The Annals of Probability
1992, Vol. 20, No. 3, 1549–1566

AUTOMORPHISM INVARIANT MEASURES ON TREES[1]

BY ROBIN PEMANTLE

Cornell University

Consider a collection of real-valued random variables indexed by the integers. It is well known that such a process can be stationary, that is, translation invariant, and ergodic and yet have very strong associations: The one-sided tail field may determine the sample; the measure may fail to be mixing in any sense; the weak law of large numbers may fail on some infinite subset of the integers. The main result of this paper is that this cannot happen if the integers are replaced by an infinite homogeneous tree and the translations are replaced by all graph automorphisms. In fact, any automorphism-invariant process indexed by the tree is a mixture of extremal processes whose one-sided tail fields are trivial, from which the mixing properties follow.

1. Introduction. An infinite, homogeneous n-ary tree is a loopless undirected graph in which every vertex has precisely $n + 1$ neighbors. For example, a unary tree ($n = 1$) is just the integers with the usual nearest neighbor edges and a binary tree ($n = 2$) looks like it is supposed to as in Figure 1.

This paper considers processes indexed by a homogeneous n-ary tree; throughout the paper, n is at least 2 unless otherwise stated. Without loss of generality, such processes are assumed to take values in the standard measure space $[0, 1]$, that is to say, the process is a measure on the set Ω of functions ξ: $\mathbf{T} \to [0, 1]$. The reason no generality is lost is that the main result (triviality of certain tail fields) is measure theoretic, so $[0, 1]$ could be replaced by any measure space without affecting the results. The convenience of $[0, 1]$ is simply to be able to take expectations directly instead of taking expectations of indicator functions.

One class of examples of tree-indexed processes is the class of Markov random fields [10]. Spitzer's 1975 paper [16] is a thorough study of Markov random fields on homogeneous trees gotten by extending integer-indexed reversible Markov chains. Higuchi [9] has elaborated on this for ferromagnetic Ising models on a homogeneous tree, giving representations of measures satisfying local conditional probability criteria as limits of Gibbs states with specified boundary conditions. Criteria for the existence of multiple Gibbs states with the same local conditional probabilities (phase transition) are given; see also [12] for results on the Ising model on a general tree. Another set of examples, from where my interest in this problem derived initially, is the

Received November 1989; revised January 1991.
[1]Research supported by a NSF postdoctoral fellowship and by a Mathematical Sciences Institute postdoctoral fellowship.
AMS 1980 *subject classifications*. Primary, 05C25, 28D99, 60K35.
Key words and phrases. Stationary, tree, exchangeable, partially exchangeable, mixing, tail field.

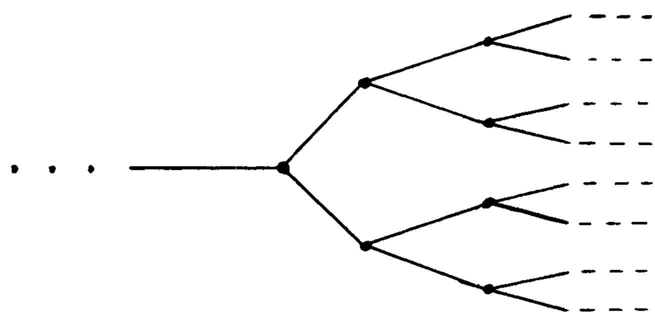

FIG. 1. *Binary tree.*

invariant measures for interacting particle systems. Classically, interacting particle systems have been studied on the integer lattices $\mathbf{Z}^n$ [11], [6]. More recently, there have been studies of the contact process on trees [13] and unoriented percolation on a tree times $\mathbf{Z}$ [8].

This paper is concerned with measures μ that are invariant under graph automorphisms of $\mathbf{T}$ (formally defined in the next section). The stationary measures for particle systems (e.g., the upper invariant measure for the contact process) are often automorphism-invariant because the rules for evolution are. In the case of the Ising model, the stationary distributions are Markov random fields and are automorphism-invariant as long as their local characteristics are. At any rate, measures that are invariant under graph automorphisms are natural objects of study and the object of this paper is to shed some light on them.

What kind of general results can we expect? Consider for comparison the case of random variables $\{X_i: i \in \mathbf{Z}\}$ indexed by the integers. Suppose the X_i are *exchangeable*, that is, their joint distribution is unaffected by permuting the indices. Then de Finetti's famous theorem ([7], Chapter VII, Section 4) says that the X_i are distributed as a mixture of i.i.d. random variables. In particular, an extremal element in the class of exchangeable measures on $\mathbf{R}^{\mathbf{Z}}$ is just an i.i.d. product measure. On the other hand, suppose only that the variables X_i are *stationary reversible*, that is, their joint distribution is unaffected by the permutations of the indices $i \mapsto i + 1$ and $i \mapsto -i$. Now there is no neat characterization of the extremal measures and very little can be said about the presence or lack of long range correlations. As an example of what bad behavior may occur, consider the following process. Let X_0 be uniform on $[0, 1)$ and let $X_j = X_0 + j\alpha_{\bmod 1}$, where α is some fixed irrational number. Let $Y_i = 1$ if $X_i \in [0, 1/10)$ and 0 otherwise. It is easy to see that the Y process is an extremal stationary reversible process but that it is not mixing and is in fact completely determined by its tail. Between exchangeability and stationarity lie various notions of *partial exchangeability*, invariance under groups of permutations of intermediate size. The more invariance you assume, the stronger the characterization you obtain; see [1] or [5] for further examples.

Returning to the case of a homogeneous n-ary tree, note that when $n = 1$, the tree is just $\mathbf{Z}^1$ and automorphism-invariance is just stationarity and reversibility, so we get nothing new. In contrast to this is the infinitary tree ($n = \infty$). Aldous [1] uses exchangeability to characterize the extremal automorphism-invariant processes; these turn out to be functions of reversible Markov chains (any transition kernel for a reversible Markov chain has a natural realization as a tree-indexed process, also called a Markov chain; see the example in Section 5). Mixing and triviality of a certain tail field follow from the Markov characterization.

The main result of this paper is to show that in fact for any $n \geq 2$, automorphism-invariance and extremality imply that the one-sided tail fields (as defined in the next section) are trivial. A mixing property follows from this as does a weak law of large numbers. The preliminary work for this result will include a discussion of what can and cannot be done in an automorphism-invariant way on the boundary of a tree and on the *horocycles* of a tree.

Horocycles on a tree are a discrete analogue of horocycles on a Riemannian surface which are circles centered at infinity. In fact much of the previous work that has been done regarding tree automorphisms and harmonic functions on trees has been motivated by the tree structure of $SL_1(K)$ for finite fields, which is a discrete analogue of the Riemannian structure on $SL_2(\mathbf{R})$. In this regard, see, for example, [15] and [3]. Since harmonic functions on any graph have a random walk interpretation, study of harmonic functions on trees and Martin boundaries has appeared in the context of random walk problems (see, for example, [14], in which it is shown among other things that the Martin potential kernel on vertices v for a particular boundary point of a homogeneous tree depends only on which horocycle centered at that boundary point contains v). The results and methods in this paper are not explicitly based on consideration of SL_2, but the use of random walk arguments in the proofs of Theorems 2 and 3 bears unmistakable resemblance to the previous work of Cartier [3], Chen [4] and others.

Note added in proof. I have recently learned from Shahar Mozes and Alex Lubotzky (personal communication) that the main result of this paper may be proved by a completely different method, namely, by adapting the Howe–Moore theorem to the group of even automorphisms of a homogeneous tree.

2. Definitions and results. Let $\mathbf{T}$ be a homogeneous n-ary tree, in other words, every vertex (also called a node) has $n + 1$ neighbors and there are no loops. Let $\overline{xy}$ denote the unique path connecting x to y, so the graph distance $d(x, y)$ is the number of edges in the path $\overline{xy}$. As a graph $\mathbf{T}$ is bipartite, which means that the set of vertices of $\mathbf{T}$ divides into two equivalence classes, where $x \sim y$ if $d(x, y)$ is even. It will be helpful to label the classes (arbitrarily) EVENS and ODDS. For any set $A \subseteq \mathbf{T}$, define the hull of A, denoted hull(A), to be $\bigcup_{x, y \in A} \overline{xy}$.

