Statistical Science
1992, Vol. 7, No. 1, 69–122

Statistics, Probability and Chaos

L. Mark Berliner

Abstract. The study of chaotic behavior has received substantial attention in many disciplines. Although often based on deterministic models, chaos is associated with complex, "random" behavior and forms of unpredictability. Mathematical models and definitions associated with chaos are reviewed. The relationship between the mathematics of chaos and probabilistic notions, including ergodic theory and uncertainty modeling, are emphasized. Popular data analytic methods appearing in the literature are discussed. A major goal of this article is to present some indications of how probability modelers and statisticians can contribute to analyses involving chaos.

Key words and phrases: Dynamical systems, ergodic theory, nonlinear time series, stationary processes, prediction.

1. INTRODUCTION

Chaos is associated with complex and unpredictable behavior of phenomena over time. Such behavior can arise in deterministic dynamical systems. Many examples are based on mathematical models for (discrete) time series in which, after starting from some initial condition, the value of the series at any time is a specified, nonlinear function of the previous value. (Continuous time processes are discussed in Section 2.) These processes are intriguing in that the realizations corresponding to different, although extremely close, initial conditions typically diverge. The practical implication of this phenomenon is that, despite the underlying determinism, we cannot predict, with any reasonable precision, the values of the process for large time values; even the slightest error in specifying the initial condition eventually ruins our attempt. Later in this article, indications that realizations of such dynamical systems can display characteristics typically associated with randomness are presented. A major theme of this study is that this connection with randomness suggests that statistical reasoning may play a crucial role in the analysis of chaos.

Strong interest has recently been shown in numerous literatures in the areas of nonlinear dynamical systems and chaos. However, rather than attempting to provide an overview of the applications of chaos, I offer a review of the basic notions of chaos with emphasis on those aspects of particular interest to statisticians and probabilists. Many of the references given here provide indications of the breath of interest in chaos. Jackson (1989) provides an introduction and an extensive bibliography [also, see Shiraiwa (1985)]. Berge, Pomeau and Vidal (1984), Cooper (1989) and Rasband (1990) discuss applications of chaos in the physical sciences and engineering. Valuable sources for work on chaos in biological and medical science include May (1987), Glass and Mackey (1988) and Basar (1990). Wegman (1988) and Chatterjee and Yilmaz (1992) present reviews of particular interest to statisticians. Finally, useful, "general audience" introductions to chaos include Crutchfield, Farmer, Packard and Shaw (1986), Gleick (1987), Peterson (1988) and Stewart (1989).

Section 2 presents discussion of the standard mathematical setup of nonlinear dynamical systems. Definitions of chaos are reviewed and chaotic behavior is explained mathematically, as well as by example. Next, I review relationships between chaos and probability. Two key points in this discussion are: (i) the role of ergodic theory and (ii) the suggestion of uncertainty modeling and analysis by probabilistic methods. Statistical analyses related to chaos are discussed in Sections 3 and 4. In Section 3, the emphasis is on some "data analytic" methods for analyzing chaotic data. The goals of these techniques basically involve attempts at understanding the structure and qualitative aspects of models and data displaying chaotic behavior. [Specifically, the notions of (i) estimation of dimension, (ii) Poincaré maps and (iii) reconstruction by time delays are reviewed.] Although statisticians

L. Mark Berliner is Associate Professor, Department of Statistics, Ohio State University, 141 Cockins Hall, 1958 Neil Avenue, Columbus, Ohio 43210.

are now beginning to make contributions along these lines, the methods described in Section 3 have been developed primarily by mathematicians and physicists. In Section 4, I discuss some possible strategies for methods of chaotic data analysis based on main stream techniques for statistical modeling and inference. Finally, Section 5 is devoted to general remarks concerning statistics and chaos.

2. MATHEMATICS, PROBABILITY AND CHAOS

2.1 The Complexity of Nonlinear Dynamical Systems

A simple deterministic dynamical system may be defined as follows. For a discrete time index set, $T = \{0, 1, 2, \dots\}$, consider a time series $\{x_t; t \in T\}$. Assume that x_0 is an initial condition and that $x_{t+1} = f(x_t)$, for some function f that maps a domain D into D. (D is typically a compact subset of a metric space). Chaotic behavior may arise when f is a nonlinear function.

To begin, some numerical examples for one of the more popular examples of dynamical systems, the logistic map, are given. The dynamical system is obtained by iterating the function $f(x) = ax(1 - x)$, where a is a fixed parameter in the interval [0,4]. Let x_0 be an initial point in the interval [0,1]; note that then all future values of the system also lie in [0,1]. To get a bit of the flavor of this map, example computations are presented for an important value of a: namely, $a = 4.0$. Figure 1 presents time series plots of the first 500 iterates of the logistic map corresponding to the initial values 0.31, 0.310001 and 0.32. The first thing to notice about these series is that their appearance is "complex." Indeed, one might be tempted to suggest these series are "random." Also, despite the similarity in the initial conditions, visual inspection of the series indicates that they are not quantitatively similar. To make the point, I have included scatterplots of these series, matched by time. The first 25 iterates of the maps in these plots are indicated by a different symbol from the rest. Points falling on the 45° line in these plots suggest time values at which the corresponding values of the systems are quite close. We see that quite early in time, the three series "predict" each other reasonably well. However, the similarity in the series diminishes rapidly as time increases. (The rate of this "separation" is in fact exponential in time.) Note that, except for very early times, the series corresponding to $x_0 =$ 0.310001 is really no better at predicting the 0.31 series than is the 0.32 series. Also, predictions based on $x_0 = 0.31$ when the correct value is 0.310001 are very poor, even though the error in specifying the initial condition is in the sixth decimal place. This sort of behavior, known as *sensitivity to initial conditions,* is one of the key components of chaos. To amplify on this phenomenon, the first frame of Figure 2 presents dot plots, at selected time values, of the dynamical system corresponding to 18 initial conditions equally spaced in the interval [0.2340, 0.2357]. There are two messages in this plot. First, note that the images of these 18 points are quickly attracted to the unit interval. Second, the initial conditions appear to get "mixed" up in an almost noncontinuous manner. (However, for the logistic map, x_t is, of course, a continuous function of x_0 for all t). The second frame of Figure 2 is a scatterplot of the values of the logistic map after 2000 iterates against the corresponding initial conditions for 4000 initials equally spaced in the interval [0.10005, 0.3]. There is clearly essentially no meaningful statements about the relationship between x_{2000} and x_0, even though x_{2000} is a well-defined polynomial function (of admittedly high order) of x_0. (Note that presenting this graph for 2000 iterates is a bit of "overkill." Corresponding scatterplots after even a 100 or so iterates would look quite the same.)

2.1.1 Some mathematics for nonlinear dynamical systems

This discussion is intended to provide some flavor of the mathematics concerning the appearance of complex or chaotic behavior in nonlinear dynamical systems. The presentation is a bit quick, and until Section 2.1.3, considers one-dimensional maps only. More complete details may be found in Collet and Eckmann (1980), Rasband (1990) and Devaney (1989). We begin by considering the long-run behavior of a dynamical system generated by a nonlinear function f. The study begins with the consideration of *fixed points* of f; namely, those points that are solutions to $f(x) = x$. The key result in this context is the following proposition. Using conventional notation, let $f^n(\cdot)$ denote the n-fold composition of f.

PROPOSITION 2.1. *Let p be a fixed point of f. If $|f'(p)| < 1$, then there exists an open interval U about p such that, for all x in U, $\lim_{n\to\infty} f^n(x) = p$.*

Under the conditions of this proposition, p is an *attracting* fixed point and the set U is a *stable set.* It is also true that if $|f'(p)| > 1$, p is a repelling fixed point. (In these two cases, p is said to be *hyperbolic*; if $|f'(p)| = 1$, p acts as a saddle point and more delicate analyses are required. The rest of this discussion focuses on hyperbolic points.)

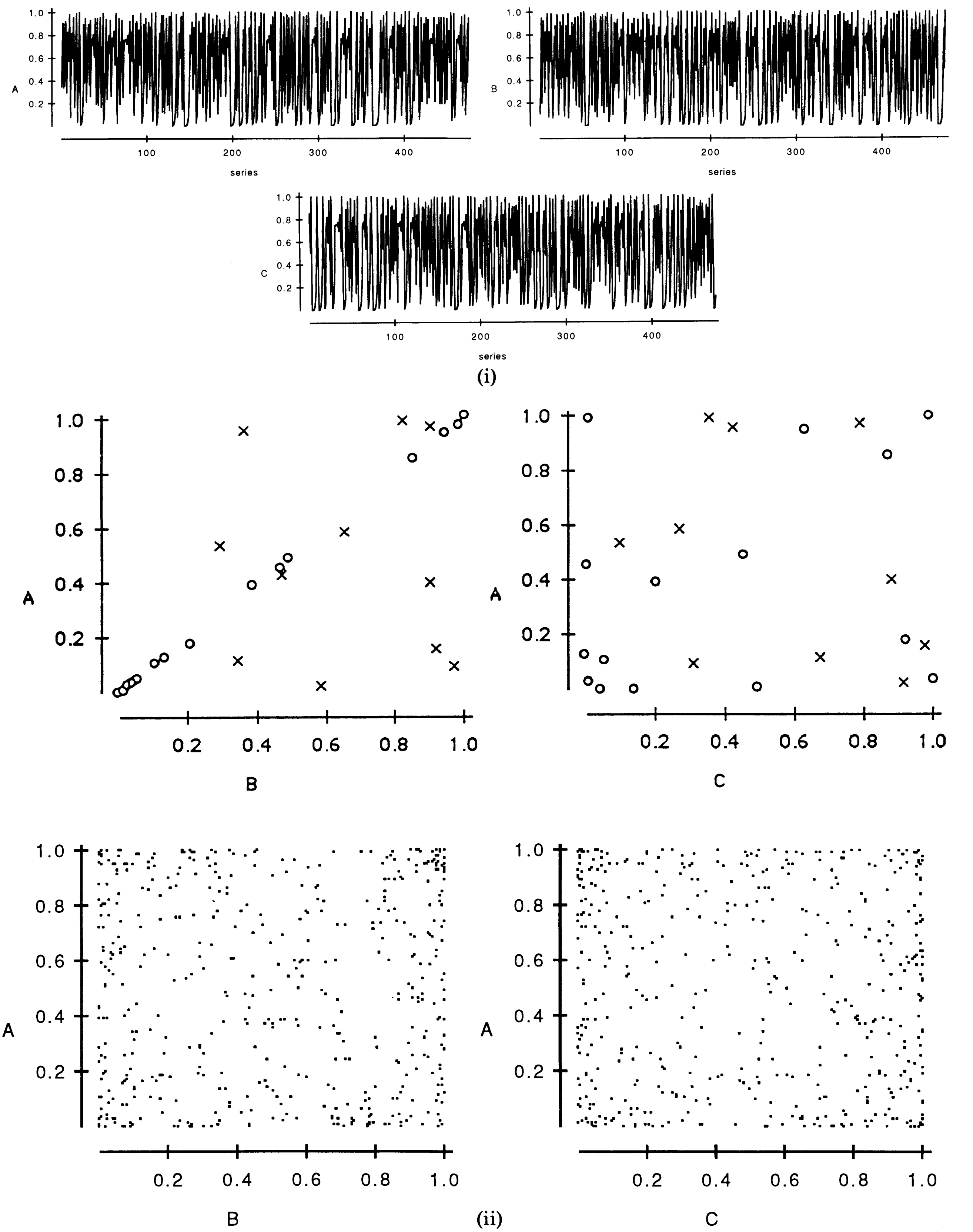
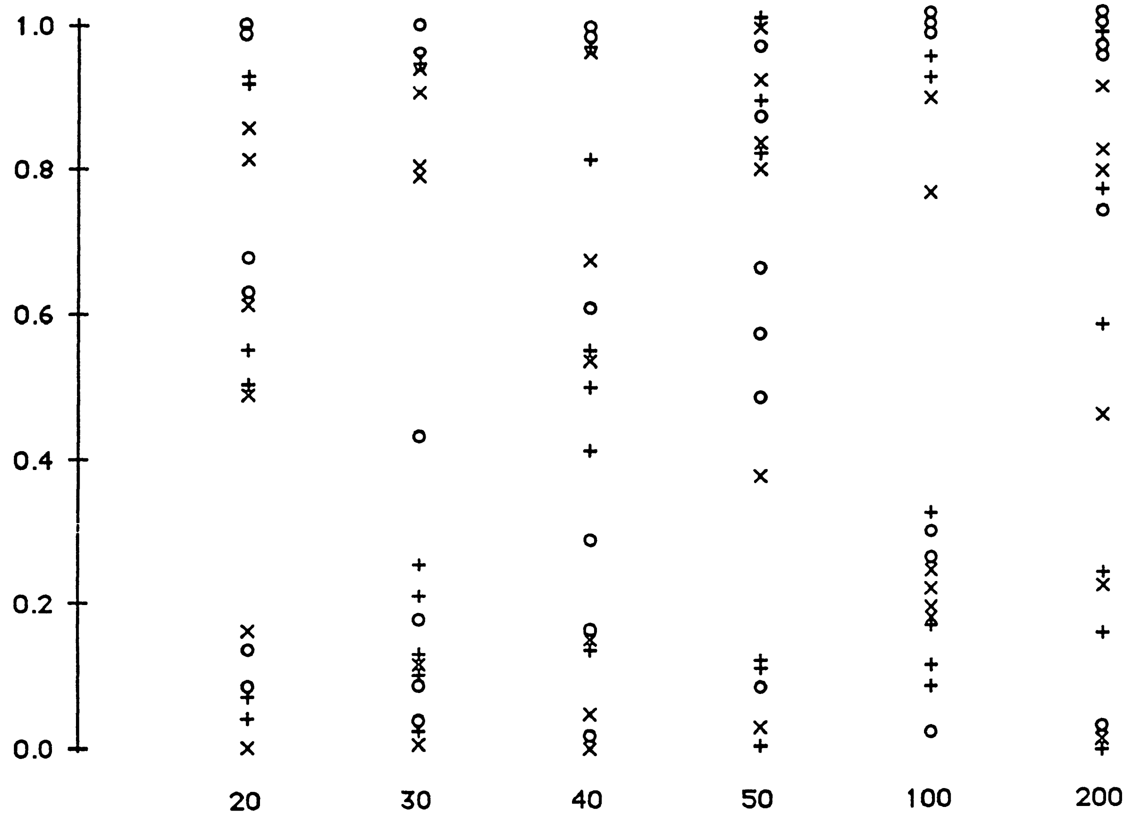
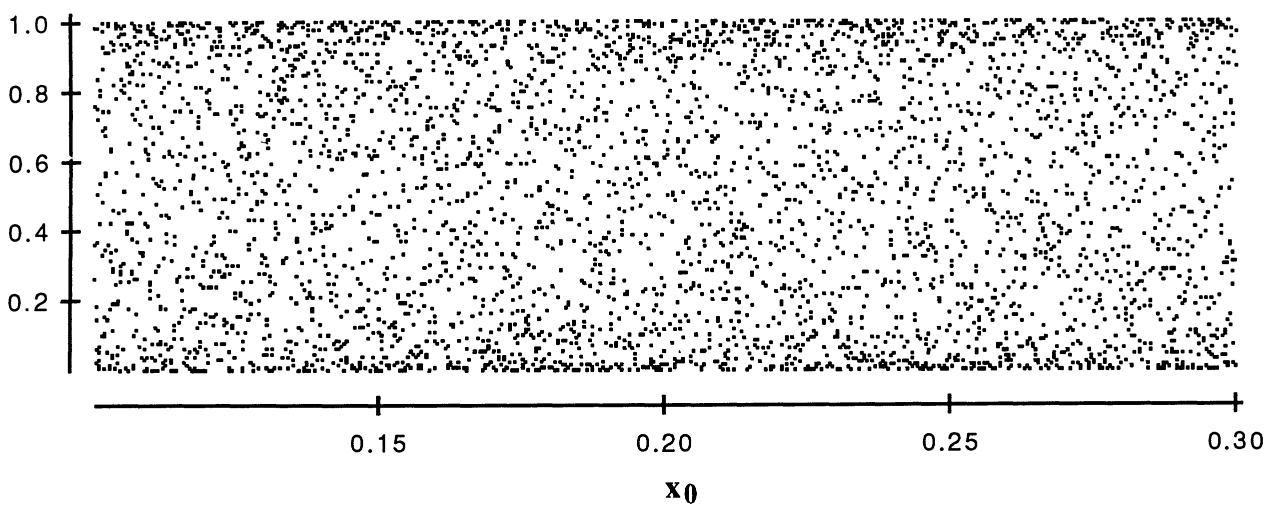


