Combinatorics, Probability and Computing (2003) 12, 547–563. © 2003 Cambridge University Press
DOI: 10.1017/S0963548303005893 Printed in the United Kingdom

Markov Processes and Ramsey Theory for Trees

HILLEL FURSTENBERG and BENJAMIN WEISS

Institute of Mathematics, Hebrew University, Givat Ram, Jerusalem 91904, Israel
(e-mail: harry@math.huji.ac.il, weiss@math.huji.ac.il)

Received 10 February 2003; revised 15 June 2003

We consider analogues of van der Waerden's theorem and Szemerédi's theorem, where arithmetic progressions are replaced by binary trees with a fixed distance between successive vertices. The proofs are based on some novel recurrence properties for Markov processes.

1. Introduction

We think of a (finite) tree as a connected directed graph with every vertex the endpoint of a unique path leading to it from a distinguished vertex (the root). A finite tree with no branching is simply a linearly ordered finite sequence of vertices:

$$0 \to 1 \to 2 \to \cdots \to n,$$

which we can regard as the paradigmatic arithmetic progression. Our purpose here is to show that large trees with positive branching exhibit Ramsey-theoretic phenomena analogous to those making their appearance in long arithmetic progressions.

Each vertex v of a tree has a *level*, $\ell(v)$ denoting the length of the path connecting v to the root of the tree.

We will say a tree has *uniform height* h if every vertex belongs to a maximal directed path of length h. If there are $2^{\alpha h}$ such paths, we say that the tree has *branching* 2^{α}. A *binary* tree is one for which each vertex leads to exactly 2 vertices unless it is a terminal vertex. Thus a binary tree of uniform height has branching 2. The structure we will find replicated in large trees of positive growth is that of an *arithmetic binary sub-tree*. This term will refer to a subset of vertices of the original tree consisting of $m+1$ subsets $V_0, V_1, \ldots, V_m$ of vertices with V_0 consisting of a single vertex, V_1 of 2 vertices, and generally V_i of 2^i vertices with each vertex $v \in V_i$, $i < m$, leading to exactly 2 vertices in V_{i+1} of level $\ell(v) + d$ for a fixed length d which is independent of i. Moreover, we require that the two paths leading out of a vertex in V_i, $i < m$ split at the first step. Thus the geometric subgraph formed by these vertices and the edges joining them has itself the appearance of a binary tree.

https://doi.org/10.1017/S0963548303005893 Published online by Cambridge University Press

Throughout our discussion we will assume that all trees have bounded degree; that is, there exists a fixed s such that no vertex in the graph is adjacent to more than $s+1$ edges (one leading to the vertex and the rest leading out).

One statement we will prove is as follows.

Theorem A. *$\exists H_1(r,h)$ for $h,r \in \mathbb{N}$ such that, if τ is a binary tree of uniform height $\geqslant H_1(r,h)$ and the vertices of τ are coloured with r colours, then τ contains a monochromatic arithmetic binary subtree of height h.*

Theorem A will follow from a 'density version' in the same spirit. To formulate this we need an appropriate notion of density for subsets of vertices in a tree of uniform height n. If we restrict a set $S \subset \mathrm{Vert}(\tau)$ to a single directed path p of length n, we can set $D_p(S) = |S \cap p|/(n+1)$. We then say that the density of S is

$$\sum_p D_p(S) \Big/ \sum_p 1,$$

where the sums are taken over all the maximal paths p in τ. The denominator is then $2^{\alpha n}$, where 2^α is the branching of τ. We now state the following result.

Theorem B. *$\exists H_2(\delta,h)$ for $\delta > 0$ and $h \in \mathbb{N}$ such that, if τ is a binary tree of uniform height $> H_2(\delta,h)$ and S is any subset of vertices of τ with density $\geqslant \delta$, then S contains an arithmetic binary subtree of height h.*

These theorems bear an obvious resemblance to the theorem of van der Waerden and Szemerédi on (ordinary) arithmetic progressions [3]. Indeed, by choosing colourings of vertices depending only on their level, Theorem A is seen to imply van der Waerden's theorem, and similarly, considering sets S in Theorem B consisting of unions of levels spaced with positive density, we deduce Szemerédi's result from Theorem B. We emphasize that in our definition of binary subtrees of, say, a family tree, the two forefathers of a particular individual must be, respectively, on the paternal side and on the maternal side. Without insisting on the immediate branching, as well as the uniform spacing between generations, our results could be deduced from the classical versions using standard combinatorial considerations.

Finally we formulate another result in the same spirit for trees of any branching greater than 1.

Theorem C. *$\exists H_3(\alpha,s,h)$ for all $\alpha > 0$ and $s,h \in \mathbb{N}$, such that, if τ is any tree with maximal degree $s+1$ of branching $\geqslant 2^\alpha$ and of uniform height $\geqslant H_3(\alpha,h)$, then τ contains an arithmetic binary tree of height h.*

The proofs of these results will depend on establishing 'multiple' recurrence results for Markov processes in which each point in the state space has positive probability for transition to any one of several points in the state space (but no more than s points). Note that for such a Markov system we can attach to each point x of the state space a

https://doi.org/10.1017/S0963548303005893 Published online by Cambridge University Press

tree of points, that is, a tree whose vertices are points of the state space with x the root of the tree, and whose edges represent transitions of positive probability. In the usual ergodic-theoretic setup with a single measure-preserving transformation T on a measure space $(X, \mathscr{B}, \mu)$, the multiple recurrence theorem proved in [2] (see also [3]) states that for a set $A \subset X$ with $\mu(A) > 0$, there will be an $m > 0$ and a point $x \in A$ that recurs repeatedly, so that for a preassigned k, we have $T^m x \in A, T^{2m} x \in A, \ldots, T^{km} x \in A$. In [2] it is shown that this statement is equivalent to Szemerédi's theorem on arithmetic progressions. Our main tool consists of an extension of this recurrence phenomenon to the Markovian situation. In the Markovian situation, consider the infinite 'transition' tree whose vertices are points of the state space and whose edges represent transitions of positive probability. An arithmetic binary subtree then consists of a 'tree' of points in the space that can be reached starting from one point in steps that are multiples of some $d > 0$. We will show that if A is a subset of the state space of positive 'stationary measure' such that, for each point in A, there are transitions of positive probability to each of two disjoint sets B_0, B_1 in the state space, then for a preassigned k there is an arithmetic binary subtree of height k, such that the vertices of the subtree are points of A and such that the two paths emerging out of these vertices pass through B_0 and B_1 respectively, in the next transition. Appropriate applications of this result will give us Theorems B and C.

