Inferential Stats Simulation Exercise

Introduction

An investigation of the exponential distribution in R and comparison with the Central Limit Theorem.

I will compare the Sample Mean and the Theoretical Mean, as well as Sample Variance versus Theoretical Variance.

Finally, I will demonstrate that the distribution approximatees normal distribution.

Sample Mean vs. Theoretical Mean

The variable `m` contains 1,000 means of 40 exponentials generated by `rexp(40, .2)`, where .2 is the rate parameter, or lambda. The theoretical mean and standard deviation of these exponential means should be 1/lambda = 1/.2 = 5.

The solid red vertical line shows the actual mean value, rounded to three places. The dotted black line shows the theoretical mean value of 1/lambda.

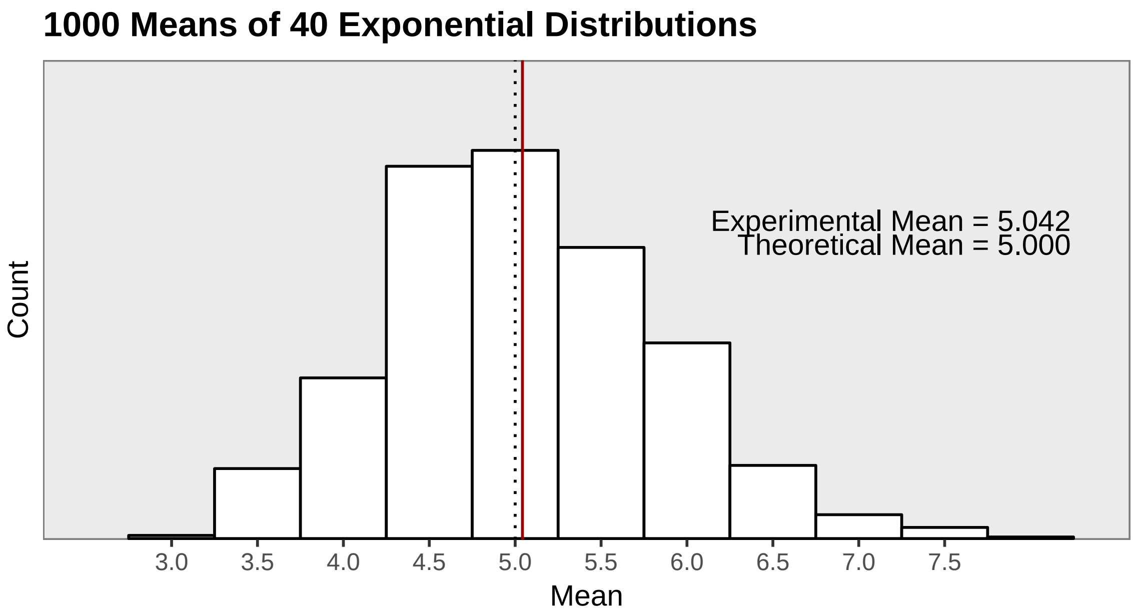
```
m <- numeric() # Container for distribution means
v <- numeric() # Container for distribution variances

for(i in 1:1000) {
  d <- rexp(40, .2)
  m[i] <- mean(d)
  v[i] <- var(d)
  }

p <- ggplot(mapping = aes(x = m)) +
  geom_histogram(binwidth = .5, color = 'black', fill = 'white') +
  ggtitle("1000 Means of 40 Exponential Distributions") + xlim(.5, NA) +
  scale_x_continuous(breaks = seq(floor(min(m)), max(m), .5), name = "Mean") +
  scale_y_discrete(name = "Count") + coord_cartesian(xlim = c(min(m)-.5,max(m)+.5 ), ylim = c(0,300)) +
  theme(panel.grid.minor = element_blank(), panel.grid.major = element_blank(),
        panel.background = element_rect(color = 'grey50'),
        plot.title = element_text(lineheight = .8, face = 'bold'),
        axis.text.y = element_blank(), axis.title.y = element_text("")
        )
```

```
## Scale for 'x' is already present. Adding another scale for 'x', which
## will replace the existing scale.
```

```
p + geom_vline(xintercept = mean(m), color = '#990000') +
    geom_vline(xintercept = 5.00, color = 'black', linetype = 'dotted') +
    geom_text(aes(mean(m) + 2*sd(m) + 1.6 ,200), label = paste0('Experimental Mean = ', round(mean(m),3
            hjust = 1) +
    geom_text(aes(mean(m) + 2*sd(m) + 1.6 ,185), label = paste0('Theoretical Mean = 5.000'),
            hjust = 1)
```



