PROBATIONARY FACULTY DEVELOPMENT GRANT-FINAL REPORT
SANTOSH KANDEL

Project Title:
A study of inference problems for non-linear functional data

Project Background and Objectives:
Alzheimer's disease is an age-related neurodegenerative disorder, which disturbs communication between various parts of the brain (Hani et al., 2019). This disease progresses from a healthy brain condition to a mild cognitive impairment and finally to irreversible neuronal loss. Hence, it is important to detect mild cognitive impairment at the early stage. Functional magnetic resonance imaging (fMRI) is considered a promising tool to detect Alzheimer's disease.The fMRI data - data consisting of signals of fMRI scans - is one of the most popular data used in the literature to study fMRI's use for Alzheimer's disease detection. Most of the existing consider the fMRI data as "linear" in nature and use the readily available tools from Functional Data Analysis (FDA). However, several studies (Worsley, 2005 and references therein) suggest that the fMRI data are "non-linear" in nature. Recently, this point of view has gotten a lot of attention (Petersen & Mueller, 2016). When the data is non-linear, the tools from FDA are not directly applicable. Recently, several methods have been proposed to study different aspects of non-linear data.
In this project, we propose two methodologies to study inference problems for non-linear data, in particular, to construct confidence regions for some parameters of interest. The main goal is to test the effectiveness of these methodologies using simulated data. The future goal is to use the proposed methodologies to study the fMRI data and get new insights.

Proposed Methodologies:
Recently, the study of non-linear data has attracted attention, and several problem specific methods are proposed to study such data. Hron et al. (2016) proposed using a particular transformation to transform non-linear data into linear data, analyze the data using tools from FDA, and use the inverse transformation to return back to the non-linear setting. Bigot et al. (2017) and Petersen and Mueller (2016) took a similar approach but proposed different transformations. Most of the studies in the literature are concerned with the so-called principal component analysis of non-linear data. They do not address any questions related to inference problems, such as estimation and construction of confidence regions. Recently, Petersen et al. (2021) initiated the study of inference problems for non-linear functional data. But, this study has several weaknesses: (1) it assumes the size of the data is large which may not be true for real-world data; (2) it is computationally expensive due to the large data size; and (3) the proposed transformations are ad-hoc and do not use the geometry of the data. In this project, we propose two methodologies to address the existing issues of large data size assumptions and look for data transformations. The first approach is to use the transformations considered by Petersen and Mueller (2016) to transform the non-linear data into linear data but apply new tools developed by Lopes et al. (2021) to study the inference problem in the linear setting and then apply the inverse transformation. The tools developed by Lopes et al. (2021) avoid issues related to the data size. The second approach is to use the so-called "Log" transformation to transform the non-linear data into linear data, then proceed as in the first approach.

Research Progress and Preliminary Results:
Here, we describe the current research activities to achieve the goal of the project. In this project, we studied how to construct a confidence region for the "mean" of non-linear data. The implementation of methodologies can be divided into three stages: (i) map the non-linear data into linear data, (ii) construct a confidence region of the transformed data in the linear setting, and (iii) use an inverse map to go back to the non-linear setting. In this project, we considered the data consisting of probability density functions supported on a compact interval. We successfully implemented stages (i) and (ii) for the first approach where we used the so-called log-quantile density (lqd) map (Petersen and Mueller (2016)) for the stage (i). The implementation of (iii) and the second approach is still ongoing research.

To implement stages (i) and (ii) for the first approach, we considered the following two scenarios: (a) The simulated data consisted of probabilities densities of the truncated normal distribution TN$(\mu, 1)$ where μ is random with uniform distribution on $[3, 4]$, (see Figure 1); and (b) the probability densities of the truncated normal distribution TN$(0.5, \sigma^2)$ with σ random and $\log(\sigma)$ with uniform distribution of $[-0.5, 0.5]$ (see Figure 2). All these numerical experiments were implemented through the programming language R. New R codes were written to conduct majority parts of experiments.

