Pacific Institute for the Mathematical Sciences

PIMS Distinguished Chair Lectures

DANIEL J. RUDOLPH
University of Maryland

PIMS Distinguished Chair
Collaborative Research Group
on Dynamics and Related Topics
University of Victoria
October, 2004

Entropy and Orbit Equivalence

ENTROPY AND ORBIT EQUIVALENCE

DANIEL J. RUDOLPH

Abstract. In these notes we first offer an overview of two core areas in the dynamics of probability measure preserving systems, the Kolmogorov-Sinai theory of entropy and the theory of orbit equivalence. Entropy is a nontrivial invariant that, said simply, measures the exponential growth rate of the number of orbits in a dynamical system, a very rough measure of the complexity of the orbit structure. On the other hand, the core theorem of the orbit theory of these systems, due to Henry Dye, says that any two free and ergodic systems are orbit equivalent, that is to say can be regarded as sitting on the same set of orbits. The goal we set out to reach now is to explain and understand the seeming conflict between these two notions.

1. Introduction

We begin with a brief sketch of what we will be doing. The study of dynamical systems comes in many flavors. The one we consider here is that of probability measure preserving dynamics. Thus the underlying state space we consider will simply be a standard probability space $(X, \mathcal{F}, \mu)$ and the dynamics will take the form of a μ preserving measurable bijection T of X. Such systems arise quite naturally. Any time one has a compact metric space as state space and a homeomorphism giving the dynamics there will be invariant Borel probability measures on the state space. It is important to realize that in changing perspective from topological to a measure theoretic one drops to a far weaker and more easily manipulated category. This is both a loss and a gain as results will only hold in this category, but the results will be quite substantial. Some good places to learn more about dynamics and ergodic theory are Brin and Stuck [1], Hasselblatt and Katok [3], Petersen [9] , Walters [12], and Rudolph [10].

Our focus will be on two particular aspects of this theory, the Kolmorogorov-Sinai entropy theory and the theory of orbit equivalence. We will go into more detail later, but suffice it here to say that the entropy of a measure preserving transformation T is a numerical value (perhaps ∞) giving the exponential growth rate in n of the number of measurably distinct orbits of length n of T. For example, the entropy of any irrational rotation of the circle is 0. This is true in fact of any topologically transitive isometry of a compact metric space relative to their unique invariant measure. As another example, consider the hyperbolic toral automorphism given by

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

Date: August 24, 2005.

acting on $\mathbb{R}^2/\mathbb{Z}^2$. This preserves Lebesgue measure. Its entropy is $\ln(\frac{-1+\sqrt{5}}{2})$, the log of the norm of its single eigenvalue outside the unit circle. More generally the entropy of any hyperbolic toral automorphism is the sum of the norms of the eigenvalues outside the unit circle. As a third example, consider the space of all Brownian paths passing through the origin. Here we consider $-\infty < t < \infty$ and are placing Wiener measure on the space of all continuous functions $\mathbb{R} \to \mathbb{R}$ with $f(0) = 0$. This is a standard probability space. One can place on it a measure preserving map by defining $T(f)(t) = f(t+1) - f(1)$, the left shift on the Brownian path relocated to pass through $(0, 0)$. This can be shown to be of infinite entropy. In the general study of entropy these are quite basic examples and not difficult to verify. As we continue we will learn and use much more about entropy. The point we make here is that it is a nontrivial invariant and can be used to distinguish distinct systems as truly different even in the measurable category.

A fundamental fact about standard probability spaces is that there is just about only one of them. Stated more precisely, suppose $(X, \mathcal{F}, \mu)$ and $(Y, \mathcal{G}, \nu)$ are two nonatomic standard probability spaces. One can prove then that there is a measure-preserving bijection $\phi : X \to Y$. If the spaces are Borel, that is to say are formally presented as compact metric spaces with Borel measures on them, then ϕ can be chosen to be Borel. Now standard probability spaces can have atoms and the masses of those atoms can be an obstacle to the existence of such a ϕ but those masses are the only obstacles. Now suppose one has dynamics on these two spaces, given by measure preserving maps T and S and moreover require they act ergodical (more on this later). An "orbit" of T is simply a list of points $\{T^j(x)\}_{j=-\infty}^{\infty}$. A natural way to describe this is to place on X an equivalence relation where two points x_1 and x_2 are considered equivalent iff for some $j \in \mathbb{Z}$, $T^j(x_1) = x_2$. In this vocabulary an orbit is then an equivalence class of the orbit relation. An "orbit equivalence" from T to S is a measure preserving bijection $\phi : X \to Y$ which almost surely takes an orbit of T to an orbit of S. It is important here to realize we are not asking that ϕ preserve the time order of points on an orbit. If ϕ did preserve time order on orbits then we would call it a "conjugacy". A very significant step in our understanding of measure-preserving dynamics is Dye's theorem from 1959 [2] saying that any two ergodic and free systems, like our T and S, are orbit equivalent. That is to say, they can be regarded as sitting on the same space of orbits just ordered differently by time. This in its own right is quite startling. Not only is their only one nonatomic standard probability space but there is also only one ergodic orbit relation on it. All of measure preserving dynamics then can be thought of as the study of how one might measurably order the points in the equivalence classes.

These two notions, a nontrivial invariant that measures the complexity of the orbit structure and and theorem that states there is only one orbit structure can be viewed either as a seeming contradiction or, more accurately, as an opportunity to explore the orbit structure at a more subtle level. That is our goal.

We will present an overview of the theory of restricted orbit equivalence [4] as a rather general approach to exploring the orbit structure of these systems more subtly than Dye. The idea here follows that of Dye. The "full group" of a measure preserving transformation T consists of all bijections ϕ with $\phi(x)$ on the T orbit of x. One can use elements of the full group to "perturb" or "rearrange" the orbit of T by conjugation to $\phi^{-1}T\phi$. Such perturbations do not change the conjugacy class of the map but they do rearrange the orbit. One could now consider sequences ϕ_i from the full group for which $\phi_i^{-1}T\phi_i$ converges pointwise to some other map T'. In this case notice that T' would have the same orbits as T. This is precisely what Dye does, and we will go into some detail on this. He shows that he can construct such a sequence ϕ_i so that one converges pointwise to an action conjugate to the dyadic adding machine. One can view this argument as the construction of a certain completion for the full group relative to an appropriate choice of pseudometric. The general restricted orbit equivalence machinery simply axiomatizes what is needed of such a pseudometric.

What we next will see is that one can construct a pseudometric very similar to the computation of entropy. We can regard it then as a measure of the complexity of the perturbation of the orbits of T when they are conjugated by some ϕ from the full group. Such pseudometrics always generate an equivalence relation of two actions being reachable, one from the other, by Cauchy sequences of perturbations. The final goal of our work is to show that for the pseudometric we describe this relation is precisely equality of entropy. This result can be found in [11] within the context of restricted orbit equivalence and for discrete amenable group actions [4].

By restricting our work to actions of $\mathbb{Z}$ we hope the ideas will be more accessible. We aim the core of these notes at an advanced graduate student and broad mathematical audience and will try to provide enough details to satisfy an educated but non-expert mathematician. We do also intend to offer an essentially complete proof of our core result without reference to the restricted orbit equivalence machinery. Hence in the final sections we must use substantial background material and at this point our work is for the expert and those pursuaded by the earlier material to learn the necessary ideas in the theory of entropy and Ornstein theory of Bernoulli shifts.

1.1. **Measure Preserving Dynamics.** Formally we will be studying measure preserving and invertible transformations of a standard probability space. Said more precisely, our space will be written $(X, \mathcal{F}, \mu)$ where X is some state space, $\mathcal{F}$ is a σ-algebra of measurable sets and μ is a probability measure defined on $\mathcal{F}$. To be "standard" for us means that X is a compact metric space, μ is a Borel measure, i.e. is defined at least on the Borel σ-algebra, and $\mathcal{F}$ is complete. This is an extremely natural context for dynamics. In particular whenever one has a compact metric space X and an homeomorphism T there will be T invariant Borel probability measures on X that arise by taking averages along orbits.

Let's describe that in a bit more detail. The signed Borel measures on a compact metric space are precisely the dual of the continuous functions $C(X)$. That is to say, a bounded linear functional on the real valued continuous functions has the form $\phi(f) = \int_X f\, d\mu$ for a signed measure μ. Any such finite Borel signed measure μ is of the form $a\mu^+ - b\mu^-$ where a and b are nonnegative reals and μ^+ and μ^- are both Borel probability measures. As the unit ball in the weak* topology is compact, we get a compact topology on the Borel probability measures. Now for any point $x \in X$ and $n \in \mathbb{N}$ we get a linear functional on $C(X)$ by the calculation $A_n(f) = \frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x))$. By compactness there must be limit points for these averages and such limit points will always be invariant probability measures. Such limit measures embody at least a part of the statistical behavior of the orbit of x. It is certainly interesting and natural to consider the statistics of orbits and this argument tells us that at the least the asymptotic statistics can be investigated through the study of invariant measures.

A second and perhaps more important notion arises here now in that once we have settled on an invariant measure we will be interested only in properties "up to measure zero" with respect to that measure. For example:

Definition 1.1. *Suppose we have two standard probability spaces $(X, \mathcal{F}, \mu)$ and $(Y, \mathcal{G}, \nu)$. We say they are measurably isomorphic (just isomorphic for short) if there are subsets $X_0 \subseteq X$ and $Y_0 \subseteq Y$, each with $\mu(X_0) = \nu(Y_0) = 1$ and a measure preserving and invertible map $\phi : X_0 \to Y_0$.*

Proposition 1.2. *Up to isomorphism there is only one nonatomic standard probability space. More generally, the masses of the atoms of the atomic part of a standard probability measure are a complete invariant of isomorphism of such spaces.*

We won't prove this rather general result (see [10]). Our discussion of Dye's theorem later will provide substantial insight into how it can be proven. Notice that if one has a measure preserving map T acting on a probability space $(X, \mathcal{F}, \mu)$ then T must preserve the atomic part of the measure. Moreover any atom must be moved by T to another atom of the same mass and hence all atoms are parts of periodic orbits. The space X then breaks up naturally into T invariant pieces on which it is either atomic and periodic or is nonatomic. All our work will occur on nonatomic actions. Thus for our purposes all of our dynamics could be considered as occuring on the same probability space, the unit interval with Lebesgue measure for example. We now strengthen our notion of isomorphism to include the dynamics.

Definition 1.3. *We say measure preserving and invertible maps T on $(X, \mathcal{F}, \mu)$ and S on $(Y, ,\nu)$ are conjugate or isomorphic if a conjugacy ϕ between the two measure spaces exists for which $S\phi = \phi T$.*

It is definitely not the case that there is only one dynamical system up to conjugacy. We now introduce the ergodic theorems of Birkhoff and von Neumann. This result is a kind of converse to our description above of how to construct invariant measures.

Theorem 1.4. *Suppose T acting on $(X, \mathcal{F}, \mu)$ is measure preserving and suppose $f \in L^1(\mu)$. Then$\frac{1}{n}\sum_{j=0}^{n-1} f \circ T^j$ converges in L^1 and pointwise a.s. to the conditional expectation of f given the σ-algebra of T invariant sets.*

When the algebra of T-invariant sets is trivial, i.e. consists just of sets of measure 0 and 1, we say either T is ergodic or μ is an ergodic measure for T.

The conditional expectation given the algebra $\mathcal{I}(T)$ of T invariant sets has a more subtle structure than just being a projection. The probability space $(X, \mathcal{F}, \mu)$ actually decomposes as an integral of ergodic measures. That is to say, X is conjugate to a space we can write as a disjoint union $\cup_{\alpha \in A} X_\alpha$ where each $(X_\alpha, \mathcal{F}_\alpha, \mu_\alpha)$ is a standard probability space, each $X_\alpha \subseteq X$ is T invariant and μ_α is a T invariant and ergodic measure and $\mu = \int \mu_\alpha \, d\mu(\alpha)$. This disintigration of the measure space is called the Rokhlin decomposition of X over the σ-algebra $\mathcal{I}(T)$ (see [10]).

From now on we will assume μ is an ergodic measure for T. Even though we have this decomposition of an arbitrary measure into ergodic components it is not usually trivial to translate results about ergodic measures to general invariant measures but it is clear that we will be obtaining results that apply to each ergodic component separately.

