UNIVERSITÉ FRANÇOIS-RABELAIS DE TOURS

École Doctorale MIPTIS
Laboratoire de Mathématique et Physique Théorique

Thèse présenté par:
Marco Tusche
en cotutelle avec L'UNIVERSITÉ DE BOCHUM

soutenue le: 29 novembre 2013

pour obtenir le grade de: **Docteur de l'université François-Rabelais de Tours**
Discipline / Spécialité: Mathématiques / Probabilités

Empirical Processes of Multiple Mixing Data

Processus empiriques de données à mélange multiple

Thèse dirigèe par:

M. Herold DEHLING	Professeur, Ruhr-Universität Bochum
M. Emmanuel LESIGNE	Professeur, Université François Rabelais de Tours

Co-Encadrant:

M. Olivier DURIEU	MCF, Université François Rabelais de Tours

Rapporteurs:

M. Jérôme DEDECKER	Professeur, Université Paris Descartes
M. Manfred DENKER	Professeur, Pennsylvania State University

Jury:

M. Jérôme DEDECKER	Professeur, Université Paris Descartes
M. Herold DEHLING	Professeur, Ruhr-Universität Bochum
M. Manfred DENKER	Professeur, Pennsylvania State University
M. Olivier DURIEU	MCF, Université François Rabelais de Tours
M. Gerhard KNIEPER	Professeur, Ruhr-Universität Bochum
M. Emmanuel LESIGNE	Professeur, Université François Rabelais de Tours

D | Ruhr-Universität Bochum
Fakultät für Mathematik

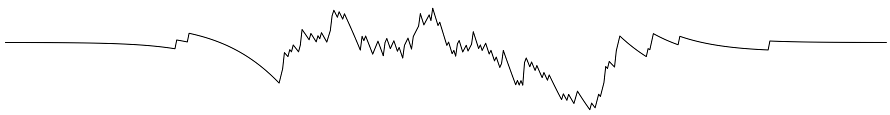
F | Université François Rabelais de Tours
Laboratoire de Mathématiques et Physique Théorique

PhD-Thesis

Empirical Processes of Multiple Mixing Data

Marco Tusche

January 7, 2014



Advisers:	Herold Dehling,	Ruhr-Universität Bochum
	Emmanuel Lesigne,	Université François-Rabelais de Tours
Co-Adviser:	Olivier Durieu,	Université François-Rabelais de Tours
Referees:	Jérôme Dedecker,	Université Paris Descartes
	Manfred Denker,	Pennsylvania State University
Examiners:	Jérôme Dedecker,	Université Paris Descartes
	Herold Dehling,	Ruhr-Universität Bochum
	Manfred Denker,	Pennsylvania State University
	Olivier Durieu,	Université François-Rabelais de Tours
	Gerhard Knieper,	Ruhr-Universität Bochum
	Emmanuel Lesigne,	Université François-Rabelais de Tours
Thesis defense:	November 29, 2013	

Empirical Processes of Multiple Mixing Data

The present thesis studies weak convergence of empirical processes of multiple mixing data. It is based on the articles Durieu and Tusche (2012), Dehling, Durieu, and Tusche (2012), and Dehling, Durieu, and Tusche (2013). A process $(X_i)_{i\in\mathbb{N}}$ is called *multiple mixing* if the covariances of two products $f(X_{i_0})\cdot\ldots\cdot f(X_{i_{q-1}})$ and $f(X_{i_q})\cdot\ldots\cdot f(X_p)$ with $0 \leq i_0 \leq \ldots \leq i_p$ can be bounded, for all f in some normed function space $\mathcal{C}$ with $\sup_x |f(x)| \leq 1$, by a term of the type $\|f(X_0)\|_s \, \|f\|_{\mathcal{C}}^{\ell} \, Q(i_0 - i_1, \ldots, i_p - i_{p-1}) \, \theta^{i_q - i_{q-1}}$ with $\ell, s \geq 1$, $\theta \in (0,1)$ and a polynomial Q.

We follow the approximating class approach introduced by Dehling, Durieu, and Volný (2009) and Dehling and Durieu (2011), who established empirical central limit theorems for dependent $\mathbb{R}$- and $\mathbb{R}^d$-valued random variables, respectively. Extending their technique, we generalize their results to arbitrary state spaces and to empirical processes indexed by classes of functions. Moreover we study sequential empirical processes. Our results apply to $\mathcal{B}$-geometrically ergodic Markov chains, iterative Lipschitz models, dynamical systems with a spectral gap on the Perron–Frobenius operator, and ergodic torus automorphisms. We establish conditions under which the empirical process of such processes converges weakly to a Gaussian process.

As our limit theorems are stated in the general context of empirical processes indexed by functions, they involve entropy conditions using adapted bracketing numbers. We present a range of classes that satisfy these entropy conditions. Examples of those classes are the indicators of finite and semi-finite rectangles, of ellipsoids, and some parametric class of monotone functions.

Furthermore, we consider the empirical process of random variables of a slower type of multiple mixing. Here the decay of the covariances is only of a polynomial instead of an exponential rate, which is required in the concept of usual multiple mixing. We establish an abstract empirical central limit theorem with applications to causal functions of i.i.d. processes such as linear processes and time delay vectors.

Keywords: Limit theorems, Weak convergence, Multivariate empirical processes, Multivariate sequential empirical processes, Empirical processes indexed by classes of functions, Dependent data, Multiple mixing, Markov chains, Dynamical systems, Spectral gap, Ergodic torus automorphism, Chaining, Change-point problems

Mathematics Subject Classification (2010): 60F05, 60F17, 60G10, 62G30, 60J05, 28D05

Processus empirique avec données à mélange multiple

Cette thèse présente des résultats sur la convergence faible de processus empiriques avec données à mélange multiple. Elle est basée sur les articles Durieu et Tusche (2012), Dehling, Durieu et Tusche (2012) et Dehling, Durieu et Tusche (2013). On dit qu'un processus $(X_i)_{i\in\mathbb{N}}$ est à mélange multiple si la covariance des deux produits $f(X_{i_0}) \cdot \ldots \cdot f(X_{i_{q-1}})$ et $f(X_{i_q}) \cdot \ldots \cdot f(X_p)$, avec $0 \leq i_0 \leq \ldots \leq i_p$, peut être bornée par un terme du type $\|f(X_0)\|_s \, \|f\|_{\mathcal{C}}^{\ell} \, Q(i_0 - i_1, \ldots, i_p - i_{p-1}) \, \theta^{i_q - i_{q-1}}$ avec $\ell, s \geq 1$, $\theta \in (0,1)$ et un polynôme Q pour tout f dans un espace de fonctions normé $\mathcal{C}$, avec $\sup_x |f(x)| \leq 1$.

Nous suivons l'approche utilisant des classes approximantes introduites par Dehling, Durieu et Volný (2009) et Dehling et Durieu (2011) qui ont pu obtenir des théorèmes limites centraux empiriques pour des variables aléatoires dépendantes à valeurs dans $\mathbb{R}$ et dans $\mathbb{R}^d$, respectivement. En développant leur technique, nous généralisons leurs résultats à des espaces d'états arbitraires et au cas de processus empiriques indexés par une classe de fonctions. De plus, nous étudions le cas de processus empiriques séquentiels. Nous donnons des conditions garantissant la convergence en loi du processus empirique vers un processus gaussien. Nos résultats s'appliquent aux processus de Markov $\mathcal{B}$-géométriquement ergodiques, aux modèles itératifs lipschitziens, aux systèmes dynamiques dont l'opérateur de Perron Frobenius prsente un trou spectral et aux automorphismes ergodiques du tore. Comme nos théorèmes limite sont énoncés dans le cadre général de processus empiriques indexés par des fonctions nous introduisons une notion d'entropie avec crochets adaptée à cette situation. Nous présentons plusieurs classes qui satisfont ces conditions d'entropie. Parmi ces classes, on trouve l'exemple des indicatrices de rectangles finis et semi-finis, des indicatrices d'ellipsoïdes et une classe paramétrique de fonctions monotones.

De plus, nous considérons le processus empirique de variables aléatoires avec un mélange multiple plus lent. Ici, la covariance décroît avec un ordre polynomial au lieu d'un ordre exponentiel. Nous établissons un théorème limite central empirique abstrait qui s'applique à des fonctions causales de processus i.i.d., comme les processus linéaires et les vecteurs à temps retardé.

Mots clés: Théorème limite, Convergence faible, Processus empiriques multivariés, Processus empiriques séquentiels multivariés, Processus empiriques indexés par une classe de fonctions, Données dépendantes, Mélange multiple, Chaîne de Markov, Système dynamique, Trou spectral, Automorphismes ergodiques du tore, Chaînage, Points de ruptures

Classification AMS (2010): 60F05, 60F17, 60G10, 62G30, 60J05, 28D05

Empirische Prozesse multipel mischender Zufallsvariablen

Die vorliegende Arbeit basiert auf den Arbeiten Durieu und Tusche (2012), Dehling, Durieu und Tusche (2012) und Dehling, Durieu und Tusche (2013). Wir untersuchen die schwache Konvergenz des empirischen Prozesses multipel mischender Zufallsvariablen. Ein Prozess heißt *multipel mischend*, falls die Kovarianzen zweier Produkte $f(X_{i_0}) \cdot \ldots \cdot f(X_{i_{q-1}})$ und $f(X_{i_q}) \cdot \ldots \cdot f(X_p)$ mit $0 \leq i_0 \leq \ldots \leq i_p$ gleichmäßig für alle f mit $\sup_x |f(x)| \leq 1$ aus einem normierten Funktionenraum $\mathcal{C}$ durch einen Term der Form $\|f(X_0)\|_s \, \|f\|_{\mathcal{C}}^{\ell} \, Q(i_0 - i_1, \ldots, i_p - i_{p-1}) \, \theta^{i_q - i_{q-1}}$ mit $\ell, s \geq 1$, $\theta \in (0,1)$ und einem Polynom Q beschränkt sind.

Dehling, Durieu und Volný (2009) entwickelten eine Approximationsklassen-Technik um empirische zentrale Grenzwertsätze abhängiger eindimensionaler und später multivariater Zufallsvariablen (Dehling und Durieu (2011)) zu beweisen. Unter Zuhilfenahme dieser Technik erweitern wir ihre Ergebnisse auf funktionenklassenindizierte empirische und sequentielle empirische Prozesse abhängiger Zufallsvariablen in beliebigen Ereignisräumen. Unsere Grenzwertsätze können auf $\mathcal{B}$-geometrisch ergodische Markov-Ketten, iterative Lipschitz Modelle, dynamische Systeme deren Perron–Frobenius Operator eine Spektrallücke aufweist und ergodische Torusautomorphismen angewandt werden. Wir entwickeln Bedingungen für die schwache Konvergenz des empirischen bzw. sequentiellen empirischen Prozesses solcher Prozesse gegen einen Gauß'schen Prozess.

Da wir durch Funktionenklassen indizierte empirische Prozesse betrachten, beinhalten unsere Grenzwertsätze bestimmte Entropie-Bedingungen. Diese verwenden eine angepasste Art von *Bracketing* Zahlen. Wir stellen eine Reihe von passenden Funktionenklassen vor, darunter die Klasse der Indikatoren von endlichen und halb-endlichen Rechtecken, von Ellipsoiden, sowie eine parametrische Klasse monotoner Funktionen.

Weiterhin untersuchen wir empirische Prozesse von Zufallsvariablen mit einer langsameren Mischungsrate. In dieser Situation ist der Abfall der Kovarianzen nur von polynomialer anstatt von exponentieller Ordnung. Wir entwickeln einen abstrakten zentralen Grenzwertsatz der auf kausale Funktionen unabhängig identisch verteilter Prozesse wie zum Beispiel lineare Prozesse und *Time Delay Vectors* angewandt werden kann.

Stichworte: Grenzwertsätze, Schwache Konvergenz, Multivariate empirische Prozesse, Multivariate sequentielle empirische Prozesse, Funktionenklassenindizierte empirische Prozesse, Abhängige Zufallsvariablen, Multiple Mischungseigenschaft, Markov-Ketten, Dynamische Systeme, Spektrallücken, Ergodische Torusautomorphismen, Chaining, Strukturbruchprobleme

AMS Klassifikation (2010): 60F05, 60F17, 60G10, 62G30, 60J05, 28D05

Acknowledgements

Sincere thanks are given to my advisers Herold Dehling, Emmanuel Lesigne, and Olivier Durieu. Our mutual discussions were a huge benefit for my understanding of the matter and advances in my research. I had a very exciting time working with Herold and Olivier on our common articles. Thanks to the co-workers I had the pleasure to meet both in Tours and Bochum, I encountered a great working atmosphere at both places.

Schließlich möchte ich mich bei meinen Freunden und meiner Familie – insbesondere bei Simone, Renate, Siegmund und Nicolai – für ihre moralische Unterstützung während der Zeit meines Studiums bedanken.

Contents

1. Introduction

1.1. Empirical Central Limit Theorems

This thesis is dedicated to the study of limit theorems for *empirical processes* of *multiple mixing* data. Let $(X_i)_{i\in\mathbb{N}}$ be a stationary real-valued process with marginal distribution μ. The *empirical measure* μ_n is given by $\mu_n := n^{-1}\sum_{i=1}^n \delta_{X_i}$, where δ_x denotes the Dirac measure given by $\delta_x(A) := \mathbf{1}_A(x)$. For a fixed interval $(-\infty, x]$, $x \in \mathbb{R}$, and an ergodic process $(X_i)_{i\in\mathbb{N}}$, $\mu_n((-\infty,,x])$ converges almost surely to $\mu((-\infty,,x])$. This motivates the concept of a *central limit theorem for the empirical process* or an *empirical central limit theorem*. We say that an *empirical central limit theorem* holds if the *empirical process* $\mathrm{U}_n = \big(\mathrm{U}_n(x)\big)_{x\in\mathbb{R}}$ given by

$$\mathrm{U}_n(x) := \sqrt{n}\big(\mu_n((-\infty,x]) - \mu((-\infty,x])\big) \tag{1.1}$$

converges in distribution to a tight centred Gaussian process. Here, we consider U_n as an element of the càdlàg space $\mathbb{D}(\mathbb{R})$.[1]

Empirical CLTs find application in non-parametric statistics. For instance, the empirical CLT can be used to derive the asymptotic distribution of the Kolmogorov–Smirnov goodness-of-fit test. This test is used to determine if a stationary process has some given marginal distribution μ_0, that is to test the hypothesis

$\mathbb{H}_0$: *"the process has marginal distribution μ_0"*,

against the alternative

$\mathbb{H}_A$: *"the process has a marginal distribution different from μ_0"*.

The associated test statistic is

$$D_n = \sup_{x\in\mathbb{R}} \sqrt{n}\big|\mu_n((-\infty,x]) - \mu_0((-\infty,x])\big|.$$

By the continuous mapping theorem, the asymptotic null distribution of D_n is given by $\sup_{x\in\mathbb{R}}|W(x)|$, where W denotes the limit distribution of U_n.

Donsker (1952) showed that in the case of an independently uniformly distributed underlying process $(X_i)_{i\in\mathbb{N}}$, the empirical process $\big(\mathrm{U}_n(x)\big)_{x\in[0,1]}$ converges in distribution to a Brownian bridge process W. An empirical CLT for dependent data was given by Billingsley (1968), who considered φ-mixing (or uniformly strong mixing) processes.

Dudley (1966) and Bickel and Wichura (1971) were among the first to study empirical process CLTs for multi-dimensional independent and identically distributed (i.i.d.) data. Here

[1] The space of right continuous functions with left limits from $\mathbb{R}$ on $\mathbb{R}$.

U_n is given by (1.1), where $x \in \mathbb{R}^d$ and $(-\infty, x]$ denotes the multi-dimensional rectangle $(-\infty, x_1] \times \ldots \times (-\infty, x_d]$. Philipp and Pinzur (1980), Philipp (1984), and Dhompongsa (1984) treat weak convergence of multi-dimensional empirical processes of mixing data.

A process $(X_i)_{i\in\mathbb{N}}$ is called α-mixing (or strongly mixing), if there is a sequence of non-negative real numbers $(\alpha(n))_{n\in\mathbb{N}^*}$ such that $\alpha(n) \to 0$ as $n \to \infty$ and

$$\big|\mathbf{P}(A \cap B) - \mathbf{P}(A)\,\mathbf{P}(B)\big| \le \alpha(n)$$

for any $k \in \mathbb{N}^*$ and every $A \in \sigma(X_0, \ldots, X_k)$ and $B \in \sigma(X_{k+n}, X_{k+n+1}, \ldots)$.[2] φ-mixing is a stronger version of α-mixing and corresponds to the case, where α can be replaced by $\varphi(n)\,\mathbf{P}(A)$ with $\varphi(n) \to 0$ as $n \to \infty$.

Davydov (1970) showed that in the case of an α-mixing process, for every $k \in \mathbb{N}^*$, the covariance of every $\sigma(X_0, \ldots, X_k)$-measurable real-valued function ξ and every $\sigma(X_{k+n}, X_{k+n+1}, \ldots)$-measurable real-valued function η can be bounded by

$$10\,(\mathbf{E}\,|\xi|^s)^{1/s}\,(\mathbf{E}\,|\eta|^m)^{1/m}\,\alpha(n)^{1-(1/s+1/m)} \tag{1.2}$$

for all $s, m \ge 1$ such that $1/s + 1/m < 1$ and $\mathbf{E}\,|\xi|^s$, $\mathbf{E}\,|\eta|^m < \infty$.

We work with a different mixing property. We assume directly that a similar bound like (1.2) with $\alpha(n) = \theta^n$, $\theta \in (0,1)$ is satisfied for a very particular choice of ξ and η, namely $\xi = f(X_{i_0}) \cdot \ldots \cdot f(X_{i_{q-1}})$ with $0 \le i_0 \le \ldots \le i_{q-1} \le k$ and $\eta = f(X_{i_q}) \cdot \ldots \cdot f(X_{i_p})$ with $k+n \le i_q \le \ldots \le i_p$, $q \le p \in \mathbb{N}^*$, where f is an element of some specific normed space $(\mathcal{C}, \|\cdot\|_{\mathcal{C}})$ of real-valued functions. We say that $(X_i)_{i\in\mathbb{N}}$ is *multiple mixing* if there is a $\theta \in (0,1)$ which does not depend on p such that for every $k, q \in \mathbb{N}^*$ the covariance $|\,\mathbf{Cov}(\xi, \eta)|$ can be bounded by

$$(\mathbf{E}\,|f(X_0)|^s)^{1/s}\,\|f\|_{\mathcal{C}}^{\ell}\,Q(i_1 - i_0, \ldots, i_p - i_{p-1})\,\theta^n \tag{1.3}$$

with $s, \ell \ge 1$, uniformly for all $f \in \mathcal{C}$ with $\sup_x |f(x)| \le 1$, where Q is a polynomial function that depends only on p with total degree of Q not larger than some fixed $d_0 \ge 0$. This bound is in many aspects weaker than the bound (1.2) with exponential α, since we strongly restrict the choice of ξ and η and allow that (1.3) grows with the size of the gaps $i_k - i_{k-1}$ and the $\mathcal{C}$-norm $\|f\|_{\mathcal{C}}$ of the function f (in our applications, $\|f\|_{\mathcal{C}}$ is always large).

The multiple mixing property holds for instance in the setting of data arising from Markov chains or dynamical systems with a spectral gap of the corresponding transfer operator on some space of regular functions containing $\mathcal{C}$. Durieu (2008a) established the multiple mixing property of the ergodic torus automorphism.

Dehling, Durieu, and Volný (2009) developed a technique (cf. Section 1.3) to treat the one-dimensional empirical processes of Markov chains with a spectral gap on the space of bounded Lipschitz functions. Dehling and Durieu (2011) extend this result to multidimensional data of multiple mixing data with applications to dynamical systems of ergodic automorphisms

[2] $\sigma\big(\{X_i : i \in I\}\big)$, $I \subset \mathbb{N}$, denotes the σ-algebra generated by $\big\{X_i : i \in I\big\}$.

of the torus and Markov chains with a spectral gap on the space of bounded α-Hölder functions.

In Part II of this thesis, we provide results for processes that satisfy a weaker version of the multiple mixing property which we call *slow multiple mixing*. This concept allows us to treat processes that only have a polynomial decay in (1.3) instead of the exponential decay given by θ^n. We establish an empirical CLT for slowly multiple mixing processes that can be applied to causal functions of i.i.d. processes such as linear processes and time delay vectors.

Dudley (1978) initiated the study of generalized empirical processes, that is the study of the process given by $\sqrt{n}\big(\mu_n(A) - \mu(A)\big)$, where A is an element of some specific subclass of $\mathfrak{B}(\mathbb{R}^d)$.[3] Ossiander (1987) used bracketing conditions to derive CLTs for empirical processes, indexed by classes of L^2 functions. Peškir and Yukich (1994) established function class indexed Glivenko–Cantelli theorems for stationary processes using weighted random entropy conditions.

Let $(X_i)_{i\in\mathbb{N}}$ be a stationary process in a measurable space $(\mathcal{X}, \mathcal{A})$ with marginal distribution μ and let $\mathcal{F}$ be a class of real-valued measurable functions on $\mathcal{X}$. We denote the integral $\int_{\mathcal{X}} f \, d\mu$ with respect to a measure μ by μf. The *$\mathcal{F}$-indexed empirical process* $\mathrm{U}_n = \big(\mathrm{U}_n(f)\big)_{f\in\mathcal{F}}$ is given by

$$\mathrm{U}_n(f) := \sqrt{n}\big(\mu_n(f) - \mu f\big), \quad f \in \mathcal{F}. \tag{1.4}$$

In this setting, we consider U_n as a random element in the space $\ell^\infty(\mathcal{F})$ of bounded real-valued functions on $\mathcal{F}$.[4] We say that an *empirical CLT* or a *CLT for the $\mathcal{F}$-indexed empirical process* holds, if U_n converges in distribution in $\ell^\infty(\mathcal{F})$ to some tight Gaussian process. Ossiander (1987) established a CLT for the empirical process indexed by a class of functions for i.i.d. data. An empirical CLT for α-mixing (or strongly mixing) data has been established by Andrews and Pollard (1994). Doukhan, Massart, and Rio (1995) and Rio (1998) study the empirical processes indexed by a class functions for β-mixing (or absolute regularly mixing) data. For further results, see the survey article of Dehling and Philipp (2002).

In the setting of Markov chains and dynamical systems, the spectral gap method is a useful tool to establish CLTs for processes $(f(X_i))_{i\in\mathbb{N}}$, where f is an element of the function class $\mathcal{B}$ on which one has the spectral gap.

If $\mathcal{F}$ is contained in $\mathcal{B}$, one can deduce the finite-dimensional CLT from the one-dimensional CLT via the Cramér-Wold theorem. If further one has sufficient moment bounds of the partial sums $\sum_{i=1}^n (f(X_i) - \mu f)$, $f \in \mathcal{F}$, to establish the tightness of U_n, one can apply the classical "finite-dimensional convergence plus tightness" approach to prove the empirical CLT for $(\mathrm{U}_n)_{f\in\mathcal{F}}$. Collet, Martinez, and Schmitt (2004) proved an empirical CLT for expanding maps of the unit interval this way.

The technique of Dehling et al. (2009) and Dehling and Durieu (2011) can be applied to situations, where the one-dimensional CLT and the moment bounds are not directly available under the functions $f \in \mathcal{F}$ but for a different class of functions $\mathcal{C}$. This is useful for instance in the setting of a spectral gap on a space that does not contain the indexing class $\mathcal{F}$. As an

[3] $\mathfrak{B}(\mathbb{R}^d)$ denotes the Borel σ-algebra, that is the σ-algebra, generated by the open sets in $\mathbb{R}^d$.

[4] $\ell^\infty(\mathcal{F})$ is equipped with the supremum norm $\|U\|_\infty = \sup_{f\in\mathcal{F}} |U(f)|$.

example, Gouëzel (2009) provides samples of dynamical systems that support a spectral gap on the Lipschitz functions but not on the bounded variation functions.

Dehling et al. (2009) and Dehling and Durieu (2011) treat exclusively classical empirical processes, that is the case, where $\mathcal{F} := \{(-\infty, x] : x \in \mathbb{R}^d\}$. This thesis extends the techniques developed by Dehling et al. (2009) to empirical processes indexed by classes of functions. We establish a CLT for $\mathcal{F}$-indexed empirical processes of multiple mixing data with applications to the ergodic automorphism of the torus. The general results of this work have been used by Durieu (2013) to establish an empirical CLT for iterative Lipschitz models that contract on average.

1.2. Sequential Empirical Central Limit Theorems

Beyond empirical CLTs, Part III of this thesis studies so called *sequential empirical processes* of multiple mixing data. In the classical case of $\mathbb{R}$-valued data, the sequential empirical process is defined as the $\mathbb{R} \times [0,1]$ indexed process $\mathrm{V}_n = (\mathrm{V}_n(x,t))_{(x,t)\in\mathbb{R}\times[0,1]}$ given by

$$\mathrm{V}_n(x,t) = \frac{[nt]}{\sqrt{n}} \sum_{i=1}^{[nt]} \big(\mu_{[nt]}((-\infty, x]) - \mu((-\infty, x])\big),$$

where $[\cdot]$ denotes the lower Gauss bracket, i.e. $[x] := \sup\{z \in \mathbb{Z} : z \leq x\}$. The process V_n is also known as the two-parameter empirical process.

The study of sequential empirical processes was initiated independently by Müller (1970) and Kiefer (1972). 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

$$\mathbf{E}\left(K(x,s)K(y,t)\right) = \min\{s,t\}\big(\mu((-\infty, \min\{x,y\}]) - \mu((-\infty, x])\mu((-\infty, y])\big).$$

The limit process K is called Kiefer process, or Kiefer-Müller process. In the style of the preceding section, we say that a *sequential empirical central limit theorem* holds, if V_n converges in distribution to a tight Gaussian process.

Sequential empirical CLTs play an important role in the statistical analysis of change-point problems. Assume we have a sample $X_1, \ldots, X_n$ of some stochastic process, say of i.i.d. random variables. In most cases there will be some more or less strong oscillation around the expectation of the corresponding random variables. It can happen that the mean value of $X_i, \ldots, X_{k^*}$ differs notably from the mean value of $X_{k^*+1}, \ldots, X_n$ for some $k^* \in \{1, \ldots, n-1\}$. A change point test is used to test the hypothesis that this difference occurred only by chance against the alternative that the underlying process has changed after the first k^* realizations.

Let $(X_i)_{i\in\mathbb{N}}$ be a stochastic process with marginal distributions $(\mu_i)_{i\in\mathbb{N}}$. We want to test the null hypothesis

$\mathbb{H}_0$: *"the process is stationary"*

against the alternative

$\mathbb{H}_A$: *"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".*

For this change point problem, we propose the test statistic

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

where $\mu_{k+1,n}$ denotes the empirical distribution of $X_{k+1}, \ldots, X_n$ given by $(n-k)^{-1}\sum_{i=k+1}^n \delta_{X_i}$ (for convenience, let μ_0 and $\mu_{n+1,n}$ be the constant zero measure). In order to determine the asymptotic null 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

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

If the sequential empirical CLT holds, we obtain under the null hypothesis $\mathbb{H}_0$ that

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

where K is the centred Gaussian process with covariance structure

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

This process is also referred to as a Kiefer process. The proof is given in Chapter 12 (see Lemma 12.1). Now, applying the continuous mapping theorem to the supremum-functional, we obtain the asymptotic null distribution of the test statistic T_n, that is

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

In fact this result remains true for general $\mathcal{F}$-indexed empirical processes, see Proposition 12.1.

Berkes and Philipp (1977) established a sequential empirical CLT for strongly mixing random variables. In a recent paper Dedecker, Merlevède, and Pène (2013) proved a sequential empirical CLT for dynamical systems given by an ergodic automorphism of the torus.

In Part III, we extend our empirical process techniques to the sequential case and develop a general sequential empirical CLT for multiple mixing processes. As results, we present sequential empirical CLTs for Markov chains and dynamical systems with a spectral gap and for ergodic torus automorphisms. These results are given in the setting of a generalized version of sequential empirical processes indexed by a class of functions in the style of (1.4).

1.3. The Approximating Class Approach for Empirical CLTs

The standard approach to prove empirical process CLTs uses techniques that go back to Prohorov (1956). Prohorov's Theorem states that a tight family of probability measures is relatively compact. Thus any subsequence $Y_{n'}$ of a tight sequence of random variables Y_n contains a convergent subsequence $Y_{n''}$. Assume now, that one can show that the finite dimensional distributions of a sequence of processes Y_n converge to the finite dimensional distributions of some process Y. Then the finite dimensional distributions of the limit Z of the convergent subsequence $Y_{n''}$ must coincide with those of Y and thus Z and Y are identically distributed. We conclude that every subsequence of Y_n has a convergent subsequence with limit Y, which implies $Y_n \xrightarrow{d} Y$.

A sufficient condition for tightness of a process $(Y_n)_{n\in\mathbb{N}^*}$ in the càdlàg space $\mathbb{D}([0,1])$ is e.g. that for every $\varepsilon, \eta > 0$, there exist some $a, \delta > 0$ and an integer $n_0 \in \mathbb{N}^*$ such that

$$\sup_{n\geq 1} \mathbf{P}\Big(|Y_n(0)| > a\Big) \leq \eta, \tag{1.5}$$

$$\sup_{n\geq n_0} \mathbf{P}\Bigg(\sup_{|t-s|\leq\delta} |Y_n(t) - Y_n(s)| \geq \varepsilon\Bigg) \leq \eta, \tag{1.6}$$

cf. Theorem 15.5 in Billingsley (1968).

The "finite-dimensional convergence plus tightness" technique can be applied to the case, where the underlying process is i.i.d.. One of the first who applied this technique to dependent data was Billingsley (1968), who considered the classical empirical process of φ-mixing processes of $[0,1]$-valued random variables. In this setting the finite dimensional convergence of the empirical process $(\mathrm{U}_n)_{n\in\mathbb{N}^*}$ given by (1.1) is equivalent to a multidimensional CLT under the class of indicator functions $\mathbf{1}_{[0,t]}$, $t \in [0,1]$, that is

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(\mathbf{1}_{[0,t_1]}(X_i), \ldots, \mathbf{1}_{[0,t_k]}(X_i)\Big) \xrightarrow{d} N(0,\Sigma) \quad \text{for all } k \in \mathbb{N}^*,\ t_1, \ldots, t_k \in [0,1]. \tag{1.7}$$

Here $N(0,\Sigma)$ is a k-dimensional normal distribution with covariance matrix Σ which may depend on $t_1, \ldots, t_k$. Since the application of a measurable function conserves the φ-mixing property of a stochastic process, (1.7) can be directly deduced from the CLT for φ-mixing random variables (cf. e.g. Ibragimov and Linnik (1971)) and the Cramér Wold device.

Billingsley (1968) verifies condition (1.6) to establish the tightness of $(U_n)_{n\in\mathbb{N}^*}$ (note that (1.5) is always satisfied for an empirical process). He uses the following fourth moment bound for sums of φ-mixing random variables (Lemma 1 in Chapter 22 in Billingsley (1968)): If $(\xi_i)_{i\in\mathbb{N}}$ is a φ-mixing process with $|\xi_0| \in [0,1]$ a.s., $\mathbf{E}\,\xi_0 = 0$, and mixing coefficients φ_k such that $\sum_{k=1}^{\infty} k^2 \sqrt{\varphi_k} < \infty$, then

$$\mathbf{E}\Bigg(\Big|\sum_{i=1}^{n} \xi_i\Big|^4\Bigg) \leq K\big(n^2(\mathbf{E}\,\xi_0^2)^2 + n\,\mathbf{E}\,\xi_0^2\big) \quad \text{for some fixed } K > 0. \tag{1.8}$$

Since his proof is given for uniformly distributed random variables—which suffices to establish the empirical CLT for $[0, 1]$-valued random variables with continuous distribution functions—(1.8) yields

$$\mathbf{E}\Big(\big|\mathrm{U}_n(t) - \mathrm{U}_n(s)\big|^4\Big) \le \frac{2K_1}{\varepsilon}(t-s)^2 \quad \text{for all } \varepsilon > 0 \text{ such that } \varepsilon < n|t-s|. \tag{1.9}$$

As a result of some maximal inequality (Theorem 12.2 in Billingsley (1968)) which involves an application of Markov's inequality on the fourth moments of the increments of $\mathrm{U}_n(t)$ with respect to t, Billingsley shows that there is a $\delta > 0$ and a $n_0 \in \mathbb{N}^*$ such that

$$\mathbf{P}\bigg(\sup_{t\in[s,s+\delta]} \big|\mathrm{U}_n(t) - \mathrm{U}_n(s)\big| \ge 4\varepsilon\bigg) < \eta\delta$$

for $s \in [0, 1]$ and all $n \in \mathbb{N}^*$, which implies (1.6).

The Approximating Class Approach of Dehling, Durieu, and Volný

In the classical approach, one uses finite-dimensional convergence and bounds for increments of the empirical process indexed by the class of indicators of rectangles. However, there are situations where the aforementioned properties are not directly available, but where one can establish a finite dimensional CLT and moment bounds under a class of regular functions. This occurs e.g. in the case of data arising from Markov chains or dynamical systems, for which the Markov or Perron–Frobenius operator, respectively, has a spectral gap on a specific class of functions (cf. Section 2.1 and Section 10.2).

The technique of Dehling et al. (2009) is qualified for such situations. It is based on the assumption that a CLT and specific moment bounds hold for functionals of the underlying process under some class of regular functions. Using a modified theorem of Billingsley (1968) (Theorem 4.2), they obtain the (classical) empirical CLT without having to establish the finite-dimensional convergence and tightness of the empirical process directly.

Billingsley's Theorem 4.2 applies to random variables ξ_n, $\xi_n^{(q)}$, $\xi^{(q)}$, ξ (where $n, q \in \mathbb{N}^*$) with values in a separable metric space (S, ρ) satisfying

(a) $\xi_n^{(q)} \xrightarrow{d} \xi^{(q)}$ as $n \to \infty$, for all $q \ge 1$,

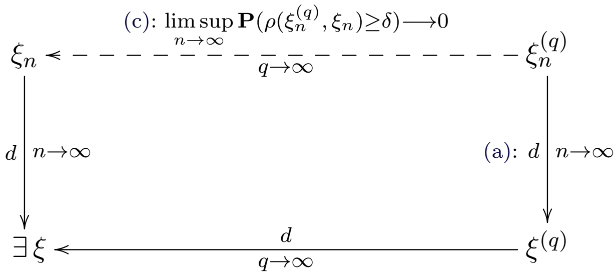
(b) $\xi^{(q)} \xrightarrow{d} \xi$ as $q \to \infty$ and

(c) $\limsup_{n\to\infty} \mathbf{P}\big(\rho(\xi_n^{(q)}, \xi_n) \ge \delta\big) \longrightarrow 0$ as $q \to \infty$ for all $\delta > 0$.

It states that under these conditions ξ_n converges in distribution to ξ.

Dehling et al. (2009) proved that this result holds without condition (b), provided that S is a complete separable metric space (cf. Theorem 2 in Dehling et al. (2009)). More precisely, they could show that in this situation (a) and (c) together imply the existence of a random variable ξ satisfying (b), and thus by Billingsley's theorem $\xi_n \xrightarrow{d} \xi$.

The theorem is illustrated by the following diagram.



Now, let $(X_i)_{i\in\mathbb{N}}$ be a stationary process with values in a measurable space $(\mathcal{X}, \mathcal{A})$ and let $\mathcal{F}$ be a uniformly bounded class of real-valued measurable functions on $\mathcal{X}$. Assume that U_n is measurable, and that there is a uniformly bounded class $\mathcal{G}$ of real-valued measurable functions on $\mathcal{X}$ with the following properties.

(1.A) $\mathrm{U}_n(f) \xrightarrow{d} N(0, \sigma_f^2)$ for every $f \in \mathrm{Vect}(\mathcal{G})$.

(1.B) For every $q \in \mathbb{N}^*$ there is a function $\pi_q : \mathcal{F} \longrightarrow \mathcal{G}$ with $\#\pi_q(\mathcal{F}) < \infty$ such that $\xi_n^{(q)} = (\mathrm{U}_n(\pi_q f))_{f\in\mathcal{F}}$ and $\xi_n = (\mathrm{U}_n(f))_{f\in\mathcal{F}}$ satisfy (c).[5]

In this situation one can show that (a) holds for some piecewise constant Gaussian process $\xi^{(q)}$ (cf. Proposition 3.1). By Dehling et al. (2009) if the state space S of $\xi_n^{(q)}$ and ξ_n is separable and complete this implies that $\xi^{(q)}$ converges in distribution and that ξ_n converges to the same limit ξ as $\xi^{(q)}$. As the $\xi^{(q)}$ are Gaussian, their limit ξ must also be Gaussian and we conclude that the empirical CLT holds (cf. Section 3.2).

An Approximating Class Approach for Empirical Processes Indexed by Classes of Functions

Dehling et al. (2009) and Dehling and Durieu (2011) only consider classical empirical processes, i.e. empirical processes indexed by a class $\mathcal{F} = \{\mathbf{1}_{[-\infty,x]} : x \in \mathbb{R}^d\}$ of semi-finite rectangles in $[-\infty,\infty]^d$. As in this case U_n takes values in the space $(\mathbb{D}([-\infty,\infty]^d), d_S)$ of multidimensional *càdlàg*[6] functions on $[-\infty,\infty]^d$ equipped with the corresponding Skorokhod metric d_S, they can set $(S,\rho) = (\mathbb{D}([-\infty,\infty]^d), d_S)$. In this case, the separability of S and the measurability of U_n (w.r.t. the Borel σ-algebra on $\mathbb{D}([-\infty,\infty]^d)$) are always guaranteed.

In the present thesis however, we generalize to arbitrary uniformly bounded indexing sets $\mathcal{F}$ and consider U_n as a random element in the non-separable space $S = \ell^\infty(\mathcal{F})$ of uniformly bounded real functions on $\mathcal{F}$ equipped with the supremum norm $\|\cdot\|_\infty$. It is well known that, in general, $\mathrm{U}_n = (\mathrm{U}_n(f))_{f\in\mathcal{F}}$ is not measurable and thus the usual theory of weak convergence

[5] $\#A$ denotes the cardinality of a set A.

[6] For definition of $(\mathbb{D}([-\infty,\infty]^d), d_S)$ see Neuhaus (1971, p.1286 ff.). Note that $(\mathbb{D}([-\infty,\infty]^d), d_S)$ is a complete and separable space (more precisely, Neuhaus (1971) and Straf (1972) proved this for the space $\mathbb{D}([0,1]^d)$, but—since $[0,1]$ and $[-\infty,\infty]$ are homeomorphic—the metric on $\mathbb{D}([-\infty,\infty]^d)$ can be naturally extended to a metric on $\mathbb{D}([-\infty,\infty]^d)$ which conserves all relevant properties (cf. Dehling and Durieu (2011, p.1081f.))).

of random variables does not apply. We therefore use the theory based on convergence of outer expectations; see van der Vaart and Wellner (1996).

We call any (not necessarily measurable) function on a probability space $(\Omega, \mathfrak{S}, \mathbf{P})$ a *random element*. If it is also measurable we call it a *random variable*. For a real-valued random element X, the *outer expectation* or *outer integral* $\mathbf{E}^* X$ is defined as

$$\mathbf{E}^* X := \inf\big\{\mathbf{E}\, Y : Y \text{ is measurable, } Y \geq X, \text{ and, } \mathbf{E}\, Y \text{ exists}\big\}.$$

By Lemma 1.2.1 in van der Vaart and Wellner (1996) for a random element X there exists a *measurable cover function* X^* satisfying $X^* \geq X$ and $\mathbf{E}^* X = \mathbf{E}\, X^*$, provided that $\mathbf{E}^* X < \infty$. The *outer probability* $\mathbf{P}^*$ of a probability measure $\mathbf{P}$ is defined as

$$\mathbf{P}^*(A) := \inf\big\{\mathbf{P}(B) : B \in \mathfrak{S}, A \subset B, \big\} \quad A \subset \Omega.$$

$\mathbf{P}^*$ coincides with $\mathbf{P}$ on $\mathfrak{S}$ and we have that $\mathbf{P}^*(A) = \mathbf{E}^*(\mathbf{1}_A)$ for all $A \subset \Omega$ (cf. Lemma 1.2.3 in van der Vaart and Wellner (1996)).

Let (S, ρ) be a metric space equipped with the Borel σ-algebra generated by the open sets. We say that a sequence of S-valued random elements $(\xi_n)_{n\in\mathbb{N}}$ *converges in distribution* to a measurable random element ξ (write $\xi_n \xrightarrow{d} \xi$), if

$$\mathbf{E}^*(\varphi(\xi_n)) \to \mathbf{E}(\varphi(\xi)), \quad \text{as } n \to \infty$$

for all bounded and continuous functions $\varphi : S \to \mathbb{R}$.

For more details and properties of random elements see the book of van der Vaart and Wellner (1996).

To deal with non-measurable processes in the setting of the theory of outer expectation, in our paper Dehling, Durieu, and Tusche (2012) we established the following theorem which can be seen as an extension of the theorem of Billingsley (1968) and its adaptation by Dehling et al. (2009) discussed earlier.

Theorem 1.1. *Let $\xi_n, \xi_n^{(q)}, \xi^{(q)}$ (where $n, q \geq 1$) be random elements with values in a complete metric space (S, ρ), and suppose that $\xi^{(q)}$ is measurable and separable, i.e. there is a separable set $S^{(q)} \subset S$ such that $\mathbf{P}(\xi^{(q)} \in S^{(q)}) = 1$. If the conditions*

$$\xi_n^{(q)} \xrightarrow{d} \xi^{(q)} \qquad \text{as } n \to \infty, \quad \text{for all } q \geq 1, \tag{1.10}$$

$$\limsup_{n\to\infty} \mathbf{P}^*\big(\rho(\xi_n, \xi_n^{(q)}) \geq \delta\big) \longrightarrow 0 \quad \text{as } q \to \infty, \text{ for all } \delta > 0 \tag{1.11}$$

are satisfied, then there exists an S-valued, separable random variable ξ such that $\xi^{(q)} \xrightarrow{d} \xi$ as $q \to \infty$, and

$$\xi_n \xrightarrow{d} \xi \quad \text{as } n \to \infty.$$

The proof is postponed to the appendix (see Section A.4).

In the classical approach, $\mathcal{F}$-indexed empirical process CLTs require some control of the size indexing class $\mathcal{F}$ to establish the tightness. This size is usually measured by *covering* or *bracketing numbers*. Let $(\mathcal{V}, \|\cdot\|_\mathcal{V})$ be a normed space of real-valued measurable functions defined on a probability space $(\mathcal{X}, \mathcal{A}, \mu)$. Usually, the covering number of $\mathcal{F}$ is defined as the minimum number $N_c(\varepsilon, \mathcal{F}, \|\cdot\|_\mathcal{V})$ of $\|\cdot\|_\mathcal{V}$-balls of radius ε needed to cover $\mathcal{F}$. The classical bracketing number $N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|_\mathcal{V})$ is analogously defined as the smallest number of $\|\cdot\|_\mathcal{V}$-brackets of size ε that cover $\mathcal{F}$. Here, a $\|\cdot\|_\mathcal{V}$-bracket is defined a set $[l,u] := \{f : \mathcal{X} \to \mathbb{R} : l \leq f \leq u\}$ with $l \leq u \in \mathcal{F}$ and the $\|\cdot\|_\mathcal{V}$-size of $[l,u]$ is given by $\|u-l\|_\mathcal{V}$.[7] Covering and bracketing conditions are frequently formulated via the existence of the integral of certain functionals of the corresponding covering or bracketing numbers, respectively. They are usually addressed to as *entropy* conditions. For instance, van der Vaart and Wellner (1996) provide an empirical CLT for the i.i.d. case under the condition that $\int_0^\infty \sqrt{\log(N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|_2))}\, d\varepsilon < \infty$, where $\|\cdot\|_2$ is given by $\|f\|_2 = (\int_\mathcal{X} |f|^2 \, d\mu)^{1/2}$. This theorem essentially corresponds to an earlier result of Ossiander (1987), who considered indexing classes of functions.

We apply the approximating class approach introduced above. To approximate the indexing class $\mathcal{F}$, we use the intersection $\mathcal{G}$ of some $\|\cdot\|_\infty$-ball with some normed vector space $(\mathcal{C}, \|\cdot\|_\mathcal{C})$ of real-valued, measurable functions on $\mathcal{X}$. As discussed earlier in this section, in order to apply the approximating class technique, we need to show that for every $m \in \mathbb{N}^*$ there is a function $\pi_m : \mathcal{F} \longrightarrow \mathcal{G}$ taking only finitely many values, such that for all $\varepsilon > 0$

$$\limsup_{n\to\infty} \mathbf{P}^*\Big(\sup_{f\in\mathcal{F}} \big|\mathrm{U}_n(\pi_m f) - \mathrm{U}_n(f)\big| > \varepsilon\Big) \longrightarrow 0 \quad \text{as } m \to \infty. \tag{1.12}$$

To establish this property, we apply chaining arguments, eventually allowing us to estimate the probability in (1.12) with the help of moment bounds of expressions of the type $U_n(g)$, $g \in \mathcal{G} \cup (\mathcal{G} - \mathcal{G})$. Our technique is designed to work with processes $(X_i)_{i\in\mathbb{N}}$ that are multiple mixing with respect to $\mathcal{C}$ (cf. Chapter 2). In this situation, for every $p \in \mathbb{N}^*$ one has a moment bound of the $2p$th moments of $\mathrm{U}_n(g)$ with $g \in \mathcal{G} \cup (\mathcal{G} - \mathcal{G})$, cf. Proposition 2.1. These bounds involve logarithmic terms of the $\mathcal{C}$-norm of g. We therefore use a different kind of covering numbers that measure how well $\mathcal{F}$ can be approximated by the function class $\mathcal{G}$ in $\mathrm{L}^s(\mu)$-norm, with a simultaneous control of the $\|\cdot\|_\mathcal{C}$-size of the approximating functions.

We developed the following adapted notion of bracketing. Let $\varepsilon, A > 0$, $s \geq 1$ and $\|\cdot\|_s$ be given by $\|f\|_s = (\int_\mathcal{X} |f|^s \, d\mu)^{1/s}$. For a class $\mathcal{G}$ of measurable real-valued functions defined on $\mathcal{X}$, we call a set $\{f : \mathcal{X} \longrightarrow \mathbb{R} : l \leq f \leq u\}$ an $(\varepsilon, A, \mathcal{G}, \|\cdot\|_s)$-bracket, if $l \leq u \in \mathcal{G}$, $\max\{\|u\|_\mathcal{B}, \|l\|_\mathcal{B}\} \leq A$ and $\|u-l\|_s \leq \varepsilon$. We define the bracketing number $N(\varepsilon, A, \mathcal{F}, \mathcal{G}, \|\cdot\|_s)$ of $\mathcal{F}$ as the smallest number of $(\varepsilon, A, \mathcal{G}, \|\cdot\|_s)$-brackets that are needed to cover $\mathcal{F}$. In applications, we usually require an entropy condition which involves these bracketing numbers with A given by an exponential term of the reciprocal value of the first argument of the bracket. Examples for classes $\mathcal{F}$ that satisfy such a condition are provided in Chapter 4.

Our definition is close to the definition of bracketing numbers given by Ossiander (1987) (see

[7]for two real valued functions f, g on $\mathcal{X}$, we wrtite $f \leq g$ if $f(x) \leq g(x)$ for all $x \in \mathcal{X}$.

also van der Vaart and Wellner (1996)), but different in a vital point. In Ossiander (1987), no assumptions are made on the upper and lower functions of the bracket other than that they are close in L^2. Here, our moment bound (2.2) below forces us to require the extra condition that u and l belong to the class $\mathcal{G}$ and that their $\mathcal{C}$-norms are controlled. Obviously, our bracketing numbers are always larger than the ones defined in Ossiander (1987) and naturally our condition on the size of $\mathcal{F}$ are stronger. On the other hand, our results apply to dependent data, while Ossiander (1987) and van der Vaart and Wellner (1996) treat i.i.d. data.

1.4. Overview of Main Results

In this thesis, we establish a collection of abstract empirical and sequential empirical central limit theorem. These theorems are adapted for the situation where the underlying process satisfies a *multiple mixing* condition. We call a stationary process $(X_i)_{i\in\mathbb{N}}$ *multiple mixing* with respect to some normed vector space $(\mathcal{C}, \|\cdot\|_{\mathcal{C}})$ of real-valued functions, defined on the state space of $(X_i)_{i\in\mathbb{N}}$, if there are an $s \geq 1$ and a $\theta \in (0,1)$ such that for any $p \in \mathbb{N}^*$, all covariances of the type

$$\mathbf{Cov}\Big(f(X_{i_0})\cdots f(X_{i_{q-1}}), f(X_{i_q})\cdots f(X_{i_p})\Big)$$

can be bounded by a term

$$\|f\|_s \|f\|_{\mathcal{C}}^{\ell} Q(i_1 - i_0, \dots, i_p - i_{p-1})\theta^{i_q - i_{q-1}} \tag{1.13}$$

uniformly for all $f \in \mathcal{C}$ with $\mathbf{E}\, f(X_0) = 0$ and $|f(X_0)| \leq 1$ *almost surely*, $i_0 \leq \ldots \leq i_p \in \mathbb{N}$, and $q \leq p$ (cf. Definition 2.1). Here Q is some polynomial, the exponent ℓ is greater or equal to 1 and $\|\cdot\|_s$ denotes the L^s-norm given by $\|f\|_s = (\mathbf{E}\,|f(X_0)|^s)^{1/s}$.

Examples for such processes are $\mathcal{B}$-geometrically ergodic Markov chains, iterative Lipschitz models that contract on average, dynamical systems with a spectral gap on the Perron–Frobenius operator and ergodic automorphisms of the torus. These processes are introduced in Chapter 2.

In Part I we establish the following CLT for empirical processes of multiple mixing data that are indexed by a uniformly bounded class of functions $\mathcal{F}$.

