First-Order Predictive Sequences and Induced Transformations

Andrew B. Nobel *

November 29, 1999

Abstract

General conditions are given under which an individual sequence $\mathbf{x} = x_1, x_2, \ldots$ taking values in a complete separable metric space X will induce a measure preserving transformation $T : X \to X$. The results here generalize earlier work on D-sequences, and are established by different methods, based on some elementary facts concerning sequences with limiting relative frequencies of finite order.

Department of Statistics Technical Report #2367
University of North Carolina, Chapel Hill

*Andrew Nobel is with the Department of Statistics, University of North Carolina, Chapel Hill, NC 27599. Email: nobel@stat.unc.edu. His research was supported in part by NSF Grant DMS-9501926.

1 Introduction

Motivated by recent work on chaos and non linear dynamics, there has been increasing interest in statistical inference from deterministic systems that exhibit random behavior. Repeated measurements of a given system made at discrete time instants can be represented by a sequence

$$\mathbf{x} = x_1, x_2, \ldots \qquad x_i \in X \tag{1}$$

where x_i is the value of the i'th measurement, and X is the state space of the system. Of interest here is the study of $\mathbf{x}$ as an *individual sequence*, apart from any probabilistic ensemble from which it may have been generated. The primary goal of the paper is to give general conditions under which $\mathbf{x}$ may be called deterministic.

The discrete time evolution of a deterministic system is typically modeled by a map $T : X \to X$ that acts on the state space of the system, and that preserves a measure μ on X. The evolution of the system beginning at some initial state $x \in X$ is then given by the forward trajectory

$$\mathbf{x} = x, Tx, T^2x, \ldots \tag{2}$$

of T starting at x. Though it is common in statistical analyses to assume that measurements of the system are subject to observational or dynamical noise, estimates of T, μ, and other quantities of interest can be obtained directly from $\mathbf{x}$. Observation noise can, in some situations, be removed by judicious averaging (see Lalley (1998)).

Suppose that $\mathbf{x}$ is a fixed sequence of noisy or noiseless measurements of the form (1), and assume to avoid trivialities that no two elements of $\mathbf{x}$ are the same. In light of (2) one might define $\mathbf{x}$ to be deterministic if it arises as the trajectory of some measure preserving transformation $T : X \to X$. This is sensible if T is assumed to be continuous, but leaves open the question of which measure, if any, T should preserve. More problematic is the fact that any measurable transformation $T : X \to X$ can be modified on the countable set $\{x_1, x_2, \ldots\}$ in order to ensure that $x_{i+1} = Tx_i$ for $i \geq 1$. If T preserves a non-atomic measure μ, then the modification is negligible. Thus if this proposed definition is to be meaningful, some constraints must be placed on T.

As it happens, it is better to turn the question around: rather than ask if $\mathbf{x}$ is determined by a measure preserving transformation, ask instead whether there is a measure preserving transformation determined by $\mathbf{x}$. The latter question, which is the principal subject of the paper, is meaningful in very general settings and leads naturally to joint specification of the transformation and the measure it preserves. The question was first considered by

Maharam (1965) who defined and studied determining (D) sequences taking values the unit interval.

1.1 D-sequences

The study of D-sequences is based on shift extension sets and continuous restrictions of the next-element function. Recall that the lower density of any set $M \subseteq \mathcal{N} = \{1, 2, \ldots\}$ is defined by

$$d_*(M) = \liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I\{i \in M\}.$$

Let $\mathbf{x} = x_1, x_2, \ldots$ be a sequence of numbers $x_i \in [0, 1]$, no two of which are the same. Define $\mathbf{x}(M) = \{x_i : i \in M\}$. A subset $M \subseteq \mathcal{N}$ is said to be a *shift extension* set if there exists a homeomorphism ϕ from $\overline{\mathbf{x}(M)}$ onto a subset of $[0, 1]$ such that

$$\phi(x) = x_{n+1} \;\text{ if and only if }\; x = x_n\,.$$

Let λ denote Lebesgue measure on $[0, 1]$. Then $\mathbf{x}$ is said to be a D-sequence if

(D1) For every $\epsilon > 0$ there is a $\delta > 0$ such that $d_*(M) > 1 - \delta$ implies $\lambda(\overline{\mathbf{x}(M)}) > 1 - \epsilon$.

(D2) There exist shift extension sets $M_1 \subseteq M_2 \subseteq \cdots$ such that $d_*(M_r) \to 1$ as $r \to \infty$.

Maharam (1965, Theorems 1 and 2) established the following result.

Theorem A *To every D-sequence $\mathbf{x}$ there corresponds an almost everywhere invertible Borel measurable transformation $T : [0, 1] \to [0, 1]$ such that*

(a) $T(x_n) = x_{n+1}$

(b) For each $\epsilon > 0$ there exists $M \subseteq \mathcal{N}$ with $d_*(M) > 1 - \epsilon$ such that if $x_{m_k} \to x$ with $m_k \in M$ then $x_{m_k+1} \to T(x)$.

If T' is any other transformation satisfying (b) then $\lambda\{T = T'\} = 1$. Moreover, if

$$\frac{1}{n} \sum_{i=1}^{n} I\{x_i \in [a, b)\} \;\to\; \lambda([a, b)) \tag{3}$$

for every $[a, b) \subseteq [0, 1)$ then $\lambda(T^{-1}A) = \lambda(A)$ for each Borel set $A \subseteq [0, 1]$.

Maharam also defined the notion of an E-sequence. She showed that every E-sequence determines an ergodic Lebesgue-measure preserving transformation of $[0, 1]$, and that almost every sample sequence of such a transformation is an E sequence. Bick (1967) extended

Maharam's results to sequences taking values in $\mathbb{R}$ that are relatively uniformly distributed with respect to Lebesgue measure. His arguments readily generalize to sequences with values in $\mathbb{R}^l$. Kappos and Papadopoulou (1967) further extended Maharam's results to sequences taking values in a locally compact Hausdorff space, of possibly infinite measure, with a countable base.

Sun (1995) showed that there is an almost everywhere continuous isomorphism between any complete separable metric space X with a non-atomic measure μ and the unit interval with Lebesgue measure. He defined D-sequences in X and showed that any such sequence corresponds, via the isomorphism, to a D-sequence in $[0, 1]$. In particular if $\mathbf{x}$ is a D-sequence in X such that $n^{-1} \sum_{i=1}^{n} g(x_i) \to \int g d\mu$ for every function $g : X \to \mathbb{R}$ that is continuous at μ-almost every $x \in X$, then there is a corresponding μ-preserving transformation $T : X \to X$ satisfying conditions (a) and (b) above.

