Central Limit Theorem comparison for exponential distribution

yogarjun77

Thursday, May 21, 2015

Overview:

An exponential distribution can be simulated in R using function rexp(n, lambda) where n is the rate parameter. Theoretically, both the mean and standard deviation of an exponential distribution is 1/lambda. A sample of 1000 simulations of the averages of 40 exponentials are generated and in this experiment and compared with the theoretical values for mean and variance. The distribution is also compared with a normal distribution.

Preparation:
a. Load libraries and determine parameters to be used

```
library(ggplot2)
lambda = 0.2
n = 40
sim = 1000 #total simulations
```

b. Simulation of sample distribution

```
set.seed(333)
exp2 = NULL
for (i in 1 :sim){
exp2 = c(exp2, mean(rexp(n,lambda)))}
str(exp2)
```

```
##  num [1:1000] 5.15 5.07 5.63 5.62 3.49 ...
```

1 - Sample mean vs Theoretical Mean

Mean of sample set of 1000 averages of 40 exponentials

```
        mean_exp2 <- round(mean(exp2),3)
        print(mean_exp2)
```

```
## [1] 5.045
```

Theoretical mean(population mean)

```
      1/lambda
```

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

The sample mean (5.045) is almost equal to the theoretical value(5), fulfilling the first condition for Central Limit Theorem $E[\bar{X}] = \mu$.

2 - Sample Variance vs Theoretical Variance

Variance of sample set of 1000 averages of 40 exponentials:

```
        var_exp2 <- round(var(exp2),3)
        print(var_exp2)
```

```
## [1] 0.636
```

Theoretical variance (population variance) = σ^2/n

```
        var_exp2 <- (round((((1/lambda)^2)/n),3))
        print(var_exp2)
```

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

Again, the sample variance is almost equal to the theoretical value, fulfilling the second condition for CLT: $\mathrm{Var}[\bar{X}] = \sigma^2/\mathrm{n}$

3 - Proving that the distribution is approximately normal

a)Comparing with a distribution of 1000 exponentials.

Create a sample distribution of 1000 exponentials

```
set.seed(333)
exp_th <- rexp(1000, 0.2)
```

Mean & Variance of 1000 exponential

```
round(mean(exp_th),3)
```

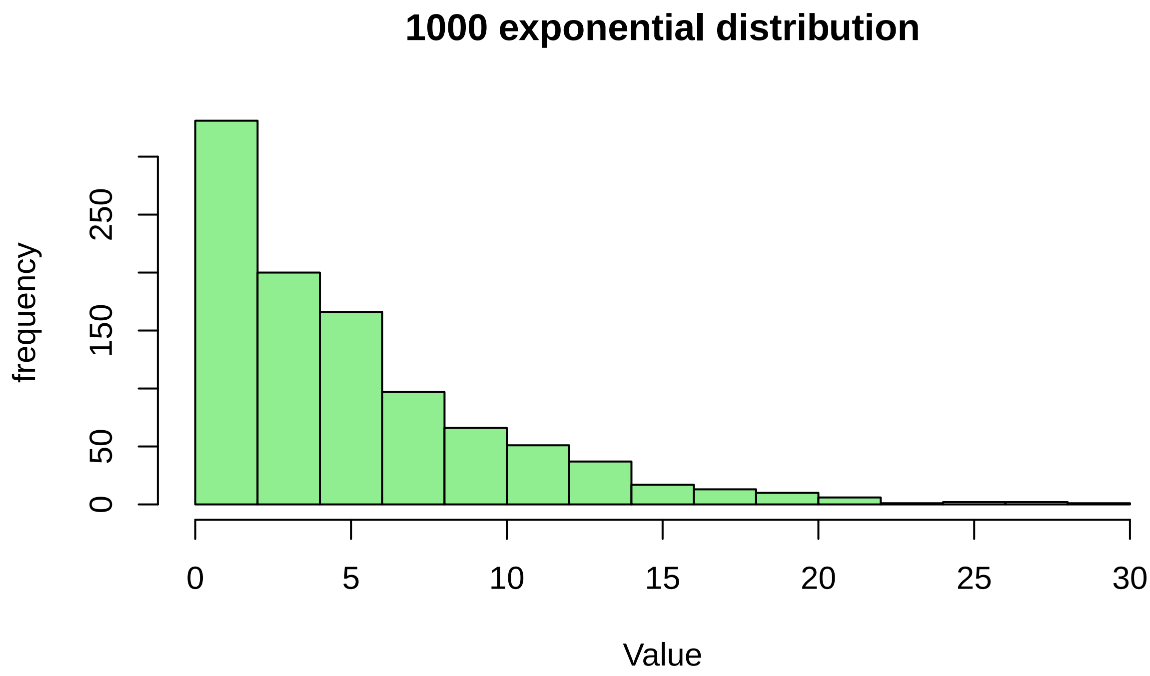
```
## [1] 4.971
```

```
round(var(exp_th)/1000, 3)
```

