Simulation and Inferential Analysis of the Exponential Distribution

Steve Myles

July 22, 2015

Contents

Overview **1**

Simulations **1**

Sample Mean vs. Theoretical Mean **2**

Sample Variance vs. Theoretical Variance **2**

Distribution **3**

Appendix **4**

1: Simulations 4

2: Histogram of simulation results 4

3: Histogram of simulation results with normal distribution 4

Overview

This report is an investigation of the exponential distribution in R. It involves a comparison with the Central Limit Theorem. Lambda (λ) is equal to 0.2 and 1000 simulations of the distribution of means of 40 exponentials are performed.

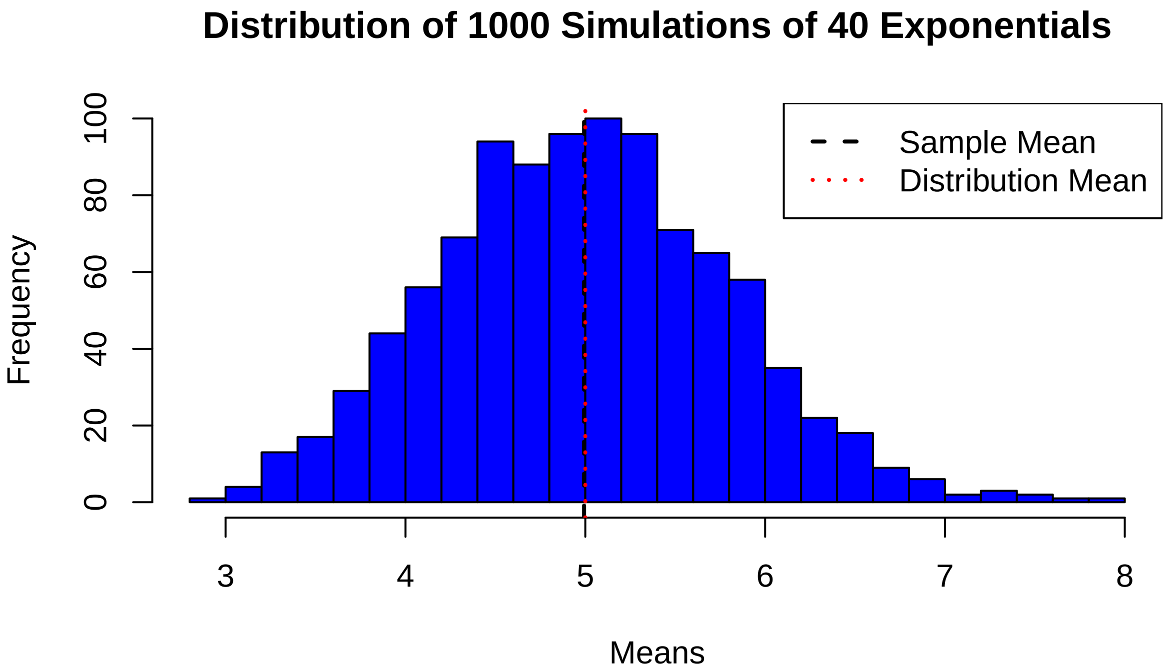
Simulations

As given in the assignment, the means of 1000 simulations of 40 samples from the exponential distribution with $\lambda = 0.2$ are taken. The theoretical mean of the distribution is calculated as 5 while its standard deviation is 5 and its variance is 25.

The sample mean (i.e., the mean of the 1000 simulations) is 4.9931694 and the theoretical variance and standard deviation of a normal distribution with mean = $1/\lambda$ and n = 40 are 0.625 and 0.7905694, respectively.

Sample Mean vs. Theoretical Mean

As can be seen from the following plot of the simulation results (code is available in the Appendix), the sample mean (dashed black line) is very close to the distribution mean (dotted red line):



