Statistical Inference Assessment: Simulation Exercise

Jason Collins

Overview

In this report we compare the theoretical and sample mean and variance of 1000 simulations of 40 exponential distributions. The sample and theoretical means and variance are, as expected, close in value, with the distribution of the means approximating a normal distribution.

Simulations

For the analysis in this report, I performed 1000 simulations of 40 exponential distributions with $\lambda = 0.2$.

```
#set simulation parameters
lambda <- 0.2
n <- 40

#simulate and place in 1000 x 40 matrix
set.seed(20170301)
simulation <- matrix(rexp(1000*n, rate = lambda), ncol = n)
```

Sample Mean Versus Theoretical Mean

In this section I compare the theoretical mean, being that we would expect based on the parameters used in the simulation, with the sample mean. The theoretical mean can be calculated from the simulation parameters.

```
#theoretical mean
tmean <- 1/lambda
tmean
```

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

```
#sample mean, based on mean of the mean of each simulation
means <- rowMeans(simulation)
smean <- mean(means)
smean
```

```
## [1] 5.017577
```

The expected value of a sample from an exponential distribution with $\lambda = 0.2$ is 5. The sample mean is 5.0176, close to the theoretical mean.

This is shown in the histogram at the bottom of this report, with the red line representing the sample mean (the center of the sample distribution) and the blue line the theoretical mean based on the parameters used in the simulation.

Sample Variance Versus Theoretical Variance

In this section I compare the theoretical variance of the means of each simulation, being that we would expect based on the parameters used in the simulation, with the sample variance

```r
#theoretical standard deviation and variance
tsd <- 1 / lambda / n^0.5
tvar <- tsd^2
tvar
```

```
## [1] 0.625
```

```r
#sample variance
svar <- var(means)
svar
```

```
## [1] 0.6206553
```

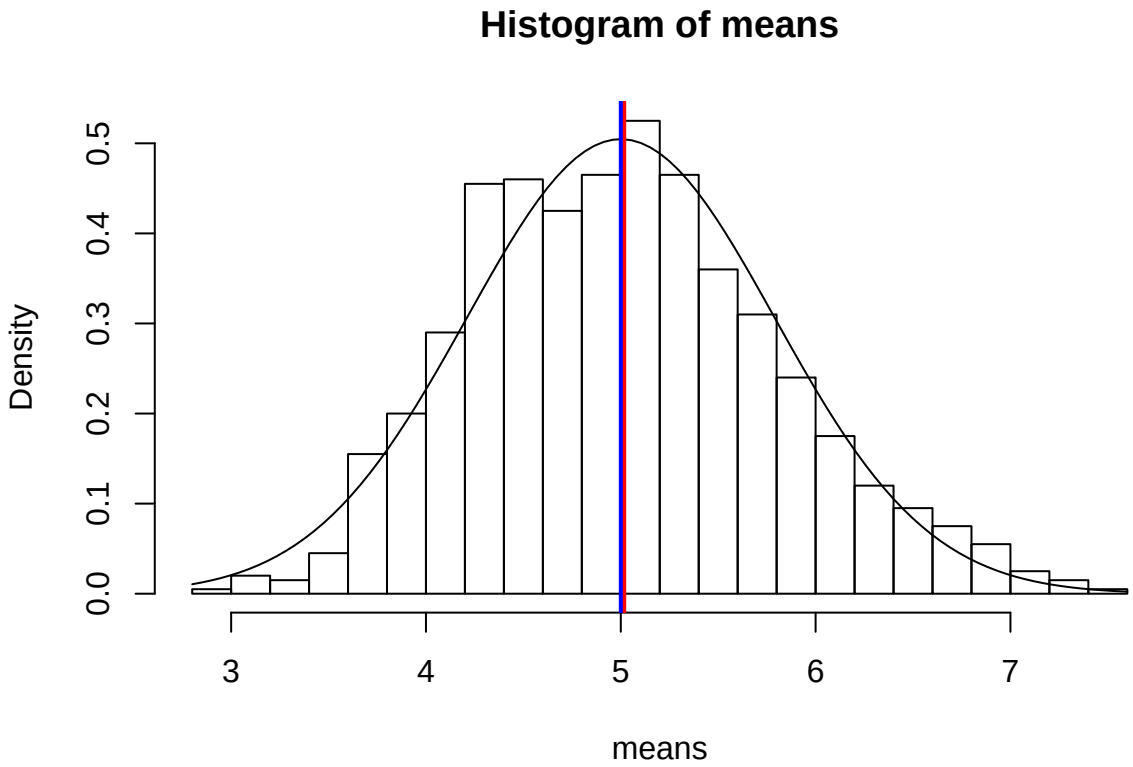
The theoretical variance of the averages of 40 exponentials drawn from an exponential distribution with $\lambda = 0.2$ is 0.625. Across the 1000 simulations of 40 exponentials, the sample variance of the means is 0.6207, similar to the theoretical variance.

Distribution

In this final section, we explore the normality of the distribution of the 1000 averages. We do this by plotting a histogram of the means and draw over it a normal distribution with the theoretical mean and variance.

```r
#histogram of means
hist(means, breaks = 20, probability = TRUE)

#add normal distribution and lines for theoretical and sample mean
curve(dnorm(x, tmean, tsd), add=TRUE)
abline(v = smean, col = "red", lwd = 2)
abline(v = tmean, col = "blue", lwd = 2)
```



Visually there is considerable similarity between the distribution of the averages and the normal distribution

based on the theoretical mean and variance. The sample mean is drawn on the plot as a red line, and the theoretical mean as a blue line.

Appendix

```
sessionInfo()
```

```
## R version 3.3.2 (2016-10-31)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS Sierra 10.12.3
##
## locale:
## [1] en_AU.UTF-8/en_AU.UTF-8/en_AU.UTF-8/C/en_AU.UTF-8/en_AU.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## loaded via a namespace (and not attached):
##  [1] backports_1.0.5  magrittr_1.5     RevoUtils_10.0.2 rprojroot_1.2
##  [5] formatR_1.4      tools_3.3.2      htmltools_0.3.5  yaml_2.1.13
##  [9] Rcpp_0.12.7      stringi_1.1.2    rmarkdown_1.3    knitr_1.14
## [13] stringr_1.1.0    digest_0.6.10    evaluate_0.10
```