Entrance and escape dynamics for the typical set

Schuyler B. Nicholson,[1] Jonah S. Greenberg,[1] and Jason R. Green[1,2,3,*]
[1]*Department of Chemistry, University of Massachusetts Boston, Boston, Massachusetts 02125, USA*
[2]*Department of Physics, University of Massachusetts Boston, Boston, Massachusetts 02125, USA*
[3]*Center for Quantum and Nonequilibrium Systems, University of Massachusetts Boston, Boston, Massachusetts 02125, USA*

(Received 5 September 2017; published 29 January 2018)

According to the asymptotic equipartition property, sufficiently long sequences of random variables converge to a set that is typical. While the size and probability of this set are central to information theory and statistical mechanics, they can often only be estimated accurately in the asymptotic limit due to the exponential growth in possible sequences. Here we derive a time-inhomogeneous dynamics that constructs the properties of the typical set for all finite length sequences of independent and identically distributed random variables. These dynamics link the finite properties of the typical set to asymptotic results and allow the typical set to be applied to small and transient systems. The main result is a geometric mapping—the triangle map—relating sequences of random variables of length n to those of length $n+1$. We show that the number of points in this map needed to quantify the properties of the typical set grows linearly with sequence length, despite the exponential growth in the number of typical sequences. We illustrate the framework for the Bernoulli process and the Schlögl model for autocatalytic chemical reactions and demonstrate both the convergence to asymptotic limits and the ability to reproduce exact calculations.

DOI: 10.1103/PhysRevE.97.012146

I. INTRODUCTION

Typical behaviors lie at the heart of statistical mechanics [1]. Asymptotic theories, such as large deviation theory [2–4] and equilibrium statistical mechanics [5,6], are effective, in part, because random variables converge to their "typical" value in the appropriate asymptotic limits. Take a monatomic gas of N atoms in thermal equilibrium with a heat bath. For this system, the relative standard deviation of the energy is $\sigma(E)/\langle E\rangle = O(N^{-1/2})$ [7]. As the number of atoms becomes large, the size of deviations from the mean become relatively small. For example, when the number of atoms is 10^{18}, the relative error is $\sigma(E) \simeq 10^{-9}\langle E\rangle$. Only when this relative error is small can systems be well described by their mean or typical behavior. Many tools are available to describe systems at and away from equilibrium, prominent examples being fluctuation theorems [8–10] and maximum entropy approaches [11,12]. What remains open is how to precisely and accurately quantify the typical states of systems that cannot solely be described by their mean behavior. Systems, including molecular machines [13–15] and single molecules [16,17], can exhibit large fluctuations [18] in structure, energy, or position.

Another definition of typicality exists in information theory. There, the asymptotic equipartition property (AEP) [19,20] says that sequences of random variables converge to a high-probability subset: the typical set, A_ϵ^n. Take a system described by a finite set of random variables or states, ω, of size $|\omega| = M$. The states could represent the sides of a coin, different chemical species [21], or the coarse-grained regions of a partitioned dynamical system [22]. The AEP states that for sequences of length n, $\hat{\omega}_n = (\omega_1,\omega_2,\ldots,\omega_n)$, in the limit where $n \to \infty$, the sample entropy of the typical sequences converges to the entropy rate h_μ: $-n^{-1}\ln\mu(\hat{\omega}_n) \to h_\mu$. These sequences occur with probability $\sum_{\hat{\omega}_n\in A_\epsilon^n}\mu(\hat{\omega}_n) \sim 1$, constitute the typical set, and determine average behavior. In dynamical systems theory language, the entropy rate h_μ is the Kolmogorov-Sinai (KS) entropy [23,24]. There are recent applications of the AEP to irreversibility in stationary Markov processes [25], relations to the Fisher information [26], and the harnessing of fluctuations for thermodynamic function [27]. These results all rely on an asymptotic limit, a situation we avoid here.

The existence of the typical set was first shown for finite alphabets generating independent identically distributed (i.i.d.) sequences by Shannon [28] and McMillan [29]. It was generalized to stationary-ergodic processes by Breiman for finite alphabets. Chung extended the typical set to countably infinite alphabets under the condition $h_\mu < \infty$ [30,31]. As an aside, caution is necessary for infinite alphabets [32] and correlated finite alphabets [19] where there are examples of divergent entropy rates. The typical set is fundamental to information theory, where it is essential to limits on the coding and transmission of information. For example, the logarithm of the size of the typical set (per symbol) is a bound on the rate that information can be transmitted [33].

Though underappreciated, the typical set does exist for finite sequences. However, there are challenges to an accurate and predictive theory for its properties. One challenge is that the convergence rate theories based on statistical moments do not always give accurate bounds [34–37]. Another challenge is that for long, but finite sequences, there is an exponential growth in possibilities and explicitly generating each sequence becomes intractable. Given a system with M states, the number of possible sequences often grows as $M^n = e^{nh_{\text{top}}}$, where n is the

*jason.green@umb.edu

length of a sequence of states and $h_{\text{top}} = \ln M$ is the growth rate or topological entropy rate [38–40]. Given only $M = 8$ states and $n = 12$, there are $8^{12} = 4.4 \times 10^{12}$ possible sequences, which is comparable to the number of galaxies in the known universe [41]. In this work, we address these challenges.

Here we present a quantitative framework for the typical set [19] that bridges the gap between brute force calculations of all possible sequences and asymptotic approximations. In this framework, we evolve densities of sequences which represent observables of the typical set, such as size $|A_\epsilon^n|$ or probability $\mu\{A_\epsilon^n\}$. Previous work introduced a variational version of the typical set [20,42] that avoids asymptotic limits. The complementary formalism introduced here avoids the combinatorial explosion of sequences. We will introduce the framework in several parts. The space of finite length sequences will be partitioned into three sets or "macrosequences": one typical and two atypical sets. These macrosequences describe how ensembles of typical and atypical sequences change as a function of the sequence length. Transitions in and out of the typical set are described by a time-inhomogeneous dynamics. For these dynamics, we define a discrete dynamical system with a geometric interpretation that maps the exact size and probability of the typical set over n. Together the macrosequences and associated dynamics form an object similar to ϵ-machines in computational mechanics [43–45]. Both our geometric construction and ϵ-machines provide a simplified description of a system by encoding all possible histories into possible futures.

II. BACKGROUND AND NOTATION

Consider a particular sequence $\hat{\omega}_n$ generated by some dynamical process with joint probability $\mu(\hat{\omega}_n)$. Though we will refer to n as the length of the sequence, it could also be a dimensionless measure of time, $n = t/\Delta t$. Individual states will be labeled by $j = 1,2,\ldots,M$. We will assume all states are independent and the probability distribution over the states p_j, such that $\sum_j p_j = 1$, is stationary with respect to n. Sequences are then independent, identically distributed (i.i.d.) random variables. While the marginal probability p is stationary, the joint probability over sequences need not be stationary. The dynamics can generate M^n possible sequences, and only in the infinite limit does the difference in probability between any two sequences go to zero through the AEP:

$$\lim_{n\to\infty} \mu(\hat{\omega}_{n+1}) - \mu(\hat{\omega}_n) \sim e^{-nh_\mu}(e^{h_\mu} - 1) \sim 0. \tag{1}$$

For i.i.d. random variables, the entropy growth rate h_μ is equivalent to the Shannon entropy, $h_\mu = H = -\sum_j p_j \ln p_j$, which depends only on p_j and measures the average surprise of observing state j. More generally, however, the entropy rate,

$$h_\mu \equiv h_\mu(n) = -n^{-1} \sum_{\hat{\omega}_n} \mu(\hat{\omega}_n) \ln \mu(\hat{\omega}_n), \tag{2}$$

depends on the joint distribution. The entropy rate is central to the definition of the typical set for all sequences of length n [19]:

$$A_\epsilon^n \equiv \left\{ e^{-n(h_\mu+\epsilon)} \leqslant \mu(\hat{\omega}_n) \leqslant e^{-n(h_\mu-\epsilon)} \right\}. \tag{3}$$

The parameter $\epsilon \in \mathbb{R}^+$ is fixed and, together with n, defines the neighborhood of typical sequences around the entropy rate.

