Project Report 1

Yilun Zhang

Wednesday, October 22, 2014

Simulation

This is the code to generate random number with seed 10086 and simulate the mean of 40 exponential distribution with rate 0.2 for 1000 times.

```
set.seed(10086)
n = 1000
means <- c()
for (i in 1:n){
  x <- rexp(40,0.2)
  means <- c(means,mean(x))
}
```

Mean value of the 1000 simulated means

Calculate the mean of the generated 1000 mean values.

```
mean(means)
```

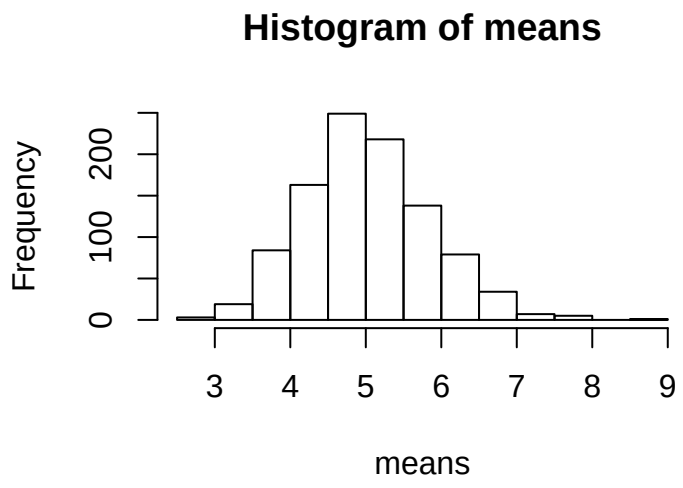
```
## [1] 5.031191
```

The mean value from the simulation is 5.031191.\ The theoretical mean of exponential distribution with rate 0.2 is 1/0.2=5, so the simulation gave us a very reasonable and close mean value.\

Histogram of the distribution of means

The following histogram shows the distribution of the mean values.

```
hist(means)
```



Variation in the mean values

This is the code to calculate the standar deviation of the simulated means.

```
# simulated sd
sd(means)
```

```
## [1] 0.8222948
```

```
# theoretical sd
1/0.2/sqrt(40)
```

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

The simulated standard deviation of the mean is 0.8222948.\ Theoretically, the standard deviation would be the 0.7905694 since it is the distribution of the mean of 40 exponential distributions, and we need to divide it by the square root of sample size.\ By comparision, there is a ~0.03 difference between the simulated sd and theoretical sd, this difference is believed to be smaller if we simulate more times (more than 1000).

Comparing the distribution to normal distribution

By looking at the histogram created above, we can see a bell shape centered around 5 with an outlier at around 9, but overall it looks approximately normal.\ However by conducting the formal normal distribution test, the Shapiro-Wilk test, the p-value is less than 0.05 (0.0001196) and we reject the hypothesis that it is normal.

```
shapiro.test(means)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  means
## W = 0.993, p-value = 0.0001196
```

95% Confidence Interval of the means

The following code produces the 95% confidence interval of the mean value of 40 exponential random variable with rate 0.2.

```
mean(means) + c(-1,1)*1.96*5/sqrt(40)
```

```
## [1] 3.481675 6.580707
```