performance\_pd4$Annual.Return,3200,theta)

results.combined <- bootstrap(stock\_performance\_all$Annual.Return,3200,theta)

#standard error of sampling statistic

se.boot.pd1=sqrt(var(results.pd1$thetastar))

se.boot.pd2=sqrt(var(results.pd2$thetastar))

se.boot.pd3=sqrt(var(results.pd3$thetastar))

se.boot.pd4=sqrt(var(results.pd4$thetastar))

se.boot.combined=sqrt(var(results.combined$thetastar))

# Bootstrap confidence intervals

#period1

normal.ci.pd1<-c(mu.hat.pd1-2\*se.boot.pd1, mu.hat.pd1+2\*se.boot.pd1)

pivatol.ci.pd1<-c(2\*mu.hat.pd1-quantile(results.pd1$thetastar,0.975), 2\*mu.hat.pd1-quantile(results.pd1$thetastar,0.025))

quantile.ci.pd1<-quantile(results.pd1$thetastar, c(0.025, 0.975))

normal.ci.pd1;pivatol.ci.pd1;quantile.ci.pd1

#period2

normal.ci.pd2<-c(mu.hat.pd2-2\*se.boot.pd2, mu.hat.pd2+2\*se.boot.pd2)

pivatol.ci.pd2<-c(2\*mu.hat.pd2-quantile(results.pd2$thetastar,0.975), 2\*mu.hat.pd2-quantile(results.pd2$thetastar,0.025))

quantile.ci.pd2<-quantile(results.pd2$thetastar, c(0.025, 0.975))

normal.ci.pd2;pivatol.ci.pd2;quantile.ci.pd2

#period3

normal.ci.pd3<-c(mu.hat.pd3-2\*se.boot.pd3, mu.hat.pd3+2\*se.boot.pd3)

pivatol.ci.pd3<-c(2\*mu.hat.pd3-quantile(results.pd3$thetastar,0.975), 2\*mu.hat.pd3-quantile(results.pd3$thetastar,0.025))

quantile.ci.pd3<-quantile(results.pd3$thetastar, c(0.025, 0.975))

normal.ci.pd3;pivatol.ci.pd3;quantile.ci.pd3

#period4

normal.ci.pd4<-c(mu.hat.pd4-2\*se.boot.pd4, mu.hat.pd4+2\*se.boot.pd4)

pivatol.ci.pd4<-c(2\*mu.hat.pd4-quantile(results.pd4$thetastar,0.975), 2\*mu.hat.pd4-quantile(results.pd4$thetastar,0.025))

quantile.ci.pd4<-quantile(results.pd4$thetastar, c(0.025, 0.975))

normal.ci.pd4;pivatol.ci.pd4;quantile.ci.pd4

#combined period

normal.ci.combined<-c(mu.hat.combined-2\*se.boot.combined, mu.hat.combined+2\*se.boot.combined)

pivatol.ci.combined<-c(2\*mu.hat.combined-quantile(results.combined$thetastar,0.975), 2\*mu.hat.combined-quantile(results.combined$thetastar,0.025))

quantile.ci.combined<-quantile(results.combined$thetastar, c(0.025, 0.975))

normal.ci.combined;pivatol.ci.combined;quantile.ci.combined

#density plots

myData <- data.frame(results.pd1$thetastar,results.pd2$thetastar,results.pd3$thetastar,

results.pd4$thetastar,results.combined$thetastar)

dens <- apply(myData, 2, density)

plot(NA, xlim=range(sapply(dens, "[", "x")), ylim=range(sapply(dens, "[", "y")),xlab = 'Annual Returns',ylab = 'Density')

mapply(lines, dens, col=1:length(dens))

i <- c('annual.return.period1','annual.return.period2','annual.return.period3','annual.return.period4','annual.return.combined')

legend("topright",legend=i, fill=1:length(i))

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# 3. MLE point estimation and asymptotic distributions

#identify the distribution of random variables: 'Annual return'

par(mfrow=c(2,3))

qqnorm(stock\_performance\_pd1$Annual.Return,main = 'annual.return.period1')

qqline(stock\_performance\_pd1$Annual.Return)

qqnorm(stock\_performance\_pd2$Annual.Return,main = 'annual.return.period2')

qqline(stock\_performance\_pd2$Annual.Return)

qqnorm(stock\_performance\_pd3$Annual.Return,main = 'annual.return.period3')

qqline(stock\_performance\_pd3$Annual.Return)

qqnorm(stock\_performance\_pd4$Annual.Return,main = 'annual.return.period4')

qqline(stock\_performance\_pd4$Annual.Return)

qqnorm(stock\_performance\_all$Annual.Return,main = 'annual.return.combined')

qqline(stock\_performance\_all$Annual.Return)