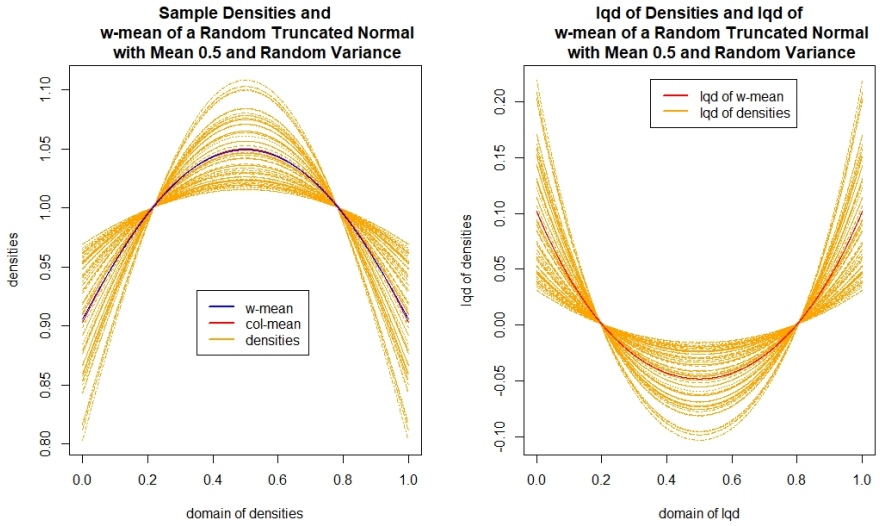

FIGURE 1. The figure on the left consists of sample densities drawn from $TN(0.5, \sigma^2)$ truncated on $[0, 1]$ with $\log(\sigma) \in \mathcal{U}[-0.5.0.5]$. The figure on the right consists of the sample paths of lqd transformation of the truncated normal densities.

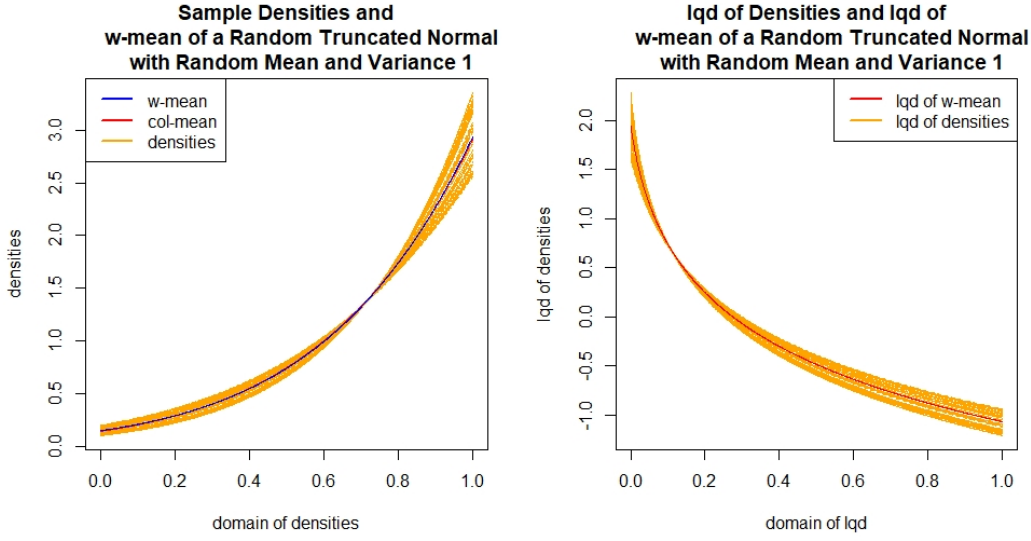

FIGURE 2. The figure on the left consists of sample densities drawn from $TN(\mu, 1)$ truncated on $[0, 1]$ with $\mu \sim \mathcal{U}[3, 4]$. The figure on the right consists of sample paths of lqd transformation of the truncated normal densities.

To compute the confidence region for the so-called Wasserstein mean, we considered confidence level $1 - \alpha$ where $\alpha \in \{0.01, 0.05, 0.1\}$ and sample size $n \in \{50, 200\}$. After applying the lqd, we computed the confidence region for the image of the Wasserstein mean and also the coverage probability this image. The preliminary results from the numerical experiment are shown in Table 1.

