Available online at www.sciencedirect.com
ScienceDirect

Journal of Multivariate Analysis 98 (2007) 40–56

Journal of Multivariate Analysis

www.elsevier.com/locate/jmva

Asymptotic distributions of nonparametric regression estimators for longitudinal or functional data

Fang Yao

Department of Statistics, Colorado State University, Fort Collins, CO 80523, USA

Received 28 January 2005
Available online 28 September 2006

Abstract

The estimation of a regression function by kernel method for longitudinal or functional data is considered. In the context of longitudinal data analysis, a random function typically represents a subject that is often observed at a small number of time points, while in the studies of functional data the random realization is usually measured on a dense grid. However, essentially the same methods can be applied to both sampling plans, as well as in a number of settings lying between them. In this paper general results are derived for the asymptotic distributions of real-valued functions with arguments which are functionals formed by weighted averages of longitudinal or functional data. Asymptotic distributions for the estimators of the mean and covariance functions obtained from noisy observations with the presence of within-subject correlation are studied. These asymptotic normality results are comparable to those standard rates obtained from independent data, which is illustrated in a simulation study. Besides, this paper discusses the conditions associated with sampling plans, which are required for the validity of local properties of kernel-based estimators for longitudinal or functional data.

AMS 1991 subject classification: 62G08

Keywords: Asymptotic distribution; Covariance; Functional data; Longitudinal data; Regression; Within-subject correlation

1. Introduction

Modern technology and advanced computing environments have facilitated the collection and analysis of high-dimensional data, or data that are repeatedly measured for a sample of subjects. The repeated measurements are often recorded over a period of time, say on an closed and bounded interval $\mathcal{T}$. It also could be a spacial variable, such as in image or geoscience applications.

E-mail addresses: fyao@utstat.toronto.edu, fyao@stat.colostate.edu.

0047-259X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2006.08.007

When the data are recorded densely over time, often by machine, they are typically termed functional or curve data with one observed curve or function per subject, while in longitudinal studies the repeated measurements usually take place on a few scattered observational time points for each subject. A significant intrinsic difference between two settings lies in the perception that functional data are observed in the continuum without noise [2,3], whereas longitudinal data are observed at sparsely distributed time points and are often subject to experimental error [4]. However, in practice functional data are analyzed after smoothing noisy observations [10], which indicates that the differences between two data types related to the way in which a problem is perceived are arguably more conceptual than actual. Therefore in this paper, kernel-based regression estimators obtained from observations at discrete time points contaminated with measurement errors, rather than observations in the continuum, are considered for these realistic reasons. In the context of kernel-based nonparametric regression, the effects of sampling plans on the statistical estimators are also investigated.

A vast literature has been developed in the past decade on the kernel-based regression for independent and identically distributed data, for summary, see Fan and Gijbels [5]. There has been substantial recent interest in extending the existing asymptotic results to functional or longitudinal data [8,11,14,13,9]. The issues caused by the within-subject correlation are rigorously addressed in this paper. Hart and Wehrly [8] studied the Gasser–Müller estimator of the mean function for repeated measurements observed on a regular grid by assuming stationary correlation structure, and showed that the influence of the within-subject correlation on the asymptotic variance is of smaller order compared to the standard rate obtained from independent data and will disappear when the correlation function is differentiable at zero. Our asymptotic distribution result is in fact consistent with that in Hart and Wehrly [8] and applicable for general covariance structure without stationary assumption. This problem was also discussed by Staniswalis and Lee [12] and Lin and Carroll [9], where they used the heuristic arguments of the local property of local polynomial estimation and intuitively ignored the within-subject correlation while deriving the asymptotic variances. This paper derives appropriate conditions that are required for the validity of the local property of kernel type estimators obtained from longitudinal or functional data. These conditions also provide practical guidelines for various sampling procedures.

The contribution of this paper is the derivation of general asymptotic distribution results in both one-dimensional and two-dimensional smoothing context for real-valued functions with arguments which are functionals formed by weighted averages of longitudinal or functional data. These asymptotic normality results are comparable to those obtained for identically distributed and independent data. These results are applied to the kernel-based estimators of the mean and covariance functions, which yields asymptotic normal distributions of these estimators. In particular, to the best of our knowledge, no asymptotic distribution results are available up to date for nonparametric estimation of covariance functions obtained from longitudinal or functional data contaminated with measurement error. By comparison, Hall et al. [6,7] investigated asymptotic properties of nonparametric kernel estimators of autocovariance, where the measurements were only observed from a single stationary stochastic process or random field. Although the asymptotic distributions are derived for random design in this paper, the arguments can be extended to fixed design and other sampling plans with appropriate modifications, and asymptotic bias and variance terms can also be obtained in similar manner. This will provide theoretical basis and practical guidance for the nonparametric analysis of functional or longitudinal data with important potential applications which are based on the asymptotic distributions. Typical examples include the construction of asymptotic confidence bands for regression functions and confidence regions for covariance surfaces, and also fast selection of bandwidth for covariance surface estimation based on asymptotic mean squared error. Other applications in the context of smoothing independent data can be explored for the smoothing of longitudinal or functional data using kernel-based estimators.

The remainder of the paper is organized as follow. In Section 2 we derive the general asymptotic distributions of one- and two-dimensional smoothers obtained from longitudinal or functional data for random design. These general asymptotic results are applied to commonly used kernel-type estimators of the mean curve and covariance surface in Section 3. Extension to fixed design is discussed in Section 4. A simulation study is presented to evaluate the derived asymptotic results for correlated data in Section 5, while discussions, including potential applications of the resulting asymptotic normality, are offered in Section 6.

2. General results of asymptotic distributions for random design

In this section we will define general functionals that are kernel-weighted averages of the data for one-dimensional and two-dimensional smoothing. The introduced general functionals include the most commonly used types of kernel-based estimators as special cases, such as Gasser–Müller estimator, Nadaraya–Waston estimator, local polynomial estimator, etc. Since Nadaraya–Waston and local polynomial estimators are mostly used in practice, their asymptotic behaviors in terms of bias and variance for independent data have been thoroughly studied in existing literature. However, for longitudinal or functional data, particularly in regard to covariance surface estimators, the asymptotic behaviors of bias and variance of these two commonly used estimators are still largely unknown. Therefore in Section 3, the general asymptotic results developed in this section are applied to Nadaraya–Waston and local polynomial estimators in both one-dimensional and two-dimensional smoothing settings. In particular, the lack of asymptotic results for the covariance surface estimators of longitudinal or functional data is an additional motivation for the definition of the two-dimensional general functional that can be applied to develop the asymptotic distributions for these estimators.

We first consider random design while extension to other sampling plans is deferred to Section 4. In classical longitudinal studies, measurements are often intended to be on a regular time grid. However, since individuals may miss scheduled visits, the resulting data usually become sparse, where only few observations are obtained for most subjects, with unequal numbers of repeated measurements per subject and different measurement times T_{ij} per individual. This sampling plan motivates the following assumptions which are applicable for a number of longitudinal data encountered in practice.

Let $X = \{X(t), t \in [0, \mathcal{T}]\}$, $\mathcal{T} < \infty$, be a continuous-time stochastic process defined on a probability space $(\Omega, \mathcal{A}, P)$ with finite variance, $E(\int_0^{\mathcal{T}} X^2(t)\, dt) < \infty$, and X_i be independently and identically distributed (i.i.d.) realizations of X which can be modelled as follows in general. The trajectories X_i have the mean function $\mu(t)$ and uncorrelated random coefficients ξ_{ik} with mean zero, variances λ_k satisfying $\sum_{k=1}^{\infty} \lambda_k < \infty$ and corresponding basis functions $\phi_k(t)$, where the within-subject covariance function is determined as $C(s, t) = cov(X_i(s), X_i(t)) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s)\phi_k(t)$. A typical example is the Karhunen–Loève representation of stochastic processes, where λ_k and ϕ_k correspond to eigenvalues and eigenfunctions. Suppose that one has data $\{(T_{ij}, Y_{ij}), 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant N_i\}$, where Y_{ij} are observations drawn from the process X_i at time T_{ij}, incorporating the uncorrelated measurement errors ε_{ij} that have mean zero and a constant variance σ^2,

$$Y_{ij} = X_i(T_{ij}) + \varepsilon_{ij} = \mu(T_{ij}) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k(T_{ij}) + \varepsilon_{ij}, \quad T_{ij} \in \mathcal{T}, \tag{1}$$

where $E\varepsilon_{ij} = 0$, $var(\varepsilon_{ij}) = \sigma^2$, and the number of observations, $N_i(n)$ depending on the sample size n, are considered random. We make the following assumptions,

(A1.1) The number of observations $N_i(n)$ made for the ith subject or cluster, $i = 1, \ldots, n$, is a r.v. with $N_i(n) \overset{\text{i.i.d}}{\sim} N(n)$, where $N(n) > 0$ is a positive integer-valued random variable with $\limsup_{n\to\infty} EN(n)^2/[EN(n)]^2 < \infty$ and $\limsup_{n\to\infty} EN(n)^4/EN(n)EN(n)^3 < \infty$.

In the sequel the dependence of $N_i(n)$ and $N(n)$ on the sample size n is suppressed for simplicity; i.e., $N_i = N_i(n)$ and $N(n) = N$. The observation times and measurements are assumed to be independent of the number of measurements, i.e., for any subset $J_i \subseteq \{1, \ldots, N_i\}$ and for all $i = 1, \ldots, n$,

(A1.2) $(\{T_{ij} : j \in J_i\}, \{Y_{ij} : j \in J_i\})$ is independent of N_i.

Writing $\boldsymbol{T}_i = (T_{i1}, \ldots, T_{iN_i})^T$ and $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{iN_i})^T$, it is easy to see that the triples $\{\boldsymbol{T}_i, \boldsymbol{Y}_i, N_i\}$ are i.i.d..

2.1. Asymptotic normality of one-dimensional smoother

To assume appropriate regularity conditions that are used to derive asymptotic properties, we define a new type of continuity that differs from those which are commonly used. We say that a real function $f(x, y) : \Re^{p+q} \to \Re$ is continuous on $x \in A \subseteq \Re^p$ uniformly in $y \in \Re^q$, provided that for any $x \in A$ and $\varepsilon > 0$, there exists a neighborhood of x not depending on y, saying $U(x) \subseteq \Re^p$, such that $|f(x', y) - f(x, y)| < \varepsilon$ for all $x' \in U(x)$ and $y \in \Re^q$.

For random design, (T_{ij}, Y_{ij}) are assumed to have the identical distribution as (T, Y) with joint density $g(t, y)$. Assume that the observation times T_{ij} are i.i.d. with the marginal density $f(t)$, but dependence is allowed among Y_{ij} and Y_{ik} that are observations made for the same subject or cluster. Also denote the joint density of (T_j, T_k, Y_j, Y_k) by $g_2(t_1, t_2, y_1, y_2)$, where $j \neq k$. Let v, k be given integers, with $0 \leqslant v < k$. We assume regularity conditions for the marginal and joint densities, $f(t)$, $g(t, y)$, $g_2(t_1, t_2, y_1, y_2)$ and the mean function of the underlying process $X(t)$, i.e., $E[X(t)] = \mu(t)$, with respect to a neighborhood of a interior point $t \in \mathcal{T}$, assuming that there exists a neighborhood $U(t)$ of t such that:

(B1.1) $\frac{d^k}{du^k} f(u)$ exists and is continuous on $u \in U(t)$, and $f(u) > 0$ for $u \in U(t)$;

(B1.2) $g(u, y)$ is continuous on $u \in U(t)$ uniformly in $y \in \Re$; $\frac{d^k}{du^k} g(u, y)$ exists and is continuous on $u \in U(t)$ uniformly in $y \in \Re$;

(B1.3) $g_2(u, v, y_1, y_2)$ is continuous on $(u, v) \in U(t)^2$ uniformly in $(y_1, y_2) \in \Re^2$;

(B1.4) $\frac{d^k}{du^k} \mu(u)$ exists and is continuous on $u \in U(t)$.

Let $K_1(\cdot)$ be nonnegative univariate kernel functions in one-dimensional smoothing. The assumptions for kernels $K_1 : \Re \to \Re$ are as follows. We say that a univariate kernel function K_1 is of order (v, k), if

$$\int u^\ell K_1(u)\, du = \begin{cases} 0, & 0 \leqslant \ell < k, \quad \ell \neq v, \\ (-1)^v v!, & \ell = v, \\ \neq 0, & \ell = k, \end{cases} \tag{2}$$

(B2.1) K_1 is compactly supported, $\|K_1\|^2 = \int K_1^2(u)\,du < \infty$;
(B2.2) K_1 is a kernel function of order (v, ℓ).

Let $b = b(n)$ be a sequences of bandwidths that are used in one-dimensional smoothing. We develop asymptotics as $n \to \infty$, and require
(B3) $b \to 0$, $n(EN)b^{v+1} \to \infty$, $b(EN) \to 0$, and $n(EN)b^{2k+1} \to d^2$ for some d with $0 \leqslant d < \infty$.
One could see in the proof of Theorem 1 that the assumptions (B3) combined with (A1.1) provide the condition such that the local property of kernel-type estimators holds for longitudinal or functional data with the presence of within-subject correlation.

Let $\{\psi_\lambda\}_{\lambda=1,\dots,l}$ be a collection of real functions $\psi_\lambda : \Re^2 \to \Re$, which satisfy:

(B4.1) $\psi_\lambda(t, y)$ are continuous on $\{t\}$ uniformly in $y \in \Re$;
(B4.2) $\frac{d^k}{dt^k}\psi_\lambda(t, y)$ exists for all arguments (t, y) and are continuous on $\{t\}$ uniformly in $y \in \Re$.

Then we define the general weighted averages

$$\Psi_{\lambda n} = \frac{1}{nENb^{v+1}} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \psi_\lambda(T_{ij}, Y_{ij}) K_1\left(\frac{t - T_{ij}}{b}\right), \quad \lambda = 1, \dots, l.$$

and

$$\mu_\lambda = \mu_\lambda(t) = \frac{d^v}{dt^v} \int \psi_\lambda(t, y) g(t, y)\,dy, \quad \lambda = 1, \dots, l.$$

Let

$$\sigma_{\kappa\lambda} = \sigma_{\kappa\lambda}(t) = \int \psi_\kappa(t, y)\psi_\lambda(t, y) g(t, y)\,dy \|K_1\|^2, \quad 1 \leqslant \lambda, \kappa \leqslant l,$$

and $H : \Re^l \to \Re$ be a function with continuous first order derivatives. We denote the gradient vector $((\partial H/\partial x_1)(v), \dots, (\partial H/\partial x_l)(v))^T$ by $DH(v)$ and $\bar{N} = \sum_{i=1}^n N_i/n$.

Theorem 1. *If the assumptions* (A1.1), (A1.2) *and* (B1.1)–(B4.2) *hold*, *then*

$$\sqrt{n\bar{N}b^{2v+1}}[H(\Psi_{1n}, \dots, \Psi_{ln}) - H(\mu_1, \dots, \mu_l)] \xrightarrow{\mathcal{D}} \mathcal{N}(\beta, [DH(\mu_1, \dots, \mu_l)]^T \Sigma[DH(\mu_1, \dots, \mu_l)]), \tag{3}$$

where

$$\beta = \frac{(-1)^k d}{k!} \int u^k K_1(u)\,du \sum_{\lambda=1}^{l} \frac{\partial H}{\partial \mu_\lambda}\{(\mu_1, \dots, \mu_l)^T\} \frac{d^{k-v}}{dt^{k-v}}\mu_\lambda(t), \quad \Sigma = (\sigma_{\kappa\lambda})_{1 \leqslant \kappa, \lambda \leqslant l}.$$

Proof. It is seen that $\bar{N}$ can be replaced with *EN* by Slutsky Theorem under (A1.1). We now show that

$$\sqrt{n(EN)b^{2v+1}}[H(E\Psi_{1n}, \dots, E\Psi_{ln}) - H(\mu_1, \dots, \mu_l)] \longrightarrow \beta. \tag{4}$$

Since (A1.1) and (A1.2) hold, and K_1 is of order (v, k), using Taylor expansion to order k, one obtains

$$E\Psi_{\lambda n} = \frac{1}{nb^{v+1}} E\left\{\sum_{i=1}^{n} \frac{1}{EN} \sum_{j=1}^{N_i} \psi_\lambda(T_{ij}, Y_{ij}) K_1\left(\frac{t-T_{ij}}{b}\right)\right\}$$

$$= \frac{1}{b^{v+1}EN} E\left\{\sum_{j=1}^{N} E\left[\psi_\lambda(T_j, Y_j) K_1\left(\frac{t-T_j}{b}\right)\middle| N\right]\right\}$$

$$= \frac{1}{b^{v+1}} E\left\{\psi_\lambda(T, Y) K_1\left(\frac{t-T}{b}\right)\right\}$$

$$= \mu_\lambda + \frac{(-1)^k}{k!} \int u^k K_1(u)\, du \, \frac{d^{k-v}}{dt^{k-v}} \mu_\lambda(t) b^{k-v} + o(b^{k-v}). \tag{5}$$

Then (4) follows from an l-dimensional Taylor expansion of H of order 1 around $(\mu_1, \ldots, \mu_l)^T$, coupled with (5). If we can show

$$\sqrt{n(EN)b^{2v+1}}[(\Psi_{1n}, \ldots, \Psi_{ln})^T - (E\Psi, \ldots, E\Psi_{ln})^T] \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma), \tag{6}$$

in analogy to Bhattacharya and Müller [1], and continuity of DH at $(\mu_1, \ldots, \mu_l)^T$ and applying similar arguments used in (5), we find $DH(E\Psi_{1n}, \ldots, E\Psi_{ln}) \to DH(\mu_1, \ldots, \mu_l)$. Then Carmér–Wold device yields

$$\sqrt{n(EN)b^{2v+1}}[H(\Psi_{1n}, \ldots, \Psi_{ln}) - H(E\Psi, \ldots, E\Psi_{ln})] \xrightarrow{\mathcal{D}} \mathcal{N}(0, DH(\mu_1, \ldots, \mu_l)^T$$

$$\Sigma DH(\mu_1, \ldots, \mu_l)), \tag{7}$$

combined with (4), leading to (3).

It remains to show (6). Observing (A1.1) and (A1.2), one has

$$n(EN)b^{2v+1} cov(\Psi_{\lambda n}, \Psi_{\kappa n})$$

$$= \frac{1}{b} E\left\{\frac{1}{EN}\left[\sum_{j=1}^{N} \psi_\lambda(T_j, Y_j) K_1\left(\frac{t-T_j}{b}\right)\right]\left[\sum_{k=1}^{N} \psi_\kappa(T_k, Y_k) K_1\left(\frac{t-T_k}{b}\right)\right]\right\}$$

$$-\frac{EN}{b} E\left[\frac{1}{EN} \sum_{j=1}^{N} \psi_\lambda(T_j, Y_j) K_1\left(\frac{t-T_j}{b}\right)\right]$$

$$\times E\left[\frac{1}{EN} \sum_{k=1}^{N} \psi_\kappa(T_k, Y_k) K_1\left(\frac{t-T_k}{b}\right)\right]$$

$$\equiv I_1 - I_2.$$

It is obvious that $I_2 = O(b) = o(1)$ from the derivation of (5). For I_1, it can be written as

$$I_1 = \frac{1}{b}E\left[\frac{1}{EN}\sum_{j=1}^{N}\psi_\lambda(T_j,Y_j)\psi_\kappa(T_j,Y_j)K_1^2\left(\frac{t-T_j}{b}\right)\right]$$
$$+\frac{1}{b}E\left[\frac{1}{EN}\sum_{1\leqslant j\neq k\leqslant N}\psi_\lambda(T_j,Y_j)\psi_\kappa(T_k,Y_k)K_1\left(\frac{t-T_j}{b}\right)K_1\left(\frac{t-Y_k}{b}\right)\right]$$
$$\equiv Q_1 + Q_2.$$

Applying (A1.1) and (A1.2), one has

$$Q_1 = \frac{1}{b}E\left\{\frac{1}{EN}\sum_{j=1}^{N}E\left[\psi_\lambda(T_j,Y_j)\psi_\kappa(T_j,Y_j)K_1^2\left(\frac{t-T_j}{b}\right)\middle| N\right]\right\}$$
$$=\frac{1}{b}E\left[\psi_\lambda(T,Y)\psi_\kappa(T,Y)K_1^2\left(\frac{t-Y}{b}\right)\right] = \sigma_{\lambda\kappa} + o(1).$$

Then (4) will hold, observing (A1.1) and the following argument that guarantees the local property of the kernel-based estimators with the presence of within-subject correlation in longitudinal or functional data,

$$Q_2 = \frac{1}{bEN}E\left\{\sum_{1\leqslant j\neq k\leqslant N}^{N}E\left[\psi_\lambda(T_j,Y_j)\psi_\kappa(T_k,Y_k)K_1\left(\frac{t-T_j}{b}\right)K_1\left(\frac{t-T_k}{b}\right)\middle| N\right]\right\}$$
$$=\frac{EN(N-1)}{bEN}E\left[\psi_\lambda(T_1,Y_1)\psi_\kappa(T_2,Y_2)K_1\left(\frac{t-T_1}{b}\right)\right]K_1\left(\frac{t-T_2}{b}\right)$$
$$=\frac{bEN(N-1)}{EN}\int_{\Re^4}\psi_\lambda(t-ub,y_1)\psi_\kappa(t-vb,y_2)K_1(u)K_2(v)$$
$$\times g_2(t-ub,t-vb,y_1,y_2)\,du\,dv\,dy_1\,dy_2$$
$$=\frac{bEN(N-1)}{EN}\int_{\Re^2}\psi_\lambda(t,y_1)\psi_\kappa(t,y_2)g_2(t,t,y_1,y_2)\,dy_1\,dy_2 + o(b) = o(1),$$

i.e., the within-subject correlation can be ignored while deriving the asymptotic variance. □

2.2. Asymptotic normality of two-dimensional smoother

The general asymptotic result can be extended to two-dimensional smoothing. Let $(\boldsymbol{v},\boldsymbol{k})$ denote the multi-indices $\boldsymbol{v} = (v_1, v_2)$ and $\boldsymbol{k} = (k_1, k_2)$, where $|\boldsymbol{v}| = v_1 + v_2$ and $|\boldsymbol{k}| = k_1 + k_2$. In two-dimensional smoothing, more regularity assumptions are needed for joint densities. Let $f_2(s,t)$ be the joint density of (T_j, T_k), and $g_4(s,t,s',t',y_1,y_2,y_1',y_2')$ the joint density of $(T_j,T_k,T_{j'},T_{k'},Y_j,Y_k,Y_{j'},Y_{k'})$ where $j \neq k$, $(j,k) \neq (j',k')$. Denote the covariance surface by $C(s,t) = cov(X(T_j), X(T_k)|T_j = s, T_k = t)$. The following regularity conditions are assumed, where $U(s,t)$ is some neighborhood of $\{(s,t)\}$,

(C1.1) $\frac{d^{|\boldsymbol{k}|}}{du^{k_1}\,dv^{k_2}}f_2(u,v)$ exists and is continuous on $(u,v) \in U(s,t)$, and $f_2(u,v) > 0$ for $(u,v) \in U(s,t)$;

(C1.2) $g_2(u, v, y_1, y_2)$ is continuous on $(u, v) \in U(s, t)$ uniformly in $(y_1, y_2) \in \Re^2$; $\frac{d^{|\boldsymbol{k}|}}{du^{k_1}\, dv^{k_2}} g_2(u, v, y_1, y_2)$ exits and is continuous on $(u, v) \in U(s, t)$ uniformly in $(y_1, y_2) \in \Re^2$;

(C1.3) $g_4(u, v, u', v', y_1, y_2, y_1', y_2')$ is continuous on $(u, v, u', v') \in U(s, t)^2$ uniformly in $(y_1, y_2, y_1', y_2') \in \Re^4$;

(C1.4) $\frac{d^{|\boldsymbol{k}|}}{du^{k_1}\, dv^{k_2}} C(u, v)$ exists and is continuous on $(u, v) \in U(s, t)$.

Let K_2 be nonnegative bivariate kernel functions used in the two-dimensional smoothing. The assumptions for kernels K_2 are as follows,

(C2.1) K_2 is compacted supported with $\|K_2\|^2 = \int_{\Re^2} K_2^2(u, v)\, du\, dv < \infty$, and is symmetric with respect to coordinates u and v.

(C2.2) K_2 is a kernel function of order $(|\boldsymbol{v}|, |\boldsymbol{k}|)$, i.e.,

$$\sum_{\ell_1+\ell_2=|\boldsymbol{l}|} \int_{\Re^2} u^{\ell_1} v^{\ell_2} K_2(u, v)\, du\, dv = \begin{cases} 0, & 0 \leqslant |\boldsymbol{l}| < |\boldsymbol{k}|, |\boldsymbol{l}| \neq |\boldsymbol{v}|, \\ (-1)^{|\boldsymbol{v}|}|\boldsymbol{v}|!, & |\boldsymbol{l}| = |\boldsymbol{v}|, \\ \neq 0, & |\boldsymbol{l}| = |\boldsymbol{k}|. \end{cases} \tag{8}$$

Let $h = h(n)$ be a sequence of bandwidths used in two-dimensional smoothing, while it is possible that the bandwidths used for two arguments may be different. Since we will focus on the estimator of the covariance surface that is symmetric about the diagonal, it is sufficient to consider the identical bandwidths for the two arguments. The asymptotics is developed as $n \to \infty$ as follows:

(C3) $h \to 0$, $nEN^2 h^{|\boldsymbol{v}|+2} \to \infty$, $hEN^3 \to 0$, and $nE[N(N-1)]h^{2|\boldsymbol{k}|+2} \to e^2$ for some $0 \leqslant e < \infty$.

Similar to the one-dimensional smoothing case, assumptions (C3) and (A1.1) guarantee the local property of the bivariate kernel-based estimators with the presence of within-subject correlation.

Let $\{\phi_\lambda\}_{\lambda=1,\ldots,l}$ be a collection of real functions $\phi_\lambda : \Re^4 \to \Re$, $\lambda = 1, \ldots, l$, satisfying

(C4.1) $\phi_\lambda(s, t, y_1, y_2)$ are continuous on $\{(s, t)\}$ uniformly in $(y_1, y_2) \in \Re^2$;

(C4.2) $\frac{d^{|\boldsymbol{k}|}}{ds^{k_1} dt^{k_2}} \phi_\lambda(s, t, y_1, y_2)$ exit for all arguments (s, t, y_1, y_2) and are continuous on $\{(s, t)\}$ uniformly in $(y_1, y_2) \in \Re^2$.

Then the general weighted averages of two-dimensional smoothing are defined by, for $1 \leqslant \lambda \leqslant l$,

$$\Phi_{\lambda n} = \Phi_{\lambda n}(t, s) = \frac{1}{nE[N(N-1)]h^{|\boldsymbol{v}|+2}} \sum_{i=1}^{n} \sum_{1 \leqslant j \neq k \leqslant N_i} \phi_\lambda(T_{ij}, T_{ik}, Y_{ij}, Y_{ik}) \times K_2\left(\frac{s - T_{ij}}{h}, \frac{t - T_{ik}}{h}\right).$$

Let

$$m_\lambda = m_\lambda(s, t) = \sum_{v_1+v_2=|\boldsymbol{v}|} \frac{d^{|\boldsymbol{v}|}}{ds^{v_1}\, dt^{v_2}} \int_{\Re^2} \phi_\lambda(s, t, y_1, y_2) g_2(s, t, y_1, y_2)\, dy_1\, dy_2, \quad 1 \leqslant \lambda \leqslant l,$$

and

$$\omega_{\kappa\lambda} = \omega_{\kappa\lambda}(s,t) = \int_{\Re^2} \phi_\kappa(s,t,y_1,y_2)\phi_\lambda(s,t,y_1,y_2)g_2(s,t,y_1,y_2)dy_1dy_2\|K_2\|^2,$$

$$1 \leqslant \kappa, \lambda \leqslant l,$$

and $H : \Re^l \to \Re$ is a function with continuous first order derivatives as previously defined.

Theorem 2. *If assumptions* (A1.1), (A1.2) *and* (C1.1)–(C4.2) *hold*, *then*

$$\sqrt{n\bar{N}(\bar{N}-1)h^{2|\boldsymbol{\nu}|+2}}[H(\Phi_{1n},\ldots,\Phi_{ln}) - H(m_1,\ldots,m_l)]$$

$$\xrightarrow{\mathcal{D}} \mathcal{N}(\gamma, [DH(m_1,\ldots,m_l)]^T\Omega[DH(m_1,\ldots,m_l)]), \tag{9}$$

where

$$\gamma = \frac{(-1)^{|\boldsymbol{k}|}e}{|\boldsymbol{k}|!}\sum_{\lambda=1}^{l}\left\{\sum_{k_1+k_2=|\boldsymbol{k}|}\int_{\Re^2} u^{k_1}v^{k_2}K_2(u,v)\,du\,dv\frac{d^{|\boldsymbol{k}|}}{ds^{k_1}\,dt^{k_2}}\right.$$

$$\left.\times\int_{\Re^2}\phi_\lambda(s,t,y_1,y_2)g_2(s,t,y_1,y_2)\,dy_1\,dy_2\right\}$$

$$\times\left\{\frac{\partial H}{\partial m_\lambda}(m_1,\ldots,m_l)^T\right\},$$

$$\Omega = (\omega_{\kappa\lambda})_{1\leqslant\kappa\leqslant l}.$$

The proof of Theorem 2 essentially follows that of Theorem 1 with appropriate modifications which are required for two-dimensional smoothing.

3. Applications to nonparametric regression estimators for functional or longitudinal data

Although various versions of kernel-based estimators have been introduced in literature, Nadaraya–Waston and local polynomial, especially local linear estimators, are the most commonly used non-parametric smoothing techniques in longitudinal or functional data analysis. Due to within-subject correlation, the asymptotic behavior in terms of bias and variance of these estimators for noisily observed longitudinal or functional data has yet been as well understood as for i.i.d. data. Especially, asymptotic results for covariance estimators do not exist. Therefore in this section, we apply the asymptotic results developed for general functionals to Nadaraya–Waston and local linear estimators of regression function and covariance surface to obtain their asymptotic distributions.

3.1. Asymptotic distributions of mean estimators

We apply Theorem 1 to the local asymptotic distributions of the commonly used Nadaraya–Waston kernel estimator $\hat{\mu}_{\mathrm{N}}(t)$ and local linear estimator $\hat{\mu}_{\mathrm{L}}(t)$ for functional/longitudinal data:

$$\hat{\mu}_{\mathrm{N}}(t)=\left[\sum_{i=1}^{n}\sum_{j=1}^{N_i}K_1\left(\frac{t-T_{ij}}{b}\right)Y_{ij}\right]\Bigg/\left[\sum_{i=1}^{n}\sum_{j=1}^{N_i}K_1\left(\frac{t-T_{ij}}{b}\right)\right], \tag{10}$$

$$\hat{\mu}_{\mathrm{L}}(t)=\hat{\alpha}_0(t)=\underset{(\alpha_0,\alpha_1)}{\arg\min}\left\{\sum_{i=1}^{n}\sum_{j=1}^{N_i}K_1\left(\frac{t-T_{ij}}{b}\right)[Y_{ij}-(\alpha_0+\alpha_1(T_{ij}-t))]^2\right\}. \tag{11}$$

Corollary 1. *If assumptions* (A1.1), (A1.2), *and* (B1.1)–(B3) *hold with* $v=0$ *and* $k=2$, *then*

$$\sqrt{n\bar{N}b}[\hat{\mu}_{\mathrm{N}}(t)-\mu(t)]\xrightarrow{\mathcal{D}}\mathcal{N}\left(\frac{d}{2}\frac{\mu^{(2)}(t)f(t)+2\mu^{(1)}(t)f^{(1)}(t)}{f(t)}\sigma_{K_1}^2,\ \frac{var(Y|T=t)\|K_1\|^2}{f(t)}\right), \tag{12}$$

where d is as in (B3), $\sigma_{K_1}^2=\int u^2K_1(u)\,du$ *and* $\|K_1\|^2=\int K_1^2(u)\,du$.

One can see that the variance function $var(Y|T=t)$ can be easily derived from model (1) as $var(Y|T=t)=\sum_{k=1}^{\infty}\lambda_k\phi_k^2(t)+\sigma^2$, where λ_k and ϕ_k are the variances of the uncorrelated random coefficients and corresponding basis functions, and σ^2 is the variance of the measurement error.

Proof. Choose $v=0$, $k=2$, $\psi_1(u,v)=v$, $\psi_2(u,v)\equiv 1$ and $H(x_1,x_2)=x_1/x_2$ in Theorem 1, then $\hat{\mu}_{\mathrm{N}}(t)=H(\Psi_{1n},\Psi_{2n})$. To compute β, use $DH(\mu_1,\mu_2)=(1/\mu_2,-\mu_1/\mu_2^2)$, and note $\mu_1(t)=\int tg(t,v)\,dv=f(t)\mu(t)$. It is easy to show that $(d^2/dt^2)\mu_1(t)=[f^{(2)}\mu+2f^{(1)}\mu^{(1)}+f\mu^{(2)}](t)$, and $(d^2/dt^2)\mu_2(t)=f^{(2)}(t)$, leading to the bias term in (12). For the asymptotic variance, note that $\sigma_{11}=\|K_1\|^2\int v^2g(t,v)\,dv=\|K_1\|^2E(Y^2|T=t)f(t)$, $\sigma_{12}=\sigma_{21}=\|K_1\|^2\mu(t)f(t)$, $\sigma_{22}=\|K_1\|^2f(t)$, and $DH(\mu_1,\mu_2)=(1/\mu_2,-\mu_1/\mu_2^2)$, yielding the variance term in (12). □

Concerning the local linear estimator $\hat{\mu}_{\mathrm{L}}(t)$, one obtains

Corollary 2. *If assumptions* (A1.1), (A1.2), *and* (B1.1)–(B3) *hold with* $v=0$ *and* $k=2$, *then*

$$\sqrt{n\bar{N}b}[\hat{\mu}_{\mathrm{L}}(t)-\mu(t)]\xrightarrow{\mathcal{D}}\mathcal{N}\left(\frac{d}{2}\mu^{(2)}(t)\sigma_{K_1}^2,\ \frac{var(Y|T=t)}{f(t)}\|K_1\|^2\right), \tag{13}$$

where d is as in (B3), $\sigma_{K_1}^2=\int u^2K_1(u)\,du$ *and* $\|K_1\|^2=\int K_1^2(u)\,du$.

Proof. The local linear estimator $\hat{\mu}_{\mathrm{L}}(t)$ of the mean function $\mu(t)$ can be explicitly written as

$$\hat{\mu}_{\mathrm{L}}(t)=\hat{\alpha}_0(t)=\frac{\sum_i\frac{1}{EN}\sum_j w_{ij}Y_{ij}}{\sum_i\frac{1}{EN}\sum_j w_{ij}}-\frac{\sum_i\frac{1}{EN}\sum_j w_{ij}(T_{ij}-t)}{\sum_i\frac{1}{EN}\sum_j w_{ij}}\hat{a}_1(t), \tag{14}$$

where

$$\hat{\alpha}_1(t)=\frac{\sum_i\frac{1}{EN}\sum_j w_{ij}(T_{ij}-t)Y_{ij}-(\sum_i\frac{1}{EN}\sum_j w_{ij}(T_{ij}-t)\sum_i\frac{1}{EN}\sum_j w_{ij}Y_{ij})/(\sum_i\frac{1}{EN}\sum_j w_{ij})}{\sum_i\frac{1}{EN}\sum_j w_{ij}(T_{ij}-t)^2-(\sum_i\frac{1}{EN}\sum_j w_{ij}(T_{ij}-t))^2/(\sum_i\frac{1}{EN}\sum_j w_{ij})}. \tag{15}$$

Here $w_{ij} = K_1((t - T_{ij})/b)/(nb)$, where K_1 is a kernel function of order $(0, 2)$, satisfying (B2.1) and (B2.2), and $\hat{\alpha}_1(t)$ is an estimator for the first derivative $\mu'(t)$ of μ at t.

Observing that Corollary 1 implies $\hat{\mu}_{\mathrm{N}}(t) \xrightarrow{p} \mu(t)$, let $\hat{f}(t) = \sum_i \sum_j w_{ij}/N_i$, it is easy to show $\hat{f}(t) \xrightarrow{p} f(t)$ in analogy to Corollary 1. We proceed to show $\hat{a}_1(t) \xrightarrow{p} \mu'(t)$. Denote $\sigma^2_{K_1} = \int u^2 K_1(u)\,du$, the kernel function $\widetilde{K}_1(t) = -tK_1(t)/\sigma^2_{K_1}$, and define $\Psi_{\lambda n}$, $1 \leqslant \lambda \leqslant 3$ by $\psi_1(u, y) = y$, $\psi_2(u, y) \equiv 1$, $\psi_3(u, y) = u - t$. Observe that $\widetilde{K}_1$ is of order $(1, 3)$, $\hat{f}(t) \xrightarrow{p} f(t)$, and define

$$\widetilde{H}(x_1, x_2, x_3) = \frac{x_1 - x_2\hat{\mu}_{\mathrm{N}}(t)}{x_3 - bx_2^2/\hat{f}(t) \cdot \sigma^2_{K_1}} \quad \text{and} \quad H(x_1, x_2, x_3) = \frac{x_1 - x_2\mu(t)}{x_3}.$$

Then

$$\begin{aligned}\hat{\alpha}_1(t) &= \widetilde{H}(\Psi_{1n}, \Psi_{2n}, \Psi_{3n}) \\ &= \left[H(\Psi_{1n}, \Psi_{2n}, \Psi_{3n}) + \frac{\Psi_{2n}(\mu(t) - \hat{\mu}_{\mathrm{N}}(t))}{\Psi_{3n}}\right] \frac{\Psi_{3n}}{\Psi_{3n} + b^2\Psi^2_{2n}/\hat{f}(t) \cdot \sigma^2_{K_1}}.\end{aligned}$$

Note that $\mu_1 = (\mu' f + mf')(t)$, $\mu_2 = f'(t)$, and $\mu_3 = f(t)$, implying $\Psi_{\lambda n} - \mu_\lambda = O_p(1/\sqrt{n\bar{N}b^3})$, for $\lambda = 1, 2, 3$, by Theorem 1. Using *Slutsky's* Theorem, $|\widetilde{H}(\Psi_{1n}, \Psi_{2n}, \Psi_{3n}) - \mu'(t)| = O_p(1/\sqrt{n\bar{N}b^3})$ follows.

For the asymptotic distribution of $\hat{\mu}_{\mathrm{L}}$, note that

$$\hat{\mu}_{\mathrm{L}}(t) = \frac{\sum_i \frac{1}{EN}\sum_j w_{ij}Y_{ij} - \sum_i \frac{1}{EN}\sum_j w_{ij}(T_{ij} - t)\hat{a}_1(t)}{\sum_i \frac{1}{EN}\sum_j w_{ij}}.$$

Considering $\sqrt{n\bar{N}b}\sum_i \frac{1}{EN}\sum_j w_{ij}(T_{ij} - t) = \sqrt{n\bar{N}b}\sigma^2_{K_1} b^2 \Psi_{2n}$. Since $\tilde{K}_1$ is of order $(1, 3)$, Theorem 1 implies $\Psi_{2n} = f'(t) + O_p(1/\sqrt{n\bar{N}b^3})$, which yields $\sqrt{n\bar{N}b}\sigma^2_{K_1} b^2 \Psi_{2n} = \sqrt{n\bar{N}b^5}\sigma^2_{K_1} f'(t) + \sigma^2_{K_1} O_p(b) = o_p(1)$ by observing $n\bar{N}b^5 \to d^2$ for $0 \leqslant d < \infty$. Since $\hat{f}(t) \xrightarrow{p} f(t)$ and $|\hat{\alpha}_1(t) - \mu'(t)| = O_p(1/\sqrt{n\bar{N}b^3}) = o_p(1)$, we find

$$\begin{aligned}\lim_{n\to\infty} \sqrt{n\bar{N}b}[\hat{\mu}_{\mathrm{L}}(t) - \mu(t)] &\overset{\mathcal{D}}{=} \lim_{n\to\infty} \sqrt{n\bar{N}b} \\ &\times \left\{ \frac{\sum_i \frac{1}{EN}\sum_j w_{ij}Y_{ij} - \mu'(t)\sum_i \frac{1}{EN}\sum_j w_{ij}T_{ij} + t\mu'(t)\sum_i \frac{1}{EN}\sum_j w_{ij}}{\sum_i \frac{1}{EN}\sum_j w_{ij}} - \mu(t) \right\}.\end{aligned}$$

Using the kernel K_1 of order $(0, 2)$, we re-define $\Psi_{\lambda n}$, $1 \leqslant \lambda \leqslant 3$, through $\psi_1(u, y) = y$, $\psi_2(u, y) = u$ and $\psi_3(u, y) \equiv 1$, setting $v = 0$, $k = 2$, $l = 3$ and $H(x_1, x_2, x_3) = [x_1 - \mu'(t)x_2 + t\mu'(t)x_3]/x_3$. Then (13) follows by applying Theorem 1. □

3.2. Asymptotic distributions of covariance estimators

Note that in model (1), $cov(Y_{ij}, Y_{ik}|T_{ij}, T_{ik}) = cov(X(T_{ij}), X(T_{ik})) + \sigma^2\delta_{jk}$, where δ_{jl} is 1 if $j = k$ and 0 otherwise. Let $C_{ijk} = (Y_{ij} - \hat{\mu}(T_{ij}))(Y_{ik} - \hat{\mu}(T_{ik}))$ be the "raw" covariances, where $\hat{\mu}(t)$ is the estimated mean function obtained from the previous step, for instance, $\hat{\mu}(t) = \hat{\mu}_{\mathrm{N}}(t)$ or $\hat{\mu}(t) = \hat{\mu}_{\mathrm{L}}(t)$. It is easy to see that $E[C_{ijk}|T_{ij}, T_{ik}] \approx cov(X(T_{ij}), X(T_{ik})) + \sigma^2\delta_{jk}$. Therefore, the diagonal of the raw covariances should be removed, i.e., only C_{ijk}, $j \neq k$, should be included as input data for the covariance surface smoothing step, as previously observed in Staniswalis and Lee [12] and Yao et al. [15].

Commonly used nonparametric regression estimators of the covariance surface, $C(s,t) = E\{[X(T_1) - \mu(T_1)][X(T_2) - \mu(T_2)|T_1 = s, T_2 = t]\}$, are the two-dimensional Nadaraya–Waston estimator and local linear estimator defined as follows:

$$\widehat{C}_{\mathrm{N}}(s,t) = \left[\sum_{i=1}^{n} \sum_{j \neq k} K_2\left(\frac{s - T_{ij}}{h}, \frac{t - T_{ik}}{h}\right) C_{ijk}\right] \Bigg/ \left[\sum_{i=1}^{n} \sum_{j \neq k} K_2\left(\frac{s - T_{ij}}{h}, \frac{t - T_{ik}}{h}\right)\right], \tag{16}$$

$$\widehat{C}_{\mathrm{L}}(s,t) = \hat{\beta}_0(s,t) = \arg\min_{\boldsymbol{\beta}} \left\{\sum_{i=1}^{n} \sum_{j \neq k} K_2\left(\frac{s - T_{ij}}{h}, \frac{t - T_{ik}}{h}\right) \times [C_{ijk} - f(\boldsymbol{\beta}, (s,t), (T_{ij}, T_{ik}))]^2\right\}, \tag{17}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$ and $f(\boldsymbol{\beta}, (s,t), (T_{ij}, T_{ik})) = \beta_0 + \beta_1(T_{ij} - s) + \beta_2(T_{ik} - t)$. Applying Theorem 2 to the above estimators yields the asymptotic distributions.

Corollary 3. *If assumptions* (A1.1), (A1.2), *and* (C1.1)–(C3) *hold with* $|\boldsymbol{v}| = 0$ *and* $|\boldsymbol{k}| = 2$, *then*

$$\sqrt{n\bar{N}(\bar{N}-1)h^2}[\widehat{C}_{\mathrm{N}}(s,t) - C(s,t)] \xrightarrow{\mathcal{D}} \mathcal{N}\left(\gamma_{\mathrm{N}}(s,t), \frac{v(s,t)\|K_2\|^2}{f_2(s,t)}\right), \tag{18}$$

where e is as in (C3),

$$\gamma_{\mathrm{N}}(s,t) = \frac{e}{4}\sigma_{K_2}^2 \left\{\frac{d^2C(s,t)}{ds^2} f_2(s,t) + \frac{d^2C(s,t)}{dt^2} f_2(s,t) + 2\left[\frac{df_2(s,t)}{dt}\frac{dC(s,t)}{dt} + \frac{df_2(s,t)}{ds}\frac{dC(s,t)}{ds}\right]\right\} \Bigg/ f_2(s,t),$$

$$v(s,t) = var\{(Y_1 - \mu(T_1))(Y_2 - \mu(T_2))|T_1 = s, T_2 = t\},$$

$$\sigma_{K_2}^2 = \int_{\Re^2} (u^2 + v^2) K_2(u,v)\, du\, dv, \quad \|K_2\|^2 = \int_{\mathcal{R}^2} K_2^2(u,v)\, du\, dv.$$

Proof. For two-dimensional Nadaraya–Waston estimator $\widehat{C}_{\mathrm{N}}(s,t)$ of the covariance $C(s,t)$, one uses the raw observations, $C_{ijk} = (Y_{ij} - \hat{\mu}(T_{ij}))(Y_{ik} - \hat{\mu}(T_{ik}))$. Let $\tilde{C}_{ijk} = (Y_{ij} - \mu(T_{ij}))(Y_{ik} - \mu(T_{ik}))$. Note that

$$C_{ijk} = \tilde{C}_{ijk} + (Y_{ij} - \mu(T_{ij}))(\mu(T_{ik}) - \hat{\mu}(T_{ik})) + (Y_{ik} - \mu(T_{ik}))(\mu(T_{ij}) - \hat{\mu}(T_{ij})) + (\mu(T_{ij}) - \hat{\mu}(T_{ij}))(\mu(T_{ik}) - \hat{\mu}(T_{ik})).$$

Applying Lemma 1 in Yao et al. [16], one has the weak uniform convergence rate $\sup_{t \in \mathcal{T}} |\hat{\mu}(t) - \mu(t)| = O_p(1/(\sqrt{n}b))$ for both possible choice of $\hat{\mu}(t) = \hat{\mu}_{\mathrm{N}}(t)$ and $\hat{\mu}(t) = \hat{\mu}_{\mathrm{L}}(t)$. Letting

$\phi_1(t_1, t_2, y_1, y_2) = (y_1 - \mu(t_1))(y_2 - \mu(t_2))$, $\phi_2(t_1, t_2, y_1, y_2) = y_1 - \mu(t_1)$, and $\phi_3(t_1, t_2, y_1, y_2) \equiv 1$, then $\sup_{t,s\in\mathcal{T}} |\Phi_{pn}| = O_p(1)$, for $p = 1, 2, 3$, by Lemma 1 of Yao et al. [16]. This implies that $\sup_{t,s\in\mathcal{T}} |\Phi_{2n}| O_p(1/(\sqrt{n}b)) = O_p(1/(\sqrt{n}b))$ and $\sup_{t,s\in\mathcal{T}} |\Phi_{3n}| O_p(1/(\sqrt{n}b)) = O_p(1/(\sqrt{n}b))$. Since $\sup_{t\in\mathcal{T}} |\hat{\mu}(t) - \mu(t)|^2 = O_p(1/(nb))$ are negligible compared to Φ_{1n}, the Nadaraya–Waston estimator $\widehat{C}_{\mathrm{N}}(s, t)$, of $C(s, t)$ obtained from C_{ijk} is asymptotically equivalent to that obtained from $\tilde{C}_{ijk}$, denoted by $\tilde{C}_{\mathrm{N}}(t, s)$.

Therefore, it is sufficient to show that the asymptotic distribution of $\tilde{C}_{\mathrm{N}}(s, t)$ follows (18). Choose $\boldsymbol{v} = (0, 0)$, $|\boldsymbol{k}| = 2$, $\phi_1(s, t, y_1, y_2) = (y_1 - \mu(s))(y_2 - \mu(t))$, $\phi_2(s, t, y_1, y_2) \equiv 1$ and $H(x_1, x_2) = x_1/x_2$ in Theorem 2, then $\tilde{C}_{\mathrm{N}}(s, t) = H(\Psi_{1n}, \Psi_{2n})$. To compute $\gamma_{\mathrm{N}}(s, t)$, use $DH(m_1, m_2) = (1/m_2, -m_1/m_2^2)$, and note $m_1(s, t) = \int_{\Re^2} (y_1 - \mu(s))(y_2 - \mu(t)) g_2(s, t, y_1, y_2)\, dy_1 dy_2 = f_2(s, t)C(s, t)$ and $m_2(s, t) = f_2(s, t)$. One has $(d^2/dt^2)m_1(s, t) = [(d^2 f_2/dt^2)C + 2(df_2/dt)(dC/dt) + f_2(d^2C/dt^2)](s, t)$, $(d^2/d^2t)m_2(s, t) = d^2 f_2(s, t)/dt^2$ and similar derivatives with respect to the argument s leading to the bias term in (12). For the asymptotic variance, note that $\omega_{11} = \|K_2\|^2 \int_{\Re^2} (y_1 - \mu(s))^2 (y_2 - \mu(t))^2 g_2(s, t, y_1, y_2) dy_1 dy_2 = E[(Y_1 - \mu(T_1))^2 (Y_2 - \mu(T_2))^2 | T_1 = s, T_2 = t) f_2(s, t)\|K_2\|^2$, $\omega_{12} = \omega_{21} = \|K_2\|^2 f_2(s, t)C(s, t)$, $\omega_{22} = \|K_2\|^2 f_2(s, t)$, and $DH(m_1, m_2) = (1/m_2, -m_1/m_2^2)$, yielding the variance term in (12). □

Corollary 4. *If the assumptions* (A1.1), (A1.2), *and* (C1.1)–(C3) *hold with* $|\boldsymbol{v}| = 0$ *and* $|\boldsymbol{k}| = 2$, *then*

$$\sqrt{n\bar{N}(\bar{N} - 1)h^2}[\widehat{C}_{\mathrm{L}}(s, t) - C(s, t)]$$
$$\xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{e}{4}\sigma_{K_2}^2[d^2C(s, t)/ds^2 + d^2C(s, t)/dt^2], \frac{v(s, t)\|K_2\|^2}{f_2(s, t)}\right), \tag{19}$$

where e is as in (C3), $v(s, t) = var\{(Y_1 - \mu(T_1))(Y_2 - \mu(T_2))|T_1 = s, T_2 = t)$, $\sigma_{K_2}^2 = \int_{\Re^2} (u^2 + v^2) K_2(u, v)\, du\, dv$, $\|K_2\|^2 = \int_{\mathcal{R}^2} K_2^2(u, v)\, du\, dv$.

Proof. In analogy to the proof of Corollary 3, the local linear estimator $\widehat{C}_{\mathrm{L}}(s, t)$ obtained from C_{ijk} is asymptotically equivalent to that obtained from $\tilde{C}_{ijk}$, denoted by $\tilde{C}_{\mathrm{L}}(t, s)$. Also denote the solution to (17), after substituting $\tilde{C}_{ijk}$ for C_{ijk}, by $\tilde{\boldsymbol{\beta}}(s, t) = (\tilde{\beta}_0(s, t), \tilde{\beta}_1(s, t), \tilde{\beta}_2(s, t))$, and in fact $\tilde{\beta}_0(s, t) = \tilde{C}_{\mathrm{L}}(s, t)$. For simplicity, let $W_{ijk} = K_2((s - T_{ij})/h, (t - T_{ik})/h)/(nh^2)$ and "$\sum_{i,j\neq k}$" is abbreviation of "$\sum_{i=1}^n \sum_{j\neq k}$". Algebra calculations yield that

$$\tilde{C}_{\mathrm{L}} = \frac{\sum_{i,j\neq k} \tilde{C}_{ijk} W_{ijk} - \tilde{\beta}_1 \sum_{i,j\neq k} W_{ijk} T_{ij} + \tilde{\beta}_1 \sum_{i,j\neq k} W_{ijk} s - \tilde{\beta}_2 \sum_{i,j\neq k} W_{ijk} T_{ik} + \tilde{\beta}_2 \sum_{i,j\neq k} W_{ijk} t}{\sum_{i,j\neq k} W_{ijk}},$$

$$\tilde{\beta}_1 = \frac{R_{00}(S_{10}S_{02} - S_{01}S_{11}) + R_{10}(S_{00}S_{02} - S_{01}S_{20}) - R_{01}(S_{00}S_{11} - S_{10}S_{02})}{S_{00}S_{20}S_{02} - S_{00}S_{11}^2 - S_{10}^2 S_{02} + S_{10}S_{01}S_{11} + S_{20}S_{10}S_{11} - S_{01}S_{20}^2},$$

$$\tilde{\beta}_2 = \frac{R_{00}(S_{10}S_{11} - S_{01}S_{02}) - R_{10}(S_{00}S_{11} - S_{01}S_{20}) + R_{01}(S_{00}S_{20} - S_{10}^2)}{S_{00}S_{20}S_{02} - S_{00}S_{11}^2 - S_{10}^2 S_{02} + S_{10}S_{01}S_{11} + S_{20}S_{10}S_{11} - S_{01}S_{20}^2},$$

where

$$R_{pq} = \sum_{i,j\neq k} W_{ijk}(T_{ij} - s)^p (T_{ik} - t)^q \tilde{C}_{ijk}, \quad S_{pq} = \sum_{i,j\neq k} W_{ijk}(T_{ij} - s)^p (T_{ik} - t)^q.$$

Note that $\tilde{\beta}_1$ and $\tilde{\beta}_2$ are local linear estimators of the partial derivatives of $C(s,t)$, $dC(s,t)/ds$ and $dC(s,t)/dt$, respectively. In analogy to the proof of Corollary 2, it can be shown that $|\tilde{\beta}_1(s,t) - dC(s,t)/ds| = O_p(1/\sqrt{nEN(N-1)h^4})$ and $|\tilde{\beta}_2(s,t) - dC(s,t)/dt| = O_p(1/\sqrt{n\bar{N}(\bar{N}-1)h^4})$ by applying Theorem 2. Then one can substitute $dc(s,t)/ds$, $dC(s,t)/dt$ for $\tilde{\beta}_1(s,t)$, $\tilde{\beta}_2(s,t)$ in $\tilde{C}_{\mathrm{L}}(s,t)$, and denote the resulting estimator by $C^*_{\mathrm{L}}(s,t)$. It is easy to see that

$$\lim_{n\to\infty} \sqrt{n\bar{N}(\bar{N}-1)h^2}[C_{\mathrm{L}}(s,t) - C(s,t)] \overset{\mathcal{D}}{=} \lim_{n\to\infty} \sqrt{n\bar{N}(\bar{N}-1)h^2}[C^*_{\mathrm{L}}(s,t) - C(s,t)].$$

We define $\Phi_{\lambda n}$, $1 \leqslant \lambda \leqslant 4$, through $\phi_1(s,t,y_1,y_2) = (y_1 - \mu(s))(y_2 - \mu(t))$, $\phi_2(s,t,y_1,y_2) = s$, $\phi_3(s,t,y_1,y_2) \equiv 1$ and $\phi_4(s,t,y_1,y_2) = t$. Put $\boldsymbol{v} = (0,0)$, $|\boldsymbol{k}| = 2$ and $H(x_1,x_2,x_3,x_4) = [x_1 - x_2 dC(s,t)/ds - x_4 dC(s,t)/dt]/x_3 + s \cdot dC(s,t)/ds + t \cdot dC(s,t)/dt$. Then result (19) follows by applying Theorem 2. □

4. Extension to fixed design

In the studies of functional data, it is often the case that the repeated measurements are recorded densely and regularly over time by machine. For realistic reason, the measurements are assumed to be contaminated with experimental error. The asymptotic distribution results developed for random design that is often encountered in longitudinal data can be extended to fixed equally spaced design for typical functional data with slight modifications. Now the density functions $g(t,y)$ is defined as the density of $Y(t)$ at the fixed time t, $g_2(s,t,y_1,y_2)$ is the joint density of $(Y(s), Y(t))$ at (s,t), and $g_4(s,t,s',t',y_1,y_2,y'_1,y'_2)$ is the joint density of $(Y(s), Y(t), Y(s'), Y(t'))$ at (s,t,s',t'), while $f(t)$ for the random observation time T is not needed. The fixed equally spaced design considered here is as follows:

(A1*) $N_i(n) = N(n)$, $T_{i,j+1} - T_{i,j} = T_{i,j'+1} - T_{i,j'}$ for $1 \leqslant j, j' \leqslant N$, and $T_{ij} = T_{i'j}$ for $1 \leqslant i, i' \leqslant n$ and $1 \leqslant j \leqslant N$.

The extension that will be discussed in this section is also applicable to the case that the vector of observation times $\boldsymbol{T}_i$ are independent and identically distributed with $N_i = N$ and $T_{i,j+1} - T_{i,j} = T_{i,j'+1} - T_{i,j'}$, which is in fact a case that lies between random unbalanced and fixed equally spaced designs.

For the general result in Theorem 1, the weighted averages $\Psi_{\lambda n}$ should be re-defined by

$$\Psi_{\lambda n} = \frac{1}{nNb^{\nu+1}} \sum_{i=1}^{n} \sum_{j=1}^{N} \psi_\lambda(T_{ij}, Y_{ij}) K_1\left(\frac{t - T_{ij}}{b}\right).$$

Assumption (B3) now becomes
(B3*) $b \to 0$, $nNb^{\nu+1} \to \infty$, $bN \to 0$ and $nNb^{2k+1} \to d^2$ with $0 \leqslant d < \infty$.
Then in analogy to the proof of Theorem 1 with appropriate modifications, Theorem 1 is still valid if the assumptions (A1*), (B1.1)–(B2.2), (B3*), and (B4.1)–(B4.2) hold.

For the general result of two-dimensional smoothing in Theorem 2, one modifies assumption (C3) as follows:
(C3*) $h \to 0$, $nN^2 h^{|\boldsymbol{v}|+2} \to 0$, $hN^3 \to 0$, and $nN(N-1)h^{2|\boldsymbol{k}|+2} \to e^2$ for $0 \leqslant e < \infty$.
Then Theorem 2 holds under assumptions (A1*), (C1.1)–(B2.2), (C3*), and (C4.1)–(C4.2).

Therefore, Theorems 1 and 2 can be applied to obtain asymptotic distributions of kernel-based estimators for the mean and covariance functions under fixed equally spaced design. For instance, the asymptotic normal distributions of the local linear estimators $\hat{\mu}_{\mathrm{L}}(t)$ and $\widehat{C}_{\mathrm{L}}(s,t)$ are similar to those in Corollaries 3 and 4, with $f(t)$ replaced by $1/|\mathcal{T}|$ and $f(s,t)$ replaced by $1/|\mathcal{T}|^2$, where $|\mathcal{T}|$ is the length of the interval.

5. Simulation study

A numerical study is conducted to evaluate the derived asymptotic properties. The key finding in this paper is that the asymptotic results for functional or longitudinal are comparable to those obtained from independent data, i.e., the influence of within-subject covariance does not play significant role in determining the asymptotic bias and variance. For simplicity, we focus on the local polynomial mean estimators which are often superior to the Nadaraya–Waston estimators.

We first generated $M = 200$ samples consisting of $n = 50$ i.i.d. random trajectories each. Following model (1), the simulated process has a mean function $\mu(t) = (t - 1/2)^2$, $0 \leqslant t \leqslant 1$ which has a constant second derivative $\mu^{(2)}(t) = 2$, and a constant within-subject covariance function derived from a random intercept $\xi_1 \overset{\text{i.i.d.}}{\sim} N(0, \lambda_1)$, where $\lambda_1 = 0.01$ and$\phi_1(t) = 1$, $0 \leqslant t \leqslant 1$. The measurement error in (1) was set $\varepsilon_{ij} \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$, where $\sigma^2 = 0.01$. A random design was used, where the numbers of observations for each subject N_i were chosen from $\{2, 3, 4, 5\}$ with equal likelihood and the locations of the observations were uniformly distributed on $[0, 1]$, i.e., $T_{ij} \overset{\text{i.i.d.}}{\sim} U[0, 1]$. For comparison, we generated $M = 200$ samples of $n = 50$ i.i.d. random trajectories which have the same structure as in model (1) but no within-subject correlation. Letting $\xi_{i1} = 0$ and $\varepsilon_{ij} \overset{\text{i.i.d.}}{\sim} N(0, \sqrt{\lambda_1 + \sigma^2})$ leads to independent data with the same mean and variance functions. Therefore, the two sets of data have the same asymptotic distribution for the local polynomial mean estimators. We also generated $M = 200$ correlated and independent samples, respectively, consisting of $n = 200$ trajectories each for demonstrating the asymptotic behavior with the increasing sample size n.

Here we use the Epanechnikov kernel function, i.e., $K_1(u) = 3/4(1 - u^2)\mathbf{1}_{[-1,1]}(u)$, where $\mathbf{1}_A(u) = 1$ if $u \in A$ and 0 otherwise for any set A. Note that $n(EN)b^{2k+1} \to d^2$ in (B3), $\mu^{(2)}(t) = 2$, $var(Y|T = t) = \lambda_1 + \sigma^2 = 0.02$, and the design density $f(t) = 1$, where $k = 2$ for local polynomial estimators and b is the bandwidth used for the mean estimation. From the above construction, one can calculate the asymptotic variance and bias of the local polynomial mean estimators $\mu_{\mathrm{L}}(t)$ using Corollary 2 which is in fact applicable for both correlated and independent data. Since the bias and variance terms are both constant in our simulation framework, for convenience we compare the asymptotic integrated squared bias and variance with the empirical integrated squared bias and variance obtained using Monte Carlo average from $M = 200$ simulated samples based on $\int_0^1 E[\{\hat{\mu}_{\mathrm{L}}(t) - \mu(t)\}^2]\,dt = \int_0^1 \{\hat{\mu}_{\mathrm{L}}(t) - E[\hat{\mu}_{\mathrm{L}}(t)]\}^2\,dt + \int_0^1 \{E[\hat{\mu}_{\mathrm{L}}(t)] - \mu(t)\}^2\,dt$. The asymptotic integrated squared bias and variance are given by

$$\text{AIBIAS} = \frac{1}{2}\sigma_{K_1}^2 b^4, \quad \text{AIVAR} = \frac{0.02 \times \|K_1\|^2}{n\bar{N}b}, \tag{20}$$

and the asymptotic integrated mean squared error AIMSE = AIBIAS + AIVAR, where $\sigma_{K_1}^2 = \int u^2 K_1(u)\,du$, $\|K_1\|^2 = \int K_1^2(u)\,du$ and $\bar{N} = (1/n)\sum_{i=1}^n N_i$, while the empirical integrated squared bias, variance and mean squared error are denoted by EIBIAS, EIVAR and EIMSE,

The asymptotic and empirical quantities, such as the integrated squared bias, variance and mean squared error, are shown in Fig. 1 for the correlated/independent data with sample size $n = 50/n = 200$, respectively. From Fig. 1, it is obvious that the asymptotic approximation is improved by increasing the sample size. The asymptotic quantities AIBIAS, AIVAR and AIMSE agree with the

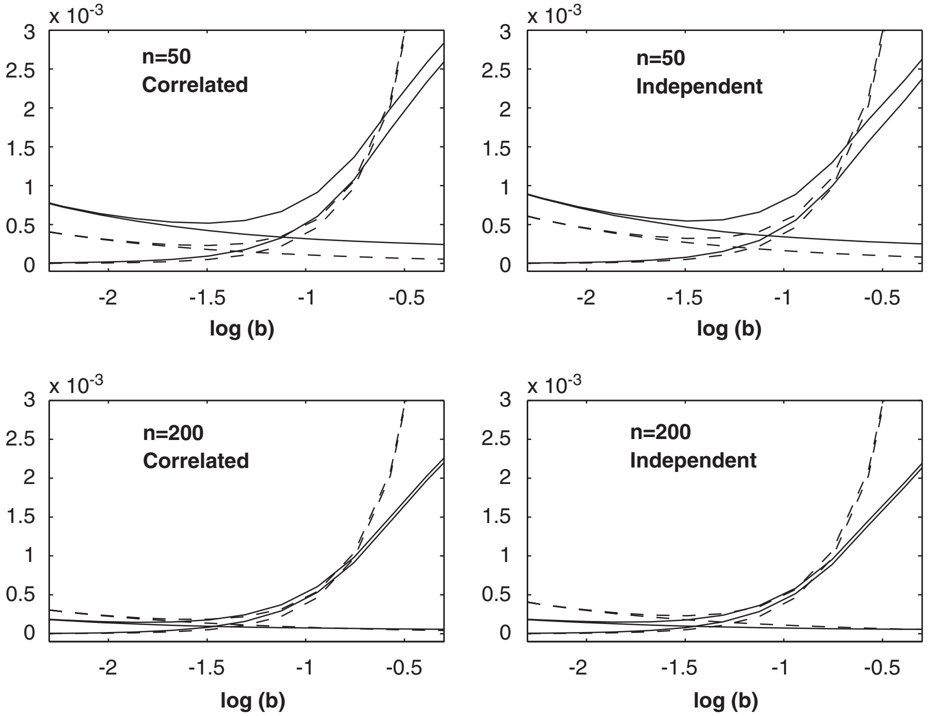

Fig. 1. Shown are the empirical quantities (solid, including EIBIAS, EIVAR, EIMSE) and asymptotic quantities (dashed, including AIBIAS, AIVAR, AIMSE) versus $\log(b)$ for correlated (left panels) and independent (right panels) data with different sample sizes $n = 50$ (top panels) and $n = 200$ (bottom panels), where b is the bandwidth used in the smoothing. In each panel, the integrated squared bias is the one with increasing pattern, the integrated variance is the one with decreasing pattern, and they cross each other, while the integrated mean squared error, which is larger than both integrated squared bias and variance for any bandwidth b, usually decreases first and then increases after reaching a minimum.

empirical quantities EIBIAS, EIVAR and EIMSE for both correlated and independent data. For the simulated data with the same sample size n, such asymptotic approximations for correlated and independent data are well comparable in pattern and magnitude. This provides the evidence that the within-subject correlation indeed does not have obvious influence on the asymptotic behavior of the local polynomial estimators compared to the standard rate obtained from independent data, which is consistent with our theoretical derivations.

6. Discussion

In this paper, the asymptotic distributions of kernel-based nonparametric regression estimators for functional or longitudinal data are studied. In particular, it derives general results for asymptotic distributions of real-valued functions with arguments which are functionals formed by weighted averages obtained from longitudinal or functional data. These asymptotic distribution results are comparable to those obtained from identically distributed and independent data. The conditions for the validity of local property of kernel-based estimators are provided in (B3), (C3), (B3*) and (C3*) with the presence of within-subject correlation in such data, under random unbalanced design described in (A1.1) and (A1.2), fixed equally spaced design described in (A1*), and some case lying between them. The proposed results could also be extended to more complicated cases, such as "panel data" where observations for different subjects are obtained at a series of common time points during a longitudinal follow-up. If considering random design, the density of the jth observation time T_j could be assumed to be $f_j(t)$, then the results are readily applied to this case with appropriate modifications with respect to the different marginal densities.

The general asymptotic distribution results in univariate and bivariate smoothing settings are applied to the kernel-based estimators of the mean and covariance functions, which yields asymptotic normal distributions of these estimators. To the best of our knowledge, there are no asymptotic distribution results available in literature for nonparametric estimators of covariance function obtained from observed noisy longitudinal or functional data. This provides theoretical basis and practical guidance for the nonparametric analysis of functional or longitudinal data with important potential applications that are based on the asymptotic distributions. For example, asymptotic confidence bands or regions for the regression curve or the covariance surface can be constructed based on their asymptotic distributions. Since, due to their heavy computational load, commonly used procedures (such as cross-validation) for bandwidth selection in two-dimensional settings are not feasible, one important research problem is to seek efficient approaches for choosing such smoothing parameters. Also functional principal component analysis, an increasingly popular tool for functional data analysis, is based on eigen-decomposition of the estimated covariance function. Thus, the influence of the asymptotic properties of covariance estimators on the estimated eigenfunctions is another potential research of interest.

References

[1] P.K. Bhattacharya, H.G. Müller, Asymptotics for nonparametric regression, Sankhyā 55 (1993) 420–441.
[2] G. Boente, R. Fraiman, Kernel-based functional principal components, Statist. Probab. Lett. 48 (1999) 335–345.
[3] H. Cardot, F. Ferraty, P. Sarda, Functional linear model, Statist. Probab. Lett. 45 (1999) 11–22.
[4] P. Diggle, P. Hergerty, K.Y. Liang, S. Zeger, Analysis of Longitudinal Data, Oxford Unversity Press, Oxford, 2002.
[5] J. Fan, I. Gijbels, Local Polynomial Modelling and Its Applications, Chapman & Hall, London, 1996.
[6] P. Hall, N.I. Fisher, B. Hoffmann, On the nonparametric estimation of covariance functions, Ann. Statist. 22 (1994) 2115–2134.
[7] P. Hall, P. Patil, Properties of nonparametic estimators of autocovariance for stationary random fields, Probab. Theory Related Fields 99 (1994) 399–424.
[8] J.D. Hart, T.E. Wehrly, Kernel regression estimation using repeated measurements data, J. Amer. Statist. Assoc. 81 (1986) 1080–1088.
[9] X. Lin, R.J. Carroll, Nonparametric function estimation for clustered data when the predictor is measured without/with error, J. Amer. Statist. Assoc. 95 (2000) 520–534.
[10] J.O. Ramsay, J.B. Ramsey, Functional data analysis of the dynamics of the monthly index of nondurable goods production, J. Econom. 107 (2002) 327–344.
[11] T.A. Sererini, J.G. Staniswalis, Quasi-likelihood estimation in semiparametric models, J. Amer. Statist. Assoc. 89 (1994) 501–511.
[12] J.G. Staniswalis, J.J. Lee, Nonparametric regression analysis of longitudinal data, J. Amer. Statist. Assoc. 93 (1998) 1403–1418.
[13] A.P. Verbyla, B.R. Cullis, M.G. Kenward, S.J. Welham, The analysis of designed experiments and longitudinal data using smoothing splines (with discussion), Appl. Statist. 48 (1999) 269–311.
[14] C.J. Wild, T.W. Yee, Additive extensions to generalized estimating equation methods, J. Roy. Statist. Soc., Ser B 58 (1996) 711–725.
[15] F. Yao, H.G. Müller, A.J. Clifford, S.R. Dueker, J. Lin Follett, B.A.Y. Buchholz, J.S. Vogel, Shrinkage estimation for functional principal component scores with application to the population kinetics of plasma folate, Biometrics 59 (2003) 676–685.
[16] F. Yao, H.G. Müller, J.L. Wang, Functional data analysis for sparse longitudinal data, J. Amer. Statist. Assoc. 100 (2005) 577–590.