Selected Exercises from G&S Section 1.1:

14. Consider the following game:
A fair coin will be tossed until the first time it comes up heads. If this occurs on the *j*th toss, you are paid 2^j dollars. The question remains, however, how much should be paid to play this game.

Consider the following simulation of the game intended to find a reasonable amount that you would be willing to pay, per game, if you will be allowed to make a large number of plays of the game.

```
% Exercise 1.1.14
% Rules of the game:
% (1) A fair coin is tossed until it comes up heads.
% (2) If this occurs on the jth toss, you are paid j dollars
% The question remains: how much would you be willing to pay to play this
% game if you will be allowed to make a large number of plays.

% Objective - determine the average payout after playing the game a large
% number of times.

P=0.5; %probability of a heads
n=100000000 %number of plays of the game
earnings = zeros(1,n); %create an array to sum the earnings of playing the game

for i=1:n
  fortune = 0;
  j=1; %the number of times that the coin has been flipped
  while(fortune == 0)
    if(rand<P) %get a heads
      fortune = 2^j;
    else
      j = j+1;
    end
  end
  earnings(1,i)=fortune;
end

avgwinnings = sum(earnings)/n
```

Running this program several times with *n* = 100,000,000 – the average return varied from $90.1672 to $29.0768. This is astounding! And it implies that coming up with a fair price to pay to play the game under the idealized rules will take more consideration than just the simulation above. It also implies something counter-intuitive to the problem – that someone should pay more than $10 (and potentially any value) to play the game. This does not depend upon the number of plays that you will be allowed (since a large *n* can be seen as averaging any lesser number of plays).

Speaking strictly theoretically (that there can be an infinite number of plays of the game, that the person offering the game has infinite wealth and that the bettor has infinite wealth) it follows that the expected return of the game is given as:

$E = \sum_{i \in \Omega} x_i p_i$ when $\Omega = \{1,2,3,\ldots\}$ is the number of flips after which there is a heads.

$E = \sum_{1}^{\infty} \left(2^i\right)\left(\frac{1}{2^i}\right)$ which is a divergent sum.

This implies that **there isn't necessarily an answer** to how much someone might pay.

This situation, realistically, is immaterial since there is not infinite wealth (and not enough time to play the game that many times). Thus, this requires constraining the game. Assuming that the person offering the game is only willing to pay out a certain amount on a single game, the game has a finite Expected Value which depends on that number and it could be calculated or simulated. Tweaking the program above, if the max is $1024, the fair value is somewhere around $11 and if the max is $2048, then the fair value is somewhere around $12 (so a prudent investor would pay less than that).

15. Tversky reports in a study on the shooting of the Philadelphia 76ers that the number of hot and cold streaks for a shooter was about one would expect by purely random effects. Assuming that a player has a fifty-fifty chance of making a shot and takes 20 shots a game, consider the simulation provided by this program:

```
% Exercise 1.1.15

% Assuming a 50-50 chance of making a basket in basketball
% Rules:
% (1) A player will take 20 shots in a game

% Objective: discern in what fraction of games, the player will have a
% streak of 5 or more baskets.

P=0.5; %probability of making a basket
games = 10000;
streaks = zeros(1,games);
shots = 20;

for i = 1:games
  current = 0; %record the current streak
  highstreak = 0; %record the highest streak in the game
  for j=1:shots
    make = rand;
    if(make < P) % makes a basket
      current = current + 1;
      if(current > highstreak)
```

```
            highstreak = current;
        end
    else
            current = 0;
        end
    end

    if(highstreak > 4) %had 5 or more hits in a row in the game
        streaks(1,i)=1;
    else
        streaks(1,i)=0;
    end
end

streakfrac = sum(streaks) / games
```

Running this program might return that the fraction of games in which the player had a streak of 5 or more baskets is 0.2452. This would seem to imply that if shooting is completely random during all games, that a player would have about a **25%** probability of having a streak of 5 or more baskets within a single game.

16. Estimate, by simulation, the average number of children there would be in a family if all people had children until they had a boy. Do the same if all people had children until they had at least one boy and at least one girl. How many more children would you expect to find under the second scheme that under the first in 100,000 families?

