Statistical Inference Course Project - Part 1

Arash Farazdaghi

August 2014

Assignment Details

The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also also 1/lambda. Set lambda = 0.2 for all of the simulations. In this simulation, you will investigate the distribution of averages of 40 exponential(0.2)s. Note that you will need to do a thousand or so simulated averages of 40 exponentials.

Global Directives

Environment setup including the libraries and knitr parameters required.

```
knitr::opts_chunk$set(echo=TRUE)
```

Setup Simulation

Setup the simulation based on the given parameters (lambda=0.2, 1000 runs of 40 each).

```
# set values of given parameters
sim.lambda <- 0.2
sim.nexp <- 40
sim.nrun <- 1000

# run simulation to create data set
set.seed(1234)
sim.data <- data.frame(cbind(1:sim.nrun, sapply(1:sim.nrun,
                      function(x) {mean(rexp(sim.nexp, sim.lambda))})))
names(sim.data) <- c("Run","Mean")
```

Data Exploration

Question 1

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

```
# distribution center
mean.dist <- mean(sim.data$Mean)
# theoretical center
mean.exp <- (1/sim.lambda)
```

The center of the distribution, 4.9742, very close to the theoretical center of 5.

Question 2

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

```
# distribution center
var.dist <- var(sim.data$Mean)
# theoretical center
var.exp <- ((1/sim.lambda)/sqrt(sim.nexp))^2
```

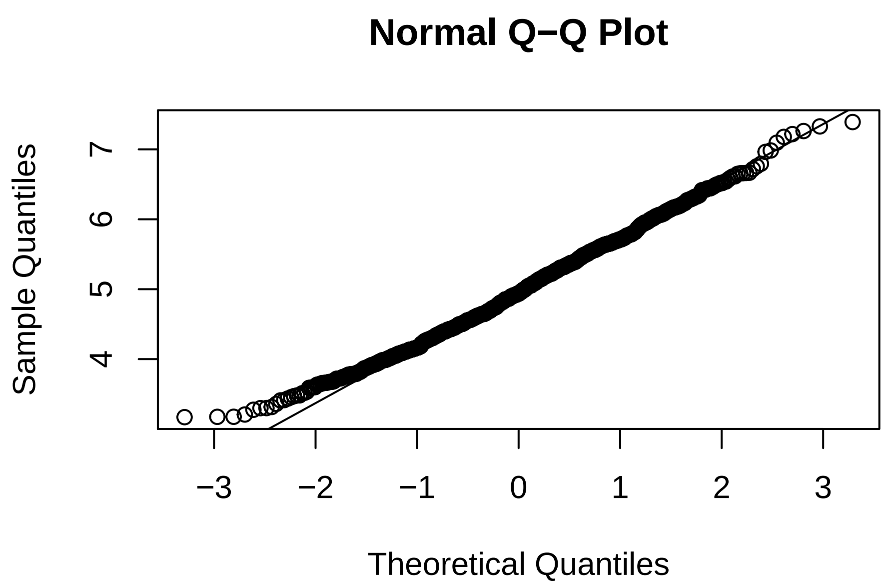
The variance of the distribution, 0.5707, very close to the theoretical variance of 0.625.

Question 3

Show that the distribution is approximately normal.

In order to evaluate whether the distribution is normal, we use the qqnorm function within R. This compares the sample and theoretical quantiles. If the data points on the plot fall along the line, then we can conclude that the ditribution is approximately normal.

```
qqnorm(sim.data$Mean)
qqline(sim.data$Mean)
```



As can be seen in the plot above, the sample distribution from the simulation is approximately normal.

Question 4

Evaluate the coverage of the confidence interval for 1/lambda: X ± 1.96 * (S/sqrt(n)).

```
conf <- mean(sim.data$Mean) + c(-1,1)*1.96*sd(sim.data$Mean)/sqrt(sim.nrun)
```

The 95% confidence interval is 4.92741753203567, - , 5.02106001046738.