Bootstrapping sums of independent but not identically distributed continuous processes with applications to functional data

Chung Chang\textsuperscript{a,\ast}, R. Todd Ogden \textsuperscript{b}

\textsuperscript{a} Department of Mathematical Sciences, New Jersey Institute of Technology, University Heights, Newark, NJ 07102-1982, United States
\textsuperscript{b} Department of Biostatistics, Columbia University, 722 W. 168th Street, 6th floor, New York, NY 10032, United States

\textbf{A R T I C L E I N F O}

\textbf{Article history:}
Received 24 August 2007
Available online 3 December 2008

\textbf{AMS subject classification:}
62F40

\textbf{Keywords:}
Functional central limit theorem
Multiplier bootstrap
Multiple-hypothesis testing
Functional data

\textbf{A B S T R A C T}

In many areas of application, the data are of functional nature, such as (one-dimensional) spectral data and two- or three-dimensional imaging data. It is often of interest to test for the significance of some set of factors in the functional observations (e.g., test for the mean differences between two groups). Testing hypotheses point-by-point (voxel-by-voxel in neuroimaging studies) results in a severe multiple-comparisons problem as the number of measurements made per observation is typically much larger than the number of observations ("large \textit{p}, small \textit{n}"). Thus solutions to this problem should take into account the spatial correlation structure inherent in the data. Popular approaches in such a setting include the general Statistical Parametric Mapping (SPM) approach and the permutation test, but these rely on strong parametric and exchangeability assumptions. In situations in which these assumptions are not satisfied, a non-parametric multiplier bootstrap approach may be used. Motivated by this problem, we present general results for multiplier bootstraps for sums of independent but not identically distributed processes. We also consider the application of these results to an imaging setting and provide sufficient conditions that will ensure asymptotic control of the familywise error rate.

\textcopyright 2008 Elsevier Inc. All rights reserved.

1. Introduction

In modern practice, data analysts often have to deal with many thousands of hypothesis testing problems simultaneously, a problem that often arises in functional data analysis (FDA; Ramsay and Silverman \cite{ramsay}). For example, in a typical Positron Emission Tomography (PET) imaging study, given two groups of subjects (depressed and normal control), it is of interest to test for a difference between groups in the average density of a target neuroreceptor at each location in the brain. We might consider binding potential (BP), a measure of the density of the receptor of interest, as the response variable and disease status, as well as age, sex, and other covariates, as explanatory variables in a multiple-regression setting.

Another practical example of testing on functional data is to compare two groups of curves. For instance, Ferraty et al. \cite{ferraty} compared two groups of meat data (fat lower than 20\% vs. fat higher than 20\%) obtained from food industry quality control. For each meat sample, a 100-channel spectrum of absorbances within a certain wavelength range is measured. Interest centers on testing for a difference in average absorbance at each level of the spectrum between the two groups of curves. Our primary motivation is in imaging applications, but the theory developed in this article can be applied to functional data of any dimensionality.

In the PET imaging example, the complicated spatial correlation structure inherent in these data makes the multiple-comparisons problem difficult. Perhaps the most prevalent approach for dealing with this problem is so-called Statistical...
Parametric Mapping (SPM; see, e.g., [3]). The validity of SPM applications rests on some strong assumptions about the data: that the data, at each voxel, are normally distributed and, across voxels, are derived from continuous random fields with a stationary covariance structure. Both the distributional and the stationary assumptions are not always satisfied in practice, so a nonparametric method is far preferable in some applications.

Nichols and Holmes [4] used a nonparametric permutation test to address this multiple-comparisons problem. This procedure requires that under the null hypothesis of no experimental effect (e.g., no effect of group on BP), the labels are exchangeable. But in multiple-regression models it is not possible to build exchangeable labels, as adding a single covariate would make the permutation test unusable. A more general applicable multiplier bootstrap procedure was proposed by Zhu et al. [5] for addressing the multiple-comparisons problem in a multiple-regression setting. The validity of the inference of this multiplier bootstrap approach depends on observing a large number of subjects. In imaging applications and in other FDA situations, the number of voxels is typically much larger than the number of subjects (“large p, small n”) so this must be considered in determining whether such a procedure is appropriate in practice. The purpose of this article is to investigate sufficient conditions for ensuring that the bootstrap approach is appropriate for functional data. Here, we simplify the approach proposed by Zhu et al. [5] and study its theoretical properties.

To deal with the “large p and small n” issue for the bootstrap approach, instead of considering the collection of test statistics as a discrete random field, we consider it as an interpolating random field indexed in \( \mathbb{R}^2 \) (g is the dimension of the image). If the field is smooth and the resolution of the image is high, then the interpolating random field can be approximated by a continuous one, which can be expressed as a sum of independent but not identically distributed stochastic processes. The limiting distribution for such sums has been intensively studied, and the key result is known as the “functional central limit theorem” (FCLT).

The multiplier bootstrap approach is a resampling scheme which helps one to compute the approximate distribution of the sum. Our multiplier bootstrap theorem, Theorem 2, is motivated by the Jain–Marcus FCLT [6,7]. There are certainly many other choices for FCLT such as that given in [8], which is more general than the Jain–Marcus theorem. However, the Jain–Marcus theorem focuses on the stochastic processes with certain smoothness conditions, which is appropriate for brain imaging study and other applications on functional data. Kosorok [9] also proposed multiplier bootstraps, motivated by the FCLT of Pollard [10]. The manageability condition in Kosorok’s paper is quite general but it is not clear whether this condition would be satisfied for applications in imaging and other functional data.

In Section 2, we introduce the framework of brain imaging and state the relevant multiple-hypothesis testing problem as well as the multiplier bootstrap procedure that deals with this problem. In Section 3, we provide some necessary background and state the Jain–Marcus theorem, and then in Section 4 we present a theorem concerning multiplier bootstraps of sums of independent stochastic processes. In Section 5, we will present results for our multiple-comparisons problem, specifying sufficient conditions, including spatial smoothing and the resolution of imaging, to ensure that the multiplier bootstrap procedure controls the familywise error rate.

### 2. Framework and bootstrap method of brain imaging

Assume there are \( n \) subjects in the study and for each subject there are \( p \) voxels in a rectangle region \( T \) in \( \mathbb{R}^2 \), where \( p = \prod_{i=1}^{d} n_i \), Fix \( t \in T \); consider a linear model:

\[
y(t) = X \beta(t) + \epsilon(t),
\]

where \( y(t) = (y_1(t), \ldots, y_n(t)) \), \( X \) is an \( n \times k \) design matrix with ith row \( x_i \), and \( \beta(t) \) is a \( k \times 1 \) vector of unknown parameters. The errors \( \epsilon_1(t), \ldots, \epsilon_n(t) \) are assumed to be i.i.d. with mean zero and variance \( \sigma^2(t) \). The model (1) is a functional linear model, which takes each response variable \( y_i \) as a function and the corresponding covariates \( x_i \) as a vector. In many applications, our interest is in testing that the \( k \)th element of \( \beta \) is equal to zero everywhere, i.e.,

\[
H_0 : \beta_k(t) = 0, \quad \forall t \in T.
\]

In practice, we can only observe \( y \) in every voxel \( v \in V \). Here, \( V \) is a collection of voxels, e.g., a whole brain or a region of interest. For a fixed voxel \( v \in V \), the null hypothesis is

\[
H_0 : \beta_k(v) = 0.
\]

The least squares (LS) estimator of \( \beta(v) \) is given by \( \hat{\beta}(v) = (X^T X)^{-1} X^T y(v) \) and note that we can write \( \hat{\beta}(v) - \beta(v) = (X^T X)^{-1} X^T \epsilon(v) \). To test the hypothesis in (3), we can choose the test statistic to be \( r_k \hat{\beta}_k(v) \), where \( \hat{\beta}_k(v) \) is the \( k \)th element of \( \hat{\beta}(v) \) and \( r_k \) is the \( k \)th diagnosed element of \( (X^T X)^{1/2} \). Under the null hypothesis (3), the test statistic may be expressed as

\[
S_n(v) = r_k \hat{\beta}_k(v) = \sum_{i=1}^{n} b_{n,i} \epsilon_i(v),
\]

where \( b_{n,i} = r_k (X^T X)^{-1} X_{k,i} \). The dependence of \( b_{n,i} \) on \( k \) is suppressed for clarity of notation.