A sequence $\mathbf{x}$ is said to be strongly uniformly distributed (s.u.d.) if for each pair of closed intervals $U, V \subseteq [0, 1]$ and each $k \geq 1$, the limit

$$R_k(U, V) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I\{x_i \in U\} I\{x_{i+k} \in V\}$$

exists and is such that $\lim_{m\to\infty} m^{-1} \sum_{k=1}^{m} R_k(U, V) = \mu(U)\mu(V)$. Coffey (1989, Theorem 3.1) showed that if $\mathbf{x}$ is a uniformly distributed D-sequence in $[0, 1]$ then its associated transformation T is ergodic if and only if $\mathbf{x}$ is strongly uniformly distributed. Sun (1995) extends this result to sequences taking values in a complete separable metric space. Bick and Coffey (1991) give an explicit construction for a class of s.u.d. D-sequences whose associated transformations have entropy zero and are not weakly mixing.

1.2 Predictive Sequences

The results of this paper generalize those of Maharam and others on D-sequences, and are established by different methods. They are based on some elementary facts about sequences with limiting relative frequencies of finite order. One possible definition of a deterministic sequence is in terms of the predictive property described below. Predictive sequences include D-sequences, the trajectories of ergodic measure preserving transformations, and more general sequences that cannot be represented in the form (2).

Let (X, d) be a complete, separable metric space with Borel sigma field $\mathcal{B}$. Let X^k have the metric $d_k(w, w') = \sum_{i=1}^{k} d(w_i, w'_i)$ and Borel sigma field $\mathcal{B}_k$. Let $\mathcal{C}_b(X^k)$ be the collection of all bounded, continuous functions $g : X^k \to \mathbb{R}$ equipped with the supremum norm $||g|| = \sup_{y\in X^k} |g(y)|$.

A Borel measurable transformation $T : X \to X$ is said to preserve a measure μ on $(X, \mathcal{B})$ if $\mu(T^{-1}A) = \mu(A)$ for each $A \in \mathcal{B}$, and in this case T is said to be a measure preserving transformation of $(X, \mathcal{B}, \mu)$. Henceforth all transformations are assumed to be Borel measurable, and all measures are assumed to be finite. A measure preserving transformation T of $(X, \mathcal{B}, \mu)$ is called ergodic if $A = T^{-1}A$ implies $\mu(A) = 0$ or 1, *i.e.* any set left fixed by the action of T has measure zero or one.

To decide whether a given sequence $\mathbf{x}$ determines a transformation it is useful to look for a functional relationship between its successive components. Roughly speaking, a D-sequence is one for which the next element operation $x_i \to x_{i+1}$ is, for arbitrarily large sets of indices i, compatible with a homeomorphism. For the purposes of establishing that $\mathbf{x}$ determines a transformation the following weaker conditions will suffice.

Definition: A sequence $\mathbf{x} = x_1, x_2, \ldots$ with $x_i \in X$ will be called *first order predictive* if

$$\Lambda(g) \;=\; \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g(x_i) \tag{4}$$

exists for every $g \in \mathcal{C}_b(X)$, and for every $\epsilon > 0$ there is a compact set K and a continuous function $h : K \to X$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I\{x_i \notin K\} \;\leq\; \epsilon. \tag{5}$$

and

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I\{x_i \in K \text{ and } d(h(x_i), x_{i+1}) \geq \epsilon\} \;\leq\; \epsilon. \tag{6}$$

Condition (4) ensures that $\mathbf{x}$ has stable first order relative frequencies. Condition (5) ensures that the elements of $\mathbf{x}$ are concentrated on compact sets, and is a natural assumption in view of the fact that every finite measure μ on $(X, \mathcal{B})$ is necessarily tight (see e.g. Billingsley (1968)). Condition (6) ensures that, on select compact sets, one can predict x_{i+1} by a continuous function of x_i with small average error. Predictive sequences of order $k \geq 2$ can be defined in an analogous fashion by extending (4) to functions $g \in C_b(X^k)$, and requiring that x_{i+k} be well approximated on average by a continuous function of $x_i, \ldots, x_{i+k-1}$. Predictive sequences of higher order will be considered elsewhere; in what follows first order predictive sequences will be referred to simply as predictive.

Every D-sequence satisfying (4) is predictive. Conditions (5) and (6) weaken the definition of a D-sequence in several respects. For $\mathbf{x}$ to be predictive there must be a continuous function h such that $h(x_i)$ is close to x_{i+1} for many indices i. Unlike a D-sequence, however,

$h(x_i)$ need not equal x_{i+1}, nor need h be invertible, nor need the set of "good" indices be nested in any way.

To every predictive sequence $\mathbf{x}$ there corresponds a sequence of continuous functions $h_1, h_2, \ldots$, where h_j satisfies (6) with $\epsilon = 1/j$. It is natural to conjecture that these functions form a Cauchy sequence under an appropriate metric, and have a limit which is the transformation determined by $\mathbf{x}$. This is established in Theorem 2 below, where it is shown, in particular, that to every predictive sequence $\mathbf{x}$ there corresponds a measure μ on $(X, \mathcal{B})$ and a unique μ-preserving transformation $T : X \to X$ such that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g(x_i, x_{i+1}) \;=\; \int g(u, Tu) d\mu \quad \text{for every} \quad g \in C_b(X^2). \tag{7}$$

Theorem 2 generalizes the results of Maharam (1965) and Sun (1995) on induced transformations in spaces of finite measure. The proof of Theorem 2 makes use of some elementary facts about sequences with stable relative frequencies, and is based on the existence of one and two dimensional stationary distributions for $\mathbf{x}$. The proofs of Theorem A and its extensions give more direct constructions of the induced transformation, based on conditions like (D1), (D2), and (3) above, and it should be noted that the connection (7) between a predictive sequence and its induced transformation T is weaker than that guaranteed in conclusion (b) of Theorem A and its extensions. However, if a predictive sequence $\mathbf{x}$ is also a D-sequence, the transformations in Theorems A and Theorem 2 preserve the same measure μ and are equal μ-almost everywhere.