1.2. **Entropy.** We now give a quick understanding of the Kolmogorov-Sinai entropy of a measure preserving transformation. Fix $(X, \mathcal{F}, \mu)$ and a measure preserving map T. For $P = \{s_1, s_2, \ldots, s_k\}$ a finite partition of X, let $P(x) = i$ if $x \in s_i$. Now let

$$P_n(x) = \{P(x), P(T(x)), \ldots, P(T^{n-1}(x))\}.$$

We call this list of symbols the "T,P,n-name" of x.

Apology: One regularly lets a "name" $\{i_1, i_1, \ldots, i_{n-1}\}$ represent both the name itself and the set of points x which have it as their T, P, n-name.

If one thinks vaguely of P as chopping X up into "small" pieces, then these names give an approximate picture of the orbit structure of T. One very rough measure of the complexity of this orbit structure is simply the number of names. As we are interested in statistical notions, we will ignore a small set of names and count the number that remain. Said more precisely, set

$$N(T, P, \varepsilon, n) = \text{min. number of } T, P, N\text{-names it takes to cover all but } \varepsilon \text{ in measure of } X.$$

We expect this number to grow exponentially in n and so attempt to extract the exponent:

$$h(T, P) = \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{\log_2(N(T, P, \varepsilon, n))}{n}.$$

Lastly, one takes as the entropy of T itself $h(t) = \sup_P h(T, P)$. From this rather sparce beginning one builds up a profound theory. We will mention some of this as we proceed. We do take a moment now to describe some examples:

Suppose T acts as a transitive isometry of a compact metric space. This situation can only arise actually if the space X is a compact abelian group and T is rotation of X by some element whose powers are dense in X. Such actions are uniquely ergodic in that they have only one invariant probability measure, Haar measure on the group. Irrational rotations of the circle, or of higher dimensional tori, are of this sort. All such isometric actions have zero entropy.

Suppose T is a hyperbolic automorphism of a finite dimensional torus, for example multiplication of $\mathbb{R}^2/\mathbb{Z}^2$ by the matrix

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

This action preserves Lebesgue measure on the torus. It is certainly not uniquely ergodic. It has an expanding eigen-direction with eigenvalue $\frac{-1+\sqrt{5}}{2}$ and a contracting eigen-direction with eigenvalue $\frac{-1-\sqrt{5}}{2}$. The entropy of T is in fact the log of the larger eigenvalue. In general, such hyperbolic actions may have a number of eigenvalues outside and inside the unit circle. The sum of the log's of the norms of those outside the unit circle is the entropy.

As a last example, take as measure space all continuous functions $\mathbb{R} \to \mathbb{R}$ passing through the origin with Wiener measure. This is the space of doubly-infinite Brownian paths passing through the origin. One can define the map T by $T(f)(t) = f(t+1) - f(1)$.

This is a rather natural shift map on these Brownian paths. It is offered here simply as an example of an infinite entropy measure preserving system.

Although much much more deserves to be said about entropy in the abstract and about its concrete calculation in examples like we have described we end with the simple observation that this is not a trivial invariant. It can be zero, it can be positive and finite, it can be infinite.

1.3. **Orbit Equivalence and Dye's Theorem.** We now consider another classical approach to the orbit structure of a measure preserving dynamical system. Suppose T_1 and T_2 act on the same space $(X, \mathcal{F}, \mu)$. To say they "have the same orbits" means

$$T_2(x) = T_1^{j(x)}(x) \text{ and } T_1(x) = T_2^{k(x)}(x) \text{ -a.s..}$$

Definition 1.5. *Suppose T and S are measure preserving invertible actions on $(X, \mathcal{F}, \mu)$ and $(Y, \mathcal{G}, \nu)$ respectively. We say they are "orbit equivalent" if there is a measure preserving bijection $\phi : X \to Y$ (almost surely) so that $\phi T \phi^{-1}$ and S have the same orbits (again, almost surely). We refer to such a map ϕ as an "orbit equivalence" between T and S.*

Quite amazingly H. Dye proved in 1959 [2], predating the theory of entropy of course, that any two ergodic actions on nonatomic spaces are orbit equivalent. Thus not only is there only one nonatomic standard probability space but there is really only one space of orbits. Essentially all measure preserving dynamics just comes down to choosing different ways of ordering the points of the orbits. In particular irrational rotations of the circle, hyperbolic toral automorphisms and the shift map on Brownian paths, all described earlier, can be regarded as simply different ways of walking on the same space of orbits.

Between these two notions, that the growth rate of the number of distinct orbits is an interesting and nontrivial invariant of a measure preserving system, and that there is essentially only one space of orbits there appears to be at least a philosophical conflict. More to the point perhaps there appears a need to explore further to understand how these two ideas might be placed on common ground.

To do this we first sketch a bit of how Dye's theorem is proven.

Definition 1.6. *By the full group $\Gamma = \Gamma(T)$ of a transformation T we mean the space of all measure preserving bijections of the form $\phi(x) = T^{j(x)}(x)$, that is to say, which map a point to another point of its T orbit.*

One can regard an element in Γ as perturbing the orbit structure of T by modifying $T \longrightarrow T' = \phi^{-1} T \phi$. Now T' has the same orbits as T. It is also trivially seen to be conjugate to T but it is not identical to T if ϕ is nontrivial. One can put a natural metric topology on these perturbations by setting

$$d_0(T, T') = \mu(\{x | T(x) \neq T'(x)\}).$$

Now if $T_i = \phi_i^{-1} T \phi_i$ is a d_0 Cauchy sequence of such perturbations it is not difficult to see that the T_i will converge in probability to a map T' with the same orbits as T but no longer necessarilly isomorphic to T. In fact this is precisely what Dye does to prove his result. He shows that one can construct maps ϕ_i in $\Gamma(T)$ so that this limit is conjugate to a dyadic adding machine (perhaps the simplest possible ergodic action on a nonatomic space). In later sections we will give more detail on how Dye did this.

1.4. **The "Complexity" of an Orbit Equivalence.** Having seen a sketch of how Dye showed there was only one orbit structure we can now consider how to integrate the counting ideas of entropy with the Dye theory. For $\phi \in \Gamma(T)$, ϕ acts as a permutation of each orbit of T. We can write each such as a permutation of $\mathbb{Z}$ by writing $h_x^{T,\phi}(i)$ to be that element of $\mathbb{Z}$ with $\phi(T^i(x)) = T^{h_x^{T,\phi}(i)}(x)$. This map $h_x^{T,\phi}$ is a bijection of $\mathbb{Z}$ that describes how the orbit of x is rearranged by ϕ. We begin with an important fact that follows from the ergodic theorem:

Lemma 1.7. *Suppose T acting on $(X, \mathcal{F}, \mu)$ is a measure preserving invertible map and $\phi \in \Gamma(T)$ is an element of the full group. We conclude*

$$\lim_{n\to\infty} \frac{\#\{x, T(x), \ldots, T^{n-1}(x)\} \triangle \phi(\{x, T(x), \ldots, T^{n-1}(x)\})}{n} = 0 \text{ a.s.}$$

which is the same as saying

$$\lim_{n\to\infty} \frac{\#\{0, 1, \ldots, n-1\} \triangle h_x^{T,\phi}(\{0, 1, \ldots, n-1\})}{n} = 0$$

Proof. Remember that $\phi(x) = T^{h(x)}(x)$ for some measurable function h. For $\varepsilon > 0$ choose a value H so that $\mu(\{x||h(x)| \geq H\}) < \varepsilon$. Now split X into two subsets $A = \{x||h(x)| < H\}$ and $B = \{x||h(x)| \geq H\}$. By the ergodic theorem, for a.e. x, once n is large enough

$$\#\{0 \leq i < n | T^i(x) \in B\} < \varepsilon n.$$

Now once $\varepsilon n > H$, for any i with $\varepsilon n < i < (1-\varepsilon)n$ with $T^i(x) \notin B$ we must have $\phi(T^i(x)) = T^{h(T^i(x)+i}(x)$ with $0 \leq i + h(T^i(x)) < n$ and we conclude

$$\frac{\#\{x, T(x), \ldots, T^{n-1}(x)\} \triangle \phi(\{x, T(x), \ldots, T^{n-1}(x)\})}{n} < 3\varepsilon.$$

□

This lemma tells us that on long finite blocks the map $h_x^{T,\phi}$ is essentially a permutation. We introduce a metric to measure this closeness. For two maps f_1 and f_2 with domains containing $\{0, 1, \ldots, n-1\}$ set

$$d_n(f_1, f_2) = \frac{\#\{0 \leq i < n | f_1(i) \neq f_2(i)\}}{n}$$

the fraction of values in this block where f_1 and f_2 differ. Now also set $S(n)$ to be the group of all permutations of $\{0, 1, \ldots, n-1\}$. We can interpret the lemma as saying that as n grows, the set of maps $h_x^{t,\phi}$ get ever closer in d_n to $S(n)$. In particular, for any $\varepsilon > 0$, once n is large enough, all but ε in measure of the $h_x^{T,\phi}$ will be within ε of $S(n)$. Thinking of the $S(n)$ then as approximations for these maps we can attempt to count how many permutations are needed to achieve the action of ϕ "up to ε". To do this we follow the format of the construction we described of entropy. For each choice of n and ε we can seek the minimal number of permutations in $S(n)$ whose ε neighborhoods in d_n cover all but ε of X. When n is small there may be no such cover, but the lemma guarantees that once n is large enough there will be such covers. Define $N(T, \phi, n, \varepsilon)$ to be this minimum number of permutations needed. It may not be completely obvious but even though $\#S(n) = n!$

the lemma again tells us we expect this number to grow exponentially and we would like to extract the exponential rate so we define

$$\mathcal{C}(T,\phi) = \lim_{\varepsilon\to 0}\limsup_{n\to\infty} \frac{\log_2(N(T,\phi,n,\varepsilon))}{n}.$$

This is a natural asymptotic measure of the complexity of the perturbation of the orbits of T by ϕ.

Theorem 1.8.

$$\mathcal{C}(T,\phi) \leq h(T).$$

Proof. Fix $\varepsilon > 0$ and a large value of n and pick a minimal collection $S \subseteq S^n$ of permutations so that for all but $\varepsilon/6$ of the $x \in X$ there is a $\pi \in S$ with $d_n(h_x^{T,\phi},\pi) < \varepsilon/6$. Let P be a partition of X according to the value π chosen at x, if it is, and an error set E where no permutation is assigned.

For large values N we can use the T,P,N-name of a point to construct a permutation close to $h_x^{T,\phi}$. Starting at x move forward on its T orbit until you hit a first point $T^{j_1}(x) \notin E$. Permute the following block of n points $j_1, j_1+1, \ldots, j_1+n-1$ by the permutation assigned to $T^{j_1}(x)$ by P. Continue forward along the orbit starting at $T^{j_1+n}(x)$ until you find another point $T^{j_1}(x) \notin E$ and again permute the following collection of n indices by the corresponding permutation. Continue this until you reach $T^{N-n-1}(x)$. Define the permutation to fix points outside these blocks. So long as the long permutated blocks cover most of the names our new permutation will be close to the action of $h_x^{T,\phi}$.

To see this more precisely, pick values $\varepsilon/2 > \varepsilon' > 0$. By the ergodic theorem, once N is large enough, for all but ε' of the $x \in X$, all but at most $\varepsilon N/2$ of the set of points $x, T(x), \ldots, T^{N-1}(x)$ is covered by full disjoint blocks of consecutive points $x_i, T(x_i), \ldots, T^n(x_i)$ where $x_i \notin E$ as $\mu(E) < \varepsilon/3$ and points in a partial block would have measure $\leq n/N$.

