StatInf_Simulation

wdewit

September 11, 2016

Overview

In this report we will investigate how closely the distribution of the mean of 40 random variables, drawn from an exponential distribution with lamba=0.2, approximates the normal distribution with mean=5 and standard deviation=0.79 as stated by the Central Limit Theorem. In order to obtain a distribution of the mean we will repeat the drawing of the 40 random variables a thousand times.

Simulations

We first draw 40000 observations from an exponential distribution with lamba=0.2 and store them in a matrix of 1000 rows and 40 columns.

```
simul <- 1000
n <- 40
lambda <- 0.2
set.seed(3)
rdraws <- rexp(n*simul,lambda)
rmatrix <- matrix(rdraws, nrow=simul, ncol=n)
```

Sample mean versus theoretical mean

For each of the 1000 samples of 40 observations from the exponential distribution we calulated their average value. The average of these 1000 sample means (dotted orange line), 4.99 is very close to the theoretical mean which is 1/lambda, or 5 (solid red line).

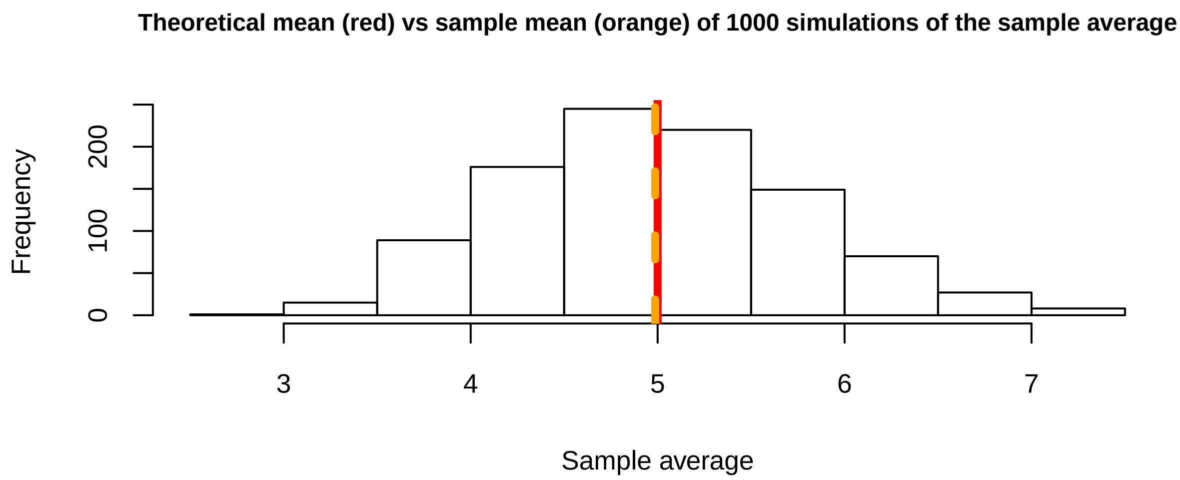
```
smeans <- apply(rmatrix,1,mean)
smean <- mean(smeans)
theormean <- 1/lambda
print(c('The sample mean is: ',round(smean,2),'.'))
```

```
## [1] "The sample mean is: " "4.99"                 "."
```

```
print(c('The theoretical mean is: ',theormean,'.'))
```

```
## [1] "The theoretical mean is: " "5"
## [3] "."
```

```
par(cex.main=0.75, cex.lab=0.80, cex.axis=0.80)
hist(smeans,xlab="Sample average",
     main="Theoretical mean (red) vs sample mean (orange) of 1000 simulations of the sample average")
abline(v=theormean, lw=4, col="red")
abline(v=smean, lw=4, col="orange", lty=2)
```