FIG. 1. *Examples of the logistic map: $a = 4.0$. Initial conditions: (A) $x_0 = .31$; (B) $x_0 = .310001$; (C) $x_0 = .32$. (i) Time series plots for times 26 – 500. (ii) Scatterplots matched by time. In top plots, ○ denotes times 1 – 15 and × denotes times 15 – 25; bottom plots show times 26 – 500.*



(i)



(ii)

FIG. 2. *Mixing behavior of the logistic map*: $a = 4.0$. (*i*) *Dot plots for* 18 *initial conditions*. ○ *denotes* $x_0 = .2340, .2341, .2342, .2343, .2344, .2345$; × *denotes* $x_0 = .2346, .2347, .2348, .2349, .2350, .2351$; + *denotes* $x_0 = .2352, .2353, .2354, .2355, .2356, .2357$. (*ii*) *Scatterplot of logistic map at time* 2000 *against* x_0 *for* 4000 x_0*'s in* [.10005, .3].

Consider now the logistic map defined earlier. The fixed points of this map, i.e., solutions to the equation $x = ax(1 - x)$, are easily seen to be 0 and $p_a = 1 - 1/a$. Note that $f'(x) = a(1 - 2x)$. We now consider the behavior of iterates of the logistic for various a.

1. $a < 1$: Because $f'(0) = a < 1$, 0 is an attracting fixed point. Furthermore, the iterates $x_n = f^n(x)$ are monotonically decreasing for all x in (0,1], but bounded from below by zero, and so, since f is continuous, $\lim_{n\to\infty} f^n(x) = 0$. Note that this argument also happens to cover the nonhyperbolic case when $a = 1$.

2. $1 < a < 3$: First, note that now 0 is a repelling

fixed point. The transition from attraction to repulsion occurred as $f'(0)$ increased to one at $a = 1$ and became greater than one for $a > 1$. This is a flag that there is a potential change in behavior. Consider the other fixed point p_a. Note that $|f'(p_a)| = |2 - a| < 1$, and so p_a is an attracting fixed point if $1 < a < 3$. Thus, for all $0 < x < 1$ there is a chance that $\lim_{n\to\infty} f^n(x) = p_a$. This conclusion is already guaranteed for the stable set of p_a. It is not hard to demonstrate that iterates of any point in (0,1) eventually enter the stable set. The conclusion is that for almost all, in the sense of Lebesgue measure, x in [0,1], $\lim_{n\to\infty} f^n(x) = p_a$. The only exceptions are the endpoints 0 and 1.

3. $a > 3$: Now both 0 and p_a are repelling fixed points. Because $f'(p_a) > 1$, p_a can no longer attract iterates of the map except for x's such that there exists an n such that $f^n(x) = p_a$. This set is the collection of *preimages* of p_a. This set is *countable*, because there is a "denumerable" algorithm for finding the preimages: Namely, first find the solutions to $f(x) = p_a$. In this step the solutions are p_a and $1 - p_a$. The next step is to "invert" $1 - p_a$ to obtain two more preimages, etc. Also, note that for all a, the preimages of 0 are 0 and 1, but 1 has no preimages. To see what happens to other points, consider the fixed points of $f^2(x)$. (Fixed points of f^2 are *periodic points of period 2* for f). Now, we must solve

$$a\big(ax(1-x)\big)\big(1 - ax(1-x)\big) - x = 0, \tag{2.1}$$

a fourth degree polynomial. However, we can recognize that both 0 and p_a are solutions. Accounting for these solutions, we can find the remaining two by solving the quadratic

$$\begin{aligned} &a^3x^2 + (p_a - 2)a^3x \\ &\quad + a^3\big(1 + p_a(p_a - 2)\big) + a^2 = 0. \end{aligned} \tag{2.2}$$

The solutions are

$$p_u = \left[a + 1 + \sqrt{a^2 - 2a - 3}\right]/2a \text{ and}$$
$$p_l = \left[a + 1 - \sqrt{a^2 - 2a - 3}\right]/2a.$$

Note that, of course, $f(p_u) = p_l$. First, these roots are real and distinct if $a^2 - 2a - 3 > 0$ or if $a > 3$. At $a = 3$, $p_a = p_u = p_1$. That is, we observe a *bifurcation* at $a = 3$, in which the attracting fixed-point p_a splits into two pieces. Next, we ask for what a are these roots attracting fixed-points of f^2? Note that

$$f^{2\prime}(x) = a^2(1 - 2x)\big(1 - 2ax(1-x)\big),$$

and so $f^{2\prime}(p_u) = f^{2\prime}(p_l) = 1 - (a^2 - 2a - 3)$. Thus, $|1 - (a^2 - 2a - 3)| < 1$ if $3 < a < 1 + \sqrt{6}$. Further, we apply Proposition 1.1 to f^2 to infer the existence of a stable set, U_u, corresponding to p_u such that for all x in U_u, $\lim_{n\to\infty} f^{2n}(x) = p_u$. An analogous claim is true for p_l. A bit more work then yields that for all x (0,1), except for the preimages of 0 and p_a, and for $3 < a < 1 + \sqrt{6}$, x_n is asymptotically attracted to oscillate between p_u and p_l. Note that, theoretically, x_n need not actually ever reach p_u or p_l exactly, although numerical computations typically indicate exact oscillation on this two-point *attractor*.

Because $|f^{2\prime}(p_u)| = 1$ at $a = 1 + \sqrt{6}$ and $|f^{2\prime}(p_u)| > 1$ for $a > 1 + \sqrt{6}$, we (correctly) anticipate another bifurcation at $a = 1 + \sqrt{6}$ in which the period two attractor splits into a period four attractor. However, the analysis based on the analogs of (2.1) and (2.2) for this, and later, bifurcations is no longer feasible. More delicate tools are needed to mathematically describe the behavior. The result is that the period doubling bifurcations of periods 4 to 8, etc., continue as a increases, on to "period 2^∞" at $a_\infty = 3.56994\ldots$. The fundamental mathematics explaining the "period doubling cascade" is due to the physicist, M. Feigenbaum (see Feigenbaum, 1978). For $a > a_\infty$, the asymptotic behavior of the logistic becomes even more complex than period doubling cascade and is still not completely understood. We observe more period doubling as well as aperiodic behavior. For some a x_n appears to wander on a "singular" attractor (i.e., a Cantor set), whereas for other a, particularly, $a = 4$, as x_0 varies, x_n wanders on a "continuous" set. One interesting phenomenon occurs in a small interval of a's near $3.83\ldots$. In this region, a period 3 attractor appears, and then quickly under goes its own period doubling cascade. This fact implies further complications in that, if a is such nontrivial three-cycle behavior is possible, then all other cycles have solutions. That is, if $f^3(x) = x$ has nontrivial solutions, so does $f^n(x) = x$, for all n. This result is a consequence of Sarkovskii's Theorem [see Devaney (1989), Section 1.10]. Also, see Li and Yorke (1975). On the other hand, for each fixed a, there is at most one attracting periodic cycle. That is, except for a set of Lebesgue measure zero, all initial values lead to paths that are attracted to the same periodic cycle, if there is attracting cycle. We have seen a version of this property above: for $3 < a < 1 + \sqrt{6}$, there are nontrivial solutions to both $f(x) = x$ and $f^2(x) = x$. However, only the period 2 cycle is attracting. The period one cycle is only observable for the preimages of p_a and 0.

To illustrate some of the ideas concerning periodic behavior, consider the case of $a = 3.83001$. This value of a corresponds to an attracting three-cycle with attractor 0.1561 . . . → 0.5046 . . . → 0.9574 . . . . Figure 3 presents a scatterplot of the values of x_{500} against x_0 for 800 x_0's equally spaced in the interval [0.17, 0.3]. After 500 iterates, all of these initial points lead to paths that approximately cycle through the attractor; but the phase of the cycling varies in a complicated fashion. This suggests a form of sensitivity to initial conditions, at least for points outside of stable sets corresponding to values in the attractor, in the periodic case. By exploring other initial conditions, we can in fact estimate the stable sets corresponding to the three points of the attractor; that is, the three attracting fixed points of f^3. For example, my computations indicate that the stable set corresponding to the point 0.1561 . . . is approximately $I_1 = [0.14545, 0.16357]$. This estimate can be verified by checking that $f^3(x) \in I_1$ for $x \in I_1$. This criterion was numerically satisfied for 2000 equally spaced points in I_1.

To summarize the asymptotic behavior of the logistic map, consider the plot in Figure 4. This plot is intended to indicate the attractor of the logistic map as a function of a. For a grid of values of a from 3.45 to 4, iterates 101 to 221 of the logistic have been plotted. The result, known as an *orbit diagram*, is an interesting, complex object.

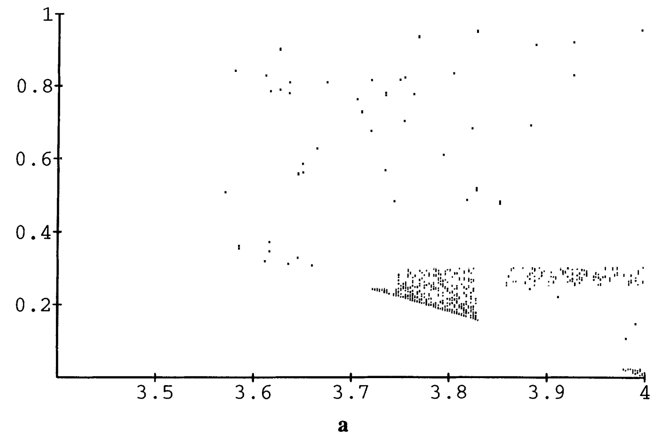


FIG. 4. *Iterates of the logistic map.*

2.1.2 What is mathematical chaos?

There does not appear to be a universally accepted, mathematical definition of chaos. There are different ways to quantify what one might mean by complex or unpredictable behavior. The primary concept appears to be the notion sensitivity to initial conditions, typically quantified as:

DEFINITION 2.1. f: $D \to D$ displays *sensitivity to initial conditions* if there exists $\delta > 0$ such that for any x in D and any neighborhood V of x, there exists a y in V and $n \geq 0$ such that $|f^n(x) - f^n(y)| > \delta$.

