Probability Distribution of Functional Principal Component Scores and Modal Function

Adhiraj Mandal, Surajit Ray

School of Mathematics and Statistics, University of Glasgow

October 21, 2024

1 Abstract

This paper develops a comprehensive framework for deriving the probability distribution of functional principal component (FPC) scores for random functions, expressed in the $\mathcal{L}_2$ space via orthonormal basis expansion. We investigate two distinct approaches for representing these scores: (1) utilizing a random number of principal components, and (2) employing a finite but deterministic number of principal components. In the first setup, we show that the stochastic nature of the number of components results in a distribution function of mixed type, which is neither fully discrete nor continuous. Consequently, a probability density function cannot be defined in this case. However, in the second setup—where the number of components is fixed and finite—we can gradually increase the number of components to infinity. This approach yields a continuous distribution function for the FPC scores, enabling us to define and obtain a probability density function.

Additionally, we extend the analysis by leveraging the derived density function to introduce a new method for estimating the functional mode (or modal function) for functional data. This definition allows for a novel mode estimation technique that avoids the need for explicitly calculating principal component scores. Instead, the modal function is determined directly from the coefficients of the orthonormal basis expansion. This approach simplifies the computational process while providing a robust framework for characterizing modes in complex functional data. The theoretical results and methodology are applicable in various domains of functional data analysis, offering new insights into the probabilistic structure of FPC scores and facilitating more efficient computations of key statistical features.

2 Introduction

Functional data analysis (popularly abbreviated as FDA) is a branch of Statistics that attempts to analyse information that varies over a continuum. Such processes are often considered to be functions of a parameter. The parameter is often referred to as 'time', however, it can be any other domain, such as energy, wavelength, spatial location etc. In functional data analysis, each curve is considered to be an individual entity, instead of a number of observations along the curve. The origins of techniques for functional data analysis can be traced back to studies conducted by Grenander (1950). It is a relatively new field of study, and the term itself was coined by Ramsay and Dalzell (1991). In 2008, Ramsay and Silverman (see Ramsay and Silverman (2008)) developed a theoretical framework for analysing functional data and utilised the strategies for analysing real-world datasets. An example of such an analysis of a real-world data set can be found at Ramsay et al. (1996), where functional data analysis techniques were applied to movements of lips during speech was studied. Such statistical techniques can be important to a range of different disciplines including automatic speech recognition, experimental phonetics, speech motor control and language pathology. The R-package `fda` was developed in 2014 (see Ramsay et al. (2014)) to aid analysis of functional data.

However, the theoretical framework for statistically analysing functional data is requires further development. One of the major drawbacks of functional data is that unlike univariate or multivariate data, a probability density function is not very clearly defined for random functions. Delaigle and Hall (2010) showed that a general probability density function can not be defined for all functional data. This makes extension of standard multivariate statistical analysis techniques in the realm of functional data very challenging. Lin et al. (2018) showed that when imposed with six regularity conditions and considered to be elements of a mixture inner product space, a probability density function can be defined for functional data.

The popular measures of central tendency (mean, median and mode) for univariate and multivariate data are heavily reliant on the probability density functions.Suppose, X is a n-dimensional continuous multivariate random variable with joint probability density function $f(x)$. The measures of central tendency are defined as follows.

Mean : $\mu = \mathbb{E}(x) = \int_{\mathbb{R}^n} x.f(x)dx$.

Median : $m = \arg\min_{z\in\mathbb{R}^n} \mathbb{E}[||X - z||] = \arg\min_{z\in\mathbb{R}^n} \int_{\mathbb{R}^n} ||x - z||.f(x)dx$.

Mode : $M = \arg\max_{x\in\mathbb{R}^n} f(x)$.

Thus, the difficulty in definition of the probability density function f for functional data translates immediately to difficulty in the definition of the measures of central tendency. Nevertheless, popular statistical analysis techniques, such as Principal Component Analysis (PCA) are readily extendable to Functional Principal Component Analysis (FPCA). Delaigle and Hall (2010) defined the functional mode (alternatively known as modal function) by exploiting the concept of functional principal component scores. In this article we shall first obtain a theoretical probability density function for functional principal component scores and then use the definition of Delaigle and Hall (2010) to derive a closed form for the modal function.

3 Defining Modal Function

Suppose f is the density of a functional random variable X defined on the $\mathcal{L}^2$ space. Lin et al. (2018) showed that such a density can be defined on the $\mathcal{L}^2$ space, by considering them to be elements of a mixture inner product space. Ferraty and Vieu (2006, pg. 130), defined functional mode to be the solution (under the assumption of existence) of $\sup_X f(X)$. However, computation of the mode calls for the estimation of density function f along with the calculation of it's supremum. For the estimation of the density function, an implementation of a pseudo kernel-type estimator has been proposed by Ferraty and Vieu (2006, pg. 130). Let $X_{[1]}, X_{[2]}, \ldots, X_{[n]}$ be a random sample. $\forall\chi$, we define,

$$\tilde{f}(\chi) = \frac{1}{Q(K,h)} \sum_{i=1}^{n} K\left(\frac{1}{h} d(\chi, X_{[i]})\right),$$

where, $Q(K,h)$ is a positive quantity independent of X, K is an appropriate kernel and d is a semi-metric defined on the space of functions. We should note that $\tilde{f}$ is not an estimator, as $Q(K,h)$ is not known. However, this aids in the estimation of the modal function as the maximisation of $\tilde{f}$ is essentially the maximisation of $\sum_{i=1}^{n} K(h^{-1}d(\chi, X_{[i]}))$. An estimation of the mode on the sample space $\mathcal{S}$ is given by,

$$X_{mode,\mathcal{S}} = \arg\max_{\xi\in\mathcal{S}} \tilde{f}(\xi). \tag{1}$$

Delaigle and Hall (2010) however, exploited the concept of functional principal component scores to provide an alternative technique for estimation of functional mode. We assume that the functional random variable X is supported on a compact interval $\mathcal{I}$. For most of our practical purposes, this assumption is quite reasonable. Suppose, $\theta_1 \geq \theta_2 \geq \ldots$ are the eigenvalues and ϕ_j are the respective orthonormal eigenvectors of the linear operator with kernel K. $\{\phi_1, \phi_2, \ldots\}$ is a basis for the space of all functions. Denoting the j^{th} functional principal component score by X_j, we can express X as

$$X = \sum_{j=1}^{\infty} \theta_j^{\frac{1}{2}} X_j \phi_j, \quad \text{where} \quad X_j = \theta_j^{-\frac{1}{2}} \int_{\mathcal{I}} X\phi_j. \tag{2}$$