Theorem 1.2 (An Empirical Central Limit Theorem for Multiple Mixing Processes)**.** *Let $(X_i)_{i\in\mathbb{N}}$ be multiple mixing with respect to some normed vector space $\mathcal{C}$ of measurable functions with $s \geq 1$ and total degree of the multivariate polynomial Q not larger than d_0. If $\big(f(X_i)\big)_{i\in\mathbb{N}}$ satisfies the central limit theorem for all $f \in \mathcal{C}$ and if there exists some uniformly bounded subclass $\mathcal{G} \subset \mathcal{C}$ such that for some $\gamma > \max\{1, d_0\}$ the entropy condition*

$$\int_0^1 \varepsilon^r \sup_{\varepsilon\leq\delta\leq 1} N^2\big(\delta, \exp(C\delta^{-1/\gamma}), \mathcal{F}, \mathcal{G}, \|\cdot\|_s\big) d\varepsilon < \infty \qquad \text{for some } r > -1, \text{ and } C > 0 \tag{1.14}$$

holds, then the empirical process $U_n = (U_n(f))_{f\in\mathcal{F}}$ converges in distribution in $\ell^\infty(\mathcal{F})$ to a tight centred Gaussian process W.

Under slightly stronger assumptions, involving an exponential decay of covariances of $\varphi(X_0)$ and $f(X_k)$ with $f \in \mathcal{C}$ and $\varphi \in \mathcal{F} \cup (\mathcal{F} - \mathcal{G})$ as $k \to \infty$, we can also determine the exact covariance structure of the limit process W.

We show that property (1.14) holds under reasonable assumption on the distribution function of X_0 for the classes of indicators of finite and semi-finite rectangles, bounded balls and ellipsoids, balls of arbitrary metric, and some parametric class of monotone functions, where we use some bounded subclass $\mathcal{G}$ of the real-valued bounded Hölder continuous functions.

Among others, we have the following statements about our entropy condition (1.14). Let μ denote a measure on $\mathbb{R}^d$ with distribution function F and let $\mathcal{G}$ be the class of real-valued α-Hölder functions on $\mathbb{R}^d$ that are uniformly bounded by 1.

(i) If there are some $s \in [1, \infty]$ and $\gamma > 1$ such that the modulus of continuity ω_{F} of F satisfies $\omega_{\mathrm{F}}(x) = \mathcal{O}(|\log(x)|^{-s\gamma})$ as $x \to \infty$,[8] then condition (1.14) holds (with the same s and γ) for $\mathcal{F} := \{\mathbf{1}_{(t,u]} : t, u \in \mathbb{R}^d, t \le u\}$, the class of indicators of finite and semi-finite rectangles on $\mathbb{R}^d$.

(ii) Let $\mathcal{F} := \{\mathbf{1}_{E(x,r)} : x \in [0,1]^d, r \in [0, D]^d\}$ for any fixed $D > 0$, where $E(x, r)$ denotes the ellipsoid $E(x, r) := \{y \in \mathbb{R}^d : \sum_{i=1}^d (x_i - y_i)^2 / r_i^2 \le 1\}$. If μ has a bounded density with respect to the Lebesgue measure, then condition (1.14) is satisfied for all $s \in [1, \infty], \gamma > 1$.

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

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

- $0 \le f_t(x) \le 1$ for all $t \in [0, 1]$ and $x \in \mathbb{R}$,
- $f_s \le f_t$ for all $0 \le s \le t \le 1$,
- f_t is monotone increasing on $\mathbb{R}$ for all $t \in [0, 1]$, and
- $G_\mu(t) = \mu f_t$ is Lipschitz,

If there are some $s \in [1, \infty]$ and $\gamma > 1$ such that the distribution function F of μ satisfies $\omega_{\mathrm{F}}(x) = \mathcal{O}(|\log(x)|^{-s\gamma})$ as $x \to \infty$, then (1.14) holds (with the same s and γ).

As an application we prove an empirical CLT for ergodic automorphisms of the torus. Let $\mathbb{T}^d = \mathbb{R}^d / \mathbb{Z}^d$ denote the multidimensional torus identified with $[0, 1]^d$ and let A denote a $d \times d$ matrix with integer coefficients, determinant ± 1, and with no eigenvalue that is a root of unity. Then $T : \mathbb{T}^d \longrightarrow \mathbb{T}^d$ given by $Tx = Ax \mod 1$ induces a measure preserving ergodic automorphism on $\mathbb{T}^d$ equipped with the Lebesgue measure λ.

We establish the empirical CLT for processes $(\varphi(T^i))_{i \in \mathbb{N}^*}$ with a Hölder function φ.

Theorem 1.3 (An Empirical CLT for Ergodic Automorphisms of the Torus)**.** *If* (1.14) *holds for some uniformly bounded subspace $\mathcal{G}$ of $\mathcal{C}$ with d_0 equal to the size of the biggest Jordan block*

[8] Recall that $\omega_F(t) := \sup\{|F(x) - F(y)| : |x - y| \le t\}$, where $|\cdot|$ denotes the corresponding Euclidean norm.

of A restricted to its neutral subspace, then U_n *given by*

$$\mathrm{U}_n(f) = \frac{1}{\sqrt{n}} \sum_{i=1} \left(f \circ \varphi(T^i) - \int f \circ \varphi \, d\lambda \right) \quad f \in \mathcal{F}$$

satisfies the empirical CLT.

This result can be applied for any of the indexing classes given in (i)-(iv) under different assumptions on the distribution $\lambda \circ \varphi^{-1}$. For instance, we show that under the assumption that the distribution function F of $\lambda \circ \varphi^{-1}$ satisfies $\omega_{\mathrm{F}}(x) = \mathcal{O}(|\log(x)|^{-\gamma})$ for some $\gamma > \max\{1, d_0\}$, the empirical process indexed by the class of finite and semi-finite rectangles given in (i) satisfies the empirical CLT. Under the stronger assumption that $\lambda \circ \varphi^{-1}$ has a bounded density, we establish an empirical CLT for the empirical process indexed by the class of bounded ellipsoids, given in (ii).

In Part II, we extend the concept of multiple mixing to the situation, where $\theta^{i_q - i_{q-1}}$ in the covariance bound (1.13) is replace by $\Theta(i_q - i_{q-1})$ for some non-negative function Θ with $\sum_{i=1}^{\infty} \Theta(i) < \infty$ (cf. Definition 6.1). We call a process that satisfies this condition *slowly multiple mixing.* In order to treat slow multiple mixing processes we restrict to empirical processes indexed by semi-finite rectangles. We have the following result for slowly multiple mixing processes processes.

Theorem 1.4 (An Empirical CLT for Slowly Multiple Mixing Processes)**.** *Let $(X_i)_{i \in \mathbb{N}}$ be a slowly multiple mixing process on $\mathbb{R}^d$ with respect to the space of bounded real-valued α-Hölder continuous functions with $s \geq 1$ and $\Theta : \mathbb{N} \longrightarrow \mathbb{R}_0^+$ such that there exists a $p \in \mathbb{N}^*$ with $p > sd$ and $\sum_{i=0}^{\infty} i^{2p-2} \Theta(i) < \infty$. If, for every bounded $\mathbb{R}$-valued α-Hölder continuous function f, the process $(f(X_i))_{i \in \mathbb{N}^*}$ satisfies the CLT, and if the distribution function* F *of X_0 is β-Hölder with $\beta > \alpha sp/(p - sd)$, then there is a centred Gaussian process $W = (W(t))_{t \in [-\infty,\infty]^d}$ with almost surely continuous sample paths such that $\mathrm{U}_n \xrightarrow{d} W$ in the space $\mathbb{D}([-\infty, \infty]^d)$.*

This theorem applies e.g. to 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 the process $(X_i)_{i \in \mathbb{N}}$ given by $X_i := G((\xi_j)_{j \leq i})$, where G is a measurable $\mathbb{R}^d$-valued function. The dependence structure of $(X_i)_{i \in \mathbb{N}}$ can be measured by the physical dependence measure due to Dedecker and Prieur (2005) (see also Wu (2005)), which is given by

$$\delta_{i,m} := \left(\mathbf{E} \left| X_i - G(\xi_i, \xi_{i-1}, \ldots, \xi_1, \xi'_0, \xi'_{-1}, \ldots) \right|^m \right)^{\frac{1}{m}},$$

where $(\xi'_i)_{i \in \mathbb{Z}}$ is an independent copy of $(\xi_i)_{i \in \mathbb{Z}}$ and $m \geq 1$. We show, that $(G((\xi_j)_{j \leq i}))_{i \in \mathbb{N}^*}$ satisfies the empirical CLT, given that the assumptions of Theorem 1.4 hold with $\Theta(i) = (\delta_{i,(s-1)/s})^{\alpha}$.

As concrete examples we apply this result to linear processes and time delay vectors. Let $(a_j)_{j \in \mathbb{N}}$ be a family of linear operators from the state space $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ of the ξ_i to $\mathbb{R}^d$. For a linear process $X_i := \sum_{j=0}^{\infty} a_j(\xi_{i-j})$, the condition on the dependence measure $\delta_{i,m}$ reduce to

$$\sum_{j=i}^{\infty} \sup\{|a_j(x)| : x \in \mathcal{X}, \|x\|_{\mathcal{X}} \leq 1\} = \mathcal{O}(i^{-b}) \quad \text{with } b > \min_{p\in\mathbb{N},\ p>sd} \frac{s}{\theta} \frac{(2p-1)p}{p-sd}.$$

The last part of this thesis is dedicated to the study of sequential versions of empirical processes of multiple mixing data. Here, we focus again on exponential decay of the covariances, but also give an abstract theorem that can be applied in the case of slower mixing rates. For exponential multiple mixing, it turns out that the only assumption that significantly changes, is the CLT assumption on $\big(f(X_i)\big)_{i\in\mathbb{N}}$. Here we need the stronger assumption that

$$\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{[nt_1]}(f_1(X_i) - \mu f_1)\ , \ \ldots \ , \ \sum_{i=1}^{[nt_k]}(f_k(X_i) - \mu f_k)\right) \xrightarrow{d} N(0,\Sigma) \quad \text{as } n \to \infty \tag{1.15}$$
$$\text{for all } k \in \mathbb{N}^*,\ f_1,\ldots,f_k \in \mathcal{C},\ t_1,\ldots,t_k \in [0,1].$$

Our sequential empirical CLT is the following.

Theorem 1.5 (A Sequential Empirical CLT for Multiple Mixing Processes)**.** *Let $(X_i)_{i\in\mathbb{N}}$ be multiple mixing with respect to some normed vector space $\mathcal{C}$ of measurable functions with $s \geq 1$, total degree of the multivariate polynomial Q not larger than d_0. If further the sequential finite-dimensional CLT (1.15) holds and if there exists some uniformly bounded subclass $\mathcal{G} \subset \mathcal{C}$ such that the entropy condition (1.14) holds, then the sequential empirical process $\mathrm{V}_n = (\mathrm{V}_n(f,t))_{(f,t)\in\mathcal{F}\times[0,1]}$ given by*

$$\mathrm{V}_n(f,t) := \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} \big(f(X_i) - \mathbf{E}\, f(X_0)\big), \quad (f,t) \in \mathcal{F} \times [0,1],$$

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

We provide a mixing condition, that allows to deduce property (1.15) from the one-dimensional CLT (cf. Lemma 11.1).[9] This condition holds for the ergodic automorphism of the torus without additional assumptions and thus we can extend our empirical CLT for the ergodic torus automorphism Theorem 1.3 to a sequential empirical CLT, where our assumptions are the same as in the non-sequential version.

Further we give a direct proof of condition (1.15) for $\mathcal{B}$-geometrically ergodic Markov chains and dynamical systems with a spectral gap on the Perron–Frobenius operator (cf. Theorem 10.1). As a result, we can also apply Theorem 1.5 in this situations.

Let $(X_i)_{i\in\mathbb{N}}$ be a Markov chain with Markov operator P and invariant measure ν and let $(\mathcal{B}, \|\cdot\|_{\mathcal{B}})$ denote a complex Banach space of $\mathbb{C}$-valued functions on the state space of the X_i. We assume that $\mathcal{B}$ satisfies the following conditions:

[9] Note that since we consider dependent data, this property can not be directly computed from the one-dimensional CLT using the Cramér-Wold device.

(a) $\mathbf{1}_\mathcal{X} \in \mathcal{B}$, $|f|$ and $\overline{f} \in \mathcal{B}$ for all $f \in \mathcal{B}$, and for every point x in the domain of the functions in $\mathcal{B}$, the mapping $f \mapsto f(x)$ is continuous on $\mathcal{B}$.

(b) There is an $m \in [1, \infty]$ and some $K > 0$ such that $(\int |f|^m \, d\nu)^{1/m} < K\|f\|_\mathcal{B}$ for all $f \in \mathcal{B}$.

(c) $\mathcal{B}$ is a Banach algebra, that is the inner multiplication "$\cdot$" satisfies $\|f \cdot g\|_\mathcal{B} \leq \|f\|_\mathcal{B} \cdot \|g\|_\mathcal{B}$.

We say, that $(X_i)_{i\in\mathbb{N}}$ is $\mathcal{B}$-geometrically ergodic if there is some $\kappa > 0$, $\theta \in [0, 1)$ such that

$$\|P^n f - \mathbf{E}\, f(X_0)\|_\mathcal{B} \leq \kappa \|f\|_\mathcal{B} \theta^n \quad \text{for all } f \in \mathcal{B}. \tag{1.16}$$

This property corresponds to a spectral gap of P on $\mathcal{B}$.[10] We have the following sequential empirical CLT for such Markov chains.

Theorem 1.6 (A Sequential Empirical CLT for $\mathcal{B}$-Geometrically Ergodic Markov Chains)**.** *If $(X_i)_{i\in\mathbb{N}}$ is a $\mathcal{B}$-geometrically ergodic Markov chain and if there is some uniformly bounded subset $\mathcal{G}$ of the real-valued functions in $\mathcal{B}$ such that entropy condition (1.14) holds with $s = m/(m-1)$ for some $r > -1$, $\gamma > 1$, and $C > 0$, then the sequential empirical V_n process converges in distribution in $\ell^\infty(\mathcal{F} \times [0, 1])$ to a tight Gaussian process K with covariance structure given by*

$$\mathbf{Cov}\big(K(f, t), K(g, u)\big) = \min\{t, u\} \left\{ \sum_{k=0}^{\infty} \mathbf{Cov}\big(f(X_0), g(X_k)\big) + \sum_{k=1}^{\infty} \mathbf{Cov}\big(f(X_k), g(X_0)\big) \right\}.$$

As an application, we consider iterative Lipschitz models. An iterative Lipschitz model is a Markov chain $(X_i)_{i\in\mathbb{N}^*}$ with a transition probability P of the form

$$P(x, A) = \sum_{i=0}^{\infty} p_i(x)\, \mathbf{1}_A(T_i(x)),$$

where $\{T_i : i \in \mathbb{N}\}$ is a family of Lipschitz continuous transformations of the state space of the X_i and $\{p_i : i \in \mathbb{N}\}$ is a family of $[0, 1]$-valued Lipschitz functions on that same space such that $\sum_{i=0}^{\infty} p_i = 1$. We say that the Lipschitz model $(X_i)_{i\in\mathbb{N}}$ contracts on average, if exists a $\rho \in (0, 1)$ such that $\sum_{i=0}^{\infty} d(T_i(x), T_i(y)) p_i(x) < \rho d(x, y)$ for all x, y. Under certain technical assumptions, such processes satisfy property (1.16) with respect to the space of weighted Lipschitz functions. This space satisfies the conditions (a)–(c). Further it contains a space of bounded Hölder continuous functions entropy results apply.

The techniques used to establish Theorem 1.6 also apply in the case of a measure preserving dynamical system. In this case, we consider the corresponding Perron–Frobenius operator given instead of the Markov operator. Apart from that, we obtain the sequential empirical CLT under the same conditions as in Theorem 1.6.

[10] I.e. the constant functions are the only eigenvectors with eigenvalue of modulus 1 and all other eigenvalues are contained in a disc of radius strictly smaller 1.

1.5. Structure of this Thesis

The remainder of this thesis is structured as follows.

In Chapter 2, we introduce the concept of multiple mixing process. We establish an increment bound of the $2p$-th moments for multiple mixing processes and present some examples of such processes.

In Part I, we establish a CLT for $\mathcal{F}$-indexed empirical processes of multiple mixing data (Chapter 3). We discuss a range of indexing classes that satisfy our entropy condition (Chapter 4). As an example we demonstrate the application of our empirical CLT to dynamical systems given by an ergodic automorphism of the torus (Chapter 5). This part of the thesis corresponds mainly to the article Dehling, Durieu, and Tusche (2012).

Processes with a weaker mixing property (called slowly multiple mixing here) are discussed in Part II. We provide an empirical CLT (Chapter 7) with applications to causal functions such as linear processes and time delay vectors (Section 6.2). This part is based on the article Durieu and Tusche (2012).

Part III contains the results from the article Dehling, Durieu, and Tusche (2013). Here, we provide a sequential empirical CLT for multiple mixing data (Chapter 9) with applications to $\mathcal{B}$-geometrically ergodic Markov chains, dynamical systems with a spectral gap on the Perron–Frobenius operator (Chapter 10) and ergodic automorphisms of the torus (Chapter 11).

2. Multiple Mixing Processes

Let $(X, \mathcal{A})$ be a measurable space. For a positive measure λ on $\mathcal{X}$ and a λ-integrable complex-valued function f on $\mathcal{X}$, we 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 λ-integrable complex-valued functions on $\mathcal{X}$. This space is equipped with the norm $\|f\|_s = \big(\lambda(|f|^s)\big)^{1/s}$. Further, we denote the space of essentially bounded complex-valued measurable functions on $\mathcal{X}$ with respect to λ by $\mathrm{L}^\infty(\lambda)$ 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 manuscript it will always be clear which measure we refer to.

Let $(X_i)_{i\in\mathbb{N}}$ be a stationary stochastic process with state space $(\mathcal{X}, \mathcal{A})$ and marginal distribution μ and let $(\mathcal{E}, \|\cdot\|_{\mathcal{E}})$ denote a real or complex normed vector space of functions on $\mathcal{X}$ with values in $\mathbb{R}$ or $\mathbb{C}$, respectively.

Definition 2.1 (Multiple Mixing)**.** We say that a process $(X_i)_{i\in\mathbb{N}}$ is *multiple mixing* with respect to $\mathcal{E}$ if there exist a $\theta \in (0, 1)$, as $s \geq 1$, and an integer $d_0 \in \mathbb{N}$ such that for all $p \in \mathbb{N}^*$ there is an integer ℓ and a multivariate polynomial Q of total degree not larger than d_0 such that

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

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

An important property of multiple mixing processes is that, for all $p \in \mathbb{N}^*$, they allow some increment bound of the $2p$-th moments of $\mathrm{U}_n(f)$ with $f \in \{f \in \mathcal{C} : \|f\|_\infty \leq 1\}$. These bounds are presented in the following section.

2.1. Moment bounds for Multiple Mixing Processes

Dehling and Durieu (2011) investigate moment bounds for multiple mixing processes. Here, we state and prove a version their result (Theorem 4 in Dehling and Durieu (2011)). Note, that this version contains a correction of a slight computational error in Dehling and Durieu (2011) that occurs when the polynomials in the multiple mixing property are of a strictly positive degree, and that lead to a wrong exponent in the logarithmic part on the r.h.s. of the corresponding moment bound. Let $(\mathcal{C}, \|\cdot\|_{\mathcal{C}})$ be a normed real vector space of $\mathbb{R}$-valued measurable functions on $\mathcal{X}$. We have the following moment bound for multiple mixing processes.

Proposition 2.1. *Let $(X_i)_{i\in\mathbb{N}^*}$ be multiple mixing for some $\theta \in (0,1)$, $d_0 \in \mathbb{N}$, and $s \geq 1$ with respect to $\mathcal{C}$. If $\mathbf{1}_{\mathcal{X}} \in \mathcal{C}$, then for every $p \in \mathbb{N}^*$, there exists a $C_p > 0$ such that for all $n \in \mathbb{N}^*$*

$$\mathbf{E}\Big(\sum_{i=1}^{n} f(X_i) - \mu f\Big)^{2p} \leq C_p \sum_{i=1}^{p} n^i \|f\|_s^i \log^{2p+ai}(\|f\|_{\mathcal{C}} + 1), \quad \text{for all } f \in \mathcal{C} \text{ with } \|f\|_\infty \leq 1, \tag{2.2}$$

where $a := \max\{-1, d_0 - 2\}$.

Proof. Let $f \in \mathcal{C}$ with $\|f\|_\infty \leq 1$ and set $g := f - \mu f$. For non-negative integers $j_1, j_2, \ldots$ and $q \in \mathbb{N}^*$, we use the abbreviation $j_q^* := \sum_{k=1}^{q} j_k$. Further, we introduce the following notation:

$$I_n(0) := 0, \qquad I_n(p) := \sum_{\substack{0 \leq j_1, \ldots, j_p \leq n-1 \\ j_p^* \leq n-1}} \left|\mathbf{E}\big(g(X_0) g(X_{j_1^*}) \ldots g(X_{j_p^*})\big)\right|, \quad p \in \mathbb{N}^*.$$

By stationarity we have that

$$\mathbf{E}\Big(\sum_{i=1}^{n} g(X_i)\Big)^{p} \leq p! n I_n(p-1) \quad \text{for all } p \in \mathbb{N}^*. \tag{2.3}$$

We will show by complete induction, that for all $p \in \mathbb{N}^*$ there is a constant $c_p > 0$ such that

$$I_n(2p) \leq c_p \sum_{i=1}^{p} n^{i-1} \|g(X_0)\|_r^i \log^{2p+ai+1}(\|g\|_{\mathcal{C}} + \theta^{-1}), \tag{2.4}$$

$$I_n(2p-1) \leq c_p \sum_{i=1}^{p} n^{i-1} \|g(X_0)\|_r^i \log^{2p+ai}(\|g\|_{\mathcal{C}} + \theta^{-1}). \tag{2.5}$$

To do so, we need the following technical lemma.

Lemma 2.1. *Let*

$$J_n(p,q) := \sum_{j_q=0}^{n-1} \sum_{\substack{0 \leq j_1, \ldots, j_{q-1}, j_{q+1}, \ldots, j_p \leq j_q \\ j_p^* \leq n-1}} \left|\mathbf{E}\big(g(X_0) g(X_{j_1^*}) \ldots g(X_{j_p^*})\big)\right|, \quad p \in \mathbb{N}^*, \; q \in \{1, \ldots, p\}.$$

Then for all $p \in \mathbb{N}^*$ and $q \in \{1, \ldots, p\}$ there exists a constant $c_p > 0$ such that

$$J_n(p,q) \leq c_p \, \|g(X_0)\|_r \log^{p+a+1}(\|g\|_{\mathcal{C}} + \theta^{-1}) + n I_n(q-1) I_n(p-q),$$

where $a := \max\{-1, d_0 - 2\}$.

Proof of Lemma 2.1. Let n_0 be a positive integer such that

$$\frac{\log(\|g\|_{\mathcal{C}} + \theta^{-1})}{\log(\theta^{-1})} < n_0 \leq \frac{\log(\|g\|_{\mathcal{C}} + \theta^{-1})}{\log(\theta^{-1})} + 1.$$

Note, that therefore $\theta^{n_0}\|g\|_{\mathcal{C}} \leq 1$ and $n_0 \geq 2$. Now, let

$$\begin{aligned} A_{j_1,\ldots,j_p} &:= \left|\mathbf{Cov}\big(g(X_0)g(X_{j_1^*})\ldots g(X_{j_{q-1}^*}), g(X_{j_q^*})g(X_{j_{q+1}^*})\ldots g(X_{j_p^*})\big)\right|, \\ B_{j_1,\ldots,j_p} &:= \left|\mathbf{E}\big(g(X_0)g(X_{j_1^*})\ldots g(X_{j_{q-1}^*})\big)\right|\left|\mathbf{E}\big(g(X_0)g(X_{j_{q+1}})\ldots g(X_{j_p^*-j_q^*})\big)\right|. \end{aligned}$$

With this notation, we have

$$J_n(p,q) \leq \sum_{j_q=0}^{n-1} \sum_{\substack{0\leq j_1,\ldots,j_{q-1},j_{q+1},\ldots,j_p\leq j_q \\ j_p^*\leq n-1}} A_{j_1,\ldots,j_p} + B_{j_1,\ldots,j_p}$$

Since $\|g\|_\infty \leq 1$ we have $A_{j_1,\ldots,j_p} \leq 2\|g(X_0)\|_1$ and therefore

$$\sum_{\substack{0\leq j_1,\ldots,j_{q-1},j_{q+1},\ldots,j_p\leq j_q \\ j_p^*\leq n-1}} A_{j_1,\ldots,j_p} \leq 2(j_q+1)^{p-1}\|g(X_0)\|_1. \tag{2.6}$$

Furthermore, using the multiple mixing property (2.1), we also have that

$$\sum_{\substack{0\leq j_1,\ldots,j_{q-1},j_{q+1},\ldots,j_p\leq j_q \\ j_p^*\leq n-1}} A_{j_1,\ldots,j_p} \leq \|g(X_0)\|_r\|g\|_{\mathcal{C}}^{\ell}\theta^{j_q} \sum_{0\leq j_1,\ldots,j_{q-1},j_{q+1},\ldots,j_p\leq j_q} Q(j_1,\ldots,j_p).$$

Since $\deg Q \leq d_0$, $Q(j_1,\ldots,j_p)$ can be bounded above by $|Q|(j_q)^{d_0}$, where $|Q|$ is the sum of the positive parts of all coefficients of Q. This implies

$$\begin{aligned} \sum_{\substack{0\leq j_1,\ldots,j_{q-1},j_{q+1},\ldots,j_p\leq j_q \\ j_p^*\leq n-1}} A_{j_1,\ldots,j_p} &\leq |Q|\,\|g(X_0)\|_r\|g\|_{\mathcal{C}}^{\ell}\theta^{j_q}(j_q+1)^{p+d_0-1} \\ &\leq |Q|\,\|g(X_0)\|_r\theta^{j_q-n_0\ell}(j_q+1)^{p+d_0-1} \end{aligned} \tag{2.7}$$

where we used that $|\{0\leq j_1,\ldots,j_{q-1},j_{q+1},\ldots,j_p\leq j_q\}| \leq (j_q+1)^{p-1}$. Using (2.6) for $j_q=0$ to $n_0\ell-2$ and (2.7) for $j_q \geq n_0\ell-1$, we obtain

$$\begin{aligned} &\sum_{j_q=0}^{n-1} \sum_{\substack{0\leq j_1,\ldots,j_{q-1},j_{q+1},\ldots,j_p\leq j_q \\ j_p^*\leq n-1}} A_{j_1,\ldots,j_p} \\ &\leq 2\sum_{j_q=0}^{n_0\ell-2} \|g(X_0)\|_1(j_q+1)^{p-1} + |Q|\,\|g(X_0)\|_r \sum_{j_q=n_0\ell-1}^{n-1} \theta^{j_q-n_0\ell}(j_q+1)^{p+d_0-1} \\ &\leq 2\|g(X_0)\|_1(n_0-1)^p + \theta|Q|\,\|g(X_0)\|_r \sum_{j_q=0}^{n-n_0\ell} \theta^{j_q}(j_q+n_0\ell)^{p+d_0-1} \\ &\leq 2\|g(X_0)\|_r(n_0-1)^p + c_p\theta|Q|\,\|g(X_0)\|_r(n_0-1)^{p+d_0-1}\sum_{j_q=0}^{\infty}\theta^{j_q}(j_q)^{p+d_0-1}, \end{aligned}$$

for some constant $b_p > 0$, depending only on p.

Thus, since $n_0 - 1 \leq \log(\|g\|_{\mathcal{C}} + \theta^{-1})/\log(\theta^{-1})$ and $\sum_{j_q=0}^{\infty} \theta^{j_p}(j_q)^{p+d_0-1} < \infty$, there is a constant $c_p \geq 2$ such that

$$\sum_{j_q=0}^{n-1} \sum_{\substack{0\leq j_1,\dots,j_{q-1},j_{q+1},\dots,j_p\leq j_q \\ j_p^*\leq n-1}} A_{j_1,\dots,j_p} \leq c_p \|g(X_0)\|_r \log^{p+a+1}(\|g\|_{\mathcal{C}} + \theta^{-1}),$$

with $a = \max\{-1, d_0 - 2\}$. On the other hand,

$$\sum_{\substack{0\leq j_1,\dots,j_{q-1},j_{q+1},\dots,j_p\leq j_q \\ j_p^*\leq n-1}} B_{j_1,\dots,j_p} \leq I_n(q-1)I_n(p-q).$$

Therefore $J_n(p,q) \leq c_p \|g(X_0)\|_r \log^{p+a+1}(\|g\|_{\mathcal{C}} + \theta^{-1}) + nI_n(q-1)I_n(p-q)$. □

Obviously we have $I_n(0) = 0$ and $I_n(1) = |\mathbf{E}(g(X_0))|$ and thus (2.4) and (2.5) hold with $p = 0$, $p = 1$. Now, with

$$I_n(p) \leq \|g(X_0)\|_r \log^{p+a+1}(\|g\|_{\mathcal{C}} + \theta^{-1}) + \sum_{q=2}^{p-1} nI_n(q-1)I_n(p-q),$$

and Lemma 2.1 one can carry out the inductive step. With (2.5) established for all $p \in \mathbb{N}^*$, (2.2) follows from (2.3). □

Let us now introduce some classes of multiple mixing processes, that will be investigated in the following parts of this thesis.

2.2. $\mathcal{B}$-Geometrically Ergodic Markov Chains

Let $(X_i)_{i\in\mathbb{N}}$ be a time-homogeneous Markov chain on a measurable state space $(\mathcal{X}, \mathcal{A})$ with 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

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

We assume that there exists a complex Banach space $(\mathcal{B}, \|\cdot\|_{\mathcal{B}})$ 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 operators from $\mathcal{B}$ to $\mathcal{B}$.

We call $(X_i)_{i\in\mathbb{N}^*}$ *$\mathcal{B}$-geometrically ergodic* (cf. Meyn and Tweedie (1993) and Hervé and Pène (2010)), if the corresponding Markov operator on $\mathcal{B}$ satisfies

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

Remark 2.1. This property is also often referred to as strong or geometric ergodicity with respect to $\mathcal{B}$. Given that $\mathbf{1}_{\mathcal{X}} \in \mathcal{B}$, this property corresponds to a *spectral gap* property of P

acting on $\mathcal{B}$, that is 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 $\mathcal{L}(\mathcal{B})$, given by

$$P = \Pi + N \tag{2.8}$$

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 w.r.t. $\mathcal{B}$, given by

$$\|P\|_{\mathcal{L}(\mathcal{B})} := \sup_{f\in\mathcal{B}\setminus\{0\}} \frac{\|P(f)\|_{\mathcal{B}}}{\|f\|_{\mathcal{B}}}.$$

To establish the multiple mixing property, analogously to Lemma 3 in Dehling and Durieu (2011), we make the following three assumptions on the space $\mathcal{B}$:

(2.B) $\mathbf{1}_{\mathcal{X}} \in \mathcal{B}$, $|f|$ and $\overline{f} \in \mathcal{B}$ for all $f \in \mathcal{B}$, and for every $x \in \mathcal{X}$, the mapping $f \mapsto f(x)$ is continuous on $\mathcal{B}$.

Moreover for some $m \in [1, \infty]$,

(2.C) $\mathcal{B}$ is continuously included in $\mathrm{L}^m(\nu)$, i.e. $\mathcal{B} \subset \mathrm{L}^m(\mu)$ and there is a $K > 0$ such that $\|f\|_m \le K\|f\|_{\mathcal{B}}$ for all $f \in \mathcal{B}$.

Lastly, we use the condition that

(2.D) there exist some $C > 0$ 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 $\|fg\|_{\mathcal{B}} \le C \max\{\|f\|_{\mathcal{B}}, \|g\|_{\mathcal{B}}\}^{\ell}$.

Remark 2.2. Note that if $\mathcal{B}$ is a Banach algebra,[1] condition (2.D) holds with $\ell = 2$.

Lemma 2.2. *Under the conditions (2.A), (2.B), (2.C), and (2.D), the Markov chain $(X_i)_{i\in\mathbb{N}}$ 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 $\|f\|_\infty \le 1$ and set $s = m/(m-1)$. For all $p > q > 0$, for all $i_0 < i_1 < \ldots < i_p$, we write $g = fP^{i_{q+1}-i_q}\big(\ldots fP^{i_{p-1}-i_{p-2}}\big(fP^{i_p-i_{p-1}}(f)\big)\ldots\big)$. By (2.D), g belongs to $\mathcal{B}$. Let σ_j denote the σ-algebra generated by $X_0, \ldots, X_j$. Using Hölder's inequality, we obtain

$$\begin{aligned}
&\big|\mathbf{Cov}(f(X_{i_0})\cdots f(X_{i_{q-1}}), f(X_{i_q})\cdots f(X_{i_p}))\big| \\
&\quad= \mathbf{E}\Big\{f(X_{i_0})\cdots f(X_{i_{q-1}})\cdot\Big(\mathbf{E}\Big[f(X_{i_q})\ldots\,\mathbf{E}\big[f(X_{i_{p-1}})\,\mathbf{E}[f(X_{i_p})|\sigma_{i_{p-1}}]\;\big|\;\sigma_{i_{p-2}}\big]\ldots\;\Big|\;\sigma_{i_{q-1}}\Big] \\
&\qquad - \mathbf{E}\big(f(X_{i_q})\cdots f(X_{i_p})\big)\Big)\Big\} \\
&\quad\le \|f(X_{i_0})\cdots f(X_{i_{q-1}})\|_s\,\|P^{i_q-i_{q-1}}(g) - (\nu g)\,\mathbf{1}_{\mathcal{X}}\,\|_m.
\end{aligned} \tag{2.9}$$

[1] I.e. $(\mathcal{B}, +, \|\cdot\|_{\mathcal{B}})$ is a complete normed vector space, $(\mathcal{B}, +, \cdot)$ is an associative algebra over $\mathbb{C}$ or $\mathbb{R}$ (which is always the case for spaces of $\mathbb{C}$- or $\mathbb{R}$-valued functions), and the inner multiplication "$\cdot$" in $\mathcal{B}$ satisfies $\|f\cdot g\|_{\mathcal{B}} \le \|f\|_{\mathcal{B}}\cdot\|g\|_{\mathcal{B}}$.

Using (2.A), (2.C), and $\|f\|_\infty \leq 1$, we infer

$$\left|\mathbf{Cov}(f(X_{i_0})\cdots f(X_{i_{q-1}}), f(X_{i_q})\cdots f(X_{i_p}))\right| \leq K\|f\|_s\|g\|_{\mathcal{B}}\theta^{i_q-i_{q-1}}.$$

By (2.A), P has a decomposition given in (2.8) and thus

$$\|P^k f\|_{\mathcal{B}} \leq \|N^k f\|_{\mathcal{B}} + \|\Pi\|_{\mathcal{L}(B)}\|f\|_{\mathcal{B}} \leq c\|f\|_{\mathcal{B}} \quad \text{for all } k \in \mathbb{N}^*$$

for some $c \geq 1$ since N has a spectral radius that is strictly smaller 1. Now, using (2.D), 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. □

The following section introduces bounded Hölder continuous and weighted Lipschitz functions. We present some basic properties of this spaces that will be used later. Especially, we show that the weighted Lipschitz functions satisfy condition (2.B), (2.C), and (2.D).

2.3. Hölder Spaces and Weighted Lipschitz Functions

Let $\alpha \in (0,1]$ and $\mathbb{K} = \mathbb{C}$ or $\mathbb{K} = \mathbb{R}$. For a metric space $(\mathcal{X}, d)$, we denote the space of *bounded α-Hölder continuous functions* on $\mathcal{X}$ with values in $\mathbb{K}$ by $\mathcal{H}_\alpha(\mathcal{X}, \mathbb{K})$. This space is equipped with the norm

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

where

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

Note that this space is a Banach algebra. Further we introduce the space $\mathcal{H}_{\alpha,\beta}(\mathcal{X}, \mathbb{K})$ of Lipschitz functions with weights. Let $\alpha \in (0,1]$, $\beta \in [0,1]$ and $x_0 \in \mathcal{X}$. $\mathcal{H}_{\alpha,\beta}(\mathcal{X}, \mathbb{K})$ is defined as the space of continuous function from $\mathcal{X}$ to $\mathbb{K}$ with the norm $\|\cdot\|_{\mathcal{H}_{\alpha,\beta}} = N_\beta(\cdot) + m_{\alpha,\beta}(\cdot) < \infty$, where

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

Observe that $\mathcal{H}_{\alpha,0}(\mathcal{X}, \mathbb{K}) = \mathcal{H}_\alpha(\mathcal{X}, \mathbb{K})$, $\|\cdot\|_{\mathcal{H}_{\alpha,0}} = \frac{1}{2}\|\cdot\|_{\mathcal{H}_\alpha}$, and that $\mathcal{H}_\alpha(\mathcal{X}, \mathbb{K})$ is a subspace of $\mathcal{H}_{\alpha,\beta}(\mathcal{X}, \mathbb{K})$ for every $\beta > 0$. Further, $\mathcal{H}_\alpha(\mathcal{X}, \mathbb{C})$ satisfies the following basic properties.

Lemma 2.3. *If ν has a finite first moment (that is $\int_{\mathcal{X}} d(x,x_0)\, d\nu(x) < \infty$ for some $x_0 \in \mathcal{X}$) then for all α and $\beta \in [0,1]$,*

(a) 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 $\|fg\|_{\mathcal{H}_{\alpha,\beta}} \leq \|f\|_{\mathcal{H}_\alpha}\|g\|_{\mathcal{H}_{\alpha,\beta}}$,

(b) $(\mathcal{H}_{\alpha,\beta}(\mathcal{X}, \mathbb{C}), \|\cdot\|_{\mathcal{H}_{\alpha,\beta}})$ is a Banach space which satisfies condition (2.B),

(c) there exists a $C > 0$ such that $\|f\|_{1/\beta} \leq CN_\beta(f)$ for every $f \in \mathcal{H}_{\alpha,\beta}(\mathcal{X}, \mathbb{C})$ (where we agree to let $1/\beta = \infty$ for $\beta = 0$),

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

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

The proof only requires straight forward calculations and is therefore postponed to the appendix (cf. Section A.5). As a direct consequence one obtains

Proposition 2.2. *If ν has a finite first moment, then for all $\alpha \in (0,1]$ and $\beta \in [0,1]$ the space $\mathcal{B} = \mathcal{H}_{\alpha,\beta}(\mathcal{X}, \mathbb{C})$ satisfies (2.B), (2.C), and (2.D) with $m = \beta^{-1}$ if $\beta \neq 0$ and $m = \infty$ else.*

In the following section, we present iterative Lipschitz models as an example for Markov chains satisfying condition (2.A) w.r.t. the space of weighted Lipschitz functions.

2.4. Iterative Lipschitz Models that Contract on Average

Consider the following experiment. A mouse is confronted with a sequence of possible events $E_0, \ldots, E_m$. After each event, the behaviour of the mouse alters, the mouse "learns". We are interested in this learning process. Assume the "conscience" or the "state of learning" of the mouse is represented by an element x of some appropriate space $\mathcal{X}$ which may change during the experiment. The learning process the mouse undergoes after experiencing event E_i shall be represented by a transformation T_i on $\mathcal{X}$. Assume further, that the probability of the occurrence of the event E_i is given by p_i, where $p_0, \ldots, p_m$ is a probability vector depending only on the current learning state of the mouse. Such mathematical learning models have first been studied by Bush and Mosteller (1951), Bush and Mosteller (1953) and Karlin (1953).

In this section we consider so called iterative Lipschitz models that contract on average. Assume that $(\mathcal{X}, d)$ is a (not necessarily compact) metric space in which every closed ball is compact and suppose that $\mathcal{X}$ is equipped with the Borel σ-algebra $\mathfrak{B}(\mathcal{X})$. Let $\{T_i, i \in \mathbb{N}\}$ and $\{p_i,\ i \in \mathbb{N}\}$ be families of Lipschitz functions that map from $\mathcal{X}$ to $\mathcal{X}$ and $\mathcal{X}$ to $[0,1]$, respectively, where the p_i satisfy

$$\sum_{i=1}^{\infty} p_i(x) = 1 \quad \text{for all } x \in \mathcal{X}.$$

An *iterative Lipschitz model* is then given by a Markov chain $(X_i)_{i\in\mathbb{N}}$ with state space $\mathcal{X}$ and a transition probability P of the form

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

Thus—as in the learning model—each step of the Markov chain corresponds to the application of one of the transformations T_i which is chosen randomly with respect to the discrete probability distribution given by $p_0(x), p_1(x), \ldots$ depending only on the actual state x of the chain. We further assume that this model has a property of *contraction on average*, that is that there exists a $\rho \in (0,1)$ such that

$$\sum_{i=0}^{\infty} d(T_i(x), T_i(y))p_i(x) < \rho d(x,y) \quad \text{for all } x, y \in \mathcal{X}. \tag{2.10}$$

Such Markov chains have been studied e.g. by Barnsley, Demko, Elton, and Geronimo (1988), Peigné (1993), Walkden (2007), and by Durieu (2013). As in these papers, we work with the following technical properties in the setting of Markov chains.

For some fixed $x_0 \in \mathcal{X}$, suppose

$$\sup_{\substack{x,y,z\in\mathcal{X},\\ y\neq z}} \sum_{i=0}^{\infty} \frac{d(T_i(y), T_i(z))}{d(y,z)} p_i(x) < \infty, \tag{2.11}$$

$$\sup_{x,y\in\mathcal{X}} \sum_{i=0}^{\infty} \frac{d(T_i(y), x_0)}{1 + d(y, x_0)} p_i(x) < \infty, \tag{2.12}$$

$$\sup_{x\in\mathcal{X}} \sum_{i=0}^{\infty} \frac{d(T_i(x), x_0)}{1 + d(x, x_0)} \sup_{y,z\in\mathcal{X}, y\neq z} \frac{|p_i(y) - p_i(z)|}{d(y,z)} < \infty. \tag{2.13}$$

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

$$d\big(T_{i_n} \circ \ldots \circ T_{i_1}(x)\,,\, T_{j_n} \circ \ldots \circ T_{j_1}(y)\big)\big(1 + d\big(T_{j_n} \circ \ldots \circ T_{j_1}(x)\,,\, x_0\big)\big) \longrightarrow 0 \quad \text{as } n \to \infty \tag{2.14}$$

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 (2.11) – (2.13) are verified when the family of maps T_i is finite and (2.14) is verified when (2.10) – (2.13) hold and each p_i is positive. See Peigné (1993) for a discussion on these assumptions.

Under the conditions (2.10) – (2.14), Peigné (1993) 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 starting distribution ν (i.e. $X_0 \sim \nu$) and transition probability P.

Now, according to Theorem 1 in Peigné (1993), we immediately obtain the following proposition.

Proposition 2.3. *If $(X_i)_{i\in\mathbb{N}^*}$ is an iterative Lipschitz model with values in $\mathcal{X}$ and satisfies (2.10) – (2.14), then P is a bounded linear operator on $\mathcal{B} = \mathcal{H}_{\alpha,\beta}(\mathcal{X}, \mathbb{C})$ that satisfies (2.A) for any $\alpha, \beta \in (0, 1/2)$ such that $\alpha < \beta$.*

An application of Proposition 2.2 and Proposition 2.3 to Lemma 2.2 yields the following corollary.

Corollary 2.1. *Assume that $(X_i)_{i\in\mathbb{N}^*}$ is an iterative Lipschitz model with values in $\mathcal{X}$ and satisfies (2.10) – (2.14) for some $x \in \mathcal{X}$. Then for all $\alpha, \beta \in (0, 1/2)$ with $\alpha < \beta$ the process $(X_i)_{i\in\mathbb{N}^*}$ is multiple mixing w.r.t. $\mathcal{H}_{\alpha,\beta}(\mathcal{X}, \mathbb{C})$ with $d_0 = 0$ and $s = (1-\beta)^{-1}$.*

2.5. Dynamical Systems with a Spectral Gap

Another example of multiple mixing processes are dynamical systems that satisfy the same assumptions as the Markov chains introduced in Section 2.2, but where P is the Perron–Frobenius operator instead of the Markov operator.

Let $(\mathcal{X}, \mathcal{A}, \nu)$ be a probability space and let T be a measure preserving transformation of $\mathcal{X}$, that is $\nu(T^{-1}(A)) = \nu(A)$ for all $A \in \mathcal{A}$. The Perron–Frobenius operator P is defined on $L^1(\nu)$ by the equation

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

The assumption on the Markov chain in section Section 2.2 can also be applied on the Perron–Frobenius operator P. It is easy to see, that an analogue result to Lemma 2.2 holds holds for dynamical systems

Lemma 2.4. *Let $\mathcal{B}$ be a complex Banach space of $\mathbb{C}$-valued functions on $\mathcal{X}$. If a measure preserving dynamical system $(\mathcal{X}, \mathcal{A}, \nu, T)$ satisfies (2.A), (2.B), (2.C), and (2.D), then the process $(T^i)_{i\in\mathbb{N}}$ is multiple mixing w.r.t. $\mathcal{B}$ with $s = m/(m-1)$ and $d_0 = 0$.*

Proof. Follow the proof of Lemma 2.2. In the computation of (2.9) instead of conditional expectations use the following consideration. Let $f, g \in \mathcal{B}$ with $\|f\|_\infty, \|g\|_\infty < \infty$, then

$$\begin{aligned}
\mathbf{Cov}\Big(f \cdot f(T^i)\,,\, g(T^{i+n})\Big) &= \nu\Big(f \cdot f(T^i) \cdot g(T^{i+n})\big) - \nu\big(f \cdot f(T^i)\big)\,\nu\big(g(T^{i+n})\Big)\\
&= \nu\Big(P^n\big((P^i f) \cdot f\big) \cdot g\Big) - \nu\big(P^i f \cdot f\big)\,\nu g\\
&= \nu\Big(\big(P^n\big((P^i f) \cdot f\big) - \nu(P^i f \cdot f)\big) \cdot g\Big)\\
&\le \big\|P^n\big((P^i f) \cdot f\big) - \big(\nu\big((P^i f) \cdot f\big)\big)\,\mathbf{1}_{\mathcal{X}}\big\|_m \|g\|_s.
\end{aligned}$$

□

In the preceding sections, the multiple mixing property was always established by using a spextral gap property of the operator. However, there are multiple mixing processes that do not necessarily have a spectral gap property. One example for such processes is the ergodic automorphism of the torus, which we introduce in the following section.

2.6. Ergodic Automorphisms of the Torus

For $d \ge 2$, we define the d-dimensional Torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$ as the quotient space of $\mathbb{R}^d$ equipped with the usual euclidean metric and the equivalence relation $x \sim y$ if and only if $x - y \in \mathbb{Z}^d$. In the following, we will always identify $\mathbb{T}^d$ with the d-dimensional intervall $[0,1]^d$. If A is a square matrix of dimension d with integer coefficients and determinant ± 1, then the transformation $T : \mathbb{T}^d \longrightarrow \mathbb{T}^d$ defined by

$$Tx = Ax \mod 1$$

is an automorphism of $\mathbb{T}^d$ that preserves the Lebesgue measure λ. Thus $(\mathbb{T}^d, \mathfrak{B}(\mathbb{T}^d), \lambda, T)$ is a measure preserving dynamical system.[2] It is ergodic if and only if the matrix A has no eigenvalue which is a root of unity.[3] A result of Kronecker shows that in this case, A always has at least one eigenvalue which has modulus different than 1.

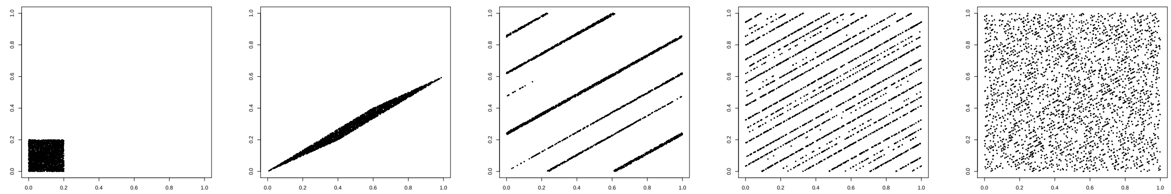


Figure 2.1.: The graphics illustrate how the ergodic torus automorphism given by $T(x, y) = (3x + 2y, 2x + y) \mod 1$ acts on a sample of $3,000$ uniform distributed points in $[0, 0.2]^2$ in four steps.

The hyperbolic automorphisms of the torus (i.e. A has no eigenvalue of modulus 1) are particular cases of Anosov diffeomorphisms. Their properties are better understood than in the general case. However, the general case of ergodic automorphisms is an example of a partially hyperbolic system. A central limit theorem for regular observables of such automorphisms of the torus has been established by Leonov (1960), see also Le Borgne (1999) for refinements.

For an ergodic automorphism of the torus given by A, one has a decomposition of $\mathbb{R}^d$ as the direct sum $\mathbb{R}^d = E^s \oplus E^u \oplus E^c$ of the stable subspace E^s, the unstable subspace E^u, and the central (or neutral) subspace E^c of $\mathbb{R}^d$. These spaces can be characterized by the property that there is some fixed $C_s, C_u, C_c > 0$ and $\kappa > 1$ such that $|A^n v|_{\max} \leq C_s \kappa^{-n} |v|_{\max}$ for all $v \in E^s$, $|A^n v|_{\max} \geq C_u \kappa^n |v|_{\max}$ for all $v \in E^u$, and $|A^n v|_{\max} \leq C_c n^J |v|_{\max}$ for all $v \in E^c$, where $|\cdot|_{\max}$ is the maximum norm on $\mathbb{R}^d$ and J is the size of the biggest Jordan block of T restricted to E^c. Dehling and Durieu (2011) proved a CLT for the $\{\mathbf{1}_{[0,x]} : x \in [0,1]^d\}$-indexed empirical process of the ergodic automorphism of the torus. They also showed that for any $\alpha \in (0, 1]$ the process $(X_i)_{i \in \mathbb{N}^*}$ given by

$$X_0 \sim \lambda, \qquad\qquad X_i = T(X_{i-1}), \quad i \in \mathbb{N}^*.$$

is multiple mixing w.r.t. the space $\mathcal{H}_\alpha(\mathbb{T}^d, \mathbb{R})$ of $\mathbb{R}$-valued bounded α-Hölder continuous functions on the torus with $s = 1$ and $d_0 = J$ (see Proposition A.2). Note, that therefore one also has the multiple mixing property for the process $(\varphi(X_i))_{i \in \mathbb{N}^*}$ for every Hölder continuous function $\varphi : \mathbb{T}^d \longrightarrow \mathbb{R}^\ell$, $\ell \in \mathbb{N}^*$.

