Statistical Inference: Exponential Distribution Simulation

Jacques Botes

September 2014

Problem Statement

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.

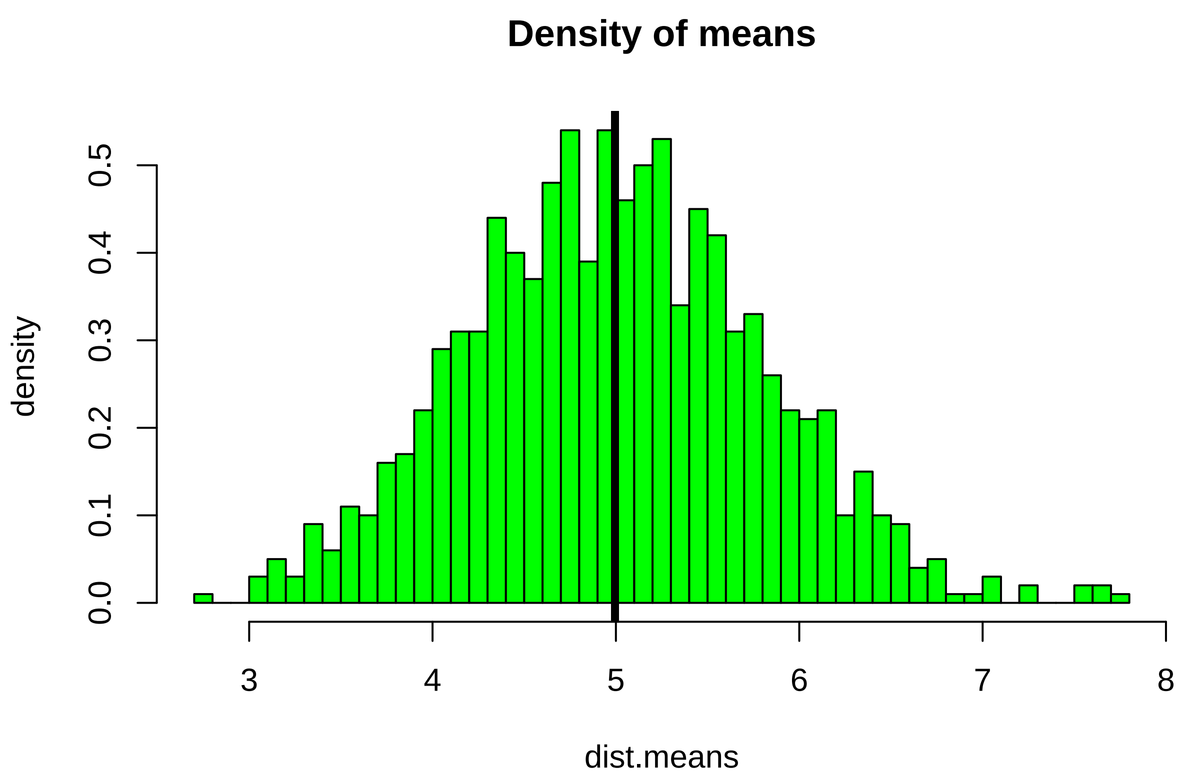
1. Show where the distribution is centered at and compare it to the theoretical center of the distribution.
2. Show how variable it is and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.
4. Evaluate the coverage of the confidence interval for $1/\lambda = \bar{X} \pm 1.96\frac{S}{\sqrt{n}}$.

```
## Loading required package: ggplot2
```

First setup the simulation and get the means of the 1000 simulations

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

```r
hist(dist.means,breaks=sample.size,prob=T,col="green",main="Density of means",ylab="density")
abline(v=mean(dist.means),col="black",lwd=4)
```



```r
center.ac <- mean(dist.means)
center.th <- 1/lambda
```

The Theoretical center of the distribution is calculated as $1/\lambda = 1/0.2 = 5$. The center of the distribution is 4.9951. The black line in the above plot displays the center.