Infact for any square integrable function x on $\mathcal{I}$, we can similarly write,

$$x = \sum_{j=1}^{\infty} \theta_j^{\frac{1}{2}} x_j \phi_j, \quad \text{where} \quad x_j = \theta_j^{-\frac{1}{2}} \int_{\mathcal{I}} x \phi_j. \tag{3}$$

We should note that owing to the orthogonality of the ϕ_j's, X_j's are uncorrelated. Then we can define the modal function as

$$x_{\text{mode}} = \sum_{j=1}^{\infty} \theta_j^{\frac{1}{2}} m_j \phi_j, \tag{4}$$

where, $\forall j$, m_j denotes the mode of the density f_j of x_j. We can estimate the modal function as

$$\hat{x}_{\text{mode}} = \sum_{j=1}^{T} \hat{\theta}_j^{\frac{1}{2}} \hat{m}_j \hat{\phi}_j, \tag{5}$$

where $\hat{\phi}_j$ and $\hat{\theta}_j$ are the estimators of ϕ_j and θ_j respectively, $\hat{m}_j$ is the mode of $\hat{f}_j$, the estimator of f_j, and T is a point of truncation.

Having the final target of clustering in mind and due to simplicity of estimation, we shall be using (5) as a definition of mode for the remainder of the project. In the following subsection we shall derive the theoretical distribution of the functional principal component scores. We shall later use this theoretical distribution to calculate the modal function explicitly.

3.1 Theoretical Distribution of Functional Principal Component Scores

Following the works of Lin et al. (2018) we shall first construct a random function in the $\mathcal{L}^2$ space, which is the space of all square integrable functions. Then we shall use the definition of functional principal score (see Delaigle and Hall, 2010) to obtain a theoretical probability distribution function for functional principal component scores. It should be noted that though the following results are obtained for the $\mathcal{L}^2$ space, the results can be easily extended to any Hilbert space $\mathcal{H}$.

Let $\Phi = \{\phi_1, \phi_2, \dots\}$ be a completely orthonormal basis of $\mathcal{L}^2$.

Following the work of Lin et al. (2018) we start with a sequence $\tilde{\xi}_1, \tilde{\xi}_2, \tilde{\xi}_3, \dots$ of uncorrelated and centred random variables such that the joint probability densities $\tilde{f}_k$ of $\tilde{\xi}_1, \tilde{\xi}_2, \dots, \tilde{\xi}_k$ exist for all k.

Also suppose K is a positive random integer with distribution $\pi = (\pi_1, \pi_2, \pi_3, \dots)$, where $\pi_k = \mathbb{P}(K = k)$ and K is independent of $\tilde{\xi}_1, \tilde{\xi}_2, \tilde{\xi}_3, \dots$. So, we need to have $\sum_{k=1}^{\infty} \pi_k = 1$. Furthermore, we define, $\pi_j^* = \sum_{k=j}^{\infty} \pi_k$.

Within this framework we can construct a random element X as

$$X = \mu + \sum_{k=1}^{K} \tilde{\xi}_k \phi_k, \tag{6}$$

where $\mu \in \mathcal{H}$. However, without any loss of generality we can assume $\mu = 0$, which enables us to rewrite (6) as

$$X = \sum_{k=1}^{K} \tilde{\xi}_k \phi_k. \tag{7}$$

This leads us to the following theorem.

Theorem 3.1. *Under the above setup, the distribution function of the j^{th} functional principal component score is given by*

$$\mathbb{P}[X_j \leq u] = \begin{cases} \pi_j^* \tilde{G}_j(u_0) & \text{if } u < 0 \\ (1 - \pi_j^*) + \pi_j^* \tilde{G}_j(u_0) & \text{if } u \geq 0, \end{cases}$$

where, $\tilde{G}_j$ denotes the distribution function of $\tilde{\xi}_j$ and $u_0 = \theta_j^{\frac{1}{2}} u$.

Proof. The functional principal component scores are given by Delaigle and Hall (2010)

$$X_j = \theta_j^{-\frac{1}{2}} \int X \phi_j, \tag{8}$$

where, $\theta_j = \pi_j^* var(\tilde{\xi}_j)$. Then we can rewrite (8) using (7) as,

$$X_j = \theta_j^{-\frac{1}{2}} \int \left(\sum_{k=1}^{K} \tilde{\xi}_k \phi_k \right) \phi_j. \tag{9}$$

Thus, exploiting the orthogonality of the basis functions, we have,

$$X_j = \begin{cases} \theta_j^{-\frac{1}{2}} \tilde{\xi}_j, & \text{if } \ j \leq K \\ 0, & \text{if } \ j > K. \end{cases}$$

This enables us to write

$$X_j = \theta_j^{-\frac{1}{2}} \tilde{\xi}_j 1_{\{j \leq K\}}, \tag{10}$$

where $1_{\{j \leq K\}}$ denotes the indicator function of the set $\{j \leq K\}$.

To obtain the distribution function we consider,

$$\begin{aligned} \mathbb{P}[X_j \leq u] &= \mathbb{P}[\theta_j^{-\frac{1}{2}} \tilde{\xi}_j 1_{\{j \leq K\}} \leq u] \\ &= \mathbb{P}[\tilde{\xi}_j 1_{\{j \leq K\}} \leq u_0] \text{ where, } \ u_0 = \theta_j^{\frac{1}{2}} u \\ &= \mathbb{P}[\tilde{\xi}_j 1_{\{j \leq K\}} \leq u_0 | j \leq K]\mathbb{P}[j \leq K] + \mathbb{P}[\tilde{\xi}_j 1_{\{j \leq K\}} \leq u_0 | j > K]\mathbb{P}[j > K] (\text{theorem of total probability}) \\ &= \pi_j^* \mathbb{P}[\tilde{\xi}_j 1_{\{j \leq K\}} \leq u_0 | j \leq K] + (1 - \pi_j^*)\mathbb{P}[\tilde{\xi}_j 1_{\{j \leq K\}} \leq u_0 | j > K]. \end{aligned} \tag{11}$$

We shall now consider two cases separately.

