0
Introduction

1 Origins

Statistical mechanics as developed by Gibbs [1], grew out of an attempt to apply systematically probability theory (not yet itself systematised) to conservative mechanical systems with many degrees of freedom. Earlier investigations along these lines by Maxwell and Boltzmann were based on highly specialised assumptions concerning the interaction of particles for the purpose of explicating thermodynamics. Gibbs sought a theory which made as few assumptions as possible concerning the 'nature' of particles yet which, at least by analogy and perhaps by direct application, embraced thermodynamical problems. His theory rested on Hamiltonian formalism and thus included an invariant (Liouville) distribution of phase.

If a system of k particles moves in a region of three-dimensional space, at each point of time the system is described by the position $q^i = (q^i_1, q^i_2, q^i_3)$ and momentum $p^i = (p^i_1, p^i_2, p^i_3)$ coordinates for $i = 1, 2, \ldots, k$. These data are summarised by a point x in some region X of $6k$-dimensional space, the phase space. Assuming the system to be conservative, Liouville's theorem asserts the existence of a volume m on X 'smoothly' related to the differential structure of X, which is invariant through the course of time, i.e. a subregion will evolve through successive subregions of the same volume. If x_t $(t \in \mathbb{R})$ represents the time evolution of a point in phase space, then T_s: $x_t \to x_{t+s}$ $(s \in \mathbb{R})$ defines a one-parameter group of transformations $(T_t \circ T_s = T_{t+s}, T_0 = \text{identity})$ which preserves the measure m.

If F is a sufficiently smooth real-valued function which is invariant with respect to the motion defined by $\{T_t\}$ $(F \circ T_t = F$ for all $t \in \mathbb{R})$, then the study of the dynamical system on X reduces to the study of the dynamical system on the invariant subspaces $F_a = \{x : F(x) = a\}$ and each F_a possesses an invariant volume m_a canonically related to m. For a Hamiltonian system the Hamiltonian function H (or total

energy) is such a function or *integral* and, in general, m_a is a finite invariant measure.

Modern ergodic theory grew out of the so-called ergodic problem or hypothesis, viz. phase averages and time averages coincide on each surface of constant energy:

$$\frac{1}{m_a(H_a)} \int_{H_a} f \, \mathrm{d}m_a = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} f \circ T_t \, \mathrm{d}t. \tag{0.1}$$

Boltzmann thought that each surface of constant energy might be entirely filled by a single trajectory – a hypothesis which would ensure (0.1), but which is rarely satisfied. The quasi-ergodic hypothesis (of P. and T. Ehrenfest) that each trajectory is dense in an energy surface is not enough to ensure (0.1) and in any case excludes numerous important examples.

We now know that (0.1) holds for almost all $x \in H_a$ (or in the $L^2(H_a)$ sense of convergence) if and only if $\{T_t : t \in \mathbb{R}\}$ is *ergodic* on (H_a, m_a) (a term to be defined later). These statements, which amount to Birkhoff's [1] and von Neumann's [1] ergodic theorems respectively, ushered into mathematics the subject of ergodic theory proper. Both theorems were initially formulated for conservative Hamiltonian systems: in other words in the setting of ordinary differential equations, and involved difficulties which have subsequently been removed. von Neumann's theorem was based on Koopman's observation that these mechanical systems, preserving as they do a natural measure, induce on the Hilbert space of square integrable functions a one-parameter group of unitary operators. Thus, von Neumann was able to exploit the spectral theory in the development of which he had played such a major role. Subsequently Hopf provided a simplified proof which avoided spectral theory.

Even though von Neumann's theorem is intrinsically on a more elementary level than Birkhoff's, it was all that was needed (given ergodicity) for the original purposes of statistical mechanics as von Neumann argued and as Birkhoff (and Koopman) partially conceded: 'In view of these facts, it is of interest to decide which of the two formulations, (1) or (2) [mean convergence or a.e. convergence] corresponds to the actual physical problem of the ergodic hypothesis. It turns out that (1) is sufficient – that it, indeed, is the precise

mathematical equivalent of the physical state of affairs.' (von Neumann [2], p. 275.)

'With regard to the scope of this theorem, we may make the following remarks:

1. From the point of view of the gross statistics on Ω (classical kinetic theory), it is equivalent in its implications to the Mean Ergodic Theorem.