To test the hypothesis (2), it is natural to construct a test statistic by constructing a functional of the test statistics for individual voxels. Let \( F \) be any continuous functional and \( \{S_n(v), v \in V\} \) be a collection of test statistics for all voxels, which
we will express simply as $[S_n(v)]$. Then the test statistic for (2) can be constructed using $F([S_n(v)])$, expressed simply as $F_n$.

One typical example for $F_n$ is the maximum of absolute values, i.e., $F_n = \max_{v \in V} |S_n(v)|$.

Our purpose is to get the limiting distribution of $F_n$ and thus compute an approximate $p$-value for testing (2). In imaging studies, we typically assume some smoothness conditions on the error processes $\{\epsilon_i(t), t \in T, i = 1, \ldots, n\}$, so it is natural to consider the discrete process $\{\sum_{i=1}^n b_{n,i}\epsilon_i(v), v \in V\}$ in (4) as an interpolating process

$$ X_n^*(t) \equiv \sum_{i=1}^n b_{n,i}\epsilon_i(s_n(t)), \quad (5) $$

where $s_n(t)$ is the center of the voxel that is closest to $t$. The dependence of $s_n(t)$ on $n$ is for convenience, to derive the asymptotic properties for $X_n^*(t)$. The asymptotic theory that we derive for $X_n^*(t)$ is related to the so-called “infill asymptotics” in spatial statistics; i.e., the field $T$ is fixed and the maximum diameter for all voxels approaches zero (the resolution approaches perfection) as $n$ approaches infinity. Since we know that for any fixed voxel $v \in V$, the distribution of the test statistic in (4) can be approximated by a normal distribution, it is of interest to know the conditions that will ensure that the interpolating processes $X_n^*(t)$ can be approximated by a Gaussian process $X$ indexed by $T$. If $X_n^*(t)$ converges to $X$ in probability, then by the continuous mapping theorem the distribution of test statistics $F_n = \max_{v \in V} |S_n(v)| = \max_{v \in T} [X_n^*(t)]$ approaches that of $F(X) \equiv \sup_{t \in T} X(t)$. Therefore, if we can characterize the distribution of $F(X)$, we can compute the approximate $p$-value for an observed $F_n$.

However, the challenge is that the distribution of $F(X)$ is usually very difficult to characterize because of the complicated spatial correlation structure in $\{\epsilon_i(t), i = 1, \ldots, n, t \in T\}$ inherent in brain imaging. To avoid making parametric assumptions about the spatial correlation structure, we follow Zhu et al. [5] and take a multiplier bootstrap approach to estimate its distribution. The bootstrap procedure is as follows:

(B1) Estimate $\hat{\beta}(v)$ for each voxel $v$ and calculate the test statistic $F_n$.

(B2) Resampling step:

(B2.1) For each voxel $v$, calculate the LS estimator $\hat{\beta}(v)$ under the null hypothesis (i.e. its $k$th element is zero).

(B2.2) For each bootstrap sample $b = 1, \ldots, B$, randomly generate multipliers $c_1(b), \ldots, c_n(b)$ from a distribution with mean zero and variance one.

(B2.3) For each bootstrap sample $b$ and each voxel $v$, calculate bootstrap response variables $y^b_i(v) = x_i\hat{\beta}(v) + c_i(b)\bar{r}_i(v)$, for $i = 1, \ldots, n$, and $\bar{r}_i(v) = y_i(v) - x_i\hat{\beta}(v)$.

(B2.4) Let $\hat{\beta}^b(v)$ be the $k \times 1$ vector of the LS estimator for bootstrap sample $b$ and voxel $v$ and $\hat{\beta}^b_i(v)$ be its $k$th element.

For every bootstrap sample $b$, calculate $S^b_n(v) = r_k\hat{\beta}^b_i(v) = \sum_{i=1}^n b_{n,i}c_i(b)\bar{r}_i(v)$ for each voxel $v$ in $V$ and calculate the test statistics $F^b_n = \max_{v \in V} |S^b_n(v)|$.

(B3) Calculate $\frac{\#\{b \mid b, b = 1, \ldots, B\}}{B}$ as an approximate $p$-value of the test statistic $F_n$.

In order to evaluate this bootstrap approach, for the same reason as we mentioned above in this section, we consider the discrete processes $\{\sum_{i=1}^n b_{n,i}c_i(b)\bar{r}_i(v), v \in V\}$ in step (B2.4) as an interpolating stochastic process:

$$ \hat{X}_n^*(t) \equiv \sum_{i=1}^n b_{n,i}c_i(b)\bar{r}_i(s_n(t)). \quad (6) $$

The key step is proving that $\hat{X}_n^*(t)$ converges conditionally to the process $X$, the limiting process of $X_n^*(t)$. (Note: Throughout this paper, “converge conditionally” means conditional weak convergence in probability, which we will define right before Theorem 2 in Section 4.) Then we can prove that the distribution of $F^b_n = \max_{v \in V} |S^b_n(v)| = \max_{v \in T} \hat{X}_n^*(t)$ can estimate that of $F(X)$ conditionally (see details in Remark 3); hence the bootstrap approach asymptotically controls the familywise error rate.

From the discussion above, in order to estimate the distribution of $F_n$, it suffices to prove the following:

(a) $X_n^*$ converges to a Gaussian process $X$, and

(b) $\hat{X}_n^*$ converges conditionally to the same process $X$.

Furthermore, if the resolution of the brain image is good and the error processes are smooth enough, it can be proved that $X_n^*(t)$ in (a) can be approximated by the process $X_n(t)$, where

$$ X_n(t) \equiv \sum_{i=1}^n b_{n,i}\epsilon_i(t); \quad (7) $$

$\hat{X}_n^*(t)$ in (b) can be approximated by the process $\tilde{X}_n(t)$, where

$$ \tilde{X}_n(t) \equiv \sum_{i=1}^n c_i b_{n,i}\epsilon_i(t). \quad (8) $$
Both \( X_n(t) \) and \( \bar{X}_n(t) \) involve sums of independent but not identically distributed stochastic processes; this motivates us to study asymptotic properties of bootstraps of independent sums. In Section 4, we present a bootstrap theorem for sums of independent but not identically distributed stochastic processes with a smoothing condition which will be useful for imaging applications in Theorem 3. Although in our imaging studies \( T \) is a rectangle in \( \mathbb{R}^s \), in Theorem 2, we let \( T \) be any subset in a general semimetric space for other potential applications. Before presenting the bootstrap theorem, we first provide some background and state the functional central limit theorem that we use in this paper, the Jain–Marcus theorem.

3. Preliminaries and functional central limit theorems

We are interested in estimating the limiting distribution of sums of the form \( \sum_{i=1}^n Z_{n,i}(\omega, t) \), where the real-valued stochastic processes \( \{Z_{n,i}(\omega, t), t \in T, 1 \leq i \leq n\} \), for \( n \geq 1 \), are independent within rows on the probability space \( (\Omega, \mathcal{A}, \mathbb{I}) \), indexed by a common semimetric space \( (T, \rho) \). For the computation of outer expectations, independence will always be understood to imply that the projections of the probability spaces involved for fixed \( n \) are products of the probability spaces corresponding to each independent process. We assume the usual pointwise measurability of these stochastic processes, i.e., \( Z_{n,i}(\cdot, t) \) is measurable for each \( t \in T, n \geq 1, \) and \( 1 \leq i \leq n \). To deal with the non-measurability issue, we denote as \( \mathbb{E}^* \) and \( \mathbb{P}^* \) the outer expectation and outer probability, respectively, consistently with the notation in van der Vaart and Wellner [7] (hereafter abbreviated as VW). A stochastic process \( X \) is called a Gaussian process if each finite-dimensional marginal \( (X(t_1), \ldots, X(t_s)) \) has a multivariate normal distribution on Euclidean space. Let \( \mathbb{F}^\infty(T) \) denote the space of all uniformly bounded, real functions on \( T \) with the uniform metric defined by the supremum norm \( \| f \| \equiv \sup_{t \in T} | f(t) | \). Sometimes we need to use the Euclidean norm, so to avoid confusion with notation, unless otherwise specified, \( \| \cdot \| \) means the Euclidean norm. The stochastic processes \( X_n \) converges weakly to a Borel process \( X \) in \( \mathbb{F}^\infty \) if

\[
E^* f(X_n) \to \int X f(x) \text{ for every continuous and bounded function } f \text{ defined in } \mathbb{F}^\infty.
\]

