Name: Syed Bilal Hoda

ID: sh02918

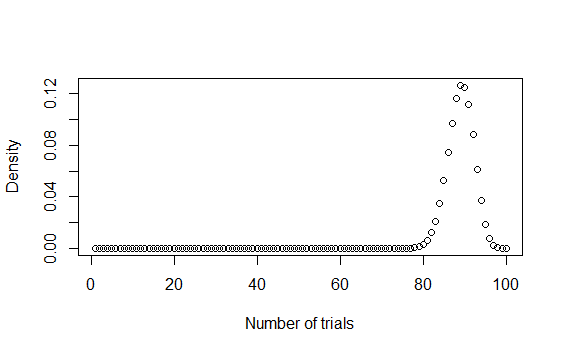
Date: 28/2/2020

Question 1:

Plot a vector of 100 numbers that are drawn from a probability mass function of a binomial distribution with a 0.5 probability of success

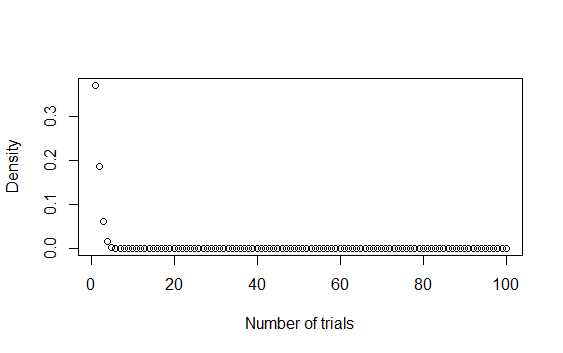
1(a) What happens when the probability is 0.01, 0.1, and 0.89? (2 points)

**Answer:**



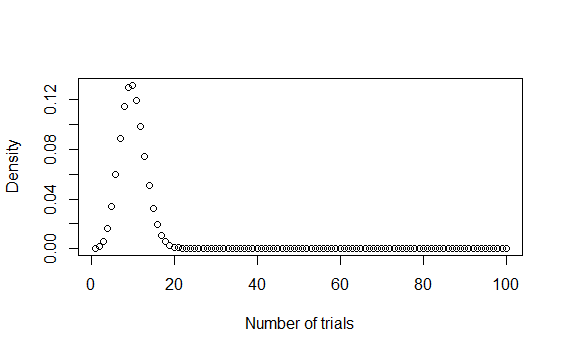
When the probability is 0.5 the plot is symmetric so we can say the probability of 50 successes in 100 trials is the highest.

**Probability 0.01:**



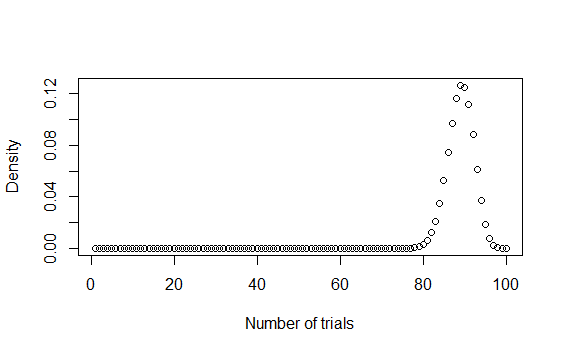
The probability of getting 1-3 successes would be higher whereas getting 10 or more successes would be 0.

**Probability 0.1:**



The probability of getting 1-20 successes would be higher whereas getting more than 22 successes would be 0.

**Probability 0.89:**



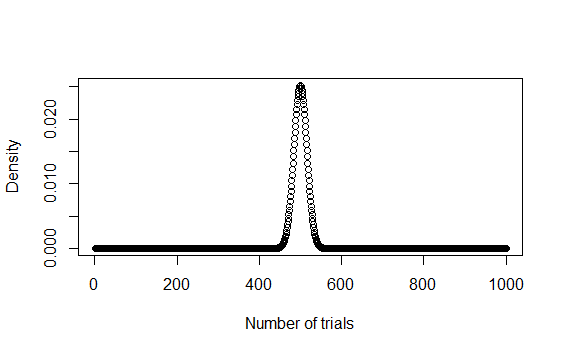
When the probability is 0.89 the probability of getting 80-90 successes would be higher in a 100 trials.

