ISRAEL JOURNAL OF MATHEMATICS, Vol. 36, Nos. 3–4, 1980

VERSIK PROCESSES: FIRST STEPS

BY
ARTHUR ROTHSTEIN

ABSTRACT

A. M. Versik defined a property of finite-state stationary stochastic processes that formally is slightly weaker than D. S. Ornstein's very weak Bernoulli (*VWB*) property. Versik conjectured his property was equivalent with *VWB*. The conjecture is false. We construct a zero entropy Versik process that is not loosely Bernoulli (*LB*) and, by taking a skew product of this process with the two-shift, we produce a *K* process that is Versik but not *VWB*. Our technique for constructing a non-*LB* process differs from the few known methods.

§1. Introduction

We are interested here in stationary processes (T, P). T is bijective, bimeasurable, measure preserving map on a probability space $(\Omega, \mathcal{F}, m)$ and $P = \{P^1, \cdots, P^n\}$ is a partition of Ω into pairwise disjoint measurable sets. We freely identify Ω with $\{1, \cdots, n\}^Z$ via $x(i) = j$ if $T^i x \in P^j$. In this identification T becomes the shift on the sequence space. We say x and y agree at time i (under T — this qualifier is used when required for clarity) if $x(i) = y(i)$.

The process (T', P') is a factor of the process (T, P) if there is a surjective, measure preserving map ϕ from the underlying space of (T', P') such that $T' \circ \phi = \phi \circ T$. The processes are isomorphic if ϕ^{-1} also makes (T, P) a factor of (T', P'). Two processes can be factors of each other without being isomorphic (see [15]). Note that these definitions avoid mention of P. Indeed, much of ergodic theory can be developed, and is more naturally formulated, in terms of T alone. Only recently was it proved ([8]) that for a large class of T this formulation achieves no greater generality.

If $P = \{P^1, \cdots, P^n\}$ is a partition, define $I(P)(x) = -\log m(P^i)$ if $x \in P^i$. (All logarithms are taken base 2.) Define $H(P) = \int I(P)$, called the entropy of the process (T, P). If (T, P) is a process, $(1/n)H(\bigvee_{i=1}^n T^i P)$ decreases to a limit, written $h(T, P)$ and called the entropy of the process (T, P). Shannon and

Received February 11, 1980

McMillan proved that $(1/n)I(\bigvee_{i=1}^{n} T^iP)$ converges in L^1, and Breiman extended the result to pointwise convergence (see [2] or [14]). If T is described without a partition, the entropy of T, written $h(T)$, is defined as $\sup h(T, P)$ where the supremum is taken over all partitions P. If $\bigvee_{i \in Z} T^iP = \mathcal{F}$ for some partition P, then $h(T) = h(T, P) \leqq H(P) < \infty$; such a P is called a (finite) generator. Krieger proved ([8], see also [4, chapter 28]): If $h(T) < \log n$, then T has a generator P with n atoms.

Entropy is an isomorphism invariant for processes: If (T', P') is a factor of (T, P), then $h(T', P') \leqq h(T, P)$. For an important class of processes, the converse question held great interest for many years: If $h(T', P') = h(T, P)$, are (T', P') and (T, P) isomorphic? The classes in question were the K (for Kolmogorov) and B (for Bernoulli) processes. A process (T, P) is K if the intersections $\bigcap_{n=k}^{\infty} \bigvee_{i=n}^{\infty} T^iP$ decrease to the trivial σ-algebra (ϕ, Ω). Rohlin and Sinai proved (see [14]) (T, P) is K if and only if every factor (T, Q) has positive entropy. A process (T, P) is B if the partitions $\{T^iP,\ i \in Z\}$ are independent. That B processes are K follows from Kolmogorov's zero-one law. Ornstein proved B processes are classified (i.e., their isomorphism classes are determined) by their entropy, but K processes are not (see [11]).

If (T, Q) is a factor of a K process (T, P), then (T, Q) has no factors of zero entropy, hence (T, Q) is K. The case for factors of B processes is not as nice. If (T, P) is B, then $(T, P \vee TP)$ is not B, even though it is isomorphic to (T, P). Ornstein found a property, *VWB* (for "very weak Bernoulli"), which is held by precisely those processes which are isomorphic to B processes. With this result, an earlier result of Ornstein ([11, section I.5]) reads that factors of *VWB* processes are *VWB*.

Suppose the sequences $x = (x^1, \cdots, x^n)$ and $y = (y^1, \cdots, y^n)$ are drawn from a finite alphabet. Define their distance as $d(x, y) = (1/n)\#\{i : x^i \neq y^i\}$. (The d and f we use to metrize symbols usually appear in the literature as $\bar{d}$ and $\bar{f}$. We have removed the overscore for typographical convenience.) Suppose m and m' are probabilities on $\Omega = \{1, \cdots, k\}^n$. Regard m and m' as nonatomic by regarding Ω as the unit interval partitioned into k^n subintervals. Suppose ϕ is a measure preserving $(m \circ \phi^{-1} = m')$ map on Ω, and consider $\int d(x, \phi x) m(dx)$. Define the distance between m and m', written $d(m, m')$, as the infimum of this integral over all such ϕ.

Suppose (T, P) is a process. Fix n and for each $x \in \Omega$, consider the distance between two probabilities on $\bigvee_{i=1}^{n} T^{-i}P$, m and $m(\cdot \mid P^+)(x)$. Since the distance depends measurably on x, we can integrate over Ω. Call the result $VWB^n(T, P)$. (T, P) is *VWB* if $VWB^n(T, P) \to 0$ as $n \to \infty$.

Consider (T, P) afresh. Fix n and consider two probabilities on

$$Q = \bigvee_{i=1}^{n} T^{-i}P \times \bigvee_{i=1}^{n} T^{i}P = \bigvee_{i=1}^{n} (T^{-1} \times T)^{i}(P \times P), \tag{1.1}$$

$$A \times B \mapsto m(A)m(B) \quad \text{and} \quad A \times B \mapsto m(A \cdot B).$$

Identify Q with $\{1, \cdots, k^2\}^n$, then denote the distance between these probabilities by $V^n(T, P)$. Since the image of ϕ is the diagonal of $\Omega \times \Omega$, we may just as soon consider ϕ as mapping $\Omega \times \Omega$ to Ω. We say (T, P) is Versik (after A. M. Versik) if $V^n(T, P) \to 0$ as $n \to \infty$.

Think of (T, P) as a teletype which every second prints a character according to some stationary probability law. An observer, watching the teletype, has seen all outputs since time began. Knowing these outputs and the probability law, he wants to predict the next n characters. Another observer also wants to predict the future, but he knows only the probability law. To complicate matters, the print mechanism on the teletype has just become defective. A small percentage of the subsequent outputs will be illegible. The probability law of the teletype is *VWB* if the observer who has seen the prior outputs has no advantage over his companion. If the teletype was always defective — the first observer has seen most but not all of the prior outputs — and if he has no advantage, then the probability law of the teletype is Versik. These heuristics suggest *VWB* processes are Versik. Formally the argument runs as follows.