A process \( X \) is tight in \( \mathbb{F}^\infty(T) \) if its corresponding probability measure \( L \) is tight in \( \mathbb{F}^\infty(T) \), i.e., for any \( \epsilon > 0 \) there exists a compact set \( K \) in \( \mathbb{F}^\infty(T) \) such that \( L(K) \geq 1 - \epsilon \). Also, for any semimetric \( v \) on \( T \), let

\[
U_v \equiv \{ z \in \mathbb{F}^\infty(T) : z \text{ is uniformly } v\text{-continuous} \}.
\]

The covering number \( N(\epsilon, T, \rho) \) is the minimal number of balls \( \{ t : \rho(s, t) < \epsilon \} \) of radius \( \epsilon \) needed to cover the set \( T \).

We state the Jain–Marcus functional central limit theorem (see Theorem 2.11.13 in VW) for the stochastic process

\[
X_n(\omega, t) \equiv \sum_{i=1}^n [Z_{n,i}(\omega, t) - EZ_{n,i}(t)].
\]

The \( EZ_{n,i}(t) \) can be absorbed into the process \( Z_{n,i} \), from now on without loss of generality we assume that \( Z_{n,i}(\omega, t) \) are mean zero processes for \( n \geq 1 \) and \( 1 \leq i \leq n \). In our imaging application, \( \{Z_{n,i}\} \) is the processes \( \{b_{n,i}(t), t \in T \} \) defined in (7) and the Jain–Marcus theorem will be used to prove (a) in Section 2.

**Theorem 1** (Jain–Marcus Theorem). Suppose the stochastic processes from the triangular array \( \{Z_{n,i}(w, t), t \in T, i = 1, \ldots, n\}, \) \( n \geq 1, \) are mean zero and independent within rows such that for almost every \( \omega \in \Omega \) satisfy:

(A) \( |Z_{n,i}(\omega, s) - Z_{n,i}(\omega, t)| \leq M_{n,i}(\omega) \rho(s, t) \) for every \( s, t, \) and for some independent random variables \( M_{n,1}, \ldots, M_{n,n} \) and a semimetric \( \rho \) such that

\[
\int_0^\infty \sqrt{\log N(\epsilon, T, \rho)} d\epsilon < \infty,
\]

\[
\sum_{i=1}^n EM_{n,i}^2 = O(1);
\]

(B) (Lindeberg condition for norms) \( \sum_{i=1}^n E \| Z_{n,i} \|^2 \mathbb{I}_{\{ \| Z_{n,i} \|_T > \eta \}} \to 0, \) for every \( \eta > 0 \), where \( \{A\} \) is the indicator of \( A \);

(C) \( H(s, t) = \lim_{n \to \infty} EX_n(s)X_n(t) \) exists for every \( s, t \in T \).

Then \( X_n \) converges weakly in \( \mathbb{F}^\infty(T) \) to a tight Gaussian process \( X \) with covariance function \( H(s, t) \).

**Remark 1.** The inequality (9) implies that \( T \) is totally bounded with respect to \( \rho \). The \( Z_{n,i} \) processes are all measurable because their sample paths are uniformly continuous almost surely and the index set \( T \) is totally bounded. Therefore, no outer expectation is needed in condition (B). Condition (A) is the smoothness condition for the trajectories of the stochastic processes. Since \( \rho \) can be any metric, it is convenient to set up suitable smoothness conditions for many applications. For example, in Euclidean space if we let \( \rho(s, t) = \| s - t \|^\alpha, \) for \( \alpha > 0 \), then condition (A) represents that the processes satisfy the Hölder condition of order \( \alpha \). The inequality (9) in condition (A) represents the entropy condition which characterizes the size of the index set \( T \). Condition (B), the Lindeberg condition for norms, implies the Lindeberg condition for marginals. Marginal weak convergence to a Gaussian process follows from the Lindeberg condition for marginals and condition (C) that the covariance function converges. The proof can be found in Theorem 2.11.13 of VW.
4. Main bootstrap results

To state the conditional bootstrap theorem, we need notation for conditional expectation and an equivalent definition for weak convergence of stochastic processes. In the following, let multipliers \( \{c_i\} \) be defined on a probability space \( \{\Omega_c, \Lambda_c, \Pi_c\} \) and \( E_c \) denote taking expectation over \( \{c_i\} \) conditional on the data \( \{Z_{n,i}\} \). Also, for a metric space \( (D, d) \) let \( BL_1(D) \) be the space of real-valued functions on \( D \) with Lipschitz norm bounded by 1, i.e., for any \( f \in BL_1(D) \), \( \sup_{x \in D} |f(x)| \leq 1 \) and \( |f(x) - f(y)| \leq d(x, y) \) for all \( x, y \in D \). In our imaging application, \( D \) is the rectangle \( T \) and \( \{Z_{n,i}\} \) is again the processes \( \{\epsilon_{ni}(t), t \in T\} \) defined in (7). Our interest is in the conditional weak convergence of the multiplier processes:

\[
\tilde{X}_n(\omega, t) \equiv \sum_{i=1}^{n} c_i Z_{n,i}(\omega, t),
\]

defined on the product probability space \( \{\Omega \times \Omega_c, \Lambda \times \Lambda_c, \Pi \times \Pi_c\} \). This conditional weak convergence will be used to prove (b) in Section 2.

As described in the discussion following Theorem 1.12.2 in VW and the continuous mapping theorem. Therefore, this version of the bootstrap theorem is sufficient for our purposes.

**Theorem 2** (Main Bootstrap Theorem). Suppose the stochastic processes in the triangular array \( \{Z_{n,i}(\omega, t), t \in T, i = 1, \ldots, n\} \), \( n \geq 1 \), are mean zero and independent within rows and satisfy conditions (A), (B), and (C) in Theorem 1 as well as the following two additional conditions:

(D) \( \lim_{n \to \infty} \sum_{i=1}^{n} E\|Z_{n,i}\|_T < \infty \).

(E) The multiplier random variables \( c_i, i = 1, \ldots, n \), are i.i.d. with mean zero and variance one and independent of the \( Z_{n,i} \) processes and the \( M_{n,i} \) random variables.

Then

\[
\sup_{h \in BL_1(\\infty(T))} |E_h(\tilde{X}_n(\omega, \cdot)) - E_h(X)| \to 0
\]

in outer probability.

Proofs of this result and **Theorem 3** (to be presented in Section 5) are given in the Appendix.

**Remark 2.** In terminology just introduced, **Theorem 2** states that \( \tilde{X}_n \) converges conditionally to \( X \). The proof of this theorem is given in Appendix A.1. Van der Vaart and Wellner call this “conditional weak convergence” in probability. **Theorem 2** implies that, conditional on the data \( Z_{n,i} \), the multiplier process characterizes the distribution of sample paths of \( X \). It has many useful applications; for example, for any continuous functional \( F, P(F(\tilde{X}_n) \leq a(\omega)) \to P(F(\tilde{X}_n) \leq a(\omega)) \) in outer probability for every continuity point \( a \in \mathbb{R} \) of \( F(\cdot) \), i.e., \( P(F(\tilde{X}_n) \leq a) = P(F(X) < a) \), which follows from Lemma 1.9.2 and Theorem 1.12.2 in VW and the continuous mapping theorem. Therefore, this version of the bootstrap theorem is sufficient for our purposes.

5. Main results for application to functional data

In this section we consider the application of **Theorem 2** in application to imaging studies. First, we specify some sufficient conditions that will ensure that the bootstrap approach is appropriate for brain imaging data. Some comments on the conditions are given in the remark below the statement of the theorem. The proof of this theorem is given in Appendix A.2.

