Annex A
Mathematical Framework

This annex is a summary of the mathematical background needed for this book.

A.1 Dynamical Systems

In this book we only consider two kinds of "discrete-time" dynamical systems: continuous and measure-preserving systems. Roughly speaking, the first are the basic objects of topological dynamics and the second ones play a major role in the study of statistical properties.

Definition 7 A continuous (or topological) dynamical system is a pair (M,f), where M is a topological space and f:$M \to M$ a continuous map.

Let Ω be a non-empty set, $\mathcal{B}$ a sigma-algebra of subsets of Ω, and μ:$\mathcal{B} \to \mathbb{R} \cup \{+\infty\}$ a positive measure on the measurable space $(\Omega, \mathcal{B})$. A typical example of measurable space is a topological space endowed with the *Borel sigma-algebra*, i.e., the sigma-algebra generated by the open sets. The *measure space* $(\Omega, \mathcal{B}, \mu)$ is called a finite-measure space if $\mu(\Omega) < \infty$. A measurable map (function, transformation) f:$\Omega \to \Omega$ is said to preserve the measure μ, or to be μ-preserving, if $\mu(f^{-1}(B)) = \mu(B)$ for all $B \in \mathcal{B}$. Equivalently, the measure μ is said to be f-invariant. Sometimes $(\Omega, \mathcal{B}, \mu)$ is called the *state space* of the dynamic f.

Definition 8 Let $(\Omega, \mathcal{B}, \mu)$ be a finite-measure space and f:$\Omega \to \Omega$ a μ-preserving map. Then $(\Omega, \mathcal{B}, \mu, f)$ is called a measure-preserving dynamical system.

If $(\Omega, \mathcal{B}, \mu, f)$ is a measure-preserving dynamical system, we can assume without loss of generality that $\mu(\Omega) = 1$, i.e., that $(\Omega, \mathcal{B}, \mu)$ is a *probability space*. In this light, Ω is the space of elementary events, $\mathcal{B}$ comprises all outcomes we might be interested in, and $\mu(B)$ is the probability of the outcome $B \in \mathcal{B}$.

Given a measurable map f:$\Omega \to \Omega$, it is very difficult in practice to prove that f preserves the measure μ since, in general, not all elements $B \in \mathcal{B}$ are explicitly known. In general, all we know is a semi-algebra $\mathcal{S}$ generating $\mathcal{B}$. For example, if $\mathcal{B}$ is the Borel sigma-algebra of the interval $[0, 1] \subset \mathbb{R}$ with the standard topology, then $\mathcal{S}$ can be taken to be the collection of all subintervals of $[0, 1]$, or just the collection

J.M. Amigó, *Permutation Complexity in Dynamical Systems,*
Springer Series in Synergetics, DOI 10.1007/978-3-642-04084-9,

of subintervals of the forms $[0, b]$ and $(a, b]$, $0 \leq a < b \leq 1$. It can be proved [202] that if (i) $\mathcal{S}$ is a semi-algebra which generates $\mathcal{B}$ and (ii) for every $A \in \mathcal{S}$, $f^{-1}(A) \in \mathcal{B}$ and $\mu(f^{-1}(A)) = \mu(A)$, then f preserves the measure μ.

Exercise 13 Prove that $\mathcal{S} = \{[a, b) : 0 \leq a < b < 1\}$ is a semi-algebra of subsets of the interval $[0, 1)$ that generates the Borel sigma-algebra of $[0, 1)$.

Example 22 Suppose $\Omega = [0, 1)$, $\mathcal{B}$ is the Borel sigma-algebra of $[0, 1)$, and λ is the Lebesgue measure on $[0, 1)$. Furthermore, let $f : \Omega \to \Omega$ be the map given by $f(x) = Nx \bmod 1$, where $N \in \mathbb{Z}$, $|N| \geq 2$. Then f preserves λ. Indeed, for every half-open interval $[a, b) \subset [0, 1)$,

$$f^{-1}([a, b)) = \bigcup_{i=0}^{N-1} \left[\frac{a+i}{N}, \frac{b+i}{N}\right)$$

if $N \geq 2$ and

$$f^{-1}([a, b)) = \bigcup_{i=1}^{|N|} \left(\frac{i-b}{|N|}, \frac{i-a}{|N|}\right]$$

if $N \leq -2$. Hence,

$$\lambda\left(f^{-1}[a, b)\right) = \sum_{i=0}^{N-1} \frac{b-a}{N} = \sum_{i=1}^{|N|} \frac{b-a}{|N|} = b - a = \lambda([a, b)).$$

