Group pattern detection of longitudinal data using functional statistics

Rongjiao Ji, Alessandra Micheletti, Natasa Krklec Jerinkic, Zoranka Desnica

Abstract Estimations and evaluations of the main patterns of time series data in groups benefit large amounts of applications in various fields. Different from the classical auto-correlation time series analysis and the modern neural networks techniques, in this paper we propose a combination of functional analysis of variance (FANOVA) and permutation tests in a more intuitive manner for a limited sample size. First, FANOVA is applied in order to separate the common information and to dig out the additional categorical influence through paired group comparison, the results of which are secondly analyzed through permutation tests to identify the time zones where the means of the different groups differ significantly. Normalized kernel functions of different groups are able to reflect remarkable mean characteristics in grouped unities, also meaningful for deeper interpretation and group-wise classification. In order to learn whether and when the proposed method of FANOVA and permutation F-test works precisely and efficiently, we compare the estimated kernel results with the ground truth on simulated data. After the confirmation of the model’s efficiency from simulation, we apply it also to the RAVDESS facial dataset to extract the emotional behaviors of humans based on facial muscles contractions (so-called action units (AU) technically in computer graphics), by comparing the neutral performances with emotional ones.

Key words: functional analysis of variance (FANOVA), permutation F-test, emotion, expression evolution, action units

Rongjiao Ji, Alessandra Micheletti
Università degli Studi di Milano, e-mail: rongjiao.ji@unimi.it,alessandra.micheletti@unimi.it

Natasa Krklec Jerinkic
University of Novi Sad, e-mail: natasa.krklec@dmi.uns.ac.rs

Zoranka Desnica
3Lateral DOO, e-mail: zoranka.desnica@3lateral.com
1 Introduction

Many natural, economical, or technological phenomena can be represented as real functions \( f : \mathbb{R} \rightarrow \mathbb{R} \) where the domain can be time or another one-dimensional variable, and data of many copies of these functions can be collected, usually in a discrete format.

We here consider the case where we collect data of a set of random functions describing a real phenomenon, which is characterized by the presence of groups, where each group is characterized by a different mean underlying pattern that we want to detect.

Examples of such situations are: the collection of temperature evolution in different regions of the world, in the same time window, where we want to study the underlying temperature evolution functions characterizing each region and their main differences; the collection of stock price evolution, with stocks related with goods of different economical sectors, or different segments of the same sector, and we want to extract and compare the underlying evolution patterns of the sector or segments; etc.

In this paper we will face the problem of underlying pattern detection and comparison of groups of random functions, using a specific driving example: the detection of human emotions from functions describing the movements of facial muscles during a video.

The study of human facial expressions and emotions never stops in our daily life while we communicate with others. Following the increased interest in automatic facial behavior analysis and understanding, the need of a semantic interpretation of the evolution of facial expressions and of human emotions has become of interest in recent years \[2\]. In this paper, based on a work cooperated with the Serbian company 3Lateral, which has special expertise on building visual styles and designs in animation movies, we want to explore functional statistical instruments to identify the emotions while analyzing the expressions through recorded videos of human faces. The final aim of this research is to use this information to better and more realistically establish virtual digital characters, able to interact autonomously with real humans.

The data that we consider are multivariate longitudinal data, showing the evolution in time of different face muscles contraction. Functional Data Analysis (FDA) offers the possibility to analyze the entire expression evolution process over time and to gain detailed and in-depth insight into the analysis of emotion patterns. The basic idea in functional data analysis is that the measured data are noisy observations coming from a smooth function. Ramsay and Silverman \[4\] describe the main features of FDA, that can be used to perform exploratory, confirmatory or predictive data analysis. Ullah and Finch \[5\] published a systematic review on the applications of functional data analysis, where they included all areas where FDA was applied.

Here we will apply models of functional regression and functional analysis of variance (FANOVA) to identify the mean functional patterns related with different emotions. Additionally we setup a technique, based on permutation tests, to identify the time zones where the identified patterns are significantly different. The identified
patterns can then be used to perform an automatic detection and classification of the emotions expressed by a human face.

The methods here proposed are also tested on simulated data to highlight their effectiveness in identifying the right underlying patterns and the correct time zones in which the patterns differ, also in relationship with the level of noise present in the collected data.

In spite of the specific motivating case study here considered, the methods here proposed and the related software can be applied to many other similar situations in different fields of application.

