ISRAEL JOURNAL OF MATHEMATICS 79 (1992), 33–45

FINITARILY DETERMINISTIC GENERATORS FOR ZERO ENTROPY SYSTEMS

BY

STEVEN KALIKOW

Department of Mathematics, Cornell University, Ithaca, N.Y., U.S.A.

AND

YITZHAK KATZNELSON

Department of Mathematics, Stanford University Stanford, CD 94305-2125, U.S.A.

AND

BENJAMIN WEISS

Department of Mathematics, Hebrew University, Jerusalem, Israel

ABSTRACT

Zero entropy processes are known to be deterministic—the past determines the present. We show that each is isomorphic, as a system, to a finitarily deterministic one, i.e., one in which to determine the present from the past it suffices to scan a finite (of random length) portion of the past. In fact we show more: the finitary scanning can be done even if the scanner is noisy and passes only a small fraction of the readings, provided the noise is independent of our system.

The main application we present here is that any zero entropy system can be extended to a random Markov process (namely one in which the conditional distribution of the present given the past is a mixture of finite state Markov chains). This allows one to study zero entropy transformations using a procedure completely different from the usual cutting and stacking.

Received March 3, 1991

1. Introduction

Consider an irrational rotation of a circle. Partition the circle into the upper half, A, and the lower half, B. If a point is selected randomly on the circle and moved back and forth in time in accordance with this irrational rotation we get a doubly infinite stationary process in A's and B's. The process has zero entropy because the past determines the present. However, the process is better than merely zero entropy. The past determines the present in a finitary manner (as we look back in time, after looking only finitely far back, we know the present). Moreover, if only a portion of the past is available—on account of random noise which eradicates an arbitrarily large proportion of the readings—one can still determine the present in terms of the available information finitarily. We refer to this property as *FDN* (Finitarily Deterministic under independent Noise); we prove that every process of zero entropy is isomorphic to one which is *FDN*.

The applications we present deal with random Markov processes. In [1], two seemingly different kinds of processes, **random Markov processes** and **uniform martingales** are defined. For the sake of completeness we repeat the definitions below. Because of the way random Markov processes are constructed, and the very nice property of uniform martingales, it is of interest to compare ergodic processes with them.

Definition of a Uniform Martingale. A uniform martingale is a stationary process on a finite alphabet Q, $\ldots x_{-2}x_{-1}x_0x_1x_2\ldots$, each $x_i \in Q$, such that

$$\forall\, \varepsilon > 0\; \exists\, N\colon \forall\, n > N, \| P(x_0 = q \| x_{-1}x_{-2}\ldots x_{-n}) - P(x_0 = q \| x_{-1}x_{-2}\ldots) \,| < \varepsilon.$$

for all $q, x_{-1}, x_{-2}\ldots$. In other words, a stationary process $\{x_i\}_{i\in \mathbf{Z}}$ is a uniform martingale if the martingale

$$P(x_0 = q\,|x_{-1}), P(x_0 = q\,|x_{-1}x_{-2}), P(x_0 = q\,|x_{-1}x_{-2}x_{-3}), \ldots$$

converges uniformly on all sequences $x_{-1}, x_{-2}, \ldots$.

Doeblin and Fortet [3] introduced the notion of a g-function. Keane [4] introduced the notion of a"continuous" g-function, which is the same as a uniform martingale. He showed that under certain conditions g-functions have unique measures which are mixing. Berbee [2] developed general uniqueness results. Petit [5] extended Keane's work, where Keane's "continuous" notion was replaced by differentiable, and under certain conditions was able to show that the measures obtained were weak Bernoulli. The new idea in [1] was that of random Markov processes.