1.3 Overview of Paper

The definition and basic properties of averaging sequences are given in the next section. Section 3 is devoted to the properties of predictive sequences, and the proof of Theorem 2. Several corollaries and extensions of Theorem 2 are presented in Section 3.2, including simple conditions for the continuity and invertibility of the induced transformation. Finitary estimation of the induced transformation is briefly discussed in Section 3.4.

In Section 4 two alternative definitions of a deterministic sequence are given. These definitions are shown to be equivalent to the predictive property when X is a normed linear space. As a corollary, a connection between the two dimensional distribution of a sequence and its averaging properties is obtained.

2 Basic Properties of Averaging Sequences

Definition: A sequence $\mathbf{x} = x_1, x_2, \ldots$ with $x_i \in X$ is *tight* if for each $\epsilon > 0$ there exists a compact set $K \subseteq X$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I\{x_i \notin K\} < \epsilon. \tag{8}$$

A sequence $\mathbf{x}$ is *k-averaging* if it is tight and

$$\Lambda(g) \;=\; \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g(x_i, \ldots, x_{i+k-1}) \tag{9}$$

exists for every $g \in \mathcal{C}_b(X^k)$.

Tightness of $\mathbf{x}$ is equivalent to the tightness of the measures $\nu_n(\cdot) = n^{-1}\sum_{i=1}^{n} I\{x_i \in \cdot\}$. It may readily be shown that tightness in one dimension implies tightness in higher dimensions.

The condition (9) is taken from Furstenberg's (1960) extensive study of prediction from individual sequences that take values in a compact Hausdorff space. There the sequences under consideration are assumed to be k-averaging for every $k \geq 1$, and to satisfy additional regularity conditions not required here. Masani (1963) gives an overview of some of Furstenberg's work. Note that if $\mathbf{x}$ is k-averaging then it is l-averaging for $l \leq k$. Beginning with the work of Weyl (1916), there is a substantial literature on the existence, construction, and properties of 1-averaging sequences, which are commonly referred to as uniformly distributed. For an overview, see Kuipers and Niederreiter (1973).

A k-averaging sequence is one that obeys the ergodic theorem for functions in $C_b(X^k)$. While no reference is made to the value of the limits $\Lambda(g)$ in (9), their existence is sufficient to ensure that $\Lambda(g)$ is the integral of g with respect to a unique k-dimensional probability measure μ_k. Furstenberg (1960) noted that, when X is compact and Hausdorff, the existence of μ_k follows directly from the Riesz representation theorem. A similar connection can be established for tight sequences.

Theorem 1 *If $\mathbf{x}$ is tight and k-averaging then there is a unique Borel probability measure μ_k on $(X^k, \mathcal{B}_k)$ such that $\Lambda(g) = \int g d\mu_k$ for every $g \in C_b(X^k)$.*

Proof: Note that $C_b(X^k)$ is closed under scaler multiplication, and the pointwise minimum, maximum and addition of finitely many functions. Thus $C_b(X^k)$ is a vector lattice of functions on X^k. Evidently $\Lambda(1) = 1$, $\Lambda(ag) = a\Lambda(g)$ and $\Lambda(g + g') = \Lambda(g) + \Lambda(g')$ for $a \in \mathbb{R}$ and $g, g' \in C_b(X^k)$, so $\Lambda(\cdot)$ is a normalized linear functional on $C_b(X^k)$.

Let $g_1 \geq g_2 \geq \cdots$ be a decreasing sequence of functions in $C_b(X^k)$ such that $\lim_n g_n(v) = 0$ for every $v \in X^k$. Fix $\epsilon > 0$ and let $\tilde{K} \subseteq X^k$ be a compact set such that

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I\{(x_i, \ldots, x_{i+k-1}) \notin \tilde{K}\} < \epsilon.$$

The functions g_r converge uniformly to zero on $\tilde{K}$, and therefore when r is sufficiently large,

$$\Lambda(g_r) \;\leq\; \limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \left[\sup_{v\in\tilde{K}} g_r(v) + ||g_1|| I\{(x_i, \ldots, x_{i+k-1}) \notin \tilde{K}\} \right] \;\leq\; (1 + ||g_1||)\epsilon.$$

Therefore $\Lambda(g_r) \to 0$ and thus the functional $\Lambda(\cdot)$ is a Daniell integral. It then follows from the Daniell-Stone representation theorem (see Royden (1988, p.432)) that there is unique measure μ_k on $(X^k, \mathcal{S})$ such that $\Lambda(g) = \int_{X^k} g(v) d\mu_k(v)$ for every $g \in C_b(X^k)$. Here $\mathcal{S}$ is the least σ-algebra with respect to which every function in $C_b(X^k)$ is measurable, so that $\mathcal{S} \subseteq \mathcal{B}_k$. On the other hand, $\mathcal{S}$ contains every closed ball in X^k, hence every open ball, and as X^k is separable, $\mathcal{B}_k \subseteq \mathcal{S}$. □

Definition: The measure μ_k appearing in Theorem 1 will be called the k-dimensional distribution of $\mathbf{x}$.

If $T : X \to X$ is a measure preserving transformation of $(X, \mathcal{B}, \mu)$ then μ-almost every trajectory $\mathbf{x} = x, Tx, T^2x, \ldots$ is tight and k-averaging for $k \geq 1$. More generally, almost every sample sequence of a stationary stochastic process taking values in X is tight and k-averaging for $k \geq 1$. The asymptotic behavior of an averaging sequence is similar to that of a sample sequence from a stationary process. In some cases, statistical inference from individual averaging sequences is possible, without the need for stochastic assumptions. Prediction from individual sequences was studied by Furstenberg (1960). Estimating the induced transformation of an individual sequence is discussed in Section 3.4.

The limiting distribution of a k-averaging sequence $\mathbf{x}$ is unchanged under suitable perturbations, insertions, and deletions.

Proposition 1 *Let $\mathbf{x}$ be a tight, k-averaging sequence with k-dimensional distribution μ_k.*

(a) Let $r_n(\mathbf{x}, \mathbf{y})$ be the least number of insertions, deletions, and coordinate-wise changes needed to transform $\mathbf{y}$ into a new sequence $\mathbf{y}'$ such that $y'_1 = y_1, \ldots, y'_n = x_n$. If $n^{-1} r_n(\mathbf{x}, \mathbf{y}) \to 0$ then $\mathbf{y}$ is k-averaging and has k-dimensional distribution μ_k.

