

# Inferring the mixing properties of a stationary ergodic process from a single sample-path

Azadeh Khaleghi\*, Gábor Lugosi†

**Abstract**—We propose strongly consistent estimators of the  $\ell_1$  norm of the sequence of  $\alpha$ -mixing (respectively  $\beta$ -mixing) coefficients of a stationary ergodic process. We further provide strongly consistent estimators of individual  $\alpha$ -mixing (respectively  $\beta$ -mixing) coefficients for a subclass of stationary  $\alpha$ -mixing (respectively  $\beta$ -mixing) processes with summable sequences of mixing coefficients. The estimators are in turn used to develop strongly consistent goodness-of-fit hypothesis tests. In particular, we develop hypothesis tests to determine whether, under the same summability assumption, the  $\alpha$ -mixing (respectively  $\beta$ -mixing) coefficients of a process are upper bounded by a given rate function. Moreover, given a sample generated by a (not necessarily mixing) stationary ergodic process, we provide a consistent test to discern the null hypothesis that the  $\ell_1$  norm of the sequence  $\alpha$  of  $\alpha$ -mixing coefficients of the process is bounded by a given threshold  $\gamma \in [0, \infty)$  from the alternative hypothesis that  $\|\alpha\| > \gamma$ . An analogous goodness-of-fit test is proposed for the  $\ell_1$  norm of the sequence of  $\beta$ -mixing coefficients of a stationary ergodic process. Moreover, the procedure gives rise to an asymptotically consistent test for independence.

**Index Terms**—stationary ergodic process, mixing coefficients, long-range dependence, consistency, estimation, hypothesis testing

## I. INTRODUCTION

**M**IXING is a fundamental notion in the theory of stochastic processes. Roughly speaking, a stochastic process  $\mathbf{X} = \langle X_t \rangle_{t \in \mathbb{N}}$  indexed by “time” is mixing if events separated by long time intervals are approximately independent. There are various notions to quantify such asymptotic independence, including  $\alpha$ -mixing,  $\beta$ -mixing,  $\phi$ -mixing,  $\rho$ -mixing,  $\psi$ -mixing; see Bradley [1] for a general survey. These notions of mixing are particularly useful in time-series analysis where non-asymptotic concentration inequalities are available for empirical averages of stationary processes, see, for example, Ibragimov [2], Doukhan [3], Rio [4], Bertail, Doukhan, and Soulier [5], Bradley [1], Bosq [6]. In order to be able to take advantage of these tools, in statistical studies one often assumes that the process is not only mixing but that the mixing coefficients are such that the desired concentration inequalities hold. Despite the widespread use of such assumptions, little attention has been paid to validating these conditions. For instance, in the context of sequential decision making in the presence of long-range dependencies, Grünewälder and Khaleghi [7] propose algorithms with finite-time guarantees which require known upper-bounds on the  $\ell_1$  norm of the

$\varphi$ -mixing coefficients of the pay-off distributions. They also show (see, [7, Proposition 9]) that individual mixing coefficients naturally appear in theoretical bounds on the quality of approximation of the optimal strategies by simpler and more computationally tractable approaches. Indeed, consistent estimators of mixing coefficients can serve as indispensable tools in modern statistics.

In this paper, we study the problem of estimating  $\alpha$  and  $\beta$ -mixing coefficients. More precisely, we address the following problem:

*Given a finite sample-path of a real-valued, stationary ergodic process, how can one consistently estimate the  $\alpha$ -mixing and  $\beta$ -mixing coefficients of the process?*

Since the  $\ell_1$  norm of the sequence of these coefficients plays an important role in quantifying how much sample averages differ from their expectations (see, e.g. [4, Corollary 1.1]), we pay special attention to estimating the  $\ell_1$  norm. Our main results show that consistent estimation of these norms is indeed possible, under the only assumption that the process is stationary ergodic. (If the process is not mixing, the estimators diverge to infinity.) We also show how these estimates can be used to derive consistent hypothesis tests on the mixing coefficients.

The main difficulty of the estimation problem stems from the fact that all mixing coefficients are inherently asymptotic quantities, yet one only has a finite sample available. This makes estimation a nontrivial task, especially when no a-priori properties are assumed apart from stationarity and ergodicity of the process. Our focus in this paper is on  $\alpha$  and  $\beta$ -mixing coefficients, as these are arguably the most widely used notions of mixing with numerous statistical applications. One may, of course, ask the analogous questions on other measures of mixing. A particularly interesting notion is  $\phi$  (or uniform) mixing (see, e.g. Bradley [1] for a definition) since under such conditions one has Hoeffding-type exponential inequalities in terms of the  $\ell_1$  norm of the sequence of  $\phi$ -mixing coefficients, see Rio [4, Corollary 2.1] and Samson [8]. However, it is unclear if the  $\ell_1$  norm of the sequence of  $\phi$ -mixing coefficients can be consistently estimated. In this case, a key challenge lies in conditioning on potentially rare events whose probabilities may be arbitrarily small. We leave this interesting challenge for future research.

## A. $\alpha$ and $\beta$ mixing

We start by defining the notions of mixing relevant to this paper. Let  $(\Omega, \mathfrak{F}, \mu)$  be a probability space, and suppose that

\* ENSAE - CREST, Institut Polytechnique de Paris, 5 av. Le Chatelier, 91120 Palaiseau, France, azadeh.khaleghi@ensae.fr

† Department of Economics and Business, Pompeu Fabra University; ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain; Barcelona Graduate School of Economics, gabor.lugosi@gmail.com

$\mathfrak{U}$  and  $\mathfrak{V}$  are two  $\sigma$ -subalgebras of  $\mathfrak{F}$  respectively. A classical measure of dependence between  $\mathfrak{U}$  and  $\mathfrak{V}$ , introduced by Rosenblatt [9], is given by

$$\alpha(\mathfrak{U}, \mathfrak{V}) := \sup_{U \in \mathfrak{U}, V \in \mathfrak{V}} |\mu(U \cap V) - \mu(U)\mu(V)|$$

which gives rise to the sequence of dependence coefficients  $\alpha := \langle \alpha(m) \rangle_{m \in \mathbb{N}}$  (called the  $\alpha$ -mixing coefficients) of a stochastic process  $\mathbf{X} = \langle X_t \rangle_{t \in \mathbb{N}}$ , where  $\alpha(m)$  is defined as,

$$\sup_{j \in \mathbb{N}} \alpha(\sigma(\{X_t : 1 \leq t \leq j\}), \sigma(\{X_t : t \geq j + m\})).$$

The  $\beta$ -dependence  $\beta(\mathfrak{U}, \mathfrak{V})$  between  $\mathfrak{U}$  and  $\mathfrak{V}$  was originally introduced by Volkonskii and Rozanov [10], [11] as follows (see also Rio [4]). Let  $\iota(\omega) \mapsto (\omega, \omega)$  be the injection map from  $(\Omega, \mathfrak{F})$  to  $(\Omega \times \Omega, \mathfrak{U} \otimes \mathfrak{V})$ , where  $\mathfrak{U} \otimes \mathfrak{V}$  is the product sigma algebra generated by  $\mathfrak{U} \times \mathfrak{V}$ . Let  $\mu_{\otimes}$  be the probability measure defined on  $(\Omega \times \Omega, \mathfrak{U} \otimes \mathfrak{V})$  obtained as the pushforward measure of  $\mu$  under  $\iota$ . Let  $\mu_{\mathfrak{U}}$  and  $\mu_{\mathfrak{V}}$  denote the restriction of  $\mu$  to  $\mathfrak{U}$  and  $\mathfrak{V}$  respectively. Then

$$\beta(\mathfrak{U}, \mathfrak{V}) := \sup_{W \in \sigma(\mathfrak{U} \times \mathfrak{V})} |\mu_{\otimes}(W) - \mu_{\mathfrak{U}} \times \mu_{\mathfrak{V}}(W)|,$$

where  $\mu_{\mathfrak{U}} \times \mu_{\mathfrak{V}}$  is the product measure on  $(\Omega \times \Omega, \mathfrak{U} \otimes \mathfrak{V})$  obtained from  $\mu_{\mathfrak{U}}$  and  $\mu_{\mathfrak{V}}$ . This measure of dependence leads to the sequence  $\beta := \langle \beta(m) \rangle_{m \in \mathbb{N}}$  of  $\beta$ -mixing coefficients of a process  $\mathbf{X}$ , where  $\beta(m)$  is given by

$$\sup_{j \in \mathbb{N}} \beta(\sigma(\{X_t : 1 \leq t \leq j\}), \sigma(\{X_t : t \geq j + m\})).$$

A stochastic process is said to be  $\alpha$ -mixing or *strongly mixing*, if it exhibits an asymptotic independence of the form  $\lim_{m \rightarrow \infty} \alpha(m) = 0$ . It is said to be  $\beta$ -mixing or *absolutely regular* if  $\lim_{m \rightarrow \infty} \beta(m) = 0$ . It is straightforward to check that  $\beta(\mathfrak{U}, \mathfrak{V}) \geq \alpha(\mathfrak{U}, \mathfrak{V})$  so that absolute regularity implies strong mixing. Moreover, summability of  $\alpha$  (respectively  $\beta$ ) is clearly a sufficient but not necessary condition for a process to be  $\alpha$ -mixing (respectively  $\beta$ -mixing).

### B. The estimation problem

Regardless of whether or not a process is mixing, its mixing coefficients provide a measure of its dependence structure. Moreover, as mentioned above, individual  $\alpha$ -mixing (respectively  $\beta$ -mixing) coefficients and/or their  $\ell_1$ -norm  $\|\alpha\| = \sum_{m \in \mathbb{N}} \alpha(m)$  (respectively  $\|\beta\| = \sum_{m \in \mathbb{N}} \beta(m)$ ) commonly appear in concentration inequalities for dependent processes. Thus, in order to use these bounds in a statistical problem where the samples may be dependent, knowledge of the sequences  $\alpha$ ,  $\beta$  or at least of their  $\ell_1$  norms is required.

For certain subclasses of dependent processes, bounds on the mixing coefficients are known. For example, conditions for the geometric ergodicity of Markov chains have been well studied, see Meyn and Tweedie [12] and references therein. More recently, some upper bounds on the mixing rates of non-stationary ARCH processes were proposed by Fryzlewicz and Rao [13]. However, for larger classes of stationary processes, the mixing coefficients are typically unknown, and surprisingly little research has been devoted to the problem of estimating mixing coefficients. One exception is the work of McDonald,

Shalizi, and Schervish [14], [15] who provide estimators for  $\beta$ -mixing coefficients and show consistency of their estimators. Unfortunately, due to a lack of precision in the presentation, we were unable to verify some of the main claims of these papers. Our work does not build upon these results and we consider a more general setting where, apart from ergodicity, no assumptions are required on the process distributions (as opposed to some implicit assumptions on the existence and smoothness of finite-dimensional densities made in [14], [15]).

Towards an adaptive approach, given a sample generated by a stationary ergodic process, our objective in this paper is to estimate its dependence structure as reflected by its  $\alpha$ -mixing and  $\beta$ -mixing coefficients. To this end, in Section III-A, we first focus on a subclass of stationary ergodic processes which are  $\alpha$ -mixing with the additional property that their sequences of  $\alpha$ -mixing coefficients are summable. For this class, without knowledge of any upper-bounds on  $\|\alpha\|$  and merely using the fact that  $\|\alpha\| < \infty$ , we provide asymptotically consistent estimators of the individual  $\alpha$ -mixing coefficients as well as for  $\|\alpha\|$ . The consistency results for these estimators are established via Theorem 1 and Theorem 2 respectively. We rely on Rio's covariance inequality [4, Corollary 1.1] to control the variance of partial sums. Next, we propose an alternative approach for consistently estimating  $\|\alpha\|$  of a (not necessarily mixing) stationary ergodic process where we no longer require  $\alpha$  to be summable. In this case, if the process happens to be  $\alpha$ -mixing with  $\|\alpha\| < \infty$ , then the proposed estimator converges to  $\|\alpha\|$ , otherwise it diverges to infinity. The weak and strong consistency properties of these estimators follow from Theorems 3 and 4. The results in this section rely on Lemma 1 and Proposition 1 which together show that the approximation given by (4) converges to  $\alpha(m)$ .

In Section III-B we provide analogous results for the sequence of  $\beta$ -mixing coefficients of a stationary ergodic process. Most of the arguments are similar to those given in Section III-A, and in particular since  $\alpha(m) \leq \beta(m)$ , the same covariance inequality is sufficient to control the variance of partial sums in the estimation. However, a key challenge in this case is that unlike  $\alpha(\mathfrak{U}, \mathfrak{V})$ , the  $\beta$ -dependence between  $\mathfrak{U}$  and  $\mathfrak{V}$  is defined for the product space  $(\Omega \times \Omega, \mathfrak{U} \otimes \mathfrak{V}, \mu_{\otimes})$ . In order to propose an approximation that can be estimated from a single sample-path, we rely on the identity,

$$\beta(\mathfrak{U}, \mathfrak{V}) = \sup \frac{1}{2} \sum_{i \in \mathbb{N}} \sum_{j \in \mathbb{N}} |\mu(U_i \cap V_j) - \mu(U_i)\mu(V_j)|$$

where the supremum is taken over all pairs of countable partitions  $\{U_1, U_2, U_3, \dots\}$  and  $\{V_1, V_2, V_3, \dots\}$  of  $\Omega$  such that  $U_i \in \mathfrak{U}$  and  $V_j \in \mathfrak{V}$  for each  $i, j \in \mathbb{N}$ ; see Bradley [1, Vol. 1, P. 67, Note 2]. This, together with [1, Vol. 1 Proposition 3.21] gives rise to Lemma 3 which in turn leads to Proposition 4. Analogues of Lemma 1 and Proposition 1, these results show that the approximation (15) that is based on the cylinder sets converges to  $\beta(m)$ .

