Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 2297–2301, March 1991
Mathematics

Approximate entropy as a measure of system complexity

(statistic/stochastic processes/chaos/dimension)

STEVEN M. PINCUS

990 Moose Hill Road, Guilford, CT 06437

Communicated by Lawrence Shepp, December 7, 1990 (received for review June 19, 1990)

ABSTRACT **Techniques to determine changing system complexity from data are evaluated. Convergence of a frequently used correlation dimension algorithm to a finite value does not necessarily imply an underlying deterministic model or chaos. Analysis of a recently developed family of formulas and statistics, approximate entropy (ApEn), suggests that ApEn can classify complex systems, given at least 1000 data values in diverse settings that include both deterministic chaotic and stochastic processes. The capability to discern changing complexity from such a relatively small amount of data holds promise for applications of ApEn in a variety of contexts.**

In an effort to understand complex phenomena, investigators throughout science are considering chaos as a possible underlying model. Formulas have been developed to characterize chaotic behavior, in particular to encapsulate properties of strange attractors that represent long-term system dynamics. Recently it has become apparent that in many settings nonmathematicians are applying new "formulas" and algorithms to experimental time-series data prior to careful statistical examination. One sees numerous papers concluding the existence of deterministic chaos from data analysis (e.g., ref. 1) and including "error estimates" on dimension and entropy calculations (e.g., ref. 2). While mathematical analysis of known deterministic systems is an interesting and deep problem, blind application of algorithms is dangerous, particularly so here. Even for low-dimensional chaotic systems, a huge number of points are needed to achieve convergence in these dimension and entropy algorithms, though they are often applied with an insufficient number of points. Also, most entropy and dimension definitions are discontinuous to system noise. Furthermore, one sees interpretations of dimension calculation values that seem to have no general basis in fact—e.g., number of free variables and/or differential equations needed to model a system.

The purpose of this paper is to give a preliminary mathematical development of a family of formulas and statistics, approximate entropy (ApEn), to quantify the concept of changing complexity. We ask three basic questions: (*i*) Can one certify chaos from a converged dimension (or entropy) calculation? (*ii*) If not, what are we trying to quantify, and what tools are available? (*iii*) If we are trying to establish that a measure of system complexity is changing, can we do so with far fewer data points needed, and more robustly than with currently available tools?

I demonstrate that one can have a stochastic process with correlation dimension 0, so the answer to *i* is No. It appears that stochastic processes for which successive terms are correlated can produce finite dimension values. A "phase space plot" of consecutive terms in such instances would then demonstrate correlation and structure. This implies neither a deterministic model nor chaos. Compare this to figures 4 a and b of Babloyantz and Destexhe (1).

If one cannot hope to establish chaos, presumably one is trying to distinguish complex systems via parameter estimation. The parameters typically associated with chaos are measures of dimension, rate of information generated (entropy), and the Lyapunov spectrum. The classification of dynamical systems via entropy and the Lyapunov spectra stems from work of Kolmogorov (3), Sinai (4), and Oseledets (5), though these works rely on ergodic theorems, and the results are applicable to probabilistic settings. Dimension formulas are motivated by a construction in the entropy calculation and generally resemble Hausdorff dimension calculations. The theoretical work above was not intended as a means to effectively discriminate dynamical systems given finite, noisy data, or to certify a deterministic setting. For all these formulas and algorithms, the amount of data typically required to achieve convergence is impractically large. Wolf *et al.* (6) indicate between 10^d and 30^d points are needed to fill out a d-dimensional strange attractor, in the chaotic setting. Also, for many stochastic processes, sensible models for some physical systems, "complexity" appears to be changing with a control parameter, yet the aforementioned measures remain unchanged, often with value either 0 or ∞.

To answer question *iii*, I propose the family of system parameters ApEn(m, r), and related statistics ApEn(m, r, N), introduced in ref. 7. Changes in these parameters generally agree with changes in the aforementioned formulas for low-dimensional, deterministic systems. The essential novelty is that the ApEn(m, r) parameters can distinguish a wide variety of systems, and that for small m, especially $m = 2$, estimation of ApEn(m, r) by ApEn(m, r, N) can be achieved with relatively few points. It can potentially distinguish low-dimensional deterministic systems, periodic and multiply periodic systems, high-"dimensional" chaotic systems, stochastic, and mixed systems. In the stochastic setting, analytic techniques to calculate ApEn(m, r), estimate ApEn(m, r, N), and give rates of convergence of the statistic to the formula all are reasonable problems for which a machinery can be developed along established probabilistic lines.

