

Approximating class approach for empirical processes of dependent sequences indexed by functions

HEROLD DEHLING^{1,*}, OLIVIER DURIEU² and MARCO TUSCHE^{1,**}

¹*Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany.*

*E-mail: *herold.dehling@rub.de; **marco.tusche@rub.de*

²*Laboratoire de Mathématiques et Physique Théorique UMR-CNRS 7350, Fédération Denis Poisson FR-CNRS 2964, Université François-Rabelais de Tours, Parc de Grandmont, 37200 Tours, France.*

E-mail: olivier.durieu@lmpt.univ-tours.fr

We study weak convergence of empirical processes of dependent data $(X_i)_{i \geq 0}$, indexed by classes of functions. Our results are especially suitable for data arising from dynamical systems and Markov chains, where the central limit theorem for partial sums of observables is commonly derived via the spectral gap technique. We are specifically interested in situations where the index class \mathcal{F} is different from the class of functions f for which we have good properties of the observables $(f(X_i))_{i \geq 0}$. We introduce a new bracketing number to measure the size of the index class \mathcal{F} which fits this setting. Our results apply to the empirical process of data $(X_i)_{i \geq 0}$ satisfying a multiple mixing condition. This includes dynamical systems and Markov chains, if the Perron–Frobenius operator or the Markov operator has a spectral gap, but also extends beyond this class, for example, to ergodic torus automorphisms.

Keywords: Empirical processes indexed by classes of functions; dependent data; Markov chains; dynamical systems; ergodic torus automorphism; weak convergence

1. Introduction

Let $(X_i)_{i \geq 0}$ be a stationary stochastic process of \mathbb{R} -valued random variables with marginal distribution μ . We denote the empirical measure of order n by $\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$. The classical empirical process is defined by $U_n(t) = \sqrt{n}(\mu_n((-\infty, t]) - \mu((-\infty, t]))$, $t \in \mathbb{R}$. In the case of i.i.d. processes, the limit behavior of the empirical process was first investigated by Donsker [15], who proved that $(U_n(t))_{t \in \mathbb{R}}$ converges weakly to a Brownian bridge process. This result, known as Donsker’s empirical process central limit theorem, confirmed a conjecture of Doob [16] who had observed that certain functionals of the empirical process converge in distribution towards the corresponding functionals of a Brownian bridge. Donsker’s empirical process CLT has been generalized to dependent data by many authors. One of the earliest results is Billingsley [4], who considered functions of mixing processes, with an application to the empirical distribution of the remainders in a continued fraction expansion.

Empirical processes play a very important role in large sample statistical inference. Many statistical estimators and test statistics can be expressed as functionals of the empirical distri-

bution. As a result, their asymptotic distribution can often be derived from empirical process limit theorems, combined with the continuous mapping theorem or a functional delta method. A well-known example is the Kolmogorov–Smirnov goodness-of-fit test, which uses the test statistic $D_n := \sup_{t \in \mathbb{R}} \sqrt{n} |\mu_n((-\infty, t]) - \mu_0((-\infty, t])|$ in order to test the null hypothesis that μ_0 is the marginal distribution of X_1 . Under the null hypothesis, the limit distribution of D_n is given by the supremum of the Gaussian limit of the empirical process. Another example are Von-Mises-statistics, also known as V-statistics. These are defined as $V_n := \frac{1}{n^2} \sum_{1 \leq i, j \leq n} h(X_i, X_j)$, where $h(x, y)$ is a symmetric kernel function. Specific examples include the sample variance and Gini’s mean difference, where the kernel functions are given by $(x - y)^2/2$ and $|x - y|$, respectively. V-statistics can be expressed as integrals with respect to the empirical distribution function, namely $V_n = \iint h(x, y) d\mu_n(x) d\mu_n(y)$. The asymptotic distribution of V_n can then be derived via a functional delta method from an empirical process central limit theorem; see, for example, Beutner and Zähle [2] for some recent results.

Empirical process CLTs for \mathbb{R}^d -valued i.i.d. data $(X_i)_{i \geq 0}$ have first been studied by Dudley [19], Neuhaus [27], Bickel and Wichura [3] and Straf [33]. These authors consider the classical d -dimensional empirical process $\sqrt{n}(\mu_n((-\infty, t]) - \mu((-\infty, t]))$, where $(-\infty, t] = \{x \in \mathbb{R}^d: x_1 \leq t_1, \dots, x_d \leq t_d\}$, $t \in \mathbb{R}^d$, denotes the semi-infinite rectangle in \mathbb{R}^d . Philipp and Pinzur [31], Philipp [30] and Dhompongsa [13] studied weak convergence of the multivariate empirical process in the case of mixing data.

Dudley [20] initiated the study of empirical processes indexed by classes of sets, or more generally by classes of functions. This approach allows the study of empirical processes for very general data, not necessarily having values in Euclidean space. CLTs for empirical processes indexed by classes of functions require entropy conditions on the size of the index set. For i.i.d. data, Dudley [20] obtained the CLT for empirical processes indexed by classes of sets satisfying an entropy condition with inclusion. Ossiander [29] used an entropy condition with bracketing to obtain results for empirical processes indexed by classes of functions. For the theory of empirical processes of i.i.d. data, indexed by classes of functions, see the book by van der Vaart and Wellner [34]. Limit theorems for more general empirical processes indexed by classes of functions have also been studied under entropy conditions for general covering numbers, for example, by Nolan and Pollard [28] who investigate empirical U -processes.

In the case of strongly mixing data, Andrews and Pollard [1] were the first to obtain CLTs for empirical processes indexed by classes of functions. Doukhan, Massart and Rio [18] and Rio [32] study empirical processes for absolutely regular data. Borovkova, Burton and Dehling [5] investigate the empirical process and the empirical U -process for data that can be represented as functionals of absolutely regular processes. For further results, see the survey article by Dehling and Philipp [12], the book by Dedecker *et al.* [8], as well as the paper by Dedecker and Prieur [9].

A lot of research has been devoted to the study of statistical properties of data arising from dynamical systems or from Markov chains. A very powerful technique to prove CLTs and other limit theorems is the spectral gap method, using spectral properties of the Perron–Frobenius operator or the Markov operator on an appropriate space of functions; see Hennion and Hervé [24]. When the space of functions under consideration contains the class of indicator functions of intervals, standard tools can be used to establish the classical empirical process CLT. Finite-dimensional convergence of the empirical process follows from the CLT for $\sum_{i=1}^n 1_{(-\infty, t]}(X_i)$,

and tightness can be established using moment bounds for $\sum_{i=1}^n 1_{(s,t]}(X_i)$. Collet, Martinez and Schmitt [7] used this approach to establish the empirical process CLT for expanding maps of the unit interval.

The situation differs markedly when the CLT and moment bounds are not directly available for the index class of the empirical process, but only for a different class of functions. Recently, Dehling, Durieu and Volný [11] developed techniques to cover such situations. They were able to prove classical empirical process CLTs for \mathbb{R} -valued data when the CLT and moment bounds are only available for Lipschitz functions. Dehling and Durieu [10] extended these techniques to \mathbb{R}^d -valued data satisfying a multiple mixing condition for Hölder continuous functions. Under this condition, they proved the CLT for the empirical process indexed by semi-infinite rectangles $(-\infty, t]$, $t \in \mathbb{R}^d$. The multiple mixing condition is strictly weaker than the spectral gap condition. For example, ergodic torus automorphisms satisfy a multiple mixing condition, while generally they do not have a spectral gap. Dehling and Durieu [10] proved the empirical process CLT for ergodic torus automorphisms. Durieu and Tusche [23] provide very general conditions under which the classical empirical process CLT for \mathbb{R}^d -valued data holds.

The above mentioned papers study exclusively classical empirical processes, indexed by semi-infinite intervals or rectangles. It is the goal of the present paper to extend the techniques developed by Dehling, Durieu and Volný [11] to empirical processes indexed by classes of functions. Let $(\mathcal{X}, \mathcal{A})$ be a measurable space, let $(X_i)_{i \geq 0}$ be a stationary process of \mathcal{X} -valued random variables, and let \mathcal{F} be a uniformly bounded class of real-valued functions on \mathcal{X} . We consider the \mathcal{F} -indexed empirical process $(\frac{1}{\sqrt{n}} \sum_{i=1}^n (f(X_i) - \mathbb{E}f(X_1)))_{f \in \mathcal{F}}$. As in the above mentioned papers, we will assume that there exists some Banach space \mathcal{B} of functions on \mathcal{X} such that the CLT and a moment bound hold for partial sums $\sum_{i=1}^n g(X_i)$, for all g in some subset of \mathcal{B} ; see Assumptions 1 and 2. These conditions are satisfied, for example, when the Perron–Frobenius operator or the Markov operator acting on \mathcal{B} has a spectral gap. Again, if the index class \mathcal{F} is a subset of \mathcal{B} , standard techniques for proving empirical process CLTs can be applied. In many examples, however, \mathcal{B} is some class of regular functions, while \mathcal{F} is a class of indicators of sets. It is the goal of the present paper to provide techniques suitable for this situation.

