Statistical Inference Course Project Part 1 - Exponential Distribution Simulation Statistics - (Johns Hopkins Data Science Certification)

MD Alamgir, Ph.D.

October 22, 2016

Overview

The exponential distribution is the probability distribution that describes the time between events in a Poisson process, i.e. a process in which events occur continuously and independently at a constant average rate. Per project instructions, here we investigate the exponential distribution in R and compare it with the Central Limit Theorem.The central limit theorem states that the sampling distribution of the mean of any independent, random variable will be normal or nearly normal, if the sample size is large enough.

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.We investigate the distribution of averages of 40 exponentials. One thousand simulations are performed. Appendix at the end of the report shows “Approach” uses and relevant plots.

Assumptions

- The rate lambda is constant, a requirement of exponential distribution, and specied here by the value 0.2.
- Theoretical standard deviation is estimated as s/sqrt(n), where s is the standard deviation for exponential distribution, and n is the sample size.

Set working directory

```
setwd("C:/Users/MD/Documents/Week4_Statistical_Inference")
```

- Load libraries needed for data processing and analysis

```
library("dplyr", verbose=FALSE)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Simulations

- Rstudio, and knitr tools used. Set parameter values per problem instructions(rate=lambda=0.2, nexp= number of exponentials=40, nsim=number of simulations=1000, Simulate using the R function rexp(n, lambda)

```
set.seed(9999) #set a seed value to be able to reproduce randomiztions
lambda=0.2 # Set value for rate lambda
nexp=40 # Set value for n
nsim=1000 # Set value for number of simulations

# Use replicate function in R with nsim and rexp(nexp, lambda) as arguments to run the simulatio
ns
expsim <- replicate(nsim, rexp(nexp, lambda))
msim <-apply(expsim, 2, mean) # Evaluate the "means" of the simulations.
```

Comparison of sample mean and Theoretical Mean

The theoretical mean of the exponential distribution is 1/lambda. Given lambda=0.2, the theoretical mean is 5.0 A scatter plot of the mean of the 1000 simulations are shown in the Appendix The calculated mean is seen to be very close to, and within 0.42% of the theoretical mean 5.

```
# Scatter plot of the means for the 1000 simulations

TMean<-1/lambda # Theoretical mean of the exponential distribution
round(TMean, 2)
```

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

```
AMean<-mean(msim)# Actual Mean of sample means
round(AMean, 2)
```

```
## [1] 4.98
```

```
DiffMeanPct<-100*(AMean-TMean)/TMean
round(DiffMeanPct, 2)
```

```
## [1] -0.42
```

Sample Variance vs.Theoretical Variance