Let (T, P) be any process. Fix n and for each y in Ω let ϕ_y be a map on Ω that achieves $d(m, m(\cdot \mid P^+)(y))$. The map $\phi : (x, y) \mapsto \phi_y(x)$ preserves the measures (1.1), and for all (x, y),

$$d_{T^{-1} \times T}(x, y ; \phi(x, y), \phi(x, y)) = d_T(x, \phi(x, y)),$$

which upon double integration yields

$$\iint d_{T^{-1} \times T}(x, y ; \phi(x, y), \phi(x, y)) m(dx) m(dy) = VWB^n(T, P).$$

We conclude $V^n(T, P) \leqq VWB^n(T, P)$. Thus *VWB* processes are Versik.

Versik processes share several important properties with B processes and K processes: They are mixing of all orders; they are closed under taking products, factors, powers and roots; if $\{T^t, t \in R\}$ is an ergodic flow and T^1 is Versik, then T^t is Versik for all t. In a 1976 letter to Ornstein, Versik conjectured his property was equivalent with *VWB*. In fact, it does not even imply K. There are Versik processes with zero entropy. Even if a Versik process is K, it still may fail to be *VWB*. Our goal in this thesis is to construct such processes. By taking advantage of the Equivalence Theory (see [18] for a presentation) we can do both jobs with a single construction. We explain how.

If x is a string of symbols from a finite alphabet, let $l(x)$ denote its length. If x and y are two symbol strings of arbitrary length, define $f(x, y)$ as

$$\frac{n}{l(x)+l(y)},$$

where n is the minimal number of deletions from x and y required to make the strings identical. Clearly, if $l(x)=l(y)$, $f(x,y)\leqq d(x,y)$. If (T,P) is a process, define $LB^n(T,P)$ for each n by substituting f for d in the definition of $VWB^n(T,P)$. A process (T,P) is loosely Bernoulli (LB) if $LB^n(T,P)\to 0$ as $n\to\infty$. Thus VWB processes are LB. When $h(T,P)=0$, (T,P) is LB if and only if for every (x,y) in $\Omega\times\Omega$, $f((x(1),\cdots,x(n)),(y(1),\cdots,y(n)))\to 0$. Among the zero entropy LB processes are irrational rotations of the circle. All positive entropy LB processes are equivalent (see Chapter 2 for a definition) as are all zero entropy LB processes. Loose Bernoullicity is an equivalence invariant. Factors of LB processes are LB.

Suppose (T,P) is an ergodic process, i.e.,

$$\frac{1}{n}\sum_{i=1}^{n} m(A\cdot T^iA)\to m(A)^2 \qquad \text{for all } A.$$

Let (U,Q) be the two-shift. Consider the process $(T\times U, P\times Q)$, then replace $T\times U$ by

$$S:(x,y)\mapsto\begin{cases}(x,Uy), & \text{if } y\in Q^0;\\ (Tx,Uy), & y\in Q^1.\end{cases} \tag{1.2}$$

A process constructed this way is called a skew product. Meilijson proved ([9]) $(S,P\times Q)$ is K. Feldman proved ([5]) (T,P) is LB if $(S,P\times Q)$ is LB. In particular, (T,P) is LB if $(S,P\times Q)$ is VWB. (The converse is not true. Burton ([3]) constructed a LB process (T,P) for which $(S,P\times Q)$ is not VWB.) To this we add the observation, proved below, that $(S,P\times Q)$ is Versik if (T,P) is Versik.

Thus we will have both our desired counterexamples once we construct a zero entropy non-LB Versik process: If (T,P) is Versik and not LB, then $(S,P\times Q)$ is Versik, K and not VWB. We construct this (T,P) in §6.

Our basic idea, heuristically, is this. We construct a process (T,P) inductively by cutting and stacking. Zero blocks have relatively prime lengths, so we can avoid periodicity in (T,P) without adding mass at any step of the induction. We would like to form n-blocks by independently concatenating $n-1$-blocks. We cannot quite do this, however, since we want to reduce the entropy, if only slightly. So we stack some number $n \cdot a(n)$ of $n-1$-blocks independently, followed by $a(n)$ $n-1$-blocks stacked in some very deterministic way. This is enough to give (T, P) zero entropy, and if $a(n)$ is sufficiently large with respect to $a(1), \cdots, a(n-1)$, then (T, P) is still Versik. Moreover, we can make (T, P) *LB* or non-*LB* by our choice of the deterministic stacking on the $a(n)$ $n-1$-blocks. Feldman's original zero entropy non-*LB* construction ([5]) used periodicity to enforce f-separation of symbol strings. Our construction does just the opposite, using independence.

The construction in §6 combines two techniques: how to construct a zero entropy Versik process, and how to construct a zero entropy non-*LB* process. We develop these techniques separately, the former in §4, the latter in §5. In §2 we lay bare the language and some of the assumptions we will use in our constructions. §3 collects the little general information we have about Versik processes.

§7 surveys where the subject has gone since our first results in early 1978. All but one of the new results are due to M. Gerber, a student of J. Feldman at Berkeley; they will be included in her thesis. Gerber is also responsible for the current form of Theorem 4.1. In its original version the theorem imposed additional constraints on the deterministic part of the stacking.

This work benefits from many conversations I have had with D. Ornstein, D. Rudolph and J. Feldman. Ornstein, M. Gerber and B. Weiss helped me in trying to make this paper readable. To all of them go my heartfelt thanks.

§2. Background information

Facts about the f distance

Suppose x is a juxtaposition of two symbol strings, $x = x^1x^2$, where null strings are not excluded. Then any given symbol string y may be decomposed into a juxtaposition, $y = y^1y^2$, such that

$$f(x, y) = \sum_{i=1}^{2} f(x^i, y^i) \frac{l(x^i) + l(y^i)}{l(x) + l(y)}.$$

Indeed, let y^2 begin with the first symbol in y that matches the first non-deleted symbol in x^2. This extends by induction: If x is a juxtaposition of n symbol strings, $x = x^1 \cdots x^n$, then any given symbol string y may be decomposed as $y = y^1 \cdots y^n$ such that

$$f(x,y)=\sum_{i=1}^{n} f(x^i,y^i)\frac{l(x^i)+l(y^i)}{l(x)+l(y)}.$$

For all symbol strings x and y,

$$\frac{l(x)-l(y)}{l(x)+l(y)}\leqq f(x,y)\leqq 2\frac{\inf(l(x),l(y))}{l(x)+l(y)}.$$

The left-hand inequality implies

$$\frac{l(x)}{l(y)}\leqq\frac{1+f(x,y)}{1-f(x,y)}. \tag{2.1}$$

Primer on Kakutani equivalence

Suppose (T,P) is an ergodic process and $0<m(A)<1$. Define $T_A:A\to A$ by mapping x to $T^{n(x)}x$, where $n(x)$ is the first time x returns to A. Under the probability $B\mapsto m(A\cdot B)/m(A)$, T_A is a bijective, bimeasurable, measure preserving, ergodic map. T_A has no natural finite generator, so we do not speak of an induced process. Abramov ([1]) proved $h(T_A)=h(T)/m(A)$.

