Statistical Inference - Course Project (Task 1)

Joscha Frischherz

September 25, 2015

The first task of this projece will address the comparison of an exponential distribution with the Central Limit Theorem.The investigation will occur for averages of 40 exponentials, over 1000 simulations. The following assumptions are made about the exponential distribution:

- mean = 1/lambda
- standard deviation = 1/lambda
- lambda =0.2

R code for initialisation of variables:

```
lam <- 0.2
mn <- 1/lam
s <- 1/lam
```

First, we compare the sample mean and compare it to the theoretical mean of the distribution:

```
##sample mean
mean(rexp(40,lam))
```

```
## [1] 5.400085
```

```
##theoretical mean
mn
```

```
## [1] 5
```

We note that the sample mean for 40 exponentials is close but not exactly the given theoretical mean. Furthermore if we take multiple samples that mean varies:

```
##sample means for multiple samples (3)
mean(rexp(40,lam))
```

```
## [1] 5.095044
```

```
mean(rexp(40,lam))
```

```
## [1] 4.797686
```

```
mean(rexp(40,lam))
```

```
## [1] 5.224225
```

The above sample means vary above and below the theoretical mean. Simulating the mean for 1000 samples and taking the mean of the simulation will approximate the theoretical mean of 5 (1/0.2).

```r
##Simulation of 1000 samples of 40 exponentials
mns <- NULL
for (i in 1 : 1000) mns = c(mns, mean((rexp(40,lam))))
##mean of simulation means
mean(mns)
```

```
## [1] 5.011717
```

We essentially show that the fluctuation of means above and below the theoretical mean over a large number of simulations averages to the theoretical mean.

Second, to indicate how variable the samlple is, we calculate the variance of the above simulation as well as the theoretical variance for the distribution:

```r
##variance of simulated means
var(mns)
```

```
## [1] 0.6191656
```

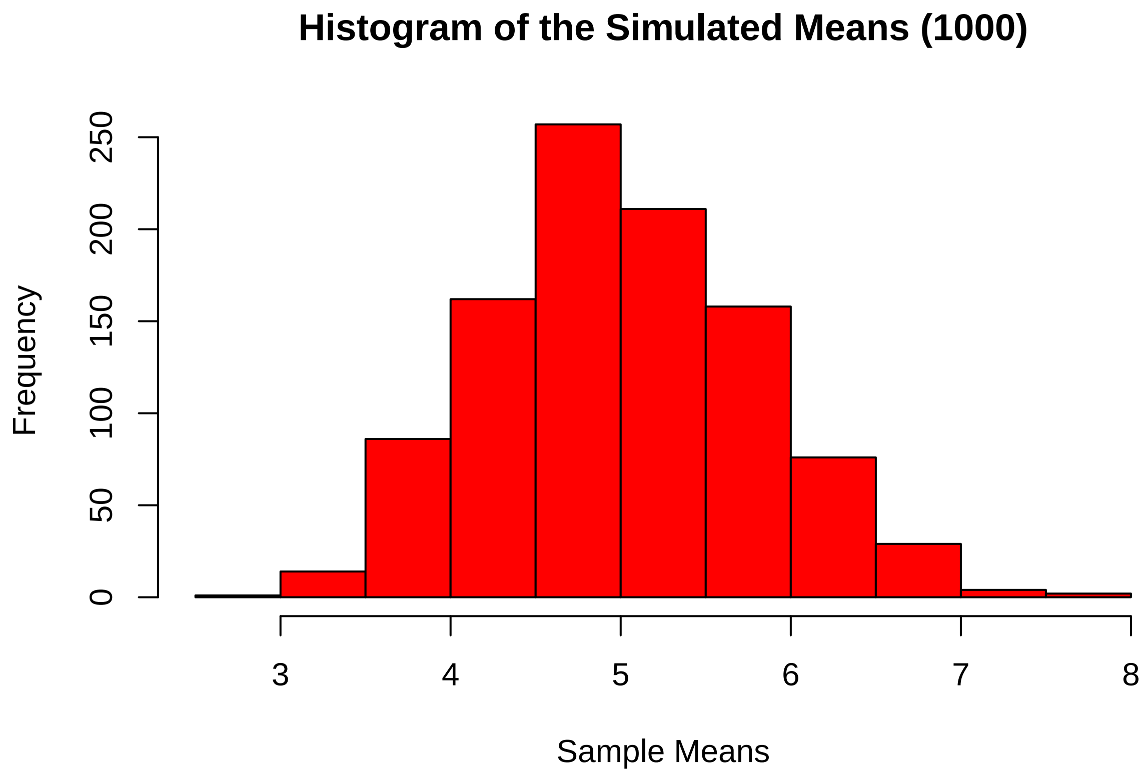
```r
##theoretical variance for 40 exponentials
1/lam^2/40
```

```
## [1] 0.625
```

We see again that a large number of simulations i.e. 1000 approximates the theoretical variance of 0.625.

Third, we now plot the sample means into a histogram to demonstrate that the distribution of the means is approximately normal:

```r
##display a histogramm of the 1000 means
hist(mns, col = "red", main = "Histogram of the Simulated Means (1000)", xlab = "Sample Means")
```



The histogram is bell-curved which indicates a normal distribution. In other words the means of the means for the 40 exponentials are normally distributed, which indicates, that the highest frequency of them occurs around the theoretical mean (center) with a more or less equal distribution to each side of the tails.