[2] $\mathfrak{B}(\mathbb{T}^d)$ denotes the Borel σ-algebra on $\mathbb{T}^d$.
[3] I.e. if v is an eigenvalue of A, then there is no $n \in \mathbb{N}^*$ such that $v^n = 1$.

Part I.

Empirical Central Limit Theorems for Multiple Mixing Processes

Based on the article
Herold Dehling, Olivier Durieu, and Marco Tusche (2012): Approximating class approach for empirical processes of dependent sequences indexed by functions. Preprint. To appear in Bernoulli. arXiv: 1201.2256

Keywords: Empirical processes indexed by classes of functions, Dependent data, Markov chains, Dynamical systems, Ergodic torus automorphism, Weak convergence

Mathematics Subject Classification (2010): 60F05, 60F17, 60G10, 62G30, 28D05

3. An ECLT for Multiple Mixing Processes

Let $(X_i)_{i\in\mathbb{N}}$ be a stationary stochastic process with state space $(\mathcal{X},\mathcal{A})$ and marginal distribution μ and let $\mathcal{F}$ be a uniformly bounded class of real-valued measurable functions defined on $\mathcal{X}$. For a measure Λ on $(\mathcal{X},\mathcal{A})$, we use the notation $\Lambda f = \int_{\mathcal{X}} f \, d\Lambda$. We denote the space of bounded real-valued functions on $\mathcal{F}$ by $\ell^\infty(\mathcal{F})$. This space is equipped with the supremum norm and the Borel σ-algebra generated by the open sets. The $\mathcal{F}$-indexed empirical process $\mathrm{U}_n = (\mathrm{U}_n(f))_{f\in\mathcal{F}}$ is the $\ell^\infty(\mathcal{F})$-valued random element given by

$$\mathrm{U}_n(f) := \sqrt{n}\big(\mu_n(f) - \mu f\big), \quad f \in \mathcal{F},$$

where $\mu_n(f) := n^{-1}\sum_{i=1}^n f(X_i)$.[1] Recall that we can not assume that U_n is measurable here and therefore have to use the theory of outer expectation and integrals (cf. Section 1.3).

The goal of the following section is to establish an empirical process CLT for processes $(X_i)_{i\in\mathbb{N}}$ that are multiple mixing w.r.t. some function space $\mathcal{C}$ (see Chapter 2). The content of this part corresponds mainly to the article Dehling et al. (2012) and is supplemented by some later results.

In what follows, we will frequently make two assumptions concerning the process $(f(X_i))_{i\in\mathbb{N}^*}$, where $f : \mathcal{X} \to \mathbb{R}$ belongs to some normed vector space $(\mathcal{C}, \|\cdot\|_{\mathcal{C}})$ of measurable real-valued functions on $\mathcal{X}$. The precise choice of $\mathcal{C}$ will depend on the specific example. Often, we take $\mathcal{C}$ to be the space of all bounded Lipschitz or α-Hölder continuous functions.

Assumption 3.I (CLT for $\mathcal{C}$-Observables)**.** For all $f \in \mathcal{C}$, there exists a $\sigma_f^2 \geq 0$ such that

$$U_n(f) \xrightarrow{d} N(0, \sigma_f^2), \tag{3.1}$$

where $N(0,\sigma_f^2)$ denotes the normal law with mean zero and variance σ_f^2.

Assumption 3.II (Moment Bounds for $\mathcal{C}$-Observables)**.** For some $s \geq 1$ and $a \in \mathbb{R}$ for all $p \geq 1$, there exists a constant $C_p > 0$ such that for all $f \in \mathcal{C}$ with $\|f\|_\infty \leq 1$

$$\mathbf{E}\bigg(\sum_{i=1}^n \big(f(X_i) - \mu f\big)\bigg)^{2p} \leq C_p \sum_{i=1}^p n^i \|f\|_s^i \log^{2p+ai}(\|f\|_{\mathcal{C}} + 1). \tag{3.2}$$

Remark 3.1. Note, that if Assumption 3.II holds, then for fixed $M > 0$ inequality (3.2) also holds uniformly for all $f \in \mathcal{C}_M := \{f \in \mathcal{C} : \|f\|_\infty \leq M\}$. In this case, the constants C_p may differ from those for case where $\|f\|_\infty \leq 1$ is assumed. To see this, observe that every $f \in \mathcal{C}_M$

[1] U_n takes values in $\ell^\infty(\mathcal{F})$ since $\mathcal{F}$ is supposed to be uniformly bounded.

can be represented as $f = Mf_1$ for some $f_1 \in \mathcal{C}$ with $\|f_1\|_\infty \leq 1$. Then inequality (2.2) for general $f \in \mathcal{C}_M$ can be directly deduced from the special case, where $\|f\|_\infty \leq 1$.

Both Assumption 3.I and Assumption 3.II have been established by many authors for a wide range of stationary processes. Concerning the CLT, see e.g. the three-volume monograph by Bradley (2007) for mixing processes, Dedecker, Doukhan, Lang, León, Louhichi, and Prieur (2007) for so-called weakly dependent processes in the sense of Doukhan and Louhichi (1999), and Hennion and Hervé (2001) for many examples of Markov chains and dynamical systems.

Durieu (2008b) proved 4th moment bounds of the type (3.2) for Markov chains or dynamical systems for which the Markov operator or the Perron-Frobenius operator acting on $\mathcal{C}$ has a spectral gap. It was generalized to $2p$-th moment bounds for multiple mixing processes by Dehling and Durieu (2011), see Proposition 2.1 in this thesis. Note that the multiple mixing condition implies the moment bound (3.2) with $a = \max\{-1, d_0 - 2\}$ (see Proposition 2.1). This applies e.g. in the case of ergodic automorphisms of the torus (see Section 2.6 and Proposition A.2). Further, the spectral gap property leads to the multiple mixing condition with $d_0 = 0$, and thus to the moment bound (3.2) with $a = -1$, see Lemma 2.2 and Lemma 2.4.

Reminding ourselves of the discussion about the approximating class approach in Section 1.3, Assumption 3.I corresponds to property (1.A), while property (1.B) will be derived from Assumption 3.II and a specific entropy condition on our bracketing numbers, which are defined as follows.

Definition 3.1. Let μ be a probability measure on a measurable space $(\mathcal{X}, \mathcal{A})$, let $\mathcal{G}$ be a subclass of a normed vector space $(\mathcal{C}, \|\cdot\|_\mathcal{C})$ of real-valued measurable function on $\mathcal{X}$, $s \geq 1$, and $\varepsilon, A > 0$. For two measurable functions $l, u : \mathcal{X} \longrightarrow \mathbb{R}$ with $l(x) \leq u(x)$ for all $x \in \mathcal{X}$ we call the bracket

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

an $(\varepsilon, A, \mathcal{G}, \mathrm{L}^s(\mu))$-bracket if $l, u, \in \mathcal{G}$ and if

$$\|u - l\|_s \leq \varepsilon,$$
$$\max\{\|u\|_\mathcal{B}, \|l\|_\mathcal{B}\} \leq A.$$

For a class of measurable functions $\mathcal{F}$ defined 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 $\mathcal{F}$.

3.1. Statement of Results

We can now state the first abstract main result of our work.

Theorem 3.1 (Empirical CLT)**.** *Let $(\mathcal{X}, \mathcal{A})$ be a measurable space, 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 uniformly bounded class of measurable functions on $\mathcal{X}$. Suppose that for some normed vector space $\mathcal{C}$ of measurable real-valued functions on $\mathcal{X}$, $a \in \mathbb{R}$, and $s \geq 1$, Assumption 3.I and Assumption 3.II hold. Moreover, assume that there exist some uniformly bounded subclass $\mathcal{G} \subset \mathcal{C}$, some constants $r > -1$, $\gamma > \max\{2 + a, 1\}$, and $C > 0$ such that*

$$\int_0^1 \varepsilon^r \sup_{\varepsilon \leq \delta \leq 1} N^2\big(\delta, \exp(C\delta^{-1/\gamma}), \mathcal{F}, \mathcal{G}, \mathrm{L}^s(\mu)\big) d\varepsilon < \infty. \tag{3.3}$$

Then the empirical process U_n converges in distribution in $\ell^\infty(\mathcal{F})$ to a tight centred Gaussian process W.

The proof is given in Section 3.2.

Remark 3.2. (i) Note that the bracketing number $N(\delta, \exp(C\delta^{-1/\gamma}), \mathcal{F}, \mathcal{G}, \mathrm{L}^s(\mu))$ might not be a monotone function of δ. This is the reason why we take the supremum in the integral (3.3).

(ii) The proof of Theorem 3.1 shows that the statement also holds if condition (3.2) only holds for some integer p satisfying

$$p > \frac{(r+1)\gamma}{\gamma - \max\{2 + a, 1\}}.$$

(iii) If for some $r' \geq 0$,

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

condition (3.3) is satisfied for all $r > 2r' - 1$.

(iv) Examples of classes of functions satisfying condition (3.3) are provided in Chapter 4. Among the examples are indicators of multidimensional rectangles, of ellipsoids, and of balls of arbitrary metrics, as well as a class of monotone functions.

(v) Among the possible applications of Theorem 3.1 are dynamical systems with a spectral gap on the Perron–Frobenius operator and $\mathcal{B}$-geometrically ergodic Markov chains such as iterative Lipschitz models that contract on average (see Section 2.2, 2.4, and Section 2.5). Durieu (2013) applied Theorem 3.1 for iterative Lipschitz models that contract on average satisfying the conditions (2.10) – (2.14). In Chapter 5, we provide applications of Theorem 3.1 to ergodic torus automorphisms, indexed by various classes of indicator functions.

In the general setting of Theorem 3.1, we cannot precise the covariance structure of the limit process W. The following lemma identifies the covariance structure of W under additional conditions.

Lemma 3.1. *In the situation of Theorem 3.1, assume that*

(i) Assumption 3.I holds with variance σ_f^2 given by

$$\sigma_f^2 = \mathbf{Var}\big(f(X_0)\big) + 2\sum_{k=1}^{\infty} \mathbf{Cov}\big(f(X_0), f(X_k)\big), \tag{3.4}$$

(ii) there are a $\rho \in (0,1)$ and a $C' > 0$ such that

$$\big|\mathbf{Cov}\big(\varphi(X_0), f(X_k)\big)\big| \le C' \|\varphi\|_\infty \|f\|_{\mathcal{C}}\, \rho^k$$

for all $f \in \mathcal{C}$ and all $\varphi \in \mathcal{F} \cup (\mathcal{F} - \mathcal{G})$

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

$$\mathbf{Cov}\big(W(f), W(g)\big) = \sum_{k=0}^{\infty} \mathbf{Cov}\big(f(X_0), g(X_k)\big) + \sum_{k=1}^{\infty} \mathbf{Cov}\big(f(X_k), g(X_0)\big) \quad f, g \in \mathcal{F}. \tag{3.5}$$

The proof of Lemma 3.1 is given in Section 3.3.

As shown in Chapter 2, multiple mixing processes with total degree of the the polynomial term not larger than $d_0 \in \mathbb{N}$ satisfy Assumption 3.II with $a = \max\{-1, d_0 - 2\}$. Thus, for multiple mixing processes, we have the following version of Theorem 3.1.

Theorem 3.2 (Empirical CLT for Multiple Mixing Processes)**.** *Let $(X_i)_{i\in\mathbb{N}}$ be a stationary stochastic process on a state space $(\mathcal{X}, \mathcal{A})$ with marginal distribution μ, and let $\mathcal{F}$ be a uniformly bounded class of measurable functions on $\mathcal{X}$. Assume that the process $(X_i)_{i\in\mathbb{N}}$ is multiple mixing w.r.t. to some normed vector space $\mathcal{C}$ of measurable functions on $\mathcal{X}$ with $s \ge 1$, total degree of the multivariate polynomial Q in (2.1) not larger than d_0 and satisfies Assumption 3.I. If there exist some uniformly bounded subclass $\mathcal{G} \subset \mathcal{C}$, some constants $r > -1$, $\gamma > \{1, d_0\}$, and $C > 0$ such that (3.3) holds, then the empirical process U_n converges in distribution in $\ell^\infty(\mathcal{F})$ to a tight centred Gaussian process W.*

If further the variance σ_f^2 in Assumption 3.I is given by (3.4) and if there are constants $\rho \in (0,1)$ and $C' > 0$ such that for all $f \in \mathcal{C}$ and all $\varphi \in \mathcal{F} \cup (\mathcal{F} - \mathcal{G})$

$$\big|\mathbf{Cov}\big(\varphi(X_0), f(X_k)\big)\big| \le C' \|\varphi\|_\infty \|f\|_{\mathcal{C}}\, \rho^k,$$

then the covariance structure of the limit process W is given by (3.5).

3.2. Proof of Theorem 3.1

As discussed in Section 1.3 we apply Theorem 1.1. First, let us construct a process $U_n^{(q)}$ corresponding to $\xi_n^{(q)}$ in Theorem 1.1 an the complete metric space $S = \ell^\infty(\mathcal{F})$.

For all $q \ge 1$, let

$$N_q := N(2^{-q}, \exp(C2^{\frac{q}{\gamma}}), \mathcal{F}, \mathcal{G}, \mathrm{L}^s(\mu))$$

There exist two sets of functions $\{g_{q,1}, \ldots, g_{q,N_q}\} \subset \mathcal{G}$ and $\{g'_{q,1}, \ldots, g'_{q,N_q}\} \subset \mathcal{G}$, such that $\|g_{q,i} - g'_{q,i}\|_s \le 2^{-q}$, $\|g_{q,i}\|_{\mathcal{C}} \le \exp(C2^{\frac{q}{\gamma}})$, $\|g'_{q,i}\|_{\mathcal{C}} \le \exp(C2^{\frac{q}{\gamma}})$ and for all $f \in \mathcal{F}$, there exists an i such that $g_{q,i} \le f \le g'_{q,i}$. Further, by (3.3),

$$\sum_{q\ge1} 2^{-(r+1)q} N_q^2 < \infty. \tag{3.6}$$

For all $q \geq 1$, we can build a partition $\mathcal{F} = \bigcup_{i=1}^{N_q} \mathcal{F}_{q,i}$ of the class $\mathcal{F}$ into N_q subsets such that for all $f \in \mathcal{F}_{q,i}$, $g_{q,i} \leq f \leq g'_{q,i}$. To see this define $\mathcal{F}_{q,1} = [g_{q,1}, g'_{q,1}]$ and $\mathcal{F}_{q,i} = [g_{q,i}, g'_{q,i}] \backslash (\cup_{j=1}^{i-1} \mathcal{F}_j)$.

In the sequel, we will use the notation $\pi_q f = g_{q,i}$ and $\pi'_q f = g'_{q,i}$ if $f \in \mathcal{F}_{q,i}$. For each $q \geq 1$, we introduce the process

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

which is constant on each $\mathcal{F}_{q,i}$. Further, if $f \in \mathcal{F}_{q,i}$, we have

$$\mu_n^{(q)}(f) \leq \mu_n(f) \leq \mu_n(\pi'_q f)$$

We introduce

$$U_n^{(q)}(f) := \mathrm{U}_n(\pi_q f) = \sqrt{n}(\mu_n^{(q)}(f) - \mu(\pi_q f)), \quad f \in \mathcal{F}.$$

Now we establish assumption (1.10) and (1.11) of Theorem 1.1 in two separate propositions.

Proposition 3.1. *For all $q \geq 1$, the sequence $(U_n^{(q)}(f))_{f \in \mathcal{F}}$ converges in distribution in $\ell^\infty(\mathcal{F})$ to a piecewise constant Gaussian process $(U^{(q)}(f))_{f \in \mathcal{F}}$ as $n \to \infty$.*

Proof. Since $\pi_q f \in \mathcal{G}$ and $\mathcal{G}$ is a subset of $\mathcal{C}$, by assumption (3.1), the CLT holds and $U_n^{(q)}(f)$ converges to a Gaussian law for all $f \in \mathcal{F}$. We can apply the Cramér-Wold device to get the finite-dimensional convergence: for all $k \geq 1$, for all $f_1, \ldots, f_k \in \mathcal{F}$, $(U_n^{(q)}(f_1), \ldots, U_n^{(q)}(f_k))$ converges in distribution to a Gaussian vector $(U^{(q)}(f_1), \ldots, U^{(q)}(f_k))$ in $\mathbb{R}^k$. Since $U_n^{(q)}$ is constant on each element $\mathcal{F}_{q,i}$ of the partition, the finite-dimensional convergence implies the weak convergence of the process. Indeed, consider the function $\tau_q : \mathbb{R}^{N_q} \to \ell^\infty(\mathcal{F})$ that maps a vector $x = (x_1, \ldots, x_{N_q})$ to the function $\mathcal{F} \to \mathbb{R}$, $f \mapsto x_i$ such that $f \in \mathcal{F}_{q,i}$. For $f_1 \in \mathcal{F}_{q,1}, \ldots, f_{N_q} \in \mathcal{F}_{q,N_q}$ we have $U_n^{(q)} = \tau_q(U_n^{(q)}(f_1), \ldots, U_n^{(q)}(f_{N_q}))$ and thus the continuous mapping theorem guarantees that $U_n^{(q)}$ converges weakly to the random variable $U^{(q)} = \tau_q(U^{(q)}(f_1), \ldots, U^{(q)}(f_{N_q}))$ which is constant on each $\mathcal{F}_{q,i}$. □

Proposition 3.2. *For all $\varepsilon > 0$, $\eta > 0$ there exists a q_0 such that for all $q \geq q_0$*

$$\limsup_{n \to \infty} \mathbf{P}^*(\sup_{f \in \mathcal{F}} |\,\mathrm{U}_n(f) - U_n^{(q)}(f)| > \varepsilon) \leq \eta.$$

Proof. For a random variable Y let $\overline{Y}$ denote its centring $\overline{Y} := Y - \mathbf{E}\,Y$. If for arbitrary random variables Y_l, Y, Y_u, we have $Y_l \leq Y \leq Y_u$ then

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

Using $\mu_n^{(q+K)}(f) \leq \mu_n(f) \leq \mu_n(\pi'_{q+K} f)$ and $\mathbf{E}\,|\mu_n(\pi'_{q+K} f) - \mu_n^{(q+K)}(f)| \leq 2^{-(q+K)}$ for all

$f \in \mathcal{F}$, we obtain

$$
\begin{aligned}
|\,\mathrm{U}_n(f) - U_n^{(q)}(f)| &= \left| \left\{ \sum_{k=1}^{K} U_n^{(q+k)}(f) - U_n^{(q+k-1)}(f) \right\} + \mathrm{U}_n(f) - U_n^{(q+K)}(f) \right| \\
&\leq \left\{ \sum_{k=1}^{K} \left| U_n^{(q+k)}(f) - U_n^{(q+k-1)}(f) \right| + \left| \mathrm{U}_n(\pi'_{q+K} f) - U_n^{(q+K)}(f) \right| \right\} \\
&\quad + \sqrt{n} 2^{-(q+K)}.
\end{aligned}
$$

In order to assure $\frac{\varepsilon}{4} \leq 2^{-(q+K)}\sqrt{n} \leq \frac{\varepsilon}{2}$, for fixed n and q, choose $K = K_{n,q}$, where

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

Here $[\cdot]$ denotes the lower Gauss bracket given by $[x] := \sup\{z \in \mathbb{Z} : z \leq x\}$. For each $i \in \{1, \ldots, N_q\}$, we obtain

$$
\begin{aligned}
\sup_{f \in \mathcal{F}_{q,i}} |\,\mathrm{U}_n(f) - U_n^{(q)}(f)| &\leq \sum_{k=1}^{K_{n,q}} \sup_{f \in \mathcal{F}_{q,i}} |U_n^{(q+k)}(f) - U_n^{(q+k-1)}(f)| \\
&\quad + \sup_{f \in \mathcal{F}_{q,i}} |\,\mathrm{U}_n(\pi'_{q+K_{n,q}} f) - U_n^{(q+K_{n,q})}(f)| + \frac{\varepsilon}{2}.
\end{aligned}
$$

By taking $\varepsilon_k = \frac{\varepsilon}{4k(k+1)}$, $\sum_{k\geq 1} \varepsilon_k = \frac{\varepsilon}{4}$ and we get for each $i \in \{1, \ldots, N_q\}$,

$$
\begin{aligned}
\mathbf{P}^*\left(\sup_{f \in \mathcal{F}_{q,i}} |\,\mathrm{U}_n(f) - U_n^{(q)}(f)| \geq \varepsilon \right) &\leq \sum_{k=1}^{K_{n,q}} \mathbf{P}^*\left(\sup_{f \in \mathcal{F}_{q,i}} |U_n^{(q+k)}(f) - U_n^{(q+k-1)}(f)| \geq \varepsilon_k \right) \\
&\quad + \mathbf{P}^*\left(\sup_{f \in \mathcal{F}_{q,i}} |\,\mathrm{U}_n(\pi'_{q+K_{n,q}} f) - U_n^{(q+K_{n,q})}(f)| \geq \frac{\varepsilon}{4} \right).
\end{aligned}
$$

Notice that the suprema in the r.h.s. are in fact maxima over finite numbers of functions, since the functionals π_q and π'_q (and thus $U_n^{(q)}$) are constant on the $\mathcal{F}_{q,i}$. Therefore, we can work with standard probability theory from this point: the outer probabilities can be replaced by usual probabilities on the right-hand side. For each k, choose a set F_k composed of at most $N_{k-1}N_k$ functions of $\mathcal{F}$ in such a way that F_k contains one function in each non empty $\mathcal{F}_{k-1,i} \cap \mathcal{F}_{k,j}$, $i = 1, \ldots, N_{k-1}$, $j = 1, \ldots, N_k$. Then, for each $i \in \{1, \ldots, N_q\}$, we have

$$
\begin{aligned}
\mathbf{P}^*\left(\sup_{f \in \mathcal{F}_{q,i}} |\,\mathrm{U}_n(f) - U_n^{(q)}(f)| \geq \varepsilon \right) &\leq \sum_{k=1}^{K_{n,q}} \sum_{f \in \mathcal{F}_{q,i} \cap F_{q+k}} \mathbf{P}\left(|U_n^{(q+k)}(f) - U_n^{(q+k-1)}(f)| \geq \varepsilon_k \right) \\
&\quad + \sum_{f \in \mathcal{F}_{q,i} \cap F_{q+K_{n,q}}} \mathbf{P}\left(|\,\mathrm{U}_n(\pi'_{q+K_{n,q}} f) - U_n^{(q+K_{n,q})}(f)| \geq \frac{\varepsilon}{4} \right).
\end{aligned}
$$

Now using Markov's inequality at the order $2p$ (p will be chosen later) and assumption (3.2), we infer

$$\mathbf{P}^*\left(\sup_{f\in\mathcal{F}_{q,i}} |\,\mathrm{U}_n(f) - U_n^{(q)}(f)| \geq \varepsilon\right)$$
$$\leq C_p \sum_{k=1}^{K_{n,q}} \sum_{f\in\mathcal{F}_{q,i}\cap F_{q+k}} \frac{1}{\varepsilon_k^{2p}} \sum_{j=1}^{p} n^{j-p} \|\pi_{q+k}f - \pi_{q+k-1}f\|_s^j \log^{2p+aj}(\|\pi_{q+k}f - \pi_{q+k-1}f\|_{\mathcal{C}} + 1)$$
$$+ C_p \sum_{f\in\mathcal{F}_{q,i}\cap F_{q+K_{n,q}}} \left(\frac{4}{\varepsilon}\right)^{2p} \sum_{j=1}^{p} n^{j-p} \|\pi_{q+K_{n,q}}f - \pi'_{q+K_{n,q}}f\|_s^j$$
$$\cdot \log^{2p+aj}(\|\pi_{q+K_{n,q}}f - \pi'_{q+K_{n,q}}f\|_{\mathcal{C}} + 1).$$

At this point, without loss of generality, we can assume that $a \geq -1$ (if not, take a larger a) and thus the assumption on γ reduces to $\gamma > 2 + a$.

Note that by construction, for each $k \geq 1$,

$$\begin{aligned}
\|\pi_{q+k}f - \pi_{q+k-1}f\|_s &\leq \|\pi_{q+k}f - f\|_s + \|\pi_{q+k-1}f - f\|_s \leq 3\cdot 2^{-(q+k)} \\
\|\pi_{q+k}f - \pi'_{q+k}f\|_s &\leq 2^{-(q+k)} \\
\|\pi_{q+k}f - \pi_{q+k-1}f\|_{\mathcal{C}} &\leq 2\exp(C2^{\frac{q+k}{\gamma}}) \\
\|\pi_{q+k}f - \pi'_{q+k}f\|_{\mathcal{C}} &\leq 2\exp(C2^{\frac{q+k}{\gamma}}).
\end{aligned}$$

Thus,

$$\mathbf{P}^*\left(\sup_{f\in\mathcal{F}_{q,i}} |\,\mathrm{U}_n(f) - U_n^{(q)}(f)| \geq \varepsilon\right)$$
$$\leq 2^{2p+1} C_p \sum_{j=1}^{p} \sum_{k=1}^{K_{n,q}} \#(\mathcal{F}_{q,i}\cap F_{q+k}) \frac{(k(k+1))^{2p}}{\varepsilon^{2p}} n^{j-p} 2^{-j(q+k)} \log^{2p+aj}(2\exp(C2^{\frac{q+k}{\gamma}}) + 1),$$

and if q is large enough,

$$\begin{aligned}
&\mathbf{P}^*\left(\sup_{f\in\mathcal{F}} |\,\mathrm{U}_n(f) - U_n^{(q)}(f)| \geq \varepsilon\right) \\
&\leq \sum_{i=1}^{N_q} \mathbf{P}^*\left(\sup_{f\in\mathcal{F}_{q,i}} |\,\mathrm{U}_n(f) - U_n^{(q)}(f)| \geq \varepsilon\right) \\
&\leq D \sum_{i=1}^{N_q} \sum_{j=1}^{p} \sum_{k=1}^{K_{n,q}} \#(\mathcal{F}_{q,i}\cap F_{q+k}) \frac{(k(k+1))^{2p}}{\varepsilon^{2p}} n^{j-p} 2^{-j(q+k)} 2^{(2p+aj)\frac{q+k}{\gamma}}.
\end{aligned}$$

Here D is a new constant which depends on p, C, and C_p. Since $(\mathcal{F}_{q,i})_{i=1,\ldots,N_q}$ is a partition of $\mathcal{F}$, we have

$$\sum_{i=1}^{N_q} \#(\mathcal{F}_{q,i}\cap F_{q+k}) = \#(F_{q+k}) \leq N_{q+k-1}N_{q+k}$$

and thus

$$\begin{aligned}
&\mathbf{P}^*\left(\sup_{f\in\mathcal{F}}|\,\mathrm{U}_n(f)-U_n^{(q)}(f)|\geq\varepsilon\right)\\
&\leq D'\sum_{j=1}^{p}\frac{n^{j-p}}{\varepsilon^{2p}}\sum_{k=1}^{K_{n,q}}N_{q+k-1}N_{q+k}k^{4p}2^{(2p+(a-\gamma)j)\frac{q+k}{\gamma}}\\
&\leq D'\sum_{j=1}^{p}\frac{n^{j-p}}{\varepsilon^{2p}}2^{(p-j)(\gamma+2+a)\frac{q+K_{n,q}}{\gamma}}\sum_{k=1}^{K_{n,q}}N_{q+k-1}N_{q+k}k^{4p}2^{((-a-\gamma)p+(2+2a)j)\frac{q+k}{\gamma}}\\
&\leq D''\sum_{j=1}^{p-1}\frac{n^{(j-p)\frac{\gamma-(2+a)}{2\gamma}}}{\varepsilon^{2p+(p-j)\frac{\gamma+2+a}{\gamma}}}\sum_{k=1}^{\infty}N_{q+k-1}N_{q+k}k^{4p}2^{(2+a-\gamma)p\frac{q+k}{\gamma}}\\
&\qquad+\frac{D'}{\varepsilon^{2p}}\sum_{k=1}^{\infty}N_{q+k-1}N_{q+k}k^{4p}2^{(2+a-\gamma)p\frac{q+k}{\gamma}},
\end{aligned}\tag{3.7}$$

where D' and D'' are positive constants also depending on p, C, and C_p. Note that we used that we can assume without loss of generality that $a\geq -1$ and thus $(2+2a)j\leq(2+2a)p$ in the last inequality. As $p\frac{2+a-\gamma}{\gamma}\to-\infty$ when p tends to infinity, there exists some $p>1$ such that $p\frac{2+a-\gamma}{\gamma}<-(r+1)$ and thus by (3.6),

$$\sum_{k=2}^{\infty}N_{k-1}N_kk^{4p}2^{p(2+a-\gamma)\frac{k}{\gamma}}\leq\sum_{k=2}^{\infty}N_{k-1}^2k^{4p}2^{p(2+a-\gamma)\frac{k}{\gamma}}+\sum_{k=2}^{\infty}N_k^2k^{4p}2^{p(2+a-\gamma)\frac{k}{\gamma}}<\infty.$$

Therefore the first summand of (3.7) goes to zero as n goes to infinity and the second summand of (3.7) goes to zero as q goes to infinity. □

Proposition 3.1 and Proposition 3.2 establish condition (1.10) and (1.11) of Theorem 1.1 for the random elements U_n, $U_n^{(q)}$, $U^{(q)}$ with value in the complete metric space $\ell^\infty(\mathcal{F})$. This completes the proof of Theorem 3.1. □

3.3. Proof of Lemma 3.1

For $f\in\mathcal{F}$, let π_qf, $q\in\mathbb{N}^*$, be the approximating functions defined in Section 3.2 The entropy condition in Theorem 3.1, yields for every $q\in\mathbb{N}^*$

$$\|f-\pi_qf\|_s\leq 2^{-q}\tag{3.8}$$

$$\|\pi_qf\|_{\mathcal{C}}\leq\exp\big(C2^{\frac{q}{\gamma}}\big),\tag{3.9}$$

where $s\geq 1$ is given in the assumptions of Theorem 3.1. Let $b\in(1,\gamma)$. For all $g\in\mathcal{F}$, and $k\in\mathbb{N}^*$ there exist some $g_k\in\mathcal{G}$ satisfying

$$\|g_k-g\|_s\leq k^{-b}\tag{3.10}$$

$$\|g_k\|_{\mathcal{C}}\leq\exp\big(Ck^{\frac{b}{\gamma}}\big).\tag{3.11}$$

Let $U^{(q)}$ denote the limit process given in Proposition 3.1. By an application the Cramér-Wold device, we deduce from (i) that for all $f, g \in \mathcal{F}$ and $q \in \mathbb{N}^*$

$$\mathbf{Cov}\big(U^{(q)}(f), U^{(q)}(g)\big) = \sum_{k=0}^{\infty} \mathbf{Cov}\big(\pi_q f(X_0), \pi_q g(X_k)\big) + \sum_{k=1}^{\infty} \mathbf{Cov}\big(\pi_q g(X_0), \pi_q f(X_k)\big).$$

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 W of $U^{(q)}$ is given by $\mathbf{Cov}(W(f), W(g)) = \lim_{q\to\infty} \mathbf{Cov}(U^{(q)}(f), U^{(q)}(g))$. It is therefore sufficient to show that for all $f, g \in \mathcal{F}$

$$\Big|\sum_{k=0}^{\infty} \mathbf{Cov}\big(\pi_q f(X_0), \pi_q g(X_k)\big) - \mathbf{Cov}\big(f(X_0), g(X_k)\big)\Big| \tag{3.12}$$
$$+ \Big|\sum_{k=1}^{\infty} \mathbf{Cov}\big(\pi_q g(X_0), \pi_q f(X_k)\big) - \mathbf{Cov}\big(g(X_0), f(X_k)\big)\Big| \longrightarrow 0 \quad \text{as } q \to \infty.$$

By symmetry, both series can be treated the same way. Let $k(q) := 2^{\frac{q}{b}}$. We consider the series in line (3.12). We have

$$\Big|\sum_{k=0}^{\infty} \mathbf{Cov}\big(\pi_q f(X_0), \pi_q g(X_k)\big) - \mathbf{Cov}\big(f(X_0), g(X_k)\big)\Big|$$
$$\leq \sum_{k=0}^{k(q)} \big|\mathbf{Cov}\big(\pi_q f(X_0) - f(X_0), \pi_q g(X_k)\big)\big| + \sum_{k=0}^{k(q)} \big|\mathbf{Cov}\big(f(X_0), \pi_q g(X_k) - g(X_k)\big)\big| \tag{3.13}$$
$$+ \sum_{k=k(q)+1}^{\infty} \big|\mathbf{Cov}\big(\pi_q f(X_0) - f(X_0), \pi_q g(X_k)\big)\big| \tag{3.14}$$
$$+ \sum_{k=k(q)+1}^{\infty} \big|\mathbf{Cov}\big(f(X_0), \pi_q g(X_k) - g(X_k)\big)\big|. \tag{3.15}$$

Let us treat the terms separately. Recall that both $\mathcal{F}$ and $\mathcal{G}$ are uniformly bounded in $\|\cdot\|_\infty$-norm and set

$$M := \sup_{f \in \mathcal{F} \cup \mathcal{G}} \|f\|_\infty.$$

For the term in line (3.13), we know by Hölder's inequality, (3.8), and the fact that $b > 1$ that

$$\sum_{k=0}^{k(q)} \big|\mathbf{Cov}\big(\pi_q f(X_0) - f(X_0), \pi_q g(X_k)\big)\big| + \sum_{k=0}^{k(q)} \big|\mathbf{Cov}\big(f(X_0), \pi_q g(X_k) - g(X_k)\big)\big|$$
$$\leq 2M \sum_{k=0}^{k(q)} \big(\|\pi_q f - f\|_s + \|\pi_q g - g\|_s\big)$$
$$\leq 4M k(q) 2^{-q} = 2^{-(1-\frac{1}{b})q} \longrightarrow 0 \quad \text{as } q \to \infty.$$

For the term in line (3.14), by condition (ii) and inequality (3.9) we obtain

$$\sum_{k=k(q)+1}^{\infty} \big|\mathbf{Cov}\big(\pi_q f(X_0) - f(X_0), \pi_q g(X_k)\big)\big| \leq C' \|\pi_q f - f\|_\infty \|\pi_q g\|_{\mathcal{C}} \sum_{k=k(q)+1}^{\infty} \rho^k$$
$$\leq K \exp\big(C 2^{\frac{q}{\gamma}} + \log(\rho) 2^{\frac{q}{b}}\big) \longrightarrow 0 \quad \text{as } q \to \infty,$$

where K denotes some finite non-negative constant and where we used that $\rho \in (0,1)$ and $b \in (1, \gamma)$. It only remains to show, that the term in line (3.15) goes to zero as $q \to \infty$. We have

$$\sum_{k=k(q)+1}^{\infty} \big|\mathbf{Cov}\big(f(X_0), \pi_q g(X_k) - g(X_k)\big)\big|$$
$$\leq \sum_{k=k(q)+1}^{\infty} \big|\mathbf{Cov}\big(f(X_0), \pi_q g(X_k) - g_k(X_k)\big)\big| \tag{3.16}$$
$$+ \sum_{k=k(q)+1}^{\infty} \big|\mathbf{Cov}\big(f(X_0), g_k(X_k) - g(X_k)\big)\big|. \tag{3.17}$$

First, consider the term in line (3.16). By (ii), (3.9), and (3.11)

$$\sum_{k=k(q)+1}^{\infty} \big|\mathbf{Cov}\big(f(X_0), \pi_q g(X_k) - g_k(X_k)\big)\big|$$
$$\leq C' \sum_{k=k(q)+1}^{\infty} \|f\|_\infty \, \|\pi_q g - g_k\|_{\mathcal{C}} \, \rho^k$$
$$\leq C' \left(\|\pi_q g\|_{\mathcal{C}} \sum_{k=k(q)+1}^{\infty} \rho^k \right) + C' \left(\sum_{k=k(q)+1}^{\infty} \|g_k\|_{\mathcal{C}} \, \rho^k \right)$$
$$\leq K \exp\big(C 2^{\frac{q}{\gamma}} + \log(\rho) 2^{\frac{q}{b}}\big) + C' \sum_{k=k(q)+1}^{\infty} \exp\big(C k^{\frac{b}{\gamma}} + \log(\rho) k\big) \longrightarrow 0 \quad \text{as } q \to \infty$$

for some finite $K > 0$, where we used that $\rho \in (0,1)$ and $b \in (1, \gamma)$ and thus the series on the r.h.s. in the last line converges for each $q \in \mathbb{N}^*$. To treat the term in line (3.17), we apply Hölder's inequality and (3.10), which yields

$$\sum_{k=k(q)+1}^{\infty} \big|\mathbf{Cov}\big(f(X_0), g_k(X_k) - g(X_k)\big)\big| \leq 2M \sum_{k=k(q)+1}^{\infty} \|g_k - g\|_s$$
$$\leq 2M \sum_{k=k(q)+1}^{\infty} k^{-b} \longrightarrow 0 \quad \text{as } q \to \infty,$$

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

4. Entropy of Some Indexing Classes

In many examples that satisfy Assumption 3.I and Assumption 3.II, the normed vector space $\mathcal{C}$ is the space of bounded Lipschitz or α-Hölder continuous functions, see examples in Dehling et al. (2009) and Dehling and Durieu (2011). In this chapter, we therefore restrict our attention to the case where $\mathcal{C}$ is a space of bounded Hölder functions and give several examples of classes $\mathcal{F}$ which satisfy the entropy condition (3.3). Further, we assume that $(\mathcal{X}, d)$ is a metric space equipped with the corresponding Borel σ-algebra. We choose $\mathcal{C} = \mathcal{H}_\alpha(\mathcal{X}, \mathbb{R})$ for some fixed $\alpha \in (0, 1]$. As the approximating class we use the subclass $\mathcal{G} = \mathcal{H}_\alpha(\mathcal{X}, [0, 1]) := \{f \in \mathcal{H}_\alpha(\mathcal{X}, \mathbb{R}) : 0 \leq f \leq 1\}$ of $\mathcal{C}$.

Throughout this chapter, we use the following notations: For an increasing function F from $\mathbb{R}$ to $\mathbb{R}$, F^{-1} denotes the pseudo-inverse function defined by $F^{-1}(t) := \sup\{x \in \mathbb{R} : F(x) \leq t\}$ where $\sup \emptyset := -\infty$. The modulus of continuity of F is defined by

$$\omega_F(\delta) = \sup\{|F(x) - F(y)| : |x - y| \leq \delta\}.$$

Constants that only depend on fixed parameters $p_1, \ldots, p_k$ will be denoted with these parameters in the subscript, such as $c_{p_1,\ldots,p_k}$. Furthermore the notation $f(x) = O_{p_1,\ldots,p_k}(g(x))$ as $x \to 0$ or $x \to \infty$ means that there exists a constant $c_{p_1,\ldots,p_k}$ such that $f(x) \leq c_{p_1,\ldots,p_k} g(x)$ for all x sufficiently small or large, respectively. Except in Section 4.5, in all examples we consider the case where $\mathcal{X}$ is a subset of $\mathbb{R}^d$ equipped with the Euclidean norm denoted by $|\cdot|$, where $d \geq 1$ is some fixed integer.

In most of the examples, we use the transition function given in the following definition which uses the notations

$$d_A(x) := \inf_{a \in A} d(x, a) \qquad \text{and} \qquad d(A, B) := \inf_{a \in A, b \in B} d(a, b),$$

for any element $x \in \mathcal{X}$ and sets A, $B \subset \mathcal{X}$, where we define $\inf \emptyset = +\infty$.

Definition 4.1. Let A, B be subsets of $\mathcal{X}$ such that $d(A, B) > 0$. We define the transition function $T[A, B] : \mathcal{X} \to \mathbb{R}$ by

$$T[A, B](x) := \frac{d_B(x)}{d_B(x) + d_A(x)},$$

if A and B are non-empty, $T[A, B] := 0$ if $A = \emptyset$, and $T[A, B] := 1$ if $B = \emptyset$ but $A \neq \emptyset$.

Observe, that we have $T[A, B](\mathcal{X}) \subset [0, 1]$, $T[A, B](x) = 1$ for all $x \in A$ and $T[A, B](x) = 0$ for all $x \in B$.

Lemma 4.1. *For any subsets A, B of $\mathcal{X}$ such that $d(A,B) > 0$, the transition function $T[A,B]$ is a bounded α-Hölder continuous function and we have*

$$\|T[A,B]\|_\alpha \leq 1 + \left(\frac{3}{d(A,B)}\right)^\alpha.$$

Proof. By the triangle inequality, we have for all $x, y \in \mathcal{X}$ that

$$|d_B(x) - d_B(y)| \leq d(x,y) \qquad \text{and} \qquad d_B(x) + d_A(x) \geq d(A,B).$$

Therefore,

$$\begin{aligned}
&|T[A,B](x) - T[A,B](y)| \\
&= \frac{\big|\big(d_B(x) - d_B(y)\big)\big(d_B(y) + d_A(y)\big) \;+\; d_B(y)\big(d_B(y) + d_A(y)\big) - d_B(y)\big(d_B(x) + d_A(x)\big)\big|}{\big(d_B(x) + d_A(x)\big)\big(d_B(y) + d_A(y)\big)} \\
&= \frac{|d_B(x) - d_B(y)|}{d_B(x) + d_A(x)} \;+\; \frac{d_B(y)}{d_B(y) + d_A(y)} \frac{\big|\big(d_B(y) - d_B(x)\big) + \big(d_A(y) - d_A(x)\big)\big|}{d_B(x) + d_A(x)} \\
&\leq 3\, \frac{d(x,y)}{d(A,B)}
\end{aligned}$$

and thus

$$\begin{aligned}
\|T[A,B]\|_\alpha &:= \|T[A,B]\|_\infty + \sup_{x \neq y} \frac{\big|T[A,B](x) - T[A,B](y)\big|}{d(x,y)^\alpha} \\
&\leq 1 + \sup_{x \neq y} \left(\frac{\big|T[A,B](x) - T[A,B](y)\big|}{d(x,y)}\right)^\alpha \big|T[A,B](x) - T[A,B](y)\big|^{1-\alpha} \\
&\leq 1 + \left(\frac{3}{d(A,B)}\right)^\alpha.
\end{aligned}$$

□

4.1. Indicators of Rectangles

Here, we consider $\mathcal{X} = \mathbb{R}^d$. In its classical form, the empirical process is defined by the class of indicator functions of left infinite rectangles, i.e. the class $\{1_{(-\infty,t]} : t \in \mathbb{R}^d\}$, where $(-\infty, t]$ denotes the set of points x such that $x \leq t$.[1] This case was treated under similar assumptions by Dehling and Durieu (2011). We will see that Theorem 3.1 covers the results of that paper.

The following proposition gives an upper bound for the bracketing number of the larger class

$$\mathcal{F} = \{1_{(t,u]} : t, u \in [-\infty, +\infty]^d, t \leq u\},$$

where $(t, u]$ denotes the rectangle which consists of the points x such that $t < x$ and $x \leq u$.

[1] On $\mathbb{R}^d$, we use the partial order : $x \leq t$ if and only if $x_i \leq t_i$ for all $i = 1, \ldots, d$.

Proposition 4.1. *Let $s \geq 1$, $\mathcal{G} = \mathcal{H}_\alpha(\mathbb{R}^d, [0,1])$, and let μ be a probability distribution on $\mathbb{R}^d$ with distribution function* F*.*

(i) If F *satisfies*

$$\omega_{\mathrm{F}}(x) = \mathcal{O}(|\log(x)|^{-s\gamma}) \quad \text{as } x \to 0 \tag{4.1}$$

for some $\gamma > 1$, then there exists a constant $C = C_{\mathrm{F}} > 0$ such that

$$N\big(\varepsilon, \exp(C\varepsilon^{-1/\gamma}), \mathcal{F}, \mathcal{G}, \mathrm{L}^s(\mu)\big) = \mathcal{O}_d\left(\varepsilon^{-2ds}\right) \quad \text{as } \varepsilon \to 0.$$

(ii) If F *is β-Hölder continuous for some $\beta \in (0,1]$, then exists a constant $C = C_{\mathrm{F}} > 0$ such that stronger bracketing condition*

$$N\big(\varepsilon, C\varepsilon^{-\frac{s\alpha}{\beta}}, \mathcal{F}, \mathcal{G}, L^s(\mu)\big) = \mathcal{O}_d\left(\varepsilon^{-2ds}\right) \quad \text{as } \varepsilon \to 0$$

holds.

Proof. (i) Let $\varepsilon \in (0,1)$ and $m = [6d\varepsilon^{-s} + 1]$. For all $i \in \{1, \dots, d\}$ and $j \in \{0, \dots, m\}$, we define the quantiles

$$t_{i,j} := \mathrm{F}_{(i)}^{-1}\left(\frac{j}{m}\right)$$

where $\mathrm{F}_{(i)}^{-1}$ is the pseudo-inverse of the marginal distribution function $\mathrm{F}_{(i)}$.[2] Now, if $j = (j_1, \dots, j_d) \in \{0, \dots, m\}^d$, we write

$$t_j = (t_{1,j_1}, \dots, t_{d,j_d}).$$

In the following definitions, for convenience, we will also denote by $t_{i,-1}$ or $t_{i,-2}$ the points $t_{i,0}$ and by $t_{i,m+1}$ the points $t_{i,m}$. We introduce the brackets $[l_{k,j}, u_{k,j}]$, $k \in \{0, \dots, m\}^d$, $j \in \{0, \dots, m\}^d$, $k \leq j$, given by the bounded α-Hölder functions

$$l_{k,j}(x) := T\left[[t_{k+1}, t_{j-2}], \mathbb{R}^d \setminus [t_k, t_{j-1}]\right](x),$$

and

$$u_{k,j}(x) := T\left[[t_{k-1}, t_j], \mathbb{R}^d \setminus [t_{k-2}, t_{j+1}]\right](x),$$

where we have used the convention that $[s,t] = \emptyset$ if $s \not\leq t$ and that the addition of an integer to a multi-index is the addition of the integer to every component of the multi-index.

For each $k \leq j$, we have

$$\begin{aligned} \|l_{k,j} - u_{k,j}\|_s^s &\leq \mu\left([t_{k-2}, t_{j+1}] \setminus [t_{k+1}, t_{j-2}]\right) \\ &\leq \sum_{i=1}^d |\,\mathrm{F}_{(i)}(t_{i,k_i+1}) - \mathrm{F}_{(i)}(t_{i,k_i-2})| + |\,\mathrm{F}_{(i)}(t_{i,j_i+1}) - \mathrm{F}_{(i)}(t_{i,j_i-2})| \leq 2\frac{3d}{m}, \end{aligned}$$

[2] $\mathrm{F}_{(i)}(t) := \mu(\mathbb{R} \times \cdots \times \mathbb{R} \times (-\infty, t] \times \mathbb{R} \times \cdots \times \mathbb{R})$.

and thus $\|l_{k,j} - u_{k,j}\|_s \leq \varepsilon$. Moreover, since for $a < b < b' < a'$,

$$d([b,b'], \mathbb{R}^d \setminus [a,a']) = \inf_{i=1,\dots,d} \inf \left\{ |a_i - b_i|, |a'_i - b'_i| \right\},$$

using Lemma 4.1 and (4.1), we have

$$\begin{aligned}
\|l_{k,j}\|_{\mathcal{H}_\alpha} &\leq 1 + 3^\alpha \left(\inf_{i=1,\dots,d} \min\{|t_{i,k_i} - t_{i,k_i+1}|, |t_{i,j_i-1} - t_{i,j_i-2}|\} \right)^{-\alpha} \\
&\leq 1 + 3^\alpha \left[\inf \left\{ x > 0 : \exists i \in \{1,\dots,d\}, \exists t, \mathrm{F}_{(i)}(t+x) - \mathrm{F}_{(i)}(t) \geq \frac{1}{m} \right\} \right]^{-\alpha} \\
&\leq 1 + 3^\alpha \left[\inf \left\{ x > 0 : c_{\mathrm{F}} |\log(x)|^{-s\gamma} \geq \frac{1}{m} \right\} \right]^{-\alpha} \\
&\leq 1 + 3^\alpha \exp\left(\alpha (c_{\mathrm{F}} m)^{\frac{1}{s\gamma}} \right),
\end{aligned}$$

where c_{F} is given by (4.1). The same bound holds for $\|u_{k,j}\|_\alpha$.

Thus, there exists a new constant $C_{\mathrm{F}} > 0$ such that for all $k \leq j \in \{0,\dots,m\}^d$, $[l_{k,j}, u_{k,j}]$ is an $(\varepsilon, \exp(C_{\mathrm{F}} \varepsilon^{-\frac{1}{\gamma}}), \mathcal{G}, \mathrm{L}^s(\mu))$-bracket. It is clear that for each function $f \in \mathcal{F}$ there exists a bracket of the form $[l_{k,j}, u_{k,j}]$ which contains f. Further, we have at most $(m+1)^{2d}$ such brackets, which proves (i).

(ii) Parallel the proof of (i). When computing the bound for $\|l_{k,j}\|_{\mathcal{H}_\alpha}$, use that a β-Hölder function has $|\cdot|^\beta$ as a modulus of continuity. Then one obtains $\|l_{k,j}\|_{\mathcal{H}_\alpha} \leq 1 + 3^\alpha m^{\alpha/\beta}$ and thus $\|l_{k,j}\|_{\mathcal{H}_\alpha} \leq 1 + \left(3^\alpha (6d)^{\alpha/\beta}\right) \varepsilon^{-s\alpha/\beta}$. □

Notice that under the assumptions of the proposition, condition (3.3) is satisfied and therefore Theorem 3.1 may be applied to empirical processes indexed by the class of indicators of rectangles, taking $\mathcal{C}$ to be the class of bounded Hölder functions.

Corollary 4.1. *Let $(X_i)_{i \geq 0}$ be an $\mathbb{R}^d$-valued stationary process. Let $\mathcal{F}$ be the class of indicator functions of rectangles in $\mathbb{R}^d$ and let $\mathcal{G} = \mathcal{H}_\alpha(\mathbb{R}^d, [0,1])$. Assume that, for some $s \geq 1$, $a \in \mathbb{R}$, and $\gamma > \max\{2+a, 1\}$, Assumptions 1 and 2 hold, and that the distribution function* F *of the X_i satisfies (4.1). Then the empirical process $(\mathrm{U}_n(f))_{f \in \mathcal{F}}$ converges in distribution in $\ell^\infty(\mathcal{F})$ to a tight centred Gaussian process.*

Remark 4.1. By regarding the class of indicator functions of left infinite rectangles as a sub-class of $\mathcal{F}$, we obtain Theorem 1 of Dehling and Durieu (2011) as a particular case of the preceding corollary.

