Exponential distribution data simulation and analysis

Yilin

February 12, 2016

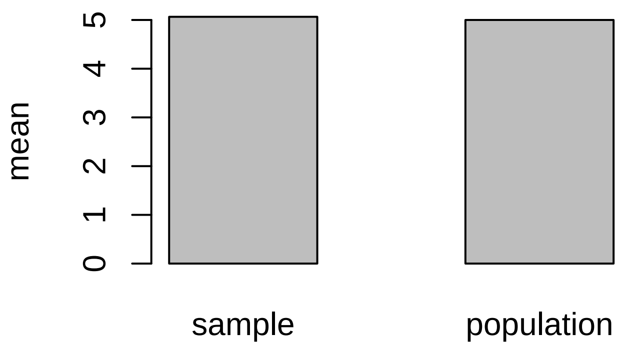
Overview: Simulate 40*1000 exponential distribution data with lambda=0.2, calculate the mean and standard deviation of the 40 exponennials' mean. Compare the result with the theoretical result, mean=1/lambda, sd=1/lambda.

1.1000*40 simulations of exponential distribution data

```
data<-rexp(n=1000*40,rate=0.2)
```

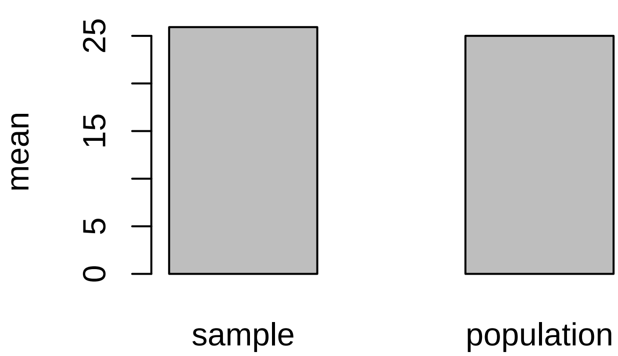
2.Compare the sample mean with the theoretical mean of the population

```
sample_mean=mean(data); population_mean=5
par(mar=c(8,10,8,10))
barplot(height=c(sample=sample_mean,population=population_mean),width=0.5,space=1,
        ylab="mean",axes=TRUE)
```



Sample mean is mean=5.00 and theoretical mean of the distribution is 1/lambda=5, they are identical. This example demonstrate the central limit theorem, the sample mean approximates the population mean if the sample size is large
3.Compare the sample variance with the theoretical variance of the population

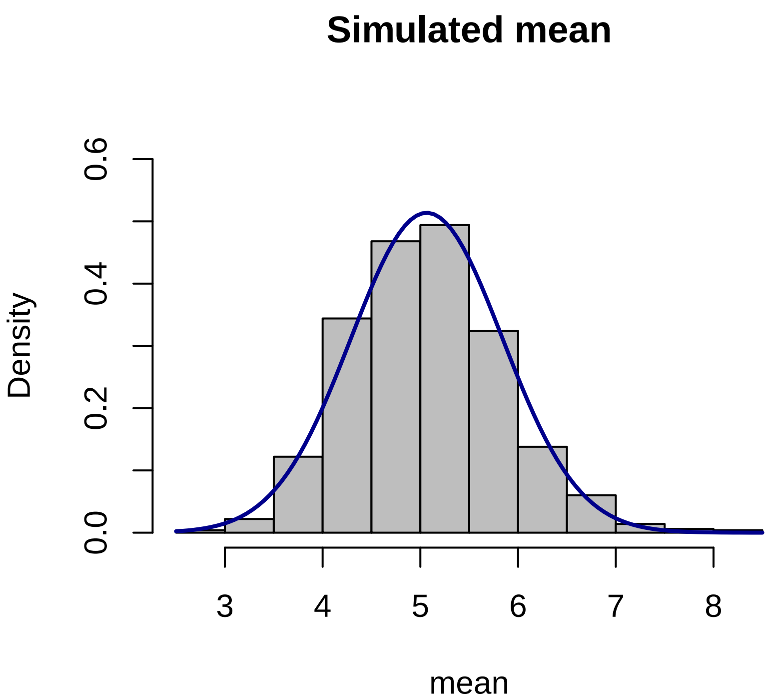
```
sample_var=var(data); population_var=25
par(mar=c(8,10,8,10))
barplot(height=c(sample=sample_var,population=population_var),width=0.5,space=1,
       ylab="mean",axes=TRUE)
```



Variance of the sample is var=25.06, theoretical variance of the distribution is 1/(lambda)^2=25, they are also identical. The central limit theorem is approved again that the sample variance approximates the population variance if the sampel size is large

4.Histogram of simulated mean

```
data_n<-matrix(data, nrow=1000, ncol=40)
meandata<-apply(data_n,1,mean)
m<-mean(meandata); std=sd(meandata)
par(mar=c(6,8,6,8))
hist(meandata,main="Simulated mean",ylim=c(0,0.6),xlab="mean",col="grey",prob=TRUE)
curve(dnorm(x,mean=m,sd=std),col="darkblue",lwd=2,add=TRUE)
```



The distribution of the simulated mean is approximately normal. This example demonstrates that sample mean of non Gaussian distributed data has Gaussian distribution.