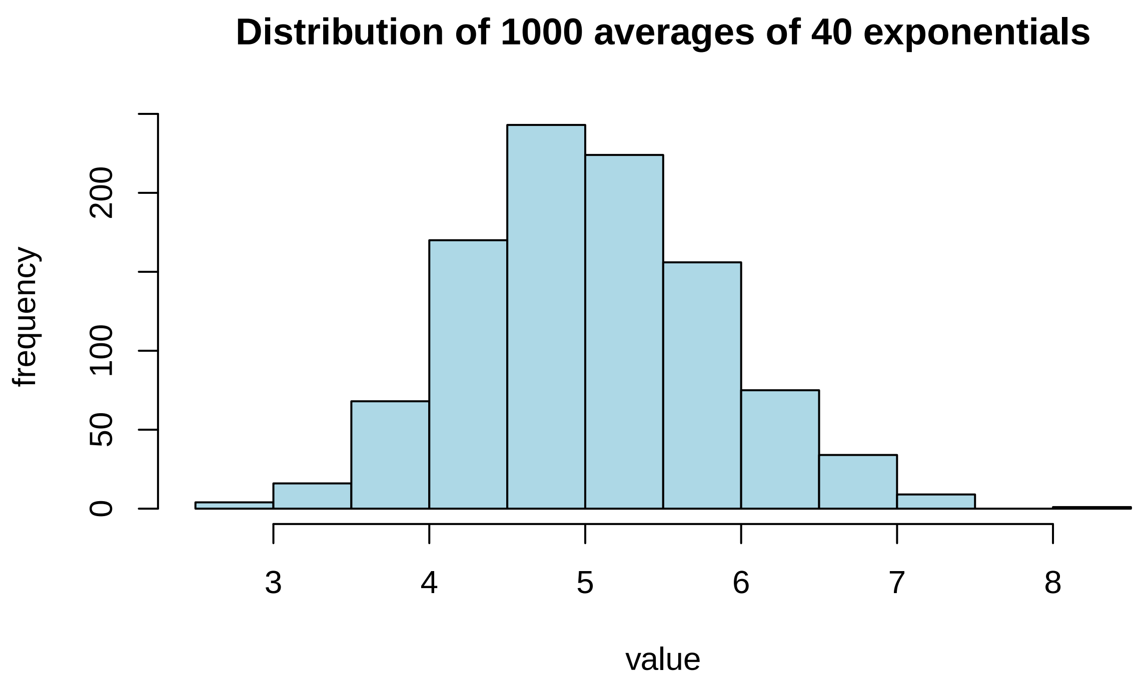
```
## [1] 0.022
```

```
#Plot histogram for both distributions for visual comparison

hist(exp_th, col = "lightgreen", xlab = "Value",
     ylab = "frequency", main = "1000 exponential distribution")
```



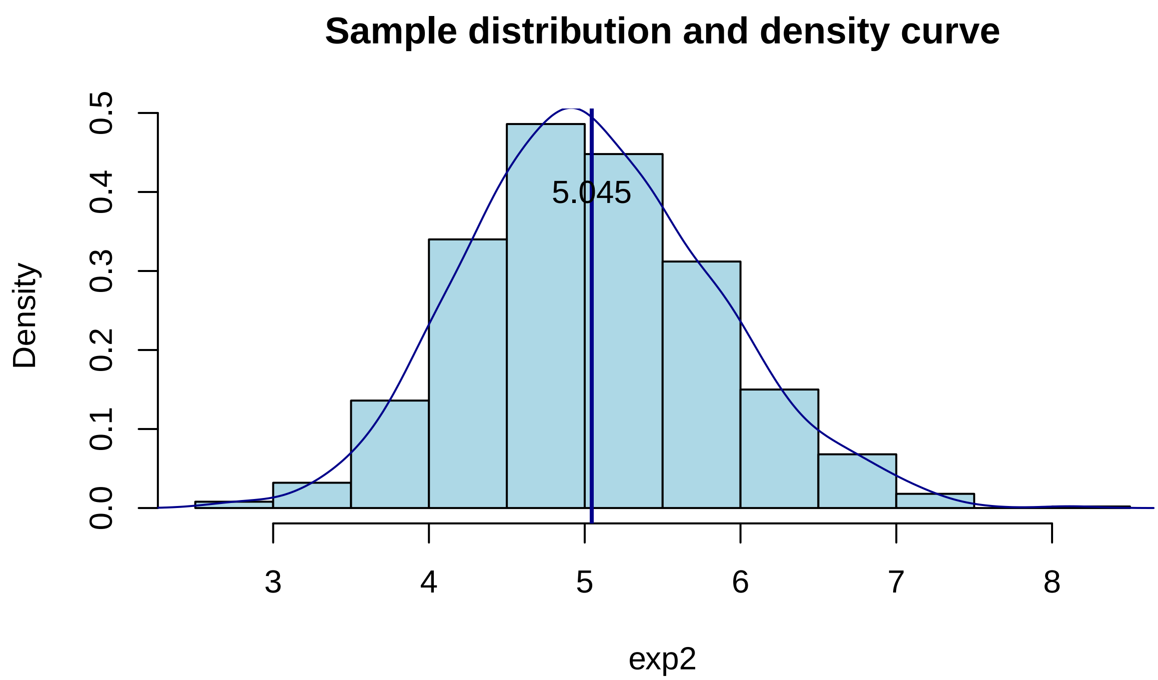

```r
hist(exp2, col = "lightblue", xlab = "value", ylab = "frequency",
     main = "Distribution of 1000 averages of 40 exponentials")
```



