

Statistical Inference Course Project - Part 1

Amber Beasock

27 December 2015

Synopsis

This is the project for the statistical inference class. In it, you will use simulation to explore inference and do some simple inferential data analysis. The project consists of two parts:

1. A simulation exercise.
 2. Basic inferential data analysis.
-

Instructions

The exponential distribution can be simulated in R with `rexp(n, lambda)` where `lambda` is the rate parameter. The mean of exponential distribution is $1/\lambda$ and the standard deviation is also $1/\lambda$. Set `lambda = 0.2` for all of the simulations. In this assignment, you will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations. Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials.

Set-up

```
# set seed for reproducibility
set.seed(43)

# set lambda to 0.2
lambda <- 0.2

# 40 samples
n <- 40

# 1000 simulations
nosim <- 1000

# Simulation
sim <- matrix(rexp(nosim * n, rate = lambda), nosim)

# Simulation Means
sim_means <- apply(sim, 1, mean)
```

Question 1

Show the sample mean and compare it to the theoretical mean of the distribution.

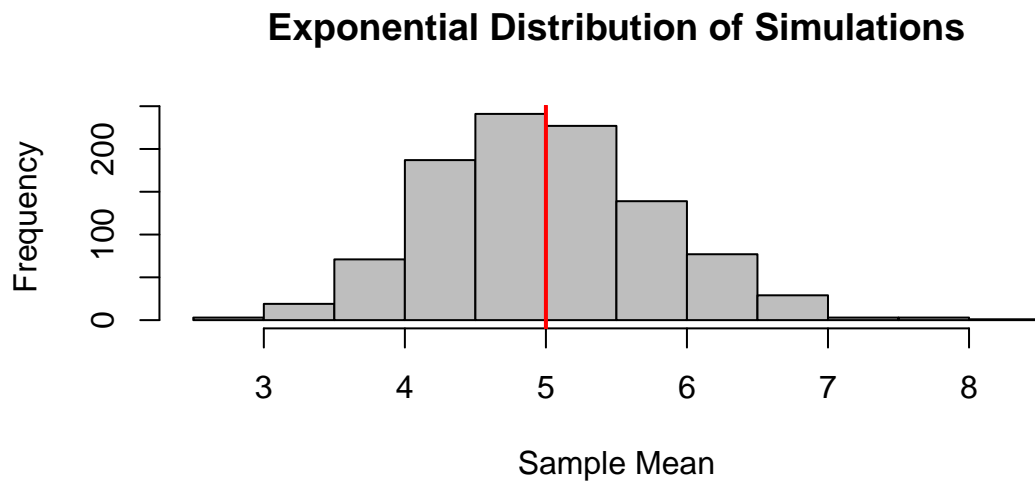
```
# Theoretical Mean
theoretical_mean <- 1/lambda
print(paste("Theoretical Mean =", theoretical_mean))

## [1] "Theoretical Mean = 5"

# Sample Mean
sample_mean <- mean(sim_means)
print(paste("Sample Mean=", sample_mean))

## [1] "Sample Mean= 5.00079782063507"

# Visualization
hist(sim_means, xlab="Sample Mean", main="Exponential Distribution of Simulations", col="grey")
# Vertical line on histogram representing the theoretical mean
abline(v = theoretical_mean, col = "red", lwd = 2)
```



Answer 1

The sample mean is 5.0007978 and the theoretical mean is 5. Thus there is no statistically significant difference between the sample mean and the theoretical mean.

Question 2

Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

```
# Theoretical Standard Deviation
tsd <- 1/lambda/sqrt(n)

# Theoretical variance
tvar <- 1/lambda^2/n
print(paste("Theoretical Variance =", tvar))
```

```
## [1] "Theoretical Variance = 0.625"
```

```
#Sample Standard Deviation
sim_sd <- sd(sim_means)

# Sample variance
sim_var <- var(sim_means)
print(paste("Sample Variance =", sim_var))
```

```
## [1] "Sample Variance = 0.636872886765362"
```

Answer 2

The theoretical variance is 0.625 and the sample variance of the distribution is 0.6368729.

Question 3

Show that the distribution is approximately normal.

```
hist(sim_means, breaks=n, prob=T, col="lightgrey", xlab="Sample Means", ylab="Density", main="Distribut
  lines(density(sim_means))
  xfit<-seq(min(sim_means),max(sim_means),length=100)
  yfit<-dnorm(xfit,mean=1/lambda,sd=1/lambda/sqrt(n))
  lines(xfit, yfit, col="red", lty=2)
  abline(v = theoretical_mean, col = "blue", lwd = 2)
  legend('topright', c("Simulation Density", "Theoretical Density", "Theoretical Mean"), lty=c(1,2), c
```

Distribution of Sample Means

