Sample Project

Alexander Klonov

Exponential distribution and Central Limit Theorem

Investigate the exponential distribution and compare it with the Central Limit Theorem. Based on CLT, mean and variance of the exponential distribution will be compared to the theoretical ones.

The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. For this case lambda is 0.2. Basic preparations:

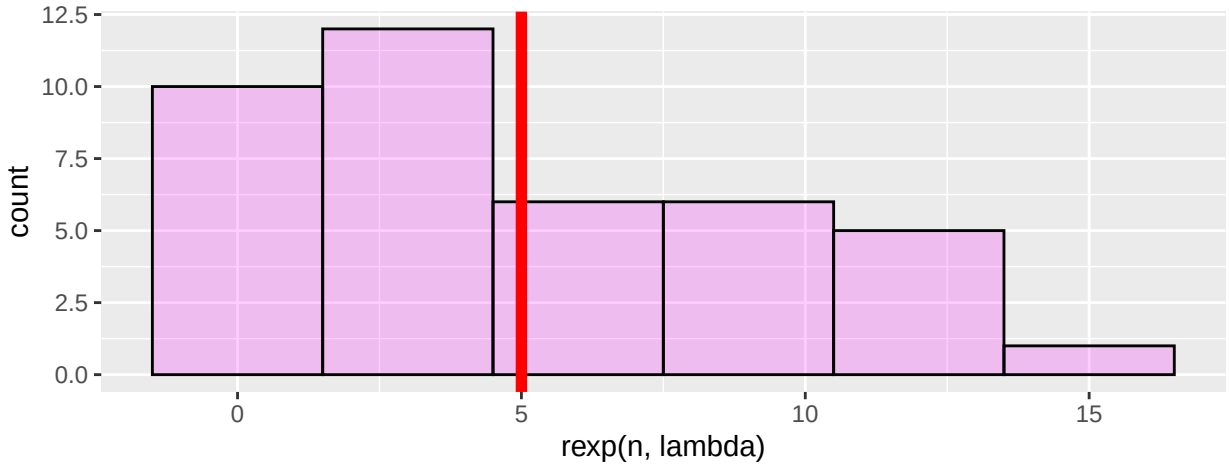
```
library(ggplot2)
lambda<-0.2
n<-40
simN<-1000
```

Theoretical mean is equal to 5. However, one sample mean is different from theoretical:

```
mean(rexp(n, lambda))
```

```
## [1] 5.578634
```

```
g<-ggplot(data=as.data.frame(rexp(n, lambda),
                             aes = (x =rexp(n, lambda) )))
g<-g+geom_histogram(fill="magenta",alpha = .20,
                    binwidth=3, color = "black",
                    aes(x =rexp(n, lambda) ))+
  geom_vline(xintercept = 5, size = 2, col="red")
g
```



Same goes for variance. Theoretical:

```
1/lambda^2
```

```
## [1] 25
```

And from one sample:

```
var(rexp(n, lambda))
```

```
## [1] 17.34798
```

Making a one thousand simulations

Central Limit Theorem is quite useful here - it can be used to find mean and variance.

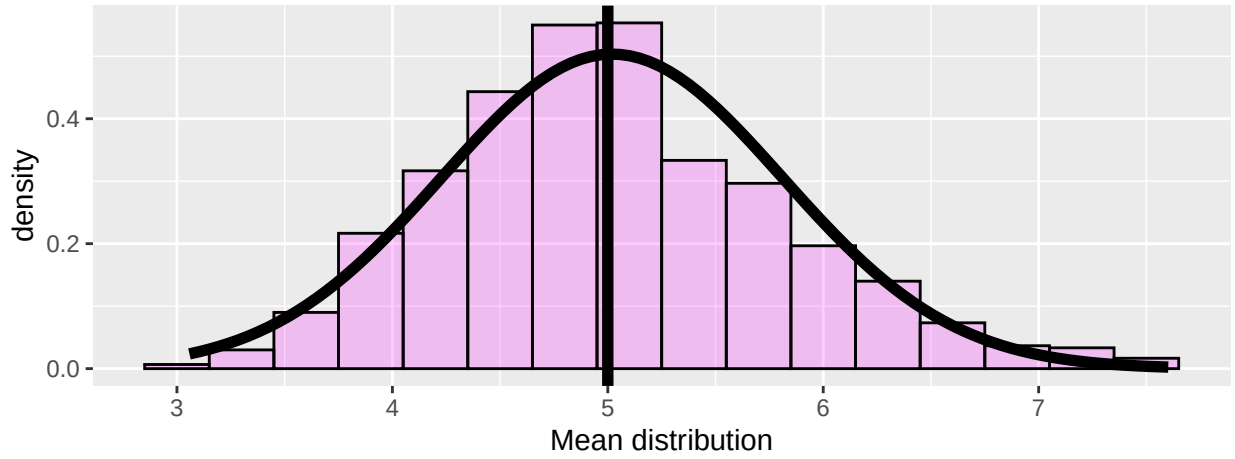
Here, we are finding mean of the 1000 exponential distributions:

```
mns<-NULL
for(i in 1:simN) mns<-c(mns, mean(rexp(n, lambda)) )
```

