Macroeconomic Dynamics, **3**, 1999, 84–114. Printed in the United States of America.

MD SURVEY

NONLINEAR DYNAMICS AND CHAOS PART II: ERGODIC APPROACH

ALFREDO MEDIO
University "Ca' Foscari" of Venice

This is the second part of a two-part survey of the modern theory of nonlinear dynamical systems. We focus on the study of statistical properties of orbits generated by maps, a field of research known as ergodic theory. After introducing some basic concepts of measure theory, we discuss the notions of invariant and ergodic measures and provide examples of economic applications. The question of attractiveness and observability, already considered in Part I, is revisited and the concept of natural, or physical, measure is explained. This theoretical apparatus then is applied to the question of predictability of dynamical systems, and the notion of metric entropy is discussed. Finally, we consider the class of Bernoulli dynamical systems and discuss the possibility of distinguishing orbits of deterministic chaotic systems and realizations of stochastic processes.

Keywords: Nonlinearity, Ergodic Theory, Chaos, Bernoulli Dynamical Systems

1. INTRODUCTION

In Part I of this survey, we discussed dynamical systems from a geometric or topological point of view. The geometric approach is intuitively appealing and lends itself to suggestive graphical representations. Therefore, it has been tremendously successful in the study of low-dimensional systems: continuous-time systems with one or two variables; discrete-time systems with one or perhaps two variables. For higher-dimensional systems, however, the approach has encountered rather formidable obstacles, and rigorous results and classifications are few. Thus, it is sometimes convenient to change perspective and adopt a different approach, based on the concept of measure and aimed at the investigation of the statistical properties of orbits. This requires the use and understanding of some basic notions and results to which we devote Part II. We see that the ergodic theory of dynamical systems often parallels its geometric counterpart and many concepts discussed in Part I (e.g., invariant, indecomposable, attracting sets; attractors; Lyapunov characteristic exponents) are reconsidered in a different light, greatly enhancing our

The author gratefully acknowledges the financial support of the Italian Ministry of the University (MURST) and the National Council of Research (CNR). Address correspondence to Alfredo Medio, Department of Economics, University "Ca' Foscari" of Venice, 873 S. Giobbe, Venice 30121, Italy; e-mail: medio@unive.it.

 1365-1005/99 $9.50

understanding of them. We see that the ergodic approach is most powerful and effective for dealing with basic issues, such as chaotic behavior and predictability, and for investigating the relationship between deterministic and stochastic systems.

2. SOME ELEMENTARY MEASURE THEORY

After these preliminary considerations, we now need to define a certain number of concepts and methods that we use in the remainder of this paper. Ergodic theory discusses dynamical systems in terms of two fundamental mathematical objects: a *measure space* $(X, \mathcal{F}, \mu)$ and a *measure-preserving* map $T : X \to X$ (or, for short, $X \hookleftarrow$). In this context, T is often called "transformation."

X is some arbitrary set; $\mathcal{F}$ is a collection of subsets of X, which is closed under the operations of complementation, countable union, and intersection; it is called a σ-algebra. If X is a metric space, i.e., a space, endowed with a distance function such as $\mathbf{R}^n$, the most natural choice for $\mathcal{F}$ is the so-called *Borel σ-algebra*, $\mathcal{B}$. By definition, this is the smallest σ-algebra containing open subsets of X, where "smallest" means that any other σ-algebra that contains open subsets of X also contains any set contained in $\mathcal{B}$. In what follows, unless stated otherwise, it is always assumed that $\mathcal{F} = \mathcal{B}$. The quantity μ is a measure. In general, a measure $\mu : \mathcal{F} \to \mathbf{R}^+$ is a set function, i.e., a function that assigns nonnegative values to sets. The integral notation

$$\mu(A) = \int_A d\mu(x) \qquad A \in \mathcal{B}(X)$$

is often used.

We are interested here in finite measures (i.e., $0 \leq \mu < \infty$). In this case, we can always normalize μ so that $\mu(X) = 1$ and we then call μ a probability measure. The smallest closed subset of X that is μ-measurable and has a μ-null complement is called the *support* of μ. When μ is *absolutely continuous* (a property that is discussed later), the concept of probability measure may be related to that of *probability density* by the following equation:

$$\mu(A) = \int_A \rho(x)\, dx, \tag{1}$$

where $\rho : X \to \mathbf{R}^+$ is the probability density function, defined on the entire state space. Notice that μ is a function of a set, whereas ρ is a function of the coordinates of the points belonging to the set.

A transformation T is said to be *measurable* if $[A \in \mathcal{F}] \Rightarrow [T^{-1}A = \{x : Tx \in A\} \in \mathcal{F}]$. T is said to be measure-preserving with respect to μ or, equivalently, μ is said to be T-invariant whenever $\mu(T^{-1}(A)) = \mu(A)$ for all sets $A \in \mathcal{F}$.[1] Thus, T-invariant measures are compatible with T in the sense that sets of a certain size (in terms of the selected measure) are mapped by T into sets of the same size.

In the applications with which we are concerned, X denotes the state space and usually we have $X \subset \mathbf{R}^n$; the sets $A \in \mathcal{F}$ denote configurations of the state space

https://doi.org/10.1017/S1365100599010044 Published online by Cambridge University Press

of special interest, such as fixed or periodic points, limit cycles, strange attractors, or subsets of them; the transformation T is the law governing the time evolution of the system. We often refer to the quadruplet $(X, \mathcal{F}, \mu, T)$, or even to the triplets $(X, \mathcal{F}, \mu)$ or $(X, \mathcal{F}, T)$ as "dynamical systems."

Our discussion is concentrated on the study of invariant measures for reasons that can be explained as follows. Because we want to study the statistical, or probabilistic, properties of the orbits of dynamical systems, we need to calculate averages over time. As a matter of fact, certain basic quantities such as Lyapunov characteristic exponents (which, as we saw earlier, measure the rate of divergence of nearby orbits) or metric entropy (which, as we see later, measures the rate of information generated by observations of a system) can be looked at as time averages. For this purpose, it is necessary that orbits $\{x, Tx, T^2x, \ldots\}$ generated by a transformation T possess statistical regularity and certain time limits exist.

More specifically, we often wish to give a meaningful answer to the following question: How often does an orbit originating from a given point of the state space visit a given region of it?

Formally, the problem can be represented thusly: Let T be a transformation of the space X preserving μ, and let A be an element of $\mathcal{F}$. Then we define

$$V_n(x) \equiv \#\{i : 0 \leq i < n, T^i x \in A\}, \tag{2}$$

$$v_n(x) \equiv \frac{1}{n} V_n(x), \qquad x \in X. \tag{3}$$

Therefore, $V_n(x)$ denotes the number of visits of a set A after n interations of T and $v_n(x)$ denotes the *average* number of visits. Now, let n become indefinitely large. We would like to know whether the limit

$$\hat{v}(x) \equiv \lim_{n\to\infty} v_n(x) \tag{4}$$

exists. This can be established by means of a basic result in ergodic theory, known as Birkhoff–Khinchin (B–K) ergodic theorem. In its generality, the ergodic theorem states that, if $T : X \hookleftarrow$ preserves a probability measure μ and f is any function integrable on X, then the limit

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) = \hat{f}(x) \tag{5}$$

exists for μ-almost[2] every point $x \in X$ and that $\hat{f}(x)$ is T-invariant, i.e., $\hat{f}(Tx) = \hat{f}(x)$.

The reader can verify easily that, if we choose $f = \chi_A$, where χ_A denotes the so-called characteristic function, or indicator, of A, i.e.,

$$\chi_A(x) = \begin{cases} 1, & \text{if} \quad x \in A \\ 0, & \text{if} \quad x \notin A \end{cases},$$

https://doi.org/10.1017/S1365100599010044 Published online by Cambridge University Press

then $(1/n)\sum_{i=0}^{n-1} f(T^i x) = v_n(x)$ and therefore the B–K theorem guarantees the existence of limit (4).

2.1. Invariant, Ergodic Measures

In general, however, the limit (5) in the ergodic theorem depends on x, which means that the time averages in question may be different for orbits originating from different initial states. This happens, for example, when the space X is decomposable under the action of T and there exist two subspaces X_1 and X_2, both invariant with respect to T, i.e., when T maps points of X_1 only to X_1 and points of X_2 only to X_2.

The dynamic decomposability of the system—a geometric or topological fact—is reflected in the existence of a T-invariant measure μ that is decomposable in the sense that it can be represented as a weighted average of invariant measures μ_1 and μ_2; i.e., we have $\mu = \alpha\mu_1 + (1-\alpha)\mu_2$, where $\alpha \in (0, 1)$ and μ_1 and μ_2 may or may not in their turn be decomposable.

In general, we are not interested in the properties of a single orbit starting from an arbitrary initial point, but in the overall properties of ensembles of orbits originating from all possible initial conditions in a certain given region of the space. Thus, it would be desirable that the average calculated along a particular "history" of the system should be equal to the averages evaluated over all possible histories. There exists a fundamental class of invariant measures that satisfy the requirement of indecomposability in this sense and are called *ergodic measures*.[3] Several equivalent characterizations of ergodicity exist, of which we select two.

DEFINITION 1. *Given a dynamical system $(X, \mathcal{F}, \mu, T)$, the T-invariant measure μ is called ergodic if*

(*i*) *whenever $T^{-1}(A) = A$ for some $A \in \mathcal{F}$, then either $\mu(A) = 1$ or $\mu(A) = 0$ or, equivalently*,

(*ii*) *the limiting function $\hat{f}$ defined in the Birkhoff–Khinchin ergodic theorem is a constant, and we have*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) = \int_X f(x)\, d\mu(x) \qquad \textbf{(6)}$$

for μ-almost every x.[4]

Definition 1(ii) sometimes is described summarily by saying that "time average equals space average." Again using $f = \chi_A$, where A is a measurable subset of X, the ergodicity of μ implies that the average number of visits to a region A of an orbit originating from almost every point is equal to the size that the ergodic measure assigns to that region. In formal terms, we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_A(T^i x) = \int_X \chi_A(x)\, d\mu(x) = \int_A d\mu(x) = \mu(A) \qquad \textbf{(7)}$$

for μ-almost every $x \in X$.

Before proceeding further in our analysis, we first need to define two special and extremely useful types of measure.

DEFINITION 2. *Let us fix a point* $x \in X$. *We call the* Dirac measure (*centered on* x) *the probability measure* μ *that assigns value* 1 *to all of the subsets* A *of* X *that contain* x, *and value* 0 *to those subsets that do not contain it. Formally*, *we have*

$$\mu(A) = \begin{cases} 1 & \text{if} \quad x \in A \\ 0 & \text{if} \quad x \notin A \end{cases}.$$

The Dirac measure is also called the Dirac delta and usually is denoted by δ_x.