2 Methods

2.1 Functional Data Analysis

Functional Data Setting Let \( \Omega \) be a population space of statistical units (\( \omega \in \Omega \)) and \( s_n = \{\omega_1, \omega_2, \ldots, \omega_n\} \) be a random sample of size \( n \) drawn from \( \Omega \). Let \( Y \) be a \( L_2 \)-continuous multivariate stochastic process defined on \( \Omega \) associating to a set of \( D \) curves, \( D \geq 2 \), where each curve in one dimension defined on a given finite interval \( \mathcal{T} = [0, T], 0 < T < \infty \), i.e., \( Y(\omega) = [y_{\omega,1}(t), \ldots, y_{\omega,D}(t)]^T \) (denoted as \( Y = (y_1, \ldots, y_D)^T \) for simplicity), \( t \in \mathcal{T} \), such that

\[
\lim_{h \to 0} \mathbb{E}[||Y(t+h) - Y(t)||^2] = \lim_{h \to 0} \int_0^T \sum_{d=1}^D \mathbb{E}[(y_{d}(t+h) - y_{d}(t))^2]dt = 0.
\]

The observations of \( Y \) on the sample \( s_n \) are called multivariate functional data, providing independent realizations as samples \( \{Y(\omega_1), \ldots, Y(\omega_n)\} \) with multivariate curves \( 3 \). The mean functions of \( y_d \) and \( Y \) are

\[
\mathbb{E}[y_d(t)] = \mu_d(t), t \in [0, T], \text{ and } \mathbb{E}[Y] = \mu(t) = (\mu_1(t), \ldots, \mu_D(t))^T.
\]

The covariance operator of \( Y \) is then defined as an integral operator \( \mathcal{C} \) with kernel

\[
\mathcal{C}(t,s) = \mathbb{E}[(Y(t) - \mu(t)) \otimes (Y(s) - \mu(s))], \forall t,s \in [0, T],
\]

where \( \otimes \) is the tensor product on \( \mathbb{R}^D \). Thus, \( \mathcal{C}(t,s) \) is a \( D \times D \) matrix with elements

\[
\mathcal{C}(t,s)[i,j] = \text{Cov}(Y_i(t), Y_j(s)), i,j = 1, \ldots, D.
\]

Functional Principal Component Analysis (FPCA) Under the hypothesis of \( L_2 \)-continuity, \( \mathcal{C} \) is an Hilbert-Schmidt operator, i.e., it is compact and self-adjoint. Meanwhile, the spectral analysis of \( \mathcal{C} \) provides a countable set of positive eigenvalues \( \{\eta_{i,j} \geq 1\} \) associated to an orthonormal basis of multivariate eigen-functions \( \{f_{i,j} \geq 1\}, f_i = [f_{i,1}, \ldots, f_{i,D}]^T, \) such that
$$\mathbf{C} f_j = \eta_j f_j,$$

with $\eta_1 \geq \eta_2 \geq \ldots$ and $\langle f_i, f_j \rangle_{L^2([0,T])} = \delta_{i,j}$ with $\delta_{i,j} = 1$ if $i = j$, otherwise $\delta_{i,j} = 0$. Assume that the principal components $\eta_j$ of $Y$ are zero-mean random variables, as the projections of $Y$ on the $j$th eigenfunction of $\mathbf{C}$:

$$\eta_j = \int_0^T (Y(t) - \mu(t), f_j(t)) dt = \int_0^T \sum_{d=1}^D (Y_d(t) - \mu_d(t)) f_{j,d}(t) dt,$$

which means $Y(t)$ varies around the mean function with random amplitude variations in the directions of $\{f_j\}_{j \geq 1}$. Therefore, following Karhunen-Loeve expansion, it holds that

$$Y(t) = \mu(t) + \sum_{j \geq 1} \eta_j f_j(t) \text{ and } y_d(t) = \mu_d(t) + \sum_{j \geq 1} \eta_j f_{j,d}(t)$$

are the best approximations of this form under the mean square criterion.

**Functional Registration** In practice, it is essential to align observed functions into a common registered internal timeline to isolate the phase variability of the functions, but keeping, at the same time, the amplitude-phase unchanged to maintain the information under interests. The phase variation is normally represented by a random change of time scale, which is mostly a non-linear transformation. We use the warping functions $T_i : [0,T] \rightarrow [0,T], i = 1, \ldots, n$, assuming that they are increasing functions independent of amplitude variation. They map unregistered chronological time $t^*_i$ to registered internal time $t$ so that $T_i^{-1}(t^*_i) = t$, with $E[T_i(t)] = t$.

