COLLOQUIUM MATHEMATICUM

VOL. 101 2004 NO. 1

TAIL FIELDS GENERATED BY SYMBOL COUNTS IN MEASURE-PRESERVING SYSTEMS

BY

KARL PETERSEN (Chapel Hill, NC) and JEAN-PAUL THOUVENOT (Paris)

In memory of Anzelm Iwanik

Abstract. A finite-state stationary process is called (one- or two-sided) *super-K* if its (one- or two-sided) *super-tail field*—generated by keeping track of (initial or central) symbol counts as well as of arbitrarily remote names—is trivial. We prove that for every process (α, T) which has a direct Bernoulli factor there is a generating partition β whose one-sided super-tail equals the usual one-sided tail of β. Consequently, every K-process with a direct Bernoulli factor has a one-sided super-K generator. (This partially answers a question of Petersen and Schmidt.)

1. Introduction and statement of results. A bilateral finite-state ergodic stationary process $\ldots, x_{-1}, x_0, x_1, \ldots$ corresponds to an invertible measure-preserving transformation $T : X \to X$ on a nonatomic Lebesgue probability space $(X, \mathcal{B}, \mu)$ and a finite measurable partition $\alpha = \{A_1, \ldots, A_r\}$ of X according to the relationship $x_i = j \in \{1, \ldots, r\}$ if and only if $T^i x \in A_j$, $i \in \mathbb{Z}$. In a convenient notation, we will also write $\alpha(T^i x) = A_j$, or even $\alpha(T^i x) = j$. With such a process are associated the *tail fields*

$$
\begin{aligned}
\mathcal{T}^+(\alpha) &= \bigcap_{n\ge 0} \mathcal{B}\{x_n, x_{n+1}, \ldots\},\\
\mathcal{T}^-(\alpha) &= \bigcap_{n\ge 0} \mathcal{B}\{x_{-n}, x_{-n-1}, \ldots\},\\
\mathcal{T}^\pm(\alpha) &= \bigcap_{n\ge 0} \mathcal{B}\{x_j : |j| \ge n\}.
\end{aligned}
\tag{1.1}
$$

The system $(X, \mathcal{B}, \mu, T)$ is called K or *Kolmogorov* if for every partition α the tail field $\mathcal{T}^+(\alpha)$ is trivial, i.e. consists only of sets of measure 0 or 1. For a unilateral stationary process $x_1, x_2, \ldots$ corresponding to an endomorphism of $(X, \mathcal{B}, \mu)$, this is equivalent to ergodicity of the measure μ for the action of the group Γ of changes to finitely many coordinates of x (or the *odometer*),

2000 *Mathematics Subject Classification*: 37A05, 37A35, 37A50.

Key words and phrases: tail field, K-system, super-K system, Bernoulli system, exchangeable algebra, entropy.

and to ergodicity of μ for the *Gibbs equivalence relation* for which two unilateral sequences on the finite alphabet $\{1, \ldots, r\}$ are equivalent if and only if they differ in only finitely many coordinates. The Kolmogorov Zero-One Law states that independent identically distributed (i.i.d.) processes have this property; in fact, bilateral i.i.d. processes also have trivial two-sided tail fields. We deal mainly with bilateral processes and reserve the terms "one-sided" and "two-sided" to refer to tail fields.

We now consider some finer tail fields, which keep track not only of which cell of α is entered at times arbitrarily far out, but also of how many times each cell of α has been entered up to that time. For this purpose we use the vectors $v_n^m(x) \in \{0, 1, \ldots\}^r$ defined by

$$v_n^m(x)(i) = \#\{j : n \leq j \leq m \text{ and } x_j \in A_i\}. \tag{1.2}$$

We will also abbreviate $v_m = v_0^m$, and for any set of coordinates H, let

$$v_H(x)(i) = \#\{j \in H : x_j \in A_i\}. \tag{1.3}$$

Now define the *super-tail fields* by

$$\begin{aligned} \mathcal{F}^+(\alpha) &= \bigcap_{n \geq 0} \mathcal{B}\{v_n, v_{n+1}, \ldots\}, \\ \mathcal{F}^-(\alpha) &= \bigcap_{n \geq 0} \mathcal{B}\{v_{-n}^0, v_{-n-1}^0, \ldots\}, \\ \mathcal{F}^\pm(\alpha) &= \bigcap_{n \geq 0} \mathcal{B}\{v_{-j}^j : j \geq n\}. \end{aligned} \tag{1.4}$$

Evidently also

$$\mathcal{F}^+(\alpha) = \bigcap_{n \geq 0} \mathcal{B}\{v_n, x_{n+1}, x_{n+2}, \ldots\}, \tag{1.5}$$

and similarly in the other cases. These are special cases of *cocycle-generated* tail fields determined by a function ϕ from $\{1, \ldots, r\}$ to a group G: if in the above definitions $v_n^m(x)$ is replaced by

$$v(\phi)_n^m(x) = \phi(x_m) \cdots \phi(x_n), \tag{1.6}$$

then the resulting fields are denoted $\mathcal{F}_\phi^+$, $\mathcal{F}_\phi^-$, and $\mathcal{F}_\phi^\pm$, respectively. (The special case $G = \mathbb{Z}^r$ and $\phi(j) = e_j$, the jth standard basis vector, produces the previous super-tail fields.)

We will call the process (α, T) *super-*K^+ in case $\mathcal{F}^+(\alpha)$ is trivial; super-K^- and super-$K^\pm$ are defined analogously. This idea was introduced in [7], where it was also noted that triviality of the appropriate super-tail field corresponds to ergodicity under the group Π of *permutations* of finitely many coordinates, and, in the unilateral case, to ergodicity of the adic transformation. There it was proved that many Gibbs measures, including mixing Markov measures, are super-$K^\pm$—more generally, many Gibbs measures are quasi-invariant (nonsingular) and ergodic for countable equivalence relations on subshifts of finite type generated by certain kinds of cocycles taking values in discrete groups.

