Statistical Inference Project Part 1

Overview

This project looks at the premise of the Central Limit Theorem (CLT) and compares it to the exponential distribution. The CLT states that under certain conditions, with a large enough sample size, the means of these samples with be roughly normally distributed. Hence, if we generate an exponential distribution, with a mean of λ and a standard deviation also of λ, and we plot the means, the plot will be roughly normal.

Simulation

We ran 10000 simulations, with a sample size of 40 and $\lambda = 0.2$.

```
sim <- replicate(10000, rexp(40, 0.2))
```

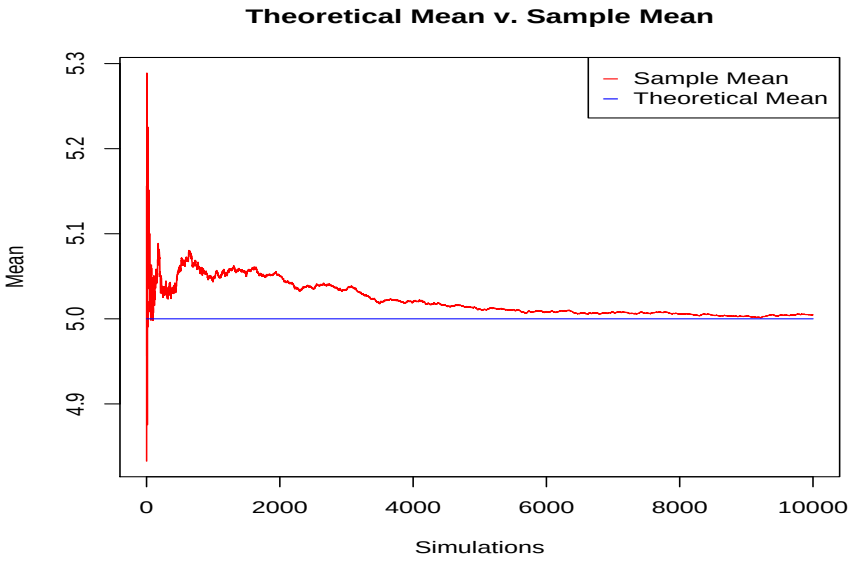
Comparing Theoretical and Sample Means

The following code compares the theoretical mean, $1/\lambda$, which is 5, with the cumulative means of the simulation.

```
theoretical_mean <- 1 / lambda
converging_means <- cumsum(col_means) / (1:sim_num)

plot(x = 1:sim_num, y = converging_means, type = "l",
     main = "Theoretical Mean v. Sample Mean",
     xlab = "Simulations", ylab = "Mean", col = "red")
lines(x = c(1, sim_num), y = c(theoretical_mean, theoretical_mean),
      type = "l", col = "blue")
legend("topright", pch = "_", col = c("red", "blue"),
       legend = c("Sample Mean", "Theoretical Mean"))
```

The graph is:



suggesting that the CLT applies.

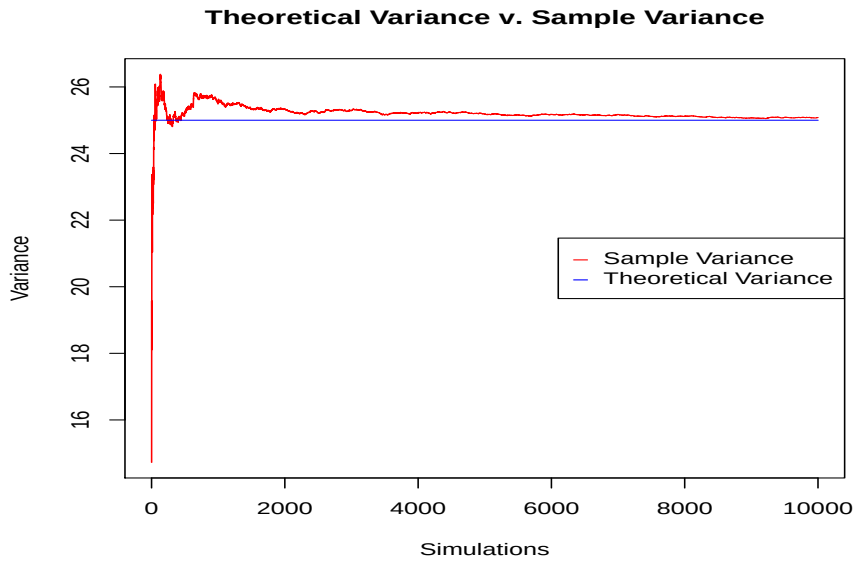
Comparing Theoretical and Sample Variances

