Commun. math. Phys. 29, 311—317 (1973)

Mixing Transformations on Metric Spaces

T. Erber
Department of Physics, Illinois Institute of Technology, Chicago, Illinois, USA

B. Schweizer
Department of Mathematics and Statistics, University of Massachusetts, Amherst, Massachusetts, USA

A. Sklar
Department of Mathematics, Illinois Institute of Technology, Chicago, Illinois, USA

Received September 5, 1972

Abstract. It is proved that if a metric space is subjected to a mixing transformation, then there exists a positive number x_0 such that the probability that any arbitrary set of positive measure is asymptotically mapped into a set of diameter less than x_0 is zero. Physical implications of this result, in particular the interpretation of Poincaré recurrence, are discussed.

In this note we consider a metric space (Ω, d) usually assumed separable, on which a probability measure P is defined. The domain of P is assumed to include all Borel sets in Ω; in particular, therefore, all spherical neighborhoods (open balls) are P-measurable. We call P *non-singular* if there exist two open balls Q_1, Q_2, and a positive number x_0 such that $P(Q_1) > 0$, $P(Q_2) > 0$, and $d(w_1, w_2) > x_0$ for all w_1 in Q_1, w_2 in Q_2. We call P *pervasive* if $P(Q) > 0$ for all open balls Q in Ω. If Ω has more than one point, then clearly any pervasive measure is non-singular.

As usual, a transformation T of Ω onto itself is called (strongly) *mixing* if the inverse image $T^{-1}A$ of any P-measurable set A is P-measurable, and

$$\lim_{n \to \infty} P(T^{-n} A \cap B) = P(A) P(B) \tag{1}$$

for any two P-measurable subsets A, B in Ω. The *diameter* of a subset S of Ω, i.e., the supremum of the set of all distances $d(w_1, w_2)$ for w_1, w_2 both in S, will be denoted by $\operatorname{diam}(S)$. If S is P-measurable, then the *essential diameter* of S, denoted by $\operatorname{ess\,diam}(S)$, is the infimum of the diameters of all P-measurable sets S' such that $S' \subseteqq S$ and $P(S') = P(S)$. Both $\operatorname{diam}(S)$ and $\operatorname{ess\,diam}(S)$ may be infinite.

Our principal aim is to prove the following:

Theorem 1. *Let T be mixing with respect to a non-singular probability measure P on a separable metric space (Ω, d). For any subset S of Ω and any $x > 0$, let $N(S, T, x)$ denote the set of positive integers m such that* $\operatorname{diam}(T^m S) < x$, *and let $\chi_{N(S,T,x)}$ denote the characteristic function of $N(S, T, x)$. Then there is a positive number x_0 such that*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \chi_{N(S,T,x_0)}(\mathrm{m}) = 0 \tag{2}$$

for any P-measurable set S with $P(S) > 0$. If P is pervasive then x_0 in (2) *can be taken to be any number less than* $\operatorname{ess\,diam}(\Omega)$. *Furthermore, if TA is measurable for any measurable set A, and if* $\operatorname{ess} N(T, S, x)$ *denotes the set of positive integers m such that* $\operatorname{ess\,diam}(T^m S) < x$, *then* (2) *holds with $N(S, T, x_0)$ replaced by* $\operatorname{ess} N(S, T, x_0)$.

Loosely speaking, Theorem 1 says that on the average a mixing transformation never (well, hardly ever!) maps a set of positive measure into a set of small (essential) diameter. Such a result can have considerable physical significance: there is a brief discussion of this at the end of the note, immediately upon the conclusion of the proof of Theorem 1. The proof itself is based on, and therefore preceded by, Lemmas 1 and 2 and Theorem 3.

We begin the formal discussions with the consideration of a special case. If T, in addition to being mixing, is *invertible* (one-one and such that the image of any P-measurable set is P-measurable) then Theorem 1 is a corollary of the stronger result established in the following:

Theorem 2. *Let T be invertible and mixing with respect to a non-singular probability measure P on a (not necessarily separable) metric space (Ω, d). Then there is a positive number x_0 such that for any P-measurable set S with $P(S) > 0$ there exists a positive integer $n(S)$ such that:*

$$\operatorname{ess\,diam}(T^m S) > x_0 \quad \textit{for all} \quad m \geqq n(S). \tag{3}$$

If P is pervasive then x_0 can be taken to be any number less than $\operatorname{ess\,diam}(\Omega)$.

