Convergence rate of principal component analysis with local-linear smoother for functional data under a unified weighing scheme

Xingyu Yan, Xiaolong Pu, Yingchun Zhou & Xiaolei Xun

Published online: 20 Aug 2019.

Submit your article to this journal

View related articles

View Crossmark data

Taylor & Francis
Taylor & Francis Group

Check for updates

Convergence rate of principal component analysis with local-linear smoother for functional data under a unified weighing scheme

Xingyu Yan[a], Xiaolong Pu[a], Yingchun Zhou[a] and Xiaolei Xun[b]

[a]Key Laboratory of Advanced Theory and Application in Statistics, and Data Science – MOE, School of Statistics, East China Normal University, Shanghai, People's Republic of China; [b]School of Data Science, Fudan University, Shanghai, People's Republic of China

ABSTRACT

The unified weighing scheme for the local-linear smoother in analysing functional data can deal with data that are dense, sparse or of neither type. In this paper, we focus on the convergence rate of functional principal component analysis using this method. Almost sure asymptotic consistency and rates of convergence for the estimators of eigenvalues and eigenfunctions have been established. We also provide the convergence rate of the variance estimation of the measurement error. Based on the results, the number of observations within each curve can be of any rate relative to the sample size, which is consistent with the earlier conclusions about the asymptotic properties of the mean and covariance estimators.

1. Introduction

In this article, we consider the typical functional data setting, where a sample of n curves are observed over the time range $\mathcal{T}$, each at m_i discrete points for $i = 1, \ldots, n$. When analysing such data, sparsity of the time grid at which the measurements are observed should be taken into account, and proper estimation procedures would be adopted accordingly. Conventionally, pre-smoothing the observations from each subject is viable for dense data before subsequent analysis, whereas the subjects are pooled to borrow information for sparse data; furthermore, two types of estimation procedures present different asymptotic properties (Zhang & Wang, 2016). For the essential problem of estimating the mean and covariance functions, we refer to, for example, Ferraty and Vieu (2006), Ramsay and Silverman (2005), Rice and Silverman (1991), Staniswalis and Lee (1998), Zhou, Lin, and Liang (2017) and references therein.

Here we focus on local-linear smoother, which has high popularity due to its conceptual simplicity, attractive local features and ability for automatic boundary correction (Fan & Gijbels, 1996). To ensure that the effect of each curve on the optimisers is not overly affected by the denseness of observations, different weighing schemes have been proposed. A scatter plot smoother is employed by Yao, Müller, and Wang (2005), which assigns the same weight to each observation for sparse functional data analysis, referred to as the 'OBS' scheme. Alternatively, Li and Hsing (2010) suggested

a unified framework in which the number of observations within each curve can be of any rate relative to the sample size, where each subject received the same weight, referred to as the 'SUBJ' scheme. More recently, Zhang and Wang (2016) proposed a more general weighing scheme which includes the previous two commonly used schemes as special cases, and provided a comprehensive and unifying analysis of the asymptotic properties for estimation.

In addition to the essential estimation problem, functional principal component analysis (FPCA) has become a common part in functional data analysis, for example, to achieve dimension reduction of functional data by summarising the data in a few functional principal component (FPC) scores, or to interpret the varying trend of individual trajectories with the eigenfunctions. For a comprehensive discussion on FPCA, one may refer to Greven, Crainiceanu, Caffo, and Reich (2010), Hall, Müller, and Wang (2006), James, Hastie, and Sugar (2000), Jiang and Wang (2010), Yao and Lee (2006), and the references therein. Although there has been a lot of literature on FPCA, only a few theoretical studies on FPCA have been made, such as in Hall and Hosseini-Nasab (2006), Hall et al. (2006) and Li and Hsing (2010), and they are all based on 'OBS' or 'SUBJ' scheme. The convergence rates of the eigenvalues and eigenfunctions for the FPCA have not been studied under the general weighing scheme. The theoretical results in the paper not only provide the upper bounds of the convergence rates of

eigenvalues and eigenfunctions under a unified weighing scheme but also provide bases for further theoretical studies on functional clustering and classification based on the FPCA method.

The work of Zhang and Wang (2016) established the asymptotic normality, L^2 convergence, and uniform convergence for the mean and covariance estimators, but not the asymptotic properties of the FPC. Under the general weighing scheme, we provide in this article the almost sure convergence rate for eigenvalues and eigenfunctions, and further the convergence rate for the variance estimation of the measurement errors. The rest of the article is organised as follows. Notations, model and methodology including FPCA are included in Section 2. The main results about convergence rate in eigenvalues, eigenfunctions and the variance estimation of measurement errors are established in Section 3, with all technical proofs left to the Appendix. Simulation studies to verify the theoretical results are shown in Section 4. The concluding remarks are given in Section 5.

2. Model and methodology

Consider a random process $X(t)$ defined on a fixed interval $\mathcal{T} = [0, 1]$ with mean function $\mu(t) = E\{X(t)\}$ and covariance function $\gamma(s, t) = \text{cov}\{X(s), X(t)\}$. Denote with Y_{ij} the error-prone observations of the random process at random points T_{ij}, that is

$$Y_{ij} = X_i(T_{ij}) + \varepsilon_{ij}, \ i = 1, \ldots, n, \ j = 1, \ldots, m_i;$$

where the X_is are realisations of X, $\varepsilon_{ij} = \varepsilon_i(t_{ij})$ are identically distributed measurement errors with mean zero and variance σ^2, and all the X_is, T_{ij}s and ε_{ij}s are assumed to be independent.

2.1. Local-linear smoother

A local-linear estimator of the mean function $\widehat{\mu}(t) = \widehat{\beta}_0$ is obtained by minimising the weighted least squares

$$\sum_{i=1}^{n} \omega_i \sum_{j=1}^{m_i} K_{h_1}(T_{ij} - t) \left\{ Y_{ij} - \beta_0 - \beta_1(T_{ij} - t) \right\}^2,$$

with respect to (β_0, β_1), where $K_h(\cdot) = h^{-1}K(\cdot/h)$ is a kernel with bandwidth h. It was proposed in Zhang and Wang (2016) to assign weight ω_i to each observation for the ith subject such that $\sum_{i=1}^{n} m_i\omega_i = 1$, which is the general weighing scheme. Specifically, assignment $\omega_i = 1/\sum_{i=1}^{n} m_i$ along with $\nu_i = 1/\sum_{i=1}^{n} m_i(m_i - 1)$ leads to the OBS scheme; assignment $\omega_i = 1/(nm_i)$ along with $\nu_i = 1/nm_i(m_i - 1)$ leads to the SUBJ scheme.

To estimate the covariance function $\gamma(s, t)$, we first estimate $G(s, t) = E\{X(s)X(t)\}$, and then it follows that

$$\widehat{\gamma}(s, t) = \widehat{G}(s, t) - \widehat{\mu}(s)\widehat{\mu}(t). \qquad (1)$$

Similarly as before, a local-linear estimator $\widehat{G}(s, t) = \widehat{\beta}_0$ is obtained by minimising the weighted least squares

$$\sum_{i=1}^{n} \nu_i \sum_{1 \leq j \neq l \leq m_i} K_{h_2}(T_{ij} - s)K_{h_2}(T_{il} - t)$$

$$\left\{ Y_{ij}Y_{il} - \beta_0 - \beta_{11}(T_{ij} - s) - \beta_{12}(T_{il} - t) \right\}^2,$$

where weight ν_i is attached to each $Y_{ij}Y_{il}$ for the ith subject such that $\sum_{i=1}^{n} m_i(m_i - 1)\nu_i = 1$.

Finally to estimate the variance of measurement errors, we start by a local-linear estimator $\widehat{V}(t) = \widehat{\beta}_0$ of $V(t) := G(t, t) + \sigma^2$, obtained by minimising

$$\sum_{i=1}^{n} \omega_i \sum_{j=1}^{m_i} K_{h_3}(T_{ij} - t)\{Y_{ij}^2 - \beta_0 - \beta_1(T_{ij} - t)\}^2.$$

We then estimate σ^2 by $\widehat{\sigma}^2 = \int_0^1 \{\widehat{V}(t) - \widehat{G}(t, t)\}\mathrm{d}t$. Throughout this article, we select the bandwidths h_1, h_2 and h_3 by using the leave-one-out cross-validation method.

2.2. Functional principal component analysis

We consider a spectral decomposition of $\gamma(s, t)$ and its approximation. According to Mercer's theorem, the covariance function has the spectral decomposition

$$\gamma(s, t) = \sum_{j=1}^{\infty} \lambda_j\phi_j(s)\phi_j(t),$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ are the eigenvalues of $\gamma(\cdot, \cdot)$, and the ϕ_j's are the corresponding eigenfunctions, i.e., the principal components, which form an orthonormal system on the space of square-integrable functions on [0,1]. Following the Karhunen–Loève expansion, $X_i(t)$ is represented as $X_i(t) = \mu(t) + \sum_{j=1}^{\infty} \xi_{ij}\phi_j(t)$, where $\xi_{ij} = \int_0^1 X_i(t)\phi_j(t)\mathrm{d}t$ is referred to as the jth FPC score of the ith subject. For each i, the ξ_{ij} s are uncorrelated random variables with $E(\xi_{ij}) = 0$ and $E(\xi_{ij}^2) = \lambda_j$.

With the local-linear estimate $\widehat{\gamma}(s, t)$, we can approximate it with

$$\widehat{\gamma}(s, t) = \sum_{j=1}^{K} \widehat{\lambda}_j\widehat{\phi}_j(s)\widehat{\phi}_j(t),$$

where $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \cdots \geq 0$ are the estimated eigenvalues, the $\widehat{\phi}_j$s are the corresponding estimated eigenfunctions, and K is the number of principal components selected. We refer to Hall et al. (2006) and Yao et al. (2005) for comprehensive discussions about computation of the eigenvalues and eigenfunctions of an integral operator with a symmetric kernel. We refer to Li, Wang, and Carroll (2013) for intensive discussions about the choice of K and the underlying theory of the functional principal component analysis.

3. Main results about convergence rates

A few more notations are introduced first. Given a positive integer k, we denote $M_{Sk} = \sum_{i=1}^{n} m_i^k$, $\overline{M} = \sum_{i=1}^{n} m_i/n$, $\overline{M}_{Sk} = M_{Sk}/n$ and $\overline{M}_{Hk} = (n^{-1} \sum_{i=1}^{n} m_i^{-k})^{-1}$, where the subscript 'H' in $\overline{M}_H$ suggests a harmonic mean. The asymptotic behaviour of estimated eigenvalues and eigenfunctions is given in Theorem 3.1.

Theorem 3.1: *Suppose that the regularity conditions (A1)–(A2), (B1)–(B4), (C1)–(C2) and (D1)–(D2) in the Appendix hold. For any fixed j, we have the following results :*

(a) *convergence rate of estimated eigenvalue*

$$
\widehat{\lambda}_j - \lambda_j = O\Bigg((\log(n)/n)^{1/2} + h_1^2 + h_2^2 + \log(n)
$$
$$
\left[\frac{\sum_{i=1}^{n} m_i \omega_i^2}{h_1} + \sum_{i=1}^{n} m_i(m_i - 1)\omega_i^2 \right]
$$
$$
+ \log(n) \left[\frac{\sum_{i=1}^{n} m_i(m_i - 1)v_i^2}{h_2^2} \right.
$$
$$
+ \frac{\sum_{i=1}^{n} m_i(m_i - 1)(m_i - 2)v_i^2}{h_2}
$$
$$
+ \sum_{i=1}^{n} m_i(m_i - 1)(m_i - 2)(m_i - 3)v_i^2 \Bigg] \Bigg)
$$
$$
a.s.
$$

(b) *convergence rate of estimated eigenfunction*

$$
\sup_{t\in[0,1]} |\widehat{\phi}_j(t) - \phi_j(t)|
$$
$$
= O\Bigg(h_1^2 + \left\{ \log(n) \left[\frac{\sum_{i=1}^{n} m_i \omega_i^2}{h_1} \right. \right.
$$
$$
+ \sum_{i=1}^{n} m_i(m_i - 1)\omega_i^2 \Bigg]\Bigg\}^{1/2} + h_2^2
$$
$$
+ \left\{ \log(n) \left[\frac{\sum_{i=1}^{n} m_i \omega_i^2}{h_2} + \sum_{i=1}^{n} m_i(m_i - 1)\omega_i^2 \right] \right\}^{1/2}
$$
$$
+ \log(n) \left[\frac{\sum_{i=1}^{n} m_i(m_i - 1)v_i^2}{h_2^2} \right.
$$
$$
+ \frac{\sum_{i=1}^{n} m_i(m_i - 1)(m_i - 2)v_i^2}{h_2}
$$
$$
+ \sum_{i=1}^{n} m_i(m_i - 1)(m_i - 2)(m_i - 3)v_i^2 \Bigg] \Bigg) \quad a.s.
$$

In the following, we state in Corollary 3.2 and Corollary 3.3 the specialised results for the OBS and SUBJ schemes, respectively. For either the OBS or SUBJ scheme, convergence rate depends on the order of $\overline{M}_{Sk}$ and $\overline{M}_{Hk}$ relative to n and the order of bandwidth.

Corollary 3.2: *Suppose that the conditions in Theorem 3.1 hold, along with two additional assumptions (C3) and (D3). Then under the OBS scheme, for any fixed j,*

(a)
$$
\widehat{\lambda}_j - \lambda_j = O\Bigg((\log(n)/n)^{1/2} + h_1^2 + h_2^2
$$
$$
+ \left(\frac{1}{\overline{M}h_1} + \frac{\overline{M}_{S2}}{(\overline{M})^2} \right) \frac{\log(n)}{n}
$$
$$
+ \left(\frac{1}{\overline{M}_{S2}h_2^2} + \frac{\overline{M}_{S3}}{(\overline{M}_{S2})^2 h_2} \right.
$$
$$
+ \frac{\overline{M}_{S4}}{(\overline{M}_{S2})^2} \Bigg) \frac{\log(n)}{n} \Bigg) \quad a.s.
$$

(b)
$$
\sup_{t\in[0,1]} |\widehat{\phi}_j(t) - \phi_j(t)|
$$
$$
= O\Bigg(h_1^2 + \sqrt{ \left(\frac{1}{\overline{M}h_1} + \frac{\overline{M}_{S2}}{(\overline{M})^2} \right) \frac{\log(n)}{n} } + h_2^2
$$
$$
+ \sqrt{ \left(\frac{1}{\overline{M}h_2} + \frac{\overline{M}_{S2}}{(\overline{M})^2} \right) \frac{\log(n)}{n} }
$$
$$
+ \left(\frac{1}{\overline{M}_{S2}h_2^2} + \frac{\overline{M}_{S3}}{(\overline{M}_{S2})^2 h_2} \right.
$$
$$
+ \frac{\overline{M}_{S4}}{(\overline{M}_{S2})^2} \Bigg) \frac{\log(n)}{n} \Bigg) \quad a.s.
$$

Corollary 3.3: *Suppose that the conditions in Theorem 3.1 hold. Then under the SUBJ scheme, for any fixed j,*

(a)
$$
\widehat{\lambda}_j - \lambda_j
$$
$$
= O\Bigg((\log(n)/n)^{1/2} + h_1^2 + h_2^2
$$
$$
+ \left(\frac{1}{\overline{M}_H h_1} + 1 \right) \frac{\log(n)}{n}
$$
$$
+ \left(\frac{1}{\overline{M}_{H2}h_2^2} + \frac{1}{\overline{M}_H h_2} + 1 \right) \frac{\log(n)}{n} \Bigg) \quad a.s.
$$

(b)
$$
\sup_{t\in[0,1]} |\widehat{\phi}_j(t) - \phi_j(t)|
$$
$$
= O\Bigg(h_1^2 + \sqrt{ \left(\frac{1}{\overline{M}_H h_1} + 1 \right) \frac{\log(n)}{n} } + h_2^2
$$
$$
+ \sqrt{ \left(\frac{1}{\overline{M}_H h_2} + 1 \right) \frac{\log(n)}{n} }
$$
$$
+ \left(\frac{1}{\overline{M}_{H2}h_2^2} + \frac{1}{\overline{M}_H h_2} + 1 \right) \frac{\log(n)}{n} \Bigg) \quad a.s.
$$

In addition, we provide below the convergence rate of the estimated variance of the measurement error under the general weighing scheme, as well as the special cases of the OBS and SUBJ schemes.

Theorem 3.4: *Assume that the conditions in Theorem 3.1, and conditions (C1') and (C2') hold. Then under the general weighing framework,*
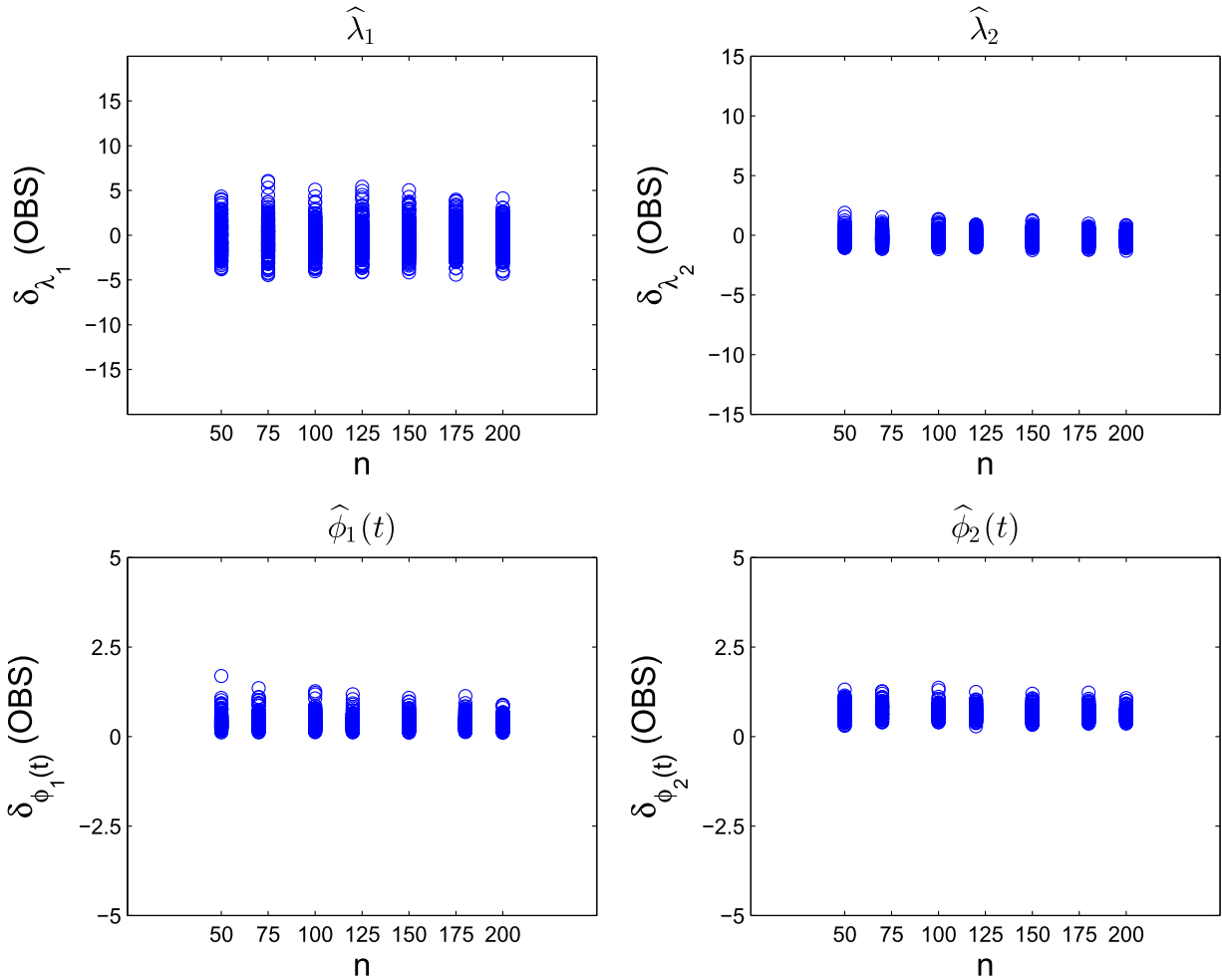
$$\widehat{\sigma}^2 - \sigma^2$$

$$= O\left(h_2^2 + \left\{ \log(n)\left[\frac{\sum_{i=1}^{n} m_i \omega_i^2}{h_1} \right.\right.\right.$$

$$\left.\left. + \sum_{i=1}^{n} m_i(m_i - 1)\omega_i^2 \right] \right\}^{1/2}$$

$$+ \log(n)\left[\frac{\sum_{i=1}^{n} m_i(m_i - 1)\nu_i^2}{h_2^2} \right.$$

$$+ \frac{\sum_{i=1}^{n} m_i(m_i - 1)(m_i - 2)\nu_i^2}{h_2}$$

$$\left. + \sum_{i=1}^{n} m_i(m_i - 1)(m_i - 2)(m_i - 3)\nu_i^2 \right]$$

$$+ h_3^2 + \log(n)\left[\frac{\sum_{i=1}^{n} m_i \omega_i^2}{h_3} \right.$$

$$\left.\left. + \sum_{i=1}^{n} m_i(m_i - 1)\omega_i^2 \right] \right) \quad a.s.$$

Corollary 3.5: *Suppose that the conditions in Theorem 3.4 hold.*

(a) *OBS: With an additional assumption (C3),*

$$\widehat{\sigma}^2 - \sigma^2$$

$$= O\left(h_2^2 + \sqrt{\left(\frac{1}{\overline{M}h_2} + \frac{\overline{M}_{S2}}{(\overline{M})^2} \right)\frac{\log(n)}{n}} \right.$$

$$+ \left(\frac{1}{\overline{M}_{S2}h_2^2} + \frac{\overline{M}_{S3}}{(\overline{M}_{S2})^2 h_2} + \frac{\overline{M}_{S4}}{(\overline{M}_{S2})^2} \right)\frac{\log(n)}{n}$$

$$\left. + h_3^2 + \left(\frac{1}{\overline{M}h_3} + \frac{\overline{M}_{S2}}{(\overline{M})^2} \right)\frac{\log(n)}{n} \right) \quad a.s.$$

Figure 1. This is for the 'OBS' case. The top two panels present the values of δ_{λ_1} and δ_{λ_2} for 200 replications as n increases, respectively. The bottom two panels present the values of $\delta_{\phi_1(t)}$ and $\delta_{\phi_2(t)}$ for 200 replications as n increases, respectively.

(b) SUBJ:

$$\widehat{\sigma}^2 - \sigma^2$$

$$= O\left(h_2^2 + \sqrt{\left(\frac{1}{\overline{M}_H h_2} + 1 \right) \frac{\log(n)}{n}} \right.$$

$$+ \left(\frac{1}{\overline{M}_{H2} h_2^2} + \frac{1}{\overline{M}_H h_2} + 1 \right) \frac{\log(n)}{n} + h_3^2$$

$$\left. + \left(\frac{1}{\overline{M}_H h_3} + 1 \right) \frac{\log(n)}{n} \right).$$

4. Simulation

To illustrate the theoretical results in Section 2, we now turn to the numerical performance of the estimators as sample size increases. Choices on ω_i and ν_i that satisfy the conditions in the general weighing scheme must be made. Since the 'OBS' and 'SUBJ' cases are the two most commonly used choices, and our corollaries showed the specific results regarding these two cases, we use these two cases as examples to illustrate the theoretical results. The data are generated from the following model:

$$Y_{ij} = X_i(T_{ij}) + \varepsilon_{ij},$$

where $X_i(t) = \mu(t) + \sum_{k=1}^{2} \xi_{ik} \phi_k(t)$, $\xi_{ik} \sim N(0, \lambda_k)$ for $k = 1, 2$ and $\varepsilon_{ij} \sim N(0, \sigma^2)$. Let $\mu(t) = t + \sin(t) + \cos(t)$, $\phi_1(t) = -\sqrt{2} \cos(\pi t)$, $\phi_2(t) = \sqrt{2} \sin(\pi t)$, set $(\lambda_1, \lambda_2, \sigma^2) = (0.6, 0.3, 0.2)$.

The observation times are generated in the following way. Each individual has a set of 'scheduled' time points, $\{1, 2, \ldots, 20\}$, and each scheduled time has a 20% probability of being skipped. The actual observation time is a random perturbation of a scheduled time: a uniform [0,1] random variable is added to a nonskipped scheduled time. This results in different observed time points T_{ij} per subject.

To illustrate that the convergence rates of the estimated $\widehat{\lambda}_k$, $\widehat{\phi}_k$ and $\widehat{\sigma}^2$ have the orders of magnitude shown in Section 2, let $\delta_{\lambda_k} = (\widehat{\lambda}_k - \lambda_k)/a_n$, where a_n is the derived convergence rate (e.g., under the OBS scheme, $a_n = (\log(n)/n)^{1/2} + h_1^2 + h_2^2 + \left(\frac{1}{\overline{M} h_1} + \frac{\overline{M}_{S2}}{(\overline{M})^2} \right) \frac{\log(n)}{n} + \left(\frac{1}{\overline{M}_{S2} h_2^2} + \frac{\overline{M}_{S3}}{(\overline{M}_{S2})^2 h_2} + \frac{\overline{M}_{S4}}{(\overline{M}_{S2})^2} \right) \frac{\log(n)}{n}$), if the estimated $\widehat{\lambda}_k$ is actually consistent with λ_k by this order of magnitude, the range of the term δ_{λ_k} should decrease or remain constant as n increases. Here we set $n = 50, 75, 100, 125, 150, 175, 200$ for both 'OBS' and 'SUBJ' cases, and 200 replications were done for each sample size.
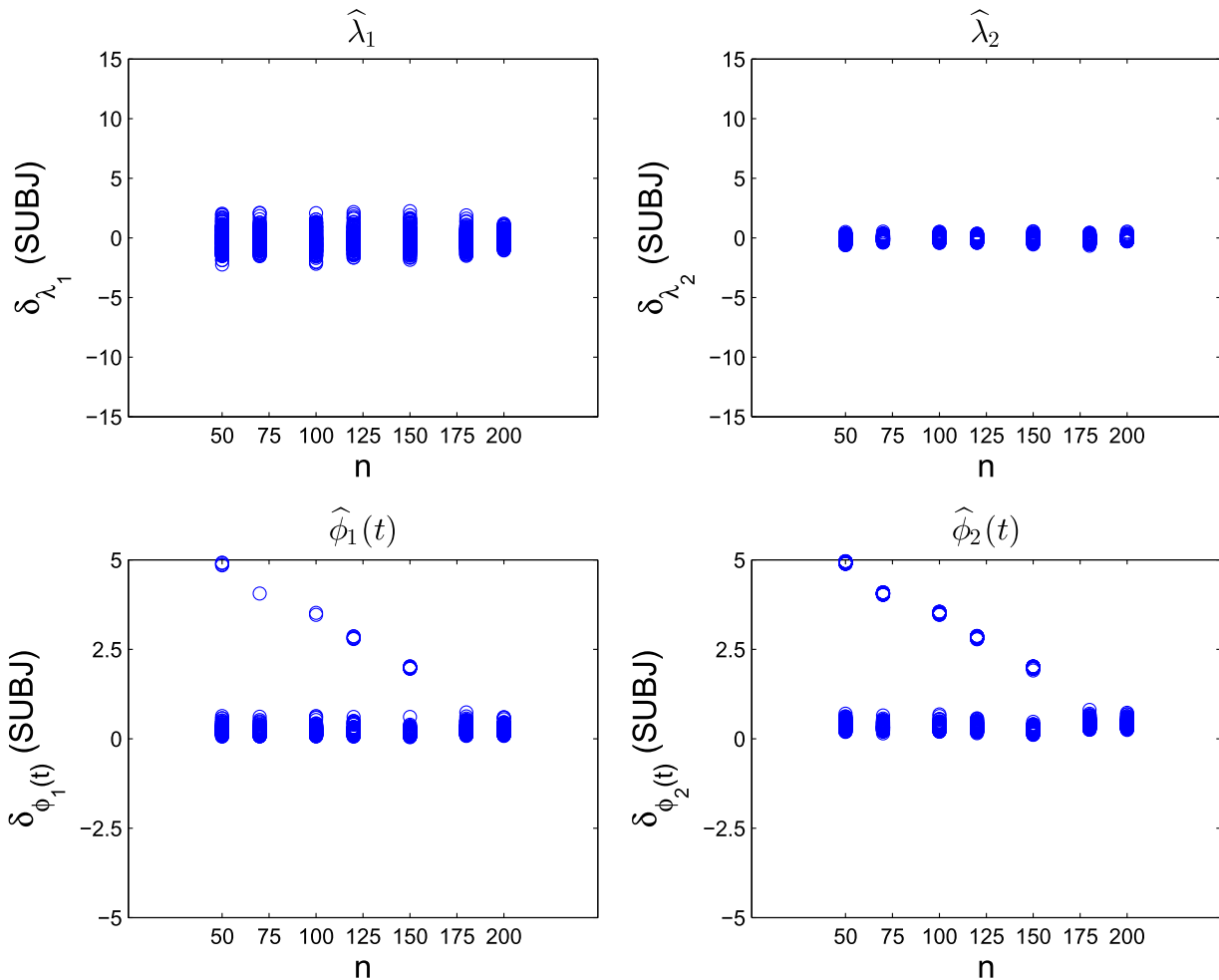
Figure 2. This is for the 'SUBJ' case. The top two panels present the values of δ_{λ_1} and δ_{λ_2} for 200 replications as n increases, respectively. The bottom two panels present the values of $\delta_{\phi_1(t)}$ and $\delta_{\phi_2(t)}$ for 200 replications as n increases, respectively.
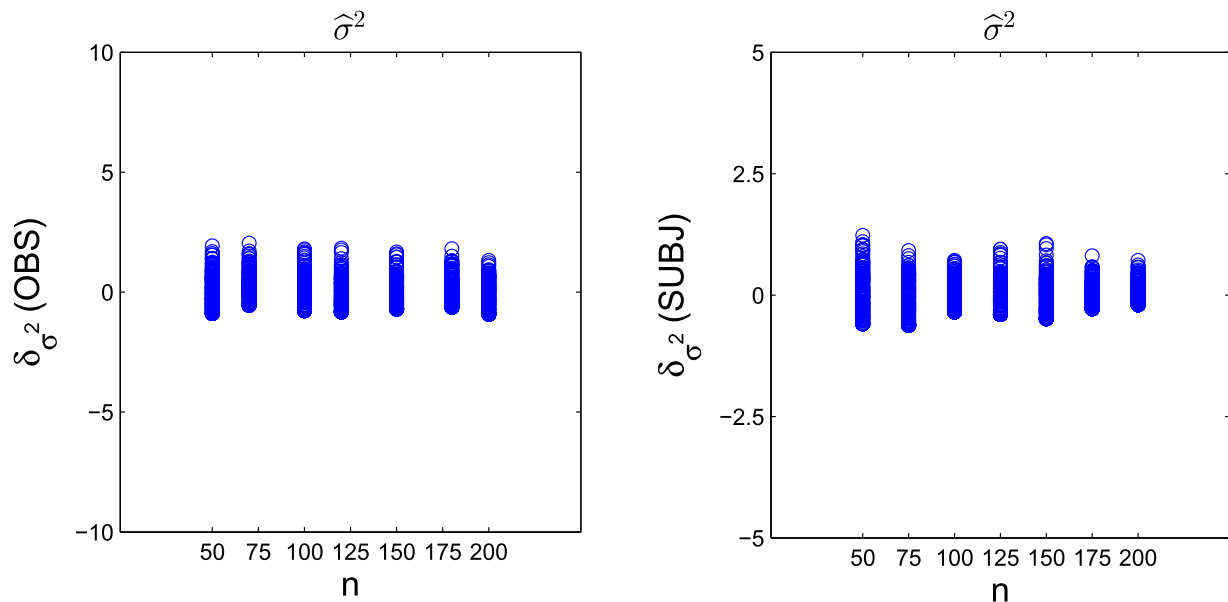
Figure 3. Plots of δ_{σ^2} for 'OBS' (the left panel) and 'SUBJ' (the right panel), respectively.

The procedure is visualised in Figures 1–3. The number of K in each replication was chosen by the 90% fraction of explained variance. As shown in Figure 1, the ranges (under 'OBS' scheme) for δ_{λ_k} (top two panels) and $\delta_{\phi_k(t)}$ (bottom two panels) for $k = 1, 2$ tend to be stable or go down, as the sample size n increases. This demonstrates that the derived order of magnitude of convergence rate in Corollary 3.2 is reasonable.

A similar phenomenon under the 'SUBJ' scheme can be observed in Figure 2, which shows the converging processes of the estimators $\widehat{\lambda}_k$ and $\widehat{\phi}_k$ for $k = 1$, 2. Although there are some abnormal values in the estimates of eigenfunctions (shown in the left bottom and right bottom panel), generally the fluctuation range tends to be stable. And as expected, the abnormal values disappear as the sample size increases. This shows that the derived convergence rates in Corollary 3.3 is reasonable.

Furthermore, the δ_{σ^2} under the 'OBS' and 'SUBJ; schemes are plotted in Figure 3. Observe that as the sample size increases, the range of variation of the 200 replications tends to be stable, which demonstrates the results of Corollary 3.5.

5. Discussion

It is common in functional data analysis literature that a method focuses on either dense or sparse data, while discussions about data of neither type are much less. Due to different behaviours of the two types of methods, one needs to choose properly the analysis method when dealing with real data, which is not necessarily as easy as it looks like. For example, we may face a mixture of densely and sparsely observed curves, or even it may be difficult to decide a sampling density. In this sense, methods that can handle any type of data are

appreciated. And the method we consider in this article belongs to this category. Specifically, we investigate the almost sure convergence of functional principal component analysis following Zhang and Wang (2016) and complement the unified theoretical framework they set up. We also note that the special case of Corollary 3.3 under SUBJ scheme is consistent with Theorem 3.6 in Li and hsing (2010). The convergence rate of Corollary 3.2 under OBS scheme is better than that of Corollary 1 in Yao et al. (2005), due to different techniques of proofs. We prove with asymptotic expansions of the eigenvalues and eigenfunctions of estimated covariance function (Hall and Hosseini-Nasab (2006) and Hall et al. (2006)) and strong uniform convergence rate of $\hat{\gamma}(s, t)$ by Lemma 5.1 in this article, which lead to a better convergence rate. It is also of great interest to establish the asymptotic distribution and optimal convergence rate of $\widehat{\phi}_j(t)$ under the general weighing framework, which we left for future work. Furthermore, the general weighing framework may be used in functional data regression, classification, clustering, etc., and hence the theoretical results here could be extended to those cases as well. This will also be pursued as future work.

Notes on contributors

Xingyu Yan is currently a Ph.D. candidate in the School of Statistics at East China Normal University. He is interested in functional data analysis.

Xiaolong Pu is currently Professor of statistics in the School of Statistics at East China Normal University. He is interested in applied statistics, particularly in statistical testing via sampling, statistical process control, sequential analysis and reliability. He received a Ph.D. in Statistics from East China Normal University.

Yingchun Zhou is currently Professor of statistics in the School of Statistics at East China Normal University. She is interested in functional data analysis and biostatistics. She received a Ph.D. in Statistics from Boston University and worked as a postdoc at National Institute of Statistical Sciences before moving to East China Normal University.

Xiaolei Xun is currently holding a visiting position in the School of Data Science at Fudan University. She is interested in functional data analysis, Bayesian methods, modeling of high-dimensional data and biostatistics. She received a Ph.D. in Statistics from Texas A&M University and worked in Novartis for five years as biometrician and statistical methodologist before moving to Fudan University.

References

Fan, J. Q., & Gijbels, I. (1996). *Local polynomial modelling and its applications*. London: Chapman and Hall.

Fan, J. Q., & Zhang, W. Y. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scandinavian Journal of Statistics*, 27, 715–731.

Ferraty, F., & Vieu, P. (2006). *Nonparametric functional data analysis: Theory and practice*. Berlin: Springer.

Greven, S., Crainiceanu, C. M., Caffo, B. S., & Reich, D. (2010). Longitudinal functional principal component analysis. *Electronic Journal of Statistics*, 4, 1022–1054.

Hall, P., & Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society, Series B*, 68, 109–126.

Hall, P., Müller, H.-G., & Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Annals of Statistics*, 34, 1493–1517.

James, G. M., Hastie, T. J., & Sugar, C. A. (2000). Principal component models for sparse functional data. *Biometrika*, 87, 587–602.

Jiang, C. R., & Wang, J. -L. (2010). Covariate adjusted functional principal components analysis for longitudinal data. *Annals of Statistics*, 38, 1194–1226.

Li, Y. H., & Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Annals of Statistics*, 38, 3321–3351.

Li, Y., Wang, N., & Carroll, R. J. (2013). Selecting the number of principal components in functional data. *Journal of the American Statistical Association*, 108, 1284–1294.

Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis*. New York: Springer.

Rice, J. A., & Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society, Series B*, 53, 233–243.

Staniswalis, J. G., & Lee, J. J. (1998). Nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association*, 93, 1403–1418.

Yao, F., & Lee, T. C. M. (2006). Penalized spline models for functional principal component analysis. *Journal of the Royal Statistical Society, Series B*, 68, 3–25.

Yao, F., Müller, H.-G., & Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100, 577–590.

Zhang, X. K., & Wang, J. -L. (2016). From sparse to dense functional data and beyond. *Annals of Statistics*, 44, 2281–2321.

Zhou, L., Lin, H. Z., & Liang, H (2017). Efficient estimation of the nonparametric mean and covariance functions for longitudinal and sparse functional data. *Journal of the American Statistical Association*.doi: 10.1080/01621459.2017.1356317.

Zhu, H. T., Li, R. Z., & Kong, L. L. (2012). Multivariate varying coefficient model for functional responses. *Annals of Statistics*, 40, 2634–2666.

Appendix

The following regularity conditions that are used to establish the asymptotic properties of the proposed estimators are imposed mainly for mathematical simplicity and may be modified as necessary. In the following, h_1, h_2 and h_3 are bandwidths used in estimating $\mu(t)$, $\gamma(s,t)$ and $V(t)$, respectively.

A.1 Regularity conditions

Kernel function

(A1) $K(\cdot)$ is a symmetric probability density function on $[-1, 1]$ and

$$\sigma_K^2 = \int u^2 K(u) du < \infty, \quad \|K\|^2 = \int K(u)^2 du < \infty.$$

(A2) $K(\cdot)$ is Lipschitz continuous: There exists $0 < L < \infty$ such that $|K(u) - K(v)| \leq L|u - v|$, for any $u, v \in [0, 1]$. This implies $K(\cdot) \leq M_K$ for a constant M_K.

Time points and true functions

(B1) $\{T_{ij} : i = 1, \ldots, n, j = 1, \ldots, N_i\}$, are *i.i.d.* copies of a random variable T defined on $[0,1]$. The density $f(\cdot)$ of T is bounded from below and above: $0 < m_f \leq \min_{t \in [0,1]} f(t) \leq \max_{t \in [0,1]} f(t) \leq M_f < \infty$. Furthermore, $f^{(2)}(\cdot)$, the second derivative of $f(\cdot)$, is bounded.

(B2) X is independent of T and ε is independent of T and U.

(B3) $\mu^{(2)}(t)$, the second derivative of $\mu(t)$, is bounded on $[0,1]$.

(B4) $\partial^2\gamma(s,t)/\partial s^2$, $\partial^2\gamma(s,t)/\partial s\partial t$ and $\partial^2\gamma(s,t)/\partial t^2$ are bounded on $[0,1]^2$.

Conditions for mean estimation

(C1) $h_1 \to 0$, $\log(n) \sum_{i=1}^n m_i\omega_i^2/h_1 \to 0$, $\log(n) \sum_{i=1}^n m_i(m_i - 1)\omega_i^2 \to 0$.

(C1') $h_3 \to 0$, $\log(n) \sum_{i=1}^n m_i\omega_i^2/h_3 \to 0$, $\log(n) \sum_{i=1}^n m_i(m_i - 1)\omega_i^2 \to 0$.

(C2) For some $\alpha > 2$, $E \sup_{t \in [0,1]} |X(t)|^\alpha < \infty$, $E|\varepsilon|^\alpha < \infty$ and

$$n\left[\sum_{i=1}^n m_i\omega_i^2 h_1 + \sum_{i=1}^n m_i(m_i - 1)\omega_i^2 h_1^2\right]\left[\frac{\log(n)}{n}\right]^{2/\alpha - 1} \to 0.$$

(C2') For some $\alpha > 2$, $E \sup_{t \in [0,1]} |X(t)|^\alpha < \infty$, $E|\varepsilon|^\alpha < \infty$ and

$$n\left[\sum_{i=1}^n m_i\omega_i^2 h_3 + \sum_{i=1}^n m_i(m_i - 1)\omega_i^2 h_3^2\right]\left[\frac{\log(n)}{n}\right]^{2/\alpha - 1} \to 0.$$

(C3) $\sup_n(n \max_i m_i\omega_i) \leq B < \infty$.

Conditions for covariance estimation

(D1)$h_2 \to 0$, $\log(n) \sum_{i=1}^{n} m_i(m_i - 1)v_i^2/h_2^2 \to 0$, $\log(n)$
$\sum_{i=1}^{n} m_i(m_i - 1)(m_i - 2)v_i^2/h_2 \to 0$, $\log(n) \sum_{i=1}^{n} m_i(m_i - 1)(m_i - 2)(m_i - 3)v_i^2 \to 0$.

(D2) For some $\beta > 2$, $E \sup_{t \in [0,1]} |X(t)|^{2\beta} < \infty$, $E|\varepsilon|^{2\beta} < \infty$
and

$$n\left[\sum_{i=1}^{n} m_i(m_i - 1)v_i^2 h_2^2 + \sum_{i=1}^{n} m_i(m_i - 1)(m_i - 2)v_i^2 h_2^3 \right.$$

$$\left. + \sum_{i=1}^{n} m_i(m_i - 1)(m_i - 2)(m_i - 3)v_i^2 h_2^4 \right]$$

$$\times \left[\frac{\log(n)}{n} \right]^{2/\beta - 1} \to \infty.$$

(D3) $\sup_n (n \max_i m_i(m_i - 1)v_i) \leq B' < \infty$.

The above conditions (A1)–(A2) and (B1)–(B4) are commonly used in the literature of the functional data and longitudinal data (see, e.g.. Fan Zhang (2000); Zhu, Li, Kong (2012)). Conditions (C1)–(C3) are used to guarantee the consistency of the mean estimators. Conditions (D1)–(D3) are used to guarantee the consistency of the covariance estimators. For the SUBJ estimators, (C3) and (D3) are automatically satisfied, similar versions of (C2) and (D2) were adopted by Li and Hsing (2010). In addition, (C1') and (C2') are used to guarantee the consistency of measurement error variance estimation.

A.2 Proof

To begin, let us give some notations that will be used in the sequel. Denote

$$S_r = \sum_{i=1}^{n} \omega_i \sum_{j=1}^{m_i} K_{h_1}(T_{ij} - t)\left(\frac{T_{ij} - t}{h_1} \right)^r,$$

$$R_r = \sum_{i=1}^{n} \omega_i \sum_{j=1}^{m_i} K_{h_1}(T_{ij} - t)\left(\frac{T_{ij} - t}{h_1} \right)^r Y_{ij},$$

$$Q_r = \sum_{i=1}^{n} \omega_i \sum_{j=1}^{m_i} K_{h_3}(T_{ij} - t)\left(\frac{T_{ij} - t}{h_3} \right)^r Y_{ij}^2,$$

where $r = 0, 1, 2$, and

$$S_{pq} = \sum_{i=1}^{n} v_i \sum_{1 \leq j \neq l \leq m_i} K_{h_2}(T_{ij} - s)K_{h_2}(T_{il} - t)$$

$$\times \left(\frac{T_{ij} - s}{h_2} \right)^p \left(\frac{T_{il} - t}{h_2} \right)^q$$

and

$$R_{pq} = \sum_{i=1}^{n} v_i \sum_{1 \leq j \neq l \leq m_i} K_{h_2}(T_{ij} - s)K_{h_2}(T_{il} - t)$$

$$\times \left(\frac{T_{ij} - s}{h_2} \right)^p \left(\frac{T_{il} - t}{h_2} \right)^q Y_{ij}Y_{il}$$

where for $p, q = 0, 1, 2$. For any univariate function $\varphi(\cdot) \in [0,1]$ and a bivariate function $\Psi(\cdot, \cdot) \in [0,1]^2$, define the L^2 norm by $\|\varphi\| = [\int \varphi(t)^2 dt]^{1/2}$ and the Hilbert–Schmidt norm by $\Psi = [\int \int \Psi(s,t)^2 ds dt]^{1/2}$. The domains of the integrals $[0,1]$ are omitted unless otherwise specified.

First, we present the convergence rate of $\widehat{\gamma}$ obtained in (1).

Lemma A.1: *Under the conditions in Appendix A.1, we have*

$$\sup_{s,t \in [0,1]} |\widehat{\gamma}(s,t) - \gamma(s,t)|$$

$$= O\left(h_1^2 + \left\{ \log(n) \left[\frac{\sum_{i=1}^{n} m_i \omega_i^2}{h_1} + \sum_{i=1}^{n} m_i(m_i - 1)\omega_i^2 \right] \right\}^{1/2} \right.$$

$$\times h_2^2 + \left\{ \log(n) \left[\frac{\sum_{i=1}^{n} m_i(m_i - 1)v_i^2}{h_2^2} \right. \right.$$

$$+ \frac{\sum_{i=1}^{n} m_i(m_i - 1)(m_i - 2)v_i^2}{h_2}$$

$$\left. \left. + \sum_{i=1}^{n} m_i(m_i - 1)(m_i - 2)(m_i - 3)v_i^2 \right\}^{1/2} \right) \quad a.s.$$

Proof: From (1), We note that

$$\sup_{s,t \in [0,1]} |\widehat{\gamma}(s,t) - \gamma(s,t)|$$

$$\leq |\widehat{G}(s,t) - G(s,t)| + \mu(s) \sup_{t \in [0,1]} |\widehat{\mu}(t) - \mu(t)|$$

$$+ \widehat{\mu}(t) \sup_{s \in [0,1]} |\widehat{\mu}(s) - \mu(s)|.$$

By (B3) and Theorem 5.1 in Zhang and Wang (2016), we have $\mu(s) \sup_{t \in [0,1]} |\widehat{\mu}(t) - \mu(t)|$, $\widehat{\mu}(t) \sup_{s \in [0,1]} |\widehat{\mu}(s) - \mu(s)| = O(\{\log(n)[\frac{\sum_{i=1}^{n} m_i \omega_i^2}{h_1} + \sum_{i=1}^{n} m_i(m_i - 1)\omega_i^2]\}^{1/2})$ *a.s.*. Thus the result can be derived from Theorem 5.2 in Zhang and Wang (2016). ∎

The following lemma is similar with Lemma 6 of Li and Hsing (2010) and will be used in our following proof repeatedly. Let Δ be the integral operator with kernel $\widehat{\gamma} - \gamma$.

Lemma A.2: *For any bounded measurable function ϕ on [0,1],*

$$\sup_{t \in [0,1]} |(\Delta \phi)(t)|$$

$$= O\left(h_1^2 + \left\{ \log(n) \left[\frac{\sum_{i=1}^{n} m_i \omega_i^2}{h_1} \right. \right. \right.$$

$$\left. \left. + \sum_{i=1}^{n} m_i(m_i - 1)\omega_i^2 \right] \right\}^{1/2}$$

$$+ h_2^2 + \left\{ \log(n) \left[\frac{\sum_{i=1}^{n} m_i \omega_i^2}{h_2} \right. \right.$$

$$\left. \left. + \sum_{i=1}^{n} m_i(m_i - 1)\omega_i^2 \right] \right\}^{1/2}$$

$$+ \log(n) \left[\frac{\sum_{i=1}^{n} m_i(m_i - 1)v_i^2}{h_2^2} \right.$$

$$+ \frac{\sum_{i=1}^{n} m_i(m_i - 1)(m_i - 2)v_i^2}{h_2}$$

$$\left. \left. + \sum_{i=1}^{n} m_i(m_i - 1)(m_i - 2)(m_i - 3)v_i^2 \right] \right) \quad a.s.$$

Proof: By (1), it follows that $(\Delta \phi)(t) = A_{n1} + A_{n2}$, where $A_{n1} = \int (\widehat{G} - G)(s,t)\phi(s)ds$ and $A_{n2} = \int \{\mu(s)\mu(t) - \widehat{\mu}(s)$

$\widehat{\mu}(t)\}\phi(s)ds$. With minor derivation, we have

$$A_{n1} = \int (\mathcal{A}_1 R_{00}^* - \mathcal{A}_2 R_{10}^* - \mathcal{A}_3 R_{01}^*)\mathcal{B}^{-1}\phi(s)ds,$$

where $\mathcal{A}_1 = S_{20}S_{02} - S_{11}^2$, $\mathcal{A}_2 = S_{10}S_{02} - S_{01}S_{11}$, $\mathcal{A}_3 = S_{01}S_{20} - S_{10}S_{11}$, $\mathcal{B} = \mathcal{A}_1 S_{00} - \mathcal{A}_2 S_{10} - \mathcal{A}_3 S_{01}$ and

$$R_{pq}^* = R_{pq} - G(s,t)S_{pq} - h_2 \partial G(s,t)/\partial s S_{p+1,q}$$
$$- h_2 \partial G(s,t)/\partial t S_{p,q+1} \text{ for } p,q = 0,1.$$

Further, by Taylor's expansion,

$$R_{pq}^* = \sum_{i=1}^n v_i \sum_{1 \le j \ne l \le m_i} \left(\frac{T_{ij} - s}{h_2}\right)^p$$
$$\times \left(\frac{T_{il} - t}{h_2}\right)^q K_{h_2}(T_{ij} - s)K_{h_2}(T_{il} - t)$$
$$\times [Y_{ij}Y_{il} - G(T_{ij}, T_{ij})] + O(h_2^2). \tag{A1}$$

Applying the proof of Theorem 5.2 in Zhang and Wang (2016), we obtain, uniformly in s, t,

$$R_{pq}^* = O(h_2^2 + b_n/h_2) \quad a.s. \tag{A2}$$

With further calculation,

$$\mathcal{A}_1 = f^2(s)f^2(t)(\sigma_K^2)^2 + O(h_2 + b_n/h_2) \quad a.s. \tag{A3}$$

and

$$\mathcal{B} = f^3(s)f^3(t)(\sigma_K^2)^2 + O(h_2 + b_n/h_2) \quad a.s., \tag{A4}$$

where $b_n = \{\log(n)[\sum_{i=1}^n m_i(m_i - 1)v_i^2 h_2^2 + \sum_{i=1}^n m_i(m_i - 1)(m_i - 2)v_i^2 h_2^3 + \sum_{i=1}^n m_i(m_i - 1)(m_i - 2)(m_i - 3)v_i^2 h_2^4]\}^{1/2}$.

We focus on $\int \mathcal{A}_1 \mathcal{R}_{00}^* \mathcal{B}^{-1}\phi(s)ds$ since the other two terms can be dealt with similarly. Specifically,

$$\int \mathcal{A}_1 \mathcal{R}_{00}^* \mathcal{B}^{-1}\phi(s)ds$$

$$= \frac{1}{f(t)}\sum_{i=1}^n \omega_i \sum_{1 \le j \ne l \le m_i} \{Y_{ij}Y_{il} - G(T_{ij}, T_{il})\}K_{h_2}(T_{il} - t)$$

$$\times \int K_{h_2}(T_{ij} - s)\phi(s)f(s)^{-1}ds$$

$$+ O\left(h_2^2 + \log(n)\left[\frac{\sum_{i=1}^n m_i(m_i - 1)v_i^2}{h_2^2}\right.\right.$$

$$+ \frac{\sum_{i=1}^n m_i(m_i - 1)(m_i - 2)v_i^2}{h_2}$$

$$\left.\left.+ \sum_{i=1}^n m_i(m_i - 1)(m_i - 2)(m_i - 3)v_i^2\right]\right).$$

Note that

$$\left|\int_0^1 K_{h_2}(T_{ij} - s)\phi(s)f(s)^{-1}ds\right|$$

$$\le \sup_{s \in [0,1]} (|\phi(s)|f(s)^{-1}) \int_{-1}^1 K(u)du.$$

Similarly to the proof of Lemma 5 in Zhang and Wang (2016), we can prove the following almost sure uniform rate:

$$\frac{1}{f(t)}\sum_{i=1}^n \omega_i \sum_{1 \le j \ne l \le m_i} \{Y_{ij}Y_{il} - G(T_{ij}, T_{il})\}K_{h_2}(T_{il} - t)$$

$$\times \int K_{h_2}(T_{ij} - s)\phi(s)f(s)^{-1}ds$$

$$= O\left(\left\{\log(n)\left[\frac{\sum_{i=1}^n m_i\omega_i^2}{h_2}\right.\right.\right.$$

$$\left.\left.\left.+ \sum_{i=1}^n m_i(m_i - 1)\omega_i^2\right]\right\}^{1/2}\right) \quad a.s.$$

Thus

$$\int \mathcal{A}_1 \mathcal{R}_{00}^* \mathcal{B}^{-1}\phi(s)ds$$

$$= O\left(\left\{\log(n)\left[\frac{\sum_{i=1}^n m_i\omega_i^2}{h_2} + \sum_{i=1}^n m_i(m_i - 1)\omega_i^2\right]\right\}^{1/2}\right.$$

$$+ h_2^2 + \log(n)\left[\frac{\sum_{i=1}^n m_i(m_i - 1)v_i^2}{h_2^2}\right.$$

$$+ \frac{\sum_{i=1}^n m_i(m_i - 1)(m_i - 2)v_i^2}{h_2}$$

$$\left.\left.+ \sum_{i=1}^n m_i(m_i - 1)(m_i - 2)(m_i - 3)v_i^2\right]\right) \quad a.s.$$

The term A_{n2} can be written as

$$A_{n2} = \int \{\mu(t) - \widehat{\mu}(t)\}\int \mu(s)\phi(s)ds$$
$$+ \widehat{\mu}(t)\int \{\mu(s) - \widehat{\mu}(s)\}\phi(s)ds.$$

Following Theorem 5.1 in Zhang and Wang (2016), it is easy to see that

$$A_{n2} = O\left(h_1^2 + \left\{\log(n)\left[\frac{\sum_{i=1}^n m_i\omega_i^2}{h_1}\right.\right.\right.$$

$$\left.\left.\left.+ \sum_{i=1}^n m_i(m_i - 1)\omega_i^2\right]\right\}^{1/2}\right) \quad a.s.$$

Hence, the lemma follows. ∎

Proof of Theorem 3.1: Following Hall and Hosseini-Nasab (2006) and Bessel's inequality, we have

$$\|\widehat{\phi}_j - \phi_j\| \le C(\|\Delta\phi_j\| + \|\Delta\|^2),$$

where $\|\Delta\phi_j\| = (\int\int\{\widehat{\gamma}(s,t) - \gamma(s,t)\phi_j(s)\}^2 dsdt)^{1/2}$ and $\|\Delta\| = (\int\int\{\widehat{\gamma}(s,t) - \gamma(s,t)\}^2 dsdt)^{1/2}$. Lemma A.1 and Lemma A.2 lead to

$$\|\widehat{\phi}_j - \phi_j\| = O\left(h_1^2 + \left\{\log(n)\left[\frac{\sum_{i=1}^n m_i\omega_i^2}{h_1}\right.\right.\right.$$

$$\left.\left.+ \sum_{i=1}^n m_i(m_i - 1)\omega_i^2\right]\right\}^{1/2}$$

$$+ h_2^2 + \left\{\log(n)\left[\frac{\sum_{i=1}^n m_i\omega_i^2}{h_2}\right.\right.$$

$$\left.\left.\left.+ \sum_{i=1}^n m_i(m_i - 1)\omega_i^2\right]\right\}^{1/2}$$

$$+ \log(n) \left[\frac{\sum_{i=1}^n m_i(m_i - 1)v_i^2}{h_2^2} \right.$$

$$+ \frac{\sum_{i=1}^n m_i(m_i - 1)(m_i - 2)v_i^2}{h_2}$$

$$\left. \left. + \sum_{i=1}^n m_i(m_i - 1)(m_i - 2)(m_i - 3)v_i^2 \right] \right) \quad a.s.$$

By (4.9) in Hall et al. (2006), we have

$$\widehat{\lambda}_j - \lambda_j = B_{n1} + B_{n2} + O(\|\Delta\phi_j\|^2),$$

where

$$B_{n1} = \int \int (\mathcal{A}_1 R_{00}^* - \mathcal{A}_2 R_{10}^* - \mathcal{A}_3 R_{01}^*) \mathcal{B}^{-1} \phi_j(s)\phi_j(t) \mathrm{d}s \mathrm{d}t$$

and

$$B_{n2} = \int \{\mu(s) - \widehat{\mu}(s)\}\phi_j(s)\mathrm{d}s \int \widehat{\mu}(t)\phi_j(t)\mathrm{d}t$$

$$+ \int \mu(s)\phi_j(s)\mathrm{d}s \int \{\mu(t) - \widehat{\mu}(t)\}\phi_j(t)\mathrm{d}t.$$

For B_{n1}, again it suffices to focus on $\int \int \mathcal{A}_1 R_{00}^* \mathcal{B}^{-1} \phi_j(s)\phi_j(t)$ $\mathrm{d}s\mathrm{d}t$. First, we have

$$\int \int \mathcal{A}_1 R_{00}^* \mathcal{B}^{-1} \phi_j(s)\phi_j(t)\mathrm{d}s\mathrm{d}t$$

$$= \sum_{i=1}^n v_i \sum_{1 \le j \le l \le m_i} \{Y_{ij}Y_{il} - G(T_{ij}, T_{il})\}$$

$$\times \int \int K_{h_1}(T_{ij} - s)K_{h_2}(T_{jl} - t)$$

$$\times \phi_j(s)\phi_j(t)\{f(s)f(t)\}^{-1}\mathrm{d}s\mathrm{d}t$$

$$+ O\left(h_2^2 + \log(n) \left[\frac{\sum_{i=1}^n m_i(m_i - 1)v_i^2}{h_2^2} \right. \right.$$

$$+ \frac{\sum_{i=1}^n m_i(m_i - 1)(m_i - 2)v_i^2}{h_2}$$

$$\left. \left. + \sum_{i=1}^n m_i(m_i - 1)(m_i - 2)(m_i - 3)v_i^2 \right] \right) \quad a.s.$$

By Lemma 5 in Li and Hsing (2010), we have

$$B_{n1} = O\left((\log(n)/n)^{1/2} + h_2^2 \right.$$

$$+ \log(n) \left[\frac{\sum_{i=1}^n m_i(m_i - 1)v_i^2}{h_2^2} \right.$$

$$+ \frac{\sum_{i=1}^n m_i(m_i - 1)(m_i - 2)v_i^2}{h_2}$$

$$\left. \left. + \sum_{i=1}^n m_i(m_i - 1)(m_i - 2)(m_i - 3)v_i^2 \right] \right) \quad a.s.$$

Following Theorem 5.1 in Zhang and Wang (2016), it can be shown similarly that

$$B_{n2} = O\left((\log(n)/n)^{1/2} + h_1^2 + \log(n) \left[\frac{\sum_{i=1}^n m_i\omega_i^2}{h_1} \right. \right.$$

$$\left. \left. + \sum_{i=1}^n m_i(m_i - 1)\omega_i^2 \right] \right) \quad a.s.$$

This completes the proof for assertion (a). For assertion (b), we have, for any $t \in [0, 1]$,

$$\lambda_j|\widehat{\phi}_j(t) - \phi_j(t)|$$

$$= |\widehat{\lambda}_j\widehat{\phi}_j(t) - \lambda_j\phi_j(t) - (\widehat{\lambda}_j - \lambda_j)\widehat{\phi}_j(t)|$$

$$\le \int \{\widehat{\gamma}(s, t) - \gamma(s, t)\}\phi_j(t)\mathrm{d}s$$

$$+ \int \widehat{\gamma}(t, t)\{\widehat{\phi}_j(s) - \phi_j(s)\}\mathrm{d}s + |\widehat{\lambda}_j - \lambda_j||\widehat{\phi}_j(t)|$$

$$\le \sup_{t \in [0,1]} |(\Delta\phi)(t)| + O(\|\widehat{\phi}_j - \phi_j\|)$$

$$+ |\widehat{\lambda}_j - \lambda_j| \sup_{t \in [0,1]} |\widehat{\phi}_j(t)|,$$

where the last inequality is established by Cauchy-Schwarz inequality. Based on Lemma 6 and assertion (a), assertion (b) holds. ∎

Proof of Theorem 3.4: By rearranging the terms, we have

$$\widehat{\sigma}^2 - \sigma^2 = \int \{\widehat{V}(t) - V(t)\}\mathrm{d}t - \int \{\widehat{G}(t) - G(t)\}\mathrm{d}t.$$

First, we consider $\widehat{V}(t) - V(t)$. Similarly to (C.1) in Zhang and Wang (2016), we have

$$\widehat{V}(t) - V(t) = \frac{Q_0^* S_2 - Q_1^* S_1}{S_0 S_2 - S_1^2}, \qquad (A5)$$

where $Q_q^* = Q_q - V(t)S_q - hV^{(1)}S_{q+1}$ for $q = 0, 1$. By (D.1) in Zhang and Wang (2016), Q_q^* has the uniformly rate $O(h_3^2 + a_n/h_3)$ a.s. and we can derive that $S_2 = f(t)\sigma_K^2 + O(h_3 + a_n/h_3)$, where $a_n = \{\log(n)[\sum_{i=1}^n m_i\omega_i^2 h_3 + \sum_{i=1}^n m_i(m_i - 1)\omega_i^2 h_3^2]\}^{1/2}$. One can see that

$$\widehat{V}(t) - V(t) = \frac{1}{f(t)} \sum_i \omega_i \sum_{j=1}^{m_i} K_{h_3}(T_{ij} - t)\{Y_{ij}^2 - V(T_{ij})\}$$

$$+ O(h_3^2 + (a_n/h_3)^2) \quad a.s.$$

Thus

$$\int_0^1 \widehat{V}(t) - V(t)\mathrm{d}t$$

$$= \sum_i \omega_i \sum_{j=1}^{m_i} \{Y_{ij}^2 - V(T_{ij})\} \int_0^1 K_{h_3}(T_{ij} - t)f^{-1}(t)\mathrm{d}t$$

$$+ O(h_3^2 + (a_n/h_3)^2) \quad a.s.$$

We apply (A5) but will focus on the leading term $\frac{Q_0^* S_2}{S_0 S_2 - S_1^2}$, since the other term is of lower order and can be dealt with similarly. Note that

$$\left| \int_0^1 K_{h_3}(T_{ij} - t)f^{-1}(t)\mathrm{d}t \right| \le \sup_t f^{-1}(t),$$

and by Lemma 5 in Li and Hsing (2010), we have

$$\int_0^1 \widehat{V}(t) - V(t)\mathrm{d}t = O((\log(n)/n)^{1/2} + h_3^2 + (a_n/h_3)^2).$$

(A6)

Next, to consider $\widehat{G}(t) - G(t)$ we follow the similar expression (C.2) in Zhang and Wang (2016). Again we focus on $R_{00}^* \mathcal{A}_1 \mathcal{B}^{-1}$. Applying (A2), we obtain, uniformly in s, t,

$$R_{00}^* = O(h_2^2 + b_n/h_2)$$

and

$$\mathcal{A}_1 \mathcal{B}^{-1} = [f(s)f(t)]^{-1} + O(h_2 + b_n/h_2),$$

where $b_n = \{\log(n)[\sum_{i=1}^n m_i(m_i - 1)v_i^2 h_2^2 + \sum_{i=1}^n m_i(m_i - 1)(m_i - 2)v_i^2 h_2^3 + \sum_{i=1}^n m_i(m_i - 1)(m_i - 2)(m_i - 3)v_i^2 h_2^4]\}^{1/2}$. Similarly to the proof of Theorem 3.4 in Li and Hsing (2010), it can be shown that

$$\int_0^1 \{\widehat{G}(t) - G(t)\}\mathrm{d}t = O(h_2^2 + a_n/h_2 + (b_n/h_2)^2) \quad a.s. \tag{A7}$$

The theorem follows from (A6) and (A7). ∎