The observed time-warped curves, represented through a Karhunen-Loeve expansion based on a functional basis $f_{j,d}$, are

$$\tilde{y}_{i,d}(t) = y_{i,d}(T_i^{-1}(t^*_i)) = \mu_d(T_i^{-1}(t^*_i)) + \sum_{j \geq 1} \eta_j f_{j,d}(T_i^{-1}(t^*_i)).$$

The literature is rich of proper methods to do this, and a more in-depth review would be beyond the scope of this report. Briefly speaking, we followed the principal components based registration method [7] with a generative process, by alternating the process between using the generalized FPCA to calculate template functions and estimating smooth time warping functions that map observed curves to templates [1].

**Approximation into finite basis of functions** Often in practice, data are observed at discrete time points and with some noise. In order to get the functional feature of data, researchers normally use some smoothing and interpolation methods to rebuild the consecutive function from the discrete data, while assuming that the latent function belongs to a finite dimensional functional space spanned by some functional basis. In this way, functional data objects are constructed from time series data by specifying a set of basis functions and a set of coefficients defining a linear combination of these basis functions. For non-periodic time series, B-spline basis functions which contain polynomial segments jointed end-to-end at argument val-

1 Codes are available in the R package “registr” [6]
ues are frequently considered, benefiting the most from fast computational ability and great flexibility [4].

### 2.2 Multiple Multivariate Function-on-Scalar Regression

Let \( y_{g,d,k}(t) \) denote the evolution of one specific variate \( d \in \{1, \ldots, D\} \) in sample \( k \in \{1, \ldots, K\} \) for one group \( g \in \{0, \ldots, G\} \). We can assume that

\[
y_{g,d,k}(t) = \mu_{d,0}(t) + \alpha_{d,g}(t) + \varepsilon_{g,d,k}(t),
\]

where \( \mu_{d,0}(t) \) is the grand mean function independent from groups. The term \( \alpha_{d,g}(t) \) is the additional effect function of group \( g \) on the considered variate \( d \), while \( \varepsilon_{g,d,k}(t) \) represents the unexplained zero-mean variation, specific of the \( k \)-th sample within group \( g \). In order to identify the parameter functions uniquely, we require the constraint \( \sum_{g=0}^{G} \alpha_{d,g}(t) = 0, \forall t \).

By grouping the samples from the same group, we can define a \((GK + 1) \times (G + 1)\) design matrix \( Z_d \) for univariate model, with suitable 0 and 1 entries, as described in [4] Section 9.2, and rewrite Equation (3) in matrix form

\[
y_d(t) = Z_d \beta_d(t) + \varepsilon_d(t), \forall t,
\]

where

\[
y_d(t) = \begin{bmatrix} y_{0,d,1}(t) \\ \vdots \\ y_{0,d,K}(t) \\ y_{1,d,1}(t) \\ \vdots \\ y_{1,d,K}(t) \\ \vdots \\ y_{G,d,1}(t) \\ \vdots \\ y_{G,d,K}(t) \\ 0 \end{bmatrix}, Z_d = \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 1 & 0 & 0 & \cdots & 1 \\ 0 & 1 & 1 & 1 & \cdots & 1 \end{bmatrix}, \beta_d(t) = \begin{bmatrix} \mu_{d,0}(t) \\ \alpha_{d,0}(t) \\ \alpha_{d,1}(t) \\ \vdots \\ \alpha_{d,G}(t) \\ \beta_{d,0}(t) \\ \beta_{d,1}(t) \\ \vdots \\ \beta_{d,G+1}(t) \end{bmatrix}, \varepsilon_d(t) = \begin{bmatrix} \varepsilon_{0,d,1}(t) \\ \vdots \\ \varepsilon_{0,d,K}(t) \\ \varepsilon_{1,d,1}(t) \\ \vdots \\ \varepsilon_{1,d,K}(t) \\ \vdots \\ \varepsilon_{G,d,1}(t) \\ \vdots \\ \varepsilon_{G,d,K}(t) \end{bmatrix}.
\]

Here \( \beta_d \) is defined as the corresponding set of \( G + 1 \) regression functions for mean and \( G \) additional groups, and \( \varepsilon_d \) is the matrix of noise with the same dimension of \( y_d \).

