Statistical Inference Project - Part 1

Alan C Bonnici

Tuesday, September 16, 2014

Synopsis

This project attempts to answer the following question:

> 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.
>
> Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponential(0.2)s. You should > 1. Show where the distribution is centered at and compare it to the theoretical center of the distribution. 2. Show how variable it is and compare it to the theoretical variance of the distribution. 3. Show that the distribution is approximately normal. 4. Evaluate the coverage of the confidence interval for 1/lambda: $\bar{X} \pm 1.96\frac{S}{\sqrt{n}}$.

Generating the Data

To ensure code reproducability, the random seed is set and data to be used in the simulation is generated.

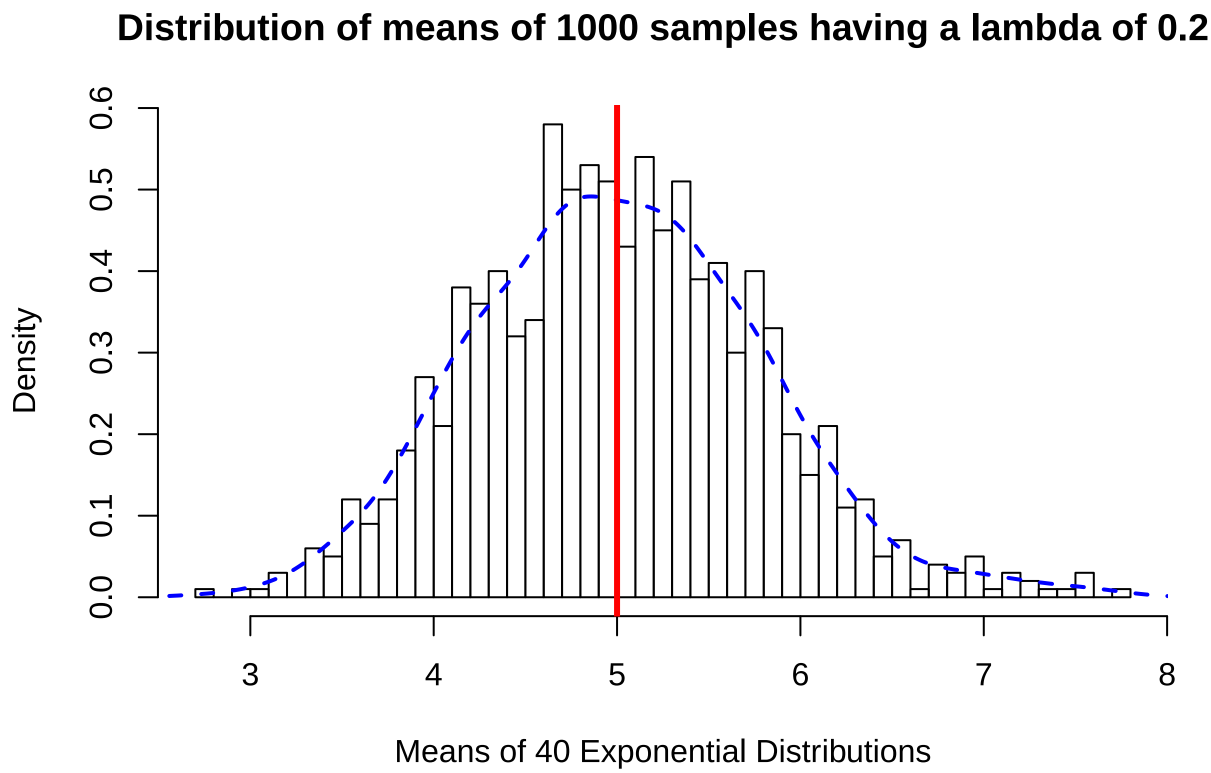
```
set.seed(123)
lambda <- 0.2
numTests <- 1000
testCount <- 40
data <- matrix(rexp(numTests * testCount, rate=lambda), nrow = numTests)
means <- rowMeans(data)

# Compute the mean of each row of the generated data.
meanDist <- apply(data, 1, mean)
```

The mean (center) of the distribution from the generated data is 5.012. This compares very well to the theoretical mean of $\frac{1}{\lambda}$ which works out to 5.

