Exponential Distribution in R vs Central Limit Theorem

Rajesh Sankar

22-Aug-2015

Overview

In this project we will investigate the exponential distribution in R and compare it with the Central Limit Theorem (CLT). Sample mean, variance and distribution will approximate population mean, variance and distribution as sample size grows large. Regardless of the population distribution, the Central Limit Theorem states that sample means will be normally distributed about the population mean with variance inversely proportional to the sample size.

Simulations:

We will run a series of 1000 simulations to create a data set for comparison to theory. Each simulation will contain 40 observations and the expoential distribution function will be rexp(n, lambda) where lambda is the rate parameter. The mean of the exponential distribution and the standard deviation are both 1/lambda.

```
library(ggplot2)
lambda <- 0.2
n <- 40
nosim <- 1000
distExp = NULL
mean_exp <- 1/lambda
sd_exp <- 1/lambda
```

The figure below shows the exponential distribution generated using lambda as 0.2. The x-axis values range from 0 through 10.

```
# Exponential Distribution

xval <- seq(0.0,10.0,0.01)
p <- function(xval,lambda) lambda*exp(-lambda*xval)
exp_prob <- p(xval,0.2)
dat <- data.frame(xval, exp_prob)
qplot(xval,exp_prob,data=dat,geom="line",xlab="x", ylab="p(x)",main="Exponential
Distribution")
```
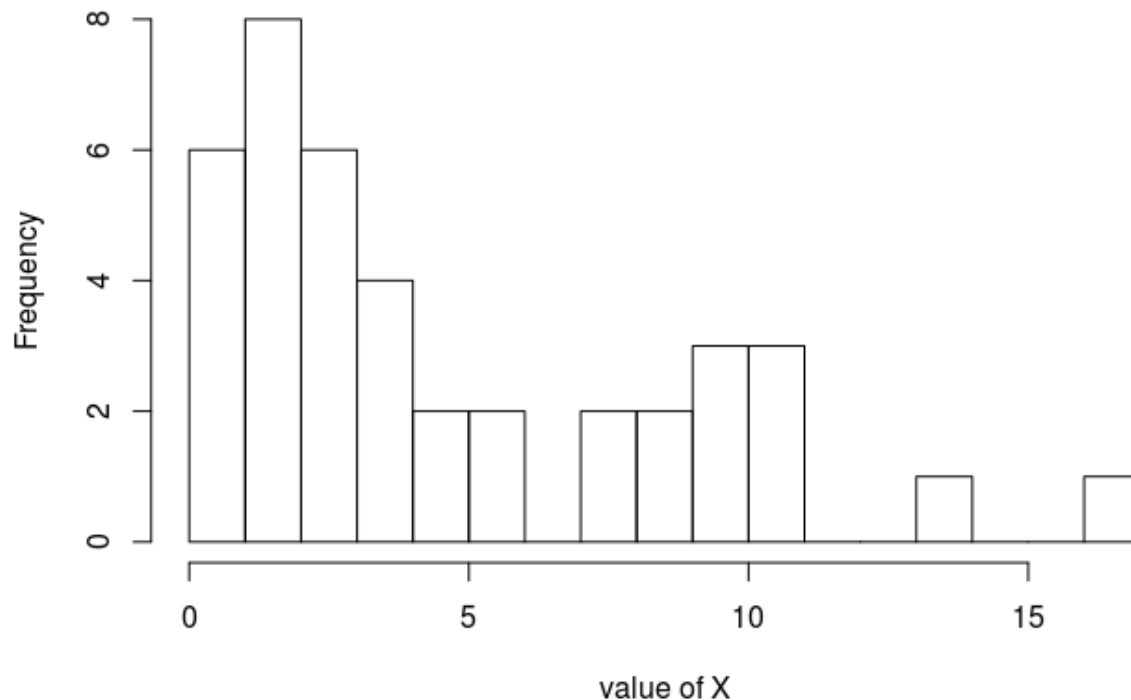
Exponential Distribution
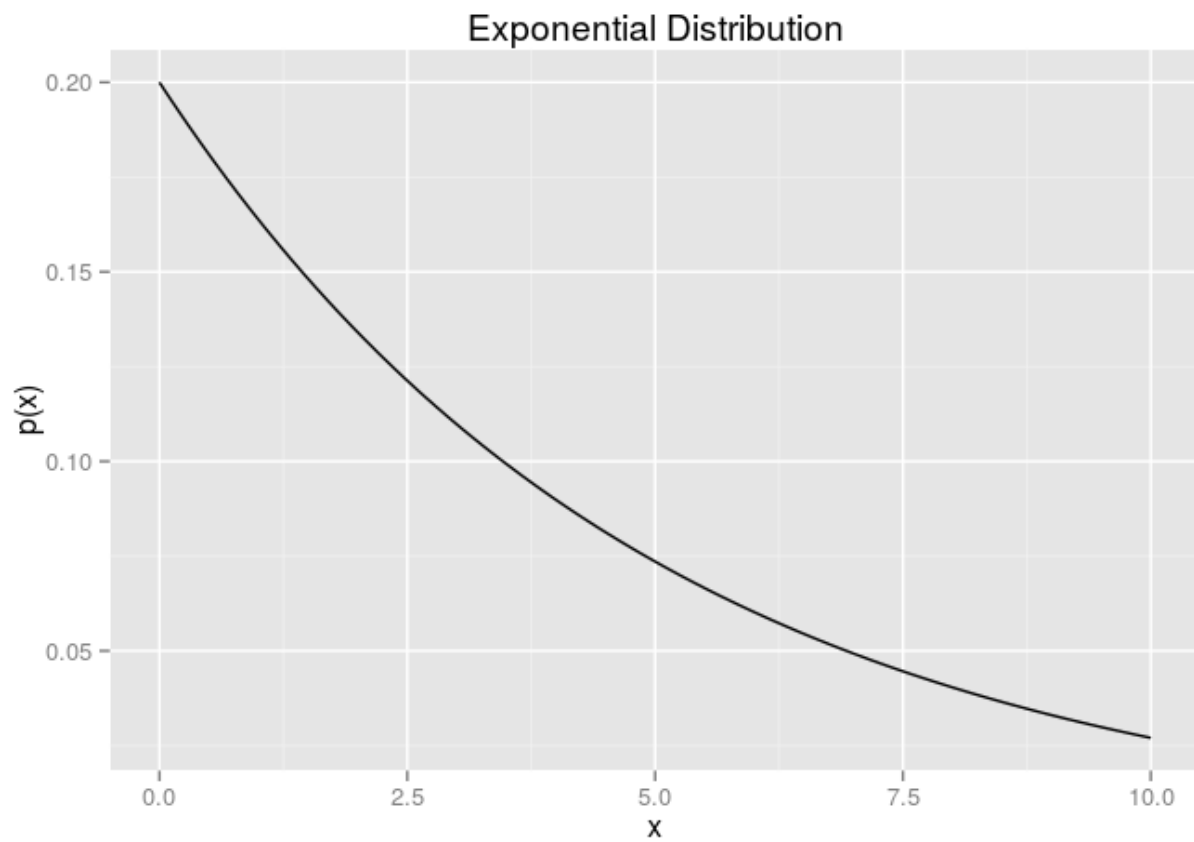
The figure below shows a histogram with 40 random samples taken from the exponential distribution.

```
# Average of 40 exponential distrubtions with lambda = 0.2
distExp <- rexp(n,lambda)
hist(distExp, main="40 Exponential Distributions", xlab="X", breaks=20)
```

