Statistical Inference Course Project- Exponential Distribution and Central Limit Theorem

Hizban Sheik

2022-09-19

Synopsis

The Aim of this work is to investigate the exponential distribution and compare it with **Central Limit Theorem** to see if it follows the theorem or not.

Simulation for Sampling Distriution of Exponential Distribution

```
nsim <- 1000
sample.dist <- vector(length=nsim)
nsamples <- 40
lambda <- 0.2
true.mean <- 1/lambda
true.variance <- (1/lambda)^2
variance <- vector(length=nsim)
for (i in 1:nsim){

  sample <- rexp(nsamples,rate=lambda)
  sample.dist[i] <- mean(sample)
  variance[i] <- var(sample)
}
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.1.3
```

```
sample.mean <- mean(sample.dist)
variance.mean <- mean(variance)
data <- data.frame(x=sample.dist)
g <- ggplot(data, aes(x = x)) + geom_histogram(binwidth=.3, colour = "black",
                                               aes(y = ..density..))
g <- g + geom_vline(xintercept=true.mean,color='red',size=2) +
geom_vline(xintercept=sample.mean,color='blue',size=2) +
ggtitle("Distribution of Sample Means") +
theme(plot.title = element_text(hjust = 0.5))
g
```

Blue Vertical Line shows the mean of the sample distribution and red line is the true mean of the exponential distribution i.e 1/lambda.The Sample Mean is a statistic trying to estimate true mean of distribution.The

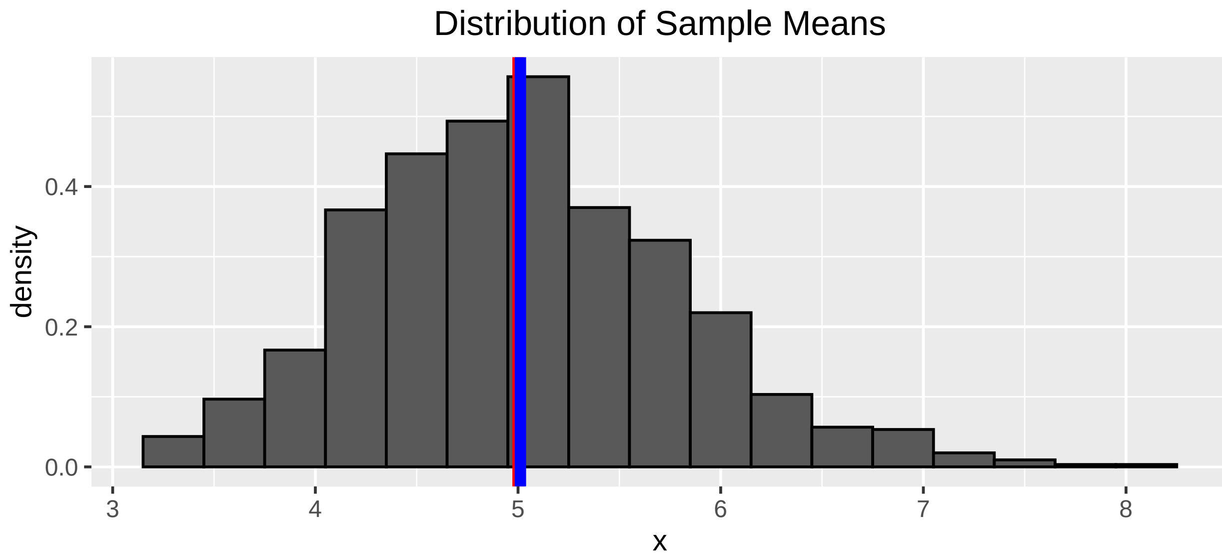

Figure 1: Density Plot for Distribution of Sample Means

Central Limit Theorem states that if you have a population with mean mu and standard deviation sigma and take sufficiently large random samples from the population with replacement , then the distribution of the sample means will be approximately normally distributed and its mean will be equal to true mean of distribution.From above simulation plot this theory is confirmed as the mean of sample means is almost equal to true mean.

```
print(paste0("The True Mean is ",true.mean))
```

```
## [1] "The True Mean is 5"
```

```
print(paste0("The Mean of Sample Means is ",sample.mean))
```

```
## [1] "The Mean of Sample Means is 5.01176147616864"
```