Assuming that boys and girls are equally likely, then consider the following program:

```
% Exercise 1.1.16

% Objective 1: Estimate, by simulation, the average number of children there
% would be in a family if all people had children until they had a boy.

% Objective 2: Do the same if all people had children until they had at
% least one boy and at least one girl.

% How many more children would you expect to find under the second
% objective than under the first in 100,000 families.


n = 100000; %Number of families
P = 0.5; %Assume boys and girls are equally likely.

%Objective 1:
totals = zeros(1,n); %keep track of how many children each family has
for i=1:n
    boy = 0;
```

```
   child = 0;
   while(boy == 0)
      if(rand < P)
         child = child+1; %had a girl
      else
         child = child+1; %had a boy
         boy = 1;
      end
   end
   totals(1,i) = child;
end

average1 = sum(totals) / n

%Objective 2:
people = zeros(1,n);
for j=1:n
   boys = 0;
   girls = 0;
   rascals = 0;
   while((boys+girls) ~= 2)
      if(rand < P)
         rascals = rascals+1;
         girls = 1;
      else
         rascals = rascals+1;
         boys = 1;
      end
   end
   people(1,j) = rascals;
end

average2 = sum(people) / n

difference = sum(people) - sum(totals)
```

□

Running this program yields that the average number of children under the first scheme is **2.0018 or ~2** and that the average number of children under the second scheme is **3.0022 or ~3**. This means that within the 100,000 families, the difference in the total number of children is 100,033 or ~100,000 (more under scheme 2 than scheme 1).

Selected Exercises from G&S Section 1.2:

12. You offer 3 : 1 odds that your friend Smith will be elected mayor of your city.

If the odds are r : s then this means that you are assigning the probability

$P(E) = \frac{r/s}{(r/s)+1} = \frac{3}{4} = 0.75$ to the event that Smith wins.

13. In a horse race, the odds that Romance will win are listed as 2 : 3 and that Downhill will win are 1 : 2. The odds that should be given for the event that either Romance or Downhill wins proceed as follows:

The probability that either Romance or Downhill wins is given as the sum of the probability that Romance wins and the probability that Downhill wins.

$P(E_R) = \frac{2/3}{2/3+1} = \frac{2}{5}$ and $P(E_D) = \frac{1/2}{1/2+1} = \frac{1}{3}$ thus,

$P(E_R)+P(E_D) = \frac{11}{15} = \frac{r/s}{r/s+1}$ and $\frac{r}{s} - \frac{11}{15}\left(\frac{r}{s}\right) = \frac{11}{15}$ or $\frac{r}{s} = \frac{11/15}{1-11/15} = \frac{11}{4}$.

Thus, the odds that should be given to the event that either Romance or Downhill wins are given as **11 : 4**.

16. In a fierce battle, not less than 70 percent of the soldiers lost one eye, not less than 75 percent lost one ear, not less than 80 percent lost one hand, and not less than 85 percent lost one leg. The minimal possible percentage of those who simultaneously lost one ear, one eye, one hand, and one leg is given by considering these traits using an iterative algorithm as follows:

If A% of all soldiers lost an eye, then the minimum percentage that lost an eye is trivially given as *A*.
The minimum overlap of these A percent soldiers with the trait B, that they lost an ear, is given by first assigning losing an ear to the soldiers that did not lose an eye:

$O_{AB} = B-(100-A) = 75-(100-70) = 45$

It follows that the minimum number of soldiers that lost an eye, ear, and hand is given first by assigning trait B to those all of those who do not have both A and B.

$O_{ABC} = C-(100-O_{AB}) = 80-(100-45) = 25$ and it follows that

$O_{ABCD} = D-(100-O_{ABC}) = 85-(100-25) = 10$

Thus, the minimum number of soldiers that simultaneously lost one ear, one eye, one hand, and one leg is **10%**.

17. Assume that the probability of a "success" on a single experiment with n outcomes is $1/n$. Let *m* be the number of experiments necessary to make it a favorable bet that at least one success will occur.

a) In *m* trials, the probability that there are no successes is given by:

$$P\left(\tilde{E}_m\right) = \left(1 - \frac{1}{n}\right)^m$$

If the probability of a success within one trial is given as $P(E) = \frac{1}{n}$ then the probability of *not* obtaining a success is given as $P\left(\tilde{E}\right) = 1 - P(E) = 1 - \frac{1}{n}$.
In *m* trials, these probabilities multiply, since the probability of *n* independent repetitions of an experiment with probability p is p^n yielding:

$$P\left(\tilde{E}_m\right) = \underbrace{\left(1 - \frac{1}{n}\right)\cdots\left(1 - \frac{1}{n}\right)}_{m} = \left(1 - \frac{1}{n}\right)^m$$

