Statistical Inference Course Project

Nour Qweder

8/13/2020

1.2.) Simulations

```
## Check for missing dependencies
library(ggplot2)
library(lemon)
```

```
##
## Attaching package: 'lemon'
```

```
## The following objects are masked from 'package:ggplot2':
##
##     CoordCartesian, element_render
```

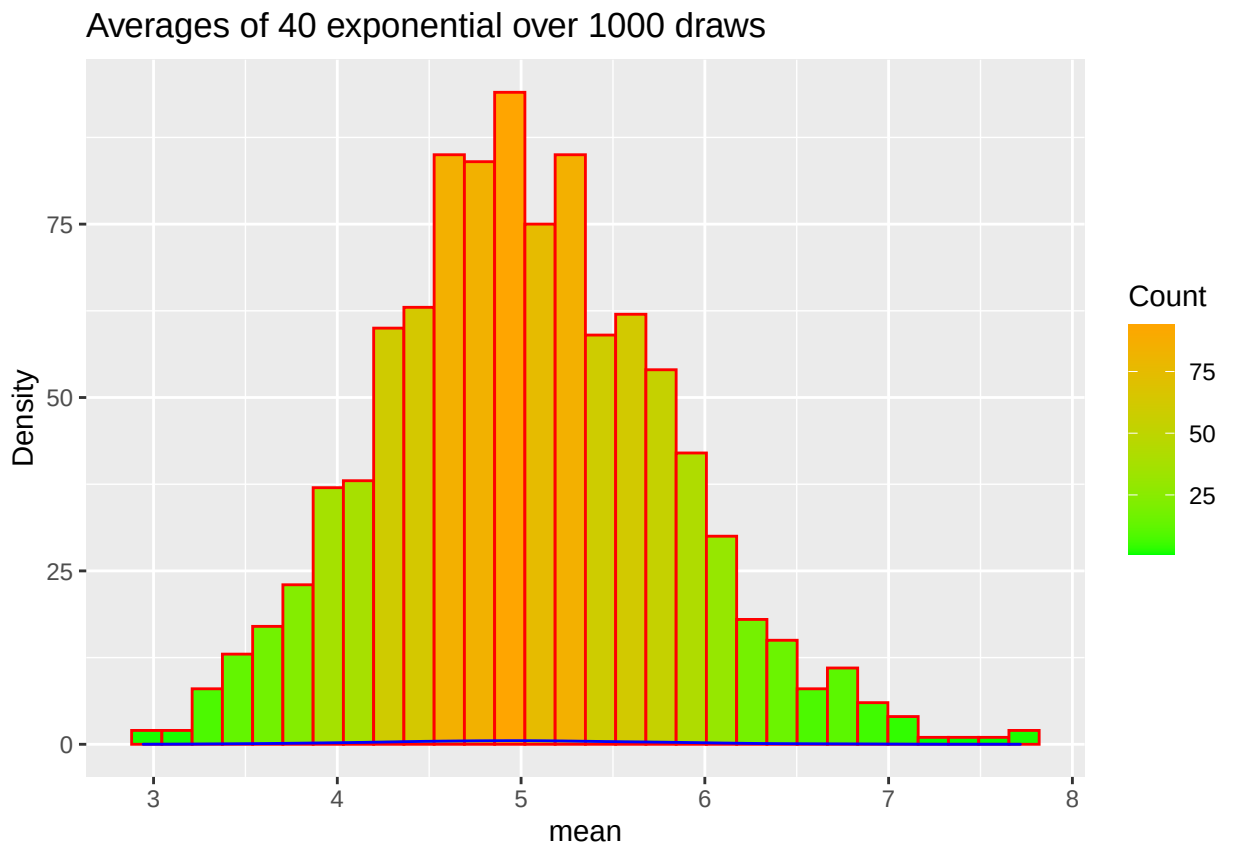
```
#install.packages("lemon")
knit_print.data.frame <- lemon_print
## init
lambda <- 0.2
n <- 40
sims <- 1:1000
set.seed(123)

# Simulate the draws
draws <- data.frame(x=sapply(sims, function(x) {mean(rexp(n, lambda))}))
#?rexp
# Plot



p1 <- ggplot(draws, aes(x=x)) +
  geom_histogram(
    col="red",
    aes(y=..count.., fill=..count..)) +
    scale_fill_gradient("Count", low="green", high="orange")+
    geom_density(colour="blue") +
  labs(title="Averages of 40 exponential over 1000 draws", y="Density", x="mean")
p1
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



1.3.) Sample mean vs theoretical mean

```
# Tabulating the Sample Mean & Theoretical Mean
sample.mean <- mean(draws$x)
theo.mean <- 1/lambda
mean.df1<-cbind(sample.mean, theo.mean)
knitr::kable(mean.df1)
```

sample.mean	theo.mean
5.011911	5

