MA 402: Project 2

Matthew Murray, Johnathan Rhyne

Fall 2019

Instructions:

- Detailed instructions regarding submission are available on the class website[1].
- The zip file should contain three files project2.pdf, project2.tex, classnotes.sty.

1) (10 points) The infamous RANDU generator was part of the Scientific Subroutine Package on IBM main-frame computers in the 1960s; the generator corresponds to

$$I_{n+1} \equiv (aI_n + c) \bmod m, \qquad n = 0, 1, \ldots$$

with $a = 65539$, $c = 0$ and $m = 2^{31}$.

(a) Show that

$$I_{n+2} - 6I_{n+1} + 9I_n \equiv 0 \bmod m.$$

Hint: Recall that $a \equiv b$ mod m means $a = km + b$ for some integer k. Another useful fact is that $a = 65539 = 2^{16} + 3$.

Solution: Assume $I_n \in \mathbb{Z}$

Given: $I_{n+1} \equiv (aI_n)\text{mod} \implies I_{n+1} = mk_1 + aI_n$

$$\begin{aligned}
I_{n+2} \equiv aI_{n+1} \bmod m &\implies I_{n+2} = k_2 m + aI_{n+1} \\
&\implies I_{n+2} = k_2 m + a(mk_1 + aI_n) \\
&\implies I_{n+2} = k_2 m + amk_1 + a^2 I_n \\
Given: a^2 &= (2^{16} + 3)^2 = 2^{32} + 6 * 2^{16} + 9 \\
&\implies I_{n+2} = k_2 m + amk_1 + I_n(2^{32} + 6 \cdot 2^{16} + 9)
\end{aligned}$$

Let: $I_{n+2} - 6I_{n+1} + 9I_n \equiv (a)\text{mod} m \ni a \in [0, m)$

[1] `https://github.ncsu.edu/asaibab/ma402_fall_2019/blob/master/projects.md`

$\therefore$ We know that:

$$I_{n+2} - 6I_{n+1} + 9I_n \equiv (b)\text{mod}m$$
$$\implies k_2m + amk_1 + I_n(2^{32} + 6 \cdot 2^{16} + 9) - 6mk_1 - 6 \cdot 2^{16}I_n - 18I_n + 9I_n \equiv b\text{mod}m$$
$$\implies m(k_2 + ak_1 - 6k_1) + 2^{32}I_n \equiv b\text{mod}m$$
$$\implies m(k_2 + ak_1 - 6k_1) + 2^{32}I_n = k_3m + b$$
$$Given : k_4 = k_2 + ak_1 - 6k_1$$
$$\implies mk_4 + 2 \cdot mI_n = k_3m + b$$
$$\implies 2I_n = k_3 - k_4 + \frac{b}{m}$$

$Let : k_5 = k_3 - k_4$ and since $b < m$ we know that $\dfrac{b}{m} < 1$

$$\implies 2I_n = k_5 + \frac{b}{m}$$

Since $I_n \in \mathbb{Z}$ we know that b must be 0 as an Integer plus a non-Integer is not an Integer

$\therefore k_5 = 2$ and we have shown that:

$$I_{n+2} - 6I_{n+1} + 9I_n = km \ni k \in \mathbb{Z}$$
$$\implies I_{n+2} - 6I_{n+1} + 9I_n \equiv (0)\text{mod}m$$

■

(b) Implement RANDU and verify graphically its severe lack of equi-distribution by creating a three dimensional plot of the 10,000 points for some odd I_0 of your choice. More precisely, plot (in 3D)

$$\{(I_{n-1}, I_n, I_{n+1})/m\} \qquad n = 1, \ldots, 10,000.$$

Solution:

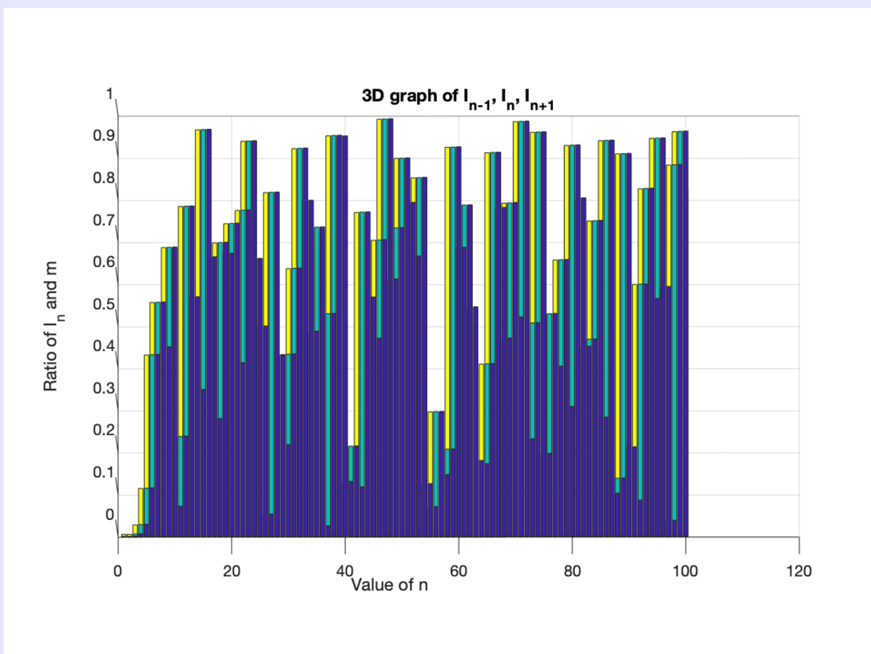


Figure 1: Graph with n=100 for better visibility and $I_0 = 35$

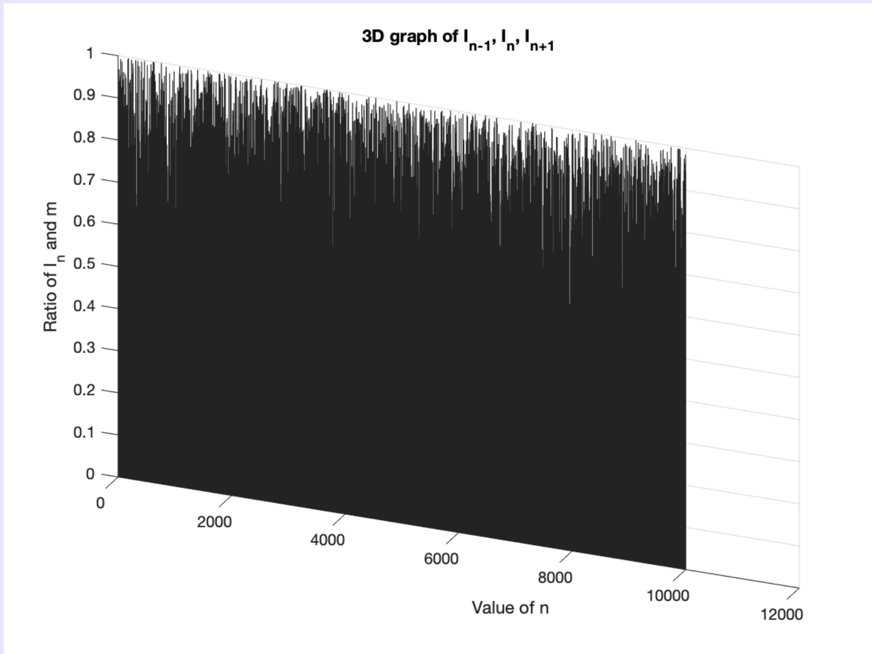
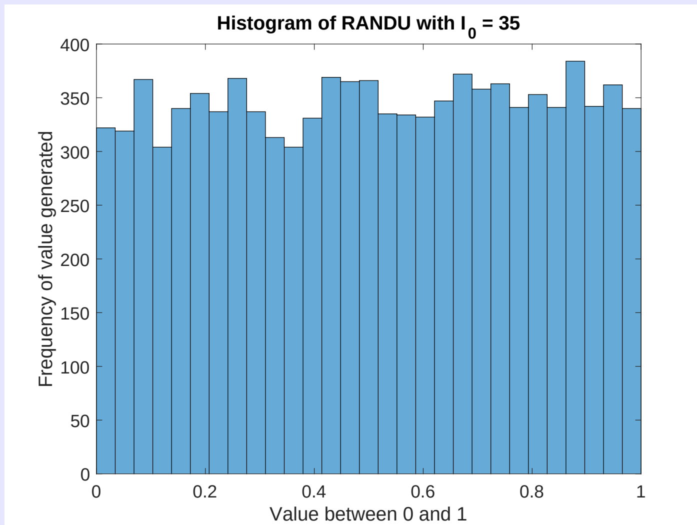


Figure 2: Graph with n=10000 for problem requirement and $I_0 = 35$

(c) Plot $\{I_n/m\}$ for $n = 1, \ldots, 10000$ as a histogram with 30 bins.

