Edward Moore

ECE 3522: Stochastic Processes in Signals and Systems

Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA 19122

# Problem Statement

Two sets of data are to be analyzed using a window and frame analysis technique of varying sizes to find the mean and variance. The two sets of data are Google’s stock price since inception and an audio file sampled at 8kHz. The results of these analyses are then compared to some global statistics on these data sets such as minimum and maximum values, mean, median and variance. This type of analysis should give some good insight to the short, and more importantly, long term behavior.

# Approach and Results

Finding global variables is important when analyzing data but it doesn’t give much perspective as to its behavior over time. In order to do this we break up the data into smaller chunks and perform the same analyses. Looking at how these values change over time gives a better idea of how the data is trending.

Since Googles stock inception there have been 2616 business days as of this paper meaning we have that many points of data. The plot of this data is shown below in figure 1. Breaking this up into windows of 7 and 31 and frames of 1, 7, 14 and 30 should give a good indication how the stock has behaved week to week and month to month. Figures 2 and 3 below give the mean and variance calculated over these window and frame sizes.



Figure 1

Min = 49.95

Max = 609.47

Mean = 286.74

Median = 264.83

Variance = 1.62\*104

**Google Stock Means: Window, Frame Analysis**

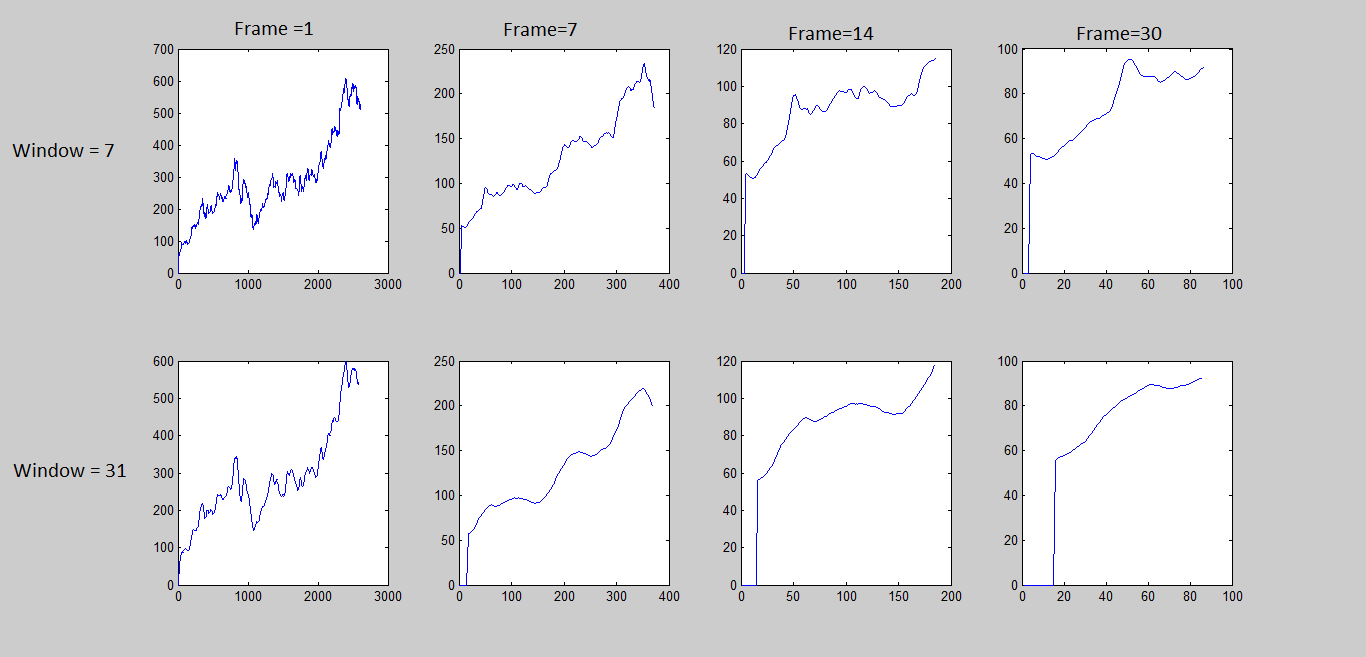


Figure 2

**Google Stock Variances: Window, Frame Analysis**

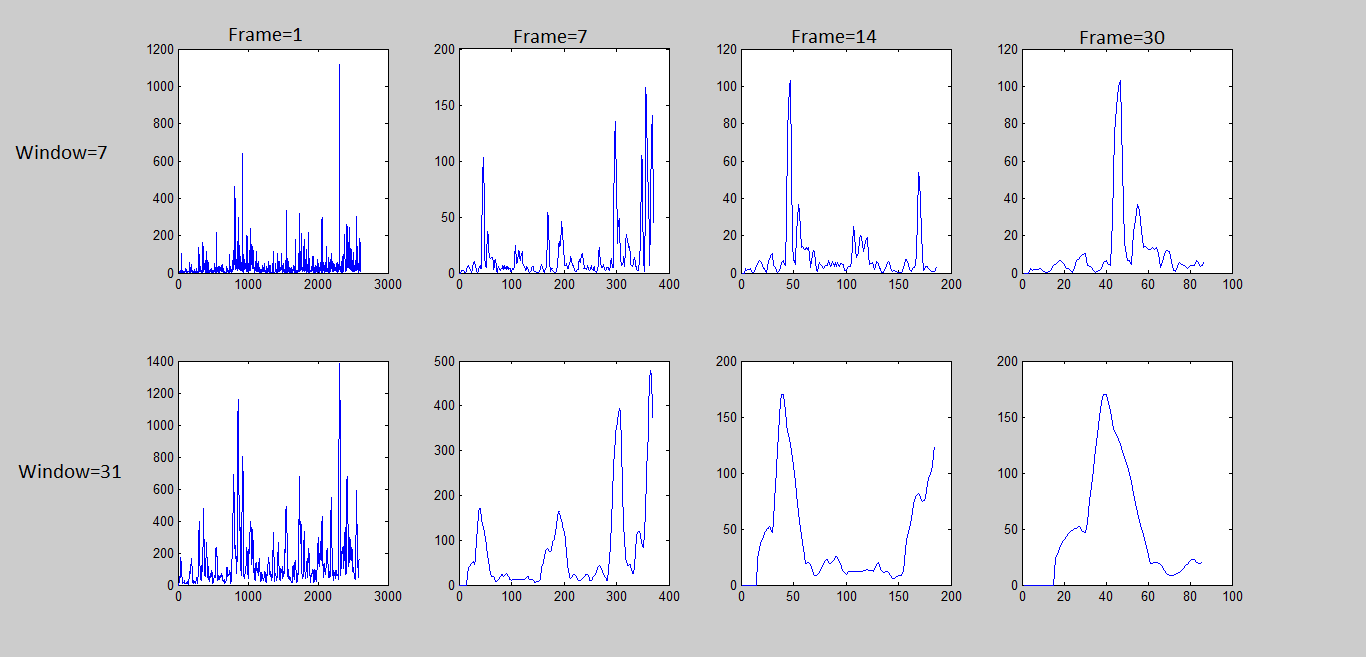


Figure 3

There are a number of things we can tell from the graphs of mean and variance. In terms of prediction we can tell from the graphs with smaller frame sizes that we can expect in the short term that the stock price will most likely drop since the mean and variance are experiencing a downward trend. Conversely, when looking at graphs with bigger frame sizes we can see that the long term trend shows the price is increasing. When predicting stock prices what’s going to happen over the long term is clearly what’s most important so it seems that it would be a safe bet that while the stock might drop for the next few days, over the coming months the price is likely to increase overall. This is illustrated by the global variance and in the variance plots. The global variance and corresponding variance graphs are what should be expected since dealing with stocks which often vary greatly.

After converting the audio file to WAV format and loading it into Matlab, an array of 82425 data points was created. The absolute value must be taken before analyzing the data. After doing this the data was then plotted. This is shown in figure 4 below. The same type of window, frame analysis performed earlier is then used again. Results of these analyses are shown below in figures 5 and 6.