### C. Hypothesis testing

In Section IV we show how our estimators can be used to construct goodness-of-fit tests. First, given a sample generated

by a stationary  $\alpha$ -mixing (respectively  $\beta$ -mixing) process  $\mu$ , we use our estimator of  $\alpha(m)$  (respectively  $\beta(m)$ ) to test the null hypothesis that the  $\alpha$ -mixing (respectively  $\beta$ -mixing) coefficients of  $\mu$  are bounded by a given *rate function*  $\gamma : \mathbb{N} \mapsto [0, 1]$  against the alternative hypothesis  $H_1$  that there exists some  $m \in \mathbb{N}$  such that  $\alpha(m) > \gamma(m)$  (respectively  $\beta(m) > \gamma(m)$ ). The consistency of these tests follow from Theorems 7 and 8, provided that the sequence of  $\alpha$ -mixing (respectively  $\beta$ -mixing) coefficients of the process is summable. Interestingly, Nobel [16] used hypothesis testing to estimate polynomial decay rates for covariance-based mixing conditions. By contrast, we do not require the rate functions to be polynomial or to belong to any specific function class. Moreover, given a sample generated by a (not necessarily mixing) stationary ergodic process  $\mu$ , we construct tests to discern the null hypothesis  $H_0$  that  $\|\alpha\|$  (respectively  $\|\beta\|$ ) is bounded by a given threshold  $\gamma \in [0, \infty)$  from the alternative hypothesis  $H_1$  that it exceeds  $\gamma$ . The consistency of these procedures follow from Theorem 9 (for  $\alpha$ -mixing coefficients) and Theorem 10 (for  $\beta$ -mixing coefficients). As a direct consequence, we obtain strongly consistent tests for independence, obtaining an alternative proof of the main result of Morvai and Weiss [17]; this is stated as Corollary 1.

## II. PRELIMINARIES

In this section we fix our notation and introduce some basic definitions. Let  $\mathcal{X} = [0, 1]$  and denote by  $\mathfrak{B}_{\mathcal{X}}$  its Borel  $\sigma$ -algebra. Denote by  $\mathfrak{B}^{(k)}$  the product  $\sigma$ -algebra on  $\mathcal{X}^k$ ,  $k \in \mathbb{N}$ . Let  $\mathcal{X}^{\mathbb{N}}$  be the set of all  $\mathcal{X}$ -valued infinite sequences indexed by  $\mathbb{N}$ . A (discrete-time) stochastic process is a probability measure  $\mu$  on the space  $(\mathcal{X}^{\mathbb{N}}, \mathfrak{F})$  where  $\mathfrak{F}$  denotes the Borel  $\sigma$ -algebra on  $\mathcal{X}^{\mathbb{N}}$  generated by the cylinder sets. Associated with the process is a sequence of random variables  $\mathbf{X} := \langle X_t \rangle_{t \in \mathbb{N}}$  with process distribution  $\mu$  where  $X_t : \mathcal{X}^{\mathbb{N}} \rightarrow \mathcal{X}$  are coordinate projections such that  $X_t(\mathbf{a}) = a_t$  for  $\mathbf{a} = \langle a_t \rangle_{t \in \mathbb{N}} \in \mathcal{X}^{\mathbb{N}}$ . We use the term (*stochastic process*) to refer to either the measure  $\mu$  or to its corresponding sequence of random variables  $\mathbf{X}$ ; the distinction should be clear from the context. A process is  $\mathbf{X}$  is stationary if for all  $i, k \in \mathbb{N}$  and all  $B \in \mathfrak{B}^{(k)}$  we have  $\Pr((X_1, \dots, X_k) \in B) = \Pr((X_{1+i}, \dots, X_{k+i}) \in B)$ . Let  $S : \mathcal{X}^{\mathbb{N}} \rightarrow \mathcal{X}^{\mathbb{N}}$  denote the (left) shift transformation on  $\mathcal{X}^{\mathbb{N}}$  which maps  $\mathbf{a} := (a_1, a_2, \dots) \in \mathcal{X}^{\mathbb{N}}$  to  $S\mathbf{a} = (a_2, a_3, \dots)$ . It is continuous relative to the product topology on  $\mathcal{X}^{\mathbb{N}}$  and defines a set transformation  $S^{-1}$  given by  $S^{-1}A := \{\mathbf{a} \in \mathcal{X}^{\mathbb{N}} : S\mathbf{a} \in A\}$ ,  $A \subseteq \mathcal{X}^{\mathbb{N}}$ . Thus, it is straightforward to check that  $S^{-1}$  is Borel measurable and that stationarity of  $\mathbf{X}$  translates to the condition that  $\mu(S^{-1}B) = \mu(B)$  for all  $B \in \mathfrak{F}$ . A stationary process is ergodic if every shift-invariant measurable set has measure 0 or 1 so that if  $S^{-1}B = B$  for some  $B \in \mathfrak{F}$  then  $\mu(B) \in \{0, 1\}$ . Recalling the definition of a strongly mixing process given in the introduction, note the well-known fact that a stationary mixing process is ergodic, see, for example, Bradley [1].

The following notation will be used throughout the paper. Let  $\Delta_{k,\ell}$  be the set of dyadic cubes in  $\mathcal{X}^k$ ,  $k \in \mathbb{N}$  of side-

length  $2^{-\ell}$ . That is,

$$\Delta_{k,\ell} := \left\{ \left[ \frac{i_1}{2^\ell}, \frac{i_1+1}{2^\ell} \right) \times \dots \times \left[ \frac{i_k}{2^\ell}, \frac{i_k+1}{2^\ell} \right), \right. \\ \left. \text{where } i_j \in \{0, \dots, 2^\ell - 1\}, j \in \{1, \dots, k\} \right\} \quad (1)$$

For each  $k, \ell \in \mathbb{N}$  we denote by  $\mathcal{D}_{k,\ell} := \mathcal{P}(\Delta_{k,\ell})$  the power-set of  $\Delta_{k,\ell}$ . For a given  $B \in \mathfrak{B}^{(k)}$ ,  $k \in \mathbb{N}$ , we denote the event  $\{X_i, \dots, X_{i+k-1} \in B\}$ ,  $i \in \mathbb{N}$  by  $[B]_i^{i+k}$ . For  $m, \ell \in \mathbb{N}$ ,  $n > m \in \mathbb{N}$  and each  $j \in \{1, \dots, n-m\}$  define the  $\sigma$  algebra generated by the sets  $[A]_1^j$  for  $A \in \mathcal{D}_{j,\ell}$  by

$$\mathfrak{F}_1^j(\ell) := \sigma \left( \left\{ [A]_1^j : A \in \mathcal{D}_{j,\ell} \right\} \right) \quad (2)$$

and similarly let

$$\mathfrak{F}_{j+m}^n(\ell) := \sigma \left( \left\{ [B]_{j+m+1}^n : B \in \mathcal{D}_{j',\ell}, \right. \right. \\ \left. \left. j' := n - m - j \right\} \right) \quad (3)$$

**Remark 1** (Choice of  $\mathcal{X}$ ). In order to keep the notation simple and avoid uninteresting technicalities, we have chosen to take  $\mathcal{X} = [0, 1]$ . However, extensions to more general spaces including to  $\mathbb{R}$  and  $\mathbb{R}^d$  are straightforward. For instance, an extension to  $\mathcal{X} = \mathbb{R}$  can be readily achieved by modifying the discretization in (1) such that instead of  $[0, 1]^k$ , sets of the form  $[-2^\ell, 2^\ell]^k$  are partitioned into cubes of side-length  $2^{-\ell}$ .

## III. ESTIMATION

Consider a stationary ergodic process  $\mu$  with corresponding sequence of random variables  $\mathbf{X} = \langle X_t \rangle_{t \in \mathbb{N}}$  and sequences of  $\alpha$ -mixing and  $\beta$ -mixing coefficients  $\alpha := \langle \alpha(m) \rangle_{m \in \mathbb{N}}$  and  $\beta := \langle \beta(m) \rangle_{m \in \mathbb{N}}$  respectively. In this section we introduce estimators of  $\|\alpha\|$  and  $\alpha(m)$ , as well as of  $\|\beta\|$  and  $\beta(m)$ , and establish their consistency; these are the main results of the paper.

We consider the somewhat simpler case of  $\alpha$ -mixing coefficients in Section III-A. The corresponding estimators for  $\beta$ -mixing coefficients are given Section III-B. For the most part, the estimators – and proofs – concerning the  $\beta$ -mixing coefficients are analogous to their  $\alpha$ -mixing counterparts. However, as discussed in Sections I-A and I-B, there is a subtle distinction between the measurable spaces on which the two mixing coefficients are defined. This calls for a slightly different treatment in the case of  $\beta$ -mixing coefficients, giving rise to a number of technical challenges which we address as part of our analysis in Section III-B.

### A. Approximating and estimating the $\alpha$ -mixing quantities

Since the mixing coefficients  $\alpha(m)$  of  $\mathbf{X}$  are inherently asymptotic, we first approximate them by quantities that only depend on finite-dimensional projections of the process and can therefore be estimated from finite sample-paths. Subsection III-A1 below is dedicated to this task. Equipped with the approximation results, we present our estimators for the  $\alpha$ -mixing parameters in Subsection III-A2. The proofs and technical results are deferred to Section VI.

1) *Approximating  $\alpha(m)$  and  $\|\alpha\|$* : For each  $m, \ell \in \mathbb{N}$ ,  $n > m \in \mathbb{N}$  and  $j \in \{1, \dots, n - m\}$ , define  $\alpha_{n,j}^\ell(m)$  as

$$\sup_{\substack{A \in \mathcal{D}_{j,\ell} \\ B \in \mathcal{D}_{j',\ell}}} \left| \mu([A]_1^j \cap [B]_{j+m+1}^n) - \mu([A]_1^j) \mu([B]_{j+m+1}^n) \right| \quad (4)$$

where  $j' := n - m - j + 1$ . In Lemma 1 below we show that to approximate the  $\alpha$ -dependence between  $\mathfrak{F}_1^j(\ell)$  and  $\mathfrak{F}_{j+m}^n(\ell)$ , it suffices to use  $\alpha_{n,j}^\ell(m)$  where the supremum is taken over the smaller classes of sets  $\mathcal{D}_{j,\ell}$  and  $\mathcal{D}_{j',\ell}$ , with  $j' := n - m - j$ .

**Lemma 1.** For  $m, \ell \in \mathbb{N}$ ,  $n > m \in \mathbb{N}$  and  $j \in \{1, \dots, n - m\}$  we have

$$\alpha_{n,j}^\ell(m) = \alpha(\mathfrak{F}_1^j(\ell), \mathfrak{F}_{j+m}^n(\ell)).$$

For  $m, \ell \in \mathbb{N}$ ,  $n > m \in \mathbb{N}$  let  $\alpha_n^\ell(m) = \max_{j \in \{1, \dots, n - m\}} \alpha_{n,j}^\ell(m)$ . The next proposition shows that  $\alpha_n^\ell(m)$  approximates  $\alpha(m)$  for sufficiently large values of  $\ell$  and  $n$ .

**Proposition 1.** For every  $m \in \mathbb{N}$  we have

$$\lim_{n, \ell \rightarrow \infty} \alpha_n^\ell(m) = \alpha(m).$$

2) *Estimating  $\alpha(m)$  and  $\|\alpha\|$* : We are now in a position to introduce the natural empirical estimates of the approximate mixing coefficients  $\alpha_n^\ell(m)$ . For  $t \in \mathbb{N}$  let  $\mu_t(\mathbf{X}, \cdot) : \mathfrak{B}^{(k)} \rightarrow [0, 1]$ ,  $k \in \mathbb{N}$  denote the empirical measure

$$\mu_t(\mathbf{X}, B) := \frac{1}{t} \sum_{i=0}^{t-1} \chi_B\{X_{ik+1}, \dots, X_{(i+1)k}\}, \quad (5)$$

where  $\chi$  is the indicator function. An empirical estimate of  $\alpha_n^\ell(m)$ ,  $m, \ell \in \mathbb{N}$ ,  $n > m \in \mathbb{N}$  can be obtained as  $\hat{\alpha}_{t,n}^\ell(\mathbf{X}, m)$  given by

$$\max_{j \in \{1, \dots, n - m\}} \max_{\substack{A \in \mathcal{D}_{j,\ell} \\ B \in \mathcal{D}_{j',\ell}}} \left| \gamma_{t,n}^{m,j}(\mathbf{X}, A, B) - \mu_t(\mathbf{X}, A) \mu_t(\mathbf{X}, B) \right| \quad (6)$$

where  $j' := n - m - j$ ,  $\mu_t(\mathbf{X}, \cdot)$  is given by (5) and  $\gamma_{t,n}^{m,j}(\mathbf{X}, A, B)$  is specified as

$$\frac{1}{t} \sum_{i=0}^{t-1} \chi_A(X_{in+1}, \dots, X_{in+j}) \chi_B(X_{in+j+m}, \dots, X_{(i+1)n}) \quad (7)$$

with  $t \geq n$ . First, observe that when  $m, n, \ell \in \mathbb{N}$  are fixed, we only have finitely many cylinder sets to consider in (6), hence the ergodic theorem leads to the following result.

**Proposition 2.** Let  $\mathbf{X}$  be a (not necessarily mixing) stationary ergodic process with process distribution  $\mu$  and sequence of  $\alpha$ -mixing coefficients  $\alpha = \langle \alpha(m) \rangle_{m \in \mathbb{N}}$ . For every  $m, \ell, n \in \mathbb{N}$  it holds that

$$\lim_{t \rightarrow \infty} \hat{\alpha}_{t,n}^\ell(\mathbf{X}, m) = \alpha_n^\ell(m), \quad \mu - \text{almost surely.}$$

However, this in itself is clearly not sufficient to lead to a consistent estimate of  $\alpha(m)$ . To achieve a consistent estimate of the  $\alpha$ -mixing coefficient we proceed as follows.

a) *Estimation under finite  $\|\alpha\|$* : Let us start by considering the case where the sequence of  $\alpha$  mixing coefficients is summable, i.e.  $\|\alpha\| < \infty$ . Let  $\langle \ell_t \rangle, \langle n_t \rangle$  for  $t \in \mathbb{N}$  be non-decreasing unbounded sequences of positive integers. Let the sequence of positive numbers  $\langle \delta_t \rangle_{t \in \mathbb{N}}$  be such that  $\sum_{t=1}^{\infty} \delta_t < \infty$ . For each  $m, \ell, n \in \mathbb{N}$  define the constant  $C_{m,\ell,n}$  by

$$C_{m,\ell,n} := m(2^{2^{n\ell} + 2^{m\ell+1} + 1}). \quad (8)$$

Let  $\langle \epsilon_t \rangle_{t \in \mathbb{N}}$  be another sequence of positive numbers such that  $\lim_{t \rightarrow \infty} \epsilon_t = 0$ . For a fixed  $m \in \mathbb{N}$  and each  $t \in \mathbb{N}$  define

$$\tau_t := \frac{C_{m,\ell_t,n_t}}{m\epsilon_t^2\delta_t} \quad (9)$$

We propose an estimator for  $\alpha(m)$  as

$$\hat{\alpha}_t(\mathbf{X}, m) := \hat{\alpha}_{\tau_t, n_t}^{\ell_t}(\mathbf{X}, m). \quad (10)$$

**Theorem 1** ( $\hat{\alpha}_t$  is strongly consistent). For each  $m \in \mathbb{N}$  and any stationary  $\alpha$ -mixing process  $\mathbf{X}$  with process distribution  $\mu$  and sequence of  $\alpha$ -mixing coefficients  $\alpha$  such that  $\|\alpha\| < \infty$  we have

$$\lim_{t \rightarrow \infty} \hat{\alpha}_t(\mathbf{X}, m) = \alpha(m), \quad \mu - \text{almost surely.}$$

The proof of Theorem 1 relies on the observation that when the  $\ell_1$  norm of the sequence of  $\alpha$ -mixing coefficients is finite, we have the following nonasymptotic inequality for the empirical version of the approximation of the mixing coefficients.