DEFINITION 3. *The* k*-dimensional* Lebesgue measure *is the measure that assigns to each* k*-dimensional open box in* $\mathbf{R}^k$ *its volume.*[5]

The Lebesgue measure, which we henceforth denote by m, corresponds to the intuitive notion of length (for a one-dimensional set) and volume (for k-dimensional ones). It also provides an intuitive and physically relevant notion of probability.

We now provide some examples that will give the rather abstract questions discussed so far a more intuitive flavor and also will establish a bridge between the geometric and the ergodic approaches to dynamical systems.

Example 1

Consider a dynamical system in $\mathbf{R}^n$, characterized by a unique, fixed point x_0. In this case, the Dirac measure δ_{x_0} is invariant and ergodic.

Now, consider the case illustrated in Figure 1 and showing the phase space of the system of ODE's ($\dot{x} = x - x^3$; $\dot{y} = -x$). In this case, the measure $\mu = \alpha\delta_{x_1} + (1 - \alpha)\delta_{x_2}$, where $\alpha \in [0, 1]$, $x_1 = -1$, and $x_2 = 1$, is invariant. However, as the reader can establish easily, μ is not ergodic. It can be decomposed into the two measures δ_{x_1} and δ_{x_2}, which are also invariant. Accordingly, the time averages of orbits starting in the basin of attraction of x_1 are different from those of orbits originating in the basin of attraction of x_2. The system is clearly decomposable from both a geometric and an ergodic point of view.

Example 2

Consider a discrete-time dynamical system characterized by a periodic orbit of period k, $\{x, Tx, \ldots, T^k x = x\}$. In this case, the measure that assigns the value $1/k$ to each point of the orbit is invariant and ergodic.

Example 3

Consider a continuous-time dynamical system characterized by a limit cycle $\Gamma = \{\phi_t(x) : 0 < t \leq \tau\}$, where ϕ_t is the flow map associated with the solution of the system, τ is the period of the cycle, and x is any point on Γ. In this case, the following measure is invariant and ergodic:

$$\mu = \frac{1}{\tau} \int_0^\tau \delta_{\phi_t x}\, dt.$$

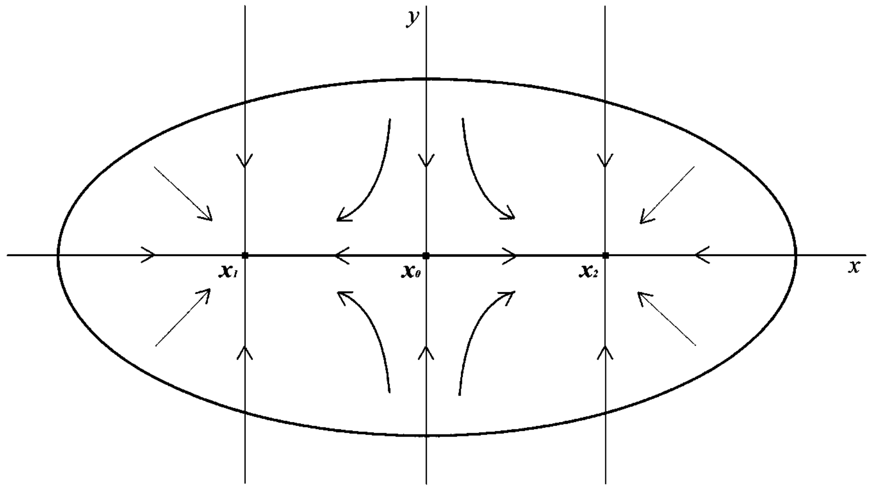

FIGURE 1. A decomposable invariant set.

This means that the probability is spread over the cycle Γ, which is the support of μ, according to the time parameter.

Example 4

Consider the map $T_C : S^1 \to S^1$, where S^1 denotes the unit circle and $T_C(z) = cz$, $c = e^{i\alpha 2\pi}$, $\alpha \in [0, 1)$. A point on S^1 is identified by the angle formed by the abscissa and the line joining the point with the origin. The map T_C rotates points on the circle by an angle $\alpha 2\pi$. It easy to see that T_C preserves the so-called "circular Lebesgue measure" (or "Lebesgue measure on the circle") defined by

$$\hat{m} = m \circ \theta^{-1},$$

where, as usual, m is the Lebesgue measure and $\theta : [0, 1) \to S^1$ is defined by $\theta(x) = e^{i2\pi x}$. In words: When we apply the circular Lebesgue measure to a subset of the circle, first we map it to a corresponding subinterval of [0, 1) and then we assign to that subinterval a value equal to its length.[6] If α is irrational, the measure $\hat{m}$ on the unit circle is ergodic.

When dynamical systems are chaotic, invariant measures seldom can be defined in a precise manner. This difficulty parallels that of precisely locating chaotic invariant sets. We discuss here a well-known case in which the formula for the density of the invariant ergodic measure can be written exactly, i.e., the logistic map $T_L : [0, 1] \hookleftarrow$, $T_L(x) = rx(1 - x)$ with $r = 4$, which we have already discussed at length in Part I. For this purpose, we proceed in steps:

- Step 1. We can verify that the map T_L is related to the (symmetrical) "tent" map

$$T_\Lambda : [0, 1] \to [0, 1],$$

$$T_\Lambda(y) = \begin{cases} 2y & \text{for} \quad 0 \le y \le 1/2 \\ 2 - 2y & \text{for} \quad 1/2 < y \le 1 \end{cases}$$

https://doi.org/10.1017/S1365100599010044 Published online by Cambridge University Press

by the relation

$$T_L \circ \theta = \theta \circ T_\Lambda,$$

where θ denotes the homeomorphism

$$\theta : [0, 1] \hookleftarrow, x = \theta(y) = \sin^2\left(\frac{\pi}{2}y\right).$$

Thus, the two maps are topologically equivalent.

- Step 2. Considering that the counterimage of each subinterval I of [0, 1] under T_Λ consists of two subintervals whose lengths are half the length of I, one can promptly conclude that the tent map T_Λ preserves the Lebesgue measure m. Therefore, T_L must preserve a measure ρ such that

$$\int_{[0,1]} dy = \int_{\theta([0,1])} \rho(dx) = \int_{[0,1]} |\theta'(y)|\rho(dx). \qquad \textbf{(8)}$$

Hence, by the definition of $\theta(y)$ [whence $\theta'(y) = \pi \sin(\frac{\pi}{2}y)\cos(\frac{\pi}{2}y)$] and recalling that $\cos^2(\cdot) = [1 - \sin^2(\cdot)]$, we obtain

$$\rho(dx) = \frac{dy}{|\theta'(y)|} = \frac{dy}{|\pi \sin(\pi y/2)\cos(\pi y/2)|} = \frac{dx}{\pi\sqrt{x(1-x)}}\cdot \qquad \textbf{(9)}$$

We make use of this result later when we discuss the notion of isomorphism.

3. LYAPUNOV CHARACTERISTIC EXPONENTS REVISITED

We can use the ideas discussed earlier to reformulate the definition of Lyapunov characteristic exponents (LCE's) and reconsider the question of the dependence of LCE's on initial conditions.

In the simpler one-dimensional case, consider a transformation $T : X \hookleftarrow$ with X an interval of $\mathbf{R}$, and a T-invariant probability measure μ.

From our discussion of ergodicity, we know that, if μ is ergodic, for μ-almost all x the time average of an integrable function f, i.e., $\bar{f} = \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^i x)$ is equal to the expected value of f, $\hat{f} = \int_X f(x)\,d\mu(x)$. If we now choose $f(x) = \ln|T'(x)|$, the LCE of a map T can be written as

$$\lambda = \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \ln|T'(T^i(x))| = \int_X \ln|T'(x)|\,d\mu(x), \qquad \textbf{(10)}$$

which is independent of the initial condition. In this simple case, the interpretation of (10) is clear. The quantity $\ln|T'(x)|$—the ln of the absolute value of the slope of the curve generated by the map T in the (x_{x+1}, x_n) plane—measures the (exponential) rate at which small discrepancies between trajectories (or small errors) are amplified by the action of the map. In general, that slope varies with x and its different values are weighted by μ. Values of the slope obtaining over sets of x whose μ-measure is zero do not affect the final results.

All of this can be illustrated easily by two simple examples.

Example 5

The asymmetric tent map

$$T_{\hat{\Lambda}} : [0,1] \to [0,1]$$

$$T_{\hat{\Lambda}}(x) = \begin{cases} x/a & \text{for} \quad 0 \le x \le a \\ (1-x)/(1-a) & \text{for} \quad a < x \le 1 \end{cases}. \qquad \textbf{(11)}$$

It is easily seen that (11) preserves the length of subintervals of [0,1] and therefore it preserves the Lebesgue measure. In this case we have $d\mu(x) = dx$. Hence

$$\begin{aligned} \lambda &= \int_0^1 \ln|T'_{\hat{\Lambda}}(x)|\,dx = \int_0^a \ln\left(\frac{1}{a}\right)dx + \int_a^1 \ln\left(\frac{1}{1-a}\right)dx \\ &= a\ln\left(\frac{1}{a}\right) + (1-a)\ln\left(\frac{1}{1-a}\right). \end{aligned} \qquad \textbf{(12)}$$

Clearly, for $a = 1/2$, we are in the case of the symmetric tent map and $\lambda = \ln 2$.

Example 6

The logistic map $T_L(x) = 4x(1-x)$. Recalling the result (9), we have for almost all points x,

$$\lambda = \int_{[0,1]} \ln|T'_L(x)| \frac{dx}{\pi[(x(1-x)]^{1/2}} = \int_{[0,1]} \frac{\ln|4-8x|}{\pi[x(1-x)]^{1/2}}\,dx = \ln 2. \qquad \textbf{(13)}$$

(However, notice that, if we choose the special initial point $x = 0$, we have $\lambda(0) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \ln|T'_L(x_i)| = \ln|T'_L(0)| = \ln 4$).

In the general n-dimensional case, the ergodic theorem cannot be applied as simply and directly as in the one-dimensional case. Broadly speaking, the root of the difficulty is the fact that, when T is multidimensional, the quantities $DT(x)$ (where D is the partial derivative operator) are not scalars but noncommuting matrices and, consequently,

$$\ln\left\| \prod_{i=0}^{n-1} DT(x_i) \right\| \neq \sum_{i=0}^{n-1} \ln\|DT(x_i)\|.$$

However, the celebrated multiplicative ergodic theorem, first proved by Oseledec,[7] provides an extension of the ergodic theorem to products of matrices. Let us define the LCEs as

$$\lambda(x,w) = \lim_{n\to\infty} \frac{1}{n}\ln\|DT^n(x)w\|, \qquad \textbf{(14)}$$

https://doi.org/10.1017/S1365100599010044 Published online by Cambridge University Press

where w is a vector in the tangent space at x. [Notice the slight, inessential difference from the definition (17) of Part I of this survey.]

