To appear in the
Special Issue on Complexity
of **Chaos, Solitons, and Fractals**,
W. Ebeling, editor (1993).

SFI Technical Report 93-05-028

Fluctuation Spectroscopy

Karl Young
Space Sciences Division*
NASA Ames Research Center
Mail Stop 245–3, Moffet Field, California 94035 USA

James P. Crutchfield
Physics Department†
University of California
Berkeley, California 94720 USA

Abstract

We review the thermodynamics of estimating the statistical fluctuations of an observed process. Since any statistical analysis involves a choice of model class — either explicitly or implicitly — we demonstrate the benefits of a careful choice. For each of three classes a particular model is reconstructed from data streams generated by four sample processes. Then each estimated model's thermodynamic structure is used to estimate the typical behavior and the magnitude of deviations for the observed system. These are then compared to the known fluctuation properties. The type of analysis advocated here, which uses estimated model class information, recovers the correct statistical structure of these processes from simulated data. The current alternative — direct estimation of the Renyi entropy from time series histograms — uses neither prior nor reconstructed knowledge of the model class. And, in most cases, it fails to recover the process's statistical structure from finite data — unpredictability is overestimated. In this analysis, we introduce the fluctuation complexity as a measure of a process's total range of allowed statistical variation. It is a new and complementary characteristic in that it differs from the process's information production rate and its memory capacity.

* KY's Internet address is karl@gojira.berkeley.edu
† JPC's Internet address is chaos@gojira.berkeley.edu

Contents

Figures iii
List of Tables vii
1 Introduction 1
2 Sequence Distributions and Fluctuations 3
2.1 Introduction and Definitions 3
2.2 Fluctuation Spectra Directly From Sequence Histograms 4
2.3 Large Deviation Theory 6
2.4 Renyi Entropy 7
2.5 Fluctuation Spectra from Renyi Entropy 9
2.6 Large Deviations and Free Energy 11
3 ϵ-Machines 12
3.1 Introduction and Definitions 12
3.2 Data Sources 14
3.3 ϵ-Machine Thermodynamics 18
3.4 Thermodynamic Complexities 21
4 Estimated Fluctuation Spectra 23
4.1 Histograms as Models 24
4.2 Empirical ϵ-Machines 28
4.3 Spectroscopy 32
5 Conclusions 39
6 Acknowledgments 40
Appendix A Thermodynamic Details 42
A.1 Biased Coin Process 42
A.2 Golden Mean Process 44
A.3 Even Process 46
Bibliography 47

Figures

1 Labeled, directed graph with transition probabilities for modeling tosses of a biased coin. The branching probabilities are $\Pr(s = 0) = .4$ and $\Pr(s = 1) = .6$. The vertices of the graph represent the "knowledge states" of the process as discussed in the text; in this case there is only one, $\mathbf{V} = \{\mathbf{A}\}$. We represent the start state or "state of total ignorance" with a double circle. State-to-state transitions are represented by the graph edges. These are labeled $s|p$, where $s \in \mathcal{A}$ is a symbol in the measurement alphabet and $p = p_{v_i \underset{s}{\rightarrow} v_j}$ is a transition probability. 16

2 The golden mean process — so-called since the growth rate of the number of sequences as a function of length is the logarithm of the golden mean. In fact, the total number of sequences at length L is given by the Fibonacci number F_{L+2}. In simplest terms, the golden mean process generates all binary sequences except those containing two consecutive 0s. We have chosen a particular statistical bias so that, for example, $\Pr(s = 1|v = \mathbf{A}) = 0.6$. See Figure 1 for explanation of the representation. 17

3 The even process generates all binary sequences in which 1's occur in even length blocks bounded by 0's. The statistical bias is set so that $\Pr(s = 1|v = \mathbf{A}) = 0.6$. See Figure 1 for explanation of the representation. . . 17

4 Biased coin process: Mosaic of sequence histograms for sequences of lengths $L \in [1, 9]$. (After .) Each histogram plots $\log_2 P(s^L)$ versus s^L, where $P(s^L)$ is the probability density and s^L is evaluated as a binary fraction. Each histogram was obtained from a data stream consisting of a binary sequence of length $k = 10^7$ generated by a random walk through the stochastic machine shown in Figure 1. The random walk is biased according to the transition probabilities. The self-similar structure of the distribution is easily discernible. And this suggests that the fluctuation spectrum will be easy to model for the biased coin process. . . 25

5 Golden mean process: Sequence histogram mosaic as in Figure 4 but obtained from a data stream consisting of a binary sequence of length $k = 10^7$ generated by a random walk through the machine of Figure 2. Compared to the biased coin process, the scaling behavior is visually more complicated; though some regularities in the bin heights and in the distribution's support are discernible across different sequence lengths. Here there are excluded sequences seen as "holes" in the distribution's support. These occur in bins associated with the set of sequences containing subsequences $\mathbf{w} \in \{00\}$. 26

6 The even process: Sequence histogram mosaic as in Figure 4 but obtained from a data stream consisting of a binary sequence of length $k = 10^7$ generated by a random walk through the labeled, directed graph shown in Figure 3. The even process is more complicated still. The sequence distribution's support consists of a countable infinity of Cantor sets, for example. 27

7 The Logistic map at the Misiurewicz parameter: Sequence histogram mosaic as in Figure 4 but obtained from a data stream consisting of a binary sequence of length $k = 10^7$ generated by observing iterates of the logistic map — Eq. (53) — with a binary measuring instrument. There seems to be little apparent scaling structure in the mosaic, either in the bin heights or in the "holes" in the support. 28

8 Reconstructed ϵ-machine for the biased coin process obtained from a binary sequence of length $k = 10^7$ generated by a random walk through the machine of Figure 1. 29

9 Reconstructed ϵ-machine for the golden mean process obtained from a binary sequence of length $k = 10^7$ generated by a random walk through the machine of Figure 2. 29

10 Reconstructed ϵ-machine for the even process obtained from a binary sequence of length $k = 10^7$ generated by a random walk through the machine of Figure 3. 30

11 The Misiurewicz machine: the ϵ-machine reconstructed from a binary sequence of $k = 10^7$ iterates of the logistic map $f(x) = rx(1 - x)$ at the Misiurewicz parameter value $r = r_M \approx 3.9277370017867516\ldots$. The symbols 0 and 1 of the measurement alphabet correspond to the left and right halves, respectively, of a binary partition of the unit interval $x \in [0, 1]$ with partition divider at $x = 0.5$. 30

12 Biased coin fluctuation spectra: The histogram spectrum (vertical lines) and the Renyi spectrum (solid curve with superposed dots) were estimated from histograms as in Figure 4, but at sequence length $L = 10$. In this and the following figures, the histogram spectrum used 300 uniform-width energy bins between $\mathcal{U}_{\min}$ and $\mathcal{U}_{\max}$. The Renyi spectrum is given over $\beta \in [-30, 30]$. The machine fluctuation spectrum (solid curve) was numerically estimated using the reconstructed machine shown in Figure 8 over a range $\beta \in [-30, 30]$. The Renyi and machine spectra are essentially identical apart from the small discrepancy for high-$\mathcal{U}$, low-probability sequences, as discussed in the text. Since the Renyi spectrum is a very good approximation to the actual fluctuation spectrum for the biased coin, the topological and metric entropies are essentially the same: $h^R = h^M = h^{R,M}$ and $h_\mu^R = h_\mu^M = h_\mu^{R,M}$, as noted in the figure. 34

13 Golden mean process fluctuation spectra: The histogram spectrum (vertical lines) and the Renyi spectrum (solid curve with superposed dots) were estimated from histograms as in Figure 5, but at sequence length $L = 10$. The Renyi spectrum is shown over the range $\beta \in [-100, 100]$. The machine fluctuation spectrum (solid curve) was numerically estimated using the reconstructed machine shown in Figure 9 over the range $\beta \in [-80, 120]$. The $\mathcal{U}$ axis is expanded as compared to the other examples due to the smaller range of fluctuations as already noted in the histogram mosaic of Figure 5. Despite adequate topological and metric entropy estimates in the Renyi spectrum, it overestimates $s(\mathcal{U})$ at both high and low $\mathcal{U}$. Note that at low energy the Renyi spectrum curve intersects the top of a histogram spectrum bin. 36
14 Even process fluctuation spectra: The histogram spectrum (vertical lines) and the Renyi spectrum (solid curve with superposed dots) were estimated from histograms as in Figure 6, but at sequence length $L = 10$. The Renyi spectrum is shown over the range $\beta \in [-70, 30]$. The machine fluctuation spectrum (solid curve) was numerically estimated using the reconstructed machine shown in Figure 10 over the range $\beta \in [-20, 20]$. The Renyi spectrum estimate of the topological and metric entropies is in substantially larger error than for the golden mean process. Worse, one sees that $h_\mu^R > h^M$. Additionally, the Renyi spectrum method overestimates $s(\mathcal{U})$ at large energy, but underestimates $s(\mathcal{U})$ at low energy. 37
15 Logistic map process fluctuation spectra: The histogram spectrum (vertical lines) and the Renyi spectrum (solid curve with superposed dots) were estimated from histograms as in Figure 7, but at sequence length $L = 10$. The Renyi spectrum is shown over the range $\beta \in [-50, 50]$. The machine fluctuation spectrum (solid curve) was numerically estimated using the reconstructed Misiurewicz machine shown in Figure 11 over the range $\beta \in [-100, 40]$. As seen for the even process the Renyi spectrum poorly approximates the topological and metric entropies and gives $h_\mu^R > h^M$. It and the direct histogram spectrum substantially overestimate $\mathcal{U}_{\min}$ and $\mathcal{U}_{\max}$. They also overestimate the high-energy, low-probability sequence entropy, while underestimating the low-energy, high-probability sequence entropy. 38

List of Tables

Table 1 — The second column gives the connection matrices T, as defined in Eq. (42), for the three discrete-state processes discussed in this section. Each connection matrix represents a Markov chain with elements corresponding to transition probabilities between the process's states. The third column lists the left eigenvector, defined by Eq. (46) and normalized in probability, associated with the largest eigenvalue — which is unity for stochastic matrices. The elements are the asymptotic state probabilities. The appendix gives a different, "split state" representation for the biased coin. 15

Table 2 — The second and third columns give the topological and metric entropies — defined by Eqs. (44) and (47), respectively — for the three prototype discrete-state processes. The second column in the fourth row gives the topological entropy for the logistic map computed using the kneading determinant with 100 terms and estimating its smallest zero to 1 part in 10^6. As an estimate of h_μ the third column gives the Lyapunov exponent λ for the logistic map averaged over 10^8 iterates. Note that the differences $h - h_\mu$ or $h - \lambda$, i.e. the difference between the values in the second and third columns, gives a rough measure of the inhomogeneity in the asymptotic sequence distribution. The fourth and fifth columns give the ϵ-machine topological h^M and metric h_μ^M entropies, defined by Eqs. (44) and (47), for the four reconstructed machines shown in Figures 8 through 11. The Renyi estimates — h^R and h_μ^R — are also given in the next columns. All values have been rounded in the last decimal place. 31

Table 3 — The minimum and maximum energy densities for the prototype processes and estimated from the machine and Renyi fluctuation spectra for the four example processes. $\mathcal{U}_{\min}$ and $\mathcal{U}_{\max}$ values are obtained from the exact expressions given in the Appendix. The machine quantities are $\mathcal{U}_{\min}^M$ and $\mathcal{U}_{\max}^M$; the Renyi quantities are $\mathcal{U}_{\min}^R$ and $\mathcal{U}_{\max}^R$. 31

Table 4 Complexities for the original processes and those estimated from the reconstructed machines and their fluctuation spectra. C_μ is the statistical complexity from Eq. (49) for the original process. C_μ^M is the estimate obtained from the reconstructed machines. Ξ_0 and Ξ_1 are the two fluctuation complexities estimated from the reconstructed machines' fluctuation spectra. They are calculated via Eqs. (75) and (77), respectively. 32

Introduction

When investigating the behavior of a system with many degrees of freedom one is generally forced to use a statistical description. The information required to explain the behavior in mechanical terms (say) is intractably large, overwhelming both formal and quantitative analyses. At the heart of such a statistical analysis is the attempt to uncover the system's "typical" behavior as well as the likelihood of deviations from this "typical" behavior. This is the basis of recent extensions of large deviation theory and traditional statistical mechanical techniques to the study of state space distributions of dynamical systems.[1–9] In this context "typical" simply refers to the expectation value of some quantity, e.g. the energy, given a phase space distribution. More generally, in the following we will refer to deviations from an expected value as fluctuations and the set of probabilities associated with the occurrence of deviations as the fluctuation spectrum. This use of "fluctuation" connotes the variation in probability over a space of events, which is distinguished from the empirical variation in an event's estimated probability.

Always implicit in these investigations is the notion of a model. In this paper we demonstrate that, when trying to determine a process's typical behavior and fluctuations from measurement data, a crucial step is to use that data to build a model of the underlying system and to do so relative to an explicitly chosen class of models. In terms of the model classes we will discuss, there are then well-defined mathematical and numerical techniques for estimating typical behavior and fluctuation spectra which, as we shall show, are simply not obtainable directly from a finite data set alone. In particular we will focus on models determined by the recently proposed technique of ϵ-machine reconstruction.[10] Results from these models will be contrasted with sequence histogram and Renyi entropy estimation. In the most general terms, our methodology should be contrasted with the not unusual approach of first choosing a statistic — (say) box counting for the correlation dimension[11] — and then estimating it from data. This apparently common sense approach ignores the implicit model class of the statistic and of its estimation algorithm. One consequence is that systematic biases, due to the model class's inherent limitations, often become conflated with the process's intrinsic fluctuation structure.

We will assume that an essentially one-to-one representation between a process's state space trajectories and sequences of discrete symbols from a finite alphabet has been established. In accord with the general program of ϵ-machine reconstruction it is the statistics of these symbol sequences which we will study. We note that while there is a clear relationship between the techniques described here and those that study the statistics of the geodesics on a compact surface of constant negative curvature[12] — and generally to the study of the statistical structure of unstable periodic orbits of a dynamical system[13] — our techniques are not restricted to such cases.

We also wish to emphasize the distinction between studying the statistics of trajectories, as we do here, and studying fluctuations in the state space distribution.[5–9] In those cases for which

trajectories can be effectively identified with their initial conditions — such as deterministic dynamics in the absence of external noise — the results are the same. In limiting the scope of analysis to a given probability distribution over some set of events — as done by focusing only on the state space distribution — one is implicitly assuming that the events occur independently. The consequence is that correlations are ignored and temporal structure is lost. In contrast to this, a distribution over state space trajectories of a given duration allows for the determination of temporal correlations and even an approximation of the effective equations of motion.[14] Then the decay in average correlations as a function of increasing trajectory duration provides some measure of the convergence rate to an asymptotic distribution over trajectories and hence over states. And it also leads to estimates of a process's complexity.[15,16] Simply stated, a trajectory distribution determines a unique state space distribution but a state space distribution certainly does not determine a unique trajectory distribution. Going somewhat beyond this, we will show that if one has identified an approximate model, estimated within an explicit model class, then not only can one obtain direct information about correlations and convergence rates but a good estimate of the asymptotic trajectory distribution itself.

We study "data" generated by relatively simple, but illustrative processes. Complete determination of the fluctuation spectra of well-understood processes is certainly a prerequisite to analyzing experimental data sets and, as we demonstrate even for simple processes, model identification is crucial. In fact, reconstructed ϵ-machines allow not only for the determination of asymptotic fluctuation spectra but for the estimation of the statistics of symbol sequences of any length via straightforward enumeration techniques[17] and for the estimation of various complexities.

In the next section we review the basic notions of statistical fluctuations in sequence distributions and the standard Renyi entropy methods used to obtain the fluctuation spectrum. Following this, we recount those elements of ϵ-machines and their reconstruction necessary for calculating the fluctuation spectrum. At that point three model classes are identified — histograms, Renyi scaling, and ϵ-machines. Four prototypical processes are then selected as example data sources. Using data from these, models within each class are reconstructed. Finally, detailed comparisons between the prototypes' fluctuation spectra and the spectra obtained from the models are presented.

The set of techniques used for calculating the fluctuation spectra of dynamical systems is often referred to as the thermodynamic formalism.[2–4] The spirit and notation we adopt in the following emphasize not just the formal aspect, but also the direct connection with equilibrium thermodynamics. As it happens, the thermodynamic interpretation turns on a single definition that relates information to energy.

Sequence Distributions and Fluctuations

Introduction and Definitions

Consider a discrete time dynamical system $\mathcal{D}(T, X)$ with state space X and dynamic $T : X \to X$. The temporal evolution of a state $x_t \in X$ is governed by the equations of motion $x_{t+1} = Tx_t$. Starting from an initial condition $x_0 \in X$ successive application of the dynamic results in a state sequence, or trajectory, $\mathrm{x} = x_0 x_1 x_2 x_3 \ldots$. We assume that the dynamical system is observed via a finite partition $\mathcal{P}$ on the state space. This is a finite collection of non-overlapping subsets that cover X with each element labeled by a symbol $s \in \mathcal{A}$, where $\mathcal{A}$ is the "measurement" alphabet. The symbol sequence $s_0 s_1 s_2 \ldots$ associated with a given trajectory $\mathrm{x} = x_0 x_1 x_2 \ldots$ is defined by returning the label of that element of $\mathcal{P}$ in which the state x_t is at time t. We further assume that the partition is generating. This means that there is a one-to-one mapping between the set of allowed trajectories $\{\mathrm{x}\}$ and the set of observed symbol sequences $\{s_0 s_1 s_2 \ldots\}$.* We denote the set of infinite symbol sequences as $\mathcal{A}^\infty$ and the set of length L sequences as $\mathcal{A}^L$. When we have a particular realization of the dynamical system's time evolution, we will refer to the symbol sequence as a data stream $\mathrm{s} = s_0 s_1 s_2 s_3 \ldots$. In what follows we will assume the system producing the data stream is stationary; that is, neither $\mathcal{D}$ nor the probabilities of trajectories it produces depend on time. A word $\mathrm{w} = s_0 s_1 \ldots s_{L-1}$ is a string of alphabet symbols whose length L we denote by $|\mathrm{w}|$. We will usually think of a data stream s as composed of sets of subwords occurring somewhere within it. We denote the size of a set S as $\|S\| =$ card S.

We turn to the study of distributions $\Pr(\omega)$ over infinite measurement sequences $\omega \in \mathcal{A}^\infty$ where $\Pr(\omega)$ is independent of the structure of the distribution's support. We will focus on variations in sequence probability amplitude over the set of ω. Recall that we refer to these variations as fluctuations and that this use of "fluctuations" is distinct from empirical variations due to finite sampling effects, for example. By support of the distribution we simply mean the set of infinite sequences over which the probability distribution is defined. By independence of the support we mean to ignore any structure associated with the arrangement of the sequences. This is analogous to examining the properties of a probability distribution defined over a Cantor set which are independent of the Cantor set structure. For example a trivial case in the current context would be a uniform distribution over a Cantor set. No matter how intricate the structure of the Cantor set there are no fluctuations associated with the distribution, since all sequences of equal length have the same probability, and hence the fluctuation spectrum is uninteresting.

