p1a

george soilis

8 June 2015

Introduction

In this report we investigate the exponential distribution and compare it with the Central Limit Theorem. Both the mean and standard deviation of the exponential distribution are 1/lambda. We are using a lambda value of 0.2 and we investigate the properties of the distribution of averages of 40 exponentials by 1000 simulations.

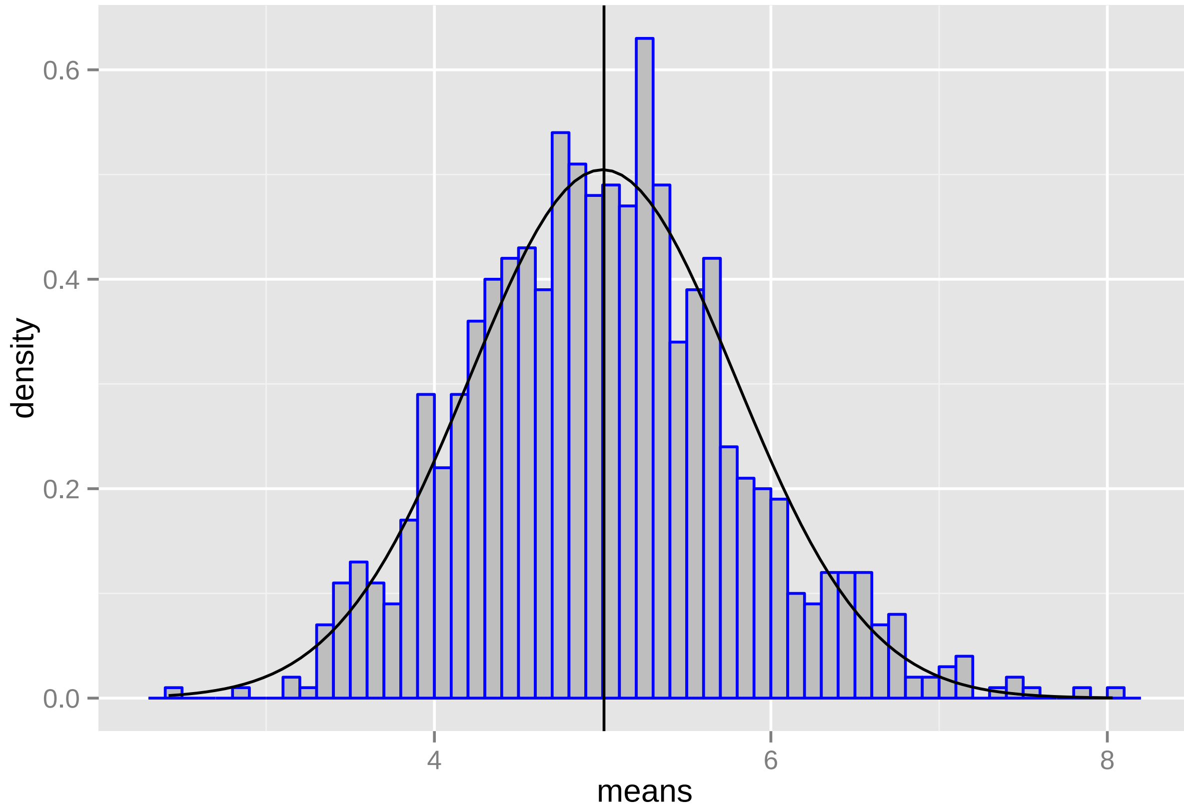
We first show the mean and compare it to the theoretical mean of the distribution answering the question :

1. Show the sample mean and compare it to the theoretical mean of the distribution.

Generating the data

```
library(ggplot2)

set.seed(215)
lambda <- 0.2
expectedMean <- expectedSd <- 1/lambda
numSamples <- 40
numSims <- 1000
sampleData <- matrix(rexp(numSamples * numSims, lambda), ncol = numSamples)
means <- apply(sampleData, 1, mean)
variances <- apply(sampleData, 1, var)
meanOfMeans <- mean(means)
df <- as.data.frame(means)
plot <- ggplot(df, aes(x=means))
plot <- plot + geom_histogram(aes(y=..density..), colour="blue", fill="grey", binwidth=0.1)
plot <- plot + stat_function(fun=dnorm, args=list(mean=expectedMean, sd=expectedSd/sqrt(40)))
plot <- plot + geom_vline(xintercept = meanOfMeans)
print(plot)
```



Mean of the means

The theoretical mean is 1/lamda = 5 , and the sample means is

```
mean(means)
```

```
## [1] 5.008104
```

which is indeed close (as shown on our histogram by the overlayed grey line.