The Oseledec theorem proves that the limit (14) exists for μ-almost all x, where, as usual, μ indicates a T-invariant probability measure. If w is chosen at random, the limit (14) will be equal to the *largest* LCE. As $w \neq 0$ changes in the tangent space, $\lambda(x, w)$ will take $s \leq n$ different values. The theorem also proves that, if μ is ergodic, the λ's are μ-almost everywhere constant (i.e., they do not depend on the initial conditions x).

4. NATURAL, ABSOLUTELY CONTINUOUS, SRB MEASURES

To establish that an invariant measure is ergodic may not be enough, however, to make it interesting. In our discussion of dynamical systems from a geometric point of view, we first considered the properties of invariance and indecomposability of a set under the action of a map (or a flow). We also tried to establish the conditions under which an invariant set is, in an appropriate sense, observable. In so doing, however, we encountered some conceptual difficulties. In particular, we remarked that attractors, i.e., sets toward which orbits converge, are not necessarily attracting (asymptotically stable) sets. We now reconsider the question of observability afresh.

Consider again the system described by Figure 1. The stable fixed points x_1 and x_2 are observable: If we pick the initial conditions at random and we plot the orbit of the system on a computer screen, we have a nonnegligible chance of seeing either of them on the screen. On the contrary, the unstable fixed point $x_0 = 0$ is not observable unless we start from points on the y axis—which is an unlikely event—and there are no errors or disturbances. In this case, although all three measures $\delta_{x_1}, \delta_{x_2}, \delta_{x_0}$ are invariant and ergodic, only the first two are physically relevant.

The basic reason is that the set defined by "μ-almost initial conditions" in (7) may be too small vis-à-vis the state space and therefore negligible in a physical sense. To avoid this situation, we would like to find a measure that is determined by the time averages (7) for randomly chosen initial conditions x_0. A natural way of making the concept of "randomly chosen" more precise is to require that $x_0 \in B$, where B is a set of Lebesgue measures $m(B) > 0$. Measures satisfying this requirement are essentially unique and sometimes are called *natural invariant measures* and the corresponding density is called *natural invariant density*; the phrase *physical measure* also is used.

Whereas the properties characterizing a natural measure are verified easily in the simpler cases for which the dynamics are not too complicated, in the general case including complex or chaotic systems the determination of the natural invariant measure is a hard and not entirely solved problem. Notice this fact, however: An invariant ergodic measure μ that is absolutely continuous with respect to the Lebesgue measure will automatically satisfy the requirement for a natural measure. To see this, let us consider the following definition.

DEFINITION 4. *Given a measurable space* $(X, \mathcal{F})$ *and two probability measures* μ_1 *and* μ_2, *we say that* μ_1 *is* absolutely continuous *with respect to* μ_2 *(denoted sometimes as* $\mu_1 \ll \mu_2$*) if, for any set* $B \in \mathcal{F}$, $[\mu_2(B) = 0] \Rightarrow [\mu_1(B) = 0]$.

Absolute continuity with respect to the Lebesgue measure m (often simply called "absolute continuity") of a measure μ excludes therefore the possibility that sets that are negligible with respect to the Lebesgue measure (and thus negligible in a physical sense) are assigned a positive value by μ because, from Definition 4, it follows that if $\mu \ll m$, then $[\mu(B) > 0] \Rightarrow [m(B) > 0]$.

Unfortunately, dissipative systems, which form a class of systems often studied in the applications that interest us here, cannot have invariant absolutely continuous measures. By definition, such systems contract volumes of initial conditions, and thus their attractors must have a (k-dimensional) Lebesgue measure equal to zero.

Although no general existence theorems for natural invariant measures exist, it seems that in most cases the experimental, or computer-generated, trajectories of systems automatically produce well-defined time averages. Operationally, an invariant physical measure should describe the distribution in space of points generated by the time evolution of the system. Formally, such measure can be defined as the time average of Dirac deltas centered at the points visited by the system, as follows:

$$\mu = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \delta_{T^i x} \tag{15}$$

(discrete time), where T is the map governing the motion of the system, or

$$\mu = \lim_{\tau\to\infty} \frac{1}{\tau} \int_0^\tau \delta_{\varphi_t x}\, dt \tag{16}$$

(continuous time), where φ_t denotes the relevant flow map.

Because experimental, or computer-generated, dynamics of systems often seem to provide a natural selection among the many (even uncountably many) invariant measures, the question arises whether it is possible to define rigorously a class of invariant measures (and thereby a class of dynamical systems) possessing the desired properties of physical measures. An answer to this question is provided by the so-called Sinai–Ruelle–Bowen (SRB) measures, broadly defined as measures that are absolutely continuous *along the unstable directions*.[8] For systems possessing an SRB measure ρ, there exists a subset S of the state space of positive Lebesgue measure such that, for all orbits originating in S, the SRB measure is given by the time averages (15) or (16).

Although dissipative systems cannot have absolutely continuous invariant measures, they can have SRB measures. These are smooth (have densities) in the stretching directions, but are rough (have no densities) in the contracting directions.

Another way of providing a selection of the natural measure is based on the observation that, owing to the presence of noise—determined by physical disturbances or by computer roundoff errors—the time evolution of a system can be looked at as a stochastic process. Under commonly verified assumptions, the latter has a unique stationary measure ρ_ε, which is a function of the level of noise ε. *If this measure tends to a definite limit as the level of noise tends to zero,* this limit—sometimes called a *Kolmogorov measure*—can be taken as a description of the natural measure. For some, but not all, systems the Kolmogorov and the SRB measures actually coincide.

5. ATTRACTORS AS INVARIANT MEASURES

The ideas and results discussed in the preceding sections suggest a characterization of attractors that integrates their geometric and ergodic features. The definition below follows Milnor's (1985, p. 179) and Palis and Takens' (1993, p. 138) discussion on the subject and somewhat modifies that provided in Part I of this survey.

DEFINITION 5. *Suppose a dynamical system is described by a transformation* $T : X \hookleftarrow$. *We say that a compact set* $A \subset X$ *is an* attractor *if* (i) *the basin of attraction* $B(A)$*—i.e., the set of points* $x \in X$ *that are asymptotic to some point* $y \in A$ *in the sense that the distance* $d(T^n x, T^n y) \to 0$, *as* $n \to \infty$*—has positive Lebesgue measure*; *and* (ii) T *is transitive on* A. (*Transitivity of* T *on* A *means that* T *has an orbit that is dense in* A *and therefore the dynamics on the attractor are indecomposable.*)

Suppose now that there exists an ergodic SRB measure ρ preserved by T with support on the attractor A. This is known to be the case for certain classes of dynamical systems with a *hyperbolic structure*. Roughly speaking, this requires that the evolution of the system can be decomposed neatly into expanding and contracting directions. In this case, for initial conditions $x \in B(A)$ and all integrable functions f, we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) = \int f \, d\rho. \tag{17}$$

Thus, we have a nice correspondence between the geometric and the ergodic properties of attractors, represented respectively by the set A and the measure ρ with support in A. Although the system's attractor may have Lebesgue measure zero and therefore, strictly speaking, may not be observable, we can study its statistical properties by looking at the time evolution of an observable ensemble of orbits originating in a basin of attraction of a nonnegligeable size (positive Lebesgue measure). These orbits behave in a statistically regular manner and this regularity is governed by the invariant measure ρ. Some authors [cf. MacKay (1992, p. 5)] would take the ergodic approach one step further and suggest the following definition:

DEFINITION 6. *An attractor is an invariant, ergodic measure ρ such that the set of generic points x—i.e., the points for which* (17) *holds—has positive Lebesgue measure.*

6. PREDICTABILITY, ENTROPY

The rather formidable apparatus described above will allow us to discuss the question of predictability of chaotic systems in a rigorous manner. In so doing, however, we first must remove a possible source of confusion. The ergodic approach analyzes dynamical systems by means of probabilistic methods. One might immediately point out that the outcome of deterministic dynamical systems, such as those represented by differential or difference equations and discussed in Part I, are not random events but, under usually assumed conditions, they are uniquely determined by initial values. Consequently, one might conclude that measure and probability theories are not the appropriate tools of analysis. Prima facie, this seems to be a convincing argument. If we know the equations of motion of a deterministic system *and we can monitor its state with infinite precision*, then there is nothing left to discuss: The future of the system can be forecast exactly.

Infinite precision of observation is a purely mathematical expression, however, and it has no physical counterpart. When dynamical systems theory is applied to real problems, a distinction therefore must be made between *states* of a system, i.e., points in a state space, and *observable states*, i.e., subsets (or cells) of the state space, whose (nonzero) size reflects our limited power of observation. For example, we cannot verify by observation the statement that the length of an object is π cm (a number with an infinitely long string of decimals, thus containing an infinite amount of information). Under normal circumstances, however, we can easily verify the statement that the length of the object is, for example, between 3 and 4 cm. Alternatively, we can think of the situation occurring when we plot the orbits of a system on the screen of our computer: What we see are not actual points but pixels of small but nonzero size; the greater the resolution of the graphics environment, the smaller are the pixels.

On the other hand, in real systems, perfect foresight only makes sense when it is interpreted as an asymptotic state of affairs that is approached as observers (e.g., economic agents) accumulate information and learn about the position of the system. Much of what follows concerns the conditions under which prediction is possible, given precise knowledge of the equations of the system (i.e., given a deterministic system), but an imprecise, albeit arbitrarily accurate, observation of its state.

As becomes apparent in the discussion that follows, the distinction between state and observable state is unimportant for systems whose orbit structure is simple (e.g., systems characterized by a stable fixed point or a stable limit cycle). That is, for these systems, the assumption of infinite precision of observation is a convenient simplification and all of the interesting results of the investigation still hold qualitatively if that unrealistic assumption is removed. The distinction,

https://doi.org/10.1017/S1365100599010044 Published online by Cambridge University Press

however, is essential for complex, or chaotic, systems. Indeed, one might even say that many of the characterizing features of chaotic dynamics—above all, their lack of predictability—can be understood only by taking into account the basic physical fact that observation can be made arbitrarily, but not infinitely, precise. Finally, we see that the study of the dynamics of observable states provides the essential link between deterministic and stochastic systems.