(b) (De Moivre) If $m = n\log 2$ then it follows by the Power Law that

$$\lim_{n\to\infty}\left(1 - \frac{1}{n}\right)^m = \lim_{n\to\infty}\left(1 - \frac{1}{n}\right)^{n\log 2} = \left(\lim_{n\to\infty}\left(1 - \frac{1}{n}\right)^n\right)^{\log 2} = \left(e^{-1}\right)^{\log 2} = 2^{-1} = \frac{1}{2}$$

(c) For de Méré's two bets: that a 6 (one die) or boxcars (two dice) would turn up in a certain number of rolls, this method implies:
n_1= 6 and $n_2 = 36$
Then DeMoivre would have found:
$m_1 = 6\log 2 \approx 4.159$ and $m_2 = 36\log 2 \approx 24.953$

DeMoivre would have found that 4 rolls was appropriate for finding a 6, and that 25 rolls was appropriate for finding boxcars – something that de Méré had gotten incorrect.

18. (a) *Proposition*
For events $A_1,\ldots,A_n$:
$P\left(A_1 \cup \ldots \cup A_n\right) \le P\left(A_1\right) + \ldots + P\left(A_n\right)$
Proof
The left side of the inequality, $P\left(A_1 \cup \ldots \cup A_n\right)$ is the sum of $m(\omega)$ for ω either in A_1 or A_2 or ... or A_n. It follows that if $A_1,\ldots,A_n$ are pairwise disjoint subsets of Ω then (as stated in Theorem 1.2 of G&S p.23) $P\left(A_1 \cup \ldots \cup A_n\right) = P\left(A_1\right) + \ldots + P\left(A_n\right)$.
However, if the events are not pairwise disjoint subsets and there is some ω in more than one of $A_1,\ldots,A_n$ then the right side of the inequality will sum $m(\omega)$ more than exactly once, and since $m(\omega) \ge 0$, if $m(\omega) \ne 0$ then the right side of the equation will be greater than that in the left side of the equation – proving the inequality.
(b) *Proposition*
For events *A* and *B*
$P(A \cap B) \ge P(A) + P(B) - 1$
Proof

This inequality can be viewed as
$\sum_{\omega\in A\cap B} m(\omega) \geq \sum_{\omega\in A} m(\omega) + \sum_{\omega\in B} m(\omega) - \sum_{\omega\in\Omega} m(\omega)$ since $\sum_{\omega\in\Omega} m(\omega) = 1$.
The left side of the inequality represents the sum of $m(\omega)$ for ω in both *A* and *B*. It suffices to show that the right side which represents the sums of $m(\omega)$ for ω in *A*, $m(\omega)$ for ω in *B*, and all $m(\omega)$ for ω in Ω, respectively, will be equal to or less than the left side. If $A \cup B = \Omega$, then the right side of the equation counts each ω in precisely one of *A* or *B* only once and subtracts it once. Under the same circumstances, it will count each ω in both *A* and *B* twice, and subtract it once – maintaining equality with the left side. However, if there is some $\omega \in \Omega$ where $\omega \notin A,B$ such that $m(\omega) \neq 0$ then it follows by the previous reasoning that this value will be subtracted exactly once (without being counted); this means that the right side of the equation will be less than the left. Thus, the proposition holds.

19. *Proposition*
If *A*, *B*, and *C* are any three events, then
$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(C \cap A) + P(A \cap B \cap C)$
Proof
The left side of this equation, $P(A \cup B \cup C)$, is the sum of $m(\omega)$ for ω either in *A* or *B* or *C*, thus it suffices to show that the right side of this equation also sums $m(\omega)$. It follows by this reasoning that if an outcome ω is in exactly one of *A*, *B*, or *C*, then it will be counted exactly once in the sums $P(A) + P(B) + P(C)$. If an outcome ω is in precisely two of *A*, *B*, and *C*, then it will be counted twice in those sums, and, thus, the terms $P(A \cap B)$, $P(B \cap C)$, and $P(C \cap A)$ must be subtracted. However, if an outcome is in precisely all three events *A*, *B*, and *C* then it will have been counted three times in the first three terms, subtracted three times in the subsequent three terms (that adjust for elements in two events), and thus must be added back in $P(A \cap B \cap C)$. Thus, each $m(\omega)$ is counted exactly once in the sum of the right side of the equation, and the proposition holds.