In ergodic theory parlance the set of infinite sequences which share a particular length L subsequence $\mathrm{w} = w_0 w_1 \ldots w_{L-1}, w_i \in \mathcal{A}$, is called an L-cylinder

$$s_{\mathbf{w}} = \{\omega : \omega_0 = w_0, \ldots, \omega_{L-1} = w_{L-1}; \omega \in \mathcal{A}^\infty\} \tag{1}$$

* In fact, for most of our discussion a finite-to-one mapping is permitted.

Note that the cylinder sets $s_{\mathbf{w}}$ are disjoint. Formally the cylinder measure is

$$\Pr(s_{\mathbf{w}}) = \frac{N(s_{\mathbf{w}})}{N(\mathcal{S}^L)} \tag{2}$$

where $N(s_{\mathbf{w}}) = \|s_{\mathbf{w}}\|$, $N(\mathcal{S}^L) = \left\|\mathcal{S}^L\right\|$, and $\mathcal{S}^L = \bigcup_{\mathbf{w}\in\mathcal{A}^L} s_{\mathbf{w}}$ consists of all of the L-cylinder sets. If all infinite sequences are allowed, then $\mathcal{S}^L = \mathcal{A}^\infty$ and $\Pr(s_{\mathbf{w}=0}) = \Pr(s_{\mathbf{w}=1}) = \frac{1}{2}$. The number of distinct L-cylinders allowed is denoted

$$N(L) = \left\| \left\{ \mathbf{w} : \mathbf{w} \in \mathcal{A}^L, \Pr(s_{\mathbf{w}}) > 0 \right\} \right\| \tag{3}$$

And this is the number of cylinder sets which partition $\mathcal{S}^L$.

For a finite length data stream $\mathbf{s} = s_0 s_1 \ldots s_{k-1}$, the above definitions must be modified. In particular, we associate the occurrences of a given word with the indices at which it begins within the data stream $\mathbf{s}$. In analogy with the cylinder measure, the natural estimator for a word's probability is then

$$\Pr(\mathbf{w}|\mathbf{s}) \approx \frac{N(\mathbf{w})}{N(\mathcal{S}^L)}, \quad |\mathbf{w}| = L \leq k \tag{4}$$

where $N(\mathbf{w})$ is the finite count of $\mathbf{w}$'s appearance in $\mathbf{s}$, $\mathcal{S}^L$ is the set of length L words observed in $\mathbf{s}$, and $N(\mathcal{S}^L) = k - L + 1$ is the total number. The estimate's accuracy clearly depends on the data stream's length and the nature of the source.

When it simplifies presentation in the following each finite measurement sequence of length L, independent of when it is observed, will be referred to as an L-cylinder. Thus, we conflate a sequence $\mathbf{w}$ with the ergodic theory L-cylinder $s_{\mathbf{w}}$. Henceforth, we will use s^L to denote a length L word $\mathbf{w}$, keeping in mind that this is distinct from the set which $s_{\mathbf{w}}$ indexes within $\mathcal{S}^L$. For stationary processes this is a harmless indiscretion.

Fluctuation Spectra Directly From Sequence Histograms

In what follows we will consider cylinder histograms — plots of $\log_2 \Pr(s^L)$ versus s^L — and variations in histogram bin probabilities. Given a distribution over sequences our goal is to glean useful "macroscopic" properties from the population of "microscopic" sequences by investigating the fluctuation spectrum. In particular we seek a description of the complete measure found in the thermodynamic limit, $L \to \infty$. We define the following "intensive" quantities. Intensive here refers to the fact that the quantities are independent of cylinder length and presumed to be convergent in the thermodynamic limit.

The energy density of a length L sequence s^L is defined as

$$\mathcal{U}_{s^L} = -\frac{\log_2 \Pr(s^L)}{L} \tag{5}$$

For the infinite sequence $\omega \in \mathcal{A}^\infty$ such that $s^L \underset{L\to\infty}{\to} \omega$ the energy density is

$$\mathcal{U}_\omega = -\lim_{L\to\infty} \frac{\log_2 \Pr(s^L)}{L} \tag{6}$$

This definition is often treated as an assumption about the asymptotic scaling of the sequence distribution with sequence length, or state space distribution with coarse-graining size, for a given system and is referred to as the "scaling ansatz".[18] In the approach advocated here we find it useful to simply define the energy density as above. In cases for which its thermodynamic limit exists this energy can be thought of as a scaling exponent. Defining the energy density at the outset as a scaling exponent precludes a meaningful interpretation of the energy of finite sequences. Perhaps more importantly, our definition allows us to isolate the contact with physics to the relationship between energy and information — a subject which is still problematic in our view.

The thermodynamic entropy density $s(\mathcal{U})$ is defined by counting cylinders with a given value of $\mathcal{U}$. Normalizing, we obtain

$$s(\mathcal{U}, L) = \frac{\log_2 N(\mathcal{S}^L_\mathcal{U})}{L} \tag{7}$$

where $\mathcal{S}^L_\mathcal{U} = \left\{s^L : s^L \in \mathcal{A}^L, \mathcal{U}_{s^L} = \mathcal{U}\right\}$ is the level set at energy $\mathcal{U}$ and $N(\mathcal{S}^L_\mathcal{U}) = \left\|\mathcal{S}^L_\mathcal{U}\right\|$ is its size. For infinite sequences with $s^L \underset{L\to\infty}{\to} \omega$ the entropy density is

$$s(\mathcal{U}) = \lim_{L\to\infty} \frac{\log_2 N(\mathcal{S}^L_\mathcal{U})}{L} \tag{8}$$

Thus, the first and crudest approximation to a process's fluctuation spectrum $s(\mathcal{U})$ from finite sets of finite length L cylinders is to just use the definitions of $\mathcal{U}$ and $s(\mathcal{U})$ given above in Eqs. (6) and (8). These quantities are approximated by the equivalent quantities for finite cylinders of length L. They are then plotted, parametrically over the set of probability amplitudes on a graph of $s(\mathcal{U})$ versus $\mathcal{U}$. Example plots will be given later.

It is important to note that sequence histograms, considered as a model class, contain the assumption of block independence in approximating a sequence distribution. In other words, length L sequence histograms represent the joint distribution $\Pr(s_0 s_1 s_2 s_3 \ldots)$ as a product of independent distributions over length L sequences. That is,

$$\Pr(s_0 s_1 s_2 s_3 \ldots) = \Pr(s_0 s_1 \ldots s_{L-1}) \Pr(s_L s_{L+1} \ldots s_{2L-1}) \Pr(s_{2L} s_{2L+1} \ldots s_{3L-1}) \ldots \tag{9}$$

In many cases of physical interest — such as in the context of the thermodynamic limit ($L \to \infty$) — the assumption that infinite sequences can be treated as consisting of independent subsequences is reasonable. But this is less than clear when approximating asymptotic distributions with histograms estimated from finite data streams. And this is particularly acute for processes — like those found at the onset of chaos and at phase transitions — in which correlations exist in sequences of any length.

Large Deviation Theory

With the energy and entropy densities just defined we can begin to go beyond their simple empirical estimates via sequence histograms to probe more deeply into a process's fluctuation properties. To do this, we appeal to a generalization of the Shannon-McMillan theorem.[19,20] The latter indicates that for sufficiently long words w (i) there are two sets of sequences, "typical" — those which one is likely to observe — and "atypical", and (ii) the probability of typical sequences decreases with increasing length at an exponential rate

$$\Pr(\mathrm{w}) \propto 2^{-h_\mu |\mathrm{w}|} \tag{10}$$

where h_μ is a constant independent of $|\mathrm{w}|$, when $|\mathrm{w}|$ is sufficiently large. This constant is called the Shannon entropy rate or the metric entropy, depending on the area of application. It will be defined more directly below. Atypical sequences decay more rapidly and so typical sequences dominate what one observes. To take a simple example, a typical sequence for a biased coin that has a probability of 0.6 of producing a head on a single toss is one with 60% heads; although the sequence consisting of all heads is atypical. We will study this example in more detail below. Despite the simplicity indicated by the Shannon-McMillan theorem, the typical-atypical dichotomy is too coarse for our needs. Here we consider a generalized relation for energy-parametrized subsets of sequences and look at how these subsets' probabilities scale.

The total probability of the class of sequences with a given energy depends not only on the sequences' individual probabilities, but also on their number. This interdependence is captured by the large deviation rate function $\mathcal{I}(\mathcal{U})$.[4,8,21,22] It is defined in terms of the entropy and energy densities as

$$\mathcal{I}(\mathcal{U}) = \mathcal{U} - s(\mathcal{U}) \tag{11}$$

Very roughly, from Eqs. (6) and (8) one sees that the rate function is a measure of the informational mismatch between the size — as measured by $s(\mathcal{U})$ — of an energy level set and the probability of its individual sequences — as measured by $\mathcal{U}$.

The Gartner-Ellis theorem[4,21] expresses the rate function in terms of the likelihood of observing a sequence with energy $\mathcal{U}$

$$\mathcal{I}(\mathcal{U}) = -\lim_{L \to \infty} \frac{\log_2 Pr(\mathcal{U}_{s^L})}{L} \tag{12}$$

where

$$\Pr(\mathcal{U}_{s^L}) = \sum_{s^L \in \mathcal{S}^L_{\mathcal{U}}} \Pr\left(s^L\right) \tag{13}$$

in which the sum is taken over the sequences in the probability level set $\mathcal{S}^L_{\mathcal{U}}$. The interpretation of the rate function is more direct if we write it in the form

$$\Pr(\mathcal{U}_{s^L}) \sim 2^{-L\mathcal{I}(\mathcal{U}_{s^L})} \tag{14}$$

which is true for large enough L. Using Eq. (11) we rewrite this as

$$\begin{aligned} \Pr(\mathcal{U}_{s^L}) &\sim 2^{-L(\mathcal{U}_{s^L} - s(\mathcal{U}_{s^L}))} \\ &= \Pr\left(s^L\right) \mathrm{N}\left(\mathcal{S}^L_{\mathcal{U}_{s^L}}\right) \end{aligned} \tag{15}$$

The first factor is the "intrinsic" probability of a length L sequence s^L with energy $\mathcal{U}_{s^L}$. The second factor is the number of sequences of length L with energy $\mathcal{U}_{s^L}$.

Thus, we see that the large deviation rate function is closely related to the fluctuation spectrum. And either view — information theoretic or thermodynamic — analyzes the range of fluctuations produced by a process via the trade-off between the number of events and their probability.

Renyi Entropy

There is a third and closely related view popular in the dynamics literature. It is based on a generalized entropy introduced by Renyi.[23] It is best understood, for our purposes at least, in terms of several basic properties of Shannon's original entropy H.[19] This is defined for general distributions $\mathbf{P} = \{p_i : i = 0, 1, 2, \ldots\}$ as

$$H(\mathbf{P}) = -\sum_i p_i \log_2 p_i \tag{16}$$

As Shannon notes, H measures the amount of information obtained by independent samples of the distribution $\mathbf{P}$. Independence plays a central role in the utility of information. In particular, Shannon's entropy is additive for independent processes. If $\mathbf{R}$ is a joint distribution which factors into two independent distributions $\mathbf{P}$ and $\mathbf{Q}$, then from Eq. (16) one sees that

$$H(\mathbf{R}) = H(\mathbf{P}) + H(\mathbf{Q}) \tag{17}$$

In the case of a sequence distribution, if the measurements at different times are independent, then there is a simple linear form for the joint entropy

$$H\left(\Pr\left(s^L\right)\right) = L \cdot H(\Pr(s)) \tag{18}$$

and so a simple scaling of the sequence probabilities holds

$$\Pr\left(s^L\right) \underset{L\gg 1}{\propto} 2^{-L\cdot H(\Pr(s))} \tag{19}$$

If one assumes that successive samples of a process are independent when they are not, however, then the apparent Shannon entropy will be higher than the process's rate of producing information.[15] And the process will appear more random than it is. Examples of this phenomenon will be given later on.

Renyi introduced the entropy

$$H_{\mathbf{P}}(\beta) = \frac{1}{\beta - 1} \log_2 \sum_i p_i^\beta \tag{20}$$

as a geometrically averaged information. It was intended as an extension of Shannon's entropy which uses a linear averaging of the self-information, $-\log_2 p_i$.[23] In fact, the Renyi entropy is the most general average which is additive over independent distributions. As is the case for Shannon entropy, underlying the Renyi entropy is the notion of independent events; as is easily verified from Eq. (20). Thus, these entropies, when used as statistics to study a process, entail assumptions of independence, which may or may not be appropriate.

Returning to sequence distributions, a generalized Renyi entropy rate can be defined as

$$h(\beta) = \frac{1}{\beta - 1} \lim_{L\to\infty} \frac{1}{L} \log_2 \sum_{s^L \in \mathcal{A}^L} \Pr^\beta\left(s^L\right) \tag{21}$$

When $\beta = 1$ it becomes the metric entropy or Shannon entropy rate

$$h_\mu = \lim_{L\to\infty} -\frac{1}{L} \sum_{s^L \in \mathcal{A}^L} \Pr\left(s^L\right) \log_2 \Pr\left(s^L\right) \tag{22}$$

This gives the asymptotic growth rate of the total Shannon information

$$H\left(Pr\left(s^L\right)\right) \underset{L\gg 1}{\approx} C + h_\mu L \tag{23}$$

where C is a constant. When $\beta = 0$ the Renyi entropy rate becomes the topological entropy h of the sequences. And this is simply the asymptotic growth rate of the number of distinct sequences as $L \to \infty$

$$h = \lim_{L\to\infty} \frac{\log_2 N(L)}{L} \tag{24}$$

Although the Renyi entropy rate embodies an implicit assumption of block independence, since it is based on histogram estimates of sequence probabilities, it adds the significant additional assumption of incremental scaling. According to Eq. (9) the block independence dictated by histograms leads to a scaling only at multiples of L. Unlike block independence, incremental scaling as in Eq. (23) imposes the additional constraint of self-similarity between histograms at consecutive lengths. One also sees that incremental scaling implies $h_\mu \gg \frac{C}{L}$ there. Thus, scaling holds when the constant of proportionality C is independent of length or, at least, decays rapidly. The onset of chaos and phase transitions are examples of when this isn't the case.

Fluctuation Spectra from Renyi Entropy

The current method for calculating the fluctuation spectrum of a cylinder histogram[18] begins with a "scaling ansatz" — our definition of energy in Eq. (6) — and the definition of Renyi entropy in Eq. (21). These are used to derive, via either Lagrange multipliers or the method of steepest descents, the following relations between Renyi entropy and the thermodynamic entropy, Eq. (8), and energy densities

$$h(\beta) = \frac{\mathcal{U}\beta - s(\mathcal{U})}{\beta - 1} \tag{25}$$

and

$$\mathcal{U}(\beta) = \frac{\partial}{\partial \beta}(\beta - 1)h(\beta) \tag{26}$$

Then the fluctuation spectrum is the entropy density

$$s(\mathcal{U}) = \mathcal{U}\beta - (\beta - 1)h(\beta) \tag{27}$$

as a function of the energy density.

$\mathcal{U}$ and $s(\mathcal{U})$ were defined previously for sequences and sequence classes in Eqs. (6) and (8), respectively. Here, though, $\mathcal{U}$ is considered a parameter and $s(\mathcal{U})$ a function of that parameter. This generalizes the earlier definitions in that they now no longer refer directly to sequences and classes.

And so, at this point, we have two ways of approximating $s(\mathcal{U})$ from finite sets of finite length L cylinders. The first was to use empirical estimates of $\mathcal{U}$ and $s(\mathcal{U})$ in Eqs. (5) and (7) and then plot $s(\mathcal{U})$ versus $\mathcal{U}$ over the range of empirically determined energies.

The second method approximates the fluctuation spectrum defined by Eqs. (25) and (26) using estimated finite cylinder probabilities. To implement this, we need the finite length approximation to Eq. (27)

$$s(\mathcal{U}(\beta, L), L) = \mathcal{U}(\beta, L)\beta - (\beta - 1)h(\beta, L) \tag{28}$$

in which

$$h(\beta, L) = \frac{1}{\beta - 1}\,\frac{1}{L}\log_2 \mathcal{Z}(\beta, L) \tag{29}$$

is a finite L approximation to Eq. (21). In the latter expression we have introduced the partition function

$$\mathcal{Z}(\beta, L) = \sum_{s^L \in \mathcal{A}^L} Pr^{\beta}\left(s^L\right) \tag{30}$$

to simplify notation. With the finite L equivalent of Eq. (26)

$$\mathcal{U}(\beta, L) = \frac{\partial}{\partial \beta}(\beta - 1)h(\beta, L) \tag{31}$$

we have

$$\mathcal{U}(\beta, L) = -\frac{1}{L} \sum_{s^L \in \mathcal{A}^L} \frac{\Pr^{\beta}\left(s^L\right)}{\mathcal{Z}(\beta, L)} \log_2 \Pr\left(s^L\right) \tag{32}$$

Note that by expressing the energy density directly in terms of the cylinder probabilities a new distribution

$$Q_{\beta}\left(s^L\right) = \frac{Pr^{\beta}\left(s^L\right)}{\mathcal{Z}(\beta, L)} \tag{33}$$

has appeared. This is referred to as the "twisted" distribution in large deviation theory.[21] In a sense, at each β the original distribution is shifted to a new one. It describes a stochastic process for which sequences with energy $\mathcal{U}(\beta, L)$ are the most likely — more precisely, they are the typical sequences. Finally, note that the thermodynamic entropy density can also be expressed in terms of the twisted distribution — it is the Shannon entropy rate of the twisted distribution,

$$s(\mathcal{U}) = \lim_{L \to \infty} \frac{1}{L} H\left(Q_{\beta}\left(s^L\right)\right) \tag{34}$$

The Renyi method to estimate the fluctuation spectrum is preferable to the histogram method because varying β allows us to vary the weights of different energy level sets and so interpolate over a continuous range of fluctuation with a smooth, convex function $s(\mathcal{U})$. In pragmatic terms, Eqs. (29) through (32) average over more energy levels and so over more data, since all cylinders contribute at each β. By way of comparison, the direct histogram method uses data only from cylinders isolated to narrow energy ranges. The result is that the Renyi method gives better estimates of entropy as a function of energy.