To be more precise, we formalize the notion of partition. A *finite partition* $\mathcal{P}$ of X is a collection $\{P_1, \ldots, P_N\}$ of disjoint sets whose union is equal to X. A partition also can be viewed as a function—if you wish, an "observation function"—$\mathcal{P} : X \to \{P_1, \ldots, P_N\}$ such that, for each point of the state space $x \in X, \mathcal{P}(x)$ is the element of the partition, the cell, or atom, of X, in which x is contained. Consider now the action of a transformation T of X. The compound function $T^{-1} \circ \mathcal{P}$ gives us the preimage under T of each point contained in the cell $\mathcal{P}(x)$ and, in so doing, defines another partition of X.[9] Next, consider the operation $\mathcal{P}_1 \bigvee \mathcal{P}_2$: It consists of all possible intersections of the elements of the partitions $\mathcal{P}_1$ and $\mathcal{P}_2$ and is called a "join." Then, the join $T^{-1} \circ \mathcal{P} \bigvee \mathcal{P}$ forms a partition of the space $X \times X \equiv X^2$ of all *sequences* of two states in X occupied by the system under the action of T. Once again, the partition can be viewed as a function that assigns to each point in X^2, i.e., to each two-state sequence, a corresponding sequence of two cells P_i, P_j $(i, j = 1, \ldots, N)$. As we repeat the backward iteration of T and the join operation n times, we obtain the partition

$$\bigvee_{i=0}^{n-1} T^{-i}(\mathcal{P}) \equiv \mathcal{P} \bigvee T^{-1}(\mathcal{P}) \bigvee \cdots \bigvee T^{-n+1}(\mathcal{P}), \tag{18}$$

which is a function that, for each sequence of n states of the system, assigns a corresponding sequence of n cells. In what follows, we study the probabilistic properties of dynamics of systems observed in a partitioned state space.

Each observation can be looked at as an experiment whose outcome is uncertain. The uncertainty of the experiment or, equivalently, the amount of information contained in one observation, can be measured by means of a well-defined quantity called *entropy*. Broadly speaking, if we consider a random variable ξ taking a finite number N of values with probability $p_1, \ldots, p_N$, we define the entropy of ξ by the quantity[10]

$$H(\xi) = -\sum_{i=1}^{N} p_i \ln(p_i), \tag{19}$$

where we take the convention that $0 \ln 0 = 0$. Notice that H is maximum $(\ln N)$ when $p_i = 1/N$ for all i, and minimum (zero) when one of the p's is equal to 1, the others being 0. Thus, if we consider a game of dice, the maximum entropy of a throw (the maximum uncertainty about its outcome) obtains when each of the six facets of a die has the same probability $(1/6)$ of turning up. An unfair player can reduce the uncertainty (to himself) of the outcome by "loading" the dice and thereby increasing the probability of one or more of the six faces (and, correspondingly, decreasing the probability of the others).

When dealing with the predictability of a dynamical system, we are not interested in the entropy of a partition of the state space (the information contained in a single experiment) but with the entropy of the system (the rate at which replications of the experiments, i.e., repeated, finite-precision observations of the system as it evolves in time, produce information), when the number of observations becomes very large.

To make this idea more precise, let us consider again a dynamical system (X, T, μ), where the state space is restricted to the support of the ergodic, T-invariant measure μ, and a finite, μ-measurable partition of X, $\mathcal{P} = (P_1, \ldots, P_N)$. As we discussed earlier, because μ is ergodic, $\mu(P_i)$ measures the probability of finding the system in the cell P_i. Thus, the entropy of $\mathcal{P}$ will be equal to

$$H(\mathcal{P}) = -\sum_{i=1}^{N} \mu(P_i) \ln \mu(P_i) \tag{20}$$

and the entropy of the super partition $\bigvee_{i=0}^{n-1} T^{-i}(\mathcal{P})$, namely,

$$H\left(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}\right), \tag{21}$$

can be calculated analogously, summing over all of the cells of $\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}$. If we now divide $H(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P})$ by the number of observations n, we obtain the *average* amount of information contained in—the average amount of uncertainty about—the super experiment consisting of the repeated observation of the system along a typical orbit. If we increase the number of observations indefinitely, we obtain[11]

$$h(\mu, \mathcal{P}) = \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}\right), \tag{22}$$

which is the entropy of the system with respect to the partition $\mathcal{P}$.

The RHS of equation (22) is (the limit of) a fraction: The numerator is the entropy of a partition obtained by iterating T and the denominator is the number of iterations. Loosely speaking, if when the number of iterations increases, $H(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P})$ remains bounded, limit (22) will be zero; if it grows linearly with n, the limit will be a finite, positive value; if it grows more than linearly, the limit will be infinite. To interpret this result, consider that each cell of the partition $\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}$ corresponds to a sequence of length n of cells of $\mathcal{P}$, i.e., to an orbit of length n of the system, observed with $\mathcal{P}$ precision. The quantity $H(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P})$ will increase or not with n according to whether, increasing the number of observations, the number of possible sequences also increases. From this point of view, it is easy to understand why simple systems, e.g., those characterized by attractors that are fixed points or periodic orbits, have zero entropy. Transients apart, for those systems the possible sequences of states are limited and their number does not increase with the number of observations. Complex systems are precisely those for which the number of possible sequences of states grows with the number of observations, in such a way that limit (22) tends to a positive value. For finite-dimensional, deterministic systems characterized by bounded attractors, entropy is bounded above by the sum of the positive Lyapunov exponents and is therefore finite.

The entropy of a system with respect to a given partition can be given an alternative, very illuminating formulation by making use of the auxiliary concept of *conditional entropy* of $\mathcal{A}$ given $\mathcal{B}$, defined by

$$H(\mathcal{A} \mid \mathcal{B}) = -\sum_{A,B} \mu(A \cap B) \ln \mu(A \mid B), \tag{23}$$

where A, B denote elements of the partitions $\mathcal{A}$ and $\mathcal{B}$, respectively. If we think of a partition as an experiment whose outcome is uncertain, then conditional entropy can be viewed as the amount of uncertainty of the experiment $\mathcal{A}$ when the outcome of the experiment $\mathcal{B}$ is known.

It can be shown that [cf. Billingsley (1965, pp. 79–82)]

$$\lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}\right) = \lim_{n\to\infty} H\left(\mathcal{P} | \bigvee_{i=1}^{n} T^{-i}\mathcal{P}\right). \tag{24}$$

Equation (24) provides another useful interpretation of $h(\mu, \mathcal{P})$): It is the amount of uncertainty of—the amount of information contained in—an observation of the system in the partitioned state space, conditional upon the (finite-precision) knowledge of its states over the infinitely remote past. Therefore, zero entropy means that knowledge of the past removes all uncertainty about the future; i.e., the system is entirely predictable. On the contrary, positive entropy means that, no matter how long we observe the evolution of the system, additional observations still have a positive information content; i.e., the system is not entirely predictable.

So far, we have been talking about entropy relative to a specific partition. The entropy of a system then is defined to be

$$h(\mu) = \sup_{\mathcal{P}} h(\mu, \mathcal{P}), \tag{25}$$

where the supremum is taken over all finite partitions.[12]

The quantity $h(\mu)$ also is known as Kolmogorov–Sinai (K–S), or metric entropy. Unless we indicate differently, by entropy we mean K–S entropy.

Remark 1. In the mathematical literature, as well as in economic applications, one can find a related concept, known as *topological entropy*. Consider a transformation T of the state space M onto itself, together with a partition $\mathcal{P}$ of M. Let $N(\mathcal{P})$ be the number of elements of $\mathcal{P}$. The topological entropy of $\mathcal{P}$ is defined as

$$H_{\mathrm{TOP}}(\mathcal{P}) = \ln N(\mathcal{P}). \tag{26}$$

https://doi.org/10.1017/S1365100599010044 Published online by Cambridge University Press

Then, the topological entropy of T with respect to $\mathcal{P}$ is

$$h_{\text{TOP}}(T, \mathcal{P}) = \lim_{n\to\infty} \frac{1}{n} H_{\text{TOP}}\left(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}\right). \tag{27}$$

Finally, the topological entropy of T is defined as

$$h_{\text{TOP}}(T) = \sup_{\mathcal{P}} h_{\text{TOP}}(T, \mathcal{P}). \tag{28}$$

Comparing (20–25) and (26–28), the reader will notice that, in the computation of $H_{\text{TOP}}(\mathcal{P})$ and consequently of $h_{\text{TOP}}(T)$, we have not taken into account the probability measure of the elements of the relevant partition. If M is compact, there is a simple relation between topological and metric entropies, namely:

$$h_{\text{TOP}}(T) = \sup_{\mathcal{M}} h(\mu),$$

where $\mathcal{M}$ is the set of the ergodic measures invariant with respect to T. A positive topological entropy indicates the presence of an invariant ergodic measure and a corresponding invariant set over which the dynamics are chaotic (unpredictable). However, when $h_{\text{TOP}}(T) > 0$ but the metric entropy with respect to the natural measure is zero, chaos may take place over a region of the state space that is too small to be observed. This phenomenon is nicknamed thin chaos.

Actual computation of the metric entropy $h(\mu)$ directly from its definition looks a rather desperate project. Fortunately, a result due to Kolmogorov and Sinai guarantees that, under conditions often verified in specific problems, the entropy of a system $h(\mu)$ can be obtained from the computation of its entropy relative to a given partition, $h(\mu, \mathcal{P})$. Formally, we have the following[13]

THEOREM 1 (Kolmogorov–Sinai). *Let $(X, \mathcal{F}, \mu)$ be a measure space, T a transformation preserving μ, and $\mathcal{P}$ a partition of $(X, \mathcal{F}, \mu)$ with finite entropy. If $\bigvee_{i=0}^{\infty} T^{-1}\mathcal{P} = \mathcal{F}$ mod 0, then $h(\mu, \mathcal{P}) = h(\mu)$.*

In this case, $\mathcal{P}$ is called a *generating partition* or a *generator*. Intuitively, a partition $\mathcal{P}$ is a generator if, given a tranformation T acting on a state space X, to each point $x \in X$ there corresponds a unique infinite sequence of cells of $\mathcal{P}$, and vice versa. In what follows, we repeatedly apply this powerful result.