**Proposition 3.** Let  $\mathbf{X}$  be a stationary ergodic process with process distribution  $\mu$  and sequence of  $\alpha$ -mixing coefficients  $\alpha = \langle \alpha(m) \rangle_{m \in \mathbb{N}}$ . For every  $m, \ell, n, t \in \mathbb{N}$  and every  $\epsilon > 0$  we have

$$\Pr(|\hat{\alpha}_{t,n}^\ell(\mathbf{X}, m) - \alpha_n^\ell(m)| \geq \epsilon) \leq \frac{\|\alpha\| C_{m,\ell,n}}{m t \epsilon^2}.$$

Furthermore, for each  $M \in \mathbb{N}$  it holds that,

$$\Pr\left(\left|\sum_{m=1}^M \hat{\alpha}_{t,n}^\ell(\mathbf{X}, m) - \sum_{m=1}^M \alpha_n^\ell(m)\right| \geq \epsilon\right) \leq \frac{\|\alpha\| C_{M,\ell,n}}{t \epsilon^2}.$$

A key ingredient of the proof of Proposition 3, is Lemma 2 below, which builds upon Rio's variance bound [4, Corollary 1.1] to provide a simple concentration inequality on the empirical measure of a cylinder set  $[D]_1^k$  for any  $D \in \mathcal{D}_{k,\ell}$ ,  $k, \ell \in \mathbb{N}$ .

**Lemma 2.** For all  $D \in \mathcal{D}_{k,\ell}$ ,  $k, \ell, t \in \mathbb{N}$  and for every  $\epsilon > 0$ , we have

$$\Pr(|\mu_t(\mathbf{X}, D) - \mu([D]_1^k)| \geq \epsilon) \leq \frac{4\|\alpha\|}{t \epsilon^2}.$$

Next, we introduce an estimator for  $\|\alpha\|$  and prove its consistency under the assumption that the process is stationary  $\alpha$ -mixing, with a summable sequence  $\alpha$ . For  $t \in \mathbb{N}$  recall the parameters  $\langle \ell_t \rangle, \langle n_t \rangle, \langle \delta_t \rangle, \langle \epsilon_t \rangle$  defined earlier, and let  $\langle M_t \rangle_{t \in \mathbb{N}}$  be an increasing sequence of positive integers. For each  $t \in \mathbb{N}$  define,

$$\kappa_t := \frac{C_{M_t, \ell_t, n_t}}{\epsilon_t^2 \delta_t} \quad (11)$$

$$\theta_t(\mathbf{X}) := \sum_{m=1}^{M_t} \hat{\alpha}_{\kappa_t, n_t}^{\ell_t}(\mathbf{X}, m). \quad (12)$$

**Theorem 2** ( $\theta_t$  is strongly consistent). For any stationary  $\alpha$ -mixing process  $\mathbf{X}$  with process distribution  $\mu$  and sequence of  $\alpha$ -mixing coefficients  $\alpha$  such that  $\|\alpha\| < \infty$ , it holds that

$$\lim_{t \rightarrow \infty} \theta_t(\mathbf{X}) = \|\alpha\|, \quad \mu - \text{almost surely.}$$

*b) Estimation under arbitrary  $\|\alpha\|$ :* In this section we address the case when one does not have the guarantee that the  $\ell_1$  norm of the sequence of  $\alpha$ -mixing coefficients is finite. This setting is significantly more complex, since the conclusion of Proposition 3 is potentially vacuous, irrespective of the sample size. To circumvent this difficulty, we construct a non-decreasing sequence of estimates of  $\|\alpha\|$  – starting with the trivial lower bound 0 – by performing a sequence of tests. The value of the estimate is only increased if there is sufficient evidence that the value of  $\|\alpha\|$  is indeed larger than the current estimate. This is possible since if  $\|\alpha\|$  happens to be small, then its empirical estimator must be small thanks to Proposition 3.

With the objective to make the presentation more transparent, we start with a simpler estimator that is “weakly consistent” in the sense that if one prescribes a tolerated probability of error  $\delta$ , then one may construct a corresponding estimator that converges to  $\|\alpha\|$  with probability at least  $1 - \delta$ . Then we introduce a more complex estimator that is almost surely consistent.

In order to introduce the “weak” estimator, let  $\delta \in (0, 1)$  be the allowed probability of error, and let  $\delta_t$  be positive numbers such that  $\sum_{t=1}^{\infty} \delta_t = \delta$ . Fix a countable dense subset  $\mathcal{S}$  of  $[0, \infty)$  and let  $\mathcal{S} = \{s_1, s_2, \dots\}$  be an arbitrary enumeration of  $\mathcal{S}$ . Given a process  $\mathbf{X}$ , set  $\psi_0(\mathbf{X}) := 0$ , and for each  $t \in \mathbb{N}$  define

$$\psi_t(\mathbf{X}) = \begin{cases} \max\{s_t, \psi_{t-1}(\mathbf{X})\} & \text{if } \theta_t(\mathbf{X}) > s_t + \epsilon_t \sqrt{s_t} \\ \psi_{t-1}(\mathbf{X}) & \text{otherwise,} \end{cases} \quad (13)$$

with  $\epsilon_t$  being a decreasing positive sequence converging to zero and  $\theta_t$  specified by (12). Clearly,  $\langle \psi_t(\mathbf{X}) \rangle_{t \in \mathbb{N}}$  form a non-decreasing sequence that is either convergent or diverges to infinity.

**Theorem 3** ( $\psi_t$  is weakly consistent). Let  $\mathbf{X}$  be a (not necessarily mixing) stationary ergodic process with process distribution  $\mu$  and sequence of  $\alpha$ -mixing coefficients  $\alpha$ . With probability at least  $1 - \delta$ ,

$$\lim_{t \rightarrow \infty} \psi_t(\mathbf{X}) = \|\alpha\|.$$

We are ready to present the main result of the paper: an estimator for  $\|\alpha\|$  which, as shown below, is strongly consistent regardless of whether this norm is finite or infinite. To this end, we build upon the construction of  $\psi_t(\mathbf{X})$ , which gives rise to our third estimator of  $\|\alpha\|$ . Recall that  $\mathcal{S}$  is a countable dense subset of  $[0, \infty)$ . Define a sequence  $\langle u_t \rangle$  such that  $u_t \in \mathcal{S}$  for all  $t$  and for every  $s \in \mathcal{S}$ , the set

$\{t \in \mathbb{N} : u_t = s\}$  is infinite. The sequence  $\langle u_t \rangle$  “visits” each element of  $\mathcal{S}$  infinitely often. One such example is:

$$\begin{aligned} u_1 &= s_1, \\ u_2 &= s_1, u_3 = s_2, \\ u_4 &= s_1, u_5 = s_2, u_6 = s_3, \\ &\vdots \end{aligned}$$

The estimator  $\xi_t$  given by (14) below keeps track of a “bit value”  $b_t(s) \in \{0, 1\}$  assigned to each  $s \in \mathcal{S}$ . At each time instance,  $b_t(s)$  marks the outcome of the test the last time the value  $s$  was tested. With  $b_0(s) = 1$  for all  $s \in \mathcal{S}$ , and  $\xi_0(\mathbf{X}) := 0$ , for each  $t \in \mathbb{N}$  let

$$b_t(s) = \begin{cases} 1 & \text{if } u_t = s \text{ and } \theta_t(\mathbf{X}) \leq s + \epsilon_t \sqrt{s} \\ 0 & \text{if } u_t = s \text{ and } \theta_t(\mathbf{X}) > s + \epsilon_t \sqrt{s} \\ b_{t-1}(s) & \text{otherwise} \end{cases}$$

and define

$$\xi_t(\mathbf{X}) = \inf \{s \in \mathcal{S} : b_t(s) = 1\}, \quad (14)$$

where the infimum is defined to be zero if the set is empty.

**Theorem 4** ( $\xi_t$  is strongly consistent). Let  $\mathbf{X}$  be a (not necessarily mixing) stationary ergodic process with process distribution  $\mu$  and sequence of  $\alpha$ -mixing coefficients  $\alpha$ . We have,

$$\lim_{t \rightarrow \infty} \xi_t(\mathbf{X}) = \|\alpha\|, \quad \mu - \text{almost surely.}$$

### B. Approximating and estimating the $\beta$ -mixing quantities

In this section, we derive estimators concerning the  $\beta$ -mixing coefficients of a stationary ergodic process. For the most part, the results are analogous to those given in Section III-A. An overview of the similarities and differences between the two settings is provided in the Introduction.

We start by defining an approximation of the  $\beta$ -mixing coefficients of  $\mathbf{X}$ . Recall the notation  $\mathcal{D}_{k,\ell}$ ,  $k, \ell \in \mathbb{N}$  introduced in the beginning of Section III which corresponds to the power-set  $\mathcal{P}(\Delta_{k,\ell})$  of the set of dyadic cubes in  $\mathcal{X}^k$  of side-length  $2^{-\ell}$ . For  $m, \ell \in \mathbb{N}$ ,  $n > m \in \mathbb{N}$  and  $j \in \{1, \dots, n - m\}$ , let  $\beta_{n,j}^\ell(m)$  be defined as

$$\frac{1}{2} \sum_{A \in \mathcal{D}_{j,\ell}} \sum_{B \in \mathcal{D}_{j',\ell}} \left| \mu([A]_1^j \cap [B]_{j+m+1}^n) - \mu([A]_1^j) \mu([B]_{j+m+1}^n) \right| \quad (15)$$

where  $j' := n - m - j + 1$ . The intuition behind this choice follows from the discussion provided in Section I-B. Recall the  $\sigma$ -subalgebras of  $\mathfrak{F}$ , denoted  $\mathfrak{F}_1^j(\ell)$  and  $\mathfrak{F}_{j+m}^n(\ell)$ , with  $m, \ell \in \mathbb{N}$ ,  $n > m \in \mathbb{N}$  and  $j \in \{1, \dots, n - m\}$ , as given by (2) and (3) respectively. In a manner analogous to the estimation of the  $\alpha$ -mixing coefficients, we show via Lemma 3 below that to approximate the  $\beta$ -dependence between  $\mathfrak{F}_1^j(\ell)$  and  $\mathfrak{F}_{j+m}^n(\ell)$ , it suffices to use  $\beta_{n,j}^\ell(m)$  given by (15). Note that the  $\pi - \lambda$  argument used in the proof of Lemma 1 does not carry over to this case where summations are involved in (15) in place of the suprema in (4).

**Lemma 3.** For  $m, \ell \in \mathbb{N}$ ,  $n > m \in \mathbb{N}$  and  $j \in \{1, \dots, n-m\}$  we have

$$\beta_{n,j}^\ell(m) = \beta(\mathfrak{F}_1^j(\ell), \mathfrak{F}_{j+m}^n(\ell)).$$

For  $m, \ell \in \mathbb{N}$ ,  $n > m \in \mathbb{N}$ , let  $\beta_n^\ell(m) = \max_{j \in \{1, \dots, n-m\}} \beta_{n,j}^\ell(m)$  define an approximation of  $\beta(m)$ .

**Proposition 4.** For every  $m \in \mathbb{N}$  we have

$$\lim_{n, \ell \rightarrow \infty} \beta_n^\ell(m) = \beta(m).$$

An empirical estimate of  $\beta_n^\ell(m)$ ,  $m, \ell \in \mathbb{N}$ ,  $n > m \in \mathbb{N}$ , denoted  $\hat{\beta}_{t,n}^\ell(\mathbf{X}, m)$ , can be obtained as

$$\max_{j \in \{1, \dots, n-m\}} \frac{1}{2} \sum_{A \in \mathcal{D}_{j,\ell}} \sum_{B \in \mathcal{D}_{j',\ell}} \left| \gamma_{t,n}^{m,j}(\mathbf{X}, A, B) - \mu_t(\mathbf{X}, A) \mu_t(\mathbf{X}, B) \right| \quad (16)$$

where  $j' := n - m - j$ ,  $\mu_t(\mathbf{X}, \cdot)$  and  $\gamma_{t,n}^{m,j}(\mathbf{X}, \cdot, \cdot)$  are given by (5) and (7) respectively. As with Lemma 2, since for fixed  $m, n, \ell \in \mathbb{N}$  we only have finitely many cylinder sets to consider in (16), a simple application of the ergodic theorem gives the following lemma.

**Lemma 4.** Let  $\mathbf{X}$  be a (not necessarily mixing) stationary ergodic process with process distribution  $\mu$  and sequence of  $\beta$ -mixing coefficients  $\beta = \langle \beta(m) \rangle_{m \in \mathbb{N}}$ . For every  $m, \ell, n \in \mathbb{N}$  it holds that

$$\lim_{t \rightarrow \infty} \hat{\beta}_{t,n}^\ell(\mathbf{X}, m) = \beta_n^\ell(m), \quad \mu - \text{almost surely.}$$

For each  $m, \ell, k \in \mathbb{N}$  let the constant  $\tilde{C}_{m,\ell,k}$  be given by

$$\tilde{C}_{m,\ell,k} := m(2^{2k\ell+1+2m\ell+1}+2). \quad (17)$$

An argument analogous to that given for Proposition 3 leads to the following result.