(b) If $\mathbf{y}$ is tight and such that $n^{-1} \sum_{i=1}^{n} I\{d(x_i, y_i) > \epsilon\} \to 0$ for every $\epsilon > 0$, then $\mathbf{y}$ is k-averaging and has k-dimensional distribution μ_k.

Proof: If $n^{-1} r_n(\mathbf{x},\mathbf{y}) \to 0$ one may readily show that $\mathbf{y}$ is tight. Moreover for every function $g \in C_b(X^k)$,

$$\left| \frac{1}{n}\sum_{i=1}^n g(x_i,\ldots,x_{i+k-1}) \;-\; \frac{1}{n}\sum_{i=1}^n g(y_i,\ldots,y_{i+k-1}) \right| \;\leq\; \frac{||g||\, k\, r_n(\mathbf{x},\mathbf{y})}{n}.$$

Thus the limiting averages along $\mathbf{x}$ and $\mathbf{y}$ agree, and $\mathbf{y}$ has k-dimensional distribution μ_k. To establish (b), fix $g \in C_b(X^k)$ and $\epsilon > 0$. Let K be a compact subset of X such that

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^n (\, I\{x_i \notin K\} \,+\, I\{y_i \notin K\}\,) < \epsilon\,,$$

and let $\delta > 0$ be such that $d_k(v_1, v_2) < \delta$ implies $|g(v_1) - g(v_2)| < \epsilon$ for every $v_1, v_2 \in K^k$. For each $i \geq 1$ the difference $|g(x_i,\ldots,x_{i+k-1}) \,-\, g(y_i,\ldots,y_{i+k-1})|$ is at most

$$\epsilon \;+\; 2\,||g|| \sum_{j=i}^{i+k-1} (I\{x_i \notin K\} \,+\, I\{y_i \notin K\} \,+\, I\{d(x_j,y_j) \geq \delta\}).$$

It then follows from the assumptions that

$$\limsup_{n\to\infty} \left| \frac{1}{n}\sum_{i=1}^n g(x_i,\ldots,x_{i+k-1}) \;-\; \frac{1}{n}\sum_{i=1}^n g(y_i,\ldots,y_{i+k-1}) \right| \;\leq\; (2k||g|| + 1)\epsilon\,.$$

As both $\epsilon > 0$ and $g \in C_b(X_k)$ were arbitrary, $\mathbf{y}$ is k-averaging with k-dimensional distribution μ_k. □

Proposition 2 *If* $\mathbf{x}$ *is tight and k-averaging then its* k *and* $(k-1)$ *dimensional stationary distributions satisfy* $\mu_{k-1}(A) = \mu_k(A \times X) = \mu_k(X \times A)$ *for each* $A \in \mathcal{B}_{k-1}$.

Proof: Define the measure $\nu(A) = \mu_k(A \times X)$ for $A \in \mathcal{B}_{k-1}$. Then

$$\int I_A(u_1,\ldots,u_{k-1}) d\nu \;=\; \int I_A(u_1,\ldots,u_{k-1}) 1(u_k)\, d\mu_k\,,$$

where $1(u) = 1$ for each $u \in X$, and it follows by standard arguments that $\int \phi\, d\nu = \int \phi \times 1\, d\mu_k$ for every bounded measurable function $\phi : X^{k-1} \to \mathbb{R}$. In particular, for every $g \in C_b(X^{k-1})$,

$$\int g\, d\nu \;=\; \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n g(x_i,\ldots,x_{i+k-2}) 1(x_{i+k-1}) \;=\; \int g\, d\mu_{k-1},$$

and therefore $\nu = \mu_{(k-1)}$. The proof of the second equality is similar. □

For each $A \subseteq X^k$ let $\overline{A}$ and A^o denote, respectively, the closure and interior of A, and let $\partial A = \overline{A} \setminus A^o$ be the boundary of A. The next lemma follows from Theorem 1 and the Portmanteau Theorem for weak convergence of probability measures (c.f. Billingsley (1968)).

Lemma 1 *If* $\mathbf{x}$ *is a tight* k*-averaging sequence with stationary distribution* μ_k*, then the following hold.*

(a) For every set $A \in \mathcal{B}_k$ *such that* $\mu_k(\partial A) = 0$*,*

$$\mu_k(A) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I_A(x_i, \ldots, x_{i+k-1})$$

(b) For every open set $G \subseteq X^k$*,*

$$\mu_k(G) \;\le\; \liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I_G(x_i, \ldots, x_{i+k-1})$$

(c) For every closed set $F \subseteq X^k$*,*

$$\mu_k(F) \;\ge\; \limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I_F(x_i, \ldots, x_{i+k-1}).$$

3 Predictive Sequences

3.1 Averaging Properties and Induced Transformations

By definition every predictive sequence $\mathbf{x}$ is 1-averaging. In fact, predictive sequences have limiting relative frequencies of every finite order.

Proposition 3 *If* $\mathbf{x}$ *is predictive then* $\mathbf{x}$ *is* k*-averaging for each* $k \ge 1$*.*

Proof: The argument proceeds by induction on k: suppose that $\mathbf{x}$ is $(k-1)$-averaging for some $k \ge 2$. Fix a function $g \in C_b(X^k)$ and a number $\epsilon > 0$. Select a compact set $K \subseteq X$ and a continuous function $h : K \to X$ such that (5) and (6) hold. Let $\delta \in (0, \epsilon)$ be such that $v_1, v_2 \in K^k$ and $d_k(v_1, v_2) < \delta$ implies $|g(v_1) - g(v_2)| < \epsilon$. Let $\tilde{g} : X^{k-1} \to \mathbb{R}$ be any continuous function extending $g(u_1, \ldots, u_{k-1}, h(u_{k-1}))$ on $X^{k-2} \times K$, and such that $||\tilde{g}|| = ||g||$. (The existence of $\tilde{g}$ is a consequence of the Tietze Extension Theorem.) For each $u_1, \ldots, u_k \in X$,

$$\begin{aligned} &|\tilde{g}(u_1, \ldots, u_{k-1}) - g(u_1, \ldots, u_k)| \\ &\quad\le\; \epsilon \;+\; 2\,||g||\left[\sum_{j=1}^{k} I\{u_j \notin K\} \;+\; I\{u_{k-1} \in K,\, d(h(u_{k-1}), u_k) \ge \epsilon\}\right] \end{aligned}$$

