15.075 Statistical Thinking and Data Analysis
Computer Exercises 2
Due November 10, 2011

Instructions: Please solve the following exercises using MATLAB. One simple way to present your solutions is to copy all of your code and results (plots, numerical answers, etc.) into a Word document, which you should submit in class. There is no online submission. **Do NOT share code!**

Exercise 1
In this exercise, we demonstrate the Central Limit Theorem, where each X_i is exponential.

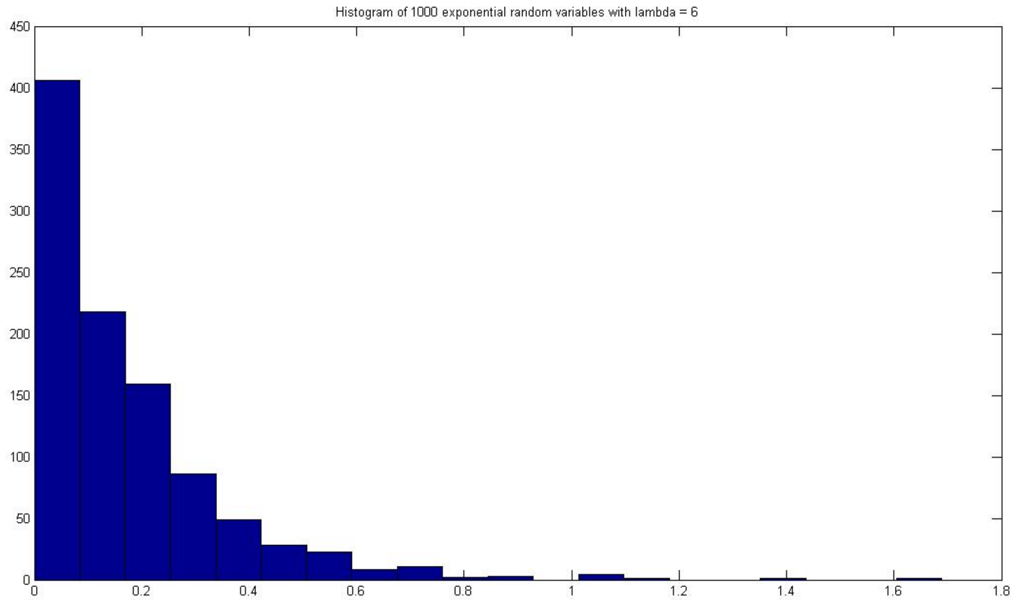
a. Generate 1000 random numbers from the exponential distribution with $\lambda = 6$, and plot them in a histogram. This should give you an idea of what the exponential distribution looks like.
b. For $n = 2$, repeat the following 1000 times: Generate a random sample of n numbers from the exponential distribution with $\lambda = 6$.
c. Compute the sample mean of the n numbers and standardize it using the true mean and standard deviation of the distribution.
d. Make a histogram and normal plot of the 1000 sample means.
e. Repeat (b)-(d) for $n = 10$, 20, and 100. Put all four histograms and all four normal plots in the same window. Comment on their shapes.

