Investigation of Expoenential distribution

Muralidharan S

22 August 2015

Overview

The exponential distribution is being discussed below. There is a R code embedded which talks about the exponential distribution and also talks about how the averages of 40 exponential distributions behave when simulated over a large number of times (in our case a thousand times). The behaviour should be in line with the Centaral Limit Theorem.

Central Limit Theorem (definition from wikipedia)

In probability theory, the central limit theorem (CLT) states that, given certain conditions, the arithmetic mean of a sufficiently large number of iterates of independent random variables, each with a well-defined expected value and well-defined variance, will be approximately normally distributed, regardless of the underlying distribution.

Simulations

The following R code gets the averages of 40 exponential distributions simulated a thousand times. The distribution of the averages should resemble a normal distribution as per the Central Limit Theorem.

```r
# R code for investigation of exponential distribution

# Set mean value to be null initially
mns=NULL

# Store the means of 40 exponentials in thousand simulations in the variable mns
for (i in 1 : 1000) mns = c(mns, mean(rexp(40,0.2)))

# This part answers the first question of part 1
# Get the mean of the sample
average<-mean(mns)

# Theoretical mean of the distribution
Theoretical_mean<-1/0.2

# The mean of the distribution of averages of 40 expoenentials simulated over
# a thousand times is very close to the theoretical mean. This is an important
# axiom from the Central Limit Theorem
```

```r
# This part answers the second question of part 1
# Get the variance of the distribution
Variance<-var(mns)

##Theoretical variance of the distributionn
Theoretical_Variance<-(1/0.2)^2

# As we are looking at the distribution of the averages simulated over a large
# number of times the variance of the sample greatly reduces as compared to the
# theoretical variance of the distribution.

##This part answers the third question of part 1

# From the histogram we observe that the distribution of averages appears
# like a bell shaped normal distribution with the mean of this distribution
# of averages very close to the theoretical mean of the actual distribution

Median<-median(mns)

# We observe that the median of the distribution is very close the mean of
# the distribution

# From the below 3 steps we see what percentage of the distribution lies within
# (mean + 1 standard deviation),(mean + 1 standard deviations),
# and (mean + 1 standard deviations). If this is a normal distribution 68% of the
# values will lie within (mean + 1 standard deviation), 95% within
# (mean + 2 standard deviations) and 99.7% within (mean + 3 standard deviations)
sum(mns<=(mean(mns)+sd(mns)) & mns>=(mean(mns)-sd(mns)))

## [1] 686

#686
sum(mns<=(mean(mns)+sd(mns)*2) & mns>=(mean(mns)-sd(mns)*2))

## [1] 951

#952
sum(mns<=(mean(mns)+sd(mns)*3) & mns>=(mean(mns)-sd(mns)*3))

## [1] 998

#997



# From the above tests we can safely say that the distribution of averages
# of 40 exponential distributions simulated a thousand times is a normal
# distribution.
```

Theoretical mean and sample mean

Here we will look at how the theoretical mean and the sample mean compare with each other. As observed below both the values are very close to each other.

```
cat("The theoretical mean of the distribution is", (1/0.2))

## The theoretical mean of the distribution is 5

cat("The sample mean of the averages distribution is", mean(mns))

## The sample mean of the averages distribution is 4.990905
```

Theoretical Variance and sample Variance

Here we will look at how the theoretical mean and the sample mean compare with each other.

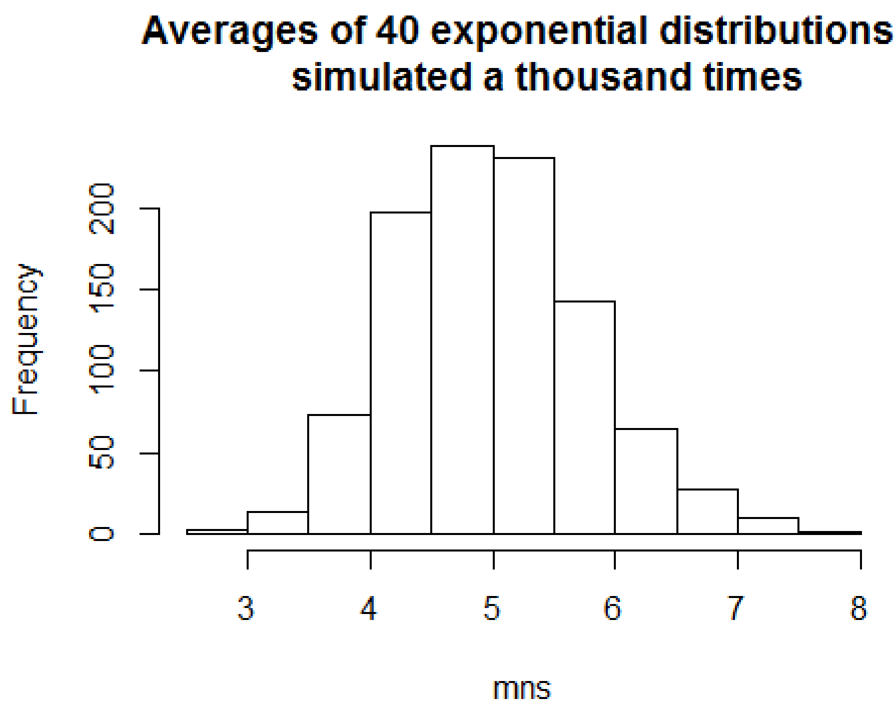
```
cat("The theoretical variance of the distribution is", ((1/0.2)^2))

## The theoretical variance of the distribution is 25

cat("The sample variance of the averages distribution is", var(mns))

## The sample variance of the averages distribution is 0.6007072
```

Distribution

The salient features of a normal distribution is as follows, * The distribution appears like a bell curve. * The mean and the median are equal. * 68% of the values lie between (mean + 1 std deviation) and (mean- 1 std deviation). Similarly 95% in 2 std deviations and 99.7% in 3 std deviations.

```
cat("The sample mean of the averages distribution is", mean(mns))

## The sample mean of the averages distribution is 4.990905

cat("The sample median of the averages distribution is", median(mns))

## The sample median of the averages distribution is 4.95658

cat("The % of values that lie within 1 standard deviation", sum(mns<=(mean(mns)+s
d(mns)) & mns>=(mean(mns)-sd(mns)))/10)

## The % of values that lie within 1 standard deviation 68.6

cat("The % of values that lie within 2 standard deviation", sum(mns<=(mean(mns)+s
d(mns)*2) & mns>=(mean(mns)-sd(mns)*2))/10)

## The % of values that lie within 2 standard deviation 95.1

cat("The % of values that lie within 3 standard deviation", sum(mns<=(mean(mns)+s
d(mns)*3) & mns>=(mean(mns)-sd(mns)*3))/10)

## The % of values that lie within 3 standard deviation 99.8
```

Appendix



Averages of 40 exponential distributions
simulated a thousand times
Frequency
0
50
100
150
200
3
4
5
6
7
8
mns