2. From the viewpoint of the detailed statistics along an individual path curve, it is fundamentally more far-reaching; in it is proved for the first time that the relative time of sojourn along almost every individual path curve *exists*, a result often assumed implicitly in the writing of physicists, but never proved.' (Birkhoff and Koopman [1], p. 281.)

The above reference to time averages on (almost all) trajectories is not without force. Birkhoff's theorem allows us to speak of these averages, in connection with the duration spent by a point in a particular region, and we may conclude, almost surely (when the system is ergodic), that this average time is proportional to the volume of the region. In other words, not only do we know that almost all points of a region return infinitely often to that region, but we can state the proportion of time spent in that region. In this connection the former non-quantitative version of recurrence was known to Poincaré [1] and had been 'proved' by Gibbs [1] (Chapter XII). Clearly then, ergodic-type questions had been examined considerably earlier than by Birkhoff and von Neumann. Indeed, in a rather different way (although not entirely, considering its relation to the problem of Lagrange (cf. Arnold and Avez [1]; Sternberg [1])), we may consider Weyl's theorem on the uniform distribution mod 1 of $n\alpha_1, \ldots, n\alpha_k$ (when $\alpha_1, \ldots, \alpha_k$, 1 are rationally independent), as the first special ergodic theorem and we may note that it bears the same relationship to Kronecker's theorem on the density of this sequence, as does Birkhoff's theorem to Poincaré's recurrence theorem.

Following the publication of the two ergodic theorems a dichotomy in the development of dynamical systems can be discerned. On the one hand investigations were carried out into the topology of dynamical systems – topological dynamics (Gottschalk and Hedlund

[1]; Krylov and Bogolioubov [1]; Ellis [1]) and the qualitative theory of differential equations (Nemytskii and Stepanov [1]) – programmes initiated by Birkhoff [2] and Poincaré [1], respectively – and on the other, the theory of measure-preserving transformations or ergodic theory was pursued abstractly. The two branches proceeded on more or less independent paths.

Ergodic theory itself shot off in various directions. The basic ergodic theorems were clarified by Khintchine [1]; Hopf [1]; Kakutani and Yosida [1]; and new ergodic theorems were proved by Hopf [2], [3]; Hurewicz [1]; Chacon and Ornstein [1] (cf. also Garsia [1], [2]). The problem of the existence of finite and σ-finite invariant measures was pursued (Hopf [4]; Dowker [1], [2]; Hajian and Kakutani [1]; Ornstein [1]), and dynamical systems were decomposed and represented (Ambrose [1]; Halmos [2]; Ambrose, Halmos and Kakutani [1]; Ambrose and Kakutani [1]; Rohlin [1]). As these results are not related to the main themes of this work, we refer the reader to the addresses and surveys of Halmos [3]; Kakutani [1] and Rohlin [2].

One result in particular calls for special emphasis as it has played such an important role in motivating the research of the past three decades. I am referring to the classification theory of Halmos and von Neumann (von Neumann [3]; Halmos and von Neumann [1]) for ergodic measure-preserving transformations with discrete spectrum. Two such transformations on (Lebesgue) probability spaces are *isomorphic* (cf. Chapter 4) if and only if they have the same set of eigenvalues (in this case, are spectrally equivalent). Halmos and von Neumann displayed all such transformations as compact abelian group translations. Very little further progress on the classification problem was made until 1958, although von Neumann and Anzai [1] constructed interesting non-isomorphic but spectrally equivalent transformations with mixed spectrum.

In 1954 Kolmogorov [1] directed the attention of mathematicians to various unsolved problems in classical mechanics connected with Hamiltonian systems and the decomposition of these systems into invariant tori. Concerning these tori he enquired after the possibility of changing flows into the well-known 'irrational' flows by an analytic change of coordinates. These problems involve questions of approxi-

mability of irrational numbers by rationals (the so-called problem of 'small denominators') and are allied to problems of structural stability. Further progress on these questions was made by Kolmogorov [2]; Arnold [1], [2], [3], [4]; Moser [1]; Herman [1]. The concept of structural stability just referred to is due to Andronov and Pontrjagin [1] and concerns the problem of what basic features of a dynamical system remain intact when the coefficients of that system undergo a small perturbation. If the system remains qualitatively the same, it is said to be structurally stable. Investigations into this form of stability were revived by the works of Peixoto [1]; Arnold and Sinai [1]. Indeed it was principally these works which led to a mushrooming of activity in differentiable dynamical systems and classical mechanics in recent years (cf. Smale [1] [2], Anosov [1]).

