

Comment

Colleen D. Cutler

As someone who has been working in the area of chaos and statistics for the past few years, I find it very gratifying to witness the beginnings of growing interest among statisticians in this field. As noted by the authors of both papers, the area of chaos has grown very rapidly, spilling over into many different disciplines (notably the physical and biological sciences) and generating much controversy and a wealth of ad hoc techniques (many statistical in nature) for the analysis of chaotic data. The authors of both papers are to be commended for providing stimulating overviews of the theory, techniques and applications of chaos, a far from easy task given the explosive growth of literature in this area.

There is considerable opportunity for statisticians to make an impact in this field (by supplying practitioners, usually scientists, with appropriate methodology), as well as to make an impact on statistics itself (by incorporating features of chaos, such as nonlinear deterministic models, into data analysis). Recent statistical work in chaos (on topics ranging from estimation of dimension and Lyapunov exponents to nonlinear prediction) include Denker and Keller (1986), Cutler and Dawson (1989, 1990), Nychka, McCaffrey, Ellner and Gallant (1990), Wolff (1990), Berliner (1991), Børgsted Hansen (1991), Cheng and Tong (1991), Cutler (1991) and Smith (1991, 1991b). I hope this issue of *Statistical Science* will encourage more statisticians to consider future work in this area.

Numerous topics are addressed by both papers. I will limit my discussion to a few areas in which I have experience or special interest.

PROBABILITY DISTRIBUTIONS ASSOCIATED WITH CHAOS

At least one major source of difficulty in statistical analyses of chaotic systems is the demand by practitioners for techniques that are applicable to an enormous variety of dynamical systems. Whereas "chaos" itself has certain specific defining properties (such as sensitive dependence on initial conditions), the types of probability distributions

arising from chaotic models do not. Here, I assume an ergodic system with probability measure P arising as the limiting empirical distribution along a trajectory (corresponding to a particular set of initial conditions). Specifically

$$(1) \quad P(B) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} I_B(T^k(x))$$

for P -almost all x

where T is the map being iterated. The attractor of a chaotic system need not be fractal (e.g., the attractor of the logistic map $T(x) = 4x(1-x)$ is the unit interval $[0, 1]$), although a fractal structure is a frequent feature. (In the next section, I will indicate how the presence or absence of an underlying fractal structure can affect dimension estimation procedures.) The distribution P , which is supported on the attractor, may have very different behaviors for different systems (indeed, even for different parameter values of the same system). In the case where the attractor A is a smooth subset (e.g., a manifold) of \mathbf{R}^N , P may be absolutely continuous with respect to the Lebesgue measure on A , thus possessing an invariant density g . This density g may or may not be bounded. For example, in the logistic case with $a = 4$, the density $g(x) = \pi^{-1}[x(1-x)]^{-1/2}$ has singularities at both 0 and 1, whereas the "tent" map (discussed by Chatterjee and Yilmaz) features the uniform distribution as invariant measure. The presence of singularities in the density (or even regions of bounded yet steep density) can adversely affect estimation procedures. More typically, at least for dissipative systems in higher dimensions, the attractor is a fractal and, thus, P is necessarily singular with respect to Lebesgue measure. It is interesting that, in the area of chaos, continuous singular distributions arise as natural objects of study, rather than as examples of mathematical pathology (as frequently portrayed in mathematical statistics courses). In fact, due to the possible complexity of attracting sets and distributions, a strong interplay between mathematics and statistics is often required to analyze a chaotic system properly. In practice, the governing equations of a system are rarely known. [Even when they are, it is often difficult or impossible to prove rigorous results about the system asymptotics. The "simple" two-dimensional Hénon mapping defied rigorous analysis from 1976 until

Colleen D. Cutler is Associate Professor, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1.

the recent paper of Benedicks and Carleson (1991).] Thus, statistical analyses may involve a fair amount of “shooting in the dark.” (The problems of observational noise, random physical system perturbations and computational errors become important here. I will elaborate on these in the last section of my comment.) It will be interesting to see the extent to which the “parametric” nonlinear models proposed by Berliner (in Section 4.1) will be useful in the modeling and prediction of physical systems. In any case, at this point in time, there is a demand for statistical techniques that are either highly robust (and, therefore, necessarily inefficient in various cases) or that enable the experimentalist, via data analysis, to reliably distinguish between different types of systems in order that more efficient, specialized techniques can be used.

