

Analysis of the exponential distribution

Sid Reddy

20 Dec 2014

Analysis of the mean

In this section, we will identify the sample and the population means, and visualize them with a plot.

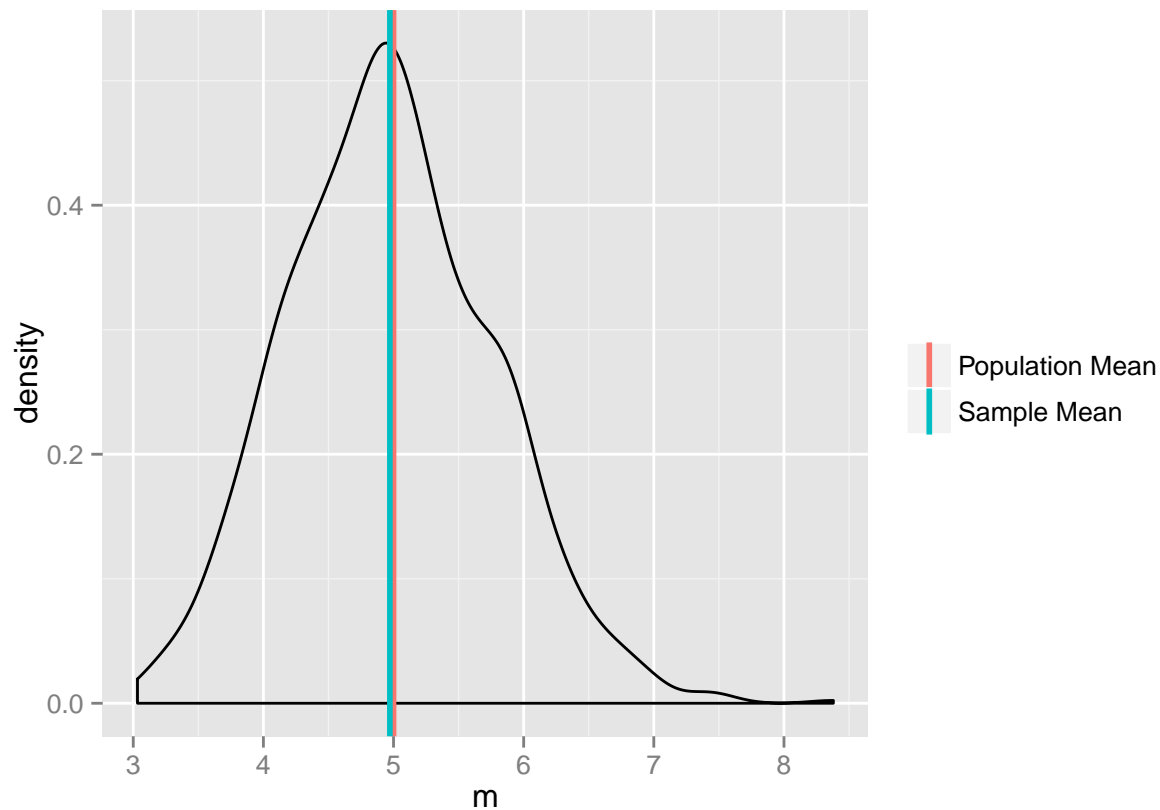
```
library(ggplot2)

# Set seed for reproducibility
set.seed(12345)

# Initialize all parameters: lambda, number of samples per simulation,
# number of simulations
l <- 0.2
n <- 40
ns <- 1000

# Obtain the distribution of the mean for each simulation, and the actual
# population mean
m <- apply(matrix(rexp(ns * n, l), ns), 1, mean)
mm <- mean(m)

# Plot the probability density function of the distribution above, along
# with the sample and population means
qplot(m, geom = "density") + geom_vline(aes(xintercept = c(mm, 1/l), color = c("Sample Mean",
  "Population Mean")), size = 1, show_guide = TRUE) + labs(color = "")
```



Answering the first function, the sample mean is 4.971972, and the population mean is 5.