Figure 4

Min = 0

Max = .4576

Mean = .0412

Median = .0249

Variance = .0022

**Audio Data Means: Window, Frame Analysis**

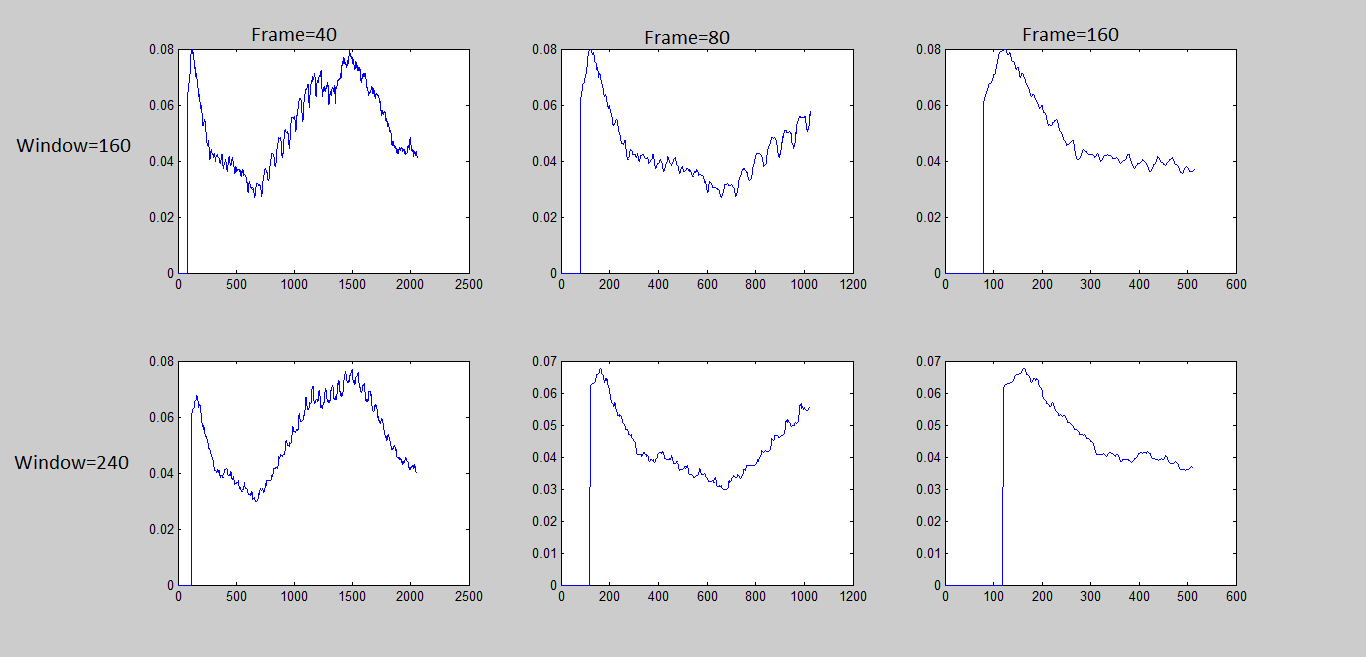


Figure 5

**Audio Data Variances: Window, Frame Analysis**

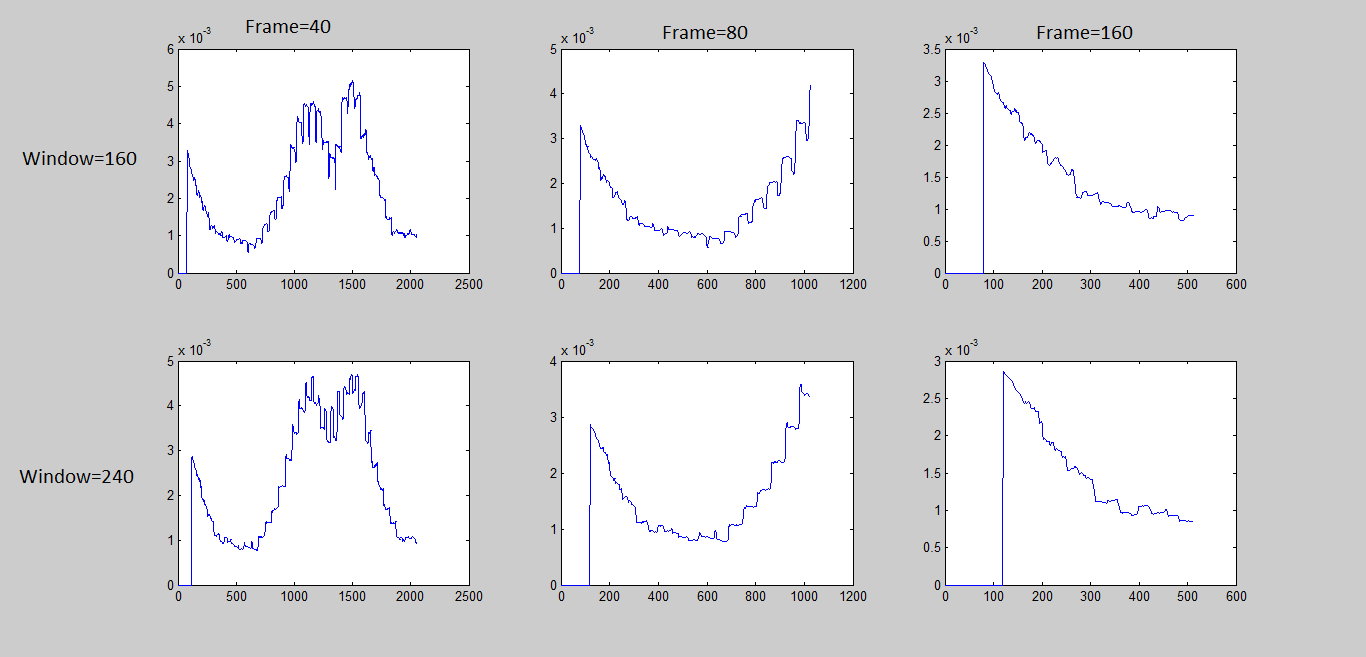


Figure 6

Since the data set is much larger this time, larger frame and window sizes needed to be used. The frames sizes used corresponded to 5, 10 and 20msec and the windows to 20 and 30msec. This gives us a good idea how the data is behaving over these time frames. From global variance we expect the graph to be fairly consistent with most values not deviating far from the mean. This makes sense since we have an audio file, the output (volume) is intuitively all around the same level.