Proof. If T is invertible, then (1), as is well-known (and readily verified) is equivalent to:

$$\lim_{n\to\infty} P(T^n A \cap B) = P(A)\, P(B) \tag{4}$$

for any P-measurable sets A, B. If P is non-singular there exist open balls Q_1, Q_2 and an $x_0 > 0$ such that $P(Q_1) > 0$, $P(Q_2) > 0$, and $d(w_1, w_2) > x_0$ for all w_1 in Q_1, w_2 in Q_2. Clearly, if P is pervasive, then x_0 can be any number less than $\operatorname{ess\,diam}(\Omega)$.

Now let S be any P-measurable set with $P(S) > 0$. By virtue of (4), there is an integer $n(S)$ such that both $P(T^m S \cap Q_1) > 0$ and $P(T^m S \cap Q_2) > 0$ for $m \geqq n(S)$. Since $d(w'_m, w''_m) > x_0$ for any w'_m in $T^m S \cap Q_1$ and w''_m in $T^m S \cap Q_2$, it follows that ess diam$(T^m S) > x_0$. Hence (3) holds and the proof of Theorem 2 is complete.

In Lemma 1, and subsequently, we shall use the standard facts concerning product measures as found, e.g., in [5], § 6.2. We shall also need the notion of a *product transformation:* if T_1 transforms Ω_1 into itself, and T_2 transforms Ω_2 into itself, then their product is the transformation $T_1 \times T_2$ defined on $\Omega_1 \times \Omega_2$ by $(T_1 \times T_2)(w_1, w_2) = (T_1 w_1, T_2 w_2)$ for all pairs (w_1, w_2) in $\Omega_1 \times \Omega_2$.

Lemma 1. *Let T_1 be mixing with respect to a probability measure P_1 on a set Ω_1, and T_2 be mixing with respect to a probability measure P_2 on a set Ω_2. Then the product transformation $T_1 \times T_2$ is mixing with respect to the product measure of P_1 and P_2 on $\Omega_1 \times \Omega_2$.*

Proof. A *Cartesian rectangle* in $\Omega_1 \times \Omega_2$ is a set of the form $S_1 \times S_2$, where S_1 is a P_1-measurable subset of Ω_1 and S_2 a P_2-measurable subset of Ω_2. By direct computation, one verifies that (1) holds, first of all when A and B are each Cartesian rectangles, and then when A and B are each a union of two disjoint Cartesian rectangles. Since the intersection of two Cartesian rectangles is a Cartesian rectangle, it follows that (1) holds whenever A and B are in the field generated by the Cartesian rectangles in $\Omega_1 \times \Omega_2$. A standard result (see e.g., [1], Theorem 1.2) now permits us to extend (1) to arbitrary A and B in the σ-field generated by the Cartesian rectangles in $\Omega_1 \times \Omega_2$; hence $T_1 \times T_2$ is mixing with respect to the product measure of P_1 and P_2.

For any positive integer n, we denote the n-fold product $T \times T \times \cdots \times T$ of a transformation T with itself by $T^{(n)}$, and the n-fold product of a probability measure P with itself by $P^{(n)}$. An induction then yields the following:

Corollary. *If T is mixing with respect to P on Ω, then $T^{(n)}$ is mixing with respect to $P^{(n)}$ on $\Omega^n = \Omega \times \cdots \times \Omega$ for each positive integer n.*

Lemma 2. *Let P be a probability measure on a separable metric space (Ω, d). For any real number x, let D_x denote the set of all pairs (w_1, w_2) of points in Ω such that $d(w_1, w_2) < x$. Then D_x is $P^{(2)}$-measurable for each x, and the real function F defined by:*

$$F(x) = P^{(2)}(D_x) \tag{5}$$

is a distribution function such that $F(x) = 0$ for $x \leqq 0$. If P is non-singular, then there is a positive number x_0 such that $F(x_0) < 1$; if P is pervasive, then $F(x) < 1$ for all $x <$ ess diam(Ω).

