Exponential Distribution Simulation

Trent Parkinson

December 23, 2017

Overview

This project 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 $\frac{1}{\lambda}$ and the standard deviation is also $\frac{1}{\lambda}$. Setting $\lambda = \frac{1}{5}$ for all of the simulations. The project will investigate the distribution of averages of 40 exponentials over a thousand simulations.

Setting the environment

Loading necessary libraries to perform loading and plotting. Setting seed so project is fully reproducible.

```
library(data.table)
library(ggplot2)

set.seed(1990)
```

Simulation

Setting the parameters for the simulated exponential distribution, and taking the mean of 40 outcomes, over 1000 trials and storing the means in a `data.table`.

```
lambda <- 1/5
simulations <- 1000
n <- 40

expo_simu <- matrix(rexp(simulations * n, lambda), nrow = simulations)
expo_means <- data.table(means = apply(expo_simu,1,mean))
```

Means

The theoretical mean of an exponential distribution is $\mu = \frac{1}{\lambda} = \frac{1}{1/5} = 5$. In the simulation the mean was found to be.

```
exp_mean <- mean(expo_means$means)
exp_mean
```

```
## [1] 5.035035
```

The difference in the theoretical mean and experimental mean as a percentage of error is only,

```
theor_mean <- 1/lambda
mean_pe <- abs((theor_mean - exp_mean)/theor_mean)*100
cat("Percentage of Error\n",paste(round(mean_pe,digits = 3),"%"))
```

```
## Percentage of Error
##  0.701 %
```

Computing the 95% confidence interval shows the theoretical mean is with in the bounds.

```
lower_conf <- t.test(expo_means$means, mu = 1/lambda)$conf[1]
upper_conf <- t.test(expo_means$means, mu = 1/lambda)$conf[2]

cat("95% Confidence Interval\n",paste0(round(lower_conf,digits = 3),
                                       ", ", round(upper_conf, digits = 3)))
```

```
## 95% Confidence Interval
##  4.986, 5.084
```

Variance

As with the mean, the variance can be compared in the same manner. The theoretical variance (σ^2) is

$$\sigma^2 = \left(\frac{\frac{1}{\lambda}}{\sqrt{n}}\right)^2 = \left(\frac{\frac{1}{1/5}}{\sqrt{40}}\right)^2 = \frac{25}{5 \cdot 8} = \frac{5}{8} = .625$$

```
exp_sd <- sd(expo_means$means)
exp_var <- exp_sd^2
exp_var
```

```
## [1] 0.6171751
```

The percentage of error for the the sample variance is,

```
theor_sigma <- (1/lambda)/sqrt(n)
theor_var <- theor_sigma^2
var_pe <- abs((theor_var - exp_var)/theor_var)*100
cat("Percentage of Error\n",paste(round(var_pe,digits = 3),"%"))
```

