JHU_SI Project

mbh

Thursday, June 04, 2015

Rubric

In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution 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.

Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should 1. Show the sample mean and compare it to the theoretical mean of the distribution. 2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution. 3. Show that the distribution is approximately normal.

In point 3, focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

Title

Investigation of the distribution of the mean of samples (n=40) of exponential distirbutions

Author : Michael Hunt

Overview

In a few (2-3) sentences explain what is going to be reported on.

Simulations

Include English explanations of the simulations you ran, with the accompanying R code. Your explanations should make clear what the R code accomplishes.

Sample Mean versus Theoretical Mean

```
lambda<-0.2
means<-apply(matrix(rexp(40000,lambda),1000, 40), 1, mean)
hist(means)
```

Histogram of means
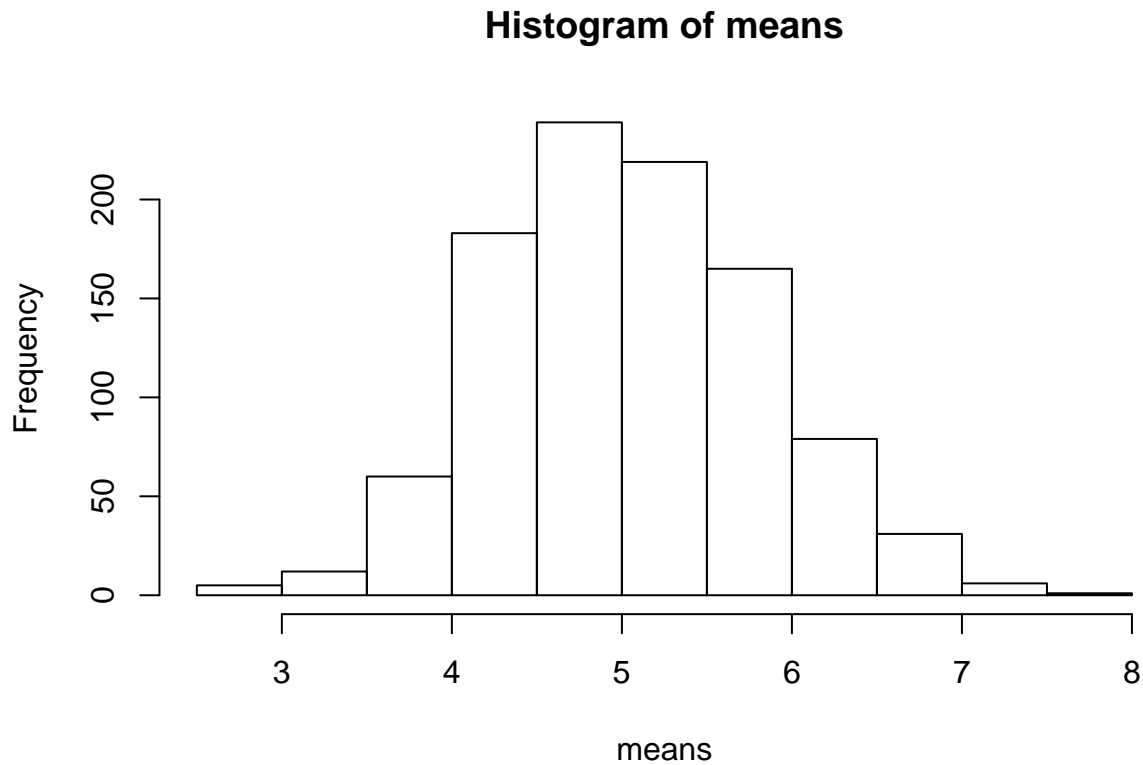

```r
mean(means)
```

```
## [1] 5.041933
```

```r
sd(means)
```

```
## [1] 0.7826139
```

The theoretical mean of the exponential distribution $f(x) = \lambda e^{(-\lambda/x)}$ is given by

$\mu = \int_0^\infty x f(x) dx = \int_0^\infty x \lambda e^{(-\lambda/x)} dx = \frac{1}{\lambda}$

Hence for $\lambda = 0.2, \mu = 5$.