4.2. Indicators of Multidimensional Balls in the Unit Cube

Here, we consider the class $\mathcal{F}$ of indicator functions of balls on $\mathcal{X} = [0,1]^d$, i.e.

$$\mathcal{F} := \{1_{\mathrm{B}(x,r)} : x \in [0,1]^d, r \geq 0\}$$

where $\mathrm{B}(x,r) = \{y \in [0,1]^d : |x-y| < r\}$. We have the following upper bound.

Proposition 4.2. *Let μ be a probability distribution on $[0,1]^d$ with a density bounded by some $B > 0$ and let $s \geq 1$. Then there exists a constant $C = C_{d,B} > 0$ such that*

$$N\big(\varepsilon, C\varepsilon^{-\alpha s}, \mathcal{F}, \mathcal{G}, \mathrm{L}^s(\mu)\big) = \mathcal{O}_{d,B}\left(\varepsilon^{-(d+1)s}\right) \quad \textit{as } \varepsilon \to 0,$$

where $\mathcal{G} = \mathcal{H}_\alpha([0,1]^d, [0,1])$.

Note that the second argument in the bracketing number is different from the one appearing in the condition (3.3). In this situation we have a stronger type of bracketing number than in (3.3).

Proof. Let $\varepsilon > 0$ be fixed and $m = [\varepsilon^{-s}]$. For all $i = (i_1, \dots, i_d) \in \{0, \dots, m\}^d$, we denote by c_i the centre of the rectangle $[\frac{i_1-1}{m}, \frac{i_1}{m}] \times \cdots \times [\frac{i_d-1}{m}, \frac{i_d}{m}]$. Then we define, for $i \in \{1, \dots, m\}^d$ and $j \in \{0, \dots, m\}$, the functions

$$l_{i,j}(x) := T\left[\mathrm{B}\left(c_i, \frac{j-2}{m}\sqrt{d}\right), [0,1]^d \setminus \mathrm{B}\left(c_i, \frac{j-1}{m}\sqrt{d}\right)\right](x)$$

and

$$u_{i,j}(x) := T\left[\mathrm{B}\left(c_i, \frac{j+2}{m}\sqrt{d}\right), [0,1]^d \setminus \mathrm{B}\left(c_i, \frac{j+3}{m}\sqrt{d}\right)\right](x),$$

where we use the convention that a ball with negative radius is the empty set.

By Lemma 4.1, these functions are α-Hölder and, since $d(\mathrm{B}(x,r), \mathbb{R}^d \setminus \mathrm{B}(x,r')) = r' - r$, we have

$$\|l_{i,j}\|_{\mathcal{H}_\alpha} \leq 1 + \left(\frac{3m}{\sqrt{d}}\right)^\alpha \leq 1 + 3\varepsilon^{-s\alpha}.$$

The same bound holds for $\|u_{i,j}\|_{\mathcal{H}_\alpha}$. Since μ has a bounded density with respect to Lebesgue measure, we also have

$$\begin{aligned}\|l_{i,j} - u_{i,j}\|_s^s &\leq \mu\left(\mathrm{B}\left(c_i, \frac{j+3}{m}\sqrt{d}\right) \setminus \mathrm{B}\left(c_i, \frac{j-2}{m}\sqrt{d}\right)\right) \\ &\leq Bc_d\left(\left(\frac{j+3}{m}\sqrt{d}\right)^d - \left(\frac{j-2}{m}\sqrt{d}\right)^d\right),\end{aligned}$$

where c_d is the constant $\frac{\pi^{d/2}}{\Gamma(d/2+1)}$ (Γ is the gamma function). Hence,

$$\|l_{i,j} - u_{i,j}\|_s \leq c_{d,B}^{1/s}\varepsilon$$

as $\varepsilon \to 0$, where $c_{d,B}$ is a constant depending only on d and B.

Now, if f belongs to $\mathcal{F}$, then $f = 1_{\mathrm{B}(x,r)}$ for some $x \in [0,1]^d$, and $0 \leq r \leq \sqrt{d}$. Thus, there exist some $i = (i_1, \dots, i_d) \in \{0, \dots, m\}^d$ and $j \in \{0, \dots, m\}$ such that

$$x \in \left[\frac{i_1-1}{m}, \frac{i_1}{m}\right) \times \cdots \times \left[\frac{i_d-1}{m}, \frac{i_d}{m}\right) \quad \text{and} \quad \frac{j}{m}\sqrt{d} \leq r \leq \frac{j+1}{m}\sqrt{d}.$$

We then have $l_{i,j} \leq f \leq u_{i,j}$.

Thus, the $(m+1)m^d$ brackets $[l_{i,j}, u_{i,j}]$, $i \in \{1, \dots, m\}^d$ and $j \in \{0, \dots, m\}$, cover the class $\mathcal{F}$. Therefore, $N(c_{d,B}^{1/s}\varepsilon, 4\varepsilon^{-\alpha s}, \mathcal{F}, \mathcal{G}, \mathrm{L}^s(\mu)) = \mathcal{O}_{d,B}(\varepsilon^{-(d+1)s})$ as $\varepsilon \to 0$, which implies that there exists a constant $C_{d,B} > 0$, for which $N(\varepsilon, C_{d,B}\varepsilon^{-\alpha s}, \mathcal{F}, \mathcal{G}, \mathrm{L}^s(\mu)) = \mathcal{O}_{d,B}(\varepsilon^{-(d+1)s})$ as $\varepsilon \to 0$. □

4.3. Indicators of Uniformly Bounded Multidimensional Ellipsoids Centred in the Unit Cube

Set $\mathcal{X} = \mathbb{R}^d$. Here, we consider the class of ellipsoids which are aligned with the coordinate axes, have their centre in $[0,1]^d$, and their parameters bounded by some constant $D > 0$. Without loss of generality, we assume that $D \in \mathbb{N}^*$. For $x = (x_1, \dots, x_d) \in [0,1]^d$ and all $r = (r_1, \dots, r_d) \in [0, D]^d$, we set

$$\mathrm{E}(x, r) := \left\{ y \in \mathbb{R}^d : \sum_{i=1}^{d} \frac{(y_i - x_i)^2}{r_i^2} \leq 1 \right\}.$$

We denote by $\mathcal{F}$ the class of indicator functions of these ellipsoids, i.e.

$$\mathcal{F} := \{ 1_{\mathrm{E}(x,r)} : x \in [0,1]^d, r \in [0, D]^d \}.$$

We have the following upper bound.

Proposition 4.3. *Let μ be a probability distribution on $\mathbb{R}^d$ with a density bounded by some $B > 0$. Then there exists a constant $C = C_{d,B,D} > 0$ such that*

$$N\big(\varepsilon, C\varepsilon^{-2\alpha s}, \mathcal{F}, \mathcal{G}, \mathrm{L}^s(\mu)\big) = \mathcal{O}_{d,B}\left(\varepsilon^{-2ds}\right) \quad \text{as } \varepsilon \to 0,$$

where $\mathcal{G} = \mathcal{H}_\alpha(\mathbb{R}^d, [0,1])$.

Proof. Let $\varepsilon > 0$ be fixed and $m = [\varepsilon^{-s}]$. For all $i = (i_1, \dots, i_d) \in \{0, \dots, m\}^d$, we denote by I_i the rectangle $[\frac{i_1 - 1}{m}, \frac{i_1}{m}] \times \cdots \times [\frac{i_d - 1}{m}, \frac{i_d}{m}]$. Then, for $i \in \{1, \dots, m\}^d$ and $j = (j_1, \dots, j_d) \in \{0, \dots, Dm-1\}^d$, we define the sets

$$U_{i,j} = \bigcup_{x \in I_i} \mathrm{E}\left(x, \frac{j}{m}\right) = \left\{ y \in \mathbb{R}^d : \min_{x \in I_i} \sum_{k=1}^{d} \frac{(y_k - x_k)^2}{j_k^2} \leq \frac{1}{m^2} \right\}$$

and

$$L_{i,j} = \bigcap_{x \in I_i} \mathrm{E}\left(x, \frac{j}{m}\right) = \left\{ y \in \mathbb{R}^d : \max_{x \in I_i} \sum_{k=1}^{d} \frac{(y_k - x_k)^2}{j_k^2} \leq \frac{1}{m^2} \right\}.$$

We introduce the bracket $[l_{i,j}, u_{i,j}]$ given by

$$l_{i,j}(x) := T\left[L_{i,j-1}, \mathbb{R}^d \setminus L_{i,j}\right](x) \;\text{ and }\; u_{i,j}(x) := T\left[U_{i,j+1}, \mathbb{R}^d \setminus U_{i,j+2}\right](x),$$

where we use the convention that an ellipsoid with one negative parameter is the empty set.

By Lemma 4.1, these functions are α-Hölder. Further, we have the following lemma which is proved in the appendix (cf. Section A.6):

Lemma 4.2. *For all* $j \in \{0, \ldots, Dm-1\}^d$, $x \in \mathbb{R}^d$, *we have*

$$d\left(\mathrm{E}\left(x, \frac{j}{m}\right), \mathbb{R}^d \setminus \mathrm{E}\left(x, \frac{j+1}{m}\right)\right) \geq D^{-1}m^{-2}.$$

As a consequence we infer that the distance between $U_{i,j}$ and $\mathbb{R}^d \setminus U_{i,j+1}$ is at least $D^{-1}m^{-2}$ and the distance between $L_{i,j}$ and $\mathbb{R}^d \setminus L_{i,j+1}$ is at least $D^{-1}m^{-2}$. Thus, by Lemma 4.1, we have

$$\|l_{i,j}\|_{\mathcal{H}_\alpha} \leq 1 + 3^\alpha D^\alpha m^{2\alpha} \leq 1 + 3D\varepsilon^{-2\alpha s},$$

and the same bound holds for $\|u_{i,j}\|_{\mathcal{H}_\alpha}$.

Now, to bound $\|u_{i,j} - l_{i,j}\|_s$ we need to estimate the Lebesgue measures of $U_{i,j}$ and $L_{i,j}$. Recall that, if $j = (j_1, \ldots, j_d) \in \mathbb{R}^d_+$ and $x \in \mathbb{R}^d$, the Lebesgue measure of the ellipsoid $\mathrm{E}(x, j)$ is given by

$$\lambda(\mathrm{E}(x, j)) = c_d \prod_{k=1}^{d} j_k,$$

where c_d is the constant $\frac{\pi^{d/2}}{\Gamma(d/2+1)}$.[3] The set $U_{i,j}$ can be seen as the set constructed as follows: start from an ellipsoid of parameters j/m centred at the centre of I_i, cut it along its hyperplanes of symmetry, and shift each obtained component away from the centre by a distance of $1/2m$ in every direction; $U_{i,j}$ is then the convex hull of these 2^d components (see Figure 4.1 for the dimension 2).

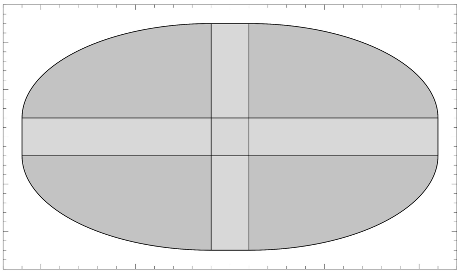


Figure 4.1.: $U_{i,j}$ in dimension 2

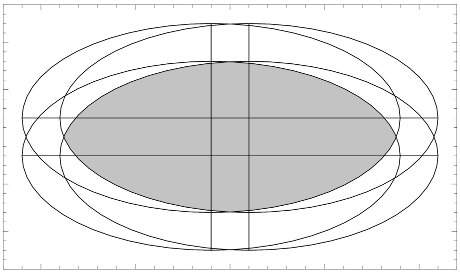


Figure 4.2.: $L_{i,j}$ in dimension 2

Let us denote by $V_{i,j}$ the set that has been added to the 2^d components to obtain the convex hull. We can bound the volume of $U_{i,j}$ by the volume of the ellipsoid plus a bound on the volume of $V_{i,j}$, that is

$$\lambda(U_{i,j}) \leq c_d \prod_{k=1}^{d} \frac{j_k}{m} + \sum_{k=1}^{d} \frac{1}{m} \prod_{l \neq k} \frac{2j_l + 1}{m}.$$

The set $L_{i,j}$ can be seen an the intersection of the 2^d ellipsoids of parameters j/m centred at each corner of the hypercube I_i (see Figure 4.2 for the dimension 2). Its volume is larger than

[3] Here Γ denotes the gamma function.

the volume of an ellipsoid of parameters j/m minus the volume of $V_{i,j}$. We thus have

$$\lambda(L_{i,j}) \geq c_d \prod_{k=1}^{d} \frac{j_k}{m} - \sum_{k=1}^{d} \frac{1}{m} \prod_{l \neq k} \frac{2j_l + 1}{m}.$$

Since μ has a bounded density with respect to Lebesgue measure, we have

$$\begin{aligned} \|l_{i,j} - u_{i,j}\|_s^s &\leq \mu\left(U_{i,j+2} \setminus L_{i,j-1}\right) \\ &\leq B\lambda(U_{i,j+2}) - B\lambda(L_{i,j-1}) \end{aligned}$$

We infer $\|l_{i,j} - u_{i,j}\|_s = c_{d,B}^{1/s}(\varepsilon)$, as $\varepsilon \to 0$, where the constant $c_{d,B}$ only depends on d and B.

Now, if f belongs to $\mathcal{F}$, then $f = 1_{\mathrm{E}(x,r)}$ for some $x \in \mathcal{X}$, and $r \in [0, D]^d$. Thus, there exist some $i = (i_1, \ldots, i_d) \in \{0, \ldots, m\}^d$ and $j \in \{0, \ldots, Dm-1\}^d$ such that

$$x \in \left[\frac{i_1 - 1}{m}, \frac{i_1}{m}\right) \times \cdots \times \left[\frac{i_d - 1}{m}, \frac{i_d}{m}\right)$$

and for each $k = 1, \ldots, d$,

$$\frac{j_k}{m} \leq r_k \leq \frac{j_k + 1}{m}.$$

We then have $l_{i,j} \leq f \leq u_{i,j}$.

Thus, the $D^d m^{2d}$ brackets $[l_{i,j}, u_{i,j}]$, $i \in \{1, \ldots, m\}^d$ and $j \in \{0, \ldots, Dm-1\}^d$, cover the class $\mathcal{F}$. Therefore, there exists a $C_{d,B,D} > 0$, such that $N(\varepsilon, C_{d,B,D}\varepsilon^{-\alpha s}, \mathcal{F}, \mathcal{G}, \mathrm{L}^s(\mu)) = \mathcal{O}_{d,B}(\varepsilon^{-2ds})$, as $\varepsilon \to 0$. □

4.4. Indicators of Uniformly Bounded Multidimensional Ellipsoids

In Section 4.3, we only considered indicators of ellipsoids centred in a compact subset of $\mathbb{R}^d$, namely the unit square. The following lemma will allow us to extend such results to indicators of sets in the whole $\mathbb{R}^d$, at the cost of a moderate additional assumption and a marginal increase of the bracketing numbers.

Lemma 4.3. *Let μ be a measure with continuous distribution function* F*, and $s \geq 1$. Furthermore let $\mathcal{F} := \{1_S : S \in \mathcal{S}\}$, where $\mathcal{S}$ is a class of measurable sets of diameter not larger than $D \geq 1$, and $\mathcal{G} = \mathcal{H}_\alpha(\mathbb{R}^d, [0,1])$. Assume that there are constants $p, q \in \mathbb{N}^*$, $C > 0$, and a function $f : \mathbb{R}_+ \to \mathbb{R}_+$, such that for any $K > 0$ we have*

$$N\big(\varepsilon, f(\varepsilon), \mathcal{F}_K, \mathcal{G}, \mathrm{L}^s(\mu)\big) \leq CK^p \varepsilon^{-q}, \tag{4.2}$$

for sufficiently small ε, where $\mathcal{F}_K := \{1_S : S \in \mathcal{S},\ S \subset [-K, K]^d\}$. If there are some constants $b, \beta > 0$ such that

$$\mu(\{x \in \mathbb{R}^d : |x| > t\}) \leq bt^{-\frac{1}{\beta}}, \tag{4.3}$$

for all sufficiently large t, then

$$N\left(\varepsilon, \max\left\{f(\varepsilon), 4\sqrt{d}(\omega_{\mathrm{F}}^{-1}(2^{-(d+1)}\varepsilon^s))^{-\alpha}\right\}, \mathcal{F}, \mathcal{G}, \mathrm{L}^s(\mu)\right) = O_{\beta,b,C,D,p}(\varepsilon^{-(\beta ps+q)}) \quad \text{as } \varepsilon \to 0,$$

where ω_{F} is the modulus of continuity of F.

The proof is postponed to the appendix (see Section A.7)

Proposition 4.4. *Let $\mathcal{F}$ denote the class of indicators of ellipsoids of diameter uniformly bounded by $D > 0$, which are aligned with coordinate axes (and arbitrary centres in the whole space $\mathbb{R}^d$). If μ is a measure on $\mathbb{R}^d$ with a density bounded by $B > 0$ and if furthermore (4.3) holds for some $\beta > 0$ and $b > 0$, then there exists a constant $C = C_{d,B,D} > 0$ such that*

$$N\left(\varepsilon, C\varepsilon^{-2\alpha s}, \mathcal{F}, \mathcal{G}, \mathrm{L}^s(\mu)\right) = O_{\beta,b,d,B,D,s}\left(\varepsilon^{-(\beta s+2)ds}\right) \quad \text{as } \varepsilon \to 0,$$

where $\mathcal{G} = \mathcal{H}_\alpha(\mathbb{R}^d, [0,1])$.

Proof. In the situation of Section 4.3 change the set of the centres of the ellipsoids $[0,1]^d$ to $[-K,K]^d$ and apply Lemma 4.3. Following the proof of Proposition 4.3 we can easily see that condition (4.2) holds for $p = ds$, $q = 2ds$ and $f(\varepsilon) = C_{d,B,D}\varepsilon^{-2\alpha s}$. Note that since we have a bounded density, we have $\omega_{\mathrm{F}}(x) \leq Bx$ and therefore $4\sqrt{d}(\omega_{\mathrm{F}}^{-1}(2^{-(d+1)}\varepsilon^s))^{-\alpha} \leq 4\sqrt{d}(2^{d+1}B)^\alpha\varepsilon^{-\alpha s} \leq C_{d,B,D}\varepsilon^{-2\alpha s}$ for sufficiently small ε. □

Remark 4.2. In the situation of Proposition 4.4 for the class $\mathcal{F}'$ of indicators of balls in $\mathbb{R}^d$ with uniformly bounded diameter, we can obtain the slightly sharper bound

$$N\big(\varepsilon, C\varepsilon^{-\alpha s}, \mathcal{F}', \mathcal{G}, \mathrm{L}^s(\mu)\big) = O_{\beta,b,d,B,D,s}(\varepsilon^{-((\beta+1)ds+1)s}) \quad \text{as } \varepsilon \to 0$$

for some $C = C'_{d,B} > 0$ by applying Lemma 4.3 directly to the situation in Section 4.2 and using the same arguments as in the previous example.

4.5. Indicators of Balls of an Arbitrary Metric with Common Centre

Let $(\mathcal{X}, d)$ be a metric space and fix $x_0 \in \mathcal{X}$. An x_0-centred ball is given by

$$\mathrm{B}(t) := \{x \in \mathcal{X} : d(x_0, x) \leq t\}.$$

We have the following bound on the bracketing numbers of the class $\mathcal{F} := \{1_{\mathrm{B}(t)} : t > 0\}$.

Proposition 4.5. *Let $s \geq 1$, $\mathcal{G} = \mathcal{H}_\alpha(\mathcal{X}, [0,1])$, and let μ be a probability measure on $\mathcal{X}$ with distribution function* F

(i) If the modulus of continuity ω_G of the function $G(t) := \mu(\mathrm{B}(t))$ satisfies

$$\omega_G(x) = \mathcal{O}(|\log x|^{-s\gamma}) \quad \text{as } x \to 0 \tag{4.4}$$

for some $\gamma > 1$*, then there is a constant* $C = C_G > 0$ *such that*

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

(ii) If G *is* β*-Hölder continuous for some* $\beta \in (0,1]$ *then there is a constant* $C = C_G > 0$ *such that*

$$N\big(\varepsilon, C\varepsilon^{-\frac{s\alpha}{\beta}}, \mathcal{F}, \mathcal{G}, L^s(\mu)\big) = \mathcal{O}(\varepsilon^{-s}) \quad \text{as } \varepsilon \to 0,$$

Remark 4.3. Note that in the case that $\mathcal{X} = \mathbb{R}^2$, $d\mu(t) = \rho(t)dt$, the metric d is given by the Euclidean norm, and $x_0 = 0$, an equivalent condition to (4.4) is

$$\sup_{r \geq 0} \int_r^{r+x} t \int_0^{2\pi} \rho(te^{i\varphi})\; d\varphi\; dt = \mathcal{O}(|\log x|^{-s\gamma}) \quad \text{as } x \to 0.$$

Proof of Proposition 4.5. (i) Fix $\varepsilon > 0$ and choose $m = [1 + 3\varepsilon^{-s}]$. Let G^{-1} denote the pseudo-inverse of G and set for $i \in \{1, \ldots, m\}$

$$r_i := G^{-1}\Big(\frac{i}{m}\Big), \qquad\qquad \mathrm{B}_i := \mathrm{B}(r_i).$$

For convenience set $\mathrm{B}_{-1}, \mathrm{B}_0 := \emptyset$ and $\mathrm{B}_{m+1} = \mathcal{X}$. Define

$$l_i(x) := T\left[\mathrm{B}_{i-2}, \mathcal{X} \setminus \mathrm{B}_{i-1}\right](x) \;\; \text{and} \;\; u_i(x) := T\left[\mathrm{B}_i, \mathcal{X} \setminus \mathrm{B}_{i+1}\right](x)$$

The system $\{[l_i, u_i] : i \in \{1, \ldots, m\}\}$ is a covering for $\mathcal{F}$. Obviously

$$\|u_i - l_i\|_s^s \leq \mu(B_{i+1} \setminus B_{i-2}) \leq \frac{3}{m} \leq \varepsilon^s.$$

By Lemma 4.1, we have

$$\|u_i\|_{\mathcal{H}_\alpha} \leq 1 + \frac{3^\alpha}{d(\mathrm{B}_i, \mathcal{X} \setminus \mathrm{B}_{i+1})^\alpha} \leq 1 + \frac{3^\alpha}{(r_{i+1} - r_i)^\alpha}.$$

Since by condition (4.4)

$$\begin{aligned} r_{i+1} - r_i &\geq \inf\Big\{x > 0 : \exists t \in \mathbb{R} \text{ such that } G(t+x) - G(t) \geq \frac{1}{m}\Big\} \\ &\geq \inf\Big\{x > 0 : \exists t \in \mathbb{R} \text{ such that } \omega_G(x) \geq \frac{1}{m}\Big\} \\ &\geq \exp(-c_G m^{\frac{1}{s\gamma}}) \end{aligned}$$

for some constant $c_G > 0$, there is a constant $C_G > 0$ such that

$$\|u_i\|_{\mathcal{H}_\alpha} \leq 1 + 3^\alpha \exp(\alpha c_G m^{\frac{1}{s\gamma}}) \leq \exp(C_G m^{\frac{1}{s\gamma}}) \leq \exp(C_G \varepsilon^{-\frac{1}{\gamma}}).$$

Analogously, we can show that $\|l_i\|_{\mathcal{H}_\alpha} \leq \exp(C_G\varepsilon^{-\frac{1}{\gamma}})$. This implies that all brackets $[l_i, u_i]$ are $(\varepsilon, \exp(C_G\varepsilon^{-\frac{1}{\gamma}}), \mathcal{F}, \mathcal{G}, \mathrm{L}^s(\mu))$-brackets and thus (i) is proved.

(ii) To prove (ii), follow the proof of (i). In the computation of the bound for the α-norm, use that G is β-Hölder and thus has modulus of continuity $|\cdot|^\beta$. □

4.6. A Class of Monotone Functions.

In this example, we choose $\mathcal{X} = \mathbb{R}$. We consider the case of a one-parameter class of functions $\mathcal{F} = \{f_t : t \in [0,1]\}$, where f_t are functions from $\mathbb{R}$ to $\mathbb{R}$ with the properties:

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

(ii) for all $0 \le s \le t \le 1$, $f_s \le f_t$;

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

Note that all the sequel remains true if in (iii), increasing is replaced by decreasing. Further, for a probability measure μ on $\mathbb{R}$, we define

$$G_\mu : [0,1] \longrightarrow \mathbb{R}, \quad G_\mu(t) := \mu f_t$$

and we say that G_μ is Lipschitz with Lipschitz constant $\lambda > 0$ if $|G_\mu(t) - G_\mu(s)| \le \lambda |t-s|$ for all $s, t \in [0,1]$.

Empirical processes indexed by a 1-parameter class of functions arise, e.g. in the study of empirical U-processes; see Borovkova, Burton, and Dehling (2001). The empirical U-distribution function with kernel function $g(x,y)$ is defined as

$$\mathrm{U}_n(t) = \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} 1_{\{g(X_i, X_j) \le t\}}.$$

Then, the first order term in the Hoeffding decomposition is given by

$$\sum_{i=1}^n g_t(X_i),$$

where $g_t(x) = P(g(x, X_1) \le t)$. For this class of functions, conditions (i) and (ii) are automatically satisfied. Condition (iii) holds, if $g(x,y)$ is monotone in x. This is e.g. the case for the kernel $g(x,y) = y - x$, which arises in the study of the empirical correlation integral; see Borovkova et al. (2001).

Proposition 4.6. *Let $s \ge 1$, $\mathcal{G} = \mathcal{H}_\alpha(\mathbb{R}, [0,1])$, and let μ be a probability measure on $\mathbb{R}$ with distribution function* F*. Assume that G_μ is Lipschitz with Lipschitz constant $\lambda > 0$.*

(i) If F *satisfies*

$$\omega_{\mathrm{F}}(x) = \mathcal{O}(|\log(x)|^{-s\gamma}) \quad \text{as } x \to 0 \tag{4.5}$$

for some $\gamma > 1$, then there exists a $C = C_{\mathrm{F}} > 0$, such that

$$N\big(\varepsilon, \exp(C\varepsilon^{-\frac{1}{\gamma}}), \mathcal{F}, \mathcal{G}, \mathrm{L}^s(\mu)\big) = O_\lambda\left(\varepsilon^{-s}\right) \quad \text{as } \varepsilon \to 0$$

(ii) If F *is* β*-Hölder continuous with* $\beta \in (0,1]$*, then there exists a* $C = C_{\mathrm{F}} > 0$*, such that*

$$N\big(\varepsilon, C\varepsilon^{-\frac{s\alpha}{\beta}}, \mathcal{F}, \mathcal{G}, L^s(\mu)\big) = \mathcal{O}(\varepsilon^{-s}) \quad \text{as } \varepsilon \to 0.$$

Proof. (i) Let $\varepsilon > 0$ and $m = [(\lambda+4)\varepsilon^{-s} + 1]$. For $i = 0, \dots, m$, we set

$$t_i = \frac{i}{m} \text{ and } x_i = \mathrm{F}^{-1}\left(\frac{i}{m}\right).$$

We always have $x_m = +\infty$, but x_0 could be finite or $-\infty$. In order to simplify the notation, in the first case, we change to $x_0 = -\infty$.

We define, for $j \in \{1, \dots, m\}$, the functions l_j and u_j as follows. If $k \in \{1, \dots, m-1\}$, we set $l_j(x_k) = f_{t_{j-1}}(x_{k-1})$ and $u_j(x_k) = f_{t_j}(x_{k+1})$, where we have to understand $f(\pm\infty)$ as $\lim_{x\to\pm\infty} f(x)$. If $k \in \{0, \dots, m-1\}$ and $x \in (x_k, x_{k+1})$, we define $l_j(x)$ and $u_j(x)$ by the linear interpolations,

$$l_j(x) = l_j(x_k) + (x - x_k)\frac{l_j(x_{k+1}) - l_j(x_k)}{x_{k+1} - x_k},$$

$$u_j(x) = u_j(x_k) + (x - x_k)\frac{u_j(x_{k+1}) - u_j(x_k)}{x_{k+1} - x_k},$$

with the exceptions that $l_j(x) = l_j(x_1) = f_{t_{j-1}}(-\infty)$ if $x \in (-\infty, x_1)$ and $u_j(x) = u_j(x_{m-1}) = f_{t_j}(+\infty)$ if $x \in (x_{m-1}, +\infty)$. Then it is clear that for all $t_{j-1} \le t \le t_j$, we have $l_j \le f_t \le u_j$, i.e. f_t belongs to the bracket $[l_j, u_j]$.

Further, being piecewise affine functions, l_j and u_j are α-Hölder continuous functions with Hölder norm

$$\|l_j\|_{\mathcal{H}_\alpha} \le 1 + \max_{k=1,\dots,m} \frac{l_j(x_k) - l_j(x_{k-1})}{(x_k - x_{k-1})^\alpha} \le 1 + \max_{k=1,\dots,m} \frac{1}{(x_k - x_{k-1})^\alpha} \le 1 + \exp\left(C_{\mathrm{F}} m^{\frac{1}{s\gamma}}\right).$$

Here we have used the condition (4.5) and the same computation as for the class of indicators of rectangles. Analogously, the same bound holds for $\|u_j\|_{\mathcal{H}_\alpha}$.

Now,

$$\|u_j - l_j\|_s^s \le \|u_j - l_j\|_1 \le \|u_j - f_{t_j}\|_1 + \|f_{t_j} - f_{t_{j-1}}\|_1 + \|l_j - f_{t_{j-1}}\|_1.$$

First, since G_μ is Lipschitz, we have

$$\|f_{t_j} - f_{t_{j-1}}\|_1 \le G(t_j) - G(t_{j-1}) \le \lambda(t_j - t_{j-1}) = \frac{\lambda}{m}.$$

For $x \in [x_{k-1}, x_k]$, since f_t is increasing, we have $u_j(x) \le f_{t_j}(x_{k+1})$ and $u_{t_j}(x) \ge f_{t_j}(x_{k-1})$, thus

$$\begin{aligned}\|u_j - f_{t_j}\|_1 &\le \sum_{k=1}^{m-1} |f_{t_j}(x_{k+1}) - f_{t_j}(x_{k-1})| \mu([x_k, x_{k+1}]) \\ &\le \frac{1}{m}\sum_{k=1}^{m-1} (|f_{t_j}(x_{k+1}) - f_{t_j}(x_k)| + |f_{t_j}(x_k) - f_{t_j}(x_{k-1})|) \le \frac{2}{m}\end{aligned}$$

since, by monotonicity, $\sum_{k=0}^{m-1} |f_{t_j}(x_{k+1}) - f_{t_j}(x_k)| \leq 1$. In the same way we get $\|l_j - f_{t_{j-1}}\|_1 \leq \frac{2}{m}$ and we infer

$$\|u_j - l_j\|_s \leq \left(\frac{\lambda + 4}{m}\right)^{1/s} \leq \varepsilon.$$

Thus, the number of $(\varepsilon, \exp(C_{\mathrm{F}}\varepsilon^{-\frac{1}{\gamma}}), \mathcal{G}, \mathrm{L}^s(\mu))$-brackets needed to cover the class $\mathcal{F}$ is bounded by m, which proves the proposition.

(ii) Part (ii) can be shown analogously, where one uses that β-Hölder continuous functions have modulus of continuity $|\cdot|^\beta$. □

5. An ECLT for the Ergodic Automorphism of the Torus

In this section we establish an empirical CLT for processes whose behaviour is determined by an ergodic automorphism on the d-dimensional torus. Dehling and Durieu (2011) established an empirical CLT where the underlying random variables of the empirical process are directly the iterates of the automorphism themselves and where the empirical process is indexed by the class of rectangles of the form $[0,t]$, $t \in [0,1]^d$. Here, we generalize this result to an empirical CLT with different indexing classes and to the case, where the underlying random variables are functionals of the iterate of the automorphism under some regular function.

Let $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$ be the torus of dimension $d \geq 2$, which is identified with $[0,1]^d$ and equipped with the Lebesgue measure λ. Recall, that the automorphism of the torus $T : \mathbb{T}^d \to \mathbb{T}^d$ introduced in Section 2.6 is given by

$$Tx = Ax \mod 1,$$

where A is a square matrix of dimension d with integer coefficients and determinant ± 1 and such that no eigenvalue of A is a root of unity. Recall that T is measure preserving and ergodic w.r.t. λ (cf. Section 2.6). In the remainder of this chapter, we denote by $\|\cdot\|_1$ the $\mathrm{L}^1(\lambda)$-norm given by $\|f\|_1 := \lambda(|f|)$. Further, for the ith iterate of T, we write T^i. By T^0 we denote the identity map on $\mathbb{T}^d$.

We extend the definition of the $\mathbb{R}$- or $\mathbb{C}$-valued bounded α-Hölder continuous functions from Section 2.3 to bounded α-Hölder functions with values in $\mathbb{R}^\ell$, $\ell \in \mathbb{N}^*$, by replacing the absolute value (or modulus) in the definition of $\|\cdot\|_{\mathcal{H}_\alpha}$ by the corresponding euclidean norm. We denote the space of such functions defined on a space $\mathcal{X}$ by $\mathcal{H}_\alpha(\mathcal{X}, \mathbb{R}^\ell)$.

As an application of Theorem 3.2, we establish the following proposition.

Theorem 5.1 (Empirical CLT for Ergodic Automorphisms of the Torus)**.** *Let $\mathcal{F}$ be a uniformly bounded class of functions on $\mathbb{R}^\ell$, $\ell \in \mathbb{N}^*$, $\varphi \in \mathcal{H}_\beta(\mathbb{T}^d, \mathbb{R}^\ell)$, $\beta \in (0,1]$, and let d_0 denote the size of the biggest Jordan block of T restricted to its neutral subspace. If the entropy condition (3.3) holds with $\mu = \lambda \circ \varphi^{-1}$ and $s = 1$ for some uniformly bounded subset $\mathcal{G}$ of $\mathcal{H}_\alpha(\mathbb{R}^\ell, \mathbb{R})$ with $\alpha \in (0,1]$, $r \geq -1$, $C > 0$ and $\gamma > \max\{1, J\}$, then the empirical process $\mathrm{U}_n = (\mathrm{U}_n(f))_{f \in \mathcal{F}}$ given by*

$$\mathrm{U}_n(f) := \frac{1}{\sqrt{n}} \left(\sum_{i=1}^{n} f \circ \varphi(T^i) - \lambda(f \circ \varphi) \right)$$

converges in distribution in $\ell^\infty(\mathcal{F})$ to a tight centred Gaussian process W.

Proof. We use Theorem 3.2 with $\mathcal{X} = \mathbb{R}^\ell$, $\mathcal{C} = \mathcal{H}_\alpha(\mathbb{R}^\ell, \mathbb{R})$, and $X_i = \varphi(T^i)$. For $f \in \mathcal{H}_\alpha(\mathbb{R}^\ell, \mathbb{R})$ and $\varphi \in \mathcal{H}_\beta(\mathbb{T}^d, \mathbb{R}^\ell)$ the function $f \circ \varphi$ is an element $\mathcal{H}_{\alpha\beta}(\mathbb{T}^d, \mathbb{R})$. Dehling and Durieu (2011) showed that $(T^i)_{i\in\mathbb{N}}$ is multiple mixing w.r.t. $\mathcal{H}_{\alpha\beta}(\mathbb{T}^d, \mathbb{R})$ for any $\alpha, \beta \in (0, 1]$, $s = 1$, and d_0 not larger than the size J of the biggest Jordan block of T restricted to its neutral subspace (see Proposition A.2). Therefore $(X_i)_{i\in\mathbb{N}}$ is multiple mixing w.r.t. $\mathcal{H}_\alpha(\mathbb{R}^\ell, \mathbb{R})$ with the same s and d_0. By a result of Leonov (1960) (see also Le Borgne (1999)) the CLT holds under all functions in $\mathcal{H}_{\alpha\beta}(\mathbb{T}^d, \mathbb{R})$. Then $(X_i)_{i\in\mathbb{N}} = (\varphi \circ T^i)_{i\in\mathbb{N}}$ satisfies the CLT under $\mathcal{H}_\alpha(\mathbb{R}^\ell, \mathbb{R})$ and thus Theorem 3.2 applies. □

Theorem 5.1 provides no information about the covariance structure of the limiting process. If in the situation of Theorem 5.1 we assume that furthermore there is some (possibly infinite) covering of $\mathcal{F}$ by $\mathrm{L}^1(\mu)$-balls with centres in $\mathcal{G}$ that provide some polynomial control of the α-norm of those central functions, we can identify the covariance structure of the limiting process as the following lemma shows.

Lemma 5.1. *If the assumptions of Theorem 5.1 are satisfied with $\gamma > 2$ and if furthermore for every $f \in \mathcal{F}$ there exits a $\tilde{f}_k \in \mathcal{G}$ such that*

$$\|(f - \tilde{f}_k) \circ \varphi\|_1 \le k^{-1}, \tag{5.1}$$

$$\|\tilde{f}_k\|_{\mathcal{H}_\alpha} \le C' k^a \tag{5.2}$$

for some $a, C' > 0$, then the covariance structure of the limiting process W is given by

$$\mathbf{Cov}\big(W(f), W(g)\big) = \sum_{k=0}^{\infty} \mathbf{Cov}\big(f(\varphi), g(\varphi(T^k))\big) + \sum_{k=1}^{\infty} \mathbf{Cov}\big(f(\varphi(T^k)), g(\varphi)\big) \quad f, g \in \mathcal{F}. \tag{5.3}$$

Proof. For $i \in \mathbb{N}$ let $X_i = \varphi(T^i)$ and follow the proof of Lemma 3.1 in Section 3.3. Condition (i) of Lemma 3.1 is satisfied due to Leonov (1960) (cf. Le Borgne (1997)). However we can not assume that (ii) holds. We therefore need to show that the terms (3.14) and (3.15) vanish as $q \to \infty$. All remaining terms can be treated the same way as before. Observe that in the case of the ergodic automorphism of the torus, there exists a $\rho \in (0, 1)$ and a $K > 0$ such that for all $f \in \mathcal{F}$ and all $h \in \mathcal{H}_\alpha(\mathbb{R}^\ell, \mathbb{R})$

$$\big|\mathbf{Cov}\big(f(X_0), h(X_n)\big)\big| \le K \|h\|_{\mathcal{H}_\alpha}\, \rho^n. \tag{5.4}$$

To see this let $\tilde{f}_k$, $k \in \mathbb{N}^*$ satisfy (5.1) and (5.2), then by Lemma A.1 and Hölder's inequality there is a $D > 0$ and a $\theta \in (0, 1)$ such that

$$\begin{aligned}\big|\mathbf{Cov}\big(f(X_0), h(X_n)\big)\big| &\le \big|\mathbf{Cov}\big(\tilde{f}_k(X_0), h(X_n)\big)\big| + \big|\mathbf{Cov}\big(f(X_0) - \tilde{f}_k(X_0), h(X_n)\big)\big| \\ &\le D\big(\|\tilde{f}_k\|_{\mathcal{H}_\alpha} \|h\|_{\mathcal{H}_\alpha} \theta^n + \|(f - \tilde{f}_k) \circ \varphi\|_1 \, \|h \circ \varphi\|_\infty\big) \quad \text{for all } k \in \mathbb{N}^*.\end{aligned}$$

Here we used that $\|h \circ \varphi\|_{\mathcal{H}_{\alpha\beta}} \le \|h\|_{\mathcal{H}_\alpha}(1 + \|\varphi\|^\alpha_{\mathcal{H}_\beta})$ for all $h \in \mathcal{H}_\alpha(\mathbb{R}^\ell, \mathbb{R})$ and $\varphi \in \mathcal{H}_\beta(\mathbb{R}^d, \mathbb{R}^\ell)$.

Now, setting $k = [\theta^{-n/(2a)}] + 1$, by (5.1) and (5.2) there is a $K > 0$ such that

$$\left|\mathbf{Cov}\big(f(X_0), h(X_n)\big)\right| \leq 2^{-1} K\big(\|h\|_{\mathcal{H}_\alpha}\theta^{\frac{n}{2}} + \|h\|_{\mathcal{H}_\alpha}\theta^{\frac{n}{2a}}\big).$$

This implies (5.4) with $\rho = \max\{\theta^{1/2}, \theta^{1/(2a)}\}$.

Now consider the term (3.14). Recall that $k(q) = 2^{q/b}$, where b was an arbitrary real number in $(1, \gamma)$. Here, since γ is assumed to be larger than 2, we may chose $b \in (1, \gamma/2)$. Using Lemma A.1 and (5.4), we have

$$\begin{aligned}
&\sum_{k=k(q)+1}^{\infty} \left|\mathbf{Cov}\big((\pi_q f)(X_0) - f(X_0), (\pi_q g)(X_k)\big)\right| \\
&\leq \sum_{k=k(q)+1}^{\infty} \left|\mathbf{Cov}\big((\pi_q f)(X_0), (\pi_q g)(X_k)\big)\right| + \sum_{k=k(q)+1}^{\infty} \left|\mathbf{Cov}\big(f(X_0), (\pi_q g)(X_k)\big)\right| \\
&\leq \Big(D\|\pi_q f\|_{\mathcal{H}_\alpha}\|\pi_q g\|_{\mathcal{H}_\alpha} \sum_{k=k(q)+1}^{\infty} \theta^k\Big) + \Big(K\|\pi_q g\|_{\mathcal{H}_\alpha} \sum_{k=k(q)+1}^{\infty} \rho^k\Big) \\
&\leq K_\varphi\Big(\exp\big(C2^{\frac{2q}{\gamma}} + \log(\theta)2^{\frac{q}{b}}\big) + \exp\big(C2^{\frac{q}{\gamma}} + \log(\rho)2^{\frac{q}{b}}\big)\Big) \longrightarrow 0 \quad \text{as } q \to \infty,
\end{aligned}$$

where $K_\varphi > 0$ is some finite constant and where we use that $\rho, \theta \in (0, 1)$ and $b \in (1, \gamma/2)$. With similar arguments we obtain for (3.15) that

$$\begin{aligned}
&\sum_{k=k(q)+1}^{\infty} \left|\mathbf{Cov}\big(f(X_0), \pi_q g(X_k) - g_k(X_k)\big)\right| \\
&\leq \sum_{k=k(q)+1}^{\infty} \left|\mathbf{Cov}\big(f(X_0), \pi_q g(X_k)\big)\right| + \sum_{k=k(q)+1}^{\infty} \left|\mathbf{Cov}\big(f(X_0), g_k(X_k)\big)\right| \\
&\leq \sum_{k=k(q)+1}^{\infty} \exp\big(C2^{\frac{q}{\gamma}}\big)\rho^k + \sum_{k=k(q)+1}^{\infty} \exp\big(Ck^{\frac{b}{\gamma}}\big)\rho^k \longrightarrow 0 \quad \textit{as } q \to \infty,
\end{aligned}$$

which finishes the proof. □

Remark 5.1. The assumptions in Lemma 5.1 and the entropy condition (3.3) in Theorem 5.1 can be simplified to the following stronger condition: Let there be constants $r > -1$, $a, C' > 0$ such that

$$\int_0^1 \varepsilon^r \sup_{\varepsilon \leq \delta \leq 1} N^2\big(\delta, C'\delta^{-a}, \mathcal{F}, \mathcal{G}, \mathrm{L}^1(\mu)\big) d\varepsilon < \infty.$$

Under this condition (3.3) is easily satisfied for any $\gamma > 1$. Further, this condition even gives us a finite covering of $\mathcal{F}$ by $\big(\varepsilon, C'\varepsilon^{-a}, \mathcal{G}, \mathrm{L}^1(\mu)\big)$-brackets, where Lemma 5.1 only requires a (possibly infinity) covering by corresponding balls and thus (5.1) and (5.2) are satisfied.

Applying the results from Chapter 4, Theorem 5.1 and Remark 5.1 yields the following corollary.

Corollary 5.1. *Let $d \geq 2$ and T be an ergodic d-dimensional automorphism of the torus $\mathbb{T}^d$ with J the size of the biggest Jordan block of T restricted to its neutral subspace. Let $\varphi \in \mathcal{H}_\beta(\mathbb{T}^d, \mathbb{R}^\ell)$, $\ell \in \mathbb{N}^*$, and let F denote the distribution function of $\mu := \lambda \circ \varphi^{-1}$.*

(i) If the modulus of continuity ω_{F} of F satisfies $\omega_{\mathrm{F}}(x) = \mathcal{O}(|\log(x)|^{-\gamma})$ as $x \to 0$ for some $\gamma > \max\{1, J\}$ then the empirical CLT holds w.r.t. to the class of indicators of finite and infinite rectangles on $\mathbb{R}^\ell$ given in Section 4.1.

The empirical CLT holds further w.r.t. the 1-parameter class $\mathcal{F} = \{f_t : t \in [0,1]\}$ of monotone function introduced in Section 4.6 if $G_\mu : [0,1] \longrightarrow \mathbb{R},\ t \mapsto \mu f_t$ is Lipschitz.

(ii) If F is β'-Hölder continuous w.r.t. some $\beta' \in (0,1]$ then the empirical CLT holds w.r.t. the class of indicators of finite and infinite rectangles given in Section 4.1 with covariance structure given by (5.3).

If further G_μ is Lipschitz, then the empirical CLT holds w.r.t. the 1-parameter class of monotone function introduced in that section, where the covariance structure of the limiting process is given by (5.3).

(iii) If μ has a bounded density, then the empirical CLT holds w.r.t. the classes of indicators of multidimensional balls and indicators of uniformly bounded multidimensional ellipsoids introduced in Section 4.2, 4.3, and 4.4 with covariance structure given by (5.3).

□

Part II.

Empirical Central Limit Theorems for Slowly Multiple Mixing Processes

Based on the article
Olivier Durieu and Marco Tusche (2012): An empirical process central limit theorem for multidimensional dependent data. J. Theoret. Probab.. DOI 10.1007/s10959-012-0450-3

Keywords: Multivariate empirical processes, Limit theorems, Multiple mixing, Chaining

Mathematics Subject Classification (2010): 62G30, 60F17, 60G10

6. Slowly Multiple Mixing Processes

In the earlier sections, we considered multiple mixing processes with some exponential decrease of the covariances. This property led to an increment bound of $2p$th moments of the approximation process (cf. Proposition 2.1). The aim of the following chapters is to extend our approximating class approach (see Section 1.3) to situations, where we have a much slower decrease of the covariances that appear in the definition of the multiple mixing property. We consider processes, where the decay of the covariances can be described by a summable sequence. In order to treat such processes, we develop a version of Theorem 3.1 which can be applied when only weaker moment bounds than in Part I are available. This is achieved mainly by introducing balancing conditions between the distribution of X_0 and properties of the approximating class of functions. As a concrete application, we obtain results for causal functions of i.i.d. processes.

We define the slow multiple mixing property as follows.

Definition 6.1 (Slow Multiple Mixing)**.** Let $(X_i)_{i\in\mathbb{N}}$ be a stationary stochastic process of $\mathbb{R}^d$-valued random variables, and let $\mathcal{C}$ be a space of measurable real-valued functions defined on $\mathbb{R}^d$ and equipped with a semi-norm $\|\cdot\|_{\mathcal{C}}$. For integers $i_1, ..., i_j \in \mathbb{N}$, we write $i_j^* := i_1 + \ldots + i_j$. We say that $(X_n)_{n\in\mathbb{N}}$ is slowly multiple mixing with respect to $\mathcal{C}$ if there exist a constant $s \geq 1$ and a decreasing function $\Theta : \mathbb{N} \longrightarrow \mathbb{R}_0^+$ such that for any $p \in \mathbb{N}^*$, there is a constant $K_p < \infty$ satisfying

$$\begin{aligned} &\left|\mathbf{Cov}\big(f(X_0)f(X_{i_1^*})\cdot\ldots\cdot f(X_{i_{q-1}^*})\,,\,f(X_{i_q^*})f(X_{i_{q+1}^*})\cdot\ldots\cdot f(X_{i_p^*})\big)\right| \\ &\quad\leq K_p\|f(X_0)\|_s\|f\|_{\mathcal{C}}\Theta(i_q) \end{aligned} \tag{6.1}$$

for all $f \in \mathcal{C}$ with $\|f\|_\infty \leq 1$ and $\mathbf{E}(f(X_0)) = 0$ and all $i_1, \ldots, i_p \in \mathbb{N}$, $q \in \{1, \ldots, p\}$.

Remark 6.1. From the abstract definition, the notion of "slow" may seem a bit irritating, since Θ is not specified here. However, in applications we will use this notion only in the context, where $\Theta(k)$ goes to zero with a slower rate than θ^k with $\theta \in (0, 1)$. Note that we also do not allow a polynomial term Q of strictly positive degree on the r.h.s. of (6.1), thus the notion of slow multiple mixing does not include every case of (exponential) multiple mixing as introduced in the earlier chapters of this thesis.

The following section establishes moment bounds for slowly multiple mixing processes.

