Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 4432–4436, May 1992
Mathematics

Approximating Markov chains

(stochastic processes/dynamical systems/turbulence/weak convergence/approximate entropy)

STEVEN M. PINCUS

990 Moose Hill Road, Guilford, CT 06437

Communicated by Peter D. Lax, January 24, 1992

ABSTRACT **A common framework of finite state approximating Markov chains is developed for discrete time deterministic and stochastic processes. Two types of approximating chains are introduced: (*i*) those based on stationary conditional probabilities (time averaging) and (*ii*) transient, based on the percentage of the Lebesgue measure of the image of cells intersecting any given cell. For general dynamical systems, stationary measures for both approximating chains converge weakly to stationary measures for the true process as partition width converges to 0. From governing equations, transient chains and resultant approximations of all *n*-time unit probabilities can be computed analytically, despite typically singular true-process stationary measures (no density function). Transition probabilities between cells account explicitly for correlation between successive time increments. For dynamical systems defined by uniformly convergent maps on a compact set (e.g., logistic, Henon maps), there also is weak continuity with a control parameter. Thus all moments are continuous with parameter change, across bifurcations and chaotic regimes. Approximate entropy is seen as the information-theoretic rate of entropy for approximating Markov chains and is suggested as a parameter for turbulence; a discontinuity in the Kolmogorov–Sinai entropy implies that in the physical world, some measure of coarse graining in a mixing parameter is required.**

I aim to develop a framework of finite state space approximating (m, r) Markov chains for discrete-time stochastic and deterministic processes. The motivation derives from needs: (*i*) to assess claims of deterministic chaos, from time-series analysis; (*ii*) to produce a tractable, general procedure for "solving" stochastic and deterministic difference equations; and (*iii*) to address meaningful questions for dynamical systems where there is sensitivity to initial conditions.

In many reports of chaos (e.g., refs. 1 and 2), it appears that investigators may be observing a correlated, possibly stochastic process with a stationary measure. To evaluate paradigms other than chaotic processes and independent, identically distributed random variables as candidate models for data, we need first to assess the behavior of continuous-state processes, on a partition, in a statistically valid manner. Process approximation by a low-order Markov chain on a coarse partition will provide this validity. Second, for both stochastic and deterministic differential equations, analytic solution techniques are often nonexistent, so the utility of a family of easily solved approximating processes, converging to a true solution, is apparent. Formally, *m*th-order difference equations are *m*th-order Markov processes, continuous-state space. The idea here is to approximate these systems by *m*th-order discrete-state space Markov chains, which are well understood, with straightforward procedures to calculate stationary measures and rates of convergence to steady state. Third, if a dynamical system or differential equation is sensitive to initial conditions, a transient calculation is inappropriate, since two arbitrarily close initial conditions can produce divergent orbits. Also, steady-state probabilities do not tell the whole story, ignoring correlation between values at successive time points.

For deterministic differential equations, the approximating-chain approach contrasts fundamentally with classical solution methods (3, 4). For a finite-difference approximation, one solves for grid values at time $t + \Delta t$, *deterministically*, in terms of grid values at time t, Δt, the mesh dimensions, and nonlinear differential operators; in the present approach, grid values at time $t + \Delta t$ are *probabilistically specified* from the aforementioned data. Anticipated advantages here are that approximating Markov chains will (*i*) provide a probabilistic analogue of a transient solution; (*ii*) account explicitly for correlation between successive time increments; and (*iii*) have nice "stability" properties, for classically unstable processes, in that stationary measures for these Markov chains will be weakly continuous with perturbations.

The approximating (m, r) Markov chains will be given by explicit transition matrices, with the elements p_{ij} well-defined approximations of local-process behavior. One can then calculate the stationary measure $\{\pi_i\}$ for the approximating chain, and use the parameters $\{p_{ij}\}$ and $\{\pi_i\}$ in a variety of ways—e.g., (*i*) to establish that two processes are different, by establishing that their respective approximating chains are different, and (*ii*) to estimate true process parameters by related parameters for the approximating chain. Greater approximation accuracy (larger m and smaller r) requires greater data input, in the spirit of analogous requirements for Taylor and Fourier series.

Approximating Markov Chains and Related Parameters

For deterministic and stochastic discrete-time processes, support on some interval $[A, B]$, I define *approximating* (m, r) Markov chains. These chains are mth-order, state space $\{A + r/2, A + 3r/2, \ldots, B - r/2\}$. We assume equilibrium (stationary) behavior throughout most of this discussion. Recall the following. *Definition 1*: The Markov chain (or process) $\{X_n\}$ is of order m if the conditional probability $P\{X_n \in A_n \| X_k = a_k, k < n\}$ is independent of the values a_k for $k < n - m$.

Definition 2: For a stationary discrete time, continuous state-space stochastic process, with $A \leq X_n \leq B$ almost surely, define an approximating (m, r, A, B) Markov chain as follows: (*a*) Divide $[A, B]$ into $(B - A)/r$ cells; the ith cell $C(i) = [x, x + r)$, where $x = A + (i - 1)r$; (*b*) define $\text{mid}(i) = A + (i - 1)r + r/2$; (*c*) define $p_{i\text{vect},j}$ for all length m vectors of integers $i$vect and integers j, $i\text{vect} = (i_1, i_2, \cdots, i_m)$, $1 \leq i_k \leq (B - A)/r$ for all k, $1 \leq j \leq (B - A)/r$. $p_{i\text{vect},j}$ = {conditional probability that $X_k \in C(j)$, given that $X_{k-1} \in C(i_1)$, $X_{k-2} \in C(i_2)$, . . . , and $X_{k-m} \in C(i_m)$}. By stationarity, this probability is constant for all k. When probability $\{X_{k-1} \in C(i_1)$,

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.

