Analysis of the Exponential Distribution

Cleitus

Wednesday, January 14, 2015

1. Overview

In this paper we will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with `rexp(n, lambda)` where λ is the rate parameter. The mean of the exponential distribution, as well as its standard deviation, is $1/\lambda$. Let $\lambda = 0.2$ for all of the simulations. we will study the distribution of averages of 40 exponentials over 1000 simulations.

2. Simulations

Start by loading all the needed packages.

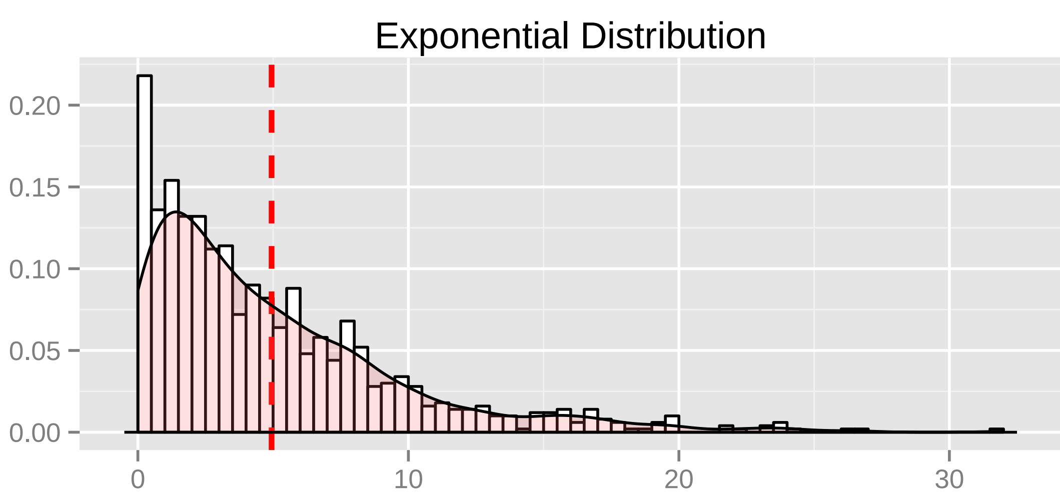
```
# Load necessary library
library(ggplot2)
```

To get a feel of the Exponential Distribution lets plot a histogram of a representative sample.
Like we already said, we're gonna assume that $\lambda = 0.2$.

```
# For reproducibility
set.seed(1729)

lambda = 0.2
df = data.frame(expo = rexp(1000, lambda))

ggplot(df, aes(x=expo)) + labs(x = "", y = "") + ggtitle("Exponential Distribution") +
  geom_histogram(aes(y=..density..), binwidth=.5, colour="black", fill="white") +
  geom_vline(aes(xintercept=mean(expo)), color="red", linetype="dashed", size=1) +
  geom_density(alpha=.2, fill="#FF6666")
```



Now to investigate the distribution of 1000 averages of 40 exponentials.

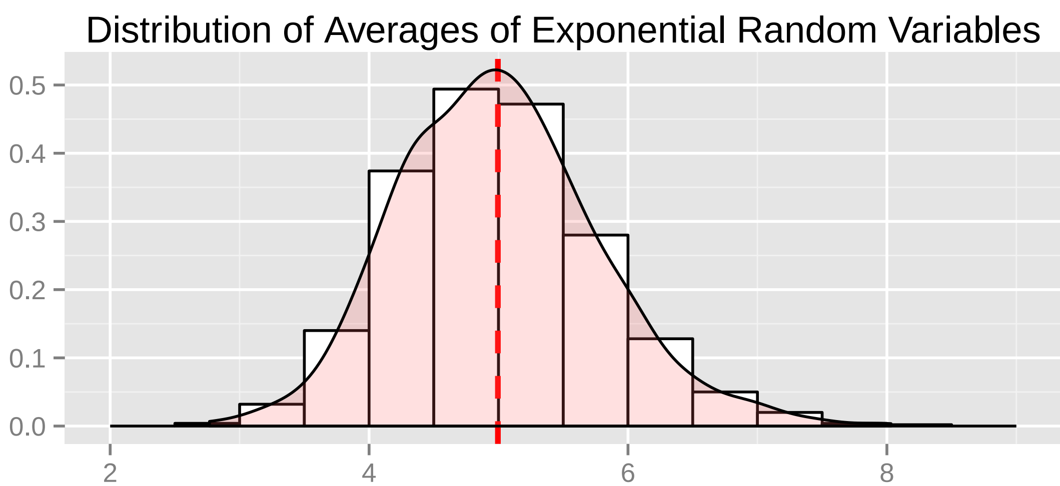
```
df = data.frame(expmean = mean(rexp(40, lambda)))
for (i in 2 : 1000) df <- rbind(df, mean(rexp(40, lambda)))
```