Invariant Measures and Algorithms to Classify Them

A mathematical foundation for a strange attractor of a dynamical system is provided by considering the underlying distribution as an invariant measure. This requires the existence of a limiting ergodic *physical measure*, which represents experimental time averages (8). Chaos researchers have developed algorithms to estimate this measure, and associated parameters, from data, but explicit analytic calculations are generally impossible, resulting in numerical calculations as normative and in several algorithms to compute each parameter. Representative of the dimension algorithms (9) are capacity dimension, information dimension, correlation dimension, and the Lyapunov dimension. The most com-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: ApEn, approximate entropy; K–S, Kolmogorov–Sinai; E–R, Eckmann–Ruelle.

monly used entropy algorithms are given by the K–S entropy (8), K_2 entropy [defined by Grassberger and Procaccia (10)], and a marginal redundancy algorithm given by Fraser (11). Wolf *et al.* (6) have provided the most commonly used algorithm for computing the Lyapunov spectra.

Other developments further confound a single intuition for each of these concepts. Hausdorff dimension, defined for a geometric object in an n-dimensional Euclidean space, can give fractional values. Mandelbrot (12) has named these nonintegral dimension objects "fractals" and has extensively modeled them. Intuitively, entropy addresses system randomness and regularity, but precise settings and definitions vary greatly. Classically, it has been part of the modern quantitative development of thermodynamics, statistical mechanics, and information theory (13, 14). In ergodic theory, an entropy definition for a measure-preserving transformation was invented by Kolmogorov, originally to resolve the problem of whether two Bernoulli shifts are isomorphic (3). It is distinct from the concept of metric entropy, also invented by Kolmogorov (15), in which a purely metric definition is given. Ellis (16) discusses level 1, 2 (Kullback-Leibler), and 3 entropies, which assess the asymptotic behavior of large deviation probabilities.

Invariant measures have been studied apart from chaos throughout the last 40 years. Grenander (17) developed a theory of probabilities on algebraic structures, including laws of large numbers and a central limit theorem for stochastic Lie groups involving these measures. Furstenberg (18) proved a strong law of large numbers for the norm of products of random matrices, in terms of the invariant measures. Subsequently Oseledets (5) proved the related result that a normalized limit of a product of random matrices, times its adjoint, converges to a nonnegative definite symmetric matrix. This latter result, often associated with dynamical systems, is proved for random matrices in general, and it allows one to deduce the Lyapunov exponents as the eigenvalues of the limiting matrix. Pincus (19) analytically derived an explicit geometric condition for the invariant measures associated with certain classes of random matrices to be singular and "fractal-like" and a first term in an asymptotic expansion for the largest Lyapunov exponent in a Bernoulli random matrix setting (20). Thus noninteger dimensionality and the classification of system evolution by the Lyapunov spectra make sense in a stochastic environment.

The above discussion suggests that great care must be taken in concluding that properties true for one dimension or entropy formula are true for another, intuitively related, formula. Second, since invariant measures can arise from stochastic or deterministic settings, in general it is not valid to infer the presence of an underlying deterministic system from the convergence of algorithms designed to encapsulate properties of invariant measures.

Correlation Dimension, and a Counterexample

A widely used dimension algorithm in data analysis is the correlation dimension (21). Fix m, a positive integer, and r, a positive real number. Given a time-series of data $u(1), u(2), \ldots, u(N)$, from measurements equally spaced in time, form a sequence of vectors $x(1), x(2), \ldots, x(N-m+1)$ in $\mathbf{R}^m$, defined by $x(i) = [u(i), u(i+1), \ldots, u(i+m-1)]$. Next, define for each i, $1 \le i \le N-m+1$,

$$C_i^m(r) = \text{(number of } j \text{ such that } d[x(i), x(j)] \le r)/(N-m+1). \quad [1]$$

We must define $d[x(i), x(j)]$ for vectors $x(i)$ and $x(j)$. We follow Takens (22) by defining

$$d[x(i), x(j)] = \max_{k=1,2,\ldots,m} (|u(i+k-1) - u(j+k-1)|). \quad [2]$$

From the $C_i^m(r)$, define

$$C^m(r) = (N-m+1)^{-1} \sum_{i=1}^{N-m+1} C_i^m(r) \quad [3]$$

and define

$$\beta_m = \lim_{r\to 0} \lim_{N\to\infty} \log C^m(r)/\log r. \quad [4]$$

The assertion is that for m sufficiently large, β_m is the correlation dimension. Such a limiting slope has been shown to exist for the commonly studied chaotic attractors.