But now the number of T,P,N-names it takes to cover all but ε' of X is $N(T,P,\varepsilon',N)$ and this tells us that $N(T,\phi,\varepsilon,N) \leq N(T,P,\varepsilon',N)$ so

$$\limsup_{N\to\infty} \frac{\log_2(N(T,\phi,\varepsilon,N)}{N} \leq \lim_{\varepsilon'\to 0}\limsup_{N\to\infty} \frac{\log_2(N(T,P,\varepsilon',N)}{N} = h(T,P).$$

Now letting $\varepsilon \longrightarrow 0$ we are done. □

Lemma 1.9. *Given T ergodic, $\phi \in \Gamma(T)$ and P a finite partition,*

$$|h(\phi^{-1}T\phi,P) - h(T,P)| \leq \mathcal{C}(T,\phi).$$

Proof. The two processes $(\phi^{-1}T\phi,P)$ and $(T,\phi(P))$ are precisely identical and hence have the same entropy. Any T,P,n-name can be permuted into a lot of $T,\phi(P),n$-names, and vice versa, but up to small measure, each gives rise to at most something like $N(T,\phi,\varepsilon,n)$ names in the other, one for each permutation. The rest of the argument is to manage the asymptotics. □

We now want to define a notion of a very tame sequence of perturbations.

Definition 1.10. *Suppose we have a sequence of $\phi_i \in \Gamma(T)$ with $\phi_i^{-1}T\phi_i$ converging in probability to some map T'. We say the sequence of rearrangements T, ϕ_i has "zero asymptotic complexity" if for all $\varepsilon > 0$, once $i < j$ are large enough,*

$$\mathcal{C}(\phi_i^{-1}T\phi_i, \phi_i^{-1}\phi_j) < \varepsilon.$$

Theorem 1.11. *Suppose T, ϕ_i has zero asymptotic complexity and $\phi_i^{-1}T\phi_i$ converges pointwise to T'. Then $h(T) \leq h(T')$*

Proof. Set $\varepsilon > 0$. First as all $\phi_i^{-1}T\phi_i$ are conjugate, they all have the same entropy. Hence we can assume that we start far enough out in the sequence of perturbations that $\mathcal{C}(T, \phi_i) < \varepsilon$. But now for all partitions P,

$$h(\phi_i^{-1}T\phi_i, P) \geq h(T, P) - \varepsilon.$$

As we have pointwise convergence to T', upper semi-continuity of entropy on finite state processes tells us $h(T', P) \geq \limsup_i h(\phi_i^{-1}T\phi, P)$ giving the result. □

Notice that if $\phi_i^{-1}T\phi_i$ converges in probability to T' then automatically $\phi_i T' \phi_i^{-1}$ converges in probability to T. If we asked that these have zero asymptotic complexity in both directions then necessarilly we would have $h(T) = h(T')$. The highlight of our plans now are to get to the following theorem:

Theorem 1.12. *If T and S are ergodic and of equal entropy, then there are elements $\phi_i \in \Gamma(T)$ so that $\phi_i^{-1}T\phi_i$ converges in probability to some T' that is conjugate to S and the sequences of perturbations in both directions are of zero asymptotic complexity.*

To say this in more casual terms, if T and S are of equal entropy then the orbits of T can be perturbed to look like those of S using an exponentially small number of permutations. This to a degree then explains the issue relating orbit equivalence to entropy. One can capture the entropy by controlling the complexity of the orbit changes allowed and this is precisely sufficient to capture it.

In the following sections we will go into more detail on this fact both embedding it into a more general picture of restricted orbit equivalence and showing more detail on how the result is proven.

2. Restricted Orbit Equivalence

In this section we put the ideas sketched in the previous into a general context (see [4] for a complete discussion). Fix $(X, \mathcal{F}, \mu)$, a standard probability space, and on it an orbit relation $\mathcal{O} = \mathcal{O}(U) = \{(x, U^j(x)) | j \in \mathbb{Z}\}$ where U is an ergodic and aperiodic action. Set $\mathcal{A} = \{T | \mathcal{O}(T) = \mathcal{O}(U)\}$. Lastly let Γ be the full group of the orbit relation $\mathcal{O}$, that is to say the collection of all invertible maps ϕ with $(x, \phi(x)) \in \mathcal{O}$ a.s.

For $T \in \mathcal{A}$ and $\phi \in \Gamma$ we call the pair T, ϕ a "rearrangement pair" as we think of ϕ as "rearranging" or "perturbing" the orbits of T to those of $\phi^{-1}T\phi$.

We now define an abstract notion of the "size" of a rearrangment pair, written $m(T, \phi) \in \mathbb{R}^+ \cup \{0, \infty\}$ and satisfies three properties:

(1) For $\varepsilon > 0$ there is a $\delta > 0$ so that if $m(T,\phi) < \delta$ then
$$\mu(\{x|\phi(T(x)) \neq T(\phi(x))\}) < \varepsilon.$$
Notice this last calculation is measuring the degree to which ϕ fails to commute with T.

(2) For $T \in \mathcal{A}$ define
$$m_T(\phi_1,\phi_2) = m(\phi_1^{-1}T\phi_1, \phi_1^{-1}\phi_2)$$
and this should be a pseudometric on Γ, that is to say should satisfy the triangle inequality.

(3) For a pair T,ϕ, where $\phi(x) = T^{h(x)}(x)$ we can construct a map $g = g^{T,\phi} : X \to \mathbb{Z}^{\mathbb{Z}}$ given by $g(j) = h(T^j(x))$. Now $g_*\mu$ is a shift invariant measure on $\mathbb{Z}^{\mathbb{Z}}$. As $\mathbb{Z}^{\mathbb{Z}}$ is Polish in the product topology, the space $\mathcal{M}(\mathbb{Z}^{\mathbb{Z}})$ of shift invariant Borel probability measures on $\mathbb{Z}^{\mathbb{Z}}$ is weak* Polish. For all T,ϕ and $\varepsilon > 0$ there must be a neighborhood O of $g_*\mu$ in $\mathcal{M}(\mathbb{Z}^{\mathbb{Z}})$ so that if $g_*^{T'\phi'}(\mu') \in O$ then
$$m(T',\phi') < m(T,\phi) + \varepsilon.$$
We call this "weak* upper-semicontinuity of m."

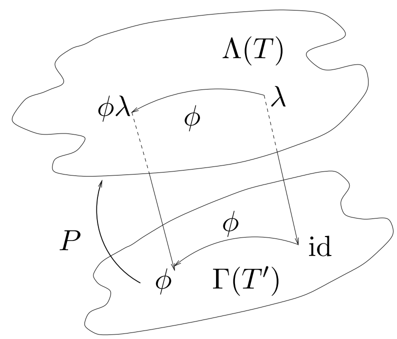
A large number of examples of sizes are known, here are two.

Example 2.1. *Set $m^0(T,\phi) = \mu(\{x|T\phi(x) \neq \phi T(x)\})$. All three conditions for a size are rather easy to show for m^0. Notice that condition (1) tells us this is the weakest possible size, and a rearrangement that is δ-small by some size must be ε-small by m^0.*

Example 2.2. *The complexity of a rearrangement $\mathcal{C}(T,\phi)$ satisfies conditions (2) and (3) but definitely does not satisfy (1). The triangle inequality for condition (2) follows from realizing that the number of permutations used in the composition of two full group elements will be bounded by the product of the number used by each separately. Condition three is a consequence of counting arguments essentially identical to the proof of upper semi-continuity of entropy for finite state systems. To get condition (1) we do a very simple thing, we define a size by*
$$m^e(T,\phi) = \mathcal{C}(T,\phi) + m^0(T,\phi)$$
we just add on the weakest size m^0.

We now give a quick description of how one uses a size to develop an equivalence relation on $\mathcal{A}$, the collection of maps with the same orbits as T. Fix an action T and a size m. If we take $\Gamma(T)$ modulo the equivalence relation of being at an m_T distance 0 we get a separable metric space. Separability follows from condition (3) and the weak* separability of $\mathcal{M}(\mathbb{Z}^{\mathbb{Z}})$. Now let Λ_T be the m_T closer of Γ modulo the equivalence relation of being at m_T-distance zero. This now is a complete and separable metric space. Γ still acts on Λ_T – in fact it acts isometrically. To see this note that
$$\phi_1^{-1}T\phi_1 \longrightarrow \phi_2^{-1}T\phi_2$$
$$\phi^{-1}\phi_1^{-1}T\phi_1\phi = (\phi^{-1}\phi_1\phi)^{-1}(\phi^{-1}T\phi)(\phi^{-1}\phi_1\phi) \longrightarrow \phi^{-1}\phi_1^{-1}T\phi_2\phi$$
are identical weak*. Hence the map $\phi_i \to \phi^{-1}\phi_i$ is an m_T isometry and extends as an isometry to the closure.

Next notice that by (1) if the ϕ_i are m_T-Cauchy then the maps $\phi_i^{-1}T\phi_i$ converge in probability to some $T' \in \mathcal{A}$. That is to say, for any $\lambda = \{\phi_i\} \in \Lambda(T)$ we obtain an element $T' \in \mathcal{A}$. Now $\Gamma(T') = \Gamma(T)$. Define a map P that takes $\mathrm{id} \in \Gamma(T') \to \lambda \in \Lambda(T)$ and extend this map P to all of $\Gamma(T')$ by setting $P(\phi) = \phi\lambda$.



Condition (3) tells us that P is a contraction from $m_{T'}$ to m_T and so it extends to a contraction from $\Lambda(T') \to \Lambda(T)$. Although the space $\Lambda(T)$ is an abstract closure think of the point λ as representing T' and the identity as representing T. In this sense then P is actually the identity as it carries $\mathrm{id}(= T')$ in $\Lambda(T')$ to $\lambda(= T')$ in $\Lambda(T')$.

Suppose now that the identity is in $P(\Lambda(T'))$. That is to say, T can be reached by a Cauchy sequence of rearrangements in $\Lambda(T')$. In this case then P is in fact an isometry. This tells us that if $\lambda = \{\phi_i\}$ then $\{\phi_i^{-1}\}$ is $m_{T'}$ Cauchy, giving an explicit Cauchy sequence of rearrangements taking T' back to T. When this is true we say T and T' are m-equivalent. This is true of a residual subset of $\Lambda(T)$. That is to say, one can show that the collection of $\lambda \in \Lambda(T)$ for which the corresponding maps P are isometries are a residual subset of $\Lambda(T)$. This means that the equivalence classes are topologized by m_T as Polish spaces.

We say two actions T and S are m-equivalent if $S \cong T' \in \mathcal{A}$ and T' is m-equivalent to T in the sense we just described. One might ask if it is really necessary to consider the possibility that the identity might fail to be in $P(\Lambda(T'))$. Our size m^e in fact says it is. If we started with a choice of T that is of zero entropy then on $\Gamma(T)$, m^e will simply be m^0. It is possible to construct m_T^0 Cauchy sequences of perturbations that lead to actions T' that are of positive entropy. In fact all of $\mathcal{A}$ is reachable by such m_T^0 Cauchy sequences of rearrangements by Dye's theorem. But we could never get an m_T^e Cauchy sequence leading back from T' to T.

Having seen that any size gives rise to an equivalence relation one can naturally ask for any particular choice of size what is this equivalence relation? The two sizes we want to think about in this context are m^0 and m^e. Our goal is to gain some insight into why any two free ergodic actions are m^0 equivalent (Dye's Theorem) and that any two actions of the same entropy are m^e equivalent (our main result).

3. Rokhlin's Lemmas and Dye's Theorem

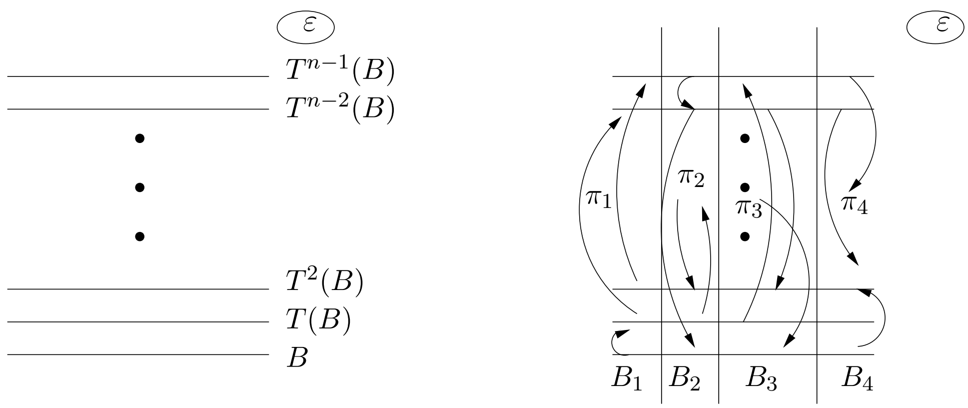
In this section we will develop some basic machinery and use it to give a sketch of how one would prove Dye's Theorem building on the notions of the previous section. To begin we remind the reader of Rokhlin's lemmas

Lemma 3.1 (Rokhlin)**.** *For T an ergodic transformation of a nonatomic probability space $(X, \mathcal{F}, \mu)$ and $n \in \mathbb{N}$ and $1 \geq \varepsilon > 0$ there exists a subset $B \in \mathcal{F}$ with $B, T(B), \ldots, T^{n-1}(B)$ all disjoint and with*

$$\mu(\cup_{i=0}^{n-1} T^t(B)) = 1 - \varepsilon.$$

This statement is just a little different from what one usually sees in that we ask that the measure of the union be precisely $1 - \varepsilon$ where one usually only asks that it be $\geq 1 - \varepsilon$. It is a simple thing to "shave" down the tower to get this extra detail.