These problems are not central to ergodic theory as such, but the direction of this research is particularly interesting to us as it is possible to discern a return to the kind of problems which gave rise to ergodic theory. Moreover, the impetus which ergodic theory received in 1958 moved it in a direction which relates increasingly to the classical mechanical problems. Thus, the divergent strands of dynamical systems are beginning to converge again.

In 1958 Kolmogorov [2] introduced the concept of entropy (borrowed from Shannon's information theory [1]) into ergodic theory and thereby solved an outstanding problem: to find two spectrally equivalent dynamical systems with *continuous* spectrum which are not (spatially) isomorphic. In fact entropy is an isomorphism invariant assigning easily computable numbers to Bernoulli systems (independent processes – cf. Chapter 3) and all Bernoulli systems have the same continuous spectral characteristics. Since that time ergodic theory has made enormous progress in the hands of Rohlin [3], [4], [5], and Sinai [1], [2], [3], [4], initially, and latterly with Ornstein's outstanding results (Ornstein [2], [3]). Sinai employed the new concept in his analyses of classical dynamical systems [2] and, in particular, in his investigations of hard spherical gases [3]. More recently Ornstein classified *all* Bernoulli systems [3] and, together with Weiss (Ornstein and Weiss [1]), showed that the geodesic flows on surfaces of constant negative curvature are Bernoulli flows. The research of recent years is so abundant that

we can do no better than refer the reader to the books and surveys – Shields [1]; Ornstein [2]; Friedman and Ornstein [1]; Weiss [1]; Moser, Phillips and Varadhan [1].

These, then, are some of the main trends in present day ergodic theory. In the early days there were several stages of abstraction implicit in the development of the subject: thermodynamics, statistical mechanics, conservative mechanical systems, measure-preserving transformations. Today, with its new connections with differentiable dynamical systems and aided by the relatively new ideas from information theory, ergodic theory continues to fulfil its earlier promises.

2 Preliminaries

A *probability space* $(X, \mathscr{B}, m)$ is a triple with X a set, $\mathscr{B}$ a σ-algebra of subsets of X, and m a probability measure (a non-negative countably additive function defined on $\mathscr{B}$ with $m(X) = 1$).

A *measure-preserving transformation* (*or endomorphism*) T of $(X, \mathscr{B}, m)$ is a surjective map $T: X \to X$ such that $T^{-1}\mathscr{B} \subset \mathscr{B}$ (i.e. $T^{-1}B \in \mathscr{B}$ whenever $B \in \mathscr{B}$) and $m(T^{-1}B) = m(B)$ for all $B \in \mathscr{B}$.

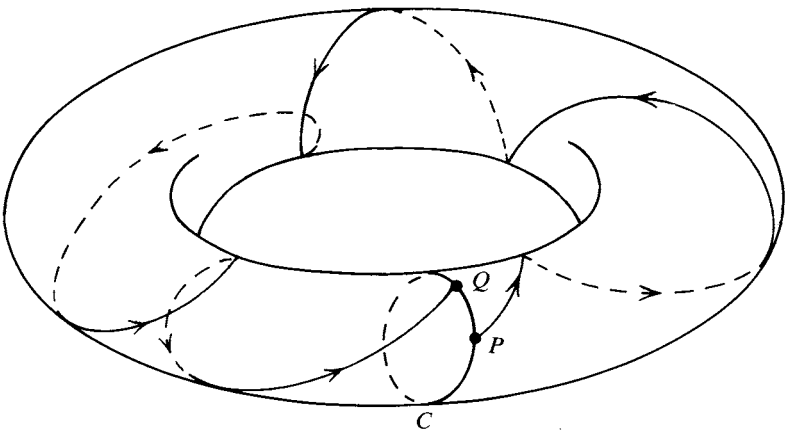
A measure-preserving transformation T is said to be *invertible* (or an *automorphism*) if T is one – one and $T^{-1}\mathscr{B} = \mathscr{B}$ so that the inverse T^{-1} is also a measure-preserving transformation.

