Central Limit Theory and the Exponential Distribution

Andrew Crisp

November 13, 2015

Overview

Through a simulation, the Central Limit Theory will be demonstrated. The mean and variance of the exponential distribution will be simulated and compared to calculated values.

Simulations

During the simulation, lambda will be set to a controlled 0.2. 1000 passes will be made, collecting 40 random exponential distribution samples. After collecting the sample values, the mean of each group will be collected and indexed for easy inspection.

```
lambda <- 0.2
nosims <- 1000
n <- 40
rexpValues <- NULL
means <- NULL
for (i in 1:nosims) rexpValues <- rbind(rexpValues,rexp(n = n, rate = lambda))
for (i in 1:nosims) means <- rbind(means, mean(rexpValues[i,]))
means <- as.numeric(means)
```

Sample mean versus Theoretical mean

The mean of an exponential distribution is 1/lambda. This theoretical mean can be calculated easily and compared with the actual collected sample mean.

```
calcMean <- 1/lambda
simMean <- mean(rexpValues)
CLTMean <- mean(means)
```

The calculated mean is 5.

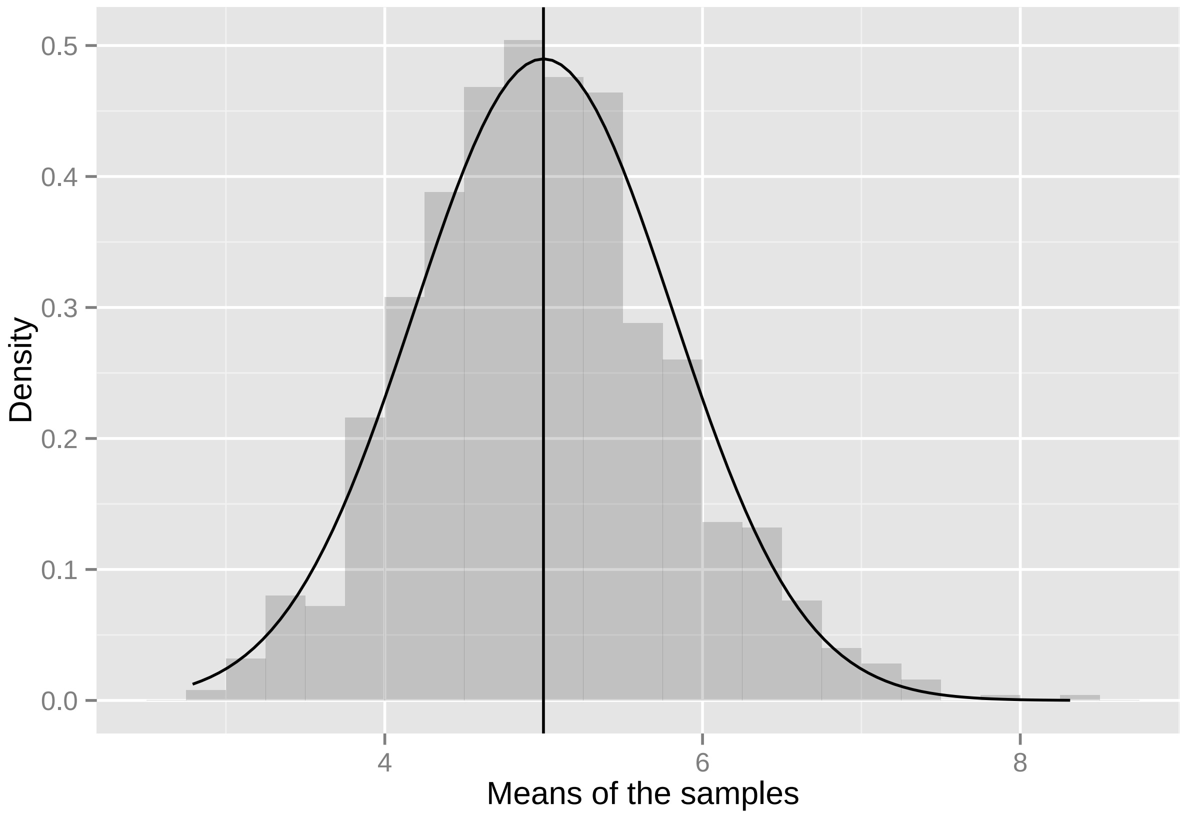
The Central Limit Theory says that the mean of the a sample should approximate the mean of the population. The mean of all 1000 means simulated is 4.9988197.

