A Sequential Empirical CLT for Multiple Mixing Processes with Application to $\beta$-Geometrically Ergodic Markov Chains

Herold Dehling† Olivier Durieu‡ Marco Tusche‡

February 6, 2014

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.

MSC: 60F05, 60F17, 60G10, 62G30, 60J05

Keywords: Multivariate sequential empirical processes, Limit theorems, Multiple mixing, Spectral gap, Dynamical systems, Markov chain, Change-point problems

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 (1952). This theorem, conjectured by Doob (1949), 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 $\omega^2$ criterion, and more generally von Mises Statistics.

*Research supported by the German Science Foundation, Grant DE 370-4 (New Techniques for Empirical Processes of Dependent Data).
†Fakultät für Mathematik, Ruhr-Universität Bochum, Universitätsstrae 150, 44780 Bochum, Germany
‡Laboratoire de Mathématiques et Physique Théorique, UMR-CNRS 7350, Fédération Denis Poisson FR-CNRS 2964, Université François Rabelais de Tours, Parc de Grandmont, 37200 Tours, France
Ciesielski and Kesten (1962) 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 (1968) 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 processes of dependent data, see Dehling and Philipp (2002).

Müller (1970), and independently Kiefer (1972), initiated the study of the sequential empirical process, defined as

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

where $F(x) = P(X_1 \leq x)$. 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

$$E(K(x, s)K(y, t)) = \min(s, t)(F(\min(x, y)) - F(x)F(y)).$$

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

Komlós, Major, and Tusnády (1975), using a technique due to Csörgő and Révész (1975), 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 (1981) or the survey article by Gänssler and Stute (1979).

Many authors have studied extensions of the sequential empirical process CLT to dependent data, e.g. Berkes and Philipp (1977) and Philipp and Pinzur (1980) for strongly mixing processes and Berkes, Hörmann, and Schauer (2009) for $S$-mixing processes. Dehling and Taqqu (1989) determined the asymptotic distribution of the sequential empirical process in the case of long-range dependent data.

Recently, Dehling, Durieu, and Volný (2009) 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 (2011), Durieu and Tusche (2012) and Dehling, Durieu, and Tusche (2012) 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, 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.

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 $H_0$: “the process is stationary with marginal distribution $\mu$” against the alternative $H_A$: “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

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

where $F_k(x) = P(X_1 \leq x)$. 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

$$E(K(x, s)K(y, t)) = \min(s, t)(F(\min(x, y)) - F(x)F(y)).$$

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

Komlós, Major, and Tusnády (1975), using a technique due to Csörgő and Révész (1975), 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 (1981) or the survey article by Gänssler and Stute (1979).

Many authors have studied extensions of the sequential empirical process CLT to dependent data, e.g. Berkes and Philipp (1977) and Philipp and Pinzur (1980) for strongly mixing processes and Berkes, Hörmann, and Schauer (2009) for $S$-mixing processes. Dehling and Taqqu (1989) determined the asymptotic distribution of the sequential empirical process in the case of long-range dependent data.

Recently, Dehling, Durieu, and Volný (2009) 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 (2011), Durieu and Tusche (2012) and Dehling, Durieu, and Tusche (2012) 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, 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.

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 $H_0$: “the process is stationary with marginal distribution $\mu$” against the alternative $H_A$: “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

$$T_n := \max_{0 \leq k \leq n} \sup_x \frac{k}{n} \left(1 - \frac{k}{n}\right) \sqrt{n} \left|F_k(x) - F_{k+1,n}(x)\right|,$$
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
\[
R_n(x,t) = \sqrt{n} t (1 - t) \left( F_{nt}(x) - F_{[nt]+1,n}(x) \right).
\]

As proved in the appendix (Theorem 7), assuming "convergence of the sequential empirical process", we obtain under the null hypothesis \( H_0 \) that
\[
R_n \Rightarrow (K(x,t) - tK(x,1))_{(x,t) \in \mathbb{R} \times [0,1]},
\]
where \( K \) is the centred Gaussian process with covariance structure
\[
\text{Cov}(K(x,t), K(y,s)) = \min\{s,t\} \left\{ \sum_{k=0}^\infty \text{Cov}(1\{X_0 \leq x\}, 1\{X_k \leq y\}) + \sum_{k=1}^\infty \text{Cov}(1\{X_0 \leq y\}, 1\{X_k \leq x\}) \right\}.
\]

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

Note that, in fact this result remains true for general \( \mathcal{F} \)-indexed empirical processes, (see Proposition 5).

The remainder of this paper is organized as follows: in Section 2 we present sequential empirical CLTs for \( \mathcal{B} \)-geometrically ergodic Markov chains (Theorem 1) and dynamical systems that have a spectral gap on the transfer operator (Theorem 3). More abstract results, such as a sequential empirical CLT for multiple mixing random variables (Theorem 5), are stated in Section 3. These results will be the foundation for the later proof of the theorems from Section 2. The proofs of the aforementioned results can be found in Section 4, 5 and 6. The asymptotic distribution of the test statistic \( T_n \) (Proposition 5) is given in the appendix.

## 2 Sequential Empirical CLTs under Spectral Gap

### 2.1 Definitions and Notations

Let \( (\mathcal{X}, \mathcal{A}) \) be a measurable space. For a positive measure \( \lambda \) on \( \mathcal{X} \) and a \( \lambda \)-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 \( L^s(\lambda) \) 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. \( \lambda \) by \( L^\infty(\lambda) \) and the corresponding (essential) supremum norm by \( \| \cdot \|_\infty \). Note that this norms depend heavily on the choice of the measure \( \lambda \), 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 \(\mu\) 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 \(F_n : \mathcal{F} \to \mathbb{R}\), 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

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

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

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 \(\sigma\)-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 (1996)).

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 \(\varphi : \ell^\infty(\mathcal{F} \times [0,1]) \to \mathbb{R}\). 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 (1987) and van der Vaart and Wellner (1996, 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, Durieu, and Tusche (2012).

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

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

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 \(\varepsilon > 0, A > 0, \text{ and } s \in [1, \infty]\), we call \([l,u]\) an \((\varepsilon, A, \mathcal{G}, L^s(\mu))\)-bracket, if \(l, u \in \mathcal{G}\) and

\[
\|u - l\|_s \leq \varepsilon, \quad \|u\|_{\mathcal{C}} \leq A, \quad \|l\|_{\mathcal{C}} \leq A.
\]

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

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

as the smallest number of \((\varepsilon, A, \mathcal{G}, L^s(\mu))\)-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 rate of the size of the brackets in $L^s$-norm, but also with a control of the increasing rate of the $\| \cdot \|_{C}$-size of the bracketing functions as the $L^s$-norm goes to zero.

### 2.2 $\mathcal{B}$-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 $\nu$. We assume that the Markov chain starts with initial distribution $\nu$, i.e. that the distribution of $X_0$ is $\nu$. This makes $(X_i)_{i \in \mathbb{N}}$ a stationary sequence. We also denote by $P$ the associated Markov operator defined by

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

We assume that there exists a complex Banach space $(\mathcal{B}, \| \cdot \|_B)$ of measurable functions from $\mathcal{X}$ to $\mathbb{C}$ such that $P$ is a bounded linear operator on $\mathcal{B}$. We denote by $\mathcal{L}(\mathcal{B})$ the space of bounded linear operator from $\mathcal{B}$ to $\mathcal{B}$. We will need the following properties of the space $\mathcal{B}$:

(A) $1_{\mathcal{X}} \in \mathcal{B}$, $|f|$ and $\overline{f}$ for all $f \in \mathcal{B}$, and the Dirac measures $\delta_x$ are continuous on $\mathcal{B}$.

Moreover 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 $\| \cdot \|_m \leq K \| \cdot \|_B$.

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

(C) $\|P^n f - (\nu f) 1_{\mathcal{X}}\|_B \leq \kappa \|f\|_B \theta^n$ 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 (1993), Hervé (2008), and Hervé and Pène (2010)).

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

For a function $f$ from $\mathcal{X}$ to $\mathbb{R}$, using Fourier kernels, we introduce the perturbed operators

$$(P_{f,t}\varphi)(x) = P(e^{itf}\varphi)(x) = \int_{\mathcal{X}} e^{itf(y)}\varphi(y)P(x, dy), \quad t \in \mathbb{R}.$$ 

In order to apply the Nagaev method (c.f. Hennion and Hervé (2001)), we also need the following condition for some real vector space $C$ of functions from $\mathcal{X}$ to $\mathbb{R}$, which will be specified later in the applications.
(D) For all \( f \in \mathcal{C} \), for \( t \) in a neighbourhood of 0 we have that \( P_{f,t} \in \mathcal{L}(\mathcal{B}) \) and further that \( t \mapsto P_{f,t} \) is two times continuous differentiable with derivative given by

\[
\left( \frac{d^k}{dt^k} P_{f,t} \right)_{t=0} \varphi = P((i f)^k \varphi), \quad \varphi \in \mathcal{B}.
\]

We will see that the conditions (A) – (D) guarantee a sequential finite dimensional CLT for functions in \( \mathcal{C} \) (see Proposition 4 in Section 6). Remark that, in application, \( \mathcal{C} \) will be chosen as a subset of \( \mathcal{B} \).

Now to establish a tightness property of the empirical process, the following further condition on the space \( \mathcal{B} \) is useful.

(E) 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 \( \| fg \|_{\mathcal{B}} \leq C \max\{\| f \|_{\mathcal{B}}, \| g \|_{\mathcal{B}}\}^\ell \).

Note that if \( \mathcal{B} \) is a Banach algebra, condition (E) holds with \( \ell = 2 \). If further \( \mathcal{C} \) is a subset of \( \mathcal{B} \), then for every \( f \in \mathcal{C} \), the mapping \( t \mapsto P_{f,t} \) is an entire function and therefore condition (D) is also satisfied.

To derive a CLT for a \( \mathcal{F} \)-indexed empirical process, we now have to precise the relation between the class \( \mathcal{F} \) and the Banach space \( \mathcal{B} \) or, more precisely, between \( \mathcal{F} \) and the vector space \( \mathcal{C} \) which satisfies condition (D). Note that, in the particular case where \( \mathcal{F} \) is a subset of \( \mathcal{C} \), from (A) – (D) 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 (1987). Nevertheless, in many examples, the functions of \( \mathcal{F} \) do not belong to the space \( \mathcal{B} \). To overcome this difficulty, we have to measure how the functions of \( \mathcal{F} \) are well approximated by the functions of \( \mathcal{B} \). 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 \( \mathcal{B} \). Since \( \mathcal{F} \) is composed of real-valued function, we concentrate on real-valued function on the space \( \mathcal{B} \). We denote by \( \mathcal{B}_R \) the subset of \( \mathcal{B} \) composed by real-valued function. Note that \( (\mathcal{B}_R, \| \cdot \|_{\mathcal{B}}) \) is a real Banach space. Our conditions on the Markov chain (in particular condition (C)) enable us to deal with bracketing numbers allowing an exponential growth of the \( \mathcal{B} \)-norm of the bracket functions as the \( \| \cdot \|_{s} \)-size of the bracket goes to zero. This leads the following entropy condition.