This procedure has frequently been applied to experimental data; investigators seek a "scaling range" of r values for which $\log C^m(r)/\log r$ is nearly constant for large m, and they infer that this ratio is the correlation dimension (21). In some instances, investigators have concluded that this procedure establishes deterministic chaos.

The latter conclusion is not necessarily correct: a converged, finite correlation dimension value does not guarantee that the defining process is deterministic. Consider the following stochastic process. Fix $0 \le p \le 1$. Define $X_j = \alpha^{-1/2} \sin(2\pi j/12)$ for all j, where α is specified below. Define Y_j as a family of independent identically distributed (i.i.d.) real random variables, with uniform density on the interval $[-\sqrt{3}, \sqrt{3}]$. Define Z_j as a family of i.i.d. random variables, $Z_j = 1$ with probability p, $Z_j = 0$ with probability $1-p$. Set

$$\alpha = \left(\sum_{j=1}^{12} \sin^2(2\pi j/12)\right)/12, \quad [5]$$

and define $\mathrm{MIX}_j = (1-Z_j)X_j + Z_jY_j$. Intuitively, MIX($p$) is generated by first ascertaining, for each j, whether the jth sample will be from the deterministic sine wave or from the random uniform deviate, with likelihood $(1-p)$ of the former choice, then calculating either X_j or Y_j. Increasing p marks a tendency towards greater system randomness.

We now show that almost surely (a.s.) β_m in Eq. **4** equals 0 for all m for the MIX(p) process, $p \ne 1$. Fix m, define $k(j) = (12m)j - 12m$, and define $N_j = 1$ if $(\mathrm{MIX}_{k(j)+1}, \ldots, \mathrm{MIX}_{k(j)+m}) = (X_1, \ldots, X_m)$, $N_j = 0$ otherwise. The N_j are i.i.d. random variables, with the expected value of N_j, $E(N_j)$, $\ge (1-p)^m$. By the Strong Law of Large Numbers, a.s.

$$\lim_{N\to\infty} \sum_{j=1}^{N} N_j/N = E(N_1) \ge (1-p)^m.$$

Observe that $(\Sigma_{j=1}^N N_j/12mN)^2$ is a lower bound to $C^m(r)$, since $x_{k(i)+1} = x_{k(j)+1}$ if $N_i = N_j = 1$. Thus, a.s. for $r < 1$

$$\limsup_{N\to\infty} \log C^m(r)/\log r \le (1/\log r) \lim_{N\to\infty} \log \left(\sum_{j=1}^{N} N_j/12mN\right)^2 \le \log((1-\mathrm{p})^{2m}/(12m)^2)/\log r.$$

Since $(1-p)^{2m}/(12m)^2$ is independent of r, a.s. $\beta_m = \lim_{r\to 0} \lim_{N\to\infty} \log C^m(r)/\log r = 0$. Since $\beta_m \ne 0$ with probability 0 for each m, by countable additivity, a.s. for all m, $\beta_m = 0$.

The MIX(p) process can be motivated by considering an autonomous unit that produces sinusoidal output, surrounded by a world of interacting processes that in ensemble produces output that resembles noise relative to the timing of the unit. The extent to which the surrounding world interacts with the unit could be controlled by a gateway between the two, with a larger gateway admitting greater apparent noise to compete with the sinusoidal signal.

It is easy to show that, given a sequence X_j, a sequence of i.i.d. Y_j, defined by a density function and independent of the

X_j, and $Z_j = X_j + Y_j$, then Z_j has an infinite correlation dimension. It appears that correlation dimension distinguishes between correlated and uncorrelated successive iterates, with larger estimates of dimension corresponding to more uncorrelated data. For a more complete interpretation of correlation dimension results, stochastic processes with correlated increments should be analyzed.

