Exponential Distribution Analysis

Enrique El Gato

Overview
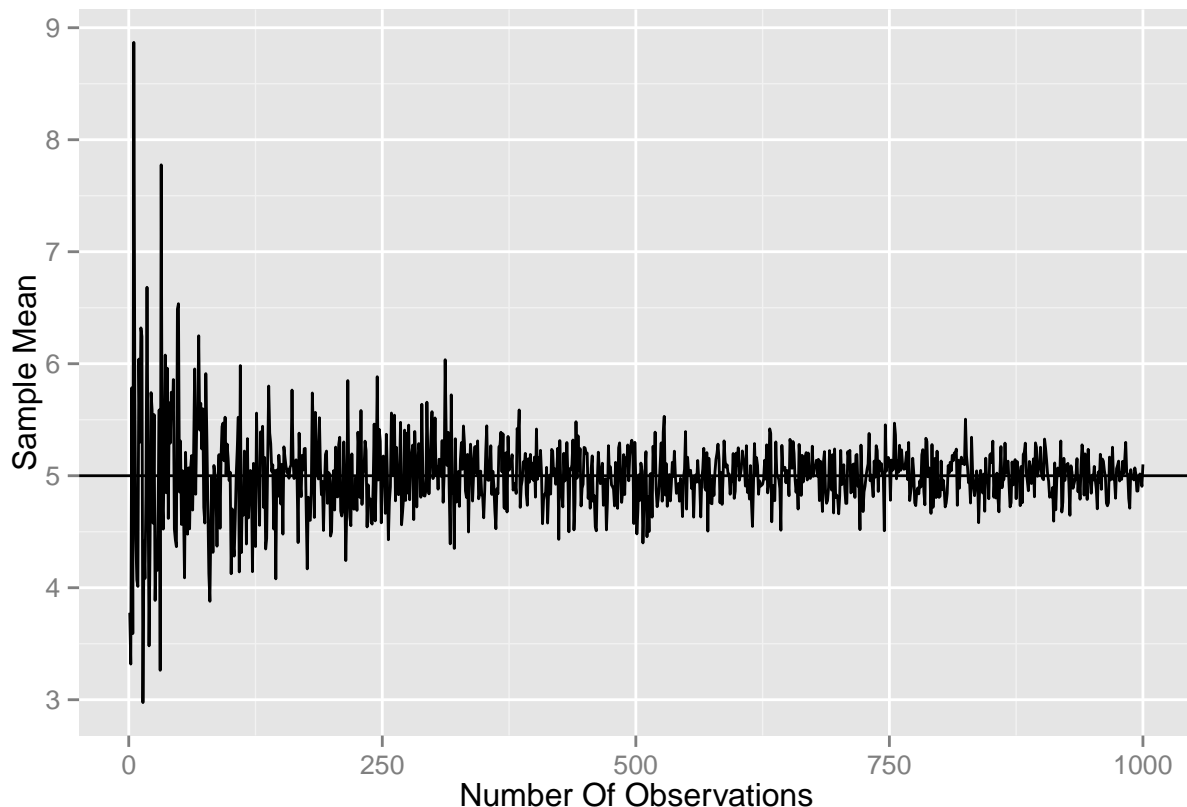
This analysis will examine the Exponential Distribution with regard to:

- That per the Law of Large Numbers (LLN) sample means and variances off iid samples are consistent estimators and will converge on their population values as samples sizes are increased.
- That per the Central Limit Theorem distribution of averages of iid variables become standard normal as sample sizes increase.

