Statistical Inference - Course Project - Part 1 - Simulation Exercise

Jose Maria Vilaverde (jmvilaverde)

Saturday, May 23, 2015

Overview

In this report is investigated the exponential distribution in R and compare it with the Central Limit Theorem.

The CLT states that the distribution of averages of iid variables (properly normalized) becomes that of a standard normal as the sample size increases.

The result is that $\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n-\mu)}{sd} = \frac{Estimate-Meanofestimate}{Std.Err.ifestimate}$ has a distribution like that of a standard normal for large n.

Conditions for investigation

For this investigation, the exponential distribution is simulated with R function rexp(sampleSize, lambda), where lambda is the rate parameter and sampleSize is equal to the sample size.

The distribution of averages of 40 exponentials are investigated in a thousand simulations.

Parameters:

- The sample size is 1000.
- The mean of estimate (μ) of exponential distribution is $\frac{1}{\lambda}$: * $\mu = \frac{1}{\lambda} = \frac{1}{0.2} = 5$
- The standard deviation (σ) is also $\frac{1}{\lambda}$: $\sigma = \frac{1}{\lambda} = \frac{1}{0.2} = 5$
- Lambda (λ) is set to 0.2 for all simulations: $\lambda = 0.2$
- The number of exponentials used per average simulation is 40 exponentials: n = 40
- The number of simulations is one thousand: simulations = 1000

Steps

1. Sample mean and comparation with the theoretical mean of the distribution.

Sample mean:

Is obtained a sample of one thousand exponentials and is calculated the sample mean:

> **Parameters**: $\lambda = 0.2$, sampleSize = 1000, number of exponentials per mean n = 40
>
> **Sample mean**: $\bar{X}_n = \frac{\Sigma_{i=1}^{n=1000} X_i}{n} = 5.0161985$ ***(View Code 1 & 2)***

Theoretical mean:

Is calculated the theoretical mean of the distribution:

> **Parameters**: $\lambda = 0.2$, $\mu = \frac{1}{\lambda}$
>
> **Theoretical mean** of the distribution: $\mu = \frac{1}{\lambda} = \frac{1}{0.2} = 5$ ***(View Code 3)***

Comparation:***(View Figure 1)***

Sample mean	Theoretical mean
5.0161985	5

2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

Sample variance:

> **Sample variance**: $\mathrm{Var}(\bar{X}) = \sqrt{\frac{\Sigma_{i=1}(X_i - \bar{X})^2}{n-1}} = 4.1750379$ ***(View Code 4)***

Theoretical variance of the distribution:

> **Parameters**: $\sigma = \frac{1}{\lambda}$, $\lambda = 0.2$
>
> **Theoretical variance**: $Var(\bar{X}) = \frac{\sigma^2}{\sqrt{n}} = \frac{(\frac{1}{\lambda})^2}{\sqrt{n}} = \frac{(\frac{1}{0.2})^2}{\sqrt{40}} = 3.9528471$ ***(View Code 5)***

Comparation: ***(View Figure 2)***

Sample variance	Theoretical variance
4.1750379	3.9528471

3. Show that the distribution is approximately normal.

The CLT states that the distribution of averages of iid variables (properly normalized) becomes that of a standard normal as the sample size increases.

View ***figure 3*** for comparative between normalized distribution of averages of iid variables and standard normal. In this comparative is analyzed the relation between expected quantiles and quantiles calculated for the normalized distribution of averages.

The distribution aproximates to expected values.

Simulations and figures

1. Sample mean and comparation with the theoretical mean of the distribution.

```
##Code 1: Calculate the sample
lambda = 0.2
n = 40 #number of means
sampleSize = 1000 #size of the sample
sample = NULL
#for (i in 1:sampleSize) sample = c(sample, rexp(sampleSize,lambda))
sample = for (i in 1 : sampleSize) sample = c(sample, mean(rexp(n=40,lambda)))

#Code 2: Calculate the mean of the sample
sampleMean = mean(sample)
```

sampleMean = 5.0161985

```
#Code 3: Calculate the theoretical mean
lambda = 0.2
mu = 1/lambda
theoMean = mu
```

theoMean = 5

```
#Code Figure 1: Generation of graphic of sample, sample mean and theoretical mean
hist(sample, main="Distribution of sample", xlab="Exponential values")
title(sub="Figure 1.")
legend("topright", legend = c("Sample mean", "Theoretical mean"), col = c("red", "blue"), pch="l", cex=
abline(v=sampleMean, col="red")
abline(v=theoMean, col="blue")
```