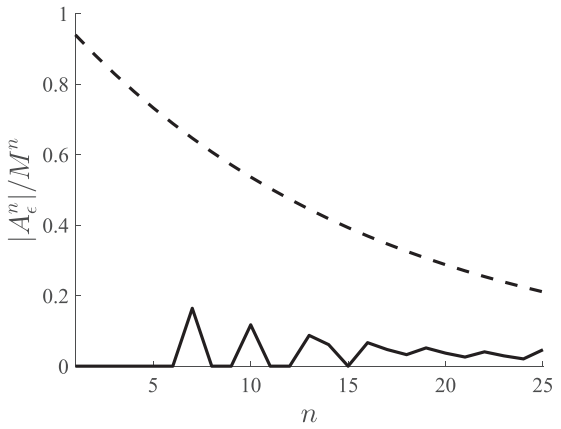

FIG. 1. Comparison of the exact fraction of sequences in the typical set, $|A_\epsilon^n|/M^n$, as a function of the sequence length, n, from enumeration (solid line) to the asymptotic upper bound, $|A_\epsilon^n|/M^n \sim e^{n(h_\mu+\epsilon-\ln M)}$ (dashed line) for a biased coin with $p = [0.7,0.3]$ and $\epsilon = 0.02$.

The choice of ϵ is arbitrary, so long as n is sufficiently large. In our calculations, to avoid the trivial solution of an empty typical set of finite length sequences, we choose ϵ so that $\mu\{A_\epsilon^n\} \neq 0$ for all n.

An asymptotic upper bound on the size of the typical set is [19]

$$\left|A_\epsilon^n\right| \leqslant e^{n(h_\mu+\epsilon)}. \tag{4}$$

In the infinite n limit, ϵ can be made arbitrarily small and $|A_\epsilon^n| \sim e^{nh_\mu}$. For $n \ll \infty$, the upper bound can be a poor approximation of the size of the typical set. Figure 1 shows the normalized size of the typical set, $|A_\epsilon^n|/M^n$ (solid line), for a biased coin with the probability of heads being 0.7 and 0.3 for tails. The normalized upper bound $\exp[n(h_\mu + \epsilon - \ln M)]$ is a monotonic function of n (dashed line) but $|A_\epsilon^n|$ is not: sequences enter and escape from the typical set. To account for the fluctuating size of the typical set, we next introduce a partition over the space of sequences.

III. EVOLUTION OF TYPICAL AND ATYPICAL SEQUENCES

The biased coin example highlights the need to predict the typical set for sequences that are longer than those accessible from direct enumeration and shorter than those well approximated by the asymptotic limit. Fluctuations in properties of the typical set arise from the interplay between the changing sample entropy $-n^{-1}\ln\mu(\hat{\omega}_n)$ and the bounds $h_\mu - \epsilon$ and $h_\mu + \epsilon$. A consequence of this interplay is that typical (atypical) sequences of length n can give rise to atypical (resp. typical) sequences at $n+1$. Since the number and probability of sequences entering and leaving the typical set is a function of n, we can represent changes in these quantities through transition probabilities. The transition probabilities are between groups of sequences we call "macrosequences." All sequences belonging to the same macrosequence have the same average behavior. The typical set is one macrosequence. For an alphabet of size M, sequences of length n can be subsequences to at most M sequences of length $n+1$. In this way, sequences

can be seen as transitioning from one macrosequence to another through their offspring. If a sequence transitions from typical to atypical, for example, the size of the typical set will decrease by $1/M^{n+1}$. The probability in the typical set will decrease by the joint probability of that sequence, which is not necessarily $1/M^{n+1}$. Consequently, both the number of sequences in each macrosequence and their corresponding probabilities evolve under two different dynamics.

We next make the relationship between $\hat{\omega}_n$, $\mu(\hat{\omega}_n)$ and the macrosequences more precise. Doing so will allow us to predict which sequences will be typical at n. We begin by defining a partition using the typical set over the space of all sequences of length n, Ω^n.

A. Macrosequence dynamics

For each n, a natural partition over the sequences uses A_ϵ^n and its complement, which represents all atypical sequences

$$C_\epsilon^n \equiv \left\{\Omega^n \backslash A_\epsilon^n\right\}. \tag{5}$$

We can further divide the complement C_ϵ^n into the lower complement, $C_l^n = \{\hat{\omega}_n : \mu(\hat{\omega}_n) < e^{-n(H+\epsilon)}\}$, and the upper complement, $C_u^n = \{\hat{\omega}_n\colon \mu(\hat{\omega}_n) > e^{-n(H-\epsilon)}\}$. The union of the three *macrosequences* cover Ω^n,

$$\begin{aligned}\Omega^n &= \bigcup_{\alpha=1}^{3} S_\alpha^n = C_l^n \bigcup A_\epsilon^n \bigcup C_u^n,\\ S_\alpha^n \cap S_\beta^n &= \emptyset \qquad \text{for} \qquad \alpha \neq \beta,\end{aligned} \tag{6}$$

where S_α^n represents an arbitrary macrosequence. Every sequence belonging to the same macrosequence has qualitatively the same average behavior. Each atypical macrosequence has a distinct average behavior, motivating the definition of two atypical macrosequences, C_l^n and C_u^n, instead of just one. For example, it is often the case that $\Pr[C_l^{n+1}|C_l^n] = 1$ and C_l acts as an absorbing state for relatively small n. In comparison, from the geometric structure to follow, the self-transition probability for the other atypical macrosequence is often $\Pr[C_u^{n+1}|C_u^n] = 1-\delta$ where $\delta \ll 1$, meaning this macrosequence will continually leak probability, even for large n. The macrosequences provide an alternate dynamics for the sequences. Every sequence can be generated, and the properties of the typical set calculated directly or, as we show here, the macrosequences can be evolved to compute the properties of A_ϵ^n (Fig. 2).

To describe how the number of sequences in each macrosequence evolves with n, we need to make the idea of transitions between macrosequences more precise. Every sequence is given by an ordered list of states, $\hat{\omega}_n = (\omega_1, \omega_2, \ldots, \omega_n)$. All sequences of length $n+1$ are created by appending ω_{n+1} to $\hat{\omega}_n$. Since the states are ordered, the sequence $\hat{\omega}_n$ will be a subsequence of at most M sequences of length $n+1$. We call this set of length $n+1$ sequences the "children" of $\hat{\omega}_n$:

$$\mathcal{C}(\hat{\omega}_n) = \{\hat{\omega}_{n+1}\colon \hat{\omega}_{n+1} = \hat{\omega}_n\omega_{n+1}\}. \tag{7}$$

In this nomenclature, sequences transition between the typical and atypical sets by producing "offspring." The second generation children are

$$\mathcal{C}^2(\hat{\omega}_n) = \{\hat{\omega}_{n+2}\colon \hat{\omega}_{n+2} = \hat{\omega}_{n+1}\omega_{n+2}, \hat{\omega}_{n+1} \in \mathcal{C}(\hat{\omega}_n)\}. \tag{8}$$

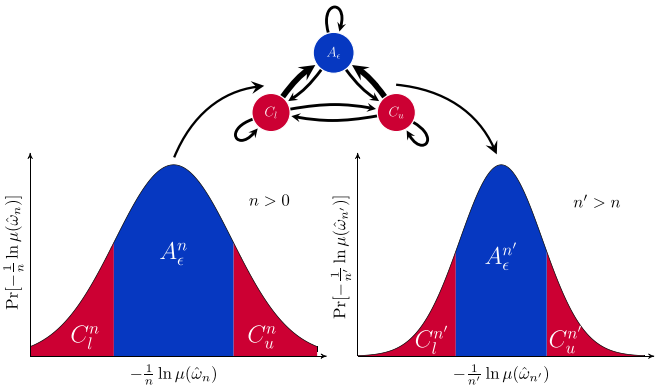

FIG. 2. With increasing n, the joint probability tends towards a uniform distribution and the sample entropy, $-n^{-1}\ln\mu(\hat{\omega}_n)$, distribution concentrates. The changes in distribution occur as individual sequences move between typical (blue) and atypical (red) macrosequences. The dynamics of these macrosequences are an alternative route to quantify the size and probability in the typical set.

