Uniform Convergence of Vapnik-Chervonenkis Classes Under Ergodic Sampling

Terrence M. Adams * and Andrew B. Nobel †

August 5, 2009

Abstract

We show that if $\mathcal{X}$ is a complete separable metric space and $\mathcal{C}$ is a countable family of Borel subsets of $\mathcal{X}$ with finite VC dimension, then for every stationary ergodic process with values in $\mathcal{X}$, the relative frequencies of sets $C \in \mathcal{C}$ converge uniformly to their limiting probabilities. Beyond ergodicity, no assumptions are placed on the sampling process, and no regularity conditions are imposed on the elements of $\mathcal{C}$. The result extends existing work of Vapnik and Chervonenkis and others, who have studied uniform convergence for i.i.d. and strongly mixing processes. Our method of proof is new and direct: it does not rely on symmetrization techniques, probability inequalities, or mixing conditions. The uniform convergence of relative frequencies for VC-major and VC-graph classes of functions under ergodic sampling is established as a corollary of the basic result for sets.

Updated version. Original submitted for publication January 2008.

1 Introduction

The strong law of large numbers and its extension to dependent processes via the ergodic theorem is one of the central results of Probability. The strong law connects sampling and population based quantities, and is one of the basic tools for establishing the consistency of statistical inference procedures. Uniform laws of large numbers extend the strong law by guaranteeing the uniform convergence of averages to their limiting expectations over a given

*Terrence Adams is with the Department of Defense, 9800 Savage Rd. Suite 6513, Ft. Meade, MD 20755

†Andrew Nobel is with the Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, NC 27599-3260. Email: nobel@email.unc.edu. Corresponding Author.

family of functions. Uniform laws of large numbers have been widely used and extensively studied in a number of fields, including statistics, where they play a foundational role in the theory of empirical processes and machine learning. In the latter, they underly many results on consistency and rates of convergence for classification and regression procedures.

The majority of work on uniform laws of large numbers to date has considered independent, identically distributed samples, though there is also a substantial literature concerned with dependent sequences satisfying a variety of mixing conditions. The primary focus of this paper is the uniform convergence of relative frequencies over a family of sets for general ergodic processes. In particular, we show that a sufficient condition for uniform convergence in the i.i.d. case, namely having finite Vapnik-Chervonenkis (VC) dimension, is sufficient to ensure uniform convergence in the ergodic case as well. The VC-dimension is a combinatorial quantity that describes the ability of a collection of sets to pick apart finite subsets of points. It can be defined without reference to metrics, epsilon-coverings, metric entropies, or standard notions of vector space dimension.

Let $\mathbf{X} = X_1, X_2, \ldots$ be a stationary sequence of random variables taking values in a complete separable metric space $\mathcal{X}$ equipped with its associated Borel sigma-field $\mathcal{S}$. Under the standard definition $\mathbf{X}$ is ergodic if its invariant sigma field is trivial (*c.f.* Definition 6.30 in Breiman [4]). An equivalent, mixing-based definition of ergodicity can be formulated as follows. For each $k \geq 1$ let $\mathcal{S}^k$ denote the usual product sigma field on $\mathcal{X}^k$. Then the process $\mathbf{X}$ is ergodic if for each $k \geq 1$ and every $A, B \in \mathcal{S}^k$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{P}(X_1^k \in A, X_{i+1}^{i+k} \in B) \;\to\; \mathbb{P}(X_1^k \in A)\, \mathbb{P}(X_1^k \in B), \tag{1}$$

where X_1^k denotes the k-tuple $(X_1, \ldots, X_k)$. The condition simply states that, on average, the present and the future of $\mathbf{X}$ become independent as the gap between them grows.

Suppose that $\mathbf{X}$ is ergodic. Here and in what follows we let X denote a random variable independent of $\mathbf{X}$ and having the same distribution as X_1. For each set $C \in \mathcal{S}$, the ergodic theorem ensures that the relative frequency $m^{-1}\sum_{i=1}^m I_C(X_i)$ of C converges almost surely to the probability $\mathbb{P}(X \in C)$ as m tends to infinity. Of interest here are families of sets over which this convergence is uniform. To this end, define the random variables

$$\Gamma_m(\mathcal{C} : \mathbf{X}) \;\triangleq\; \sup_{C\in\mathcal{C}} \left| \frac{1}{m} \sum_{i=1}^m I(X_i \in C) \,-\, \mathbb{P}(X \in C) \right|, \quad m \geq 1. \tag{2}$$

A countable family $\mathcal{C}$ of Borel measurable sets is said to be a *Glivenko-Cantelli* class for $\mathbf{X}$ if the relative frequencies of $C \in \mathcal{C}$ converge uniformly to their limiting probabilities, in the sense that

$$\Gamma_m(\mathcal{C} : \mathbf{X}) \;\to\; 0 \;\text{ with probability one as } m \to \infty. \tag{3}$$

Note that the uniformity here is over the family $\mathcal{C}$, and not the underlying sample space; following standard usage the term "uniform convergence" is used rather than the more traditional "equiconvergence". The assumption that $\mathcal{C}$ is countable ensures that the supremum in (3) is measurable. Uncountable families are discussed briefly below.

Vapnik and Chervonenis [25] established necessary and sufficient conditions for (3) under i.i.d. sampling. Their work provides a connection between uniform convergence and the combinatorial complexity of a family $\mathcal{C}$, where the latter is measured by the ability of the family to break apart finite sets of points. Let $\mathcal{C}$ be any collection of subsets of $\mathcal{X}$, and let $D \subseteq \mathcal{X}$ be any finite set of points. The shatter coefficient (or index) of $\mathcal{C}$ with respect to D is defined by

$$S(D : \mathcal{C}) = |\{C \cap D : C \in \mathcal{C}\}|, \tag{4}$$

and is simply the number of distinct subsets of D that can be captured by sets $C \in \mathcal{C}$. Clearly $S(D : \mathcal{C}) \leq 2^{|D|}$. When equality holds, $\mathcal{C}$ is said to shatter the set D. The result of Vapnik and Chervonenkis can be stated as follows.

Theorem A (Vapnik and Chervonenkis [25])**.** *If $X_1, X_2, \ldots$ are i.i.d., then the uniform strong law (3) holds if and only if*

$$\frac{1}{n} \log S(\{X_1, \ldots, X_n\} : \mathcal{C}) \to 0$$

in probability as n tends to infinity.

In subsequent work, Vapnik and Chervonenkis [26] characterized uniform convergence for classes of real-valued functions through the related notion of metric entropy. Talagrand [23] later provided a characterization of uniform convergence in the i.i.d. case that strengthens these results, and is focused on what happens when uniform convergence fails. For non-atomic distributions his results show that (3) fails to hold if and only if there is a set $A \in \mathcal{S}$ with $P(A) > 0$ such that, for almost every realization of $\mathbf{X}$, the family $\mathcal{C}$ shatters the set $\{X_{n_1}, X_{n_2}, \ldots\}$ consisting of those X_i that lie in A.

Definition: The VC-dimension of a family $\mathcal{C}$, denoted here by $\dim(\mathcal{C})$, is the largest integer $k \geq 1$ such that $S(D : \mathcal{C}) = 2^k$ for some k-element subset D of $\mathcal{X}$. If for every $k \geq 1$ the family $\mathcal{C}$ can shatter some k-element set, then $\dim(\mathcal{C}) = +\infty$.

A family $\mathcal{C}$ is said to be a VC class if $\dim(\mathcal{C})$ is finite. The following combinatorial result of Sauer provides polynomial bounds on the shatter coefficients of VC classes in terms of their combinatorial dimension.

Lemma A (Sauer [21])**.** *If* $\dim(\mathcal{C}) = V < \infty$ *then* $S(D : \mathcal{C}) \leq \sum_{j=0}^{V} \binom{m}{j} \leq (m+1)^V$ *for every* $m \geq V$ *and every* $D \subseteq \mathcal{X}$ *of cardinality* m.

It follows from Lemma A and Theorem A that if $V = \dim(C) < \infty$ then $\mathcal{C}$ is a Glivenko-Cantelli class for every i.i.d. process $\mathbf{X}$. Indeed, one may establish an exponential inequality of the form $\mathbb{P}(\Gamma_m(\mathcal{C} : \mathbf{X}) > t) \leq c_1\,(m+1)^V\, e^{-c_2\, m t^2}$ for every $t > 0$ and $m \geq 1$, where c_1 and c_2 are constants that are independent of m, $\mathcal{C}$ and the distribution of $\mathbf{X}$ (*c.f.* [5]). The notions of VC-class and the VC dimension play a central role in modern central limit and empirical process theory, see [24, 6] and the references therein.

1.1 Principal Result

In this paper we show that the uniform strong law (3) holds for VC classes under general ergodic sampling schemes. No mixing conditions are imposed beyond ergodicity, and no conditions are imposed on the elements of $\mathcal{C}$. Under these circumstances, the convergence guaranteed by the ergodic theorem can be arbitrarily slow, and we cannot hope to obtain distribution free probability bounds like those discussed above for the i.i.d. case. Nevertheless, asymptotic results are still possible. Our principal result is the following theorem; its proof can be found in Sections 2 and 3 below.

Theorem 1. *Let* $\mathcal{X}$ *be a complete separable metric space equipped with its Borel measurable subsets* $\mathcal{S}$*, and let* $\mathcal{C} \subseteq \mathcal{S}$ *be any countable family of sets. If* $\dim(\mathcal{C}) < \infty$*, then for every stationary ergodic process* $\mathbf{X} = X_1, X_2, \ldots$ *taking values in* $(\mathcal{X}, \mathcal{S})$*,*

$$\Gamma_m(\mathcal{C} : \mathbf{X}) \;=\; \sup_{C \in \mathcal{C}} \left| \frac{1}{m} \sum_{i=1}^{m} I(X_i \in C) \;-\; \mathbb{P}(X \in C) \right| \;\to\; 0 \;\; \textit{wp1} \tag{5}$$

as m *tends to infinity. In other words,* $\mathcal{C}$ *is a Glivenko-Cantelli class for every stationary ergodic process.*

1.2 Uncountable Families of Sets

The assumption that the family $\mathcal{C}$ is countable ensures that the suprema $\Gamma_m(\mathcal{C} : \mathbf{X})$ are measurable, and is required for the construction of the isomorphism in Lemma 6. In addition, countability of $\mathcal{C}$ is used in the proof of Propositon 3 to ensure that no sample X_i takes values in the boundary of any set $C \in \mathcal{C}$.

