Statistical inference Project 1 – Simulation

Yangjing Long

19.09.2014

Introduction

The aim of this simulation is to investigate the distribution of averages of 40 exponential(0.2)s. A thousand simulated averages of 40 exponentials is done.

Pre-processing

First we set seed and intialize the parameters.

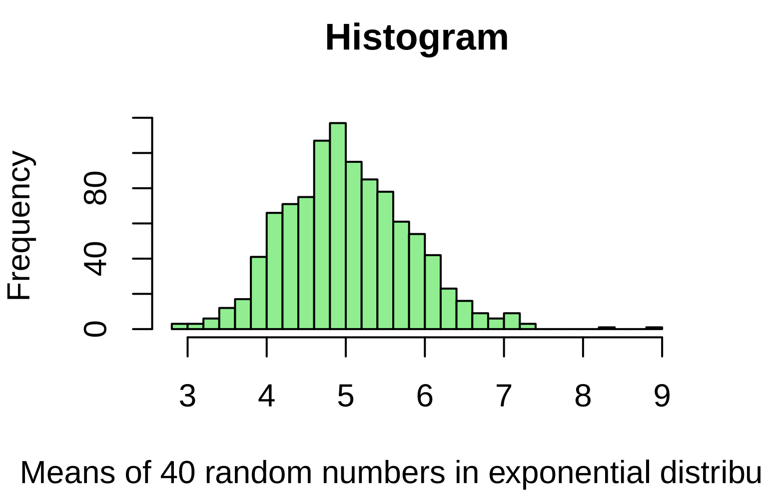
```
set.seed(2000)
lambda = 0.2; n=40
```

Simulation

Each time we randomly generate 40 numbers in exponential distribution with parameter 0.2, we repeat it for 1000 times. We record the data and plot a histogram to observe its distribution. The histogram looks like that the the 1001 means is normal distributed.

```
a<-  mean(rexp(n, lambda))
  for (i in 1:1000){
  a <- append(a, mean(rexp(n, lambda)))
  }

# histogram
hist(a, main="Histogram", xlab= "Means of 40 random numbers in exponential distribution", breaks=40, col
```



Question 1

where the distribution is centered at and compare it to the theoretical center of the distribution?

```
cat("Theoretical Center(Mean)=",1/lambda,"Actual Center(Mean)=",mean(a),"\n")
```

```
## Theoretical Center(Mean)= 5 Actual Center(Mean)= 5.029
```

Question 2

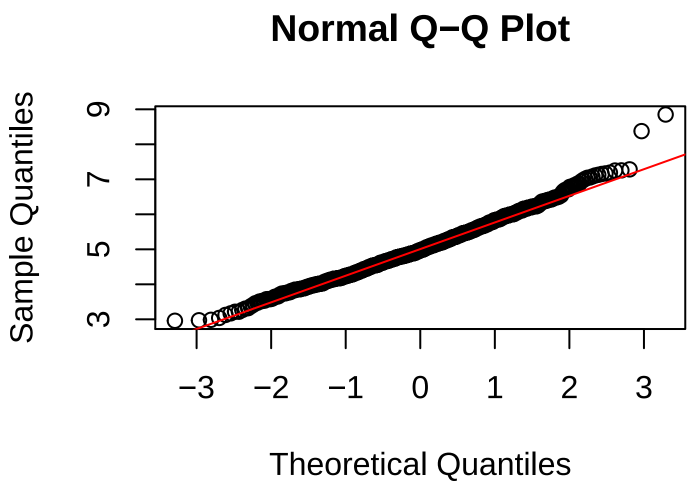
how variable it is and compare it to the theoretical variance of the distribution.

```
cat("Theoretical Variance = ",((1/lambda)^2)/n,"Actual Variance = ",var(a),"\n")
```

```
## Theoretical Variance =  0.625 Actual Variance =  0.6164
```

Question 3 Approximately normal.

We can use Q-Q plots coupled with histograms in Answer 1 to check the assumption that the underlying population is normally distributed. QQ plot shows the observations follow approximately a normal distribution, the resulting plot is roughly a straight line with a positive slope, thus showing that data does seem to follow approximately a normal distribution.



Question 4 Evaluate the coverage of the confidence interval for 1/lambda.

Theoritially, assuming the data is nomally distributed, than the 95% confidential interval is for 1/lambda is caculated as required. The interval is (3.450484, 6.549516). We test how many of our 1001 simulation means are in this interval, it turns out that 959 out of 1001 means are in this interval, which is 95.8%. It is quite confidential.

```
interval <- 1/lambda + c(-1,1)*1.96*(1/lambda)/sqrt(n)
quant <- sum(interval[1]<= a & a <= interval[2])
library('scales')
percent(quant / 1001)
```

```
## [1] "95.8%"
```