CASE 01 : $\mathbf{u_0 \geq 0}$

Then, $\tilde{\xi}_j 1_{\{j \leq K\}} \leq u_0$ if and only if $\tilde{\xi}_j \leq u_0$, provided $j \leq K$.
Hence, $\mathbb{P}[\tilde{\xi}_j 1_{\{j \leq K\}} \leq u_0 | j \leq K] = \mathbb{P}[\tilde{\xi}_j \leq u_0] = \tilde{G}_j(u_0)$, (say) where, $\tilde{G}_j$ denotes the marginal distribution function of $\tilde{\xi}_j$.

Also, if $j > K$, then $\tilde{\xi}_j 1_{\{j \leq K\}} = 0$. Which implies, $\mathbb{P}[\tilde{\xi}_j 1_{\{j \leq K\}} \leq u_0 | j > K] = 1$.

Thus, in this case we have,

$$\mathbb{P}[X_j \leq u] = \pi_j^* \tilde{G}_j(u_0) + (1 - \pi_j^*). \tag{12}$$

CASE 02 : $\mathbf{u_0 < 0}$

Then, $\mathbb{P}[\tilde{\xi}_j 1_{\{j \leq K\}} \leq u_0 | j \leq K] = \mathbb{P}[\tilde{\xi}_j \leq u_0] = \tilde{G}_j(u_0)$.

And, $\mathbb{P}[\tilde{\xi}_j 1_{\{j \leq K\}} \leq u_0 | j > K] = 0$.

Thus in this case we have,

$$\mathbb{P}[X_j \leq u] = \pi_j^* \tilde{G}_j(u_0). \tag{13}$$

Hence, from (12) and (13) we obtain the distribution function as

$$\mathbb{P}[X_j \leq u] = \begin{cases} \pi_j^* \tilde{G}_j(u_0) & \text{if,} \quad u < 0 \\ (1 - \pi_j^*) + \pi_j^* \tilde{G}_j(u_0) & \text{if,} \quad u \geq 0 \end{cases} \tag{14}$$

This completes the proof. □

We note the following comments from the above distribution function.

Remark 3.2. *The distribution function obtained in Theorem 3.1 so obtained is increasing, right continuous and converges to* 0 *and* 1 *respectively as* u *approaches* $-\infty$ *and* ∞*. So, the distribution function obtained is indeed a valid a distribution function.*

Remark 3.3. *The distribution function in Theorem 3.1 is not continuous. It has a jump at* 0*, thus indicating that the random variable* X_j *is a mixed random variable (i.e., neither discrete nor continuous).*

Remark 3.4. *Since the random variable is one of mixed type, this does not have a probability mass function or a probability density function.*

The jump at 0 can possibly be attributed to the presence of two sources of randomness, namely K and the $\tilde{\xi}_k$'s. We can try to solve this problem by removing the randomness from K. Instead we derive the distribution of the random functions as a limit of the distribution functions of a sequence of random functions. The setup remains almost similar, with slight modifications. For the sake of convenience of the reader we shall restate it.

3.2 Distribution function of Functional Principal Component Scores as a limiting distribution

Here we try to derive a distribution function for functional principal component scores. We focus on random functions defined over a compact interval $[a, b]$, ie, consider functions in the $\mathcal{L}^2[a, b]$ space. For most of our practical purposes the restriction of functions on some compact interval $[a, b]$ is justified.

Suppose, $\{\psi_1, \psi_2, \dots\}$ constitute an orthonormal bounded basis for the $\mathcal{L}^2[a, b]$ space of square-integrable functions on the compact interval $[a, b]$, ie, $\forall n \in \mathbb{N}$, ψ_n is bounded. An example of such a basis would be to consider the polynomial basis obtained by applying Gram-Scmidt orthoganilastion on the basis $\{1, t, t^2, \dots\}$. The resulting basis will still be consisted of polynomial functions, and hence bounded on a closed interval. Another example of such a basis would be to consider the Fourier basis functions, consisting of the functions $\{1, sin(\Upsilon t), cos(\Upsilon t), sin(2\Upsilon t), cos(2\Upsilon t), sin(3\Upsilon t), cos(3\Upsilon t), \dots\}$, where the parameter Υ determines the period $\frac{2\pi}{\Upsilon}$. The Fourier basis is orthogonal and can be scaled appropriately to obtain an orthonormal basis (see **?**, pg-45). Also, let $\xi_1, \xi_2, \dots$ be real valued random variables defined on some appropriate probability space, such that $\mathbb{E}(\xi_i) = 0$ for all $i \in \mathbb{N}$. Let us define,

$$X(t, \omega) = \mu(t) + \sum_{k=1}^{\infty} \xi_k(\omega)\psi_k(t) \tag{15}$$

and

$$X^{(K)}(t, \omega) = \mu(t) + \sum_{k=1}^{K} \xi_k(\omega)\psi_k(t) \tag{16}$$

Here, $X(t, \omega)$ and $X^{(K)}(t, \omega)$ are random functions defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The function $\mu(t)$ denotes the mean function. We keep in mind the definition of the j^{th} functional principal component scores as

$$X_j(\omega) = \theta_j^{-\frac{1}{2}} \int X(t,\omega)\psi_j(t)dt \tag{17}$$

and,

$$X_j^{(K)}(\omega) = \theta_j^{-\frac{1}{2}} \int X^{(K)}(t,\omega)\psi_j(t)dt \tag{18}$$

Throughout the remainder of the report we shall omit the parameters ω and t as far as practicable and refer to $X(t,\omega)$ and $X^{(K)}(t,\omega)$, $X_j(\omega)$, $X_j^{(K)}(\omega)$ and $\mu(t)$ defined in (15),(16),(17) and (18) respectively as X and $X^{(K)}$, X_j, $X_j^{(K)}$ and μ. We shall also refer to the $\mathcal{L}_2[a,b]$ space as $\mathcal{L}_2$ for the remainder of the report.

Now, we plan to obtain the distribution of X_j as a limit of the distribution of $X_j^{(K)}$. We follow the steps below to reach the result

- **STEP-01 :** We shall first show that
$$X^{(K)} \xrightarrow{\text{a.s.}} X.$$
Here '$\xrightarrow{a.s.}$' denotes almost sure convergence.

- **STEP-02 :** We shall show that the mapping $X \to X\psi_j$ is continuous $\forall j \in \mathbb{N}$.

- **STEP-03 :** We shall show that the mapping $\int : \mathcal{L}^2[a,b] \to \mathbb{R}$ is continuous.

- **STEP-04 :** We shall use the 'Continuous Mapping Theorem' to show that $X_j^{(K)} \xrightarrow{a.s.} X_j$.