A *measure-preserving flow* is a one-parameter group of measure-preserving transformations $\{T_t : t \in R\}$ (each T_t $(t \in R)$ is measure-preserving, and $T_t \circ T_s = T_{t+s}$, $T_0 =$ identity) such that the map $X \times \mathbb{R} \to X, (x, t) \to T_t x$ is measurable.

As we have said in § 1, measure-preserving flows arise naturally in the study of conservative Hamiltonian dynamical systems. Their investigation can be facilitated by a further reduction – to single measure-preserving transformations – on at least two counts:

(1) If the dynamical system or measure-preserving flow possesses a *global cross-section* as illustrated in the diagram below, then associated with that flow we have a measure-preserving transformation of the cross-section and its study will provide considerable information about the total flow.

A *global cross-section* C has the following property: each orbit (in-



dicated by arrowed lines) passes through C infinitely many times and leaves C immediately. We then have a transformation S (the Poincaré map) of C which maps P to the point Q of first return. (There is then a measure on C, which is canonically associated with the original measure on the total space, which is preserved by S.) Detailed knowledge of S can then be used in an analysis of the original dynamical system.

(2) If we choose $\varepsilon > 0$ very small, then we can expect the 'discrete' flow $\{T_{n\varepsilon} : n = 0, \pm 1, \ldots\}$ to be a good approximation to the flow $\{T_t : t \in \mathbb{R}\}$, and thus it is worthwhile studying the single transformation T_ε and its iterates $T_{n\varepsilon} = T_\varepsilon^n$ where, inductively, $T_\varepsilon^n = T_\varepsilon \circ T_\varepsilon^{n-1}$ ($T_\varepsilon^0 =$ identity).

Except for an example in Chapter 5 and some exercises, we shall be concerned exclusively with single transformations (and their iterates) rather than with measure-preserving flows.

3 Conventions

Throughout this work we shall adopt the following conventions concerning relationships modulo sets of measure zero:

Where several measures are involved, such as in § 1, Chapter 1, equations between functions, especially continuous functions, have to be interpreted strictly. For most of this work, where only one measure

is specified, equations and inequalities between functions and between sets are to be interpreted up to a set of measure zero (i.e. they are strictly true after the deletion of some set of measure zero). If $(X, \mathscr{B}, m)$ is a probability space and $\mathscr{A}_1$, $\mathscr{A}_2$ are two sub-σ-algebras, we shall interpret $\mathscr{A}_1 \subset \mathscr{A}_2$ to mean that, for every $A_1 \in \mathscr{A}_1$, there exists $A_2 \in \mathscr{A}_2$ with $m(A_1 \Delta A_2) = 0$. $(A_1 \Delta A_2 = A_1 \cup A_2 - A_1 \cap A_2.)$ $\mathscr{A}_1 = \mathscr{A}_2$ means $\mathscr{A}_1 \subset \mathscr{A}_2$ and $\mathscr{A}_2 \subset \mathscr{A}_1$. A similar convention holds for partitions α_1, α_2 (see Chapters 2 and 4). To make absolutely sure that our relationships are interpreted correctly, we have in many places stressed the convention with the abbreviation a.e. (almost everywhere) or (a.e.) $[m]$ when we wish to specify the measure m.

For a space X and subset B, B^c will denote the complement of B $(B^c = X - B)$.

For a vector space V and subspaces V_1, V_2, $V_1 + V_2$ will denote the subspace consisting of vectors $v_1 + v_2$ with $v_1 \in V_1$, $v_2 \in V_2$. We reserve the convention $V_1 \oplus V_2$ for the case when V_1, V_2 are mutually orthogonal subspaces of a Hilbert space. If $V = V_1 \oplus V_2$, then $V_1^{\perp} = V_2 = V \ominus V_1$.

1

The principal ergodic theorems

1 Uniform distribution (mod 1) and some topological dynamics

We begin with a result of H. Weyl [1] which may be regarded as the first ergodic theorem to be discovered. Admittedly the result is special. It asserts that the sequence $\{x_n\}$ $(x_n = n\alpha)$ is *uniformly distributed* (*mod* 1) when α is irrational, i.e.

$$\frac{1}{N}\sum_{n=0}^{N-1} \chi_I(\,(x_n)\,) \to |I| \quad \text{for all intervals } I \subset [0, 1], \tag{1.1}$$

where $|I|$ denotes the length of the interval I, $(\quad)$ denotes the fractional part of a number, and χ_I denotes the indicator function of I.