If S is isomorphic to T_A, we say S is a derivative of T and T is a primitive of S. S and T are called (Kakutani) equivalent if they have a common derivative. S and T have a common derivative if and only if they have a common primitive. From this it follows that this notion of equivalence is indeed an equivalence relation. By Abramov's formula, a positive entropy process cannot be equivalent to a zero entropy process.

Given a process (S,P), we may construct processes (T,Q) which are primitives of (S,P). Let $(\Omega,\mathcal{F},m)$ be the underlying space of (S,P), and suppose $g:\Omega\to Z^+$ has finite integral. Let Ω' be the graph of g, $\mathcal{F}'$ the product σ-algebra restricted to this set, and m' the product measure, normalized to one. Define T on Ω' by sending (ω,n) to $(\omega,n+1)$ if $(\omega,n+1)$ is in Ω', else to $(S\omega,0)$. For each i let Q^i be the set of (ω,n) in Ω' with $\omega\in P^i$. We say (T,Q) is exduced from (S,P). If $A(n)$ is the set of ω in Ω with (ω,n) in Ω', we say the nth level of (T,P) is built over $A(n)$.

Primer on stacking

Our construction method in §§4, 5 and 6 is a technique commonly known as cutting and stacking. This technique has seen wide use in recent years as Ornstein and his school. have used it to construct counterexamples to many conjectures in ergodic theory (see, e.g., [3], [10], [12], [15], [16]). While the technique is geometric in origin (see, e.g., the presentation in [7]), recent usage has shown it worthwhile also to think of stacking as a way of concatenating symbol strings to form the generic points of an ergodic process. We take this point of view here, and in so doing we present a formalism for stacking that is tailored for our subsequent arguments.

A block is a string of symbols drawn from a finite alphabet. A tower is a finite collection of distinct blocks weighted by a probability distribution.

Let τ be a tower with blocks $x^1, \cdots, x^n$, where block x^i has probability p^i. Let r be a positive integer. For $1 \leqq i \leqq n$, let f^i be the function on $\Omega = \{1, \cdots, n\}^r$ which assigns to the n-tuple x the number of occurrences of i in x. Suppose m is a probability distribution on Ω satisfying $\Sigma_{x \in \Omega} f^i(x) \cdot m(x) = r \cdot p^i$ for all i between 1 and n. Define a tower φ by assigning probability $m(i(1), \cdots, i(r))$ to the block $x^{i(1)} \cdots x^{i(r)}$. We say φ is a stacking of r copies of τ according to the distribution m, or briefly, φ is a stacking of τ. We will denote this relationship by $\tau < \varphi$. This is an equivalence relation.

Suppose τ and φ are towers. The ith block in τ is x^i and it occurs with probability p^i. The ith block in φ is y^i and it occurs with probability q^i. By the independent product $\tau \times \varphi$ we mean the tower which assigns probability $p^i q^j$ to the block $x^i y^j$. This notion of product is associative. When there is no danger of confusing an exponent with a superscript, let τ^n denote the n-fold product $\tau \times \cdots \times \tau$.

Lemma 2.2. *If $\varphi < \tau^1$ and $\varphi < \tau^2$ then $\varphi < \tau^1 \times \tau^2$.*

A sequence of towers $\{\tau(n), n \geqq 0\}$, with $\tau(n) < \tau(n+1)$ for all n, under very mild conditions which we will always meet, uniquely determines an ergodic process (T, P), where P has as many atoms as $\tau(0)$ has symbols in its alphabet. We write $(T, P) = \lim \uparrow \tau(n)$ or $\tau(n) \uparrow (T, P)$. We may refer to $\tau(n)$ as an n-tower, to the blocks in $\tau(n)$ as n-blocks. We may regard each trajectory x of (T, P) as a concatenation of zero blocks.

Suppose all blocks in the tower τ have length n, and that for each i the ith block has probability p^i. Define the entropy of τ, $h(\tau)$, as $1/n$ times the entropy of the distribution $(p^1, \cdots, p^s)$. If blocks in τ^1 and τ^2 have length n^1 and n^2, respectively, then

$$h(\tau^1 \times \tau^2) = \frac{n^1 \cdot h(\tau^1) + n^2 \cdot h(\tau^2)}{n^1 + n^2}. \tag{2.3}$$

Lemma 2.4. *If $(T, P) = \lim \uparrow \tau(n)$ and for each n all blocks in $\tau(n)$ have the same length, then* $\lim h(\tau(n))$ *exists and equals* $h(T, P)$.

Suppose τ is a tower. For each block in τ, replace each occurrence of the symbol 1 by the pair 11. Call the resulting tower τ^f the fattening of τ. Fattening commutes with products: $(\tau \times \varphi)^f = \tau^f \times \varphi^f$; in particular, $(\tau^n)^f = (\tau^f)^n$. Fattening also commutes with stacking: If $\tau < \varphi$ then $\tau^f < \varphi^f$. Suppose $(T, P) = \lim \uparrow \tau(n)$. Then (T^f, P^f), defined as $\lim \uparrow \tau(n)^f$, is equivalent to (T, P). Indeed, we get (T^f, P^f) by building a single level over P^1 and exducing.

Throughout this paper we use $\{0, 1\}$ to denote the tower with two blocks, 0 and 1, each with probability 1/2.

A connection between stacking and LB

Suppose all blocks in the tower τ have length $l(\tau)$. The ith block in τ is x^i and it occurs with probability p^i. Define $f(\tau) = \Sigma_{i,j} p^i p^j \cdot f(x^i, x^j)$. If all blocks in the tower φ have length $l(\varphi)$, then

$$f(\tau \times \varphi) \leqq \frac{l(\tau)f(\tau) + l(\varphi)f(\varphi)}{l(\tau) + l(\varphi)}. \tag{2.5}$$

Lemma 2.6. *Suppose $(T, P) = \lim \uparrow \tau(n)$ has zero entropy and all blocks in each $\tau(n)$ have the same length. Then (T, P) is LB if and only if $f(\tau(n)) \to 0$.*

Proof. For x and y in Ω, write

$$f((x(1), \cdots, x(n)), (y(1), \cdots, y(n)))$$

as $f_n(x, y)$. Since $h(T, P) = 0$, $LB^n(T, P) = \int f_n(x, y) m(dx) m(dy)$.