The theoretical mean of a sample of n exponential distributions will also tend towards the same mean, as $n-> \infty$ since it is a **consistent** estimator of the population mean. This is the Law of Large Numbers.

```r
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:stats':
##
```

```
##      filter
##
## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union
```

```
nosim<-1000
dat <- data.frame(
  x = c(apply(matrix(rexp(4*nosim,lambda),nosim, 4), 1, mean),
        apply(matrix(rexp(40*nosim,lambda),nosim, 40), 1, mean),
        apply(matrix(rexp(400*nosim,lambda),nosim, 400), 1, mean)
        ),
  size = factor(rep(c(4, 40, 400), rep(nosim, 3))))
g <- ggplot(dat, aes(x = x, fill = size)) + geom_density(alpha = .20,colour = "black")
g
```
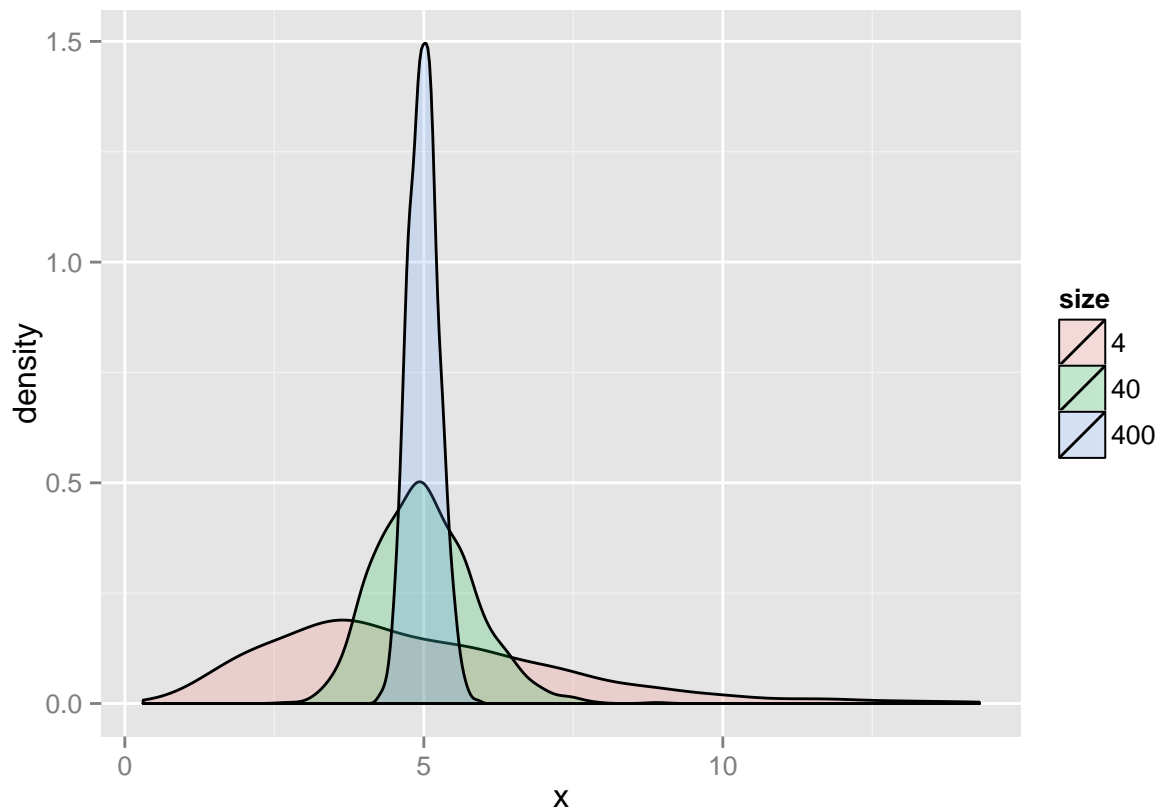


```
#g <- g + geom_vline(xintercept = 0, size = 1)
#g+facet_grid(.~ size)
```

```
library(ggplot2)
library(dplyr)
nosim<-1000
dat <- data.frame(
  x = c(apply(matrix((rexp(4*nosim,lambda)-1/lambda)/((1/lambda) / sqrt(4)),nosim, 4), 1, mean),
