Github repo for the Course: Statistical Inference Github repo for Rest of Specialization: Data Science Coursera

### Instructions

1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

### Loading Libraries

`{r DataLoading} library("data.table") library("ggplot2")`

### Task

"'{r Stuff} # set seed for reproducability set.seed(31)

# set lambda to 0.2

lambda <- 0.2

# 40 samples

n <- 40

# 1000 simulations

simulations <- 1000

# simulate

simulated_exponentials <- replicate(simulations, rexp(n, lambda))

# calculate mean of exponentials

means_exponentials <- apply(simulated_exponentials, 2, mean) “‘

## Question 1

Show where the distribution is centered at and compare it to the theoretical center of the distribution.

`{r} analytical_mean <- mean(means_exponentials) analytical_mean`

`{r} # analytical mean theory_mean <- 1/lambda theory_mean`

`{r} # visualization hist(means_exponentials, xlab = "mean", main = "Exponential Function Simulations") abline(v = analytical_mean, col = "red") abline(v = theory_mean, col = "orange")`

The analytics mean is 4.993867 the theoretical mean 5. The center of distribution of averages of 40 exponentials is very close to the theoretical center of the distribution.

## Question 2

Show how variable it is and compare it to the theoretical variance of the distribution..

`{r} # standard deviation of distribution standard_deviation_dist <- sd(means_exponentials) standard_deviation_dist`

`{r} # standard deviation from analytical expression standard_deviation_theory <- (1/lambda)/sqrt(n) standard_deviation_theory`

`{r} # variance of distribution variance_dist <- standard_deviation_dist^2 variance_dist`

`{r} # variance from analytical expression variance_theory <- ((1/lambda)*(1/sqrt(n)))^2 variance_theory`

Standard Deviation of the distribution is 0.7931608 with the theoretical SD calculated as 0.7905694. The Theoretical variance is calculated as ((1 / ??) * (1/???n))2 = 0.625. The actual variance of the distribution is 0.6291041

## Question 3

Show that the distribution is approximately normal.

`{r} xfit <- seq(min(means_exponentials), max(means_exponentials), length=100) yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(n)))` hist(means_exponentials,breaks=n,prob=T,col="orange",xlab = "means",main="Density of means",ylab="density") lines(xfit, yfit, pch=22, col="black", lty=5)

{r} # compare the distribution of averages of 40 exponentials to a normal distribution qqnorm(means_exponentials) qqline(means_exponentials, col = 2)

Due to Due to the central limit theorem (CLT), the distribution of averages of 40 exponentials is very close to a normal distribution.