Simulation Exercise

Shaaguunz

5/18/2020

** First part of Statistical Inference Project. **

Question to answer

Show the sample mean and compare it to the theoretical mean of the distribution. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution. Show that the distribution is approximately normal. In point 3, focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

let's begin

```
library(ggplot2)
```

prepare the data

```
lambda<-0.2
n<-40
simulation<-1000
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)
svar<-var(data)
smedian<-median(data)
ssd<-sd(data)
```

1. Show the sample mean and compare it to the theoretical mean of the distribution.

```
smean
```

```
## [1] 4.984598
```

```
tmean
```

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

2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

```
svar
```

```
## [1] 0.6769548
```

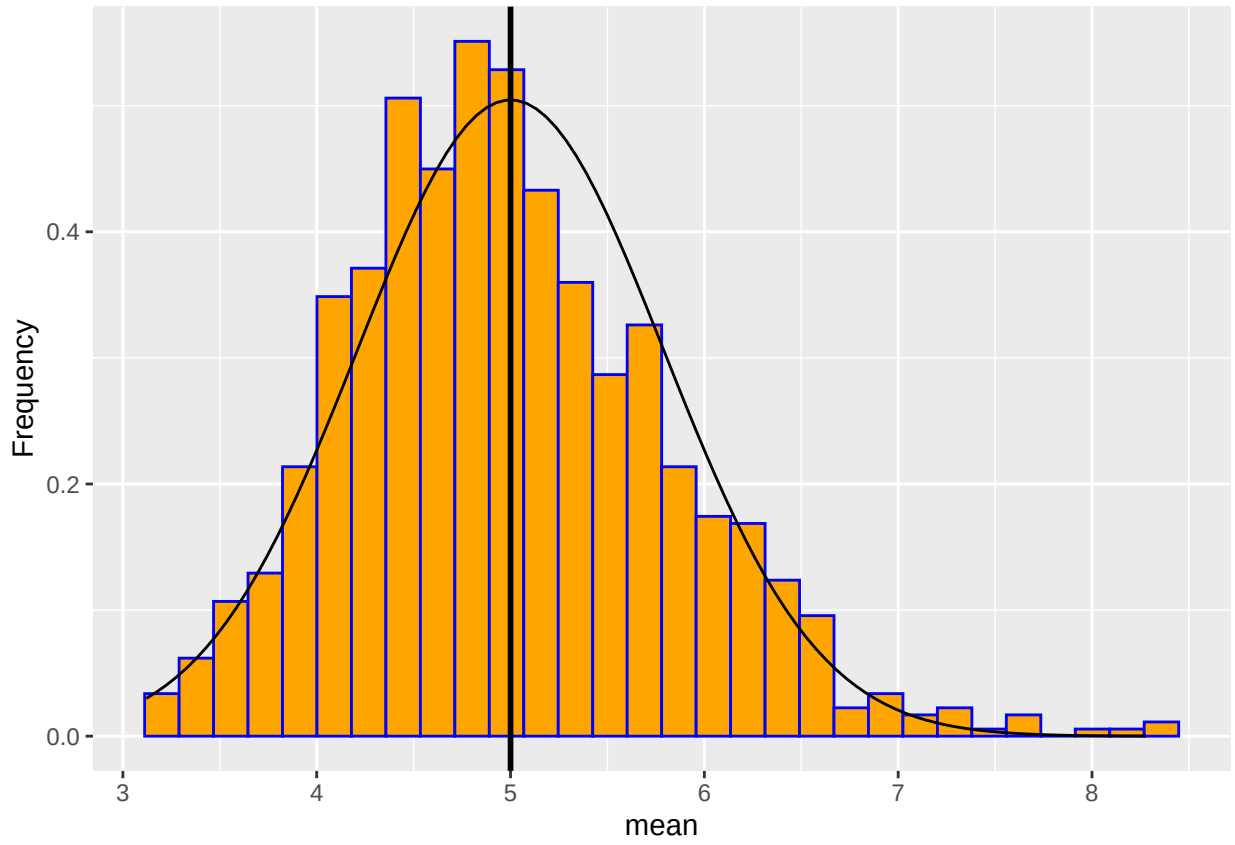
```
tvar
```

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

3.focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

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

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



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

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

```
## [1] 5.239578 4.729619
```