Every sequence $\hat{\omega}_n$ has M children in the subsequent generation, M^2 children in the next generation, and so on.

The dynamics for the number of sequences in a macrosequence is given by the transition probability:

$$R_{\alpha\beta}(n) = \Pr[\mathcal{C}(\hat{\omega}_n) \in S_\alpha^{n+1} |\ \hat{\omega}_n \in S_\beta^n]. \tag{9}$$

The transition matrix is right stochastic, $\sum_\alpha R_{\alpha\beta}(n) = 1$. The probability a sequence occupies each macrosequence is

$$s_n(\alpha) = \frac{\left|\hat{\omega}_n \in S_\alpha^n\right|}{M^n} \tag{10}$$

such that $\sum_{\alpha=1}^{3} s_n(\alpha) = 1$. The quantity $|\hat{\omega}_n \in S_\alpha^n|$ is the number of sequences in the macrosequence S_α^n. The transition matrices can be used to evolve the occupation probabilities s_n from n to n',

$$s_{n'} = R(n'-1)R(n'-2)\cdots R(n+1)R(n)s_n, \tag{11}$$

where $n' > n$. Recall that the rate of growth of all sequences is given by the topological entropy, h_{top}. Using h_{top}, the size of each macrosequence is $|S_\alpha^n| = s_n(\alpha)e^{nh_{\text{top}}}$.

B. Probability of macrosequences

Unless $\mu(\hat{\omega}_n)$ is a uniform distribution, the total joint probability in a macrosequence is not equal to $s_n(\alpha)$. Just as we did in the last section, we need to find the overlap of $\mu(\hat{\omega}_n)$ and the three macrosequences. As the length of the sequences tends to infinity, the joint probability $\mu(\hat{\omega}_n) \sim e^{-nh_\mu}$ tends to zero due to conservation of probability. The number of sequences, however, grows exponentially $|A_\epsilon^n| \sim e^{nh_\mu}$. The entropy rate h_μ then uniquely determines the growth in the number of sequences and the decay of the individual sequence probability [27]. As n grows, these asymptotic results for the growth in the number of sequences and the decay of probability hides the more subtle dynamics between macrosequences. Thus, we scale the joint probability

$$\bar{\mu}(\hat{\omega}_n) = \mu(\hat{\omega}_n)p_{\max}^{-n} \tag{12}$$

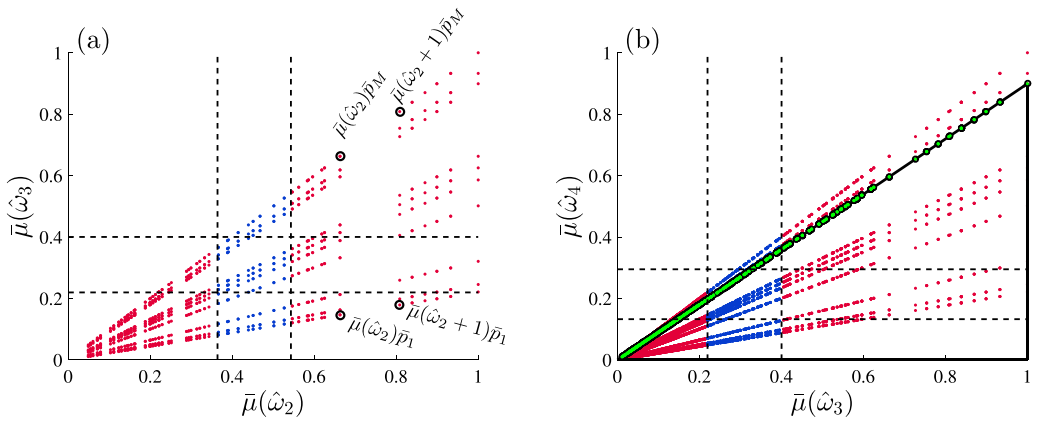

FIG. 3. Plotting $\mathcal{C}_p[\bar{\mu}(\hat{\omega}_n)]$ against $\bar{\mu}(\hat{\omega}_n)$ in (a) for a random distribution with $M = 10$ states. Every point represents the probability of one of the M^{n+1} children. Blue points denote children in A^n_ϵ and red denotes C^n_l and C^n_u. The vertical dashed lines mark the scaled bounds on A^n_ϵ, e^{-nI_l}, and e^{-nI_u}. Horizontal dashed lines are the scaled bounds for A^{n+1}_ϵ, $e^{-(n+1)I_l}$, and $e^{-(n+1)I_u}$. These bounds divide $\mathcal{C}_p[\bar{\mu}(\hat{\omega}_n)]_n$ into nine cells which determine the transition probabilities for both $R(n)$ and $Q(n)$. The left lower and upper two points highlighted with black circles are $\bar{\mu}(\hat{\omega}_2)\bar{p}_1$ and $\bar{\mu}(\hat{\omega}_2)\bar{p}_M$, respectively. The right-hand circles highlight $\bar{\mu}(\hat{\omega}_2+1)\bar{p}_1$ and $\bar{\mu}(\hat{\omega}_2+1)\bar{p}_M$, which become stretched vertically because $\bar{\mu}(\hat{\omega}_2) < \bar{\mu}(\hat{\omega}_2)$, thereby creating the triangle structure. (b) The hypotenuses of one of the M triangles intersects the probability of M^n children (highlighted with black circles).

to fix $\max[\bar{\mu}(\hat{\omega}_n)] = 1$, and we scale the marginal $\bar{p}_j = p_j/p_{\max}$ so that $\max[\bar{p}_j] = 1$. Scaled distributions are indicated by overbars. The scaled joint is evolved through the discrete map $\mathcal{C}_p : [0,1] \mapsto [0,1]$,

$$\mathcal{C}_p[\bar{\mu}(\hat{\omega}_n)] = \{\bar{\mu}(\hat{\omega}_n) \otimes \bar{p}\}, \tag{13}$$

where $\otimes$ is the standard Kronecker product, in this case, between two vectors. The total joint occupation probability of belonging to a macrosequence, S^n_α,

$$q_n(\alpha) = p^n_{\max} \sum_{\hat{\omega}_n \in S^n_\alpha} \bar{\mu}(\hat{\omega}_n), \tag{14}$$

is normalized so $\sum_{\alpha=1}^{3} q_n(\alpha) = 1$. Using the definition of children in this case, Eq. (13), the time-dependent transition probabilities are

$$Q_{\alpha\beta}(n) = \Pr\{\mathcal{C}_p[\bar{\mu}(\hat{\omega}_n)] \in S^{n+1}_\alpha \mid \bar{\mu}(\hat{\omega}_n) \in S^n_\beta\}. \tag{15}$$

These right-stochastic transition matrices evolve the marginal probability of each macrosequence S^n_α forward in n,

$$q_{n'} = Q(n'-1)Q(n'-2)\cdots Q(n+1)Q(n)q_n, \tag{16}$$

where $n' > n$. Together, the set of transition matrices $\mathcal{R} = \{R(n), R(n-1), \ldots, R(1)\}$ and the set of macrosequences, $\mathcal{S} = \{S^n_\alpha, S^{n-1}_\alpha, \ldots, S^1_\alpha\}$ describe how the number of both typical and atypical sequences change as a function of n. Likewise, $\mathcal{Q} = \{Q(n), Q(n-1), \ldots, Q(1)\}$ and $\mathcal{S}$ together describe how probability moves in and out of the macrosequences with n. These two ordered pairs, $(\mathcal{R},\mathcal{S})$ and $(\mathcal{Q},\mathcal{S})$, are what we wish to calculate for a given system.