For any node x, define a *ray from* x to be a sequence $x, x_1, x_2, \ldots$ such that $d(x_i, x_{i+1}) = 1$ and $x_{i+2} \neq x_i$ for all i. (Here and after, x_0 will always mean

x.) The set of rays $\mathscr{L}$ from a fixed node x can be given a natural measure m_x, defined by $m_x\{\mathscr{L}: y \in \mathscr{L}\} = (n+1)^{-1} n^{1-d(x,y)}$ for each y. This measure is called the *Hausdorff measure at* x since it is the Hausdorff measure for the metric d_x on rays from x defined by $d(\mathscr{L}_1, \mathscr{L}_2) = (n+1)^{-1} n^{-k}$, where k is the number of vertices other than x shared by $\mathscr{L}_1$ and $\mathscr{L}_2$. Hausdorff measure is defined on the σ-field $\mathscr{B}_x$ generated by these sets as y varies (this is just the Borel σ-field with respect to the sequence topology). Another useful way of thinking of m_x is as the Poisson measure for the symmetric random walk potential kernel [3].

The boundary of the tree, denoted $\partial\mathbf{T}$, is a natural object that can be defined (following [17]) as the set of equivalence classes of rays where two are equivalent if their set theoretic difference has finite cardinality. For any fixed vertex x, the boundary is in one-to-one correspondence with the set of rays from x. From this correspondence, $\partial\mathbf{T}$ inherits the measure m_x on the σ-field $\mathscr{B} = \mathscr{B}_x$, the subscript being dropped because all σ-fields $\mathscr{B}_x$ are identical on $\partial\mathbf{T}$. Each m_x is absolutely continuous with respect to each other, so there are well-defined null sets on $\mathscr{B}$.

The next definition follows [3] but in different notation. For any $\alpha \in \partial\mathbf{T}$ and any vertex x, define the *horocycle in direction* α *through* x, denoted $h(\alpha, x)$, to be the set of vertices y such that $d(x,z) = d(y,z)$ for all but finitely many vertices z on any ray in α. (Terminology is from hyperbolic geometry where a horocycle is a circle centered at infinity.) Write $\mathscr{H}$ for the set of horocycles. The horocycles in direction α are a partition of vertices of $\mathbf{T}$ into equivalence classes. The classes have the same order type as the integers in the sense that each horocycle h in direction α has a successor, namely the horocycle of neighbors of elements of h that are in rays from α from points in h. Since every horocycle is a subset of either EVENS or ODDS, horocycles are termed even and odd accordingly. See Figure 2 for an illustration of some of these definitions.

Let $\Omega = [0,1]^{\mathbf{T}}$ denote the set of $[0,1]$-valued functions on $\mathbf{T}$ and let $\mathscr{F}$ be the usual Borel σ-field on Ω making each coordinate measurable. Typical elements of Ω are denoted ξ. Let AUT be the group of graph-automorphisms of $\mathbf{T}$, which is to say that $\pi \in$ AUT if and only if π is a bijection from the vertices of $\mathbf{T}$ to the vertices of $\mathbf{T}$ and $\pi(x)$ neighbors $\pi(y)$ if and only if x neighbors y. Elements of AUT can be classified according to whether they fix at least one vertex, permute the two endpoints of an edge or act as a nondegenerate translation along a doubly-infinite path ([17], Proposition 3.2). Let G be the subgroup of index two in AUT (called AUT^+ in [17]) generated by permutations π that fix at least one vertex. Alternatively, $\pi \in G$ if and only if π preserves the parity classes, that is, π(EVENS) = EVENS and π(ODDS) = ODDS.

There is a natural action of G on Ω defined by $\pi(\xi) = \xi \circ \pi^{-1}$. Say a probability measure μ on $\mathscr{F}$ is G-invariant if $\mu(A) = \mu\{\pi(\xi): \xi \in A\}$ for all $A \in \mathscr{F}$ and $\pi \in G$. Use the notation $\mathbf{P}_\mu(\xi \cdots)$ to denote $\mu\{\xi: \cdots\}$.

This paper is concerned with probability measures on $\mathscr{F}$ that are invariant under the entire group of even automorphisms, G. Since G is uncountable,

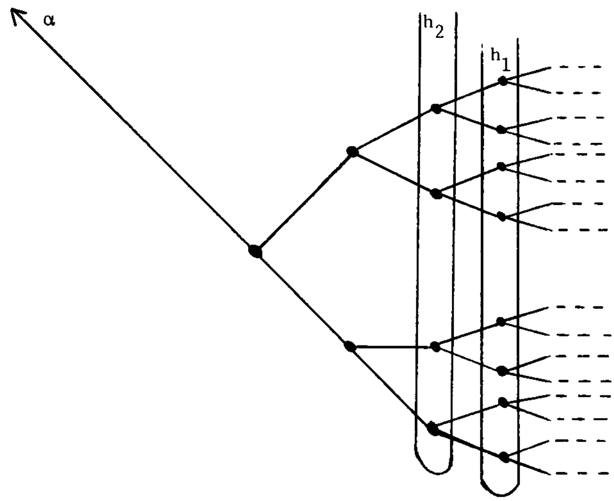

FIG. 2. *h_1 and h_2 are horocycles in direction α with h_2 the successor of h_1. Both extend infinitely upward.*

this seems like a lot to ask so here is a way to find a countable subgroup $G_0 \subseteq G$ such that any measure on $\mathscr{F}$ invariant under G_0 is invariant under the whole group G. List all pairs of finite sequences of vertices which can be mapped to each other by elements of G and choose for each pair such a map. Then G_0 is the group generated by these selected maps. Let μ be a G_0-invariant measure on $\mathscr{F}$. Let $A \in \mathscr{F}$ be a cylinder event, that is, one that depends on finitely many values of ξ. Any map $\pi \in G$ is the pointwise limit of maps in $\pi_k \in G_0$. Since π_k is eventually constant on A and μ is π_k-invariant, it follows that $\mu(A) = \mu(\pi(A))$. The cylinder events determine the measure, so μ must be G-invariant.

The space $\mathscr{M}_G$ of G-invariant probability measures is a closed, convex subset of the linear space of measures on Ω, which is compact in the weak topology. It is easy to check that the conditions of the Krein–Milman theorem are satisfied and hence that $\mathscr{M}_G$ is the closed convex hull of its extreme points (i.e., those G-invariant probability measures that are not convex combinations of other distinct G-invariant probability measures). In particular, there is no generality lost in restricting attention to extremal elements of $\mathscr{M}_G$, since every element of $\mathscr{M}_G$ has representation as an integral of extremal elements of $\mathscr{M}_G$ and the main results of this paper for extremal G-invariant measures are easily translated into results for general G-invariant measures.

REMARK. It may seem strange to limit G to those maps preserving parity, since then the measure concentrating on ξ_e is G-invariant, where ξ_e is 1 on EVENS and 0 on ODDS. The alternative is to allow the full group of tree automorphisms, in which case the measure that puts mass 1/2 on ξ_e and 1/2 on $1 - \xi_e$ is an extremal G-invariant measure. Then the tail fields in the main

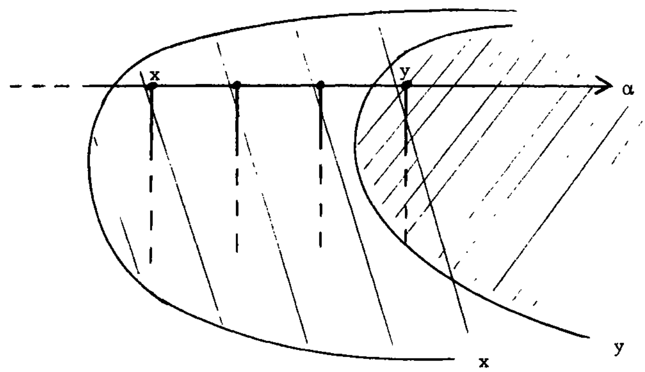

FIG. 3. *F_α is the intersection of σ-fields such as F_x and F_y, as the σ-fields move farther and farther to the right.*

theorem would have one bit of information instead of being trivial; I find this even less aesthetic than restricting G.

