



HAL
open science

Weak Convergence of the Conditional Set-Indexed Empirical Process for Missing at Random Functional Ergodic Data

Salim Bouzebda, Youssouf Souddi, Fethi Madani

► **To cite this version:**

Salim Bouzebda, Youssouf Souddi, Fethi Madani. Weak Convergence of the Conditional Set-Indexed Empirical Process for Missing at Random Functional Ergodic Data. *Mathematics*, 2024, 12 (3), pp.448. 10.3390/math12030448. hal-04427506

HAL Id: hal-04427506

<https://hal.science/hal-04427506>

Submitted on 25 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Article

Weak Convergence of the Conditional Set-Indexed Empirical Process for Missing at Random Functional Ergodic Data

Salim Bouzebda ^{1,*} , Youssouf Souddi ²  and Fethi Madani ²

¹ Laboratoire de Mathématiques Appliquées de Compiègne (L.M.A.C.), Université de Technologie de Compiègne, 60200 Compiègne, France

² Laboratory of Stochastic Models, Statistics and Applications, University of Saida-Dr. Moulay Tahar, P.O. Box 138 EN-NASR, Saïda 20000, Algeria

* Correspondence: salim.bouzebda@utc.fr

Abstract: This work examines the asymptotic characteristics of a conditional set-indexed empirical process composed of functional ergodic random variables with missing at random (MAR). This paper's findings enlarge the previous advancements in functional data analysis through the use of empirical process methodologies. These results are shown under specific structural hypotheses regarding entropy and under appealing situations regarding the model. The regression operator's asymptotic $(1 - \alpha)$ -confidence interval is provided for $0 < \alpha < 1$ as an application. Additionally, we offer a classification example to demonstrate the practical importance of the methodology.

Keywords: conditional distribution; small ball probability; missing at random; empirical process; ergodic functional data; semi-metric space; covering number

MSC: 62G20; 62G05; 62G32; 62G08; 62G35; 62G07; 62E20



Citation: Bouzebda, S.; Souddi, Y.; Madani, F. Weak Convergence of the Conditional Set-Indexed Empirical Process for Missing at Random Functional Ergodic Data. *Mathematics* **2024**, *12*, 448. <https://doi.org/10.3390/math12030448>

Academic Editor: Alexander Dudin

Received: 18 January 2024

Revised: 25 January 2024

Accepted: 29 January 2024

Published: 30 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

There are several strategies for solving problems in statistics, among which empirical process techniques are considered the best. Historically, many limit theorems for the empirical process have been established in finite dimension frameworks (see, e.g., Refs. [1–3] for exhaustive, self-contained texts with a variety of statistical applications) together under mixing conditions and independent identically distributed framework, in the setting of independent variables [4] characterized modulo measurability, the classes \mathcal{C} of sets for which the Glivenko–Cantelli theorem holds, we may also cite Refs. [5–15]. Under various mixing conditions, empirical processes based on dependent data have been investigated; for instance, the authors of Ref. [16] established the asymptotic normality of sequences undergoing *phi*-mixing. Regarding these areas of investigation concerning an alternative form of mixing, it is possible to refer to Refs. [17–20]. Nevertheless, the author of [21] identified a bracketing condition that could occur due to vigorous mixing. The function-indexed empirical procedure for beta-mixing sequences was investigated by Ref. [22]. Uniform convergence and asymptotic normality of a set-indexed conditional empirical process within a strictly stationary and strong mixing framework have been established by Ref. [23]. Over the past few decades, there has been a growing interest in the statistical literature regarding matters concerning functional random variables, which are variables with values that exist in an infinite-dimensional space. As is the case, for example, in meteorology, medicine, satellite imagery, and numerous other scientific disciplines, the proliferation of data collected on an ever-increasingly precise temporal and spatial grid has inspired the development of this research topic. Numerous complex theoretical and numerical inquiries were thus engendered by the statistical modeling of these data, which were perceived as stochastic functions. The monographs of Refs. [24,25] provide comprehensive surveys of functional data analysis, encompassing both theoretical and practical aspects. These

monographs discuss linear models for random variables that take values in a Hilbert space, scalar-on-function and function-on-function linear models, parametric discriminant analysis, and functional principle component analysis, respectively. To access the most recent findings on FDA and related subjects, we may consult the bibliographic reviews provided by sources such as Refs. [26–31], among others. For scalar-on-function nonlinear regression models, the authors of [32] emphasized nonparametric techniques, particularly kernel-type estimation. Such tools were subsequently expanded to include discrimination and classification analysis. An intriguing statistical concept that was extended to the functional data framework was examined by Ref. [33]. These concepts included the portmanteau test, change detection, and goodness-of-fit tests. Good overviews of this literature can be found in Refs. [20,34–41], and, more recently, Ref. [42] gave the first results of the conditional set-indexed empirical process in functional data. Considerable effort has been devoted to developing a convergence theory for empirical processes involving functional random variables, although these topics are well beyond the purview of the paper discussed in Ref. [23]. A theoretical framework of this nature is imperative for contemporary statistical analysis. For over six decades, functional data analysis has been acknowledged in the statistical literature and has since become the focus of numerous works. We observe the extreme limitedness of the outcomes produced by empirical processes utilizing functional frameworks. We may refer for recent references to Refs. [43–47], who achieved numerous valuable outcomes regarding set-indexed conditional empirical processes inside the functional setting of the ergodic framework. One should avoid overlooking the possibility that some pairings of observations may be incomplete in numerous practical applications, including sampling surveys, pharmaceutical tracing tests, and reliability tests. Such instances are commonly referred to as “missing data”. Others in the fields of data science and analytics will attest to the fact that missing data is a common issue. MAR (Missing At Random) indicates that while there may be systematic differences between the missing and observed values, these discrepancies can be fully accounted for by other observed variables. The situation changes significantly when predictors are present; for instance, the authors of [48–58] provide some examples of this in finite dimensionality, as recent references to Refs. [59,60]. In a recent study, the authors of [61] examined the linear quantile regression model in the presence of missing response data that occur randomly. The study utilized the inverse probability weight method. The authors developed a mathematical equation for estimating unknown parameters using quantile regression. They also introduced a standard estimator for quantile regression. Simultaneously, they formulated the empirical likelihood (EL) ratio function for the unknown parameter and established a maximum EL estimator for the unknown parameter. There is a scarcity of work that examines the statistical characteristics of functional nonparametric models for missing data. The kernel estimator of the conditional quantile was introduced by Ref. [62] under the assumptions of ergodicity and random censorship. The author also demonstrated strong consistency (with rate) and defined the asymptotic distribution of the estimator. Additionally, they applied the estimator to forecast the peak electricity demand interval using smart meter data, details of which have been omitted. In their study, the authors of [63] developed a type of estimator for the regression operator in the context of functional stationary ergodic data with missing at random (MAR) responses. They also established the asymptotic properties of the estimator, including its convergence rate in probability and asymptotic normality. For further references, we suggest consulting Refs. [64,65].

Our findings extend upon a prior study [44] by establishing more precise limits under less stringent limitations. This offers a new perspective of the empirical processes theory for random variables with general dependencies. This work addresses a problem that has not been thoroughly examined thus far. The framework of ergodic functional data was introduced by Ref. [66], who established consistencies with rates along with the asymptotic normality of the regression function estimate and provided some examples. For recent papers on the subject, we refer to Ref. [43], where the authors extended Ref. [66] to a more

general framework. Some motivations to consider ergodic dependence structure in the data rather than a mixing one are discussed in Refs. [67,68].

The objective of this study is to enhance the development of a practical methodology for addressing MAR samples in functional nonparametric situations. We want to examine the estimation of conditional set-indexed empirical processes in the presence of both missing at random (MAR) data and ergodicity.

The structure of this paper is outlined as follows. In Section 2, we introduce the notation and definitions, along with the conditional empirical process. Our main results are presented in Section 3. Section 3.1 is dedicated to discussing the procedure for selecting the bandwidth. In Section 4, we apply our main result to classification. Concluding remarks and potential future developments are discussed in Section 5. To maintain a smooth presentation flow, all proofs are consolidated in Appendix A.

2. The Set Indexed Conditional Empirical Process

To enhance clarity, let us delve into the definition of the ergodic property for processes. Consider a measurable space (S, \mathcal{J}) , and denote by $S^{\mathbb{N}}$ the space of all functions $s : \mathbb{N} \rightarrow S$. If s_j represents the value of the function s at $j \in \mathbb{N}$, define H_j as the j -th coordinate map, i.e., $H_j(s) = s_j$. Now, consider $H_j^{-1}(\mathcal{J})$ for $j \in \mathbb{N}$; a random process $Z = Z_j : j \in \mathbb{N}$ can be viewed as a random variable defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, taking values in $(S^{\mathbb{N}}, \mathcal{J}^{\mathbb{N}})$. For any $B \in \mathcal{F}$, a set is termed invariant if there exists a set $\mathcal{A} \in \mathcal{J}^{\mathbb{N}}$ such that $B = (Z_n, Z_{n+1}, \dots) \in \mathcal{A}$ holds for every $n \geq 1$. The process Z is then considered ergodic when, for any invariant set B , we have $\mathbb{P}(B) = 0$ or $\mathbb{P}(\Omega | B) = 0$. As per the ergodic theorem, it is well-known that for a stationary ergodic process Z , the following convergence holds almost surely:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n Z_i = \mathbb{E}(Z_1), \text{ almost surely.} \tag{1}$$

Therefore, the ergodic property in our setting is formulated based on the statement (1). We consider a sample of random elements $(X_1, Y_1), \dots, (X_n, Y_n)$, each drawn from the joint distribution of (X, Y) , where X takes values in a space \mathcal{E} and Y in \mathbb{R}^d . The functional space \mathcal{E} is endowed with a semi-metric $d_{\mathcal{E}}(\cdot, \cdot)$. Our goal is to investigate the relationships between X and Y by estimating functional operators associated with the conditional distribution of Y given X . One such operator is the regression operator for a measurable set C in a class of sets \mathcal{C} :

$$\mu(C | x) = \mathbb{E}(\mathbb{1}_{\{Y \in C\}} | X = x).$$

To address this, we employ a Nadaraya–Watson-type conditional empirical distribution, as proposed by Refs. [42,44,69,70]. We introduce the term MAR (Missing mechanism with MAR) for the response variable. In an available incomplete sample of size from (X, Y, δ) , denoted as $(X_i, Y_i, \delta_i), 1 \leq i \leq n$, X_i is fully observed, $\delta_i = 1$ if Y_i is observed, and $\delta_i = 0$ otherwise. The Bernoulli random variable δ satisfies:

$$\mathbb{P}(\delta = 1 | X = x; Y = y) = \mathbb{P}(\delta = 1 | X = x) = \mathcal{P}(x),$$

where $\mathcal{P}(x)$ is a function operator, termed the conditional probability of observing the response given the predictor, often unknown. This mechanism implies that δ and Y are conditionally independent given X , akin to the finite-dimensionality case in Ref. [48].

The Nadaraya–Watson-type conditional empirical distribution function is given by:

$$\mu_n(C, x) = \frac{\sum_{i=1}^n \delta_i \mathbb{1}_{\{Y_i \in C\}} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}{\sum_{i=1}^n \delta_i K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}, \tag{2}$$

where $K(\cdot)$ is a real-valued kernel function from $[0, \infty)$ into $[0, \infty)$, h_n is a smoothing parameter satisfying $h_n \rightarrow 0$ as $n \rightarrow \infty$, C is a measurable set, and $x \in \mathcal{E}$. When choosing $C = (-\infty, z]$, where $z \in \mathbb{R}^d$, it reduces to the conditional empirical distribution function $F_n(z|x) = \mu_n((-\infty, z], x)$, as referenced in Refs. [71–73]. However, the corresponding class \mathcal{C} is defined as $\{(-\infty, z], z \in \mathbb{R}^d\}$. Regarding the semi-metric topology on \mathcal{E} , we introduce the notation

$$B(x, t) = \{x_1 \in \mathcal{E} : d_{\mathcal{E}}(x_1, x) \leq t\},$$

which denotes the ball in \mathcal{E} with center x and radius t . This concept is commonly referred to as the small ball probability function in the literature, especially when t tends to zero. The significance of this notion is both theoretically and practically profound, as the concept of a ball is intricately connected with the semi-metric $d(\cdot, \cdot)$. The selection of this semi-metric becomes pivotal when dealing with data in infinite-dimensional spaces.