#shapiro.test for normal

shapiro.test(stock\_performance\_pd1$Annual.Return)

shapiro.test(stock\_performance\_pd2$Annual.Return)

shapiro.test(stock\_performance\_pd3$Annual.Return)

shapiro.test(stock\_performance\_pd4$Annual.Return)

shapiro.test(stock\_performance\_all$Annual.Return)

# The QQ plot and MLE estimate indicates that annual return in all the periods except period 3 is

# approximately normally distributed, MLE is calculated as per the formula used below:

n <- length(stock\_performance\_pd1$Annual.Return)

mu\_mle.pd1<-mean(stock\_performance\_pd1$Annual.Return)

mu\_mle.pd2<-mean(stock\_performance\_pd2$Annual.Return)

mu\_mle.pd4<-mean(stock\_performance\_pd4$Annual.Return)

mu\_mle.combined<-mean(stock\_performance\_all$Annual.Return)

sigma\_hat.pd1<-sqrt((1/n)\*sum((stock\_performance\_pd1$Annual.Return-mu\_mle.pd1)^2))

sigma\_hat.pd2<-sqrt((1/n)\*sum((stock\_performance\_pd2$Annual.Return-mu\_mle.pd2)^2))

sigma\_hat.pd3<-sqrt((1/n)\*sum((stock\_performance\_pd3$Annual.Return-mu\_mle.pd3)^2))

sigma\_hat.pd4<-sqrt((1/n)\*sum((stock\_performance\_pd4$Annual.Return-mu\_mle.pd4)^2))

sigma\_hat.combined<-sqrt((1/n)\*sum((stock\_performance\_all$Annual.Return-mu\_mle.combined)^2))

#Parametric bootstrapping using MLE estimates

pboot <- function (mu,sigma){

theta.hat\_Pbootstrap = vector()

for(i in 1:3000){

X\_i=rnorm(n,mu,sigma)

theta.hat\_Pbootstrap[i] = mean(X\_i)

}

theta.hat\_Pbootstrap

}

theta.hat.period1=pboot(mu\_mle.pd1,sigma\_hat.pd1)

theta.hat.period2=pboot(mu\_mle.pd2,sigma\_hat.pd2)

theta.hat.period3=pboot(mu\_mle.pd3,sigma\_hat.pd3)

theta.hat.period4=pboot(mu\_mle.pd4,sigma\_hat.pd4)

theta.hat.combined=pboot(mu\_mle.combined,sigma\_hat.combined)

theta.hat.period1.se=sd(theta.hat.period1)

theta.hat.period2.se=sd(theta.hat.period2)

theta.hat.period3.se=sd(theta.hat.period3)

theta.hat.period4.se=sd(theta.hat.period4)

theta.hat.combined.se=sd(theta.hat.combined)

Pbootstrap\_CI.period1<-c(mu\_mle.pd1-2\*theta.hat.period1.se,mu\_mle.pd1+2\*theta.hat.period1.se)

Pbootstrap\_CI.period2<-c(mu\_mle.pd2-2\*theta.hat.period2.se,mu\_mle.pd2+2\*theta.hat.period2.se)

Pbootstrap\_CI.period3<-c(mu\_mle.pd3-2\*theta.hat.period3.se,mu\_mle.pd3+2\*theta.hat.period3.se)

Pbootstrap\_CI.period4<-c(mu\_mle.pd4-2\*theta.hat.period4.se,mu\_mle.pd4+2\*theta.hat.period4.se)

Pbootstrap\_CI.combined<-c(mu\_mle.combined-2\*theta.hat.combined.se,mu\_mle.combined+2\*theta.hat.combined.se)

mu\_mle.pd1;mu\_mle.pd2;mu\_mle.pd3;mu\_mle.pd4;mu\_mle.combined

theta.hat.period1.se;theta.hat.period2.se;theta.hat.period3.se;theta.hat.period4.se;theta.hat.combined.se

Pbootstrap\_CI.period1;Pbootstrap\_CI.period2;Pbootstrap\_CI.period3;Pbootstrap\_CI.period4;Pbootstrap\_CI.combined