        apply(matrix((rexp(40*nosim,lambda)-1/lambda)/((1/lambda) / sqrt(40)),nosim, 40), 1, mean),
```

```
        apply(matrix((rexp(400*nosim,lambda)-1/lambda)/((1/lambda) / sqrt(400)),nosim, 400), 1, mean)
        ),
  size = factor(rep(c(4, 40, 400), rep(nosim, 3))))
g <- ggplot(dat, aes(x = x, fill = size)) + geom_density(alpha = .20,colour = "black")
g <- g + geom_vline(xintercept = 0, size = 1)
g+facet_grid(.~ size)
```
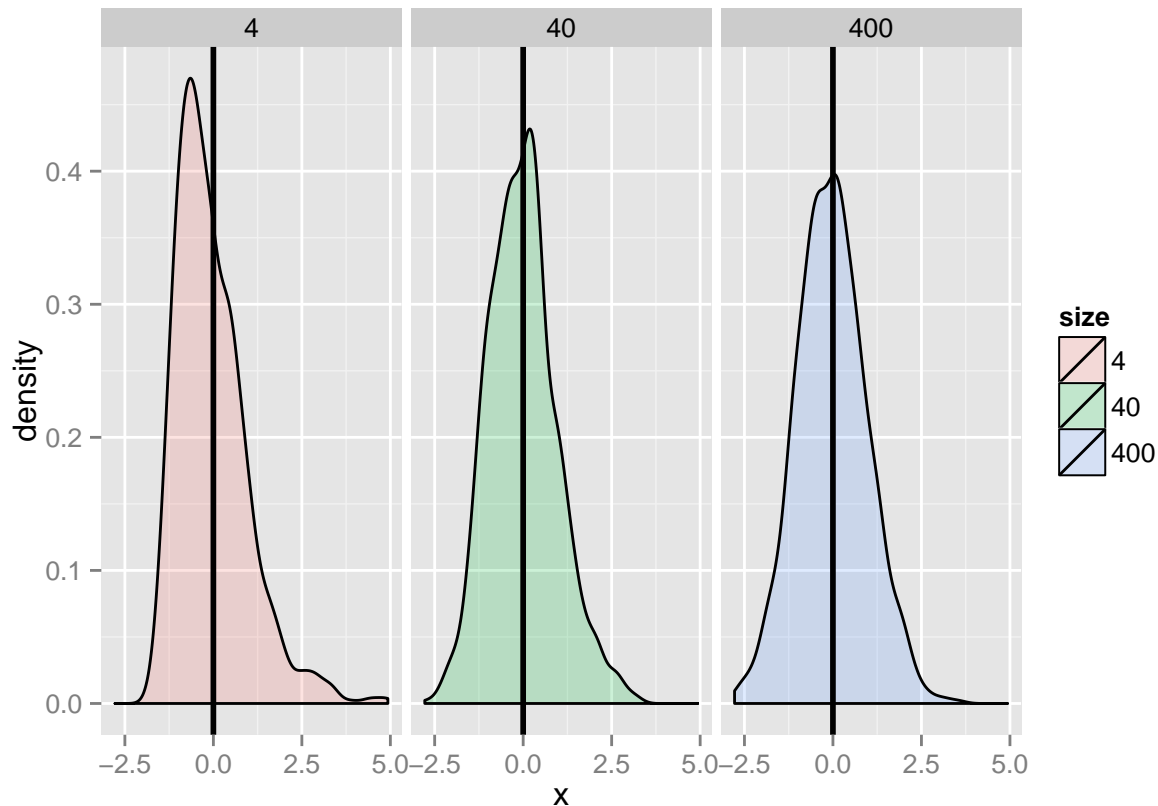

Include figures with titles. In the figures, highlight the means you are comparing. Include text that explains the figures and what is shown on them, and provides appropriate numbers.

Sample Variance versus Theoretical Variance

Include figures (output from R) with titles. Highlight the variances you are comparing. Include text that explains your understanding of the differences of the variances.

Distribution

Via figures and text, explain how one can tell the distribution is approximately normal.