Following the notations above, we can write the multivariate matrix form for the multiple regression functions, as
\[ Y(t) = Z\beta(t) + \epsilon(t), \]

where

\[
Y(t) = \begin{bmatrix}
y_1(t) \\
\vdots \\
y_d(t) \\
y_D(t)
\end{bmatrix},
Z = \begin{bmatrix}
Z_1 \\
\vdots \\
Z_d \\
Z_D
\end{bmatrix},
\beta(t) = \begin{bmatrix}
\beta_1(t) \\
\vdots \\
\beta_d(t) \\
\beta_D(t)
\end{bmatrix},
\epsilon(t) = \begin{bmatrix}
\epsilon_1(t) \\
\vdots \\
\epsilon_d(t) \\
\epsilon_D(t)
\end{bmatrix}.
\]

Under the functional least squares fitting criterion on \( R^{GD_K \times T} \) space, we can estimate the parameters

\[ \hat{\beta}(t) = \hat{B}\Theta(t), \]

where \( \Theta(t) = (\theta_1(t), \ldots, \theta_q(t), \ldots, \theta_Q(t))^T \) denotes a linear combination of basis functions with \( J_{\theta \theta} \) as the inner product of each pair of functional basis, and \( \hat{B} \) is the estimated coefficient matrix with

\[ \text{vec}(B^T) = [(Z \otimes J_{\theta \theta}^\frac{1}{2})(Z \otimes J_{\theta \theta}^\frac{1}{2})^{-1}(Z \otimes J_{\theta \theta}^\frac{1}{2})^T \text{vec}(J_{\theta \theta}^\frac{1}{2}A^T)]. \]

Here, \( A \) is the coefficient matrix of samples \( Y(t) \) with a fixed B-spline functional basis and more computation details in matrix format are shown in Appendix 3.1.

### 2.3 Functional ANOVA and F-type tests

After establishing the model for functional evolution processes and estimating the additional parameter functions affected by group categories, next goal is to investigate whether and when group labels are significantly influencing the change of variables. ANOVA-related methods are suitable tools to determine the effects of independent categorical variables on continuous dependent variables, by comparing inter-group differences between centroids represented as the vectors of the mean values of the dependent variables. Moreover, in order to construct feasibly more general hypothesis tests for group means, contrasts, defined as a linear combination of regression coefficients where the values of the linear combination sum up to 0, are applied by specifying the values on different pairs of groups gradually.

For each subgroup of variable \( d \in \{1, \ldots, D\} \) and group \( g \in \{0, \ldots, G\} \), assume that the \( k \)-th sample curve in this subgroup \( y_{d,g,k}(t) \sim \mathcal{N}(\mu_{d,g}(t), \sigma^2) \), \( k \in \{1, \ldots, K\} \) with group mean \( \mu_{d,g}(t) \) and constant variance \( \sigma^2 \), and therefore the subgroup sample mean \( Y_{d,g}(t) = \frac{1}{K} \sum_{k=1}^K y_{d,g,k}(t) \sim \mathcal{N}(\mu_{d,g}(t), \frac{\sigma^2}{K}) \). Here denote \( \mu(t) \) as the sequence of subgroup means \( \mu_{d,g}(t) \) for all \( d \in \{1, \ldots, D\} \) and \( g \in \{0, \ldots, G\} \). We construct one test for each pair of groups in comparison (one control group \( g = 0 \) and another group under study \( g = \bar{g} \) for one variable \( d \)) and in \( i \)-th test
(\(i \in \{1, \ldots, DG\}\)) we design a vector \(a = [0, \ldots, 0, 1, 0, \ldots, 0, -1, 0, \ldots, 0]\) in length \(D(G + 2)\) (same as the length of \(\beta(t)\) in last section) with the entries pointed to the pair of groups in comparison being 1 and -1 respectively, i.e., \(a_{d(G+2)+2} = 1\), \(a_{d(G+2)+2+g} = -1\), and 0 in entries remained. We define the contrast in \(i\)-th test as \(c = a^T \mu(t) = \alpha_{d,0}(t) - \alpha_{d,g}(t) = (\alpha_{d,0}(t) + \mu_{d,0}) - (\alpha_{d,g}(t) + \mu_{d,0})\) for \(DG\) different group comparisons, where the sum of weights in \(a\) is zero, and then the difference of subgroup sample means \(Y_{d,0}(t) - Y_{d,g}(t)\) is an unbiased estimator of the contrast \(c\). Following this setting, the null hypothesis at each time point \(t\) for the control group and the group under study \(g\) in variable \(d\) is