# MATLAB Code

**Google Stock Price**

filename = 'google\_v00.xlsx';

close = xlsread(filename,'E:E');

%2616 array length

a = sort(close,'ascend'); %gives min value

x = a(1)

d = sort(close, 'descend'); %gives max value

y = d(1)

m = mean(close)

med = median(close)

v = var(close)

figure(1);

plot(close)

title('Close Values')

M = [1 7 14 30]; %(days)

N = [7 31]; %(days)

figure(2);

l = 1; %counter for subplots

for j = 1:2 %loops through each window size

for k = 1:4 %loops through each frame size

startPt = (N(j) + 1)/ 2; %picks a start point by moving over half the size of a window, avoids zero stuffing the ends

newLength = floor((2616 - (N(j)-1))/M(k)); %calculates the array length of means based on window & frame size

windowMean = zeros(1, newLength); %creates an empty array to hold the means

for i = startPt:newLength

windowMean(i) = mean(close( (i-((N(j)-1)/2)): (i+((N(j)-1)/2))));%finds the mean over the specified window

end

subplot(2, 4, l)

l = l+1;

plot(windowMean)

end

end

figure(3);

l = 1; %counter for the subplots

for j = 1:2 %loops through each window size

for k = 1:4 %loops through each frame size

startPt = (N(j) + 1)/ 2; %picks a start point by moving over half the size of a window, avoids zero stuffing the ends

newLength = floor((2616 - (N(j)-1))/M(k));%calculates the array length of variances based on window & frame size

windowVar = zeros(1, newLength); %%creates an empty array to hold the variances

for i = startPt:newLength

windowVar(i) = var(close((i-((N(j)-1)/2)): (i+((N(j)-1)/2))));%finds the variance over the specified window

end

subplot(2, 4, l)

l = l+1;

plot(windowVar)

end

end

**Audio Data**

[y, Fs] = audioread('rec\_01\_speech.wav');

y = abs(y);

%82425 array length

a = sort(y,'ascend'); %gives min value

x = a(1)

d = sort(y, 'descend'); %gives max value

z = d(1)

m = mean(y)

med = median(y)

v = var(y)

figure(1);

plot(y)

title('Magnitude of Audio Signal')

M = [40 80 160]; %40 (5 msec), 80 (10 msec), 160 (20 msec);

N = [160 240]; %160 (20 msec), 240 (30 msec)

figure(2);

l = 1; %counter for subplots

for j = 1:2 %loops through each window size

for k = 1:3 %loops through each frame size

startPt = (N(j))/ 2; %picks a start point by moving over half the size of a window, avoids zero stuffing the ends

newLength = floor((82425 - (N(j)-1))/M(k)); %calculates the array length of means based on window & frame size

windowMean = zeros(1, newLength); %creates an empty array to hold the means

for i = startPt:newLength

windowMean(i) = mean(y( (i-((N(j)-1)/2)): (i+((N(j)-1)/2))));%finds the mean over the specified window

end

subplot(2, 3, l)

l = l+1;

plot(windowMean)

end

end

figure(3);

l = 1; %counter for subplots

for j = 1:2 %loops through each window size

for k = 1:3 %loops through each frame size

startPt = (N(j))/ 2; %picks a start point by moving over half the size of a window, avoids zero stuffing the ends

newLength = floor((82425 - (N(j)-1))/M(k)); %calculates the array length of variances based on window & frame size

windowVar = zeros(1, newLength); %creates an empty array to hold the variances

for i = startPt:newLength

windowVar(i) = var(y((i-((N(j)-1)/2)): (i+((N(j)-1)/2))));%finds the variance over the specified window

end

subplot(2, 3, l)

l = l+1;

plot(windowVar)

end

end

# Conclusions

We can only get so much out of global statistics. In order to determine short and long term trends other analyses must be used. In this assignment, a series of window and frame sizes were defined and used to determine these behaviors. Short term behaviors are dictated by the response of small frame sizes while larger sizes represent longer term behaviors. The mean can be used to determine which way a value is trending while variance gives an idea of how stable the systems behavior is.