Note: You should submit two different plots for this problem.

Solution:



e

Figure 3: Histogram of our RANDU generator with $I_0 = 35$

2) (10 points) Certain Boeing 757 models are configured with 168 economy seats. Experience shows that only 90% of economy-class ticket holders actually show up to board the plane. If an airline sells 178 tickets, then what is the probability of overbooking, so that some passengers do not get the seat they paid for? Compute using

(a) the Binomial distribution;

Solution: Let X_j be the decision of an economy-class ticket holder to show up and board their flight: (1-will show up), (0-won't show up)

$$X_j = \begin{cases} 1 & , p = 0.9 \\ 0 & , 1-p = 0.1 \end{cases}$$

With this definition, $S_n = \sum_{j=1}^{n} X_j$ is the number of people of people who show up to board the plane. We know that $n = 178$ is the number of people who bought a ticket. The eact solution to this takes advantage of the fact that $S_n \sim Binomial(n = 178, p = 0.9)$. Using Matlab,

$$\implies \mathbb{P}(S_{178} > 168) \approx 0.01325$$

(b) Laplace's approximation to the Binomial distribution.

Solution: In the language of this problem, Laplace's Approximation to the Binomial Theorem states that:

$$\lim_{n\to\infty} \mathbb{P}\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) = \Phi(b) - \Phi(a), \text{where } \frac{S_n - np}{\sqrt{np(1-p)}} \approx N(0,1).$$

$$\therefore \mathbb{P}\left(\frac{168 - 178(0.9)}{\sqrt{178(0.9)(0.1)}} \leq \frac{S_{178} - np}{\sqrt{(178)(0.9)(0.1)}} \leq \infty\right) = 1 - \Phi\left(\frac{168 - 178(0.9)}{\sqrt{178(0.9)(0.1)}}\right) \approx 0.02566$$

(c) Laplace's approximation to the Binomial distribution (with end correction).

Solution:

$$\mathbb{P}\left(\frac{168 - 178(0.9)}{\sqrt{178(0.9)(0.1)}} + 0.5 \leq \frac{S_{178} - np}{\sqrt{(178)(0.9)(0.1)}} \leq \infty\right) = 1 - \Phi\left(\frac{168 - 178(0.9)}{\sqrt{178(0.9)(0.1)}} + 0.5\right) \approx 0.007167$$

You may use MATLAB/Python scripts or a calculator for the necessary computations.

3) (10 points) Let X be the double exponential random variable, which has the pdf

$$f(x) = e^{-2|x|} = \begin{cases} e^{2x} & -\infty < x < 0 \\ e^{-2x} & 0 \leq x < \infty \end{cases} .$$

(a) Derive the CDF $F(x)$ and the inverse CDF $F^{-1}(x)$.

Solution: The CDF $F(x)$ is the integral $\int_{-\infty}^{x} f(z)\,\mathrm{d}z$ which has 2 cases
Which are as follows:

$$x < 0 \implies f(x) = e^{2x} \implies F(x) = \int_{-\infty}^{x} e^{2z}\,\mathrm{d}z = \frac{e^{2x}}{2}$$

$$x \geq 0 \implies f(x) = e^{-2x} \implies F(x) = 1 - \frac{e^{-2x}}{2}$$

We show that this is a valid CDF and pdf. To do so, we must show
$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = F(\infty) - F(-\infty) = 1$
$\lim_{x\to-\infty} \dfrac{e^{2x}}{2} = 0$
$\lim_{x\to\infty} 1 - \dfrac{e^{-2x}}{2} = 1$
$1 - 0 = 1 \therefore F(\infty) - F(-\infty) = 1$

The range of $F(x)$ for $x < 0$ is $[0, \frac{1}{2})$
The range of $F(x)$ for $x \geq 0$ is $[\frac{1}{2}, 1)$
By flipping the values of x and $F(x)$ and replacing $F(x)$ with $F^{-1}(x)$ we get:
$x \in [0, \frac{1}{2}) \implies F^{-1}(x) = \ln\sqrt{2x}$
$x \in [\frac{1}{2}, 1) \implies F^{-1}(x) = \ln\sqrt{2 - 2x}$

(b) By using the "inverting the CDF" technique and a Uniform$(0,1)$ random generator, show how to generate a random variable that has the double exponential distribution.