Empirical process invariance principles require a control on the size of the index class \mathcal{F} , as measured by covering or bracketing numbers; see, for example, van der Vaart and Wellner [34]. In this paper, we will consider coverings of \mathcal{F} by \mathcal{B} -brackets, that is, brackets bounded by functions $l, u \in \mathcal{B}$. Because of the specific character of our moment bounds, we have to impose conditions on the \mathcal{B} -norms of l and u . We will thus introduce a notion of bracketing numbers by counting how many \mathcal{B} -brackets of a given L^s -size and with a given control on the \mathcal{B} -norms of the upper and lower functions are needed to cover \mathcal{F} . The main theorem of the present paper establishes an empirical process CLT under an integral condition on this bracketing number.

This paper is organized as follows: Section 2 contains precise definitions as well as the statement of the main theorem. In Section 3, we will specifically consider the case when \mathcal{B} is the space of Hölder continuous functions. We will give examples of classes of functions which satisfy the bracketing number assumption. In Section 4, we will give applications to ergodic torus automorphisms which extend the empirical process CLT of Dehling and Durieu [10] to more general classes of sets. Section 5 contains the proof of our main theorem, while proofs of technical aspects of the examples can be found in the [Appendix](#).

2. Main result

Let $(\mathcal{X}, \mathcal{A})$ be a measurable space, and let $(X_i)_{i \in \mathbb{N}}$ be an \mathcal{X} -valued stationary stochastic process with marginal distribution μ . Let \mathcal{F} be a uniformly bounded class of real-valued measurable functions defined on \mathcal{X} . If Q is a signed measure on $(\mathcal{X}, \mathcal{A})$, we use the notation $Qf = \int_{\mathcal{X}} f \, dQ$. We define the map $F_n : \mathcal{F} \rightarrow \mathbb{R}$, induced by the empirical measure,

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

The \mathcal{F} -indexed empirical process of order n is given by

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

We regard the empirical process $(U_n(f))_{f \in \mathcal{F}}$ as a random element on $\ell^\infty(\mathcal{F})$; this holds as \mathcal{F} is supposed to be uniformly bounded. $\ell^\infty(\mathcal{F})$ is equipped with the supremum norm and the Borel σ -field generated by the open sets. It is well known that, in general, $(U_n(f))_{f \in \mathcal{F}}$ is not measurable and thus the usual theory of weak convergence of random variables does not apply. We use here the theory which is based on convergence of outer expectations; see van der Vaart and Wellner [34]. Given a Borel probability measure L on $\ell^\infty(\mathcal{F})$, we say that $(U_n(f))_{n \geq 1}$ converges in distribution to L if

$$\mathbb{E}^*(\varphi(U_n)) \rightarrow \int \varphi(x) \, dL(x)$$

for all bounded and continuous functions $\varphi : \ell^\infty(\mathcal{F}) \rightarrow \mathbb{R}$. Here \mathbb{E}^* denotes the outer integral. Note that $\mathbb{E}^*(X) = \mathbb{E}(X^*)$, where X^* denotes the measurable cover function of X ; see Lemma 1.2.1 in van der Vaart and Wellner [34].

In what follows, we will frequently make two assumptions concerning the process $(f(X_i))_{i \in \mathbb{N}}$, where $f : \mathcal{X} \rightarrow \mathbb{R}$ belongs to some Banach space $(\mathcal{B}, \|\cdot\|_{\mathcal{B}})$ of measurable functions on \mathcal{X} , respectively, to some subset $\mathcal{G} \subset \mathcal{B}$. The precise choice of \mathcal{B} , as well as of \mathcal{G} , will depend on the specific example. Often, we take \mathcal{B} to be the space of all Lipschitz or Hölder continuous functions, and \mathcal{G} the intersection of \mathcal{B} with an $\ell^\infty(\mathcal{X})$ -ball.

Assumption 1 (CLT for \mathcal{B} -observables). For all $f \in \mathcal{B}$, there exists a $\sigma_f^2 \geq 0$ such that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (f(X_i) - \mu f) \xrightarrow{\mathcal{D}} N(0, \sigma_f^2), \tag{2.1}$$

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

Assumption 2 (Moment bounds for \mathcal{G} -observables). For some subset $\mathcal{G} \subset \mathcal{B}$, $s \geq 1$, and $a \in \mathbb{R}$, for all $p \geq 1$, there exists a constant $C_p > 0$ such that for all $f \in \mathcal{G} - \mathcal{G} := \{g_1 - g_2 : g_1, g_2 \in \mathcal{G}\}$,

$$\mathbb{E} \left[\left(\sum_{i=1}^n (f(X_i) - \mu f) \right)^{2p} \right] \leq C_p \sum_{i=1}^p n^i \|f\|_s^i \log^{2p+ai} (\|f\|_{\mathcal{B}} + 1), \tag{2.2}$$

where $\|f\|_s = (\int_{\mathcal{X}} |f|^s d\mu)^{1/s}$ denotes the L^s -norm of f .

Both Assumptions 1 and 2 have been established by many authors for a wide range of stationary processes. Concerning the CLT, see, for example, the three-volume monograph by Bradley [6] for mixing processes, Dedecker *et al.* [8] for so-called weakly dependent processes in the sense of Doukhan and Louhichi [17], and Hennion and Hervé [24] for many examples of Markov chains and dynamical systems. Durieu [21] proved 4th moment bounds of the type (2.2) for Markov chains or dynamical systems for which the Markov operator or the Perron–Frobenius operator acting on \mathcal{B} has a spectral gap. It was generalized to $2p$ th moment bounds by Dehling and Durieu [10]. More generally, they gave similar moment bounds for processes satisfying a multiple mixing condition, that is, assuming that there exist a $\theta \in (0, 1)$ and an integer $d_0 \in \mathbb{N}$ such that for all integers $p \geq 1$, there exist an integer ℓ and a multivariate polynomial P of total degree smaller than d_0 such that

$$\begin{aligned} & \left| \text{Cov}(f(X_{i_0}) \cdots f(X_{i_{q-1}}), f(X_{i_q}) \cdots f(X_{i_p})) \right| \\ & \leq \|f\|_s \|f\|_{\mathcal{B}}^{\ell} P(i_1 - i_0, \dots, i_p - i_{p-1}) \theta^{i_q - i_{q-1}} \end{aligned} \tag{2.3}$$

holds for all $f \in \mathcal{B}$ with $\mu f = 0$ and $\|f\|_{\infty} \leq 1$, all integers $i_0 \leq i_1 \leq \dots \leq i_p$ and all $q \in \{1, \dots, p\}$. See Theorem 4 and the examples in Dehling and Durieu [10]. Note that this multiple mixing condition implies the moment bound (2.2) with for $\mathcal{G} = \{f \in \mathcal{B} : \|f\|_{\infty} \leq 1\}$ and $a = d_0 - 1$. Further, the spectral gap property leads to the multiple mixing condition with $d_0 = 0$, and thus to the moment bound (2.2) with $a = -1$, see Dehling and Durieu [10], Section 4.

We will derive a general statement about weak convergence of the empirical process $(U_n(f))_{f \in \mathcal{F}}$ under the two assumptions (2.1) and (2.2). Empirical process central limit theorems require bounds on the size of the class of functions \mathcal{F} , usually measured by the number of ε -balls required to cover \mathcal{F} . Here we will introduce a covering number adapted to the fact that (2.1) and (2.2) hold only for $f \in \mathcal{B}$ or $f \in \mathcal{G}$, respectively, and that both the \mathcal{B} -norm as well as the $L^s(\mu)$ -norm enter on the right hand side of the bound (2.2). In our approach, we use \mathcal{B} -brackets to cover the class \mathcal{F} , which leads to the following definition.

Definition. Let $(\mathcal{X}, \mathcal{A})$ be a measurable space, and let μ be a probability measure on $(\mathcal{X}, \mathcal{A})$. Let \mathcal{B} be some Banach space of measurable functions on \mathcal{X} , $\mathcal{G} \subset \mathcal{B}$ and $s \geq 1$.

(i) Given two functions $l, u : \mathcal{X} \rightarrow \mathbb{R}$ satisfying $l(x) \leq u(x)$, for all $x \in \mathcal{X}$, we define the bracket

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

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

$$\begin{aligned} \|u - l\|_s &\leq \varepsilon, \\ \|u\|_{\mathcal{B}} &\leq A, \quad \|l\|_{\mathcal{B}} \leq A, \end{aligned}$$

where $\|\cdot\|_s$ denotes the $L^s(\mu)$ -norm.

