Statistical Inference Simulation Exercise by PP

Overview

In this project we investigate the exponential distribution and compare it with the Central Limit Theorem. We show the sample mean and compare it to the theoretical mean of the distribution; demonstrate how variable the sample is (via variance) and compare it to the theoretical variance of the distribution; finally we show that the distribution is approximately normal.

Simulations

We take a 1000 simulations of the mean of exponential distributions of sample size 40. Our lambda will be 0.2 throughout the entire project. Note that we set the seed for the random selection so that it is reproducible.

```
set.seed(1)
lambda = 0.2;
sampleMean = NULL;
for (i in 1 :1000) sampleMean = c(sampleMean, mean(rexp(40,lambda)));
```

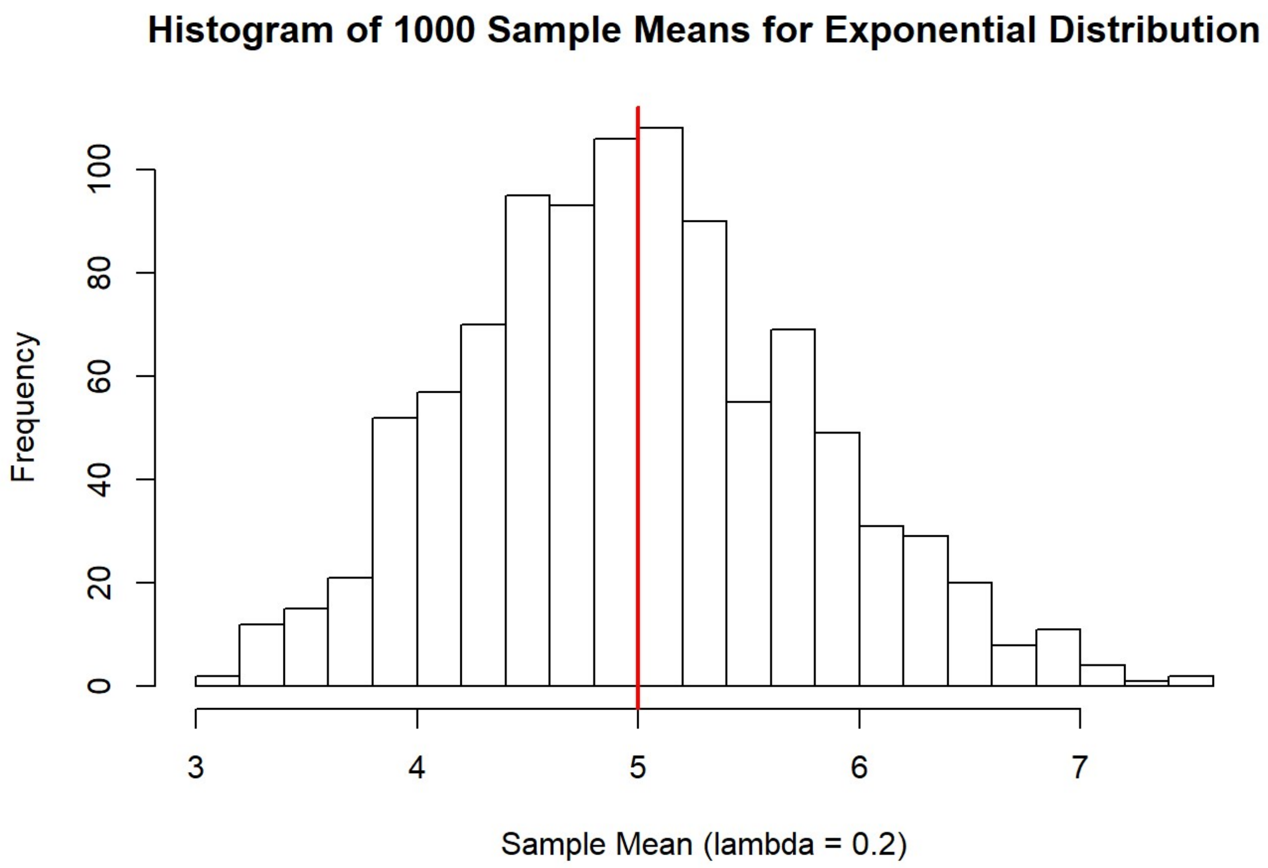
We take another 1000 simulations of the variance of exponential distributions of sample size 40. We use the same seed as last time time so that we are considering the variances for the same samples for which we examined the means.

```
set.seed(1)
sampleVariance = NULL;
for (i in 1 :1000) sampleVariance= c(sampleVariance, var(rexp(40,lambda)));
```

Sample Mean versus Theoretical Mean

We see from the following graph that the sample mean follows a gaussian distribution with a mean of about 5. This fits with the theory since the central limit theorem states that as the sample size increases the distribution of the sample mean will have a normal distribution about the expected value. The expected value of the sample mean is the expectation of the original exponential variable, which in this case is 5. We have included a red vertical line to illustrate this.

