Statistical Inference Part 1

Viktor M

Thursday, April 22, 2015

My github page

Synopsis

In this theoritical research we explore exponentinal distribution and compare it's result with CLT (Central Limit Theorem) For exponential distribution we will use rexp(n, lambda) in R, where lambda is the rate parameter. The mean of rexp is 1/lambda and the standard deviation is also also 1/lambda. We wil setlambda = 0.2 as required for all of the test simulations. We will investigate the distribution of averages of total 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. Presented result will illustrate sample mean and compare it with with theoretical mean. We also show variable sample (using it's variance) and compare to the theoretical variance of the distrubtion. And finally we will demonstrate that disttribution can be considered Normal.

Preparing Test Data

Here we assign all variables and prepare data set to be used in our research.

```
set.seed (100)

lambda <- 0.2

n <- 40
# we do 1000 simulations
simulations <- 1000

# simulate data
simulated_exponentials <- replicate(simulations, rexp(n, lambda))

# calculate mean
means_exponentials <- apply(simulated_exponentials, 2, mean)
```

Question 1.

Show the sample mean and compare it to the theoretical mean of the distribution.

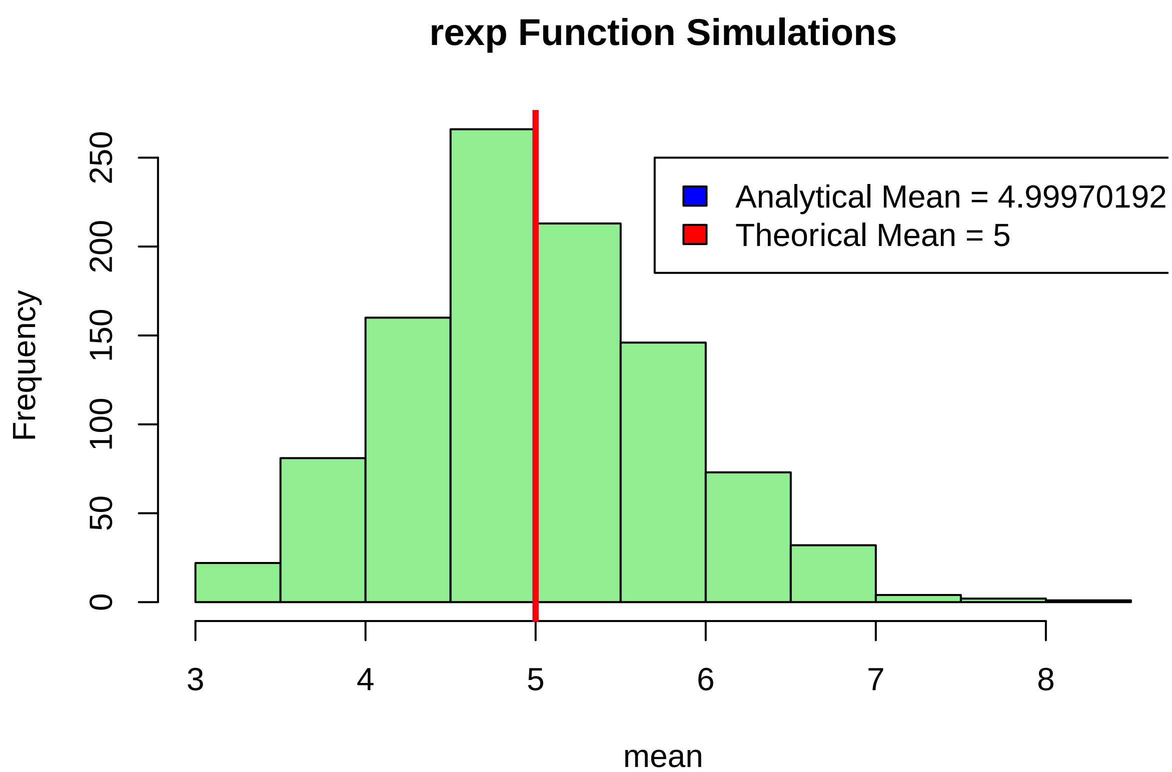
```
# distrribution mean
analytical_mean <- mean(means_exponentials)
analytical_mean
```

```
## [1] 4.999702
```

```
# analytical mean
theory_mean <- 1/lambda
theory_mean
```