(ii) For a class of measurable functions \mathcal{F} , defined on \mathcal{X} , we define the bracketing number $N(\varepsilon, A, \mathcal{F}, \mathcal{G}, L^s(\mu))$ as the smallest number of $(\varepsilon, A, \mathcal{G}, L^s(\mu))$ -brackets needed to cover \mathcal{F} .

Our definition is close to the definition of bracketing numbers given by Ossiander [29], but different. In Ossiander [29], no assumptions are made on the upper and lower functions of the bracket other than that they are close in L^2 . Here, the moment bound (2.2) forces us to require the extra condition that u and l belong to the space \mathcal{B} and that their \mathcal{B} -norms are controlled. Obviously, our bracketing numbers are always larger than the ones defined in Ossiander [29], and naturally our condition on the size of \mathcal{F} are stronger. On the other hand, our results apply to dependent data, while Ossiander [29] studies i.i.d. data.

We can now state the main theorem of the present paper. The proof will be given in Section 5.

Theorem 2.1. *Let $(\mathcal{X}, \mathcal{A})$ be a measurable space, let $(X_i)_{i \geq 1}$ 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 Banach space \mathcal{B} of measurable functions on \mathcal{X} , some subset $\mathcal{G} \subset \mathcal{B}$, $a \in \mathbb{R}$, and $s \geq 1$, Assumptions 1 and 2 hold. Moreover, assume that there exist constants $r > -1$, $\gamma > \max\{2 + a, 1\}$ and $C > 0$ such that*

$$\int_0^1 \varepsilon^r \sup_{\varepsilon \leq \delta \leq 1} N^2(\delta, \exp(C\delta^{-1/\gamma}), \mathcal{F}, \mathcal{G}, L^s(\mu)) \, d\varepsilon < \infty. \tag{2.4}$$

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

Remark 2.2. (i) Note that the bracketing number $N(\delta, \exp(C\delta^{-1/\gamma}), \mathcal{F}, \mathcal{G}, L^s(\mu))$ might not be a monotone function of δ . This is the reason why we take the supremum in the integral (2.4).

(ii) The proof of Theorem 2.1 shows that the statement also holds if condition (2.2) is only satisfied for some integer p satisfying

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

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

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

as $\varepsilon \rightarrow 0$, condition (2.4) is satisfied for all $r > 2r' - 1$.

In the next section, we will present examples of classes of functions satisfying condition (2.4). Among the examples are indicators of multidimensional rectangles, of ellipsoids, and of balls of arbitrary metrics, as well as a class of monotone functions. In Section 4, we give applications to ergodic torus automorphisms, indexed by various classes of indicator functions.

3. Examples of classes of functions

In many examples that satisfy Assumptions 1 and 2, the Banach space \mathcal{B} is the space of Lipschitz or Hölder continuous functions, see examples in Dehling, Durieu and Volný [11], Dehling and Durieu [10], or Durieu and Tusche [23]. Thus, in this section, we will restrict our attention to the case where \mathcal{B} is a space of Hölder functions and give several examples of classes \mathcal{F} which satisfy the entropy condition (2.4).

In this section, we consider a metric space (\mathcal{X}, d) . Let $\alpha \in (0, 1]$ be fixed. We denote by $\mathcal{H}_\alpha(\mathcal{X})$ the space of bounded α -Hölder continuous functions on \mathcal{X} with values in \mathbb{R} . This space is equipped with the norm

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

For this section, we chose $\mathcal{B} = \mathcal{H}_\alpha(\mathcal{X})$. As the approximating class we use the subclass $\mathcal{G} = \mathcal{H}_\alpha(\mathcal{X}, [0, 1]) := \{f \in \mathcal{H}_\alpha(\mathcal{X}) : 0 \leq f \leq 1\}$ of \mathcal{B} . Except in Example 5, in all examples we will 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 will use the transition function given in the following definition which uses the notations

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

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

Definition. Let A, B be subsets of \mathcal{X} such that $d(A, B) > 0$. We define the transition function $T[A, B]: \mathcal{X} \rightarrow \mathbb{R}$ 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 3.1. 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

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

This lemma is proved in the [Appendix](#).

We also use the following notations: For a non-decreasing 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, \dots, p_k will be denoted with these parameters in the subscript, such as c_{p_1, \dots, p_k} . Furthermore, the notation $f(x) = O_{p_1, \dots, p_k}(g(x))$ as $x \rightarrow 0$ or $x \rightarrow \infty$ means that there exists a constant c_{p_1, \dots, p_k} such that $f(x) \leq c_{p_1, \dots, p_k}g(x)$ for all x sufficiently small or large, respectively.

3.1. Example 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, that is, the class $\{1_{(-\infty, t]}: t \in \mathbb{R}^d\}$, where $(-\infty, t]$ denotes the set of points x such that $^1 x \leq t$. Under similar assumptions as in the present paper, this case was treated by Dehling and Durieu [10]. We will see that Theorem 2.1 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 $x \leq u$.

Proposition 3.2. *Let $s \geq 1$, $\gamma > 1$, and let μ be a probability distribution on \mathbb{R}^d whose distribution function F satisfies*

$$\omega_F(x) = O(|\log(x)|^{-s\gamma}) \quad \text{as } x \rightarrow 0. \tag{3.1}$$

Then there exists a constant $C = C_F > 0$ such that

$$N(\varepsilon, \exp(C\varepsilon^{-1/\gamma}), \mathcal{F}, \mathcal{G}, L^s(\mu)) = O_d(\varepsilon^{-2ds}) \quad \text{as } \varepsilon \rightarrow 0,$$

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

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

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

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

where F_i^{-1} is the pseudo-inverse of the marginal distribution function² F_i . 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 α -Hölder functions

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

and

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

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

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

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

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

using Lemma 3.1 and (3.1), we have

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

where c_F is given by (3.1). The same bound holds for $\|u_{k,j}\|_\alpha$.

² $F_i(t) = \mu(\mathbb{R} \times \dots \times \mathbb{R} \times (-\infty, t] \times \mathbb{R} \times \dots \times \mathbb{R})$.

Thus, there exists a new constant $C_F > 0$ such that for all $k \leq j \in \{0, \dots, m\}^d$, $[l_{k,j}, u_{k,j}]$ is an $(\varepsilon, \exp(C_F \varepsilon^{-1/\gamma}), \mathcal{G}, L^s(\mu))$ -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 the proposition. \square

Notice that under the assumptions of the proposition, condition (2.4) is satisfied and therefore Theorem 2.1 may be applied to empirical processes indexed by the class of indicators of rectangles, taking \mathcal{B} to be the class of bounded Hölder functions.

Corollary 3.3. *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 of the X_i satisfies (3.1). Then the empirical process $(U_n(f))_{f \in \mathcal{F}}$ converges in distribution in $\ell^\infty(\mathcal{F})$ to a tight Gaussian process.*

Remark 3.4. By regarding the class of indicator functions of left infinite rectangles as a subclass of \mathcal{F} , we obtain Theorem 1 of Dehling and Durieu [10] as a particular case of the preceding corollary.

3.2. Example 2: Indicators of multidimensional balls in the unit cube

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

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

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

Proposition 3.5. *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*

$$N(\varepsilon, C\varepsilon^{-\alpha s}, \mathcal{F}, \mathcal{G}, L^s(\mu)) = O_{d,B}(\varepsilon^{-(d+1)s}) \quad \text{as } \varepsilon \rightarrow 0,$$

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 (2.4). In this situation, we have a stronger type of bracketing number than in (2.4).

Proof of Proposition 3.5. Let $\varepsilon > 0$ be fixed and $m = \lceil \varepsilon^{-s} \rceil$. For all $i = (i_1, \dots, i_d) \in \{0, \dots, m\}^d$, we denote by c_i the center of the rectangle $[\frac{i_1-1}{m}, \frac{i_1}{m}] \times \dots \times [\frac{i_d-1}{m}, \frac{i_d}{m}]$. Then we define, for $i \in \{1, \dots, m\}^d$ and $j \in \{0, \dots, m\}$, the functions

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

and

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

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

By Lemma 3.1, these functions are α -Hölder and, since $d(\mathbf{B}(x, r), \mathbb{R}^d \setminus \mathbf{B}(x, r')) = r' - r$, we have

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

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

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

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

Now, if f belongs to \mathcal{F} , then $f = 1_{\mathbf{B}(x,r)}$ 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

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

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, $N(c_{d,B}^{1/s} \varepsilon, 4\varepsilon^{-\alpha s}, \mathcal{F}, \mathcal{G}, L^s(\mu)) = O_{d,B}(\varepsilon^{-(d+1)s})$ as $\varepsilon \rightarrow 0$, which implies that there exists a constant $C_{d,B} > 0$, for which $N(\varepsilon, C_{d,B} \varepsilon^{-\alpha s}, \mathcal{F}, \mathcal{G}, L^s(\mu)) = O_{d,B}(\varepsilon^{-(d+1)s})$ as $\varepsilon \rightarrow 0$. □

3.3. Example 3: Indicators of uniformly bounded multidimensional ellipsoids centered 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 center 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

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

We denote by \mathcal{F} the class of indicator functions of these ellipsoids, that is,

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

We have the following upper bound.