Abbreviation: ApEn, approximate entropy.

$X_{k-2} \in C(i_2), \ldots,$ and $X_{k-m} \in C(i_m)\} = 0$, define $p_{ivect,j} = 0$; and (*d*) define the approximating (m, r, A, B) Markov chain Y_i as an mth-order chain, state space $\{\text{mid}(i)\}$, by the above transition probabilities: $P\{Y_k = \text{mid}(j) \| Y_{k-1} = \text{mid}(i_1), \ldots,$ and $Y_{k-m} = \text{mid}(i_m)\} := P\{X_k \in C(j) \| X_{k-1} \in C(i_1), \ldots,$ and $X_{k-m} \in C(i_m)\}$.

In instances in which A and B are tacitly set, I refer to the above as approximating (m, r) Markov chains and make an analogous definition for deterministic processes. First, we have the following.

Definition 3: A deterministic process $\{X_n\}$ is of order m if for all n, $X_n = f(X_{n-1}, X_{n-2}, \ldots, X_{n-m})$, with f a single-valued function.

Such processes arise, e.g., in explicit time-step approximations to mth-order differential equations. For the next definition, a stationary process is required, which we form as follows. Step A: for an order m deterministic process, assign $\{X_1, X_2, \ldots, X_m\}$ to have a joint probability distribution given by a (selected) stationary measure for f. Step B: for all $k > m$, define $X_k = f(X_{k-1}, X_{k-2}, \ldots, X_{k-m})$. $\{X_n\}$ is then a stationary stochastic process. We typically are interested in the *physical* (Kolmogorov) stationary measure (ref. 5, p. 626) for f. For a wide class of deterministic processes, this measure is unique, given by well-defined time averages.

Definition 4: For a deterministic process of order m, with a preselected stationary measure and with $A \le X_i \le B$, define an approximating (m, r, A, B) Markov chain by (*i*) forming an associated stationary stochastic process by using steps A and B and then (*ii*) applying *Definition 2*.

I also consider an alternative Markov chain approximation to a deterministic map f. This *transient* chain can be calculated analytically from f, without knowledge of the stationary measure. Transition probabilities are defined by the *fraction* of the Lebesgue measure *of the image* of the conditioning set that intersects a given interval. We assume throughout the minimal restriction on f that there exists no collection of cells on which f is constant. This ensures nonzero denominators and hence well-defined $p_{ivect,j}$'s, in the following.

Definition 5: A transient (m, r, A, B) Markov chain for a deterministic map f of order m is defined on the same state space as for *Definitions 2* and *4*. The terminology and *definitions* (*2a*, *2b*, and *2d*) apply but the conditional probabilities are formed differently: (*c*) $p_{ivect,j} = \lambda(C(j) \cap f(C(i_1), C(i_2), \ldots, C(i_m)))/\lambda(f(C(i_1), C(i_2), \ldots, C(i_m)))$, where λ is the Lebesgue measure on $\mathbf{R}$.

Definition 6: Approximate entropy (ApEn) is defined as follows. Fix $r > 0$ and m a positive integer. Given a realization $\{x_i\}$ of a process $\{X_i\}$, define $v_i = (x_i, x_{i+1}, \ldots, x_{i+m-1})$. Define $C_i^m(r)$: = (number of $1 \le j \le N - m + 1$ such that $d[v_i, v_j] \le r)/(N - m + 1)$, where we define $d[v_i, v_j] = \max_{k=1,2,\ldots m}(|x_{i+k-1} - x_{j+k-1}|)$. Define $\Phi^m(r) = (N - m + 1)^{-1} \sum_{i=1}^{N-m+1} \log C_i^m(r)$, and if it exists almost surely, $\text{ApEn}(m, r) = \lim_{N\to\infty} [\Phi^m(r) - \Phi^{m+1}(r)]$.

ApEn has been developed as an efficient parameter of complexity, with both theoretical (6) and clinical utility (8–11) demonstrated for 1000 data points. Since it is generally finite, ApEn provides the capacity to distinguish many processes that Kolmogorov–Sinai entropy cannot distinguish (6), including correlated stationary stochastic processes. Since an mth-order Markov chain is a first-order chain, suitably recast, theorem 3 of ref. 6 can be applied to approximating (m, r) Markov chains. Let $\Gamma := \{\text{mid}(i)\}$, $i = 1, \ldots, (B - A)/r$. Define $\Gamma^m :=$ {all sequences of vectors $(i_1, i_2, \ldots, i_m)$ with $i_k \in \Gamma$ for each k}. We then immediately deduce *Theorem 1*; in this discrete setting, the right-hand side of Eq. 1 is well-known to information theorists as the entropy rate.

