StatInf Project 1: Simulation

Christopher Jones

November 21, 2017

Overview

In this report we compare the exponential distribution with the Central Limit Theorem (CLT) via simulation. We compare sample vs. theoretical mean, sample vs.theoretical variance, and overall goodness of fit. The exponential distribution models interarrival times for a Poisson process, with pdf given by $f(x) = \lambda e^{-\lambda x}$. Integration by parts once/twice gives the mean and variance as $\mu = 1/\lambda$ and $\sigma^2 = 1/\lambda^2$.

Simulations

This is a run of 1000 simulations of taking the mean from 40 samples of the exponential distribution.

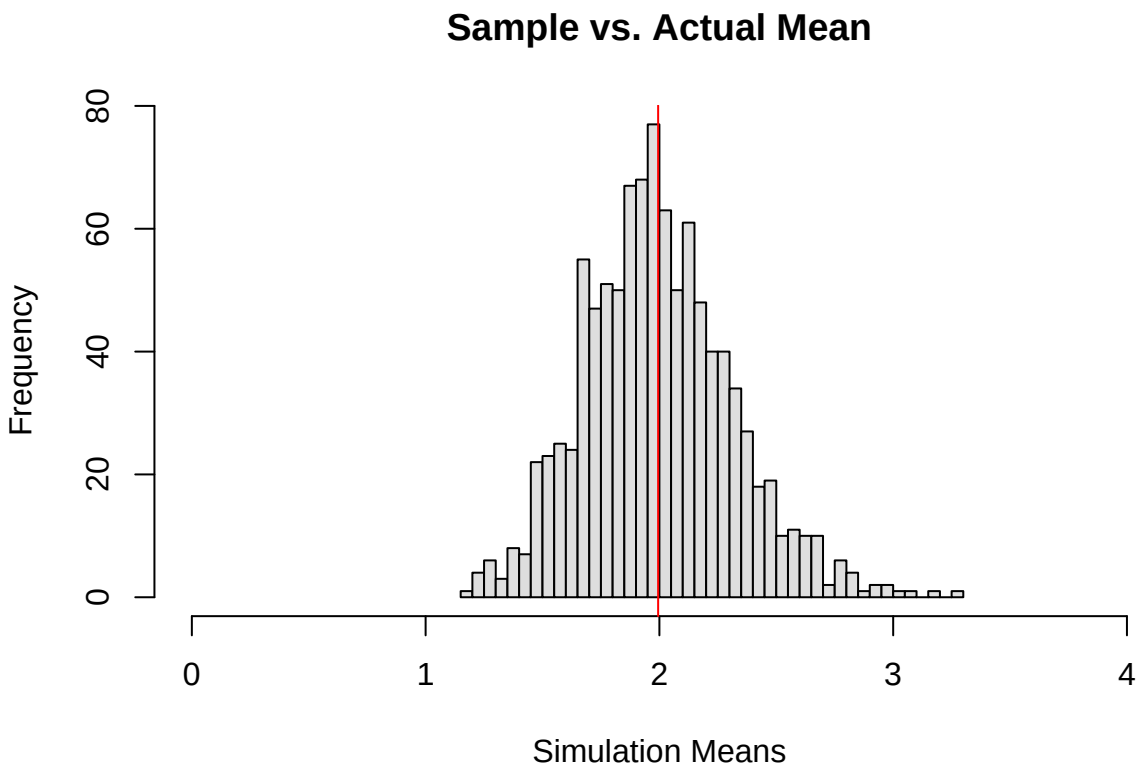
```
# assign simulation parameters
set.seed(42)
lam = .5
n <- 40
simulations <- 1000

# run
exp_means <- NULL
for (i in 1:simulations) {
  exp_means <- c(exp_means, mean(rexp(n, lam)))
}
```

Sample vs. Theoretical Means

Below is a histogram of the simulation means, along their overall mean. For our parameter $\lambda = .5$, we have the theoretical mean $\mu = 1/\lambda = 2$.

```
# Plot the results of the experiment
hist(exp_means
     , col="gray87"
     , main="Sample vs. Actual Mean"
     , xlim = c(0, 4)
     , breaks = 40
     , xlab = "Simulation Means")
abline(v = mean(exp_means), lwd = "1", col = "red")
```



The observed sample mean 1.9946033 agrees very well with the theoretical value of 2.

