ST 501 Final R Project - Fall 2018
Bayesian Inference and Analysis

`Due Date:` **December 3, 2018 (5:00PM, EST).**

Sign the **Honor Pledge**: *I have neither given nor received unauthorized aid on this assignment* on your cover page that includes your **full name** and **last four digits of Student ID**, before you submit the solution.

This is a take-home open book and open notes exam, however, you are not allowed to consult or discuss about this exam with anyone except your team member(s) and the instructor. Each team should consist of no more than three students. The solution to this exam should be prepared as thorough but concise, professional quality technical report consisting of double spaced pages. The solution should be turned in as a **single PDF** attachment with following filename formatting: `Lastname1_Lastname2-ST501-Rproject.pdf` by the deadline stated above. Codes (preferably written in `R`) should also be submitted as a **single plain text** attachment with filename `Lastname1_Lastname2-ST501-Codes.txt`

`Total points : 100` `Total time : 3 weeks`

This project will require you to install the R packages `fitdistrplus` and `goftest`. Use these to determine an approximate distribution for daily log-return values of some tech companies. Use the command `install.packages('firdistrplus')` in R to install the package and similarly the other package.

Read the materials available in the following website to familiarize yourself with the use of functions `plotdist, descdist, fitdist, gofstat` from `fitdistrplus` package and the use of `ad.test` from `goftest` package

`http://www.di.fc.ul.pt/~jpn/r/distributions/fitting.html`

A demonstration on downloading the data and illustrations of the R packages will be presented in class, to ensure that same set of data sets are being used by each team. **Please notify the instructor by Nov 19th, the finalized team members**.

1. Consider the historical data available at `https://finance.yahoo.com/` and search for `GOOGL` and click the link `Historical Data` to download the daily transactions for `Adj Close` between Nov 11, 2017 - Nov 11, 2018 (for a total of 251 trading days). You may directly use the alternative weblink `https://finance.yahoo.com/quote/GOOG/history?p=GOOG`

   Let P_t = Adjusted closing price of t-th day for $t = 0, 1, 2, \ldots, 250$ denotes trading days, such that $t = 0$ represents Nov 13th, 2017 and $t = 250$ denotes Nov 9th, 2018. Compute the log-return $R_t = \log(P_t/P_{t-1})$ for $t = 1, 2, \ldots, 250$

   (a) Obtain the 0.05 sample quantile $Q_{0.05}$ of R_t's (using the `quantile` function in R) and also obtain the sample estimate of $E[R|R < Q_{0.05}]$, mean the of those R_t values for which $R_t < Q_{0.05}$.

   (b) Plot the histogram of log-return values using the `plotdist` function from `fitdistrplus` package. Next use the `descdist` function to make a good guess of the probability distribution of R_t's and specify the name of the chosen family of distributions.

   (c) Use the function `fitdist` to obtain the parameters of the chosen distribution and then use `ad.test` function from `goftest` package to determine the suitability of your chosen distribution. A p-value above 0.1 is acceptable (the higher the p-value the more points a team will receive!)

   (d) Let $f_R(r)$ denotes the chosen probability density function (PDF) of the log-returns and write down its exact expression. Show that for any constant c,

   $$E(R|R < c) = \int_{-\infty}^{c} \frac{r f_R(r)}{F_R(c)} dr \quad \text{where } F_R(c) = \int_{-\infty}^{c} f_R(r) dr$$

   Use your chosen PDF and the above formula to compute the following summaries (using Monte Carlo or numerical integration methods if needed):
   $Q_{0.05}(R) = F_R^{-1}(0.05)$ and $E[R|R < Q_{0.05}(R)]$
   and compare the values with the corresponding empirical values that you obtained in (a).

   [5+10+15+10 = 40 points]

   Remark: *The larger the p-value in (c) the better match you'd expect to find between the summary values in (a) and (d). A team will get more points by achieving higher p-values and/or better match between (a) and (d).*

2. Consider again the historical data available at `https://finance.yahoo.com/quote/AMZN/history?p=AMZN` and download the daily transactions for `Adj Close` of `AMZN` between Nov 11, 2017 - Nov 11, 2018 (for a total of 251 trading days).

