Statistical Inference Project: Part 1

Question

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 also 1/lambda. Set lambda = 0.2 for all of the simulations. In this simulation, you will investigate the distribution of averages of 40 exponential(0.2)s. Note that you will need to do a thousand or so simulated averages of 40 exponentials.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponential(0.2)s. You should 1. Show where the distribution is centered at and compare it to the theoretical center of the distribution. 2. Show how variable it is and compare it to the theoretical variance of the distribution. 3. Show that the distribution is approximately normal. 4. Evaluate the coverage of the confidence interval for 1/lambda: $\bar{X} \pm 1.96 S_n$.

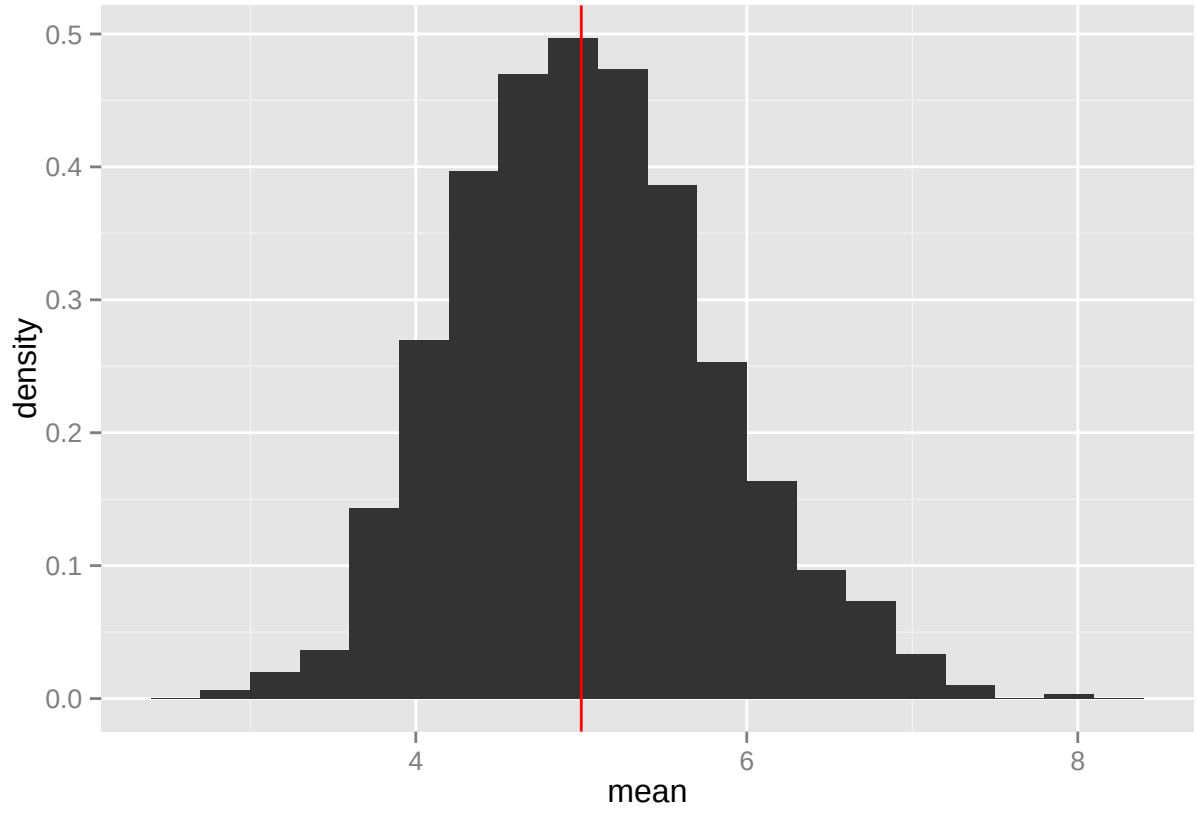
Answers

