Statistical Inference Course Project

Part 1: Simulation Exercise

Dennis van den Berg

25/07/2015

Overview

By running simulations we investigated the properties of the distribution of the mean of 40 exponentials. Specifically, we looked at:

1. its sample mean
2. its sample variance
3. the shape of the distribution

and compared them to their theoretical counterparts.

Simulations

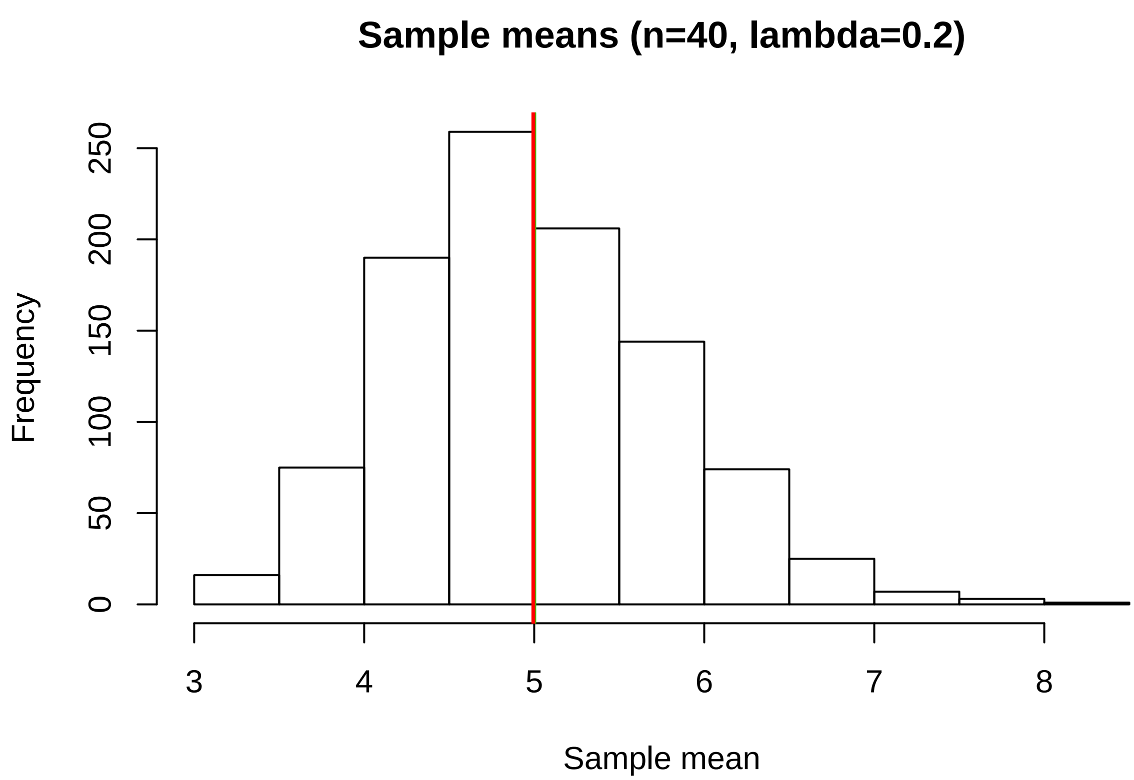
We looked at exponential distributions with λ=.2, taking samples of 40 measurements and repeating this 1000 times:

```
set.seed(1234567890)
lambda=.2
n=40
repeats=1000
sample_matrix <- matrix(rexp(n*repeats, rate=lambda), nrow=repeats, ncol=n)
```

Sample Mean versus Theoretical Mean

First we determined sample means for each sample and plotted these in a histogram. The mean of the sample means is plotted as a red line and the theoretical mean ($1/\lambda$) for exponential distributions in green:

```
sample_means <- rowMeans(sample_matrix)
hist(sample_means, main=paste("Sample means (n=", n, ", lambda=", lambda, ")", sep=""), xlab="Sample mea
theoretical_mean_exp <- 1/lambda
abline(v=theoretical_mean_exp, lwd=2, col="Green")
mean_of_sample_means <- mean(sample_means)
abline(v=mean_of_sample_means, lwd=2, col="Red")
paste("Theoretical mean (green): ", theoretical_mean_exp, sep="")
paste("Mean of sample means (red): ", mean_of_sample_means, sep="")
```




```
## [1] "Theoretical mean (green): 5"
```