\[ H_0: a^T \beta(t) = 0 \Leftrightarrow \alpha_{d,0}(t) - \alpha_{d,g}(t) = 0, \]

and the related F statistics under null hypothesis is

\[ F_{d,g}(t) = \frac{(Y_{d,0}(t) - Y_{d,g}(t))^2}{S_p^2}, \]

where the sum of sample variances is \(S_p^2 = \frac{1}{N-DG} \sum_{g=1}^G \sum_{d=1}^D \sum_{k=1}^K (y_{d,k,g}(t) - Y_{d,g}(t))^2\)

with sample size \(N\). When assuming the functional data \(y_{d,k,g}(t)\) as realizations of Gaussian processes, for any given \(t \in \mathcal{T}\), \(F_{d,g}(t) \sim F_{1, N-DG}\), where \(F_{1, N-DG}\) denotes \(F\)-distribution with 1 and \(N-DG\) degrees of freedom.

The statistic \(F_{d,g}(t)\) should be compared with a critical value of \(F_{1, N-DG}\) for any predetermined significant level \(\alpha\). The critical value of \(F_{1, N-DG}(\alpha)\) represents the \(\alpha\)-fraction of rejection area, where \(F_{1, N-DG}(\alpha)\) denotes the upper 100\(\alpha\) percentile of \(F_{1, N-DG}\). We write the \(1 - \alpha\) quantile of the distribution of \(F\) statistic by \(c_{1-\alpha}\), and given \(c_{1-\alpha}\), we reject the null hypothesis when \(F_{d,g}(t) > c_{1-\alpha}\), which indicates the dissimilarity of group means under comparison. First, we can test the statistic at all points of \(\mathcal{T}\) using the same critical value from standard \(F\)-Distribution table with correlated degrees of freedom. Since complex assumptions and convoluted formula are not often satisfied in practice for most classic statistical tests, we consider also a nonparametric test, permutation test, which is light on assumptions, widely applicable and intuitive friendly. The permutation distribution is obtained by shuffling the labels within the paired groups in comparison and calculate the \(F\) statistic \(F_{d,g}(t)\) many times to approximate to the statistic’s distribution. More inference of permutation test can be seen in Appendix 4.

### 2.4 Kernel-based Classification

After the explanation of functional linear regressions and test methods, we are able to obtain the ground mean functions and the additional group-wise functions revealing the latent group effects on variable curves. Most importantly, we can detect the time zones where the additional effects impact significantly in statistical sense. To
Fig. 1: Post hoc analysis by classifying from a given video

Further examine the efficiency of the information, a natural attempt would be to apply a post hoc analysis: do classifications using the knowledge obtained and check the correctness. Before the classification, four steps are needed in our framework to project the input functions into normalized scores in lower dimensions:

- First, centralize the input evolution by subtracting the estimated mean function for each sample $k$ on variable $d$.

$$\tilde{y}_{d,k}(t) = y_{d,k}(t) - \mu_{d,0}(t)$$

- Secondly, use the centralized evolution to make inner product with $G$ estimated additional group-wise functions (also called kernels for simplicity) in turn, and thus the evolution function is projected to a vector space with $G$ dimensions. Let $S_{g,d,k}$ represent the score of group $g$ on variable $d$ in sample $k$:

$$S_{g,d,k} = (\tilde{y}_{d,k}(t), \hat{\alpha}_{g,d}(t)) = \int_0^T w_{g,d}\tilde{y}_{d,k}(t)\hat{\alpha}_{g,d}(t)dt = \int_0^I w_{g,d}\tilde{y}_{d,k}(t)\hat{\alpha}_{g,d}(t)dt,$$

where $I$ represents the union of relevant intervals, and $w_{g,d}$ is the weights in the integral process, constant in time.
Furthermore, normalize the scores across all groups and variables from one sample $k$, so that $\tilde{S}_{g,d,k} \in [0,1]$, in order to keep the relative rate/percentage across groups.

