FRE6083, Midterm Examination, Due on October 30th, 11:59pm

1. Number of pages including this one: 4
2. For this examination, you may use the class lecture notes, homework assignments and their solutions as well as download data on yahoo finance. You may not talk to your peers and you must write your answers independently as well as justify them carefully and clearly. If an argument lacks clarity or rigor, you may not get full credit.
3. Please, include in your paper the following academic honesty statement and sign it: I affirm that I will not give or receive any unauthorized help on this exam and that all work will be my own. I will complete this exam in a fair, honest, respectful, responsible, and trustworthy manner. Student's Signature: Utkarshbhanu Ganesh Andurkar

Problem 1 (40 points)

1. (5 points) Download from Yahoo Finance the last year of daily data

for the price of a share of Bank of America Corporation (BOFA). You

will use the closing prices. The goal of this exercise is to compute and

visualize the linear autocorrelations, as well as determine whether they

are significant or not by using a statistical test. I suggest that you use

Matlab but you may also alternatively use R or Python.

In [1]:
```
import pandas as pd
```

In [2]:
```
df = pd.read_csv("BAC.csv")
```

In [3]:
```
df = df[['Close']]
```

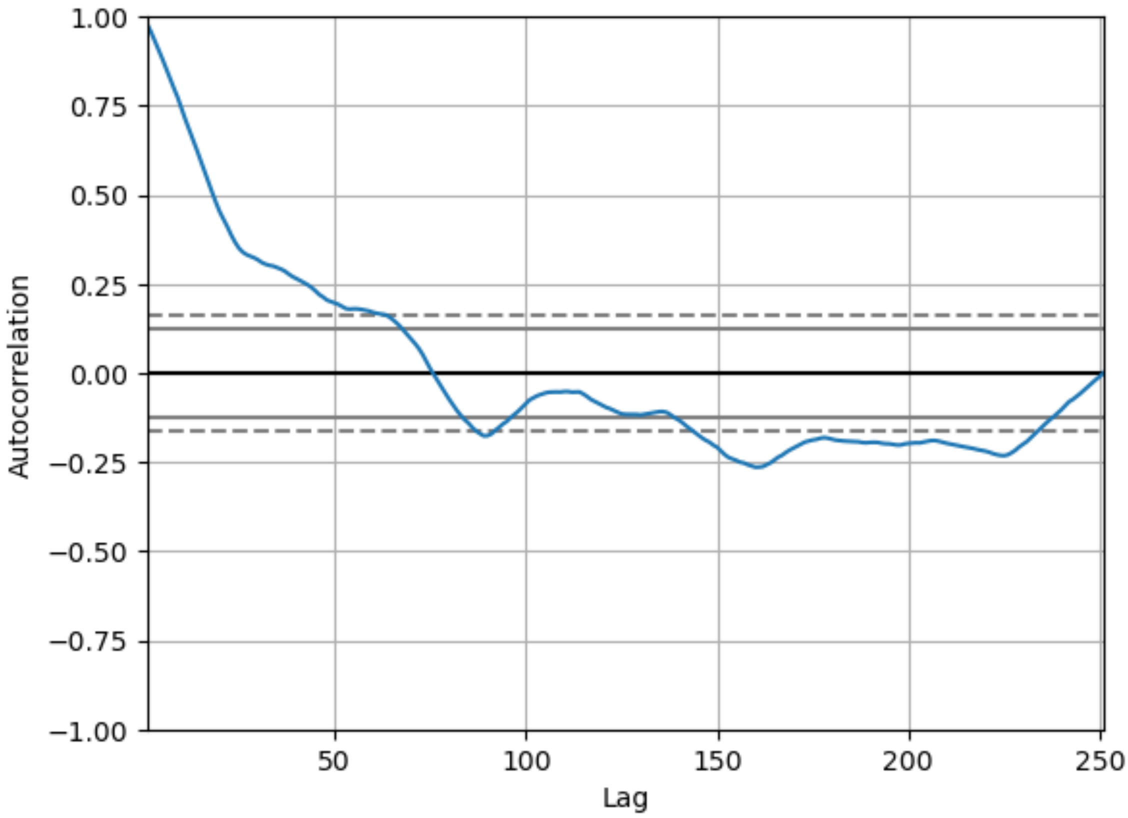
In [4]:
```
df.head()
```

Out[4]:

	Close
0	36.180000
1	36.040001
2	36.200001
3	36.090000
4	35.889999

In [5]:
```
pd.plotting.autocorrelation_plot(df)
```

Out[5]: <Axes: xlabel='Lag', ylabel='Autocorrelation'>



In [6]:
```
# If we plot autocorrelation on prices directly,
#it suggests presence of dependency as the prices
#are not stationary.
#So in further section we will use log
#returns instead of prices
```

In [7]:
```
from statsmodels.stats.stattools import durbin_watson
import numpy as np
```

In [8]:
```
durbin_watson(df['Close'])
```

Out[8]: 0.0002633682690462184

In [9]:
```
# As the value of durbin watson statistic is below 2,
# there is positive autocorrelation present in the prices
```

In [ ]:

2. (5 points) We denote by S(t) the stock price at the close and by X(t) =

logS(t) the corresponding log price where log represents the natural

logarithm. The data frequency is Δt = 1 day. The time unit is ”day”.

The daily log return is then defined as

r(t) = X(t + 1) − X(t).

Compute the daily log returns for the entire data set and plot them

as a function of the time t. We assume that the log returns are strict

sense stationary for the remainder of the problem. However, keep in

mind that the potential presence of non stationarities could threaten

the validity of your results.

In [10]:
```
Xt = pd.DataFrame([], columns=['log_return'])
```

In [11]:
```
Xt['log_return'] = np.log(df['Close'])
```

In [12]:
```
Xt.head()
```

Out[12]:

	log_return
0	3.588506
1	3.584629
2	3.589059
3	3.586016
4	3.580459

In [13]:
```
rt = Xt.diff()
```

In [14]:
```
rt = rt[1:]
```

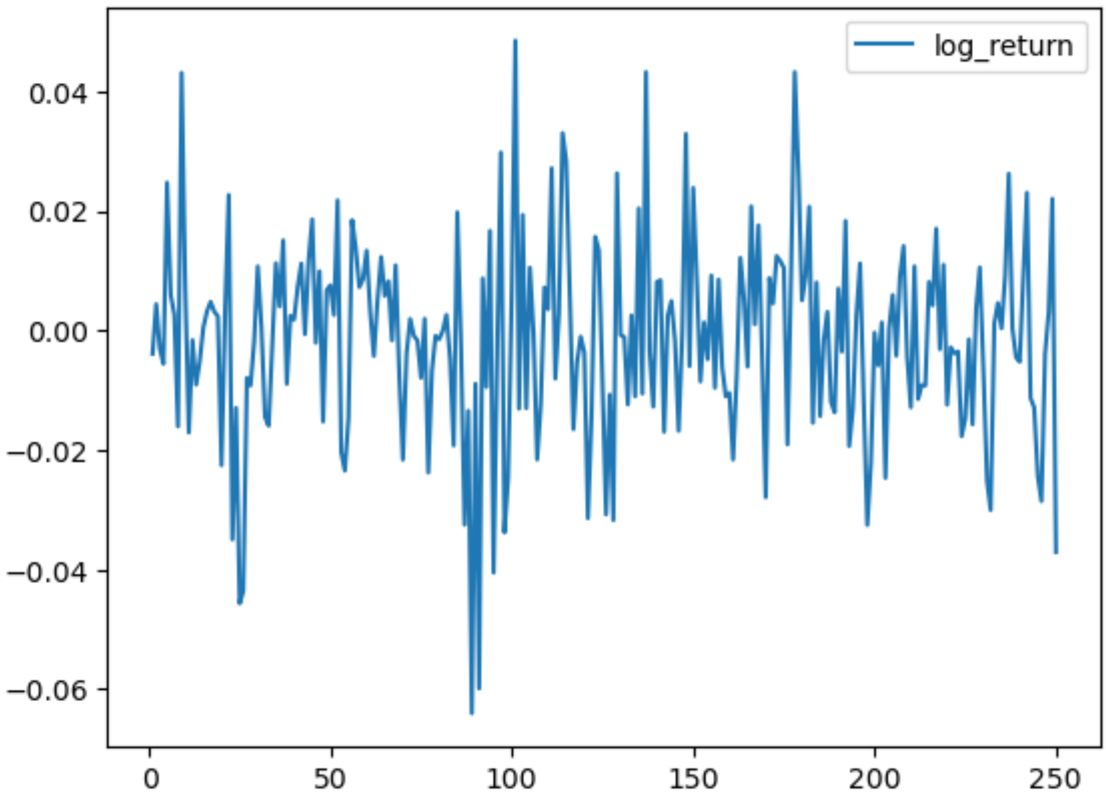
In [15]:
```
rt.head()
```

Out[15]:

	log_return
1	-0.003877
2	0.004430
3	-0.003043
4	-0.005557
5	0.024767

In [16]:
```
rt.plot()
```

Out[16]: <Axes: >



In [ ]:

3. (10 points) We denote by τ the time lag. It is a multiple of $\Delta t = 1$.

Choose a fixed time t and compute the autocorrelation function

corr[(r(t + τ), r(t))],

which represents the correlation between the daily return at time t and

the daily return τ days later by using the sample autocorrelation in

Matlab autocorr and plot it as a function of the lag τ ranging from 0

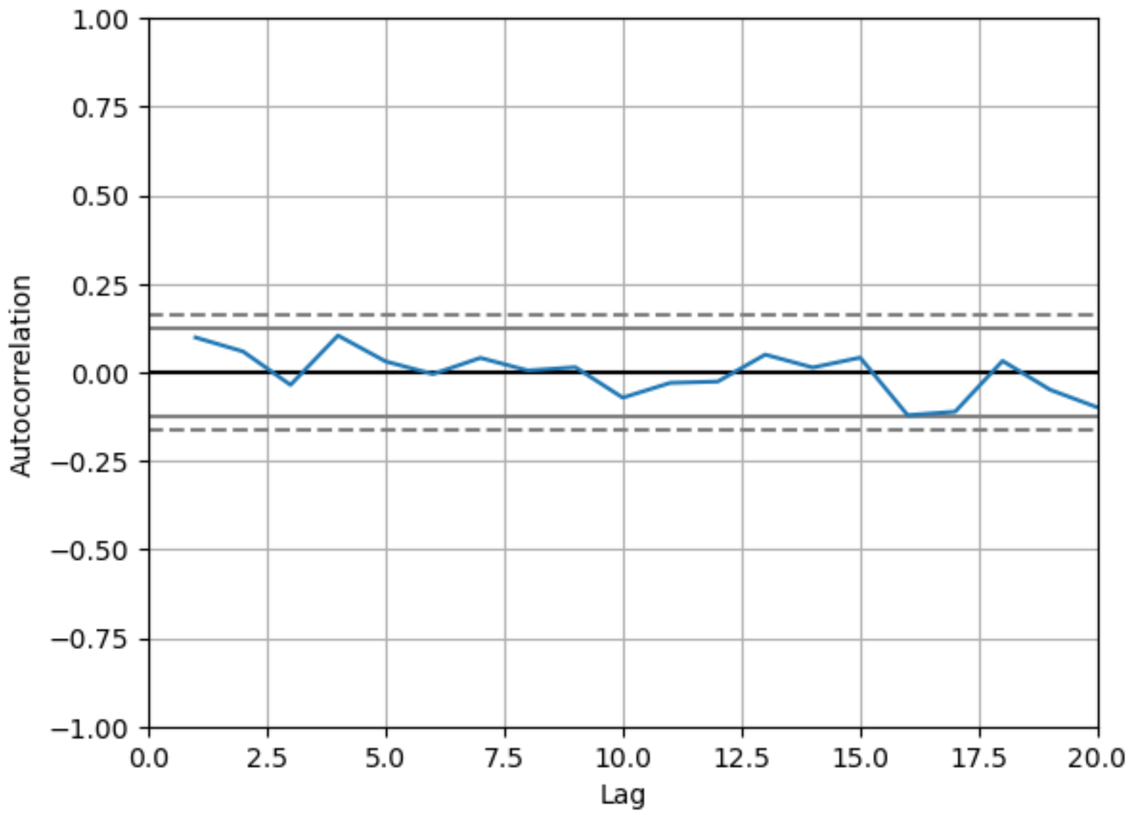
to 20.

In [17]:

```
ax = pd.plotting.autocorrelation_plot(rt)
ax.set_xlim([0, 20])
# plotting autocorrelation using
#autocorrelation plot directly
```

Out[17]:

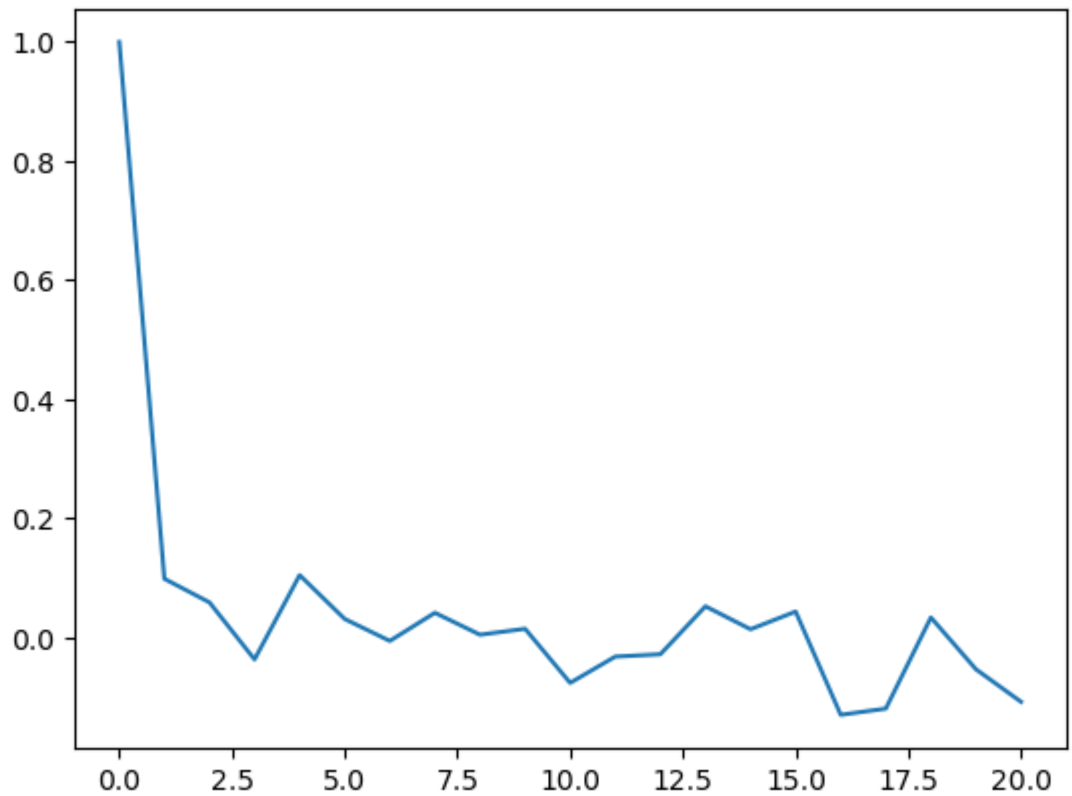
```
(0.0, 20.0)
```



In [18]:
```
#Calculating autocorrelation first
#and then plotting it seperately
autoCorr = []
for i in range(0, 21):
    autoCorr.append(rt['log_return'].autocorr(lag = i))
```

In [19]:
```
import matplotlib.pyplot as plt
```

In [20]:
```
xpoints = np.array([i for i in range(21)])
ypoints = np.array(autoCorr)
plt.plot(xpoints, ypoints)
plt.show()
```



4. (10 points) Perform a visual inspection of the results. Do the auto-

correlations seem significant? If needed, justify your answer further by

using the Ljung-Box Q-Test statistical test (lbqtest in Matlab).