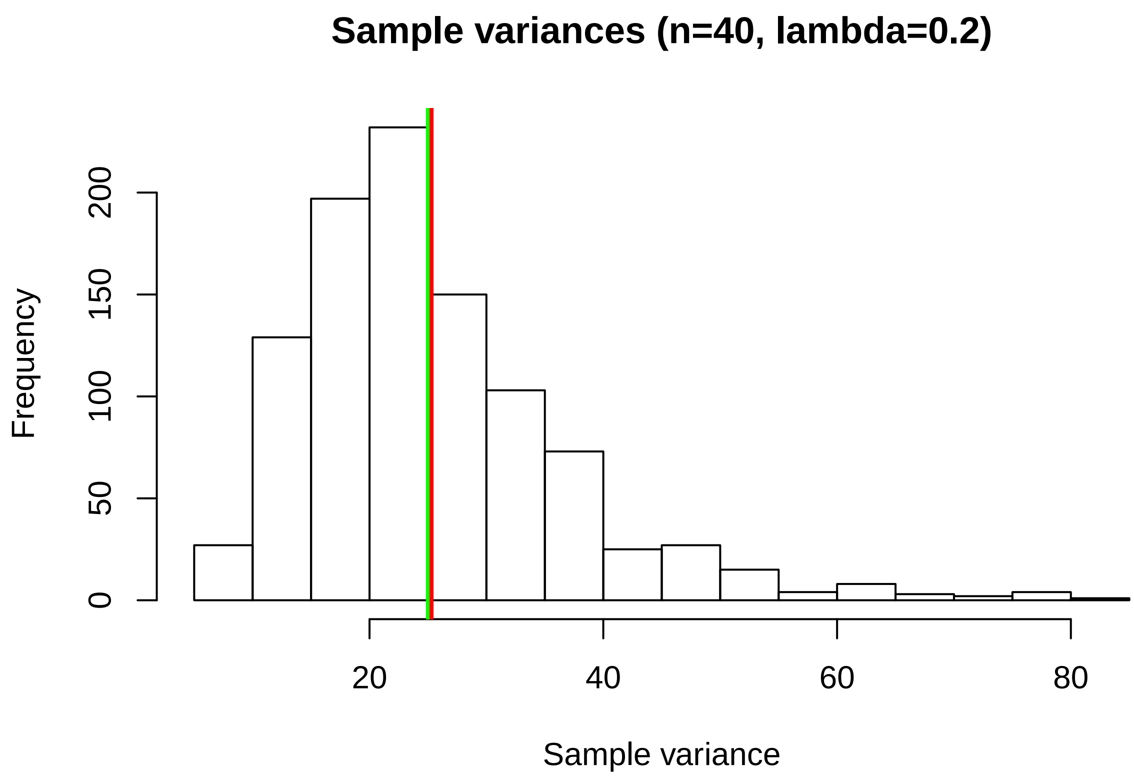
```
## [1] "Mean of sample means (red): 4.99527420490266"
```

From this histogram we conclude that the distribution of sample means is centered around the theoretical mean $1/\lambda$ (green line) of the exponential distribution. The mean of all the sample means (red line) is very close to this value. This is what we would expect from the Law of Large Numbers.

Sample Variance versus Theoretical Variance

From this same data we determined sample variances, plotted these in a histogram as well and again added their theoretical value ($1/\lambda^2$) in green and mean variance of the samples in red.

```
sample_variances <- apply(sample_matrix, 1, var)
hist(sample_variances, main=paste("Sample variances (n=", n, ", lambda=", lambda, ")", sep=""), xlab="Sa
theoretical_variance_exp <- 1/lambda^2
abline(v=theoretical_variance_exp, lwd=2, col="Green")
mean_of_sample_variances <- mean(sample_variances)
abline(v=mean_of_sample_variances, lwd=2, col="Red")
paste("Theoretical variance (green): ", theoretical_variance_exp, sep="")
paste("Mean of sample variances (red): ", mean_of_sample_variances, sep="")
```




```
## [1] "Theoretical variance (green): 25"
```

```
## [1] "Mean of sample variances (red): 25.3088125457824"
```

We conclude from this plot that the distribution of sample variances is centered around the theoretical variance $1/\lambda^2$ (green line) of the exponential distribution. The mean of all sample variances (red line) is very close to this value. This is what we would expect from the Law of Large Numbers.