The actual mean of all the simulated values is 4.9988197.

```
require(ggplot2)
```

```
ggplot() +
  aes(means) +
  geom_histogram(aes(y = ..density..), binwidth=.25, alpha=.2) +
  geom_vline(xintercept=mean(means)) +
  stat_function(geom="line",
```

```
                fun=dnorm,
                arg=list(mean=mean(means),sd=sd(means))) +
  xlab("Means of the samples") +
  ylab("Density")
```



The average of all the means is demonstrated to be at 4.9988197 and that the distribution of all the averages is approximately normal.

Sample variance versus Theoretical variance

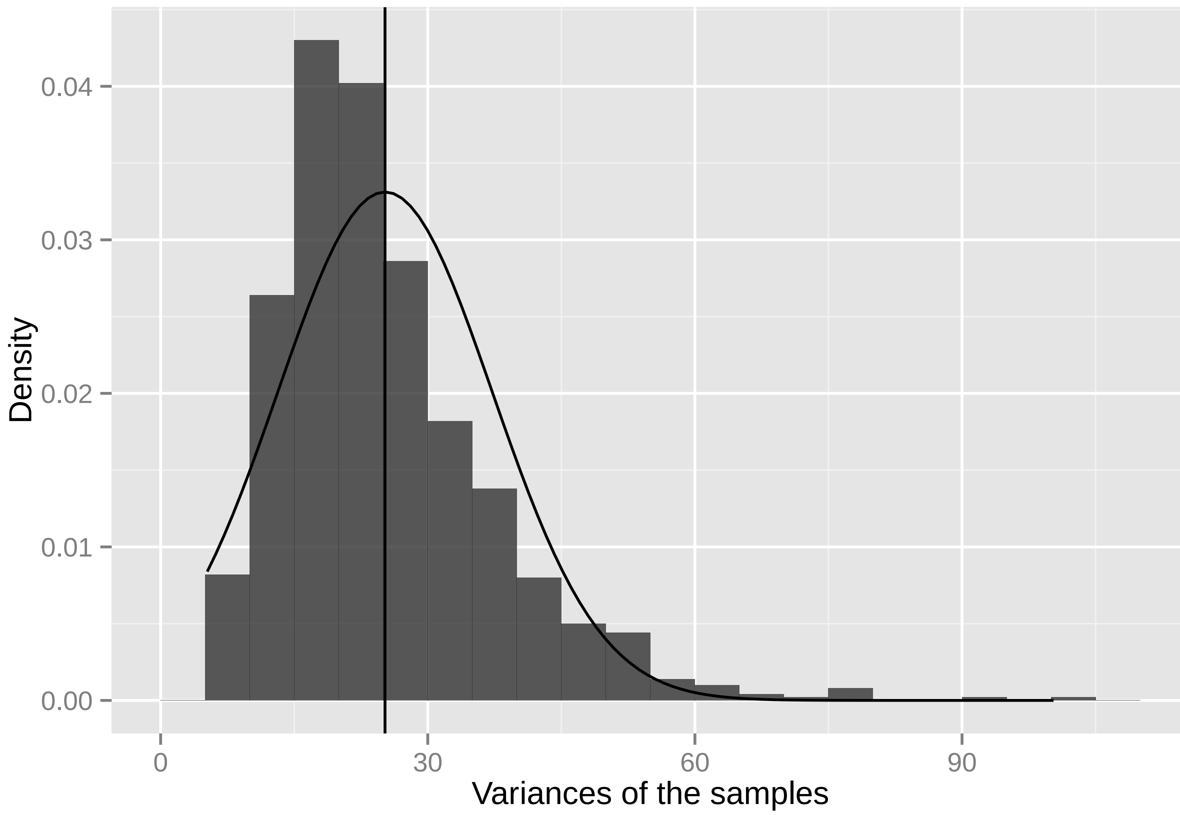
Similar to the mean, the standard deviation of an exponential distribution is 1/lambda. The variance is the square of the standard deviation. However, the variance of the samples should approach that of a normal distribution.

```
calcVariance <- (1/lambda)^2
n <- 40
n40CollectionOfVariances <- NULL
for (i in 1:nosims) n40CollectionOfVariances <- rbind(n40CollectionOfVariances, var(rexpValues[i,]))
simVariance <-mean(n40CollectionOfVariances)
```

The calculated (theoretical) variance is 25.

