Anton LeKang

ECE 3522: Stochastic Processes in Signals and Systems

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

CA10

# Problem Statement

The objective of this computer assignment is to look at a hypothesis testing problem, and to incorporate the ideas of testing with what has been learned from class. Although this problem comes straight from the book, there are many useful parts that can be developed and observed. Another key objective of this assignment is to build multiple functions that can be called upon at any time. Functions make programming simple, and users should always want to perform the simplest code. First a Gaussian distribution is made to observe random variables with the increase in variables size. Then a function is created to cheek the significance of the random variables compared to the known statistics given from the problem. This is done from incorporating the Z confidence interval. The next part of the computer assignment determines the best number of variables that complies with the desired significance level. Based on the level of significance for the tailed test, the number of random variables will influence the results. This computer assignment will use multiple functions, and different variables sizes in MATLAB to observe the effects on hypothesis testing.

# Approach and Results

Null Hypothesis H0: Lightbulbs will last for 1600 hours

Alternate Hypothesis H1: Lightbulbs will not last for 1600 houra

Part 1: Create an n Gaussian distributed random variables



Figure 1 – Gaussian When there was 10 data points

This is the first plot of the Gaussian and it is very noisy because there are not a lot of data points.

Figure 2 – Gaussian with 1000 data points

Figure 2 – Gaussian with 100 data points

Notice how figures 2 and 3 look very different than figure 1. This is because 1 has a very limited amount of data variables. Figure 3 is the best interpretation of the Gaussian and a very nice looking plot has been made. This figure looks the most Gaussian compared to the other two plots.

Part 2 and 4: Look at the confidence levels of 0.8, 0.9, 0.95, 0.99, 0.999

When looking at the hypothesis test, each confidence level was looked at individually for each number of random variables. I learned my lesson that this was extremely high on the time consuming portion, and will not happen again.

Figure 3 - Confidence test for n = 10, std = 120, and levels at 0.8, 0.9, 0.95, 0.99, 0.999

Figure 3 represents the first set of interval testing which was done with the n value set to 10, and the standard deviation set to 120. Notice how each of the status values say 0, this means that the hypothesis test did not pass, and the intervals were outside the confidence levels for that value. This means that we would reject the null hypotheses because the lightbulbs did not perform to the expected statistics.

Figure 4- Confidence test for n = 100, std = 120, and levels at 0.8, 0.9, 0.95, 0.99, 0.999

Figure 4 represents the next set of hypotheses testing for when the data size is set to 100, and the standard deviation is left the same at 120. Notice how in the first two test at .8, and .9, the hypotheses was accepted. This means that the outputted values fell within the significance level of the test. When larger than 0.9, the test failed, and the null hypothesis was rejected.

Figure 5 - Confidence test for n = 1000, std = 120, and levels at 0.8, 0.9, 0.95, 0.99, 0.999

Figure 5 shows the hypothesis test for when the sample size was set to 1000. All of the level passed and the null was accepted at each level of significance. This happens because as the data values increase, they get closer and closer to the actual mean. More variables, the better the results will be.

Standard Deviation is now set at 240 hours!

Figure 6 - Confidence test for n = 10, std = 240, and levels at 0.8, 0.9, 0.95, 0.99, 0.999

This figure shows the first hypothesis set of tests where the number of samples is at 10, and the standard deviation is set to 240. Again just like figure 3 there are not enough data values, and the null hypothesis must be rejected.

Figure 7 - Confidence test for n = 100, std = 240, and levels at 0.8, 0.9, 0.95, 0.99, 0.999

This figure shows the second hypothesis test where the number of data points increases from 10, to 100. Notice how each of the confidence levels produce a 1, so the null hypothesis is accepted, and the data values fall within the given parameters of the lightbulbs.

Figure 8 shows the last hypothesis test in which the standard deviation was 240, and the number of samples was 1000. This test also accepted all of the null hypotheses because the number of points was so large.

Figure 8 - Confidence test for n = 1000, std = 240, and levels at 0.8, 0.9, 0.95, 0.99, 0.999

Part 3: Looking at the minimum value for n

As seen in each of the figures from part 2 and 4, there was an N value associated with the output. That N values is the values in which the difference would be statistically significant. This values is found by knowing the confidence level, and standard deviation. It is the same proves as finding whether or not the hypotheses will be accepted, but this time the objective is to solve for N to make the variable statistical significant. It makes sense that as the confidence level increase, there needs to be a larger N value to impact the significance of the test.

Part 5: Maximum Value of Variance

This part of the assignment is to find which variable of standard deviation has a larger impact on the significance of the test. As seen from the results in part 2 and 5, when the standard deviation is larger more values of the null can be accepted. This is because on the known rules of standard deviation with respect to a Gaussian curve. One of the most basic rules of statistics is the 68, 95, 99.7 rule. This means that within the first standard deviation away from the mean 68% of the variables fall within that value. So 34% fall above the mean, and 34% fall below the main. This is seen when the standard deviation is set to 240 and more values are accepted compared to the null hypotheses.

