Statistical Inference Course Project 1

RossAW

27 September 2015

Overview

This project consists of two parts, a simulation exercise and a basic inferential data analysis.

```
library(ggplot2)
```

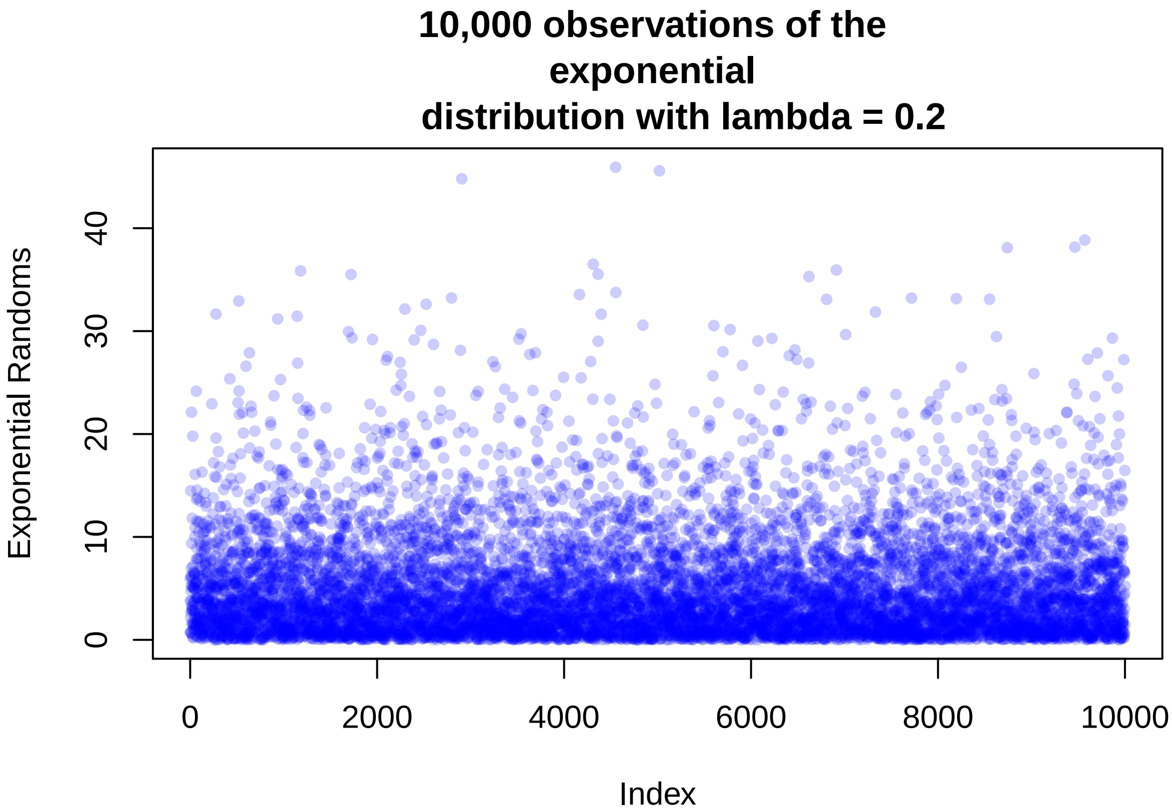
Simulations

The Exponential Distribution

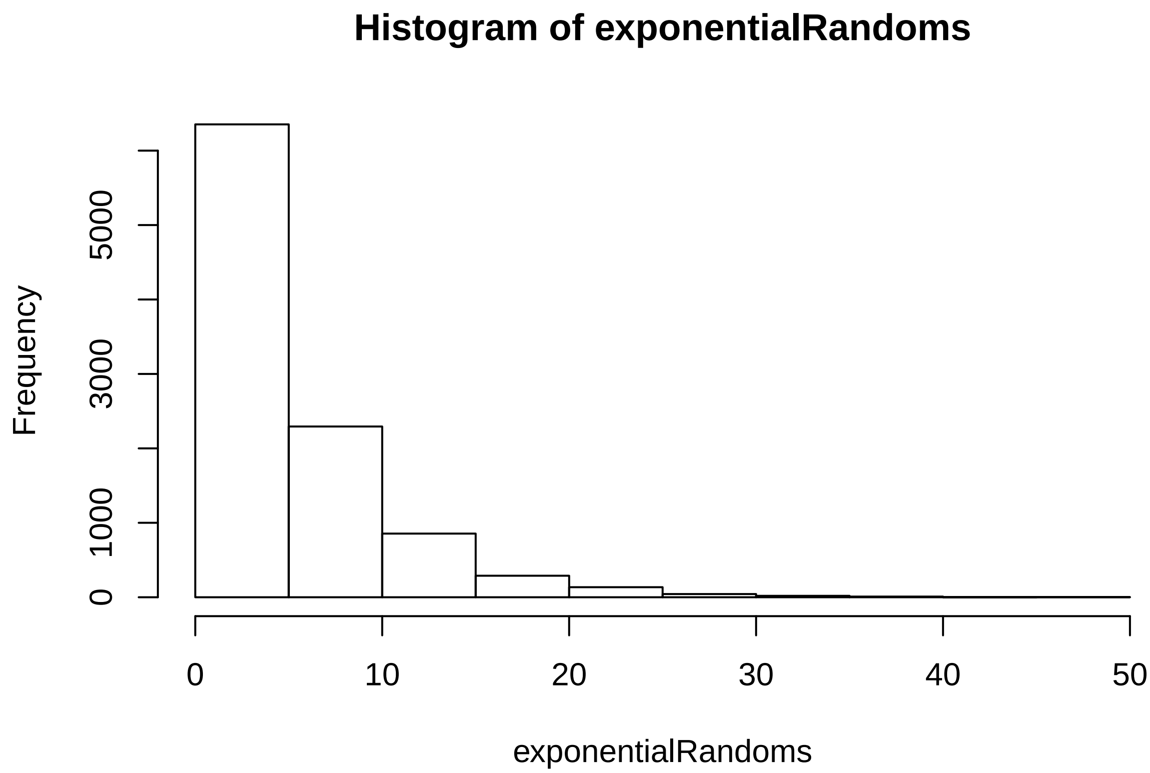
The Exponential Distribution is given by

Lets have a look at what that looks like.

```
set.seed(1)
lambda <- 0.2

NumberOfSims <- 10000

exponentialRandoms<-rexp(NumberOfSims,lambda)

plot(exponentialRandoms, pch=20, col=rgb(0,0,1,0.2),
     main = "10,000 observations of the \nexponential
     distribution with lambda = 0.2",
     ylab= "Exponential Randoms")
```




```
hist(exponentialRandoms)
```



To answer the remaining questions we need to take the average of 40 random exponentials, 1000 times. This will give us 1000 averages of exponential Randoms.

```
n <- 40 # Average of n random exponentials
AveragOfExponentialRandoms<-NULL

for(i in 1:1000){

    AveragOfExponentialRandoms=c(AveragOfExponentialRandoms,mean(rexp(n,0.2)))
}
```

Sample Mean vs Theoretical Mean

This compares the mean of the sample means to the theoretical mean.

```
TheoreticalMean<-1/lambda
TheoreticalMean
```

```
## [1] 5
```

```
SampleMean<-mean(AveragOfExponentialRandoms)
SampleMean
```

```
## [1] 5.025866
```

As predicted by the CLT the mean of the sample means is near identical to the population mean.