Several kinds of σ-fields might reasonably be called the tail field. Define the big tail field $\mathscr{F}^*$ to be the intersection over all finite sets of the σ-field of information given by the configuration outside the finite set. This is an analogue of the two-sided tail on the integers. There is also an analogue of the one-sided tail field, only instead of there being two, there is one for each boundary point. For each vertex x and each ray $\mathscr{L}$ from x, let $\mathscr{F}(\mathscr{L})$ be the σ-field of information at all vertices z for which $\overline{zx}$ intersects $\mathscr{L}$ in some point other than x. For each $\alpha \in \partial \mathbf{t}$, define $\mathscr{F}(\alpha)$ to be the intersection over rays $\mathscr{L}$ in α of $\mathscr{F}(\mathscr{L})$. Call $\mathscr{F}(\alpha)$ the tail in direction α. The main result can now be stated.

THEOREM 1. *Let μ be a probability measure on the Borel subsets of Ω that is invariant under G. Suppose μ is extremal in that class. Then for any $\alpha \in \partial \mathbf{T}$, the tail $\mathscr{F}(\alpha)$ in direction α is trivial.*

Consequences of this theorem are that extremal invariant measures on $\mathscr{F}$ are mixing, that events become independent uniformly as they become distant and that there is a very general weak law of large numbers that holds over large sets of arbitrary shape; these will be spelled out in Section 5.

REMARK. When $n = 2k - 1$, the tree is just the right Cayley graph for the free group on k generators, $FG(k)$. The left action of this group on itself preserves the graph, so $FG(k)$ is a subgroup of the automorphism group of T. It is natural to ask whether Theorem 1 is true with G replaced by the smaller group $FG(k)$. The answer is trivially no: $FG(k)$ has a quotient isomorphic to $\mathbf{Z}$, so any behavior on $\mathbf{Z}$ such as extremality plus nontrivial tails may be lifted to $FG(k)$.

3. Preliminary results on random functions. The first idea in the proof of Theorem 1 is that random functions on $\partial\mathbf{T}$ whose distributions are G-invariant are almost surely constant, that is to say, with probability 1, they are constant modulo null sets of $\partial\mathbf{T}$. For this section, let $(\Omega, \mathscr{F})$ be any measure space with probability measure μ and revert to ω instead of ξ for sample points. Let $f\colon (\Omega, \mathscr{F}) \times (\partial\mathbf{T}, \mathscr{B}) \to \mathbf{R}$ be any bounded jointly measurable function. Then $f(\omega, \cdot\,)$ is what I call a random bounded measurable function on $\partial\mathbf{T}$. Another description for f is as a random field indexed by $\partial\mathbf{T}$. Later, it will be useful to have a notion of almost sure constancy jointly in μ and $\mathscr{B}$, by which will be meant constancy modulo $\mu \times m_x$-null sets for any, hence every x. For $\omega \in \Omega$ and $x \in \mathbf{T}$, let

$$e_x(f, \omega) = \int f(\omega, \alpha)\, dm_x(\alpha)$$

be the average of f on $\partial\mathbf{T}$ in the Hausdorff measure at x. Each e_x is measurable and is hence a bounded, real-valued random variable. The following theorem is a consequence of the fact that G acts ergodically on $\partial\mathbf{T}$. A direct probabilistic proof (similar to an argument in [4] but with a different twist) is included as a warm-up to the proof of Theorem 3.

THEOREM 2. *Let $f\colon (\Omega, \mathscr{F}, \mu) \times (\partial\mathbf{T}, \mathscr{B}) \to \mathbf{R}$ be a bounded random function, that is, bounded and jointly measurable. Suppose that f is G-invariant in the sense that $\mathbf{P}_\mu(e_{x_i}(f, \omega) \in A_i) = \mathbf{P}_\mu(e_{\pi x_i}(f, \omega) \in A_i)$ for any finite collection of x_i in $\mathbf{T}$, Borel sets $A_i \subseteq \mathbf{R}$ and $\pi \in G$. Then f is μ-almost surely constant modulo $\mathscr{B}$-null sets.*

PROOF. From the relation $m_x = (n+1)^{-1}\sum_{y\colon d(x,y)=1} m_y$, it follows that $e_x(f, \omega)$ is harmonic in x for each ω. By G-invariance, $\mathbf{P}_\mu(|e_x f - e_y f| > \varepsilon)$ is the same for any x, y with $d(x,y) = 1$ and $x \in$ EVENS. Let $p(\varepsilon)$ denote this quantity. Let $z_0, z_1, \dots$ be a sample random walk on $\mathbf{T}$ independent of f with $z_0 \in$ EVENS with probability 1. Write ν for the law of the random walk. Now for each ω, $\{e_{z_k}\}$ is a martingale in the random walk filtration, hence converges almost surely. But $P_{\mu\times\nu}(|e_{z_{2k}} - e_{z_{2k+1}}| > \varepsilon)$ is easily seen to be $p(\varepsilon)$ by Fubini's theorem and conditioning on the pair (z_{2k}, z_{2k+1}). The almost sure convergence implies $P_{\mu\times\nu}(|e_{z_{2k}} - e_{z_{2k+1}}| > \varepsilon) \to 0$ as $k \to \infty$, hence $p(\varepsilon) = 0$ for any $\varepsilon > 0$. Thus $\mathbf{P}_\mu(e_x(f, \omega) = e_y(f, \omega)) = 1$ for any neighbors x and y, hence for all vertices of $\mathbf{T}$. Then $e_x(f, \omega)$ is independent of x for almost every ω. When this occurs, $f(\omega, \cdot\,)$ is constant modulo a null set, since the values of e_x determine a function modulo null sets. □

The following is a digression in the sense that it is irrelevant to the eventual proof of Theorem 1. For some fixed $\beta \in \partial\mathbf{T}$, let $f(\omega, \alpha)$ be independent of ω and equal to 1 if $\alpha = \beta$ and 0 otherwise. Then the distribution of f is G-invariant in the sense of Theorem 2, since null sets of $\Omega \times \partial\mathbf{T}$ are ignored and f is indeed almost surely constant modulo null sets. Clearly, however, f

is never constant, so the modulo null sets cannot be dropped. Can this be fixed by strengthening the G-invariance? By requiring the finite-dimensional marginals of f to have G-invariant distributions it is easy to see (exercise for the reader) that f cannot almost surely be constant with a finite exception set. No form of G-invariance, however, is strong enough to force f to be almost surely constant, as shown by the following example. For $x \in \mathbf{T}$, let $(\Omega_x, \mathscr{F}_x, \mu_x) = (\partial\mathbf{T}, \mathscr{B}, m_x)$ and let $(\Omega, \mathscr{F}, \mu)$ be the product of these. A typical $\omega \in \Omega$ is a function from $\mathbf{T}$ to $\partial\mathbf{T}$; the coordinates are independent, each with a different Hausdorff measure for its law. Let $f(\omega, \alpha) = 1$ if there is an $x \in \mathbf{T}$ with $\omega(x) = \alpha$ and 0 otherwise; in other words, f is the indicator function of the union of the single points picked out from Hausdorff measure at each point. Then f is never constant, having almost surely a countably infinite exception set, but the distribution of f is G-invariant in the strongest sense: There is a measure preserving G-action on Ω such that $f(\pi\omega, \pi\alpha) = f(\omega, \alpha)$; the action is defined by $(\pi\omega)(x) = \pi(\omega(\pi^{-1}x))$.