Proposition 3.6. *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*

$$N(\varepsilon, C\varepsilon^{-2\alpha s}, \mathcal{F}, \mathcal{G}, L^s(\mu)) = O_{d,B}(\varepsilon^{-2ds}) \quad \text{as } \varepsilon \rightarrow 0,$$

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

Proof. Let $\varepsilon > 0$ be fixed and $m = \lfloor \varepsilon^{-s} \rfloor$. For all $i = (i_1, \dots, i_d) \in \{0, \dots, m\}^d$, we denote by I_i the rectangle $[\frac{i_1-1}{m}, \frac{i_1}{m}] \times \dots \times [\frac{i_d-1}{m}, \frac{i_d}{m}]$. 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

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

and

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

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

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

where we use the convention that an ellipsoid with one negative parameter is the empty set. By Lemma 3.1, these functions are α -Hölder. Further, we have the following lemma which is proved in the Appendix.

Lemma 3.7. *For all $j \in \{0, \dots, Dm - 1\}^d$, $x \in \mathbb{R}^d$, we have*

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

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 3.1, we have

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

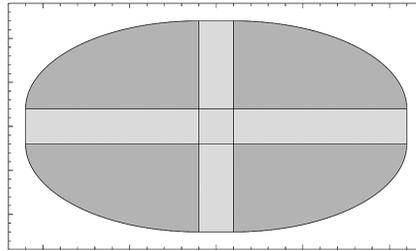


Figure 1. $U_{i,j}$ in dimension 2.

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

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 $j = (j_1, \dots, j_d) \in \mathbb{R}_+^d$ and $x \in \mathbb{R}^d$, the Lebesgue measure of the ellipsoid $E(x, j)$ is given by

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

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

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

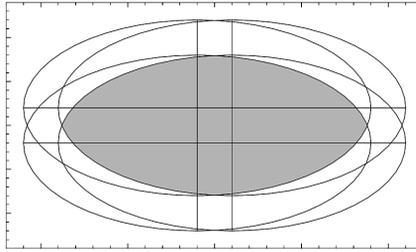


Figure 2. $L_{i,j}$ in dimension 2.

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

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

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

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

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

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

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

3.4. Example 4: Indicators of uniformly bounded multidimensional ellipsoids

In Example 3, we only considered indicators of ellipsoids centered 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 3.8. *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 $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, such that for any $K > 0$ we have*

$$N(\varepsilon, f(\varepsilon), \mathcal{F}_K, \mathcal{G}, L^s(\mu)) \leq CK^p \varepsilon^{-q} \tag{3.2}$$

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

$$\mu(\{x \in \mathbb{R}^d: |x| > t\}) \leq bt^{-1/\beta} \tag{3.3}$$

for all sufficiently large t , then

$$\begin{aligned} N(\varepsilon, \max\{f(\varepsilon), 4\sqrt{d}(\omega_F^{-1}(2^{-(d+1)}\varepsilon^s))^{-\alpha}\}, \mathcal{F}, \mathcal{G}, L^s(\mu)) \\ = O_{\beta,b,C,D,p}(\varepsilon^{-(\beta ps+q)}) \quad \text{as } \varepsilon \rightarrow 0, \end{aligned}$$

where ω_F is the modulus of continuity of F .

The proof is postponed to the [Appendix](#).

Proposition 3.9. *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 centers 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 (3.3) holds for some $\beta > 0$ and $b > 0$, then there exists a constant $C = C_{d,B,D} > 0$ such that*

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

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

Proof. In the situation of Example 3 change the set of the centers of the ellipsoids $[0, 1]^d$ to $[-K, K]^d$ and apply Lemma 3.8. Following the proof of Proposition 3.6, we can easily see that condition (3.2) 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 $\omega_F(x) \leq Bx$ and therefore $4\sqrt{d}(\omega_F^{-1}(2^{-(d+1)}\varepsilon^s))^{-\alpha} \leq 4\sqrt{d}(2^{d+1}B)^\alpha \varepsilon^{-\alpha s} \leq C_{d,B,D}\varepsilon^{-2\alpha s}$ for sufficiently small ε . \square

Remark 3.10. In the situation of Proposition 3.9 for the class \mathcal{F}' of indicators of balls in \mathbb{R}^d with uniformly bounded diameter, we can obtain the slightly sharper bound

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

for some $C = C'_{d,B} > 0$ by applying Lemma 3.8 directly to the situation in Example 2 and using the same arguments as in the previous example.

3.5. Example 5: Indicators of balls of an arbitrary metric with common center

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

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

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

Proposition 3.11. *Let $s \geq 1$ and $\gamma > 1$. If for the probability measure μ on \mathcal{X} the modulus of continuity ω_G of the function $G(t) := \mu(\mathbf{B}(t))$ satisfies*

$$\omega_G(x) = O(|\log x|^{-s\gamma}) \quad \text{as } x \rightarrow 0, \tag{3.4}$$

then there is a constant $C = C_G > 0$ such that

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

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

Remark 3.12. 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 (3.4) is

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

Proof of Proposition 3.11. Fix $\varepsilon > 0$ and choose $m = \lfloor 3\varepsilon^{-s} + 1 \rfloor$. Let G^{-1} denote the pseudo-inverse of G and set for $i \in \{1, \dots, m\}$

$$r_i := G^{-1}\left(\frac{i}{m}\right), \quad \mathbf{B}_i := \mathbf{B}(r_i).$$

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

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

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

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

By Lemma 3.1, we have

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

Since by condition (3.4)

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

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

$$\begin{aligned} \|u_i\|_\alpha &\leq 1 + 3^\alpha \exp(\alpha c_G m^{1/(s\gamma)}) \leq \exp(C_G m^{1/(s\gamma)}) \\ &\leq \exp(C_G \varepsilon^{-1/\gamma}). \end{aligned}$$

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

3.6. Example 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 $x \in \mathbb{R}$, $0 \leq f_t(x) \leq 1$;
- (ii) for all $0 \leq s \leq t \leq 1$, $f_s \leq f_t$;
- (iii) for all $t \in [0, 1]$, f_t is non-decreasing on \mathbb{R} .

Note that all the sequel remains true if in (iii), non-decreasing is replaced by non-increasing. Further, for a probability measure μ on \mathbb{R} , we define $G_\mu(t) = \mu f_t$ and we say that G_μ is Lipschitz with Lipschitz constant $\lambda > 0$ if $|G_\mu(t) - G_\mu(s)| \leq \lambda|t - s|$, for all $s, t \in [0, 1]$.

Empirical processes indexed by a 1-parameter class of functions arise, for example, in the study of empirical U-processes; see Borovkova, Burton and Dehling [5]. The empirical U-distribution function with kernel function $g(x, y)$ is defined as

$$U_n(t) = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} 1_{\{g(X_i, X_j) \leq t\}}.$$

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

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

where $g_t(x) = P(g(x, X_1) \leq t)$. 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, for example, the case for the kernel $g(x, y) = y - x$, which arises in the study of the empirical correlation integral; see Borovkova, Burton and Dehling [5].

Proposition 3.13. *Let $s \geq 1$ and $\gamma > 1$. Let μ be a probability measure on \mathbb{R} such that its distribution function F satisfies*

$$\omega_F(x) = O(|\log(x)|^{-s\gamma}) \quad \text{as } x \rightarrow 0 \tag{3.5}$$

and such that G_μ is Lipschitz with Lipschitz constant $\lambda > 0$. Then there exists a $C = C_F > 0$, such that

$$N(\varepsilon, \exp(C\varepsilon^{-1/\gamma}), \mathcal{F}, \mathcal{G}, L^s(\mu)) = O_\lambda(\varepsilon^{-s}) \quad \text{as } \varepsilon \rightarrow 0,$$

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

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

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

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 $\lim_{x \rightarrow \pm\infty} f(x)$. 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 $t_{j-1} \leq t \leq t_j$, we have $l_j \leq f_t \leq u_j$, that is, 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

$$\|l_j\|_\alpha \leq 1 + \max_{k=1, \dots, m} \frac{l_j(x_k) - l_j(x_{k-1})}{(x_k - x_{k-1})^\alpha} \leq 1 + \max_{k=1, \dots, m} \frac{1}{(x_k - x_{k-1})^\alpha} \leq 1 + \exp(C_F m^{1/(s\gamma)}).$$

Here we have used the condition (3.5) and the same computation as for the class of indicators of rectangles. Analogously, the same bound holds for $\|u_j\|_\alpha$.