A similar process can be followed for comparing the variances. The code to compare the theoretical variance with the cumulative variances of the simulation is:

```
theoretical_var <- 1 / lambda^2
sim <- replicate(sim_num, var(rexp(sam_size,lambda)))
converging_var <- cumsum(sim) / (1:sim_num)

plot(x = 1:sim_num, y = converging_var, type = "l",
     main = "Theoretical Variance v. Sample Variance",
     xlab = "Simulations", ylab = "Variance", col = "red")
lines(x = c(1, sim_num), y = c(theoretical_var, theoretical_var),
     type = "l", col = "blue")
legend("right", pch = "_", col = c("red", "blue"),
       legend = c("Sample Variance", "Theoretical Variance"))
```

giving the graph:



Again, indicating that the CLT applies.

Comparing Theoretical and Sample Distributions

To compare the distributions, we plot the normal curve over the plot of the exponential distribution samples, using this code:

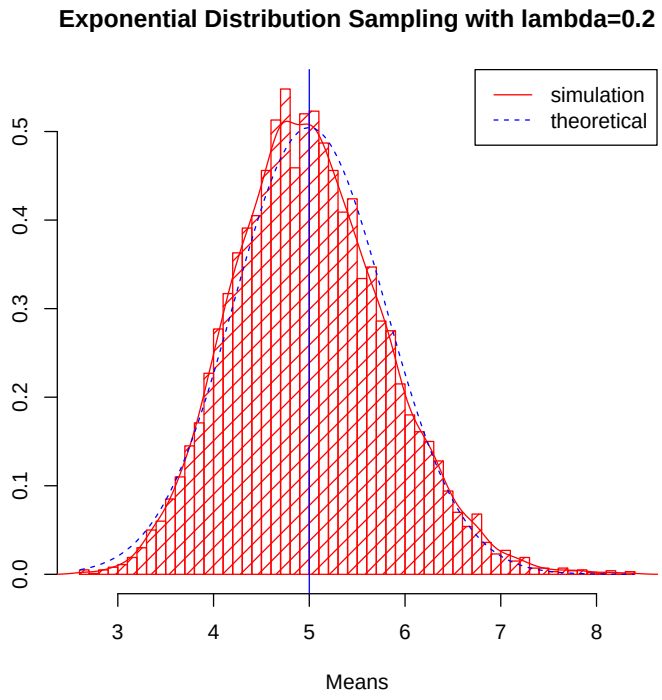
```
# plot the histogram of the averages calculated above
hist(col_means, breaks=60, freq=FALSE,
     main="Exponential Distribution Sampling with lambda=0.2",
     ylab="", xlab="Means", density=10, col="red")
```

```
# density of the averages of samples
lines(density(col_means), col="red")

# theoretical mean of distribution
abline(v=1/lambda, col="blue")

# theoretical density of the averages of samples
xfit <- seq(min(col_means), max(col_means), length=100)
yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(sam_size)))
lines(xfit, yfit, pch=22, col="blue", lty=2)
legend('topright', c("simulation", "theoretical"), lty=c(1,2), col=c("red", "blue"))
```

giving the graph:



Conclusion

This small example in no way contradicts the CLT. It shows means, variances and distributions of the exponential distribution tending to follow a normal curve as we take more and more samples.