Proof. If $x \leqq 0$, then D_x is empty, therefore automatically $P^{(2)}$-measurable with $P^{(2)}(D_x) = F(x) = 0$. If $x > 0$, then it readily follows from the separability of (Ω, d) that D_x may be expressed as a countable union of open Cartesian rectangles of the form $Q \times R$, where Q is an open ball of small diameter (compared to x) and R is the set of all points w such that $d(w, w') < x$ for all points w' in Q. Hence D_x is $P^{(2)}$-measurable and $F(x)$ is well-defined for all x. That F is a distribution function follows from the standard properties of the probability measure P. If P is non-singular, we take x_0 as in the proof of Theorem 2. We then have:

$$F(x_0) \leqq 1 - P^{(2)}(Q_1 \times Q_2) = 1 - P(Q_1)\, P(Q_2) < 1\,. \tag{6}$$

If P is pervasive, then, as before, we note that x_0 in (6) can be taken to be any number less than ess diam(Ω), and this observation completes the proof.

If the metric space (Ω, d) is not separable, then the product measure $P^{(2)}$ may not be defined for all Borel sets in $\Omega \times \Omega$. Hence the sets D_x may not be $P^{(2)}$-measurable even though the distance function d remains continuous [1].

Theorem 3. *Let T be mixing with respect to a probability measure P on a separable metric space (Ω, d). For a fixed positive integer k, an arbitrary $2k$-tuple $\mathbf{w} = (w_1, w_2, \ldots, w_{2k-1}, w_{2k})$ of points in Ω, and an arbitrary $x > 0$, let $M(k, \mathbf{w}, T, x)$ denote the set of positive integers m such that*

$$d(T^m w_1, T^m w_2) < x, \ldots, d(T^m w_{2k-1}, T^m w_{2k}) < x\,.$$

Let $\chi_{M(k,\mathbf{w},T,x)}$ be the characteristic function of $M(k, \mathbf{w}, T, x)$ and F the distribution function defined in Lemma 2. Then there is a set W_k in Ω^{2k}, with $P^{(2k)}(W_k) = 1$, such that for all $\mathbf{w}$ in W_k we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \chi_{M(k,\mathbf{w},T,x)}(m) = (F(x))^k \quad \textit{for all} \quad x > 0\,. \tag{7}$$

Proof. We first note that

$$\chi_{M(k,\mathbf{w},T,x)}(m) = \chi_{(D_x)^k}[(T^{(2k)})^m(\mathbf{w})]\,,$$

where $D_x = \{(u, v) \mid d(u, v) < x\}$. Now, by the Corollary to Lemma 1, $T^{(2k)}$ is mixing on Ω^{2k} with respect to $P^{(2k)}$. Hence (see [1], pp. 12–15), $T^{(2k)}$ is ergodic on Ω^{2k} and it follows from the Birkhoff ergodic theorem that for any $x > 0$ there is a set $W_k(x)$ in Ω^{2k}, with $P^{(2k)}(W_k(x)) = 1$, such that for all $\mathbf{w}$ in $W_k(x)$ the limit in (7) exists and is equal to $P^{(2k)}(D_x{}^k)$. But by the definition of product measure and (5),

$$P^{(2k)}(D_x{}^k) = [P^{(2)}(D_x)]^k = (F(x))^k\,.$$

[1] For a discussion of these points, see pp. 224–225 in P. Billingsley; Convergence of Probability Measures, Wiley, 1968.

It remains to eliminate the dependence on x of the set $W_k(x)$. To this end let $\{x_i\}$ be a countable dense subset of the real line, containing any discontinuity points of F; let $\{W_k(x_i)\}$ be the corresponding sets of measure one on which the limit in (7) exists; and let

$$W_k = \bigcap_{i=1}^{\infty} W_k(x_i).$$

Then $P^{(2k)}(W_k)=1$ and the limit in (7) exists for all x in $\{x_i\}$ and all $\boldsymbol{w}$ in W_k. But since the distribution function F is completely determined by its behavior on $\{x_i\}$ the limit exists for all x. This completes the proof of Theorem 3.