The super-K property on its face depends on the choice of the partition, unlike the K property. In [7] the question was raised whether every K-system has a super-K generator. In the one-sided case, when there is a (positive-entropy) direct Bernoulli factor, we can answer this affirmatively.

THEOREM 1.1. *Let $T : X \to X$ be a m.p.t. on a nonatomic Lebesgue probability space and α a generating finite measurable partition of X. Suppose that the process (α, T) is isomorphic to the direct product of a positive-entropy Bernoulli system (B, σ) and another system (Y, S). Then there is a generating partition β for (X, T) such that $\mathcal{F}^+(\beta) = \mathcal{T}^+(\beta) = \mathcal{T}^+(\alpha)$. Thus every K process with a direct Bernoulli factor has a super-K^+ generator.*

We remark that the Ornstein–Shields K-automorphisms [5], the Feldman non-loosely-Bernoulli K-automorphisms [1], Kalikow's T, T^{-1} example [4], and the examples produced by Hoffman's K counterexample machine [3] all have direct Bernoulli factors.

Our investigation actually began with the two-sided case, when we noticed how to recode any process (α, T) to an isomorphic one (β, T) with two-sided super-tail $\mathcal{F}^\pm(\beta)$ equal to the ordinary two-sided tail $\mathcal{T}^\pm(\alpha)$. Soon this was obviated by a result of Schmidt [8], according to which α itself already has this property—and in fact $\mathcal{T}^\pm(\alpha) = \mathcal{F}^\pm_\phi(\alpha)$ for one-to-one ϕ taking values in any countable discrete group with finite conjugacy classes.

There are also some results that go in the opposite direction, in that they show how to recode a process so as to produce an isomorphic one with a highly nontrivial tail field. According to a striking result of Ornstein and Weiss ([6], see also [2]), every process has an isomorphic recoding for which the two-sided tail field $\mathcal{T}^\pm$ equals $\mathcal{B}$; thus even if the original process is K (hence "completely nondeterministic", at least in the one-sided sense), it can be recoded to an isomorphic process (which is of course still K) which is "two-sided deterministic":

THEOREM 1.2 ([6]). *Given a m.p.t. $T : X \to X$ on a nonatomic Lebesgue probability space and a finite measurable partition α of X, there is a refinement β of α such that $\mathcal{T}^\pm(\beta) \supset \alpha$. Thus if α is a generator, (β, T) is isomorphic to (α, T) and two-sided deterministic, in that $\mathcal{T}^\pm(\beta) = \mathcal{B}$.*

We can establish a one-sided analogue of this for the super-tails. (Of course a one-sided version involving the ordinary tails is not possible, since for K-processes, $\mathcal{T}^+(\beta)$ will be trivial for every β.)

PROPOSITION 1.3. *Given a m.p.t. $T : X \to X$ on a nonatomic Lebesgue probability space and a finite measurable partition α of X, there is a refinement β of α such that $\mathcal{F}^+(\beta) \supset \alpha$.*

Proof. This proof is actually very easy: we just take $\beta = \alpha \vee T^{-1}\alpha$, the recoding of the process (α, T) by 2-blocks (α 2-names).

Let us see that $\mathcal{F}^+(\beta) \supset \alpha$. Consider the finite directed graph whose vertices are the elements A_i of α and for which there is an edge from A_i to A_j if and only if $\mu(A_i \cap T^{-1}A_j) > 0$; these edges are naturally labeled by, indeed correspond to, the elements of β. We now claim that if we know, for a point x and some possibly very large n, the β-symbol count $v_n(x)$ from time 0 up to time n, and the β-symbol x_n at time n, then we can determine the cell of α to which x belongs (at time 0).

From looking at $v_m(x)$ we know how many times each edge of the graph has been traversed from time 0 up to time n, and hence we know how many times each vertex has been entered as well as how many times it has been left. We also know the vertex at which the path terminates. Thus there are two cases to consider. Either there is one vertex which has been departed from one more time than it has been entered, and this must then be the initial vertex; or else each vertex has been entered the same number of times as it has been left, and then the initial vertex must be the same as the final one. ∎

Remark 1.4. Unlike the situation with the ordinary tails, for the fine tails we can have $\mathcal{F}^+(\beta) \supsetneq \mathcal{F}^\pm(\beta)$: if α is the time-0 partition of a Bernoulli shift and $\beta = \alpha \vee T^{-1}\alpha$, then $\mathcal{F}^+(\beta) \supset \alpha$ by the preceding proposition, while $\mathcal{F}^\pm(\beta)$ is trivial [7].

2. Proof of Theorem 1.1

2.1. *Stability of probabilities of probable count vectors.* A key ingredient of the proof is the asymptotic local flatness of the symbol count distribution for Bernoulli shifts (a strong version of the Hewitt–Savage Zero-One Law), as expressed in the following lemma, in which $|s| = |s_1| + \cdots + |s_q|$ denotes the L^1 norm of a vector $s \in \mathbb{Z}^q$ and $s \cdot 1 = s_1 + \cdots + s_q$. The lemma asserts that a high-probability set of points ω have accumulated symbol count vectors $v_0^n(\omega)$ whose probabilities are fairly stable when the vector is translated by a bounded amount.