`df` here is a dataframe with one column `expmean` containing 1000 observations of average values of 40 random exponentials.

3. Sample Mean VS Theoretical Mean

Now we plot the distribution of those averages.

```
ggplot(df, aes(x=expmean)) + labs(x = "", y = "") +
  ggtitle("Distribution of Averages of Exponential Random Variables") +
  geom_histogram(aes(y=..density..), binwidth=.5, colour="black", fill="white") +
  geom_vline(aes(xintercept=mean(expmean)), color="red", linetype="dashed", size=1) +
  geom_density(alpha=.2, fill="#FF6666")
```




```
mean(df$expmean)
```

```
## [1] 4.994989
```

The Theoretical mean being $1/\lambda = 5$, we conclude that the Sample mean is equal to the Theoretical mean.

4. Sample Variance VS Theoretical Variance

The Sample variance is :

```
var(df$expmean)
```

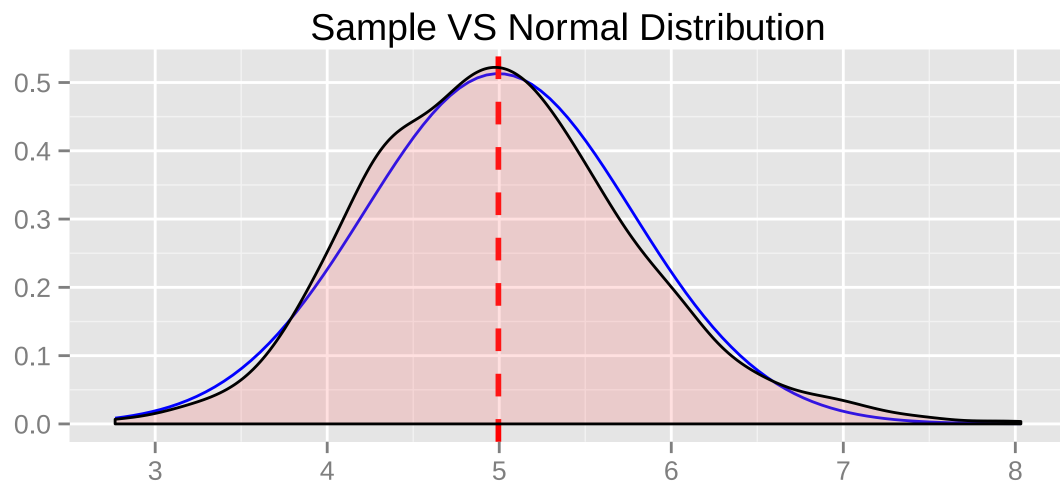
```
## [1] 0.604644
```

The Theoretical variance is $1/\lambda^2 \times 1/40 = 0.625$. And so in turn, the Sample variance closely fits its Theoretical value.

5. Distribution

Now we show that the distribution is approximately normal.
There are many ways to do that, first let's plot the sample curve against the theoretical normal curve with respective mean and standard deviation.

```
ggplot(df, aes(x=expmean)) + labs(x='', y='') + ggtitle("Sample VS Normal Distribution") +
  stat_function(fun=dnorm,args=list(mean=mean(df$expmean),sd=sd(df$expmean)),color="blue") +
  geom_vline(aes(xintercept=mean(expmean)), color="red", linetype="dashed", size=1) +
  geom_density(alpha=.2, fill="#FF6666")
```



The blue line represents the theoretical normal distribution. Then we can visually confirm that the sample distribution is roughly normal.
Furthermore, we can check with a qqplot the validity of our assumption.

```
qplot(sample = df$expmean, stat="qq")
```