- **STEP-05 :** Noting that 'almost sure convergence' implies 'convergence in distribution', we shall conclude that the distribution function of $X_j^{(K)}$ converges to the distribution function of X_j for all continuity points of the later.

We shall now state and prove a series of results, which are required to derive the final distribution functions of the principal component scores.

Lemma 3.5. $X^{(K)} \xrightarrow{a.s.} X$

Proof. First we observe that since X and μ belong to the $\mathcal{L}^2$ space, we have $(X-\mu) \in \mathcal{L}^2$. So, we have $||X-\mu|| < \infty$.

Now,

$$\begin{aligned} ||X-\mu||^2 &= ||\sum_{k=1}^{\infty} \xi_k\psi_k||^2 \\ &= \langle \sum_{k=1}^{\infty} \xi_k\psi_k, \sum_{k=1}^{\infty} \xi_k\psi_k \rangle \\ &= \sum_{k=1}^{\infty} \xi_k^2 ||\psi_k||^2 + \sum_{\substack{k=1 \\ k\neq l}}^{\infty} \sum_{l=1}^{\infty} \xi_k\xi_l \langle \psi_k, \psi_l \rangle \end{aligned} \tag{19}$$

Now, since $\{\psi_1, \psi_2, \dots\}$ constitute an orthonormal basis for the $\mathcal{L}^2$ space, we have $||\psi_k||^2 = 1$, for all k and $\langle \psi_k, \psi_l \rangle = 0$ for all $k \neq l$. Thus, we can rewrite (19) as

$$||X-\mu|| = \sum_{k=1}^{\infty} \xi_k^2 \tag{20}$$

Since, $||X - \mu|| < \infty$, we have, $\sum_{k=1}^{\infty} \xi_k^2 < \infty$

A similar calculation shows that $||X - X^{(K)}||^2 = \sum_{k=K+1}^{\infty} \xi_k^2$.

Therefore, $\forall \omega \in \Omega$ the series $\sum_{k=1}^{\infty} \xi_k^2$ converges, and hence we have

$$\lim_{K\to\infty} \sum_{k=K+1}^{\infty} \xi_k^2 = 0 \tag{21}$$

Equation (21) shows that for all $\omega \in \Omega$, $\lim_{K\to\infty} ||X - X^{(K)}||^2 = 0$, and hence $\lim_{K\to\infty} ||X - X^{(K)}|| = 0$. In other words,

$$\mathbb{P}(\lim_{K\to\infty} X^{(K)} = X) = 1 \tag{22}$$

Equation (22) shows that

$$X^{(K)} \xrightarrow{a.s.} X \tag{23}$$

This completes the proof. ☐

Lemma 3.6. *The mapping $X \to X\psi_j$ is continuous $\forall j \in \mathbb{N}$.*

Proof. Let us fix $j \in \mathbb{N}$. We need to show that $\forall \epsilon > 0, \exists \delta > 0$ such that $||X\psi_j - Y\psi_j|| < \epsilon$ whenever, $||X - Y|| < \delta$. We fix $\epsilon > 0$ and remember that as the basis function ψ_j is bounded, $\exists \gamma_j \in \mathbb{R}_+$, such that $\forall t \in [a, b]$, we have $|\phi_j(t)| \leq \gamma_j$. This gives,

$$\begin{aligned} ||X\psi_j - Y\psi_j|| &= ||(X - Y)\psi_j|| \\ &= [\int_a^b (X(t) - Y(t))^2 \psi_j^2(t) dt]^{\frac{1}{2}} \\ &\leq [\gamma_j^2 \int_a^b (X(t) - Y(t))^2 dt]^{\frac{1}{2}} \\ &= \gamma_j ||X - Y|| \end{aligned}$$

Thus, whenever we have, $||X - Y|| < \gamma_j^{-1}\epsilon$, we have, $||X\psi_j - Y\psi_j|| < \epsilon$. Thus, we can choose δ to be $\gamma_j^{-1}\epsilon$ to complete the proof. ☐

Lemma 3.7. *The mapping $\int : \mathcal{L}^2[a, b] \to \mathbb{R}$ is continuous.*

Proof. We note that, by the Cauchy-Schwartz Inequality,

$$(\int_a^b [f(x) - g(x)]dx)^2 \leq \int_a^b [f(x) - g(x)]^2 dx \int_a^b dx = |b - a| \times ||f - g||^2 \tag{24}$$

Therefore, whenever, $||f - g|| < |b - a|^{-\frac{1}{2}}\epsilon$, we have, $\int_a^b [f(x) - g(x)]dx < \epsilon$. This shows that the integral mapping is continuous. ☐

Lemma 3.8. $X_j^{(K)} \xrightarrow{a.s.} X_j$ *as* $K \to \infty$.

Proof. The proof follows from the continuity of the above mappings and applying the Continuous Mapping Theorem (see Wellner et al., 2013, pg-20). ☐

Corollary 3.8.1. $X_j^{(K)} \xrightarrow{d} X_j$ *as* $K \to \infty$.

Proof. The proof follows from observing that 'almost sure convergence' implies 'convergence in distribution' (see Billingsley, 2013, pg-207). ☐

Since we have established that $X_j^{(K)} \xrightarrow{d} X_j$, we can say that the distribution function $F_j^{(K)}$ of $X_j^{(K)}$ converges point-wise to the distribution function F_j of X_j for all continuity points of F_j. So, we shall obtain the distribution

function of the j^{th} principal component score as limit of the distribution functions of $X_j^{(K)}$.
From (16) and (18) we see that

$$X^{(K)}\psi_j(t) = \mu(t)\psi_j(t) + \sum_{k=1}^{K} \xi_k(\omega)\psi_k(t)\psi_j(t)$$

which gives,

$$\begin{aligned}\int X^{(K)}\psi_j(t)dt &= \int [\mu(t)\psi_j(t)]dt + \int \left[\sum_{k=1}^{K} \xi_k\psi_k(t)\psi_j(t)\right] dt \\ &= \int [\mu(t)\psi_j(t)]dt + \sum_{k=1}^{K} \xi_k \int [\psi_k(t)\psi_j(t)]dt \qquad (25)\end{aligned}$$

Owing to the orthonormality of the ψ_k''s, we have,

$$\int X^{(K)}\psi_j(t)dt = \begin{cases} c_j + \xi_j & \text{if } j \le K; \\ c_j & \text{if } j > K \end{cases} \qquad (26)$$

where, the constant $c_j = \int [\mu(t)\psi_j(t)]dt$. This gives us (using (18) and (26)),

$$X_j^{(K)} = \begin{cases} \theta_j^{-\frac{1}{2}}(c_j + \xi_j) & \text{if } j \le K; \\ \theta_j^{-\frac{1}{2}} c_j & \text{if } j > K \end{cases} \qquad (27)$$

Now we need to obtain the distribution function $F_j^{(K)}$ of the j^{th} Principal Component Score $X_j^{(K)}$ of $X^{(K)}$. Towards that direction, we consider the following two cases.