This definition suggests that there exist points arbitrarily close to x that separate from x during the time evolution of the dynamical system. However, the definition does not say all points must separate, apparently leaving open the possibility that sensitivity to initial conditions can occur with Lebesgue measure 0. Maps for which all points must separate under iteration are said to be *expansive*. However, expansiveness is typically too restrictive for most maps. (For example, consider the logistic map when $a = 4$. The initial conditions x_0 and $1 - x_0$ result in identical realizations of the map. Therefore, since x_0 can be chosen arbitrarily close to 0.5, expansiveness cannot be claimed). However, the set of all x leading to any periodic behavior when $a = 4$ is a set of Lebesgue measure zero. This sort of phenomenon is related to another component of the mathematical definition of chaos, namely the set of x's leading to periodic behavior is dense in D. That is, complex, aperiodic behavior can arise despite the existence of densely distributed opportunities for well-ordered behavior. This is implicit in the claim of Li and Yorke (1975) that "period three implies chaos." I think the way most people would like to interpret sensitivity is as if the map were almost everywhere, typically, in the sense of Lebesgue measure, expansive.

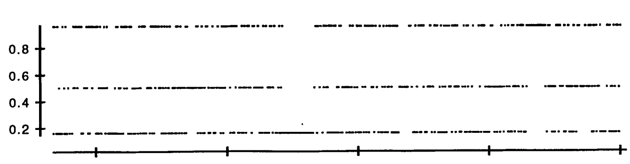


FIG. 3. *Periodic behavior of the logistic map*: $a = 3.83001$. *Scatterplot of* x_{500} *versus* x_0 *for* 800 *points in* [.17, .3].

DEFINITION 2.2. f:$D \to D$ is *almost everywhere expansive* if there exists $\delta > 0$ such that for almost all x in D and almost all y in D, there exists $n \geq 0$ such that $|f^n(x) - f^n(y)| > \delta$.

Although I have not located discussion of such a definition, there may be relationships to some of the definitions relating chaos and randomness discussed below.

Another mathematical concept associated with definitions of chaos intuitively involves the richness of chaotic paths.

DEFINITION 2.3. f:$D \to D$ is *topologically transitive* if for any pair of open sets U, V in D there exists $n > 0$ such that $f^n(U) \cap V$ is not empty.

This definition suggests that some paths of the dynamical system generated by f eventually visit all regions of D. In fact, it turns out that f is topologically transitive if and only if the map possesses a path that is dense in D.

To this point, I have only listed properties associated with chaos, but have not given an explicit definition of chaos. Indeed, different definitions exist. The popularized view of chaos revolves around sensitivity to initial conditions as in Definition 2.1. A now standard and more complete, mathematically motivated definition is given in Devaney (1989, page 50). Namely, a map is "chaotic" if it has the properties of: (i) sensitivity to initial conditions (Definition 2.1); (ii) topological transitivity (Definition 2.3); and (iii) periodic points are dense. Other related definitions of chaos (positive Liapunov exponents, as introduced in Section 3.1.1, and the existence of continuous ergodic distributions, as introduced in Section 2.2.3) involve notions of ergodic theory. See Collet and Eckmann (1980) for further discussion.

2.1.3 Dissipative systems and chaos

Much of the complex behavior of the logistic is a result of its noninvertibility. Indeed, noninvertibility is required to observe chaos for one-dimensional dynamical systems, as defined here. However, everywhere invertible maps in two or more dimensions can also exhibit chaotic behavior. Among the many interesting facets of dynamical systems, one area that receives much attention is the study of *strange attractors*. The basis issue is the long-run behavior of the system. As time proceeds, the trajectories of systems may become trapped in certain bounded regions of the state space of the system. As noted even for the logistic map, these trapping regions or *attractors* can display remarkable oddities. An important example in two dimensions is the Hénon map. This map can display the property of having a strange attractor; that is, the attractor "appears to be locally the product of a two-dimensional manifold by a Cantor set." This quote, along with a motivation of the map, may be found in Hénon (1976). Also, see the previous references for discussion. The Hénon map is given by the following equations:

$$(2.3) \quad x_{t+1} = 1 + y_t - a x_t^2 \text{ and } y_{t+1} = b x_t$$

for fixed values of a and b and $t = 0, 1, \ldots$. This invertible map can not only possess strange attractors, but also display strong sensitivity to initial conditions as encountered earlier. [The rigorous verification of many of the properties of the Hénon map is actually very difficult; see Benedicks and Carleson (1991).] To get a feel for this map, Figure 5 presents a scatterplot of the first 2000 iterations (the attractor) of the map, as well as a time series plot of the "x" series. These computations were based on the conditions $a = 1.4$, $b = 0.3$, $x_0 = 0.4$ and $y_0 = 0.3$.

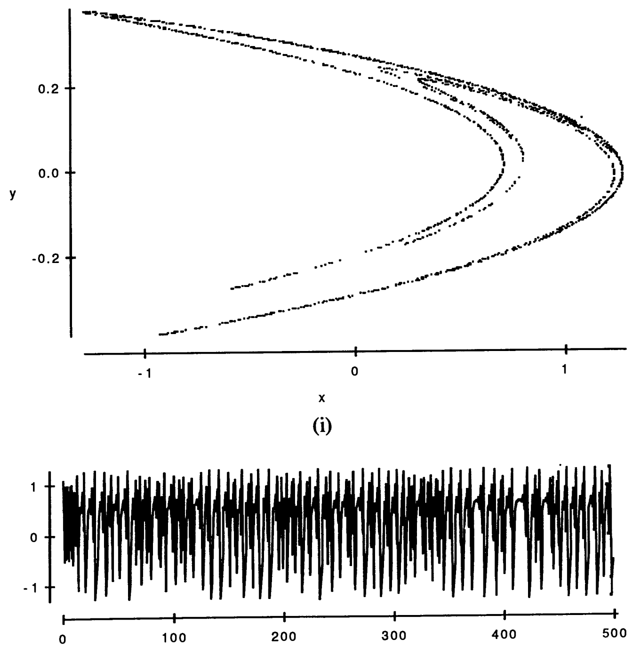


FIG. 5. *Plots for Hénon map. (i) Scatterplot of first* 2000 *iterates. (ii) First* 500 *iterates of x.*

This example illustrates two important aspects of chaotic behavior. First, note that the complex geometrical structure of the Hénon attractor. This object is of *fractal* dimension. Such objects appear in the study of many dynamical systems. The mathematics that suggest such behavior are as follows. The Hénon map, viewed as a transformation from $\mathbf{R}^2$ to $\mathbf{R}^2$, has Jacobian equal to $-b$. If $0 < b < 1$, we make the geometrical observation that Hénon map *contracts* the areas of sets to which it is applied. More generally, such maps are said to be *dissipative*. (Maps that maintain area under iteration are *conservative*.) Intuitively, the complex limiting behavior of chaotic, dissipative dynamical systems is the result of two competing mathematical trends. Dissipativeness suggests that iterates tend to collapse to sets of Lebesgue measure zero. However, an effect of chaos is to prohibit periodic behavior. The natural results consistent with these two phenomena is for the system to be attracted to an infinite, singular set of Lebesgue measure zero (in an appropriate manifold of $\mathbf{R}^k$). Such attracting sets are known as *strange attractors*. (The above heuristics are not complete. For example, conservative systems can display chaotic behavior without being attracted to singular attractors. Rather, conservative systems can have attractors that are colorfully called "fat fractals"; that is, complex geometrical objects that are essentially "space-filling." Such systems will not be considered further in this article. The reader may consult the references for discussion.)

Second, note that there is a relationship between the roles of time and dimension in the definition of dynamical systems. In particular, (2.3) can be written as a one dimensional relationship if we allow consideration of more lags of time:

$$(2.4) \qquad x_{t+1} = 1 + bx_{t-1} - ax_t^2.$$

Statisticians familiar with more conventional time series modeling might see a kinship between (2.4) and a "nonlinear autoregressive model of order 2." I will return to such concepts in Section 4.

2.1.4 Continuous time dynamical systems: Differential equations

Continuous time dynamical systems arise naturally in many applications in which the time evolution of the quantities of interest, composing a k-dimensional vector $\mathbf{x}(t)$, are modeled via differential equations. In particular, consider a initial value problem where $\mathbf{x}(0)$ is an initial condition and the dynamics of the system are quantified by the differential equation

$$(2.5) \qquad d\mathbf{x}(t)/dt = \mathbf{F}(\mathbf{x}(t)),\ t > 0.$$

The value $\mathbf{x}(t)$ describes the state of the system at time t; the domain of possible values of $\mathbf{x}(\cdot)$ is called the phase space. A specific solution to (2.5), "plotted" in the phase space, is known as an orbit. As in the case of discrete time, solutions to (2.5) can be chaotic in the sense that some views of the solutions may appear "random," solutions display sensitivity to initial conditions and, in the case of dissipative systems, now indicated when $\Sigma_{i=1}^{k} \partial F_i / \partial x_i < 0$, orbits are attracted to strange attractors.

Traditional methods for numerically solving differential equations typically involve discrete time approximations. The simplest method, in the context of (2.5) uses the approximation $\mathbf{x}(t_{n+1}) = h\,\mathbf{F}(\mathbf{x}(t_n)) + \mathbf{x}(t_n)$, where the stepsize, h, is small and $t_{n+1} = h + t_n$. Thus, numerical solutions to differential equations are themselves typically discrete time dynamical systems. For the actual computations in this article, I used a more sophisticated approximation known as the four-point Runge–Kutta method. This method appears to be considered a standard method for solving differential equations. Further details are not relevant here. The reader can find a valuable discussion of the basics in Press, Flannery, Teukolsky and Vetterling (1986).

The following example of (2.5) is used in Section 4. The differential equation, known as the Lorenz system, is extremely popular in the literature on chaos. The system, given component-wise, is

$$(2.6) \qquad \begin{aligned} dx/dt &= \sigma(y - x), \\ dy/dt &= -xz + rx - y, \\ dz/dt &= xy - bz, \end{aligned}$$

where σ, r and b are constants. Lorenz (1963) considered this system as a rough approximation to aspects of the dynamics of the Earth's atmosphere. Figure 6 presents plots in phase space of a numerical approximation of a solution to (2.6) where $\sigma = 10$, $r = 28$ and $b = 8/3$. (I tried to indicate why the attractor is nicknamed "The Butterfly.") For this famous choice of the parameters, the solutions (see Figure 7) display sensitive dependence to initial conditions and "unpredictable" fluctuations. For almost all initial conditions, orbits are attracted to the object displayed in the various panels of Figure 6. Note that I have plotted points from the discrete time numerical approximation in this figure. "True" solutions to (2.6) are continuous and are attracted to a "continuous," yet strange attractor of Lebesgue measure zero, since the Lorenz system is dissipative.

2.2 Randomness and Chaos

This section reviews various relationships between chaos and randomness. The key ideas involve the interrelations between sensitivity to initial conditions, uncertainty modeling and ergodic theory. The discussion emphasizes ideas, but, for the sake of brevity, not rigor.

2.2.1 Uncertainty, chaos and randomness

The main topic of this section is how uncertainty, especially in the presence of complexity, naturally leads to the use of random or probabilistic methods. I will begin with a historical perspective. An early and persuasive suggestion that deterministic models may be of limited value is the following discussion of Laplace, circa 1800, (from Laplace, 1951, page 4):

> Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it—an intelligence sufficiently vast to submit these data to analysis—it would embrace in the same formula the movements of the greatest bodies of the uni-

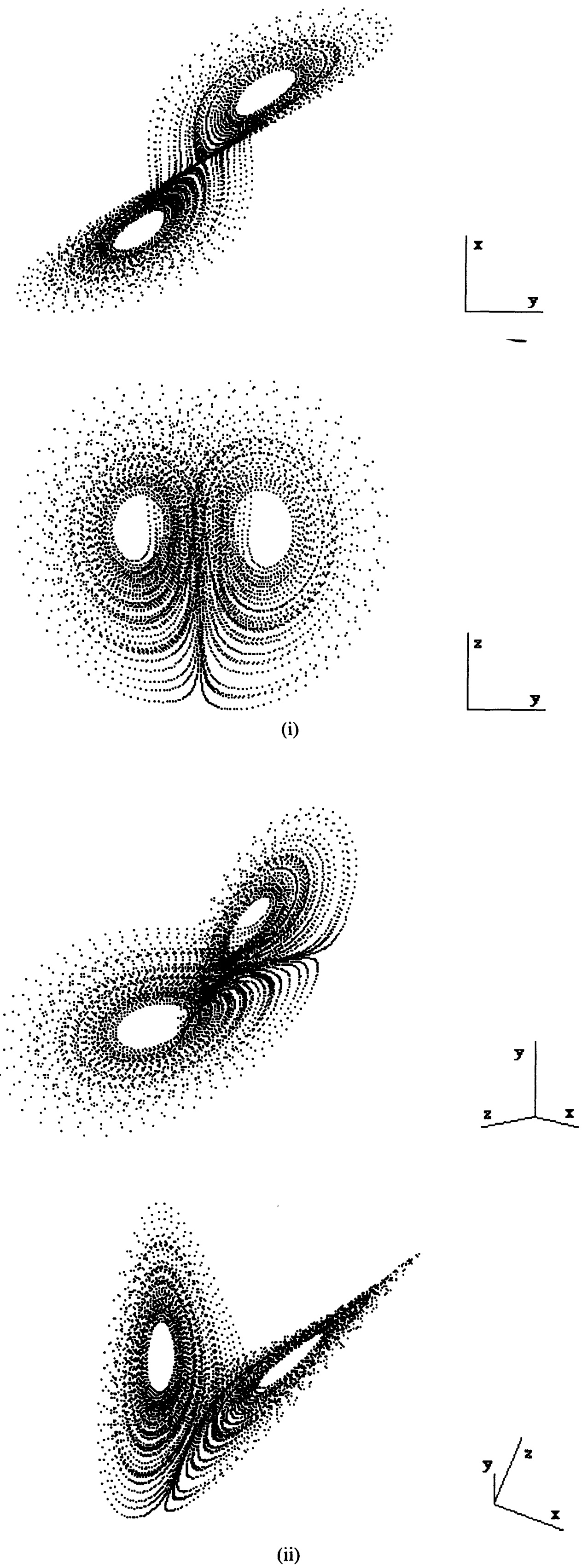


FIG. 6. *The Lorenz attractor: Four views of the Lorenz attractor based on iterates* 1001–8000 *of the four-point Runge–Kutta algorithm with stepsize* = .01.