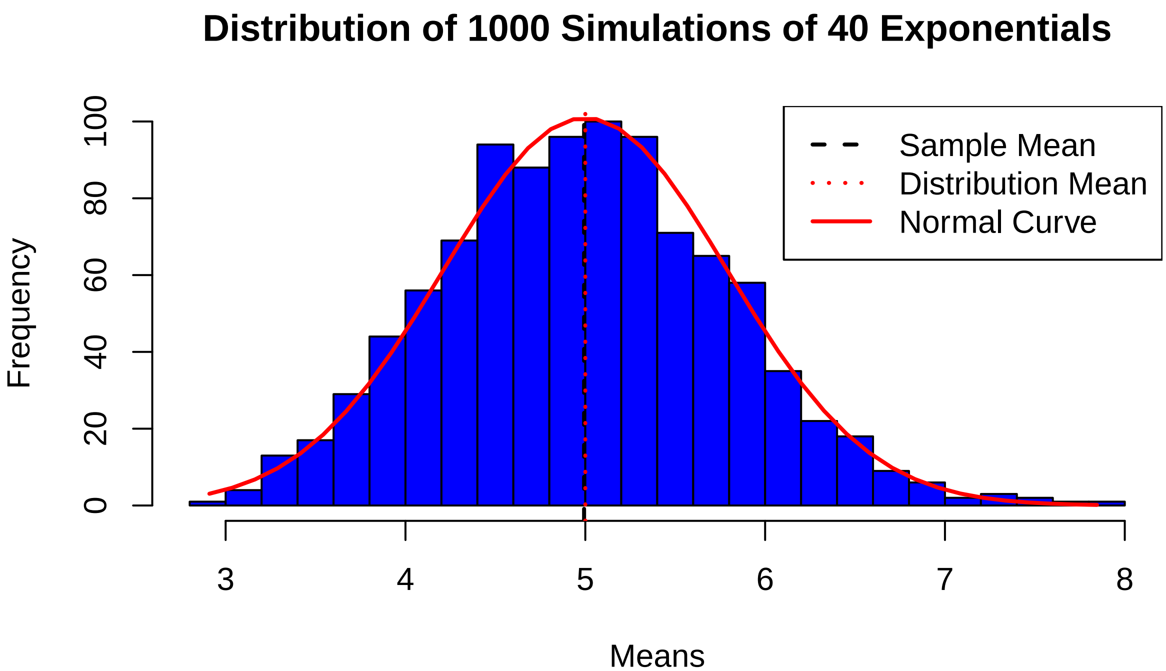
The theoretical mean of the distribution is 5 and the sample mean is 4.9931694. (These were calculated in the R code chunks above and are displayed in the preceding sentence using R Markdown.) Obviously, these are very close, which makes sense given the Central Limit Theorem.

Sample Variance vs. Theoretical Variance

The theoretical variance of the distribution is 0.625 (the variance of the exponential distribution with $\lambda = 0.2$ divided by the sample size (`samples`)) and the sample variance is 0.6299566. (These were calculated in the R code chunks above and are displayed in the preceding sentence using R Markdown.) As in the case of the sample and theoretical means, these are very close.

Distribution

Due to the central limit theorem, the distribution of these exponential means is approximately normal. This can be determined visually by recreating the above histogram and overlaying a normal curve with the appropriate mean (5) and standard deviation (0.7905694):



Given the similarity in mean and variance (and thus standard deviation), as well as the visual inspection provided above (the distribution is bell-shaped with a single peak and is approximately symmetrical about its mean), the $N(\mu = 5, \sigma^2 = 0.625)$ normal distribution (mean = 5 and variance 0.625 or standard deviation = 0.7905694) is a good approximation for the means of 1000 simulations of 40 exponentials with $\lambda = 0.2$. If either the number of simulations or the number of exponentials sampled were increased, this normal distribution would be an even better approximation.

Appendix

1: Simulations

```
## the distribution mean and standard deviation are 1/lambda for the
## exponential distribution.  the variance is the square of the std. deviation
lambda <- 0.2 ## exponential distribution parameter
samples <- 40 ## number of samples per simulation
nosim <- 1000 ## number of simulations
distMean <- distSd <- 1/lambda ## exponential distribution mean and std. dev.
## additionally, distMean is the mean of the associated normal distribution
distVar <- distSd^2 ## exponential distribution variance
ndistVar <- distVar/samples ## associated normal distribution variance
ndistSd <- sqrt(ndistVar) ## associated normal distribution standard deviation

## set mns (means) equal to NULL and conduct the simulations, then calculate
## the sample mean, standard deviation, and variance
mns <- NULL
for (i in 1:nosim) mns = c(mns, mean(rexp(samples, lambda)))
sampMean <- mean(mns)
sampSd <- sd(mns)
sampVar <- var(mns)
```

2: Histogram of simulation results

```
## set the title of the histogram
title <- paste("Distribution of", nosim, "Simulations of", samples,
           "Exponentials")

## create the histogram, addlines for the sample and distribution means and a
## legend
hist(mns, main = title, xlab = "Means", col = "blue", breaks = 20)
abline(v = sampMean, col = "black", lwd = 2, lty = 2)
abline(v = distMean, col = "red", lwd = 2, lty = 3)
legend("topright", c("Sample Mean", "Distribution Mean"), lwd = 2,
       col = c("black", "red"), lty = c(2,3))
```

3: Histogram of simulation results with normal distribution

```
## re-plot the histogram, again adding lines for the sample and dist.means
h <- hist(mns, main = title, xlab = "Means",
          col = "blue", breaks = 20)
abline(v = sampMean, col = "black", lwd = 2, lty = 2)
abline(v = distMean, col = "red", lwd = 2, lty = 3)
legend("topright", c("Sample Mean", "Distribution Mean", "Normal Curve"),
       lwd = 2, col = c("black", "red", "red"), lty = c(2, 3, 1))

## calculate the normal curve using the dnorm and diff functions then add it to
```

```r
## the plot using lines
xfit <- seq(min(mns), max(mns), length = 40)
yfit <- dnorm(xfit, mean = distMean, sd = ndistSd)
yfit <- yfit * diff(h$mids[1:2]) * length(mns)
lines(xfit, yfit, col = "red", lwd = 2)
```

(The normal curve was added to this plot based on Robert I. Kabacoff's example, found here.)