C. Triangle map

A brute force approach to calculating $(\mathcal{R},\mathcal{S})$ and $(\mathcal{Q},\mathcal{S})$ is to explicitly generate all M^n sequences. To bypass a complete enumeration, we introduce a geometric picture of $\mu(\hat{\omega}_n)$ and the resulting children for a simplified description of q_n and s_n.

Every sequence has M children. The probability of each child is iterated through $\mathcal{C}_p$, Eq. (13). Because we order the joint probabilities such that $\bar{\mu}(\hat{\omega}_n) \leqslant \bar{\mu}(\hat{\omega}_n + 1)$, plotting $\mathcal{C}_p[\bar{\mu}(\hat{\omega}_n)]$ against $\bar{\mu}(\hat{\omega}_n)$ gives a picture like that shown in Figs. 3(a)–3(b). Most striking from this picture of the joint probabilities is the triangular form of the forward mapping in the case of i.i.d. random variables.

Up to M lines can be drawn from the origin $(0,0)$ to $(1,\bar{p}_j)$. Each line will intersect the rescaled joint probability of M^n sequences (see the Appendix), and the children can be thought of as lying on the hypotenuse of a triangle [46]. One triangle is shown in Fig. 3(b) in black with the intersecting children highlighted in green. This geometric picture, where the probability of sequences lies on the hypotenuses of triangles, will help us to calculate the number, $s_n(\alpha)$, and probability, $q_n(\alpha)$, of sequences in each macrosequence. We will call the mapping $\mathcal{C}_p[\bar{\mu}(\omega_n)]$ the *triangle map*.

The upper and lower bounds of the typical set at n and $n+1$ over $\mathcal{C}_p[\bar{\mu}(\hat{\omega}_n)]$ divides the space $\bar{\mu}(\omega_{n+1}) \times \bar{\mu}(\omega_n)$ into nine regions (dashed lines in Fig. 3). These regions can be used to define a dynamics for entrance into and escape from the macrosequences, including the typical set. The number of points in each cell defines the transition probability for $R(n)$ between any two macrosequences. The total probability in each cell defines the transition probability for $Q(n)$ between any two macrosequences, S^n_β to S^{n+1}_α. Vertical dashed lines in Fig. 3 are given by the scaled bounds of the typical set A^n_ϵ: the lower bound e^{-nI_l}, where $I_l = H + \epsilon + \ln p_{\max}$, and the upper bound e^{-nIu}, where $I_u = H - \epsilon + \ln p_{\max}$. The horizontal dashed lines are the scaled bounds at $n+1$: $e^{-(n+1)I_l}$ and $e^{-(n+1)I_u}$.

Mapping consecutive joint distributions to a triangle give a geometric representation of the macrosequence dynamics. These dynamics are what we wish to predict. One difficulty is that points along any hypotenuse are not uniform. Accounting for this distribution of points is qualitatively explained in the next section. Additional details of our derivation and the construction of $(\mathcal{R},\mathcal{S})$ and $(\mathcal{Q},\mathcal{S})$ are in the Appendix.

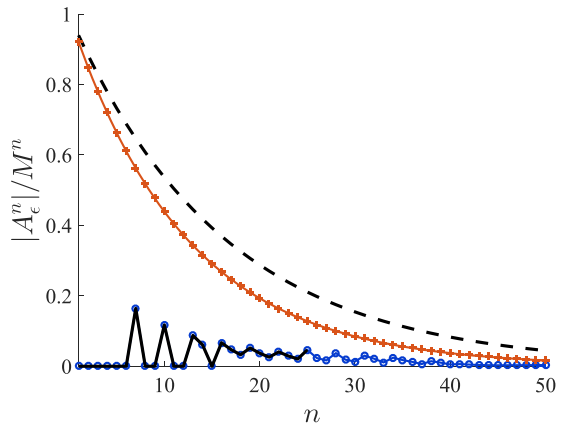

FIG. 4. The dashed black line is the upper bound $e^{n(H+\epsilon-\ln M)}$ and crosses are the $\epsilon = 0$ bound $e^{n(H-\ln M)}$ that is guaranteed only for $n \to \infty$. The fractional size of $|A_\epsilon^n|/M^n$ from enumeration (solid black line) and the calculation from our framework (blue line). The parameter ϵ is 0.02 in all cases. The method here matches the brute force calculation of sequences exactly up to where enumeration is tractable, $n \approx 25$, and agrees with the asymptotic limit near $n \approx 50$.

IV. EXAMPLES

Now we apply the framework to examples that will illustrate how the formalism generates the exact size and probability of the macrosequences over n. These examples also serve to show how other observables can be calculated from the formalism.

Let us briefly summarize the procedure: Given p_j and the entropy rate $h_\mu = H$, the triangle map $\mathcal{C}_p[\bar{\mu}(\hat{\omega}_1)]$ can be calculated. The upper and lower bounds (e^{-I_l}, e^{-I_u}) and (e^{-2I_l}, e^{-2I_u}) follow and divide $\mathcal{C}_p[\bar{\mu}(\hat{\omega}_1)]$ into nine cells defining the transition probabilities, Eq. (9) and Eq. (15). In the Appendix we show that at most $(2M+3)n$ values of the triangle map must be known to construct the transition probabilities. From the structure of the triangle map and the known distribution p_j, the total probabilities of each macrosequence, q_n and s_n, can, in principle, be calculated exactly to any desired n.

A. Bernoulli process and the redundancy

The Bernoulli process is a benchmark for the typical set that can be enumerated completely such that, for sufficiently large n, the asymptotic bounds begin to converge. This fact allows us to test the framework. Figure 4 again shows results for a biased coin where the probability of heads is 0.7 and tails is 0.3, $p = [0.7, 0.3]$. The joint probability of a sequence is $\mu(\hat{\omega}_n) = \prod_{i=1}^n p(\omega_i)$. The asymptotic bounds for the size of the typical set can be quite poor for $n \ll \infty$. As n becomes larger, though, both the bound and our measure of $|A_\epsilon^n|$ converge to the asymptotic limit (dashed line with crosses). In the limit $\mu(\hat{\omega}_n) \simeq e^{-nh_\mu}$, so the upper bound [dashed line representing Eq. (4)] will have an error of at least $e^{n\epsilon}$. In this example, we use $(2M+3)n = 300$ points on $\mathcal{C}_p[\bar{\mu}(\hat{\omega}_n)]$ to calculate $|A_\epsilon^n|$ for $n = 50$. In contrast, $2^{50} \approx 10^{14}$ sequences would have to be enumerated by brute force calculation.

While the asymptotic upper bound $|A_\epsilon^n| \leqslant e^{n(h_\mu+\epsilon)}$ is poor for $n \ll \infty$, it does motivate an important observable—the redundancy [28]:

$$r = \ln M - h_\mu. \tag{17}$$

The redundancy measures the information carrying capacity of the alphabet. If $r \neq 0$, there are correlations in the sequences [45]. Loosely speaking, the redundancy measures how closely a process is to maximizing the information rate over the alphabet. In general, for finite i.i.d. sequences, the maximum of the entropy h_μ is $\ln M$ [19]. The redundancy for finite n can then be defined as

$$r_n = \ln M - \frac{1}{n} \ln \left|A_\epsilon^n\right|. \tag{18}$$

The quantity r_n is a measure of the information per symbol used by the sequences in the typical set of length n. In the limit $n \to \infty$, $|A_\epsilon^n| \approx e^{nh_\mu}$ and ϵ can be set arbitrarily close to zero, meaning $r_n \to r$. In the next section, the Schlögl model illustrates how observables, such as the redundancy, can be calculated with this framework beyond where enumeration is tractable.