The next job is to replicate Theorem 2 but for the set of horocycles $\mathscr{H}$ instead of $\partial\mathbf{T}$. The argument will be a little bit trickier. Let $\mathscr{H}_x \subseteq \mathscr{H}$ denote the horocycles containing x. Define Borel sets of horocycles to be the sets $A \subseteq \mathscr{H}$ for which each $A \cap \mathscr{H}_x$ is Borel, where Borel subsets of $\mathscr{H}_x$ are defined by the correspondence $\alpha \leftrightarrow h(\alpha, x)$ between $\partial\mathbf{T}$ and $\mathscr{H}_x$. Similarly, a null set is a set N for which each $N \cap \mathscr{H}_x$ corresponds to a null set. A random function is a jointly measurable map from $(\Omega, \mathscr{F}) \times (\mathscr{H}, \mathscr{B})$ to the reals for some space Ω with a probability measure μ on it. For $\omega \in \Omega$ and $x, y \in \mathbf{T}$, let $I(x, y)$ be the set of $\alpha \in \partial\mathbf{T}$ for which $y \in \alpha^x$, in other words, $I(x,y)$ is the set of boundary points past y as seen by x. Let

$$e^*_{xyz}(f, \omega) = \int_{I(x,y)} f(\omega, h(\alpha, z))\, dm_x(\alpha)/m_x(I(x,y))$$

be the average of f on horocycles through z whose directions are in $I(x, y)$. Write e^*_{xy} for e^*_{xyx}.

Theorem 3. *Let f: $(\Omega, \mathscr{F}, \mu) \times (\mathscr{H}, \mathscr{B}) \to \mathbf{R}$ be a bounded random function. Suppose the distribution of f is G-invariant in the sense that the joint distribution of a finite collection $\{e^*_{x_i y_i z_i}\}$ is the same as the joint distribution of $\{e^*_{\pi x_i \pi y_i \pi z_i}\}$. Suppose further that the finite-dimensional marginals of f have G-invariant distribution, that is, the joint distribution of $f(h_i)$ is the same as the joint distribution of $f(\pi h_i)$ for a finite collection of horocycles. Then f is μ-almost surely constant on even horocycles modulo null sets, and the same for odd horocycles.*

Remark. The assumption of G-invariance of the finite-dimensional marginals is unnecessary but is included to shorten the proofs since it holds in the application and in all interesting cases I know of.

Begin with two preliminary lemmas.

LEMMA 4. *Let h_1 and h_2 be any two horocycles that intersect. Then for any $x \in h_1 \cap h_2$, the cardinality of $\mathscr{L}_1 \cap \mathscr{L}_2 \setminus \{x\}$ is some value $c(h_1, h_2)$ independent of x, where $\mathscr{L}_i$ is the ray from x in α_i, the direction of h_i. If $c = 0$, the cardinality of $h_1 \cap h_2$ is 1. If $c > 0$, the cardinality of $h_1 \cap h_2$ is $(n-1)n^{c-1}$. (Recall that n is one less than the degree of the vertices.) There is an automorphism mapping the pair (h_1, h_2) into another pair of intersecting horocycles (h'_1, h'_2) if and only if $c(h_1, h_2) = c(h'_1, h'_2)$.*

PROOF. Figure 4 illustrates the overlap of two horocycles h_1 and h_2 in the two cases $c(h_1, h_2) = 0$ and $c(h_1, h_2) = k > 0$. From these the assertions of the lemma should be clear. □

LEMMA 5. *Let α be any direction and let x and z be any vertices with z at distance two from x and on the ray from x to α. Then for any k, there is a sequence of four horocycles $h_1 = h(\alpha, x)$, h_2, h_3, $h_4 = h(\alpha, z)$ such that $h_i \cap h_{i+1}$ has cardinality at least $(n-1)n^{k-1}$ for $i = 1, 2, 3$.*

PROOF. Let y be the point on the ray from x in α at distance $2k+2$ from x. Let $\beta \in \partial\mathbf{T}$ be a direction for which the ray from x in β intersects the ray from x in α in precisely $k+1$ points and let $\gamma \in \partial\mathbf{T}$ be such that the ray from x in γ intersects the ray from x in α in precisely $k+2$ points. Let $h_2 = h(\beta, x) = h(\beta, y)$ and let $h_3 = h(\gamma, y) = h(\gamma, z)$. This works (see Figure 5). □

PROOF OF THEOREM 3. Fix $x \in \mathbf{T}$ and let N be a positive integer. Let $\alpha = x_0, x_1, \dots$ and $\beta = y_0, y_1, \dots$ be random rays from $x = x_0 = y_0$ with the following properties: (i) the law of α is m_x; (ii) the law of β is m_x; (iii) the law ν of (α, β) is $m_x \times m_x$ restricted to the set where $x_i = y_i$ if and only if $i \le N$; (iv) f is independent of α and β.

Observe that $e^*_{x_0x_k}(f, \omega) = \mathbf{E}_\nu(f(\omega, h(\alpha, x)) | x_0, \dots, x_k)$ and is therefore a martingale in k for fixed ω converging ν-almost surely to $f(\omega, h(\alpha, x) |$ all $x_i) = f(\omega, h(\alpha, x))$. Thus

$$\mathbf{P}_\nu\Big(|e^*_{x_0x_k}(f, \omega) - f(\omega, h(\alpha, x))| > \varepsilon\Big) \to 0 \tag{1}$$

for each $\varepsilon > 0$, $\omega \in \Omega$ as $k \to \infty$. Hence for any $\varepsilon > 0$, there is an $N(\varepsilon)$ such that for $k \ge N(\varepsilon)$,

$$\mathbf{P}_{\nu\times\mu}\Big(|e^*_{x_0x_k}(f, \omega) - f(\omega, h(\alpha, x))| > \varepsilon\Big) < \varepsilon. \tag{2}$$

Now restrict the choice of N above to be at least $N(\varepsilon)$. Then

$$\begin{aligned} &\mathbf{P}_{\nu\times\mu}(|f,(\omega, h(\alpha, x)) - f(\omega, h(\beta, x))| > 2\varepsilon) \\ &\quad \le \mathbf{P}_{\nu\times\mu}\Big(|e^*_{x_0x_N}(f, \omega) - f(\omega, h(\alpha, x))| > \varepsilon\Big) \\ &\qquad + \mathbf{P}_{\nu\times\mu}\Big(|e^*_{x_0x_N}(f, \omega) - f(\omega, h(\beta, x))| > \varepsilon\Big) \\ &\quad < 2\varepsilon. \end{aligned} \tag{3}$$