**Theorem 3.** Let \( T \) be a rectangle in \( \mathbb{R}^g \) and \( \{\epsilon_i, i = 1, \ldots, n, t \in T\} \) be i.i.d. stochastic processes with mean zero. Let \( X_n(\omega, t) \) be \( \sum_{i=1}^{n} b_{n,i} \epsilon_i(\omega, s_{ni}(t)) \) as in (5), where \( s_{ni}(t) = \arg \min_{\epsilon \in \{t_1, \ldots, t_{p(n)}\}} \|t - \epsilon\|_2 \), where \( \{t_1, \ldots, t_{p(n)}\} \) are the centers of the corresponding voxels. Let \( a \) be a positive number. Assume the following conditions are satisfied.

1. \( B_n = \max_{1 \leq i \leq n} |b_{n,i}| \to 0. \)
2. \( \lim_{n \to \infty} \sum_{i=1}^{n} b_{n,i}^2 < \infty. \)
3. For almost all \( \omega \in \Omega \), \( |\epsilon_i(\omega, s) - \epsilon_i(\omega, t)| \leq M_i(\omega) \|t - s\|^a \) for every \( s, t \) in \( T \) and for some i.i.d. random variables \( M_1, \ldots, M_n \) such that \( EM_i^2 < \infty. \)
4. The total number of voxels \( p(n) \) is \( \prod_{i=1}^{g} n_{i} \) and \( \{n_1, \ldots, n_g\} \) are all \( O(n^{1/a}) \). All voxels have the same size.
5. \( E\|\epsilon_i\|_T^2 < \infty \) for every \( i = 1, \ldots, n. \)
(16) The multiplier random variables \( \{c_i, i = 1, \ldots, n\} \) are i.i.d. and independent of the \( \epsilon_i \) processes and the \( M_i \) random variables. In addition, \( c_i = 1 \) with probability \( 1/2 \) and \( c_i = -1 \) with probability \( 1/2 \) for any \( i = 1, \ldots, n \).

(17) Let \( X_{i,k} \) be the submatrix of \( X \) which contains the first \( k - 1 \) columns of \( X \). Also let \( x_{i,-k} \) be the subvector of \( x_i \) which contains the first \( k - 1 \) elements of \( x_i \). \( \lim_{n \to \infty} \frac{X_{i,k}^T X_{i,k}}{n} = D \) where \( D \) is a \((k - 1) \times (k - 1)\) positive definite matrix.

\[
\sup_{n \geq 1} \sup_{i = 1, \ldots, n} \sup_{j = 1, \ldots, k - 1} |X_{ij}| \leq L < \infty.
\]

Then:

(i) \( X_n^\ast \) converges weakly to \( X \);

(ii) \( \sup_{h \in \mathbb{R}_1} \sup_{t \in \mathbb{R}_T} \mathbb{E}_c[h(X_n^\ast) - Eh(X)] \to 0 \) in outer probability.

**Remark 3.** In the above theorem, we define \( s_n(t) \) as the center of the voxel that contains \( t \), but actually \( s_n(t) \) can be defined as any point in the voxel that contains \( t \). From the above results, using the argument in **Remark 2**, \( P(F(\hat{X}_n^\ast) \leq a(\omega) \to P(F(X) \leq a) \) in outer probability for every continuity point \( a \in \mathbb{R} \) of \( F(X) \), i.e., \( P(F(X) \leq a) = P(F(X) < a) \); hence the bootstrap approach is appropriate in our imaging setting. Conditions I1 and I2 are common sufficient conditions for checking the Linderberg condition in linear regression models. Condition I3 is the smoothness condition (Hölder condition of order \( \alpha \) for processes of images. For example, in imaging studies, it is common to assume the processes to be random fields resulting from convolving independent random variables with a smoothing filter. In this case, it certainly satisfies the Hölder condition of any positive order. Even for Brownian motion, whose sample paths are not smooth, it also satisfies the Hölder condition of any order \( \alpha \) with \( 0 < \alpha < 1/2 \). Therefore, condition I3 includes a lot of common examples. Condition I4 is concerned with the resolution of the images, associated with condition I3. We can see that for smoother processes, less resolution is needed to prove asymptotics. In condition I6, we specify the distribution of the multipliers \( c_1, \ldots, c_n \) but the theorem can be extended to allow the \( c_i \)'s to easily be any independent, identically distributed, symmetric, and bounded random variables. Condition I8 is about uniform boundedness of the covariate matrix \( X \). We note that conditions I7 and I8 only involve the first \( k - 1 \) columns of the covariate matrix \( X \) because we want to derive asymptotics under the null hypothesis, i.e., \( \beta_k(t) = 0 \).

6. Discussion

Our bootstrap theorem differs from that in [9] in that the manageability condition in his paper is replaced with a smoothness and entropy condition (condition (A) in Theorem 2). Although the manageability condition is very general, the smoothness condition is more appropriate for application to functional data, such as those in spatial statistics or one-dimensional curve data because it directly characterizes the properties of sample paths of the processes. In addition, it is useful to use a smoothness condition to prove some infill asymptotics in functional data analysis because in practice, we can only observe the discrete processes instead of the continuous processes.

The hypothesis that we choose in this paper (2) is one specific example of the more general hypotheses \( a^T \beta(t) = 0, \forall t \in T \), where \( a \) is a \( k \times 1 \) constant vector. We choose this hypothesis because it is perhaps most commonly used in practice. Using Theorem 2 and the structure of the proof for Theorem 3, one can derive sufficient conditions for more general hypotheses as well.

**Acknowledgments**

The authors are grateful for helpful comments by Prof. Ian McKeague, Prof. Hongtu Zhu, referees and the Editor, which led to great improvement of the presentation.

**Appendix**

A.1. **Proof of Theorem 2**

In order to prove the main bootstrap theorem, we need to state and prove several lemmas. The approach that we take is to first prove an unconditional multiplier central limit theorem (**Lemma 1**), then a conditional central limit theorem for random vectors (**Lemma 2**), which can be considered as stochastic processes with finite index set, and then extend these to prove the conditional central limit theorem for stochastic processes with general index set which will also need two additional results (**Lemmas 3 and 4**).

First, we prove an unconditional multiplier central limit theorem (**Lemma 1**), i.e., that the multiplier processes \( \tilde{X}_n \), defined in (10), converge weakly to the same process \( X \) in \( l^\infty(T) \).

**Lemma 1.** Suppose the stochastic processes from the triangular array \( \{Z_{n,i}(w, t), t \in T, i = 1, \ldots, n\}, n \geq 1 \), are mean zero and independent within rows. If conditions (A), (B), and (C) in Theorem 1 and Conditions (D) and (E) in Theorem 2 hold, then the unconditional multiplier process \( \tilde{X}_n \) in (10) converges weakly in \( l^\infty(T) \) to the same limiting process \( X \) as in Theorem 1.

**Proof.** In order to prove this theorem, we need to check conditions (A), (B), and (C) in Theorem 1 for \( \tilde{X}_n \). For condition (A) in Theorem 1, \( |c_i(Z_{n,i}(\omega, s) - Z_{n,i}(\omega, t))| \leq |c_i|M_{n,i}(\omega)\rho(s, t) \) for \( s, t \) almost surely in the product probability space such that
For condition (B) in Theorem 1, for any $\eta > 0$, we have for any $G > 0$ that
\[ \sum_{i=1}^{n} E c_i^2 \mathbb{P}(\|Z_{n,i}\|_T > \eta) \leq E \sum_{i=1}^{n} E \|Z_{n,i}\|_T^2 + G^2 \sum_{i=1}^{n} E \|Z_{n,i}\|_T^2 \mathbb{P}(\|Z_{n,i}\|_T > \eta/G). \]

Since the second term on the right-hand side goes to zero and the first term can be made arbitrarily small by choice of $G$, condition (B) is also satisfied.

