

Homework 1

Assignment Info

Homework #: HW1
Description: Density Histogram
Course: EN.553.636 Introduction to Data Science
Semester: Spring 2023, Homewood Campus
Instructor: Tamas Budavari
TA: Matthew Tivnan
Date: January 27, 2023

Student Info

Name: Amir Hossein Daraie
JHED-ID: adaraie1
Email: adaraie1@jhu.edu (<mailto:adaraie1@jhu.edu>) \

We load data on monthly stock returns from 1926 to 2021 ([source \(https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html\)](https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html)).

```
In [2]: import numpy as np
import pandas as pd

stocks = pd.read_csv("stocks.csv")
ret = stocks["Mkt-RF"].values
```

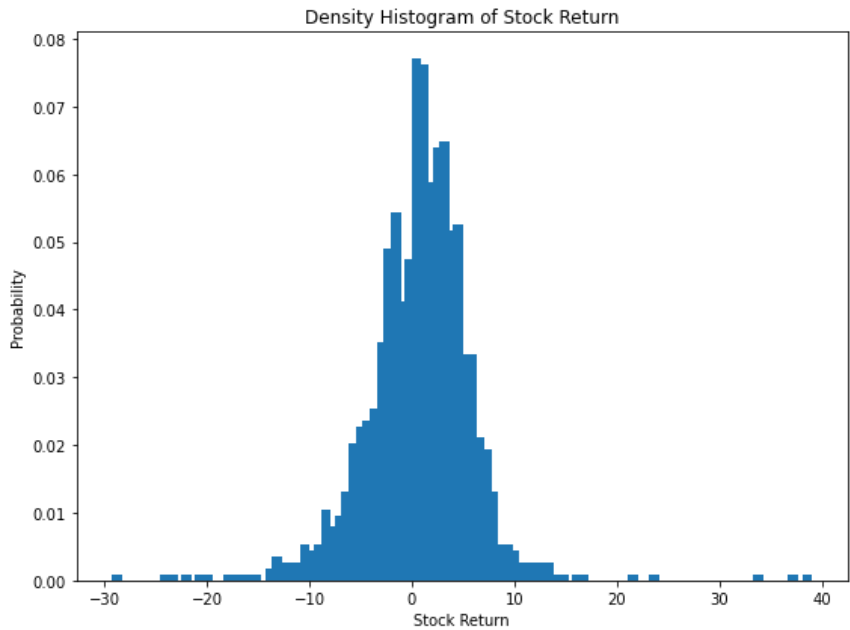
A *density histogram* is a histogram that is normalized so that the total area under the bars sums to unity. Argue that the stock return is not normally distributed as follows:

- 1. Plot a density histogram of the returns `ret` with 100 bins. [2 pts]
- 2. Over this histogram, plot the PDF of a normal distribution with mean equal to the sample mean of the returns and standard deviation equal to the sample standard deviation of the returns. [2 pts]

```
In [35]: # Import librarie we need
import matplotlib.pyplot as plt
from scipy.stats import norm
```

```
In [34]: # Calculate histogram and bins
npoints = ret.shape[0]
H, bins = np.histogram(ret, bins=100)
H = H/npoints # convert each bar from count to probability
bins = (bins[:-1] + bins[1:])/2

# Plot density histogram
plt.figure(figsize=(8,6))
plt.bar(bins,H,width=1)
plt.title("Density Histogram of Stock Return")
plt.xlabel("Stock Return")
plt.ylabel("Probability")
plt.tight_layout()
plt.show()
```

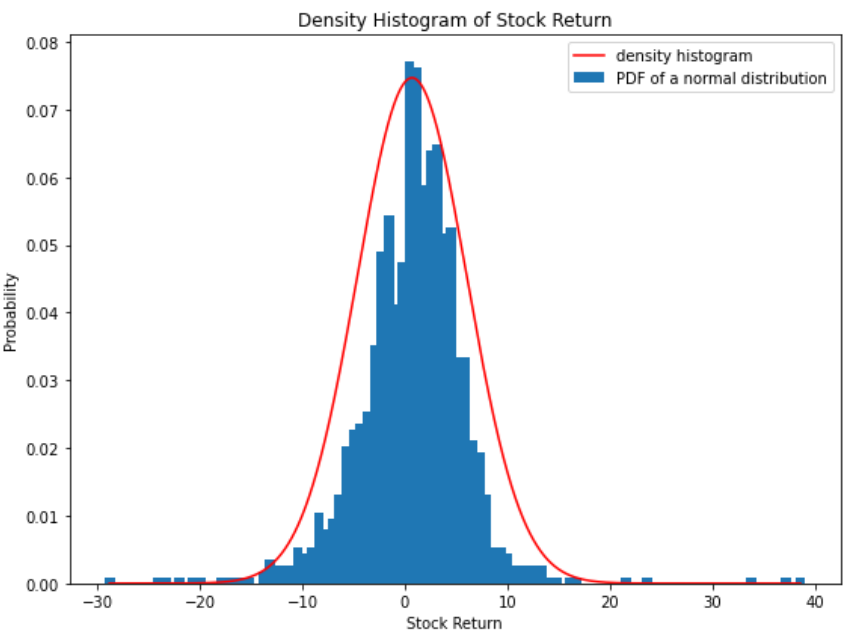


```
In [40]: # plot the PDF of a normal distribution

# Plot between a min and max value on x with 0.01 steps.
x_axis = np.arange(bins.min(), bins.max(), 0.01)

# Calculating mean and standard deviation
mean = np.mean(ret)
sd = np.std(ret)

# Plot density histogram & PDF of a normal distribution
plt.figure(figsize=(8,6))
plt.bar(bins,H,width=1)
plt.plot(x_axis, norm.pdf(x_axis, mean, sd), c='r')
plt.title("Density Histogram of Stock Return")
plt.xlabel("Stock Return")
plt.ylabel("Probability")
plt.legend(["density histogram","PDF of a normal distribution"])
plt.tight_layout()
plt.show()
```



3. Explain why stock returns are not normally distributed. [2 pts]

If we assume that they are normally distributed, it comes with a big cost. In terms of generality, it's a very severe simplifying assumption, especially because those deviation from normality (red plot) more often than not, they can be pretty severe.

Henceforth, because of many deviations from normality, it is difficult to simplify them as a set of numbers from normal distribution. This deviation can be calculated as follows:

```
In [45]: dev = norm.pdf(bins, mean, sd) - H
         np.sum(dev)
```

Out[45]: 0.47102087153281097

And as you can appreciate here with numbers, it is something hard to ignore.