Nonetheless it must be kept in mind that the Renyi method only gives a finite L approximation to the asymptotic fluctuation spectrum $s(\mathcal{U})$. In the case of the measure sofic systems considered as examples below, however, estimating a model in the form of an ϵ-machine allows for the direct calculation of the asymptotic spectrum. Anticipating somewhat, the main reason for introducing this alternative approach is that the methods just reviewed implicitly use a histogram as their effective model. And that "model class", by definition, neglects correlations between length L sequences in the construction of the fluctuation spectrum. Employing ϵ-machines as models, however, we make explicit use of long and even infinite correlations in the sequences. In one of the final sections we will illustrate the use of all three methods on particular measure sofic systems. But first we need to introduce a number of concepts related to this model class.

Large Deviations and Free Energy

Before closing this section we point out another connection between thermodynamics and large deviations. This then leads to a particularly simple interpretation of the rate function.

First, we define the (Helmholtz) free energy density $\mathcal{F}(\beta)$ via a Legendre transform of the thermodynamic entropy density

$$-\beta\mathcal{F}(\beta) = s(\mathcal{U}) - \mathcal{U}\beta(\mathcal{U}) \tag{35}$$

where

$$\beta(\mathcal{U}) = \frac{\partial s(\mathcal{U})}{\partial \mathcal{U}} \tag{36}$$

when $s(\mathcal{U})$ is differentiable — as is the case for stationary finite-memory processes. Substituting Eqs. (28) and (29) into the finite L version of Eq. (35) and taking the limit, we see that

$$\mathcal{F}(\beta) = \lim_{L\to\infty} \frac{-1}{\beta L} \log_2 \mathcal{Z}(\beta, L) \tag{37}$$

Note that this is close in form to — but not the same as — the Renyi entropy, Eq. (21). The free energy resulting from the Legendre transform is an explicit function of β and an implicit function of energy.

Having introduced the parameter β and the twisted distribution we can reinterpret the large deviation rate function. From Eqs. (11) and (35)

$$\mathcal{I}(\mathcal{U}(\beta)) = (1-\beta)\mathcal{U}(\beta) + \beta\mathcal{F}(\beta) \tag{38}$$

Plugging in the definitions of $\mathcal{F}(\beta)$ from Eq. (37) and $\mathcal{U}(\beta)$ from Eq. (26), we find

$$\begin{aligned} \mathcal{I}(\mathcal{U}(\beta)) &= \lim_{L\to\infty} \frac{1}{L} \sum_{s^L \in \mathcal{A}^L} Q_\beta\Big(s^L\Big) \log_2 \frac{Q_\beta\big(s^L\big)}{\Pr(s^L)} \\ &= \lim_{L\to\infty} \frac{1}{L} D\Big(Q_\beta\Big(s^L\Big) \| \Pr\Big(s^L\Big)\Big) \end{aligned} \tag{39}$$

where $D(Q\|P)$ is the information gain — or Kullback information "distance" — between the distributions P and Q. Thus, the rate function is simply the informational "distance" between the estimated distribution $\Pr\big(s^L\big)$ and the twisted distribution.

This completes our review of two possible approaches to studying fluctuations. We have seen, though somewhat briefly, various relationships between thermodynamics, information theory, and large deviation theory as applied to the study of temporal fluctuations. In the next section we develop an approach to fluctuation spectrum estimation that employs prior knowledge of a model class that is richer than histograms.

ϵ-Machines

Introduction and Definitions

In this section we review finitary ϵ-machines — the stochastic model class that we will use to describe distributions over sequences of measurement symbols. Elsewhere we have shown how an ϵ-machine can be reconstructed from a time series.[10,24] The goal of the reconstruction is to discover "hidden" states and state transition structure from measurements that are at best indirect reflections of some unknown internal dynamics. Here we assume a machine has been obtained and ask what properties it captures of the underlying data source. Equivalently, when using it as a generator of sequences, we ask for the statistics of the output strings in terms of the machine's calculable properties.

A finitary ϵ-machine $M = \{\mathbf{V}, \mathbf{E}, \mathcal{A}, \mathrm{T}\}$ describes the structure of strings over symbols in some measurement alphabet, $s \in \mathcal{A}$. The machine consists of a finite set of states, or vertices $\mathbf{V} = \{v_i : i = 0, 1, \ldots, V-1\}$. A set of labeled, directed edges $\mathbf{E} = \left\{e_{ij} : e_{ij} = \left(v_i \underset{s}{\rightarrow} v_j\right); v_i, v_j \in \mathbf{V}, s \in \mathcal{A}\right\}$ gives the state to state transitions over a single discrete time step. $E = \|\mathbf{E}\|$ is the total number of edges and $V = \|\mathbf{V}\|$ is the total number of vertices. For an ϵ-machine there is a unique state $v_0 \in \mathbf{V}$ specified as the initial or start state. On top of the bare connectivity structure — referred to as the machine's "shape" — $\|\mathbf{V}\| \times \|\mathbf{V}\|$ stochastic transition matrices give the conditional transition probabilities

$$\mathrm{T} = \left\{T^{(s)} : \left(T^{(s)}\right)_{ij;s} = p_{v_i \underset{s}{\rightarrow} v_j} \in [0,1], \quad i,j = 0, \ldots, V-1, v_i \in \mathbf{V}, s \in \mathcal{A}\right\} \tag{40}$$

where $p_{v_i \underset{s}{\rightarrow} v_j} = p(v_j, s|v_i) = p(v_j|v_i, s)p(s|v_i)$ is the probability of the transition to state v_j on symbol s given that the machine is in state v_i. If we strip off the transition probabilities the resulting machine is a deterministic finite automaton.[25] This is consistent with the general definition of an ϵ-machine as a causal model. This means that transitions leaving each automaton state are uniquely labeled by symbols and so the symbol determines the successor state. In the stochastic ϵ-machine this means that, for the v_i to v_j transition on symbol s, $p(v_j|v_i, s) = 1$ or, equivalently, $p_{v_i \underset{s}{\rightarrow} v_j} = p(s|v_i)$. Note that the transitions from each state are normalized

$$\sum_{j=0}^{V-1} \sum_{s \in \mathcal{A}} p(v_j, s|v_i) = 1 \quad , \quad \forall i \tag{41}$$

And so the connection matrix given by

$$T = \sum_{s \in \mathcal{A}} T^{(s)} \tag{42}$$

is a stochastic matrix. If we ignore the edge labels, it describes a Markov chain over the machine states $\mathbf{V}$. In fact, the Markov chain is irreducible: the recurrent states are strongly connected. The machine's "shape", or connectivity structure is given by the $\|\mathbf{V}\| \times \|\mathbf{V}\|$ $0-1$ matrix, denoted T_0. T_0 has a 0 in the elements corresponding to the 0 elements of T and a 1 in the elements corresponding to the nonzero elements of T. That is,

$$(T_0)_{ij} = \begin{cases} 1 & \text{if} \quad T_{ij} \neq 0 \\ 0 & \text{if} \quad T_{ij} = 0 \end{cases} \tag{43}$$

Given the largest eigenvalue λ_{max} of T_0, the topological entropy for the ϵ-machine is given by[26]

$$h = \log_2 \lambda_{max} \tag{44}$$

This quantity gives the asymptotic growth rate in the number of sequences, as a function of increasing length, and is equal to the topological entropy as defined in Eq. (24) applied to the sequences produced by the ϵ-machine.

The stationary probability

$$\vec{p}_{\mathbf{V}} = \left\{ p_{v_i} \in [0,1] : \sum_{i=0}^{V} p_{v_i} = 1, v_i \in \mathbf{V} \right\} \tag{45}$$

of the machine states is given by the left eigenvector associated with the largest eigenvalue of T

$$\vec{p}_{\mathbf{V}} T = \vec{p}_{\mathbf{V}} \tag{46}$$

Recall that the maximal eigenvalue for the stochastic matrix of an irreducible Markov chain is unity. The eigenvector must be normalized in probability. With the state probabilities in hand, the measure-theoretic entropy h_μ for an ϵ-machine is directly computed[19,26] via

$$h_\mu = -\sum_{i=0}^{V-1} p_{v_i} \sum_{j=0}^{V-1} \sum_{s \in \mathcal{A}} p_{v_i \underset{s}{\to} v_j} \log_2 p_{v_i \underset{s}{\to} v_j} \tag{47}$$

It is equal to the quantity defined in Eq. (22) applied to the sequences produced by the ϵ-machine and so gives the asymptotic growth rate of Shannon information in the sequence distribution $\Pr(s^L)$ as a function of increasing length. Unlike Eq. (22), Eq. (47) gives the entropy rate in a finite form. By the Shannon-McMillan theorem h_μ estimates the size of the set of length L typical sequences — that is, those upon which most of the probability distribution is concentrated — via

$$N_{\text{typical}}(L) \approx 2^{h_\mu L} \tag{48}$$

We note that the difference $h - h_\mu$ gives a rough measure of the fluctuations in the asymptotic sequence distribution. More specifically it measures the inhomogeneity in that distribution.[6,27]

The statistical complexity C_μ — the informational size of the machine — is defined as

$$C_\mu = -\sum_{i=0}^{V-1} p_{v_i} \log_2 p_{v_i} \tag{49}$$

It is the average amount of information that an observer gains by inferring the current machine state. In other words, for the task of predicting a process's measurement sequences, the statistical complexity indicates the knowledge an observer has about the source's hidden states.[10,24,28] For the statistical complexity to measure these aspects, the ϵ-machine representing the process must be minimal in the sense of having the smallest number of states necessary to reproduce the process's behavior. This is, in fact, what ϵ-machine reconstruction provides.

Data Sources

ϵ-Machines as Statistical Models

A finitary ϵ-machine is a statistical model for a data source in the sense that it describes a unique distribution over the sequences it generates. Consider a particular sequence $s^L = s_0 s_1 \ldots s_{L-1}$. Its probability is given directly in terms of the machine by first writing the probability of the sequence as the product of the conditional distributions for successive symbols. We then use the conditional independence of the machine states — as with states of a Markov chain. This factors the joint distribution $\Pr(s^L)$ into a product of conditionally-independent transition probabilities in the following manner

$$\begin{aligned}
\Pr(s^L) &= \Pr(s_0 s_1 \ldots s_{L-1}) \\
&= \Pr(s_0)\Pr(s_1|s_0)\Pr(s_2|s_0 s_1)\cdots\Pr(s_{L-1}|s_0\cdots s_{L-2}) \\
&= p(v_\lambda)p(s_0|v_\lambda)p(s_1|v_{s_0})p(s_2|v_{s_0 s_1})\cdots p(s_{L-1}|v_{s_0\cdots s_{L-2}}) \\
&= p(v_0)p(s_0|v_0)p(s_1|v_1)p(s_2|v_2)\cdots p(s_{L-1}|v_{L-1}) \\
&= p(s_0|v_0)p(s_1|v_1)p(s_2|v_2)\cdots p(s_{L-1}|v_{L-1})
\end{aligned} \tag{50}$$

where the notation $v_{s_0 s_1 \cdots s_{k-1}}$ in the third line refers to the state v_k to which the machine is brought upon following the sequence of transitions selected by the string $s_0 s_1 \cdots s_{k-1}$. λ is the null string. It is used above to indicate the initial time before any symbols have been observed or generated. We refer to the state v_0 as the state of total ignorance. The last line follows from the penultimate line because all strings begin in the state of total ignorance v_0 with probability one. The latter should not be confused with v_0's asymptotic probability.

An ϵ-machine is also a model in the sense that it gives an explicit representation of one mechanism by which the observed data could have been produced. Finally, and of equal importance, it generalizes from the observed data to unobserved sequences.*

* In fact, the generalization implicit in machine reconstruction handles the distinction between measure subshifts of finite type and measure Sofic systems with equal facility.

To illustrate the preceding definitions this section's remainder introduces, in Figures 1 through 3 and in Table 1, three labeled, directed graphs associated with the discrete-state stochastic processes we will use to generate data streams for part of our comparative fluctuation analysis. We also introduce another data source, a continuum-state dynamical system — the well-known logistic map — observed with a binary measuring instrument. We will reconstruct ϵ-machines directly from realizations of these data sources. The goal in this is to compare the fluctuation spectra obtained via reconstructed ϵ-machines with those obtained from the histogram-based techniques.

Process	T	$\vec{p}_{\mathbf{V}}$
Biased Coin	(1)	(1)
Golden Mean	$\begin{pmatrix} 0.6 & 0.4 \\ 1.0 & 0 \end{pmatrix}$	$(0.7143, 0.2857)$
Even	$\begin{pmatrix} 0.75 & 0.25 & 0 \\ 0 & 0.4 & 0.6 \\ 0 & 1.0 & 0 \end{pmatrix}$	$(0.0, 0.625, 0.375)$

Table 1 The second column gives the connection matrices T, as defined in Eq. (42), for the three discrete-state processes discussed in this section. Each connection matrix represents a Markov chain with elements corresponding to transition probabilities between the process's states. The third column lists the left eigenvector, defined by Eq. (46) and normalized in probability, associated with the largest eigenvalue — which is unity for stochastic matrices. The elements are the asymptotic state probabilities. The appendix gives a different, "split state" representation for the biased coin.

As before, the measurement alphabet will be $\mathcal{A} = \{0, 1\}$. In Table 1 we give the stochastic connection matrix, Eq. (42), and the asymptotic state distribution, Eq. (46). In Table 2, which appears in a later section, we list the topological and metric entropies as calculated from Eqs. (44) and (47), respectively. That table also gives the topological entropy and, as an approximation of the metric entropy, the Lyapunov exponent for the Logistic map. These will be explained below. Table 4, also in a later section, gives the process's statistical complexity as calculated from Eq. (49). Note that there is no direct way to calculate the statistical complexity from the real-valued logistic map.

Each machine state represents the observer's complete current knowledge of what sequences can be observed in the future and the probabilities with which they can occur.[28] We represent the state of total ignorance, v_0, before any observations have been made, with a double circle in the figures and this is the unique start state for a process. In general, machines will consist

of both transient and recurrent states. By recurrent state we mean a state that can be reached from any other state via some path. By transient state we mean one for which this is not the case. The structure of the transient states governs the decay time of the process to its "steady state". The (strictly) Sofic systems of ergodic theory are those for which there are cycles in the transient states.* In the figures the edges are labeled with the symbol which is observed or emitted — depending on whether the machine is interpreted as either a recognizer or generator — and the branching probability associated with that edge. As a recognizer, one distinguishes sets of sequences which are accepted and which are rejected. Examples of these accepted and rejected sequence sets will be given below.

In Figure 1 we show the single state process representing the tosses of a biased coin that produces sequences with 60% heads. Heads are denoted with 1s and tails 0s.

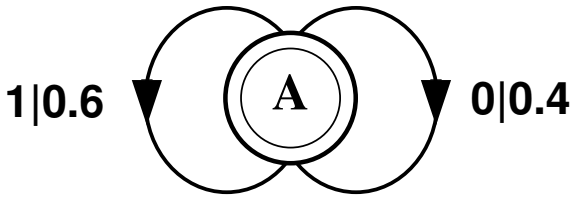

Figure 1 Labeled, directed graph with transition probabilities for modeling tosses of a biased coin. The branching probabilities are $\Pr(s=0)=.4$ and $\Pr(s=1)=.6$. The vertices of the graph represent the "knowledge states" of the process as discussed in the text; in this case there is only one, $\mathbf{V}=\{\mathbf{A}\}$. We represent the start state or "state of total ignorance" with a double circle. State-to-state transitions are represented by the graph edges. These are labeled $s|p$, where $s \in \mathcal{A}$ is a symbol in the measurement alphabet and $p = p_{v_i \to v_j}^{s}$ is a transition probability.

Figure 2 shows the graph for the golden mean process. Its name derives from the fact that its topological entropy is the logarithm of the golden mean, $\phi = \frac{1}{2}\left(1+\sqrt{5}\right)$. (See Table 2). If, as discussed above, the machine is being used to test whether a given sequence was produced by the process, one begins in the start state and travels on the sequence of states, or path, selected by the symbols in the sequence. If a symbol occurs in the sequence for which there is no transition in the graph that sequence is rejected; otherwise, it is accepted. For example, in the case of the golden mean system the sequence 1010 begins in the start state $\mathbf{A}$, and traces the path $\mathbf{AABAB}$ through the graph ending in state $\mathbf{B}$. It is accepted by the machine. The sequence 1001 begins in state $\mathbf{A}$ and follows the path $\mathbf{AAB}$ on the subsequence 10. Since there is no edge labeled with $s=0$ leaving state $\mathbf{B}$, when the second 0 in the sequence is encountered there is no transition and the sequence is rejected.

The set of sequences that are rejected by a finite machine often can be compactly described in terms of the list of smallest rejected subsequences — the set $\mathbf{F}$ of irreducible forbidden words. For the golden mean process, there is a particularly simple scaling structure of rejected sequences

* More precisely, the transient cycles appear explicitly in a process's semigroup graph.[29] They need not appear in the minimal ϵ-machine representation. The ϵ-machine approximation of the logistic process, the Misiurewicz machine presented below, is one example of the latter situation. It is a measure (strictly) Sofic system without explicit ϵ-machine transient states.

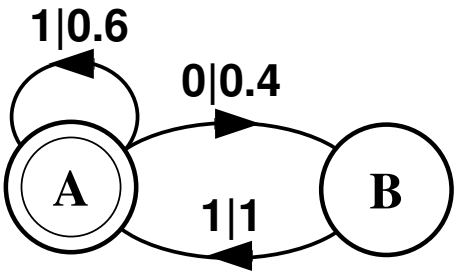

Figure 2 The golden mean process — so-called since the growth rate of the number of sequences as a function of length is the logarithm of the golden mean. In fact, the total number of sequences at length L is given by the Fibonacci number F_{L+2}. In simplest terms, the golden mean process generates all binary sequences except those containing two consecutive 0s. We have chosen a particular statistical bias so that, for example, $\Pr(s = 1|v = \mathbf{A}) = 0.6$. See Figure 1 for explanation of the representation.

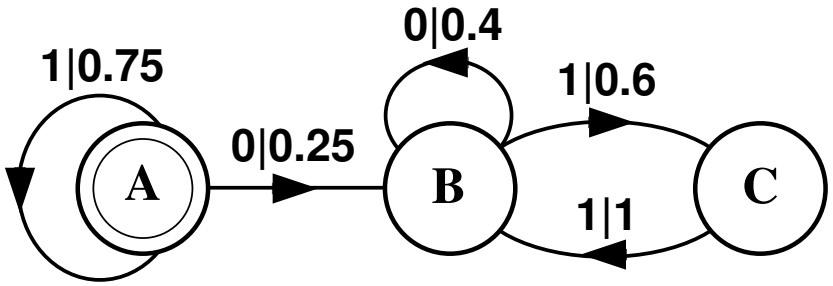

Figure 3 The even process generates all binary sequences in which 1's occur in even length blocks bounded by 0's. The statistical bias is set so that $\Pr(s = 1|v = \mathbf{A}) = 0.6$. See Figure 1 for explanation of the representation.

with increasing length. Namely, the list of irreducible forbidden words consists of the single length 2 sequence 00. All rejected sequences of greater length are simply those containing 00 as a subsequence. Thus, the golden mean process produces all binary sequences except those with consecutive 0s. In the next example we will see that $\mathbf{F}$ can be infinite, even though the machine is finite.

