STAT572 - Homework Assignment 2

Juan Carlos Apitz, ID 012523821

September 25, 2014

In-Class Exercises

Exercise 2: Generate a Random Sample from the Gamma(3,2)

For this exercise I used the in-class procedure where first we randomly generate uniformly a number $0 < u < 1$. Then we choose $x = min\{x : F(x) \geq u\}$. Where $F(x)$ is the CDF of the Gamma distribution with parameters $n = 3$ and $\lambda = 2$. Compared with the output from the MATLAB Gamma random sample generating function [command: $hist(gamrnd(3,2,1,1000))$], we see that the proposed code works well. See figures 1, 3 and 2 below.

Editor - /home/jcapitz/Documents/Stat572/labs_inclass/9_11_14/gammars.m

gammars.m | csgammp.m | question3pt1.m | ex4_7.m

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% In-Class Exercise, September 11, 2014                              %
% Generate a sample of size 1000 from a Gamma distribution with      %
% parameters n and lambda.                                           %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%initialize parameters
n=3;
lambda=2;
x=linspace(0,30);
F=zeros(1,length(x));
rs=zeros(1,1000);
u=rand(1,1000);
% Loop thru uniforms and x vector picking x's that meet criteria
for i=1:1000
    for j=1:length(x)
        if gamcdf(x(j),n,lambda)<u(i)
            F(j)=1;
        else
            F(j)=gamcdf(x(j),n,lambda);
        end
    end
    l=find(F==min(F));
    rs(i)=x(l);
end
figure
hist(rs)
title('Frequecy Histogram of Gamma Random Sample')
```

Figure 1: Code of the in-class Gamma random sample exercise.

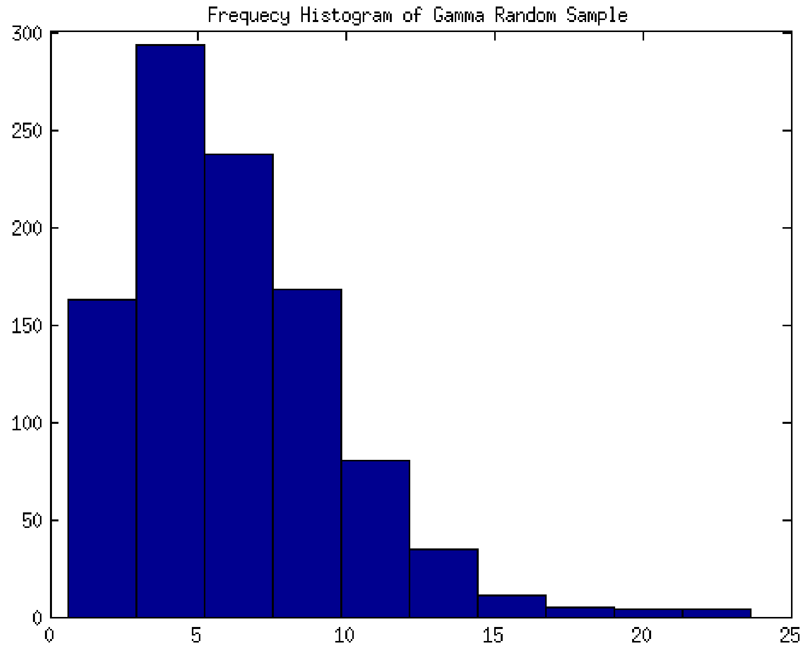


Figure 2: Histogram of the in-class Gamma random sample exercise, $n = 3$ and $\lambda = 2$.

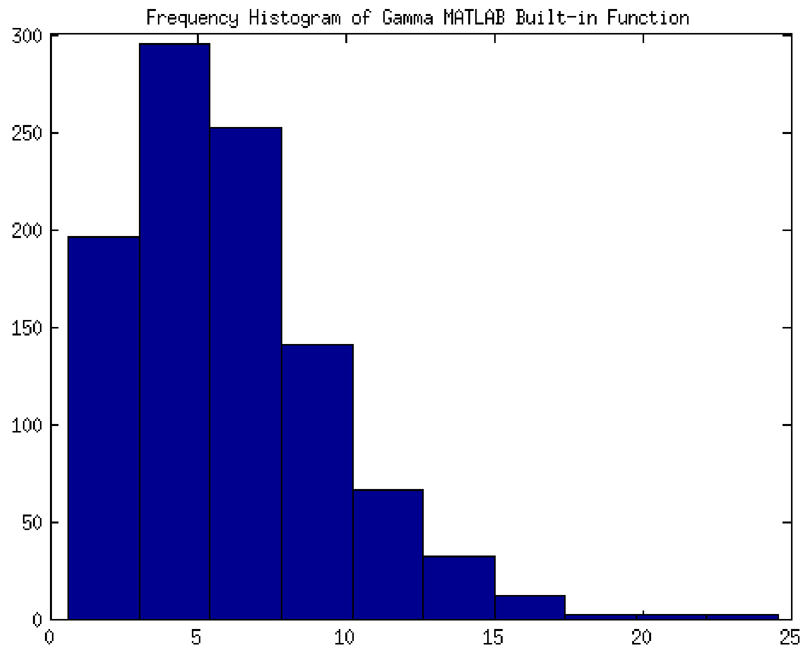


Figure 3: Histogram of a Gamma random sample generated by MATLAB's built-in function, $n = 3$ and $\lambda = 2$.

Homework 2

Exercise 3.1

Code for Exercise 3.1 and 3.5.

