MATH 185 – Homework 1
Due Wednesday, 04/11, by 11:59 PM

Send your code here. For Homework 1, write "MATH 185 - HW 1" *in subject line and nothing else in the body. There should only be one file attached, with the name* hw1-lastname-firstname.R. *Make sure your code is clean, commented and running. Keep your code simple, using packages only if really necessary. If your code does not run, include an explanation of what is going on.*

Problem 1. For the chi-squared goodness-of-fit statistic, we evaluate by simulation how accurate the large-sample approximation by the chi-squared distribution is. Our scope is modest, however. Consider the special situation where we observe a sample $X_1, \ldots, X_n$ taking values in $\{1, \ldots, k\}$, and want to test for uniformity, meaning, test whether the underlying distribution is the uniform distribution on $\{1, \ldots, k\}$.

A. Write an R function chisq.sim.hist(n, k, B). For each $b = 1, \ldots, B$, the function first simulates $X_1^b, \ldots, X_n^b$ iid from the uniform distribution on $\{1, \ldots, k\}$ and computes the corresponding chi-squared statistic D_b. It then produces a histogram of $D_1, \ldots, D_B$ and overlays the density of the chi-squared distribution with $k - 1$ degrees of freedom.

B. Try this function in the setting where $k = 5$, and $n \in \{10, 20, 50, 100\}$. Choose $B =$ 1e4. Produce a plot with 4 subplots, one for each of these settings.

C. Write a function chisq.sim.critical(n, k, B) that returns the critical value that would be used for testing at the 5% level. [This can be done with a simple modification of your function in Part A.]

D. Use this function to compare the critical value obtained by simulation with the theoretical critical value — which is the 95% quantile of the chi-squared distribution with $k - 1$ degrees of freedom — when $k = 5$ and $n \in \{10, 20, 30, \ldots, 90, 100\}$. Choose $B =$ 1e4. Produce a plot.

Problem 2. Are boys and girls equally likely to be born any month of the year? Go to http://wonder.cdc.gov/natality.html and extract data by Gender and Month for the year 2016. Enter the data in R. You can do so by hand or click on "Export", edit the resulting text file and then use the function read.table. Use these data to answer the question the best you can. Start by formalizing the question into a hypothesis testing problem, display relevant summary statistics and graphics, and then perform an appropriate test. Conclude with in a sentence or two.