Statistical Inference Project

José Gildardo Rojas-Nandayapa

Saturday, September 20, 2014

Part 1: Simulation exercises 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`. For all of the simulations, `lambda = 0.2`.

The goal is to investigate the distribution of averages of 40 exponential(0.2)s and illustrate via simulation and associated explanatory text, the properties of the distribution of the mean of 40 exponential(0.2)s.

```r
# Set seed, sample size (n), number of simulations (numsim), lambda.
set.seed(1)
n<-40
numsim<-1000
lambda<-0.2
# Create matrix with 1000 simulations and 40 exponential samples.
dataset<-matrix(rexp(numsim*n, lambda),numsim)
```

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

```r
# Calculate means (1000)
means<-rowMeans(dataset)
# Calculate mean of means (Center of distribution)
mean(means)
```

```
## [1] 4.99
```

```r
# Calculate mean (Theoretical center of the distribution)
1/lambda
```

```
## [1] 5
```

First we calculated the 1000 means and then the mean of them to get the Center of Distribution very close to the actual theoretical center of the distribution `1/lambda`.

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

```r
# Calculate variance of means (Distance from center of distribution)
var(means)
```

```
## [1] 0.6177
```

```r
# Calculate variance (Theoretical variance of the distribution)
(1/lambda)^2/n
```

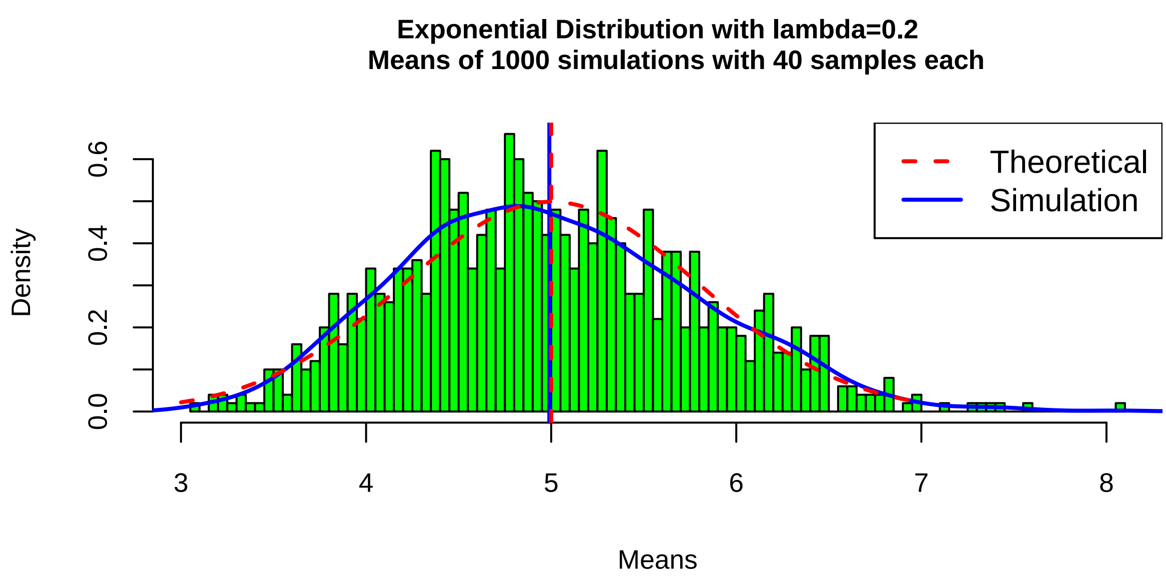
```
## [1] 0.625
```

We can see that the variance of the means (simulation) is nearly the same as the calculation (theoretical).