Lemma 2.1. *Fix a Bernoulli system $\mathcal{B}(p_1, \dots, p_q)$ with shift-invariant probability measure P and let $L \in \mathbb{N}$. Given $\varepsilon > 0$ there is $N_1 \in \mathbb{N}$ such that if $n \geq N_1$ then*

$$P\Big\{\omega : \textit{for all } s \in \mathbb{Z}^q \textit{ with } |s| \leq L, \quad \left| \frac{P\{\xi : v_0^n(\xi) = v_0^n(\omega)\}}{P\{\xi : v_0^{n+s\cdot 1}(\xi) = v_0^n(\omega) + s\}} - 1 \right| < \varepsilon \Big\} > 1 - \varepsilon. \tag{2.1}$$

Proof. The distribution of the symbol count vector v_0^n is given by the multinomial probabilities

$$P\{\xi : v_0^n(\xi) = (t_1, \ldots, t_q)\} = \frac{(n+1)!}{t_1! \cdots t_q!} p_1^{t_1} \cdots p_q^{t_q}, \tag{2.2}$$

$(t_i \geq 0,\ t_1 + \ldots + t_q = n+1)$. Assume for convenience that all $s_i \geq 0$ and define

$$a_0 = 0, \qquad a_j = \sum_{i=1}^{j} s_i \quad \text{for } j = 1, \ldots, q. \tag{2.3}$$

Then, abbreviating $v_j = v_0^n(\omega)(j)$ for $j = 1, \ldots, q$, the quotient of the probabilities appearing in the statement of the lemma is

$$\prod_{j=1}^{q} \frac{v_j + s_j}{(n+1) + a_j} \frac{v_j + s_j - 1}{(n+1) + a_j - 1} \cdots \frac{v_j + 1}{(n+1) + a_{j-1} + 1} \frac{1}{p_j^{s_j}}. \tag{2.4}$$

Now the result follows from the Weak Law of Large Numbers for Bernoulli processes, according to which for large enough n and for all ω not in a set of small measure, each $v_j/n \approx p_j$. The case when not necessarily all $s_i \geq 0$ is an immediate consequence. ∎

2.2. *Asymptotic conditional independence.* We would like to construct towers and code within them, taking precautions against the possible persistence of symbol count information from the beginnings of names. For example, even in an aperiodic subshift of finite type there might exist symbols a and b for which one cannot find a symbol z and blocks U and V such that the blocks aUz and bVz occur in the subshift and are permutations of one another. In such a case, knowing the symbol count v_0^n and x_n could determine x_0, thereby forcing nontriviality of $\mathcal{F}^+$ (see §4 of [7]). We want to recode so that this sort of thing cannot happen.

We want to construct a finite partition β which generates under T the same measure-preserving system as (α, T) and such that $\mathcal{F}^+(\beta) = \mathcal{T}^+(\beta) = \mathcal{T}^+(\alpha)$. Since β and α generate the same process, each has tail equal to the Pinsker algebra $\mathcal{P}(T)$ of that process: $\mathcal{T}^+(\beta) = \mathcal{T}^+(\alpha) = \mathcal{P}(T)$. So we need to show that $\mathcal{F}^+(\beta) \subset \mathcal{T}^+(\beta)$.

Recall that for finite partitions γ and η,

$$\gamma \perp^{\varepsilon} \eta \quad \text{means} \quad \sum_{G \in \gamma,\, N \in \eta} |\mu(G \cap N) - \mu(G)\mu(N)| < \varepsilon. \tag{2.5}$$

For σ-algebras $\mathcal{G}$ and $\mathcal{N}$,

(2.6) $\mathcal{G} \perp^{\varepsilon} \mathcal{N}$ means that $\gamma \perp^{\varepsilon} \eta$ for every finite partition γ by $\mathcal{G}$-measurable sets and η by $\mathcal{N}$-measurable sets;

equivalently (perhaps with a slightly different ε),

$$\mathcal{G} \perp^{\varepsilon} \mathcal{N} \quad \text{means} \quad |H(\mathcal{G}) + H(\mathcal{N}) - H(\mathcal{G} \vee \mathcal{N})| < \varepsilon. \tag{2.7}$$

Definitions when we condition on a third σ-algebra $\mathcal{F}$ are analogous:

$$\mathcal{G} \perp^{\varepsilon}_{\mathcal{F}} \mathcal{N} \quad \text{means} \quad |H(\gamma \,|\, \mathcal{F}) + H(\eta \,|\, \mathcal{F}) - H(\gamma \vee \eta \,|\, \mathcal{F})| < \varepsilon \tag{2.8}$$

for all finite $\mathcal{G}$-measurable partitions γ and finite $\mathcal{N}$-measurable partitions η. Definitions employing conditional expectations or measure disintegrations over $\mathcal{F}$ can also be stated, e.g. for all finite partitions γ and η as above,

$$\sum_{\substack{G \in \gamma \\ N \in \eta}} \| E(\chi_G \,|\, \mathcal{F}) E(\chi_N \,|\, \mathcal{F}) - E(\chi_G \chi_N \,|\, \mathcal{F}) \|_1 < \varepsilon. \tag{2.9}$$

Write $v_0^n(\beta)$ for the partition generated by the β-symbol counts $v_0^n(x)$ and, as usual, $\beta_n^m = \bigvee_{k=n}^m T^{-k}\beta$.

Lemma 2.2. *Suppose that a generating partition β has the following asymptotic independence property*:

(2.10) *given $k \in \mathbb{N}$ and $\varepsilon > 0$, there is $N \in \mathbb{N}$ such that $n \geq N$ implies that $\beta_{-k}^k \perp^{\varepsilon}_{\beta_n^\infty} v_0^n(\beta)$,*

or equivalently,

$$H(v_0^n(\beta) \,|\, \beta_n^\infty) - H(v_0^n(\beta) \,|\, \beta_{-k}^k \vee \beta_n^\infty) < \varepsilon \quad \textit{if } n \geq N. \tag{2.11}$$

Then

$$\mathcal{F}^+(\beta) \subset \mathcal{P}(T) = \mathcal{T}^+(\beta). \tag{2.12}$$

Proof. It is known (see [9]) that every system is relatively K over its Pinsker factor:

$$\text{if } n \geq N \text{ is large enough then } \beta_{-k}^k \perp^{\varepsilon}_{\mathcal{P}(T)} \beta_n^\infty. \tag{2.13}$$