Example 23 Let the measure space $(\Omega, \mathcal{B}, \mu)$ be as in the previous example and $f : \Omega \to \Omega$ be given now by $f(x) = x + r \bmod 1$, with $r > 0$. This transformation preserves also the Lebesgue measure λ since, for every $[a, b) \subset [0, 1)$,

$$\begin{aligned}
f^{-1}([a, b)) &= [a - r, b - r) && \text{if } a \geq r, \\
f^{-1}([a, b)) &= [a + 1 - r, b + 1 - r) && \text{if } b \leq r, \\
f^{-1}([a, b)) &= [0, b - r) \cup [a + 1 - r, 1) && \text{if } a < r < b.
\end{aligned}$$

In any case,

$$\lambda\left(f^{-1}([a, b))\right) = b - a = \lambda([a, b)).$$

A perhaps more natural way of dealing with this example views f as a rotation on the circle. The f-invariance of λ is then straightforward.

More generally, the Lebesgue measure on $\mathbb{R}^n$ is invariant under translations and rotations in $\mathbb{R}^n$. More sophisticated examples of invariant measures include the Haar measure on a locally compact topological group, the map being the action of the group. In the next section we will meet invariant measures on product spaces.

Exercise 14 Let $f:[0,1) \to [0,1)$ be the Gauss transformation,

$$f(x) = \begin{cases} 0 & \text{if } x = 0, \\ \frac{1}{x} \pmod 1 & \text{if } x \neq 0. \end{cases}$$

Show that f preserves the measure

$$\mu(B) = \frac{1}{\ln 2} \int_B \frac{dx}{1+x}, \tag{A.1}$$

where B is a Borel set of $[0,1)$. Hint:

$$f^{-1}([a,b)) = \bigcup_{n=1}^{\infty} \left(\frac{1}{b+n}, \frac{1}{a+n} \right].$$

Krylov and Bogolioubov showed that invariant measures exist under quite general conditions.

Theorem 20 [202] *Let Ω be a compact metric space and $f:\Omega \to \Omega$ a continuous map. Then there exists an f-invariant probability measure μ on $(\Omega, \mathcal{B})$, where $\mathcal{B}$ is the Borel sigma-algebra of Ω.*

In general, there can exist more than one f-invariant measure and, besides, some of them can be rather "pathological." For instance, if δ_p is the Dirac measure at p, i.e.,

$$\delta_p(B) = \begin{cases} 1 & \text{if } p \in B \\ 0 & \text{if } p \notin B \end{cases},$$

$B \in \mathcal{B}$, and x is a period-n point for f, then

$$\mu(B) = \frac{1}{n} \sum_{k=0}^{n-1} \delta_{f^k(x)}(B)$$

($f^0(x) := x$ and $f^i(x) = f(f^{i-1}(x))$ for $i \geq 1$) is an atomic measure supported on the points $\{x, f(x), \ldots, f^{n-1}(x)\}$. A set $E \subset \Omega$ is said to be the (unique) *support* of μ if (i) E is closed in Ω, (ii) $\mu(E \cap U) > 0$ if $E \cap U \neq \emptyset$ and U is open in Ω, and (iii) $\mu(E') = 0$, where $E' = \Omega \backslash E$ is the complement of E.

In general, the ordered set $\{f^i(x) : i \geq 0\}$ is called the orbit or trajectory of the point (state, initial condition, etc.) $x \in \Omega$ under the "discrete-time" dynamic f and denoted by $\mathcal{O}_f(x)$. In the case of invertible maps, one writes $\mathcal{O}_f^+(x) = \{f^i(x) : i \geq 0\}$ for the "forward" orbit, while orbit means $\mathcal{O}_f(x) = \{f^i(x) : i \in \mathbb{Z}\}$.