THEOREM 1. *For an approximating* (m, r, A, B) *Markov chain with* s < r, *almost surely*

$$ApEn(\text{m, s}) = -\sum_{ivect\in\Gamma^m}\sum_{j\in\Gamma} \pi(ivect)p_{ivect,j}\log(p_{ivect,j}), \quad [1]$$

where π *is a stationary measure for this Markov chain.*

Example 1—An (m, r) approximating chain for independent, identically distributed uniform random variables $\{X_i\}$ on [0, 1]: The state space $\Gamma = \{r/2, 3r/2, \ldots, 1 - r/2\}$, and the transition probabilities $p_{ivect,j} = 1/r$ for all $ivect \in \Gamma^m$ and $j \in \Gamma$.

Example 2—The chaotic map $f(x) = 3.6x(1 - x)$ on [0, 1]: Transition probability matrices for both the approximating (1, 1/10) Markov chain (MAT) and the transient (1, 1/10) Markov chain (TMAT) for $f(x)$ are shown in Table 1 with the (i, j)th entry corresponding to a transition from $C(i)$ to $C(j)$. Stationary probabilities are: (0 0 0 0.217 0.147 0.129 0.007 0.048 0.452 0) for MAT and (0 0 0 0.123 0.162 0.138 0.060 0.108 0.409 0) for TMAT. As shown in *Theorem 3*, stationary probabilities of $\{\text{mid}(i)\}$ for MAT agree with time-average probabilities for the $\{C(i)\}$ given by iterations of $f(x)$. To ensure that all rows have probabilities that sum to 1 in MAT, we should delete cells from the state space with 0 stationary probability.

Convergence of Approximating Chains

The (m, r) approximating chains can be used to "solve" deterministic and stochastic mth-order difference equations. The orientation is computational; we are interested in moments of system variables, the percentage of time spent in prescribed domains, and measures of correlation between contiguous observations. Stationary measures provide this information, so they become the objects of study. Below, it is shown that under general conditions, stationary measures for the approximating (m, r) chains and the transient (m, r) chains converge weakly to stationary measures for a given mth order process as $r \to 0$. We can thus estimate much about the behavior of a dynamical system by using straightforward approximating chain computations. Weak convergence results are most interesting in chaotic settings, where some neighboring orbits ultimately diverge. Transient information does not make sense for such systems, but we can still inquire about $\pi_1(A)$, the probability spent in A, or $\pi_3(Z)$, where $Z =$

Table 1. Transition probability matrices for the approximating (1, 1/10) Markov chain and the transient (1, 1/10) Markov chain for $f(x)$

MAT

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0.223	0.777	0
0	0	0	0	0	0	0	0	1.0	0
0	0	0	0	0	0	0	0	1.0	0
0	0	0	0	0	0	0	0	1.0	0
0	0	0	0	0	0.864	0.136	0	0	0
0	0	0	0.480	0.327	0.193	0	0	0	0
0	0	0	0	0	0	0	0	0	0

TMAT

0.309	0.309	0.309	0.073	0	0	0	0	0	0
0	0	0	0.302	0.396	0.302	0	0	0	0
0	0	0	0	0	0.133	0.556	0.311	0	0
0	0	0	0	0	0	0	0.407	0.593	0
0	0	0	0	0	0	0	0	1.0	0
0	0	0	0	0	0	0	0	1.0	0
0	0	0	0	0	0	0	0.407	0.593	0
0	0	0	0	0	0.133	0.556	0.311	0	0
0	0	0	0.302	0.396	0.302	0	0	0	0
0.309	0.309	0.309	0.073	0	0	0	0	0	0

$\{X_n \in A_1 \,\&\, X_{n+1} \in A_2 \,\&\, X_{n+2} \in A_3\}$, the probability that three successive observations fall into prescribed sets. We now indicate an analytic method for estimating the latter quantity above for the map $x_{n+1} = 3.6x_n(1 - x_n)$, with $A_1 = [0.8, 0.9]$, $A_2 = [0.4, 0.5]$, and $A_3 = [0.8, 0.9]$; the theorems below establish the convergence of the estimate to the true value as the mesh size goes to 0. By computer experiment based on 10^6 points, $\pi_3(Z) = 0.146$. For the transient (1, 0.1) chain approximation, $\pi_{TR}(Z) = \Pr(X_{n+2} \in [0.8, 0.9] \| X_{n+1} \in [0.4, 0.5] \,\&\, X_n \in [0.8, 0.9])\Pr(X_{n+1} \in [0.4, 0.5] \| X_n \in [0.8, 0.9])\pi_{TR}(X_n \in [0.8, 0.9])$, where Pr and π_{TR} refer to the approximating chain. Since [0.4, 0.5] and [0.8, 0.9] are single atoms in the (1, 0.1) partition, $\pi_{TR}(Z) = p_{59}\, p_{95}\, \pi_{TR}(9)$, with cell i corresponding to $[0.1(i - 1), 0.1i]$. From *Example 2*, we can conclude that $\pi_{TR}(Z) = (1.0)(0.396)(0.409) = 0.162$. Note that $\pi_{TR}(Z)$ is much closer to $\pi_3(Z)$ than is $\pi_1(A_1)\pi_1(A_2)\pi_1(A_3) = (0.452)(0.147)(0.452) = 0.030$, the latter the product of exact steady-state probabilities, which would equal $\pi_3(Z)$ if successive observations were uncorrelated. To analytically approximate the measure of an n-time unit event, (*i*) calculate all $p_{\text{ivect},j}$ for a transient chain approximation TR for the given system, (*ii*) compute all π_{ivect} by raising TR to a high power, and (*iii*) derive all n-time unit probabilities by using the Chapman–Kolmogorov equations.