In [21]:
```
#Autocorrelations do not seem
#significant as there is no
#trend in autocorrelation
```

In [22]:
```
import statsmodels.api as sm
```

In [23]:
```
sm.stats.acorr_ljungbox(rt, lags=[i for i in range(21)], return_df=True)
```

Out[23]:

	lb_stat	lb_pvalue
0	21.025313	NaN
1	2.433614	0.118759
2	3.304888	0.191581
3	3.632047	0.304036
4	6.382730	0.172331
5	6.624959	0.250060
6	6.632202	0.356203
7	7.057098	0.422959
8	7.063133	0.529837
9	7.114862	0.625162
10	8.472523	0.582783
11	8.710456	0.648599
12	8.894544	0.711910
13	9.551194	0.730158
14	9.598550	0.790906
15	10.053619	0.816354
16	14.016741	0.597467
17	17.374585	0.429280
18	17.650551	0.478887
19	18.316301	0.501411
20	21.025313	0.395644

In [24]:

```
# Here all the p values are
#much greater than significance
#level of 5% ie 0.05.
# Hence ljung box q test result
#suggests that the log returns
#are not linearly dependent
```

In [ ]:

5. (10 points) Next, we turn to the nonlinear autocorrelation functions

corr(|r(t + τ)|, |r(t)|),

where | · | denotes the absolute value, and

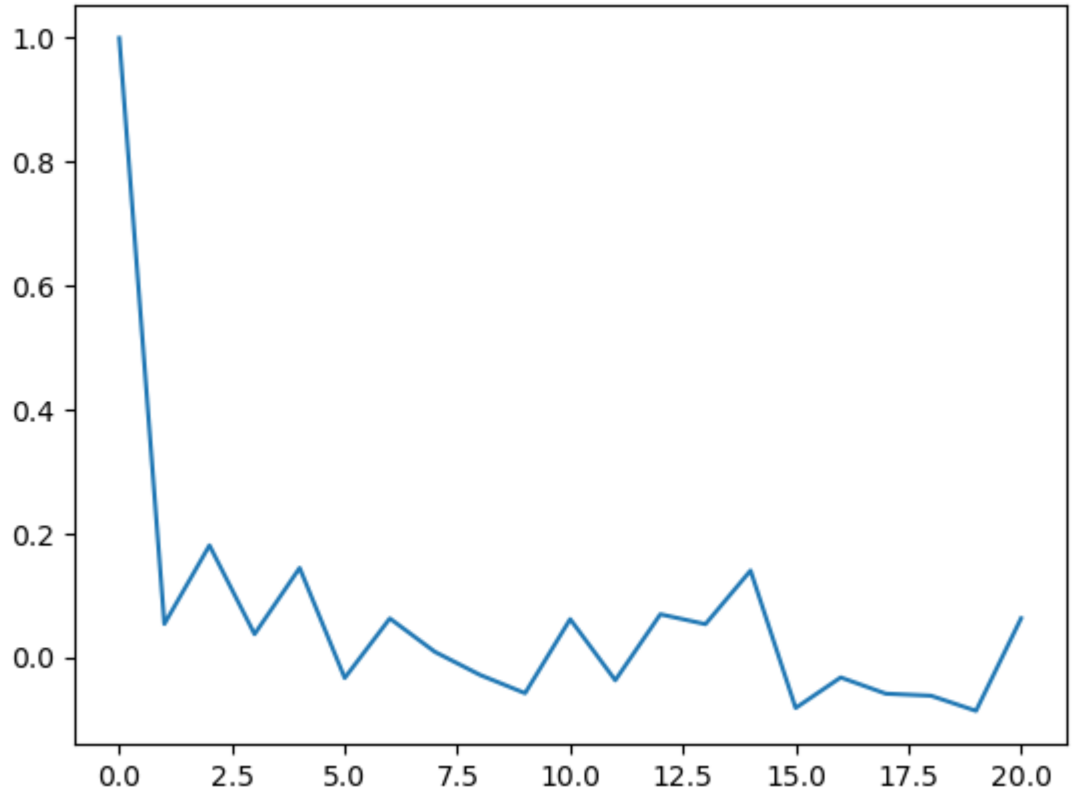
corr(r(t + τ)2, r(t)2).

Apply to these two nonlinear autocorrelation functions the same steps

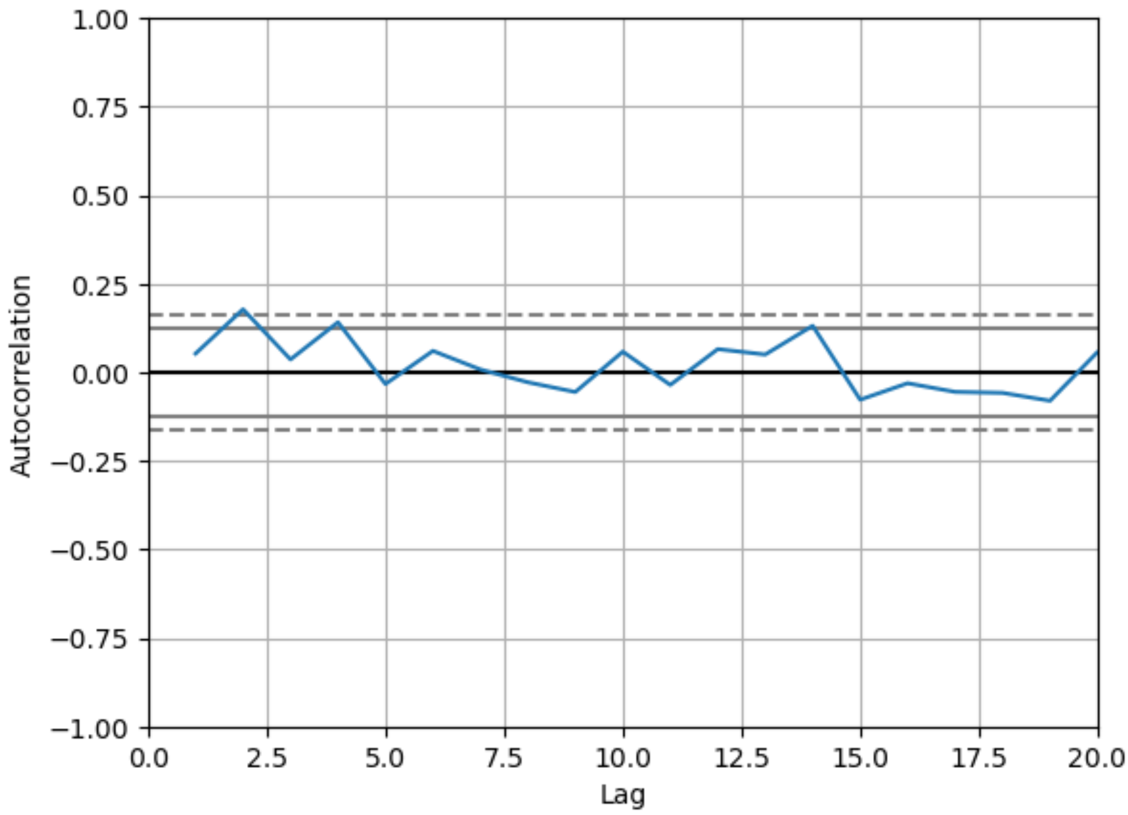
as in parts 3-4. Discuss the results and conclude.