Suppose $f(\tau(n)) \to 0$. Given $\varepsilon > 0$, choose n sufficiently large that $f(\tau(n)) < \varepsilon^2$. Then there is a set A of n-blocks, with probability at least $1 - \varepsilon$ such that the f-distance of any two blocks in A is less than 2ε. Choose points x and y in Ω. Suppose r is a large integer. By removing less than $l(\tau(n))$ symbols from each end of $(x(1), \cdots, x(r))$, we may consider x as a concatenation of n-blocks. For r sufficiently large, at least $1 - 2\varepsilon$ of these n-blocks are in A. Remove the others and enough additional n-blocks so that exactly $1 - 2\varepsilon$ of the n-blocks remain. (Without real loss we assume $1 - 2\varepsilon$ times the number of n-blocks is an integer.) Do the same for y. What remains in both x and y are n-blocks that are all within 2ε of each other in f-distance. Thus $\limsup f_r(x, y) \leqq 4\varepsilon$. Since ε was arbitrary, we conclude $f_r(x, y) \to 0$. Thus (T, P) is LB.

Now suppose (T, P) is LB. Suppose $\varepsilon < 1/2$, choose n large, let l be the length of an n-block, and set $r = 2l - 1$. For n sufficiently large, $LB^r(T, P) < \varepsilon^4$. Regard each x in Ω as a concatenation of n-blocks. Partition Ω into at most l sets A^i; x is in A^i if $(x(i), \cdots, x(i + l))$ is an n-block, and x is not in A^j for any $j < i$.

Suppose $(x, y) \in A^i \times A^j$, $|i-j| \leqq l\varepsilon$, z^s is the n-block starting at $x(i)$, and z^t is the n-block starting at $y(j)$. If $f_r(x, y) < \varepsilon$, then $f(z^s, z^t) < 7\varepsilon$. Indeed, match x to y, by removing symbols, to achieve $f_r(x, y)$. We remove $u \leqq r\varepsilon$ symbols from each. Without loss of generality assume $i \leqq j$. Then at least $l - 3u + i - j$ of the symbols in $(x(j+u), \cdots, x(i+l-u))$ remain, and they match to symbols in z^t. Thus

$$f(z^s, z^t) \leqq \frac{3j-i}{l} < 7\varepsilon.$$

For a collection of $A^i \times A^j$ with $\Sigma\, m(A^i)m(A^j) > 1 - \varepsilon$,

$$\int_{A^i} \int_{A^j} f_r(x, y) m(dx) m(dy) < \varepsilon^2 m(A^i) m(A^j). \tag{2.7}$$

Since, for n sufficiently large, the union of the $A^i \times A^j$ with $|i-j| \leqq l\varepsilon$ has measure greater than $2\varepsilon(1-\varepsilon)$, one of these $A^i \times A^j$ satisfies (2.7). So for at least $1-\varepsilon$ of the points (x, y) in $A^i \times A^j$, $f_r(x, y) < \varepsilon$. Thus for at least $1-\varepsilon$ of the pairs (z^s, z^t), $f(z^s, z^t) < 7\varepsilon$. So $f(\tau(n)) < 8\varepsilon$. Since ε was arbitrary, we conclude $f(\tau(n)) \to 0$, which completes the proof.

A property of VWB processes

The following result is an elementary consequence of the not-so-elementary Ornstein theory. The key to its proof is the fact that *VWB* processes satisfy a property called "finitely determined" (see [11, part I]).

If (T, P) is a process, let ν^s denote the distribution on s-strings determined by $\bigvee_{i=0}^{s-1} T^i P$.

Lemma 2.8. *Suppose (T, P) is VWB. Given $\varepsilon > 0$ there exist $n(\varepsilon)$ and $\delta(\varepsilon)$ such that if $s > n(\varepsilon)$ and ν is a distribution on s-strings with $f(\nu, \nu^s) < \delta(\varepsilon)$, then $d(\nu, \nu^s) < \varepsilon$.*

§3. Elementary results for Versik processes

Asymmetric Versik

We may attempt to strengthen the Versik condition as follows. Given (T, P) and positive integers r and s. Consider two probabilities on

$$Q = \bigvee_{i=1}^{r} T^{-i} P \times \bigvee_{i=1}^{s} T^i P,$$

$A \times B \mapsto m(A)m(B)$ and $A \times B \mapsto m(A \cdot B)$. Denote the distance between them by $V^{r,s}(T, P)$, where the distance between the sequences

$$(x(-s),\cdots,x(-1),x(1),\cdots,x(r))$$

and

$$(y(-s),\cdots,y(-1),y(1),\cdots,y(r))$$

is measured as $1/s$ times the number of errors in the interval $[-s,-1]$ plus $1/r$ times the number of errors in the interval $[1,r]$. We say (T,P) is asymmetric Versik if $V^{r,s}(T,P)\to 0$ as $r\to\infty$ and $s\to\infty$. Since

$$V^n(T,P)\leqq V^{n,n}(T,P)\leqq 2V^n(T,P)$$

for all n, asymmetric Versik implies Versik. The converse is also true. A proof, due to whom we don't know, follows.

If r and s remain comparable as they increase — say $1/c<r/s<c$ for fixed c — then $V^{r,s}(T,P)$ is approximated within a factor of c by $V^{t,t}(T,P)$, where $t=\max(r,s)$. Thus it suffices to prove $V^{r,s}(T,P)\to 0$ as $r\to\infty$ and $s/r\to\infty$. Suppose r and s/r are large. Then $V^{s,s}(T,P)$ is small, say ε^2. The d error in the interval $[1,s]$ is, if we ignore insignificant end effects, an average of $s-r$ d errors: The error on $[1,r]$, on $[2,r+1]$, etc. At least $1-\varepsilon$ of these averages must be less than ε. Thus for some $i<s\varepsilon$, the error on $[i+1,i+r]$ is less than ε. Shift the time origin by i units: Group the coordinates $1,\cdots,i$ with $-s,\cdots,-1$. Since $i<s\varepsilon$, this does not materially affect the d estimate on these latter coordinates. We conclude $V^{r,s}(T,P)$ is on the order of ε, which completes the proof.

Factors of Versik processes are Versik

Suppose (T,P) and (U,Q) are processes. Given n, identify $\bigvee_{i=0}^{n-1}T^iP$ and $\bigvee_{i=0}^{n-1}U^iQ$ with $\{1,\cdots,k\}^n$ for some k. Denote by $d_n((T,P),(U,Q))$ the distance between the respective probabilities on these n-tuples. Define $d((T,P),(U,Q))$ as $\limsup d_n((T,P),(U,Q))$. This lim sup is actually a limit ([11, appendix C]). If (T,Q) is a factor of (T,P), then for partitions $Q_n\leqq\bigvee_{i=-n}^{n}T^iP$, $d((T,Q),(T,Q_n))\to 0$. Now suppose (T,P) is Versik. Then each $(T,\bigvee_{i=-n}^{n}T^iP)$ is Versik, so each (T,Q_n) is Versik, and by the triangle inequality we get (T,Q) is Versik.

Powers

By the symmetry of the Versik property, if (T,P) is Versik, then so is (T^{-1},P). Suppose (T,P) is Versik and $k>1$. Match the distributions of (1.1). At worst, all the errors occur at times T^{nk}, $n\in Z$. This means $V^n(T^k,P)\leqq k\cdot V^{nk}(T,P)$, which shows (T^k,P) is Versik.