Error estimates in dimension calculations are commonly seen. In statistics, one presumes a specified underlying stochastic distribution to estimate misclassification probabilities. Without knowing the form of a distribution, or if the system is deterministic or stochastic, one must be suspicious of error estimates. There often appears to be a desire to establish a noninteger dimension value, to give a fractal and chaotic interpretation to the result, but again, prior to a thorough study of the relationship between the geometric Hausdorff dimension and the time series formula labeled correlation dimension, it is speculation to draw conclusions from a noninteger correlation dimension value.

K–S Entropy and ApEn

Shaw (23) recognized that a measure of the rate of information generation of a chaotic system is a useful parameter. In 1983, Grassberger and Procaccia (10) developed a formula, motivated by the K–S entropy, to calculate such a rate from time series data. Takens (22) varied this formula by introducing the distance metric given in Eq. 2; and Eckmann and Ruelle (8) modify the Takens formula to "directly" calculate the K–S entropy for the *physical* invariant measure presumed to underlie the data distribution. These formulas have become the "standard" entropy measures for use with time-series data. We next indicate the Eckmann–Ruelle (E–R) entropy formula, with the terminology as above.

$$\text{Define } \Phi^m(r) = (N-m+1)^{-1} \sum_{i=1}^{N-m+1} \log C_i^m(r). \quad [6]$$

$$\text{E–R entropy} = \lim_{r\to 0} \lim_{m\to\infty} \lim_{N\to\infty} [\Phi^m(r) - \Phi^{m+1}(r)]. \quad [7]$$

Note that

$$\Phi^{m+1}(r) - \Phi^m(r)$$

$$= \text{average over } i \text{ of log[conditional probability that}$$

$$|u(j+m) - u(i+m)| \le r, \text{ given that } |u(j+k) - u(i+k)|$$

$$\le r \text{ for } k = 0, 1, \ldots, m-1]. \quad [8]$$

The E–R entropy and variations have been useful in classifying low-dimensional chaotic systems. In other contexts, its utility appears more limited, as it exhibits the statistical deficiencies noted in the Introduction. Since E–R entropy is infinity for a process with superimposed noise of any magnitude (7), for use with experimental data an approximation of Eq. 7 must be employed with a meaningful range of "r" (vector comparison distance) established. As we see below, a converged "entropy" calculation for a fixed value of r no longer ensures a deterministic system. Also, E–R entropy does not distinguish some processes that appear to differ in complexity; e.g., the E–R entropy for the MIX process is infinity, for all $p \neq 0$.

Fix m and r in Eq. 6; define

$$\text{ApEn}(m, r) = \lim_{N\to\infty} [\Phi^m(r) - \Phi^{m+1}(r)]. \quad [9]$$

Given N data points, we implement this formula by defining the statistic (introduced in ref. 7)

$$\text{ApEn}(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r). \quad [10]$$

Heuristically, E–R entropy and ApEn measure the (logarithmic) likelihood that runs of patterns that are close remain close on next incremental comparisons. ApEn can be computed for any time series, chaotic or otherwise. The intuition motivating ApEn is that if joint probability measures (for these "constructed" m-vectors) that describe each of two systems are different, then their marginal distributions on a fixed partition are likely different. We typically need orders of magnitude fewer points to accurately estimate these marginals than to perform accurate density estimation on the fully reconstructed measure that defines the process.

A nonzero value for the E–R entropy ensures that a known deterministic system is chaotic, whereas ApEn cannot certify chaos. This observation appears to be the primary insight provided by E–R entropy and not by ApEn. Also, despite the algorithm similarities, ApEn(m, r) is not intended as an approximate value of E–R entropy. In instances with a very large number of points, a low-dimensional attractor, and a large enough m, the two parameters may be nearly equal. It is essential to consider ApEn(m, r) as a *family* of formulas, and ApEn(m, r, N) as a *family* of statistics; system comparisons are intended with fixed m and r.

ApEn for $m = 2$

I demonstrate the utility of ApEn(2, r, 1000) by applying this statistic to two distinct settings, low-dimensional nonlinear deterministic systems and the MIX stochastic model.