We can now prove Theorem 1. Let S be any P-measurable subset of Ω with $P(S)>0$. Then for any positive integer k, S^{2k} is $P^{(2k)}$-measurable and $P^{(2k)}(S^{2k})>0$. Let W_k be as in Theorem 3: since $P^{(2k)}(W_k)=1$, it follows that $P^{(2k)}(S^{2k}\cap W_k)=P^{(2k)}(S^{2k})>0$, whence $S^{2k}\cap W_k$ is not empty. Let $\boldsymbol{w}$ be a $2k$-tuple in $S^{2k}\cap W_k$: from the definitions of the sets $N(S,T,x)$ (in Theorem 1) and $M(k,\boldsymbol{w},T,x)$ (in Theorem 3) it is clear that $N(S,T,x)$ is a subset of $M(k,\boldsymbol{w},T,x)$ for any given $x>0$. Hence

$$\chi_{N(S,T,x)}(m) \leqq \chi_{M(k,\boldsymbol{w},T,x)}(m) \quad \text{for all} \quad m \geqq 1,$$

whence (7) yields:

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \chi_{N(S,T,x)}(m) \leqq (F(x))^k \quad \text{for all} \quad x>0, \tag{8}$$

and for any positive integer k. If x is such that $F(x)<1$, then the right-hand side of (8) can be made arbitrarily small by taking k sufficiently large. In particular, by Lemma 2 there is such an $x>0$ if P is non-singular, while any $x<\operatorname{ess\,diam}(\Omega)$ will do if P is pervasive; in both cases it follows that the left-hand side of (8) is 0, which yields (2). Finally, if TE is measurable for any measurable set E, then $T^m S$ is measurable for any positive integer m. Let m be in $\operatorname{ess} N(S,T,x_0)$. Then there is a subset V_m of $T^m S$ such that $P(V_m)=P(T^m S)$ and $\operatorname{diam} V_m < x_0$. Let $Z_m = T^m S \setminus V_m$. Since T is measure-preserving, $P(T^{-m} Z_m)=P(Z_m)=0$. Consequently, since $T^{-m}Z_m$ and $T^{-m}V_m$ are disjoint and $S \subseteqq T^{-m}T^m S = T^{-m}(V_m \cup Z_m) = T^{-m}V_m \cup T^{-m}Z_m$, we have:

$$\begin{aligned} P(S) &= P(S\cap T^{-m}T^m S) = P((S\cap T^{-m}V_m)\cup(S\cap T^{-m}Z_m)) \\ &= P(S\cap T^{-m}V_m) + P(S\cap T^{-m}Z_m) = P(S\cap T^{-m}V_m). \end{aligned}$$

Now let S_0 be the intersection of the sets $S\cap T^{-m}V_m$ for all m in $\operatorname{ess} N(S,T,x_0)$. Then $P(S_0)=P(S)>0$ and $\operatorname{diam}\, T^m S_0 < x_0$ for any m in $\operatorname{ess} N(S,T,x_0)$. Thus, as above, for any $\boldsymbol{w}$ in ${S_0}^{2k}\cap W_k$, the set $\operatorname{ess} N(S,T,x_0)$ is contained in $M(k,\boldsymbol{w},T,x_0)$, whence the rest of the argument concludes as before. This completes the proof of Theorem 1.

Mixing transformations were originally introduced [3] as mathematical models of the physical processes which occur when solutes are homogeneously dispersed in solvents by prolonged stirring. Up to now, most informal discussions of mixing transformations have focused on the homogenizing effects; Theorems 1 and 2 now make the equally important dispersive effects explicit. One approach to a treatment of such dispersive effects is furnished by Theorem 3, which admits the following interpretation: for almost all pairs of points, the distance between the images of the points under a mixing transformation behaves like a random variable whose distribution is given by the function F of Lemma 2. Furthermore, for almost all finite sets of pairs of points, the corresponding random variables are independent and identically distributed. Thus mixing transformations on metric spaces give rise to a new class of probabilistic metric (PM-) spaces (for a survey of the theory of PM-spaces, see [6]); these may, by analogy with the equilateral PM-spaces (see [7]), be called *almost equilateral.*