CASE-01 : $j \le K$.

$$\begin{aligned}F_j^{(K)}(x) &= \mathbb{P}[X_j^{(K)} \le x] \\ &= \mathbb{P}[\theta_j^{-\frac{1}{2}}(\xi_j + c_j) \le x] \\ &= \mathbb{P}[\xi_j + c_j \le \theta_j^{\frac{1}{2}} x] \\ &= \mathbb{P}[\xi_j \le \theta_j^{\frac{1}{2}} x - c_j] \\ &= G_j(\theta_j^{\frac{1}{2}} x - c_j) \qquad (28)\end{aligned}$$

where, G_j denotes the distribution function of ξ_j.

CASE-02 : $j > K$.

$$\begin{aligned}F_j^{(K)}(x) &= \mathbb{P}[X_j^{(K)} \le x] \\ &= \begin{cases} 0 & \text{if } \theta_j^{-\frac{1}{2}} x \le c_j \\ 1 & \text{if } \theta_j^{-\frac{1}{2}} x > c_j \end{cases} \\ &= 1_{\{x > \theta_j^{-\frac{1}{2}} c_j\}} \qquad (29)\end{aligned}$$

Thus, (28) and (29) give us the distribution function as

$$F_j^{(K)}(x) = \begin{cases} G_j(\theta_j^{\frac{1}{2}} x - c_j) & \text{if } j \le K \\ 1_{\{x > \theta_j^{-\frac{1}{2}} c_j\}} & \text{if } j > K \end{cases} \qquad (30)$$

This leads us to the following lemma.

Lemma 3.9. *For every point $x \in \mathbb{R}$, we have $\lim_{K\to\infty} F_j^{(K)}(x) = G_j(\theta_j^{\frac{1}{2}}x - c_j)$.*

Proof. Let us fix any $x_0 \in \mathbb{R}$ and $\epsilon > 0$. We need to show that $\exists N \in \mathbb{N}$ such that whenever $K > N$, we have $|F_j^{(K)}(x_0) - G_j(\theta_j^{\frac{1}{2}}x_0 - c_j)| < \epsilon$.
Let us take $N = j$. Then for any $K > N = j$, we have $F_j^{(K)}(x_0) = G_j(\theta_j^{\frac{1}{2}}x_0 - c_j)$, implying, $|F_j^{(K)}(x_0) - G_j(\theta_j^{\frac{1}{2}}x_0 - c_j)| = 0 < \epsilon$. Hence, we have

$$\lim_{K\to\infty} F_j^{(K)}(x) = G_j(\theta_j^{\frac{1}{2}}x - c_j) \tag{31}$$

This completes the proof. □

Corollary 3.9.1. *If x is a point of continuity of the distribution function F_j of the j^{th} principal component score X_j of the random function X, then we have $F_j(x) = G_j(\theta_j^{\frac{1}{2}}x - c_j)$.*

Proof. The proof follows immediately from the corollary 3.8.1 and lemma 3.9. □

Remark 3.10. *Since for $K \geq j$ we have $F_j^K(x) = G_j(\theta_j^{\frac{1}{2}}x - c_j)$, we have $F_j(x) = G_j(\theta_j^{\frac{1}{2}}x - c_j)$. Thus if the random variables ξ_j are continuous, the distribution function F_j is also continuous.*

Thus we have derived the distribution functions of the principal component scores of the random function X. Next, we attempt to derive an expression for the mode of X_j and hence an expression for the modal function X_{mode}.

3.3 Derivation of the Modal Function

We assume that the random variables ξ_k have a continuous distributions. We have the distribution function F_j of the j^{th} principal component score given by

$$F_j(x) = G_j(\theta_j^{\frac{1}{2}}x - c_j) \tag{32}$$

This gives us the density function f_j of the j^{th} principal component score as

$$\begin{aligned} f_j(x) &= \frac{d}{dx}F_j(x) \\ &= \frac{d}{dx}G_j(\theta_j^{\frac{1}{2}}x - c_j) \\ &= \theta_j^{\frac{1}{2}}g_j(\theta_j^{\frac{1}{2}}x - c_j) \end{aligned} \tag{33}$$

To obtain the mode of X_j, we equate the derivative of f_j to zero. This gives,

$$g_j'(\theta_j^{\frac{1}{2}}x - c_j) = 0$$

This shows that $(\theta_j^{\frac{1}{2}}x - c_j)$ is a mode of g_j.

Thus if m_{ξ_j} is a mode of g_j and, then we can obtain a mode m_j of f_j as,

$$\theta_j^{\frac{1}{2}}m_j - c_j = m_{\xi_j}$$

which gives us the mode m_j as,

$$m_j = \theta_j^{-\frac{1}{2}}(m_{\xi_j} + c_j) \tag{34}$$

Thus following the work of Delaigle and Hall (2010), we obtain the modal function as,

$$
\begin{aligned}
X_{\text{mode}}(t) &= \sum_{j=1}^{\infty} \theta_j^{\frac{1}{2}} m_j \psi_j(t) \\
&= \sum_{j=1}^{\infty} \theta_j^{\frac{1}{2}} \theta_j^{-\frac{1}{2}} (m_{\xi_j} + c_j)\psi_j(t) \\
&= \sum_{j=1}^{\infty} (m_{\xi_j} + c_j)\psi_j(t) \\
&= \mu(t) + \sum_{j=1}^{\infty} m_{\xi_j} \psi_j(t) \qquad (35)
\end{aligned}
$$

The equality in the last line follows by observing that $c_j = \int [\mu(t)\psi_j(t)]dt$ is the coefficient of ψ_j in the basis expansion of μ. This completes the derivation.

In the following section, we shall do a numerical simulation to generate random functions and estimate the modal functions. We shall also perform modal clustering using the MEM algorithm (see Li et al., 2007) for clustering the so generated random functions.