(*i*) Three frequently studied systems: a Rossler model with superimposed noise, the Henon map, and the logistic map. Numerical evidence (24) suggests that the following system of equations, Ross(R) is chaotic for $R = 1$:

$$dx/dt = -z - y$$

$$dy/dt = x + 0.15y$$

$$dz/dt = 0.20 + R(zx - 5.0). \quad [11]$$

Time series were obtained for R = 0.7, 0.8, and 0.9 by integration via an explicit time-step method with increment 0.005. The y values were recorded at intervals of $\Delta t = 0.5$. Noise was superimposed on each y value by the addition of i.i.d. gaussian random variables, mean 0, standard deviation 0.1. The respective system dynamics are given by noise superimposed on a twice-periodic, four-times-periodic, and chaotic limit cycle. The logistic map is given by

$$x_{i+1} = Rx_i(1 - x_i). \quad [12]$$

Time series were obtained for R = 3.5, 3.6, and 3.8. R = 3.5 produces periodic (period four) dynamics, and R = 3.6 and R = 3.8 produce chaotic dynamics. A parametrized version of the Henon map is given by

$$x_{i+1} = Ry_i + 1 - 1.4x_i^2$$

$$y_{i+1} = 0.3Rx_i. \quad [13]$$

Time series for x_i were obtained for R = 0.8 and 1.0, both of which correspond to chaotic dynamics. All series were generated after a transient period of 500 points. For each value of R and each system, ApEn(2, r, N) was calculated for time series of lengths 300, 1000, and 3000, for two values of r. The sample means and standard deviations were also calculated for each system. Table 1 shows the results.

Notice that for each system, the two choices of r were constant, though the different systems had different r values. One can readily distinguish any Rossler output from Henon output, or from logistic output, on the basis of the quite

Table 1. ApEn(2, r, N) calculations for three deterministic models

Model type	Control parameter	Input noise SD	Mean	SD	r	ApEn(2, r, N) N = 300	N = 1000	N = 3000	r	ApEn(2, r, N) N = 300	N = 1000	N = 3000
Rossler	0.7	0.1	−1.278	5.266	0.5	0.207	0.236	0.238	1.0	0.254	0.281	0.276
Rossler	0.8	0.1	−1.128	4.963	0.5	0.398	0.445	0.459	1.0	0.429	0.449	0.448
Rossler	0.9	0.1	−1.027	4.762	0.5	0.508	0.608	0.624	1.0	0.511	0.505	0.508
Logistic	3.5	0.0	0.647	0.210	0.025	0.0	0.0	0.0	0.05	0.0	0.0	0.0
Logistic	3.6	0.0	0.646	0.221	0.025	0.229	0.229	0.230	0.05	0.205	0.206	0.204
Logistic	3.8	0.0	0.643	0.246	0.025	0.425	0.429	0.445	0.05	0.424	0.427	0.442
Henon	0.8	0.0	0.352	0.622	0.05	0.337	0.385	0.394	0.1	0.357	0.376	0.385
Henon	1.0	0.0	0.254	0.723	0.05	0.386	0.449	0.459	0.1	0.478	0.483	0.486

different sample means and standard deviations. Generally, sample means and standard deviations converge to a limiting value much more quickly (in N) than ApEn does. Greater utility for ApEn arises when the means and standard deviations of evolving systems show little change with system evolution. Different r values were chosen for the three systems to provide the ApEn statistics a good likelihood of distinguishing versions of each system from one another.

For each of the three systems, ApEn(2, r, N) values were markedly different for different R values. ApEn(2, r, 300) gave a first-order approximation of ApEn(2, r, 3000) in these systems, with an average approximate difference of 10% for the $r \approx 0.1$ SD choice and 3.5% for the $r \approx 0.2$ SD choice. The approximation of ApEn(2, r, 1000) to ApEn(2, r, 3000) was good for both choices of r, with an average difference of less than 2% for both choices; we thus infer that ApEn(2, r, 1000) $\approx$ ApEn(2, r) for these r values.

These calculations illustrate many of the salient properties of ApEn as it pertains to evolving classes of dynamical systems. ApEn(2, r, N) appears to correspond to intuition—e.g., apparently more complex Ross(R) systems produced larger ApEn values. ApEn(2, 1.0, 1000) for Ross(0.7) is greater than 0, and equals 0.262 for the noiseless version of this twice-periodic system. Thus a positive ApEn value does not indicate chaos. Contrastingly, ApEn distinguishes the systems Ross(R), R = 0.7, 0.8, and 0.9 from each other. The converged E–R entropy for the Ross(0.7) and Ross(0.8) systems is 0, hence E–R entropy does not distinguish between these systems. The capability to distinguish multiply periodic systems from one another appears to be a desirable attribute of a complexity statistic. Also, the 0.1 intensity superimposed noise on the Rossler system did not interfere with the ability of ApEn to establish system distinction.