A simple example will help to clarify these rather difficult ideas. Consider the already-mentioned symmetrical tent map T_Λ on the interval [0, 1] and the partition consisting of the two subintervals located, respectively, to the left and to the right of the $1/2$ point (remember that measure zero sets "do not count"). Thus, we have a partition $\mathcal{P} = \{P_1, P_2\}$ of [0, 1], where $P_1 = \{0 < x < 1/2\}$ and $P_2 = \{1/2 < x < 1\}$. Then the atoms of $T^{-1}P_1$ are the two subintervals $\{0 < x < 1/4\}$ and $\{3/4 < x < 1\}$ and the atoms of $T^{-1}P_2$ are the two subintervals $\{1/4 < x < 1/2\}$ and $\{1/2 < x < 3/4\}$. Hence, taking all possible intersections of subintervals, the join $\{T^{-1}\mathcal{P} \bigvee \mathcal{P}\}$ consists of the four subintervals $\{0 < x < 1/4\}$, $\{1/4 < x < 1/2\}$, $\{1/2 < x < 3/4\}$, $\{3/4 < x < 1\}$. Repeating the same

https://doi.org/10.1017/S1365100599010044 Published online by Cambridge University Press

operation, at the $(n-1)$th step the join $\{\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}\}$ is formed by 2^n subintervals of equal length 2^{-n}, defined by $\{x : (j-1)/2^n < x < j/2^n\}$, $1 \le j \le 2^n$. Moreover, it is easy to see that, for the tent map, if we use the Borel σ-algebra, the selected partition is a generator. Hence, we can apply the Kolmogorov–Sinai theorem and have

$$h(\mu) = h(\mu, \mathcal{P}).$$

Finally, taking into account the fact that the tent map preserves the Lebesgue measure m, we conclude that the K–S entropy of the map is equal to

$$h(m) = \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}\right) = \lim_{n\to\infty} \frac{1}{n}[-2^n(2^{-n}\ln(2^{-n}))] = \ln 2.$$

Before concluding this section, we would like to recall that entropy is closely linked with another type of statistical invariants, the LCE's, which, roughly speaking, measure the rates of divergence of nearby orbits, or if you prefer, the sensitivity to initial conditions. It can be shown that, in general, we have the following inequality:

$$h(\mu) \le \sum_{i:\lambda_i>0} \lambda_i, \tag{29}$$

where λ denotes an LCE. For systems characterized by an SRB measure, strict equality holds.[14] As we have seen before, the equality indeed holds for the tent map.

The close relation between entropy and LCE is not surprising. We have already observed that entropy crucially depends on the rate at which the number of new possible sequences of coarsed-grained states of the system grows as the number of observations increases. But this rate is strictly related to the rate of divergence of nearby orbits, or, if you prefer, to the rate of amplification of errors, which, in turn, is measured by the LCE's. Thus, the presence of one positive LCE on the attractor signals positive entropy and unpredictability of the system.

7. ISOMORPHISM

In the discussion of dynamical systems from a geometric point of view, we have encountered the notion of topological equivalence. Analogously, there exists a fundamental type of equivalence relation between measure-preserving transformations, called *isomorphism*, which plays a very important role in ergodic theory and which we shall use in the sequel.

DEFINITION 7. *Let T and $\hat{T}$ be two transformations acting, respectively, on the state spaces X and $\hat{X}$ and preserving, respectively, the measures μ and $\hat{\mu}$. We say that T and $\hat{T}$ are* isomorphic *if an invertible map $\theta : X \to \hat{X}$ exists such that (excluding perhaps certain sets of measure zero):*

(*i*) *The following diagram commutes*

https://doi.org/10.1017/S1365100599010044 Published online by Cambridge University Press

$$\begin{tikzcd} X \arrow[r, "T"] \arrow[d, "\theta"'] & X \arrow[d, "\theta"] \\ \hat{X} \arrow[r, "\hat{T}"'] & \hat{X} \end{tikzcd}$$

i.e., we have $\hat{T} \circ \theta = \theta \circ T$.

(*ii*) *The map* θ *preserves the probability structure; i.e., if* I *and* $\hat{I}$ *are, respectively, measurable subsets of* X *and* $\hat{X}$, *then* $\mu(I) = \hat{\mu}(\theta(I))$ *and* $\hat{\mu}(\hat{I}) = \mu(\theta^{-1}(\hat{I}))$.
Maps having the properties (i) *and* (ii) *are called* isomorphisms.

Certain properties such as ergodicity and entropy are invariant under isomorphism. Consequently, isomorphic transformations have the same entropy. The reverse is true only for a certain class of transformations called Bernoulli of which more later.

8. APERIODIC AND CHAOTIC DYNAMICS IN ECONOMIC MODELS

As we mentioned before, it can be ascertained easily that "simple" systems, e.g., systems whose attractors are fixed points or periodic orbits, all have zero entropy and their dynamics therefore can be predicted. We then apply the ideas discussed in the preceding sections to complex systems, i.e., systems whose attractors are aperiodic. These systems have one feature in common with stochastic processes: Their asymptotic behavior can be discussed by considering densities of states, rather than orbits. However, we shall see that complex systems in this sense are not necessarily chaotic in the sense relevant to our economic discussion; i.e., they are not necessarily unpredictable. To do so, we distinguish between two fundamental classes of behavior: quasiperiodic (aperiodic but not chaotic) and chaotic.

8.1. Quasiperiodic Dynamics

Aperiodic nonchaotic (quasiperiodic) behavior arises in a number of models in economics, of which we mention here three main classes, all of them formulated in a discrete-time setting. First, we have the models describing optimal growth.[15] The second class comprises models of overlapping generations with production.[16] Finally, we have models of Keynesian (or perhaps Hicksian) derivation, describing the dynamics of a macroeconomic system characterized by nonlinear multiplier–accelerator mechanisms.[17]

In the works mentioned above, it has been shown that, under certain not unreasonable conditions on the parameters, these models can undergo a *Neimark bifurcation*. The latter describes a situation in which a stable equilibrium loses its stability when, because of the change of a controlling parameter, the modulus of a pair of complex conjugate eigenvalues of the Jacobiam matrix at equilibrium crosses the unit circle. If we except certain special resonance cases, a Neimark bifurcation generates an invariant circle the dynamics on which can be periodic or quasiperiodic according to whether a certain quantity called "rotation number" is rational or irrational. In the quasiperiodic (as well as in the periodic) case, entropy is zero and the dynamics are predictable.

To discuss these statements in a more precise manner, let us now state the basic result for the present question, which is sometimes referred to as the (discrete-time) Hopf bifurcation theorem, but was in fact first stated by Neimark (1959) (whence the more accurate name of Neimark bifurcation) and subsequently rigorously proved by Sacker (1964). The following discussion is based on Iooss (1979).

THEOREM 2 (Neimark, Sacker). *Let* $T_\mu : \mathbf{R}^2 \to \mathbf{R}^2$ *be a map of class* $C^k, k \geq 5$, *depending on a parameter* μ *so that* $x \in \mathbf{R}^2 = 0$ *is a fixed point of* T_μ *and let the following conditions be satisfied*:

(*i*) *For* μ *near zero, the Jacobian matrix* $D_x T_\mu$ *has two nonreal, conjugated eigenvalues* $\lambda(\mu)$ *and* $\hat{\lambda}(\mu)$, *with* $|\lambda(0)| = 1$;
(*ii*) $\frac{d|\lambda(0)|}{d_\mu} \neq 0$;
(*iii*) $\lambda^i \neq 1$ *for* $i = 1, 2, 3,$ *and* 4.

Then, after a trivial change of the μ *coordinate and a smooth*, μ-*dependent coordinate change of* **R**,

(*i*) *we can write the following approximation of the map* T_μ *in polar coordinates*:

$$\left.\begin{aligned} r_{n+1} &= (1+\mu)r_n - \alpha(\mu)r_n^3 \\ \phi_{n+1} &= \phi_n + \beta(\mu) + \gamma(\mu)r_n^2 \end{aligned}\right\} + \textit{higher-order terms}, \qquad \textbf{(30)}$$

where α, β, γ, *are smooth functions of* μ.

(*ii*) *Moreover*, *for* $\alpha > 0$ (*respectively*, *for* $\alpha < 0$) *and in a sufficiently small right* (*left*) *neighborhood of* $\mu = 0$, *there exists an invariant attractive (repelling) circle* Γ_μ *for the map* T_μ, *bifurcating from the fixed point* 0.

Assuming now that $\alpha > 0$ and the invariant circle is (locally) attractive, the behavior of T_μ restricted to the set Γ can be approximated arbitrarily well by iterations of a homeomorphism of the circle $f_\mu : S^1 \to S^1$. The dynamics of the latter crucially depend on an invariant of f_μ called *rotation number*, defined as (the index $_\mu$ is omitted henceforth)

$$\rho(f) = \lim_{n\to\infty} \frac{\hat{f}^n(x) - x}{n},$$

where the map $\hat{f}$ is the so-called *lift* of f; i.e., a map $\hat{f} : \mathbf{R} \to \mathbf{R}$ such that, putting $P : \mathbf{R} \to S^1$, $P(x) = e^{i2\pi x}$, we have $P \circ \hat{f} = f \circ P$. The lift f is unique up to the addition of an integer. The limit $\rho(f)$ exists and is independent of x and the choice of $\hat{f}$.

There are two basic possibilities:

(1) The limit $\rho(f) = p/q$, p and q being two integers; i.e., ρ is rational. In this case, the map f has periodic orbits of period q.
(2) The limit $\rho(f)$ is irrational, which is the case that interests us here. In this case, a known result[18] states that if $f \in C^2$, then it is topologically conjugate (and therefore

https://doi.org/10.1017/S1365100599010044 Published online by Cambridge University Press

it is dynamically equivalent) to a fixed rotation of the circle; that is, there exists a homeomorphism $\theta : S^1 \to S^1$ such that $\theta \circ f = T_\rho \circ \theta$, where

$$T_\rho(s) = s + \rho, \quad \mod 2\pi \tag{31}$$

and ρ is the rotation number. The map T_ρ is, of course, the same as T_C, which we discussed earlier in Example 4,[19] where we showed that it preserves the circular Lebesgue measure. Equation (31), in the irrational case, is the prototype of models generating nonperiodic, nonchaotic dynamics. It can be promptly established that (31) is isomorphic—via the map $\theta : [0, 1) \to S^1, \theta(x) = e^{i2\pi x}$—to the transformation of the interval

$$\begin{gathered} T_I : [0,1) \to [0,1), \\ T_I(x) = x + \alpha, \qquad \mod 1, \qquad \alpha \text{ irrational}, \end{gathered} \tag{32}$$

which, of course, preserves the Lebesgue measure m.

From a measure-theoretic point of view, the two maps (31) and (32) are in fact the same. In particular, they have the same ergodicity properties and the same entropy. Given their extreme simplicity, those transformations can be studied thoroughly and the discussion has a great pedagogic value for understanding the issues of complexity and predictability.