Distribution

Investigating the shape of the distribution, we plotted histograms for different sample sizes n:

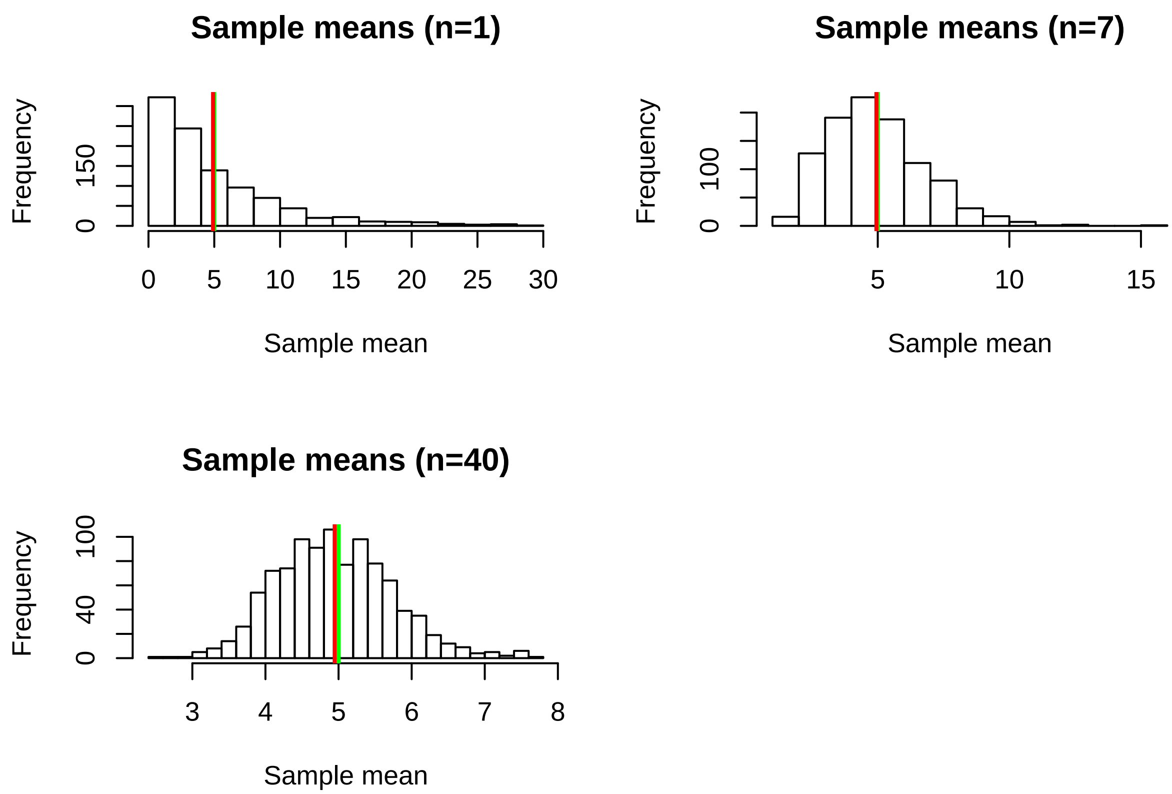
```
par(mfrow=c(2,2))

n=1
sample_matrix <- matrix(rexp(n*repeats, rate=lambda), nrow=repeats, ncol=n)
hist(rowMeans(sample_matrix), breaks=20, main=paste("Sample means (n=", n, ")", sep=""), xlab="Sample me
theoretical_mean_exp <- 1/lambda
abline(v=theoretical_mean_exp, lwd=2, col="Green")
mean_of_sample_means <- mean(rowMeans(sample_matrix))
abline(v=mean_of_sample_means, lwd=2, col="Red")

n=7
sample_matrix <- matrix(rexp(n*repeats, rate=lambda), nrow=repeats, ncol=n)
hist(rowMeans(sample_matrix), breaks=20, main=paste("Sample means (n=", n, ")", sep=""), xlab="Sample me
theoretical_mean_exp <- 1/lambda
```

```
abline(v=theoretical_mean_exp, lwd=2, col="Green")
mean_of_sample_means <- mean(rowMeans(sample_matrix))
abline(v=mean_of_sample_means, lwd=2, col="Red")

n=40
sample_matrix <- matrix(rexp(n*repeats, rate=lambda), nrow=repeats, ncol=n)
hist(rowMeans(sample_matrix), breaks=20, main=paste("Sample means (n=", n, ")", sep=""), xlab="Sample me
theoretical_mean_exp <- 1/lambda
abline(v=theoretical_mean_exp, lwd=2, col="Green")
mean_of_sample_means <- mean(rowMeans(sample_matrix))
abline(v=mean_of_sample_means, lwd=2, col="Red")
```



From these plots we see that the distribution of the sample means approximates the normal distribution when sample size n is increasing. This is in agreement with the Central Limit Theorem.