Figure 3 shows the graph for the so-called even process. The set of sequences produced by this process is referred to as the even system.[30] It consists of the set of all sequences containing only even strings of ones bounded by zeros. Note that, if edge labels are ignored, recurrent parts of the golden mean and even processes are identical. We will discuss this similarity further in a later section.

For the even system the shortest forbidden sequence is 010; it is also an irreducible forbidden word. Though there are length 4 forbidden sequences, the next longest irreducible forbidden word is 01110. Note that it does not contain any shorter forbidden sequences and so is irreducible. This turns out to be the case for the infinite set of sequences containing odd numbers of 1's sandwiched between 0's. That is, for the even system $\mathbf{F} = \left\{01^{2n+1}0 : n = 0, 1, 2, \ldots\right\}$.

A Continuum-State Source

The final data source that we consider here is derived from a trajectory of a continuum-state dynamical system, the logistic map, observed with a very coarse measuring instrument. The trajectory is generated by iterating the logistic map

$$x_{n+1} = f(x_n) \tag{53}$$

with $f(x) = rx(1-x)$. The control parameter is set to the Misiurewicz parameter value $r = r_M \approx 3.9277370017867516\ldots$ at which $f^4(x_c) = f^5(x_c)$ and $x_c = \frac{1}{2}$ — the location of the map's maximum. The trajectory $\mathbf{x} = x_0 x_1 x_2 x_3 \ldots$ is converted to a binary sequence by observing via the generating binary partition

$$\mathcal{P} = \{s = 0 \sim x_n \in [0, x_c), s = 1 \sim x_n \in [x_c, 1]\} \tag{54}$$

The generating property means that sufficiently long binary sequences identify arbitrarily small segments of initial conditions. Due to this, the information processing in the logistic map can be studied using the "coarse" measuring instrument $\mathcal{P}$. As independent checks on several statistics that will be used in our comparison, Table 2 gives the topological entropy for the logistic map. It was estimated using the kneading determinant[31,32] with 100 terms and estimating its smallest zero to 1 part in 10^6. As an estimate of the metric entropy h_μ the table also gives the Lyapunov exponent λ for the logistic map averaged over 10^8 iterates. For the logistic map λ is a good estimator of h_μ.

ϵ-Machine Thermodynamics*

To study the fluctuations in sequences, we focus on the variation in their probabilities. To this end, we introduce parametrized symbol transition matrices $\left\{T_\beta^{(s)} : s \in \mathcal{A}\right\}$ where

$$\left(T_\beta^{(s)}\right)_{ij} = e^{\beta \ln p_{v_i \to v_j}^{\ s}} \tag{55}$$

As in the preceding section on Renyi entropy, we think of each setting of the formal parameter β as emphasizing a different set of sequences. First, β directly modifies the transition weights. This determines the effective weights of paths taken through the ϵ-machine. And this, in turn, reweights the energy level sets over the sequences. The result is a new typical set. In this way each fixed setting of β in Eq. (55) is associated with equiprobability level subsets within the set of all sequences. The exponential form of the variation implies, in the sense of maximum entropy, that the entropy density and average energy density provide the only constraints on the transition weights.[19,26,33]

* This section corrects and extends the very brief description of ϵ-machine thermodynamics given in [10,24].

The associated connection matrix is defined, as before, by summing over the symbol alphabet

$$T_\beta = \sum_{s \in \mathcal{A}} T_\beta^{(s)} \tag{56}$$

The maximum eigenvalue λ_β and associated left $\vec{l}_\beta$ and right $\vec{r}_\beta$ eigenvectors are determined by the linear equations

$$\begin{aligned} \vec{l}_\beta T_\beta &= \lambda_\beta \vec{l}_\beta \\ T_\beta \vec{r}_\beta &= \lambda_\beta \vec{r}_\beta \end{aligned} \tag{57}$$

The eigenvectors are chosen so that the dot product $\vec{l}_\beta \cdot \vec{r}_\beta$ is unity. This normalizes their components in probability, yielding an effective state probability

$$\vec{\mathrm{p}}_\beta = \left\{ \left(\vec{\mathrm{p}}_\beta\right)_i = \left(\vec{l}_\beta\right)_i \left(\vec{r}_\beta\right)_i : i = 0, 1, ..., V-1 \right\} \tag{58}$$

with

$$\sum_{i=0}^{V-1} \left(\vec{\mathrm{p}}_\beta\right)_i = 1 \tag{59}$$

The Renyi entropy density, cf. Eq. (21), is given in terms of the machine by

$$h(\beta) = \frac{\log_2 \lambda_\beta}{1-\beta} \tag{60}$$

This should not be confused with the entropy rate of the sequences generated by the machine with transition weights biased by a given β. That rate will be given shortly. Note that for $\beta = 0$ in Eq. (60) we obtain the topological entropy, cf. Eq. (24),

$$h = h(0) = \log_2 \lambda_0 \tag{61}$$

The free energy density is

$$F(\beta) = -\beta^{-1} \log_2 \lambda_\beta \tag{62}$$

The similarity to the Renyi entropy $h(\beta)$ is again apparent; but it is not the same, as noted in a previous section.

In general T_β is not a stochastic matrix — the rows do not sum to unity. The equivalent stochastic process with transition probabilities weighted according to T_β is given by the stochasticized version[19,26] S_β of T_β

$$\left(\mathrm{S}_\beta\right)_{ij} = \frac{\left(T_\beta\right)_{ij} \left(\vec{r}_\beta\right)_j}{\lambda_\beta \left(\vec{r}_\beta\right)_i} \tag{63}$$

Note that $\sum_j \left(\mathrm{S}_\beta\right)_{ij} = 1$. The left eigenvector associated with the largest eigenvalue, $\lambda_{\mathrm{S}_\beta} = 1$, is given by

$$\left(\vec{\mathrm{p}}_\beta\right)_i = \left(\vec{l}_\beta\right)_i \left(\vec{r}_\beta\right)_i \tag{64}$$

This is identical to the vector defined above in Eq. (58).

Elsewhere[28,34] we show that the ϵ-machine thermodynamic entropy density is given by

$$s(\mathcal{U}(\beta)) = -\sum_{i,j} \left(\vec{\mathrm{p}}_\beta\right)_i \left(\mathrm{S}_\beta\right)_{ij} \log_2 \left(\mathrm{S}_\beta\right)_{ij} \tag{65}$$

and the machine energy density by

$$\mathcal{U}(\beta) = \beta^{-1}\left(h_\mu\left(\mathrm{S}_\beta\right) - \log_2 \lambda_\beta\right) \tag{66}$$

From Eq. (65), and recalling Eq. (47), we see that the thermodynamic entropy density $s(\mathcal{U}(\beta))$ for the machine at parameter value β is just the metric entropy h_μ of the stochasticized machine S_β

$$s(\mathcal{U}(\beta)) = h_\mu\left(\mathrm{S}_\beta\right) \tag{67}$$

In analogy to Eq. (34) it is the Shannon entropy rate of the infinite sequence twisted distribution

$$Q_\beta(\omega) = \lim_{L\to\infty} Q_\beta\left(s^L\right) \tag{68}$$

with $s^L \underset{L\to\infty}{\to} \omega$. It is important to emphasize that the fluctuation spectrum is determined completely by the ϵ-machine's stochastic connection matrix — in fact, just its recurrent component — and not by the edge symbol labeling, which is responsible for the observed sequences. Consequently, it reflects properties of the Markovian structure of the internal states, and not directly properties of the sequences. The equivalence of the sequence fluctuation spectrum — Eq. (27) — and that just given for the internal states follows from the determinism of ϵ-machines.

Above, β was simply a parameter that was varied to emphasize different energy level subsets of cylinders. It also plays the role of inverse temperature $\beta = \mathcal{T}^{-1}$ as in statistical mechanics — cf. Eqs. (27) and (66). The thermodynamic interpretation of varying β, then, is that the stochastic process is put into contact with an infinite reservoir at "temperature" β^{-1}. The contact shifts the mean energy to $\mathcal{U}(\beta)$. This emphasizes the associated paths with energy $\mathcal{U}(\beta)$ in the ϵ-machine. And those paths correspond to sequences in the Shannon typical set for the twisted distribution $Q_\beta(\omega)$.

The energy extremes

$$\mathcal{U}_{\min} = \lim_{\beta\to\infty} \mathcal{U}(\beta) \tag{69}$$

and

$$\mathcal{U}_{\max} = \lim_{\beta \to -\infty} \mathcal{U}(\beta) \tag{70}$$

give the lower and upper bounds on the range of fluctuations. The ground state, $\mathcal{U}_{\min}$, is associated with the most probable sequences; the antiground state, $\mathcal{U}_{\max}$, with the least probable sequences. All the states at negative temperature, i.e. negative β, such as that at $\mathcal{U}_{\max}$, can be thought of as population-inverted states of high energy. They are analogous to population-inverted states in condensed matter systems with bounded energy. The degree of degeneracy of these states is measured by $s(\mathcal{U}_{\min})$ and $s(\mathcal{U}_{\max})$, respectively. They can be estimated using either the histogram, Renyi, or ϵ-machine fluctuation spectrum estimation methods. For the histogram method, although there is no explicit β, one simply takes the highest and lowest energies.

The main point of this section has been to show that the functions $\mathcal{U}(\beta)$ and $s(\mathcal{U}(\beta))$ are given directly in terms of the principle eigenvalue and eigenvectors of S_β and T_β. In this way the asymptotic fluctuation spectrum can be calculated directly from an ϵ-machine using Eqs. (65) and (66). And this is our third and last technique for studying fluctuations.

Thermodynamic Complexities

An object's complexity is typically associated with the size of its description in some chosen representation. The Kolmogorov-Chaitin complexity of a binary sequence, for example, is the length in bits of the shortest program that reproduces the sequence when run on a deterministic universal Turing machine.[35,36] As intended by Kolmogorov in the search for an algorithmic basis of probability theory,[35] this complexity is closely related to Shannon's entropy rate of a process which produces the sequence in question.[20] And as such it is a measure of the amount of ideal randomness in the process. By way of contrast, we close this review of thermodynamic properties by commenting on several notions of complexity appropriate for measuring the amount of structure beyond ideal randomness.

Free Energy

First we note the analogy of the above relations in Eqs. (66) and (62) with the standard equilibrium thermodynamic relation for the free energy F, internal energy U, and entropy S

$$\mathrm{F} = \mathrm{U} - \mathcal{T}\,\mathrm{S} \tag{71}$$

where $\mathcal{T}$ is the temperature. This suggests that the free energy density is an informational measure of the mismatch between the constraints that define the process — and so its equilibrium distribution — and the actual probability distribution at the given $\mathcal{T}$. At $\beta = 1$, the free energy vanishes. Thus, the free energy is a measure of the amount of probabilistic structure — i.e. variation in the distribution — beyond that in the equilibrium distribution.

Statistical Complexity Spectrum

We now introduce a somewhat more novel quantity — and one not, apparently, part of thermodynamics proper — the statistical complexity spectrum. Given that we have a minimal machine, as we have been assuming and as machine reconstruction provides, then the statistical complexity C_μ — Eq. (49) — measures the average amount of memory in the process. Recall that this is complementary to the rate at which information is produced — the metric entropy h_μ. Analogous to the fluctuation spectrum $s(\mathcal{U}(\beta))$ we have the parametrized statistical complexity

$$\mathcal{C}_\mu(\beta) = -\sum_{i=0}^{V-1} \left(\vec{\mathrm{p}}_\beta\right)_i \log_2 \left(\vec{\mathrm{p}}_\beta\right)_i \tag{72}$$

in which $\left(\vec{\mathrm{p}}_\beta\right)_i$ is given by Eq. (64). The statistical complexity spectrum $C_\mu(\mathcal{U})$ is then obtained by simply changing from inverse temperature to energy density coordinates

$$C_\mu(\mathcal{U}) = \mathcal{C}_\mu(\beta(\mathcal{U})) \tag{73}$$

The complexity spectrum gives the apparent amount of information required to produce sequences at a given level — i.e. energy density — of fluctuation. Unlike $\mathcal{U}$ and $s(\mathcal{U})$, C_μ and $C_\mu(\mathcal{U})$ cannot be directly estimated from a data stream. They require a minimal machine and, ultimately, are based on a notion of a process's causal structure. By way of comparison, the histogram-based quantities derive from a notion of predictability, which is only indirectly related to a process's structure.

Fluctuation Complexity

How can we characterize the total range of fluctuations that a process generates? Here we introduce the fluctuation complexity Ξ — an attempt to capture this global property.*

The fluctuation spectrum $s(\mathcal{U})$ gives the upper bound on the population density of fluctuations at energy density $\mathcal{U}$. Often, however, there is more structure in a process than is revealed by $s(\mathcal{U})$. For example, the combinatorial structure of a machine's paths leads to subsets with a range of sizes, even at a given energy. Thus, sampling sequences from a process leads to a complicated set of points in the entropy-energy plane $(\mathcal{U}, s)$. A detailed account of this requires a discussion of how to enumerate the ways a particular set of edges in the graph representing the process can occur.[34] Although such an account is beyond our current scope, we can introduce several simple quantities that capture the gross structure of the $(\mathcal{U}, s)$ set.

For the most general fluctuation complexity Ξ we view the entropy-energy plane as the support of a distribution $\Pr(\mathcal{U}, s)$.† We define it to be

$$\Xi = -\int\int d\mathcal{U} ds \Pr(\mathcal{U}, s) \log_2 \Pr(\mathcal{U}, s) \tag{74}$$

* Note that this is not the quantity of the same name used in [37].

† This distribution is referred to Lebesgue measure on the entropy-energy plane — and not to the asymptotic invariant measure on sequences.

It gives the amount of information in this distribution. Alternatively, it measures the average amount of information obtained in observing a particular subpopulation at a given energy.

There are several crude approximations appropriate to the simplified discussion here. The first is the "box" approximation to the $s(\mathcal{U})$ curve:

$$\Xi_0 = h(\mathcal{U}_{\max} - \mathcal{U}_{\min}) \tag{75}$$

As a statistic, this is easy to check given the fluctuation spectrum. The next approximation is the integral

$$\Xi_1 = \int_{\mathcal{U}_{\min}}^{\mathcal{U}_{\max}} d\mathcal{U} s(\mathcal{U}) \tag{76}$$

which estimates the total area under $s(\mathcal{U})$. Since $s(\mathcal{U})$ is often only implicitly defined we can also use the Legendre transformed approximation

$$\Xi_1 = \int_{+\infty}^{-\infty} d\beta \frac{\partial \mathcal{U}}{\partial \beta} s(\mathcal{U}(\beta)) \tag{77}$$

It is clear that $\Xi_0 \geq \Xi_1 \geq \Xi$. These two complexities miss important features of the complete fluctuation spectrum. They tend to overestimate the fluctuation complexity due to restrictions on the allowed fluctuations. The restrictions can appear (i) as gaps in the area under $s(\mathcal{U})$ and (ii) in the variation of probability over the interior of $s(\mathcal{U})$. Nonetheless, they do give some measure of the range of allowed fluctuations. In using them, we ignore in a sense fluctuations of the fluctuations. More precisely, we are assuming that (i) $\Pr(\mathcal{U}, s)$ is uniform, (ii) there are subpopulations at each allowed energy with less than exponential size, and (iii) $\Pr(\mathcal{U}, s)$'s support is simply connected. In this case, all fluctuations $(\mathcal{U}', s')$ with $\mathcal{U}_{\min} \leq \mathcal{U}' \leq \mathcal{U}_{\max}$ and $0 \leq s' \leq S(\mathcal{U}')$ are allowed.

The simplest example is a fair coin. It has no fluctuations and so $\Xi_0 = \Xi_1 = \Xi = 0$. The biased coin, though, has positive fluctuation complexity. We will give more interesting examples in the following. In particular, we discuss a process with zero fluctuation complexity, but finite memory, $C_\mu > 0$, and one with finite Ξ and $C_\mu = 0$. Thus, Ξ and C_μ measure different properties.

Estimated Fluctuation Spectra

The preceding sections established the basic theory and methods for the study of fluctuation spectra, and introduced the prototype processes. This section now compares the three techniques

— based on histograms, the Renyi entropy, and ϵ-machines — for the four prototype processes. The goal, of course, is to see how well the different model classes capture the processes' internal structure and, ultimately, how well the spectra are estimated. As a reference, the appendix gives the derivation of the thermodynamic properties for the first three processes: the biased coin, golden mean, and even processes. The comparison proceeds in three steps. The first constructs sequence histograms from the various data streams; the second reconstructs ϵ-machines from the same data, and the last juxtaposes the estimated spectra and draws conclusions.

Histograms as Models

Each process was used to generate a binary data stream of length $k = 10^7$ for input to the model estimation step. For the discrete-state processes shown in Figures 1, 2, and 3, each data stream was generated via a random walk through the labeled, directed graph that was biased according to the transition probabilities. For the logistic process we generated a binary data stream of length $k = 10^7$ by iterating the map at the Misiurewicz parameter value $r = r_M$ and observing the resulting trajectory through a binary partition. By most experimental standards these are long data streams — though [38] analyzes one exception. This length was chosen solely for the benefit of the histogram and Renyi fluctuation spectrum estimation methods. The following sections will demonstrate clearly how they do even with such generous amounts of data.

Figures 4, 5, 6, and 7 show histogram mosaics estimated from the four selected processes's data streams. The mosaics consist of semi-log plots of probability density $P\left(s^L\right)$ versus s^L over a range of cylinder lengths $L \in [1, 9]$. Each histogram was obtained from the data stream by determining the frequencies f_{s^L} of the subsequences at the given length and then forming the probability density $P\left(s^L\right) = 2^L \cdot f_{s^L}$, which is normalized so that $\int_o^1 dx P(x) = 1$. The horizontal axis presents the binary sequences as binary fractions in the interval. The consequence of this is that the bin widths decrease exponentially with increasing cylinder length.

The biased coin generates all binary sequences and this is reflected in Figure 4: all bins have positive probability. A simple scaling structure is evident across the different L histograms. Scaling structure, in this case, means that the pattern of bin heights over contiguous bins in a given histogram exactly resembles the bin height pattern in shorter L histograms under suitable renormalization of the axes. This indicates that there are no correlations among sequences of any length L introduced by restrictions on subsequences. Considered as a model — a "look up table" for the sequences — the histogram requires that an exponentially large number — $\approx 2^{hL}$ — of parameters be estimated from the data. The evident scaling indicates that far fewer parameters are actually necessary.