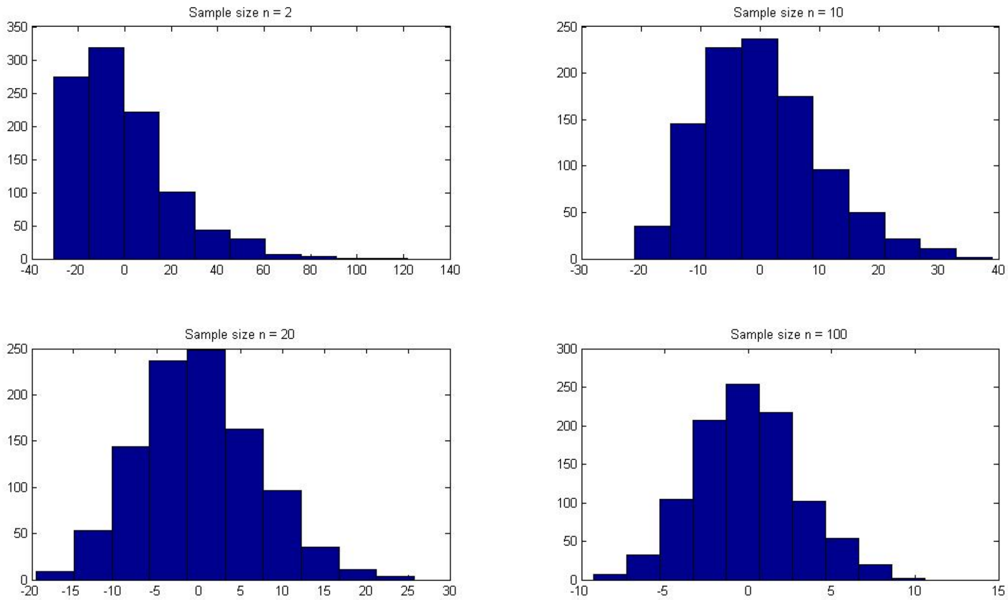
Solution

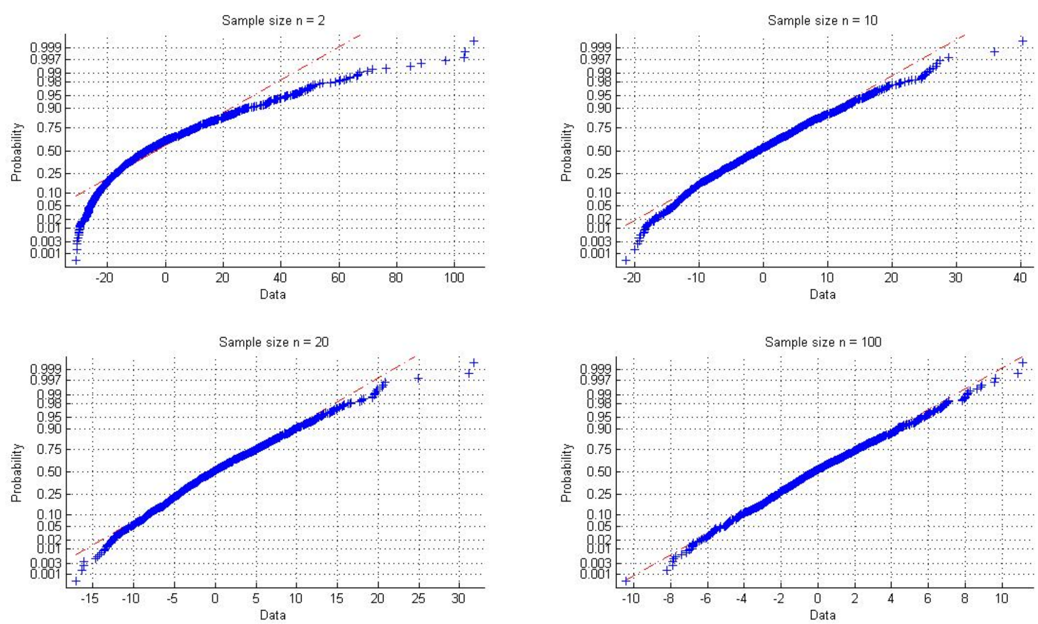
Here is the script and the results for my iterations:

```
clear all;
clc;
close all;

num_samples = 1000;
lambda = 6;
mu = 1/lambda;
sigma = mu;

% Part a
% Generate 1000 random numbers from the exponential distribution with
% lambda=6
x = exprnd(mu,num_samples,1);

% Plot a histogram
hist(x,20);
title('Histogram of 1000 exponential random variables with lambda = 6');

% Part b,c,d,e
sample_sizes = [2 10 20 100];
figure;
for i = 1:length(sample_sizes)
    sample_size = sample_sizes(i);
    x = exprnd(mu,num_samples,sample_size);
    sample_mean = mean(x');
    z_score = (sample_mean - mu)/(sigma/sqrt(num_samples));
```

```
    % Plot a histogram z-scores
    subplot(2,2,i);
    hist(z_score);
    title(['Sample size n = ', num2str(sample_size)])

    % Plot a normal plot of z-scores
    subplot(2,2,i);
    normplot(z_score);
    title(['Sample size n = ', num2str(sample_size)])
end
```







The normal plot for *n*=2 shows that the distribution of sample means is still skewed right since the exponential distribution has a very long positive tail. As *n* increases, the distribution of sample means becomes increasingly symmetric. With *n*=100, the curve in the normal plot is fairly straight, implying the distribution of sample means is approximately normal.