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

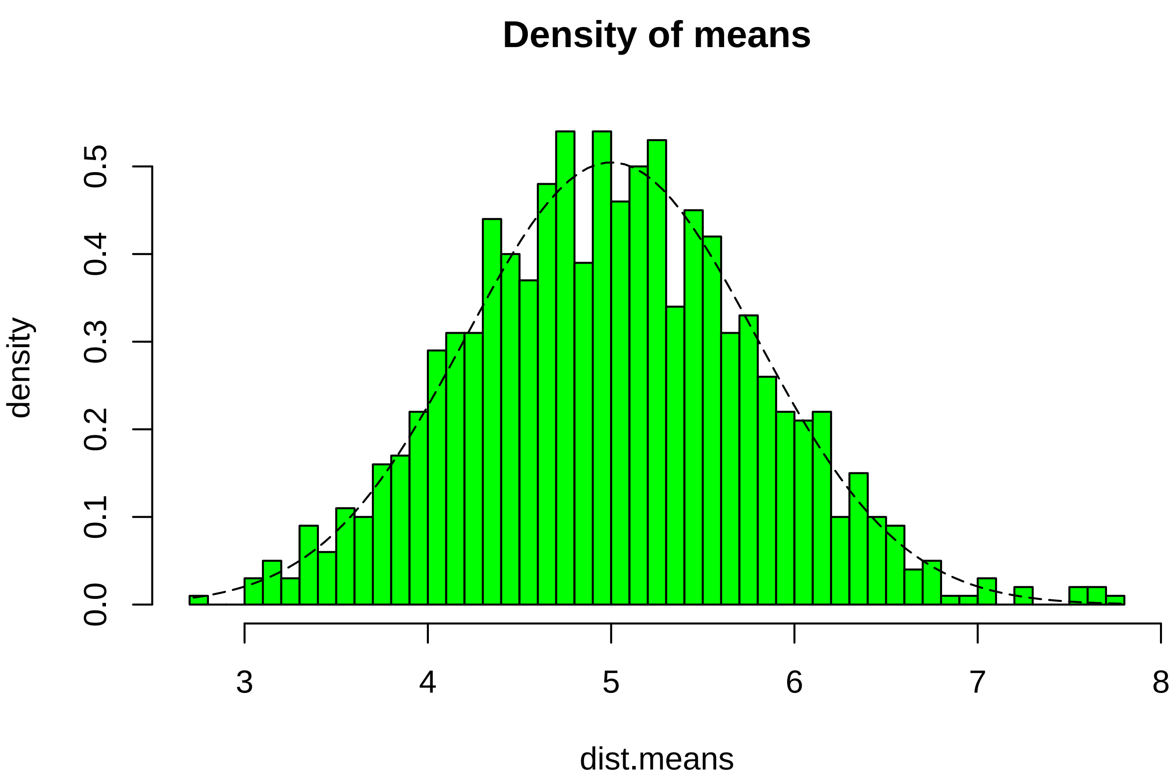
```r
sd.ac <- sd(dist.means)
sd.th <- (1/lambda)*(1/sqrt(sample.size))

var.ac <- sd.ac^2 ## = var(dist.means)
var.th <-((1/lambda)*(1/sqrt(sample.size)))^2
```

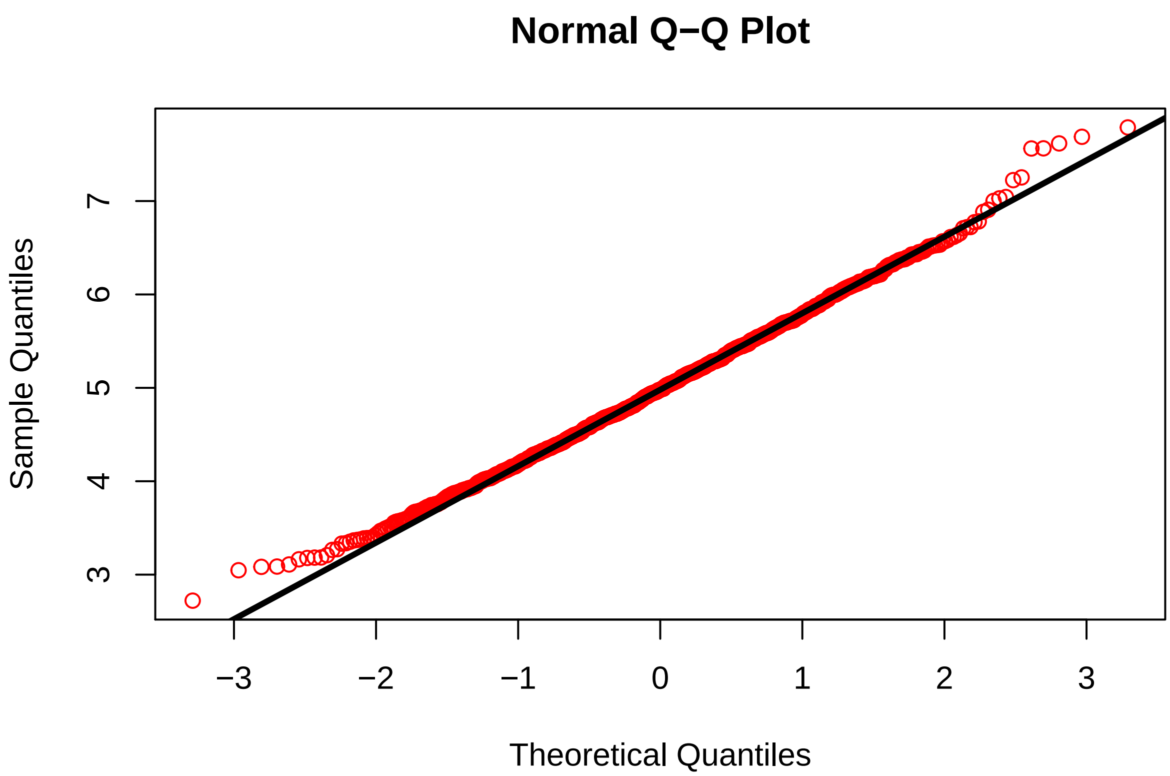
Standard Deviation of the distribution is 0.7985 with the theoretical SD calculated as 0.7906. The Theoretical variance is calculated as $(\frac{1}{\lambda} * \frac{1}{\sqrt{n}})^2 = 0.625$. Actual variance of the distribution is 0.6376

