Problem Set 3

In this problem set, you will continue to be asked to work with the gamma distribution and you may find it useful to refer back to the previous problem set and/or solutions. I recommend Wikipedia as a reference, `https://en.wikipedia.org/wiki/Gamma_distribution`.

Keep your rendered `.pdf` to no more than 4 pages. Only provide code in the rendered `.pdf` when it is specifically requested.

1. Continue using four sets of values for the shape parameter α and rate parameter λ: $(\alpha_1, \lambda_1) = (0.5, 0.5)$, $(\alpha_2, \lambda_2) = (2, 1)$, $(\alpha_3, \lambda_3) = (0.5, 1)$, and $(\alpha_4, \lambda_4) = (2, 0.5)$, and store these values in a matrix `Y` with 4 rows and 2 columns, where the first column corresponds to the shape parameter, the second column corresponds to the rate parameter, and the rows correspond to the parameters in the order provided above.

(a) Consider a sequence of values for the x-axis of a plot given by `x <- seq(0.01, 6, by = 0.01)`. Create a matrix `A` with the same number of rows as `Y` and the same number of columns as the length of `x` with all elements equal to `0`. Print the code you use to create this matrix to the rendered `.pdf`.

(b) Fill each row of `A` with the value of the gamma density evaluated at `x` for the parameter values in the same row of `Y`. Print the code you use to create this matrix to the rendered `.pdf`.

(c) Provide the maximum observed density value, i.e. the value of the largest element in `A`.

(d) Make a plot that is almost identical to the plots you made on Problem Set 2, with one change - make the limits of the y-axis 0 and the maximum density value you observe in `A`.

(e) Using `which.max` applied to each row of `A` and `x`, create a vector of length 4 named `maxx` with elements that describe the value of `x` at which the density is largest for each set of parameter values. Print the code you use to create this vector to the rendered `.pdf`.

(f) Using the `ifelse` command and some simple mathematical operations applied to `Y` as well as what you know about the Gamma distribution, create a vector of length 4 named `modex` with elements that correspond to the mode of a Gamma distribution with the shape and rate parameters in each row of `A`. Print the code you use to create this vector to the rendered `.pdf`.

(g) Summarize what you find in (e) and (f) using a 4×4 table made using `kable`, with a row for each pair of values named "Choice 1," "Choice 2," "Choice 3," and "Choice 4" one column for the α values named "Shape," one column for the λ values named "Rate," one column for the corresponding mode named "Mode" and one column for the corresponding value of `x` where the density is maximized named "Approximate Mode". Use at most 2 significant digits (you can adjust the number of digits printed by changing the argument `digits` that can be applied when using the `kable` function).

(h) In at most two sentences, explain how the "Mode" and "Approximate Mode" columns compare and justify any differences you observe.

2. In class, we did a little exercise where we created our own data. Given a vector `x <- seq(1, 9, by = 1)`, we created a vector `y` by adding a random integer between -3 and 3 to each element of `x`, `y <- x + c(2, 2, -3, 0, -1, -2, 2, 3, 1)`.

(a) Create a matrix with 9 rows and 100 columns called `Y` with elements given by `sample((-3):3, 9*100, replace = TRUE)`. Print the code you use to create this matrix to the rendered `.pdf`.

(b) Read the documentation of the function `sample` found by entering `help(sample)` into `R` and, given what you have learned about vectors and matrices, explain what `Y` is and how a row of `Y` relates to what we did in class in at most one sentence.

(c) In class, we smoothed `y` by averaging the pairs of consecutive observations (`z` with `z[1] = y[1]`, `z[2] = (y[1] + y[2])/2`, etc.) and then triads of consecutive observations (`w` with `w[1] = y[1]`, `w[2] = (y[1] + y[2])/2`, `w[3] = (y[1] + y[2] + y[3])/3`, etc.). Create matrices `Z` and `W` of the same dimensions as `Y`, but with rows that contain smoothed rows of `Y` obtained by averaging pairs of consecutive observations (for `Z`) and triads of consecutive observations (for `W`). Print the code you use to create this matrix to the rendered `.pdf`.

(d) Create vectors `ymean`, `zmean`, and `wmean` which contain the average of each row of `Y`, `Z`, and `W`, respectively. Print the code you use to create this matrix to the rendered `.pdf`.

(e) Create vectors `ysd`, `zsd`, and `wsd` which contain the standard deviation of each row of `Y`, `Z`, and `W`, respectively. Print the code you use to create this matrix to the rendered `.pdf`.

(f) Create a pair of plots by prefacing your `plot` commands with `par(mfrow = c(2, 1))`. In the first plot, plot `x` against the row means you constructed in (d). Use points with `pch = 16`. Make sure to use a different color for set of row means and annotate them with a legend. In the second plot, plot `x` against the row standard deviations you constructed in (e). Use points with `pch = 16`. Make sure to use a different color for set of row standard deviations and annotate them with a legend.