40 exponential distributions

Mean value of 40 random samples of the exponential distribution is roughly near the expected value of 1/lambda (which is equal to 5 for a value of 0.2). The Standard Deviation is also roughly equivalent to the value predicted by the exponential distribution value of 1/lambda

```
print(mean(distExp))
## [1] 4.618711
print(sd(distExp))
## [1] 4.097471
print(1/lambda)
## [1] 5
```

A simulation was performed with 1000 runs of 40 random samples taken from the exponential distribution.

```
# Simulation of 1000 runs of 40 exponontial distributions
distExp_sim <- NULL
for (i in 1 : nosim) distExp_sim = c(distExp_sim, mean(rexp(n,lambda)))
```

Sample Mean versus Theoretical Mean:

With over 1000 simulations performed, sample mean of the simulated exponential distributions approaches the value of 1/lambda.

```
# mean of 1000 simulated exponential distributions
print(mean(distExp_sim))
## [1] 5.051124
# predicted value using the central limit theorem (CLT) = 1/lambda
print(1/lambda)
## [1] 5
```

Sample Variance versus Theoretical Variance:

The sample variance of the simulated exponenital distributions becomes narrower and converges to the theoretical central limit theorem (CLT) values.

```
# Standard variance of the simulated exponential distributions
print(sd(distExp_sim))
## [1] 0.8007799
# The theoretical central limit theorem value for variance is predicted by using the
equation sigma/sqrt(n) where n=40, and sigma is 1/lambda for the exponential
distribution
CLT_sd <- (1/lambda)/sqrt(40)
print(CLT_sd)
## [1] 0.7905694
```

Distribution

As can be shown in this plot, the simulation of 1000 runs of 40 random samples of the exponential distribution follows a normal distribution. The distribution clearly following a normal distribution. These results conform to the central limit theorem (CLT), which enables the use of the sample mean and sample variance to be predicted. As shown, the mean of this simulation aligns approximately to the predicted value defined by 1/lambda.

```
hist(as.numeric(distExp_sim),breaks=50,main="1000 runs of 40 Exponential
Distributions",xlab="Mean of Exponential Distribution")
mean_distExp_sim <- mean(distExp_sim)
abline(v=mean_distExp_sim, col="blue", lwd=4, lty=20)
```

1000 runs of 40 exponential distributions