A Comparison of Exponential Distribution in R and the Central Limit Theorem

Harold Trammel

August 23, 2015

Overview

The purpose of this project is to nvestigate 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.

Simulations

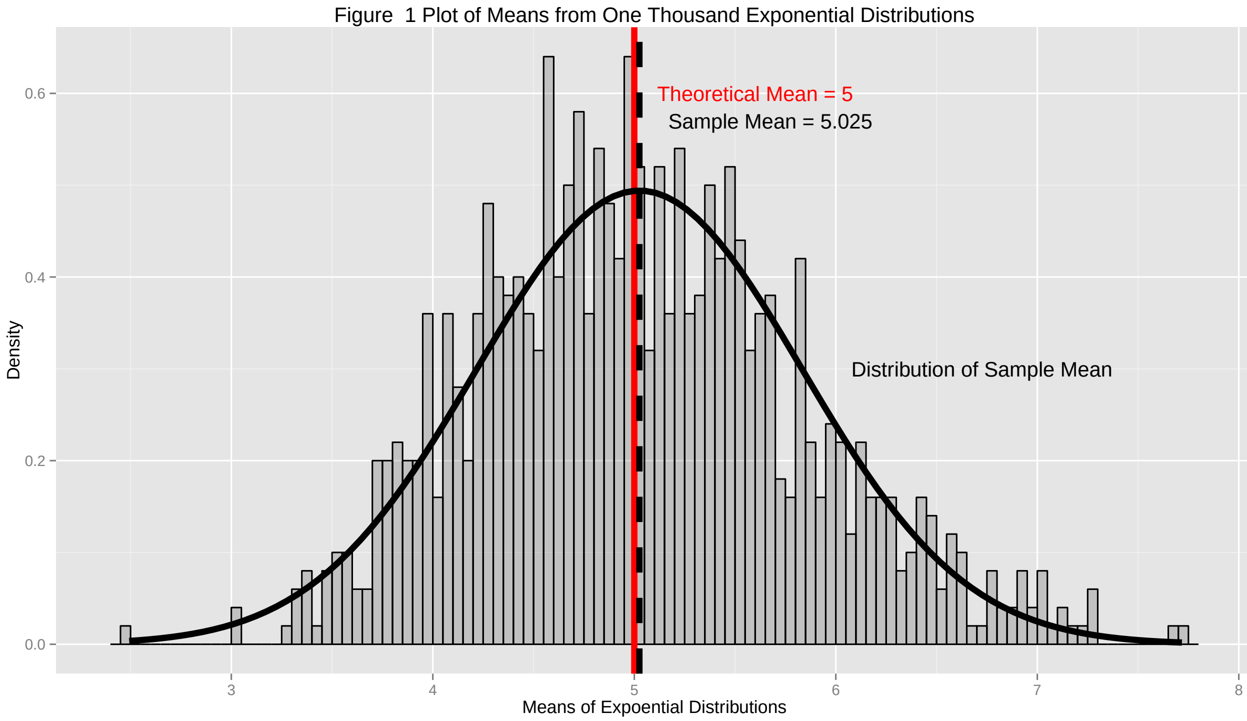
A set of one thousand simulations of forty expoential distributions was compiled. The mean, standard deviation, and variance for each simulation was determined and stored for analysis.

Sample Mean versus Theoretical Mean

The theoretical mean was 1/lambda or $1/0.2 = 5$. The theoretical standard deviation was also 1/lambda or $1/0.2 = 5$.

The combined sample mean was 5.03 which differed from the theoretical mean by -0.02. The standard deviation was 0.81, or a 0.31 difference from the theoretical standard deviation.

