Vaughn Wiernicki

ECE 3522: Stochastic Processes in Signals and Systems

Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA 19122

# Problem Statement

We were tasked with sampling a lightbulb production line and testing the mean lifetime of the light bulbs. Using different sample sizes of lightbulbs we are to determine if our results are statistically significant for various confidence levels. Using MATLAB we generated a few functions to generate Gaussian random numbers, to check the significance and to determine the minimum sample size at which the results would be significant.

# Approach and Results

Using MATLAB we generated our functions for Gaussian random numbers, checking significance and determining the minimum sample size for significance. Our first function was one line of code, being MATLAB’s “normrnd” function. To check significance I used some “if” statements to go through and determine what the confidence value and appropriate k value for that confidence level. Using the same equation for statistical significance we can rearrange the terms to give us a value of n. This value of n tells us the smallest sample size for statistical significance. We then do this for 5 different significance levels (0.8, 0.9, 0.95, 0.99, 0.999), 3 different sample sizes (10, 100, 1000) and then 2 different standard deviation values (120, 240). The results for the fifteen trials are reported below.

**σ = 120, n = 10:**

The first set of trials was for a standard deviation of 120 hours of life for the lightbulbs. For the three various sample sizes (10, 100, 1000) we get the following results. For a sampling of ten lightbulbs, we found that the new mean of ~1600 hours is only statistically significant at an 80% confidence level. The smallest level for significance is 7 lightbulbs for 80%, 11 lightbulbs for 90%, 16 lightbulbs for 95%, 27 lightbulbs for 99% and finally 44 lightbulbs for a 99.9% significance level.

**σ = 120, n = 100:**

At a sampling of one hundred lightbulbs we found that the new mean has statistical significance at 95% at most. This can be explained by a change in the mean due to a higher sample size. With more samples, we get a mean that it closer to the specified mean of 1600. For one hundred samples the smallest level for significance is 38 lightbulbs for 80%, 61 lightbulbs for 90%, 88 lightbulbs for 95%, 151 lightbulbs for 99% and finally 246 lightbulbs for a 99.9% significance level.

**σ = 120, n = 1000:**

At a sampling size of one thousand lightbulbs we found that the mean has statistical significance at all confidence levels. So naturally, we choose the highest claim of 99.9% confident that the new mean is statistically significant. For one thousand samples the smallest level for significance is 29 lightbulbs for 80%, 48 lightbulbs for 90%, 68 lightbulbs for 95%, 118 lightbulbs for 99% and finally 191 lightbulbs for a 99.9% significance level. We can see that the minimum numbers for statistical significance have decreased with a larger sample size, this could be due to the larger sample size evening the mean out and making it closer to the true value.

**σ = 240, n = 10:**

The second set of trials was for a standard deviation of 240 hours of life for the lightbulbs being tested. This means that the Gaussian will be more widely dispersed as 60% of the lightbulb’s lifetimes will now be within ±240 hours as opposed to 120 hours. We found that similar to the previous trial, the new mean, at a sample size of ten, is statistically significant at an 80% confidence level. The smallest numbers to make that claim at various confidence levels are 9 lightbulbs for 80%, 15 lightbulbs for 90%, 21 lightbulbs for 95%, 37 lightbulbs for 99% and finally 59 lightbulbs for a 99.9% significance level.

**σ = 240, n = 100:**

Now using a sample size of one hundred lightbulbs we found that our mean of 1600 is statistically significant with a 90% confidence level. This is a small confidence level than the trial with the previous standard deviation value and could be attributed to the values being more dispersed. The smallest numbers to make that claim at various confidence levels are 58 lightbulbs for 80%, 95 lightbulbs for 90%, 135 lightbulbs for 95%, 234 lightbulbs for 99% and finally 380 lightbulbs for a 99.9% significance level.

**σ = 240, n = 1000:**

Finally, with a sample size of one thousand lightbulbs and our new standard deviation we get a confidence level of 95%. The new standard deviation is really starting to hurt the confidence of statistical significance for the new measured mean. The smallest numbers to make that claim at various confidence levels are 253 lightbulbs for 80%, 414 lightbulbs for 90%, 592 lightbulbs for 95%, 1025 lightbulbs for 99% and finally 1666 lightbulbs for a 99.9% significance level.