It then follows from (5), (6), and the choice of K that

$$\limsup_{n\to\infty} \left| \frac{1}{n} \sum_{i=1}^{n} g(x_i, \ldots, x_{i+k-1}) - \frac{1}{n} \sum_{i=1}^{n} \tilde{g}(x_i, \ldots, x_{i+k-2}) \right| \;\le\; (2||g||(k+1)+1)\epsilon.$$

The average of $\tilde{g}$ on $\mathbf{x}$ is convergent since $\mathbf{x}$ was assumed to be $(k-1)$-averaging. As $\epsilon > 0$ was arbitrary, the average of g is convergent as well. □

In what follows it will be convenient to consider the metric $d'(u,v) = \min\{d(u,v), 1\}$, which is bounded and equivalent to $d(\cdot,\cdot)$ on X. The next lemma shows that the two-dimensional distribution of a predictive sequence is concentrated near the graph of a measurable function.

Lemma 2 *If $\mathbf{x}$ is predictive then for every $\gamma > 0$ there exists a function $f : X \to X$ such that $\int d'(f(u), v)\, d\mu_2(u,v) \leq \gamma$.*

Proof: Fix $\gamma > 0$ and let $\epsilon = \gamma/4$. Let $K \subseteq X$ be a compact set and $h : K \to X$ a continuous function such that (5) and (6) hold. Let $f : X \to X$ be any function agreeing with h on the set K. By an application of Proposition 2,

$$\int d'(f(u), v)\, d\mu_2 \;\leq\; \int_{K\times X} d'(f(u), v)\, d\mu_2 \;+\; \mu(K^c)$$

Lemma 1 and inequality (5) together imply that $\mu(K^c) \leq \epsilon$. Let $\tilde{g}$ be any element of $C_b(X^2)$ that extends $d'(h(u), v)$ on $K \times X$, and is such that $||\tilde{g}|| = 1$. Then

$$\begin{aligned}
\int_{K\times X} d'(f(u), v)\, d\mu_2 \;&\leq\; \int \tilde{g}(u,v)\, d\mu_2 \\
&= \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \tilde{g}(x_i, x_{i+1}) \\
&\leq \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} I\{x_i \in K\}\, d'(h(x_i), x_{i+1}) \;+\; \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} I\{x_i \notin K\} \\
&\leq \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} I\{x_i \in K,\, d'(h(x_i), x_{i+1}) \geq \epsilon\} \;+\; 2\epsilon \;\leq\; 3\epsilon
\end{aligned}$$

Thus $\int d'(f(u), v)\, d\mu_2(u,v) \leq 4\epsilon = \gamma$ and the result follows. □

The conclusion of Lemma 2 can be strengthened using the completeness of the set of measurable functions $f : X \to X$ under the metric $\alpha(f_1, f_2) = \int d'(f_1(u), f_2(u))\, d\mu$. The following Lemma may be established by a routine modification of the proof of Theorem 19.1 in Billingsley (1995).

Lemma A *Let μ be a probability measure on $(X, \mathcal{B})$ and let $f_1, f_2, \ldots : X \to X$ be measurable functions. If for every $\epsilon > 0$ there is an integer $N = N(\epsilon)$ such that $\int d'(f_n, f_m)\, d\mu \leq \epsilon$ when $n, m \geq N$, then there is a measurable function $f : X \to X$ for which $\int d'(f_n, f)\, d\mu \to 0$.*

Theorem 2 *If* $\mathbf{x}$ *is a predictive sequence with distribution* μ *then there is a corresponding* μ*-preserving transformation* $T : X \to X$ *such that*

$$\mu_2\{(u,v) : v = Tu\} = 1 \tag{10}$$

and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g(x_i, x_{i+1}) \;=\; \int g(u,Tu)d\mu(u) \qquad \text{each } g \in C_b(X^2). \tag{11}$$

If S *is any other transformation satisfying (11) then* $\mu\{S = T\} = 1$*, so* T *is essentially unique.*

Proof: By Lemma 2 there is, for each $n \geq 1$, a function $h_n : X \to X$ such that $\int d'(h_n(u),v)d\mu_2 \leq 1/n$. It then follows from Proposition 2 that when $n, m \geq N$,

$$\int d'(h_n, h_m)\, d\mu \;\leq\; \int d'(h_n(u),v)\, d\mu_2 + \int d'(h_n(u),v)\, d\mu_2 \;\leq\; 2/N.$$

By Lemma A there exists a measurable function $T : X \to X$ such that $\int d'(h_n, T)\, d\mu \to 0$. For each $n \geq 1$,

$$\int d'(Tu, v)\, d\mu_2 \;\leq\; \int d'(T, h_n)\, d\mu + \int d'(h_n(u), v)\, d\mu_2.$$

Letting $n \to \infty$ it follows that $\int d'(Tu,v)d\mu_2 = 0$, which is equivalent to (10). In particular, if $g \in C_b(X^2)$ then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g(x_i, x_{i+1}) \;=\; \int g(u,v)\, d\mu_2 \;=\; \int g(u,Tu)\, d\mu_2 \;=\; \int g(u,Tu)\, d\mu,$$

which establishes (11). For each $U, V \in \mathcal{B}$,

$$\mu_2(U \times V) \;=\; \int I\{u \in U, v \in V\}\, d\mu_2 \;=\; \int I\{u \in U, Tu \in V\}\, d\mu \;=\; \mu(U \cap T^{-1}V) \tag{12}$$

so that $\mu(T^{-1}V) = \mu_2(X \times V) = \mu(V)$ by Proposition 2. Therefore T preserves μ.