DEFINITION OF A RANDOM MARKOV PROCESS. An random Markov process is a stationary process $\{x_n\}$ on a finite alphabet Q with the property that the transition probabilities

$$P(x_0 = q \mid x_{-1}, x_{-2}, \dots x_{-n}, \dots)$$

can be calculated on a random mixture of k-order Markov processes. This means that for some distribution π_i on the positive integers and transition probabilities $t_n(q\colon x_{-1}, \dots x_{-n})$ we can write

$$P(x_0 = q \mid x_{-1}, x_{-2}, \dots x_{-n}, \dots) = \sum_{i=1}^{\infty} \pi_i t_i(q\colon x_{-1}, \dots x_{-i})$$

In [1] it was shown that these two notions coincide. The first is an intrinsic property of the process whereas the second, giving a concrete representation which in a sense finitizes the infinity in the usual conditioning on the infinite past, is not canonical. Processes can of course have many represenations as random Markov processes. We will actually work (formally) with the uniform martingale characterization but the ideas of a random Markov process guided us in finding the proof of our main result. Several open problems will be mentioned in the final section.

This is what we know about the structure of the class of uniform martingales:

(1) With the exception of finite state rotations, no zero entropy ergodic process is isomorphic to a uniform martingale [1].

(2) Every zero entropy ergodic transformation can be extended to an uniform martingale.

The reader should be able to prove (1) himself; (2) is the subject of this paper. If the reader does not wish to read such a difficult proof immediately, the proof that there is a uniform martingale which is K and not Bernoulli in [1] is the basic idea of (2) in this paper, and much simpler.

Here is an example of how this result could be used. It is known that the question of whether or not twofold mixing implies threefold mixing in zero entropy is equivalent to the question of whether or not twofold mixing implies threefold mixing in general. Using this paper, we can show that whether or not twofold mixing implies threefold mixing for uniform martingales is equivalent to whether or not twofold mixing implies threefold mixing in general. Suppose T is a twofold mixing and not threefold mixing zero entropy transformation. Let $\hat{T}$ be its uniform martingale extension constructed in this paper. Because of the way $\hat{T}$ is constructed, it is easily seen that $\hat{T}$ is mixing also. However, since $\hat{T}$ is an extension of $T, \hat{T}$ will also fail to be threefold mixing. This reduces the twofold threefold mixing problem to uniform martingales. In §2 we prove our result about FDN's, in §3 we will prove (2) while §4 is devoted to some concluding remarks.

2. Finitarily Deterministic Generators for Zero Entropy

In a zero entropy process, the past determines the present. However, sometimes one does not have to know the entire past to get the present. We now define a situation where we need only a finitary part of an arbitrary thin subsequence of the past to get the present.

Definition: A zero entropy stationary process, $X_0, X_{-1}, X_{-2}, \ldots$ is said to be **Finitarily Deterministic under random Noise** (abbreviated *FDN*) if for every $\varepsilon > 0$, the process $(X_0, a_0), (X_{-1}, a_{-1}), \ldots$ where the a's are independent of each other and of the X_i's, each $a_i \in \{0,1\}$, and each $a_i = 1$ with probability ε, has the property that, with probability 1, there exists n such that just the values of $\{X_i\}$ such that $a_i = 1$ and $-n \le i \le -1$ are sufficient to determine X_0.

The term process is used nowadays in several (related) contexts and in order to avoid possible confusion we make the following distinction: a **measure-preserving system** is a quadruple $(\Omega, \mathcal{B}, \mu, T)$ with $(\Omega, \mathcal{B}, \mu)$ a probability measure space and T an invertible, measurable, μ-preserving transformation on $(\Omega, \mathcal{B})$. A **stationary process** (on a finite alphabet $\mathcal{P}$) is a sequence of P-valued random variables with stationary joint distribution. A measure preserving system $(\Omega, \mathcal{B}, \mu, T)$ with a finite partition $\mathcal{P}$ of $(\Omega, \mathcal{B})$ defines a stationary process (with alphabet $\mathcal{P}$) by: $X_n(\omega)$ is the element of $\mathcal{P}$ which contains $T^n\omega$. Conversely, given a stationary process with values in $\mathcal{P}$ one can map the underlying probability space Ω into $\mathcal{P}^{\mathbf{Z}}$ by setting $\varphi(\omega) = \{X_n(\omega)\}$. The image under φ of the probability in Ω is a measure μ on $\mathcal{P}^{\mathbf{Z}}$, and the stationarity of the process $\{X_n\}$ is equivalent to the invariance of μ under the shift on $\mathcal{P}^{\mathbf{Z}}$. Thus a stationary process can be viewed as a measure preserving system $(\Omega, \mathcal{B}, \mu, T)$ with a specified finite partition $\mathcal{P}$. Setting $\mathcal{B}_0 = \bigvee_{-\infty}^{\infty} T^{-j}\mathcal{P}$ we refer to $(\Omega, \mathcal{B}_0, \mu, T)$ as the underlying system of the process. Two processes with isomorphic underlying systems will be referred to as isomorphic. Notice that the property *FDN* is a property of the process (or the generator) and not of the system.