   Follow the steps in the previous problem and compute the log-return values for the `AMZN` series and repeat the parts (a) - (d) for the log return values of `AMZN` series. You don't need to prove the formula for $E(R|R < c)$ again and just use the formula to compute the theoretical quantiles and expected short fall.

   [5+10+15+5 = 35 points]

3. Let P_t^G and P_t^A denote the adjusted closing prices of Google and Amazon, respectively for $t = 1, 2, \ldots, 250$ (as described in problems 1 & 2 above). Consider a portfolio that combines the two prices as $P_t(\omega) = \omega P_t^G + (1-w)P_t^A$ where $w \in [0, 1]$ is a weight to be determined.

   Let $R_t(\omega) = \log(P_t(\omega)/P_{t-1}(\omega))$ denote the log return of the portfolio for an arbitrary value of ω.

   (a) For each $\omega \in [0, 1]$ obtain the 0.05 sample quantile $Q_{0.05}(\omega)$ of the $R_t(\omega)$ and plot $Q_{0.05}(\omega)$ as a function of ω on grid of equally spaced 50 values in $[0, 1]$.

   (b) For each $\omega \in [0, 1]$ obtain sample estimate of $ES(\omega) = E[R(\omega)|R(\omega) < Q_{0.05}(\omega)]$ and plot $ES(\omega)$ as a function of ω on grid of equally spaced 50 values in $[0, 1]$.

   (c) Find value(s) of ω for which $ES(\omega)$ is minimized and maximized based on a grid search of 50 values.

   [10+10+5 = 25 points]

I have neither given nor received unauthorized aid on this assignment

Ian Huntress ID: 5264

I have neither given nor received unauthorized aid on this assignment.

Deepak Kumar Tiwari ID - 1197

Introduction

This report is for the project that we did as part of our course ST-501. We are thankful to our instructor Dr. Sujit K Ghosh for allowing us to use R programming language to complete this project. This project intends to find Value at Risk of stock returns of Google and Amazon individually and then portfolio having these two stocks. Wikipedia defines Value at Risk as **Value at risk** (**VaR**) is a measure of the risk of loss for investments. It estimates how much a set of investments might lose (with a given probability), given normal market conditions, in a set time period such as a day. VaR is typically used by firms and regulators in the financial industry to gauge the number of assets needed to cover possible losses. Typically, Value at Risk is the lowest 5% returns that could be possible as same as in our project. We have used data of Google and Amazon from finance.yahoo.com. Our data contains Adjusted Closing Price of Stocks for different dates from November 11, 2017, to November 11, 2018. We would be calculating log returns of our investment on stocks. Log returns are simply the natural logarithmic value of closing price for any day minus the logarithmic value of the closing price of the previous day. As we don't have the previous day closing price for November 13, 2017, so, we won't be including log return of stocks for the first day. After finding log returns of each day, we would try to find 5^{th} quantile of our returns and its expected value, that is a value any investor would not want but still want to keep funds or be ready to recover from that loss situation. We would also try to find statistical distribution for our log returns that fits well with the highest p-value. Wikipedia defines p-value as, In statistical hypothesis testing, the ***p*-value** or **probability value** or **asymptotic significance** is the probability for a given statistical model that, when the null hypothesis is true, the statistical summary (such as the sample mean difference between two compared groups) would be greater than or equal to the actual observed results. After finding the appropriate distribution, we would be using Monte Carlo method to find 5^{th} quantile and its expected value for GOOGLE and AMAZON stock individually as well as a portfolio for different weight of both the stocks.

We would be using the R programming language to do this project because it's easy to use and can analyze large numbers of data almost instantly.