**Proposition 5.** Let  $\mathbf{X}$  be a stationary ergodic process with process distribution  $\mu$  and sequence of  $\beta$ -mixing coefficients  $\beta = \langle \beta(m) \rangle_{m \in \mathbb{N}}$ . For every  $m, \ell, n, t \in \mathbb{N}$  and every  $\epsilon > 0$  we have

$$\Pr(|\hat{\beta}_{t,n}^\ell(\mathbf{X}, m) - \beta_n^\ell(m)| \geq \epsilon) \leq \frac{\|\beta\| \tilde{C}_{m,\ell,n}}{mt\epsilon^2}$$

with  $\tilde{C}_{m,\ell,n}$  given by (17). Moreover, for each  $M \in \mathbb{N}$ ,

$$\Pr\left(\left|\sum_{m=1}^M \hat{\beta}_{t,n}^\ell(\mathbf{X}, m) - \beta_n^\ell(m)\right| \geq \epsilon\right) \leq \frac{\|\beta\| \tilde{C}_{M,\ell,n}}{t\epsilon^2}.$$

As in the previous section, let  $\langle \ell_t \rangle, \langle n_t \rangle$  for  $t \in \mathbb{N}$  be non-decreasing unbounded sequences of positive integers. Let the sequence of positive real numbers  $\langle \delta_t \rangle_{t \in \mathbb{N}}$  be such that  $\sum_{t=1}^\infty \delta_t < \infty$ . Let  $\langle \epsilon_t \rangle_{t \in \mathbb{N}}$  be another sequence of positive real numbers such that  $\lim_{t \rightarrow \infty} \epsilon_t = 0$ . Furthermore let  $\langle M_t \rangle_{t \in \mathbb{N}}$  be an increasing sequence of positive integers. For each  $t \in \mathbb{N}$  let

$$\tilde{\tau}_t := \frac{\tilde{C}_{m,\ell_t,n_t}}{m\epsilon_t^2\delta_t}, \quad m \in \mathbb{N} \quad \text{and} \quad \tilde{\kappa}_t := \frac{\tilde{C}_{M_t,\ell_t,n_t}}{\epsilon_t^2\delta_t}, \quad (18)$$

with the constant  $\tilde{C}_{m,\ell,n}$  given by (17). Define

$$\begin{aligned} \hat{\beta}_t(\mathbf{X}, m) &:= \hat{\beta}_{\tau_t, n_t}^{\ell_t}(\mathbf{X}, m), \\ \tilde{\theta}_t(\mathbf{X}) &:= \sum_{m=1}^{M_t} \hat{\beta}_{\tau_t, n_t}^{\ell_t}(\mathbf{X}, m). \end{aligned} \quad (19)$$

Using Proposition 4, Proposition 5 and based on the choice of parameters specified by (18), arguments analogous to those given for Theorem 1 and Theorem 2, yield the following result.

**Theorem 5** ( $\hat{\beta}_t$  and  $\tilde{\theta}_t$  are strongly consistent). For each  $m \in \mathbb{N}$  and any stationary  $\beta$ -mixing process  $\mathbf{X}$  with process distribution  $\mu$  and sequence of  $\beta$ -mixing coefficients  $\beta$  such that  $\|\beta\| < \infty$  we have

$$\lim_{t \rightarrow \infty} \hat{\beta}_t(\mathbf{X}, m) = \beta(m)$$

and

$$\lim_{t \rightarrow \infty} \tilde{\theta}_t(\mathbf{X}) = \|\beta\|, \quad \mu - \text{almost surely.}$$

Next, we introduce two estimators for  $\|\beta\|$ , which are analogous to those given by (13) and (14) for the estimation of  $\|\alpha\|$ . As in the previous section, let  $\mathcal{S} := \{s_1, s_2, \dots\}$  be a countable dense subset of  $[0, \infty)$ . Given a process  $\mathbf{X}$ , set  $\tilde{\psi}_0(\mathbf{X}) := 0$ , and for each  $t \in \mathbb{N}$  define

$$\tilde{\psi}_t(\mathbf{X}) = \begin{cases} \max\{s_t, \tilde{\psi}_{t-1}(\mathbf{X})\} & \text{if } \tilde{\theta}_t(\mathbf{X}) > s_t + \epsilon_t \sqrt{s_t} \\ \tilde{\psi}_{t-1}(\mathbf{X}) & \text{otherwise,} \end{cases}$$

with  $\tilde{\theta}_t$  specified by (18). Define a sequence  $\langle u_t \rangle$  such that  $u_t \in \mathcal{S}$  for all  $t$  and for every  $s \in \mathcal{S}$ , the set  $\{t \in \mathbb{N} : u_t = s\}$  is infinite. The sequence  $\langle u_t \rangle$  “visits” each element of  $\mathcal{S}$  infinitely often. As in the case of  $\xi_t$  given by (14), the estimator  $\xi_t$  given by (20) below keeps track of the the outcome of the test when the value  $s \in \mathcal{S}$  was last tested. Set  $\tilde{b}_0(s) = 1$  for all  $s \in \mathcal{S}$ , and  $\xi_0(\mathbf{X}) := 0$ ; for each  $t \in \mathbb{N}$ , let

$$\tilde{b}_t(s) = \begin{cases} 1 & \text{if } u_t = s \text{ and } \tilde{\theta}_t(\mathbf{X}) \leq s + \epsilon_t \sqrt{s} \\ 0 & \text{if } u_t = s \text{ and } \tilde{\theta}_t(\mathbf{X}) > s + \epsilon_t \sqrt{s} \\ \tilde{b}_{t-1}(s) & \text{otherwise.} \end{cases}$$

and define

$$\tilde{\xi}_t(\mathbf{X}) = \inf \left\{ s \in \mathcal{S} : \tilde{b}_t(s) = 1 \right\} \quad (20)$$

where the inf is defined to be zero if the set is empty. Thanks to Theorem 5, Proposition 5, Proposition 4, and Lemma 4, and based on arguments analogous to those given for Theorem 3 and Theorem 4, it can be shown that  $\tilde{\psi}_t$  and  $\tilde{\xi}_t$  are consistent, regardless of the summability of  $\beta$ . More specifically, we have the following results.

**Theorem 6** ( $\tilde{\psi}_t$  and  $\tilde{\xi}_t$  are weakly and strongly consistent respectively). Let  $\mathbf{X}$  be a (not necessarily mixing) stationary ergodic process with process distribution  $\mu$  and sequence of  $\beta$ -mixing coefficients  $\beta$ . With  $\mu$ -probability at least  $1 - \delta$ , we have

$$\lim_{t \rightarrow \infty} \tilde{\psi}_t(\mathbf{X}) = \|\beta\|.$$

Moreover, it holds that

$$\lim_{t \rightarrow \infty} \tilde{\xi}_t(\mathbf{X}) = \|\alpha\|, \quad \mu - \text{almost surely.}$$

#### IV. GOODNESS-OF-FIT TESTING

In this section we use our estimators to construct consistent hypothesis tests for mixing rates as well as for upper bounds on  $\|\alpha\|$  and  $\|\beta\|$ .

##### A. Test for mixing rates

Consider a stationary ergodic process  $\mathbf{X}$  with process distribution  $\mu$  and sequences of  $\alpha$ -mixing and  $\beta$ -mixing coefficients  $\alpha = \langle \alpha(m) \rangle_{m \in \mathbb{N}}$  and  $\beta = \langle \beta(m) \rangle_{m \in \mathbb{N}}$  respectively. A mapping  $\gamma : \mathbb{N} \rightarrow [0, 1]$  is called an  $\alpha$ -mixing rate function for  $\mu$  if for each  $m \in \mathbb{N}$  we have  $\alpha(m) \leq \gamma(m)$ . Similarly, it is said to be a  $\beta$ -mixing rate function for  $\mu$  if for each fixed  $m \in \mathbb{N}$  it holds that  $\beta(m) \leq \gamma(m)$ . We wish to test the null hypothesis  $H_0$  that  $\gamma$  is an  $\alpha$ -mixing (respectively  $\beta$ -mixing) rate function for  $\mu$  against the alternative hypothesis  $H_1$  that there exists some  $m \in \mathbb{N}$  such that  $\alpha(m) > \gamma(m)$  (respectively  $\beta(m) > \gamma(m)$ ). More formally, let  $\mathcal{C}^\alpha$  be the class of stationary  $\alpha$ -mixing processes whose sequence of  $\alpha$ -mixing coefficients  $\alpha = \langle \alpha(m) \rangle_{m \in \mathbb{N}}$  is summable, that is, such that  $\|\alpha\| < \infty$ . Similarly, let  $\mathcal{C}^\beta$  be the class of stationary  $\beta$ -mixing processes whose sequence of  $\beta$ -mixing coefficients is such that  $\|\beta\| < \infty$ . For a rate function  $\gamma$ , let  $\mathcal{R}_\gamma^\alpha$  (respectively  $\mathcal{R}_\gamma^\beta$ ) be the class of all processes which have  $\gamma$  as an  $\alpha$ -mixing (respectively  $\beta$ -mixing) rate function. We construct a sequence  $\langle g_t \rangle_{t \in \mathbb{N}}$  of functions  $g_t : \mathcal{X}^\mathbb{N} \rightarrow \{+1, -1\}$ ,  $t \in \mathbb{N}$  each measurable with respect to the filtration  $\mathfrak{F}_t$  such that given a sample  $\mathbf{X}$  generated by  $\mu \in \mathcal{C}^\alpha$ , with probability 1, produce

$$\lim_{t \rightarrow \infty} g_t(\mathbf{X}) = \begin{cases} +1 & \text{if } \mu \in \mathcal{C}^\alpha \cap \mathcal{R}_\gamma^\alpha \\ -1 & \text{if } \mu \in \mathcal{C}^\alpha \setminus \mathcal{R}_\gamma^\alpha. \end{cases} \quad (21)$$

We call such a sequence of function satisfying (21) a strongly consistent goodness-of-fit test for rate function  $\gamma$  on  $\mathcal{C}^\alpha$ . Analogously, we construct a sequence  $\langle \tilde{g}_t \rangle_{t \in \mathbb{N}}$  of  $\mathfrak{F}_t$ -measurable functions  $\tilde{g}_t : \mathcal{X}^\mathbb{N} \rightarrow \{+1, -1\}$ ,  $t \in \mathbb{N}$  such that with probability 1,

$$\lim_{t \rightarrow \infty} \tilde{g}_t(\mathbf{X}) = \begin{cases} +1 & \text{if } \mu \in \mathcal{C}^\beta \cap \mathcal{R}_\gamma^\beta \\ -1 & \text{if } \mu \in \mathcal{C}^\beta \setminus \mathcal{R}_\gamma^\beta. \end{cases} \quad (22)$$

To construct our tests, let us recall some relevant notation from the previous sections. For  $t \in \mathbb{N}$  let  $\langle M_t \rangle, \langle \ell_t \rangle, \langle n_t \rangle$  be increasing sequences of positive integers. Take  $\delta \in (0, 1)$  and let the sequence of positive real numbers  $\langle \delta_t \rangle_{t \in \mathbb{N}}$  be such that  $\sum_{t=1}^\infty \delta_t = \delta$ . Let  $\langle \epsilon_t \rangle_{t \in \mathbb{N}}$  be another sequence of positive numbers such that  $\lim_{t \rightarrow \infty} \epsilon_t = 0$ . For each  $m \in \mathbb{N}$  and  $t \in \mathbb{N}$  recall the estimators  $\hat{\alpha}_t(\mathbf{X}, m) := \hat{\alpha}_{\tau_t, n_t}^{\ell_t}(\mathbf{X}, m)$  and  $\hat{\beta}_t(\mathbf{X}, m) := \hat{\beta}_{\tau_t, n_t}^{\ell_t}(\mathbf{X}, m)$  respectively specified by (10) and (19), with  $\tau_t$  given by (9) and  $\tilde{\tau}_t$  by (18). For a given process  $\mathbf{X}$  and every  $t \in \mathbb{N}$  define

$$g_t(\mathbf{X}) := \begin{cases} +1 & \text{if } \hat{\alpha}_t(\mathbf{X}, m) \leq \gamma(m) + \epsilon_t, \quad m \in \{1, \dots, M_t\} \\ -1 & \text{otherwise} \end{cases} \quad (23)$$

and

$$\tilde{g}_t(\mathbf{X}) := \begin{cases} +1 & \text{if } \hat{\beta}_t(\mathbf{X}, m) \leq \gamma(m) + \epsilon_t, \quad m \in \{1, \dots, M_t\} \\ -1 & \text{otherwise} \end{cases} \quad (24)$$

**Theorem 7.** The sequence of functions  $g_t$ ,  $t \in \mathbb{N}$  given by (23) give rise to a goodness-of-fit test on  $\mathcal{C}^\alpha$  that is strongly consistent in the sense of (21).

It is straightforward to check that a similar argument based on Proposition 4 and Proposition 5 leads to the following analogous result concerning  $\beta$ -mixing rate functions.

**Theorem 8.** The sequence of functions  $\tilde{g}_t$ ,  $t \in \mathbb{N}$  defined by (24) gives rise to a goodness-of-fit test on  $\mathcal{C}^\beta$  that is strongly consistent in the sense of (22).

##### B. Test for an upper bounds on $\|\alpha\|$ and $\|\beta\|$

We propose an asymptotically consistent test for an upper-bound on  $\|\alpha\|$  (respectively  $\|\beta\|$ ) of a stationary ergodic process that is not necessarily mixing. Denote by  $\mathcal{E}$  the set of all  $\mathcal{X}$ -valued stationary ergodic processes. As in Section IV-A, let  $\mathcal{C}^\alpha$  (respectively  $\mathcal{C}^\beta$ ) be the class of stationary  $\alpha$ -mixing (respectively  $\beta$ -mixing) processes whose sequence of  $\alpha$ -mixing (respectively  $\beta$ -mixing) coefficients is summable. Observe that  $\mathcal{C}^\beta \subset \mathcal{C}^\alpha \subset \mathcal{E}$ . Given a sample  $\mathbf{X}$  generated by  $\mu \in \mathcal{E}$ , we wish to test the null hypothesis  $H_0$  that  $\|\alpha\| \leq \gamma$  (respectively  $\|\beta\| \leq \gamma$ ) for some fixed threshold  $\gamma \in [0, \infty)$  against the alternative hypothesis that  $\|\alpha\| > \gamma$  (respectively  $\|\beta\| > \gamma$ ). Let  $\langle \zeta_t \rangle_{t \in \mathbb{N}}$  be a decreasing sequence of positive real numbers such that  $\lim_{t \rightarrow \infty} \zeta_t = 0$ . For a sample  $\mathbf{X}$ , define the sequences  $\langle f_t \rangle_{t \in \mathbb{N}}$  and  $\langle \tilde{f}_t \rangle_{t \in \mathbb{N}}$  of  $\mathfrak{F}_t$ -measurable functions

$$f_t(\mathbf{X}) := \begin{cases} +1 & \text{if } \xi_t(\mathbf{X}) \leq \gamma + \zeta_t \\ -1 & \text{otherwise} \end{cases} \quad (25)$$

and

$$\tilde{f}_t(\mathbf{X}) := \begin{cases} +1 & \text{if } \tilde{\xi}_t(\mathbf{X}) \leq \gamma + \zeta_t \\ -1 & \text{otherwise,} \end{cases} \quad (26)$$

where  $\xi_t(\mathbf{X})$  is given by (14) and  $\tilde{\xi}_t(\mathbf{X})$  is given by (20). For each  $\gamma \in [0, \infty)$  denote by  $\mathcal{C}_\gamma^\alpha$  the subclass of  $\mathcal{C}^\alpha$  corresponding to stationary  $\alpha$ -mixing processes whose sequences of  $\alpha$ -mixing coefficients sum to at most  $\gamma$ . A simple argument which relies on Theorem 4 yields the following result.