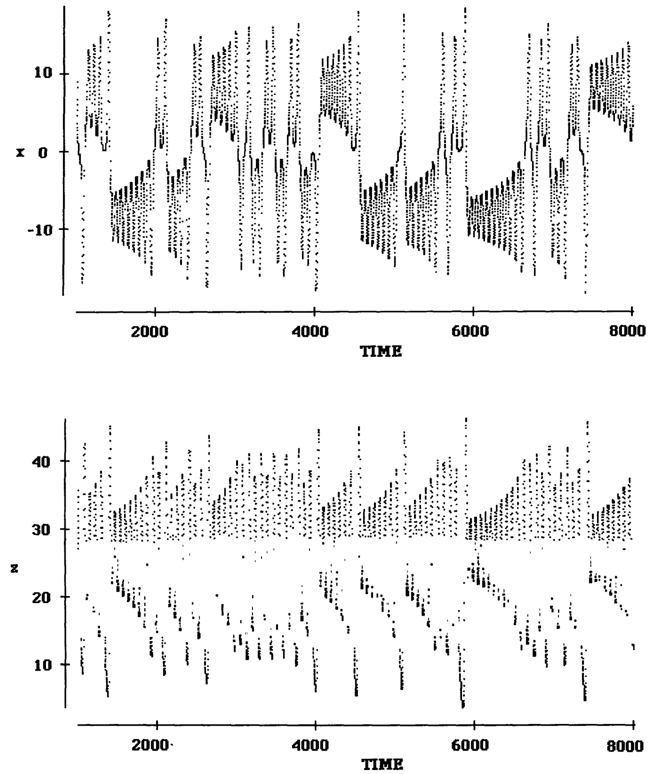


FIG. 7. *Time series plots for the Lorenz system.*

> verse and those of the lightest atom; for it, nothing would be uncertain and the future, as the past, would be present to its eyes.

After briefly reviewing some of the success of science, Laplace continues:

> All these efforts in the search for truth tend to lead [human intelligence] back continually to the vast intelligence which we have just mentioned, but from which it will always remain infinitely removed.

Laplace beautifully itemizes the need for perfect knowledge of the natural laws and initial conditions in deterministic analysis. However, he also clearly questions the relevance of his "vast intelligence" vis-à-vis human efforts. [Laplace has often been misunderstood in this regard. Readers of Laplace who emphasize the first portion of the above quote seem to believe that Laplace was a strict determinist. For example, see Stewart (1989, pages 11–12). For a particularly unjust appraisal of Laplace, see Gleick (1987, page 14) Laplace also wrote, "It is remarkable that ... [the theory of probabilities] should be elevated to the rank of the most important subjects of human knowledge" (Laplace, 1951, page 195). These are hardly the words of a strict determinist.]

A primary example of the use of probabilistic methods to partially overcome difficulties in a complex, deterministic setting is statistical mechanics. The key issue is the study of the motions of a large number (on the order of 10^{24}) of particles floating around in a box. Suppose that the system is such that we can assume that the motion of all the particles are governed by the deterministic laws of motion of classical physics. In principle, we should then be able to compute, given all of the requisite initial conditions, the exact future behavior of the entire system. However, we immediately recognize a problem: can we ever claim, for such a large system, that we actually know, with sufficient accuracy to perform the calculations, all of the initial conditions? At least for the last 140 years, the commonly agreed upon answer is no. Indeed, in the presence of uncertainty concerning these initial conditions, the motion of the particles (or, at least, observable functions of these motions) not only appear "random," but are successfully modeled via stochastic techniques.

The second pertinent class of historical examples directly relates to nonlinear dynamical chaos. For example, probabilistic methods have long been suggested in the area of fluid dynamics, particularly, turbulence; see Grenander and Rosenblatt (1984, Chapter 5) for pertinent discussion. The genius of Poincaré anticipated a great deal of the current interest in nonlinear dynamics. Consider the now frequently cited comments of Poincaré, circa 1900 (from Poincaré, 1946, pages 397–398):

> A very small cause, which escapes us, determines a considerable effect which we cannot help seeing, and then we say that the effect is due to chance. If we could know exactly the laws of nature and the situation of the universe at the initial instant, we should be able to predict the situation of this same universe at a subsequent instant. But even when the natural laws should have no further secret for us, we could know the initial situation only *approximately*. If that permits us to foresee the succeeding situation with the *same degree of approximation,* that is all we require, we say the phenomenon had been predicted, that it is ruled by laws. But it is not always the case; it may happen that slightly differences in the initial conditions produce very great differences in the final phenomena; a slight error in the former would make an enormous error in the latter. Prediction becomes impossible, and we have the fortuitous phenomenon.

This eloquent passage captures the key points for our discussion. It includes the mathematical problem of sensitivity to initial conditions. More germane to my current thesis, however, is Poincaré's explicit suggestion that we can perceive "chance" to be at works as a result of our uncertainty. Poincaré went further by suggesting an operational approach for dealing with such problems that basically involve the treatment of unknown initial conditions as random. (This idea is known as the "Method of Arbitrary Functions" and will be discussed further in Section 5.)

2.2.2 *Defining chaotic to be random*

I will only give a simple description, based on Jackson (1989), of the basic idea of what is known as *symbolic dynamics*. Consider a discrete time dynamical system such as the logistic map. At each iterate, x_n, $n > 0$, of the process, we will associate a simple indicator function based on the location of x_n. In particular, suppose we let $Y_n = 1$ if $x_n \in S$ and $Y_n = 0$ otherwise, where S is some subset of D. That is, each initial condition of the system is now associated with an infinite sequence of 0's and 1's.

DEFINITION 2.4. A map is chaotic if for any sequence of 0's and 1's, there exists an initial condition yielding the same sequence of Y_n's defined above for some fixed S.

The idea here is that if this definition is satisfied, the deterministic dynamical system is, in a sense, at least as rich as Bernoulli coin tossing. For example, it turns out that the logistic map is chaotic as long as a is large enough ($a > 3.83\ldots$) to permit period three (and, thus, all higher periods) behavior. There is a direct relationship between Definition 2.4 and the mathematical definitions relating to periodic points being dense, as described in Section 2.1. I will return to symbolic dynamics in the following subsection.

2.2.3 *Ergodic theory and chaos*

Perhaps the strongest relationships between deterministic chaos and randomness are found through consideration of ergodic theory. [Only a cursory presentation is given here. Some valuable references include Breiman (1968); Cornfield, Fomin and Sinai (1982); Eckmann and Ruelle (1985); and Ornstein (1988).] To motivate the central ideas, imagine that some arbitrary system (stochastic or deterministic) is under study. Consider one experiment in which we will observe the evolution of some variable of the systems over time. In another experiment, we will observe the same variable as in the first, but for several "similar" replicates of the same system at a given time point. Ergodic theory seeks to answer the question, "When can we expect the average of the data, over time, in the first experiment, to be the same (in expectation) as the average of the data, over the replicates at a fixed time?" There is an intriguing relationship between the question of ergodic theory and the familiar question of "independent versus repeated measures" observations. Of particular interest to Bayesians, the role of exchangeability in ergodic theory deserves attention. Pursuit of these issues is beyond the scope of this paper.

To see the role of ergodic theory in deterministic chaos, we will need a bit of formalism. Consider a probability triple (X, F, P), where X is a sample space for some experiment, F is a σ-algebra of subsets of X and P is a probability measure. Assume X is a subset of R^1 and, thus, permit P to represent a probability measure or corresponding distribution function. Next, we introduce a function, f which maps X to X. For X distributed according to P, consider the random variable $f(X)$. The function, or transformation, f is said to be *measure preserving* (or *invariant*) *with respect* to P if the random variable $f(X)$ has distribution P. To solidify this definition by example, let $X = [0,1]$ and let P be the arc-sine [or beta(0.5, 0.5)] distribution, with probability density function $p(x) = \{\pi\sqrt{x(1-x)}\}^{-1}$. It is easy to check that the logistic function, $f(x) = 4x(1-x)$, is invariant with respect to P.

A few more definitions are needed to relate these ideas to dynamical systems. For an invertible transformation f, a subset $A \in F$ is *invariant* if $f^{-1}(A) = A$. For a noninvertible f, invariance of A means $f^t(A) = A$ for all $t > 0$. Further, f is said to be *ergodic* if for every invariant subset, $A \in F$, $P(A) = 0$ or 1. That is, if f is ergodic, sample paths of the dynamical system obtained from f do not become trapped in proper subsets of the support of P, but rather mix over its support. (Note the correspondence to topological transitivity in Definition 2.3.) Furthermore, if f is ergodic with respect to two probability measures P_1 and P_2 on the same measure space, then either $P_1 = P_2$ or P_1 is orthogonal to P_2. With this structure, we can state a simple version of the ergodic theorem:

If f is measure-preserving and ergodic on (X, F, P) and Y is any random variable such that $E(|Y|) < \infty$, then

$$\frac{1}{n}\sum_{i=1}^{n} Y(f^i(x_0)) \to E_P(Y)\ a.s.(P),\ as\ n \to \infty.$$

To illustrate this result, again consider the logistic map with $a = 4$ and let P correspond to the arc sin law. To see that f is ergodic, it is not hard to

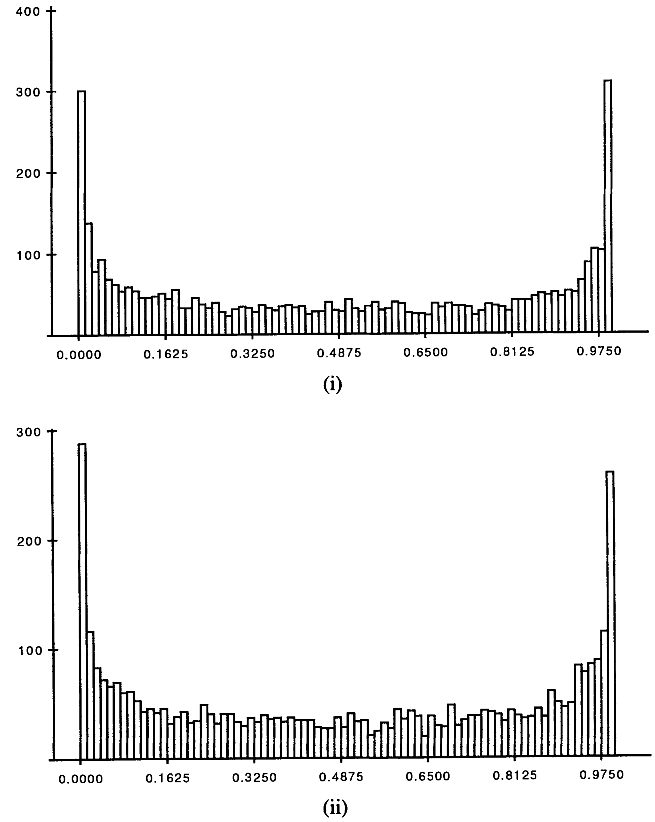


FIG. 8. *Example of ergodic behavior: Logistic map,* $a = 4.0$. (*i*) *Histogram of* 4000 *iterates of* $x_0 = .20005$. (*ii*) *Histogram of the logistic map at time* 2000 *for* 4000 x_0*'s in* [.10005, .30005].

show that except for X, for which $P(X) = 1$, any invariant subset must be denumerable and thus have P-measure 0. Ergodic behavior is suggested by considering sample paths of the logistic. Figure 8 (i) presents a histogram of 4000 iterates of this map, beginning at the initial condition $x_0 = 0.20005$. Figure 8 (ii) presents the histogram of the values of the logistic map at time 2000 for 4000 different initial conditions. In both cases, we see the appearance of something like the arc-sine density. In the construction of Figure 8 (ii) I have used what is perhaps the most important feature of the application of the ergodic theorem to deterministic dynamical systems. The key point is to note the power of the "almost sure" convergence of the theorem. That is, the initial condition x_0 of the system need not be randomly generated according to P for ergodicity to apply to the resulting dynamical system. Indeed, the initial condition need not be "randomly generated." We must only avoid sets of P-measure zero. (For the arc-sine distribution, this means Lebesgue measure zero.) Furthermore, for any logistic with a large ($a > 3.83\ldots$) enough to admit periodic cycles of all orders, each periodic cycle generates a discrete ergodic distribution that assigns equal mass to the components of the cycle.

Note that all these distributions are mutually orthogonal. An implication of these considerations suggest that, to observe ergodic behavior corresponding to the arc sine for the logistic with $a = 4$, we must particularly avoid the preimages (see Section 2) of all periodic points.