The scaling pattern for the golden mean process is more complicated. (See Figure 5.) First off, forbidden sequences are apparent as empty bins. Also, the variation in bin heights seems

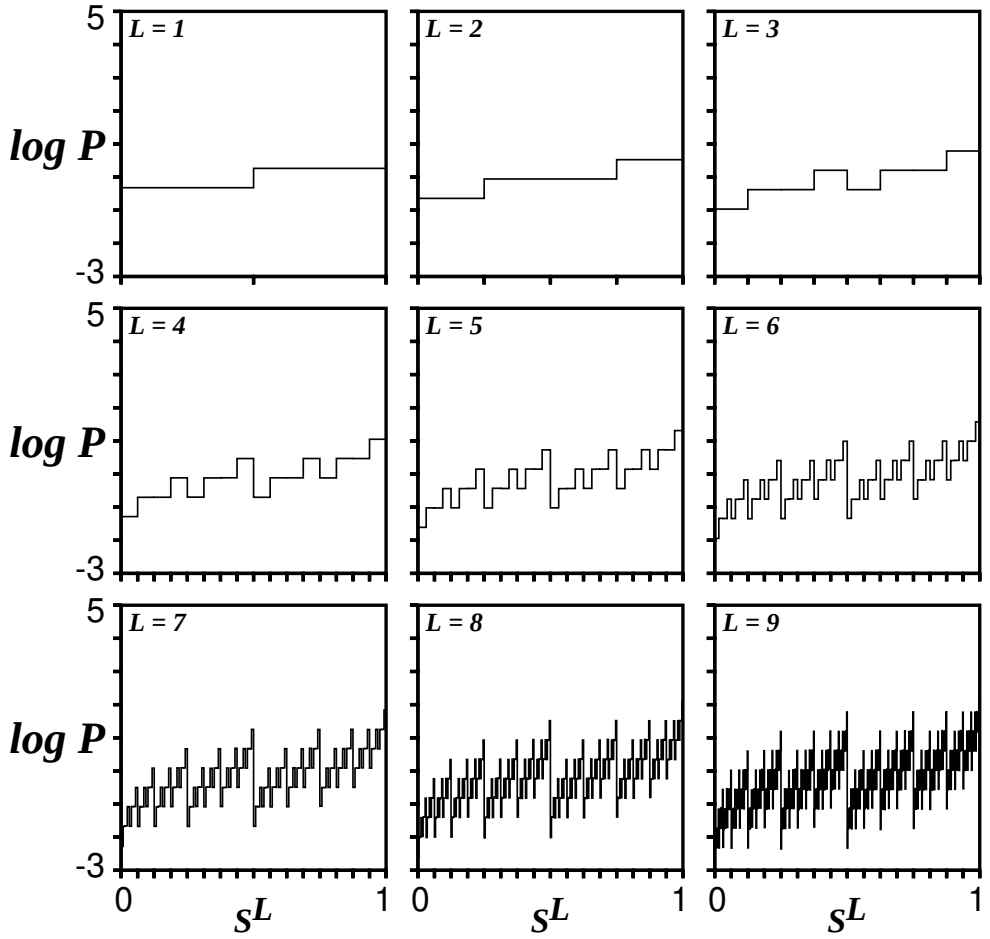

Figure 4 Biased coin process: Mosaic of sequence histograms for sequences of lengths $L \in [1, 9]$. (After [27].) Each histogram plots $\log_2 P(s^L)$ versus s^L, where $P(s^L)$ is the probability density and s^L is evaluated as a binary fraction. Each histogram was obtained from a data stream consisting of a binary sequence of length $k = 10^7$ generated by a random walk through the stochastic machine shown in Figure 1. The random walk is biased according to the transition probabilities. The self-similar structure of the distribution is easily discernible. And this suggests that the fluctuation spectrum will be easy to model for the biased coin process.

unstructured. Despite this, some scaling structure is nonetheless discernible. As discussed in a previous section, the first restricted sequence is the word 00 and this is seen as a "hole" in the histogram support at $L = 2$. The scaling structure of the histogram support with increasing length is simply the result of excluding sequences containing the irreducible forbidden word 00. The support as $L \to \infty$ is a single Cantor set with dimension equal to the golden mean's topological entropy, $h = \log_2 \phi$. As for the distribution, there is smaller range of bin heights within which the histogram values fluctuate as compared to the biased coin process. This indicates that the golden mean process has a more statistically homogenous sequence distribution. This is confirmed by

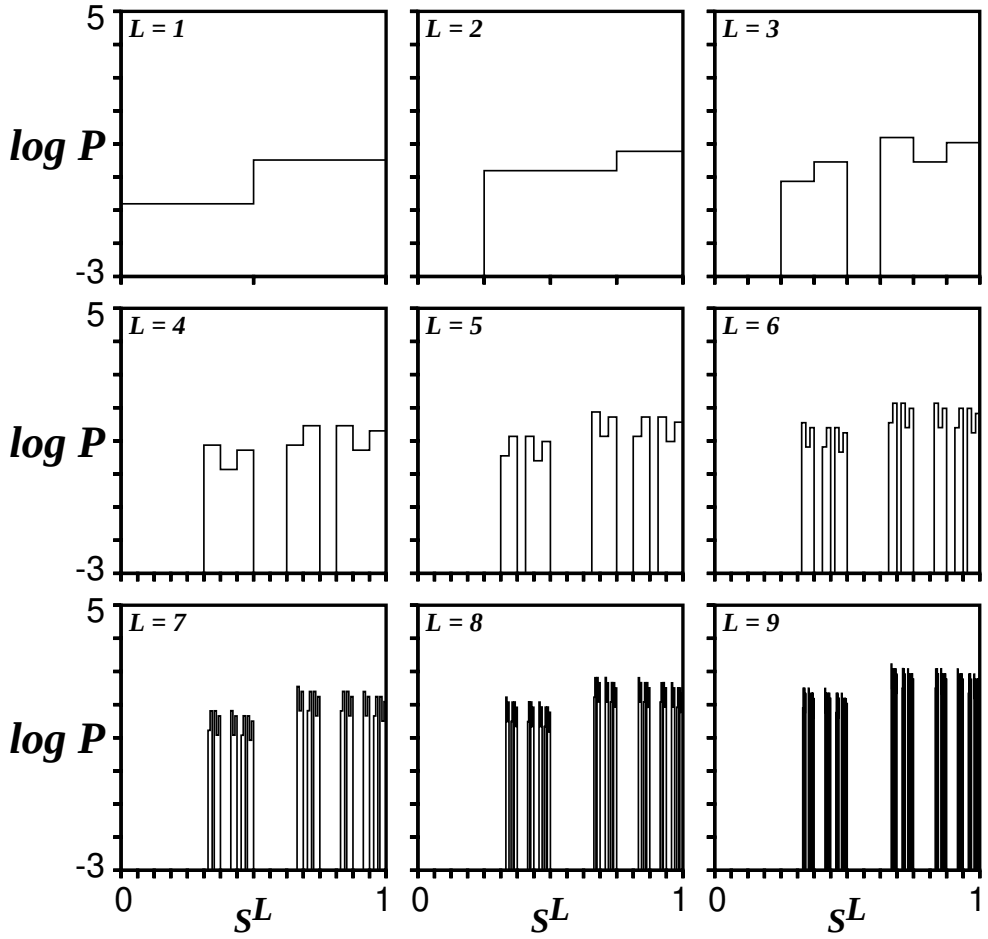

Figure 5 Golden mean process: Sequence histogram mosaic as in Figure 4 but obtained from a data stream consisting of a binary sequence of length $k = 10^7$ generated by a random walk through the machine of Figure 2. Compared to the biased coin process, the scaling behavior is visually more complicated; though some regularities in the bin heights and in the distribution's support are discernible across different sequence lengths. Here there are excluded sequences seen as "holes" in the distribution's support. These occur in bins associated with the set of sequences containing subsequences $\mathbf{w} \in \{00\}$.

comparing the "inhomogeneity parameter" for the biased coin, $h - h_\mu = 0.029$, and for the golden mean system, $h - h_\mu = 0.0007$. The difference is more than an order of magnitude.

The histogram mosaic for the even process is shown in Figure 6. Although its scaling structure is harder to discern than for the golden mean process and certainly for the biased coin, it can still be observed in this series of histograms. The pattern of "holes" is harder to identify for the even process. This is due to the fact that the holes are associated with the set of sequences containing subsequences $\mathbf{w} \in \left\{01^{2n+1}0 : n = 0, 1, 2, \ldots\right\}$ — a countably infinite number of irreducible forbidden words. Recall that there was only one such word for the golden

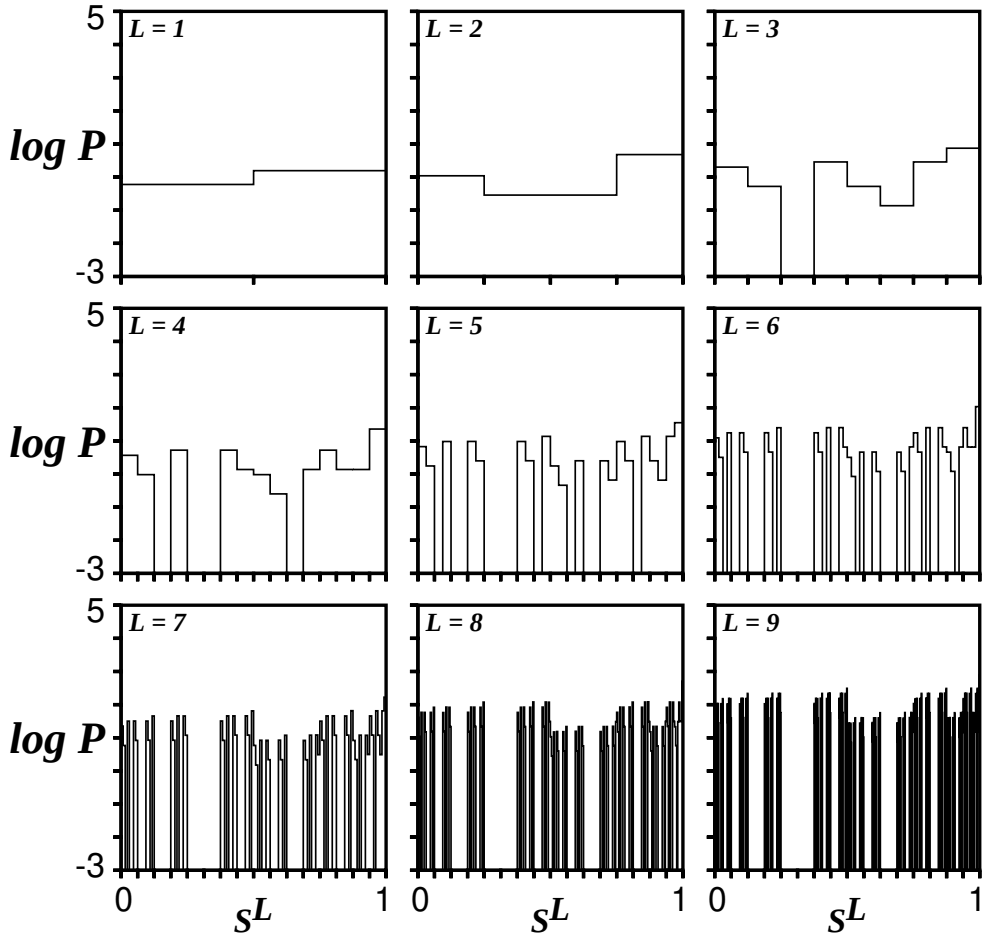

Figure 6 The even process: Sequence histogram mosaic as in Figure 4 but obtained from a data stream consisting of a binary sequence of length $k = 10^7$ generated by a random walk through the labeled, directed graph shown in Figure 3. The even process is more complicated still. The sequence distribution's support consists of a countable infinity of Cantor sets, for example.

mean process, which lead to the creation of a single Cantor set as $L \to \infty$. The support of the even process, by contrast, has an infinite number of Cantor sets — one for each $\mathbf{w} \in \mathbf{F}$. As expected from the value of the inhomogeneity parameter for the even system, $h - h_\mu = 0.087$ the range of bin heights within which the histogram values fluctuate is larger than for the golden mean process.

For the Logistic map process (Figure 7), however, the scaling structure — at least up to $L = 9$ — is very hard to discern, suggesting that correlations are still important in sequences of this length. Additionally, the variation in bin heights appears rather unstructured. Since there seems to be little regularity in the bin heights, estimation of the bulk of the histogram model "parameters" — the bin heights — appear to be necessary to properly model the process.

Compression of the histogram model appears unlikely. From this it might be expected that this process's fluctuation spectrum would be the hardest of the examples to estimate. The range within which the histogram bin heights vary is slightly smaller than for the biased coin — which is corroborated by the logistic map's inhomogeneity parameter $h - \lambda = 0.0345$.

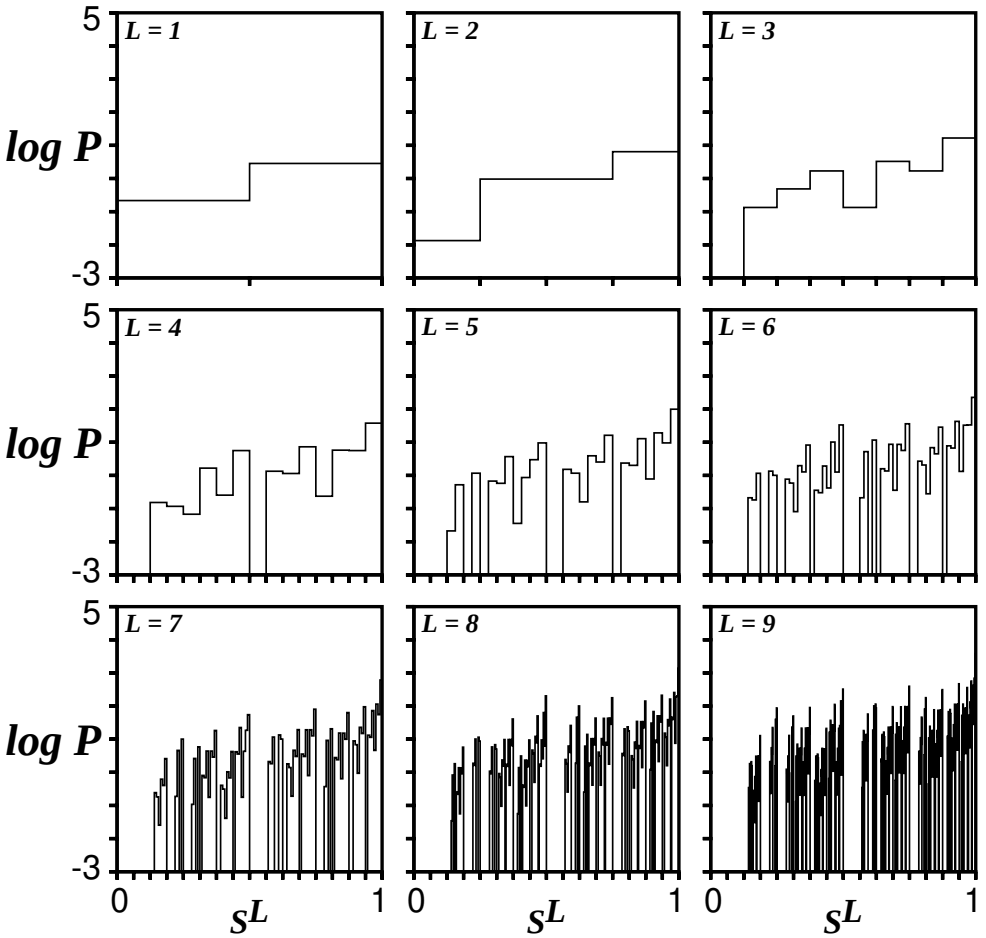

Figure 7 The Logistic map at the Misiurewicz parameter: Sequence histogram mosaic as in Figure 4 but obtained from a data stream consisting of a binary sequence of length $k = 10^7$ generated by observing iterates of the logistic map — Eq. (53) — with a binary measuring instrument. There seems to be little apparent scaling structure in the mosaic, either in the bin heights or in the "holes" in the support.

Empirical ϵ-Machines

ϵ-machines were reconstructed from the data streams according to the methods described

in [10,24].* The estimated machines for the biased coin, golden mean, and even systems are shown in Figures 8, 9, and 10 with transition probabilities quoted to one part in 10^6. The Misiurewicz machine shown in Figure 11 was reconstructed from the logistic map data stream. Its connection matrix is

$$T = \begin{pmatrix} 0.636 & 0.364 & 0 & 0 \\ 0.724 & 0 & 0.276 & 0 \\ 0 & 0 & 0 & 1.0 \\ 0 & 0.521 & 0.479 & 0 \end{pmatrix} \tag{78}$$

and its state probability vector is

$$\vec{p}_{\mathrm{V}} = (0.4913, 0.2470, 0.1309, 0.1309) \tag{79}$$

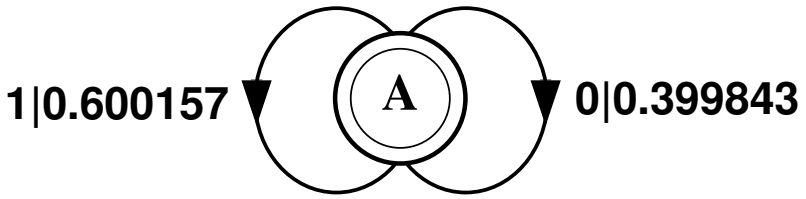

Figure 8 Reconstructed ϵ-machine for the biased coin process obtained from a binary sequence of length $k = 10^7$ generated by a random walk through the machine of Figure 1.

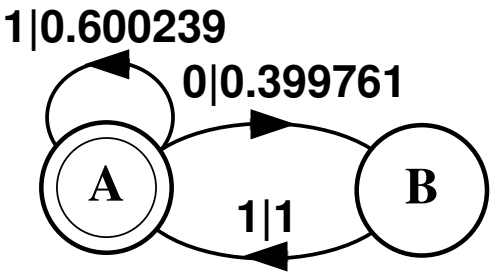

Figure 9 Reconstructed ϵ-machine for the golden mean process obtained from a binary sequence of length $k = 10^7$ generated by a random walk through the machine of Figure 2.

In closing this section we give estimates for the various thermodynamic quantities that follow from the machines and their fluctuation spectra, including the fluctuation complexities. These are in Tables 2, 3, and 4.