Variance of samples and theoretical variance

```
sample.number = (1:length(variance))
variance.data <- data.frame(var=variance,sample.number=sample.number)
ggplot(data=variance.data) + geom_line(aes(y=var,x=sample.number)) + geom_hline(yintercept=(1/lambda)^2
ggtitle("Variance of every sample")
```

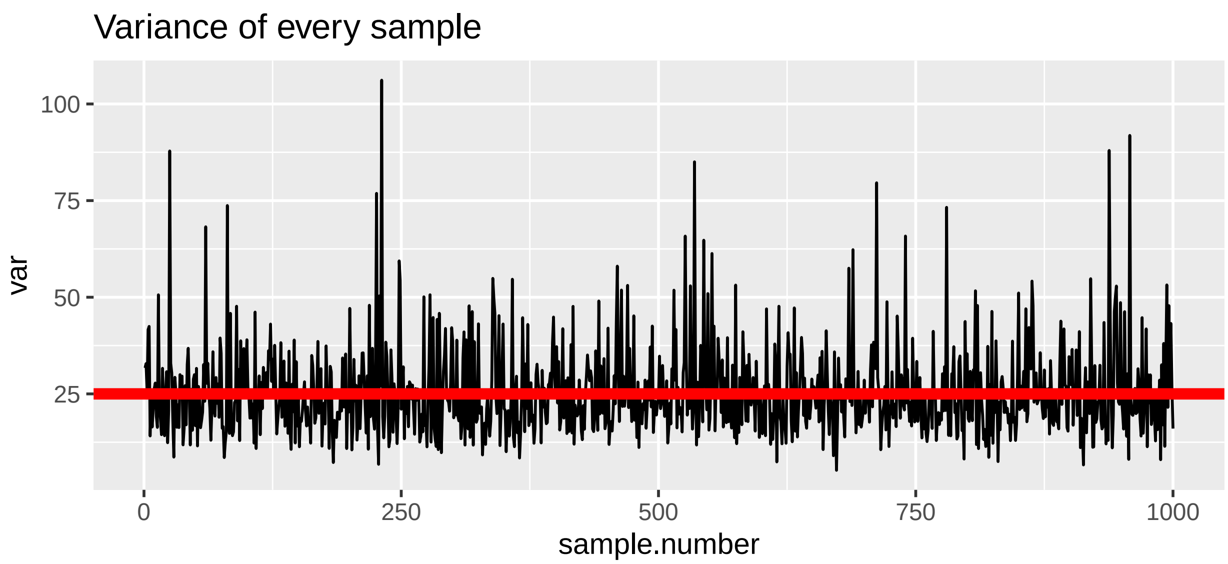

Figure 2: Plot for variances of every sample

Above Plot, shows the variances of all the thousand samples and red line is the theoretical variance of the distribution

```
print(paste0("The True Variance is ",true.variance))
```

```
## [1] "The True Variance is 25"
```

```
print(paste0("The Mean of Sample Variances is ",variance.mean))
```

```
## [1] "The Mean of Sample Variances is 25.4742688515032"
```

```
plot.data <- data.frame(x=variance)
g <- ggplot(plot.data, aes(x = x)) + geom_histogram(binwidth=3,
  colour = "black", aes(y = ..density..))
g <- g + geom_vline(xintercept=(1/lambda)^2,color='red',size=2) + geom_vline(xintercept=variance.mean,c
ggtitle("Distribution of Sample Variance") +
theme(plot.title = element_text(hjust = 0.5))
g
```