**Theorem 9.** For each  $\gamma \in [0, \infty)$ , and given a sample  $\mathbf{X}$  generated by a (not necessarily mixing) stationary ergodic process  $\mu \in \mathcal{E}$ , with probability 1 the test  $\langle f_t \rangle_{t \in \mathbb{N}}$  given by (25) has the property that

$$\lim_{t \rightarrow \infty} f_t(\mathbf{X}) = \begin{cases} +1 & \text{if } \mu \in \mathcal{C}_\gamma^\alpha \\ -1 & \text{if } \mu \in \mathcal{E} \setminus \mathcal{C}_\gamma^\alpha. \end{cases}$$

Similarly, an argument based on Theorem 6, leads to the following analogous result.

**Theorem 10.** For each  $\gamma \in [0, \infty)$ , and given a sample  $\mathbf{X}$  generated by a (not necessarily mixing) stationary ergodic process  $\mu \in \mathcal{E}$ , with probability 1 the test  $\langle \tilde{f}_t \rangle_{t \in \mathbb{N}}$  given by (26) has the property that

$$\lim_{t \rightarrow \infty} \tilde{f}_t(\mathbf{X}) = \begin{cases} +1 & \text{if } \mu \in \mathcal{C}_\gamma^\beta \\ -1 & \text{if } \mu \in \mathcal{E} \setminus \mathcal{C}_\gamma^\beta. \end{cases}$$

### C. Test for independence

Using Theorem 9, we can readily obtain a consistent test for independence on  $\mathcal{E}$ . Consider a process  $\mathbf{X}$  with distribution  $\mu$ . Recall that  $\mathcal{J}$  denotes the set of all  $\mathcal{X}$ -valued i.i.d. processes, and observe that if  $\mu \in \mathcal{J}$  then clearly  $\|\alpha\| = 0$ . Thus, if  $\mu \in \mathcal{E}$  we can use (25) with  $\gamma = 0$  to test the null hypothesis  $H_0$  that  $\mu \in \mathcal{J}$  against the alternative hypothesis  $H_1$  that  $\mu \in \mathcal{E} \setminus \mathcal{J}$  so that the process is stationary ergodic but it is not i.i.d.. More specifically, let  $\langle \zeta_t \rangle$  be a decreasing sequence of positive real numbers as before, such that  $\lim_{t \rightarrow \infty} \zeta_t = 0$ . For a given process  $\mathbf{X}$  and each  $t \in \mathbb{N}$  define

$$\bar{f}_t(\mathbf{X}) := \begin{cases} +1 & \text{if } \hat{\alpha}_t(\mathbf{X}) \leq \zeta_t \\ -1 & \text{otherwise} \end{cases} \quad (27)$$

Observing that  $\|\alpha\| = 0$  for all i.i.d. processes  $\mu \in \mathcal{J}$ , the consistency of (27) trivially follows from Theorem 9, leading to the following corollary.

**Corollary 1** (Morvai and Weiss [17]). There exists a strongly consistent test for independence on  $\mathcal{E}$ . That is, for a stationary ergodic process  $\mathbf{X}$  with process distribution  $\mu \in \mathcal{E}$  with probability one it holds that

$$\lim_{t \rightarrow \infty} \bar{f}_t(\mathbf{X}) = \begin{cases} +1 & \text{if } \mu \in \mathcal{E} \cap \mathcal{J} \\ -1 & \text{if } \mu \in \mathcal{E} \setminus \mathcal{J}. \end{cases}$$

This recovers, albeit via a different approach, the main result of Morvai and Weiss [17].

### V. OUTLOOK

We studied the problem of estimating the mixing coefficients of a real-valued, stationary ergodic process from a finite sample-path. As an initial step, we proposed strongly consistent estimators of individual  $\alpha$ -mixing coefficients for a subclasses of stationary processes whose sequences of  $\alpha$ -mixing coefficients are summable. We also provided strongly consistent estimators of the  $\ell_1$  norm of the sequence  $\alpha$ -mixing coefficients for this class of processes, and extended the results to the case where the stationary process is not necessarily  $\alpha$ -mixing. Our main inspiration for estimating  $\|\alpha\|$  has been that the  $\ell_1$  norms of the sequences of mixing coefficients naturally appear in concentration of measure inequalities for dependent processes. We further proposed analogous results concerning  $\beta$ -mixing coefficients. Finally, building upon our estimators we developed strongly consistent hypothesis tests for mixing rates as well as for upper bounds on  $\|\alpha\|$  and  $\|\beta\|$ . This, in turn, gave rise to a strongly consistent test for independence, recovering the main result of Morvai and Weiss [17] as a corollary. A particularly interesting open problem is to estimate the  $\ell_1$  norm of the sequence of  $\phi$ -mixing coefficients. Indeed, as mentioned in the Introduction, when a process is  $\phi$ -mixing, one can obtain exponential concentration inequalities in terms of  $\|\phi\|$ . However, the main difficulty in estimating the  $\phi$ -mixing coefficients is to consistently estimate the conditional probabilities, as some events may be of arbitrarily small measure. We conjecture that this challenge may be overcome by generalizing the plug-in estimators of Grünewälder [18] which are designed for the i.i.d. setting.

### VI. PROOFS

In this section we provide proofs of our technical results.

*Proof of Lemma 1.* Fix  $m, \ell \in \mathbb{N}$ ,  $n > m \in \mathbb{N}$  and  $j \in \{1, \dots, n-m\}$ . It is clear that  $\alpha_{n,j}^\ell(m) \leq \alpha(\mathfrak{F}_1^j(\ell), \mathfrak{F}_{j+m}^n(\ell))$ . Therefore, it suffices to show that  $\alpha(\mathfrak{F}_1^j(\ell), \mathfrak{F}_{j+m}^n(\ell)) \leq \alpha_{n,j}^\ell(m)$ . We show this using Dynkin's  $\pi - \lambda$  theorem as follows. Define  $\mathcal{A}_1^j(\ell) := \{[A]_1^j : A \in \mathcal{D}_{j,\ell}\}$  and

$$\mathcal{A}_{j+m}^n(\ell) := \{[B]_{j+m+1}^n : B \in \mathcal{D}_{j',\ell}, j' := n-m-j\}.$$

It trivially holds that  $\mathcal{A}_1^j(\ell)$  forms a  $\pi$ -system, since it is non-empty and closed under finite intersections. On the other hand, by definition we have

$$\sup_{\substack{U \in \mathcal{A}_1^j(\ell) \\ V \in \mathcal{A}_{j+m}^n(\ell)}} |\mu(U \cap V) - \mu(U)\mu(V)| = \alpha_{n,j}^\ell(m). \quad (28)$$

Let  $\mathcal{L}$  be the largest algebra of subsets of  $\mathfrak{F}$  such that

$$\sup_{\substack{U \in \mathcal{L} \\ V \in \mathcal{A}_{j+m}^n(\ell)}} |\mu(U \cap V) - \mu(U)\mu(V)| \leq \alpha_{n,j}^\ell(m).$$

It is straightforward to verify that  $\mathcal{L}$  forms a  $\lambda$ -system. To see this, first note that since  $\alpha_{n,j}^\ell(m) \geq 0$ , it clearly holds that  $\mathcal{X}^\mathbb{N} \in \mathcal{L}$ . Next, take  $U_1 \subseteq U_2 \in \mathcal{L}$ . Since  $\mathcal{L}$  is an algebra, it is closed under complementation as well as under pairwise unions and intersections. Therefore,  $(U_2 \setminus U_1) \in \mathcal{L}$ . Finally, consider a countable sequence of increasing subsets  $U_i \subseteq U_{i+1} \in \mathcal{L}$ ,  $i \in \mathbb{N}$  and define  $U := \bigcup_{i=1}^\infty U_i$ . It follows from the continuity of probability measure that  $U \in \mathcal{L}$ . To see this, for  $V \in \mathcal{A}_{j+m}^n(\ell)$  define  $\bar{U}_i := U_i \cap V$ ,  $i \in \mathbb{N}$ , and let  $\bar{U} := \bigcup_{i=1}^\infty \bar{U}_i$ . Observe that  $\bar{U}_i \subseteq \bar{U}_{i+1}$  so that  $\lim_{n \rightarrow \infty} \mu(\bar{U}_n) = \mu(\bar{U})$ . Similarly, it holds that  $\lim_{n \rightarrow \infty} \mu(U_n) = \mu(U)$ . Fix  $\epsilon > 0$  there exist  $N_\epsilon, N'_\epsilon$  such that for all  $n \geq \max\{N_\epsilon, N'_\epsilon\}$  we have  $|\mu(U_n) - \mu(U)| \leq \epsilon$  and  $|\mu(\bar{U}_n) - \mu(\bar{U})| \leq \epsilon$ . Therefore,

$$\begin{aligned} & |\mu(U \cap V) - \mu(U)\mu(V)| \\ &= |\mu(\bar{U}) - \mu(U)\mu(V)| \\ &\leq |\mu(\bar{U}_n) - \mu(U_n)\mu(V)| + 2\epsilon \\ &\leq \alpha_{n,j}^\ell(m) + 2\epsilon. \end{aligned}$$

Since the choice of epsilon is arbitrary, it follows that  $U \in \mathcal{L}$ . Therefore  $\mathcal{L}$  is a  $\lambda$ -system. Moreover, by (28) we have  $\mathcal{A}_1^j(\ell) \subseteq \mathcal{L}$ . Thus, as follows from Dynkin's  $\pi - \lambda$  theorem, we can deduce that  $\sigma(\mathcal{A}_1^j(\ell)) \subseteq \mathcal{L}$  so that

$$\sup_{\substack{U \in \sigma(\mathcal{A}_1^j(\ell)) \\ V \in \mathcal{A}_{j+m}^n(\ell)}} |\mu(U \cap V) - \mu(U)\mu(V)| \leq \alpha_{n,j}^\ell(m). \quad (29)$$

On the other hand, let  $\mathcal{L}'$  be the largest algebra of subsets of  $\mathfrak{F}$  such that

$$\sup_{\substack{U \in \sigma(\mathcal{A}_1^j(\ell)) \\ V \in \mathcal{L}'}} |\mu(U \cap V) - \mu(U)\mu(V)| \leq \alpha_{n,j}^\ell(m).$$

In much the same way as with  $\mathcal{L}$ , it is straightforward to check that  $\mathcal{L}'$  is a  $\lambda$ -system. Moreover, by (29) and an argument analogous to that concerning  $\mathcal{A}_1^j(\ell)$ , we can conclude that



$\mathcal{A}_{j+m}^n(\ell) \subseteq \mathcal{L}'$ . Hence, by another application of the  $\pi - \lambda$  theorem, we obtain  $\sigma(\mathcal{A}_{j+m}^n(\ell)) \subseteq \mathcal{L}'$ . This gives

$$\sup_{\substack{U \in \sigma(\mathcal{A}_1^j(\ell)) \\ V \in \sigma(\mathcal{A}_{j+m}^n(\ell))}} |\mu(U \cap V) - \mu(U)\mu(V)| \leq \alpha_{n,j}^\ell(m).$$

Noting that  $\sigma(\mathcal{A}_1^j(\ell)) = \mathfrak{F}_1^j(\ell)$  and  $\sigma(\mathcal{A}_{j+m}^n(\ell)) = \mathfrak{F}_{j+m}^n(\ell)$ , we obtain  $\alpha(\mathfrak{F}_1^j(\ell), \mathfrak{F}_{j+m}^n(\ell)) \leq \alpha_{n,j}^\ell(m)$ .  $\square$

*Proof of Proposition 1.* Fix  $m, \ell \in \mathbb{N}$  and  $n > m$ . Observe that for each  $j \in \{1, \dots, n-m\}$  we have  $\mathfrak{F}_1^j(\ell) \subset \mathfrak{F}_1^j(\ell+1)$ ,  $\ell \in \mathbb{N}$  and  $\mathfrak{F}_{j+m}^n(\ell) \subset \mathfrak{F}_{j+m}^{n+1}(\ell+1)$ ,  $n, \ell \in \mathbb{N}$  are each a sequence of  $\sigma$ -algebras with  $\bigvee_{\ell}^{\infty} \mathfrak{F}_1^j(\ell) = \sigma(X_1, \dots, X_j)$  and  $\bigvee_{n,\ell}^{\infty} \mathfrak{F}_{j+m}^n(\ell) = \sigma(\{X_t : t \geq j+m+1\})$ . Therefore, by Bradley [1, Vol. 1, Proposition 3.18]  $\lim_{n,\ell \rightarrow \infty} \alpha(\mathfrak{F}_1^j(\ell), \mathfrak{F}_{j+m}^n(\ell))$  is given by

$$\alpha(\sigma(\{X_t : t \in \{1, \dots, j\}\}), \sigma(\{X_t : t \geq j+m+1\})). \quad (30)$$

It is straightforward to check that for each  $m \in \mathbb{N}$  we have

$$\sup_{n,\ell} \max_{j \in \{1, \dots, n-m\}} \alpha_{n,j}^\ell(m) = \sup_{j \in \mathbb{N}} \sup_{n \geq j+m+1} \sup_{\ell} \alpha_{n,j}^\ell(m). \quad (31)$$

