

Part 1: Simulation Exercise

Andrew Perelson

March 15, 2017

Overview

This project investigates the exponential distribution in R and compares it with the Central Limit Theorem. Lambda will be set to 0.2 for all of the simulations and the report investigates the distribution of averages of 40 exponentials for a thousand simulations.

Simulations

Comparing The Sample and Theoretical Means

Set the seed (for repeatability), lambda and exponentials and load required library:

```
library(ggplot2)
set.seed(2000)
lambda <- 0.2
exp <- 40
sim <- data.frame(Mean=numeric(0))
```

Run the required simulations with the above set variables

```
for (i in 1:1000) { sim <- rbind(sim, mean(rexp(exp, lambda))) }
names(sim) <- "Mean"
```

Determine the sample mean:

```
sampleMean <- mean(sim$Mean)
print(sprintf("The sample mean converged to: %f", sampleMean))
```

```
## [1] "The sample mean converged to: 5.029410"
```

Determine the theoretical mean:

The theoretical mean of an exponential distribution is λ^{-1} .

```
theoreticalMean <- lambda^-1
print(sprintf("The theoretical mean converged to: %f", theoreticalMean))
```

```
## [1] "The theoretical mean converged to: 5.000000"
```

Comparison of the sample vs theoretical means:

There is only a slight difference between the sample mean and the theoretical mean.

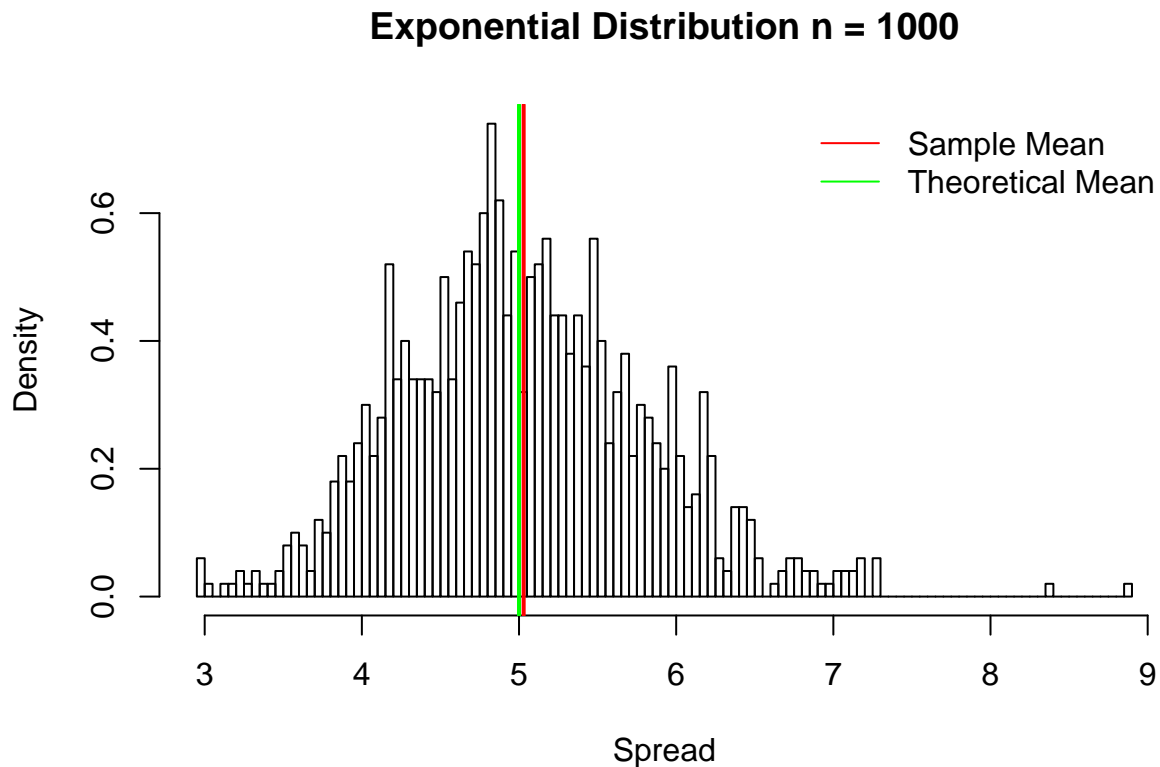
```
abs(sampleMean-theoreticalMean)
```

```
## [1] 0.02941033
```

Diagram plot indicating mean distribution with sample and theoretical means:

```
hist(sim$Mean,
     breaks = 100,
     prob = TRUE,
     main="Exponential Distribution n = 1000",
     xlab="Spread")
abline(v = sampleMean,col="red",lwd = 2)
abline(v = theoreticalMean,col="green",lwd = 2)

legend('topright', c("Sample Mean", "Theoretical Mean"),
      bty = "n", lty = c(1,1), col = c(col = "red", col = "green"))
```