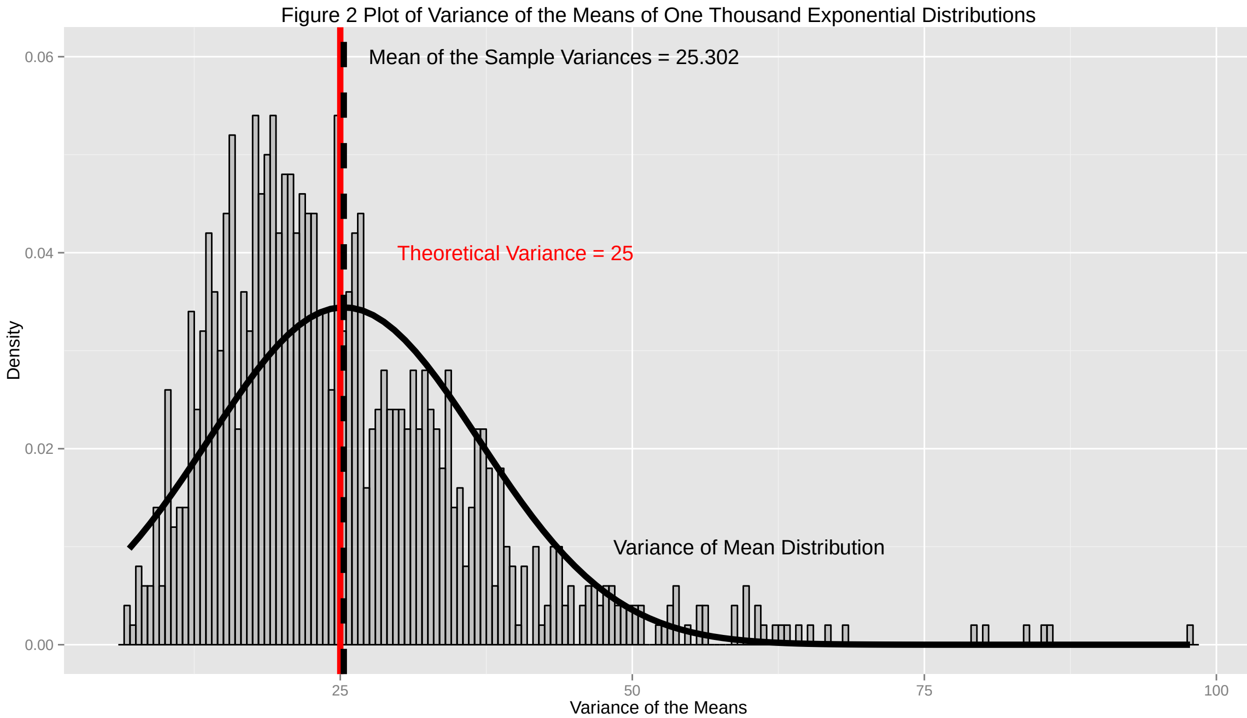
Figure 1 shows the distribution of the sample means. The graph suggests a normal distribution and a plot of a normal density based on the combined sample mean and standard deviation of the means was overlayed. The theoretical mean is shown as a red vertical line. The sample mean is shown as a black dashed line.



Sample Variance versus Theoretical Variance

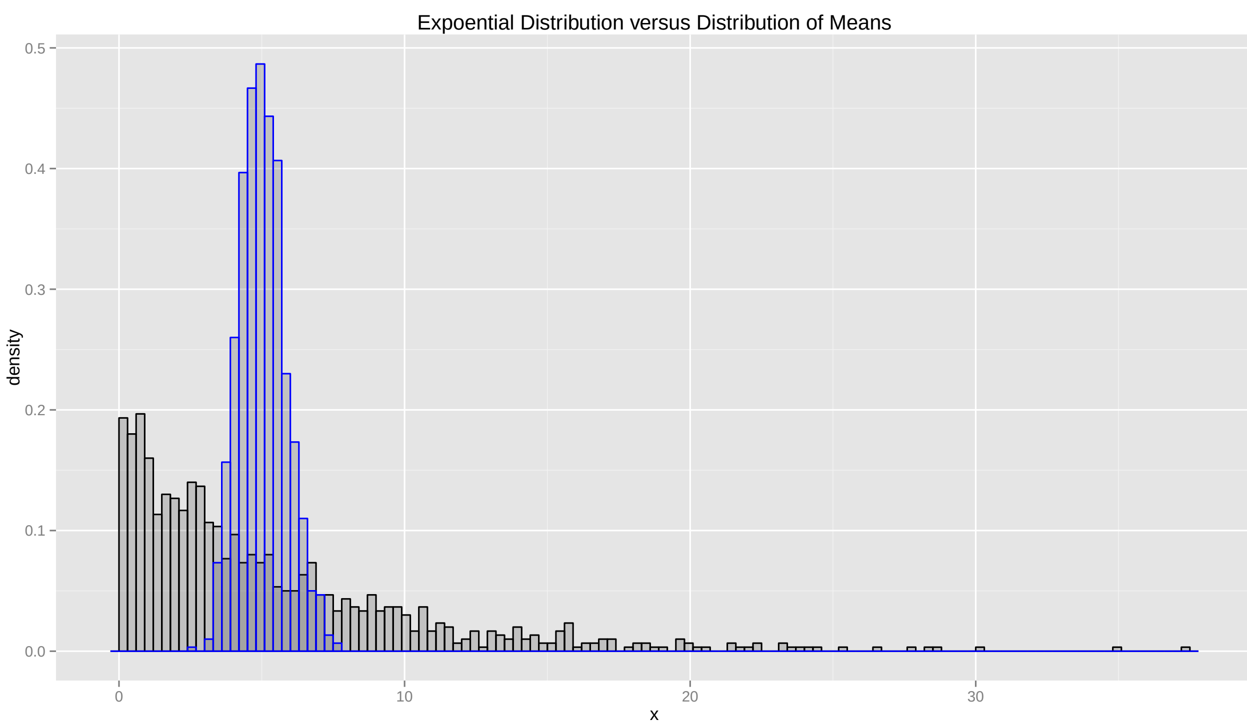
The theoretical variance of the simulations was 25 (the theoretical standard deviation squared). The variance of the simulations were averaged to get the mean of the sample variances, which is 25.3.