In many cases, the probability function for the small ball can be roughly represented as the multiplication of two independent functions with respect to variables x and h . This insight is illustrated in several examples found in Proposition 1 of [74]:

1. $\phi(h_n) = Ch_n^v$ for some $v > 0$ with $\tau_0(s) = s^v$;
2. $\phi(h_n) = Ch_n^v \exp(-Ch_n^{-p})$ for some $v > 0$ and $p > 0$ with $\tau_0(s)$ is the Dirac's function;
3. $\phi(h_n) = C|\ln(h_n)|^{-1}$ with $\tau_0(s) = \mathbb{1}_{]0,1]}(s)$ the indicator function in $]0, 1]$.

Define the following σ -fields: \mathfrak{F}_i and \mathfrak{G}_i Let

$$\mathfrak{F}_i = \sigma((X_i, Y_i, \delta_i) : 0 \leq i \leq n),$$

$$\mathfrak{G}_i = \sigma((X_i, Y_i, \delta_i) : 0 \leq i \leq n),$$

where \mathfrak{F}_i be the σ -field generated by $((X_1, Y_1, \delta_1), \dots, (X_i, Y_i, \delta_i))$ and \mathfrak{G}_i that generated by $((X_1, Y_1, \delta_1), \dots, (X_i, Y_i, \delta_i), X_{i+1})$. Let $B(x, u)$ be a ball centered at $x \in \mathcal{E}$ with radius u . Let $D_i = d(x, X_i)$ so that D_i is a nonnegative real-valued random variable. Operating within the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, consider

$$F_x(u) = \mathbb{P}(D_i \leq u) = \mathbb{P}(X_i \in B(x, u)),$$

and $F_x^{\mathfrak{F}_{i-1}} = \mathbb{P}(X_i \in B(x, u) \mid \mathfrak{F}_{i-1})$ to be the distribution function and the conditional distribution function, respectively, given the σ -field \mathfrak{F}_{i-1} of $(D_i)_{i \geq 1}$. Here, $B(x, u)$ denotes the ball in the space \mathcal{E} centered at x with radius u . Let $o_{a.s.}(u)$ represent a real random function $l(\cdot)$ such that $l(u)/u$ converges to zero almost surely as $u \rightarrow 0$. In a similar vein, define $\mathcal{O}_{a.s.}(u)$ as a real random function $l(\cdot)$ such that $l(u)/u$ is almost surely bounded. In what follows, we implicitly assume the ergodicity of the sequence of random elements $(X_i, Y_i), i = 1, \dots, n$.

2.1. Assumptions and Notation

In this paper, the variable x is a constant element within the functional space \mathcal{E} . We present the metric entropy with inclusion as a means to quantify the richness or complexity of the set class \mathcal{C} . For any given $\varepsilon > 0$, the covering number is defined as:

$$\begin{aligned} \mathcal{N}(\varepsilon, \mathcal{C}, \mu(\cdot \mid x)) &= \inf\{n \in \mathbb{N} : \exists C_1, \dots, C_n \in \mathcal{C} \text{ such that } \forall C \in \mathcal{C} \exists 1 \leq i, j \leq n \\ &\text{with } C_i \subset C \subset C_j \text{ and } \mu(C_j \setminus C_i \mid x) < \varepsilon\}. \end{aligned}$$

The term $\log(\mathcal{N}(\varepsilon, \mathcal{C}, \mu(\cdot \mid x)))$ is referred to as the metric entropy with inclusion of \mathcal{C} with respect to $\mu(\cdot \mid x)$. For numerous classes, estimates for these covering numbers are well-documented; refer, for instance, to Ref. [75]. Below, we frequently make the assumption

that either $\log \mathcal{N}(\varepsilon, \mathcal{C}, \mu(\cdot | x))$ or $\mathcal{N}(\varepsilon, \mathcal{C}, \mu(\cdot | x))$ exhibit behaviors reminiscent of powers of ε^{-1} . We affirm that condition (R_γ) is satisfied when

$$\log \mathcal{N}(\varepsilon, \mathcal{C}, \mu(\cdot | x)) \leq H_\gamma(\varepsilon), \text{ for all } \varepsilon > 0, \tag{3}$$

where

$$H_\gamma(\varepsilon) = \begin{cases} \log(A\varepsilon) & \text{if } \gamma = 0, \\ A\varepsilon^{-\gamma} & \text{if } \gamma > 0, \end{cases}$$

for some constants $A, r > 0$. As emphasized in Ref. [23], it is notable that the condition (3), where $\gamma = 0$, is fulfilled by intervals, rectangles, balls, ellipsoids, and by classes derived from these through finite set operations of union, intersection, and complement. The class of convex sets in \mathbb{R}^d ($d \geq 2$) satisfies the condition (3) with $\gamma = (d - 1)/2$. Various other sets that satisfy (3) with $\gamma > 0$ are elaborated upon in Ref. [75]. We give now further notation. For $j \geq 1$, set

$$M_j = K^j(1) - \int_0^1 (K^j)'(u) \tau_0(u) du.$$

In this section, we establish the weak convergence of the process $v_n(C, x) : C \in \mathcal{C}$ as defined by

$$v_n(C, x) := \sqrt{n\phi(h_n)}(\mu_n(C, x) - \mathbb{E}\mu_n(C, x)). \tag{4}$$

In the course of our analysis, we will rely on the following assumptions.

(H1) For every $x \in \mathcal{E}$, there exists a sequence of nonnegative bounded random functionals $(f_{i,1})_{i \geq 1}$, a sequence of random functions $(g_{i,x})_{i \geq 1}$, a deterministic nonnegative bounded functional f_1 , and a nonnegative real function ϕ such that $\phi(h_n) \rightarrow 0$ as $h \rightarrow 0$, as $h \rightarrow 0$, such that

- (i) $F_x(u) = \phi(u)f_1(x) + o(\phi(u))$ as $u \rightarrow 0$.
- (ii) For any $i \in \mathbb{N}$, $F_x^{\mathfrak{F}^{i-1}}(u) = \phi(u)f_{i,1}(x) + g_{i,x}(u)$ with $g_{i,x}(u) = o_{a.s.}(\phi(u))$ as $u \rightarrow 0$, $g_{i,x}(u)/\phi(u)$ almost surely bounded and $n^{-1} \sum_{i=1}^n g_{i,x}^j(u) = o_{a.s.}(\phi^j(u))$ as $n \rightarrow \infty, j = 1, 2$.
- (iii) $n^{-1} \sum_{i=1}^n f_{i,1}^j(x) \rightarrow f_1^j(x)$ almost surely as $n \rightarrow \infty$, for $j = 1, 2$.
- (iv) There exists a nondecreasing bounded function $\tau_0(u)$ that uniformly holds for all $u \in (0, 1)$.

$$\tau_0(u) + o(1) = \frac{\phi(ru)}{\phi(r)},$$

$$\text{as } r \downarrow 0 \text{ and } 1 \leq j \leq 2 + \delta \text{ with } \delta > 0, \int_0^1 (K^j(u))' \tau_0(u) du < \infty.$$

(H2) There exist positive constants $\beta > 0$ and $\eta_1 > 0$ such that for all $x_1, x_2 \in N_x$, a neighborhood of x , the following holds

$$|\mu(C | x_1) - \mu(C | x_2)| \leq \eta_1 d_{\mathcal{E}}^\beta(x_1, x_2).$$

(H3)(i) The conditional mean of $\mathbb{1}_{Y_i} \in C$ given the σ -field \mathfrak{G}_{i-1} depends solely on X_i , meaning that for any $i \geq 1$, $\mathbb{E}(\mathbb{1}_{Y_i} \in C | \mathfrak{G}_{i-1}) = \mu(X_i)$ almost surely. The conditional mean of $\mathbb{1}_{Y_i} \in C$ given the σ -field \mathfrak{G}_{i-1} also depends only on X_i , i.e., for any $i \geq 1$,

$$\mathbb{E} \left(\left(\mathbb{1}_{\{Y_i \in C\}} - \mu(X_i) \right)^2 \mid \mathfrak{G}_{i-1} \right) = W(X_i),$$

almost surely.

- (ii) Furthermore, the functions $W(\cdot)$ and $\mathcal{P}(\cdot)$ are continuous in a neighborhood of x , namely,

$$\sup_{\{u:d(x,u)\leq h\}} |W(u) - W(x)| = o(1) \text{ as } h \rightarrow 0,$$

$$\sup_{\{u:d(x,u)\leq h\}} |\mathcal{P}(u) - \mathcal{P}(x)| = o(1) \text{ as } h \rightarrow 0.$$

- (iii) $\exists \delta > 0$ such that we let

$$\overline{W}_{2+\delta}(u) = \mathbb{E}\left(|(\mathbb{1}_{\{Y_1 \in C\}} - \mu(x))|^{2+\delta} \mid X_1 = u\right)$$

be continuous in a neighborhood of x for $u \in \mathcal{E}$.

- (H4) For any $(y_1, y_2) \in \mathbb{R}^{2d}$ and positive constants $b_3 > 0$ and $\eta_4 > 0$, the following holds for the conditional density $f(\cdot)$ of Y given $X = x$:

$$|f(y_1) - f(y_2)| \leq \eta_4 \|y_1 - y_2\|^{b_3}.$$

- (H5) The kernel function $K(\cdot)$ has support within the interval $(0, 1)$ and possesses a continuous first derivative on $(0, 1)$. It satisfies the condition $K'(t) < 0$ for all $t \in (0, 1)$. Moreover,

$$\left| \int_0^1 (K^j)'(u) du \right| < \infty, \text{ for } j = 1, 2.$$

- (H6) Suppose that the set class \mathcal{C} adheres to condition (3);

- (H7) The smoothing parameter (h_n) fulfills the following criterion: $h_n \rightarrow 0$ and $n\phi(h_n) \rightarrow \infty$ as $n \rightarrow \infty$.

2.2. Comments on the Assumptions

The significance of condition (H1) extends to both the ergodic and functional aspects addressed in this paper. The condition utilized here shares similarities with that employed in Ref. [66]. The functions $f_{i,1}(\cdot)$ and $f_1(\cdot)$ play roles analogous to the conditional and unconditional densities in the finite-dimensional scenario. In the meantime, $\phi(u)$ describes the influence of the radius u on the small ball probability as u tends to zero, as illustrated in Ref. [66]. Conditions (H2)(i) are standard in nonparametric regression estimation. (H3)(i) is essential for establishing consistency, reflecting the Markovian nature of the functionally stationary ergodic data. This condition aligns with that used in Ref. [63]. (H3)(ii,iii) serve as continuous local conditions, necessary for the main results and for conciseness in this paper. Condition (H4) on the density $f(\cdot)$ conforms to a classical Lipschitz-type nonparametric functional model. Assumption (H5) relates to the choice of the kernel $K(\cdot)$, a common practice in nonparametric functional estimation. It is worth noting that the Parzen symmetric kernel is unsuitable in this context due to the positivity of the random process $d(x, X)$. Hence, we consider $K(\cdot)$ with support $[0, 1]$, a natural generalization of the assumption usually made in the multivariate case, where $K(\cdot)$ is expected to be a spherically symmetric density function. The conditions $K(1) > 0$ and $K'(\cdot) < 0$ ensure that $M_1 > 0$ for all limit functions τ_0 . The condition $K(1) > 0$ is necessary for defining the moments M_2 , which, in this case, are determined by the value $\mathcal{K}(1)$. (H7) provides a condition on the bandwidths, acknowledging that consistency cannot be guaranteed without it.