Solution: Now we check the end points that there will be no issues in our next step.
$F^{-1}(0) = \ln 0$ which is not defined
$F^{-1}(\frac{1}{2}) = \ln \sqrt{2-1} = \ln 1 = 0$ which is defined
Since any particular value of a continuous distribution has a near zero chance, we can ignore 0 and with a generator, just regenerate the value if it is a 0

If we are given a random variable distributed as a uniform(0,1) then we will plug these values into $F^{-1}(x)$

(c) Generate 1000 random variables from the double exponential distribution. In a single plot, with two subplots: (1) plot the histogram of the samples, and (2) plot the pdf of the double exponential distribution.

Solution:

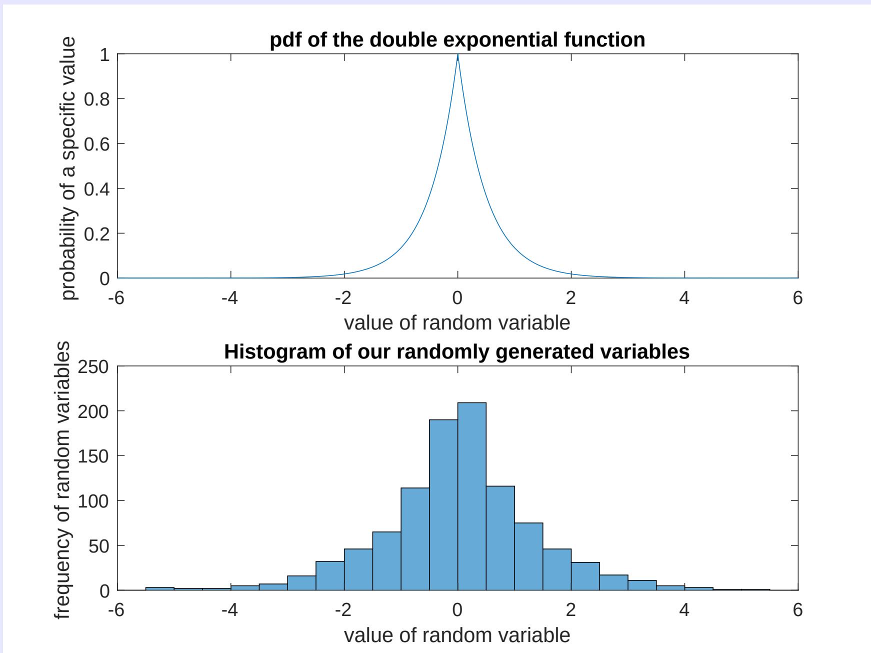


Figure 4: pdf of the double exponential compared to our histogram of our randomly generated value

We notice that both distributions follow roughly the same shape leading us to believe that we generated our variable correctly

4) (10 points) Many people believe that the daily change of price of a company's stock on the stock market is a random variable with mean 0 and variance σ^2 . That is, if Y_n represents the price of the stock on the nth day, then

$$Y_n = Y_{n-1} + X_n \qquad n \geq 1$$

where $X_1, X_2, \ldots$ are independent and identically distributed random variables with mean 0 and variance

σ^2. Suppose that the stock's price today is 100 and assume $\sigma^2 = 1$.

(a) Plot 20 different trajectories of the stock prices over the next 10 days. You may use MATLAB's `randn` or Python's `numpy.random.randn` to generate normal random numbers.

