

# Statistical Inference Part 1 - Simulation

*Sunday, July 26, 2015*

## Overview

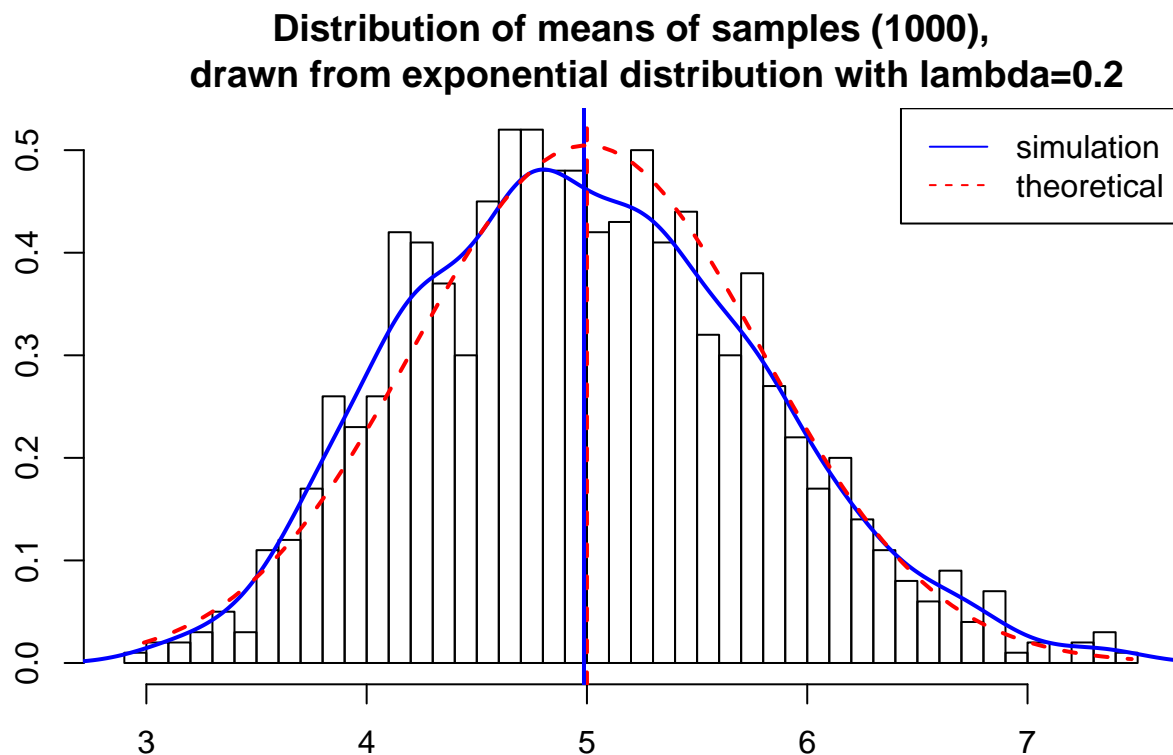
In this course project we will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with `rexp(n, lambda)` where `lambda` is the rate parameter. The mean of exponential distribution is  $1/\lambda$  ( $1/\lambda$ ) and the standard deviation is also  $1/\lambda$  ( $1/\lambda$ ). We will set  $\lambda = 0.2$  for all of the simulations. We will investigate the distribution of averages of 40 exponentials.

## Simulations

Let's do a thousand simulated averages of 40 exponentials and then 10000 simulated averages of 40 exponentials.

```
set.seed(3) # This is to ensure the same random sample is selected each time
lambda <- 0.2 #set our lamda
number_of_simulations <- 1000
sample_size <- 40
simulation <- matrix(rexp(number_of_simulations*sample_size, rate=lambda),
                     number_of_simulations, sample_size)
row_means <- rowMeans(simulation)
```