# MATLAB Code

clear; clc; clf; close all;

stdev = 120;

% stdev = 240;

mean = 1600;

n = [10, 100, 1000];

Mmu = 1570;

conf = [80, 90, 95, 99, 99.9];

%% For n = 10:

X1 = gen\_grv(mean, stdev, n(1));

sum = 0;

for i = 1:1:length(X1)

sum = sum + X1(i);

end

mu = sum/length(X1);

status1 = zeros(1, length(conf));

significance1 = zeros(1, length(conf));

var1 = zeros(1, length(conf));

for i = 1:1:length(conf)

status1(i) = check\_significance(mu, Mmu, stdev, n(1), conf(i));

significance1(i) = determine\_significance(mu, Mmu, stdev, n(1), conf(i));

var1(i) = vars(mu, Mmu, n(1), conf(i));

end

%% For n = 100:

X2 = gen\_grv(mean, stdev, n(2));

sum = 0;

for i = 1:1:length(X2)

sum = sum + X2(i);

end

mu = sum/length(X2);

status2 = zeros(1, length(conf));

significance2 = zeros(1, length(conf));

var2 = zeros(1, length(conf));

for i = 1:1:length(conf)

status2(i) = check\_significance(mu, Mmu, stdev, n(2), conf(i));

significance2(i) = determine\_significance(mu, Mmu, stdev, n(2), conf(i));

var2(i) = vars(mu, Mmu, n(2), conf(i));

end

%% For n = 1000:

X3 = gen\_grv(mean, stdev, n(3));

sum = 0;

for i = 1:1:length(X3)

sum = sum + X3(i);

end

mu = sum/length(X3);

status3 = zeros(1, length(conf));

significance3 = zeros(1, length(conf));

var3 = zeros(1, length(conf));

for i = 1:1:length(conf)

status3(i) = check\_significance(mu, Mmu, stdev, n(3), conf(i));

significance3(i) = determine\_significance(mu, Mmu, stdev, n(3), conf(i));

var3(i) = vars(mu, Mmu, n(3), conf(i));

end

----------------------------------------------------------------

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

X = normrnd(mean, stdev, 1, n);

end

----------------------------------------------------------------

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

if (conf == 99.9)

k = 3.29;

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

end

if (conf == 99)

k = 2.58;

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

end

if (conf == 95)

k = 1.96;

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

end

if (conf == 90)

k = 1.64;

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

end

if (conf == 80)

k = 1.28;

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

end

end

----------------------------------------------------------------

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

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

status = 0;

if (conf == 99.9)

k = 3.29;

if (k > z)

status = 1;

end

end

if (conf == 99)

k = 2.58;

if (k > z)

status = 1;

end

end

if (conf == 95)

k = 1.96;

if (k > z)

status = 1;

end

end

if (conf == 90)

k = 1.64;

if (k > z)

status = 1;

end

end

if (conf == 80)

k = 1.28;

if (k > z)

status = 1;

end

end

end

----------------------------------------------------------------

function [var] = vars(mean1, mean2, n, conf)

if (conf == 99.9)

k = 3.29;

var = n \* (mean1 - mean2)^2/(k)^2;

end

if (conf == 99)

k = 2.58;

var = n \* (mean1 - mean2)^2/(k)^2;

end

if (conf == 95)

k = 1.96;

var = n \* (mean1 - mean2)^2/(k)^2;

end

if (conf == 90)

k = 1.64;

var = n \* (mean1 - mean2)^2/(k)^2;

end

if (conf == 80)

k = 1.28;

var = n \* (mean1 - mean2)^2/(k)^2;

end

end

# Conclusions

We can see that the sample size has a lot to do with our confidence level. Statistical significance changes with respect to the number of sampled units as well as the dispersion of our data. Typically our data can be represented as a Gaussian function, but that still leaves room for changes in the mean and the standard deviation. For our application standard deviation’s effect on our confidence level isn’t worlds apart, but at the same time is not negligible. However, in a more industrial application, standard deviation can be very important. A reduced confidence level can do poorly for business.