For some \( s \in [1, \infty] \) and \( \mathcal{G} \subset \mathcal{B}_R \),

(F) there exist \( C > 0 \), \( r > -1 \), and \( \gamma > 1 \) such that

\[
\int_0^1 \varepsilon^r \sup_{\varepsilon \leq \delta \leq 1} N^2(\delta, \exp(\delta^{-\frac{1}{2}}), \mathcal{F}, \mathcal{G}, L^s(\nu)) d\varepsilon < \infty \quad (1)
\]

Remark 2. Observe that for \( r' \geq 0 \), inequality (1) holds for all \( r > 2r' - 1 \), if

\[
N(\varepsilon, \exp(\varepsilon^{-\frac{1}{2}}), \mathcal{F}, \mathcal{G}, L^s(\mu)) = O(\varepsilon^{-r'}) \quad \text{as } \varepsilon \to 0.
\]

Note further, that the supremum appears to deal with the possible non-monotonicity of the bracketing number.
We can now state the sequential empirical central limit theorem, which is proved in Section 6.

**Theorem 1** (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 (E) hold. If there is a $\| \cdot \|_\infty$-bounded subset $\mathcal{G} \subset \mathcal{B}_\mathbb{R}$ such that (D) is satisfied for $\mathcal{C} = \text{Vect}_\mathbb{R}(\mathcal{G})$, the smallest real vector space containing $\mathcal{G}$, and if (F) 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

$$\text{Cov}(K(f_1, t_1), K(f_2, t_2)) = \min\{t_1, t_2\} \left\{ \sum_{k=0}^{\infty} \text{Cov}(f_1(X_0), f_2(X_k)) + \sum_{k=1}^{\infty} \text{Cov}(f_1(X_k), f_2(X_0)) \right\}. \tag{2}$$

Remark 3. A centred Gaussian process $K$ with covariance structure (2) is often referred to as a Kiefer process.

Now, let us give an example by applying Theorem 1 to random iterative Lipschitz models.

**2.3 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 $\sigma$-algebra $\mathcal{B}(\mathcal{X})$. 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

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

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 in average, that is that there exists a $\rho \in (0, 1)$ such that

$$\sum_{i \geq 0} d(T_i(x), T_i(y))p_i(z) < \rho d(x, y), \quad \forall x, y, z \in \mathcal{X}. \tag{3}$$

Statistical properties of such models have been studied by Dubins and Freedman (1966), Barnsley and Elton (1988), Hennion and Hervé (2001), Wu and Shao (2004), Hervé (2008), and by Hervé and Pène (2010) in the case of constant functions $p_i$ and by Doeblin and Fortet (1937), Karlin (1953), Barnsley, Demko, Elton, and Geronimo (1988), Peigné (1993), Pollicott (2001), and by Walkden (2007) in the case of variable functions $p_i$.

As in many of the cited papers, we need the following technical properties. For some fixed
where \( X \) was proved by \( X \), \( \alpha, \beta \) ergodicity of the chain. For every \( p \) attractive \( P \) verified when \((X)\) (see also \( X \)). The following theorem extends this result to the sequential empirical processes.

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

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

(7)

with \( p_{i_n}(T_{i_n-1} \circ \ldots \circ T_{i_1}(x)) \ldots p_{i_1}(x) > 0 \) and \( p_{j_n}(T_{j_n-1} \circ \ldots \circ T_{j_1}(y)) \ldots p_{j_1}(x) > 0 \). Note that conditions (4) – (6) are verified when the family of maps \( T_i \) is finite and (7) is verified when (3) – (6) hold and each \( p_i \) is positive. See Peigné (1993) for a discussion on these assumptions.

Under the conditions (3) – (7), Peigné (1993) proved that the Markov chain has an attractive \( \mathcal{P} \)-invariant probability measure \( \nu \) with existing first moment. We define the stationary process \((X_i)_{i \geq 0}\) on \( X \) as the Markov chain started with distribution \( \nu \), 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 (2013) (see also Wu and Shao (2004) in the case of constant functions \( p_i \)). The following theorem extends this result to the sequential empirical processes.

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

\[
\| \cdot \|_\alpha := \| \cdot \|_\infty + m_\alpha(\cdot),
\]

where

\[
m_\alpha(f) := \sup_{x, y \in X \atop x \neq y} \frac{|f(x) - f(y)|}{d(x, y)^\alpha}.
\]

**Theorem 2.** Let (3) – (7) hold 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 \( \mathcal{H}_\alpha(X, \mathbb{R}) \) for some \( \alpha < \frac{s-1}{s} \) such that (F) holds. Then the \( \mathcal{F} \)-indexed sequential empirical process \((U_n(f, t))_{t \in [0, 1]}\) 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 Kiefer process with covariance given by (2).

**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 \( \mathcal{H}_{\alpha, \beta}(X, \mathbb{C}) \) denote the space of continuous function from \( X \) to \( \mathbb{C} \) with \( \| f \|_{\alpha, \beta} = N_\beta(f) + m_{\alpha, \beta}(f) < \infty \), where

\[
N_\beta(f) = \sup_{x \in X} \frac{|f(x)|}{1 + d(x, x_0)^\beta} \quad \text{and} \quad m_{\alpha, \beta}(f) = \sup_{x, y \in X \atop x \neq y} \frac{|f(x) - f(y)|}{d(x, y)^\alpha(1 + d(x, x_0)^\beta)}.
\]
In particular, the space $\mathcal{H}_\alpha(\mathcal{X}, \mathbb{C}) := \mathcal{H}_{\alpha,0}(\mathcal{X}, \mathbb{C})$ is the space of bounded $\alpha$-Hölder functions from $\mathcal{X}$ to $\mathbb{C}$ and we have $\| \cdot \|_{\alpha,0} = \| \cdot \|_\alpha$. It is a subspace of $\mathcal{H}_{\alpha,\beta}(\mathcal{X}, \mathbb{C})$ for all $\beta > 0$. The following properties are straightforward and given without proofs.

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

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

(ii) for every bounded $f, g \in \mathcal{H}_{\alpha,\beta}(\mathcal{X}, \mathbb{C}),$ we have that

$$\|fg\|_{\alpha,\beta} \leq \|f\|_\infty \|g\|_{\alpha,\beta} + \|g\|_\infty \|f\|_{\alpha,\beta},$$

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

(iv) there exists a $C > 0,$ for every $f \in \mathcal{H}_{\alpha,\beta}(\mathcal{X}, \mathbb{C}),$ $\|f\|_\beta \leq C N_{\beta}(f)^{1/\beta}.$

Therefore condition (B) holds with $m = 1/\beta$ as a consequence of (iv) and condition (E) is satisfied due to (ii). Now, according to Theorem 1 in Peigné (1993), we obtain for all $\alpha, \beta \in (0, 1/2)$ with $\alpha < \beta$ that $P$ is a bounded linear operator on $\mathcal{H}_{\alpha,\beta}(\mathcal{X}, \mathbb{C})$ which satisfies condition (C).

It remains to verify condition (D). We consider the space $\mathcal{H}_\alpha(\mathcal{X}, \mathbb{R})$ of bounded real-valued $\alpha$-Hölder functions on $\mathcal{X}$. Let $f$ be a function of this space and consider the perturbed operator defined by $P_{f,t} \varphi = P(e^{itf} \varphi).$ Using $|e^{ia} - e^{ib}| \leq |a - b|,$ we get that $e^{itf} \in \mathcal{H}_\alpha(\mathcal{X}, \mathbb{C})$ for all $t \in \mathbb{R}.$ Thus, for every $\varphi \in \mathcal{H}_{\alpha,\beta}(\mathcal{X}, \mathbb{C})$ and $t \in \mathbb{R},$ by condition (iii) of Lemma 1, we have $e^{itf} \varphi \in \mathcal{H}_{\alpha,\beta}(\mathcal{X}, \mathbb{C}).$ Since $P \in \mathcal{L}(\mathcal{H}_{\alpha,\beta}(\mathcal{X}, \mathbb{C})),\,$ we infer that $P_{f,t} \in \mathcal{L}(\mathcal{H}_{\alpha,\beta}(\mathcal{X}, \mathbb{C}))$ for all $t \in \mathbb{R}.$ Further, using again condition (iii) of Lemma 1, we see that $t \mapsto P_{f,t}$ is an analytic function from $\mathbb{R}$ to $\mathcal{L}(\mathcal{H}_{\alpha,\beta}(\mathcal{X}, \mathbb{C})),\,$ given by $P_{f,t} \varphi = \sum_{k \geq 0} P((itf)^k \varphi) t^k / k!.$ We infer that (D) holds over the space $\mathcal{H}_\alpha(\mathcal{X}, \mathbb{R}).$

We now apply Theorem 1. Let $s, \alpha,$ and $\mathcal{G}$ be as in the statement of Theorem 2. By choosing $\beta = (s - 1)/s < \frac{1}{2},$ we have $\alpha < \beta$ and thus (A) – (E) hold for the space $\mathcal{B} = \mathcal{H}_{\alpha,\beta}(\mathcal{X}, \mathbb{C})$ with $m = 1/\beta$ and $\mathcal{C} = \mathcal{H}_\alpha(\mathcal{X}, \mathbb{R}).$ Further, for any $g \in \mathcal{G},$ we have $g \in \mathcal{H}_{\alpha,\beta}(\mathcal{X}, \mathbb{C})$ and $\|g\|_{\alpha,\beta} \leq \|g\|_\alpha.$ Therefore, condition (F) is also satisfies with respect to the $\mathcal{H}_{\alpha,\beta}(\mathcal{X}, \mathbb{C})$-norm.

### 2.4 Dynamical Systems with a Spectral Gap

Let us mention that, as usual, the proof of Theorem 1 can be adapted to deal with dynamical systems using the Perron–Frobenius operator in place of the Markov operator. Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a probability space and let $T$ be a measure preserving transformation of $\mathcal{X},$ that is $\mu(T^{-1}A) = \mu(A)$ for all $A \in \mathcal{A}.$ The Perron–Frobenius operator $P$ is defined on $L^1(\mu)$ by the equation

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

Further, for a function $f$ on $\mathcal{X},$ we define the perturbed operator by $P_{f,t} \varphi = P(e^{itf} \varphi).$ We have the following result, for which the proof follows the one of Theorem 1 and is left to the reader.
**Theorem 3** (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 a real number $m \geq 1$ such that the conditions (A), (B), (C), and (E) hold with respect to the Perron–Frobenius operator and replacing $\nu$ by $\mu$. If there exists a $\| \cdot \|_\infty$-bounded subset $\mathcal{G} \subset \mathcal{B}$ such that (D) holds for the space $\mathcal{C} = \text{Vect}_{\mathbb{R}}(\mathcal{G})$ and (F) holds for $s = \frac{m}{m-1}$, then the process $(U_n(f,t))_{t \in [0,1]}$, defined by $U_n(f,t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} (f \circ T^i - \mu f)$, converges in distribution in $\ell^\infty(\mathcal{F} \times [0,1])$ to a centred Gaussian process $K$ with covariance structure given by

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

As a possible application, we can extend the empirical CLT proved by Collet, Martinez, and Schmidt (2004) for a class of expanding maps of the interval, to a sequential empirical CLT. In the situation considered in Collet, Martinez, and Schmidt (2004), the spectral gap property can be established on the space of bounded variation functions. Gouëzel (2009) 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 conditions (D) and (E) are trivially satisfied. Further, the usual class of the indicator functions of intervals can be well approximated by Lipschitz functions, and the condition (F) is verified for this class, see also Section 2.5.