In these applications the points of the state space are themselves infinite trees with some added structure, to which we give the name 'labelled M-tree'. Given an infinite tree, there are natural transitions to the trees obtained by shifting the root to one of its successors, and in an 'M-tree' these transitions will have built-in probabilities which we can think of as attached to the edges of the tree. Finally, a 'labelled M-tree' will carry further data which will translate to a continuous function on the state space.

We develop these notions in Section 2. In Section 3 we discuss general Markov systems, and in Section 4 we show how to construct stationary measures from finite data and define entropy. In Section 5 we introduce endomorphic Markov systems and prove a multiple recurrence theorem, and finally in Section 6 we use the machinery developed to prove statements regarding large finite trees. We remark that some of these ideas were discussed in [1].

2. From trees to labelled Markov trees

2.1. Trees

Suppose we have a directed graph as described in the Introduction and we label the edges by letters of a finite alphabet (of, say, s letters). Then we can identify vertices of the tree with words whose letters correspond to the route taken from the root of the tree to the given vertex. It will be convenient to define a tree *ab initio* as a set of words. In the following A is an alphabet with s letters: $A = \{a_1, a_2, \ldots, a_s\}$. $A^* = \bigcup_{n \geqslant 0} A^n$, with A^0 consisting of the 'empty word' ϕ of length 0. A^* is a semigroup under the natural multiplication by juxtaposition. If $w \in A^n$ we define its length $\ell(w) = n$. u will be called a *predecessor* of w if $w = uv$ for some $v \in A^*$.

https://doi.org/10.1017/S0963548303005893 Published online by Cambridge University Press

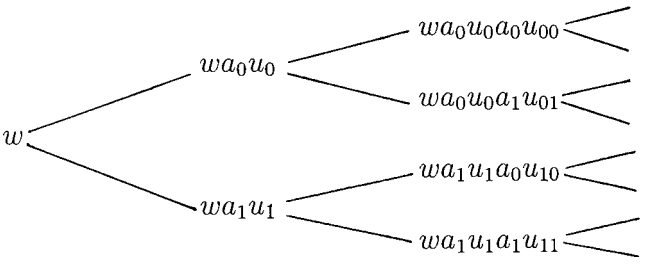


Figure 1.

Definition 1. A *tree* (over A) is a subset $\tau \in A^*$ satisfying:

(i) if $w \in \tau$ and u is a predecessor of w then $u \in \tau$,
(ii) if $w \in \tau$ then for some $a \in A$, $wa \in \tau$.

Condition (ii) forces our trees to be infinite. We shall also consider finite trees which violate (ii) and only satisfy (i). In a finite tree, a word is *terminal* if it is not a proper predecessor of any other word. If all terminal words have the same length, the finite tree is said to have *uniform height*.

We can make precise the notion of an *arithmetic binary subtree* in this terminology. It is a set of words forming a scheme as shown in Figure 1, where a_0 and a_1 are distinct letters of A and all the u's have the same length.

Given the alphabet A we can also consider infinite words, $\Omega^+ = A^{\mathbb{N}}$, which we can regard as the 'boundary' of A^*: $\Omega^+ = \partial A^*$. Any tree τ has a non-empty boundary $\partial\tau \in \Omega^+$, which we may identify with infinite paths inside τ. $\partial\tau$ is a closed subset of Ω^+ and τ is determined by $\partial\tau$. Moreover, any closed subset $Q \subset \Omega^+$ is the boundary of some tree: $Q = \partial\tau$, where τ is uniquely determined by Q.

We shall denote by $\mathscr{T}(= \mathscr{T}_A)$ the space of all trees over A. We can define a metric on $\mathscr{T}$ by

$$d(\tau_1, \tau_2) = \inf\left\{\frac{1}{N+1} : \tau_1 \cap A^N = \tau_2 \cap A^N\right\}. \tag{2.1}$$

It is easy to see that if we define the distance between closed subsets of Ω^+ by

$$D(\partial\tau_1, \partial\tau_2) = d(\tau_1, \tau_2),$$

then D is equivalent to the Hausdorff metric on the closed subsets of Ω^+.

2.2. M-trees

We remarked in the previous section that trees over A correspond to closed subsets of $\Omega^+ = A^{\mathbb{N}}$. In this section we define M-trees; these will correspond to *probability measures* on Ω^+. (M stands for Markov.)

https://doi.org/10.1017/S0963548303005893 Published online by Cambridge University Press

Definition 2. An M-tree with support $\tau \in \mathscr{T}$ consists of the following data: for each $w \in \tau$ and $a \in A$ with $wa \in \tau$, we are given $p(w;a)$, $0 < p(w;a) \leqslant 1$ such that, for each w,

$$\sum_{a \in A} p(w;a) = 1. \tag{2.2}$$

$\mathscr{M}_A$ or $\mathscr{M}$ will denote the space of all M-trees. If $\pi \in \mathscr{M}$ then $|\pi| \in \mathscr{T}$ will denote the support of π.

If π consists of the data $\{p(w;a)\}$ we may write p_π for p. Given the data $\{p(w;a)\}$, which we regard as transition probabilities from w to the successor wa, we can also form $\{q(w)\} = \{q_\pi(w)\}$ for $w \in \tau$ defined by

$$q(a_{i_1}a_{i_2}\cdots a_{i_\ell}) = p(\phi;a_{i_1})p(a_{i_1};a_{i_2})p(a_{i_1}a_{i_2};a_{i_3})\cdots p(a_{i_1}\cdots a_{i_{\ell-1}};a_{i_\ell}). \tag{2.3}$$

It is readily seen that the q-data determines and is determined by a probability measure on Ω^+ whose support is $\partial|\pi|$.

Let $\pi \in \mathscr{M}$ and assume $b \in A$ with $p_\pi(\phi;b) = q_\pi(b) > 0$. We can form a new $\mathscr{M}$-tree, denoted by π^b defined by

$$p_{\pi^b}(w;a) = p_\pi(bw;a); \tag{2.4}$$

$q_{\pi^b}(w)$ will then be the conditional probability

$$q_{\pi^b}(w) = \frac{q_\pi(bw)}{q_\pi(b)}. \tag{2.5}$$

The function $q_\pi(w)$ can be extended from the support of π to all of A^* by setting $q_\pi(w) = 0$ for $w \notin |\pi|$. We can then identify $\mathscr{M}$ with the closed subset of $[0,1]^{A^*}$ satisfying

$$q(w) = \sum_{a \in A} q(wa), q(\phi) = 1. \tag{2.6}$$

In particular, any function $q(w)$ satisfying (2.6) corresponds to an M-tree whose supporting tree is the set of w for which $q(w) > 0$. With this identification, $\mathscr{M}$ can be regarded as a compact, metric topological space. In particular we will have $\pi_n \to \pi$ in $\mathscr{M}$ if and only if, for each word $w \in A^*$, $q_{\pi_n}(w) \to q_\pi(w)$. Note that $\pi_n \to \pi$ does not imply that $|\pi_n| \to |\pi|$. Alternatively, by identifying Markov trees with probability measures on A^N, so that $\pi \in \mathscr{M}$ corresponds to the measure ν determined by

$$q_\pi(a_{i_1}a_{i_2}\cdots a_{i_e}) = \nu\{\text{cylinder set with } X_1 = a_{i_1}, \ldots, X_e = a_{i_e}\}, \tag{2.7}$$

the topology in question corresponds to the usual weak topology on the probability measure on the compact space $A^{\mathbb{N}}$.