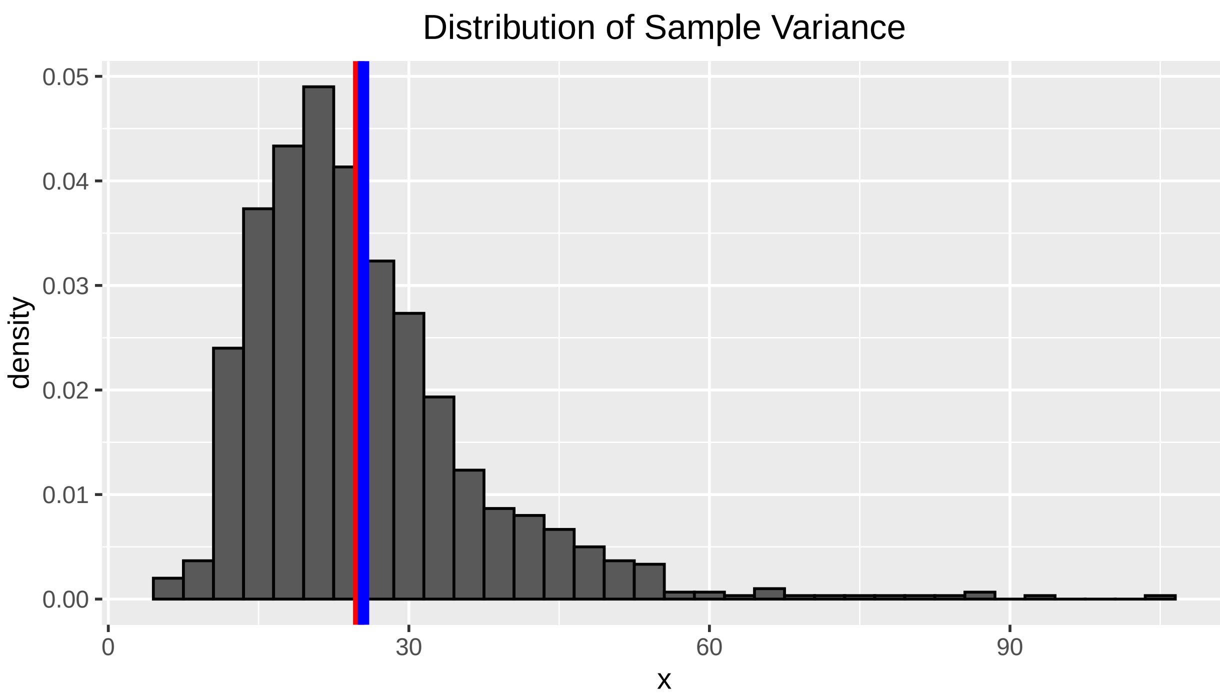

Figure 3: Density Plot for Variance of the samples

Blue Line Represents the Mean of Sample Variances and Red Line represents the True Variance of the distribution. As for mean sample means was estimating true mean of distribution for variance the sample variance will estimate the true variance of distribution. So again the mean of sample variances is almost equal to true variance of the distribution so Central Limit Theorem is again correct in this case as well

Comparing Normal Distribution and Sampling Distribution of Mean

```
exp.dist <- rexp(nsim,rate=lambda)
exp.dist.data <- data.frame(x=exp.dist)
ggplot(exp.dist.data, aes(x = x)) + geom_histogram(binwidth=2, colour = "black",
aes(y = ..density..)) + ggtitle("Exponential Distribution") +
theme(plot.title = element_text(hjust = 0.5))
```