In [25]:
```
def getAbsAutoCorr(df):
    dfabs = pd.DataFrame()
    dfabs['log_return'] = df['log_return'].abs()

    #Calculating autocorrelation first and then plotting it seperately
    autoCorr = []
    for i in range(0, 21):
        autoCorr.append(dfabs['log_return'].autocorr(lag = i))
    xpoints = np.array([i for i in range(21)])
    ypoints = np.array(autoCorr)
    plt.plot(xpoints, ypoints)
    plt.show()

    # plotting autocorrelation using autocorrelation plot directly
    ax = pd.plotting.autocorrelation_plot(dfabs)
    ax.set_xlim([0, 20])

    res = sm.stats.acorr_ljungbox(dfabs, lags=[i for i in range(21)], ret

    autoCorrelationSignificantCount = 0
    for i in range(len(res['lb_pvalue'])):
        if(res['lb_pvalue'][i] < 0.05):
            autoCorrelationSignificantCount += 1
    print(res)
    print('autocorrelation present/significant for ' + str(autoCorrelatio
```

In [26]:
```
getAbsAutoCorr(rt)
```




```
      lb_stat  lb_pvalue
0   30.281740        NaN
1    0.699962   0.402797
2    8.711116   0.012835
3    9.043254   0.028721
4   14.135250   0.006876
5   14.419293   0.013154
6   15.353618   0.017678
7   15.369015   0.031550
8   15.578985   0.048818
9   16.393201   0.059111
10  17.274604   0.068504
11  17.613332   0.090997
12  18.738482   0.095038
13  19.402219   0.111151
14  23.989864   0.045952
15  25.604418   0.042392
16  25.865863   0.055945
17  26.699369   0.062627
18  27.621957   0.068046
19  29.393147   0.060047
20  30.281740   0.065415
autocorrelation present/significant for 9 Lag values at 95% confidence le
vel
```

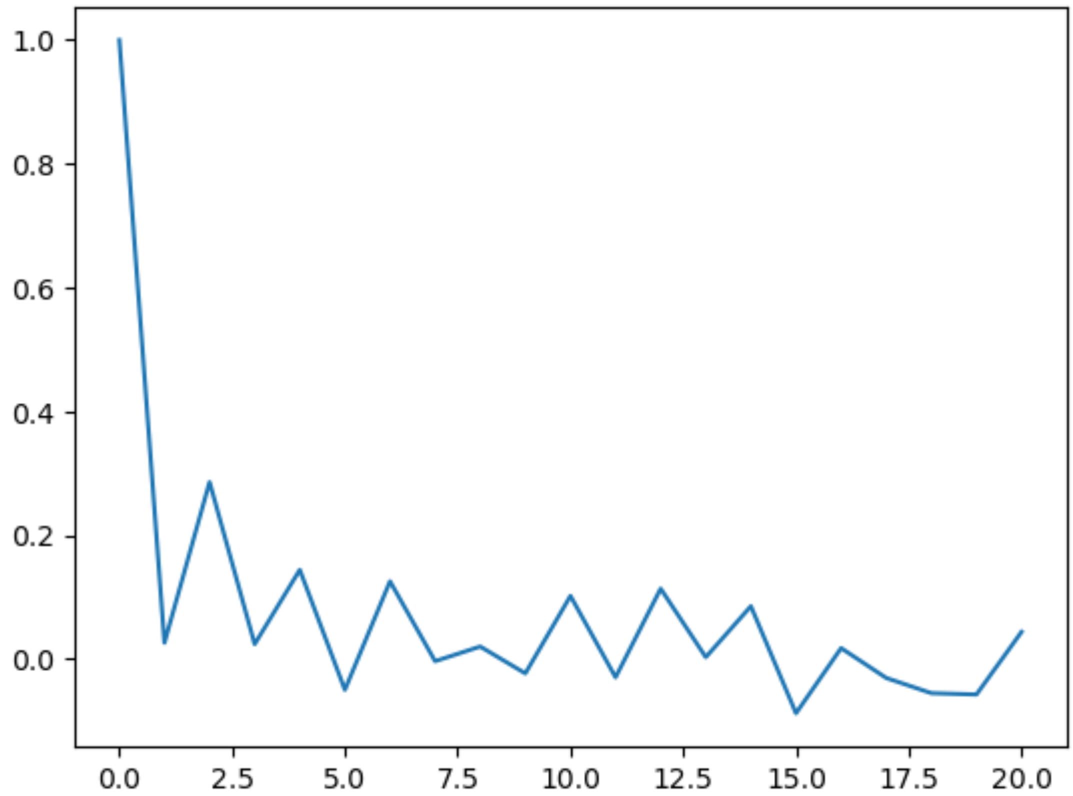


In [27]:
```
# By visual inspection for some of
#the values of lag we are getting
# linear dependence relationship
```

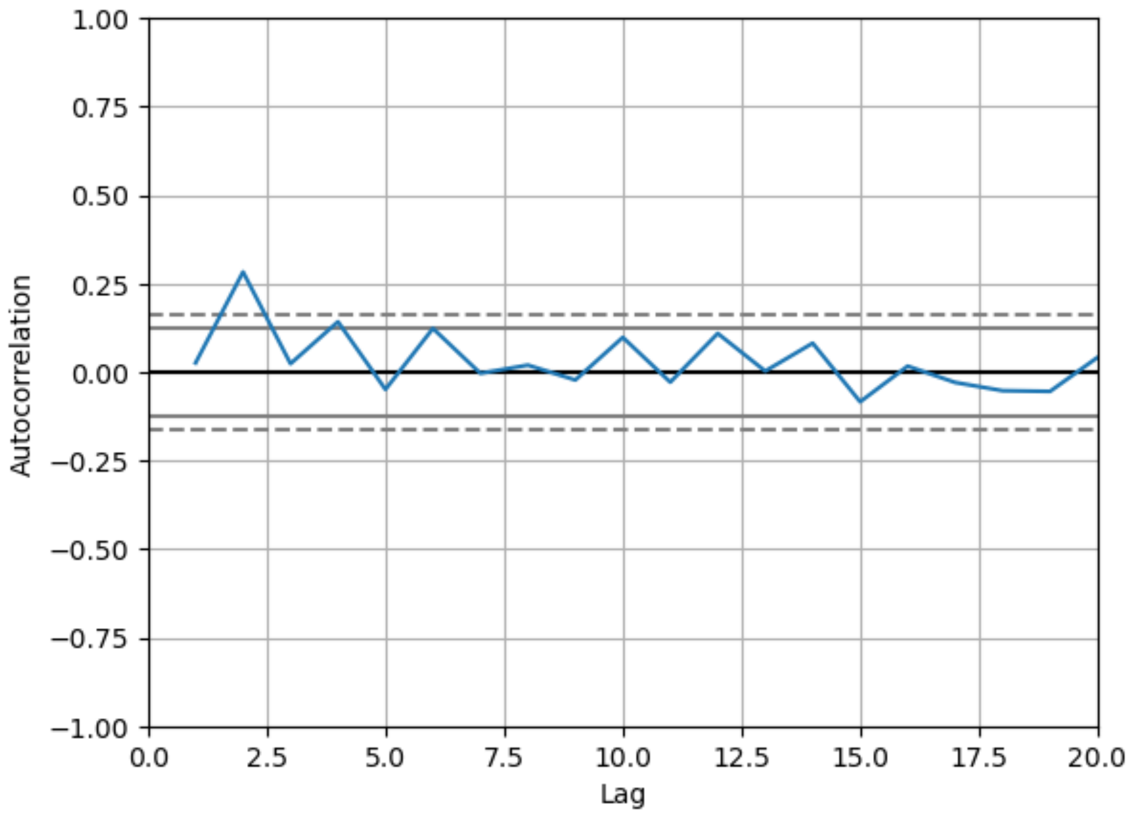
In [28]:
```
#if we use the first non linear
# correlation involving absolute values,
# for 9 values of the lag, the test
# results suggest that log returns have
# linear dependence
```