(*ii*) The family of MIX processes discussed above. For each of 100 values of p equally spaced between 0 and 1, a time series $\{MIX_j, j = 1, \ldots, N\}$ was obtained as a realization of the random processes. For each value of p, ApEn(2, r, N) was calculated for (r, N) = (0.1, 1000), (0.18, 300), (0.18, 1000), and (0.18, 3000).* Fig. 1 illustrates the results. The intuition that ApEn(2, r, N) should distinguish the processes MIX(p_1) from MIX(p_2) via a larger ApEn value for the larger of the p_i was verified for $p < 0.5$ for all selected statistics. A near-monotonicity of ApEn(2, 0.18, N) with p is seen for $0 < p < 0.7$ for N = 1000, and for $0 < p < 0.75$ for N = 3000. The much larger percentage difference between ApEn(2, 0.18, 300) and ApEn(2, 0.18, 1000), and between ApEn(2, 0.18, 1000) and ApEn(2, 0.18, 3000), for $p > 0.4$ than for corresponding differences for the deterministic models above, suggests that larger values of N are needed in this model to closely approximate E(ApEn(2, r) by ApEn(2, r, N).

*The value r = 0.18 was chosen to ensure that ApEn(2, r, N) MIX(0) = 0 for all N. This occurs when $r < r_{\min} = \min(|X_j - X_k|, X_j \neq X_k) = \sqrt{2(1 - \sqrt{3/2})}$, for X_j defined in MIX. For $r > r_{\min}$, there is similar near-monotonicity in ApEn(2, r, N) with p to that for $r < r_{\min}$.

The values r = 0.1 and r = 0.18 correspond to 10% and 18% of the MIX(p)-calculated standard deviation, for all p. Defining $S_n = MIX_1 + MIX_2 + \ldots + MIX_n$, and $V_n = (MIX_1 - S_n/n)^2 + (MIX_2 - S_n/n)^2 + \ldots + (MIX_n - S_n/n)^2$, straightforward calculations show that $E(S_n/n) = 0$ and $E(V_n) = n$, for n a multiple of 12, for all p. Hence, one cannot distinguish the MIX(p) processes by their sample means and standard deviations.

The ApEn statistics also have been applied to the analysis of heart rate data (N = 1000), and they effectively discriminated between healthy and sick groups of neonates (7). For each of several distinct ApEn(m, r, 1000) statistics, the lowest subgroup of ApEn values consistently corresponded to subjects in the sick group; these values were markedly lower than any values from the healthy group (table 2 of ref. 7).

On the basis of calculations that included the above theoretical analysis, I drew a preliminary conclusion that, for m = 2 and N = 1000, choices of r ranging from 0.1 to 0.2 SD of the $u(i)$ data would produce reasonable statistical validity of ApEn(m, r, N). For smaller r values, one usually achieves poor conditional probability estimates in Eq. **8**, while for larger r values, too much detailed system information is lost. To avoid a significant contribution from noise in an ApEn calculation, one must choose r larger than most of the noise.

ApEn and Analytics

For many stochastic processes, we can analytically evaluate ApEn(m, r) a.s. We next do so for several models. Assume a stationary process $u(i)$ with continuous state space. Let $\mu(x, y)$ be the joint stationary probability measure on $\mathbf{R}^2$ for this process (assuming uniqueness), and $\pi(x)$ be the equilibrium probability of x. Then a.s.

THEOREM 1. *ApEn*(1, r) =

$$-\int \mu(x, y)\, log \left(\int_{z=y-r}^{y+r} \int_{w=x-r}^{x+r} \mu(w, z) dw\, dz \Big/ \int_{w=x-r}^{x+r} \pi(w)\, dw \right) dx\, dy. \quad [14]$$

Proof: By stationarity, it suffices to show that the negative of the right-hand side of Eq. **14** is equal to $E(\log(C_j^2(r)/C_j^1(r)))$, which equals $E(\log P(|x_{j+1} - x_2| \le r \,\|\, |x_j - x_1| \le r))$. Since $P\{|x_{j+1} - x_2| \le r \,\|\, |x_j - x_1| \le r\} = P\{|x_{j+1} - x_2| \le r \;\&\; |x_j - x_1| \le r\}/P\{|x_j - x_1| \le r\}$, Eq. **14** follows at once by jointly conditioning on x_1 and x_2.

Similarly, we have the following.

