mathematics

MDPI

Article

Estimation and Hypothesis Test for Mean Curve with Functional Data by Reproducing Kernel Hilbert Space Methods, with Applications in Biostatistics

Ming Xiong [1], Ao Yuan [2,*], Hong-Bin Fang [2], Colin O. Wu [3] and Ming T. Tan [2,*]

[1] School of Mathematics and Statistics, Central China Normal University, Wuhan 430079, China
[2] Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University, Washington, DC 20057, USA
[3] National Heart, Lung and Blood Institute, Office of Biostatistics Research, Bethesda, MD 20892, USA
* Correspondence: ay312@georgetown.edu (A.Y.); mtt34@georgetown.edu (M.T.T.)

Abstract: Functional data analysis has important applications in biomedical, health studies and other areas. In this paper, we develop a general framework for a mean curve estimation for functional data using a reproducing kernel Hilbert space (RKHS) and derive its asymptotic distribution theory. We also propose two statistics for testing the equality of mean curves from two populations and a mean curve belonging to some subspace, respectively. Simulation studies are conducted to evaluate the performance of the proposed method and are compared with the major existing methods, which shows that the proposed method has a better performance than the existing ones. The method is then illustrated with an analysis of the growth data from the National Growth and Health Study (NGHS) project sponsored by the NIH.

Keywords: functional data; hypothesis testing; kernel function; mean curve estimation; reproducing kernel Hilbert space

MSC: 62G08

check for updates

Citation: Xiong, M.; Yuan, A.; Fang, H.-B.; Wu, C.O.; Tan, M.T. Estimation and Hypothesis Test for Mean Curve with Functional Data by Reproducing Kernel Hilbert Space Methods, with Applications in Biostatistics. *Mathematics* **2022**, *10*, 4549. https://doi.org/10.3390/math10234549

Academic Editor: Manuel Franco

Received: 2 November 2022
Accepted: 23 November 2022
Published: 1 December 2022

1. Introduction

Functional data analysis with the objectives of estimating and testing mean curves over time has been extensively used in biomedical, health science and other areas of study. Functional data are random elements in the Banach/Hilbert space, and there are no density functions or parametric models for functional data. Thus, estimates and hypothesis tests for the mean curves are mostly based on nonparametric methods, without relying on potentially unrealistic parametric model assumptions. The commonly used methods for a functional data analysis and reviews of the existing work can be found in [1–5]. A popular postulate for nonparametric inferences with functional data is that the mean curves belong to some "structured space" [6], which can be approximated by expansions of a set of known basis functions, so that the estimation and testing procedures can be constructed through the unknown coefficients of the basis expansions. The existing results based on various basis approximation methods can be found in [7–10] proposed the functional principal components method via basis expansions, and [11] studied a likelihood ratio test for longitudinal and functional data. Ref. [12] gave a comprehensive review of the developments in this area. Ref. [13] studied the general properties of the mean curve estimation, under common and independent observation points, and obtained the optimal minimax convergence rates for both cases. Ref. [14] considered a multivariate functional principal method. Ref. [15] constructed a control chart for functional data. Ref. [16] proposed a cross-component registration method. Ref. [17] considered a random projection method. Ref. [18] studied a functional linear mixed model.

In practice, sometimes the observed functional data are rather "irregular" in that observation time points are unbalanced; they are dense in some time intervals, sparse in other time intervals. Such data often arise from medical studies, for example, patients can be observed on regular schedules during their treatments initially, and their subsequent visits to hospital become less frequent and gradually thin out as the patients' conditions improve or become incurable.

To illustrate the general structure of such data in this research, Figure 1 depicts the growth data from the National Growth and Health Study (NGHS), sponsored by the National Heart, Lung, and Blood Institute, from 1985 to 2000. In these data, the observed time points are relatively dense at the beginning of the treatment and then become sparse, gradually thinning out near the study end.

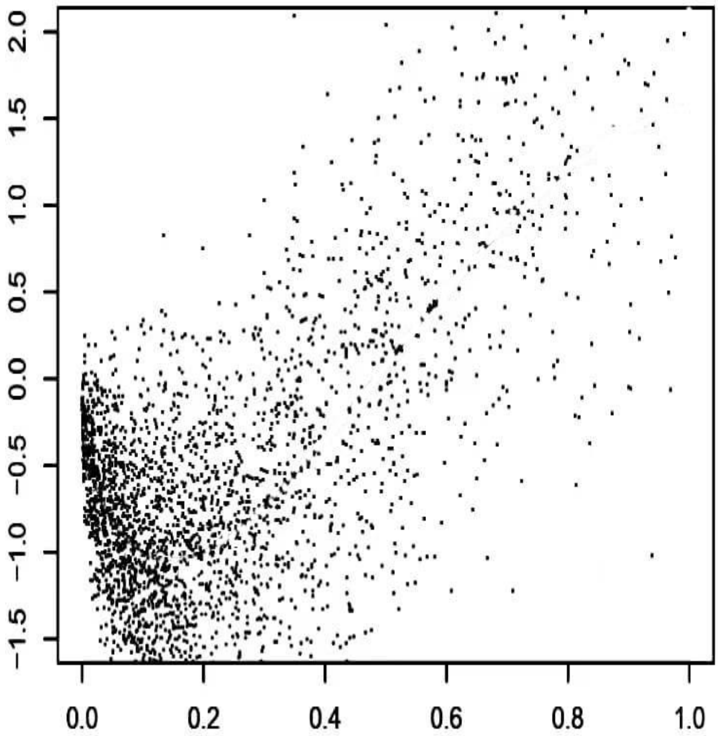

Figure 1. Raw data in the NGHS study.

Apparently, the curve estimation obtained by the common smoothing techniques aforementioned cannot be regarded as "true" observed data curves due to the unbalanced observations. In fact, our simulation studies in Section 4 show that those methods may result in a biased estimation or a relatively large variance of the estimator when the observation time points are unbalanced. Hence, some effective methods for estimating the mean curves of functional data should be developed.

In this paper, we take interpolated curves at the longitudinal times points for each individual as the estimated–observed curves; the goal is to estimate the underlying true mean curve and testing hypotheses about it. The simplest method for a mean curve estimation is just to take the empirical mean of the observed interpolated curves, or any other non-smoothing (i.e., without a roughness penalty) functional estimates. However, as the time points for the observations are generally sparse, the interpolated curves and their mean are not sufficiently smooth, often with a wiggly shape and large variances, a poor performance in the interval with sparse observations. To overcome these issues, there are several commonly used methods, such as kernel smoothing, spline and the method of the reproducing kernel Hilbert space (RKHS). The method of the RKHS has the following advantages. Instead of specifying a set of (orthogonal) basis functions and the number of bases, one only needs to choose the kernel(s) of the RKHS. Moreover, with this method, any bounded linear functional can be written as a representer (the Rize representer with respect to the inner product of the RKHS) in a closed form in terms of the kernel of the RKHS, as the estimator can often be formulated as a linear functional of the data in a closed form. Another consideration is the computation. It is known that for non-smoothing methods, the computation is often of order $O(n)$, where n is the sample size of the data, while for smoothing methods, the amount of computation may substantially exceed $O(n)$ and become computationally extensive. Thus, for smoothing methods, it is important to find a method with $O(n)$ computation load. To achieve this, for spline methods, the basis should have local-only support, i.e., nonzero only locally. The RKHS method is a special case of spline with this property and can achieve the $O(n)$ computation for many functional estimation problems, which is called the optimal basis theorem in ([1], p.363) and the representer theorem in [19]. More specifically, the RHKS $\mathbb{H}$ is a Hilbert space of functions, equipped with an inner product $\langle\cdot,\cdot\rangle$. On $\mathbb{H}$, there is a kernel $K(\cdot,\cdot)$, a bivariate function, such that $\langle K(t,\cdot), h(\cdot)\rangle = h(t)$, for all $h \in \mathbb{H}$, and so the name reproducing kernel. To apply this method to the functional estimation, like the other smoothing methods, a penalty term will be specified along with the object functional. The $\mathbb{H}$ can be divided into a null space $\mathbb{H}_0$, with a kernel K_0, corresponding to the penalty term, and its orthogonal complement $\mathbb{H}_1$ with a kernel K_1. $\mathbb{H}_0$ is a finite-dimensional space spanned by some basis $g_1,\ldots,d_d$, and an estimator $\hat{\mu}(\cdot)$ of the mean curve of data $Y_1(\cdot),\ldots,Y_n(\cdot)$ has the form, for some constants $a_1,\ldots,a_d$ and $b_1,\ldots,b_n$,

$$\hat{\mu}(\cdot) = \sum_{j=1}^{d} a_j g_j(\cdot) + \sum_{i=1}^{n} b_i \langle K_1, Y_i\rangle(\cdot).$$

Given the above reasons, we adopt the method of the RKHS for its ease of use and computational efficiency, and other well-known properties.

Recently, the RKHS method has been studied by many researchers. Ref. [20] used the method in the spline model, Ref. [21] studied the quantile regression using this method and [19] used it for functional linear regressions. However, to the best of our knowledge, no asymptotic distribution theory for the mean estimation with functional data using the RKHS exists in the literature, as we do here. The simulation studies indicate an apparent advantage of the proposed method compared to some commonly used methods for this type of data. The rest of this paper is organized as follows. In Section 2, we describe the general RKHS method for the mean curve estimation and derive the theoretical results and the asymptotic distributions of the mean curve estimations are investigated with a special construction of the RKHS. And two statistics for testing mean curves are proposed in Section 3. Section 4 provides the results of the simulation studies and the application of the proposed methods to real functional data from the National Growth and Health Study (NGHS) of the NIH. We conclude with a discussion in Section 5, and the proofs of the main results are given in the Appendix A.