We can now use this lemma to see how to construct maps $\phi \in \Gamma(T)$. Here is how. Having constructed the set B as described, partition B into subsets $B_1, B_2, \ldots, B_t$ and pick some list of permutations $\pi_1, \pi_2, \ldots, \pi_t \in S(n)$. Now for any $x_0 \in B_k$, $x = T^j(x_0)$ and $0 \leq j < n$ set $\phi(x) = T^{\phi_k(j)}(x_0)$.

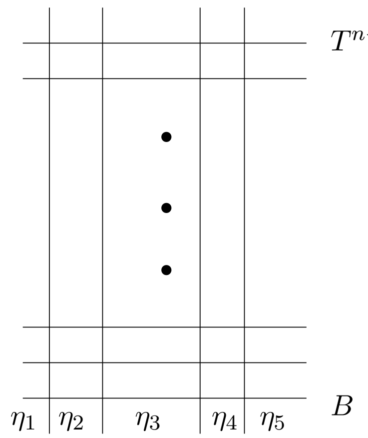


To use this effectively we need a stronger statement of Rokhlin's lemma

Lemma 3.2 (Strong Rokhlin)**.** *For P a finite partition of X and $\varepsilon > 0$ the set B in the statement of the Rokhlin lemma can be chosen so that*

$$B \perp \bigvee_{i=0}^{n-1} T^{-i}(P).$$

the picture here is that one cuts the set B according to T, P, n-names, which we can view as cutting the tower vertically into columns that, at each height in the tower lie in some fixed element of P. What the Strong Rokhlin Lemma tells us is that B can be chosen so that the relative sizes in B of each of these names is precisely the same as its size in X.



Rokhlin stack cut according to T, P, n-names

The two pictures, one of how we construct elements $\phi \in \Gamma(T)$ and the other of the Strong Rokhlin Lemma strongly suggest how we will proceed to understand Dye's Theorem. We will be choosing the permutations π_i to act on the names η_i so as to model some target transformation T by a rearrangement of S.

Here is how we start to construct such a "model" of one system inside another. Suppose we have T acting on $(X, \mathcal{F}, \mu)$ and S acting on $(Y, \mathcal{G}, \nu)$ and we take a partition $P = \{s_1, \ldots, s_t\}$ of X as some extremely rough first picture of the space X. As Y is nonatomic we can construct a partition $P' = \{s'_1, \ldots, s'_t\}$ of Y so that $\mu(s_i) = \nu(s_i)$. This "0th order" model of X inside Y captures nothing of the dynamics and very little of the measure space. In particular as we look at S, P', n-names, they need look nothing like T, P, n-names. Well that is not quite true. Once $n >> 1$ the T, P, n-names and S, P', n-names will have one thing in common, from the Ergodic Theorem. The fractions of most S, P', n-names occupied by a particular symbol i will be very close to $\nu(s'_i)$ which will be very close to the fractions of most T, P, n-names occupied by i. Stated more precisely:

Lemma 3.3. *For $\varepsilon > 0$ once n is large enough, all but ε in measure of the T, P, n-names and ε in measure of the S, P', n-names will give densities to all the corresponding symbols i within a fraction ε of $\mu(s_i) = \nu(s'_i)$.*

We describe the next step toward improving our model. Using the Strong Rokhlin Lemma construct towers of height n in each system with bases B and B' where

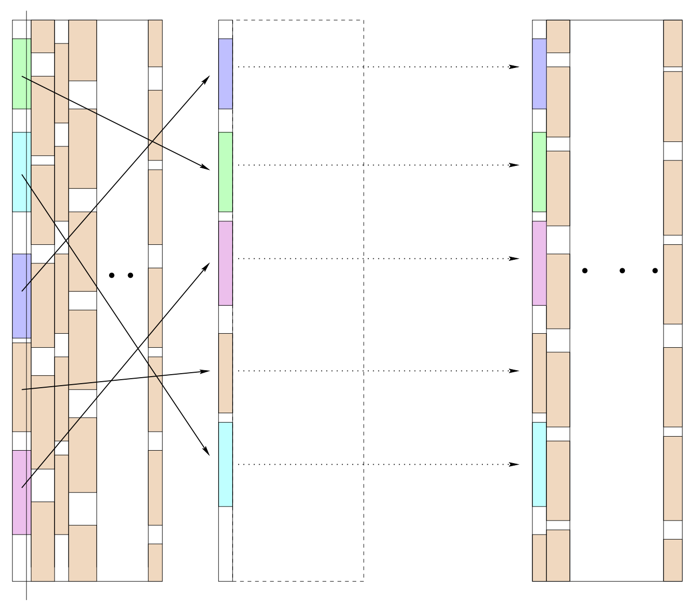
$$B \perp \bigvee_{i=0}^{n-1} T^{-i}(P) \text{ and } B' \perp \bigvee_{i=0}^{n-1} S^{-i}(P').$$

Cut B' into subsets of relative measures $\mu(\eta_1), \mu(\eta_2), \ldots$ where the η_i are the T, P, n-names. Now B' is cut in two ways, first into S, P', n-names η'_i and now into subsets of relative measure $\mu(\eta_i)$. Let $B'_1, B'_2, \ldots$ be the span of these two partitions of B'. Now associated to each B'_j we have two names, an η'_k and an η_j. The η'_k is the actual name up the tower and the η_j is the name we would like to see to improve our model of T.

For all but a subset of the B'_j of relative measure 2ε, the density of occurrences of any symbol i in η'_k and in η_j differ by a fraction at most 2ε. So away from this set of "bad" B_j's we can select a permutation π_j so that the permutated name $\pi_j \eta'_k$ and η_j agree at all but a fraction 2ε of their indices.

We now make two "perturbations" to improve our model. First, we construct $\phi_1 \in \Gamma(S)$ using these permutations so that the $\phi_1^{-1} S \phi_1, P', n$-names up the Rokhlin tower, away from the bad names, now differ on a fraction at most 2ε of indices from the paired T, P, n-name. Second we modify the partition P' to a new partition P'_1 by changing the label of at most 4ε of the points (points in bad names, and points in good names that are in error) so that the distribution of $\phi^{-1} S \phi, P'_1, n$-names on the Rokhlin tower and T, P, n-names on its Rokhlin tower are identical. This "perturbation" takes us to our first level approximation given as a pair $\phi_1 \in \Gamma(S)$ and P'_1.

What we now need to see is that we can continue making such modifications where the ϕ_i will be m^0-Cauchy and the permutations P'_i will be Cauchy in the partition metric. What one should notice is that in our first step we have modified the Rokhlin tower in Y by rearranging the action and modifying the partition to create $\phi_1^{-1} S \phi, P'_1, n$-names that are an exact copy of the Rokhlin tower in X labeled by T, P, n-names. Our program now is to continue through a series of such constructions inductively. Each inductive step will start with a pair of towers of height n_i labeled by names whose distributions on the two towers are identical. Our problem then is to build a much taller tower of height n_{i+1} and see how to modify the names on these towers to be identical. The following diagram is meant to give a schematic of how this will be accomplished.



Suppose we have finished stage i by creating a new action S_i that is conjugate to S by an element of $\Gamma(S)$ and a new partition P'_i and we have two Rokhlin towers on which the distribution of T, P, n_i-names and S_i, P'_i, n_i-names are identical. Now to move to step $i+1$ first partition the two towers according to the n_i-name up the tower containing a point. We think of this as coloring stripes up the tower in distinct colors, one for each name. As colored then these two towers are identical. Call these auxiliary partitions H_i

and H_i'. Now build a much taller tower, of height n_{i+1} using the strong Rokhlin lemma for both S_i and T with the corresponding bases chosen independant of the S_i, H_i', n_{i+1}-and T, H_i, P-names and so that the towers cover precisely $1 - \varepsilon_{i+1}$ of their corresponding spaces. When you partition the $i+1$ stage towers according to the H_i' and H_i names you will see something like what the diagram shows. A name up the tower will be a sequence of colored blocks, mostly complete, perhaps one at the top and one at the bottom only partially in the tower. It is only the first names, η_1 and η_1' where we have shown blocks with distinct colors. If you set a value $\varepsilon_{i+1} > 0$ and now choose n_{i+1} large enough then by the ergodic theorem, in all but a fraction ε_{i+1} of the names up each tower, the numbers of occurences of blocks of any particular color will be within a fraction ε_{i+1} of the number $n_{i+1} \times$ (the measure of the block of that color in the previous tower). (To claim this we must use the fact that the previous towers covered precisely $1 - \varepsilon_i$ of their measure spaces.) That is to say, for most names in these new towers the numbers of blocks of each color is about the same. We want to rearrange the S_i, H_i, n_{i+1}-names to look very much like the T, H_i, n_{i+1}-names. This says that for all but $2\varepsilon_{i+1}$ in measure of the S_i, H_i', n_{i+1}-names, we can do this by translating most of the colored blocks as a rigid block. To be precise, on all but a fraction $2\varepsilon_{i+1}$ of the base of the S_i tower, all but $(\varepsilon_i + 3\varepsilon_{i+1} + 2n_i/n_{i+1})n_{i+1}$ of the indices will be moved as rigid colored blocks. We have to omit the at most $(\varepsilon_i + \varepsilon_{i+1})n_{i+1}$ fraction that lie outside of colored blocks and then the $2\varepsilon_{i+1}n_{i+1}$ indices that lie in colored blocks where we cannot match the color and then perhaps $2n_i$ levels in partial colored blocks at the top and bottom of the tower. We choose ε_{i+1} small enough and n_{i+1} large enough that this is at most $2\varepsilon_i n_{i+1}$ The diagram shows us cutting off a piece of η_1' of the proper width and then rearranging the colors by ϕ_i to look like the sequence of colors in η_1. After such a permutation, the $\phi_{i+1}^{-1} S_i \phi_{i+1}, P_i', n_{i+1}$-name up this piece of the tower differs in at most $2\varepsilon_i n_{i+1}$ indices from the corresponding T, P_1, n_{i+1}-name. Thus we can now replace P_i' by a P_{i+1}' with $\nu(P_i' \triangle P_{i+1}') < 2\varepsilon_i$ and have our new tower identical to our old.

The fact that we rearranged so much of the space in long rigid colored blocks allows us to conclude that $m^0(S_i, \phi_{i+1}) < 2\varepsilon_{i+1} + 2\varepsilon_i + 1/n_i$. By choosing the ε_i to decay fast enough and the n_i to grow fast enough, we can ensure this is $< 3\varepsilon_i$. Of course we also note that after this modification the new labeled towers are again identical in distribution and we are ready to continue the induction.

Setting $\psi_k = \phi_1 \phi_2 \dots \phi_k$ and assuming the ε_i are summable, we obtain ψ_k is m_S^0-Cauchy and hence $\psi_k^{-1} S \psi_k$ converges in probability to an action S'. Also, as the P_i' are Cauchy in the partition metric, they converge to some limiting partition P'' and S', P'' will be identical in distribution to T, P. This then gives us a conjugacy of these actions restricted to the sub-σ-algebras these partitions generate.

This doesn't prove the theorem as we do not have either P or P' are necessarilly generating partitions. How do we deal with this final "detail"? It is in fact not very hard. At any stage i in the construction we can bring in another partition, either Q of X or R of Y, and further refine the corresponding tower by the T, Q, n_i-names or the S_i, R, n_i-names. Having done this the towers are no longer identical, but we can create a partition Q' of the tower in X or R' of the tower in Y by painting appropriate names on the tower so that the refined towers are identical in distribution. What does this tell us? Well it means we can actually have refining and generating sequences of partitions Q_j of X and R_j of Y so that there are corresponding partitions R'_j of X and Q'_j of Y and for all j the $T, Q_j \vee R'_j$ and $S', Q'_j \vee R_j$ are identical in distribution. This now is enough to imply that T and S' are conjugate.