Below, the space A will always be a compact subset of $\mathbf{R}^m$. I wish to compare stochastic processes, particularly Markov processes, to one another and to deterministic maps, and so consider a space of transition probabilities on A, Tr(A).

Definition 7: Given $A \subset \mathbf{R}^m$, define $t \in \mathrm{Tr}(A)$, t = {probability measures t_a on A for all $a \in A$, such that for all $B \subset A$ Borel measurable, the map $a \to t(a, B)$ is a Borel-measurable function}.

For a (deterministic) function $f: A \to A$, $(t_f)_a$ will be the point mass $\delta_{f(a)}$ for all a.

Definition 8—Action of Tr(A): For $t \in \mathrm{Tr}(A)$ and μ, a measure on A, define $t * \mu$ as a measure on A given by $\int f(z)d(t * \mu)(z) = \int [\int f(y)t(z, dy)]d\mu(z)$, for all Borel-measurable functions f.

Definition 9: A probability measure μ on A is *stationary* for $t \in \mathrm{Tr}(A)$ if $t * \mu = \mu$.

Definition 9 agrees with the standard notion for deterministic f and for Markov chains. By lemma 1.2 of ref. 12, there exists at least one such stationary measure for $t \in \mathrm{Tr}(A)$. Below, I do not presume absolute continuity of stationary measures with respect to Lebesgue measure; typically these measures are singular. For $\{t_n\}$, $t \in \mathrm{Tr}(A)$, we say that t_n converges to t if whenever μ_n converges to μ weakly on A, then $t_n\mu_n$ converges to $t\mu$ weakly on A.

The next, central lemma requires that two conditions be satisfied to conclude weak convergence. These conditions are often easy to verify (see below), ensuring applicability.

LEMMA 1 *(weak convergence of processes). Assume we are given* A *a compact subset of* $\mathbf{R}^n$ *and transition probabilities* t_n *and* t *on Tr*(A). *Furthermore, assume the following.*

Condition A: The map on $A: a \to t(a, dy)$ is weakly continuous [i.e., $\int_A f(y)\, t(a, dy)$ is a continuous function of a for every continuous bounded f on A].

Condition B (uniformity in weak convergence): Given any $\delta > 0$ and any bounded continuous f, there exists N such that for all $n > N$ and for all $a \in A$, $|\int_A f(y)t_n(a, dy) - \int_A f(y)t(a, dy)| < \delta$. Then t_n converges to t.

Proof: Choose probability measures ν_n that converge weakly to ν on A. Choose f continuous; then $|\int f\, d(t_n * \nu_n) - \int f\, d(t * \nu)| \le |\int f\, d(t_n * \nu_n) - \int f\, d(t * \nu_n)| + |\int f\, d(t * \nu_n) - \int f\, d(t * \nu)|$. For the first term on the right-hand side, $|\int f\, d(t_n * \nu_n) - \int f\, d(t * \nu_n)| = |\int [\int f(y)t_n(z, dy)]d\nu_n(z) - \int [\int f(y)t(z, dy)]d\nu_n(z)| \le \int \sup |\int_A f(y)t_n(z, dy) - \int_A f(y)t(z, dy)|d\nu_n(z)$, which, by *Condition B*, converges to 0 as $n \to \infty$. For the second term on the right-hand side, $|\int f\, d(t * \nu_n) - \int f\, d(t * \nu)| = |\int [\int f(y)t(z, dy)]d\nu_n(z) - \int [\int f(y)t(z, dy)]d\nu(z)|$. The integral in brackets is a continuous function of z, by *Condition A*, hence the second term converges to 0 as $n \to \infty$ by the weak convergence of ν_n to ν. Therefore $t_n * \nu_n$ converges weakly to $t * \nu$, hence *Lemma 1*.

THEOREM 2 *(weak convergence of stationary measures). Assume* A *is a compact subset of* $\mathbf{R}^n$, *transition probabilities* t_n *converge to* t *on Tr*(A), *and* t *has a unique stationary measure* ν *on* A. *For each* n, *choose* ν_n *stationary for* t_n *on* A. *Then* ν_n *converges weakly to* ν.

Proof: Since A is compact, the $\{\nu_n\}$ are a tight family and have a subsequence $\{\nu_{n(i)}\}$ that converges weakly to some probability measure Ψ on A (theorem 6.1 of ref. 13). I claim that Ψ is stationary for t. Since $t_{n(i)}$ converges to t, $t_{n(i)} * \nu_{n(i)}$ converges weakly to $t * \Psi$. But $t_{n(i)} * \nu_{n(i)} = \nu_{n(i)}$ by stationarity, so $\nu_{n(i)}$ converges weakly to $t * \Psi$. Since $\nu_{n(i)}$ converges weakly to Ψ, I conclude that $t * \Psi = \Psi$ (as claimed), and by uniqueness of the stationary measure for t, that $\Psi = \nu$. This establishes convergence for some subsequence. Suppose ν_n does not converge weakly to ν; then for some f in $C(A)$ and some positive ε, $|\int f\, d\nu_{n(i)} - \int f\, d\nu| > \varepsilon$ for all $\nu_{n(i)}$ in some subsequence. Mimicking the above argument, since the $\{\nu_{n(i)}\}$ are a tight family, there is a further subsequence $\{\nu_{n(i(m))}\}$ that converges weakly to a probability measure ξ on A, with ξ stationary for t. So $\xi = \nu$, contradicting the bounding away of the above by ε. This completes the proof.