2. Mean Curve Estimation via RKHS

We consider the stochastic processes $Y(t)$ indexed by the time point $t \in (0, T]$ for some $0 < T < \infty$. At any given $t \in (0, T]$, $Y(t) \in R$ is the real-valued outcome variable. Assume that there are n independent subjects and each subject is observed at randomly selected distinct time points. Let $Y_{ij} = Y_i(t_{ij})$ be the observation of subject i at time t_{ij} for $i = 1,\ldots,n$ and $j = 1,\ldots,m_i$. Denote the mean function of $Y(t)$ by $\mu_0(t)$, the model is assumed as

$$Y_i(t_{ij}) = \mu_0(t_{ij}) + \epsilon_i(t_{ij}). \text{ for } j = 1,\ldots,m_i, \text{ and } i = 1,\ldots,n, \tag{1}$$

where the $\epsilon_i(\cdot)$ is i.i.d. measurement error with mean zero and variance function $\sigma^2(t)$. Furthermore, we assume that $\mu_0(\cdot)$ is a continuous function on $(0, T]$. Note that the $Y_i(\cdot)$ is observed only at times t_{ij}. In much of the functional data analysis literature, the observed data are interpolated and then treated as observed curves, which is not realistic. Here, we deal with the data in a realistic way. A second-order differential interpolated curve $\hat{Y}_i(\cdot)$ is used for each subject i, such as the cubic spline interpolation [22], then $\hat{Y}_i(t_{ij}) = Y_i(t_{ij})$ at all times t_{ij}. The second-order differential interpolation is chosen as needed in the asymptotic study. We assume the following model for the $\hat{Y}_i(\cdot)$ is

$$\hat{Y}_i(t) = \mu_0(t) + \hat{\epsilon}_i(t), \quad E[\hat{\epsilon}_i(t)] = 0, \quad E[\hat{\epsilon}_i^2(t)] = \sigma^2(t), \quad (i = 1, \dots, n). \tag{2}$$

To estimate the true mean function $\mu_0(\cdot)$, the RKHS approach is employed. Let $\mathbb{H}$ be a Hilbert space consisting of square integrable functions $\ell_2(T)$ on $[0, T]$, with a given inner product $< \cdot, \cdot >_{\mathbb{H}}$. For any mapping $G : [0, T]^2 \mapsto \ell_2(T)$, $K(s, t) :=< G(s, \cdot), G(t, \cdot) >_{\mathbb{H}}$ is a reproducing kernel for $\mathbb{H}$ and any reproducing kernel of $\mathbb{H}$ can be expressed in this form ([23], Theorem 4, p. 22). Thus, for a given Hilbert space, its reproducing kernel K is non-unique. The choice of an adequate kernel for a specific statistical inference is important (for details, see Section 3 below). However, a reproducing kernel under one inner product may not be a reproducing kernel under another inner product on the same space $\mathbb{H}$. Assume $\mu_0 \in \mathbb{H}$ and there is some RKHS with a known kernel $K(\cdot, \cdot)$. For any $h \in \mathbb{H}$, define $Kh = (Kh)(\cdot) =< K(\cdot,), h >_{\mathbb{H}}$. Let $< \cdot, \cdot >$ be another norm on $\mathbb{H}$, $||h||^2 =< h, h >$ for all $h \in \mathbb{H}$ (typically $||h||^2 = \int_T h^2(t)dt$). We estimate $\mu_0(\cdot)$ by

$$\hat{\mu}_{n,\lambda}(\cdot) = \arg \inf_{\mu \in \mathbb{H}} \left\{ \frac{1}{n} \sum_{i=1}^{n} ||\hat{Y}_i - \mu||^2 + \lambda J(\mu) \right\}, \tag{3}$$

where λ is a smoothing parameter, and $J(\mu) = ||K\mu||^2_{\mathbb{H}}$ is a penalty functional for some kernel K to be addressed. In the spline method, the penalty is of the form $||\mu^{(r)}||^2$ for some order r derivative of μ, which is a special case of the RKHS methods (see below).

Let $\mathbb{H}_0 = \{h \in \mathbb{H} : J(h) = 0\} \subset \mathbb{H}$ be the null space for the penalty functional, and let $\mathbb{H}_1$ be its orthogonal complement (with respect to the inner product $< \cdot, \cdot >_{\mathbb{H}}$). Then, $\mathbb{H} = \mathbb{H}_0 \oplus \mathbb{H}_1$, i.e., $\forall h \in \mathbb{H}$, it has the decomposition $h = h_0 + h_1$, with $h_0 \in \mathbb{H}_0$ and $h_1 \in \mathbb{H}_1$, and there are two kernel functions K_0 and K_1 such that $K = K_0 + K_1$, $\forall h \in \mathbb{H}$, $(K_0 h)(\cdot) :=< K_0(\cdot,), h >_{\mathbb{H}} \ \in \mathbb{H}_0$ and $(K_1 h)(\cdot) :=< K_1(\cdot,), h >_{\mathbb{H}} \ \in \mathbb{H}_1$. Here, $\mathbb{H}_1$ is also an RKHS with the reproducing kernel $K_1(\cdot, \cdot)$ on $\mathbb{H}_1$ and $(K_1 h) = h$, $\forall h \in \mathbb{H}_1$. Because $K_0\mu \in \mathbb{H}_0$ and $K_1\mu \in \mathbb{H}_1$, we have $||K_0\mu||^2_{\mathbb{H}} = 0$, $< K_0\mu, K_1\mu >_{\mathbb{H}} = 0$, and

$$\begin{aligned} J(\mu) = ||K\mu||^2_{\mathbb{H}} = ||K_0\mu + K_1\mu||^2_{\mathbb{H}} = ||K_0\mu||^2_{\mathbb{H}} + 2 < K_0\mu, K_1\mu >_{\mathbb{H}} + ||K_1\mu||^2_{\mathbb{H}} \\ = ||K_1\mu||^2_{\mathbb{H}} =< K_1\mu, K_1\mu >_{\mathbb{H}} . \end{aligned} \tag{4}$$

Note that the inner product $< \cdot, \cdot >_{\mathbb{H}}$ of the RKHS often is not the inner product $< \cdot, \cdot >$ used in the optimization objective, and the latter is often chosen as the L_2 norm. Thus, the expression of $J(\mu)$ in (4) does not hold under the inner product $< \cdot, \cdot >$. Often the norm $< \cdot, \cdot >$ in (3) is more suitable for statistical interpretation while the norm $< \cdot, \cdot >_{\mathbb{H}}$ is chosen for convenience of the computation of the penalty term $J(\mu)$.

The RKHS estimator often has a closed form solution called representer theorem. Such a result was known for decades. Ref. [13] presented such results in their case. Here, we present it in our case. Let $d = dim(\mathbb{H}_0)$ and $g_1, \dots, g_d$ be an orthonormal basis of $\mathbb{H}_0$.

Theorem 1. *Assume $\mu_0(\cdot), \hat{Y}_i(\cdot) \in \mathbb{H}$ for $i = 1, \dots, n$. Then, for the given penalty functional $J(\mu) = ||K_1\mu||^2_{\mathbb{H}}$ and fixed λ, there are constants $\mathbf{a} = (a_1, \dots, a_d)'$ and $\mathbf{b} = (b_1, \dots, b_n)'$ such that $\hat{\mu}_{n,\lambda}$ given in (3) has the following representation*

$$\hat{\mu}_{n,\lambda}(t) = \sum_{j=1}^{d} a_j g_j(t) + \sum_{i=1}^{n} b_i (K_1 \hat{Y}_i)(t), \quad t \in (0, T].$$

Thus, instead of searching a function in a Hilbert space, for minimizing (3), only the two parametric vectors **a** and **b** are to be estimated based on this represent theorem, which is called the optimal basis theorem in ([1], p. 363).

The λ in (3) is a smoothing parameter and $0 < \lambda < \infty$. Unlike functional regression estimation, when $\lambda \to 0$, the estimation of function μ will not break down, and $\hat{\mu}_{n,0}$ is obtained as the sample mean of the $\hat{Y}_i$ which does not have desirable smoothness. When $\lambda \to \infty$, the above procedure is equivalent to minimize $J(\mu)$, and in the case of $dim(\mathbb{H}_0) = 2$, the estimator of $\hat{\mu}_{n,\infty}$ is linear in t. The most commonly used method for the choice of the smoothing parameter is the cross-validation (CV) such that the λ minimizes

$$n^{-1}\sum_{i=1}^{n}\int_T \big(\hat{Y}_i(t) - \hat{\mu}_{n,\lambda,i}(t)\big)^2 dt,$$

where $\hat{\mu}_{n,\lambda,i}(\cdot)$ is the estimator in (3) without using the i-th observation $\hat{Y}_i$. However, this method is computationally intensive. An improved version of this method is the generalized cross-validation (GCV) proposed by [24,25]. For a given linear operator A on $\mathbb{H}$, let $\{\eta_i : i = 1, 2, \ldots\}$ be the eigenvalues of A such that $|\eta_1| \geq |\eta_2| \geq \ldots$. For integer m, define

$$||A||_m = \sum_{j=1}^{m}\eta_j, \quad MSE(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\int_T \big(\hat{Y}_i(t) - \hat{\mu}_{n,\lambda}(t)\big)^2 dt.$$

By Theorem 1, the estimator of $\mu_0(\cdot)$ can be written in the form $\hat{\mu}_{n,\lambda}(t) == (K_\lambda \hat{Y}_n)(t)$, and K_λ is a linear combination of K_0 and K_1. Let $\mathbf{I}$ be the identity operator on $\mathbb{H}$, and the smoothing parameter λ is chosen by minimizing the following $GCV(\cdot)$,

$$GCV(\lambda) = \big[\lim_{m\to\infty} m^{-1}||\mathbf{I} - K_\lambda||_m\big]^{-2} MSE(\lambda).$$

Obviously, the smoothing parameter λ above is dependent on the sample size n, and $\lambda(n) \to 0$ as $n \to \infty$. For simplicity, we will use λ instead of $\lambda(n)$ through this paper.

To obtain the asymptotic distribution of the proposed estimator in (3), a specific kernel function $K(\cdot,\cdot)$ has to be chosen.

Recall $d = dim(\mathbb{H}_0)$. A common choice is $d = 2$, $\mathbb{H}_0 = \{h : h^{(2)} \equiv 0\}$ is spanned by $g_1(t) \equiv 1$ and $g_2(t) = t$, and K_1 for $\mathbb{H}_1$ is chosen by

$$K_1(s,t) = \frac{1}{(2!)^2}B_2(s)B_2(t) - \frac{1}{4!}B_4(\lfloor s - t \rfloor),$$

where $B_r(\cdot)$ is the r-th Bernoulli polynomial, $\lfloor t \rfloor = t - [t]$ is the fractional part of t and $[t]$ is the integer part of t [19,20]. However, with this kernel function, the penalty term in (3) is $J(\mu) = \mathbf{b}'\Omega\mathbf{b}$ where $\Omega = (\omega_{ij})$ is a $n \times n$ matrix with $\omega_{ij} = \int_T\int_T \hat{Y}_i(s)K_1(s,t)\hat{Y}_j(t)dsdt$ (see proof of Theorem 1 in Appendix A). Then, the computation for $\hat{\mu}_{n,\lambda}(\cdot)$ in (3) suffers from the inverse of a $n \times n$ matrix, which is a hurdle for a large n, and is difficult to obtain the asymptotic distribution of the estimator.

