Distribution of exponential samples mean

Lukasz Konczyk

26 kwietnia 2018

Part One

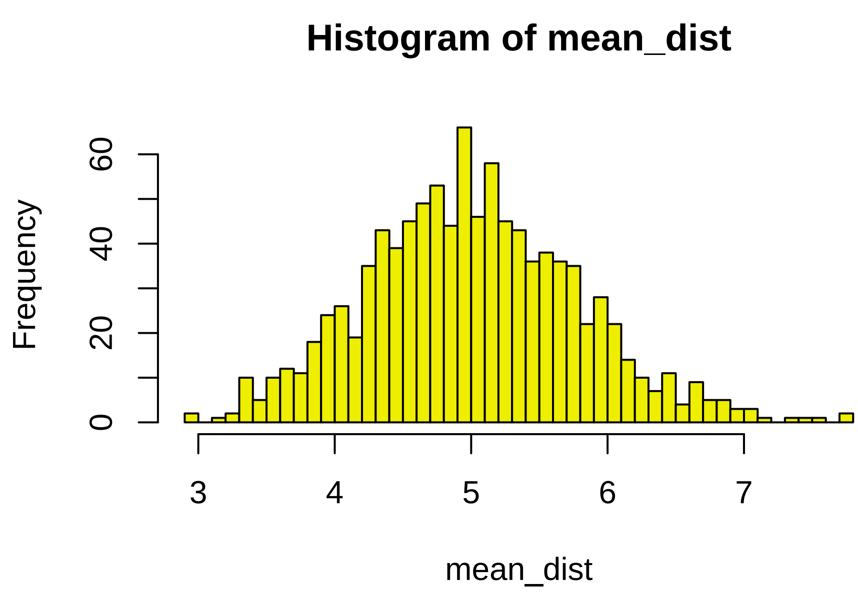
In this part of project we will investigate how distribution of exponential samples mean converges to normal distribution according to Central Limit Theorem. Our sample contains 40 exponential. We will simulate sample 1000 times. Exponential distribution has one parameter - lambda. The mean of distribution is equal to 1/lambda and variance is equal to 1/lambda also. Let's set lambda=0.2. Below code perform simulation to produce distribution of exponential mean.

```
set.seed(123)
sims_matrix<-matrix(rep(0,40*1000),ncol=1000)
for (i in 1:1000){
  sims_matrix[,i]<-rexp(40,0.2)
}
mean_dist<-apply(sims_matrix,2,mean)
```

1. Mean of distribution

In this part we will focus on mean of distribution. Firstly, have a look on histogram and see the date.

```
hist(mean_dist,breaks=50, col="yellow2")
```

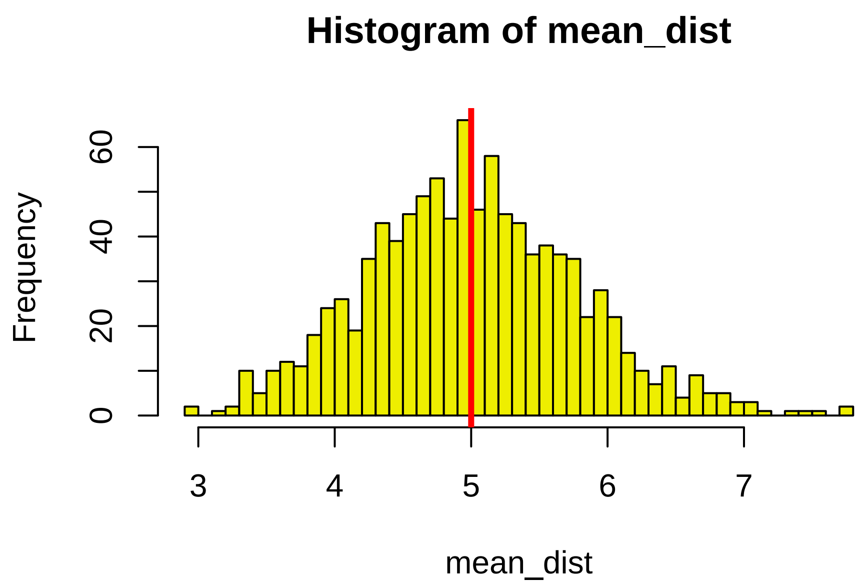


Now we calculate mean in sample and compare with theoritical value.

```
mean1<-mean(mean_dist)
mean2<-1/0.2
```

