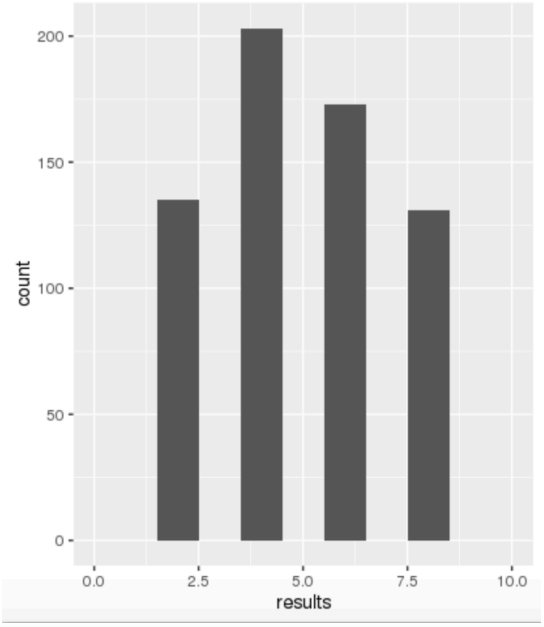
ST 516 - Homework 3
student Paul ReFalo 10/14/17

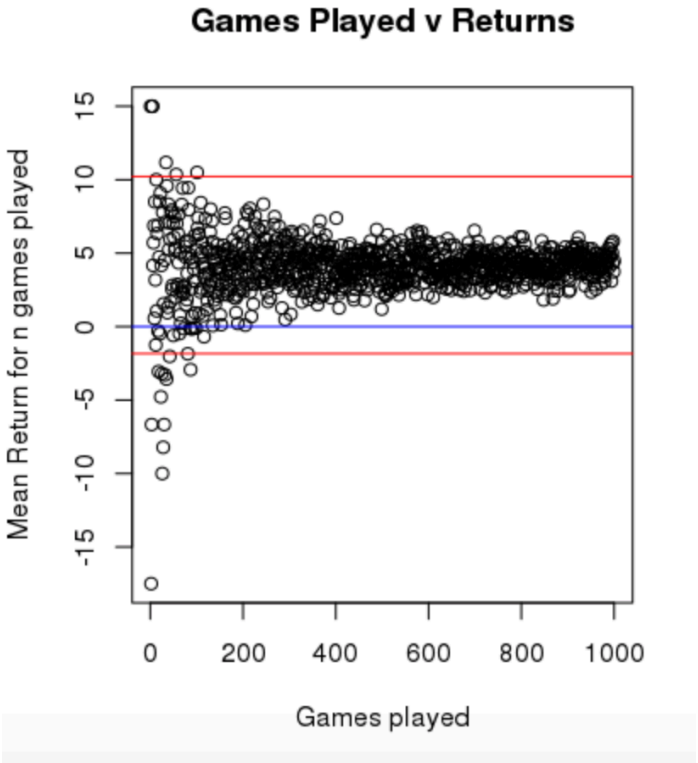
1. Consider this game: You roll one die, and lose $50 if you roll a 1, but win $15 if you roll anything else.

Use R simulations to estimate your expected win/loss value for one roll.




```
> results <- replicate(1000,
mean(replicate(30,
play(silent=TRUE))))
> qplot(results, binwidth = 1,
xlim = c(0,10))
> mean(results)
[1] 4.1905
```

The average here is ~ 4.17 for one roll.



I also make another plot of Returns as a function of games played which gives this plot. Extra credit plot?

```
> for (n in xValues)
{loopResults[n] <-
mean(replicate(n,
play(silent=TRUE)))}

> plot(xValues, loopResults,
xlab="Games played",
ylab="Mean Return for n games
played", main="Games Played v
Returns")
```

How many times did you play to find your estimate?

I first played 30 times, took the avgerage and did that 1000 times. So in the end, I played 30,000 games.

How precise do you think your answer is?