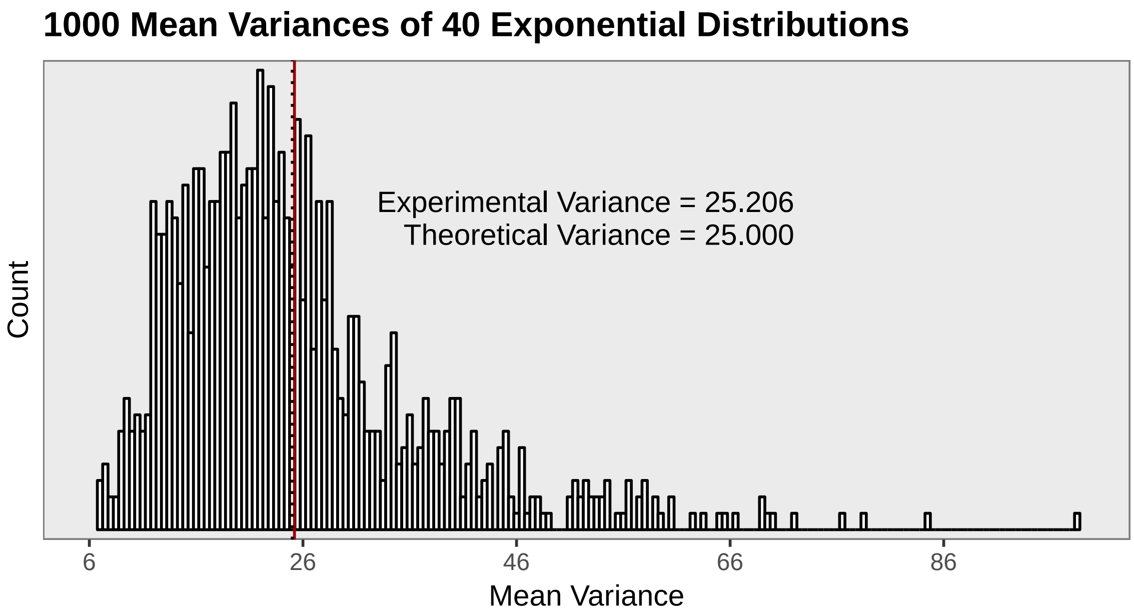
Sample Variance v. Theoretical Variance

We now create a histogram of the variances of each of our sample runs, and we find that the mean variance is 25.206. We have already noted that the theoretical standard deviation is 5. Since we know that `variance = sd^2`, the theoretical variance is 5^2, or 25.

This is very similar to our mean experimental variance.

```
q <- ggplot(mapping = aes(x = v)) +
  geom_histogram(binwidth = .5, color = 'black', fill = 'white') +
  ggtitle("1000 Mean Variances of 40 Exponential Distributions") +
  scale_x_continuous(breaks = seq(floor(min(v)), max(v), 20), name = "Mean Variance") +
  scale_y_discrete(name = "Count") +
  coord_cartesian(xlim = c(min(v)-.5,max(v)+.5 )) +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        panel.background = element_rect(color = 'grey50'),
        plot.title = element_text(lineheight = .8, face = 'bold'),
        axis.text.y = element_blank(), axis.title.y = element_text("")
        )

q  + geom_vline(xintercept = mean(v), color = '#990000') +
     geom_vline(xintercept = 25, color = 'black', linetype = 'dotted') +
     geom_text(aes(72,20), label = paste0('Experimental Variance = ', round(mean(v),3)), hjust = 1) +
     geom_text(aes(72,18),label = paste0('Theoretical Variance = 25.000'), hjust = 1)
```



Normality of the Distribution

By plotting a normal distribution along with the histogram of `m`, we can see that the distribution of means of exponential distribution is approximately normal, and therefore behaves as predicted by the Central Limit Theorem.

```
p + stat_function(fun = function(x, mean, sd, n, bw){
                            dnorm(x = x, mean = mean, sd = sd) * n * bw},
                  args = list(mean = mean(m), sd = sd(m), n = 1000, bw = .5), color = 'red') +
  geom_text(aes(mean(m) + 2*sd(m) + 1.6 ,225, color = 'red'),
            label = paste0('Normal Distribution, Mean = 5, SD = 5'),
            hjust = 1, show.legend = FALSE)
```