In [29]:
```
def getSquaredAutoCorr(df):
    dfsq = pd.DataFrame()
    dfsq['log_return'] = df['log_return']**2

    #Calculating autocorrelation first and then plotting it seperately
    autoCorr = []
    for i in range(0, 21):
        autoCorr.append(dfsq['log_return'].autocorr(lag = i))
    xpoints = np.array([i for i in range(21)])
    ypoints = np.array(autoCorr)
    plt.plot(xpoints, ypoints)
    plt.show()

    # plotting autocorrelation using autocorrelation plot directly
    ax = pd.plotting.autocorrelation_plot(dfsq)
    ax.set_xlim([0, 20])

    res = sm.stats.acorr_ljungbox(dfsq, lags=[i for i in range(21)], retu
    autoCorrelationSignificantCount = 0
    for i in range(len(res['lb_pvalue'])):
        if(res['lb_pvalue'][i] < 0.05):
            autoCorrelationSignificantCount += 1
    print(res)
    print('autocorrelation present/significant for ' + str(autoCorrelatio
```

In [30]:
```
getSquaredAutoCorr(rt)
```




```
      lb_stat  lb_pvalue
0   42.392954        NaN
1    0.166101   0.683600
2   20.445058   0.000036
3   20.585154   0.000128
4   25.743586   0.000036
5   26.368821   0.000076
6   30.274488   0.000035
7   30.278320   0.000084
8   30.375961   0.000181
9   30.509135   0.000359
10  33.018710   0.000270
11  33.235130   0.000482
12  36.371635   0.000282
13  36.373345   0.000519
14  38.141343   0.000494
15  40.039960   0.000447
16  40.113448   0.000749
17  40.349132   0.001155
18  41.107911   0.001471
19  41.927090   0.001813
20  42.392954   0.002457
autocorrelation present/significant for 19 Lag values at 95% confidence l
evel
```



In [31]:
```
# By visual inspection for some of
#the values of lag we are getting
#dependent relationship/trend
```

In [32]:
```
# if we use the second non
#linear correlation involving squares,
# for almost all (19) values of the
#lag we can say that log returns
# are dependent
```

In [33]:
```
# So whether or not there is presence
# of autocorrelation (dependency) will
# depend on the measure we choose to
# compute the autocorrelation(linear or non linear)
```

In [ ]:

Mid term

Problem 2 (20 points).

We consider the aggregate loss distribution seen in class

$$L(t) = \sum_{i=1}^{N(t)} X_i,$$

where N is a Poisson distribution with parameter ; process

$\lambda > 0$, and $X_1, X_2, X_3 \dots$, the claim amounts are i.i.d random variables, independent of process N. Furthermore, we also assume here that X_i has an exponential distribution with mean μ and variance μ^2, that is

$$F_X(x) = 1 - e^{-\frac{x}{\mu}}$$

1. (5 points) Compute the expected loss.

$$\to \quad E[L(t)] = E\Big[\sum_{i=1}^{\overset{\text{random}}{N(t)}} \underset{\text{random}}{X_i}\Big]$$

to take the Σ out we have to condition on one of the random variable's.

$$\therefore E[L(t)] = \sum_{k=0}^{\infty} E\Big[\sum_{i=1}^{N(t)} X_i \mid N(t)=k\Big] P[N(t)=k]$$

$$\therefore E[L(t)] = \sum_{k=0}^{\infty} E\left[\sum_{i=1}^{N(t)} X_i \mid N(t)=k \right] \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$

We can remove the condition as it is already incorporate in main argument.

$$\therefore E[L(t)] = \sum_{k=0}^{\infty} E\left[\sum_{i=1}^{k} X_i \right] \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$

$$= \sum_{k=0}^{\infty} \sum_{i=1}^{k} \underbrace{E[X_i]}_{\mu \text{ (given)}} \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$

$$= \sum_{k=0}^{\infty} \left(\sum_{i=1}^{k} \mu \right) \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$

$$= \sum_{k=0}^{\infty} k\mu \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$

When $k=0$ whole term is 0

$$\therefore E[L(t)] = \mu \sum_{k=1}^{\infty} \frac{e^{-\lambda t}(\lambda t)^k}{(k-1)!}$$

$$= \mu \lambda t \sum_{k=1}^{\infty} \frac{e^{-\lambda t}(\lambda t)^{k-1}}{(k-1)!}$$

using change of index,

$$E[L(t)] = \mu\lambda t \underbrace{\sum_{k=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^k}{k!}}$$

This is sum of all the probabilities of poisson distribution

$$\therefore E[L(t)] = \mu\lambda t \cdot (1) = \lambda\mu t$$

$$\boxed{\therefore \text{ Expected Loss} = \lambda\mu t}$$

2. (5 points) Compute the variance of loss.

$$\rightarrow \quad Var[L(t)] = E[(L(t))^2] - (E[L(t)])^2$$

$$E[(L(t))^2] = E\left[\left(\sum_{i=1}^{N(t)} x_i\right)^2\right]$$

$$= \sum_{k=0}^{\infty} E\left[\left(\sum_{i=1}^{N(t)} x_i\right)^2 \Big| N(t)=k\right] P[N(t)=k]$$

$$= \sum_{k=0}^{\infty} E\left[\left(\sum_{i=1}^{k} x_i\right)^2\right] \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$

$$\overline{\quad}\ (1)$$

$$E\left[\left(\sum_{i=1}^{k} x_i\right)^2\right] = E\left[\sum_{i=1}^{k} x_i^2 + \sum\sum_{\substack{j,m=1 \\ j \neq m}}^{k} x_j x_m\right]$$

As $E\left[\sum_{i=1}^{k} x_i^2\right] = \sum_{i=1}^{k} E[x_i^2]$

& $E[x_i^2] = \mu^2 + \mu^2$ [From given]

$= 2\mu^2$

$\therefore E\left[\sum_{i=1}^{k} x_i^2\right] = 2k\mu^2$

& $E\left[\sum_{j=1}^{k} \sum_{\substack{m=1 \\ j \neq m}}^{k} x_j x_m\right] = \sum_{j=1}^{k} \sum_{\substack{m=1 \\ j \neq m}}^{k} E[x_j \cdot x_m]$