THEOREM 1: *Every zero entropy process is isomorphic to one which is FDN.*

More precisely we show:

If $(\mathcal{P}^{\mathbf{Z}}, \mathcal{B}, \mu, T)$ has zero entropy ($\mathcal{P}$ finite; $\mathcal{B}$ the product sigma-algebra; T the shift; and μ a T-invariant probability measure on $\mathcal{B}$) then there exists a partition $\mathcal{Q}$ of $(\mathcal{P}^{\mathbf{Z}}, \mathcal{B})$ such that $\bigvee_{-\infty}^{\infty} T^{-j}\mathcal{Q} = \mathcal{B}$ and the process determined by $\mathcal{Q}$ is *FDN*.

Note that $\bigvee_{-\infty}^{\infty} T^{-j}\mathcal{Q} = \mathcal{B}$ will follow if we can show $\mathcal{P} \subset \bigvee_{-\infty}^{\infty} T^{-j}\mathcal{Q}$.

Before giving the proof in detail let us describe it briefly. We shall construct, via a sequence of Rohlin towers, a partition of the space $\Omega = \mathcal{P}^{\mathbf{Z}}$ into three sets labeled by $\{0, 1, 2\}$ so that the process determined by $\{0, 1, 2\}$ is an *FDN*. We construct an increasing sequence of Rohlin towers. The base of the n-th Rohlin tower will be identified by labelling the first L_n rungs by 1, the next L_n rungs by 0, and the next L_n rungs by 1, where L_n is some value much smaller than the height of the tower. The coding of the information needed to recover the process on the entire n-th tower will take place in the next L_n rungs. Taken together this set of $4L_n$ rungs will be denoted by θ_n, and the sets θ_n can be made disjoint. The predictability of the process under low density sampling will be ensured by performing the coding at the n-th stage with a great deal of redundancy.

Here are the details: let ε_i be a sequence of numbers approaching 0 rapidly and N_i be a sequence of numbers approaching ∞ rapidly. We will see how rapidly ε_i approaches 0 and N_i approaches ∞ as the proof continues. For each i we choose a Rohlin tower of height N_i and error set with measure smaller than $2\varepsilon_i + 2/10^i$. By being careful when we choose these towers and N_i, we can make sure that:

(3) Each N_i is divisible by $4i(10)^i$.

(4) $\begin{pmatrix}\text{error set}\\ \text{of } N_i \text{ tower}\end{pmatrix} \subset \begin{pmatrix}\text{error set}\\ \text{of } N_{i-1} \text{ tower}\end{pmatrix}$ for all i.

(5) All points in the bottom $N_i/10^i$ rungs of each N_i tower are in the error set of the N_{i-1} tower, henceforth refered to as the N_{i-1} error-set, for all i.

(6) All points in the top rung of the N_i tower are in the N_{i-1} error set.

Condition (4) and (6) can be easily established using well known techniques. Since we are allowing our error set to be so big, we can obtain (5) using the same method that we obtain (6).