To construct an adequate RKHS $\mathbb{H}$ on $L_2[0,T]$, we consider $\mathbb{H}_0 = \{h : h^{(2)}(\cdot) \equiv 0\}$ with inner product $< f, g >_{\mathbb{H},0}$, and its orthogonal complement $\mathbb{H}_1 = \{h : h^{(j)}(0) = 0, j = 0,1; \int_0^T h^{(2)}(t)dt < \infty\}$ with inner product $< f, g >_{\mathbb{H},1}$, where

$$< f, g >_{\mathbb{H},0} = \sum_{j=0}^{1} f^{(j)}(0)g^{(j)}(0), \qquad < f, g >_{\mathbb{H},1} = \int_0^T f^{(2)}(t)g^{(2)}(t)dt. \tag{5}$$

The inner product on $\mathbb{H}$ is defined as $< \cdot,\cdot >_{\mathbb{H}} = < \cdot,\cdot >_{\mathbb{H},0} + < \cdot,\cdot >_{\mathbb{H},1}$. Kernels for the RKHS with more general K_0 for $\mathbb{H}_0$ and K_1 for $\mathbb{H}_1$ with these inner products can be found in ([26], pp. 33–34). More general methods for construction of kernels K_0 and K_1 can be found in ([1] Section 20.3). In particular, we propose

$$K_0(s,t) = 1 + st, \quad K_1(s,t) = \int_0^T (s-u)_+(t-u)_+ du = (s \wedge t)^2\big(3(s \vee t) - (s \wedge t)\big)/6. \quad (6)$$

With the inner product given in (5), let $K = K_0 + K_1$, then $\forall h \in \mathbb{H}$, $h(t) =< K(t,\cdot), h(\cdot) >_{\mathbb{H}}$, and $\mathbb{H}_0$ and $\mathbb{H}_1$ are orthogonal to each other with respect to $< \cdot,\cdot >_{\mathbb{H}}$.

Let $\mathbf{g}(t) = (1,t)'$; $(K_0h)(t) =< K_0(t,\cdot), h(\cdot) >_{\mathbb{H},0}$, $(K_1h)(t) =< K_1(t,\cdot), h(\cdot) >_{\mathbb{H},1}$, $\mathbf{X}_n(t) = (K_1\hat{\mathbf{Y}})(t) = \big((K_1\hat{Y}_1)(t), \ldots, (K_1\hat{Y}_n)(t)\big)'$, $\bar{Y}_n(t) = \frac{1}{n}\sum_{i=1}^n \hat{Y}_i(t)$, $R =< \mathbf{g}(\cdot), \mathbf{g}'(\cdot) >_{\mathbb{H},0}$, $U_n =< \bar{Y}_n(\cdot), \mathbf{g}(\cdot) >_{\mathbb{H},0}$, $V_n =< \mathbf{g}(\cdot), \mathbf{X}_n'(\cdot) >_{\mathbb{H}}$, $S_n =< \bar{Y}_n(\cdot), \mathbf{X}_n(\cdot) >_{\mathbb{H}}$, $W_n =< \mathbf{X}_n(\cdot), \mathbf{X}_n'(\cdot) >_{\mathbb{H}} =< \mathbf{X}_n(\cdot), \mathbf{X}_n'(\cdot) >_{\mathbb{H},1}$ and

$$\Omega = (\omega_{ij})_{n\times n}, \quad \text{with} \quad \omega_{ij} =< \hat{Y}_i(\cdot), (K_1\hat{Y}_j)(\cdot) >_{\mathbb{H}} =< \hat{Y}_i(\cdot), (K_1\hat{Y}_j)(\cdot) >_{\mathbb{H},1}.$$

Denote $K_1^{(2)}(t,s) = \partial^2 K_1(t,s)/\partial s^2$, then $K_1^{(2)}(t,s) = t - s$ if $s \le t$ and 0 if $s > t$, and

$$(K_1\bar{Y}_n)(t) =< K_1(t,\cdot), \bar{Y}_n(\cdot) >_{\mathbb{H},1} = \int_0^t (t-s)\bar{Y}_n^{(2)}(s)ds = \bar{Y}_n(t) - \bar{Y}_n(0) - t\bar{Y}_n^{(1)}(0).$$

By Theorem 1, $\hat{\mu}_{n,\lambda}$ has the expression,

$$\hat{\mu}_{n,\lambda}(t) = \mathbf{a}'\mathbf{g}(t) + \mathbf{b}'\mathbf{X}_n(t), \quad (7)$$

and the coefficients **a** and **b** satisfy

$$\begin{cases} \mathbf{0} = -U_n + R\mathbf{a} + V_n\mathbf{b} \\ \mathbf{0} = -S_n + V_n'\mathbf{a} + (\lambda\Omega + W_n)\mathbf{b}, \end{cases} \quad (8)$$

because $\hat{\mu}_{n,\lambda}$ in (7) minimizes (4). The solution of (8) is given by

$$\begin{pmatrix} \hat{\mathbf{a}}_n \\ \hat{\mathbf{b}}_n \end{pmatrix} = \begin{pmatrix} R & V_n \\ V_n' & \lambda\Omega + W_n \end{pmatrix}^{-1} \begin{pmatrix} U_n \\ S_n \end{pmatrix}$$

and the estimate $\hat{\mu}_{n,\lambda}$ in (7) for fixed λ becomes

$$\begin{aligned} \hat{\mu}_{n,\lambda}(t) &= \textstyle\sum_{j=1}^2 \hat{a}_j g_j(t) + \sum_{i=1}^n \hat{b}_i (K_1\hat{Y}_i)(t) \\ &= (\mathbf{g}'(t), \mathbf{X}_n'(t)) \begin{pmatrix} R & V_n \\ V_n' & \lambda\Omega + W_n \end{pmatrix}^{-1} \begin{pmatrix} U_n \\ S_n \end{pmatrix}. \end{aligned}$$

Because each component of $\mathbf{X}_n(t)$ is an element of $\mathbb{H}_1$, each component of $g(t)$ is in $\mathbb{H}_0$, and $\mathbb{H}_0$ and $\mathbb{H}_1$ are orthogonal with respect to the inner product $< \cdot,\cdot >_{\mathbb{H}}$; then, $V_n =< \mathbf{g}, \mathbf{X}_n' >_{\mathbb{H}} = \mathbf{0}$, $W_n = \Omega$, and

$$S_n = \tfrac{1}{n} < \mathbf{X}_n(\cdot), \hat{\mathbf{Y}}'(t)1_n >_{\mathbb{H}} = \tfrac{1}{n} < \mathbf{X}_n(\cdot), \hat{\mathbf{Y}}'(\cdot) >_{\mathbb{H}} 1_n = \tfrac{1}{n}\Omega 1_n,$$

where 1_n is the n-dimensional vector of 1. Thus, we have

$$\hat{\mu}_{n,\lambda}(t) = \mathbf{g}'(t)U_n + (1+\lambda)^{-1}\frac{1}{n}\mathbf{X}_n'(t)1_n.$$

By definitions of $\mathbf{X}_n(t)$ and U_n, we have $\frac{1}{n}\mathbf{X}_n'(t)1_n = (K_1\bar{Y}_n)(t)$, $\mathbf{g}'(t)U_n = \mathbf{g}'(t) < \bar{Y}_n, \mathbf{g} >_{\mathbb{H},0} = (1,t)(\bar{Y}_n(0), \bar{Y}_n^{(1)}(0))' = \bar{Y}_n(0) + \bar{Y}_n^{(1)}(0)t =< K_0, \bar{Y}_n >_{\mathbb{H},0} (t)$, and (7) becomes

$$\begin{aligned} \hat{\mu}_{n,\lambda}(t) &= (K_0\bar{Y}_n)(t) + (1+\lambda)^{-1}(K_1\bar{Y}_n)(t) \\ &= (K_0\bar{Y}_n)(t) + (K_1\bar{Y}_n)(t) - \tfrac{\lambda}{1+\lambda}(K_1\bar{Y}_n)(t) \\ &= \bar{Y}_n(t) - \tfrac{\lambda}{1+\lambda}(K_1\bar{Y}_n)(t), \end{aligned} \quad (9)$$

because $K = K_0 + K_1$ is a reproducing kernel of $\mathbb{H}$, $[(K_0 + K_1)\bar{Y}_n](\cdot) = \bar{Y}_n(\cdot)$.

As a curve-smoothing curve estimation, $\hat{\mu}_{n,\lambda}(\cdot)$ is a biased estimator of $\mu_0(\cdot)$ and its bias is $\lambda(K_1\mu_0)(\cdot)$ (unless $\lambda = 0$ and so no smoothing regularization). Below, we consider the asymptotic normality of $\hat{\mu}_{n,\lambda}$. Denote $\xrightarrow{D}$ is the convergence in distribution. Let $l^\infty([0,T])$ be the space of bounded functions on $[0,T]$ equipped with the supreme norm, and $\overset{D}{\Rightarrow}$ for weak convergence in the space $l^\infty([0,T])$.

(C1). $E\int_T \big(\hat{Y}(t) - E[\hat{Y}(t)]\big)^2 dt < \infty$.
(C2). $\int_T \mu_0^2(t)dt < \infty$.
(C3). $\delta_n := \max_{j=1,\ldots,m_i-1; i=1,\ldots,n}(t_{i,j+1} - t_{ij}) \to 0$ (a.s.) as $n \to \infty$.

Theorem 2. *(i) Assume (C1)–(C3), and Ω defined above is invertible for all large n, then as $n \to \infty$ (also, $\lambda \to 0$),*

$$||\hat{\mu}_{n,\lambda} - \mu_0|| \to 0 \quad (a.s.)$$

(ii) In addition, $\mu_0(\cdot)$ is twice differentiable with its second-order derivative $\ddot{\mu}_0(\cdot)$, then

$$n^{1/2}\big[\hat{\mu}_{n,\lambda}(t) - \mu_0(t) - b_n(t)\big] \xrightarrow{D} N(0, \sigma^2(t)),$$

where $\sigma^2(t) = Var[Y(t)]$ and with $t_j \in \mathcal{S}$,

$$b_n(t) = \ddot{\mu}_0(t_j)(t_{j+1} - t)(t - t_j) - \frac{\lambda}{1+\lambda}(K_1\bar{Y}_n)(t) + o(t_{j+1} - t_j)^2, \quad \text{for} \quad t \in [t_j, t_{j+1}).$$

(iii) If we assume further that $\hat{Y}_i(\cdot), \mu_0(\cdot) \in \mathbb{H}_{(\alpha)}$ for all i for some $\alpha > 0$, then

$$\mathbb{G}_n(\cdot) := n^{1/2}\big[\hat{\mu}_{n,\lambda}(\cdot) - \mu_0(\cdot) - b_n(\cdot)\big] \overset{D}{\Rightarrow} \mathbb{G}(\cdot),$$

where $\mathbb{G}$ is the zero mean Gaussian process on $[0,T]$ with covariance function $R(s,t) = Cov[Y(s), Y(t)]$.