# MATLAB Code

Part 1:

function CA\_10

mean1 = 1600

mean2 = 1570

stdev = 240

n = 100

r = gen(mean1, stdev, n);

mu = mean(r);

%sets the first level of significance

i = .002

 confidence = find\_confidence(i);

 status = check\_significance(mu, mean2, stdev, n, confidence)

 N = determine\_significance(confidence, stdev, i)

% status = check\_significance(mean1, mean2, stdev, n, confidence)

%

% N = determine\_significance(confidence, stdev, level\_sig)

%

end

% function Y = gen\_grv(mean, stdev, n)

%

% p1 = -.5 \* ((D - mean)/stdev) .^ 2;

% p2 = (stdev \* sqrt(2\*pi));

% Y = exp(p1) ./ p2;

%

% end

%confidence level

function confidence = find\_confidence(i);

if i == .002

 confidence = [-3.89 3.89]

 else if i == .005

 confidence = [-3.29 3.29]

 else if i == .01

 confidence = [-2.58 2.58]

 else if i == .05

 confidence = [-1.96 1.96]

 else if i == .1

 confidence = [-1.64 1.64]

 end

 end

 end

 end

end

end

%part 3

function status = check\_significance(mu, mean2, stdev, n, confidence)

status1 = (mu-mean2)/(stdev\*(sqrt(n)))

status2 = (mu+mean2)/(stdev\*(sqrt(n)))

if status1 < confidence(1) || status2 > confidence(2)

 status = 0

else status = 1

end

end

%part 4

function N = determine\_significance(confidence, stdev, i)

cl = find\_cl(i);

N = ((confidence(2)\*stdev)/(1 - cl))^2

end

function cl = find\_cl(i)

if i == .002

 cl = .8

 else if i == .005

 cl = .9

 else if i == .01

 cl = .95

 else if i == .05

 cl = .99

 else if i == .1

 cl = .999

 end

 end

 end

 end

end

end

function r = gen(mn, stdev, n)

r = normrnd(mn,stdev, [1, n]);

plot(r)

end

This script is extremely long, but is pretty straight forward to follow. There are basically four main functions to this program and they are: r = gen(mean1, stdev, n); confidence=find\_confidence(i);status =check\_significance(mu, mean2, stdev, n, confidence) and N = determine\_significance(confidence, stdev, i). The first function creates the Gaussian of the data with the specific number of random variables for the known mean and standard devotion. The second function finds the K value of the significance based on the significance level. The next function determines if the significance falls with within that significance level. This is the part where the program either accepts or reject the hypothesis. The last function looks at the confidence, and the significance to determine the n value for the value that would be statistically significant. These functions are called multiple time and they are used to help speed of the process of codding.

# Conclusions

The whole purpose of this computer assignments was to observe a scenario with hypothesis testing. Hypothesis testing is very useful in the real world because it can be done to see if a product really preforms on how it was advertised. As seen from this assignment there are many variables which can change the outcome of a hypothesis test. In this test, the company claims that a lightbulb can last for 1600 hours, and this is the null hypothesis. The alternate hypothesis is that the lightbulb will either fall shorter, or longer than the advertised time. In the first part of this lab a Gaussian function is created, and observed with different number of inputs. As the number of inputted variables increase, the Gaussian gets more and more like a normal Gaussian curve. These random numbers from the Gaussian are used in the hypothesis test for this computer assignment. Next a hypothesis test is done for the number of samples set to 10, and it is observed with the significance at 0.8, 0.9, 0.95, 0.99, and 0.999, those number will be used to find the K values in the significance test. As it turns out, the null hypothesis is rejected when more often when the number of samples is low. All of the hypothesis are reject when the sample size is 10, and all are accepted when the sample size is 100. This shows the importance of having a larger number or variables and not testing a small sample. Next the standard deviation is increased to 240 for the test. This has a major impact on the number of test accepted because the standard deviation is doubled. This is seen in the second test where more values are accepted than when the standard deviation was set to 120. The standard deviation is extremely important, especially in hypothesis testing. Just from these two examples it is seen that based on the standard deviation, the hypotheses for the mean is significantly changed. This is due to the 68, 95, and 99.7 rule which was talked about in part 5 of this activity. As the standard deviation increased, the range of values that are significant become larger and larger. Thus the curve of the Gaussian also becomes wider and wider. Yes more values will be accepted, but these values could be farther off from the mean. This computer assignment was extremely valuable in combining the topics of Gaussian, intervals, and significance with hypothesis testing. This assignment has shown how hypothesis is important especially with real life examples like the advertised use of the lightbulb.