This is true because $\beta_n^\infty \searrow \mathcal{P}(T)$ implies that for large enough n,

$$H(\beta_{-k}^k \,|\, \mathcal{P}) - H(\beta_{-k}^k \,|\, \beta_n^\infty \vee \mathcal{P}) < \varepsilon. \tag{2.14}$$

Combining (2.10) and (2.13), for large enough n we will have

$$\beta_{-k}^k \perp^{2\varepsilon}_{\mathcal{P}(T)} (\beta_n^\infty \vee v_0^n(\beta)). \tag{2.15}$$

Since

$$\mathcal{F}^+(\beta) = \bigcap_{n=0}^\infty (\beta_n^\infty \vee v_0^n(\beta)), \tag{2.16}$$

this latter statement implies that

$$\beta_{-\infty}^\infty \perp_{\mathcal{P}(T)} \mathcal{F}^+(\beta), \tag{2.17}$$

and hence

$$\mathcal{F}^+(\beta) \subset \mathcal{P}(T) = \mathcal{T}^+(\beta). \quad \blacksquare \tag{2.18}$$

Thus our goal is produce a partition β that satisfies (2.10).

2.3. *New alphabets.* By hypothesis the system (X, T, μ) has a positive-entropy direct Bernoulli factor (B, σ, P), so that

$$(X, T, \mu) \approx (Y, S, \nu) \times (B, \sigma, P). \tag{2.19}$$

Let γ be a finite generating partition for (Y, S) and ϱ the (independent) time-0 partition of the Bernoulli factor (B, σ). Trying to avoid unnecessarily complicated notation, we regard elements of a partition, which are sets, also as symbols comprising the alphabet of the associated symbolic system; and usually the alphabets will be $\{1, \dots, n\}$ for some $n \in \mathbb{N}$. Thus each $x \in X$ is a string $(x_k) = (y_k, \omega_k)$, $k \in \mathbb{Z}$, with $y_k \in \gamma$ and $\omega_k \in \varrho$ for all k. We denote the length of any block C by $|C|$, and for a sequence x and an interval $I = [i, i+1, \dots, i+l]$, we denote the length of I by $|I|$ and the block $x_i x_{i+1} \dots x_{i+l}$ by x_I.

First we need a new alphabet, $\beta_0 = \{1, \dots, q\}$, for Y, with q large enough that, with a value of t to be determined below, for all $i = 0, 1, \dots, q|\varrho| - 1$ each γ-name of length $i + tq|\varrho|$ can be matched to a different permutation of a *single* β_0-name $E = (1^t\, 2^t \dots q^t)^{|\varrho|}$ in which each symbol appears the same number of times. And we will not use arbitrary permutations of E, but concatenations of permutations of selected sub-blocks, matching γ-names to β_0-names of the form

$$E' = (1^t\, 2^t \dots q^t)'(1^t\, 2^t \dots q^t)' \dots (1^t\, 2^t \dots q^t)', \tag{2.20}$$

where the primes indicate arbitrary, possibly different, permutations of the blocks to which they are applied.

Using Stirling's Formula, for large t a block of length qt on t of each of q symbols has on the order of $q^{(qt+1/2)} t^{(1-q)/2}$ permutations. Thus if q and t are large enough (say choose $q \gg 4|\gamma \times \varrho|$ and then t very large), we will have

$$(q^{qt+1/2} t^{(1-q)/2})^{|\varrho|} > |\gamma \times \varrho|^{(t+1)q|\varrho|}, \tag{2.21}$$

which is what we will need to be able to set up our one-to-one correspondences between γ-blocks and special β_0-blocks.

Our new alphabet for the system $Y \times B$ will be

$$\beta = (\beta_0 \times \varrho) \cup \{(0,0)_i = (0_i, 0_i) : i = 0, 1, \dots, q|\varrho| - 1\}. \tag{2.22}$$

The symbols $(0,0)_i$ are fillers for residual blocks after "free intervals" are cut into sub-blocks of length $q|\varrho|$.

2.4. *Correspondences between blocks.* Having fixed a new alphabet $\beta_0 = \{1, \dots, q\}$ for Y with q large enough, we set up two correspondences.

First, denote by $\mathcal{F}_1$ the family of all β_0-blocks

$$E' = [\tau_1(1^t\, 2^t \dots q^t)][\tau_2(1^t\, 2^t \dots q^t)] \dots [\tau_{|\varrho|}(1^t\, 2^t \dots q^t)], \tag{2.23}$$

where $\tau_1, \dots, \tau_{|\varrho|}$ are permutations acting on blocks of length tq. We have chosen q and t large enough that there exists a one-to-one function

$$\phi_1 : \bigcup_{i=0}^{q|\varrho|-1} \gamma^{i+tq|\varrho|} \to \mathcal{F}_1. \tag{2.24}$$

Next, let us enumerate the alphabet of $\beta_0 \times \varrho$ as $\{g_0, \dots, g_{q|\varrho|-1}\}$, define

$$H = g_0 \dots g_{q|\varrho|-1}, \tag{2.25}$$

so that H contains exactly one of each $\beta_0 \times \varrho$-symbol, and let $\mathcal{F}_2$ denote the set of all permutations of H. Our choice of q is large enough that there exists a one-to-one function

$$\phi_2 : (\gamma \times \varrho)^{q|\varrho|} \to \mathcal{F}_2. \tag{2.26}$$

2.5. *The marker block in B.* We form a special marker block W over the alphabet ϱ of the Bernoulli factor B:

$$W = 1^{tq}\, 2^{tq} \dots |\varrho|^{tq}. \tag{2.27}$$

Notice that this block has length $|W| = tq|\varrho|$ and that it consists of a concatenation of strings of consecutive repetitions, tq times, of the symbols of the ϱ-alphabet in order. In particular, W cannot overlap itself in any sequence in B (no prefix of W equals any suffix of W).

