Part 1 - Simulation Exercise

Peter Maneykowski

4/15/2018

In this exercise I investigate the *exponential distribution* in R. There are functions in R to generate data for an exponential distribution, and additionally I create an exponential distribution from scratch using R's `data.frame` style arithmetic. This exercise is impactful because it lets me apply probability theory in practice using real data, then comparing the results to the theoretical expectation.

Below is the code to create a `data.frame` of exponential distributions (of dimension `(1000, 40)`) and the corresponding means of the columns of that data frame (of dimension `(1000, 1)`).

```
library(ggplot2)

# Constants - specified in the directions
lambda <- 0.2  # rate parameter
n <- 40  # number of exponentials
nsimulations <- 1000 # number of simulations

set.seed(2018)

# Exponential distribution as specified in the directions
exponential_distribution <-
  matrix(data=rexp(n * nsimulations, lambda), nrow=nsimulations)
exponential_distribution_means <-
  data.frame(means=apply(exponential_distribution, 1, mean))
```

Compare Sample Mean to Theoretical Mean

As the intructions specified, the theoretical mean of the distribution is `1/lambda`. Shown:

```
1/lambda
```

```
## [1] 5
```

The mean of our experimental distribution is shown as:

```
mean(exponential_distribution_means$means)
```

```
## [1] 5.020107
```

These values are extremely close which means there is a small error.

Compare Variance of Sample to Theoretical Variance

The instructions specify that the theoretical standard deviation is `1/lambda`, and since variance is calculated as σ^2, we show the variance is:

```
(1/lambda)^2
```

```
## [1] 25
```

Since the sample standard deviation is calculated the same as the mean, `1/lambda` for both cases, we may use the same value of μ as before to calculate the sample variance:

```r
mean(exponential_distribution_means$means)^2
```

```
## [1] 25.20147
```

These two value are extremely close which means that that there is some experimental error between the two values.

Show the Distribution is Approximately Normal

The frame of exponential distribution means is broken down in to many "buckets" (using an arbitrary `binwidth` of `0.05`), and it looks quite similar to the normal distribution. Below I have programmed a simple histogram showing the counts of means in the exponential distribution.

```r
min_x = min(exponential_distribution_means$means)
max_x = max(exponential_distribution_means$means)
ggplot(exponential_distribution_means, aes(x=means)) +
  geom_histogram(binwidth=0.05, fill = "salmon") +
  scale_x_continuous(breaks=round(seq(min_x, max_x, by=1))) +
  ggtitle("Exponential distribution means (approximately normal)")
```