Although it can be weakened in many cases (see the discussion below), the assumption that $\mathcal{C}$ is countable cannot be dropped altogether, as it excludes somewhat pathological examples that may arise in the dependent setting. To illustrate, let μ be a non-atomic measure on $(\mathcal{X}, \mathcal{S})$, and let $T : \mathcal{X} \to \mathcal{X}$ be an ergodic μ-measure preserving bijection of $\mathcal{X}$. (More concretely, one may take T to be an irrational rotation of the unit circle with its uniform measure.) Let T^i denote the i-fold composition of T with itself if $i \geq 1$, the i-fold composition of T^{-1} with itself if $i \leq -1$ and the identity if $i = 0$. For each $x \in \mathcal{X}$ let $C_x = \cup_{i=-\infty}^{\infty} \{T^i x\}$ be the trajectory of x under T, and define the family $\mathcal{C} = \{C_x : x \in \mathcal{X}\}$. It is easy to see that for any two points $x_1, x_2 \in \mathcal{X}$, either $C_{x_1} = C_{x_2}$, or $C_{x_1} \cap C_{x_2} = \emptyset$, and therefore the VC-dimension of $\mathcal{C}$ equals one. Now let $X_i = T_i X_0$, where $X_0 \in \mathcal{X}$ is distributed according to μ. Then the process $\mathbf{X} = X_0, X_1, \ldots$ is stationary and ergodic. Moreover, the μ-measure of the countable set C_x is zero for every x, and it is easy to see that $\Gamma_m(\mathcal{C} : \mathbf{X}) = 1$ with probability one for each $m \geq 1$. Thus (5) fails to hold.

In spite of such negative examples, Theorem 1 can be extended in a straightforward way to uncountable classes $\mathcal{C}$ under a natural approximation condition. Call an uncountable family $\mathcal{C} \subseteq \mathcal{S}$ "nice" for a given process $\mathbf{X}$ if $\Gamma_m(\mathcal{C} : \mathbf{X})$ is measurable for each $m \geq 1$, and for every $\epsilon > 0$ there exists a countable sub-family $\mathcal{C}_0 \subseteq \mathcal{C}$ such that $\limsup_m \Gamma_m(\mathcal{C} : \mathbf{X}) \leq \limsup_m \Gamma_m(\mathcal{C}_0 : \mathbf{X}) + \epsilon$ with probability one. If $\mathcal{C}$ has finite VC-dimension, then (5) holds for every ergodic process $\mathbf{X}$ such that $\mathcal{C}$ is nice for $\mathbf{X}$.

Theorem 1 can also be extended to the case in which the elements of $\mathcal{C}$ belong to the completion of the Borel sigma field of $\mathcal{X}$ with respect to the common distribution of the X_i.

1.3 Families of Functions

Theorem 1 can be used to establish two related uniform convergence results for families of functions. These results are presented below. In each case, the results can be extended to uncountable families $\mathcal{F}$ under approximation conditions like those above for families of sets.

A countable family $\mathcal{F}$ of Borel-measurable functions $f : \mathcal{X} \to \mathbb{R}$ is said to be a Glivenko-Cantelli class for a stationary ergodic process $\mathbf{X}$ if the relative frequencies of functions in f converge uniformly to their limiting expectations, *i.e.*,

$$\Gamma_m(\mathcal{F} : \mathbf{X}) \triangleq \sup_{f \in \mathcal{F}} \left| \frac{1}{m} \sum_{i=1}^{m} f(X_i) - Ef(X) \right| \to 0 \text{ wp1 as } m \to \infty. \tag{6}$$

Here we assume that the expectation $Ef(X)$ is well defined for each $f \in \mathcal{F}$. Recall that a measurable function $F : \mathcal{X} \to [0, \infty)$ is said to be an envelope for $\mathcal{F}$ if $|f(x)| \leq F(x)$ for each $x \in \mathcal{X}$ and $f \in \mathcal{F}$. In particular, $\mathcal{F}$ is bounded if it has constant envelope $F = M < \infty$.

1.3.1 VC Major Classes

Let $L_f(\alpha) = \{x : f(x) \leq \alpha\}$ denote the α level-set of a function $f : \mathcal{X} \to \mathbb{R}$. A family of functions $\mathcal{F}$ is said to be a VC-major class if

$$\dim_{VC}(\mathcal{F}) \;=\; \sup_{\alpha \in \mathbb{R}} \dim(\{L_f(\alpha) : f \in \mathcal{F}\})$$

is finite. The following result is established in Section 4.1.

Proposition 1. *Let $\mathcal{F}$ be a countable family of Borel-measurable functions $f : \mathcal{X} \to \mathbb{R}$ with envelope F. If $\mathcal{F}$ is a VC-major class, then (6) holds for every stationary ergodic process $\mathbf{X}$ such that $EF(X)$ is finite.*

1.3.2 VC Graph Classes

The graph of a function $f : \mathcal{X} \to \mathbb{R}$ is the set $G_f \subseteq \mathcal{X} \times \mathbb{R}$ defined by $G_f = \{(x, s) : 0 \leq s \leq f(x) \;\text{ or }\; f(x) \leq s \leq 0\}$. A family $\mathcal{F}$ of functions $f : \mathcal{X} \to \mathbb{R}$ is said to be a VC-graph class if

$$\dim_G(\mathcal{F}) = \dim(\{G_f : f \in \mathcal{F}\})$$

is finite. The following result is established in Section 4.2.

Proposition 2. *Let $\mathcal{F}$ be a bounded, countable family of Borel-measurable functions $f : \mathcal{X} \to \mathbb{R}$. If $\mathcal{F}$ is a VC-graph class, then (6) holds for every stationary ergodic process $\mathbf{X}$.*

1.4 Related Work

Steele [22] used sub-additive ergodic theory to establish that both $\Gamma_m(\mathcal{C} : \mathbf{X})$ (see (2) and the entropy $n^{-1} \log S(\{X_1, \ldots, X_n\} : \mathcal{C})$ (see (4)) converge with probability one to non-negative constants whenever $\mathbf{X}$ is ergodic. In addition, he obtained refined necessary and sufficient conditions for uniform strong laws in the i.i.d. case. Nobel [11] showed that the conditions of Theorem A and Talagrand [23] do not characterize uniform convergence for ergodic processes, and in particular, that standard random entropy conditions do not ensure uniform convergence in the ergodic case.

Yukich [28] established rates of convergence for $\Gamma_m(\mathcal{F} : \mathbf{X})$ when $\mathbf{X}$ is ϕ-mixing and $\mathcal{F}$ satisfies suitable bracketing entropy conditions. Yu [27] extends these results to β-mixing (absolutely regular) processes $\mathbf{X}$ and and classes $\mathcal{F}$ satisfying metric entropy conditions. (See Bradley [3] for more on ϕ- and β-mixing conditions.) For VC-Classes $\mathcal{C}$, the results of Yu imply the uniform law (5) when the mixing coefficients β_k decrease as k^{-r} for some

$r > 0$. Work of Peskir and Yukich [18] extends this conclusion to β-mixing processes with $\beta_k = (\log k)^{-2}$.

Nobel and Dembo [12] showed that one may extend uniform strong laws from i.i.d. processes to β-mixing processes with the same one dimensional marginal distribution. Their result implies that (5) holds for any VC-class $\mathcal{C}$ and any β-mixing process $\mathbf{X}$. Peligrad [14] established an analogous result for processes satisfying a modified ϕ-mixing condition. Karandikar and Vidyasagar [9] extended the results of [12] to families of processes, and established rates of convergence depending on the behavior of the mixing coefficients.

Extending earlier work of Hoffmann-Jorgensen [8] in the i.i.d. case, Peskir and Weber [17] show that the uniform ergodic theorem (6) holds if and only if the family $\mathcal{F}$ is, in their terminology, eventually totally bounded in mean. They also note the equivalence of different notions of convergence, as in Steele's work. Peskir [16] investigated conditions for uniform mean square ergodic theorems for families of weak sense stationary processes.

Andrews [1] investigated sufficient conditions under which laws of large numbers can be extended from individual functions to classes of functions, with particular emphasis on stochastically equicontinuous classes indexed by totally bounded parameter spaces. The bibliography of his paper provides a good overview of related work.

1.5 Overview

In the absence of independence or standard uniform mixing conditions, a direct approach to Theorem 1 using symmetrization and exponential type inequalities, or a more indirect approach carried out by coupling with the independent case, does not appear to be possible. Instead, we establish, without reference to independence or mixing conditions, the contrapositive of Theorem 1: if the relative frequencies of sets $C \in \mathcal{C}$ fail to converge uniformly, then for each $L \geq 1$ we can find L points $x_1, \ldots, x_L \in \mathcal{X}$ that are shattered by $\mathcal{C}$, and consequently $\dim(\mathcal{C}) = \infty$. For this we require only the almost sure convergence guaranteed by the ergodic theorem for individual sets. Rather than working directly with the shatter coefficients $S(\cdot : \mathcal{C})$, we consider joins (partitions) generated by finite sub-collections of $\mathcal{C}$, which are defined in Section 2 below.

We begin in the next section with a special case of Theorem 1 in which $\mathcal{X} = [0, 1]$, each X_i is uniformly distributed on $\mathcal{X}$, and each element of $\mathcal{C}$ is equal to a finite union of intervals. This preliminary result, which is the core of the paper, is contained in Proposition 3. The general case of Theorem 1 is established in Section 3 using Proposition 3 and a series of three reductions. The first reduction (contained in Lemma 5) shows that it is enough to consider processes $\mathbf{X}$ for which the marginal distribution of the X_i is non-atomic. The second reduction maps the random variables in $\mathbf{X}$ and the elements of $\mathcal{C}$ to the unit interval with Lebesgue measure via a standard measure space isomorphism. The final reduction (contained in Lemma 6) makes use of an additional measure space isomorphism that maps each element of $\mathcal{C}$ into a set that is equal, up to a set of measure zero, to a finite union of intervals.

2 Classes Containing Finite Unions of Intervals

In this section we establish a version of Theorem 1 in which $\mathcal{X} = [0,1]$ and each element of $\mathcal{C}$ is a finite union of intervals. In the proof, we work with the joins of selected members of $\mathcal{C}$, which act as a surrogate for more commonly used shatter coefficients.

Definition: The join of k sets $A_1, \ldots, A_k \subseteq \mathcal{X}$, denoted $J = \bigvee_{i=1}^k A_i$, is the collection of all *non-empty* intersections $\tilde{A}_1 \cap \cdots \cap \tilde{A}_k$ where $\tilde{A}_i \in \{A_i, A_i^c\}$ for $i = 1, \ldots, k$. Note that J is a partition of $\mathcal{X}$. The join of $A_1, \ldots, A_k$ is said to be full if it has (maximal) cardinality 2^k.

The next Lemma makes an elementary connection between full joins and the VC dimension. A similar result appears in [10] as Lemma 10.3.4. We include a short proof here for completeness.