Sample vs Theoretical Variance

The sample variance of the exponential distribution is $\frac{1}{n\lambda^2} = \frac{1}{40 \times .5^2} = .1$

The observed variance is:

```
var(exp_means)
```

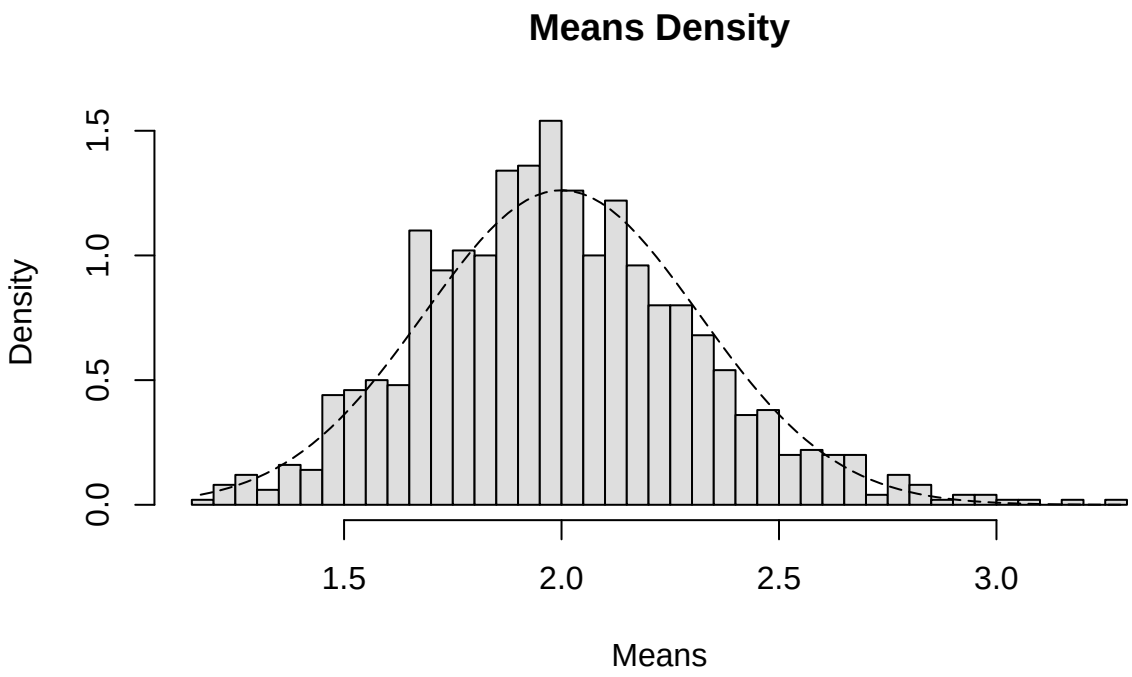
```
## [1] 0.1015105
```

So again we have very good agreement between the theoretical and observed values.

Distribution Comparison

We illustrate the conclusion of CLT visually that the observed exponential sampled means are close to normal via 2 visualizations:

```
# Sampled means vs theoretical normal distribution visualized
x_points <- seq(min(exp_means), max(exp_means), length=100)
y_points <- dnorm(x_points, mean=1/lam, sd=(1/lam/sqrt(n)))
hist(exp_means, breaks = n,prob = TRUE,col = "gray87",xlab = "Means",main = "Means Density", ylab = "Den
lines(x_points, y_points, pch=24, col="black", lty=5)
```




```
# distribution quantiles compared - sampled exponential means vs normal
qqnorm(exp_means); qqline(exp_means, col = 3)
```