Weyl gave the following criterion for uniform distribution (mod 1):

$$\text{for each integer } k \neq 0,\ \frac{1}{N}\sum_{n=0}^{N-1} \exp(2\pi i k x_n) \to 0. \tag{1.2}$$

In fact it is easy to see that if (1.1) holds then

$$\frac{1}{N}\sum_{n=0}^{N-1} f(x_n) \to \int_0^1 f(y)\,\mathrm{d}y \tag{1.3}$$

for all continuous functions f on $[0, 1]$ with $f(0) = f(1)$, since (1.3) will hold for step functions (by taking linear combinations of characteristic functions) and uniform approximation will lead to (1.3). The condition (1.3) will hold for complex functions when it holds for real functions. Hence, as a special case, (1.2) follows from (1.1) (and (1.3)). Conversely, if (1.2) holds, then by taking linear combinations we have (1.3). Here we use Weierstrass's approximation theorem. It is now easy to deduce (1.1) from (1.3). For $I \subset [0, 1]$ choose continuous real-valued functions f, g with $f \geq \chi_I \geq g$ such that

$$f(0) = f(1), \quad g(0) = g(1), \quad \int_0^1 (f - g)\,\mathrm{d}y < \varepsilon.$$

The statement (1.3) for f and g (depending on $\varepsilon > 0$) leads quickly to (1.1). We have shown that:

(1.1), (1.2), (1.3) *are pairwise equivalent.*

Using (1.2) we now see that $\{n\alpha\}$, and more generally $\{n\alpha + x\}$, is uniformly distributed (mod 1) when α is irrational. In fact for $k \neq 0$

$$\frac{1}{N}\sum_{n=0}^{N-1} \exp\left[2\pi i k(n\alpha + x)\right] = \exp(2\pi i k x)\cdot\frac{1}{N}\left[\frac{\exp(2\pi i N\alpha) - 1}{\exp(2\pi i\alpha) - 1}\right] \to 0.$$

We can place these facts into the context of topological dynamics as follows. Define $Tx = x + \alpha$ (mod 1) on $[0, 1]$ with 0, 1 identified. Evidently T is a homeomorphism of a circle and $T^n x = x + n\alpha$ (mod 1).

In this way we see, through the equivalence of (1.2) and (1.3), that

$$\frac{1}{N}\sum_{n=0}^{N-1} f(T^n x) \to \int_0^1 f(y)\,dy \tag{1.4}$$

for all continuous functions defined on the circle. It is clear that T has an equivalent representation as a homeomorphism of the circle $K = \{z \in \mathbb{C} : |z| = 1\}$ given by $Tz = e^{2\pi i\alpha} z$.

In higher dimensions we define $Tx = x + \alpha$ (mod 1), with $x = (x_1, \ldots, x_k)$, $x_i \in [0, 1]$, $\alpha = (\alpha_1, \ldots, \alpha_k)$ and again we identify 0 with 1 in each coordinate. T is a homeomorphism of the k-dimensional torus. (Equivalently, $T(z_1, z_2, \ldots, z_k) = (e^{2\pi i\alpha_1} z_1, \ldots, e^{2\pi i\alpha_k} z_k)$ on the torus $K \times \cdots \times K$ (k times).)

As before (the proof is the same) the following statements are equivalent:

$$\frac{1}{N}\sum_{n=0}^{N-1} \chi_I(T^n x) \to |I| \tag{1.5}$$

for all rectangles $I = I_1 \times \cdots \times I_k$, where $|I|$ denotes the volume of I;

$$\frac{1}{N}\sum_{n=0}^{N-1} \exp 2\pi i \langle h, T^n x\rangle \to 0 \tag{1.6}$$

when $h = (h_1, \ldots, h_k) \neq (0, \ldots, 0)$ (a lattice point), and where $\langle h, x\rangle = h_1 x_1 + \cdots + h_k x_k$;

$$\frac{1}{N}\sum_{n=0}^{N-1} f(T^n x) \to \int_0^1 \cdots \int_0^1 f(y_1, \ldots, y_k)\,dy_1 \cdots dy_k \tag{1.7}$$