Proceeding in answering the next question:

2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

Variance of the means

The theoretical variance is lambdaˆ-2/n = 25/40 = .625, and the practical is

```
var(means)
```

```
## [1] 0.6400734
```

which is pretty close, too.

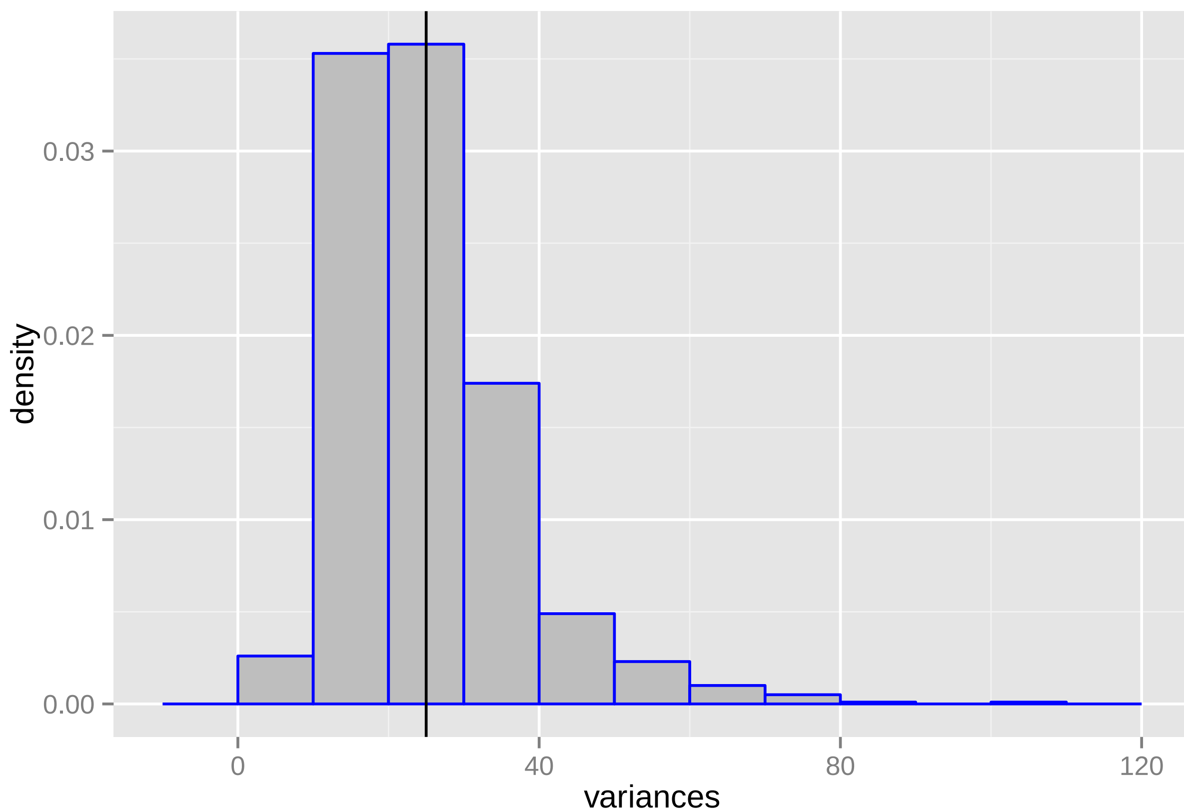
```
print(paste("Our sample variance = ", var(means)))
```

```
## [1] "Our sample variance =  0.640073396461169"
```

```
print(paste("Our expected sample variance =", 25/40))
```

```
## [1] "Our expected sample variance = 0.625"
```

```
df <- as.data.frame(variances)
plot <- ggplot(df, aes(x=variances))
plot <- plot + geom_histogram(aes(y=..density..), colour="blue", fill="grey", binwidth=10)
plot <- plot + geom_vline(xintercept = (1/lambda)^2)
print(plot)
```