2.3. *LM*-trees

In this section we add a final bit of data to give us 'labelled Markov trees'. Once again the alphabet A is fixed, but an additional alphabet B is provided.

Definition 3. An LM-tree λ consists of a pair $\lambda = (\pi, \xi)$ with $\pi \in \mathscr{M}$ and $\xi \in B^{A^*}$. The set of LM-trees associated with the alphabets A, B will be denoted by $\mathscr{LM}_{A,B}$, and $\mathscr{LM}_{A,B} = \mathscr{M} \times B^{A^*}$ can again be regarded as a compact metric space.

https://doi.org/10.1017/S0963548303005893 Published online by Cambridge University Press

The function $\xi(w)$ will be thought of as labelling (among others) the vertices of the M-tree π by elements of B. The values of $\xi(w)$ outside of the support $|\pi|$ will play a secondary role in the sequel.

We shall say that a labelling is *discriminating* if, for two words of the same length differing only in their last letter, it takes on different values. In particular, for distinct single letters it takes on distinct values. Hereafter we will deal only with discriminating labellings.

If $\xi \in B^{A^*}$ and $b \in A$, we can define ξ^b by setting

$$\xi^b(w) = \xi(bw).$$

Note that discriminating labellings remain discriminating in this transition. Also the discriminating labellings form a closed subset of all the labellings from a given alphabet.

In particular, if $\lambda = (\pi, \xi) \in \mathscr{LM}_{A,B}$ and $q_\pi(b) > 0$ then the 'transition' $\pi \to \pi^b$ is defined by (2.5), and we can also define the transition $\lambda \to \lambda^b$, setting

$$\lambda^b = (\pi^b, \xi^b).$$

More generally, for $u \in A^*$ we will define $\lambda^u = (\pi^u, \xi^u)$, where

$$q_{\pi^u}(w) = \frac{q_\pi(uw)}{q_\pi(u)}, \quad \xi^u(w) = \xi(uw) \tag{2.8}$$

provided $q_\pi(u) > 0$.

Note that if A consists of a single letter a, then there is a unique tree in $\mathscr{T}_A$:

$$\tau : \phi \to a \to a^2 \to a^3 \to \cdots \to a^n \to \cdots.$$

There is therefore also a unique M-tree in $\mathscr{M}$, and $\mathscr{LM}_{A,B}$ may be identified with the space of B sequences. The transition $\lambda \to \lambda^b$ is then the shift on $B^{\mathbb{N}}$, the space of B sequences. In the next section we extend notions of symbolic dynamics to 'Markov systems', which will form the appropriate context for studying LM-trees.

2.4. Finite trees

Clearly the trees appearing in the preceding paragraphs could also be finite trees with the data restricted to words of length $\leqslant L$ for some L. If λ is such a finite M-tree, or LM-tree we will say that λ is of height L. Suppose that λ is an LM-tree of height L, $\lambda = (\pi, \xi)$ and that ξ is numerical-valued. We will define the average of ξ on λ as follows.

Definition 4. The average $\pi(\xi)$ is defined as

$$\frac{1}{L+1} \sum_{\ell(w) \leqslant L} q_\pi(w)\xi(w). \tag{2.9}$$

Note that if ξ is a constant, say $\xi = c$, then $\pi(\xi) = c$.

Now, for each word w of length L we can consider the $L+1$ initial subwords of w and denote this set $I(w) = \{u : w = uv, \text{ for some } v\}$. We can form the average of ξ along

https://doi.org/10.1017/S0963548303005893 Published online by Cambridge University Press

these words:

$$A_w(\xi) = \frac{1}{L+1} \sum_{u \in I(w)} \xi(u).$$

Lemma 2.1.

$$\pi(\xi) = \sum_{\ell(w)=L} q_\pi(w) A_w(\xi).$$

Proof. For any word u of length $l \leqslant L$, we have

$$q_\pi(u) = \sum_{\ell(v)=L-\ell} q_\pi(uv). \tag{2.10}$$

We thus obtain

$$\begin{aligned} \sum_{\ell(w)=L} q_\pi(w) A_w(\xi) &= \sum_{\ell(w)=L} \left(\frac{1}{L+1} \sum_{uv=w} q_\pi(uv) \xi(u) \right) \\ &= \frac{1}{L+1} \sum_{\ell(u) \leqslant L} \xi(u) \sum_{\ell(v)=L-\ell(u)} q_\pi(uv) \\ &= \frac{1}{L+1} \sum q_\pi(u) \xi(u), \end{aligned}$$

which proves our assertion. □

Having a notion of average of $\xi(w)$ over the finite M-trees π also gives us a notion of density of a subset S of the vertices of π. Namely, we take as ξ the characteristic function $\mathbf{1}_S(w)$ and define the density

$$\Delta_\pi(S) = \pi(\mathbf{1}_S). \tag{2.11}$$

Suppose that we are given a finite tree of uniform height L and branching 2^β. We can turn it into a finite M-tree by setting $q(w) = 2^{-\beta L}$ for all words of the maximal length L, and using formula (2.10) to define q for all shorter words. The foregoing notion of density then coincides with that referred to in the formulation of Theorem B in the Introduction.

3. Markov systems

One can speak generally of a 'Markov system' in which a state space is specified, and for each point x of the space a measure p_x is specified representing transition probabilities from x to sets in the space. We will be interested in the situation where for each point only a finite number of transitions are possible, each of these being assigned a positive probability. We can then speak of paths of positive probability as those for which each transition has positive probability. The following restricted definition will encompass the situations we will deal with.

Definition 5. A *Markov system* consists of a pair $(M, p(\cdot,\cdot)$ where M is a compact metric space, p is a function on $M \times M$, and the following are satisfied:

https://doi.org/10.1017/S0963548303005893 Published online by Cambridge University Press

(i) $p(x, y) \geqslant 0$,
(ii) there exists an $s \in \mathbb{N}$ such that, for each $x \in M$, $p(x, y) = 0$ except for at most s values of $y \in M$,
(iii) $\sum_y p(x, y) = 1$ for all $x \in M$,
(iv) if the operator P is defined on functions on M by

$$Pf(x) = \sum_{y \in M} p(x, y) f(y), \tag{3.1}$$

then P takes continuous functions to continuous functions.