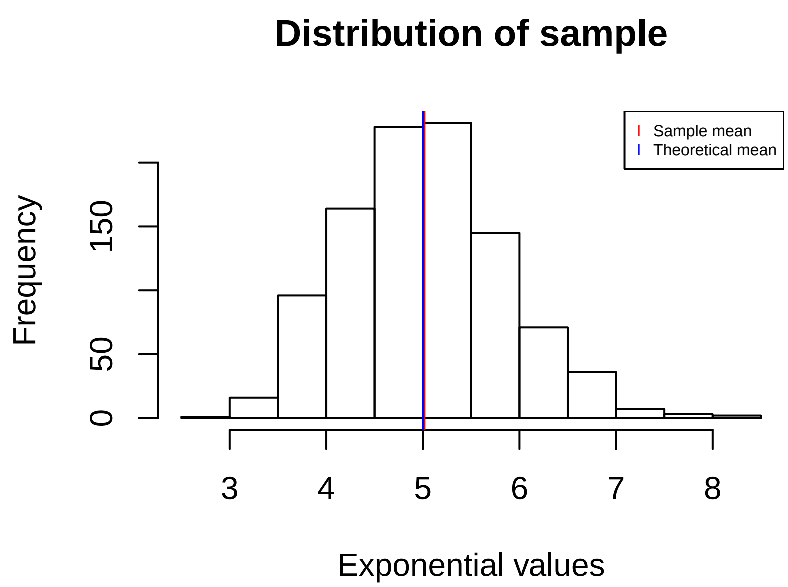

Figure 1.

2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

```
#Code 4: Sample variance:
n=40
sampleVar = sqrt(sum((sample-sampleMean)^2)/(n-1))
```

sampleVar = 4.1750379

```
#Code 5: Theoretical variance of the distribution
n = 40
lambda = 0.2
sigma = (1/lambda)
theoVar = sigma^2/sqrt(n)
```

theoVar = 3.9528471

```
#Code Figure 2: Comparative of variances
#Generate histogram
h <- hist(sample, main="Comparative of variance",
     xlab="Exponential values")
title(sub="Figure 2.")
#Calculate normal distribution for the sample
x <- seq(min(sample),max(sample),length=1000)
y <- dnorm(x,mean=sampleMean, sd=1)
y <- y*(h$counts/h$density)
#Print normal distribution
lines(x, y, col="red", lwd=1)
abline(v=theoMean, col="green")
```

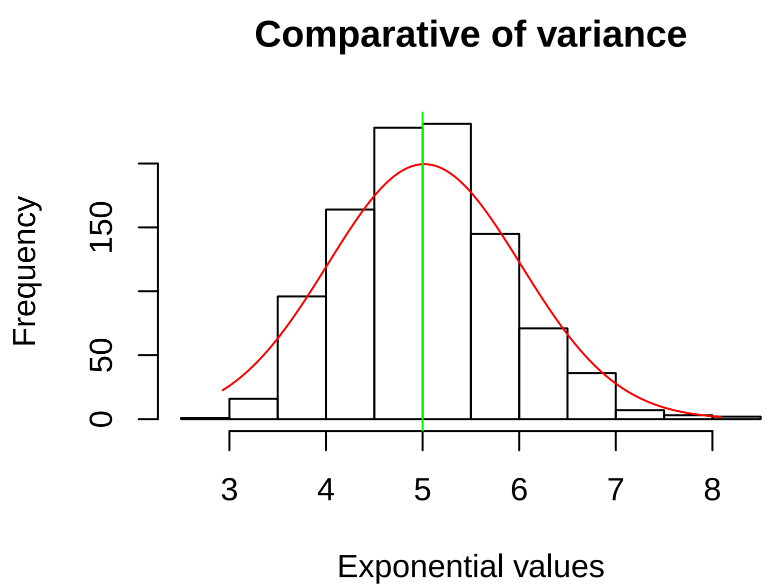

Figure 2.

3. Show that the distribution is approximately normal.

```
#Code Figure 3: Comparative qqplot between large collection of random exponentials
#Generate figure
qqnorm(sample, main = "Averages of exponentials")
qqline(sample, col="red")
```