3. Hypothesis Tests for Mean Curves

In this section, two types of tests for mean curves of functional data are considered: one is to test the hypothesis of equal mean curves from two populations and another one is the hypothesis that the mean function $\mu(\cdot)$ belongs to some subspace $\mathcal{H}^0$ of $\mathcal{H}$.

3.1. Test the Equality of Two Mean Curves

Suppose two observed samples are $\{\hat{Y}_{1,i} : i = 1, \ldots n_1\}$ i.i.d. from Y_1 and $\{\hat{Y}_{2,i} : i = 1, \ldots, n_2\}$ i.i.d. from Y_2, with their mean curves $\mu_1(\cdot)$ and $\mu_2(\cdot)$, respectively. The two samples are assumed to be independent. For the RKHS $\mathbb{H}$ on $L_2[0,T]$ with inner product (5) and kernel (6), their mean curve estimates are given by (9) as

$$\hat{\mu}_{j,\lambda_j} = \bar{Y}_{n_j,j}(t) - \frac{\lambda_j}{1+\lambda_j}(K_1\bar{Y}_{n_j,j})(t), \quad (j = 1,2).$$

Let $|T|$ be the Lebesgue measure of the set $[0,T]$. We are to test the null hypothesis

$$H_0 : \mu_1(\cdot) = \mu_2(\cdot) \ \ (a.e.) \ \ vs \ \ H_1 : \mu_1(\cdot) \neq \mu_2(\cdot).$$

In the above, (a.e.) means almost everywhere, and $\mu_1(\cdot) \neq \mu_2(\cdot)$ means not equal on some set with nonzero Lebesque measure. For this, we propose the test statistic

$$T_n = \frac{1}{|T|}\left|\left|\frac{\sqrt{n_1 n_2}}{\sqrt{n_1+n_2}}(\hat{\mu}_{1,\lambda_1} - \hat{\mu}_{2,\lambda_2})\right|\right|^2 = \frac{1}{|T|}\frac{n_1 n_2}{n_1+n_2}\int_0^T \big[\hat{\mu}_{1,\lambda_1}(t) - \hat{\mu}_{2,\lambda_2}(t)\big]^2 dt,$$

where λ_1 and λ_2 are determined by (6).

Theorem 3. *Assume the conditions of Theorem 2 (iii) for the two samples,* $0 < n_1/(n_1+n_2) < 1$, *and* $(\lambda_1+\lambda_2)n_1n_2/(n_1+n_2) \to 0$. *Then, under* H_0,

$$T_n \xrightarrow{D} \frac{1}{|T|}\sum_{j=1}^{\infty}\gamma_j Z_j^2 := W,$$

where the Z_j *are i.i.d.* $N(0,1)$ *random variables, and the* γ_j *are the eigenvalues of*

$$R(s,t) = \sum_{j=1}^{2}\alpha_j^2 Cov[\hat{Y}_j(s), \hat{Y}_j(t)], \quad s,t \in [0,T],$$

and $\alpha_j = \lim \sqrt{n_j}/\sqrt{n_1+n_2}$, $(j = 1,2)$.

Theorem 3 can be viewed as a generalization of Mahalanobis statistic for finite-dimensional statistics; it is analogous to the result in ([27], p. 66). In fact, if we observe the $\mathbf{Y}_{1i} = (Y_{1i,1},\ldots,Y_{1i,k})'$ $(i = 1,\ldots,n_1)$ and $\mathbf{Y}_{2i} = (Y_{2i,1},\ldots,Y_{2i,k})'$ $(i = 1,\ldots,n_2)$ at fixed k time points, with corresponding mean values $(\mu_{1,1},\ldots,\mu_{1,k})'$ and $(\mu_{2,1},\ldots,\mu_{2,k})'$. We take $||\cdot||_{\mathbb{H}}$ as the L_2-norm and with no penalty, i.e., $\lambda_1 = \lambda_2 = 0$, then $\hat{\mu}_{1,j}$ and $\hat{\mu}_{2,j}$ are just the corresponding sample mean $(j = 1,2)$, and Theorem 4 reduces to

$$T_n = \frac{1}{k}\sum_{j=1}^{k}\frac{n_1n_2}{n_1+n_2}\left[\hat{\mu}_{1,j} - \hat{\mu}_{2,j}\right]^2 \xrightarrow{D} \sum_{j=1}^{k}\gamma_j Z_j^2,$$

with γ_j being the eigenvalues of $R = (r_{ij})_{1\le i,j\le k}$, $r_{ij} = \sum_{a,b=1}^{2}\alpha_a\alpha_b E[Y_{a,i}Y_{b,j}]$.

In practice, the eigenvalues of $R(s,t)$ above cannot be perfectly computed. As approximation, we compute the eigenvalues $\hat{\gamma}_j$ $(j = 1,\ldots,m)$ of the matrix $R_n = (r_{ij})_{1\le i,j\le m}$ for some specified large integer m, where $r_{ij} = \sum_{a,b=1}^{2}\alpha_a\alpha_b\bar{Y}_a(t_i)\bar{Y}_b(t_j)]$, $\bar{Y}_1(t_i) = n_1^{-1}\sum_{j=1}^{n_1}Y_{1j}(t_i)$ and $\bar{Y}_1(t_i) = n_2^{-1}\sum_{j=1}^{n_2}Y_{2j}(t_i)$.

If k is relatively large, only the first p largest eigenvalues are needed for good approximation, with some chosen p $(< k)$. Let $\lambda_1,\ldots,\lambda_p$ be the p largest eigenvalues, and $\hat{\lambda}_j$ be their estimates. Then, by Theorem 2.7 in ([27], p. 31), $E(\hat{\lambda}_j - \lambda_j)^2 = O(n^{-1})$, for all $1 \le j \le p$, i.e., the estimates are good up to order $O(n^{-1})$.

3.2. Test Mean Curve in Some Subspace

This type of test has been systematically studied since the late 1980s. Ref. [28] developed such a test for regression function, and many related references can be found therein. Without loss of generality, we only consider that the subspace $\mathcal{H}^0$ is of polynomials of degree no more than three, and the hypothesis

$$H_0 : \mu(\cdot) \in \mathcal{H}^0 \quad \text{versus} \quad H_1 : \mu(\cdot) \notin \mathcal{H}^0.$$

To test H_0, we need the penalized estimate $\tilde{\mu}_{n,\lambda}$ of μ in $\mathcal{H}^0$

$$\tilde{\mu}_{n,\lambda} = \arg\min_{\mu\in\mathcal{H}^0}\left[\frac{1}{n}\sum_{i=1}^{n}||\hat{Y}_i - \mu||^2 + \lambda J(\mu)\right].$$

Let $\mathbb{H}_{12}$ be the subspace spanned by $\{t^4\}$, $\mathbb{H}_{11} = \mathbb{H}_1 \setminus \mathbb{H}_{12}$, and $\mathbb{H}_1 = \mathbb{H}_{11} \oplus \mathbb{H}_{12}$. Let K_{11} and K_{12} be the reproducing kernels for $\mathbb{H}_{11}$ and $\mathbb{H}_{12}$. Let $\tilde{\Omega} = (\tilde{\omega}_{ij})$ be the $n \times n$ matrix with $\tilde{\omega}_{ij} =< \hat{Y}_i, (K_{11}\hat{Y}_j) >_{\mathbb{H}}$. Denote $\tilde{X}_n(t) = \left((K_{11}\hat{Y}_1)(t),\ldots,(K_{11}\hat{Y}_n)(t)\right)'$, $\tilde{V}_n =< g, \tilde{X}_n' >_{\mathbb{H}}$,

$\tilde{S}_n =< \bar{Y}_n, \tilde{X}_n >_{\mathbb{H}}$, $\tilde{W}_n =< \tilde{X}_n, \tilde{X}_n' >_{\mathbb{H}}$, and U_n, $\bar{Y}_n(t)$ and R as before. For $h \in \mathcal{H}$, let $(K_{11}h)(t) =< K_{11}(\cdot,t), h(\cdot) >_{\mathbb{H}_1}$ and $(K_{12}h)(t) =< K_{12}(\cdot,t), h(\cdot) >_{\mathbb{H}_1}$. Then,

$$\tilde{\mu}_{n,\lambda}(t) = \sum_{j=1}^{d} a_j g_j(t) + \sum_{i=1}^{n} b_i (K_{11}\hat{Y}_i)(t), \quad t \in [0,T].$$

and the coefficients $(\mathbf{a}, \mathbf{b})$ minimizing (18) are

$$\begin{pmatrix} \tilde{\mathbf{a}}_n \\ \tilde{\mathbf{b}}_n \end{pmatrix} = \begin{pmatrix} R & \tilde{V}_n \\ \tilde{V}_n' & \lambda\tilde{\Omega} + \tilde{W}_n \end{pmatrix}^{-1} \begin{pmatrix} U_n \\ \tilde{S}_n \end{pmatrix}$$

and the estimate $\tilde{\mu}_{n,\lambda}$, for fixed λ, is

$$\begin{aligned} \tilde{\mu}_{n,\lambda}(t) &= \textstyle\sum_{j=1}^{2} \tilde{a}_j g_j(t) + \sum_{i=1}^{n} \tilde{b}_i (K_{11}\hat{Y}_i)(t) \\ &= (g'(t), X_n'(t)) \begin{pmatrix} R & \tilde{V}_n \\ \tilde{V}_n' & \lambda\tilde{\Omega} + \tilde{W}_n \end{pmatrix}^{-1} \begin{pmatrix} U_n \\ \tilde{S}_n \end{pmatrix}, \qquad t \in [0,T]. \end{aligned}$$

When H_0 is true, $\hat{\mu}_{n,\lambda}(\cdot)$ and $\tilde{\mu}_{n,\lambda}(\cdot)$ are expected to be close, and so large observed absolute value of any monotone functional of their difference will be evidence against H_0. The following result characterizes such a difference and can be used to test H_0.