The transition matrix of a finite state Markov chain defines a Markov system with finite M. In the sequel we shall be interested in the example

$$M = \mathscr{L}\mathscr{M}, p(\lambda, \lambda^b) = q_\pi(b) \tag{3.2}$$

where $\lambda = (\pi, \xi)$ and $\lambda^b = (\pi^b, \xi^b)$, and $p(\lambda, \lambda') = 0$ when λ' is not of the form λ^b.

The operator P acts on $C(M)$, the space of continuous function on M, and P^* will denote the adjoint operator on the space of borel measures on M. P^* takes probability measure to probability measures, and since M is compact there necessarily exists a probability measure μ on M satisfying $P^*\mu = \mu$. We note for later reference that for Dirac measures

$$P^*\delta_x = \sum_y p(x, y)\delta_y. \tag{3.3}$$

Definition 6. A probability measure μ on M is *stationary* with respect to a Markov system $(M, p(\cdot, \cdot))$ if $P^*\mu = \mu$, *i.e.*, if for every continuous $f \in C(M)$

$$\int f(x) d\mu(x) = \int \sum_y p(x, y) f(y) d\mu(x). \tag{3.4}$$

A *stationary Markov system* consists of a triple $(M, p(\cdot, \cdot), \mu)$ where μ is a stationary measure with respect to the Markov system $(M, p(\cdot, \cdot))$.

A stationary measure μ for a Markov system $(M, p(\cdot, \cdot))$ defines a stationary M-valued process $\{W_n\}_{n \geqslant 0}$ with

$$\begin{aligned} &\Pr(W_0 \in A_0, W_1 \in A_1, \ldots, W_k \in A_k) \\ &\quad = \int \sum_{x_i \in A_i} p(x_0, x_1) p(x_1, x_2) \cdots p(x_{k-1}, x_k) d\mu(x_0). \end{aligned} \tag{3.5}$$

The random variables W_k are defined on the space

$$\Omega = M \times M \times M \times \cdots,$$

which carries a measure $\tilde{\mu} = \Pr$ defined by (3.5). The natural shift transformation $\theta : \Omega \to \Omega$ defined by

$$\theta(x_1, x_2, \ldots, x_n, \ldots) = (x_2, x_3, \ldots, x_{n+1}, \ldots) \tag{3.6}$$

https://doi.org/10.1017/S0963548303005893 Published online by Cambridge University Press

preserves the measure $\tilde{\mu}$ (this is equivalent to the stationarity of μ, which in turn is equivalent to the stationarity of the process $\{W_n\}_{n\geqslant 0}$) and consequently we can apply ergodic-theoretic results to the system $(\Omega, \tilde{\mu}, \theta)$. This will constitute a major step in our proof of Theorems A, B, and C.

3.1. Markov systems on $\mathscr{M}$ and $\mathscr{LM}$

The Markov systems that will occupy us here are those in the space of Markov trees and labelled Markov trees for an alphabet A and an alphabet B. On these spaces there is a natural transition function $p(x, y)$ satisfying the conditions of Definition 5. We begin with $\mathscr{M}$. Here, for $\pi \in \mathscr{M}$ and $a \in A$ we set

$$p(\pi, \pi') = \sum_{\substack{a\in A \\ \pi^a=\pi'}} q_\pi(a) = \sum_{\substack{a\in A \\ \pi^a=\pi'}} p_\pi(\phi, a). \tag{3.7}$$

By (2.2) with $w = \phi$, we find that

$$\sum_{\pi'\in\mathscr{M}} p(\pi, \pi') = 1.$$

From the definition of the topology on $\mathscr{M}$ we see that, for any $b \in A$, $q_\pi(b)$ depends continuously on π. Using (2.5) we see that, provided $q_\pi(b) > 0$, the expression $q_{\pi^b}(w)$ depends continuously on π; hence $\pi \to \pi^b$ is continuous in the domain $q_\pi(b) > 0$. If P is the operator determined by the foregoing $p(\pi, \pi')$, we will have

$$Pf(\pi) = \sum_b f(\pi^b) q_\pi(b),$$

which will depend continuously on π. This shows that $(\mathscr{M}, p(\pi, \pi'))$ is a Markov system.

Next we consider the discriminatingly labelled A, B Markov trees. The possible transitions will be from a labelled tree (π, ξ) to its successors (π^a, ξ^a) a in A. Now if $\xi' = \xi^a$, then necessarily

$$\xi'(\varnothing) = \xi(a),$$

and since the labelling is discriminating, there is a unique letter a which can appear here. This simplifies the situation, and we have

$$p((\pi, \xi), (\pi^a, \xi^a)) = q_\pi(a). \tag{3.8}$$

Here (π^a, ξ^a) is defined provided $q_\pi(a) > 0$, and in this domain (π^a, ξ^a) depends continuously on π. (Note that ξ^b depends continuously on ξ for every b.) As a result the corresponding operator P takes continuous functions to continuous functions. As indicated earlier, this implies the existence of stationary measures for the Markov system. We shall speak of the Markov systems on $\mathscr{M}_A$ and $\mathscr{LM}_{A,B}$ as the *intrinsic* Markov systems on these spaces, and as we shall see, the sets of stationary measures for the aforementioned 'intrinsic' Markov systems are very large.

3.2. Prediction measure for stationary processes

In this section we present one example of a stationary Markov system on the space of LM-trees that arises naturally in the context of finite-valued stationary processes. Namely,

https://doi.org/10.1017/S0963548303005893 Published online by Cambridge University Press

in that context we can regard a tree as a set of possible 'future' sequences attached to a 'past', and a Markov tree can be regarded as a probability distribution over the set of futures. We can speak of the distribution of the future conditional on a particular past, *i.e.*, the prediction measure, and the distribution of various 'pasts' produces a stationary measure on the space of Markov trees. The labelling that we will use will record the present value of the stationary process.

Specifically, let A be a finite alphabet and let $\{Z)_n\}_{n\in\mathbb{Z}}$ be any stationary A-valued stochastic process. We set

$$\Omega^- = A^{\mathbb{Z}^-},$$

where $\mathbb{Z}^-$ denotes the negative integers with 0. The variables $\{Z_n\}_{n\leqslant 0}$ are defined on Ω^- and determine a measure Pr on Ω^-. For any word $w = a_{i_1}a_{i_2}\cdots a_{i_\ell}$ the conditional probabilities $\Pr(Z_1 = a_{i_1}, Z_2 = a_{i_2}, \ldots, Z_\ell = a_{i_\ell}|Z_0, Z_{-1}, \ldots)(\omega)$ are defined for almost all $\omega \in \Omega^-$. These define a measurable map, $\omega \mapsto \pi(\omega)$ from Ω^- to $\mathscr{M}_A$. We now set $B = A$, and attach to each ω the following labelling: for each word of positive length

$$w = a_{i_1}a_{i_2}\cdots a_{i_l},$$

we set $\xi(w) = a_{i_l}$, the terminal letter of w. This is independent of ω. However, we set $\xi(\varnothing) = Z_0(\omega)$. Let μ be the image of Pr in $\mathscr{L}\mathscr{M}_{A,A}$ under the map $\omega \mapsto \lambda(\omega)$ taking each ω to the labelled tree described. We claim that μ is a stationary measure for the transition probabilities given in (3.8).

