Due Date: October 3rd, 2019
Problems Due: 1, 2(a), (b).

1. Problems in R - pre-requisite
   (a) Write a `function` in R that will compute the L_1 distance between two vectors.
   (b) Consider a suitable sequence of 100 points and obtain a plot of Gamma$(1, 3)$, Gamma$(3, 5)$, Gamma$(1.5, 4.5)$
   (c) The following are well known properties of Gamma distribution:[1]
   - If $X \sim \Gamma(\alpha, \beta)$ then the probability density function of X is given by
   $$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \qquad x > 0.$$
   - $E[X] = \frac{\alpha}{\beta}$ and $\text{Var}[X] = \frac{\alpha}{\beta^2}$

   Consequently Simulate 100 replications of 50 samples from Gamma$(3, 5)$. Using the data X in each replicate and let $\hat{\alpha} = \frac{\bar{X}^2}{S^2}$, $\hat{\beta} = \frac{\bar{X}}{S^2} = \frac{\hat{\alpha}}{\bar{X}}$. Finally estimate the true shape and rate parameters by providing the average from the one obtained from each replicate.

2. *From Rajesh Sundaresan Lecture:*
   - Slides of talk are available here
   - Search times data (searchtimes.csv file is there is the shared dropbox folder.) This file contains data on search times on certain image pairs.
     i. There are two sets of data. Set 1 has six groups of experiments and Set 3 has three groups of experiment. Each group is for a fixed oddball image against its distracting pair.
     ii. Column A contains search times of individuals for the "Set 1 - Colour" pair in lecture slides on page 9. The oddball image is the picture on the left in the picture slide (Set 1 - Colour, left image with red on top of green). This column is labeled Oddball L. Column B contains the search times when the oddball image is the one on the right (Oddball R, green on top). Each individual was presented this oddball test 12 times in random order. The data is similarly arranged for columns C through R (Sets 1 and Sets 2).
     iii. You may think of each column entry as s_{ij} for oddball image i and distractor image j. These are not symmetric.
   - Firing rate data (firingrates.csv file is there is the shared dropbox folder.) This file contains data on the firing rates of neurons.
     i. Each set has a different number of neurons whose firing rates were recorded.
     ii. Consider columns A,B. Column A records the average firing rates of 114 neurons on "Set 1 - Colour - left image with red on top", while Column B refers to "Set 1 - Colour - right image with green on top".
     iii. You may think of each column entry as λ_i for image i.

   (a) In Sets 1, and 2, for each image pair on which the behavioural tests were conducted:
     i. Using searchtimes.csv compute the average search delay and using firingrates.csv compute the L_1 distance between the firing rates per neuron. *When computing average search delays, remember to subtract the baseline reaction time of 328 ms.*
     ii. Now, with L_1 distance on the x-axis and inverse of search delays on the y-axis, find the best straight line passing through the origin.
   (b) In this part, again with Sets 1 and 2, we will try to fit a Gamma distribution to the search delays.

[1] The Gamma distribution Gamma(α,β) has two parameters (shape α and rate β). It also has the property that the standard deviation to mean ratio is not arbitrary, but tied to the shape parameter α. If the shape parameter is 1, we get the exponential distribution. If the shape parameter is 2, we get a random variable that is the sum of two exponential random variables with the same rate parameter, and so on. In general, the shape parameter need not be an integer.

i. Estimate $\hat{\alpha}$: Pick all "Oddball L" groups. Plot the standard deviation of the search times against their means. Estimate the shape parameter (See Slides page 19,20).

ii. Estimate $\hat{\beta}$: For each "Oddball R" group, take the first 36 samples and estimate the rate for that group.

iii. Using the 36 samples of the "Oddball R" group that were not used:

A. Plot a density estimate of the data and see how it compares with the true density of $\Gamma(\hat{\alpha}, \hat{\beta})$

B. *(Extra Credit)* Perform the Kolmogorov-Smirnov test using `ks.test` to decide if the search delays obey the Gamma distribution using 5% as level of confidence.

(b') Extra Credit*Benne Dosai Challenge*[2]: In this part, again with Sets 1 and 2, we will try to fit a Gamma distribution to the search delays using randomisation.

i. Randomly select half the number of groups. Plot the standard deviation of the search times against their means. Estimate the shape parameter (See Slides page 19,20).

ii. On each of the groups that did not contribute to the shape parameter estimation, randomly select one half of the samples and estimate the rate parameter. Justify your estimation procedure.

iii. On each of the groups of (b')(ii), taking the samples that did not contribute to the rate parameter estimation, plot the empirical cumulative distribution function (cdf). On the same figure, plot the Gamma(α, β) cdf where α is the shape parameter estimated in (b')(i) and β is the rate parameter estimated in (b')(ii) on this group.

iv. Perform the Kolmogorov-Smirnov test using `ks.test` to decide if the search delays obey the Gamma distribution using 5% as level of confidence. Explain the test and what its result means

[2]If I get a complete solution to this question [*From two students*] along with H.W by Tuesday October 1st [*Earlier than due date*] and it is written up properly, along with a clear summary from Rajesh's talk then I will sponsor breakfast for everyone who has 90% attendance in the class at world famous branch of Davangere Benne Dosai in RR Nagar.