We see that both figures are pretty close each other. Analitical mean is equal to 5.0119113 and theoritical mean is equal to 5. It is true according to Central Limit Theory, because mean of analitical and theoritical distribution should be equal. Let's see it on the plot below. The red line is theoritical mean.

```
hist(mean_dist,breaks=50, col="yellow2")
abline(v=mean2, col="red", lwd=3)
```



2. Standard deviation and variance

Now let's focus on standard deviation. Theoritical formula for exponential variable standard deviaton in also 1/lambda. Accoriding to Central Limit Theorem, sample standard deviaton converges to theoritical standard deviation divided by square root of sample size (40). In our data sample standard deviation is equal:

```
sd1<-sd(mean_dist)
```

By comparision with theoritical value:

```
sd2<-(1/0.2)/sqrt(40)
```

We can observe, tha difference (theoritical 0.7905694 vs analitical 0.7749147) is not so big. Of course we can increase number of simulation, and then we should observe closer values each other. Look also at variance - it is also pretty close (first analitical, second theoritical).

```
sd1^2
```

```
## [1] 0.6004928
```

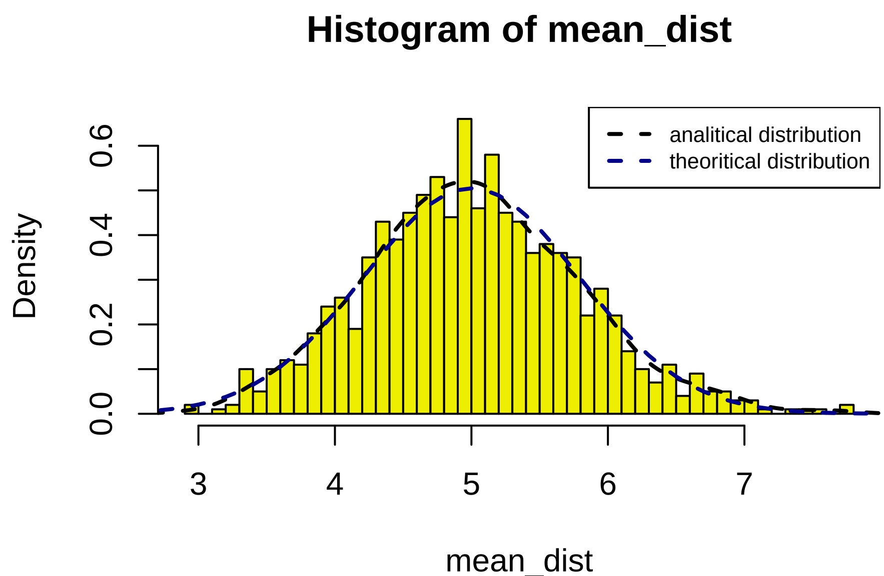
```
sd2^2
```

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

3. Density

Finally, let's look on theoritical density in comparision to analitical density.

```r
x<-seq(1,9,by=0.1)
y<-dnorm(x,mean=mean2,sd=sd2)
hist(mean_dist,breaks=50, col="yellow2", probability=TRUE)
lines(density(mean_dist),lty=2, lwd=2)
lines(x,y,lty=2, col="blue4", lwd=2)
legend("topright", legend=c("analitical distribution","theoritical distribution"),
       col=c("black","blue4"), lty=2,lwd=2, cex=0.7)
```



Looking at plot below, we can see that both density line are very close to each other. We can conlude that in this sample Central Theorem Limit works very good. Please remember, theoritical distribution, according to CTL, is normal. So even if sample hasn't normal distribution, if we have enough big number of simulation, we can construct normal distribution.