Lemma 1. *Let $\mathcal{C}$ be any collection of subsets of $\mathcal{X}$. If for some $k \geq 1$ there exists a collection $\mathcal{C}_0 \subseteq \mathcal{C}$ of 2^k sets having a full join, then VC-dim$(\mathcal{C}) \geq k$.*

Proof: Indexing the elements of $\mathcal{C}_0$ in an arbitrary manner by subsets of $[k] := \{1, \ldots, k\}$ we may write $\mathcal{C}_0 = \{C(U) : U \subseteq [k]\}$. For $i = 1, \ldots, k$, let x_i be any element of the intersection

$$\left(\bigcap_{U \subseteq [k], i \in U} C(U) \right) \cap \left(\bigcap_{U \subseteq [k], i \notin U} C(U)^c \right),$$

which is non-empty by assumption. For each subset $V \subseteq [k]$ it is easy to see that $x_i \in C(V)$ if and only if $i \in V$. Thus $\mathcal{C}_0$, and hence $\mathcal{C}$, shatters $\{x_1, \ldots, x_k\}$.

Now let $\mathbf{X} = X_1, X_2, \ldots$ be a stationary ergodic process defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, such that each X_i takes values in $[0,1]$, is Borel-measurable, and has distribution equal to Lebesgue measure $\lambda(\cdot)$ on $[0,1]$.

Proposition 3. *Let $\mathcal{C}_0$ be a countable family of subsets of $[0,1]$, each of whose elements is a finite union of intervals. Suppose that*

$$\limsup_m \Gamma_m(\mathcal{C}_0 : \mathbf{X}) \, > \, 0$$

with positive probability. Then for each integer $L \geq 1$ there exist sets $D_1, D_2, \ldots, D_L \in \mathcal{C}_0$ such that the join $K_L = D_1 \vee D_2 \vee \cdots \vee D_L$ is full, and each cell of K_L has positive Lebesgue measure.

Remarks: It follows from Lemma 1 that the family $\mathcal{C}_0$ in Proposition 3 has infinite VC-dimension. The additional fact that each cell of the joins has positive measure will be needed in the proof of Theorem 1 as we may then ignore sets of measure zero that arise in the application of Lemma 6. The assumption that $C \in \mathcal{C}_0$ is a finite union of intervals guarantees that its boundary has Lebesgue measure zero. Excluding such boundary points from the process $\mathbf{X}$ plays an important role in the final part of the proof of Lemma 3.

Proof: In what follows we will need to examine the difference between the relative frequency and probability of subsets of the unit interval. To this end, for each $\omega \in \Omega$, each $A \subseteq [0,1]$ and each $m \geq 1$ define

$$\Delta^\omega(A : m) \; \triangleq \; \left| \frac{1}{m} \sum_{i=1}^{m} I(X_i(\omega) \in A) - \lambda(A) \right| \tag{7}$$

to be the discrepancy of A with respect to the first m elements of the sample sequence $X_i(\omega)$. Let B^o, $\overline{B}$ and $\partial B = \overline{B} \setminus B^o$ denote, respectively, the interior, closure and boundary of a set $B \subseteq [0,1]$.

For $n \geq 1$ let $\mathcal{D}_n = \{[\, k\, 2^{-n}, (k+1)\, 2^{-n}] : 0 \leq k \leq 2^n - 1\}$ be the set of closed dyadic intervals of order n. Let $\mathcal{D}$ be the union of the families $\mathcal{D}_n$ and let $\mathcal{C} = \mathcal{C}_0 \cup \mathcal{D}$. Then $\mathcal{C}$ and the set $A_0 = \cup_{C \in \mathcal{C}} \partial C$ of all endpoints of elements of $\mathcal{C}$ are countable. In particular $\lambda(A_0) = 0$. By removing a $\mathbb{P}$-null set of outcomes from our sample space we can, and do, assume that $X_i(\omega) \in A_0^c$ for every $i \geq 1$ and every $\omega \in \Omega$.

Recalling the definition (2), we see that $\Gamma_m(\mathcal{C} : \mathbf{X}) \geq \Gamma_m(\mathcal{C}_0 : \mathbf{X})$, and therefore $\limsup_m \Gamma_m(\mathcal{C} : \mathbf{X}) > 0$ with positive probability. In particular, there exists an $\eta > 0$ and a set $E \in \mathcal{F}$ with $\mathbb{P}(E) > 0$ such that

$$\limsup_{m \to \infty} \left[\sup_{C \in \mathcal{C}} \Delta^\omega(C : m) \right] \, > \, \eta \;\; \text{for each } \omega \in E. \tag{8}$$

(Using the results of Steele [22], or alternatively the invariance of E, it follows that that $\mathbb{P}(E) = 1$, but we do not require this stronger result here.) Fix $0 < \delta \leq \min\{\eta/12, \mathbb{P}(E)\}$.

The remainder of the proof proceeds as follows. We first construct a sequence of "splitting sets" $R_1, R_2, \ldots \subseteq [0,1]$, in stages, from the sets in $\mathcal{C}$. At the kth stage the splitting set R_k is obtained from a sequential procedure that makes use of the splitting sets $R_1, \ldots, R_{k-1}$ produced at previous stages. Once obtained, the splitting sets are used to identify, for any $L \geq 1$, a collection of L sets in $\mathcal{C}$ that have full join, and it is easy to show that at most one member of such a collection can come from $\mathcal{D}$. The final step of the proof requires that we keep track of the process by which each splitting set R_k is produced; this requirement is reflected in the notation adopted below.

Construction of R_1. We first choose a sequence of sets $C_1, C_2, \ldots \in \mathcal{C}$ in such a way that a significant fraction of the cells in the join of $C_1, \ldots, C_n$ will intersect both C_{n+1} and its complement. Let C_1 be any set in $\mathcal{C}$. Suppose that $C_1, \ldots, C_n \in \mathcal{C}$ have already been selected, and we wish to choose C_{n+1}. Let $J_n = \mathcal{D}_n \vee C_1 \vee \cdots \vee C_n$ be the join of the previously selected sets and the dyadic intervals of order n. Since the process $\mathbf{X}$ is ergodic and J_n is finite, there exists an integer M and a set F with $\mathbb{P}(F) > 1 - \delta$ such that

$$\Delta^{\omega}(A : m) \;\leq\; \delta\, \lambda(A) \;\text{ for each }\; \omega \in F,\, A \in J_n \text{ and } m \geq M. \tag{9}$$

As $\delta < \mathbb{P}(E)$, the set $E \cap F$ has positive $\mathbb{P}$-measure, and is therefore non-empty. Let ω_{n+1} be any point in $E \cap F$. As $\omega_{n+1} \in E$, it follows from (8) there exists a set $C_{n+1} \in \mathcal{C}$ and an integer $m_{n+1} \geq M$ such that $\Delta^{\omega_{n+1}}(C_{n+1} : m_{n+1}) > \eta$. From C_{n+1}, one may construct the join $J_{n+1} = \mathcal{D}_{n+1} \vee C_1 \vee \cdots \vee C_{n+1}$, and then select C_{n+2} in the same manner as C_{n+1}. Continuing in this fashion, we obtain joins $J_{n+1}, J_{n+2}, \ldots$ and sets $C_{n+2}, C_{n+3}, \ldots \in \mathcal{C}$. We note that the sample points ω_n may vary from step to step, and that there is no requirement that m_{n+1} be greater than m_n.

The choice of the set C_{n+1} ensures that it cannot be well-approximated by a union of elements of J_n, or equivalently, that the collection of cells $A \in J_n$ containing points in C_{n+1} and C_{n+1}^c must have non-vanishing probability. To make this idea precise, define the family

$$H_n \;=\; \left\{ A \in J_n : \Delta^{\omega_{n+1}}(A \cap C_{n+1} : m_{n+1}) \;>\; \frac{\eta}{2}\, \lambda(A) \right\}.$$

The next lemma shows that the elements of $H_n \subseteq J_n$ occupy a non-vanishing fraction of the unit interval.

Lemma 2. *If $G_n = \cup H_n$ is the union of the sets $A \in H_n$, then $\lambda(G_n) \geq \eta/6$.*

Proof: Let $\omega = \omega_{n+1}$, $C = C_{n+1}$, and $m = m_{n+1}$. By decomposing $\Delta^{\omega}(C : m)$ among the elements of J_n, we obtain the following bound:

$$\begin{aligned}
\eta &< \Delta^\omega(C:m) \;\le\; \sum_{A\in J_n} \Delta^\omega(C\cap A:m) \\
&= \sum_{A\in H_n} \Delta^\omega(C\cap A:m) \;+\; \sum_{A\in J_n\setminus H_n} \Delta^\omega(C\cap A:m).
\end{aligned} \tag{10}$$

By definition of H_n, the second term in (10) is at most

$$\sum_{A\in J_n\setminus H_n} \frac{\eta}{2}\,\lambda(A) \;\le\; \frac{\eta}{2}.$$

Moreover, the first term in (10) can be bounded as follows:

$$\begin{aligned}
&\sum_{A\in H_n} \Delta^\omega(C\cap A:m) \\
&\le \sum_{A\in H_n} \frac{1}{m_{n+1}} \sum_{i=1}^m I(X_i(\omega)\in C\cap A) \;+\; \sum_{A\in H_n} \lambda(C\cap A) \\
&\le \sum_{A\in H_n} \frac{1}{m} \sum_{i=1}^m I(X_i(\omega)\in C\cap A) \;+\; \lambda(G_n) \\
&\le \sum_{A\in H_n} \Delta^\omega(A:m) \;+\; 2\,\lambda(G_n) \\
&\le (\delta+2)\lambda(G_n) \;\le\; 3\,\lambda(G_n),
\end{aligned}$$

where the penultimate inequality follows from (9) and the fact that $\omega_{n+1}\in F$. Combining the final expressions in the three preceding displays yields the result.

Let the sets $G_n = \cup H_n$, $n\ge 1$, be derived from the inductive procedure described above. For each $n\ge 1$ define a sub-probability measure $\lambda_n(B) = \lambda(B\cap G_n)$ on $([0,1],\mathcal{B})$. The collection $\{\lambda_n\}$ is necessarily tight, and therefore has a subsequence $\{\lambda_{n_r}\}$ that converges weakly to a sub-probability ν_1 on $([0,1],\mathcal{B})$, in the sense that $\int_0^1 g\,d\lambda_{n_r} \to \int_0^1 g\,d\nu_1$ as $r\to\infty$ for every (bounded) continuous function $g:[0,1]\to\mathbb{R}$. It is easy to see that ν_1 is absolutely continuous with respect to λ, and that

$$\nu_1([0,1]) \;\ge\; \limsup_{r\to\infty} \lambda_{n_r}([0,1]) \;\ge\; \eta/6.$$

In particular, the Radon-Nikodym derivative $d\nu_1/d\lambda$ is well defined, and is bounded above by 1. Define $R_1 = \{x : (d\nu_1/d\lambda)(x) > \delta\}$. From the previous remarks it follows that

$$\begin{aligned}
\frac{\eta}{6} &\le \nu_1([0,1]) \;=\; \int_0^1 \frac{d\nu_1}{d\lambda}\,d\lambda \;=\; \int_{R_1} \frac{d\nu_1}{d\lambda}\,d\lambda + \int_{R_1^c} \frac{d\nu_1}{d\lambda}\,d\lambda \\
&\le \int_{R_1} 1 d\lambda + \int_{R_1^c} \delta\,d\lambda \;\le\; \lambda(R_1)+\delta.
\end{aligned} \tag{11}$$

As $\delta < \eta/12$ by assumption, we conclude that $\lambda(R_1) \geq \eta/12 > 0$.