### 2.5 Indexing Classes of Functions

To conclude this section, we present some classes of functions for which an estimate of the bracketing number can be computed. These examples, which satisfy condition (F), come from the paper by Dehling, Durieu, and Tusche (2012).

For vectors $x = (x_1, \ldots, x_d), y = (y_1, \ldots, y_d) \in \mathbb{R}^d$ write $x \leq y$ if $x_i \leq y_i$ for all $i \in \{1, \ldots, d\}$. Further, denote the modulus of continuity of a real-valued function $F$ defined on a subset of $\mathbb{R}^d$ by $w_F$. Recall that $w_F(t) := \sup\{|F(x) - F(y)| : |x - y| \leq t\}$, where $\| \cdot \|$ denotes the corresponding Euclidean norm.

**Proposition 1.** For a metric space $\mathcal{X}$, equipped with a probability measure $\mu$, set $\mathcal{B} = \mathcal{H}_\mu(\mathcal{X}, \mathbb{R})$ and $\mathcal{G} = \{ f \in \mathcal{B} : 0 \leq f \leq 1 \}$. We have the following statements about our entropy condition (F).

(i) Let $\mathcal{X} = \mathbb{R}^d$, $\mathcal{F} := \{1_{[t,u]} : t, u \in \mathbb{R}^d, t \leq u\}$, and let $F$ denote the distribution function of $\mu$. If there are some $s \in [1, \infty]$ and $\gamma > 1$ such that $w_F(x) = O(\|\log(x)\|^{-s\gamma})$ as $x \to \infty$, then condition (F) is satisfied.

(ii) Let $\mathcal{X} = \mathbb{R}^d$ and $\mathcal{F} := \{1_{E(x,r)} : x \in [0,1]^d, r \in [0,D]^d\}$ for any fixed $D > 0$, where $E(x,r)$ denotes the ellipsoid $E(x,r) := \{ y \in \mathbb{R}^d : \sum_{i=1}^{d} (x_i - y_i)^2/r_i^2 \leq 1 \}$. If $\mu$ has a bounded density w.r.t. the Lebesgue measure, then condition (F) is satisfied for all $s \in [1, \infty]$. 

10
(iii) In the situation of (ii), \(F\) can be replaced by \(\{1_{E(x,r)} : x \in \mathbb{R}^d, r \in [0,D]^d\}\), if one furthermore assumes that \(\mu(\{x \in \mathbb{R}^d : |x| > t\}) = O(t^{-\beta})\) as \(t \to \infty\) for some \(\beta \in (0,1)\).

(iv) For an arbitrary metric space \((\mathcal{X}, \rho)\) and \(F = \{B(t) : t \geq 0\}\), where \(B(t) := \{x \in \mathcal{X} : \rho(x_0, x)\}\) for some fixed \(x_0 \in \mathcal{X}\), let \(G := \mu(B(\text{id}_{\mathbb{R}^d}))\). Then condition (F) holds if there are some \(s \in [1,\infty]\) and \(\gamma > 1\) such that \(w_G(x) = O(|\log(x)|^{-s\gamma})\) as \(x \to \infty\).

(v) Let \(\mathcal{X} = \mathbb{R}, F = \{f_t : t \in [0,1]\}\), where the \(f_t\) are functions from \(\mathbb{R}\) to \(\mathbb{R}\) which satisfy

(i) \(0 \leq f_t(x) \leq 1\) for all \(t \in [0,1]\) and \(x \in \mathbb{R}\),

(ii) \(f_s \leq f_t\) for all \(0 \leq s \leq t \leq 1\),

(iii) \(f_t\) is monotone increasing on \(\mathbb{R}\) for all \(t \in [0,1]\), and

(iv) \(G_{\mu}(t) = \mu f_t\) is Lipschitz.

Further, let \(F\) denote the distribution function of \(\mu\). If there are some \(s \in [1,\infty]\) and \(\gamma > 1\) such that \(w_F(x) = O(|\log(x)|^{-s\gamma})\) as \(x \to \infty\), then condition (F) holds.

3 Sequential Empirical CLTs for Multiple Mixing Processes

In this section, we present a more general result which can be applied in the setting of Section 2. In particular, the approach used here is useful when the indexing class \(F\) is disjoint from the space of functions on which we have good properties. In the more abstract setting, our technique requires two basic assumptions concerning the process \((f(X_i))_{i \in \mathbb{N}}\), where \(f : \mathcal{X} \rightarrow \mathbb{R}\) belongs to some normed vector space \((C, \|\cdot\|_C)\) of functionals on \(\mathcal{X}\). We assume that for some \(\|\cdot\|_\infty\)-bounded subset \(\mathcal{G} \subset C\) the following two properties hold.

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

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

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 (Moment bounds for \(\mathcal{G}\)-observables). For fixed \(p \in \mathbb{N}^*, s \geq 1\), and monotone increasing functions \(\Phi_1, \ldots, \Phi_p : \mathbb{R}_+ \rightarrow \mathbb{R}_+\), we consider the \(2p\)-th moment bound

\[
E \left[ \left(\sum_{i=1}^{n} (f(X_i) - \mu f)\right)^{2p} \right] \leq \sum_{i=1}^{p} n^i \|f\|_i^2 \Phi_i(\|f\|_C) \quad \text{for all} \quad f \in \mathcal{G} \cup (\mathcal{G} - \mathcal{G}).
\]  (8)

With these assumptions we can show the following abstract sequential empirical CLT.

Theorem 4. Let \((\mathcal{X}, \mathcal{A})\) be a measurable space, let \((X_i)_{i \in \mathbb{N}}\) be a \(\mathcal{X}\)-valued stationary process with marginal distribution \(\mu\), and let \(F\) be a uniformly bounded class of measurable functions on \(\mathcal{X}\). Suppose that for some normed vector space \(C\) of measurable functions on \(\mathcal{X}\), some
subset $\mathcal{G}$ of $\mathcal{C}$ which is bounded in $\|\cdot\|_\infty$-norm, $p \in \mathbb{N}^*$, $s \geq 1$ and some monotone increasing functions $\Phi_1, \ldots, \Phi_p : \mathbb{R}_+ \to \mathbb{R}_+$, Assumption 1 and Assumption 2 hold. Moreover, assume that there exist a constant $r > -1$ and a monotone increasing function $\Psi : \mathbb{R}_+ \to \mathbb{R}_+$ such that
\[
\int_{0}^{1} \varepsilon^r \sup_{\varepsilon \leq \delta \leq 1} N^2(\delta, \Psi(\delta^{-1}), \mathcal{F}, \mathcal{G}, L^s(\mu)) d\varepsilon < \infty.
\] (9)

If
\[
\Phi_i(2\Psi(x)) = O(x^{\gamma_i}),
\] (10)
for some non-negative constants $\gamma_i$ such that
\[
\gamma_i < 2p - (i + r + 2),
\] (11)
then the sequential empirical process $U_n$ converges in distribution in $\ell^\infty(\mathcal{F} \times [0,1])$ to a tight Gaussian process $K$.

The proof can be found in Section 4.

Remark 4. Note that the entropy bounds presented in Proposition 1 are strong versions of entropy conditions of the type in Theorem 4.

In the general setting of Theorem 4, we cannot precise the covariance structure of the limit process. The next lemma shows that under additional conditions, the limit process of $U_n$ is indeed a Kiefer process (c.f. Remark 3).

Lemma 2. In the situation of Theorem 4, assume that

(i) Assumption 1 holds with covariance matrix $\Sigma$ given by
\[
\Sigma_{i,j} = \min\{t_i, t_j\} \left\{ \sum_{k=0}^{\infty} \text{Cov}(f_i(X_0), f_j(X_k)) + \sum_{k=1}^{\infty} \text{Cov}(f_j(X_0), f_i(X_k)) \right\},
\]

(ii) there is a function $\Theta : \mathbb{N} \to \mathbb{R}_+$ and a constant $\beta > 1$ satisfying
\[
\sum_{k=1}^{\infty} \Psi(k^\beta) \Theta(k) < \infty
\]
such that for all $f \in \mathcal{G} \cup (\mathcal{G} - \mathcal{G})$ and all $\varphi \in L^s(\mu)$
\[
|\text{Cov}(\varphi(X_0), f(X_k))| \leq \|\varphi\|_s \|f\|_c \Theta(k).
\]

Then the covariance structure of the limit process $K$ is given by (2).

The proof is given in Section 5.

Note that Proposition 4 in Section 6 shows that Assumption 1 can be established in the setting of Section 2.2. Assumption 2 has been verified for $p = 2$, $\Phi_1(x) = \log^3(x + 1)$, and $\Phi_2(x) = \log^2(x + 1)$ by Durieu (2008), who considers Markov chains and dynamical systems that support a spectral gap property. In a later work, Dehling and Durieu (2011) generalized this result to general $p \in \mathbb{N}^*$ and $\Phi_1(x) = \log^p(x + 1)$. More general, they show that for a process which satisfies the so called multiple mixing condition w.r.t. $C$, for every $\|\cdot\|_\infty$-bounded $\mathcal{G} \subset \mathcal{C}$ and every $p \in \mathbb{N}^*$, there is a $c > 0$ such that Assumption 2 holds with $\Phi_i(x) = c \log^{2p+1}(d_0-1)i(x + 1)$. The multiple mixing condition is defined as follows.
**Definition** (Multiple Mixing Processes). We say that a process \((X_i)_{i \in \mathbb{N}}\) is multiple mixing with respect to \(C\) if there exist a \(\theta \in (0, 1)\) and an integer \(d_0 \in \mathbb{N}\) such that for all \(p \in \mathbb{N}^*\), there exist an integer \(\ell\) and a multivariate polynomial \(P\) of total degree not larger than \(d_0\) such that

\[
|\text{Cov}(f(X_{i_0}) \cdots f(X_{i_{q-1}}), f(X_{i_q}) \cdots f(X_{i_p}))| \leq \|f\|_s \|f\|_c^p (i_1 - i_0, \ldots, i_p - i_{q-1})^{\theta^{q-1}}
\]  

(12)

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

Note that in the setting of Section 2.2, this property with \(d_0 = 0\) can be derived from the spectral gap property, see Lemma 5. For multiple mixing processes, we have the following version of Theorem 4.

**Theorem 5** (Sequential empirical CLT for multiple mixing random variables). Let \((\mathcal{X}, A)\) be a measurable space, let \((X_i)_{i \in \mathbb{N}}\) be a \(\mathcal{X}\)-valued stationary process with marginal distribution \(\mu\), and let \(F\) be a uniformly bounded class of measurable functions on \(\mathcal{X}\). Suppose that for some \(s \geq 1\), the process \((X_i)_{i \in \mathbb{N}}\) is multiple mixing w.r.t. a normed vector space \(C\) of measurable functions on \(\mathcal{X}\), where for every \(p \in \mathbb{N}^*\) the multivariate polynomial \(P\) in inequality (12) is of total degree not larger than \(d_0\). If further Assumption 1 and condition (F) hold for some \(\|\cdot\|_\infty\)-bounded subset \(G\) of \(C\) and \(\gamma > d_0 + 1\), then the sequential empirical process \(U_n\) converges in distribution in \(l^\infty(F \times [0, 1])\) to a tight Gaussian process \(K\).

