Rachel King

ECE 3522: Stochastic Processes in Signals and Systems

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

# Problem Statement

This computer assignment deals with hypothesis testing and the calculations that come along with it. Given an example from our text, we are asked to analyze it using MATLAB as a tool. We will be checking the significance between the estimated and the actual mean, finding the number of computations necessary, and finally cycling through many variances to find the maximum value for which the difference in the mean is significant. Ultimately, the use of MATLAB function to organize the code will help us to build tools that will solve a hypothesis testing problem in an accurate, time efficient fashion given multiple inputs

# Approach and Results

1. *Generate* n *random numbers with a mean of 1600 hours and a standard deviation of 120 hours for* n = 10, 100, 1000.

In main:

for m = [10, 100, 1000]

 R = gen\_grv(1600, 120, m)

 figure(l)

 hist(R, m)

 l = l+1;

end

In my Gaussian distributed RV generator:

%generates n Gausian distributed RV's

function R = gen\_grv(mn, stdev, n)

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

end



Figure 1 - Gaussian RV's, n = 10



Figure 2 - Gaussian RV, n = 100



Figure 3 - Gaussian RV's, n = 1000

1. *Using your tools above, determine if the difference in the mean is significant for confidence levels of* 0.8, 0.9, 0.95, 0.99, 0.999.

In main:

j = 1;

for i = [0.8, 0.90, 0.95, 0.99, 0.999, 0.9999]

 status(j) = check\_sig(mu, mn, stdev, n, i);

 j = j+1;

end

In my status checking function

function status = check\_sig(mn1, mn2, stdev, n, CL)

%mn1 = estimated(hrs), mn2 = actual mean(mn)

k = determine\_k(CL);

%judge status of the test

min = mn2 - k\*stdev/sqrt(n);

max = mn2 + k\*stdev/sqrt(n);

if mn1<max && mn1>min

 status = 1;

else status = 0;

end

end

In my determine k function:

%equivilent of reading from k-value chart,

%but in code form!

function k = determine\_k(CL)

switch CL

 case 0.8

 k = 1.28;

 case 0.90

 k = 1.64;

 case 0.95

 k = 1.96;

 case 0.99

 k = 2.58;

 case 0.999

 k = 3.29;

 otherwise

 k = 3.89;

end

end



Figure 4 - Results of Question 2

1. *Determine the minimum value of n for which these differences would be statistically significant.*

In main:

j = 1;

for i = [0.8, 0.90, 0.95, 0.99, 0.999, 0.9999]

% status(j) = check\_sig(mu, mn, stdev, n, i);

 N(j) = det\_sig(mu, mn, stdev, i);

 j = j+1;

end

In my determine significance function:

%determines the value of N so that the difference

%between the two means are statistically significant

function N = det\_sig(mn1, mn2, stdev, CL)

k = determine\_k(CL)

N = ( (k\*stdev)/(1 - CL) )^2;

end



Figure 5 - Number of computations needed per confidence level (in order)

1. *Repeat no. 2 for a standard deviation of 240 hours.*



Figure 6 - Status results, variance 240

1. *Find the maximum value of the variance for which the difference in the mean is significant for* [n = 10, 100, 1000] *and* [confidence = 0.8, 0.9, 0.95, 0.99, 0.999].

In main:

n = [10, 100, 1000];

CL = [0.8, 0.9, 0.95, 0.99, 0.999];

for i = 1:length(n)

 for j = 1:length(CL)

 X(i, j) = max\_stdev(mu, mn, n(i), CL(j));

 end

end

 In my function ‘max\_stdev’

%finds the maximum value of variance for which the

%difference in mean is significant

function S = max\_stdev(mn1, mn2, n, CL)

s = 1:10:10e6;

for i = length(s):-1:1;

 status = check\_sig(mn1, mn2, i, n, CL);

 if status == 1

 S = s(i);

 break

 end

end

end



Figure 7 - Results, standard deviation testing

# Conclusions

This assignment was a test of not only our programming skills, but also our analytical skills; it required very good knowledge of hypothesis testing. When generating the normally distributed random variables, the increase in the number of variables is what gave us the figure that most closely resembled the Gaussian curve that we’ve grown to know (figures 1-3). When we want to determine whether or not a mean is significant, we first need to find the range that it passes. To do that, we need to get the value of the constant, k, from the significance level, which I organized into a separate function. From that, we can find our min and max range using the actual mean which is passed to the function, and seeing is the estimated mean falls within that function. Calculating the number of computations necessary to make the values significant, we simply use the same formula that we did in one of our quizzes. The higher the significance level, I found, the higher number of computations necessary to meet the requirement. Increasing the standard deviation resulted in it being easier for the means to fall into the range specified and be statistically significant. Finally, when asked to find the max standard deviation for each value of n and the confidence level, my results were obviously not right; all combinations resulted in the same number. However, I think my process was; I checked the status, and then if it was one, I saved that variance value and broke from the loop. Overall, this computer assignment was a good exercise in practicing hypothesis testing.