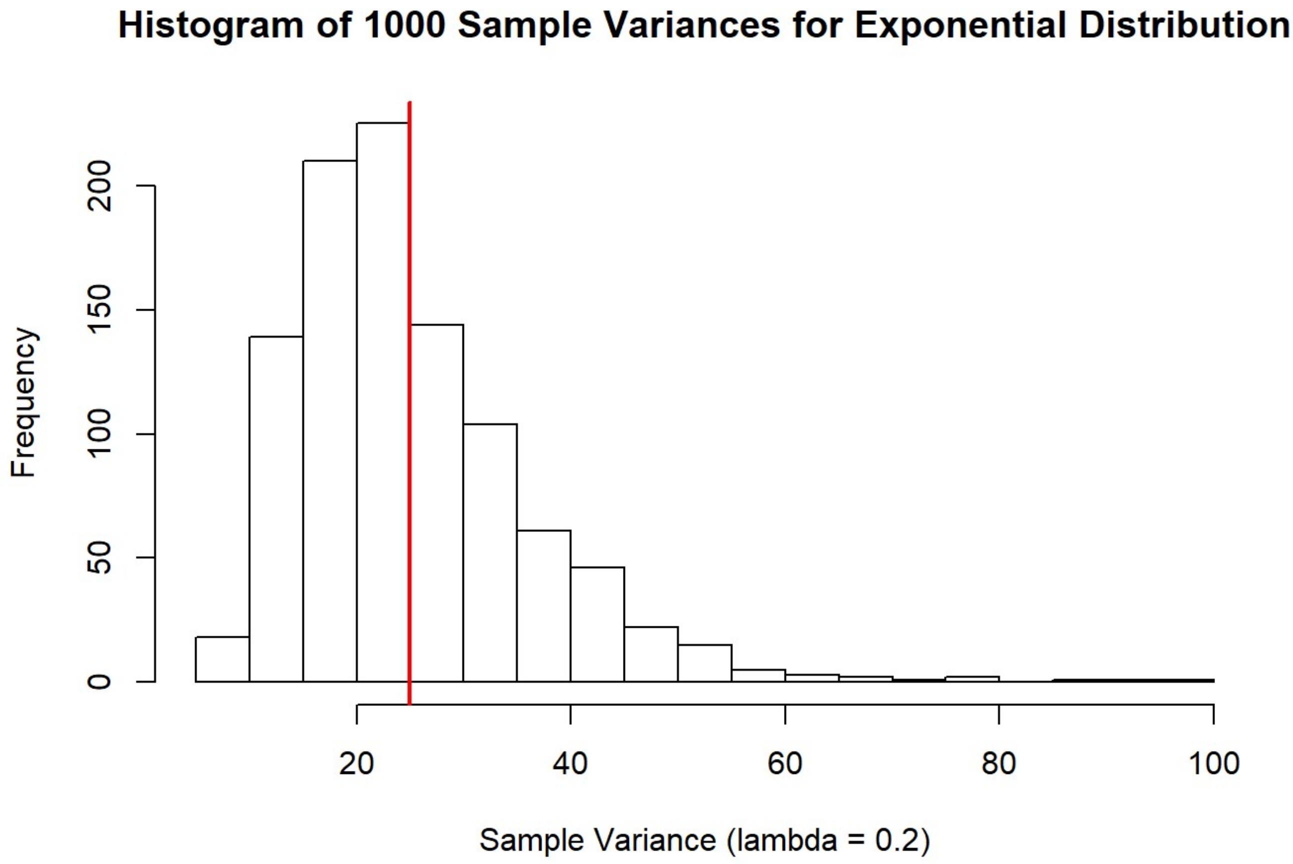
```
hist(sampleMean, main = "Histogram of 1000 Sample Means for Exponential Distributio
n", xlab = "Sample Mean (lambda = 0.2)", breaks = 20);
abline(v=5, col = "red", lwd = 2)
```



Sample Variance versus Theoretical Variance

We plot the sample variance for the distribution. The distribution is approximately normal with a peak at 25. Since the sample variance is an unbiased estimator of the population variance, the theory states that the mean of this distribution is the variance of the original exponential distribution which is (1/0.2)^2 = 25.