Actually, the above arguments offer only a partial explanation of why we see ergodic, apparently random behavior in computer-generated dynamical systems. Electronic computers can only represent a finite, although, fortunately, quite large number of numbers and thus no numerically computed dynamical system can truly be aperiodic. In the logistic map example of Figure 8, the initial conditions were necessarily only truncated real numbers, and thus, lie in a set of Lebesgue measure zero. A possible explanation of why we still can observe behavior expected under the ergodic theorem involves the so-called *shadowing property*. The idea is that the computer-generated orbit of the system is, in a sense, an approximation to some true orbit. The result is that we can be reasonably confident that, if proper care is taken, computer results, especially aggregated results such as the histograms of Figure 8, do in fact capture the correct qualitative features of the system. The required care alluded to in the previous statement refers to the roundoff error present in the numerical computation of the nonlinear function f. Further discussion of computational issues for dynamical systems is beyond the scope of this paper [see Guckenheimer and Holmes (1983); Hammel, Yorke and Grebogi (1987); and Parker and Chua (1989)].

The story of ergodic theory, chaos and randomness is still not complete. A most intimate relationship accrues from the following result. Under the assumptions of the ergodic theorem, if the initial condition x_0 is generated according to P, the sequence $\{Y(f^i(x_0)) = Y_i(x_0),\ i > 0\}$ is a stationary stochastic process. (The probabilistic structure of this stochastic process varies with the ergodic distribution used to generate x_0.) Thus, if we choose Y to be the identity transformation, $Y(x) = x$, the deterministic dynamical system with initial condition somehow chosen, with care to avoid sets of P-measure zero, is a realization of a stochastic process. This result is also important in symbolic dynamics; indeed, the previous statement provides a natural generalized definition of chaos in the spirit of Definition 2.4. Consider the symbolic definition of Section 2.3 for the logistic when $a = 4$ and $S = [0,.5]$. In this case, if the initial condition has the arc sine distribution, it can be shown that the Y's are actually independent, identically distributed ("equally likely") Bernoulli random variables. [For related discussion, see Breiman (1968, page 108).] In such a case, the ergodic theorem coincides with the strong law of large numbers. Pursuing the deterministic argument, I will leave it to philosophers to debate the meaning of the claim, "Starting at initial condition $x_0 = .2958672\ldots$, the probability that $Y_n = 1$ and $Y_{n+1} = 0$ is 0.5^2 if $n = 10^{100}$." Although there is nothing "random" here, the probability statement still makes sense to me.

3. DATA ANALYSIS AND CHAOS

In this section I will review some of the problems associated with chaotic models and data that are of particular to statisticians.

3.1 Measuring Chaos

Important questions arise in attempting to characterize what a chaotic time series of data should look like. Intuitively, a chaotic series should look "random," but this intuition is not necessarily easy to quantify in terms of the mathematical or probabilistic definitions discussed in Section 2.

3.1.1 Liapunov exponents

One of the most popular measures of chaos, *Liapunov exponents*, are based on mathematics associated with the sensitivity to initial conditions concept described in Section 2. [Nearly all of the general references given here present discussions; especially see Eckmann and Ruelle (1985).] Consider a univariate, discrete time dynamical system where $x_n f^n(x_0)$. To study the impact of varying initial conditions, it is natural to consider derivatives dx_n/dx_0. The Liapunov exponent, say $\lambda(s_0)$, is defined as $\lambda(x_0) = \lim_{n-\infty} (1/n) \log[\,|dx_n/dx_0|\,]$. Note that via the chain rule, dx_n/dx_0 can be represented as a product and so $\lambda(x_0)$, under appropriate conditions, may be subject to the ergodic theorem. If so, then $\lambda(x_0) = \lambda$, independent of x_0, almost surely with respect to an appropriate ergodic distribution. Under such circumstances λ is a quantitative measure of the dynamical system's degree of sensitivity to initial conditions. In particular, the approximation

$$dx_n \approx e^{\lambda n} dx_0 \tag{3.1}$$

suggests that for large λ, small changes in initial conditions result in separation of paths at an exponential rate as n grows.

Note that, in the ergodic case, λ can be estimated from a single time series. This means we can assess sensitivity to initial conditions even though the data is based on a single x_0. Of course, this assumes that ergodicity applies and that the x_0 that generated the data is in support of the appropriate ergodic distribution. (Recall the ergodic distributions, although orthogonal, need not be unique.)

Furthermore, by using an appropriate independent sample of initial conditions, we can potentially combine estimates of λ to obtain more precise estimates and associated standard errors.

I have defined λ for a univariate dynamical system; extensions to higher dimensions can be found in the references. Also, the work of Nychka, McCaffrey, Ellner and Gallant (1990) on estimating Liapunov exponents with nonparametric regression techniques is of special interest to statisticians.

3.1.2 The geometric structure of chaos

A second class of measures of chaos involves the notion of attractors of dynamical systems mentioned in Section 2.4. Attempts at characterizing complex geometrical objects have a long history in mathematics. Spurred on by the modern work of B. Mandelbrot and others, as well as the relationships to chaotic dynamical systems, there has been considerable research in the general area of fractal geometry and its applications. Valuable references, at various levels of mathematical sophistication, include Mandelbrot (1982), Barnsley (1988), Peitgen and Saupe (1988), Kaye (1989) and Edgar (1990). The presentation here follows those in Ruelle (1989), Baker and Gollub (1990) and Rasband (1990). Also, see Farmer, Ott and Yorke (1983), Guckenheimer (1984) and Takens (1985).

The starting point is an attempt generalize notions of geometric "size" of sets lying in $\mathbf{R}^k$, from the conventional ideas of "length" ($k = 1$), "area" and "arc length" ($k = 2$), and "volume" and "surface area" ($k > 2$), in cases in which the complexity of the sets of interest prohibits meaningful categorization by these familiar measures. (The sort of set to keep in mind is the Cantor set; this "large" set has Lebesgue measure zero.) The most readily understood class of measures involves the notion of trying to "cover" the set of interest, say A, lying in a compact subset of $\mathbf{R}^k$, with k-dimensional boxes with sides of length ε, small number. If $k = 1$ and A is simply an interval of length L, clearly the "number" of "boxes" used to cover the interval is approximately, ignoring integer part corrections, $N(A, \varepsilon) = L/\varepsilon$. For A, a k-dimensional cube with side L, we have that $N(A, \varepsilon) = (L/\varepsilon)^k$. For such nice sets A, a little algebra suggests the usual interpretation of the dimension of A:

$$(3.2) \qquad k = \lim_{\varepsilon \to 0} \frac{\log N(A, \varepsilon)}{\log(1/\varepsilon)}.$$

Different measures of dimension are based on this notion of "covering" A. [For general A, the limit in (3.2) may not exist.] For example, for an arbitrary, compact set A lying entirely in $\mathbf{R}^k$ to be covered by k-dimensional cubes, define the (Kolmogorov) capacity of A as

$$(3.3) \qquad d_c = \limsup_{\varepsilon \to 0} \frac{\log N(A, \varepsilon)}{\log(1/\varepsilon)}.$$

A second measure of the size of a set is the *Hausdorff dimension*, d_H. Although in a sense it is mathematically more pleasing than versions of (3.3), it definition requires additional development and will be omitted. It should be noted that various authors defined d_c (or one of its cousins) as the *fractal dimension* of A, while others, including Mandelbrot, define the fractal dimension to be d_H. This is unfortunate because these measures can differ: in general, $d_H \le d_c$. For the Cantor set in one dimension, $d_H = d_c = \log 2/\log 3$, suggesting that the set lies in some meaningful fraction, although of Lebesgue measure zero, of the unit interval. The fractal dimension of the Lorenz attractor introduced in Section 2 is about 2.04, suggesting that the attractor lies entirely in a manifold of $\mathbf{R}^3$, but not $\mathbf{R}^2$, and is a fractal.

There are other measures of dimension in addition to those mentioned above. Some of the measures often studied in the chaos literature may be motivated by the suggestion that one relate the geometry of attractors, the structure of ergodic distributions and the mathematical properties of chaos. To motivate the potential of such interrelationships, pretend we are faced with a problem in which a discrete time, chaotic and ergodic system evolves from one of two possible initial conditions. As we watch the evolution, we actually gain information about which initial condition was the true one, because sensitivity to initial conditions implies that the paths separate quickly, no matter how close the two candidates are. In this sense some authors claim that chaotic paths "create information" about initial conditions. [Berliner (1991) presents an analogous argument in terms of statistical information.] Alternatively, sensitivity to initial conditions also suggests that we have decreasing information about the future of a dynamical system as time increases. Consider the following heuristic argument. Suppose that the compact phase space of a univariate dynamical system is discretized, at least for our observation of it, into J cells of equal and small size, say dx. Further, we will agree to summarize our uncertainty in the state of the system through a discrete distribution on the J cells. A familiar measure of our uncertainty about a random variable distributed according a probability distribution is the *entropy* function; for any distribution, P, on the J cells above, the entropy of P is $\mathrm{Ent}(P) = -\Sigma_{i=1}^{J} P_i \log P_i$, where P_i is the probability of the ith cell. Next, suppose that at a given point in time we know that the system lies in a particular cell. The corresponding P has entropy zero. (Actually, a better argument would quantify our uncertainty in the initial condition via a probability distribution over the cell known to contain the system. That is, our entropy, as a measure of uncertainty, is virtually never zero. The basic idea can be conveyed without this correction.) At a large future time point, say t, chaos suggests that the process will be in one of approximately, ignoring integer part corrections, $\exp(\lambda t)$ cells, where λ is the positive Liapunov exponent of the system [recall (3.1)]. If a uniform distribution on these cells is appropriate, the resulting distribution has entropy approximately equal to λt. Thus, entropy increases as the forecast time increases. Providing rigor for this argument is not easy; a key point is that the future distribution need not be uniform on the candidate cells, but rather should involve the appropriate ergodic distribution. Nevertheless, this intuition suggests that the asymptotic rate of change in entropy, represented by the information dimension, introduced below, provides information about the structure of the problem. Further, entropy or the information dimension, may, in some settings, be directly related to the Liapunov exponents of the system. The theory is incomplete, but discussion can be found in the references under the topic of the Kaplan–Yorke Conjecture.

We now turn to descriptions of other measures and to problems of estimation of these measures of dimension. These two aspects are intimately related in the sense that a given measure is often defined as the result of a given estimation procedure. The typical set-up is one in which "long" realizations of the system under study are available, but the true attractor and corresponding ergodic distribution are unknown. This suggests problems of statistical estimation based on data. The standard methods are based on the assumption that the process is ergodic and that the data covers a sufficiently long time period to be representative of the ergodic distribution. For example, suppose we are to estimate a functional; say ξ, such as entropy, of the ergodic distribution, P. The first step is to partition the phase space of the system into $J(\varepsilon)$ "boxes," each of side ε. Ignoring events on the boundaries of the boxes, define $\hat{p}_i(\varepsilon)$ as the probability under the ergodic distribution of the ith box; let $\hat{p}(\varepsilon)$ represent the corresponding (multinomial) approximation to P. Assuming the attractor lies in a compact set, it is reasonable to assume that

$$(3.4) \qquad \xi(P) = \lim_{\varepsilon\to 0} \xi(\hat{p}(\varepsilon)).$$

To unify various measures discussed so far, the following class of rescaled, generalized Renyi information measures are considered:

$$(3.5) \quad \xi_q(P) = \frac{1}{q-1} \lim_{\varepsilon\to 0} \frac{\log\left(\sum_{i=1}^{J(\varepsilon)}(\hat{p}_i(\varepsilon))^q\right)}{\log \varepsilon}.$$

Note that $\xi_q(\cdot)$ is nonincreasing in q. For $q = 1$, a limiting argument yields the representation

$$(3.6) \qquad \xi_1(P) = -\lim_{\varepsilon\to 0} \frac{\operatorname{Ent}(\hat{p}(\varepsilon))}{\log \varepsilon},$$

which is known as the *information dimension.* [$\xi_1(P)$ can also be related to Hausdorff dimension; see Ruelle (1989).] Furthermore, the capacity d_c coincides with $\xi_0(P)$. The key notion of the rescaling involves consideration of the rates at which the attractor fills the appropriate space. In particular, note that, unlike (3.4), (3.2), (3.3), (3.5) and (3.6), respectively, all involve limits that are scaled by $\log \varepsilon$. For example, the information dimension is not equal to the entropy of P. Distributions with different entropies, but similar structure, can have the same information dimension.

To estimate $\xi_p(P)$ from a finite set of data, consider a decreasing sequence of ε_j's. For each ε_j and corresponding partition of the phase space into $J(\varepsilon_j)$ "boxes," we need to estimate all of the $\hat{p}_i$. For a discrete time dynamical system, it is natural to estimate the $\hat{p}_i$'s by the corresponding proportions of the data in the boxes. Most procedures proceed by graphing $\log(\sum_{i=1}^{J(\varepsilon_j)}(\hat{p}_i(\varepsilon_j))^q)$ versus $\log \varepsilon_j$ and estimating $\xi_q(P)$ by the slope of the least-squares linear fitted line through these points. This approach seems to reasonable in principle, but some art is involved. In particular, ε_j's which are too large will not capture the structure of the hypothesized attractor. Also, ε_j's which are too small lead to too many empty cells, resulting in a loss of structure. After all, the true dimension for a finite set of points of zero.