If further the covariance matrix \(\Sigma\) in Assumption 1 is given by

\[
\Sigma_{i,j} = \min\{t_i, t_j\} \left\{ \sum_{k=0}^{\infty} \text{Cov}(f_i(X_0), f_j(X_k)) + \sum_{k=1}^{\infty} \text{Cov}(f_j(X_0), f_i(X_k)) \right\},
\]

and if there are constants \(\theta \in (0, 1)\) and \(D > 0\) such that for all \(f \in G \cup (G - G)\) and all \(\varphi \in L^s(\mu)\)

\[
|\text{Cov}(\varphi(X_0), f(X_k))| \leq D \|\varphi\|_s \|f\|_c \theta^k,
\]

then the covariance structure of the limit process \(K\) is given by (2).

**Proof.** As aforementioned, multiple mixing processes satisfy Assumption 2 with \(\Phi_1(x) = c \log^{2p+(d_0-1)/\gamma}(x + 1)\) for some \(c > 0\) depending only on \(p\). Thus choosing \(\Psi := \exp(C \text{id}^{1/\gamma})\) for some \(C > 0\) and \(\gamma > 1\) (which gives a quite relaxed entropy condition concerning the \(\|\cdot\|_\text{c-size}\), we have \(\Phi_1(\Psi(x)) = O(x^{(2p+(d_0-1)/\gamma)})\). Therefore condition (11) holds for sufficiently large \(p \in \mathbb{N}^*\) if \(\gamma > d_0 + 1\). The covariance structure of the limit process is a direct consequence of Lemma 2 with \(\Theta(k) = \theta^k\) and \(\beta \in (1, \gamma)\).

**Multiple Mixing of Lower Rate** Processes of a lower mixing rate have been studied by Durieu and Tusche (2012), who consider a multiple mixing condition on \(C = \mathcal{H}_\alpha(\mathbb{R}^d, \mathbb{R})\), \(\alpha \in (0, 1]\), where \(\theta^{q-1}\) in (12) is replaced by a term \(\Theta(i_q - i_{q-1})\) with a monotone decreasing function \(\Theta : \mathbb{N} \to \mathbb{R}_+\) such that \(\sum_{i=0}^{\infty} i^{2p-2} \Theta(i) < \infty\) and \(d_0 = 0\). In this case they were able to prove the validity of Assumption 2 with \(\Phi_1 = c \text{id}^3\), \(c > 0\), see Proposition 1 in Durieu and Tusche (2012).

Such a mixing type is given e.g. for multidimensional causal functions of i.i.d. processes. A causal function of an i.i.d. process \((\xi_i)_{i \in \mathbb{Z}}\) is defined as a process \((X_i)_{i \in \mathbb{N}}\) given by \(X_i = \)
$G(\xi_1, \xi_{-1}, \ldots)$, where $G : \mathcal{X}^N \rightarrow \mathbb{R}^d$ is a measurable function. The physical measure of dependence $\delta_{i,m}$ (introduced by Wu (2005)) is defined by

$$\delta_{i,m} := \mathbb{E}\{\|G(\xi_1, \xi_{-1}, \ldots) - G(\xi_i, \xi_{-1}, \ldots, \xi_1, \xi_0, \xi_{-1}, \ldots)\|^m\}^{\frac{1}{m}},$$

where $(\xi'_i)_{i \in \mathbb{Z}}$ is an independent copy of $(\xi_i)_{i \in \mathbb{Z}}$. Durieu and Tusche (2012) showed that a causal function of an i.i.d. process has the aforementioned mixing property with $\Theta(i) = \delta_{i,m}^s$, where $m = s/(s - 1)$.

As an example consider the following MA-process. Let $(\xi_i)_{i \in \mathbb{Z}}$ be an i.i.d. process in a normed vector space $(\mathcal{Y}, \|\cdot\|_Y)$, let $(a_j)_{j \in \mathbb{N}}$ be a family of $\mathbb{R}^d$-valued linear functionals on $\mathcal{Y}$, $|a|_s := \sup\{|a(y)| : \|y\|_Y \leq 1\}$, and define the process $(X_i)_{i \in \mathbb{E}}$ by $X_i = \sum_{j=1}^{\infty} a_j \xi_{i-j}$. In this case we have $\delta_{i,m} \leq (2\|X_0\|_s)^\alpha \sum_{j=1}^{\infty} |a_j|_s$, and thus, assuming that $\|X_0\|_s$ and $\sum_{j=1}^{\infty} |a_j|_s$ are finite, $(X_i)_{i \in \mathbb{E}}$ has the multiple mixing property with $\Theta(i) = \sum_{j=1}^{\infty} |a_j|_s$.

Note that as a consequence of working with $\Phi_i = \epsilon \text{id}^l$ in this setting, in order to satisfy condition (10) in Theorem 4, we cannot choose $\Psi = \exp(C \text{id}^{1/\gamma})$, but need to work with a polynomial type of $\Psi$, which on the other hand requires stronger types of bracketing numbers than in condition (F). These can be achieved mainly using stronger conditions on $\mu$ or restrictions on $F$. Observe that the bracketing numbers presented in (ii) and (iii) of Proposition 1 are actually also available for polynomial $\Psi$ (c.f. Dehling, Durieu, and Tusche (2012)) without further restrictions of $\mu$ or $F$. Further, bracketing numbers for indicators of semifinite rectangles of the type $[-\infty, t]$, $t \in \mathbb{R}^d$, with polynomial $\Psi$ are implicitly given in Durieu and Tusche (2012) at the cost of stronger assumption on the marginal distribution $\mu$.

4 Proof of Theorem 4

The main idea of the proof is to introduce some approximation $U^{(q)}_n$ for the original process $U_n$, which is based on functions in $G$ and thus can be controlled by Assumption 1 and 2. The approximation can be constructed as follows: For all $q \geq 1$, there exist two sets of $N_q := N(2^{-q}, \Psi(2^q), F, G, L^s(\mu))$ functions $\{g_{q,1}, \ldots, g_{q,N_q}\} \subset G$ and $\{g'_{q,1}, \ldots, g'_{q,N_q}\} \subset G$, such that

$$\|g_{q,i} - g'_{q,i}\|_s \leq 2^{-q}, \quad \|g_{q,i}\|_c \leq \Psi(2^q), \quad \|g'_{q,i}\|_c \leq \Psi(2^q) \quad (13)$$

and for all $f \in F$, there exists some $i$ such that $g_{q,i} \leq f \leq g'_{q,i}$. Further, by (9),

$$\sum_{q \geq 1} 2^{-(r+1)q} N_q^2 2^q < \infty \quad (14)$$

To approximate the indexing function $f \in F$, construct a partition of $F$ into $N_q$ subsets $F_{q,i}$ such that for each $f \in F_{q,i}$ 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 $f \in F_{q,i^{*}}$. To approximate the time parameter we use the partition of $[0, 1]$ into subsets $T_{j,i}$, $j = 1, \ldots, 2^q$, given by $T_{j,i} := [(j-1)2^{-q}, j2^{-q})$ for $j < 2^q$ and $T_{2^q,i} := [1 - 2^{-q}, 1]$. For $t \in [0, 1]$ we define $\tau_q t := \max\{(j-1)2^{-q} \leq t : j = 1, \ldots, 2^q\}$ and further $\tau_t := \tau_q t + 2^{-q}$. We extend the notation introduced in Section 2 to arbitrary $\mu$-integrable functions $f : \mathcal{X} \rightarrow \mathbb{R}$ by setting

$$F_n(f) := \frac{1}{n} \sum_{i=1}^{n} f(X_i)$$
and for $t \in [0, 1]$

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

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

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

Note that this process is constant on each $F_{q, i} \times T_{q, j}$.

To draw the connection between the weak asymptotic behaviour of the original process $U_n$ and the approximating process $U_n(q)$, we use an altered version of Theorem 4.2 in Billingsley (1968, p.25):

**Theorem 6.** Let $X_n, X_n^{(q)}, X^{(q)}, q, n \geq 1$ be random elements with values in the Banach space $(\ell^\infty(F \times [0, 1]), \| \cdot \|_\infty)$ and suppose that $X^{(q)}$ is measurable and separable\(^1\). If the conditions

$$X_n^{(q)} \rightsquigarrow X^{(q)} \quad \text{as } n \to \infty \text{ for all } q \geq 1,$$

$$\limsup_{n \to \infty} P^*(\|X_n - X^{(q)}\|_\infty \geq \delta) \to 0 \quad \text{as } q \to \infty \text{ for all } \delta > 0 \quad (16)$$

are satisfied, then there exists an $\ell^\infty(F \times [0, 1])$-valued, separable random variable $X$ such that $X^{(q)} \rightsquigarrow X$ as $q \to \infty$ and

$$X_n \rightsquigarrow X \quad \text{as } n \to \infty.$$ 

For the proof and further details see Theorem 2.1 in Dehling, Durieu, and Tusche (2012).

We will prove Theorem 4 by establishing condition (15) and (16) in the two following propositions:

**Proposition 2.** For all $q \in \mathbb{N}^*$ the process $(U_n^{(q)}(f, t))_{(f, t) \in F \times [0, 1]}$ converges in distribution to a piecewise constant Gaussian process $(U^{(q)}(f, t))_{(f, t) \in F \times [0, 1]}$ as $n \to \infty$.

**Proposition 3.** Assume that Assumption 2 holds for some $p \in \mathbb{N}^*$, $s \geq 1$ and some monotone increasing functions $\Phi_1, \ldots, \Phi_p : \mathbb{R}_+ \to \mathbb{R}_+$. Moreover, suppose there exists a constant $r > -1$ and an monotone increasing function $\Psi : \mathbb{R}_+ \to \mathbb{R}_+$ such that (9) holds. If (10) holds for some non-negative constants $\gamma_i$ satisfying (11), then for all $\varepsilon, \eta > 0$ there exists some $q_0$ such that for all $q \geq q_0$

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

\(^1\)Since the objects we work with involve supremum over the non separable space $\ell^\infty(F \times [0, 1])$, measurability can not always be guaranteed. Thus we need to use the theory of outer probability as presented in van der Vaart and Wellner (1996). In this context we call any (not necessarily measurable) functions on a probability space a random element and we call it a random variable if it is also measurable. We denote the outer probability with respect to a probability measure $P$ by $P^*$. Furthermore a random variable $X$ with values in some space $S$ is called separable, if there exists some separable subset $S'$ of $S$ such that such that $P(X \in S') = 1$. 

15
Proof of Theorem 4. We can now apply Theorem 6 with \(X_n = U_n, \, X_n^{(q)} = U_n^{(q)}, \, X^{(q)} = U^{(q)}\). By Proposition 2 the convergence (15) holds, while (16) is satisfied due to Proposition 3. Therefore \(U_n\) converges in distribution to \(W\). 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 (1996)).

Proof of Proposition 2. Since by construction \(\pi_q f \in \mathcal{G}\) for all \(f \in \mathcal{F}\), due to Assumption 1, the finite dimensional process \(U^{(q)}_n(f_1, t_1), \ldots, U^{(q)}_n(f_k, t_k)\) converges in distribution to some multi-dimensional normal distributed random variable \(U^{(q)}(f_1, t_1), \ldots, U^{(q)}_n(f_k, t_k)\) for all fixed \(k \in \mathbb{N}^*, \, f_1, \ldots, f_k \in \mathcal{F}, \, t_1, \ldots, t_k \in [0,1]\). All \(U^{(q)}_n, \, n \in \mathbb{N}^*, \) are constant on each \(\mathcal{F}_{q,i} \times \mathcal{T}_{q,j}, \, i = 1, \ldots, N^q, \, j = 1, \ldots, 2^q\). 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 \(U^{(q)}_n(f, t))_{(f, t) \in \mathcal{F} \times [0,1]}\).