```
t.test(draws$x)[4]
```

```
## $conf.int
## [1] 4.963824 5.059998
## attr(,"conf.level")
## [1] 0.95
```

From previous block, it can be noticed that sample mean and the theoretical mean approximately close. Plus, after computing the confidence interval on the mean, it observed that the sample mean is between *[4.963824 5.059998]* at 95% confidence interval interval,

1.4.) Sample Variance vs theoretical Variance

```
# Tabulating the Sample Mean & Theoretical Mean
sample.var <- var(draws$x)
```

```r
theo.var <- (1/(lambda^2))/n
var.df<-cbind(sample.var, theo.var)
knitr::kable(var.df)
```

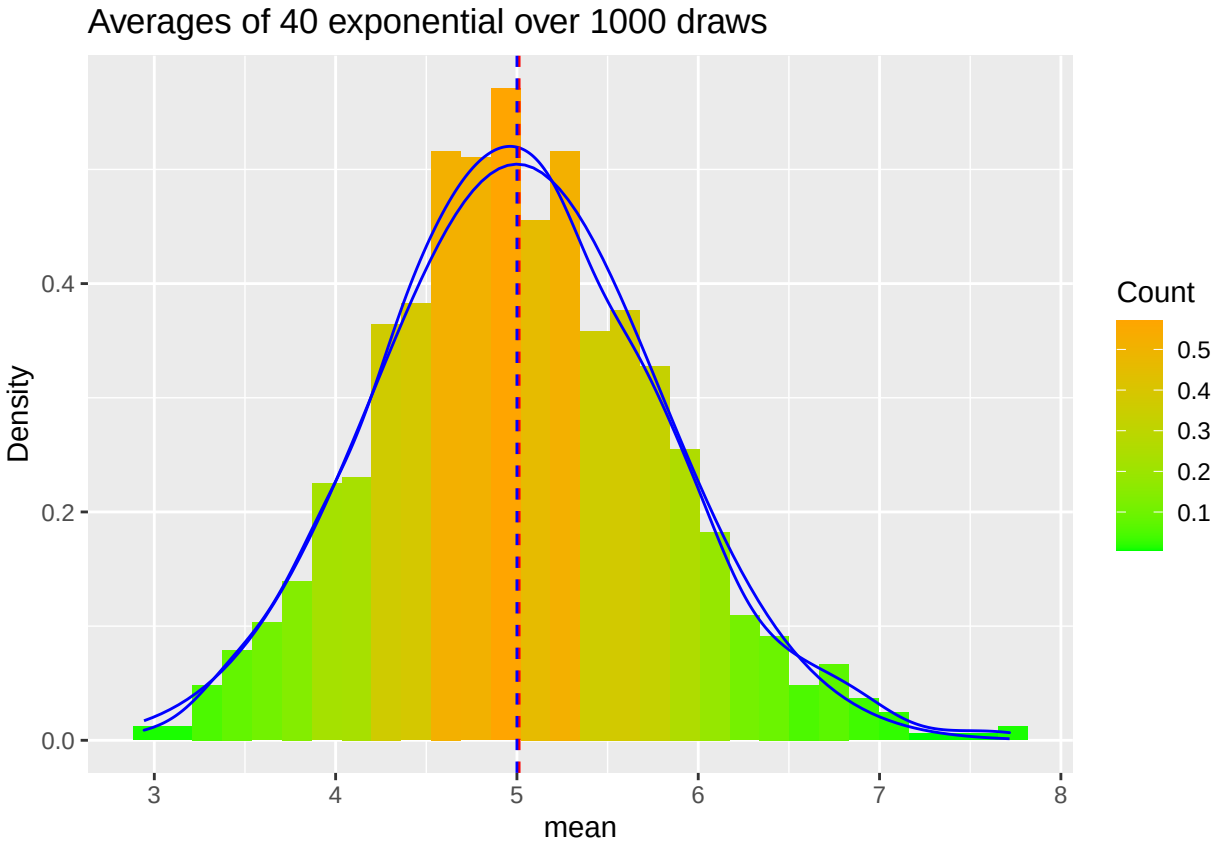
sample.var	theo.var
0.6004928	0.625

From previous block, it can be noticed that sample variance and the theoretical variance approximately close. Plus, after computing the confidence interval on the variance, it observed that the sample variance is between *[0.6004928 0.625]* at 95% confidence interval interval,

1.5.) Distribution

```r
# Plotting Sample Mean & Varience vs Theoretical Mean & Varience
p2 <- ggplot(draws, aes(x=x)) +
  geom_histogram(aes(y=..density.., fill=..density..)) +
  scale_fill_gradient("Count", low="green", high="orange")+
  labs(title="Averages of 40 exponential over 1000 draws", y="Density", x="mean") +
  geom_density(colour="blue") +
  geom_vline(xintercept=sample.mean, colour="red", linetype=2) +
  stat_function(fun=dnorm,args=list( mean=1/lambda, sd=sqrt(theo.var)),color = "blue") +
  geom_vline(xintercept=theo.mean, colour="blue", linetype=2)
p2
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```


Averages of 40 exponential over 1000 draws
Density
0.0
0.2
0.4
3
4
5
6
7
8
mean
Count
0.5
0.4
0.3
0.2
0.1