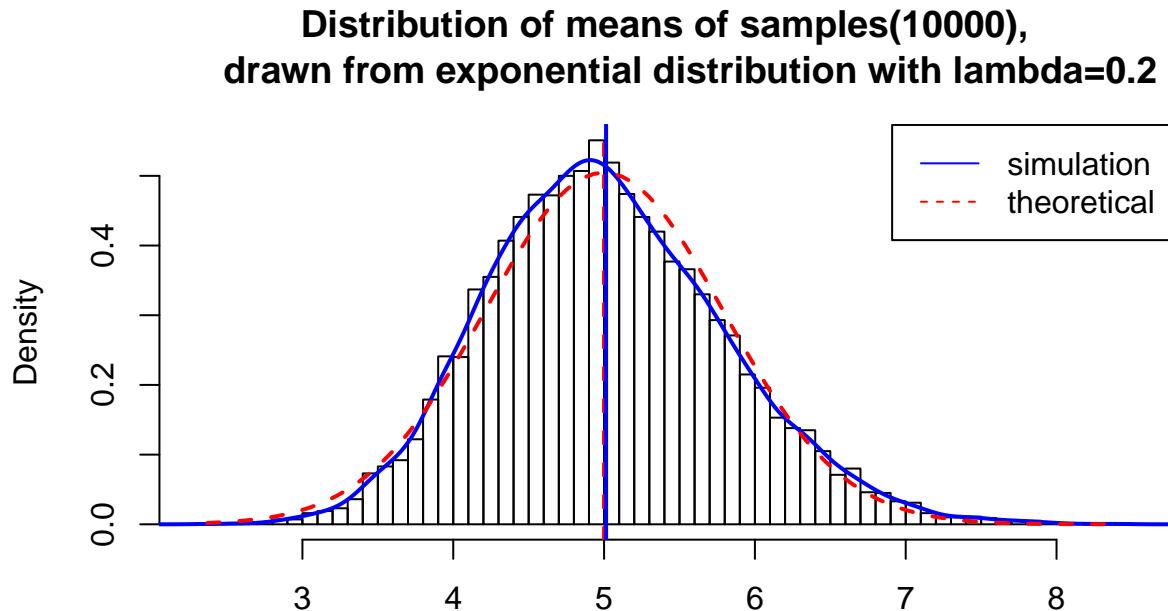
## The distribution of sample means



As you can see from the graph below, the calculated distribution of means of random sampled exponential

distributions, overlaps quite nice with the normal distribution with the expected values based on the given lambda. The graph shows the the sample (simulated in blue) and theoretical (in red) distribution. As you can see the graph is approximately normally distributed. The means are very close together (centre of distribution) as demonstrated by the red and blue vertical lines for theoretical and simulated lines respectively.

At 10000 simulations the means are almost identical while the distributions are also almost the same.



## Sample Mean versus Theoretical Mean

The expected mean  $\mu$  of a exponential distribution of rate  $\lambda$  is

$$\mu = \frac{1}{\lambda}$$

```
mu <- 1/lambda
mu
```

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

Let  $\bar{X}$  be the average sample mean of 1000 simulations of 40 randomly sampled exponential distributions.

```
meanOfMeans <- mean(row_means)
meanOfMeans
```

```
## [1] 5.013331
```

As you can see the expected mean and the average sample mean are very close. This is also shown in the graph above with the expected mean in Red vertical line and the sample mean in the Blue vertical line.

## Sample Variance versus Theoretical Variance

The expected standard deviation  $\sigma$  of a exponential distribution of rate  $\lambda$  is

$$\sigma = \frac{1/\lambda}{\sqrt{n}}$$

```
sd <- 1/lambda/sqrt(sample_size)
sd
```

```
## [1] 0.7905694
```

The variance  $Var$  of standard deviation  $\sigma$  is

$$Var = \sigma^2$$

```
Var <- sd^2
Var
```

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

Let  $Var_x$  be the variance of the average sample mean of 1000 simulations of 40 randomly sampled exponential distribution, and  $\sigma_x$  the corresponding standard deviation.

```
sd_x <- sd(row_means)
sd_x
```

```
## [1] 0.7887496
```

```
Var_x <- var(row_means)
Var_x
```

```
## [1] 0.6221259
```

As you can see the standard deviations are very close Since variance is the square of the standard deviations, minor differences will be enhanced, but are still pretty close.