Although a number of complications enter into this proof the core idea is a succession of rearrangements constructed on Rokhlin towers that, for the most part, translate long contiguous blocks. The two main tools used were the Rokhlin lemma to give us large towers to work on and the ergodic theorem to tell us some basic information about what names up the tower looked like. In the process of the proof we learned a few tricks about how to manipulate towers, both by rearranging the orders of points up the tower and changing the partition names on sets in the tower. What we now want to consider is how we might control the number of permutations we need to use in order to carry out the steps in this induction and how we can use entropy to control this number.

4. Entropy and the Shannon-McMillan-Breiman Theorem

We have discussed and used two of the core tools in measure preserving dynamics already, the ergodic theorem and Rokhlin lemmas. We now introduce the third, The Shannon-McMillan-Breiman Theorem. To begin we remind the reader that for T a measure preserving transformation of $(X, \mathcal{F}, \mu)$ and a finite partition P of X, by $P_n(x)$ we mean both the set in $\vee_{i=0}^{n-1} T^{-i}(P)$ containing x and the T, P, n-name of that set. Recall that entropy $h(T, P)$ measures the exponential growth rate of the number of such names. The Shannon-McMillan-Breiman Theorem concerns the dual issue of the size of such names.

Theorem 4.1 (Shannon-McMillan-Breiman)**.** *For T ergodic and acting on $(X, \mathcal{F}, \mu)$ and P a finite partition, for a.e. $x \in X$*

$$\lim_{n\to\infty} \frac{\log_2(\mu(P_n(x)))}{n} = -h(T, P).$$

This tells us that for $\varepsilon > 0$ and most x once n is large enough

$$\mu(P_n(x)) = 2^{-(h(T,P)\pm\varepsilon)n} = \left(2^{-h(T,P)n}\right)\left(2^{\pm\varepsilon}\right)^n.$$

It is very important to understand the error value $(2^{\pm\varepsilon})^n$ as it can be both extremely large and extremely close to 0. So the theorem is not telling us that the sizes of most names are all about equal. It does though give us substantial control and at the orders of magnitude one would expect.

As a first discussion we wish to tie this result to our ability to make small changes in the names on a Rokhlin tower. The point of doing this will be to gain some understanding of how one might use the Shannon-McMillan-Breiman (SMB) theorem on towers. Suppose we have an ergodic action T on $(X, \mathcal{F}, \mu)$ and a finite partition P with $h(T, P) > 0$. Fix a value $1/4 > \varepsilon > 0$ and choose a value n so that all but ε of the $x \in X$ have $\mu(P_m(x)) = 2^{-(h(T,P)(1\pm\varepsilon)m}$ for all $m \geq n/2$. For convenience assume ε is rational and $\varepsilon n \in \mathbb{N}$.

Now construct a Rokhlin tower of height n and base $B \perp \vee_{i=0}^{n-1} T^{-1}(P)$. As usual we partition B according to the T, P, n-names η_i. We have from the SMB theorem that outside a "bad" set $E \subseteq B$ with $\mu(E) < \varepsilon\mu(B)$ we have some control on the sizes of the remaining η_i. Now take the top $2\varepsilon n$ levels of the tower ($T^i(B)$ for $n(1-2\varepsilon) \leq i < n$) and replace the labels of the sets here with some fixed name, say $1, 1, 1, \ldots, 1$. This is a small perturbation of P to a partition $\overline{P}$. Now the partition of B into new $T, \overline{P}, n$-names is just the old partition but into $T, P, (1-2\varepsilon)n$-names. Now outside of the "bad" set E of $T, P, (1-2\varepsilon)n$-names, we have that the relative sizes of the names η_i in $B - E$ is

$$2^{-(h(T,P)(1\pm\varepsilon)(1-2\varepsilon)n}$$

and this says

$$2^{-h(T,P)(1-\varepsilon)n} \leq \mu(\eta_i) \leq 2^{-h(T,P)(1-2\varepsilon)n}.$$

That is to say, all our good $\overline{P}$ names are on the large side of $2^{-h(T,P)n}$ but not too far on the large side exponentially speaking.

Now we do the following. Take each of the good $\overline{P}$ names η_i partitioning the base B and cut them into as many whole pieces precisely of size $2^{-h(T,P)n}$ as you can with perhaps a bit of each left over of size $< 2^{-h(T,P)n}$. Now take all these leftovers and we will add them onto E. How much of B is in these leftovers? Well there are at most $2^{h(T,P)(1-\varepsilon)n}$ good names η_i and each gives a leftover of at most $2^{-h(T,P)n}$. Multiplying we see that the leftovers have total mass at most $2^{-h(T,P)\varepsilon n}$. We ask that n be chosen large enough that this is at most ε. The picture we are left with is of a base set $B - E$ for a tower covering all but 3ε of X and it is cut into pieces all of precisely the size $2^{-h(T,P)n}$. Each of these pieces has a fixed $T, \overline{P}, n$ name where P and $\overline{P}$ only differ in the top $2\varepsilon n$ levels of the tower. This construction gives some insight into how one can use the exponential nature of the error in the SMB theorem to have small changes have substantial effects.

5. A Counting Argument

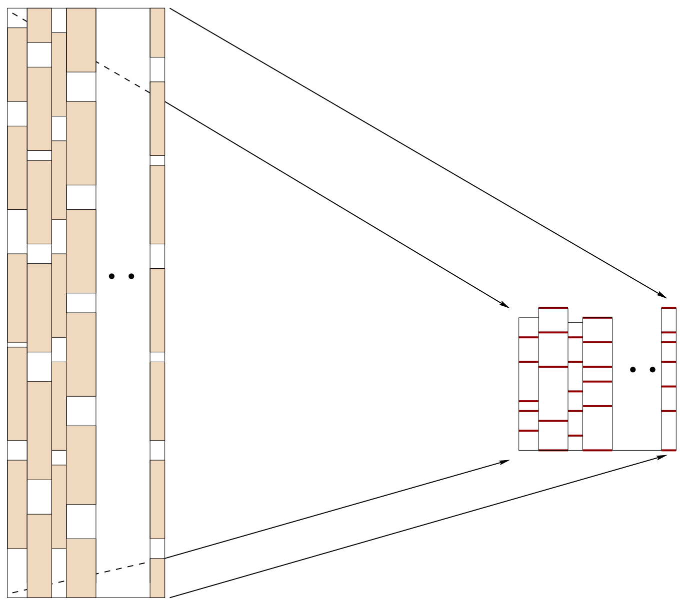
In this section we give a picture of the core combinatorial argument needed to prove our main theorem. What we will start with now is an overview of the core step in our proof of Dye's Theorem and a combinatorial estimates from that picture. We will not be precise in this argument as it is not an ingredient of our proof. Rather it is a heuristic investigation of the number of permutations a priori used at each stage in our proof of Dye's Theorem and indicates why entropy has no reason to be preserved.

Recall the induction step of our proof of Dye's Theorem. We had two towers, one in each measure space, of heights n_i, covering exactly $1 - \varepsilon_i$ of each space, and labeled by partitions P and P'_i so that the distributions of names on the two towers were identical. We colored the strips through the tower corresponding to each of these names a distinct color and then built much taller towers, now of heights n_{i+1} and in each we cut the towers vertically again into names which now take the form of sequences of colored blocks. Consider now the rearranging we do of these colored columns. In most names, most of the colored blocks will be moved rigidly. View this as taking place in three steps. First we collapse the colored blocks to consist of just one level. Then we permute these levels in any way we choose. Then we re-expand the colored levels to be a full block. When we collapse down the colored blocks we change the height of the tower to be $n'_{i+1}(x)$. This height is variable, depending on $x \in B$. We can estimate this value using the ergodic theorem. First off, there will be around $\varepsilon_i n_{i+1}$ levels in a name not in colored blocks. Second, when we collapse blocks they shrink by a factor $1/n_i$ so the collapsed blocks occupy around $(1 - \varepsilon_i)n_{i+1}/n_i$ levels.

Lastly the error from the ergodic theorem, on all but a fraction ε_{i+1} of B will be $\varepsilon_{i+1}n_{i+1}$. This tells us that for all but a fraction ε_{i+1} of B,

$$n'_{i+1}(x) = (\varepsilon_i + \frac{(1-\varepsilon_i)}{n_i} \pm \varepsilon_{i+1})n_{i+1} = \delta_i n_{i+1}$$

where δ_i is a very small but essentially constant value. We write it δ_i as its value is essentially determined before the value n_{i+1} is determined.



How many permutations might we use on each name?

$$(\delta_i n_{i+1})! \approx 2^{\delta_i n_{i+1} \log_2(\delta_i n_{i+1})}.$$

This is superexponential in n_{i+1} and so even though we are moving these long blocks rigidly there is no a priori reason that the ϕ_i constructed in our proof should have small complexity.

With that we will now move on to the real combinatorics of our problem. We present this argument in a synthetic form without reference to our towers and names but the connection to that picture should be fairly evident.

Fix a (large) value n and let $k_1, k_2, \ldots, k_t$ be given and fixed with $\sum k_i = n$. Consider the set of all lists (words) $\mathcal{L}$ of length n in the symbols $\{1, 2, \ldots, t\}$ subject to the constraint that symbol i appears precisely k_i times.

There are $N = \binom{n}{k_1, k_2, \ldots, k_t} = \frac{n!}{k_1!k_2!\ldots k_t!}$ such names.

Two facts:

(1) For $\varepsilon > 0$, once n is large enough, $N \approx 2^{(h(\alpha)\pm\varepsilon)n}$ where $\alpha = \{k_1/n, k_2/n, \ldots, k_t/n\}$ and $h(\alpha) = -\sum_{i=1}^{t} \alpha_i \log_2 \alpha_i$.

(2) For any choice of a single $\eta \in \mathcal{L}$, if you act on this word by all permutations $\pi \in S(n)$ the image will be all of $\mathcal{L}$ covered uniformly. That is to say for all $\eta' \in \mathcal{L}$,

$$\#\{\pi \in S(n) | \pi(\eta) = \eta'\} = n!/N = k_1!k_2! \dots k_t!.$$

Continuing our work, consider two lists of precisely N words chosen from $\mathcal{L}$,

$$L_1 = \{\eta_1, \eta_2, \dots, \eta_N\} = \mathcal{L} \text{ is 1-1 and}$$

$$L_2 = \{\eta'_1, \eta'_2, \dots, \eta'_N\} \text{ allows multiplicities.}$$

We are looking for lists of permutations π_i, $i = 1, 2, \dots, N$ and a bijection $H : \{1, 2, \dots, N\} \hookleftarrow$ so that $\pi_i(\eta'_i) = \eta_{H(i)}$. That is to say we want to permute the names in our second list to create a list of all possible names. We know we can do this. Call such a pair $\{\pi_i\}, H$ a "matching of L_2 to L_1". The issue for us though is to minimize the number of permutations π_i needed. Here is what we can prove.

Theorem 5.1. *Suppose the words in L_2 occur with maximum multiplicity K and $\varepsilon > 0$. One can then find a matching $\{\pi_i\}, H$ of L_2 to L_1 so that all but a fraction of at most ε of $1, 2, \dots, N$ are covered by fewer than K/ε^2 distinct choices for π_i.*

Proof. The argument is by a greedy algorithm. We will seek inductively to choose a new value π_i that will allow us to cover as many uncovered names as possible. Suppose we have a collection of $s > \varepsilon N$ elements $S_1 \in L_1$ and $S_2 \in L_2$ that have not yet been assigned permutations. S_2 can have multiplicities so select a maximal subset $S'_2 \subseteq S_2$ of distinct names. Now $\#S'_2 \geq \varepsilon N/K$. Act on each element of $\eta'_i \in S'_2$ by all of $S(n)$ and for each η'_i we will obtain a uniform cover of L_1 and so a uniform cover of S_1. That is to say a fraction of at least ε of the $\pi \in S(n)$ put $\pi(\eta'_i)$ in S_1. Thus from among the $n!\#S'_2$ pairs π, η'_i a fraction $\#S_1/N$ of them have $\pi(\eta'_i) \in S_1$. This means that for some choice of π we must have at least this fraction of $\pi(S'_2)$ in S_1. Now π acting on S'_2 is 1-1 so

$$\#\{\eta'_i \in S'_2 | \pi(\eta'_i) \in S_1\} \geq \#S'_2\#S_1/N \geq \varepsilon N/K \times \varepsilon N/N = \varepsilon^2 N/K$$

We will now set $\pi = \pi_i$ for those $\eta_i \in S'_2$ and $\pi(\eta'_i) \in S_1$. This extends our definition to at least $\varepsilon^2 N/K$ further words η'_i.