B. Schlögl model

Biological and chemical systems often obey intricate relationships, across many spatial and time scales, making these systems good candidates for the application of the typical set at finite n. Schlögl's second model [47] is a well-studied set of chemical reactions [48,49] defined by

$$\mathrm{A} + 2\mathrm{X} \rightleftharpoons 3\mathrm{X}, \tag{19}$$

$$\mathrm{X} \rightleftharpoons \mathrm{B}. \tag{20}$$

The second equation is modified [50] from Schlögl's initial work. The intermediate species X is commonly the one of interest when the reactant A and product B have fixed concentrations, a and b. Applying the law of mass action, the kinetic equation for the concentration, x, of X is an ordinary differential equation:

$$\frac{dx}{dt} = k_1 a x^2 - k_2 x^3 - k_3 x + k_4 b. \tag{21}$$

Setting this equation equal to zero gives the steady-state solutions. The number of real, steady-state solutions comes from the discriminant:

$$4k_1^3 a^3 k_3 b - k_1^2 a^2 k_4^2 + 4k_2 k_4^3 - 18 k_2 k_1 a k_4 k_3 b + 27 k_2^2 k_3^2 b^2. \tag{22}$$

Fixing the rate constant parameters, $k_i, i = 1,2,\ldots,4$, but varying the concentrations a and b, changes the number of real solutions (Fig. 5). Here, we will look at the bistable region (green), where two stable steady states are separated by a single unstable steady state [51]. Though the bistable region will be our focus, the method is applicable to any region. For input into the method, we construct a marginal distribution for each fixed point from the concentrations:

$$p(s) = \frac{s}{a+b+x}, \quad \text{where} \quad s = \{a, b, x\}. \tag{23}$$

We use concentrations of X that correspond to a particular zero of Eq. (21). From this marginal probability distribution,

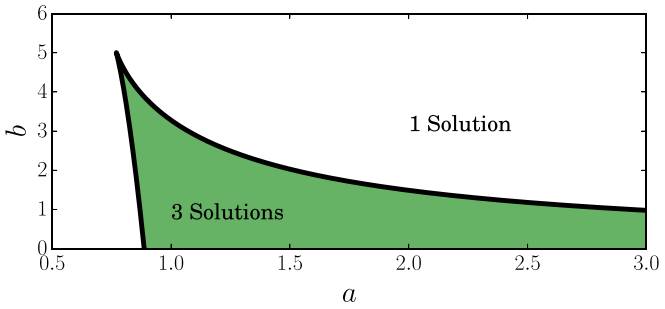

FIG. 5. Holding k_1 through k_4 fixed and varying a and b, Eq. (21) generates one, two, or three real fixed points. There are two solutions at each point on the black line that divides the regions with one and three steady-state solutions.

we construct the macrosequence dynamics of $q_n(\alpha)$ and $|S_\alpha^n|$ with $\epsilon = 0.1$, as shown in Fig. 6. As an example, we take one of the stable fixed points x, and the parameter values of $k_1 = 3, k_2 = 0.6, k_3 = 0.25, k_4 = 2.95, a = 1$, and $b = 1$. Figures 6(a)–6(b) show that we reproduce the exact values for S_α^n through enumeration (black) from the triangle-map construction of the dynamics (color). The dashed lines with circles are the typical set, A_ϵ^n, the lines with open circles are the macrosequence C_l^n, and the lines with stars are the macrosequence C_u^n. Figure 6(a) shows that $|A_\epsilon^n|$ is small, meaning the redundancy is near $\ln M$. This steady state then has large correlations between the states in a sequence. It should be noted that the second stable fixed point (not shown) exhibits qualitatively the same results.

The unstable fixed point for the same parameter values gives a different picture [Figs. 6(c) and 6(d)]. Now $|A_\epsilon^n|M^{-n} \approx \Pr[A_\epsilon^n] \approx 1$ for $n > 3$, almost all sequences are typical, and $r \approx 0$. The joint distribution for the unstable fixed point is almost uniform, and there is a lack of correlations in the sequences. While both examples are fixed points of the steady-state solution, they illustrate that the information content of their sequences is quite different.

V. CONCLUSIONS

The probability and size of the typical set are of fundamental importance to statistical mechanics and information theory. However, away from asymptotic limits, the tractable calculation of the typical set is limited by the exponential growth in the sequence space. Here we have shown that the dynamics of macrosequences circumvents this exponential growth and avoids both enumeration and asymptotic limits. For independent and identically distributed random variables, these dynamics, and therefore the future properties of the typical set, are entirely determined by a single marginal distribution. We found that the number of points needed to quantify the macrosequences grows linearly in both the number of states and the length of sequences as $\leqslant (2M + 3)n$. As a consequence, this method could be applied to systems with a larger state space, or to longer sequence lengths, than the proof-of-principle examples shown here. The method is computationally efficient, applies to the entire class of i.i.d. systems, and enables the calculation of information-theoretic observables, such as the redundancy, for finite length sequences without asymptotic approximations.

ACKNOWLEDGMENTS

This material is based upon work supported by the U.S. Army Research Laboratory and the U.S. Army Research Office under Grant No. W911NF-14-1-0359.

APPENDIX

There are three ingredients in the exact construction of the transition probabilities $R(n)$ and $Q(n)$ that avoid the need to evaluate all M^n sequences. First, all children can be thought of as lying on the hypotenuse of as many as M triangles. Using this continuous geometric representation as a set of M triangles compactly describes all children. Second, we show that at most $2M + 3$ values of the triangle map must be known to calculate each transition matrix, $R(n)$ and $Q(n)$, for a given n. To describe the probability and size of macrosequences up to length n, $(2M + 3)n$ values of the triangle map are necessary. Third, we use two cumulative density functions (CDFs) to find these $2M + 3$ points. We derive exact formulas for the CDFs at n in terms of the marginal distribution, p.

1. Children of each sequence intersect similar triangles

We now prove that each child $\mathcal{C}_p[\bar{\mu}(\hat{\omega}_n)]$ falls on the hypotenuse of a triangle, one of (at most) M triangles, which will be useful later. To prove this, we use the scaled variables [Eq. (12)] and the fact that the hypotenuse of each triangle is given by

$$l_j = \bar{p}_j x, \tag{A1}$$

when the continuous variable $x \in [0,1]$ and the index $j = 1,2,\ldots,M$. Note, later we will also refer to the continuous variable $y \in [0,1]$, which will always belong to the y axis of $\mathcal{C}_p[\bar{\mu}(\hat{\omega}_n)]$. Define the triangle, $\triangle_j$, through the points $(0,0), (1,0)$, and $(1,\bar{p}_j)$. The length of the hypotenuse for $\triangle_j$ is $r_j = \sqrt{1+\bar{p}_j^2}$, and the angles are $\theta_j = \cos^{-1}[(1+\bar{p}_j^2)^{-1/2}]$, 90, and $180 - 90 - \theta_j$.

