Harish Boggarapu

ECE 3522: Stochastic Processes in Signals and Systems

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

# Problem Statement

In this assignment we are asked to generate n Gaussian distributed random numbers and find the mean of the data generated. We also have to find if the true mean and the estimated mean of the generated data are statically significant based on the n random numbers and the confidence level. We have to determine the value of n for the difference between the two means to be statistically significant.

1. Approach and Results

**1.**

For the first task, I created a function, which generates Gaussian distributed random numbers based on mean, standard deviation, and n numbers. The mean of the data generated for n = 10, mean = 1600 and stdev = 120.

 

For n = 100



 For n = 1000

 

**2.**

For n =100, with estimated mean equal to 1577.4, at confidence level of 80%, the difference is **statistically significant**.

At 90%, the difference is **statistically significant**.

At 95%, the difference is **not statistically significant**.

At 99%, the difference is **not statistically significant**.

At 99.9%, the difference is **not statistically significant**.

At 99.99%, the difference is **not statistically significant**.

**3.**

For confidence level of 80%, the minimum value of n had to be to **47** for the difference to be statistically significant.

For 90%, n has to be **76**.

For 95%, n has to be **109**.

For 99%, n has to be **188**.

For 99.9%, n has to be **302**.

For 99.99%, n has to be **427**.

**4.**

For standard deviation of 240

For n =100, with estimated mean equal to 1587.3, at confidence level of 80%, the difference is **not** **statistically significant**.

At 90%, the difference is **not statistically significant**.

At 95%, the difference is **not statistically significant**.

At 99%, the difference is **not statistically significant**.

At 99.9%, the difference is **not statistically significant**.

At 99.99%, the difference is **not statistically significant**.

**5.**

For n = 10 with estimated mean of 1561.2

The maximum value of variance at 80% is **9.1885e+03**

At 90%, value is **5.5973e+03**

At 95%, value is **3.9188e+03**

At 99%, value is **2.2616e+03**

At 99.9%, value is **1.3908e+03**

At 99.99%, value is **994.8652**

For n = 100 with estimated mean of 1577.4

At 80%, value is **3.1174e+04**

At 90%, value is **1.8990e+04**

At 95%, value is **1.3296e+04**

At 99%, value is **7.6732e+03**

At 99.9%, value is **4.7187e+03**

At 99.99%, value is **3.3753e+03**

For n = 1000 with estimated mean of 1601.3

At 80%, value is **1.0315e+03**

At 90%, value is **628.3462**

At 95%, value is **439.9209**

At 99%, value is **253.8910**

At 99.9%, value is **156.1331**

At 99.99%, value is **111.6831**

1. MATLAB Code

**This is the main section, where we compute all the results**

clear; close all; clc;

% Create a function that generates n Gaussian distributed random variables

X = gen\_grv(1600,120,10);

% Compute the mean of X

ux = mean(X)

% Tests whether the difference between the true mean and

%the estimated mean are statistically significant

status = check\_significance(ux, 1600, 120, 100, 80);

% find n value

N = determine\_significance(ux, 1600, 120, 99.99);

% find max variance

max\_var = max\_variance(ux, 1600, 10, 80);

**This section computes whether the difference between the means is statistically significant.**

function [ status ] = check\_significance(mean1, mean2, stdev, n, confidence)

status = abs((mean1 - mean2)/((stdev)/sqrt(n)));

k = confidence\_index(confidence);

if status > k

 fprintf('\nStatistically significant\n');

else

 fprintf('\nNot statistically significant\n');

end

end

**This section computes the minimum value of n for the difference between the means to be statistically significant.**

function [ N ] = determine\_significance(mean1, mean2, stdev, confidence)

k = confidence\_index(confidence);

N = abs((((k\*stdev)/(mean1-mean2))^2))

end

**This section computes the max variance**

function [ max\_var ] = max\_variance(mean1, mean2, n, confidence)

k = confidence\_index(confidence);

max\_var = (((mean1-mean2)\*sqrt(n))/k)^2

end

This section generates the Gaussian distribution

function [ X ] = gen\_grv(mean,stdev, n)

X = mean+ stdev\*randn(n,1);

end

**This section converts confidence into k value**

function [ k ] = confidence\_index( confidence )

if confidence == 99.99

 k = 3.89;

elseif confidence == 99.9

 k = 3.29;

elseif confidence == 99

 k = 2.58;

elseif confidence == 95

 k = 1.96;

elseif confidence == 90

 k = 1.64;

elseif confidence == 80

 k = 1.28;

else

 k = 0;

end

end

#  Conclusions

The difference between estimated mean and true was larger for small n numbers compared to large n numbers. This is expected because of the small sample size. As the sample size increases, the estimated mean gets closer to the true mean. Also, for higher standard deviation, in the case of 240 hours compared to 120 hours, the difference is smaller. For n = 100, the higher the confidence level, the higher the probability of the difference between the means of falling in the acceptance region. Also, the minimum value of n increases with the confidence level because as the confidence level increases, the probability of the estimated mean falling in the acceptance region increases.