This argument can be applied at most K/ε^2 times without a contradiction. This means that at or before we have applied it this many times, hence used at most K/ε^2 distinct permutations, we must find the remaining words occupy less than a fraction ε of the full list. This completes the proof. □

Applying this to complete our theorem entails two things. First we need to understand how to link this counting argument to the picture we have of towers and the rearranging of blocks along those towers. Second we need to transform the estimate from the theorem above to that context and see what it gives.

The first issue is relatively clear. At each stage of our proof of Dye's theorem we had towers where the densities of occurrences of colored blocks up most names were essentially constant. By a small change in the names we could make them actually constant. Ignoring a lot of details then what we see is much like the content of our combinatorial lemma, telling us how many permutations we need to modify these colored names to create a listing of all possible sequences with this distribution of colors. Our estimate on the number of permutations looks like K/ε^2. What will K be? Our goal is to make it look like $2^{\varepsilon n}$ where our tower has height n. On the ε of the names where we might have to use more permutations than the theorem guarantees we opt to change the name rather than permute the existing name. The ε^2 in the denominator is exponentially insignificant. These permutations and partition changes move our picture to a "general position" of all possible orderings. We can do the same to the target process, moving it to this same general position. Combining these two moves will modify our original tower to look like our target tower. The number of permutations used on the tower will then be bounded by the square of a small exponential and hence will still be a small exponential. This will tell us the ϕ we construct is of small complexity. There is much to do to make this hand-waving discussion precise. It will require substantial basic work and is intended for those with sufficient background. In the next section we will see how to greatly simplify our efforts by aiming our rearrangements for a very precise kind of target, a Bernoulli system.

6. Ornstein's Theory of Bernoulli Systems

Restricted orbit equivalence is a natural extension of the groundbreaking work of D.S. Ornstein on the structure of Bernoulli systems. In fact to carry out our work here from first principals we would be forced to reconstruct much of that theory. We can though simply take its core results and use them to focus our work on a single construction.

Definition 6.1. *By a Bernoulli system, transformation or shift we mean any measure preserving transformation that is conjugate to the shift map on some i.i.d. sequence of random variables.*

Ornstien's premier result is the *Isomorphism Theorem*[5]:

Theorem 6.2. *If two Bernoulli transformations have the same entropy, finite or infinite, then they are conjugate.*

This theory is most easily understood in the realm of processes. For us a process is a measure preserving transformation T together with a finite partition P of the measure space X. We regard a finite partition as a map $P : X \to \mathcal{S}$ where $\mathcal{S}$ is some labeling space for the partition elements. We can usually just take it to be $\{1, 2, \dots, t\}$ for some $t \in \mathbb{N}$. The doubly infinite T, P names give a map $\eta : X \to \mathcal{S}^{\mathbb{Z}}$ that intertwines T with the left shift map σ. Thus $\eta^*(\mu)$ is a σ invariant Borel measure on $\mathcal{S}^{\mathbb{Z}}$. The space of all such measures in the weak* topology is a compact, convex metrizable space.

We can put an explicit metric on this space as

$$\mathrm{dist}(\nu, \nu') = \frac{1}{2} \sum_{n=1}^{\infty} \frac{1}{2^n} \sum_{\vec{s} \in \mathcal{S}^n} |\nu(\vec{s}) - \nu'(\vec{s})|.$$

That is to say, we take the cylinder sets of length n and compare the mass ν and ν' give them respectively and then create this weighted sum. The leading 1/2 here is simply to bound the computation by 1.

For any two processes (T, P) and (T', P') labeled by the same symbol set $\mathcal{S}$ we set

$$\mathrm{dist}(T, P; T', P') = \mathrm{dist}(\eta^*(\mu), \eta'^*(\mu')).$$

This is not a metric, just a pseudometric. For two processes then to be close in distribution simply means that for some large n the cylinder sets of length n with corresponding names have very nearly the same masses.

A second metric plays a critical role in the Bernoulli theory and more generally in the theory of conjugacy of measure preserving systems. This is the $\overline{d}$ metric and again it is on processes. Suppose ν and ν' are two shift invariant measures on $\mathcal{S}^{\mathbb{Z}}$. By a joining $\mathcal{J}$ of ν and ν' we means a shift invariant measure on $(\mathcal{S} \times \mathcal{S})^{\mathbb{Z}}$ whose two coordinate projections are ν and ν' respectively. This is a compact and convex set of Borel measures and is nonempty as it contains $\nu \times \nu'$. We call it $J(\nu, \nu')$. For a given joining $\mathcal{J}$ we can calculate how closely the symbols of the two processes are matched by calculating

$$D(\mathcal{J}) = \mathcal{J}(\{(\vec{s}, \vec{s}')|s_0 \neq s_0'\}).$$

This is a continuous linear functional and is nonnegative on the space of joinings and hence achieves its minimum on the boundary of $J(\nu, \nu')$. We set

$$\overline{d}(T, P; T', P') = \min_{\mathcal{J} \in J(\eta^*(\mu), \eta^*(\mu'))} D(\mathcal{J}).$$

If ν and ν' are ergodic measures then the boundary of the space of joinings is the set of ergodic joinings. This means that the $\overline{d}$ distance is attained on an ergodic joining. Hence what $\overline{d}$ is measuring is the density along $\mathcal{J}$ a.e. orbit of the positions where the two coordinate symbols disagree.

Even if T itself is not Bernoulli we can say (T, P) is Bernoulli if T restricted to the sub algebra $\vee_{i=-\infty}^{\infty} T^{-i}(P)$ is Bernoulli. This is equivalent to saying $\eta^*(\mu)$ on $\mathcal{S}^{\mathbb{Z}}$ makes σ a Bernoulli shift.

We now list some core results from the Ornstein theory.

Theorem 6.3 (Ornstein [8])**.** *For $\mathcal{S}$ a finite symbol set, the shift invariant measures ν on $\mathcal{S}^{\mathbb{Z}}$ which are Bernoulli form a $\overline{d}$-closed set.*

Theorem 6.4 (Ornstein and Shields [6])**.** *For $\mathcal{S}$ a finite symbol set, the shift invariant mixing Markov measures on $\mathcal{S}^{\mathbb{Z}}$ are Bernoulli.*

Theorem 6.5 (Ornstein [8])**.** *Suppose T acting on $(X, \mathcal{F}, \mu)$ is measure preserving and we have T invariant subalgebras $\mathcal{H}_i \nearrow \mathcal{F}$ on which T is Bernoulli. Then T acting on $\mathcal{F}$ is Bernoulli.*

We need one last ingredient here, the Ornstein and Weiss characterization of the Bernoulli processes.

Definition 6.6. *We say a process (T, P) is finitely determined if for all $\varepsilon > 0$ there is a $\delta > 0$ so that for any ergodic (T', P') satisfying*

i) $h(T', P') > h(T, P) - \delta$ *and*
ii) *dist*$(T, P; T', P') < \delta$ *then*
iii) $\overline{d}(T, P; T', P') < \varepsilon$.

Theorem 6.7 (Ornstein and Weiss [7])**.** *A process (T, P) is Bernoulli iff it is finitely determined.*

What these results do for us is to create a target for our orbit equivalence. What we will do is to show that for any ergodic T of positive entropy, T is m^e equivalent to a Bernoulli shift of the same entropy. The isomorphism theorem will now complete our work, telling us that any two transformations of the same entropy are m^e equivalent to isomorphic Bernoulli actions.

There is a strong connection between m^e and the notion of finitely determined process that we express through the following two lemmas.

Lemma 6.8. *For all $\varepsilon > 0$ there is a $\delta > 0$ so that for any process (T, P), if $m^0(T, \phi) < \delta$ then*

$$dist(T, P; \phi^{-1}T\phi, P) < \varepsilon.$$

Proof. Notice that if $T^i\phi(x) = \phi T^i(x)$ for $i = 0, \dots, n-1$ then the T, P, n-name of x and the $\phi^{-1}T\phi, P, n$-name of $\phi^{-1}(x)$ are identical. Hence once $m^0(T, \phi)$ is small enough, ϕ^{-1} carries the distribution of T, P, n-names almost identically to the distribution of $\phi^{-1}T\phi, P, n$ names. This implies these two processes are close in distribution. One also sees from this that the value δ does not depend on the process. □

Lemma 6.9. *For all $\varepsilon > 0$ there is a δ so that for any process (T, P), if $\mathcal{C}(T, \phi) < \delta$ then*

$$h(\phi^{-1}T\phi, P) > h(T, P) - \varepsilon.$$

Proof. One can calculate the rather elaborate inequality

$$N(\phi^{-1}T\phi, P, n, 2\varepsilon) \le N(T, P, n, \varepsilon)N(T, \phi, n, \varepsilon)\binom{n}{\varepsilon n}(\#P)^{\varepsilon n}.$$

by counting the number of $\phi^{-1}T\phi, P, n$-names one might create from a single T, P, n-name. Taking logarithms and limits gives the result. Once more notice the independence here of the process (T, P). □

What these two lemmas tell us is that a small m^e rearrangement moves a process by a small amount in both senses i) and ii) of the finitely determined condition. We can note an interesting corollary to this:

Corollary 6.10. *In any m^e equivalence class that contains a Bernoulli shift, the Bernoulli shifts form a residual set.*

Proof. If a class contains a Bernoulli shift T, as the maps $\phi^{-1}T\phi$ are dense, the Bernoulli shifts are dense. On the other hand, our two lemmas tell us that for any Bernoulli shift, any partition P and any $\varepsilon > 0$ those S with $\overline{d}(T, P; S, P) < \varepsilon$ contains an m^e neighborhood of T. Hence those S in the $\overline{d}$ closure of the (T, P) where T is Bernoulli contains a residual set, i.e. those S for which (S, P) is Bernoulli contain a residual subset. Now intersect these sets over a countable dense collection of partitions P. Any S in this intersection will be Bernoulli for all partitions and hence Bernoulli. □

We can now pursue our goal to show that every positive entropy m^e equivalence class contains a Bernoulli shift.

7. The Final Step

In this last section we give a fairly complete outline of how to complete the proof. It is intended for those who really want to understand and uses a lot of ideas from entropy theory, symbolic dynamics and the Ornstein theory.

To start, let's outline now the path to our conclusion that two ergodic transformations are m^e equivalent if they have the same entropy. First, as noted earlier, we already know this if the transformations have zero entropy so we will assume our transformations have positive entropy. Starting with an ergodic action T on $(X, \mathcal{F}, \mu)$ with $h(T) > 0$ we select a sequence of partitions Q_k which refines to points in X. Hence lumping states of these partitions will give us a dense family of partitions. Our plan is to construct an m^e Cauchy sequence of rearragements $T_k = \phi_k^{-1} T_{k-1} \phi_k$. Stage k will start with the next partition Q_k and restrictions from the previous stages that set an upper bound on ε_k and a lower bound on n_k. These parameters will be set in the construction of stage k. As in Dye's theorem it will be convenient to work with a partition P_k of the tower into colored blocks, one for each T, Q_k, n_k-name. This partition will generate the partition $\overline{Q}_k$ that consists of the Q_k names erased outside the tower.