Let $S : X \to X$ be another measurable transformation, and define functions $\beta_1(U,V) = \mu(U \cap T^{-1}V)$ and $\beta_2(U,V) = \mu(U \cap S^{-1}V)$. For each $U, V \in \mathcal{B}$ and $i = 1, 2$ the functions $\beta_i(\cdot, V)$ and $\beta_i(U, \cdot)$ are measures on $\mathcal{B}$. If S satisfies (11) then $\int g(u,Tu)d\mu = \int g(u,Su)d\mu$ for every $g \in C_b(X^2)$ and it follows that $\beta_1(U,V) = \beta_2(U,V)$ for each $U, V \in \{A \in \mathcal{B} : \mu(\partial A) = 0\}$. As this collection generates $\mathcal{B}$ and is closed under intersections, $\beta_1(U,V) = \beta_2(U,V)$ for each U, V. Therefore for each $V \in \mathcal{B}$,

$$\mu(S^{-1}V \setminus T^{-1}V) \;=\; \mu(T^{-1}V \setminus S^{-1}V) \;=\; 0. \tag{13}$$

Let $u_1, u_2, \ldots$ is a countable dense subset of X. Then

$$\{T \neq S\} \;=\; \bigcup_{i=1}^{\infty} \bigcup_{r>0} [\, T^{-1}B(u_i, r) \setminus S^{-1}B(u_i, r) \,] \,,$$

where the second union is restricted to the positive rationals. It follows from (13) that each term in brackets above has μ-measure zero. □

It is shown in the proof of Theorem 2 above that if $\mathbf{x}$ is 2-averaging then (10) implies (11). Conversely, if (11) holds, then (12) holds for U, V with $\mu(\partial U) = \mu(\partial V) = 0$, and one may then deduce (10) from arguments like those used to establish the uniqueness of T above.

Definition: A 2-averaging sequence $\mathbf{x}$ will be said to induce a transformation $T : X \to X$ if either of the equivalent conditions (10) or (11) holds.

Proposition 1 shows that suitable modification of a predictive sequence leaves its asymptotic sample averages unchanged, and therefore yields another predictive sequence with the same induced transformation.

3.2 Corollaries and Extensions

As a corollary of Theorem 2 one may derive a simple necessary and sufficient condition under which a 1-averaging sequence induces a continuous transformation.

Corollary 1 *A 1-averaging sequence* $\mathbf{x}$ *induces a continuous transformation* $S : X \to X$ *if and only if* $n^{-1} \sum_{i=1}^{n} d'(Sx_i, x_{i+1}) \to 0$.

Proof: Let $\mathbf{x}$ be 1-averaging. If $\mathbf{x}$ induces a continuous transformation S then $g(u, v) = d'(Su, v) \in C_b(X^2)$ so that by (11)

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} d'(Sx_i, x_{i+1}) \;=\; \int d'(Su, Su) \;=\; 0.$$

Conversely, if for some continuous $S : X \to X$ the limit of $n^{-1} \sum_{i=1}^{n} d'(Sx_i, x_{i+1})$ is zero, then $\mathbf{x}$ is predictive and Proposition 1 implies that for each $g \in C_b(X^2)$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g(x_i, x_{i+1}) \;=\; \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g(x_i, Sx_i) \;=\; \int g(u, Su)\, d\mu(u).$$

It then follows from the uniqueness part of Theorem 2 that S is an induced transformation of $\mathbf{x}$.

The cited references on D-sequences consider invertible transformations. The transformation T induced by a predictive sequence $\mathbf{x}$ need not be invertible. However 'reversing' the predictive condition gives a simple sufficient condition for the invertibility of T:

> A predictive sequence $\mathbf{x}$ induces an invertible transformation $T : X \to X$ if for every $\epsilon > 0$ there is a compact set $K \subseteq X$ satisfying (5) and a continuous map $h : K \to X$ such that $\limsup_n n^{-1} \sum_{i=1}^n I\{d(x_i, h(x_{i+1})) \geq \epsilon,\, x_{i+1} \in K\} \leq \epsilon$.

Equation (11) is extended to larger numbers of coordinates below. This generalizes a similar result for D-sequences established by Coffey (1989, Lemma 4.2).

Proposition 4 *Let* $\mathbf{x}$ *be a predictive sequence with induced transformation* T *and distribution* μ*. For every* $k \geq 3$ *and every choice of* $g \in C_b(X^k)$*,*

$$\frac{1}{n} \sum_{i=1}^n g(x_i, \ldots, x_{i+k-1}) \;\to\; \int g(x, Tx, \ldots, T^{k-1}x)\, d\mu(x). \tag{14}$$

Proof: The case $k = 2$ is a consequence of Theorem 2. Consider the case $k = 3$; the argument for larger values of k is similar. Let μ_3 be the 3-dimensional distribution of $\mathbf{x}$. Fix $g \in C_b(X^3)$ and define sets $A = \{(u, v, w) : w = Tv\}$ and $B = \{(u, v, w) : v = Tu, w = Tv\}$. Note that

$$\int g(u, v, w)\, d\mu_3 \;=\; \int_A g(u, v, w)\, d\mu_3 \;+\; \int_{A^c} g(u, v, w)\, d\mu_3 \;=\; \int_A g(u, v, w)\, d\mu_3$$

as the the third term is bounded in absolute value by $||g||\, \mu_3(A^c) = ||g||\, \mu_2\{(v, w) : w \neq Tv\} = 0$. Therefore,

$$\int g(u, v, w)\, d\mu_3 \;=\; \int_B g(u, v, w)\, d\mu_3 \;+\; \int_{A \setminus B} g(u, v, w)\, d\mu_3.$$

Since $I\{v \neq Tu, w = Tv\} \leq I\{v \neq Tu\}$, the absolute value of the second term is at most $||g||\, \mu_3(A \setminus B) \leq ||g||\, \mu_2\{(u, v) : v \neq Tu\} = 0$. Therefore

$$\int g(u, v, w)\, d\mu_3 \;=\; \int_B g(u, v, w)\, d\mu_3 \;=\; \int g(u, Tu, T^2u)\, d\mu_3 \;=\; \int g(u, Tu, T^2u)\, d\mu.$$

The result follows immediately by identifying $\int g(u, v, w)\, d\mu_3$ with limiting sample averages of $\mathbf{x}$. □

3.3 Ergodicity of the Induced Transformation

The trajectories of ergodic maps provide a motivating example for the study of deterministic sequences. As expected, such trajectories are predictive. The following proposition may be established by combining Proposition 20 of Sun (1995) with Theorem 5 of Maharam (1965).