To invoke *Theorem 2* directly, a limit process with a unique stationary measure is required. In general, stochastic perturbations of dynamical systems have unique stationary measures (14). We next see that we can estimate a stationary measure for a dynamical system by finding the stationary measure for an approximating chain, with small r.

THEOREM 3. *Given* f: [A, B] → [A, B], *select a stationary measure* μ *for* f *and define a* (1, r) *approximating chain* A_r *on* {*mid*(i)} *given by* Definitions 2 *and* 4. *Suppose there exists* r_0 *such that for all* $r < r_0$, A_r *is irreducible, when restricted to those cells with positive* μ *measure. Then* ν_r, *the unique stationary measure for* A_r, *converges weakly to* μ *as* $r \to 0$.

Proof: Uniqueness of ν_r on {mid(i): $\mu(C_i) > 0$} follows from the irreducibility assumption on A_r. We see the following: $\nu_r(\mathrm{mid}(i)) = \mu(C_i)$, for C_i with positive μ measure. Invoking stationarity, $\mu(C(j)) = \mu(f^{-1}C(j))$. This latter quantity $= \Sigma_i\, \mu(f^{-1}(C(j)) \cap C(i)) = \Sigma_i [\mu(f^{-1}(C(j)) \cap C(i))/\mu(C(i))]\mu(C(i)) = \Sigma_i\, p_{\mathrm{mid}(i),\mathrm{mid}(j)}\, \mu(C(i))$, for all j. Since $\nu_r(\mathrm{mid}(j)) := \mu(C_j)$ satisfies $\nu_r(\mathrm{mid}(j)) = \Sigma_i\, p_{\mathrm{mid}(i),\mathrm{mid}(j)}\, \nu_r(\mathrm{mid}(i))$, ν_r is stationary on {mid(i): $\mu(C_i) > 0$}, and by the uniqueness of ν_r, the relationship $\nu_r(\mathrm{mid}(i)) = \mu(C_i)$ is verified. To establish weak convergence of ν_r to μ, it suffices to show that the distribution functions $F_r(x) = \nu_r([A, x])$ converge to $F(x) = \mu([A, x])$ at all continuity points x of F (13). By this relationship, $F(x) - F_r(x) = \mu((\mathrm{mid}(i)_{r,x}, x])$, with $\mathrm{mid}(i)_{r,x}$ the largest midpoint in the $(1, r)$ partition $\le x$. Thus $|F(x) - F_r(x)| \le \mu((x - 1/r, x])$, which converges to 0 as $r \to 0$ since x is a continuity point of F.

To see that the irreducibility assumption is necessary, consider $f(x) = 3.6x(1 - x)$. Let a be the fixed point of f in (0, 1), and let b and c be the fixed points of f^2; $f(b) = c$ and $f(c) = b$. The measure $(\delta_a + \frac{1}{2}(\delta_b + \delta_c))/2$ is stationary for f. For sufficiently small r, the approximating chain for this measure is supported on three points, with nonzero transition probabilities $p_{a,a} = 1$, $p_{b,c} = 1$, and $p_{c,b} = 1$ (associating a, b, and c with the respective cell midpoints). For each r, choose δ_a as a stationary measure for the approximating chain. Then the weak $\lim_{r \to 0} \delta_a \ne (\delta_a + \frac{1}{2}(\delta_b + \delta_c))/2$.

For many dynamical systems, including irreducible axiom A systems (15), unique physical measures exist (16) and agree with Sinai–Ruelle–Bowen (SRB) measures (17). If no physical measure exists, there is no ergodic behavior (5), a terrible state of affairs (18). Fortunately, in both computer experiments and the physical world, a small, "uncertain" pertur-

bation is a component in system evolution. The result below verifies the convergence of transient approximating chains to the true process on Tr(*A*). We expect that for topologically transitive axiom A systems, the corresponding unique stationary measures converge weakly to the physical measure of *f*.

THEOREM 4. *Given* f:[A, B] → [A, B] *continuous and* A_r *a transient Markov chain for* f, *define* $t_r \in Tr([A, B])$, *recalling the notation of* Definitions 2 *and* 5: *for* a ∈ [A, B], *choose the unique* i *with* a ∈ C(i). *Define* $(t_r)_a$ *to agree with* $p_{i,j}$ *for* A_r; *namely* $(t_r)_a$ *has the atomic probability distribution* $\Sigma_j\, p_{i,j}\, \delta_{mid(j)}$. *Then* t_r *converges to* t_f *on* Tr[A, B] *as* r → 0.

Proof: We verify *Conditions A* and *B* of *Lemma 1*. For *Condition A*, choose *g* continuous. Then $\int_{[A,B]} g(y)\, t_f(a, dy) = g(f(a))$, continuous in *a* since *f* is continuous. For *Condition B*, choose continuous *g* and $\delta > 0$. Since [*A*, *B*] is compact, *g* is uniformly continuous, hence there exists ξ such that $|x - y| < \xi$ implies that $|g(x) - g(y)| < \delta$. Since *f* is uniformly continuous, there exists *s* such that $|x - y| < s$ implies that $|f(x) - f(y)| < \xi/2$. Choose $r < \min(s, \xi)$. Then for arbitrary $a \in A$, $|\int_{[A,B]} g(y)t_r(a, dy) - \int_{[A,B]} g(y)t_f(a, dy)| = |(\Sigma_j p_{i,j}\, g(\text{mid}(j))) - g(f(a))|$. Observe that if $p_{i,j}$ for $(t_r)_a$ has nonzero mass, then $C(j) \cap f(C(i))$ is nonempty, and hence $|\text{mid}(j) - f(a)| \le r/2 + |f(x) - f(a)|$, for some $x \in C(i)$. For this *x*, since $|x - a| \le r < s$, $|f(x) - f(a)| < \xi/2$. Hence $|\text{mid}(j) - f(a)| < (r/2) + \xi/2 < \xi$, and thus $|g(\text{mid}(j)) - g(f(a))| < \delta$. This establishes *Condition B* and *Theorem 4*.

