Coursera.org

Statistical Inference Assignment December 2014

Martin Sawtell

Part 1: Simulation

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.

1. Show where the distribution is centered at and compare it to the theoretical center of the distribution.

1000 repetitions of the rexp function, each time calculating the mean of 40 exponential(0.2) results.

```
> set.seed(123)
> lambda = 0.2
> numsims = 1:1000
> n = 40
> simdata <- data.frame(x = sapply(numsims,
function(x) {mean(rexp(n, lambda))}))
> head(simdata)
```

Rstudio output:

```
        x
1 4.811212
2 5.360077
3 4.592871
4 4.900051
5 5.516619
6 5.612835
```

Take the mean of the simulated means:

```
> mean(simdata$x)
```

```
[1] 5.011911
```

Question 1 ANSWER:
Theoretical result: "The mean of exponential distribution is 1/lambda" == 1/0.2 == 5
-This seems to be approximated by the sim result as we know that the mean of the sample converges to the mean of the population.

2. Show how variable it is and compare it to the theoretical variance of the distribution.

Calculate the variance:

```
> sd(simdata$x)^2
```

```
[1] 0.6004928
```

Theoretical variance of rexp:

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

```
[1] 0.625
```

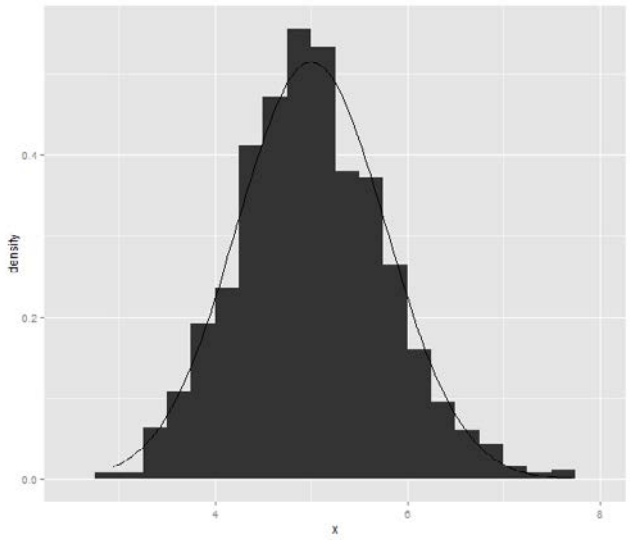
Question 2 ANSWER:
The calculated variance is similar to the theoretical variance, as demonstrated with the above calculations.

3. Show that the distribution is approximately normal.

Lets draw it as a histogram along with a normal distribution curve:

```
> library(ggplot2)
> ggplot(data = simdata, aes(x = x))
        + geom_histogram(aes(y=..density..), bin
        width = 0.25)
        + stat_function(fun = dnorm, arg =
        list(mean = 5, sd = sd(simdata$x)))
```

Rstudio output:



Question 3 ANSWER:
As you can see, the simdata distribution follows/approximates a normal distribution.