We now develop a condition (7) on the N_i towers. Let b_i be the base of the N_i tower and $\bar{b}_i$ be the complement of b_i. Let $B_i = \{b_i, \bar{b}_i\}$. Let $\mathcal{P}_i = \mathcal{P} \vee B_{i-1} \vee B_{i-2} \vee \cdots \vee B_1$. Since $(\mathcal{P}, T)$ has entropy zero and each B_i is a factor of $(\mathcal{P}, T)$ it follows that $(\mathcal{P}_i, T)$ has entropy zero. It follows that by choosing N_i large enough we can ensure that after removing a set of atoms of measure less than ε_i from $\bigvee_{j=1}^{N_i} T^{-i}\mathcal{P}_i$ there are fewer than $2^{\varepsilon_i N_i}$ atoms left. Since the N_i error set is allowed to have measure more than $2\varepsilon_i$ this allows us to ensure (by making the base independent of $\bigvee_{j=1}^{N_i} T^{-j}\mathcal{P}_i$ and then throwing bad columns into the error set) that

(7) The N_i tower has less than $2^{\varepsilon_i N_i}$ $\mathcal{P}_i$-columns,

where a $\mathcal{P}_i$ column is a set of the form $S \cup TS \cup \cdots \cup T^{N_i} S$ where S is b_i intersected with an atom of $\bigvee_{j=1}^{N_i} T^{-i}\mathcal{P}_i$.

Note that the only significance of the fact that the error set of the N_i tower has measure less than $2\varepsilon_i + \frac{2}{10^i}$ is that ε_i is chosen small enough to make $2\varepsilon_i + \frac{2}{10^i}$ summable and hence with probability 1, all points are in all but finitely many towers.

Let θ_i be the bottom $N_i/10^i$ rungs of the N_i tower. Let $L_i = N_i/4(10)^i$. By (5), the θ_i are disjoint. Now we break θ_i into 4 equal sized pieces, the bottom L_i rungs, the next L_i rungs, etc. The bottom L_i levels are labelled one, then next L_i levels are labeled zero, and the following L_i levels are again labeled one. Break the last L_i levels into blocks of size i rungs apiece $\theta_{i,1}, \theta_{i,2}, \ldots, \theta_{i,L_i/i}$. For each j and each $\mathcal{P}_i$-column, the intersection of that column with $\theta_{i,j}$ can be labeled with a zero label (alternating zeros and ones starting with a zero) or a one label (alternating zeros and ones starting with a one.)

We define the notion of the mean Hamming distance (abbreviated $\bar{h}$ distance) between two labels of $\theta_{i,j}$. Note that the word "label" is being used ambiguously to refer either to a label of a $\theta_{i,j}$ or to a collection of L_i/i such labels, one for each value of j. The number of the set of j such that the two labels differ on $\theta_{i,j}$ divided by L_i/i is the $\bar{h}$ distance between the two labels. We would like to label the $\mathcal{P}_i$-columns (of the N_i tower) intersected with θ_i with different labels so that no two labels are closer than .01 in $\bar{h}$.

Let H be a maximal collection of such labels, pairwise farther apart than .01 in $\bar{h}$. We will label each distinct $\mathcal{P}_i$-column intersect $\theta_{i,4}$ with a different member of H. All that is necessary is to show

LEMMA: *$\#H \geq$ number of $\mathcal{P}_i$-columns.*

If you consider each member of H to be a center of a disk of radius .02 in $\bar{h}$, then these disks would cover all the labels (otherwise H would not be maximal). Furthermore all these disks would have the same size. That size is

$$\binom{m}{n} + \binom{m}{n-1} + \ldots + \binom{m}{0} \leq 2\binom{m}{n},$$

where $m = L_i = N_i/4i10^i$ and $n = (.02)L_i$. Also

$$\begin{aligned} 2\binom{m}{n} &= 2\frac{m(m-1)\cdots(m-n+1)}{n!} \\ &\leq 2m^n\left(\frac{n}{e}\right)^{-n} \leq \left(2e\frac{m}{n}\right)^n = (2e50)^n \leq 280^n \leq 2^{9n} = 2^{.18L_i}. \end{aligned}$$

Therefore, since these disks cover, we have

$$\begin{aligned} (\#(H))\,2^{.18L_i} \geq & \text{ total number of labels } = 2^{L_i} \\ \#(H) \geq & \; 2^{(.82)L_i} \end{aligned}$$

On the other hand recall that the number of columns is smaller than $2^{\varepsilon_i N_i}$ (see (7)) and that we can let ε_i be as small as we like. Let $\varepsilon_i < \frac{.82}{4i10^i}$. The lemma follows. ∎

Thus, we can label each column with a different label on $\theta_{i,4}$, each two such labels farther apart than .01 in $\bar{h}$. Thus we have assigned every point in every θ_i to either 0 or 1. The remainder of the space not yet labelled is assigned to the set 2. This defines the partition $\mathcal{Q} = \{0,1,2\}$.