As all x_is are independent

$$= \sum_{j=1}^{k} \sum_{\substack{m=1 \\ j \neq m}}^{k} E[x_j] E[x_m]$$

$$= \sum_{j=1}^{k} \sum_{\substack{m=1 \\ j \neq m}}^{k} \mu^2$$

$$= k(k-1)\mu^2$$

$$\therefore E\left[\left(\sum_{i=1}^{k} x_i\right)^2\right] = 2k\mu^2 + k(k-1)\mu^2$$

$$= (k^2 - k + 2k)\mu^2 = (k^2+k)\mu^2$$

Substituting in eq^n (1)

$$E\left[(L(t))^2\right] = \sum_{k=0}^{\infty} (k^2 - k + 2k)\,\mu^2 \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$

$$= \mu^2 \left[\sum_{k=0}^{\infty} k^2 \frac{e^{-\lambda t}(\lambda t)^k}{k!} + \sum_{k=0}^{\infty} k \frac{e^{-\lambda t}(\lambda t)^k}{k!} \cancel{+ \underbrace{2\sum_{k=0}^{\infty} k \frac{e^{-\lambda t}(\lambda t)^k}{k!}}_{1\ \text{(sum of all probabilities)}}} \right]$$

also $\displaystyle\sum_{k=0}^{\infty} k \frac{e^{-\lambda t}(\lambda t)^k}{k!}$ = Expectation of poisson distribution

$= \lambda t$

$\displaystyle\sum_{k=0}^{\infty} k^2 \frac{e^{-\lambda t}(\lambda t)^k}{k!}$ = Variance of poisson distribution + (Expectation of posson distribution)^2

$$= \lambda t + (\lambda t)^2$$

$$\therefore E\left[(L(t))^2\right] = \mu^2\left[\lambda t + (\lambda t)^2 + \lambda t\right]$$

$$= \mu^2(\lambda t)^2 + 2\mu^2 \lambda t$$

as

$$Var(L(t)) = E[(L(t))^2] - (E[(L(t))])^2$$

$$\therefore Var(L(t)) = (\mu\lambda t)^2 + 2\mu^2\lambda t - (\lambda\mu t)^2$$

$$\boxed{\therefore Var(L(t)) = 2\mu^2\lambda t}$$

3. (5 points) We recall that the moment generating function of the exponentially distributed variable X_i is given by

$$\phi_{X_i}(x) = \frac{\frac{1}{\mu}}{\frac{1}{\mu} - x},$$ and that the

moment generating function of a gamma distribution with parameters n and a is

$$\phi_{\Gamma}(x) = \left(\frac{a}{a-x}\right)^n.$$

Show that the conditional density function of $\sum_{i=1}^{n} X_i$, for a given $n \geq 1$ is

$$f(\lambda|n) = \frac{(\lambda/\mu)^{n-1}}{\mu(n-1)!} e^{-\lambda/\mu}, \quad \lambda \geq 0,$$

by using the above moment generating functions.

→ For a given n,

$$L(t) = \sum_{i=1}^{n} X_i \,, \text{ given } \phi_{X_i}(x) = \frac{\frac{1}{\mu}}{\frac{1}{\mu} - x} = E\left[e^{-tX_i}\right] \quad -(1)$$

Finding moment generating function of L,

$$E\left[e^{-tL}\right] = E\left[e^{-t\sum_{i=1}^{n} X_i}\right]$$

$$= E\left[\prod_{i=1}^{n} e^{-tX_i}\right]$$

$$= \prod_{i=1}^{n} E\left[e^{-tX_i}\right]$$

by independence of all X_is ⇒ independence of all e^{-tX_i}s

$$= \prod_{i=1}^{n} \frac{\frac{1}{\mu}}{\frac{1}{\mu} - x} = \left(\frac{\frac{1}{\mu}}{\frac{1}{\mu} - x}\right)^n \quad -(2)$$

We can rename x as l as it is a particular value/level

$$\therefore E\left[e^{-tL}\right] = \left(\frac{\frac{1}{\mu}}{\frac{1}{\mu} - l}\right)^n$$

but ~~as~~ we know that a gamma distribution with parameters n and a, then its moment generating function is ~~gamma distribution~~

$$\phi_T(x) = \left(\frac{a}{a - x}\right)^n \quad -(3)$$

[Given]

As moment generating function is unique for a particular distribution & moment generating function of Loss $L(t)$ is same form as moment generating function of gamma distribution → Hence Loss follows gamma distribution

As ~~[illegible]~~ eqⁿ (2) & (3) are ~~[illegible]~~ equal,

$$\therefore \left(\frac{\frac{1}{\mu}}{\frac{1}{\mu}-x}\right)^n = \left(\frac{a}{a-x}\right)^n$$

$$\therefore a = \frac{1}{\mu}, \quad n = n$$

Probability density function of gamma distribution with parameters n and a is

$$f_X(x) = \begin{cases} \dfrac{a^n \cdot x^{n-1} e^{-ax}}{\Gamma(n)} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

in our case, $a = \frac{1}{\mu}$, $x = \lambda$ & $\Gamma(n) = (n-1)!$

$$f(\lambda|n) = \begin{cases} \dfrac{\left(\frac{1}{\mu}\right)^n \cdot \lambda^{n-1} \cdot e^{-\frac{1}{\mu}\cdot\lambda}}{(n-1)!} & \lambda \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\therefore f(\lambda|n) = \frac{\left(\frac{\lambda}{\mu}\right)^{n-1} \cdot e^{-\frac{\lambda}{\mu}}}{\mu\,(n-1)!}, \quad \lambda \geq 0$$

4. (5 points) Give a formula for the first-order density function $g(t, \lambda)$, of $L(t)$, at time t. Note that it will contain an infinite sum that you should not attempt to compute explicitly.

→ At time t, we would have observed the value of poisson process, which would be $N(t)$ ← not random as observed at time t.

∴ ~~We can~~ as $L(t) = \sum_{i=1}^{N(t)} X_i$, where X_i are exponentiall distributed with mean μ.

$L(t)$ will be sum of $N(t)$ exponential distributions, which is a gamma distribution with parameters $n = N(t)$ and $a = \frac{1}{\mu}$ as shown in previous example.

The cumulative density function is (by total probability rule)

$$G(t,\lambda) = \sum_{n=0}^{N(t)} f(\lambda \mid n) \, ,$$ [illegible]

$$\therefore G(t,\lambda) = \frac{\sum_{n=0}^{N(t)} \left(\frac{\lambda}{\mu}\right)^{n-1} e^{-\left(\frac{\lambda}{\mu}\right)}}{\mu (n-1)!} \, , \quad \lambda \geq 0$$

$\therefore$ first order density function is

$$\frac{d}{d\lambda} G(t,\lambda) = G'(t,\lambda)$$

$$\therefore \; g(t,\lambda) = G'(t,\lambda)$$

$$= \sum_{n=0}^{N(t)} \frac{1}{\mu (n-1)!} \frac{d}{d\lambda} \left(\frac{\lambda}{\mu}\right)^{n-1} e^{-\left(\frac{\lambda}{\mu}\right)} \quad \text{---} (1)$$

as
$$\frac{d}{d\lambda}\left[\left(\frac{\lambda}{\mu}\right)^{n-1} e^{-\left(\frac{\lambda}{\mu}\right)}\right] = \left(\frac{\lambda}{\mu}\right)^{n-1} \frac{d}{d\lambda} e^{-\left(\frac{\lambda}{\mu}\right)} + e^{-\left(\frac{\lambda}{\mu}\right)} \frac{d}{d\lambda}\left(\frac{\lambda}{\mu}\right)^{n-1}$$

$$= \left(\frac{\lambda}{\mu}\right)^{n-1} e^{-\left(\frac{\lambda}{\mu}\right)} \cdot \frac{-1}{\mu} + e^{-\left(\frac{\lambda}{\mu}\right)} \cdot \times \frac{n-1}{\mu^{n-1}} \cdot \lambda^{n-2}$$

$$\therefore \frac{d}{d\lambda}\left[\left(\frac{\lambda}{\mu}\right)^{n-1} e^{-\left(\frac{\lambda}{\mu}\right)}\right] = \frac{-1}{\mu}\cdot\left(\frac{\lambda}{\mu}\right)^{n-1} e^{-\left(\frac{\lambda}{\mu}\right)} + \frac{n-1}{\mu^{n-1}}\cdot \lambda^{n-2} e^{-\left(\frac{\lambda}{\mu}\right)}$$

$$= \frac{n-1}{\mu}\left(\frac{\lambda}{\mu}\right)^{n-2} e^{-\left(\frac{\lambda}{\mu}\right)} - \frac{1}{\mu}\left(\frac{\lambda}{\mu}\right)^{n-1} e^{-\left(\frac{\lambda}{\mu}\right)}$$

$$= \frac{1}{\mu}\, e^{-\left(\frac{\lambda}{\mu}\right)}\left(\frac{\lambda}{\mu}\right)^{n-2}\left[n-1-\frac{\lambda}{\mu}\right]$$

putting this in equation 1.