Now,

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

First, since G_μ is Lipschitz, we have

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

For $x \in [x_{k-1}, x_k]$, since f_t is non-decreasing, we have $u_j(x) \leq f_{t_j}(x_{k+1})$ and $u_j(x) \geq f_{t_j}(x_{k-1})$, thus

$$\begin{aligned} \|u_j - f_{t_j}\|_1 &\leq \sum_{k=1}^{m-1} |f_{t_j}(x_{k+1}) - f_{t_j}(x_{k-1})| \mu([x_k, x_{k+1}]) \\ &\leq \frac{1}{m} \sum_{k=1}^{m-1} (|f_{t_j}(x_{k+1}) - f_{t_j}(x_k)| + |f_{t_j}(x_k) - f_{t_j}(x_{k-1})|) \\ &\leq \frac{2}{m} \sum_{k=0}^{m-1} |f_{t_j}(x_{k+1}) - f_{t_j}(x_k)| \\ &\leq \frac{2}{m} \end{aligned}$$

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 $(\varepsilon, \exp(C_F \varepsilon^{-1/\gamma}), \mathcal{G}, L^s(\mu))$ -brackets needed to cover the class \mathcal{F} is bounded by m , which proves the proposition. □

4. Application to ergodic torus automorphisms

We can apply Theorem 2.1 to the empirical process of ergodic torus automorphisms. Let $\mathbb{T}^d = \mathbb{R}^d / \mathbb{Z}^d$ be the torus of dimension $d > 1$, which is identified with $[0, 1]^d$. If A is a square matrix of dimension d with integer coefficients and determinant ± 1 , then the transformation $T : \mathbb{T}^d \rightarrow \mathbb{T}^d$ defined by

$$Tx = Ax \text{ mod } 1$$

is an automorphism of \mathbb{T}^d that preserves the Lebesgue measure λ . Thus $(\mathbb{T}^d, \mathcal{B}(\mathbb{T}^d), \lambda, T)$ is a measure preserving dynamical system. It is ergodic if and only if the matrix A has no eigenvalue which is a root of unity. A result of Kronecker shows that in this case, A always has at least one eigenvalue which has modulus different than 1. The hyperbolic automorphisms (i.e., 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 for which strong results can be proved. The central limit theorem for regular observables has been proved by Leonov [26], see also Le Borgne [25] for refinements. Other limit theorems can be found in Dolgopyat [14]. The one-dimensional empirical process, for \mathbb{R} -valued regular observables, has been studied by Durieu and Jouan [22]. Dehling and Durieu [10] proved weak convergence of the classical empirical process (indexed

by indicators of left infinite rectangles). We can now generalize this result to empirical processes indexed by further classes of functions. We can get the following proposition, as a corollary of Theorem 2.1 and the results of the preceding section.

Theorem 4.1. *Let T be an ergodic d -torus automorphism and let \mathcal{F} be one of the following classes:*

- *the class of indicators of rectangles of \mathbb{T}^d ;*
- *the class of indicators of Euclidean balls of \mathbb{T}^d ;*
- *the class of indicators of ellipsoids of bounded diameter of \mathbb{T}^d ;*

Then the empirical process

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

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

Proof. Let \mathcal{F} be one of the classes of functions and \mathcal{B} be the class of α -Hölder functions for some $\alpha \in (0, 1]$. We set \mathcal{G} the subclass of \mathcal{B} given by the functions bounded by 1. We consider the \mathbb{T}^d -valued stationary process $X_i = T^i$. Since the distribution of X_0 is the Lebesgue measure on \mathbb{T}^d , Propositions 3.2, 3.5 and 3.6 show that the condition (2.4) holds for every possible choice of class \mathcal{F} . For all $f \in \mathcal{B}$, the central limit theorem (2.1) holds; see Leonov [26] and Le Borgne [25]. Dehling and Durieu [10], Proposition 3, show that the ergodic automorphisms of the torus satisfy the multiple mixing property (2.3) for functions of the class \mathcal{G} , and with the constants $\ell = 1$ and d_0 the size of the biggest Jordan block of T restricted to its neutral subspace. Thus, the $2p$ th moment bound (2.2) holds, and Theorem 2.1 can be applied to conclude. \square

5. Proof of the main theorem

In the proof of Theorem 2.1, we need a generalization of Theorem 4.2 of Billingsley [4]. Billingsley considers random variables $X_n, X_n^{(m)}, X^{(m)}, X, m, n \geq 1$, with values in a separable metric space (S, ρ) satisfying (a) $X_n^{(m)} \xrightarrow{\mathcal{D}} X^{(m)}$ as $n \rightarrow \infty$, for all $m \geq 1$, (b) $X^{(m)} \xrightarrow{\mathcal{D}} X$ as $m \rightarrow \infty$ and (c) $\forall \delta > 0, \limsup_{n \rightarrow \infty} P(\rho(X_n^{(m)}, X_n) \geq \delta) \rightarrow 0$ as $m \rightarrow \infty$. Theorem 4.2 of Billingsley [4] states that then $X_n \xrightarrow{\mathcal{D}} X$. Dehling, Durieu and Volný [11] proved that this result holds without condition (b), provided that S is a complete separable metric space. More precisely, they could show that in this situation (a) and (c) together imply the existence of a random variable X satisfying (b), and thus by Billingsley’s theorem $X_n \xrightarrow{\mathcal{D}} X$. Here, we will generalize this theorem to possibly non-measurable random elements with values in non-separable spaces. Regarding convergence in distribution of non-measurable random elements, we use the notation of van der Vaart and Wellner [34]. In accordance with the terminology of van der Vaart and Wellner [34], we will call a not necessarily measurable function with values in a measurable space a random element.

Theorem 5.1. Let $X_n, X_n^{(m)}, X^{(m)}, m, n \geq 1$, be random elements with values in a complete metric space (S, ρ) , and suppose that $X^{(m)}$ is measurable and separable, that is, there is a separable set $S^{(m)} \subset S$ such that $P(X^{(m)} \in S^{(m)}) = 1$. If the conditions

$$X_n^{(m)} \xrightarrow{\mathcal{D}} X^{(m)} \quad \text{as } n \rightarrow \infty, \text{ for all } m \geq 1, \tag{5.1}$$

$$\limsup_{n \rightarrow \infty} P^*(\rho(X_n, X_n^{(m)}) \geq \delta) \rightarrow 0 \quad \text{as } m \rightarrow \infty, \text{ for all } \delta > 0 \tag{5.2}$$

are satisfied, then there exists an S -valued, separable random variable X such that $X^{(m)} \xrightarrow{\mathcal{D}} X$ as $m \rightarrow \infty$, and

$$X_n \xrightarrow{\mathcal{D}} X \quad \text{as } n \rightarrow \infty.$$

The proof is postponed to the [Appendix](#).

Proof of Theorem 2.1. For all $q \geq 1$, there exist two sets of $N_q := N(2^{-q}, \exp(C2^{q/\gamma}), \mathcal{F}, \mathcal{G}, L^s(\mu))$ functions $\{g_{q,1}, \dots, g_{q,N_q}\} \subset \mathcal{G}$ and $\{g'_{q,1}, \dots, g'_{q,N_q}\} \subset \mathcal{G}$, such that $\|g_{q,i} - g'_{q,i}\|_s \leq 2^{-q}$, $\|g_{q,i}\|_{\mathcal{B}} \leq \exp(C2^{q/\gamma})$, $\|g'_{q,i}\|_{\mathcal{B}} \leq \exp(C2^{q/\gamma})$ and for all $f \in \mathcal{F}$, there exists an i such that $g_{q,i} \leq f \leq g'_{q,i}$. Further, by (2.4),

$$\sum_{q \geq 1} 2^{-(r+1)q} N_q^2 < +\infty. \tag{5.3}$$

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 $\mathcal{F}_{q,i} = [g_{q,i}, g'_{q,i}] \setminus (\bigcup_{j=1}^{i-1} \mathcal{F}_j)$.

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

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

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

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

We introduce

$$U_n^{(q)}(f) := U_n(\pi_q f) = \sqrt{n}(F_n^{(q)}(f) - \mu(\pi_q f)); \quad f \in \mathcal{F}.$$

Proposition 5.2. 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 $n \rightarrow \infty$.