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

```r
# Build graphs
am <- paste("Analytical Mean =",analytical_mean ,sep = " ")
tm <- paste("Theorical Mean =",theory_mean ,sep = " ")
hist(means_exponentials, xlab = "mean", main = "rexp Function Simulations",col="lightgreen")
abline(v = analytical_mean, col="blue",lwd=3)
abline(v = theory_mean, col = "red",lwd=3)
legend(x=5.7,y=250,legend=c(am,tm),fill=c("blue","red"))
```



Answer 1

As we can see from legend both means are very close to each other, on histogram they actually overlapped.

Question 2.

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

```r
# standard deviation of distribution
standard_deviation_dist <- sd(means_exponentials)
standard_deviation_dist
```

```
## [1] 0.8020251
```

```r
# standard deviation from analytical expression
standard_deviation_theory <- (1/lambda)/sqrt(n)
standard_deviation_theory
```

```
## [1] 0.7905694
```

```r
# variance of distribution
variance_dist <- standard_deviation_dist^2
variance_dist
```

```
## [1] 0.6432442
```

```r
# variance from analytical expression
variance_theory <- ((1/lambda)*(1/sqrt(n)))^2
variance_theory
```

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

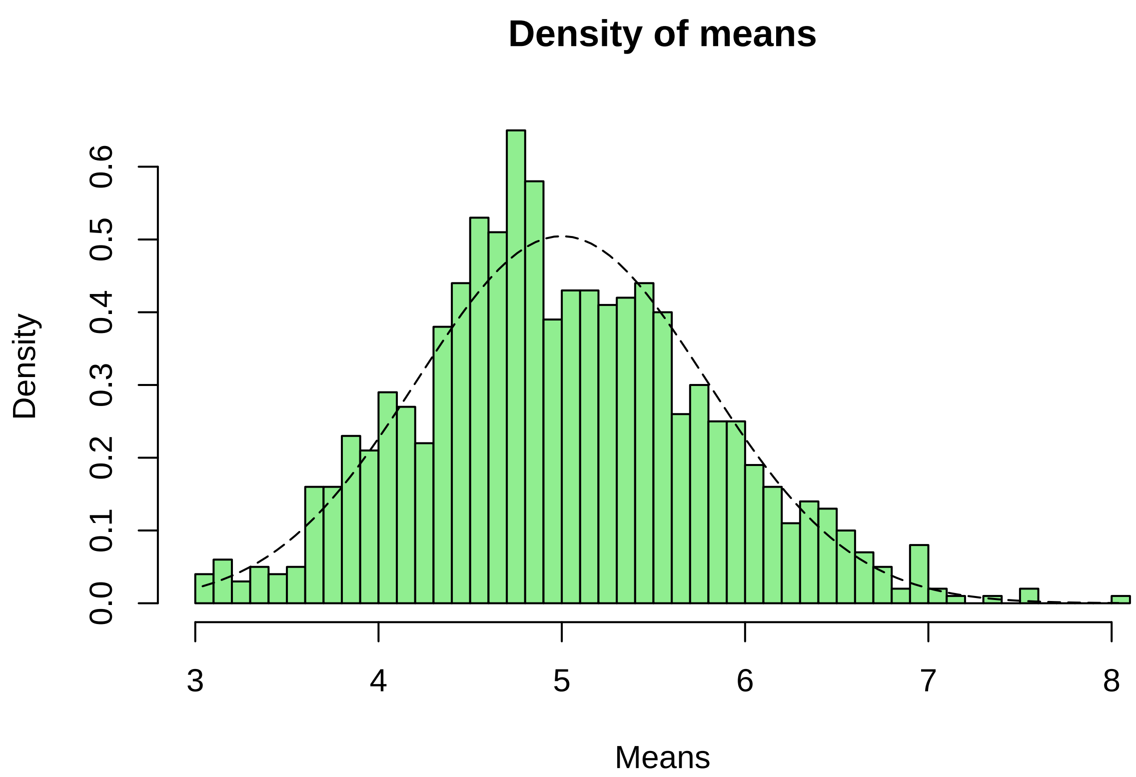
Answer 2

Standard Deviation: Distribution = 0.7931608 vs Theoretical SD calculated as 0.7905694. Variance: Theoretical variance = 0.625. vs Actual variance of the distribution = 0.6291041

Question 3.

