Comment: Relation Between Statistics and Chaos

Richard L. Smith

The involvement of statisticians in the field of chaos is relatively recent, but rapidly growing. Howell Tong's book (Tong, 1990) did much to make statisticians aware of the field. The Royal Statistical Society has hosted discussion papers by Bartlett (1990), and papers from a recent one-day meeting will appear in a special issue of the *Journal of the Royal Statistical Society, Series B* in 1992. I am delighted to see that *Statistical Science* is also taking a lead in developing this fertile source of statistical problems.

Both of the articles under discussion, Berliner (1992) and Chatterjee and Yilmaz (1992), are essentially expository, outlining the theory of chaos in a manner that is oriented toward statistical application. Berliner's article in particular shows how notions of ergodic theory, which tends to be regarded as being "at the hard end" of deterministic dynamical systems theory, have simple probabilistic interpretations that make the theory appealing to statisticians, even though it is essentially describing deterministic systems.

In developing some more specific themes on which I can comment in some detail, I would like to concentrate particularly on the contribution that statisticians can make to the interpretation of data from dynamical systems.

There is an extensive literature on the mathematical properties of systems, such as the logistic map or Lorenz's system of differential equations, and there are also physical systems such as Taylor–Couette flow where the underlying dynamics of the system is sufficiently well understood for a direct association to be made between mathematical theory and experimental observation. But in areas such as ecology or economics, it is impossible to know the detailed mathematical equations governing the system, and the whole of the evidence for "chaos," if indeed there is any evidence at all, comes from the interpretation of experimental data.

Richard L. Smith is Professor of Statistics, University of North Carolina at Chapel Hill, CB 3260, Phillips Hall, Chapel Hill, North Carolina 27599-3260.

Ruelle (1990) gave a particularly witty demonstration of how easy it is to misinterpret such data.

I will focus on just one of the numerous techniques proposed, namely the estimation of correlation dimension. Suppose we have a univariate time series $\{X_n\}$, and form d-dimensional embedded vectors $\mathbf{Y}_n = (X_{n-d+1}, X_{n-d+2}, \ldots, X_n)$ or more generally $\mathbf{Y}_n = (X_{n-(d-1)\tau}, X_{n-(d-2)\tau}, \ldots, X_n)$, where τ is a *lag parameter*. Define

$$C_N(r) = \frac{\sum_{i=2}^{N} \sum_{j=1}^{i} I(\|\mathbf{Y}_i - \mathbf{Y}_j\| < r)}{N(N-1)/2}$$

where N is the number of observations, I denotes the indicator function and $\|\cdot\|$ is a norm. The limit

$$C(r) = \lim_{N\to\infty} C_N(r)$$

is called the *correlation integral*. The correlation dimension, when it is defined, is given by

$$\nu = \lim_{r\to 0} \frac{\log C(r)}{\log r}. \tag{1}$$

In the context of fractals, these formulas give a relatively straightforward way of determining a dimension of a fractal. In the context of chaotic time series, if it is possible to estimate a correlation dimension, which for large enough embedding dimension d is independent of d within the limits of statistical error, then this is often taken as an indicator of deterministic chaos as opposed to randomness.

Most current algorithms for calculating ν from experimental or observational data essentially consist of regressing $\log C_N(r)$ on $\log r$ over a suitable range of r. An alternative technique is the following. First, we strengthen (1) to

$$C(r) \sim ar^{\nu} \quad \text{as } r\downarrow 0. \tag{2}$$

This gives an asymptotic power-law tail for the distribution of distance between two arbitrarily chosen points of the attractor. In practice, we may choose to simplify this even further to

$$C(r) = ar^{\nu} \quad \text{for } r < \varepsilon \tag{3}$$

for some threshold ε, which will be considered further below. A second assumption is that the

$N(N-1)/2$ interpoint distances formed from N **Y** vectors are in fact *independent* random variables. This assumption, which at first sight appears ridiculous, in fact has a lot to support it when considering the lower tail of the distribution. Theiler (1990) made an extensive study of this phenomenon, which he called the *independent distance hypothesis* or IDH. In the probability literature, such ideas are implicit in a number of approximate techniques; for example, the work of Silverman and Brown (1978) on nearest-neighbor distances in spatial patterns, which Barbour and Eagleson (1984) showed could be handled more elegantly by means of what is now called the Stein–Chen (or Chen–Stein) method of Poisson approximation (see Arratia, Goldstein and Gordon, 1990).