Construction of R_k for $k \geq 2$. The splitting sets $R_2, R_3, \ldots$ are defined in order, following the general iterative procedure used to construct R_1. The critical difference between the first and subsequent stages is that the sets $R_1, \ldots, R_{k-1}$ produced at stages 1 through $k-1$ are included in the join used at stage k to define R_k. In what follows, let $C_k(n)$, $J_k(n)$, $\omega_k(n)$, $m_k(n)$, $H_k(n)$ and $G_k(n)$ denote the quantities appearing at the nth step of the kth stage. In particular, let $C_1(n) = C_n$, $n \geq 1$, be the elements of $\mathcal{C}$ considered in stage 1, and define $J_1(n)$, $\omega_1(n)$, $m_1(n)$, $H_1(n)$ and $G_1(n)$ in a similar fashion.

Suppose that stages 1 through $k-1$ have been completed, and that we wish to construct the splitting set R_k at stage k. Let $C_k(1)$ be any element of $\mathcal{C}$, and suppose that $C_k(2), \ldots, C_k(n)$ have already been selected. Define the join

$$J_k(n) \ = \ \mathcal{D}_n \vee \bigvee_{j=1}^{k-1} R_j \vee \bigvee_{i=1}^{n} C_k(i). \tag{12}$$

By the ergodic theorem, there exists an integer M and a set F with $\mathbb{P}(F) > 1 - \delta$ such that (9) holds with J_n replaced by $J_k(n)$. As before, it follows from these inequalities and (8) that there exists a sample point $\omega_k(n+1) \in E \cap F$, a set $C_k(n+1) \in \mathcal{C}$, and an integer $m_k(n+1) \geq M$ such that

$$\Delta^{\omega_k(n+1)}(A : m_k(n+1)) \ \leq \ \delta\, \lambda(A) \text{ for each } A \in J_k(n), \tag{13}$$

and simultaneously,

$$\Delta^{\omega_k(n+1)}(C_k(n+1) : m_k(n+1)) \ > \ \eta. \tag{14}$$

Using these quantities, define the family

$$H_k(n) \ = \ \Big\{ A \in J_k(n) : \Delta^{\omega_k(n+1)}(C_k(n+1) \cap A : m_k(n+1)) \ > \ \frac{\eta}{2} \lambda(A) \Big\} \tag{15}$$

and let $G_k(n) = \cup H_k(n)$ be the union of the elements of $H_k(n)$.

Defining $J_k(n+1)$ as in (12) and continuing in the same fashion, we obtain a sequence $C_k(n+2), C_k(n+3), \ldots \in \mathcal{C}$ and a corresponding sequence of sets $G_k(n+1), G_k(n+2), \cdots \subseteq [0,1]$. Lemma 2 ensures that $\lambda(G_k(n)) \geq \eta/6$ for each $n \geq 1$. As before, there is a sequence of integers $n_k(1) < n_k(2) < \cdots$ such that the measures $\lambda(B \cap G_k(n_k(r)))$ converge weakly as $r \to \infty$ to a sub-probability measure ν_k on $([0,1], \mathcal{B})$ that is absolutely continuous with respect to $\lambda(\cdot)$. Define $R_k = \{x : (d\nu_k/d\lambda)(x) > \delta\}$. The argument in (11) shows that $\lambda(R_k) \geq \eta/12$. The arguments below require that we consider density points of the splitting sets. With this in mind, for $k \geq 1$ let

$$\tilde{R}_k \;=\; \left\{ x \in R_k : \lim_{\alpha \to 0} \frac{\lambda((x-\alpha, x+\alpha) \cap R_k)}{2\alpha} \;=\; 1 \right\}$$

be the set of Lebesgue points of R_k. By standard results on differentiation of integrals (c.f. Theorem 31.3 of Billingsley (1995)), $\lambda(\tilde{R}_k) = \lambda(R_k) \geq \eta/12$. The sets $\tilde{R}_k$ are used to construct full joins in the next step of the proof.

Construction of Full Joins. Fix an integer $L \geq 2$. As the measures of the sets $\tilde{R}_k$ are bounded away from zero, there exist positive integers $k_1 < k_2 < \ldots < k_L$ such that $\lambda(\bigcap_{j=1}^{L} \tilde{R}_{k_j}) > 0$. Define the intersections

$$Q_r = \bigcap_{j=1}^{L-r} \tilde{R}_{k_j}$$

for $r = 0, 1, \ldots, L-1$. Note that $Q_0 \subseteq Q_1 \subseteq \cdots \subseteq Q_{L-1}$. Recall that B^o, $\overline{B}$ and ∂B denote, respectively, the interior, closure and boundary of a set $B \subseteq [0,1]$.

Lemma 3. *There exist sets $D_1, D_2, \ldots, D_{L-1} \in \mathcal{C}$ such that for each $l = 1, \ldots, L-1$ the join $K_l = D_1 \vee D_2 \vee \cdots \vee D_l$ satisfies $|K_l| = 2^l$, and for each $B \in K_l$ the intersection $B^o \cap Q_l$ is non-empty. In particular, each cell of K_l has positive Lebesgue measure.*

Proof: We establish the result by induction on l, beginning with the case $l = 1$. In particular, we show that there exists a set $D_1 \in \mathcal{C}$ such that $D_1^o \cap Q_1$ and $(D_1^c)^o \cap Q_1$ are non-empty. To this end, choose $x_1 \in Q_0$, which is non-empty by assumption, and let $\epsilon = \delta/2(\delta+1)$. By definition of the sets $\tilde{R}_{k_j}$, there exists $\alpha_1 > 0$ such that the interval $I_1 \triangleq (x_1 - \alpha_1, x_1 + \alpha_1)$ satisfies

$$\lambda(I_1 \cap Q_0) \;\geq\; (1-\epsilon)\lambda(I_1) \;=\; 2\alpha_1(1-\epsilon). \tag{16}$$

To simplify notation, let $\kappa = k_L$. It follows from the last display and the definition of $R_\kappa \supseteq Q_0$ that

$$\nu_\kappa(I_1 \cap R_\kappa) \;=\; \int_{I_1 \cap R_\kappa} \frac{d\nu_\kappa}{d\lambda}\, d\lambda \;>\; \delta\, \lambda(I_1 \cap R_\kappa) \;\geq\; 2\alpha_1(1-\epsilon)\delta. \tag{17}$$

Now let $\{n_\kappa(r) : r \geq 1\}$ be the subsequence used to define the sub-probability ν_κ. As I_1 is an open set, the portmanteau theorem and (17) imply that

$$\liminf_{r \to \infty} \lambda(I_1 \cap G_\kappa(n_\kappa(r))) \;\geq\; \nu_\kappa(I_1) \;\geq\; \nu_\kappa(I_1 \cap R_\kappa) \;>\; 2\alpha_1(1-\epsilon)\delta.$$

Choose r sufficiently large so that $\lambda(I_1 \cap G_\kappa(n_\kappa(r))) > 2\alpha_1(1-\epsilon)\delta$ and $2^{-n_\kappa(r)} < \delta\,\alpha_1/4$. We require the following subsidiary lemma.

Lemma 4. *There exists a set $A \in H_\kappa(n_\kappa(r))$ such that $A \subseteq I_1$ and $\lambda(A \cap Q_1) > 0$. Moreover, A is contained in Q_1.*

Proof: Let $G = G_\kappa(n_\kappa(r))$. The choice of $n_\kappa(r)$ ensures that

$$\begin{aligned}
(1-\epsilon)\,\delta\,\lambda(I_1) &< \lambda(I_1 \cap G) \\
&= \lambda(I_1 \cap Q_1 \cap G) \;+\; \lambda(I_1 \cap Q_1^c \cap G) \\
&\leq \lambda(I_1 \cap Q_1 \cap G) \;+\; \lambda(I_1 \cap Q_1^c) \\
&\leq \lambda(I_1 \cap Q_1 \cap G) \;+\; \epsilon\lambda(I_1).
\end{aligned}$$

where the first inequality follows from our choice of r, and the final inequality follows from (16) and the fact that $Q_0 \subseteq Q_1$. The last display and the definition of ϵ imply that $\lambda(I_1 \cap Q_1 \cap G) \geq \delta\alpha_1$. As the collection of sets used to define $J_\kappa(n_\kappa(r))$ includes the dyadic intervals of order $n_\kappa(r)$, each element A of the join has diameter (and Lebesgue measure) bounded by $2^{-n_\kappa(r)} < \delta\,\alpha_1/4$. These last two inequalities imply that

$$\delta\,\alpha_1 \;\leq\; \lambda(I_1 \cap Q_1 \cap G) \;\leq\; \sum_A \lambda(Q_1 \cap A) \,+\, 2\,\frac{\delta\,\alpha_1}{4},$$

where the sum is over sets $A \in H_\kappa(n_\kappa(r))$ such that $A \subseteq I_1$. In particular, it is clear that the sum is necessarily positive, and the first part of the claim follows. Note that $A \in H_\kappa(n_\kappa(r))$ implies $A \in J_\kappa(n_\kappa(r))$. Thus the inclusion of the sets $R_1, \ldots, R_{\kappa-1}$ in the join ensures that A is contained in either R_{k_j} or $R_{k_j}^c$, but not both, for each $j = 1, \ldots, L-1$. If $\lambda(A \cap Q_1) > 0$, then necessarily $A \cap Q_1 \neq \emptyset$, and the containment relations imply that $A \subseteq Q_1$. This completes the proof of Lemma 4

Let $D_1 = C_\kappa(n_\kappa(r)+1) \in \mathcal{C}$, where r is the index appearing in Lemma 4. Recall that D_1 is a finite union of intervals, and that no random variables X_i takes values in the finite set ∂D_1. In addition, ∂D_1 has Lebesgue measure zero. Let $A \in H_\kappa(n_\kappa(r))$ be the set identified in Lemma 4, and note that $\lambda(A) > 0$. We argue by contradiction that A (and therefore Q_1) has non-empty intersection with the interiors of D_1 and D_1^c. Suppose first that $A \cap D_1^o = \emptyset$. In this case,

$$\Delta^\omega(A \cap D_1 : m) \;=\; \Delta^\omega(A \cap D_1^o : m) \;=\; 0$$

for every $m \geq 1$ and every $\omega \in \Omega$. However, as $A \in H_\kappa(n_\kappa(r))$ (see (15)) and $\lambda(A) > 0$, we know that $\Delta^\omega(A \cap D_1 : m) > 0$ when $\omega = \omega_\kappa(n_\kappa(r)+1)$ and $m = m_\kappa(n_\kappa(r)+1)$. Thus we arrive at a contradiction.