4 Numerical Studies

The following simulation was performed using the R programming language. The main goals of this simulation are generation of random functions, estimation of the modal function and demonstration of the almost sure convergence results theoretically obtained in lemma 3.5. The necessary R codes have been included in the Appendix.
50 random functions were generated using the Fourier basis function system. Each of the functions were generated using 101 coefficients. To introduce clustering we have generated the coefficients from five different types of distribution functions with varying parameters and 10 functions were created using each of them. The distributions explored were, $\mathcal{N}(\frac{5}{k^2}, \frac{0.1}{k})$, $\mathcal{N}(\frac{tan(k)}{k^5}, \frac{0.1}{k})$, $\mathcal{N}(\frac{sin(k)}{6k}, \frac{0.1}{k^2})$, $\mathcal{N}(\frac{(-1)^k.k}{2^k}, \frac{1}{k})$ and $\mathcal{N}(\frac{(-1)^k}{k!}, \frac{0.1}{k})$ for $k \in \{1, 2, 3, \ldots, 101\}$. Figure 1 shows the random functions so obtained and the principal component scores. For further analysis and clustering using the MEM algorithm it is essential to determine the number of principal components to be retained. For this the percentage of variability explained by the principal components was used. The minimum number of principal components required to explain 99.9% of the variability were retained. For the random functions generated by this simulation, 9 principal components explained 99.9% of the variability, and hence the clustering was performed using only 9 principal components. Figure 2 shows the principal component functions and the cumulative percentage of variability explained the principal components. Having determined the number of principal components to be used for the clustering, we can perform modal clustering of the functional data obtained. Figure 3 depicts the HMAC (Hierarchical Mode Association Clustering) Dendrogram for the clustering and the clustered functions. The different colours denote the different clusters. Some colours also denote more than one cluster. That is because of the limited availability of colours in R compared to the number of clusters identified, 13 in this case. However, such instances can be visually identified by observing the locations of the clusters. Table 1 shows the expected clustering compared against the clustering obtained by the MEM algorithm.

Expected Clusters	1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,2,3,3,3,3,3,3,3,3,3,3,4,4,4,4,4,4,4,4,4,4,5,5,5,5,5,5,5,5,5,5
Obtained Clusters	1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,3,4,5,6,7,8,9,10,11,12,13,13,13,13,13,13,13,13,13,13

Table 1: Expected versus Obtained Clusters

From table 1 and figure 3 we can see that the first cluster of functions have been properly identified, and the second and the third clusters have been merged owing to their similarity in behaviour. An excessive amount of variance in the fourth cluster in comparison to the mean function has resulted in identification of each of the individual functions as singleton clusters. The fifth cluster has again been identified appropriately.
Our next target in this simulation is to demonstrate the almost sure convergence result obtained in 3.5. Towards