#asymptotic normality

par(mfrow = c(2,2))

hist(theta.hat.period1,xlab="annual.return.period.1",main="Sampling distribution of MLE estimate",probability = T)

lines(density(theta.hat.period1, adjust=2))

hist(theta.hat.period2,xlab="annual.return.period.2",main="Sampling distribution of MLE estimate",probability = T)

lines(density(theta.hat.period2, adjust=2))

hist(theta.hat.period4,xlab="annual.return.period.4",main="Sampling distribution of MLE estimate",probability = T)

lines(density(theta.hat.period4, adjust=2))

hist(theta.hat.combined,xlab="annual.return.combined",main="Sampling distribution of MLE estimate",probability = T)

lines(density(theta.hat.combined, adjust=2))

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# 5. Hypothesis Testing

annual.return.period1 <- stock\_performance\_pd1$Annual.Return

annual.return.period2 <- stock\_performance\_pd2$Annual.Return

annual.return.period3 <- stock\_performance\_pd3$Annual.Return

annual.return.period4 <- stock\_performance\_pd4$Annual.Return

annual.return.combined <- stock\_performance\_all$Annual.Return

#wilcoxon test for comparing median between period 3 and combined period

wilcox.test(annual.return.period3,annual.return.combined,paired = TRUE, exact = F,alternative = "less")

#paired t-tests for comapring average annual returns

t.test(stock\_performance\_pd1$Annual.Return,stock\_performance\_all$Annual.Return, paired = TRUE, alternative = "greater",conf.level = 0.95)

t.test(stock\_performance\_pd2$Annual.Return,stock\_performance\_all$Annual.Return, paired = TRUE, alternative = "greater",conf.level = 0.95)

t.test(stock\_performance\_pd3$Annual.Return,stock\_performance\_all$Annual.Return, paired = TRUE, alternative = "greater",conf.level = 0.95)

t.test(stock\_performance\_pd4$Annual.Return,stock\_performance\_all$Annual.Return, paired = TRUE, alternative = "greater",conf.level = 0.95)

# t-test for testing whether average excess returns are greater than 0

t.test(x = stock\_performance\_pd1$Excess.Return, alternative = "greater", mu = 0,

conf.level = 0.95)

t.test(x = stock\_performance\_pd2$Excess.Return, alternative = "greater", mu = 0,

conf.level = 0.95)

t.test(x = stock\_performance\_pd3$Excess.Return, alternative = "greater", mu = 0,

conf.level = 0.95)

t.test(x = stock\_performance\_pd4$Excess.Return, alternative = "greater", mu = 0,

conf.level = 0.95)

t.test(x = stock\_performance\_all$Excess.Return, alternative = "greater", mu = 0,

conf.level = 0.95)

#hypothesis test for testing average absolute winning rate is greater than 0

wilcoxon.test

t.test(x = stock\_performance\_pd1$Abs..Win.Rate, alternative = "greater", mu = 0.7,

conf.level = 0.95)

t.test(x = stock\_performance\_pd2$Abs..Win.Rate, alternative = "greater", mu = 0.7,

conf.level = 0.95)

t.test(x = stock\_performance\_pd3$Abs..Win.Rate, alternative = "greater", mu = 0.7,

conf.level = 0.95)

t.test(x = stock\_performance\_pd4$Abs..Win.Rate, alternative = "greater", mu = 0.7,

conf.level = 0.95)

t.test(x = stock\_performance\_all$Abs..Win.Rate, alternative = "greater", mu = 0.7,

conf.level = 0.95)

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# 5. Bayesian Analysis

install.packages("Bolstad")

library("Bolstad")

par(mfrow=c(1,1))

# Unknown parameter: Mean of Annual Return in period 1

se <- sd(stock\_performance\_pd1$Annual.Return)

x<-stock\_performance\_pd1$Annual.Return

##posterior density for mu, the mean of a normal distribution, with a discrete prior on mu

# find the posterior density with a uniform prior on mu

normdp(x,sigma.x=se)

## find the posterior density with a non-uniform prior on mu

mu = seq(-2,2,by=0.2)

length(mu)

mu.prior = runif(length(mu))

mu.prior = sort(mu.prior/sum(mu.prior))

normdp(x,sigma.x=se,mu,mu.prior)

##posterior density for mu, the mean of a normal distribution, with a normal prior on mu

#find the posterior density with a N(0,1) prior on mu

normnp(x,sigma=1)

## find the posterior density with N(0.5,3) prior on mu

normnp(x,0.5,3,1)

# varying the standard deviation of the prior: increase in se\_prior results in a decrease in se\_posterior

# varying the mean did not produce any significant difference

normnp(x,0.5,1,1)

normnp(x,0.5,3,1)

normnp(x,0.5,5,1)

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*