Modeled on this colored tower we will construct a mixing Markov chain (T'_k, P'_k) which will again have colored blocks which will then encode a partition $\overline{Q}'_k$. We will now construct the full group element ϕ_k and establish bounds to take forward for $\varepsilon_{k'}$ and $n_{k'}$, $k' > k$ that will guarantee:

i) The T_k converge in probability to some $S \overset{m^e}{\sim} T$.
ii) For each k, the partitions $\overline{Q}_k$ and $\overline{Q}'_k$ satisfy

$$\overline{d}(S, \overline{Q}_k; T'_k \overline{Q}'_k) < \varepsilon_k.$$

We need to discuss each of these conditions a bit. Condition i) is not simply that the rearrangements (T, ψ_k) are m^e Cauchy as this does not ensure that the reverse sequence (S, ψ_k^{-1}) is m^e Cauchy. But as the $S \in \Lambda(T)$ with $T \in \Lambda(S)$ are a residual subset, if we choose the ε_k small enough (depending on $\phi_1, \phi_2, \ldots, \phi_{k-1}$) then the limiting S will lie in this residual set. Thus obtaining i) is just a matter of setting upper bounds inductively on the ε_k.

Condition ii) tells us that any lumping of states H of a $\overline{Q}_k$ will still have

$$\overline{d}(S, H; T'_k, H') < \varepsilon_k$$

where H' is the corresponding lumping in the Markov process. Such partitions H are dense in the partition metric. Hence all partitions P of X have (S, P) in the $\overline{d}$ closure of Bernoulli systems and this tells us S must be Bernoulli. Now obtaining ii) is a matter of constructing the Markov chain $(T'_k, \overline{Q}'_k)$ which will be Bernoulli and hence satisfy the finitely determined condition. This then gives us a δ_k so that if we obtain

i) $\operatorname{dist}(S, \overline{Q}_k; T'_k, \overline{Q}'_k) < \delta_k$ and
ii) $h(S, \overline{Q}_k) > h(T'_k, \overline{Q}'_k) < \delta_k$ then we will have
iii) $\overline{d}(S, \overline{Q}_k; T'_k, \overline{Q}_k) < \varepsilon_k$.

Obtaining i) and ii) has two parts. First we need to get

i)$'$ $\mathrm{dist}(T_{k+1}, \overline{Q}_k; T'_k, \overline{Q}'_k) < \delta_k/2$ and
ii)$'$ $h(T_{k+1}, \overline{Q}_k) > h(T'_k, \overline{Q}'_k) < \delta_k/2$

explicitely in step k. With this in hand we can take forward inductively bounds on $\varepsilon_{k'}$ and $n_{k'}$ for $k' > k$ that will guarantee we lose at most $\delta_k/2$ in these inequalities as we move to the limit. We control i)$'$ with m^0 and ii)$'$ with $\mathcal{C}$.

We can now remove all the k's from the description and say we have an ergodic transformation T and partition Q with $h(T, Q) > 0$. We will assume in fact that Q is a generating partition which just means we will do our construction inside the σ algebra it generates under T. We are also given an upper bound for ε and a lower bound for n but can set further bounds in the process of the construction.

What we now describe is how to actually choose n and build our Rokhlin tower and then construct the model Markov process and set up and manipulate our Rokhlin towers much as we did in the proof of Dye's theorem in such a way as to apply our combinatorics in order to get conditions i) and ii) of the finitely determined condition. We will work with just one tower, labeled by colored blocks, and attempt to manipulate it into a "general position". Then considering the second tower on our Markov chain as also moved to this general position.

We set some notation. For any probability vector $P = \{p_i, \dots, p_t\}$ set $H(P) = -\sum_i p_i \log_2(p_i)$, the classical entropy function. For any partition $R = \{r_i\}$ set $h(R) = H(\{\mu(r_i)\})$.

To start we will assume some basic facts about the picture. Assume we are looking at the tower with base B of height n cut into T, Q, n-names constructed from the Strong Rokhlin lemma. A small corollary to the Strong Rokhlin Lemma will help. This uses much classical entropy theory.

Corollary 7.1. *For T ergodic, Q a finite partition and n and ε fixed one can find bases B_m for Strong Rokhlin towers of height n omitting precisely ε of the space so that if R_m is the two set partition into the tower and its compliment and P_m is Q on the tower itself and a single set outside the tower, we will have*

$$\lim_{m\to\infty} h(T, R_m) = 0 \text{ and}$$

$$\lim_{m\to\infty} h(T, P_m) = h(T, Q)(1 - \varepsilon).$$

Proof. From the classical theory $h(T, Q) = \lim_{n\to\infty} h(Q| \vee_{i=-1}^{-\infty} T^{-i}(Q))$ and we know $\mathrm{dist}(Q| \vee_{i=-1}^{-\infty} T^{-i}(Q))$ converges pointwise a.e., by the Martingale theorem. Now construct a sequence of Strong Rokhlin towers of heights mn $m \nearrow$, with error set of size ε. From each of these one can construct a tower of height n by cutting this one into m blocks of height n. But now as mn grows, the partition R_m into the tower and its compliment becomes ever more invariant and independant of the algebra generated by Q. Hence we get $h(T, R_m) \to 0$. Further, when m is large most x in the tower give us a good estimate for the conditional entropy of the Q present given the Q past. Points outside the tower are almost past measurable and hence give us a conditional entropy near zero.

□

We define some auxiliary partitions. Take P_n to be the partition that colors T, Q, n-name columns up the tower in distinct colors and for which the set outside the tower is uncolored, i.e. is a separate partition element u. Let $R_n = \{u, v\}$ be a two set partition into the tower and its complement. Let Δ_n be the normalized distribution of colored names on the base of the tower, i.e. a probability vector $\{\Delta_n(c)\}$ of conditional measures of the various colors c in P_n.

It is classical that

$$\lim_n \frac{1}{n} H(\Delta_n) \nearrow h(T, Q).$$

Moreover if $\frac{1}{n} H(\Delta_n) = h(T, Q)$ then (T, Q) is already i.i.d. and we need do nothing. Hence we will assume

$$\frac{1}{n} H(\Delta_n) - h(T, Q) = d > 0.$$

We will assume here that n is large enough that $d < \frac{\varepsilon h(T,Q)}{10 \log_2(\# Q)}$ for later purposes.

We now construct our model Markov measures. Suppose we create an alphabet whose elements are of the form (c, i), the various colors paired with an index i from 0 to $n-1$ and a last "uncolored" letter u. Now give a transition matrix on this that says $(c, i) \to (c, i+1)$ if $i < n-1$, $(c, n-1) \to (c', 0)$ and u, for all colors c' and $u \to \{u$ and all $(c, 0)\}$. This transition matrix gives a mixing subshift of finite type. We can consider the closed convex set of all shift invariant measures on this subshift. Consider the subset $\mathcal{M}$ of such measures for which u has measure ε and the colored sets (c, i) have measures $(1-\varepsilon)\Delta_n(c)/n$. This is a closed subset of measures and consists of all measures that model our tower picture for (T, P_n). We can create a compatible Markov matrix $M_{\max}$ by setting the probability of all allowed transitions into u to be ε and all transitions into $(c', 0)$ to be $(1-\varepsilon)\Delta_n(c')$. The corresponding mixing Markov chain $(T_n^{\max}, P_n^{\max})$ gives the unique measure of maximal entropy in $\mathcal{M}$. Hence

$$h(T, P_n) \leq h(T_n^{\max}, P_n^{\max}).$$

There is also a zero entropy Markov measure in $\mathcal{M}$ given by $(c, n-1) \to (c, 0)$ and $u \to u$ identically. This is not even ergodic, but suppose its transition matrix is M_0. Now consider the line of Markov matrices $\alpha M_{\max} + (1-\alpha) M_0$ for $0 < \alpha \leq 1$. These all give mixing Markov chains as there is a nontrivial $M_{\max}$ component and all give Markov measures in $\mathcal{M}$. As α varies, the entropy of these Markov chains varies continuously from 0 to $h(T_n^{\max}, P_n^{\max})$. Hence for some choice α the corresponding mixing Markov chain (T', P_n') will have

$$h(T', P_n') = h(T, P_n) = h.$$

This is our model. From it and the value ε we now obtain a value δ from the finitely determined condition, which it satisfies. We will assume $\delta < d$. We now need to construct ϕ and modify P_n by a small amount to P_n''. Our goal is to get $(\phi^{-1} T \phi, P_n'')$ within $\delta/2$ of (T', P_n') in both entropy and distribution.

Our work so far has already forced n to be large and ε small but as d and δ are determined solely by n we can still choose our tower so that

$$h(T, R_n) < \frac{\varepsilon \delta}{10}$$

and so that

$$h(T,P_n) = h(T,Q)(1-\varepsilon) \pm \frac{\varepsilon\delta}{10}.$$

We will construct ϕ on a much taller tower, of height n'. We select the size of n' to control a large number of error terms. Toward that end we will use the notation $e(n')$ to represent a generic error term that tends to zero in n'. So, for example, we can write $\sqrt{e(n')} = e(n')$.

We have four quanitities growing at distinct exponential rates in n'. We set conditions as follows on them:

1) The size of all but $e(n')$ of the T, R_n, n'-names is at most
$$2^{-(\frac{\delta\varepsilon}{10} - e(n'))n'}.$$
2) The size of all but $e(n')$ of the T, P_n, n'-names is of order
$$2^{-h(T,P_n)\pm e(n'))n'}.$$
3) All but $e(n')$ of the T, P_n, n'-names have colored blocks that occur with densities within $e(n')$ of Δ_n.
4) The size of all but $e(n')$ of the T, Q, n'-names is of order
$$2^{-(h(T,Q)\pm e(n'))n'}.$$
5) For $n_0 \geq \frac{\delta n'}{2}$ and divisible by n, the number of ways to color n_0/n blocks of length n with distribution exactly Δ_n is
$$N(n_0) = \binom{\frac{n_0}{n}}{\frac{k(c_1)n_0}{mn}, \ldots, \frac{k(c_t)n_0}{mn}} \quad \text{and is of the form} \quad 2^{(\frac{1}{n}H(\Delta_n)\pm e(n'))n_0}.$$

We have bounds that tell us the latter three of these grow at distinct exponential rates as

$$h(T,Q)(1-\varepsilon) \pm \frac{\varepsilon\delta}{10} = h(T,P_n)$$

and

$$h(T,Q) = \frac{1}{n}H(\Delta_n) - d.$$

We now carry out some "surgery" on a strong Rokhlin tower of height n' in a series of steps in preparation for constructing ϕ. Formally the value n' is not yet set as we will carry error terms $e(n')$. Setting bounds on these errors we will determine n'. Consider such a tower, based on a set B' and with our list of bounds above all given relative to the set B'. Statements 1), 2), 3), and 4) all have error sets in that each has a small set of names where the stated bounds fail to hold. Working on this tower of height n', for each separately we could delete this error set from the tower and have the stated bounds holding for all names up the tower. We need to accomplish this for all the conditions simultaneously.

Step I) Starting at 4) we delete all the Q names in the error set. Notice that for all but $\sqrt{e(n')}$ of either the P_n or R_n names have a fraction less than $\sqrt{e(n')}$ in this error set. Hence with a modified $e(n')$ we have maintained 1), 2) and 3). We now can delete the P_n names in the error sets of conditions 2) and 3) and as we saw modify $e(n')$ and maintain 1). As a P_n name is a union of Q names we do not lose 4). We now can move to 1) and delete the bad R_n names. As these are unions of P_n names we maintain the other conditions. In this step we have shaved $e(n')$ mass off the tower.

Step II) As a next step we want to eliminate partial colored blocks at the top and bottom of the names. We move these points into u. The new P_n names we create are lumpings of at most $(\#Q)^{2n} = 2^{e(n')n'}$ old P_n names and we keep all our bounds.