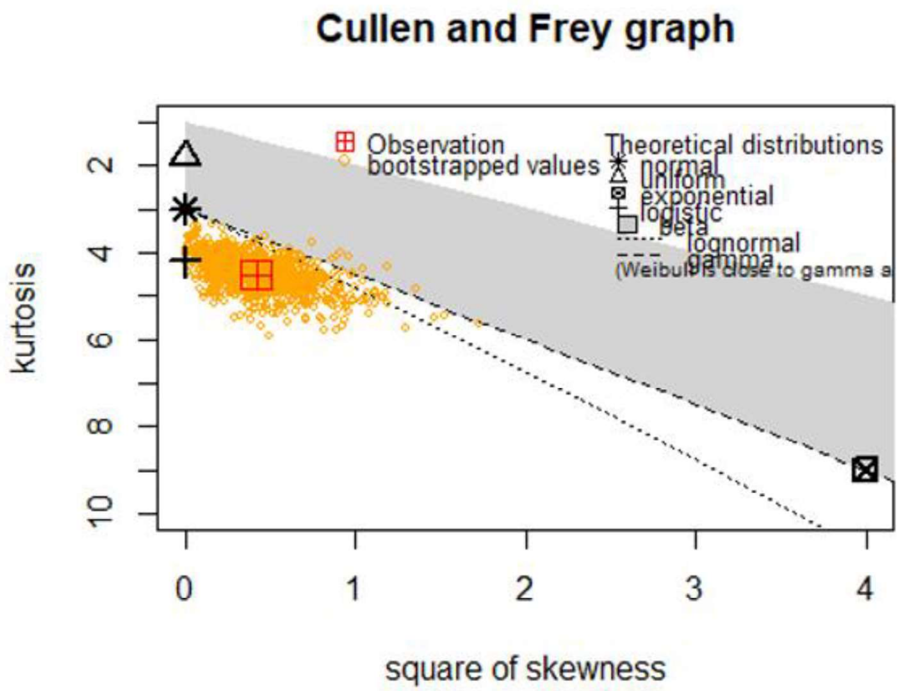
This project report contains a short summary of our results for GOOGLE, AMAZON, and portfolio of both the stocks separately followed by codes that we used to arrive at the result followed by reference page containing all references for materials that we used to complete this project.

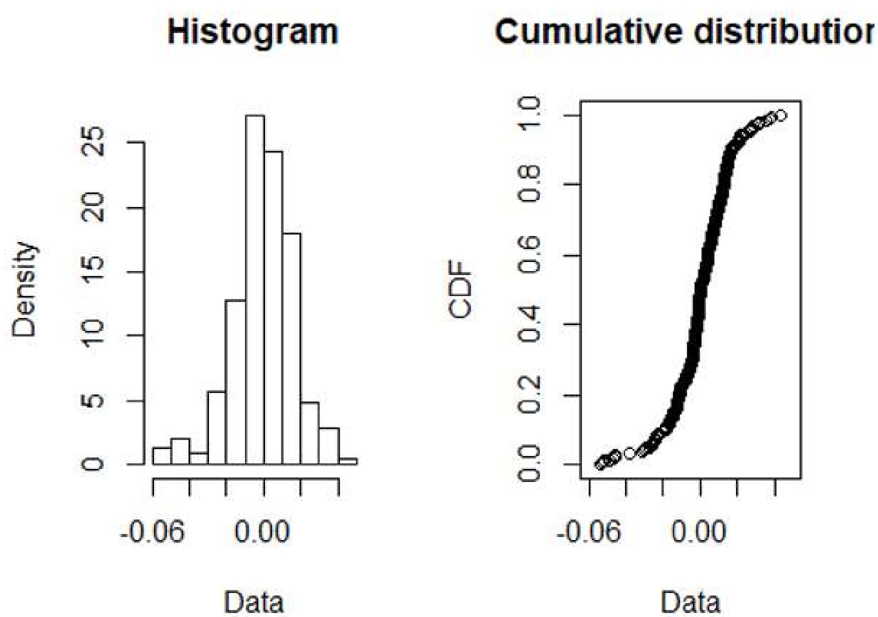
GOOGLE

We found that 0.05th quantile of log return of GOOGLE stock is -0.0269222423214955. An investor can interpret this number as exp (-0.0269222423214955), which is 0.97343675468913806441697063756728, so the closing price for any day is 97.34% approximately of previous trading day which he would want to avoid and want the price to be at least above this number. The expected value of returns below 0.05th quantile is -0.0422122933577563, which means on an average closing price of any day having fewer returns than 0.05th quantile would be 0.95866624049298366722807515690989 or 95.866% of the previous trading day.

We plotted the histogram and Cullen and Frey graph for our log returns and found out that logistic distributions might fit well as it is closer to our data, we tried to find p values for different distributions and found out that p-value for logistic is highest at 0.3967. Using this distribution, we tried to find the 0.05th quantile and expected values below it by using the Monte Carlo method to see how well our chosen distribution fits. Our calculated quantile and expected value came out to be -0.0246741601396428 and -0.0335791338286787 respectively.