And for the last question :

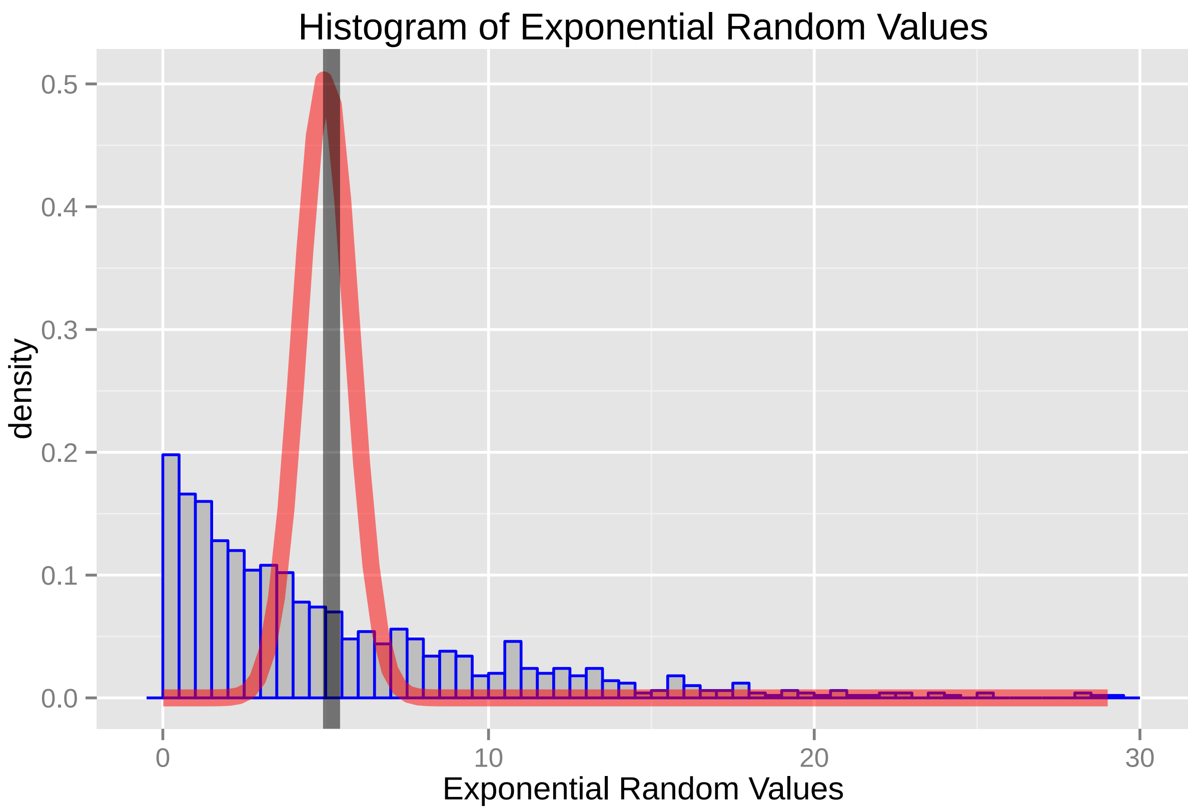
3. Show that the distribution is approximately normal.