Select a point $p \in \mathcal{P}^{\mathbf{Z}}$. With probability one, p is in all but finitely many of the N_i-towers and p is in at most one θ_i. Thus, select i such that p is in the N_i tower but not θ_i. There is a $\mathcal{P}_i$ column of the N_i tower containing p. Our immediate goal will be to determine that $\mathcal{P}_i$ column by randomly sampling the past $\mathcal{Q}$ name of $\mathcal{P}$.

Suppose we know where θ_i is, i.e. suppose we know the set of $k < N_i$ such that $t^{-k}(p) \in \theta_i$. For sufficiently large i, even if you are only sampling the past $\mathcal{Q}$ name of p with probability ε (review the definition of *FDN*) you will probably know the top L_i rungs of θ_i very well in $\bar{h}$. This is because these rungs rigidly alternate for stretches of length i so that even if you know only one member of such a stretch, you know the whole stretch. Even though we are sampling each

$T^{-j}(p)$ (to see where it is in $\mathcal{Q}$) with probability ε, if i is much larger than $\frac{1}{\varepsilon}$, we will probably sample at least one member of most of the stretches, thereby enabling us to see these rungs well in $\bar{h}$, and hence enabling us to know which column p is in.

Let us sum up what we have just established. Suppose $p \in \mathcal{P}^{\mathbf{Z}}$. If $T^{-j}(p)$ is sampled (to see which member of $\mathcal{Q}$ it is in) for all $j > 0$ with probability ε, then with probability one there is an i such that p is in the N_i tower, $p \notin \theta_i$, and if you know where θ_i is, then you know what $\mathcal{P}_i$-column of the N_i tower p is in.

Knowing where θ_i is, is equivalent to knowing which rung of the N_i tower p is on. We now show that if you know which rung p is on and which $\mathcal{P}_i$-column of the N_i tower p is on then

(8) You know which atom of $\mathcal{P}$ p is in.

(9) You know which atom of $\mathcal{Q}$ p is in.

Knowing the $\mathcal{P}_i$-column means knowing which atom of $\mathcal{P}_i$ each rung of that column is a subset of. Since you know which of those rungs p is in, you know which atom of $\mathcal{P}_i$ p is in. Since $\mathcal{P}_i$ is finer than $\mathcal{P}$, (8) follows.

We now analyze which atom of $\mathcal{Q}$ p is in, i.e. whether p is labeled 0, 1, or 2. Since $\mathcal{P}_i$ is finer than B_j for all $j < i$, we know, for all $j < i$, which atom of B_j each rung of the $\mathcal{P}_i$-column of the N_i tower p is in. Thus we can determine which rung of the N_j tower p is on. Hence you know, for each $j < i$, whether or not p is in θ_j (it is impossible for $p \in \theta_j$ for $j > i$ because otherwise p would be in the N_i error set). If it is in none of the θ_j then p is labeled 2. Now select $j < i$ and suppose $p \in \theta_j$. Since $\mathcal{P}_i$ is finer than $\mathcal{P}_j$ and B_j we know which $\mathcal{P}_j$ column of the N_j tower p is in. However, for a given $\mathcal{P}_j$ column of the N_j tower, we know which atom of $\mathcal{Q}$ each rung of θ_j is in. Thus we know whether p is assigned to either 0 or 1, (9) is established.