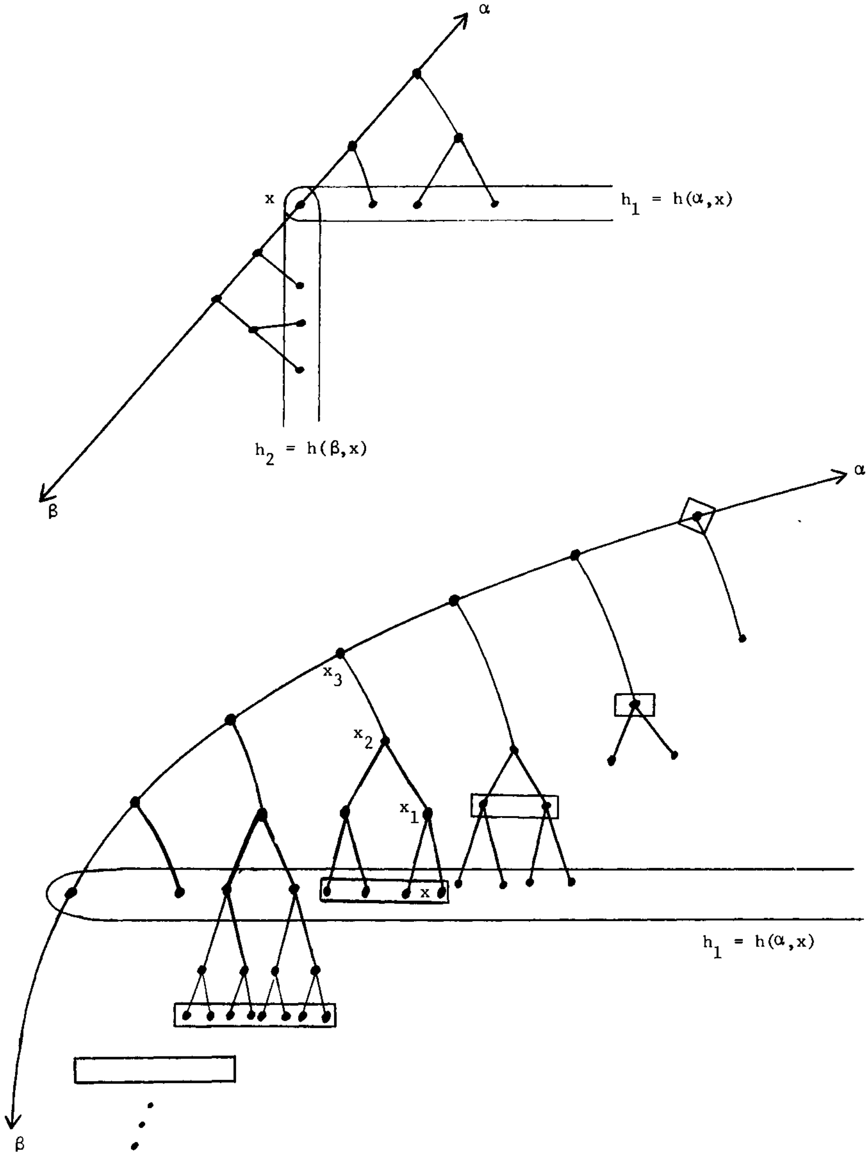

FIG. 4. *(a) The case* $c(h_1, h_2) = 0$. *(b) The case* $c(h_1, h_2) = 3$; *here* $\mathscr{L}_1 \cap \mathscr{L}_2 \setminus \{x\} = \{x_1, x_2, x_3\}$. $h_2 = h(\beta, x)$ *represented by vertices in rectangular boxes.*

On the other hand, G-invariance of the finite-dimensional marginals of f, together with Lemma 4, implies there is a $p(N, \varepsilon)$ such that for any h_1, h_2 with $c(h_1, h_2) = N$,

$$\mathbf{P}_\mu(|f(\omega, h_1) - f(\omega, h_2)| > 2\varepsilon) = p(N, \varepsilon). \tag{4}$$

The rays α and β are chosen always to satisfy $c(h(\alpha, x), h(\beta, x)) = N$. Thus,

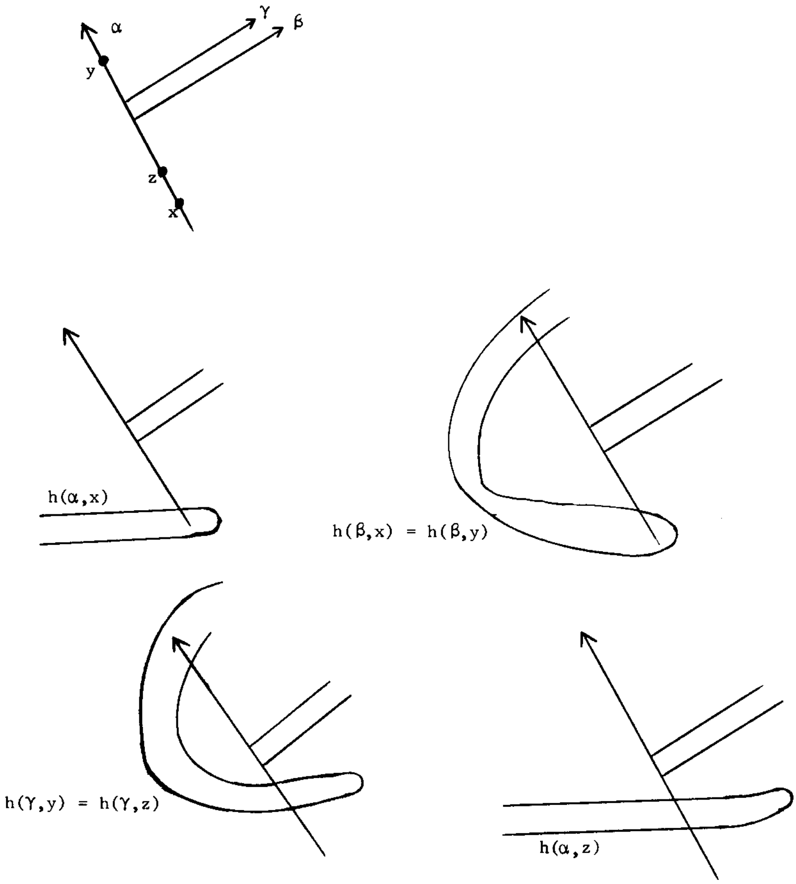

FIG. 5.

conditioning on α and β in (3) gives $p(N, \varepsilon) < 2\varepsilon$ for $N \geq N(\varepsilon)$. Informally, horocycles that share enough vertices are arbitrarily likely to have arbitrary close values of f.

Now let $h(\alpha, x)$, $h(\alpha, \omega)$, $h(\alpha, z)$ be successive horocycles in direction α. Apply Lemma 5 to the horocycles $h_1 = h(\alpha, x)$ and $h_4 = h(\alpha, z)$ to get h_2 and h_3 with $c(h_i, h_{i+1}) \geq N(\varepsilon)$ for $i = 1, 2, 3$. Then

$$\begin{aligned}\mathbf{P}_\mu(|f(\omega, h_1) - f(\omega, h_4)| > 6\varepsilon) &\leq \sum_{i=1}^{3} \mathbf{P}_\mu(|f(\omega, h_i) - f(\omega, h_{i+1})| > 2\varepsilon)\\ &= \sum_{i=1}^{3} p(c(h_i, h_{i+1}), \varepsilon)\\ &< 6\varepsilon.\end{aligned}$$

This is true for any $\varepsilon > 0$, hence $\mathbf{P}_\mu(f(\omega, h_1) = f(\omega, h_4)) = 1$. Since h_1 and h_4

can be any consecutive even horocycles it follows that for each α, $\mathbf{P}_\mu(f$ is constant on even horocycles in direction $\alpha) = 1$. Then by Fubini's theorem,

$$\mathbf{P}_\mu\big(\{\alpha\colon f(h_1) \neq f(h_2) \text{ for even horocycles } h_i \text{ in direction } \alpha\} \text{ is a null set}\big) = 1. \tag{5}$$

Let $f_e\colon \Omega \times \partial\mathbf{T} \to \mathbf{R}$ be defined by $f_e(\omega, \alpha) = 0$ if there are vertices $x, y \in$ EVENS with $f(\omega, h(\alpha, x)) \neq f(\omega, h(\alpha, y))$ and otherwise be defined by $f_e(\omega, \alpha) = f(\omega, h(\alpha, x))$ for any $x \in$ EVENS. Define f_o analogously for odd horocycles. It is trivial to check that f_e and f_o are measurable. To check that they are G-invariant, pick any $z \in$ EVENS and write

$$\begin{aligned} e_{xy}(f_e, \omega) &= \int_{I(x,y)} f_e(\omega, \alpha)\, dm_x(\alpha)/m_x(I(x,y)) \\ &= \int_{I(x,y)} f(\omega, h(\alpha, z))\, dm_x(\alpha)/m_x(I(x,y)) \\ &= e^*_{xyz}(f, \omega), \end{aligned}$$

where z is any even vertex and the last two equalities are μ-almost sure; the fact that f may be substituted for f_e under the integral sign follows from (5). Now $e_{\pi x \pi y}(f_e, \omega) = e^*_{\pi x \pi y \pi z}(f, \omega)$ μ-almost surely, so G-invariance of f_e follows from G-invariance of f. Theorem 2 now implies that f_e is almost surely constant modulo null sets. Of course, the same is true of f_o. This proves Theorem 3. □

4. Proof of Theorem 1. To show that $\mathscr{F}(\alpha)$ is trivial, it suffices to show that $\mathbf{P}_\mu(A | \mathscr{F}(\alpha))(\cdot)$ is constant for cylinder events A. Since versions of $\mathbf{P}_\mu(A | \mathscr{F}(\alpha))(\cdot)$ differ on null sets, it suffices to find for each A and α a version that is μ-almost surely constant. Here is an outline of the reasoning.