```
Editor - /home/jcapitz/Documents/Stat572/homework/hw2/question3pt1.m
question3pt5.m  question3pt6.m  question3pt9.m  question3pt1.m

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Exercise 3.1: Generate 500 random samples of size n:                      %
% and analyze the behavior of the sample mean and variance.                 %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

n = [2,15,45];
size = {'n = 2 ';'n = 15';'n = 45'};
Mean =  zeros(3,1);
Variance = zeros(3,1);

for i=1:length(n)
    x = randn(n(i),500);
    xbar = mean(x);
    Mean(i) = mean(xbar);
    Variance(i) = var(xbar);

    % plot the data
    figure(i)
    histfit(xbar)
    title(['Frequency Histogram, ' 'n=' num2str(n(i))])
end

Table = table(Mean,Variance,'RowNames',size);
disp(Table)
```

Figure 4: Code Exercise 3.1.

Discussion of Mean and Variance.

In figure 5 below we can see the resulting values for the sample mean and sample variance of the calculated mean at each sample size. As predicted by the Central Limit Theorem, we see that the distribution of the sample mean, $\bar{x}$, looks approximately normal, the larger the sample size (see histograms,figure 6, 7, and 8). Additionally, we see that as n gets larger, $\bar{x}$ approaches $\mu = 0$, and $s^2_{\bar{x}}$ approaches $\frac{\sigma^2}{n} = \frac{1}{n} \to 0$ the larger the sample size. This is precisely what is predicted by CLT.

```
Command Window

>> question3pt1
                 Mean       Variance
              ________     ________

    n = 2     0.010139      0.47093
    n = 15    0.021819     0.063246
    n = 45    0.001743       0.0202