We use the marker W to decompose the index set $\mathbb{Z}$ into intervals of two kinds. An interval of coordinates $[j, \dots, j + l(W) - 1]$ across which W appears in $\omega \in B$ will be called a *marked interval*, and the remaining places $[j + l(W), \dots, j + m - 1]$ before the next appearance of W will be called a *free interval*. Numbering these intervals consecutively so that M_0 is the first marked interval that includes nonnegative numbers and F_j is the free interval immediately to the right of M_j for all j, we have $\mathbb{Z} = \bigcup_j (M_j \cup F_j)$.

2.6. *The idea behind the coding.* Our coding will be accomplished by working on each $M_j \cup F_j$ separately. On M_j the B-coordinates will not be changed, and the Y-coordinates will be changed by using ϕ_1 (taking into account also a few (i) extra entries in Y), in such a way that every $\beta_0 \times \varrho$-symbol appears the *same* number of times across M_j—see Property 1 below. The free interval F_j is cut into subintervals of length $q|\varrho|$, on each of which the $\gamma \times \varrho$-name is changed by applying ϕ_2, thereby changing both the Y and B entries. If $|F_j| > 0$, we add one extra $\beta_0 \times \varrho$-symbol, g_i, with $i = |F_j| \bmod q|\varrho|$, plus enough extra filler symbols $(0_i, 0_i)$ to make up the rest of $|F_j|$ (see Property 2). Note that the extra symbol g_i depends only on the length of F_j, which is determined by the appearances of W in the sequence ω in B.

The two different coding mechanisms just described work as follows. Any information residing in β-symbol counts across marked intervals is removed by making the $\beta_0 \times \varrho$-symbol count vector across each marked interval a constant vector. Across a union of free intervals the $\beta_0 \times \varrho$-symbol count has for excess from a constant vector (multiple of $(1, 1, \ldots, 1)$) a function of a finite-state Bernoulli process (W', σ'_W, P'_W) (described in the next section), and the count of the filler symbols $(0,0)_i$ is also a function of this process. This will allow us to show that the pair symbol count across a long interval that is an exact union of marked and free intervals is asymptotically flat.

If we start and stop our symbol counting at times interior to marked or free intervals, we will obtain a (pair) β-symbol count that is a vector translate of a symbol count across a complete union of marked and free intervals $M_j \cup F_j$. With high probability, W appears with bounded gap, so the norm of this translate is bounded, as is the norm of the translate for the corresponding symbol counts of the finite-state Bernoulli process (W', σ'_W, P'_W), and Lemma 2.1 applies. Thus as long as $\sum |F_j|$ is sufficiently large, which will be the case if we have hit W enough times, i.e. if we are summing over a long enough interval of coordinates, the distribution of *all* such symbol counts will be approximately flat.

2.7. *The countable-alphabet process in the gaps and its clumping.* In the Bernoulli factor (B, σ, P), the first-return map $\sigma_W : [W] \to [W]$ to the cylinder set $[W]$, together with the normalized measure $P_W = P/P[W]$, is isomorphic to a Bernoulli system on the countable alphabet ϱ_W consisting of all ϱ-blocks WU, with U a ϱ-block (possibly the empty one ε) not containing W. This can be checked as follows. Write each $\omega \in [W]$ as

$$\omega = \ldots W\omega_{F_{-1}}.W\omega_{F_0}W\omega_{F_1}W\omega_{F_2}\ldots, \tag{2.28}$$

with each $W\omega_{F_j} \in \varrho_W$, so that

$$\sigma_W\omega = \ldots W\omega_{F_0}.W\omega_{F_1}W\omega_{F_2}W\omega_{F_3}\ldots. \tag{2.29}$$

Recall that W does not overlap itself, so there is no confusion in this representation. The probability of a symbol WU_i in this first-return system is $P(WU_iW)/P(W) = P(WU_i)$, while the probability of a block $WU_{i_1}\ldots WU_{i_r}$ is $P_W(WU_{i_1}\ldots WU_{i_r}W) = \prod_{i=1}^r P(WU_i)$, so the independence of cylinder sets with nonoverlapping coordinate ranges follows.

We define a surjective map $\psi : \varrho_W \to \beta_0 \times \varrho$, i.e. a clumping of this countable alphabet onto the main part $\beta_0 \times \varrho$ of the finite alphabet β, as follows. We have fixed a numbering $\beta_0 \times \varrho = \{g_i : i = 0, \ldots, q|\varrho| - 1\}$. Let $\psi(W\varepsilon) = (0,0)_0$. For each nonempty ϱ-block V that does not contain W, let

$$\psi(WV) = g_i, \qquad \text{with } i \equiv |V| \bmod q|\varrho|. \tag{2.30}$$

Define $\Psi(\omega)_0 = \psi(W\omega_{F_0})$. Then Ψ determines a factor map from the countable-state Bernoulli system $([W], \sigma_W, P_W)$ onto a finite-state Bernoulli system (W', σ'_W, P'_W).

For a free interval F, the extra symbol that will occur in the recoding of the string $(y,\omega)_F$ will then be $\psi(W\omega_F)$ (for $F \neq \varepsilon$).

2.8. *The recoding and the new partition β.* Consider M_jF_j as above, a marked interval followed by a free interval, and let us suppress the subscript j. Let $d = |F|/(q|\varrho|)$ and $i = |F| \bmod q|\varrho|$. Regarding F as a string of integers (as well as an interval), factor F into a concatenation $F = F_0F_1 \dots F_d$, with $|F_0| = i$ and $|F_r| = q|\varrho|$ for $r = 1, \dots, d$. Define $g_0\,(0_0, 0_0)^{-1} = \varepsilon$, the empty block. Our recoding of the $\gamma\times\varrho$-block across MF to a β-block across MF is defined by

$$\begin{aligned} x_M x_F &= (y_M, \omega_M)(y_{F_0}, \omega_{F_0})\,(y_{F_1}, \omega_{F_1}) \dots (y_{F_d}, \omega_{F_d}) \\ &\to [\phi_1(y_M y_{F_0}), \omega_M)][g_i\,(0_i, 0_i)^{i-1}][\phi_2(y_{F_1}, \omega_{F_1})] \dots [\phi_2(y_{F_d}, \omega_{F_d})]. \end{aligned} \tag{2.31}$$

Applying this procedure on each M_jF_j defines a map

$$\phi : (\gamma \times \varrho)^{\mathbb{Z}} \to \beta^{\mathbb{Z}}, \qquad \phi(y,\omega) = (\widetilde{y}, \widetilde{\omega}). \tag{2.32}$$

Note that no new appearances of W can be created in the B coordinate, and, as stated earlier, previous appearances of W are preserved. The symbols $(0,0)_0$ and $(0,0)_1$ are not used, but that is all right.