6.1. Moment Bounds for Slowly Multiple Mixing Processes

Proposition 6.1. *Let $(X_n)_{n\in\mathbb{N}}$ be slowly multiple mixing w.r.t. $\mathcal{C}$ with $s \geq 1$ and a function Θ such that for some $p \in \mathbb{N}^*$*

$$\sum_{i=0}^{\infty} i^{2p-2}\Theta(i) < \infty. \tag{6.2}$$

Then there is a finite constant $C > 0$ such that for all $f \in \mathcal{C}$ with $\|f\|_\infty \leq 1$ and all $n \in \mathbb{N}^*$, we have

$$\begin{aligned}\mathbf{E}\Big(\Big|\sum_{i=1}^{n} f(X_i) - \mathbf{E}(f(X_0))\Big|^{2p}\Big)& \\ \leq C\sum_{i=1}^{p} n^i \left\|f(X_0) - \mathbf{E}(f(X_0))\right\|_s^i \left(\|f - \mathbf{E}(f(X_0))\|_{\mathcal{C}}\right)^i&\end{aligned} \tag{6.3}$$

Proof. Without loss of generality, assume that $\mathbf{E}(f(X_0)) = 0$. By stationarity, we have

$$\begin{aligned}\Big|\mathbf{E}\Big(\Big(\sum_{i=1}^{n} f(X_i)\Big)^p\Big)\Big| &= \Big|\sum_{1\leq i_1,\dots,i_p\leq n} \mathbf{E}\big(f(X_{i_1})\cdot\ldots\cdot f(X_{i_p})\big)\Big| \\ &\leq p!n\Big|\sum_{\substack{0\leq i_1,\dots,i_{p-1}\leq n-1\\ i^*_{p-1}\leq n-1}} \mathbf{E}\big(f(X_0)f(X_{i_1^*})\cdot\ldots\cdot f(X_{i^*_{p-1}})\big)\Big|.\end{aligned}$$

Using the notations $I_n(0) := \big|\mathbf{E}(f(X_0))\big| = 0$ and

$$I_n(p) := \sum_{\substack{0\leq i_1,\dots,i_p\leq n-1\\ i^*_p\leq n-1}} \big|\mathbf{E}\big(f(X_0)f(X_{i_1^*})\cdot\ldots\cdot f(X_{i^*_p})\big)\big|, \tag{6.4}$$

for $p \in \mathbb{N}^*$, we therefore have

$$\Big|\mathbf{E}\Big(\Big(\sum_{i=1}^{n} f(X_i)\Big)^p\Big)\Big| \leq p!nI_n(p-1). \tag{6.5}$$

Decomposing the sum in (6.4) with respect to the highest increment of indices i_q, $q \in \{1, \dots, p\}$, we receive a bound

$$I_n(p) \leq \sum_{q=1}^{p} J_n(p, q),$$

where

$$J_n(p, q) = \sum_{i_q=0}^{n-1} \sum_{\substack{0\leq i_1,\dots,i_{q-1},i_{q+1},\dots,i_p\leq i_q\\ i^*_p\leq n-1}} \big|\mathbf{E}\big(f(X_0)f(X_{i_1^*})\cdot\ldots\cdot f(X_{i^*_p})\big)\big|.$$

We successively treat the terms $J_n(p,q)$ and $I(p)$ in two separate lemmata.

Lemma 6.1. *Let $p \in \mathbb{N}^*$. If $(X_n)_{n\in\mathbb{N}}$ is slowly multiple mixing w.r.t. $\mathcal{C}$ with $s \geq 1$ and $\Theta : \mathbb{N} \longrightarrow \mathbb{R}_0^+$ such that*

$$\sum_{i=0}^{\infty} i^{p-1}\Theta(i) < \infty, \tag{6.6}$$

then for all $q \in \{1,\dots,p\}$ there exists a constant $K' > 0$ such that

$$J_n(p,q) \leq K'\|f(X_0)\|_s\|f\|_{\mathcal{C}} + nI_n(q-1)I_n(p-q) \quad \text{for all } n \in \mathbb{N}^* \text{ and } f \in \mathcal{C}.$$

Proof. Set

$$\begin{aligned} A_{i_1,\dots,i_p} &:= \left|\mathbf{Cov}\left(f(X_0)f(X_{i_1^*})\cdot\ldots\cdot f(X_{i_{q-1}^*})\,,\, f(X_{i_q^*})f(X_{i_{q+1}^*})\cdot\ldots\cdot f(X_{i_p^*})\right)\right| \\ B_{i_1,\dots,i_p} &:= \left|\mathbf{E}\left(f(X_0)f(X_{i_1^*})\cdot\ldots\cdot f(X_{i_{q-1}^*})\right)\right|\cdot\left|\mathbf{E}\left(f(X_0)f(X_{i_{q+1}})\cdot\ldots\cdot f(X_{i_p^*-i_q^*})\right)\right|, \end{aligned}$$

where we used the stationarity of $(X_i)_{i\in\mathbb{N}}$ in the last line. We have

$$\begin{aligned} &J_n(p,q) \\ &\quad\leq \sum_{i_q=0}^{n-1} \sum_{\substack{0\leq i_1,\dots,i_{q-1},i_{q+1},\dots,i_p\leq i_q \\ i_p^*\leq n-1}} A_{i_1,\dots,i_p} + \sum_{i_q=0}^{n-1} \sum_{\substack{0\leq i_1,\dots,i_{q-1},i_{q+1},\dots,i_p\leq i_q \\ i_p^*\leq n-1}} B_{i_1,\dots,i_p}. \end{aligned}$$

An application of the slow multiple mixing property (6.1) yields

$$\begin{aligned} \sum_{i_q=0}^{n-1} \sum_{\substack{0\leq i_1,\dots,i_{q-1},i_{q+1},\dots,i_p\leq i_q \\ i_p^*\leq n-1}} A_{i_1,\dots,i_p} &\leq K\|f(X_0)\|_s\|f\|_{\mathcal{C}} \sum_{i_q=0}^{n-1}(i_q+1)^{p-1}\Theta(i_q) \\ &\leq K'\|f(X_0)\|_s\|f\|_{\mathcal{C}} \end{aligned}$$

for some constant $K' < \infty$, since $\sum_{i_q=0}^{\infty} i_q^{p-1}\Theta(i_q) < \infty$ by (6.6). Finally,

$$\begin{aligned} \sum_{\substack{0\leq i_1,\dots,i_{q-1},i_{q+1},\dots,i_p\leq i_q \\ i_p^*\leq n-1}} B_{i_1,\dots,i_p} \leq &\sum_{\substack{0\leq i_1,\dots,i_{q-1}\leq n-1 \\ i_{q-1}^*\leq n-1}} \left|\mathbf{E}\left(f(X_0)f(X_{i_1^*})\cdot\ldots\cdot f(X_{i_q^*})\right)\right| \\ &\cdot \sum_{\substack{0\leq i_{q+1},\dots,i_p\leq n-1 \\ i_p^*-i_q^*\leq n-1}} \left|\mathbf{E}\left(f(X_0)f(X_{i_{q+1}})\cdot\ldots\cdot f(X_{i_p^*-i_q^*})\right)\right| \end{aligned}$$

and thus

$$\sum_{i_q=0}^{n-1} \sum_{\substack{0\leq i_1,\dots,i_{q-1},i_{q+1},\dots,i_p\leq i_q \\ i_p^*\leq n-1}} B_{i_1,\dots,i_p} \leq nI_n(p-1)I_n(p-q).$$

□

From now on, let $\lceil \cdot \rceil$ denote the upper Gauss bracket $\lceil x \rceil := \min\{z \in \mathbb{Z} : z \geq x\}$.

Lemma 6.2. *Let $p \in \mathbb{N}^*$. Assume that $(X_n)_{n\in\mathbb{N}}$ is slowly multiple mixing with $s \geq 1$ and Θ such that (6.6) holds. Then there is a constant $K_p < \infty$ such that*

$$I_n(p) \leq K_p \sum_{i=1}^{\lceil p/2 \rceil} n^{i-1} \|f(X_0)\|_s^i \|f\|_{\mathcal{C}}^i \tag{6.7}$$

for all $f \in \mathcal{C}$ with $\|f\|_\infty \leq 1$ and $\mathbf{E}(f(X_0)) = 0$.

Proof. We will use complete induction to prove the lemma. By Lemma 6.1 we can easily see that

$$I_n(1) \leq K_1 \|f(X_0)\|_r \, \|f\|_{\mathcal{C}}$$

for some constant $K_1 < \infty$ if (6.6) is satisfied. Now consider an arbitrary $\tilde{p} \geq 2$ satisfying (6.6) and assume that (6.7) holds for all $p \leq \tilde{p} - 1$. We have

$$\begin{aligned}
I_n(\tilde{p}) &\leq \sum_{q=1}^{\tilde{p}} J_n(\tilde{p}, q) \\
&\leq \sum_{q=1}^{\tilde{p}} \Big(K' \|f(X_0)\|_s \|f\|_{\mathcal{C}} + n I_n(q-1) I_n(\tilde{p}-q) \Big) \\
&\leq \tilde{p} K' \|f(X_0)\|_s \|f\|_{\mathcal{C}} \\
&\quad + n \sum_{q=1}^{\tilde{p}} \Big(K_{q-1} \sum_{i=1}^{\lceil \frac{q-1}{2} \rceil} n^{i-1} \|f(X_0)\|_s^i \|f\|_{\mathcal{C}}^i \Big) \Big(K_{\tilde{p}-q} \sum_{j=1}^{\lceil \frac{\tilde{p}-q}{2} \rceil} n^{j-1} \|f(X_0)\|_s^j \|f\|_{\mathcal{C}}^j \Big) \\
&\leq K' \tilde{p} \|f(X_0)\|_s \|f\|_{\mathcal{C}} + n \sum_{q=1}^{\tilde{p}} K'' \sum_{i=2}^{\lceil \frac{q-1}{2} \rceil + \lceil \frac{\tilde{p}-q}{2} \rceil} n^{i-2} \|f(X_0)\|_s^i \|f\|_{\mathcal{C}}^i \\
&\leq K'' \Big\{ \|f(X_0)\|_s \|f\|_{\mathcal{C}} + n \tilde{p} \sum_{i=2}^{\lceil \tilde{p}/2 \rceil} n^{i-2} \|f(X_0)\|_s^i \|f\|_{\mathcal{C}}^i \Big\} \\
&\leq K_{\tilde{p}} \sum_{i=1}^{\lceil \tilde{p}/2 \rceil} n^{i-1} \|f(X_0)\|_s^i \|f\|_{\mathcal{C}}^i
\end{aligned}$$

for some constants $K', K'', K_{\tilde{p}} < \infty$, since $\lceil (q-1)/2 \rceil + \lceil (\tilde{p}-q)/2 \rceil \leq \lceil \tilde{p}/2 \rceil$. □

By (6.5) and Lemma 6.2 we immediately obtain

$$\mathbf{E}\Big(\Big| \sum_{i=1}^{n} f(X_i) \Big|^{2p} \Big) \leq (2p)! n I_n(2p-1) \leq K_p \sum_{i=1}^{p} n^i \|f(X_0)\|_s^i \|f\|_{\mathcal{C}}^i$$

since (6.2) implies that (6.6) holds with p replaced by $2p-1$, which completed the proof of 6.1. □

Remark 6.2. Recall that sharper moment bounds are available under multiple mixing with exponential decay (cf. Section 2.1). However, the moment bounds given by Proposition 6.1 are sufficient to apply the approximating class approach as we will show in Chapter 7.

The next section is dedicated to a specific class of slowly multiple mixing processes, the class of causal functions of i.i.d. processes.

6.2. Causal Functions of I.I.D. Processes

One example of processes that feature the slow multiple mixing property (6.1) and that can be treated by our methods is the class of causal functions of i.i.d. processes, which are defined as follows.

Definition 6.2 (Causal function)**.** Let $(\xi_j)_{j\in\mathbb{Z}}$ be an independent identically distributed process with values in a Banach space $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$. We call $(X_i)_{i\in\mathbb{N}}$ a causal function of $(\xi_j)_{j\in\mathbb{Z}}$ if there is a measurable function $G : \mathcal{X}^{\mathbb{N}} \to \mathbb{R}^d$ such that each X_i is of the form

$$X_i := G((\xi_{i-j})_{j\in\mathbb{N}}).$$

Let us now introduce a measure of the dependence structure of a causal function of an i.i.d. process $(\xi_j)_{j\in\mathbb{Z}}$. Set

$$\dot{X}_i := G(\xi_i, \xi_{i-1}, \ldots, \xi_1, \xi_0', \xi_{-1}', \ldots),$$

where $(\xi_j')_{j\in\mathbb{Z}}$ is an independent copy of $(\xi_j)_{j\in\mathbb{Z}}$, i.e. $(\xi_j)_{j\in\mathbb{Z}}$ and $(\xi_j')_{j\in\mathbb{Z}}$ are identically distributed, and both processes are independent from each other. We can now define for $i \in \mathbb{N}^*$ and $m \geq 1$,

$$\delta_{i,m} = \|X_i - \dot{X}_i\|_m := \mathbf{E}\big(|X_i - \dot{X}_i|^m\big)^{\frac{1}{m}}, \tag{6.8}$$

where $|\cdot|$ denotes the euclidean norm in $\mathbb{R}^d$. This physical dependence measure was introduced by Dedecker and Prieur (2005) (see also Wu (2005) and Dedecker and Prieur (2007)).

Proposition 6.2. *Let $(X_i)_{i\in\mathbb{N}}$ be an $\mathbb{R}^d$-valued causal function of an i.i.d. process. Then $(X_i)_{i\in\mathbb{N}}$ is slowly multiple mixing w.r.t. $\mathcal{H}_\alpha(\mathbb{R}^d, \mathbb{R})$ with $s \geq 1$ and $\Theta(i) = (\delta_{i,m})^\alpha$ for every $\alpha \in (0, 1]$, and $s \in [1, \infty)$, $m \in (1, \infty]$ with $\frac{1}{s} + \frac{1}{m} = 1$. As a consequence of Proposition 6.1, if*

$$\sum_{i=1}^{\infty} i^{2p-2}(\delta_{i,m})^\alpha < \infty \tag{6.9}$$

for some $p > sd$, then the moment bound (6.3) holds for all $f \in \mathcal{H}_\alpha(\mathbb{R}^d, \mathbb{R})$ such that $\|f\|_\infty \leq 1$, with p, s as above.

Proof. Since $(X_i)_{i\in\mathbb{N}}$ is a causal function of an i.i.d. process, we can write $X_i = G(\xi_i, \xi_{i-1}, \ldots)$, with $G : \mathcal{X}^{\mathbb{N}} \to \mathbb{R}^d$. Let $(\xi_j')_{j\in\mathbb{Z}}$ and $(\xi_j'')_{j\in\mathbb{Z}}$ be copies of the underlying process $(\xi_j)_{j\in\mathbb{Z}}$ such that all three processes are independent. Set

$$\dot{X}_i^{(k)} := G(\xi_i, \xi_{i-1}, \ldots, \xi_{i-k+1}, \xi'_{i-k}, \xi'_{i-k-1}, \ldots),$$
$$\ddot{X}_i^{(k)} := G(\xi_i, \xi_{i-1}, \ldots, \xi_{i-k+1}, \xi''_{i-k}, \xi''_{i-k-1}, \ldots),$$

and note that therefore $(X_i)_{i\in\mathbb{N}} \stackrel{d}{=} (\dot{X}_i^{(k)})_{i\in\mathbb{N}} \stackrel{d}{=} (\ddot{X}_i^{(k)})_{i\in\mathbb{N}}$. We have

$$\begin{aligned}
&\big|\mathbf{Cov}\big(f(X_0)\ldots f(X_{i^*_{q-1}})\,,\, f(X_{i^*_q})\ldots f(X_{i^*_p})\big)\big| \\
&\quad\le \big|\mathbf{Cov}\big(f(X_0)\ldots f(X_{i^*_{q-1}}) - f(\dot{X}_0^{(k)})\ldots f(\dot{X}_{i^*_{q-1}}^{(k)})\,,\, f(X_{i^*_q})\ldots f(X_{i^*_p})\big)\big| \\
&\qquad + \big|\mathbf{Cov}\big(f(\dot{X}_0^{(k)})\ldots f(\dot{X}_{i^*_{q-1}}^{(k)})\,,\, f(X_{i^*_q})\ldots f(X_{i^*_p}) - f(\ddot{X}_{i^*_q}^{(k)})\ldots f(\ddot{X}_{i^*_p}^{(k)})\big)\big| \\
&\qquad + \big|\mathbf{Cov}\big(f(\dot{X}_0^{(k)})\ldots f(\dot{X}_{i^*_{q-1}}^{(k)})\,,\, f(\ddot{X}_{i^*_q}^{(k)})\ldots f(\ddot{X}_{i^*_p}^{(k)})\big)\big|.
\end{aligned}\tag{6.10}$$

Since $f(\dot{X}_0^{(k)})\cdot\ldots\cdot f(\dot{X}_{i^*_{q-1}}^{(k)})$ is $\sigma\big(\{\xi_j : j \le i^*_{q-1}\} \cup \{\xi'_j : j \in \mathbb{Z}\}\big)$-measurable while $f(\ddot{X}_{i^*_q}^{(k)})\cdot\ldots\cdot f(\ddot{X}_{i^*_p}^{(k)})$ is $\sigma\big(\{\xi_j : j > i^*_q - k\} \cup \{\xi''_j : j \in \mathbb{Z}\}\big)$-measurable, the functions in the last covariance on the right-hand side of (6.10) are independent as soon as $k \le i_q$ and thus the last summand is equal to 0 in this case. Recall that we only consider such f that satisfy $\|f\|_\infty \le 1$. If we apply Hölder's inequality to equation (6.10) we obtain for s, m satisfying $\frac{1}{s} + \frac{1}{m} = 1$,

$$\begin{aligned}
&\big|\mathbf{Cov}\big(f(X_0)\cdot\ldots\cdot f(X_{i^*_{q-1}}) - f(\dot{X}_0^{(k)})\cdot\ldots\cdot f(\dot{X}_{i^*_{q-1}}^{(k)})\,,\, f(X_{i^*_q})\cdot\ldots\cdot f(X_{i^*_p})\big)\big| \\
&\quad\le 2\|f(X_0)\cdot\ldots\cdot f(X_{i^*_{q-1}}) - f(\dot{X}_0^{(k)})\cdot\ldots\cdot f(\dot{X}_{i^*_{q-1}}^{(k)})\|_m \|f(X_{i^*_q})\cdot\ldots\cdot f(X_{i^*_p})\|_s \\
&\quad\le 2q\|f(X_0)\|_s\|f(X_0) - f(\dot{X}_0^{(k)})\|_m
\end{aligned}\tag{6.11}$$

where we used that $\Big|\prod_{i=1}^n a_i - \prod_{i=1}^n b_i\Big| \le \sum_{i=1}^n \big|a_i - b_i\big|$. for $a_i, b_i \in [-1,1]$. Since $|f(x) - f(y)| \le \|f\|_\alpha \|x-y\|^\alpha$, an application of Jensen's inequality to (6.11) yields

$$\begin{aligned}
&\big|\mathbf{Cov}\big(f(X_0)\cdot\ldots\cdot f(X_{i^*_{q-1}}) - f(\dot{X}_0^{(k)})\cdot\ldots\cdot f(\dot{X}_{i^*_{q-1}}^{(k)})\,,\, f(X_{i^*_q})\cdot\ldots\cdot f(X_{i^*_p})\big)\big| \\
&\quad\le 2q\|f(X_0)\|_s\|f\|_{\mathcal{H}_\alpha}\big(\|X_0 - \dot{X}_0^{(k)}\|_m\big)^\alpha \\
&\quad= 2q\|f(X_0)\|_s\|f\|_{\mathcal{H}_\alpha}(\delta_{k,m})^\alpha.
\end{aligned}$$

Analogously, we can show that

$$\begin{aligned}
&\big|\mathbf{Cov}\big(f(\dot{X}_0^{(k)})\ldots f(\dot{X}_{i^*_{q-1}}^{(k)})\,,\, f(X_{i^*_q})\ldots f(X_{i^*_p}) - f(\ddot{X}_{i^*_q}^{(k)})\ldots f(\ddot{X}_{i^*_p}^{(k)})\big)\big| \\
&\quad\le 2(p-q)\|f(X_0)\|_s\|f\|_{\mathcal{H}_\alpha}\|(\delta_{k,m})^\alpha,
\end{aligned}$$

thus for $k = i_q$, we have

$$\big|\mathbf{Cov}\big(f(X_0)\ldots f(X_{i^*_{q-1}})\,,\, f(X_{i^*_q})\ldots f(X_{i^*_p})\big)\big| \le 2q\|f(X_0)\|_s\|f\|_{\mathcal{H}_\alpha}(\delta_{i_q,m})^\alpha,$$

which proves the slow multiple mixing property □

7. An Empirical CLT for Slowly Multiple Mixing Processes

In this chapter, we establish an empirical CLT for slowly multiple mixing processes. We are especially interested in the case, where the decay of the covariances is given by a term Θ that converges to zero with sub-exponential rate. We therefore develop a general theorem that can be applied to weaker moment bounds than Theorem 3.1 in Part I. As our main intention here is to extend the application of our approximating class approach (see Section 1.3) to situations where the underlying process only satisfies a weaker version of the multiple mixing property, we restrict our attention to $\mathbb{R}^d$-valued stationary processes $(X_i)_{i\in\mathbb{N}^*}$ and consider the classical empirical processes indexed by the class of indicators of semi-finite rectangles $\{\mathbf{1}_{[-\infty,t]} : t \in \mathbb{R}^d\}$.

For sets such as $\{x \in [-\infty,\infty]^d : a \leq x < b\}$ with $a \leq b \in \mathbb{R}^d$, we write $(a,b]$, where $\leq$, $<$, ... used in $\mathbb{R}^d$ are to be understood component-wise.[1] The *empirical distribution function* $\mathrm{F}_n : [-\infty,\infty]^d \longrightarrow \mathbb{R}$ is given by

$$\mathrm{F}_n(t) := \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{[-\infty,t]}(X_i)$$

Let F denote the (multidimensional) distribution function of $\mathcal{X}_0$. We consider the empirical process $\mathrm{U}_n = (\mathrm{U}_n(t))_{t\in[-\infty,\infty]^d}$ as the random element given by

$$\mathrm{U}_n(t) := \sqrt{n}\big(\mathrm{F}_n(t) - \mathrm{F}(t)\big), \quad t \in [-\infty,\infty]^d.$$

which takes valued in the càdlàg space $\mathbb{D}([\infty,\infty]^d)$ equipped with the Skorokhod metric d_S.[2]

Before coming to the statement of our main results, let us have a look at our central assumptions. Similar as in Chapter 3, we assume that $(X_i)_{i\in\mathbb{N}}$ satisfies a Central Limit Theorem under a class of functions $\mathcal{C}$.

Assumption 7.I (CLT for $\mathcal{C}$-Observables.)**.** For every $f \in \mathcal{C}$ such that $\mathbf{E}(f(X_0)) = 0$, there exists $\sigma_f^2 \geq 0$ such that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f(X_i) \xrightarrow{d} N(0,\sigma_f^2). \tag{7.1}$$

Further, we assume that the following generalized moment bounds hold under the same function space $\mathcal{C}$ for which the CLT (7.1) is satisfied.

[1] i.e., for $a = (a_1,\ldots,a_d)$ and $b = (b_1,\ldots,b_d) \in [-\infty,\infty]^d$, write $a \leq b$ if and only if $a_i \leq b_i$ for all $i = 1,\ldots,d$.
[2] For more details, see page 8.

Assumption 7.II (Moment Bounds for $\mathcal{C}$-Observables)**.** There are finite constants $C > 0$, $s \geq 1$, $p \in \mathbb{N}^*$ and increasing functions $\Phi_1, \dots, \Phi_p : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ such that for all $f \in \mathcal{C}$ with $\|f\|_\infty \leq 1$ and all $n \in \mathbb{N}^*$, we have

$$\mathbf{E}\Big(\sum_{i=1}^{n}\big(f(X_i) - \mathbf{E}(f(X_0))\big)\Big)^{2p} \\ \leq C \sum_{i=1}^{p} n^i \, \|f(X_0) - \mathbf{E}(f(X_0))\|_s^i \, \Phi_i(\|f - \mathbf{E}(f(X_0))\|_{\mathcal{C}}). \tag{7.2}$$

Recall that this condition is met for instance for processes satisfying a slow multiple mixing property with $\Phi(x) = x^i$ (see Proposition 6.1).

Since in this setting, the indexing class of the empirical process is a fixed class of semi-finite rectangles, we do not apply our bracketing technique in full generality here. Instead the bracketing will be directly included in the proof of property (1.B) of our approximating class approach.

Control of the $\|\cdot\|_{\mathcal{C}}$-Size of the Approximating Functions

Assumption 7.I and Assumption 7.II refer to the processes $(f(X_i))_{i\in\mathbb{N}}$ for $f \in \mathcal{C}$. In order to obtain results for the empirical process U_n indexed by indicator functions of semi-finite rectangles, we use approximations by functions of the space $\mathcal{C}$. Here, we use a function that provides a control over the $\|\cdot\|_{\mathcal{C}}$-size of the approximating functions. This function replaces the entropy condition used in Part I. It corresponds in a way to the second argument in the bracketing numbers in that section and can be seen as an adapted version of the bracketing technique used there.

Definition 7.1 ($\|\cdot\|_{\mathcal{C}}$-Control)**.** Let F be a (multidimensional) distribution function on $\mathbb{R}^d$ and let $\mathcal{C}$ be some vector space of $\mathbb{R}$-valued measurable functions on $\mathbb{R}^d$, equipped with a semi-norm $\|\cdot\|_{\mathcal{C}}$. We call a increasing function $\Psi : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ a $\|\cdot\|_{\mathcal{C}}$*-control* (with respect to F) if for every $a < b \in [-\infty, \infty]^d$, there is a function $\varphi_{(a,b)} \in \mathcal{C}$ such that for any $x \in \mathbb{R}^d$,

$$\mathbf{1}_{(-\infty,a]} \leq \varphi_{(a,b]} \leq \mathbf{1}_{(-\infty,b]} \,. \tag{7.3}$$

and such that

$$\|\varphi_{(a,b)}\|_{\mathcal{C}} \leq \Psi\Big(\frac{1}{\min_{i=1,\dots,d} \omega_{\mathrm{F}_{(i)}}(b_i - a_i)}\Big), \tag{7.4}$$

where $\mathrm{F}_{(i)}$ denotes the i-th marginal distribution function of X and where $\omega_{\mathrm{F}_{(i)}}$ is the modulus of continuity of $\mathrm{F}_{(i)}$.[3]

If such a function Ψ exists, we say that $\mathcal{C}$ *approximates the indicator functions of semi-finite rectangles* $\mathcal{R} := \{[-\infty, t] : t \in \mathbb{R}^d\}$ with $\|\cdot\|_{\mathcal{C}}$-control Ψ.

[3]Recall that the modulus of continuity ω_g of a real-valued function g is defined by $\omega_g(\delta) := \sup\{|g(t) - g(s)| : s, t \in \mathbb{R}^d, \ \|t - s\| \leq \delta\}$.

Example 7.1. As an example, consider the space of bounded α-Hölder functions $\mathcal{H}_\alpha(\mathbb{R}^d, \mathbb{R})$ introduced in Section 2.3. Choose $\varphi_{(a,b)} \in \mathcal{H}_\alpha(\mathbb{R}^d, \mathbb{R})$ as

$$\varphi_{(a,b)}(x_1, \ldots, x_d) := \prod_{i=1}^{d} \varphi\Big(\mathbf{1}_{(-\infty,\infty)^2}(a_i, b_i) \cdot \frac{x_i - b_i}{b_i - a_i}\Big),$$

where $\varphi : \mathbb{R} \to \mathbb{R}$ is given by $\varphi(x) = \mathbf{1}_{[-\infty,-1]}(x) - x\,\mathbf{1}_{(-1,0]}(x)$. Obviously, this choice of $\varphi_{(a,b)}$ satisfies (7.3). Let us now check condition (7.4). Since for all $j = 1, \ldots, d$, $\omega_{\mathrm{F}_{(j)}}(\delta) \le \omega_{\mathrm{F}}(\delta)$, we have

$$b_j - a_j \ge \inf\big\{\delta > 0 : \omega_{\mathrm{F}}(\delta) \ge \min_{i=1,\ldots,d} \omega_{\mathrm{F}_{(i)}}(b_i - a_i)\big\}.$$

Thus, by the definition of $\varphi_{(a,b)}$ we obtain

$$\begin{aligned} \|\varphi_{(a,b)}\|_{\mathcal{H}_\alpha} &\le d \max_{j=1,\ldots,d} \mathbf{1}_{(-\infty,\infty)^2}(a_j, b_j) \cdot \frac{1}{(b_j - a_j)^\alpha} + 1 \\ &\le d\Big(\omega_{\mathrm{F}}^{\leftarrow}\Big(\min_{i=1,\ldots,d} \omega_{\mathrm{F}_{(i)}}(b_i - a_i)\Big)\Big)^{-\alpha} + 1, \end{aligned}$$

where $\omega_{\mathrm{F}}^{\leftarrow}(y) := \inf\big\{\delta > 0 : \omega_{\mathrm{F}}(\delta) \ge y\big\}$.Hence, (7.4) is satisfied for the increasing function Ψ given by

$$\Psi(z) := d\big(\omega_{\mathrm{F}}^{\leftarrow}(z^{-1})\big)^{-\alpha} + 1 \tag{7.5}$$

and thus Ψ defines an $\|\cdot\|_{\mathcal{H}_\alpha}$-control, which gives us the following lemma:

Lemma 7.1. *The space of bounded α-Hölder functions $\mathcal{H}_\alpha(\mathbb{R}^d, \mathbb{R})$ approximates the indicator functions of semi-finite rectangles $\mathcal{R}$ with $\|\cdot\|_{\mathcal{H}_\alpha}$-control Ψ w.r.t.* F *given by* (7.5)*.*

Under these assumptions, we can establish the following theorems.

7.1. Statement of Results

Theorem 7.1. *Let $(X_i)_{i\in\mathbb{N}}$ be a stationary process of $\mathbb{R}^d$-valued random vectors with continuous multidimensional distribution function* F*. Assume that there is a vector space $\mathcal{C}$ of measurable functions $\mathbb{R}^d \to \mathbb{R}$, containing the constant functions, equipped with a semi-norm $\|\cdot\|_{\mathcal{C}}$, and satisfying the following conditions:*

(i) For every $f \in \mathcal{C}$ such that $\|f\|_\infty < \infty$, the CLT (7.1) *holds.*

(ii) $\mathcal{C}$ approximates the indicator functions of semi-finite rectangles $\mathcal{R}$ with $\|\cdot\|_{\mathcal{C}}$-control Ψ w.r.t. F*.*

(iii) There are constants $s \ge 1$, $p > sd$, $\gamma_1, \ldots, \gamma_p \in \mathbb{R}$ satisfying

$$0 \le \gamma_i < \frac{i}{s} + 2(p - i) - d \quad \textit{for all } i = 1, \ldots, p \tag{7.6}$$

and some increasing functions $\Phi_1, \dots, \Phi_p : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ *satisfying*

$$\Phi_i(2\Psi(z)) = \mathcal{O}(z^{\gamma_i}) \quad \text{as } z \to \infty \tag{7.7}$$

such that for every $f \in \mathcal{C}$ *with* $\|f\|_\infty \le 1$, *the moment bound* (7.2) *holds.*

Then there is a centred Gaussian process $W = (W(t))_{t \in [-\infty,\infty]^d}$ *with almost surely continuous sample paths such that* $\mathrm{U}_n \xrightarrow{d} W$ *in the space* $\mathbb{D}([-\infty,\infty]^d)$.

Remark 7.1. (i) In fact, the functions $\Phi_1, \dots, \Phi_p$ and the function Ψ only have to be increasing for sufficiently large arguments. Note that condition (7.2) has only to be satisfied for a certain subclass of $\mathcal{C}$, see Remark 7.5.

(ii) Although, the focus of this chapter is to consider processes with a sub-exponential decay, the assumptions of Theorem 7.1 are quite general and can also be applied in the exponential case. For instance, choosing $\mathcal{C} = \mathcal{H}_\alpha(\mathbb{R}^d, \mathbb{R})$ and $\gamma_i = 2p - i$ we obtain Theorem 3 in Dehling and Durieu (2011), setting further $\alpha = d = s = 1$ covers Theorem 1 in Dehling et al. (2009).

As a consequence of this abstract theorem, we can give a statement for slowly multiple mixing processes, for which conditions are more easily verifiable.

Theorem 7.2 (Empirical CLT for Slowly Multiple Mixing Data)**.** *Let* $(X_i)_{i \in \mathbb{N}}$ *be a stationary* $\mathbb{R}^d$*-valued process with continuous multidimensional distribution function* F. *Assume there is a vector space* $\mathcal{C}$ *of measurable functions* $\mathbb{R}^d \to \mathbb{R}$, *containing the constant functions, equipped with a semi-norm* $\|\cdot\|_{\mathcal{C}}$ *that satisfies the following conditions:*

(i) For every $f \in \mathcal{C}$ *such that* $\|f\|_\infty < \infty$ *the CLT* (7.1) *holds.*

(ii) The process $(X_i)_{i \in \mathbb{N}^*}$ *is slowly multiple mixing w.r.t. to* $\mathcal{C}$ *for some* $s \ge 1$ *and* $\Theta : \mathbb{N} \longrightarrow \mathbb{R}_0^+$ *such that there exists a* $p > sd$ *satisfying* $\sum_{i=0}^{\infty} i^{2p-2}\Theta(i) < \infty$.

(iii) $\mathcal{C}$ *approximates the indicator functions of semi-finite rectangles* $\mathcal{R}$ *with* $\|\cdot\|_{\mathcal{C}}$*-control* Ψ *w.r.t.* F *such that* $\Psi(z) = \mathcal{O}(z^{1/\gamma})$ *for some* $\gamma > \frac{sp}{p-sd}$.

Then there is a centred Gaussian process $W = (W(t))_{t \in [-\infty,\infty]^d}$ *with almost surely continuous sample paths such that* $\mathrm{U}_n \xrightarrow{d} W$ *in the space* $\mathbb{D}([-\infty,\infty]^d)$.

Proof. By Proposition 6.1, condition Assumption 7.II holds with $\Phi_i(x) = x^i$. Then, taking $\gamma_i = i/\gamma$, the condition (iii) of Theorem 7.1 is satisfied. □

Remark 7.2. In the situation, where $\mathcal{C}$ is the space $\mathcal{H}_\alpha(\mathbb{R}^d, \mathbb{R})$ with $\alpha \in (0, 1]$, by Lemma 7.1, (iii) of Theorem 7.2 can be replaced by

$$\left(\omega_{\mathrm{F}}^{\leftarrow}\left(\frac{1}{z}\right)\right)^{\alpha} = \mathcal{O}\left(z^{\frac{1}{\gamma}}\right) \quad \text{as } z \to \infty, \text{ for some } \gamma > \frac{sp}{p - sd}.$$

This is certainly satisfied (for instance for $\gamma = \theta/\alpha$) if F is θ-Hölder with

$$\theta > \frac{\alpha sp}{p - sd}. \tag{7.8}$$

7.2. Proof of Theorem 7.1

To prove Theorem 7.1, we apply Theorem 1.1 where we choose $S = \mathbb{D}([-\infty,\infty]^d)$, the càdàg space on $[-\infty,\infty]^d$ equipped with the Skorokhod metric d_S and $\xi_n = \mathrm{U}_n$ denotes the empirical process. We therefore need to find a process $U_n^{(q)}$ which approximates U_n as $q \to \infty$ in the sense of (1.11) and to show that this process is convergent in distribution for each q as $n \to \infty$.

Remark 7.3. Note that in this situation one can also apply the version of Theorem 1.1 used by Dehling et al. (2009), since the relevant processes are measurable and S is separable.

Following the techniques presented in Dehling and Durieu (2011, p.1078 ff), we begin by introducing a partition for $[-\infty,\infty]^d$. Let $\mathrm{F}_{(i)}$ be the i-th marginal distribution of F, $0 = r_0^{(q)} < r_1^{(q)} < \ldots < r_q^{(q)} = 1$ a partition of $[0,1]$, and set for $i \in \{1,\ldots,d\}$ and $j_i \in \{0,\ldots,q\}$,

$$t_{i,j_i}^{(q)} := \mathrm{F}_{(i)}^{-1}(r_{j_i}^{(q)}),$$

where $\mathrm{F}_{(i)}^{-1}(y) := \sup\{x \in [-\infty,\infty] : \mathrm{F}_{(i)}(x) \le y\}$. Note that the $\mathrm{F}_{(i)}^{-1}$ are injective since the $\mathrm{F}_{(i)}$ are continuous. For convenience, we also define $t_{i,q+1}^{(q)} := t_{i,q}^{(q)}$. For $j \in \{0,\ldots,q+1\}^d$, set

$$t_j^{(q)} := (t_{1,j_1}^{(q)},\ldots,t_{d,j_d}^{(q)}) = (\mathrm{F}_{(1)}^{-1}(r_{j_1}^{(q)}),\ldots,\mathrm{F}_{(d)}^{-1}(r_{j_d}^{(q)})).$$

To keep notation short, denote $(x,\ldots,x) \in [-\infty,\infty]^d$ by $\overline{x}$.

We can construct a $\mathcal{C}$-approximation of the indicator function $\mathbf{1}_{[-\overline{\infty},t_{j-\overline{1}}^{(q)}]}$ by setting for $j \in \{1,\ldots,q\}^d$,

$$\varphi_j^{(q)} := \begin{cases} \varphi_{\left(t_{j-\overline{2}}^{(q)},t_{j-\overline{1}}^{(q)}\right)} & \text{if } j \ge \overline{2}, \\ 0 & \text{if } j \not\ge \overline{2}, \end{cases} \tag{7.9}$$

where $\varphi_{(t_{j-\overline{2}}^{(q)},t_{j-\overline{1}}^{(q)})} \in \mathcal{C}$ satisfies (7.3) and (7.4). Observe that $t_{j-\overline{2}}^{(q)} < t_{j-\overline{1}}^{(q)}$, since all $\mathrm{F}_{(i)}^{-1}$ are injective.

To approximate the empirical distribution function, we introduce

$$F_n^{(q)}(t) := \sum_{j\in\{1,\ldots,q\}^d} \left(\frac{1}{n}\sum_{i=1}^n \varphi_j^{(q)}(X_i)\right) \mathbf{1}_{[t_{j-\overline{1}}^{(q)},t_j^{(q)})}(t).$$

Note that for t in any fixed rectangle $[t_{j-\overline{1}}^{(q)}, t_j^{(q)})$ we have the simple form

$$F_n^{(q)}(t) = \frac{1}{n}\sum_{i=1}^n \varphi_j^{(q)}(X_i).$$

By the definition of the $\varphi_j^{(q)}$ it is easy to see that therefore

$$\mathrm{F}_n(t_{j-\overline{2}}^{(q)}) \le F_n^{(q)}(t) \le \mathrm{F}_n(t_{j-\overline{1}}^{(q)}) \quad \forall t \in [t_{j-\overline{1}}^{(q)}, t_j^{(q)}).$$

Thus, it is natural to approximate (as $q \to \infty$) U_n by

$$U_n^{(q)} := \Big(\sqrt{n}\big(F_n^{(q)}(t) - F^{(q)}(t)\big)\Big)_{t\in[-\infty,\infty]^d},$$

where

$$F^{(q)}(t) := \mathbf{E}\big(F_n^{(q)}(t)\big) = \sum_{j\in\{1,\dots,q\}^d} \mathbf{E}\big(\varphi_j^{(q)}(X_0)\big)\, \mathbf{1}_{[t^{(q)}_{j-\bar{1}}, t^{(q)}_j)}(t).$$

Remark 7.4. Notice that at this point the $\varphi_j^{(q)}$ and thus $U_n^{(q)}$ depend heavily on the chosen partition $r_0, \dots, r_q$ on $[0,1]$. Therefore the notation with the superscript q may be misleading at first glance, but since whenever the choice of the partition matters, we will only use equidistant partitions of $[0,1]$, and thus in all relevant situations the partitions will be uniquely defined by q.

As the central idea to prove Theorem 7.1 is to use Theorem 1.1, we need to check (1.10) and (1.11) for $S = \mathbb{D}([-\infty,\infty]^d)$. This is done in the next two lemmas.

Lemma 7.2. *For every partition $0 = r_0^{(q)} < \dots < r_q^{(q)} = 1$ of $[0,1]$, $U_n^{(q)}$ converges weakly to some centred Gaussian process $W^{(q)} \in \mathbb{D}([-\infty,\infty]^d)$ whose sample paths are constant on each of the rectangles $[t^{(q)}_{j-\bar{1}}, t^{(q)}_j)$, $j \in \{1,\dots,q\}^d$.*

Proof. Since all the $U_n^{(q)}$ are constant on each of the rectangles $[t^{(q)}_{j-\bar{1}}, t^{(q)}_j)$, it suffices to show weak convergence of the sequence of vectors

$$\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\varphi_j^{(q)}(X_i) - \mathbf{E}\left(\varphi_j^{(q)}(X_i)\right)\right)\right)_{j\in\{1,\dots,q\}^d},$$

which is a consequence of the CLT (7.1) and the Cramér–Wold device. □

Lemma 7.3. *Let $0 = r_0^{(q)} < r_1^{(q)} < \dots < r_q^{(q)} = 1$ be the partition of $[0,1]$ defined by $r_k^{(q)} = \frac{k}{q}$. Then for all $\varepsilon, \eta > 0$ there is a $q_0 \in \mathbb{N}^*$ such that for all $q \geq q_0$,*

$$\limsup_{n\to\infty} \mathbf{P}\left(\sup_{t\in[-\infty,\infty]^d} |\,\mathrm{U}_n(t) - U_n^{(q)}(t)| > \varepsilon\right) \leq \eta.$$

Proof. Let us consider $\varepsilon, \eta > 0$ fixed for the rest of this proof. Consider the partition $0 = r_0^{(q)} < \dots < r_q^{(q)} = 1$ of $[0,1]$ defined in the statement of the lemma and set $h = \frac{1}{q}$. For each $k \in \mathbb{N}^*$, consider the refined partition

$$r^{(q)}_{m-1} = s^{(k)}_{m,0} < s^{(k)}_{m,1} < \dots < s^{(k)}_{m,2^k} = r^{(q)}_m$$

of $[r^{(q)}_{m-1}, r^{(q)}_m]$, where

$$s^{(k)}_{m,\ell} := r^{(q)}_{m-1} + \ell \cdot \frac{h}{2^k},$$

$\ell \in \{0,\ldots,2^k\}$ and $m \in \{0,\ldots,q\}$. Setting for $i \in \{1,\ldots,d\}$, $j_i \in \{1,\ldots,q\}$, $l_i \in \{0,\ldots,2^k\}$

$$s^{(k)}_{i,j_i,l_i} = \mathrm{F}^{-1}_{(i)}(s^{(k)}_{j_i,l_i}),$$

we obtain partitions

$$t^{(q)}_{i,j_i-1} = s^{(k)}_{i,j_i,0} < s^{(k)}_{i,j_i,1} < \ldots < s^{(k)}_{i,j_i,2^k} = t^{(q)}_{i,j_i}$$

of $[t^{(q)}_{i,j_i-1}, t^{(q)}_{i,j_i}]$. To simplify the notation in the following calculations, we set

$$s^{(k)}_{i,j_i,-1} := s^{(k)}_{i,j_i-1,2^k-1} \text{ for } j_i > 1, \text{ and } s^{(k)}_{i,j_i,2^k+1} := s^{(k)}_{i,j_i+1,1}, \text{ for } j_i < q.$$

Let us now focus on a fixed rectangle $[t^{(q)}_{j-\bar{1}}, t^{(q)}_{j})$ for some $j = (j_1,\ldots,j_d) \in \{1,\ldots,q\}^d$. Our aim is to construct a chain to link the point $t^{(q)}_{j-\bar{1}}$ to some arbitrary point $t \in [t^{(q)}_{j-\bar{1}}, t^{(q)}_{j})$. Therefore, we set

$$l_{i,j_i}(k,t) = \max\bigl\{\ell \in \{0,\ldots,2^k\} : s^{(k)}_{i,j_i,\ell} \le t_i\bigr\} \in \{0,\ldots,2^k-1\}.$$

Since we consider j to be fixed, we may drop the index j in order to simplify further notation. More precisely, we set

$$s^{(k)}_{l} := (s^{(k)}_{1,j_1,l_1},\ldots,s^{(k)}_{d,j_d,l_d}) \text{ and } l(k,t) := (l_{1,j_1}(k,t),\ldots,l_{d,j_d}(k,t)).$$

In this way for any $k \in \mathbb{N}^*$, we obtain an ($[-\infty,\infty]^d$- valued) chain

$$t^{(q)}_{j-\bar{1}} = s^{(0)}_{l(0,t)} \le s^{(1)}_{l(1,t)} \le \ldots \le s^{(k)}_{l(k,t)} \le t \le s^{(k)}_{l(k,t)+\bar{1}}.$$

Now set $\psi^{(0)}_{\bar{0}} := \varphi_{(t^{(q)}_{j-\bar{1}}, t^{(q)}_{j})}$ and choose for every $k \in \mathbb{N}^*$ and $l \in \{0,\ldots,2^k+1\}^d$, a function $\psi^{(k)}_{l} \in \mathcal{C}$ such that[4]

$$\psi^{(k)}_{l} = \begin{cases} 0 & \text{if } \exists i \in \{1,\ldots,d\} : j_i = 1 \text{ and } l_i = 0, \\ 1 & \text{if } (\exists i \in \{1,\ldots,d\} : j_i = q \text{ and } l_i = 2^k+1) \\ & \text{and } (\nexists i \in \{1,\ldots,d\} : j_i = 1 \text{ and } l_i = 0), \\ \varphi_{(s^{(k)}_{l-\bar{1}}, s^{(k)}_{l})} & \text{else,} \end{cases} \tag{7.10}$$

where $\varphi_{(s^{(k)}_{l-\bar{1}}, s^{(k)}_{l})}$ satisfies (7.3) and (7.4). By this definition we have for every $t \in [-\infty,\infty]^d$ and $l \in \{0,\ldots,2^k\}^d$, the following inequalities:

$$\mathbf{1}_{[-\overline{\infty}, s_{l-\bar{1}}]} \le \psi^{(k)}_{l} \le \mathbf{1}_{[-\overline{\infty}, s_l]}, \tag{7.11}$$

[4]the reference to the indices j and q is omitted, since these are considered to be fixed.

$$\varphi_j^{(q)} \leq \psi_{l(1,t)}^{(1)} \leq \ldots \leq \psi_{l(k,t)}^{(k)} \leq \mathbf{1}_{[-\overline{\infty},t]} \leq \psi_{l(k,t)+\overline{2}}^{(k)}. \tag{7.12}$$

Using inequality (7.12), we obtain for $t \in [t_{j-\overline{1}}^{(q)}, t_j^{(q)})$ and $K \in \mathbb{N}^*$, the telescopic-sum representation

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\Big(\mathbf{1}_{[-\overline{\infty},t]}(X_i) - F_n^{(q)}(t)\Big) \\
&\quad = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{[-\overline{\infty},t]}(X_i) - \frac{1}{n}\sum_{i=1}^{n}\varphi_j^{(q)}(X_i) \\
&\quad = \sum_{k=1}^{K}\frac{1}{n}\sum_{i=1}^{n}\Big(\psi_{l(k,t)}^{(k)}(X_i) - \psi_{l(k-1,t)}^{(k-1)}(X_i)\Big) + \frac{1}{n}\sum_{i=1}^{n}\Big(\mathbf{1}_{[-\overline{\infty},t]}(X_i) - \psi_{l(K,t)}^{(K)}(X_i)\Big).
\end{aligned} \tag{7.13}$$

Let us now consider

$$\mathrm{U}_n(t) - U_n^{(q)}(t) = \sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{[-\overline{\infty},t]}(X_i) - \mathrm{F}(t)\Big) - \sqrt{n}\Big(F_n^{(q)}(t) - F^{(q)}(t)\Big).$$

Equation (7.13) yields

$$\begin{aligned}
\mathrm{U}_n(t) - U_n^{(q)}(t) = &\sum_{k=1}^{K}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big(\psi_{l(k,t)}^{(k)}(X_i) - \mathbf{E}\,\psi_{l(k,t)}^{(k)}(X_0)\big) - \big(\psi_{l(k-1,t)}^{(k-1)}(X_i) - \mathbf{E}\,\psi_{l(k-1,t)}^{(k-1)}(X_0)\big) \\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big(\mathbf{1}_{[-\overline{\infty},t]}(X_i) - \mathrm{F}(t)\big) - \big(\psi_{l(K,t)}^{(K)}(X_i) - \mathbf{E}\,\psi_{l(K,t)}^{(K)}(X_0)\big).
\end{aligned} \tag{7.14}$$

Applying the inequalities in (7.12), we gain the following upper bounds for the last sum on the right-hand side of the above inequality. For every $K \in \mathbb{N}^*$

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big(\big(\mathbf{1}_{[-\overline{\infty},t]}(X_i) - \mathrm{F}(t)\big) - \big(\psi_{l(K,t)}^{(K)}(X_i) - \mathbf{E}\,\psi_{l(K,t)}^{(K)}(X_0)\big)\big) \\
&\quad \geq -\sqrt{n}\Big(\mathrm{F}(t) - \mathbf{E}\,\psi_{l(K,t)}^{(K)}(X_0)\Big) \geq -\sqrt{n}\Big(\mathbf{E}\,\psi_{l(K,t)+\overline{2}}^{(K)}(X_0) - \mathbf{E}\,\psi_{l(K,t)}^{(K)}(X_0)\Big)
\end{aligned} \tag{7.15}$$

and

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(\big(\mathbf{1}_{[-\overline{\infty},t]}(X_i) - \mathrm{F}(t)\big) - \big(\psi_{l(K,t)}^{(K)}(X_i) - \mathbf{E}\,\psi_{l(K,t)}^{(K)}(X_0)\big)\Big) \\
&\quad \leq \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(\big(\psi_{l(K,t)+\overline{2}}^{(K)}(X_i) - \mathbf{E}\,\psi_{l(K,t)+\overline{2}}^{(K)}(X_0)\big) - \big(\psi_{l(K,t)}^{(K)}(X_i) - \mathbf{E}\,\psi_{l(K,t)}^{(K)}(X_0)\big) \\
&\qquad + \sqrt{n}\Big(\mathbf{E}\,\psi_{l(K,t)+\overline{2}}^{(K)}(X_0) - \mathrm{F}(t)\Big) \\
&\quad \leq \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(\big(\psi_{l(K,t)+\overline{2}}^{(K)}(X_i) - \mathbf{E}\,\psi_{l(K,t)+\overline{2}}^{(K)}(X_0)\big) - \big(\psi_{l(K,t)}^{(K)}(X_i) - \mathbf{E}\,\psi_{l(K,t)}^{(K)}(X_0)\big) \\
&\qquad + \sqrt{n}\Big(\mathbf{E}\,\psi_{l(K,t)+\overline{2}}^{(K)}(X_0) - \mathbf{E}\,\psi_{l(K,t)}^{(K)}(X_0)\Big).
\end{aligned} \tag{7.16}$$

For convenience, let $s^{(k)}_{i,q,2^k+1} := s^{(k)}_{i,q,2^k}$. Using equation (7.11) and the continuity of F,[5] we obtain

$$\begin{aligned}
&\sqrt{n}\Big|\mathbf{E}\big(\psi^{(K)}_{l(K,t)+\bar{2}}(X_0)\big) - \big(\mathbf{E}\,\psi^{(K)}_{l(K,t)}(X_0)\big)\Big| \\
&\quad\le \sqrt{n}\Big|\mathbf{E}\big(\mathbf{1}_{[-\overline{\infty},s_{l(K,t)+\bar{2}})}(X_0)\big) - \mathbf{E}\big(\mathbf{1}_{[-\overline{\infty},s_{l(K,t)-\bar{1}})}(X_0)\big)\Big| \\
&\quad= \sqrt{n}\Big(\mathrm{F}(s_{l(K,t)+\bar{2}}) - \mathrm{F}(s_{l(K,t)-\bar{1}})\Big) \\
&\quad\le \sqrt{n}\Big(d \max_{i=1,\dots,d}\big\{\mathrm{F}_{(i)}(s_{i,j_i,l_{i,j_i}(K,t)+\bar{2}}) - \mathrm{F}_{(i)}(s_{i,j_i,l_{i,j_i}(K,t)-\bar{1}})\big\}\Big) = \frac{3d\sqrt{n}h}{2^K},
\end{aligned}$$

and thus, if we choose

$$K = K_n := \left\lceil \log_2\Big(\frac{2^4 d}{\varepsilon}\sqrt{n}h\Big)\right\rceil, \tag{7.17}$$

we obtain

$$\Big|\mathbf{E}\,\psi^{(K)}_{l(K,t)+\bar{2}}(X_0) - \mathbf{E}\,\psi^{(K)}_{l(K,t)}(X_0)\Big| < \frac{\varepsilon}{2}. \tag{7.18}$$

In summary, using (7.15), (7.16) and (7.18) in equation (7.14) yields, for all $n \in \mathbb{N}^*$,

$$\begin{aligned}
&\big|\mathrm{U}_n(t) - U^{(q)}_n(t)\big| \\
&\quad\le \Bigg|\sum_{k=1}^{K_n}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(\psi^{(k)}_{l(k,t)}(X_i) - \mathbf{E}\,\psi^{(k)}_{l(k,t)}(X_0)\Big) - \Big(\psi^{(k-1)}_{l(k-1,t)}(X_i) - \mathbf{E}\,\psi^{(k-1)}_{l(k-1,t)}(X_0)\Big)\Bigg| \\
&\qquad+ \Bigg|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(\psi^{(K_n)}_{l(K_n,t)+\bar{2}}(X_i) - \mathbf{E}\,\psi^{(K_n)}_{l(K_n,t)+\bar{2}}(X_0)\Big) - \Big(\psi^{(K_n)}_{l(K_n,t)}(X_i) - \mathbf{E}\,\psi^{(K_n)}_{l(K_n,t)}(X_0)\Big)\Bigg| + \frac{\varepsilon}{2}.
\end{aligned} \tag{7.19}$$

Now, consider the maximum of the terms in (7.19) over all $t \in [t^{(q)}_{j-\bar{1}}, t^{(q)}_j)$. By the definition of the $l(k,t)$ we have

$$\left\lfloor\frac{l(k,t)}{2}\right\rfloor := \left(\left\lfloor\frac{l_{1,j_1}(k,t)}{2}\right\rfloor,\dots,\left\lfloor\frac{l_{d,j_d}(k,t)}{2}\right\rfloor\right) = l(k-1,t).$$

We therefore obtain

$$\begin{aligned}
&\sup_{t\in[t^{(q)}_{j-\bar{1}},t^{(q)}_j)} |\,\mathrm{U}_n(t) - U^{(q)}_n(t)| \\
&\le \sum_{k=1}^{K_n}\frac{1}{\sqrt{n}}\max_{l\in\{0,\dots,2^k-1\}^d}\Bigg|\sum_{i=1}^{n}\Big(\psi^{(k)}_l(X_i) - \mathbf{E}\,\psi^{(k)}_l(X_0)\Big) - \Big(\psi^{(k-1)}_{\lfloor l/2\rfloor}(X_i) - \mathbf{E}\,\psi^{(k-1)}_{\lfloor l/2\rfloor}(X_0)\Big)\Bigg| \\
&\quad+ \frac{1}{\sqrt{n}}\max_{l\in\{0,\dots,2^{K_n}-1\}^d}\Bigg|\sum_{i=1}^{n}\Big(\psi^{(K_n)}_{l+\bar{2}}(X_i) - \mathbf{E}\,\psi^{(K_n)}_{l+\bar{2}}(X_0)\Big) - \Big(\psi^{(K_n)}_l(X_i) - \mathbf{E}\,\psi^{(K_n)}_l(X_0)\Big)\Bigg| + \frac{\varepsilon}{2}.
\end{aligned}$$

[5] note that for continuous F, we have $\mathrm{F}\circ\mathrm{F}^{-1}(x) = x$ for all $x \in [0,1]$.