The above argument supposes that you will know where θ_i is. We now give a method of knowing where θ_i is. Sample the past $\mathcal{Q}$ name (sampling each letter with probability ε). Suppose you sample two successive terms and they turn out to be 1,0. Then exactly L_i rungs later you sample two successive terms and they turn out to be 0,1 and between that 1,0 and 0,1 you see a stretch of more than L_{i-1} zeros in succession. Then it follows that the $0, 1$ are precisely the L_i and L_{i+1} rungs of θ_i. ε^4 is the probability that when sampling with probability ε you will see the 0,1 and the 1,0. If the towers grow rapidly enough, we can insure that we see the desired stretch of zeros (in between the 1,0 and the 0,1) with probability exceeding 9/10 because L_{i-1} is so much smaller than L_i that we are bound to see a stretch of L_{i-1} zeros in a stretch of length L_i. This gives us a probability of $(9/10)\varepsilon^4$ of knowing where θ_i is. There are infinitely many i to try this experiment on, and the events that "you see 1,0, then 0,1 exactly L_i later, and the appropriate stretch of zeros in between" are independent as i runs, each with probability $(9/10)\varepsilon^4$. Thus you will know where θ_i is for infinitely many i. Thus we see that if we randomly sample the Q name into the past, after sampling finitely far, we can determine (8) and (9). (8) implies that $\mathcal{P} \subset \bigvee_{i\in\mathbb{Z}} T^i Q$ and (9) implies that $(Q^{\mathbb{Z}}, T)$ is *FDN*. The proof of Theorem 1 is complete. ∎

3. Extending Some Zero Entropy Processes to Uniform Martingales

In this section we prove:

THEOREM 2: *FDN implies that the process can be extended to a uniform martingale.*

The definition of *FDN* says essentially that a small randomly chosen past will, at some finitarily chosen point, predict the present. The following lemma says that this property will still hold true if, for some fixed N, we are forbidden to look at times $-1, -2, \ldots, -N$.

LEMMA: *Let $\{X_i\}$ be a FDN. Then for any fixed integer N, and any $\varepsilon > 0$, the process $\left(X_{-(N+1)}, a_{-(N+1)}\right), \left(X_{-(N+2)}, a_{-(N+2)}\right), \ldots$, where the $\{a_i\}$ process is i.i.d., $\{0,1\}$ valued, independent of the $\{X_i\}$ process, $P\left(a_i = 1\right) = \varepsilon$, has the property that with probability one there exists M such that those values X_i with $-N < i < -(N+M)$ and $a_i = 1$ determines X_0.*

Proof: Consider the entire process $\left(X_{-1}, a_{-1}\right), \left(X_{-2}, a_{-2}\right), \ldots$. It is possible, with positive probability, for $a_{-1} = a_{-2} = \cdots = a_{-N} = 0$. If the lemma were false, then the *FDN* property would be violated. ∎

For the proof of theorem 2 we shall need some definitions leading to that of **The Final Factor** (see [1], proof that T, T^{-1} can be extended to a uniform martingale). Let $\{N_i\}_{i\in\mathbb{N}}$ be a rapidly increasing sequence of integers. We will see how rapidly increasing as the proof continues. The **lookback variable** is a random variable M such that $P(M = N_i) = 1/2^{i+1}$. The **lookback process** is an i.i.d. doubly infinite process $\{M_i\}_{i\in\mathbb{Z}}$ such that each M_i is distributed like M. The **0-order process** is the product of $\{X_i\}_{i\in\mathbb{Z}}$ with $\{M_i\}_{i\in\mathbb{Z}}$ where the former is a *FDN* and the latter is a lookback process. The **1-order process** is obtained as follows. Start with the **0-order process.** For each i, ask whether or not $X_{i-M_i}, X_{i-M_i+1}, \ldots, X_{i-1}$ determines X_i. If the answer is yes, leave X_i as it is. If the answer is no let X_i take on the symbol "?", where "?" is interpreted as conveying no information. In general, let the n **order process** be defined from the $n-1$ order process in the same way that the 1 order process is defined from the 0 order process. Thus the set of n-order question marks is an increasing set with n. Hence, as n approaches ∞, the n-order process converges to a limit. That limit is called the **final process.** The factor of the final process you get by simply removing the lookback process is called the **final factor.**