For this purpose, we can use two standard results of ergodic theory and state the following:

PROPOSITION 1. *In the irrational case, the transformation* (31) *is ergodic* (*with respect to the invariant measure* $\hat{m}$).

For a proof see, for example, Doob (1994, pp. 120–121).

PROPOSITION 2. *Transformation* (31) *has zero entropy.*

For a proof, see Mañé (1987, p. 222).

Then, because of isomorphism, Proposition 3 follows.

PROPOSITION 3. *Transformation* (32) *is ergodic* (*with respect to the Lebesgue measure*) *and has zero entropy.*

The results, very interesting in themselves, also help to clarify an element of confusion in the economic literature. Several recent articles in economic journals make use of the concept of *ergodic chaos* (EC) to identify complex, or chaotic, dynamics [see, e.g., Grandmont (1988), Boldrin (1989), and Bala and Majumdar (1992)]. In these works, EC is characterized by the existence of a unique, invariant probability measure that is absolutely continuous and by the property that we have defined above as ergodicity [see, e.g., the definition used by Bala and Majumdar (1992, pp. 439–440)]. The reader can verify easily that maps (31) and (32)—in the irrational case—comply with all of the requirements for EC. Starting from an arbitrarily given initial point on the relevant set, their orbits move over the set, filling it up, without ever returning to the same point in finite time. However, there is nothing particularly complex about their dynamics which, as we have just demonstrated, are in fact perfectly predictable. If economic time series were chaotic in this sense, there would be little trouble in forecasting their future values!

8.2. Chaotic Dynamics

We now contrast the preceding example with another, equally simple, transformation of the interval, namely, the celebrated and already discussed logistic map:

$$\begin{gathered} T_L : [0,1] \to [0,1] \\ T_L(x) = rx(1-x) \qquad 1 \leq r \leq 4. \end{gathered} \tag{33}$$

Map (33) and other analogous one-dimensional, "one-hump" maps often have been used as mathematical idealizations of problems arising in economics including, among others, macroeconomic models [e.g, Stutzer (1980); Day (1982)]; models of rational consumption [e.g., Benhabib and Day (1981)]; models of overlapping generations [e.g., Benhabib and Day (1982), Grandmont (1985)]; models of optimal growth [e.g., Deneckere and Pelikan (1986)]. Recent overviews of the matter with further instances of one-hump functions derived from economic problems can be found in Baumol and Benhabib (1989), Lorenz (1989), Boldrin and Woodford (1990), and Scheinkman (1990). For a continuous-time generalization of one-dimensional maps, see Invernizzi and Medio (1991).

It is known that for system (33) there exists a nonnegligible (positive Lebesgue measure) set of values of the parameter r for which the dynamics of (33) are chaotic. This, in particular, is true for $r = 4$. The behavior of the logistic map in this case can be better studied by considering an even simpler map, the tent map, which we have already discussed above. This is one of the very few transformations for which exact results are available and, in the preceding sections, we could evaluate exactly its K–S entropy and the (equal) LCE. On the other hand, it can be shown that the logistic (with $r = 4$) and tent maps are isomorphic, and therefore they have the same entropy (and the same Lyapunov exponent).

Recalling Definition 7 and the results established in Section 2, we can state the following:

PROPOSITION 4. *The logistic map, with parameter* $r = 4$, T_L, *and the tent map* T_Λ *are isomorphic.*

Because isomorphism preserves entropy, we can conclude that the logistic map has entropy equal to $\ln 2 > 0$ and its dynamics are therefore unpredictable. Notice that, in this case, the metric entropy and the unique LCE are equal.

For economics, the implications of the results just obtained are puzzling. For example, consider the case in which models of optimal growth give rise to dynamic equations of the logistic type with chaotic parameter [cf. Deneckere and Pelikan (1986)]. The sequences thus generated are optimal in the sense that they solve a problem of intertemporal maximization of rational agents, in an economy satisfying the requirements of competitive equilibrium at each point of time. In the absence of exogenous, random disturbances, along optimal trajectories, agents' expectations are supposed to be always fulfilled, i.e., we assume agents' perfect foresight. Although the latter assumption may be acceptable when the dynamics of the system are simple (e.g., convergence to a steady state or to a periodic orbit), it makes little sense if the dynamics are chaotic, in the sense discussed here. In our case, the information set on which agents base their predictions consists of the observations of past values of the relevant variables. But we have just demonstrated that—if the entropy of the system is positive—knowledge of the infinitely remote past with arbitrarily (but not infinitely) great precision of measurement is not sufficient to forecast future values correctly.

9. BERNOULLI DYNAMICS: DETERMINISTIC VERSUS STOCHASTIC SYSTEMS

But this is not all. In fact, not all unpredictable (positive entropy) systems are equally unpredictable. Whereas zero entropy implies that the dynamics of a system are predictable with regard to any possible finite partition, positive entropy simply means that the system is unpredictable with regards to *at least one* partition. As we see in a moment, however, there exist systems that are totally unpredictable in the sense that they are unpredictable for *any* possible partition. Among the latter, there exist a special class called *Bernoulli*, which is the most chaotic, or the least predictable, of all. As we see in a moment, Bernoulli systems are fundamental in at least two ways: first, they are the core of chaotic (positive entropy) systems; and, second, the output of deterministic Bernoulli systems cannot be distinguished from that of certain stochastic processes. In what follows, we try to make these statements more precise.

The first step in our reasoning is to provide a unified characterization of *abstract dynamical systems*, which will permit us to discuss the comparison between random processes and deterministic chaotic systems rigorously and effectively.

Consider a stochastic process $\mathcal{M}$, defined on some probability space $(\Omega, \mathcal{F}, \mu)$ and taking values in a partitioned metric space M. The distribution of the process is defined by the measure on $X \equiv M^{\mathbf{Z}}$ (the space of all possible sample paths $\{x_n\}_{-\infty}^{\infty}$), given by the image of μ under the mapping $\omega \mapsto \{x(\omega)\}_{-\infty}^{\infty}$. If $\mathcal{M}$ is stationary, the shift map σ on X, $x_n(\omega) = x_{n-1}(\sigma(\omega))$, preserves the probability measure μ. We can then define a measure-preserving dynamical system (X, σ, μ), as well as a function $\mathcal{P}$ on X, $\mathcal{P}(\omega) = x_0(\omega)$. The latter defines a partition of X, a cell of which is the subset of X such that x_0 (i.e., the value of x at time zero) belongs to a given cell of M. $\mathcal{P}$ is a function from X to M and can be thought of as the result of an (finitely precise) observation on X.

The approach sketched in the preceding paragraphs provides a unified treatment of "concrete" deterministic systems (i.e., systems observed with finite precision) and (stationary) stochastic processes, both of which then can be looked at as abstract dynamical systems whose realizations can be compared meaningfully. Consider, for example, a deterministic system $\bar{\mathcal{M}}$, whose evolution on a state space is observed with finite precision—i.e., the system is endowed with an observation function (partition) $\bar{\mathcal{P}}$. If the ranges of the functions (partitions) $\mathcal{P}$ and $\bar{\mathcal{P}}$ are defined on the same metric space M, realizations of the stochastic processes $\mathcal{M}$ and trajectories of the (concrete) deterministic system $\bar{\mathcal{M}}$ can be compared and their distance on M evaluated.

The conceptual apparatus of abstract dynamical systems can be used to discuss *Bernoulli processes*, i.e., finite-valued stationary processes with independent and identically distributed terms, which play a crucial role in probability theory and the theory of stochastic processes. In this context, Bernoulli processes are defined as processes whose model is a *Bernoulli shift on an arbitrary partition*. The latter can be characterized as follows: Consider again an abstract dynamical system $(X, \sigma, \mu, \mathcal{P})$, where the notation is the same as before. The term "cylinder" (or sometimes "thin cylinder") is used to define a subset of X consisting of all sequences having specified entries for a finite number of specified coordinates. In mathematical terms a cylinder is denoted thus:

$$\{\omega : x_l(\omega) = m_l \quad n \le l < n + k\},$$

where x_l is the lth coordinate of an element of X and m_l is an element of the partitioned state space M. Now, let the σ-invariant probability measure on a cylinder have the following property:

$$\mu\{\omega : x_l(\omega) = m_l \quad n \le l < n + k\} = \prod_{l=n}^{n+k-1} \mu(m_l).$$

Considering the definition of σ and $\mathcal{P}$ given above, this can be written equivalently as

$$\mu\{\omega : \mathcal{P}(\sigma^l \omega) = m_l, \quad n \le l < n + k\} = \prod_{l=n}^{n+k-1} \mu(m_l),$$

where σ^l is the lth iterate of σ. The process defined by $(X, \sigma, \mu, \mathcal{P})$ is called a *Bernoulli shift*. Correspondingly, a flow ϕ_t will be called a *Bernoulli flow* if and only if the corresponding time-one-map ϕ_1 is Bernoulli.

Analogous definitions are used for one-sided sequences. In this case, if the terms of a sequence are ordered with subscripts increasing from left to right, the map σ acts by shifting the sequence one step to the left and dropping the element with the lowest subscript.

The simplest example of an experiment whose representation is a Bernoulli process is given by a repeated tossing of a fair coin, with $1/2$ probability of heads, and $1/2$ probability of tails. In this case, the probability of any *given* n-sequence of heads and tails is equal to $(1/2)^n$. The process is commonly denoted by $B(1/2, 1/2)$.

Processes that are isomorphic to a Bernoulli process are called *B-processes*. They include both deterministic (discrete- or continuous-time) dynamical systems [e.g., the logistic map (for $r = 4$), the tent map, the Lorenz geometric model, the

https://doi.org/10.1017/S1365100599010044 Published online by Cambridge University Press

Šilnikov model] and stochastic processes (e.g., the i.i.d. processes, the continuous-time Markov processes). In what follows, the phrase Bernoulli system is used broadly when we refer to a deterministic system.

A detailed, rigorous treatment of the theoretical issues underlying the study of Bernoulli processes is totally out of the question here. We only recall a few results that are directly relevant to our present discussion, omitting the proofs. The interested reader can find excellent discussions of the main topics and proofs of the most important results in Ornstein (1974) and Ornstein and Weiss (1991), from which we have drawn heavily.

We first need the following preliminary definition.

DEFINITION 8. *Let $T_1 : X \to X$ and $T_2 : Y \to Y$ be two transformations of measurable spaces X and Y preserving the measures μ_1 and μ_2, respectively. If there exists a map $\theta : Y \to X$ such that (i) $\mu_2(\theta^{-1}(A)) = \mu_1(A)$ for all measurable subsets of X (except possibly sets of measure zero) and (ii) $\theta \circ T_2 = T_1 \circ \theta$ almost everywhere, then T_1 is said to be a* factor *of T_2.*

Then, we state the following propositions.