To see this, note that by setting $U_n = (\ldots Z_{n-k}, \ldots, Z_{n-1}, Z_n)$, the variables $\{U_n\}$ form a Markov process with values in Ω^- and with transition probabilities

$$\Pr(U_{n+1} = (\omega, b)|(U_n = \omega) = P(Z_1 = b|Z_0, Z_{-1}, \ldots)(\omega).$$

This process is stationary; consequently $\{\lambda(U_n)\}_{n>0}$ is a stationary process. It can now be verified that $\{\lambda(U_n)\}_{n>0}$ is the stationary Markov process on $\mathscr{L}\mathscr{M}_{A,A}$ for which μ is the stationary measure and for the transitions of the intrinsic Markov system. This follows from the fact that $\lambda(U_1)$ is the Markov tree with

$$q_{\pi(U_1)}(a_{i_1}\cdots a_{i_\ell}) = \Pr(Z_2 = a_{i_1}, Z_3 = a_{i_2}, \ldots, Z_{\ell+1} = a_{i_\ell}|Z_1, Z_0, \ldots, Z_{-n}, \ldots). \tag{3.9}$$

If $Z_1 = b$ so that $U_1 = (U_0, b)$, then the conditional probability in (3.9) is the quotient

$$\frac{\Pr(Z_1 = b, Z_2 = a_{i_1}, \ldots, Z_{\ell+1} = a_{i_\ell}|Z_0, Z_{-1}, \ldots)}{\Pr(Z_1 = b|Z_0, Z_{-1}, \ldots)} = \frac{q(ba_{i_1}\cdots a_{i_\ell})}{q(b)}$$

with $q = q_{\pi(U_0)}$, so that $q_{\pi(U_1)} = q^{Z_1}_{\pi(U_0)}$ according to (2.5). Since

$$\Pr(Z_1 = b|U_0) = q_{\pi(U_0)}(b),$$

which depends only on $\lambda(U_0)$, we can also write

$$\Pr(Z_1 = b|\lambda(U_0)) = q_{\pi(U_0)}(b),$$

and the latter represents the transition probability from $\lambda(U_0)$ to $\lambda(U_1)$ inside $\mathscr{L}\mathscr{M}_{A,A}$.

https://doi.org/10.1017/S0963548303005893 Published online by Cambridge University Press

4. Stationary measures and entropy

4.1. Stationary measures from finite data

Suppose we are given a labelled Markov tree (π, ξ) and a distinguished height L. We consider all the labelled Markov trees 'descending' (ascending) from (π, ξ) up to the Lth generation: $(\pi^u, \xi^u), 0 \leqslant l(u) \leqslant L$. Assign to each of these a weight $(L+1)^{-1} q_\pi(u)$; we obtain a measure on $\mathscr{LM}_{A,B}$:

$$\mu_{(\pi,\xi;L)} = \frac{1}{L+1} \sum_{0 \leqslant l(u) \leqslant L} q_\pi(u) \delta_{(\pi^u, \xi^u)}. \tag{4.1}$$

Since $\sum_{l(u)=j} q_\pi(u) = 1$ for each j, $0 \leqslant j \leqslant L$, (4.1) defines a probability distribution. Using equations (2.8), (3.2) and (3.3), we find

$$\begin{aligned} P^* \mu_{(\pi,\xi;L)} &= \frac{1}{L+1} \sum_{0 \leqslant l(u) \leqslant L} q_\pi(u) P^* \delta_{(\pi^u, \xi^u)} \\ &= \frac{1}{L+1} \sum_{0 \leqslant l(u) \leqslant L} q_\pi(u) \sum_{a \in A} q_{\pi^u}(a) \delta_{(\pi^{ua}, \xi^{ua})} \\ &= \frac{1}{L+1} \sum_{0 \leqslant l(u) \leqslant L} \sum_{a \in A} q_\pi(ua) \delta_{(\pi^{ua}, \xi^{ua})} \\ &= \mu_{(\pi,\xi;L)} + \frac{1}{L+1} \sum_{l(u)=L+1} q_\pi(u) \delta_{(\pi^u, \xi^u)} - \frac{1}{L+1} \delta_{(\pi,\xi)}. \end{aligned} \tag{4.2}$$

From (4.2) we see that as $L \to \infty$, the measure $\mu_{(\pi,\xi,L)}$ is asymptotically invariant under P^*. Thus, for any sequence $\{(\pi_n, \xi_n, L_n)\}$ with $L_n \to \infty$, any weak limit of $\{\mu_{(\pi_n,\xi_n,L_n)}\}$ will be a stationary measure on $\mathscr{LM}_{A,B}$.

Now let (π_n, ξ_n) denote finite labelled Markov trees of uniform height L_n. We extend each of them arbitrarily to an infinite labelled Markov tree $(\hat{\pi}_n, \hat{\xi}_n)$ and we consider a weak limit μ^* of $\{\mu_{(\hat{\pi}_n,\hat{\xi}_n,L_n)}\}$.

Proposition 4.1. *If $\{\mu_{(\hat{\pi}_n,\hat{\xi}_n,L_n)}\} \to \mu^*$, then μ^* depends only on $\{(\pi_n, \xi_n)\}$ and is independent of the extensions $\{(\hat{\pi}_n, \hat{\xi}_n)\}$ that are chosen.*

Proof. A dense family of continuous functions on $\mathscr{LM}_{A,B}$ is provided by functions depending on a finite set of coordinates $\{\pi(w), \xi(w); l(w) \leqslant l_0\}$. For the $\delta_{(\hat{\pi}^u, \hat{\xi}^u)}$ appearing in (4.1), these functions depend on the chosen extensions for only $l_0/L_n + 1$ of the total expressions, and since $L_n \to \infty$, $\int f d\mu^*$ for f in this family is independent of the extensions chosen. □

4.2. The entropy of a Markov system

Given a Markov system $(M, p(\cdot,\cdot))$, we associate with each point $x \in M$ the information function

$$H(x) = -\sum_y p(x,y) \log p(x,y) \quad (\log = \log_2), \tag{4.3}$$

which measures the amount of indeterminism in the system at that point.

https://doi.org/10.1017/S0963548303005893 Published online by Cambridge University Press

Definition 7. The entropy of a stationary Markov system is

$$H(M, p(\cdot,\cdot), \mu) = \int_M H(x)\, d\mu(x). \tag{4.4}$$

Theorem 4.2. *Let (π_n, ξ_n) be a sequence of finite discriminately labelled Markov trees with uniform heights $L_n \to \infty$, and let*

$$\mu^* = \lim \mu_{(\hat{\pi}_n, \hat{\xi}_n; L_n)} \tag{4.5}$$

as above. Then the entropy of the Markov system in $\mathscr{LM}_{A,B}$, with stationary measure μ^* is given by

$$H(\mathscr{LM}_{A,B}, \mu^*) = -\lim \frac{1}{L_n + 1} \sum_{l(w)=L_n} q_{\pi_n}(w) \log q_{\pi_n}(w). \tag{4.6}$$

In this formulation, we have suppressed the canonical transition probabilities given by (3.7).