Comparing The Sample and Theoretical Variance

Determine the sample variance:

```
sampleVariance <- var(sim$Mean)
print(sprintf("The sample variance: %f", sampleVariance))
```

```
## [1] "The sample variance: 0.616808"
```

Determine the theoretical variance:

The theoretical variance of an exponential distribution is $(\lambda * \sqrt{n})^2$.

```
theoreticalVariance <- (lambda * sqrt(exp))^-2
print(sprintf("The theoretical variance: %f", theoreticalVariance))
```

```
## [1] "The theoretical variance: 0.625000"
```

Comparison of the sample vs theoretical variance:

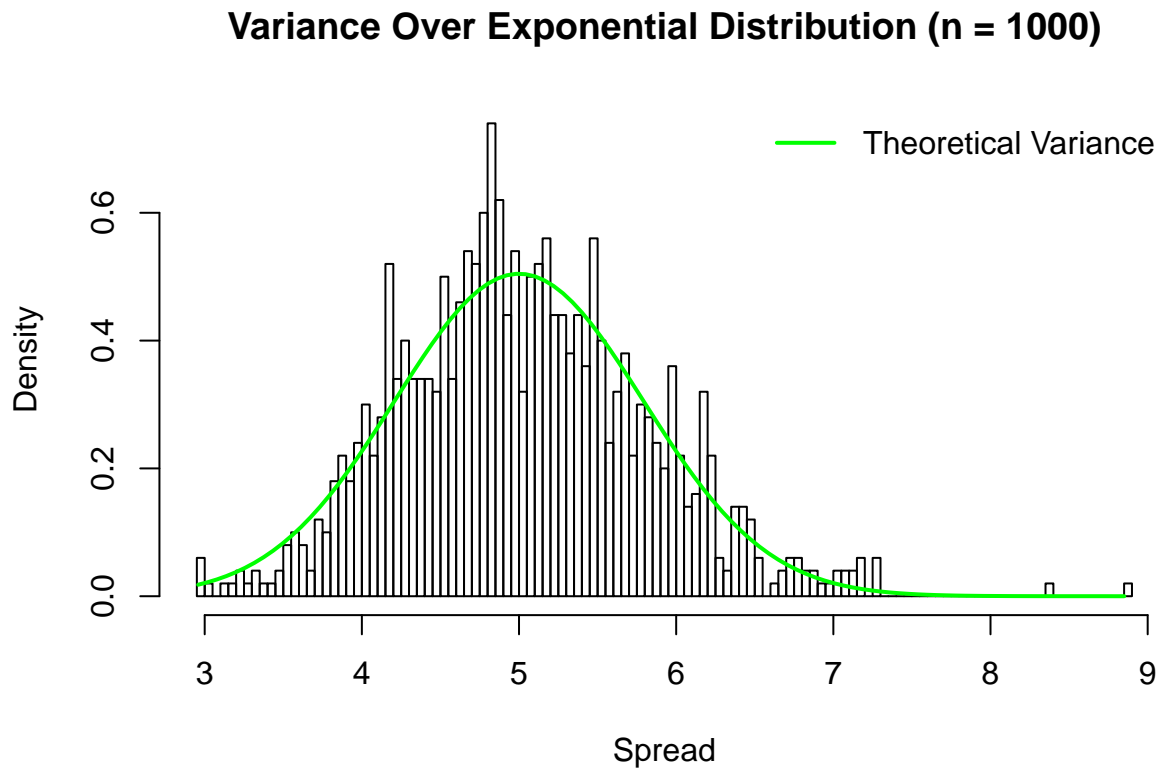
There is a slight difference between the simulation sample variance and the exponential distribution theoretical variance:

```
abs(sampleVariance-theoreticalVariance)
```

```
## [1] 0.008191757
```

Diagram plot indicating variance compared to theoretical variance:

```
hist(sim$Mean,
     breaks = 100,
     prob = TRUE,
     main = "Variance Over Exponential Distribution (n = 1000)",
     xlab = "Spread")
xfit <- seq(min(sim$Mean), max(sim$Mean), length = 100)
yfit <- dnorm(xfit, mean = 1/lambda, sd = (1/lambda/sqrt(40)))
lines(xfit, yfit, pch = 22, col = "green", lwd = 2)
legend('topright', c("Theoretical Variance"),
     lty = 1, lwd = 2, bty = "n", col = "green")
```



Showing that the distribution is approximately normal:

This is a density histogram of the 1000 simulations. There is an overlay with a normal distribution that has a mean of λ^{-1} and standard deviation of $(\lambda \sqrt{n})^{-1}$, the theoretical normal distribution for the simulations.

```
ggplot(data.frame(y=sim$Mean), aes(x=y)) +  
  geom_histogram(aes(y=..density..), binwidth=0.2, fill="wheat", color="black") +  
  stat_function(fun=dnorm, args=list(mean=lambda^-1, sd=(lambda*sqrt(exp))^-1)) +  
  labs(title="Distribution compared to normal", x="Mean", y="Density")
```