Of particular interest to physicists and applied mathematicians are the situations under which the attracting distribution P may be classified as a Bowen–Ruelle measure. Such measures, although usually singular (but not always; the attracting arcsine distribution of the logistic map is a Bowen–Ruelle measure), attract initial conditions from sets of positive Lebesgue measure. [Note this differs from simple ergodicity as presented in (1).] This notion is important because such systems may be considered “physically observable.” Another characteristic of Bowen–Ruelle measures is that they usually combine both absolutely continuous and fractal properties; specifically, whereas being singular and concentrated on a fractal set, the measure will have conditional densities along the unstable manifolds (direction of stretching) of the attractor. The existence of Bowen–Ruelle measures has been proven for Axiom A systems (Bowen and Ruelle, 1975), as well as in certain specific other cases. A readable introduction to these ideas can be found in Young (1983), as well as in Ruelle (1989).

ESTIMATION OF DIMENSION

The authors of both papers have included discussions of various dimension concepts and corresponding estimation techniques. In fact, more work in this area has already been done, and I will elaborate here on some of the results and problems. The chief reference here is Cutler (1991), which includes a summary of the mathematical results in Cutler (1990a, b, c), as well as some statistical methods.

Suppose we have an ergodic system consisting of a map T , attractor A and invariant distribution P . Assume we have collected data W_1, \dots, W_n (in the form of a time series) from an orbit of this system. The geometric dimension of A may be of interest,

as it provides a measure of the complexity or “fractalness” of A , as well as a rough idea of the number of degrees of freedom necessary to model the system. However, as noted by all three authors, the definition of dimension here is not unique and is sometimes interpreted as the capacity (or box-counting) dimension d_C or as the Hausdorff dimension d_H . A third definition is possible, that of “packing dimension” d_P , based on the relatively recent work of Tricot and Taylor [for a review see Taylor (1986)]. Hausdorff and packing dimension are closely related and, mathematically, both are more satisfying than capacity dimension. For “reasonable” sets A (including all sorts of irregular and fractal sets), $d_H(A) = d_P(A)$, although this equality need not extend to $d_C(A)$. (The distinction between d_H and d_P becomes important when considering the measure P , which I will do shortly.) Estimation of geometric dimension from time series can be very difficult. Insufficient data, due to the high dimensionality of A , is a frequent problem. Another important problem (although less often discussed) is the fact that the observations W_1, \dots, W_n are distributed according to P (rather than uniformly across A), and if P is highly nonuniform or singular over A we may get a very distorted view of the attractor. A study of the effect of nonuniformity of P on the estimation of $d_C(A)$ is provided by Hunt (1990). This is one reason (among others) motivating the study of dimensions defined in terms of P .

Geometric dimension is determined by scaling properties of the set A ; measure-dependent dimensions are defined in terms of scaling properties of P . We may consider local (or pointwise) scaling as well as global scaling of P . The lower and upper pointwise dimension maps at x are defined, respectively, as

$$(2) \quad \begin{aligned} \sigma^-(x) &= \liminf_{r \rightarrow 0} \frac{\log P(B(x, r))}{\log r}, \\ \sigma^+(x) &= \limsup_{r \rightarrow 0} \frac{\log P(B(x, r))}{\log r} \end{aligned}$$

where $B(x, r)$ is the ball of radius r centered at x . Thus, $\sigma^-(x)$ and $\sigma^+(x)$ indicate the manner in which the probability mass scales at x . Cutler (1990b, c) has shown (for P ergodic and T sufficiently smooth) the existence of unique values α and β such that

$$(3) \quad \sigma^-(x) = \alpha \text{ P-a.s.} \quad \text{and} \quad \sigma^+(x) = \beta \text{ P-a.s.}$$