Comparing the values in Table 2, the entropies for the processes and for the reconstructed machines show extremely good agreement. Indeed, the shape of all of the reconstructed machines

* The tree depth $D = 16$, the morph depth $L = 8$, and no probabilistically distinguished states were reconstructed — i.e. $\delta = 1$. There was only a small variation in the estimates for a range of L about this choice.

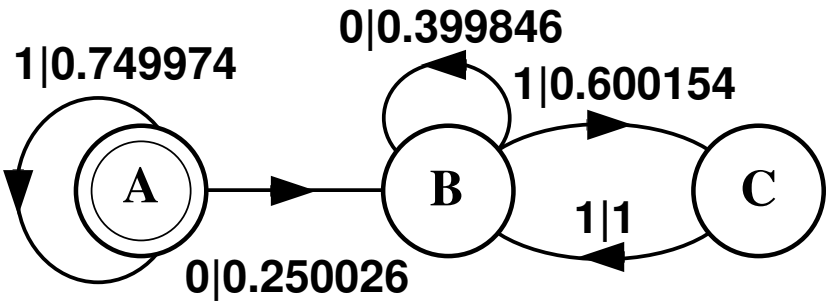

Figure 10 Reconstructed ϵ-machine for the even process obtained from a binary sequence of length $k = 10^7$ generated by a random walk through the machine of Figure 3.

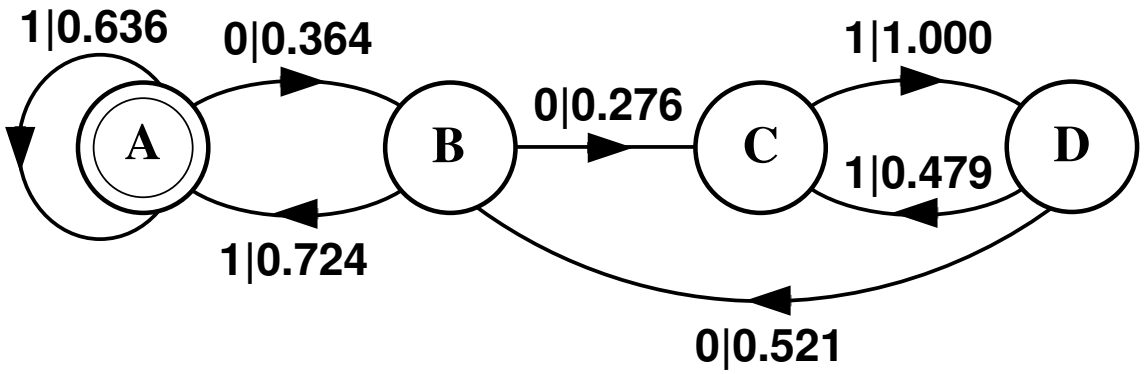

Figure 11 The Misiurewicz machine: the ϵ-machine reconstructed from a binary sequence of $k = 10^7$ iterates of the logistic map $f(x) = rx(1-x)$ at the Misiurewicz parameter value $r = r_M \approx 3.9277370017867516\ldots$. The symbols 0 and 1 of the measurement alphabet correspond to the left and right halves, respectively, of a binary partition of the unit interval $x \in [0, 1]$ with partition divider at $x = 0.5$.

was identical to that of the data generating processes shown in Figures 1, 2, and 3. Additionally, the Misiurewicz machine had the correct shape according to the kneading matrix for the logistic map at $r = r_M$.[16,32,39] Hence the topological entropy for the reconstructed machines is equal to that for the data generating processes. This is in contrast to the Renyi spectrum method which, as we will see, significantly overestimates the topological entropy in all but the simplest cases. Finally, note that the metric entropy of the Misiurewicz machine is a good approximation of the logistic map Lyapunov exponent λ quoted in Table 2. This is expected from (i) the theorem[40] stating that when the invariant measure is absolutely continuous with respect to Lebesgue measure, then $h_\mu = \lambda$, and (ii) at $r = r_M$ the logistic map's invariant measure is absolutely continuous. Notice, though, that the relative error between the Misiurewicz machine h_μ and the logistic map's λ is larger than the error between the discrete-state source and reconstructed ϵ-machine metric entropies.

The minimum and maximum energies for the prototype processes and the Renyi and machine methods are given in Table 3. For the first three processes the appendix gives exact expressions

Process	h	h_μ	h^M	h_μ^M	h^R	h_μ^R
Biased Coin	1	0.9710	1.0000	0.9709	0.9999	0.9710
Golden Mean	0.6942	0.6935	0.6942	0.6936	0.7170	0.7105
Even	0.6942	0.6068	0.6942	0.6067	0.7858	0.6948
Logistic $r = r_M$	0.8232	$\lambda = 0.7887$	0.8232	0.8054	0.8670	0.8387

Table 2 The second and third columns give the topological and metric entropies — defined by Eqs. (44) and (47), respectively — for the three prototype discrete-state processes. The second column in the fourth row gives the topological entropy for the logistic map computed using the kneading determinant with 100 terms and estimating its smallest zero to 1 part in 10^6. As an estimate of h_μ the third column gives the Lyapunov exponent λ for the logistic map averaged over 10^8 iterates. Note that the differences $h - h_\mu$ or $h - \lambda$, i.e. the difference between the values in the second and third columns, gives a rough measure of the inhomogeneity in the asymptotic sequence distribution.[6,27] The fourth and fifth columns give the ϵ-machine topological h^M and metric h_μ^M entropies, defined by Eqs. (44) and (47), for the four reconstructed machines shown in Figures 8 through 11. The Renyi estimates — h^R and h_μ^R — are also given in the next columns. All values have been rounded in the last decimal place.

Process	$\mathcal{U}_{\min}$	$\mathcal{U}_{\max}$	$\mathcal{U}_{\min}^M$	$\mathcal{U}_{\max}^M$	$\mathcal{U}_{\min}^R$	$\mathcal{U}_{\max}^R$
Biased Coin	0.7370	1.3219	0.7366	1.3225	0.7377	1.3285
Golden Mean	0.6610	0.7370	0.6614	0.7364	0.6509	0.8293
Even	0.3685	1.3219	0.3683	1.3225	0.3685	1.3857
Logistic $r = r_M$	—	—	0.5310	0.9619	0.6062	1.1379

Table 3 The minimum and maximum energy densities for the prototype processes and estimated from the machine and Renyi fluctuation spectra for the four example processes. $\mathcal{U}_{\min}$ and $\mathcal{U}_{\max}$ values are obtained from the exact expressions given in the Appendix. The machine quantities are $\mathcal{U}_{\min}^M$ and $\mathcal{U}_{\max}^M$; the Renyi quantities are $\mathcal{U}_{\min}^R$ and $\mathcal{U}_{\max}^R$.

for these quantities. The values are quoted in the second and third columns. The machine-based energies, in the fourth and fifth columns, are computed in a similar manner, but with the empirically estimated transition probabilities. The sixth and seventh columns give the Renyi estimates. Generally, the machine quantities are reasonably accurate; the Renyi estimates, substantially less so. We shall refer back to these results with the next section's discussion of the full fluctuation spectra.

Finally, Table 4 presents the various thermodynamic complexities. The second column gives the statistical complexity C_μ computed from Eq. (49) using the eigenvector expressions derived in the appendix for the prototype processes. The third column lists C_μ^M, the statistical complexity computed from Eq. (49) using the reconstructed machines' state probability vector. The next two columns give the fluctuation complexities Ξ_0 and Ξ_1 from Eqs. (75) and (77), respectively. The latter was estimated by numerical integration of each reconstructed machine's fluctuation spectrum. Notice that the biased coin has positive fluctuation complexities, but zero statistical complexity. It produces fluctuations without any internal memory. By way of contrast, the appendix shows that the golden mean and even process transition probabilities can be changed

so that the opposite is true. That is, in this case they have $C_\mu > 0$, but $\Xi = 0$. Thus, the statistical complexity and fluctuation complexity measure different properties of a process.

Process	C_μ	C_μ^M	Ξ_0	Ξ_1
Biased Coin	0	0.0000	0.5850	0.4220
Golden Mean	0.8631	0.8630	0.0527	0.0380
Even	0.9544	0.9545	0.6619	0.4767
Logistic $r = r_M$	—	1.7699	0.3548	0.2433

Table 4 Complexities for the original processes and those estimated from the reconstructed machines and their fluctuation spectra. C_μ is the statistical complexity from Eq. (49) for the original process. C_μ^M is the estimate obtained from the reconstructed machines. Ξ_0 and Ξ_1 are the two fluctuation complexities estimated from the reconstructed machines' fluctuation spectra. They are calculated via Eqs. (75) and (77), respectively.

Spectroscopy

Figures 12, 13, 14, and 15 compare for each process the histogram — Eqs. (5) and (7) — and Renyi — Eqs. (28) and (31) — fluctuation spectra at $L = 10$ with the asymptotic spectra calculated directly from the ϵ-machines — Eqs. (65) and (66). With $L = 10$, data streams of length $k = 10^7$ give a few percent expected variation in the empirical estimates of the sequence probabilities. This variation can be estimated roughly as $\sqrt{k^{-1}2^{hL}}$, which gives 0.5% to 1.0% variation for the various processes. The expected variation for the least probable sequences can be estimated in a similar fashion, except that $\mathcal{U}_{\max}$ is used instead of h, and this yields a range of variation from 0.5% to 3%. This range of empirical variation is the basis of our choice of $L = 10$, given that we have set $k = 10^7$.

The vertical lines in the figures represent normalized histogram counts for various values of $\mathcal{U}$. The Renyi spectra are plotted as dotted-line curves and the machine fluctuation spectra as smooth curves. The normalization used in the spectra is such that for an unbiased coin, whose fluctuation spectrum would consist of the single point $(s(\mathcal{U}),\mathcal{U}) = (1,1)$, all estimation methods agree. That is, the Renyi and machine spectra consist of this single point and the histogram spectrum consists of a single bin at $\mathcal{U} = 1$ with height 1.

The figures do not show the fluctuation spectra of the discrete-state prototype processes as derived in the appendix. As an alternative, the preceding tables gave numerical values for a number of thermodynamic quantities related to these spectra. The main reason for omitting the sources' spectra is that they are indistinguishable from the spectra of the reconstructed ϵ-machines at the figures' resolution.

Recall from the discussion of the thermodynamics that β is the slope of the $s(\mathcal{U})$ versus $\mathcal{U}$ curve. The solid, diagonal line in the figures represents the identity $s(\mathcal{U}) = \mathcal{U}$ and by

construction it is tangent to $s(\mathcal{U})$ at $\beta = 1$. From Eqs. (60), (65), and (67), we see that at this point $h_\mu = h(1) = s(\mathcal{U}(1))$. We also note — from Eqs. (27) and (67) — that the maximum $s(\mathcal{U}(0))$ of the fluctuation spectrum gives the topological entropy $h = h(0) = s(\mathcal{U}(0))$. In the figures the topological and metric entropies as obtained from the Renyi spectrum are labeled h^R and h_μ^R, respectively; those obtained from the machine spectrum are labeled h^M and h_μ^M, respectively. Their numerical values were quoted in Table 2.

Recall that the difference between the identity and the $s(\mathcal{U})$ curve is the rate function $\mathcal{I}(\mathcal{U})$, Eq. (11), which gives the probability decay rate of the energy $\mathcal{U}$ level set. The minimum of the rate function occurs at $\beta = 1$ and identifies the typical set, whose probability decay is zero. Nonetheless, the probability of individual sequences in the typical set decays at a rate which is the metric entropy. From the figures it is apparent that the most populated energy level set occurs at $\beta = 0$. The sequences in this most populous set have energy $\mathcal{U}(0)$ and probability decay rate equal to the topological entropy h.

As one might expect, there is good agreement between the various methods for the biased coin. (See Figure 12.) The Renyi and machine fluctuation spectra obtained are nearly identical. This was anticipated in the arguments in previous sections about the efficacy of the various techniques in the presence of correlations and the lack of correlation as exhibited in the biased coin histograms of Figure 4.

First, let's consider the histogram spectrum — the vertical lines in Figure 12. For the biased coin the number of energy levels is simply equal to the number of binomial coefficients of order L, which is $L + 1$. Therefore, in the example with $L = 10$ there are 11 energy levels. Recall that the entropy density is proportional to the logarithm of the binomial coefficient itself. Figure 12 seems to show only 9 distinct peaks, or vertical lines, for the histogram spectrum. This is due to the fact that there is only a single way of obtaining the most probable sequence — all 1's — or the least probable sequence — all 0's. This gives two binomial coefficients equal to 1 and so the entropy for these energy levels is $s(\mathcal{U}) = \log_2 1 = 0$. The result is that these two levels are not visible as peaks in the figure.

There are several systematic biases in the Renyi fluctuation spectrum. The first is the consistent overestimation of $s(\mathcal{U})$ at $\beta = 0$ — i.e. an overestimation of the topological entropy. When $\beta = 0$ the Renyi entropy simply counts all sequences. The overestimation is not apparent for the biased coin Renyi spectrum since the lack of correlation leads to a rapid convergence with L in the entropy estimates. This is directly related to the overestimation of the fractal dimension by the state space Renyi spectrum techniques.[5,6]

More interestingly, there are fluctuations about the high-energy histogram peaks as compared to the low-energy peaks. These fluctuations are due to low-probability sequences whose empirical frequency in the data stream is slightly higher or lower than would be obtained by calculating their probability directly from the binomial coefficients. The result is a spread of "spurious" energy levels about the peaks — that is, these are empirical fluctuations of the intrinsic energy

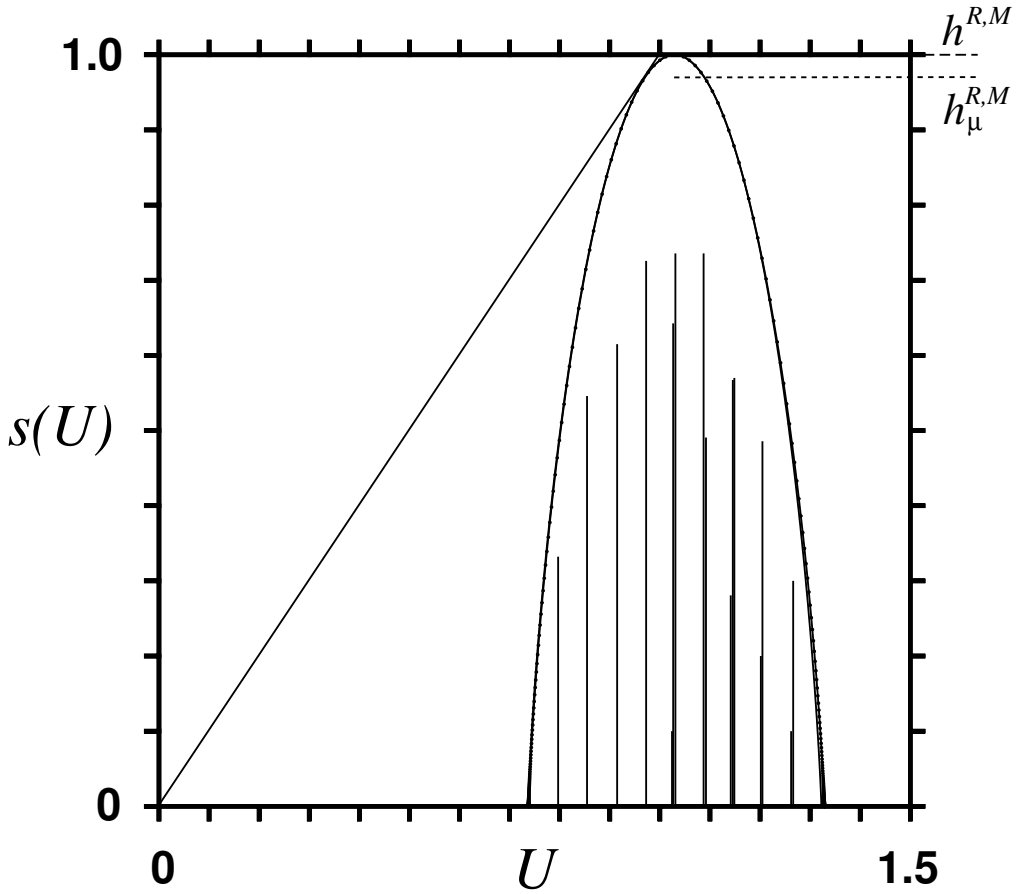

Figure 12 Biased coin fluctuation spectra: The histogram spectrum (vertical lines) and the Renyi spectrum (solid curve with superposed dots) were estimated from histograms as in Figure 4, but at sequence length $L = 10$. In this and the following figures, the histogram spectrum used 300 uniform-width energy bins between $\mathcal{U}_{\min}$ and $\mathcal{U}_{\max}$. The Renyi spectrum is given over $\beta \in [-30, 30]$. The machine fluctuation spectrum (solid curve) was numerically estimated using the reconstructed machine shown in Figure 8 over a range $\beta \in [-30, 30]$. The Renyi and machine spectra are essentially identical apart from the small discrepancy for high-$\mathcal{U}$, low-probability sequences, as discussed in the text. Since the Renyi spectrum is a very good approximation to the actual fluctuation spectrum for the biased coin, the topological and metric entropies are essentially the same: $h^R = h^M = h^{R,M}$ and $h_\mu^R = h_\mu^M = h_\mu^{R,M}$, as noted in the figure.

fluctuations. Spurious energy levels lead to a second systematic bias in the Renyi method's fluctuation spectrum — an overestimate of $\mathcal{U}_{\max}$. This is a consequence of the fact that the Renyi estimate $\mathcal{U}_{\max}^R$ at $\beta \ll 0$ is tied to the least probable sequence in the entire data stream. That sequence's probability is usually underestimated and sometimes grossly so. This results in an increase in its apparent energy.

These biases lead to an overestimate of the entropy density for high-energy, low-probability sequences. This is evident already in the biased coin, though not very pronounced in Figure 12. There are several contributions and so this problem is often the most noticeable. The first two contributions are the biases just outlined — the general increase in the Renyi estimates of

the entropy near $\beta = 0$ and the overestimate of $\mathcal{U}_{\max}$. The overestimate of entropy pulls the spectrum to higher entropy and the highest, empirically-determined energy level pulls the Renyi spectrum to higher energy. Combined with these biases, the convexity of the Renyi fluctuation spectrum produces a smooth interpolation and overestimation.

A third systematic bias in the Renyi spectrum can occur. It results in either an over- or an underestimation of $\mathcal{U}_{\min}$ — i.e. of the most probable sequences. Note that in the low-energy regime the energy levels are quite well-approximated. The spurious energy level effect is much reduced, since the levels are associated with the more probable sequences which are well sampled. The third systematic bias is a finite size effect. Even with exact sequence probabilities, finite L limits the accuracy of the asymptotic energy level estimates. This derives from a mismatch between L and the most probable cycles' length l. If $L = l$ the most probable sequences' energy density is equal to the most probable cycle's energy density $\mathcal{U}_{\min}$. In fact, this occurs when L is any multiple of l. However, if L is not a multiple of l, the most probable sequences' energy density can be an over- or underestimate of the most probable cycle's energy density. And so, on the one hand, $\mathcal{U}_{\min}$ overestimation competes with $s(\mathcal{U})$ overestimation and sometimes wins, resulting in decreased entropy at low energy. On the other hand, $\mathcal{U}_{\min}$ underestimation can exacerbate $s(\mathcal{U})$ overestimation.