Mean distribution And now, we are going to plot this along with normal distribution line.

```
Mmns<-mean(mns)
smns<-sd(mns)

g<-ggplot(data=as.data.frame(mns), aes(x=mns))
g<-g+geom_histogram(fill = "magenta",alpha = .20,
    binwidth=.3, color = "black", aes(y = ..density..))+
  geom_vline(xintercept = 5, size = 2)
g<-g+stat_function(fun= function(mns)
  dnorm(mns, mean = Mmns, sd = smns), size =2)+
  xlab("Mean distribution")
g
```



As we can see, distribution looks like normal, the more simulations lead to more "normal" like distribution.

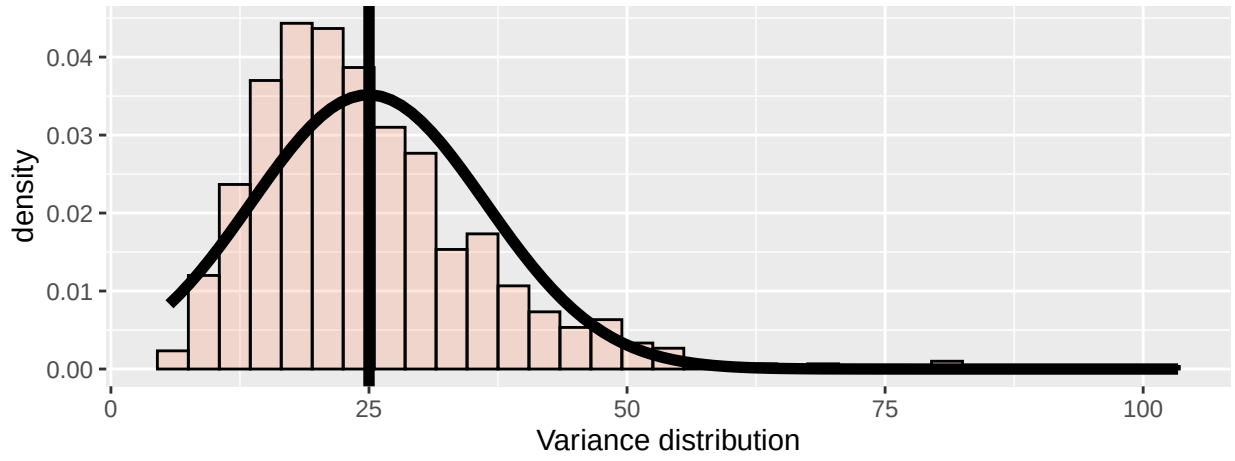
Variance distribution Same principles goes for variance distribution:

Making a 1000 simulations

```
mns<-NULL
for(i in 1:simN) mns<-c(mns, var(rexp(n, lambda)) )
```

Histogram:

```
Vmns<-mean(mns)
smns<-sd(mns)
g<-ggplot(data=as.data.frame(mns), aes(x=mns))
g<-g+geom_histogram(fill = "coral",alpha = .20,
                    binwidth=3, color = "black", aes(y = ..density..))+
  geom_vline(xintercept = 25, size = 2)
g<-g+stat_function(fun= function(mns)
  dnorm(mns, mean = Vmns, sd = smns), size =2)+
  xlab("Variance distribution")
g
```



Conclusion

In this task, we used a CLT to find mean and variance of the exponential distribution. By one thousand simulations we got result very close to theoretical one. Based on this, CLT may come very helpful in work with small samples.