fx >>
```

Figure 5: Results, mean and variance.

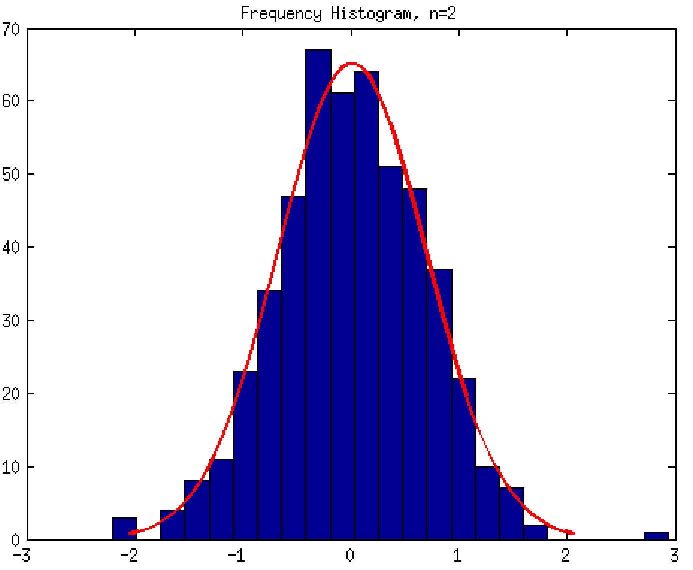


Figure 6: Frequency Histogram of Sample Means, n=2.

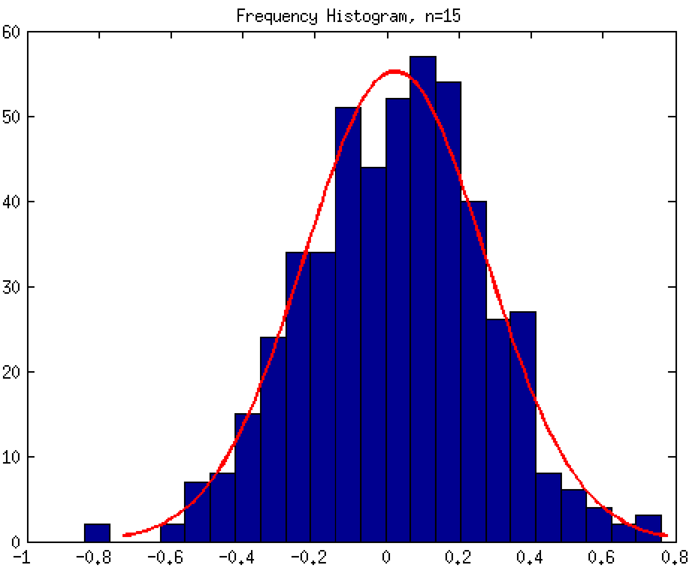


Figure 7: Frequency Histogram of Sample Means, n=15.

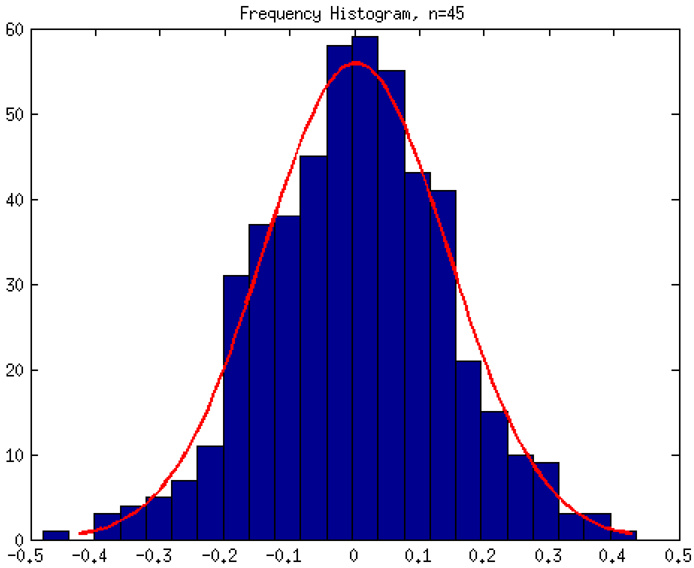


Figure 8: Frequency Histogram of Sample Means, n=45

Exercise 3.5

Discussion of Skewness and Kurtosis.

In figure 9 we can see the results for the normal distribution's skewness and kurtosis . Since the data comes from a normal distribution, we expect the skewness coefficient to be near 0 by symmetry. This is the case, specially when $n = 45000$, which is expected as the larger the sample size the closer it will resemble its true distribution. The kurtosis of a normal distribution is in theory equal to 3. As we can observe in figure 9 the coefficient of kurtosis tends to 3.

Editor - /home/jcapitz/Documents/Stat572/homework/hw2/question3pt5.m

question3pt5.m | question3pt6.m | question3pt9.m | question3pt1.m

```matlab
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Exercise 3.5: Generate normal random samples of size n            %
% and analyze skewness, and kurtosis.                                %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

n = [2000,15000,45000];
size = {'n = 2000 ';'n = 15000';'n = 45000'};
Skewness = zeros(3,1);
Kurtosis = zeros(3,1);

for i = 1:length(n)
    % generate a random sample from the uniform dist
    x = randn(1,n(i));
    % find the mean of the sample
    mu=mean(x);

    % find the skewness
    num = (1/n(i))*sum((x-mu).^3);
    den = (1/n(i))*sum((x-mu).^2);
    Skewness(i) = num/den^(3/2);

    % find the kurtosis
    num=(1/n(i))*sum((x-mu).^4);
    den = (1/n(i))*sum((x-mu).^2);
    Kurtosis(i) = num/den^2;
end

Table = table(Skewness,Kurtosis,'RowNames',size);
disp(Table)
```

Command Window

```
>> clear
>> question3pt5
                  Skewness     Kurtosis
                 __________    ________

    n = 2000      0.012747     3.0982
    n = 15000     0.014156      2.933
    n = 45000   -0.0074426     3.0297