Show that the distribution is approximately normal.

```r
xfit <- seq(min(means_exponentials), max(means_exponentials), length=100)
yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(n)))
hist(means_exponentials,breaks=n,prob=T,col="lightgreen",xlab = "Means",main="Density of means",ylab="De
lines(xfit, yfit, pch=30, col="black", lty=5)
```



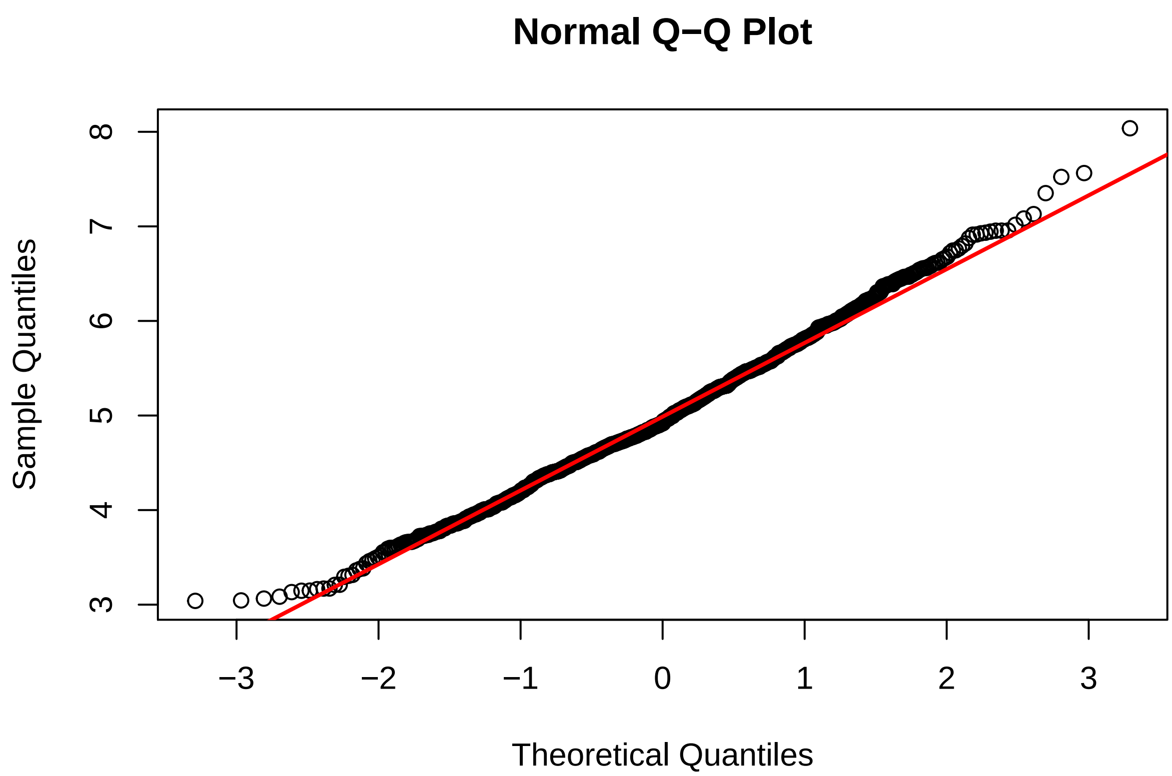
Answer 3

As we can see from graphs distrubtion is tend to be normal.

Q4. Motivating example

Compare the distribution of 1000 random uniforms and the distribution of 1000 averages of 40 random uniforms

```
require(graphics)
qqnorm(means_exponentials)
qqline(means_exponentials, col = 2,lwd=2)
```



Answer 4

To confirm CLT (Central Limit Theorem) we can see thatthe distribution of averages of 40 exponentials is very close to a normal distribution.