3. Main Results

Below, we note $Z \stackrel{\mathcal{D}}{=} \mathcal{N}(\mu, \sigma^2)$ when the random variable Z is distributed according to a normal distribution with mean μ and variance σ^2 . The symbol $\stackrel{\mathcal{D}}{\rightarrow}$ represents convergence in distribution, while $\xrightarrow{\mathbb{P}}$ indicates convergence in probability.

Theorem 1 (Uniform Consistency). Assume that the conditions (H1)–(H7) are satisfied. Consider a class of measurable sets \mathcal{C} for which

$$\mathcal{N}(\varepsilon, \mathcal{C}, \mu(\cdot | x)) < \infty,$$

for any $\varepsilon > 0$. Moreover, assume that for every $C \in \mathcal{C}$

$$|\mu(C, y)f(y) - \mu(C, x)f(x)| \rightarrow 0, \text{ as } y \rightarrow x.$$

If $n\phi(h_n) \rightarrow \infty$ and $h_n \rightarrow 0$ as $n \rightarrow \infty$, then

$$\sup_{C \in \mathcal{C}} |\mu_n(C, x) - \mathbb{E}(\mu_n(C, x))| \xrightarrow{\mathbb{P}} 0.$$

Note that the proof of Theorem 1 follows directly from the decomposition

$$\begin{aligned} \mu_n(C, x) - \mathbb{E}(\mu_n(C, x)) &= \frac{1}{\mathbb{E}(\widehat{f}_n(x))} [\widehat{\varphi}_n(C, x) - \mathbb{E}(\widehat{\varphi}_n(C, x))] \\ &\quad - \frac{\mu_n(C, x)}{\mathbb{E}(\widehat{f}_n(x))} [\widehat{f}_n(x) - \mathbb{E}(\widehat{f}_n(x))], \\ &= \frac{Q_n(x)}{\mathbb{E}(\widehat{f}_n(x))}, \end{aligned}$$

where

$$Q_n(x) = [\widehat{\varphi}_n(C, x) - \mathbb{E}(\widehat{\varphi}_n(C, x))] - \mu_n(C, x) [\widehat{f}_n(x) - \mathbb{E}(\widehat{f}_n(x))].$$

and

$$\begin{aligned} \widehat{\varphi}_n(C, x) &= \frac{1}{n\phi(h_n)} \sum_{i=1}^n \delta_i \mathbb{1}_{\{Y_i \in C\}} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right), \\ \widehat{f}_n(x) &= \frac{1}{n\phi(h_n)} \sum_{i=1}^n \delta_i K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right). \end{aligned}$$

Let $\Delta_i(x) = K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)$. We have

$$\begin{aligned} \widehat{\varphi}_n(C, x) &= \frac{1}{n\phi(h_n)} \sum_{i=1}^n \mathbb{1}_{\{Y_i \in C\}} \delta_i \Delta_i(x), \\ \widehat{f}_n(x) &= \frac{1}{n\phi(h_n)} \sum_{i=1}^n \delta_i \Delta_i(x). \end{aligned}$$

Henceforth, for $x \in \mathcal{E}$, let us denote

$$\mathbb{E}(\widehat{\varphi}_n(C, x)) = \frac{1}{n\mathbb{E}(\Delta_1(x))} \sum_{i=1}^n \mathbb{E}(\delta_i \mathbb{1}_{\{Y_i \in C\}} \Delta_i(x) | \mathfrak{F}_{i-1}),$$

and

$$\mathbb{E}(\widehat{f}_n(x)) = \frac{1}{n\mathbb{E}(\Delta_1(x))} \sum_{i=1}^n \mathbb{E}(\delta_i \Delta_i(x) | \mathfrak{F}_{i-1}),$$

here, $\mathbb{E}(X | \mathfrak{F})$ represents the conditional expectation of the random variable X given the σ -field \mathfrak{F} .

To establish asymptotic normality, define the “bias” term as

$$B_n(x) = \frac{\mathbb{E}(\widehat{f}_n(x)) - \mu_n(C, x)\mathbb{E}(\widehat{\varphi}_n(C, x))}{\mathbb{E}(\widehat{\varphi}_n(C, x))}$$

The subsequent result presents the weak convergence. It is important to note that $f_1(x)$ is specified in **(H1)**.

Theorem 2 (Asymptotic normality). *Assuming **(H1)–(H7)**, as $n \rightarrow \infty$, for $m \geq 1$ and $C_1, \dots, C_m \in \mathcal{C}$, we have*

$$\{v_n(C_i, x)_{i=1, \dots, m}\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma),$$

where $\Sigma = \sigma_{ij}(x), i, j = 1, \dots, m$ and

$$\sigma_{ij}(x) = \frac{M_2}{\mathcal{P}(x)M_1^2 f_1(x)} \left(\mathbb{E}(\mathbb{1}_{\{Y \in C_i \cap C_j\}} \mid X = x) - \mathbb{E}(\mathbb{1}_{\{Y \in C_i\}} \mid X = x)\mathbb{E}(\mathbb{1}_{\{Y \in C_j\}} \mid X = x) \right),$$

whenever $f_1(x) > 0$ and

$$M_1 = K(1) - \int_0^1 K'(u)\tau_0(u)du, \quad M_2 = K^2(1) - \int_0^1 (K^2)'(u)\tau_0(u)du.$$

To obtain the density of the process, it is essential to introduce the following function, which provides insights into the asymptotic behavior of the modulus of continuity:

$$\Lambda_\gamma(\sigma^2, n) = \begin{cases} \sqrt{\sigma^2 \log \frac{1}{\sigma^2}}, & \text{if } \gamma = 0; \\ \max\left((\sigma^2)^{(1-\gamma)/2}, n\phi(h_n)^{(3\gamma-1)/(2(3\gamma+1))}\right), & \text{if } \gamma > 0. \end{cases}$$

Theorem 3. *Assume that **(H1)–(H7)** are satisfied. For every $\sigma^2 > 0$, consider $\mathcal{C}_\sigma \subset \mathcal{C}$ as a class of measurable sets with*

$$\sum_{t=1}^n \sup_{C \in \mathcal{C}_\sigma} \mu(C, x) \leq \sigma^2 \leq 1,$$

and suppose that \mathcal{C} fulfils (3) with $\gamma \geq 0$. Additionally, we assume that $\phi(h_n) \rightarrow 0$ and $n\phi(h_n) \rightarrow +\infty$ as $n \rightarrow +\infty$, such that

$$n\phi(h_n) \leq \left(\Lambda_\gamma(\sigma^2, n)\right)^2,$$

and as $n \rightarrow +\infty$, we have

$$\frac{n\phi\left(\sigma^2 \log\left(\frac{1}{\sigma^2}\right)\right)^{1+\gamma}}{\log(n)} \rightarrow \infty.$$

Furthermore, we assume that $\sigma^2 \geq h^2$. For $\gamma > 0$ and $d = 1, 2$, the latter has to be replaced by $\sigma^2 \geq \phi(h_n) \log\left(\frac{1}{\phi(h_n)}\right)$. Under the conditions of Theorem 2, the process converges in law to a Gaussian process $\{v(C, x) : C \in \mathcal{C}\}$, which possesses a version with uniformly bounded and uniformly continuous paths with respect to the $\|\cdot\|_2$ -norm. The covariance is given by $\sigma_{ij}(x)$ as specified in Theorem 2.

Remark 1. *The distance of two measures μ_1, μ_2 in the Prokhorov metric is defined as (see, e.g., Refs. [76–79])*

$$\rho_P(\mu_1, \mu_2) := \inf\{\varepsilon > 0 \mid \mu_1(B) \leq \mu_2(B^\varepsilon) + \varepsilon, \forall \text{ Borel sets } B \subset \Omega\}$$

Here $B^\varepsilon = \{x \mid d(x, B) < \varepsilon\}$, where $d(x, B)$ is the distance of x to B , i.e., $d(x, B) = \inf_{z \in B} \|x - z\|$. The distance of two random variables ξ_1, ξ_2 in the Ky Fan metric is defined as [80]

$$\rho_K(\xi_1, \xi_2) := \inf\{\varepsilon > 0 \mid \mu\{\omega \in \Omega \mid d(\xi_1(\omega), \xi_2(\omega)) > \varepsilon\} < \varepsilon\}.$$

It is worthwhile to establish an adequate link of our findings to these distances in the conditional setting.

Remark 2. Central limit theorems are frequently utilized to establish confidence intervals for the target being estimated. In the realm of non-parametric estimation, the asymptotic variance $\Sigma(x) := \sigma_{i,j}(x)$ in the central limit depends on certain unknown functions. Consequently, in practical scenarios, only approximate confidence intervals can be derived, even when $\Sigma(x)$ is functionally specified. Notably, according to Theorem 2, the limiting variance incorporates the unknown function $f_1(\cdot)$ and the normalization is contingent on the function $\phi(\cdot)$, which is not explicitly identifiable in practice. Furthermore, the quantities $W(\cdot)$ and τ_0 need to be estimated. The corollary below, a slight modification of Theorem 2, permits a practical form of the results to be used, as typically the conditional variance $W(x)$ is estimated similarly to what is obtained by Ref. [63].

Let

$$\begin{aligned} W_n &= \frac{\sum_{i=1}^n (\delta_i \mathbb{1}_{\{Y_i \in C\}} - \mu_n(x))^2 K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h}\right)}{\sum_{i=1}^n \delta_i K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h}\right)} \\ &= \frac{\sum_{i=1}^n (\delta_i \mathbb{1}_{\{Y_i \in C\}} - \mu_n(x))^2 K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h}\right)}{\sum_{i=1}^n \delta_i K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h}\right)} - (\mu_n(x))^2 \\ &= \widehat{g}_n(x) - (\mu_n(x))^2. \end{aligned}$$

Let us introduce the following estimation

$$\mathcal{F}_{x,n}(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{d(x, X_i) \leq t\}}.$$

By employing the decomposition of $\tau_0(\cdot)$ in (H1)(i) and (H1)(i,iv), one can estimate $\tau_0(\cdot)$ as

$$\tau_n(t) = \frac{\mathcal{F}_{x,n}(th)}{\mathcal{F}_{x,n}(h)}.$$

Subsequently, for a given kernel $K(\cdot)$ and the quantities M_1 and M_2 can be estimated as follows

$$M_{1,n} = K(1) - \int_0^1 K'(s) \tau_n(s) ds, \quad M_{2,n} = K^2(1) - \int_0^1 (K^2)'(s) \tau_n(s) ds.$$

Finally, the estimator of $\mathcal{P}(x)$ is denoted by

$$\mathcal{P}_n(x) = \frac{\sum_{i=1}^n \delta_i K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}.$$

Corollary 1. Suppose that conditions (H1)–(H7) are satisfied, where K' and $(K^2)'$ are integrable functions. Additionally, assume that $n\mathcal{F}_x(h) \rightarrow \infty$ and $h^{\beta}(n\mathcal{F}_x(h))^{1/2} \rightarrow 0$ as $n \rightarrow \infty$. Then, for any $x \in \mathcal{E}$ such that $f_1(x) > 0$, we have

$$\frac{M_{1,n}}{\sqrt{M_{2,n}}} \sqrt{\frac{n\mathcal{F}_{x,n}(h_n)\mathcal{P}_n(x)}{W_n(x)}} (\mu_n(C, x) - \mu(C, x)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Using Corollary (1) the asymptotic $100(1 - \alpha)$ confidence band given by

$$\left[\mu_n(C, x) - c_\alpha \frac{M_{1,n}}{\sqrt{M_{2,n}}} \sqrt{\frac{W_n(x)}{n\mathcal{F}_{x,n}(h)\mathcal{P}_n(x)}}, \mu_n(C, x) + c_\alpha \frac{M_{1,n}}{\sqrt{M_{2,n}}} \sqrt{\frac{W_n(x)}{n\mathcal{F}_{x,n}(h)\mathcal{P}_n(x)}} \right].$$

where c_α is the upper $\frac{\alpha}{2}$ quantile of the Normal distribution $\mathcal{N}(0, 1)$

3.1. The Bandwidth Selection Criterion

Several approaches have been devised and refined to formulate asymptotically optimal bandwidth selection rules for nonparametric kernel estimators, particularly for the Nadaraya–Watson regression estimator. Some noteworthy contributions include [81–87]. Choosing this parameter appropriately is essential, whether in the conventional finite-dimensional case or within the infinite-dimensional framework, to guarantee favorable practical performance. Let us define the leave-out- (X_j, Y_j, δ_j) estimator for the regression function

$$\mu_{n,j}(C, x) = \frac{\sum_{i=1, i \neq j}^n \delta_i \mathbb{1}_{\{Y_i \in C\}} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}{\sum_{i=1}^n \delta_i K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}. \tag{5}$$

To minimize the quadratic loss function, we introduce the following criterion, where we have a (known) nonnegative weight function $\mathcal{W}(\cdot)$:

$$CV(C, h) := \frac{1}{n} \sum_{j=1}^n \left(\delta_j \mathbb{1}_{\{Y_j \in C\}} - \mu_{n,j}(C, X_j) \right)^2 \mathcal{W}(X_j). \tag{6}$$

Building upon the concepts developed by Ref. [83], a natural approach for selecting the bandwidth is to minimize the preceding criterion. Thus, let us choose \hat{h}_n , as the minimizer over h :

$$\sup_{C \in \mathcal{C}} CV(C, h).$$

One can replace (6) by

$$CV(C, h_n) := \frac{1}{n} \sum_{j=1}^n \left(\delta_j \mathbb{1}_{\{Y_j \in C\}} - \mu_{n,j}(C, X_j) \right)^2 \widehat{\mathcal{W}}(X_j, x). \tag{7}$$

In practice, one takes, for $j = 1, \dots, n$, the uniform global weights $\mathcal{W}(X_j) = 1$, and the local weights

$$\widehat{\mathcal{W}}(X_j, x) = \begin{cases} 1 & \text{if } d(X_j, x) \leq h_n, \\ 0 & \text{otherwise.} \end{cases}$$

For brevity, we have concentrated on the most popular method, namely, the cross-validated selected bandwidth. This approach can be extended to any other bandwidth selector, such as the bandwidth based on Bayesian ideas [88].

4. Applications to Classification with Partially Labeled Data

In this section, we apply the results developed in the previous sections to the problem of statistical classification. We consider a sample of random elements $(X_1, Y_1), \dots, (X_n, Y_n)$ drawn from the joint distribution of (X, Y) , where X takes values in a space \mathcal{E} and Y in \mathbb{R}^d . In classification, the objective is to predict the integer-valued label Y based on the covariate vector X . More formally, we aim to find a function (classifier) $\theta : \mathcal{E} \rightarrow \mathbb{R}^d$ for

which the probability of misclassification error (incorrect prediction), i.e., $\mathbb{P}(\theta(X) \neq Y)$, is minimized. Let

$$\mathcal{P}_k(x) = \mathbb{P}(Y = k \mid X = x), \quad x \in \mathcal{E}, 1 \leq k \leq n.$$

Demonstrating that the optimal classifier, i.e., the one with the minimum probability of error, is given by

$$\theta_B(x) = \arg \max_{1 \leq k \leq n} \mathcal{P}_k(x),$$

i.e., the best classifier θ_B satisfies

$$\max_{1 \leq k \leq n} \mathcal{P}_k(x) = \mathcal{P}_{\theta_B(x)}(x).$$

As θ_B is unknown, the data is utilized to construct estimates of θ_B . Specifically, let $\mathbb{D}_n = (X_1, Y_1), \dots, (X_n, Y_n)$ represent a random sample from the distribution of (X, Y) , where each (X_i, Y_i) is fully observable. Let $\hat{\theta}_n$ be any sample-based classifier. In other words, $\hat{\theta}_n(X)$ is the predicted value of Y , based on \mathbb{D}_n and X . Let

$$L_n(\hat{\theta}_n) = \mathbb{P}(\hat{\theta}_n(X) \neq Y \mid \mathbb{D}_n),$$

be the conditional probability of error of the sample-based classifier $\hat{\theta}_n$. Then $\hat{\theta}_n$ is said to be consistent if $L_n(\hat{\theta}_n) \rightarrow L_n(\theta_B) = \mathbb{P}(\theta_B(X) \neq Y)$ as $n \rightarrow \infty$, for $k = 1, \dots, n$. Let $\hat{\mathcal{P}}_k(x)$ be any sample-based estimators of $\mathcal{P}_k(x) = \mathbb{P}(Y = k \mid X = x)$ and define the classification rule $\hat{\theta}_n$ by

$$\hat{\theta}_n(x) = \arg \max_{1 \leq k \leq n} \hat{\mathcal{P}}_k(x).$$

In other words, $\hat{\theta}_n$ satisfies

$$\max_{1 \leq k \leq n} \hat{\mathcal{P}}_k(x) = \hat{\mathcal{P}}_{\hat{\theta}_n(x)}(x),$$

to show $L_n(\hat{\theta}_n) - L_n(\theta_B) \rightarrow 0$ it is sufficient to show that $\hat{\mathcal{P}}_k(x) - \mathcal{P}_k(x) \rightarrow 0$ by posing $\delta_i = \hat{\mathcal{P}}_k(x)$, we have

$$\mu_n(C, x) = \frac{\sum_{i=1}^n \hat{\mathcal{P}}_k(x) \mathbb{1}_{\{Y_i \in C\}} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}{\sum_{i=1}^n \hat{\mathcal{P}}_k(x) K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}. \tag{8}$$

Theorem 4. Under the conditions of Theorem 3, we have the convergence

$$L_n(\hat{\theta}_n) - L_n(\theta_B) \rightarrow 0.$$

5. Concluding Remarks

In this investigation, we have examined the asymptotic properties of the conditional set-indexed empirical process involving ergodic functional data that are missing at random (MAR). Our findings are obtained under assumptions pertaining to the richness of the index class \mathcal{C} of sets in terms of metric entropy with bracketing. Our contribution is two-fold: first, we have developed a functional methodology for addressing MAR samples in non-parametric problems, and second, we have extended our non-parametric conditional methodology by incorporating the ergodicity concepts introduced in Ref. [44]. Several challenging open questions remain in this context, including potential extensions to other types of non-parametric predictors such as functional local linear predictors, functional k NN predictors, and others. Furthermore, exploring extensions to problems beyond prediction, such as the estimation of variance error, is an interesting avenue for future research. Another direction for future exploration is the consideration of reducing the predictor’s dimensionality by employing a Single Functional Index Model (SFIM) to estimate the

regression, as discussed in Refs. [89,90]. SFIM has shown its effectiveness in improving the consistency of the regression operator estimator.

Author Contributions: Conceptualization, S.B.; methodology, S.B.; validation, S.B., Y.S. and F.M.; formal analysis, S.B. and Y.S.; investigation, S.B. and Y.S.; original draft preparation, S.B. and Y.S.; writing—review and editing, S.B. and Y.S.; supervision, S.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data is contained within the article.

Acknowledgments: The authors would like to thank the Editor-in-Chief, an Associate-Editor, and the three referees for their extremely helpful remarks, which resulted in a substantial improvement of the original form of the work and a presentation that was more sharply focused.

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

Appendix A

The proofs of our results are presented in this section. The notation introduced earlier is also utilized in the subsequent sections.

Lemma A1. Assume that conditions (H1(i))–(H1(ii))–(H1(iv))–(H5) hold true for any real numbers $1 \leq j \leq 2 + \delta$ and $1 \leq k \leq 2 + \delta$ with $\delta > 0$. As $n \rightarrow \infty$, we have:

- (i) $\frac{1}{\phi(h)} \mathbb{E}(\Delta_i^j(x) \mid \mathfrak{F}_{i-1}) = M_j f_{i,1}(x) + \mathcal{O}_{a.s.}\left(\frac{g_{i,x}(h)}{\phi(h)}\right)$;
- (ii) $\frac{1}{\phi(h)} \mathbb{E}(\Delta_1^j(x)) = M_j f_1(x) + o(1)$;
- (iii) $\frac{1}{\phi^k(h)} (\mathbb{E}(\Delta_1(x)))^k = M_1^k f_1^k(x) + o(1)$.

Proof of Lemma A1. For the proof of Lemma A1, the reader is directed to Ref. [66]. □

Lemma A2. Assume that the hypotheses (H1) and (H5), along with condition (H7), are satisfied. As $n \rightarrow \infty$, for every fixed neighborhood $N_{\mathcal{E}}$ of x in the functional space \mathcal{E} , we have:

$$\sum_{t=1}^n \lim_{n \rightarrow \infty} \sup_{x \in N_{\mathcal{E}}} \hat{f}_n(x) = \sum_{t=1}^n \lim_{n \rightarrow \infty} \mathbb{E}(\hat{f}_n(x)) = \mathcal{P}(x).$$

Proof of Lemma A2. We shall prove that

$$\hat{f}_n(x) \xrightarrow{\mathbb{P}} \mathcal{P}(x). \tag{A1}$$

We employ the identical proof as presented in Ref. [63]. See that.

$$\hat{f}_n(x) = R_{1,n}(x) + R_{2,n}(x),$$

where

$$R_{1,n}(x) = \frac{1}{n \mathbb{E}(\Delta_1(x))} \sum_{t=1}^n (\delta_t \Delta_t(x) - \mathbb{E}(\delta_t \Delta_t(x) \mid \mathfrak{F}_{t-1})),$$

$$R_{2,n}(x) = \frac{1}{n \mathbb{E}(\Delta_1(x))} \mathbb{E}[\delta_t \Delta_t(x) \mid \mathfrak{F}_{t-1}].$$

First, we need to establish under the assumption (H1)(i–iii) and (H3)(i) and for $n \rightarrow \infty$ as $n\phi(h) \rightarrow \infty$, we have

$$R_{2,n}(x) \xrightarrow{\mathbb{P}} \mathcal{P}(x),$$

as $n \rightarrow \infty$. Using the properties of conditional expectation and the missing at random (MAR) mechanism, and combining assumptions **(H1)**(ii,iii) and **(H3)**(i) with the continuity property of $\mathcal{P}(x)$ along with Lemma A1, we derive:

$$\begin{aligned}
 R_{2,n}(x) &= \frac{1}{n\mathbb{E}(\Delta_1(x))} \sum_{t=1}^n (\mathbb{E}\{\mathbb{E}[\delta_i\Delta_i(x) \mid \mathfrak{F}_{i-1}] \mid \mathfrak{G}_{i-1}\}) \\
 &= \frac{1}{n\mathbb{E}(\Delta_1(x))} \sum_{t=1}^n (\mathbb{E}[\mathcal{P}(x) + o(1)\delta_i\Delta_i(x) \mid \mathfrak{F}_{i-1}]) \\
 &= (\mathcal{P}(x) + o(1)) \frac{1}{n\mathbb{E}(\Delta_1(x))} \sum_{t=1}^n (\phi(h)M_1f_{i,1}(x) + \mathcal{O}(g_{ix}(h))) \\
 &= (\mathcal{P}(x) + o(1)) \frac{\phi(h)}{\mathbb{E}(\Delta_1(x))} \left(\frac{1}{n} \sum_{t=1}^n M_1f_{i,1}(x) + \frac{1}{n} \sum_{t=1}^n \mathcal{O}_{as}\left(\frac{g_{i,x}(h)}{\phi(h)}\right) \right) \\
 &= (\mathcal{P}(x) + o(1)) \frac{1}{M_1f_1(x) + o(1)} (M_1(f_1(x) + o(1)) + o_{as}(1)) \\
 &\rightarrow \mathcal{P}(x).
 \end{aligned}$$

Second, we will prove that as $n \rightarrow \infty$,

$$R_{1,n}(x) \xrightarrow{\mathbb{P}} 0.$$

On the one hand, we define $\eta_{n,i} = \delta_i\Delta(x) - \mathbb{E}(\delta_i\Delta(x))$ for $i = 1, \dots, n$. Thus, $\eta_{n,i}, 1 \leq i \leq n$ forms a triangular array of martingale differences with respect to the σ -field \mathfrak{F}_{i-1} and

$$R_{1,n}(x) = \frac{1}{n\mathbb{E}(\Delta_1(x))} \sum_{t=1}^n \eta_{n,i}(x).$$

By combining Burkholder’s inequality [91] and Jensen’s inequality, we establish that for any $\epsilon > 0$, there exists a constant C_0 such that:

$$\begin{aligned}
 \mathbb{P}(|R_{1,n}(x)| > \epsilon) &= \mathbb{P}\left(\left|\sum_{t=1}^n \eta_{n,i}(x)\right| > \epsilon n\mathbb{E}(\Delta_1(x))\right) \\
 &\leq C_0 \frac{E(\eta_{n,i}^2(x))}{\epsilon^2 n(\mathbb{E}(\Delta_1(x)))^2} < C_0 \frac{E(\delta_1\Delta_1^2(x))}{\epsilon^2 n(\mathbb{E}(\Delta_1^2(x)))} \rightarrow 0,
 \end{aligned}$$

as $n \rightarrow \infty$, where we use the results from lemma (A1). Since $n\phi(h) \rightarrow \infty$ as $n \rightarrow \infty$ we then conclude that

$$R_{1,n}(x) = o_{\mathbb{P}}(1).$$

Thus, the proof is complete. \square

We will utilize arguments akin to those employed in the work of Ref. [63] to establish the asymptotic normality of the process $Q_n(x)$ defined as:

$$Q_n(x) = [\widehat{\varphi}_n(C, x) - \mathbb{E}(\widehat{\varphi}_n(C, x))] - \mu_n(C, x) [\widehat{f}_n(x) - \mathbb{E}(\widehat{f}_n(x))].$$

Lemma A3. Assuming that the hypotheses **(H1)**–**(H7)** are fulfilled, we can state that for any $x \in \mathcal{E}$ such that $f_1(x) > 0$, we have:

$$\sqrt{n\phi(h_n)}Q_n(x) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_0^2(x)), \text{ as } n \rightarrow \infty. \tag{A2}$$

where

$$\sigma_0^2(x) = \frac{M_2W(x)\mathcal{P}(x)}{M_1^2f_1(x)},$$

whenever $f_1(x) > 0$

Proof of Lemma A3. Let us introduce some notation. We put

$$\eta_{ni} = \left(\frac{\phi(h)}{n}\right)^{1/2} \delta_i(\mathbb{1}_{\{Y_i \in C\}} - \mu(x)) \frac{\Delta_i(x)}{\mathbb{E}(\Delta_1(x))}, \tag{A3}$$

and define $\xi_{ni} = \eta_{ni} - \mathbb{E}(\eta_{ni} \mid \mathfrak{F}_{i-1})$. It is easily seen that

$$(n\phi(h))^{1/2} Q_n(x) = \sum_{t=1}^n \xi_{ni}. \tag{A4}$$

Here, for any fixed $x \in \mathcal{E}$, the terms in (A4) form a triangular array of stationary martingale differences with respect to the σ -field \mathfrak{F}_{i-1} . This allows us to apply the central limit theorem for discrete-time arrays of real-valued martingales (refer to Ref. [92], page 23) to establish the asymptotic normality of $Q_n(x)$. This can be accomplished by verifying the following statements:

(a)

$$\sum_{t=1}^n \mathbb{E}(\xi_{ni}^2 \mid \mathfrak{F}_{i-1}) \longrightarrow \sigma_0^2(x),$$

(b)

$$n\mathbb{E}(\xi_{ni}^2 \mathbb{1}_{|\eta_{ni}| > \epsilon}) = o(1),$$

holds for any $\epsilon > 0$ (Lindeberg condition).

Proof of Part (a). Observe first that

$$\left| \sum_{t=1}^n \mathbb{E}(\eta_{ni}^2 \mid \mathfrak{F}_{i-1}) - \sum_{t=1}^n \mathbb{E}(\xi_{ni}^2 \mid \mathfrak{F}_{i-1}) \right| \leq \sum_{t=1}^n (\mathbb{E}(\eta_{ni} \mid \mathfrak{F}_{i-1}))^2.$$

Making use of the condition (H2) and Lemma A1, one has

$$\begin{aligned} & \mathbb{E}(\eta_{ni} \mid \mathfrak{F}_{i-1}) \\ &= \frac{1}{\mathbb{E}(\Delta_i)} \left(\frac{\phi(h)}{n}\right)^{1/2} |\mathbb{E}((\mu(X_i) - \mu(x))\Delta_i(x)\mathcal{P}(X_i) \mid \mathfrak{F}_{i-1})| \\ &\leq \frac{1}{\mathbb{E}(\Delta_i)} \left(\frac{\phi(h)}{n}\right)^{1/2} \sum_{t=1}^n \sup_{u \in B(x,h)} |\mu(X_i) - \mu(x)| \mathbb{E}(\Delta_i(x) \mid \mathfrak{F}_{i-1}) h^\beta (o(1) + \mathcal{P}(x)) \\ &\leq \mathcal{O}(h^\beta) \left(\frac{\phi(h)}{n}\right)^{1/2} \left(\frac{f_{i,1}(x)}{f_1(x)} + \mathcal{O}_{a.s}\left(\frac{g_{i,x}(h)}{\phi(h)}\right)\right) h^\beta (o(1) + \mathcal{P}(x)). \end{aligned}$$

Thus, by (H1)(ii,iii), we have

$$\begin{aligned} \sum_{t=1}^n (\mathbb{E}(\eta_{ni} \mid \mathfrak{F}_{i-1}))^2 &= \mathcal{O}_{a.s}(h^{2\beta}\phi(h)) \left(\frac{1}{f_1^2(x)} \frac{1}{n} + \sum_{t=1}^n f_{i,1}^2(x) + o_{a.s}(1)\right) \\ &\quad \times (o(1) + \mathcal{P}(x))^2 \\ &= \mathcal{O}_{a.s}(\phi(h)h^{2\beta}). \end{aligned} \tag{A5}$$

The statement (a) follows then if we show that

$$\sum_{t=1}^n \lim_{n \rightarrow \infty} \sum_{t=1}^n \mathbb{E}(\eta_{ni}^2 \mid \mathfrak{F}_{i-1}) = \sigma_0^2. \tag{A6}$$

To prove (A6), observe that

$$\begin{aligned} \sum_{t=1}^n \lim_{n \rightarrow \infty} \sum_{t=1}^n \mathbb{E}(\eta_{ni}^2 \mid \mathfrak{F}_{i-1}) &= \frac{\phi(h)}{n(\mathbb{E}(\Delta_1(x)))^2} \sum_{t=1}^n \mathbb{E} \left[(\mathbb{1}_{\{Y_i \in C\}} - \mu(x))^2 \delta_i \Delta_i^2(x) \mid \mathfrak{F}_{i-1} \right] \\ &= J_{1n} + J_{2n}, \end{aligned}$$

where

$$J_{1n} = \frac{\phi(h)}{n(\mathbb{E}(\Delta_1(x)))^2} \sum_{t=1}^n \mathbb{E} \left[\mathbb{E} \left((\mathbb{1}_{\{Y_i \in C\}} - \mu(X_i))^2 \delta_i \Delta_i^2(x) \mid \mathfrak{F}_{i-1} \right) \right],$$

and

$$J_{2n} = \frac{\phi(h)}{n(\mathbb{E}(\Delta_1(x)))^2} \sum_{t=1}^n \mathbb{E} \left[(\mu(X_i) - \mu(X))^2 \delta_i \Delta_i^2(x) \mid \mathfrak{F}_{i-1} \right].$$

Hence, leveraging the properties of conditional expectation, we derive:

$$\begin{aligned} J_{1n} &= \frac{\phi(h)}{n(\mathbb{E}(\Delta_1(x)))^2} \sum_{t=1}^n \mathbb{E} \left\{ \mathbb{E} \left[(\mathbb{1}_{\{Y_i \in C\}} - \mu(X_i))^2 \delta_i \Delta_i^2(x) \mid \mathfrak{B}_{i-1} \right] \mid \mathfrak{F}_{i-1} \right\} \\ &= \frac{\phi(h)}{n(\mathbb{E}(\Delta_1(x)))^2} \sum_{t=1}^n \mathbb{E} \left\{ \Delta_i^2(x) \mathbb{E} \left[(\mathbb{1}_{\{Y_i \in C\}} - \mu(X_i))^2 \delta_i \mid X_i \right] \mid \mathfrak{F}_{i-1} \right\} \\ &= \frac{\phi(h)}{n(\mathbb{E}(\Delta_1(x)))^2} \sum_{t=1}^n \mathbb{E} \left[W(X_i) \mathcal{P}(X_i) \Delta_i^2(x) \mid \mathfrak{F}_{i-1} \right]. \end{aligned}$$

Likewise, with the assumptions (H2)(ii,iii) and (H4)(i), along with the aid of Lemma A1 once more, it follows that, as $n \rightarrow \infty$:

$$\begin{aligned} J_{1n} &= \frac{\phi(h)}{n(\mathbb{E}(\Delta_1(x)))^2} \sum_{t=1}^n \mathbb{E} \left[(o(1) + W(x))(o(1) + \mathcal{P}(x)) \Delta_i^2(x) \mid \mathfrak{F}_{i-1} \right] \\ &= \frac{1}{\frac{(\mathbb{E}(\Delta_1(x)))^2}{\phi^2(h)}} \frac{1}{n} \frac{1}{\phi(h)} \sum_{t=1}^n (o(1) + W(x))(o(1) + \mathcal{P}(x)) \\ &\quad \times (M_2 \phi(h) f_{i1}(x) + \mathcal{O}_{a.s.}(g_{i,x}(h))) \\ &\rightarrow \frac{M_2 W(x) \mathcal{P}(x)}{M_1^2 f_1(x)} = \sigma_0(x)^2. \end{aligned}$$

Again, combining Lemma A1 with conditions (H1)(ii), and (H3)(ii,iii), it is evident that:

$$\sum_{t=1}^n \lim_{n \rightarrow \infty} J_{1n} = \frac{M_2 W(x) \mathcal{P}(x)}{M_1^2 f_1(x)},$$

almost surely, whenever $f_1(x) > 0$. Consider now the term J_{2n} . Utilizing conditions (H1)(ii,iii) and (H2)(i) alongside Lemma A1, we can express, as $n \rightarrow \infty$:

$$\begin{aligned} |J_{2n}| &= \mathcal{O}(h^{2\beta}) \frac{\phi(h)}{n(\mathbb{E}(\Delta_1(x)))^2} \sum_{t=1}^n \mathbb{E} \left(\delta_i \Delta_i^2(x) \mid \mathfrak{F}_{i-1} \right) \\ &= \mathcal{O}(h^{2\beta}) \left(\frac{M_2}{M_1^2 f_1(x)} + o_{a.s.}(1) \right) \rightarrow 0, \text{ almost surely,} \end{aligned} \tag{A7}$$

whenever $f_1(x) > 0$, this completes the proof of Part (a).

Proof of Part (b). The Lindeberg condition results from Corollary 9.5.2 in Ref. ([93]), which implies that

$$n \mathbb{E}(\zeta_{ni}^2 \mathbb{1}(|\zeta_{ni}| > \varepsilon)) \leq 4n \mathbb{E} \left(\eta_{ni}^2 \mathbb{1}(|\eta_{ni}| > \varepsilon/2) \right).$$

Let $a > 1$ and $b > 1$ such that $\frac{1}{a} + \frac{1}{b} = 1$. Applying Hölder and Markov inequalities, one can express, for all $\varepsilon > 0$:

$$\mathbb{E}\left(\eta_{ni}^2 \mathbb{1}(|\eta_{ni}| > \varepsilon/2)\right) \leq \frac{\mathbb{E}|\eta_{ni}|^{2a}}{(\varepsilon/2)^{2a/b}},$$

where C_0 is a positive constant and $2a = 2 + \delta$. Utilizing δ from the condition **(H3)**(iii) of conditional moments, we obtain:

$$\begin{aligned} & 4n\mathbb{E}\left(\eta_{ni}^2 \mathbb{1}(|\eta_{ni}| > \varepsilon/2)\right) \\ & \leq C_0 \left(\frac{\phi(h)}{n}\right)^{(2+\delta)/2} \frac{n}{(\mathbb{E}(\Delta_1(x)))^{2+\delta}} \mathbb{E}([\mathbb{1}_{\{Y_i \in C\}} - \mu(x)|\delta_i \Delta_i(x)]^{2+\delta}) \\ & \leq C_0 \left(\frac{\phi(h)}{n}\right)^{(2+\delta)/2} \frac{n}{(\mathbb{E}(\Delta_1(x)))^{2+\delta}} \mathbb{E}\left(\mathbb{E}\left(|\mathbb{1}_{\{Y_i \in C\}} - \mu(x)|^{2+\delta} \delta_i (\Delta_i(x))^{2+\delta} \mid X_i\right)\right) \\ & \leq C_0 \left(\frac{\phi(h)}{n}\right)^{(2+\delta)/2} \frac{n}{(\mathbb{E}(\Delta_1(x)))^{2+\delta}} \mathbb{E}\left((\Delta_i(x))^{2+\delta} \mathcal{P}(X_i) \overline{W}_{2+\delta}(X_i)\right) \\ & \leq C_0 \left(\frac{\phi(h)}{n}\right)^{(2+\delta)/2} \frac{n}{\mathbb{E}(\Delta_1(x))^{2+\delta}} \mathbb{E}\left[(\Delta_1(x))^{2+\delta} (\mathcal{P}(x) + o(1)) (\overline{W}_{2+\delta}(x) + o(1))\right] \\ & \leq C_0 (n\phi(h))^{-\delta/2} \frac{(M_{2+\delta} f_1(x) + o(1))}{(M_1^{2+\delta} f_1^{2+\delta}(x) + o(1))} (\mathcal{P}(x) \overline{W}_{2+\delta}(x) + o(1)) \\ & = \mathcal{O}((n\phi(h))^{-\delta/2}), \end{aligned}$$

where the last equality follows from Lemma A1. This concludes the proof of part (b) as $n\phi(h) \rightarrow \infty$ when $n \rightarrow \infty$. Thus, the proof is complete. \square

Proof of Theorem 1. By Lemma A3 it follows that

$$\sqrt{nf(h_n)}Q_n(x) = \mathcal{O}_{\mathbb{P}}(1).$$

Thus, by Lemma A2 the proof is valid. \square

Proof of Theorem 2. The proof follows from A1, A2, and Slutsky’s Theorem, so the proof is valid. \square

Proof of Theorem 3. Let us recall some facts. Let $f(\cdot) = \delta_i \mathbb{1}\{\cdot \in C_1\}$ and $g(\cdot) = \delta_i \mathbb{1}\{\cdot \in C_2\}$. Given random measures λ on $(\mathbf{X}, \mathcal{X})$, we define

$$d_\lambda^{(2)}(f, g) := \left[\lambda(f - g)^2\right]^{1/2}.$$

Say that a class of functions \mathcal{F} has uniformly integrable entropy with respect to \mathcal{L}_2 -norm if

$$\int_0^\infty \sup_{\gamma \in M(\mathbf{X}, \mathcal{F})} \left[\ln N\left(\epsilon \left[\gamma(F^2)\right]^{1/2}, \mathcal{F}, d_\gamma^{(2)}\right)\right]^{1/2} d\epsilon < \infty,$$

where

$$d_\gamma^{(2)}(f, g) := \left[\int_{\mathbf{X}} (f - g)^2 d\gamma\right]^{1/2}.$$

If the class \mathcal{F} possesses uniformly integrable entropy, $(\mathcal{F}, d_\gamma^{(2)})$ is totally bounded for any measure γ . Let κ be an envelope of \mathcal{F} , i.e., κ is a measurable function mapping \mathcal{F} to $[0, \infty)$ such that:

$$\sup_{f \in \mathcal{F}} |f(t)| \leq \kappa(t), \quad \text{for all } t \in \mathbb{R}.$$

Let $M(\mathbb{R}, \kappa)$ be the set of all measures γ on $(\mathbb{R}, \mathcal{F})$ with

$$\gamma(\kappa) := \int_{\mathbb{R}} \kappa^2 d\gamma < \infty, \tag{A8}$$

and

$$d_\gamma^{(r)}(f, g) := \left[\int_{\mathbb{R}} (f - g)^r d\gamma \right]^{1/r}.$$

Given random measures λ on $(\mathbb{R}, \mathcal{F})$, we define

$$d_\lambda^{(2)}(f, g) := [\lambda(f - g)^2]^{1/2}.$$

Let us introduce the uniform entropy integral

$$J(\delta, \mathcal{F}, d_\gamma^{(2)}) = \int_0^\delta \sup_{\gamma \in (\mathbb{R}, \mathcal{F})} \left[\log \left(\mathcal{N} \left(\epsilon [\gamma(\kappa^2)]^{1/2}, \mathcal{F}, d_\gamma^{(2)} \right) \right) \right]^{1/2} d\epsilon.$$

We say that \mathcal{F} has uniformly integrable entropy with respect to L_2 -norm if

$$J(\infty, \mathcal{F}, d_\gamma^{(2)}) < \infty. \tag{A9}$$

If the class \mathcal{F} possesses uniformly integrable entropy, $(\mathcal{F}, d_\gamma^{(2)})$ is totally bounded for any measure γ . Let $\mathbb{B}(\varphi) : \varphi \in \mathcal{F}$ be a Gaussian process whose sample paths are contained in

$$U_b(\mathcal{F}, d_\gamma^{(2)}) := \left\{ f \in \ell^\infty(\mathcal{F}) : f \text{ is uniformly continuous with respect to } d_\gamma^{(2)} \right\}.$$

Let $\mathcal{L}(\bullet)$ denote the law of \bullet . Notice that obtaining a uniform CLT essentially means that we show the following convergence

$$\left\{ \mathcal{L}(A_{n,\varphi}) : \varphi \in \mathcal{F} \right\} \rightarrow \left\{ \mathcal{L}(\mathbb{B}(\varphi)) : \varphi \in \mathcal{F} \right\},$$

where the processes are indexed by \mathcal{F} and considered as random elements of the bounded real-valued functions on \mathcal{F} defined by

$$\ell^\infty(\mathcal{F}) := \left\{ f : \mathcal{F} \rightarrow \mathbb{R} : \|f\|_{\mathcal{F}} := \sup_{\varphi \in \mathcal{F}} |f(\varphi)| < \infty \right\}, \tag{A10}$$

which is a Banach space equipped with the sup norm. In the following, we employ the weak convergence in the sense of Ref. [94], which we recap in the following definition. Throughout the paper, \mathbb{E}^* denotes the upper expectation with respect to the outer probability \mathbb{P}^* ; for further details and discussion, refer to Ref. [1] (p. 6) and Ref. [95] (§6.2, p. 88). □

Definition A1. A sequence of $\ell^\infty(\mathcal{F})$ -valued random functions $\{T_n : n \geq 1\}$ converges in law to a $\ell^\infty(\mathcal{F})$ -valued Borel measurable random function T whose law concentrates on a separable subset of $\ell^\infty(\mathcal{F})$, denoted $T_n \rightsquigarrow T$, if,

$$\mathbb{E}g(T) = \lim_{n \rightarrow \infty} \mathbb{E}^*g(T_n), \quad \forall g \in C(\ell^\infty(\mathcal{F}), \|\cdot\|_{\mathcal{F}}),$$

where $C(\ell^\infty(\mathcal{F}), \|\cdot\|_{\mathcal{F}})$ is the set of all bounded $\|\cdot\|_{\mathcal{F}}$ -continuous functions from $(\ell^\infty(\mathcal{F}), \|\cdot\|_{\mathcal{F}})$ into \mathbb{R} .

We set

$$\eta_{n,i}(f, x) := \eta_{n,i}(C_1, x) := \left(\frac{\phi(h)}{n} \right)^{1/2} \left(\delta_i \mathbb{1}_{\{Y_i \in C_1\}} - \mu(C, x) \right) \frac{\Delta_i(x)}{\mathbb{E}(\Delta_i(x))},$$

with $\Delta_i(x) = K(h^{-1}d(x, X_i))$, and define $\eta_{n,i}(g, x)$ in a similar way. Let

$$\xi_{n,i}(f, x) := \eta_{n,i}(f, x) - \mathbb{E}(\eta_{n,i}(f, x) \mid \mathfrak{F}_{i-1}).$$

Let us define

$$\sigma_n^2(f, g) = \sum_{i=1}^n (\xi_{n,i}(f, x) - \xi_{n,i}(g, x))^2.$$

To establish Theorem 3, we can rely on Theorem 2 of [96] (see also Refs. [10,13,15]). It is sufficient to demonstrate that, for all constant $L > 0$, as n tends to infinity:

$$\mathbb{P}^* \left\{ \sup_{f, g \in \mathcal{F}} \frac{\sigma_n^2(f, g)}{(d_{\mu_n}^{(2)}(f, g))^2} > L \right\} \rightarrow 0, \tag{A11}$$

which is implied by the following,

$$\mathbb{E}^* \sup_{d^{(2)}(f, g) \leq \delta_n} \sum_{i=1}^n \frac{\mathbb{E}((\xi_{n,i}(f, x) - \xi_{n,i}(g, x))^2 \mid \mathfrak{F}_{i-1})}{(d^{(2)}(f, g))^2} \rightarrow 0, \text{ as } \delta_n \rightarrow 0,$$

where we recall

$$d^{(2)}(f, g) := \left[\int_{\mathbb{R}} (f - g)^2 d\mathbb{P} \right]^{1/2}.$$

In the rest of the proof, denote by $\beta_n(x) = \frac{\sqrt{\phi(h)}}{\mathbb{E}[\Delta_1(x)]}$, and

$$\zeta(f, x) = \zeta(C_1, x) := \left(\delta_i \mathbb{1}_{\{Y_i \in C_1\}} - \mu(C, x) \right) \Delta_i(x).$$

Therefore, we have the following

$$\begin{aligned} & \sum_{i=1}^n \frac{\mathbb{E}((\xi_{n,i}(f, x) - \xi_{n,i}(g, x))^2 \mid \mathfrak{F}_{i-1})}{d^{(2)}(f, g)} \\ &= \frac{\beta_n^2(x)}{nd^{(2)}(f, g)} \sum_{i=1}^n \mathbb{E} \left[\left((\zeta(f, x) - \zeta(g, x)) - \mathbb{E}[\zeta(f, x) - \zeta(g, x) \mid \mathcal{F}_{i-1}] \right)^2 \mid \mathfrak{F}_{i-1} \right] \\ &\leq \frac{\beta_n^2(x)}{nd^{(2)}(f, g)} \sum_{i=1}^n 2\mathbb{E} \left[\left(\zeta(f, x) - \zeta(g, x) \right)^2 \mid \mathfrak{F}_{i-1} \right] \\ &\quad - 2\mathbb{E} \left\{ \left[\mathbb{E} \left[\left(\zeta(f, x) - \zeta(g, x) \right) \mid \mathfrak{F}_{i-1} \right] \right]^2 \right\} \\ &:= T_{1,n} + T_{2,n}. \end{aligned}$$