Under these two assumptions—power-law tail and IDH—the estimation of the constants a and ν, based on all distances less than some threshold ε, becomes a simple problem of maximum likelihood estimation, with a dimension estimate

$$\hat{\nu} = M \Big/ \sum_{j=1}^{M} \log(\varepsilon / d_j),$$

where M is the number of interpoint distances less than ε and $d_1, \ldots, d_M$ are those distances. The earliest reference to this, for estimating the tail of a distribution, is Hill (1975), but it has several times been rediscovered in the chaos literature (Takens, 1984; Ellner, 1988; Broomhead and Jones, 1989).

A key parameter in all of this is the threshold ε. In practice, this is not well defined but must be chosen through some form of bias-variance trade-off. In Smith (1991), I adapted known results in extreme value theory, especially Hall (1982), to do this. My calculations assumed the IDH, but used the correct correlation integral $C(r)$ based on a continuous distribution of data in d-dimensional space (for which $\nu = d$). In other words, my calculations were for the null hypothesis of no chaos. The results lead to some interesting comparisons on the growth in required sample size as d increases. For example, suppose the true distribution of the **Y** vectors is independent standard normal in d-dimensional space. To achieve a root mean squared error (in $\hat{\nu}$) of 0.1 requires approximate sample sizes (based on an asymptotic calculation) of 5.1×10^3, 3.2×10^6, 2.5×10^9 and 2.6×10^{12}, when the true values are $\nu = 5, 10, 15$ and 20. Clearly, this is a very rapid exponential growth in sample size with true dimension, and strengthens results of Smith (1988) and Ruelle (1990) who have drawn similar conclusions from somewhat cruder arguments. However, if we relax the requirement a little, say to root mean squared error of 1 instead of 0.1, the required approximate sample sizes become more reasonable—30, 1000, 4500 and 2.6×10^6 for the four values of ν quoted above. The conclusion appears to be that it is possible to estimate moderate values of ν based on sample sizes of a few thousand observations, but only if we are not too stringent about the accuracy of the estimates.

These results at least provide a start in obtaining some rigorous statistical theory for the detection of chaos in real-time series. I think they also have a philosophical interpretation for the deterministic chaos versus randomness debate. It has been argued that the majority of "real" systems correspond to deterministic systems in which our notion of randomness arises primarily as an expression of ignorance about the exact state of this system. (Exceptions to this point of view include quantum mechanics and Mendelian genetics.) However, few real systems could be described by low-dimensional dynamics, even if we did assume the system to be essentially deterministic. We would not expect to describe a complicated real system by a correlation dimension of less than 20, say. Calculations such as those above show that the sample sizes needed to estimate such dimensions accurately are enormous, and lend support to the notion that, *in practice*, such high-dimensional systems are indistinguishable from random ones. This may be a way of quantifying Berliner's (Section 5.2) notional of an *operational* distinction between chaos and randomness: the distinction is only meaningful if we can measure it, which we cannot if ν is too large.

Returning to the technical issues, there are many features that the above calculations ignore. Perhaps the most important is that these calculations are essentially only for the null hypothesis of continuously distributed data, and so do not describe the state of affairs when there is a truly chaotic system. Here, we immediately run into a problem, because it is known that even equation (2) may not hold for a fractal. In general, the constant a must be replaced by an oscillating function (Theiler, 1988). So, we need a more general estimation method straight away. Cutler and Dawson (1989, 1990) are among only very few authors to have provided rigorous results for estimation in the case of genuinely fractal systems. Other issues include the question of whether the IDH is valid; the quite separate issue of how to handle time series correlations in the original X series (Theiler, 1986, argued that this could be an important issue, one to which I return below in the discussion of fractional Brownian motion), whether the lag parameter τ should be taken as 1 or some number bigger than 1, and whether correlation dimension is in fact the most appropriate of the various fractal dimensions to consider. Broomhead and Jones (1989) proposed estimation procedures for the class of *generalized Renyi dimensions* (Section 3.1.2. of Berliner's paper) in which, in particular, the cases $q = 2$ and $q = 1$ correspond, respectively, to the correlation dimension and the information dimension. Arguments presented by Cutler (1991) seem to me to be making the case that these should not be considered equivalent concepts and there are, in fact, grounds for regarding information dimension as the right one to try to estimate.