The following examples illustrate more clearly these systematic biases. Although subtle in our example of the biased coin, the "spurious" energy level effect for low-probability sequences can be seen as a slight separation between the Renyi and machine spectrum estimates at the high energy end. (Cf. the extremal energy estimates in Table 3.) In the following examples this effect is much more pronounced. The effect is harder to illustrate there since one has to consider constrained multinomial coefficients rather than simple unconstrained binomial coefficients. The cause is the same, though: fluctuations in low probability sequences.

Figure 13 shows the fluctuation spectra for the golden mean process. The energy fluctuations occur in a tighter range due to the relatively small value of the inhomogeneity parameter $h - h_\mu$ as noted in the discussion of Figure 5. The agreement between the Renyi and machine estimates of the topological and metric entropies (cf. Table 2), while clearly not as good as in the case of the biased coin, might still seem reasonable at $L = 10$. But the energy of the antiground state at $\mathcal{U}_{\max}$ is very poorly approximated here. This effect is a result of spurious energy levels, as just discussed. Here low-probability sequences occur with low empirical frequency in the data stream and this yields a high $\mathcal{U}_{\max}$. As anticipated, there is a general sensitivity of the Renyi spectrum to fluctuations in the high-$\mathcal{U}$, low-probability sequences as manifested in the large overestimation of $s(\mathcal{U})$ there.

At the opposite end of the energy range, low $\mathcal{U}$, rather than underestimating $s(\mathcal{U})$ as discussed above, $s(\mathcal{U})$ is actually overestimated as with high $\mathcal{U}$. The overestimation is a consequence of the underestimation of $\mathcal{U}_{\min}$ in concert with the overestimation of topological entropy. It is important to note that these biases are avoided for the machine spectrum, since the reconstructed machine

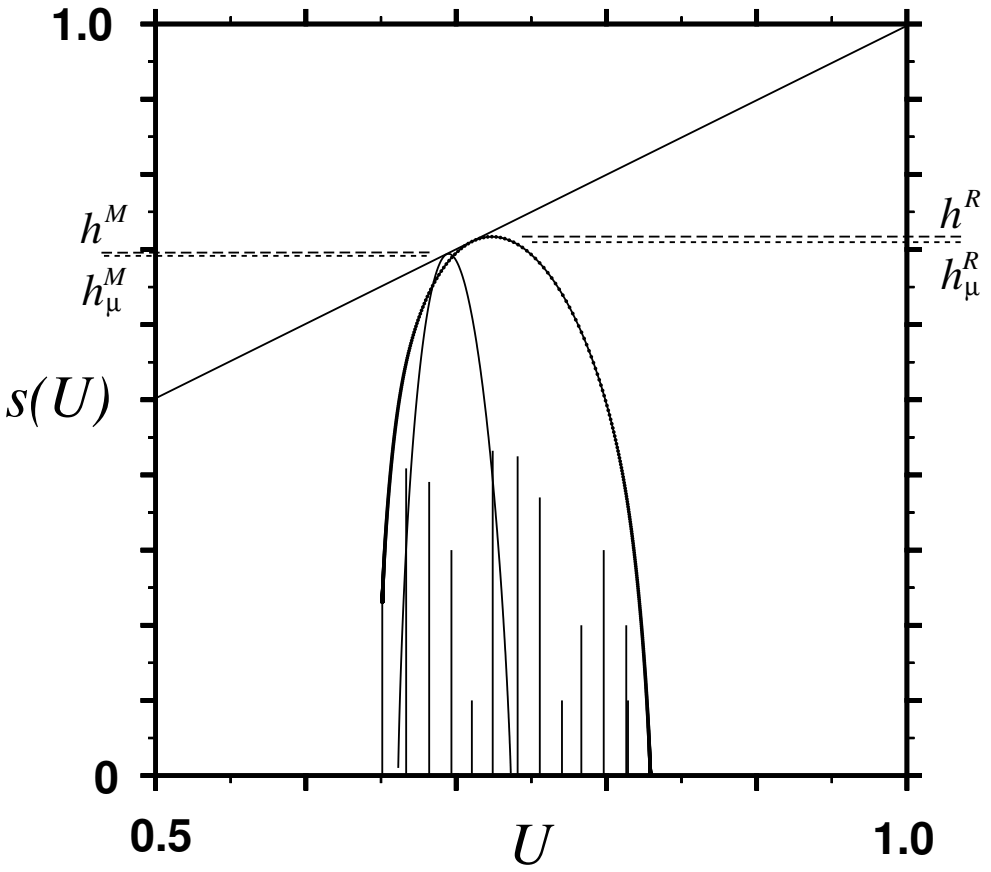

Figure 13 Golden mean process fluctuation spectra: The histogram spectrum (vertical lines) and the Renyi spectrum (solid curve with superposed dots) were estimated from histograms as in Figure 5, but at sequence length $L = 10$. The Renyi spectrum is shown over the range $\beta \in [-100, 100]$. The machine fluctuation spectrum (solid curve) was numerically estimated using the reconstructed machine shown in Figure 9 over the range $\beta \in [-80, 120]$. The $\mathcal{U}$ axis is expanded as compared to the other examples due to the smaller range of fluctuations as already noted in the histogram mosaic of Figure 5. Despite adequate topological and metric entropy estimates in the Renyi spectrum, it overestimates $s(\mathcal{U})$ at both high and low $\mathcal{U}$. Note that at low energy the Renyi spectrum curve intersects the top of a histogram spectrum bin.

allows for a direct, L-independent approximation of the asymptotic spectrum. Sequence length is a relevant parameter in reconstructing the machine. But once this has been accomplished, the asymptotic spectrum is approximated directly from the transition matrices without reference to L. A consequence, already noted, of the narrow energy range and small inhomogeneity parameter $h - h_\mu$ is that the Renyi method yields fairly good approximations for the topological and metric entropies.

In Figure 13 the low energy end of the Renyi spectrum does not extend all the way to the $\mathcal{U}$ axis, but intersects a histogram energy bin. The Renyi spectrum does not intersect the $\mathcal{U}$ axis due to the slow convergence for large positive values of β. In more general cases, which we will not present here, the actual asymptotic spectrum might not go to zero entropy at the

extremal energies due to a degeneracy in the ground and antiground states. That is, for stationary processes described by irreducible Markov chains one can observe either or both $s(\mathcal{U}_{\max}) > 0$ and $s(\mathcal{U}_{\min}) > 0$. As shown in the appendix, though, these values vanish for the golden mean process.

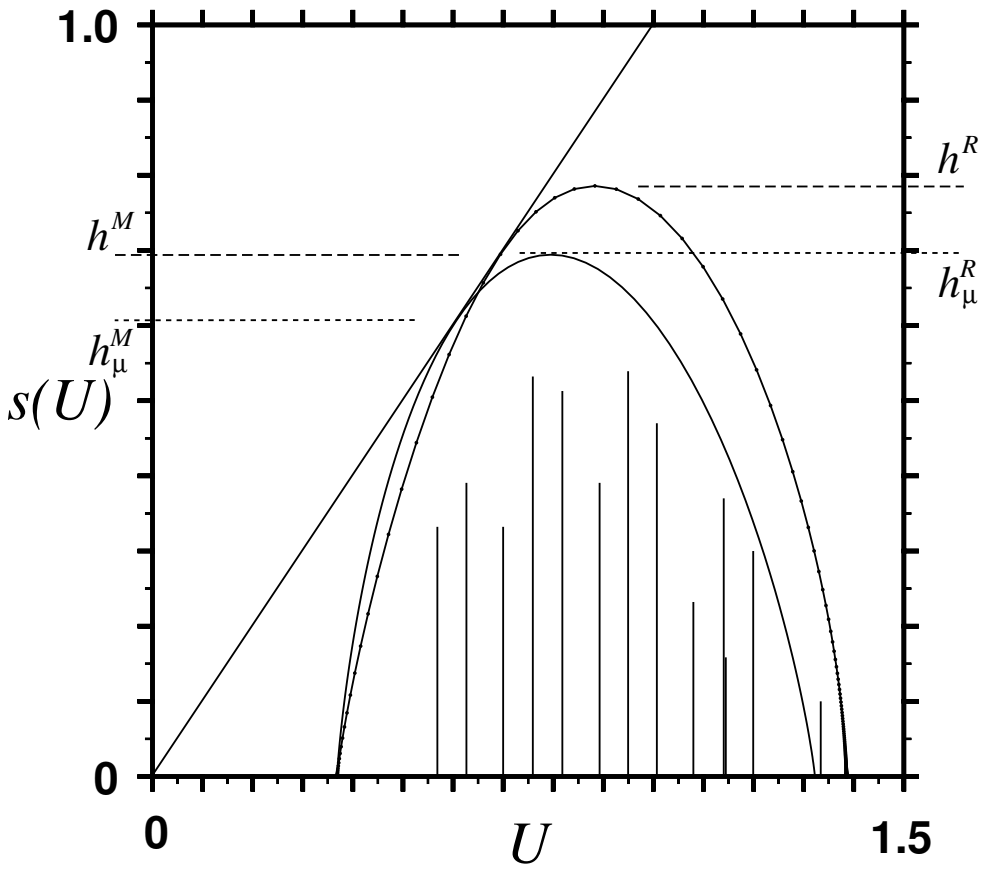

Figure 14 Even process fluctuation spectra: The histogram spectrum (vertical lines) and the Renyi spectrum (solid curve with superposed dots) were estimated from histograms as in Figure 6, but at sequence length $L = 10$. The Renyi spectrum is shown over the range $\beta \in [-70, 30]$. The machine fluctuation spectrum (solid curve) was numerically estimated using the reconstructed machine shown in Figure 10 over the range $\beta \in [-20, 20]$. The Renyi spectrum estimate of the topological and metric entropies is in substantially larger error than for the golden mean process. Worse, one sees that $h^R_\mu > h^M$. Additionally, the Renyi spectrum method overestimates $s(\mathcal{U})$ at large energy, but underestimates $s(\mathcal{U})$ at low energy.

For the even system, shown in Figure 14, we see the expected high-$\mathcal{U}$ overestimation of $s(\mathcal{U})$ by the Renyi method similar to, but more pronounced than, that observed for the previous cases. (Cf. the difference in energy ranges for the plots.) Also note that the machine and Renyi estimates of $\mathcal{U}_{\min}$ agree. This plus the Renyi method's overestimation of the topological entropy leads to the slight, but consistent underestimation of $s(\mathcal{U})$ for the low-$\mathcal{U}$, high-probability fluctuations. For the even process the topological and metric entropies are poorly approximated by the Renyi

spectrum for the length $k = 10^7$ data stream and length $L = 10$ sequences considered here. (Cf. Table 2). Note also that h_μ^R is slightly larger than h^M. Since h^M is an excellent approximation to the actual topological entropy — see Table 2 — the h_μ^R overestimation violates the theoretical inequality $h_\mu \leq h$. The Renyi method seems to be indicating not only that there is a problem in accounting for the information in the sequence distribution, but also that there are more sequences than are actually there.

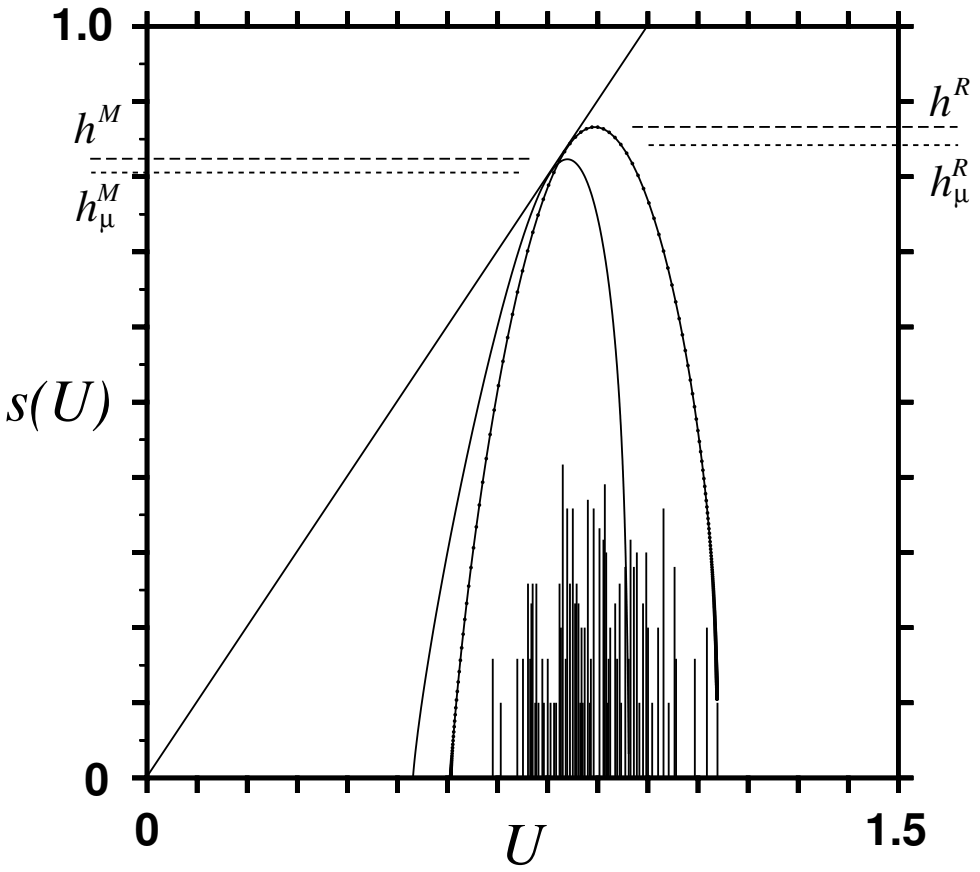

Figure 15 Logistic map process fluctuation spectra: The histogram spectrum (vertical lines) and the Renyi spectrum (solid curve with superposed dots) were estimated from histograms as in Figure 7, but at sequence length $L = 10$. The Renyi spectrum is shown over the range $\beta \in [-50, 50]$. The machine fluctuation spectrum (solid curve) was numerically estimated using the reconstructed Misiurewicz machine shown in Figure 11 over the range $\beta \in [-100, 40]$. As seen for the even process the Renyi spectrum poorly approximates the topological and metric entropies and gives $h_\mu^R > h^M$. It and the direct histogram spectrum substantially overestimate $\mathcal{U}_{\min}$ and $\mathcal{U}_{\max}$. They also overestimate the high-energy, low-probability sequence entropy, while underestimating the low-energy, high-probability sequence entropy.

Figure 15 shows the fluctuation spectra for the Logistic map process. Unlike the prototype discrete-state processes, one cannot compute a first-principle $s(\mathcal{U})$ for the logistic map considered as a continuum-state process. Instead, the calculation of its fluctuation spectrum typically uses

either a generating or a Markov partition of the interval. Even in the latter case, in which one could directly calculate $s(\mathcal{U})$, numerical estimates of the transition probabilities would be required. ϵ-machine reconstruction from a generating partition, though, implements the necessary construction of a Markov partition through its inference of hidden states and its estimation of transition probabilities. There are alternatives to studying the fluctuation spectrum this way. One of these requires that the partition element size vanish and that the sequence length diverge. Another, which studies the fluctuations in Lyapunov exponent, requires knowledge of the equations of motion and numerical simulation.[6,9]

As seen in Figure 15, the Renyi method gives a generally poor approximation of the fluctuation spectrum. This example shows an extreme case of poor high-energy approximation. However, the Renyi method not only consistently overestimates the high-$\mathcal{U}$, low-probability fluctuations as do the previous examples but also shows an extreme case of finite size effects by substantially underestimating the low-$\mathcal{U}$, high-probability fluctuations. This results in a significant overestimation of both the ground state energy $\mathcal{U}_{\min}$ and the antiground state energy $\mathcal{U}_{\max}$. (Cf. Table 3). Here, as in the case of the even system, $h_\mu^R > h^M$, but the discrepancy is even larger. While better approximations using the histogram and Renyi methods are to be expected for larger L, Figure 15 clearly indicates the systematic biases of such methods. They fail to give good approximations to the asymptotic fluctuation spectrum even with $k = 10^7$ and $L = 10$ for even this relatively low complexity system.

Conclusions

Using some fairly simple, low-statistical-complexity processes the preceding comparative study of fluctuation spectra estimation has illustrated the central importance of carefully choosing a model class when analyzing data. While there are other, perhaps more important reasons for taking care when choosing a model class, we have been concerned with showing that the proper choice is essential when approximating the statistical fluctuations for an observed process. At some basic level, the very notion of a "fluctuation" depends on prior explicit or implicit modeling assumptions. Choosing histograms as a model class — as done explicitly in the histogram spectra methods and as done implicitly in the Renyi methods — only the most simplistic extrapolations of the data are possible. Introducing a slightly subtler class, ϵ-machines, led to direct methods for obtaining good approximations of asymptotic fluctuation spectra.

Via large deviation theory we introduced the scaling indices for probabilities as thermodynamic potentials. We produced specific algorithms for computing the fluctuation spectra in terms of sequence histograms and Renyi entropy and, most importantly, for estimating the asymptotic fluctuation spectrum using the Markovian thermodynamic properties of ϵ-machines. We noted the distinction between these techniques and those used for estimating the fluctuation spectrum from state space distributions and Lyapunov spectra. The general efficacy of ϵ-machines turns

on their ability to represent processes without imposing absolute statistical independence of or uniform scaling over finite length sequences. Their states are defined, in fact, in terms of conditional independence. This allows for the proper accounting of a stationary process's structure and of the effect the process's statistical complexity has on the convergence of various statistics.

Model classes which do not allow for conditional independence typically require more data and larger, more complex estimated models. Indeed, the spectrum estimation problems we have described will be substantially exacerbated compared to the present examples in experimental situations with less data and for processes with higher statistical and fluctuation complexities. Conversely, the relative advantages of ϵ-machines will become more apparent.