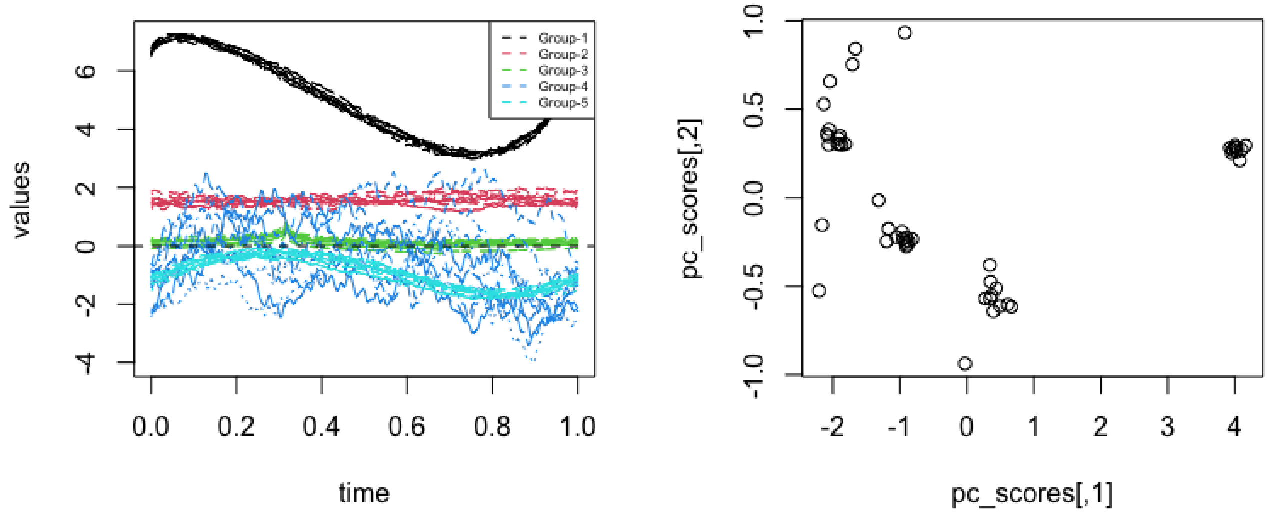

Figure 1: The Random Functions and Principal Component Scores

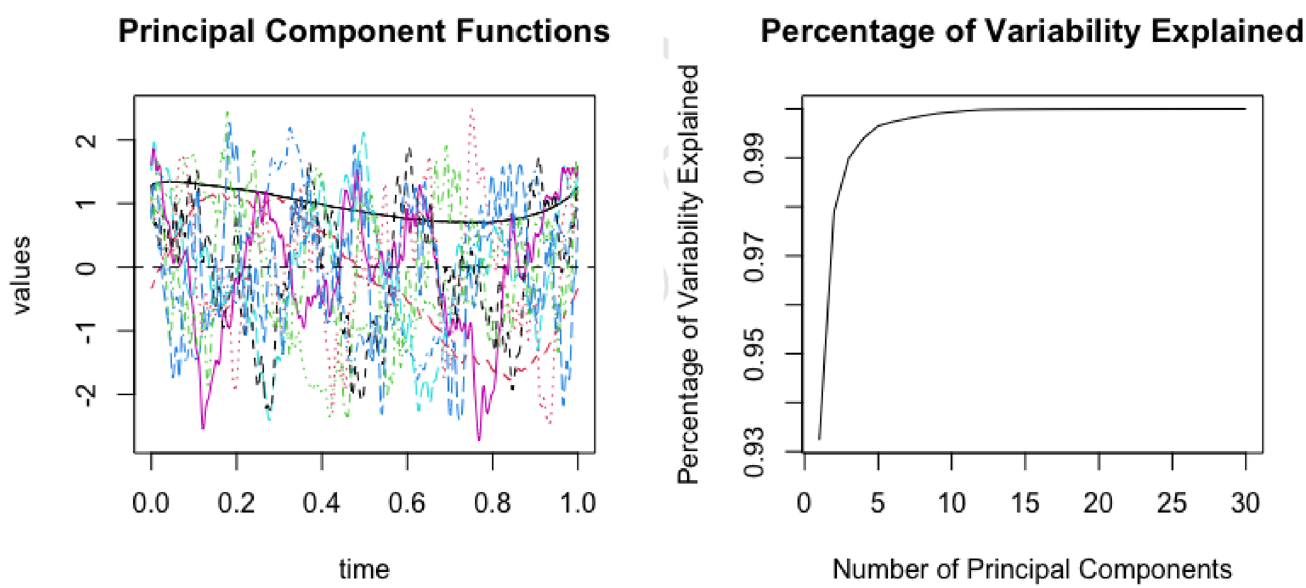

Figure 2: Principal Component Functions and Percentage of Variability Explained

that direction, a random function has been generated using random coefficients having a $\mathcal{N}(\frac{tan(k)}{k^5}, \frac{0.1}{k})$, and plotted in red. Also, functions have been generated by the partial sums of products of coefficients and the basis functions have been plotted in black against the red curve. Figure 4 shows that though the black curves initially behave quite differently from the red curve,they gradually converge to the red curve, thus demonstrating the almost sure convergence result obtained in Lemma 3.5. We shall next be estimating and plotting the modal function. A comparative study of the modal function proposed in equation 4 and its estimated value in equation 35 has been done by plotting the two estimated modal functions against each other for the third group of functions. The two estimations, as can be seen from figure 5 almost align with each other. The differences between the two estimations

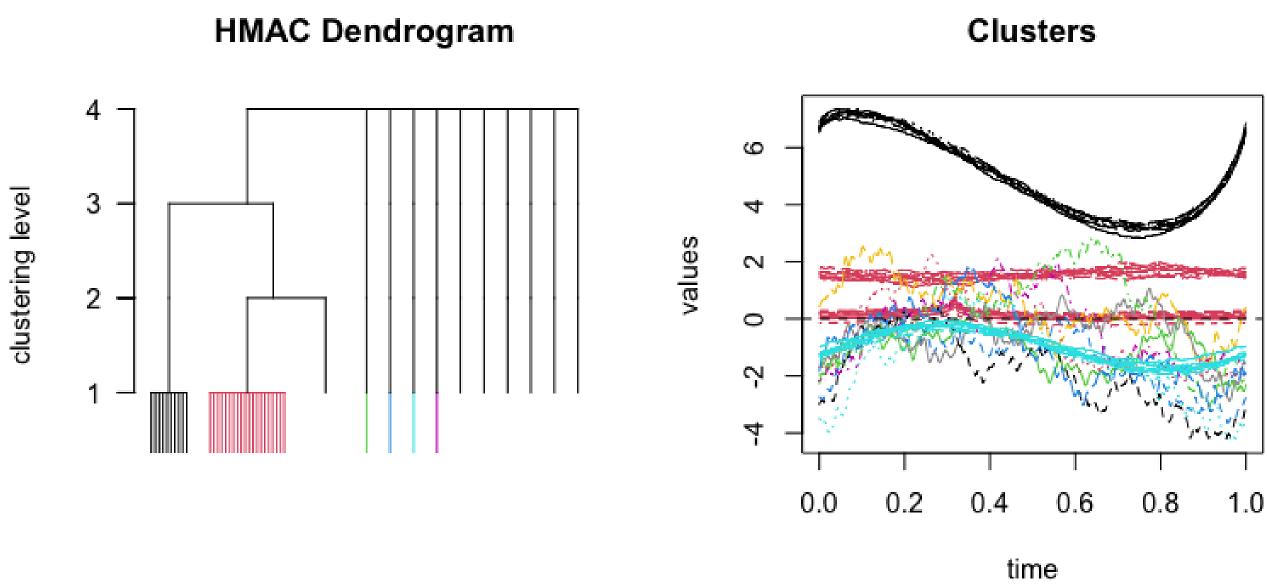

Figure 3: Clustering using MEM Algorithm

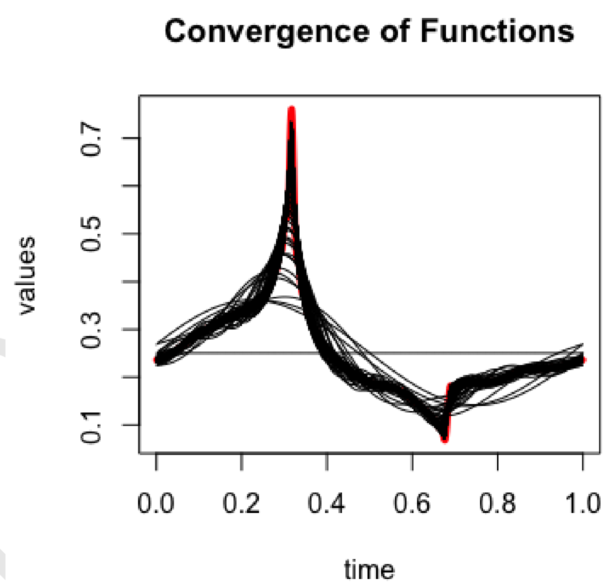

Figure 4: Demonstrating Almost Sure Convergence of Functions

can be attributed to the number of components being used for the estimation. This shows that 35 is a good estimation of the modal function.

5 Conclusion

In this paper, we developed a rigorous framework for deriving the probability distribution of functional principal component (FPC) scores for random functions expressed in the $\mathcal{L}_2$ space using orthonormal basis expansions. Our work focused on understanding the structure of FPC scores under two distinct representations of the number of

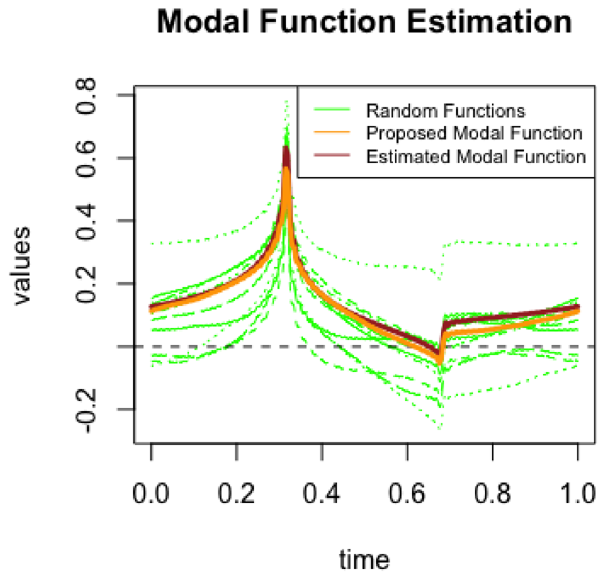

Figure 5: Estimation of the Modal Function

principal components: one where the number of components is random, and another where it is fixed and finite. These two approaches allowed us to explore both the theoretical challenges and the practical implications associated with FPC score distributions in functional data analysis (FDA).