The standard deviation of the distribution from the generated data is 0.78. This compares very well to the expected sd of $\frac{1}{\lambda\sqrt{n}}$ which works out to 0.791.

The variance of the distribution from the generated data is 0.609. This compares to the expected variance of 0.625.

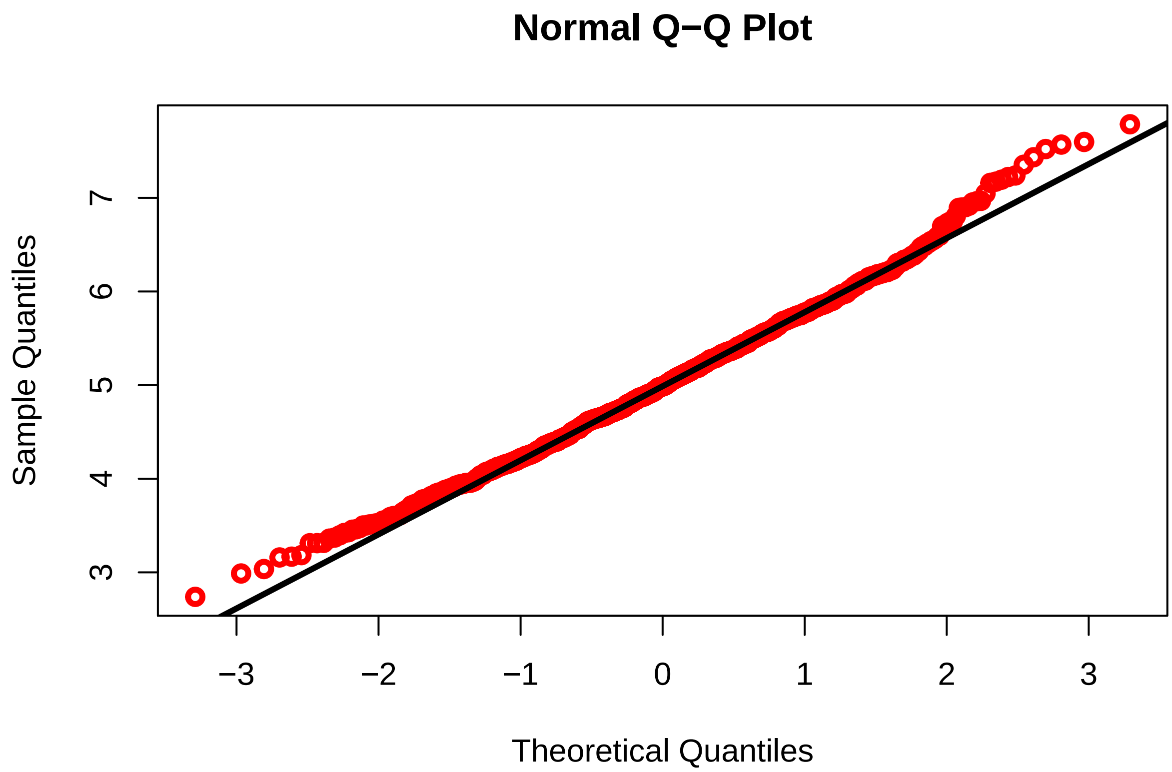
```
# plot the histogram of the means
hist(means, breaks=50, prob=TRUE,
     main="Distribution of means of 1000 samples having a lambda of 0.2",
     xlab="Means of 40 Exponential Distributions",
     ylab="Density")
lines(density(means), col="blue", lwd=2, lty=2)
# show the mean of distribution
abline(v=1/lambda, col="red", lwd=3)
```



The plot above indicates that this the samples follow a normal distribution as indicated by the central limit theorem. The computed statistics converge to their theoretical equivalents. The red line shows the theoretical mean while the blue dashed line is the density curve.

the quantile-quantile plot below show that although the tails diverge from the norm, on the whole the two plots superimpose one another thereby supporting the fact that the result follows a normalization distribution as described by the CLT.