- If $S_{g,d,k} < 0$ for some $g$, move them to positive numbers by subtracting the minimum ($\min\{0, \min_g S_{g,d,k}\}$).
- Then, rescale them to be smaller than 1 across groups.

$$\tilde{S}_{g,d,k} = \frac{S_{g,d,k} - \min\{0, \min_g S_{g,d,k}\}}{\sum_{g=1}^G [S_{g,d,k} - \min\{0, \min_g S_{g,d,k}\}]}.$$

In this way $\tilde{S}_{g,d,k}$ represents the probability that variable $d$ in sample $k$ behaves as in (has been generated by) group $g$. If the sample $k$ that we are considering is labeled with $g$, more importance is given to the variables that assign a highest probability $\tilde{S}_{g,d,k}$ to the group $g$.

- Finally, combine the scores across variables for sample $k$ (labeled with $g$) by a linear convex combination with the trained weights $r_{l,g}$, such that, if the label of sample $k$ is $g$, then $S_{g,k}$ is the maximum score.

$$S_{\text{comb},g,k} = r_{1,g} \tilde{S}_{g,1,k} + r_{2,g} \tilde{S}_{g,2,k} + \cdots + r_{L,g} \tilde{S}_{g,L,k},$$

with $r_{l,g} \in [0,1]$ and $\sum_{l=1}^L r_{l,g} = 1$. For sample $k$, $[S_{1,k}, S_{2,k}, \ldots, S_{G,k}]$ contains the scores of all groups under study, and all the elements of this vector fall in region $[0,1]$.

For classification based on scores, we apply SVM method on scores $[S_{1,k}, S_{2,k}, \ldots, S_{G,k}]$ to classify the samples and examine the kernels efficiency. A further study on the classification methods is important but is skipped here for space saving because the classification is not the focus of this paper, although we will explore further in next study.

3 Details of Methods

3.1 Approximation into a finite basis of functions

Assume that each curve $y_{g,d,k}(d \in \{1, \ldots, D\}$ for $k$-th video can be expressed as a linear combination of basis functions $\Theta(t) = (\theta_1(t), \ldots, \theta_q(t), \ldots, \theta_Q(t))^T$. We define $J_{\theta\theta}$ to represent the inner product of each pair of functional basis.

$$y_{g,d,k}(t) = \sum_{q=1}^Q a_{g,d,k,q} \theta_q(t) = a_{g,d,k} \Theta(t),$$
\[ y_{g,d}(t) = A_{g,d} \Theta(t), \quad A_{g,d} = [a_{g,d,1}, \ldots, a_{g,d,K}]^T \]

\[ Y_g(t) = A_g \Theta(t), \quad A_g = [A_{g,1}, \ldots, A_{g,d}, \ldots, A_{g,D}]^T \]

\[ Y = A \Theta(t), \quad A = [A_1, \ldots, A_g, \ldots, A_G]^T \]

with \( A_g = [A_{g,1}, \ldots, A_{g,d}, \ldots, A_{g,D}]^T \), \( A_{g,d} = \begin{bmatrix} a_{g,d,1,1} \cdots a_{g,d,1,Q} \\ \vdots & \vdots & \vdots \\ a_{g,d,K,1} \cdots a_{g,d,K,Q} \end{bmatrix} \).

Similarly, the coefficient function \( \beta = [\beta_1(t), \ldots, \beta_D(t)]^T \) can be decomposed with the same functional basis in the following way:

\[ \beta = B \Theta(t), \]

with \( B = [B_1, \ldots, B_d, \ldots, B_D]^T \), \( B_d = \begin{bmatrix} b_{1,d,1} \cdots b_{1,d,Q} \\ \vdots & \vdots & \vdots \\ b_{G,d,1} \cdots b_{G,d,Q} \end{bmatrix} \).

Consider that in multivariate linear regression setting

\[ Y = Z \beta + E, \]

by substituting the functional curves with coefficient matrix and functional basis, we have

\[ A \Theta(t) = ZB \Theta(t) + E \]