>>
```

Figure 9: Code and results for analysis of skewness and kurtosis - normal distribution.

Exercise 3.6

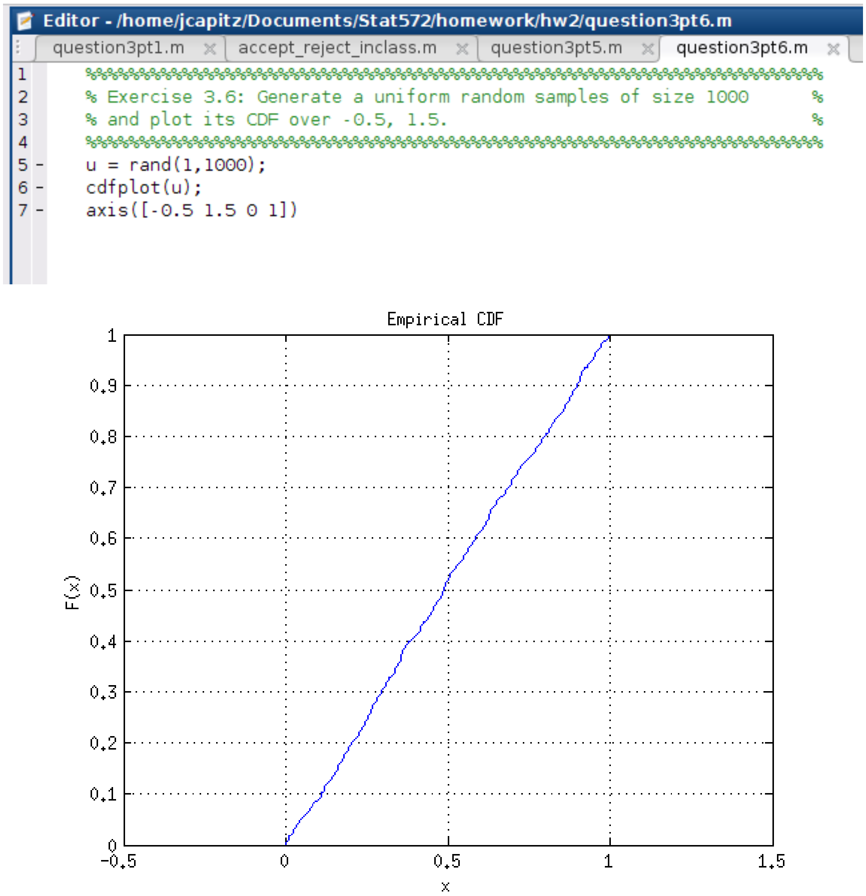


Figure 10: Code and CDF plot for uniform sample, $n = 1000$.

Exercise 3.9

Quantile Estimation Discussion.

In theory, we expect, that given a random variable $u \sim Unif(0,1)$, its $q_{33} = 0.33$, $q_{40} = 0.40$, $q_{63} = 0.63$, $q_{90} = 0.90$. In this exercise, we get estimates that are close to the theoretical results: $\hat{q}_{33} = 0.3329$, $\hat{q}_{40} = 0.4033$, $\hat{q}_{63} = 0.6422$, $\hat{q}_{90} = 0.859$. See figure 11 below for results.

```
Editor - /home/jcapitz/Documents/Stat572/homework/hw2/question3pt9.m
question3pt5.m   question3pt6.m   question3pt9.m   question3pt1.m   ex3_6.m

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Exercise 3.9: Generate a uniform random samples of size 100       %
% and calculate quantiles p=.33, p=.40, p=.63, p=.90.               %
% Modified code from Martinez, example 3.6                          %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% First generate some uniform(0,1) data.
x = rand(100,1);
% Now get the order statistics. These will serve
% as the observed values for the ordinate (Y_obs).
xs=sort(x);
% Now get the observed values for the abscissa (X_obs).
n=length(x);
phat = ((1:n)-0.5)/n;
% We want to get the quartiles.
p = [0.33, 0.40, 0.63, 0.90];
% The following provides the estimates of the quartiles
% using linear interpolation.
qhat=interp1(phat,xs,p);
```