Some Results for Google Stock:


Cullen and Frey graph
Observation
bootstrapped values
Theoretical distributions
normal
uniform
exponential
logistic
beta
lognormal
gamma
kurtosis
square of skewness
0
1
2
3
4
2
4
6
8
10

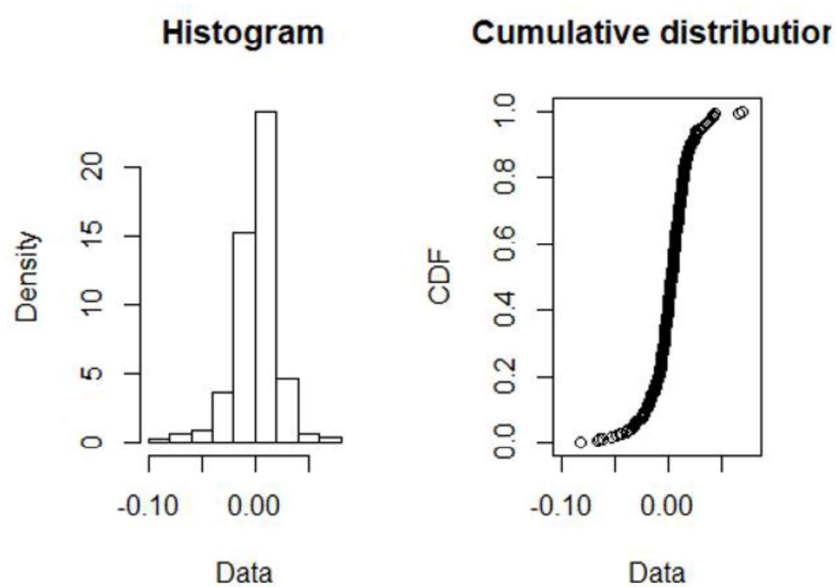

Histogram
Density
Data
0
5
10
15
20
25
-0.06
0.00
Cumulative distributior
CDF
Data
0.0
0.2
0.4
0.6
0.8
1.0
-0.06
0.00

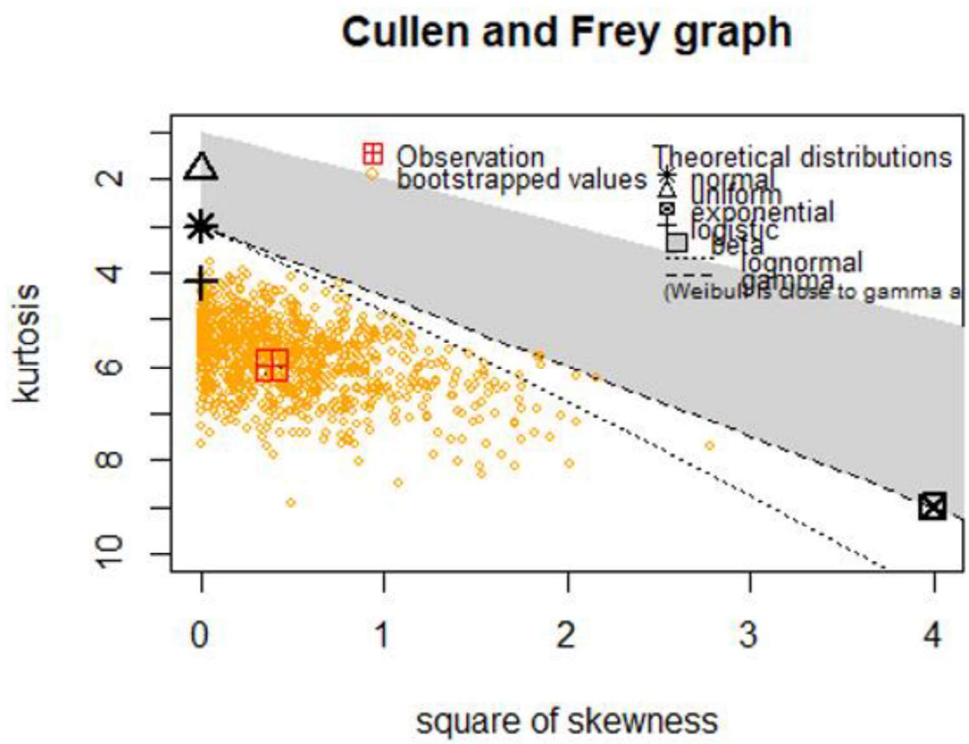
AMAZON

We found the 0.05th quantile of AMAZON stock as -0.03228565926968. An investor can interpret this number same as describes above for GOOGLE stock. The expected value of log returns less than this value is -0.0489619506516627.