Theorem 4. *Under H_0, we have (i) Assume conditions of Theorem 2 (i), then*

$$n^{1/2}\left[\hat{\mu}_{n,\lambda}(t) - \tilde{\mu}_{n,\lambda}(t)\right] \xrightarrow{D} N(0, \tau^2(t)),$$

where $\tau^2(t) = E[D^2(t)]$, $D(t) = (K_{12}[Y - \mu_0])(t)$.
(ii) Assume conditions of Theorem 2 (iii), then

$$\mathbb{D}_n(\cdot) := n^{1/2}\left(\hat{\mu}_{n,\lambda}(\cdot) - \tilde{\mu}_{n,\lambda}(\cdot)\right) \overset{D}{\Rightarrow} \mathbb{D}(\cdot),$$

where $\mathbb{D}(\cdot)$ is the mean zero Gaussian process on T with covariance function $Q(s,t) = Cov[D(s), D(t)]$.

In application, $\tau^2(\cdot)$ is estimated by $\hat{\tau^2}(t) = (n-1)^{-1} \sum_{i=1}^{n} \{(K_{12}[Y_i - \tilde{\mu}_{n,\lambda}])(t)\}^2$, and $Q(s,t)$ is estimated by $\hat{Q}(s,t) = n^{-1} \sum_{i=1}^{n} K_{12}[Y_i - \tilde{\mu}_{n,\lambda}])(s) K_{12}[Y_i - \tilde{\mu}_{n,\lambda}])(t)$.

4. Numerical Analysis

To investigate the finite sample properties of the proposed procedures, we perform two simulation studies. The first study is to compare the proposed RKHS estimator of mean curves with the conventional local linear smooth and spline methods. The second study is to examine the performance of statistic T_n for testing the equality of two mean curves. Then, a real data analysis illustrates the performance of our proposed procedures in this paper well.

Simulation 1. *To compare with the commonly used local linear fit (R-package lowess) and spline smoother (R-package smooth.spline), we consider the estimator $\hat{\mu}_{n,\lambda}(\cdot)$ with the spacial choices of the kernel and inner products given in (22). We assume that the underlying individual curve i at time points $t \in T = [0,10]$ is generated from $y_i(t) \sim N(\mu(t), \sigma^2(t))$, where $\mu(t) = t\sin(5+3t)$, $\sigma^2(t) = t^2$. For each subject i, the number of observation time points m_i is assumed to generate from the uniform distribution on $\{5,6,\ldots,20\}$ and the observation time points $\mathbf{t}_i = (t_{i1}, \ldots, t_{i,m_i})$ are generated from $Exp(\lambda)$ with $\lambda = 0.6$. Then, interpolate the $y(t_{ij})$ on T to obtain $\hat{y}_i(\cdot)$.*

The fitted results are presented in Figure 2 with sample sizes of 50, 100 and 200, respectively.

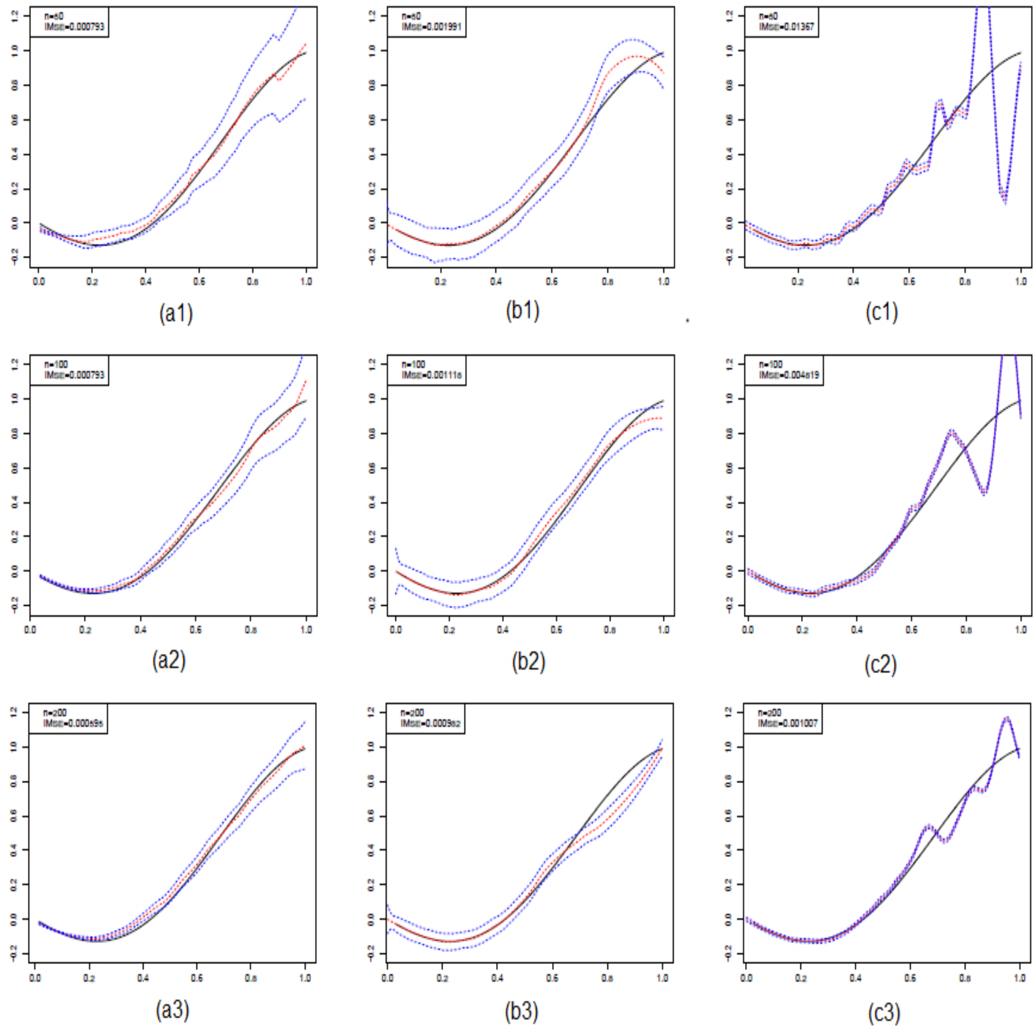

Figure 2. Solid line: true curve; dotted middle blue line: estimated curve; dotted lower and upper blue lines: 95% confidence bands. The first, second and third rows are for sample size of 50, 100 and 200, respectively. In each row, the left panel is the proposed method, the middle panel is the local linear smoother and the right panel is the spline estimate.

The simulation shows that the proposed method (fitted curves (a1)–(a3) has a better performance than the other two methods. The RKHS estimator has a relatively stable performance; it has narrower confidence bands at the relatively dense region, and it becomes wider at the sparse region. For the local linear smoother (b1)–(b3), the width of its confidence bands has no apparent difference when the observation points change from relatively dense to sparse. The estimated curves with the local linear smoothing method have biases due to the sparse observation time points. The spline has a very good fit and confidence band when the data observation points are relatively dense, but the estimated curves have a large bias when the data become sparse, as seen in (c1)–(c3), near the right end of the x-axis, and the spline estimates are not stable, with some moderate to large fluctuations.

Simulation 2. *In this simulation study, we examine the performance of the statistic T_n for testing the hypothesis H_0, the equality of the mean curves of two stochastic processes with the alternative hypothesis that two mean curves are not equal. We assume that the samples are generated as $x_i(t) \sim N(\mu(t), \sigma^2(t))$, where $\mu(t) = t\sin(5 + (3t))$, $\sigma^2(t) = t^2$, $(i = 1, \ldots, 50)$, and $y_i(t) \sim N(\eta(t), \sigma^2(t))$, where $\eta(t) = \mu(t) + C\cos(2 + (2t))$, $(i = 1, \ldots, 30)$, respectively, and C is a turning parameter to measure the amount of difference between the two samples. We take the* *parameter C to be 0, 0.5, 0.7, 0.8, 0.9 and 1.0 with corresponding meaningful differences* $\Delta_1 = |T|^{-1}\int_T (\mu(t) - \eta(t))^2 dt$ *which are to be 0, 0.179, 0.352, 0.460, 0.582 and 0.718. For each subject i, the number of observation time points* m_i *and the observation time points* $\mathbf{t}_i = (t_{i1}, \ldots, t_{i,m_i})$ *are generated as in Simulation 1 above. The simulation results are presented in Table 1 based on 10,000 replications. Table 1 shows that the computed type I error is slightly less than the nominal type I error of 0.05 (first row with* $C = 0$*) and powers for* $C \geq 0.5$*.*

Table 1. Power/type I error of T_n-based simulated 10,000 replications.

Turning Parameter (C)	Δ	T_n	Power
0	0	0.070	0.044
0.5	0.179	7.111	0.120
0.7	0.352	14.690	0.293
0.8	0.460	18.446	0.504
0.9	0.582	23.548	0.872
1.0	0.718	30.616	0.999

Real Data Analysis

With the proposed method, we analyze the growth data from the National Growth and Health Study (NGHS) project (https://biolincc.nhlbi.nih.gov/studies/nghs/, accessed on 1 March 2016), sponsored by the National Heart, Lung, and Blood Institute, from 1985 to 2000. The main purpose of the study is to investigate the differences between Caucasian and African-American girls in the development of obesity in pubescent females due to psychosocial, socioeconomic and other environmental factors.

The NGHS is an epidemiological study of the cardiovascular risk factors in 1166 Caucasian and 1213 African-American girls during adolescence. In this longitudinal study, starting from age 9, each subject had a baseline examination and annual examinations. The study was renewed twice to continue the longitudinal investigation until the subjects reached the age of 19 to 20. We deleted those individuals (about 30 for each race) with only 1 or 2 observations, as our method is for longitudinal data, and the total number of subjects is $n = n_1 + n_2 = 1136 + 1183 = 2319$. The number of follow-up visits for each subject varies from 3 to 10. The ages vary between 9 and 19 years, and we use the age range $T = [9, 19)$. The body mass index (BMI) is used as the response $y(t)$ as a function of age t. The mean curves are estimated using (22) for the two groups separately. The results are presented in Figure 3 which suggest that the girls' BMI increases with age and shows that African-American girls tend to have higher BMIs than Caucasian girls.

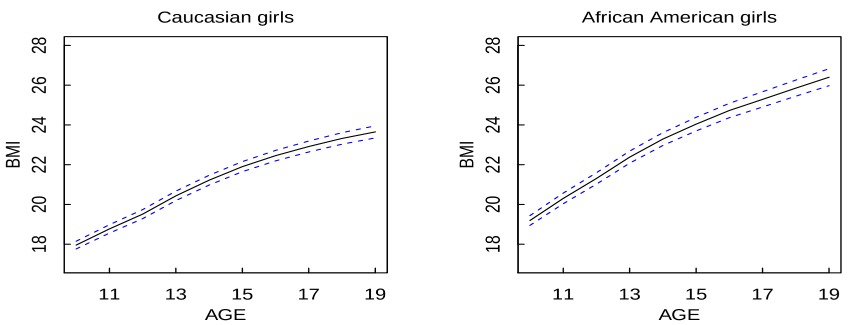

Figure 3. The estimated mean curves of girls' BMI.

