Statistical Inference Course Project Part 1 – Simulation

Report Summary

The following is a statistical analysis of 1,000 simulations generating 40 random exponential variables. The mean and variance of the means were calculated and compared to theoretical values. The distribution of means was also plotted to show convergence to normal distribution, thereby demonstrating the central limit theorem.

First, create a vector containing means of 40 random exponentially distributed values ($\lambda = 0.2$). Vector length is equal to the number of simulations, which is 1000.

```
set.seed(1)
mns <- NULL
for (i in 1:1000) mns <- c(mns, mean(rexp(40, 0.2)))
str(mns)
```

```
##  num [1:1000] 4.86 5.96 4.28 4.7 5.2 ...
```

The theoretical mean for an exponential distribution is equal to $1/\lambda$.

```
1/0.2
```

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

According to the law of large numbers, the actual mean of `mns` should approach the theoretical mean. Based on the computation below, this appears to be true.

```
mean(mns)
```

```
## [1] 4.990025
```

The variance of the distribution of means is estimated as $1/(\lambda^2 * n)$, where n is the sample size.

```
1/(0.2^2*40)
```

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

The actual variance of the means should approach the estimated value. As seen below, the actual variance is close to its theoretical value.

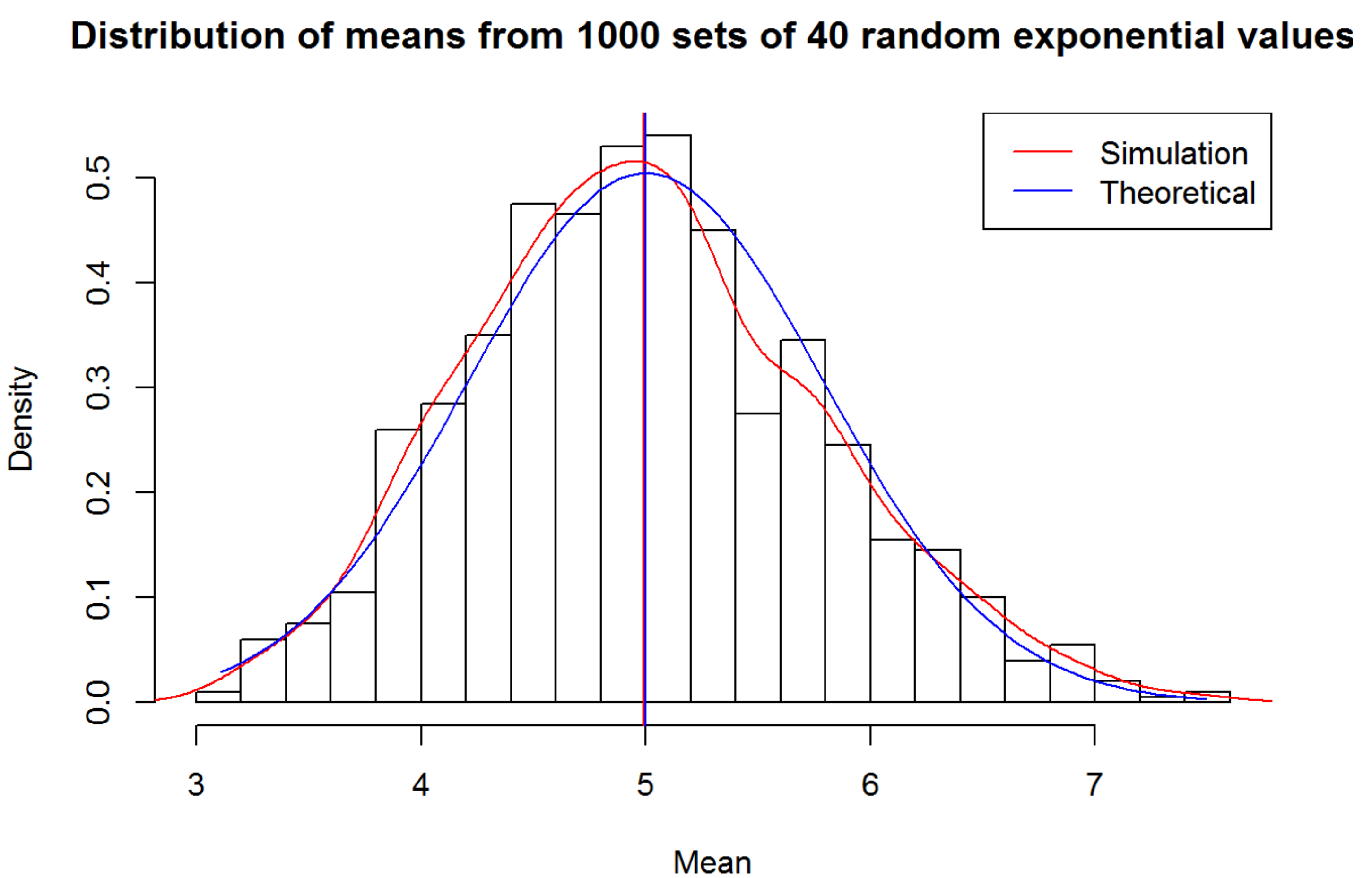
```
var(mns)
```

```
## [1] 0.6111165
```

The central limit theorem states that the distribution of sample means should approach a normal distribution given a large enough number of samples. The plot below shows the distribution of means from 1000 simulations of 40 randomly generated exponential values. The vertical lines represent the actual and theoretical means. Overlays of the distribution densities, both actual and theoretical, illustrate that the simulated distribution of means is approximately normal.

```
hist(mns, breaks=25, prob=T, main='Distribution of means from 1000 sets of 40 random exponential valu
es', xlab='Mean')
abline(v=1/0.2, col='blue')
abline(v=mean(mns), col='red')
lines(density(mns), col='red')

# Draw a normal distribution density with mean=1/lambda and sd=1/lambda/sqrt(n)
x <- seq(min(mns), max(mns), length=100)
y <- dnorm(x, mean=1/0.2, sd=1/0.2/sqrt(40))
lines(x, y, col='blue')
legend('topright', c('Simulation','Theoretical'), lty=1, col=c('red','blue'))
```



Compare the above distribution to that of 1000 randomly generated, exponentially distributed values. In this example, which is plotted below, the distribution is clearly not normal.

```
set.seed(2)
exp <- rexp(1000, 0.2)
hist(exp, breaks=25, prob=T, main='Distribution of 1,000 random exponential values', xlab='')
lines(density(exp), col='red')
```