3. Show that the distribution is approximately normal.

```r
xfit <- seq(min(dist.means), max(dist.means), length=100)
yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(sample.size)))
hist(dist.means,breaks=sample.size,prob=T,col="green",main="Density of means",ylab="density")
lines(xfit, yfit, pch=22, col="black", lty=5)
```




```r
qqnorm(dist.means,col="red" ) #normal QQ plot
qqline(dist.means,lwd=3) ##theoretical line
```

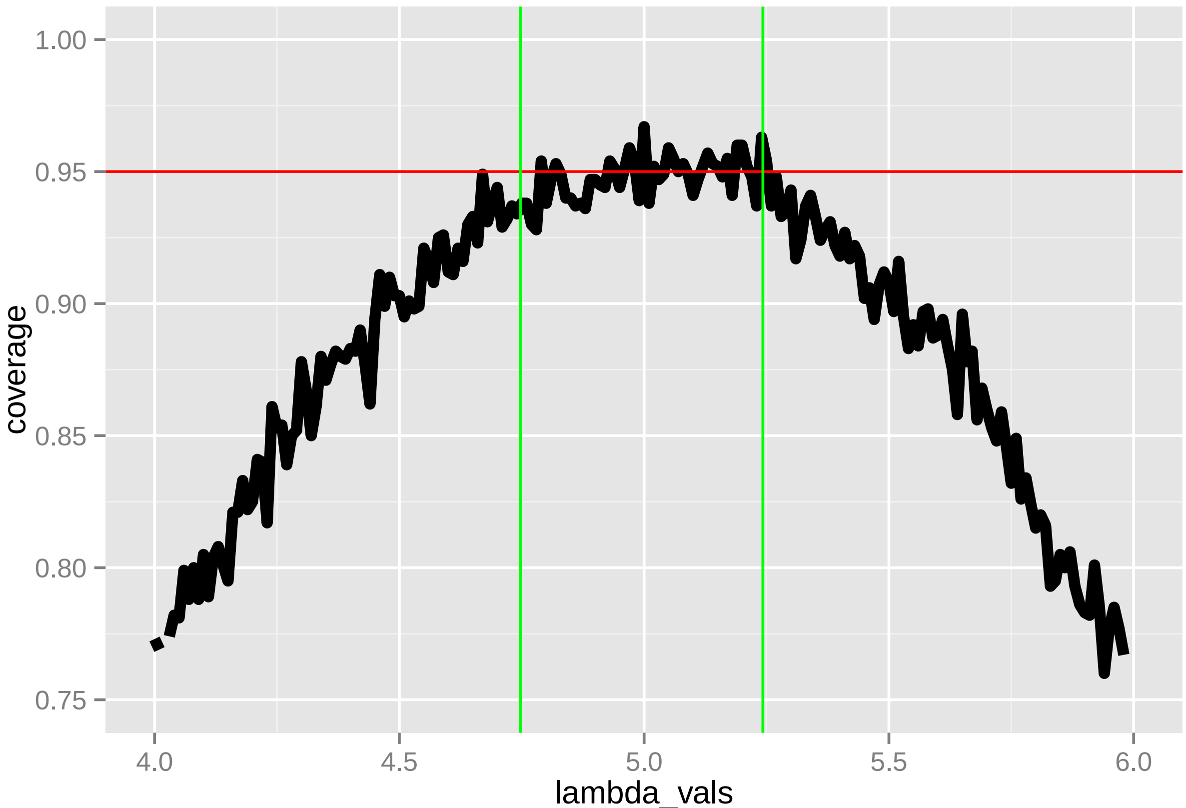


In the first plot we have overlayed a normal distribution (in black) over the density plot taken from the means of the exponential distribution. To confirm the distribution we use a qqnorm plot and overlay the theoretical line. We notice on teh QQ plot that most of the red is on the theoretical normal line and only deviates at the begging and end due to the skewness of the exponential distribution.

```
lambda_vals <- seq(4, 6, by=0.01) ##we know the center is around 5
coverage <- sapply(lambda_vals, function(x) {
    mu.hats <- rowMeans(matrix(rexp(sample.size*n, rate=0.2),ncol = sample.size, nrow=n))
    int <- qnorm(0.975) * sqrt(1/lambda**2/sample.size)
    ll <- mu.hats - int
    ul <- mu.hats + int
    mean(ll < x & ul > x)
})

v <- mean(dist.means) + c(-1,1)*1.96*sd(dist.means)/sqrt(sample.size)

ggplot(data.frame(lambda_vals, coverage),
       aes(x = lambda_vals, y = coverage)) +
        geom_line(size = 2) +
        geom_hline(yintercept = 0.95,col="red")+ ylim(.75, 1.0) +
        geom_vline(xintercept=v,col="green")
```



The interval above 95% is between 4.7476, 5.2426 and is depicted in the plot above using the green lines. The red line shows the 95% confidence interval line