Solution:

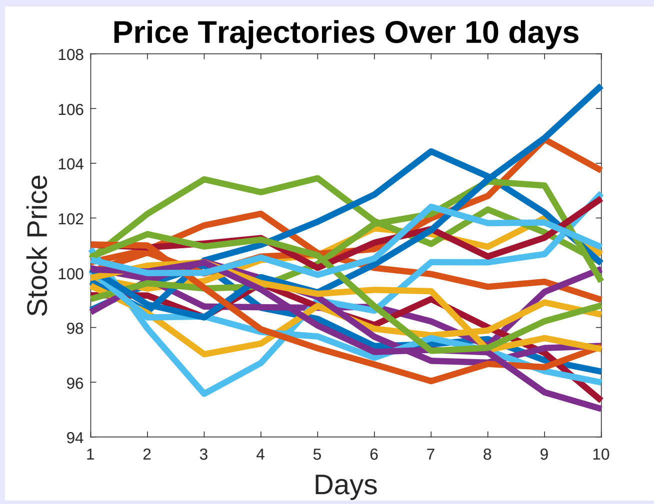


Figure 5: Plot of the Stock Price Trajectories Over a Period of 10 days

(b) Using Monte Carlo, estimate the probability that the stock's price will exceed 105 after 10 days. Report the results using $10, 100, 500$ samples.

Hint: Let $X = Y_{10} = Y_0 + \sum_{i=1}^{10} X_i$. Use the fact that $P(X \geq 105) = E[I_{X\geq 105}]$, where $I_{X\geq 105}$ is an indicator random variable

$$I_{X\geq 105} = \begin{cases} 1 & X \geq 105 \\ 0 & \text{otherwise} \end{cases}.$$

Solution: We want to find $\mathbb{P}(X \geq 105)$ after 10 days. We do so by the folling process:

i. Generate $X_i \sim N(0,1)$, $i = 1, 2, ..., 10$ and Compute $\{X_k\}_{k=1}^{N} = Y_0 + \sum_{i=1}^{10} X_i$, where $N =$ 10, 100, 500

ii. Perform the deterministic calculation $g(X_k) = I_{X_k \geq 105} = \begin{cases} 1 & X_k \geq 105 \\ 0 & \text{otherwise} \end{cases}$; which assigns each price a value of 1 or 0 depending on whether it is > 105.

iii. Aggregate the results by computing $\mathbb{P}_N = \frac{\sum_{k=1}^{N} g(X_k)}{N}$. The computations are shown at the end of the document. The results are the following:

$$\mathbb{P}_{10} = 0.10$$

$$\mathbb{P}_{100} = 0.05$$

$$\mathbb{P}_{500} = 0.052$$

(c) Compute this probability analytically. You can use the fact that if X_i's are independent normal random variables with mean μ and variance σ^2, then $\sum_{i=1}^{n} X_i$ is also a normal random variable with mean $n\mu$ and variance $n\sigma^2$.

Solution: We want to find $\mathbb{P}(X \geq 105)$ after 10 days.

By the Central Limit Theorem, As $N \to \infty$, $\mathbb{P}\Big(\sqrt{N}\frac{X-\mu}{\sigma} \leq a\Big) \to \Phi(a)$. We know that $X = 100 + \sum_{i=1}^{10} X_i$.

$$\implies \mu = E[X] = E[100] + E\Big[\sum_{i=1}^{10} X_i\Big] = 100 + 10(0) = 100$$

$$\sigma^2 = Var(X) = Var\Big(100 + \sum_{i=1}^{10} X_i\Big)$$

The inital stock price is independent of the r.v. X_i.

$$\implies Var(X) = Var(100) + Var\Big(\sum_{i=1}^{10} X_i\Big)$$

$$= 0 + 10(1) = 10.$$

$$\implies \mathbb{P}(X \geq 105) = \mathbb{P}\Big(\sqrt{10}\frac{X-100}{10} \geq \sqrt{10}\frac{105-100}{10}\Big) \approx 1 - \Phi\Big(\sqrt{10}\frac{105-100}{10}\Big) \approx 0.056923149$$

5) (10 points) The county hospital is located at the center of a square whose sides are 3 miles wide. If an accident occurs within this square, then the hospital sends out an ambulance. The road network is rectangular, so the travel distance from the hospital, whose coordinates are $(0,0)$, to the point (x_1, x_2) is $|x_1| + |x_2|$ (that is, the 1-norm or the Manhattan distance).

(a) (5 points) If an accident occurs at a point that is uniformly distributed in the square, find the expected travel distance of the ambulance.