Proof of Proposition 3. Let \(\overline{Z} := Z - \mathbf{E}Z\) denote the centring of a random variable \(Z\) and observe that for any random variables \(Y_1 \leq Y \leq Y_2\) the inequality

\[
|Y - \mathbf{E}Y| \leq |Y_1 - \mathbf{E}Y_1| + \mathbf{E}|Y_2 - Y_1|
\]

holds. Since for \(f \in \mathcal{F}, \, k \in \mathbb{N}\) we have \(F_{[nt]}(\pi_{q+k} f, t) \leq F_{[nt]}(f, t) \leq F_{[nt]}(\pi_{q+k} f, t), \) using that \(\|\cdot\|_1 \leq \|\cdot\|_s\) for \(s \geq 1\) and applying (13), we obtain

\[
|U_n^{(q)}(f, t) - U_n(\pi_{q+k} f, t)| \\
\leq |U_n^{(q)}(\pi_{q+k} f, t) - U_n(\pi_{q+k} f, t)| + \frac{\sqrt{n}}{\sqrt{n}} \mathbf{E}|F_{[nt]}(\pi_{q+k} f - \pi_{q+k} f)| \\
\leq |U_n(\pi_{q+k} f, t) - U_n(\pi_{q+k} f, t)| + \sqrt{n} 2^{-(q+k)}.
\]

Moreover, for all \(n \geq 2^{q+k}\) and \(g \in \mathcal{G}\)

\[
|U_n(g, t) - U_n(g, \tau_{q+k} t)| = \frac{1}{\sqrt{n}} \left| \sum_{i=\lfloor n \tau_{q+k} t \rfloor + 1}^{\lfloor nt \rfloor} g(X_i) - \mu(g) \right| \\
\leq 2Mn^{-\frac{1}{2}} (\lfloor nt \rfloor - \lfloor n \tau_{q+k} t \rfloor) \\
\leq 4M \sqrt{n} 2^{-(q+k)},
\]

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

\(U^{(q)}_n(f, t) := U_n(\pi_{q'} f, \tau_{q'} t)\).

An application of the triangle inequality, (17), and (18) yields

\[
|U_n(f, t) - U_n^{(q+k)}(f, t)| \leq |U_n^{(q+k)}(f, t) - U_n^{(q+k)}(f, t)| + (4M + 1) \sqrt{n} 2^{-(q+k)}.
\]
Combining (19) with a telescopic sum argument, one obtains for any $K \geq 1$

\[
\left| U_n(f, t) - U_n^{(q)}(f, t) \right| \\
= \left\{ \sum_{k=1}^{K} \left| U_n^{(q+k)}(f, t) - U_n^{(q+k-1)}(f, t) \right| + U_n(f, t) - U_n^{(q+K)}(f, t) \right\} \\
\leq \left\{ \sum_{k=1}^{K} \left| U_n^{(q+k)}(f, t) - U_n^{(q+k-1)}(f, t) \right| \right\} + \left| U_n^{(q+K)}(f, t) - U_n^{(q+K)}(f, t) \right| \\
+ (4M + 1) \sqrt{n} 2^{-(q+K)}. \tag{20}
\]

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

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

For each $i = 1, \ldots, N_q$, $j = 1, \ldots, 2^q$, inequality (20) implies

\[
\sup_{t \in T_{q,j}} \sup_{f \in \mathcal{F}_{q,i}} \left| U_n(f, t) - U_n^{(q)}(f, t) \right| \leq \left\{ \sum_{k=1}^{K_{n,q}} \sup_{t \in T_{q,j}} \sup_{f \in \mathcal{F}_{q,i}} \left| U_n^{(q+k)}(f, t) - U_n^{(q+k-1)}(f, t) \right| \right\} \\
+ \sup_{t \in T_{q,j}} \sup_{f \in \mathcal{F}_{q,i}} \left| U_n^{(q+K)}(f, t) - U_n^{(q+K)}(f, t) \right| + \frac{\varepsilon}{2}.
\]

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

\[
P^* \left( \sup_{t \in T_{q,j}} \sup_{f \in \mathcal{F}_{q,i}} \left| U_n(f, t) - U_n^{(q)}(f, t) \right| \geq \varepsilon \right) \\
\leq \left\{ \sum_{k=1}^{K_{n,q}} P^* \left( \sup_{t \in T_{q,j}} \sup_{f \in \mathcal{F}_{q,i}} \left| U_n^{(q+k)}(f, t) - U_n^{(q+k-1)}(f, t) \right| \geq \varepsilon_k \right) \right\} \\
+ P^* \left( \sup_{t \in T_{q,j}} \sup_{f \in \mathcal{F}_{q,i}} \left| U_n^{(q+K)}(f, t) - U_n^{(q+K)}(f, t) \right| \geq \frac{\varepsilon}{4} \right). \tag{21}
\]

Recall that $(\pi_{q+k}, \tau_{q+k})$ and thus $U_n^{(q+k)}$ and $U_n^{(q+k)}$ are constant on each $\mathcal{F}_{q+k,i} \times T_{q+k,j}$, $i = 1, \ldots, N_{q+k}$, $j = 1, \ldots, 2^{q+k}$, and thus the suprema on the r.h.s. of inequality (21) 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

\[
F_{k,q,i} := \mathcal{F}_{q,i} \cap \mathcal{F}(q + k) \\
T_{k,q,j} := \{(j-1)2^{-q} + (m - 1)2^{-(q+k)} : m \in \{1, \ldots, 2^k\}\}.
\]
Inequality (21) and the triangle inequality imply

\[
P^*(\sup_{t \in T_{q,j}} \sup_{f \in F_{q,i}} |U_n(f, t) - U_n^{(q)}(f, t)| \geq \varepsilon)
\]
\[
\leq \left\{ \sum_{k=1}^{K_{n,q}} \sum_{t \in T_{k,q,j}} \sum_{f \in F_{k,q,i}} (\varepsilon_k/2)^{-2p} \left( E\|U_n(\pi_{q+k} f, \tau_{q+k-1} t) - U_n(\pi_{q+k-1} f, \tau_{q+k-1} t)\|^{2p} \right) + E\|U_n(\pi_{q+k} f, \tau_{q+k} t) - U_n(\pi_{q+k} f, \tau_{q+k} t)\|^{2p} \right\} + \left\{ \sum_{t \in T_{K_{n,q},q,j}} \sum_{f \in F_{K_{n,q},q,i}} (\varepsilon/8)^{-2p} \left( E\|U_n(\pi_{q+K_{n,q}} f, \tau_{q+K_{n,q}} t) - U_n(\pi_{q+K_{n,q}} f, \tau_{q+K_{n,q}} t)\|^{2p} \right) + E\|U_n(\pi_{q+K_{n,q}} f, \tau_{q+K_{n,q}} t) - U_n(\pi_{q+K_{n,q}} f, \tau_{q+K_{n,q}} t)\|^{2p} \right\}.
\]

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

\[
P^*(\sup_{t \in T_{q,j}} \sup_{f \in F_{q,i}} |U_n(f, t) - U_n^{(q)}(f, t)| \geq \varepsilon)
\]
\[
\leq \left\{ \sum_{k=1}^{K_{n,q}} \sum_{t \in T_{k,q,j}} \sum_{f \in F_{k,q,i}} (\varepsilon_k/2)^{-2p} \left( E\|U_n(\pi_{q+k} f, \tau_{q+k-1} t) - U_n(\pi_{q+k-1} f, \tau_{q+k-1} t)\|^{2p} \right) + E\|U_n(\pi_{q+k} f, \tau_{q+k} t) - U_n(\pi_{q+k} f, \tau_{q+k} t)\|^{2p} \right\} + \left\{ \sum_{t \in T_{K_{n,q},q,j}} \sum_{f \in F_{K_{n,q},q,i}} (\varepsilon/8)^{-2p} \left( E\|U_n(\pi_{q+K_{n,q}} f, \tau_{q+K_{n,q}} t) - U_n(\pi_{q+K_{n,q}} f, \tau_{q+K_{n,q}} t)\|^{2p} \right) + E\|U_n(\pi_{q+K_{n,q}} f, \tau_{q+K_{n,q}} t) - U_n(\pi_{q+K_{n,q}} f, \tau_{q+K_{n,q}} t)\|^{2p} \right\}.
\]

We will treat the expected values on the r.h.s. of inequality (22) separately now by using Assumption 2 and properties of our brackets used to cover F. Recall that by (13) we have

\[
\|\pi_{q+k} f - \pi_{q+k-1} f\|_s \leq \|\pi_{q+k} f - f\|_s + \|\pi_{q+k-1} f - f\|_s \leq 3 \cdot 2^{-(q+k)}
\]
\[
\|\pi_{q+k} f - \pi'_{q+k} f\|_s \leq 2^{-(q+k)}
\]
\[
\|\pi_{q+k} f - \pi_{q+k-1} f\|_c \leq 2\Psi(2^{q+k})
\]
\[
\|\pi_{q+k} f - \pi'_{q+k} f\|_c \leq 2\Psi(2^{q+k}).
\]

For convenience, throughout the rest of the proof we 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
Applying successively (8), (23), (24), and (10) we have
\[
\mathbb{E}\left| U_n(\pi_{q+k} f, \tau_{q+k-1} t) - U_n(\pi_{q+k-1} f, \tau_{q+k-1} t) \right|^{2p} \ll n^{-p} \sum_{\ell=1}^{p} n^{\ell} \|\pi_{q+k} f - \pi_{q+k-1} f\|_{\Phi_\ell(\|\pi_{q+k} f - \pi_{q+k-1} f\|_{C})} \\
\ll \sum_{\ell=1}^{p} n^{-(p-\ell)2(\gamma_\ell-\ell)(q+k)}
\] (25)
and analogously
\[
\mathbb{E}\left| U_n'(\pi_{q+k+n,q} f, \tau_{q+k+n,q} t) - U_n(\pi_{q+k+n,q} f, \tau_{q+k+n,q} t) \right|^{2p} \ll \sum_{\ell=1}^{p} n^{-(p-\ell)2(\gamma_\ell-\ell)(q+k)}.
\] (26)

For fixed \( g \in \mathcal{G} \) we have by stationarity
\[
\mathbb{E}\left| U_n(g, \tau_{q+k} t) - U_n(g, \tau_{q+k-1} t) \right|^{2p} = n^{-p} \mathbb{E} \left[ \left( \sum_{i=1}^{\lfloor n\tau_{q+k-1} t \rfloor - [n\tau_{q+k-1} t]} (g(X_i) - \mu g) \right)^{2p} \right],
\] (27)
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
\[
[n\tau_{q+k} t] - [n\tau_{q+k-1} t] \leq n2^{-(q+k)} + 1 \quad \text{for all } n \geq 2^{q+k}.
\]
Applying (8), (13), and (10) to (27) we obtain
\[
\mathbb{E}\left| U_n(\pi_{q+k} f, \tau_{q+k} t) - U_n(\pi_{q+k} f, \tau_{q+k-1} t) \right|^{2p} \ll n^{-p} \sum_{\ell=1}^{p} (n2^{-(q+k)})^\ell \|\pi_{q+k} f\|_{\Phi_\ell(\|\pi_{q+k} f\|_{C})} \\
\ll \sum_{\ell=1}^{p} n^{-(p-\ell)2(\gamma_\ell-\ell)(q+k)}
\] (28)
and analogously
\[
\mathbb{E}\left| U_n'(\pi_{q+k+n,q} f, \tau_{q+k+n,q} t) - U_n'(\pi_{q+k+n,q} f, \tau_{q+k+n,q} t) \right|^{2p} \ll \sum_{\ell=1}^{p} n^{-(p-\ell)2(\gamma_\ell-\ell)(q+k)}.
\] (29)