TABLE 1. Simulation Results

Setting	Random Component	Random Density	α	n	Coverage Probability
1	$\mu \sim \mathcal{U}[3,4]$	TN$(\mu, 1)$ on $[0,1]$	0.05	50	0.954
2	$\mu \sim \mathcal{U}[3,4]$	TN$(\mu, 1)$ on $[0,1]$	0.05	200	0.9365
3	$\mu \sim \mathcal{U}[3,4]$	TN$(\mu, 1)$ on $[0,1]$	0.01	50	0.985
4	$\mu \sim \mathcal{U}[3,4]$	TN$(\mu, 1)$ on $[0,1]$	0.01	200	0.978
5	$\mu \sim \mathcal{U}[3,4]$	TN$(\mu, 1)$ on $[0,1]$	0.1	50	0.916
6	$\mu \sim \mathcal{U}[3,4]$	TN$(\mu, 1)$ on $[0,1]$	0.1	200	0.875
7	$\log(\sigma) \sim \mathcal{U}[-0.5,.5]$	TN$(0.5, \sigma^2)$ on $[0,1]$	0.1	200	0.919
8	$\log(\sigma) \sim \mathcal{U}[-0.5,.5]$	TN$(0.5, \sigma^2)$ on $[0,1]$	0.1	50	0.919
9	$\log(\sigma) \sim \mathcal{U}[-0.5,.5]$	TN$(0.5, \sigma^2)$ on $[0,1]$	0.05	50	0.945
10	$\log(\sigma) \sim \mathcal{U}[-0.5,.5]$	TN$(0.5, \sigma^2)$ on $[0,1]$	0.05	200	0.953
11	$\log(\sigma) \sim \mathcal{U}[-0.5,.5]$	TN$(0.5, \sigma^2)$ on $[0,1]$	0.01	50	0.9865
12	$\log(\sigma) \sim \mathcal{U}[-0.5,.5]$	TN$(0.5, \sigma^2)$ on $[0,1]$	0.01	200	0.9875

To interpret the results, we compared the coverage probability, on the last column of Table 1, with $1-\alpha$, where α in the third last column. The coverage probabilities are closer to $1-\alpha$ which suggests that stage (i) and stage (ii) give satisfactory results for these numerical experiment. These preliminary results signal an optimism for the first approach.

Dissemination of Results:
I have written an abstract to present the result at Joint Mathematics Meetings 2024 in San Francisco.

Benefits to Students:
Two student researchers were hired to assist with the project. Student researchers learned R programming language and assisted in conducting numerical experiments.

Outlook:
We will continue working on this project. In the future, we plan to examine the efficacy of the second approach proposed in the project through simulation studies.

Acknowledgments:
I would like to thank the Probationary Faculty Development Grant Awards program for support.

References:
Bigot, J., Gouet, R., Klein, T., and Lopez, A. (2017). Geodesic PCA in the Wasserstein space by convex PCA. Annales de l'Institut Henri Poincare B: Probability and Statistics, 53, 1–26.

Degras, D. A. (2011). Simultaneous confidence bands for nonparametric regression with functional data. Statistica Sinica, 1735–1765.

Hani, H. S., Ata, E., Abbas, B.-F. (2019). Identification of the early stage of Alzheimer's disease using structural MRI and resting-state fMRI. Frontiers in Neurology, 10.

Hron, K., Menafoglio, A., Templ, M., Hruzova, K., and Filzmoser, P. (2016). Simplicial principal component analysis for density functions in Bayes spaces. MOX-Report, 25.

Lopes, M. E., Lin, Z., and Mueller, H.-G. (2021). Bootstrapping max statistics in high dimensions: Near parametric rates under weak variance decay and application to functional and multinomial data. Annals of Statistics, 48, 1214–1229.

Petersen, A., and Mueller, H.-G. (2016). Functional data analysis for density functions by transformation to Hilbert space. Annals of Statistics, 44, 183–218.

Petersen, A., Liu, X., and Divani, A. A. (2021). 'Wasserstein f-tests and confidence bands for the Fréchet regression of density response curves.' The Annals of Statistics, 49, 590–611.

Worsley, K. J., Chen, J.-I., Lerch, J. and Evans, A. C. (2005). Comparing functional connectivity via thresholding correlations and singular value decomposition. Philosophical Transactions of the Royal Society B: Biological Sciences 360, 913–920.