In the first approach, we considered a random number of principal components, which introduced significant stochastic variability into the representation of FPC scores. We demonstrated that in such cases, the distribution of FPC scores is of mixed type, meaning it has both discrete and continuous elements. This mixed nature arises because the number of components itself is not fixed, leading to a distribution that cannot be described solely by a probability density function (PDF). The mixed-type distribution is an important finding because it reveals a fundamental limitation in standard probabilistic approaches when dealing with situations where the dimensionality of the principal component expansion is uncertain or random. This contributes to our understanding of the complexities that arise when attempting to analyze functional data in a non-deterministic framework, where the number of components cannot be easily controlled or predicted.

In contrast, our second approach—where the number of principal components is fixed and finite—enabled a clearer analysis. We showed that as the number of principal components increases and approaches infinity, the distribution of FPC scores converges to a continuous distribution, allowing for the definition of a proper probability density function. This result is especially relevant for practical applications in FDA where a fixed and finite number of components is often used to approximate the underlying functions. By gradually increasing the number of components, we can achieve a more refined approximation of the true distribution, thus enabling a more robust statistical analysis of the FPC scores. This finding suggests that in many real-world applications, where the number of principal components is chosen based on model selection criteria (such as explained variance), the resulting continuous distribution provides a solid foundation for making probabilistic inferences about the underlying functional data.

One of the major contributions of our work is the introduction of a novel method for estimating the functional mode (or modal function) based on the coefficients of the orthonormal basis expansion, without the need to explicitly compute the FPC scores. Traditional approaches to mode estimation in functional data require direct computation of these scores, which can be computationally expensive and prone to errors when the number of components is large or when the data is noisy. Our method, on the other hand, circumvents this step by leveraging the coefficients directly, thus simplifying the computational process while maintaining accuracy. By focusing on the coefficients of the orthonormal basis, our approach provides a more efficient and stable way to estimate the mode, which is particularly useful in high-dimensional functional data applications.

We further extended our theoretical analysis by connecting our new definition of the functional mode to existing approaches, such as those proposed by Peter Hall and others. We demonstrated that our definition is consistent with the classical understanding of the modal function while offering a more flexible and computationally efficient approach. Through this connection, we showed that our mode estimation technique retains the desirable properties of established methods, such as robustness to small perturbations in the data and consistency as the number of components increases. This consistency reinforces the validity of our approach and highlights its potential for broader applications in FDA.

To validate our theoretical findings, we conducted extensive numerical studies. These studies confirmed that our method for estimating the modal function not only aligns with traditional definitions but also performs well in practice, particularly in cases where the data is noisy or where the number of principal components is large. The numerical results demonstrated that our proposed framework yields accurate and reliable estimates of the mode, even when the functional data exhibits complex patterns or non-standard behavior. This provides strong empirical support for the practical utility of our methodology, suggesting that it can be effectively applied to a wide range of functional data analysis problems, from clustering and classification to anomaly detection and prediction.

In addition to the contributions related to FPC score distributions and mode estimation, our work has broader implications for the field of functional data analysis. By providing a clear understanding of the probabilistic structure of FPC scores, we offer new tools for characterizing functional data in a probabilistic framework. This is particularly important for tasks that involve uncertainty quantification, such as Bayesian FDA, where the ability to define probability distributions over functional data is crucial. Our results lay the groundwork for future research in these areas, providing a foundation for further exploration into the probabilistic properties of functional data.

Overall, the framework we developed in this paper offers several key contributions to the theory and practice of FDA. First, by clarifying the nature of FPC score distributions under different representations of principal components, we provide new insights into the challenges and opportunities associated with probabilistic analysis in FDA. Second, our novel mode estimation technique simplifies the computation of key statistical features, making it more accessible and practical for applied researchers. Finally, our theoretical and empirical results demonstrate the robustness and applicability of our methods across a wide range of functional data contexts.

Looking ahead, we envision several potential extensions of this work. One promising direction is to explore how our framework can be integrated with advanced FDA techniques, such as functional regression and functional classification, where the estimation of principal component scores and modes plays a central role. Another area for future research is to investigate how our approach can be adapted to handle functional data with more complex structures, such as sparsely observed functions or functions with missing data. In conclusion, we believe that the methodologies and insights developed in this paper provide a valuable foundation for advancing the field of FDA and open up new avenues for both theoretical exploration and practical application in a wide range of scientific domains.

References

P. Billingsley. *Convergence of probability measures.* John Wiley & Sons, 2013.

A. Delaigle and P. Hall. Defining probability density for a distribution of random functions. *The Annals of Statistics*, pages 1171–1193, 2010.

F. Ferraty and P. Vieu. *Nonparametric functional data analysis: theory and practice*, volume 76. Springer, 2006.

U. Grenander. Stochastic processes and statistical inference. *Arkiv för matematik*, 1(3):195–277, 1950.

J. Li, S. Ray, and B. G. Lindsay. A nonparametric statistical approach to clustering via mode identification. *Journal of Machine Learning Research*, 8(8), 2007.

Z. Lin, H.-G. Müller, and F. Yao. Mixture inner product spaces and their application to functional data analysis. *The Annals of Statistics*, 46(1):370–400, 2018.

J. O. Ramsay and C. Dalzell. Some tools for functional data analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(3):539–561, 1991.

J. O. Ramsay and B. W. Silverman. Functional data analysis. *İnternet Adresi: http*, 2008.

J. O. Ramsay, K. G. Munhall, V. L. Gracco, and D. J. Ostry. Functional data analyses of lip motion. *The Journal of the Acoustical Society of America*, 99(6):3718–3727, 1996.

J. O. Ramsay, H. Wickham, S. Graves, and G. Hooker. fda: Functional data analysis. *R package version*, 2(4):142, 2014.

J. Wellner et al. *Weak convergence and empirical processes: with applications to statistics.* Springer Science & Business Media, 2013.