Statistical Inference Course Project

JLei

5/1/2020

This project consists of two parts:

- A simulation exercise
- Basic inferential data analysis

Part 1 - Simulation Exericse

This part allows one to investigate the exponential distribution in R and compare it with the Central Limit Theorem.

The exponential distribution can be simulated in R with *rexp(n, lambda)* where lambda is the rate parameter. The mean and the standard distribution of exponential distribution are both 1/lambda. For the simulation, lambda is set to **0.2**, n is set to **40**, and the number of simulation is set to **1000** to investigate the distribution of averages of 40 exponentials.

```
set.seed(123456)
## simulate 1000 averages of 40 exponentials
mns <- NULL
for (i in 1:1000) {
        mns <- c(mns, mean(rexp(40, 0.2)))
}

## simulate a normal distribution with mean of 5 and sd of 5/sqrt(n) with 1000 simulations (rnorm())
exmns <- rnorm(1000, 5, 5/sqrt(40))

dataset <- data.frame(mns, exmns)
```

1. What is the sample mean and how does it compare to the theoretical mean of the distribution (which is 1/lambda, 1/0.2=5)?

```
mean(mns)
```

```
## [1] 5.022915
```

Answer: The sample mean is 5.023, approximating the theoretical mean of 5 very closely.

2. What is the sample variance and how does it compare to the theoretical variance of the distribution?

```
## sample variance
sd(mns)^2
```

```
## [1] 0.6570391
```

```
## theoretical variance of the distribution of averages of 40 exponentials
(1/0.2/sqrt(40))^2
```

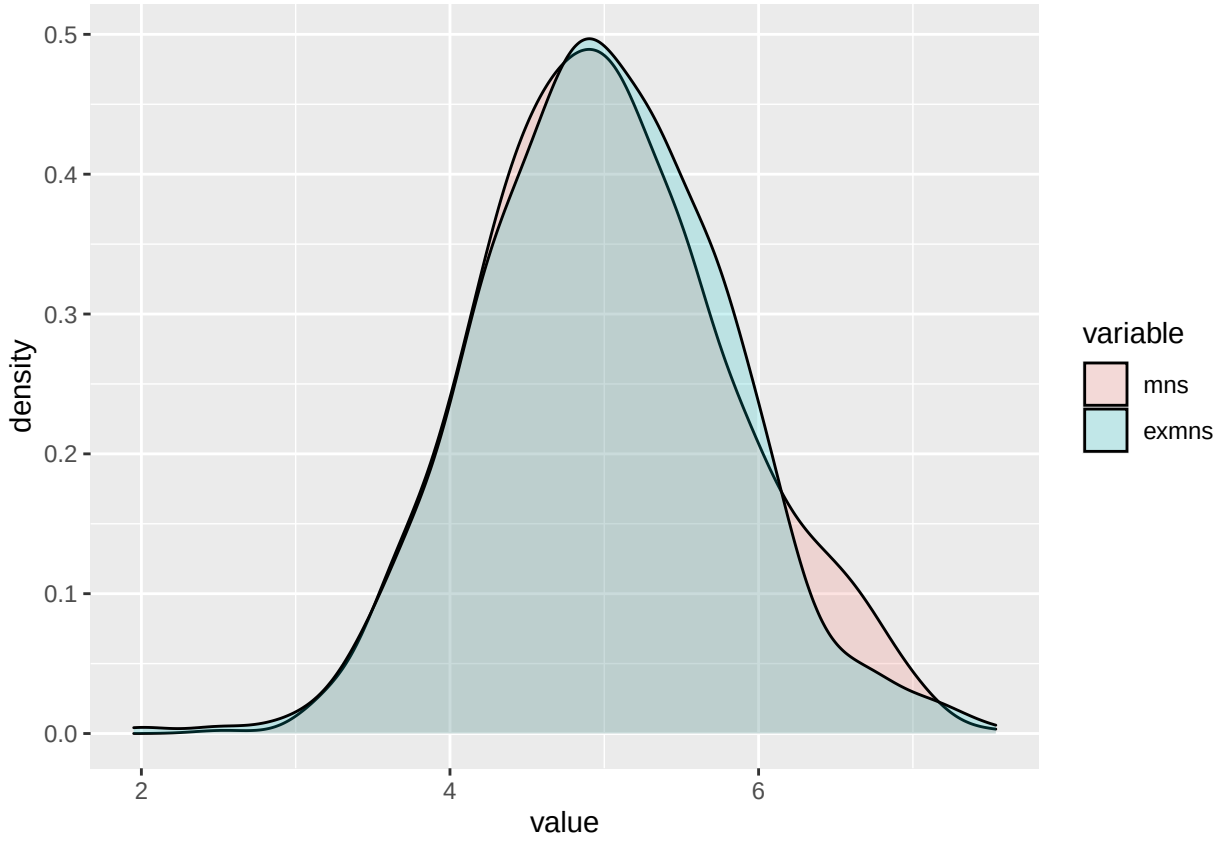
```
## [1] 0.625
```

Answer: The sample variance is 0.657 and it's close to the theoretical variane of the distribution 0.625.

3. Show that the distribution is approximately normal.

```
library(ggplot2)
library(reshape2)
ggplot(melt(dataset), aes(x=value, fill=variable))+geom_density(alpha=0.2)
```

```
## No id variables; using all as measure variables
```



Comment: The density curve for sample mean distribution **mns** (pink color) is approximately normal, which is plotted in blue.