Since t_r and A_r are identical upon iteration, they have identical stationary measures, supported on the {mid(*i*)}, that are straightforward to calculate. As in *Theorem 3*, we require irreducibility on recurrent states of the transient chains for convergence of these stationary measures to a unique physical measure. This irreducibility holds for general classes of Markov chains; in *Example 2*, the recurrent states are {mid(*i*), $4 \le i \le 9$}, and these are easily seen to communicate with one another.

As shown in Fig. 1, even coarse mesh approximating and transient chains can produce good estimates for the physical measure of a dynamical system. Stationary probability distribution functions are shown for the physical measure for the map $x_{n+1} = 3.6x_n(1 - x_n)$, for the (1, 0.1) and (1, 0.04) approximating Markov chains and for the (1, 0.1) transient Markov chain. To generate this figure, I assumed a uniform density on each cell *C*(*i*), rather than a point mass at {mid(*i*)}. The (1, 0.04) approximating chain produces a very accurate stationary measure approximation; observe also that the (1, 0.1) approximating chain produces a more accurate estimate of the true process stationary measure than does the (1, 0.1) transient chain, as expected.

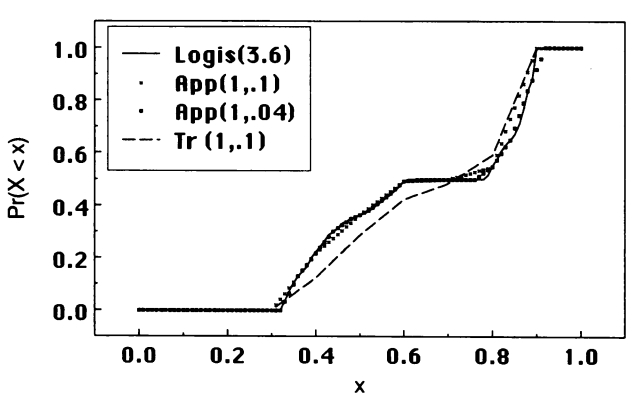

FIG. 1. Stationary probability distribution functions for the physical measure for the logistic map $x_{n+1} = 3.6x_n(1 - x_n)$ [Logis (3.6)], the (1, 0.1) and (1, 0.04) approximating Markov chains [App(1, .1) and App (1, .04), respectively], and the (1, 0.1) transient Markov chain [Tr(1, .1)].

The weak convergence results above are true much more generally than for uniform partitions, which were chosen for pedagogic simplicity. *Theorem 5* generalizes *Theorem 4* for transient approximating chains. The proof, omitted, is virtually identical to that for *Theorem 4* (recall that *Lemma 1* has been established for $A \subset \mathbf{R}^m$). We say a finite collection $\{P_i\}$ of subsets of $\mathbf{R}^m$ form a partition of *A* if *A* is the union of the mutually disjoint connected P_i.

THEOREM 5. *Assume* f:A → A *is continuous on compact* A ⊂ $\mathbf{R}^m$. *Choose a sequence of partitions* $\{P_i\}_r$ *of* A *for* r > 0 *such that dsup*(r) = $max_i diam(P_i)_r \to 0$ *as* r → 0. *For each* i *and* r, *choose an arbitrary element* $(m_i)_r \in (P_i)_r$. *For each* r, *define a transient Markov chain for* f *on* $\{(m(i))_r\}$, *as in* Definition 5, *by the percentage of the image of a cell* P_i *that intersects each specified cell* P_j. *Define* $t_r \in Tr(A)$: *for* a ∈ A, *choose the unique* i *with* a ∈ $(P_i)_r$. *Define* $(t_r)_a$ *to agree with* $p_{i,j}$ *for* A_r: $(t_r)_a$ *has the atomic probability distribution* $\Sigma_j\, p_{i,j}\, \delta_{m(j)}$. *Then* t_r *converges to* t_f *on* Tr(A) *as* r → 0.

This result allows us to apply approximating chains to first-order Markov maps *f*:$\mathbf{R}^m$ to $\mathbf{R}^m$ (e.g., the Henon map, *m* = 2). The standard procedure of turning a *k*th-order Markov process *f*:$\mathbf{R}^m$ to $\mathbf{R}^m$ into a first-order process f_{assoc}:$\mathbf{R}^{mk}$ to $\mathbf{R}^{mk}$ implies that *k*th-order Markov processes on $\mathbf{R}^m$ can be estimated by approximating chains, with weak convergence as dsup → 0.