Any vertex y in the horocycle $h(\alpha, x)$ is indistinguishable from x when viewed from α (i.e., there is a $\pi \in G$ fixing α and mapping x to y) so $\mathscr{F}(\alpha)$ carries the same information about y as about x. The conditional distribution of $\xi(x)$ given $\mathscr{F}(\alpha)$ depends therefore only on the $h(\alpha, x)$ and not on x itself and thus defines a function from horocycles in the direction α to probability distributions on $[0, 1]$. As α varies, this defines a function from $\mathscr{H}$ to probability distributions on $[0, 1]$. Since the function is defined in a G-invariant way, it must be almost surely constant modulo null sets, thus the conditional distribution of $\xi(x)$ given $\mathscr{F}(\alpha)$ is constant. Applying a similar argument to more general finite-dimensional marginals shows that $\mathscr{F}(\alpha)$ is trivial.

Begin for real by constructing the conditional probabilities. A *configuration* on a finite set S is just a Borel subset of $[0, 1]^S$, that is, a specification of a range of values for ξ to take at each vertex in S. For each finite $S \subseteq \mathbf{T}$, each $t \in \mathbf{Z}^+$ and each $x \in \mathbf{T} \setminus \text{hull}(S)$, let $\mathscr{N}(S, x, t)$ denote the set of $z \in \mathbf{T}$ for which $x \in \text{hull}(S \cup \{z\})$ and $1 \leq d(x, z) \leq t$.

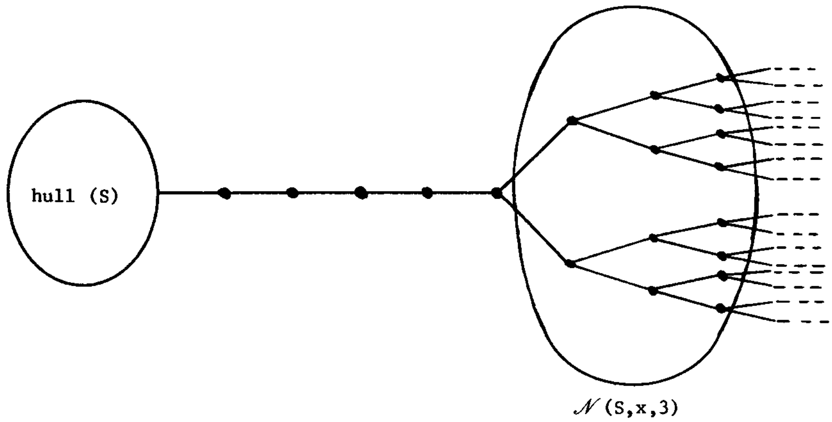

FIG. 6.

The first goal is to define conditional probabilities

$$\mathbf{P}_\mu(\xi \text{ is in configuration } \eta | \xi(z) \text{ for } z \in \mathcal{N}(S, x, t)) \tag{6}$$

in a G-invariant way for each fixed configuration η on each finite set S. Fix an $x_0 \in \mathbf{T}$. Each configuration η on a finite set S has finitely many images under automorphisms of $\mathbf{T}$ fixing x_0 (see [17]). Choose one from each orbit and for each chosen S, η, each $t \in \mathbf{Z}^+$ and each θ: $\mathcal{N}(S, x, t) \to [0, 1]$, define $W_0(S, \eta, x, t)(\theta)$ to be an arbitrary version of (6) evaluated at a ξ that agrees with θ on $\mathcal{N}(S, x_0, t)$. Let $W(S, \eta, x_0, t)(\theta)$ be the average of $W_0(S, \eta, x_0, t)(\theta')$ over the finitely many ϕ' that are $\theta \circ \pi^{-1}$ for some $\pi \in G$ that fixes S, x_0 and $\mathcal{N}(S, x_0, t)$. Now extend W to arbitrary S, η and x by

$$W(S, \eta, x, t)(\theta) = W(\pi S, \pi\eta, \pi x, t)(\theta \circ \pi^{-1}), \tag{7}$$

where $\pi \in G$ maps x to x_0 and η to the only configuration on which W_0 is defined. The right-hand side of (7) is well defined, due to the averaging procedure, even though the term $\theta \circ \pi^{-1}$ may not be. Clearly, W is a version of (6) and is G-invariant in the sense that (7) holds for all $\pi \in G$.

For each finite $S \subseteq \mathbf{T}$ and each $\alpha \in \partial\mathbf{T}$, let $x_r(S, \alpha)$ be the unique $x \in \mathbf{T}$ such that $d(x, \text{hull}(S)) = r$ and any ray in α from a point in S contains x. Let

$$V(S, \eta, \alpha)(\xi) = \lim_{r \to \infty} \lim_{t \to \infty} W(S, \eta, x_r(S, \alpha), t)(\xi|_{\mathcal{N}(S, x_r(S, \alpha), t)}), \tag{8}$$

where any limit that does not exist is defined to be zero. Recall that $\mathcal{F}(\alpha) = \bigcap_r \mathcal{F}(\alpha^r)$, where $\mathcal{F}(\alpha^r)$ is the σ-field generated by the values of ξ on the $\mathcal{N}(S, x_r(S, \alpha), \infty)$. Then the inner limit on the right-hand side of (8) exists μ-almost surely and is a version of $\mathbf{P}_\mu(\xi$ is in configuration $\eta | \mathcal{F}(\alpha^r))$ and the outer limit exists μ-almost surely and is a version of $\mathbf{P}_\mu(\xi$ is in configuration $\eta | \mathcal{F}(\alpha))$.

The conditional probabilities are now suitably defined. It is time to see why they are μ-almost surely constant. Fix S and η. Fix a horocycle $h(x_0, \alpha_0)$ disjoint from hull(S). Let Q: $\Omega \times \mathscr{H} \to [0,1]$ be defined by

$$Q(\xi, h) = V(\pi S, \pi\eta, \pi\alpha_0)(\xi), \tag{9}$$

where π is any map in G mapping h_0 to h.

LEMMA 6. *Q is well defined, jointly measurable, G-invariant and has G-invariant finite-dimensional marginals.*

PROOF. The argument $\pi\alpha_0$ is well defined because it is the direction of h. The pair $(\pi S, \pi\eta)$ is defined up to mapping under elements of G that fix h, hence fix α. Such maps fix all but finitely many vertices in any ray in α, so the pair $(\pi S, \pi\eta)$ ranges over the union over r of the finite orbits of some particular pair $(\pi_0 S, \pi_0\eta)$ under maps fixing all points on rays from S_0 in α at distance r or greater from hull(S_0). By construction, W is constant on each of these orbits, hence on their union, since all the orbits intersect at $(\pi_0 S, \pi_0\eta)$. It follows from (8) that V is also constant on the union of the orbits and therefore that Q is well defined.

To check measurability for fixed S and η, note that for fixed r and t, the right-hand side of (8) depends on α only through the points $x_i(\alpha)$: $i = 1, \dots, r$, so it is a simple function partitioning $\partial\mathbf{T}$ into clopen sets. For each of these sets, the right-hand side of (8) depends on ξ only through the values of ξ on the finite set $\mathscr{N}(S, x_r(S, \alpha), t)$, so it is a simple function in ξ as well, thus jointly measurable. Joint measurability is closed under taking limits and assigning the value zero when limits do not exist.

To check G-invariance, note that

$$Q(\xi \circ \pi^{-1}, \pi h) = Q(\xi, h)$$

as an immediate consequence of definitions (7), (8) and (9). Since μ is G-invariant, this is a strong form of G-invariance implying both the ones in the lemma. □

Returning to the proof of Theorem 1, Theorem 3 and Lemma 6 together imply that Q is μ-almost surely constant modulo null sets on even horocycles and on odd horocycles. The constant modulo null set value of $Q(\xi, \cdot\,)$ on even horocycles is a.s. the same as for $Q(\pi\xi, \cdot\,)$, so it is a G-invariant random variable, hence almost surely constant by the assumption of extremality. The same is true of the odd horocycles, so Q is $\mu \times \mathscr{B}$-almost surely constant on parity classes, meaning that the exceptional set is a $\mu \times m_x$ null set for any x. Then $Q(\cdot, \alpha)$ is constant modulo μ-null sets for almost every α, hence for every α by the obvious G-invariance, and the triviality of $\mathscr{F}(\alpha)$ has been shown.