Proof. Since $\pi_q f \in \mathcal{G}$ and \mathcal{G} is a subset of \mathcal{B} , by assumption (2.1), 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 $f_1, \dots, f_k \in \mathcal{F}$, $(U_n^{(q)}(f_1), \dots, U_n^{(q)}(f_k))$ converges in distribution to a Gaussian vector $(U^{(q)}(f_1), \dots, U^{(q)}(f_k))$ 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 $\tau_q: \mathbb{R}^{N_q} \rightarrow \ell^\infty(\mathcal{F})$ that maps a vector $x = (x_1, \dots, x_{N_q})$ to the function $\mathcal{F} \rightarrow \mathbb{R}$, $f \mapsto x_i$ such that $f \in \mathcal{F}_{q,i}$. For $f_1 \in \mathcal{F}_{q,1}, \dots, f_{N_q} \in \mathcal{F}_{q,N_q}$ we have $U_n^{(q)} = \tau_q(U_n^{(q)}(f_1), \dots, U_n^{(q)}(f_{N_q}))$ and thus the continuous mapping theorem guarantees that $U_n^{(q)}$ converges weakly to the random variable $U^{(q)} = \tau_q(U^{(q)}(f_1), \dots, U^{(q)}(f_{N_q}))$ which is constant on each $\mathcal{F}_{q,i}$. \square

Proposition 5.3. *For all $\varepsilon > 0$, $\eta > 0$ there exists a q_0 such that for all $q \geq q_0$*

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

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

$$|\bar{Y} - \bar{Y}_l| \leq |\bar{Y}_u - \bar{Y}_l| + \mathbb{E}|Y_u - Y_l|.$$

Using $F_n^{(q+K)}(f) \leq F_n(f) \leq F_n(\pi'_{q+K} f)$ and $\mathbb{E}|F_n(\pi'_{q+K} f) - F_n^{(q+K)}(f)| \leq 2^{-(q+K)}$ for all $f \in \mathcal{F}$, we obtain

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

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

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

For each $i \in \{1, \dots, N_q\}$, we obtain

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

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

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

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 by 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 $\mathcal{F}_{k-1,i} \cap \mathcal{F}_{k,j}$, $i = 1, \dots, N_{k-1}$, $j = 1, \dots, N_k$. Then, for each $i \in \{1, \dots, N_q\}$, we have

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

Now using Markov's inequality at the order $2p$ (p will be chosen later) and assumption (2.2), we infer

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

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 $k \geq 1$,

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

Thus,

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

and if q is large enough,

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

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

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

thus we have

$$\begin{aligned} &P^* \left(\sup_{f \in \mathcal{F}} |U_n(f) - U_n^{(q)}(f)| \geq \varepsilon \right) \\ &\leq D' \sum_{j=1}^p \frac{n^{j-p}}{\varepsilon^{2p}} \sum_{k=1}^{K_{n,q}} N_{q+k-1} N_{q+k} k^{4p} 2^{(2p+(a-\gamma)j)(q+k)/\gamma} \end{aligned} \tag{5.4}$$

$$\begin{aligned}
 &\leq D' \sum_{j=1}^p \frac{n^{j-p}}{\varepsilon^{2p}} 2^{(p-j)(\gamma+2+a)(q+K_{n,q})/\gamma} \\
 &\quad \times \sum_{k=1}^{K_{n,q}} N_{q+k-1} N_{q+k} k^4 p^2 2^{((-a-\gamma)p+(2+2a)j)(q+k)/\gamma} \\
 &\leq D'' \sum_{j=1}^{p-1} \frac{n^{(j-p)(\gamma-(2+a))/(2\gamma)}}{\varepsilon^{2p+(p-j)(\gamma+2+a)/\gamma}} \sum_{k=1}^{\infty} N_{q+k-1} N_{q+k} k^4 p^2 2^{(2+a-\gamma)p(q+k)/\gamma} \\
 &\quad + \frac{D'}{\varepsilon^{2p}} \sum_{k=1}^{\infty} N_{q+k-1} N_{q+k} k^4 p^2 2^{(2+a-\gamma)p(q+k)/\gamma},
 \end{aligned}$$

because $a \geq -1$ and thus $(2 + 2a)j \leq (2 + 2a)p$, and where D' and D'' are positive constants also depending on p, C , and C_p . As $p^{\frac{2+a-\gamma}{\gamma}} \rightarrow -\infty$ when p tends to infinity, there exists some $p > 1$ such that $p^{\frac{2+a-\gamma}{\gamma}} < -(r + 1)$ and thus by (5.3),

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

Therefore, the first summand of (5.4) goes to zero as n goes to infinity and the second summand of (5.4) goes to zero as q goes to infinity. □

Propositions 5.2 and 5.3 establish for the random elements $U_n, U_n^{(q)}, U^{(q)}$ with value in the complete metric space $\ell^\infty(\mathcal{F})$ conditions (5.1) and (5.2) of Theorem 5.1, respectively. Thus, Theorem 5.1 completes the proof of Theorem 2.1. □

Appendix

Proof of Lemma 3.1. By the triangle inequality, we have for all $x, y \in \mathcal{X}$ that

$$\begin{aligned}
 |d_B(x) - d_B(y)| &\leq d(x, y), \\
 d_B(x) + d_A(y) &\leq d(A, B).
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 &|T[A, B](x) - T[A, B](y)| \\
 &= \left| \frac{(d_B(x) - d_B(y))(d_B(y) + d_A(y)) + d_B(y)(d_B(y) + d_A(y)) - d_B(y)(d_B(x) + d_A(x))}{(d_B(x) + d_A(x))(d_B(y) + d_A(y))} \right|
 \end{aligned}$$

$$\begin{aligned}
 &= \left| \frac{d_B(x) - d_B(y)}{d_B(x) + d_A(x)} \right| + \frac{d_B(y)}{d_B(y) + d_A(y)} \left| \frac{(d_B(y) - d_B(x)) + (d_A(y) - d_A(x))}{d_B(x) + d_A(x)} \right| \\
 &\leq 3 \frac{d(x, y)}{d(A, B)}
 \end{aligned}$$

and thus

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

Proof of Lemma 3.7. 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, \dots, 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 $|D_v|_* = \max_{i=1, \dots, d} |v_i|$. We can characterize $E(0, \frac{j}{m})$ and $\mathbb{R}^d \setminus E(0, \frac{j}{m} + \frac{1}{m})$ by

$$E\left(0, \frac{j}{m}\right) = \{z \in \mathbb{R}^d: |D_{j/m}^{-1}z| \leq 1\}$$

and

$$\mathbb{R}^d \setminus E\left(0, \frac{j}{m} + \frac{1}{m}\right) = \{y \in \mathbb{R}^d: |D_{j/m+1/m}^{-1}y| > 1\},$$

respectively. Thus, for any $z \in E(0, \frac{j}{m})$ and $y \in \mathbb{R}^d \setminus E(0, \frac{j}{m} + \frac{1}{m})$,

$$\begin{aligned}
 |y - z| &\geq |D_{j/m+1/m}^{-1}|_*^{-1} |D_{j/m+1/m}^{-1}y - D_{j/m+1/m}^{-1}D_{j/m}D_{j/m}^{-1}z| \\
 &\geq |D_{j/m+1/m}^{-1}|_*^{-1} (|D_{j/m+1/m}^{-1}y| - |D_{j/m+1/m}^{-1}D_{j/m}|_* |D_{j/m}^{-1}z|) \\
 &> |D_{j/m+1/m}^{-1}|_*^{-1} (1 - |D_{j/m+1/m}^{-1}D_{j/m}|_*) \\
 &= \min_{j_i=1, \dots, d} \left\{ \frac{j}{m} + \frac{1}{m} \right\} \left(1 - \max_{i=1, \dots, d} \left\{ \frac{j_i/m}{j_i/m + 1/m} \right\} \right) \\
 &\geq \frac{1}{Dm^2}
 \end{aligned}$$

since $j_i \in \{0, \dots, Dm - 1\}$. □

Proof of Lemma 3.8. For any $\varepsilon > 0$, set $K_\varepsilon = \sup\{K > 0: \mu([-K, K]^d) \leq 1 - \varepsilon\}$. We will denote the function $(0, 1) \rightarrow \mathbb{R}^+$, $\varepsilon \mapsto K_\varepsilon$ by K_\bullet . Now, introduce the bracket $[L, U_\varepsilon]$, given by

$$L \equiv 0 \quad \text{and} \quad U_\varepsilon := T[\mathbb{R}^d \setminus [-K_{\varepsilon^s/2}, K_{\varepsilon^s/2}]^d, [-K_{\varepsilon^s}, K_{\varepsilon^s}]^d].$$