Figure 2 shows a plot of the sample variances. The graph suggests a normal distribution. The *fitdistr* function was used to estimate the mean and standard deviation of the plot. A plot of a normal density based on the estimated mean and standard deviation was overlayed. The theoretical variance is shown as a red vertical line and mean of sample variances is shown as a black vertical line.



Distribution

The following shows a large exponential distribution with a distribution of sample means.



Appendix

Data Simulations

```
set.seed(4)
dat <- data.frame(replicate(num_sim, rexp(num_samp, 0.2)))
pop_df <- data.frame(
    pop_mean = c(apply(dat, 2, mean)),
    pop_sd = c(apply(dat, 2, sd)),
    pop_var = c(apply(dat, 2, var))
    )
d_mean <- mean(pop_df$pop_mean)
d_sd <- sd(pop_df$pop_mean)
d_var <- var(pop_df$pop_mean)
```

Plot of Means from One Thousand Exponential Distributions

```
fig <- fig + 1
g1 <- ggplot(pop_df, aes(x = pop_mean)) +
        geom_histogram(alpha = .20, binwidth=.05, colour = "black", aes(y = ..density..)) +
        ggtitle(paste("Figure ", fig,
                      "Plot of Means from One Thousand Exponential Distributions"))  +
        geom_vline(stat = "vline", xintercept = theo_mean,
            size = 2, color = "red", linetype = 1) +
        annotate("text", x = theo_mean + 0.6, y = 0.6, colour="red",
            label="Theoretical Mean = 5") +
        geom_vline(stat = "vline", xintercept = samp_mean, size = 2, linetype = 2) +
        annotate("text", x = samp_mean + 0.65, y = 0.57,
            label = paste("Sample Mean =", round(samp_mean,3))) +
        stat_function(fun = dnorm, size = 2, linetype = 1,
            args = list(mean = samp_mean, sd = samp_sd)) +
        annotate("text", x = samp_mean+1.7, y=.3, label="Distribution of Sample Mean")
print(g1)
```

Plot of Variance of the Means of One Thousand Exponential Distributions

```
var_fit <- fitdistr(pop_df$pop_var, "normal")

fig <- fig + 1
g2 <- ggplot(pop_df, aes(x = pop_var)) +
    geom_histogram(alpha = .20, binwidth=.5, colour = "black", aes(y = ..density..)) +
    ggtitle(paste("Figure", fig,
        "Plot of Variance of the Means of One Thousand Exponential Distributions")) +
    stat_function(fun = dnorm, size = 2, linetype = 1, size = 2,
                  args = list(mean = var_fit$estimate[1], sd = var_fit$estimate[2])) +
    annotate("text", x = 60, y=0.01, label="Variance of Mean Distribution") +
    geom_vline(stat = "vline", xintercept = theo_var, size = 2, color = "red", linetype = 1) +
    annotate("text", x = theo_var + 15, y = 0.04, colour="red",
                label="Theoretical Variance = 25") +
    geom_vline(stat = "vline", xintercept = samp_var, size = 2, linetype = 2) +
    annotate("text", x = samp_var + 18, y = 0.06,
                label = paste("Mean of the Sample Variances =", round(samp_var,3)))
print(g2)
```

Plot Comparing Expoential Distribution versus Distribution of Means

```
set.seed(4)
samp <- data.frame(rexp(num_sim, 0.2))
names(samp) <- "x"
samp_fit <- fitdistr(samp$x, "exponential")
samp_rate <- samp_fit$estimate[1]

g3 <- ggplot(NULL) +
    geom_histogram(data=samp,
                   alpha = .20,
                   binwidth=.3,
                   colour = "black",
                   aes(x, y = ..density..)) +
    geom_histogram(data=pop_df,
                   alpha = .20,
                   binwidth=.3,
                   colour = "blue",
                   aes(pop_mean, y = ..density..)) +
    ggtitle("Expoential Distribution versus Distribution of Means")
print(g3)
```