Proposition A *If $S : X \to X$ is an ergodic measure preserving transformation of $(X, \mathcal{B}, \mu)$ then for μ-almost every $x \in X$ the trajectory $\mathbf{x} = x, Sx, S^2x, \ldots$ is a predictive sequence with distribution μ and induced transformation $T = S$.*

Some predictive sequences, *e.g.* those with repeated terms, may not be representable as the generic trajectory of an ergodic transformation. More importantly, the natural converse to Proposition A does not hold: the transformation induced by a predictive sequence $\mathbf{x}$ need not be ergodic. Maharam (1965) describes a predictive sequence taking values in $[0, 1]$ that determines the identity map $T(x) = x$. To sketch a simpler example of how T may fail to be ergodic, let $\mathbf{x}$ and $\mathbf{y}$ be two predictive sequences whose respective distributions ν_1 and ν_2 have disjoint supports. Define $s_1 = 1$ and $s_l = s_{l-1} + l^{1/2}$ for $l > 1$. Divide $\mathbf{x}$ and $\mathbf{y}$ into non-overlapping blocks $b_l = x_{s_{l-1}}, \ldots, x_{s_l - 1}$ and $c_l = y_{s_{l-1}}, \ldots, y_{s_l - 1}$, and define a new sequence $\mathbf{z} = b_1, c_1, b_2, c_2, \ldots$ by interleaving the blocks. It can readily be shown that $\mathbf{z}$ is a predictive sequence with distribution $\nu = \nu_1/2 + \nu_2/2$. Moreover, the supports of ν_1 and ν_2 will be non-trivial invariant sets for the induced transformation T of $\mathbf{z}$, and therefore T fails to be ergodic. This construction may be carried with any finite number of initial sequences.

Let $\mathbf{x}$ be a 1-averaging predictive sequence with distribution μ and induced transformation T. For each pair $g_1, g_2 \in C_b(X)$, define

$$R_k(g_1, g_2) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} g_1(x_i) g_2(x_{i+k}) \,.$$

It follows immediately from standard results (e.g. Petersen (1989, Proposition 5.3)) that the induced transformation T of $\mathbf{x}$ is ergodic if and only if

$$\lim_{m \to \infty} \frac{1}{m} \sum_{k=1}^{m} R_k(g_1, g_2) \;=\; \int g_1 \, d\mu \cdot \int g_2 \, d\mu$$

for each $g_1, g_2 \in C_b(X)$. Analogous characterizations for D-sequences were given by Coffey (1989) and Sun (1995). Note that the conditions, involving first a limit with increasing sample size, and then a limit with increasing separation, do not provide an effective means for establishing the ergodicity of the induced transformation.

3.4 Estimation of the Induced Transformation

Theorem 2 shows that predictive sequence $\mathbf{x}$ induces a unique measure preserving transformation $T : X \to X$. The sense in which $\mathbf{x}$ induces T is infinitary: given the entire sequence $\mathbf{x}$ one can, in principle, construct the induced transformation T. However, when the one-dimensional distribution μ of $\mathbf{x}$ is comparable to a known reference measure, it is possible estimate T in a *finitary* fashion from the initial sequences of $\mathbf{x}$. Let μ_0 be a fixed, non-atomic probability measure on $(X, \mathcal{B})$. For each $\kappa > 1$ define the set of measures $\mathcal{D}(\mu_0, \kappa) = \{\mu : \kappa^{-1} \leq d\mu/d\mu_0 \leq \kappa\}$. The following theorem is given in Adams and Nobel (1999).

Theorem B *Given μ_0 and κ there exist measurable maps $\hat{T}_n : X^{n+1} \to X$, $n \geq 1$, such that for every predictive sequence $\mathbf{x}$ with distribution $\mu \in \mathcal{D}(\mu_0, \kappa)$,*

$$\mu(\hat{T}_n^{-1} A \,\Delta\, T^{-1} A) \to 0 \quad \textit{for every} \quad A \in \mathcal{B}\,.$$

Here T is the induced transformation of $\mathbf{x}$, and $\hat{T}_n(x) = \hat{T}_n(x : x_1, \ldots, x_n)$ is an estimate of T based on the first n elements of $\mathbf{x}$.

4 Strong Predictive and Singular Sequences

Two alternative families of "deterministic" sequences are defined below, one formally smaller than the family of predictive sequences, and one larger. It is shown that all three families coincide if X is a normed linear space. The first alternative family is defined by requiring that the functions h appearing in the definition of a predictive sequence be continuous on all of X.

Definition: A sequence $\mathbf{x} = x_1, x_2, \ldots$ with $x_i \in X$ will be called *strong predictive* if it is 1-averaging and for every $\delta > 0$ there is a continuous function $h : X \to X$ such that

$$\limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} I\{d(h(x_i), x_{i+1}) \geq \delta\} \;\leq\; \delta. \tag{15}$$

The second alternative family is defined solely in terms of the two dimensional distribution of $\mathbf{x}$, without (explicit) reference to its limiting sample averages. Recall that two measures ν_1 and ν_2 on $(X, \mathcal{B})$ are mutually singular, written $\nu_1 \perp \nu_2$, if they are supported on disjoint sets, *i.e.* if there exists a set $W \in \mathcal{B}$ such that $\nu_1(W^c) = \nu_2(W) = 0$.

Definition: A 2-averaging sequence $\mathbf{x}$ will be called *singular* if its two dimensional distribution μ_2 is such that

$$\mu_2(\cdot \times V) \perp \mu_2(\cdot \times V^c) \quad \text{for each} \quad V \in \mathcal{B}\,, \tag{16}$$

in other words, the one-dimensional measures $\mu_2(\cdot \times V)$ and $\mu_2(\cdot \times V^c)$ have disjoint supports.

Using general results on the connection between set mappings and point mappings (c.f. Royden (1988)) one may show that every singular sequence induces a measure preserving transformation (c.f. Rudolph (1990)). A special case of this fact follows from Proposition 5 below, where it is shown that the three types of sequences defined above coincide if X satisfies the following condition.

Assumption I: For every finite measure μ, every finite-valued Borel measurable function $f : X \to X$ and every $\epsilon > 0$ there exists a continuous function $h : X \to X$ such that $\int d'(f,h)\, d\mu < \epsilon$.

It follows readily from Lusin's Theorem that Assumption I holds when $X = \mathbb{R}^l$, $l \geq 1$, and when $X = \mathbb{R}^\infty$ with metric $d(\mathbf{a}, \mathbf{b}) = \sum_{i \geq 1} 2^{-i} |a_i - b_i|$. A direct argument, using Urysohn's Lemma and the regularity of Borel measures, shows that Assumption I holds for any separable Banach space.