Proof. The theorem follows by evaluating $\int H(x) d\mu_{(\hat{\pi}, \hat{\xi}; L)}(x)$ according to (4.1).

For any (π', ξ'),

$$H(\pi', \xi') = -\sum_{a \in A} q_{\pi'}(a) \log q_{\pi'}(a),$$

and applying this to the points $(\hat{\pi}^u, \hat{\xi}^u)$ of (4.1), we see that only for $l(u) = L$ does $H(\hat{\pi}^u, \hat{\xi}^u)$ depend on the extension of (π, ξ). Now

$$\begin{aligned}
&\frac{1}{L+1} \sum_{0 \leqslant l(u) < L} q_\pi(u) H(\pi^u, \xi^u) \\
&\quad = \frac{1}{L+1} \sum_{0 \leqslant l(u) < L} q_\pi(u) \left\{ \sum_{a \in A} \frac{q_\pi(ua)}{q_\pi(u)} \log \frac{q_\pi(u)}{q_\pi(ua)} \right\} \\
&\quad = \frac{1}{L+1} \sum_{0 \leqslant l(u) < L} \left\{ q_\pi(u) \log q_\pi(u) - \sum q_\pi(ua) \log q_\pi(ua) \right\} \\
&\quad = -\frac{1}{L+1} \sum_{l(w)=L} q_\pi(w) \log q_\pi(w),
\end{aligned}$$

and this gives the desired result. □

5. Endomorphic systems and multiple recurrence

5.1. Endomorphic Markov systems

A special class of Markov systems, that we will call *endomorphic Markov systems*, will play a key role in our multiple recurrence theorem. These are systems $(M, p(\cdot,\cdot))$ which may be characterized by a continuous, single valued transformation on the state space M, $S : M \to M$ with the property that $p(x, y) > 0$ implies $S(y) = x$.

https://doi.org/10.1017/S0963548303005893 Published online by Cambridge University Press

The transformation S by itself does not determine the values of $p(x,y)$; however, if we add to the data (M,S) a P-stationary measure μ, then for μ, a.e. x, $p(x,y)$ is determined. To see this recall that $P : C(M) \to C(M)$ is defined by

$$(Pf)(x) = \sum_y P(x,y) f(y),$$

and if we define $S : C(M) \to C(M)$ by

$$(Sf)(x) = f(S(x)),$$

then we can compute for $f, g \in C(M)$:

$$\begin{aligned} P\left((Sf)g\right)(x) &= \sum_y p(x,y) f\left(S(y)\right) g(y) \\ &= \sum_y p(x,y) f(x) g(y) \\ &= \left(f \cdot Pg\right)(x), \end{aligned}$$

to obtain the formula $P(Sf \cdot g) = f \cdot Pg$. In particular, taking $g = \mathbf{1}$ and using the fact that $P\mathbf{1} = \mathbf{1}$, we see that $PS = I$. If μ is a stationary measure then we can integrate the formula we obtained to see that

$$\int Sf \cdot g \, d\mu = \int f \cdot Pg \, d\mu,$$

which shows that P and S are adjoint operators in $L^2(M, d\mu)$. This shows that S and μ determine the operator P and thus the transition kernel $p(x,y)$, on the support of the measure μ.

While not all Markov systems are endomorphic Markov systems – indeed the finite Markov chain example treated above is not – any Markov system has an extension that is an endomorphic system. It is this universality that will enable us to reduce the recurrence theorem to endomorphic systems. We will show this later on.

Let us denote by H_0 the full Hilbert space $L^2(M,\mu)$, with μ a fixed stationary measure, and set $H_n = S^n(H_0)$. Since S is multiplicative, H_n is characterized by a σ-algebra, say $\mathscr{B}_n$, and the orthogonal projection Q_n, onto H_n, is given by the conditional expectation operator onto $L^2(M, \mathscr{B}_n, \mu)$. We denote by H_∞ the intersection $\bigcap_{n=1}^\infty H_n$, and $\mathscr{B}_\infty = \bigcap_{n=1}^\infty \mathscr{B}_n$.

On H_∞, P is an invertible operator with S its inverse. This is because any $f \in H_\infty$ is of the form $f = Sf_1$ for some $f_1 \in H_\infty$, and then since $PS = I$ we conclude that

$$Pf = PSF_1 = f_1.$$

Thus P maps H_∞ to H_∞, and applying S we get $SPf = Sf_1 = f$, or $SP = I$.

Denote by M_∞ the factor measure space $M/\mathscr{B}_\infty$, and by $\Pi : M \to M_\infty$ the canonical projection. On M_∞, S defines an invertible mapping and we will denote S^{-1} by T. We will now see that M is *relatively mixing* over M_∞.

This relative mixing follows from the fact that $Q_n f \to Q_\infty f$, from which it follows that if $f \perp H_\infty$ then $\|Q_n f\|_{L^2} \to 0$. Now we can verify that $Q_n = S^n P^n$, by checking that $S^n P^n$

https://doi.org/10.1017/S0963548303005893 Published online by Cambridge University Press

is a self-adjoint projection and for $f \in H_n$, $f = S^n f_0$ for some f_0, and thus

$$S^n P^n f = S^n P^n S^n f_0 = S^n (P^n S^n) f_0 = S^n f_0 = f.$$

Hence $S^n P^n$ is the identity on H_n and equals Q_n.

Now for $f \perp H_\infty$, $\|Q_n f\|_{L^2} \to 0$ can be written as $\|S^n P^n f\|_{L^2} \to 0$, and since μ is stationary, we can apply P^n without changing the norm and finally deduce

$$\|P^n f\|_{L^2} \to 0 \quad \text{for all } f \perp H_\infty,$$

which is to say that P is mixing relative to M_∞.

