

# Application of Central Limit Theorem to exponential distribution

Alex G.

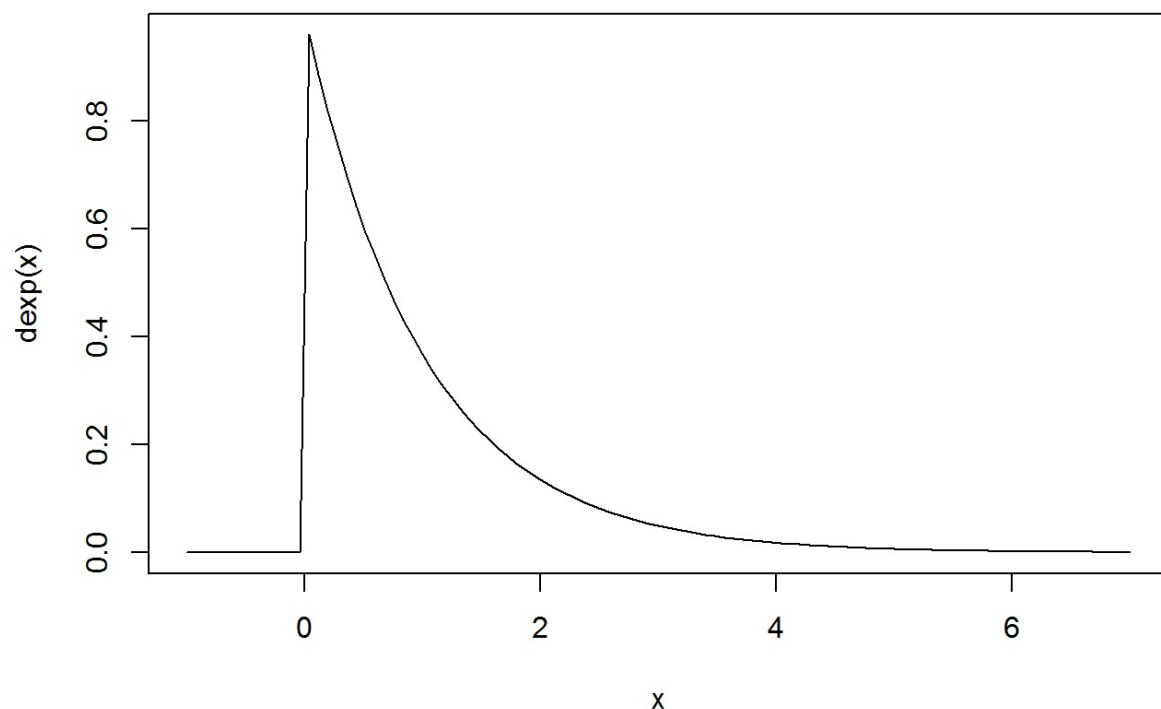
## Overview

Central Limit Theorem states that if we take many samples of some distribution, then means of these samples will be distributed in a different way - they will follow normal distribution, and their average (mean of means) will be centered on theoretical mean of the original distribution.

The following simulation demonstrates how this applies to exponential distribution. Exponential distribution is described by a rate parameter  $\lambda$ , and then its theoretical mean and standard deviation will both be  $1/\lambda$ .

This is shape of exponential distribution. We can look at it by plotting its density function.

```
curve(dexp(x), -1, 7)
```