The recoding is shift-invariant and determines a partition β of $X = Y\times B$ according to the time-0 symbol. The original $\gamma \times \varrho$-sequence is recoverable from the β-coding since our correspondences ϕ_1 and ϕ_2 are one-to-one, so β generates the full σ-algebra of $Y \times B$ under $S \times \sigma$.

2.9. *Properties of the recoded system.* The recoding has been constructed so as to have the following properties:

Property 1. Across each marked interval M in a recoded point $(\widetilde{y}, \widetilde{\omega})$, each β-symbol appears the same number r of times, except for the special symbols $(0,0)_i$, which do not appear at all.

Property 2. Across each free interval F in a recoded point $(\widetilde{y}, \widetilde{\omega})$, each β-symbol appears the same number $r'(|F|)$ of times—*except* that if $|F| > 0$ and $i = |F| \bmod |\varrho|q$, then the special symbol $\psi(W\omega_F) = g_i$ appears $r'(|F|) + 1$ times and the filler symbol $(0,0)_i$ appears $i-1$ times.

For a sequence $\omega \in B = \varrho^{\mathbb{Z}}$ and an interval $[k,k'] \subset \mathbb{Z}$, denote by $v^{\psi}_{[k,k']}(\omega)$ the vector that counts the W'-symbols $\psi(\omega_F)$ as F runs through the free subintervals of $[k,k']$.

Property 3. Consider a long interval $[k,k'] \subset \mathbb{Z}$ and points $x = (y,\omega)$ for which $\omega_{[k,k']}$ begins and ends with the marker W. Each β-symbol count $v_{[k,k']}(x)$ (which is actually a function only of ω) uniquely determines the W'-symbol count $v^{\psi}_{[k,k']}(\omega)$, and vice versa.

Proof. Given a β-symbol count v across this interval, for each $i = 1, \ldots, q|\varrho| - 1$ the number $v((0,0)_i)$ of appearances of the filler symbol $(0,0)_i$ is a multiple of $i-1$, and for $i = 2, \ldots, q|\varrho|-1$ the quotient $n_i = v((0,0)_i)/(i-1)$ gives the number of times that g_i was used as the "extra" symbol in recoding $x_{[k,k']}$ (these entries constitute what we call the "excess vector"). Thus the difference $b(i) = v_{[k,k']}(x)(g_i) - v((0,0)_i)/(i-1)$ is constant in $i = 2, \ldots, q|\varrho|-1$. Moreover, g_0 is never used as an "extra" symbol, so automatically also $v(g_0) = b(i)$, $i = 2, \ldots, q|\varrho| - 1$. If we also define $b(1)$ to take this same constant value, then the vector b gives the constant "base" count of the $\beta_0 \times \varrho$-symbols which does not include the special symbols $\psi(\omega_F)$. Therefore

$$v(g_i) - b(i) = v^{\psi}_{[k,k']}(\omega)(i), \qquad i = 0, \ldots, q|\varrho| - 1, \tag{2.33}$$

the count of the "extra" symbols $\psi(\omega_F)$ over all free subintervals F of $[k, k']$.

Conversely, a symbol count $v^{\psi}_{[k,k']}(\omega)$ of the $\psi(\omega_F)$ over all free subintervals F of $[k, k']$ specifies the number n_i of times that each length congruence class $i = 1, \ldots, q|\varrho|-1$ mod $q|\varrho|$ appears among the nonempty free intervals, equivalently the number of each "extra" g_i, $i = 1, \ldots, q|\varrho|-1$. The n_i are the entries in the corresponding "excess" β-symbol count vector, and they also determine the number $n_i(i-1)$ of appearances of each filler symbol $(0,0)_i$, $i = 1, \ldots, q|\varrho|-1$. The remaining entries in the β-symbol count $v_{[k,k']}(x)$ are apportioned equally among all the $\beta_0 \times \varrho$-symbols (the number of entries still undetermined being necessarily divisible by $q|\varrho|$). ∎

Property 4. Translation of a β-symbol count vector v by at most L in each entry produces a translation in the corresponding W'-symbol count vector $f(v)$ by at most $3L$ in each entry. (In this setting it is more convenient to use the L^∞ norm rather than the equivalent L^1 norm on $\mathbb{Z}^d$ in connection with Lemma 2.1.)

2.10. *How to verify asymptotic conditional independence.* Now we will verify that the property (2.10) holds for β. Let $\varepsilon > 0$, $\delta \ll \varepsilon$, and first of all choose K such that the columns of the return tower over W that have height less than or equal to K cover all but δ of X:

$$\text{if} \quad B_K = \pi_B^{-1}\Big[\bigcup_{j=0}^{K} \sigma^{-j}W \cap \bigcup_{j=0}^{K} \sigma^{j}W\Big], \quad \text{then} \quad \mu(B_K) > 1 - \delta. \tag{2.34}$$

Apply Lemma 2.1 to the Bernoulli process (W', σ'_W, P'_W) (the clumping of blocks across free intervals) with the translate bound $L = 2K + 2k + 1$ to find an N_1 such that if $n \geq N_1$, then most symbol count vectors (i.e., coming from a set $B(\delta, L)$ of $\omega \in B$ of measure greater than $1 - \delta$) resulting from any n observations $\psi(W\omega_{F_0}), \ldots, \psi(W\omega_{F_{n-1}})$ have very nearly the same probability (their quotient is within a distance δ of 1) as their translates by vectors of size no more than $3L$.