1(b) What happens when the size is increased to 1000, 10000, 100000? (2 points)

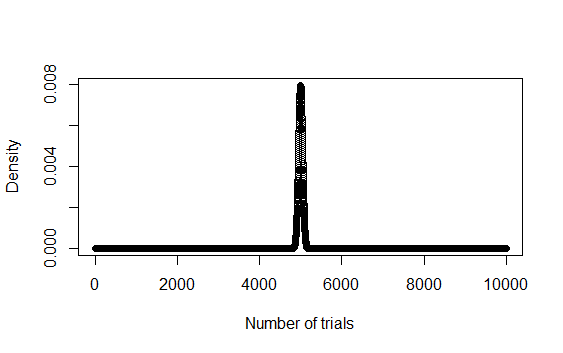
**Answer:**

We keep the probability fixed (0.5) and change the sizes.

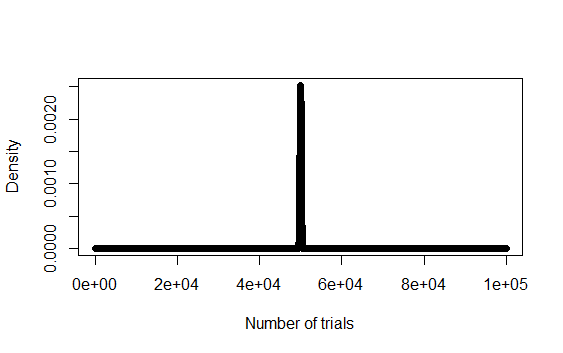
**Size 1000:**



**Size 10000:**



**Size 100000:**



As the size increases the distributions peak at the same fractional distance from the origin, N/2. The peak in the distribution gets sharper and the width of the curve also reduces in other words the standard deviation reduces.

Reference: http://www.pas.rochester.edu/~stte/phy104-F00/notes-5.html

Question 2

Suppose widgit weights produced at Acme Widgit Works have weights that are normally distributed with mean 17.46 grams and variance 375.67 grams.

2a) What is the probability that a randomly chosen widgit weighs more than 19 grams? Hint: What is P (X > 19) when X has the N (17.46, 375.67) distribution? (1 point)

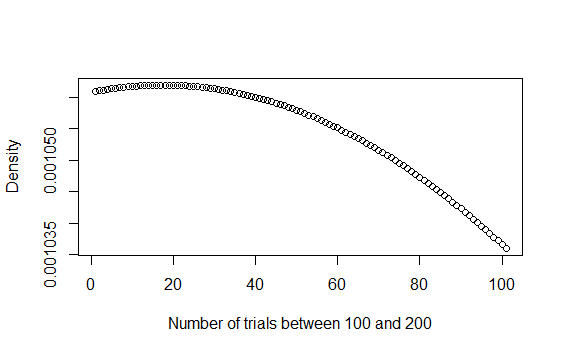
Note: R wants the s. d. as the parameter, not the variance.

**Answer:**

The probability is 0.4983646.

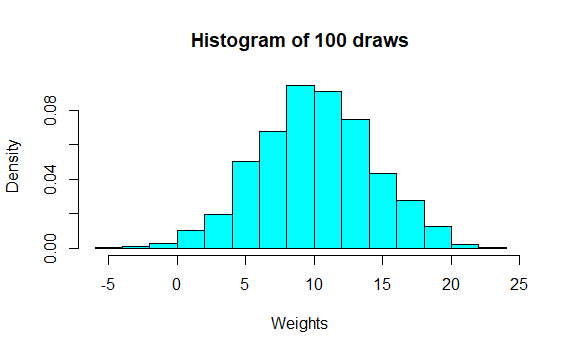
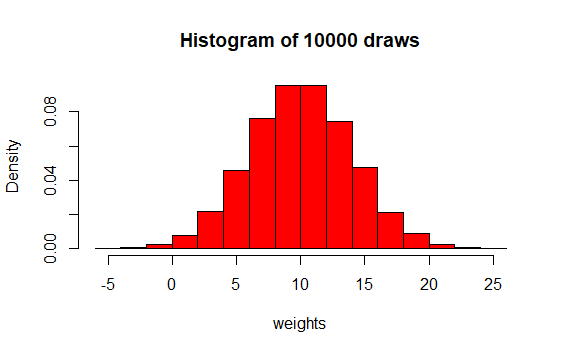
2b) Please plot the probabilities of outcomes for 100 values between 2 and 200? (1 point)