The comparison of the fluctuation spectra demonstrated that ϵ-machine reconstruction allowed for an exact calculation of the topological entropy. This was not possible for the other methods. The asymptotic fluctuation spectra were calculated solely on the basis of the Markov structure of the ϵ-machines. Thus, the inference of "hidden" states plays a central role in estimating properties of the observed sequences. The comparison also showed that the machine methods neither overestimated the high-energy sequences nor underestimated the low-energy sequences; as did the other techniques. Even with the large simulated data sets — data sets far larger than would be expected in typical experiments — the conventional histogram and Renyi methods fail badly in terms of producing acceptable estimates of entropies and fluctuations. ϵ-machine reconstruction also gave far better estimates of the ground and antiground states — the highest and lowest probability sequences, respectively — than did the other methods. This is important in the analysis of extreme fluctuations — that is, of events with high or low probability. The problematic performance of the histogram and Renyi methods is made all the more dramatic once one realizes that the errors are in scaling exponents. Thus, the errors in metric entropy (say) are magnified exponentially and indicate huge misestimations of the probabilities for even moderately long sequences.

Regarding the thermodynamic analysis of ϵ-machines, we note finally that the asymptotic fluctuation spectra are determined solely by the underlying Markov process structure. And the latter is independent of the edge labels and measurement symbols. In turn, the spectra are independent of the detailed structure of the sequences. This indicates that fluctuation spectra are by no means complete invariants for a process. A similar limitation applies to the thermodynamic approach in general. It doesn't even account for statistical complexity.[28] To determine important structural characteristics associated with a process — e.g. strict soficity or intrinsic computational capability — means other than fluctuation spectra and the thermodynamic formalism are necessary.[29,34]

Acknowledgments

The authors thank John Ashley, Jim Hanson, Brian Marcus, Jeff Scargle, and Dan Upper for

useful discussions. This work was funded in part by AFOSR 91-0293. Portions were completed while KY was funded by a National Research Council Postdoctoral fellowship.

Appendix A Thermodynamic Details

This appendix collects together the calculation of a number of thermodynamic quantities for the biased coin, golden mean, and even processes, with common variable $p \equiv \Pr(s = 1|v)$.

Biased Coin Process

The biased coin as represented (say) in Figure 1 has a single state and two edges. To calculate the thermodynamic properties as a function of the bias p we change the representation to the edge graph. The edge graph of a machine is a machine whose states are the edges of the original machine. In this representation the biased coin connection matrix is

$$T = \begin{pmatrix} p & q \\ p & q \end{pmatrix} \tag{80}$$

where $p = \Pr(s = 1|v = \mathbf{A})$ and $q = 1 - p$. The parameterized connection matrix is then

$$T_\beta = \begin{pmatrix} e^{\beta \ln p} & e^{\beta \ln q} \\ e^{\beta \ln p} & e^{\beta \ln q} \end{pmatrix} \equiv \begin{pmatrix} \hat{p} & \hat{q} \\ \hat{p} & \hat{q} \end{pmatrix} \tag{81}$$

To calculate $s(\mathcal{U})$ there are several items required as a function of β: the principal eigenvalue λ_β of T_β and its left and right eigenvectors as defined via Eqs. (57). One finds

$$\lambda_\beta = \hat{p} + \hat{q} \tag{82}$$

and

$$\vec{l}_\beta = l(\hat{p}, \hat{q}) \tag{83}$$

and

$$\vec{r}_\beta = r \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{84}$$

where r and l are undetermined constants. Then from Eq. (63) the stochasticized connection matrix is

$$\mathrm{S}_\beta = (\hat{p} + \hat{q})^{-1} \begin{pmatrix} \hat{p} & \hat{q} \\ \hat{p} & \hat{q} \end{pmatrix} \tag{85}$$

The stationary state probability distribution is given by Eq. (64) which here is

$$\vec{\mathrm{p}}_\beta = (\hat{p} + \hat{q})^{-1} (\hat{p}, \hat{q}) \tag{86}$$

The edge graph is not a minimal representation of the biased coin, Figure 1 is, however. Therefore, computing the Shannon entropy of the state distribution in Eq. (86) yields too high a

value for the process's memory capacity. That is, this value is not the statistical complexity of the biased coin, which is zero. Note that $\lambda_0 = 2$ and $\vec{\mathrm{p}}_0 = \left(\frac{1}{2}, \frac{1}{2}\right)$, and that $\lambda_1 = 1$ and $\vec{\mathrm{p}}_1 = (p, q)$.

With the state probabilities and the stochasticized machine, the entropy rate $\mathcal{S}(\beta) = s(\mathcal{U}(\beta)) = h_\mu(\mathrm{S}_\beta)$ follows directly

$$\mathcal{S}(\beta) = \frac{-\hat{p}\log_2\hat{p} - \hat{q}\log_2\hat{q}}{\hat{p}+\hat{q}} + \log_2(\hat{p}+\hat{q}) \tag{87}$$

At $\beta = 0$ we verify that $\mathcal{S}(0) = 1$ and at $\beta = 1$ that

$$\mathcal{S}(1) = -p\log_2 p - q\log_2 q \tag{88}$$

The internal energy density follows from Eq. (66)

$$\mathcal{U}(\beta) = \frac{-\hat{p}\log_2\hat{p} - \hat{q}\log_2\hat{q}}{\beta(\hat{p}+\hat{q})} \tag{89}$$

Note that at $\beta = 0$ we have

$$\mathcal{U}(0) = -\log_2 p - \log_2 q \tag{90}$$

and at $\beta = 1$, we verify that $\mathcal{U}(1) = \mathcal{S}(1)$, as it should.

The energy extremes

$$\mathcal{U}_{\min} = \lim_{\beta\to\infty} \mathcal{U}(\beta) \tag{91}$$

and

$$\mathcal{U}_{\max} = \lim_{\beta\to-\infty} \mathcal{U}(\beta) \tag{92}$$

require a little care. In particular, the most and least probable paths exchange identity at a value of $p^* = \frac{1}{2}$. And this affects the ratios of terms important to the above limits. If $p < p^*$, then we find

$$\mathcal{U}_{\min} = -\log_2 q \tag{93}$$

and

$$\mathcal{U}_{\max} = -\log_2 p \tag{94}$$

If $p > p^*$, we have the opposite association of minimum and maximum energies.

We note that $\lim_{\beta\to\infty} \mathcal{S}(\beta) = \lim_{\beta\to-\infty} \mathcal{S}(\beta) = 0$. And so the minimum and maximum energy states have a small number of configurations — in fact, at most one: either the edge sequence $\mathrm{AAAAAAAA}\ldots$ or the edge sequence $\mathrm{BBBBBBBB}\ldots$, in which edge A is associated with transition probability p and B with q. That is, the ground and antiground states are nondegenerate.

Golden Mean Process

Recall that the stochastic connection matrix for our version of the golden mean process is

$$T = \begin{pmatrix} p & q \\ 1 & 0 \end{pmatrix} \tag{95}$$

where $p = \Pr(s = 1 | v = \mathbf{A})$ and $q = 1 - p$. The parameterized connection matrix is then

$$T_\beta = \begin{pmatrix} e^{\beta \ln p} & e^{\beta \ln q} \\ 1 & 0 \end{pmatrix} \equiv \begin{pmatrix} \hat{p} & \hat{q} \\ 1 & 0 \end{pmatrix} \tag{96}$$

To calculate $\mathcal{S}(\beta) = s(\mathcal{U}(\beta))$ there are several items required as a function of β. First, the principal eigenvalue λ_β of T_β and its left and right eigenvectors as defined via Eqs. (57). One finds

$$\lambda_\beta = \frac{\hat{p} + \sqrt{\hat{p}^2 + 4\hat{q}}}{2} \tag{97}$$

and

$$\vec{l}_\beta = l\big(\lambda_\beta \hat{q}^{-1}, 1\big) \tag{98}$$

and

$$\vec{r}_\beta = r \begin{pmatrix} \lambda_\beta \\ 1 \end{pmatrix} \tag{99}$$

where r and l are again undetermined constants. The stationary state probability distribution is given by Eq. (64) which yields

$$\vec{\mathrm{p}}_\beta = \big(\lambda_\beta^2/\hat{q} + 1\big)^{-1} \big(\lambda_\beta^2/\hat{q}, 1\big) \tag{100}$$

Note that $\lambda_0 = \big(1 + \sqrt{5}\big)/2$ — the golden mean ϕ — and $\vec{\mathrm{p}}_0 = \big(1 + \phi^2\big)^{-1} \big(\phi^2, 1\big)$, and that $\lambda_1 = 1$ and $\vec{\mathrm{p}}_1 = (1 + q)^{-1}(1, q)$. Then from Eq. (63) the stochasticized connection matrix is

$$\mathrm{S}_\beta = \begin{pmatrix} \lambda_\beta^{-1} \hat{p} & \lambda_\beta^{-2} \hat{q} \\ 1 & 0 \end{pmatrix} \tag{101}$$

With the state probabilities and the stochasticized machine, the entropy rate $\mathcal{S}(\beta) = h_\mu\big(\mathrm{S}_\beta\big)$ follows directly

$$\mathcal{S}(\beta) = \frac{-1}{\hat{q} + \lambda_\beta^2} \big\{ \hat{p} \lambda_\beta \log_2 \hat{p} + \hat{q} \log_2 \hat{q} - \big(\hat{p} \lambda_\beta + 2\hat{q}\big) \log_2 \lambda_\beta \big\} \tag{102}$$

At $\beta = 0$ we verify that $\mathcal{S}(0) = \log_2 \phi$ and at $\beta = 1$ that

$$\mathcal{S}(1) = \frac{-p \log_2 p - q \log_2 q}{1 + q} \tag{103}$$

The internal energy density follows from Eq. (66)

$$\mathcal{U}(\beta) = \frac{-\hat{p}\lambda_\beta \log_2 \hat{p} - \hat{q}\log_2 \hat{q}}{\beta\left(\hat{q} + \lambda_\beta^2\right)} \tag{104}$$

Note that at $\beta = 0$ we have

$$\mathcal{U}(0) = -\frac{\phi \log_2 p + \log_2 q}{1+\phi^2} \tag{105}$$

and at $\beta = 1$, we verify that $\mathcal{U}(1) = \mathcal{S}(1)$.

The energy extremes

$$\mathcal{U}_{\min} = \lim_{\beta\to\infty} \mathcal{U}(\beta) \tag{106}$$

and

$$\mathcal{U}_{\max} = \lim_{\beta\to-\infty} \mathcal{U}(\beta) \tag{107}$$

require more care than before. In particular, the most and least likely paths exchange identity at a value of $p^* = \phi - 1$. This value can be obtained by considering all paths, beginning in the start state $\mathbf{A}$ and returning to state $\mathbf{A}$, i.e. paths of the form $\mathbf{A}...\mathbf{A}$. Paths of this type at any length will consist of concatenations of the elementary cycles $\mathbf{AA}$ and $\mathbf{ABA}$. The energies of these paths are

$$\mathcal{U} = -\log_2 p \tag{108}$$

and

$$\mathcal{U} = -\frac{1}{2}\log_2 q \tag{109}$$

respectively. Therefore the minimum and maximum energy paths, and all others, will have the same energy when the elementary cycles have the same energy. This condition determines the crossover energy and the value of p for which it occurs. The latter is calculated as follows.

$$\begin{aligned} \log_2 p^* &= \frac{1}{2}\log_2 q^* \\ &= \frac{1}{2}\log_2 (1-p^*) \end{aligned} \tag{110}$$

and so

$$(p^*)^2 + p^* - 1 = 0 \tag{111}$$

The solution for which $p^* > 0$ is $p^* = \phi - 1$.

Whether $p < p^*$ or $p > p^*$ affects the ratios of terms important to the above limits. If $p < p^*$, then we find

$$\mathcal{U}_{\min} = -\frac{1}{2}\log_2 q \tag{112}$$

and

$$\mathcal{U}_{\max} = -\log_2 p \tag{113}$$

If $p > p^*$, we have the opposite association of minimum and maximum energies.

At $p = p^*$, the fluctuation complexity vanishes since $\mathcal{U}_{\min} = \mathcal{U}_{\max}$, yet the statistical complexity — and so the amount of memory in the process — is positive: $C_\mu \approx 0.8505$ bits, but $\Xi_i = 0$.

We note that $\lim_{\beta\to\infty} \mathcal{S}(\beta) = \lim_{\beta\to-\infty} \mathcal{S}(\beta) = 0$. And so the minimum and maximum energy states have a small number of configurations — in fact, at most two: either the state sequence $\mathbf{AAAAAAA}\ldots$ or the sequences $\mathbf{ABABABA}\ldots$ and $\mathbf{BABABAB}\ldots$.

Even Process

The even and golden mean processes share the same Markovian state transition structure for their recurrent states. We can ignore the even process's transient state structure in the limit of infinitely long sequences. And so, asymptotically, the even process's thermodynamic properties are given by the golden mean analysis just outlined. The sole difference is that we must swap p and q in the expressions above for the golden mean process if we use $p = \Pr(s = 1|v = \mathbf{B})$ for the even process. We also note that the even process has an infinite number of configurations at the extremal energies. Taking into account the different naming of the recurrent states for the even process compared to that for the golden mean process, the even process's extremal configurations have infinitely long tails of the golden mean extremal sequences preceded by arbitrary length, but transient state sequences of the form $\mathbf{A}^n\mathbf{B}, n = 1, 2, 3, \ldots$. Thus, there is still a relatively small — subexponential — number of extremal configurations.

Bibliography

1. B. B. Mandelbrot. Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier. *J. Fluid. Mech.*, 62:331, 1974.
2. R. Bowen. *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms*, volume 470 of *Lect. Notes Math.* Springer-Verlag, Berlin, 1975.
3. D. Ruelle. *Thermodynamic Formalism*. Addision Wesley, Reading, MA, 1978.
4. R. S. Ellis. *Entropy, Large Deviations, and Statistical Mechanics*, volume 271 of *Grund. der Math. Wissens.* Springer-Verlag, New York, 1985.
5. T.C. Halsey, M.H. Jensen, L.P. Kadanoff, I. Procaccia, and B.I. Shraiman. Fractal measures and their singularities: the characterization of strange sets. *Phys. Rev. A*, 33:1141, 1986.
6. G. Paladin and A. Vulpiani. Anomolous scaling laws in multifractal objects. *Phy. Rep.*, 156:148, 1987.
7. B. B. Mandelbrot. Fractals in geophysics. *Pure App. Geop.*, 131:5, 1989.
8. Y. Oono. Large deviation and statistical physics. *Prog. Theo. Phys.*, 99:165, 1989.
9. R. Badii. Conservations laws and thermodynamic formalism for dissipative dynamical systems. *Riv. del Nuovo Cimento*, 12:1, 1989.
10. J. P. Crutchfield and K. Young. Inferring statistical complexity. *Phys. Rev. Let.*, 63:105, 1989.
11. P. Grassberger and I. Proccacia. Characterization of strange attractors. *Phys. Rev. Lett.*, 50:346, 1983.
12. T. Bedford, M. Keane, and C. Series. *Ergodic theory, symbolic dynamics, and hyperbolic spaces*. Oxford University Press, New York, 1991.
13. P. Cvitanovic. Periodic orbits as the skeleton of classical and quantum chaos. *Physica D*, 51:138, 1991.
14. J. P. Crutchfield and B. S. McNamara. Equations of motion from a data series. *Complex Systems*, 1:417, 1987.
15. J. P. Crutchfield and N. H. Packard. Symbolic dynamics of noisy chaos. *Physica*, 7D:201, 1983.
16. P. Grassberger. Toward a quantitative theory of self-generated complexity. *Intl. J. Theo. Phys.*, 25:907, 1986.
17. P. Billingsley. Statistical methods in markov chains. *Ann. Math. Stat.*, 32:12, 1961.
18. H. G. Schuster. *Deterministic Chaos: An Introduction*. VCH Publishing, New York, 1988.
19. C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, Champaign-Urbana, 1962.

20. T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, New York, 1991.
21. J. A. Bucklew. *Large Deviation Techniques in Decision, Simulation, and Estimation*. Wiley-Interscience, New York, 1990.
22. K. Young. Large deviations in dynamical systems. unpublished, 1992.
23. A. Renyi. Some fundamental questions of information theory. In *Selected Papers of Alfred Renyi, Vol. 2*, page 526. Akademii Kiado, Budapest, 1976.
24. J. P. Crutchfield and K. Young. Computation at the onset of chaos. In W. Zurek, editor, *Entropy, Complexity, and the Physics of Information*, volume VIII of *SFI Studies in the Sciences of Complexity*, page 223. Addison-Wesley, 1990.
25. J. E. Hopcroft and J. D. Ullman. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, Reading, MA, 1979.
26. W. Parry and S. Tuncel. *Classification Problems in Ergodic Theory*, volume 67 of *London Mathematical Society Lecture Notes Series*. Cambridge University Press, London, 1982.
27. J. P. Crutchfield and N. H. Packard. Symbolic dynamics of one-dimensional maps: Entropies, finite precision, and noise. *Intl. J. Theo. Phys.*, 21:433, 1982.
28. J. P. Crutchfield. Semantics and thermodynamics. In M. Casdagli and S. Eubank, editors, *Nonlinear Modeling and Forecasting*, volume XII of *Santa Fe Institute Studies in the Sciences of Complexity*, page 317, Reading, Massachusetts, 1992. Addison-Wesley.
29. K. Young. *The Grammar and Statistical Mechanics of Complex Physical Systems*. PhD thesis, University of California, Santa Cruz, 1991. published by University Microfilms Intl, Minnesota.
30. B. Weiss. Subshifts of finite type and sofic systems. *Monastsh. Math.*, 77:462, 1973.
31. P. Collet, J. P. Crutchfield, and J. P. Eckmann. Computing the topological entropy of maps. *Comm. Math. Phys.*, 88:257, 1983.
32. J. Milnor and W. Thurston. On iterated maps of the interval. *Springer Lecture Notes*, 1342:465 – 563, 1988.
33. J. W. Gibbs. *Elementary Principles of Statistical Mechanics*. Dover, New York, 1960.
34. J. P. Crutchfield and K. Young. ϵ-machine spectroscopy. in preparation, 1993.
35. A. N. Kolmogorov. Three approaches to the concept of the amount of information. *Prob. Info. Trans.*, 1:1, 1965.
36. G. Chaitin. On the length of programs for computing finite binary sequences. *J. ACM*, 13:145, 1966.
37. J. E. Bates and H. K. Shepard. Information fluctuation as a measure of complexity. Technical Report preprint, University of New Hampshire, Durham, 1991.

38. J. D. Scargle, D. L. Donoho, J. P. Crutchfield, T. Steiman-Cameron, J. Imamura, and K. Young. The quasi-periodic oscillations and very low frequency noise of Scorpius X-1 as transient chaos: A dripping handrail? *Astrophy. J. Lett.*, 411:L91, 1993.
39. J. P. Crutchfield. *Noisy Chaos*. PhD thesis, University of California, Santa Cruz, 1983. published by University Microfilms Intl, Minnesota.
40. J.-P. Eckmann and D. Ruelle. Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.*, 57:617, 1985.