In general, large values of q are useful in relating geometric structure to probabilistic structure. An important example, known as the correlation dimension and due to Grassberger and Procaccia (1983), is given by $\xi_2(P)$. To motivate this dimension define, for a given ergodic distribution, the function

$$(3.7) \qquad \Pr(|X - Y| \le r),$$

for $r > 0$, and where X and Y are independent, identically distributed random variables generated from the ergodic distribution under study, provides some probabilistic information concerning the structure of this distribution. To empirically estimate (3.7) based on a set of data, $\{y_i\}_{i=1,N}$, from a discrete ergodic distribution, Grassberger and Procaccia (1983) consider the quantity

$$C(r) = \frac{1}{N^2} \sum_{i,j} I_{[0,r]}(|y_i - y_i|),$$

where I represents the usual indicator function. The *correlation dimension* is defined as

$$d_{GP} = \lim_{r \to 0} \frac{\log C(r)}{\log r}.$$

In conclusion, most of the traditional procedures described in the literature strike me as what a statistician would describe as "nonparametric method of moments" procedures. Also, the effects of observation error, arising from various plausible error models, is not completely understood. For newer methods and further discussion along these lines, see Wolff (1990) and Smith (1992) in the "statistics literature" and Ellner (1988), Möller et al. (1989) and Ramsey and Yuan (1989) in the "physics literature."

3.2 Data Analysis: The Search for Structure

To set the stage for the topics of this section, consider the following "experiment." Suppose I provided you with a data set $\{x_n\}_{n=1,1000}$ consisting of computer-generated iterates of the logistic map for a randomly chosen initial condition and with $a = 4$, but told you nothing about how the data was generated. As a data analyst, you might begin by looking at a simple time series plot of the data. As indicated earlier, the resulting time series plot would suggest that the data are indeed random. However, no reader of this paper would be fooled into making such a conclusion. For example, if you considered fitting an autoregressive model to the data, you would probably first look at a scatterplot of x_{n+1} versus x_n. Of course, this plot would look like a plot of the function $y = 4x(1 - x)$. That is, by simply considering the "right way" to look at the data, the deterministic structure of the logistic map would simply appear despite the random appearance of the original data. On the other hand, if instead of the original data, I provided you with a time series of the corresponding symbolic dynamic of the data (i.e., $\{y_n\}_{n=1,1000}$ where $y_n = 1$, if $x_n >$ 0.5, and 0, otherwise), then there is nothing that you could do to find the now hidden deterministic structure of the underlying logistic map.

Whereas this example falls far short of indicating the complexity associated with chaotic data analysis, it does suggest one of the fundamental questions: What operations can we perform to chaotic data to "find" any underlying determinism? The second part of the example presents a warning concerning the "design of experiments" for chaotic data analysis: we must try to observe variables that can be informative about any underlying structure.

3.2.1 Poincaré Maps

Suppose a particular continuous time dynamical system is under study with the intent of trying to understand some features of the dynamics. The particular system may be a known chaotic system being studied via computer experiments or an unknown system displaying complex, "random" behavior being observed (without error!) in nature. Faced with complex behavior, perhaps in high dimensions, we study a subset of the data carefully chosen to provide information about the underlying dynamics. Specifically, one defines a *Poincaré surface of section*, as some manifold in the phase space, which the realization of the system strikes "transversally." We then consider the successive values of some subset of variables of the system each time the system passes through the section.

The above suggestion is easily illustrated for the Lorenz system. We consider the Poincaré section consisting of the two-dimensional plane, $\{(x, y, z): x = y\}$. Next, simply construct a "time series" $\{\ldots, x(\tau), x(\tau+1), \ldots\}$ of the values, in time order, of x each time the solution passes through the section. The fundamental point is that the new, lower dimensional series is a discrete time dynamical system, although on a different time scale from the original, that inherits important qualitative features from the original. Indeed, there actually exists a function, say f, such that the iterates of the new series may be represented as $x(\tau + 1) = f(x(\tau))$; f is known as a Poincaré Map. The Poincaré Map for our example is numerically suggested in Figure 9. To obtain this map, I simply recorded the values of x at each intersection of the system with the chosen section and then scatterplotted $x(\tau + 1)$ against $x(\tau)$. Although numerical errors are of course present, note that the points appear to fall on some function quite tightly. (The data for this figure are based on 20,000 iterates of the numerical solution, which resulted in 429 intersections.) To indicate the nature of the map, I also scatterplotted $x(\tau + 2)$ against $x(\tau)$ and $x(\tau + 10)$ against $x(\tau)$. The results are simply (numerical estimates of) the appropriate compositions of the Poincaré Map.

The value of the above type of analysis may be more than theoretical in suggesting interesting ideas for data analysis. For example, imagine analyzing data from an unknown, yet chaotic-looking system. If we were fortunate enough to find a Poincaré Map such as the above through data analysis, we made great strides in understanding the system. In particular, some deterministic element of the system is identified. Further, in the Lorenz case, note that the Poincaré Map is a noninvertible

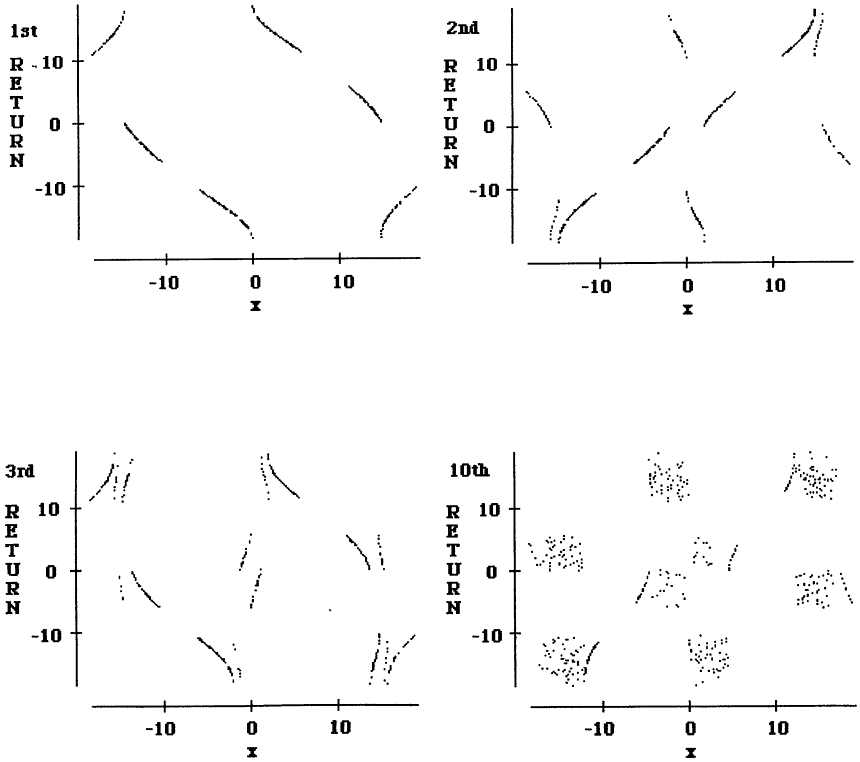


FIG. 9. *Poincaré Map and its iterates.*

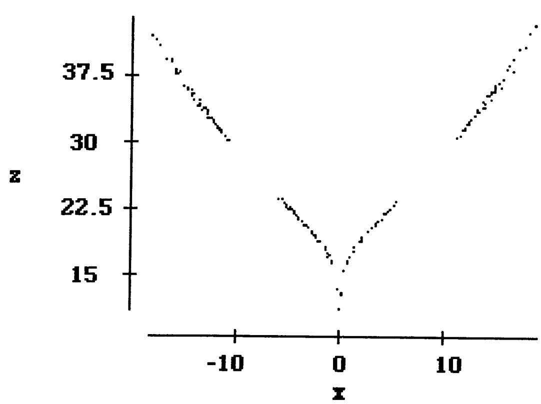


FIG. 10. *Scatterplot of z versus x on the Poincaré section.*

univariate map, as was the logistic map, thereby suggesting a source of the apparent chaotic behavior. Note that there appears to be some potential for limited predictions (also see Nese, 1989). If intersections of the system with the Poincaré section are practically interesting in the context of the problem under study, the empirically obtained Poincaré Map offers very precise predictions of successive values of some variable at points of intersection. To amplify the predictive idea, consider a bivariate time series $\{\ldots, (x(\tau), z(\tau)), \ldots\}$ in the Lorenz system example. A scatterplot of this series is given in Figure 10. Note the very "tight" relationship between x and z on the Poincaré section. This suggests that predictions of the value of z at an intersection time can be made given only the corresponding value of x. Finally, to enhance the value of either of these types of prediction, we might combine forecasts with data analytically obtained information concerning the waiting times, measured on the original time scale, until intersections. A plot of the values of x at returns are plotted as a function of return number in Figure 11. Histograms of the waiting times for the data used to obtain the Poincaré Map are presented in Figure 12.

3.2.2 Reconstruction by time delays

The study of experimental time series data, in an attempt to understand important features of the underlying, but unknown, dynamics driving the process of interest, offers challenges of interest to the statistician. In particular, one may not be able to observe all of the important dynamical variables; indeed, we may not even know which variables are important. A potential approach to data analysis in such cases is usually known as *reconstruction by time delays*. Primary originators of the main ideas are Packard, Crutchfield, Farmer and Shaw (1980), Takens (1981) and D. Ruelle (see Ruelle, 1989).

To motivate the idea, consider a simple dynamical system with a two-dimensional phase space (see Section 2.1.4). The first coordinate of the system is some function over time, say $x_1(t)$. The other coordinate of the system is $x_2(t) = dx_1(t)/dt$. In such a

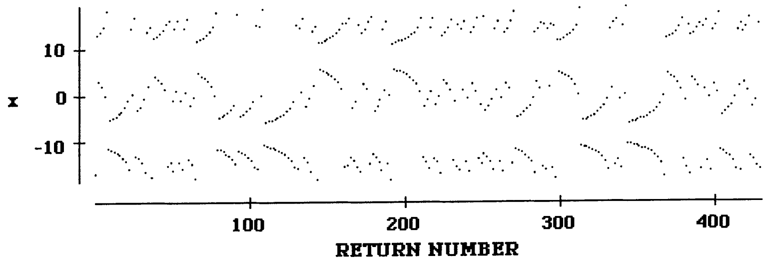


FIG. 11. *Time series plot of returned x by return number.*

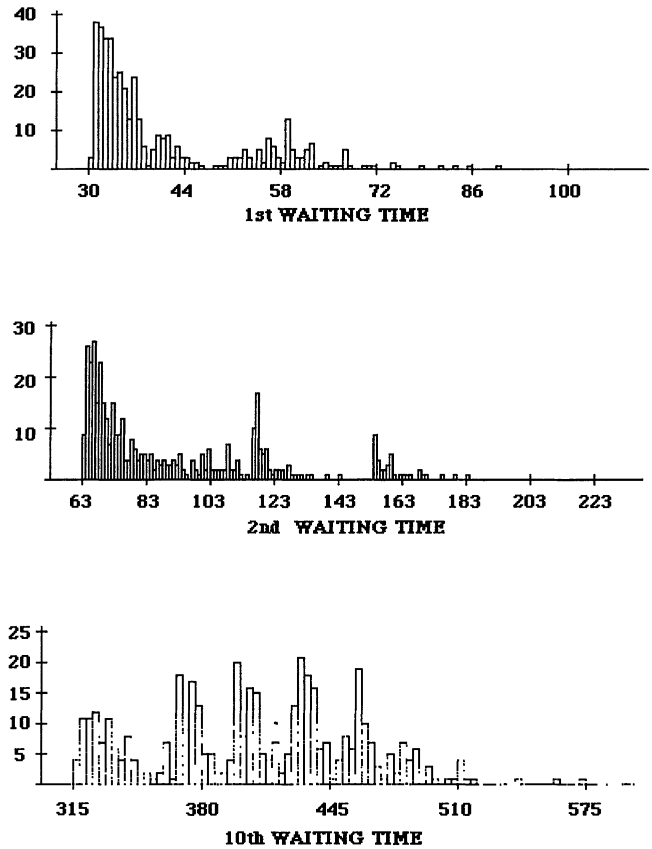


FIG. 12. *Histograms of waiting times until return to Poincaré section.*

case, we quickly recognize that we can learn a great deal about the structure of the behavior of this process, in particular, characteristics of the phase space plot of $x_2(t)$ plotted against $x_1(t)$, even if we can only observe $x_1(\cdot)$ at a discrete collection of time points. This is possible since, for small h, $x_2(t) \approx [x_1(t+h) - x_1(t)]/h$. Therefore, a large number of observations, close to each other temporally, of the first coordinate of the system tell us about some characteristics of the entire system. A second example is based on the Hénon map. Although the map is defined in (2.3) is in $\mathbf{R}^2$, (2.4) suggests that all the dynamics of the process are contained in the single "x" coordinate if only we view the data appropriately.