A quite separate class of procedures to distinguish chaos from noise is concerned essentially with direct reconstruction of the map that generates the process. The basic model for a deterministic nonlinear series may be written in the form

$$(4) \qquad X_n = F(X_{n-d}, X_{n-d+1}, \dots, X_{n-1})$$

for some nonlinear function F.

In practice, we will want to permit some form of noise in (4), since even deterministic systems will be affected by measurement error to some extent. This suggests at once two different models: one in which the series $\{X_n\}$ is generated as in (4), but we observe

$$(5) \qquad W_n = X_n + \varepsilon_n,$$

$\{\varepsilon_n, n \geq 1\}$ being a stationary noise process which, in its simplest form, we may assume to consist of independent random variables. I call this the *observational noise* model. A second class of models is the *system noise* model, in which (4) is replaced by the system

$$(6) \qquad W_n = F(W_{n-d}, W_{n-d+1}, \dots, W_{n-1}) + \varepsilon_n,$$

in which the errors $\{\varepsilon_n, n \geq 1\}$ propagate through the system. In the current literature, even assuming the need for a noise term has been accepted, the distinction between (5) and (6) is often not made clear. However, returning to the theme of dimension estimation for a moment, in Smith (1991) I have shown that the two models behave in quite a different way in their effect on a dimension estimate. In more general terms, the distinction between different kinds of noise may be expected to be important for a variety of estimation procedures.

Statistical problems abound. At one level, there is the direct problem of how well F can be estimated, which is essentially a problem in nonlinear regression. Sugihara and May (1990) presented a simple prediction procedure based on this, which was refined by Casdagli (1992). The essence of Casdagli's proposal is as follows. Suppose we want to predict X_{n+T} from $\{X_i, i \leq n\}$, for a given forecasting step T. He performs linear regression of X_{j+T} on $X_{j-(i-1)\tau}$, $1 \leq i \leq d$, as j ranges over a suitable set of indices for which the vectors $(X_{j-(d-1)\tau}, X_{j-(d-2)\tau}, \dots, X_j)$ are suitably close to $(X_{n-(d-1)\tau}, X_{n-(d-2)\tau}, \dots, X_n)$. Specifically, Casdagli uses the k nearest neighbors among the data in d-dimensional space. Here, d and τ are, respectively, the embedding dimension and the lag parameter, and k is an arbitrary choice that effectively controls the amount of smoothing. Sugihara and May (1990) only considered the case that k takes its minimum possible value $d + 1$.

This procedure is consistent with modern ideas in nonparametric regression (Fan, 1990). However, there are problems in applying them to large embedding dimensions, because of the well-known "curse of dimensionality." Perhaps this is the counterpart, for function reconstruction algorithms, of the restriction to relatively small embedding dimensions referred to above in the context of dimension estimation, although here the problems usually occur at around $d = 5$. The most complete discussion of these problems that I am aware of is the work by Nychka, Ellner, McCaffrey and Gallant (1992), who used spline and neural net techniques for function reconstruction.

The main focus of the paper by Nychka, Ellner, McCaffrey and Gallant (1992) is in the estimation of the largest Lyapunov exponent, for which they state a number of rigorous results and conjectures. Both of the papers under discussion mention Lyapunov exponents, and their estimation does indeed seem to be one of the major issues of the statistical approach to chaos. However, the problem is tough; there being few rigorous results even about the consistency of estimators, and hardly any about their asymptotic distributions. Wolff (1992) has made some progress in the case of one-dimensional maps, but there is much more to do. The extension of the problem from estimating the largest Lyapunov exponent to estimating the entire Lyapunov spectrum is another matter again.

