Statistical Inference - Course Project, Part 1

ACBLimehouse

7/16/2017

Statistical Inference Course Project 1

Overview

In this project I will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution is simulated here 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. Lambda is set to <0.2> for all of the simulations. I will investigate the distribution of averages of 40 exponentials. In this investigation, there are 10,000 simulations.

Simulations

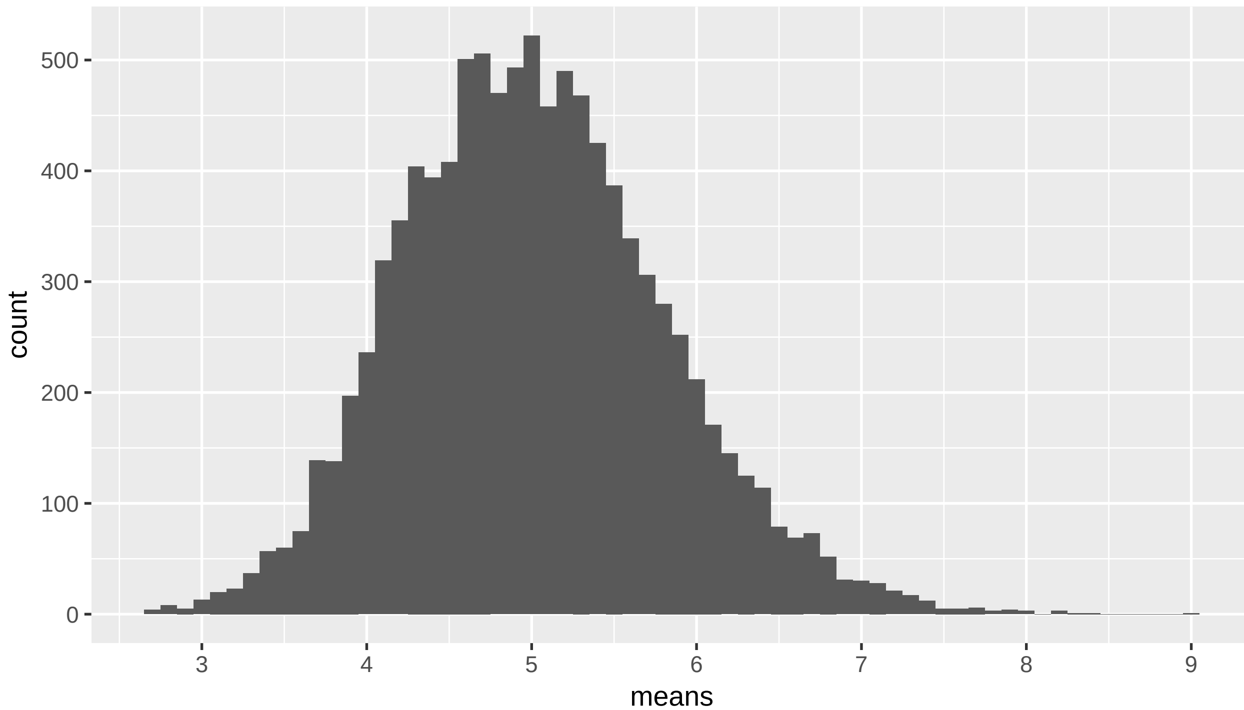
```
# load neccesary libraries
library(ggplot2)

# set constants
lambda <- 0.2 # lambda for rexp
n <- 40 # number of exponetials
numberOfSimulations <- 10000 # number of tests

# set the seed to create reproducability
set.seed(1105823)

# run the test resulting in n x numberOfSimulations matrix
exponentialDistributions <- matrix(data=rexp(n * numberOfSimulations, lambda),
                                   nrow=numberOfSimulations)
exponentialDistributionMeans <- data.frame(means=apply(exponentialDistributions, 1, mean))
```

```
# plot the means
ggplot(data = exponentialDistributionMeans, aes(x = means)) +
  geom_histogram(binwidth=0.1) +
  scale_x_continuous(breaks=round(seq(min(exponentialDistributionMeans$means),
                                      max(exponentialDistributionMeans$means), by=1)))
```



Sample Mean versus Theoretical Mean

The expected mean μ of a exponential distribution of rate λ is

$\mu = \frac{1}{\lambda}$

```
mu <- 1/lambda
mu
```

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

Let $\bar{X}$ be the average sample mean of 10,000 simulations of 40 randomly sampled exponential distributions.

```
meanOfMeans <- mean(exponentialDistributionMeans$means)
meanOfMeans
```

```
## [1] 5.016326
```

As you can see the expected mean and the avarage sample mean are very close.

