Statistical Inference Course Project 1

Avinash

October 25, 2015

Overview

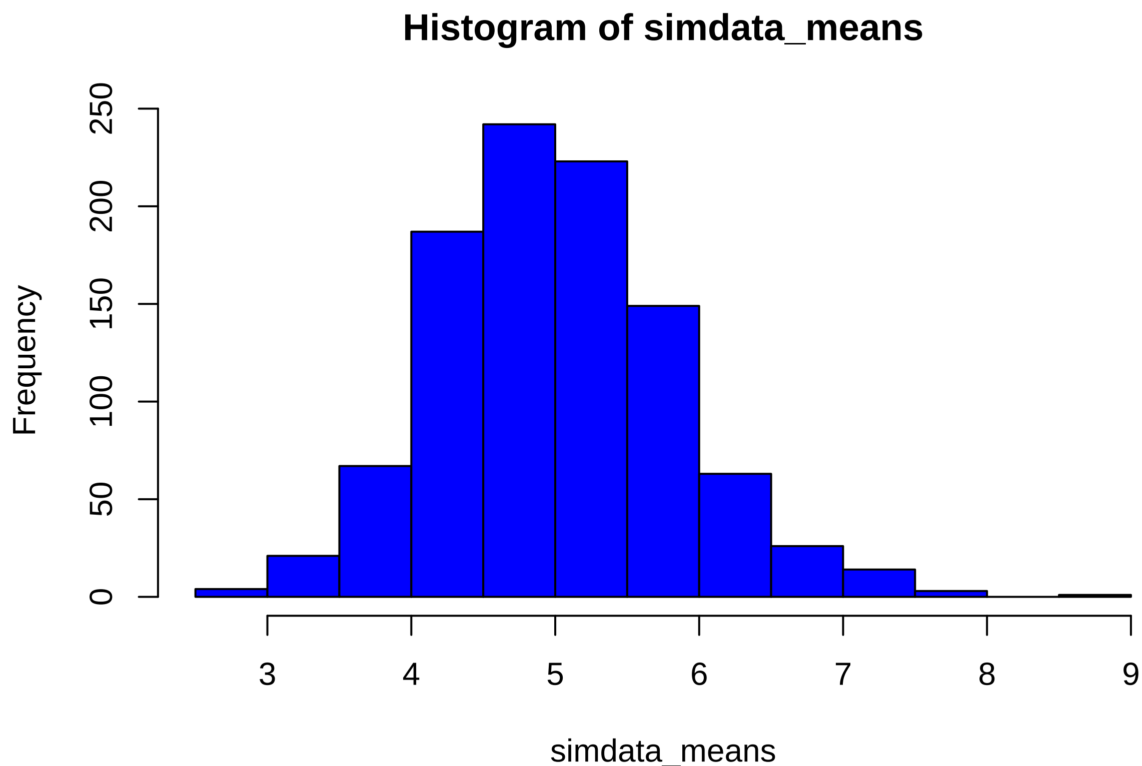
We investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution is simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. We set lambda = 0.2 for all of the simulation.We will investigate the distribution of averages of 40 exponentials over a thousand simulations.

Simulation

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.2.2
```

```
nosim <- 1000;
n <- 40;
lambda <- 0.2;
## create a matrix with 1000 simulations each with 40 rexp generated
## values
set.seed(234)
simdata <- matrix(rexp(nosim * n, rate=lambda), nosim);
## for each of the 1000 simulations calculate the average of across the
## 40 values
simdata_means <- apply(simdata, 1, mean);
hist(simdata_means, col="blue");
```



Sample Mean versus Theoretical Mean

The expected mean μ of a exponential distribution of rate λ is $\mu = \frac{1}{\lambda}$ Let $\bar{X}$ be the average sample mean of 1000 simulations of 40 randomly sampled exponential distributions. As you can see the expected mean and the avarage sample mean are very close

```
theoretical_mean <- 1/lambda;
print (paste("Theoretical center of the distribution = ", theoretical_mean));
```

```
## [1] "Theoretical center of the distribution =  5"
```

```
print (paste("Actual center of the distribution based on the simulations = ", round(mean(simdata_means)
```

```
## [1] "Actual center of the distribution based on the simulations =  5.002"
```

Sample Variance versus Theoretical Variance

The expected standard deviation σ of a exponential distribution of rate λ is $\sigma = \frac{1/\lambda}{\sqrt{n}}$ The e The variance Var of standard deviation σ is $Var = \sigma^2$ Let Var_x be the variance of the average sample mean of 1000 simulations of 40 randomly sampled exponential distribution, and σ_x the corresponding standard deviation. As you can see the standard deviations are very close Since variance is the square of the standard deviations, minor differnces will we enhanced, but are still pretty close.

