Statistical Inference: Peer Graded Assignment

Author: Brynjólfur Gauti Jónsson

Summary

The purpose of this assignment is to compare theoretical parameters of the Exponential Distribution to empirically obtained measurements sampled via simulation. The Exponential Distribution samples a continuous Poisson process.

Sampling the Distribution

We will sample the distribution in R using rexp(n, lambda) where lambda is the rate. We will sample 40 exponentials 1000 times and store the mean of each of the samples.

```
set.seed(101)
exp <- NULL
for (i in 1:1000) exp <- c(exp, mean(rexp(40, 0.2)))
head(exp)
```

```
## [1] 4.034012 4.994885 3.809412 4.670817 5.168187 5.396491
```

Now that we have obtained a sample of exponentials we can go ahead with our calculations.

1. The Mean

The theoretical mean of the Exponential Distribution is calculated such:

$\mu = 1/\lambda$.

Let's compare that to the mean of our distribution of means.

```
# Empirical Mean
mean(exp)
```

```
## [1] 5.012603
```

```
# Theoretical Mean
1/0.2
```

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

```
# Difference in Means
abs(mean(exp)-1/0.2)
```

```
## [1] 0.01260258
```

We aproximated the theoretical mean very closely.

2. The Variancee

The theoretical variance is given by:

$\sigma^2 = \frac{1}{(\lambda * \sqrt{n})^2}$.

Now we compare it to an empirically obtained variance.

```
# Empirical Variance
var(exp)
```

```
## [1] 0.5985383
```

```
# Theoretical Variance
(0.2 * sqrt(40))^-2
```

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

```
# Difference in Variances
abs(var(exp) - (0.2 * sqrt(40))^-2)
```

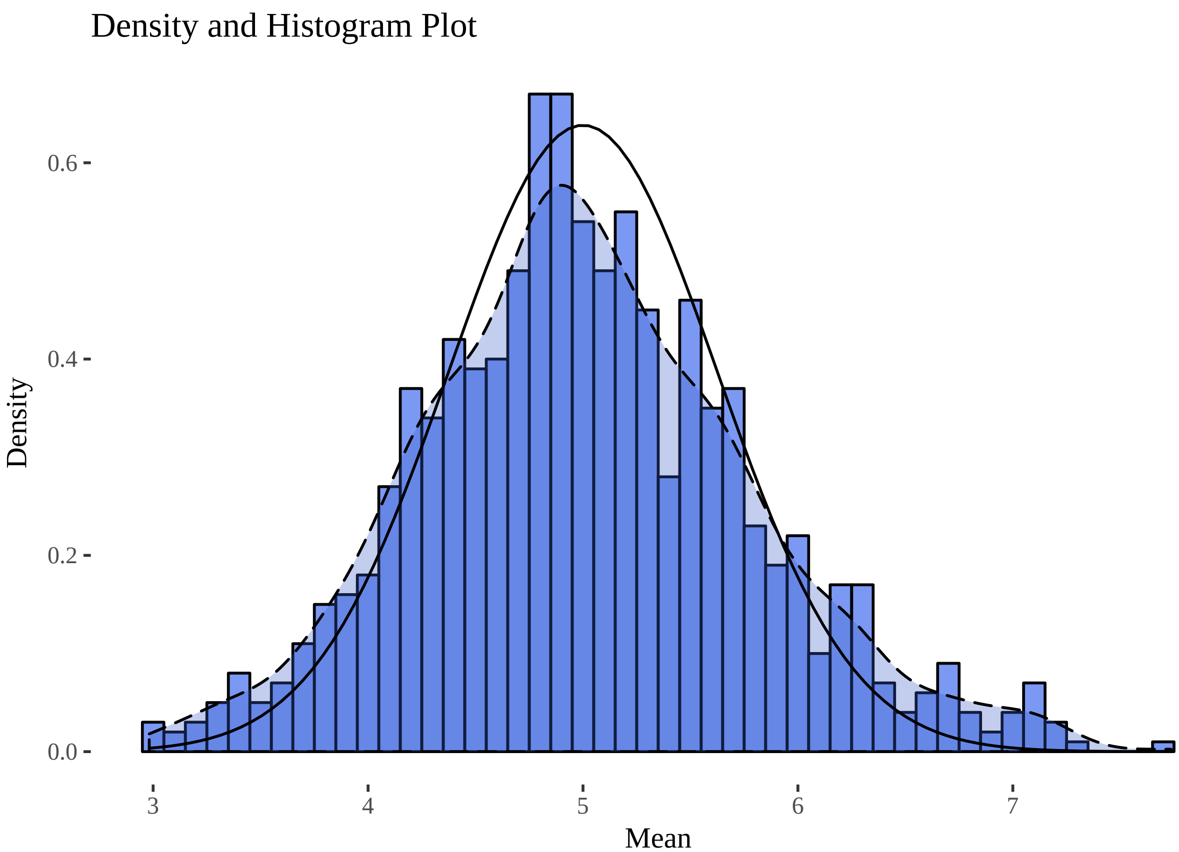
```
## [1] 0.02646167
```

3. Normality of Means

We plot a histogram with the density curve of the sample plotted by a broken line, and a normal density curve plotted with a solid line.

```r
library(ggplot2)
library(ggthemes)
g <- ggplot(data.frame(x = exp), aes(x=x)) + theme_tufte()

g + geom_histogram(aes(y = ..density..), fill = 'royalblue2', color='black',
                   alpha = 0.7, binwidth = .1) +
    geom_density(aes(y=..density..), fill='royalblue3', alpha=0.3, linetype = 5) +
    stat_function(fun = 'dnorm', args=list(mean=1/0.2, sd = (0.2 * sqrt(40))^-2)) +
    xlab('Mean') + ylab('Density') + ggtitle('Density and Histogram Plot')
```



We assumed in this assignment that a large enough sample of means gathered from a large enough sample of exponentials would be approximately normally distributed. We have shown that to be so.