Compare Exponential Distribution and Central Limit Theorem

Ramana Sonti

2/4/2017

Overview:

In this report, the distribution of 40 random exponentials has been compared against the distribution of their means in an attempt to prove what Central Limit Theorem states. A set of 1000 means have been simulated and plotted to show that the distribution approximates to normal when a collection of a large enough sample set is considered in the study.

Part 1: Simulation Exercise

Given: In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem. 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 1/lambda. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

```
## Loading required package: pastecs

## Loading required package: boot

## Loading required package: R.utils

## Loading required package: R.oo

## Loading required package: R.methodsS3

## R.methodsS3 v1.7.1 (2016-02-15) successfully loaded. See ?R.methodsS3 for help.

## R.oo v1.20.0 (2016-02-17) successfully loaded. See ?R.oo for help.

##
## Attaching package: 'R.oo'

## The following objects are masked from 'package:methods':
##
##     getClasses, getMethods

## The following objects are masked from 'package:base':
##
##     attach, detach, gc, load, save

## R.utils v2.4.0 (2016-09-13) successfully loaded. See ?R.utils for help.

##
## Attaching package: 'R.utils'

## The following object is masked from 'package:pastecs':
##
##     extract
```

```
## The following object is masked from 'package:utils':
##
##     timestamp
```

```
## The following objects are masked from 'package:base':
##
##     cat, commandArgs, getOption, inherits, isOpen, parse, warnings
```

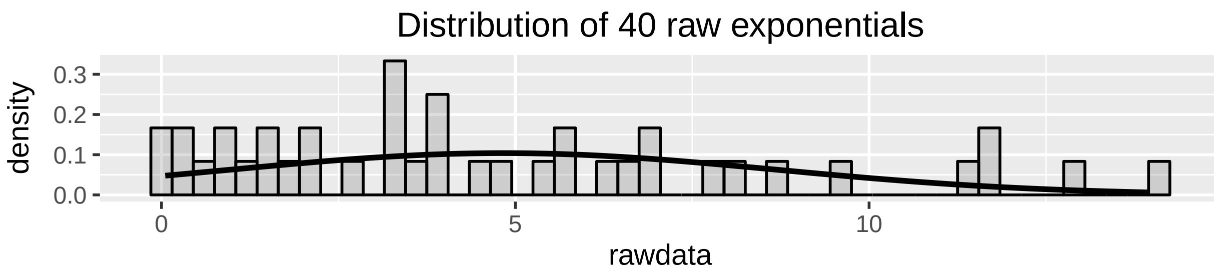
```
## Loading required package: dtplyr
```

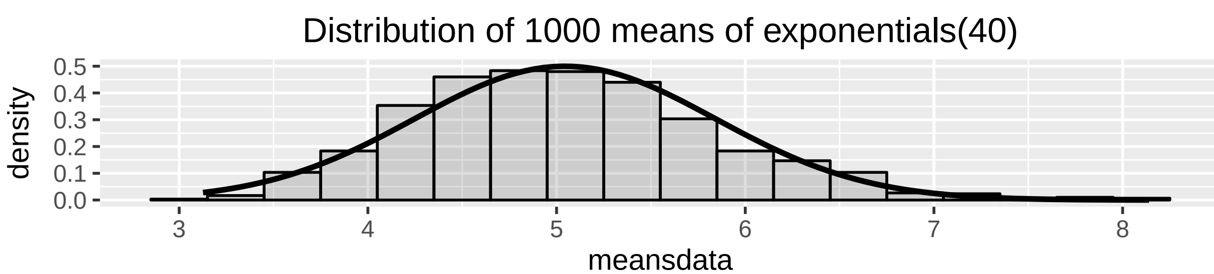
```
## Loading required package: ggplot2
```

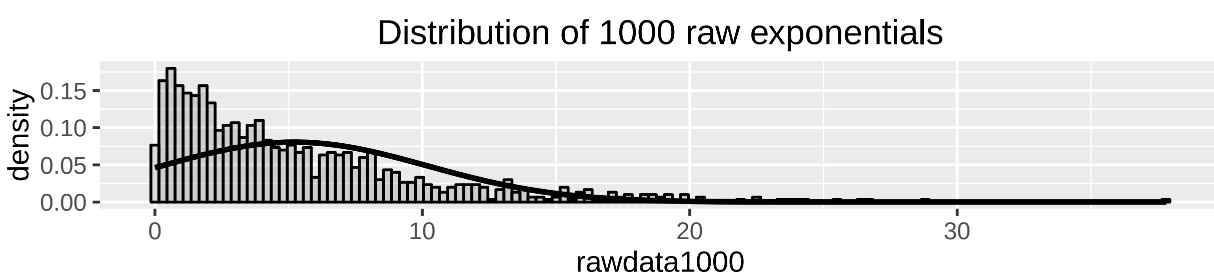
```
## Loading required package: grid
```

```
## [1] Stats on 1000 means :
```

```
##      nbr.val     nbr.null       nbr.na          min          max
## 1.000000e+03 0.000000e+00 0.000000e+00 3.126236e+00 8.140064e+00
##        range          sum       median         mean      SE.mean
## 5.013828e+00 5.045060e+03 4.987271e+00 5.045060e+00 2.524390e-02
## CI.mean.0.95          var      std.dev     coef.var
## 4.953715e-02 6.372544e-01 7.982821e-01 1.582305e-01
```








```
## [1] Mean of raw data                                   : 4.8326211489293
```

```
## [1] Sample mean (mean of 1000 means) from simulation    : 5.04505958753827
```

```
## [1] Theoretical mean (1/lambda)                        : 5
```

```
## [1] Variance of raw data                               : 14.7234142944959
```

```
## [1] Sample variance (var of 1000 means) from simulation : 0.637254390249431
```

```
## [1] Theoretical variance (1/(lambda^2)/n)              : 0.625
## [1] Standard Deviation of raw data                                : 3.8371101488615
## [1] Sample Standard deviation (sd of 1000 means) from simulation : 0.798282149524484
## [1] Theoretical standard error ((1/lambda)/sqrt(n))              : 0.790569415042095
## [1] Mean of 1000 raw exponentials     : 5.21984780345108
## [1] Standard Deviation of 1000 raw exponentials     : 4.94878350590842
## [1] Variance of 1000 raw exponentials     : 24.4904581883512
```

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

3 From the results and the plot above, it is evident that the simulated sample mean (5.04 : means of 1000 samples) is very close to theretical mean (5).

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

From the results above, the sample variance (0.637 : variance of 1000 menas) is very close to theoretical variance (0.625). It shoud get much closer to the theretical variance if we increase number of simulations.

3. Show that the distribution is approximately normal.

As shown in the third plot, the distribution is not normal even if we consider a large collection of exponentials. However, the sample means of the sample set (40) that is not so large produces the distribution that very close to normal distribution as shown in plot 2. The distribution is a classic "bell" shaped curve and is centered around population mean (~5) and is asymptotic.