It can happen that for almost all x in a set $U \subset \Omega$ with positive Lebesgue measure, its orbit is bounded and, moreover, the sequences of probability measures

$$\frac{1}{n}\sum_{k=0}^{n-1}\delta_{f^k(x)}$$

converge weakly to a measure μ, i.e., for almost all $x \in U$ and any continuous map $\varphi:\Omega \to \Omega$,

$$\lim_{k\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\varphi(f^k(x)) = \int_\Omega \varphi d\mu$$

holds. Then μ is an f-invariant measure that is usually called the natural or physical measure for its relevance in physics and computer simulations [72].

An important issue in measure-preserving dynamical systems is the existence of absolutely continuous invariant measures. A measure μ on a topological space Ω is said to be *absolutely continuous* (with respect to the Lebesgue measure dx), if $\mu(dx) = \rho(x)dx =: d\mu$, where the *density function* $\rho:\Omega \to \Omega$ (also called the Radon–Nikodym derivative of μ with respect to the Lebesgue measure, $d\mu/dx$) is continuous. For example, if μ is measure (A.1) on the interval $[0, 1)$ endowed with the Borel sigma-algebra, then

$$\mu(dx) = \frac{1}{\ln 2}\frac{dx}{1+x} \quad \text{or} \quad \frac{d\mu}{dx} = \frac{1}{\ln 2}\frac{1}{1+x}.$$

In general there are few results on the existence of absolutely continuous invariant measures. In the case of self-maps of one-dimensional intervals, there are some general conditions that appear in the usual theorems on existence of such measures.

Recall that a *partition* of a measure space $(\Omega, \mathcal{B}, \mu)$ is a disjoint collection of elements of $\mathcal{B}$ whose union is Ω.

Definition 9 Let $\alpha = \{I_i\}_{i=1}^d$ be a partition of the interval $I = [a, b] \subset \mathbb{R}$ into subintervals I_i. Given the map $f:I \to I$, assume that $f|_{I_i}$ is C^k ($k \geq 1$) for each i.

(a) f is said to be C^k *piecewise expanding* if there exists $\lambda > 1$ such that $|f'(x)| > \lambda$ for all $x \in I_i$ and each i.
(b) f is said to be C^k Markov if $f(\mathring{I}_i) \supset \mathring{I}_j$ whenever $f(\mathring{I}_i) \cap \mathring{I}_j \neq \emptyset$ ("Markov property"), where $\mathring{I}_i$ stands for the interior of I_i, $1 \leq i \leq d$. In this case, α is called a *Markov partition* for f. The matrix $A = (A_{ij})_{1\leq i, j\leq d}$ with

$$A_{i,j} = \left\{\begin{array}{ll} 1 & \textit{if } f(\mathring{I}_i) \supset \mathring{I}_j, \\ 0 & \textit{if } f(\mathring{I}_i) \cap \mathring{I}_j = \emptyset, \end{array}\right\} \tag{A.2}$$

is called the transition matrix for f.

See, for instance, [37, Chap. 5] and [105] for results concerning the existence of absolutely continuous invariant measures for piecewise expanding and/or Markov transformations (complying with additional conditions).

Exercise 15 Prove that the logistic map $g(x) = 4x(1-x)$, $0 \leq x \leq 1$, has an invariant measure with density function

$$\rho(x) = \frac{1}{\pi\sqrt{x(1-x)}}, \tag{A.3}$$

i.e., $\int_0^1 \rho(x)dx = 1$, and

$$\int_{[a,b]} \rho(x)dx = \int_{g^{-1}[a,b]} \rho(x)dx,$$

for all $0 \leq a < b \leq 1$. Figure A.1 shows the plot of the function $\rho(x)$. Is $g(x)$ piecewise expanding? Is $g(x)$ Markovian?