Suppose now that $(D_1^c)^o \cap A = \emptyset$. In this case $A \subseteq \overline{D}_1$, and with the choice of $\omega = \omega_\kappa(n_\kappa(r)+1)$ and $m = m_\kappa(n_\kappa(r)+1)$ we have

$$\frac{\eta}{2}\,\lambda(A) \;<\; \Delta^\omega(A \cap D_1 : m) \;=\; \Delta^\omega(A \cap \overline{D_1} : m) \;=\; \Delta^\omega(A : m) \;\leq\; \delta\lambda(A).$$

Here the first inequality follows from the fact that $A \in H_\kappa(n_\kappa(r))$, and the second follows from (13). Comparing the first and last terms above, the fact that $\delta \leq \eta/12$ again yields a contradiction. We note that the argument above applies to any set $A \in H_\kappa(n_\kappa(r))$ having positive Lebesgue measure.

Suppose now that we have identified sets $D_1, \ldots, D_l \in \mathcal{C}$, with $l \leq L-2$, such that the join $K_l = D_1 \vee \cdots \vee D_l$, satisfies the conditions of Lemma 3. Let $K_l = \{B_j : j \in [2^l]\}$, and let $x_j \in B_j^o \cap Q_l$ for each $j \in [2^l]$. Select $\alpha_{l+1} > 0$ such that for each j the interval $I_j \triangleq (x_j - \alpha_{l+1}, x_j + \alpha_{l+1})$ is contained in B_j^o and satisfies

$$\lambda(I_j \cap Q_l) \;\geq\; (1-\epsilon)\lambda(I_j) \;=\; 2\alpha_{l+1}(1-\epsilon).$$

Let $\kappa' = k_{L-l}$ and let $\{n_{\kappa'}(l) : l \geq 1\}$ be the subsequence used to define the sub-probability $\nu_{\kappa'}$. For each interval I_j,

$$\liminf_{r\to\infty} \lambda(I_j \cap G_{\kappa'}(n_{\kappa'}(r))) \;\geq\; \nu_{\kappa'}(I_j) \;\geq\; \nu_{\kappa'}(I_j \cap R_{\kappa'}) \;>\; 2\alpha_{l+1}(1-\epsilon)\delta,$$

where the last inequality follows from the previous display, and the fact that $Q_l \subseteq R_{\kappa'}$. Choose r sufficiently large so that $\lambda(I_j \cap G_{\kappa'}(n_{\kappa'}(r))) > 2\alpha_{l+1}(1-\epsilon)\delta$ for each j, and $2^{-n_{\kappa'}(r)} < \delta\,\alpha_{l+1}/4$.

By applying the proof of Lemma 4 to each interval I_j, it is easy to see that there exist sets $A_j \in H_{\kappa'}(n_{\kappa'}(r))$ such that $A_j \subseteq I_j \subseteq B_j^o$, $\lambda(A_j \cap Q_{l+1}) > 0$, and $A_j \subseteq Q_{l+1}$. Let $D_{l+1} = C_{\kappa'}(n_{\kappa'}(r)+1) \in \mathcal{C}$. Arguments identical to the case $l = 1$ above show that for each j the intersections $A_j \cap D_{l+1}^o$ and $A_j \cap (D_{l+1}^c)^o$ are non-empty. This completes the inductive step, and the proof, of Lemma 3.

Given any two dyadic intervals, they are disjoint, intersect at one point, or one contains the other. Therefore, among the sets $D_1, \ldots, D_{L-1}$ of Lemma 3, at most one can be a dyadic interval; the remainder are contained in $\mathcal{C}_0$ and together have a full join whose cells have positive Lebesgue measure. This completes the proof of Proposition 3

3 Reductions and Proof of Theorem 1

As noted in the introduction, Theorem 1 is derived from Proposition 3 via a series of three reductions. Two of these reductions are based on the following lemmas, whose proofs can be found in Appendix A. The third follows from standard results on measure space isomorphisms. In what follows, $A \triangle B = (A \setminus B) \cup (B \setminus A)$ is the standard symmetric difference of two sets.

Lemma 5. *Let $\mathbf{X} = X_1, X_2, \ldots$ be a stationary ergodic process taking values in $(\mathcal{X}, \mathcal{S})$, and let $\mathcal{C} \subseteq \mathcal{S}$ be a countable family of sets such that $\limsup_m \Gamma_m(\mathcal{C} : \mathbf{X}) > 0$ with positive probability. Then $\mathcal{X}$ is necessarily uncountable, and there exists a stationary ergodic process $\tilde{\mathbf{X}} = \tilde{X}_1, \tilde{X}_2, \ldots$ with values in $(\mathcal{X}, \mathcal{S})$ such that $\mathbb{P}(\tilde{X}_i = x) = 0$ for each $x \in \mathcal{X}$ and $\limsup_m \Gamma_m(\mathcal{C} : \tilde{\mathbf{X}}) > 0$ with positive probability.*

Lemma 6. *Let $\mathcal{C} = \{C_1, C_2, \ldots\}$ be a countable collection of Borel subsets of $[0,1]$ such that the maximum diameter of the elements of the join $J_n = \bigvee_{i=1}^n C_i$ tends to zero as $n \to \infty$. Then there exists a Borel-measurable map $\phi : [0,1] \to [0,1]$ and a Borel set $V_1 \subseteq [0,1]$ of measure one such that: (i) ϕ preserves Lebesgue measure and is 1:1 on V_1, (ii) the image $V_2 = \phi(V_1)$ and the inverse map $\phi^{-1} : V_2 \to V_1$ are Borel measurable, (iii) ϕ^{-1} preserves Lebesgue measure, and (iv) for every set $C \in \mathcal{C}$ there is a set $U(C)$, equal to a finite union of intervals, such that $\lambda(\phi(C) \triangle U(C)) = 0$.*

3.1 Proof of Theorem 1

We establish the contrapositive of Theorem 1 via a reduction to Proposition 3. Suppose that $\limsup_m \Gamma_m(\mathcal{C} : \mathbf{X}) > 0$ with positive probability. Let $\mu()$ denote the one-dimensional marginal distribution of $\mathbf{X}$. By Lemma 5, we may restrict our attention to the case in which $\mu(\cdot)$ is non-atomic, and $\mathcal{X}$ is uncountable. It then follows from standard measure space isomorphism results [20] that there exist Borel measurable sets $\mathcal{X}_0 \subseteq \mathcal{X}$ and $I_0 \subseteq [0,1]$ with $\mu(\mathcal{X}_0) = \lambda(I_0) = 1$ and an invertible map $\psi : \mathcal{X}_0 \to I_0$ such that ψ and ψ^{-1} are measurable with respect to the restricted sigma algebras $\mathcal{S} \cap \mathcal{X}_0$ and $\mathcal{B} \cap I_0$, respectively, and $\mu(A) = \lambda(\psi(A))$ for each $A \in \mathcal{S} \cap \mathcal{X}_0$. The event $E = \{X_i \in \mathcal{X}_0^c \text{ for some } i \geq 1\}$ has probability zero, so by removing E from the underlying sample space, we may assume that $X_i(\omega) \in \mathcal{X}_0$ for each sample point ω and each $i \geq 1$.

Define $Y_i = \psi(X_i)$ for $i \geq 1$ and let $\mathcal{C}_1 = \{\psi(C \cap \mathcal{X}_0) : C \in \mathcal{C}\}$ be the (Borel) images in $[0,1]$ of the elements of $\mathcal{C}$. The process $\mathbf{Y} = Y_1, Y_2, \ldots$ is stationary and ergodic with marginal distribution λ. If $C_1 = \psi(C \cap \mathcal{X}_0)$ is an element of $\mathcal{C}_1$, then $\lambda(C_1) = \mu(C \cap \mathcal{X}_0) = \mu(C)$ as $\mu(\mathcal{X}_0) = 1$, and $I(Y_i \in C_1) = I(\psi(X_i) \in \phi(C \cap \mathcal{X}_0)) = I(X_i \in C)$ as $\phi(\cdot)$ is one-to-one. Moreover, if $\mathcal{C}_1$ shatters points $u_1, \ldots, u_k \in [0,1]$, then $\mathcal{C}$ shatters

$\psi^{-1}(u_1), \ldots, \psi^{-1}(u_k)$. It follows that $\Gamma_m(\mathcal{C}_1 : \mathbf{Y}) = \Gamma_m(\mathcal{C} : \mathbf{X})$ with probability one (actually, for every ω) and that $\dim(\mathcal{C}_1) \leq \dim(\mathcal{C})$.

Let $\mathcal{C}_2 = \mathcal{C}_1 \cup \mathcal{D}$, where $\mathcal{D}$ denotes the closed dyadic subintervals of $[0,1]$. Then $\Gamma_m(\mathbf{Y} : \mathcal{C}_2) \geq \Gamma_m(\mathbf{Y} : \mathcal{C}_1)$ and an easy argument shows that $\dim(\mathcal{D}) = 2$. Using Lemma A (c.f. Exercise 4.1 of [5]) one may show that $\dim(\mathcal{C}_2) \leq \dim(\mathcal{C}_1) + \dim(\mathcal{D}) + 1 \leq \dim(C_1) + 3$. As the family $\mathcal{C}_2$ includes $\mathcal{D}$, it satisfies the conditions of Lemma 6 above: let V_1, V_2 and $\phi : [0,1] \to [0,1]$ be the associated sets and point mapping in the lemma. Define $Z_i = \phi(Y_i)$ for $i \geq 1$, and let $\mathcal{C}_3 = \{\phi(C \cap V_1) : C \in \mathcal{C}_2\}$. Arguments like those above show that $\Gamma_m(\mathcal{C}_3 : \mathbf{Z}) = \Gamma_m(\mathcal{C}_2 : \mathbf{Y})$ with probability one, and that $\dim(\mathcal{C}_3) \leq \dim(\mathcal{C}_2)$.

By Lemma 6, for each set $C \in \mathcal{C}_3$ there is a set $U(C)$ that is equal to a finite union of intervals and is such that $\lambda(C \triangle U(C)) = 0$. Let $\mathcal{U} = \{U(C) : C \in C_3\}$. Then $\Gamma_m(\mathcal{U} : \mathbf{Z}) = \Gamma_m(\mathcal{C}_3 : \mathbf{Z})$ with probability one, and it follows from the other relations established above that $\limsup_m \Gamma_m(\mathcal{U} : \mathbf{Z}) > 0$ with positive probability. Fix $L \geq 1$. By Proposition 3 there exist sets $U(C_1), \ldots, U(C_L) \in \mathcal{U}$ such that their join has 2^L cells, and each cell has positive probability. It follows that the join $J_L = C_1 \vee \cdots \vee C_L$ is full as well. As L was arbitrary, Lemma 1 implies that $\mathcal{C}_3$ has infinite VC-dimension, and the same is therefore true of $\mathcal{C}$. This completes the proof of Theorem 1.