Now, apply (25), (26), (28), and (29) to (22). We infer
\[
P^*(\sup_{t \in \bar{T}_{q,i}} \sup_{\ell \in \mathcal{F}_{q,i}} \left| U_n(\pi_{q+k} f, \tau_{q+k} t) - U_n^{(q)}(f, t) \right| \geq \varepsilon) \\
\ll \sum_{k=1}^{K_{n,q}} \# T_{k,q,j} \# F_{k,q,t} \frac{(k(k+1))^{2p}}{\varepsilon^{2p}} \sum_{\ell=1}^{p} n^{-(p-\ell)2(\gamma_\ell-\ell)(q+k)}.
\] (30)
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 \( \sum_{i=1}^{N_q} \# F_{k,q,i} = \# \mathcal{F}(q + k) \leq N_{q+k-1} N_{q+k} \). Therefore (30) yields

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

This implies that for any \( \eta > 0 \)

\[
P^* \left( \sup_{t \in [0,1]} \sup_{f \in \mathcal{F}} \left| U_n(f, t) - U_n^{(q)}(f, t) \right| > \varepsilon \right)
\ll \sum_{\ell=1}^{p} n^{-(p-\ell)} \max \left\{ 1, 2^{(p-\ell)+r+2+\eta}(q+K_{n,q}) \right\} \sum_{k=1}^{K_{n,q}} N_{q+k-1} N_{q+k} k^{4p} 2^{-(r+1+\eta)(q+k)}
\ll \max \left\{ 1, \max_{\ell=1,\ldots,p} n^{\frac{1}{2}(r+\ell-2p+r+2+\eta)} \right\} \sum_{k=q+1}^{\infty} N_{k-1} N_k k^{4p} 2^{-(r+1+\eta)k}. \tag{31}
\]

By (11) we can choose \( \eta \) small enough to assure \( \gamma_{\ell_0} + \ell_0 - 2p + r + 2 + \eta < 0 \) for all \( \ell_0 = 1, \ldots, p \). Thus the factor in front of the sum is uniformly bounded w.r.t. \( n \). Using (14), we obtain

\[
\sum_{k=1}^{\infty} N_{k-1} N_k k^{4p} 2^{-(r+1+\eta)k} \leq \sum_{k=1}^{\infty} 2^{-(r+1)k} N_{k-1}^2 k^{4p} 2^{-(r+1)k} + \sum_{k=1}^{\infty} 2^{-(r+1)k} N_k^2 k^{4p} 2^{-(r+1)k} < \infty
\]

for sufficiently small \( \eta > 0 \) which implies that the series in (31) goes to zero as \( q \to \infty \). \( \square \)

5 Proof of Lemma 2

For \( f \in \mathcal{F} \), recall the definition of the approximating functions \( \pi_q f \) in Section 4 and note that as a consequence of the entropy condition in Theorem 4, we know that for every \( q \in \mathbb{N}^* \)

\[
\| f - \pi_q f \|_s \leq 2^{-q} \tag{32}
\]
\[
\| \pi_q f \|_c \leq \Psi(2^q). \tag{33}
\]

Similarly, for all \( g \in \mathcal{F} \) and \( k \in \mathbb{N}^* \) there exist some \( g_k \in \mathcal{G} \) satisfying

\[
\| g_k - g \|_s \leq k^{-\beta} \tag{34}
\]
\[
\| g_k \|_c \leq \Psi(k^\beta). \tag{35}
\]
Let $U^{(q)}$ denote the limit process given in Proposition 2. Condition (i) implies that for all $f, g \in \mathcal{F}$, $t, u \in [0, 1]$ and $q \in \mathbb{N}^*$

$$\text{Cov}(U^{(q)}(f, t), U^{(q)}(g, u)) = \min\{t, u\} \left\{ \sum_{k=0}^{\infty} \text{Cov}(\pi_q f(X_0), \pi_q g(X_k)) + \sum_{k=1}^{\infty} \text{Cov}(\pi_q g(X_0), \pi_q f(X_k)) \right\}.$$ 

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

$$\sum_{k=0}^{\infty} \text{Cov}(\pi_q f(X_0), \pi_q g(X_k)) = \text{Cov}(f(X_0), g(X_k)) \quad (36)$$

$$+ \sum_{k=1}^{\infty} \text{Cov}(\pi_q g(X_0), \pi_q f(X_k)) - \text{Cov}(g(X_0), f(X_k)) \longrightarrow 0 \quad \text{as} \quad q \to \infty.$$ 

By symmetry, both series can be treated the same way. Let $k(q) := 2^{q/\beta}$. We consider the series in line (36). We have

$$\sum_{k=0}^{\infty} \text{Cov}(\pi_q f(X_0), \pi_q g(X_k)) = \text{Cov}(f(X_0), g(X_k)) \quad (37)$$

$$\leq \sum_{k=0}^{k(q)} \text{Cov}(\pi_q f(X_0) - f(X_0), \pi_q g(X_k)) + \sum_{k=0}^{k(q)} \text{Cov}(f(X_0), \pi_q g(X_k) - g(X_k))$$

$$+ \sum_{k=k(q)+1}^{\infty} |\text{Cov}(\pi_q f(X_0) - f(X_0), \pi_q g(X_k))|$$

$$+ \sum_{k=k(q)+1}^{\infty} |\text{Cov}(f(X_0), \pi_q g(X_k) - g(X_k))|.$$ 

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 (37), we know by Hölder’s inequality, (32), and the fact that $\beta > 1$ that

$$\sum_{k=0}^{k(q)} \text{Cov}(\pi_q f(X_0) - f(X_0), \pi_q g(X_k)) + \sum_{k=0}^{k(q)} \text{Cov}(f(X_0), \pi_q g(X_k) - g(X_k))$$

$$\ll \sum_{k=0}^{k(q)} (\|\pi_q f - f\|_s + \|\pi_q g - g\|_s)$$

$$\ll k(q) 2^{-q} = 2^{-(1-\beta)q} \longrightarrow 0 \quad \text{as} \quad q \to \infty,$$ 

21
where again, we write $x \ll y$ if there is a constant $C \in (0, \infty)$ depending only on global parameters such that $x \leq Cy$. For the term in line (38), by (ii), (32), and (33) we obtain

$$
\sum_{k=k(q)+1}^{\infty} \left| \text{Cov} \left( \pi_q f(X_0) - f(X_0), \pi_q g(X_k) \right) \right|
\leq \|\pi_q f - f\|_s \sum_{k=k(q)+1}^{\infty} \|\pi_q g\| c \Theta(k)
\ll 2^{-q} \sum_{k=k(q)+1}^{\infty} \Psi(2^q) \Theta(k) \rightarrow 0 \text{ as } q \rightarrow \infty,
$$

where we used that $\Psi$ is increasing and condition (ii) in the last step. It only remains to show, that the term in line (39) goes to zero as $q \rightarrow \infty$. We have

$$
\sum_{k=k(q)+1}^{\infty} \left| \text{Cov} \left( f(X_0), \pi_q g(X_k) - g(X_k) \right) \right|
\leq \sum_{k=k(q)+1}^{\infty} \left| \text{Cov} \left( f(X_0), \pi_q g(X_k) - g_k(X_k) \right) \right|
+ \sum_{k=k(q)+1}^{\infty} \left| \text{Cov} \left( f(X_0), g_k(X_k) - g(X_k) \right) \right|.
$$

First, consider the term in line (40). By (ii), (33), and (35)

$$
\sum_{k=k(q)+1}^{\infty} \left| \text{Cov} \left( f(X_0), \pi_q g(X_k) - g_k(X_k) \right) \right|
\ll \sum_{k=k(q)+1}^{\infty} \|f\|_s \|\pi_q g - g_k\| c \Theta(k)
\ll \left( \sum_{k=k(q)+1}^{\infty} \|\pi_q g\| c \Theta(k) \right) + \left( \sum_{k=k(q)+1}^{\infty} \|g_k\| c \Theta(k) \right)
\ll \left( \sum_{k=k(q)+1}^{\infty} \Psi(2^q) \Theta(k) \right) + \left( \sum_{k=k(q)+1}^{\infty} \Psi(k^\beta) \Theta(k) \right) \rightarrow 0 \text{ as } q \rightarrow \infty,
$$

where we used that $\Psi$ is increasing and applied condition (ii) in the last line. To treat the term in line (41), we use Hölder’s inequality and (34). We obtain

$$
\sum_{k=k(q)+1}^{\infty} \left| \text{Cov} \left( f(X_0), g_k(X_k) - g(X_k) \right) \right|
\ll \sum_{k=k(q)+1}^{\infty} \|g_k - g\|_s
\ll \sum_{k=k(q)+1}^{\infty} k^{-\beta} \rightarrow 0 \text{ as } q \rightarrow \infty,
$$

since $\beta > 1$ and thus $\sum_{k=1}^{\infty} k^{-\beta} < \infty$, which completes the proof. \(\square\)
6 Proof of Theorem 1

Let $(X_i)$ and $(B, \| \cdot \|)$ be the Markov chain and the Banach space introduced in Section 2.2. To prove Theorem 1, we shall apply Theorem 5. We begin by showing, that Assumption 1 holds with covariance structure (2). To this aim, we will partially follow the lines of the proof of Theorem A in Hennion and Hervé (2001). For a measurable real-valued function $f$ on $\mathcal{X}$ and a real number $t \in [0, 1]$, we introduce the notation

$$S_n(f, t) := \sum_{i=1}^{[nt]} f(X_i).$$

The following proposition gives a sequential finite dimensional CLT under (A) – (D).

**Proposition 4.** Suppose that for some $m \in [1, \infty]$, (A), (B), (C) hold. Let $k$ be a positive integer and $t_1, \ldots, t_k \in [0, 1]$. Let $f_1, \ldots, f_k$ be real-valued functions on $\mathcal{X}$ such that $\nu(|f_i|^2) < \infty$ and (D) holds for the space $C = \text{Vect}_\mathbb{R}(f_1, \ldots, f_k)$, the smallest real vector space containing $f_1, \ldots, f_k$. Then, we have

$$\frac{1}{\sqrt{n}}(S_n(f_1 - \nu f_1, t_1), \ldots, S_n(f_k - \nu f_k, t_k)) \rightsquigarrow N(0, \Sigma) \text{ as } n \to \infty,$$

where $N(0, \Sigma)$ is a normal distribution in $\mathbb{R}^k$ with mean 0 and covariance matrix $\Sigma = (\Sigma_{i,j})_{1 \leq i,j \leq k}$. If furthermore $f_1, \ldots, f_k \in L^s(\nu) \cap B_\mathbb{R}$ with $s = m/(m-1)$, then the covariance matrix is given by

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

This proposition will show that Assumption 1 holds with covariance structure (2) since by assumption, $\mathcal{G}$ is only composed by bounded real-valued functions from $B$.

