Statistical Inference Course Project Part 1 - Kevin O'Leary

Part 1 of the Statistical Inference course project investigates the Central Limit Theorem (CLT), specifically whether the arithmetic mean of a sufficiently large number of iterates of exponential distributions will be approximately normally distributed.

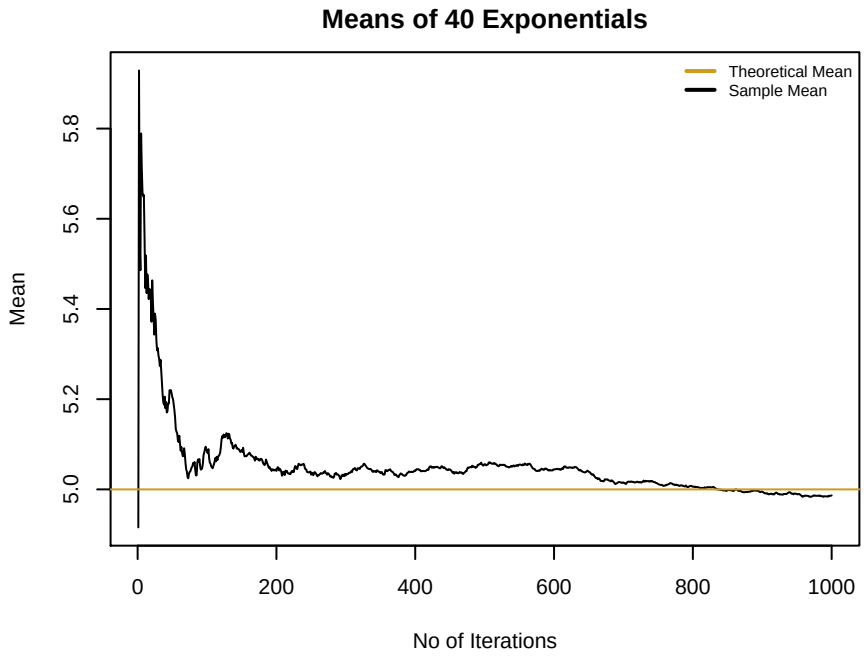
1. Show the sample mean and compare it to the theoretical mean of the distribution.

First we need to find the mean of 40 exponential distributions and iterate 1000 times. Then we can simply take the mean of this vector and compare against the theoretical mean of 1/lambda. Here we find close agreement of **4.986508** and **5**, respectively.

```
set.seed(42) ##specify random seed state
lambda=.2 ##set lambda value
expno=40 ##set number of exponentials
simno=1000 ##set number of simulations

exdis = rep(NA,simno) ##create an empty vector with 1000 elements

for (i in 1:simno){ ##for every element in exdis...
  exdis[i] = mean(rexp(expno,lambda)) ##populate with the mean of 40 exponentials
}

samean=mean(exdis) ##assign the mean of exdis to samean
theomean=1/lambda ##assign the theoretical mean to theomean
```

To appreciate this graphically, we first need to calculate, then plot, the moving average, or cumulative mean of our exponentials. The CLT suggests that this cumulative mean should converge on the theoretical mean as the iterations increase, which is in agreement with the graphic below.

```
cumean = cumsum(exdis)/seq_along(exdis) ##mean of each cumulative sum

par(mar=c(3,11,1,1)+1.5, cex=0.7) ##set margins and font size

plot(seq_along(exdis), cumean, type="l",
     main="Means of 40 Exponentials",xlab="No of Iterations",
     ylab="Mean")
       abline(h=theomean, col="goldenrod3") ##theoretical mean line
     legend("topright", legend=c("Theoretical Mean","Sample Mean"),
       col=c("goldenrod3","black"), lwd=c(2,2), cex=.7, bty="n")
```

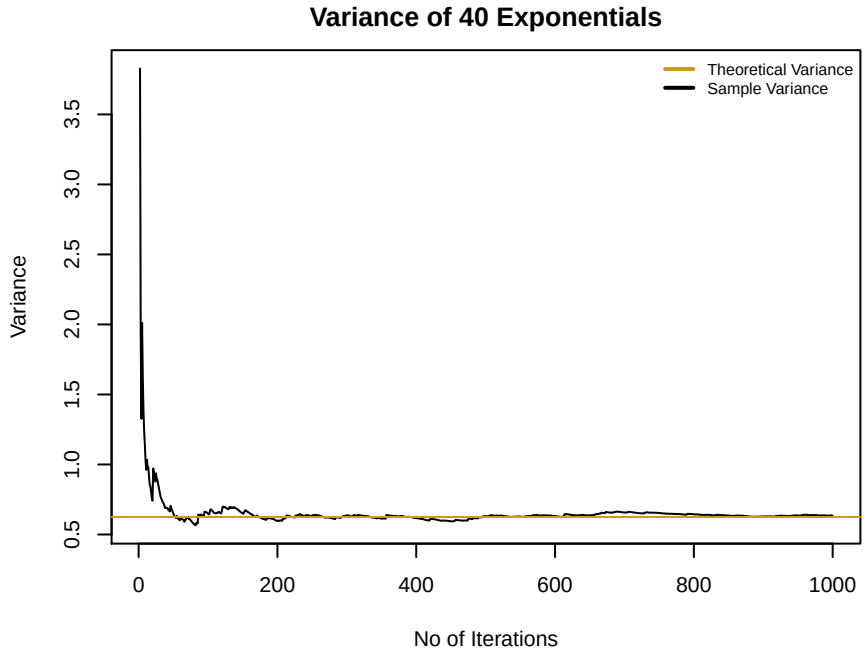


2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

Similar to above, we can take the variance of our vector and compare against the theoretical variance of 40 distributions. In this case we find that **0.6344405** and **0.625** match closely.

Next we need to calculate the cumulative variance to plot the change in variance with iteration. As was seen above, the greater the number of iterations, the closer the sample variance gets to the theoretical variance.

```
samvar = var(exdis) ##assign sample variance to samvar
theovar = 1/((lambda^2)*expno) ##theoretical variance for 40 distributions
cumvar = cumsum((exdis-samean)^2)/(seq_along(exdis)-1) ##cumulative variance

par(mar=c(3,11,1,1)+1.5, cex=0.7) ##set margins and font size

plot(seq_along(exdis),cumvar,type="l",
     main="Variance of 40 Exponentials",
     xlab="No of Iterations", ylab="Variance")
abline(h=theovar, col="goldenrod3") ##theoretical variance line
legend("topright", legend=c("Theoretical Variance","Sample Variance"),
       col=c("goldenrod3","black"), lwd=c(2,2),cex=.7,bty="n")
```



3. Show that the distribution is approximately normal.

Below we have superimposed a normal distribution curve with the theoretical mean and standard deviation over a histogram of our data. We can see that both are in general agreement with the data, meaning that the data is approximately normal.

```
par(mar=c(3,11,1,1)+1.5, cex=0.7) ##set margins and font size

hist(exdis,breaks=30, freq=FALSE, col="slategray2",
     main="Comparison of Sample Exponentials to Normal Distribution", xlab="Value",ylab="Density")
curve(dnorm(x, mean=theomean, sd=sqrt(theovar)),
      add=TRUE,lwd=2, col="goldenrod3") ##normal distribution curve
abline(v=theomean,lwd=2,col="violetred3") ##theoretical mean line
legend("topright", legend=c("Theoretical Mean","Theoretical Distribution"),
       col=c("violetred3","goldenrod3"), lty=c(1,1),lwd=c(2,2),cex=.8,bty="n")
```