4 Proof of VC-Major and VC-Graph Results

4.1 Proof of Proposition 1

Let $\mathbf{X}$ be a stationary ergodic process. Suppose first that $\mathcal{F}$ is bounded, with constant envelope $M < \infty$. Fix $\epsilon > 0$ and select an integer K such that $2M/K \leq \epsilon$. For each $f \in \mathcal{F}$ define the approximation

$$\overline{f}(x) \;=\; M - \frac{2M}{K} \sum_{j=1}^{K} I(f(x) \leq M - 2Mj/K).$$

Note that $\overline{f}(x) - \epsilon \leq f(x) \leq \overline{f}(x)$ for each $x \in \mathcal{X}$ and thus, by an elementary bound,

$$0 \;\leq\; \Gamma_m(\mathcal{F} : \mathbf{X}) \;\leq\; 2\epsilon + \Gamma_m(\overline{\mathcal{F}} : \mathbf{X}),$$

where $\overline{\mathcal{F}} = \{\overline{f} : f \in \mathcal{F}\}$. It follows readily from Theorem 1 and the assumption that $\dim_{VC}(\mathcal{F})$ is finite that $\Gamma_m(\overline{\mathcal{F}} : \mathbf{X}) \to 0$ with probability one as n tends to infinity. As $\epsilon > 0$ was arbitrary, we conclude that $\Gamma_m(\mathcal{F} : \mathbf{X}) \to 0$ with probability one as well.

Suppose now that $\mathcal{F}$ has an envelope F such that $EF(X) < \infty$. Fix $0 < M < \infty$ and for each $f \in \mathcal{F}$ define $f_M(x) = f(x) I(F(x) \leq M)$. Let $\mathcal{F}_M = \{f_M : f \in \mathcal{F}\}$. Then, by an elementary bound and an application of the ergodic theorem to $F(x)I(F(x) \leq M)$,

$$0 \leq \limsup_{m\to\infty} \Gamma_m(\mathcal{F} : \mathbf{X}) \leq \limsup_{m\to\infty} \Gamma_m(\mathcal{F}_M : \mathbf{X}) + 2E[F(X)I(F(X) > M)].$$

A straightforward argument shows that $\mathcal{F}_M$ is a VC-major class and therefore, by the result above, the first term on the right hand side is equal to zero. The second term can be made arbitrarily small by choosing M sufficiently large.

4.2 Proof of Proposition 2

Let $\mathbf{X}$ be a stationary ergodic process with one-dimensional marginal distribution μ. Let $M < \infty$ be an envelope for $\mathcal{F}$. Replacing each $f \in \mathcal{F}$ by $(f + M)/2M$, we may assume without loss of generality that each $f \in \mathcal{F}$ takes values in $[0, 1]$, and therefore

$$G_f = \{(x, s) : x \in \mathcal{X} \text{ and } 0 \leq s \leq f(x) \leq 1\}.$$

Let $Y_1, Y_2, \ldots \in [0, 1]$ be independent, uniformly distributed random variables defined on the same probability space as, and independent of, the process $\mathbf{X}$. For $i \geq 1$ define $Z_i = (X_i, Y_i) \in \mathcal{X} \times [0, 1]$. It follows from standard results in ergodic theory (c.f. [15])) that the process $\mathbf{Z} = Z_1, Z_2, \ldots$ is stationary and ergodic. Let $Z = (X, Y)$ be distributed as Z_1. By an application of Fubini's theorem, for each $f \in \mathcal{F}$,

$$\mathbb{P}(Z \in G_f) = (\mu \otimes \lambda)(G_f) = \int_{\mathcal{X}} \lambda((G_f)_x)\, d\mu(x) = \int_{\mathcal{X}} f(x)\, d\mu(x) = Ef(X), \quad (18)$$

where $G_x = \{s : (x, s) \in G\}$ denotes the x-section of G. Moreover,

$$\frac{1}{m}\sum_{i=1}^m I(Z_i \in G_f) = \frac{1}{m}\sum_{i=1}^m I(Y_i \leq f(X_i)). \quad (19)$$

By an elementary bound, $\Gamma_m(\mathcal{F} : \mathbf{X}) \leq \Gamma_m^1(\mathcal{F} : \mathbf{Z}) + \Gamma_m^2(\mathcal{F} : \mathbf{Z})$ where

$$\Gamma_m^1(\mathcal{F} : \mathbf{Z}) = \sup_{f\in\mathcal{F}} \left| \frac{1}{m}\sum_{i=1}^m I(Y_i \leq f(X_i)) - Ef(X) \right|$$

and

$$\Gamma_m^2(\mathcal{F} : \mathbf{Z}) = \sup_{f\in\mathcal{F}} \left| \frac{1}{m}\sum_{i=1}^m [I(Y_i \leq f(X_i)) - f(X_i)] \right|.$$

It follows from (18) and (19) that

$$\Gamma_m^1(\mathcal{F} : \mathbf{Z}) = \sup_{G\in\mathcal{G}} \left| \frac{1}{m}\sum_{i=1}^m I(Z_i \in G) - \mathbb{P}(Z \in G) \right|,$$

which tends to zero with probability one by Theorem 1 and the assumption that $\mathcal{G}$ is a VC-class. To analyze the second supremum, note that when $X_1 = x_1, \ldots, X_m = x_m$ are fixed,

$$\Gamma_m^2(\mathcal{F} : (x_1, Y_1), \ldots, (x_n, Y_n)) \; = \; \sup_{f \in \mathcal{F}} \left| \frac{1}{m} \sum_{i=1}^{m} [I(Y_i \leq f(x_i)) - \mathbb{P}(Y_i \leq f(x_i))] \right| ,$$

and that $Y_1, \ldots, Y_n$ remain independent under this conditioning. By a routine modification of standard empirical process arguments like those in Theorem 3.1 of Devroye and Lugosi [5], one may establish that

$$E[\,\Gamma_m^2(\mathcal{F}, \mathbf{Z}) \,|\, X_1^n\,] \; \leq \; 2 \left(\frac{\ln 2 S_m(\mathcal{G})}{m} \right)^{1/2} \; \triangleq \; L_m.$$

Here $S_m(\mathcal{G})$ is the (maximal) shatter coefficient of $\mathcal{G}$ defined by

$$S_m(\mathcal{G}) \; = \; \max |\{G \cap \{z_1, \ldots, z_m\} : G \in \mathcal{G}\}| ,$$

where the maximum is taken over all m-sequences $z_1, \ldots, z_m \in \mathcal{X} \times [0, 1]$. As $\mathcal{G}$ has finite VC-dimension, V say, it follows from Sauer's Lemma A above that $S_m(\mathcal{G}) \leq (m + 1)^V$, and consequently, $L_m = O((\ln m/m)^{1/2})$. An straightforward application of McDiarmid's bounded difference inequality (*c.f.* Theorem 2.2 of [5]) shows that for $t > 0$,

$$\mathbb{P}(\Gamma_m^2(\mathcal{F} : \mathbf{Z}) \geq L_m + t \,|\, X_1^n) \; \leq \; e^{-2mt^2}.$$

Taking expectations, the same bound holds for the unconditional probability, and it then follows from a simple application of the first Borel-Cantelli Lemma that $\Gamma_m^2(\mathcal{F} : \mathbf{Z})$ tends to zero with probability one as m tends to infinity.

A Appendix

A.1 Proof of Lemma 5

Following arguments like those in Breiman [4], we may assume without loss of generality that $\mathbf{X} = \{X_i : -\infty < i < \infty\}$ is a two-sided process, and that $\mathbf{X}$ is defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ via a left shift transformation and a projection map. Specifically, Ω is the set of all bi-infinite sequences $\omega = (\omega_i)_{i=-\infty}^{\infty}$ with $\omega_i \in \mathcal{X}$ for each i, and $\mathcal{F} = \otimes_{i=-\infty}^{\infty} \mathcal{S}$ is the usual product sigma field. We may further assume that $X_i(\omega) = X_0(T^i \omega)$, where $X_0 : \Omega \to \mathcal{X}$ is the coordinate projection $X_0(\omega) = \omega_0$ and $T : \Omega \to \Omega$ is the standard left-shift transformation defined by $(T\omega)_i = \omega_{i-1}$. The stationarity of $\mathbf{X}$ implies that T and

T^{-1} preserve $\mathbb{P}(\cdot)$. Ergodicity of $\mathbf{X}$ ensures that T is ergodic: if $TA = A$ then $\mathbb{P}(A) = 0$ or 1.

As noted by Steele [22], the subadditive ergodic theorem implies that the random variables $\Gamma_m(\mathcal{C} : \mathbf{X})$ converge with probability one to a constant. In particular, if $\limsup_m \Gamma_m(\mathcal{C} : \mathbf{X}) > 0$ with positive probability then it follows that

$$\liminf_m \Gamma_m(\mathcal{C} : \mathbf{X}) > 0 \;\text{ with probability one.} \tag{20}$$

This stronger converse of the Glivenko-Cantelli property will be needed in what follows.

Let $A = \{x \in \mathcal{X} : \mu(\{x\}) = 0\}$ contain the non-atomic points of $\mathcal{X}$. If $A^c = \emptyset$ then $\mathcal{X}$ is uncountable, and there is nothing else to prove. Assume then that $A^c \neq \emptyset$. As A^c consists of the (finite or countable) set of points in $\mathcal{X}$ having positive μ-measure, $A \in \mathcal{S}$. Given $\epsilon > 0$, we may express A^c as a disjoint union $A_1 \cup A_2$ such that the cardinality of A_1 is finite and $\mu(A_2) < \epsilon$. Let $\hat{\mu}_m(A) = m^{-1} \sum_{i=1}^m I(X_i \in A)$ denote the empirical measure of $X_1, \ldots, X_m$. By an elementary bound,

$$\Gamma_m(\mathcal{C} : \mathbf{X}) \;\leq\; \Gamma_m(\mathcal{C} \cap A : \mathbf{X}) \;+ \sum_{x \in A_1} |\hat{\mu}_m(\{x\}) \,-\, \mu(\{x\})| \;+\; \hat{\mu}_m(A_2) \,+\, \mu(A_2).$$

As m tends to infinity, the second term above tends to zero, and the last two terms are together less than 2ϵ. As $\epsilon > 0$ was arbitrary, we conclude that $\mu(A) > 0$, so that $\mathcal{X}$ is uncountable. Moreover, (20) implies that $\liminf_m \Gamma_m(\mathcal{C} \cap A : \mathbf{X}) > 0$ with probability one.