We first evaluate $T_{1,n}$. We have

$$\begin{aligned} T_{1,n} &\leq \frac{2\beta_n^2(x)}{nd^{(2)}(f, g)} \sum_{i=1}^n 2\mathbb{E} \left[\Delta_i^2(x) (\delta_i f(Y_i) - \delta_i g(Y_i))^2 \mid \mathfrak{F}_{i-1} \right] \\ &\quad + 2\mathbb{E} \left[\delta_i \Delta_i^2(x) (\mu(C_1, x) - \mu(C_2, x))^2 \mid \mathfrak{F}_{i-1} \right] \\ &:= T_{1,n,1} + T_{1,n,2}. \end{aligned}$$

Using the fact that $\mathbb{E}(\Delta_1^2(x)) = \mathcal{O}(\phi(h))$ (as indicated in Lemma A1), and taking into account that the class of functions \mathcal{F} has a constant envelope and $K(\cdot)$ is both bounded and bounded away from zero, one can obtain the following upper bound for the last equation, where C is a positive constant:

$$\begin{aligned}
 T_{1,n,1} &\leq \frac{C\sqrt{\phi(h)}}{d^{(2)}(f,g)} \mathbb{E}[\Delta_1(x)(f(Y_1) - g(Y_1))] \\
 &\leq \frac{C\sqrt{\phi(h)}}{d^{(2)}(f,g)} \mathbb{E}[\Delta_1(x)^2]^{1/2} \mathbb{E}[(f(Y_1) - g(Y_1))^2]^{1/2} \\
 &= \frac{C\sqrt{\phi(h)}}{\bar{G}^2(\zeta)} \mathbb{E}[\Delta_1(x)^2]^{1/2} \\
 &= \mathcal{O}(\phi(h)) = o(1).
 \end{aligned}$$

Making use of similar arguments, we infer that

$$T_{1,n,2} = \frac{C\phi(h)^{3/2}}{d^{(2)}(f,g)} (\mathbb{E}[\delta(f(Y) - g(Y))|X = x])^2 = \mathcal{O}(\phi(h)^{3/2}) = o(1).$$

We readily obtain that,

$$T_{1,n} = o(1).$$

By employing arguments akin to those utilized in the proof of the previous statement, we can establish that

$$T_{2,n} = o(1).$$

Using the Lindeberg conditions from the preceding proof and (A11), along with Theorem 1 of [96], we deduce that for a given $\varepsilon > 0$ and $\gamma > 0$, there exists $\eta > 0$, such that:

$$\limsup_{n \rightarrow \infty} \mathbb{P}^* \left\{ \sup_{d(C_1, C_2) \leq \eta} |\nu_n(C_1, x) - \nu_n(C_2, x)| \geq 5\gamma \right\} \leq 3\varepsilon. \tag{A12}$$

Now, the proof of the theorem is completed by combining this last equation with Theorem 3.

References

- van der Vaart, A.W.; Wellner, J.A. *Weak Convergence and Empirical Processes*; Springer Series in Statistics; With applications to statistics; Springer: New York, NY, USA, 1996; pp. xvi+508. [\[CrossRef\]](#)
- Shorack, G.R.; Wellner, J.A. *Empirical Processes with Applications to Statistics*; Classics in Applied Mathematics; Society for Industrial and Applied Mathematics (SIAM): Philadelphia, PA, USA, 2009; Volume 59, pp. xli+956. [\[CrossRef\]](#)
- Dudley, R.M. *Uniform Central Limit Theorems*; Cambridge Studies in Advanced Mathematics; Cambridge University Press: Cambridge, UK, 1999; Volume 63, pp. xiv+436. [\[CrossRef\]](#)
- Vapnik, V.N.; Červonenkis, A.J. The uniform convergence of frequencies of the appearance of events to their probabilities. *Teor. Veroyatnost. i Primenen.* **1971**, *16*, 264–279.
- Dudley, R.M. Central limit theorems for empirical measures. *Ann. Probab.* **1978**, *6*, 899–929. [\[CrossRef\]](#)
- Giné, E.; Zinn, J. Some limit theorems for empirical processes. *Ann. Probab.* **1984**, *12*, 929–998. [\[CrossRef\]](#)
- Le Cam, L. A remark on empirical measures. In *A Festschrift for Erich Lehmann in Honor of His Sixty-Fifth Birthday*; Wadsworth Statist./Probab. Ser.; UC Berkeley Statistics: Wadsworth, OH, USA; Belmont, CA, USA, 1983; pp. 305–327.
- Pollard, D. A central limit theorem for empirical processes. *J. Aust. Math. Soc. Ser. A* **1982**, *33*, 235–248. [\[CrossRef\]](#)
- Bass, R.F.; Pyke, R. A strong law of large numbers for partial-sum processes indexed by sets. *Ann. Probab.* **1984**, *12*, 268–271. [\[CrossRef\]](#)
- Bouzebda, S.; Soukariéh, I. Renewal type bootstrap for U -process Markov chains. *Markov Process. Relat. Fields* **2022**, *28*, 673–735.
- Alvarez-Andrade, S.; Bouzebda, S.; Lachal, A. Strong approximations for the p -fold integrated empirical process with applications to statistical tests. *Test* **2018**, *27*, 826–849. [\[CrossRef\]](#)
- Bouzebda, S. Some applications of the strong approximation of the integrated empirical copula processes. *Math. Methods Stat.* **2016**, *25*, 281–303. [\[CrossRef\]](#)
- Soukariéh, I.; Bouzebda, S. Renewal type bootstrap for increasing degree U -process of a Markov chain. *J. Multivar. Anal.* **2023**, *195*, 105143. [\[CrossRef\]](#)
- Bouzebda, S.; Soukariéh, I. Limit theorems for a class of processes generalizing the U -empirical process. *Stochastics* **2024**, 1–36.
- Soukariéh, I.; Bouzebda, S. Exchangeably Weighted Bootstraps of General Markov U -Process. *Mathematics* **2022**, *10*, 3745. [\[CrossRef\]](#)
- Yoshihara, K.I. Conditional empirical processes defined by ϕ -mixing sequences. *Comput. Math. Appl.* **1990**, *19*, 149–158. [\[CrossRef\]](#)

17. Eberlein, E. Weak convergence of partial sums of absolutely regular sequences. *Stat. Probab. Lett.* **1984**, *2*, 291–293. [[CrossRef](#)]
18. Nobel, A.; Dembo, A. A note on uniform laws of averages for dependent processes. *Stat. Probab. Lett.* **1993**, *17*, 169–172. [[CrossRef](#)]
19. Yu, B. Rates of convergence for empirical processes of stationary mixing sequences. *Ann. Probab.* **1994**, *22*, 94–116. [[CrossRef](#)]
20. Bouzebda, S.; Nemouchi, B. Central limit theorems for conditional empirical and conditional U -processes of stationary mixing sequences. *Math. Methods Stat.* **2019**, *28*, 169–207. [[CrossRef](#)]
21. Andrews, D.W.K.; Pollard, D. An Introduction to Functional Central Limit Theorems for Dependent Stochastic Processes. *Int. Stat. Rev. Rev. Int. Stat.* **1994**, *62*, 119–132. [[CrossRef](#)]
22. Doukhan, P.; Massart, P.; Rio, E. Invariance principles for absolutely regular empirical processes. *Ann. Inst. H. Poincaré Probab. Stat.* **1995**, *31*, 393–427.
23. Polonik, W.; Yao, Q. Set-indexed conditional empirical and quantile processes based on dependent data. *J. Multivar. Anal.* **2002**, *80*, 234–255. [[CrossRef](#)]
24. Bosq, D. *Linear Processes in Function Spaces; Lecture Notes in Statistics; Theory and Applications; Springer: New York, NY, USA, 2000; Volume 149, pp. xiv+283.* [[CrossRef](#)]
25. Ramsay, J.O.; Silverman, B.W. *Functional Data Analysis, 2nd ed.; Springer Series in Statistics; Springer: New York, NY, USA, 2005; pp. xx+426.*
26. Cuevas, A. A partial overview of the theory of statistics with functional data. *J. Stat. Plan. Inference* **2014**, *147*, 1–23. [[CrossRef](#)]
27. Goia, A.; Vieu, P. An introduction to recent advances in high/infinite dimensional statistics [Editorial]. *J. Multivar. Anal.* **2016**, *146*, 1–6. [[CrossRef](#)]
28. Aneiros, G.; Cao, R.; Fraiman, R.; Genest, C.; Vieu, P. Recent advances in functional data analysis and high-dimensional statistics. *J. Multivar. Anal.* **2019**, *170*, 3–9. [[CrossRef](#)]
29. Ling, N.; Vieu, P. Nonparametric modelling for functional data: Selected survey and tracks for future. *Statistics* **2018**, *52*, 934–949. [[CrossRef](#)]
30. Chowdhury, J.; Chaudhuri, P. Multi-sample comparison using spatial signs for infinite dimensional data. *Electron. J. Stat.* **2022**, *16*, 4636–4678. [[CrossRef](#)]
31. Chowdhury, J.; Chaudhuri, P. Convergence rates for kernel regression in infinite-dimensional spaces. *Ann. Inst. Stat. Math.* **2020**, *72*, 471–509. [[CrossRef](#)]
32. Ferraty, F.; Vieu, P. *Nonparametric Functional Data Analysis; Springer Series in Statistics; Theory and Practice; Springer: New York, NY, USA, 2006; pp. xx+258.*
33. Horváth, L.; Kokoszka, P. *Inference for Functional Data with Applications; Springer Series in Statistics; Springer: New York, NY, USA, 2012; pp. xiv+422.* [[CrossRef](#)]
34. Bosq, D.; Blanke, D. *Inference and Prediction in Large Dimensions; Wiley Series in Probability and Statistics; John Wiley & Sons, Ltd.: Chichester, UK; Dunod, Scotland; Paris, France, 2007; pp. x+316.* [[CrossRef](#)]
35. Shi, J.Q.; Choi, T. *Gaussian Process Regression Analysis for Functional Data; CRC Press: Boca Raton, FL, USA, 2011; pp. xx+196.*
36. Zhang, J.T. *Analysis of Variance for Functional Data; Monographs on Statistics and Applied Probability; CRC Press: Boca Raton, FL, USA, 2014; Volume 127, pp. xxiv+386.*
37. Bongiorno, E.G.; Goia, A.; Salinelli, E.; Vieu, P. An overview of IWFOs'2014. In *Contributions in Infinite-Dimensional Statistics and Related Topics; Esculapio: Bologna, Italy, 2014; pp. 1–5.*
38. Hsing, T.; Eubank, R. *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators; Wiley Series in Probability and Statistics; John Wiley & Sons, Ltd.: Chichester, UK, 2015; pp. xiv+334.* [[CrossRef](#)]
39. Aneiros, G.; Bongiorno, E.G.; Cao, R.; Vieu, P. (Eds.) Functional statistics and related fields. In *Proceedings of the 4th International Workshop on Functional and Operational Statistics, IWFOs, Corunna, Spain, 15–17 June 2017; Springer: Cham, Switzerland, 2017; pp. xxiv+288.*
40. Berrahou, N.; Bouzebda, S.; Douge, L. Functional uniform-in-bandwidth moderate deviation principle for the local empirical processes involving functional data. *Math. Methods Stat.* **2024**, *33*, 1–43.
41. Poryvaĭ, D.V. An invariance principle for conditional empirical processes formed by dependent random variables. *Izv. Ross. Akad. Nauk Ser. Mat.* **2005**, *69*, 129–148. [[CrossRef](#)]
42. Bouzebda, S.; Madani, F.; Souddi, Y. Some Asymptotic Properties of the Conditional Set-Indexed Empirical Process Based on Dependent Functional Data. *Int. J. Math. Stat.* **2022**, *22*, 77–105.
43. Bouzebda, S.; Chaouch, M. Uniform limit theorems for a class of conditional Z -estimators when covariates are functions. *J. Multivar. Anal.* **2022**, *189*, 104872. [[CrossRef](#)]
44. Souddi, Y.; Madani, F.; Bouzebda, S. Some characteristics of the conditional set-indexed empirical process involving functional ergodic data. *Bull. Inst. Math. Acad. Sin. (New Ser.)* **2021**, *16*, 367–399. [[CrossRef](#)]
45. Bouzebda, S.; Soukarieh, I. Nonparametric conditional U -processes for locally stationary functional random fields under stochastic sampling design. *Mathematics* **2022**, *10*, 16. [[CrossRef](#)]
46. Soukarieh, I.; Bouzebda, S. Weak Convergence of the Conditional U -statistics for Locally Stationary Functional Time Series. *Stat. Inference Stoch. Process* **2024**, *16*, 1–78. [[CrossRef](#)]
47. Bouzebda, S.; Nezzal, A. Uniform in number of neighbors consistency and weak convergence of k NN empirical conditional processes and k NN conditional U -processes involving functional mixing data. *AIMS Math.* **2024**, *9*, 4427–4550. [[CrossRef](#)]