PROPOSITION 1: *The final factor is a uniform martingale.*

Proof: Consider the final process. M_0 is independent of $X_{-1}, X_{-2}, \ldots$. Select $\varepsilon > 0$. There must exist K_ε such that $P(M_0 \geq K_\varepsilon) < \varepsilon$. For any $\{a_i\}_{i=1}^\infty$ and any $\{b_i\}_{i=1}^\infty$ in the alphabet of the $\{X_i\}$ process, and for any a,

$$\begin{aligned} &P(X_0 = a \mid M_0 < K_\varepsilon, X_{-1} = a_1, \ldots, X_{-K_\varepsilon} = a_{K_\varepsilon}, X_{-K_\varepsilon+1} = a_{K_\varepsilon+1}, \ldots) \\ =&P(X_0 = a \mid M_0 < K_\varepsilon, X_{-1} = a_1, \ldots, X_{-K_\varepsilon} = a_{K_\varepsilon}, X_{-K_\varepsilon+1} = b_1, \ldots) \end{aligned}$$

so

$$\begin{aligned} \Big\| &P(X_0 = a \mid X_{-1} = a_1, \ldots, X_{-K_\varepsilon} = a_{K_\varepsilon}, X_{-K_\varepsilon+1} = a_{K_\varepsilon+1}, \ldots) \\ &- P(X_0 = a \mid X_{-1} = a_1, \ldots, X_{-K_\varepsilon} = a_{K_\varepsilon}, X_{-K_\varepsilon+1} = b_1, \ldots) \Big\| \leq \varepsilon \end{aligned}$$ ∎

PROPOSITION 2: *If the $\{N_i\}$ increase rapidly enough, then the final factor is an extension of the FDN*

Proof: First we show that the final process is an extension of the *FDN*. By the *FDN* property, we can select a random set $\hat{\theta}_1$, subset of the negative integers, such that,

(10a) for $i \in \hat{\theta}_1, M_i = N_1$,

(10b) $\hat{\theta}_1$ is finite with probability one, and

(10c) $\{X_i\}_{i\in\hat{\theta}_1}$ determines X_0. Note that we have not actually chosen N_1 yet. Pick N_0 such that

$$P\left(\hat{\theta}_1 \subset \{-1, -2, \ldots, -N_0\}\right) \geq 1 - 10^{-1}.$$

Let $\theta_1 = \hat{\theta}_1 \cap \{-1, -2, \ldots, -N_0\}$. Then

(11) $P\left(\theta_1 = \hat{\theta}_1\right) \geq 1 - 10^{-1}$.

When $\theta_1 = \hat{\theta}_1$, in the 1-order process X_0 will not become a question mark. Similarly, there is a random set $\hat{\theta}_2$ such that,

(12a) for all $i \in \hat{\theta}_2, M_i = N_2$.

By the lemma we can insist that:

(12b) all of $\hat{\theta}_2$ is less than $-N_0$,

(12c) $\hat{\theta}_2$ is finite with probability one,

and

(12d) $\{X_i\}_{i\in\hat{\theta}_2}$ determines X_0 and all $\{X_i\}_{i\in\theta_1}$.

Choose N_1 so that

$$P\left(\hat{\theta}_2 \subset \{-N_0 - 1, -N_0 - 2, \ldots, -N_1\}\right) \geq 1 - 10^{-2}$$

Let $\theta_2 = \hat{\theta}_2 \cap \{-N_0 - 1, -N_0 - 2, \ldots, -N_1\}$. Then

(13) $P\left(\theta_2 = \hat{\theta}_2\right) \geq 1 - 10^{-2}$.

If the event $\theta_2 = \hat{\theta}_2$ takes place, then the $X_i, i \in \theta_1$ will not become question marks in the 1-order process and if also $\theta_1 = \hat{\theta}_1$ then X_0 will not become a question mark in the 2-order process.