```
Command Window
>> question3pt9
>> qhat

qhat =

    0.3329    0.4033    0.6422    0.8590

>>
```

Figure 11:

Exercise 3.11

For this exercise I increased the sample size to $n = 100,000$ for example 3.5 and to $n = 500,000$ for example 3.6. The results are shown in figure 12. As we can see, these results are much closer to the true quartiles for both the Uniform and Normal distributions. For reference:

Uniform(0,1): $q_1 = 0.25$, $q_2 = 0.50$, $q_3 = 0.75$;

Normal(0,1): $q_1 = -0.6745$, $q_2 = 0$, $q_3 = 0.6745$.

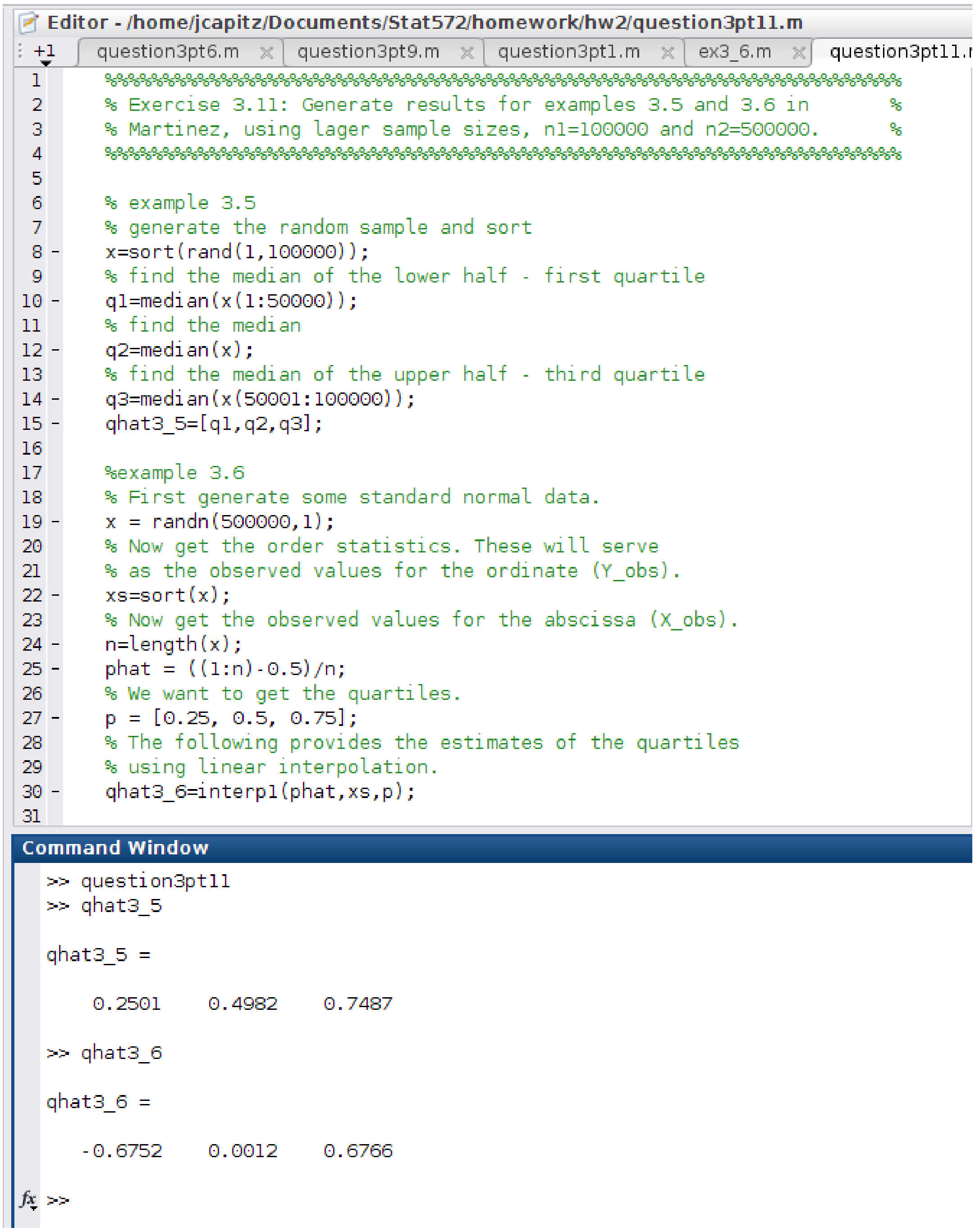
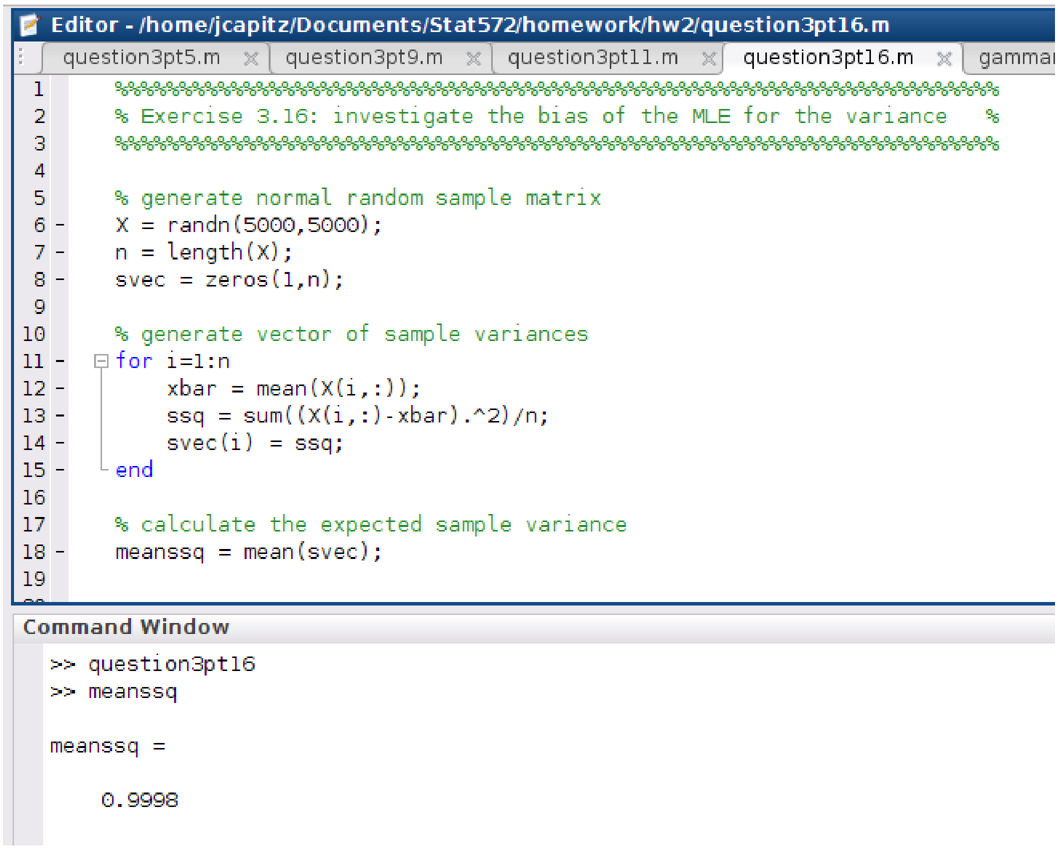


Figure 12:

Exercise 3.16

After completing the procedure required, the mean of the variances calculated is 0.998, which gives us an absolute difference of $|0.998-1| = 2.2202\times 10^{-4}$, so very small. This is because I used a sample of size 5,000 which makes the bias of the estimator very small. From mathematical statistics we know that the estimator $S^2 = \sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{n}$ has a bias of $\frac{n-1}{n}$ since $E(S^2) = \frac{n-1}{n}\sigma^2$.

For this exercise the theoretical value of the bias is $\frac{4999}{5000} = 0.9998$ which precisely matches our our result, i.e. $E(S^2) = \frac{n-1}{n}\sigma^2 = \frac{4999}{5000}\times 1 = 0.9998$. See figure 13 below.



Editor - /home/jcapitz/Documents/Stat572/homework/hw2/question3pt16.m

question3pt5.m | question3pt9.m | question3pt11.m | question3pt16.m | gamma

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Exercise 3.16: investigate the bias of the MLE for the variance    %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% generate normal random sample matrix
X = randn(5000,5000);
n = length(X);
svec = zeros(1,n);

% generate vector of sample variances
for i=1:n
    xbar = mean(X(i,:));
    ssq = sum((X(i,:)-xbar).^2)/n;
    svec(i) = ssq;
end

% calculate the expected sample variance
meanssq = mean(svec);
```

Command Window

```
>> question3pt16
>> meanssq

meanssq =

    0.9998
```

Figure 13: