Electronic Journal of Probability

Electron. J. Probab. **19** (2014), no. 87, 1–26.
ISSN: 1083-6489 DOI: 10.1214/EJP.v19-3216

A sequential empirical CLT for multiple mixing processes with application to $\mathcal{B}$-geometrically ergodic Markov chains*

Herold Dehling† Olivier Durieu‡ Marco Tusche†‡

Abstract

We investigate the convergence in distribution of sequential empirical processes of dependent data indexed by a class of functions $\mathcal{F}$. Our technique is suitable for processes that satisfy a multiple mixing condition on a space of functions which differs from the class $\mathcal{F}$. This situation occurs in the case of data arising from dynamical systems or Markov chains, for which the Perron–Frobenius or Markov operator, respectively, has a spectral gap on a restricted space. We provide applications to iterative Lipschitz models that contract on average.

Keywords: Multivariate Sequential Empirical Processes; Limit Theorems; Multiple Mixing, Spectral Gap; Dynamical Systems; Markov chain; Change-Point Problems.
AMS MSC 2010: 60F05; 60F17; 60G10; 62G30; 60J05.
Submitted to EJP on December 20, 2013, final version accepted on September 15, 2014.

1 Introduction

The asymptotic behaviour of empirical processes has been studied for more than 60 years. The first rigorous result was the empirical process central limit theorem for i.i.d. data, established by Donsker [22]. This theorem, conjectured by Doob [23], made it possible to derive the asymptotic distribution of a large number of test statistics and estimators that can be represented as functionals of the empirical process, by an application of the continuous mapping theorem. Among the examples are the Kolmogorov-Smirnov goodness of fit test, the Cramér-Von Mises ω^2 criterion, and more generally von Mises statistics.

Ciesielski and Kesten [6] were among the first to extend Donsker's empirical process CLT to weakly dependent data, studying the empirical distribution of remainders in the dyadic expansion of a random number $\omega \in [0,1]$. Billingsley [5] proved the first general result for dependent data, namely an empirical process CLT for data that can be represented as functionals of a mixing process. For an overview of the literature on empirical

*Support: German Sci. Foundation DE 370-4 (*New Techniques for Empirical Processes of Dependent Data*).
†Fakultät für Mathematik, Ruhr-Universität Bochum, Germany.
E-mail: Herold.Dehling@RUB.de E-mail: Marco.Tusche@RUB.de
‡Laboratoire de Mathématiques et Physique Théorique UMR-CNRS 7350, Fédération Denis Poisson FR-CNRS 2964, Université François-Rabelais de Tours, France. E-mail: olivier.durieu@lmpt.univ-tours.fr

processes of dependent data, see Dehling and Philipp [19], Dedecker, Doukhan, Lang, León, Louhichi, and Prieur [10].

Müller [37], and independently Kiefer [33], initiated the study of the sequential empirical process, defined as

$$U_n(x,t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} \left(\mathbf{1}_{\{X_i \le x\}} - F(x) \right),$$

where $F(x) = P(X_1 \le x)$. The process $U_n(x,t)$ is also known as the two-parameter empirical process. Kiefer and Müller showed that for i.i.d. data, the sequential empirical process converges in distribution to a mean zero Gaussian process $K(x,t)$ with covariance structure

$$\mathbf{E}\left(K(x,t)K(y,u)\right) = \min(t,u)(F(\min(x,y)) - F(x)F(y)).$$

The limit process $K(x,t)$ is called Kiefer process, or Kiefer-Müller process.

Komlós, Major, and Tusnády [34], refining a technique originally invented by Csörgő and Révész [8], established the almost sharpest possible bounds for the error in the approximation of the sequential empirical process by the Kiefer process in the i.i.d. case so far. For an overview of this topic, see the book by Csörgő and Révész [9] or the survey article by Gänssler and Stute [27].

Many authors have studied extensions of the sequential empirical process CLT to dependent data, e.g. Berkes and Philipp [4] and Philipp and Pinzur [40] for strongly mixing processes and Berkes, Hörmann, and Schauer [3] for S-mixing processes. Recently, Dedecker, Merlevède, and Rio [13] proved strong approximation results for the sequential empirical process of some stationary sequences, see also Dedecker, Merlevède, and Rio [14] in the case of functions of absolutely regular sequences. Dehling and Taqqu [20] determined the asymptotic distribution of the sequential empirical process in the case of long-range dependent data.

Recently, Dehling, Durieu, and Volný [18] have developed a technique to prove empirical process CLTs for Markov chains and dynamical systems that do not necessarily satisfy any of the standard mixing conditions. The technique has been extended by Dehling and Durieu [16], Durieu and Tusche [26] and Dehling, Durieu, and Tusche [17] to multivariate empirical processes and to empirical processes indexed by classes of functions. Among the examples that could be treated by the new techniques are $\mathcal{B}$-geometrically ergodic Markov chains, dynamical systems with a spectral gap on the transfer operator and ergodic automorphisms of the d-dimensional torus, for which the empirical process CLT could be established. It is the goal of the present paper to extend these techniques to the sequential empirical process, with a special focus on $\mathcal{B}$-geometrically ergodic Markov chain. To this aim, we shall develop a sequential empirical CLT under multiple mixing (see definition in Section 2.2) that can be applied to this situation.

To illustrate our results, we present applications to a number of concrete examples. E.g., we establish a new sequential empirical process CLT for a class of Lipschitz models that contract on average; see Section 3.2. We also present an application to ergodic torus automorphisms, and to expanding maps of the unit interval. These last two examples have recently also been investigated by Dedecker, Merlevède, and Pène [12] and by Dedecker et al. [13], who obtained results similar to ours.

Sequential empirical process CLTs can be applied to the study of the asymptotic distribution of change-point tests based on the empirical distribution function. Suppose $(X_i)_{i\in\mathbb{N}}$ is a stochastic process with marginal distribution functions $\mu_1, \mu_2, \ldots$.

Given the observations $X_1, \ldots, X_n$, we want to test the hypothesis $\mathbf{H}_0$: *"the process is stationary with marginal distribution μ"* against the alternative $\mathbf{H}_A$: *"there exists a $k^* \in \{1, \ldots, n-1\}$ such that $(X_1, \ldots, X_{k^*})$ and $(X_{k^*+1}, \ldots, X_n)$ are both stationary with different marginal distributions"*. We propose the test statistic

$$T_n := \max_{0 \le k \le n} \sup_x \frac{k}{n}\left(1 - \frac{k}{n}\right)\sqrt{n}\big|F_k(x) - F_{k+1,n}(x)\big|,$$

where F_k denotes the empirical distribution function of the observations $X_1, \ldots, X_k$ and $F_{k+1,n}$ denotes the empirical distribution function of $X_{k+1}, \ldots, X_n$ (set $F_0 = F_{n+1,n} = 0$). In order to determine the asymptotic distribution of T_n, we study the $\ell^\infty(\mathbb{R} \times [0,1])$-valued process $R_n = (R_n(x,t))_{(x,t) \in \mathbb{R} \times [0,1]}$ given by

$$R_n(x,t) = \sqrt{n}t(1-t)\big(F_{[nt]}(x) - F_{[nt]+1,n}(x)\big).$$

As proved in Section 4 (Proposition 4.1), assuming "convergence of the sequential empirical process", we obtain under the null hypothesis $\mathbf{H}_0$ that

$$R_n \rightsquigarrow \big(K(x,t) - tK(x,1)\big)_{(x,t) \in \mathbb{R} \times [0,1]},$$

where K is the centred Gaussian process with covariance structure

$$\begin{aligned}&\mathbf{Cov}\big(K(x,t), K(y,u)\big)\\ &\quad = \min\{t,u\}\left\{\sum_{k=0}^{\infty} \mathbf{Cov}\big(\mathbf{1}_{\{X_0 \le x\}}, \mathbf{1}_{\{X_k \le y\}}\big) + \sum_{k=1}^{\infty} \mathbf{Cov}\big(\mathbf{1}_{\{X_0 \le y\}}, \mathbf{1}_{\{X_k \le x\}}\big)\right\}.\end{aligned}$$

This process is also referred to as a Kiefer process. Applying the continuous mapping theorem to the supremum-functional, we obtain the asymptotic distribution of the test statistic T_n under the null hypothesis, that is

$$T_n \rightsquigarrow \sup_{x \in \mathbb{R},\ t \in [0,1]} |K(x,t) - tK(x,1)|.$$

Note that, in fact this result remains true for general $\mathcal{F}$-indexed empirical processes, (see Theorem 4.3).

The remainder of this paper is organized as follows: In Section 2, we recall some definitions and give the statement of a sequential empirical CLT for multiple mixing processes (Theorem 2.5). We also discuss an application of our general technique to the situation of the ergodic automorphisms of the torus. In Section 3, as application, we present sequential empirical CLTs for $\mathcal{B}$-geometrically ergodic Markov chains (Theorem 3.5) and dynamical systems with a transfer operator having a spectral gap (Theorem 3.8). A concrete application of Theorem 3.5 to Lipschitz iterative models that contract on average (Corollary 3.6) is also given in this section. The asymptotic distribution of the test statistic T_n (Theorem 4.3) is given in Section 4. The proofs of the main results are postponed to Section 5 and Section 6.

2 A Sequential Empirical CLT for Multiple Mixing Processes

2.1 Definitions and Notations

Let $(\mathcal{X}, \mathcal{A})$ be a measurable space. For a positive measure λ on $\mathcal{X}$ and a λ-integrable complex-valued function f on $\mathcal{X}$, we will use the notation $\lambda f := \int_{\mathcal{X}} f \, d\lambda$. For $s \in [1, \infty)$, we denote by $\mathrm{L}^s(\lambda)$ the Lebesgue space of s-th power integrable complex-valued functions on $\mathcal{X}$. This space is equipped with the norm $\|f\|_s = (\lambda(|f|^s))^{1/s}$. Further, we denote the space of essentially bounded measurable functions on $\mathcal{X}$ w.r.t. λ by $L^\infty(\lambda)$ and the corresponding (essential) supremum norm by $\|\cdot\|_\infty$. Note that these norms depend heavily on the choice of the measure λ; however throughout this paper it will always be clear which measure we refer to.

Let $(X_i)_{i\in\mathbb{N}}$ be an $\mathcal{X}$-valued stationary stochastic process with marginal distribution μ and let $\mathcal{F}$ be a class of real-valued measurable functions on $\mathcal{X}$ which is uniformly bounded w.r.t. the $\|\cdot\|_\infty$-norm. For $n \in \mathbb{N}^*$, we define the map $F_n : \mathcal{F} \longrightarrow \mathbb{R}$, induced by the empirical measure, by

$$F_n(f) := \frac{1}{n}\sum_{i=1}^{n} f(X_i), \quad f \in \mathcal{F}.$$

The *sequential empirical process* of the n-th order of $(X_i)_{i\in\mathbb{N}}$ is then the $\mathcal{F}\times[0,1]$-indexed process $U_n := (U_n(f,t))_{(f,t)\in\mathcal{F}\times[0,1]}$ given by

$$U_n(f,t) := \frac{[nt]}{\sqrt{n}}\big(F_{[nt]}(f) - \mu f\big) = \frac{1}{\sqrt{n}}\sum_{i=1}^{[nt]}\big(f(X_i) - \mu f\big), \quad (f,t) \in \mathcal{F}\times[0,1],$$

where $[\cdot]$ denotes the lower Gauss bracket, i.e. $[x] := \sup\{z \in \mathbb{Z} : z \le x\}$.

For fixed $n \in \mathbb{N}^*$, we consider U_n as a random element in the metric space $\ell^\infty(\mathcal{F}\times[0,1])$ of bounded real-valued functions on $\mathcal{F}\times[0,1]$, equipped with the supremum norm and the corresponding Borel σ-field. Since $\mathcal{F}\times[0,1]$ is uncountable, here we cannot assume that U_n is measurable and thus standard techniques of weak convergence do not apply. We will therefore use the theory of outer probability and expectation (see van der Vaart and Wellner [43]).

Let $\mathbf{E}^* X$ denote the outer expectation of a possibly non-measurable random element X, let U be measurable, and let $U, U_0, U_1, \ldots$ take values in $\ell^\infty(\mathcal{F}\times[0,1])$. We define convergence in distribution or weak convergence $U_n \rightsquigarrow U$ in $\ell^\infty(\mathcal{F}\times[0,1])$ as the convergence $\mathbf{E}^*(\varphi(U_n)) \to \mathbf{E}(\varphi(U))$ of all bounded and continuous functions $\varphi : \ell^\infty(\mathcal{F}\times[0,1]) \longrightarrow \mathbb{R}$. We say that the process $(X_i)_{i\in\mathbb{N}}$ satisfies a sequential empirical CLT if the process U_n converges in distribution in $\ell^\infty(\mathcal{F}\times[0,1])$ to a tight centred Gaussian process.

Empirical CLTs usually require some bound of the size of the indexing class $\mathcal{F}$. This size is usually measured by counting certain sets, e.g. balls or brackets of a given $\|\cdot\|_s$-size, needed to cover $\mathcal{F}$ (c.f. Ossiander [38] and [43, p.83 ff.]). In our upcoming setting, we will only deal with properties for functions of a restricted class which could be disjoint of the class $\mathcal{F}$. We thus need an adapted notion of bracketing numbers. This notion was introduced in Dehling et al. [17].

Definition 2.1. *Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a probability space. For two functions $l, u : \mathcal{X} \to \mathbb{R}$ such that $l(x) \le u(x)$ for all $x \in \mathcal{X}$, we define the bracket*

$$[l,u] := \{f : \mathcal{X} \to \mathbb{R} : l(x) \le f(x) \le u(x), \text{ for all } x \in \mathcal{X}\}.$$

Let $\mathcal{G}$ be a subset of a normed real vector space $(\mathcal{C}, \|\cdot\|_{\mathcal{C}})$ of measurable real-valued functions on $\mathcal{X}$. For given $\varepsilon > 0$, $A > 0$, and $s \in [1,\infty]$, we call $[l,u]$ an $(\varepsilon, A, \mathcal{G}, \mathrm{L}^s(\mu))$-bracket, if $l, u \in \mathcal{G}$ and

$$\|u - l\|_s \le \varepsilon$$
$$\|u\|_{\mathcal{C}} \le A, \quad \|l\|_{\mathcal{C}} \le A.$$

For a class of real-valued functions $\mathcal{F}$ on $\mathcal{X}$, we define the bracketing number

$$N(\varepsilon, A, \mathcal{F}, \mathcal{G}, \mathrm{L}^s(\mu))$$

as the smallest number of $(\varepsilon, A, \mathcal{G}, \mathrm{L}^s(\mu))$-brackets needed to cover $\mathcal{F}$.

This notion of brackets allows to control the number of brackets needed to cover $\mathcal{F}$ not only with respect to the decreasing size of the brackets in L^s-norm, but also with a control of the increasing $\|\cdot\|_{\mathcal{C}}$-size of the bracketing functions as the L^s-norm goes to zero.

2.2 Multiple mixing processes and the main result

In this section, we present a general result which will be applied to $\mathcal{B}$-geometrically ergodic Markov chains in Section 3. We consider stationary sequences $(X_i)_{i\in\mathbb{N}}$ which satisfy a multiple mixing condition with respect to some space of functions. Consider some normed vector space $(\mathcal{C}, \|\cdot\|_{\mathcal{C}})$ of functions on $\mathcal{X}$. The multiple mixing condition is defined as follows.

Definition 2.2 (Multiple Mixing Processes)**.** *We say that a process $(X_i)_{i\in\mathbb{N}}$ is* multiple mixing *with respect to $\mathcal{C}$ if there exist a real $\theta \in (0,1)$, a real $s \geq 1$, and an integer $d_0 \in \mathbb{N}$ such that for all $p \in \mathbb{N}^*$, there exist an integer ℓ and a multivariate polynomial P of total degree not larger than d_0 such that*

$$\begin{aligned}&\left|\mathbf{Cov}(f(X_{i_0})\cdots f(X_{i_{q-1}}), f(X_{i_q})\cdots f(X_{i_p}))\right| \\ &\quad \leq \|f\|_s \|f\|_{\mathcal{C}}^{\ell} P(i_1 - i_0, \ldots, i_p - i_{p-1})\theta^{i_q - i_{q-1}}\end{aligned} \tag{2.1}$$

holds for all $f \in \mathcal{C}$ with $\mu f = 0$ and $\|f\|_\infty \leq 1$, all integers $i_0 \leq i_1 \leq \ldots \leq i_p$ and all $q \in \{1, \ldots, p\}$.

As proved in Dehling and Durieu [16], multiple mixing processes satisfy a moment bound which is particularly useful to establish empirical CLTs.

The approach developed here is useful when the indexing class $\mathcal{F}$ is different from the space $\mathcal{C}$. In the following we shall require the two following assumptions concerning the processes $(f(X_i))_{i\in\mathbb{N}}$, where $f : \mathcal{X} \longrightarrow \mathbb{R}$ belongs to $(\mathcal{C}, \|\cdot\|_{\mathcal{C}})$.

Assumption 2.3 (Finite dimensional sequential CLT for $\mathcal{C}$-observables)**.** *For any choice of $f_1, \ldots, f_k \in \mathcal{C}$ and $t_1, \ldots, t_k \in [0,1]$*

$$\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{[nt_1]}(f_1(X_i) - \mu f_1)\ ,\ \ldots\ ,\ \sum_{i=1}^{[nt_k]}(f_k(X_i) - \mu f_k)\right) \rightsquigarrow N(0, \Sigma),$$

where $N(0,\Sigma)$ denotes some k-dimensional normal distribution with mean zero and covariance matrix $\Sigma = (\Sigma_{i,j})_{1\leq i,j\leq k}$.

Assumption 2.4 (Multiple mixing w.r.t. $\mathcal{C}$)**.** *The process $(X_i)_{i\in\mathbb{N}}$ is multiple mixing with respect to $\mathcal{C}$, and with parameters $\theta \in (0,1)$, $s \geq 1$, and $d_0 \in \mathbb{N}$.*

To derive a CLT for an $\mathcal{F}$-indexed empirical process, we now have to precise the relation between the class $\mathcal{F}$ and the space $\mathcal{C}$. Note that, in the particular case where $\mathcal{F}$ is a subset of $\mathcal{C}$, from Assumption 2.3 we can infer the finite dimensional convergence of the process $(U_n)_{n\in\mathbb{N}}$. Then, the tightness can be established under an entropy condition on $\mathcal{F}$ that uses the usual bracketing number defined as in Ossiander [38]. Nevertheless, in many examples, the functions of $\mathcal{F}$ do not belong to the space $\mathcal{C}$. To overcome this difficulty, we have to measure how the functions of $\mathcal{F}$ are well approximated by the functions of $\mathcal{C}$. We will use the bracketing numbers introduced in the preceding section to obtain a control on the size of $\mathcal{F}$ which depends on the possibility of approximation by the space $\mathcal{C}$.

We can show the following sequential empirical CLT.

Theorem 2.5. *Let $(\mathcal{X}, \mathcal{A})$ be a measurable space, let $(X_i)_{i\in\mathbb{N}}$ be an $\mathcal{X}$-valued stationary process with marginal distribution μ, and let $\mathcal{F}$ be a uniformly bounded class of measurable functions on $\mathcal{X}$. Suppose that, for some normed vector space $\mathcal{C}$ of measurable functions on $\mathcal{X}$, Assumptions 2.3 and 2.4 hold.*

If there exist a subset $\mathcal{G}$ of $\mathcal{C}$ which is bounded in $\|\cdot\|_\infty$-norm, $C > 0$, $r > -1$, and $\gamma > d_0 + 1$ such that

$$\int_0^1 \varepsilon^r \sup_{\varepsilon\le\delta\le 1} N^2\big(\delta, \exp\big(C\delta^{-\frac{1}{\gamma}}\big), \mathcal{F}, \mathcal{G}, \mathrm{L}^s(\mu)\big) d\varepsilon < \infty \tag{2.2}$$

then the sequential empirical process U_n converges in distribution in $\ell^\infty(\mathcal{F}\times[0,1])$ to a tight Gaussian process K.

Observe that for $r' \ge 0$, inequality (2.2) holds for all $r > 2r' - 1$, if

$$N\big(\varepsilon, \exp\big(C\delta^{-\frac{1}{\gamma}}\big), \mathcal{F}, \mathcal{G}, \mathrm{L}^s(\mu)\big) = O(\varepsilon^{-r'}) \quad \text{as } \varepsilon \to 0.$$

Note further, that the supremum in (2.2) appears in order to deal with the possible non-monotonicity of the bracketing number.

Let us also mention that several classes of functions $\mathcal{F}$ which satisfy the condition (2.2) with respect to a space of bounded Hölder functions are presented in Dehling et al. [17]. Among these classes are indicators of rectangles, indicators of balls, indicators of ellipsoids, and a class of monotone functions in dimension 1.

In this general setting of Theorem 2.5, we are unable to specify the covariance structure of the limit process. The next corollary shows that under additional conditions, the limit process of U_n is indeed a Kiefer process.

Corollary 2.6. *In the situation of Theorem 2.5, assume further that*

(i) Assumption 2.3 holds with covariance matrix Σ given by

$$\Sigma_{i,j} = \min\{t_i, t_j\}\left\{\sum_{k=0}^\infty \mathbf{Cov}\big(f_i(X_0), f_j(X_k)\big) + \sum_{k=1}^\infty \mathbf{Cov}\big(f_j(X_0), f_i(X_k)\big)\right\}, \tag{2.3}$$

(ii) there exists a constant $D > 0$ such that for all $f \in \mathcal{G}\cup(\mathcal{G}-\mathcal{G})$ and all $\varphi \in \mathcal{F}\cup(\mathcal{F}-\mathcal{G})$

$$\big|\mathbf{Cov}\big(\varphi(X_0), f(X_k)\big)\big| \le D\|\varphi\|_\infty \|f\|_{\mathcal{C}}\theta^k, \tag{2.4}$$

Then the covariance structure of the limit process K is given by

$$\begin{aligned}&\mathbf{Cov}\big(K(f_1,t_1), K(f_2,t_2)\big)\\ &\quad= \min\{t_1,t_2\}\left\{\sum_{k=0}^\infty \mathbf{Cov}\big(f_1(X_0), f_2(X_k)\big) + \sum_{k=1}^\infty \mathbf{Cov}\big(f_1(X_k), f_2(X_0)\big)\right\},\end{aligned} \tag{2.5}$$

for all $f_1, f_2 \in \mathcal{F}$, $t_1, t_2 \in [0,1]$.

The proof of Theorem 2.5 and Corollary 2.6 are given, respectively, in Section 5 and Section 6.

Remark 2.7. *A centred Gaussian process K with covariance structure* (2.5) *is often referred to as a Kiefer process.*

In Section 3, we will apply Theorem 2.5 to prove a sequential empirical CLT for $\mathcal{B}$-geometrically ergodic Markov chains, which is the main motivation of the paper. Before, we would like to mention that other applications of Theorem 2.5 are possible.

Ergodic Automorphism of the Torus Let T be an ergodic automorphism of the d dimensional torus $\mathbb{T}^d$ as introduced in Section 4 of Dehling et al. [17]. Following the ideas of Dehling et al. [17], we can extend their theorem to a sequential empirical CLT. Let $\mathcal{G}$ be a bounded subset of $\mathcal{H}_\alpha(\mathbb{T}^d, \mathbb{R})$, $\alpha \in (0,1]$, let $\mu = \lambda$ be the Lebesgue measure on $\mathbb{T}^d$ and assume further that $\mathcal{F}$ is a uniformly bounded class of functions from $\mathbb{T}^d$ to $\mathbb{R}$. We denote by d_0 the size of the biggest Jordan block of T restricted to its neutral subspace. We can establish the following result which is not proved here.

Corollary 2.8. *Assume that the class $\mathcal{F}$ satisfies the condition* (2.2) *for some $\gamma > d_0+1$. Assume further that there exist $C > 0$ and $a > 0$ such that for all $f \in \mathcal{F}$ and $k \in \mathbb{N}^*$, there exists $g_k \in \mathcal{G}$ satisfying $\|f - g_k\|_1 \leq k^{-1}$ and $\|g_k\| \leq Ck^a$. Then the sequential empirical process given by*

$$U_n(f,t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} (f \circ T^i - \lambda f), \quad f \in \mathcal{F}, t \in [0,1]$$

converges in distribution in $\ell^\infty(\mathcal{F} \times [0,1])$ to a Kiefer process.

Note that both assumptions on $\mathcal{F}$ are satisfied e.g. if $\mathcal{F}$ is the class of indicators of rectangles, balls, or ellipsoids (to see this, follow the proof of Proposition 3.2, 3.5 and 3.6 in Dehling et al. [17]).

This result is proved in details in Tusche [42] by application of Theorem 2.5. We just notice here that, in this situation, the multiple mixing property holds (see Dehling and Durieu [16]) and that Assumption 2.3 can be derived from the classical CLT (see Tusche [42], Lemma 11.1). Further, assumption (ii) of Corollary 2.6 is not straightforward. Instead, using the second assumption of the proposition, we can show that there exist some $c > 0$ and $\theta \in (0,1)$ such that for all $f \in \mathcal{F}$ and $g \in \mathcal{H}_\alpha(\mathbb{T}^d, \mathbb{R})$, $|\mathbf{Cov}(f, g \circ T^n)| \leq c\|g\|_\alpha \theta^n$, which is sufficient to conclude as in Corollary 2.6.

Remark 2.9. *As mentioned earlier, for ergodic torus automorphisms Dedecker et al. [12] have investigated the sequential empirical process indexed by a class of the form $\{1_{(-\infty,t]} \circ f : t \in \mathbb{R}^l\}$, where $f : \mathbb{T}^d \to \mathbb{R}^l$ is fixed. Under some regularity assumptions on f, and using techniques different from ours, Dedecker et al. [12] obtain weak convergence to a Kiefer process. They also develop a tightness criterion (Proposition 3.13) that can be applied to many other examples, e.g. those given in Dedecker and Prieur [15].*

Multiple Mixing of Lower Rate Processes of a lower mixing rate have been studied by Durieu and Tusche [26]. They consider a multiple mixing condition w.r.t. the space of bounded α-Hölder functions on $\mathbb{R}^d$, $\alpha \in (0,1]$, where the term $\theta^{i_q - i_{q-1}}$ in (2.1) is replaced by a general term $\Theta(i_q - i_{q-1})$ with a monotone decreasing function $\Theta : \mathbb{N} \to \mathbb{R}_+$. Under the condition that $\sum_{i=0}^{\infty} i^{2p-2}\Theta(i) < \infty$, they were able to establish an empirical CLT. This could also be extend to a sequential version. Since it is not needed for our application, we decide to not develop this very general setting here. We can just remark that, in this situation, a stronger entropy condition will be needed. In particular, the second parameter in the bracketing number which appears in (2.2) should be replaced by a polynomial function of ε^{-1}.

3 A Sequential Empirical CLT under Spectral Gap

We now present an application of Theorem 2.5 to establish a sequential empirical CLT for Markov chain having a spectral gap property.

3.1 $\mathcal{B}$-geometrically ergodic Markov chains

In the following, let $(X_i)_{i\in\mathbb{N}}$ be a time homogeneous Markov chain on a measurable state space $(\mathcal{X},\mathcal{A})$ with a probability transition P and an invariant measure ν. We assume that the Markov chain starts with initial distribution ν, i.e that the distribution of X_0 is ν. This makes $(X_i)_{i\in\mathbb{N}}$ a stationary sequence. We also denote by P the associated Markov operator defined by

$$Pf = \int_{\mathcal{X}} f(y)\, P(\cdot, dy).$$

Now, let $(\mathcal{B}, \|\cdot\|_{\mathcal{B}})$ a Banach space of measurable functions from $\mathcal{X}$ to $\mathbb{R}$. We will assume that P is a bounded linear operator on $\mathcal{B}$, and we denote by $\mathcal{L}(\mathcal{B})$ the space of all bounded linear operators from $\mathcal{B}$ to $\mathcal{B}$.

We will need the following properties on the Banach space $\mathcal{B}$:

(A) $\mathbf{1}_{\mathcal{X}} \in \mathcal{B}$ and $P \in \mathcal{L}(\mathcal{B})$.

For some $m \in [1,\infty]$,

(B) $\mathcal{B}$ is continuously included in $L^m(\nu)$, i.e. there is a $K>0$ such that $\|\cdot\|_m \le K\|\cdot\|_{\mathcal{B}}$.

Further we consider processes such that the action of the corresponding Markov operator on $\mathcal{B}$ satisfies

(C) $\|P^n f - (\nu f)\,\mathbf{1}_{\mathcal{X}}\|_{\mathcal{B}} \le \kappa \|f\|_{\mathcal{B}} \theta^n$ for some $\kappa > 0$, $\theta \in [0,1)$, and all $f \in \mathcal{B}$.

This property is often referred to as strong or geometric ergodicity with respect to $\mathcal{B}$ (c.f. Meyn and Tweedie [36], Hervé [30], and Hervé and Pène [31]).

Remark 3.1. *Note that condition* (C) *corresponds to a spectral gap property of* P *acting on* $\mathcal{B}$*, i.e.* 1 *is the only eigenvalue of modulus one, it is simple, and the rest of the spectrum is contained in a disk of radius strictly smaller than one. Further, in this case there exists a decomposition of the linear operator* P *in* $\mathcal{L}(\mathcal{B})$*,*

$$P = \Pi + N,$$

such that $\Pi f = (\nu f)\,\mathbf{1}_{\mathcal{X}}$ *is a projection on the eigenspace of* 1*,* $N \circ \Pi = \Pi \circ N = 0$*, and* $\rho(N) := \lim_{n\to\infty} \|N^n\|_{\mathcal{L}(\mathcal{B})}^{1/n} < 1$*, where* $\|\cdot\|_{\mathcal{L}(\mathcal{B})}$ *denotes the operator norm on* $\mathcal{B}$*.*

We first show below that the conditions (A) – (C) guarantee a sequential finite dimensional CLT for functions in $\mathcal{B}$.

Actually, we will establish a k-dimensional Donsker invariance principle, which of course implies the desired result by a projection.

Proposition 3.2. *Suppose that for some* $m \in [1,\infty]$*,* (A)*,* (B)*,* (C) *hold. Let* k *be a positive integer and* $f_1,\dots,f_k \in \mathcal{B} \cap L^s(\nu)$*, with* $s = m/(m-1)$*. Then*

$$\big(U_n(f_1,t),\dots,U_n(f_k,t)\big)_{t\in[0,1]} \rightsquigarrow W \tag{3.1}$$

in $(\ell^\infty[0,1])^k$*, where* $W := \big(W_1(t),\dots,W_k(t)\big)_{t\in[0,1]}$ *is a centred Gaussian process with covariances*

$$\mathbf{Cov}\big(W_i(t),W_j(u)\big) = \min\{t,u\}\left\{\sum_{k=0}^{\infty}\mathbf{Cov}\big(f_i(X_0),f_j(X_k)\big) + \sum_{k=1}^{\infty}\mathbf{Cov}\big(f_j(X_0),f_i(X_k)\big)\right\}.$$

In particular this proposition shows that Assumption 2.3 holds with covariance structure (2.3).

Proof. To prove this proposition, we will use a result of Dedecker and Merlevède [11]. An application of their Corollary 2 (see also Theorem 2 in Dedecker and Merlevède [11]) yields that a sufficient condition for the convergence (3.1) is that the centred random vector $Z_i := \big(f_1(X_i) - \nu f_1, \dots, f_k(X_i) - \nu f_k\big)$ satisfies

$$\sum_{i=0}^{\infty} \mathbf{E}\big(|Z_0|\,|\,\mathbf{E}(Z_i|X_0)|\big) < \infty. \tag{3.2}$$

Here $|\cdot|$ denotes the Euclidean norm on $\mathbb{R}^k$. By Hölders inequality one has

$$\mathbf{E}\big(|Z_0|\,|\,\mathbf{E}(Z_i|X_0)|\big) \le \mathbf{E}(|Z_i|^s)^{\frac{1}{s}}\,\mathbf{E}\big(\big|\mathbf{E}(Z_i|X_0)\big|^m\big)^{\frac{1}{m}}.$$

The assumption that the f_j belong to $L^s(\nu)$ gives that $\mathbf{E}(|Z_i^s|)^{\frac{1}{s}} < \infty$. Applying (B) and (C), we finally obtain

$$\begin{aligned}&\mathbf{E}\big(\big|\mathbf{E}(Z_i|X_0)\big|^m\big)^{\frac{1}{m}}\\ &\quad\le \sum_{j=1}^{k}\big\|\mathbf{E}(f_j(X_i) - \nu f_j|X_0)\big\|_m \le K\sum_{j=1}^{k}\|P^i f_j - \nu f_j\|_{\mathcal{B}} \le K\kappa\sum_{j=1}^{k}\|f_j\|_{\mathcal{B}}\theta^i,\end{aligned}$$

which shows that (3.2) holds and thus proves the proposition. □

Note that without the assumption that the f_i belong to $L^s(\nu)$, it is still possible to prove a finite-dimensional sequential CLT (Assumption 2.3) using the Nagaev method consisting of operator perturbations. However, without $f_i \in L^s(\nu)$ we do not obtain a characterization of the covariance matrix (see Tusche [42] for details).

Now, to apply Theorem 2.5 to prove a sequential empirical CLT, we need to show the multiple mixing property of $(X_i)_{i\in\mathbb{N}}$. To this aim, the following further condition on the space $\mathcal{B}$ is useful.

(D) There exist $C > 0$ and $\ell \in \mathbb{N}^*$ such that, if $f \in \mathcal{B}$ and $g \in \mathcal{B}$ are bounded by 1, then $fg \in \mathcal{B}$ and $\|fg\|_{\mathcal{B}} \le C\max\{\|f\|_{\mathcal{B}}, \|g\|_{\mathcal{B}}\}^{\ell}$.

Note that if $\mathcal{B}$ is a Banach algebra, condition (D) holds with $\ell = 2$.

The following lemma is now a straightforward extension of Lemma 3 in Dehling and Durieu [16].

Lemma 3.3. *Under the conditions* (A), (B), (C), *and* (D), $(X_i)_{i\in\mathbb{N}}$ *satisfies the multiple mixing property w.r.t.* $\mathcal{B}$ *with* $d_0 = 0$ *and* $s = m/(m-1)$.

Eventually, observe that the second assumption of Corollary 2.6 is also satisfied as it is shown by the following lemma.

Lemma 3.4. *Under the conditions* (A), (B), *and* (C), *for all* $f \in \mathcal{B}$ *and all* $g \in \mathrm{L}^s(\nu)$, *with* $s = \frac{m}{m-1}$, *we have*

$$|\,\mathbf{Cov}(g(X_0), f(X_k))| \le C\|g\|_s\|f\|_{\mathcal{B}}\theta^k.$$

Proof. Applying successively Hölder's inequality, (B), and (C), we get

$$\begin{aligned}|\,\mathbf{Cov}(g(X_0), f(X_k))| &\le \mathbf{E}\,|g(X_0)\,\mathbf{E}(f(X_n) - \nu f|X_0)|\\ &\le \|g\|_s\|P^k f - (\nu f)\,\mathbf{1}_{\mathcal{X}}\,\|_{\mathcal{B}}\\ &\le C\|g\|_s\|f\|_{\mathcal{B}}\theta^k.\end{aligned}$$

□

As a conclusion, we thus have the following sequential empirical central limit theorem as a corollary of Theorem 2.5, Corollary 2.6, Proposition 3.2, and Lemma 3.3.

Theorem 3.5 (Sequential empirical CLT for $\mathcal{B}$-geometrically ergodic Markov chains)**.** *Let $\mathcal{F}$ be a $\|\cdot\|_\infty$-bounded class of functions from $\mathcal{X}$ to $\mathbb{R}$. Assume that for some $m \in [1,\infty]$, the conditions* (A), (B), (C), *and* (D) *hold. If there is a $\|\cdot\|_\infty$-bounded subset $\mathcal{G} \subset \mathcal{B}$ such that* (2.2) *is satisfied with $s = m/(m-1)$, then the sequential empirical process converges in distribution in $\ell^\infty(\mathcal{F}\times[0,1])$ to a centred Gaussian process K with covariance structure given by* (2.5).

Now, let us give an example by applying Theorem 3.5 to random iterative Lipschitz models.

3.2 Iterative Lipschitz models that contract on average

In this section, we assume that $(\mathcal{X},d)$ is a (not necessarily compact) metric space in which every closed ball is compact. Further we assume that $\mathcal{X}$ is equipped with the Borel σ-algebra $\mathfrak{B}(\mathcal{X})$. Let $\{T_i,\ i\geq 0\}$ be a family of Lipschitz maps from $\mathcal{X}$ to $\mathcal{X}$. We consider the Markov chain with state space $\mathcal{X}$ and transition probability P given by

$$P(x,A) = \sum_{i\geq 0} p_i(x)\,\mathbf{1}_A(T_i(x)), \quad x\in\mathcal{X},\ A\in\mathfrak{B}(\mathcal{X}),$$

where the p_i are Lipschitz functions from $\mathcal{X}$ to $[0,1]$ which satisfy $\sum_{i\geq 0} p_i(x) = 1$ for all $x\in\mathcal{X}$. Thus, each step of the Markov chain corresponds to the application of one of the maps T_i which is chosen randomly with respect to a probability distribution which depends on the actual state of the chain. We assume that this model has a property of contraction on average, that is that there exists a $\rho\in(0,1)$ such that

$$\sum_{i\geq 0} d(T_i(x),T_i(y))p_i(x) < \rho d(x,y), \quad \forall x,y\in\mathcal{X}. \tag{3.3}$$

Statistical properties of such models have been studied by Dubins and Freedman [24], Barnsley and Elton [2], Hennion and Hervé [29], Wu and Shao [45], Hervé [30], and by Hervé and Pène [31] in the case of constant functions p_i and by Döblin and Fortet [21], Karlin [32], Barnsley, Demko, Elton, and Geronimo [1], Peigné [39], Pollicott [41], and by Walkden [44] in the case of variable functions p_i.

As in many of the cited papers, we need the following technical properties. For some fixed $x_0\in\mathcal{X}$, suppose

$$\sup_{\substack{x,y,z\in\mathcal{X},\\ y\neq z}} \sum_{i\geq 0} \frac{d(T_i(y),T_i(z))}{d(y,z)} p_i(x) < \infty, \tag{3.4}$$

$$\sup_{x,y\in\mathcal{X}} \sum_{i\geq 0} \frac{d(T_i(y),x_0)}{1+d(y,x_0)} p_i(x) < \infty, \tag{3.5}$$

$$\sup_{x\in\mathcal{X}} \sum_{i\geq 0} \frac{d(T_i(x),x_0)}{1+d(x,x_0)} \sup_{y,z\in\mathcal{X}, y\neq z} \frac{|p_i(y)-p_i(z)|}{d(y,z)} < \infty. \tag{3.6}$$

Moreover assume that for all $x,y\in\mathcal{X}$, there exist sequences of integers $(i_n)_{n\geq 1}$ and $(j_n)_{n\geq 1}$ such that

$$d\big(T_{i_n}\circ\ldots\circ T_{i_1}(x)\,,\,T_{j_n}\circ\ldots\circ T_{j_1}(y)\big)\big(1+d\big(T_{j_n}\circ\ldots\circ T_{j_1}(x)\,,\,x_0\big)\big)\to 0 \quad \text{as } n\to\infty \tag{3.7}$$

with $p_{i_n}(T_{i_{n-1}}\circ\ldots\circ T_{i_1}(x))\cdot\ldots\cdot p_{i_1}(x) > 0$ and $p_{j_n}(T_{j_{n-1}}\circ\ldots\circ T_{j_1}(y))\cdot\ldots\cdot p_{j_1}(x) > 0$. Note that conditions (3.4) – (3.6) are verified when the family of maps T_i is finite and (3.7) is verified when (3.3) – (3.6) hold and each p_i is positive. See Peigné [39] for a discussion on these assumptions.

Under the conditions (3.3) – (3.7), Peigné [39] proved that the Markov chain has an attractive P-invariant probability measure ν with existing first moment. We define the stationary process $(X_i)_{i\in\mathbb{N}}$ on $\mathcal{X}$ as the Markov chain with transition probability P starting with distribution ν, that is $X_0 \sim \nu$.

A central limit theorem for the empirical process associated to the Markov chain $(X_i)_{i\geq 0}$ was proved by Durieu [25] (see also Wu and Shao [45] in the case of constant functions p_i). The following theorem extends this result to the sequential empirical processes.

For $\alpha \in (0,1]$, we consider the space $\mathcal{H}_\alpha(\mathcal{X})$ of bounded α-Hölder continuous functions on $\mathcal{X}$ with values in $\mathbb{R}$, equipped with the norm

$$\|\cdot\|_{\mathcal{H}_\alpha} := \|\cdot\|_\infty + m_\alpha(\cdot),$$

where

$$m_\alpha(f) := \sup_{\substack{x,y\in\mathcal{X}\\ x\neq y}} \frac{|f(x)-f(y)|}{d(x,y)^\alpha}.$$

Corollary 3.6. *Let* (3.3) – (3.7) *hold,* $(X_i)_{i\in\mathbb{N}}$ *be the Markov chain with transition probability* P *starting under the invariant distribution* ν*, and consider a* $\|\cdot\|_\infty$*-bounded class of functions* $\mathcal{F}$*. Let* $s \in (1,2)$ *and* $\mathcal{G}$ *be a* $\|\cdot\|_\infty$*-bounded subset of the space* $\mathcal{H}_\alpha(\mathcal{X})$ *for some* $\alpha < \frac{s-1}{s}$ *such that* (2.2) *holds. Then the* $\mathcal{F}$*-indexed sequential empirical process* $(U_n(f,t))_{\mathcal{F}\times[0,1]}$ *associated to the process* $(X_i)_{i\geq 0}$ *converges in distribution in the space* $\ell^\infty(\mathcal{F}\times[0,1])$ *to a centred Gaussian process with covariance given by* (2.5)*.*

Proof. First, we introduce spaces of Lipschitz functions with weights that give the geometric ergodicity of the chain. For every $\alpha,\beta \in [0,1]$, let $\mathcal{H}_{\alpha,\beta}(\mathcal{X})$ denote the space of continuous function from $\mathcal{X}$ to $\mathbb{R}$ with $\|f\|_{\mathcal{H}_{\alpha,\beta}} = N_\beta(f) + m_{\alpha,\beta}(f) < \infty$, where

$$N_\beta(f) = \sup_{x\in\mathcal{X}} \frac{|f(x)|}{1+d(x,x_0)^\beta} \quad \text{and} \quad m_{\alpha,\beta}(f) = \sup_{x,y\in\mathcal{X},x\neq y} \frac{|f(x)-f(y)|}{d(x,y)^\alpha(1+d(x,x_0)^\beta)}.$$

In particular, the space $\mathcal{H}_\alpha(\mathcal{X}) := \mathcal{H}_{\alpha,0}(\mathcal{X})$ is the space of bounded α-Hölder functions from $\mathcal{X}$ to $\mathbb{R}$ and we have $\|\cdot\|_{\mathcal{H}_{\alpha,0}} = 2^{-1}\|\cdot\|_{\mathcal{H}_\alpha}$. It is a subspace of $\mathcal{H}_{\alpha,\beta}(\mathcal{X})$ for all $\beta > 0$. The following properties are straightforward and given without proof.

Lemma 3.7. *For all* α *and* $\beta \in [0,1]$*,*

(i) the space $(\mathcal{H}_{\alpha,\beta}(\mathcal{X}), \|\cdot\|_{\mathcal{H}_{\alpha,\beta}})$ *is a Banach space which satisfies condition* (A)*,*

(ii) for every bounded functions $f,g \in \mathcal{H}_{\alpha,\beta}(\mathcal{X})$*, we have that* $\|fg\|_{\mathcal{H}_{\alpha,\beta}} \leq \|f\|_\infty\|g\|_{\mathcal{H}_{\alpha,\beta}} + \|g\|_\infty\|f\|_{\mathcal{H}_{\alpha,\beta}}$*,*

(iii) for every $f \in \mathcal{H}_\alpha(\mathcal{X})$ *and* $g \in \mathcal{H}_{\alpha,\beta}(\mathcal{X})$*, we have that* $\|fg\|_{\mathcal{H}_{\alpha,\beta}} \leq \|f\|_{\mathcal{H}_\alpha}\|g\|_{\mathcal{H}_{\alpha,\beta}}$*,*

(iv) there exists a $C > 0$*, for every* $f \in \mathcal{H}_{\alpha,\beta}(\mathcal{X})$*,* $f \in L^{\frac{1}{\beta}}(\nu)$ *and* $\|f\|_{\frac{1}{\beta}} \leq CN_\beta(f)$*.*

Therefore condition (B) holds with $m = 1/\beta$ as a consequence of (iv), and condition (D) is satisfied due to (ii). Now, according to Theorem 1 in Peigné [39], we obtain for all $\alpha,\beta \in (0,1/2)$ with $\alpha < \beta$ that P is a bounded linear operator on $\mathcal{H}_{\alpha,\beta}(\mathcal{X})$ which satisfies condition (C).

We now apply theorem 3.5. Let s, α, and $\mathcal{G}$ be as in the statement of Corollary 3.6. By choosing $\beta = (s-1)/s < \frac{1}{2}$, we have $\alpha < \beta$ and thus (A) – (D) hold for the space $\mathcal{B} = \mathcal{H}_{\alpha,\beta}(\mathcal{X})$ with $m = 1/\beta$. Further, for any $g \in \mathcal{G}$, we have $g \in \mathcal{H}_{\alpha,\beta}(\mathcal{X})$ and $\|g\|_{\mathcal{H}_{\alpha,\beta}} \leq \|g\|_{\mathcal{H}_\alpha}$. Therefore, condition (2.2) is also satisfied with respect to the $\mathcal{H}_{\alpha,\beta}(\mathcal{X})$-norm. ☐

3.3 Dynamical Systems with a Spectral Gap

Let us mention that the result of Section 3.1 can be adapted to deal with dynamical systems using the Perron–Frobenius operator in place of the Markov operator. Let $(\mathcal{X}, \mathcal{A})$ be a measurable space and let T be a measurable transformation on $\mathcal{X}$ which preserves a probability measure μ on $(\mathcal{X}, \mathcal{A})$. Let P be the associated Perron–Frobenius operator defined on $L^1(\mu)$ by the equation

$$\mu(f \cdot Pg) = \mu(f \circ T \cdot g), \quad \forall f \in L^\infty(\mu), g \in L^1(\mu).$$

We have the following result which can be derived from theorem 3.5 using relativized kernel as in Hennion and Hervé [29], Chapter XI.

Theorem 3.8 (Sequential empirical CLT for dynamical systems with a spectral gap)**.** *Let $\mathcal{F}$ be a $\|\cdot\|_\infty$-bounded class of functions from $\mathcal{X}$ to $\mathbb{R}$. Assume that there exist a Banach space $\mathcal{B}$ and $m \in [1, \infty]$ such that the conditions* (A)*,* (B)*,* (C)*, and* (D) *hold with respect to the Perron–Frobenius operator and replacing ν by μ. If there exists a $\|\cdot\|_\infty$-bounded subset $\mathcal{G} \subset \mathcal{B}$ such that* (2.2) *holds for $s = \frac{m}{m-1}$, then the process $(U_n(f,t))_{\mathcal{F}\times[0,1]}$, defined by $U_n(f,t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} \left(f \circ T^i - \mu f\right)$, converges in distribution in $\ell^\infty(\mathcal{F} \times [0,1])$ to a centred Gaussian process K with covariance structure given by*

$$\mathbf{Cov}(K(f,t), K(g,u)) = \min\{t,u\} \left(\sum_{k=0}^{\infty} \mathbf{Cov}(f, g \circ T^k) + \sum_{k=1}^{\infty} \mathbf{Cov}(f \circ T^k, g) \right).$$

As a possible application, we can extend the empirical CLT proved by Collet, Martinez, and Schmitt [7] for a class of expanding maps of the interval, to a sequential empirical CLT. In the situation considered in Collet et al. [7], the spectral gap property can be established on the space of functions of bounded variation.

We consider a piecewise C^2 expanding map T of the interval $[0,1]$ which is topologically mixing. We assume that there is a finite partition of $[0,1]$ by intervals such that T is monotone on each interval and further $\inf_{x\in[0,1]} |(T^n)'(x)| \geq CK^n$ for some $C > 0$ and $K > 1$. As noted in Collet et al. [7] (see also Lasota and Yorke [35]), there is a unique ergodic invariant probability measure μ such that $d\mu = h(x)d\lambda$, where λ is the Lebesgue measure on $[0,1]$. The function h belongs to the Banach algebra BV of functions of bounded variation. By application of Theorem 3.8, we obtain the following result.

Corollary 3.9. *Assume that $\frac{1}{h}\mathbf{1}_{h>0} \in BV$, and let $\mathcal{F}$ be a $\|\cdot\|_\infty$-bounded class of functions such that there exists a subset $\mathcal{G}$ of BV for which* (2.2) *holds for some $s \geq 1$. Then the process $(U_n(f,t))_{\mathcal{I}\times[0,1]}$, defined by $U_n(f,t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} \left(f \circ T^i - \mu f\right)$, converges in distribution in $\ell^\infty(\mathcal{I} \times [0,1])$ to a centred Kiefer process.*

Note that, in the usual case where $\mathcal{F} = \{\mathbf{1}_{[0,t]} \mid t \in [0,1]\}$, this result is not new. It can be derived from the result of Section 3 in Dedecker et al. [13], since the coefficient $\beta_{2,X}(n)$ which is considered in that paper decreases exponentially fast in our setting (see Dedecker and Prieur [15], Section 6.3). In Dedecker et al. [13], the result is stronger since a strong approximation by a Kiefer process is proved. This implies our weak convergence result.

Proof. Recall (see Hennion and Hervé [29]) that the Perron-Frobenius operator P associated with T and λ has a spectral gap on BV: 1 is a simple eigenvalue with eigenfunction h, and the rest of the spectrum is in a disk of radius strictly smaller than 1. Further, the space BV satisfies assumptions (A) and (B) (with $m = +\infty$) and the operator P satisfies $\|P^n f - (\lambda f)h\|_{BV} \leq \kappa \|f\|_{BV} \theta^n$ for some $\kappa > 0$, $\theta \in [0,1)$, and all $f \in BV$. BV being a Banach algebra, condition (D) is also satisfied.

In general, the Lebesgue measure is not the invariant measure, i.e. h is not 1. Thus, we define the set $I_h = \{x \in [0,1] \mid h(x) > 0\}$ and for functions defined on I_h, we introduce the operator P_h defined by $P_h f(x) = \frac{1}{h(x)} P(fh)(x)$. Note that, since $\mu(I_h) = 1$, every function f defined on $[0,1]$ is μ almost surely equal to the function defined by f on I_h and 0 on $[0,1]\backslash I_h$. With this remark, we can easily check that $\mu(f \cdot P_h g) = \mu(f \circ T \cdot g)$, for all $f \in L^\infty(\mu)$ and $g \in L^1(\mu)$. Then P_h is the Perron-Frobenius operator associated with T and μ. Now, if a function f is defined on I_h, the function fh can be considered on $[0,1]$ by giving the value 0 on $[0,1]\backslash I_h$. We introduce the space $\mathcal{B}_h = \{f : I_h \to \mathbb{R} \mid fh \in BV\}$ equipped with the norm $\|f\|_h = \|fh\|_{BV}$. Let us check that the assumptions of Theorem 3.8 are satisfied for P_h and $\mathcal{B}_h$.

Clearly, $\mathcal{B}_h$ satisfies the condition (A). The fact that $\frac{1}{h}\mathbf{1}_{h>0} \in BV$ gives (B) (with $m = +\infty$) and (D). From the spectral decomposition of P we derive the spectral decomposition of P_h and we obtain the condition (C) on the space $\mathcal{B}_h$ (with μ instead of ν). Thus Theorem 3.8 can be applied in this situation. □

As a simple example, we can consider any class of functions $\mathcal{F} = \{f_t \mid t \in [0,1]\}$ indexed by a parameter $t \in [0,1]$ that satisfies:

- for all $t \in [0,1]$, f_t is a non-increasing function bounded by 1,
- for all $0 \le t \le u \le 1$, $f_t \le f_u$,
- the function $t \in [0,1] \mapsto \mu f_t$ is α-Hölder for some $\alpha \in (0,1]$.

Indeed, in this situation the choice $\mathcal{G} = \mathcal{F}$ is possible. For all $t \in [0,1]$, f_t is BV with $\|f_t\|_{BV} \le 2$. Now, fix $\varepsilon > 0$ and choose $m = \lfloor \varepsilon^{-\frac{1}{\alpha}} \rfloor$. Let $t_i = \frac{i}{m}$, $i = 0, \dots, m$. For all $t \in [0,1]$, there exists $i \in \{1, \dots, m\}$ such that $t_i \le t \le t_{i+1}$ and thus $f_{t_i} \le f_t \le f_{t_{i+1}}$. Further,

$$\|f_{t_i} - f_{t_{i+1}}\|_1 = \mu f_{t_i} - \mu f_{t_{i+1}} \le C(\frac{1}{m})^\alpha \le C\varepsilon.$$

This shows that $N(\varepsilon, 2, \mathcal{F}, \mathcal{F}, \mathrm{L}^1(\mu)) = O(\varepsilon^{-\frac{1}{\alpha}})$ as $\varepsilon \to 0$ and gives (2.2).

Gouëzel [28] gave examples of expanding maps of the interval for which the Perron-Frobenius operator does not act on the space of bounded variation functions, but acts on the space of Lipschitz functions with a spectral gap property. These examples also satisfy the assumptions of our theorem and thus sequential empirical CLTs can be proved. Note that the space of Lipschitz functions is a Banach algebra and thus condition (D) is trivially satisfied. Further, the usual class of the indicator functions of intervals can be well approximated by Lipschitz functions, and the condition (2.2) is verified for this class.

4 Statistical applications

As mentioned in the introduction, sequential empirical CLTs can be applied to derive asymptotic distributions in change-point tests based on the empirical distribution function. We shall consider below the natural generalization of the process T_n, introduced in Section 1, to processes taking values in a measurable space $\mathcal{X}$. Let $(X_i)_{i \in \mathbb{N}}$ be a $\mathcal{X}$-valued stationary process, and $\mathcal{F}$ be a class of function on $\mathcal{X}$. As before, we denote the empirical measure by $F_n(f) := n^{-1}\sum_{i=1}^n f(X_i)$, $n \in \mathbb{N}^*$, and we set $F_0(f) = 0$. For $j \in \{1, \dots, n\}$, we define $F_{j,n}(f) := (n-j+1)^{-1}\sum_{i=j}^n f(X_i)$ and set $F_{n+1,n}(f) := 0$. Consider the $\ell^\infty(\mathcal{F} \times [0,1])$-valued process $R_n = (R_n(f,t))_{(f,t) \in \mathcal{F} \times [0,1]}$ given by

$$R_n(f,t) := \sqrt{n}\frac{[nt]}{n}\frac{n-[nt]}{n}\big(F_{[nt]}(f) - F_{[nt]+1,n}(f)\big).$$

The following theorem gives the asymptotic distribution of R_n.

Proposition 4.1. *Assume that $(X_i)_{i\in\mathbb{N}}$ satisfies the sequential empirical CLT with indexing class $\mathcal{F}$ and limit process K, that is, $U_n \rightsquigarrow K$ in $\ell^\infty(\mathcal{F}\times[0,1])$ as $n\to\infty$, where K denotes a tight centred Gaussian process. Then*

$$R_n \rightsquigarrow (K(f,t) - tK(f,1))_{(f,t)\in\mathcal{F}\times[0,1]}$$

in $\ell^\infty(\mathcal{F}\times[0,1])$ to as $n\to\infty$.

Proof. Let μ denote the distribution function of the X_i. For $t\in[1/n,1)$ we have

$$\begin{aligned}
&F_{[nt]}(f) - F_{[nt]+1,n}(f)\\
&= \frac{1}{[nt]}\sum_{i=1}^{[nt]} f(X_i) - \frac{1}{n-[nt]}\sum_{i=[nt]+1}^{n} f(X_i)\\
&= \frac{1}{[nt]}\sum_{i=1}^{[nt]} (f(X_i)-\mu f) - \frac{1}{n-[nt]}\sum_{i=[nt]+1}^{n} (f(X_i)-\mu f)\\
&= \left(\frac{1}{[nt]} + \frac{1}{n-[nt]}\right)\sum_{i=1}^{[nt]} (f(X_i)-\mu f) - \frac{1}{n-[nt]}\sum_{i=1}^{n} (f(X_i)-\mu f)\\
&= \frac{1}{\sqrt{n}}\frac{n}{[nt]}\frac{n}{n-[nt]}\,U_n(f,t) - \frac{1}{\sqrt{n}}\frac{1}{t}\frac{n}{n-[nt]}\,tU_n(f,1).
\end{aligned}\tag{4.1}$$

Further, by definition we have $R_n(f,1)=0$ and $R_n(f,t)=0$ for $t\in[0,1/n)$. Since also $U_n(f,t)=0$ for $t\in[0,1/n)$, we obtain with (4.1) that

$$\begin{aligned}
R_n(f,t) &= U_n(f,t) - \frac{[nt]}{n}\,U_n(f,1),\\
&= U_n(f,t) - tU_n(f,1) + \frac{nt-[nt]}{n}U_n(f,1) \quad \text{for all } t\in[0,1].
\end{aligned}\tag{4.2}$$

Let A_n denote the $\mathcal{F}\times[0,1]$-indexed processes given by $A_n(f,t) := \big((nt-[nt])/n\big)U_n(f,t)$. Since $\sup_{t\in[0,1]}|(nt-[nt])/n| \to 0$ as $n\to\infty$, by Slutsky's Theorem and the sequential empirical CLT, A_n converges in distribution (and thus in probability) to zero. Another application of Slutsky's theorem and the sequential empirical CLT on (4.2) yields

$$R_n = \big(U_n(f,t) - tU_n(f,1)\big)_{(f,t)\in\mathcal{F}\times[0,1]} + A_n \rightsquigarrow \big(K(f,t) - tK(f,1)\big)_{(f,t)\in\mathcal{F}\times[0,1]}.$$

Here we have applied the continuous mapping theorem in the final step. □

Remark 4.2. *Note that, in the setting of Theorem 3.5 and Theorem 3.8, the process K is a Kiefer process (that is the covariance structure is given by (2.5)).*

An application of the continuous mapping theorem with the supremum-functional to the above theorem yields the following proposition about the asymptotic distribution of the test statistic T_n defined by

$$T_n := \max_{0\le k\le n}\sup_{f\in\mathcal{F}} \frac{k}{n}\Big(1-\frac{k}{n}\Big)\sqrt{n}\big|F_k(f) - F_{k+1,n}(f)\big|.$$

Theorem 4.3. *If $(X_i)_{i\in\mathbb{N}^*}$ satisfies the sequential empirical CLT, then under the null hypothesis $\mathbf{H}_0$ we have the convergence*

$$T_n \rightsquigarrow \sup_{f\in\mathcal{F},\ t\in[0,1]} |K(f,t) - tK(f,1)|.$$

Proof. $R_n(f,\cdot)$ is obviously constant on the intervals $\big[k/n,(k+1)/n\big)$, $k=0,\ldots,n-1$ and further $R_n\big(f,k/n\big)=k/n(1-k/n)\sqrt{n}\big(F_k(f)-F_{k+1,n}(f)\big)$ for $k=0,\ldots,n$. Thus $T_n=\sup_{f\in\mathcal{F},t\in[0,1]}R_n(f,t)$ and we can apply the continuous mapping theorem with

$$\ell^\infty(\mathcal{F}\times[0,1])\longrightarrow\mathbb{R},\quad \varphi\mapsto\sup_{f\in\mathcal{F},\ t\in[0,1]}|\varphi(f,t)|.$$

□

5 Proof of Theorem 2.5

As proved in Dehling and Durieu [16], multiple mixing processes satisfy the following $2p$-th moment bound.

Assumption 5.1 (Moment bounds for $\mathcal{C}$-observables)**.** *There exist $p\in\mathbb{N}^*$, $s\geq 1$, and monotone increasing functions $\Phi_1,\ldots,\Phi_p:\mathbb{R}_+\longrightarrow\mathbb{R}_+$,*

$$\mathbf{E}\left[\left(\sum_{i=1}^n(f(X_i)-\mu f)\right)^{2p}\right]\leq\sum_{i=1}^p n^i\|f\|_s^i\Phi_i(\|f\|_{\mathcal{C}})\quad\textit{for all } f\in\mathcal{C}\textit{ with }\|f\|_\infty\leq 1.\tag{5.1}$$

We shall obtain Theorem 2.5 as a consequence of the more general following result.

Theorem 5.2. *Let $(X_i)_{i\in\mathbb{N}}$ be an $\mathcal{X}$-valued stationary process with marginal distribution μ and let $\mathcal{F}$ be a uniformly bounded class of measurable functions on $\mathcal{X}$. Suppose that for some normed vector space $\mathcal{C}$ of measurable functions on $\mathcal{X}$, Assumptions 2.3 and 5.1 hold. Moreover, assume that there exist a subset $\mathcal{G}$ of $\mathcal{C}$ which is bounded in $\|\cdot\|_\infty$-norm, a constant $r>-1$ and a monotone increasing function $\Psi:\mathbb{R}_+\longrightarrow\mathbb{R}_+$ such that*

$$\int_0^1\varepsilon^r\sup_{\varepsilon\leq\delta\leq 1}N^2\big(\delta,\Psi\left(\delta^{-1}\right),\mathcal{F},\mathcal{G},\mathrm{L}^s(\mu)\big)d\varepsilon<\infty.\tag{5.2}$$

If

$$\Phi_i(2\Psi(x))=O(x^{\gamma_i}),\tag{5.3}$$

for some non-negative constants γ_i such that

$$\gamma_i<2p-(i+r+2),\tag{5.4}$$

then the sequential empirical process U_n converges in distribution in $\ell^\infty(\mathcal{F}\times[0,1])$ to a tight Gaussian process K.

Proof of Theorem 2.5. Under multiple mixing (Assumption 2.4), Assumption 5.1 holds for all $p\geq 1$ and we can specify that $\Phi_i(x)=c\log^{2p+(d_0-1)i}(x+1)$ for some $c>0$ depending only on p, see Dehling and Durieu [16]. Observe that, choosing $\Psi:=\exp(C\,\mathrm{id}^{1/\gamma})$ for some $C>0$ and $\gamma>1$, we have $\Phi_i(2\Psi(x))=O(x^{(2p+(d_0-1)i)/\gamma})$. Therefore, the conditions (5.3) and (5.4) hold for sufficiently large $p\in\mathbb{N}^*$ as soon as $\gamma>d_0+1$. With this choice of Ψ, the condition (5.2) is exactly the condition (2.2). Thus Theorem 2.5 is a consequence of Theorem 5.2. □

The proof of Theorem 5.2 extends the idea introduced in Dehling et al. [17], taking into account the time parameter due to the sequential case. The main idea is to introduce some approximation $U_n^{(q)}$ for the original process U_n, which is based on functions in

$\mathcal{G}$ and thus can be controlled by Assumptions 2.3 and 5.1. The approximation can be constructed as follows: For all $q \geq 1$, there exist two sets of $N_q := N(2^{-q}, \Psi(2^q), \mathcal{F}, \mathcal{G}, \mathrm{L}^s(\mu))$ functions $\{g_{q,1}, \dots, g_{q,N_q}\} \subset \mathcal{G}$ and $\{g'_{q,1}, \dots, g'_{q,N_q}\} \subset \mathcal{G}$, such that

$$\|g_{q,i} - g'_{q,i}\|_s \leq 2^{-q}, \qquad \|g_{q,i}\|_{\mathcal{C}} \leq \Psi(2^q), \qquad \|g'_{q,i}\|_{\mathcal{C}} \leq \Psi(2^q) \tag{5.5}$$

and for all $f \in \mathcal{F}$, there exists some i such that $g_{q,i} \leq f \leq g'_{q,i}$. Further, by (5.2),

$$\sum_{q \geq 1} 2^{-(r+1)q} N_q^2 < \infty. \tag{5.6}$$

To approximate the indexing function $f \in \mathcal{F}$, construct a partition of $\mathcal{F}$ into N_q subsets $\mathcal{F}_{q,i}$ such that for each $f \in \mathcal{F}_{q,i}$ one has $g_{q,i} \leq f \leq g'_{q,i}$. We use the notation $\pi_q f = g_{q,i^*}$ and $\pi'_q f = g'_{q,i^*}$, where i^* is the uniquely defined integer such that $f \in \mathcal{F}_{q,i^*}$. To approximate the time parameter we use the partition of $[0,1]$ into subsets $\mathcal{T}_{q,j}$, $j = 1\dots, 2^q$, given by $\mathcal{T}_{q,j} := [(j-1)2^{-q}, j2^{-q})$ for $j < 2^q$ and $\mathcal{T}_{q,2^q} := [1-2^{-q}, 1]$. For $t \in [0,1]$ we define $\tau_q t := \max\{(j-1)2^{-q} \leq t : j = 1, \dots, 2^q\}$ and further $\tau'_q t := \tau_q t + 2^{-q}$. We extend the notation introduced in Section 2.1 to arbitrary μ-integrable functions $f : \mathcal{X} \longrightarrow \mathbb{R}$ by setting

$$F_n(f) := \frac{1}{n} \sum_{i=1}^{n} f(X_i)$$

and for $t \in [0,1]$

$$U_n(f,t) := \frac{[nt]}{\sqrt{n}} \big(F_{[nt]}(f) - \mu(f)\big) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} \big(f(X_i) - \mu(f)\big).$$

For each $q \geq 1$, we introduce the approximating process

$$U_n^{(q)}(f,t) := U_n(\pi_q f, \tau_q t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[n\tau_q t]} \left(\pi_q f(X_i) - \mu(\pi_q f)\right).$$

Note that this process is constant on each $\mathcal{F}_{q,i} \times \mathcal{T}_{q,j}$.

The approximating processes $U_n^{(q)}$ will help us to establish the weak convergence of the process U_n. Using Theorem 2.1 in Dehling et al. [17], we see that it is sufficient to show that there exist processes $U^{(q)} \in \ell^\infty(\mathcal{F} \times [0,1])$, $q \geq 1$, such that

$$U_n^{(q)} \rightsquigarrow U^{(q)} \quad \text{as } n \to \infty \text{ for all } q \geq 1, \tag{5.7}$$

and

$$\limsup_{n\to\infty} \mathrm{P}^*\big(\|U_n - U_n^{(q)}\|_\infty \geq \delta\big) \longrightarrow 0 \quad \text{as } q \to \infty \text{ for all } \delta > 0. \tag{5.8}$$

We will establish the conditions (5.7) and (5.8) in the two following propositions:

Proposition 5.3. *If Assumption 2.3 holds, then the process* $(U_n^{(q)}(f,t))_{(f,t)\in\mathcal{F}\times[0,1]}$ *converges for all* $q \in \mathbb{N}^*$ *in distribution to some piecewise constant Gaussian process* $(U^{(q)}(f,t))_{(f,t)\in\mathcal{F}\times[0,1]}$ *as* $n \to \infty$.

Proposition 5.4. *Assume that Assumption 5.1 holds for some* $p \in \mathbb{N}^*$, $s \geq 1$ *and some monotone increasing functions* $\Phi_1, \dots, \Phi_p : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$. *Moreover, suppose there exists a constant* $r > -1$ *and a monotone increasing function* $\Psi : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ *such that* (5.2) *holds. If* (5.3) *holds for some non-negative constants* γ_i *satisfying* (5.4), *then for all* $\varepsilon, \eta > 0$ *there exists some* q_0 *such that for all* $q \geq q_0$

$$\limsup_{n\to\infty} \mathrm{P}^*\left(\sup_{t\in[0,1]} \sup_{f\in\mathcal{F}} \left|U_n(f,t) - U_n^{(q)}(f,t)\right| > \varepsilon\right) \leq \eta.$$

Proof of Theorem 5.2. By Proposition 5.3 the convergence (5.7) holds, while (5.8) is satisfied due to Proposition 5.4. Therefore, by Theorem 2.1 in Dehling et al. [17], U_n converges in distribution to an $\ell^\infty(\mathcal{F}\times[0,1])$-valued, separable random variable K. Furthermore, we know that $U^{(q)}$ is a piecewise constant Gaussian process which converges in distribution to K. Thus K is Gaussian, too. Since $\ell^\infty(\mathcal{F}\times[0,1])$ is complete, the tightness of K follows from the separability (c.f. Lemma 1.3.2 in van der Vaart and Wellner [43]). □

Proof of Proposition 5.3. Since by construction $\pi_q f \in \mathcal{G}$ for all $f \in \mathcal{F}$, due to Assumption 2.3, the finite dimensional process $(U_n^{(q)}(f_1,t_1),\dots,U_n^{(q)}(f_k,t_k))$ converges in distribution for all fixed $k \in \mathbb{N}^*$ to some multi-dimensional normal distributed random variable $(U^{(q)}(f_1,t_1),\dots,U^{(q)}(f_k,t_k))$, $f_1,\dots,f_k \in \mathcal{F}$, $t_1,\dots,t_k \in [0,1]$. All $U_n^{(q)}$, $n \in \mathbb{N}^*$, are constant on each $\mathcal{F}_{q,i}\times\mathcal{T}_{q,j}$, $i=1,\dots,N^q$, $j=1,\dots,2^q$. Therefore $U^{(q)}$ is constant on all $\mathcal{F}_{q,i}\times\mathcal{T}_{q,j}$, too. Since these sets form a partition of $\mathcal{F}\times[0,1]$, the finite dimensional convergence yields the convergence in distribution of the whole process $(U_n^{(q)}(f,t))_{(f,t)\in\mathcal{F}\times[0,1]}$. □

Proof of Proposition 5.4. Let $\overline{Z} := Z - \mathbf{E}\,Z$ denote the centring of a random variable Z and observe that for any random variables $Y_l \le Y \le Y_u$ the inequality

$$|\overline{Y}-\overline{Y_l}| \le |\overline{Y_u}-\overline{Y_l}| + \mathbf{E}\,|Y_u - Y_l|$$

holds. Since for $f\in\mathcal{F}$, $k\in\mathbb{N}$ we have $F_{[nt]}(\pi_{q+k}f,t)\le F_{[nt]}(f,t)\le F_{[nt]}(\pi'_{q+k}f,t)$, using that $\|\cdot\|_1 \le \|\cdot\|_s$ for $s\ge 1$ and applying (5.5), we obtain

$$\begin{aligned}
&\big|U_n(f,t)-U_n(\pi_{q+k}f,t)\big| \\
&\quad\le \big|U_n(\pi'_{q+k}f,t)-U_n(\pi_{q+k}f,t)\big| + \frac{[nt]}{\sqrt{n}}\,\mathbf{E}\big|F_{[nt]}(\pi'_{q+k}f-\pi_{q+k}f)\big| \\
&\quad\le \big|U_n(\pi'_{q+k}f,t)-U_n(\pi_{q+k}f,t)\big| + \sqrt{n}2^{-(q+k)}. \qquad (5.9)
\end{aligned}$$

Moreover, for all $n\ge 2^{q+k}$ and $g\in\mathcal{G}$

$$\begin{aligned}
\big|U_n(g,t)-U_n(g,\tau_{q+k}t)\big| &= \frac{1}{\sqrt{n}}\Bigg|\sum_{i=[n\tau_{q+k}t]+1}^{[nt]} g(X_i)-\mu(g)\Bigg| \\
&\le 2Mn^{-\frac{1}{2}}([nt]-[n\tau_{q+k}t]) \\
&\le 4M\sqrt{n}2^{-(q+k)}, \qquad (5.10)
\end{aligned}$$

where $M := \sup\{\|g\|_\infty : g\in\mathcal{G}\}$ is finite by assumption. Analogously to the processes $U_n^{(q)}$, we introduce the processes $U_n'^{(q)}$ given by

$$U_n'^{(q)}(f,t) := U_n(\pi'_q f, \tau'_q t).$$

An application of the triangle inequality, (5.9), and (5.10) yields

$$\Big|U_n(f,t)-U_n^{(q+k)}(f,t)\Big| \le \Big|U_n'^{(q+k)}(f,t)-U_n^{(q+k)}(f,t)\Big| + (4M+1)\sqrt{n}2^{-q+k}. \qquad (5.11)$$

Combining (5.11) with a telescopic sum argument, one obtains for any $K \geq 1$

$$\begin{aligned}
&\left|U_n(f,t) - U_n^{(q)}(f,t)\right| \\
&= \left|\left\{\sum_{k=1}^{K} U_n^{(q+k)}(f,t) - U_n^{(q+k-1)}(f,t)\right\} + U_n(f,t) - U_n^{(q+K)}(f,t)\right| \\
&\leq \left\{\sum_{k=1}^{K} \left|U_n^{(q+k)}(f,t) - U_n^{(q+k-1)}(f,t)\right|\right\} + \left|U_n'^{(q+K)}(f,t) - U_n^{(q+K)}(f,t)\right| \\
&\quad + (4M+1)\sqrt{n}2^{-(q+K)}. \qquad (5.12)
\end{aligned}$$

To assure $\varepsilon/4 \leq (4M+1)\sqrt{n}2^{-(q+K)} \leq \varepsilon/2$, choose $K = K_{n,q}$, given by

$$K_{n,q} := \left[\log_2\left(\frac{4(4M+1)\sqrt{n}}{2^q \varepsilon}\right)\right].$$

For each $i = 1, \ldots, N_q$, $j = 1, \ldots, 2^q$, inequality (5.12) implies

$$\begin{aligned}
\sup_{t\in\mathcal{T}_{q,j}} \sup_{f\in\mathcal{F}_{q,i}} |U_n(f,t) - U_n^{(q)}(f,t)| \leq &\left\{\sum_{k=1}^{K_{n,q}} \sup_{t\in\mathcal{T}_{q,j}} \sup_{f\in\mathcal{F}_{q,i}} \left|U_n^{(q+k)}(f,t) - U_n^{(q+k-1)}(f,t)\right|\right\} \\
&+ \sup_{t\in\mathcal{T}_{q,j}} \sup_{f\in\mathcal{F}_{q,i}} \left|U_n'^{(q+K)}(f,t) - U_n^{(q+K)}(f,t)\right| + \frac{\varepsilon}{2}.
\end{aligned}$$

Set $\varepsilon_k = \varepsilon/(4k(k+1))$. Then $\sum_{i=1}^{\infty} \varepsilon_k = \varepsilon/4$ and for all $i = 1, \ldots, N_q$ we have

$$\begin{aligned}
&\mathrm{P}^*\left(\sup_{t\in\mathcal{T}_{q,j}} \sup_{f\in\mathcal{F}_{q,i}} |U_n(f,t) - U_n^{(q)}(f,t)| \geq \varepsilon\right) \\
&\leq \left\{\sum_{k=1}^{K_{n,q}} \mathrm{P}^*\left(\sup_{t\in\mathcal{T}_{q,j}} \sup_{f\in\mathcal{F}_{q,i}} \left|U_n^{(q+k)}(f,t) - U_n^{(q+k-1)}(f,t)\right| \geq \varepsilon_k\right)\right\} \\
&\quad + \mathrm{P}^*\left(\sup_{t\in\mathcal{T}_{q,j}} \sup_{f\in\mathcal{F}_{q,i}} \left|U_n'^{(q+K)}(f,t) - U_n^{(q+K)}(f,t)\right| \geq \frac{\varepsilon}{4}\right). \qquad (5.13)
\end{aligned}$$

Recall that (π_{q+k}, τ_{q+k}) and thus $U_n^{(q+k)}$ and $U_n'^{(q+k)}$ are constant on each $\mathcal{F}_{q+k,i} \times \mathcal{T}_{q+k,j}$, $i = 1, \ldots N_{q+k}$, $j = 1, \ldots, 2^{q+k}$, and thus the suprema on the r.h.s. of inequality (5.13) are in fact maxima over finite numbers of functions. Therefore the outer probabilities may be replaced by usual probabilities here. Now, for each $k \in \mathbb{N}^*$, choose a set $\mathcal{F}(k)$ of at most $N_{k-1}N_k$ functions in $\mathcal{F}$, such that $\mathcal{F}(k)$ contains at least one function in each non empty $\mathcal{F}_{k,i} \cap \mathcal{F}_{k-1,i'}$, $i = 1, \ldots, N_k$, $i' = 1, \ldots, N_{k-1}$. For $q \in \mathbb{N}^*$ and $i \in \{1, \ldots, N_q\}$, define

$$\begin{aligned}
F_{k,q,i} &:= \mathcal{F}_{q,i} \cap \mathcal{F}(q+k) \\
T_{k,q,j} &:= \left\{(j-1)2^{-q} + (m-1)2^{-(q+k)} : m \in \{1, \ldots, 2^k\}\right\}.
\end{aligned}$$

Inequality (5.13) implies

$$\begin{aligned}
&\mathrm{P}^*\Big(\sup_{t\in\mathcal{T}_{q,j}}\sup_{f\in\mathcal{F}_{q,i}}|U_n(f,t)-U_n^{(q)}(f,t)|\geq\varepsilon\Big)\\
&\quad\leq\left\{\sum_{k=1}^{K_{n,q}}\sum_{t\in T_{k,q,j}}\sum_{f\in F_{k,q,i}}\mathrm{P}\Big(\Big|U_n^{(q+k)}(f,t)-U_n^{(q+k-1)}(f,t)\Big|\geq\varepsilon_k\Big)\right\}\\
&\qquad+\sum_{t\in T_{K_{n,q},q,j}}\sum_{f\in F_{K_{n,q},q,i}}\mathrm{P}\Big(\Big|U_n'^{(q+K_{n,q})}(f,t)-U_n^{(q+K_{n,q})}(f,t)\Big|\geq\frac{\varepsilon}{4}\Big)\\
&\quad\leq\left\{\sum_{k=1}^{K_{n,q}}\sum_{t\in T_{k,q,j}}\sum_{f\in F_{k,q,i}}\mathrm{P}\Big(\Big|U_n(\pi_{q+k}f,\tau_{q+k-1}t)-U_n(\pi_{q+k-1}f,\tau_{q+k-1}t)\Big|\geq\frac{\varepsilon_k}{2}\Big)\right.\\
&\qquad\quad\left.+\mathrm{P}\Big(\Big|U_n(\pi_{q+k}f,\tau_{q+k}t)-U_n(\pi_{q+k}f,\tau_{q+k-1}t)\Big|\geq\frac{\varepsilon_k}{2}\Big)\right\}\\
&\qquad+\sum_{t\in T_{K_{n,q},q,j}}\sum_{f\in F_{K_{n,q},q,i}}\mathrm{P}\Big(\Big|U_n(\pi'_{q+K_{n,q}}f,\tau_{q+K_{n,q}}t)-U_n(\pi_{q+K_{n,q}}f,\tau_{q+K_{n,q}}t)\Big|\geq\frac{\varepsilon}{8}\Big)\\
&\qquad\quad+\mathrm{P}\Big(\Big|U_n(\pi'_{q+K_{n,q}}f,\tau'_{q+K_{n,q}}t)-U_n(\pi'_{q+K_{n,q}}f,\tau_{q+K_{n,q}}t)\Big|\geq\frac{\varepsilon}{8}\Big).
\end{aligned}$$

Applying Markov's inequality on the $2p$-th moments, we obtain

$$\begin{aligned}
&\mathrm{P}^*\Big(\sup_{t\in\mathcal{T}_{q,j}}\sup_{f\in\mathcal{F}_{q,i}}|U_n(f,t)-U_n^{(q)}(f,t)|\geq\varepsilon\Big)\\
&\quad\leq\left\{\sum_{k=1}^{K_{n,q}}\sum_{t\in T_{k,q,j}}\sum_{f\in F_{k,q,i}}\Big(\frac{\varepsilon_k}{2}\Big)^{-2p}\Big(\mathbf{E}\Big|U_n(\pi_{q+k}f,\tau_{q+k-1}t)-U_n(\pi_{q+k-1}f,\tau_{q+k-1}t)\Big|^{2p}\right.\\
&\qquad\quad\left.+\mathbf{E}\Big|U_n(\pi_{q+k}f,\tau_{q+k}t)-U_n(\pi_{q+k}f,\tau_{q+k-1}t)\Big|^{2p}\Big)\right\}\\
&\qquad+\sum_{t\in T_{K_{n,q},q,j}}\sum_{f\in F_{K_{n,q},q,i}}\Big(\frac{\varepsilon}{8}\Big)^{-2p}\Big(\mathbf{E}\Big|U_n(\pi'_{q+K_{n,q}}f,\tau_{q+K_{n,q}}t)-U_n(\pi_{q+K_{n,q}}f,\tau_{q+K_{n,q}}t)\Big|^{2p}\\
&\qquad\quad+\mathbf{E}\Big|U_n(\pi'_{q+K_{n,q}}f,\tau'_{q+K_{n,q}}t)-U_n(\pi'_{q+K_{n,q}}f,\tau_{q+K_{n,q}}t)\Big|^{2p}\Big).
\end{aligned}\tag{5.14}$$

We will treat the expected values on the r.h.s. of inequality (5.14) separately now by using Assumption 5.1 and properties of our brackets used to cover $\mathcal{F}$. Recall that by (5.5) we have

$$\begin{aligned}
\|\pi_{q+k}f-\pi_{q+k-1}f\|_s&\leq\|\pi_{q+k}f-f\|_s+\|\pi_{q+k-1}f-f\|_s\leq 3\cdot 2^{-(q+k)}\\
\|\pi_{q+k}f-\pi'_{q+k}f\|_s&\leq 2^{-(q+k)}
\end{aligned}\tag{5.15}$$

$$\begin{aligned}
\|\pi_{q+k}f-\pi_{q+k-1}f\|_{\mathcal{C}}&\leq 2\Psi(2^{q+k})\\
\|\pi_{q+k}f-\pi'_{q+k}f\|_{\mathcal{C}}&\leq 2\Psi(2^{q+k}).
\end{aligned}\tag{5.16}$$

For convenience, throughout the rest of the proof will write $x\ll y$ if there is some finite constant $C\in(0,\infty)$ such that $x\leq Cy$, where C may only depend on global parameters of the corresponding statement. Applying successively (5.1), (5.15), (5.16), and (5.3) we have

$$\begin{aligned}
&\mathbf{E}\big|U_n(\pi_{q+k}f,\tau_{q+k-1}t)-U_n(\pi_{q+k-1}f,\tau_{q+k-1}t)\big|^{2p}\\
&\quad\ll n^{-p}\sum_{\ell=1}^{p} n^{\ell}\|\pi_{q+k}f-\pi_{q+k-1}f\|_s^{\ell}\Phi_\ell(\|\pi_{q+k}f-\pi_{q+k-1}f\|_{\mathcal{C}})\\
&\quad\ll \sum_{\ell=1}^{p} n^{-(p-\ell)}2^{(\gamma_\ell-\ell)(q+k)} \qquad (5.17)
\end{aligned}$$

and analogously

$$\mathbf{E}\big|U_n(\pi'_{q+K_{n,q}}f,\tau_{q+K_{n,q}}t)-U_n(\pi_{q+K_{n,q}}f,\tau_{q+K_{n,q}}t)\big|^{2p}\ll\sum_{\ell=1}^{p} n^{-(p-\ell)}2^{(\gamma_\ell-\ell)(q+K_{n,q})}. \qquad (5.18)$$

For fixed $g\in\mathcal{G}$ we have by stationarity

$$\mathbf{E}\big|U_n(g,\tau_{q+k}t)-U_n(g,\tau_{q+k-1}t)\big|^{2p}=n^{-p}\,\mathbf{E}\left[\left(\sum_{i=1}^{[n\tau_{q+k}t]-[n\tau_{q+k-1}t]}\big(g(X_i)-\mu g\big)\right)^{2p}\right], \qquad (5.19)$$

where we consider $\sum_{i=1}^{0}\ldots=0$. Note that by construction $\tau_{q+k}t-\tau_{q+k-1}t\in\{0,2^{-(q+k)}\}$ for every $t\in[0,1]$ and therefore

$$[n\tau_{q+k}t]-[n\tau_{q+k-1}t]\le n2^{-(q+k)}+1\quad\text{for all } n\ge 2^{q+k}.$$

Applying (5.1), (5.5), and (5.3) to (5.19) we obtain

$$\begin{aligned}
\mathbf{E}\big|U_n(\pi_{q+k}f,\tau_{q+k}t)-U_n(\pi_{q+k}f,\tau_{q+k-1}t)\big|^{2p}&\ll n^{-p}\sum_{\ell=1}^{p}\big(n2^{-(q+k)}\big)^{\ell}\|\pi_{q+k}f\|_s^{\ell}\Phi_\ell(\|\pi_{q+k}f\|_{\mathcal{C}})\\
&\ll\sum_{\ell=1}^{p} n^{-(p-\ell)}2^{(\gamma_\ell-\ell)(q+k)} \qquad (5.20)
\end{aligned}$$

and analogously

$$\mathbf{E}\big|U_n(\pi'_{q+K_{n,q}}f,\tau'_{q+K_{n,q}}t)-U_n(\pi'_{q+K_{n,q}}f,\tau_{q+K_{n,q}}t)\big|^{2p}\ll\sum_{\ell=1}^{p} n^{-(p-\ell)}2^{(\gamma_\ell-\ell)(q+K_{n,q})}. \qquad (5.21)$$

Now, apply (5.17), (5.18), (5.20), and (5.21) to (5.14). We infer

$$\begin{aligned}
&\mathrm{P}^*\left(\sup_{t\in\mathcal{T}_{q,j}}\sup_{f\in\mathcal{F}_{q,i}}\Big|U_n(f,t)-U_n^{(q)}(f,t)\Big|\ge\varepsilon\right)\\
&\quad\ll\sum_{k=1}^{K_{n,q}}\#T_{k,q,j}\,\#F_{k,q,i}\,\frac{(k(k+1))^{2p}}{\varepsilon^{2p}}\sum_{\ell=1}^{p} n^{-(p-\ell)}2^{(\gamma_\ell-\ell)(q+k)}. \qquad (5.22)
\end{aligned}$$

Recall that by construction of the partitions of $\mathcal{F}$ and $[0,1]$ at the beginning of this section, we have $\sum_{j=1}^{2^q}\#T_{k,q,j}=2^{q+k}$ and $\sum_{i=1}^{N_q}\#F_{k,q,i}=\#\mathcal{F}(q+k)\le N_{q+k-1}N_{q+k}$.

Therefore (5.22) yields

$$\begin{aligned}
&\mathrm{P}^*\left(\sup_{t\in[0,1]}\sup_{f\in\mathcal{F}}\left|U_n(f,t)-U_n^{(q)}(f,t)\right|>\varepsilon\right)\\
&\quad\ll\sum_{\ell=1}^{p}\sum_{k=1}^{K_{n,q}}\sum_{j=1}^{2^q}\#T_{k,q,j}\sum_{i=1}^{N_q}\#F_{k,q,i}k^{4p}n^{-(p-\ell)}2^{(\gamma_\ell-\ell)(q+k)}\\
&\quad\ll\sum_{\ell=1}^{p}\sum_{k=1}^{K_{n,q}}N_{q+k-1}N_{q+k}k^{4p}n^{-(p-\ell)}2^{(\gamma_\ell-\ell+1)(q+k)}.
\end{aligned}$$

This implies that for any $\eta>0$

$$\begin{aligned}
&\mathrm{P}^*\left(\sup_{t\in[0,1]}\sup_{f\in\mathcal{F}}\left|U_n(f,t)-U_n^{(q)}(f,t)\right|>\varepsilon\right)\\
&\ll\sum_{\ell=1}^{p}n^{-(p-\ell)}\max\left\{1\ ,\ 2^{(\gamma_\ell-\ell+r+2+\eta)(q+K_{n,q})}\right\}\sum_{k=1}^{K_{n,q}}N_{q+k-1}N_{q+k}k^{4p}2^{-(r+1+\eta)(q+k)}\\
&\ll\max\left\{1\ ,\ \max_{\ell=1,\dots,p}n^{\frac{1}{2}(\gamma_\ell+\ell-2p+r+2+\eta)}\right\}\sum_{k=q+1}^{\infty}N_{k-1}N_kk^{4p}2^{-(r+1+\eta)k}. \qquad (5.23)
\end{aligned}$$

By (5.4) we can choose η small enough to assure $\gamma_\ell+\ell-2p+r+2+\eta<0$ for all $\ell=1,\dots,p$. Thus the factor in front of the sum is uniformly bounded w.r.t. n. Using (5.6), we obtain

$$\sum_{k=1}^{\infty}N_{k-1}N_kk^{4p}2^{-(r+1+\eta)k}\le\sum_{k=1}^{\infty}2^{-(r+1)k}N_{k-1}^2\,k^{4p}2^{-\eta k}+\sum_{k=1}^{\infty}2^{-(r+1)k}N_k^2\,k^{4p}2^{-\eta k}<\infty$$

for sufficiently small $\eta>0$ which implies that the series in (5.23) goes to zero as $q\to\infty$. □

6 Proof of Corollary 2.6

In order to simplify the expressions, set $\Psi(x)=\exp(Cx^{1/\gamma})$, where C and γ are given by Theorem 2.5. Choose $b\in(1,\gamma)$ and observe that,

$$\sum_{k=1}^{\infty}\Psi(k^b)\theta^k<\infty. \qquad (6.1)$$

For $f\in\mathcal{F}$, recall the definition of the approximating functions π_qf from Section 5 and note that, as a consequence of the entropy condition in Theorem 2.5, we know that for every $q\in\mathbb{N}^*$,

$$\|f-\pi_qf\|_s\le 2^{-q} \qquad (6.2)$$

$$\|\pi_qf\|_{\mathcal{C}}\le\Psi(2^q), \qquad (6.3)$$

where $s\ge1$ is given in the assumptions of theorem 2.5. Similarly, for all $g\in\mathcal{F}$ and $k\in\mathbb{N}^*$ there exist some $g_k\in\mathcal{G}$ satisfying

$$\|g_k-g\|_s\le k^{-b} \qquad (6.4)$$

$$\|g_k\|_{\mathcal{C}}\le\Psi(k^b). \qquad (6.5)$$

Let $U^{(q)}$ denote the limit process given in Proposition 5.3. Condition (i) implies that for all $f, g \in \mathcal{F}$, $t, u \in [0,1]$ and $q \in \mathbb{N}^*$

$$\begin{aligned}
&\mathbf{Cov}\big(U^{(q)}(f,t), U^{(q)}(g,u)\big) \\
&\quad = \min\{t,u\}\left\{\sum_{k=0}^{\infty}\mathbf{Cov}\big(\pi_q f(X_0), \pi_q g(X_k)\big) + \sum_{k=1}^{\infty}\mathbf{Cov}\big(\pi_q g(X_0), \pi_q f(X_k)\big)\right\}.
\end{aligned}$$

Since the auto-covariance functions of a converging Gaussian process converge to the auto-covariance functions of the limit process, the covariance structure of the limit process K of $U^{(q)}$ is given by $\mathbf{Cov}(K(f,t), K(g,u)) = \lim_{q\to\infty}\mathbf{Cov}(U^{(q)}(f,t), U^{(q)}(g,u))$. Thus it suffices to show that

$$\begin{aligned}
&\Big|\sum_{k=0}^{\infty}\mathbf{Cov}\big(\pi_q f(X_0), \pi_q g(X_k)\big) - \mathbf{Cov}\big(f(X_0), g(X_k)\big)\Big| \qquad (6.6)\\
&\quad + \Big|\sum_{k=1}^{\infty}\mathbf{Cov}\big(\pi_q g(X_0), \pi_q f(X_k)\big) - \mathbf{Cov}\big(g(X_0), f(X_k)\big)\Big| \longrightarrow 0 \quad \text{as } q \to \infty.
\end{aligned}$$

By symmetry, both series can be treated the same way. Let $k(q) := 2^{q/b}$. We consider the series in line (6.6). We have

$$\begin{aligned}
&\Big|\sum_{k=0}^{\infty}\mathbf{Cov}\big(\pi_q f(X_0), \pi_q g(X_k)\big) - \mathbf{Cov}\big(f(X_0), g(X_k)\big)\Big| \\
&\quad \le \sum_{k=0}^{k(q)}\big|\mathbf{Cov}\big(\pi_q f(X_0) - f(X_0), \pi_q g(X_k)\big)\big| + \sum_{k=0}^{k(q)}\big|\mathbf{Cov}\big(f(X_0), \pi_q g(X_k) - g(X_k)\big)\big| \qquad (6.7)\\
&\qquad + \sum_{k=k(q)+1}^{\infty}\big|\mathbf{Cov}\big(\pi_q f(X_0) - f(X_0), \pi_q g(X_k)\big)\big| \qquad (6.8)\\
&\qquad + \sum_{k=k(q)+1}^{\infty}\big|\mathbf{Cov}\big(f(X_0), \pi_q g(X_k) - g(X_k)\big)\big|. \qquad (6.9)
\end{aligned}$$

Let us treat the terms separately. Recall that both $\mathcal{F}$ and $\mathcal{G}$ are uniformly bounded in $\|\cdot\|_\infty$-norm. For the term in line (6.7), we know by Hölder's inequality, (6.2), and the fact that $b > 1$ that

$$\begin{aligned}
&\sum_{k=0}^{k(q)}\big|\mathbf{Cov}\big(\pi_q f(X_0) - f(X_0), \pi_q g(X_k)\big)\big| + \sum_{k=0}^{k(q)}\big|\mathbf{Cov}\big(f(X_0), \pi_q g(X_k) - g(X_k)\big)\big| \\
&\quad \ll \sum_{k=0}^{k(q)}\big(\|\pi_q f - f\|_s + \|\pi_q g - g\|_s\big) \\
&\quad \ll k(q)2^{-q} = 2^{-(1-\frac{1}{b})q} \longrightarrow 0 \quad \text{as } q \to \infty,
\end{aligned}$$

where again, we write $x \ll y$ if there is a constant $C \in (0,\infty)$ depending only on global parameters such that $x \le Cy$. For the term in line (6.8), by (2.4), (6.2), and (6.3) we obtain

$$\begin{aligned}
&\sum_{k=k(q)+1}^{\infty} \left|\mathbf{Cov}\big(\pi_q f(X_0) - f(X_0), \pi_q g(X_k)\big)\right| \\
&\leq D\|\pi_q f - f\|_\infty \sum_{k=k(q)+1}^{\infty} \|\pi_q g\|_{\mathcal{C}}\, \theta^k \\
&\ll \sum_{k=k(q)+1}^{\infty} \Psi(2^q)\theta^k \longrightarrow 0 \quad \text{as } q \to \infty,
\end{aligned}$$

where we used that Ψ is increasing and condition (6.1) in the last step. It only remains to show, that the term in line (6.9) goes to zero as $q \to \infty$. We have

$$\begin{aligned}
&\sum_{k=k(q)+1}^{\infty} \left|\mathbf{Cov}\big(f(X_0), \pi_q g(X_k) - g(X_k)\big)\right| \\
&\leq \sum_{k=k(q)+1}^{\infty} \left|\mathbf{Cov}\big(f(X_0), \pi_q g(X_k) - g_k(X_k)\big)\right| \qquad (6.10)\\
&\quad + \sum_{k=k(q)+1}^{\infty} \left|\mathbf{Cov}\big(f(X_0), g_k(X_k) - g(X_k)\big)\right|. \qquad (6.11)
\end{aligned}$$

First, consider the term in line (6.10). By (2.4), (6.3), and (6.5)

$$\begin{aligned}
&\sum_{k=k(q)+1}^{\infty} \left|\mathbf{Cov}\big(f(X_0), \pi_q g(X_k) - g_k(X_k)\big)\right| \\
&\ll \sum_{k=k(q)+1}^{\infty} \|f\|_\infty \, \|\pi_q g - g_k\|_{\mathcal{C}}\, \theta^k \\
&\ll \left(\sum_{k=k(q)+1}^{\infty} \|\pi_q g\|_{\mathcal{C}}\, \theta^k\right) + \left(\sum_{k=k(q)+1}^{\infty} \|g_k\|_{\mathcal{C}}\, \theta^k\right) \\
&\ll \left(\sum_{k=k(q)+1}^{\infty} \Psi(2^q)\theta^k\right) + \left(\sum_{k=k(q)+1}^{\infty} \Psi(k^b)\theta^k\right) \longrightarrow 0 \quad \text{as } q \to \infty,
\end{aligned}$$

where we used that Ψ is increasing and applied condition (6.1) in the last line. To treat the term in line (6.11), we use Hölder's inequality and (6.4). We obtain

$$\begin{aligned}
\sum_{k=k(q)+1}^{\infty} \left|\mathbf{Cov}\big(f(X_0), g_k(X_k) - g(X_k)\big)\right| &\ll \sum_{k=k(q)+1}^{\infty} \|g_k - g\|_s \\
&\ll \sum_{k=k(q)+1}^{\infty} k^{-b} \longrightarrow 0 \quad \text{as } q \to \infty,
\end{aligned}$$

since $b > 1$ and thus $\sum_{k=1}^{\infty} k^{-b} < \infty$, which completes the proof. □

References

[1] Michael F. Barnsley, Stephen G. Demko, John H. Elton, and Jeffrey S. Geronimo, *Invariant measures for Markov processes arising from iterated function systems with place-dependent probabilities*, Ann. Inst. H. Poincaré Probab. Statist. **24** (1988), no. 3, 367–394. MR-0971099

[2] Michael F. Barnsley and John H. Elton, *A new class of Markov processes for image encoding*, Adv. in Appl. Probab. **20** (1988), no. 1, 14–32. MR-0932532

[3] István Berkes, Siegfried Hörmann, and Johannes Schauer, *Asymptotic results for the empirical process of stationary sequences*, Stochastic Process. Appl. **119** (2009), no. 4, 1298–1324. MR-2508575

[4] István Berkes and Walter Philipp, *An almost sure invariance principle for the empirical distribution function of mixing random variables*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **41** (1977), no. 2, 115–137. MR-0464344

[5] Patrick Billingsley, *Convergence of probability measures*, John Wiley & Sons Inc., New York, 1968. MR-0233396

[6] Zbigniew Ciesielski and Harry Kesten, *A limit theorem for the fractional parts of the sequence* $\{2^k t\}$, Proc. Amer. Math. Soc. **13** (1962), 596–600. MR-0138612

[7] Pierre Collet, Servet Martinez, and Bernard Schmitt, *Asymptotic distribution of tests for expanding maps of the interval*, Ergodic Theory Dynam. Systems **24** (2004), no. 3, 707–722. MR-2060995

[8] Miklós Csörgő and Pál Révész, *A new method to prove strassen type laws of invariance principle. II*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **31** (1975), no. 4, 261–269. MR-1554018

[9] ———, *Strong approximations in probability and statistics*, Probability and Mathematical Statistics, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1981. MR-0666546

[10] Jérôme Dedecker, Paul Doukhan, Gabriel Lang, José Rafael León, Sana Louhichi, and Clémentine Prieur, *Weak dependence: with examples and applications*, Lecture Notes in Statistics, vol. 190, Springer, New York, 2007. MR-2338725

[11] Jérôme Dedecker and Florence Merlevède, *The conditional central limit theorem in Hilbert spaces*, Stochastic Process. Appl. **108** (2003), no. 2, 229–262. MR-2019054

[12] Jérôme Dedecker, Florence Merlevède, and Françoise Pène, *Empirical central limit theorems for ergodic automorphisms of the torus*, ALEA - Lat. Am. J. Probab. Math. Stat. **10** (2013), no. 2, 731–766. MR-3125745

[13] Jérôme Dedecker, Florence Merlevède, and Emmanuel Rio, *Strong approximation results for the empirical process of stationary sequences*, Ann. Probab. **41** (2013), no. 5, 3051–3696. MR-3127895

[14] ———, *Strong approximation of the empirical distribution function for absolutely regular sequences in* $\mathbb{R}^d$, Electron. J. Probab. **19** (2014), no. 9, 1–56. MR-3164762

[15] Jérôme Dedecker and Clémentine Prieur, *An empirical central limit theorem for dependent sequences*, Stochastic Process. Appl. **117** (2007), no. 1, 121–142. MR-2287106

[16] Herold Dehling and Olivier Durieu, *Empirical processes of multidimensional systems with multiple mixing properties*, Stochastic Process. Appl. **121** (2011), no. 5, 1076–1096. MR-2775107

[17] Herold Dehling, Olivier Durieu, and Marco Tusche, *Approximating class approach for empirical processes of dependent sequences indexed by functions*, Bernoulli **20** (2014), no. 3, 1372–1403. MR-3217447

[18] Herold Dehling, Olivier Durieu, and Dalibor Volný, *New techniques for empirical processes of dependent data*, Stochastic Process. Appl. **119** (2009), no. 10, 3699–3718. MR-2568292

[19] Herold Dehling and Walter Philipp, *Empirical process techniques for dependent data*, Empirical process techniques for dependent data, Birkhäuser Boston, Boston, MA, 2002, pp. 3–113. MR-1958777

[20] Herold Dehling and Murad S. Taqqu, *The empirical process of some long-range*

dependent sequences with an application to U-statistics, Ann. Statist. **17** (1989), no. 4, 1767–1783. MR-1026312

[21] Wolfgang Döblin and Robert Fortet, *Sur des chaînes à liaisons complètes*, Bull. Soc. Math. France **65** (1937), 132–148. MR-1505076

[22] Monroe D. Donsker, *Justification and extension of Doob's heuristic approach to the Komogorov-Smirnov theorems*, Ann. Math. Statistics **23** (1952), 277–281. MR-0047288

[23] Joseph L. Doob, *Heuristic approach to the Kolmogorov-Smirnov theorems*, Ann. Math. Statistics **20** (1949), 393–403. MR-0030732

[24] Lester E. Dubins and David A. Freedman, *Invariant probabilities for certain Markov processes*, Ann. Math. Statist. **37** (1966), 837–848. MR-0193668

[25] Olivier Durieu, *Empirical processes of iterated maps that contract on average*, Stat. Probabil. Lett. **83** (2013), no. 11, 2454–2458. MR-3144025

[26] Olivier Durieu and Marco Tusche, *An empirical process central limit theorem for multidimensional dependent data*, J. Theoret. Probab. **27** (2014), 249–277. MR-3174225

[27] Peter Gänssler and Winfried Stute, *Empirical processes: a survey of results for independent and identically distributed random variables*, Ann. Probab. **7** (1979), no. 2, 193–243. MR-0525051

[28] Sébastien Gouëzel, *An interval map with a spectral gap on Lipschitz functions, but not on bounded variation functions*, Discrete Contin. Dyn. Syst. **24** (2009), no. 4, 1205–1208. MR-2505699

[29] Hubert Hennion and Loïc Hervé, *Limit theorems for Markov chains and stochastic properties of dynamical systems by quasi-compactness*, Lecture Notes in Mathematics, vol. 1766, Springer-Verlag, Berlin, 2001. MR-1862393

[30] Loïc Hervé, *Vitesse de convergence dans le théorème limite central pour des chaînes de Markov fortement ergodiques*, Ann. Inst. Henri Poincaré Probab. Stat. **44** (2008), no. 2, 280–292. MR-2446324

[31] Loïc Hervé and Françoise Pène, *The Nagaev-Guivarc'h method via the Keller-Liverani theorem*, Bull. Soc. Math. France **138** (2010), no. 3, 415–489. MR-2729019

[32] Samuel Karlin, *Some random walks arising in learning models. I*, Pacific J. Math. **3** (1953), 725–756. MR-0058910

[33] Jack C. Kiefer, *Skorohod embedding of multivariate RV's, and the sample DF*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **24** (1972), 1–35. MR-1554013

[34] János Komlós, Péter Major, and Gábor Tusnády, *An approximation of partial sums of independent* RV*'s and the sample* DF*. I*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **32** (1975), 111–131. MR-0375412

[35] Andrzej Lasota and James A. Yorke, *On the existence of invariant measures for piecewise monotonic transformations*, Trans. Amer. Math. Soc. **186** (1973), 481–488. MR-0335758

[36] Sean P. Meyn and Richard L. Tweedie, *Markov chains and stochastic stability*, Communications and Control Engineering Series, Springer-Verlag London Ltd., London, 1993. MR-1287609

[37] Dietrich Werner Müller, *On Glivenko-Cantelli convergence*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **16** (1970), 195–210. MR-0388616

[38] Mina Ossiander, *A central limit theorem under metric entropy with L_2 bracketing*, Ann. Probab. **15** (1987), no. 3, 897–919. MR-0893905

[39] Marc Peigné, *Iterated function systems and spectral decomposition of the associated Markov operator*, Fascicule de probabilités, Publ. Inst. Rech. Math. Rennes, vol. 1993, Univ. Rennes I, Rennes, 1993, p. 28. MR-1347702

[40] Walter Philipp and Laurence Pinzur, *Almost sure approximation theorems for the multivariate empirical process*, Z. Wahrsch. Verw. Gebiete **54** (1980), no. 1, 1–13. MR-0595473

[41] Mark Pollicott, *Contraction in mean and transfer operators*, Dyn. Syst. **16** (2001), no. 1, 97–106. MR-1835908

[42] Marco Tusche, *Empirical processes of multiple mixing data*, Ph.D. thesis, Ruhr-Universität Bochum, Université François-Rabelais de Tours, 2014.

[43] Aad W. van der Vaart and Jon A. Wellner, *Weak convergence and empirical processes*, Springer-Verlag, New York, 1996. MR-1385671

[44] Charles P. Walkden, *Invariance principles for iterated maps that contract on average*, Trans. Amer. Math. Soc. **359** (2007), no. 3, 1081–1097 (electronic). MR-2262842

[45] Wei Biao Wu and Xiaofeng Shao, *Limit theorems for iterated random functions*, J. Appl. Probab. **41** (2004), no. 2, 425–436. MR-2052582

Acknowledgments. The authors thank the referee for her/his very careful reading of an earlier version of this paper, and for many thoughtful comments. The suggestions made by the referee helped to improve the presentation of the paper.