Define a second triangle formed from $\bar{\mu}(\hat{\omega}_n)$ and one child from $\mathcal{C}_p[\bar{\mu}(\hat{\omega}_n)]$ as $\triangle'$ with points $(0,0)$, $(\bar{\mu}(\hat{\omega}_n),0)$, and $(\bar{\mu}(\hat{\omega}_n),\bar{\mu}(\hat{\omega}_n)\bar{p}_j)$. The length of the hypotenuse for $\triangle'$ is $r' = \sqrt{\bar{\mu}(\hat{\omega}_n)^2 + [\bar{\mu}(\hat{\omega}_n)\bar{p}_j]^2}$. Meaning that the angles of $\triangle'$ are, θ', 90, and $180 - 90 - \theta'$ where

$$\theta' = \cos^{-1}\left[\frac{\bar{\mu}(\hat{\omega}_n)}{\sqrt{\bar{\mu}(\hat{\omega}_n)^2 + (\bar{\mu}(\hat{\omega}_n)\bar{p}_j)^2}}\right] = \theta_j. \tag{A2}$$

Therefore, $\triangle'$ and $\triangle_j$ are similar, meaning at least one child intersects the hypotenuse, $\mathcal{C}_p[\mu(\hat{\omega}_n)] \in l_j$. The consequence of forming similar triangles is that, since $\mathcal{C}_p[\bar{\mu}(\hat{\omega}_n)] = \bar{p}_j\bar{\mu}(\hat{\omega}_n)$, for each $\bar{p}_j$ there are at least M^n points intersecting one line l_j. And, together, the set of lines $\{l_j\}$ must intersect all M^{n+1} points on $\mathcal{C}_p[\bar{\mu}(\hat{\omega}_n)]$.

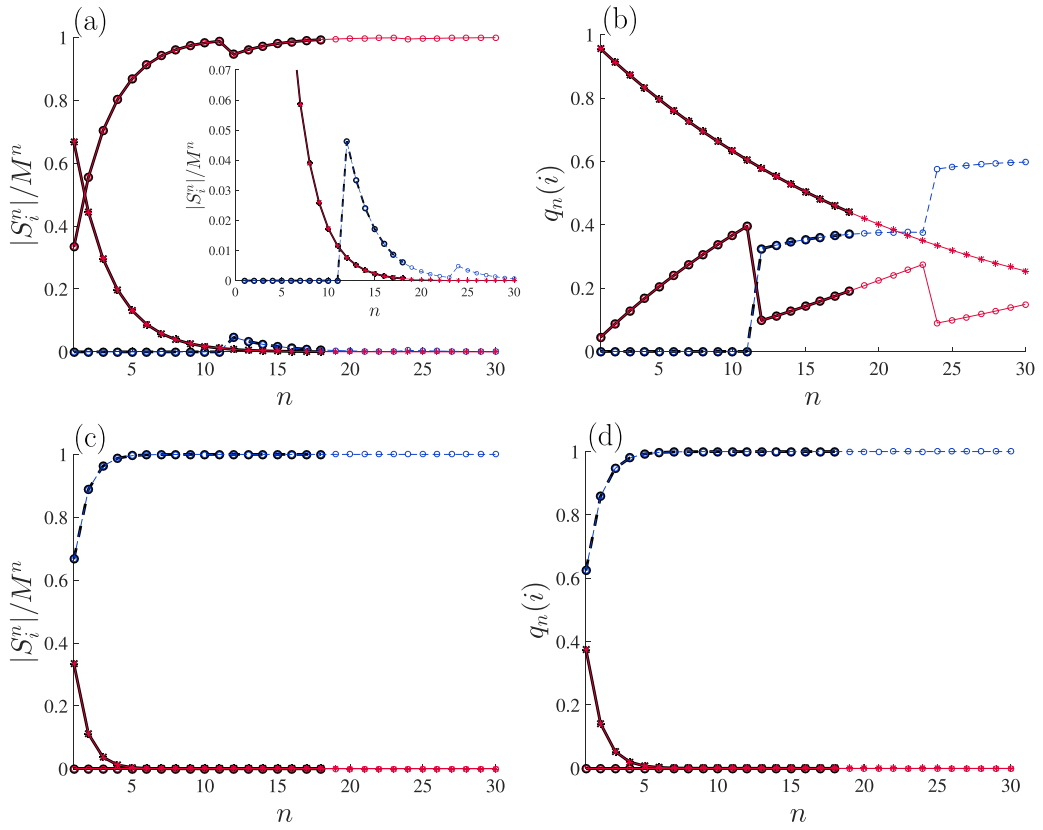

FIG. 6. The (a) size of and (b) the amount of probability in S_α^n for a stable fixed point of the Schlögl model (black lines) when $\epsilon = 0.1$: C_l^n (open circles), C_u^n (stars), and A_ϵ^n (black dashed line). Colored symbols are the results from the method here. The (c) size and (d) the probability of S_α^n for the unstable fixed point at the same parameter values. There is a second stable fixed point that exhibits qualitatively the same high redundancy behavior as the stable fixed point (data not shown).

2. Exact construction of macrosequence dynamics

Now we derive the exact construction of the transition probabilities $(R(n), Q(n))$ and show that at most $(2M+3)n$ values of the triangle map can describe the size and probability of the macrosequences for any n, as opposed to the possible M^n sequences normally needed in a brute force approach.

As described in the main text, the boundaries of the typical set at n and $n+1$, $\{e^{-nI_l}, e^{-nI_u}, e^{-(n+1)I_l}, e^{-(n+1)I_u}\}$, divide the triangle map into nine cells (Fig. 7). Each hypotenuse can cross a typical set boundary only once. To count how many sequences or how much probability is in each cell, we need the points where l_j enters and exits each boundary. We use the location of the intersections mapped to the x axis, x_j^o and x_j^f. These intersection points of l_j are given by the logical rules in Table I. For example, the contribution l_j makes to the transition probability, $\Pr[C_l^{n+1}|A_\epsilon^n]$, is determined by where l_j crosses the boundaries at the two points:

$$x_j^f = \min[\bar{p}_j e^{-nI_u}, e^{-(n+1)I_l}]\bar{p}_j^{-1} = l_j(x)\bar{p}_j^{-1},$$
$$x_j^o = e^{-nI_l} = l_j(x). \tag{A3}$$

Cells of the triangle map, marked by bounds of the typical set, define the transition probabilities between the macrosequences S_β^n and S_α^{n+1}. To calculate $R_{\alpha\beta} = \Pr[S_\alpha^{n+1}|S_\beta^n]$ the number of sequences in each cell must be counted. The goal is to find $R(n)$ given the scaled distributions for each state $\bar{p}_j$ and each sequence $\bar{\mu}(\hat{\omega}_n)$. For the joint distribution, there is a cumulative distribution function (CDF) $\rho_n(x)$, $x \in [0,\infty)$ given by

$$\rho_n(x) = \begin{cases} M^{-n} \sum_{\hat{\omega}_n} \int_0^x \delta[\bar{\mu}(\hat{\omega}_n) - s]\, ds & \text{if } x \leqslant 1 \\ 1 & \text{if } x > 1 \end{cases}. \tag{A4}$$

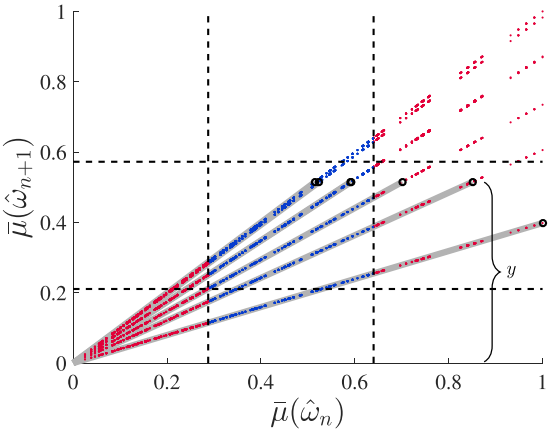

FIG. 7. The cumulative density $\rho_{n+1}(y)$ is proportional to all the sequences which lie below a particular y. Since each sequence lies on the function l_j, the point on $l_j(y)$ can be mapped to $\rho_n(x)$. If $y > \bar{p}_j$, as is shown with the rightmost gray line covering l_1, then the contribution from ρ_n at this point is one.

TABLE I. Rules determining the two points x_j^f and x_j^o, which determine the contribution each l_j makes to the transition probabilities $R(n)$ and $Q(n)$.