**Proof.** First let $f$ be a function as in the statement of the proposition. By the Perturbation Theorem (see Theorem III.8 in Hennion and Hervé (2001)), there exist a neighbourhood $I_f$ of 0 and $0 < \theta < \eta < 1$ such that for all $t \in I_f$, there exist operators $\Pi_{f,t}$ and $N_{f,t}$ and complex numbers $\lambda_{f,t}$ such that

$$P_{f,t} = \lambda_{f,t} \Pi_{f,t} + N_{f,t}$$

with

$$\Pi_{f,t} = \Pi_{f,t}, \quad N_{f,t} \circ \Pi_{f,t} = \Pi_{f,t} \circ N_{f,t} = 0, \quad \rho(N_{f,t}) < \theta, \quad |\lambda_{f,t}| \geq \eta \quad \text{for all } t \in I_f.$$

Moreover, $\lambda_{f,0} = 1, \Pi_{f,0} = \Pi, N_{f,0} = N$ and the maps $t \mapsto \lambda_{f,t}, t \mapsto \Pi_{f,t}$ and $t \mapsto N_{f,t}$ have continuous second derivatives on $I_f$. We thus have for all $n \geq 1$

$$P_{f,t}^n = \lambda_{f,t}^n \Pi_{f,t} + N_{f,t}^n.$$

Further, if $\nu(f) = 0$, by Lemma IV.4' in Hennion and Hervé (2001) the Taylor expansion of $\lambda_{f,t}$ as $t$ goes to 0 is given by

$$\lambda_{f,t} = 1 - \frac{t^2}{2} \sigma_f^2 + o(t^2) \quad (42)$$
Lemma 3.

For every function \(X \in B\), and every measurable function \(g, s \in C\), \(s \in (0, 1)\) and set \(S_n = S_n(f, g, s)\). The following lemma gives us an expression of the corresponding characteristic function.

**Lemma 3.** For every function \(\varphi \in B\), \(t \in \mathbb{R}\), and \(n \geq 1\),

\[
E(e^{itS_n} \varphi(X_n)) = \nu \left( P_{f,t}^{[ns]} P_{g,t}^{n-[ns]} \varphi \right).
\]

In particular, the characteristic function of \(n^{-\frac{1}{2}}S_n\) is given by

\[
E\left( e^{itn^{-\frac{1}{2}}S_n} \right) = \nu \left( P_{f,t}^{[ns]} P_{g,t}^{n-[ns]} 1_\chi \right). \tag{44}
\]

**Proof of Lemma 3.** For every \(k \geq 1\) and every measurable function \(F : \chi^{k-1} \to \mathbb{R}\), we have

\[
E\left( e^{it(F(X_1, \ldots, X_k) + f(X_k))} \varphi(X_k) \right) = E\left( e^{itF(X_1, \ldots, X_{k-1})} E\left( e^{itf(X_k)} \varphi(X_k | X_{k-1}, \ldots, X_1) \right) \right)
\]

and the same equation with \(g\) instead of \(f\). The Lemma can now be proved by induction. \(\square\)

To study the weak convergence of \(S_n\), we have to compute the limit of

\[
P_{f,t}^{[ns]} P_{g,t}^{n-[ns]} 1_\chi = \lambda_{f,t}^{[ns]} \lambda_{g,t}^{n-[ns]} \Pi_{f,t}^{[\chi]} \Pi_{g,t}^{[\chi]} 1_\chi + \lambda_{f,t}^{[ns]} \Pi_{g,t}^{[\chi]} N_{f,t}^{n-[ns]} 1_\chi
\]

and

\[
\lambda_{g,t}^{n-[ns]} \Pi_{f,t}^{[\chi]} 1_\chi + N_{g,t}^{[ns]} N_{f,t}^{n-[ns]} 1_\chi.
\]

\[\text{(43)}\]
By (42), we infer that
\[ \lambda_{f, \nu}^{[n]} \to \exp\left(-\frac{t^2}{2} s \sigma_f^2\right) \quad \text{and} \quad \lambda_{g, \nu}^{[n]} \to \exp\left(-\frac{t^2}{2} (1-s) \sigma_g^2\right) \quad \text{as} \ n \to \infty, \]
where \( \sigma_f \) and \( \sigma_g \) are given by (43). Further, since \( \rho(N_{f,t}) < 1 \) and \( \rho(N_{g,t}) < 1 \), we have that \( \|N_{f,t}^n\|_{L(B)} \to 0 \) and \( \|N_{g,t}^n\|_{L(B)} \to 0 \) uniformly in \( t \in I_f \cap I_g \) as \( n \to \infty \). By continuity, we also have \( \Pi_{f, \nu}^{\frac{1}{\sqrt{n}}} 1_x \to 1_x \) and \( \Pi_{g, \nu}^{\frac{1}{\sqrt{n}}} 1_x \to 1_x \) as \( n \to \infty \). We therefore obtain
\[
\left\| \lambda_{f, \nu}^{[n]} \Pi_{f, \nu}^{\frac{1}{\sqrt{n}}} \lambda_{g, \nu}^{[n]} \Pi_{g, \nu}^{\frac{1}{\sqrt{n}}} 1_x \right\|_B \leq \left| \lambda_{f, \nu}^{[n]} \right| \left\| \Pi_{f, \nu}^{\frac{1}{\sqrt{n}}} \|_{L(B)} \right\| \left\| N_{g, \nu}^{[n]} \|_{L(B)} \right\| 1_x \|_B \to 0, \\
\left\| \lambda_{g, \nu}^{[n]} \Pi_{g, \nu}^{\frac{1}{\sqrt{n}}} \lambda_{f, \nu}^{[n]} \Pi_{f, \nu}^{\frac{1}{\sqrt{n}}} 1_x \right\|_B \leq \left| \lambda_{g, \nu}^{[n]} \right| \left\| \Pi_{g, \nu}^{\frac{1}{\sqrt{n}}} \|_{L(B)} \right\| \left\| N_{f, \nu}^{[n]} \|_{L(B)} \right\| 1_x \|_B \to 0, \\
\left\| N_{f, \nu}^{[n]} \lambda_{f, \nu}^{[n]} \lambda_{g, \nu}^{[n]} \Pi_{g, \nu}^{\frac{1}{\sqrt{n}}} 1_x \right\|_B \leq \left\| N_{f, \nu}^{[n]} \\|_{L(B)} \right\| \left\| N_{g, \nu}^{[n]} \|_{L(B)} \right\| 1_x \|_B \to 0, \\
\lambda_{f, \nu}^{[n]} \lambda_{g, \nu}^{[n]} \Pi_{f, \nu}^{\frac{1}{\sqrt{n}}} \Pi_{g, \nu}^{\frac{1}{\sqrt{n}}} 1_x \to \exp\left(-\frac{t^2}{2} s \sigma_f^2\right) \exp\left(-\frac{t^2}{2} (1-s) \sigma_g^2\right) 1_x \quad \text{as} \ n \to \infty. 
\]
This we infer
\[
\lim_{n \to +\infty} P_{f, \nu}^{[n]} P_{g, \nu}^{[n]} 1_x = \exp\left(-\frac{t^2}{2} (s \sigma_f^2 + (1-s) \sigma_g^2)\right) 1_x \quad \text{as} \ n \to \infty, 
\]
which, using (44), gives the weak convergence of \( \frac{S_{[n]}}{\sqrt{n}} \) to a centred normal distribution with variance given by \( \sigma_{f,g,s} = s \sigma_f^2 + (1-s) \sigma_g^2 \). By (43), we obtain that Proposition 4 holds with the covariance matrix \( \Sigma \) given by
\[
\Sigma_{i,j} = \min\{t_i, t_j\} \frac{1}{2} \left( \sigma_{i+f,j}^2 - \sigma_{i,j}^2 - \sigma_{i+f,j}^2 \right). \tag{45}
\]

**Lemma 4.** Under the conditions (A), (B), and (C), for all \( f \in B \) and all \( g \in L^s(\nu) \), with 
\[ s = \frac{m}{m-1}, \]
we have
\[ |\text{Cov}(g(X_0), f(X_k))| \leq C\|g\|_s \|f\|_{B^k}. \]

**Proof.** Applying successively Hölder inequality, (B), and (C), we get
\[
|\text{Cov}(g(X_0), f(X_k))| \leq E|g(X_0)| E|f(X_n) - \nu f|X_0| \leq \|g\|_s \|P^n f - (\nu f) 1_X\|_B \leq C\|g\|_s \|f\|_{B^k}. 
\]

The preceding Lemma shows that the series \( \sum_{k=0}^{\infty} \text{Cov}(f(X_0), f(X_k)) \) converges for \( f \in B_{\mathbb{R}} \cap L^s(\nu) \). Thus, using Kronecker’s Lemma, equation (43) becomes
\[ \sigma_f^2 = \nu(f^2) + 2 \sum_{k \geq 1} \text{Cov}(f(X_0), f(X_k)), \]
which, with (45), completes the proof of Proposition 4. \( \square \)
Proof of Theorem 1. Now, take $\mathcal{C} = \text{Vect}_\mathbb{R}(\mathcal{G})$. By Proposition 4, Assumption 1 is satisfied. In order to apply Theorem 5 to prove Theorem 1, it remains to show the multiple mixing property of $(X_i)_{i \in \mathbb{N}}$. The following lemma, which is basically Lemma 3 in Dehling and Durieu (2011), gives this property w.r.t. the Banach space $\mathcal{B}$ containing $\mathcal{C}$.

Lemma 5. Under the conditions (A), (B), (C), and (E), $(X_i)_{i \in \mathbb{N}}$ satisfies the multiple mixing property w.r.t. $\mathcal{B}$ with $d_0 = 0$ and $s = m/(m - 1)$.

Proof. Let $f \in \mathcal{B}$ such that $\| f \|_\infty \leq 1$ and set $s = m/(m - 1)$. For all $p > q > 0$, for all $i_0 < i_1 < \ldots < i_p$, we write $g = f P^{i_q + 1 - i_q} (f P^{i_{q+1} - i_q} (\ldots (f P^{i_p} (f))))$. By (E), $g$ belongs to $\mathcal{B}$. Using Hölder’s inequality, we obtain
\[
  |\text{Cov}(f(X_{i_0}) \cdots f(X_{i_{q-1}}), f(X_{i_q}) \cdots f(X_{i_p}))| \leq \mathbb{E}(|f(X_{i_0}) \cdots f(X_{i_{q-1}})|^s)^{1/s} \| P_{q+1}^{i_q} (g) - \nu(g) \|_m.
\]

Using (B), (C), and $\| f \|_\infty \leq 1$, we infer
\[
  |\text{Cov}(f(X_{i_0}) \cdots f(X_{i_{q-1}}), f(X_{i_q}) \cdots f(X_{i_p}))| \leq \kappa \| f \|_s \| g \|_\mathcal{B} |i_q - i_{q+1}|^{\ell}.
\]

Now, since the spectral radius of $P \in \mathcal{L}(\mathcal{B})$ is 1, there exists a $c \geq 1$, which does not depend on $f$, such that $\| P^k f \|_\mathcal{B} \leq c \| f \|_\mathcal{B}$ for all $k \in \mathbb{N}^*$. By (E), we obtain two constants $C > 0$ and $\ell \in \mathbb{N}^*$, depending only on $p$, such that $\| g \|_\mathcal{B} \leq C \| f \|_\mathcal{B}^{\ell}$. This completes the proof of the lemma.