Roots

We say a measure preserving transformation T is Versik if (T, P) is Versik for all finite partitions P. If T^k is Versik and P is a finite partition, then $(T^k, \bigvee_{i=0}^{k-1} T^iP)$ is Versik. This matches the distributions in (1.1) except for k coordinates; i.e.,

$$V^{nk}(T, P) \leqq \frac{1}{n} + V^n\left(T^k, \bigvee_{i=0}^{k-1} T^iP\right).$$

Since a finite number of coordinates exert little influence on $V^s(T, P)$ for large s, we conclude (T, P) is Versik. Since P was arbitrary, we conclude T is Versik.

Flows

Suppose $\{T^t, t \in R\}$ is an ergodic flow and T^1 is Versik. By the above, T^r is Versik for all nonzero rational r. Fix a positive irrational α and a finite partition P. Choose a positive rational r, sufficiently small that for most points x, most of the points T^tx for $0 < t < r$ lie in the same atom of P as does x. For each integer n, let $k(n)$ be the smallest integer with $r \cdot k(n) \geqq n\alpha$. For n large, $V^n(T^r, P)$ is small. Since the sequences $\{k(n), n > 0\}$ and $\{k(n), n < 0\}$ both have density r/α, the $\bar{d}$ error is small even when we restrict our attention to the partitions $T^{k(i)r}P$. Replace each $T^{k(i)r}P$ by $T^{i\alpha}P$. By the choice of r, this does not materially affect the $\bar{d}$ estimate. Thus $V^s(T^\alpha, P)$ is small, where $s \approx nr/\alpha$. We conclude (T^α, P) is Versik. Since P was arbitrary, T^α is Versik.

Versik processes are mixing

A process (T, P) is mixing of order r if for all $(A^1, \cdots, A^r)$,

$$m\left(\bigcap_{i=1}^r T^{n(i)}A^i\right) \to \prod_{i=1}^r m(A^i) \qquad \text{as } |n(i) - n(j)| \to \infty, \quad i \neq j.$$

Mixing of order 2 is simply called mixing. (A long-unsolved problem is whether these distinctions are necessary.) To prove (T, P) is mixing it suffices to show $m(A \cdot T^nA) \to m(A)^2$ whenever (T, Q) is a factor of (T, P) and A is an atom of Q. Since factors of Versik processes are Versik, we need only show $m(A \cdot T^nA) \to m(A)^2$ whenever (T, P) is Versik and A is an atom of P. Fix n. Identify $\bigvee_{i=1}^n T^{-i}P \times \bigvee_{i=1}^n T^iP$ with $\{1, \cdots, k\}^{2n}$. Let μ be the measure $B \times C \mapsto m(B)m(C)$ in (1.1). There is a map ϕ on $\Omega \times \Omega$, preserving the measures of (1.1), such that

$$\int \frac{1}{n} \# \{i : x(i) \neq \phi x(i) \quad \text{or} \quad x(-i) \neq \phi x(-i)\} \mu(dx) = V^n(T, P).$$

Then for some i,

$$\mu\{x : x(i) \neq \phi x(i) \quad \text{or} \quad x(i-n) \neq \phi x(i-n)\} \leqq 2V^n(T,P).$$

This measure dominates

$$|m(T^{-i}A \cdot T^{n-i}A) - m(T^{-i}A)m(T^{n-i}A)|.$$

Since m is T-invariant we conclude

$$|m(A \cdot T^n A) - m(A)^2| \leqq 2V^n(T,P).$$

Now let $n \to \infty$.

This method extends to prove (T, P) is mixing of all orders.

A Versik skew product

$(S, P \times Q)$ is defined as in (1.2) above.

LEMMA 3.1. *$V^n(S, P \times Q) < 12/n + 7\sqrt{V^n(T,P)}$. In particular, $(S, P \times Q)$ is Versik if (T, P) is.*

PROOF. We may think of $(S, P \times Q)$ as a random walk along the trajectories of (T, P). If at any time n the walk y is at $x(i)$, then at time $n+1$, y advances to $x(i+1)$ only if $y(n) = 1$; otherwise y remains at $x(i)$.

Without loss of generality let $(\Omega, \mathcal{F}, m)$ be the probability space underlying (T, P) and (U, Q). Let $\mu = m \times m$, the product probability on $(\Omega \times \Omega, \mathcal{F} \otimes \mathcal{F})$. Fix n and let $\phi : \Omega \times \Omega \to \Omega$ be the map that achieves $V^n(T, P)$. Set $\varepsilon = \sqrt{V^n(T,P)}$. Without real loss assume $n\varepsilon$ is an integer. There is a set $E \subset \Omega \times \Omega$ with $\mu(E) > 1 - \varepsilon$ such that for each (x, y) in E and at least $n(1-\varepsilon)$ of the i between 1 and n, (x, y) agrees with $(\phi(x,y), \phi(x,y))$ under $T \times T^{-1}$.

To compute $V^n(S, P \times Q)$ we must consider the probabilities on

$$\bigvee_{i=1}^{n} S^{-i}(P \times Q) \times \bigvee_{i=1}^{n} S^{i}(P \times Q),$$

$$A \times B \times C \times D \mapsto \mu(A \times B \times C \times D) = m(A)m(B)m(C)m(D)$$

and

$$A \times B \times C \times D \mapsto \mu((A \times B) \cdot (C \times D)) = m(A \cdot C)m(B)m(D).$$

The map $(w, x; y, z) \mapsto (w, x'; y, z')$, where $x' = z' = \phi(x, z)$, clearly commutes with these probabilities. Our estimate on $V^n(S, P \times Q)$ will follow once we produce a set $F \subset \Omega \times \Omega$ with $\mu(F) > 1 - 12/n$ such that for each (w, y) in F, each (x, z) in E and at least $n(1 - 6\varepsilon)$ of the i between 1 and n, $(w, x; y, z)$ agrees with

$(w, \phi(x,z); y, \phi(x,z))$ under $S \times S^{-1}$. By symmetry it suffices to find a set $A \subset \Omega$ with $m(A) > 1 - 6/n$ such that for each w in A, each (x,z) in E and at least $n(1-3\varepsilon)$ of the i between 1 and n, $S^i(w,x)$ agrees with $S^i(w, \phi(x,z))$.