Let Ω_A denote the set of $\omega \in \Omega$ such that $\omega_0 \in A$ and both index sets $\{i \geq 1 : w_i \in A\}$ and $\{i \leq -1 : w_i \in A\}$ are infinite. By the ergodic theorem, $\mathbb{P}(\Omega_A) = \mu(A) > 0$. For $\omega \in \Omega_A$, define $\tau(\omega) = \min\{k \geq 1 : T^k\omega \in A\}$ (which is finite) and the induced transformation $\tilde{T} : \Omega_A \to \Omega_A$ by $\tilde{T}\omega = T^{\tau(\omega)}\omega$. Routine arguments from ergodic theory [15] show that $\tilde{T}$ is invertible, that $\tilde{T}$ is measurable on the restricted sigma field $\mathcal{F}_A = \mathcal{F} \cap \Omega_A$, and that $\tilde{T}$ preserves the normalized measure $\mathbb{P}_A(\cdot) = \mathbb{P}(\cdot)/\mathbb{P}(\Omega_A)$ on $(\Omega_A, \mathcal{F}_A)$ and is ergodic. For the sake of completeness, we provide a sketch of the proofs using a geometric argument from ergodic theory known as the Kakutani skyscraper. For each positive integer k, define $A_k = \{\omega \in \Omega_A : \tau(\omega) = k\}$. Then the sets $A_1, A_2, \ldots$ partition Ω_A. Moreover, $\cup_{k=1}^{\infty} \cup_{i=0}^{k-1} T^i A_k$ is a disjoint union containing almost every point in Ω. The Kakutani skyscraper of Ω_A is created by stacking the sets $T^1 A_k, \ldots, T^{k-1} A_k$ above A_k for each $k \geq 1$.

The measurability of $\tilde{T}$ follows from the fact that each A_k is measurable and that $\tilde{T}$ restricted to A_k equals T^k restricted to A_k. Invertibility of $\tilde{T}$ follows directly from the invertibility of T and the construction of the Kakutani skyscraper. In particular, let $\omega_1 \neq \omega_2$ be points in Ω_A. Then $\tilde{T}(\omega_1) = T(T^{\tau(\omega_1)-1}\omega_1)$ and $\tilde{T}(\omega_2) = T(T^{\tau(\omega_2)-1}\omega_2)$. As T is invertible, and $T^{\tau(\omega_1)-1}(\omega_1)$ and $T^{\tau(\omega_2)-1}(\omega_2)$ are distinct points in the Kakutani skyscraper, it follows that $\tilde{T}(\omega_1) \neq \tilde{T}(\omega_2)$. The measure preserving property of $\tilde{T}$ follows from the fact that T is measure preserving on each of the sets A_k. To establish ergodicity, suppose that $B \subset \Omega_A$ is a set of positive measure that is invariant for $\tilde{T}$. The set $C = \cup_{i=-\infty}^{\infty} T^i B$ is invariant under T, and $C \cap A = B$ since B is invariant for $\tilde{T}$. As T is ergodic, C contains A and therefore $A = B$. It follows that $\tilde{T}$ is ergodic.

Define $\tilde{X}_0 : \Omega_A \to \mathcal{X}$ by $\tilde{X}_0(\omega) = \omega_0$ and $\tilde{X}_i(\omega) = \tilde{X}_0(\tilde{T}^i \omega)$ for $-\infty < i < \infty$. Then the process $\tilde{\mathbf{X}} = \{\tilde{X}_i\}$ defined on $(\Omega_A, \mathcal{F}_A, \mathbb{P}_A)$ is stationary and ergodic, takes values in $(\mathcal{X}, \mathcal{S})$, and has marginal distribution $\mu_A(\cdot) = \mu(\cdot)/\mu(A)$ with no point masses.

We wish to show that $\limsup_m \Gamma_m(\mathcal{C} : \tilde{\mathbf{X}}) > 0$ with positive $\mathbb{P}_A$-probability. To this end, for each $\omega \in \Omega_A$, define $\tau_0(\omega) = 0$, $\tau_1(\omega) = \tau(\omega)$, and $\tau_{l+1}(\omega) = \min\{k > \tau_l(\omega) : \omega_k \in A\}$. By definition of Ω_A, each function τ_l is finite. For each $m \geq 1$, $C \in \mathcal{C}$ and $\omega \in \Omega_A$,

$$
\begin{aligned}
\frac{1}{m} \sum_{i=0}^{m-1} I(\tilde{X}_i(\omega) \in C) &= \frac{1}{m} \sum_{j=0}^{\tau_{m-1}(\omega)} I(X_j(\omega) \in C \cap A) \\
&= \frac{1}{\mu(A)} W_m(\omega) \frac{1}{\tau_{m-1}(\omega)} \sum_{j=0}^{\tau_{m-1}(\omega)} I(X_j(\omega) \in C \cap A), \qquad (21)
\end{aligned}
$$

where we have defined $W_m = \mu(A)\, \tau_{m-1}/m$. By the ergodic theorem, for $\mathbb{P}_A$-almost every $\omega \in \Omega_A$,

$$
\frac{m}{\tau_{m-1}(\omega)} = \frac{1}{\tau_{m-1}(\omega)} \sum_{j=0}^{\tau_{m-1}(\omega)} I(X_j(\omega) \in C \cap A) \ \to \ \mu(A)
$$

as m tends to infinity, and therefore $W_m \to 1$ with $\mathbb{P}_A$ probability one. Omitting the dependence on ω, it follows from (21) and the definition of $\mu_A(\cdot)$ that

$$
\begin{aligned}
\Gamma_m(\mathcal{C} : \tilde{\mathbf{X}}) &= \sup_{C \in \mathcal{C}} \left| \frac{1}{m} \sum_{i=0}^{m-1} I(\tilde{X}_i \in C) \ - \ \mu_A(C) \right| \\
&= \frac{1}{\mu(A)} \sup_{C \in \mathcal{C}} \left| W_m \frac{1}{\tau_{m-1}} \sum_{j=0}^{\tau_{m-1}} I(X_j \in C \cap A) \ - \ \mu(C \cap A) \right| \\
&\geq \frac{1}{\mu(A)} \Gamma_{\tau_{m-1}}(\mathcal{C} \cap A : \mathbf{X}) \ - \ |W_m - 1| \sup_{C \in \mathcal{C}} \left| \frac{1}{\tau_{m-1}} \sum_{j=0}^{\tau_{m-1}} I(X_j \in C \cap A) \right| \\
&\geq \frac{1}{\mu(A)} \Gamma_{\tau_{m-1}}(\mathcal{C} \cap A : \mathbf{X}) \ - \ |W_m - 1|.
\end{aligned}
$$

The first inequality above follows by writing $W_m = 1 + (W_m - 1)$, and then using the elementary bound $\sup_\alpha |a_\alpha - b_\alpha| \geq \sup_\alpha |a_\alpha| - \sup_\alpha |b_\alpha|$. It follows from the last display that

$$\begin{aligned}
\limsup_m \Gamma_m(\mathcal{C} : \tilde{\mathbf{X}}) &\geq \liminf_m \Gamma_m(\mathcal{C} : \tilde{\mathbf{X}}) \\
&\geq \frac{1}{\mu(A)} \liminf_m \Gamma_{\tau_{m-1}}(\mathcal{C} \cap A : \mathbf{X}) \\
&\geq \frac{1}{\mu(A)} \liminf_m \Gamma_m(\mathcal{C} \cap A : \mathbf{X})
\end{aligned}$$

and the argument above shows that the final term is positive with $\mathbb{P}_A$-probability one. This completes the proof.

A.2 Proof of Lemma 6

The isomorphism ϕ is defined as a limit of isomorphisms ϕ_n. The maps ϕ_n are defined inductively. To begin, let

$$\phi_1(x) = \begin{cases} \lambda([0,x] \cap C_1) & \text{if } x \in C_1 \\ \lambda(C_1) + \lambda([0,x] \cap C_1^c) & \text{if } x \in C_1^c \end{cases}$$

Then ϕ_1 maps C_1 into $[0, \lambda(C_1)]$ and C_1^c into $[\lambda(C_1), 1]$. By standard arguments, ϕ_1 is Lebesgue measure preserving, and a bijection almost everywhere.

Suppose now that maps $\phi_1, \ldots, \phi_n$ have been defined in such a way that (i) for each element A of the join $J_n = \bigvee_{i=1}^n C_i$ and each $x \in A$, $\phi_n(x) = \beta_n(A) + \lambda([0,x] \cap A)$, where $\beta_n(A)$ is a constant, and (ii) the intervals $\{[\beta_n(A), \beta_n(A) + \lambda(A)) : A \in J_n\}$ form a disjoint covering of $[0,1)$. For each each $A \in J_n$ and each $x \in A$ define

$$\phi_{n+1}(x) = \begin{cases} \beta_n(A) + \lambda([0,x] \cap A \cap C_{n+1}) & \text{if } x \in A \cap C_{n+1} \\ \beta_n(A) + \lambda(A \cap C_{n+1}) + \lambda([0,x] \cap A \cap C_{n+1}^c) & \text{if } x \in A \cap C_{n+1}^c. \end{cases}$$

With these definitions, properties (i) and (ii) hold for J_{n+1} and ϕ_{n+1}. Moreover, $\phi_1, \phi_2, \ldots$ have the property that for each n, each cell $A \in J_n$, and each $m \geq n$, the function ϕ_m is a Lebesgue measure preserving almost everywhere bijection from A into $[\beta_n(A), \beta_n(A) + \lambda(A)]$. In particular, for each $A \in J_n$ and each $m \geq n$,

$$\mathrm{cl}(\phi_n(A)) \;=\; \mathrm{cl}(\phi_m(A)) \;=\; [\beta_n(A), \beta_n(A) + \lambda(A)],$$

where $\mathrm{cl}(U)$ denotes the closure of U.

Fix $x \in [0,1]$ for the moment, and for $n \geq 1$, let $A_n(x)$ be the cell of J_n containing x. Note that the sequence $\phi_n(x), \phi_{n+1}(x), \ldots$ is contained in the interval $\mathrm{cl}(\phi_n(A_n(x)))$, whose diameter is equal to $\lambda(A_n(x)) \leq \mathrm{diam}(A_n(x))$. By assumption, the latter quantity tends to zero as $n \to \infty$, and therefore $\{\phi_n(x) : n \geq 1\}$ is a Cauchy sequence. Let $\phi(x)$ denote its limit. Then $\phi(\cdot)$ is a limit of measurable functions, hence measurable.

