Anton LeKang

ECE 3522: Stochastic Processes in Signals and Systems

Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA 1912

CA5

# Problem Statement

The objective of this computer assignment is to learn the basics of covariance and correlation coefficient. The first part of the assignment looks at the correlation coefficient of an audio file. The coefficient correlation gives a prediction of value of one variable based on another variable. This problem looks at the relationship between X and Y, which compares the direction, plus the degree in how the variable move. The second part of the computer assignment looks at covariance. The covariance is a simpler form of the coefficient and provides how two variables are related. Covariance will only state positive or negative for relativity. This computer assignment will use an audio file as the main inputted source of the data. MATLAB will be used to compute the covariance and correlation.

# Approach and Results

Part 1:Plot the Correlation Coefficient

The first part of the computer assignment shows two plots of the correlation coefficient of an audio file. In the first Figure(1) it shows the correlation coefficient when the X variable starts at .9secs. The program stores the first 240 samples starting at .9secs. The Y variable is the next 240 samples after the X variable. The correlation coefficient provides a measure of how closely two variables are related. The Y-axis shows that the scale for correlation is between -1 and 1. If the point is a 1, then the variables have a perfect positive correlation. This means that not only are they positive, but they are move at the same proportional rate or are linear. If the values is -1 then the variables are perfectly negative. This means that both of the variables are negative, and they are moving linearly. It one variable moves a degree, and then the other variable moves at that same negative degree. If the value is 0, then that means there is no correlation between the two variables. In the figure below the highest positive number is around .8, and the lowest negative number is around -.5. When there are the peaks the variables have more linearity than when the value is flat.

Figure 1 – Correlation Coefficient starting at .9 Sec

The figure(3) below shows the correlation coefficient of the audio file starting at 3secs. The X variable takes the first 240 samples starting at 3secs, and the Y variable takes the next 240 samples after that. This figure has a lot lower values than the figure above. That means these variables are not as linear. The highest positive value is around .3, while the highest negative value is around -.4. When the numbers are closer to 0 the linearity gets smaller and smaller.

Figure 2 – Correlation starting at 3secs

Part 2: Covariance

This part of the lab looks at the covariance of two variables. Just like the correlation this calculation gives an indication on how close two variables are related. This calculation with only give whether the variables are positively related or negatively. It not gives to a specific degree from -1 to 1 like the correlation. The part looked at the covariance of the first 16 samples starting at .9 secs, 1.1secs, and 3secs. Below is one of the 16 by 16 matrix, followed by the figures of each sample.

Figure 3 – Matrix Covariance

The figure above shows the first four columns of the 16 by 16 matrix. Each number is either positive or negative and that relates to the correlation between the two data values. If the number is positive then the variables have a positively related, and if the numbers are negative then they are negatively related.

Figure 4 – Covariance at .9secs

Figure 5 – Covariance at 1.1sec

Figure 6 – Covariance at 3 secs

These three figures above show the covariance for the time values .9secs, 1.1secs, and 3secs. Each of the figures have 16 lines, and 16 points within each line for the different covariance for that row. The covariance takes its two variables and compares them either positively or negatively. The matrix with the first four columns relates to the covariance in figure 4. For the first column the largest number is .9826E5 which is represented by the green line at the top, the smallest number is-1.51E6 which represented by the orange line. The only thing that these values mean is if the relationship is positive or negative, it does not give a degree to how linear the two variables are.

# MATLAB Code

Part 1:

%Rachel King and Anton Lekang

%Computer Assignment 5

function CA\_5

fp = fopen('rec\_01\_speech.raw','r');

sig = fread(fp,inf,'int16');

[z,q] = size(sig);

%length of parsed file

p = 240;

%index starting at 0.9 sec = 900 msec

r = 7200;

%index starting at 3 sec = 3000 msec

r1 = 24,000;

x = zeros(p);

x1 = zeros(p);

x = sig(r:r+p);

x1 = sig(r1:r1+p);

y = zeros(p);

for k = 1:512

 y = sig(r+k:r+p+k);

 y1 = sig(r1+k:r1+p+k);

 C(k) = corr(x, y);

 C1(k) = corr(x1, y);

end

figure(1)

plot(C)

ylabel('Amplitude')

xlabel('Time (sec)')

title('Correlation Coefficient starting at .9 Sec')

figure(2)

plot(C1)

title('Correlation Coefficient starting at 3 Sec')

ylabel('Amplitude')

xlabel('Time (sec)')

end

Part 2: Covariance

%Rachel King and Anton Lekang

%Computer Assignment 5

function CA\_5

fp = fopen('rec\_01\_speech.raw','r');

sig = fread(fp,inf,'int16');

[z,q] = size(sig);

%length of parsed file

p = 16;

%index starting at 0.9 sec = 900 msec

q1 = 900; %starting point, in msec

r1 = q1\*8;

%starting point, 1100 msec

q2 = 1100;

r2 = q2\*8;

%starting point, 3000 msec

q3 = 3000;

r3 = q3\*8;

%parse signals, length 16

x1 = zeros(p);

x1 = sig(r1:r1+p);

x2 = zeros(p);

x2 = sig(r2:r2+p);

x3 = zeros(p);

x3 = sig(r3:r3+p);

N = 16;

for i = 0:15

 for j = 0:15

 C1(i+1, j+1) = (1/(N-1))\*sum(x1(N-i)\*x1(N-j));

 C2(i+1, j+1) = (1/(N-1))\*sum(x2(N-i)\*x2(N-j));

 C3(i+1, j+1) = (1/(N-1))\*sum(x3(N-i)\*x3(N-j));

 end

end

C1

figure(1)

plot(C1)

figure(2)

plot(C2)

figure(3)

plot(C3)

end

These are the scripts for each part of the computer assignment. In the first part of the assignment the X values start at .9sec for 240 variables, and then the Y values take the next 240 variables. The correlation instruction is used, and then they are each individually plotted. In part 2, the covariance is derived from using the three different time samples of .9sec, 1.1secs, and 3secs. The X value is used to compare between the covariance at 240 samples from a 16 by 16 matrix. The calculations for covariance are done at C1, C2, and C3.

# Conclusions

Calculating the correlation and covariance of two variables can be a very valuable tool when looking at data. Both of these analyses provide a way to compare variables and to see they are related to one another. The correlation is computed first in this computer assignment and it is a very useful. Not only does it tell how two variables X and Y are related to each other positively or negatively, it also give the degree of value. The range is between -1 and 1, and if the value is 1 then they have perfect positive linearity, and if it is equaled to -1 then the variables perfect negative linearity. The covariance can also relate variables to positive and negative, but it does not give a degree of value. This can be a very useful tool especially in statistics when looking at the economic growth of a company. These formulas will help tell if the company is performing well, and the relationship between two variables, like growth and days. This was a very knowledgably computer assignment, and I like how it correlated with what we are learning in class.