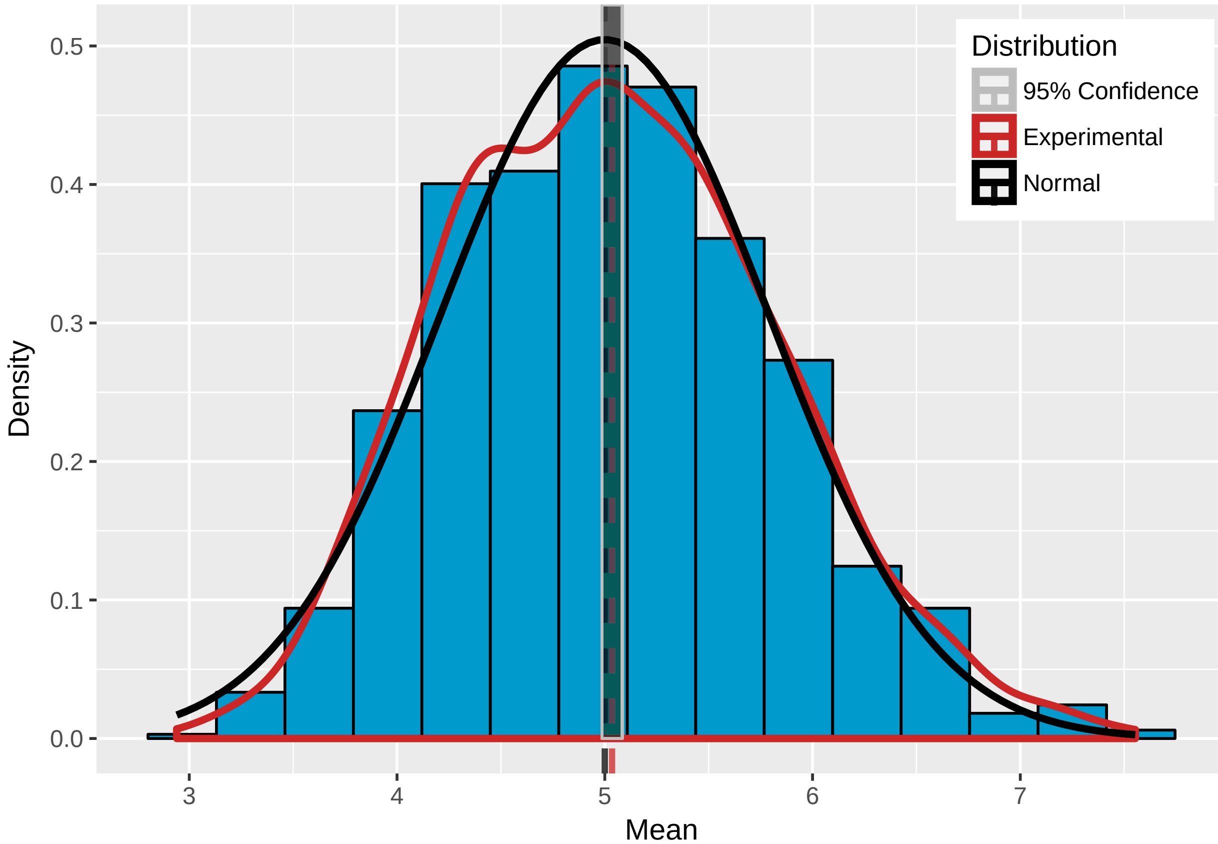
```
## Percentage of Error
##  1.252 %
```

Distribution

The simulation produces an approximate to the normal distribution, with the experimental mean and variance in close proximity to the theoretical values. The graph below shows the density plot of the means of the 1000 simulations as it compares to the normal distribution.

```
plot1 <- ggplot(data = expo_means, aes(means)) +
    geom_histogram(aes(y = ..density..), fill = "deepskyblue3",
                   color = "black", bins = 15) +
    geom_density(aes(color = "Experimental"), lwd = 1.3) +
    stat_function(fun = "dnorm", aes(color = "Normal"), lwd = 1.3,
                  args = list(mean = theor_mean, sd = theor_sigma)) +
    geom_vline(aes(xintercept = exp_mean, color = "Experimental"),
               lwd = 1.1, linetype = "dashed", alpha = .75) +
    geom_vline(aes(xintercept = theor_mean, color = "Normal"),
               lwd = 1.1, linetype = "dashed", alpha = .75) +
    scale_color_manual(name = "Distribution",
                       values = c("Normal" ="black",
                                  "Experimental"= "firebrick3",
                                  "95% Confidence" = "grey")) +
    geom_rect(aes(xmin=lower_conf, xmax=upper_conf, ymin=0, ymax=Inf,
```

```
                    color = "95% Confidence"), alpha = .01) +
    theme(legend.position = c(.875,.85)) +
    xlab("Mean") + ylab("Density")
print(plot1)
```



As can be seen in the graph the calculated distribution of means of randomly sampled exponential distributions closely approximates the normal distribution. The theoretical mean is within the 95% confidence interval of the experimental mean. The difference in the two distributions could be reduced if the number of simulations increased. With an increase in the number of averages the experimental distribution would get closer to the normal distribution.