Statistical Inference Course Project

Mohammad Ali Mahmoudzadeh

November 17, 2015

1 Overview

This report examines the central limit theorm on exponential distribution. The distribution on averages of 40 exponentials will be ploted and their closeness to a normal distribution will be discussed.

2 simulation

The values of mean and varaince of samples are stored in columns of a data frame `sampleData`. Theoretical values of mean and variance are stored in `expMean` `expVar`. After 1000 sampling, the mean of these samples are measured and ploted.

```
library(ggplot2)
lambda<-0.2
n<-40
noSim<-1000
set.seed(100)
samplesData=data.frame(mean=numeric(),var=numeric())
expMean<-1/lambda
expVar<-(1/lambda)^2
for (i in 1 : noSim) {
  samples<-rexp(n,lambda)
  samplesData <- rbind(samplesData, data.frame(mean=mean(samples),var=var(samples)))
}
```

3 Sample Mean versus Theoretical Mean

The following R code generate Figure 1 which shows the distribution of mean outcomes. The black vertical line presents the theoretical mean of the exponential distribution and the dashed red line is that of the sampled data. The mean value of sample means is 4.9997019 and the theoretical mean is 5.

```
g<-NULL
g <- ggplot(samplesData, aes(x = mean))
g <- g + geom_histogram(binwidth=.3,fill = "lightblue", colour = "black", aes(y = ..density..))
g <- g + geom_vline(xintercept = expMean, size = 1)
g <- g + geom_vline(aes(xintercept=mean(samplesData$mean)),color="red", linetype="dashed", size=1)
g <- g + ggtitle("Distributions of means of sampled data")
g
```

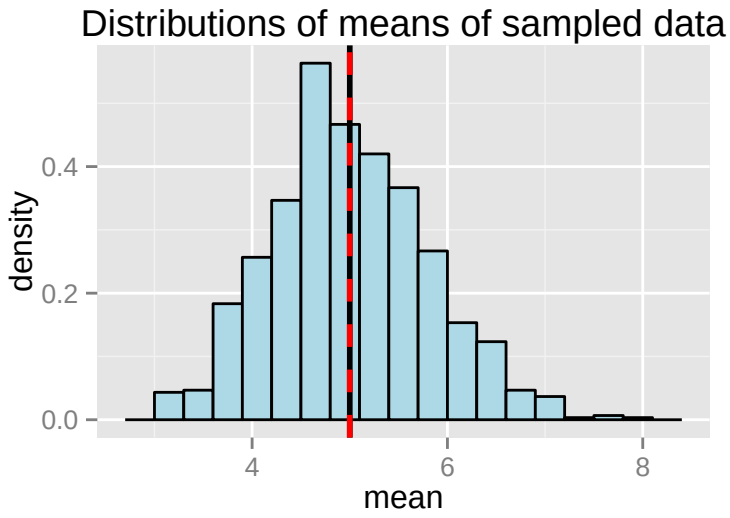

Figure 1: Distributions of means of sampled data. The vertical line represents the theoretical mean value.

4 Sample Variance versus Theoretical Variance

The next figure shows the distribution of variance for each batch of sample outcomes. The vertical black line represents the theoretical variance of this exponential distribution and the dashed red line is that of the samples means. The mean value of sample variances is 25.3728703 and the theoretical mean is 25.

```
g <- ggplot(samplesData, aes(x = var))
g <-g + geom_histogram(binwidth=.3,fill = "lightblue", aes(y = ..density..))
g <-g + geom_vline(xintercept = expVar, size = 1)
g <-g + geom_vline(aes(xintercept=mean(samplesData$var)),color="red", linetype="dashed", size=1)
g <-g+ggtitle("Distributions of variances of sampled data")
g
```