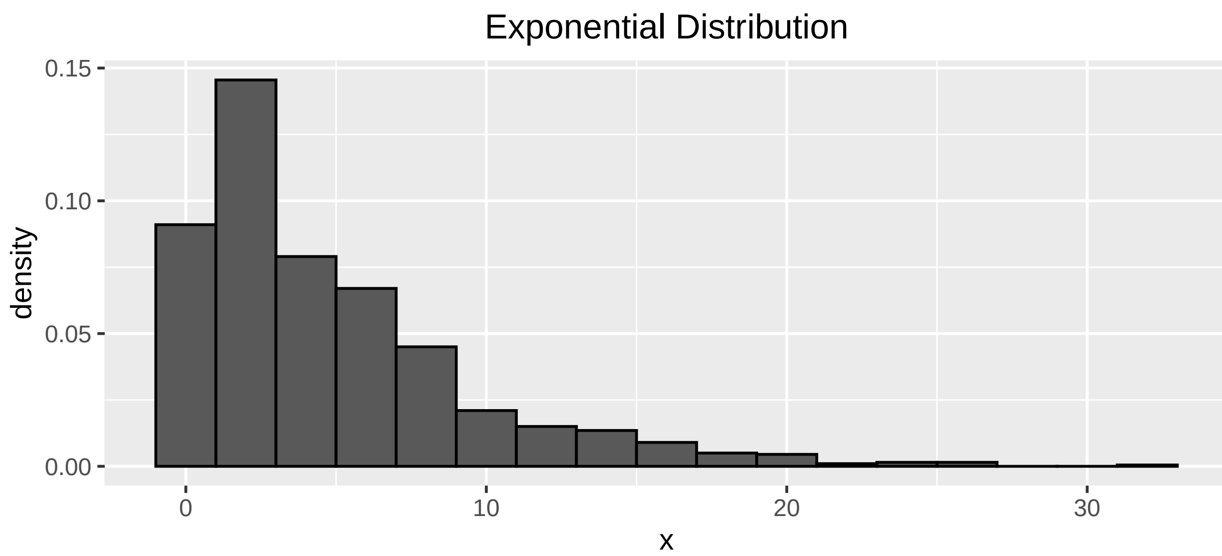

Figure 4: Desnity Plot for Exponential Distribution

```
g <- ggplot(data, aes(x = x)) + geom_histogram(binwidth=.3, colour = "black",
aes(y = ..density..))
g <- g + stat_function(fun=dnorm,args=list(mean=5,sd=5/(nsamples)^0.5),size=2)+
ggtitle("Comparing Normal Distribution to Distribution of sample means") +
theme(plot.title = element_text(hjust = 0.5))
g
```

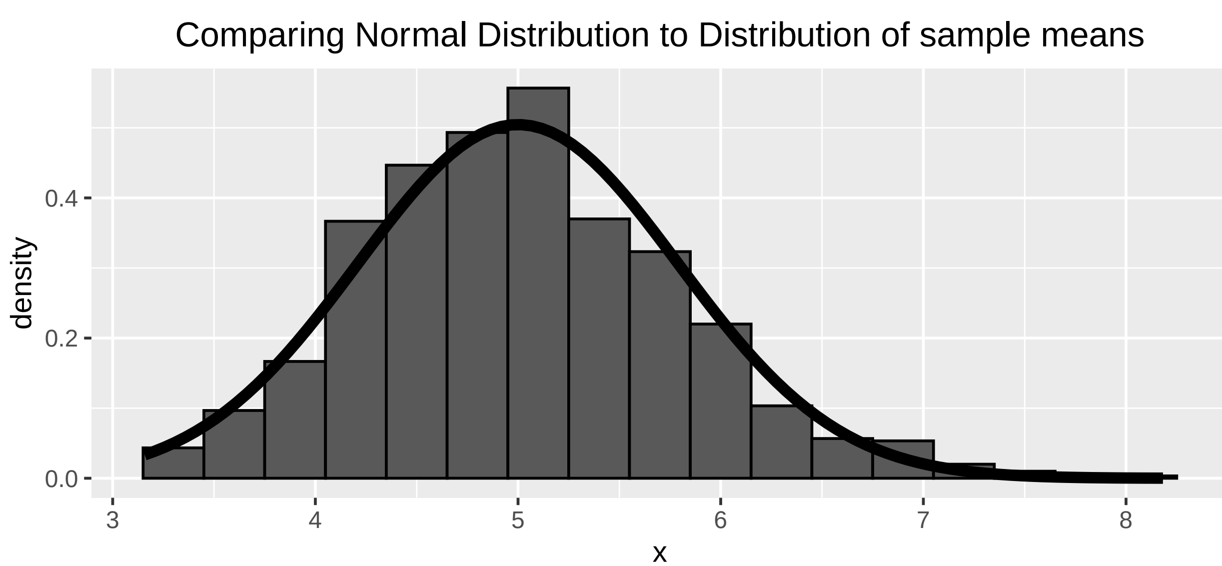

Figure 5: Distribution of Sample Means

The Distribution of Sample Means is obtained by sampling samples of size = 40 from the exponential distribution and taking averages of them.Central Limit Theorem states that the distribution of sample mean is normal with mean = population mean and Standard Deviation = Standard Error of Mean.

The above plot compares the normal distribution and sample mean distribution and it clearly visible that sample mean distribution is approximately normal,also if we increase the sample size then it will fit even better.