Then, we test the null hypothesis of the equality of two mean curves for the two samples (Caucasian and African-American girls). The value of the statistic T_n is 2576.302, the 95% upper quartile of W in Theorem 6 is 938.439 and the corresponding p-value is 0.0015. Thus, the difference between Caucasian and African-American girls is statistically significant based on the observed data. The conclusion is consistent with other studies on these data. For example, Ref. [29] analyzed the same data. They used a varying coefficient model to analyze the regression relationship between systolic blood pressure (SBP) and age and race. Their conclusion is that African-American girls tend to have higher probabilities of "SBP $>$ 100 mmHg" than Caucasian girls so that race is a factor affecting the SBP. It is known that SBP is strongly related to BMI.

5. Concluding Remarks

We have proposed and studied the reproducing kernel Hilbert space method for the analysis of functional data, motivated by a practical problem, in which the observation points are relatively dense in some time intervals and sparse in other time intervals. The unbalanced observed time points result in a biased estimation or a large variance of the estimation based on the current methods. The simulation studies indicate an apparent advantage of the proposed method compared to some commonly used methods for this type of data. We derived extensive theoretical results for the RKHS estimation, including convergence rates of estimates with two commonly used norms.

To use the RKHS methods for a functional data analysis, the key is to choose an adaptive kernel. Different kernels may result in a very different computational efficiency. In this paper, we proposed a special kernel and the corresponding estimator of the mean curve for functional data has a very simple expression. The asymptotic distribution of the estimator is also given. Furthermore, we proposed two statistics for testing the hypothesis of equal mean curves from two populations and the hypothesis that the mean function belongs to some subspace. The finite sample performance of the proposed methods is evaluated by the simulation studies and the methods provided new insight in the analysis of functional growth data in the NIH NGHS study.

As future works, we can extend the RKHS method to the case of case-control studies with observational functional data. With observational data, treatment assignment is often not randomized as in the ideal case; it is known that the naive estimate of a treatment–effect curve is biased and a causal inference method is needed. A popular such method is the doubly robust estimator commonly used in ordinary data. To construct such an estimator, a propensity score model and an outcome model will be specified, and as long as one of the models is correctly specified, the resulting estimator will be unbiased. To extend this estimator to functional data is non-trivial and will be our future work. Another possible extension is to consider the missing responses in the longitudinal data, with the case of missing not at random (MNAR), which is a topic of general interest.

Author Contributions: Conceptualization, A.Y. and H.-B.F.; methodology, A.Y.; software, M.X.; validation, A.Y., M.X. and H.-B.F.; formal analysis, M.X.; investigation, M.X., A.Y., H.-B.F.; resources, A.Y., H.-B.F., C.O.W. and M.T.T.; data curation, C.O.W., H.-B.F. and M.X.; writing—original draft preparation, A.Y.; writing—review and editing, A.Y. and H.-B.F.; visualization, M.X., A.Y. and H.-B.F.; supervision, M.T.T.; project administration, M.T.T.; funding acquisition, M.T.T. All authors have read and agreed to the published version of the manuscript

Funding: The research of Xiong, Yuan, Fang and Tan is partially supported by the National Cancer Institute (NCI) grants R01CA164717.

Informed Consent Statement: Patient consent was waived as the data is opened to public.

Data Availability Statement: The growth data from the national growth and health study (NGHS) project is available at https://biolincc.nhlbi.nih.gov/studies/nghs/, accessed on 1 March 2016.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Proof of Theorem 1. Let $\mathcal{H}$ be the (closed) subspace linear spanned by $\{g_j(\cdot), (K_1\hat{Y}_i)(\cdot) : j = 1, \ldots, d; i = 1, \ldots, n.\}$ with respect to the norm $< \cdot, \cdot >$. Denote $\mathbf{a} = (a_1, \ldots, a_d)'$, $\mathbf{g}(t) = (g_1(t), \ldots, g_d(t))'$, $\mathbf{b} = (b_1, \ldots, b_n)'$ and $(K_1\hat{\mathbf{Y}})(t) = ((K_1\hat{Y}_1)(t), \ldots, (K_1\hat{Y}_n)(t))'$. Let $r(\cdot) \in \mathbb{H} \setminus \mathcal{H}$ be orthogonal to $\mathbf{g}(\cdot)$ and $(K_1\hat{\mathbf{Y}})(\cdot)$ with respect to the norm $< \cdot, \cdot >$. We rewrite $\hat{\mu}_{n,\lambda}(\cdot)$ as

$$\hat{\mu}_{n,\lambda}(t) = \mathbf{a}'\mathbf{g}(t) + \mathbf{b}'(K_1\hat{\mathbf{Y}})(t) + r(t). \tag{A1}$$

We need only to show $r \equiv 0$ if $\hat{\mu}_{n,\lambda}(\cdot)$ in (A1) minimizes the right hand of (4).

Because K is the kernel function on $\mathbb{H}$, we have that $K = K_0 + K_1$, and $\forall h \in \mathbb{H}$, $h = (Kh) = (K_0h) + (K_1h) = h_0 + h_1$, $h_j = K_jh \in \mathbb{H}_j$ $(j = 0, 1)$. As $\hat{Y}_i \in \mathbb{H}$, $K_0\hat{Y}_i \in \mathbb{H}_0$, and by definition of $\mathbb{H}_0$, $||K_0\hat{Y}_i||^2_{\mathbb{H}} = 0$, so $K_0\hat{Y}_i = 0$. Thus, $\hat{Y}_i = K\hat{Y}_i = K_0\hat{Y}_i + K_1\hat{Y}_i = K_1\hat{Y}_i \in \mathcal{H}$ $(i = 1, \ldots, n)$. Then, $r(\cdot)$ is also orthogonal to $\hat{Y}_i(\cdot)$ $(i = 1, \ldots, n)$ (with respect to the norm $< \cdot, \cdot >$, because $r \in \mathbb{H} \setminus \mathcal{H}$), and

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n ||\hat{Y}_i - \hat{\mu}_{n,\lambda}||^2 &= \frac{1}{n}\sum_{i=1}^n ||\hat{Y}_i - \sum_{j=1}^d a_j g_j - \sum_{j=1}^n b_j(K_1\hat{Y}_j) - r||^2 \\
&= \frac{1}{n}\sum_{i=1}^n \Big(||\hat{Y}_i - \sum_{j=1}^d a_j g_j - \sum_{j=1}^n b_i(K_1\hat{Y}_j)||^2 - 2 < \hat{Y}_i, r > \\
&\quad -2\sum_{j=1}^d a_j < g_j, r > -2\sum_{j=1}^n b_j < K_1\hat{Y}_j, r > + ||r||^2 \Big) \\
&= \frac{1}{n}\sum_{i=1}^n ||\hat{Y}_i - \sum_{j=1}^d a_j g_j - \sum_{j=1}^n b_j(K_1\hat{Y}_j)||^2 + ||r||^2.
\end{aligned} \tag{A2}$$

Since $(K_1g_j) \equiv 0$, we have $(K_1\hat{\mu}_{n,\lambda}) = \mathbf{b}'(K_1\hat{\mathbf{Y}}) + (K_1r)$ and

$$\begin{aligned}
J(\hat{\mu}_{n,\lambda}) &= < K_1\hat{\mu}_{n,\lambda}, K_1\hat{\mu}_{n,\lambda} >_{\mathbb{H}} \\
&= \mathbf{b}' < K_1\hat{\mathbf{Y}}, K_1\hat{\mathbf{Y}} >_{\mathbb{H}} \mathbf{b} + 2\mathbf{b}' < K_1\hat{\mathbf{Y}}, K_1r >_{\mathbb{H}} + < K_1r, K_1r >_{\mathbb{H}} .
\end{aligned} \tag{A3}$$

Furthermore, K_1 is a linear operator $\{\mathbb{H}, < \cdot, \cdot >\} \mapsto \{K_1\mathbb{H}, < \cdot, \cdot >_{\mathbb{H}}\}$. Let $K_1^* : \{K_1\mathbb{H}, < \cdot, \cdot >_{\mathbb{H}}\} \mapsto \{\mathbb{H}, < \cdot, \cdot >\}$ be its adjoint operator. Because K_1 and K_1^* are continuous bounded linear operators, $\mathcal{H}$ is closed with respect to both the norms $|| \cdot ||$ and $|| \cdot ||_{\mathbb{H}}$. Thus, $K_1^*(K_1\hat{\mathbf{Y}}) \subset K_1^*\mathcal{H} \subset \mathcal{H}$ (with respect to the norm $< \cdot, \cdot >$), and obtain $< K_1\hat{\mathbf{Y}}, K_1r >_{\mathbb{H}} =< K_1^*(K_1\hat{\mathbf{Y}}), r >= \mathbf{0}$ due to the orthogonality of $r(\cdot)$ to $\mathcal{H}$, *i.e.*, $< h, r >= 0$, $\forall\ h \in \mathcal{H}$. (A3) becomes

$$J(\hat{\mu}_{n,\lambda}) = \mathbf{b}' < K_1\hat{\mathbf{Y}}, K_1\hat{\mathbf{Y}} >_{\mathbb{H}} \mathbf{b} + < K_1r, K_1r >_{\mathbb{H}} = \mathbf{b}'\Omega\mathbf{b} + J(r), \tag{A4}$$

where $\Omega = (\omega_{ij})_{n\times n}$ with $\omega_{ij} =< K_1\hat{Y}_i, K_1\hat{Y}_j >_{\mathbb{H}}$. From (A2) and (A4), we obtain

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n ||\hat{Y}_i - \hat{\mu}_{n,\lambda}||^2 + \lambda J(\hat{\mu}_{n,\lambda}) &= \frac{1}{n}\sum_{i=1}^n ||\hat{Y}_i - \sum_{j=1}^d a_j g_j - \sum_{j=1}^n b_i(K_1\hat{Y}_j)||^2 \\
&\quad + ||r||^2 + \lambda\left[\mathbf{b}'\Omega\mathbf{b} + J(r)\right].
\end{aligned} \tag{A5}$$

For any $(\mathbf{a}, \mathbf{b})$, (A5) is minimized when $||r||^2 + \lambda J(r) = 0$. $J(r) \geq 0$ and $\lambda > 0$ imply $r = 0$ and Theorem 1 is proven. □