To see this first let  $c := \sup_{j \in \mathbb{N}} \sup_{n \geq j+m} \sup_{\ell} \alpha_{n,j}^\ell(m)$  and fix some  $\epsilon > 0$ ; by the definition of  $\sup$  there exist  $\ell^*, j^* \in \mathbb{N}$  and  $n^* \geq j^*+m$  such that  $\alpha_{n^*,j^*}^{\ell^*}(m) \geq c - \epsilon$ . We have  $c - \epsilon \leq \alpha_{n^*,j^*}^{\ell^*}(m) \leq \sup_{n,\ell} \max_{j \in \{1, \dots, n-m\}} \alpha_{n,j}^\ell(m)$ . Similarly, let  $c' := \sup_{n,\ell} \max_{j \in \{1, \dots, n-m\}} \alpha_{n,j}^\ell(m)$ , and note that there exist some  $n', \ell' \in \mathbb{N}$  such that  $\max_{j \in \{1, \dots, n'-m\}} \alpha_{n',j}^{\ell'}(m) \geq c' - \epsilon$ . It follows that  $c' - \epsilon \leq \max_{j \in \{1, \dots, n'-m\}} \alpha_{n',j}^{\ell'}(m) \leq \sup_{j \in \mathbb{N}} \sup_{n \geq j+m} \sup_{\ell} \alpha_{n,j}^\ell(m)$ . Since the choice of  $\epsilon$  is arbitrary, (31) follows. We obtain

$$\begin{aligned} \lim_{n,\ell \rightarrow \infty} \alpha_n^\ell(m) &= \lim_{n,\ell \rightarrow \infty} \max_{j \in \{1, \dots, n-m\}} \alpha_{n,j}^\ell(m) \\ &= \sup_{n,\ell} \max_{j \in \{1, \dots, n-m\}} \alpha_{n,j}^\ell(m) \end{aligned} \quad (32)$$

$$= \sup_{j \in \mathbb{N}} \sup_{n \geq j+m} \sup_{\ell} \alpha_{n,j}^\ell(m) \quad (33)$$

$$= \sup_{j \in \mathbb{N}} \lim_{n,\ell \rightarrow \infty} \alpha_{n,j}^\ell(m) \quad (34)$$

$$= \sup_{j \in \mathbb{N}} \lim_{n,\ell \rightarrow \infty} \alpha(\mathfrak{F}_1^j(\ell), \mathfrak{F}_{j+m}^n(\ell)) \quad (35)$$

$$= \alpha(m), \quad (36)$$

where (32) and (34) follow from the fact that for a fixed  $m \in \mathbb{N}$ ,  $\alpha_{n,j}^\ell(m)$  is an increasing function of  $n, \ell$ , (33) follows from (31), (35) follows from Lemma 1, and (36) follows from (30).  $\square$

The proof of Lemma 2 is provided next. A key ingredient of the analysis is the concentration inequality of Rio, which is stated below for completeness.

**Lemma 5** (Rio [4, Corollary 1.1]). Let  $\langle Y_i \rangle_{i \geq 0}$  be a stationary sequence of  $[-1, 1]$ -valued random variables. Define  $\bar{\alpha}(m) := \alpha(\sigma(Y_0), \sigma(Y_m))$  and suppose that  $\|\bar{\alpha}\| := \sum_{m \in \mathbb{N}} \bar{\alpha}(m) < \infty$ . For each  $t \in \mathbb{N}$ , let  $S_t := Y_0 + \dots + Y_{t-1}$ . We have

$$\text{Var}(S_t) \leq 4t \|\bar{\alpha}\|.$$

*Proof of Lemma 2.* Let  $\bar{\mathbb{N}} := \mathbb{N} \cup \{0\}$ . For each  $D \in \mathcal{D}_{k,\ell}$ ,  $k, \ell \in \mathbb{N}$  define the  $[-1, 1]$ -valued sequence of random variables  $\langle Y_i \rangle_{i \in \bar{\mathbb{N}}}$  where  $Y_i := \chi_D\{X_{ik+1}, \dots, X_{(i+1)k}\} - \mu([D]_1^k)$ ,  $i \in \bar{\mathbb{N}}$ . Observe that  $\langle Y_i \rangle_{i \in \bar{\mathbb{N}}}$  is a zero-mean  $[-1, 1]$ -valued stationary process and for each  $m \in \mathbb{N}$  we have,

$$\begin{aligned} \alpha(\sigma(Y_0), \sigma(Y_m)) &\leq \alpha(\sigma(X_1, \dots, X_k), \sigma(X_{mk+1}, \dots, X_{(m+1)k})) \\ &\leq \alpha(k(m-1) + 1) \end{aligned}$$

where the first inequality follows from the fact that by definition we have  $\sigma(Y_0) \subset \sigma(X_1, \dots, X_k)$  and that  $\sigma(Y_m) \subset \sigma(X_{mk+1}, \dots, X_{(m+1)k})$ , and the second inequality follows from the definition of the  $\alpha$ -mixing coefficients of  $\mathbf{X}$ . Thus, we obtain

$$\begin{aligned} \sum_{m \in \mathbb{N}} \alpha(\sigma(Y_0), \sigma(Y_m)) &\leq \sum_{m \in \mathbb{N}} \alpha(k(m-1) + 1) \\ &\leq \sum_{m \in \mathbb{N}} \alpha(m) \\ &= \|\bar{\alpha}\| \end{aligned}$$

where, the second inequality follows from the fact that  $\alpha(m)$  is a decreasing sequence so that  $\alpha(u) \leq \alpha(v)$ ,  $u > v \in \mathbb{N}$ .

Let  $S_t := \sum_{i=0}^{t-1} Y_i$ . By Chebychev's inequality together with Lemma 5, and noting that by definition  $\bar{\alpha}(m) \leq \alpha(m)$ , we obtain

$$\begin{aligned} \Pr\left(|\mu_t(\mathbf{X}, D) - \mu([D]_1^k)| \geq \epsilon\right) &= \Pr(|S_t| \geq t\epsilon) \\ &\leq \frac{\text{Var}(S_t)}{t^2 \epsilon^2} \\ &\leq \frac{4 \|\bar{\alpha}\|}{t \epsilon^2}. \end{aligned}$$

$\square$

*Proof of Proposition 3.* Fix  $\epsilon > 0$ . For  $t, n \in \mathbb{N}$  define

$$\Omega_{t,n} := \left\{ \max_{D \in \mathcal{D}_{n,\ell}} |\mu_t(\mathbf{X}, D) - \mu([D]_1^n)| \leq \epsilon/2^{2^{m\ell}} \right\}.$$

By Lemma 2 and a union bound we obtain,

$$\Pr(\Omega_{t,n}) \geq 1 - (2^{2^{n\ell} + 2^{m\ell+1}}) \frac{\|\bar{\alpha}\|}{t \epsilon^2}.$$

For all  $\omega \in \Omega_{t,n}$ , and each  $A \in \mathcal{D}_{j,\ell}$ ,  $B \in \mathcal{D}_{j',\ell}$ ,  $j \in \{1, \dots, n-m\}$ ,  $j' := n-m-j+1$  the following inequalities hold. First we have,

$$\begin{aligned} &|\gamma_{t,n}^{m,j}(\mathbf{X}, A, B) - \mu([A]_1^j \cap [B]_{j+m+1}^n)| \\ &\leq \sum_{C \in \mathcal{D}_{m,\ell}} |\mu_t(\mathbf{X}, A \times C \times B) - \mu([A \times C \times B]_1^n)| \\ &\leq \epsilon. \end{aligned}$$

Next, it holds that

$$\begin{aligned} &|\mu_t(\mathbf{X}, A) \mu_t(\mathbf{X}, B) - \mu([A]_1^j) \mu([B]_{j+m+1}^n)| \\ &\leq |\mu_t(\mathbf{X}, A) - \mu([A]_1^j)| |\mu_t(\mathbf{X}, B) - \mu([B]_{j+m+1}^n)| \\ &\quad + \mu([A]_1^j) |\mu_t(\mathbf{X}, B) - \mu([B]_{j+m+1}^n)| \\ &\quad + \mu([B]_{j+m+1}^n) |\mu_t(\mathbf{X}, A) - \mu([A]_1^j)| \\ &\leq \epsilon^2/(2^{2^{m\ell}+1}) + 2\epsilon/(2^{2^{m\ell}}) \leq \epsilon. \end{aligned}$$

Recall the convention  $j' := j - n - m$  for  $j \in \{1, \dots, n - m\}$ . Define the events  $E_1 := \{\exists j, A \in \mathcal{D}_{j,\ell}, B \in \mathcal{D}_{j',\ell} : |\gamma_{t,n}^{m,j}(\mathbf{X}, A, B) - \mu([A]_1^j \cap [B]_{j+m+1}^n)| \geq \epsilon\}$  and  $E_2 := \{\exists j, A \in \mathcal{D}_{j,\ell}, B \in \mathcal{D}_{j',\ell} : |\mu_t(\mathbf{X}, A) - \mu_t(\mathbf{X}, B) - \mu([A]_1^j) - \mu([B]_{j+m+1}^n)| \geq \epsilon\}$ . We obtain,

$$\begin{aligned} \Pr(|\hat{\alpha}_{t,n}^\ell(\mathbf{X}, m) - \alpha_n^\ell(m)| \geq \epsilon) &\leq \Pr(E_1) + \Pr(E_2) \\ &\leq 2 \Pr(\Omega_{t,n}^c) \\ &\leq (2^{2^{n\ell} + 2^{m\ell+1} + 1}) \frac{\|\alpha\|}{t\epsilon^2}. \end{aligned}$$

Thus, (8) gives  $C_{M,\ell,n} = M(2^{2^{n\ell} + 2^{m\ell+1} + 1})$  for each  $M \in \mathbb{N}$ , and we obtain

$$\begin{aligned} \Pr(|\sum_{m=1}^M \hat{\alpha}_{t,n}^\ell(\mathbf{X}, m) - \alpha_n^\ell(m)| \geq \epsilon) \\ \leq \Pr(\sum_{m=1}^M |\hat{\alpha}_{t,n}^\ell(\mathbf{X}, m) - \alpha_n^\ell(m)| \geq \epsilon) \\ \leq \frac{\|\alpha\| C_{M,\ell,k}}{t\epsilon^2}. \end{aligned}$$

□

*Proof of Theorem 1.* Fix  $\epsilon > 0$ . As follows from Proposition 1, there exist  $L_\epsilon, N_\epsilon \in \mathbb{N}$  such that for all  $n \geq N_\epsilon$  and all  $\ell \geq L_\epsilon$  we have,

$$|\alpha_n^\ell(m) - \alpha(m)| \leq \epsilon. \quad (37)$$

Let  $T_\epsilon^{(1)} \in \mathbb{N}$  be such that  $\ell_t \geq L_\epsilon, n_t \geq N_\epsilon$  for  $t \geq T_\epsilon^{(1)}$ . Define the events

$$E_t := \{|\hat{\alpha}_t(\mathbf{X}, m) - \alpha_n^\ell(m)| \geq \epsilon_t\}, \quad t \in \mathbb{N}.$$

Since as specified by (9)  $\tau_t = \frac{C_{m,\ell_t,n_t}}{\epsilon_t^2 \delta_t}$ , then by Proposition 3 for each  $t \in \mathbb{N}$  it holds that

$$\Pr(E_t) \leq \frac{\|\alpha\| C_{m,\ell_t,n_t}}{m\tau_t \epsilon_t^2} \leq \frac{\|\alpha\| \delta_t}{m}. \quad (38)$$

Let  $\delta := \sum_{t \in \mathbb{N}} \delta_t$  and note that  $\delta < \infty$ . Define  $E := \limsup_{t \rightarrow \infty} E_t = \bigcap_{t=1}^\infty \bigcup_{t'=t}^\infty E_{t'}$ . By (38) we have  $\sum_{t \in \mathbb{N}} \Pr(E_t) \leq \frac{\|\alpha\|}{m} \sum_{t \in \mathbb{N}} \delta_t = \frac{\|\alpha\|}{m} \delta < \infty$ , since  $\|\alpha\| < \infty$ . Therefore, by the Borel-Cantelli Lemma we obtain  $\Pr(E) = 0$ . As a result, there exists some random  $T^*$  (which depends on the sample-path) such that for all  $t \geq T^*$  we have

$$|\hat{\alpha}_t(\mathbf{X}, m) - \alpha_n^\ell(m)| \leq \epsilon_t, \quad \mu - \text{almost surely}. \quad (39)$$

Let  $T_\epsilon^{(2)} \in \mathbb{N}$  be large enough such that for all  $t \geq T_\epsilon^{(2)}$  we have  $\epsilon_t \leq \epsilon/2$ . Take  $t \geq \max\{T^*, T_\epsilon^{(1)}, T_\epsilon^{(2)}\}$ , with probability one we have  $|\hat{\alpha}_t(\mathbf{X}, m) - \alpha(m)| \leq |\hat{\alpha}_t(\mathbf{X}, m) - \alpha_n^\ell(m)| + \epsilon \leq \epsilon_t + \epsilon \leq \epsilon$  where the first inequality follows from (37) and the second inequality follows from (39). □

*Proof of Theorem 2.* Fix  $\epsilon > 0$ . Since  $\|\alpha\| < \infty$ , there exists some  $M_\epsilon \in \mathbb{N}$  such that  $\sum_{m=M_\epsilon+1}^\infty \alpha(m) \leq \epsilon/5$ . Let  $\epsilon' := \frac{\epsilon}{5M_\epsilon}$ . As follows from Proposition 1, for each  $t \in \mathbb{N}$ , there exist  $L_{\epsilon'}, N_{\epsilon'} \in \mathbb{N}$  such that for all  $n \geq N_{\epsilon'}$  and all  $\ell \geq L_{\epsilon'}$  we have,

$$\max_{m=1, \dots, M_\epsilon} |\alpha_n^\ell(m) - \alpha(m)| \leq \epsilon'. \quad (40)$$

Let  $T_\epsilon^{(1)} \in \mathbb{N}$  be such that  $M_t \geq M_\epsilon, \ell_t \geq L_{\epsilon'}, n_t \geq N_{\epsilon'}$  for  $t \geq T_\epsilon^{(1)}$ . Define the sequence of events

$$E_t := \left\{ \left| \theta_t(\mathbf{X}) - \sum_{m=1}^{M_t} \alpha_{n_t}^{\ell_t}(m) \right| \geq \epsilon_t \right\}, \quad t \in \mathbb{N}.$$