The average variance of the simulations is 25.199812.

```r
ggplot() +
  aes(n40CollectionOfVariances) +
  geom_histogram(aes(y = ..density..), alpha=.8, binwidth=5) +
  geom_vline(xintercept=mean(n40CollectionOfVariances)) +
  stat_function(geom="line",
                fun=dnorm,
                arg=list(mean=mean(n40CollectionOfVariances),sd=sd(n40CollectionOfVariances))) +
  xlab("Variances of the samples") +
  ylab("Density")
```

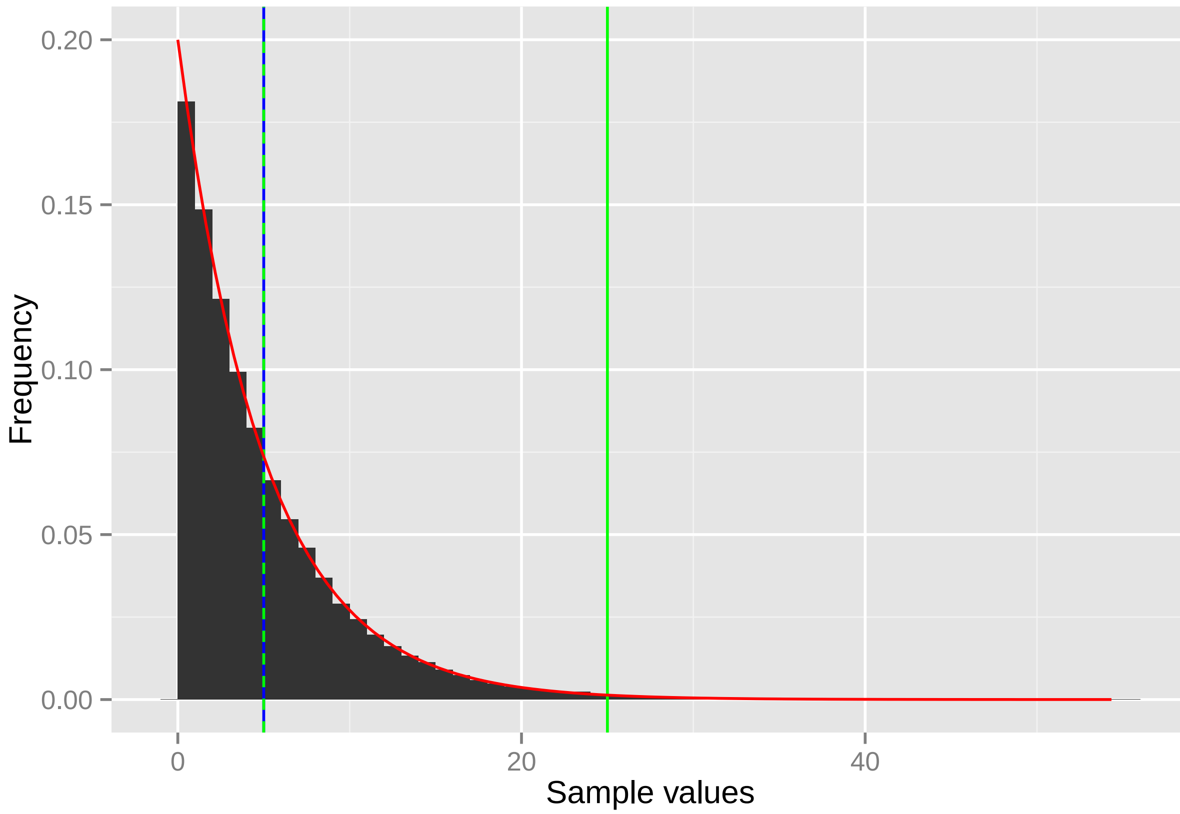


The average variance from the samples is approaching the calculated variance of the exponential distribution of 1/lambda, 25.

Distribution

```r
ggplot() +
  aes(as.numeric(rexpValues)) +
  geom_histogram(binwidth=1, aes(y=..density..)) +
  geom_vline(xintercept=calcMean, color='blue', labels="Mean") +
  geom_vline(xintercept=calcVariance, color='green', labels="Variance") +
  geom_vline(xintercept=sqrt(calcVariance),
             color='green',
             linetype='dashed',
```

```
            labels="Standard Deviation") +
  stat_function(geom="line", fun=dexp,arg=list(rate=lambda),color='red') +
  xlab("Sample values") +
  ylab("Frequency")
```



The exponential distribution can be seen to not be anywhere near a normal distribution. The central limit theorum, though, is useful in identifying the marked values, mean, standard deviation, and variance. By sampling the distribution and examining the samples, the desired descriptive values emerge. It is demonstrated that the average of sample means is approximately the population mean. Similarly, the average of sample variances is approximately the population variance. The distribution of the population proves to not influence this behaviour