28.
a) Intuitively, the probability that a “randomly chosen” positive integer is a multiple of three would be $P = \frac{1}{3}$ since it follows by counting that each third integer would be a multiple of three (i.e. $\not{1}, \not{2}, 3, \not{4}, \not{5}, 6, \ldots$ and so on).

b) Let $P_3(N)$ be the probability that an integer, chosen at random between 1 and N, is a multiple of 3 (since the sample space is finite, this is a legitimate probability. Then it follows that:

$$P_3 = \lim_{N\to\infty} P_3(N)$$

In a finite sample space, the probability that an integer chosen at random between 1 and N is a multiple can be thought of as a piece-wise defined function, whereby

$$P_3(N) = \begin{cases} \dfrac{\left(\frac{1}{3}\right)N}{N}, & N \bmod 3 = 0 \\ \dfrac{\left(\frac{1}{3}\right)(N-1)}{N}, & N \bmod 3 = 1 \\ \dfrac{\left(\frac{1}{3}\right)(N-2)}{N}, & N \bmod 3 = 2 \end{cases}$$

since given any *N*, then $(N-2)(N-1)N$ will always be divisible by 3 and exactly one of *N*-2, *N*-1, or *N* will be divisible by 3. Thus,

$$P_3 = \lim_{N\to\infty} P_3(N) = \begin{cases} \lim_{N\to\infty}\left(\frac{1}{3}\right)\frac{N}{N} \\ \lim_{N\to\infty}\left(\frac{1}{3}\right)\frac{N-1}{N} \\ \lim_{N\to\infty}\left(\frac{1}{3}\right)\frac{N-2}{N} \end{cases} = \begin{cases} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{cases} = \frac{1}{3} \quad \text{which was expected.}$$

c) If *A* is any set of positive integers, let $A(N)$ mean the number of elements of *A* which are less than or equal to *N*. Then define the "probability" of *A* as
$P(A) = \lim_{N\to\infty} \frac{A(N)}{N}$ provided this limit exists.
This implies that $A(N) = |A \cap \{1,\ldots,N\}|$
If *A* is a finite set of size, at most, *k*. Then it follows that $A(N) = |A \cap \{1,\ldots,N\}| \le k$
The probability is then given as: $P(A) = \lim_{N\to\infty} \frac{A(N)}{N} = \lim_{N\to\infty} \frac{c}{N}$ where $c \le k$. Thus,
$P(A) = \lim_{N\to\infty} \frac{c}{N} = 0$.
If *A* is the set of all positive integers, then $A(N) = |A \cap \{1,\ldots,N\}| = N$ and
$P(A) = \lim_{N\to\infty} \frac{N}{N} = 1$.
Thus, the probability of the set of all integers is not the sum of the probabilities of the individual integers in this set. In this way, the definition of probability given is not a completely satisfactory definition.

d) Let *A* be the set of all positive integers with an odd number of digits. Then $P(A)$ does not exist, as follows:

Given that $A(N) = |A \cap \{1,\ldots,N\}|$ then it follows to find an expression for the number of positive integers with an odd number of digits:

$\text{O} = \left(10^m - 1 - 10^{m-1}\right) + \ldots + \left(10^3 - 1 - 10^2\right) + \left(10^1 - 1 - 10^0\right) + 1$ within any subset $\{1,2,3,\ldots,N'\} \quad s.t. \quad 10^m - 1 \le N' \le 10^{m+1} - 1$

Simplifying,

$$\text{O} = \left(\frac{9}{10}10^m - 1\right) + \ldots + \left(\frac{9}{10}10^3 - 1\right) + \left(\frac{9}{10}10^1 - 1\right) + 1$$

As $N \to \infty$, this will also become a limit

$\text{O} = \lim_{\gamma \to \infty}\left[1 + \sum_{m=0}^{\gamma}\left(\frac{9}{10}10^{2m+1} - 1\right)\right]$ and

$P(A) = \lim_{N \to \infty} {A(N)}/{N} = \lim_{\substack{N \to \infty \\ \gamma \to \infty}} \frac{\left[1 + \sum_{m=0}^{\gamma}\left(\frac{9}{10}10^{2m+1} - 1\right)\right]}{N}$ which is indeterminate, and implies that this set will not have a probability.

Thus, under the above definition of probability, not all sets have probabilities.