## Data preparation and sample check

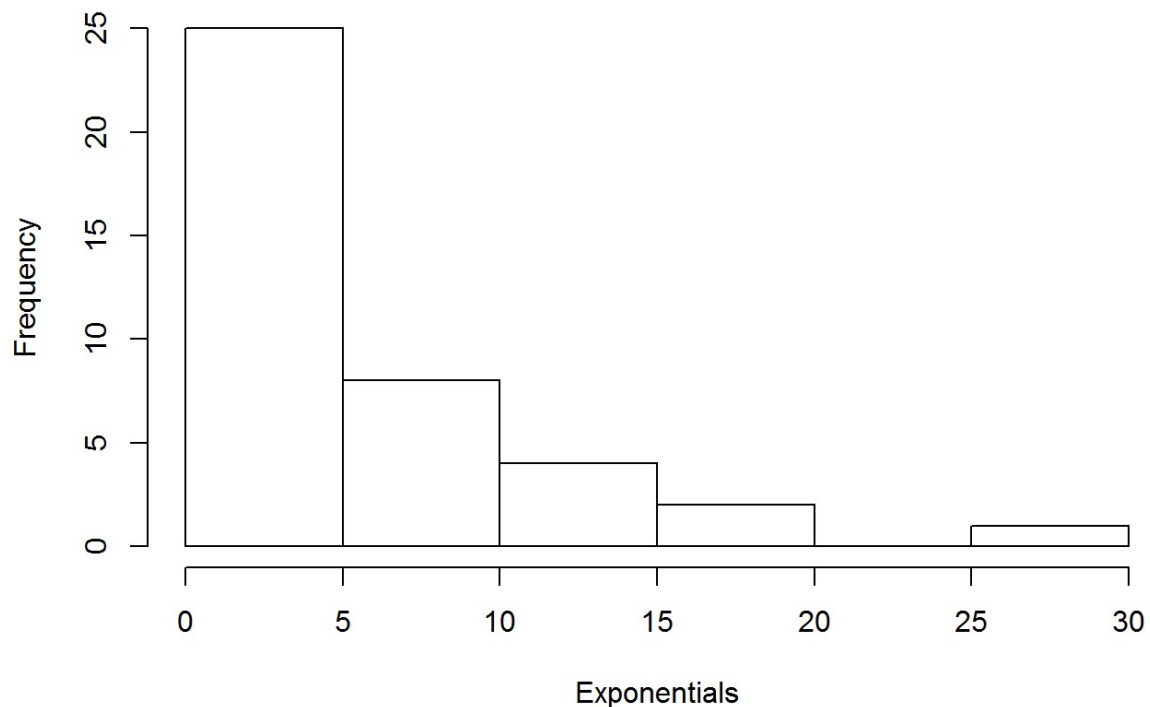
In order to do the simulation, we generate 1000 samples of 40 data points using random generation function for exponential distribution with  $\lambda = 0.2$ .

```
lambda <- 0.2
dat <- list()
set.seed(1)
for(i in 1:1000) dat[[i]] <- rexp(40, lambda)
```

If we take a random sample and see it, we should see something similar to the previous plot of the distribution function for the exponential distribution. And its mean and standard deviation should be close to 5 both .

```
set.seed(1)
random.sample <- sample(1:1000,1)
hist(dat[[random.sample]],
      main=paste("Distribution in sample", random.sample),
      xlab="Exponentials")
```

### Distribution in sample 266



```
(random.mean <- mean(dat[[random.sample]]))
```

```
## [1] 5.232086
```

```
# DifferenceWithTheory() function is hidden, it's just cosmetics to display
%
DifferenceWithTheory(random.mean, 1/lambda)
```

```
## Difference with theory is 5%
```

```
(random.sd <- sd(dat[[random.sample]]))
```

```
## [1] 5.491864
```

```
DifferenceWithTheory(random.sd, 1/lambda)
```

```
## Difference with theory is 10%
```

## Distribution of sample means

Now let's calculate mean for each of 1000 samples and see how these means are distributed themselves. We expect to see that average mean should be even closer to 5 than we saw in a single random sample above.

```
means <- sapply(dat, mean)
(sample.mean <- mean(means))
```

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

```
DifferenceWithTheory(sample.mean, 1/lambda)
```

```
## Difference with theory is 0.2%
```

It is indeed close, and much closer than for a single sample. In fact, let's see how smaller the variance of means is than variance of the original distribution.

```
(means.variance <- var(means))
```

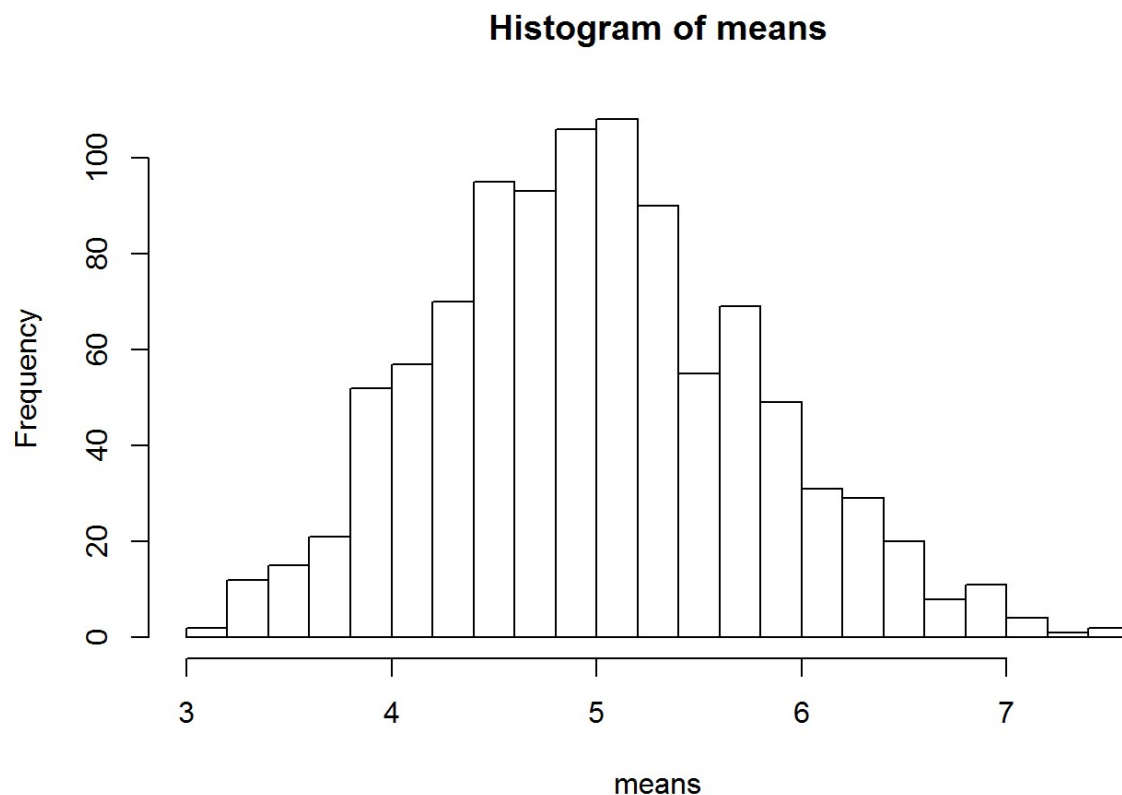
```
## [1] 0.6111165
```

```
cat(paste("Variance of the means is", round(25/means.variance),
          "times smaller than variance of the population.\n"))
```

```
## Variance of the means is 41 times smaller than variance of the population.
```