Analysis of the variance

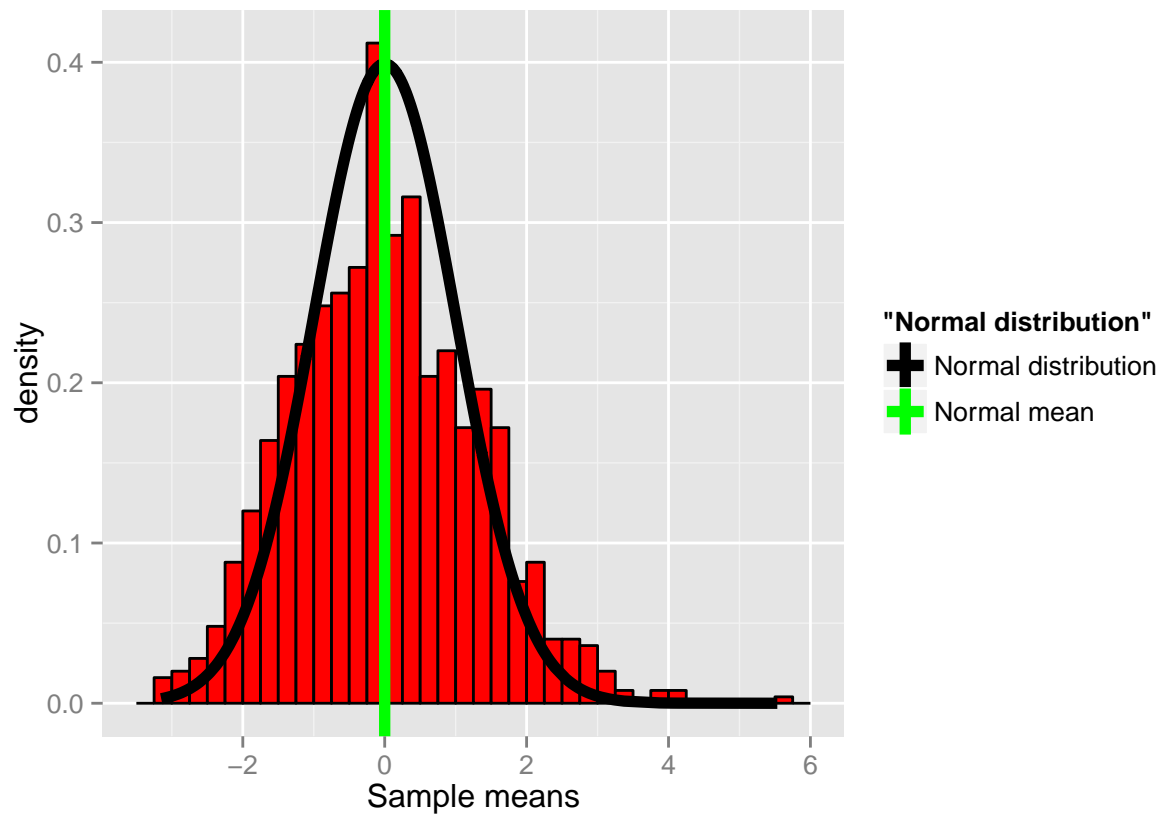
```
# Get the sample and the population variance of the distribution of means respectively
v <- var(m)
pv <- 1/(1^2 * n)
```

Answering the second question, the sample variance is 0.6157926, and the population variance is 0.625. We can notice that the variances are pretty close.

Comparison with the normal distribution

The plot below shows a histogram of the distribution of the means, along with a curve for the normal distribution. Note that the distribution is approximately normal.

```
x <- (m - mm)/v
ggplot() + geom_histogram(aes(x = x, y = ..density..), color = "black", fill = "red",
  binwidth = 0.25) + labs(x = "Sample means") + geom_line(data.frame(x = x,
  y = dnorm(x)), mapping = aes(x = x, y = y, color = "Normal distribution"),
  size = 2, show_guide = T) + geom_vline(aes(xintercept = 0, color = "Normal mean"),
  size = 2, show_guide = TRUE) + scale_color_manual(values = c(`Normal distribution` = "black",
  `Normal mean` = "green"))
```



Another way to look at this issue, is to observe the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials. You will notice that the first plot is the exponential distribution, while the distribution of averages of 40 exponentials starts looking like the normal distribution (given the central limit theorem).

```
library(gridExtra)
g1 <- qplot(rexp(1000, 0.2), geom = "histogram", binwidth = 0.25, color = I("black"),
  fill = I("red"), main = "Distribution of exponentials")
g2 <- qplot(m, geom = "histogram", binwidth = 0.25, color = I("black"), fill = I("red"),
  main = "Distribution of averages of 40 exponentials")
grid.arrange(g1, g2, ncol = 2)
```