Choose N large enough to ensure that for most $x \in X$ (all but a set of measure less than δ), within the time interval $[0, N]$ the marker W has been hit at least N_1 times, in fact that we have encountered at least N_1 nonempty free intervals. Let $n \geq N$.

NOTATION. 1. Abbreviate

$$\mathcal{A} = v_0^n(\beta), \quad \mathcal{B} = \beta_{-k}^k, \quad \mathcal{C} = \beta_n^\infty. \tag{2.35}$$

2. Write $a \approx_\delta b$ to mean that $|a/b - 1| < \delta$, equivalently (for $a, b > 0$) $(1-\delta)b < a < (1+\delta)b$ (a form well adapted for summing over a and b).

Formula (2.10) will follow if we can show that conditioned on $\mathcal{C}$ most symbol counts have probabilities that are stable under small changes of the symbols being fixed by $\mathcal{B}$ and $\mathcal{C}$. For this purpose we use Rokhlin's theory of Lebesgue spaces, complete sub-σ-algebras, and the corresponding factor spaces or partitions and disintegrations of measures. We want to show that for each cell C of the (Rokhlin) partition corresponding to the σ-algebra $\mathcal{C}$, with corresponding disintegrated measure μ_C, for a set of cells A of $\mathcal{A}$ forming a set of measure greater than $1-\delta$, for any cell $R \in \mathcal{B}$,

$$\mu_C(A \mid R) \approx_\delta \mu_C(A). \tag{2.36}$$

2.11. *Fixing some more coordinates.* We have to refine the partitions involved in order to keep track of the various possible strings that can appear between time k and the time u_1 of the next entrance to W, as well as strings starting at the time u_l of the last complete appearance of W before time n and ending at time $n-1$. (These "edge" coordinates $[k+1, u_1 - 1] \cup [u_l + l(W) + 1, n-1]$ may include parts of marked intervals as well as free intervals.)

Define

$$\begin{aligned} u_1(\omega) &= \inf\{i > k : \sigma^i \omega \in W\}, \\ u_l(\omega) &= \sup\{i \leq n - l(W) : \sigma^i \omega \in W\}. \end{aligned} \tag{2.37}$$

Fix $k_1 \in (k, k+K]$ and $k_l \in [n - l(W) - K, n - l(W)]$ and let

$$\begin{aligned} \Omega_{k_1,k_l} &= B_K \cap \pi_B^{-1}\{\omega : u_1(\omega) = k_1,\ u_l(\omega) = k_l\}, \\ \mathcal{D}_{k_1,k_l} &= \beta_{k+1}^{k_1-1} \vee \beta_{k_l+l(W)+1}^{n-1}. \end{aligned} \tag{2.38}$$

LEMMA 2.3. *To prove formula* (2.10) ($\mathcal{A} \perp_{\mathcal{C}}^{\varepsilon} \mathcal{B}$), *it is enough to show that there is $\delta > 0$ such that for each cell C of $\mathcal{C}$, for a large-measure set of atoms A of $\mathcal{A}$, for each k_1 and k_l and each choice of cells $D_1, D_2 \in \mathcal{D}_{k_1,k_l}$, and $R_1, R_2 \in \mathcal{B}$,*

$$\mu_C(A \mid \Omega_{k_1,k_l} \cap R_1 \cap D_1) \approx_\delta \mu_C(A \mid \Omega_{k_1,k_l} \cap R_2 \cap D_2). \tag{2.39}$$

Proof. From the hypothesis (2.39) it follows that

$$\mu_C(A\cap \Omega_{k_1,k_l}\cap R\cap D)\approx_\delta \mu_C(A)\mu_C(\Omega_{k_1,k_l}\cap R\cap D) \quad \text{for all } R,D. \tag{2.40}$$

Sum on $D\in\mathcal{D}_{k_1,k_l}$ to conclude that for most $A\in\mathcal{A}$ and all $R\in\mathcal{B}$,

$$\mu_C(A\cap \Omega_{k_1,k_l}\cap R)\approx_\delta \mu_C(A)\mu_C(\Omega_{k_1,k_l}\cap R). \tag{2.41}$$

Finally, sum over all k_1,k_l to conclude that for most $A\in\mathcal{A}$ and all $R\in\mathcal{B}$,

$$\mu_C(A\cap R)\approx_\delta \mu_C(A)\mu_C(R), \tag{2.42}$$

and hence (if δ is small enough and "most" is enough)

$$\mathcal{A}\perp^{\varepsilon}_{\mathcal{C}}\mathcal{B}. \ \blacksquare \tag{2.43}$$

2.12. *Proof of formula* (2.39). The key idea here is that changing the edge conditions (symbol counts over the intervals $[k+1,u_1-1]\cup[u_l+l(W)+1,n-1]$) translates interior symbol counts (over times $[u_1,u_l+|W|]$).

We define a set of "good points" as follows. The good points x are among those in the large-probability set B_K whose images under π_B hit the set W in the Bernoulli factor B within K steps in both forward and backward time and very many times (way more than N_1) during the interval from 0 to n. We also demand that $x\in\pi_B^{-1}B(\delta,L)$, so that $\omega=\pi_Bx$ has a good (stable) symbol count for the clumped process across free intervals.

Fix a symbol count $v_0^n(x)=c_A$ of such a "good" point, with A the corresponding atom of $\mathcal{A}$, and the times k_1,k_l of the first and last complete appearances of W between times k and n. Let us determine the relative probabilities of such a symbol count, given $R_1\cap D_1\in\mathcal{B}\vee\mathcal{D}_{k_1,k_l}$, versus given $R_2\cap D_2$, with respect to the measure μ_C.

