Electronic Journal of Statistics
Vol. 16 (2022) 3385–3429
ISSN: 1935-7524
https://doi.org/10.1214/22-EJS2023

Empirical process theory for nonsmooth functions under functional dependence

Nathawut Phandoidaen[1] **and Stefan Richter**[1]

[1] *Institut für angewandte Mathematik, Im Neuenheimer Feld 205, Universität Heidelberg*
e-mail: phandoidaen@math.uni-heidelberg.de, *e-mail:* stefan.richter@iwr.uni-heidelberg.de

Abstract: We provide an empirical process theory for locally stationary processes over nonsmooth function classes. An important novelty over other approaches is the use of the flexible functional dependence measure to quantify dependence. A functional central limit theorem and nonasymptotic maximal inequalities are provided. The theory is used to prove the functional convergence of the empirical distribution function (EDF) and to derive uniform convergence rates for kernel density estimators both for stationary and locally stationary processes. A comparison with earlier results based on other measures of dependence is carried out.

MSC 2010 subject classifications: Primary 60F17; secondary 60F10.
Keywords and phrases: Empirical process theory, functional dependence measure, maximal inequality, functional central limit theorem, locally stationary processes.

Received August 2021.

1. Introduction

Empirical process theory is one of the key concepts in proving uniform convergence rates and weak convergence of composite functionals. It is preferable to have a theory which can be applied to observations which are dependent but also nonstationary. Locally stationary processes allow for a smooth change of the distribution over time but can locally be approximated by stationary processes and thus provide more flexible time series models (cf. [6]). This paper extends the theory of our recent paper [14] where we have established an empirical process theory for locally stationary processes under functional dependence considering function classes that are at least Hölder-continuous. Here, we additionally allow for nonsmooth functions, in particular, our framework includes (but is by far not limited to) the empirical distribution function (EDF).

The only papers that are known to the authors that explicitly deal with functional convergence of locally stationary processes are [14] and [12]. For stationary processes, a vast range of theoretical results are available. A prominent idea to measure dependence of random variables is given by mixing (cf. [9]). The publications [2], [21] and [10] derive large deviation results and uniform central limit theorems under absolute regularity (β-mixing). In [16], refined results are available. Other general theories are based on Markov chains and other types of mixing, cf. the overview in [14].

Regarding the functional weak convergence of the EDF, more specific conditions were derived in the literature for stationary observations. [11, Theorem 4] provide functional convergence of the EDF using bounds for covariances of Hölder functions of the random variables. Another abstract concept was introduced by [3] via S-mixing (for stationary mixing), which imposes the existence of m-dependent approximations of the original observations. They then derive strong approximations and uniform central limit theorems for the EDF. Other approaches were presented in [4] and [7]. In [19] and [12] uniform central limit theorems for the EDF were derived for stationary and piece-wise locally stationary processes under functional dependence.

Our empirical process theory is derived under the dependence concept of functional dependence (cf. [18]). In combination with the theory of martingales it allows for sharp large deviation inequalities (cf. [20] or [22]). We assume that $X_i = (X_{ij})_{j=1,...,d}$, $i = 1, ..., n$, is a d-dimensional Bernoulli shift process of the form

$$X_i := X_{i,n} := J_{i,n}(\mathcal{A}_i), \tag{1.1}$$

where $\mathcal{A}_i = \sigma(\varepsilon_i, \varepsilon_{i-1}, ...)$ is the sigma-algebra generated by ε_i, $i \in \mathbb{Z}$, a sequence of i.i.d. random variables in $\mathbb{R}^{\tilde{d}}$ ($d, \tilde{d} \in \mathbb{N}$), and some measurable function $J_{i,n} : (\mathbb{R}^{\tilde{d}})^{\mathbb{N}_0} \to \mathbb{R}^d$, $i = 1, ..., n$, $n \in \mathbb{N}$. We note that X_i still depends on n, but for the sake of readibility we omit the dependence on n.

For a real-valued random variable W and some $\nu > 0$, we define $\|W\|_\nu := \mathbb{E}[|W|^\nu]^{1/\nu}$. If ε_k^* is an independent copy of ε_k, independent of $\varepsilon_i, i \in \mathbb{Z}$, we define $\mathcal{A}_i^{*(i-k)} := (\varepsilon_i, ..., \varepsilon_{i-k+1}, \varepsilon_{i-k}^*, \varepsilon_{i-k-1}, ...)$ and set $X_i^{*(i-k)} := X_{i,n}^{*(i-k)} := J_{i,n}(\mathcal{A}_i^{*(i-k)})$. The uniform functional dependence measure is then given by

$$\delta_\nu^X(k) = \sup_{i=1,...,n} \sup_{j=1,...,d} \big\| X_{ij} - X_{ij}^{*(i-k)} \big\|_\nu. \tag{1.2}$$

The value δ_ν^X measures the impact of ε_0 on X_k. The representation (1.1) actually does cover a large variety of processes. We additionally allow J to vary with i and n to cover processes which change their stochastic behavior over time. This is exactly the form of the so-called locally stationary processes discussed in [6].

Since we are working in a time series context, many applications ask for functions f that not only depend on the actual observation of the process but on the whole (infinite) past $Z_i := (X_i, X_{i-1}, X_{i-2}, ...)$. In the course of this paper, we aim to derive asymptotic properties of the empirical process

$$\mathbb{G}_n(f) := \frac{1}{\sqrt{n}} \sum_{i=1}^n \big\{ f(Z_i, \frac{i}{n}) - \mathbb{E} f(Z_i, \frac{i}{n}) \big\}, \quad f \in \mathcal{F}, \tag{1.3}$$

where

$$\mathcal{F} \subset \{ f : (\mathbb{R}^d)^{\mathbb{N}_0} \times [0,1] \to \mathbb{R} \text{ measurable} \}.$$

Let $\mathbb{H}(\varepsilon, \mathcal{F}, \|\cdot\|)$ denote the bracketing entropy, that is, the logarithm of the number of ε-brackets with respect to some distance $\|\cdot\|$ that is necessary to cover $\mathcal{F}$ (this is made precise at the end of this section). We will define a distance

V_n which guarantees weak convergence of (1.3) if the corresponding bracketing entropy integral $\int_0^1 \sqrt{\mathbb{H}(\varepsilon, \mathcal{F}, V_n)} d\varepsilon$ is finite.

Our main contributions are the following:

- We derive maximal inequalities for $\mathbb{G}_n(f)$ where the class $\mathcal{F}$ consists of nonsmooth functions.
- We state conditions to ensure asymptotic tightness and functional convergence of $\mathbb{G}_n(f)$, $f \in \mathcal{F}$.

Eventhough our theory allows for general function classes, we will have a special focus on the EDF. In particular, we derive functional convergence of the EDF under weak conditions on the moments and the dependence structure of the process X_i. We will see that our results typically pose weaker conditions on the underlying dependence structure than comparable results for the stationary case mentioned above. In particular, we compare our results with [12] where the authors discussed the EDF of piece-wise locally stationary processes.

The paper is structured as follows. In Section 2, we present our main result Theorem 2.8, the functional central limit theorem under minimal moment conditions. We then derive a version for stationary processes, and discuss its application on empirical distribution functions where the underlying process is either stationary or locally stationary. It is the aim of Section 2.6 to show a wide range of applicability of our theory. Some assumptions are postponed to Section 3, where a new multivariate central limit theorem for locally stationary processes is presented. In Section 4 we provide new maximal inequalities for $\mathbb{G}_n(f)$ in case of a finite and infinite function class $\mathcal{F}$. In Section 5 a conclusion is drawn. We postpone all detailed proofs to the Appendix, Section A.

We now introduce some basic notation. For $a, b \in \mathbb{R}$, let $a \wedge b := \min\{a, b\}$, $a \vee b := \max\{a, b\}$. For $k \in \mathbb{N}$,

$$H(k) := 1 \vee \log(k) \tag{1.4}$$

which naturally appears in large deviation inequalities. For a given finite class $\mathcal{F}$, let $|\mathcal{F}|$ denote its cardinality. We use the abbreviation

$$H = H(|\mathcal{F}|) = 1 \vee \log |\mathcal{F}| \tag{1.5}$$

if no confusion arises. For some distance $\|\cdot\|$, let $\mathbb{N}(\varepsilon, \mathcal{F}, \|\cdot\|)$ denote the bracketing numbers, that is, the smallest number of ε-brackets $[l_j, u_j] := \{f \in \mathcal{F} : l_j \leq f \leq u_j\}$ (i.e. measurable functions $l_j, u_j : (\mathbb{R}^d)^{\mathbb{N}_0} \times [0,1] \to \mathbb{R}$ with $\|u_j - l_j\| \leq \varepsilon$ for all j) to cover $\mathcal{F}$. Let $\mathbb{H}(\varepsilon, \mathcal{F}, \|\cdot\|) := \log \mathbb{N}(\varepsilon, \mathcal{F}, \|\cdot\|)$ denote the bracketing entropy. For $\nu \geq 1$, let

$$\|f\|_{\nu,n} := \Big(\frac{1}{n}\sum_{i=1}^n \big\|f\big(Z_i, \frac{i}{n}\big)\big\|_\nu^\nu\Big)^{1/\nu}. \tag{1.6}$$

2. A functional central limit theorem under functional dependence and application to empirical distribution functions

A process X_i, $i = 1, ..., n$, is called locally stationary if for each $u \in [0,1]$, there exists a stationary process $\tilde{X}_i(u)$ approximating X_i for $i = 1, ..., n$, i.e. $X_i \approx \tilde{X}_i(u)$ if $|u - \frac{i}{n}|$ is small (cf. [6]). The exact form needed is stated in Assumption 2.5. Thus, X_i behaves stationary around each fixed (rescaled) time point $u \in [0,1]$, but over the whole time period $i = 1, ..., n$ its distribution can change drastically. Deterministic properties of the process like expectation, covariance, spectral density or empirical distribution functions therefore also depend on the rescaled time $u \in [0,1]$. As an example, consider the localized empirical distribution function of X_i,

$$\hat{G}_{n,h_n}(x,v) := \frac{1}{nh_n}\sum_{i=1}^{n} K\big(\frac{i/n - v}{h_n}\big)\mathbb{1}_{\{X_i \le x\}}, \tag{2.1}$$

where $K : \mathbb{R} \to \mathbb{R}$ is a kernel function and $h_n > 0$ a bandwidth. The goal of this paper is to provide a general empirical process theory which allows to show, for instance, a functional central limit theorem of $\hat{G}_{n,h_n}(x,v)$ for fixed $v \in [0,1]$ of the form

$$\big[\sqrt{nh}\big(\hat{G}_{n,h_n}(x,v) - G(x,v)\big)\big]_{x\in\mathbb{R}} \xrightarrow{d} \mathbb{G}(x)_{x\in\mathbb{R}} \tag{2.2}$$

where $(\mathbb{G}(x))_{x\in\mathbb{R}}$ is a centered Gaussian process and $G(x,v) = \mathbb{P}(\tilde{X}_0(v) \le x)$ denotes the distribution function of $\tilde{X}_0(u)$.

Clearly, the additional localization via kernels changes the convergence rate of the empirical process. To discuss (2.1) with the general form (1.3), we therefore suppose that any $f \in \mathcal{F}$ has a representation

$$f(z,u) = D_{f,n}(u) \cdot \bar{f}(z,u), \qquad z \in (\mathbb{R}^d)^{\mathbb{N}_0}, u \in [0,1], \tag{2.3}$$

where $\bar{f}$ is independent of n and $D_{f,n}(u)$ is independent of $z = (z_j)_{j\in\mathbb{N}_0}$. For the specific example given in (2.2), we would consider

$$\mathcal{F} = \Big\{(z,u) \mapsto f_x(z,u) := \frac{1}{\sqrt{h_n}}K(\frac{u-v}{h_n}) \cdot \mathbb{1}_{\{z_0 \le x\}} : x \in \mathbb{R}\Big\},$$

and thus $D_{f_x,n}(u) = \frac{1}{\sqrt{h_n}}K(\frac{u-v}{h_n})$ and $\bar{f}_x(z,u) = \mathbb{1}_{\{z_0 \le x\}}$.

We now introduce the necessary assumptions for our empirical process theory based on the functional dependence measure. Based on the decomposition (2.3), we define the following two function classes based on $\bar{f}$, which mimic the one-step-ahead mean and variance forecast,

$$\begin{aligned} \bar{\mathcal{F}}^{(1)} &:= \{(z,u) \mapsto \mathbb{E}[\bar{f}(Z_i,u)|Z_{i-1} = z] : i \in \mathbb{Z},\, f \in \mathcal{F}\}, \\ \bar{\mathcal{F}}^{(2)} &:= \{(z,u) \mapsto \mathbb{E}[\bar{f}(Z_i,u)^2|Z_{i-1} = z]^{1/2} : i \in \mathbb{Z},\, f \in \mathcal{F}\}. \end{aligned}$$

For $s \in (0,1]$, a sequence $z = (z_j)_{j\in\mathbb{N}_0}$ of elements of $\mathbb{R}^d$ (equipped with the maximum norm $|\cdot|_\infty$) and an absolutely summable sequence $L = (L_j)_{j\in\mathbb{N}_0}$ of nonnegative real numbers, we set

$$|z|_{L,s} := \Big(\sum_{j=0}^{\infty} L_j |z_j|_\infty^s\Big)^{1/s}, \qquad |z|_L := |z|_{L,1}.$$

Definition 2.1. A class $\mathcal{G}$ is called a (L,s,R,C)-class if $L = (L_j)_{j\in\mathbb{N}_0}$ is a sequence of nonnegative real numbers, $s \in (0,1]$ and $R : (\mathbb{R}^d)^{\mathbb{N}_0} \times [0,1] \to [0,\infty)$ satisfies for all $u \in [0,1]$, $z, z' \in (\mathbb{R}^d)^{\mathbb{N}_0}$, $g \in \mathcal{G}$,

$$|g(z,u) - g(z',u)| \le |z - z'|_{L,s}^s \cdot \big[R(z,u) + R(z',u)\big].$$

Furthermore, the tuple $C = (C_R, C_{\mathcal{G}}) \in (0,\infty)^2$ satisfies $\sup_u |g(0,u)| \le C_{\mathcal{G}}$, $\sup_u |R(0,u)| \le C_R$.

2.1. Main assumptions

There are two basic assumptions on $\bar{f}$ connected to our main result. The first is a compatibility condition which connects smoothness properties of $\bar{\mathcal{F}}^{(\kappa)}$, $\kappa \in \{1,2\}$ with corresponding moment assumptions on the process X_i, $i = 1,...,n$.

Assumption 2.2 (Compatibility condition on $\mathcal{F}$)**.** *The classes $\bar{\mathcal{F}}^{(\kappa)}$, $\kappa \in \{1,2\}$, are (L,s,R,C)-classes for some L, s, R, C, and there exists $p \in (1,\infty]$, $C_X > 0$ such that*

$$\sup_{i,u} \|R(Z_{i-1},u)\|_{2p} \le C_R, \qquad \sup_{i,j} \|X_{ij}\|_{\frac{2sp}{p-1}} \le C_X.$$

Let $\mathbb{D}_n \ge 0$ and $\Delta(k) \ge 0$ such that

$$2dC_R \sum_{j=0}^{k-1} L_j (\delta^X_{\frac{2sp}{p-1}}(k-j-1))^s \le \Delta(k), \qquad \sup_{f\in\mathcal{F}} \Big(\frac{1}{n}\sum_{i=1}^n \big|D_{f,n}(\frac{i}{n})\big|^2\Big)^{1/2} \le \mathbb{D}_n.$$

Based on Assumption 2.2, we define for $f \in \mathcal{F}$,

$$V_n(f) := \|f\|_{2,n} + \sum_{k=1}^{\infty} \min\{\|f\|_{2,n}, \mathbb{D}_n\Delta(k)\}, \tag{2.4}$$

where $\|f\|_{2,n}$ is given by equation (1.6). Clearly, $V_n(f-g)$ can be interpreted as a distance between $f, g \in \mathcal{F}$, albeit it is actually a seminorm. Furthermore, let

$$\beta(q) := \sum_{j=q}^{\infty} \Delta(j). \tag{2.5}$$

The following assumptions are also required when stating a multivariate central limit theorem on the empirical process $\mathbb{G}_n(f)$.

Assumption 2.3. *Let $\bar{F}$ be an envelope function of $\{\bar{f} : f \in \mathcal{F}\}$, that is, $|\bar{f}(\cdot)| \le \bar{F}(\cdot)$ for all $f \in \mathcal{F}$. There exists $\bar{p} \in (1,\infty]$ such that $\sup_{i,u} \|\bar{F}(Z_i,u)\|_{2\bar{p}} < \infty$, $\sup_{v,u} \|\bar{F}(\tilde{Z}_0(v),u)\|_{2\bar{p}} < \infty$. Furthermore, either*

- *X_i is stationary, or*
- *for all $c > 0$ and $f \in \mathcal{F}$,*

$$\sup_{u,v\in[0,1]} \frac{1}{c^s}\mathbb{E}\Big[\sup_{|a|_{L_{\mathcal{F}},s}\le c}\big|\bar{f}(\tilde{Z}_0(v),u) - \bar{f}(\tilde{Z}_0(v)+a,u)\big|^2\Big] < \infty. \tag{2.6}$$

Additionally, (2.6) also holds for $\bar{F}$ instead of $\bar{f}$.

Assumption 2.4. *There exists some $\varsigma \in (0,1]$ such that for every $f \in \mathcal{F}$,*

$$|\bar{f}(z,u_1) - \bar{f}(z,u_2)| \le |u_1-u_2|^{\varsigma}\cdot\big(\bar{R}(z,u_1)+\bar{R}(z,u_2)\big),$$

and $\sup_{u,v}\|\bar{R}(\tilde{Z}_0(v),u)\|_2 < \infty$.

Assumption 2.5. *For each $u \in [0,1]$, there exists a process $\tilde{X}_i(u) = J(\mathcal{A}_i,u)$, $i \in \mathbb{Z}$, where J is a measurable function. Furthermore, there exists some $C_X > 0$, $\varsigma \in (0,1]$ such that for every $i \in \{1,...,n\}$, $u_1,u_2 \in [0,1]$,*

$$\|X_i - \tilde{X}_i(\frac{i}{n})\|_{\frac{2sp}{p-1}} \le C_X n^{-\varsigma}, \qquad \|\tilde{X}_i(u_1)-\tilde{X}_i(u_2)\|_{\frac{2sp}{p-1}} \le C_X|u_1-u_2|^{\varsigma}.$$

For $\tilde{Z}_i(u) = (\tilde{X}_i(u),\tilde{X}_{i-1}(u),...)$ it holds that $\sup_{v,u}\|R(\tilde{Z}_0(v),u)\|_{2p} < \infty$.

For $f \in \mathcal{F}$, let $D_{f,n}^{\infty} := \sup_{i=1,...,n} D_{f,n}(\frac{i}{n})$ where $D_{f,n}$ can be recalled in equation (2.3).

Assumption 2.6. *For all $f \in \mathcal{F}$, the function $\frac{D_{f,n}(\cdot)}{D_{f,n}^{\infty}}$ has bounded variation uniformly in n, and*

$$\sup_{n\in\mathbb{N}}\frac{1}{n}\sum_{i=1}^{n} D_{f,n}(\frac{i}{n})^2 < \infty, \qquad \frac{D_{f,n}^{\infty}}{\sqrt{n}} \to 0. \tag{2.7}$$

One of the two following cases hold.

(i) Case $\mathbb{K} = 1$ (global case): For all $f,g \in \mathcal{F}$, $u \mapsto \mathbb{E}[\mathbb{E}[\bar{f}(\tilde{Z}_{j_1}(u),u)|Z_0]\cdot \mathbb{E}[\bar{g}(\tilde{Z}_{j_2}(u),u)|Z_0]]$ has bounded variation for all $j_1,j_2 \in \mathbb{N}_0$ and the following limit exists:

$$\Sigma_{fg}^{(1)} := \lim_{n\to\infty}\int_0^1 D_{f,n}(u)D_{g,n}(u)\cdot\sum_{j\in\mathbb{Z}} Cov(\bar{f}(\tilde{Z}_0(u),u),\bar{g}(\tilde{Z}_j(u),u))du.$$

(ii) Case $\mathbb{K} = 2$ (local case): There exists a sequence $h_n > 0$ and $v \in [0,1]$ such that $\mathrm{supp} D_{f,n}(\cdot) \subset [v-h_n, v+h_n]$. It holds that

$$h_n \to 0, \qquad \sup_{n\in\mathbb{N}}(h_n^{1/2}\cdot D_{f,n}^{\infty}) < \infty.$$

The following limit exists for all $f,g \in \mathcal{F}$:

$$\Sigma_{fg}^{(2)} := \lim_{n\to\infty}\int_0^1 D_{f,n}(u)D_{g,n}(u)du\cdot\sum_{j\in\mathbb{Z}} Cov(\bar{f}(\tilde{Z}_0(v),v),\bar{g}(\tilde{Z}_j(v),v)).$$

While Assumption 2.3 asks for smoothness of $\bar{f}$ in the L^2-sense if X_i is nonstationary, Assumption 2.4 requires the function class $\mathcal{F}$ to behave smoothly in the second argument. Assumption 2.5 formulates what it means for a process to be *locally stationary* (cf. [6]). The last Assumption 2.6 mainly controls the behavior of the factor $D_{f,n}(u)$ of $f \in \mathcal{F}$ (from equation (2.3)) which does not depend on the observations. Assumption 2.6 looks rather technical but is crucial to this context. The first part including (2.7) guarantees the right normalization of $D_{f,n}(\cdot)$. The second part ensures the convergence of the asymptotic variances $\mathrm{Var}(\mathbb{G}_n(f))$ and covariances $\mathrm{Cov}(\mathbb{G}_n(f), \mathbb{G}_n(g))$ with respect to the behavior of $D_{f,n}(\cdot)$.

Assumptions 2.4, 2.5 and 2.6 are needed to allow for *very different* function classes $\mathcal{F}$ which later on becomes very important as we consider function classes of infinite cardinality. In many special cases, however, some of these assumptions are automatically fulfilled. For example,

- If $\bar{f}(z,u) = \bar{f}(z)$ does not depend on u, Assumption 2.4 is fulfilled.
- If X_i is stationary, Assumption 2.5 is fulfilled.
- If $D_{f,n}(u) = 1$, Assumption 2.6 is fulfilled.

We aslo need an additional submultiplicativity assumption on the dependence term $\beta(\cdot)$ from (2.5).

Assumption 2.7. *There exists a constant $C_\beta > 0$ such that for each $q_1, q_2 \in \mathbb{N}$,*

$$\beta(q_1 q_2) \leq C_\beta \cdot \beta(q_1)\beta(q_2).$$

It is easily seen that Assumption 2.7 is fulfilled if $\Delta(k)$, in Assumption 2.2, follows a polynomial ($\Delta(k) = ck^{-\alpha}$ for $c > 0, \alpha > 1$) or exponential decay ($\Delta(k) = c\rho^k$ for $c > 0$, $\rho \in (0,1)$), cf. [14, Lemma 7.9]. It is generally not possible to show Assumption 2.7 if $\Delta(k)$ contains a factor of the form $\frac{1}{\log(k)}$.

2.2. A functional central limit theorem

We now state our main result. Recall the definition of V_n from (2.4). In the space

$$\ell^\infty(\mathcal{F}) = \{\mathbb{G} : \mathcal{F} \to \mathbb{R} \,|\, \|\mathbb{G}\|_\infty := \sup_{f \in \mathcal{F}} |\mathbb{G}(f)| < \infty\}, \tag{2.8}$$

the following theorem holds true.

Theorem 2.8. *Suppose that $\mathcal{F}$ satisfies Assumptions 2.2, 2.3, 2.7, 2.5, 2.4, 2.6. For*

$$\psi(\varepsilon) = \sqrt{\log(\varepsilon^{-1} \vee 1)} \log\log(\varepsilon^{-1} \vee e) \tag{2.9}$$

suppose that

$$\sup_{n \in \mathbb{N}} \int_0^1 \psi(\varepsilon)\sqrt{\mathbb{H}(\varepsilon, \mathcal{F}, V_n)}d\varepsilon < \infty.$$

Then in $\ell^\infty(\mathcal{F})$,

$$\big[\mathbb{G}_n(f)\big]_{f \in \mathcal{F}} \xrightarrow{d} \big[\mathbb{G}(f)\big]_{f \in \mathcal{F}}$$

where $(\mathbb{G}(f))_{f\in\mathcal{F}}$ *is a centered Gaussian process with covariances*

$$\mathrm{Cov}(\mathbb{G}(f),\mathbb{G}(g)) = \lim_{n\to\infty}\mathrm{Cov}(\mathbb{G}_n(f),\mathbb{G}_n(g)) = \Sigma^{(\mathbb{K})}$$

and $\Sigma^{(\mathbb{K})}$ *is from Assumption 2.6.*

Here, the weak convergence is meant in the usual sense with outer probabilities. The properties of the space $\ell^\infty(\mathcal{F})$ can be found in [17], for instance.

Note that it is a result of the convergence of the finite-dimensional distributions in Section 3, Theorem 3.1, and asymptotic tightness in Section 4.2, Corollary 4.5.

If further conditions in Assumption 2.2 are imposed, we derive simpler forms of V_n which are shown in the table below. In this case we suppose that $\mathbb{D}_n \in (0,\infty)$ is independent of $n\in\mathbb{N}$ and $\Delta(k)$ is of polynomial or geometric decay. These results are proven in [14, Lemma 7.11 and Lemma 7.12].

The theorem significantly simplifies if X_i is stationary, $\bar{f}(z,u)=\bar{f}(z_0)$, depends only on one observation and no weighting is present, i.e. $D_{f,n}(u)=1$. Assumptions 2.3, 2.4, 2.5 and 2.6 are then directly fulfilled. These assumptions are needed only to provide a (pointwise) central limit theorem for locally stationary processes. They basically ask for several smoothness properties of $\bar{f}$.

In more detail, let

$$\tilde{\mathbb{G}}_n(h) := \frac{1}{\sqrt{n}}\sum_{i=1}^n\big\{h(X_i)-\mathbb{E}h(X_i)\big\},$$

where $X_i = J(\mathcal{A}_i)$, $i=1,\dots,n$, is a stationary Bernoulli shift process and $\mathcal{H}\subset\{h:\mathbb{R}^d\to\mathbb{R} \text{ measurable}\}$ with envelope function $\bar{h}$, i.e. for $h\in\mathcal{H}$ we have $|h(\cdot)|\le\bar{h}(\cdot)$, such that

$$h^{(1)}(z_0)=\mathbb{E}[h(X_1)|X_0=z_0], \qquad h^{(2)}(z_0)=\mathbb{E}[h(X_1)^2|X_0=z_0]^{1/2}$$

are Hölder continuous with exponent s and constant $L_{\mathcal{H}}$, that is, for all $z,z'\in\mathbb{R}$,

$$|h^{(1)}(z)-h^{(1)}(z')|\le L_{\mathcal{H}}|z-z'|^s, \qquad |h^{(2)}(z)-h^{(2)}(z')|\le L_{\mathcal{H}}|z-z'|^s.$$

Assumption 2.2 automatically is satisfied with $R(\cdot)=\frac{1}{2}$ and thus $C_R=\frac{1}{2}$, $L=L_{\mathcal{H}}$ as well as $C_{\mathcal{G}}=\max\{h^{(1)}(0),h^{(2)}(0)\}$. Recall $\beta(\cdot)$ in equation (2.5) and the functional dependence measure $\delta_\nu^X(k)$ from (1.2). Then we have the following corollary of Theorem 2.8 in notation of Assumption 2.2.

Corollary 2.9. *Suppose that* $\|X_1\|_{2s}<\infty$ *and put* $\mathbb{D}_n := 1$*. Let* $\Delta(k)$ *fulfill* $\Delta(k)\ge dL_{\mathcal{H}}\delta_{2s}^X(k-1)^s$ *and there exists* $C_\beta>0$ *such that for all* $q_1,q_2\in\mathbb{N}$*,*

$$\beta(q_1q_2)\le C_\beta\beta(q_1)\beta(q_2). \tag{2.10}$$

Furthermore, $\|\bar{h}(X_1)\|_{2\bar{p}}<\infty$ *for some* $\bar{p}>1$*. Assume that*

$$\sup_{n\in\mathbb{N}}\int_0^1\psi(\varepsilon)\sqrt{\mathbb{H}(\varepsilon,\mathcal{H},V_n)}d\varepsilon<\infty,$$

where $\psi(\varepsilon)$ is from (2.9). Then it holds in $\ell^\infty(\mathcal{H})$ that

$$\big[\tilde{\mathbb{G}}_n(h)\big]_{h\in\mathcal{H}} \xrightarrow{d} \big[\tilde{\mathbb{G}}(h)\big]_{h\in\mathcal{H}},$$

where $(\tilde{\mathbb{G}}(h))_{h\in\mathcal{H}}$ is a centered Gaussian process with covariances

$$\mathrm{Cov}(\tilde{\mathbb{G}}(h_1), \tilde{\mathbb{G}}(h_2)) = \sum_{k\in\mathbb{Z}} \mathrm{Cov}(h_1(X_0), h_2(X_k)).$$

2.3. Application to empirical distribution functions of stationary processes

As an example, consider the family of indicators

$$\mathcal{H} = \{h_x(z_0) := \mathbb{1}_{\{z_0 \le x\}} : x \in \mathbb{R}\},$$

which is the function class corresponding to the empirical distribution function

$$\big[\hat{G}_n(x)\big]_{x\in\mathbb{R}} = \Big[\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{X_i \le x\}}\Big]_{x\in\mathbb{R}} = \big[\tilde{\mathbb{G}}_n(h)\big]_{h\in\mathcal{H}}.$$

Suppose that X_i, $i = 1, ..., n$, is stationary. Define the conditional distribution function

$$G_z(x) = \mathbb{P}(X_1 \le x | X_0 = z).$$

Then we have the following corollary for dependence coefficients (cf. equation (1.2)) that follow a polynomial decay.

Corollary 2.10. *Suppose that X_i is stationary and $z \mapsto G_z(x)$ is Lipschitz continuous with Lipschitz constant L_G for all $x \in \mathbb{R}$. Suppose that for some $s \in (0, \frac{1}{2}]$, $\|X_1\|_{2s} < \infty$ and $\delta_{2s}^X(k) \le ck^{-\alpha}$ with $\alpha > \frac{1}{s}$, $c > 0$. Then,*

$$\big[\hat{G}_n(x)\big]_{x\in\mathbb{R}} \xrightarrow{d} \big[\tilde{\mathbb{G}}(x)\big]_{x\in\mathbb{R}},$$

where $\tilde{\mathbb{G}}(x)$ is a Gaussian process with

$$\mathrm{Cov}(\tilde{\mathbb{G}}(x), \tilde{\mathbb{G}}(y)) = \sum_{k\in\mathbb{Z}} \mathrm{Cov}(\mathbb{1}_{\{X_0 \le x\}}, \mathbb{1}_{\{X_k \le y\}}).$$

Proof of Corollary 2.10. Due to $\min\{1, w\} \le w^a$ for $a \in [0, 1]$, $w \ge 0$, we have that for any $s \in (0, \frac{1}{2}]$,

$$|G_z(x) - G_{z'}(x)| \le \min\{1, L_G|z - z'|\} \le L_G^s |z - z'|^s.$$

and

$$|G_z(x) - G_{z'}(x)|^{1/2} \le \min\{1, (L_G|z - z'|)^{1/2}\} \le L_G^s |z - z'|^s.$$

Choose $\Delta(k) = cL_G(k-1)^{-\alpha s}$, which is easily seen to satisfy (2.10) (in particular, $\beta(q) < \infty$ for $q \in \mathbb{N}$) for some $C_\beta = C_\beta(\alpha, s, c, L_G)$ chosen large enough.

Table 1

Equivalent expressions of V_n and the corresponding entropy integral under the condition that $\mathbb{D}_n \in (0,\infty)$ is independent of n. We omitted the lower and upper bound constants which are only depending on c, ρ, α and $\mathbb{D}_n$. Furthermore, $\tilde{\sigma} = \tilde{\sigma}(\sigma)$ fulfills $\tilde{\sigma} \to 0$ for $\sigma \to 0$.

	$\Delta(j)$	
	$cj^{-\alpha}$, $\alpha > 1, c > 0$	$c\rho^j$, $\rho \in (0,1)$, $c > 0$
$V_n(f)$	$\|f\|_{2,n} \max\{\|f\|_{2,n}^{-\frac{1}{\alpha}}, 1\}$	$\|f\|_{2,n} \max\{\log(\|f\|_{2,n}^{-1}), 1\}$
$\int_0^\sigma \sqrt{\mathbb{H}(\varepsilon,\mathcal{F},V_n)}d\varepsilon$	$\int_0^{\tilde{\sigma}} \varepsilon^{-\frac{1}{\alpha}}\psi(\varepsilon)\sqrt{\mathbb{H}(\varepsilon,\mathcal{F},\|\cdot\|_{2,n})}d\varepsilon$	$\int_0^{\tilde{\sigma}} \log(\varepsilon^{-1})\psi(\varepsilon)\sqrt{\mathbb{H}(\varepsilon,\mathcal{F},\|\cdot\|_{2,n})}d\varepsilon$

Note that $\mathbb{H}(\varepsilon, \mathcal{H}, \|\cdot\|_{2,n}) = O(\log(\varepsilon^{-1}))$ for a given $\varepsilon > 0$ by [17, Example 19.6], since in the stationary situation of the corollary, $\|h\|_{2,n} = \mathbb{E}[h(X_1)^2]^{1/2}$. Since $\alpha s > 1$, Table 1 implies that

$$\int_0^1 \psi(\varepsilon)\sqrt{\mathbb{H}(\gamma,\mathcal{H},V_n)}d\varepsilon = O\Big(\int_0^1 \psi(\varepsilon)\varepsilon^{-\frac{1}{\alpha s}}\sqrt{\log(\varepsilon^{-1})}d\varepsilon\Big) < \infty.$$

Corollary 2.9 now implies the assertion. □

2.4. Comparison with other functional convergence results for the empirical distribution function of stationary processes

In the literature, several functional convergence results for the empirical distribution function were already provided. Here we list some approaches which are closely related to the functional dependence measure and compare the results to Corollary 2.10.

In [4], stationary processes of the form $X_i = J(\mathcal{G}_i)$ are considered where $\mathcal{G}_i = (\varepsilon_i, \varepsilon_{i-1}, ...)$ and J is measurable. Therein, the function J itself is assumed to fulfill a (geometrically decaying) Lipschitz condition, i.e. for any sequences $(a_i), (a_i')$ with $a_i = a_i'$, $i \le k$,

$$\big|J((a_i)) - J((a_i'))\big| \le C\alpha^k \tag{2.11}$$

for some constants $C, \alpha > 0$. Based on this, 1-approximation coefficients a_k are defined as upper bounds on

$$\mathbb{E}\big\|X_0 - \mathbb{E}[X_0|\sigma(\varepsilon_0, ..., \varepsilon_k)]\big\|_1 \le a_k.$$

There is a strong connection between $\delta_1^X(k)$ and a_k, since it is possible to choose $a_k \le \sum_{j=k+1}^\infty \delta_1^X(j)$. The work of [4, Theorem 5] shows that under summability conditions on a_k, the β-mixing coefficients *and* monotonicity assumptions on $\mathcal{F} = \{f_t : t \in [0,1]\}$, a uniform central limit theorem for $(\mathbb{G}_n(f_t))_{t\in[0,1]}$ holds. Compared to our setting, (2.11) would lead to a geometrically decaying functional dependence measure $\delta^X(k)$. Thus, the result in our Corollary 2.10 is much less restrictive regarding the dependence of the underlying process.

In [7, Theorem 2.1], a uniform central limit theorem for the empirical distribution function is shown under $\beta_2(k) = O(k^{-1-\gamma})$, $\gamma > 0$, by using specifically designed dependence coefficients $\beta_2(k)$, $k \in \mathbb{N}_0$, based on the idea of absolute regularity. We now compare this result with Corollary 2.10. In [8, Section 6.1] it was shown that if $X_i = J(\mathcal{G}_i)$ is stationary and the distribution function of X_1 is Lipschitz continuous, then for any $\nu \in [0,1]$ one has

$$\beta_2(k) \le C \cdot \Big(\sum_{j=k+1}^{\infty} \delta_\nu^X(j)^{\nu'} \Big)^{\frac{\nu}{\nu'(\nu+1)}}, \qquad \nu' = \min\{\nu, 1\},$$

where $C > 0$ is a constant independent of k. The condition $\beta_2(k) = O(k^{-1-\gamma})$ now naturally provides a decay condition on $\delta_\nu^X(k)$. With $\nu = 2s$ which corresponds to the moments of the process we have given in Corollary 2.10, we see after a short calculation that $\beta_2(k) = O(k^{-1-\gamma})$ asks for

$$\alpha \ge \frac{1}{s} + \frac{\gamma}{2s} + \gamma + 1.$$

In other words, if the results from [7], [8] are transferred to the functional dependence measure setting, they need a more restrictive decay condition.

Meanwhile, [3] investigates strong approximations of the multivariate empirical distribution function process (that is, contrary to our approach, the results are limited to empirical distribution functions). They assume that the stationary process $X_i = J(\mathcal{G}_i)$ allows for approximations $(X_i^{(m)})$ such that for all m, i,

$$\mathbb{P}(|X_i - X_i^{(m)}| \ge m^{-A}) \le m^{-A} \tag{2.12}$$

with some $A > 4$, and for any disjoint intervals $I_1, ..., I_r$ of integers and any positive integers $m_1, ..., m_r$, the vectors $\{X_i^{(m_1)} : i \in I_1\}, ..., \{X_i^{(m_r)} : j \in I_r\}$ are independent provided the separation between I_k and I_l is greater than $m_k + m_l$. Under these assumptions, [3, Theorem 1, Corollary 1] shows that the empirical distribution function of X_i weakly converges to some Gaussian process.

When having knowledge about the functional dependence measure, $X_i^{(m)}$ could be chosen as $X_i^{(m)} = \mathbb{E}[X_i|\varepsilon_i, ..., \varepsilon_{i-m}]$. Then by Markov's inequality,

$$\mathbb{P}\big(|X_i - X_i^{(m)}| \ge m^{-A}\big) \le \frac{\|X_i - X_i^{(m)}\|_{2s}^{2s}}{m^{-2sA}} \le \big(m^A \cdot \sum_{j=m+1}^{\infty} \delta_{2s}^X(j)\big)^{2s},$$

so that (2.12) leads to a decay condition on $\delta_\nu^X(j)$. After a short calculation, we see that (2.12) is fulfilled if

$$\alpha \ge (\frac{1}{2s} + 1)A + 1,$$

again a more restrictive decay condition than given in Corollary 2.10.

The work of [11] discusses the functional convergence of the multivariate empirical distribution function under a general growth condition imposed on the moments of $\sum_{i=1}^n \{h(X_i) - \mathbb{E}h(X_i)\}$, where $h \in \mathcal{H}_\gamma$ are Hölder continuous functions with exponent $\gamma \in (0,1]$ approximating the indicator functions. They also relate their result to the functional dependence measure.

2.5. Application to empirical distribution functions of locally stationary processes

In this section, we apply our theory to the localized empirical distribution function $\hat{G}_{n,h_n}(x,v)$ from (2.1) on a locally stationary process as motivated in the beginning of Section 2. Afterwards, we compare our result with [12] and [19].

Suppose that X_i is locally stationary in the sense that for each $u \in [0,1]$, there exists a stationary process $\tilde{X}_i(u) = J(\mathcal{A}_i, u)$, $i \in \mathbb{Z}$, for a measurable function J such that

$$\|X_i - \tilde{X}_i(\frac{i}{n})\|_{2s} \le C_X n^{-\varsigma}, \qquad \|\tilde{X}_i(u) - \tilde{X}_i(u')\|_{2s} \le C_X |u-u'|^{\varsigma}$$

for a constant $C_X > 0$, $\varsigma \in (0,1]$, $u, u' \in [0,1]$ and $i \in \{1,...,n\}$.

Recall $G(x,v) = \mathbb{P}(\tilde{X}_1(v) \le x)$. Define the conditional distribution function of the stationary approximation of X_i,

$$G_z(x,v) = \mathbb{P}(\tilde{X}_1(v) \le x \mid \tilde{X}_0(v) = z).$$

Finally, we have to impose a regularity assumption on the distribution function $G_i(x) := \mathbb{P}(X_i \le x)$ of the locally stationary process itself.

We have the following generalization of Corollary 2.10.

Corollary 2.11. *Let $v \in (0,1)$. Suppose that there exists some $L_G > 0$ such that*

- *$z \mapsto G_z(x,v)$ is Lipschitz continuous with constant L_G for all $x \in \mathbb{R}$,*
- *$x \mapsto G(x,v)$ is Lipschitz continuous with constant L_G,*
- *$x \mapsto G_i(x)$ is Lipschitz continuous with constant L_G and $\lim_{x\to-\infty}\sup_{i,n} G_i(x) = 0$, $\lim_{x\to+\infty}\inf_{i,n} G_i(x) = 1$.*

Assume that $K : \mathbb{R} \to \mathbb{R}$ is a Lipschitz continuous kernel function with $\int K(x)dx = 1$ and support $\subset [-\frac{1}{2}, \frac{1}{2}]$.

Furthermore, for some $s \in (0, \frac{1}{2}]$ let $\sup_{i,n} \|X_i\|_{2s} < \infty$ and $\delta_{2s}^X(k) \le ck^{-\alpha}$ with $\alpha > \frac{1}{s}$, $c > 0$.

Then for $h_n n \to \infty$, $h_n \to 0$,

$$\big[\hat{G}_{n,h_n}(x,v)\big]_{x\in\mathbb{R}} \xrightarrow{d} \big[\tilde{\mathbb{G}}(x,v)\big]_{x\in\mathbb{R}},$$

where $\tilde{\mathbb{G}}(x,v)$ is a Gaussian process with

$$\mathrm{Cov}(\tilde{\mathbb{G}}(x,v), \tilde{\mathbb{G}}(y,v)) = \int K(u)^2 du \cdot \sum_{k\in\mathbb{Z}} \mathrm{Cov}(\mathbb{1}_{\{\tilde{X}_0(v)\le x\}}, \mathbb{1}_{\{\tilde{X}_k(v)\le y\}}).$$

Proof of Corollary 2.11. We verify the conditions of Theorem 2.8. By $\min\{1, w\} \le w^a$ for $a \in [0,1]$, $w \ge 0$, we have for any $s \in (0, \frac{1}{2}]$,

$$|G_z(x,v) - G_{z'}(x,v)| \le \min\{1, L_G|z-z'|\} \le L_G^s |z-z'|^s.$$

and

$$|G_z(x,v) - G_{z'}(x,v)|^{1/2} \leq \min\{1, (L_G|z-z'|)^{1/2}\} \leq L_G^s|z-z'|^s.$$

This shows Assumption 2.2 with $p = \infty$, $R(\cdot) = \frac{1}{2} = C_R$.

Choose $\Delta(k) = cL_G(k-1)^{-\alpha s}$, which can easily be seen to satisfy Assumption 2.7 (in particular, $\beta(q) < \infty$ for $q \in \mathbb{N}$) for some $C_\beta = C_\beta(\alpha, s, c, L_G)$ chosen large enough. Regarding Assumption 2.3 we first have

$$\begin{aligned}
& \frac{1}{c^s}\mathbb{E}\sup_{L_G|a|\leq c}[|\mathbb{1}_{\{\tilde{Z}_0(v)\leq x\}} - \mathbb{1}_{\{\tilde{Z}_0(v)+a\leq x\}}|^2] \\
\leq\ & \frac{1}{c^s}\mathbb{E}|\mathbb{1}_{\{\tilde{Z}_0(v)\leq x\}} - \mathbb{1}_{\{\tilde{Z}_0(v)\leq x-\frac{c}{L_G}\}}| \\
\leq\ & \frac{1}{c^s}(\mathbb{P}(\tilde{Z}_0(v)\leq x) - \mathbb{P}(\tilde{Z}_0(v)\leq x - \frac{c}{L_G})) \\
\leq\ & \frac{1}{c^s}(G_z(x,v) - G_z(x-\frac{c}{L_G}, v)) \\
\leq\ & \frac{1}{c^s}\min\{1,c\} \leq 1.
\end{aligned}$$

The envelope function is the constant 1-function and satisfies the required condition trivially. Therefore, Assumption 2.3 holds true. Assumption 2.4 is automatically satisfied for fixed $v \in (0,1)$. For Assumption 2.6, note that $D_{f,n}(u) = \frac{1}{\sqrt{h_n}}K(\frac{u-v}{h_n})$ satisfies

$$\frac{1}{n}\sum_{i=1}^n D_{f,n}(\frac{i}{n})^2 \leq \frac{1}{nh_n}\sum_{i=1}^n K(\frac{i/n-v}{h_n})^2 \leq |K|_\infty^2 < \infty,$$

and $D_{f,n}^\infty \leq \frac{1}{\sqrt{h_n}}|K|_\infty$. Thus $\frac{D_{f,n}^\infty}{\sqrt{n}} \leq \frac{|K|_\infty}{\sqrt{nh_n}} \to 0$, and the support satisfies $\mathrm{supp}[D_{f,n}(\cdot)] \subset [v-h_n, v+h_n]$. Finally, $h_n^{1/2}D_{f,n}^\infty \leq |K|_\infty < \infty$ and, since $v \in (0,1)$,

$$\lim_{n\to\infty}\int_0^1 D_{f,n}(u)D_{g,n}(u)du = \lim_{n\to\infty}\frac{1}{h_n}\int_0^1 K(\frac{u-v}{h_n})^2 du = \int K(u)^2 du.$$

This shows all conditions of Assumption 2.6 (ii).

It holds that $\mathbb{H}(\varepsilon, \mathcal{H}, \|\cdot\|_{2,n}) = O(\log(\varepsilon^{-1}))$ which is proven subsequently.

Let $\varepsilon > 0$. Since $\lim_{x\to-\infty}\sup_{i,n}G_i(x) = 0$ and $\lim_{x\to+\infty}\inf_{i,n}G_i(x) = 1$, there exists $x_N = x_N(\varepsilon) > x_1 = x_1(\varepsilon) > 0$ such that $\sup_{i,n}G_i(x_1) \leq \varepsilon$, $\inf_{i,n}G_i(x_1) \geq 1-\varepsilon$. Define $x_{j+1} := x_1 + j\cdot\frac{\varepsilon^2}{L_G}$, $j = 1,2,...,N-1$ with $N = 1 + \lceil\frac{(x_N-x_1)L_G}{\varepsilon^2}\rceil$. Put $x_0 = -\infty$ and $x_{N+1} = \infty$. Then for $j = 1,2,...,N-1$ we have

$$\begin{aligned}
& \|\mathbb{1}_{\{\cdot\leq x_{j+1}\}} - \mathbb{1}_{\{\cdot\leq x_{j+1}\}}\|_{2,n}^2 \\
\leq\ & \sup_{i=1,\ldots,n}\mathbb{E}[(\mathbb{1}_{\{X_i\leq x_{j+1}\}} - \mathbb{1}_{\{X_i\leq x_j\}})^2] = \sup_{i=1,\ldots,n}[G_i(x_{j+1}) - G_i(x_j)]
\end{aligned}$$

$$\leq \quad L_G|x_{j+1} - x_j| \leq \varepsilon^2,$$

which shows that $[\mathbb{1}_{\{\cdot \leq x_j\}}, \mathbb{1}_{\{\cdot \leq x_{j+1}\}}]$, $j = 0, ..., N$ are ε-brackets with respect to $\|\cdot\|_{2,n}$. Hence, $\mathbb{H}(\varepsilon, \mathcal{H}, \|\cdot\|_{2,n}) = O(\log(\varepsilon^{-1}))$.

Since $\alpha s > 1$, Table 1 implies that

$$\int_0^1 \psi(\varepsilon)\sqrt{\mathbb{H}(\gamma, \mathcal{H}, V_n)} d\varepsilon = O\Big(\int_0^1 \psi(\varepsilon)\varepsilon^{-\frac{1}{\alpha s}}\sqrt{\log(\varepsilon^{-1})} d\varepsilon\Big) < \infty.$$

Theorem 2.8 now implies the assertion. □

In [19] a similar result is obtained for stationary sequences and accounts for weighted empirical processes, that is, $(1+|x|)^{\gamma/2} G_z(x, v)$ for $\gamma > 0$.

The recently published work [12] considers functional convergence of the empirical distribution function of piece-wise locally stationary processes. They impose that the functional dependence measure to decay geometrically cf. [12, assumption (A5)]. In the above Corollary 2.11 we were able to provide some weaker assumptions in general. In particular, we only need polynomial decay of the dependence coefficients. For instance, let us consider the linear locally stationary processes

$$X_{i,n} = \sum_{k=0}^{\infty} a_k(\frac{i}{n})\varepsilon_{i-k},$$

where ε_i is i.i.d. with $\mathbb{E}\varepsilon_1 = 0$, $\mathbb{E}|\varepsilon|^{\nu} < \infty$ and $a_k : [0, 1] \to \mathbb{R}$ are arbitrary functions with $\sup_{u\in[0,1]} |a_k(u)| \leq \frac{C}{k^{\alpha}}$, $\alpha > 1$. Then, $\delta_{\nu}^X(k) = O(k^{-\alpha})$. In the same manner other decay rates can be realized by setting $\delta_{\nu}^X(k) = O(\sup_{u\in[0,1]} |a_k(u)|)$ for an appropriate choice of the coefficients $a_k(\cdot)$.

We note that Assumption 2.5 also covers recursively defined time-varying processes like the tvARMA model, cf. Proposition 2.4 in [5], defined by

$$\sum_{j=0}^{p} \alpha_j\big(\frac{t}{n}\big)X_{t-j}^{(n)} = \sum_{k=0}^{q} \beta_k\big(\frac{t}{n}\big)\sigma\big(\frac{t-k}{n}\big)\varepsilon_{t-k},$$

where ε_i is i.i.d. with $\mathbb{E}\varepsilon_1 = 0$, $\mathbb{E}|\varepsilon_i| < \infty$ and α_j, β_k, $\sigma : [0, 1] \to \mathbb{R}$ are of bounded variation with $\alpha_0(u) = \beta_0(u) = 1$, $\alpha_j(u) = \alpha_j(0)$, $\beta_k(u) = \beta_k(0)$ for $u < 0$ such that $\sum_{j=0}^{p} \alpha_j(u)z^j \neq 0$ for all u and $0 < |z| \leq 1 + \delta$, $\delta > 0$. In this case it is not sufficient to assume a representation $X_i^{(n)} = \tilde{X}_i^{(n)}(i/n)$.

2.6. Further applications

Our theory allows for empirical process theory of general function classes. We illustrate further applications in two short examples.

Example 1 (Distribution of residuals): Consider the locally stationary time series model which is defined recursively via

$$X_i = m(X_{i-1}, \frac{i}{n}) + \sigma(X_{i-1}, \frac{i}{n})\varepsilon_i, \qquad i = 1, ..., n,$$

where ε_i, $i \in \mathbb{Z}$, is an i.i.d. sequence of random variables with $\mathbb{E}\varepsilon_1 = 0$, $\mathrm{Var}(\varepsilon_1) = 1$ and $\sigma, m : \mathbb{R} \times [0,1] \to \mathbb{R}$. These functions can be estimated with typical Nadaraya-Watson-type estimators

$$\hat{m}(x,v) := \frac{\frac{1}{n}\sum_{i=1}^n K_{h_n}(v - \frac{i}{n})K_{h_n}\big(x - X_{i-1})X_i}{\frac{1}{n}\sum_{i=1}^n K_{h_n}(v - \frac{i}{n})K_{h_n}\big(x - X_{i-1})}$$

and

$$\hat{\sigma}^2(x,v) := \frac{\frac{1}{n}\sum_{i=1}^n K_{h_n}(v - \frac{i}{n})K_{h_n}\big(x - X_{i-1}) \cdot \big(X_i - \hat{m}(X_{i-1}, \frac{i}{n})\big)^2}{\frac{1}{n}\sum_{i=1}^n K_{h_n}(v - \frac{i}{n})K_{h_n}\big(x - X_{i-1})}$$

with some bounded kernel $K : \mathbb{R} \to \mathbb{R}$ and $K_h(\cdot) := \frac{1}{h}K(\frac{\cdot}{h})$. Besides estimation of $m(\cdot), \sigma(\cdot)$, it may also be of interest to derive the distribution function G_ε of ε_i. Following the approach of [1], we define empirical residuals $\hat{\varepsilon}_i = \frac{X_i - \hat{m}(X_{i-1}, i/n)}{\hat{\sigma}(X_{i-1}, i/n)}$. Then the convergence of $(\hat{G}_\varepsilon(x))_{x \in \mathbb{R}}$,

$$\hat{G}_\varepsilon(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{\hat{\varepsilon}_i \le x\}} = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{\varepsilon_i \le x \cdot \frac{\hat{\sigma}(X_{i-1}, i/n)}{\sigma(X_{i-1}, i/n)} + \frac{\hat{m}(X_{i-1}, i/n) - m(X_{i-1}, i/n)}{\sigma(X_{i-1}, i/n)}\}}$$

can be discussed with empirical process theory and analytic properties of $\hat{m}, \hat{\sigma}$. Following the proof of Lemma 1 in [1], we have to define

$$\mathcal{F} := \quad \{f_{x,d_1,d_2}(\varepsilon, z, u) = \mathbb{1}_{\{\varepsilon \le x \cdot d_2(z,u) + d_1(z,u)\}} : \\ -\infty < x < \infty,\, d_1 \in C_1^{1+\delta}(R, [0,1]),\, d_2 \in \tilde{C}_1^{1+\delta}(R, [0,1])\},$$

where R is the union of all domains of X_i, $i = 1, ..., n$, $\delta > 0$ and $C_1^{1+\delta}(R, [0,1])$ is the class of differentiable functions $d : R \times [0,1] \to \mathbb{R}$ (with respect to the first component) such that

$$\|d\|_{1+\delta} = \max\{\sup_x |d(x)|, \sup_x |d'(x)|\} + \sup_{x,y} \frac{|d'(x) - d'(y)|}{|x-y|^\delta} \le 1,$$

and $\tilde{C}_1^{1+\delta}(R, [0,1])$ is the class of differentiable functions $d : R \times [0,1] \to \mathbb{R}$ (with respect to the first component) such that $\|d\|_{1+\delta} \le 2$, $\inf_x\{d(x)\} \ge \frac{1}{2}$. Then, one has to show that $\frac{\hat{\sigma}}{\sigma} \in C_1^{1+\delta}(R, [0,1])$, $\frac{\hat{m}-m}{\sigma} \in \tilde{C}_1^{1+\delta}(R, [0,1])$ based on assumptions on m, σ, K and provide the entropy of $\mathcal{F}$, so that

$$\big(\sqrt{n}\big(\hat{G}_\varepsilon(x) - G_\varepsilon(x)\big)\big)_{x \in \mathbb{R}} = \big(\mathbb{G}_n(f_{x, \frac{\hat{m}-m}{\sigma}, \frac{\hat{\sigma}}{\sigma}})\big)_{f_{x, \frac{\hat{m}-m}{\sigma}, \frac{\hat{\sigma}}{\sigma}} \in \mathcal{F}}$$

can be discussed with our theory.

The following example shows how to provide results for Nadaraya-Watson-type estimators. In particular, we make use of the maximal inequality provided in Section 4, Corollary 4.2.

Example 2 (Kernel density estimation): Let $K : \mathbb{R} \to \mathbb{R}$ be some bounded kernel function which is Lipschitz continuous, satisfies $\int K(u)du = 1$ and has support $\subset [-\frac{1}{2}, \frac{1}{2}]$. For some bandwidth $h_n > 0$, put $K_h(\cdot) := \frac{1}{h}K(\frac{\cdot}{h})$.

We consider the localized density estimate of the density $g_{\tilde{X}_1(v)}$ of the stationary approximation $\tilde{X}_1(v)$,

$$\hat{g}_{n,h_n}(x,v) = \frac{1}{n}\sum_{i=1}^{n} K_{h_{1,n}}(\frac{i}{n}-v)K_{h_{2,n}}(X_i - x)$$

where $h_{1n}, h_{2n} > 0$ are bandwidths. Suppose that:

- For some $s \le \frac{1}{2}$, $\alpha > s^{-1}$, $\delta_{2s}^X(j) = O(j^{-\alpha})$ and $\sup_{i,n} \|X_i\|_{2s} < \infty$.
- There exists $p_K \ge 2s, C_K > 0$ such that for u large enough, $|K(u)| \le C_K|u|^{-p_K}$.
- There exist constants $C_\infty, L_G > 0$ such that the following holds. The conditional density $g_{X_i|X_{i-1}=z}$ of X_i given $X_{i-1} = z$ satisfies $|g_{X_i|X_{i-1}=z}|_\infty \le C_\infty$ and for any $x \in \mathbb{R}$, $z \mapsto g_{X_i|X_{i-1}=z}(x)$ is Lipschitz continuous with constant L_G.

We show that if $\log(n)\big(nh_{1,n}h_{2,n}^{\frac{\alpha(s\wedge\frac{1}{2})}{\alpha(s\wedge\frac{1}{2})-1}}\big)^{-1} = O(1)$,

$$\sup_{x\in\mathbb{R},v\in[0,1]} \big|\hat{g}_{n,h_n}(x,v) - \mathbb{E}\hat{g}_{n,h_n}(x,v)\big| = O_p\Big(\sqrt{\frac{\log(n)}{nh_{1,n}h_{2,n}}}\Big). \tag{2.13}$$

To do so, note that

$$\sqrt{nh_{1,n}h_{2,n}}\big(\hat{g}_{n,h_n}(x,v) - \mathbb{E}\hat{g}_{n,h_n}(x,v)\big) = \mathbb{G}_n(f_{x,v}),$$

with

$$\mathcal{F} = \{f_{x,v}(z,u) = \sqrt{h_{1,n}}K_{h_{1,n}}(u-v)\cdot\sqrt{h_{2,n}}K_{h_{2,n}}(z-x) : x\in\mathbb{R}, v\in[0,1]\}.$$

To obtain (2.13), we use Corollary 4.2. We have for $\kappa \in \{1,2\}$,

$$\begin{aligned}
\mu^{(\kappa)}(z) \quad &:= \quad \frac{1}{h_{2,n}}\mathbb{E}[K_{h_{2,n}}(X_i - x)^\kappa \mid X_{i-1} = z]^\kappa \\
&= \quad \frac{1}{\sqrt{h_{2,n}}}\Big(\int K\big(\frac{y-x}{h_{2,n}}\big)^\kappa f_{X_i|X_{i-1}=z}(y)dy\Big)^{1/\kappa} \\
&= \quad h_{2,n}^{\frac{1}{\kappa}-\frac{1}{2}}\Big(\int K(w)^\kappa f_{X_i|X_{i-1}=z}(x + wh_{2,n})dw\Big)^{1/\kappa}.
\end{aligned}$$

Hence,

$$\begin{aligned}
&|\mu^{(\kappa)}(z) - \mu^{(\kappa)}(z')| \\
&\quad\le h_{2,n}^{\frac{1}{\kappa}-\frac{1}{2}}\Big(\int |K(w)|^\kappa |f_{X_i|X_{i-1}=z}(x+wh_{2,n}) - f_{X_i|X_{i-1}=z'}(x+wh_{2,n})|dw\Big)^{1/\kappa}.
\end{aligned}$$

On the other hand, $|f_{X_i|X_{i-1}=z}(x+wh_{2,n}) - f_{X_i|X_{i-1}=z'}(x+wh_{2,n})| \le \min\{L_G|z-z'|, C_\infty\}$. For $s \le \frac{1}{\kappa}$, we obtain

$$|\mu^{(\kappa)}(z) - \mu^{(\kappa)}(z')| \quad \le \quad h_{2,n}^{\frac{1}{\kappa}-\frac{1}{2}}\Big(\int |K(w)|^\kappa dw\Big)^{1/\kappa}\cdot\Big[C_\infty \min\big\{1, \frac{L_G}{C_\infty}|z-z'|\big\}\Big]^{1/\kappa}$$

$$\leq \quad h_{2,n}^{\frac{1}{\kappa}-\frac{1}{2}}\Big(\int |K(w)|^{\kappa}\Big)^{1/\kappa} C_{\infty}^{\frac{1}{\kappa}-s} L_G^s |z-z'|^s.$$

This shows that Assumption 2.2 is satisfied with $R(\cdot) = \frac{1}{2} = C_R$ and $L_{\mathcal{F}} = L_G$ and $\Delta(k) = L_G(k-1)^{-\alpha s}$. As before, it is easily seen that Assumption 2.7 is satisfied.

We apply Corollary 4.2 with $\bar{F} = \frac{|K|_\infty}{\sqrt{h_{2,n}}} =: C_{\bar{F},n}$. For the grids $V_n = \{in^{-3} : i = 1, ..., n^3\}$, $\mathcal{X}_n = \{in^{-3} : i \in \{-2\lceil n^{3+\frac{1}{2s}}\rceil, ..., 2\lceil n^{3+\frac{1}{2s}}\rceil\}\}$, we obtain

$$\begin{aligned} &\sqrt{nh_{1,n}h_{2,n}} \sup_{x\in\mathcal{X}_n, v\in V_n} \big|\hat{g}_{n,h_n}(x,v) - \mathbb{E}\hat{g}_{n,h_n}(x,v)\big| \\ = &\sup_{x\in\mathcal{X}_n, v\in V_n} |\mathbb{G}_n(f_{x,v})| = O_p\big(\sqrt{\log(n)}\big). \end{aligned}$$

The discretization of (2.13) is rather standard and postponed to the Appendix, Section A.4. Under additional conditions on the process, it is possible to replace $\mathbb{E}\hat{g}_{n,h}(x,v)$ by $g_{\tilde{X}_1(v)}(x)$ by introducing a typical bias of the order $O(\sqrt{nh_{1,n}^3} + \sqrt{nh_{2,n}^3})$. However, these are purely analytical considerations and are omitted.

Remark 2.12. As mentioned in the introduction, our theory allows us to discuss functions that depend on the whole infinite past. As an example let us consider M-estimation. In the (stationary) MA(1)-model

$$X_i = \theta\varepsilon_{i-1} + \varepsilon_i$$

with i.i.d. $\varepsilon_i \sim N(0,1)$, we perform a maximum likelihood estimation of θ. Then, the quasi log-Likelihood is, up to some constants, of the form

$$L_n^\circ(\theta) = \frac{1}{n}\sum_{i=1}^n f(Z_i^\circ, \theta), \qquad Z_i^\circ = (X_i, X_{i-1}, ..., X_1, 0, 0, ...)$$

where $f : \mathbb{R}^{\mathbb{N}} \times [-1,1] \to \mathbb{R}$, $f(x,\theta) = \sum_{k=0}^\infty (-\theta)^k x_k$,. The asymptotics of $L_n^\circ(\theta)$ can then be studied via

$$L_n(\theta) = \frac{1}{n}\sum_{i=1}^n f(Z_i, \theta),$$

which now depends on infinitely many past observation X_i, $-\infty < i \leq n$. Such a replacement can be performed uniformly in θ with high probability, as long as f is summable with respect to its first argument. The advantage of this replacement lies in the fact that f does not have to depend on i.

3. A general central limit theorem for locally stationary processes

In this section, we provide a multivariate central limit theorem for $\mathbb{G}_n(f)$.

It is possible to show the following analogue of a multivariate central limit theorem for $\mathbb{G}_n(f)$ as in Theorem [14, Theorem 3.4]. The proof is similar to the proof given in [14, Theorem 3.4]; the only difference appears in [14, Lemma 7.8] for which we supply the proof in the Appendix, Section A.3, Lemma A.6 under the different Assumptions 2.3, 2.4, 2.5 and 2.6.

Theorem 3.1. *Suppose that $\mathcal{F}$ satisfies Assumptions 2.2, 2.3, 2.5, 2.4 and 2.6. Let $m \in \mathbb{N}$, $f_1, ..., f_m \in \mathcal{F}$ and $\Sigma^{\mathbb{K}} = \Sigma^{(\mathbb{K})}_{f_k,f_l})_{k,l=1,...,m}$. Then,*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{\begin{pmatrix} f_1(Z_i, \frac{i}{n}) \\ \vdots \\ f_m(Z_i, \frac{i}{n})\end{pmatrix} - \mathbb{E}\begin{pmatrix} f_1(Z_i, \frac{i}{n}) \\ \vdots \\ f_m(Z_i, \frac{i}{n})\end{pmatrix}\Big\} \xrightarrow{d} N(0, \Sigma^{\mathbb{K}}),$$

where $\Sigma^{(\mathbb{K})}$ is from Assumption 2.6.

4. Maximal inequalities and asymptotic tightness under functional dependence

We now provide an approach for empirical process theory if the class $\mathcal{F}$ consists of nonsmooth functions. Our approach is based on the decomposition

$$\mathbb{G}_n(f) = \mathbb{G}_n^{(1)}(f) + \mathbb{G}_n^{(2)}(f)$$

into a martingale

$$\mathbb{G}_n^{(1)}(f) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big\{f(Z_i, \frac{i}{n}) - \mathbb{E}[f(Z_i, \frac{i}{n})|Z_{i-1}]\big\}$$

and a process

$$\mathbb{G}_n^{(2)}(f) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big\{\mathbb{E}[f(Z_i, \frac{i}{n})|Z_{i-1}] - \mathbb{E}f(Z_i, \frac{i}{n})\big\}$$

which is smooth with respect to the arguments Z_i if Assumption 2.2 is fulfilled. The second part $\mathbb{G}_n^{(2)}$ can then be controlled in a similar way as done in [14, Section 4], therefore this term is only discussed in the Appendix. Note that [14] cannot be directly applied because Assumption 2.2 therein asks for all elements of a function class $\mathcal{F}$ to be at least Hölder continuous (or of (L, s, R, C)-class for some L, s, R, C). The term $\mathbb{G}_n^{(1)}$ is dealt with by using a Bernstein-type inequality for martingales. Observe that the conditional variance of $\mathbb{G}_n^{(1)}(f)$ on Z_{i-1} is bounded from above by

$$R_n^2(f) := \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[f(Z_i, \frac{i}{n})^2|Z_{i-1}].$$

The first step is now to bound $R_n^2(f)$ uniformly over $f \in \mathcal{F}$.

4.1. Maximal inequalities

Based on $\beta(\cdot)$ from equation (2.5), we define

$$q^*(x) := \min\{q \in \mathbb{N} : \beta(q) \le q \cdot x\}.$$

Let $D_n^\infty(u) := \sup_{f\in\mathcal{F}} |D_{f,n}(u)|$, where $D_{f,n}$ is given by equation (2.3). For $\nu \ge 2$, choose $\mathbb{D}_{\nu,n}^\infty$ such that

$$\Big(\frac{1}{n}\sum_{i=1}^n D_n^\infty(\frac{i}{n})^\nu\Big)^{1/\nu} \le \mathbb{D}_{\nu,n}^\infty. \tag{4.1}$$

Put $\mathbb{D}_n^\infty = \mathbb{D}_{2,n}^\infty$. For $\delta > 0$, let

$$r(\delta) := \max\{r > 0 : q^*(r)r \le \delta\}.$$

The specific values for $q^*(\cdot)$ and r under polynomial and exponential decaying $\Delta(\cdot)$ are given in the table below.

TABLE 2
Equivalent expressions of $q^*(\cdot)$ and $r(\cdot)$ taken from [14, Lemma 7.10]. We omitted the lower and upper bound constants which are only depending on C, ρ, α.

	$\Delta(j)$	
	$Cj^{-\alpha}$, $\alpha > 1$	$C\rho^j$, $\rho \in (0,1)$
$q^*(x)$	$\max\{x^{-\frac{1}{\alpha}}, 1\}$	$\max\{\log(x^{-1}), 1\}$
$r(\delta)$	$\min\{\delta^{\frac{\alpha}{\alpha-1}}, \delta\}$	$\min\{\frac{\delta}{\log(\delta^{-1})}, \delta\}$

Recall V_n from equation (2.4) and $H = H(|\mathcal{F}|) = 1 \vee \log|\mathcal{F}|$ as in (1.5). We have the following theorem.

Theorem 4.1 (Controlling the variance)**.** *Let $\mathcal{F}$ satisfy $|\mathcal{F}| < \infty$ and Assumption 2.2. Then there exists some universal constant $c > 0$ such that the following holds. If $\sup_{f\in\mathcal{F}}\|f\|_\infty \le M$ and $\sup_{f\in\mathcal{F}} V_n(f) \le \sigma$, then*

$$\mathbb{E}\max_{f\in\mathcal{F}}\Big|R_n^2(f) - \mathbb{E}R_n^2(f)\Big| \le c \cdot \min_{q\in\{1,\dots,n\}}\Big[\mathbb{D}_n r(\frac{\sigma}{\mathbb{D}_n})\sigma + C_\Delta(\mathbb{D}_n^\infty)^2\beta(q) + \frac{qM^2H}{n}\Big]. \tag{4.2}$$

Furthermore,

$$\mathbb{E}\max_{f\in\mathcal{F}}\Big|R_n^2(f) - \mathbb{E}R_n^2(f)\Big| \le 2c \cdot \Big[\mathbb{D}_n r(\frac{\sigma}{\mathbb{D}_n})\sigma + q^*\big(\frac{M^2H}{n(\mathbb{D}_n^\infty)^2C_\Delta}\big)\frac{M^2H}{n}\Big]. \tag{4.3}$$

Theorem 4.1 in conjunction with [14, Theorem 4.1] can be used to provide uniform convergence rates for $\mathbb{G}_n(f)$.

Corollary 4.2 (Uniform convergence rates)**.** *Suppose that $\mathcal{F}$ satisfies $|\mathcal{F}| < \infty$, Assumption 2.2 for some $\nu \ge 2$, and Assumption 2.7. Let $\bar{F} := \sup_{f\in\mathcal{F}}\bar{f}$ and assume that for some $\nu_2 \in [2,\infty]$,*

$$C_{\bar{F},n} := \sup_{i,u}\|\bar{F}(Z_i,u)\|_{\nu_2} < \infty.$$

If

$$\sup_{n\in\mathbb{N}}\sup_{f\in\mathcal{F}} V_n(f) < \infty, \qquad \sup_{n\in\mathbb{N}} \frac{\mathbb{D}^{\infty}_{\nu_2,n}}{\mathbb{D}^{\infty}_n} < \infty, \qquad \sup_{n\in\mathbb{N}} \frac{C^2_{\bar{F},n} H}{n^{1-\frac{2}{\nu_2}} r(\frac{\sigma}{\mathbb{D}^{\infty}_n})^2} < \infty, \tag{4.4}$$

then

$$\max_{f\in\mathcal{F}} |\mathbb{G}_n(f)| = O_p(\sqrt{H}).$$

4.2. Asymptotic tightness

In this section, we extend the maximal inequality from Theorem 4.1 to arbitrary (infinite) classes $\mathcal{F}$. Here, the submultiplicativity assumption on $\beta(\cdot)$ from (2.5) becomes important again.

Recall $H(k) = 1 \vee \log(k)$, $\mathbb{D}_n$ as in Assumption 2.2, $\mathbb{D}^{\infty}_n = \mathbb{D}^{\infty}_{2,n}$ as in equation (4.1) and $\mathbb{H}(\varepsilon, \mathcal{F}, V)$ as the bracketing entropy. For $n \in \mathbb{N}$, $\delta > 0$, define

$$m(n, \delta, k) := r(\frac{\delta}{\mathbb{D}_n}) \cdot \frac{\mathbb{D}^{\infty}_n n^{1/2}}{H(k)^{1/2}}. \tag{4.5}$$

Here, $m(n, \delta, k)$ represents the threshold for rare events in the chaining procedure. We have the following maximal inequality.

Theorem 4.3. *Let $\mathcal{F}$ satisfy Assumption 2.2 and 2.7, and F be some envelope function of $\mathcal{F}$. Furthermore, let $\sigma > 0$ and suppose that $\sup_{f\in\mathcal{F}} V_n(f) \le \sigma$. Let ψ be defined as in (2.9). Then there exists a universal constant $c > 0$ such that for each $\eta > 0$,*

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{f\in\mathcal{F}} \big|\mathbb{G}^{(1)}_n(f)\big| > \eta\Big) \\
\le\;& \frac{1}{\eta}\Big[c\Big(1 + \frac{\mathbb{D}^{\infty}_n}{\mathbb{D}_n} + \frac{\mathbb{D}_n}{\mathbb{D}^{\infty}_n}\Big) \cdot \int_0^{\sigma} \psi(\varepsilon)\sqrt{1 \vee \mathbb{H}\big(\varepsilon, \mathcal{F}, V\big)}\, \mathrm{d}\varepsilon \\
+\;& \sqrt{n}\big\|F \mathbb{1}_{\{F > \frac{1}{4} m(n,\sigma,\mathbb{N}(\frac{\sigma}{2},\mathcal{F},V_n))\}}\big\|_1\Big] \\
+\;& c\Big(1 + q^*(C_{\Delta}^{-1} C_{\beta}^{-2})\Big(\frac{\mathbb{D}^{\infty}_n}{\mathbb{D}_n}\Big)^2\Big)\int_0^{\sigma} \frac{1}{\varepsilon\psi(\varepsilon)^2} d\varepsilon.
\end{aligned} \tag{4.6}$$

Remark 4.4. Let $m > 0$. The chaining procedure found in [13] for martingales uses the fact that for functions f, g with $|f| \le g$ and $g(\cdot) > m$,

$$|\mathbb{G}^{(1)}_n(f)| \le |\mathbb{G}^{(1)}_n(g)| + 2\sqrt{n} \cdot \frac{1}{n}\sum_{i=1}^n \mathbb{E}[g(Z_i, \frac{i}{n})|Z_{i-1}] \le |\mathbb{G}^{(1)}_n(g)| + 2\sqrt{n}\frac{R^2_n(g)}{m}.$$

Afterwards, bounds for the conditional variance $R^2_n(g)$ are applied. In our case, these bounds are not sharp enough. We therefore employ the inequality

$$|\mathbb{G}^{(1)}_n(f)| \le |\mathbb{G}^{(1)}_n(g)| + 2|\mathbb{G}^{(2)}_n(g)| + 2\sqrt{n}\frac{\|g\|^2_{2,n}}{m}$$

and are forced to use the "smooth" chaining technique applied on $\mathbb{G}_n^{(2)}(g)$ as in [14, Theorem 4.4] and on $R_n^2(g)$ from Theorem 4.1.

We now obtain asymptotic equicontinuity of the process $\mathbb{G}_n(f)$ by using Theorem 4.3 for $\mathbb{G}_n^{(1)}$ and [14, Theorem 4.4] for $\mathbb{G}_n^{(2)}$.

Corollary 4.5. *Let $\mathcal{F}$ satisfy the Assumptions 2.2, 2.7, 2.5, 2.4 and 2.3. For ψ from (2.9), suppose that*

$$\sup_{n\in\mathbb{N}} \int_0^\infty \psi(\varepsilon)\sqrt{1 \vee \mathbb{H}(\varepsilon, \mathcal{F}, V_n)}d\varepsilon < \infty. \tag{4.7}$$

Furthermore, let $\mathbb{D}_n, \mathbb{D}_n^\infty \in (0,\infty)$ be independent of n, and

$$\sup_{i=1,\dots,n} \frac{D_n^\infty(\frac{i}{n})}{\sqrt{n}} \to 0. \tag{4.8}$$

Then, the process $\mathbb{G}_n(f)$ is equicontinuous with respect to V, that is, for every $\eta > 0$,

$$\lim_{\sigma\to 0} \limsup_{n\to\infty} \mathbb{P}\Big(\sup_{f,g\in\mathcal{F}, V(f-g)\le\sigma} |\mathbb{G}_n(f) - \mathbb{G}_n(g)| \ge \eta\Big) = 0.$$

Remark 4.6. Compared to [14, Corollary 4.5], the condition (4.7) of Corollary 4.5 is not optimal due to the additional (small, but nonconstant) $\psi(\varepsilon)$-factor (cf. (2.9)). The reason here is that we do not approximate the distance $R_n^2(\cdot)$ uniformly over the class $\mathcal{F}$ in an external step but evaluate the needed bounds for $R_n^2(\cdot)$ during the chaining process. This is also the reason why our result does not include the i.i.d. version as a special case. However, in comparison to the results of [14, Lemma 7.12] we do not lose much due to this factor in the presence of polynomial dependence. Even in the case of exponential decay, the additional factor is of the same size as the factor already contributed due to dependence.

In comparison to our approach in [14], the smoothness assumptions on f therein, allow us to directly calculate the functional dependence measure of the process $f(Z_i, \frac{i}{n})$. If f itself is nonsmooth we build up our theory upon the decomposition of $\mathbb{G}_n(f)$ into $\mathbb{G}_n^{(1)}(f)$ and $\mathbb{G}_n^{(2)}(f)$. The calculations regarding the conditional expectations $\mathbb{E}[f(Z_i, \frac{i}{n})|Z_{i-1}]$ in $\mathbb{G}_n^{(1)}(f)$ now have to be adjusted. This incorporates further technicalities in the proofs. More precisely, the tightness argument becomes more complex. However, as a martingale difference we can discuss it with Friedman's martingale inequalities, which in turn lead us to impose conditions on the variance $\mathbb{E}[f(Z_i, \frac{i}{n})^2|Z_{i-1}]^{1/2}$. At the same time, the proof of the central limit theorem for finite-dimensional distributions has to be adapted to our new setting as we also have to deal with the limit distribution of the variance. So, our theory here can be seen as a generalization of [14], although small compromises have to be made in form of the entropy condition and proof strategy.

5. Conclusion

In this paper, we have developed an empirical process theory for locally stationary processes and function classes of possibly nonsmooth functions. Here, the dependence was quantified with the functional dependence measure. We have proven maximal inequalities and functional central limit theorems. An empirical process theory for locally stationary processes is a key step to derive asymptotic and nonasymptotic results for a large class of time series.

We have shown that our theory can be applied to empirical distribution functions (EDFs) and kernel density estimators, but much more structures can be discussed. Compared to earlier papers in the context of stationary processes and the EDF, our results provide remarkable weak conditions on the dependence decay of the process. In particular, compared to [12], we could prove that functional weak convergence of the EDF holds under much simpler assumptions.

From a technical point of view, the linear and moment-based nature of the functional dependence measure has forced us to modify several approaches from [14]. A main issue was given by the fact that the dependence measure only transfers decay rates of continuous functions. The nonsmooth nature of the function class was dealt with a decomposition into a martingale and a conditional expectation part.

Appendix A: Appendix

We now provide some proof details for the main sections.

A.1. Proofs of Section 2

Lemma A.1. *Let Assumption 2.2 hold for some $\nu \geq 2$. Then for all $u \in [0,1]$,*

$$\delta_\nu^{\mathbb{E}[f(Z_i,u)|Z_{i-1}]}(k) \leq |D_{f,n}(u)| \cdot \Delta(k), \quad (A.1)$$

$$\sup_i \Big\| \sup_{f\in\mathcal{F}} \big|\mathbb{E}[f(Z_i,u)|Z_{i-1}] - \mathbb{E}[f(Z_i,u)|Z_{i-1}]^{*(i-k)}\big|\Big\|_\nu \leq D_n^\infty(u) \cdot \Delta(k), \quad (A.2)$$

$$\sup_i \|f(Z_i,u)\|_2 \leq |D_{f,n}(u)| \cdot C_\Delta. \quad (A.3)$$

Furthermore,

$$\Big\|\mathbb{E}[f(Z_i,u)^2|Z_{i-1}] - \mathbb{E}[f(Z_i,u)^2|Z_{i-1}]^{*(i-k)}\Big\|_{\nu/2} \leq 2|D_{f,n}(u)| \cdot \|f(Z_i,u)\|_\nu \cdot \Delta(k), (A.4)$$

$$\Big\| \sup_{f\in\mathcal{F}} \big|\mathbb{E}[f(Z_i,u)^2|Z_{i-1}] - \mathbb{E}[f(Z_i,u)^2|Z_{i-1}]^{*(i-k)}\big|\Big\|_{\nu/2} \leq D_n^\infty(u)^2 \cdot C_\Delta \cdot \Delta(k),$$

(A.5)

where $C_\Delta := 2\max\{d,\tilde{d}\}|L_{\mathcal{F}}|_1 C_X^s C_R + C_{\bar{f}}$.

Proof of Lemma A.1. Let $\bar{\mu}_{f,i}^{(1)}(z,u) = \mathbb{E}[\bar{f}(Z_i,u)|Z_{i-1} = z]$ and $\bar{\mu}_{f,i}^{(2)}(z,u) = \mathbb{E}[\bar{f}(Z_i,u)^2|Z_{i-1} = z]$. We have by Assumption 2.2 that

$$\begin{aligned}
&\sup_i \big\|\mathbb{E}[f(Z_i,u)|Z_{i-1}] - \mathbb{E}[f(Z_i,u)|Z_{i-1}]^{*(i-k)}\big\|_\nu \\
= \;& |D_{f,n}(u)|\cdot\sup_i \big\|\bar{\mu}_{f,i}^{(1)}(Z_{i-1},u) - \bar{\mu}_{f,i}^{(1)}(Z_{i-1}^{*(i-k)},u)\big\|_\nu \\
\le \;& |D_{f,n}(u)|\cdot\sup_i \Big\|\big|Z_{i-1} - Z_{i-1}^{*(i-k)}\big|_{L_{\mathcal{F},s}}^s\Big\|_{\frac{p\nu}{p-1}} \Big\|R(Z_{i-1},u) + R(Z_{i-1}^{*(i-k)},u)\Big\|_{p\nu} \\
\le \;& |D_{f,n}(u)|\cdot\sup_i \Big\|\sum_{j=0}^\infty L_{\mathcal{F},j}\big|X_{i-1-j} - X_{i-1-j}^{*(i-k)}\big|_\infty^s\Big\|_{\frac{p\nu}{p-1}} \\
&\qquad\times\Big(\|R(Z_{i-1},u)\|_{p\nu} + \Big\|R(Z_{i-1}^{*(i-k)},u)\Big\|_{p\nu}\Big) \\
\le \;& |D_{f,n}(u)|\cdot 2dC_R\sum_{j=0}^{k-1} L_{\mathcal{F},j}\delta_{\frac{p\nu s}{p-1}}(k-j-1)^s,
\end{aligned}$$

that is, the assertion (A.1) holds with the given $\Delta(k)$. The proof of (A.2) is similar.

We now prove (A.3). We have

$$\mathbb{E}[f(Z_i,u)^2] = \mathbb{E}[\mathbb{E}[f(Z_i,u)^2|Z_{i-1}]] = D_{f,n}(u)^2\mathbb{E}[\bar{\mu}_{f,i}^{(2)}(Z_{i-1},u)^2]$$

and thus $\|f(Z_i,u)\|_2 = |D_{f,n}(u)|\cdot\|\bar{\mu}_{f,i}^{(2)}(Z_{i-1},u)\|_2$. Since

$$|\bar{\mu}_{f,i}^{(2)}(y,u)| \le |\bar{\mu}_{f,i}^{(2)}(y,u) - \bar{\mu}_{f,i}^{(2)}(0,u)| + |\bar{\mu}_{f,i}^{(2)}(0,u)|,$$

the proof now follows the same lines as in the proof of [14, Lemma 7.3].

We now show (A.4) and (A.5). We have

$$\big|\bar{\mu}_{f,i}^{(2)}(z,u)^2 - \bar{\mu}_{f,i}^{(2)}(z',u)^2\big| = \big|\bar{\mu}_{f,i}^{(2)}(z,u) - \bar{\mu}_{f,i}^{(2)}(z',u)\big|\cdot\big[|\bar{\mu}_{f,i}^{(2)}(z,u)| + |\bar{\mu}_{f,i}^{(2)}(z',u)|\big].$$

We then have by the Cauchy Schwarz inequality that

$$\begin{aligned}
&\Big\|\sup_{f\in\mathcal{F}}\big|\bar{\mu}_{f,i}^{(2)}(Z_{i-1},u)^2 - \bar{\mu}_{f,i}^{(2)}(Z_{i-1}^{*(i-k)},u)^2\big|\Big\|_{\nu/2} \\
\le\;& \Big\|\sup_{f\in\mathcal{F}}\big|\bar{\mu}_{f,i}^{(2)}(Z_{i-1},u) - \bar{\mu}_{f,i}^{(2)}(Z_{i-1}^{*(i-k)},u)\big|\Big\|_\nu \cdot 2\Big\|\sup_{f\in\mathcal{F}}\big|\bar{\mu}_{f,i}^{(2)}(Z_{i-1},u)\big|\Big\|_\nu
\end{aligned} \tag{A.6}$$

Since $\{\bar{\mu}_{f,i}^{(2)} : f\in\mathcal{F}, i\in\{1,...,n\}\}$ forms a $(L_{\mathcal{F}},s,R,C)$-class for some $L_{\mathcal{F}}$, s, R, C, the first factor in (A.6) is bounded by $\Delta(k)$ as before. Furthermore,

$$|\bar{\mu}_{f,i}^{(2)}(z,u)| \;\le\; |\bar{\mu}_{f,i}^{(2)}(z,u) - \bar{\mu}_{f,i}^{(2)}(0,u)| + |\bar{\mu}_{f,i}^{(2)}(0,u)|$$

$$\leq \quad |z|_{L_{\mathcal{F}},s}^s (R(z,u) + R(0,u)) + |\bar{\mu}_{f,i}^{(2)}(0,u)|.$$

Note that

$$\begin{aligned}
& \Big\| |Z_{i-1}|_{L_{\mathcal{F}},s}^s \cdot \big[R(Z_{i-1},u) + R(0,u)\big]\Big\|_{\nu} \\
\leq \quad & \Big\| \sum_{j=0}^{\infty} L_{\mathcal{F},j} |Z_{i-1-j}|_{\infty}^s \Big\|_{\frac{p}{p-1}\nu} \cdot \Big(\|R(Z_{i-1},u)\|_{p\nu} + |R(0,u)|\Big) \\
\leq \quad & d|L_{\mathcal{F}}|_1 \sup_{i,j} \|X_{ij}\|_{\frac{\nu s p}{p-1}}^s \cdot (C_R + |R(0,u)|) \\
\leq \quad & 2d|L_{\mathcal{F}}|_1 C_X^s C_R.
\end{aligned}$$

We now obtain (A.5) from (A.6) with the given C_Δ.

By the Cauchy-Schwarz inequality we have for $q \geq 2$,

$$\begin{aligned}
& \delta_{\nu/2}^{\mathbb{E}[f(Z_i,u)^2|Z_{i-1}]}(k) \\
= \quad & \sup_i \Big\| \mathbb{E}[f(Z_i,u)^2|Z_{i-1}] - \mathbb{E}[f(Z_i,u)^2|Z_{i-1}]^{*(i-k)} \Big\|_{\nu/2} \\
= \quad & |D_{f,n}(u)| \cdot \sup_i \Big\| D_{f,n}(u) \big(\bar{\mu}_{f,i}^{(2)}(Z_{i-1},u)^2 - \bar{\mu}_{f,i}^{(2)}(Z_{i-1}^{*(i-k)},u)^2\big)\Big\|_{\nu/2} \\
\leq \quad & |D_{f,n}(u)| \cdot \sup_i \Big\| \bar{\mu}_{f,i}^{(2)}(Z_{i-1},u) - \bar{\mu}_{f,i}^{(2)}(Z_{i-1}^{*(i-k)},u)\Big\|_{\nu} \\
& \qquad \times 2 \Big\| D_{f,n}(u)\bar{\mu}_{f,i}^{(2)}(Z_{i-1},u)\Big\|_{\nu} \qquad\qquad (A.7)
\end{aligned}$$

Furthermore,

$$\Big\| D_{f,n}(u)\bar{\mu}_{f,i}^{(2)}(Z_{i-1},u)\Big\|_{\nu} \leq \|\mathbb{E}[f(Z_i,u)^2|Z_{i-1}]^{1/2}\|_{\nu} \leq \|f(Z_i,u)\|_{\nu}. \qquad (A.8)$$

Since Assumption 2.2 holds for $\bar{\mu}_{f,i}^{(2)}$, the first factor in (A.7) is bounded by $D_{f,n}(u)\Delta(k)$ as in the proof of [14, Lemma 7.3]. Inserting this and (A.8) into (A.7), we obtain the result (A.4). □

A.2. Proofs of Section 4.1

A.2.1. Proof of Theorem 4.1

In this section, we consider

$$W_i(f) = \mathbb{E}[f(Z_i,\frac{i}{n})^2|Z_{i-1}], \qquad S_n(f) := \sum_{i=1}^n \big\{W_i(f) - \mathbb{E}W_i(f)\big\}.$$

Then

$$R_n(f)^2 = \frac{1}{n}\sum_{i=1}^n W_i(f), \qquad R_n(f)^2 - \mathbb{E}R_n(f)^2 = \frac{1}{n}S_n(f).$$

We obtain from Lemma A.1, (A.4) and (A.5) the following results with $\nu = 2$.

Lemma A.2. *Suppose that Assumption 2.2 holds. Then for each $i = 1, ..., n$, $j \in \mathbb{N}$, $s \in \mathbb{N} \cup \{\infty\}$, $f \in \mathcal{F}$,*

$$\begin{aligned}\Big\| \sup_{f\in\mathcal{F}} \big|W_i(f) - W_i(f)^{*(i-j)}\big| \Big\|_1 &\le C_\Delta D_n^\infty(\frac{i}{n})^2\Delta(j),\\ \big\|W_i(f) - W_i(f)^{*(i-j)}\big\|_1 &\le 2|D_{f,n}(\frac{i}{n})| \cdot \|f(Z_i, \frac{i}{n})\|_2\Delta(j),\\ \big\|W_i(f)\|_s &\le \|f(Z_i, \frac{i}{n})\|_{2s}^2.\end{aligned}$$

We approximate $W_i(f)$ by independent variables as follows (cf. also [20], [22]). Let

$$W_{i,j}(f) := \mathbb{E}[W_i(f)|\varepsilon_{i-j}, \varepsilon_{i-j+1}, ..., \varepsilon_i], \qquad j \in \mathbb{N},$$

and

$$S_{n,j}(f) := \sum_{i=1}^n \{W_{i,j}(f) - \mathbb{E}W_{i,j}(f)\}.$$

Let $q \in \{1, ..., n\}$ be arbitrary. Put $L := \lfloor \frac{\log(q)}{\log(2)} \rfloor$ and $\tau_l := 2^l$ $(l = 0, ..., L-1)$, $\tau_L := q$. Then we have

$$W_i(f) = W_i(f) - W_{i,q}(f) + \sum_{l=1}^L (W_{i,\tau_l}(f) - W_{i,\tau_{l-1}}(f)) + W_{i,1}(f)$$

(in the case $q = 1$, the sum in the middle does not appear) and thus

$$S_n(f) = \big[S_n(f) - S_{n,q}(f)\big] + \sum_{l=1}^L \big[S_{n,\tau_l}(f) - S_{n,\tau_{l-1}}(f)\big] + S_{n,1}(f).$$

We write

$$\begin{aligned}S_{n,\tau_l}(f) - S_{n,\tau_{l-1}}(f) &= \sum_{i=1}^{\lfloor \frac{n}{\tau_l}\rfloor+1} T_{i,l}(f),\\ T_{i,l}(f) &:= \sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n} \big[W_{k,\tau_l}(f) - W_{k,\tau_{l-1}}(f)\big].\end{aligned}$$

The random variables $T_{i,l}(f), T_{i',l}(f)$ are independent if $|i - i'| > 1$. This leads to the decomposition

$$\begin{aligned}\max_{f\in\mathcal{F}} \Big|\frac{1}{n}S_n(f)\Big| &\le \max_{f\in\mathcal{F}} \frac{1}{n}\big|S_n(f) - S_{n,q}(f)\big|\\ &\quad + \sum_{l=1}^L \Big[\max_{f\in\mathcal{F}} \Big|\frac{1}{\frac{n}{\tau_l}} \sum_{\substack{i=1\\ i \text{ even}}}^{\lfloor \frac{n}{\tau_l}\rfloor+1} \frac{1}{\tau_l}T_{i,l}(f)\Big| + \max_{f\in\mathcal{F}} \Big|\frac{1}{\frac{n}{\tau_l}} \sum_{\substack{i=1\\ i \text{ odd}}}^{\lfloor \frac{n}{\tau_l}\rfloor+1} \frac{1}{\tau_l}T_{i,l}(f)\Big|\Big]\end{aligned}$$

$$
\begin{aligned}
&\quad + \max_{f\in\mathcal{F}} \frac{1}{n}\big|S_{n,1}(f)\big| \\
&=: \quad A_1 + A_2 + A_3. \qquad\qquad (A.9)
\end{aligned}
$$

The next result is a uniform bound on means of independent random variables.

Lemma A.3. *Assume that $Q_i(f)$, $i = 1, ..., m$ are independent variables indexed by $f \in \mathcal{F}$ which fulfill $\mathbb{E}Q_i(f) = 0$, $\frac{1}{m}\sum_{i=1}^m \|Q_i(f)\|_1 \le \sigma_Q$ and $|Q_i(f)| \le M_Q$ a.s. $(i = 1, ..., n)$. Then there exists some universal constant $c > 0$ such that*

$$
\mathbb{E}\max_{f\in\mathcal{F}}\Big|\frac{1}{m}\sum_{i=1}^m Q_i(f)\Big| \le c\Big(\sigma_Q + \frac{M_Q H}{m}\Big), \qquad (A.10)
$$

where H is defined by (1.5).

Proof of Lemma A.3. Let $Q_i = Q_i(f)$. By Bernstein's inequality, we have for each $f \in \mathcal{F}$ that

$$
\begin{aligned}
\mathbb{P}\Big(\Big|\frac{1}{m}\sum_{i=1}^m Q_i\Big| \ge x\Big) \quad &\le \quad 2\exp\Big(-\frac{1}{2}\frac{x^2}{\frac{1}{m^2}\sum_{i=1}^m \|Q_i\|_2^2 + x\frac{M_Q}{m}}\Big) \\
&\le \quad 2\exp\Big(-\frac{1}{2}\frac{x^2}{\frac{M_Q}{m}\cdot\sigma_Q + x\frac{M_Q}{m}}\Big),
\end{aligned}
$$

where we used in the last step that $\|Q_i\|_2^2 = \mathbb{E}[Q_i^2] \le M_Q\|Q_i\|_1$.

With standard arguments (cf. the proof of Lemma 19.33 in [17]), we conclude that there exists some universal constant $c_1 > 0$ with

$$
\mathbb{E}\max_{f\in\mathcal{F}}\Big|\frac{1}{m}\sum_{i=1}^m Q_i(f)\Big| \quad \le \quad c_1\Big(\sqrt{H}\big(\frac{\sigma_Q M_Q}{m}\big)^{1/2} + \frac{M_Q H}{m}\Big).
$$

The result follows by using $(\frac{H\sigma_Q M_Q}{m})^{1/2} \le 2\frac{M_Q H}{m} + 2\sigma_Q$. ☐

We now prove Theorem 4.1 based on Lemma A.2 and Lemma A.3 and the decomposition (A.9).

Proof of Theorem 4.1. We first discuss A_2. We have

$$
\sum_{l=1}^L \mathbb{E}\max_{f\in\mathcal{F}}\frac{1}{\frac{n}{\tau_l}}\Big|\sum_{1\le i\le\lfloor\frac{n}{\tau_l}\rfloor+1, i\text{ odd}}\frac{1}{\tau_l}T_{i,l}(f)\Big|.
$$

Since $\|W_{k,j}(f) - W_{k,j-1}(f)\|_1 \le 2\min\{\|W_k(f)\|_1, \delta_1^{W_k(f)}(j-1)\}$, we have for each $f \in \mathcal{F}$,

$$
\frac{1}{\tau_l}\|T_{i,l}\|_1 \quad \le \quad \sum_{j=\tau_{l-1}+1}^{\tau_l}\frac{1}{\tau_l}\Big\|\sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n}(W_{k,j} - W_{k,j-1})\Big\|_1
$$

$$
\begin{aligned}
&\leq \sum_{j=\tau_{l-1}+1}^{\tau_l} \frac{1}{\tau_l} \sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n} \Big\| W_{k,j} - W_{k,j-1} \Big\|_1 \\
&\leq 2 \sum_{j=\tau_{l-1}+1}^{\tau_l} \frac{1}{\tau_l} \sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n} \min\{\|W_k(f)\|_1, \delta_1^{W_k(f)}(j-1)\} \\
&\leq 2 \sum_{j=\tau_{l-1}+1}^{\tau_l} \min\{\frac{1}{\tau_l} \sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n} \|W_k(f)\|_1, \frac{1}{\tau_l} \sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n} \delta_1^{W_k(f)}(j-1)\} \\
&= 2 \sum_{j=\tau_{l-1}+1}^{\tau_l} \min\{\sigma_{i,l}, \Delta_{i,j,l}\},
\end{aligned}
$$

where

$$
\sigma_{i,l} := \frac{1}{\tau_l} \sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n} \|W_k(f)\|_1, \qquad \Delta_{i,j,l} := \frac{1}{\tau_l} \sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n} \delta_1^{W_k(f)}(j-1).
$$

We conclude that

$$
\begin{aligned}
&\frac{1}{\lfloor \frac{n}{\tau_l} \rfloor + 1} \sum_{i=1}^{\lfloor \frac{n}{\tau_l} \rfloor + 1} \frac{1}{\tau_l} \|T_{i,l}\|_1 \\
&\leq 2 \sum_{j=\tau_{l-1}+1}^{\tau_l} \min\{\frac{1}{\frac{n}{\tau_l}} \sum_{i=1}^{\lfloor \frac{n}{\tau_l} \rfloor + 1} \sigma_{i,l}, \frac{1}{\frac{n}{\tau_l}} \sum_{i=1}^{\lfloor \frac{n}{\tau_l} \rfloor + 1} \Delta_{i,j,l}\} \\
&\leq \sum_{j=\tau_{l-1}+1}^{\tau_l} \min\{\frac{1}{n} \sum_{i=1}^{n} \|W_i(f)\|_1, \frac{1}{n} \sum_{i=1}^{n} \delta_1^{W_i}(j)\}. \qquad \text{(A.11)}
\end{aligned}
$$

Furthermore, it holds that

$$
\frac{1}{\tau_l}|T_{i,l}| \leq 2 \sup_i \|W_i(f)\|_\infty \leq 2\|f\|_\infty^2 \leq 2M^2. \qquad \text{(A.12)}
$$

By Lemma A.3, (A.10), we have with some universal constant $c_1 > 0$ that

$$
\begin{aligned}
&\mathbb{E}A_2 \\
&\leq 2c_1 \sum_{l=1}^{L} \Big[\sup_{f\in\mathcal{F}} \Big(\frac{1}{\lfloor \frac{n}{\tau_l} \rfloor + 1} \sum_{i=1}^{\lfloor \frac{n}{\tau_l} \rfloor + 1} \frac{1}{\tau_l} \|T_{i,l}(f)\|_1 \Big) + \frac{2M^2 H}{\lfloor \frac{n}{\tau_l} \rfloor + 1} \Big] \\
&\leq 2c_1 \Big(\sum_{l=1}^{L} \sup_{f\in\mathcal{F}} \sum_{j=\tau_{l-1}+1}^{\tau_l} \min\{\frac{1}{n} \sum_{i=1}^{n} \|W_i(f)\|_1, \frac{1}{n} \sum_{i=1}^{n} \delta_1^{W_i(f)}(j)\} + \frac{qM^2 H}{n} \Big). \qquad \text{(A.13)}
\end{aligned}
$$

By Lemma A.2 and the Cauchy-Schwarz inequality for sums,

$$\begin{aligned}
&\sum_{l=1}^{L}\sup_{f\in\mathcal{F}}\sum_{j=\tau_{l-1}+1}^{\tau_l}\min\{\frac{1}{n}\sum_{i=1}^{n}\|W_i(f)\|_1, \frac{1}{n}\sum_{i=1}^{n}\delta_1^{W_i}(j)\}\\
\le\ &\sum_{l=1}^{L}\sup_{f\in\mathcal{F}}\sum_{j=\tau_{l-1}+1}^{\tau_l}\min\{\frac{1}{n}\sum_{i=1}^{n}\|f(Z_i,\frac{i}{n})\|_2^2, \frac{2}{n}\sum_{i=1}^{n}D_{f,n}(\frac{i}{n})\|f(Z_i,\frac{i}{n})\|_2\cdot\Delta(j)\}\\
\le\ &\sum_{j=1}^{\infty}\min\{\sup_{f\in\mathcal{F}}\|f\|_{2,n}^2, 2\mathbb{D}_n\sup_{f\in\mathcal{F}}\|f\|_{2,n}\cdot\Delta(j)\}\\
=\ &\sup_{f\in\mathcal{F}}\|f\|_{2,n}\cdot\bar{V}(\sup_{f\in\mathcal{F}}\|f\|_{2,n})\\
=\ &\sup_{f\in\mathcal{F}}\big(\|f\|_{2,n}\cdot\bar{V}(\|f\|_{2,n})\big)\\
\le\ &\sup_{f\in\mathcal{F}}\big[\|f\|_{2,n}V_n(f)\big],
\end{aligned}\tag{A.14}$$

where

$$\bar{V}(x) = x + \sum_{j=1}^{\infty}\min\{x, \mathbb{D}_n\Delta(j)\}\tag{A.15}$$

and in the second-to-last equality the fact that $x\mapsto x\cdot\bar{V}(x)$ is increasing in x.

We also have $\|W_{i,0}(f)-\mathbb{E}W_{i,0}(f)\|_\infty \le 2\|f\|_\infty^2 \le 2M^2$ and $\|W_{i,0}(f)-\mathbb{E}W_{i,0}(f)\|_1 \le 2\|W_i(f)\|_1$. Thus by Lemma A.3, (A.10),

$$\begin{aligned}
\mathbb{E}A_3 &\le \mathbb{E}\max_{f\in\mathcal{F}}\Big|\frac{1}{n}\sum_{i=1}^{n}(W_{i,0}(f)-\mathbb{E}W_{i,0}(f))\Big|\\
&\le 2c_1\Big(\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\|W_i(f)\|_1 + \frac{M^2H}{n}\Big) &\text{(A.16)}\\
&\le 2c_1\Big(\sup_{f\in\mathcal{F}}\|f\|_{2,n}^2 + \frac{M^2H}{n}\Big). &\text{(A.17)}
\end{aligned}$$

Finally,

$$\begin{aligned}
\mathbb{E}A_1 &\le \sum_{j=q}^{\infty}\mathbb{E}\sup_{f\in\mathcal{F}}\Big|\frac{1}{n}\sum_{i=1}^{n}(W_{i,j+1}(f)-W_{i,j}(f))\Big|\\
&\le \sum_{j=q}^{\infty}\frac{1}{n}\sum_{i=1}^{n}\big\|\sup_{f\in\mathcal{F}}|W_{i,j+1}(f)-W_{i,j}(f)|\big\|_1.
\end{aligned}$$

Since $|W_{i,j+1}(f)-W_{i,j}(f)| = |\mathbb{E}[W_i(f)^{**(i-j)} - W_i(f)^{**(i-j+1)}|\mathcal{A}_i]| \le \mathbb{E}[|W_i(f)^{**(i-j)} - W_i(f)^{**(i-j+1)}|\,|\mathcal{A}_i]$ where we use the notation $H(\mathcal{F}_i)^{**(i-j)} := H(\mathcal{F}_i^{**(i-j)})$ and $\mathcal{F}_i^{**(i-j)} = (\varepsilon_i, \varepsilon_{i-1}, ..., \varepsilon_{i-j}, \varepsilon_{i-j-1}^*, \varepsilon_{i-j-2}^*, ...)$., we have

$$\big\|\sup_{f\in\mathcal{F}}|W_{i,j+1}(f)-W_{i,j}(f)|\big\|_1$$

$$
\begin{aligned}
&\leq \big\|\mathbb{E}[\max_{f\in\mathcal{F}}|W_i(f)^{**(i-j)} - W_i(f)^{**(i-j+1)}|\,|\mathcal{A}_i]\big\|_1 \\
&\leq \big\|\sup_{f\in\mathcal{F}}|W_i(f)^{**(i-j)} - W_i(f)^{**(i-j+1)}|\big\|_1 \\
&= \big\|\sup_{f\in\mathcal{F}}|W_i(f) - W_i(f)^{*(i-j)}|\big\|_1 \leq D_n^{\infty}(\frac{i}{n})^2 C_\Delta \Delta(j), \qquad \text{(A.18)}
\end{aligned}
$$

which shows that

$$\mathbb{E}A_1 \leq (\mathbb{D}_n^{\infty})^2 C_\Delta \beta(q). \qquad \text{(A.19)}$$

Collecting the upper bounds (A.13), (A.14), (A.17) and (A.19), we obtain that

$$\mathbb{E}\max_{f\in\mathcal{F}}\Big|\frac{1}{n}S_n(f)\Big| \leq (4c_1+1)\cdot\Big[\sup_{f\in\mathcal{F}}\big[\|f\|_{2,n}V_n(f)\big] + (\mathbb{D}_n^{\infty})^2 C_\Delta\beta(q) + \frac{qM^2H}{n}\Big]. \qquad \text{(A.20)}$$

By (A.29), $V_n(f) \leq \sigma$ implies $\|f\|_{2,n}^2 \leq \mathbb{D}_n r(\frac{\delta}{\mathbb{D}_n})\|f\|_{2,n}$ and thus

$$\|f\|_{2,n} \leq \mathbb{D}_n r(\frac{\sigma}{\mathbb{D}_n}),$$

thus

$$\sup_{f\in\mathcal{F}}\big[\|f\|_{2,n}V_n(f)\big] \leq \mathbb{D}_n r(\frac{\sigma}{\mathbb{D}_n})\sigma. \qquad \text{(A.21)}$$

Inserting (A.21) into (A.20) yields the first assertion (4.2) of the lemma.

We now show (4.3) with a case distinction. We abbreviate $q^* = q^*(\frac{M^2H}{n(\mathbb{D}_n^{\infty})^2C_\Delta})$. If $q^*\frac{H}{n} \leq 1$, we have $q^* \in \{1,...,n\}$ and thus

$$
\begin{aligned}
P &\leq c\Big(\mathbb{D}_n r(\frac{\sigma}{\mathbb{D}_n})\sigma + (\mathbb{D}_n^{\infty})^2 C_\Delta\beta(q^*) + q^*\frac{M^2H}{n}\Big) \\
&\leq 2c\Big(\mathbb{D}_n r(\frac{\sigma}{\mathbb{D}_n})\sigma + q^*\frac{M^2H}{n}\Big) \\
&= 2c\Big(\mathbb{D}_n r(\frac{\sigma}{\mathbb{D}_n})\sigma + M^2\cdot\min\Big\{q^*\frac{H}{n},1\Big\}\Big). \qquad \text{(A.22)}
\end{aligned}
$$

If $q^*\frac{H}{n} \geq 1$, choose $q_0 = \lfloor\frac{n}{H}\rfloor \leq \frac{n}{H}$. By simply bounding each summand with M^2, we have

$$
\begin{aligned}
\mathbb{E}\max_{f\in\mathcal{F}}\Big|\frac{1}{n}S_n(f)\Big| &\leq M^2 \leq c\Big(\mathbb{D}_n r(\frac{\sigma}{\mathbb{D}_n})\sigma + M^2\Big) \\
&\leq 2c\Big(\mathbb{D}_n r(\frac{\sigma}{\mathbb{D}_n})\sigma + M^2\cdot\min\Big\{q^*\frac{H}{n},1\Big\}\Big). \qquad \text{(A.23)}
\end{aligned}
$$

holds. Putting the two bounds (A.22) and (A.23) together, we obtain the result (4.3). □

The following lemma is an auxiliary result to prove Corollary 4.2 and Lemma A.5.

Lemma A.4. *Let $\mathcal{F}$ be some finite class of functions. Let $R > 0$ be arbitrary and assume that $\sup_{f\in\mathcal{F}} \|f\|_\infty \le M$. Then there exists a universal constant $c > 0$ such that*

$$\mathbb{E}\max_{f\in\mathcal{F}} \big|\mathbb{G}_n^{(1)}(f)\big|\mathbb{1}_{\{R_n(f)^2\le R^2\}} \le c\Big\{R\sqrt{H} + \frac{MH}{\sqrt{n}}\Big\}, \tag{A.24}$$

where H is defined by (1.5).

Proof of Lemma A.4. By Theorem 3.3 in [15], it holds for $x, a > 0$ and a measurable function f that

$$\mathbb{P}\Big(\big|\mathbb{G}_n^{(1)}(f)\big| \ge x, R_n(f)^2 \le R^2\Big) \le 2\exp\Big(-\frac{1}{2}\frac{x^2}{R^2 + \frac{2\|f\|_\infty x}{3\sqrt{n}})}\Big).$$

Using standard arguments (cf. the proof of Lemma 19.33 in [17]), we obtain (A.24). □

Proof of Corollary 4.2. Let us define the following functions first.

For $m > 0$, define $\varphi_m^\wedge : \mathbb{R} \to \mathbb{R}$ and the corresponding "peaky" residual function $\varphi_m^\vee : \mathbb{R} \to \mathbb{R}$ via

$$\varphi_m^\wedge(x) := (x \vee (-m)) \wedge m, \qquad \varphi_m^\vee(x) := x - \varphi_m^\wedge(x).$$

Now, let $Q \ge 1$, and $\sigma := \sup_{n\in\mathbb{N}} \sup_{f\in\mathcal{F}} V_n(f) < \infty$. Put

$$M_n = \frac{\sqrt{n}}{\sqrt{H}} r\big(\frac{\sigma Q^{1/2}}{\mathbb{D}_n^\infty}\big)\mathbb{D}_n^\infty.$$

Let $F(z,u) := D_n^\infty(u)\cdot \bar{F}(z,u)$, (recall $\bar{F} = \sup_{f\in\mathcal{F}} \bar{f}$). Then

$$\begin{aligned}
&\mathbb{P}\Big(\max_{f\in\mathcal{F}} |\mathbb{G}_n(f)| > Q\sqrt{H}\Big)\\
\le\ & \mathbb{P}\Big(\max_{f\in\mathcal{F}} |\mathbb{G}_n(f)| > Q\sqrt{H}, \sup_{i=1,\dots,n} F(Z_i, \frac{i}{n}) \le M_n\Big)\\
&\quad + \mathbb{P}\Big(\sup_{i=1,\dots,n} F(Z_i,\frac{i}{n}) > M_n\Big)\\
\le\ & \mathbb{P}\Big(\max_{f\in\mathcal{F}} |\mathbb{G}_n(\varphi_{M_n}^\wedge(f))| > \frac{Q\sqrt{H}}{2}\Big)\\
&\quad + \mathbb{P}\Big(\frac{1}{\sqrt{n}}\max_{f\in\mathcal{F}}\big|\sum_{i=1}^n \mathbb{E}[f(Z_i,\frac{i}{n})\mathbb{1}_{\{|f(Z_i,\frac{i}{n})|>M_n\}}]\big| > \frac{Q\sqrt{H}}{2}\Big)\\
&\quad + \mathbb{P}\Big(\sup_{i=1,\dots,n} F(Z_i,\frac{i}{n}) > M_n\Big).
\end{aligned} \tag{A.25}$$

For the first summand in (A.25), we use the decomposition

$$\mathbb{P}\Big(\max_{f\in\mathcal{F}} |\mathbb{G}_n(\varphi_{M_n}^\wedge(f))| > \frac{Q\sqrt{H}}{2}\Big)$$

$$
\begin{aligned}
&\leq \mathbb{P}\Big(\max_{f\in\mathcal{F}}|\mathbb{G}_n^{(1)}(\varphi_{M_n}^{\wedge}(f))| > \frac{Q\sqrt{H}}{4}\Big) + \mathbb{P}\Big(\max_{f\in\mathcal{F}}|\mathbb{G}_n^{(2)}(\varphi_{M_n}^{\wedge}(f))| > \frac{Q\sqrt{H}}{4}\Big)\\
&\leq \mathbb{P}\Big(\max_{f\in\mathcal{F}}|\mathbb{G}_n^{(1)}(\varphi_{M_n}^{\wedge}(f))| > \frac{Q\sqrt{H}}{4}, \max_{f\in\mathcal{F}} R_n(\varphi_{M_n}^{\wedge}(f))^2 \leq \sigma^2\Big)\\
&\qquad + \mathbb{P}\Big(\max_{f\in\mathcal{F}} R_n(\varphi_{M_n}^{\wedge}(f)) > \sigma^2\Big)\\
&\qquad + \mathbb{P}\Big(\max_{f\in\mathcal{F}}|\mathbb{G}_n^{(2)}(\varphi_{M_n}^{\wedge}(f))| > \frac{Q\sqrt{H}}{4}\Big). \qquad \text{(A.26)}
\end{aligned}
$$

We now discuss the three terms separately. By Lemma A.4, we have

$$
\begin{aligned}
&\mathbb{P}\Big(\max_{f\in\mathcal{F}}|\mathbb{G}_n^{(1)}(\varphi_{M_n}^{\wedge}(f))| > \frac{Q\sqrt{H}}{4}, \max_{f\in\mathcal{F}} R_n(\varphi_{M_n}^{\wedge}(f))^2 \leq Q^{3/2}\sigma^2\Big)\\
\leq\ & \frac{4c}{Q\sqrt{H}}\Big[\sigma Q^{3/4}\sqrt{H} + \frac{M_n H}{\sqrt{n}}\Big] \leq \frac{4c}{Q\sqrt{H}}\Big[\sigma Q^{3/4}\sqrt{H} + \sigma\sqrt{H}Q^{1/2}\Big] \leq \frac{8c}{Q^{1/4}}.
\end{aligned}
$$

By Theorem 4.1 and (A.31),

$$
\begin{aligned}
&\mathbb{P}\Big(\max_{f\in\mathcal{F}} R_n(\varphi_{M_n}^{\wedge}(f))^2 > Q^{3/2}\sigma^2\Big)\\
\leq\ & \frac{2c}{\sigma^2 Q^{3/2}}\Big[\mathbb{D}_n r(\frac{\sigma}{\mathbb{D}_n})\sigma + q^*\Big(\frac{M^2 H}{n(\mathbb{D}_n^{\infty})^2 C_\Delta}\Big)\frac{M^2 H}{n}\Big]\\
\leq\ & \frac{2c}{\sigma^2 Q^{3/2}}\Big[\sigma^2 + q^*\Big(\frac{r(\frac{\sigma Q^{1/2}}{\mathbb{D}_n^{\infty}})^2}{C_\Delta}\Big) r(\frac{\sigma Q^{1/2}}{\mathbb{D}_n^{\infty}})^2(\mathbb{D}_n^{\infty})^2\Big]\\
\leq\ & \frac{2c}{\sigma^2 Q^{3/2}}\Big[\sigma^2 + q^*\Big(C_\Delta^{-1}C_\beta^{-2}\Big)\cdot\Big[q^*\Big(r(\frac{\sigma Q^{1/2}}{\mathbb{D}_n^{\infty}})\Big) r(\frac{\sigma Q^{1/2}}{\mathbb{D}_n^{\infty}})\Big]^2(\mathbb{D}_n^{\infty})^2\Big]\\
\leq\ & \frac{2c}{\sigma^2 Q^{3/2}}\Big[\sigma^2 + q^*\Big(C_\Delta^{-1}C_\beta^{-2}\Big)\sigma^2 Q\Big]|\\
\leq\ & \frac{2c}{Q^{1/2}}\big[1 + q^*\Big(C_\Delta^{-1}C_\beta^{-2}\Big)\big]
\end{aligned}
$$

for C_Δ defined in Lemma A.1.

By [14, Theorem 4.1] applied to $W_i(f) = \mathbb{E}[f(Z_i, \frac{i}{n})|Z_{i-1}]$,

$$
\begin{aligned}
&\mathbb{P}\Big(\max_{f\in\mathcal{F}}|\mathbb{G}_n^{(2)}(\varphi_{M_n}^{\wedge}(f))| > \frac{Q\sqrt{H}}{4}\Big)\\
\leq\ & \frac{8c}{Q\sqrt{H}}\cdot\Big[\sigma\sqrt{H} + q^*\Big(r(\frac{\sigma Q^{1/2}}{\mathbb{D}_n^{\infty}})\Big) r(\frac{\sigma Q^{1/2}}{\mathbb{D}_n^{\infty}})\mathbb{D}_n^{\infty}\Big]\\
\leq\ & \frac{8c}{Q\sqrt{H}}\big[\sigma\sqrt{H} + \sigma Q^{1/2}\sqrt{H}\big] \leq \frac{16c\sigma}{Q^{1/2}}.
\end{aligned}
$$

Inserting the upper bounds into (A.26), we obtain

$$
\mathbb{P}\Big(\max_{f\in\mathcal{F}}|\mathbb{G}_n(\varphi_{M_n}^{\wedge}(f))| > \frac{Q\sqrt{H}}{2}\Big) \leq \frac{8c}{Q^{1/4}} + \frac{2c}{Q^{1/2}}\big[1 + q^*\Big(C_\Delta^{-1}C_\beta^{-2}\Big)\big] + \frac{16c\sigma}{Q^{1/2}} \to 0
$$

for $Q \to \infty$. The second and third summand in (A.25) were already discussed in the proof of [14, Corollary 4.3] (equation (7.34) and (7.35) therein; note especially that we only need there that $\|\bar{F}(Z_i, \frac{i}{n})\|_{\nu_2} \le C_{\bar{F},n}$ instead of C_Δ which is part of the assumptions), and converge to 0 for $Q \to \infty$ under the given assumptions. ☐

The following Lemma A.5 is used to prove Theorem 4.3.

Lemma A.5 (Compatibility lemma 2)**.** *Let $\psi : (0,\infty) \to [1,\infty)$ be some function and $k \in \mathbb{N}$, $\delta > 0$. If $\mathcal{F}$ fulfills $|\mathcal{F}| \le k$ and Assumptions 2.2, 2.7, then there exists some universal constant $c > 0$ such that the following holds: If $\sup_{f\in\mathcal{F}} V_n(f) \le \delta$ and $\sup_{f\in\mathcal{F}} \|f\|_\infty \le m(n,\delta,k)$, then*

$$\mathbb{E}\max_{f\in\mathcal{F}} \big|\mathbb{G}_n^{(1)}(f)\big|\mathbb{1}_{\{R_n(f)\le 2\delta\psi(\delta)\}} \le 2c(1+\frac{\mathbb{D}_n^\infty}{\mathbb{D}_n})\cdot\psi(\delta)\delta\sqrt{H(k)}, \quad \text{(A.27)}$$

$$\mathbb{P}\Big(\sup_{f\in\mathcal{F}} R_n(f) > 2\delta\psi(\delta)\Big) \le \frac{2c(1+q^*\big(C_\Delta^{-1}C_\beta^{-2}\big)(\frac{\mathbb{D}_n^\infty}{\mathbb{D}_n})^2)}{\psi(\delta)^2}. \quad \text{(A.28)}$$

Proof of Lemma A.5. By Lemma A.4 and since $r(a) \le a$ (cf. [14, Lemma 7.5]),

$$\begin{aligned}\mathbb{E}\max_{f\in\mathcal{F}} \big|\mathbb{G}_n^{(1)}(f)\big|\mathbb{1}_{\{R_n(f)\le 2\delta\psi(\delta)\}} &\le c\Big\{2\psi(\delta)\delta\sqrt{H(k)} + \frac{m(n,\delta,k)H(k)}{\sqrt{n}}\Big\}\\ &\le 2c\cdot\big[\psi(\delta)\cdot\delta + \mathbb{D}_n^\infty r(\frac{\delta}{\mathbb{D}_n})\big]\sqrt{H(k)}\\ &\le 2c\cdot(1+\frac{\mathbb{D}_n^\infty}{\mathbb{D}_n})\cdot\psi(\delta)\delta\sqrt{H(k)},\end{aligned}$$

which shows (A.27).

For $\hat{a} = \arg\min_{j\in\mathbb{N}}\big\{\|f\|_{2,n}\cdot j + \mathbb{D}_n\beta(j)\big\}$ and since $\|f\|_{2,n} \le V_n(f) \le \delta$ we have with $r(\frac{\delta}{\mathbb{D}_n}) \ge \frac{\delta}{\mathbb{D}_n\hat{a}}$,

$$\frac{\|f\|_{2,n}^2}{\mathbb{D}_n^\infty r(\frac{\delta}{\mathbb{D}_n})} \le \frac{\mathbb{D}_n\hat{a}\|f\|_{2,n}^2}{\mathbb{D}_n^\infty\delta} \le \frac{\mathbb{D}_n V_n(f)\|f\|_{2,n}}{\mathbb{D}_n^\infty\delta} \le \frac{\mathbb{D}_n}{\mathbb{D}_n^\infty}\|f\|_{2,n}. \quad \text{(A.29)}$$

Therefore, $\|f\|_{2,n}^2 \le \mathbb{D}_n r(\frac{\delta}{\mathbb{D}_n})\|f\|_{2,n}$ and thus $\|f\|_{2,n} \le \mathbb{D}_n r(\frac{\delta}{\mathbb{D}_n})$. Note that due to $r(a) \le a$,

$$\mathbb{E}R_n(f)^2 = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[f(Z_i,\frac{i}{n})^2] \le \|f\|_{2,n}^2 \le (\mathbb{D}_n r(\frac{\delta}{\mathbb{D}_n}))^2 \le \delta^2. \quad \text{(A.30)}$$

Recall that $\beta_{norm}(q) = \frac{\beta(q)}{q}$. By Assumption 2.7, we have that for any $x_1, x_2 > 0$, $\tilde{q} = q^*(x_1)q^*(x_2)$ satisfies

$$\beta_{norm}(\tilde{q}) \le C_\beta\beta_{norm}(q^*(x_1))\beta_{norm}(q^*(x_2)) \le C_\beta x_1 x_2.$$

Thus, by definition of q^*,

$$q^*(C_\beta x_1 x_2) \le q^*(x_1)q^*(x_2). \quad \text{(A.31)}$$

We obtain that

$$q^*\Big(r(\frac{\delta}{\mathbb{D}_n})^2\frac{1}{C_\Delta}\Big) \le q^*\Big(r(\frac{\delta}{\mathbb{D}_n})\Big)^2 q^*\big(C_\Delta^{-1}C_\beta^{-2}\big). \tag{A.32}$$

By (A.30), Markov's inequality, Theorem 4.1 and (A.32),

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{f\in\mathcal{F}} R_n(f)^2 > 2\psi(\delta)^2\delta^2\Big)\\
\le\ &\mathbb{P}\Big(\sup_{f\in\mathcal{F}} |R_n(f)^2 - \mathbb{E}R_n(f)^2| > \psi(\delta)^2\delta^2\Big)\\
\le\ &\frac{2c}{\psi(\delta)^2\delta^2}\cdot\Big[\mathbb{D}_n r(\frac{\delta}{\mathbb{D}_n})\delta + q^*\Big(r(\frac{\delta}{\mathbb{D}_n})^2\frac{1}{C_\Delta}\Big)r(\frac{\delta}{\mathbb{D}_n})^2(\mathbb{D}_n^\infty)^2\Big]\\
\le\ &\frac{2c}{\psi(\delta)^2\delta^2}\cdot\Big[\delta^2 + \Big[q^*\Big(r(\frac{\delta}{\mathbb{D}_n})\Big)r(\frac{\delta}{\mathbb{D}_n})\Big]^2 q^*\big(C_\Delta^{-1}C_\beta^{-2}\big)(\mathbb{D}_n^\infty)^2\Big]\\
\le\ &\frac{2c}{\psi(\delta)^2\delta^2}\cdot\Big[\delta^2 + \delta^2 q^*\big(C_\Delta^{-1}C_\beta^{-2}\big)(\frac{\mathbb{D}_n^\infty}{\mathbb{D}_n})^2\Big]\\
\le\ &\frac{2c(1+q^*\big(C_\Delta^{-1}C_\beta^{-2}\big)(\frac{\mathbb{D}_n^\infty}{\mathbb{D}_n})^2)}{\psi(\delta)^2},
\end{aligned}$$

which shows (A.28). □

Proof of Theorem 4.3. In the following, we abbreviate $\mathbb{H}(\delta) = \mathbb{H}(\delta,\mathcal{F},V)$ and $\mathbb{N}(\delta) = \mathbb{N}(\delta,\mathcal{F},V)$. The proof follows the lines of [14, Theorem 4.4]. We present it here for completeness. Recall again that for $m > 0$, $\varphi_m^\wedge : \mathbb{R}\to\mathbb{R}$ and the corresponding "peaky" residual function $\varphi_m^\vee : \mathbb{R}\to\mathbb{R}$ via

$$\varphi_m^\wedge(x) := (x\vee(-m))\wedge m, \qquad \varphi_m^\vee(x) := x - \varphi_m^\wedge(x).$$

We choose $\delta_0 = \sigma$ and $\delta_j = 2^{-j}\delta_0$, and

$$m_j = \frac{1}{2}m(n,\delta_j,N_{j+1}),$$

as well as $M_n = \frac{1}{2}m_0$. We then use

$$\mathbb{E}\sup_{f\in\mathcal{F}}\Big|\mathbb{G}_n^{(1)}(f)\Big| \le \mathbb{E}\sup_{f\in\mathcal{F}(M_n)}\Big|\mathbb{G}_n^{(1)}(f)\Big| + \frac{1}{\sqrt{n}}\sum_{i=1}^n\mathbb{E}\big[F(Z_i)\mathbb{1}_{\{F(Z_i)>M_n\}}\big], \tag{A.33}$$

where $\mathcal{F}(M_n) := \{\varphi_{M_n}^\wedge(f) : f\in\mathcal{F}\}$.

We construct a nested sequence of partitions $(\mathcal{F}_{jk})_{k=1,\dots,N_j}$, $j\in\mathbb{N}$ of $\mathcal{F}(M_n)$ (where $N_j := \mathbb{N}(\delta_0)\cdot\ldots\cdot\mathbb{N}(\delta_j)$), and a sequence Δ_{jk} of measurable functions such that

$$\sup_{f,g\in\mathcal{F}_{jk}}|f-g| \le \Delta_{jk}, \qquad V(\Delta_{jk}) \le \delta_j.$$

In each $\mathcal{F}_{jk}$, we fix some $f_{jk}\in\mathcal{F}$, and define $\pi_j f := f_{j,\psi_j f}$ where $\psi_j f := \min\{i\in\{1,\dots,N_j\} : f\in\mathcal{F}_{ji}\}$, and put $\Delta_j f := \Delta_{j,\psi_j f}$, and

$$I(\sigma) := \int_0^\sigma \psi(\varepsilon)\sqrt{1\vee\mathbb{H}(\varepsilon,\mathcal{F},V)}d\varepsilon,$$

as well as

$$\tau := \min\Big\{j \geq 0 : \delta_j \leq \frac{I(\sigma)}{\sqrt{n}}\Big\} \vee 1. \tag{A.34}$$

For functions f, g with $|f| \leq g$, it holds that

$$\begin{aligned}
|\mathbb{G}_n^{(1)}(f)| &\leq |\mathbb{G}_n^{(1)}(g)| + 2\sqrt{n} \cdot \frac{1}{n}\sum_{i=1}^n \mathbb{E}[g(Z_i, \frac{i}{n})|Z_{i-1}] \\
&\leq |\mathbb{G}_n^{(1)}(g)| + 2|\mathbb{G}_n^{(2)}(g)| + 2\sqrt{n} \cdot \frac{1}{n}\sum_{i=1}^n \mathbb{E}[g(Z_i, \frac{i}{n})] \\
&\leq |\mathbb{G}_n^{(1)}(g)| + 2|\mathbb{G}_n^{(2)}(g)| + 2\sqrt{n}\|g\|_{1,n}.
\end{aligned}$$

Using a similar approach as in [14, Section 7.2, equations (7.8) and (7.9)] applied to $W_i(f) = f(Z_i, \frac{i}{n}) - \mathbb{E}[f(Z_i, \frac{i}{n})|Z_{i-1}]$, and the fact that $\|f - \pi_0 f\|_\infty \leq 2M_n \leq m_0$, we have the decomposition

$$\begin{aligned}
&\sup_{f\in\mathcal{F}} |\mathbb{G}_n^{(1)}(f)| \\
\leq\ & \sup_{f\in\mathcal{F}} |\mathbb{G}_n^{(1)}(\pi_0 f)| \\
&\quad + \sup_{f\in\mathcal{F}} |\mathbb{G}_n^{(1)}(\varphi^{\wedge}_{m_\tau}(f - \pi_\tau f))| + \sum_{j=0}^{\tau-1} \sup_{f\in\mathcal{F}} \Big|\mathbb{G}_n^{(1)}(\varphi^{\wedge}_{m_j - m_{j+1}}(\pi_{j+1} f - \pi_j f))\Big| \\
&\qquad + \sum_{j=0}^{\tau-1} \sup_{f\in\mathcal{F}} |\mathbb{G}_n^{(1)}(R(j))| \\
\leq\ & \sup_{f\in\mathcal{F}} |\mathbb{G}_n^{(1)}(\pi_0 f)| \\
&\quad + \Big\{ \sup_{f\in\mathcal{F}} |\mathbb{G}_n^{(1)}(\varphi^{\wedge}_{m_\tau}(\Delta_\tau f))| + 2\sup_{f\in\mathcal{F}} |\mathbb{G}_n^{(2)}(\varphi^{\wedge}_{m_\tau}(\Delta_\tau f))| \\
&\qquad\quad + 2\sqrt{n} \sup_{f\in\mathcal{F}} \|\Delta_\tau f\|_{1,n} \Big\} \\
&\quad + \sum_{j=0}^{\tau-1} \sup_{f\in\mathcal{F}} \Big|\mathbb{G}_n^{(1)}(\varphi^{\wedge}_{m_j - m_{j+1}}(\pi_{j+1} f - \pi_j f))\Big| \\
&\quad + \sum_{j=0}^{\tau-1} \Big\{ \sup_{f\in\mathcal{F}} \Big|\mathbb{G}_n^{(1)}(\min\big\{\big|\varphi^{\vee}_{m_{j+1}}(\Delta_{j+1} f)\big|, 2m_j\big\})\Big| \\
&\qquad\qquad + 2\sup_{f\in\mathcal{F}} \Big|\mathbb{G}_n^{(2)}(\min\big\{\big|\varphi^{\vee}_{m_{j+1}}(\Delta_{j+1} f)\big|, 2m_j\big\})\Big| \\
&\qquad\qquad + 2\sqrt{n} \sup_{f\in\mathcal{F}} \|\Delta_{j+1} f \mathbb{1}_{\{\Delta_{j+1} f > m_{j+1}\}}\|_{1,n} \Big\} \\
&\quad + \sum_{j=0}^{\tau-1} \Big\{ \sup_{f\in\mathcal{F}} \Big|\mathbb{G}_n^{(1)}(\min\big\{\big|\varphi^{\vee}_{m_j - m_{j+1}}(\Delta_j f)\big|, 2m_j\big\})\Big|
\end{aligned}$$

$$+2\sup_{f\in\mathcal{F}}\Big|\mathbb{G}_n^{(2)}(\min\big\{\big|\varphi_{m_j-m_{j+1}}^{\vee}(\Delta_j f)\big|,2m_j\big\})\Big|$$
$$+2\sqrt{n}\sup_{f\in\mathcal{F}}\|\Delta_j f\mathbb{1}_{\{\Delta_j f>m_j-m_{j+1}\}}\|_{1,n}\Big\} \tag{A.35}$$

We have for $f\in\mathcal{F}(M_n)$,

$$\begin{aligned}
\pi_0 f &= \varphi_{2M_n}^{\wedge}(\pi_0 f),\\
\varphi_{m_\tau}^{\wedge}(\Delta_\tau f) &\le \min\{\Delta_\tau f, 2m_\tau\},\\
\varphi_{m_j-m_{j-1}}^{\wedge}(\pi_{j+1}f-\pi_j f) &\le \min\{\Delta_j f, 2m_j\},\\
\min\{\varphi_{m_{j+1}}^{\vee}(\Delta_{j+1}f),2m_j\} &\le \min\{\Delta_j f, 2m_j\},\\
\min\{\varphi_{m_j-m_{j+1}}^{\vee}(\Delta_j f),2m_j\} &\le \min\{\Delta_j f, 2m_j\}.
\end{aligned} \tag{A.36}$$

We therefore define the event

$$\begin{aligned}
\Omega_n \;:=\; &\{\sup_{f\in\mathcal{F}(M_n)} R_n(\varphi_{2M_n}^{\wedge}(\pi_0 f))\le 2\sigma\psi(\sigma)\}\\
&\cap\bigcap_{j=1}^{\tau}\big\{\sup_{f\in\mathcal{F}(M_n)} R_n(\min\{\Delta_j f,2m_j\})\le 2\delta_j\psi(\delta_j)\big\}.
\end{aligned}$$

From (A.35) and (A.36), we obtain

$$\begin{aligned}
&\sup_{f\in\mathcal{F}(M_n)}|\mathbb{G}_n^{(1)}(f)|\mathbb{1}_{\Omega_n}\\
\le\;&\sup_{f\in\mathcal{F}(M_n)}|\mathbb{G}_n^{(1)}(\pi_0 f)|\mathbb{1}_{\{\sup_{f\in\mathcal{F}(M_n)}R_n(\pi_0 f)\le 2\sigma\psi(\sigma)\}}\\
&+\Big\{\sup_{f\in\mathcal{F}}|\mathbb{G}_n^{(1)}(\varphi_{m_\tau}^{\wedge}(\Delta_\tau f))|\\
&\qquad\times\mathbb{1}_{\{\sup_{f\in\mathcal{F}(M_n)}R_n(\min\{\Delta_\tau f,2m_\tau\})\le 2\delta_\tau\psi(\delta_\tau)\}}+2R_2\Big\}\\
&+\sum_{j=0}^{\tau-1}\sup_{f\in\mathcal{F}}\Big|\mathbb{G}_n^{(1)}(\varphi_{m_j-m_{j+1}}^{\wedge}(\pi_{j+1}f-\pi_j f))\Big|\\
&\qquad\times\mathbb{1}_{\{\sup_{f\in\mathcal{F}(M_n)}R_n(\min\{\Delta_j f,2m_j\})\le 2\delta_j\psi(\delta_j)\}}\\
&+\sum_{j=0}^{\tau-1}\sup_{f\in\mathcal{F}}\Big|\mathbb{G}_n^{(1)}(\min\big\{\big|\varphi_{m_{j+1}}^{\vee}(\Delta_{j+1}f)\big|,2m_j\big\})\Big|\\
&\qquad\times\mathbb{1}_{\{\sup_{f\in\mathcal{F}(M_n)}R_n(\min\{\Delta_j f,2m_j\})\le 2\delta_j\psi(\delta_j)\}}+2R_4\\
&+\sum_{j=0}^{\tau-1}\sup_{f\in\mathcal{F}}\Big|\mathbb{G}_n^{(1)}(\min\big\{\big|\varphi_{m_j-m_{j+1}}^{\vee}(\Delta_j f)\big|,2m_j\big\})\Big|\\
&\qquad\times\mathbb{1}_{\{\sup_{f\in\mathcal{F}(M_n)}R_n(\min\{\Delta_j f,2m_j\})\le 2\delta_j\psi(\delta_j)\}}+2R_5\\
=:\;&\tilde{R}_1+\{\tilde{R}_2+2R_2\}+\tilde{R}_3+\{\tilde{R}_4+2R_4\}+\{\tilde{R}_5+2R_5\}.
\end{aligned} \tag{A.37}$$

We now discuss the terms $\tilde{R}_i$, $i = 1, ..., 5$ separately. The terms R_i, $i \in \{2, 4, 5\}$ can be discussed similarly to the proof found in [14, Theorem 4.4]. Put

$$\tilde{C}_n := 2c(1 + \frac{\mathbb{D}_n^\infty}{\mathbb{D}_n}),$$

where c is a resulting constant from the bound in [14, Theorem 4.1 or Lemma 7.2].

- Since $|\{\pi_0 f : f \in \mathcal{F}(M_n)\}| \le \mathbb{N}(\delta_0)$, $\|\pi_0 f\|_\infty \le M_n \le m(n, \delta_0, \mathbb{N}(\delta_1))$, we have by Lemma A.5:

$$\begin{aligned}\mathbb{E}\tilde{R}_1 &= \mathbb{E} \sup_{f \in \mathcal{F}(M_n)} |\mathbb{G}_n^{(1)}(\pi_0 f)| \mathbb{1}_{\{\sup_{f\in\mathcal{F}(M_n)} R_n(\pi_0 f) \le 2\delta_0 \psi(\delta_0)\}} \\ &\le \tilde{C}_n \psi(\delta_0)\delta_0 \sqrt{1 \vee \log \mathbb{N}(\delta_1)}.\end{aligned}$$

- It holds that $|\{\varphi^\wedge_{m_\tau}(\Delta_\tau f) : f \in \mathcal{F}(M_n)\}| \le N_\tau$. If $g := \varphi^\wedge_{m_\tau}(\Delta_\tau f)$, then $\|g\|_\infty \le m_\tau \le m(n, \delta_\tau, N_{\tau+1})$. We conclude by Lemma A.5:

$$\begin{aligned}\mathbb{E}\tilde{R}_2 &\le \mathbb{E} \sup_{f \in \mathcal{F}} |\mathbb{G}_n^{(1)}(\varphi^\wedge_{m_\tau}(\Delta_\tau f))| \\ &\qquad \times \mathbb{1}_{\{\sup_{f\in\mathcal{F}(M_n)} R_n(\min\{\Delta_\tau f, 2m_\tau\}) \le 2\delta_\tau \psi(\delta_\tau)\}} \\ &\le \tilde{C}_n \psi(\delta_\tau)\delta_\tau \cdot \sqrt{1 \vee \log N_{\tau+1}}.\end{aligned}$$

- Since the partitions are nested, it holds that $|\{\varphi^\wedge_{m_j - m_{j+1}}(\pi_{j+1} f - \pi_j f) : f \in \mathcal{F}(M_n)\}| \le N_{j+1}$. If $g := \varphi^\wedge_{m_j - m_{j+1}}(\pi_{j+1} f - \pi_j f)$, we have $\|g\|_\infty \le m_j - m_{j+1} \le m_j \le m(n, \delta_j, N_{j+1})$. We conclude by Lemma A.5:

$$\begin{aligned}\mathbb{E}\tilde{R}_3 &\le \sum_{j=0}^{\tau-1} \mathbb{E} \sup_{f \in \mathcal{F}} \Big|\mathbb{G}_n^{(1)}(\varphi^\wedge_{m_j - m_{j+1}}(\pi_{j+1} f - \pi_j f))\Big| \\ &\qquad \times \mathbb{1}_{\{\sup_{f\in\mathcal{F}(M_n)} R_n(\min\{\Delta_j f, 2m_j\}) \le 2\delta_j \psi(\delta_j)\}} \\ &\le \tilde{C}_n \sum_{j=0}^{\tau-1} \psi(\delta_j)\delta_j \sqrt{1 \vee \log N_{j+1}}.\end{aligned}$$

- It holds that $|\{\min\{\varphi^\vee_{m_{j+1}}(\Delta_{j+1} f), 2m_j\} : f \in \mathcal{F}(M_n)\}| \le N_{j+1}$. If $g := \min\{\varphi^\vee_{m_{j+1}}(\Delta_{j+1} f), 2m_j\}$, we have $\|g\|_\infty \le 2m_j = m(n, \delta_j, N_{j+1})$. We conclude by Lemma A.5:

$$\begin{aligned}\mathbb{E}\tilde{R}_4 &\le \sum_{j=0}^{\tau-1} \mathbb{E} \sup_{f \in \mathcal{F}} \Big|\mathbb{G}_n^{(1)}(\min\big\{\big|\varphi^\vee_{m_{j+1}}(\Delta_{j+1} f)\big|, 2m_j\big\})\Big| \\ &\qquad \times \mathbb{1}_{\{\sup_{f\in\mathcal{F}(M_n)} R_n(\min\{\Delta_j f, 2m_j\}) \le 2\delta_j \psi(\delta_j)\}} \\ &\le \tilde{C}_n \sum_{j=0}^{\tau-1} \psi(\delta_j)\delta_j \sqrt{1 \vee \log N_{j+1}}.\end{aligned}$$

- It holds that $|\{\min\{\varphi^{\vee}_{m_j-m_{j+1}}(\Delta_j f), 2m_j\} : f \in \mathcal{F}(M_n)\}| \le N_{j+1}$. If $g := \min\{\varphi^{\vee}_{m_j-m_{j+1}}(\Delta_j f), 2m_j\}$, we have $\|g\|_\infty \le 2m_j = m(n,\delta_j,N_{j+1})$. We conclude by Lemma A.5 that:

$$\begin{aligned}\mathbb{E}\tilde{R}_5 &\le \sum_{j=0}^{\tau-1}\mathbb{E}\sup_{f\in\mathcal{F}}\Big|\mathbb{G}_n^{(1)}(\min\big\{\big|\varphi^{\vee}_{m_j-m_{j+1}}(\Delta_j f)\big|, 2m_j\big\})\Big| \\ &\qquad\qquad \times \mathbb{1}_{\{\sup_{f\in\mathcal{F}(M_n)} R_n(\min\{\Delta_j f, 2m_j\})\le 2\delta_j\psi(\delta_j)\}} \\ &\le \tilde{C}_n\sum_{j=0}^{\tau-1}\psi(\delta_j)\delta_j\cdot\sqrt{1\vee\log N_{j+1}}.\end{aligned}$$

Inserting the bounds for $\mathbb{E}\tilde{R}_i$, $i = 1,...,5$ and the bounds for R_i, $i \in \{2,4,5\}$ from the proof of [14, Theorem 4.4] into (A.37), we obtain that with some universal constant $\tilde{c} > 0$,

$$\mathbb{E}\sup_{f\in\mathcal{F}(M_n)}\Big|\mathbb{G}_n^{(1)}(f)\Big|\mathbb{1}_{\Omega_n} \le \tilde{c}(1+\frac{\mathbb{D}_n^\infty}{\mathbb{D}_n}+\frac{\mathbb{D}_n}{\mathbb{D}_n^\infty})\Big[\sum_{j=0}^{\tau+1}\psi(\delta_j)\delta_j\sqrt{1\vee\log N_{j+1}}+I(\sigma)\Big]. \tag{A.38}$$

Note that

$$\sum_{j=k}^{\infty}\delta_j\psi(\delta_j) \le 2\sum_{j=k}^{\infty}\int_{\delta_{j+1}}^{\delta_j}\psi(\delta_j)dx \le 2\int_0^{\delta_k}\psi(x)dx.$$

By partial integration, it is easy to see that there exists some universal constant $c_\psi > 0$ such that

$$\big|\int_0^{\delta_k}\psi(x)dx\big| \le c_\psi\delta_k\psi(\delta_k), \tag{A.39}$$

thus

$$\sum_{j=k}^{\infty}\delta_j\psi(\delta_j) \le 2c_\psi\delta_k\psi(\delta_k). \tag{A.40}$$

Using (A.40), we can argue as in the proof [14, Theorem 4.4] (see (7.44), (7.45) and (7.46) therein) that there exists some universal constant $\tilde{c}_2 > 0$ such that

$$\sum_{j=0}^{\infty}\psi(\delta_j)\delta_j\sqrt{1\vee\log N_{j+1}} \le \tilde{c}_2 I(\sigma).$$

Insertion of the results into (A.38) yields

$$\mathbb{E}\sup_{f\in\mathcal{F}(M_n)}\big|\mathbb{G}_n^{(1)}(f)\big|\mathbb{1}_{\Omega_n} \le \tilde{c}\cdot(3\tilde{c}_2+1)(1+\frac{\mathbb{D}_n^\infty}{\mathbb{D}_n}+\frac{\mathbb{D}_n}{\mathbb{D}_n^\infty})I(\sigma). \tag{A.41}$$

<u>Discussion of the event Ω_n:</u> We have

$$\mathbb{P}(\Omega_n^c) \le \mathbb{P}\Big(\sup_{f\in\mathcal{F}(M_n)} R_n(\varphi^{\wedge}_{2M_n}(\pi_0 f)) > 2\psi(\sigma)\sigma\Big)$$

$$
\begin{aligned}
&\quad + \sum_{j=1}^{\tau+1} \mathbb{P}\Big(\sup_{f\in\mathcal{F}(M_n)} R_n(\min\{\Delta_j f, 2m_j\}) > 2\psi(\delta_j)\delta_j \Big) \\
=: &\quad R_1^\circ + R_2^\circ. \qquad (A.42)
\end{aligned}
$$

We now discuss R_i°, $i = 1, 2$. Put

$$
C_n^\circ := 2c\Big\{1 + q^*\big(C_\Delta^{-1}C_\beta^{-2}\big)\big(\frac{\mathbb{D}_n^\infty}{\mathbb{D}_n}\big)^2\Big\},
$$

where c is from Lemma A.5.

- Since $|\{\varphi_{2M_n}^\wedge(\pi_0 f) : f \in \mathcal{F}(M_n)\}| \le \mathbb{N}(\delta_0) = \mathbb{N}(\sigma)$, $\|\varphi_{2M_n}^\wedge(\pi_0 f)\|_\infty \le 2M_n \le m(n,\sigma,\mathbb{N}(\sigma))$ and $V(\varphi_{2M_n}^\wedge(\pi_0 f)) \le V(\pi_0 f) \le \sigma$, we have by Lemma A.5:
$$
R_1^\circ \le \frac{C_n^\circ}{\psi(\sigma)^2}.
$$
- It holds that $|\{\min\{\Delta_j f, 2m_j\} : f \in \mathcal{F}(M_n)\}| \le N_{j+1}$. We have $\|\min\{\Delta_j f, 2m_j\}\|_\infty \le 2m_j = m(n,\delta_j,N_{j+1})$ and $V(\min\{\Delta_j f, 2m_j\}) \le V(\Delta_j f) \le \delta_j$. We conclude by Lemma A.5 that:
$$
R_3^\circ \le C_n^\circ \sum_{j=0}^{\tau+1} \frac{1}{\psi(\delta_j)^2}.
$$

Inserting the bounds for R_i°, $i = 1, 2$, into (A.42) yields

$$
\mathbb{P}(\Omega_n^c) \le 2C_n^\circ \sum_{j=0}^{\infty} \frac{1}{\psi(\delta_j)^2}. \qquad (A.43)
$$

We now have

$$
\sum_{j=0}^{\infty} \frac{1}{\psi(\delta_j)^2} \le 2\int_0^\sigma \frac{1}{\varepsilon\psi(\varepsilon)^2} d\varepsilon = \frac{2}{\log(\log(\sigma))}.
$$

We conclude that for each $\eta > 0$,

$$
\begin{aligned}
\mathbb{P}\Big(\sup_{f\in\mathcal{F}} |\mathbb{G}_n^{(1)}(f)| > \eta\Big) &\le \mathbb{P}\Big(\sup_{f\in\mathcal{F}} |\mathbb{G}_n^{(1)}(f)| > \eta, \Omega_n\Big) + \mathbb{P}(\Omega_n^c) \\
&\le \frac{1}{\eta}\mathbb{E}\sup_{f\in\mathcal{F}} |\mathbb{G}_n^{(1)}(f)|\mathbb{1}_{\Omega_n} + \mathbb{P}(\Omega_n^c).
\end{aligned}
$$

Insertion of (A.33), (A.41) and (A.43) gives the result. □

Proof of Corollary 4.5. We will follow the proof of [14, Corollary 4.5]. Define $\tilde{\mathcal{F}}_: = \{f - g : f, g \in \mathcal{F}\}$. We obtain

$$
\mathbb{P}\Big(\sup_{V(f-g)\le\sigma,\, f,g\in\mathcal{F}} |\mathbb{G}_n(f) - \mathbb{G}_n(g)| \ge \eta\Big)
$$

$$\leq \mathbb{P}\Big(\sup_{V(\tilde{f})\leq\sigma,\,\tilde{f}\in\tilde{\mathcal{F}}}|\mathbb{G}_n^{(1)}(\tilde{f})| \geq \frac{\eta}{2}\Big) + \mathbb{P}\Big(\sup_{V(\tilde{f})\leq\sigma,\,\tilde{f}\in\tilde{\mathcal{F}}}|\mathbb{G}_n^{(2)}(\tilde{f})| \geq \frac{\eta}{2}\Big) \quad \text{(A.44)}$$

Now let $F(z,u) := 2D_n^\infty(u)\cdot\bar{F}(z,u)$, where $\bar{F}$ is from Assumption 2.3. Then obviously, F is an envelope function of $\tilde{\mathcal{F}}$.

We now discuss the second summand on the right hand side in (A.44). By Markov's inequality and [14, Theorem 4.4] applied to $W_i(f) = \mathbb{E}[f(Z_i,\frac{i}{n})|Z_{i-1}]$, we obtain as in the proof of [14, Corollary 4.5] that

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{V(\tilde{f})\leq\sigma,\,\tilde{f}\in\tilde{\mathcal{F}}}|\mathbb{G}_n^{(2)}(\tilde{f})| \geq \frac{\eta}{2}\Big)\\
\leq\;& \frac{\tilde{c}}{(\eta/2)}\Big[2\sqrt{2}\Big(1+\frac{\mathbb{D}_n^\infty}{\mathbb{D}_n}+\frac{\mathbb{D}_n}{\mathbb{D}_n^\infty}\Big)\int_0^{\sigma/2}\sqrt{1\vee\mathbb{H}(u,\mathcal{F},V)}du\\
&\qquad+\frac{4\sqrt{1\vee\mathbb{H}(\frac{\sigma}{2})}}{r(\frac{\sigma}{\mathbb{D}_n})}\big\|F^2\mathbb{1}_{\{F>\frac{1}{4}n^{1/2}\frac{r(\sigma)}{\sqrt{1\vee\mathbb{H}(\frac{\sigma}{2})}}\}}\big\|_{1,n}\Big]. \qquad \text{(A.45)}
\end{aligned}$$

The first summand in (A.45) converges to 0 for $\sigma\to 0$ (uniformly in n) since

$$\sup_{n\in\mathbb{N}}\int_0^{\sigma/2}\sqrt{1\vee\mathbb{H}(u,\mathcal{F},V)}du \leq \sup_{n\in\mathbb{N}}\int_0^{\sigma}\psi(\varepsilon)\sqrt{1\vee\mathbb{H}(\varepsilon,\mathcal{F},V)}d\varepsilon < \infty.$$

We now discuss the second summand in (A.45). The continuity conditions from Assumption 2.3 on $\bar{F}$ yield as in the proof of Lemma A.6(ii) that for all $u,u_1,u_2,v_1,v_2\in[0,1]$,

$$\|\bar{F}(Z_i,u)-\bar{F}(\tilde{Z}_i(\frac{i}{n}),u)\|_2 \leq C_{cont}\cdot n^{-\alpha s/2}, \quad \text{(A.46)}$$

$$\|\bar{F}(Z_i(v_1),u_1)-\bar{F}(\tilde{Z}_i(v_2),v_2)\|_2 \leq C_{cont}\cdot\big(|v_1-v_2|^{\alpha s/2}+|u_1-u_2|^{\alpha s}\big). \quad \text{(A.47)}$$

In the same manner of [14, Corollary 4.5], we now obtain with (A.46) and (A.47) that

$$\big\|F^2\mathbb{1}_{\{F>\frac{1}{4}n^{1/2}\frac{r(\sigma)}{\sqrt{1\vee\mathbb{H}(\frac{\sigma}{2})}}\}}\big\|_{1,n} \to 0 \qquad \text{(A.48)}$$

for $n\to\infty$ (this is obvious if Z_i is stationary, i.e. the first part of Assumption 2.3 is fulfilled), which shows that (A.45) converges to 0 for $\sigma\to 0$, $n\to\infty$.

We now consider the first term in (A.44). By Theorem 4.3, we have with some universal constant $c>0$ that

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{V(\tilde{f})\leq\sigma,\,\tilde{f}\in\tilde{\mathcal{F}}}|\mathbb{G}_n^{(1)}(\tilde{f})| \geq \frac{\eta}{2}\Big)\\
\leq\;& \frac{2}{\eta}\Big[c\Big(1+\frac{\mathbb{D}_n^\infty}{\mathbb{D}_n}+\frac{\mathbb{D}_n}{\mathbb{D}_n^\infty}\Big)\cdot\int_0^{\sigma}\psi(\varepsilon)\sqrt{1\vee\mathbb{H}\big(\varepsilon,\tilde{\mathcal{F}},V\big)}\,\mathrm{d}\varepsilon\\
&\qquad+\frac{4\sqrt{1\vee\mathbb{H}(\frac{\sigma}{2})}}{r(\frac{\sigma}{\mathbb{D}_n})}\big\|F^2\mathbb{1}_{\{F>\frac{1}{4}m(n,\sigma,\mathbb{N}(\frac{\sigma}{2}))\}}\big\|_1\Big]\\
&\qquad+c\Big(1+q^*\big(C_\Delta^{-1}C_\beta^{-2}\big)\Big(\frac{\mathbb{D}_n^\infty}{\mathbb{D}_n}\Big)^2\Big)\int_0^{\sigma}\frac{1}{\varepsilon\psi(\varepsilon)^2}d\varepsilon. \qquad \text{(A.49)}
\end{aligned}$$

For the first summand in (A.49),

$$\begin{aligned}&\int_0^{\sigma} \psi(\varepsilon)\sqrt{1 \vee \mathbb{H}(\varepsilon, \tilde{\mathcal{F}}, V)}d\varepsilon \\ \leq \quad & 2\sqrt{2}\int_0^{\sigma/2} \psi(2\varepsilon)\sqrt{1 \vee \mathbb{H}(\varepsilon, \mathcal{F}, V)}d\varepsilon \leq 2\sqrt{2}\int_0^{\sigma/2} \psi(\varepsilon)\sqrt{1 \vee \mathbb{H}(\varepsilon, \mathcal{F}, V)}d\varepsilon.\end{aligned}$$

Note that it is easily seen that $\mathbb{N}(\varepsilon, \tilde{\mathcal{F}}, V) \leq \mathbb{N}(\frac{\varepsilon}{2}, \mathcal{F}, V)^2$ (cf. [17], Theorem 19.5), thus

$$\mathbb{H}(\varepsilon, \tilde{\mathcal{F}}, V) \leq 2\mathbb{H}(\frac{\varepsilon}{2}, \mathcal{F}, V). \tag{A.50}$$

Together with (4.7) and the uniform boundedness of $\mathbb{D}_n, \mathbb{D}_n^{\infty}$, we obtain that the first summand in (A.49) converges to 0 for $\sigma \to 0$ (uniformly in n).

The third summand in (A.49) converges to 0 for $\sigma \to 0$ (uniformly in n) since $\int_0^{\infty} \varepsilon\psi(\varepsilon)^2 d\varepsilon < \infty$ and by the uniform boundedness of $\mathbb{D}_n, \mathbb{D}_n^{\infty}$.

The second summand in (A.49) converges to 0 for $n \to \infty$ by (A.48). □

A.3. Proofs of Section 3

Lemma A.6. *Let $\mathcal{F}$ satisfy Assumptions 2.5, 2.4. Suppose that Assumptions 2.2, 2.3 hold. Then there exist constants $C_{cont} > 0, C_{\bar{f}} > 0$ such that for any $f \in \mathcal{F}$,*

(i) for any $j \geq 1$,

$$\begin{aligned}\|P_{i-j}f(Z_i, u)\|_2 &\leq D_{f,n}(u)\Delta(j), \\ \sup_{i=1,\ldots,n} \|f(Z_i, u)\|_2 &\leq C_{\Delta} \cdot D_{f,n}(u),\end{aligned}$$

$$\sup_{i,u} \|\bar{f}(Z_i, u)\|_2 \leq C_{\bar{f}}, \qquad \sup_{v,u} \|\bar{f}(\tilde{Z}_0(v), u)\|_2 \leq C_{\bar{f}}.$$

(ii) with $x = \frac{1}{2}$,

$$\|\bar{f}(Z_i, u) - \bar{f}(\tilde{Z}_i(\frac{i}{n}), u)\|_2 \leq C_{cont} \cdot n^{-\varsigma s x}, \tag{A.51}$$

$$\begin{aligned}&\|\bar{f}(\tilde{Z}_i(v_1), u_1) - \bar{f}(\tilde{Z}_i(v_2), u_2)\|_2 \\ \leq \quad & C_{cont} \cdot \big(|v_1 - v_2|^{\varsigma s x} + |u_1 - u_2|^{\varsigma s}\big).\end{aligned} \tag{A.52}$$

Proof of Lemma A.6. (i) If Assumption 2.2 is satisfied, we have by Lemma A.1 that

$$\begin{aligned}\|P_{i-j}f(Z_i, u)\|_2 &= \|P_{i-j}\mathbb{E}[f(Z_i, u)|\mathcal{A}_{i-1}]\|_2 \\ &\leq \|\mathbb{E}[f(Z_i, u)|\mathcal{A}_{i-1}] - \mathbb{E}[f(Z_i, u)|\mathcal{A}_{i-1}]^{*(i-j)}\|_2 \\ &\leq D_{f,n}(u)\Delta(j).\end{aligned}$$

The second assertion follows from Lemma A.1.

(ii) Let $\bar{C}_R := \sup_{v,u}\|\bar{R}(\tilde{Z}_0(v),u)\|_2$ and $C_R := \max\{\sup_{i,u}\|R(Z_i,u)\|_2, \sup_{u,v}\|R(\tilde{Z}_0(v),u)\|_2\}$. We first use Assumption 2.4 and Hölder's inequality to obtain

$$\begin{aligned} & \|\bar{f}(\tilde{Z}_i(v),u_1) - \bar{f}(\tilde{Z}_i(v),u_2)\|_2 \quad (A.53)\\ \leq\ & |u_1-u_2|^{\varsigma}\cdot\big(\|\bar{R}(\tilde{Z}_i(v),u_1)\|_2 + \|R(\tilde{Z}_i(v),u_2)\|_2\big)\\ \leq\ & 2\bar{C}_R|u_1-u_2|^{\varsigma}. \quad (A.54)\end{aligned}$$

Assume w.l.o.g. that

$$\sup_{u,v}\frac{1}{c^s}\mathbb{E}\Big[\sup_{|a|_{L_{\mathcal{F}},s}\leq c}\big|\bar{f}(\tilde{Z}_0(v),u) - \bar{f}(\tilde{Z}_0(v)+a,u)\big|^2\Big] \leq C_R.$$

(which is obvious if Z_i is stationary, i.e. the first part of Assumption 2.3 is fulfilled; in this case $Z_i = \tilde{Z}_i(v)$ for all v). Let $c_n > 0$ be some sequence. Let $C_{\bar{f}} := \max\{\sup_{i,u}\|f(Z_i,u)\|_{2\bar{p}}, \sup_{u,v}\|f(\tilde{Z}_0(v),u)\|_{2\bar{p}}\}$. Then we have by Jensen's inequality,

$$\begin{aligned} & \big\|\bar{f}(Z_i,u) - \bar{f}(\tilde{Z}_i(v),u)\big\|_2\\ \leq\ & \mathbb{E}\Big[\big|\bar{f}(Z_i,u) - \bar{f}(\tilde{Z}_i(v),u)\big|^2\mathbb{1}_{\{|Z_i-\tilde{Z}_i(v)|_{L_{\mathcal{F}},s}\leq c_n\}}\Big]^{1/2}\\ & \quad +\mathbb{E}\Big[(\bar{f}(Z_i,u) - \bar{f}(\tilde{Z}_i(v),u)^2\mathbb{1}_{\{|Z_i-\tilde{Z}_i(v)|_{L_{\mathcal{F}},s}> c_n\}}\Big]^{1/2}\\ \leq\ & \mathbb{E}\Big[\sup_{|a|_{L_{\mathcal{F}},s}\leq c_n}\big|\bar{f}(\tilde{Z}_i(v),u) - \bar{f}(\tilde{Z}_i(v)+a,u)\big|^2\Big]^{1/2}\\ & \quad +\big\{\big\|\bar{f}(Z_i,u)\big\|_{2\bar{p}} + \bar{f}(\tilde{Z}_i(v),u)\big\|_{2\bar{p}}\big\}\mathbb{P}(|Z_i-\tilde{Z}_i(v)|_{L_{\mathcal{F}},s} > c_n)^{\frac{\bar{p}-1}{2\bar{p}}}\\ \leq\ & C_R c_n^s + 2C_{\bar{f}}\Big(\frac{\||Z_i-\tilde{Z}_i(v)|_{L_{\mathcal{F}},s}\|_{\frac{2\bar{p}s}{\bar{p}-1}}}{c_n}\Big)^s\\ \leq\ & C_R c_n^s + 2C_{\bar{f}}C_X(|L_{\mathcal{F}}|_1 + \sum_{j=0}^{\infty}L_{\mathcal{F},j}j^{\varsigma s})\cdot\frac{\{|v-\frac{i}{n}|^{\varsigma s} + n^{-\varsigma s}\}}{c_n^s}.\end{aligned}$$

We obtain with $c_{cont} := C_R + 2C_{\bar{f}}C_X(|L_{\mathcal{F}}|_1 + \sum_{j=0}^{\infty}L_{\mathcal{F},j}j^{\varsigma s})$ that

$$\|\bar{f}(Z_i,u) - \bar{f}(\tilde{Z}_i(v),u)\|_2 \leq c_{cont}\cdot\Big[c_n^s + \frac{|v-\frac{i}{n}|^{\varsigma s} + n^{-\varsigma s}}{c_n^s}\Big]. \quad (A.55)$$

Furthermore, as above, for any $c > 0$,

$$\begin{aligned} & \|f(\tilde{Z}_i(v_1),u) - f(\tilde{Z}_i(v_2),u)\|_2\\ \leq\ & C_R c^s + 2C_{\bar{f}}\Big(\frac{\||\tilde{Z}_0(v_1)-\tilde{Z}_0(v_2)|^s_{L_{\mathcal{F}},s}\|_{\frac{2\bar{p}}{\bar{p}-1}}}{c}\Big)^s\\ \leq\ & C_R c^s + 2C_{\bar{f}}C_X|L_{\mathcal{F}}|_1\cdot\frac{|v_1-v_2|^{\varsigma s}}{c^s}. \quad (A.56)\end{aligned}$$

From (A.55), we obtain the first assertion with $v = \frac{i}{n}$. The second assertion follows from (A.56) and (A.54). □

A.4. Details of Section 2.6

We first show that the supremum over $x \in \mathbb{R}$, $v \in [0,1]$ can be approximated by a supremum over grids $x \in \mathcal{X}_n$, $v \in V_n$.

For some $Q > 0$, put $c_n = Qn^{\frac{1}{2s}}$. Define the event $A_n = \{\sup_{i=1,\dots,n} |X_i| \le c_n\}$. Then by Markov's inequality,

$$\mathbb{P}(A_n^c) \le n \cdot \frac{\|X_i\|_{2s}^{2s}}{Q^{2s} c_n^{2s}} \le \frac{C_X^{2s} n}{c_n^{2s}} \tag{A.57}$$

is arbitrarily small for Q large enough.

Put $\hat{g}^{\circ}_{n,h_n}(x,v) := \frac{1}{n}\sum_{i=1}^n K_{h_{1,n}}(i/n - v) K_{h_{2,n}}(X_i - x) \mathbb{1}_{\{|X_i| \le c_n\}}$. Then

$$\text{on } A_n, \qquad \hat{g}^{\circ}_{n,h_n}(\cdot) = \hat{g}_{n,h_n}(\cdot). \tag{A.58}$$

Furthermore,

$$\begin{aligned} & \sqrt{nh_{1,n}h_{2,n}}\big|\mathbb{E}\hat{g}_{n,h_n}(x,v) - \mathbb{E}\hat{g}^{\circ}_{n,h_n}(x,v)\big| \\ \le\ & \frac{\sqrt{nh_{1,n}h_{2,n}}|K|_\infty}{nh_{1,n}} \sum_{i=1}^n \mathbb{E}[K_{h_{2,n}}(X_i - x)\mathbb{1}_{\{|X_i| > c_n\}}] \\ \le\ & \sqrt{nh_{1,n}h_{2,n}}(h_{1,n}h_{2,n})^{-1}|K|_\infty c_n^{-2s} \sup_i \mathbb{E}[K(\frac{X_i - x}{h_{2,n}})|X_i|^{2s}] \\ \le\ & Q^{-2s}(nh_{1,n}h_{2,n})^{-1/2}|K|_\infty^2 C_X^{2s} = o(1). \end{aligned} \tag{A.59}$$

For $|x| > 2c_n$, we have $K_{h_{2,n}}(X_i - x)\mathbb{1}_{\{|X_i| \le c_n\}} \le h_n^{-1}(\frac{c_n}{h_n})^{-p_K} = h_n^{p_K - 1} c_n^{-p_K}$ and thus

$$\begin{aligned} \sqrt{nh_n}|\hat{g}^{\circ}_{n,h_n}(x,v) - \mathbb{E}\hat{g}^{\circ}_{n,h_n}(x,v)| &\le \frac{2|K|_\infty C_K}{h_{1,n}^{1/2}}(nh_{2,n})^{1/2} h_{2,n}^{p_K - 1} c_n^{-p_K} \\ &\le \frac{h_{2,n}^{p_K}}{Q^{p_K}(nh_{1,n}h_{2,n})^{1/2}} = o(1). \end{aligned} \tag{A.60}$$

By (A.58), (A.59) and (A.60), we have on A_n,

$$\begin{aligned} & \sqrt{nh_{1,n}h_{2,n}} \sup_{x\in\mathbb{R}, v\in[0,1]} |\hat{g}_{n,h_n}(x,v) - \mathbb{E}\hat{g}_{n,h_n}(x,v)| \\ =\ & \sqrt{nh_{1,n}h_{2,n}} \sup_{x\in\mathbb{R}, v\in[0,1]} |\hat{g}^{\circ}_{n,h_n}(x,v) - \mathbb{E}\hat{g}^{\circ}_{n,h_n}(x,v)| + o_p(1) \\ =\ & \sqrt{nh_{1,n}h_{2,n}} \sup_{|x|\le 2c_n, v\in[0,1]} |\hat{g}^{\circ}_{n,h_n}(x,v) - \mathbb{E}\hat{g}^{\circ}_{n,h_n}(x,v)| + o_p(1) \\ =\ & \sqrt{nh_{1,n}h_{2,n}} \sup_{|x|\le 2c_n, v\in[0,1]} |\hat{g}_{n,h_n}(x,v) - \mathbb{E}\hat{g}_{n,h_n}(x,v)| + o_p(1). \end{aligned} \tag{A.61}$$

Let $\mathcal{X}_n = \{in^{-3} : i \in \{-2\lceil c_n\rceil n^3, \dots, 2\lceil c_n\rceil n^3\}\}$ be a grid that approximates each $x \in [-2c_n, 2c_n]$ with precision n^{-3}, and $V_n = \{in^{-3} : i = 1, \dots, n^3\}$. Since K are Lipschitz continuous with constant L_K,

$$\sqrt{nh_{1,n}h_{2,n}} \sup_{|x-x'|\le n^{-3}, |v-v'|\le n^{-3}} \big|\big(\hat{g}_{n,h_n}(x,v) - \mathbb{E}\hat{g}_{n,h_n}(x,v)\big)$$

$$
\begin{aligned}
& \qquad\qquad\qquad\qquad - \big(\hat{g}_{n,h_n}(x',v) - \mathbb{E}\hat{g}_{n,h_n}(x',v)\big)\big| \\
\leq\ & 2\frac{\sqrt{n}}{\sqrt{h_{1,n}h_{2,n}}} \sup_{|x-x'|\leq n^{-3},|v-v'|\leq n^{-3}} \Big[\frac{L_K|K|_\infty|x-x'|}{h_{2,n}} + \frac{L_K|K|_\infty|v-v'|}{h_{1,n}}\Big] \\
=\ & O(n^{-1}). \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (A.62)
\end{aligned}
$$

We conclude from (A.57), (A.61) and (A.62) that

$$
\begin{aligned}
& \sqrt{nh_{1,n}h_{2,n}} \sup_{x\in\mathbb{R},v\in[0,1]} |\hat{g}_{n,h_n}(x,v) - \mathbb{E}\hat{g}_{n,h_n}(x,v)| \\
=\ & \sqrt{nh_{1,n}h_{2,n}} \sup_{x\in\mathcal{X}_n,v\in V_n} |\hat{g}_{n,h_n}(x,v) - \mathbb{E}\hat{g}_{n,h_n}(x,v)| + O_p(1) \quad (A.63)
\end{aligned}
$$

It was already shown that Assumption 2.2 is satisfied. Furthermore, we can choose $\mathbb{D}_n = |K|_\infty$, $\mathbb{D}^\infty_{\nu_2,n} = \frac{|K|_\infty}{\sqrt{h_{1,n}}}$ with $\nu_2 = \infty$, and $\bar{F}(z,u) = \sup_{f\in\mathcal{F}} \bar{f}(z,u) \leq \frac{|K|_\infty}{\sqrt{h_{2,n}}} =: C_{\bar{F},n}$. Note that

$$
\begin{aligned}
& \mathbb{E}[(\sqrt{h_{2,n}}K_{h_{2,n}}(X_i - x))^2] \\
=\ & \mathbb{E}\big[\mathbb{E}[(\sqrt{h_{2,n}}K_{h_{2,n}}(X_i - x))^2|X_{i-1}]\big] \\
=\ & \int \Big(\int K(w)^\kappa f_{X_i|X_{i-1}=z}(x + wh_{2,n})dw\Big)^{1/\kappa} d\mathbb{P}^{X_{i-1}}(z) \\
\leq\ & C_\infty \cdot (\int K(w)^2 dw)^{1/2}.
\end{aligned}
$$

therefore

$$
\|f_{x,v}\|_{2,n} \leq \mathbb{D}_n C_\infty \int K(w)^2 dw,
$$

which implies $\sigma := \sup_{n\in\mathbb{N}} \sup_{f\in\mathcal{F}} V_n(f) < \infty$. Due to $\Delta(k) = O(k^{-\alpha s})$, the last condition in (4.4) is fulfilled if

$$
\sup_{n\in\mathbb{N}} \frac{\log(n)}{nh_{2,n}h_{1,n}^{\frac{\alpha s}{\alpha s-1}}} < \infty.
$$

By Corollary 4.2, we have

$$
\begin{aligned}
& \sqrt{nh_{1,n}h_{2,n}} \sup_{x\in\mathcal{X}_n,v\in V_n} \big|\hat{g}_{n,h_n}(x) - \mathbb{E}\hat{g}_{n,h_n}(x,v)\big| \\
=\ & \sup_{f\in\mathcal{F}} |\mathbb{G}_n(f)| = O_p(\sqrt{\log|\mathcal{F}|}) = O(\sqrt{\log(n)}).
\end{aligned}
$$

With (A.63), it follows that

$$
\sqrt{nh_{1,n}h_{2,n}} \sup_{x\in\mathbb{R},v\in[0,1]} |\hat{g}_{n,h_n}(x,v) - \mathbb{E}\hat{g}_{n,h_n}(x,v)| = O_p\big(\sqrt{\log(n)}\big).
$$

References

[1] Akritas, M. G. and Van Keilegom, I. (2001). Non-parametric estimation of the residual distribution. *Scand. J. Statist.* **28** 549–567. MR1858417

[2] Arcones, M. A. and Yu, B. (1994). Central limit theorems for empirical and U-processes of stationary mixing sequences. *J. Theoret. Probab.* **7** 47–71. MR1256391

[3] Berkes, I., Hörmann, S. and Schauer, J. (2009). Asymptotic results for the empirical process of stationary sequences. *Stochastic Process. Appl.* **119** 1298–1324. MR2508575

[4] Borovkova, S., Burton, R. and Dehling, H. (2001). Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation. *Trans. Amer. Math. Soc.* **353** 4261–4318. MR1851171

[5] Dahlhaus, R. and Polonik, W. (2009). Empirical spectral processes for locally stationary time series. *Bernoulli* **15** 1–39. MR2546797

[6] Dahlhaus, R., Richter, S. and Wu, W. B. (2019). Towards a general theory for nonlinear locally stationary processes. *Bernoulli* **25** 1013–1044. MR3920364

[7] Dedecker, J. (2010). An empirical central limit theorem for intermittent maps. *Probab. Theory Related Fields* **148** 177–195. MR2653226

[8] Dedecker, J. and Prieur, C. (2007). An empirical central limit theorem for dependent sequences. *Stochastic Process. Appl.* **117** 121–142. MR2287106

[9] Doukhan, P. (1994). *Mixing. Lecture Notes in Statistics* **85**. Springer-Verlag, New York Properties and examples. MR1312160

[10] Doukhan, P., Massart, P. and Rio, E. (1995). Invariance principles for absolutely regular empirical processes. *Ann. Inst. H. Poincaré Probab. Statist.* **31** 393–427. MR1324814

[11] Durieu, O. and Tusche, M. (2014). An empirical process central limit theorem for multidimensional dependent data. *J. Theoret. Probab.* **27** 249–277. MR3174225

[12] Mayer, U., Zähle, H. and Zhou, Z. (2020). Functional weak limit theorem for a local empirical process of non-stationary time series and its application. *Bernoulli* **26** 1891–1911. MR4091095

[13] Nishiyama, Y. et al. (2000). Weak convergence of some classes of martingales with jumps. *The Annals of Probability* **28** 685–712. MR1782271

[14] Phandoidaen, N. and Richter, S. (2020). Empirical process theory for locally stationary processes. MR4337712

[15] Pinelis, I. (1994). Optimum bounds for the distributions of martingales in Banach spaces. *Ann. Probab.* **22** 1679–1706. MR1331198

[16] Rio, E. (2013). Inequalities and limit theorems for weakly dependent sequences. Lecture.

[17] van der Vaart, A. W. (1998). *Asymptotic statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge University Press, Cambridge. MR1652247

[18] Wu, W. B. (2005). Nonlinear system theory: another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154. MR2172215

[19] Wu, W. B. (2008). EMPIRICAL PROCESSES OF STATIONARY SEQUENCES. *Statistica Sinica* **18** 313–333. MR2384990

[20] Wu, W. B., Liu, W. and Xiao, H. (2013). Probability and moment inequalities under dependence. *Statist. Sinica* **23** 1257–1272. MR3114713

[21] Yu, B. (1994). Rates of convergence for empirical processes of stationary mixing sequences. *Ann. Probab.* **22** 94–116. MR1258867

[22] Zhang, D. and Wu, W. B. (2017). Gaussian approximation for high dimensional time series. *Ann. Statist.* **45** 1895–1919. MR3718156