3. Show that the distribution is approximately normal.

The histogram shows the means of the distribution with 100 breaks.

```
hist(means, breaks=100, prob=TRUE,main="Exponential Distribution with lambda=0.2
     Means of 1000 simulations with 40 samples each",col="green",xlab="Means",
     cex.lab=.8,cex.axis=.8,cex.main=.8)
lines(density(means),col="blue",lwd=2)
abline(v=mean(means),col="blue",lwd=2)
legend("topright",c("Theoretical","Simulation"),lty=c(2,1),lwd=2,col=c("red","blue"))
# Overlap a normal distribution reference plot
x=seq(3,7,length=1000)
y=dnorm(x,mean=1/lambda,sd=.8)
lines(x,y,type="l",lwd=2,lty=c(2),col="red")
# Overlaps the mean reference.
abline(v=1/lambda, col="red",lwd=2,lty=c(2))
```



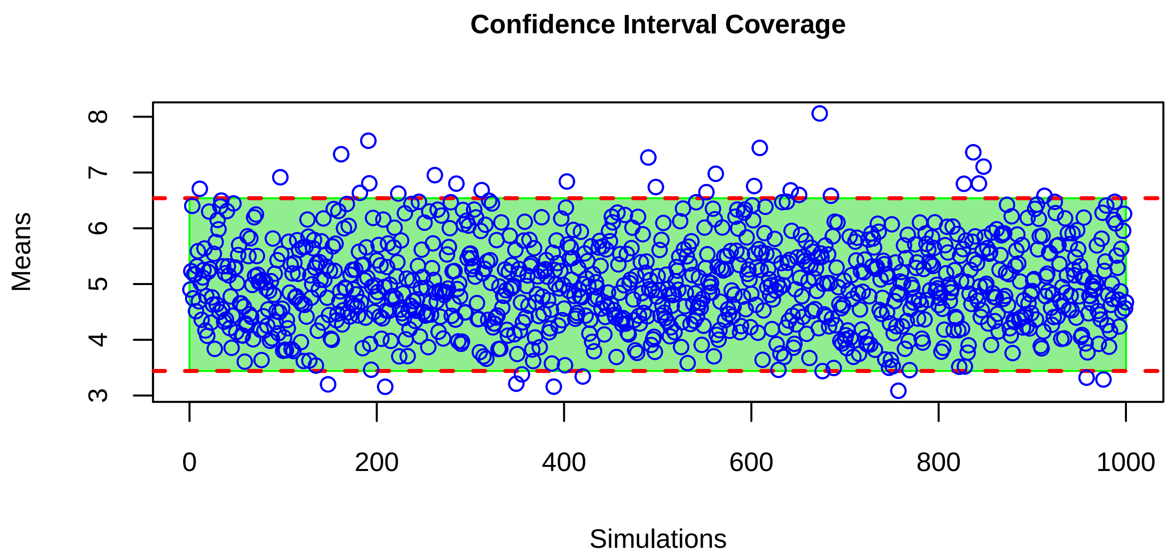
The plot shows that the average means curve (blue) is shaped nearly as a normal distribution (red). The vertical lines show the how close are the theoretical and the simulated means.

4. Evaluate the coverage of the confidence interval for 1/lambda: X¯±1.96Sn.

To evaluate the coverage we have to calculate first the confidence intervals:

```
# Calculating the standard deviation and confidence intervals
sd <- sqrt((1/lambda)^2/n)
ciLow = mean(means)-1.96*sd
ciHigh = mean(means)+1.96*sd
# Plots means for each of the 1000 simulations.
plot(means,type="p",xlab="Simulations",ylab="Means",col="blue",
     main="Confidence Interval Coverage",cex.lab=.8,cex.axis=.8,cex.main=.8)
# Plots confidence interval as a green area (area of coverage)
cord.x <- c(0,0,1000,1000)
```

```r
cord.y <- c(ciLow,ciHigh,ciHigh,ciLow)
polygon(cord.x,cord.y,col='lightgreen',border="green")
# Plots confidence interval limits in red
abline(h=ciLow, col="red",lwd=2,lty=c(2))
abline(h=ciHigh, col="red",lwd=2,lty=c(2))
# Overlaps points of means.
points(means,type="p",xlab="Simulations",ylab="Means",col="blue")
```



To calculate the coverage, we can just count the means that are within the confidence interval and then average the total count.

```r
# Initiates covered vector, counts covered means
covered <- c()
for (mn in means) {
  if ((mn>=ciLow) && (mn<=ciHigh))
      covered<-append(covered,1)
  else
      covered<-append(covered,0)
}
# Mean of count represents coverage of the confidence interval
mean(covered)
```

```
## [1] 0.964
```

Conclusions

- The distribution of the sample means is nearly normal.
- The sample means average and variance is close to the values predicted.
- The coverage of the 95% confidence interval for means is very close to the expected theoretical value.