Proposition 5 *If* $\mathbf{x} = x_1, x_2, \ldots$ *takes values in a complete separable metric space satisfying Assumption I, then the following are equivalent:*

(a) $\mathbf{x}$ *is predictive*

(b) $\mathbf{x}$ *is singular*

(c) $\mathbf{x}$ *is strong predictive*

Proof: In each case the definitions ensure that $\mathbf{x}$ is 1-averaging, and therefore tight. If $\mathbf{x}$ is predictive then it is 2-averaging by Proposition 5, and it follows readily from the relation (12) established in the proof of Theorem 2 that $\mathbf{x}$ is singular.

Suppose then that $\mathbf{x}$ is singular, and fix $\epsilon > 0$. Let $K \subseteq X$ be a compact set such that $n^{-1} \sum_{i=1}^n I\{x_i \notin K\}$ is eventually less than ϵ, and let $\pi = \{V_1, \ldots, V_r\}$ be a finite partition of X such that $\mathrm{diam}(V_i) \leq \epsilon$ if $V_i \cap K \neq \emptyset$. It follows from (16) that to each $V_i \in \pi$ there corresponds a set $U_i \subseteq X$ such that $\mu_2(U_i \times V_i^c) = \mu_2(U_i^c \times V_i) = 0$. In particular, one has

$$\mu(U_i) \;=\; \mu_2(U_i \times X) \;=\; \mu_2(U_i \times V_i) \;=\; \mu_2(X \times V_i) \;=\; \mu(V_i)$$

so that $\sum_{i=1}^r \mu(U_i) = 1$. Moreover, if $i \neq j$ then

$$\begin{aligned} \mu(U_i \cap U_j) &= \mu_2((U_i \cap U_j) \times V_i) + \mu_2((U_i \cap U_j) \times V_i^c) \\ &\leq \mu_2(U_j \times V_j^c) + \mu_2(U_i \times V_i^c) \;=\; 0. \end{aligned}$$

Thus, by removing or adding a μ-nullset from each U_i, we may assume that $\{U_1, \ldots, U_r\}$ are a partition of X.

Select points $v_i \in V_i$ and define $f : X \to X$ by setting $f(u) = v_i$ if $u \in U_i$. By Assumption I there is a continuous function $h : X \to X$ such that $\int d'(h,f)d\mu \le \epsilon$. Then $g(u,v) = d'(h(u),v)$ is bounded and continuous, and since $\mathbf{x}$ is 2-averaging,

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} d'(h(x_i),x_{i+1}) &= \int d'(h(u),v)\,d\mu_2 \\
&\le \int d'(f(u),v)\,d\mu_2 + \int d'(h(u),f(u))d\mu \\
&\le \sum_{i=1}^{r}\int_{X\times V_i} d'(f(u),v)\,d\mu_2 + \epsilon \\
&= \sum_{i=1}^{r}\int_{U_i\times V_i} d'(f(u),v)\,d\mu_2 + \epsilon \\
&= \sum_{i=1}^{r}\int_{U_i\times V_i} d'(v_i,v)d\mu_2 + \epsilon \\
&\le \sum_{i=1}^{r}[\epsilon\mu_2(U_i\times V_i) + \mu_2(K^c\times V_i)] + \epsilon \;\le\; 3\epsilon,
\end{aligned}$$

where the second equality is a consequence of the fact that $\mu_2(U_i^c \times V_i) = 0$. This last inequality and the elementary relation

$$I\{d(h(x_i),x_{i+1}) \ge \sqrt{3\epsilon}\} \;\le\; d'(h(x_i),x_{i+1})/\sqrt{3\epsilon}$$

show that (15) holds with $\delta = \sqrt{3\epsilon}$. As $\epsilon > 0$ was arbitrary, it follows that $\mathbf{x}$ is strong predictive. Finally, it is clear that every strong predictive sequence is predictive. □

Proposition 5 establishes a connection between the k-averaging properties of a sequence and general features of its two dimensional distribution.

Corollary 2 *Let $\mathbf{x}$ be a 2-averaging sequence taking values in a separable Banach space. If the 2-dimensional distribution of $\mathbf{x}$ is singular or, equivalently, is concentrated on the graph of a measurable function, then $\mathbf{x}$ is k-averaging for every $k \ge 1$.*

Acknowledgements

The author wishes to thank Idris Assani, Rajiva Karandikar, and Karl Petersen for their comments and suggestions.

References

[1] T.M. Adams and A.B. Nobel (1999). Finitary reconstruction of a measure preserving transformation. Submitted for publication.

[2] T.A. Bick (1967). On orbits in spaces of infinite measure. *Duke Math. J.*, 34(4) 717-724.

[3] T.A. Bick and J. Coffey (1991). A class of examples of D-sequences. *Ergod. Th. & Dynam. Sys.*, 11 1-6.

[4] P. Billingsley (1995). *Probability and Measure*, Third edition. John Wiley and Sons, New York.

[5] P. Billingsley (1968). *Convergence of Probability Measures* John Wiley and Sons, New York.

[6] J. Coffey (1989). Uniformly distributed D-sequences. In *Measure and Measurable Dynamics*, Contemporary Math. 94 97-112.

[7] H. Furstenberg (1960). *Stationary Processes and Prediction Theory.* Annals of Math. Studies 44, Princeton Univ. Press, Princeton, NJ.

[8] D.A. Kappos and S. Papadopoulou (1967). Transformations determined by the orbit of a general point. *Rev. Roum. Math. Pures et Appl.*, XII(9) 1297-1303.

[9] L. Kuipers and H. Niederreiter (1974). *Uniform Distribution of Sequences.* Wiley, New York.

[10] S.P. Lalley (1998). Beneath the noise, chaos. Submitted for publication.

[11] D. Maharam (1965). On orbits under ergodic measure-preserving transformations. *Trans. AMS*, 119 51-66.

[12] P.R. Masani (1963). Review of *Stationary Processes and Prediction Theory. Bull. AMS*, 69(2) 195-207.

[13] K. Petersen (1989). *Ergodic Theory.* Cambridge Univ. Press.

[14] H.L. Royden (1988). *Real Analysis*, third edition. Prentice Hall, New Jersey.

[15] D.J. Rudolph (1990). *Fundamentals of Measurable Dynamics.* Oxford University Press, New York.

[16] H. Weyl (1916). Über die Gleichverteilung von Zahlen mod. Eins. *Math. Ann.*, 77 313-352.

[17] Y. Sun (1995). Isomorphisms for convergence structures. *Adv. in Math.*, 116 322-355.