The three-step approximation performed at the beginning of this section is justified by *Theorem 6*. For $t \in \text{Tr}(A)$ and μ stationary for *t*, set $\mu_1 = \mu$ and define μ_m as a measure on A^m, $m > 1$, by $\int g(x_1, x_2, \ldots, x_m)d\mu_m(x_1, x_2, \ldots, x_m) = \int \ldots [\int[\int g(x_1, x_2, \ldots, x_m)t(x_{m-1}, dx_m)]t(x_{m-2}, dx_{m-1})]. \ldots d\mu(x_1)$, for Borel-measurable *g*. The proof (omitted) follows from a straightforward recursive argument, repeatedly applying *Conditions A* and *B* of *Lemma 1*.

THEOREM 6. *Let* A_r *be the transient* (1, r) *chain for* f:[A, B] → [A, B], *and assume there exists* r_0 *such that for all* $r < r_0$, A_r *is irreducible when restricted to recurrent cells. Assume that* ν_r, *the unique stationary measures for* A_r, *converge weakly to* ν_{phys} *as* r → 0 *and that* ν_{phys} *is nonatomic. Fix* m *and subintervals* {C(i)}, i = 1, . . . , m *of* [A, B]; *denote* D := C(1) × C(2) . . . , × C(m). *Then* $\nu_{r,m}(D) \to \nu_{phys,m}(D)$ *as* r → 0.

Next, consider the evolution of dynamical systems with control parameter. For perturbed dynamical systems, we generally have weak continuity of stationary measures. First, for continuous *f*:*A* → *A*, *A* ⊂ $\mathbf{R}^m$, define $f^\varepsilon \in \text{Tr}(A)$, a uniform perturbation of magnitude ≤ ε. Pick $a \in \mathbf{R}^m$; define the measure f^ε_a by the density function $\varphi_a(z) = K_a$ for $z \in A$ such that $\|z - f(a)\| < \varepsilon$, $\varphi_a(z) = 0$ otherwise, with ‖.‖ the Euclidean metric on $\mathbf{R}^m$ and K_a = 1/(the *m*-dimensional volume of (*A* ∩ the ε-ball around *a*)).

THEOREM 7. *Assume a family of dynamical systems is given by continuous* f_x:A → A, *where* f_s *converges uniformly on* A *to* f_r. *Choose* ε *and assume that* f^ε_r *has unique stationary measure* ν_r. *For each* s, *choose a stationary measure* ν_s *for* f^ε_s; *then* ν_s *converges weakly to* ν_r.

Proof: We verify *Conditions A* and *B* of *Lemma 1*. *Condition A* follows from the continuity of f_r. For *Condition B*, choose *g* continuous on *A* and $\delta > 0$. Define $K = \sup_{a \in A} K_a$ (finite, since K_a is continuous on a compact set). Since *A* is compact, *g* is uniformly continuous, and hence there exists τ such that $|x - y| < \tau$ implies that $|g(x) - g(y)| < \delta/K$. By uniform convergence of f_s to f_r, there exists ω such that $|r - s| < \omega$ implies $|f_s(a) - f_r(a)| < \tau$, for all $a \in A$. Then for arbitrary $a \in A$, and $|r - s| < \omega$, $|\int_A g(y)t_s(a, dy) - \int_A g(y)t_r(a, dy)| \le K \int_{\|x\| < \varepsilon} |g(f_s(a)+x) - g(f_r(a) + x)|\, dx \le K \sup_{x \in A} |g(f_s(a) + x) - g(f_r(a) + x)| \le K(\delta/K) = \delta$. This establishes *Condition B* and, by *Theorem 2*, the desired result.

Most familiar dynamical systems satisfy the uniform convergence assumption, and hence perturbations of these systems generally fall under the aegis of *Theorem 7*. Weak convergence with control parameter implies continuity of all

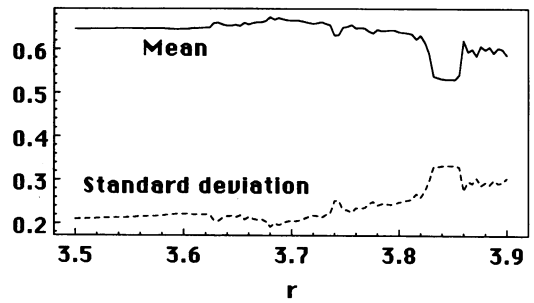

FIG. 2. Mean and standard deviation vs. control parameter r for the logistic map $x_{n+1} = rx_n(1 - x_n)$.

moment statistics, since they are integrals of polynomials with respect to the stationary measure [e.g., the mean = $\int_A x\, d\nu(x)$]. This is illustrated in Fig. 2, which demonstrates the continuity of the time-averaged mean and standard deviation as a function of the control parameter for the logistic map. Since these calculations are performed by computer, they yield statistics for a slightly perturbed version of the logistic map, which fits the context of *Theorem 7*. Compare the perspectives of the analyst/statistician and the topologist; the topologist sees structural change and instability with control parameter, through bifurcations and into chaotic regions, while the analyst sees continuity.†

I demonstrate weak convergence across bifurcations for the map $x_{n+1} = rx_n(1 - x_n)$ near $r = 3.0$, where the system changes from a single limit point to a period 2 limit. For $r < 3.0$, $s = 1 - 1/r$ is an attracting fixed point with physical measure = $\delta_{s(r)}$. For $3.0 < r < 3.2$, $s = 1 - 1/r$ is a repelling fixed point. The other fixed points of f^2, t and $u = [(1 + 1/r) \pm \sqrt{(1 - 2/r - 3/r^2)}]/2$ are attracting, and thus the physical measure = $\frac{1}{2}(\delta_{t(r)} + \delta_{u(r)})$. Since $t(r)$ and $u(r)$ both converge to $s(r) = 2/3$ at $r = 3.0$, weak convergence follows. This can be seen from the bifurcation diagram: weak convergence follows from the connectivity of the graph of the physical measure, in the multiply periodic domain.