5.2. Multiple recurrence for endomorphic systems

We continue to assume that (M, P, μ, S) is an endomorphic Markov system with stationary measure μ. Suppose that $\{A, B_0, B_1\}$ are sets of positive μ-measure such that

$$\varphi = 1_A \cdot P 1_{B_0} \cdot P 1_{B_1} \tag{5.1}$$

is positive on A. This means that for all $x \in A$ there is a positive probability of passing both to B_0 and to B_1. By a binary tree of height k and branch length n based on $\{A, B_0, B_1\}$, we will mean a set of vertices $\{x_u \in A : u \in \bigcup_{j=0}^{k} \{0,1\}^j\}$ with the edges $\{x_u, x_{ui}\}$, $i = 0, 1$, indicating that there is a positive probability of going from x_u to x_{ui} in exactly n steps with the first one being to a point in B_i. We proceed now to show how the assumption (5.1) implies the existence of binary trees of height k and some branch length n based on $\{A, B_0, B_1\}$. This is the basic multiple recurrence theorem for Markov systems from which we will deduce our results.

The constructibility of the first level of the tree will follow from the positivity of

$$1_A P\left(1_{B_0} P^{n-1} \varphi\right) P\left(1_{B_1} P^{n-1} \varphi\right).$$

This part of the argument uses Poincaré recurrence. Namely, using the representation $S^n P^n = Q_n$ we have, formally, $P^n = S^{-n} Q_n$. On H_∞, S is invertible, as we have seen with $S^{-1} = T$. For large n, Q_n is close to Q_∞ so that we can approximate as follows: set $\varphi_1 = \varphi$, and assume $\|Q_{n-1}\varphi_1 - Q_\infty \varphi_1\| < \epsilon$. Then

$$\left\|S^{n-1} P^{n-1} \varphi_1 - S^{n-1}\left(S^{-(n-1)}\right) Q_\infty \varphi_1\right\| < \varepsilon,$$

and since S is an isometry we conclude that

$$\left\|P^{n-1}\varphi - S^{-(n-1)} Q_\infty \varphi_1\right\| < \varepsilon.$$

Defining

$$\varphi_2 = 1_A P\left(1_{B_0} P^{n-1} \varphi_1\right) P\left(1_{B_1} P^{n-1} \varphi_1\right),$$
$$\psi_2 = 1_A P\left(1_{B_0} S^{-(n-1)} Q_\infty \varphi_1\right) P\left(1_{B_1} S^{-(n-1)} Q_\infty \varphi_1\right),$$

the above shows that φ_2 and ψ_2 are rather close (as $n \to \infty$). Recalling that $P(1_B S g) = g \cdot P 1_B$, we get that

$$\psi_2 = 1_A (T^n (Q_\infty \varphi_1))^2 \cdot P 1_{B_0} P 1_{B_1} = \varphi_1 \cdot (T^n Q_\infty \varphi_1)^2.$$

https://doi.org/10.1017/S0963548303005893 Published online by Cambridge University Press

It will be convenient to write $\varphi_\infty = Q_\infty \varphi_1$, and recalling that $T = S^{-1}$ is a point mapping, we have

$$\psi_2 = \varphi_1 T^n(\varphi_\infty^2).$$

By Poincaré recurrence, the integral of this is positive for appropriate n. The multiple recurrence extension of Poincaré's theorem allows us to continue.

We have seen that for n large, if R denotes the operator

$$Rf = 1_A \cdot P\left(1_{B_0} P^{n-1} f\right) \cdot P\left(1_{B_1} P^{n-1} f\right),$$

then $\varphi_2 = R\varphi_1$ is approximately equal to ψ_2 as above. Repeating this calculation, we see that the $R\varphi_2$ which we define to be φ_3 is close to $R\psi_2$, which may be approximated as before in the following way:

$$\begin{aligned} R\psi_2 &= 1_A P\left\{1_{B_0} P^{n-1}\left(\varphi_1 T^n \varphi_\infty^2\right)\right\} P\left\{1_{B_1} P^{n-1}\left(\varphi_1 T^n \varphi_\infty^2\right)\right\} \\ &\overset{\varepsilon}{\approx} 1_A P\left\{1_{B_0} T^n S Q_\infty\left(\varphi_1 T^n \varphi_\infty^2\right)\right\} P\left\{1_{B_1} T^n S Q_\infty\left(\varphi_1 T^n \varphi_\infty^2\right)\right\} \\ &= 1_A P 1_{B_0} \cdot P 1_{B_1} \cdot T^n \varphi_\infty^2 T^{2n} \varphi_\infty^4 = \varphi_1 \cdot T^n \varphi_\infty^2 T^{2n} \varphi_\infty^{2^2}. \end{aligned}$$

This latter expression we define to be ψ_3, and thus φ_3 is close to ψ_3. This will be repeated up to φ_n, ψ_n.

Since T is a point mapping and the $\psi_\infty^2, \psi_\infty^4, \psi_\infty^8, \ldots, \psi_\infty^{2^n}$ are all powers of the same positive function, we may apply the ergodic Szemerédi theorem to conclude that ψ_n, and thus φ_n, has a positive integral which gives the desired multiple recurrence theorem.

5.3. Multiple recurrence for Markov systems

Suppose $(M, p(\cdot,\cdot))$ and $(M', p'(\cdot,\cdot))$ are two Markov systems with stationary measures μ and μ', respectively. We say $(M', p'(\ ,\ ))$ is a *factor* of $(M, p(\cdot,\cdot))$ if there is a continuous map $\varphi : M \to M'$ with

$$p'(\varphi(x), \varphi(y)) = p(x, y)$$

whenever $p(x, y) > 0$, and with φ taking μ to μ'.

Proposition 5.1. *Every stationary Markov system is a factor of an endomorphic Markov system.*

Proof. Let $(M, p(\cdot,\cdot))$ be a Markov system with stationary measure μ. Let $\widetilde{M}$ be the space of 'paths':

$$\widetilde{M} = \cdots M \times M \times \cdots \times M = M^{\mathbb{Z}^-}$$

representing the past of the stationary M-valued process $\{W_n\}$ of Section 3.1. We let $\widetilde{\mu}$ be the probability measure on $M^{\mathbb{Z}^-}$ associated to the process $\{W_n\}_{n \leqslant 0}$. We take the transition probabilities $\widetilde{p}(\omega, \omega')$ to be those connected to the stationary process $\{W_n\}$:

$$\widetilde{p}(\omega, \omega') = p(x_-, x_0)$$

https://doi.org/10.1017/S0963548303005893 Published online by Cambridge University Press

provided $\omega' = \dots x_{-2}, x_{-1}, x_0$ and ω is the shift of ω': $\omega = \dots x_{-3}, x_{-2}, x_{-1}$, and $\tilde{p}(\omega, \omega') = 0$ in every other case. The measure $\tilde{\mu}$ is a stationary measure for $(\widetilde{M}, \tilde{p}(\cdot,\cdot), S)$, and moreover, letting $S : \widetilde{M} \to \widetilde{M}$ be the right shift, we see that $(\widetilde{M}, \tilde{p}(\cdot,\cdot), S)$ is an endomorphic system. Finally, setting

$$\varphi(\dots x_{-2}, x_{-1}, x_0) = x_0,$$

we find that $(M, p(\cdot,\cdot))$ is a factor of $(\widetilde{M}, \tilde{p}(\cdot,\cdot))$. □

It is now a straightforward matter to deduce a multiple recurrence theorem for any Markov system from the corresponding theorem for endomorphic systems.