Choose $\varepsilon_k = \frac{\varepsilon}{4k(k+1)}$ and note that $\sum_{k=1}^{\infty} \varepsilon_k = \varepsilon/4$. An application of Markov's inequality for the $2p$-th moments combined with condition (7.2) implies

$$\begin{aligned}
&\mathbf{P}\Big(\sup_{t\in[t_{j-1}^{(q)},t_j^{(q)})} |\,\mathrm{U}_n(t) - U_n^{(q)}(t)| \geq \varepsilon\Big) \\
&\leq \Bigg\{\sum_{k=1}^{K_n} \sum_{l\in\{0,\ldots,2^k-1\}^d} \mathbf{P}\Big(\frac{1}{\sqrt{n}}\Big|\sum_{i=1}^{n}\big(\psi_l^{(k)}(X_i) - \mathbf{E}\,\psi_l^{(k)}(X_0)\big) \\
&\qquad\qquad - \big(\psi_{[l/2]}^{(k-1)}(X_i) - \mathbf{E}\,\psi_{[l/2]}^{(k-1)}(X_0)\big)\Big| > \varepsilon_k\Big)\Bigg\} \\
&\quad + \sum_{l\in\{0,\ldots,2^{K_n}-1\}^d} \mathbf{P}\Big(\frac{1}{\sqrt{n}}\Big|\sum_{i=1}^{n}\big(\psi_{l+\overline{2}}^{(K_n)}(X_i) - \mathbf{E}\,\psi_{l+\overline{2}}^{(K_n)}(X_0)\big) \\
&\qquad\qquad - \big(\psi_l^{(K_n)}(X_i) - \mathbf{E}\,\psi_l^{(K_n)}(X_0)\big)\Big| > \frac{\varepsilon}{4}\Big) \\
&\leq \Bigg\{\sum_{k=1}^{K_n} \sum_{l\in\{0,\ldots,2^k-1\}^d} \frac{1}{\varepsilon_k^{2p} n^p}\,\mathbf{E}\Big(\Big|\sum_{i=1}^{n}\big(\psi_l^{(k)}(X_i) - \mathbf{E}\,\psi_l^{(k)}(X_0)\big) \\
&\qquad\qquad - \big(\psi_{[l/2]}^{(k-1)}(X_i) - \mathbf{E}\,\psi_{[l/2]}^{(k-1)}(X_0)\big)\Big|^{2p}\Big)\Bigg\} \\
&\quad + \sum_{l\in\{0,\ldots,2^{K_n}-1\}^d} \frac{4^{2p}}{n^p\varepsilon^{2p}}\,\mathbf{E}\Big(\Big|\sum_{i=1}^{n}\big(\psi_{l+\overline{2}}^{(K_n)}(X_i) - \mathbf{E}\,\psi_{l+\overline{2}}^{(K_n)}(X_0)\big) - \big(\psi_l^{(K_n)}(X_i) - \mathbf{E}\,\psi_l^{(K_n)}(X_0)\big)\Big|^{2p}\Big) \\
&\leq 2C\Bigg\{\sum_{k=1}^{K_n} \sum_{l\in\{0,\ldots,2^k-1\}^d} \frac{1}{\varepsilon_k^{2p} n^p} \sum_{i=1}^{p} n^i\, \|\psi_l^{(k)}(X_0) - \psi_{[l/2]}^{(k-1)}(X_0)\|_s^i\, \Phi_i\Big(2\|\psi_l^{(k)} - \psi_{[l/2]}^{(k-1)}\|_{\mathcal{C}}\Big) \\
&\quad + \sum_{l\in\{0,\ldots,2^{K_n}-1\}^d} \frac{4^{2p}}{\varepsilon^{2p} n^p} \sum_{i=1}^{p} n^i\, \|\psi_{l+\overline{2}}^{(K_n)}(X_0) - \psi_l^{(K_n)}(X_0)\|_s^i\, \Phi_i\Big(2\|\psi_{l+\overline{2}}^{(K_n)} - \psi_l^{(K_n)}\|_{\mathcal{C}}\Big)\Bigg\}. \qquad (7.20)
\end{aligned}$$

We collect the necessary auxiliary calculations in the following lemma.

Lemma 7.4. *For all $l \in \{0,\ldots,2^k-1\}^d$, $s \geq 1$, and $k,n \in \mathbb{N}^*$,*

$$\begin{aligned}
\|\psi_l^{(k)}(X_0) - \psi_{[l/2]}^{(k-1)}(X_0)\|_s &\leq \Big(\frac{3dh}{2^k}\Big)^{\frac{1}{s}}, \\
\|\psi_{l+\overline{2}}^{(K_n)}(X_0) - \psi_l^{(K_n)}(X_0)\|_s &\leq \Big(\frac{3dh}{2^{K_n}}\Big)^{\frac{1}{s}}, \\
\|\psi_l^{(k)}\|_{\mathcal{C}} &\leq \max\Big\{\Psi\Big(\frac{2^k}{h}\Big), \|1\|_{\mathcal{C}}\Big\}.
\end{aligned}$$

Proof. By (7.11) and the continuity of the $\mathrm{F}_{(i)}$,

$$\begin{aligned}
\|\psi_l^{(k)}(X_0) - \psi_{[l/2]}^{(k-1)}(X_0)\|_s &\leq \|\,\mathbf{1}_{[-\overline{\infty},s_l^{(k)}]}(X_0) - \mathbf{1}_{[-\overline{\infty},s_{[l/2]-\overline{1}}^{(k-1)}]}(X_0)\|_s \\
&\leq \Big(d\max_{i=1,\ldots,d}\big(\mathrm{F}_{(i)}(s_{i,j_i,l_i}^{(k)}) - \mathrm{F}_{(i)}(s_{i,j_i,l_i-3}^{(k)})\big)\Big)^{\frac{1}{s}} \leq \Big(\frac{3dh}{2^k}\Big)^{\frac{1}{s}}.
\end{aligned}$$

The second inequality can be proven in a similar way.

In the first two cases of the definition (7.10), $\psi_l^{(k)}$ is a constant function taking either the value zero or one for each argument. In this case the last inequality of the lemma is trivially satisfied by the conditions on $\|\cdot\|_{\mathcal{C}}$. Else, $\psi_l^{(k)}$ has a representation $\varphi_{(s_{l-\bar{1}}^{(k)}, s_l^{(k)})}$, where

$$\begin{aligned} s_l^{(k)} &= \Big(s_{1,j_1,l_1}^{(k)}, \ldots, s_{d,j_d,l_d}^{(k)}\Big) = \Big(\mathrm{F}_{(1)}^{-1}\big(s_{j_1,l_1}^{(k)}\big), \ldots, \mathrm{F}_{(d)}^{-1}\big(s_{j_d,l_d}^{(k)}\big)\Big), \\ s_{l-\bar{1}}^{(k)} &= \Big(s_{1,j_1,l_1-1}^{(k)}, \ldots, s_{d,j_d,l_d-1}^{(k)}\Big) = \Big(\mathrm{F}_{(1)}^{-1}\big(s_{j_1,l_1}^{(k)} - h2^{-k}\big), \ldots, \mathrm{F}_{(d)}^{-1}\big(s_{j_d,l_d}^{(k)} - h2^{-k}\big)\Big) \end{aligned}$$

and hence, for every $i \in \{1, \ldots, d\}$

$$\begin{aligned} s_{i,j_i,l_i}^{(k)} - s_{i,j_i,l_i-1}^{(k)} \in \big\{\delta > 0 : \exists t \in \mathbb{R},\ |\,\mathrm{F}_{(i)}(t) - \mathrm{F}_{(i)}(t-\delta)| \geq h2^{-k}\big\} \\ \subset \big\{\delta > 0 : \omega_{\mathrm{F}_{(i)}}(\delta) \geq h2^{-k}\big\}. \end{aligned} \tag{7.21}$$

To see this, set $\delta = \mathrm{F}_{(i)}^{-1}(s_{l_i}^{(k)}) - \mathrm{F}_{(i)}^{-1}(s_{l_i-1}^{(k)}) > 0$, $t = \mathrm{F}_{(i)}^{-1}(s_{l_i}^{(k)})$, and recall that the $\mathrm{F}_{(i)}^{-1}$ are injective. Now condition (7.4) yields

$$\|\varphi_l^{(k)}\|_{\mathcal{C}} \leq \Psi\left(\frac{1}{\min\limits_{i=1,\ldots,d} \omega_{\mathrm{F}_{(i)}}(s_{i,j_i,l_i}^{(k)} - s_{i,j_i,l_i-1}^{(k)})}\right) \leq \Psi\Big(\frac{2^k}{h}\Big),$$

since $\min_{i=1,\ldots,d} \omega_{\mathrm{F}_{(i)}}\big(s_{j_i,l_i}^{(k)} - s_{1,j_i,l_i-1}^{(k)}\big) \geq h2^{-k}$ by (7.21). □

An application of Lemma 7.4 to (7.20) yields

$$\begin{aligned} &\mathbf{P}\bigg(\sup_{t \in [t_{j-1}^{(q)}, t_j^{(q)})} |\,\mathrm{U}_n(t) - U_n^{(q)}(t)| \geq \varepsilon\bigg) \\ &\leq 2C\left\{\sum_{k=1}^{K_n}\sum_{i=1}^{p} \frac{2^{dk} n^{-(p-i)}}{\varepsilon_k^{2p}} \Big(\frac{3dh}{2^k}\Big)^{\frac{i}{s}} \Phi_i\Big(2\Psi\Big(\frac{2^k}{h}\Big)\Big) + \sum_{i=1}^{p} \frac{2^{dK_n} n^{-(p-i)}}{(\frac{\varepsilon}{4})^{2p}} \Big(\frac{3dh}{2^{K_n}}\Big)^{\frac{i}{s}} \Phi_i\Big(2\Psi\Big(\frac{2^{K_n}}{h}\Big)\Big)\right\} \\ &\leq 2C\sum_{i=1}^{p}\left\{(3d)^{\frac{i}{s}} n^{-(p-i)} \sum_{k=1}^{K_n} \frac{2^{dk}}{\varepsilon_k^{2p}} \Big(\frac{h}{2^k}\Big)^{\frac{i}{s}} \Phi_i\Big(2\Psi\Big(\frac{2^k}{h}\Big)\Big)\right\} \\ &\leq D\sum_{i=1}^{p}\left\{n^{-(p-i)} \sum_{k=1}^{K_n} 2^{(d-\frac{i}{s})k} k^{4p} \Phi_i\Big(2\Psi\Big(\frac{2^k}{h}\Big)\Big) h^{\frac{i}{s}}\right\} \\ &\leq D\left\{\sum_{i=1}^{p-1} n^{-(p-i)} \left(\frac{2^{K_n}}{h}\right)^{d-\frac{i}{s}} {K_n}^{4p+1} \Phi_i\left(2\Psi\left(\frac{2^{K_n}}{h}\right)\right) h^d\right\} \\ &\quad + D\left\{\sum_{k=1}^{K_n} 2^{(d-\frac{p}{s})k} k^{4p} \Phi_p\Big(2\Psi\Big(\frac{2^k}{h}\Big)\Big) h^{\frac{p}{s}}\right\} \end{aligned} \tag{7.22}$$

for every $j \in \{1, \ldots, q\}^d$, where $D > 0$ denotes some finite constant. In the second inequality we used that Ψ and Φ_i are increasing functions and that $\varepsilon/4 > \varepsilon_{K_n}$. Let us first deal with the term in the last line of (7.22). By condition (7.7) we have

$$\sum_{k=1}^{K_n} 2^{(d-\frac{p}{s})k} k^{4p} \Phi_p\Big(2\Psi\Big(\frac{2^k}{h}\Big)\Big) h^{\frac{p}{s}} \le C'' h^{\frac{p}{s}-\gamma_p} \sum_{k=1}^{\infty} 2^{(\gamma_p-(\frac{p}{s}-d))k} k^{4p},$$

where $\gamma_p < \frac{p}{s} - d$. Hence, there is a nonnegative constant $D' < \infty$ such that

$$\sum_{k=1}^{K_n} 2^{(d-\frac{p}{s})k} k^{4p} \Phi_p\Big(2\Psi\Big(\frac{2^k}{h}\Big)\Big) h^{\frac{p}{s}} \le D' h^{\frac{p}{s}-\gamma_p} = o(h^d). \tag{7.23}$$

Now consider the first summand on the right-hand side of inequality (7.22). In (7.17) we chose $K_n = [\log_2(2^4 d\sqrt{n}h/\varepsilon)]$, and hence condition (7.7) yields for any $i = 1, \dots, p-1$,

$$\begin{aligned} & n^{-(p-i)} \Big(\frac{2^{K_n}}{h}\Big)^{d-\frac{i}{s}} {K_n}^{4p+1} \Phi_i\Big(2\Psi\Big(\frac{2^{K_n}}{h}\Big)\Big) h^d \\ & \le D'' \log_2^{4p+1}\Big(\frac{2^4 d}{\varepsilon}\sqrt{n}h\Big) \cdot (\sqrt{n})^{\gamma_i - (\frac{i}{s}+2(p-i)-d)} h^d \end{aligned}$$

for some non-negative constant $D'' < \infty$. Since $\gamma_i < \frac{i}{s} + 2(p-i) - d$ for $i = 1, \dots, p-1$, by (7.7) we obtain for all $\eta > 0$ and sufficiently large $n \in \mathbb{N}^*$,

$$D \sum_{i=1}^{p-1} \Big\{ n^{-(p-i)} \Big(\frac{2^{K_n}}{h}\Big)^{d-\frac{i}{s}} {K_n}^{4p+1} \Phi_i\Big(2\Psi\Big(\frac{2^{K_n}}{h}\Big)\Big) h^d \Big\} \le \frac{1}{2}\eta h^d. \tag{7.24}$$

Finally, by (7.22), (7.23), and (7.24), for any $\eta > 0$

$$\begin{aligned} & \limsup_{n\to\infty} \mathbf{P}\Big(\sup_{t\in[-\infty,\infty]^d} |\,\mathrm{U}_n(t) - U_n^{(q)}(t)| \ge \varepsilon \Big) \\ & \le \limsup_{n\to\infty} \sum_{j\in\{1\dots,q\}^d} \mathbf{P}\Big(\sup_{t\in[t_{j-1}^{(q)}, t_j^{(q)})} |\,\mathrm{U}_n(t) - U_n^{(q)}(t)| \ge \varepsilon \Big) \\ & \le q^d \Big(o(h^d) + \frac{1}{2}\eta h^d \Big) = q^d \Big(o(q^{-d}) + \frac{1}{2}\eta q^{-d} \Big), \end{aligned}$$

since $h = 1/q$. Hence, there is a $q_0 \in \mathbb{N}^*$ such that

$$\limsup_{n\to\infty} \mathbf{P}\Big(\sup_{t\in[-\infty,\infty]^d} |\,\mathrm{U}_n(t) - U_n^{(q)}(t)| \ge \varepsilon \Big) \le \eta$$

for all $q \ge q_0$. □

With Lemma 7.2 and Lemma 7.3 established, let us finally prove Theorem 7.1. By application of Theorem 1.1 on $\mathbb{D}([-\infty,\infty]^d)$ equipped with the Skorokhod metric ρ, Lemma 7.2 (with $r_k^{(q)} := \frac{k}{q}$) and Lemma 7.3 show that U_n converges in distribution to a process W which is also the limit process of the sequence $W^{(q)}$, $q \in \mathbb{N}^*$. Since all $W^{(q)}$ are centred Gaussian processes, the limit process must also be centred Gaussian.

It remains to prove the continuity of the sample paths of W. At this point we already know that U_n converges weakly to W. Therefore, it is sufficient to show that for every $\varepsilon, \eta > 0$, there is a $\delta > 0$ such that

$$\limsup_{n\to\infty} \mathbf{P}\Big(\sup_{\|t-u\|<\delta} |\,\mathrm{U}_n(t) - \mathrm{U}_n(u)| > 3\varepsilon \Big) < 3\eta. \tag{7.25}$$

The sufficiency of this condition can be proven exactly in the same way as in the proof of Theorem 15.5 in Billingsley (1968, p.127 f).

For all $q \in \mathbb{N}^*$, by some triangle inequality arguments we obtain

$$\begin{aligned}
&\limsup_{n\to\infty} \mathbf{P}\Big(\sup_{\|t-u\|<\delta} |\,\mathrm{U}_n(t) - \mathrm{U}_n(u)| > 3\varepsilon \Big) \\
&\quad \le 2 \limsup_{n\to\infty} \mathbf{P}\Big(\sup_{t} |\,\mathrm{U}_n(t) - U_n^{(q)}(t)| > \varepsilon \Big) \\
&\qquad + \limsup_{n\to\infty} \mathbf{P}\Big(\sup_{\|t-u\|<\delta} |U_n^{(q)}(t) - U_n^{(q)}(u)| > \varepsilon \Big),
\end{aligned}$$

and thus, by Lemma 7.3, there is an $q_0 \in \mathbb{N}^*$ such that for all $q \ge q_0$,

$$\begin{aligned}
&\limsup_{n\to\infty} \mathbf{P}\Big(\sup_{\|t-u\|<\delta} |\,\mathrm{U}_n(t) - \mathrm{U}_n(u)| > 3\varepsilon \Big) \\
&\quad \le 2\eta + \limsup_{n\to\infty} \mathbf{P}\Big(\sup_{\|t-u\|<\delta} |U_n^{(q)}(t) - U_n^{(q)}(u)| > \varepsilon \Big).
\end{aligned} \tag{7.26}$$

Now set $\delta_q := \frac{1}{2} \min_{j\in\{0,\dots,q\}^d} \Big\{ \max_{i=1,\dots,d} |t_{j_i} - t_{j_i - 1}| \Big\}$ and observe that δ_q is strictly positive for any $q \in \mathbb{N}^*$, since the $\mathrm{F}_{(i)}^{-1}$ used in the construction of the t_j are strictly increasing. Obviously, for all $\delta \le \delta_q$ and $\|t - u\| < \delta$, the points $t, u \in [-\infty, \infty]^d$ must be located in adjacent (or identical) intervals of the form $[t_j, t_{j-\bar{1}})$. Since the process $U_n^{(q)}$ is constant on any of the intervals $[t_j, t_{j-\bar{1}})$ and by symmetry in the arguments t, u we obtain

$$\sup_{\|t-u\|<\delta} |U_n^{(q)}(t) - U_n^{(q)}(u)| = \max_{\substack{j\in\{0,\dots,q\}^d \\ z\in\{0,1\}^d,\ j\ge z}} |U_n^{(q)}(t_j) - U_n^{(q)}(t_{j-z})|,$$

and thus,

$$\begin{aligned}
&\mathbf{P}\Big(\sup_{\|t-u\|<\delta} |U_n^{(q)}(t) - U_n^{(q)}(u)| > \varepsilon \Big) \\
&\quad \le 2^d (q+1)^d \max_{\substack{j\in\{0,\dots,q\}^d \\ z\in\{0,1\}^d,\ j\ge z}} \mathbf{P}\Big(|U_n^{(q)}(t_j) - U_n^{(q)}(t_{j-z})| > \varepsilon \Big).
\end{aligned} \tag{7.27}$$

Recall that the functions $\varphi_j^{(q)}$ are defined in (7.9). Analogously to the calculations in Lemma 7.4, one can show that for all $j \in \{0, \dots, q\}^d$ and $z \in \{0,1\}^d$ such that $j \ge z$, we have

$$\|\varphi_{j+\bar{1}}^{(q)}(X_0) - \varphi_{j+\bar{1}-z}^{(q)}(X_0)\|_s \le \Big(\frac{3d}{q}\Big)^{\frac{1}{s}} \qquad \text{and} \qquad \|\varphi_{j+\bar{1}}^{(q)}\|_{\mathcal{C}} \le \max\Big\{ \Psi(q), \|1\|_{\mathcal{C}} \Big\}.$$

Then, by applying one after another Markov's inequality, the $2p$-th moment bounds (7.2), and the preceding inequalities, we obtain

$$\begin{aligned}
&\mathbf{P}\Big(|U_n^{(q)}(t_j) - U_n^{(q)}(t_{j-z})| > \varepsilon\Big)\\
&= \mathbf{P}\bigg(\Big|\sum_{i=1}^{n}\big(\varphi^{(q)}_{j+\bar{1}}(X_i) - \varphi^{(q)}_{j+\bar{1}-z}(X_i)\big) - \mathbf{E}\big(\varphi^{(q)}_{j+\bar{1}}(X_0) - \varphi^{(q)}_{j+\bar{1}-z}(X_0)\big)\Big|^{2p} > n^p\varepsilon^{2p}\bigg)\\
&\le n^{-p}\varepsilon^{-2p}\,\mathbf{E}\Big|\sum_{i=1}^{n}\big(\varphi^{(q)}_{j+\bar{1}}(X_i) - \varphi^{(q)}_{j+\bar{1}-z}(X_i)\big) - \mathbf{E}\big(\varphi^{(q)}_{j+\bar{1}}(X_0) - \varphi^{(q)}_{j+\bar{1}-z}(X_0)\big)\Big|^{2p}\\
&\le 2Cn^{-p}\varepsilon^{-2p}\sum_{i=1}^{p} n^i \|\varphi^{(q)}_{j+\bar{1}}(X_0) - \varphi^{(q)}_{j+\bar{1}-z}(X_0)\|_s^i \Phi_i\Big(2\|\varphi^{(q)}_{j+\bar{1}} - \varphi^{(q)}_{j+\bar{1}-z}\|_{\mathcal{C}}\Big)\\
&\le 2Cn^{-p}\varepsilon^{-2p}\sum_{i=1}^{p} n^i \Big(\frac{3d}{q}\Big)^{\frac{i}{s}}\Phi_i\Big(2\Psi(q)\Big)\\
&\le D\sum_{i=1}^{p} n^{-(p-i)} q^{\gamma_i - \frac{i}{s}}\\
&\le Dq^{\gamma_p - (\frac{p}{s})} + D\sum_{i=1}^{p-1} n^{-(p-i)} q^{\gamma_i - \frac{i}{s}},
\end{aligned}$$

where D is some finite constant. Therefore, by (7.27) there is another finite constant D' such that

$$\mathbf{P}\Big(\sup_{\|t-u\|<\delta} |U_n^{(q)}(t) - U_n^{(q)}(u)| > \varepsilon\Big) \le D'q^d\Big(q^{\gamma_p - \frac{p}{s}} + \sum_{i=1}^{p-1} n^{-(p-i)} q^{\gamma_i - \frac{i}{s}}\Big),$$

and thus,

$$\limsup_{n\to\infty}\mathbf{P}\Big(\sup_{\|t-u\|<\delta_q} |U_n^{(q)}(t) - U_n^{(q)}(u)| > \varepsilon\Big) \le D'q^{\gamma_p - (\frac{p}{s} - d)} < \eta$$

for sufficiently large $q \in \mathbb{N}^*$, say $q \ge q_1$. By (7.26) this implies that (7.25) holds for $\delta = \delta_{\max\{q_0,q_1\}}$. □

Remark 7.5. We saw in the proof that the theorem also holds if (7.2) is only satisfied for a certain subclass of functions in $\mathcal{C}$; more precisely if (7.2)holds for all functions $f \in \mathcal{C}$ of the form $f := \varphi_{(a,b)} - \varphi_{(a',b')}$, where $a, b, a', b' \in [-\infty, \infty]^d$, $a' < b$, are such that

$$\mathbf{P}(X_0 \in [a', b']) \le 2\,\mathbf{P}(X_0 \in [a, b]) \le \mathbf{P}(X_0 \in [a', b]) \le 3\,\mathbf{P}(X_0 \in [a', b']). \tag{7.28}$$

Choosing for each $q \in \mathbb{N}^*$, an $f_q := \varphi_{(a,b)} - \varphi_{(a',b')}$ such that (7.28) is satisfied for $\mathbf{P}(X_0 \in [a', b']) = 1/q$, it can be shown that

$$\|f_q(X_0)\|_s^i = \mathcal{O}\Big(\frac{q^{\gamma_i - \frac{i}{s}}}{\Phi_i(\|f_q\|_{\mathcal{C}})}\Big) \quad \text{as } q \to \infty.$$

8. Empirical CLTs for Causal Functions of I.I.D. Processes

In Chapter 7 we gave quite abstract conditions for the empirical CLT. Here, we establish an empirical CLT for causal functions of i.i.d. processes. Recall that, under reasonable assumption on the physical dependence measure $\delta_{i,m}$, such processes are slowly multiple mixing w.r.t. $\mathcal{H}_\alpha(\mathbb{R}^d,\mathbb{R})$, see Proposition 6.2. In order to apply Theorem 7.2, the second crucial point besides the slow multiple mixing property is that a CLT holds under $\mathcal{H}_\alpha(\mathbb{R}^d,\mathbb{R})$. The following proposition shows that this is true for causal functions of i.i.d. processes under weaker conditions than those that we needed for the slow multiple mixing property.

Proposition 8.1 (A CLT for Causal Functions of I.I.D. Data)**.** *If $(X_i)_{i\in\mathbb{N}}$ is an $\mathbb{R}^d$-valued causal function of an i.i.d. process and satisfies*

$$\sum_{i=1}^{\infty}(\delta_{i,m})^\alpha < \infty \tag{8.1}$$

for some $\alpha \in (0,1]$, $m \in [1,\infty]$ then the CLT (7.1) holds under $\mathcal{H}_\alpha(\mathbb{R}^d,\mathbb{R})$ with

$$\sigma_f^2 = \mathbf{E}(f(X_0)^2) + 2\sum_{i=1}^{\infty}\mathbf{E}\big(f(X_0)f(X_i)\big).$$

Proof. We use a result of Dedecker (1998) which is recalled as Proposition A.1 in the appendix. Choose an arbitrary $f \in \mathcal{H}_\alpha(\mathbb{R}^d,\mathbb{R})$ with $\mathbf{E}(f(X_0)) = 0$. The process $(Y_i)_{i\in\mathbb{N}}$ given by $Y_i := f(X_i)$ is centred, ergodic, has finite second moments, and is adapted to the filtration

$$(\mathcal{M}_i)_{i\in\mathbb{N}} := \big(\sigma(\xi_i,\xi_{i-1},\ldots)\big)_{i\in\mathbb{N}}.$$

As before, let $(\xi'_j)_{j\in\mathbb{Z}}$ be an independent copy of $(\xi_j)_{j\in\mathbb{Z}}$ and set

$$X'_i := G(\xi'_i,\xi'_{i-1},\ldots), \qquad \dot{X}'_i := G(\xi'_i,\xi'_{i-1},\ldots,\xi'_1,\xi_0,\xi_{-1},\ldots).$$

Observe that by the independence of $\mathcal{M}_0$ and $\sigma(\{\xi'_i : i\in\mathbb{Z}\})$ we have that

$$\mathbf{E}\big(f(X'_i)|\mathcal{M}_0\big) = \mathbf{E}(f(X'_i)) = 0,$$

and

$$\mathbf{E}\big(f(X_i)|\mathcal{M}_0\big) = \mathbf{E}\big(f(\dot{X}'_i)|\mathcal{M}_0\big).$$

Thus,

$$\mathbf{E}\Big\{\big|Y_0\,\mathbf{E}\big(Y_i|\mathcal{M}_0\big)\big|\Big\} \le \|f\|_\infty\,\mathbf{E}\Big\{\big|\mathbf{E}\big(f(X_i)|\mathcal{M}_0\big)\big|\Big\} = \|f\|_\infty\,\mathbf{E}\Big\{\big|\mathbf{E}\big(f(\dot{X}_i') - f(X_i')|\mathcal{M}_0\big)\big|\Big\}$$
$$\le \|f\|_\infty\,\mathbf{E}\big|f(\dot{X}_i') - f(X_i')\big|$$

and therefore

$$\mathbf{E}\Big\{\big|Y_0\,\mathbf{E}\big(Y_i|\mathcal{M}_0\big)\big|\Big\} \le \|f\|^2_{\mathcal{H}_\alpha}\,\mathbf{E}\big|\dot{X}_i' - X_i'\big|^\alpha \le \|f\|^2_{\mathcal{H}_\alpha}(\delta_{i,1})^\alpha,$$

where we used Jensen's inequality and $f \in \mathcal{H}_\alpha(\mathbb{R}^d, \mathbb{R})$ in the last steps. Therefore, by (8.1),

$$\sum_{i=1}^n Y_0\,\mathbf{E}\big(Y_i|\mathcal{M}_0\big)$$

converges in L_1, and thus Proposition A.1 applies. □

As a direct application of previous results, we obtain the following theorem.

Theorem 8.1 (Empirical CLT for Causal Functions of I.I.D. Processes)**.** *Let $(X_i)_{i\in\mathbb{N}}$ be an $\mathbb{R}^d$-valued causal function of an i.i.d. sequence. Assume that:*

(i) The distribution function F of X_0 is θ-Hölder for some $\theta \in (0,1]$.

(ii) There are some $s \in [1,\infty)$, $m \in (1,\infty]$ satisfying $\frac{1}{s}+\frac{1}{m} = 1$, an integer $p > sd$, and a positive constant $\alpha \in (0,1]$ satisfying (6.9) and (7.8).

Then there is a centred Gaussian process $W = (W(t))_{t\in[-\infty,\infty]^d}$ with almost surely continuous sample paths such that $\mathrm{U}_n \xrightarrow{d} W$ in the space $\mathbb{D}([-\infty,\infty]^d)$.

Proof. We apply Theorem 7.2 with $\mathcal{C} = \mathcal{H}_\alpha(\mathbb{R}^d, \mathbb{R})$ (see also Remark 7.2). By Proposition 8.1 the CLT (7.1) holds under $\mathcal{H}_\alpha(\mathbb{R}^d, \mathbb{R})$. Proposition 6.2 shows that $(X_i)_{i\in\mathbb{N}}$ is slowly multiple mixing w.r.t. $\mathcal{H}_\alpha(\mathbb{R}^d, \mathbb{R})$ with $\Theta(i) = (\delta_{i,m})^\alpha$ and $s \in [1,\infty)$ such that $\frac{1}{s}+\frac{1}{m} = 1$, and thus $\sum_{i=0}^\infty i^{2p-2}\Theta(i) < \infty$. □

Example 8.1 (Linear Processes)**.** Let $(X_i)_{i\in\mathbb{N}}$ be a causal linear function given by

$$X_i := \sum_{j=0^\infty} a_j \xi_{i-j},$$

where $(\xi_j)_{j\in\mathbb{Z}}$ is an i.i.d. $\mathcal{X}$-valued process, and $(a_j)_{j\in\mathbb{N}}$ is a family of linear operators from $\mathcal{X}$ to $\mathbb{R}^d$. We denote the norm of such operators by

$$\|a\|_* = \sup\{|a(x)| : x \in \mathcal{X}, \|x\|_{\mathcal{X}} \le 1\}.$$

If $\|\xi_0\|_m < \infty$ for some $m > 1$, if the distribution function F of X_0 is θ-Hölder, and if

$$\sum_{j=i}^{\infty} \|a_j\|_* = \mathcal{O}(i^{-b}) \text{ with } b > \min_{p\in\mathbb{N},\ p>sd} \frac{s}{\theta}\frac{(2p-1)p}{p-sd}$$

for $s = \frac{m}{m-1}$, then the empirical Central Limit Theorem holds.

Proof. Let $(\xi'_j)_{j\in\mathbb{Z}}$ be an independent copy of $(\xi_j)_{j\in\mathbb{Z}}$, and p an integer which realizes the minimum in the condition on b. By assumption, there is an $\varepsilon > 0$ such that $b > (1+\varepsilon)\frac{(2p-1)sp}{\theta(p-sd)}$. We can choose $\alpha = \theta(1+\varepsilon)^{-1}\frac{p-sd}{sp}$, ensuring that (7.8) is satisfied. Since

$$\delta_{i,m} = \Big\|\sum_{j=i}^{\infty} a_j(\xi_{i-j} - \xi'_{i-j})\Big\|_m \leq \|\xi_0 - \xi'_0\|_m \sum_{j=i}^{\infty} \|a_j\|_*,$$

we have

$$i^{2p-2}(\delta_{i,m})^\alpha \leq (2\|\xi_0\|_m)^\alpha i^{2p-2}\Big(\sum_{j=i}^{\infty}\|a_j\|_*\Big)^\alpha = \mathcal{O}(i^{2p-2-\alpha b}),$$

where $2p-2 < -1$ since $\alpha b > 2p-1$. Hence (6.9) holds and Theorem 8.1 applies. □

Remark 8.1. The example of causal linear processes has already been studied by several authors. The condition on the coefficients a_i are somewhat strict in the above example. For results with weak assumptions on the convergence rate of $\sum_{j=i}^{\infty}|a_j|$ see Doukhan and Surgailis (1998), Dedecker and Prieur (2007), Wu (2008), and Dedecker (2010).

Example 8.2 (Time Delay Vectors)**.** Let $(X_i)_{i\in\mathbb{N}}$ be a real-valued causal function of an i.i.d. process. We define the time delay vector process $(Y_i)_{i\in\mathbb{N}}$ of dimension $d \geq 1$ by

$$Y_i = (X_i, \ldots, X_{i+d-1}), \quad i \in \mathbb{N}.$$

If the scalar process $(X_i)_{i\in\mathbb{N}}$ satisfies (i) and (ii) of Theorem 8.1, then the Empirical Central Limit Theorem holds for the process $(Y_i)_{i\in\mathbb{N}}$.

Proof. Assume that $(X_i)_{i\in\mathbb{N}}$ satisfies (i) and (ii) of Theorem 8.1 and let us check that the process $(Y_i)_{i\in\mathbb{N}}$ also satisfies these assumptions. Denote by F_X the distribution function of X_0 and by F_Y the multidimensional distribution function of Y_0. The marginals of F_Y are all equal F_X and therefore $\omega_{F_Y} \leq d\omega_{F_X}$. Thus, F_Y is θ-Hölder. Denote by $\delta_{i,m}(X)$ and $\delta_{i,m}(Y)$ the coefficients introduced in (6.8) relative respectively to $(X_i)_{i\in\mathbb{N}}$ and $(Y_i)_{i\in\mathbb{N}}$. We can see that there exists a constant $C > 0$ such that for all $i \in \mathbb{N}$,

$$\delta_{i,m}(Y) \leq C(\delta_{i,m}(X) + \ldots + \delta_{i+d-1,m}(X)).$$

Thus, we infer that $(Y_i)_{i\in\mathbb{N}}$ satisfies (6.9) with the same constant α as for $(X_i)_{i\in\mathbb{N}}$. □

Part III.

Sequential Empirical CLTs for Multiple Mixing Processes.

Based on the article
Herold Dehling, Olivier Durieu, and Marco Tusche (2013): A sequential empirical CLT for multiple mixing processes with application to $\mathcal{B}$-geometrically ergodic Markov chains. Preprint. arXiv: 1303.4537

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

Mathematics Subject Classification (2010): 60F05, 60F17, 60G10, 62G30, 60J05

9. A SECLT for Multiple Mixing Processes

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. 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 $\mathrm{V}_n := (\mathrm{V}_n(f,t))_{(f,t)\in\mathcal{F}\times[0,1]}$ given by

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

where $\mu f := \int_{\mathcal{X}} f\, d\mu$ and $\mu_n(f) := n^{-1}\sum_{i=1}^n f(X_i)$.

For fixed $n\in\mathbb{N}^*$, we consider V_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-algebra. 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.

As in Part I we cannot assume that V_n is measurable and therefore have to use the theory of outer probability and expectation (cf. Section 1.3). Further, we make use of our adapted bracketing numbers defined in Chapter 3 (cf. Section 1.3), which allows us 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 $\mathrm{L}^s(\mu)$-norm,[1] but also with a control of the increasing rate of the $\|\cdot\|_{\mathcal{C}}$-size of the bracketing functions as the $\mathrm{L}^s(\mu)$-norm goes to zero.

Recall that for a probability space $(\mathcal{X},\mathcal{A},\mu)$, $s\geq 1$, and a subclass $\mathcal{G}$ of a normed vector space $(\mathcal{C},\|\cdot\|_{\mathcal{C}})$, an $(\varepsilon, A,\mathcal{G},\mathrm{L}^s(\mu))$-bracket is a set $[l,u] := \{f:\mathcal{X}\to\mathbb{R} : l\leq f\leq u\}$ with $l\leq u\in\mathcal{G}$, $\|u-l\|_s\leq\varepsilon$, and $\max\big\{\|l\|_{\mathcal{C}},\|u\|_{\mathcal{C}}\big\}\leq A$. The bracketing number $N(\varepsilon,A,\mathcal{F},\mathcal{G},\mathrm{L}^s(\mu))$ of a class $\mathcal{F}$ of real-valued functions on $\mathcal{X}$ w.r.t. $\mathcal{G}$ is defined as the minimum number of $(\varepsilon,A,\mathcal{G},\mathrm{L}^s(\mu))$-brackets, needed to cover $\mathcal{F}$.

9.1. Statement of Results

Let $(\mathcal{C},\|\cdot\|_{\mathcal{C}})$ be some normed vector space of function on $\mathcal{X}$. As in Chapter 3 we make two basic assumptions concerning the process $(f(X_i))_{i\in\mathbb{N}}$, $f\in\mathcal{C}$.

[1] The $\mathrm{L}^s(\mu)$-norm is given by $\|f\|_s = \mu(|f|^s)^{1/s}$.

Assumption 9.I (Finite-Dimensional Sequential CLT for $\mathcal{C}$-Observables)**.** For every choice of $f_1, \ldots, f_k \in \mathcal{C}$ and $t_1, \ldots, t_k \in [0,1]$

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

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

Assumption 9.II (Moment Bounds for $\mathcal{C}$-Observables)**.** For fixed $p \in \mathbb{N}^*$, $s \geq 1$, and monotone increasing functions $\Phi_1, \ldots, \Phi_p : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$, there exists some $C_p > 0$ such that for all $f \in \mathcal{C}$ with $\|f\|_\infty \leq 1$

$$\mathbf{E}\left(\sum_{i=1}^{n}\big(f(X_i) - \mu f\big)\right)^{2p} \leq C_p \sum_{i=1}^{p} n^i \|f\|_s^i \Phi_i(\|f\|_{\mathcal{C}}). \tag{9.1}$$

Remark 9.1. Of course, Assumption 9.II implies that for every fixed $M > 0$ inequality (9.1) holds with different constants C_p uniformly for all $f \in \mathcal{C}_M$.

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

Theorem 9.1 (Sequential Empirical CLT)**.** *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 $\mathcal{X}$, 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 $\Phi_1, \ldots, \Phi_p : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$, Assumption 9.I and Assumption 9.II hold. Moreover, assume that there exist a constant $r > -1$ and a monotone increasing function $\Psi : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ such that*

$$\int_0^1 \varepsilon^r \sup_{\varepsilon\le\delta\le1} N^2\big(\delta, \Psi\left(\delta^{-1}\right), \mathcal{F}, \mathcal{G}, \mathrm{L}^s(\mu)\big) d\varepsilon < \infty. \tag{9.2}$$

If

$$\Phi_i(2\Psi(x)) = \mathcal{O}(x^{\gamma_i}), \quad \text{as } x \to \infty \tag{9.3}$$

for some non-negative constants γ_i such that

$$\gamma_i < 2p - (i + r + 2), \tag{9.4}$$

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

The proof is given in Section 9.2.

Remark 9.2. (i) Observe that Assumption 9.I is stronger than Assumption 3.I. This comes from the fact, that in the situation of $t_1 \neq t_2$ the limit distribution of $\mathrm{V}(f_1, t_1), \mathrm{V}(f_2, t_2)$ can not be computed directly from the one-dimensional CLT via the Cramér-Wold device, since we consider non-necessarily independent processes $(X_i)_{i\in\mathbb{N}}$.

(ii) Assumption 9.II is a more general version of Assumption 3.II, which corresponds to the case where $\Phi_i = \log^{2p+ai}(\mathrm{id}_{\mathbb{R}} + 1)$. This condition can be deduced from the multiple mixing property of the underlying process w.r.t. $\mathcal{C}$. Recall that this property can be established for instance for $\mathcal{B}$-geometrically ergodic Markov chains (Section 2.2), dynamical systems with a spectral gap of the Perron–Frobenius Operator (Section 2.5), the ergodic automorphism of the torus (Section 2.6), and in a non-exponential form for causal functions of i.i.d. processes (Section 6.2).

(iii) For examples of classes $\mathcal{F}$ that satisfy (9.2) see Chapter 4.

(iv) As in Section 3.1, inequality (9.2) holds for all $r > 2r' - 1$ with $r' \geq 0$ if

$$N\big(\varepsilon, \Psi\left(\varepsilon^{-1}\right), \mathcal{F}, \mathcal{G}, \mathrm{L}^s(\mu)\big) = \mathcal{O}(\varepsilon^{-r'}) \quad \text{as } \varepsilon \to 0.$$

Theorem 9.1 holds no information about the covariance structure of the limit process K. However, under some additional conditions, it can be shown that the limit process of V_n is indeed a Kiefer process (cf. Remark 9.3).

Lemma 9.1. *In the situation of Theorem 9.1, assume that*

(i) Assumption 9.I holds with covariance matrix Σ given by

$$\Sigma_{i,j} = \min\{t_i, t_j\} \left\{ \sum_{k=0}^{\infty} \mathbf{Cov}\big(f_i(X_0), f_j(X_k)\big) + \sum_{k=1}^{\infty} \mathbf{Cov}\big(f_j(X_0), f_i(X_k)\big) \right\}, \tag{9.5}$$

(ii) there is a function $\Theta : \mathbb{N} \longrightarrow \mathbb{R}_+$ and a constant $b > 1$ satisfying

$$\sum_{k=1}^{\infty} \Psi(k^b)\Theta(k) < \infty \tag{9.6}$$

such that for all $f \in \mathcal{C}$ and all $\varphi \in \mathcal{F} \cup (\mathcal{F} - \mathcal{G})$

$$\big|\mathbf{Cov}\big(\varphi(X_0), f(X_k)\big)\big| \leq \|\varphi\|_\infty \|f\|_{\mathcal{C}} \Theta(k). \tag{9.7}$$

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

$$\begin{aligned} &\mathbf{Cov}\big(K(f,t), K(g,u)\big) \\ &\quad = \min\{t,u\} \left\{ \sum_{k=0}^{\infty} \mathbf{Cov}\big(f(X_0), g(X_k)\big) + \sum_{k=1}^{\infty} \mathbf{Cov}\big(f(X_k), g(X_0)\big) \right\}, \quad f,g \in \mathcal{F},\ t,u \in [0,1]. \end{aligned} \tag{9.8}$$

The proof is given in Section 9.3.

Remark 9.3. A centred Gaussian process K with covariance structure (9.8) is often referred to as a Kiefer process.

As Assumption 9.II can be deduced from the multiple mixing property, for this kind of processes we have the following version of Theorem 9.1.

Theorem 9.2 (Sequential Empirical CLT for Multiple Mixing Processes)**.** *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 $s \geq 1$, the process $(X_i)_{n\in\mathbb{N}}$ is multiple mixing w.r.t. a normed vector space $\mathcal{C}$ of measurable functions on $\mathcal{X}$, where for every $p \in \mathbb{N}^*$ the multivariate polynomial Q in inequality (2.1) is of total degree not larger than d_0. If Assumption 9.I holds and if there are a $\|\cdot\|_\infty$-bounded subset $\mathcal{G}$ of $\mathcal{C}$, an $r > -1$, and a $\gamma > \max\{1, d_0\}$ such that*

$$\int_0^1 \varepsilon^r \sup_{\varepsilon\leq\delta\leq 1} N^2\big(\delta, \exp(C\delta^{-1/\gamma}), \mathcal{F}, \mathcal{G}, \mathrm{L}^s(\mu)\big) d\varepsilon < \infty, \tag{9.9}$$

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

If further the covariance matrix Σ in Assumption 9.I is given by (9.5) and if there are constants $\rho \in (0,1)$ and $D > 0$ such that for all $f \in \mathcal{C}$ and all $\varphi \in \mathcal{F} \cup (\mathcal{F} - \mathcal{G})$

$$\big|\mathbf{Cov}\big(\varphi(X_0), f(X_k)\big)\big| \leq D\|\varphi\|_\infty \|f\|_{\mathcal{C}}\, \rho^k,$$

then the covariance structure of the limit process K is given by (9.8).

Proof. By Proposition 2.1, Assumption 9.II is holds for $\Phi_i(x) = c\log^{2p+ai}(x+1)$ with $a = \max\{-1, d_0 - 1\}$ and some $c > 0$ depending only on p. Thus choosing $\Psi(x) := \exp(Cx^{1/\gamma})$ for some $C > 0$ and $\gamma > 1$ (which gives a quite relaxed entropy condition concerning the $\|\cdot\|_{\mathcal{C}}$-size), we have $\Phi_i(2\Psi(x)) = O(x^{(2p+ai)/\gamma})$. Therefore condition (9.4) holds for sufficiently large $p \in \mathbb{N}^*$ if $\gamma > \max\{1, d_0\}$. The covariance structure of the limit process is a direct consequence of Lemma 9.1 with $\Theta(k) = \rho^k$ and $b \in (1, \gamma)$. ☐

Applications of Theorem 9.2 are provided in the following chapters. In Section 10.1 we establish sequential empirical CLTs for $\mathcal{B}$-geometrically ergodic Markov chains and dynamical systems with a spectral gap on the corresponding transfer operator. An application to the ergodic automorphism of the multidimensional torus is given in Chapter 11.