THEOREM 2. *For an i.i.d. process with density function* π(x), *a.s. (for any* m $\ge$ 1)

$$ApEn(m, r) = -\int \pi(y) log \left(\int_{z=y-r}^{y+r} \pi(z)\, dz \right) dy. \quad [15]$$

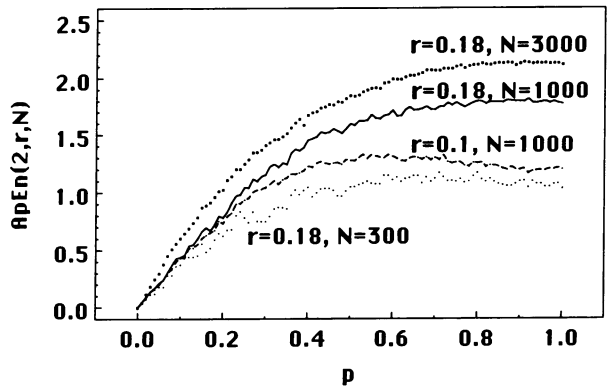

FIG. 1. ApEn(2, r, N) vs. control parameter p for MIX model.

The proof is straightforward and omitted; the i.i.d. assumption allows us to deduce that ApEn(m, r) equals the right-hand side of Eq. **14**, which simplifies to the desired result, since $\mu(x, y) = \pi(x)\pi(y)$. Thus the classical i.i.d. and "one-dimensional" cases yield straightforward ApEn calculations; ApEn also provides a machinery to evaluate less-frequently analyzed systems of nonidentically distributed and correlated random variables. We next see a result familiar to information theorists, in different terminology.

THEOREM 3. *In the first-order stationary Markov chain (discrete state space* X*) case, with* $r < min(|x - y|, x \neq y$, x *and* y *state space values), a.s. for any* m

$$ApEn(m, r) = -\sum_{x\in X}\sum_{y\in X}\pi(x)p_{xy}\,log(p_{xy}). \qquad [16]$$

Proof: By stationarity, it suffices to show that the right-hand side of Eq. **16** equals

$-E(\log(C_1^{m+1}(r)/C_1^{m}(r)))$. This latter expression =

$-E(\log P(|x_{j+m} - x_{m+1}| \le r \,\|\, |x_{j+k-1} - x_k| \le r$ for

$k = 1, 2, \ldots, m)) = -E(\log P(x_{j+m} = x_{m+1} \,\|\, x_{j+k-1} = x_k$ for

$k = 1, 2, \ldots, m)) = -E(\log P(x_{j+m} = x_{m+1} \,\|\, x_{j+m-1} = x_m))$

$$= -\sum_{x\in X}\sum_{y\in X} P(x_{j+m} = y \;\&\; x_{j+m-1} = x)(\log[P(x_{j+m} = y \;\&\; x_{j+m-1} = x)/P(x_{j+m-1} = x)]). \qquad [17]$$

Intermediate equalities in the above follow from the choice of r, and by the Markov property, respectively. This establishes the desired equality.

For example, consider the Markov chain on three points {1, 2, 3}, with transition probabilities $p_{12} = 1$, $p_{23} = 1$, $p_{33} = 1/3$, $p_{31} = 2/3$. The stationary probabilities are computed to be $\pi(1) = \pi(2) = 2/7$, $\pi(3) = 3/7$. For $r < 1$ and any m, application of *Theorem 3* yields that almost surely, $-$ApEn(m, r) = $\pi(3)p_{31}\log(p_{31}) + \pi(3)p_{33}\log(p_{33}) = (2/7)\log 2/3 + (1/7)\log 1/3$. As another example, application of *Theorem 2* to the MIX(1) process of i.i.d. uniform random variables yields that almost surely, $-$ApEn(m, r) $\approx \log(r/\sqrt{3})$ for all m.

Future Direction

Given N data points, guidelines are needed for choices of m and r to ensure reasonable estimates of ApEn(m, r) by ApEn(m, r, N). For prototypal models in which system complexity changes with a control parameter, evaluations of ApEn(m, r) as a function of the control parameter would be useful. Statistics are needed to give rates of convergence of ApEn(m, r, N) to ApEn(m, r) and for fixed N, error estimates for ApEn(m, r, N). Statistics for the stochastic setting would follow from central limit theorems for correlated random variables; verification of conditions, and computations are likely to be nontrivial, since the crucial summands are highly correlated. Monte Carlo techniques can readily be performed to numerically estimate the convergence rates and error probabilities.