We plotted the histogram and Cullen and Frey graph for our log returns and found out that logistic distributions might fit well as it is closer to our data, we tried to find p values for different distributions and found out that p-value for logistic is highest at 0.2098. Using this distribution, we tried to find the 0.05th quantile and expected values below it by using the Monte Carlo method to see how well our chosen distribution fits. Our calculated quantile and expected value came out to be -0.026468285244855 and -0.0335791338286787 respectively.

Some result for Amazon:





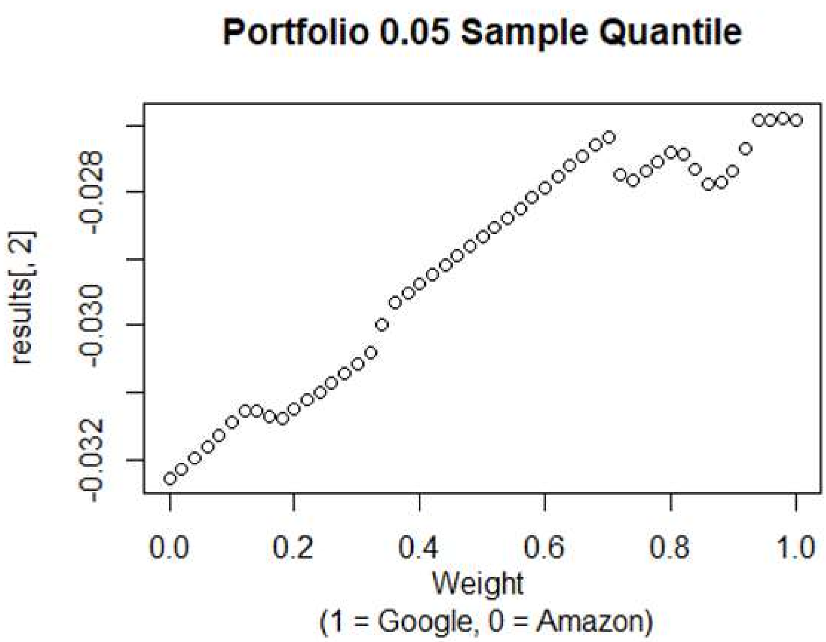
PORTFOLIO

Our Portfolio contains both stocks of GOOGLE and AMAZON. Price of value of portfolio for any day would be given by P(w)=w*(GOOGLE)+(1-w) *(AMAZON), where is a weight ranging from 0 to 1. If w equal to 0 then portfolio contains AMAZON entirely and if w equal to 1, then portfolio would contain GOOGLE entirely. We will be finding log return in the same manner as described before. We tried to find the 0.05th quantile and expected value of log returns less than these quantiles for each value of w ranging from 0 to 1 with step size of 0.02. Here are the 50 quantiles and corresponding expected values that we got for different values of w:

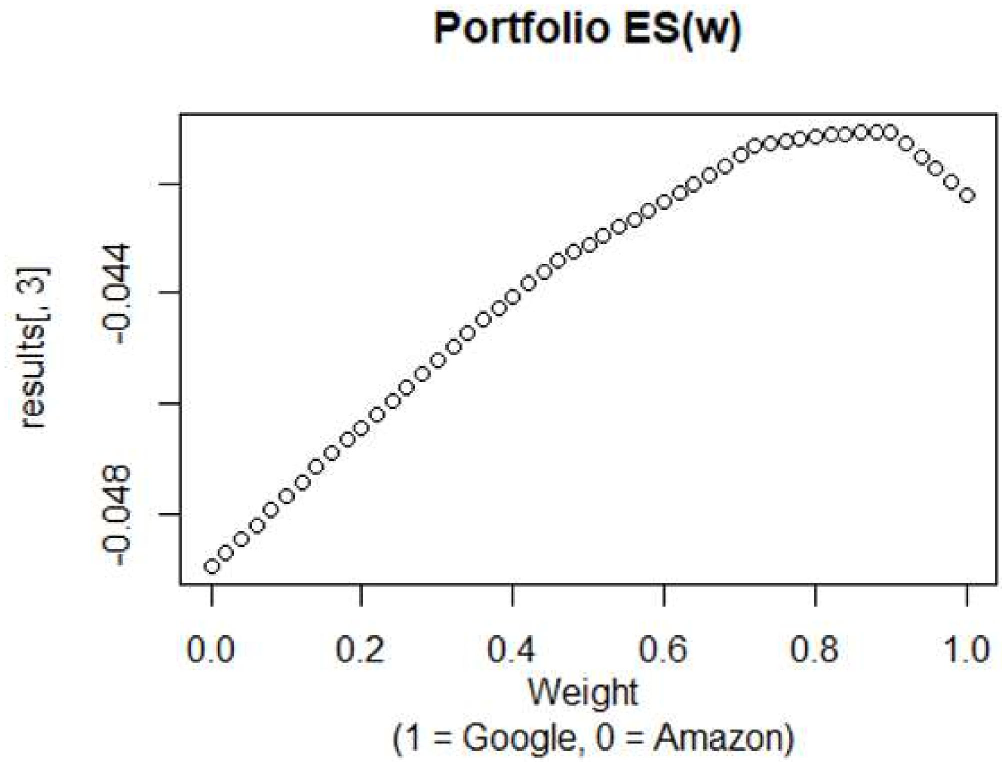
```
## [1] "0,-0.03228565926968,-0.0489619506516627"
## [1] "0.02,-0.0321430782479925,-0.0487140914885715"
## [1] "0.04,-0.0319922193136131,-0.0484632610621292"
## [1] "0.06,-0.0318150709224549,-0.0482094019436072"
## [1] "0.08,-0.0316354133963284,-0.0479524551475879"
## [1] "0.1,-0.0314531915224737,-0.0476923600772964"
## [1] "0.12,-0.0312683484138661,-0.0474290544675678"
## [1] "0.14,-0.0312658801738745,-0.0471624743253251"
## [1] "0.16,-0.0313600807985499,-0.0468970373695395"
## [1] "0.18,-0.0313886312607346,-0.0466668637104395"
## [1] "0.2,-0.0312555047950824,-0.0464336547783473"
## [1] "0.22,-0.0311206758725396,-0.0461973459462559"
## [1] "0.24,-0.0309841115231037,-0.045957870667086"
## [1] "0.26,-0.0308457779171857,-0.0457151604002326"
## [1] "0.28,-0.030705640337325,-0.045469144534669"
## [1] "0.3,-0.0305636631487717,-0.0452197503084114"
## [1] "0.32,-0.030419809768892,-0.0449669027241424"
## [1] "0.34,-0.0299942633349644,-0.0447105244607734"
## [1] "0.36,-0.0296423285047719,-0.0444739937999597"
## [1] "0.38,-0.0295100020278407,-0.0442650701731032"
## [1] "0.4,-0.0293758924249627,-0.0440527124526158"
## [1] "0.42,-0.0292399632807731,-0.0438368356806475"
## [1] "0.44,-0.0291021771778676,-0.0436173520698694"
## [1] "0.46,-0.0289388037681328,-0.0433965943246181"
## [1] "0.48,-0.0288208792056232,-0.043253178325997"
## [1] "0.5,-0.0286772871703478,-0.0431071135455648"
## [1] "0.52,-0.0285316777671717,-0.0429583274366049"
## [1] "0.54,-0.0283840080154647,-0.0428067448001878"
## [1] "0.56,-0.0282342336999794,-0.0426522876630842"
## [1] "0.58,-0.0280823093260246,-0.0424948751488635"
## [1] "0.6,-0.0279281880726449,-0.0423344233417251"
## [1] "0.62,-0.0277718217437388,-0.0421708451425879"
## [1] "0.64,-0.0276131607169775,-0.0420040501169105"
## [1] "0.66,-0.0274521538904217,-0.0418339443336858"
## [1] "0.68,-0.0272887486267167,-0.0416604301949998"
## [1] "0.7,-0.0271833424875118,-0.0414834062555097"
## [1] "0.72,-0.0277282187709412,-0.0413027670311273"
## [1] "0.74,-0.0278338776007679,-0.0412455899590589"
## [1] "0.76,-0.027690994441806,-0.0412061295117769"
## [1] "0.78,-0.027545711937794,-0.0411640697528311"
## [1] "0.8,-0.0273979691391895,-0.0411193067141158"
## [1] "0.82,-0.0274253228222344,-0.0410954513387191"
## [1] "0.84,-0.0276489596923667,-0.0410838654356778"
## [1] "0.86,-0.0278805826529057,-0.0410698489679145"
```

```
## [1] "0.88,-0.0278507185980245,-0.0410532919979561"
## [1] "0.9,-0.0276925760663188,-0.0410621710613598"
## [1] "0.92,-0.0273417977409692,-0.0412670324144819"
## [1] "0.94,-0.02693745090797,-0.0414919161540386"
## [1] "0.96,-0.0269161349019496,-0.0417186329051333"
## [1] "0.98,-0.0268974604313206,-0.0419471913083924"
## [1] "1,-0.0269222423214955,-0.0422122933577563"
```