9.2. Proof of Theorem 9.1

The main idea of the proof is to introduce some approximation $V_n^{(q)}$ for the original process V_n, which is based on functions in $\mathcal{G}$ and thus can be controlled by Assumption 9.I and 9.II. 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 $\{g_{q,1}, \ldots, g_{q,N_q}\} \subset \mathcal{G}$ and $\{g'_{q,1}, \ldots, g'_{q,N_q}\} \subset \mathcal{G}$, such that

$$\|g_{q,i} - g'_{q,i}\|_s \leq 2^{-q}, \qquad \|g_{q,i}\|_{\mathcal{C}} \leq \Psi(2^q), \qquad \|g'_{q,i}\|_{\mathcal{C}} \leq \Psi(2^q) \tag{9.10}$$

and for all $f \in \mathcal{F}$, there exists some i such that $g_{q,i} \le f \le g'_{q,i}$. Further, by (9.2),

$$\sum_{q\ge 1} 2^{-(r+1)q} N_q^2 < \infty. \tag{9.11}$$

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 $g_{q,i} \le f \le g'_{q,i}$. 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 $\tau_q t := \max\{(j-1)2^{-q} \le t : j = 1,\dots,2^q\}$ and further $\tau'_q t := \tau_q t + 2^{-q}$. We extend the notation introduced in Chapter 10 to arbitrary μ-integrable functions $f : \mathcal{X} \longrightarrow \mathbb{R}$ by setting

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

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

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

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

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

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

To draw the connection between the weak asymptotic behaviour of the original process V_n and the approximating process $V_n^{(q)}$, we use Theorem 1.1. We establish the conditions (1.10) and (1.11) in the two following propositions.

Proposition 9.1. *For all $q \in \mathbb{N}^*$ the process $(V_n^{(q)}(f,t))_{(f,t)\in\mathcal{F}\times[0,1]}$ converges in distribution to a piecewise constant Gaussian process $(V^{(q)}(f,t))_{(f,t)\in\mathcal{F}\times[0,1]}$ as $n \to \infty$.*

Proposition 9.2. *Assume that Assumption 9.II holds for some $p \in \mathbb{N}^*$, $s \ge 1$ and some monotone increasing functions $\Phi_1,\dots,\Phi_p : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$. Moreover, suppose there exists a constant $r > -1$ and an monotone increasing function $\Psi : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ such that (9.2) holds. If (9.3) holds for some non-negative constants γ_i satisfying (9.4), then for all $\varepsilon, \eta > 0$ there exists some q_0 such that for all $q \ge q_0$*

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

Proof of Theorem 9.1. We can now apply Theorem 1.1 with $\xi_n = \mathrm{V}_n$, $\xi_n^{(q)} = V_n^{(q)}$, $\xi^{(q)} = V^{(q)}$. By Proposition 9.1 the convergence (1.10) holds, while (1.11) is satisfied due to Proposition 9.2.

Therefore V_n converges in distribution to an $\ell^\infty(\mathcal{F} \times [0,1])$-valued, separable random variable W. Furthermore, we know that $V^{(q)}$ 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 (cf. Lemma 1.3.2 in van der Vaart and Wellner (1996)). □

Proof of Proposition 9.1. Since by construction $\pi_q f \in \mathcal{G}$ for all $f \in \mathcal{F}$, due to Assumption 9.I, the finite-dimensional process $(V_n^{(q)}(f_1, t_1), \ldots, V_n^{(q)}(f_k, t_k))$ converges in distribution to some multi-dimensional normal distributed random variable $(V^{(q)}(f_1, t_1), \ldots, V_n^{(q)}(f_k, t_k))$ for all fixed $k \in \mathbb{N}^*$, $f_1, \ldots, f_k \in \mathcal{F}$, $t_1, \ldots, t_k \in [0,1]$. All $V_n^{(q)}$, $n \in \mathbb{N}^*$, are constant on each $\mathcal{F}_{q,i} \times \mathcal{T}_{q,j}$, $i = 1, \ldots, N^q$, $j = 1, \ldots, 2^q$. Therefore $V^{(q)}$ 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 $(V_n^{(q)}(f,t))_{(f,t) \in \mathcal{F} \times [0,1]}$. □

Proof of Proposition 9.2. Let $\overline{Z} := Z - \mathbf{E}\, Z$ denote the centring of a random variable Z and observe that for any random variables $Y_l \leq Y \leq Y_u$ the inequality

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

holds. Since for $f \in \mathcal{F}$, $k \in \mathbb{N}$ we have $\mu_{[nt]}(\pi_{q+k} f, t) \leq \mu_{[nt]}(f,t) \leq \mu_{[nt]}(\pi'_{q+k} f, t)$, using that $\|\cdot\|_1 \leq \|\cdot\|_s$ for $s \geq 1$ and applying (9.10), we obtain

$$\begin{aligned}
&\left|\mathrm{V}_n(f,t) - \mathrm{V}_n(\pi_{q+k} f, t)\right| \\
&\quad \leq \left|\mathrm{V}_n(\pi'_{q+k} f, t) - \mathrm{V}_n(\pi_{q+k} f, t)\right| + \frac{[nt]}{\sqrt{n}}\, \mathbf{E}\left|\mu_{[nt]}(\pi'_{q+k} f - \pi_{q+k} f)\right| \\
&\quad \leq \left|\mathrm{V}_n(\pi'_{q+k} f, t) - \mathrm{V}_n(\pi_{q+k} f, t)\right| + \sqrt{n} 2^{-(q+k)}.
\end{aligned} \tag{9.12}$$

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

$$\begin{aligned}
\left|\mathrm{V}_n(g,t) - \mathrm{V}_n(g, \tau_{q+k} t)\right| &= \frac{1}{\sqrt{n}} \left| \sum_{i=[n\tau_{q+k}t]+1}^{[nt]} g(X_i) - \mu(g) \right| \\
&\leq 2Mn^{-\frac{1}{2}}([nt] - [n\tau_{q+k} t]) \\
&\leq 4M\sqrt{n} 2^{-(q+k)},
\end{aligned} \tag{9.13}$$

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

$$V_n'^{(q)}(f,t) := \mathrm{V}_n(\pi'_q f, \tau'_q t).$$

An application of the triangle inequality, (9.12), and (9.13) yields

$$\left|\mathrm{V}_n(f,t) - V_n^{(q+k)}(f,t)\right| \leq \left|V_n'^{(q+k)}(f,t) - V_n^{(q+k)}(f,t)\right| + (4M+1)\sqrt{n} 2^{-q+k}. \tag{9.14}$$

Combining (9.14) with a telescopic sum argument, one obtains for any $K \geq 1$

$$\begin{aligned}
&\left| \mathrm{V}_n(f,t) - V_n^{(q)}(f,t) \right| \\
&= \left| \left\{ \sum_{k=1}^{K} V_n^{(q+k)}(f,t) - V_n^{(q+k-1)}(f,t) \right\} + \mathrm{V}_n(f,t) - V_n^{(q+K)}(f,t) \right| \\
&\leq \left\{ \sum_{k=1}^{K} \left| V_n^{(q+k)}(f,t) - V_n^{(q+k-1)}(f,t) \right| \right\} + \left| V_n'^{(q+K)}(f,t) - V_n^{(q+K)}(f,t) \right| \\
&\quad + (4M+1)\sqrt{n} 2^{-(q+K)}. \qquad (9.15)
\end{aligned}$$

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 (9.15) implies

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

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

$$\begin{aligned}
&\mathbf{P}^* \left(\sup_{t \in \mathcal{T}_{q,j}} \sup_{f \in \mathcal{F}_{q,i}} |\mathrm{V}_n(f,t) - V_n^{(q)}(f,t)| \geq \varepsilon \right) \\
&\leq \left\{ \sum_{k=1}^{K_{n,q}} \mathbf{P}^* \left(\sup_{t \in \mathcal{T}_{q,j}} \sup_{f \in \mathcal{F}_{q,i}} \left| V_n^{(q+k)}(f,t) - V_n^{(q+k-1)}(f,t) \right| \geq \varepsilon_k \right) \right\} \\
&\quad + \mathbf{P}^* \left(\sup_{t \in \mathcal{T}_{q,j}} \sup_{f \in \mathcal{F}_{q,i}} \left| V_n'^{(q+K)}(f,t) - V_n^{(q+K)}(f,t) \right| \geq \frac{\varepsilon}{4} \right). \qquad (9.16)
\end{aligned}$$

Recall that (π_{q+k}, τ_{q+k}) and thus $V_n^{(q+k)}$ and $V_n'^{(q+k)}$ 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 (9.16) 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 (9.16) implies

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

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

$$\begin{aligned}
&\mathbf{P}^*\Big(\sup_{t\in\mathcal{T}_{q,j}}\sup_{f\in\mathcal{F}_{q,i}} |\,\mathrm{V}_n(f,t)-V_n^{(q)}(f,t)|\geq\varepsilon\Big)\\
&\leq\Big\{\sum_{k=1}^{K_{n,q}}\sum_{t\in T_{k,q,j}}\sum_{f\in F_{k,q,i}}\Big(\frac{\varepsilon_k}{2}\Big)^{-2p}\Big(\mathbf{E}\Big|\mathrm{V}_n(\pi_{q+k}f,\tau_{q+k-1}t)-\mathrm{V}_n(\pi_{q+k-1}f,\tau_{q+k-1}t)\Big|^{2p}\\
&\qquad+\mathbf{E}\Big|\mathrm{V}_n(\pi_{q+k}f,\tau_{q+k}t)-\mathrm{V}_n(\pi_{q+k}f,\tau_{q+k-1}t)\Big|^{2p}\Big)\Big\}\\
&\quad+\sum_{t\in T_{K_{n,q},q,j}}\sum_{f\in F_{K_{n,q},q,i}}\Big(\frac{\varepsilon}{8}\Big)^{-2p}\Big(\mathbf{E}\Big|\mathrm{V}_n(\pi'_{q+K_{n,q}}f,\tau_{q+K_{n,q}}t)-\mathrm{V}_n(\pi_{q+K_{n,q}}f,\tau_{q+K_{n,q}}t)\Big|^{2p}\\
&\qquad+\mathbf{E}\Big|\mathrm{V}_n(\pi'_{q+K_{n,q}}f,\tau'_{q+K_{n,q}}t)-\mathrm{V}_n(\pi'_{q+K_{n,q}}f,\tau_{q+K_{n,q}}t)\Big|^{2p}\Big).
\end{aligned}\tag{9.17}$$

We will treat the expected values on the r.h.s. of inequality (9.17) separately now by using Assumption 9.II and properties of our brackets used to cover $\mathcal{F}$. Recall that by (9.10) we have

$$\begin{aligned}
\|\pi_{q+k}f-\pi_{q+k-1}f\|_s &\leq\|\pi_{q+k}f-f\|_s+\|\pi_{q+k-1}f-f\|_s\leq 3\cdot 2^{-(q+k)} \qquad (9.18)\\
\|\pi_{q+k}f-\pi'_{q+k}f\|_s &\leq 2^{-(q+k)}\\
\|\pi_{q+k}f-\pi_{q+k-1}f\|_{\mathcal{C}} &\leq 2\Psi(2^{q+k}) \qquad (9.19)\\
\|\pi_{q+k}f-\pi'_{q+k}f\|_{\mathcal{C}} &\leq 2\Psi(2^{q+k}).
\end{aligned}$$

Notation. For convenience, from now on 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 the corresponding statement.

Applying successively Assumption 9.II and equations (9.18), (9.19), and (9.3) we have

$$\begin{aligned}
&\mathbf{E}\big|\mathrm{V}_n(\pi_{q+k}f,\tau_{q+k-1}t)-\mathrm{V}_n(\pi_{q+k-1}f,\tau_{q+k-1}t)\big|^{2p}\\
&\quad\ll n^{-p}\sum_{\ell=1}^{p} n^{\ell}\|\pi_{q+k}f-\pi_{q+k-1}f\|_s^{\ell}\Phi_\ell(\|\pi_{q+k}f-\pi_{q+k-1}f\|_{\mathcal{C}})\\
&\quad\ll \sum_{\ell=1}^{p} n^{-(p-\ell)}2^{(\gamma_\ell-\ell)(q+k)} \qquad (9.20)
\end{aligned}$$

and analogously

$$\mathbf{E}\big|\mathrm{V}_n(\pi'_{q+K_{n,q}}f,\tau_{q+K_{n,q}}t)-\mathrm{V}_n(\pi_{q+K_{n,q}}f,\tau_{q+K_{n,q}}t)\big|^{2p}\ll\sum_{\ell=1}^{p} n^{-(p-\ell)}2^{(\gamma_\ell-\ell)(q+K_{n,q})}. \qquad (9.21)$$

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

$$\mathbf{E}\big|\mathrm{V}_n(g,\tau_{q+k}t)-\mathrm{V}_n(g,\tau_{q+k-1}t)\big|^{2p}=n^{-p}\,\mathbf{E}\left[\left(\sum_{i=1}^{[n\tau_{q+k}t]-[n\tau_{q+k-1}t]}\big(g(X_i)-\mu g\big)\right)^{2p}\right], \qquad (9.22)$$

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 Assumption 9.II and equations (9.10), and (9.3) to (9.22) we obtain

$$\begin{aligned}
\mathbf{E}\big|\mathrm{V}_n(\pi_{q+k}f,\tau_{q+k}t)-\mathrm{V}_n(\pi_{q+k}f,\tau_{q+k-1}t)\big|^{2p}&\ll n^{-p}\sum_{\ell=1}^{p}\big(n2^{-(q+k)}\big)^{\ell}\|\pi_{q+k}f\|_s^{\ell}\Phi_\ell(\|\pi_{q+k}f\|_{\mathcal{C}})\\
&\ll\sum_{\ell=1}^{p} n^{-(p-\ell)}2^{(\gamma_\ell-\ell)(q+k)} \qquad (9.23)
\end{aligned}$$

and analogously

$$\mathbf{E}\big|\mathrm{V}_n(\pi'_{q+K_{n,q}}f,\tau'_{q+K_{n,q}}t)-\mathrm{V}_n(\pi'_{q+K_{n,q}}f,\tau_{q+K_{n,q}}t)\big|^{2p}\ll\sum_{\ell=1}^{p} n^{-(p-\ell)}2^{(\gamma_\ell-\ell)(q+K_{n,q})}. \qquad (9.24)$$

Now, apply (9.20), (9.21), (9.23), and (9.24) to (9.17). We infer

$$\begin{aligned}
&\mathbf{P}^*\left(\sup_{t\in\mathcal{T}_{q,j}}\sup_{f\in\mathcal{F}_{q,i}}\Big|\mathrm{V}_n(f,t)-V_n^{(q)}(f,t)\Big|\ge\varepsilon\right)\\
&\quad\ll\sum_{k=1}^{K_{n,q}}\#T_{k,q,j}\,\#F_{k,q,i}\,\frac{(k(k+1))^{2p}}{\varepsilon^{2p}}\sum_{\ell=1}^{p} n^{-(p-\ell)}2^{(\gamma_\ell-\ell)(q+k)}. \qquad (9.25)
\end{aligned}$$

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 $\sum_{i=1}^{N_q} \#F_{k,q,i} = \#\mathcal{F}(q+k) \leq N_{q+k-1}N_{q+k}$. Therefore (9.25) yields

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

This implies that for any $\eta > 0$

$$\mathbf{P}^* \left(\sup_{t\in[0,1]} \sup_{f\in\mathcal{F}} \left| \mathrm{V}_n(f,t) - V_n^{(q)}(f,t) \right| > \varepsilon \right)$$
$$\ll \sum_{\ell=1}^{p} n^{-(p-\ell)} \max\left\{1\ ,\ 2^{(\gamma_\ell - \ell + r + 2 + \eta)(q + K_{n,q})}\right\} \sum_{k=1}^{K_{n,q}} N_{q+k-1} N_{q+k} k^{4p} 2^{-(r+1+\eta)(q+k)}$$
$$\ll \max\left\{1\ ,\ \max_{\ell=1,\ldots,p} n^{\frac{1}{2}(\gamma_\ell + \ell - 2p + r + 2 + \eta)}\right\} \sum_{k=q+1}^{\infty} N_{k-1} N_k k^{4p} 2^{-(r+1+\eta)k}. \tag{9.26}$$

By (9.4) we can choose η small enough to assure $\gamma_\ell + \ell - 2p + r + 2 + \eta < 0$ for all $\ell = 1, \ldots, p$. Thus the factor in front of the sum is uniformly bounded w.r.t. n. Using (9.11), we obtain

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

for sufficiently small $\eta > 0$ which implies that the series in (9.26) goes to zero as $q \to \infty$. □

9.3. Proof of Lemma 9.1

This proof parallels the proof of Lemma 3.1. We therefore shorten calculations where the same arguments are used. For $f \in \mathcal{F}$, recall the definition of the approximating functions $\pi_q f$ in Section 9.2. By the entropy condition in Theorem 9.1, we know that for every $q \in \mathbb{N}^*$

$$\|f - \pi_q f\|_s \leq 2^{-q} \tag{9.27}$$
$$\|\pi_q f\|_{\mathcal{C}} \leq \Psi(2^q). \tag{9.28}$$

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

$$\|g_k - g\|_s \leq k^{-b} \tag{9.29}$$
$$\|g_k\|_{\mathcal{C}} \leq \Psi(k^b). \tag{9.30}$$

Let $V^{(q)}$ denote the limit process given in Proposition 9.1. Condition (i) implies that for all $f, g \in \mathcal{F}$, $t, u \in [0,1]$ and $q \in \mathbb{N}^*$

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

With the same arguments as in the proof of Lemma 3.1, it is sufficient to show that for all $f, g \in \mathcal{F}$ the term $\big|\sum_{k=0}^{\infty}\mathbf{Cov}\big(\pi_q f(X_0), \pi_q g(X_k)\big) - \mathbf{Cov}\big(f(X_0), g(X_k)\big)\big|$ converges to zero as $k \to \infty$. Let $k(q) := 2^{q/b}$. By the triangle inequality, we have

$$\begin{aligned}&\Big|\sum_{k=0}^{\infty}\mathbf{Cov}\big(\pi_q f(X_0), \pi_q g(X_k)\big) - \mathbf{Cov}\big(f(X_0), g(X_k)\big)\Big|\\
&\leq \sum_{k=0}^{k(q)}\big|\mathbf{Cov}\big(\pi_q f(X_0) - f(X_0), \pi_q g(X_k)\big)\big| + \sum_{k=0}^{k(q)}\big|\mathbf{Cov}\big(f(X_0), \pi_q g(X_k) - g(X_k)\big)\big| \qquad (9.31)\\
&\quad + \sum_{k=k(q)+1}^{\infty}\big|\mathbf{Cov}\big(\pi_q f(X_0) - f(X_0), \pi_q g(X_k)\big)\big| \qquad (9.32)\\
&\quad + \sum_{k=k(q)+1}^{\infty}\big|\mathbf{Cov}\big(f(X_0), \pi_q g(X_k) - g(X_k)\big)\big|. \qquad (9.33)\end{aligned}$$

Recall that both $\mathcal{F}$ and $\mathcal{G}$ are uniformly bounded in $\|\cdot\|_\infty$-norm. The term in line (9.31) can be treated exactly as the term (3.13) in the proof of Lemma 3.1. For the term in line (9.32), by (9.7), (9.27), (9.28), condition (9.6), and the monotonicity of Ψ, with similar calculations as for the term in line (3.14), we obtain

$$\sum_{k=k(q)+1}^{\infty}\big|\mathbf{Cov}\big(\pi_q f(X_0) - f(X_0), \pi_q g(X_k)\big)\big| \ll \sum_{k=k(q)+1}^{\infty}\Psi(2^q)\Theta(k) \longrightarrow 0 \quad \text{as } q \to \infty.$$

By the triangle inequality, the term in line (9.33) can be bounded by

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

With (9.7), (9.28), and (9.30) we obtain with the same calculation as for the term (3.16)

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

where we also used that Ψ is increasing and applied condition (9.6). Finally, the second series can be treated the same way as (3.17), which completes the proof. □

10. SECLTs for Markov Chains and Dynamical Systems with a Spectral Gap

In this chapter we present an application of Theorem 9.2 to processes given by a $\mathcal{B}$-geometrically ergodic Markov chain or a dynamical system. This leads to sequential empirical CLTs for Markov chains and dynamical systems with a spectral gap on the corresponding operator.

10.1. $\mathcal{B}$-Geometrically Ergodic Markov Chains

Let (X_i) and $(\mathcal{B}, \|\cdot\|)$ be the Markov chain and the complex Banach space of $\mathbb{C}$-valued functions on $\mathcal{X}$ introduced in Section 2.2. Recall that we call $(X_i)_{i\in\mathbb{N}^*}$ $\mathcal{B}$-geometrically ergodic, if the corresponding Markov operator P on $\mathcal{B}$ satisfies

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

Let m be a constant in $[1, \infty]$. We use the following three assumptions on the space $\mathcal{B}$, which were introduced in Section 2.2.

(10.B) $\mathbf{1}_{\mathcal{X}} \in \mathcal{B}$, $|f|$ and $\overline{f} \in \mathcal{B}$ for all $f \in \mathcal{B}$, and the mappings $f \mapsto f(x)$ are continuous on $\mathcal{B}$ for every $x \in \mathcal{X}$.

(10.C) $\mathcal{B}$ is continuously included in $\mathrm{L}^m(\nu)$, i.e. $\mathcal{B} \subset \mathrm{L}^m(\mu)$ and there is a $K > 0$ such that $\|f\|_m \leq K\|f\|_{\mathcal{B}}$ for all $f \in \mathcal{B}$.

(10.D) there exist some $C > 0$ 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 $\|fg\|_{\mathcal{B}} \leq C \max\{\|f\|_{\mathcal{B}}, \|g\|_{\mathcal{B}}\}^{\ell}$.

We use the bracketing numbers computed in Chapter 4 to obtain a control of the size of $\mathcal{F}$. Since $\mathcal{F}$ is composed of real-valued functions, we can restrict to the class $\mathcal{B}_{\mathbb{R}}$ of real-valued functions in the space $\mathcal{B}$ to obtain the bracketings for $\mathcal{F}$.[1] Our conditions on the Markov chain (in particular condition (10.A)) 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 $\mathcal{G} \subset \mathcal{B}_{\mathbb{R}}$,

(10.E) there exist some $C > 0$, $r > -1$, and $\gamma > 1$ such that

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

[1] Note that $(\mathcal{B}_{\mathbb{R}}, \|\cdot\|_{\mathcal{B}})$ is a real Banach space.

As we saw in Section 2.2 (Lemma 2.2), the assumptions (10.A), (10.B), (10.C), and (10.D) imply that $(X_i)_{i\in\mathbb{N}}$ is multiple mixing w.r.t. $\mathcal{B}_{\mathbb{R}} = \{f \in \mathcal{B} : f(\mathcal{X}) \subset \mathbb{R}\}$ with $d_0 = 0$, $s = m/(m-1)$. In order to apply Theorem 3.2, we need that sequential finite-dimensional CLT holds under $\mathcal{B}_{\mathbb{R}}$.

The next section establishes a sequential finite-dimensional CLT for $\mathcal{B}$-observables of a $\mathcal{B}$-geometrically ergodic Markov chain.

10.2. A Sequential Finite-Dimensional CLT for Real-Valued $\mathcal{B}$-Observables

It is known, that under certain regularity condition on the perturbation of the Markov Operator P, observables of real-valued functions of $\mathcal{B}$-geometrically ergodic Markov chains satisfy a finite-dimensional CLT (cf. Hennion and Hervé (2001)). Paralleling the approach of Hennion and Hervé (2001), here we show that there is also an sequential finite-dimensional CLT available for such observables.

For a measurable real-valued function f on $\mathcal{X}$ and a real number $t \in [0,1]$, we introduce the notation

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

Further, let $\mathcal{C} \subset \mathcal{B}$ be a space of measurable functions from $\mathcal{X}$ to $\mathbb{R}$. Using Fourier kernels, we introduce for a function $f : \mathcal{X} \longrightarrow \mathbb{R}$ and a real number $t \in \mathbb{R}$ the perturbed operators given by

$$P_{f,t}\varphi = P(e^{itf}\varphi) = \int_{\mathcal{X}} e^{itf(y)}\varphi(y)P(\cdot,dy).$$

We introduce the following regularity assumption on the perturbed operators of P:

(10.F) for all $f \in \mathcal{C}$, for t in a neighbourhood I_f of 0 we have that $P_{f,t} \in \mathcal{L}(\mathcal{B})$ and further that the mapping $I_f \longrightarrow \mathcal{L}(\mathcal{B})$, $t \mapsto P_{f,t}$ is two times continuous differentiable on I_f with derivative in $t = 0$ given by

$$\Big(\frac{\partial^k}{\partial t^k}P_{f,t}\Big)_{t=0}\varphi = P\big((if)^k\varphi\big) \quad k \in \{1,2\}.$$

If $\mathcal{B}$ is a Banach algebra and 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 analytic and therefore condition (10.F) is also satisfied (see Lemma A.2 for details). An example for Markov chains that satisfy condition (10.F) are iterative Lipschitz models as introduced in Section 2.4.

Proposition 10.1. *Let $\alpha \in (0,1]$, $\beta \in [0,1]$. If $(X_i)_{i\in\mathbb{N}}$ is an iterative Lipschitz model with values in $\mathcal{X}$ and satisfies (2.10) – (2.14), then (10.F) holds with $\mathcal{B} = \mathcal{H}_{\alpha,\beta}(\mathcal{X},\mathbb{C})$ and $\mathcal{C} = \mathcal{H}_\alpha(\mathcal{X},\mathbb{R})$.*

Proof. Let $f \in \mathcal{H}_\alpha(\mathcal{X},\mathbb{R})$ 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 statement (a) of Lemma 2.3, 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 (a) of Lemma 2.3, with the same arguments as in Lemma A.2 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 $P_{f,t}\varphi = \sum_{k=0}^{\infty} P\big((if)^k\varphi\big)t^k/k!$. We infer that (10.F) holds over the space $\mathcal{H}_\alpha(\mathcal{X}, \mathbb{R})$. □

We can now state our sequential finite-dimensional CLT for observables of $(X_i)_{i\in\mathbb{N}^*}$ under (10.A), (10.B), (10.C), and (10.F).

Theorem 10.1 (Sequential Finite-Dimensional CLT)**.** *Suppose that for some $m \in [1, \infty]$, (10.A), (10.B), and (10.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 measurable functions on $\mathcal{X}$ such that $\nu(|f_i|^2) < \infty$ and (10.F) holds for the space $\mathcal{C} = \mathrm{Vect}_\mathbb{R}(f_1, \ldots, f_k)$, the smallest real vector space containing $f_1, \ldots, f_k$. Then, we have*

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

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

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

Proof of Theorem 10.1. Here, we partially follow the lines of the proof of Theorem A in Hennion and Hervé (2001). First let f be a function as in the statement of the theorem. 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

$$\Pi_{f,t}^2 = \Pi_{f,t}, \quad N_{f,t} \circ \Pi_{f,t} = \Pi_{f,t} \circ N_{f,t} = 0, \quad \rho(N_{f,t}) < \theta, \quad |\lambda_{f,t}| \geq \eta \quad \text{for all } t \in I_f,$$

where $\rho(N_{f,t}) := \lim_{n\to\infty} \|N_{f,t}^n\|_{\mathcal{L}(\mathcal{B})}^{1/n}$. 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' and Lemma IV.3 in Hennion and Hervé (2001) the Taylor expansion of $\lambda_{f,t}$ as t goes to 0 is given by

$$\lambda_{f,t} = 1 - \frac{t^2}{2}\sigma_f^2 + o(t^2) \tag{10.1}$$

with

$$\sigma_f^2 := \lim_{n\to\infty} \frac{1}{n}\mathbf{E}\big(S_n(f,1)^2\big) \tag{10.2}$$

These are the main ingredients to derive a CLT for the process $(f(X_i))_{i\geq 0}$. Here we want to show a finite-dimensional sequential CLT. Without loss of generality we will treat the case $k=2$. By the Cramèr-Wold device, it is sufficient to prove the convergence of the real linear combinations $a_1 n^{-\frac{1}{2}} S_n(f_1,t_1) + a_2 n^{-\frac{1}{2}} S_n(f_2,t_2)$ of any square ν-integrable functions $f_1, f_2 \in \mathcal{C}$ to a normal distribution. Since for $t_1 < t_2$, the preceding term is equal to

$$n^{-\frac{1}{2}} S_n(a_1 f_1 + a_2 f_2, t_1) + n^{-\frac{1}{2}} \sum_{i=[nt_1]+1}^{[nt_2]} a_2 f_2(X_i)$$

and $\mathcal{C}$ is a real vector space, it is sufficient to show the convergence of all sums of the form $n^{-\frac{1}{2}} S_n$ with

$$S_n(f,g,s) = \sum_{i=1}^{[ns]} f(X_i) + \sum_{i=[ns]+1}^{n} g(X_i),$$

where $f, g \in \mathcal{C}$, $s \in (0,1)$. So, fix $f, g \in \mathcal{C}$, $s \in (0,1)$ and set $S_n = S_n(f,g,s)$. The following lemma gives us an expression of the corresponding characteristic function.

Lemma 10.1. *For every function $\varphi \in \mathcal{B}$, $t \in \mathbb{R}$, and $n \geq 1$,*

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

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

$$\mathbf{E}\Big(e^{itn^{-\frac{1}{2}}S_n}\Big) = \nu\Bigg(P_{f,\frac{t}{\sqrt{n}}}^{[ns]} P_{g,\frac{t}{\sqrt{n}}}^{n-[ns]} \mathbf{1}_{\mathcal{X}}\Bigg). \tag{10.3}$$

Proof of Lemma 10.1. For every $k \geq 1$ and every measurable function $F : \mathcal{X}^{k-1} \to \mathbb{R}$, we have

$$\begin{aligned}
\mathbf{E}\big(e^{it(F(X_1,\ldots,X_{k-1})+f(X_k))}\varphi(X_k)\big) &= \mathbf{E}\Big(e^{itF(X_1,\ldots,X_{k-1})}\,\mathbf{E}\big(e^{itf(X_k)}\varphi(X_k)|X_{k-1},\ldots,X_1\big)\Big) \\
&= \mathbf{E}\Big(e^{itF(X_1,\ldots,X_{k-1})}\,\mathbf{E}\big(e^{itf(X_k)}\varphi(X_k)|X_{k-1}\big)\Big) \\
&= \mathbf{E}\big(e^{itF(X_1,\ldots,X_{k-1})}P_{f,t}\varphi(X_{k-1})\big)
\end{aligned}$$

and the same equation with g instead of f. The Lemma can now be proved by induction. □

To study the weak convergence of $\frac{S_n}{\sqrt{n}}$, we have to compute the limit of

$$P^{[ns]}_{f,\frac{t}{\sqrt{n}}} P^{n-[ns]}_{g,\frac{t}{\sqrt{n}}} \mathbf{1}_{\mathcal{X}} = \lambda^{[ns]}_{f,\frac{t}{\sqrt{n}}} \lambda^{n-[ns]}_{g,\frac{t}{\sqrt{n}}} \Pi_{f,\frac{t}{\sqrt{n}}} \Pi_{g,\frac{t}{\sqrt{n}}} \mathbf{1}_{\mathcal{X}} + \lambda^{[ns]}_{f,\frac{t}{\sqrt{n}}} \Pi_{f,\frac{t}{\sqrt{n}}} N^{n-[ns]}_{g,\frac{t}{\sqrt{n}}} \mathbf{1}\, 1_{\mathcal{X}} + \lambda^{n-[ns]}_{g,\frac{t}{\sqrt{n}}} N^{[ns]}_{f,\frac{t}{\sqrt{n}}} \Pi_{g,\frac{t}{\sqrt{n}}} \mathbf{1}_{\mathcal{X}} + N^{[ns]}_{f,\frac{t}{\sqrt{n}}} N^{n-[ns]}_{g,\frac{t}{\sqrt{n}}} \mathbf{1}_{\mathcal{X}}\,.$$

By (10.1), we infer that

$$\lambda^{[ns]}_{f,\frac{t}{\sqrt{n}}} \longrightarrow \exp\Big(-\frac{t^2}{2} s\sigma_f^2\Big) \quad \text{and} \quad \lambda^{n-[ns]}_{g,\frac{t}{\sqrt{n}}} \longrightarrow \exp\Big(-\frac{t^2}{2}(1-s)\sigma_g^2\Big) \quad \text{as } n \to \infty,$$

where σ_f and σ_g are given by (10.2). Further, since $\rho(N_{f,t}) < 1$ and $\rho(N_{g,t}) < 1$, we have that $\|N^n_{f,t}\|_{\mathcal{L}(\mathcal{B})} \to 0$ and $\|N^n_{g,t}\|_{\mathcal{L}(\mathcal{B})} \to 0$ uniformly in $t \in I_f \cap I_g$ as $n \to \infty$. By continuity, we also have $\Pi_{f,\frac{t}{\sqrt{n}}} \mathbf{1}_{\mathcal{X}} \to \mathbf{1}_{\mathcal{X}}$ and $\Pi_{g,\frac{t}{\sqrt{n}}} \mathbf{1}_{\mathcal{X}} \to \mathbf{1}_{\mathcal{X}}$ as $n \to \infty$. We therefore obtain

$$\big\|\lambda^{[ns]}_{f,\frac{t}{\sqrt{n}}} \Pi_{f,\frac{t}{\sqrt{n}}} N^{n-[ns]}_{g,\frac{t}{\sqrt{n}}} \mathbf{1}_{\mathcal{X}}\big\|_{\mathcal{B}} \le \big|\lambda^{[ns]}_{f,\frac{t}{\sqrt{n}}}\big| \, \big\|\Pi_{f,\frac{t}{\sqrt{n}}}\big\|_{\mathcal{L}(\mathcal{B})} \big\|N^{n-[ns]}_{g,\frac{t}{\sqrt{n}}}\big\|_{\mathcal{L}(\mathcal{B})} \| \mathbf{1}_{\mathcal{X}} \|_{\mathcal{B}} \longrightarrow 0,$$
$$\big\|\lambda^{n-[ns]}_{g,\frac{t}{\sqrt{n}}} N^{[ns]}_{f,\frac{t}{\sqrt{n}}} \Pi_{g,\frac{t}{\sqrt{n}}} \mathbf{1}_{\mathcal{X}}\big\|_{\mathcal{B}} \le \big|\lambda^{n-[ns]}_{g,\frac{t}{\sqrt{n}}}\big| \, \big\|N^{[ns]}_{f,\frac{t}{\sqrt{n}}}\big\|_{\mathcal{L}(\mathcal{B})} \big\|\Pi_{g,\frac{t}{\sqrt{n}}}\big\|_{\mathcal{L}(\mathcal{B})} \| \mathbf{1}_{\mathcal{X}} \|_{\mathcal{B}} \longrightarrow 0,$$
$$\big\|N^{[ns]}_{f,\frac{t}{\sqrt{n}}} N^{n-[ns]}_{g,\frac{t}{\sqrt{n}}} \mathbf{1}_{\mathcal{X}}\big\|_{\mathcal{B}} \le \big\|N^{[ns]}_{f,\frac{t}{\sqrt{n}}}\big\|_{\mathcal{L}(\mathcal{B})} \big\|N^{n-[ns]}_{g,\frac{t}{\sqrt{n}}}\big\|_{\mathcal{L}(\mathcal{B})} \| \mathbf{1}_{\mathcal{X}} \|_{\mathcal{B}} \longrightarrow 0,$$

and

$$\lambda^{[ns]}_{f,\frac{t}{\sqrt{n}}} \lambda^{n-[ns]}_{g,\frac{t}{\sqrt{n}}} \Pi_{f,\frac{t}{\sqrt{n}}} \Pi_{g,\frac{t}{\sqrt{n}}} \mathbf{1}_{\mathcal{X}} \longrightarrow \exp\Big(-\frac{t^2}{2} s\sigma_f^2\Big) \exp\Big(-\frac{t^2}{2}(1-s)\sigma_g^2\Big) \mathbf{1}_{\mathcal{X}}$$

as $n \to \infty$. Thus we infer

$$P^{[ns]}_{f,\frac{t}{\sqrt{n}}} P^{n-[ns]}_{g,\frac{t}{\sqrt{n}}} \mathbf{1}_{\mathcal{X}} \longrightarrow \exp\Big(-\frac{t^2}{2}\big(s\sigma_f^2 + (1-s)\sigma_g^2\big)\Big) \mathbf{1}_{\mathcal{X}} \quad \textit{as } n \to \infty,$$

which, using (10.3), gives the weak convergence of $n^{-1/2} S_n$ to a centred normal distribution with variance given by $\sigma_{f,g,s} = s\sigma_f^2 + (1-s)\sigma_g^2$. By (10.2), we obtain that Theorem 10.1 holds with the covariance matrix Σ given by

$$\Sigma_{i,j} = \min\{t_i, t_j\} \frac{1}{2}\big(\sigma^2_{f_i+f_j} - \sigma^2_{f_i} - \sigma^2_{f_j}\big). \tag{10.4}$$

Lemma 10.2. *Under the conditions (10.A), (10.B), and (10.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, (10.B), (10.C), and (10.A), we obtain

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

□

The preceding lemma shows that the series $\sum_{k=0}^{\infty} \mathbf{Cov}(f(X_0), f(X_k))$ converges for $f \in \mathcal{B}_{\mathbb{R}} \cap \mathrm{L}^s(\nu)$. Thus, using Kronecker's Lemma, by (10.2) we obtain

$$\sigma_f^2 = \nu(f^2) + 2\sum_{k\geq 1} \mathbf{Cov}(f(X_0), f(X_k)),$$

which, with (10.4), completes the proof of Theorem 10.1. □

As a direct application of Proposition 10.1 to Theorem 10.1, we obtain the following corollary.

Corollary 10.1. *Assume that the Markov chain $(X_i)_{i\in\mathbb{N}^*}$ is an iterative Lipschitz model with values in $\mathcal{X}$ and satisfies (2.10) – (2.14). Then Theorem 10.1 applies for all $\alpha \in (0,1]$ and every choice of finitely many functions $f_1, \dots, f_k \in \mathcal{H}_\alpha(\mathcal{X}, \mathbb{R})$. Thus a sequential finite-dimensional CLT holds for $\mathcal{H}_\alpha(\mathcal{X}, \mathbb{R})$-observables of $(X_i)_{i\in\mathbb{N}}$.*

10.3. SECLTs for $\mathcal{B}$-geometrically Ergodic Markov Chains

We can now apply Theorem 9.2 to the results of Section 10.1. Assumption 9.I can be veryfied with the help of Theorem 10.1 and the multiple mixing property can be established using Lemma 2.2. Moreover, the extra assumptions in Theorem 9.2 concerning the covariance structure of the limit process are a direct consequence of Theorem 10.1 and Lemma 10.2.

Theorem 10.2 (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 Banach space $\mathcal{B}$ satisfies (10.A), (10.B), (10.C), and (10.D). If there is a $\|\cdot\|_\infty$-bounded subset $\mathcal{G} \subset \mathcal{B}_{\mathbb{R}}$ such that (10.F) holds for $\mathcal{C} = \mathrm{Vect}_{\mathbb{R}}(\mathcal{G})$, the smallest real vector space containing $\mathcal{G}$, and such that (10.E) holds with $s = m/(m-1)$, then the sequential empirical process V_n converges in distribution in $\ell^\infty(\mathcal{F} \times [0,1])$ to a centred Gaussian process K with covariance structure given by (9.8).*

As a result of Theorem 10.2, in the situation of iterative Lipschitz models we obtain the following proposition.

Proposition 10.2 (Sequential Empirical CLT for Iterative Lipschitz Models)**.** *Let (2.10) – (2.14) 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 $\mathcal{H}_\alpha(\mathcal{X}, \mathbb{R})$ for some $\alpha < \frac{s-1}{s}$ such that the entropy condition (10.E) holds. Then the $\mathcal{F}$-indexed sequential empirical process $(\mathrm{V}_n(f,t))_{\mathcal{F}\times[0,1]}$ 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 (9.8).*

Proof. As a direct consequence of Proposition 2.2, the condition (10.B), (10.C), and (10.D) are satisfied with $\mathcal{B} = \mathcal{H}_{\alpha,\beta}(\mathcal{X}, \mathbb{C})$. Due to Proposition 2.3, (2.10) – (2.14) yield that (10.A) is satisfied. Now, by choosing $\beta = (s-1)/s < 1/2$, we have $\alpha < \beta$ and thus by Proposition 10.1 condition (10.F) holds 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 $\|g\|_{\mathcal{H}_{\alpha,\beta}} \leq \|g\|_{\mathcal{H}_\alpha}$. Thus, by (9.9), the condition (10.E) is also satisfies with respect to the $\mathcal{H}_{\alpha,\beta}(\mathcal{X}, \mathbb{C})$-norm. □

It should be mentioned that Durieu (2013) established an empirical CLT for $\mathcal{B}$-geometrically Markov chains given by an iterative Lipschitz model, using similar arguments. Of course, in this case the sequential finite-dimensional CLT is not needed and it suffices to use a one dimensional CLT as given by Hennion and Hervé (2001).

10.4. A SECLT for Dynamical Systems with a Spectral Gap

Let us mention that the proof of Theorem 10.2 can be adapted to deal with dynamical systems using the Perron–Frobenius operator in place of the Markov operator. Recall that for a measure preserving transformation T of a probability space $(\mathcal{X}, \mathcal{A}, \nu)$, the corresponding Perron–Frobenius on $L^1(\nu)$ is given by

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

For a function f on $\mathcal{X}$, we define the perturbed operator $P_{f,t}$ by $P_{f,t}\varphi = P(e^{itf}\varphi)$.

We have the following result, for which the proof parallels the one of Theorem 10.2 (using Lemma 2.4 instead of Lemma 2.2) and is therefore omitted here.

Theorem 10.3 (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 an $m \in [1, \infty]$ such that the conditions (10.A), (10.B), (10.C), and (10.D) hold with respect to the Perron–Frobenius operator. If there exists a $\|\cdot\|_\infty$-bounded subset $\mathcal{G} \subset \mathcal{B}_\mathbb{R}$ such that (10.F) holds for the space $\mathcal{C} = \mathrm{Vect}_\mathbb{R}(\mathcal{G})$ and (10.E) holds for $s = \frac{m}{m-1}$, then the process $(U_n(f,t))_{\mathcal{F}\times[0,1]}$ defined by $U_n(f,t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} \left(f \circ T^i - \nu f\right)$ converges in distribution in $\ell^\infty(\mathcal{F} \times [0,1])$ to a centred Gaussian process K with covariance structure given by*

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

11. A SECLT for the Ergodic Automorphism of the Torus

Recall the notation from Chapter 5. In this chapter, we establish a sequential version of Theorem 5.1. In order to apply Theorem 9.2 we need to establish a sequential finite dimensional CLT for $(T^i)_{i\in\mathbb{N}^*}$, the iterates of an ergodic automorphism of the d-dimensional torus $\mathbb{T}^d$. The one dimensional CLT is already known due to Leonov (1960) (cf. Le Borgne (1999)). Here we develop a lemma that extends the finite dimensional CLT to the sequential finite-dimensional CLT under assumptions similar to the multiple mixing property.

Lemma 11.1. *Let $(X_i)_{i\in\mathbb{N}^*}$ be a stationary stochastic process with state space $\mathcal{X}$ and marginal distribution μ. Let $\mathcal{B}$ be a complex Banach space of measurable functions $\mathcal{X}\to\mathbb{C}$ and $\mathcal{B}_\mathbb{R}$ be the subclass of the $\mathbb{R}$-valued functions in $\mathcal{B}$. Suppose that*

(i) $n^{-1/2}\sum_{i=1}^n(f(X_i)-\mu f)\xrightarrow{d}N(0,\sigma_f^2)$ *for all* $f\in\mathcal{B}_\mathbb{R}$,

(ii) for all $f\in\mathcal{B}_\mathbb{R}$ *and* $x\in\mathbb{R}_+$, *we have* $\exp\big(ix(f-\mu f)\big)\in\mathcal{B}$ *and*

$$\sup_{n\in\mathbb{N}^*}\big\|\exp\big(ixn^{-1/2}(f-\mu f)\big)\big\|_\mathcal{B}<\infty,$$

(iii) there is a constant $\theta\in(0,1)$ *and a function* $C:\mathbb{N}^*\longrightarrow\mathbb{R}_+$ *with* $\log(C(n))=o(n)$ *such that for all* $\varphi,\psi\in\mathcal{B}$ *with* $\mu\psi=\mu\varphi=0$ *and* $\|\psi\|_\infty,\|\varphi\|_\infty\le 1$

$$\left|\mathbf{Cov}\left(\prod_{i=1}^q\varphi(X_i)\,,\prod_{i=q+1+k}^{q+p+k}\psi(X_i)\right)\right|\le C(p+q)\,(1\vee\|\varphi\|_\mathcal{B})(1\vee\|\psi\|_\mathcal{B})\theta^k\quad\text{for all } k,p,q\in\mathbb{N}^*.$$

Then for all $t\in[0,1]$, $f,g\in\mathcal{B}_\mathbb{R}$

$$\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{[nt]}(f(X_i)-\mu f)\right)+\frac{1}{\sqrt{n}}\left(\sum_{i=[nt]+1}^{n}(g(X_i)-\mu g)\right)\xrightarrow{d}N(0,t\sigma_f^2+(1-t)\sigma_g^2).$$

Proof. Without loss of generality, assume that C is increasing and that $C(n)\to\infty$ as $n\to\infty$. Let $k(n):=[\log(C(n))/|\log(\theta)|]+1$. Then

$$k(n)=o(n),\tag{11.1}$$
$$C(n)\theta^{2k(n)}\to 0.\tag{11.2}$$

By condition (i), (11.1), and Slutzky's Theorem

$$\frac{1}{\sqrt{n}}\Bigg(\sum_{i=[nt]-k(n)+1}^{[nt]}(f(X_i)-\mu f)\Bigg)+\frac{1}{\sqrt{n}}\Bigg(\sum_{i=[nt]+1}^{[nt]+k(n)}(g(X_i)-\mu g)\Bigg)$$
$$=\sqrt{\frac{k(n)}{n}}\Bigg\{\Bigg(\frac{1}{\sqrt{k(n)}}\sum_{i=[nt]-k(n)+1}^{[nt]}(f(X_i)-\mu f)\Bigg)+\Bigg(\frac{1}{\sqrt{k(n)}}\sum_{i=[nt]+1}^{[nt]+k(n)}(g(X_i)-\mu g)\Bigg)\Bigg\}\xrightarrow{\mathbf{P}}0.$$

Therefore it is sufficient to show that

$$Y_n:=\frac{1}{\sqrt{n}}\Bigg(\sum_{i=1}^{[nt]-k(n)}(f(X_i)-\mu f)\Bigg)+\frac{1}{\sqrt{n}}\Bigg(\sum_{i=[nt]+k(n)+1}^{n}(g(X_i)-\mu g)\Bigg)$$
$$\xrightarrow{d}N(0,t\sigma_f^2+(1-t)\sigma_g^2).$$

Now apply Lévy's continuity theorem. Denote the characteristic function of a random variable Y by Φ_Y. We have

$$\begin{aligned}\Phi_{Y_n}(x)&=\mathbf{E}\Bigg[\exp\Bigg(\frac{ix}{\sqrt{n}}\sum_{i=1}^{[nt]-k(n)}(f(X_i)-\mu f)\Bigg)\exp\Bigg(\frac{ix}{\sqrt{n}}\sum_{i=[nt]+k(n)+1}^{n}(g(X_i)-\mu g)\Bigg)\Bigg]\\&=\mathbf{E}\Bigg[\exp\Bigg(\frac{ix}{\sqrt{n}}\sum_{i=1}^{[nt]-k(n)}(f(X_i)-\mu f)\Bigg)\Bigg]\mathbf{E}\Bigg[\exp\Bigg(\frac{ix}{\sqrt{n}}\sum_{i=[nt]+k(n)+1}^{n}(g(X_i)-\mu g)\Bigg)\Bigg]\\&\quad+\mathbf{Cov}\Bigg(\prod_{i=1}^{[nt-k(n)]}\exp\Big(\frac{ix}{\sqrt{n}}(f(X_i)-\mu f)\Big),\prod_{i=[nt]+k(n)+1}^{n}\exp\Big(\frac{ix}{\sqrt{n}}(g(X_i)-\mu g)\Big)\Bigg)\end{aligned}\tag{11.3}$$

By condition (ii) and (iii) we have

$$\begin{aligned}&\mathbf{Cov}\Bigg(\prod_{i=1}^{[nt-k(n)]}\exp\Big(\frac{ix}{\sqrt{n}}(f(X_i)-\mu f)\Big),\prod_{i=[nt]+k(n)+1}^{n}\exp\Big(\frac{ix}{\sqrt{n}}(g(X_i)-\mu g)\Big)\Bigg)\\&\le C(n)\theta^{2k(n)}\Big(1\vee\big\|\exp\big(ixn^{-\frac{1}{2}}(f-\mu f)\big)\big\|_{\mathcal{B}}\Big)\\&\quad\cdot\Big(1\vee\big\|\exp\big(ixn^{-\frac{1}{2}}(g-\mu g)\big)\big\|_{\mathcal{B}}\Big)\longrightarrow0\quad\text{as }n\to\infty,\end{aligned}\tag{11.4}$$

where we used (11.2) and (ii) in the last step. Further, by stationarity and condition (i)

$$\mathbf{E}\Bigg[\exp\Bigg(\frac{ix}{\sqrt{n}}\sum_{i=1}^{[nt]-k(n)}(f(X_i)-\mu f)\Bigg)\Bigg]\longrightarrow\Phi_{N(0,t\sigma_f^2)}(x)\quad\text{as }n\to\infty$$

and

$$\mathbf{E}\Bigg[\exp\Bigg(\frac{ix}{\sqrt{n}}\sum_{i=[nt]+k(n)+1}^{n}(g(X_i)-\mu g)\Bigg)\Bigg]\longrightarrow\Phi_{N(0,(1-t)\sigma_g^2)}(x)\quad\text{as }n\to\infty.$$

With (11.3) and (11.4) this implies

$$\Phi_{Y_n}(x) \longrightarrow \Phi_{N(0,t\sigma_f^2+(1-t)\sigma_g^2)}(x) \quad \text{as } n \to \infty,$$

which, as a consequence of Lévy's continuity theorem, completes the proof. □

For $\mathcal{B} = \mathcal{H}_\alpha(\mathcal{X}, \mathbb{C})$, assumption (ii) of Lemma 11.1 is satisfied. To see this, let $f \in \mathcal{G}$, set $g = f - \mu f$ and recall that $|\exp(iz) - \exp(iy)| \leq |z - y|$ for all $y, z \in \mathbb{R}$. We therefore have that for $x \in \mathbb{R}_+$

$$\begin{aligned}\sup_{\substack{a,b\in\mathcal{X}\\ a\neq b}} \frac{\big|\exp\big(ixn^{-\frac{1}{2}}g(b)\big) - \exp\big(ixn^{-\frac{1}{2}}g(a)\big)\big|}{|b-a|^\alpha} &\leq \sup_{\substack{a,b\in\mathcal{X}\\ a\neq b}} \frac{\big|xn^{-\frac{1}{2}}g(b) - xn^{-\frac{1}{2}}g(a)\big|}{|b-a|^\alpha} \\ &\leq xn^{-\frac{1}{2}}\, m_\alpha(g)\end{aligned}$$

and further $\|\exp\big(ixn^{-1/2}g\big)\|_\infty = 1$, which implies $\exp\big(ixg\big) \in \mathcal{B}$ and

$$\|\exp\big(ixn^{-\frac{1}{2}}g\big)\|_{\mathcal{H}_\alpha} \leq 1 + xn^{-\frac{1}{2}}\, m_\alpha(f).$$

In the case of an ergodic automorphism of the torus, condition (iii) with $\mathcal{B} = \mathcal{H}_\alpha(\mathbb{T}^d, \mathbb{C})$ is a direct consequence of Lemma A.1. As aforementioned, (i) holds due to a result of Leonov (1960) with $\mathcal{B}_\mathbb{R} = \mathcal{H}_\alpha(\mathbb{T}^d, \mathbb{R})$. Therefore Lemma 11.1 applies and we obtain that Assumption 9.I holds with $\mathcal{C} = \mathcal{H}_\alpha(\mathbb{T}^d, \mathbb{R})$ for any $\alpha \in (0, 1]$. Now, let $(\mathbb{T}^d, \mathfrak{B}(\mathbb{T}^d), \lambda, T)$ be the dynamical system of an ergodic automorphism of the torus as introduced in Section 2.6. Following the arguments from Chapter 5 and applying Theorem 9.2 we obtain the following result.