$S_y^n \to S_x^{n+1}$	x_j^f	x_j^o
$C_l^n \to C_l^{n+1}$	$\min[\bar{p}_j e^{-nI_l}, e^{-(n+1)I_l}]\bar{p}_j^{-1}$	0
$C_l^n \to A_\epsilon^{n+1}$	$\min[\bar{p}_j e^{-nI_l}, e^{-(n+1)I_u}]\bar{p}_j^{-1}$	$e^{-(n+1)I_l}\bar{p}_j^{-1}$
$C_l^n \to C_u^{n+1}$	e^{-nI_l}	$e^{-(n+1)I_u}\bar{p}_j^{-1}$
$A_\epsilon^n \to C_l^{n+1}$	$\min[\bar{p}_j e^{-nI_u}, e^{-(n+1)I_l}]\bar{p}_j^{-1}$	e^{-nI_l}
$A_\epsilon^n \to A_\epsilon^{n+1}$	$\min[\bar{p}_j e^{-nI_u}, e^{-(n+1)I_u}]\bar{p}_j^{-1}$	$\max[\bar{p}_j e^{-nI_l}, e^{-(n+1)I_l}]\bar{p}_j^{-1}$
$A_\epsilon^n \to C_u^{n+1}$	e^{-nI_u}	$\max[\bar{p}_j e^{-nI_l}, e^{-(n+1)I_u}]\bar{p}_j^{-1}$
$C_u^n \to C_l^{n+1}$	$\min[\bar{p}_j, e^{-(n+1)I_l}]\bar{p}_j^{-1}$	e^{-nI_u}
$C_u^n \to A_\epsilon^{n+1}$	$\min[\bar{p}_j, e^{-(n+1)I_u}]\bar{p}_j^{-1}$	$\max[\bar{p}_j e^{-nI_u}, e^{-(n+1)I_l}]\bar{p}_j^{-1}$
$C_u^n \to C_u^{n+1}$	1	$\max[\bar{p}_j e^{-nI_u}, e^{-(n+1)I_u}]\bar{p}_j^{-1}$

The analytic expressions for the intersection points correspond to locations on the CDF $\rho_n(x)$. The transition probability for the sequence dynamics is then given by the contribution from each line l_j that enters the same transition cell:

$$R_{\alpha\beta}(n) = \frac{1}{\mathcal{Z}_\alpha^R}\sum_j \left[\rho_n\left(x_j^f\right) - \rho_n\left(x_j^o\right)\right]. \tag{A5}$$

The normalization factor $\mathcal{Z}_\alpha^R$ ensures $R(n)$ is right stochastic, $\sum_\alpha R_{\alpha\beta}(n) = 1$.

The length of the line segment l_j in a particular cell corresponds to a certain amount of cumulative probability,

$$\varrho_n(x) = \begin{cases} \sum\limits_{\hat{\omega}_n | \mu(\hat{\omega}_n) \leqslant x} \mu(\hat{\omega}_n) & \text{if } x \leqslant 1 \\ 1 & \text{if } x > 1 \end{cases}, \tag{A6}$$

or $\Pr[l_j(x_j^f) - l_j(x_j^o)] = p_j[\varrho_n(x_j^f) - \varrho_n(x_j^o)]$, where l_j is written in terms of the unbarred distribution $l_j = p_j x$, and the transition probabilities are built from this CDF:

$$Q_{\alpha\beta}(n) = \frac{1}{\mathcal{Z}_\alpha^Q}\sum_j p_j\left[\varrho_n\left(x_j^f\right) - \varrho_n\left(x_j^o\right)\right]. \tag{A7}$$

Now we show that the number of points needed to construct the transition matrices grows linearly in M and n. From Table I, the rule for calculating $C_u^n \to C_l^{n+1}$ $(\min[\bar{p}_j, e^{-(n+1)I_l}]\bar{p}_j^{-1})$ means that in addition to the boundaries of the typical set at n and $n+1$, we also need to consider the end points of each line as a boundary. Including the end points with (at most) the four other boundaries l_j can cross, we need to evaluate ρ_n and ϱ_n at the set of points $\mathcal{I} = \{e^{-nI_l}, e^{-nI_u}, 1, e^{-(n+1)I_l}/\bar{p}_j, e^{-(n+1)I_u}/\bar{p}_j\}$. Since $R(n)$ and $Q(n)$ are determined by the set of intersections $\mathcal{I}$, and the index j runs from 1 to M, at most $(2M+3)n$ points are required to determine the macrosequence dynamics up to n.

3. Calculating ρ_n and ϱ_n from p_j

In the last section, we showed that the macrosequence transition probabilities at any n can be calculated from the distributions (p_j, ρ_n, ϱ_n). Now we derive a formula for the CDFs at n in terms of the marginal and CDFs at $n = 1$.

Let us start with an important property of ρ_{n+1}. Assuming p_j and ρ_1 are known, the CDF at $n+1$, $\rho_{n+1}(y)$, is proportional to the number of children lying below the point y on $\mathcal{C}_p[\bar{\mu}(\hat{\omega}_n)]$. Figure 7 illustrates this idea for a given y value. The sequences contributing to the CDF at $n+1$ are highlighted in gray. Summing the number of points with $l_j \leqslant y$ (black circles) gives the CDF at $n+1$. Each point below y has a corresponding value of x on the $\bar{\mu}(\hat{\omega}_n)$ axis, $y/\bar{p}_j = x$. Then, for a given y value, the CDF $\rho_{n+1}(y)$ is given in terms of the previous CDF $\rho_n(x)$:

$$\rho_{n+1}(y) = \frac{1}{M}\sum_{j=1}^{M}\rho_n\left(\frac{y}{\bar{p}_j}\right). \tag{A8}$$

We note, from the definition of the CDF, Eq. (A4), if $y/\bar{p}_j > \bar{p}_j$, then $\rho_n(y/\bar{p}_j) = 1$. Letting $y' = y/\bar{p}_j$, $\rho_n(y')$ can be found the same way using $\mathcal{C}_p[\bar{\mu}(\hat{\omega}_{n-1})]$. Substituting ρ_{n-1} into ρ_n gives

$$\rho_{n+1}(y) = \frac{1}{M^2}\sum_{j=1}^{M}\sum_{k=1}^{M}\rho_{n-1}\left(\frac{y}{\bar{p}_j\bar{p}_k}\right). \tag{A9}$$

Repeating until ρ_1 gives

$$\rho_{n+1}(y) = \frac{1}{M^n}\sum_{k_1,k_2,\ldots,k_n}\rho_1\left(\frac{y}{\bar{p}_{k_1}\bar{p}_{k_2},\ldots,\bar{p}_{k_n}}\right). \tag{A10}$$

This expression gives the exact value of $\rho_n(y)$ from ρ_1 and $\bar{p}_j$. To find $R(n)$ and $Q(n)$, the points where l_j maps to the elements of $\mathcal{I}$ are needed. There are at most $2M+3$ points. Unfortunately, this CDF has M^n entries in the summation for each value of y. So while it is a way to exactly calculate the values needed for the transition probabilities, it is not practical when n is large. However, Eq. (A10) shows that $\mathcal{C}_p[\bar{\mu}(\hat{\omega}_n)]$ and ρ_n are a potential way to describe the macrosequence dynamics for i.i.d. r.v.s when ρ_1 and $\bar{p}_j$ are known.

To transform Eq. (A10) into a more tractable form, we will use the fact that the random variables are i.i.d. This property means some positions in the sum are repeated, such as $y/\bar{p}_{k_1}\bar{p}_{j_2} = y/\bar{p}_{j_1}\bar{p}_{k_2}$. Counting the number of times $\bar{p}_j$

appears, s_j, leads to a simplified form of the CDF:

$$\rho_{n+1}(y) = \frac{n!}{M^n} \sum_{k=1}^{\binom{M+n-1}{n}} \frac{\rho_1(\frac{y}{\bar{r}_k})}{N_k}. \tag{A11}$$

The sum runs over all multisets, i.e., combinations where order is ignored, of the number of times $\bar{p}_j$ appears in the constraint $\sum_j s_j = n$. The denominator $\bar{r}_k$ shifts the position where ρ_1 is evaluated, $\bar{r}_k = \prod_{j=1}^{M} \bar{p}_j^{s_j}$, and $N_k = \prod_{j=1}^{M} s_j!$. Since the number of multisets for a given M and n grows far slower than M^n, Eq. (A11) offers substantial computational savings over enumerating all possible sequences.