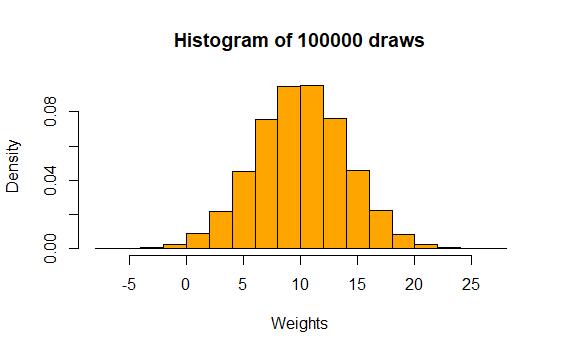
**Answer:**



2c) Increase the number of random draws from this distribution to 10000 and 100000 What does this distribution look like? In which interval does most of the data lie in? (2 points)

**Answer:**

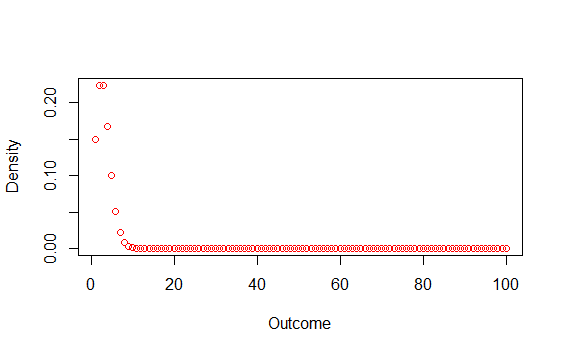
 



The width of the bars or bin size has decreased in 10000 and the width of the bars in case of 100000 has further decreased while the standard deviation is fixed. This is due to the fact as the number of draws increase more and more data gets packed into each interval. The distribution is a bell curve. Most of the data lies in the 8-12 interval.

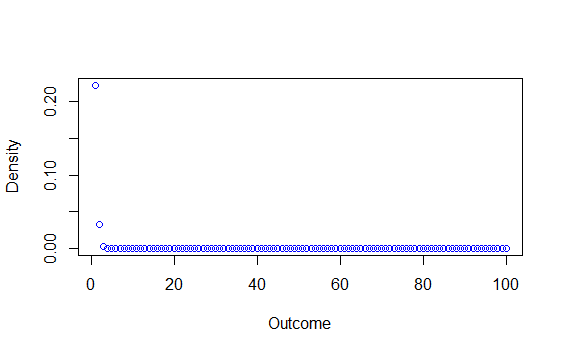
Question 3

For rate parameter 3:



3a) What happens to the shape of the distribution when the rate parameter is 0.3? (1 point)

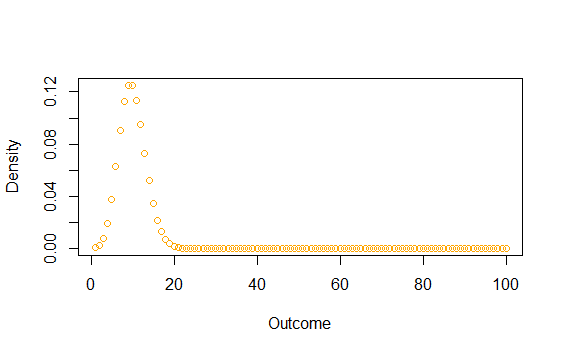
**Answer:**



For the rate parameter 0.3 the shape of the distribution is further skewed to the right. The probability of 0.3 outcomes near the arrival rate (0.3) is higher whereas the probability of rest of the outcomes i.e. more than 5 outcomes would be close to 0.