The variance of this data set is:

```
> var(results)
[1] 19.3078
```

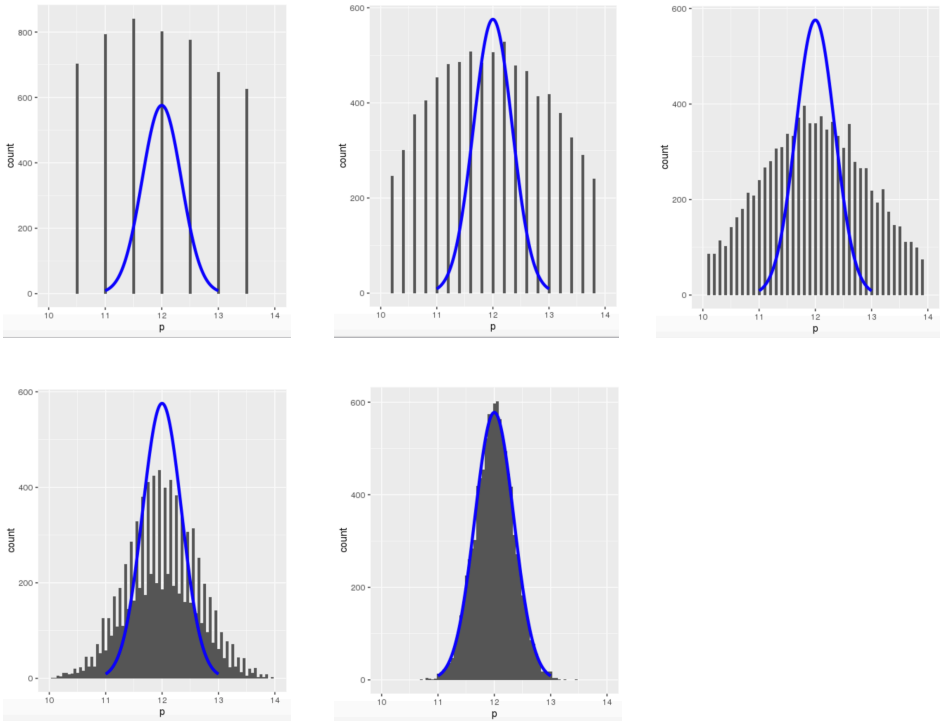
From the second plot you can see that the variance decreases as the number of games played.

How much would you be willing to pay to play this game?

I would certainly be willing to play this game since the mean result is a positive value. I would want to play as many games as possible as the distribution narrows for more games played (see graph 2 'Games Played v Returns')

Continue on next page.

2. Central Limit Theorem question. For this question, choose one of the following distributions, and replicate the exploration from the lab with sample sizes of 2, 5, 10, 30 and 100:




```
> sample_size <- 100
> p <- replicate(10000, mean(rpois(sample_size, 12)))
>
> sd_p <- sd(p)
> x1 <- seq(11, 13, length = 1000)
> y1 <- dnorm(x1, 12, sqrt(12)/10)
>
> qplot(p, binwidth = .05, xlim=c(10,14)) +
+   geom_line(aes(y = 10000*0.05*y1, x = x1), size = 1.5, color = "blue")
```

Fascinating to see how sample size affects the distribution. I picked Poisson with a parameter of 12.

Here are the plots in ascending order with sample sizes = [2, 5, 10, 30, 100]. As the sample size increases, whose mean is seen as grey bars, the shape of the distribution of these bars more closely matches a normal distribution, seen in blue, centered on the same point of 12. A Poisson distribution with a sample size of 100 closely matches the Normal distribution for the same data set.

3. Describe why we do not usually know the population mean. What statistic do we usually use to estimate the population mean and why?

We usually don't know the population mean because it is impractical to gather data on all members of the data set. For example, to know the average cholesterol of 40 men in the USA would require a enormous amount of logistics, testing, and money. But to sample this population and estimate the Population mean would be much more realistic.