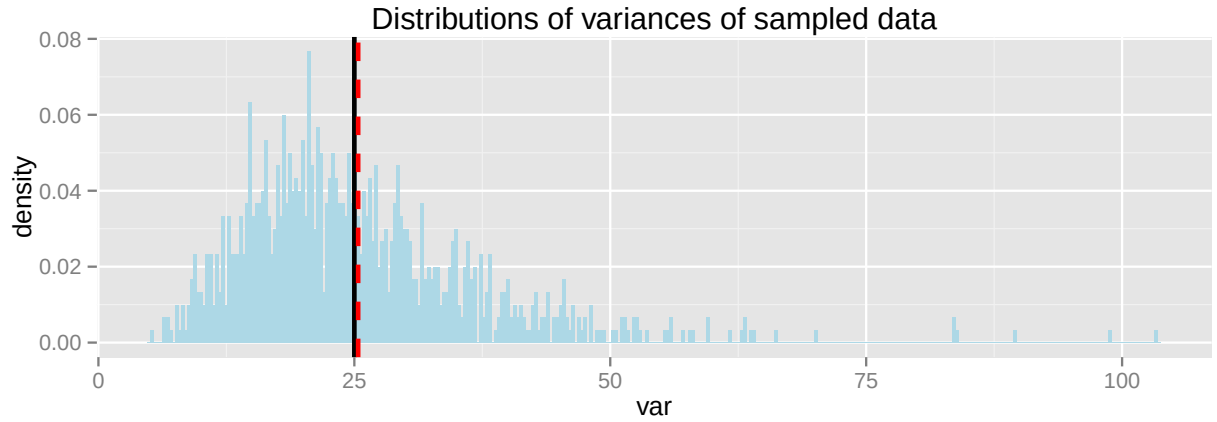

Figure 2: Distributions of variances of sampled data. The vertical black line represents the theoretical variance value.

5 Distribution

A plot of distribution of mean of samples show that the means follow the normal distribution. Figure 3 shows a maping of a normal distribution centered at the theoretical mean and with a standard deviation of $\sigma/\sqrt{(n)}$ on the distribution of means of sampled data. A very close match is obsereved between the two sets of data which is translatable to a normal distribution for the mean values i.e. the Central Limit Theorem.

```
library(ggplot2)
g <- ggplot(samplesData, aes(x = mean))
g <- g + geom_histogram(binwidth=.3,color="black",fill = "lightblue", aes(y = ..density..))
g <- g + stat_function(fun = dnorm, args = c(mean = expMean, sd = sqrt(expVar/40)), colour = "red",size=
g <- g + ggtitle("Distributions of means of sampled data")
g
```

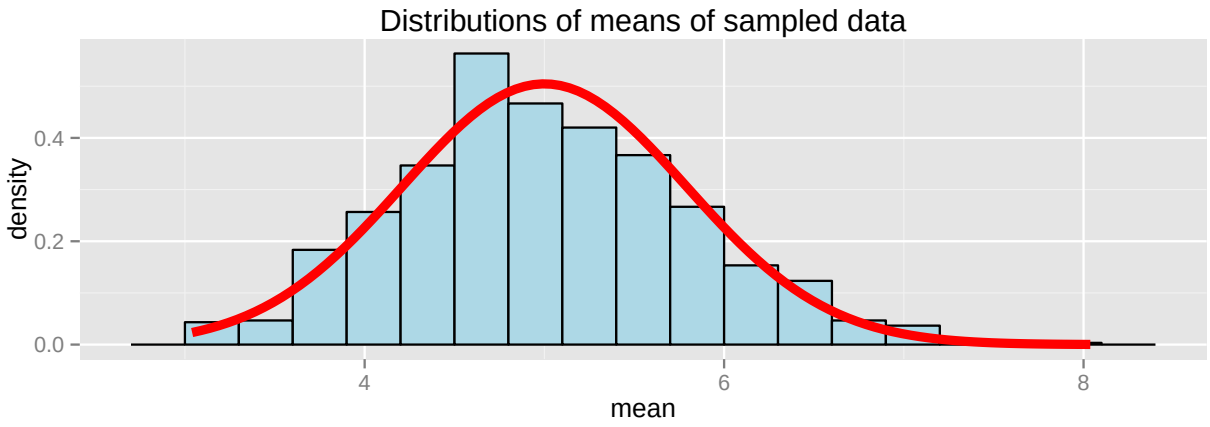

Figure 3: Distributions of means of sampled data compared to normal distribution.