Statistical Inference Course Project (Part 1)

MsGret

10 August 2020

The project consists of two parts: simulation (part 1) and basic inferential data analysis (part 2). In part 1 we investigate the exponential distribution in R and compare it with the Central Limit Theorem.

The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. We will investigate the distribution of averages of 40 exponentials.

Simulations

Set lambda for all of the simulations, number of exponentials, number of simulations:

```
lambda <- .2
nexp <- 40
ns <- 1000
```

Simulate the distribution of 1000 averages of 40 exponentials:

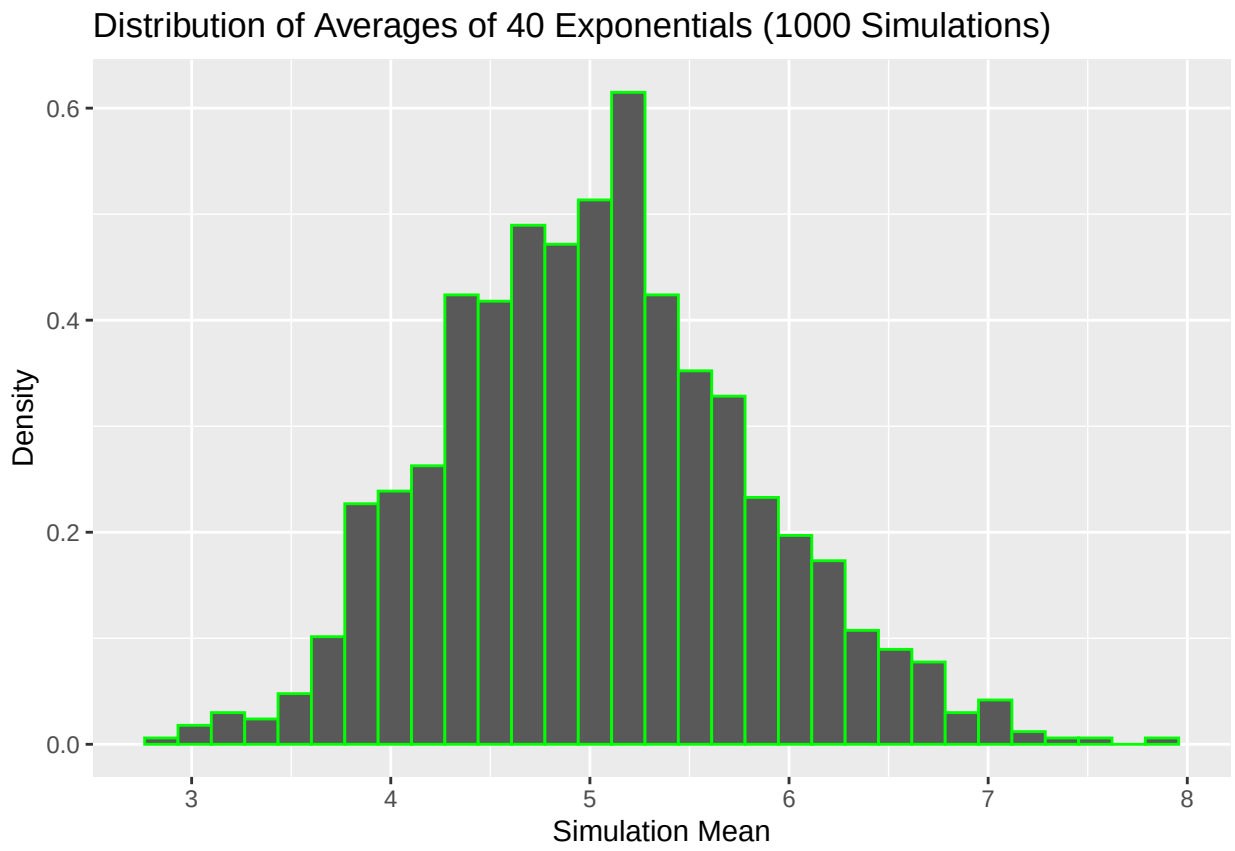
```
set.seed(2020)
mns = NULL
for (i in 1 : ns)
    mns = as.data.frame(rbind(mns, mean(rexp(n = nexp, rate = lambda))))
names(mns) <- "average"
```

Plot result of simulation:

```
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
ggplot(data = mns, aes(x = average)) +
    geom_histogram(aes(y = ..density..),
                   binwidth = (max(mns$average) - min(mns$average)) / 30,
                   colour = "green") +
    labs(x = "Simulation Mean", y = "Density",
         title = "Distribution of Averages of 40 Exponentials (1000 Simulations)")
```



Sample Mean versus Theoretical Mean

The theoretical mean of exponential distribution is 1/lambda. Find the mean from the simulation:

```
theor_mean <- 1 / lambda
theor_mean
```

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

```
sim_mean <- mean(mns$average)
sim_mean
```

```
## [1] 5.033948
```

```
diff_mean <- abs(sim_mean - theor_mean)
```

As we can see in plot bellow Sample Mean and Theoretical Mean are very close (difference is 0.0339482).

Sample Variance versus Theoretical Variance

The theoretical standard deviation of exponential distribution is 1/lambda. Find the Theoretical Variance and Sample Variance of 40 exponentials:

```
theor_var <- (1 / lambda^2) / nexp
theor_var
```

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

```
sim_var <- var(mns$average)
sim_var
```

```
## [1] 0.6070127
```

```
diff_var <- abs(sim_var - theor_var)
```

As we can see in plot bellow Sample Variance and Theoretical Variance are very close (difference is 0.0179873).

Plot Results

Plot the theoretical (red line) and simulation (yellow dashed line) means and :

```
ggplot(data = mns, aes(x = average)) +
    geom_histogram(aes(y = ..density..),
                   binwidth = (max(mns$average) - min(mns$average)) / 30,
                   colour = "green") +
    stat_function(fun = dnorm, args = list(mean = sim_mean, sd = sqrt(sim_var)),
                  colour = "yellow", size = 1.5, linetype = 2) +
    stat_function(fun = dnorm, args = list(mean = theor_mean, sd = sqrt(theor_var)),
                  size = 1, colour = "red") +
    geom_vline(xintercept = theor_mean, colour = "red", size = 1) +
    geom_vline(xintercept = sim_mean, colour="yellow", size = 1.5, linetype = 2) +
    labs(x = "Simulation Mean", y = "Density",
         title = "Theoretical Distribution versus Sample Result",
         caption = "red - theoretical mean and distribution,
         yellow - sample mean and distribution")
```