[Tyler Olivieri]

ECE 3512: Stochastic Processing in Signals and Systems

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

# Problem Statement

A lightbulb company proposes that their lightbulbs have a specific mean and standard deviation. They claim to have improved their lightbulbs. Joe tests this hypothesis by selecting 100 lightbulbs and finding their mean and standard deviation. Is the difference statistically significant?

# Approach and Results



Figure : standard dev 120

In Figure 1, it can be seen that there are a few significantly statistical differences. When a status is “1” it is statistically different. It is also important to note that there were different sets of RV generated for each n. When minN10 is less than 10, we see a significant difference. Similarly with minN1000. If it is less than 100 it is a statistically significant difference. This is because minN10 computes the minimum number of n to be statistically different. If it less than 10, it will be statistically different because 10 numbers were generated.



Figure : stdev = 240

In Figure 2, the standard deviation was changed to 240. Similar results can be seen before. On this particular simulation we can see status1000 is flagged and minN1000 has a number less than 1000 in the first position of the vector.



Figure : snippet of status for range of stdev

In Figure 3, a range of stdev can be seen. Since a new set of RV were generated each time the standard deviation changes the results were scattered. Some significantly significant differences were seen. The minimum N is less than the amount generated at the same indexes. Even as I went out to 1000 standard deviation it was mostly not significant differences.

# MATLAB Code

In function gen\_grv two Gaussian distributed RV are created with an input of the mean and standard deviation.

The function determine\_significance determines if two data sets with two means with a particular standard deviation are significantly different or not. It takes as an input the level of significance. It returns “1” if significantly different.

The function check\_significance outputs the minimum n for the two sets of data to be statistically different. It takes an input of the two means, standard deviation, and the level of significance.

In the main script, random vectors were generated, estimates of the mean were taken, and the functions were called to determine if the results were statistically different and what the minimum number of n is to make the two results statistically different.

Then the same procedure was used in a for loop that looped over standard deviation from 500:1:1000.

%Tyler Olivieri
%function gen\_grv
%inputs:
%u -mean-
%stdev - standard deviation
%n- number of gaussian distributed rv
%box-muller method
function x = gen\_grv(u,stdev,n)

u1 = rand(n,1);
u2 = rand(n,1);
%sqrt(-2\*log(u1)).\*cos(2\*pi\*u2) generates between [0 1]
x = u + stdev\*(sqrt(-2\*log(u1)).\*cos(2\*pi\*u2));

%function determine significance
%determines number of samples for certain confidence level
%inputs:
%mean1
%mean2
%stdev
%confidence in percentage/100----limited confidence inputs

function N= determine\_significance(mean1,mean2,stdev,significance,n)

switch significance
 case .001
 z\_low = 3.08;
 z\_high = 3.08;
 case .01
 z\_low = 2.58;
 z\_high = 2.58;
 case .05
 z\_low = 1.96;
 z\_high = 1.96;
 case .1
 z\_low = 1.64;
 z\_high = 1.64;
 otherwise
 disp('Invalid significance. Its a limited function man..');
end

z = (mean1-mean2)/(stdev/sqrt(n));
if (z\_low <= z && z<= z\_high)
 N = 0;
 flag = 1;
else
 flag = 0;
end
if(flag ==0)
 N = ((z\_high)\*stdev/(abs(mean1-mean2)))^2;
end

%Tyler Olivieri
%function check\_significance
%inputs:
%mean1
%mean2
%stdev
%n
%significance
%status = 1 is significant difference
%Two-tailed test

function status = check\_significance(mean1,mean2,stdev,n,significance)

%significance is compliment of confidence

% calculate z score
z = (mean1-mean2)/(stdev/sqrt(n));

%z score from significance level
%significance = .05;
switch significance
 case .001
 za = 3.08;
 case .01
 za = 2.58;
 case .05
 za = 1.96;
 case .1
 za = 1.64;
 otherwise
 disp('Invalid significance. Its a limited function man..');
end

%test significance
if(abs(z)>abs(za))
 status = 1;
else
 status = 0;
end

%%CA 10
% Tyler Olivieri
clc;clear;
%inputs/givens
mean1 = 1600;
stdev = 120;

% changing n
index = 0;
x10 = gen\_grv(mean1,stdev,10);
x100 = gen\_grv(mean1,stdev,100);
x1000 = gen\_grv(mean1,stdev, 1000);

mean\_estimate10 = sum(x10)/length(x10);
mean\_estimate100 = sum(x100)/length(x100);
mean\_estimate1000 = sum(x1000)/length(x1000);

for significance = [.1 .05 .01 .001]
 index = index +1;
 status10(index) = check\_significance(mean1,mean\_estimate10,stdev,10,significance);
 status100(index) = check\_significance(mean1,mean\_estimate100,stdev,100,significance);
 status1000(index) = check\_significance(mean1,mean\_estimate1000,stdev,1000,significance);

 minN10(index) = determine\_significance(mean1,mean\_estimate10,stdev,significance,10);
 minN100(index) = determine\_significance(mean1,mean\_estimate100,stdev,significance,100);
 minN1000(index) = determine\_significance(mean1,mean\_estimate1000,stdev,significance,1000);
end

clc;clear;
mean1 = 1600;
stdev = [500:1000];
n = [10 100 1000];

for i = 1:length(stdev)
index = 0;
x10 = gen\_grv(mean1,stdev(i),10);
x100 = gen\_grv(mean1,stdev(i),100);
x1000 = gen\_grv(mean1,stdev(i), 1000);

mean\_estimate10 = sum(x10)/length(x10);
mean\_estimate100 = sum(x100)/length(x100);
mean\_estimate1000 = sum(x1000)/length(x1000);
for significance = [.1 .05 .01 .001]
 index = index +1;
 status10(index,i) = check\_significance(mean1,mean\_estimate10,stdev(i),10,significance);
 status100(index,i) = check\_significance(mean1,mean\_estimate100,stdev(i),100,significance);
 status1000(index,i) = check\_significance(mean1,mean\_estimate1000,stdev(i),1000,significance);

 minN10(index,i) = determine\_significance(mean1,mean\_estimate10,stdev(i),significance,10);
 minN100(index,i) = determine\_significance(mean1,mean\_estimate100,stdev(i),significance,100);
 minN1000(index,i) = determine\_significance(mean1,mean\_estimate1000,stdev(i),significance,1000);
end
end

# Conclusions

The results showed that there could be both statistically significant differences and non-significant differences. It really depends on the particular data set that is generated. Since it is random, our results changed every time the simulation was run. The data made sense each time when comparing its status to the minimum number of n required to make a significant difference. I thought the data would show more significant differences the higher the standard deviation but that didn’t seem to be the case. The mean never varied too much especially with the larger data sets. This may be why not many significant differences were seen. Overall, the simulations seemed to be successful and showed that different samples of a data set can show a different level of significance. Having more samples always helps to make better estimates of the mean. Also, having a higher confidence also shows less significant differences.