PROPOSITION 5 (Sinai). *If the entropy of a dynamical system is positive (i.e., if a system is not entirely predictable), then at least one partition of the state space exists such that the system with respect to that partition is isomorphic to a Bernoulli shift. This property can be described formally by saying that all systems that are not entirely predictable have a Bernoulli shift as a factor.*

Proposition 5, together with Definition 8, says that, if a dynamical system described by the transformation T_2 has positive entropy, then we can find a map θ and a corresponding system T_1 such that T_1 is Bernoulli. Of course if θ is one-to-one and invertible, then T_1 and T_2 are isomorphic. In this case, if T_2 is Bernoulli, so is T_1 (and vice versa).

PROPOSITION 6 (Sinai, Ornstein). *Bernoulli systems have only Bernoulli factors and their dynamics are equally unpredictable on any partition of their state spaces.*

Propositions 5 and 6 imply that, although not all chaotic systems are Bernoulli, the mechanism described by Bernoulli shifts plays an essential role in the generation of chaos and unpredictability.[20]

PROPOSITION 7 (Ornstein). *Two finite-entropy B-processes are isomorphic if and only if they have the same entropy.*

Recalling that the entropy of a flow ϕ_t is the same as that of its time-one map ϕ_1 and considering that, for a given map T, the entropy of the tth iterate, $h(T^t) = |t|h(T)$, we can conclude that there exists a unique finite-entropy Bernoulli flow, up to time rescaling and isomorphism.

We now can show that the tent map (and consequently the logistic map with $r = 4$) is not only chaotic in the sense of positive entropy, but is also Bernoulli.

To prove our assertion, we first state the following:

PROPOSITION 8. *The dyadic map*

$$T_D : [0, 1) \to [0, 1)$$

$$T_D(x) = 2x \mod 1$$

is isomorphic to $B(1/2, 1/2)$. *Hence it is Bernoulli.*

Proof. Making use of the binary notation, i.e., putting $x = \sum_{i=1}^{\infty} a_i 2^{-i}$, ($a_i = 0$ or 1), we can establish a map between the interval [0, 1) and the space Σ^2 whose elements are one-sided, infinite sequences of the symbols {0, 1}; thus

$$.a_1, a_2, a_3, \ldots \leftrightarrow (a_1, a_2, a_3, \ldots).$$

[To make the mapping single-valued, we introduce the convention that, at the dyadic values $2^{-1}, 2^{-2} \ldots, 2^{-n}, \ldots,$ the corresponding sequences will end with infinite sequences of ones rather than zeros. For example, we postulate that $0.5 \to 0.0111\ldots$ (an infinite sequence of ones follows), rather than $0.5 \to 0.1000\ldots$ (an infinite sequence of zeros follows).]

Next, consider that (a) The action of T_D on [0, 1) is essentially the same as that of the one-sided Bernoulli shift σ on Σ^2. If we denote by a'_k a shifted symbol, in both cases we have $a'_k = a_{k+1}$, and a_1 is discarded. (b) The measures of sets of points in [0, 1) and that of corresponding sets of sequences in Σ^2 are the same if for the former we use the Lebesgue measure and for the latter the product measure on cylinders defined earlier in this section. For example, the Lebesgue measure of the subinterval [0, 1/4] is 0.25 and so is the measure of the corresponding cylinder $\{a : a_i = 0,\ i = 1, 2\}$, i.e., we have $\mu\{a : a_i = 0,\ i = 1, 2\} = \mu(a_1 = 0)\mu(a_2 = 0) = (0.5)(0.5) = 0.25$.

PROPOSITION 9. *The tent map* T_Λ *is a factor of the dyadic map* T_D.

Proof. By putting $\theta = T_\Lambda : [0, 1] \hookleftarrow$, we can verify that, for m-almost all points of the interval, we have $T_\Lambda \circ \theta = \theta \circ T_D$. Because both T_D and T_Λ preserve the Lebesgue measure m, the measure structure is preserved by $\theta = T_\Lambda$, i.e., for all subintervals of [0, 1], except perhaps sets of measure zero, we have $m(T_\Lambda^{-1}(I)) = m(I)$. Notice, however, that T_Λ is a two-to-one map and therefore Proposition 9 does not necessarily imply that T_D and T_Λ are isomorphic.

PROPOSITION 10. T_Λ *is Bernoulli.*

Proof. Proposition 10 follows from Propositions 6, 8, and 9.

PROPOSITION 11. T_L *is Bernoulli.*

Proof. This follows from Propositions 4 and 10.

PROPOSITION 12. *Maps T_L (with $r = 4$), T_Λ, T_D and the Bernoulli process $B(1/2, 1/2)$ all have the same entropy* $\ln 2$.

Proof. In view of the propositions already proved, and the fact that isomorphism preserves entropy, we only need to prove that the entropy $h(T_L) = h(T_\Lambda)$ is equal to $h(T_D)$. This is easily proved because exactly the same reasoning that we have used to prove that the entropy of the tent map is equal to ln 2 can be applied to the dyadic map, with the same result.

In conclusion, from a measure-theoretic point of view, the logistic map (with parameter $r = 4$), the dyadic map, the tent map, and the process $B(1/2, 1/2)$ are the same. This makes the puzzling problem we mentioned at the end of Section 8 even more puzzling. Indeed, when the map governing the system belongs to the Bernoulli class, economic agents face sequences of values of the variables that have the same probability structure as that of a random process. To postulate perfect foresight in this situation is clearly absurd.

9.1. Distinguishing Deterministic Chaos and Randomness: α-Congruence

Equivalence implicit in isomorphism concerns the probabilistic structure of orbits but not their geometry, much of which can be distorted by the map that relates two isomorphic spaces. Therefore, geometrically different systems can be isomorphic. To overcome this difficulty, the notion of α-congruence has been suggested. The following definition, which is based on Ornstein and Weiss (1991, pp. 22–23, 63), is formulated in the more general terms of flows, i.e., continuous-time systems, and can be adapted easily to discrete-time systems.

DEFINITION 9. *Consider two flows ϕ_t and $\bar{\phi}_t$, defined on the same metric space M and preserving the measures μ and $\bar{\mu}$, respectively. We say that ϕ_t and $\bar{\phi}_t$ are α-congruent if they are isomorphic via a map $\theta : M \to M$ that satisfies*

$$\mu\{x : d(x, \theta(x)) > \alpha\} < \alpha,$$

where $x \in M$ and d is a fixed metric in M.

That is to say, the isomorphism θ moves points in M by less than α, except for a set of points in M of measure less than α. Thus if α is so small that (1) we do not appreciate distances in M smaller than α and (2) we ignore events whose probability is smaller than α, then we consider α-congruent systems as actually indistinguishable. Thus, α-congruence is a form of equivalence stronger than isomorphism because it require that systems be close not only in a measure-theoretic sense but also in a geometric one.

The following results, recently established for α-congruent systems are particularly relevant to the present discussion [proofs in Radunskaya (1992)].

PROPOSITION 13 (Radunskaya). *Let ϕ_t be a flow on a manifold M that is isomorphic to the Bernoulli flow of infinite entropy. Then, for any $\alpha > 0$, there is a continuous-time, finite-state Markov process $\mathcal{M}_t$ taking measure on M, which is α-congruent to ϕ_t.*

https://doi.org/10.1017/S1365100599010044 Published online by Cambridge University Press

This is a remarkable result but is not sufficient for our present purposes. In the study of deterministic chaotic systems, we usually consider the dynamics of the system on a compact attractor whose entropy is finite and bounded by the sum of the positive LCE's. Therefore, we are interested here in flows with finite entropy. The following result is then more important:

PROPOSITION 14 (Radunskaya). *Let ϕ_t be a B-flow of finite entropy on a manifold M. Let B_t^∞ be an infinite entropy B-flow on a probability space Ω. Then, for any $\alpha > 0$, there exists a continuous-time Markov process $\mathcal{M}_t$ on a finite number of states $\{s_i\} \in M \times \Omega$ such $\mathcal{M}_t$ is α-congruent to $\bar{\phi}_t = \phi_t \times B_t^\infty$.*

These rather abstract results can be given a striking commonsense interpretation. Let us consider an observer looking at orbits generated by a deterministic Bernoulli system with finite entropy and let us suppose that observation takes place through a device (a "viewer") that distorts by less than α, with probability greater than $1 - \alpha$, where α is a positive number that we can choose as small as we wish. Proposition 13 (infinite-entropy case) tells us that the orbits as seen through the viewer are arbitrarily close to a continuous-time, finite-state Markov process. In the finite-entropy case, to compare the orbits of the deterministic system with the sample paths of the (infinite-entropy) Markov process, we need to introduce some additional entropy by "sprinkling" the deterministic system with a bit of noninterfering noise. We can again use the parable of orbits observed through a slightly distorting viewer, but now the errors are random. Proposition 14 tells us that in this case too the observed orbits are most of the time arbitrarily close to the sample paths of the Markov process.

These sharp results should produce some skepticism on the possibility of rigorously testing—e.g., by estimating the value of correlation dimension or the dominant LCE—whether a given series has been generated by a deterministic or a stochastic mechanism. In view of Propositions 13 and 14, if those tests are applied to a Markov process and to a deterministic Bernoulli system, which are α-congruent, they should give the same results for sufficiently small values of α.

Because we are concerned here only with the general consequences of the results above, we do not discuss them in any detail, but only mention a simple example of α-congruence between a discrete-time deterministic system and a stochastic process, once again making use of the tent map that we know to be Bernoulli. For this map and for any given partition of the state space $I = [0, 1]$, we can define a Markov chain on k states (the number of states depending on the partition), such that its sample paths are indistinguishable (within the prescribed resolution, or precision of observation) from the orbits of the deterministic tent map.[21] For example, if the degree of accuracy α is equal to $1/2$—i.e., we can only tell whether the state of the system is on the left or on the right of the middle point of the state space [0, 1]—then a Markov chain on two states L and R, with transition matrix

$$\begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

https://doi.org/10.1017/S1365100599010044 Published online by Cambridge University Press

will generate sample paths indistinguishable (within the prescribed degree of accuracy) from those of the map T_Λ.

If we now increase the precision of observation and put $\alpha = 2^{-k}$, $k > 1$ (i.e., we can locate a point in the state space within subintervals of length 2^{-k}), we can construct a 2^k state Markov chain with transition matrix

$$\begin{bmatrix} 1/2 & 1/2 & 0 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & \cdots & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & \cdots & 1/2 & 1/2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 1/2 & 1/2 & \cdots & \cdots & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & \cdots & \cdots & 0 & 0 \end{bmatrix}.$$

Again it will not be possible (within the precision α) to distinguish between a sample path of the Markov chain and a typical orbit of T_Λ on the k-partitioned state space.