For condition (C) in Theorem 1,
\[ \bar{H}(s, t) = \lim_{n \to \infty} E \bar{X}_n(s) \bar{X}_n(t) = \lim_{n \to \infty} \sum_{i=1}^{n} E c_i^2 E X_n(s) X_n(t) = H(s, t). \]

Therefore, $\bar{X}_n$ converges weakly to the same limiting tight Gaussian process $X$ in Theorem 1. ■

Second, we state a conditional central limit theorem for random vectors proved by [9].

**Lemma 2** ([9]). Let $\{Y_{n,i}, i = 1, \ldots, n, n \geq 1\}$ be a triangular array of mean zero real random vectors in $\mathbb{R}^g$, independent within rows; and let $\{c_i, i \leq 1\}$ be i.i.d. random variables with mean zero and variance 1 that are independent of $\{Y_{n,i}\}$. Suppose also that:

(a) $\lim_{n \to \infty} \sum_{i=1}^{n} E Y_{n,i} Y_{n,i}^T = V_0$ is a positive definite matrix;
(b) for every $\eta > 0$, $\limsup_{n \to \infty} \sum_{i=1}^{n} E \|Y_{n,i}\|^2 \mathbb{P}(\|Y_{n,i}\| > \eta) = 0$.

Then:
(i) $\sum_{i=1}^{n} Y_{n,i}$ converges weakly to $Y_0 \sim N_d(0, V_0)$;
(ii) $\sup_{h \in BL_1(\mathbb{R}^g)} |E[h(\sum_{i=1}^{n} c_i Y_{n,i}) − Eh(Y_0)]| \to 0$ in probability, as $n \to \infty$, where the metric associated with $BL_1(\mathbb{R}^g)$ is the uniform metric.

Note: The uniform metric for $\mathbb{R}^g$ mentioned here is a metric $d$ defined by $d(s, t) = \max_{i=1,\ldots,g} |s_i − t_i|$, where $s_i$ and $t_i$ are the $i$th coordinates of $s$ and $t$ in $\mathbb{R}^g$, respectively.

We now state the Hoffmann–Jorgensen inequalities (Lemma 3) which will be needed to prove Lemma 4.

**Lemma 3** (Hoffmann–Jorgensen Inequalities for Moments). Suppose that $Z_1, \ldots, Z_n$ are independent stochastic mean zero processes indexed by $T$ and let $X_n = \sum_{i=1}^{n} Z_i$. Then there exist constant $K$ and $\mu$ such that
\[ E^* \|X_n\|^2 \leq K E^* \max_{\|u\| \leq 1} \|Z_u\|^2 + G_n^{-1}(\mu^2), \]
where $G_n^{-1}$ is the quantile function of the random variable $\|X_n\|^2$.

In order to extend conditional central limit theorem for random vectors to conditional FCLT, we need to prove asymptotic equicontinuity in the first moment for $\bar{X}_n$ in (10), which is stated in Lemma 4.

**Lemma 4.** Suppose the stochastic processes from the triangular array $\{Z_{n,i}(w, t), t \in T, i = 1, \ldots, n\}, n \geq 1$, are mean zero and independent within rows. Let $X_n \equiv \sum_{i=1}^{n} Z_{n,i}$ and $X$ be a tight Gaussian process. Also let $\rho_2(s, t) = (E(X(s) − X(t))^2)^{1/2}$. Then the statement (i) $\Rightarrow$ (ii):
(i) $X_n$ converges weakly to $X$ in $L^2(T)$;
(ii) (equicontinuity in probability) $\sup_{s,t \in T : \rho_2(s, t) \leq \delta_n} |X_n(\cdot, s) − X_n(\cdot, t)| \to 0$ in outer probability for every $\delta_n \downarrow 0$.

Furthermore, if for every $\eta > 0$
\[ \lim_{n \to \infty} \sum_{i=1}^{n} E^* \|Z_{n,i}\|_T^2 \mathbb{P}(\|Z_{n,i}\|_T > \eta) = 0, \]
then (ii) $\Rightarrow$ (iii).

(iii) (equicontinuity in first moment) $E^* \sup_{s,t \in T : \rho_2(s, t) \leq \delta_n} |X_n(\cdot, s) − X_n(\cdot, t)| \to 0$ for every $\delta_n \downarrow 0$.

**Proof.** First define $T_{\delta_n} = \{(s, t), \rho_2(s, t) \leq \delta_n\}$ and let $\|X\|_{T_{\delta_n}} = \sup_{s,t \in T : \rho_2(s, t) \leq \delta_n} |X(\cdot, s) − X(\cdot, t)|$. To simplify notation, in this proof we abbreviate $\|\cdot\|_{T_{\delta_n}}$ as $\|\cdot\|_{\delta_n}$. To prove (i) $\Rightarrow$ (ii), we note that since $X$ is a tight Gaussian process, $(T, \rho_2)$ is totally bounded and almost all paths $t \mapsto X(t, \omega)$ are uniformly $\rho_2$-continuous. Therefore, (1) $\Rightarrow$ (2) by the general equicontinuous theorem of weak convergence for stochastic processes (Theorem 1.5.7 and Addendum 1.5.8 in VW).
To prove (ii)⇒(iii), for any \( \eta > 0 \),
\[
E^* \max_{1 \leq n} \|Z_n\|_{T_n}^2 = E \max_{1 \leq n} \|Z_n\|_{T_n}^2 \geq \eta^2 + \sum_{i=1}^{n} \int_{\eta}^{\infty} P(\|Z_n\|_{T_n}^2 > t)2tdt \\
\leq \eta^2 + \sum_{i=1}^{n} \int_{\eta}^{\infty} P(\|Z_n\|_{T_n}^2 > t/2)2tdt \\
\leq \eta^2 + \sum_{i=1}^{n} \int_{\eta}^{\infty} P(\|Z_n\|_{T_n}^2 > t)8tdt \\
= \eta^2 + \sum_{i=1}^{n} \int_{\eta}^{\infty} P \left( \|Z_n\|_{T_n}^2 \geq \frac{\eta}{2} > t \right) 8tdt \\
\leq \eta^2 + 4 \sum_{i=1}^{n} E \left( \|Z_n\|_{T_n}^2 \right) \geq \frac{\eta}{2} \to 0
\]
(11)
as \( n \to \infty \) and \( \eta \downarrow 0 \).

The inequality (11) follows from Problem 2.3.5 in VW. The quantile function \( G^{-1}\) of \( \|Z_n\|_{T_n}^2 \) converges to zero pointwise because \( \|Z_n\|_{T_n} \) converges to zero in outer probability (from (ii)). By Lemma 3, we prove (iii). ■

Note: Both Lemmas 3 and 4 are for general mean zero stochastic processes (not just continuous processes); therefore, the outer expectations are needed.

**Proof of Theorem 2.** Define \( \rho_2(s, t) \equiv \langle E|X(s) - X(t)\rangle^2 \rangle^2 \). From the discussion followed by Example 1.5.10 in VW, we know that a Gaussian process \( X \) in \( F^\infty(T) \) is tight if and only if \( (T, \rho_2) \) is totally bounded and almost all paths \( t \mapsto X(t, \omega) \) are uniformly \( \rho_2 \)-continuous. Therefore, \( \forall \epsilon > 0, \exists \delta > 0 \) such that we can find a finite \( \delta \)-net (w.r.t. \( \rho_2 \)) such that \( \forall t \in T, |X(\omega, M_t(t)) - X(\omega, t)| \leq \epsilon \), where \( M_t(t) \) is the closest element in the finite \( \delta \)-net assigned to each \( t \in T \).

\[
\sup_{h \in BL_1(F^\infty(T))} |E_i h(\tilde{X}(\omega, \cdot)) - Eh(X)| \leq \sup_{h \in BL_1(F^\infty(T))} |E_i h(\tilde{X}_n(\omega, \cdot)) - Eh(\tilde{X}_n(\omega, M_t(\cdot)))| \\
+ \sup_{h \in BL_1(F^\infty(T))} |E_i h(\tilde{X}_n(\omega, M_t(\cdot))) - Eh(M_t(\cdot)))| + \sup_{h \in BL_1(F^\infty(T))} |E h(X(M_t(\cdot))) - Eh(X)|.
\]
(12)

When \( \delta \to 0 \), the third term on the right-hand side of (12) goes to zero by the bounded convergence theorem. (The reason that the bounded convergence theorem can be applied is that almost all sample paths of \( X \) are uniformly \( \rho_2 \)-continuous and \( h \in BL_1(F^\infty(T)) \)).