Notice that \( B \) is the coefficient that we want to estimate. Here by applying least square method on \( R^{GDK \times T} \) space, we have
Group pattern detection of longitudinal data using functional statistics

\[ J(B) = \|E\|_{L_2}^2 = \sum_{g \in \{1, \ldots, G\}, \ d \in \{1, \ldots, D\}, \ k \in \{1, \ldots, K\}} \|y_{g,d,k}(t) - \hat{y}_{g,d,k}(t)\|_{L_2}^2 \]  

\[ = \int \|A(\Theta(t) - ZB(\Theta(t))\|^2 dt \]  

\[ = \int \| (A - ZB)\Theta(t)\|^2 dt \]  

\[ = \int trace[(((A - ZB)\Theta(t))^T ((A - ZB)\Theta(t))] dt \]  

\[ = \int trace[\Theta(t)^T (A - ZB)^T (A - ZB)\Theta(t)] dt \]  

\[ = trace[\Theta(t)^T (A - ZB)^T (A - ZB)\Theta(t)] dt \]  

\[ = trace[(A - ZB)^T (A - ZB) \int \Theta(t)^T \Theta(t) dt] \]  

\[ = tr[(((A - ZB)J_{\theta \theta})^T (A - ZB)J_{\theta \theta}^T] \]  

\[ = vec((A - ZB)J_{\theta \theta})^T vec((A - ZB)J_{\theta \theta}) \]  

\[ = \|vec((A - ZB)J_{\theta \theta})^T \|^2 \]  

\[ = \| vec(J_{\theta \theta} A^T) - vec(Z \otimes J_{\theta \theta} vec(B^T)) \|^2 \]  

\[ = \| vec(J_{\theta \theta} A^T) - (Z \otimes J_{\theta \theta} vec(B^T)) \|^2 \]  

Let \( J(B) = 0 \), we can derive that

\[ vec(B^T) = [(Z \otimes J_{\theta \theta})^T (Z \otimes J_{\theta \theta})]^{-1} (Z \otimes J_{\theta \theta})^T vec(J_{\theta \theta} A^T), \]

together with

\[ C_{mat} = (Z \otimes J_{\theta \theta})^T (Z \otimes J_{\theta \theta}) \]

\[ = (Z^T Z) \otimes (J_{\theta \theta} J_{\theta \theta}) \]

\[ D_{mat} = (Z \otimes J_{\theta \theta})^T vec(J_{\theta \theta} A^T) \]

4 Permutation Test

Let’s denote \( \Phi_n \) as the set of all possible permutations of \( 1, \ldots, n \), \( M_n \) as the cardinality of \( \Phi_n \). A permutation \( \phi \overset{A}{\equiv} (\phi_1, \ldots, \phi_n) \in \Phi. \)
$X_n^\phi$, the permuted version of $X_n$, $X_n^\phi \triangleq (X_1^\phi, \ldots, X_n^\phi)$. $F_{r_n}^\phi = F_{r_n}(X_n^\phi)$ to denote the F-test statistic computed based on the permutations $X_n^\phi$.

Let $F_{r_n}^\phi(s)$ be the permutation distribution function of $F_{r_n}^\phi$, defined as

$$F_{r_n}^\phi(s) = M_n^{-1} \sum_{\phi \in \Phi_n} 1_{F_{r_n}(X_n^\phi) \leq s}.$$

We write the $1 - \alpha$ quantile of $F_{r_n}^\phi$ by $c_{1-\alpha,n}$, defined as $c_{1-\alpha,n} \triangleq \inf\{t : F_{r_n}^\phi(t) \geq 1 - \alpha\}$.

Given the quantile $c_{1-\alpha,n}$, the permutation test rejects the null hypothesis when $Fr_n > c_{1-\alpha,n}$.

5 Acknowledgements

This work was funded by European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska Curie grant agreement No 812912 for the project BIGMATH. Computational resources have been provided by the INDACO Platform, which is a project of High Performance Computing at the University of Milan.

References

1. José Ramón Berrendero, Ana Justel, and Marcela Svarc. Principal components for multivariate functional data. *Computational Statistics & Data Analysis*, 55(9):2619–2634, 2011.
2. A.J. Fridlund. Human facial expression: An evolutionary view. *Academic Press*, 2014.
3. Julien Jacques and Cristian Preda. Clustering multivariate functional data. In *COMPSTAT 2012*, pages 353–366, 2012.
4. J. Ramsay and B.W. Silverman. *Functional data analysis*. Springer, 1997.
5. S. Ullah and F. F. Caroline. Applications of functional data analysis: A systematic review. *BMC medical research methodology*, 2013.
6. J. Wrobel. Register: Registration for exponential family functional data. *Journal of Open Source Software*, 3(22):557, 2018.
7. J. Wrobel, V. Zippunnikov, J. Schrack, and J. Goldsmith. Registration for exponential family functional data. *Biometrics*, 75(1):48–57, 2019.