

# Statistical Inference Course Project - Part 1

1. Overview In this project I 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  $1/\lambda$  and the standard deviation is also  $1/\lambda$ . I will set `lambda = 0.2` for all of the simulations. I will investigate the distribution of averages of 40 exponentials. Note that I will need to do a thousand simulations.

## 2. Simulations

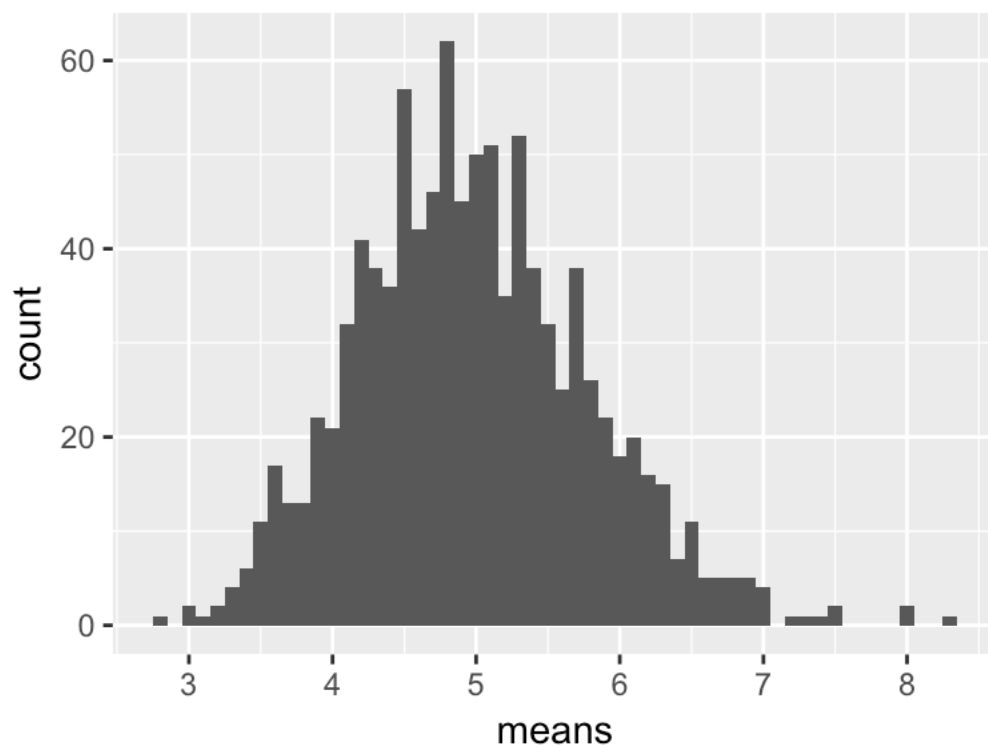
```
library(ggplot2)

# set constants
lambda <- 0.2 # lambda for rexp
n <- 40 # # exponentials
noOfSimulations <- 1000 # number of tests

# set the seed to create reproducibility
set.seed(23888)

expDistributions <- matrix(data=rexp(n * noOfSimulations, lambda), nrow=noOfSimulations)

expDistributionMean <- data.frame(means=apply(expDistributions, 1, mean))
```



## 3. Sample Mean versus Theoretical Mean

The expected mean of an exponential distribution of rate 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(expDistributionMean$means)
meanOfMeans
```

```
## [1] 4.979235
```

As you can see the expected mean and the average sample mean are very close

#### 4. Sample Variance versus Theoretical Variance

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

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

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

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

```
variance <- sd^2
variance
```

```
## [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(expDistributionMean$means)
sd_x
```

```
## [1] 0.807221
```

```
variance_x <- var(expDistributionMean$means)
variance_x
```

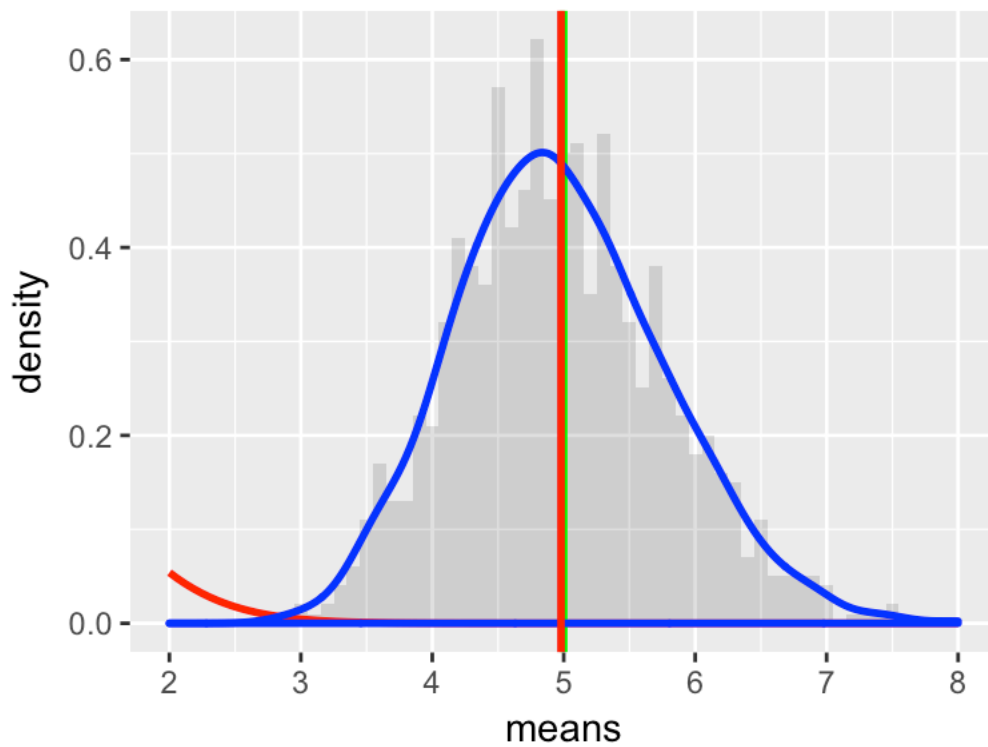
```
## [1] 0.6516057
```

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.

#### 5. Distribution

Comparing the population means & standard deviation with a normal distribution of the expected values.  
Added lines for the calculated and expected means

```
# plot the means  
ggplot(data = expDistributionMean, aes(x = means)) +  
  geom_histogram(binwidth=0.1, aes(y=..density..), alpha=0.2) +  
  stat_function(fun = dnorm, arg = list(mean = mu , sd = sd), colour = "red", size=1)  
+  
  geom_vline(xintercept = mu, size=1, colour="green") +  
  geom_density(colour="blue", size=1) +  
  geom_vline(xintercept = meanOfMeans, size=1, colour="red") +  
  scale_x_continuous(breaks=seq(mu-3,mu+3,1), limits=c(mu-3,mu+3))
```



As you can see from the graph, 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