Selected Exercises from G&S Section 2.1:

5. In order to estimate the area under the curve $y = \frac{1}{x+1}$ within the unit square, consider the following program:

```
% Exercise 2.1.5

% Objective: estimate the area under the graph of y = 1/(x+1) in the unit
% square using 10,000 points chosen at random and using the Monte Carlo
% procedure (i.e. assuming that the probability of a point being under
% the curve in the unit square is equal to the area under the curve in that region).

n = 10000; % number of points chosen
success = zeros(1,n);

for i=1:n
   x = rand;
   y = rand;
   if(y <= (1/(x+1)))
      success(1,i)=1;
   end
end

area = sum(success) / n
```

□

This program chooses *n* randomly selected points and tests whether or not the point is below the curve in question. Running the program results in: area = 0.6899.
In order to calculate the area under the curve exactly, it is given as:

$$A = \int_0^1 \frac{dx}{x+1} = \log(x+1)\Big|_0^1 = \log(2) - \log(1) = \log(2)$$

This finding implies that by the above Monte Carlo method, that $\log(2) \approx 0.6899$ Using an accepted value of 0.6931, this means that the method based in probability was accurate within about 0.005 with *n* = 10,000. This estimate becomes better as the number of random points tested is increased. With *n* = 1,000,000 for instance, the area = 0.6936 or within about 0.0005.

6. To simulate Buffon's needle problem, consider the following program:

```
% Exercise 2.1.6
% Objective: To simulate Buffon's needle problem.

% Choose independently the distance d and the angle at random with d
% between 0 and 1/2 and the angle between 0 and pi/2.
% Check whether d <= (1/2)sin(angle)
% Doing this a large number of times, we estimate pi as 2/a where a is the
% fraction of times that d <= (1/2)sin(angle)
% Run for n = 100, n = 1000, and n = 10000

n = 100; % number of experiments
crosses = zeros(1,n);

for i=1:n
  angle = rand*pi/2;
  d = rand*1/2;

  if((d <= (sin(angle)/2)))
    crosses(1,i)=1;
  end
end
frac = sum(crosses) / n;

estimate = 2/frac
```

This program returns an estimate of π. Where running the program with different values of *n* can yield the following results:

n	$\overline{\pi}$
100	3.2258
1000	3.1746
10000	3.1387
1000000	3.1414

Thus, from the pattern we see that the accuracy of the experimental approximation for π improves as the number of experiments increases.

7. For Buffon's needle problem as considered by Laplace within a grid of horizontal and vertical lines one unit apart, take the following program:

```
% Exercise 2.1.7
```

```
% For Buffon's needle problem, Laplace considered a grid with horizontal
% and vertical lines one unit apart.

% Objective: Simulate this experiment.
% Choose at random, an angle between 0 and pi/2
% and independently, two numbers d1 and d2 between 0 and L/2 (the distance
% from the center of the needle to the nearest horizontal and vertical
% line)

%The needle crosses a line if either d1 <= (L/2)sin angle or d2 <= (L/2)
%cos angle.

%Estimate pi as (4L - L^2) / a where a is the proportion of time that the
%needle crosses at least one line.

%Take L =1
n = 100; % number of experiments
crosses = zeros(1,n);
L=1;

for i=1:n
  angle = rand*pi/2;
  d1 = rand*1/2;
  d2=rand*1/2;
  if((d1 <= (sin(angle)*L/2)) || (d2 <= (cos(angle)*L/2)))
    crosses(1,i)=1;
  end

end

frac = sum(crosses) / n;

estimate = (4*L - L^2)/frac
```

□

This program returns an estimate of π where:
$\bar{\pi} = \frac{4L - L^2}{a}$ where *a* is the proportion of times that the needle crosses at least one line.

Running the program with *n* = 100 (as above) it could return: estimate = 3.1250
Running the program with *n* = 1000 it could return: estimate = 3.1612
Running the program with *n* = 10,000 it could return: estimate = 3.1351
Running the program with *n* = 1,000,000 it could return: estimate = 3.1415

Thus, we see that with relatively small numbers of experiments, the estimates still have a decent possibility of being quite far from the actual value of π (with the numbers above, *n* = 100 yielded a result off by 0.02, as did *n* = 1000). Yet, when the number of experiments is increased significantly, the estimate becomes increasingly better.

As compared with the results in Exercise 6, the results by this method are very similar and tend to occur over similar ranges. Anecdotally, Laplace's method has slightly better values, especially at small *n* – but with increasing *n*, both of them deliver increasingly better estimates.