Theorem 11.1 (Sequential Empirical CLT for Ergodic Automorphisms of the Torus)**.** *Let $\mathcal{F}$ be a uniformly bounded class of functions on $\mathbb{R}^\ell$, $\ell \in \mathbb{N}^*$, $\varphi \in \mathcal{H}_\beta(\mathbb{T}^d, \mathbb{R}^\ell)$, $\beta \in (0, 1]$, and let d_0 denote the size of the biggest Jordan block of T restricted to its neutral subspace. If the entropy condition (9.9) holds with $\mu = \lambda \circ \varphi^{-1}$ and $s = 1$ for some uniformly bounded subset $\mathcal{G}$ of $\mathcal{H}_\alpha(\mathbb{R}^\ell, \mathbb{R})$ with $\alpha \in (0, 1]$, $r \geq -1$, $C > 0$ and $\gamma > \max\{1, d_0\}$, then the empirical process $\mathrm{V}_n = (\mathrm{V}_n(f))_{(f,t)\in\mathcal{F}\times[0,1]}$ given by*

$$\mathrm{V}_n(f, t) := \frac{1}{\sqrt{n}}\Bigg(\sum_{i=1}^{[nt]} f \circ \varphi(T^i) - \lambda(f \circ \varphi)\Bigg)$$

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

Remark 11.1. Note that corresponding remarks to Theorem 5.1, Lemma 5.1 and Corollary 5.1 also hold for the sequential case. In particular, the sequential empirical CLT holds for all the classes of functions listed in Corollary 5.1. Note, that in the corresponding situation of

Lemma 5.1, here the covariance structure is given by

$$\begin{aligned}&\mathbf{Cov}\big(K(f,t),K(g,u)\big)\\&\quad=\min\{t,u\}\Big\{\sum_{k=0}^{\infty}\mathbf{Cov}\big(f(\varphi),g(\varphi(T^k))\big)+\sum_{k=1}^{\infty}\mathbf{Cov}\big(f(\varphi(T^k)),g(\varphi)\big)\Big\}.\end{aligned}$$

Remark 11.2. Dedecker et al. (2013) proved a sequential empirical for the ergodic automorphism of the torus where they restrict to the indexing class $\mathcal{F} := \{\mathbf{1}_{(-\infty,x]} : x \in \mathbb{R}^\ell\}$. They also consider a process $(\varphi(T^i))_{i\in\mathbb{N}^*}$ for some fixed function φ. The assumptions differ from ours in two points: In their approach, they require that the distribution function F of $\lambda \circ \varphi^{-1}$ in α-Hölder for some $\alpha \in (0,1]$, where we only demand that $\omega_{\mathrm{F}}(x) \leq |\log(x)|^{-\gamma}$ for $\gamma > \max\{1, d_0\}$ (cf. (i) in Corollary 5.1). Vice versa, we need that φ is α-Hölder for some $\alpha \in (0,1]$, while they only require that F satisfies $\omega_\varphi(x) \leq |\log(x)|^{-a}$ for a specific $a > 1$ that depends of ℓ and α (for instance for $\ell = \alpha = 1$ they require $a > 10/3$). However, it should be mentioned, that we also require that F is Hölder[1] if we want to identify the covariance structure of the limit process K.

[1] in this situation, the exponent of the Hölder condition can be arbitrary chosen in $(0,1]$ and has no influence on the other conditions (cf. Corollary 5.1).

12. Statistical Applications of Sequential Empirical CLTs

As discussed in the introduction, sequential empirical CLTs find application in the study of change point tests. Here, we consider the generalized setting, where the sequential empirical process is indexed by a arbitrary uniformly bounded class $\mathcal{F}$ of real-valued measurable functions on the state space of our data. Recall that we provide results for the case that $\mathcal{F}$ is the space of indicators of semi-finite and finite rectangles, bounded ellipsoids, balls of arbitrary metric and some specific parametric class of monotone functions (cf. Chapter 4).

If we want to test the hypothesis that $(X_i)_{i\in\mathbb{N}^*}$ is stationary with marginal distribution μ against the alternative that there is a $k^* \in \mathbb{N}^*$ such that $(X_1, \ldots, X_{k^*})$ and $(X_{k^*+1}, \ldots, X_n)$ are both stationary and have a different marginal distribution, we can use the test statistic

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

Under the hypothesis, the asymptotic behaviour can be derived from the limit distribution of the sequential empirical process if applicable. We have the following theorem.

Proposition 12.1. *If $(X_i)_{i\in\mathbb{N}^*}$ satisfies the sequential empirical CLT with limit distribution K, then under the null hypothesis $\mathbf{H}_0$ we have the convergence*

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

To prepare the proof of Proposition 12.1, consider the natural generalization of the process R_n introduced in Chapter 1. Let $(X_i)_{i\in\mathbb{N}}$ be an $\mathcal{X}$-valued stationary process with empirical measure $\mu_n(f) := n^{-1}\sum_{i=1}^n f(X_i)$, $n \in \mathbb{N}^*$. We set $\mu_0(f) = 0$. For $j \in \{1, \ldots, n\}$ we define $\mu_{j,n}(f) := (n-j+1)^{-1}\sum_{i=j}^n f(X_i)$ and set $\mu_{n+1,n}(f) := 0$. 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

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

The following lemma gives the asymptotic distribution of R_n.

Lemma 12.1. *Assume that $(X_i)_{i\in\mathbb{N}}$ satisfies the sequential empirical CLT with indexing class $\mathcal{F}$ and limit process K, that is, $\mathrm{V}_n \xrightarrow{d} K$ in $\ell^\infty(\mathcal{F}\times[0,1])$ as $n\to\infty$, where K denotes a tight centred Gaussian process. Then*

$$R_n \xrightarrow{d} (K(f,t) - tK(f,1))_{(f,t)\in\mathcal{F}\times[0,1]} \quad \textit{in } \ell^\infty(\mathcal{F}\times[0,1]) \textit{ to as } n\to\infty.$$

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

$$\begin{aligned}
&\mu_{[nt]}(f) - \mu_{[nt]+1,n}(f) \\
&= \frac{1}{[nt]}\sum_{i=1}^{[nt]} f(X_i) - \frac{1}{n-[nt]}\sum_{i=[nt]+1}^{n} f(X_i) \\
&= \frac{1}{[nt]}\sum_{i=1}^{[nt]} (f(X_i) - \mu f) - \frac{1}{n-[nt]}\sum_{i=[nt]+1}^{n} (f(X_i) - \mu f) \\
&= \left(\frac{1}{[nt]} + \frac{1}{n-[nt]}\right)\sum_{i=1}^{[nt]} (f(X_i) - \mu f) - \frac{1}{n-[nt]}\sum_{i=1}^{n} (f(X_i) - \mu f) \\
&= \frac{1}{\sqrt{n}}\frac{n}{[nt]}\frac{n}{n-[nt]}\,\mathrm{V}_n(f,t) - \frac{1}{\sqrt{n}}\frac{1}{t}\frac{n}{n-[nt]}\,t\,\mathrm{V}_n(f,1). \qquad (12.1)
\end{aligned}$$

Further, by definition we have $R_n(f,1) = 0$ and $R_n(f,t) = 0$ for $t \in [0, 1/n)$. Since also $\mathrm{V}_n(f,t) = 0$ for $t \in [0,1/n)$, we obtain with (12.1) that

$$\begin{aligned}
R_n(f,t) &= \mathrm{V}_n(f,t) - \frac{[nt]}{n}\,\mathrm{V}_n(f,1), \\
&= \mathrm{V}_n(f,t) - t\,\mathrm{V}_n(f,1) + \frac{nt-[nt]}{n}\,\mathrm{V}_n(f,1) \quad \text{for all } t \in [0,1]. \qquad (12.2)
\end{aligned}$$

Let A_n denote the $\mathcal{F}\times[0,1]$-indexed processes given by $A_n(f,t) := \big((nt-[nt])/n\big)\,\mathrm{V}_n(f,t)$. 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 (12.2) yields

$$R_n = \big(\mathrm{V}_n(f,t) - t\,\mathrm{V}_n(f,1)\big)_{(f,t)\in\mathcal{F}\times[0,1]} + A_n \xrightarrow{d} \big(K(f,t) - tK(f,1)\big)_{(f,t)\in\mathcal{F}\times[0,1]}.$$

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

Remark 12.1. Note that, in the setting of Theorem 10.2 and Theorem 10.3 and for a wide class of multiple mixing processes (cf. Theorem 9.2), including ergodic automorphisms of the torus (see Theorem 11.1), the covariance structure of K is given by (9.8).

Proof of Proposition 12.1. $R_n(f,\cdot)$ is obviously constant on the intervals $\big[k/n, (k+1)/n\big)$, $k = 0,\dots,n-1$ and further $R_n\big(f,k/n\big) = k/n(1-k/n)\sqrt{n}\big(\mu_k(f) - \mu_{k+1,n}(f)\big)$ for $k = 0,\dots,n$. 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

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

□

A. Appendix

A.1. A CLT of Dedecker (1998)

The following proposition corresponds to a particular case of Corollary 1 in Dedecker (1998).

Proposition A.1 (Dedecker (1998))**.** *Let* $(Y_i)_{i\in\mathbb{N}}$ *be an ergodic stationary process with* $\mathbf{E}(Y_0) = 0$ *and* $\mathbf{E}(Y_0^2) < \infty$*, which is adapted to a filtration* $(\mathcal{M}_i)_{i\in\mathbb{N}}$*. If* $\sum_{i=0}^{n} Y_0\, \mathbf{E}(Y_i|\mathcal{M}_0)$ *converges in* L_1*, then*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i \xrightarrow{d} N(0,\sigma^2) \quad \text{as } n\to\infty,$$

where $\sigma^2 = \mathbf{Var}(Y_0^2) + 2\sum_{i=1}^{\infty}\mathbf{Cov}(Y_0,Y_i) < \infty$.

A.2. Some Covariance Inequalities

The following proposition corresponds to Proposition 3 in Dehling and Durieu (2011) .

Proposition A.2 (Dehling and Durieu (2011))**.** *There exist* $C > 0$*,* $0 < \theta < 1$*, for all* $m,p\in\mathbb{N}^*$*, such that for all bounded* α*-Hölder functions* φ *(*$\alpha\in(0,1]$*) with* $\|\varphi\|_\infty \le 1$*, for all* $k_1 \le \ldots \le k_m \le 0 \le l_1 \le \ldots \le l_p$*, for all* $n\in\mathbb{N}$*,*

$$\left|\mathbf{Cov}\Big(\prod_{j=1}^{m}\varphi\circ T^{k_j}\,,\,\prod_{j=1}^{p}\varphi\circ T^{l_j+n}\Big)\right| \le C\,\|\varphi\|_1\,\|\varphi\|_{\mathcal{H}_\alpha}\,Q(k_1,\ldots,k_m)\theta^n$$

where $Q(k_1,\ldots,k_m) = \sum_{i=1}^{m}|k_i|^{d_0}$ *with* d_0 *the size of the biggest Jordan block of* T *restricted to its neutral subspace.*

Following Lemma 1.3.1. in Le Borgne and Pène (2005), Durieu (2008a) established the following covariance inequality.

Lemma A.1. *Let* T *be an ergodic automorphism of the* d*-dimensional torus* $\mathbb{T}^d$*,* $d\ge 2$*, equipped with the Lebesgue measure. Then there exist constants* $\theta\in(0,1)$*,* $C>0$*, and* $c\in\mathbb{N}^*$ *such that for all* $\varphi,\psi\in\mathcal{H}_\alpha(\mathbb{T}^d,\mathbb{C})$ *with* $\|\varphi\|_\infty, \|\psi\|_\infty \le 1$

$$\mathbf{Cov}\Big(\prod_{i=1}^{q}\varphi\circ T^i\,,\prod_{i=q+k+1}^{q+k+p}\psi\circ T^i\Big) \le Cpq^c\,\|\varphi\|_{\mathcal{H}_\alpha}\,\|\psi\|_{\mathcal{H}_\alpha}\theta^k \quad \text{for all } k,p,q\in\mathbb{N}^*.$$

□

A.3. Lemma A.2

Lemma A.2. *Let $\mathcal{B}$ be a Banach algebra of functions on a space $\mathcal{X}$ and P be a bounded linear operator on $\mathcal{B}$. If $\mathcal{B}$ satisfies (2.B), then for all $f \in \mathcal{B}$ the mapping $a_f : \mathbb{R} \longrightarrow \mathcal{L}(\mathcal{B})$ given by $t \mapsto P_{f,t}$ is analytic and has a representation*

$$a_f(t) = \sum_{k=0}^{\infty} \frac{(it)^k}{k!} P_{(k)},$$

where $P_{(k)}$ is given by $P_{(k)}\varphi := P(f^k\varphi)$.

Proof. First, we establish pointwise convergence, i.e. convergence of $a_n(t)(\varphi)$ for every $\varphi \in \mathcal{B}$. By the Banach algebra property, $\sum_{k=1}^{n}(itf)^k/k!$ converges in $\mathcal{B}$ to e^{itf} as $n \to \infty$ and therefore

$$P_{f,t}\varphi = P(e^{itf}\varphi) = P\Big(\sum_{k=0}^{\infty} \frac{(it)^k}{k!} f^k\varphi\Big).$$

Recall that a bounded linear operator is always continuous and thus

$$P\Big(\sum_{k=0}^{\infty} \frac{(it)^k}{k!} f^k\varphi\Big) = \lim_{n\to\infty} P\Big(\sum_{k=0}^{n} \frac{(it)^k}{k!} f^k\varphi\Big) = \lim_{n\to\infty} \sum_{k=0}^{n} \frac{(it)^k}{k!} P_{(k)}\varphi,$$

which gives us the pointwise convergence.

By condition (2.B), the inequality $\|P_{(k)}\|_{\mathcal{L}(\mathcal{B})} \leq \|f\|_{\mathcal{B}}^k \|P\|_{\mathcal{L}(\mathcal{B})}$ holds and thus $\sum_{k=0}^{n}(it)^k P_{(k)}/k!$ converges in $\mathcal{L}(\mathcal{B})$ as $n \to \infty$. Convergence in operator norm implies pointwise convergence, which yields

$$a_f(t)(\varphi) = \lim_{n\to\infty} \sum_{k=0}^{n} \frac{(it)^k}{k!} P_{(k)}\varphi = \Big(\sum_{k=0}^{\infty} \frac{(it)^k}{k!} P_{(k)}\Big)\varphi \quad \text{for all } \varphi \in \mathcal{B}.$$

□

A.4. Proof of Theorem 1.1

We first show that $\xi^{(q)}$ converges in distribution to some random variable ξ. We denote by $L^{(q)}$ the distribution of $\xi^{(q)}$; this is defined since $\xi^{(q)}$ is measurable. Moreover, $L^{(q)}$ is a separable Borel probability measure on S.

By Theorem 1.12.4 of van der Vaart and Wellner (1996), weak convergence of separable Borel measures on a metric space S can be metrised by the bounded Lipschitz metric, defined by

$$d_{BL_1}(L_1, L_2) = \sup_{f\in BL_1} \left| \int f(x)dL_1(x) - \int f(x)dL_2(x) \right|,$$

for any Borel measures L_1, L_2 on S. Here, $BL_1 := \{f : S \longrightarrow \mathbb{R} : \|f\|_{BL_1} \le 1\}$, where

$$\|f\|_{BL_1} := \max\left\{\sup_{x\in S}|f(x)|,\ \sup_{x\neq y\in S}\frac{f(x)-f(y)}{\rho(x,y)}\right\}.$$

In addition, the theorem states that the space of all separable Borel measures on a complete space is complete with respect to the bounded Lipschitz metric. Thus it suffices to show that $L^{(q)}$ is a d_{BL_1}-Cauchy sequence. We obtain

$$\begin{aligned} d_{BL_1}\big(L^{(q)}, L^{(r)}\big) &= \sup_{f\in BL_1} |\,\mathbf{E}\, f(\xi^{(q)}) - \mathbf{E}\, f(\xi^{(r)})| \\ &\le \sup_{f\in BL_1} \Big\{|\,\mathbf{E}\, f(\xi^{(q)}) - \mathbf{E}^*\, f(\xi_n^{(q)})| + |\,\mathbf{E}^*\, f(\xi_n^{(q)}) - \mathbf{E}^*\, f(\xi_n)| \\ &\qquad + |\,\mathbf{E}^*\, f(\xi_n) - \mathbf{E}^*\, f(\xi_n^{(r)})| + |\,\mathbf{E}^*\, f(\xi_n^{(r)}) - \mathbf{E}\, f(\xi^{(r)})|\Big\} \end{aligned}$$

for all $n \in \mathbb{N}^*$. For a Borel measurable separable random element $\xi^{(q)}$ weak convergence $\xi_n^{(q)} \xrightarrow{d} \xi^{(q)}$ as $n \to \infty$ is equivalent to $\sup_{f\in BL_1} |\,\mathbf{E}\, f(\xi^{(q)}) - \mathbf{E}^*\, f(\xi_n^{(q)})| \longrightarrow 0$; see van der Vaart and Wellner (1996, p.73). Hence by (1.10) we obtain

$$d_{BL_1}\big(L^{(q)}, L^{(r)}\big) \le \liminf_{n\to\infty} \sup_{f\in BL_1} |\,\mathbf{E}^*\, f(\xi_n^{(q)}) - \mathbf{E}^*\, f(\xi_n)| + |\,\mathbf{E}^*\, f(\xi_n) - \mathbf{E}^*\, f(\xi_n^{(r)})|.$$

Using Lemma 1.2.2 (iii) in van der Vaart and Wellner (1996), we obtain

$$|\,\mathbf{E}^*\, f(\xi_n^{(q)}) - \mathbf{E}^*\, f(\xi_n)| \le \mathbf{E}(|f(\xi_n) - f(\xi_n^{(q)})|^*)$$

and therefore

$$\begin{aligned} \sup_{f\in BL_1} |\,\mathbf{E}^*\, f(\xi_n^{(q)}) - \mathbf{E}^*\, f(\xi_n)| &\le \mathbf{E}\big(\rho(\xi_n, \xi_n^{(q)}) \wedge 2\big)^* \\ &= \int_0^\infty \mathbf{P}^*\big(\rho(\xi_n, \xi_n^{(q)}) \wedge 2 \ge t\big)\, dt, \end{aligned} \tag{A.1}$$

where we used the last statement of Lemma 1.2.2 in van der Vaart and Wellner (1996). Now, let $\varepsilon > 0$ be given. By (1.11), there exists an $q_0 \in \mathbb{N}^*$ such that for every $q \ge q_0$ there is some $n_0 \in \mathbb{N}^*$ such that for every $n \ge n_0$ we have $\mathbf{P}^*\big(\rho(\xi_n, \xi_n^{(q)}) \ge \varepsilon/3\big) \le \varepsilon/3$. Therefore

$$\mathbf{P}^*\Big(\rho(\xi_n, \xi_n^{(q)}) \wedge 2 \ge t\Big) \le \begin{cases} 1, & \text{if } t < \frac{\varepsilon}{3} \\ \frac{\varepsilon}{3}, & \text{if } \frac{\varepsilon}{3} \le t \le 2 \\ 0, & \text{if } 2 < t. \end{cases}$$

Applying this inequality to (A.1), we obtain

$$\liminf_{n\to\infty} \sup_{f\in BL_1} |\,\mathbf{E}^*\, f(\xi_n^{(q)}) - \mathbf{E}^*\, f(\xi_n)| \le \int_0^2 \frac{\varepsilon}{3} + 1_{\{t<\frac{\varepsilon}{3}\}}\, dt = \varepsilon$$

for all $q \geq q_0$. Hence for $q, r \geq q_0$ we have $d_{BL_1}(L^{(q)}, L^{(r)}) \leq 2\varepsilon$; i.e. $(L^{(q)})_{q\in\mathbb{N}^*}$ is a d_{BL_1}-Cauchy sequence in a complete metric space.

The remaining part of the proof follows closely the proof of Theorem 4.2 in Billingsley (1968), replacing the probability measure $\mathbf{P}$ by the outer measure $\mathbf{P}^*$ where necessary and making use of the Portmanteau theorem; see van der Vaart and Wellner (1996), Theorem 1.3.4 (iii), and the sub-additivity of outer measures. From part (i), we already know that there is some measurable ξ such that $\xi^{(q)} \xrightarrow{d} \xi$. Let $F \subset S$ be closed. Given $\varepsilon > 0$, we define the ε-neighbourhood $F_\varepsilon := \{s \in S\ : \inf_{x\in F} \rho(s, x) \leq \varepsilon\}$, and observe that F_ε is also closed. Since $\{\xi_n \in F\} \subset \{\xi_n^{(q)} \in F_\varepsilon\} \cup \{\rho(\xi_n^{(q)}, \xi_n) \geq \varepsilon\}$, we obtain

$$\mathbf{P}^*(\xi_n \in F) \leq \mathbf{P}^*(\xi_n^{(q)} \in F_\varepsilon) + \mathbf{P}^*(\rho(\xi_n^{(q)}, \xi_n) \geq \varepsilon),$$

for all $q \in \mathbb{N}^*$. By (1.11) we may choose q_0 so large that for all $q \geq q_0$

$$\limsup_{n\to\infty} \mathbf{P}^*(\rho(\xi_n^{(q)}, \xi_n) \geq \varepsilon) \leq \varepsilon/2.$$

As $\xi^{(q)} \xrightarrow{d} \xi$, by the Portmanteau theorem we may choose q_1 so large that for all $q \geq q_1$

$$\mathbf{P}(\xi^{(q)} \in F_\varepsilon) \leq \mathbf{P}(\xi \in F_\varepsilon) + \varepsilon/2.$$

We now fix $q \geq \max(q_0, q_1)$. By (1.10) we have $\xi_n^{(q)} \xrightarrow{d} \xi^{(q)}$ as $n \to \infty$. Thus an application of the Portmanteau theorem yields

$$\limsup_{n\to\infty} \mathbf{P}^*(\xi_n^{(q)} \in F_\varepsilon) \leq \mathbf{P}(\xi^{(q)} \in F_\varepsilon),$$

$$\limsup_{n\to\infty} \mathbf{P}^*(\xi_n \in F) \leq \mathbf{P}(\xi \in F_\varepsilon) + \varepsilon.$$

Since this holds for any $\varepsilon > 0$ and $\lim_{\varepsilon\to 0} \mathbf{P}(\xi \in F_\varepsilon) = \mathbf{P}(\xi \in F)$, we get

$$\limsup_{n\to\infty} \mathbf{P}^*(\xi_n \in F) \leq \mathbf{P}(\xi \in F),$$

for all closed sets $F \subset S$. By a final application of the Portmanteau theorem we infer $\xi_n \xrightarrow{d} \xi$.

A.5. Proof of Lemma 2.3

(a) For $f \in \mathcal{H}_\alpha(\mathcal{X}, \mathbb{C})$ and $g \in \mathcal{H}_{\alpha,\beta}(X, \mathbb{C})$ we have $N_\beta(fg) \leq \|f\|_\infty N_\beta(g)$ and

$$\begin{aligned} m_{\alpha,\beta}(fg) &\leq \sup_{\substack{x,y\in\mathcal{X}\\ x\neq y}} \left(|f(x)| \frac{|g(x) - g(y)|}{d(x,y)^\alpha (1 + d(x,x_0)^\beta)} \right) + \left(\frac{|f(x) - f(y)|}{d(x,y)^\alpha} \frac{|g(y)|}{1 + d(x,x_0)^\beta} \right) \\ &\leq \|f\|_\infty m_{\alpha,\beta}(g) + \|f\|_{\mathcal{H}_\alpha} N_\beta(g). \end{aligned}$$

Thus $\|fg\|_{\mathcal{H}_{\alpha,\beta}} = N_\beta(fg) + m_{\alpha,\beta}(fg) \leq \|f\|_{\mathcal{H}_\alpha} N_\beta(g) + \|f\|_\infty m_{\alpha,\beta}(g) \leq \|f\|_{\mathcal{H}_\alpha} \|g\|_{\mathcal{H}_{\alpha,\beta}}$.

(b) First, we show that $(\mathcal{H}_{\alpha,\beta}(\mathcal{X},\mathbb{C}), \|\cdot\|_{\mathcal{H}_{\alpha,\beta}})$ is complete. Let f_n be a Cauchy sequence in $\mathcal{H}_{\alpha,\beta}(\mathcal{X},\mathbb{C})$. Then $N_\beta(f-f_n) \longrightarrow 0$ and $m_{\alpha,\beta}(f-f_n) \longrightarrow 0$ as $n \to \infty$. $N_\beta(f-f_n) \longrightarrow 0$ implies that there is a function $f : \mathcal{X} \longrightarrow \mathbb{C}$ such that $f_n(x) \longrightarrow f(x)$ as $n \to \infty$ for every $x \in \mathcal{X}$. We have

$$N_\beta(f) = \sup_{x\in\mathcal{X}} \lim_{n\to\infty} \frac{|f_n(x)|}{1+d(x,x_0)^\beta} \le \sup_{x\in\mathcal{X}} \limsup_{n\to\infty} \frac{|f_n(x)|}{1+d(x,x_0)^\beta} \le \limsup_{n\to\infty} N_\beta(f-f_n) < \infty$$

and similarly $m_{\alpha,\beta}(f) < \infty$. This implies that $\|f\|_{\mathcal{H}_{\alpha,\beta}} < \infty$ and thus $f \in \mathcal{H}_{\alpha,\beta}(\mathcal{X},\mathbb{C})$. Further since f_n is a Cauchy sequence, we have

$$\begin{aligned} N_\beta(f-f_n) &= \sup_{x\in\mathcal{X}} \lim_{m\to\infty} \frac{|f_m(x)-f_n(x)|}{1+d(x,x_0)^\beta} \\ &\le \limsup_{m\to\infty} N_\beta(f_m-f_n) \le \sup_{m\ge n} N_\beta(f_m-f_n) \longrightarrow 0 \quad \text{as } n\to\infty \end{aligned}$$

and similarly $m_{\alpha,\beta}(f-f_n) \longrightarrow 0$ as $n \to \infty$. Therefore $f_n \to f$ in $(\mathcal{H}_{\alpha,\beta}(\mathcal{X},\mathbb{C}), \|\cdot\|_{\mathcal{H}_{\alpha,\beta}})$. The space $\mathcal{H}_{\alpha,\beta}(\mathcal{X},\mathbb{C})$ is obviously closed under composition with the modulus or conjugation functional on $\mathbb{C}$. Finally, the continuity of $f \mapsto f(x)$ is a direct consequence of the pointwise convergence of any sequence f_n that converges in $(\mathcal{H}_{\alpha,\beta}(\mathcal{X},\mathbb{C}), \|\cdot\|_{\mathcal{H}_{\alpha,\beta}})$ and thus (2.B) holds.

(c) The statement is trivial for $\beta = 0$. Recall that by assumption, the fist moment of ν exists. Let $\beta \in (0,1]$ then $|f| \le (1+d(\cdot,x_0))^\beta N_\beta(f)$ and thus

$$\|f\|_{1/\beta} \le \left\|1+d(x,x_0)^\beta\right\|_{1/\beta} N_\beta(f) \le \Big(1+\big(\nu(d(\cdot,x_0))\big)^\beta\Big) N_\beta(f).$$

(d) Let $f, g \in \mathrm{L}^\infty(\nu)$. Then

$$\begin{aligned} \|fg\|_{\mathcal{H}_{\alpha,\beta}} &\le \|f\|_\infty N_\beta(g) + \sup_{\substack{x\in\mathcal{X}\\ x\ne y}} |f(x)| \frac{|g(x)-g(y)|}{d(x,y)^\alpha(1+d(x,x_0))^\beta} + |g(y)| \frac{|f(x)-f(y)|}{d(x,y)^\alpha(1+d(x,x_0))^\beta} \\ &\le \|f\|_\infty N_\beta(g) + \|f\|_\infty m_{\alpha,\beta}(g) + \|g\|_\infty m_{\alpha,\beta}(f) \end{aligned}$$

and thus $\|fg\|_{\mathcal{H}_{\alpha,\beta}} \le \|f\|_\infty \|g\|_{\mathcal{H}_{\alpha,\beta}} + \|g\|_\infty \|f\|_{\mathcal{H}_{\alpha,\beta}}$. □

A.6. Proof of Lemma 4.2

Without loss of generality, assume that $x = 0$. For $v \in \mathbb{R}^d$, let D_v denote the diagonal $d\times d$-matrix with diagonal entries $v_1, \ldots, v_d$. We define the operator norm of the $d\times d$-matrix A by $|A|_* := \sup_{y\in\mathbb{R}^d\setminus\{0\}} |Ay|/|y|$. Observe that $|\mathrm{D}_v|_* = \max_{i=1,\ldots,d} |v_i|$. We can characterize $\mathrm{E}(0, \frac{j}{m})$ and $\mathbb{R}^d \setminus \mathrm{E}(0, \frac{j}{m} + \frac{1}{m})$ by

$$\mathrm{E}\left(0, \frac{j}{m}\right) = \left\{ z \in \mathbb{R}^d : \left|\mathrm{D}^{-1}_{\frac{j}{m}} z\right| \le 1 \right\} \quad \text{and} \quad \mathbb{R}^d \setminus \mathrm{E}\left(0, \frac{j}{m} + \frac{1}{m}\right) = \left\{ y \in \mathbb{R}^d : \left|\mathrm{D}^{-1}_{\frac{j}{m}+\frac{1}{m}} y\right| > 1 \right\}$$

respectively. Thus, for any $z \in \mathrm{E}\left(0, \frac{j}{m}\right)$ and $y \in \mathbb{R}^d \setminus \mathrm{E}\left(0, \frac{j}{m} + \frac{1}{m}\right)$,

$$\begin{aligned}
|y-z| &\geq \left|\mathrm{D}^{-1}_{\frac{j}{m}+\frac{1}{m}}\right|_*^{-1} \left|\mathrm{D}^{-1}_{\frac{j}{m}+\frac{1}{m}} y - \mathrm{D}^{-1}_{\frac{j}{m}+\frac{1}{m}} \mathrm{D}_{\frac{j}{m}} \mathrm{D}^{-1}_{\frac{j}{m}} z\right| \\
&\geq \left|\mathrm{D}^{-1}_{\frac{j}{m}+\frac{1}{m}}\right|_*^{-1} \left(\left|\mathrm{D}^{-1}_{\frac{j}{m}+\frac{1}{m}} y\right| - \left|\mathrm{D}^{-1}_{\frac{j}{m}+\frac{1}{m}} \mathrm{D}_{\frac{j}{m}}\right|_* \left|\mathrm{D}^{-1}_{\frac{j}{m}} z\right|\right) \\
&> \left|\mathrm{D}^{-1}_{\frac{j}{m}+\frac{1}{m}}\right|_*^{-1} \left(1 - \left|\mathrm{D}^{-1}_{\frac{j}{m}+\frac{1}{m}} \mathrm{D}_{\frac{j}{m}}\right|_*\right) \\
&= \min_{i=1,\ldots,d} \left\{\frac{j_i}{m} + \frac{1}{m}\right\} \left(1 - \max_{i=1,\ldots,d} \left\{\frac{\frac{j_i}{m}}{\frac{j_i}{m} + \frac{1}{m}}\right\}\right) \\
&\geq \frac{1}{Dm^2}
\end{aligned}$$

since $j_i \in \{0, \ldots, Dm-1\}$. □

A.7. Proof of Lemma 4.3

For any $\varepsilon > 0$, set $K_\varepsilon = \sup\{K > 0 : \mu([-K,K]^d) \leq 1 - \varepsilon\}$. We will denote the function $(0,1) \to \mathbb{R}^+$, $\varepsilon \mapsto K_\varepsilon$ by $K_\bullet$. Now, introduce the bracket $[L, U_\varepsilon]$, given by

$$L \equiv 0 \qquad \text{and} \qquad U_\varepsilon := T\left[\mathbb{R}^d \setminus [-K_{\varepsilon^s/2}, K_{\varepsilon^s/2}]^d, [-K_{\varepsilon^s}, K_{\varepsilon^s}]^d\right].$$

Obviously, we have $\|U_\varepsilon - L\|_s \leq \|U_\varepsilon - L\|_1^{1/s} \leq \varepsilon$.

To get a bound for the Hölder-norm of U_ε, consider the distribution function

$$G(t) := \mu\big(\{x \in \mathbb{R}^d : |x|_{\max} \leq t\}\big)$$

on $\mathbb{R}$, where $|x|_{\max} = \max\{|x_i| : i = 1, \ldots, d\}$. Observe that the pseudo-inverse G^{-1} of G is linked to $K_\bullet$ by the equality $K_\varepsilon = G^{-1}(1-\varepsilon)$. With geometrical arguments we infer

$$G(t) = \sum_{j \in \{-1,1\}^d} \sigma(j)\, \mathrm{F}(tj),$$

where $\sigma(j) := \prod_{i=1}^d j_i \in \{-1, 1\}$. Therefore

$$\begin{aligned}
\omega_G(x) &= \sup_{t\in\mathbb{R}}\{G(t+x) - G(t)\} = \sup_{t\in\mathbb{R}} \sum_{j\in\{-1,1\}^d} \sigma(j)\big(\mathrm{F}((t+x)j) - \mathrm{F}(tj)\big) \\
&\leq \sum_{j\in\{-1,1\}^d} \sup_{t\in\mathbb{R}} \big|\mathrm{F}((t+x)j) - \mathrm{F}(tj)\big| \leq \sum_{j\in\{-1,1\}^d} \omega_\mathrm{F}(\sqrt{d}x) \\
&\leq 2^d \omega_\mathrm{F}(\sqrt{d}x).
\end{aligned}$$

Now by Lemma 4.1 we obtain

$$\begin{aligned}
\|U_\varepsilon\|_{\mathcal{H}_\alpha} &\leq 1 + \frac{3^\alpha}{|G^{-1}(1-\frac{\varepsilon^s}{2}) - G^{-1}(1-\varepsilon^s)|^\alpha} \\
&\leq 1 + 3^\alpha \left(\inf\left\{x > 0 : \exists t \in \mathbb{R} \text{ such that } G(t+x) - G(t) \geq \frac{\varepsilon^s}{2}\right\}\right)^{-\alpha} \\
&\leq 1 + 3^\alpha \left(\inf\left\{x > 0 : \omega_G(x) \geq \frac{\varepsilon^s}{2}\right\}\right)^{-\alpha} \\
&\leq 1 + 3^\alpha \left(\sup\left\{x \geq 0 : \omega_{\mathrm{F}}(\sqrt{d}x) \leq \frac{\varepsilon^s}{2^{d+1}}\right\}\right)^{-\alpha} \\
&= 1 + (3\sqrt{d})^\alpha (\omega_{\mathrm{F}}^{-1}(2^{-(d+1)}\varepsilon^s))^{-\alpha},
\end{aligned}$$

where we used that ω_{F} is continuous here to replace the infimum by the supremum. Then $[L, U_\varepsilon]$ is an $(\varepsilon, 4\sqrt{d}(\omega_{\mathrm{F}}^{-1}(2^{-(d+1)}\varepsilon^s))^{-\alpha}, \mathcal{G}, \mathrm{L}^s(\mu))$-bracket for sufficiently small ε. Since $[L, U_\varepsilon]$ contains any $f \in \mathcal{F}\backslash\mathcal{F}_{K_{\varepsilon/2}+D}$, by (4.2) we obtain for all those ε the bound

$$N\Big(\varepsilon, \max\big\{f(\varepsilon), 4\sqrt{d}(\omega_{\mathrm{F}}^{-1}(2^{-(d+1)}\varepsilon^s))^{-\alpha}\big\}, \mathcal{F}, \mathcal{G}, \mathrm{L}^s(\mu)\Big) \leq C(K_{\varepsilon^s/2} + D)^p \varepsilon^{-q} + 1.$$

Let us finally consider the growth rate of $K_{\varepsilon^s/2}$ as $\varepsilon \to 0$. By assumption (4.3) and since $|\cdot|_{\max} \leq |\cdot|$, we have $1 - G(t) \leq bt^{-1/\beta}$ for sufficiently large t. Therefore,

$$G((b/\varepsilon)^\beta) \geq 1 - \varepsilon.$$

By the definition of $K_\bullet$, we therefore obtain that $K_{\varepsilon^s/2} \leq (2b/\varepsilon^s)^\beta = O_{\beta,b}(\varepsilon^{-\beta s})$ which proves the lemma. □

Nomenclature

X^* the measurable cover function of a real valued random element X, page 9

$\lfloor \cdot \rfloor$ the lower Gauss bracket, page 4

$\lceil \cdot \rceil$ the upper Gauss bracket, page 62

$|\cdot|$ the absolute value, the modulus, or the euclidean norm on $\mathbb{R}^d$

$\#A$ the cardinality of a set A

$\mathbb{D}([-\infty,\infty]^d)$ the space of multidimensional *càdlàg* functions on $[-\infty,\infty]^d$, page 8

$\xrightarrow{d}$ convergence in distribution of a sequence of random elements, page 9

$\ll$ asymptotically smaller or equal, page 92

$\|\cdot\|_{\mathcal{H}_\alpha}$ the α-Hölder norm, page 22

$\|\cdot\|_{\mathcal{H}_{\alpha,\beta}}$ the α,β-Lipschitz norm with weight in x_0, page 22

$\|\cdot\|_\infty$ the essential supremum norm w.r.t. the corresponding probability measure, page 17

$\|\cdot\|_s$ the sth moment of f w.r.t. the corresponding probability measure, page 17

$\|\cdot\|_{\mathcal{L}(\mathcal{B})}$ the operator norm w.r.t. $\mathcal{B}$, page 21

$\mathcal{B}$ a complex Banach space of measurable functions from $\mathcal{X}$ to $\mathbb{C}$

$\mathcal{B}_\mathbb{R}$ the real Banach space composed of the real-valued function in $\mathcal{B}$

$\mathcal{C}$ a normed vector space of real valued measurable functions on $\mathcal{X}$

$\mathcal{C}_M$ $:= \{f \in \mathcal{C} : \|f\|_\infty \leq M\}$ for some $M > 0$, page 29

δ_x the Dirac measure given by $\delta_x(A) = \mathbf{1}_A(x)$

$\delta_{i,m}$ a physical dependence measure, page 63

$\mathbf{E}^*(X)$ the outer expectation or outer integral of a real-valued random element X, page 9

F the (multidimensional) distribution function

$F^{-1}(t)$ $:= \sup\{x \in [-\infty,\infty] : F(x) \leq t\}$

F_n the empirical distribution function, page 65

$\mathcal{H}_\alpha(\mathcal{X}, \mathbb{K})$ the space of bounded α-Hölder continuous functions on $\mathcal{X}$ with values in $\mathbb{K}$, page 22

$\mathcal{H}_{\alpha,\beta}(\mathcal{X}, \mathbb{K})$ the space of weighted Lipschitz functions on $\mathcal{X}$ with values in $\mathbb{K}$, page 22

i.i.d. independent and identically distributed

$\mathcal{L}(\mathcal{B})$ the space of bounded linear operators from $\mathcal{B}$ to $\mathcal{B}$, page 20

$\ell^\infty(\mathcal{F})$ the space of uniformly bounded real-valued functions of $\mathcal{F}$

$\mathrm{L}^s(\lambda)$ the Lebesgue space of s-th power integrable complex-valued functions, page 17

$\mathrm{L}^\infty(\lambda)$ the space of complex-valued measurable functions that are essentially bounded w.r.t. the corresponding probability measure, page 17

μ a probability distribution, usually the marginal distribution of the stationary process $(X_i)_{i\in\mathbb{N}^*}$

$m_\alpha(f)$ $:= \sup_{x,y\in\mathcal{X},x\neq y} \frac{|f(x)-f(y)|}{d(x,y)^\alpha}$, page 22

$m_{\alpha,\beta}(f)$ $:= \sup_{x,y\in\mathcal{X},x\neq y} \frac{|f(x)-f(y)|}{d(x,y)^\alpha(1+d(x,x_0)^\beta)}$, page 22

μf $:= \int_\mathcal{X} f \, d\mu$

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

$\mathbb{N}^*$ $:= \{1, 2, \ldots\}$

$\mathbb{N}$ $:= \{0, 1, 2, \ldots\}$

$N_\beta(f)$ $:= \sup_{x,y\in\mathcal{X},x\neq y} \frac{|f(x)-f(y)|}{d(x,y)^\alpha(1+d(x,x_0)^\beta)}$, page 22

$N(\varepsilon, A, \mathcal{F}, \mathcal{G}, \mathrm{L}^s(\mu))$ the bracketing number of $\mathcal{F}$ w.r.t. $\mathcal{G}$, page 30

$N(0, \Sigma)$ normal distribution in $\mathbb{R}^k$ with mean 0 and covariance matrix Σ

$\mathcal{O}$ $f(x) = \mathcal{O}(g(x))$ as $x \to x_0$ if and only if $\limsup_{x\to x_0} \frac{|f(x)|}{|g(x)|} < \infty$

o $f(x) = o(g(x))$ as $x \to x_0$ if and only if $\lim_{x\to x_0} \frac{|f(x)|}{|g(x)|} = 0$

$\omega_F(\delta)$ the modulus of continuity of F given by $\omega_F(\delta) = \sup\{|F(x) - F(y)| : |x - y| \leq \delta\}$

$\omega_F^{\leftarrow}(\delta)$ $:= \inf\{\delta > 0 : \omega_F(\delta) \geq y\}$

$\mathbf{P}^*$ the outer probability w.r.t. a probability measure $\mathbf{P}$, page 9

U_n the empirical process given by $\mathrm{U}_n(f) := \sqrt{n}(\mu_n(f) - \mu f)$

$\mathrm{Vect}_\mathbb{R}(f_1, \ldots, f_k)$ the smallest vector space containing $f_1, \ldots, f_k$

V_n the sequential empirical process given by $\mathrm{V}_n(f,t) := \frac{[nt]}{\sqrt{n}}(\mu_{[nt]}(f) - \mu f)$, page 85

$\mathcal{X}$ an arbitrary measurable space, the state space of the process $(X_i)_{i\in\mathbb{N}}$

$(X_i)_{i\in\mathbb{N}}$ a stationary stochastic process with state space $\mathcal{X}$ and marginal distribution μ

Index

Bibliography

Andrews, D.W., Pollard, D.: An introduction to functional central limit theorems for dependent stochastic processes. International Statistical Review **62**, 119–132 (1994)

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)

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)

Bickel, P.J., Wichura, M.J.: Convergence criteria for multiparameter stochastic processes and some applications. Ann. Math. Statist. **42**, 1656–1670 (1971)

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

Borovkova, S., Burton, R., Dehling, H.: Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation. Trans. Amer. Math. Soc. **353**(11), 4261–4318 (electronic) (2001)

Bradley, R.C.: Introduction to strong mixing conditions. Vol. 1 – 3. Kendrick Press, Heber City, UT (2007)

Bush, R.R., Mosteller, F.: A mathematical model for simple learning. Psych. Rev. **58**(5), 313–323 (1951)

Bush, R.R., Mosteller, F.: A stochastic model with applications to learning. Ann. Math. Statistics **24**, 559–585 (1953)

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)

Davydov, J.A.: The invariance principle for stationary processes. Teor. Verojatnost. i Primenen. **15**, 498–509 (1970)

Dedecker, J.: A central limit theorem for stationary random fields. Probab. Theory Related Fields **110**(3), 397–426 (1998)

Dedecker, J.: An empirical central limit theorem for intermittent maps. Probab. Theory Related Fields **148**(1-2), 177–195 (2010)

Dedecker, J., Doukhan, P., Lang, G., León, J.R., Louhichi, S., Prieur, C.: Weak dependence: with examples and applications, *Lecture Notes in Statistics*, vol. 190. Springer, New York (2007)

Dedecker, J., Merlevède, F., Pène, F.: Empirical central limit theorems for ergodic automorphisms of the torus (2013). Preprint. arXiv:1210.3546

Dedecker, J., Prieur, C.: New dependence coefficients. Examples and applications to statistics. Probab. Theory Related Fields **132**(2), 203–236 (2005)

Dedecker, J., Prieur, C.: An empirical central limit theorem for dependent sequences. Stochastic Process. Appl. **117**(1), 121–142 (2007)

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. To appear in Bernoulli. arXiv: 1201.2256

Dehling, H., Durieu, O., Tusche, M.: A sequential empirical CLT for multiple mixing processes with application to B-geometrically ergodic Markov chains (2013). Preprint. arXiv: 1303.4537

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)

Dhompongsa, S.: A note on the almost sure approximation of the empirical process of weakly dependent random vectors. Yokohama Math. J. **32**(1-2), 113–121 (1984)

Donsker, M.D.: Justification and extension of Doob's heuristic approach to the Komogorov-Smirnov theorems. Ann. Math. Statistics **23**, 277–281 (1952)

Doukhan, P., Louhichi, S.: A new weak dependence condition and applications to moment inequalities. Stochastic Process. Appl. **84**(2), 313–342 (1999)

Doukhan, P., Massart, P., Rio, E.: Invariance principles for absolutely regular empirical processes. Ann. Inst. H. Poincaré Probab. Statist. **31**(2), 393–427 (1995)

Doukhan, P., Surgailis, D.: Functional central limit theorem for the empirical process of short memory linear processes. C. R. Acad. Sci. Paris Sér. I Math. **326**(1), 87–92 (1998)

Dudley, R.M.: Weak convergences of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces. Illinois J. Math. **10**, 109–126 (1966)

Dudley, R.M.: Central limit theorems for empirical measures. Ann. Probab. **6**(6), 899–929 (1978)

Durieu, O.: Comportements asymptotiques des processus stationnaires et des processus empiriques dans des systèmes dynamiques. Ph.D. thesis, Université de Rouen (2008a)

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

Durieu, O.: Empirical processes of iterated maps that contract on average. Stat. Probabil. Lett. **83**(11), 2454–2458 (2013)

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

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

Ibragimov, I.A., Linnik, Y.V.: Independent and stationary sequences of random variables. Wolters-Noordhoff Publishing, Groningen (1971). With a supplementary chapter by I. A. Ibragimov and V. V. Petrov, Translation from the Russian edited by J. F. C. Kingman

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)

Le Borgne, S.: Dynamique symbolique et propriétes stochastiques des automorphismes du tore: cas hyperbolique et quasihyperbolique. Ph.D. thesis, Université de Rennes (1997)

Le Borgne, S.: Limit theorems for non-hyperbolic automorphisms of the torus. Israel J. Math. **109**, 61–73 (1999)

Le Borgne, S., Pène, F.: Vitesse dans le théorème limite central pour certains systèmes dynamiques quasi-hyperboliques. Bull. Soc. Math. France **133**(3), 395–417 (2005)

Leonov, V.P.: On the central limit theorem for ergodic endomorphisms of compact commutative groups. Dokl. Akad. Nauk SSSR **135**, 258–261 (1960)

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)

Neuhaus, G.: On weak convergence of stochastic processes with multidimensional time parameter. Ann. Math. Statist. **42**, 1285–1295 (1971)

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)

Peškir, G., Yukich, J.E.: Uniform ergodic theorems for dynamical systems under VC entropy conditions. In: Probability in Banach spaces, 9 (Sandjberg, 1993), *Progr. Probab.*, vol. 35, pp. 105–128. Birkhäuser Boston, Boston, MA (1994)

Philipp, W.: Invariance principles for sums of mixing random elements and the multivariate empirical process. In: Limit theorems in probability and statistics, Vol. I, II (Veszprém, 1982), *Colloq. Math. Soc. János Bolyai*, vol. 36, pp. 843–873. North-Holland, Amsterdam (1984)

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

Prohorov, Y.V.: Convergence of random processes and limit theorems in probability theory. Theor. Probability Appl. **1**, 157–214 (1956)

Rio, E.: Processus empiriques absolument réguliers et entropie universelle. Probab. Theory Related Fields **111**(4), 585–608 (1998)

Straf, M.L.: Weak convergence of stochastic processes with several parameters. In: Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory, pp. 187–221. Univ. California Press, Berkeley, Calif. (1972)

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.: Empirical processes of stationary sequences. Statist. Sinica **18**(1), 313–333 (2008)