In general, there is a random set $\hat{\theta}_n$ such that, for all $i \in \hat{\theta}_n, M_i = N_n$ such that, all of $\hat{\theta}_n$ is less than $-N_{n-2}$, such that $\hat{\theta}_n$ is finite with probability one, and such that $\{X_i\}_{i\in\hat{\theta}_n}$ determines X_0 and all $\{X_i\}_{i\in\hat{\theta}_{n-1}}$. Choose N_{n-1} so that

$$P\left(\hat{\theta}_n \subset \{-N_{n-2} - 1, -N_{n-2} - 2, \ldots, -N_{n-1}\}\right) \geq 1 - 10^{-n}.$$

Let $\theta_n = \hat{\theta}_n \cap \{-N_{n-2} - 1, -N_{n-2} - 2, \ldots, -N_{n-1}\}$.

(14) $P\left(\theta_n = \hat{\theta}_n\right) \geq 1 - 10^{-n}$.

By (14), and Borel Cantelli, there exists n_0 such that

(15) for all $n \geq n_0$, $\theta_n = \hat{\theta}_n$.

It follows that for all $n \geq n_0$ and all $i \in \theta_n, X_i$ never becomes a question mark in any m-order process and hence is not a question mark in the final process. However, $\{X_i\}_{i\in\theta_{n_0}}$ determines X_0.

Although X_0 may be a question mark in the final process, the final process determines what X_0 originally was and similarly it determines what all the X_i originally were, and hence the final process is an extension of the original process.

We really didn't need to use the knowledge of the M_i's for the determination of X_0, and this shows that the final factor is an extension of the original process and the proof of Theorem 2 in concluded. ∎

THEOREM 3: *Every zero entropy can be extended to a uniform martingale.*

Proof: This is just a combination of Theorem 1 and Theorem 2. ∎

4. Final Remarks

Our first remark concerns the nature of the random noise in the definition of FDN. We worked with the simplest kind of random noise—namely, one in which the noise process itself consists of independent random variables. A more elaborate coding scheme would enable one to enlarge the admissible noise processes to a much wider class. It should be possible to carry out the construction in such a way that any process **disjoint** from the fixed zero entropy process could be the noise process. In particular, any K-system could be the noise. For the application we had in mind here these generalizations were not needed and so we kept the construction as simple as possible.

Here are some open problems:

PROBLEM 1: *Is every K process isomorphic to a uniform martingale?*

PROBLEM 2: *Can every K process (perhaps every process) be extended to a uniform martingale?*

PROBLEM 3: *Perhaps every positive entropy process is isomorphic to a uniform martingale (this seems unlikely).*

PROBLEM 4: *Every measure preserving, finite entropy, ergodic transformation is isomorphic to a stationary pair process $\{x_i, y_i\}_{i\in\mathbf{Z}}$, each x_i in some finite alphabet, each y_i in some finite alphabet, $\{x_i\}_{i\in\mathbf{Z}}$ Bernoulli and y_0 determined by $\{x_i\}_{i\in\mathbf{Z}}$ and $\{y_i\}_{i<0}$. Can we also insist that y_0 be determined by $\{x_i\}_{i<0}$ and $\{y_i\}_{i<0}$?*

This last problem needs some justification. Although it is an interesting problem in its own right, the reader can wonder what it has to do with this paper. The T, T^{-1} transformation which has the form of this problem, has already been shown to extend to a uniform martingale in [1] and if this open problem can be solved affirmatively, we may be able to extend all transformations to a uniform martingale (using the technique of the current paper).

References

[1] S. Kalikow, *Random Markov process and uniform martingales*, Israel J. Math. **71** (1990), 33–34.

[2] H. Berbee, *Chains with infinite connections: Uniqueness and Markov representation*, Probability Theory and Related Fields **76** (1987), 243–253.

[3] W. Doeblin and R. Fortet, *Sur les chains a liaisons complétes*, Bull. Soc. Math. France **65** (1937), 132–148.

[4] M. Keane, *Strongly mixing g-measures*, Invent. Math. **16** (1972), 309–324.

[5] B. Petit, *Schemes de Bernoulli et g-measure*, C.R. Acad. Sci. Paris **Ser. A** (1975), 17–20.