The second term on the right-hand side of (12) is
\[
\sup_{h \in BL_1(F^\infty(T))} |E_i h(\tilde{X}_n(\omega, M_t(\cdot))) - Eh(X(\omega, M_t(\cdot)))| \to 0
\]
in probability by Lemma 2. The first term on the right-hand side of (12) is
\[
\sup_{h \in BL_1(F^\infty(T))} |E_i h(\tilde{X}_n(\omega, \cdot)) - E_i h(\tilde{X}_n(\omega, M_t(\cdot)))| \leq \sup_{h \in BL_1(F^\infty(T))} E_i |h(\tilde{X}_n(\omega, \cdot)) - h(\tilde{X}_n(\omega, M_t(\cdot)))| \\
\leq E_i \left( \sup_{s, t \in T; t_2(s, t) \leq \delta} |\tilde{X}_n(\omega, s) - \tilde{X}_n(\omega, t)| \right)^*.
\]
(13)

Thus, the outer expectation of the left-hand side of (13) is bounded above by
\[
E^* \sup_{s, t \in T; t_2(s, t) \leq \delta} |\tilde{X}_n(\omega, s) - \tilde{X}_n(\omega, t)|.
\]

By Lemma 1 (unconditional multiplier weak convergence), \( \tilde{X}_n \) converges weakly to a tight Gaussian process \( X \). Therefore, by Lemma 4, \( E^* \sup_{s, t \in T; t_2(s, t) \leq \delta} |\tilde{X}_n(\omega, s) - \tilde{X}_n(\omega, t)| \) converges to zero as \( n \to \infty \) and \( \delta \downarrow 0 \). Hence, we can conclude that the first term converges to zero in outer probability. ■

**Remark 4.** The proof of Theorem 2 is based on the structure of Theorem 2 in [9].

**A.2. Proof of Theorem 3**

Recall the processes \( X_n^*(t) = \sum_{j=1}^{n} b_{n,j} i(s_j(t), \tilde{X}_n^*(t)) \), \( \tilde{X}_n^*(t) = \sum_{i=1}^{n} b_{n,i} \beta s_i(t), X_n(t) = \sum_{i=1}^{n} b_{n,i} (c_i(t), t), \) and \( \tilde{X}_n = \sum_{i=1}^{n} c_i b_{n,i} \), defined in equations (5), (6), (7), and (8), respectively. We note that \( s_n(t) = \arg \min_{s \in [t_1, ..., t_p]} ||t - s|| \) where \( \{t_1, ..., t_p\} \)
are the centers of the corresponding voxels. Our purpose is to prove that the process \( X_n^* \) converges weakly to a tight Gaussian process \( X \) ((i) in Theorem 3) and the process \( \tilde{X}_n^* \) converges conditionally to the same process \( X \) ((ii) in Theorem 3). For the purpose of this proof, we also let

\[
\tilde{X}_n(\omega, t) = \sum_{i=1}^{n} c_i b_{ni} \tilde{r}_i(\omega, t)
\]

and

\[
\tilde{X}_n^*(\omega, t) = \sum_{i=1}^{n} c_i b_{ni} \varepsilon_i(\omega, s_n(t)).
\]

We have that

\[
\sup_{h \in BL_1(I_T)} |E_h(\tilde{X}_n) - Eh(X)| \leq \sup_{h \in BL_1(I_T)} |E_h(\tilde{X}_n^*) - E_h(\tilde{X}_n)| + \sup_{h \in BL_1(I_T)} |E_h(\tilde{X}_n^*) - Eh(X)|
\]

\[
\leq E_\varepsilon \|\tilde{X}_n^* - \tilde{X}_n\|_T + \sup_{h \in BL_1(I_T)} |E_h(\tilde{X}_n^*) - Eh(X)|
\]

\[
\leq E_\varepsilon \|\tilde{X}_n - \tilde{X}_0\|_T + \sup_{h \in BL_1(I_T)} |E_h(\tilde{X}_n^*) - Eh(X)|.
\]

So in order to prove (ii) in Theorem 3, it suffices to prove that the last two terms on the right-hand side of inequality (16) both converge to 0 in outer probability. We will first prove that \( X_n^* \) converges to \( X \) weakly ((i) in Theorem 3) and \( \tilde{X}_n^* \) converges conditionally to the same process \( X \), i.e., the last term of the right-hand side in inequality (16) converges to zero in outer probability, in the next lemma. After the next lemma, we will prove that \( E_\varepsilon \|\tilde{X}_n - \tilde{X}_0\|_T \), the second to last term of the right-hand side in inequality (16), converges to 0 in outer probability, which completes the proof of Theorem 3.

**Lemma 5.** Let \( X_n^*, \tilde{X}_n^*, \tilde{X}_n^*, \tilde{X}_n^* \) be the processes defined in (5), (7), (8) and (15), respectively. Assume that conditions (11)–(15) in Theorem 3 are satisfied, and suppose also that:

(i6') Multiplier random variables \( c_i \), \( i = 1, \ldots, n \), are i.i.d. with mean zero and variance one and independent of the processes \( \varepsilon \).

Then:

(i) \( X_n^* - X_n \to 0 \) in probability.

(ii) \( X_n^* \) converges weakly to a tight Gaussian process \( X \).

(iii) \( \sup_{h \in \mathbb{B}L_1(I_T)} |E_h(\tilde{X}_n) - Eh(X)| \to 0 \) in outer probability.

(iv) \( \sup_{h \in \mathbb{B}L_1(I_T)} |E_h(\tilde{X}_n^*) - Eh(\tilde{X}_n^*)| \to 0 \) in outer probability.

Note: If (i), (ii), (iii), and (iv) hold, then \( X_n^* \) converges weakly to a tight Gaussian process \( X \) and \( \sup_{h \in \mathbb{B}L_1(I_T)} |E_h(\tilde{X}_n^*) - Eh(X)| \to 0 \) in outer probability.

**Proof.** To prove (i), it suffices to prove \( P(\|X_n^* - X_n^*\|_T > \epsilon) \to 0 \) as \( n \to \infty \). By conditions I3 and I4, we have that

\[
\sup_{t \in T} |X_n^* - X_n^*(\epsilon(t) - \epsilon(s_n(t)))| = \sup_{t \in T} \sum_{i=1}^{n} |b_{ni}| |\varepsilon_i(\epsilon(t) - \varepsilon(s_n(t)))|
\]

\[
\leq \sup_{t \in T} \sum_{i=1}^{n} |b_{ni}| \cdot |\varepsilon_i(\epsilon(t) - \varepsilon(s_n(t)))|
\]

\[
\leq \left( \sum_{i=1}^{n} |b_{ni}| M_i \right) \times \sup_{t \in T} \|t - s_n(t)\|^a
\]

\[
\leq \left( B_n \sum_{i=1}^{n} M_i \right) \times O \left( \frac{1}{n} \right).
\]