Since  $\kappa_t = \frac{C_{M_t,\ell_t,n_t}}{\epsilon_t^2 \delta_t}$ , then by Proposition 3 for each  $t \in \mathbb{N}$  we have

$$\Pr(E_t) \leq \frac{\|\alpha\| C_{M_t,\ell_t,n_t}}{t\epsilon_t^2} \leq \|\alpha\| \delta_t.$$

Let  $\delta := \sum_{t \in \mathbb{N}} \delta_t$  and note that  $\delta < \infty$ . Define  $E := \limsup_{t \rightarrow \infty} E_t = \bigcap_{t=1}^\infty \bigcup_{t'=t}^\infty E_{t'}$ . It follows that  $\sum_{t \in \mathbb{N}} \Pr(E_t) \leq \|\alpha\| \sum_{t \in \mathbb{N}} \delta_t = \|\alpha\| \delta < \infty$ . Therefore, by the Borel-Cantelli Lemma we obtain  $\Pr(E) = 0$ . As a result, there exists some random  $T^*$  (which depends on the sample-path) such that for all  $t \geq T^*$  we have

$$\left| \theta_t(\mathbf{X}) - \sum_{m=1}^{M_t} \alpha_{n_t}^{\ell_t}(m) \right| \leq \epsilon_t, \quad \mu - \text{almost surely}.$$

Let  $T_\epsilon^{(2)} \in \mathbb{N}$  be large enough so that  $\epsilon_t \leq \epsilon/5$  for  $t \geq T_\epsilon^{(2)}$ . With probability one, for all  $t \geq \max\{T^*, T_\epsilon^{(1)}, T_\epsilon^{(2)}\}$  obtain,

$$\begin{aligned} & \left| \theta_t(\mathbf{X}) - \|\alpha\| \right| \\ & \leq \left| \theta_t(\mathbf{X}) - \sum_{m=1}^{M_t} \alpha_{n_t}^{\ell_t}(m) \right| \\ & \quad + \left| \sum_{m=1}^{M_t} \alpha_{n_t}^{\ell_t}(m) - \sum_{m=1}^{M_t} \alpha(m) \right| + \epsilon/5 \\ & \leq \sum_{m=1}^{M_\epsilon} |\alpha_{n_t}^{\ell_t}(m) - \alpha(m)| \\ & \quad + \sum_{m=M_\epsilon+1}^{M_t} (\alpha_{n_t}^{\ell_t}(m) + \alpha(m)) + 2\epsilon/5 \quad (41) \\ & \leq \epsilon, \quad (42) \end{aligned}$$

where (41) follows from the fact that for all  $n, \ell, m \in \mathbb{N}$  we have  $\alpha_n^\ell(m) \geq 0$  and  $\alpha(m) \geq 0$ , and (42) follows from (40) and that by definition  $\alpha_n^\ell(m) \leq \alpha(m)$  for all  $n, \ell, m \in \mathbb{N}$ . □

*Proof of Theorem 3.* First, we prove that with probability at least  $1 - \delta$  we have  $\lim_{t \rightarrow \infty} \psi_t(\mathbf{X}) \leq \|\alpha\|$ . If  $\|\alpha\| = \infty$  then there is nothing to prove, so we may assume  $\|\alpha\| < \infty$ . Consider any  $s_t > \|\alpha\|$ . Then

$$\begin{aligned} & \Pr\left(\theta_t(\mathbf{X}) > s_t + \epsilon_t \sqrt{s_t}\right) \\ & \leq \Pr\left(\theta_t(\mathbf{X}) > \sum_{m=1}^{M_t} \alpha(m) + \epsilon_t \sqrt{s_t}\right) \quad (43) \end{aligned}$$

$$\leq \frac{\|\alpha\| C_{M_t,\ell_t,n_t}}{\kappa_t s_t \epsilon_t^2} \quad (44)$$

$$\leq \delta_t, \quad (45)$$

where (43) follows from the fact that  $\sum_{m=1}^{M_t} \alpha(m) \leq \|\alpha\| < s_t$ , (44) follows from Proposition 3 and (45) from the choice of  $\kappa_t$  given by (11), together with noting that  $s_t > \|\alpha\|$ . Observing that  $\sum_{t \in \mathbb{N}} \delta_t = \delta$ , the union bound implies

$\Pr(\{\exists s_t > \|\alpha\| : \theta_t(\mathbf{X}) > s_t + \epsilon_t \sqrt{s_t}\}) \leq \delta$ . It remains to prove that  $\lim_{t \rightarrow \infty} \psi_t(\mathbf{X}) \geq \|\alpha\|$  with probability one. To this end, first assume that  $\|\alpha\| < \infty$ . We show that, with probability one, for any  $\eta < \|\alpha\|$ , there exists some  $s_t > \eta$  such that  $\theta_t(\mathbf{X}) > s_t + \epsilon_t \sqrt{s_t}$ . Fix some  $\eta < \|\alpha\|$  and let  $\varrho = \|\alpha\| - \eta$ . Recall that  $\langle \epsilon_t \rangle$  is a decreasing sequence of positive real numbers with  $\lim_{t \rightarrow \infty} \epsilon_t = 0$ . Therefore, there exists some  $T_1$  such that for all  $t \geq T_1$ ,

$$\|\alpha\| - \frac{7\varrho}{8} + \epsilon_t \sqrt{\|\alpha\| - \frac{7\varrho}{8}} < \|\alpha\| - \frac{3\varrho}{4}. \quad (46)$$

By Theorem 2 there exists some random  $T_2 \in \mathbb{N}$  such that, for all  $t \geq T_2$  with probability one we have,

$$\theta_t(\mathbf{X}) \geq \|\alpha\| - \frac{3\varrho}{4}. \quad (47)$$

Moreover, with  $\mathcal{S}$  chosen to be a countable dense subset of  $[0, \infty)$  we can find some  $t \geq \max\{T_1, T_2\}$  such that  $s_t \in (\eta, \|\alpha\| - 7\varrho/8)$ . It follows that

$$s_t + \epsilon_t \sqrt{s_t} < \|\alpha\| - \frac{3\varrho}{4} \leq \theta_t(\mathbf{X}), \quad (48)$$

where the first inequality follows from the fact that  $s_t < \|\alpha\| - 7\varrho/8$ , and the second inequality follows from (47). As a result, we obtain  $\psi_t(\mathbf{X}) > \eta$  as claimed. Finally, suppose that  $\|\alpha\| = \infty$  and fix some  $\eta \in \mathbb{N}$ . Since  $\|\alpha\| = \infty$ , there exists some  $M^*$  such that  $\sum_{m=1}^{M^*} \alpha(m) \geq \eta + 1$ . By Proposition 1 there exist  $n^*, \ell^* \in \mathbb{N}$  such that  $\max_{m \in 1, \dots, M^*} |\alpha_{n^*}^{\ell^*}(m) - \alpha(m)| \leq \frac{1}{M^*}$  leading to,  $\sum_{m=1}^{M^*} \alpha_{n^*}^{\ell^*}(m) \geq \eta$ . Let  $T'_1$  be large enough so that  $M_t \geq M^*$ ,  $n_t \geq n^*$ , and  $\ell_t \geq \ell^*$  for all  $t \geq T'_1$ . Fix some  $\epsilon^* > 0$ . By Lemma 2 there exists some random  $N_{\epsilon^*}$  such that for all  $u \geq N_{\epsilon^*}$  with probability one we have  $\max_{m \in 1, \dots, M^*} |\hat{\alpha}_{u, n^*}^{\ell^*}(m) - \alpha_{n^*}^{\ell^*}(m)| \leq \epsilon^*/2M^*$  so that

$$\sum_{m=1}^{M^*} \hat{\alpha}_{u, n^*}^{\ell^*}(\mathbf{X}, m) \geq \eta - \epsilon^*/2. \quad (49)$$

Since  $\epsilon_t \downarrow 0$ , there exists some  $T'_2$  such that for all  $t \geq T'_2$  it holds that

$$\eta - \epsilon^* + \epsilon_t \sqrt{\eta - \epsilon^*} < \eta - \epsilon^*/2. \quad (50)$$

Let  $T'_3$  be large enough so that  $\kappa_t$  given by (11) is larger than  $N_{\epsilon^*}$  for all  $t \geq T'_3$ . Recalling that  $\mathcal{S}$  is a countable dense subset of  $[0, \infty)$  we can find some  $t \geq \max\{T'_1, T'_2, T'_3\}$  such that  $s_t \in (\eta - 2\epsilon^*, \eta - \epsilon^*)$ . With probability one it holds that,

$$s_t + \epsilon_t \sqrt{s_t} < \eta - \epsilon^*/2 \quad (51)$$

$$\leq \sum_{m=1}^{M^*} \hat{\alpha}_{\kappa_t, n^*}^{\ell^*}(\mathbf{X}, m) \quad (52)$$

$$\leq \sum_{m=1}^{M_t} \hat{\alpha}_{\kappa_t, n_t}^{\ell_t}(\mathbf{X}, m) \quad (53)$$

$$= \theta_t(\mathbf{X}) \quad (54)$$

where (51) follows from (50), (52) follows from (49), and (53) follows from the fact that  $t \geq T'_1$  (and hence  $M_t \geq M^*$ ,  $n_t \geq n^*$ , and  $\ell_t \geq \ell^*$ ) together with the monotonicity of  $\hat{\alpha}_{\kappa, n}^{\ell}(m)$ . Since the above holds for any  $\eta \in \mathbb{N}$ , the result follows.  $\square$

*Proof of Theorem 4.* First we show that

$$\limsup_{t \rightarrow \infty} \xi_t(\mathbf{X}) \leq \|\alpha\| \quad \mu - \text{almost surely}. \quad (55)$$

This trivially holds when  $\|\alpha\| = \infty$ . So, suppose that  $\|\alpha\| < \infty$ . By an argument similar to that given in the proof Theorem 3, whenever  $u_t > \|\alpha\|$ , we have  $\Pr(\theta_t(\mathbf{X}) > u_t + \epsilon_t \sqrt{u_t}) \leq \delta_t$ . Since  $\sum_{t \geq 1} \delta_t < \infty$ , by the Borel-Cantelli lemma, we have

$$\sum_{t: u_t > \|\alpha\|} \chi\{\theta_t(\mathbf{X}) > u_t + \epsilon_t \sqrt{u_t}\} < \infty, \quad \mu - \text{almost surely}.$$

Hence, with probability one, there exists a random time  $T_1 \in \mathbb{N}$  such that, for all  $t > T_1$ ,

$$\theta_t(\mathbf{X}) \leq u_t + \epsilon_t \sqrt{u_t} \quad \text{whenever } u_t > \|\alpha\|. \quad (56)$$

In particular, let  $E := \{s \in \mathcal{S} \cap (\|\alpha\|, \infty) : b_{T_1+1}(s) = 0\}$ , and observe that  $|E| < \infty$ . Therefore, there exists some time  $T_2 > T_1$  such that, for every  $s \in E$  there exists  $t \in \{T_1 + 1, \dots, T_2\}$  with  $u_t = s$ . As follows from (56), for all  $t \geq T_2$ , with probability one we have  $b_t(s) = 1$  for all  $s > \|\alpha\|$  and (55) follows. The fact that

$$\liminf_{t \rightarrow \infty} \xi_t(\mathbf{X}) \geq \|\alpha\|, \quad \text{almost surely}, \quad (57)$$

for  $\|\alpha\| < \infty$  follows analogously. To see this, take any  $\eta < \|\alpha\| < \infty$  and define  $\varrho := \|\alpha\| - \eta$ . As in the proof of Theorem 3, there exists some  $s \in \mathcal{S} \cap (\eta, \|\alpha\| - 7\varrho/8)$  and a random time  $T'_1$  such that for all  $t \geq T'_1$ , with probability one, we have

$$\theta_t(\mathbf{X}) > u_t + \epsilon_t \sqrt{u_t} \quad \text{whenever } u_t \leq s. \quad (58)$$

The set  $E' := \{s \in \mathcal{S} \cap (\eta, \|\alpha\| - 7\varrho/8) : b_{T'_1+1}(s) = 1\}$  can only have finitely many elements, each of which will eventually be visited; that is, we can find some  $T'_2 > T'_1$  such that for every  $s \in E'$  there exists some  $t \in \{T'_1 + 1, \dots, T'_2\}$  with  $u_t = s$ . By (58) for all  $t \geq T'_2$  with probability one it holds that  $b_t(s) = 0$  for all  $s \in \mathcal{S} \cap (\eta, \|\alpha\| - 7\varrho/8)$  and (57) follows. Now suppose that  $\|\alpha\| = \infty$ . By an argument analogous to the above, for each  $\eta \in \mathbb{N}$  and every  $\epsilon^* > 0$  there exists some  $s \in (\eta - 2\epsilon^*, \eta - \epsilon^*)$  and a random  $T''_1 \in \mathbb{N}$  such that for all  $t \geq T''_1$  with probability one we have

$$\theta_t(\mathbf{X}) > \eta - \epsilon^* \geq u_t + \epsilon_t \sqrt{u_t} \quad \text{whenever } u_t \leq s. \quad (59)$$

Again, by (59) and noting that the set  $E'' := \{s \in \mathcal{S} \cap (\eta - 2\epsilon^*, \eta - \epsilon^*) : b_{T''_1+1}(s) = 1\}$  has finite cardinality, we can find some  $T''_2$  such that for all  $t \geq T''_2$  with probability one we have  $b_t(s) = 0$  for all  $s \in \mathcal{S} \cap (\eta - 2\epsilon^*, \eta - \epsilon^*)$ . Since this holds for every  $\eta \in \mathbb{N}$  and any  $\epsilon^* > 0$ , it follows that when  $\|\alpha\| = \infty$  we have  $\liminf_{t \rightarrow \infty} \xi_t(\mathbf{X}) \geq \infty$ ,  $\mu$  - almost surely. This completes the proof.  $\square$