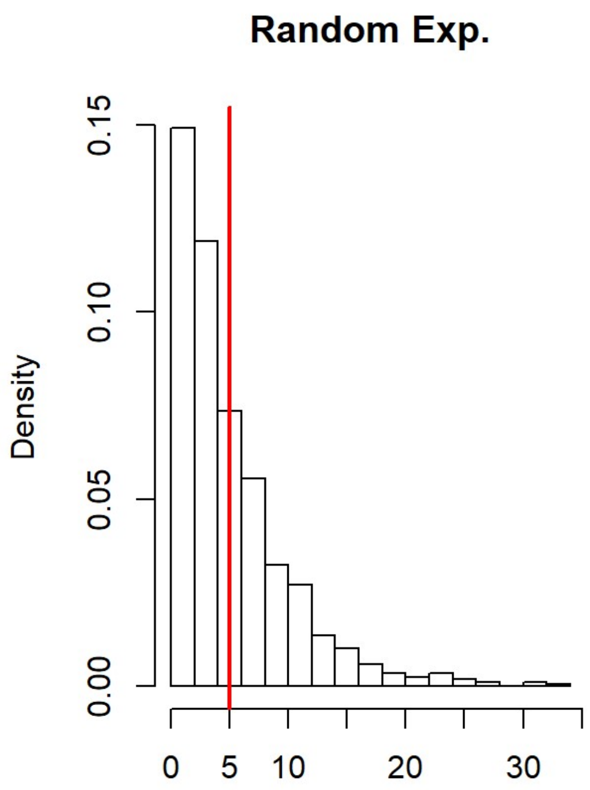
```
hist(sampleVariance, main = "Histogram of 1000 Sample Variances for Exponential Dis
tribution", xlab = "Sample Variance (lambda = 0.2)", breaks = 20);
abline(v=25, col = "red", lwd = 2)
```

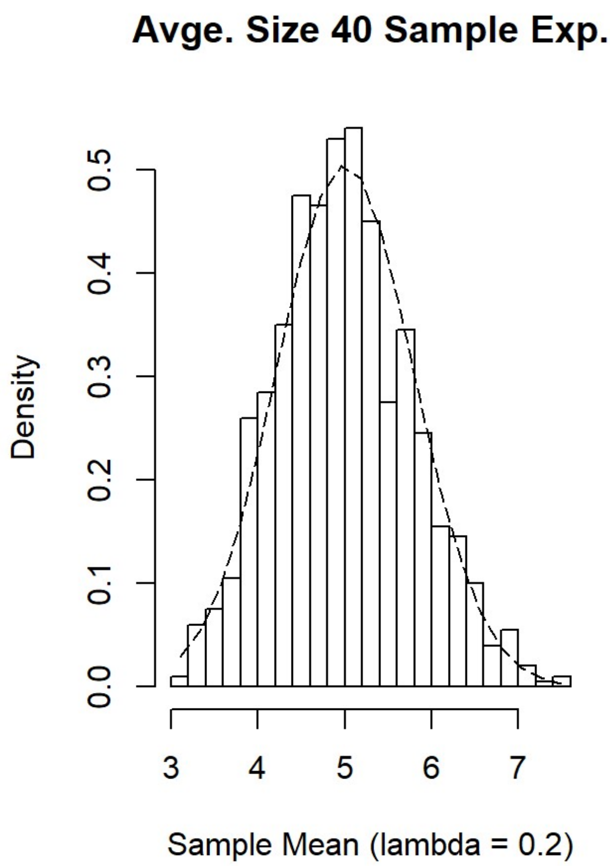


Distribution

We show that the distributions of a large enough sample size are approxmately normal. We do this by comparing a large collection of exponential distributions (i.e. 1000 selections from an exponential distribution) with a large selection of the sample averages from an exponential distribution. The right-hand graph is clearly normal.

```
set.seed(1)
distr <- rexp(1000, lambda)
par(mfrow=c(1, 2))
hist(distr, breaks=20, probability = TRUE, main="Random Exp.", xlab="")
abline(v=5, col="red", lwd=2)
hist(sampleMean, probability = TRUE, main = "Avge. Size 40 Sample Exp.", xlab = "Sa
mple Mean (lambda = 0.2)", breaks = 20);
xfit <- seq(min(sampleMean), max(sampleMean), length = 20)
yfit <- dnorm(xfit, mean = 1/lambda, sd = (1/lambda)/sqrt(40))
lines(xfit, yfit, lty = 5)
```





The dashed normal distribution on the righthand plot shows the sample mean is normally distributed.