Our considerations so far can be given a more or less optimistic interpretation according to one's point of view and temperament. The results discussed above indicate that we cannot hope to provide a generally valid test for distinguishing deterministic chaos and true randomness. This would certainly be impossible for Bernoulli systems and, to the extent that the conjecture that "most observable chaos is Bernoulli"[22] is correct, it would be generally impossible. Consequently, at least for a certain class of concrete dynamical systems, the possibility exists of representing them either as deterministic systems (plus perhaps some random, noninterfering disturbances) or as stochastic processes. The choice between the two is a matter of expedience rather than theory. In principle, a deterministic representation is superior for explanation purposes, but this is only true if the resulting model is sufficiently simple and we can provide a physical (economic) interpretation of the state variables and the functional relationships among them.

In the circumstances, we share the view of the meteorologists Vautard and Ghil (1989, p. 395), who aptly observed that

> The right question . . . is not whether a given time series is of purely deterministic or purely stochastic origin, i.e., whether very few or very many d-o-f [degrees of freedom] have interacted to produce it. It is rather, how much and what part of the observed variability is due to a few d-o-f and what part to the infinite rest. The former part can presumably be modeled deterministically, and analyzed therewith rather completely; the latter has still to be relinquished to the obscure kingdom of means and variances.

NOTES

1. For a set $A \in \mathcal{F}$ the expression $T^{-1}(A)$ denotes the *preimage* of A, i.e., the set of points that are mapped to A by the map T.

2. This, of course, means that the set of points of X for which the limit (5) does not exist is negligible in the sense that the invariant measure μ assigns zero value to that set.

3. Notice that, although in the literature one commonly encounters the expression ergodic map, this only makes sense if it is clear with respect to what invariant measure the map is indeed ergodic.

4. As we shall see, however, the set of points for which (6) does *not* hold may well include "almost every point" with respect to another measure, in particular with respect to the Lebesgue measure.

5. This definition of Lebesgue measure is rather coarse but intuitively clear and sufficient for our present purpose. The interested reader can find a more rigorous treatment of the matter, for example, in Cohen (1980) and Edgar (1990).

6. The reader will remember that $e^{i2\pi x} = (\cos 2\pi x + i \sin 2\pi x)$. So, for any value of $x \in [0, 1)$, the function $\theta(x)$ identifies a point on the unit circle and, for any subinterval $I \subset [0, 1)$, $\theta(I)$ identifies an arc $\Gamma \subset S^1$. Accordingly, the inverse function θ^{-1} maps arcs into intervals.

7. The interested reader will find a thorough discussion of the analytical and numerical questions concerning the LCEs in Benettin *et al.* (1980).

8. A rigorous definition of SBR measures can be found in Eckmann and Ruelle (1985, pp. 639–641).

9. Once again, the reason that we iterate T backward is to deal with noninvertible maps.

10. This is the celebrated formula of Shannon, the founder of the modern theory of information. It is not the only conceivable way of measuring uncertainty (or information). However, it complies with certain generally accepted axioms of information theory formulated by Khinchin [see Beck and Schlögl (1993, pp. 47–49)].

11. For a proof that this limit exists, see Billingsley (1965, pp. 81–82); Mañé (1987, p. 216).

12. When we want to emphasize the map T (and the relevant invariant measure is known), we also can denote the entropy of the system as $h(T)$.

13. For discussion and proof of the K-S theorem, see Billinglsey (*op. cit.*, pp. 84–85); Mañé (*op. cit.*, pp. 218–222).

14. For technical details, see Ornstein and Weiss (1991, pp. 78–85); see also Ruelle (1989, pp. 71–77).

15. For a recent analysis of a case relevant to our discussion, see Venditti (1996).

16. Cf., for example, Reichlin (1986), Hommes (1991), and Medio (1992).

17. Cf., for example, Hommes (1991).

18. The relevant theorem is due to Denjoy [cf. Iooss (1979, pp. 48–49); Anosov and Arnold (1988, pp. 43–44) and Whitley (1983, p. 200)].

19. In (31), s measures the angle formed by the segment joining the origin and a point on S^1, and the abscissa.

20. There exists an intermediate class of transformations, called *Kolmogorov* (K) *transformations*, that are unpredictable for any possible partition. Although all Bernoulli transformations are K, the reverse need not be true [cf. Ornstein and Weiss (1991, pp. 19–20)]. K transformations are seldom met in applications (at least in economics) and we do not deal with them here.

21. I owe this example to a private correspondence with Amy Radunskaya.

22. See Ornstein and Weiss (1991, p. 22).

REFERENCES

Anosov, D.V. & V.I. Arnold, eds. (1988) *Dynamical Systems I*. Berlin: Springer-Verlag.

Bala, V. & M. Majumdar (1992) Chaotic tatonnement. *Economic Theory* 2, 437–445.

Baumol, W.J. & J. Benhabib (1989) Chaos: Significance, mechanism, and economic applications. *Journal of Economic Perspectives* 3, 77–107.

Beck, C. & F. Schlögl (1993) *Thermodynamics of Chaotic Systems*. Cambridge, UK: Cambridge University Press.

Benettin, G., L. Galgani, A. Giorgilli & J.-M. Strelcyn (1980) Lyapunov characteristic exponents for smooth dynamical systems and for Hamiltonian systems: A method for computing all of them, Part I: Theory. *Meccanica* 15, 9–20.

Benhabib, J. & R.H. Day (1981) Rational choice and erratic behavior. *Review of Economic Studies* 48, 459–471.

Benhabib, J. & R.H. Day (1982) A characterization of erratic dynamics in the overlapping generations models. *Journal of Economic Dynamics and Control* 4, 37–55.

Billingsley, P. (1965) *Ergodic Theory and Information*. New York: John Wiley and Sons.

Boldrin, M. (1989) Path of optimal accumulation in two-sector models. In W.A. Barnett, J. Geweke & K. Shell (eds.), *Economic Complexity: Chaos, Sunspots, Bubbles, and Nonlinearity*. Cambridge, UK: Cambridge University Press.

Boldrin, M. & M. Woodford (1990) Equilibrium models displaying endogenous fluctuations and chaos: A survey. *Journal of Monetary Economics* 25, 189–223.

Bullard, J. & A. Butler (1993) Nonlinearity and chaos in economic models: Implications for policy decisions. *Economic Journal* 103, 849–867.

Cohen, D.L. (1980) *Measure Theory*. Boston: Birkhäuser.

Day, R.H. (1982) Irregular growth cycles. *American Economic Review* 72, 406–414.

Deneckere, R. & S. Pelikan (1986) Competitive chaos. *Journal of Economic Theory* 40, 13–25.

Doob, J.L. (1994) *Measure Theory*. New York: Springer-Verlag.

Eckmann, J.-P. & D. Ruelle (1985) Ergodic theory of chaos and strange attractors. *Reviews of Modern Physics* 57, 617–656.

Edgar, G. (1990) *Measure, Topology and Fractal Geometry*. Berlin: Springer-Verlag.

Grandmont, J.M. (1985) On endogenous competetive business cycles. *Econometrica* 53, 995–1045.

Grandmont, J.M. (1988) Nonlinear Difference Equations Bifurcations and Chaos: An Introduction. Working paper 8811, Centre d'Étude Prospectives d'Économie Mathématique Appliquée à la Planification.

Hommes, C.H. (1991) *Chaotic Dynamics in Economic Models. Some Simple Case-Studies*. Groningen: Wolters-Noordhoff.

Invernizzi, S. & A. Medio (1991) On lags and chaos in economic dynamic models. *Journal of Mathematical Economics* 29, 521–550.

Iooss, G. (1979) *Bifurcations of Maps and Applications*. Amsterdam: North-Holland.

Jackson, E.A. (1990) *Perspectives of Nonlinear Dynamics,* 2 Vols. Cambridge, UK: Cambridge University Press.

Lorenz, H.W. (1989) *Nonlinear Dynamical Economics and Chaotic Motion*. Berlin: Springer-Verlag.

MacKay, R.S. (1992) Nonlinear Dynamics in Economics: A Review of Some Key Features of Nonlinear Dynamical Systems. Manuscript, European University Institute.

Mañé, R. (1987) *Ergodic Theory and Differentiable Dynamics*. Berlin: Springer-Verlag.

Medio, A. (1992) *Chaotic Dynamics. Theory and Applications to Economics*. Cambridge, UK: Cambridge University Press.

Milnor, J. (1985) On the concept of attractor. *Communications of Mathematical Physics* 99, 177–195.

Neimark, J. (1959) On some cases of periodic motions depending on parameters. *Doklady Akademii Nauk SSSR*, 736–739.

Ornstein, D.S. (1974) *Ergodic Theory, Randomness, and Dynamical Systems*, Yale Mathematical Monographs 5. New Haven and London: Yale University Press.

Ornstein, D.S. & B. Weiss (1991) Statistical properties of chaotic systems. *Bulletin (New Series) of the American Mathematical Society* 24, 11–115.

Palis, J. & F. Takens (1993) *Hyperbolicity & Sensitive Chaotic Dynamics at Homoclinic Bifurcations*. Cambridge, UK: Cambridge University Press.

Radunskaya, A. (1992) *Alpha-Congruence of Bernoulli Flows and Markov Processes*: *Distinguishing Random and Deterministic Chaos.* Ph.D. Dissertation, Stanford University.

Reichlin, P. (1986) Equilibrium cycles in an overlapping generations economy with production. *Journal of Economic Theory* 40, 89–102.

Ruelle, D. (1989) *Chaotic Evolution and Strange Attractors*. Cambridge, UK: Cambridge University Press.

Sacker, R.J. (1964) On Invariant Surfaces and Bifurcation of Periodic Solutions of Ordinary Differential Equations. IMM-NYU 333, New York University.

Stutzer, M.J. (1980) Chaotic dynamics and bifurcation in a macro-model. *Journal of Economic Dynamics and Control* 2, 353–376.
Vautard, R. & M. Ghil (1989) Singular spectrum analysis in nonlinear dynamics with applications to paleoclimatic time series. *Physica* 35B, 395–424.
Venditti, A. (1996) Hopf bifurcation and quasi-periodic dynamics in discrete multisector optimal growth models. *Research in Economics/Ricerche Economiche* 50, 267–291.
Whitley, D. (1983) Discrete dynamical systems in dimensions one and two. *Bulletin of the London Mathematical Society* 15, 177–217.