The basic idea extends to general systems. Consider a k-dimensional dynamical system $\mathbf{x}(t)$ evolving in continuous time. Define a univariate system $\{y_t\}$ where the y process is given by $y_t = h(\mathbf{x}(t))$ for a function h: $\mathbf{R}^k \to \mathbf{R}$. A discrete, multivariate time series is now constructed by defining vectors $\mathbf{v}_t$ as $\mathbf{v}_t = (y_t, y_{t+\tau}, \ldots, y_{t+(n-1)\tau})$ for some choice of τ and n. The claim, based on the theory of Takens (1981), is that analysis of a sequence of say N $\mathbf{v}_t$'s, at time $t, t+\tau, t+2\tau, \ldots, t+N\tau$, permits certain inferences, asymptotically, concerning the qualitative, especially geometric, behavior of the original system. Theoretical justification of this idea stems from the topological result known as Whitney's Embedding Theorem. Of course, there are assumptions under which the methodology works. The crucial one of these is that $n \geq 2d_H + 1$, where d_H is the Hausdorff dimension of the attractor of the original system. There are also interesting statistical issues related to the use of time delays. In particular, reasonable choices for the design parameters τ, N, n (in practice, d_H is unknown) and the embedding function h are needed for data analysis. It is typically the case that h is chosen to simply "pick off" one of the components of $\mathbf{x}$. Depending on h and assuming N must have some reasonable bound, τ should be chosen to be large enough to overcome the very strong local "correlation" in the process, yet be small enough to capture the important dynamics of the process [see Liebert and Schuster (1989) for pertinent discussion]. The impact of the presence of various types of observation errors on analyses seems to be only partially understood. [See Abarbanel, Brown and Kadtke (1989) for some work in this direction.] Beyond the references already given, the reader may consult Broomhead and King (1986) and Rasband (1990) for further discussion. Nicolis and Prigogine (1989) discuss the implementation of the method in an example based on climate data.

3.2.3 Other methodologies

It seems quite natural to attempt to try to fit convenient, flexible functions to chaotic data. The goals are similar to those in nonparametric regression [see Wahba (1990) for general discussion]. Such goals include data smoothing, interpolation between observation times and short-time prediction. The potential here is actually very large. I cannot do justice to the direction here, but offer the following references to the interested reader: Lumley (1970), Crutchfield and McNamara (1987), Farmer and Sidorowich (1987), Casdagli (1989), and Kostelich and Yorke (1990). Furthermore, the reference by Nychka, McCaffrey, Ellner and Gallant (1990) is especially recommended to statisticians for its new results, as well as review in this direction.

4. STATISTICAL ANALYSES FOR CHAOS

4.1 Parametric Statistical Analysis for Chaotic Models

Geweke (1989) and Berliner (1991) are the primary references for this section. A natural class of models for which the statistician feels "at home" are based on the specification of a dynamical function driving the system under study. Assume that the function is known up to a finite collection of parameters. Specifically, the dynamical system is assumed to be driven by the relationship

$$x_{t+1} = f(x_t; \eta) \tag{4.1}$$

where f is specified. The parameter η and the initial value x_0 may both be unknown. Further, assume that, at some time points, we observe the x-process with observation error. Depending on the model we use for the errors, we an construct a likelihood function based on data for the unknown quantities. Even in very simple examples, such as the logistic map observed with independent Gaussian errors, the resulting likelihood functions can be extremely complex, intractable objects. Berliner (1991) offers some heuristics concerning the behavior of such "chaotic likelihoods." For example, it is possible to relate chaos as measured via Liapunov exponents with Fisher information concerning unknown initial conditions. Chaotic processes observed with error produce statistical information concerning initial conditions. However, the value of this information for prediction is limited. Specifically, if maximum likelihood estimates of initial conditions are sought, one is first faced with a very difficult problem of finding such estimates. Even if one were able to obtain a good estimate of x_0, sensitivity to initial conditions moderates the value of such estimates in the context of prediction.

Berliner (1991) considers Bayesian forecasting based on models as suggested in the previous paragraph. Bayesian forecasting, in the presence of unknown initial conditions, is intimately related to ergodic theory via a phenomenon called statistical regularity by Hopf. See Engel (1987) for a very valuable discussion. To quickly communicate the idea, recall the ergodic theory review of Section 2.2.3 and the corresponding example of the logistic map with $a = 4$. In this case the arc-sine distribution is a continuous, ergodic distribution. Invariance implies that if x_0 is generated according to the arc-sine, then at every time t, x_t also has an arc-sine distribution. However, some intuition suggested that x_0 need not actually be generated according to the ergodic distribution for ergodic behavior to manifest itself. Hopf's notion of statistical regularity is a rigorous result along these lines. The result is that, under mild regularity conditions, including a continuous ergodic distribution, say P, if x_0 has any distribution that is absolutely continuous with respect to P, then as t tends to infinity, x_t converges to distribution to P. That is, the initial distribution "washes out." (Note the natural correspondence between this notion and the concept of stationary, ergodic distributions in Markov processes.) The application to Bayesian forecasting is immediate. Under the appropriate conditions, if we compute a posterior distribution for unknown initial conditions and that posterior is abso- lutely continuous with respect to a continuous ergodic distribution, then our implied predictive distribution for x_t as t grows must collapse to the ergodic distribution. This is a strong, statistical reflection of the notion of unpredictability of chaotic processes. The strength of the result is that there is nothing philosophical to debate before accepting implication to forecasting. The theorem says that under perfect conditions in which the prior is agreed to be known and correctly specified, and thus, Bayesian computations are uncontroversial applications of probability theory, long-term predictions, more precise than those associated with a continuous ergodic distribution, of chaotic processes are impossible.

4.2 Statistics and Dynamical Systems

There is a huge literature devoted to statistical analyses for dynamical systems. Statisticians regularly consider the model with "system equation"

$$x_{t+1} = f(x_t; \eta) + z_t, \tag{4.2}$$

where $\{z_t\}$ is itself a stochastic process, and "observation equation" for the observable y

$$y_t = h(x_t) + e_t, \tag{4.3}$$

where h is some function and e represents measurement error. The inclusion of z in (4.2) is suggested as a natural, more realistic version of (4.1), which allows some notion of error in the specification of f, as well as the possibility of unknown environmental effects influencing the evolution of the process. Study of (4.2) and (4.3), especially when f (and h) is a linear function, has been popular for at least 30 years under the general topic of Kalman filtering. See Jazwinski (1970) and Meinhold and Singpurwalla (1983) for review. Study of generalizations of (4.2) [typically without (4.3)] that allow for dependence upon more time lags of the x process, but assume f to be linear and z to be white noise is a subset of classical time series analysis as in Box and Jenkins (1976). Also for discussion of related ideas involving state space modeling, see Aoki (1987). Key recent references on nonlinear time series include Kitagawa (1987) and Priestly (1988); also, see Gallant (1987). A key reference on Bayesian analyses of dynamic models is West and Harrison (1989).

Few would argue with the claim that relatively little of the mainstream statistical time series literature expressly deals with issues in chaos, such as the problems discussed in Section 3. [An important recent work that is, at least partially, motivated by notions of chaos is Tong (1990). This highly recommended book offers substantial review of chaos and related work in nonlinear time series.] However, chaotic data analysis should not be viewed as fundamentally distinct from mainstream statistical time series analysis. Consider (4.2) and (4.3) without any error (z and e) terms. If the analyst assumes, as is customary, that (4.2) generates an ergodic process (i.e., past its transience period), the mathematics discussed in Section 2 imply that the data driven by (4.3) form a realization of a stationary stochastic process. This conclusion applies to the original series as well as data based on any (measurable) h function. Thus, the inference base for data analytic procedures, such as the method of time delays or symbolic dynamics, is the same as in general stationary time series analysis. This common foundation for chaologists and statisticians can hold in the presence of observation errors as well. Suppose that the e_t's are iid. In models without the z terms, if the x process is ergodic, (4.3) corresponds to a stationary process. More generally, if (4.2) yields a Markov process that is stationary and ergodic, (4.3) still yields a stationary stochastic process.

5. DISCUSSION

Two discussion points concerning chaos and statistics are considered here. First, what does the emergence of chaos suggest about statistical modeling? Second, what can statisticians and probabilists bring to the practical study of chaos?

5.1 Impact of Chaos on Statistics

First, I suggest some philosophical implications of chaos for statistics. Section 2 contained discussion of two central points that relate mathematical chaos to probability theory. These were (i) Poincaré's notions relating uncertainty about aspects of deterministic process to randomness, and (ii) the various relationships between ergodic theory, stochastic processes and chaos. The first point of discussion involves foundational impacts of these notions on statistical philosophy. For example, consider the "sacred cow" example of elementary statistics courses: the probabilistic structure of fair coin tossing. What is meant by "the probability of 'heads' is 0.5?" The typical explanation revolves around the frequency interpretation of probability, but relies on some primitive notion of randomness on the part of students. [Many works of I. J. Good are very pertinent here; see Good (1983).] Of course, Bayesian teachers typically question the validity of the frequency definition and remark on the "subjective" assertion of the coin's fairness. However, one might raise the question of the existence of any randomness in this context. If we knew all the pertinent initial conditions for the coin toss, we could apply the laws of physics to determine the outcome of the toss [see Engel (1987) for precisely such computations; also, see Ford (1983)]. It can then be claimed that our use of notions of randomness in coin tossing is really a reflection of uncertainties about initial conditions.

Note that this discussion coincides with a portion of the Bayesian philosophy of statistics. In particular, a key component of the Bayesian paradigm is the modeling of uncertain quantities as if they are random. Perhaps the moral of deterministic chaos and its modeling via probability can be exemplified with a simple challenge to the reader: can you come up with a physical model or a real statistics problem that results in randomness that does not stem from some deterministic uncertainty? (The only "rule" in this mind game is that your cannot resort to quantum mechanics.) If the answer is no, or at least not without a great deal of work, then perhaps we are led to one of the common answers given by Bayesians to B. Efron's question, "Why isn't everyone a Bayesian?" (Efron, 1986). The "answer" is that the question is vacuous: everyone is a Bayesian. (Of course, I am referring to modeling. I am aware that not everyone applies Bayes' Theorem in reaching conclusions.) My point is that frequentist statisticians may find it even more difficult to adhere to their traditional view of statistics. Specifically, the notion of a fine distinction between random and unknown, but deterministic, quantities seems untenable in practice. A further problem for the frequentist involves the notion of "repeatable and identical" experiments. Most statisticians would agree to the nonexistence, at least at a very pure level, of such experiments due to ever-changing environmental effects. Beyond environmental effects, however, chaos suggests that because one could never truly claim identical settings for even controllable factors and, since even slight errors may have great influence, repeatable and identical experiments are not possible. This problem deprives the frequentist any foundation for inference.

These points also bear on discussions within the Bayesian community. Specifically, I refer to the subjective versus objective or necessarian controversy. The notions of ergodic distributions, statistical regularity and Poincaré's method of arbitrary functions all contain relevant messages in this regard. [Engel (1987) includes an excellent review of the method of arbitrary functions.] Consider the following remarks of Von Plato (1983):

> It will be interesting to see what kind of an escape from the recently established rigorous results in ergodic theory the subjectivist school on the foundations of probability theory will take.
>
> It has been traditional to think that deterministic systems allow only subjectively interpretable probabilities. There is however a different tradition, in which the 'stable frequency phenomena of nature' (to use an expression of Hopf) are accounted for by objectively interpreted probabilities, despite the view that these phenomena are supposed to be governed by laws of nature of the classical sort.

I claim that the "escape" alluded to by Von Plato is unnecessary. There is nothing to escape from. As Savage (1973) explains:

> Personal probability works well in science wherever probability seems at all relevant; in particular, the personalistic position does no violence to any genuine objectivity of science; and finally, the personalistic position does not neglect any appropriate role of frequency or of symmetry in the application of probability.

The final phrase of this quote implies that no subjectivist would claim that fundamental knowledge, such as contained in ergodic properties, concerning a system under study should be ignored. Indeed, ergodic distributions and other "objectivist" suggestions are natural candidates for benchmark analyses and starting points upon which to build subjectivist models. However, no scientist should wish to be restricted in the incorporation of additional, case-specific information. Also, I question the practical importance of the notion of "objectively interpreted probabilities," since one must first subjectively decide that the laws under which such probabilities are derived actually apply and are complete in a given setting.

Beyond philosophy, statisticians should be interested in chaos in light of the numerous opportunities for potential research contributions. Also, I believe that the techniques, such as reviewed in Section 3, being developed to deal with chaos should be considered a part of mainstream statistical time series methods. As chaos becomes more popular, such techniques will be useful and "expected" by our colleagues, in day-to-day statistical consulting and collaborative research. Finally, many of the notions and modeling strategies associated with chaos may prove useful in other statistical problems. As an example, it is interesting to speculate on the value of ideas about spatially distributed chaotic processes in statistical imaging problems. For an introduction to "spatio-temporal chaos," see Crutchfield and Kaneko (1987) and Crutchfield (1988).

5.2 Impact of Statistics on the Analysis of Chaos

One possible avenue for scientific inquiry in various settings revolves around the question of deciding whether or not a particular phenomenon or data set is deterministic, yet chaotic, or random. (For example, see Berge, Pomeau and Vidal, 1984; Brock, 1986; Bartlett, 1990; and Sugihara and May, 1990.) It is important to first decide if such a question makes sense. I have already asked whether or not "ordinary" randomness is simply a result of uncertainty in the presence of determinism. Regardless of the foundational answer, statisticians should be able to make *operational* contributions along these lines. Note that such issues are challenging, because deterministic, chaotic models and stochastic models typically both can be made to fit data. (Again, I refer in part to the relationship between ergodic, chaotic systems and stationary stochastic processes.) Specifically, recall the general structure suggested by (4.2) and (4.3). When f is chaotic and z is omitted from (4.2), the interpretation of the model $x_{t+1} = f(x_t; \eta)$ as deterministic seems to be irrelevant in terms of prediction, since meaningful predictions must still be statistical. Alternatively, the inclusion of z in (4.2) need not be based on an assertion that the phenomenon under study is innately random, but rather that f, although of some value, is not capable of capturing all the deterministic aspects of the process. Important contributions from statisticians revolve around the construction of probabilistic/statistical predictions and estimates of interesting quantities. The performance and flexibility of statistically correct prediction procedures should form powerful criteria for choosing among alternative models. For general discussions and references concerning statistical modeling, see Cox (1990), Hill (1990) and Lehmann (1990).