Proximity to the normal distribution

We first examine the distribution of a large collection of random exponentials

Plot random values from an exponential distribution

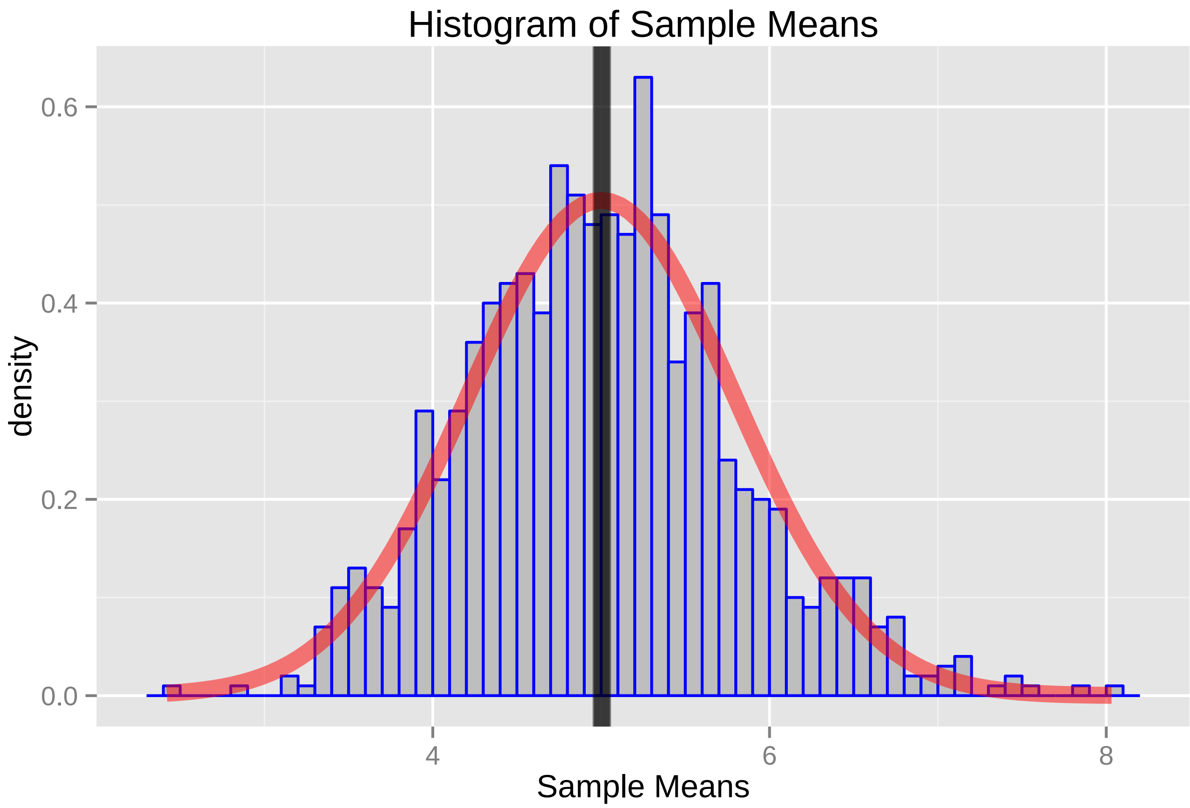
```
df <- data.frame(val=rexp(numSims, lambda))
g <- ggplot(df, aes(x=val)) +
    geom_histogram(aes(y=..density..), binwidth=0.5, color="blue", fill="grey")
g + labs(x="Exponential Random Values",
         title="Histogram of Exponential Random Values") +
  stat_function(fun=dnorm,args=list(mean=1/lambda, sd=1/(lambda*sqrt(numSamples))),
                color="red", size=3, alpha=0.5) +
    geom_vline(xintercept=mean(df$val), size=3, alpha=0.5 )
```



We then examine the distribution of a large collection of averages of 40 exponentials

Plot density of sample means and compare to theoretical normal

```
df <- data.frame(means=means)
g <- ggplot(df, aes(x=means)) +
    geom_histogram(aes(y=..density..), binwidth=0.1, color="blue",fill="grey")
g + geom_vline(xintercept=1/lambda, size=3, alpha=0.5)+labs(x="Sample Means", title="Histogram of Sampl
                     color="red", size=3, alpha=0.5) +
    geom_vline(xintercept=mean(means), size=3, alpha=0.5)
```



We can clearly see how close the histogram follows the well known "Bell Curve" of the Normal distribution.