Statistical Inference Course Project - part 1

Arkadiusz Oliwa

1 March 2018

Overview

In this project you I'll investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution is simulated in R with rexp(n, lambda) where lambda is the rate parameter equal 0.2. I will investigate the distribution of averages of 40 exponentials (thousand simulations).

```
n <- 1000
samples <- 40
lambda <- 0.2
```

Simulation

I will generate the simulation

```
dataRaw <- matrix(data = rexp(n = n * samples, rate = lambda),
                  nrow = n, ncol = samples)
dataRaw <- as.data.frame(dataRaw)
```

I will generate the distribution of the averages of 40 exponentials

```
dataMean <- apply(X = dataRaw, MARGIN = 1, FUN = mean)
```

Sample Mean versus Theoretical Mean and Sample Variance versus Theoretical Variance

The theoretical mean of the distribution

```
theoreticalMean <- 1 / lambda
theoreticalMean
```

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

The sample mean

```
sampleMean <- mean(dataMean)
sampleMean
```

```
## [1] 5.028842
```

The sample mean is close to the theoretical mean

The theoretical variance of the distribution

```
theoreticalVariance <- (1 / lambda)^2/samples
theoreticalVariance
```

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

The sample variance

```
sampleVariance <- var(dataMean)
sampleVariance
```

```
## [1] 0.6514014
```

The sample variance is close to the theoretical variance

The theoretical standard deviation of the distribution

```
theoreticalSD <- 1/lambda/sqrt(samples)
theoreticalSD
```

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

The sample standard deviation

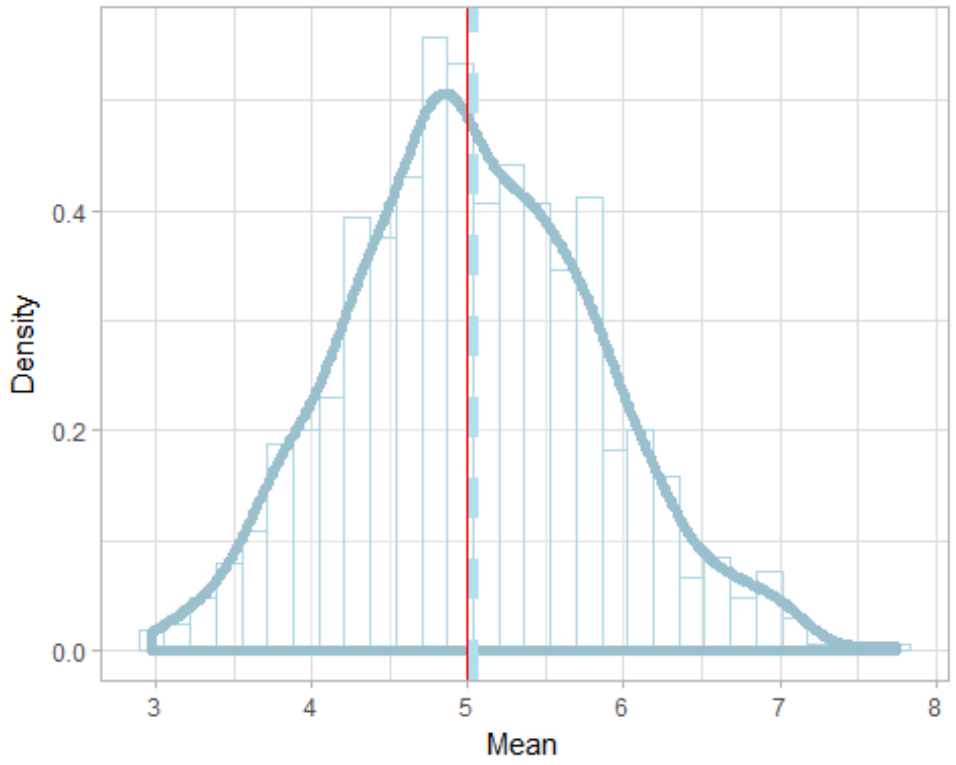
```
standardDeviation <- sd(dataMean)
standardDeviation
```

```
## [1] 0.8070944
```

The sample standard deviation is close to the theoretical standard deviation

The distribution of averages of 40 random exponentials with 1000 simulations

```
ggplot(data = as.data.frame(dataMean), aes(dataMean)) +
  geom_histogram(aes(y=..density..), fill="white", col="lightblue",
                 alpha=.3, bins = 30) +
  geom_density(col="lightblue3", lwd=2) +
  geom_vline(xintercept = sampleMean, col="lightblue2", linetype="dashed",
             lwd=2) +
  geom_vline(xintercept = theoreticalMean, col="red") +
  labs(x="Mean", y = "Density") +
  theme_light()
```



As you can see above the distribution of averages of independent and identically distributed (IID) variables becomes that of a standard normal.