We claim that $\mathrm{cl}(\phi(A)) = \mathrm{cl}(\phi_n(A))$ for every $n \geq 1$ and every $A \in J_n$. To see this, fix $A \in J_n$. If $y \in \mathrm{cl}(\phi(A))$ then there exists $x_1, x_2, \ldots \in A$ such that $\phi(x_m) \to y$. By definition of $\phi(\cdot)$, there exist integers $r_1, r_2, \ldots$ tending to infinity such that $\phi_{r_m}(x_m) \to y$. As each value $\phi_{r_m}(x_m) \in \phi_{r_m}(A) \subseteq \mathrm{cl}(\phi_n(A))$, we have $y \in \mathrm{cl}(\phi_n(A))$. Thus $\mathrm{cl}(\phi(A)) \subseteq \mathrm{cl}(\phi_n(A))$, the latter set being equal to the interval $I_A = [\beta_n(A), \beta_n(A) + \lambda(A)]$. To establish the converse, let $y_0 \in I_A^o$ and $\epsilon > 0$ be such that $(y_0 - \epsilon, y_0 + \epsilon) \subseteq I_A$. By the shrinking diameter assumption on the joins J_m, there exists an integer m and a cell $A' \in J_m$ such that $A' \subseteq A$ and $\mathrm{cl}(\phi_m(A')) \subseteq I_0$ has positive measure. Thus if $x \in A'$ then $\phi_r(x) \in \mathrm{cl}(\phi_m(A'))$ for $r \geq m$, and therefore $\phi(x) \in I_0$. As $\epsilon > 0$ was arbitrary, it follows that $I_A^o \subseteq \mathrm{cl}(\phi(A))$, and consequently $I_A \subseteq \mathrm{cl}(\phi(A))$ as well.

We now establish that the map ϕ preserves Lebesgue measure. To this end, for each $n \geq 1$ define

$$\mathcal{Q}_n \;=\; \{\mathrm{cl}(\phi(A)) : A \in J_n\} \,\cup\, \{\{\beta_n(A)\} : A \in J_n\} \cup \{\{1\}\}$$

to be the collection of intervals into which the elements of J_n are mapped, and the endpoints of these intervals. We wish to show that $\lambda(\phi^{-1}B) = \lambda(B)$ for each $B \in \mathcal{Q}_n$. Suppose first that α is the endpoint of some interval $\mathrm{cl}(\phi(A'))$ with $A' \in J_n$. Fix $\epsilon > 0$ and let $m \geq n$ be large enough that $\max\{\lambda(A) : A \in J_m\} \leq \epsilon/2$. Let $A_1, \ldots, A_r$ be those elements of J_m such that $\mathrm{cl}(\phi_m(A_j))$ contains the point α. Then $\phi^{-1}\{\alpha\} \subseteq \cup_{j=1}^r A_j$ and at most two of the sets A_j can have positive measure. It follows that $\lambda(\phi^{-1}\{\alpha\}) \leq \epsilon$, and as $\epsilon > 0$ was arbitrary we have $\lambda(\phi^{-1}\{\alpha\}) = 0$. Suppose now that $B \in \mathcal{Q}_n$ is of the form $B = \mathrm{cl}(\phi(A)) = [\alpha_1, \alpha_2]$ for some element $A \in J_n$. Then $\phi^{-1}B = A \cup \phi^{-1}\{\alpha_1\} \cup \phi^{-1}\{\alpha_2\}$, and therefore

$$\lambda(\phi^{-1}B) \;=\; \lambda(A) \;=\; \lambda(\mathrm{cl}(\phi_n(A))) \;=\; \lambda(\mathrm{cl}(\phi(A))) \;=\; \lambda(B).$$

It follows from these arguments that $\lambda(\phi^{-1}B) = \lambda(B)$ for each $B \in \cup_{m \geq 1}\mathcal{Q}_m$. As the latter collection generates the Borel sigma field of $[0,1]$ and is closed under intersections, ϕ preserves Lebesgue measure.

Next, we show that ϕ is 1:1 on a Borel subset of $[0,1]$ with full measure. Let $\mathcal{Q}^0 = \bigcup_{m=1}^{\infty}\{\beta_m(A) : A \in J_m\} \cup \{\{1\}\}$ be the (countable) set of endpoints of the intervals $\{\mathrm{cl}(\phi(A)) : A \in J_m, m \geq 1\}$. Since ϕ^{-1} preserves Lebesgue measure, $\lambda(\phi^{-1}\mathcal{Q}^0) = 0$. Define $V_1 = [0,1] \setminus \phi^{-1}\mathcal{Q}^0$, so that $\lambda(V_1) = 1$. Let x_1 and x_2 be distinct points in V_1. Since the diameters of the elements of J_n tend to zero, there exists an n such that x_1 and x_2 are contained in different elements of J_n. Thus, ϕ_n maps x_1 and x_2 to distinct intervals, which may intersect only at their endpoints. Hence, ϕ also maps x_1 and x_2 to distinct intervals. Since V_1 excludes points that map to endpoints of these intervals, $\phi(x_1) \neq \phi(x_2)$. Therefore, ϕ is a bijection on V_1, and we have established conclusion (i) of the Lemma.

Conclusion (ii) of the Lemma follows from (i) and general results concerning measurable maps of complete separable metric spaces, see Corollary 3.3 of Parthasarathy [13]. To establish (iii), note that for any measurable subset $A \subseteq V_1$, $\lambda(\phi(A)) = \lambda(\phi^{-1}(\phi(A))) = \lambda(A)$, since ϕ is measure preserving and 1:1 on V_1.

To establish conclusion (iv), let $C \in \mathcal{C}$. Then there exist positive integers k and n such that $C = \bigcup_{i=1}^k A_i$ where $A_1, A_2, \ldots, A_k$ are (disjoint) cells in J_n. Let $U(C) = \bigcup_{i=1}^k [\beta_n(A_i), \beta_n(A_i) + \lambda(A_i)]$. Then $\phi(C) = \bigcup_{i=1}^k \phi(A_i) \subseteq \bigcup_{i=1}^k \mathrm{cl}(\phi(A_i)) = U(C)$ and $\lambda(\phi(C)) = \sum_{i=1}^k \lambda(A_i) = \lambda(U(C))$. Thus, $\lambda(\phi(C) \bigtriangleup U(C)) = \lambda(U(C) \setminus \phi(C)) = 0$.

Remark: The condition that the cells of the joins have diminishing diameters, rather than measures tending to zero, is necessary. If, for example, $C_n = \bigcup_{i=0}^{2^n-1} [\frac{2i}{2^{n+1}}, \frac{2i+1}{2^{n+1}})$ for positive integers n, then the limiting map is $\phi(x) = 2x \bmod 1$.

References

[1] Andrews, D.K. (1987) Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers. *Econometrica* **55:6** 1465–1471. MR0923471 (90b:62146)

[2] Billingsley, P. (1995). *Probability and Measure*, 3rd ed., Wiley, New York. MR1324786 (95k:60001)

[3] Bradley, R.C. (1986) Basic properties of strong mixing conditions. *Dependence in Probability and Statistics: A Survey of Recent Results* 165–192 Birkhäuser, Boston, MA.

[4] Breiman, L. (1992) Probability (Classics in Applied Mathematics). *SIAM Series* **7** SIAM, Philadelphia, PA. MR1163370 (93d:60001)

[5] Devroye, L. and Lugosi, G. (2001) *Combinatorial Methods in Density Estimation* Springer. MR1843146 (2002h:62002)

[6] Dudley, R.M. (1999) *Uniform Central Limit Theorems* Cambridge Univ. Press, New York.

[7] Halmos, P. (1964) *Measure Theory*, 9th ed., D. Van Nostrand Company, New York. MR0033869 (11,504d)

[8] Hoffmann-Jørgensen, J. (1984) Necessary and sufficient condition for the uniform law of large numbers. *Probability in Banach Spaces, V* Lecture Notes in Math., **1153** 258–272, Springer, Berlin. MR0821985 (87j:60019)

[9] Karandikar, R.L. and Vidyasagar, M. (2002) Rates of uniform convergence of empirical means with mixing processes. *Statistics & Probability Letters* **58:3** 297–307. MR1921877 (2003e:60067)

[10] Matousek, J. (2002) Lectures on Discrete Geometry. *Graduate Texts in Mathematics* **212** Springer, New York. MR1899299 (2003f:52011)

[11] Nobel, A.B. (1995) A counterexample concerning uniform ergodic theorems for a class of functions. *Statistics & Probability Letters* **24** 165–168. MR1354371 (97c:60083)

[12] Nobel, A.B. and Dembo, A. (1993) A note on uniform laws of averages for dependent processes. *Statistics & Probability Letters* **17** 169–172. MR1229933 (94e:60031)

[13] Parthasarathy, K.R. (1967) *Probability Measures on Metric Spaces* Academic Press, New York. MR0226684 (37 #2271)

[14] Peligrad, M. (2001) A Note on the Uniform Laws for Dependent Processes Via Coupling. *Journal of Theoretical Probability* **14:4** 979–988. MR1860084 (2002h:60057)

[15] Petersen, K. (1983) *Ergodic Theory* Cambridge University Press, Cambridge, MA. MR0833286 (87i:28002)

[16] Peškir, G. (1998) The uniform mean-square ergodic theorem for wide sense stationary processes. *Stochastic Anal. Appl.* **16:4** 697–720. MR1632570 (2000e:60053)

[17] Peškir, G. and Weber, M. (1993) Necessary and sufficient conditions for the uniform law of large numbers in the stationary case. *Functional Analysis, IV (Dubrovnik)* 165–190, Institute of Mathematics, University of Aarhus, Denmark. MR1326386 (96b:60015)

[18] Peškir, G. and Yukich, J.E. (1994) Uniform ergodic theorems for dynamical systems under VC entropy conditions. *Probability in Banach spaces, 9 (Sandjberg, 1993)*, 105–128, Birkhäuser, Boston, MA. MR1308513 (96d:28020)

[19] Pollard, D. (1984) *Convergence of Stochastic Processes* Springer, New York. MR0762984 (86i:60074)

[20] Royden, H.L. (1988) *Real Analysis*, 3rd ed. Macmillan Publishing Company, New York. MR1013117 (90g:00004)

[21] Sauer, N. (1972) On the density of families of sets. *J. Comb. Theory (A)* **13** 145–147. MR0307902 (46 #7017)

[22] Steele, J.M. (1978) Empirical discrepancies and subadditive processes. *Ann. Probab.* **6** 118–127. MR0464379 (57 #4310)

[23] Talagrand, M. (1987) The Glivenko-Cantelli problem. *Ann. Probab.* **15** 837–870. MR0893902 (88h:60012)

[24] van der Vaart, A.W. and Wellner, J.A. (1996) *Weak Convergence and Empirical Processes*. Springer, New York.

[25] Vapnik, V.N. and Chervonenkis, A.Ya. (1971) On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Appl.* **16** 264–280. MR0627861 (83d:60031)

[26] Vapnik, V.N. and Chervonenkis, A.Ya. (1981) Necessary and sufficient conditions for the uniform convergence of means to their expectations. *Theory Probab. Appl.* **26** 532–553.

[27] Yu, B. (1994) Rates of convergence for empirical processes of stationary mixing sequences. *Ann. Probab.* **22** 94–116. MR1258867 (94m:60056)

[28] Yukich, J.E. (1986) Rates of convergence for classes of functions: the non-i.i.d. case. *J. Multivariate Anal.* **20** 175–189. MR0866068 (88f:60049)