What we now have are new partitions $\overline{R}_n$, $\overline{P}_n$ and $\overline{Q}$ of the tower into names, each within $e(n')$ of the original partitions and so that:

1) The relative size on B' of all $T, \overline{R}_n, n'$ names is at most
$$2^{-(\frac{\delta\varepsilon}{10} - e(n'))n'}.$$
2) The relative size on B' of all $T, \overline{P}_n, n'$-names is of order
$$2^{-h(T,P_n) \pm e(n'))n'}.$$
3) All of the T, P_n, n' names on the tower have colored blocks that occur with densities within $e(n')$ of Δ_n.
4) The relative size on B' of all $T, \overline{Q}, n'$ names is of order
$$2^{-(h(T,Q) \pm e(n'))n'}.$$

Step III) As a first step in building ϕ we are going to push all the colored blocks to the bottom of the tower, maintaining order separately in both the colored blocks and the complimentary points in u. The number of permutations needed to accomplish this is precisely the number of $\overline{R}_n$ names in the tower which from 1) above is bounded by
$$2^{(\frac{\delta\varepsilon}{10} + e(n'))n'}.$$

Suppose now η is a colored name obtained by this "push down". Let the colored blocks in η occupy the bottom at most $g(\eta) = (1 - \varepsilon + e(n'))n'$ levels of the tower. We choose a value $\overline{e} = \frac{\delta\varepsilon}{5}(\frac{1}{n}H(\Delta_n))^{-1} + e(n')$ so that the height
$$n'' = \left((1-\varepsilon)\left(1 - \frac{dn}{H(\Delta_n)}\right) + \overline{e}\right) n'$$
is a multiple of nm. We know
$$N(n'') = 2^{(\frac{1}{n}H(\Delta_n \pm e(n'))n''} = 2^{(h(T,P_n) \pm \frac{\delta\varepsilon}{10} \pm e(n') + \frac{\delta\varepsilon}{5})n'}$$
and so is on the order of $\delta\varepsilon/10$ larger exponentially than the reciprocal of the size of T, P_n-names. As we know $d > \delta$, if n' is large enough the value n'' will be strictly below $g(\eta)$. On the colored blocks up to level n'' we know the densities of each color is within a fraction $e(n')$ of Δ_n.

Step IV) Hence we can now modify the pushed down colored names below n'' on a fraction of their length of order $e(n')$ so that we see these n''/n colored blocks with colors distributed precisely as Δ_n. In doing this we may lump together some colored names which previously were different but now agree. The number of colored names in any lump though is of the form $2^{e(n')n'}$.

Each pushed-down colored name after all our changes is a union of at most $2^{(\frac{\delta\varepsilon}{10} + e(n'))n'}$ distinct T, P_n-names. We want to cut these T, P_n-names into subsets each of size exactly

$\mu(B')/N(n'')$. These names might not be multiples of this size but from our bounds we calculate that each T, P_n name has a size at least

$$2^{(\frac{\delta\varepsilon}{5}\pm e(n'))}(N(n''))^{-1},$$

which is larger by an exponential factor than $(N(n''))^{-1}$.

Step V) Hence by modifying the T, P_n-names on less than some $e(n')$ (decaying exponentially in n') of the tower we ensure all P_n names can be cut into pieces of size $\mu(B')/N(n'')$. We make this modification and produce this partition S of the base into small sets of uniform size $\mu(B')/N(n'')$.

We are now ready to use our combinatorics as we have a tower of height n'' colored by blocks of length n all with precisely the same distribution Δ_n and each name occupying a fraction of B' a multiple of $1/N(n'')$ in size. This then gives us a list L_2 of $N(n'')$ colors, with multiplicities. In addition each P_n name and hence colored name is partitioned by S into pieces of size $\mu(B')/N(n'')$. Our goal is to assign permutations in $S(n''/n)$ to each of the small sets in S so that using them to permute the n''/n colored blocks above each set each small set will now have a distinct coloring i.e. the list L_1 of all colorings. We want to accomplish this with an exponentially small, in n', set of permutations.

Our combinatorial work tells us how many permutations we will need to accomplish this. We know we can assign permutations to each of the sets in S so as to rearrange the colored blocks above that set in such a way that all the modified colorings are distinct and hence give all possible arrangements of the coloring. By Theorem 5.1 we can accomplish this so that all but a fraction $(\varepsilon/5)$ of the base B' is covered by sets that use at most $\frac{K}{(\varepsilon/5)^2}$ permutations where K is the maximum number of small subsets of S in any colored name up to height n'' (after the push-down of colors and all modifications).

We need to estimate K. To begin, for η a pushed down name,

$$\max(g(\eta)) - n'' \le \frac{n}{H(\Delta_n)}\left(d - \varepsilon d - \frac{\delta\varepsilon}{5} + e(n')\right)n' \le \left(\frac{d + e(n')}{h(T,Q)}\right)n'.$$

If n' is large enough, having chosen d small enough earlier, this will be $\le \dfrac{\varepsilon}{5\log_2(\#Q)}n'$. This means the maximum number of distinct Q names, and hence P_n names that occur between heights n'' and $\max(g(\eta))$ is bounded by

$$(\#Q)^{(\max(g(\eta)-n'')} \le 2^{\frac{\varepsilon}{5}n'}.$$

We have already seen that a colored name η will contain at most $2^{(\frac{3\delta\varepsilon}{10}+e(n'))n'}$ elements of S and now we have seen that each name up to height n'' contains at most $2^{\frac{\varepsilon}{5}n'}$ colored names η. Hence if n' is large enough

$$K \le 2^{\frac{\varepsilon}{4}n'}$$

and hence the number of permutations we need on all but $\varepsilon/5$ of the base is $2^{\frac{\varepsilon}{4}n'}/(\varepsilon/5)^2$. Now if n' is large enough this is $\le 2^{\frac{\varepsilon}{3}n'}$.

Step VI) Rather than permuting the names on the remaining $\varepsilon/5$ of the tower, we use no permutation and rather modify the colorings to be the names we desire. Hence combining this with the permutations used to push down the colors we can now see all possible colored names of length n'', a unique one on each set in S, and we accomplished this using at most $2^{\frac{\varepsilon}{2}n'}$ permutations, and modifying the partition P_n on a subset of size $\leq \dfrac{\varepsilon}{5} + e(n')$.

This completes one half of the picture, having modified our original T, P_n, n'-names to a general position that consists of all possible lists of n''/n colored blocks placed at the bottom of the tower, one in each element of the partition S. For $s \in S$ let the permutation we apply to the name above s be $\pi(s)$. This is one of at most $2^{\frac{\varepsilon}{3}n'}$ permutations

We can apply precisely the same series of six steps to our Markov model (T', P'_n) and modify it to this same general position. In doing this we take the values n' and n'' used in the two modifications to be the same and the total amount of mass shaved off the two towers in the course of the constructions to be the same. Let S' be the partition in the Markov model into small sets of size $\mu(B')/N(n'')$ and let $\pi'(s')$ be the permutation applied to the name above s'. Again this is one of at most $2^{\frac{\varepsilon}{3}n'}$ permutations.

Each element of S and each element of S' after our rearranging sees a distinct coloring in their bottom n'' positions and these are all ways of reordering the n''/n colored blocks. Hence we can pair up the elements of S and S' by a bijection Ψ that matches sets with the same coloring up to level n''. The names above s and $\Psi(s)$ were created by a series of steps. We have "done" those steps to the levels above each set s. We now want to "undo" the steps constructed above $\Psi(s)$ but to the name above s. Let η be the original T, P_n-name that contains s and η' be the T', P'_n-name that contains $\Psi(s)$.

For each element $s \in S$ we proceed to accomplish the "undoing". First we replace the name above s after the rearranging with the name above $\Psi(s')$ after its rearranging. This only changes the name in levels above n'' by replacing the name above s from levels n'' to $g(\eta)$ by the name of $\Psi(s)$ from levels n'' to $g'(\eta')$. This modifies the partition on a set of size at most $2(\max_\eta g(\eta) - n'')/n' \leq 2\varepsilon/5$ from our earlier estimates. Next we take the permutation $\pi'(s')$ on levels up to n' used to create the colored name above $\Psi(s)$ and we appy its inverse to the name above s. On the $\varepsilon/5$ of the tower where we changed names in step VI) we undo the change on the corresponding names. This now creates precisely the name we had above $\Psi(s)$ for all s just before we applied our combinatorics. We will ignore the small change we made in step V) to make the T', P'_n, n'-names be precisely divisible into sets of size $\mu(B')/N(n")$. Next we modify an $e(n')$ fraction of the colors above s to precisely match the pushed down colors above $\Psi(s)$ by moving away from precise distribution Δ_n undoing step IV). This is a change of order $e(n')$. Next apply the inverse of the "push down" of step III) to the name to spread out the colored blocks. We now have a name above s that agrees with η' except for possible partial blocks at the top and bottom. We ignore this minor change from step II) as well as the material shaved off in step I).

Starting from our original partition P_n then we have moved to a new partition P''_n that is within ε of P_n. In addition, to each element $s \in S$ we have obtain a permutation $\overline{\pi}(s) = \pi'^{-1}(\Psi(s))\pi(s)$ and if we permute the levels above s in the tower we obtain a full group element ϕ. For all but $e(n')$ of the points in the tower over B' that are colored, ϕ has moved blocks of colored points by a rigid translation. Hence $m^0(T, \phi) < \varepsilon + 1/n + e(n')$ is small. In addition, the number of permutations we have used on the tower to form ϕ is at most $2^{\frac{2\varepsilon}{3}n'}$ and hence $\mathcal{C}(T, \phi) \leq 2\varepsilon/3$.

We conclude that if $n > 1/\varepsilon$ and n' is large enough then $m^e(T,\phi) < 3\varepsilon$. We want to see that $(\phi^{-1}T\phi, P''_n)$ and (T', P'_n) are within δ in distribution and entropy. That we are within δ in distribution follows easily from the fact that our tower is now all but $e(n')$ covered by a copy of the distribution of T'_n, P'_n, n'-names and we can choose n' as large as we like. This is precisely what Dye had shown us how to do. Closeness in entropy is the hard part. Notice though that all of the $\phi^{-1}T\phi, P''_n$-names up the tower contain at most $2^{\frac{\delta\varepsilon}{5}+e(n'))n'}$ elements of S and these elements partition the T, P_n-names up the tower. Partition the tower vertically by an auxiliary partition V into at most $2^{\frac{\delta\varepsilon}{5}+e(n'))n'}$ sets so that $P''_n \vee V$ under the action of $\phi^{-1}T\phi$ determines the element of S containing the set and hence generates the T, P_n-name up the tower.

Now as V is a partition into vertical slices up the tower of height n',

$$h(\phi^{-1}T\phi, V) < \frac{\delta\varepsilon}{5} + e(n').$$

Furthermore, if we make the set omitted from the tower of height n' decay as $e(n')$ then

$$h(\phi^{-1}T\phi, P''_n \vee V) \geq h(T, P_n) - e(n')$$

as we generate the T, P_n names whenever we are inside the tower. Now if n' is large enough we have

$$h(\phi^{-1}T\phi, P''_n) \geq h(T, P_n) - \frac{\delta\varepsilon}{4},$$

more than enough for our purposes.

To remind the reader of what we have accomplished, we now know that for any ergodic process (T, P) and $\varepsilon > 0$ we can find a full group element ϕ and a Bernoulli process (T', P') with

i) $m^e(T,\phi) < \varepsilon$ and
ii) $\overline{d}(\phi^{-1}T\phi, P; T', P') < \varepsilon$.

This forces any m^e-equivalence class with positive entropy to contain a Bernoulli action. Hence any two ergodic actions of equal entropy are m^e-equivalent.

References

1. M. Brin and G. Stuck, *Introduction to dynamical systems*, Cambridge Univ. Press, 2002.
2. H. A. Dye, *On groups of measure preserving transformations i*, Amer. J. Math. **81** (1959), 119–159.
3. B. Hasselblatt and A. Katok, *Introduction to the modern theory of dynamical systems*, Encyclopedia of Mathematics and its applications, vol. 54.
4. J. W. Kammeyer and D. J. Rudolph, *Restricted orbit equivalence for actions of discrete amenable groups*, Cambridge University Press, 2001.
5. D. S. Ornstein, *Two bernoulli shifts with infinite entropy are isomorphic*, Adv. in Math. **5** (1970), 339–348.
6. D. S. Ornstein and P. C. Shields, *Mixing Markov shifts of kernel type are Bernoulli*, Advances in Math. **10** (1973), 143–146.
7. Donald S. Ornstein and Benjamin Weiss, *Finitely determined implies very weak Bernoulli*, Israel J. Math. **17** (1974), 94–104.
8. D.S. Ornstein, *Ergodic theory, randomness and dynamical systems*, Yale University Press, 1974.
9. K. Petersen, *Ergodic theory*, Cambridge University Press.
10. D. Rudolph, *Fundamentals of measurable dynamics*, Oxford University Press, 1990.

11. D. J. Rudolph, *An entropy preserving dye's theorem for ergodic actions*, Journal d'Analyse Math.
12. P. Walters, *An introduction to ergodic theory*, Springer-Verlag, 1982.