```r
theoretical_var <- (1/lambda)^2/n;
theoretical_sd <- 1/lambda/sqrt(n);
print (paste("Theoretical variance = ", theoretical_var));
```

```
## [1] "Theoretical variance =  0.625"
```

```r
print (paste("Actual variance based on the simulations = ", round(var(simdata_means), 3)));
```

```
## [1] "Actual variance based on the simulations =  0.663"
```

```r
print (paste("Theoretical standard deviation = ", round(theoretical_sd, 3)));
```

```
## [1] "Theoretical standard deviation =  0.791"
```

```r
print (paste("Actual standard deviation based on the simulations = ", round(sd(simdata_means), 3)));
```

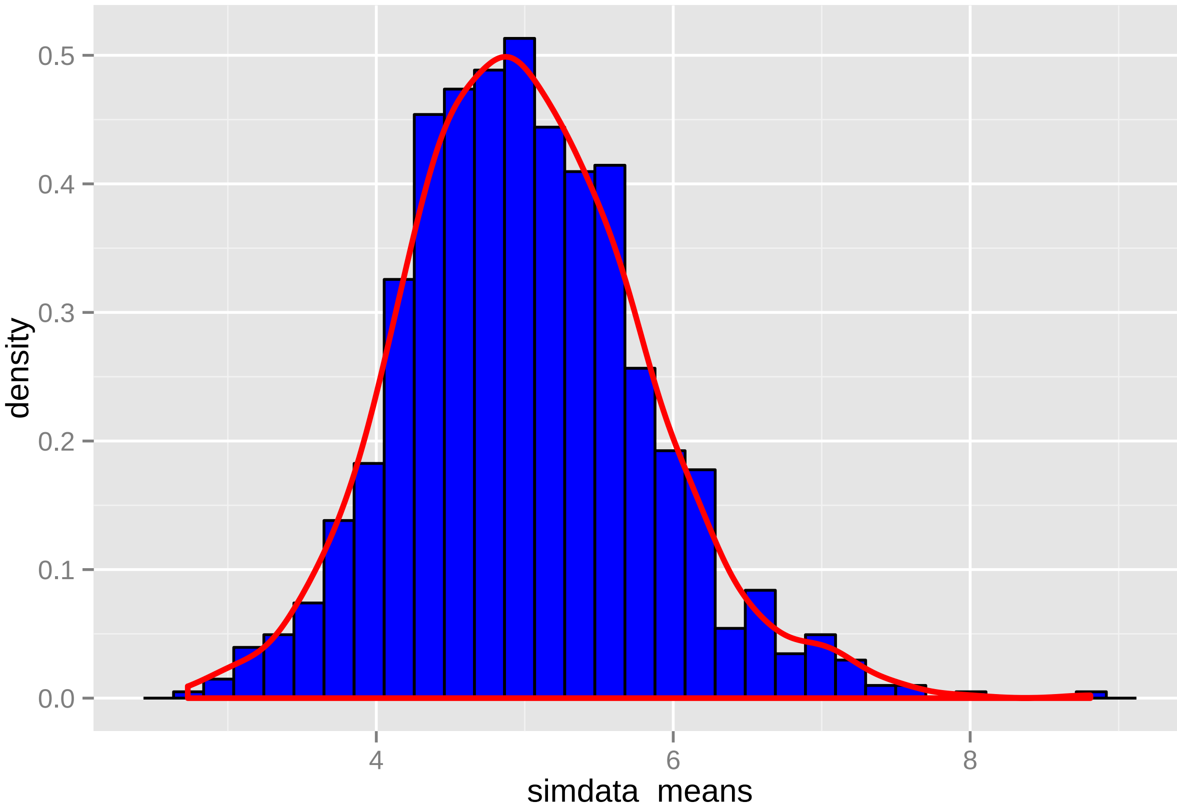
```
## [1] "Actual standard deviation based on the simulations =  0.814"
```

Distribution

Comparing the population means & standard deviation with a normal distribution of the expected values. Added lines for the calculated and expected means

```r
plotdata <- data.frame(simdata_means);
m <- ggplot(plotdata, aes(x = simdata_means));
m <- m + geom_histogram(aes(y=..density..), colour="black", fill = "blue")
m + geom_density(colour="red", size=1);
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```



As you can see from the graph, the calculated distribution of means of random sampled exponantial distributions, overlaps quite nice with the normal distribution with the expected values based on the given lamba