3b) What happens to the shape of the distribution when the rate parameter is 10? (1 point)

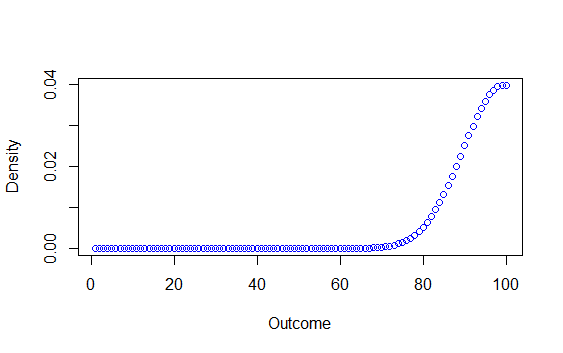
**Answer:**



For the rate parameter 10 the shape of distribution is lesser skewed to the right. We can see a proper curve is being formed. The peak of the curve is at 10 which has a probability of 0.12. The probability of outcomes near the arrival rate (10) is higher whereas the probability of rest of the outcomes i.e. more than 20 outcomes would be close to 0.

3c) What happens to the shape of the distribution when the rate parameter is 100? (1 point)

**Answer:**



For the rate parameter 100, the shape of the distribution is skewed to the left. We can see a curve is being formed from 80 - 100. The peak of the curve is at 100 which has a probability of 0.04. The probability of outcomes near the arrival rate (100) is higher whereas the probability of rest of the outcomes i.e. less than 70 outcomes would be close to 0.

**To sum up:**

When we are changing the rate parameter the mean and variance of the distribution is changing

So as we increase the rate parameter the probability of observing more events would increase

and if we decrease the rate parameter then the probability of observing less events is higher.

3d) Where are the mass of all points on the distribution? (1 point)

**Answer:**

The mass is around the rate parameter:

For lambda = 3 the mass will be from 0 to 5

For lambda = 0.3 the mass will be from 0 to 0.5

For lambda = 10 the mass will be from 0 to 20

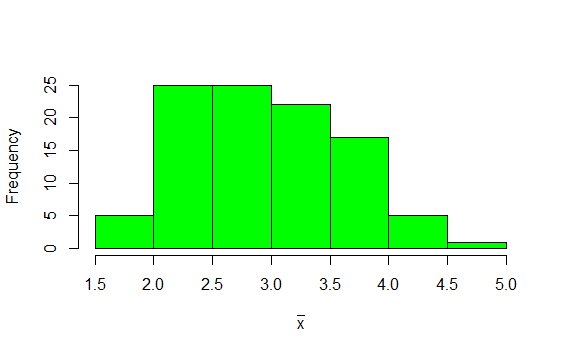
For lambda = 100 the mass will be from 80 to beyond 100.

Question 4

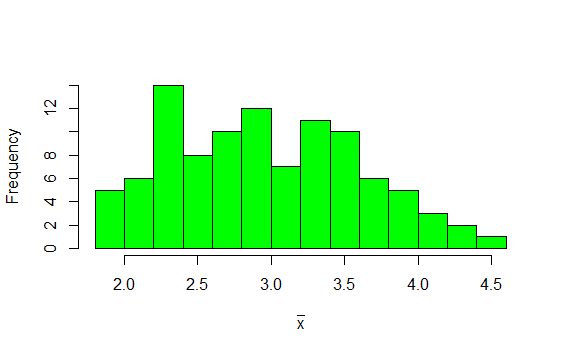
4a) What does this distribution look like? (1 point)

**Answer:**

With breaks 5:

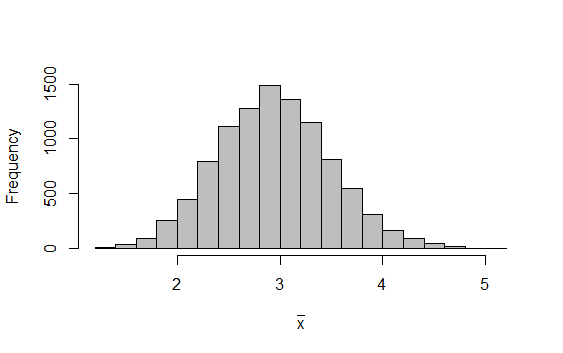


With breaks 10:



The distribution looks like a Poisson distribution because it is somewhat skewed to the right. Secondly, since we are plotting 100 means of 10 randomly sampled values from Poisson distribution the distribution is starting to become a little symmetric. For larger samples the effect of Central Limit Theorem would become more evident.