Let $I=[k_1,k_l+l(W)]$ denote the "interior" range of coordinates, consisting of the set of indices made up of full passes through marked and free intervals in the interval $[0,n]$. A key point is that replacing R_1,D_1 by R_2,D_2 changes at most $2K+2k+1$ coordinates in the β-name of x, and if we are to preserve the symbol count c_A across the interval $[0,n]$, the symbol count across I must be translated by a corresponding vector whose norm is bounded by $2K+2k+1$.

For $i=1,2$ let v_i denote the β-symbol count vector across the set of indices $[0,n]\setminus I$ determined by each point in $\Omega_{k_1,k_l}\cap R_i\cap D_i$. Recall that our recoding is done between occurrences of the marker W, so the β-symbol count across the interval I, which is just a function of the direct Bernoulli factor B, is independent of all the original $\gamma\times\varrho$-symbols, and hence of all the new β-symbols, on any range of coordinates disjoint from I. This independence also holds conditioned on C and on Ω_{k_1,k_l}. Therefore,

$$\begin{aligned}(2.44)\qquad \mu_C(A\cap R_1\cap D_1\mid \Omega_{k_1,k_l}) &= \mu_C(\{v_I=c_A-v_1\}\cap R_1\cap D_1\mid \Omega_{k_1,k_l})\\ &= \mu_C(\{v_I=c_A-v_1\}\mid \Omega_{k_1,k_l})\mu_C(R_1\cap D_1\mid \Omega_{k_1,k_l}),\end{aligned}$$

and hence

$$(2.45)\qquad \frac{\mu_C(A\cap R_1\cap D_1\cap \Omega_{k_1,k_l})}{\mu_C(R_1\cap D_1\cap \Omega_{k_1,k_l})} = \mu_C(\{v_I=c_A-v_1\}\mid \Omega_{k_1,k_l}).$$

Referring to Property 3 and denoting by f_I the $\psi(\varrho_W)$-symbol count for the finite-state Bernoulli process W' determined by the ϱ-blocks appearing across the free subintervals of I, we see that $\Omega_{k_1,k_l}\cap\{v_I=c_A-v_1\} = \Omega_{k_1,k_l}\cap\{f_I=f(c_A-v_1)\}$, and similarly for v_2. Since c_A-v_1 and c_A-v_2 are translates of one another by vectors of L^∞ size no more than L, it follows that $f(c_A-v_1)$ and $f(c_A-v_2)$ are translates of one another by vectors of size no more than $3L$ (see Property 4). Therefore we may apply the flatness of symbol counts for the process (W',σ'_W,P'_W) given by Lemma 2.1, and again using independence as above, we can complete this calculation as follows:

$$\begin{aligned}(2.46)\qquad \mu_C(A\mid R_1\cap D_1\cap \Omega_{k_1,k_l}) &= \mu_C(\{v_I=c_A-v_1\}\mid \Omega_{k_1,k_l})\\ &\approx_\delta \mu_C(\{v_I=c_A-v_2\}\mid \Omega_{k_1,k_l}) = \mu_C(A\mid R_2\cap D_2\cap \Omega_{k_1,k_l}),\end{aligned}$$

proving (2.39).

3. Questions. 1. Will the conclusion of Theorem 1.1 hold if the hypothesis of existence of a direct Bernoulli factor is removed?

2. Can our recoding be accomplished in a *unilateral* way? If so, every exact endomorphism with a finite generator would have a finite super-K generator (cf. [7, 8]).

3. Is $\mathcal{F}^+(\alpha)$ trivial if and only if $\mathcal{F}^-(\alpha)$ is trivial? Of course the analogous result for $\mathcal{T}^+$ and $\mathcal{T}^-$ is true.

4. Does there exist a K-system for which *every* generator β has $\mathcal{T}^\pm(\beta)=\mathcal{B}$? If so, exactly which systems have $\mathcal{T}^\pm(\beta)=\mathcal{B}$ (equivalently $\mathcal{F}^\pm(\beta)=\mathcal{B}$) for every generator β?

5. Is the set of super-K^+ partitions first category in every system? (From Proposition 1.3 it follows that the set of non-super-K^+ partitions is dense.)

Acknowledgments. We thank the referee for criticisms and suggestions that led to many improvements.

REFERENCES

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

[2] B. M. Gurevič [B. M. Gurevich], *On one-sided and two-sided regularity of stationary random processes*, Soviet Math. Dokl. 14 (1973), 804–808; transl. of Dokl. Akad. Nauk SSSR 210 (1973), 763–766.

[3] C. Hoffman, *A K counterxample machine*, Trans. Amer. Math. Soc. 351 (1999), 4263–4280.

[4] S. A. Kalikow, T, T^{-1} *transformation is not loosely Bernoulli*, Ann. of Math. (2) 115 (1982), 393–409.

[5] D. S. Ornstein and P. C. Shields, *An uncountable family of K-automorphisms*, Adv. Math. 10 (1973), 63–88.

[6] D. S. Ornstein and B. Weiss, *Every transformation is bilaterally deterministic*, Israel J. Math. 21 (1975), 154–158.

[7] K. Petersen and K. Schmidt, *Symmetric Gibbs measures*, Trans. Amer. Math. Soc. 349 (1997), 2775–2811.

[8] K. Schmidt, *Tail-fields of products of random variables and ergodic equivalence relations*, Ergodic Theory Dynam. Systems 19 (1999), 1325–1341.

[9] J.-P. Thouvenot, *Une classe de systèmes pour lesquels la conjecture de Pinsker est vraie*, Israel J. Math. 21 (1975), 208–214.

Department of Mathematics
University of North Carolina
CB 3250, Phillips Hall
Chapel Hill, NC 27599 U.S.A.
E-mail: petersen@math.unc.edu
http://www.math.unc.edu/Faculty/petersen

Laboratoire de Probabilités
Université Pierre et Marie Curie
4 place Jussieu, 75252 Paris, France

Received 14 July 2003;
revised 22 March 2004 (4363)