Now we turn to ϱ_n. We find ϱ_n in terms of ϱ_1 with an argument similar to that for ρ_n and ρ_1 above. Only now, the CDF $\varrho_n(y/\bar{p}_j)$ must be multiplied by p_j. Again, through $\mathcal{C}_p[\bar{\mu}(\hat{\omega}_n)]$, $\varrho_{n+1}(y)$ can be written in terms of $\varrho_n(y/\bar{p}_j)$:

$$\varrho_{n+1}(y) = \sum_j p_j \varrho_n\left(\frac{y}{\bar{p}_j}\right). \tag{A12}$$

Writing $y' = y/\bar{p}_j$ gives

$$\begin{aligned}\varrho_{n+1}(y) &= \sum_j p_j \sum_l p_l \varrho_{n-1}\left(\frac{y'}{\bar{p}_l}\right) \\ &= \sum_{j,l} p_j p_l \varrho_{n-1}\left(\frac{y}{\bar{p}_j \bar{p}_l}\right).\end{aligned} \tag{A13}$$

Continuing to ϱ_1, and again using the fact that the random variables are i.i.d., gives

$$\varrho_{n+1}(y) = n! \sum_{k=1}^{\binom{M+n-1}{n}} \frac{r_k}{N_k} \varrho_1\left(\frac{y}{\bar{r}_k}\right). \tag{A14}$$

Eq. (A14) differs from Eq. (A11) only in that we need ϱ_1 instead of ρ_1 and we have the probability $r_k = \prod_{j=1}^{M} p_j^{s_j}$ associated with each entry of ϱ_1.

[1] J. L. Lebowitz, Physica A (Amsterdam) **194**, 1 (1993).
[2] H. Touchette, Phys. Rep. **478**, 1 (2009).
[3] R. S. Ellis, Ann. Prob. **12**, 1 (1984).
[4] J. M. Meylahn, S. Sabhapandit, and H. Touchette, Phys. Rev. E **92**, 062148 (2015).
[5] G. E. Uhlenbeck and G. W. Ford, *Lectures in Statistical Mechanics* (American Mathematical Society, Providence, RI, 1963), Vol. 1.
[6] T. L. Hill, *An Introduction to Statistical Thermodynamics* (Dover, New York, 1986), Vol. 2.
[7] E. A. Jackson, *Equilibrium Statistical Mechanics* (Prentice-Hall, New York, 1968), Vol. 1.
[8] G. E. Crooks, Phys. Rev. E **60**, 2721 (1999).
[9] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[10] L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, and C. Landim, Phys. Rev. Lett. **87**, 040601 (2001).
[11] E. T. Jaynes, Phys. Rev. **106**, 620 (1957).
[12] R. Dewar, J. Phys. A **36**, 631 (2003).
[13] B. Yurke, A. Turberfield, A. P. Mills, F. C. Simmel, and J. L. Neumann, Nature (London) **406**, 605 (2000).
[14] R. Golestanian, T. B. Liverpool, and A. Ajdari, Phys. Rev. Lett. **94**, 220801 (2005).
[15] M. L. Dekhtyar and V. M. Rozenbaum, J. Chem. Phys. **134**, 044136 (2011).
[16] C. Bustamante, J. C. Macosko, and J. L. Wuite, Nat. Rev. **1**, 130 (2000).
[17] M. Manosas and F. Ritort, Biophys. J. **88**, 3224 (2005).
[18] S. Wennmalm, L. Edman, and R. Rigler, Proc. Natl. Acad. Sci. U. S. A. **94**, 10641 (1997).
[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 2006), Vol. 2.
[20] S. Nicholson, M. Alaghemandi, and J. R. Green, J. Chem. Phys. **145**, 084112 (2016).
[21] M. Alaghemandi and J. R. Green, Phys. Chem. Chem. Phys. **18**, 2810 (2016).
[22] R. L. Davidchack, Y. C. Lai, E. M. Bollt, and M. Dhamala, Phys. Rev. E **61**, 1353 (2000).
[23] A. N. Kolmogorov, Dokl. Akad. Nauk SSSR **124**, 754 (1959).
[24] Ya. G. Sinai, Dokl. Akad. Nauk SSSR **124**, 768 (1959).
[25] P. Gaspard, J. Stat. Phys. **117**, 599 (2004).
[26] P. Zegers, A. Fuentes, and C. Alarcón, Entropy **15**, 2861 (2013).
[27] J. Crutchfield and C. Aghamohammadi, arXiv:1609.02519 (2016).
[28] C. E. Shannon, Bell Syst. Tech. J. **27**, 623 (1948).
[29] B. McMillan, Ann. Math. Stat. **24**, 196 (1953).
[30] K. L. Chung, Ann. Math. Stat. **32**, 612 (1961).
[31] P. H. Algoet and T. M. Cover, Ann. Prob. **16**, 899 (1988).
[32] V. Baccetti and M. Visser, J. Stat. Mech.: Theory Exp. (2013) P04010.
[33] R. W. Yeung, *A First Course in Information Theory*, vol. 1 (Springer, 2002).
[34] H. Chernoff, Ann. Prob. **9**, 533 (1981).
[35] K.-M. Chung, H. Lam, Z. Liu, and M. Mitzenmacher, in *29th International Symposium on Theoretical Aspects of Computer Science (STACS 2012)*, Leibniz International Proceedings in Informatics (LIPIcs), edited by C. Dürr and T. Wilke, Vol. 14 (Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2012), pp. 124–135.
[36] B. P. Rao and M. Sreehari, J. Stat. Planning Inference **63**, 325 (1997).
[37] A. D. Healy, Comput. Complex. **17**, 3 (2008).
[38] R. L. Adler, A. G. Konheim, and M. H. McAndrew, Trans. Am. Math. Soc. **114**, 309 (1965).
[39] R. Bowen, Trans. Am. Math. Soc. **153**, 401 (1971).
[40] C. Beck and F. Schlögl, *Thermodynamics of Chaotic Systems* (Cambridge University Press, Cambridge, 1993).
[41] C. Conselice, A. Wilkinson, K. Duncan, and A. Mortlock, Astrophys. J. **830**, 83 (2016).
[42] S. B. Nicholson, M. Alaghemandi, and J. R. Green, J. Chem. Phys. **148**, 044102 (2018).
[43] J. P. Crutchfield and K. Young, Phys. Rev. Lett. **63**, 105 (1989).
[44] C. R. Shalizi and J. P. Crutchfield, J. Stat. Phys. **104**, 817 (2001).
[45] J. P. Crutchfield and D. P. Feldman, Chaos **13**, 25 (2003).
[46] In general, all children will lie on the hypotenuse of a triangle. An exception is the uniform distribution. Since all sequences are equally likely, every sequence is typical independent of both ϵ and n. In this case, the triangle map would be a single point.
[47] F. Schlögl, Z. Phys. A **253**, 147 (1972).

[48] A. B. Goryachev and A. V. Pokhilko, FEBS Lett. **582**, 1437 (2008).
[49] R. G. Endres, PloS ONE **10**, e0121681 (2015).
[50] M. Vellela and H. Qian, J. R. Soc. Interface **6**, 925 (2009).
[51] J. Tyson, R. Albert, A. Goldbeter, P. Ruoff, and J. Sible, J. R. Soc. Interface **5**, S1 (2008).