Statistical Inference Course Project

Investigating the Central Limit Theorem for an Exponential Distribution

Author: Gaurav Bansal

In this project we will investigate the central limit theorem via the exponential distribution, which can be 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.

For a thousand simulations, we will investigate 40 distributions with lambda=0.2.

1. First, we want to show that the sample mean is about equal to the theoretical mean. Let's first calculate the sample mean:

```
lambda <- 0.2
n <- 40
sims <- 1000
set.seed(50)
distrib <- replicate(sims, rexp(n, lambda))
means <- colMeans(distrib)
calcmean <- mean(means)
print(calcmean)
```

```
## [1] 4.969142
```

Let's now compare the number above with the theoretical mean:

```
theomean <- 1/lambda
print(theomean)
```

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

We can see that the two numbers are roughly equal.

2. Second, let's compare the calculated and theoretical variance.

```
calcvar <- var(means)
print(calcvar)
```

```
## [1] 0.6399667
```

Let's now compare the number above with the theoretical variance:

```
theovar <- (1/lambda)^2/n
print(theovar)
```

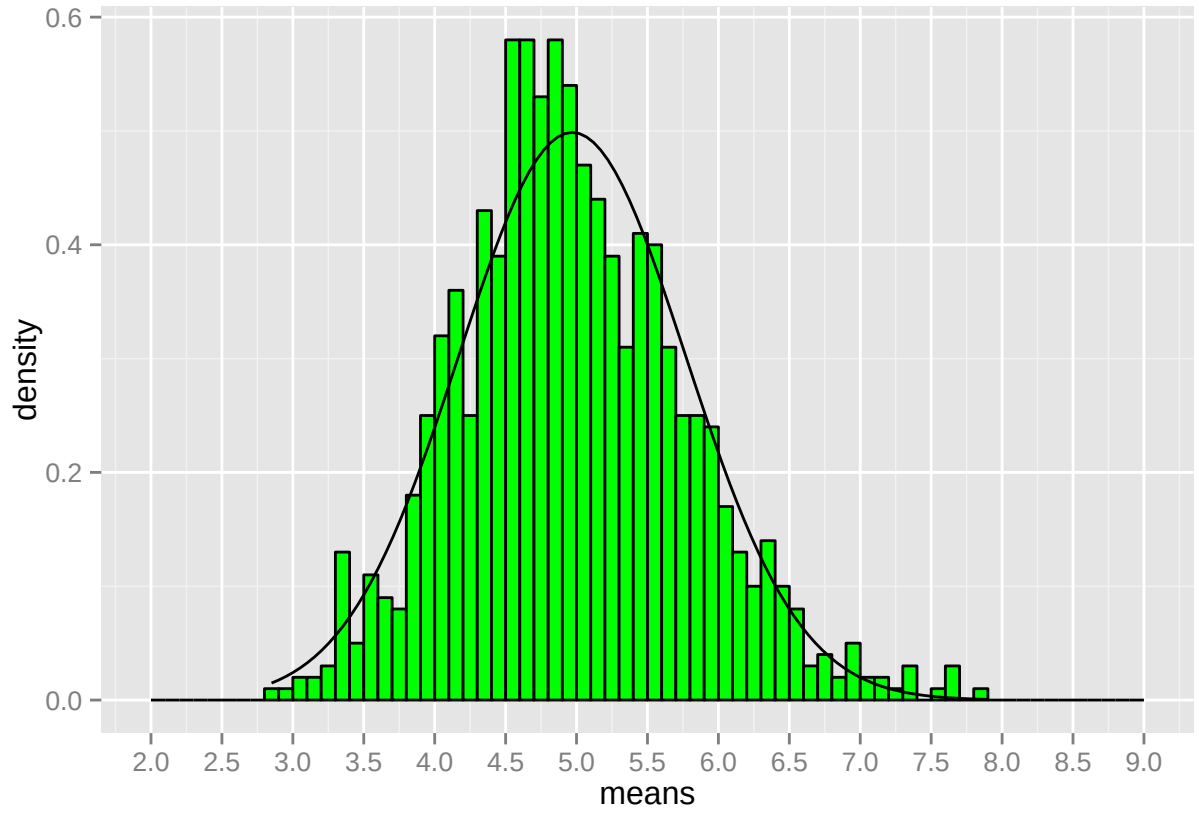
```
## [1] 0.625
```

We can see that the two numbers are roughly equal.

3. Lastly, let's show that the distribution of means is approximately normal. We can do this by simply plotting a histogram of the means and comparing it with a normal curve.

```
library(ggplot2)
normplot <- ggplot(as.data.frame(means), aes(x=means)) +
  geom_histogram(aes(y=..density..), color="black", fill="green", breaks=seq(2,9, by=.1)) +
  stat_function(fun=dnorm, args=list(mean=calcmean, sd=sqrt(calcvar))) +
  scale_x_continuous(breaks=seq(1,10, by=.5))
normplot
```

```
## Warning: position_stack requires constant width: output may be incorrect
```



Though no the best fit possible, we see that the histogram approximately fits a normal distribution.