Exponential Distributions and the Central Limit Theorem

Nick Krabbenhoeft

March 22, 2015

Summary

The Central Limit Theorem states that the distribution of means and variances of samples from a non-standard distribution tend towards a standard normal distribution. This report demonstrates the theorem with a simulation of data from the exponential distribution.

Simulation

First, we generate a set of 1000 samples of 40 observations from an exponential distribution with $\lambda = .2$ and calculate the mean and variance for each.

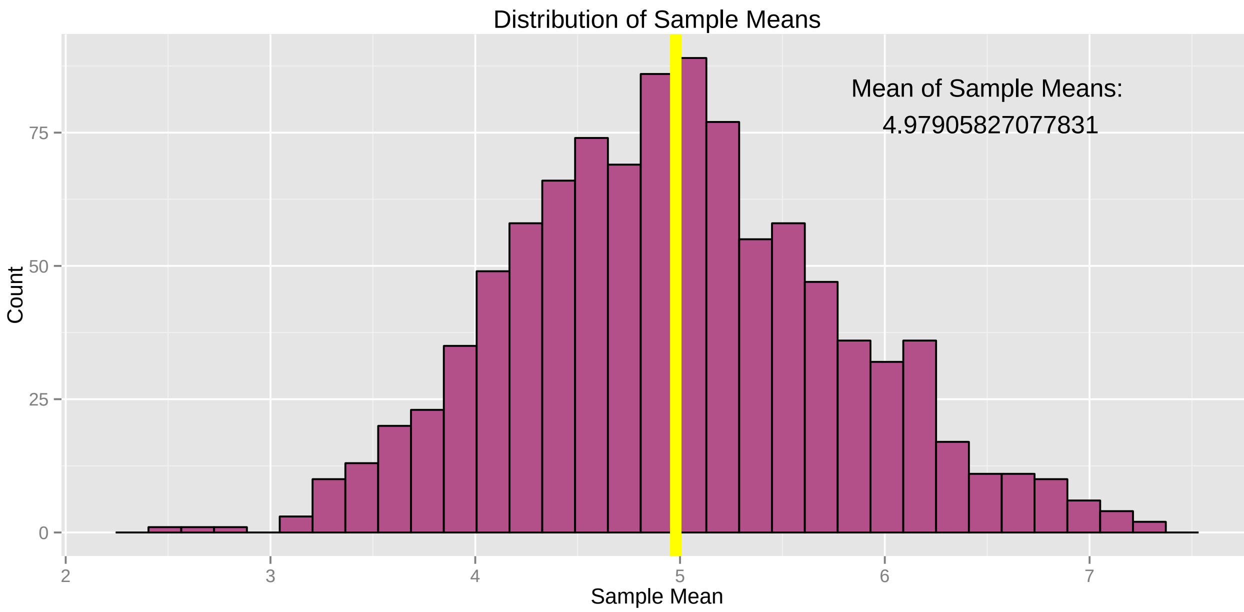
```
set.seed(11235)
sample <- replicate(1000, rexp(40, rate = .2))

# place means and variances of each sample in a data frame
df.sample <- data.frame(id = seq(1:1000))
for (i in 1:1000) df.sample$mean[i] <- mean(sample[, i])
for (i in 1:1000) df.sample$var[i] <- var(sample[, i])
```

Results

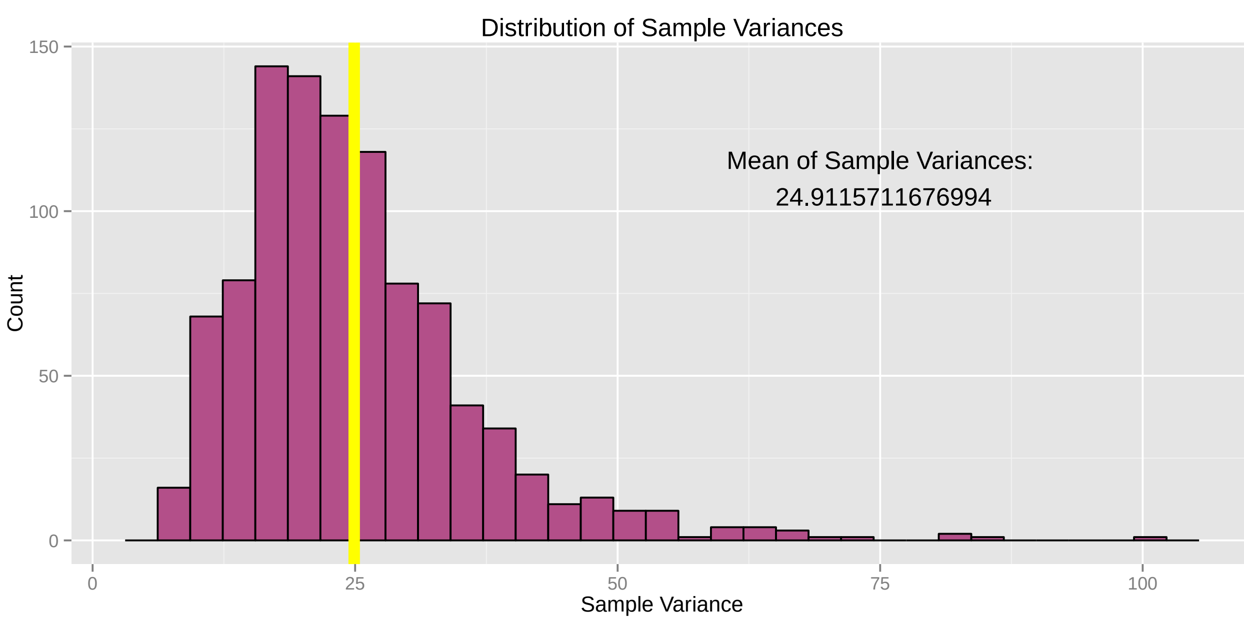
To demonstrate the Central Limit Theorem, we demonstrate that the sample means and variances tend towards the theoretical means and variances. Then we demonstrate that each tend towards normal distributions.