This means that there exists a set $B \subseteq A$ with $d_H(B) = \alpha$, $d_P(B) = \beta$ and $P(B) = 1$. In most cases, $\alpha = \beta$ [this corresponds to saying that the

pointwise limit $\sigma^-(x) = \sigma^+(x)$ exists *P*-a.s.], but not always. Ledrappier and Misiurewicz (1985) give examples of smooth maps where equality fails. When $\alpha = \beta$, we may interpret this common value as the information dimension of the system (this can then be shown to be equivalent to the definition given by Chatterjee, Yilmaz and Berliner) and estimation of α becomes possible. Note that if *P* is absolutely continuous over *A*, then $\alpha = \beta = d_H(A) = d_P(A)$ (and so the information and geometric dimensions coincide) but if *P* is singular over *A*, we will generally obtain $\alpha < d_H(A)$. This clarifies and completes the comments of Chatterjee and Yilmaz concerning the relationship between pointwise and geometric dimensions in reference to Cutler and Dawson (1989).

We may study global scaling properties of *P* by considering the function $V_r(x) = P(B(x, r))$ and the family of L^q norms ($q \neq 0$) given by

$$(4) \quad \|V_r\|_q = E(V_r(X)^q)^{1/q}$$

where the average is over all points *x* in the phase space. Note that the special case $q = 1$ reduces to the spatial correlation integral $C(r) = E(P(B(X, r)))$, the average mass in a random ball of radius *r*. The generalized Rényi dimensions (mentioned briefly by Berliner) are then obtained by

$$(5) \quad D(q+1) = \lim_{r \rightarrow 0} \frac{\log \|V_r\|_q}{\log r},$$

assuming these limits exist. The correlation dimension ν corresponds to $D(2)$ on this *q*-scale. In the case that *P* is uniform across the attractor, the Rényi dimensions all coincide and equal the information dimension. In the case of nonuniformity of *P*, the Rényi dimensions will differ among themselves to an extent that reflects the nature of the singularities of *P*. The study of singularities has become important in physics, and the Rényi dimensions play a significant role in the so-called “multifractal” analysis of dynamical systems [see Halsey et al. (1986)]. It is important to note that the correlation dimension of a system can be very different from the geometric and information dimensions. [I point this out because the correlation dimension is frequently misused. A fractal correlation dimension need not imply a fractal attractor (see Cutler, 1990a, for a detailed example).] It also remains to be seen how effectively the method of estimating dimension via time-delay embeddings can be used to distinguish chaos from true stochastic behavior. Osborne and Provenzale (1989) constructed an example of a stochastic system where the sequence of correlation dimension estimates converged to a finite value (contrary to the popular

assumption that they must diverge for a stochastic system). It is possible that information dimension might prove more useful in this regard.

One problem in dimension estimation that is peculiar to distributions on fractals is that of “lacunarity,” a topic that was not mentioned by the authors. Lacunarity refers to the failure of the mass of *P* (whether considered pointwise or globally) to scale exactly as a power law for small values of the radius *r*. This phenomenon is most serious and most prevalent for the pointwise dimension, so I will describe it in this case. Typically, we have

$$(6) \quad P(B(x, r)) = K(x, r)r^\alpha,$$

where α is the pointwise dimension and $K(x, r)$ is an oscillating function of *r*. An exception occurs when *P* is absolutely continuous over a smooth subset of \mathbf{R}^N , in which case there exists a constant $0 < K(x) < \infty$ with $K(x, r) \rightarrow K(x)$. Otherwise, at most points *x*, $K(x, r)$ will oscillate or converge to 0 or ∞ as $r \rightarrow 0$ (see Preiss, 1987). For many standard examples, we, in fact, observe $\liminf_{r \rightarrow 0} K(x, r) = 0$, whereas $\limsup_{r \rightarrow 0} K(x, r) = \infty$. This behavior can lead to inconsistency of dimension estimates. The problem of lacunarity and some coping techniques (primarily in terms of correlation dimension) are considered in Theiler (1988, 1990), Cutler (1991) and Smith (1991, 1991b). In particular, Smith uses a method based on mixture models to increase variability to compensate for lacunarity.