Yet another aspect, touched upon briefly by Berliner, is *shadowing*. The problem is inherent in equation (6): given a sequence generated by this equation, with the $\{\varepsilon_n\}$ representing either random or nonrandom (e.g., computer round-off error) noise, is there a deterministic realization $\{X_n\}$ of (4) that stays "close" to $\{W_n\}$ in some suitably defined senses? Affirmative answers were given in the 1960's by Anosov and Bowen in the case of *hyperbolic* systems, but not all dynamical systems are hyperbolic, and more recent theory such as Gregobi, Hammel and Yorke (1987) is needed to handle nonhyperbolic cases. A very interesting paper by Sauer and Yorke (1991) gives what is claimed to be an explicit method of checking whether a shadowing orbit exists, which could be used to check the realism of computer-generated solutions of chaotic systems. However, statistical aspects are only beginning to be looked at (Farmer and Sidorowich, 1990). I mention this area because it is another obvious area of interaction between chaos and statistics, where most of the work at the moment is being done by the dynamical systems experts.

Despite all these high-powered techniques becoming available, a good deal of practical discussion is still in terms of statistically very simple ideas. A case in point is the recent paper of Tilman and Wedin (1991) that was featured in the October 22, 1991 issue of *The New York Times.* This paper used annual population totals for 20 samples of grass over a 6-year period. The fertility of the soil varied from sample to sample, and it was observed that the most fertile soils produced the greatest fluctuations in growth, suggestive of chaos. The conjectured mechanism to explain this behavior was that a very high growth in one year produces a lot of litter that stifles the next year's growth. The challenge for statisticians, I suggest, is to come up with systematic methods for examining hypotheses of this type.

Some of Berliner's paper is concerned with the impact chaos might have on the age-old conflict between Bayesian and frequentist views of statistics. One example of this is in Section 4.1, where he discusses parametric inference from a chaotic model. Apart from the somewhat novel feature that the starting point of the system may itself be a parameter to be estimated, the main interest in this is the highly irregular likelihood functions produced. I believe one can argue that Bayesian methods handle such likelihoods better than classical methods, but this is largely a technical point and does not influence the more fundamental philosophical issues. The philosophical argument is given in Section 5.1. I find it interesting that the only other statistician whom I am aware of commenting on this aspect of chaos, Durbin (1987, 1990), reached the diametrically opposite conclusion from the same facts! To quote Durbin (1987), Section 4.2, "It has been shown using the modern theory of nonlinear dynamics that deterministic systems containing only a few elements can exhibit genuinely stochastic behaviour obeying the laws of probability.... If one puts these theories together with Kolmogorov and Martin-Lof's theory of randomness it appears that the case for the postulation of objective probability models in physical situations such as games of chance has been strengthened." It seems to me that Berliner's own interpretations of ergodic theory in Section 2 are closer to Durbin's position than the one in which Berliner himself subsequently adopts.

Turning now to the paper by Chatterjee and Yilmaz, this is a broad-ranging review that draws together a very large number of references to chaos from all areas of science, although I was disappointed that the authors did not give more analysis of the statistical issues raised by this vast literature. I will confine my detailed comment to one aspect of the paper, the discussion of fractional Brownian motion and fractional differencing in time series analysis. Chatterjee and Yilmaz suggest that there are close connections between these concepts and chaos and fractals. One connection is undeniable: fractional Brownian motions do generate sample paths that are fractals. However, to connect this with chaos seems to me to be wrong and the source of some confusion in the current literature. Two recent papers have served to clarify these issues.

In the first, Osborne and Provenzale (1989) analyzed the behavior of the correlation integral in stochastic time series models, with a spectrum of the form $f(\omega) \sim \omega^{-\alpha}$ for ω near 0. Fractional Brownian motions and fractional ARIMA processes are examples of such processes. Osborne and Provenzale showed that such processes can have estimated correlation dimension converging to $2/(\alpha - 1)$ for $1 < \alpha < 3$, as the length of the time series tends to infinity. This result appeared to contradict the notion that finite correlation dimension corresponds to deterministic chaotic behavior.

A reply has been given by Theiler (1991), however. The essence of his reply is that, in processes with such long-range time-series correlations, the effect of such correlations on estimators of correlation dimension is to induce a serious bias. (The terminology is unfortunate here; "correlation dimension" has nothing to do with correlations in the usual time series sense.) There are many other issues raised by Theiler's paper, but the end result is the claim that the phenomenon observed by Osborne and Provenzale would not occur with a "good" dimension algorithm. This seems to show that chaotic behavior on the one hand, and long-range dependence and self-similarity of the other, are essentially different phenomena and should not be confused.