What are the sample means and variances?

```
library(ggplot2)
mean.mean <- mean(df.sample$mean)
ggplot(df.sample) +
        geom_histogram(aes(mean), fill = '#B34F89', col = 'black') +
        geom_vline(xintercept = mean.mean, col = 'yellow', size = 3) +
        labs(x = 'Sample Mean', y = 'Count',
             title = 'Distribution of Sample Means') +
        annotate('text', x = 6.5, y = 80,
                 label = paste('Mean of Sample Means:\n', mean.mean))
```




```r
var.mean <- mean(df.sample$var)
ggplot(df.sample) +
        geom_histogram(aes(var), fill = '#B34F89', col = 'black') +
        geom_vline(xintercept = var.mean, col = 'yellow', size = 3) +
        labs(x = 'Sample Variance', y = 'Count',
             title = 'Distribution of Sample Variances') +
        annotate('text', x = 75, y = 110,
                 label = paste('Mean of Sample Variances:\n', var.mean))
```



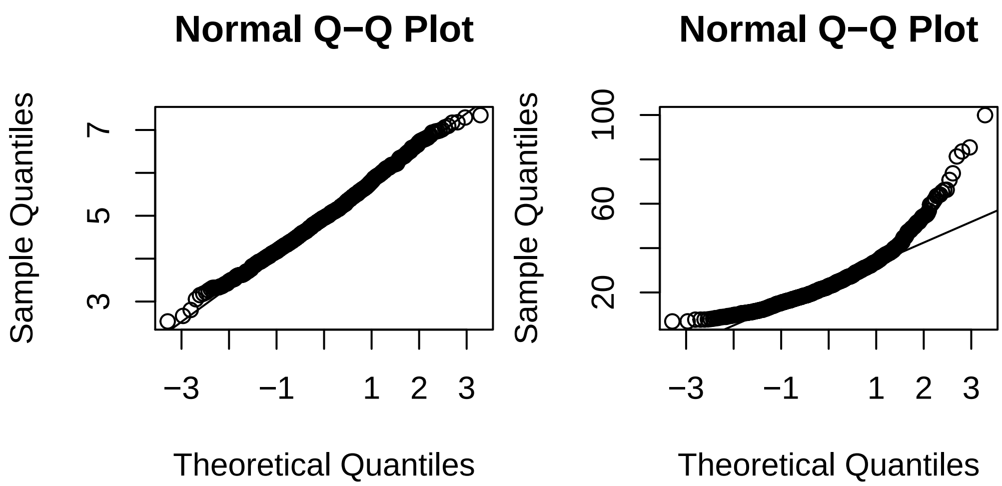
How do the sample means and variances compare to the theoretical means and variance?

The observed mean of sample means is 4.9790583. The theoretical mean is $1/\lambda = \frac{1}{.2} = 5$.

The observed mean of sample means is 24.9115712. The theoretical mean is $1/\lambda = \frac{1}{.2}^2 = 25$.

Is the distribution approximately normal?

```r
qqnorm(df.sample$mean); qqline(df.sample$mean)
qqnorm(df.sample$var); qqline(df.sample$var)
```



From the comparison of the distribution of the sample means and variances to the distribution of standard normal means and variances, we can see that the distribution of exponential samples approximates the standard normal distribution fairly well within one standard deviation of the mean. Exponential distributions have a much fatter tail then standard normal distributions, as can be seen in the long tail of the distribution of sample variances above and the difference between them above the first standard deviation from the mean of the variances in the plot on the right.