Recall the proof of the weak law of large numbers for a sequence $X_1, \cdots, X_i, \cdots$ of independent random variables with the same distribution as the square integrable random variable X: For $\sigma > 0$,

$$\sigma^2 \cdot m\left\{ \left| \frac{1}{n}\sum_{i=1}^n X_i - \int X \right| \geqq \sigma \right\} \quad \text{by Chebyshev's inequality}$$

$$\leqq \left| \frac{1}{n}\sum_{i=1}^n X_i - \int X \right|^2$$

$$(3.1.1) \qquad = \frac{1}{n}\left(\int X^2 - \left(\int X \right)^2 \right)$$

$$\leqq \frac{1}{n}\int X^2.$$

Fix (x,z) in E and consider those times i when $S^i(w,x)$ disagrees with $S^i(w, \phi(x,z))$. The walk w must be sitting at a coordinate j where x and $\phi(x,z)$ disagree. The walk remains at this coordinate for exactly k units of time with probability $1/2^k$. Since there are at most $n\varepsilon$ of these coordinates, the number of disagreements between (w,x) and $(w,\phi(x,z))$ is at most $\Sigma_{j=1}^{n\varepsilon} X^j(w)$ where the X^j are independent with common distribution X, $m(X = k) = 1/2^k$. Set $A = \{w : \Sigma_{j=1}^{n\varepsilon} X^j(w) < 3n\varepsilon\}$. Since $\int X = 2$ and $\int X^2 = 6$, (3.1.1) implies $m(A) > 1 - 6/n$. This completes the estimate of $V^n(S, P \times Q)$.

§4. A zero entropy Versik process

THEOREM 4.1. *Suppose $\{a(n), n \geqq 1\}$ and $\{b(n), n \geqq 1\}$ are sequences of positive integers with $a(n) \to \infty$ and $b(n) \to \infty$. Suppose $\{\tau(n), n \geqq 0\}$ and $\{\varphi(n), n \geqq 0\}$ are sequences of towers satisfying*

(4.1.1) *the lengths of blocks in $\tau(0)$ have greatest common divisor* 1,

(4.1.2) *$\varphi(n)$ is a stacking of $a(n+1)$ copies of $\tau(n)$, $n \geqq 0$,*

(4.1.3) *$\tau(n) = \varphi(n-1) \times \tau(n-1)^{a(n)b(n)}$, $n \geqq 1$.*

Set $(T,P) = \lim \uparrow \tau(n)$. If $a(n) \to \infty$ sufficiently fast (i.e., $a(n)$ is sufficiently large with respect to $a(n-1), \cdots, a(1)$), then (T,P) is Versik.

PROOF. For each n let (T_n, P_n) be the process determined by indefinitely stacking $\tau(n)$ independently. By (4.1.1) and (4.1.3), the lengths of the blocks in each $\tau(n)$ have greatest common divisor 1. Hence each (T_n, P_n) is a factor of a mixing finite-step Markov process. Ornstein and Friedman showed (see [11, section I.8]) mixing finite-step Markov processes are *VWB*. Thus (T_n, P_n), a factor of a *VWB* process, is *VWB* and, a fortiori, Versik. Make sure $a(n)$ is large enough that $V^s(T_{n-1}, P_{n-1}) < 1/n$ whenever $s > a(n)$.

Suppose $a(n+1) < s \leqq 2a(n+2)$. Choose a string $x = (x(0), \cdots, x(s-1))$ from $\bigvee_{i=0}^{s-1} T^i P$. Regard x as a concatenation of $\tau(n+2)$ blocks. Since s is much smaller than any $\tau(n+2)$ block, x, with very large probability (greater than $1 - 1/a(n+1)$), lies in a single $\tau(n+2)$ block. Without real loss assume this probability is 1. Decompose x into the constituent $\tau(n+1)$ blocks of this $\tau(n+2)$ block.

Momentarily set aside the $\tau(n+1)$ blocks in $\varphi(n+1)$. Of the remaining $\tau(n+1)$ blocks — the ones that were stacked independently — a small fraction of the coordinates at the start of each come from $\varphi(n)$. Overlay these coordinates with $\tau(n)$ blocks chosen independently. If the last $\tau(n)$ block overlaying a particular $\varphi(n)$ block runs past the end of the $\varphi(n)$ block, throw away the excess coordinates of the $\tau(n)$ block. Now overlay $\varphi(n+1)$ by independently chosen $\tau(n+1)$ blocks, again truncating the last one if it runs too long.

What we have done, then, is to make a small $\bar{d}$ change (roughly $2/b(n+1)$) to $\bigvee_{i=0}^{s-1} T^i P$, producing a distribution which is very close in $\bar{f}$ to $\bigvee_{i=0}^{s-1} T_n^i P_n$. This $\bar{f}$ distance is $1/s$ times the number of remaining coordinates in the $\tau(n)$ blocks that were truncated. It is essentially the ratio of the length of a $\tau(n)$ block to the length of a $\tau(n+1)$ block. This ratio varies inversely with $a(n+1)$. Thus by Lemma 2.8 and the triangle inequality, for $a(n+1)$ sufficiently large, we get $\bar{d}_s((T,P),(T_n,P_n))$ small (roughly $2/b(n+1)$).

By another application of the triangle inequality, we have for all s and n,

$$V^s(T,P) \leqq V^s(T_n,P_n) + \bar{d}_{2s}((T,P),(T_n,P_n)) + \bar{d}_s((T,P),(T_n,P_n)).$$

Given s large, choose n with $a(n+1) < s \leqq a(n+2)$. Then all three terms on the right are small, and they approach zero as s increases. We conclude (T,P) is Versik.

THEOREM 4.2. *There exists a zero entropy LB Versik process.*

PROOF. We construct by induction an increasing sequence of towers, $\{\tau(n), n \geqq 0\}$; $\tau(n)$ will have $g(n)$ blocks, all of length $l(n)$, all equiprobable. Set $\tau(0) = \{0, 1\}$.

Given $\tau(n-1)$, impose an arbitrary ordering on its blocks. Let $a(n)$ be a large integer. Let $\varphi(n-1)$ be a tower with just one block: the first $\tau(n-1)$ block written $a(n)$ times, the second written $a(n)$ times, etc. Define

$$\tau(n) = \varphi(n-1) \times \tau(n-1)^{n \cdot a(n) g(n-1)}.$$

Since

$$g(n) = g(n-1)^{n \cdot a(n) g(n-1)}$$

and

$$l(n) = (n+1) a(n) g(n-1) l(n-1),$$

we have

$$h(\tau(n)) = \frac{n}{n+1} h(\tau(n-1)).$$

By (2.5),

$$f(\tau(n)) = \frac{n}{n+1} f(\tau(n-1)).$$

Do this for every n. Then $h(\tau(n)) \to 0$ and $f(\tau(n)) \to 0$.

Define $(T, P) = \lim \uparrow \tau(n)$. By Lemma 2.4, $h(T, P) = 0$. By Lemma 2.6, (T, P) is *LB*. Since (T^f, P^f) is equivalent to (T, P), $h(T^f, P^f) = 0$ and (T^f, P^f) is *LB*. By Theorem 4.1, if $a(n) \to \infty$ sufficiently fast, (T^f, P^f) is Versik.