48. Cheng, P.E. Nonparametric estimation of mean functionals with data missing at random. *J. Am. Stat. Assoc.* **1994**, *89*, 81–87. [[CrossRef](#)]
49. Cheng, P.E.; Chu, C.K. Kernel estimation of distribution functions and quantiles with missing data. *Stat. Sin.* **1996**, *6*, 63–78.
50. Little, R.J.A.; Rubin, D.B. *Statistical Analysis with Missing Data*; Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics; John Wiley & Sons, Inc.: New York, NY, USA, 1987; pp. xvi+278.
51. Nittner, T. Missing at random (MAR) in nonparametric regression—A simulation experiment. *Stat. Methods Appl.* **2003**, *12*, 195–210. [[CrossRef](#)]
52. Tsiatis, A.A. *Semiparametric Theory and Missing Data*; Springer Series in Statistics; Springer: New York, NY, USA, 2006; pp. xvi+383.
53. Wang, Q.; Sun, Z. Estimation in partially linear models with missing responses at random. *J. Multivar. Anal.* **2007**, *98*, 1470–1493. [[CrossRef](#)]
54. Wang, Q. Probability density estimation with data missing at random when covariables are present. *J. Stat. Plan. Inference* **2008**, *138*, 568–587. [[CrossRef](#)]
55. Liang, H.; Wang, S.; Carroll, R.J. Partially linear models with missing response variables and error-prone covariates. *Biometrika* **2007**, *94*, 185–198. [[CrossRef](#)]
56. Efromovich, S. Nonparametric regression with responses missing at random. *J. Stat. Plan. Inference* **2011**, *141*, 3744–3752. [[CrossRef](#)]
57. Efromovich, S. Nonparametric regression with predictors missing at random. *J. Am. Stat. Assoc.* **2011**, *106*, 306–319. [[CrossRef](#)]
58. Tang, N.; Zhao, P.; Zhu, H. Empirical likelihood for estimating equations with nonignorable missing data. *Stat. Sin.* **2014**, *24*, 723–747. [[CrossRef](#)]
59. Müller, U.U.; Schick, A. Efficiency transfer for regression models with responses missing at random. *Bernoulli* **2017**, *23*, 2693–2719. [[CrossRef](#)]
60. Müller, U.U.; Schick, A. Efficiency for heteroscedastic regression with responses missing at random. *J. Stat. Plan. Inference* **2018**, *196*, 132–143. [[CrossRef](#)]
61. Shen, Y.; Liang, H.Y. Quantile regression and its empirical likelihood with missing response at random. *Stat. Pap.* **2018**, *59*, 685–707. [[CrossRef](#)]
62. Ferraty, F.; Sued, M.; Vieu, P. Mean estimation with data missing at random for functional covariables. *Statistics* **2013**, *47*, 688–706. [[CrossRef](#)]
63. Ling, N.; Liang, L.; Vieu, P. Nonparametric regression estimation for functional stationary ergodic data with missing at random. *J. Stat. Plan. Inference* **2015**, *162*, 75–87. [[CrossRef](#)]
64. Ling, N.; Liu, Y.; Vieu, P. Conditional mode estimation for functional stationary ergodic data with responses missing at random. *Statistics* **2016**, *50*, 991–1013. [[CrossRef](#)]
65. Wang, L.; Cao, R.; Du, J.; Zhang, Z. A nonparametric inverse probability weighted estimation for functional data with missing response data at random. *J. Korean Stat. Soc.* **2019**, *48*, 537–546. [[CrossRef](#)]
66. Laib, N.; Louani, D. Nonparametric kernel regression estimation for functional stationary ergodic data: Asymptotic properties. *J. Multivar. Anal.* **2010**, *101*, 2266–2281. [[CrossRef](#)]
67. Didi, S.; Bouzebda, S. Wavelet Density and Regression Estimators for Continuous Time Functional Stationary and Ergodic Processes. *Mathematics* **2022**, *10*, 4356. [[CrossRef](#)]
68. Didi, S.; Al Harby, A.; Bouzebda, S. Wavelet Density and Regression Estimators for Functional Stationary and Ergodic Data: Discrete Time. *Mathematics* **2022**, *10*, 3433. [[CrossRef](#)]
69. Nadaraja, E.A. On a regression estimate. *Teor. Veroyatnost. i Primen.* **1964**, *9*, 157–159.
70. Watson, G.S. Smooth regression analysis. *Sankhyā Ser. A* **1964**, *26*, 359–372.
71. Stute, W. Conditional empirical processes. *Ann. Stat.* **1986**, *14*, 638–647. [[CrossRef](#)]
72. Stute, W. On almost sure convergence of conditional empirical distribution functions. *Ann. Probab.* **1986**, *14*, 891–901. [[CrossRef](#)]
73. Horváth, L.; Yandell, B.S. Asymptotics of conditional empirical processes. *J. Multivar. Anal.* **1988**, *26*, 184–206. [[CrossRef](#)]
74. Ferraty, F.; Mas, A.; Vieu, P. Nonparametric regression on functional data: Inference and practical aspects. *Aust. N. Z. J. Stat.* **2007**, *49*, 267–286. [[CrossRef](#)]
75. Dudley, R.M. A course on empirical processes. In *École d'été de Probabilités de Saint-Flour, XII—1982*; Lecture Notes in Mathematics; Springer: Berlin, Germany, 1984; Volume 1097, pp. 1–142. [[CrossRef](#)]
76. Billingsley, P. *Convergence of Probability Measures*, 2nd ed.; Wiley Series in Probability and Statistics: Probability and Statistics; A Wiley-Interscience Publication; John Wiley & Sons, Inc.: New York, NY, USA, 1999; pp. x+277. [[CrossRef](#)]
77. Huber, P.J. *Robust Statistics*; Wiley Series in Probability and Mathematical Statistics; John Wiley & Sons, Inc.: New York, NY, USA, 1981; pp. ix+308.
78. Parthasarathy, K.R. *Probability Measures on Metric Spaces*; Reprint of the 1967 original; AMS Chelsea Publishing: Providence, RI, USA, 2005; pp. xii+276. [[CrossRef](#)]
79. Hofinger, A. The metrics of Prokhorov and Ky Fan for assessing uncertainty in inverse problems. *Österreich. Akad. Wiss. Math.-Natur. Kl. Sitzungsber. II* **2006**, *215*, 107–125. [[CrossRef](#)]
80. Fan, K. Entfernung zweier zufälligen Grössen und die Konvergenz nach Wahrscheinlichkeit. *Math. Z.* **1944**, *49*, 681–683. [[CrossRef](#)]

81. Bouzebda, S.; Nemouchi, B. Uniform consistency and uniform in bandwidth consistency for nonparametric regression estimates and conditional U -statistics involving functional data. *J. Nonparametr. Stat.* **2020**, *32*, 452–509. [[CrossRef](#)]
82. Hall, P. Asymptotic properties of integrated square error and cross-validation for kernel estimation of a regression function. *Z. Wahrsch. Verw. Geb.* **1984**, *67*, 175–196. [[CrossRef](#)]
83. Rachdi, M.; Vieu, P. Nonparametric regression for functional data: Automatic smoothing parameter selection. *J. Stat. Plan. Inference* **2007**, *137*, 2784–2801. [[CrossRef](#)]
84. Dony, J.; Mason, D.M. Uniform in bandwidth consistency of conditional U -statistics. *Bernoulli* **2008**, *14*, 1108–1133. [[CrossRef](#)]
85. Bouzebda, S. On the weak convergence and the uniform-in-bandwidth consistency of the general conditional U -processes based on the copula representation: Multivariate setting. *Hacet. J. Math. Stat.* **2023**, *52*, 1303–1348. [[CrossRef](#)]
86. Bouzebda, S.; Taachouche, N. On the variable bandwidth kernel estimation of conditional U -statistics at optimal rates in sup-norm. *Phys. A* **2023**, *625*, 129000. [[CrossRef](#)]
87. Bouzebda, S. General tests of conditional independence based on empirical processes indexed by functions. *Jpn. J. Stat. Data Sci.* **2023**, *6*, 115–177. [[CrossRef](#)]
88. Shang, H.L. Bayesian bandwidth estimation for a functional nonparametric regression model with mixed types of regressors and unknown error density. *J. Nonparametr. Stat.* **2014**, *26*, 599–615. [[CrossRef](#)]
89. Bouzebda, S.; Laksaci, A.; Mohammedi, M. The k -nearest neighbors method in single index regression model for functional quasi-associated time series data. *Rev. Mat. Complut.* **2023**, *36*, 361–391. [[CrossRef](#)]
90. Bouzebda, S.; Laksaci, A.; Mohammedi, M. Single index regression model for functional quasi-associated time series data. *Revstat* **2022**, *20*, 605–631.
91. Hall, P.; Heyde, C.C. *Martingale Limit Theory and Its Application*; Probability and Mathematical Statistics; Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers]: New York, NY, USA; London, UK, 1980; pp. xii+308.
92. Györfi, L.; Morvai, G.; Yakowitz, S.J. Limits to consistent on-line forecasting for ergodic time series. *IEEE Trans. Inf. Theory* **1998**, *44*, 886–892. [[CrossRef](#)]
93. Chow, Y.S.; Teicher, H. *Probability Theory*, 3rd ed.; Springer Texts in Statistics; Independence, interchangeability, martingales; Springer: New York, NY, USA, 1997; pp. xxii+488. [[CrossRef](#)]
94. Hoffmann-Jørgensen, J. *Stochastic Processes on Polish Spaces*; Various Publications Series (Aarhus); Aarhus Universitet, Matematisk Institut: Aarhus, Denmark, 1991; Volume 39, pp. ii+278.
95. Kosorok, M.R. *Introduction to Empirical Processes and Semiparametric Inference*; Springer Series in Statistics; Springer: New York, NY, USA, 2008; pp. xiv+483. [[CrossRef](#)]
96. Bae, J.; Jun, D.; Levental, S. The uniform CLT for martingale difference arrays under the uniformly integrable entropy. *Bull. Korean Math. Soc.* **2010**, *47*, 39–51. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.