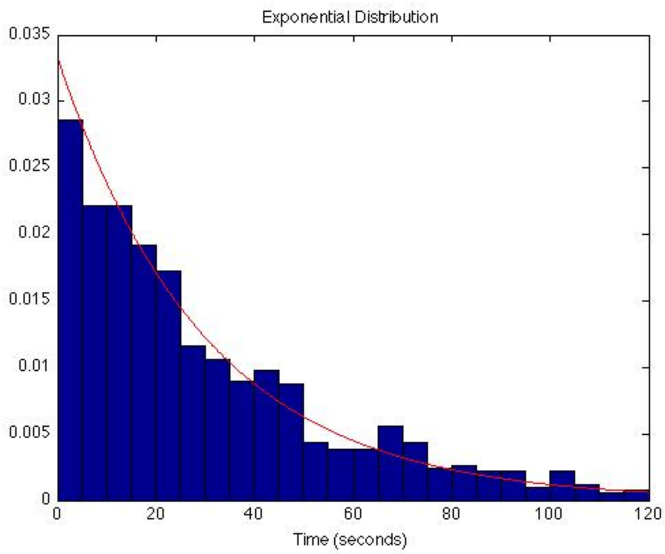
9. Consider the waiting times between cars passing on a highway where the average time between cars is 30 seconds. Assuming that this situation can be simulated by a sequence of random numbers, each of which is chosen by computing $\left(-\frac{1}{\lambda}\right)\log(rand)$ where $1/\lambda$ is the average time between cars or emissions, then consider this program which simulates the waiting time between cars and compares it against the theoretical density function, $f(x) = \lambda e^{-\lambda x} = \left(1/30\right)e^{-\left(1/30\right)x}$.

```
% Exercise 2.1.9

% Exponential Distribution
% Consider the waiting times between cars passing on a highway when the
% average time between cars is 30 seconds.

% Assume that this can be simulated using a sequence of random numbers,
% each of which is chosen by computing (-1/lambda)log(rand) where 1/lambda
% is the average time between cars.

% *This code was originally taken from
%
http://www.math.dartmouth.edu/~doyle/docs/60/matlab/ch2/normaldensitybargraph.m
% and was authored by Professor Doyle
% Objective: to plot bargraph of waiting times for cars and to compare it
% with the theoretical density; exponential distribution.

avgwait = 30; % Average wait time (1/lambda)
lambda = 1/avgwait;
n=1000; %sample size
sample=(-avgwait)*log(rand(1,n)); % Simulate wait times

binstart=0;
binstop=120; % Record from 0 seconds to 120 seconds
binwidth=5; % Make each bar 5 seconds wide
binedges=[binstart:binwidth:binstop];

bincount=histc(sample,binedges);
binprob=bincount/n;
binheight=binprob/binwidth;
bar(binedges,binheight,'histc')

% Plot the theoretical density.
```

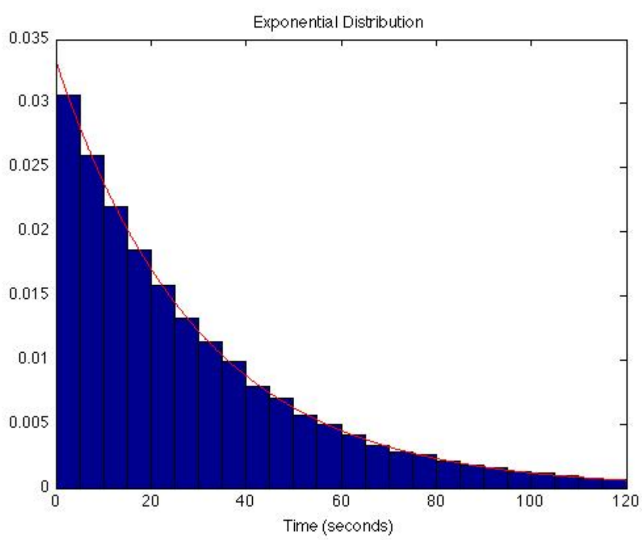
```
hold on
x=binstart:0.1:binstop;
y=(lambda)*exp((-lambda)*x);
plot(x,y,'r')
title('Exponential Distribution');
xlabel('Time (seconds)');
axis([binstart, binstop, 0, 0.035]);
hold off
```

□

Running this program with n = 1000 yields the following graph:



Running the program with n = 100,000 yields this graph:



From this, it is reasonable to see that the theoretical function models the bar graph produced in this way well, especially as the sample size increases.