There is also a need for carefully designed experiments in the context of chaotic phenomena. This avenue of research may be particularly interesting and challenging to both statisticians and probabilists, in that at least some of the basis of inference involves ergodic theory.

To summarize, I believe that statisticians and probabilists can make important contributions in the area of chaos. First, in view of the universally accepted view that long-term, precise prediction of the path of a chaotic process is impossible, useful conclusions or predictions must be statistical in nature. For example, one may be able to assess probabilities for interesting events, or subsets of the phase space of the dynamical system under study. The computation of such estimates must typically account for uncertainties in initial conditions, model specification, unknown parameters and the implications of ergodic theory. Second, assessing the effects of one dynamical system that serves as an input to a second system of interest is a natural statistical problem (see Tong, 1990, page 429). All of these points are important for active analyses of dynamical systems when the goals of analysis include prediction and "control" (see Ott, Grebogi and Yorke, 1990, and Sinha, Ramaswamy and Subba Rao, 1990, for relevant discussion of control problems). When conclusions are sought in practical problems based on data and uncertainty modeling, the analyses are exactly the business of statisticians and probabilists.

ACKNOWLEDGMENTS

I thank Steve MacEachern for both listening and sharing his insights concerning this article. My sincere thanks go to the referees for their excellent help. This work was partially supported by NSF Grant DMS-89-06787.

REFERENCES

ABARBANEL, H. D. I., BROWN, R. and KADTKE, J. B. (1989). Prediction and system identification in chaotic nonlinear systems: Time series with broadband spectra. *Phys. Lett. A* **138** 401–408.

AOKI, M. (1987). *State Space Modeling of Time Series*. Springer, New York.

BAKER, G. L. and GOLLUB, J. P. (1990). *Chaotic Dynamics*. Cambridge Univ. Press.

BARNSLEY, M. (1988). *Fractals Everywhere*. Academic, San Diego, Calif.

BARTLETT, M.S. (1990). Chance or chaos? (with discussion). *J. Roy. Statist. Soc. Ser. A* **153** 321–347.

BASAR, E., ed. (1990). *Chaos in Brain Function*. Springer, New York.

BENEDICKS, M. and CARLESON, L. (1991). The dynamics of Hénon the map. *Ann. of Math. (2)* **133** 73–169.

BERGE, P., POMEAU, Y. and VIDAL C. (1984). *Order Within Chaos*. Wiley, New York.

BERLINER, L. M. (1991). Likelihood and Bayesian prediction of chaotic systems. *J. Amer. Statist. Assoc.* **86** 938–952.

BOX, G. E. P. and JENKINS, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.

BREIMAN, L. (1968). *Probability*. Addison-Wesley, Reading Mass.

BROCK, W. A. (1986). Distinguishing random and deterministic systems. *J. Econom. Theory* **40** 168–195.

BROOMHEAD, D. S. and KING G. P. (1986). Extracting qualitative dynamics from experimental data. *Phys. D* **20** 217–236.

CASDAGLI, M. (1989). Nonlinear prediction of chaotic time series. *Phys. D* **35** 335–356.

CHATTERJEE, S. and YILMAZ, M. R. (1992). Chaos, fractals and statistics (with discussion). *Statist. Sci.* **7** 49–121.

COLLET, P. and ECKMANN, J.-P. (1980). *Iterated Maps on the Interval as Dynamical Systems*. Birkhäuser, Boston.

COOPER, N. G., ed. (1989). *From Cardinals to Chaos*. Cambridge Univ. Press.

CORNFELD, I. P., FOMIN, S. V. and SINAI, YA.G. (1982). *Ergodic Theory*. Springer, New York.

COX, D. R. (1990). Role of models in statistical analysis. *Statist. Sci.* **5** 169–174.

CRUTCHFIELD, J. P. (1988). Complexity in nonlinear image processing. *IEEE Trans. Circuits and Systems* **35** 770–780.

CRUTCHFIELD, J. P., FARMER, J. D., PACKARD, N. H. and SHAW, R. S. (1986). Chaos. *Scientific American* **255** 46–57.

CRUTCHFIELD, J. P. and KANEKO, K. (1987). Phenomenology of spatio-temporal chaos. In *Directions in Chaos* (H. Bai-Lin, ed). World Scientific Press, Singapore.

CRUTCHFIELD, J. P. and MCNAMARA, B. S. (1987). Equations of motion from a data series. *Complex Systems* **1** 417–452.

DEVANEY, R. L. (1989). *Introduction to Chaotic Dynamical Systems*, 2nd ed. Benjamin-Cummings, Menlo Park, Calif.

ECKMANN, J.-P. and RUELLE, D. (1985). Ergodic theory of chaos and strange attractors. *Rev. Modern Phys.* **57** 617–656.

EDGAR, G. E. (1990). *Measure, Topology, and Fractal Geometry*. Springer, New York.

EFRON, B. (1986). Why isn't everyone a Bayesian? (with discussion). *Amer. Statist.* **40** 1–11.

ELLNER, S. (1988). Estimating attractor dimensions from limited data: A new method, with error estimates. *Phys. Lett. A* **133** 128–133.

ENGEL E. (1987). A road to randomness. Technical Report 278, Dept. Statistics, Stanford Univ.

FARMER, J. D., OTT, E. and YORKE, J. A. (1983). The dimension of chaotic attractors. *Phys. D* **7** 153–180.

FARMER, J. D. and SIDOROWICH, J. J. (1987). Predicting chaotic time series. *Phys. Rev. Lett.* **59** 845–848.

FEIGENBAUM, M.J. (1978). Quantitative universality in a class of nonlinear transformations. *J. Statist. Phys.* **19** 25–52.

FORD, J. (1983). How random is a coin toss? *Physics Today* **1** 40–47.

GALLANT, A. R. (1987). *Nonlinear Statistical Models*. Wiley, New York.

GEWEKE, J. (1989). Inference and forecasting for deterministic nonlinear time series observed with measurement errors. In *Nonlinear Dynamics and Evolutionary Economics* (R. Day and P. Chen, eds.). To appear.

Glass, L. and Mackey, M. C. (1988). *From Clocks to Chaos*. Princeton Univ. Press.

Gleick, J. (1987). *Chaos: Making a New Science*. Viking Penguin, New York.

Good, I. J. (1983). *Good Thinking*. Univ. Minnesota Press, Minneapolis.

Grassberger, P. and Procaccia, I. (1983). Measuring the strangeness of strange attractors. *Phys. D9* 189–208.

Grenander, U. and Rosenblatt, M. (1984). *Statistical Analysis of Stationary Time Series*, 2nd ed. Chelsea, New York.

Guckenheimer, J. (1984). Dimension estimates for attractors. *Contemp. Math.* **28** 357–367.

Guckenheimer, J. and Holmes, P. (1983). *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. Springer, New York.

Hammel, S. M., Yorke, J. A. and Grebogi, C. (1987). Do numerical orbits of chaotic dynamical processes represent true orbits? *J. Complexity* **3** 136–145.

Hénon M. (1976). A two dimensional mapping with a strange attractor. *Comm. Math. Phys.* **50** 69–77.

Hill, B. M. (1990). A theory of Bayesian data analysis. In *Bayesian and Likelihood Methods in Statistics and Econometrics* (S. Geisser, J. S. Hodges, S. J. Press and A. Zellner, eds.) 49–74. North-Holland, Amsterdam.

Jackson, E. A. (1989). *Perspectives of Nonlinear Dynamics*. Cambridge Univ. Press.

Jazwinski, A. H. (1970). *Stochastic Processes and Filtering Theory*. Academic, New York.

Kaye, B. H. (1989). *A Random Walk Through Fractal Dimensions*. VCH, New York.

Kitagawa, G. (1987). Non-Gaussian state-space modelling of nonstationary time series (with discussion). *J. Amer. Statist. Assoc.* 82 1032–1063.

Kostelich, E. J. and Yorke, J. A. (1990). Noise reduction: Finding the simplest dynamical system consistent with the data. *Phys. D* **41** 183–196.

Laplace, P.S. (1951). *A Philosophical Essay on Probabilities*. Dover, New York.

Lehmann, E. L. (1990). Model specification: The views of Fisher and Neyman and later developments. *Statist. Sci.* **5** 160–168.

Li, T.-Y. and Yorke, J. A. (1975). Period three implies chaos. *Amer. Math. Monthly* **82** 985–992.

Liebert, W. and Schuster, H. G. (1989). Proper choice of the time delay for the analysis of chaotic time series. *Phys. Lett. A* **142** 107–111.

Lorenz, E. N. (1963). Determinist non-periodic flow. *J. Atmospheric Sci.* **20** 130–141.

Lumley, J. L. (1970). *Stochastic Tools in Turbulence*. Academic, New York.

Mandelbrot, B. B. (1982). *The Fractal Geometry of Nature*. W. H. Freeman, New York.

May, R. M. (1987). Chaos and the dynamics of biological populations. In *Dynamical Chaos* (M. V. Berry, I. C. Percival and N. O. Weiss, eds.) 27–43. Princeton Univ. Press.

Meinhold, R. J. and Singpurwalla, N. D. (1983). Understanding the Kalman filter. *Amer. Statist.* **37** 123–127.

Möller, M., Lange, W., Mitschke, F., Abraham, N. B. and Hübner, U. (1989). Errors from digitizing and noise in estimating dimensions. *Phys. Lett. A* **138** 176–182.

Nese, J. M. (1989). Quantifying local predictability in phase space. *Phys. D* **35** 237–250.

Nicholis, G. and Prigogine, I. (1989). *Exploring Complexity*. W. H. Freeman, New York.

Nychka, D. McCaffrey, D., Ellner, S. and Gallant, A. R. (1990). Estimating Lyapunov exponents with nonparametric regression. Inst. Statistics Mimeograph Series No. 1977, Dept. Statistics, North Carolina State Univ.

Ornstein, D. S. (1988). Ergodic theory, randomness, and "chaos." *Science* **243** 182–187.

Ott, E., Grebogi, C. and Yorke, J. A. (1990). Controlling chaos. *Phys. Rev. Lett.* **64** 1196–1199.

Packard, N. H., Crutchfield, J. P., Farmer, J. D. and Shaw, R. S. (1980). Geometry from a time series. *Phys. Rev. Lett.* **45** 712–716.

Parker, T. S. and Chua, L. O. (1989). *Practical Numerical Algorithms for Chaotic Systems*. Springer, New York.

Peitgen, H.-O. and Saupe, D., eds. (1988). *The Science of Fractal Images*. Springer, New York.

Peterson, I. (1988). *The Mathematical Tourist*. W. H. Freeman, New York.

Poincaré, H. (1946). *The Foundations of Science*. Science Press, Lancaster, U.K.

Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. (1986). *Numerical Recipes*. Cambridge Univ. Press.

Priestley, M. B. (1988). *Nonlinear and Non-stationary Time Series Analysis*. Academic, New York.

Ramsey, J. B. and Yuan, H.-J. (1989). Bias and error bars in dimension calculations and their evaluation in some simple models. *Phys. Lett. A* **134** 287–297.

Rasband, S. N. (1990). *Chaotic Dynamics of Nonlinear Systems*. Wiley, New York.

Ruelle, D. (1989). *Chaotic Evolution and Strange Attractors*. Cambridge Univ. Press.

Savage, L. J. (1973). Probability in science: A personalistic account. In *Proceedings of the Fourth International Congress for Logic, Methodology, and Philosophy of Science* (P. Suppes, L. Henkin, A. Joja, and GR. C. Mois, eds.) 417–428. North-Holland, Amsterdam.

Shiraiwa, K. (1985). *Bibliography for Dynamical Systems*. Dept. Mathematics, Nagoya Univ., Furocho, Chikusa-Ku, Nagoya, Japan.

Sinha, S., Ramaswamy, R. and Subba Rao, J. (1990). Adaptive control in nonlinear dynamics. *Phys. D* **43** 118–128.

Smith, R. L. (1992). Estimating dimensions in noisy chaotic time series (with discussion). *J. Roy. Statist. Soc. Ser. B* To appear.

Stewart, I. (1989). *Does God Play Dice*? Blackwell, London.

Sugihara, G. and May, R. M. (1990). Nonlinear forecasting as a way of distinguishing chaos from measurement errors in time series. *Nature* **344** 734–741.

Takens, F. (1981). Detecting strange attractors in turbulence. *Dynamical Systems and Turbulence. Lecture Notes in Math.* **898** 366–381. Springer, New York.

Takens, F. (1985). On the numerical determination of the dimension of an attractor. *Dynamical Systems and Bifurcations. Lecture Notes in Math.* **1125** 99–106. Springer, New York.

Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*. Oxford Univ. Press.

Von Plato, J. (1983). The method of arbitrary functions. *British J. Philos. Sci.* **34** 37–47.

Wahba, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia.

Wegman, E. J. (1988). On randomness, determinism, and computability. *J. Statist. Plann. Inference* **20** 279–294.

West, M. and Harrison J. (1989). *Bayesian Forecasting and Dynamic Models*. Springer, New York.

Wolff, R. C. L. (1990). A note on the behavior of the correlation integral in the presence of a time series. *Biometrika* **77** 689–697.