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

```
sd <- 1/lambda/sqrt(n)
sd
```

```
## [1] 0.7905694
```

The variance Var of standard deviation σ is $Var = \sigma^2$

```r
Var <- sd^2
Var
```

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

Let Var_x be the variance of the average sample mean of 10,000 simulations of 40 randomly sampled exponential distribution, and σ_x the corresponding standard deviation.

```r
sd_x <- sd(exponentialDistributionMeans$means)
sd_x
```

```
## [1] 0.7968646
```

```r
Var_x <- var(exponentialDistributionMeans$means)
Var_x
```

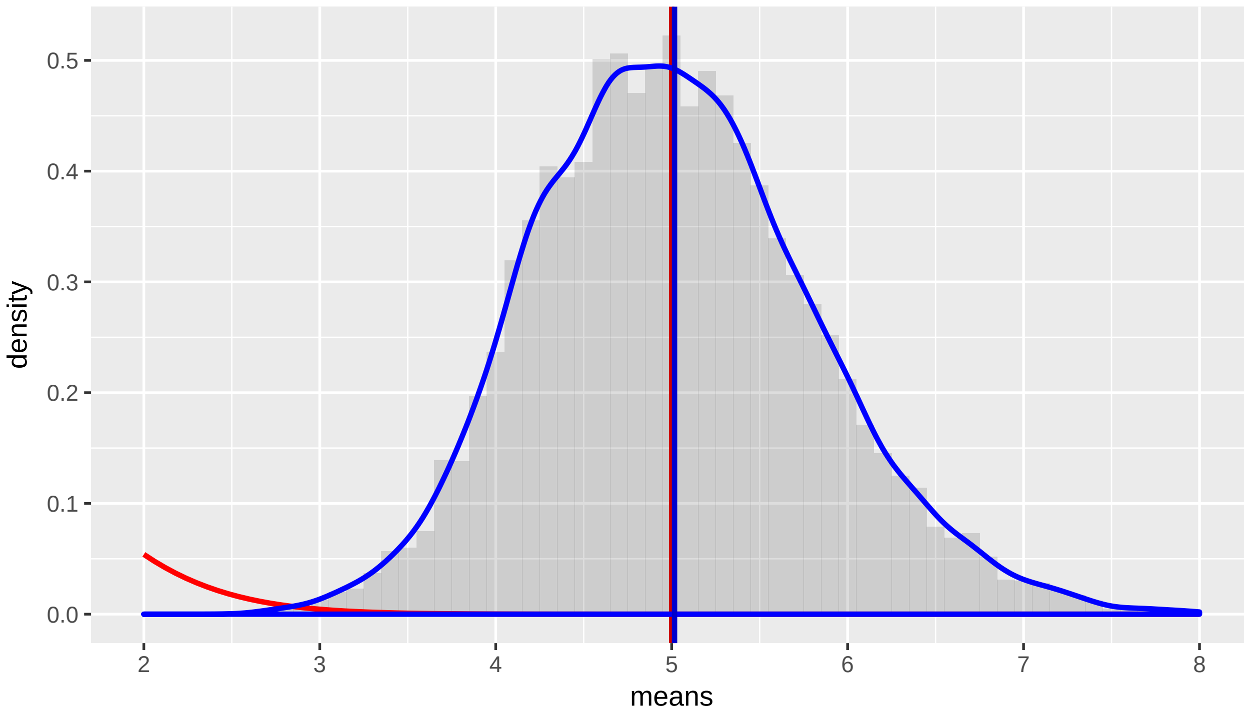
```
## [1] 0.6349932
```

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.

Distribution

The following graph compares the population means & standard deviation with a normal distribution of the expected values with added lines for the calculated and expected means:

```r
# plot the means
ggplot(data = exponentialDistributionMeans, aes(x = means)) +
  geom_histogram(binwidth=0.1, aes(y=..density..), alpha=0.2) +
  stat_function(fun = dnorm, arg = list(mean = mu , sd = sd), colour = "red", size=1) +
  geom_vline(xintercept = mu, size=1, colour="#CC0000") +
  geom_density(colour="blue", size=1) +
  geom_vline(xintercept = meanOfMeans, size=1, colour="#0000CC") +
  scale_x_continuous(breaks=seq(mu-3,mu+3,1), limits=c(mu-3,mu+3))
```



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