We run a 1000 simulations of drawing 40 samples from an exponential distrubution with *lambda = 0.2* . For every simulation we also calculation the mean, the variance and whether a 95% confidence interval on the sample contains the theoretical population mean.

```
library(ggplot2)
nosim <- 1000
ssize <- 40
lambda <- .2
μ <- 1 / lambda
stddev <- 1 / lambda
se <- stddev / sqrt(ssize)

xfunc <- function(x, μ, se) (mean(x) - μ) / se
cm <- function(x, stddev, μ) {
  civ <- mean(x) + c(-1, 1) * 1.96 * stddev / sqrt(length(x))
  civ[1] <= μ & civ[2] >= μ
}

sim = matrix(rexp(nosim * ssize, lambda), nosim)

dat <- data.frame(
  x = apply(sim, 1, xfunc, μ, se),
  var = apply(sim, 1, var),
  mean = apply(sim, 1, mean),
  civs = apply(sim, 1, cm, stddev, μ))
```

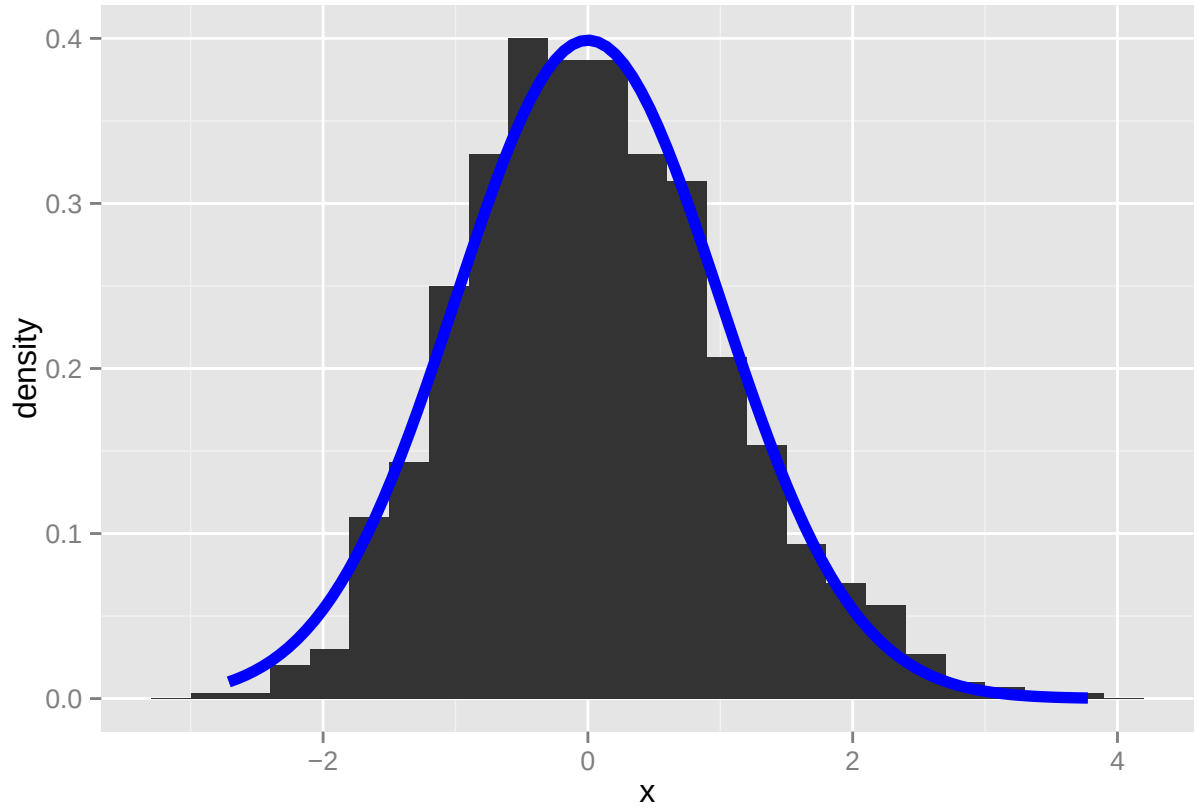
This plot shows that the sample means are centered around the theoretical population mean of 5.

```
gm <- ggplot(dat, aes(x = mean))
gm <- gm + geom_histogram(binwidth=.3, aes(y = ..density..))
gm <- gm + geom_vline(xintercept=5, linetype="solid", colour="red")

gm
```



