Homework 7 – Due 11:59 pm CT, 16 April 2021

The total points on this assignment is 75. The datasets are posted on the class website.

1. Gamma Ray Bursts (GRBs) are the brightest known electromagnetic events known to occur in space and have been studied extensively ever since their discovery in the late sixties. While the cosmological origin of GRBs is well-established, questions on their source and nature remain unresolved. The Burst and Transient Source Experiment (BATSE) Catalog of the National Aeronautics and Space Administration (NASA) provides temporal and spectral information for many GRBs. Of interest to us are the parameters:

   T_{50}: The time by which 50% of the flux arrive.

   T_{90}: The time by which 90% of the flux arrive.

   P_{64}, P_{256}, P_{1024}: The peak fluxes measured in bins of 64, 256 and 1024 milliseconds, respectively.

   F_1, F_2, F_3, F_4: The four time-integrated fluences in the 20-50, 50-100, 100-300, and $>$ 300 keV spectral channels, respectively.

   The first two variables are the time variables, the next set are the peak fluxes, and the third set are the time-integrated fluences.

   The current BATSE catalog, that is, the BATSE 4Br contains bursts from 1973 Gamma Ray Bursts but only 1599 of them are observed in all nine coordinates.

   Chattopadhyay and Maitra (2017) established that there are five distinct kinds of gamma ray bursts that are ellipsoidally-shaped in the logarithmic scale. The R data object `GRB-5groups.rda` contains this classification and the observations (in the logarithmic scale) for each of the 1599 GRBs (but we will not be using them in this exercise).

   (a) Provide a factor analysis of the dataset for $q = 1, 2, 3, 4$ factors. Use BIC to determine the number of optimal factors. [*10 points*]

   Although there are a number of packages available for doing factor analysis, I suggest the use of the `fad()` function in the R package `fad` with usage very similar to that of the `factanal()` function in base-R but with the added advantage of speed and its calculation of the BIC (as a component in the returned object). (It can also handle the case of high-dimensional, *i.e.* $p > n$, datasets.)

   (b) Describe the uniquenesses. How do they compare with the original variances? [*5 points*]

   (c) Describe the factor loadings at the optimal number of factors that are obtained by BIC. [*10 points*]

2. We will now characterize the variability of handwriting in the context of digits. Specifically, we will focus on the zip code (available in the form of a list of matrices, one for each digit at `zip-sep.rda`) and only on the digit "5" (the sixth item on the list). (For this reduced dataset, $n = 102 < 256 = p$ so that the `fad` package will need to be used.)

   (a) Use factor analysis with $q = 1, 2, \ldots, 15$ factors and BIC to decide on the number of factors. Note that if convergence is not achieved, you may need to run the function again to get around initialization issues (we will fix this in a later version). Display the factor loadings obtained and the uniquenesses by means of images. [*15 points*]

   (b) Intepret the results from (a) above. [*10 points*]

3. We go back to the psychological dataset of HW 5.

   (a) Perform a canonical correlation analysis between the academic variables and the psychological response variables. Comment. [*10 points*]

(b) Perform individual canonical correlation analyses between the academic and the psychological response sets of variables for each of the three program groups. Comment. [*15 points*]