$$g(t,\lambda) = \sum_{n=0}^{N(t)} \frac{1}{\mu(n-1)!}\cdot\frac{1}{\mu}\cdot e^{-\left(\frac{\lambda}{\mu}\right)}\left(\frac{\lambda}{\mu}\right)^{n-2}\left[n-1-\frac{\lambda}{\mu}\right]$$

$$= \sum_{n=0}^{N(t)} \frac{e^{-\left(\frac{\lambda}{\mu}\right)}}{\mu^{n}(n-1)!}\cdot\lambda^{(n-2)}\left[n-1-\frac{\lambda}{\mu}\right]$$

Problem 3 (20 points)

Consider the Markov chain with states $\{0, 1, \ldots, n\}$ and transition probabilities

$$P_{0,1} = P_{n,n-1} = 1$$

$$P_{i,i+1} = p_i \quad , \quad P_{i,i-1} = 1 - p_i$$

this implies $P_{i,i} = 0 = 1 - (P_{i,i+1} + P_{i,i-1})$

where $0 < p_i < 1$ are given for all $i = 1, \ldots, n-1$.

1. (6 points) Write the transition probability matrix P.

→

$$\underset{(n\times 1)\times(n\times 1)}{P} = \begin{array}{c|cccccccc} & 0 & 1 & 2 & \cdots & i & \cdots & n-1 & n \\ \hline 0 & 0 & 1 & 0 & \cdots & & \cdots & 0 & 0 \\ 1 & (1-p_1) & 0 & p_1 & 0 & \cdots & \cdots & 0 & 0 \\ \vdots & 0 & 0 & \cdots & & & & 0 & 0 \\ i & 0 & 0 & (1-p_i) & 0 & p_i & & 0 & 0 \\ \vdots & \vdots & & & & & & & \\ n-1 & & & & & (1-p_{n-1}) & 0 & p_{n-1} & \\ n & 0 & 0 & 0 & \cdots & & & 1 & 0 \end{array}$$

$(n+1)\times(n+1)$: Dimension

alternatively we can reprent it as

$$P = \begin{bmatrix} 0 & 1 & & & \\ & \ddots & & & 0 \\ & (1-P_i) & 0 & P_i & \\ 0 & & & \ddots & \\ & & & 1 & 0 \end{bmatrix}$$

2. (6 points) Write the system of linear equations satisfied by the candidate stationary probabilities.

→ $$\pi = \pi P$$

$$\sum_{j=0}^{n} \pi_j = 1$$

where $\pi = (\pi_0, \pi_1, \dots \pi_n)$

$\pi_0 = P_{1,0}\pi_1 = (1-P_1)\pi_1$

$\therefore$ $\pi_n = P_{n-1,n}\pi_{n-1} = P_{n-1}\pi_{n-1}$

for all other i's ~~other i's~~

$\pi_i = P_{i+1,i}\,\pi_{i+1} + P_{i-1,i}\,\pi_{i-1} = (1-P_{i+1})\pi_{i+1} + P_{i-1}\pi_{i-1}$

3. (8 points) Solve this system to compute the candidate stationary probabilities π_i for $i = 0, \ldots, n$

as $\pi = \pi P$

$$\therefore \pi = [\pi_0 \; \pi_1 \; \pi_2 \ldots \pi_n] \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ & \ddots & & & \\ & (1-P_i) & 0 & P_i & \\ & & & \ddots & \\ 0 & & & 1 & 0 \end{bmatrix}$$

$$\therefore \pi_0 = (1-P_1)\pi_1 \Rightarrow \pi_1 = \frac{\pi_0}{1-P_1}$$

$$\pi_1 = P_{0,1}\pi_0 + P_{2,1}\pi_2$$

$$= 1.\pi_0 + (1-P_2)\pi_2$$

$$\therefore \frac{\pi_0}{1-P_1} = \pi_0 + (1-P_2)\pi_2$$

$$\therefore \pi_0\left[\frac{1}{1-P_1} - 1\right] = (1-P_2)\pi_2$$

$$\therefore \pi_2 = \pi_0 \left[\frac{P_1}{(1-P_1)(1-P_2)}\right]$$

Similarly:

$$\pi_2 = (1-P_3)\pi_3 + P_1\pi_1$$

$$\therefore (1-P_3)\pi_3 = \pi_0\left[\frac{P_1}{(1-P_1)(1-P_2)}\right] - P_1 \frac{\pi_0}{(1-P_1)}$$

$$\pi_3 = \frac{\pi_0 P_1 P_2}{(1-P_1)(1-P_2)(1-P_3)}$$

Similarly

$$\Pi_3 = (1-P_4)\Pi_4 + P_2\Pi_2$$

$$\therefore (1-P_4)\Pi_4 = \Pi_0\left[\frac{P_1 P_2}{(1-P_1)(1-P_2)(1-P_3)}\right] - P_2 \cdot \frac{\Pi_0 P_1}{(1-P_1)(1-P_2)}$$

$$= \frac{\Pi_0 P_1 P_2}{(1-P_1)(1-P_2)}\left[\frac{1}{1-P_3} - 1\right]$$

$$\therefore \Pi_4 = \frac{\Pi_0 P_1 P_2 P_3}{(1-P_1)(1-P_2)(1-P_3)(1-P_4)}$$

We can recognize the patten.
Hence for any i, $i = 1 \ldots n-1$

multiply $i-1$ terms

$$\Pi_i = \Pi_0 \times \left[\frac{\prod_{j=1}^{i-1} P_j}{\prod_{j=1}^{i} (1-P_j)}\right]$$

multiply i terms

~~Use~~ $\Pi_0 = d(1-P_1)\Pi_1 = d(1-P_1)$

As

$$\pi_n = P_{n-1}\, \pi_{n-1}$$

$$\therefore \pi_n = P_{n-1} \cdot \pi_0 \times \left[\frac{\prod_{j=1}^{n-2} P_j}{\prod_{j=1}^{n-1} (1-P_j)} \right]$$

$$\therefore \pi_n = \pi_0 \times P_{n-1} \times \left[\frac{\prod_{j=1}^{n-2} P_j}{\prod_{j=1}^{n-1} (1-P_j)} \right] \quad -(1)$$

as we have all the values of P_j, $j = 1, \ldots, n-1$

we can find all π_j s in terms of π_0 by substituting the value of P_j in the equation ~~to find all π_i~~ for $j = 1, \ldots, n-1$

as ~~π_0~~ we can also find the value of π_n in terms of π_0 using eq^n (1).

