

Statistical Interface Assignment, Part A

Vivek Ranjan

November 22, 2015

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 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.

Questions 1 and 2

Show where the distribution is centered at and compare it to the theoretical center of the distribution. Show how variable it is and compare it to the theoretical variance of the distribution.

```
lambda = 0.2
n = 40
nsms = 1:1000
set.seed(820)
means <- data.frame(x = sapply(nsms, function(x) {mean(rexp(n, lambda))}))
head(means)

##           x
## 1 5.750000
## 2 3.808205
## 3 4.058154
## 4 3.999241
## 5 4.312532
## 6 4.418246

mean(means$x)

## [1] 4.998812

sd(means$x)

## [1] 0.7909422

(1/lambda)/sqrt(40)

## [1] 0.7905694

var(means$x)

## [1] 0.6255895
```

```
((1/lambda)/sqrt(40))^2
```

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

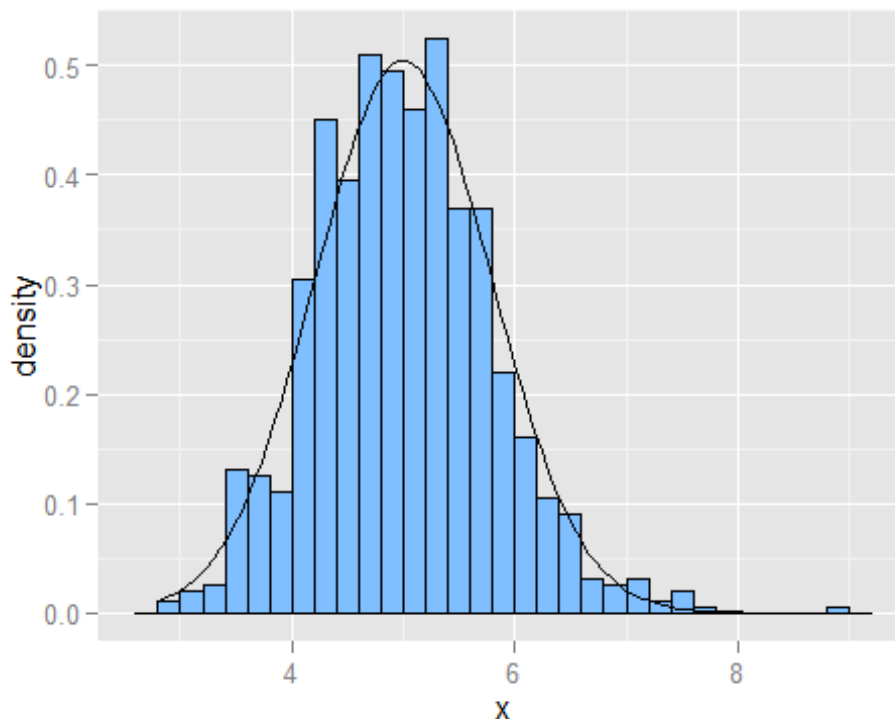
Center of the distribution: 4.9988. Expected center: 5.0. The mean of the means of the exponential of 1000 simulations of 40 exponential(0.2)s is 4.9988, which is very close to the expected mean of $1/0.2 = 5.0$.

Variability of the distribution. The standard deviation of 0.7909 is also close to the expected standard deviation of 0.79056

Question 3

Show that the distribution is approximately normal. Below is a histogram plot of the means of the 1000 simulations of `rexp(n, lambda)`. It is overlaid with a normal distribution with mean 5 and standard deviation 0.7909. Yes, the distribution of our simulations appears normal.

```
library(ggplot2)
ggplot(data = means, aes(x = x)) +
  geom_histogram(aes(y=..density..), fill = I('#80bfff'),
                 binwidth = 0.20, color = I('black')) +
  stat_function(fun = dnorm, arg = list(mean = 5, sd = sd(means$x)))
```



Question 4

Evaluate the coverage of the confidence interval

```
mean(means$x) + c(-1,1)*1.96*sd(means$x)/sqrt(nrow(means))  
## [1] 4.949789 5.047835
```

Coursera Statistical Interface by [Vivek Ranjan](#)