§5. A non-*LB* process

Suppose all the blocks in the tower τ have length k. We say τ is (a, b) separated (assume $a < 1/3$, $b < 1/2$) if there exists a positive δ with $2^{14\sqrt{k}} \delta^b < 1$ such that for every string y in every $\{0,1\}^r$, $f(x, y) \geqq a$ for $1 - \delta$ of the blocks x in τ. Note this condition implies $f(\tau) > a/2$. Thus if $(T, P) = \lim \uparrow \tau(n)$ has zero entropy and each $\tau(n)$ is $(a(n), b(n))$ separated, where $\liminf a(n) > 0$, then by Lemma 2.6 (T, P) is not *LB*.

Lemma 5.1. *Suppose τ is (a, b) separated. For n sufficiently large and towers $\varphi^1, \cdots, \varphi^n$, each a stacking of two copies of τ, the tower*

$$\varphi = \varphi^1 \times \cdots \times \varphi^n$$

is $(a(1-2b), b/2)$ separated.

We need a simple inequality for binomial coefficients.

Lemma 5.2. *If $0 < \sigma < 1$, then $\binom{n}{|n\sigma|} < 2^{3n\sqrt{\sigma}}$.*

Proof. Clearly it suffices to consider the case where $n\sigma$ is integral. Then it also suffices to consider the case $\sigma \leqq 1/2$. For all x in $(0, 1)$, $x^{4x} > 1/8$. Put a probability measure on $\{0,1\}^n$. If $(x(1),\cdots,x(n))$ has exactly k zeroes, give it measure $\sigma^k(1-\sigma)^{n-k}$. The number of strings with exactly $n\sigma$ zeroes is $\binom{n}{|n\sigma|}$. The measure of each such string is

$$\begin{aligned}\sigma^{n\sigma}(1-\sigma)^{n(1-\sigma)} &\geqq \sigma^{2n\sigma}\\ &= \sqrt{\sigma}^{4n\sqrt{\sigma}\sqrt{\sigma}}\\ &> 2^{-3n\sqrt{\sigma}}.\end{aligned}$$

Since the measure of the whole space is one, we are done.

Proof of Lemma 5.1. Let δ be the number associated with (a,b). For n sufficiently large and $\varepsilon=\varepsilon(n)=(2^{14\sqrt{k}}\delta^b)^n$ we have $2^{14\sqrt{kn}}\varepsilon^{b/2}<1$. We will use this ε to show φ is $(a(1-2b),b/2)$ separated.

Fix r, then fix y in $\{0,1\}^r$. Suppose $x\in\varphi$. Express x as a concatenation of τ blocks, $x=x^1\cdots x^{2n}$. For this x, decompose y as $y^1\cdots y^{2n}$, so that

$$f(x,y)=\sum_{i=1}^{2n} f(x^i,y^i)\frac{l(x^i)+l(y^i)}{l(x)+l(y)}.$$

For convenience write

$$\frac{l(x^i)+l(y^i)}{l(x)+l(y)}$$

as t^i. If $f(x,y)<1/3$, then $l(y)<2l(x)=4nk$, hence as x ranges over $\{f(x,y)<a(1-2b)\}$ the number of distinct decompositions (recall some y^i may be null) is less than

$$2n\cdot\binom{4nk}{2n}<2n\cdot 2^{12nk\sqrt{1/2k}}<2^{9n\sqrt{k}}$$

for large n. Fix one such decomposition. We will be done once we show that for this decomposition

$$m\left\{x\in\varphi:\sum t^if(x^i,y^i)<a(1-2b)\right\}<2^{5n\sqrt{k}}\delta^{nb}.$$

Suppose $I\subset\{1,\cdots,n\}$ has exactly $[nb]+1$ elements. Define

$$L(I)=\{x\in\varphi: f(x^{2i-1},y^{2i-1})<a\quad\text{for all}\quad i\in I\},$$

$$R(I)=\{x\in\varphi: f(x^{2i},y^{2i})<a\quad\text{for all}\quad i\in I\}.$$

By hypothesis each $\{x: f(x^{2i},y^{2i})<a\}$ has measure at most δ, and since each lies

in a distinct φ^i, they are all independent: $m(R(I)) < \delta^{nb}$. Likewise for $L(I)$. The number of such I is no greater than

$$\binom{n}{[nb]+1} < 2^{3n\sqrt{b+1/n}} < 2^{4n}.$$

Since $2 \cdot 2^{4n}\delta^{nb} \leqq 2^{5n\sqrt{k}}\delta^{nb}$, we will be done once we show that for every $x \in \varphi$ with $\Sigma\, t^i f(x^i, y^i) < a(1-2b)$ there is an I with $x \in L(I) \cup R(I)$.

Fix x with $\Sigma\, t^i f(x^i, y^i) < a(1-2b)$ and let J be the set of i with $f(x^i, y^i) < a$. Then

$$\begin{aligned} a(1-2b) &> \sum t^i f(x^i, y^i) \\ &\geqq \sum_{i \in J^c} t^i f(x^i, y^i) \\ &\geqq a \sum_{i \in J^c} t^i, \end{aligned}$$

so $\Sigma_{i\in J} t^i > 2b$. Since $a < 1/3$, we have $l(x) + l(y) > \frac{3}{2} l(x)$; whenever $i \in J$, we have $l(x^i) + l(y^i) < 3l(x^i)$, so $t^i < 1/n$ for all i in J. Hence $2b < \Sigma_{i\in J} t^i < (1/n)\#(J)$, or $\#(J) > 2nb$. Let J' be the set of even indices in J. Without loss of generality, assume $\#(J') > nb$. By removing as many additional indices as necessary pare J' down to size $[nb]+1$. Then $J' = \{2i : i \in I\}$ for some I, and $x \in R(I)$.

THEOREM 5.3. *There exists a zero entropy non-LB process.*

PROOF. We construct by induction an increasing sequence of towers, $\{\tau(n), n \geqq 0\}$. The tower $\tau(n)$ will have $g(n)$ blocks, all of length $l(n)$, all equiprobable. Fix $k > 3600$ and set $a(0) = 1/3k$. Define inductively $a(n) = (1-2^{-n})a(n-1)$. The tower $\tau(n)$ will be $(a(n), 1/2^{n+2})$ separated.

Set $\tau(0) = \{0,1\}^k$. For $x \in \tau(0)$, a string $y \in \{0,1\}^r$ satisfies $f(x,y) < a(0)$ if and only if $r = k$ and $y = x$. Thus for $\delta = 2^{-k}$ we see $\tau(0)$ is $(a(0), 1/4)$ separated.

Given $\tau(n-1)$, define $\varphi(n)$ by stacking two copies of $\tau(n-1)$, where each block in $\varphi(n)$ is a single block of $\tau(n-1)$, written twice. Define $\tau(n) = \varphi(n)^{c(n)}$ where $c(n)$ is large. By Lemma 5.1, if $c(n)$ is sufficiently large, then $\tau(n)$ is $(a(n), 1/2^{n+2})$ separated. Since $g(n) = g(n-1)^{c(n)}$ and $l(n) = 2c(n)l(n-1)$, we have $h(\tau(n)) = \frac{1}{2} h(\tau(n-1))$. Do this for every n. Then $h(\tau(n)) \to 0$.