Proof of Theorem 2. (i) Because $\mathbb{H}_0$ and $\mathbb{H}_1$ are orthogonal with respect to the inner product $< \cdot, \cdot >_{\mathbb{H}}$, each component of $\mathbf{X}_n(t)$ is an element of $\mathbb{H}_1$, and each component of $\mathbf{g}(t)$ is in $\mathbb{H}_0$, then $V_n =< \mathbf{g}, \mathbf{X}_n' >_{\mathbb{H}} = \mathbf{0}$. Denote $W_n = (w_{ij})_{n\times n}$, $w_{ij} =< K_1\hat{Y}_i, K_1\hat{Y}_j >_{\mathbb{H}}$. Because $(K_1\hat{Y}_i)(t) = \hat{Y}_i(t) - (K_0\hat{Y}_i)(t)$, $(K_1\hat{Y}_i)(t) \in \mathbb{H}_1$ and $(K_0\hat{Y}_i)(t) \in \mathbb{H}_0$,

$$\begin{aligned}
w_{ij} &= < \hat{Y}_i - K_0\hat{Y}_i, K_1\hat{Y}_j >_{\mathbb{H}} =< \hat{Y}_i, K_1\hat{Y}_j) >_{\mathbb{H}} - < K_0\hat{Y}_i, K_1\hat{Y}_j) >_{\mathbb{H}} \\
&= < \hat{Y}_i, K_1\hat{Y}_j >_{\mathbb{H}} =< \hat{Y}_i, K_1\hat{Y}_j) >_{\mathbb{H},1} = \omega_{ij},
\end{aligned}$$

i.e., $W_n = \Omega$. Now, we have

$$\begin{aligned} \hat{\mu}_{n,\lambda}(t) &= (\mathbf{g}'(t), \mathbf{X}_n'(t)) \begin{pmatrix} R & \mathbf{0} \\ \mathbf{0} & (1+\lambda)\Omega \end{pmatrix}^{-1} \begin{pmatrix} U_n \\ S_n \end{pmatrix} \\ &= \mathbf{g}'(t) R^{-1} U_n + (1+\lambda)^{-1} \mathbf{X}_n'(t) \Omega^{-1} S_n. \end{aligned} \tag{A6}$$

Let $\hat{\mathbf{Y}}(t) = (\hat{Y}_1(t), \ldots, \hat{Y}_n(t))'$, I_2 be the 2×2 identity matrix, and 1_n be the n-dimensional vector of 1. Note $R = I_2$, $\bar{Y}_n(t) = n^{-1} \hat{\mathbf{Y}}'(t) 1_n$, and

$$S_n = \frac{1}{n} < \mathbf{X}_n(\cdot), \hat{\mathbf{Y}}'(t) 1_n >_{\mathbb{H}} = \frac{1}{n} < \mathbf{X}_n(\cdot), \hat{\mathbf{Y}}'(\cdot) >_{\mathbb{H}} 1_n = \frac{1}{n} \Omega 1_n.$$

Then, (A6) becomes

$$\hat{\mu}_{n,\lambda}(t) = \mathbf{g}'(t) U_n + (1+\lambda)^{-1} \frac{1}{n} \mathbf{X}_n'(t) 1_n.$$

By definition of $\mathbf{X}_n(t)$, we have $\frac{1}{n} \mathbf{X}_n'(t) 1_n = (K_1 \bar{Y}_n)(t)$. Recall the definition of U_n, $\mathbf{g}'(t) U_n = \mathbf{g}'(t) < \bar{Y}_n, \mathbf{g} >_{\mathbb{H},0} = (1,t)(\bar{Y}_n(0), \bar{Y}_n^{(1)}(0))' = \bar{Y}_n(0) + \bar{Y}_n^{(1)}(0) t =< K_0, \bar{Y}_n >_{\mathbb{H},0} (t)$. We have that

$$\begin{aligned} \hat{\mu}_{n,\lambda}(t) &= (K_0 \bar{Y}_n)(t) + (1+\lambda)^{-1} (K_1 \bar{Y}_n)(t) \\ &= (K_0 \bar{Y}_n)(t) + (K_1 \bar{Y}_n)(t) - \tfrac{\lambda}{1+\lambda} (K_1 \bar{Y}_n)(t) = \bar{Y}_n(t) - \tfrac{\lambda}{1+\lambda} (K_1 \bar{Y}_n)(t). \end{aligned}$$

Thus, $\hat{\mu}_{n,\lambda}(\cdot)$ is the smoothing of $\bar{Y}_n(\cdot)$ by the two operators K_0 and K_1. Part i) of the theorem is proved by the fact that $\lambda = \lambda(n) \to 0$, and $E(\bar{Y}_n(\cdot)) \to E(Y(\cdot)) = \mu_0$ under conditions (C1)–(C3).
(ii) Under (C1)–(C3), the $\hat{Y}_i$'s are i.i.d. but are dependent of n as the t_{ij} is. Let $\mathcal{S}_n = \{T_i, i = 1, 2, \ldots, n\}$ and $\mu_n(\cdot) = E(\bar{Y}(\cdot) | \mathcal{S}_n)$, the conditional expectation under the given observed time points of n subjects. We have

$$\begin{aligned} n^{1/2} \big(\hat{\mu}_{n,\lambda}(t) - \mu_0(t) \big) &= n^{1/2} \bigg(\big(\bar{Y}_n(t) - \mu_n(t) \big) + \big(\mu_n(t) - \mu_0(t) - \tfrac{\lambda}{1+\lambda} (K_1 \bar{Y}_n)(t) \big) \bigg) \\ &= n^{1/2} \big(\bar{Y}_n(t) - \mu_n(t) + b_n(t) \big), \end{aligned}$$

where $b_n(t) = \big(\mu_n(t) - \mu_0(t) \big) - \frac{\lambda}{1+\lambda} (K_1 \bar{Y}_n)(t)$. Note

$$n^{1/2} \big(\bar{Y}_n(t) - \mu_n(t) \big) = n^{-1/2} \sum_{i=1}^{n} \big(\hat{Y}_i(t) - E[\hat{Y}_i(t)] \big).$$

The sequence $\hat{Y}_i(t) - E[\hat{Y}_i(t)]$ is i.i.d. with mean 0 and variance $\sigma_n^2(t) = var(\hat{Y}_i(t)) \to var(Y(t)) = \sigma^2(t)$ as $n \to \infty$. Then, using the central limit theory,

$$n^{-1/2} \sum_{i=1}^{n} \big(\hat{Y}_i(t) - E[\hat{Y}_i(t)] \big) \xrightarrow{D} N(0, \sigma^2).$$

Similarly as in Theorem 1 in [30], by the assumption $\mu_0 \in \mathcal{H}_\alpha$, we obtain $\mu_n(t) - \mu_0(t) = O(\delta_n^\alpha)$, and for $t \in [t_j, t_{j+1})$ with $t_j \in \mathcal{S}$,

$$\mu_n(t) - \mu_0(t) = \ddot{\mu}_0(t_j)(t_{j+1} - t)(t - t_j) + o(t_{j+1} - t_j)^2.$$

Thus,

$$b_n = (\ddot{\mu}_0(t_j)(t_{j+1} - t)(t - t_j) - \frac{\lambda}{1+\lambda} (K_1 \bar{Y}_n)(t) + o(t_{j+1} - t_j)^2.$$

(iii) Let $\rho(s,t) = |Y(s) - Y(t)|$ and $\rho_n(s,t) = \left(n^{-1}\sum_{i=1}^{n}(\hat{Y}_i(s) - \hat{Y}_i(t))^2\right)^{1/2}$. The condition $\hat{Y}_i(\cdot), \mu_0(\cdot) \in \mathbb{H}_{(\alpha)}$ for all i for some $\alpha > 0$, and Corollary 2.7.2 in [31], we have

$$\log N_{[\,]}(\epsilon, \mathbb{H}_{(\alpha)}, L_2) \leq C\epsilon^{-1/(2\alpha)} < \infty, \quad \text{(A7)}$$

for any $\epsilon > 0$ and some constant C. Then, $N_{[\,]}(\epsilon, \mathbb{H}_{(\alpha)}, L_1) < \infty$ for every $\epsilon > 0$. By Theorem 2.4.1 in [26], $\mathbb{H}_{(\alpha)}$ is a Gelivenko–Cantelli class, which implies $\sup_{s,t\in T}|\rho_n(s,t) - \rho(s,t)| \rightarrow 0(a.s.)$. Because $\mathbb{H}_{(\alpha)}$ is bounded, there is an envelop $M < \infty$ such that $\sup_{t\in T}\max_i\left(\hat{Y}_i(t) - E[\hat{Y}_i(t)]\right)^2 \leq M$. Moreover, $N(\epsilon, \mathbb{H}_{(\alpha)}, L_2) \leq N_{[\,]}(\epsilon/2, \mathbb{H}_{(\alpha)}, L_2)$ and (A7) imply the uniform entropy condition

$$\int_0^\infty \sup_Q \sqrt{\log N(\epsilon M, \mathbb{H}_\alpha, L_2(Q))}d\epsilon < \infty,$$

where the $\sup_Q$ is for all probability measures Q. Assume suitable measurable conditions, then by Theorem 2.8.9 in [31], $\mathbb{H}_{(\alpha)}$ is a Donsker class, i.e., $\mathbb{G}_n \overset{D}{\Rightarrow} \mathbb{G}$ in $l^\infty(T)$, and $\mathbb{G}$ is a Gaussian process with mean 0 and covariance function $R(s,t) = Cov[Y(s), Y(t)]$. □

Proof of Theorem 3. Let $\mathbb{G}_{j,n_j}(\cdot) = n_j^{1/2}(\cdot)\left(\bar{Y}_{j,n_j}(\cdot) - \mu_j(\cdot)\right)$. In the proof of Theorem 2, we see that $\hat{\mu}_{j,\lambda_j}(t) = \bar{Y}_{j,n_j}(t) - [\lambda_j/(1+\lambda_j)](K_1\bar{Y}_{j,n_j})(t)$, where $\bar{Y}_{j,n_j}(\cdot)$ is the corresponding sample mean. Under H_0, $\mu_1(\cdot) = \mu_2(\cdot)$,

$$\begin{aligned} T_n &= \frac{1}{|T|}\Bigg|\Bigg|\frac{n_2}{[n_1+n_2]^{1/2}}\mathbb{G}_{1,n_1}(\cdot) + \frac{n_1}{[n_1+n_2]^{1/2}}\mathbb{G}_{2,n_2}(\cdot) \\ &\quad - \sum_{j=1}^{2}\frac{\lambda_j}{1+\lambda_j}\frac{n_1^{1/2}n_2^{1/2}}{[n_1+n_2]^{1/2}}(K_1\bar{Y}_{j,n_j})(\cdot)\Bigg|\Bigg|^2 \\ &= \frac{1}{|T|}\Bigg|\Bigg|\alpha_1\mathbb{G}_{1,n_1}(\cdot) + \alpha_2\mathbb{G}_{2,n_2}(\cdot)\Bigg|\Bigg|^2 + o_p(1). \end{aligned}$$