In information theory, classification algorithms that are based on universal data compression schemes [e.g., see Ziv (25)] have been seen to be effective for finite state-space processes with a small alphabet. A similarly designed algorithm for the continuous state space could be considered. Also, one could intuitively consider ApEn as a measure of projected information from a finite-dimensional distribution in certain settings. Statistical analysis of projections of higher-dimensional data has been performed via projection pursuit (26), and the kinematic fundamental formulas of integral geometry allow reconstruction of size distributions of an object from lower-dimensional volume and area information (27). Yomdin (28) has used metric entropy to sharpen the Morse–Sard theorem, providing estimates for the "size" of the critical and near-critical values of a differentiable map. These estimates prove useful in geometric measure theory calculations of parameters of manifolds in terms of parameters of low-codimension projections.

I thank Burton Singer both for enlightening discussions and for two references that provided perspective, the Donsker–Varadhan and Prosser–Root papers; Penny Smith for the suggestion that the Ornstein–Uhlenbeck process may be worthy of examination in the current context, prompting a study of Markov processes and chains; and David Salsburg for the reference to projection pursuit.

1. Babloyantz, A. & Destexhe, A. (1988) *Biol. Cybern.* **58**, 203–211.
2. Zbilut, J. P., Mayer-Kress, G. & Geist, K. (1988) *Math. Biosci.* **90**, 49–70.
3. Kolmogorov, A. N. (1958) *Dokl. Akad. Nauk SSSR* **119**, 861–864.
4. Sinai, Ya. G. (1959) *Dokl. Akad. Nauk SSSR* **124**, 768–771.
5. Oseledets, V. I. (1968) *Trans. Moscow Math. Soc.* **19**, 197–231.
6. Wolf, A., Swift, J. B., Swinney, H. L. & Vastano, J. A. (1985) *Physica D* **16**, 285–317.
7. Pincus, S. M., Gladstone, I. M. & Ehrenkranz, R. A., *J. Clin. Monit.*, in press.
8. Eckmann, J. P. & Ruelle, D. (1985) *Rev. Mod. Phys.* **57**, 617–656.
9. Parker, T. S. & Chua, L. O. (1989) *Practical Numerical Algorithms for Chaotic Systems* (Springer, New York), pp. 167–199.
10. Grassberger, P. & Procaccia, I. (1983) *Phys. Rev. A* **28**, 2591–2593.
11. Fraser, A. M. (1989) *IEEE Trans. I.T.* **35**, 245–262.
12. Mandelbrot, B. B. (1977) *The Fractal Geometry of Nature* (Freeman, New York).
13. Shannon, C. E. (1948) *Bell Syst. Tech. J.* **27**, 379–423.
14. Shannon, C. E. (1948) *Bell Syst. Tech. J.* **27**, 623–656.
15. Kolmogorov, A. N. (1956) *Dokl. Akad. Nauk SSSR* **108**, 585–589.
16. Ellis, R. (1985) *Entropy, Large Deviations and Statistical Mechanics* (Springer, New York), pp. 1–27.
17. Grenander, U. (1963) *Probabilities on Algebraic Structures* (Wiley, New York).
18. Furstenberg, H. (1963) *Trans. Am. Math. Soc.* **108**, 377–428.
19. Pincus, S. M. (1983) *Ann. Prob.* **11**, 931–938.
20. Pincus, S. M. (1985) *Trans. Am. Math. Soc.* **287**, 65–89.
21. Grassberger, P. & Procaccia, I. (1983) *Physica D* **9**, 189–208.
22. Takens, F. (1983) in *Atas do 13. Col. brasiliero de Matematicas* (Rio de Janerio).
23. Shaw, R. (1981) *Z. Naturforsch. A* **36**, 80–112.
24. Rossler, O. E. (1976) *Phys. Lett. A* **57**, 397–398.
25. Ziv, J. (1988) *IEEE Trans. Inform. Theory* **IT-34**, 278–286.
26. Huber, P. J. (1985) *Ann. Stat.* **13**, 435–475.
27. Santolo, L. A. (1976) *Integral Geometry and Geometric Probability* (Addison–Wesley, Reading, MA).
28. Yomdin, Y. (1983) *Math. Ann.* **264**, 495–515.