Exercise 2

Solve Problem 6.23 from the book.

Solution

a) Obviously the α-risk is 0.05 from definition. Let's calculate the β-risk of the rule if μ=1.

The β-risk is just β = 1-π(1) = 1 - $\Phi[-z_\alpha + \frac{(\mu-\mu_0)\sqrt{n}}{\sigma}]$ = 1-Φ(-1.645+3) = 1-Φ(1.355) = 0.0877.

b) Here is the script and the results for my iterations:

```
clear all;
clc;
close all;

% Parameters
level = 0.05;
mu_0 = 0;
mu = 1;
sigma = 1;
n = 9;
num_samples = 100;
% Reject if sample mean is greater than critical_value
critical_value = mu_0+norminv(1-level)*sigma/sqrt(n);

% Generate 100 samples of size n with mean mu_0
x = sigma*randn(num_samples,n)+mu_0;
x_bar = mean(x');
a_risk = sum(x_bar>critical_value)/num_samples;


% Generate 100 samples of size n with mean mu
x = sigma*randn(num_samples,n)+mu;
x_bar = mean(x');
b_risk = sum(x_bar<critical_value)/num_samples;
```

```
>> a_risk

a_risk =

    0.0500

>> b_risk

b_risk =

    0.0700
```

The results are close to their theoretical values.