Define $(T, P) = \lim \uparrow \tau(n)$. By Lemma 2.4, $h(T, P) = 0$. Since $\liminf a(n) > 0$, the remarks at the beginning of this chapter apply: (T, P) is not LB.

§6. A non-*LB* Versik process

THEOREM 6.1. *There exists a zero entropy non-LB Versik process.*

PROOF. We construct by induction an increasing sequence of towers $\{\tau(n), n \geqq 0\}$; $\tau(n)$ will have $g(n)$ blocks, all of length $l(n)$, all equiprobable. Fix $k > 3600$ and set $a(0) = 1/3k$. Define inductively $a(n) = (1 - 2^{-n})a(n-1)$. The tower $\tau(n)$ will be $(a(n), 1/2^{n+2})$ separated.

Set $\tau(0) = \{0, 1\}^k$. As in Theorem 5.3, $\tau(0)$ is $(a(0), 1/4)$ separated.

Given $\tau(n-1)$, define $\varphi(n-1)$ by stacking two copies of $\tau(n-1)$, where each block in $\varphi(n-1)$ is a single block of $\tau(n-1)$, written twice. Define

$$\tau(n) = \varphi(n-1)^{c(n)} \times \tau(n-1)^{2n \cdot c(n)},$$

where $c(n)$ is large. By Lemma 5.1, if $c(n)$ is sufficiently large, then $\tau(n)$ is $(a(n), 1/2^{n+2})$ separated. Since

$$g(n) = g(n-1)^{(2n+1)c(n)}$$

and

$$l(n) = 2(n+1)c(n)l(n-1),$$

we have

$$h(\tau(n)) = \frac{2n+1}{2n+2} h(\tau(n-1)).$$

Do this for every n. Then $h(\tau(n)) \to 0$.

Define $(T, P) = \lim \uparrow \tau(n)$. As in Theorem 5.3, $h(T, P) = 0$ and (T, P) is not *LB*. Since (T^f, P^f) is equivalent to (T, P), $h(T^f, P^f) = 0$ and (T^f, P^f) is not *LB*. By Theorem 4.1, if $c(n) \to \infty$ sufficiently fast, then (T^f, P^f) is Versik.

§7. Extensions and open problems

Extending all counterexamples to the Versik class

During the early 1970s, Ornstein and others constructed counterexamples to a wide range of conjectures in ergodic theory (see, e.g., [10], [12], [15], [16]). Many of these results concerned properties of *VWB* processes that were not held by *K* processes. We offer a program of constructing these counterexamples in the Versik class. For instance, find an uncountable number of pairwise nonisomorphic processes with the same entropy which are Versik and *K*; find a *K* Versik process with no square root; find two *K* Versik processes which are weakly isomorphic but not isomorphic; find a Versik process which is not the direct product of a K process and a zero entropy process; find a zero entropy Versik process which commutes only with its powers.

More Versik constructions

If a transformation T is LB (i.e., (T, P) is LB for all P), $T \times T$ need not be LB (see [17]). Katok constructed a zero entropy T for which $T \times T$ is LB. His T was weakly mixing but not mixing. Gerber used Katok's method on the deterministic part of a Versik construction to get a Versik T with $T \times T\, LB$. She modified this construction to get a Versik T with all finite products $T \times \cdots \times T\ LB$. L. Swanson independently extended Katok's method to get a T with all finite products $T \times \cdots \times T\, LB$; her transformation, like Katok's is weak mixing but not mixing.

Versik flows

Say an ergodic flow $\{T^t, t \in R\}$ is Versik if T^1 is a Versik transformation. In §3 we showed that for such flows T^t is Versik for all $t \neq 0$. Are Versik flows Bernoulli? No. Gerber has constructed zero entropy Versik flows, both LB and non-LB, with our processes for cross sections.

Half-half extremality

If (T, P) is a process, let ν^n be the distribution on n-strings determined by $\bigvee_{i=0}^{n-1} T^i P$. Call (T, P) half-half extremal if for every $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon) > 0$ such that for n sufficiently large, if $\nu = \nu^n$ is almost an average of two distributions on n-strings,

$$d\left(\nu, \frac{\mu^1 + \mu^2}{2}\right) < \delta,$$

then $d(\nu, \mu^1) < \varepsilon$. This is a weak form of J. P. Thouvenot's notion of extremality (see [6]), which is equivalent with VWB. Ornstein has shown that Versik processes are half-half extremal. Does the converse hold?

References

1. L. M. Abramov, *On the entropy of a derived automorphism*, Dokl. Akad. Nauk SSSR **128** (1959), 647–650; English transl.; Amer. Math. Soc. Transl. (2) **49** (1966), 162–166.

2. P. Billingsley, *Ergodic Theory and Information*, John Wiley, 1965.

3. R. M. Burton, Jr., *A non-Bernoulli skew product*, Thesis, Stanford Univ., 1977.

4. M. Denker, C. Grillenberger and K. Sigmund, *Ergodic theory on compact spaces*, Lecture Notes in Math. **527**, 1976.

5. J. Feldman, *New K-automorphisms and a problem of Kakutani*, Israel J. Math. **24** (1976), 16–38.

6. J. Feldman, *Some remarks about finitely determined and finitely fixed processes*, unpublished (1976).

7. N. Friedman, *Introduction to Ergodic Theory*, Van Nostrand Reinhold, 1970.

8. W. Krieger, *On entropy and generators of measure-preserving transformations*, Trans. Amer. Math. Soc. **149** (1970), 453–464.

9. I. Meilijson, *Mixing properties of a class of skew-products*, Israel J. Math. **19** (1974), 266–270.

10. D. S. Ornstein, *A mixing transformation for which Pinsker's conjecture fails*, Advances in Math. **10** (1973), 103–123.

11. D. S. Ornstein, *Ergodic Theory, Randomness and Dynamical Systems*, Yale Univ. Press, 1974.

12. D. S. Ornstein and P. Shields, *An uncountable family of K automorphisms*, Advances in Math. **10** (1973), 63–68.

13. D. S. Ornstein and B. Weiss, *Finitely determined implies very weak Bernoulli*, Israel J. Math. **17** (1974), 94–104.

14. W. Parry, *Entropy and Generators in Ergodic Theory*, W. A. Benjamin, 1969.

15. S. Polit, *Weakly isomorphic transformations need not be isomorphic*, Thesis, Stanford Univ., 1974.

16. D. J. Rudolph, *Two nonisomorphic K-automorphisms with isomorphic squares*, Israel J. Math. **23** (1976), 274–287.

17. D. J. Rudolph, *Nonequivalence of measure preserving transformations*, Lecture Notes, Institute of Advanced Studies, Hebrew Univ. of Jerusalem, 1976.

18. B. Weiss, *Equivalence of Measure Preserving Transformations*, Lecture Notes, Institute of Advanced Studies, Hebrew Univ. of Jerusalem, 1976.

BANK OF AMERICA
BOX 37000
SAN FRANCISCO, CALIF 94137 USA