To estimate the population mean we often used the long-run average of many observations. As it says in the lecture slides:

"As the sample size *n* gets larger and larger, the sample mean of *n* independent observations gets closer and closer to the population mean, μ."

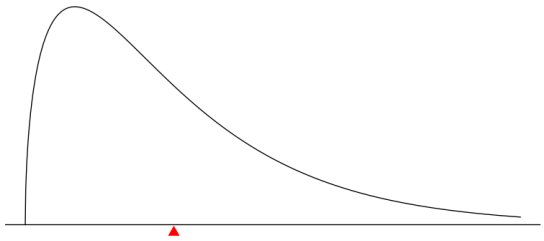
4. The above plot represents the distribution function for a random variable X:

. (a) Is this distribution skewed? Why or why not? If so, what is the direction of the skew?

Skewed because the curve is not symmetrical. The skew is to the left, or lower x-values, of center.

. (b) Describe the relationship between the mean and the median for this distribution. If they are different from each other, why?

The mean is the average, or the point at which it could be balanced on a point like this:



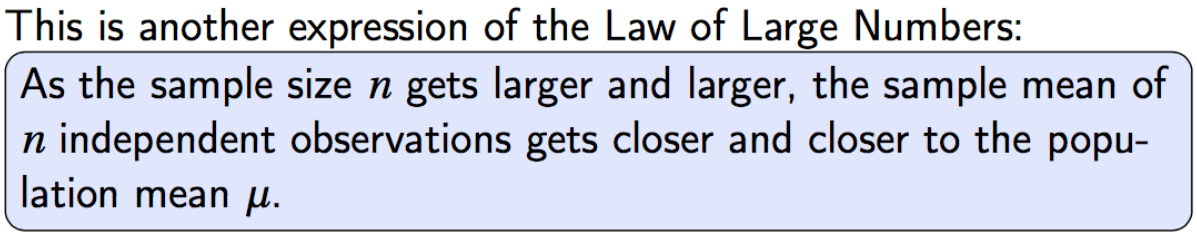
The median is the middle value in the list. For this curve, the middle value will be to the left of the mean.

. (c) How many modes does this distribution have? Is it possible for a distribution to have more than one mode, median, or mean? Why?

A distribution can have only one mean and only one median, but can have multiple modes. This distribution has one mode, or most common value, which is the highest peak in the curve. Yes, it is possible for a distribution to have multiple modes, or peaks, that tie for frequency.

6. Participants in a cake frosting race must frost identical cakes to make them look like a given picture. Their cake-frosting times are recorded and their mean cake-frosting time is calculated. Ten cake makers are selected at random and their mean cake-frosting time is 8.73 minutes. One thousand cake makers are selected at random and their mean cake-frosting time is 10.19 minutes. One hundred thousand cake makers are selected at random and their mean cake-frosting time is 9.97 minutes. What would you guess is the mean cake-frosting time of the population of cake makers? Why?

I would guess that the mean of the population is close to 9.97 minutes because of the Law of Large Numbers. From the lecture 1, slide 11:



7. Pianos are rated on discrete a scale from 0 to 10 for having the correct pitch, with 10 being the best score. If larger and larger random samples of pianos are selected for a rating, and their mean pitch score is calculated, what will the distribution of the sample mean start to look like? What is the theorem that is responsible for this property? Does the existence of this phenomenon depend on the distribution of the population?

The distribution will like more like the Normal distribution the larger the sample size gets. This is the Central Limit Theorem (CLT). The CLT does not depend on the distribution of the population. From lecture 5, slide 5.

The **Central Limit Theorem** tells us about the sampling distribution of the sample mean.

Let $X_1, X_2, \ldots, X_n$ denote an independent sample of random variables from a population that has unknown mean μ and unknown variance σ^2. Then, for large sample sizes, the sampling distribution of the sample mean, $\overline{X}$, is approximately Normal with mean μ and variance σ^2/n.

The beauty of this theorem is that it holds true *regardless of the shape* of the population distribution from which the sample is taken.