In [2], the mean distance, i.e., the integral $\int_0^\infty x\,dF(x)$, was introduced as a useful characterization of the dispersive effects of a mixing transformation. Such mean distances will not distinguish between two mixing transformations on the same space, since F depends only on the metric d and the probability measure P: nevertheless, two such transformations will generally differ in the rapidity and smoothness of the mixing process, and these differences can be gauged by the rate of approach to, and fluctuations about, the mean distance for particular sets of points. Numerical calculations of these quantities have been carried out in particular cases, and may be found in [2] and [4].

If a mixing transformation T is invertible, then its inverse T^{-1} is also mixing, and Theorem 2 applies to it. Thus we encounter dispersion if we follow a set S of small diameter and positive measure into the "distant past" as well as the "distant future". Another way of putting it is that S is the result of the coalescence of an originally completely spread-out set: this coalescence can last only a finite time and is then inevitably followed by a further, and final, dispersal. On this basis, biological, and even cosmological speculations immediately suggest themselves ... .

Perhaps the most significant implications of Theorems 1 and 2 arise in connection with a question that goes back to the very origins of ergodic theory: the famous controversy between Boltzmann and Zermelo about the relevance of the Poincaré (or Poincaré-Carathéodory) recurrence theorem (for a brief discussion and references to the original papers, see [2]). Poincaré had shown that almost all points in a space subject to a measure-preserving transformation return over and over again to positions arbitrarily close to their original position. Since states of physical systems are often represented as points in appropriate phase spaces, and the transformations these spaces undergo are measure-preserving (Liouville's theorem), the recurrence theorem holds, with an effect that seems to contradict the observed manifestations of fundamental thermodynamic principles. But, ideally, to represent a state of a physical system as a *point*, the state must be known *exactly*. Since this is effectively never the case, an obviously more realistic model, which reflects the imprecision of our measurements, is that the initial configuration of a system does not comprise a single state but rather an assembly of experimentally indistinguishable states which correspond to a set, of possibly very small, but still positive, diameter, and possibly very small, but still positive, measure. If now the physical transformation involved can be represented by a *mixing* transformation of the phase space (a far-reaching but not altogether unnatural assumption), then Theorem 1 or Theorem 2 will apply and yield the fact that there will generally be *no* recurrence: the initial set will, after a while, spread out in diameter and, with probability 1, *stay* spread out. A detailed discussion of the relation of this principle of "Stable Recurrence" to "Poincaré Recurrence" and "Exact Recurrence" is given in [2]. In this sense, Theorems 1 and 2 constitute a partial, if belated, vindication of the physical intuition of Ludwig Boltzmann.

References

1. Billingsley, P.: Ergodic theory and information. New York: Wiley 1965.
2. Erber, T., Sklar, A.: Macroscopic irreversibility as a manifestation of micro-instabilities. Modern Developments in Thermodynamics, ed. by B. Gal-Or. Jerusalem-New York: Israel Universities Press and J. Wiley and Sons 1973.
3. Hopf, E.: On causalty, statistics, and probability. J. Math. Phys. **13**, 51—102 (1934).
4. Johnson, P., Sklar, A.: Iterative, ergodic, and mixing properties of Čebyšev polynomials. To appear.
5. Kingman, J. F. C., Taylor, S. J.: Introduction to measure and probability. Cambridge: Cambridge University Press 1966.
6. Schweizer, B.: Probabilistic metric spaces — the first 25 years. New York Statistician **19**, 3—6 (1967).
7. Schweizer, B., Sklar, A.: Statistical metric spaces. Pacific J. Math. **19**, 313—334 (1960).

T. Erber
Department of Physics
Illinois Institute of Technology
Chicago, Illinois 60616, USA

B. Schweizer
Department of Mathematics and Statistics
University of Massachusetts
Amherst, Massachusetts 01003, USA

A. Sklar
Department of Mathematics
Illinois Institute of Technology
Chicago, Illinois 60616, USA