

Use of Central Limit Theorem in Simulating Exponential Random Variables

Data Analysis Series: Statistical Inference

Courtney D. Shelley

August 21, 2014

Introduction

The exponential distribution describes the arrival time of a randomly recurring independent event sequence. If μ is the mean waiting time for the next event recurrence, its probability density function is:

$$P(x) = D'(x) = \lambda * e^{-\lambda x} \mathbb{I}_{[0, \infty)}.$$

The exponential distribution can be simulated in R with `rexp(n, lambda)` where `n` is the number of `lambda` is the rate parameter. The mean of the exponential distribution is $1/\lambda$ and the standard deviation is also $1/\lambda$.

Analysis

The purpose of this exercise is to illustrate via simulation the distribution of averages of 40 exponential ($\lambda = 0.2$) simulations. Approximately 1,000 simulated averages will be conducted.

```
#### 1000 Simulations of the average of 40 exp(0.2) variables.
means<- replicate(1000, mean(rexp(40, 0.2)))
```

Goal 1: Demonstrate the center of the distribution and compare to the theoretical center of the distribution.

```
meanOfMeans <- mean(means)
expMean <- 1/0.2
```

The simulated data mean is 5.0162, which is very similar to the theoretical mean of $1/\lambda$ or 5.

Goal 2: 2. Demonstrate simulated distribution variation and compare to theoretical distribution.

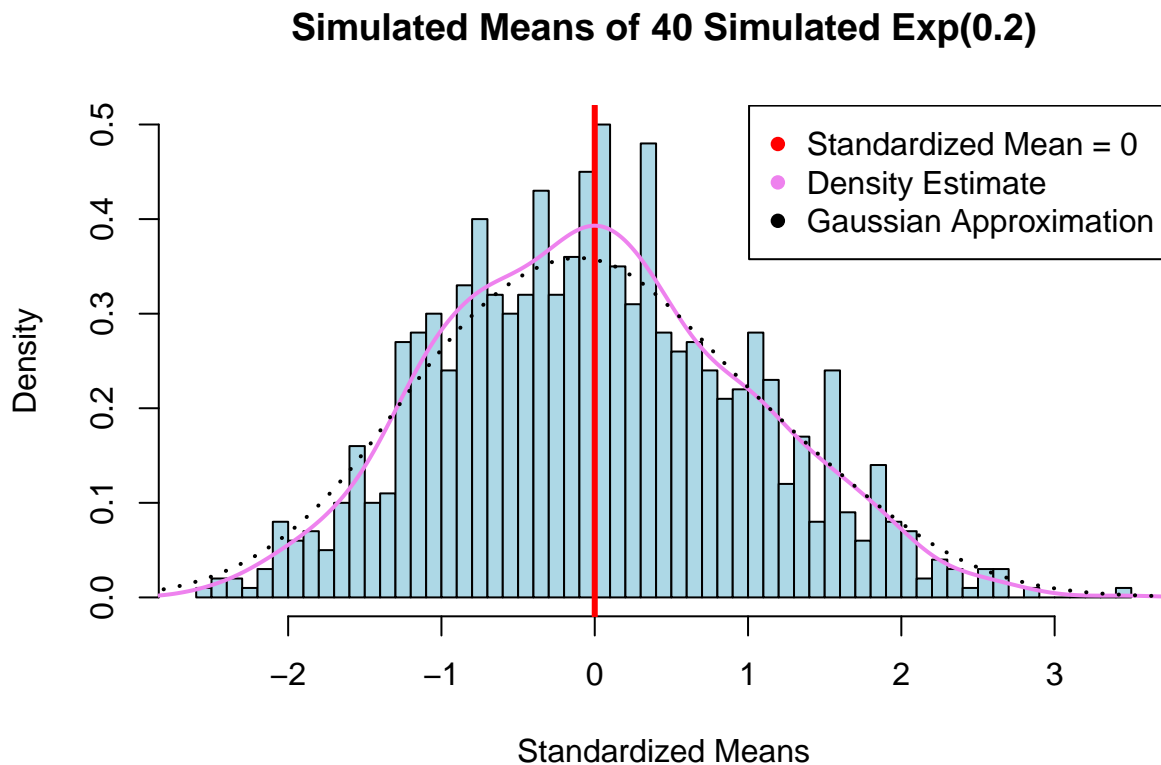
```
varOfMeans <- var(means)
sdOfMeans <- sd(means)
expVar <- (1/0.2)^2
expSD <- 1/0.2
```

The variation of the simulated data is 0.6587 while the expected theoretical variation is 25. It is not surprising that the simulated variation does not closely align with the theoretical variance, since the simulated means clearly follow a normal distribution themselves, as predicted by the Central Limit Theorem and demonstrated below.

Goal 3: Demonstrate that the simulated means follow a normal distribution.

Exploratory analysis of the simulated data by way of a histogram show that the simulated data are normally distributed about the mean = 5.0162, with a standard deviation of 0.8116.

```
Z<-(means-mean(means))/sdOfMeans
hist(Z, prob=TRUE, col="lightblue", breaks = 50, xlab = "Standardized Means", main="Simulated Means of 40 Simulated Exp(0.2)")
abline(v=0, col="red", lwd=3)
lines(density(Z), col="violet", lwd=2) # add a density estimate with defaults
lines(density(Z, adjust=2), lty="dotted", col="black", lwd=2) # add another smoother density
legend("topright", pch = 16, col = c("red", "violet", "black"), legend = c("Standardized Mean = 0", "Density Estimate", "Gaussian Approximation"))
```



Goal 4: Evaluate the coverage of the confidence interval for $1/\lambda$: $\bar{X} \pm 1.96S/\sqrt{n}$.

```
ll<-mean(means)-1.96*expSD/sqrt(40)
ul<-mean(means)+1.96*expSD/sqrt(40)
coverage<-subset(means, means>ll & means<ul)
percent <- length(coverage)/1000*100
```

The 95% confidence interval of simulated means is 3.4667-6.5658 with a 94.9% coverage. Since the method used to create the CI is low-biased, this coverage percentage makes sense.