How sample means are distributed?

```
hist(means, breaks=25)
```



Distribution of sample means looks normal - it has symmetrical bell shape and centered on 5 - theoretical mean of the population.

## Distribution of sample variances

What about variances in our samples? Is it close to variance of the population?

```
variances <- sapply(dat, var)
(mean.variance <- mean(variances))
```

```
## [1] 25.06459
```

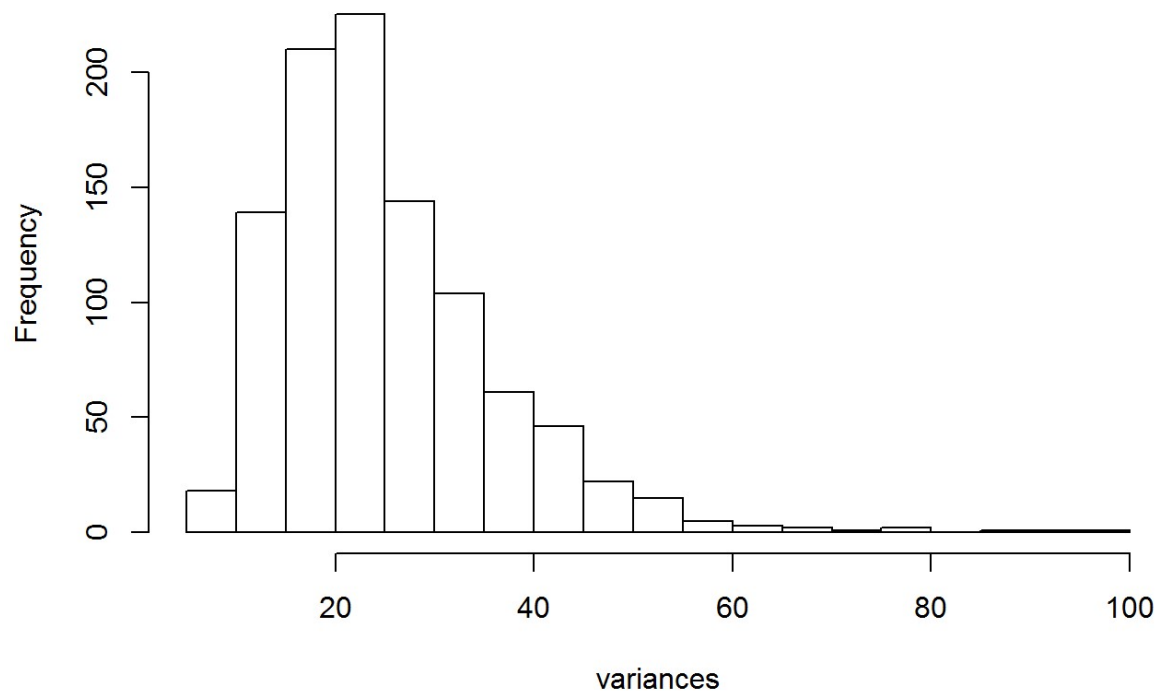
```
DifferenceWithTheory(mean.variance, 1/lambda^2)
```

```
## Difference with theory is 0.3%
```

Average variance of the samples is close to theoretical variance, but, unlike distribution of sample means, we do not have expectation to see it distributed normally.

```
hist(variances, breaks=25)
```

**Histogram of variances**



Indeed, while it's centered on theoretical variance of the population, but its shape isn't symmetrical, so it's not normal.