*Proof of Lemma 3.* Fix  $m, \ell \in \mathbb{N}$ ,  $n > m \in \mathbb{N}$  and some  $j \in \{1, \dots, n - m\}$ . Define  $\mathcal{C}_1^j(\ell) := \{[A]_1^j : A \in \mathcal{D}_{j, \ell}\}$  and  $\mathcal{C}_{j+m}^n(\ell) := \{[B]_{j+m+1}^n : B \in \mathcal{D}_{j', \ell}, j' := n - m - j\}$ . Observe that  $\sigma(\mathcal{C}_1^j(\ell)) = \mathfrak{F}_1^j(\ell)$  and that  $\sigma(\mathcal{C}_{j+m}^n(\ell)) = \mathfrak{F}_{j+m}^n(\ell)$ . Moreover,  $\mathfrak{F}_1^j(\ell)$  and  $\mathfrak{F}_{j+m}^n(\ell)$  are atomic, with the elements of the countable collections  $\mathcal{C}_1^j(\ell)$  and  $\mathcal{C}_{j+m}^n(\ell)$  as their atoms

respectively. By Bradley [1, Vol. 1, Proposition 3.21] we can write  $\beta(\mathfrak{F}_1^j(\ell), \mathfrak{F}_{j+m}^n(\ell))$  as,

$$\frac{1}{2} \sum_{U \in \mathcal{C}_1^j(\ell)} \sum_{V \in \mathcal{C}_{j+m}^n(\ell)} |\mu(U \cap V) - \mu(U)\mu(V)| = \beta_{n,j}^\ell(m).$$

□

*Proof sketch of Proposition 5.* The result follows in much the same way as in Proposition 3, by observing that since  $\beta(m) \geq \alpha(m)$ ,  $m \in \mathbb{N}$ , as follows from Lemma 2 for all  $D \in \mathcal{D}_{n,\ell}$ ,  $n, \ell, t \in \mathbb{N}$  and every  $\epsilon > 0$  we have,

$$\Pr(|\mu_t(\mathbf{X}, D) - \mu([D]_1^n)| \geq \epsilon) \leq \frac{4\|\beta\|}{t\epsilon^2}.$$

Thus, for a fixed  $\epsilon > 0$  and  $t, n \in \mathbb{N}$  we can deduce that  $\Pr(\Omega_{t,n}) \geq 1 - (2^{2^{2n\ell+1}+2^{m\ell+1}+2}) \frac{\|\beta\|}{t\epsilon^2}$  where

$$\Omega_{t,n} := \left\{ \max_{D \in \mathcal{D}_{n,\ell}} |\mu_t(\mathbf{X}, D) - \mu([D]_1^n)| \leq \epsilon / (2^{2^{2n\ell+1}+2^{m\ell+1}+2}) \right\}.$$

Next, as in the proof of Proposition 3 we can obtain

$$\begin{aligned} \Pr(|\hat{\beta}_{t,n}^\ell(\mathbf{X}, m) - \beta_n^\ell(m)| \geq \epsilon) &\leq 2\Pr(\Omega_{t,n}^c) \\ &\leq (2^{2^{2n\ell+1}+2^{m\ell+1}+2}) \frac{\|\beta\|}{t\epsilon^2}, \end{aligned}$$

which together with a union bound gives the second concentration bound as well. □

*Proof of Theorem 7.* Consider a rate function  $\gamma : \mathbb{N} \rightarrow [0, 1]$  and denote by  $\mathcal{R}_\gamma^\alpha$  the class of processes with  $\gamma$  as rate function of their  $\alpha$ -mixing coefficients. Consider a process measure  $\mu \in \mathcal{C}^\alpha$  with corresponding sequence of random variables  $\mathbf{X}$ . First, observe that  $g_t$  is  $\mathfrak{F}_t$ -measurable. To prove that  $g_t$  is consistent in the sense of (21), we proceed as follows. Consider a process  $\mathbf{X}$  with process distribution  $\mu$ . Let  $E_t := \{\exists m \in \{1, \dots, M_t\} \text{ such that } \hat{\alpha}_t(\mathbf{X}, m) > \gamma(m) + \epsilon_t\}$ . If  $\mu \in \mathcal{C}^\alpha \cap \mathcal{R}_\gamma^\alpha$ , we have

$$\begin{aligned} \Pr(E_t) &\leq \sum_{m=1}^{M_t} \Pr(\hat{\alpha}_t(\mathbf{X}, m) > \gamma(m) + \epsilon_t) \\ &\leq \sum_{m=1}^{M_t} \Pr(\hat{\alpha}_t(\mathbf{X}, m) > \alpha(m) + \epsilon_t) \end{aligned} \quad (60)$$

$$\leq \sum_{m=1}^{M_t} \Pr(|\hat{\alpha}_t(\mathbf{X}, m) - \alpha_{n_t}^\ell(m)| > \epsilon_t) \quad (61)$$

$$\leq \|\alpha\| \delta_t, \quad (62)$$

where (60) follows from the fact that  $\mu \in \mathcal{R}_\gamma^\alpha$ , (61) from observing that  $\alpha_{n_t}^\ell(m) \leq \alpha(m)$ ,  $m \in \mathbb{N}$ , and (62) follows from Proposition 3. Noting that  $\|\alpha\| < \infty$  and  $\delta_t$  is summable, we have  $\sum_{t \in \mathbb{N}} \Pr(E_t) \leq \|\alpha\| \sum_{t \in \mathbb{N}} \delta_t = \|\alpha\| \delta < \infty$ . Therefore,  $\Pr(\bigcap_{t=1}^\infty \bigcup_{t'=t}^\infty E_{t'}) = 0$  by the Borel-Cantelli Lemma. This means that there exists some random  $\tau$  (which depends on the sample-path) such that for all  $t \geq \tau$  we have  $g_t(\mathbf{X}) = +1$ ,  $\mu$ -almost surely. On the other hand, suppose  $\mu \in \mathcal{C}^\alpha \setminus \mathcal{R}_\gamma^\alpha$  and observe that there exists some  $m^* \in \mathbb{N}$  and some  $\delta^* \in (0, 1]$  such that  $\alpha(m^*) - \gamma(m^*) = \delta^*$ . Let  $\tilde{E}_t := \{\exists m \in \{1, \dots, M_t\} \text{ such that } \hat{\alpha}_t(m) \leq \gamma(m) + \epsilon_t\}$ . Recalling that  $\langle \ell_t \rangle$  and  $\langle n_t \rangle$  are increasing sequences, by

Proposition 1 there exists some  $T_1$  such that for all  $t \geq T_1$  we have

$$|\alpha_{n_t}^{\ell_t}(m^*) - \alpha(m^*)| \leq \delta^*/2. \quad (63)$$

Since  $\langle M_t \rangle$  is an increasing sequence, we can find some  $T_2 \in \mathbb{N}$  such that  $m^* \leq M_t$  for all  $t \geq T_2$ . Moreover, since  $\langle \epsilon_t \rangle$  decreases with  $t$  there exists some  $T_3$  such that  $\epsilon_t \leq \delta^*/4$  for all  $t \geq T_3$ . Let  $T^* := \max\{T_1, T_2, T_3\}$ . For  $t \geq T^*$  we have,

$$\begin{aligned} \Pr(\tilde{E}_t) &\leq \Pr(\hat{\alpha}_t(\mathbf{X}, m^*) \leq \gamma(m^*) + \epsilon_t) \\ &\leq \Pr(\hat{\alpha}_t(\mathbf{X}, m^*) \leq \gamma(m^*) + \delta^*/2 - \epsilon_t) \end{aligned} \quad (64)$$

$$\leq \Pr(\hat{\alpha}_t(\mathbf{X}, m^*) - \alpha_{n_t}^{\ell_t}(m^*) \leq -\epsilon_t) \quad (65)$$

$$\leq \Pr(|\hat{\alpha}_t(\mathbf{X}, m^*) - \alpha_{n_t}^{\ell_t}(m^*)| \geq \epsilon_t) \quad (66)$$

$$\leq \|\alpha\| \delta_t, \quad (67)$$

where (64) follows from the fact that  $\epsilon_t \leq \delta^*/4$  for  $t \geq T_3$  so that  $\delta^*/2 - \epsilon_t \geq \epsilon_t$ , (65) follows from (63) and (67) follows from Proposition 3 and the choice of  $\tau_t$ . In much the same way as above, we have  $\sum_{t=T^*}^\infty \Pr(\tilde{E}_t) \leq \|\alpha\| \delta < \infty$  and by the Borel-Cantelli Lemma we obtain  $\Pr(\bigcap_{t=1}^\infty \bigcup_{t'=t}^\infty \tilde{E}_{t'}) \leq \Pr(\bigcap_{t=T^*}^\infty \bigcup_{t'=t}^\infty \tilde{E}_{t'}) = 0$ . Thus, there is some random  $T$  such that for all  $t \geq T$  we have

$$g_t(\mathbf{X}) = -1, \quad \mu - \text{almost surely.}$$

□

*Proof sketch of Proposition 4.* Fix  $m, \ell \in \mathbb{N}$  and  $n > m$ . Observe that for each  $j \in \{1, \dots, n-m\}$  we have  $\mathfrak{F}_1^j(\ell) \subseteq \mathfrak{F}_1^j(\ell+1)$ ,  $\ell \in \mathbb{N}$  and  $\mathfrak{F}_{j+m}^n(\ell) \subseteq \mathfrak{F}_{j+m}^{n+1}(\ell+1)$ ,  $n, \ell \in \mathbb{N}$  are each a sequence of  $\sigma$ -algebras with  $\bigvee_\ell \mathfrak{F}_1^j(\ell) = \sigma(X_1, \dots, X_j)$  and  $\bigvee_{n,\ell} \mathfrak{F}_{j+m}^n(\ell) = \sigma(\{X_t : t \geq j+m+1\})$ . Therefore, by Bradley [1, Vol. 1, Proposition 3.18] we can deduce that  $\lim_{n,\ell \rightarrow \infty} \beta(\mathfrak{F}_1^j(\ell), \mathfrak{F}_{j+m}^n(\ell))$  is given by

$$\beta(\sigma(\{X_t : 1 \leq t \leq j\}), \sigma(\{X_t : t \geq j+m+1\})).$$

As in Lemma 1, it is straightforward to check that for each  $m \in \mathbb{N}$  we have

$$\sup_{n,\ell} \max_{j \in \{1, \dots, n-m\}} \beta_{n,j}^\ell(m) = \sup_{j \in \mathbb{N}} \sup_{n \geq j+m+1} \sup_{\ell} \beta_{n,j}^\ell(m). \quad (68)$$

Thus, analogously to the proof of Lemma 1, by (68) and Lemma 3 we obtain  $\lim_{n,\ell \rightarrow \infty} \beta_n^\ell(m) = \beta(m)$ . □

## VII. ACKNOWLEDGMENTS

The work of Azadeh Khaleghi was supported in part by a grant of the French National Research Agency (ANR), “Investissements d’Avenir” (LabEx Ecodec/ANR-11-LABX-0047). Gábor Lugosi acknowledges the support of Ayudas Fundacion BBVA a Proyectos de Investigación Científica 2021 and the Spanish Ministry of Economy and Competitiveness, Grant PGC2018-101643-B-I00 and FEDER, EU.

## REFERENCES

- [1] R. C. Bradley, *Introduction to Strong Mixing Conditions*. Kendrick Press, 2007, vol. 1,2, 3.
- [2] I. A. Ibragimov, "Some limit theorems for stationary processes," *Theory of Probability & its Applications*, vol. 7, no. 4, pp. 349–382, 1962.
- [3] P. Doukhan, *Mixing: Properties and Examples*. Springer Lecture Notes, 1994.
- [4] E. Rio, *Théorie asymptotique des processus aléatoires faiblement dépendants*. Springer, 1999.
- [5] P. Bertail, P. Doukhan, and P. Soulier, *Dependence in probability and statistics*. Springer, 2006.
- [6] D. Bosq, *Nonparametric statistics for stochastic processes: estimation and prediction*. Springer Science & Business Media, 2012, vol. 110.
- [7] S. Grünewälder and A. Khaleghi, "Approximations of the restless bandit problem," *The Journal of Machine Learning Research*, vol. 20, no. 1, pp. 514–550, 2019.
- [8] P.-M. Samson, "Concentration of measure inequalities for markov chains and  $\phi$ -mixing processes," *The Annals of Probability*, 2000.
- [9] M. Rosenblatt, "A central limit theorem and a strong mixing condition," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 42, no. 1, p. 43, 1956.
- [10] Y. A. Rozanov and V. Volkonskii, "Some limit theorems for random functions i," *Theory of Probability & its Applications*, vol. 4, no. 2, pp. 178–197, 1959.
- [11] V. Volkonskii and Y. A. Rozanov, "Some limit theorems for random functions. ii," *Theory of Probability & its Applications*, vol. 6, no. 2, pp. 186–198, 1961.
- [12] S. P. Meyn and R. L. Tweedie, *Markov chains and stochastic stability*. Springer Science & Business Media, 2012.
- [13] P. Fryzlewicz and S. Rao, "Mixing properties of arch and time-varying arch processes," *Bernoulli*, vol. 17, no. 1, pp. 320–346, 2011.
- [14] C. S. D. McDonald and M. Schervish, "Estimating beta-mixing coefficients," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence & Statistics*, 2011, pp. 516–524.
- [15] D. J. McDonald, C. R. Shalizi, and M. Schervish, "Estimating beta-mixing coefficients via histograms," *Electronic Journal of Statistics*, vol. 9, no. 2, pp. 2855–2883, 2015.
- [16] A. B. Nobel, "Hypothesis testing for families of ergodic processes," *Bernoulli*, vol. 12, no. 2, pp. 251–269, 2006.
- [17] G. Morvai and B. Weiss, "Testing stationary processes for independence," in *Annales de l'Institut Henri Poincaré Probabilités et statistiques*, vol. 47, no. 4, 2011, pp. 1219–1225.
- [18] S. Grünewälder, "Plug-in estimators for conditional expectations and probabilities," in *International Conference on Artificial Intelligence & Statistics*, 2018, pp. 1513–1521.

**Azadeh Khaleghi** is a professor of Statistics at École Nationale de la Statistique et de l'Administration Économique (ENSAE) - Centre de Recherche en Économie et Statistique (CREST). She received her B.Sc. and M.Sc. degrees from the University of Toronto, and her Ph.D. degree from INRIA - University of Lille I. Her research interests lie in the mathematical foundations of statistics, machine learning, and information theory.

**Gábor Lugosi** received his Ph.D. from the Hungarian Academy of Sciences in 1991. Since 1996, he has been at the Department of Economics, Pompeu Fabra University, Barcelona where in 2006 he became an ICREA research professor. His research interests include the theory of machine learning, random structures, nonparametric statistics, inequalities in probability, and information theory.