Therefore, using Chebyshev’s inequality and the above inequality, we show that

\[
P (\|X_n^* - X_n^*\|_T > \epsilon) \leq P \left( \sum_{i=1}^{n} M_i > \frac{\epsilon}{B_n \times O \left( \frac{1}{n} \right)} \right)
\]

\[
\leq \frac{E \left( \sum_{i=1}^{n} M_i \right) B_n O \left( \frac{1}{n} \right)}{\epsilon}
\]

\[
\leq B_n O(1) \to 0.
\]
To prove (ii), it suffices to check conditions (A), (B), and (C) for $X_n$ in Theorem 1. To prove that condition (A) is satisfied w.r.t the semimetric $\| \cdot \|_\alpha$ in Theorem 1, since $T$ is a rectangle in $\mathbb{R}^g$,
\[
\int_0^\infty \sqrt{\log N(\epsilon, T, \| \cdot \|_\alpha)} d\epsilon < \infty.
\]
By conditions I2 and I3, we have that
\[
\sum_{i=1}^n b_{n,i}(\epsilon_i(s) - \epsilon_i(t)) \leq \sum_{i=1}^n (|b_{n,i}|M_i) \cdot \|t - s\|_\alpha,
\]
where
\[
\sum_{i=1}^n E(|b_{n,i}|^2M_i)^2 = O(1).
\]
To check condition (B) in Theorem 1, for any $\eta > 0$
\[
\sum_{i=1}^n E\|X_n\|_T^2 \mathbb{1}\{\|X_n\|_T > \eta\} \leq \sum_{i=1}^n b_{n,i}^2E\|\epsilon_i(t)\|_T^2 \mathbb{1}\{|b_{n,i}|\|\epsilon_i(t)\|_T > \eta\}
\]
\[
\leq \left(\sum_{i=1}^n b_{n,i}^2\right) E\|\epsilon_i(t)\|^2 \mathbb{1}\{B_n \|\epsilon_i(t)\|_T > \eta\}
\]
\[
\to 0.
\]
Therefore, (11), (12), and (14) imply the Lindeberg condition for norms in Theorem 1.
To check condition (C) in Theorem 1,
\[
H(s, t) = EX_n(t)X_n(t)
\]
\[
= \sum_{i=1}^n b_{n,i}^2EX_i(t)\epsilon_i(s)
\]
\[
\to S^2E\epsilon(t)\epsilon(s),
\]
where $S^2 = \lim_{n \to \infty} \sum_{i=1}^n b_{n,i}^2 < \infty$.
To prove (iii), it suffices to check conditions (D) and (E) in Theorem 2 for $\tilde{X}_n$, which are obvious. Therefore, from Theorem 1 we have that $X_n$ converges weakly to a tight Gaussian process $X$; from Theorem 2 we have that $\tilde{X}_n$ converges conditionally to the same Gaussian process $X$.

The last step is to prove (iv), that $\sup_{h \in \mathcal{B}_1(\mathcal{L}^\infty(T))} |E_c(h(\tilde{X}_n)) - E_c(h(\tilde{X}^*_n))| \to 0$ in outer probability. It suffices to prove
\[
E_c\left(\sup_{h \in \mathcal{B}_1(\mathcal{L}^\infty(T))} |h(\tilde{X}_n) - h(\tilde{X}^*_n)|\right) \to 0 \text{ in outer probability.}
\]
\[
E_c \sup_{h \in \mathcal{B}_1(\mathcal{L}^\infty(T))} |h(\tilde{X}_n) - h(\tilde{X}^*_n)| \leq E_c\left(\left\|\tilde{X}_n - \tilde{X}^*_n\right\|_T\right)
\]
\[
= E_c\left(\left\|\sum_{i=1}^n c_i b_{n,i}(\epsilon_i(t) - \epsilon_i(s_n(t)))\right\|_T\right)
\]
\[
\leq E_c \sum_{i=1}^n |b_{n,i}| |c_i| \left\|\epsilon_i(t) - \epsilon_i(s_n(t))\right\|_T
\]
\[
\leq E_c B_n \sum_{i=1}^n |c_i| M_i \left\|\epsilon_i(t) - s_n(t)\right\|_T\alpha
\]
\[
= O\left(\frac{1}{n}\right) B_n \sum_{i=1}^n M_i E_c |c_i|
\]
\[
\to 0
\]
in probability, as $n \to \infty$, because
\[
E \frac{B_n \sum_{i=1}^n M_i}{n} \leq \sum_{i=1}^n \frac{B_n M_i}{n} \to 0. \quad \blacksquare
\]
So far, we have proved that the process $\hat{X}_n^* = \sum_{i=1}^n c_i b_{n,i}(S_n(t))$ converges conditionally to the process $X$. To prove (ii) in Theorem 3, by inequality (16), the only thing that remains is to prove that $E_c \|\hat{X}_n - \tilde{X}_n\|_T$ converges to 0 in outer probability. To do this, we need Hoeffding's inequality, which is stated in the next lemma.

**Lemma 6 (Hoeffding’s Inequality).** Let $Y_1, \ldots, Y_n$ be independent real-valued random variables with expectation zero. Suppose for all $i$, and for certain constants $r_i > 0$,

$$|Y_i| \leq r_i$$

with probability 1. Then for all $\delta > 0$,

$$P \left( \sum_{i=1}^n Y_i \geq \delta \right) \leq \exp \left( -\frac{\delta^2}{2 \sum_{i=1}^n r_i^2} \right).$$

**Proof of Theorem 3.** Before we start proving this theorem, we first define some notation. For any fixed positive integer $q$, we divide the rectangle $T$ equally into $2^{2q}$ sub-rectangles and denote the collection of their centers as $c_q$. Let $l_1, \ldots, l_s$ be the lengths of $T$ and $R = \max\{l_1, \ldots, l_s\}$. Also let $\pi_q(t) = \arg\min_{s \in q} \|t - s\|$ and let $N_q$ be $\#\{\pi_q(t), t \in T\}$. Then we have $N_q = 2^{2q}$:

$$\sup_{t \in T} \|t - \pi_q(t)\| \leq \frac{\sqrt{q}}{2} \times 2^{-q} R = 2^{-q} R_1,$$

where $R_1 = \frac{\sqrt{q}}{2} R$;

$$\sup_{t \in T} \|\pi_q(t) - \pi_{q-1}(t)\| \leq 2^{-q} R_1 + 2^{-q+1} R_1 = 3 \times 2^{-q} R_1.$$

We first note that, since

$$\hat{X}_n(\omega, t) = \sum_{i=1}^n c_i b_{n,i}(\omega, t) = \sum_{i=1}^n c_i b_{n,i}(y_i - x_i(\tilde{\beta}_n(t)))$$

and $\tilde{X}_n(\omega, t) = \sum_{i=1}^n c_i b_{n,i}(y_i - x_i(\tilde{\beta}(t)))$, we have that

$$\hat{X}_n(\omega, t) - \tilde{X}_n(\omega, t) = -\sum_{i=1}^n c_i b_{n,i} x_i(\tilde{\beta}(t) - \beta(t))$$

and

$$E_c \|\hat{X}_n - \tilde{X}_n\|_T = E_c \sup_{t \in T} \left| \sum_{i=1}^n c_i b_{n,i} x_i(\tilde{\beta}(t) - \beta(t)) \right|.$$

Therefore, for any fixed positive integer $Q$,

$$E_c \|\hat{X}_n - \tilde{X}_n\|_T \leq E_c \sup_{t \in T} \left| \sum_{i=1}^n c_i b_{n,i} x_i(\tilde{\beta}(t) - \beta(t) - (\tilde{\beta}(\pi_q(t)) - \beta(\pi_q(t)))) \right| + E_c \sup_{t \in [\pi_{q(t)}]} \left| \sum_{i=1}^n c_i b_{n,i} x_i(\tilde{\beta}(t) - \beta(t)) \right| \tag{17}$$

The second term on the right-hand side of inequality (17) converges to zero with probability 1 because $N_q$ is finite and since

$$E_c \left| \sum_{i=1}^n c_i b_{n,i} x_i(\tilde{\beta}(t) - \beta(t)) \right| \leq \left\{ E_c \left| \sum_{i=1}^n b_{n,i} x_i(\tilde{\beta}(t) - \beta(t)) \right|^2 \right\}^{1/2} \leq \left\{ \sum_{i=1}^n b_{n,i}^2 |x_i(\tilde{\beta}(t) - \beta(t))|^2 \right\}^{1/2} \rightarrow 0$$

