Simulation Exercises

Tuan Vu

Saturday, October 25, 2014

This is the first part of Statistical Inference Project.

Requirements 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 also 1/lambda. Set lambda = 0.2 for all of the simulations. In this simulation, you will investigate the distribution of averages of 40 exponential(0.2)s. Note that you will need to do a thousand or so simulated averages of 40 exponentials.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponential(0.2)s. You should 1. Show where the distribution is centered at and compare it to the theoretical center of the distribution. 2. Show how variable it is and compare it to the theoretical variance of the distribution. 3. Show that the distribution is approximately normal. 4. Evaluate the coverage of the confidence interval for 1/lambda: $\bar{X} \pm 1.96Sn$???.

Let's do some analysis

Load the package

```
library(ggplot2)
```

Prepare the data

```
lambda <- 0.2
n <- 40 # size
simulation <- 1000 # number of simulations
data <- replicate(simulation, mean(rexp(n, lambda)))
```

Theoretical Value

```
tmean <- 1/lambda
tvar <- 1/(lambda^2*n)
tsd <- (1/lambda)/sqrt(n)
```

Simulated Data Value

```
smean <- mean(data) # Mean
svar <- var(data)
smedian <- median(data) # Median
ssd <- sd(data) # Standard Deviation
```

1. Where the distribution is centered at and compare it to the theoretical center of the distribution?

```
smean # Simulated Mean
```

```
## [1] 5.035466
```

```
tmean # Theoretical Mean
```

```
## [1] 5
```

2. How variable it is and compare it to the theoretical variance of the distribution?

```
svar # Simulated value
```

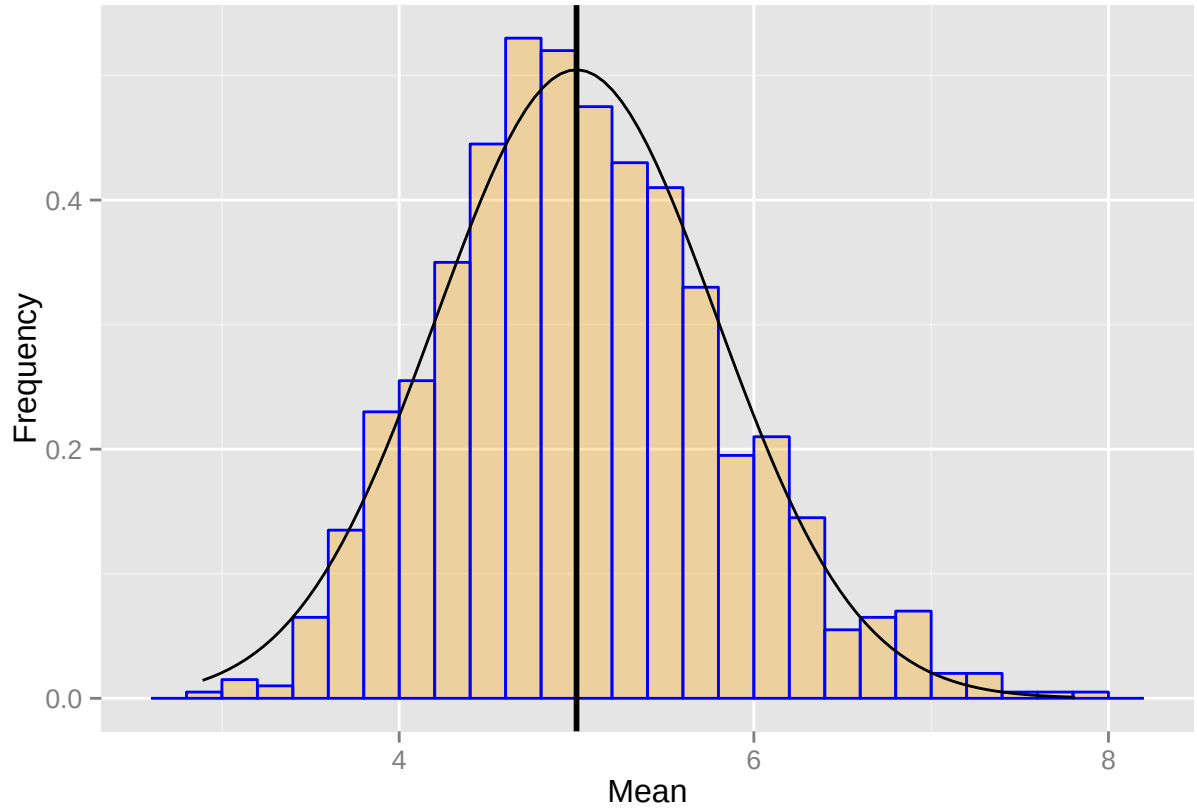
```
## [1] 0.6221573
```

```
tvar # Theoretical value
```

```
## [1] 0.625
```

3. Show that the distribution is approximately normal

```
data2 <- data.frame(data)
hist <- ggplot(data2, aes(x = data))
hist <- hist + geom_histogram(aes(y = ..density..), colour = "blue",
fill = "orange", alpha = .3,binwidth=.2)
hist <- hist + stat_function(fun = "dnorm", args = list(mean = tmean, sd = tsd))
hist <- hist + geom_vline(xintercept=tmean,size=1)
hist <- hist + xlab("Mean")+ylab("Frequency")
hist
```



4. Evaluate the coverage of the confidence interval for 1/lambda

```r
smean + (c(1,-1)*1.96 * (ssd/sqrt(n)))
```

```
## [1] 5.279908 4.791023
```