Theorem 5.2. *Let $(M, p(\cdot,\cdot), \mu)$ be a stationary Markov system, A, B_0, B_1 Borel subsets of M with $\mu(A) > 0$ and with $p(x, B_0) > 0$ for each $x \in A$, and also $p(x, B_1) > 0$ for each $x \in A$ (here $p(x, B)$ denotes $\sum_{y \in B} p(x, y)$). Then for any k there exists an n and a binary tree of height k and branch length n based on A, B_0, B_1.*

Proof. Take $(\widetilde{M}, \tilde{p}(\cdot,\cdot), \tilde{\mu})$ as the endomorphic system of which $(M, p(\cdot\cdot), \mu)$ is a factor, and let $\varphi : \widetilde{M} \to M$ be the associated map. Set $\widetilde{A} = \varphi^{-1}(A), \widetilde{B}_0 = \varphi^{-1}(B_0)$ and $\widetilde{B}_1 = \varphi(B_1)$. If $\omega \in \widetilde{A}, \omega = \dots x_{-1}, x_0$, then $x_0 \in A$ and there exists $y \in B_0$ with $p(x_0, y) > 0$. Let $\omega' = \dots x_{-1}, x_0, y$; then $\omega' \in \widetilde{B}_0$ and $p(\omega, \omega') > 0$. Similarly for $\widetilde{B}_1$. We obtain a binary tree in $\widetilde{M}$, and applying φ we obtain the desired tree in M. □

6. Existence of arithmetic binary trees

We present the proof of Theorem B. Our alphabet A has two letters $\{a_0, a_1\}$ and we suppose that for arbitrarily large L_n we can find a subset $S_n \subset A^*_{L_n} = \bigcup_{j \leqslant L_n} A^j$, each S_n having density $\geqslant \delta$, and such that no S_n has an arithmetic binary subtree of height k.

For each n we construct a finite labelled Markov tree. On the finite tree $A^*_{L_n}$ we define probabilities $g(w) = 2^{-l(w)}$ converting it to a Markov tree π_n. Now, for any word $w \neq \varnothing$, let $t(w)$ denote the terminal letter of w, $t(a_{i_1}, a_{i_2} \dots a_{i_l}) = a_{i_l}$ and set $t(\varnothing) = a_0$. Take $B = \{0, 1\} \times A$ and set

$$\xi_n(w) = (1_{S_n}(w), t(w)). \tag{6.1}$$

Note that this labelling is discriminating. We form the space $\mathscr{LM}_{A,B}$ of all the infinite labelled Markov trees (π, ξ). We set

$$E = \{(\pi, \xi) : \xi(\varnothing) = (1, a_0) \text{ or } (1, a_1)\},$$
$$F_0 = \{(\pi, \xi) : \xi(\varnothing) = (0, a_0) \text{ or } (1, a_0)\},$$
$$F_1 = \{(\pi, \xi) : \xi(\varnothing) = (0, a_1) \text{ or } (1, a_1)\}.$$

Finally, we let μ^* be a measure on $\mathscr{LM}_{A,B}$ which appears as a limit of the form (4.5). It is readily seen that $\mu^*(E) \geqslant \delta$, inasmuch as, for each n, S_n has density $\geqslant \delta$. Moreover, the transition probabilities for all the trees obtained from (π_n, ξ_n) are $1/2$ and so μ^* is supported on (π, ξ) with $p((\pi, \xi), F_0) = p((\pi, \xi), F_1) = 1/2$. We can now invoke Theorem 5.2

https://doi.org/10.1017/S0963548303005893 Published online by Cambridge University Press

to find a binary tree of points based on E, F_0, F_1. This binary tree can be approximated by subtrees of the (π_n, ξ_n) and this contradicts our hypothesis negating the validity of Theorem B.

The proof of Theorem C proceeds along similar lines. We are now working in an alphabet A with s letters, and we suppose that we have trees T_n of uniform height $L_n \to \infty$ and branching $2^{\beta_n} \geqslant 2^{\alpha}$ not containing arithmetic binary trees of height k. We convert T_n into a Markov tree π_n, setting $p_{\pi_n}(w) = 2^{-\beta_n L_n}$ for words w of the maximal length L_n, and defining $q_{\pi_n}(w)$ for shorter words by (2.10). We take $B = A$ and set $\xi_n(w) = t(w)$. We again consider a limiting measure μ^* of the form (4.5). We note that by Theorem 4.2,

$$H(\mathscr{LM}_{A,B}, \mu^*) = -\lim \frac{1}{L_n + 1} 2^{\beta_n L_n} \cdot 2^{-\beta_n L_n} \cdot (-\beta_n L_n) \geqslant \alpha. \tag{6.2}$$

This implies that for positive μ^*-measure, we will have that the probability vector $(q_\pi(a))_{a \in A}$ is not degenerate, and in particular, we can find $a, b \in A$ with $q_\pi(a) > 0$ as well as $q_\pi(b) > 0$. Let $E = \{(\pi, \xi) : q_\pi(a) > 0 \text{ and } q_\pi(b) > 0\}$. Now let $F_a = \{(\pi, \xi) : \xi(\varnothing) = a$, and $F_b = \{(\pi, \xi) : \xi(\varnothing) = b$; then, by definition, for (π, ξ) in E the transitions to F_a and F_b have positive probability. We again invoke Theorem 5.2 to obtain a binary tree of points based on E, F_a, F_b and we are led to a contradiction. This establishes Theorem C. □

References

[1] Furstenberg, H. (1970) Intersections of Cantor sets and transversality of semigroups. In *Problems in Analysis* (R. Gunning, ed.) Princeton University Press, pp. 41–59.

[2] Furstenberg, H. (1981) *Recurrence in Ergodic Theory and Combinatorial Number Theory*, Princeton University Press, Princeton.

[3] Graham, R., Rothschild, B. and Spencer, J. (1980) *Ramsey Theory*, Wiley, New York.

https://doi.org/10.1017/S0963548303005893 Published online by Cambridge University Press