The histogram and variance values show that a distribution of 1000 exponentials are not similar to a normal distribution.

b) Histogram of 1000 averages of 40 exponentials mapped onto a normal distribution

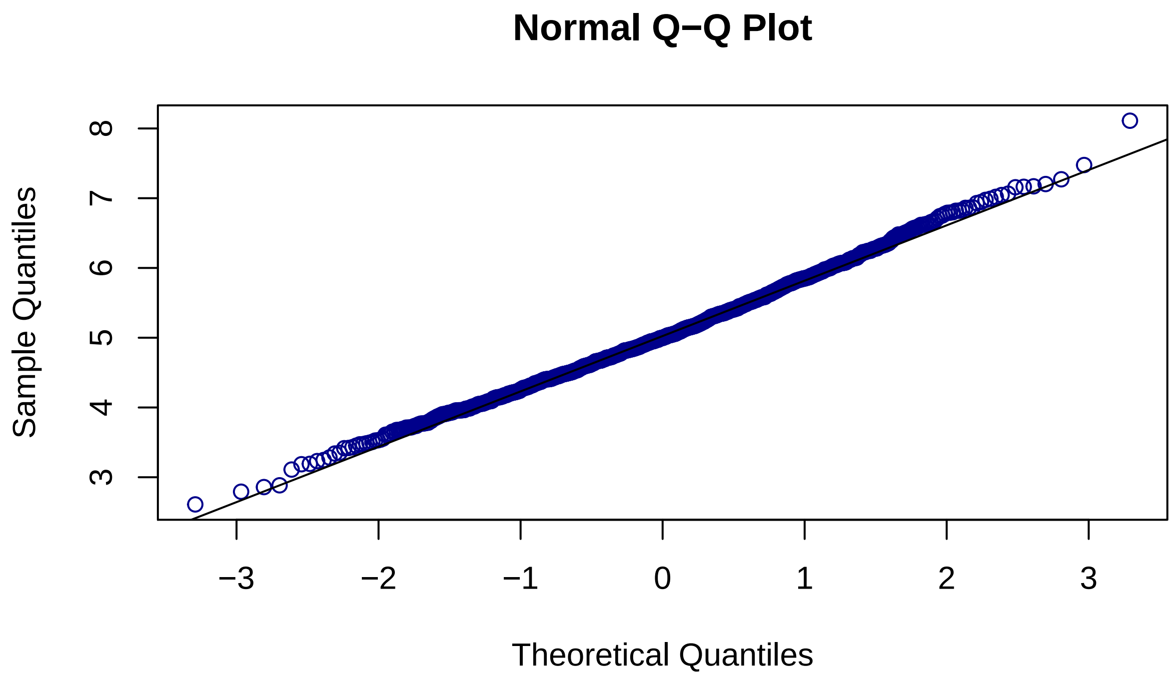
```
hist(exp2, col= "lightblue", breaks = "sturges", freq = FALSE,
     main = "Sample distribution and density curve")
abline (v=mean_exp2, col = "darkblue", lwd = 2)
text(mean_exp2,0.4, mean_exp2)
lines(density(exp2), col="darkblue")
```



The distribution exhibits a Gaussian shape almost similar to a normal distribution.

c) Normal Q-Q plot

```
qqnorm(exp2, col = "darkblue")
qqline(exp2, col = "black")
```



A Q-Q (quantile-quantile) plot is more effective to test if a sample is normally distributed. The resulting plots show that the normalized exp_2 (circles) closely fit the theoretical default normal (line).

Conclusion

The sample fulfilled 2 key conditions for the Central Limit Theorem Criteria

a) E[$\bar{X}$]=µ

b) Var[$\bar{X}$]=σ^2/n

The histogram for 1000 averages of 40 exponentials(the sample) and its Q-Q plot positively indicate that the sample exhibits characteristics of a normal distribution according to the Central Limit Theorem.

Additionally just taking 1000 exponentials with the same random seed do not yield a similar distribution.

All these points show proof that the sample is highly similar to a normal distribution based on the Central Limit Theorem.