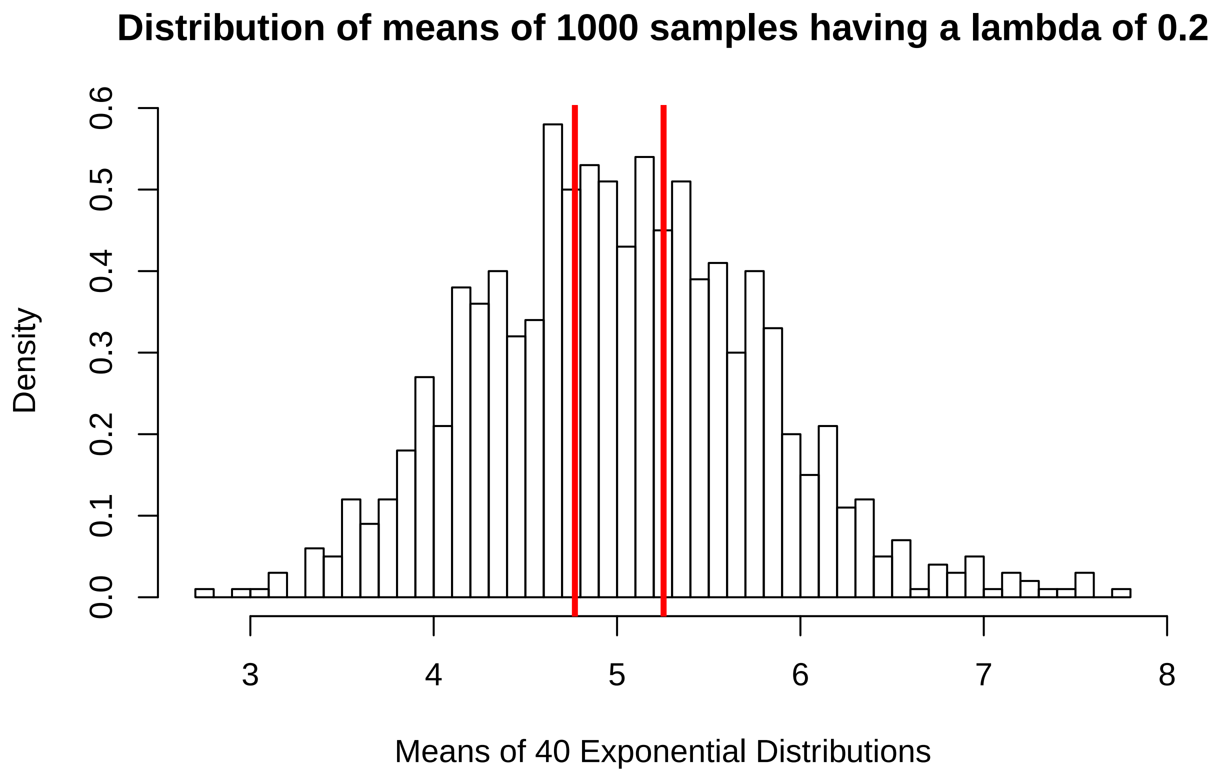
```
qqnorm(means, col="red", lwd=3)
qqline(means, col="black", lwd=3)
```




```
CI <- 1.96
ll <- mean(means) - CI * (sd(means)/sqrt(testCount))
ul <- mean(means) + CI * (sd(means)/sqrt(testCount))
```

The 95% confidence interval, spans from `4.77` and `5.254`.

```
# plot the histogram of the means
hist(means, breaks=50, prob=TRUE,
     main="Distribution of means of 1000 samples having a lambda of 0.2",
     xlab="Means of 40 Exponential Distributions",
     ylab="Density")
# show the two CIs
abline(v=c(ll, ul), col="red", lwd=3)
```


Distribution of means of 1000 samples having a lambda of 0.2
Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
3
4
5
6
7
8
Means of 40 Exponential Distributions