5. Corollaries, examples and limitations. This section describes a few easy consequences of Theorem 1. The results are basically that mixing and a weak law of large numbers follow from the triviality of the tail fields $\mathscr{F}(\alpha)$. The proofs are along the usual lines.

For events A and B, write $d(A, B) \geq j$ if there are sets V and W of vertices such that A depends only on the value of ξ at vertices in V, B depends only on the value of ξ at vertices in W and $\inf\{d(x, y): x \in \text{hull}(V), y \in \text{hull}(W)\} \geq j$.

COROLLARY 7 (Uniform independence for distant events). *Let μ be an extremal G-invariant probability measure on Ω. Then for any fixed event A,*

$$\sup_{B:\, d(A, B) \geq k} |\mu(A \cap B) - \mu(A)\mu(B)| \to 0 \quad \text{as } k \to \infty.$$

REMARK. It is not possible to get uniformity in A as well as B. For a counterexample let $\xi^{(r)}$ be independent $\{0, 1\}$-valued automorphism-invariant processes with $\mathbf{P}(\xi^{(r)}(v) = \xi^{(r)}(w)) = 0.9$ whenever $d(v, w) = r$ and let ξ be the $[0, 1]$-valued process whose rth binary digit is $\xi^{(r)}$.

PROOF. Suppose not and choose a sequence of events B_k for which

$$d(A, B_k) \geq k \quad \text{and} \quad |\mu(A \cap B_k) - \mu(A)\mu(B_k)| \geq \varepsilon \tag{10}$$

for some $\varepsilon > 0$ and all k. Let V_k and W_k be the sets of vertices witnessing $d(A, B_k) \geq k$ as in the definition of $d(A, B) \geq j$ from the second paragraph of this section. The first step is to show that all the sets V_k may be taken to be equal. Assume without loss of generality that no B_k is trivial so no W_k is empty. Then each set V_k has $\text{hull}(V_k) \neq \mathbf{T}$, so for each V_k, there is some oriented edge e_k such that V_k is included in the component of $\mathbf{T} \setminus e_k$ containing the tail of e_k and W_k is included in the component containing the head of e_k. There are two possibilities. First, suppose there are only finitely many distinct edges e_k, call them $\{e_k: k \in S\}$. Recall A is measurable with respect to each $\sigma(\xi(v): v \in V_k)$, so that A is measurable with respect to $\bigcap_{k \in S} \sigma(\xi(v): v \in V_k) = \sigma(\xi(v): v \in \bigcap_{k \in S} V_k)$. Since $\bigcap_{k \in S} V_k$ is disjoint from each W_k, each V_k may be taken to be this set. On the other hand, if there are infinitely many distinct e_k, then there is some ray $x_0, x_1, \ldots$ such that infinitely many oriented edges $\overline{x_i, x_{i+1}}$ appear among the e_k. If α is the boundary point for the ray $x_0, x_1, \ldots,$ then A is measurable with respect to $\mathscr{F}(\alpha)$. By Theorem 1, $\mathscr{F}(\alpha)$ is trivial, hence A is trivial and the corollary is trivially true in this case. Step 1 has now been accomplished and in fact something stronger has been shown: Not only can all V_k can be assumed equal, but there is an oriented edge $e = \overline{xy}$ such that V is in the component of $\mathbf{T} \setminus e$ containing x and each W_k is in the component of $\mathbf{T} \setminus e$ containing y.

Let $x_0, x_1, \ldots$ be any fixed ray. Write $\mathbf{T}^{(k)}$ for the component of $\mathbf{T} \setminus \{x_{k-1}\}$ containing x_k and $\mathbf{T}'$ for the component of $\mathbf{T} \setminus \{x_1\}$ containing x_0. Then since $\text{hull}(V)$ and $\text{hull}(W_k)$ are at distance at least k, there is some $\pi_k \in G$ mapping V into $\mathbf{T}'$ and mapping W_k into $\mathbf{T}(k)$. In fact by the stronger version of step 1, π_k may always be chosen to map the edge e to $\overline{x_0, x_1}$ and thus $\pi_k A$ will be the same event for all k. Replacing B_k by $\pi_k B_k$ and A_k by $\overline{A} = \pi_k A_k$ does not affect equation (10), so there is no loss of generality in assuming that W_k is already a subset of $\mathbf{T}^{(k)}$. The rest is easy.

Write $\mathscr{F}_k$ for $\sigma(\mathbf{T}^{(k)})$ and note that $\bigcap \mathscr{F}_k = \mathscr{F}(\alpha)$, where $\alpha \in \partial\mathbf{T}$ is the direction of the ray $x_0, x_1, \dots$. By Theorem 1, $\mathscr{F}(\alpha)$ is trivial. Then

$$|\mu(A \cap B_k) - \mu(A)\mu(B_k)| \le \int_{B_k} |\mathbf{E}_\mu(A|\mathscr{F}_k) - \mu(A)|$$
$$\le \|\mathbf{E}_\mu(A|\mathscr{F}_k) - \mu(A)\|_1.$$

Since $\bigcap \mathscr{F}_k$ is trivial, the martingale convergence theorem implies $\mathbf{E}_\mu(A|\mathscr{F}_k) - \mu(A) \to 0$ almost surely and in $L^1(\mu)$. Thus $|\mu(A \cap B_k) - \mu(A)\mu(B_k)| \to 0$ and the corollary is proved. □

Corollary 8 (Mixing). *Let μ be an extremal G-invariant probability measure on Ω. For any $\pi \in G$, let $|\pi|$ denote $\min_{x \in \mathbf{T}} d(x, \pi x)$. Then for any events A and B, $\mu(A \cap \pi B) \to \mu(A)\mu(B)$ as $|\pi| \to \infty$.*

Proof. Choose any $\varepsilon > 0$ and let A' and B' be cylinder sets with $\mu(A \Delta A') < \varepsilon$ and $\mu(B \Delta B') < \varepsilon$. Now hull$(A' \cup B')$ is finite, having some diameter D. Then $d(A', \pi B') \ge |\pi| - D \to \infty$ as $|\pi| \to \infty$. By the previous corollary, $|\mu(A' \cap \pi B') - \mu(A')\mu(B')|$ goes to zero. But

$$|\mu(A \cap \pi B) - \mu(A)\mu(B)| \le |\mu(A' \cap \pi B') - \mu(A')\mu(B')| + 4\varepsilon,$$

so $\limsup_{|\pi| \to \infty} |\mu(A \cap \pi B) - \mu(A)\mu(B)| \le 4\varepsilon$. Since ε was arbitrary, the lim sup is zero and the corollary is proved. □

Corollary 9 (Birkhoff averaging). *Again let μ be an extremal G-invariant probability measure on Ω. Let $\pi_1, \pi_2, \dots$ be any sequence of maps in G such that for any $x \in \mathbf{T}$, $d(\pi_i x, \pi_j x) \ge |i - j|$ [e.g., $\pi_k = (\pi)^k$ for some π without fixed points]. Let $f\colon \Omega \to \mathbf{R}$ be any bounded measurable function. Let*

$$f^{(k)} = \frac{1}{k} \sum_{j=1}^{k} f \circ \pi_j.$$

Then $f^{(k)} \to \mathbf{E}_\mu f$ in probability.

Proof. First consider the case where $f = \mathbf{1}_A$ for some event A. Clearly, $\mathbf{E}_\mu f^{(k)} = \mathbf{E}_\mu f$, so it suffices to show that the variance of $f^{(k)}$ goes to zero. Write the variance of $f^{(k)}$ as

$$\frac{1}{k^2} \sum_{1 \le i, j \le k} \left[\mu(\pi_i A \cap \pi_j A) - \mu(A)^2\right].$$

Use the previous corollary with $A = B$ and $\pi = \pi_i^{-1}\pi_j$ to conclude that for any $\varepsilon > 0$, there is an N such that $|i - j| > N \Rightarrow |\pi_i^{-1}\pi_j| > N \Rightarrow |\mu(\pi_i A \cap \pi_j A) - \mu(A)^2| < \varepsilon$. Then for $k > N/\varepsilon$, the fraction of summands for which $|i - j| \le N$ is at most 2ε so the variance is at most $2\varepsilon + (1 - 2\varepsilon)\varepsilon$. Thus the variance goes to zero as k goes to infinity and the special case is proved.