A Parameter for Turbulence

The study of turbulence has long been an enigma. A commonly used measure for turbulence is the Reynolds number, which has at least two deficiencies: (*i*) there is an artificial length scale imposed in the formula, and (*ii*) interpretation of the amount of turbulence given by a set value of the Reynolds number seems to be heavily shape dependent. I propose ApEn as a measure of turbulence, given a grid (partition) and a variable of interest (e.g., pressure or x-component of velocity). Once a grid and a variable have been set, define

$$\text{ApEn}(m, \text{grid}): = -\sum_{i\text{vect}\in\Gamma^m}\sum_{j\in\Gamma} \pi(i\text{vect})p_{i\text{vect},j}\log(p_{i\text{vect},j}), \quad [2]$$

with state space Γ the collection of cells. This can be generalized to n-tuples of variables by considering Γ^n as the state space. Numerical estimation of Eq. 2 from data is inexpensive, since stationary and transition counts on a grid specify this estimate. Eq. 2 captures both the stationary distribution of the flow via π and the transient (mix*ing*) effects of the flow given by the $p_{i\text{vect},j}$. Thus we distinguish the well-mixed, completely stagnant system (ApEn = 0) from the well-mixed, actively mixing system (ApEn > 0).

It is tempting to form a measure of "turbulence" by letting the grid size converge to 0 in Eq. 2, to speak of a parameter without reference to a specified grid. There is an important reason not to do so, in addition to computational cost: if there exists some ε below which process behavior cannot be ascertained, any relationship between a converged value of Eq. 2 and a true process value is coincidence. So the notion of which process of two is more "random" or complex should be tied to the choice of partition. In practice, there often appears to be a nice consistency across meshes; if ApEn(m, r)(A) ≥ ApEn(m, r)(B), then ApEn(n, s)(A) ≥ ApEn(n, s)(B) for many choices of n and s. Since the global ApEn parameter is aggregating heterogeneous local information, there is no reason to expect this behavior in general.‡

†Geometric changes in attractors appear to be manifested in the differentiability of a "weak integral" as a function of the control parameter. Thus, e.g., the mean in Fig. 2 is piecewise smooth in the multiply periodic domain and nondifferentiable at bifurcations and at returns to periodicity from chaos, most blatantly realized near 3.828, where the logistic map changes from chaotic to periodic, period 3.

‡A "flip-flop pair of processes" can be constructed to establish that even given no noise and infinite data, the determination of which of two processes is more random, turbulent, or complex must be tied to partition choice. For deterministic processes, in theory the Ornstein–Weiss guessing scheme (7) can be applied to estimate the Kolmogorov–Sinai entropy and the limiting value of Eq. 2, given *no* process noise.

For discussions that gave perspective, I thank R. Burton, T. L. Lai, P. Jones, D. Ornstein, S. R. S. Varadhan, and L. Shepp; for inspiration, I thank M. J. Minkin and W. Mozart (Clemenza di Tito).

1. Schaffer, W. M. & Kot, M. (1985) *J. Theor. Biol.* **112**, 403–427.
2. Sugihara, G. & May, R. M. (1990) *Nature (London)* **344**, 734–741.
3. Lapidus, L. & Pinder, G. F. (1982) *Numerical Solution of Partial Differential Equations in Science and Engineering* (Wiley, New York).
4. Bender, C. M. & Orszag, S. A. (1978) *Advanced Mathematical Methods for Scientists and Engineers* (McGraw–Hill, New York), pp. 61–318.
5. Eckmann, J. P. & Ruelle, D. (1985) *Rev. Mod. Phys.* **57**, 617–656.
6. Pincus, S. M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 2297–2301.
7. Ornstein, D. S. & Weiss, B. (1990) *Ann. Probab.* **18**, 905–930.
8. Pincus, S. M., Gladstone, I. M. & Ehrenkranz, R. A. (1991) *J. Clin. Monitoring* **7**, 335–345.
9. Pincus, S. M. & Keefe, D. L. (1992) *Am. J. Physiol.*, in press.
10. Pincus, S. M. & Viscarello, R. R. (1992) *Obstet. Gynecol.* **79**, 249–255.
11. Kaplan, D. T., Furman, M. I., Pincus, S. M., Ryan, S. M., Lipsitz, L. A. & Goldberger, A. L. (1991) *Biophys. J.* **59**, 945–949.
12. Furstenberg, H. (1963) *Trans. Am. Math. Soc.* **108**, 377–428.
13. Billingsley, P. (1968) *Convergence of Probability Measures* (Wiley, New York).
14. Bhattacharya, R. N. & Waymire, E. C. (1990) *Stochastic Processes with Applications* (Wiley, New York), pp. 214–215.
15. Smale, S. (1967) *Bull. Am. Math. Soc.* **73**, 747–817.
16. Kifer, Y. I. (1974) *Math. USSR Izv.* **8**, 1083–1107.
17. Young, L. S. (1990) *Trans. Am. Math. Soc.* **318**, 525–543.
18. Newhouse, S. (1979) *Publ. Math. IHES* **50**, 102–151.