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

$$G(t) := \mu(\{x \in \mathbb{R}^d: |x|_{\max} \leq t\})$$

on \mathbb{R} , where $|x|_{\max} = \max\{|x_i|: i = 1, \dots, d\}$. 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

$$G(t) = \sum_{j \in \{-1, 1\}^d} \sigma(j)F(tj),$$

where $\sigma(j) := \prod_{i=1}^d j_i \in \{-1, 1\}$. Therefore,

$$\begin{aligned} \omega_G(x) &= \sup_{t \in \mathbb{R}} \{G(t+x) - G(t)\} = \sup_{t \in \mathbb{R}} \sum_{j \in \{-1, 1\}^d} \sigma(j)(F((t+x)j) - F(tj)) \\ &\leq \sum_{j \in \{-1, 1\}^d} \sup_{t \in \mathbb{R}} |F((t+x)j) - F(tj)| \leq \sum_{j \in \{-1, 1\}^d} \omega_F(\sqrt{d}x) \\ &\leq 2^d \omega_F(\sqrt{d}x). \end{aligned}$$

Now by Lemma 3.1 we obtain

$$\begin{aligned} \|U_\varepsilon\|_\alpha &\leq 1 + \frac{3^\alpha}{|G^{-1}(1 - \varepsilon^s/2) - G^{-1}(1 - \varepsilon^s)|^\alpha} \\ &\leq 1 + 3^\alpha \left(\inf \left\{ x > 0: \exists t \in \mathbb{R} \text{ such that } G(t+x) - G(t) \geq \frac{\varepsilon^s}{2} \right\} \right)^{-\alpha} \\ &\leq 1 + 3^\alpha \left(\inf \left\{ x > 0: \omega_G(x) \geq \frac{\varepsilon^s}{2} \right\} \right)^{-\alpha} \\ &\leq 1 + 3^\alpha \left(\sup \left\{ x \geq 0: \omega_F(\sqrt{d}x) \leq \frac{\varepsilon^s}{2^{d+1}} \right\} \right)^{-\alpha} \\ &= 1 + (3\sqrt{d})^\alpha (\omega_F^{-1}(2^{-(d+1)}\varepsilon^s))^{-\alpha}, \end{aligned}$$

where we used that ω_F is continuous here to replace the infimum by the supremum.

Then $[L, U_\varepsilon]$ is an $(\varepsilon, 4\sqrt{d}(\omega_F^{-1}(2^{-(d+1)}\varepsilon^s))^{-\alpha}, \mathcal{G}, L^s(\mu))$ -bracket for sufficiently small ε . Since $[L, U_\varepsilon]$ contains any $f \in \mathcal{F} \setminus \mathcal{F}_{K_{\varepsilon^s/2}+D}$, by (3.2) we obtain for all those ε the bound

$$N(\varepsilon, \max\{f(\varepsilon), 4\sqrt{d}(\omega_F^{-1}(2^{-(d+1)}\varepsilon^s))^{-\alpha}\}, \mathcal{F}, \mathcal{G}, L^s(\mu)) \leq C(K_{\varepsilon^s/2} + D)^p \varepsilon^{-q} + 1.$$

Let us finally consider the growth rate of $K_{\varepsilon^s/2}$ as $\varepsilon \rightarrow 0$. By assumption (3.3) 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. \square

Proof of Theorem 5.1. (i) We will first show that $X^{(m)}$ converges in distribution to some random variable X . We denote by $L^{(m)}$ the distribution of $X^{(m)}$; this is defined since $X^{(m)}$ is measurable. Moreover, $L^{(m)}$ is a separable Borel probability measure on S . By Theorem 1.12.4 of van der Vaart and Wellner [34], weak convergence of separable Borel measures on a metric space S can be metrized by the bounded Lipschitz metric, defined by

$$d_{\text{BL}_1}(L_1, L_2) = \sup_{f \in \text{BL}_1} \left| \int f(x) dL_1(x) - \int f(x) dL_2(x) \right|$$

for any Borel measures L_1, L_2 on S . Here, $\text{BL}_1 := \{f : S \rightarrow \mathbb{R} : \|f\|_{\text{BL}_1} \leq 1\}$, where

$$\|f\|_{\text{BL}_1} := \max \left\{ \sup_{x \in S} |f(x)|, \sup_{x \neq y \in S} \frac{f(x) - f(y)}{\rho(x, y)} \right\}.$$

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 $L^{(m)}$ is a d_{BL_1} -Cauchy sequence. We obtain

$$\begin{aligned} d_{\text{BL}_1}(L^{(m)}, L^{(l)}) &= \sup_{f \in \text{BL}_1} |\mathbb{E}f(X^{(m)}) - \mathbb{E}f(X^{(l)})| \\ &\leq \sup_{f \in \text{BL}_1} \{ |\mathbb{E}f(X^{(m)}) - \mathbb{E}^*f(X_n^{(m)})| + |\mathbb{E}^*f(X_n^{(m)}) - \mathbb{E}^*f(X_n)| \\ &\quad + |\mathbb{E}^*f(X_n) - \mathbb{E}^*f(X_n^{(l)})| + |\mathbb{E}^*f(X_n^{(l)}) - \mathbb{E}f(X^{(l)})| \} \end{aligned}$$

for all $n \in \mathbb{N}$. For a Borel measurable separable random element $X^{(m)}$ weak convergence $X_n^{(m)} \xrightarrow{\mathcal{D}} X^{(m)}$ as $n \rightarrow \infty$ is equivalent to $\sup_{f \in \text{BL}_1} |\mathbb{E}f(X^{(m)}) - \mathbb{E}^*f(X_n^{(m)})| \rightarrow 0$; see van der Vaart and Wellner [34], page 73. Hence by (5.1), we obtain

$$d_{\text{BL}_1}(L^{(m)}, L^{(l)}) \leq \liminf_{n \rightarrow \infty} \sup_{f \in \text{BL}_1} |\mathbb{E}^*f(X_n^{(m)}) - \mathbb{E}^*f(X_n)| + |\mathbb{E}^*f(X_n) - \mathbb{E}^*f(X_n^{(l)})|.$$

Using Lemma 1.2.2(iii) in van der Vaart and Wellner [34], we obtain

$$|\mathbb{E}^*f(X_n^{(m)}) - \mathbb{E}^*f(X_n)| \leq \mathbb{E}(|f(X_n) - f(X_n^{(m)})|^*)$$

and therefore

$$\begin{aligned} \sup_{f \in \text{BL}_1} |\mathbb{E}^* f(X_n^{(m)}) - \mathbb{E}^* f(X_n)| &\leq \mathbb{E}(\rho(X_n, X_n^{(m)}) \wedge 2)^* \\ &= \int_0^\infty P^*(\rho(X_n, X_n^{(m)}) \wedge 2 \geq t) dt, \end{aligned} \tag{A.1}$$

where we used the last statement of Lemma 1.2.2 in van der Vaart and Wellner [34]. Now, let $\varepsilon > 0$ be given. By (5.2), there exists an $m_0 \in \mathbb{N}$ such that for every $m \geq m_0$ there is some $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$ we have $P^*(\rho(X_n, X_n^{(m)}) \geq \varepsilon/3) \leq \varepsilon/3$. Therefore,

$$P^*(\rho(X_n, X_n^{(m)}) \wedge 2 \geq t) \leq \begin{cases} 1, & \text{if } t < \frac{\varepsilon}{3}, \\ \frac{\varepsilon}{3}, & \text{if } \frac{\varepsilon}{3} \leq t \leq 2, \\ 0, & \text{if } 2 < t. \end{cases}$$

Applying this inequality to (A.1), we obtain

$$\liminf_{n \rightarrow \infty} \sup_{f \in \text{BL}_1} |\mathbb{E}^* f(X_n^{(m)}) - \mathbb{E}^* f(X_n)| \leq \int_0^2 \frac{\varepsilon}{3} + 1_{\{t < \varepsilon/3\}} dt = \varepsilon$$

for all $m \geq m_0$. Hence for $l, m \geq m_0$ we have $d_{\text{BL}_1}(L^{(m)}, L^{(l)}) \leq 2\varepsilon$; that is, $(L^{(m)})_{m \in \mathbb{N}}$ is a d_{BL_1} -Cauchy sequence in a complete metric space.