Sample Variance vs Theoretical Variance

This compares the sample varianace to the theoretical variance.

```
SampleVariance<-var(AveragOfExponentialRandoms)
SampleVariance
```

```
## [1] 0.6374065
```

```
TheoreticalVariance<-(1/lambda^2)/n
TheoreticalVariance
```

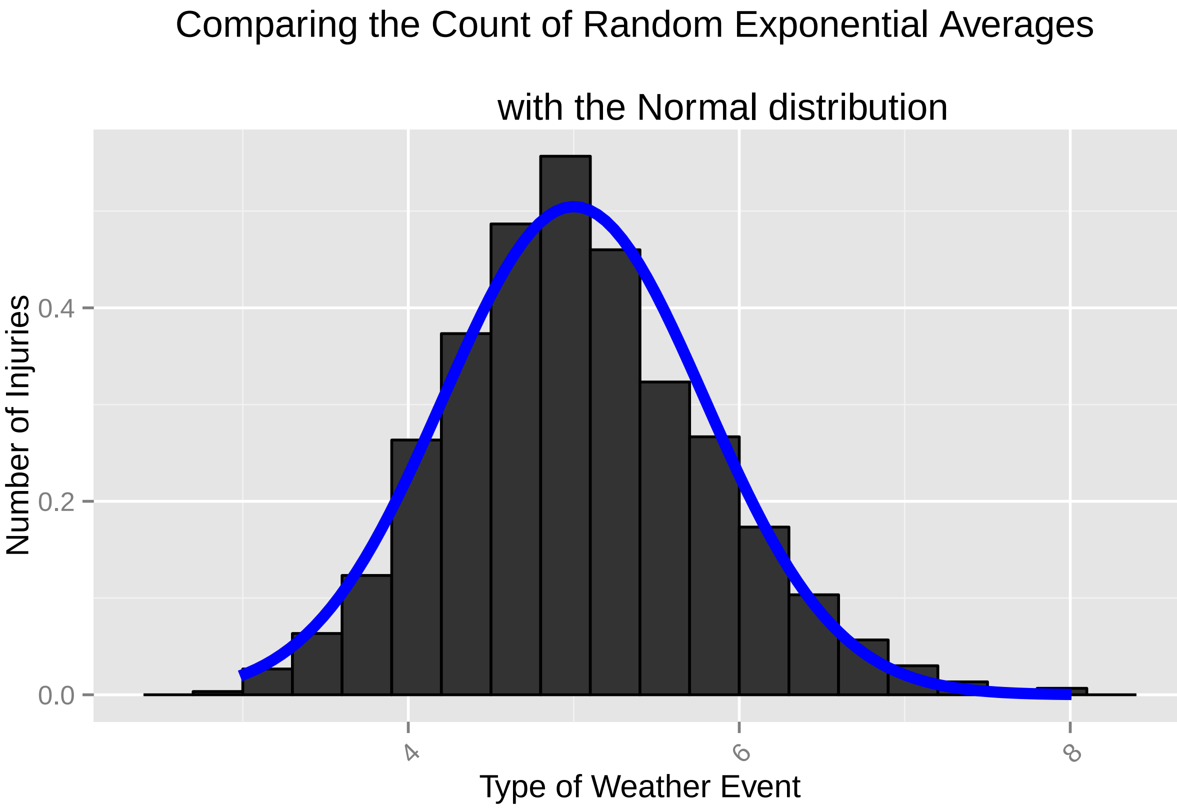
```
## [1] 0.625
```

As predicted by the CLT the sample varaiance is near identical to the population variance.

Distribution vs Normal Distribution

The CLT states that the average of a set of random samples is approximately Normally distributed with a mean given by the propluation mean and a variance given by the standard error of the mean N(mu,sig^2/n)

```
AveragOfExponentialRandoms<-data.frame(AveragOfExponentialRandoms)
TheoreticalSD<-sqrt(TheoreticalVariance)

g <- ggplot(AveragOfExponentialRandoms, aes(x = AveragOfExponentialRandoms)) +
    geom_histogram(binwidth=.3, colour = "black", aes(y = ..density..))

g <- g + stat_function(fun = dnorm, size = 2,colour="blue",
                       args = list(mean = TheoreticalMean, sd = TheoreticalSD))

g<-g+ylab("Number of Injuries")+
        xlab("Type of Weather Event")+
        theme(axis.text.x = element_text(angle = 45,hjust = 1))+
        ggtitle("Comparing the Count of Random Exponential Averages \n
                with the Normal distribution")
g
```



The above graph shows that the distribution is very well described by the normal distribution.