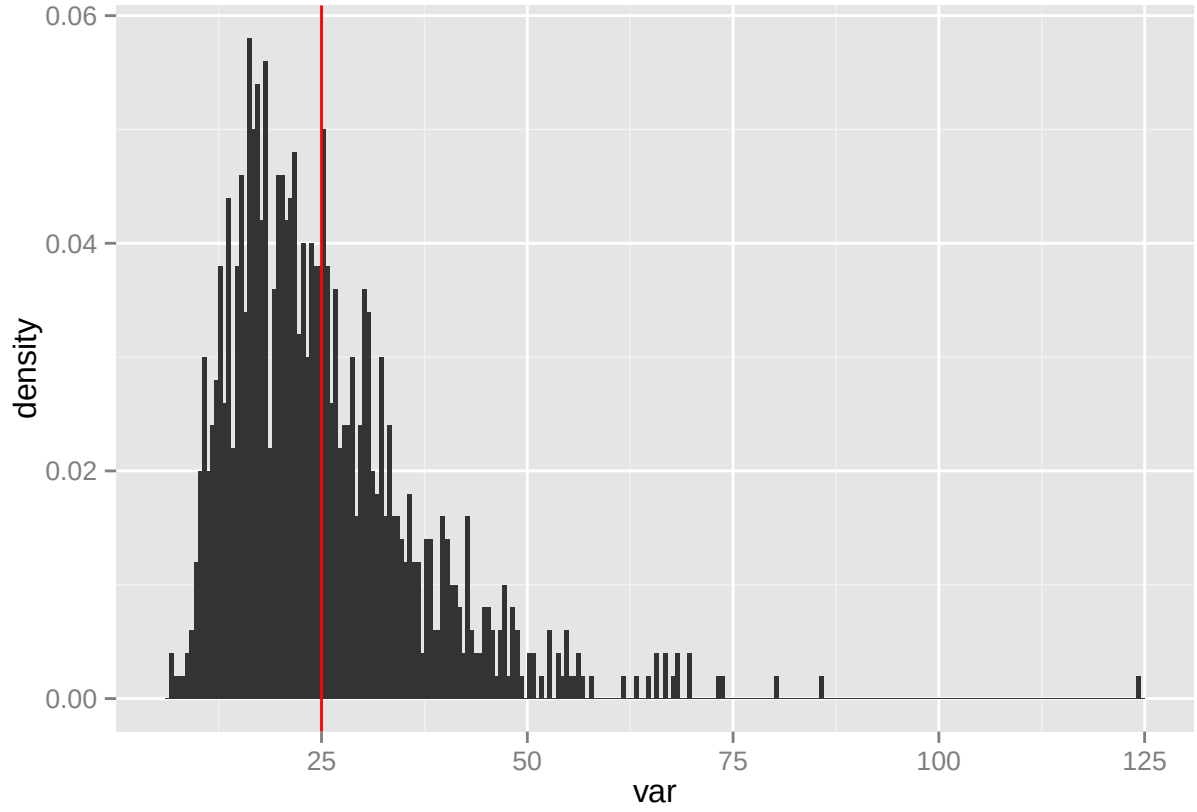
This plot shows that the resulting density is approximately normal.

```
gx <- ggplot(dat, aes(x = x))
gx <- gx + geom_histogram(binwidth=.3, aes(y = ..density..))
gx <- gx + stat_function(fun = dnorm, colour="blue", size = 2)

gx
```



This plot shows how the mass of the variance distribution roughly corresponds to the population variance.

```
gvar <- ggplot(dat, aes(x = var))
gvar <- gvar + geom_histogram(binwidth=.5, aes(y = ..density..))
gvar <- gvar + geom_vline(xintercept=stddev^2, linetype="solid", colour="red")

gvar
```



This ratio shows that roughly 95% of the samples contained the theoretical population mean within their 95% confidence interval.

```
sum(dat$civs) / length(dat$civs)
```

```
## [1] 0.949
```