By Theorem 2 (iii),

$$\alpha_1\mathbb{G}_{1,n_1}(\cdot) + \alpha_2\mathbb{G}_{2,n_2}(\cdot) \overset{D}{\Rightarrow} \mathbb{G}(\cdot),$$

where $\mathbb{G}(\cdot) = \alpha_1\mathbb{G}_1(\cdot) + \alpha_2\mathbb{G}_2(\cdot)$, $\mathbb{G}_1(\cdot)$ and $\mathbb{G}_2(\cdot)$ are independent, $\mathbb{G}_j(\cdot)$ is a mean zero Gaussian process on T, with covariance function $R_j(s,t) = Cov[Y_j(s), Y_j(t)]$ $(s,t \in T)$ $(j = 1,2)$. Thus, $\mathbb{G}(\cdot)$ is a mean zero Gaussian process on T, with covariance function

$$R(s,t) = \sum_{j=1}^{2}\alpha_j^2 R_j(s,t), \quad s,t \in T.$$

Hence, we have

$$T_n \overset{D}{\rightarrow} \frac{1}{|T|}||\mathbb{G}(\cdot)||^2.$$

Because $R(\cdot,\cdot)$ is a.e. continuous and T is bounded, then $R^2(\cdot,\cdot)$ is integrable, i.e., $\int_T\int_T R^2(s,t)dsdt < \infty$. By Mercer's Theorem (see Theorem 5.2.1 in [23], p. 208), we have

$$R(s,t) = \sum_{j=1}^{\infty}\gamma_j h_j(s)h_j(t),$$

where $\gamma_j \geq 0$ $(j = 1,2,\ldots)$ are the eigenvalues of $R(\cdot,\cdot)$, and $h_j(\cdot)$ $(j = 1,2,\ldots)$ are the corresponding orthonormal eigenfunctions ($\int_T h_i(t)h_j(t)dt = 0$ for $i \neq j$, and $\int_T h_i^2(t)dt = 1$ for all i). Let $Z_1,\ldots,Z_m,\ldots$ be i.i.d. random variables and $Z_m \sim N(0,1)$, then $\mathbb{Z}(t) =$

$\sum_{j=1}^{\infty} \sqrt{\gamma_j} Z_j h_j(t)$ is a Gaussian process on T with mean zero and covariance function $R(s,t)$. Thus, two stochastic processes $\mathbb{G}(\cdot)$ and $\mathbb{Z}(\cdot)$ have the same distribution on T, i.e.,

$$\mathbb{G}(\cdot) \stackrel{d}{=} \mathbb{Z}(\cdot) = \sum_{j=1}^{\infty} \sqrt{\gamma_j} Z_j h_j(\cdot)$$

and for $||\cdot||$ being the L_2-norm,

$$\frac{1}{|T|}||\mathbb{G}(\cdot)||^2 = \frac{1}{|T|}\int_T \mathbb{G}^2(t)dt \stackrel{d}{=} \frac{1}{|T|}\int_T \left(\sum_{j=1}^{\infty} \sqrt{\gamma_j} Z_j h_j(t)\right)^2 dt = \frac{1}{|T|}\sum_{j=1}^{\infty} \gamma_j Z_j^2.$$

□

Proof of Theorem 4. (i) As in the proof of Theorem 3 (i), $\tilde{\mu}_{n,\lambda}(t) = (K_0\bar{Y}_n)(t) + (1+\lambda)^{-1}(K_{11}\bar{Y}_n)(t)$, and we have

$$\begin{aligned}\left[\hat{\mu}_{n,\lambda}(t) - \tilde{\mu}_{n,\lambda}(t)\right] &= (1+\lambda)^{-1}([K_1 - K_{11}]\bar{Y}_n)(t) \\ &= (1+\lambda)([K_1 - K_{11}]\bar{Y}_n)(t) + O(\lambda^2) = (1+\lambda)(K_{12}\bar{Y}_n)(t).\end{aligned}$$

Note that under H_0, $E\{(K_{12}\bar{Y}_n)(\cdot)\} = ([K_1 - K_{11}]\mu_0)(\cdot) = 0$. So, under H_0,

$$\begin{aligned}n^{1/2}\left[\hat{\mu}_{n,\lambda}(t) - \tilde{\mu}_{n,\lambda}(t)\right] &= n^{1/2}\Big((1+\lambda)(K_{12}[\bar{Y}_n - \mu_0])(t) + O(\lambda^2)\Big) \\ &= (1+\lambda)n^{1/2}(K_{12}[\bar{Y}_n - \mu_0])(t) + o(1) \xrightarrow{D} N(0, \tau^2(t)),\end{aligned}$$

where $\tau^2(t) = E\left[K_{12}[Y - \mu_0])(t)\right]^2$.
(ii) The proof is similar to that of Theorem 2 (ii). □

References

1. Ramsay, J.O.; Silverman, B.W. *Functional Data Analysis*; Springer: New York, NY, USA, 2005.
2. Clarkson, D.B.; Fraley, C.; Gu, C.; Ramsay, J.O. *S+ Functional Data Analysis*; Springer: New York, NY, USA, 2005.
3. Ferraty, F.; Vieu, P. *Nonparametric Fuctional Data Analysis*; Springer: New York, NY, USA, 2006.
4. Zhang, C.; Peng, H.; Zhang, J.-T. Two samples tests for functional data. *Commun. Stat. Theory Methods* **2010**, *39*, 559–578. [CrossRef]
5. Degras, D. Simultaneous confidence bands for the mean of functional data. *WIRS Comput. Stat.* **2017**, *9*, e1397. [CrossRef]
6. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer: New York, NY, USA, 2009.
7. Shi, M.; Weiss, R.E.; Taylor, J.M.G. An analysis of paediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves. *Appl. Stat.* **1996**, *45*, 151–163. [CrossRef]
8. Rice, J.A.; Wu, C.O. Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* **2001**, *57*, 253–259. [CrossRef] [PubMed]
9. Huang, J.Z.; Wu, C.O.; Zhou, L. Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* **2002**, *89*, 111–128. [CrossRef]
10. Yao, F.; M*ü*ller, H.-G.; Wang, J.-L. Functional data analysis for sparse longitudinal data. *J. Am. Stat. Assoc.* **2005**, *100*, 577–590. [CrossRef]
11. Staicu, A.-M.; Li, Y.; Ruppert, D.; Crainiceanu, C.M. Likelihood ratio tests for dependent data with applications to longitudinal and functional data analysis. *Scand. J. Stat.* **2014**, *41*, 932–949. [CrossRef]
12. Wang, J.-L.; Chiou, J.-M.; M*ü*ller, H.-G. Functional data analysis. *Annu. Rev. Stat. Its Appl.* **2016**, *3*, 257–295. [CrossRef]
13. Cai, T.; Yuan, M. Optimal estimation of the mean functions based on discretely sampled functional data: Phase transition. *Ann. Stat.* **2011**, *39*, 2330–2355. [CrossRef]
14. Happ, C.; Greven, S. Multivariate functional principal component analysis for data observed on different (dimensional) domains. *J. Am. Stat. Assoc.* **2018**, *113*, 649–659. [CrossRef]
15. Flores, M.; Naya, S.; Fern*á*ndez-Casal, R.; Zaragoza, S.; Rana, P.; Tarrio-Saavedra, J. Constructing a control chart using functional data. *Mathematics* **2020**, *8*, 58. [CrossRef]
16. Carroll, C.; Müller, H.G.; Kneip, A. Cross-component registration for multivariate functional data, with application to growth curves. *Biometrics* **2021**, *77*, 839–851 [CrossRef]
17. Mel*é*ndez, R.; Giraldo, R.; Leiva, V. Sign, Wilcoxon and Mann-Whitney tests for functional data: An approach based on random projections. *Mathematics* **2021**, *9*, 44. [CrossRef]

18. Ran, M.; Yang, Y. Optimal estimation of large functional and longitudinal data by using functional linear mixed model. *Mathematics* **2022**, *10*, 4322. [CrossRef]
19. Yuan, M.; Cai, T. A reproducing kernel Hilbert space approach to functional linear regression. *Ann. Stat.* **2010**, *38*, 3412–3444. [CrossRef]
20. Wahba, G. *Spline Models for Observational Data*; SIAM: Philadelphia, PA, USA, 1990.
21. Li, Y.; Liu, Y.; Zhu, J. Quantile regression in reproducing kernel Hilbert space. *J. Am. Stat.* **2007**, *102*, 255–268. [CrossRef]
22. Hazewinkel, M. Spline interpolation. In *Encyclopedia of Mathematics 1*; Springer: New York, NY, USA, 2001.
23. Berlinet, A.; Thomas-Agnan, C. *Reproducing Kernel Hilbert Space in Probability and Statistics*; Kluwer Academic Publishers: : Dordrecht, The Netherlands, 2004.
24. Wahba, G. A survey of some smoothing problems and the method of generalized cross-validation for solving them. In *Applications of Statistics*; Krisnaiah, P.R., Ed.; North Holland: Amsterdam, The Netherlands, 1977; pp. 507–523.
25. Craven, P.; Wahba, G. Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.* **1979**, *31*, 377–403. [CrossRef]
26. Gu, C. *Smoothing Spline ANOVA Models*; Springer: New York, NY, USA, 2002.
27. Horváth, L.; Kokoszka, P. *Inference for Functional Data with Applications*; Springer: New York, NY, USA, 2012.
28. Stute, W. Nonparametric model checks for regression. *Ann. Stat.* **1997**, *25*, 613–641. [CrossRef]
29. Wu, C.O.; Tian, X. Nonparametric estimation of conditional distributions and rank-tracking probabilities with time-varying transformation models in longitudinal studies. *J. Am. Stat. Assoc.* **2013**, *108*, 971–982. [CrossRef]
30. Yuan, A.; Fang, H.-B.; Wu, C.O.; Tan, M.T. Hypothesis testing for multiple mean and correlation curves with functional data. *Stat. Sin.* 2019, *in press*. [CrossRef]
31. Van der Vaart, A.; Wellner, J. *Weak Convergence and Empirical Processes*; Springer: New York, NY, USA, 1996.