Solution: Suppose we randomly generate 2 random variables X, Y such that
X and Y are iid as Uniform(-1.5, 1.5)
One method of doing this is to generate X and Y such that they're Uniform(0, 1) as Matlab already implements this. Then we multiply each value by 3 and then subtract 1.5 from each value to get Uniform(-1.5, 1.5)
Since X and Y are iid, $|X|$ and $|Y|$ are also iid uniform from 0 to 1.5.
$\therefore$ so is their sum but this is instead uniform from 0 to 3.
By definition, the expected value of a uniform distribution Unif(a,b) is $\frac{b+a}{2}$, which is $\frac{3}{2} = 1.5$

(b) (5 points) Compute this expectation using Monte Carlo integration. Report the results using $10, 100, 500$ samples.

Solution: Let:

$$f(x,y) = \begin{cases} 1 & -1.5 \leq x, y \leq 1.5 \\ 0 & \text{otherwise} \end{cases} .$$

And $g(x,y) = |X| + |Y|$
Now, we have the fact that

$$\begin{aligned}
I &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)g(x,y)dxdy \\
&= \int_{-\infty}^{-1.5}\int_{-\infty}^{-1.5} f(x,y)g(x,y)dxdy + \int_{-\infty}^{-1.5}\int_{-1.5}^{1.5} f(x,y)g(x,y)dxdy + \int_{-\infty}^{-1.5}\int_{1.5}^{\infty} f(x,y)g(x,y)dxdy \\
&+ \int_{-1.5}^{1.5}\int_{-\infty}^{-1.5} f(x,y)g(x,y)dxdy + \int_{-1.5}^{1.5}\int_{-1.5}^{1.5} f(x,y)g(x,y)dxdy + \int_{-1.5}^{1.5}\int_{1.5}^{\infty} f(x,y)g(x,y)dxdy \\
&+ \int_{1.5}^{\infty}\int_{-\infty}^{-1.5} f(x,y)g(x,y)dxdy + \int_{1.5}^{\infty}\int_{-1.5}^{1.5} f(x,y)g(x,y)dxdy + \int_{1.5}^{\infty}\int_{1.5}^{\infty} f(x,y)g(x,y)dxdy \\
&= 0 + 0 + 0 + 0 + \int_{-1.5}^{1.5}\int_{-1.5}^{1.5} f(x,y)g(x,y)dxdy + 0 + 0 + 0 + 0 \\
&= \int_{-1.5}^{1.5}\int_{-1.5}^{1.5} f(x,y)g(x,y)dxdy = E(g(x,y)) \\
&\approx \frac{1}{N}\sum_{j=1}^{N} g(x,y)
\end{aligned}$$

We run Matlab to simulate N = 10, 100, and 500 and get the following information
$N = 10 \implies E(g(x,y)) \approx 1.7026$
$N = 100 \implies E(g(x,y)) \approx 1.5769$
$N = 500 \implies E(g(x,y)) \approx 1.4634$
One can see the trend that our estimation approaches our analytically computed value

Solution: For Figure 1 (Figure 2 is Figure 1 code but line 15 is 'for n=1:99')

```
syms I_0;
syms a;
```

```
syms m;
%syms n;
syms finalMatrix;
syms addRow;
syms z;
syms I_n;
syms I_n1;
I_0 = 35;
a = 65539;
m = 2^31;
%n = (1:10000);
finalMatrix = [35 / m, 2293865 / m, 13762875 / m];
for n = 1:9999
    I_n = mod(a * I_0, m);
    I_n1 = mod(a * I_n, m);
    addRow = [I_0 / m, I_n / m, I_n1 / m];
    finalMatrix = [finalMatrix; addRow];
    I_0 = I_n;
end
ax = gca;
ax.FontSize = 14;
bar3(finalMatrix);
title('3D graph of I_n - 1, I_n, I_n + 1');
xlabel('Value of I_n');
ylabel('Value of n');
zlabel('Ratio of I_n and m');
```