in probability as $n \rightarrow \infty$ because $\sum_{i=1}^n b_{n,i}^2 < \infty$ and $\tilde{\beta}(t) - \beta(t) \rightarrow 0$ in probability $\forall t \in T$. Therefore, in order to prove Theorem 3, it suffices to prove the first term on the right-hand side of inequality (17) converges to zero in probability as $Q \rightarrow \infty$. 
For any fixed $\omega \in \Omega$, define
\[
d_{2,n,\omega}(s, t) = \left\{ \sum_{i=1}^{n} b_{n,i}^2 \left( x_i \left[ \tilde{\beta}_n(t) - \beta(t) - (\tilde{\beta}_n(s) - \beta(s)) \right]^2 \right) \right\}^{1/2}.
\]
Also let $e = (e_1, \ldots, e_n)$, $M(\omega) = (M_1(\omega), \ldots, M_n(\omega))^T$, and $\lambda_i(\omega) = |x_i - k(X^T_k X_k)^{-1}X^T_k|_{\|\cdot\|_\omega M(\omega)}$ and define $|e|_{\|\cdot\|_\omega} = ((|e_1|, \ldots, |e_n|)$. Under the null hypothesis (2),
\[
d_{2,n,\omega}(s, t) = \left\{ \sum_{i=1}^{n} b_{n,i}^2 \left( |x_i - k(X^T_k X_k)^{-1}X^T_k|_{\|\cdot\|_\omega M(\omega)} \|s - t\|_\omega \right)^2 \right\}^{1/2} = \left( \sum_{i=1}^{n} b_{n,i}^2 \lambda_i^2(\omega) \right)^{1/2} \|s - t\|_\omega.
\]
Using conditions 12, 17, and 18 and that $EM^2_1 < \infty$ and by applying the Strong Law of Large Numbers, we show that $\lim_{n \to \infty} \max_{1 \leq i \leq n} \lambda_i(\omega) \leq C(L, D, k, EM_{1}) \equiv C$ a.s.

Let $\Omega_n = \{ \omega \in \Omega : \max_{1 \leq i \leq n} \lambda_i(\omega) \leq C + 1 \}$. Define $C_1 = (C + 1) \times (\lim_{n \to \infty} \sum_{i=1}^{n} b_{n,i}^2)^{1/2}/2$. If $\omega \in \Omega_n$, then $d_{2,n,\omega}(s, t) \leq C_1 \|s - t\|^a$ and
\[
P_\epsilon \left( \sup_{t \in T} \left| \sum_{i=1}^{n} c_i b_{n,i} x_i [\tilde{\beta}(\pi_q(t)) - \beta(\pi_q(t)) - (\tilde{\beta}(\pi_{q-1}(t)) - \beta(\pi_{q-1}(t)))] \right| \geq \delta \right) \\
\leq N_q^2 \times \left( 2 \exp \left( -\frac{\delta^2}{2 \sup_{t \in T} d_{2,n,\omega}^2(\pi_q(t), \pi_{q-1}(t))} \right) \right) \quad \text{(by condition 16, Lemma 6, and symmetry)} \\
\leq 2N_q^2 \times \left( \exp \left( -\frac{\delta^2}{2C_1^2 \times 9^a \times 2^{-2q_0} \times R_{q_0}^2} \right) \right) \\
\leq 2 \exp \left( \log(N_q^2) - \frac{\delta^2}{C_2 \times 2^{-2q_0}} \right),
\]
where $C_2 = 2C_1^2 \times 9^a \times R_{q_0}^2$.

Let $\eta_q = C_3 \times 2^{-\frac{3}{2}q_0}$ and choose $C_3$ such that $\sum_{q=1}^{\infty} \eta_q \leq 1$. (The idea is to choose a proper positive number $\eta_q$ such that $\sum_{q=1}^{\infty} \eta_q \leq 1$.) By inequality (18), we have that
\[
P_\epsilon \left( \sup_{t \in T} \left| \sum_{i=1}^{n} c_i b_{n,i} x_i [\tilde{\beta}(t) - \beta(t) - (\tilde{\beta}(\pi_q(t)) - \beta(\pi_q(t)))] \right| \geq \delta \right) \\
\leq \sum_{q=Q+1}^{\infty} P_\epsilon \left( \sup_{t \in T} \left| \sum_{i=1}^{n} c_i b_{n,i} x_i [\tilde{\beta}(\pi_q(t)) - \beta(\pi_q(t)) - (\tilde{\beta}(\pi_{q-1}(t)) - \beta(\pi_{q-1}(t)))] \right| \geq \eta_q \delta \right) \\
\leq \sum_{q=Q+1}^{\infty} 2 \exp \left( \log(N_q^2) - \frac{\eta_q^2 \delta^2}{C_2 \times 2^{-2q_0}} \right) \\
\leq \sum_{q=Q+1}^{\infty} 2 \exp \left( 6q \log(2) - \frac{C_3^2 \times 2^{3q} \delta^2}{C_2} \right) \\
\leq \sum_{q=Q+1}^{\infty} 2 \exp \left( 6q \log(2) - C_4 \times 2^{3q} \times \delta^2 \right),
\]
where $C_4 = \frac{C_3^2}{C_2}$. 

\[1302\]
Let $\epsilon > 0$ and choose $Q$ s.t. $\forall q \geq Q$ and $\forall \delta \leq \epsilon$, $6q \log(2) \leq \frac{1}{2} C_q \times 2^{q\alpha} \times \delta^2$. Then by inequality (19), we have that

$$P_{\epsilon} \left( \sup_{t \in T} \left| \sum_{i=1}^{n} c_i b_n, x_i \{ \tilde{\beta}(t) - \beta(t) - (\tilde{\beta}(\pi_Q(t)) - \beta(\pi_Q(t))) \} \right| \geq \delta \right) \leq \sum_{q=Q+1}^{\infty} 2 \exp \left( -\frac{1}{2} C_q 2^{q\alpha} \delta^2 \right).$$

Now, by inequality (20), the first term on the right-hand side of inequality (17) is

$$= \int_{0}^{\infty} P_{\epsilon} \left( \sup_{t \in T} \left| \sum_{i=1}^{n} c_i b_n, x_i \{ \tilde{\beta}(t) - \beta(t) - (\tilde{\beta}(\pi_Q(t)) - \beta(\pi_Q(t))) \} \right| \geq \delta \right) d\delta \leq \epsilon + \int_{\epsilon}^{\infty} P_{\epsilon} \left( \sup_{t \in T} \left| \sum_{i=1}^{n} c_i b_n, x_i \{ \tilde{\beta}(t) - \beta(t) - (\tilde{\beta}(\pi_Q(t)) - \beta(\pi_Q(t))) \} \right| \geq \delta \right) d\delta \leq \epsilon + \int_{\epsilon}^{\infty} \sum_{q=Q+1}^{\infty} P_{\epsilon} \left( \sup_{t \in T} \left| \sum_{i=1}^{n} c_i b_n, x_i \{ \tilde{\beta}(\pi_Q(t)) - \beta(\pi_Q(t)) - (\tilde{\beta}(\pi_{q-1}(t)) - \beta(\pi_{q-1}(t))) \} \right| \geq \eta q \delta \right) d\delta \leq \epsilon + \int_{\epsilon}^{\infty} \sum_{q=Q+1}^{\infty} 2 \exp \left( -\frac{1}{2} C_q 2^{q\alpha} \delta^2 \right) d\delta \leq \epsilon + \int_{\epsilon}^{\infty} \sum_{q=Q+1}^{\infty} 2 \exp(-C_q q \delta^2) d\delta \leq \epsilon + \int_{\epsilon}^{\infty} \frac{2 \exp(-C_q (Q + 1) \delta^2)}{1 - \exp(-C_q \delta^2)} d\delta \leq \epsilon + \frac{1}{1 - \exp(-C_q \delta^2)} \int_{\epsilon}^{\infty} 2 \exp(-C_q (Q + 1) \delta^2) d\delta \to 0$$

as $\epsilon \to 0$ and $Q \to \infty$, where $C_q$ is chosen to be a positive constant such that $\frac{1}{2} C_q 2^{\alpha q} \geq C_q$, $\forall q$. Using the fact that $\lim_{n \to \infty} P(\Omega_n) = 1$, the result follows. ■

**Remark 5.** From the proof above, it is clear that $c_i$ (in condition (16)) can be relaxed to any i.i.d bounded and symmetric random variables with zero mean and variance one. The technique of the proof above is called “chaining”, which is demonstrated in [11]. The idea is to use a strong inequality like Hoeffding’s inequality to make an exponential bound and then use the condition (A) in Theorem 1. Condition (A) includes two parts; one is about smoothness of sample path and the other is about covering number (potential condition).

**References**