Sample variance versus theoretical variance

We know that the variance of the sample mean is the population variance divided by n. As the variance from an exponential distribution=(1/lambda)^2, the theoretical variance of the sample mean is (1/0.2)^2/40=0.625. The variance from our simulated sample means is very near to this theoretical number: 0.63.

```
svar <- var(smeans)
theorvar <- (1/lambda)^2/n
print(c('The sample variance is: ',round(svar,2),'.'))
```

```
## [1] "The sample variance is: " "0.63"
## [3] "."
```

```
print(c('The theoretical variance is: ',theorvar,'.'))
```

```
## [1] "The theoretical variance is: " "0.625"
## [3] "."
```

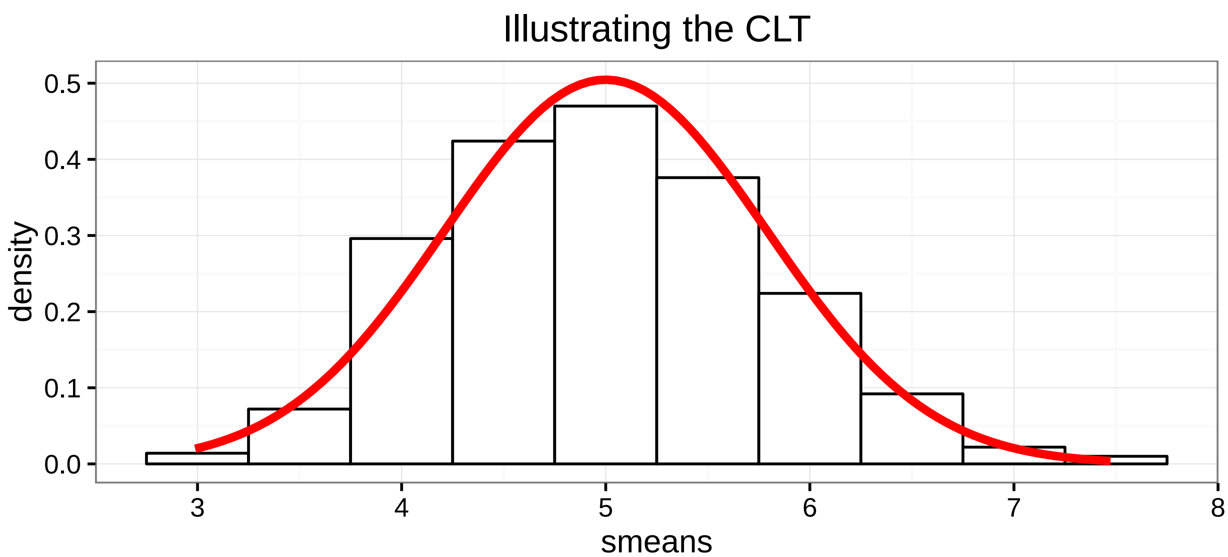
Distribution approximately normal ?

The Central Limit Theorem states that when you draw a large sample of observations from the same distribution with replacement (iid), their sample mean follows a normal distribution with the mean=population mean and SD=population SD/square root(sample size).

So, in our case the distribution from our 1000 sample means should closely follow a N(mean=5,var=0.625). In the plot below you notice that our empirical density does closely follows a N(mean=5,var=0.625)(red line).

```
library(ggplot2)
smeansdf <- as.data.frame(smeans)
g <- ggplot(data=smeansdf,aes(x = smeans))+ labs(title="Illustrating the CLT")
g <- g + geom_histogram(aes(y = ..density..), fill = "white", binwidth=0.5, colour = "black")
```

```
g + stat_function(fun=dnorm,colour="red",
                  size=1.5, args=list(mean=1/lambda, sd=(1/lambda)/sqrt(n)))+theme_bw()
```



Note that a sample of 1000 random variables drawn from our exp(0.2) distribution is far from being normal.

```
rexp1000 <- rexp(n=1000,lambda)
rexp1000df <- as.data.frame(rexp1000)
g <- ggplot(data=rexp1000df,aes(x = rexp1000))+labs(title="Distribution of 1000 draws from exp(0.2)")
g <- g + geom_histogram(aes(y = ..density..), fill = "white", binwidth=0.5, colour = "black")
g + geom_density(size=1.5, colour="red")+theme_bw()
```