(ii) The remaining part of the proof follows closely the proof of Theorem 4.2 in Billingsley [4], replacing the probability measure P by the outer measure P^* where necessary and making use of the Portmanteau theorem; see van der Vaart and Wellner [34], Theorem 1.3.4(iii), and the sub-additivity of outer measures. From part (i), we already know that there is some measurable X such that $X^{(m)} \xrightarrow{\mathcal{D}} X$. Let $F \subset S$ be closed. Given $\varepsilon > 0$, we define the ε -neighborhood $F_\varepsilon := \{s \in S: \inf_{x \in F} \rho(s, x) \leq \varepsilon\}$, and observe that F_ε is also closed. Since $\{X_n \in F\} \subset \{X_n^{(m)} \in F_\varepsilon\} \cup \{\rho(X_n^{(m)}, X_n) \geq \varepsilon\}$, we obtain

$$P^*(X_n \in F) \leq P^*(X_n^{(m)} \in F_\varepsilon) + P^*(\rho(X_n^{(m)}, X_n) \geq \varepsilon)$$

for all $m \in \mathbb{N}$. By (5.2) we may choose m_0 so large that for all $m \geq m_0$

$$\limsup_{n \rightarrow \infty} P^*(\rho(X_n^{(m)}, X_n) \geq \varepsilon) \leq \varepsilon/2.$$

As $X^{(m)} \xrightarrow{\mathcal{D}} X$, by the Portmanteau theorem we may choose m_1 so large that for all $m \geq m_1$

$$P(X^{(m)} \in F_\varepsilon) \leq P(X \in F_\varepsilon) + \varepsilon/2.$$

We now fix $m \geq \max(m_0, m_1)$. By (5.1) we have $X_n^{(m)} \xrightarrow{\mathcal{D}} X^{(m)}$ as $n \rightarrow \infty$. Thus an application of the Portmanteau theorem yields

$$\limsup_{n \rightarrow \infty} P^*(X_n^{(m)} \in F_\varepsilon) \leq P(X^{(m)} \in F_\varepsilon),$$

$$\limsup_{n \rightarrow \infty} P^*(X_n \in F) \leq P(X \in F_\varepsilon) + \varepsilon.$$

Since this holds for any $\varepsilon > 0$ and $\lim_{\varepsilon \rightarrow 0} P(X \in F_\varepsilon) = P(X \in F)$, we get

$$\limsup_{n \rightarrow \infty} P^*(X_n \in F) \leq P(X \in F)$$

for all closed sets $F \subset S$. By a final application of the Portmanteau theorem we infer $X_n \xrightarrow{D} X$. \square

Acknowledgments

The authors thank the referees and the Associate Editor for their very careful reading and for their most constructive criticism of an earlier version of this paper that greatly helped to improve the presentation.

Research supported in part by German Research Foundation Grant DE 370-4 *New Techniques for Empirical Processes of Dependent Data*, and the Collaborative Research Grant SFB 823 *Statistical Modeling of Nonlinear Dynamic Processes*.

References

- [1] Andrews, D.W. and Pollard, D. (1994). An introduction to functional central limit theorems for dependent stochastic processes. *International Statistical Review* **62** 119–132.
- [2] Beutner, E. and Zähle, H. (2012). Deriving the asymptotic distribution of U- and V-statistics of dependent data using weighted empirical processes. *Bernoulli* **18** 803–822. [MR2948902](#)
- [3] Bickel, P.J. and Wichura, M.J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Statist.* **42** 1656–1670. [MR0383482](#)
- [4] Billingsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley. [MR0233396](#)
- [5] Borovkova, S., Burton, R. and Dehling, H. (2001). Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation. *Trans. Amer. Math. Soc.* **353** 4261–4318. [MR1851171](#)
- [6] Bradley, R.C. (2007). *Introduction to Strong Mixing Conditions*. Vols 1–3. Heber City, UT: Kendrick Press. [MR2325294](#), [MR2325295](#), [MR2325296](#)
- [7] Collet, P., Martinez, S. and Schmitt, B. (2004). Asymptotic distribution of tests for expanding maps of the interval. *Ergodic Theory Dynam. Systems* **24** 707–722. [MR2060995](#)
- [8] Dedecker, J., Doukhan, P., Lang, G., León R., J.R., Louhichi, S. and Prieur, C. (2007). *Weak Dependence: With Examples and Applications*. *Lecture Notes in Statistics* **190**. New York: Springer. [MR2338725](#)
- [9] Dedecker, J. and Prieur, C. (2007). An empirical central limit theorem for dependent sequences. *Stochastic Process. Appl.* **117** 121–142. [MR2287106](#)
- [10] Dehling, H. and Durieu, O. (2011). Empirical processes of multidimensional systems with multiple mixing properties. *Stochastic Process. Appl.* **121** 1076–1096. [MR2775107](#)
- [11] Dehling, H., Durieu, O. and Volný, D. (2009). New techniques for empirical processes of dependent data. *Stochastic Process. Appl.* **119** 3699–3718. [MR2568292](#)

- [12] Dehling, H. and Philipp, W. (2002). Empirical process techniques for dependent data. In *Empirical Process Techniques for Dependent Data* 3–113. Boston, MA: Birkhäuser. [MR1958777](#)
- [13] Dhompongsa, S. (1984). A note on the almost sure approximation of the empirical process of weakly dependent random vectors. *Yokohama Math. J.* **32** 113–121. [MR0772909](#)
- [14] Dolgopyat, D. (2004). Limit theorems for partially hyperbolic systems. *Trans. Amer. Math. Soc.* **356** 1637–1689 (electronic). [MR2034323](#)
- [15] Donsker, M.D. (1952). Justification and extension of Doob’s heuristic approach to the Komogorov–Smirnov theorems. *Ann. Math. Statist.* **23** 277–281. [MR0047288](#)
- [16] Doob, J.L. (1949). Heuristic approach to the Kolmogorov–Smirnov theorems. *Ann. Math. Statist.* **20** 393–403. [MR0030732](#)
- [17] Doukhan, P. and Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Process. Appl.* **84** 313–342. [MR1719345](#)
- [18] Doukhan, P., Massart, P. and Rio, E. (1995). Invariance principles for absolutely regular empirical processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **31** 393–427. [MR1324814](#)
- [19] Dudley, R.M. (1966). Weak convergences of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces. *Illinois J. Math.* **10** 109–126. [MR0185641](#)
- [20] Dudley, R.M. (1978). Central limit theorems for empirical measures. *Ann. Probab.* **6** 899–929. [MR0512411](#)
- [21] Durieu, O. (2008). A fourth moment inequality for functionals of stationary processes. *J. Appl. Probab.* **45** 1086–1096. [MR2484163](#)
- [22] Durieu, O. and Jouan, P. (2008). Empirical invariance principle for ergodic torus automorphisms; genericity. *Stoch. Dyn.* **8** 173–195. [MR2429199](#)
- [23] Durieu, O. and Tusche, M. (2012). An empirical process central limit theorem for multidimensional dependent data. *J. Theoret. Probab.* DOI:10.1007/s10959-012-0450-3.
- [24] Hennion, H. and Hervé, L. (2001). *Limit Theorems for Markov Chains and Stochastic Properties of Dynamical Systems by Quasi-Compactness. Lecture Notes in Math.* **1766**. Berlin: Springer. [MR1862393](#)
- [25] Le Borgne, S. (1999). Limit theorems for non-hyperbolic automorphisms of the torus. *Israel J. Math.* **109** 61–73. [MR1679589](#)
- [26] Leonov, V.P. (1960). On the central limit theorem for ergodic endomorphisms of compact commutative groups. *Dokl. Akad. Nauk SSSR* **135** 258–261. [MR0171302](#)
- [27] Neuhaus, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter. *Ann. Math. Statist.* **42** 1285–1295. [MR0293706](#)
- [28] Nolan, D. and Pollard, D. (1987). U -processes: Rates of convergence. *Ann. Statist.* **15** 780–799. [MR0888439](#)
- [29] Ossiander, M. (1987). A central limit theorem under metric entropy with L_2 bracketing. *Ann. Probab.* **15** 897–919. [MR0893905](#)
- [30] Philipp, W. (1984). Invariance principles for sums of mixing random elements and the multivariate empirical process. In *Limit Theorems in Probability and Statistics, Vols I, II (Veszprém, 1982). Colloquia Mathematica Societatis János Bolyai* **36** 843–873. Amsterdam: North-Holland. [MR0807587](#)
- [31] Philipp, W. and Pinzur, L. (1980). Almost sure approximation theorems for the multivariate empirical process. *Z. Wahrsch. Verw. Gebiete* **54** 1–13. [MR0595473](#)
- [32] Rio, E. (1998). Processus empiriques absolument réguliers et entropie universelle. *Probab. Theory Related Fields* **111** 585–608. [MR1641838](#)
- [33] Straf, M.L. (1972). 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* 187–221. Berkeley, CA: Univ. California Press. [MR0402847](#)

- [34] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. New York: Springer. [MR1385671](#)

Received February 2012 and revised November 2012