By linearity, it is immediate that the corollary is also true for simple functions (finite linear combinations of indicator functions). Finally, let f be any bounded measurable function and $\varepsilon > 0$ be arbitrary. Let g be a simple function with $\|f - g\|_\infty < \varepsilon$. Then for large enough k, $\mathbf{P}_\mu(|g^{(k)} - \mathbf{E}_\mu g| > \varepsilon) < \varepsilon$. But then $\mathbf{P}_\mu(|f^{(k)} - \mathbf{E}_\mu f| > 3\varepsilon) < \varepsilon$, since $f^{(k)}$ and $\mathbf{E}_\mu f$ are within ε of $g^{(k)}$ and $\mathbf{E}_\mu g$, respectively. Thus $f^{(k)} \to \mathbf{E}_\mu f$ in probability as well. □

The weak law of large numbers is a special case of this, although it must be formulated differently to take parity into account.

COROLLARY 10 (Weak law). *Let μ be an ergodic G-invariant measure of $\Omega = [0,1]^{\mathbf{T}}$ and A be any Borel subset of $[0,1]$. For a finite set of vertices S, let $e(S) = 1 - o(S) = |S \cap EVENS| \setminus |S|$ and let $A(S)$ be the fraction of vertices in $v \in S$ for which $\xi(v) \in A$. Let x and y be any vertices in EVENS and ODDS, respectively. For any ε, there is an $N(\varepsilon)$ such that for any finite set $S \subseteq \mathbf{T}$ of cardinality at least $N(\varepsilon)$, the probability is at least $1 - \varepsilon$ that $|A(S) - [e(S)\mathbf{P}_\mu(\xi(x) \in A) + o(S)\mathbf{P}_\mu(\xi(y) \in A)]| < \varepsilon$.*

PROOF. By G-invariance, the expected value of $A(S)$ is just $e(S)\mathbf{P}_\mu(\xi(x) \in A) + o(S)\mathbf{P}_\mu(\xi(y) \in A)$. So again it suffices to show that the variance of $A(S)$ goes to zero. The correlation of the events $\xi(x) \in A$ and $\xi(y) \in A$ goes to zero as the distance between x and y goes to infinity, according to Corollary 8. Requiring S to be large enough forces each vertex in S to be at a large distance from all but an arbitrary small fraction of the vertices in S. Then the variance of $A(S)$ is $o(1)$ as the cardinality of S gets large, and the corollary is proved. □

Berger and Ye [2] have used similar results to this to define entropy for G-invariant probability measures on $[0,1]^{\mathbf{T}}$ as follows. Let h_k be the entropy of the law of μ restricted to a ball of radius k and divided by the number of vertices in the ball. Then they show that h_k decreases to a limit h which they call the entropy of the process. In a subsequent work [18], they use a version of Corollary 10 to show a version of the Shannon–McMillan–Breiman theorem for trees, namely that if μ is an extremal G-invariant measure and $h_k(\xi)$ is the log of the probability of seeing ξ on the ball of radius k divided by the number of vertices in the ball, then $h_{2k} \to h$ in probability.

Recall that there is an analogue of the two-sided tail field that is defined by $\mathscr{F}^* = \bigcap_S \sigma(\xi(x)\colon x \notin S)$, where S ranges over finite subsets of $\mathbf{T}$. The following example shows that $\mathscr{F}^*$ is not necessarily trivial even when ξ is a Markov random field. Contrast this to the case of $\mathbf{Z}$-indexed stationary processes, where irreducible aperiodic Markov chains always have trivial two-sided tails. The contrast is in a sense due to the fact that $\mathscr{F}^*$ contains the information of infinitely many tails.

EXAMPLE. Fix any $p < 1$ for which $(2p-1)^2 n > 1$. Let ξ be the Markov random field such that $\xi(x)$ is always -1 or 1 and the probability of $\xi(x)$

being 1 given that $\xi(y) = 1$ for precisely k of the $n + 1$ neighbors of x is $p^k(1 - p)^{n+1-k}/(p^k(1 - p)^{n+1-k} + (1 - p)^k p^{n+1-k})$. This is one of Spitzer's "Markov chains" [16] and can be constructed by picking some x, letting $\xi(x)$ be -1 or 1 with probability 1/2 each and defining the values of ξ outwards from x by letting $\xi(y)$ agree with $\xi(z)$ with probability p, where z is the neighbor of y that is closer to x.

Let $x_0 \in \mathbf{T}$ be a fixed vertex and let $V_k(\xi)$ be the sum of ξ over vertices at distance k from x_0. T. Kamae [9] has shown that $\mathbf{E}(V_0(\xi)V_k(\xi))/[\mathbf{E}V_k(\xi)^2]^{1/2}$ remains bounded away from zero as $k \to \infty$, which, by taking limits in L^2, implies the existence of $\mathscr{F}^*$ measurable random variables positively correlated with $\xi(x_0)$ and in particular, nondegenerate.

Acknowledgments. The example in Section 5 due to T. Kamae [9] was brought to my attention by David Aldous and its origin was pointed out by the referee; my work was enhanced by several conversations with Russ Lyons and his comments on early drafts of the work.

REFERENCES

[1] Aldous, D. (1983). Exchangeability. *École d'Été de Probabilités de Saint Flour*. *Lecture Notes in Math.* **1117**. Springer, Berlin.
[2] Berger, T. and Ye, Z. (1990). Entropic aspects of random fields on trees. *IEEE Trans. Inform. Theory* **IT-36** 1006–1018.
[3] Cartier, P. (1972). Fonctions harmonique sur an arbre. *Symp. Math.* **9** 203–270.
[4] Chen, S.-S. (1981). Limit sets of automorphism groups of a tree. *Proc. Amer. Math. Soc.* **83** 437–441.
[5] Diaconis, P. (1988). Recent progress in de Finetti's notions of exchangeability. In *Bayesian Statistics* **3** (Valencia, 1987) 111–125. Oxford Univ. Press, N.Y.
[6] Durrett, R. (1988). *Lecture Notes on Particle Systems*. Wadsworth, Belmont, Calif.
[7] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**, 2nd ed. Wiley, New York.
[8] Grimmett, G. and Newman, C. (1989). Percolation in $\infty + 1$ dimensions. In *Disorder and Physical Systems* (G. Grimmett and D. Welsh, eds.). Clarendon Press, Oxford.
[9] Higuchi, Y. (1977). Remarks on the limiting Gibbs states on a $(d + 1)$-tree. *Publ. Res. Inst. Math. Sci.* **13** 335–348.
[10] Kindermann, R. and Snell, J. L. (1980). *Markov Random Fields and Their Applications*. Amer. Math. Soc., Providence, R.I.
[11] Liggett, T. (1985). *Interacting Particle Systems*. Springer, New York.
[12] Lyons, R. (1989). The Ising model and percolation on trees and tree-like graphs. *Comm. Math. Phys.* **125** 337–353.
[13] Pemantle, R. (1992). The contact process on trees. *Ann. Probab.* **20**.
[14] Sawyer, S. (1978). Isotropic random walks in a tree. *Z. Wahrsch. Verw. Gebiete* **42** 279–292.
[15] Serre, J.-P. (1968, 1969). Arbres, amalgames et SL_2. Collège de France. Notes.
[16] Spitzer, F. (1975). Markov random fields on an infinite tree. *Ann. Probab.* **3** 387–398.
[17] Tits, J. (1970). Sur le groupe des automorphismes d'un arbre. In *Essays on Topology and Related Topics*. Springer, New York.
[18] Ye, Z. and Berger, T. (1990). Asymptotic equipartition property for random fields on trees. Unpublished manuscript.

Department of Mathematics
Oregon State University
Corvallis, Oregon 97331-4605