To conclude the proof of Theorem 1 observe, that the extra assumptions in Theorem 5, which concern the covariance structure of the limit process, are satisfied due to Proposition 4 and Lemma 4.

7 Appendix

Consider the natural generalization of the process $R_n$ introduced in Section 1. Let $(X_i)_{i \in \mathbb{N}}$ be a $\mathcal{X}$-valued stationary process with empirical measure $F_n(f) := n^{-1} \sum_{i=1}^n f(X_i), n \in \mathbb{N}^*$. We set $F_0(f) = 0$. For $j \in \{1, \ldots, n\}$ 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 $R_n = \{R_n(f,t) \mid (f,t) \in \mathcal{F} \times [0,1]\}$ given by
\[
  R_n(f,t) := \sqrt{n} \left( n \left[ n t \right] - \left[ n t \right] \right) (F_{\left[ n t \right]}(f) - F_{\left[ n t \right] + 1}(f)).
\]

The following theorem gives the asymptotic distribution of $R_n$.

Theorem 7. Assume that $(X_i)_{i \in \mathbb{N}}$ satisfies the sequential empirical CLT with indexing class $\mathcal{F}$ and limit process $K$, that is, $U_n \sim K$ in $\ell^\infty(\mathcal{F} \times [0,1])$ as $n \to \infty$, where $K$ denotes a tight centred Gaussian process. Then
\[
  R_n \sim (K(f,t) - tK(f,1))((f,t) \in \mathcal{F} \times [0,1])
\]
in $\ell^\infty(\mathcal{F} \times [0,1])$ to as $n \to \infty$.  

26
Proof. Let $\mu$ denote the distribution function of the $X_i$. For $t \in [1/n, 1)$ we have

$$F_{[nt]}(f) - F_{[nt]+1,n}(f)$$

$$= \frac{1}{\lfloor nt \rfloor} \sum_{i=1}^{\lfloor nt \rfloor} f(X_i) - \frac{1}{n - \lfloor nt \rfloor} \sum_{i=\lfloor nt \rfloor+1}^{n} f(X_i)$$

$$= \frac{1}{\lfloor nt \rfloor} \sum_{i=1}^{\lfloor nt \rfloor} (f(X_i) - \mu f) - \frac{1}{n - \lfloor nt \rfloor} \sum_{i=\lfloor nt \rfloor+1}^{n} (f(X_i) - \mu f)$$

$$= \left( \frac{1}{\lfloor nt \rfloor} + \frac{1 - n}{n - \lfloor nt \rfloor} \right) \sum_{i=1}^{\lfloor nt \rfloor} (f(X_i) - \mu f) - \frac{1}{n - \lfloor nt \rfloor} \sum_{i=1}^{n} (f(X_i) - \mu f)$$

$$= \frac{n}{\sqrt{n \lfloor nt \rfloor n - \lfloor nt \rfloor}} U_n(f, t) - \frac{1}{\sqrt{nt}} \sum_{i=1}^{n} \frac{n - \lfloor nt \rfloor}{n} U_n(f, 1). \quad (46)$$

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 $(46)$ that

$$R_n(f, t) = U_n(f, t) - \frac{\lfloor nt \rfloor}{n} U_n(f, 1),$$

$$= U_n(f, t) - tU_n(f, 1) + \frac{n - \lfloor nt \rfloor}{n} U_n(f, 1) \quad \text{for all } t \in [0, 1]. \quad (47)$$

Let $A_n$ denote the $\mathcal{F} \times [0, 1]$-indexed processes given by $A_n(f, t) := ((n-t-\lfloor nt \rfloor)/n)U_n(f, t)$. Since $\sup_{t \in [0, 1]} |(n-t-\lfloor nt \rfloor)/n| \to 0$ 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 $(47)$ yields

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

Here we have applied the continuous mapping theorem in the final step. \hfill \Box

Remark 5. Note that, in the setting of Theorem 1 and Theorem 3 and for a wide class of multiple mixing processes (c.f. Theorem 5), the covariance structure of $K$ is given by $(2)$. 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 $T_n$.

**Proposition 5.** If $(X_i)_{i \in \mathbb{N}}$ satisfies the sequential empirical CLT, then under the null hypothesis $H_0$ we have the convergence

$$T_n := \max_{0 \leq k \leq n} \frac{k}{n} \sum_{f \in \mathcal{F}} \frac{1}{\sqrt{n}} |F_k(f) - F_{k+1,n}(f)| \rightsquigarrow \sup_{f \in \mathcal{F}, t \in (0, 1]} |K(f, t) - tK(f, 1)|.$$

Proof. $R_n(f, \cdot)$ is obviously constant on the intervals $[k/n, (k+1)/n)$, $k = 0, \ldots, n-1$ and further $R_n(f, k/n) = k/n(1-k/n)\sqrt{n}(F_k(f) - F_{k+1,n}(f))$ 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

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

27
References

Barnsley, M.F., Demko, S.G., Elton, J.H., Geronimo, J.S.: Invariant measures for Markov processes arising from iterated function systems with place-dependent probabilities. Ann. Inst. H. Poincaré Probab. Statist. 24(3), 367–394 (1988)

Barnsley, M.F., Elton, J.H.: A new class of Markov processes for image encoding. Adv. in Appl. Probab. 20(1), 14–32 (1988)

Berkes, I., Hörmann, S., Schauer, J.: Asymptotic results for the empirical process of stationary sequences. Stochastic Process. Appl. 119(4), 1298–1324 (2009)

Berkes, I., Philipp, W.: An almost sure invariance principle for the empirical distribution function of mixing random variables. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 41(2), 115–137 (1977)

Billingsley, P.: Convergence of probability measures. John Wiley & Sons Inc., New York (1968)

Ciesielski, Z., Kesten, H.: A limit theorem for the fractional parts of the sequence \(\{2^k t\}\). Proc. Amer. Math. Soc. 13, 596–600 (1962)

Collet, P., Martinez, S., Schmitt, B.: Asymptotic distribution of tests for expanding maps of the interval. Ergodic Theory Dynam. Systems 24(3), 707–722 (2004)

Csörgő, M., Révész, P.: A new method to prove strassen type laws of invariance principle. II. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 31(4), 261–269 (1975)

Csörgő, M., Révész, P.: Strong approximations in probability and statistics. Probability and Mathematical Statistics. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York (1981)

Dehling, H., Durieu, O.: Empirical processes of multidimensional systems with multiple mixing properties. Stochastic Process. Appl. 121(5), 1076–1096 (2011)

Dehling, H., Durieu, O., Tusche, M.: Approximating class approach for empirical processes of dependent sequences indexed by functions (2012). Preprint. arXiv: 1201.2256

Dehling, H., Durieu, O., Volný, D.: New techniques for empirical processes of dependent data. Stochastic Process. Appl. 119(10), 3699–3718 (2009)

Dehling, H., Philipp, W.: Empirical process techniques for dependent data. In: Empirical process techniques for dependent data, pp. 3–113. Birkhäuser Boston, Boston, MA (2002)

Dehling, H., Taqqu, M.S.: The empirical process of some long-range dependent sequences with an application to \(U\)-statistics. Ann. Statist. 17(4), 1767–1783 (1989)

Doeblin, W., Fortet, R.: Sur des chaînes à liaisons complètes. Bull. Soc. Math. France 65, 132–148 (1937)

Donsker, M.D.: Justification and extension of Doob’s heuristic approach to the Komogorov-Smirnov theorems. Ann. Math. Statistics 23, 277–281 (1952)
Doob, J.L.: Heuristic approach to the Kolmogorov-Smirnov theorems. Ann. Math. Statistics 20, 393–403 (1949)

Dubins, L.E., Freedman, D.A.: Invariant probabilities for certain Markov processes. Ann. Math. Statist. 37, 837–848 (1966)

Durieu, O.: A fourth moment inequality for functionals of stationary processes. J. Appl. Probab. 45(4), 1086–1096 (2008)

Durieu, O.: Empirical processes of iterated maps that contract on average (2013). Preprint. arXiv:1206.4903

Durieu, O., Tusche, M.: An empirical process central limit theorem for multidimensional dependent data. J. Theoret. Probab. (2012). DOI 10.1007/s10959-012-0450-3

Gänsler, P., Stute, W.: Empirical processes: a survey of results for independent and identically distributed random variables. Ann. Probab. 7(2), 193–243 (1979)

Gouëzel, S.: An interval map with a spectral gap on Lipschitz functions, but not on bounded variation functions. Discrete Contin. Dyn. Syst. 24(4), 1205–1208 (2009)

Hennion, H., Hervé, L.: Limit theorems for Markov chains and stochastic properties of dynamical systems by quasi-compactness, Lecture Notes in Mathematics, vol. 1766. Springer-Verlag, Berlin (2001)

Hervé, L.: Vitesse de convergence dans le théorème limite central pour des chaînes de Markov fortement ergodiques. Ann. Inst. Henri Poincaré Probab. Stat. 44(2), 280–292 (2008)

Hervé, L., Pène, F.: The Nagaev-Guivarc’h method via the Keller-Liverani theorem. Bull. Soc. Math. France 138(3), 415–489 (2010)

Karlin, S.: Some random walks arising in learning models. I. Pacific J. Math. 3, 725–756 (1953)

Kiefer, J.: Skorohod embedding of multivariate RV’s, and the sample DF. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 24, 1–35 (1972)

Komlós, J., Major, P., Tusnády, G.: An approximation of partial sums of independent RV’s and the sample DF. I. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 32, 111–131 (1975)

Meyn, S.P., Tweedie, R.L.: Markov chains and stochastic stability. Communications and Control Engineering Series. Springer-Verlag London Ltd., London (1993)

Müller, D.W.: On Glivenko-Cantelli convergence. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 16, 195–210 (1970)

Ossiander, M.: A central limit theorem under metric entropy with $L_2$ bracketing. Ann. Probab. 15(3), 897–919 (1987)
Peigné, M.: Iterated function systems and spectral decomposition of the associated Markov operator. In: Fascicule de probabilités, *Publ. Inst. Rech. Math. Rennes*, vol. 1993, p. 28. Univ. Rennes I, Rennes (1993)

Philipp, W., Pinzur, L.: Almost sure approximation theorems for the multivariate empirical process. *Z. Wahrsch. Verw. Gebiete* 54(1), 1–13 (1980)

Pollicott, M.: Contraction in mean and transfer operators. *Dyn. Syst.* 16(1), 97–106 (2001)

van der Vaart, A.W., Wellner, J.A.: Weak convergence and empirical processes. Springer-Verlag, New York (1996)

Walkden, C.P.: Invariance principles for iterated maps that contract on average. *Trans. Amer. Math. Soc.* 359(3), 1081–1097 (electronic) (2007)

Wu, W.B.: Nonlinear system theory: another look at dependence. *Proc. Natl. Acad. Sci. USA* 102(40), 14,150–14,154 (electronic) (2005)

Wu, W.B., Shao, X.: Limit theorems for iterated random functions. *J. Appl. Probab.* 41(2), 425–436 (2004)