We tried to plot the quantiles and expected values with different values of w and got this:



We can see that quantiles increases as w increases, Also, we got this plot for expected values of log returns for different values w:



We can see from both the graphs that quantiles and Expected values increase as w increases that means an investor would prefer Amazon stocks more in their portfolio to keep Value at Risk minimum. We can see from the Expected value vs w plot that Expected value of returns less than 0.05^{th} quantile is minimum at w=0 and maximum at w=0.88

(d) $f_R(r)$ = PDF of log return

we need to prove that for any constant c,

$$E(R|R<c) = \int_{-\infty}^{c} \frac{r f_R(r)}{F_R(c)} dr.$$

where $F_R(c) = \int_{-\infty}^{c} f_R(r) dr$

Soln.

We know that Expected ~~variable~~ value of any X given Y is given by.

$$E[X|Y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx. \quad \text{—(1)}$$

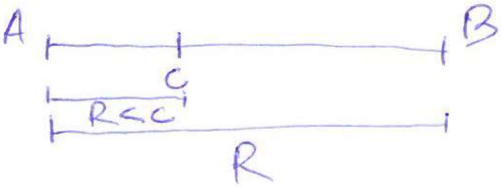
$$\text{where } f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} \quad \text{—(2)}$$

Here let's assume X = R, random variable representing log returns of all days.

Let Y be R<C, that be random variable representing log returns of only those values of R which are less than C.

Please note that R<C is actually a part of R.

We can understand it in ~~on~~ this way.



Let's say R is a random variable with range A and B

$$R \in (A, B).$$

and $R<C$ is a random variable with range A and C. so, joint function of. R and $R<C$ would be actually R.

so, $$f_{R|R<C} = \frac{f_R(r)}{f_{R<C}(r)}. \qquad -(3).$$

where $f_{R<C}(r)$ is a marginal function of R less than C

we know that marginal fnctn can be derived from joint function as:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y).dx.$$

Substituting values we will have

$$f_{R<C}(r) = \int_{-\infty}^{C} f_R(r)\, dr \text{. as } f_{R<C} \in (-\infty, C) \qquad -(4)$$

and joint function as discussed above is $f_R(r)$

Substituting (4) in (3) we will have.

$$f_{R|R<C} = \frac{f_R(r).}{\int_{-\infty}^{c} f_R(r)\, dr} \qquad -(5)$$

we are given that $F_R(c) = \int_{-\infty}^{c} f_R(r)\,dr.$

So, (5) becomes.

$$f_{R|R<c} = \frac{f_R(r)}{F_R(c)} \qquad -(6).$$

Substituting (6) in (1) we will have.

$$\boxed{E(R/R<c) = \int_{-\infty}^{c} \frac{r\, f_R(r)}{F_R(c)}\,dr.} \qquad -(7).$$

Here limit is $(-\infty, c)$ because $R<c$ is actually having range $(-\infty, c)$.

So, we have proved that (7) is actually true and can be used to find expected value of log returns of all values which are less than c. In question

$c = 0.05$.

References:

1. https://finance.yahoo.com/screener

2. http://www.stat.umn.edu/geyer/old/5101/rlook.html

3. https://en.wikipedia.org

4. ***Mathematical Statistics and Data Analysis, 3rd Edition***