Part 1 Simulation Exercise

Volha Leusha

Introduction

In this project exponential distribution is compared with the Central Limit Theorem. The properties of the distribution of the mean of fourty exponentials are illustrated via simulation and associated explanatory text.The simulation is performed one thousands times and results for mean and varience accross simulations are compared to theoretical values.

Simulations

In this paragraph, 1000 simulations of 40 exponentials are performed. Then the mean is taken accross exponentials for each simulation separatelly.

```
#Create matrix with dimensions 40 by 1000 filled with exponentialy
#distributed data and find the mean for each simulation:
n<- 40
simulations <- 1000
lambda<-.2
mu<- apply(matrix(data =rexp(n*simulations, rate= lambda),
                  nrow=n, ncol=simulations), 2, mean)
```

Sample Mean versus Theoretical Mean

The Central Limit Theorem states that the sample average is approximately normally distributed with a mean given by the population mean (as number of variables tends towards infinity).

```
#Calculate sample mean versus theoretical mean:
smean<- mean(apply(matrix(data =rexp(n*1000, rate= lambda),
                          nrow=n, ncol=simulations), 2, mean))
tmean<- 1/lambda
```

Theoretical mean is:

```
tmean
```

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

Sample mean is:

```
smean
```

```
## [1] 4.983805
```

It can be seen that sample mean is approximatelly equal to theoretical mean.

Sample Variance versus Theoretical Variance

The Central Limit Theorem states that the sample mean is approximately normally distributed with a varience a variance given by the standard error of the mean (as number of variables tends towards infinity). Theiretical varience in this case is population varience divided by sqrt(n).

```
#Calculate sample mean versus theoretical mean:
ssdt<- sd(apply(matrix(data =rexp(n*1000, rate= lambda),
                       nrow=n, ncol=simulations), 2, mean))
tsdt<- (1/lambda)/sqrt(n)
```

Theoretical mean is:

```
tsdt
```

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

Sample mean is:

```
ssdt
```

```
## [1] 0.803797
```

It can be seen that simulation sample varience is approximatelly equal to theoretical varience.

Distribution

Below data disctribution for sample means, explonentially distributed and normally distribured data is shown:

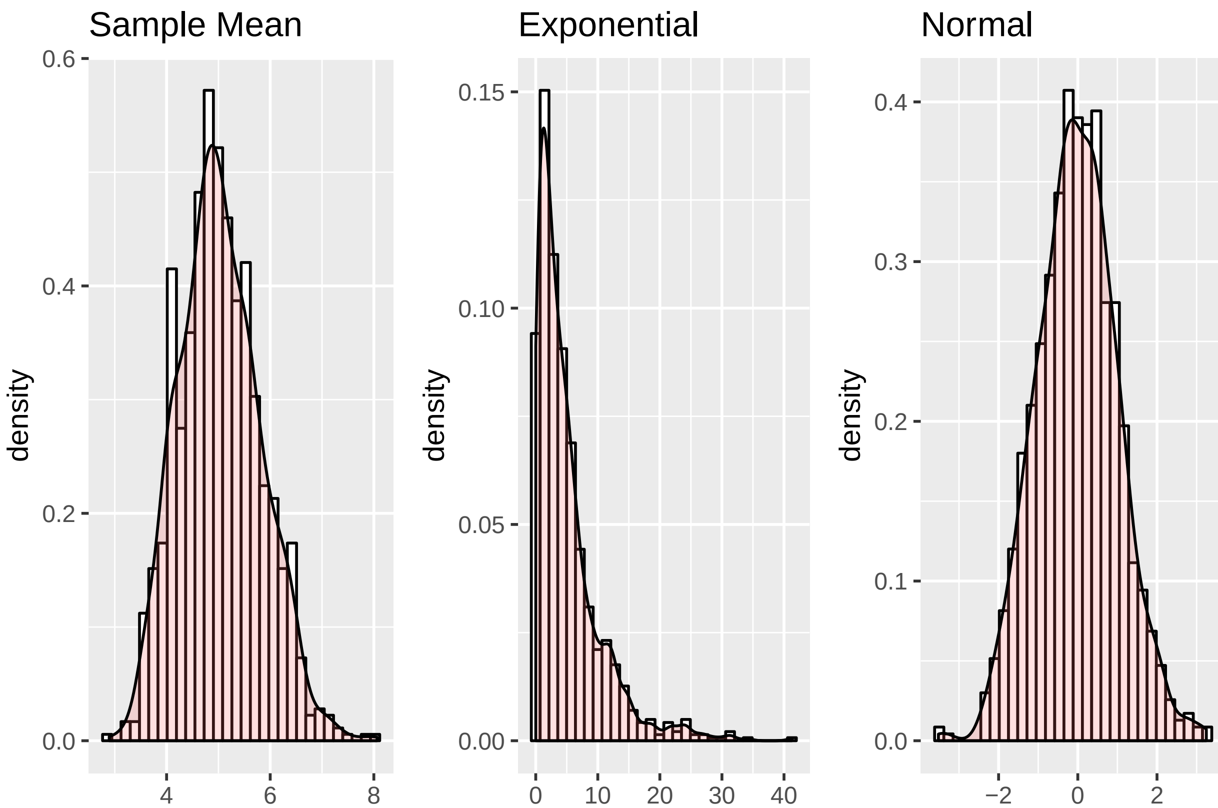
```
#create density histogram for sample means
g1<- ggplot(data=data.frame(mu), aes(mu))+
    geom_histogram(aes(y = ..density..), colour="black", fill="white") +
    geom_density(alpha = 0.2, fill = "#FF6666")+
    labs(title="Sample Mean", x="")

#create density histogram for distribution variable example:
expvar<- rexp(simulations, rate =.2)
g2<- ggplot(data=data.frame(expvar), aes(expvar))+
    geom_histogram(aes(y = ..density..), colour="black", fill="white")+
    geom_density(alpha = 0.2, fill = "#FF6666")+
    labs(title="Exponential", x="")

#create density histogram for normal distribution example:
randvar<- rnorm(simulations)
g3<- ggplot(data=data.frame(randvar), aes(randvar))+
    geom_histogram(aes(y = ..density..), colour="black", fill="white")+
    geom_density(alpha = 0.2, fill = "#FF6666")+
    labs(title="Normal", x="")

#arrange all three graphs on one page
grid.arrange(g1, g2, g3, nrow = 1)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



It can be seen that sample mean distribution looks more like normally distributed data rather than the original exponential distribution.