4b) What happens to the shape of the distribution if we draw 10000 points instead of 100? (1 point)



The shape of the distribution becomes bell shaped. In this case we get a normal distribution because we are plotting 10000 means of Poisson so the normal curve shows the effect of central limit theorem.

Question 5

5a) Based on your sampling survey, which of the distributions that we learned in class are relevant to the kind of data that you have collected? (1 points)

**Answer:**

The Poisson distribution seems relevant to my data because I noted the number of orders for four categories for a fixed interval of time for each day of the week. So the number of orders in each category essentially depends on the number of people coming in to purchase fries. Hence, the distribution relevant to my data is Poisson.

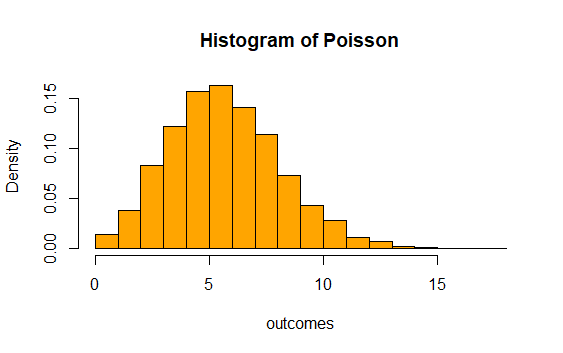
5b) Determine the mean, sd, probability of successes and/or rate parameter (depending on the type of data collected) for your data. Use this information to draw 10000 random variables from the relevant probability distribution. Plot these random draws as a histogram (2 points)

**Answer:**

The mean of the data is 6.17641.

The standard deviation is 5.615106.

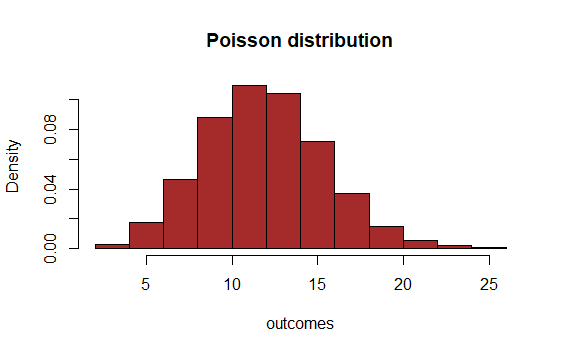
The rate parameter is 6.1764706.



5c) Based on your distribution, where do the center/mass of outcomes lie? (1 point)

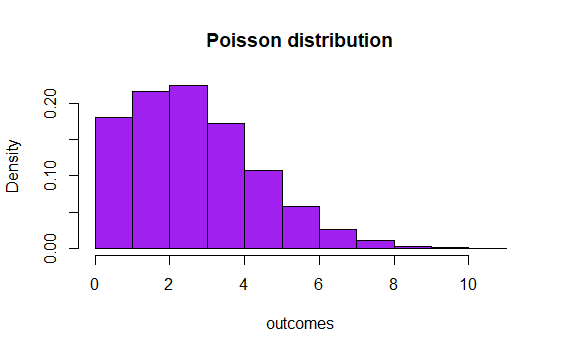
The center or the mass is around the rate parameter or one could say between the intervals 3-7

5d) What happens to the shape of your distribution when you change your mean/sd/rate parameter/probability of successes to twice that you observed in your data? (1 point)



By changing the rate parameter to twice (12.35294) the shape of the distribution becomes more symmetric and it looks more like a bell curve with some skewness to the right.

5e) What happens to the shape of your distribution when you change your mean/sd/rate parameter/probability of successes to half that you observed in your data? (1 point)



By changing the rate parameter to half (3.088235) the shape of the distribution becomes more skewed to the right. It is still a Poisson distribution. The mass is concentrated around the rate parameter which is 3.088235