For Figure 3

```
syms finalVector;
syms I_0;
syms a;
syms m;
syms I_n;
I_0 = 35;
a = 65539;
m = 2^31;
finalVector = [I_0 / m];
for n = 1:9999
    I_n = mod(a * I_0, m);
    finalVector = [finalVector, I_n / m];
    I_0 = I_n;
end
Ctrs = linspace(0, 1, 30);
h = histogram(finalVector, Ctrs);
xlim([0 1]);
ax = gca;
ax.FontSize = 14;
title('Histogram of RANDU with I_0 = 35');
xlabel('Value between 0 and 1');
ylabel('Frequency of value generated');
```

For Figure 4

```
syms X;
syms funcInput;
syms Inverse;
Inverse = linspace(1, 1000, 1000);
funcInput = linspace(-1000, 1000, 1000);
funcOutput = -2 .* abs(funcInput);
X = rand(1, 1000); %constructs the variables
for n = 1:1000
    %Ensures that the value produced will be in our domain
    %Accuracy is not lost as the probability of any
    %particular point is basically 0
    while X(n) == 0 || X(n) == 1
        X(n) = rand;
    end
end

for n = 1:1000
    if (X(n) < .5)
        Inverse(n) = log(2 * X(n));
    else
        Inverse(n) = -1 * log(2 - 2 * X(n));
    end
end
Z = -100:0.01:100; %Chose to start at -6 and end at 6 b/c it's essentially
    0
                %when not between -4 and 4
Y = exp(-2 .* abs(Z));
hold on
subplot(2, 1, 1);
plot(Z,Y);
xlim([-6 6]);
ax = gca;
ax.FontSize = 12;
title('pdf of the double exponential function')
xlabel('value of random variable')
ylabel('probability of a specific value')
subplot(2, 1, 2);
h = histogram(Inverse);
xlim([-6 6]);
title('Histogram of our randomly generated variables')
xlabel('value of random variable')
ylabel('frequency of random variables')
hold off
ax = gca;
ax.FontSize = 12;
```

Code used for Question 5: 'iterations' is passed as 1. This value is included in the event we want to average our averages

```
function Approx(size, iterations)
answer = rand(iterations,1);
for m = 1:iterations
    X = rand(size, 1);
```

```
    for n = 1:size
        X(n) = (X(n) * 3) - 1.5;
    end
    Y = rand(size, 1);
    for n = 1:size
        Y(n) = (Y(n) * 3) - 1.5;
    end
    g = rand(size, 1);
    for n = 1:size
        g(n) = abs(X(n)) + abs(Y(n));
    end
    f = sum(g);
    answer(m) = f / size;
end
disp(mean(answer));
```

Solution: For Figure 5

```
price_today = 100; %Todays stock price
T = 1:10; %Days after today
N_trajectories = 20; %Number of trajectories to plot
N_t = 10; %time over which the trajectory will cover
deltax = randn(N_trajectories,N_t); %Initialize a matrix made of 20 sets
    of X~N(0,1)
                        %with each column representing 10 days worth of
                            stock
                        %price changes
for trajectory = 1:N_trajectories
    PT = zeros(1,N_t); %Initailize the price trajectory
    PT(1) = price_today + deltax(trajectory,1); %Tomorrow's stock price
    for price = 2:10
        PT(price) = PT(price-1)+deltax(trajectory,price); %Fill the price
            trajectory(PT) at each
                                                                %row(trajectory)
                                                                     of deltax
    end
    plot(T,PT,'Linewidth',4) %plot the trajectory
    hold on
    title('Price Trajectories Over 10 days', 'fontsize',20)
    xlabel('Days','fontsize',18)
    ylabel('Stock Price','fontsize',18)
end
hold off
```

Computations for 4) part (b)

```
format long
price_today = 100; %Todays stock price
N = 500; %Number of samples: That simulate possible prices
        %the stock can have after 10 days.
```

```
P_N = zeros(1,N);
for k = 1:N
    Xi = randn(1,10); %Generate 10 changes in stock price from N(0,1)
                      %over a 10 day period
    X = price_today + sum(Xi); %stock price after 10 days
    P_N(k) = X; %N samples
end
P_N
g_X = zeros(1,N); %Intialize space for deterministic calculation
for k = 1:N
    if P_N(k) >= 105
        g_X(k) = 1; %If the sample is greater than 105 assign 1
    else
        g_X(k) = 0;%If not assign 0
    end
end
g_X
p = sum(g_X)/N %Estimated probability = # of success/ Total Trials
```