Now we can use

$$\sum_{j=0}^{n} \pi_j = 1$$

to solve for π_0. As we get the value of π_0, we can find all other π_j s $\forall j$, $j = 1 \ldots n$.

Problem 4 (20 points)

We consider the sequence of independent variables X_j that have a uniform distribution on $(-j, j)$.

1. (4 points) Compute the characteristic function $\phi_{X_j}(u)$ of X_j.

→ $\phi_{X_j}(u) = E\left[e^{-iuX_j}\right]$

As X_j follows uniform distribution its PDF is

$$f(x) = \begin{cases} \frac{1}{2j} & -j < x < j \\ 0 & \text{elsewhere.} \end{cases}$$

$$\therefore E\left[e^{-iuX_j}\right] = \int_{-j}^{j} e^{-iux} \cdot \frac{1}{2j}\, dx$$

$$= \frac{1}{2j}\left[\frac{e^{-iux}}{-iu}\right]_{-j}^{j}$$

$$= \frac{1}{2j}\left[\frac{e^{-iuj} - e^{iuj}}{-iu}\right]$$

$$E\left[e^{-iuX_j}\right] = \frac{1}{i(uj)}\left[\frac{e^{i(uj)} - e^{-i(uj)}}{2}\right]$$

as

$$\sinh(ix) = \frac{e^{ix} - e^{-ix}}{2}$$

$$\therefore E\left[e^{-iux_j}\right] = \frac{1}{i(uj)} \cdot \sinh(iuj)$$

as $\sinh(ix) = i\sin x$ [Trigonometric simiplification]

$$\therefore E\left[e^{-iux_j}\right] = \frac{1}{i(u_j)} \cdot i\sin(u_j)$$

$$\therefore E\left[e^{-iux_j}\right] = \frac{\sin(u_j)}{u_j} = \phi_{x_j}(u)$$

2. (4 points) Deduce a formula for the characteristic function $\phi_{S_n}(u)$ of the sum

$$S_n = \sum_{j=1}^{n} X_j$$

$$\rightarrow \quad \phi_{S_n}(u) = E\left[e^{-iuS_n}\right]$$

$$= E\left[e^{-iu\sum_{j=1}^{n} x_j}\right]$$

$$\therefore \phi_{S_n}(u) = E\left[\prod_{j=1}^{n} e^{-iux_j}\right]$$

As all x_j's are independent.

$$\therefore \phi_{S_n}(u) = \prod_{j=1}^{n} E\left[e^{-iux_j}\right]$$

using last Problem. sol^n to

$$\therefore \phi_{S_n}(u) = \prod_{j=1}^{n} \frac{\sin(ju)}{ju}$$

3. (4 points) Deduce by using similar arguments a formula for the characteristic function $\phi_{S_n/n^{3/2}}(u)$ of $S_n/n^{3/2}$.

$$\phi_{S_n/n^{3/2}}(u) = E\left[e^{-iu\frac{S_n}{n^{3/2}}}\right]$$

$$= E\left[e^{-\frac{iu}{n^{3/2}}\sum_{j=1}^{n} x_j}\right]$$

as all x_j's are independent.

$$= E\left[\prod_{j=1}^{n} e^{-\frac{iu}{n^{3/2}}x_j}\right]$$

$$= \prod_{j=1}^{n} E\left[e^{-\frac{iu}{n^{3/2}}x_j}\right]$$

using previous sol^n

$$\therefore \phi_{S_n/n^{3/2}}(u) = \prod_{j=1}^{n} \frac{\sin\left(\frac{u}{n^{3/2}} \cdot j\right)}{\frac{u}{n^{3/2}} \cdot j}$$

4. (5 points) Show that the limit of $\phi_{S_n/n^{3/2}}(u)$ as $n \to +\infty$ is $e^{-u^2/18}$ by

using taylor expansion.

→ As taylor series expansion of $\sin x$ is at origin

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots$$

$$\therefore \frac{\sin\left(\frac{u}{n^{3/2}} \cdot j\right)}{\frac{u}{n^{3/2}} \cdot j} = 1 - \frac{1}{3!}\left(\frac{u}{n^{3/2}} \cdot j\right)^2 + \frac{1}{5!}\left(\frac{u}{n^{3/2}} \cdot j\right)^4 - \dots$$

as $n \to \infty$ the effect of terms from term 3 onwards diminishes. Hence we can approximate the above function as

$$\frac{\sin\left(\frac{uj}{n^{3/2}}\right)}{\frac{uj}{n^{3/2}}} \approx 1 - \frac{1}{3!}\left(\frac{u \cdot j}{n^{3/2}}\right)^2$$

$$= 1 - \frac{(uj)^2}{6n^3}$$

$$\therefore \frac{\sin\left(\frac{uj}{n^{3/2}}\right)}{\left(\frac{uj}{n^{3/2}}\right)} = 1 + \frac{1}{6}\frac{(-u^2 j^2)}{n^3}$$

As $$\phi_{S_n/n^{3/2}}(u) = \prod_{j=1}^{n} \frac{\sin\left(\frac{u}{n^{3/2}} \cdot j\right)}{\frac{u}{n^{3/2}} \cdot j}$$

$$\phi_{S_n/n^{3/2}}(u) = \prod_{j=1}^{n} \left(1 + \frac{1}{6} \cdot \frac{(-u^2 j^2)}{n^3}\right)$$

as $$\lim_{x\to\infty} \left(1 + \frac{k}{x^m}\right)^x = e^{\frac{k}{m}}$$ (formula of limits)

$$\therefore \lim_{n\to\infty} \phi_{S_n/n^{3/2}}(u) = \lim_{n\to\infty} \prod_{j=1}^{n} \left(1 + \frac{1}{6} \cdot \frac{(-u^2 j^2)}{n^3}\right)$$

if we put every j as 1

as $n \to \infty$, $n^3 \to \infty$

$$= \lim_{n\to\infty} \left(1 + \frac{1}{6} \cdot \frac{(-u^2)}{n^3}\right)^n$$

$$= e^{-u^2 \times \frac{1}{6} \times \frac{1}{3}} = \underline{\underline{e^{-\frac{u^2}{18}}}}$$

Hence proved.

5. (3 points) Conclude that the sequence $S_n / n^{3/2}$ converges in distribution

as n goes to ∞ to a normal distribution with mean 0 & variance $1/9$.

→ We know

$$\cancel{\lim_{n \to \infty} \phi \frac{S_n}{n^{3/2}}}$$

$$\lim_{n \to +\infty} \phi_{S_n/n^{3/2}}(u) = e^{-u^2/18} \quad (1)$$

This means as n tends to $+\infty$, the characteristic function of $\frac{S_n}{n^{3/2}}$

converges to $e^{-u^2/18}$

& as characteristic function uniquely describes a probability distribution $e^{-u^2/18}$ should ~~the~~ ~~the~~ ~~pdf~~ of ~~the~~ . describe the distribution uniquely.

The form of characteristic function is same as that of characteristic function of normal distribution with mean 0 and variance $1/9$

We know that characteristic function of a normally distributed random variable with mean 0 & variance σ is

$$\phi(u) = e^{-\frac{u^2}{2}\cdot\sigma}$$

comparing it with eq^n (1)

$$e^{-\frac{u^2}{18}} = e^{-\frac{u^2}{2}\cdot\sigma}$$

$$\therefore \frac{\sigma}{2} = \frac{1}{18} \qquad \therefore \sigma = \frac{1}{9}$$

$\therefore$ The sequence $\frac{S_n}{n^{3/2}}$ converges in distribution as n goes to infinity to a normal distribution with mean 0 & variance $\frac{1}{9}$.