Chatterjee and Yilmaz end with the challenging claim that the theory of chaos may turn out to be as important as relativity and the uncertainty principle of quantum mechanics. I think one is bound to react to such a statement with a lot of scepticism! Nevertheless, it is clear that an awareness of chaotic phenomena can substantially change our way of thinking about time series and systems in general, and the authors of these two papers are to be congratulated for their clear exposition of these issues.

ACKNOWLEDGMENTS

I am grateful to Steve Ellner and Doug Nychka for recent conversations on some of the topics discussed herein.

Comment: Simplicity and Nonlinearity

Ruey S. Tsay

Chaos is indeed a fascinating subject. It certainly will have some important impact on statistics both in theory and in application. Further, statisticians and probabilists can definitely make significant contributions in chaos. Therefore, I congratulate Professors Chatterjee, Yilmaz and Berliner on their nice and lucid introductions of chaos to the general statistical audience.

I agree with Professor Berliner that chaos is not distinct from mainstream statistics, especially regarding to stochastic processes and time series analysis. The argument between "deterministic" and "stochastic" is misleading. It results from our propensity to dichotomize events surrounding us. From a dynamical system point of view, a "stochastic system" is merely a "deterministic one" with infinite dimension. The difference, if any, is our inability to understand the complexity of a nonlinear system and our preference, justifiably so, to use simple linear models.

Furthermore, there is a close theoretical relation between the stability of a deterministic system and the ergodicity of a stochastic system. For instance, consider the simple deterministic system,

$$(1) \qquad y_t = \begin{cases} ay_{t-1}, & \text{if } y_{t-d} \le 0, \\ by_{t-1}, & \text{if } y_{t-d} > 0, \end{cases}$$

and the stochastic model,

$$(2) \qquad x_t = \begin{cases} ax_{t-1} + e_{1,t}, & \text{if } x_{t-d} \le 0, \\ bx_{t-1} + e_{2,t}, & \text{if } x_{t-d} > 0, \end{cases}$$

where a and b are real numbers, d is a positive integer, $\{e_{1,t}\}$ and $\{e_{2,t}\}$ are independent sequences of independently and identically distributed random variates satisfying $E|e_{i,t}| < \infty$. Chen and Tsay (1991) show that the necessary and sufficient condition of geometrical ergodicity of x_t in (2) is

$$(3) \qquad \begin{gathered} a < 1, \quad b < 1, \quad ab < 1, \\ a^{r(d)}b^{s(d)} < 1, \quad a^{s(d)}b^{r(d)} < 1, \end{gathered}$$

where $r(d)$ and $s(d)$ are nonnegative integers, depending on d such that $r(d)$ and $s(d)$ are odd and even numbers, respectively. It was shown in Lim (1992) that the condition in (3) is also the stability condition of y_t in (1).

Turning to the impact of chaos on statistics, I believe that the impact is far beyond those discussed by the authors. For example, chaos is an "eye-opener" for statisticians and probabilists. It points out loudly and clearly the need to explore nonlinearity and to develop statistical methods and tools that can adequately analyze nonlinear models. The linear world is very limited. That a "tent-map" can generate a realization with autocorrelations the same as those of a particular first-order autoregressive time series illustrates this point clearly. Linear models will undoubtedly continue to play an important role in statistical analysis, but the time has come for statisticians to see the nonlinear planet.

It is natural to ask the question, can we observe attractors in practice, as raised by Professors Chatterjee and Yilmaz and by many people in studying chaos. However, this is a simple-minded question. It falls again into the dichotomous world I alluded to before. Moreover, that no one has yet observed an attractor does not prove the nonexistence of attractors in practice. The important question is that, given a finite realization, possibly noisy, and some specific objectives of analysis, can we determine the most "appropriate model," within a reasonable class, for the data? This is a pressing problem in chaos. More importantly, it is a typical problem in statistics, and the statistician's job is to provide sound methods and proper tools for answering such a question. Here, I like to emphasize the objectives of the intended analysis, which were not emphasized in the two papers, and the

Ruey S. Tsay is Professor of Statistics, Graduate School of Business, University of Chicago, 1101 East 58th Street, Chicago, Illinois 60637.