NOISE, PERTURBATIONS AND NUMERICAL METHODS

The authors of both papers have briefly mentioned the topic of “noise,” which I believe deserves some comment. There are different ways in which a dynamical system may be regarded as being contaminated by noise. The simplest case occurs when observing the evolution of a physical system; because measurements can be carried out only to a limited degree of accuracy, we automatically incur observational error. This leads to the stochastic model

$$(7) \quad Y_{n+1} = T(X_n) + Z_{n+1},$$

where *T* is the system mapping, Z_{n+1} is a random error term, X_n is the true state of the system after *n* iterations and Y_{n+1} is the observed state after *n* + 1 iterations. This situation has more ramifications for chaotic analysis than it does for more standard data analysis, because here, parameters of interest (such as dimension) are determined by the very fine limiting structure of the system (a

structure that may be obscured by even small amounts of noise). Smith (1991) has discussed the problem of dimension estimation for systems with this type of observational noise.

A different type of noise results when the system under evolution undergoes perturbations due to some external force or change. The perturbations are then propagated through the system. This also describes the situation of rounding error in numerical simulations of chaotic systems; the original rounding error is repeatedly magnified by the “stretching” behavior of the map and the computed numerical trajectory [called a pseudo-orbit by Hammel et al. (1988)] diverges far from the true path. Hammel, Yorke and Grebogi pointed out that often pseudo-orbits are in fact true orbits corresponding to different initial conditions, but even in the ergodic case, this is not necessarily reassuring. The dyadic map $T(x) = 2x \pmod{1}$ on the unit interval is ergodic and chaotic with the uniform distribution as invariant measure. However, all orbits of this

map quickly iterate to zero on the computer. These are true orbits of the system; unfortunately, they correspond to initial conditions (dyadic rationals) that are attracted to the fixed point at zero and do not exhibit “typical” system behavior. Thus, in numerical simulations, it is not always easy to determine whether observed behavior is “real” or an artifact of the simulation procedure.

Corless (1991) has looked at the related problem of approximating solutions to differential equations by numerical methods (here again the computed solution may not resemble the intended system; see Hockett, 1990 and Corless, Essex and Nerenberg, 1991) and has proposed an “operational” definition of chaos. He suggests that a system should be considered chaotic if all “nearby” solutions are chaotic (regardless of the actual properties of the system itself). The reasoning here is that perturbations will cause any physical system to be pushed into neighboring states and these should be the real objects of study.

Comment: Inference and Prediction in the Presence of Uncertainty and Determinism

John Geweke

1. INTRODUCTION

The discovery of nonlinear determinism and chaos in physical systems, the study of these phenomena by physicists and mathematicians and their consideration by investigators in a wide array of disciplines have been ably surveyed from the perspective of statistics and probability in these two articles. The authors have indicated clearly that the contributions and relevance of statistical science are still unresolved, and some basic questions are open. Because chaotic dynamics generate realizations that can be characterized as purely random, what is role of stochastic modeling? If observed deterministic nonlinear processes always interact with stochastic processes, then are the conventional tools of statistical inference any less ade-

quate here than elsewhere? The resolution of these issues will take time, and these surveys will contribute to this process by having brought the statistically relevant aspects of nonlinear determinism and chaos to a wider audience.

Independent of how these questions are answered, the models discussed bring to the practical level latent questions about the implications of determinism for the fundamental role that randomness seems to play in so much of statistics. Berliner has discussed these matters in the final section of his contribution. I have found deterministic models an enlightening vehicle for taking up these questions on a practical level, and in these brief remarks I will provide a few illustrations. The next section provides an approach to inference and prediction in the nonstochastic world of the models these authors have discussed. The likelihood function is presented for two simple models in Section 3, and the construction of predictive densities (for the past, as well as the future) is illustrated in Section 4.

John Geweke is Professor of Economics, University of Minnesota, Minneapolis, Minnesota 55455.