Sample Mean versus Theoretical Mean

The graph below shows that as the sample size increases to 1000, the sample means gravitate towards the theoretical mean of 1/lamdba which in this case is 5 as the lambda is .2.

```
sampleSize <- 1000;lambda <- .2;theoreticalMean <- 1/lambda;means = NULL;
set.seed(1)
for (i in 1 : sampleSize) means = c(means, mean(rexp(i,lambda)))
library(ggplot2)
g <- ggplot(data.frame(x=1:sampleSize,y=means),aes(x=x,y=y)) + geom_hline(yintercept=theoreticalMean) +
g
```
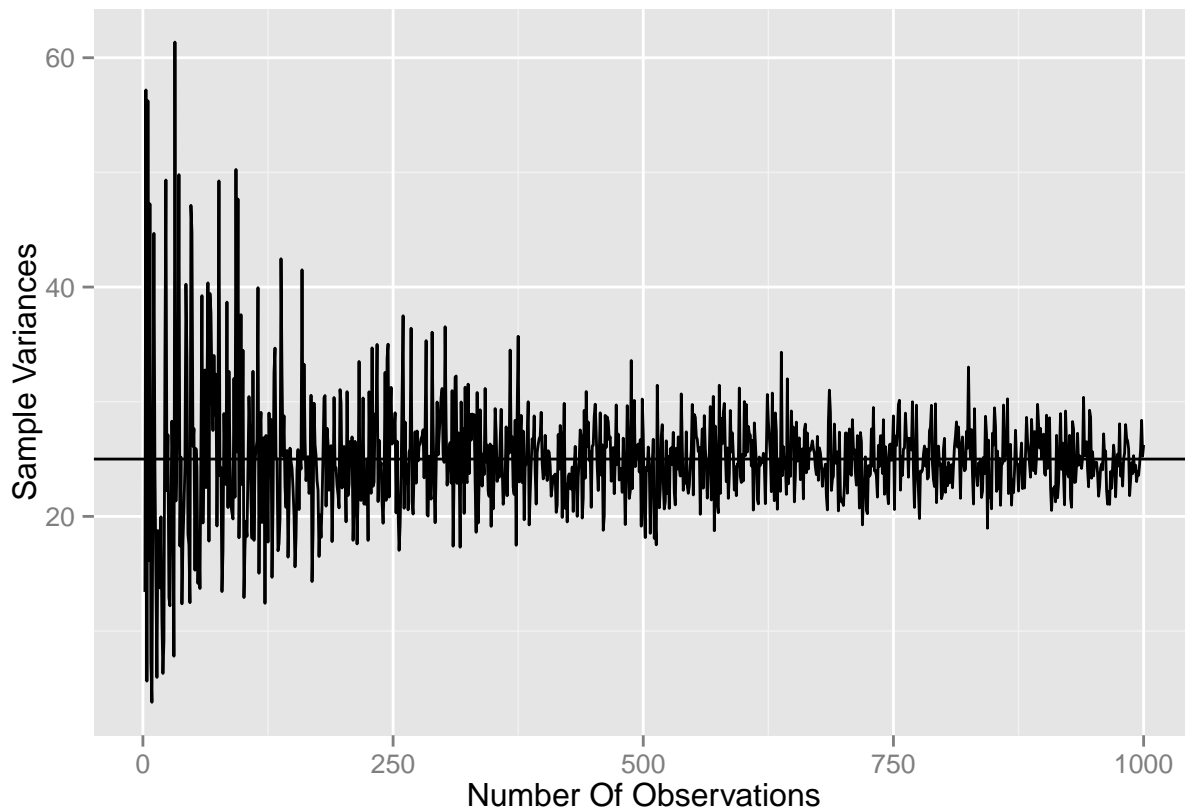
Sample Variance versus Theoretical Variance

The graph below shows that as the sample size increases to 1000, the sample variances gravitate towards the theoretical variance of 1/lamdba which in this case is 25 as the lambda is .2 and standard deviation is 5.

```r
rm(list=ls())
sampleSize <- 1000
lambda <- .2
theoreticalSD <- 1/lambda
theoreticalVAR <- theoreticalSD^2
vars = NULL
set.seed(1)
for (i in 1 : sampleSize) vars = c(vars, var(rexp(i,lambda)))
library(ggplot2)
g <- ggplot(data.frame(x=1:sampleSize,y=vars),aes(x=x,y=y))
g <- g + geom_hline(yintercept=theoreticalVAR)
g <- g + geom_line(size = .5)
g <- g + labs(x="Number Of Observations", y="Sample Variances")
g
```

```
## Warning in loop_apply(n, do.ply): Removed 1 rows containing missing values
## (geom_path).
```

Distribution

When taking a 1000 samples of size 40, the distribution of the means is approximately normal. The histogram below shows the distribution of the means. A normal distribution has been overlaid on the bar graph. The distribution of the means of the same Exponential Distributions appears to be normal (as the CLT predicts) with a mean of 5 (as the LLN predicts).

```
rm(list=ls())
set.seed(1)
means = NULL
set.seed(1)
for (i in 1 : 1000) means = c(means, mean(rexp(40,.2)))
h <- hist(means)
xfit<-seq(min(means),max(means),length=1000)
yfit<-dnorm(xfit,mean=mean(means),sd=sd(means))
yfit <- yfit*diff(h$mids[1:2])*length(means)
lines(xfit, yfit, col="red", lwd=2)
legend("topright", c("Mean Hist", "Normal Dist"), col = c("black", "red"), lty = 1,bty="n")
```

Histogram of means