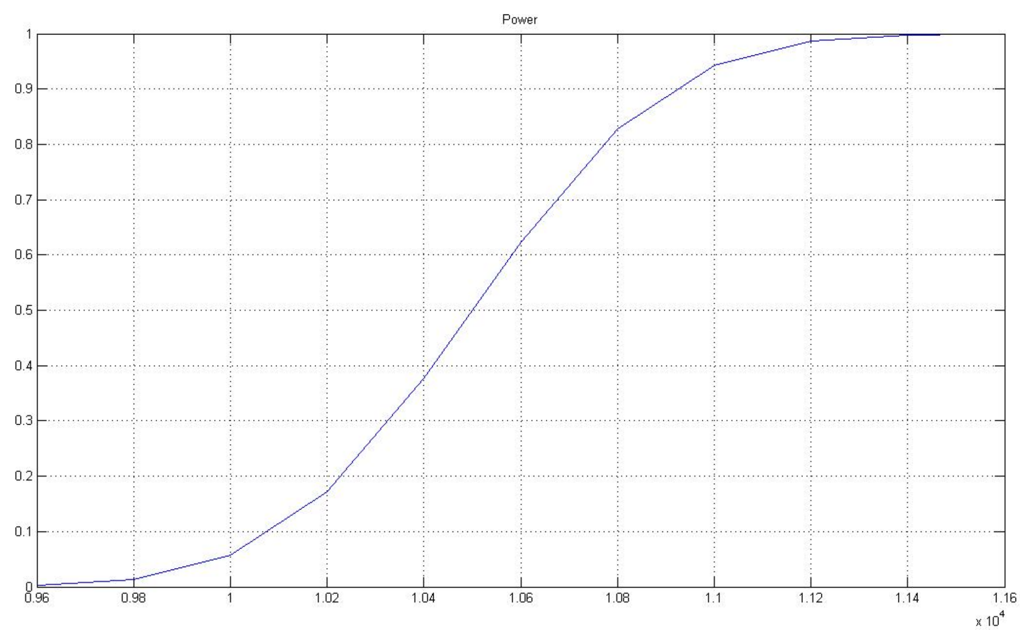
Exercise 3

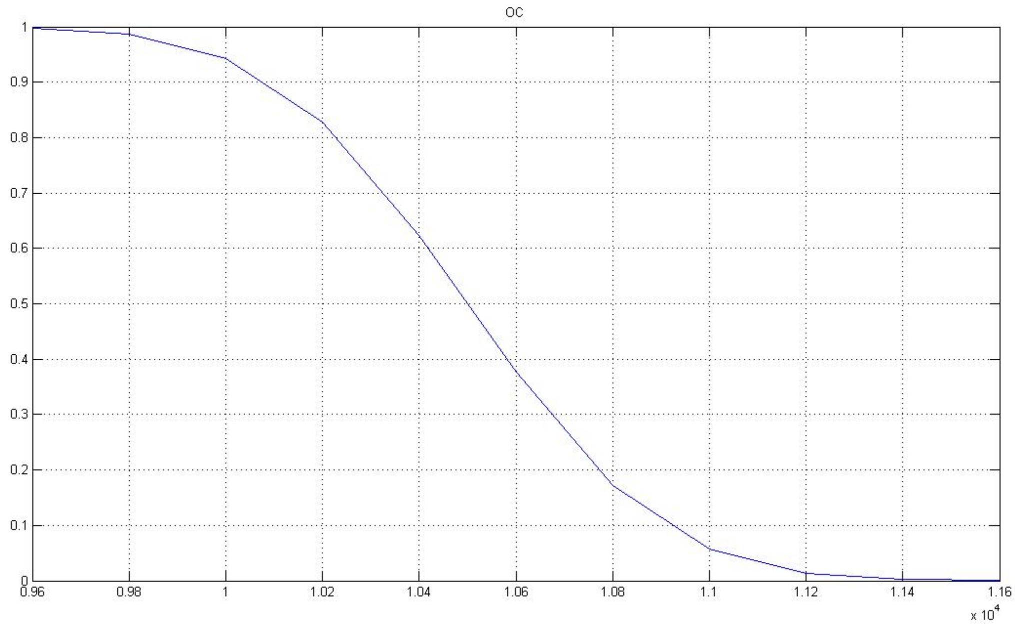
Solve Problem 6.25 from the book.

Solution

Here is the script and the results for my iterations:

```
clear all;
clc;
close all;

sigma = 1000;
n = 10;

% We reject when x_bar is greater than the critical value
critical_value = 10500;


mu = 9600:200:11600;

power = 1-normcdf((critical_value-mu)*sqrt(n)/sigma);
oc = 1-power;

plot(mu,power);
grid;
title('Power');


figure;
plot(mu,oc);
grid;
title('OC');
```




OC
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.96
0.98
1
1.02
1.04
1.06
1.08
1.1
1.12
1.14
1.16
x 10^4

Exercise 4
A thermostat used in an electrical device is to be checked for the accuracy of its design setting of 200°F. Ten thermostats were tested to determine their actual settings, resulting in the following data:

202.2	203.4	200.5	202.5	206.3	198.0	203.7	200.8	201.3	199.0

Assume the settings come from a normal distribution. Using $\alpha = 0.05$, perform a hypothesis test to determine if the mean setting is greater than 200°F. What are the null and alternative hypotheses? Which test do you use and why? Explain your conclusion using

a. An appropriate confidence interval.
b. A critical value from the distribution of the test statistic.
c. A *p*-value.

Solution

The hypotheses are
H_0: $\mu = 200$.
H_1: $\mu > 200$.

We use a *t*-test since we do not know the true variance, and the sample size is small.

```
clear all;
clc;
close all;

x = [202.2,203.4,200.5,202.5,206.3,198.0,203.7,200.8,201.3,199.0];
mu = 200;
alpha = 0.05;


[h p ci stats] = ttest(x,mu,alpha, 'right')
```

h =

1

p =

0.0227

ci =

200.3729 Inf

stats =

tstat: 2.3223
df: 9
sd: 2.4102

For all the following reasons, we reject H_0.

a. The lower 95% CI is [200.3729, ∞), which does not contain 200.
b. The observed value of the *t*-test statistic is 2.3223, which is larger than $t_{9,0.05} = 1.833$.
c. The *p*-value is 0.0227, which is smaller than the significance level $\alpha = 0.05$.

15.075J / ESD.07J Statistical Thinking and Data Analysis
Fall 2011