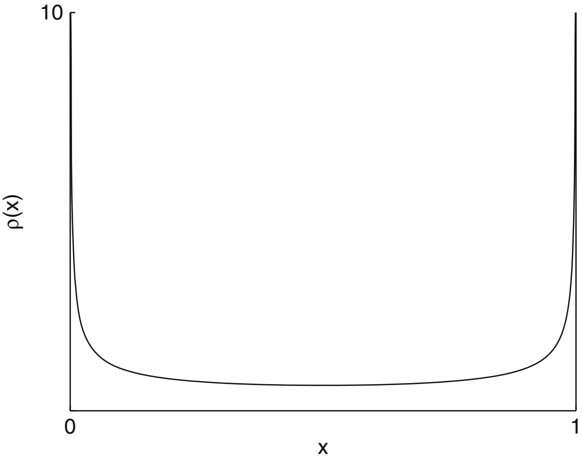

Fig. A.1 The density $\rho(x)$, (A.3)

Once we know that invariant measures are rather abundant objects, suppose that $f:\Omega \to \Omega$ is such that $f^{-1}(B) = B$ for some $B \in \mathcal{B}$. Then $f^{-1}(\Omega\backslash B) = \Omega\backslash B$ and the action of f on Ω can be decomposed into two disjoint pieces: $f|_B$ and $f|_{\Omega\backslash B}$. If f is indecomposable in the previous sense, one says that f is ergodic.

Definition 10 Let $(\Omega, \mathcal{B}, \mu, f)$ be a measure-preserving dynamical system. The map f is said to be ergodic if

$$f^{-1}(B) = B, \quad B \in \mathcal{B} \Rightarrow \mu(B) = 0 \quad \text{or} \quad \mu(B) = 1$$

Alternatively, μ is said to be an ergodic measure for f. Also, the dynamical system $(\Omega, \mathcal{B}, \mu, f)$ is said to be ergodic.

Thus an ergodic measure cannot be decomposed as a (properly weighted or "convex") sum of invariant measures. It might seem that this definition is a far cry from the original Boltzmann's *Ergodenhypothese*, which states that the trajectory of a closed thermodynamic system in the phase space (spanned by the coordinates and conjugate canonical momenta of its constituent particles) covers densely and uniformly the "energy shell," that is, the hypersurface in phase space defined by the restriction that the energy of the system is constant. But it was on the way to lying Boltzmann's proposal on a mathematically sound basis that G. Birkhoff introduced the concept of ergodicity in its modern version. Birkhoff's seminal *ergodic theorem* states the following.

Theorem 21 [202] *If* $(\Omega, \mathcal{B}, \mu, f)$ *is an ergodic dynamical system, then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \varphi(f^i(x)) = \int_{\Omega} \varphi d\mu \quad a.e. \tag{A.4}$$

for all $\varphi \in L^1(\mu)$.

As usual, "a.e." is shorthand for "almost everywhere" with respect to the relevant measure (μ here) and $L^1(\mu)$ is the space of μ-integrable functions. The property assumed by the *Ergodenhypothese* goes by the name of *topological transitivity* in the theory of discrete dynamical systems. A continuous self-map f of a compact metric space Ω is called topologically transitive if there exists some $x \in \Omega$ such that $\mathcal{O}_f(x)$ is dense in Ω (if f is invertible, then $\mathcal{O}_f(x)$ also includes the "backward" iterates $f^{-n}(x)$, $n \in \mathbb{N}$).

Let χ_B denote the *characteristic function* of the set $B \in \mathcal{B}$,

$$\chi_B(x) = \begin{cases} 1 & \text{if } x \in B \\ 0 & \text{if } x \notin B \end{cases}.$$

The substitution $\varphi = \chi_B$ in (A.4) yields then

$$\frac{1}{n} \sum_{i=0}^{n-1} \chi_B(f^i(x)) \to \mu(B) \quad \text{a.e.},$$

when $n \to \infty$. This means that if $(\Omega, \mathcal{B}, \mu, f)$ is ergodic, then the orbit of almost every initial condition $x \in \Omega$ visits the region B of the state space with asymptotic frequency $\mu(B)$. This resembles the law of large numbers in statistics and, in fact, there are plenty of deep relations between ergodic theory and statistics [31, 67].

Let Ω be a compact metrizable space Ω, and $\mathcal{B}$ the Borel sigma-algebra on Ω. A continuous map $f:\Omega \to \Omega$ is called *uniquely ergodic* if there is only one f-invariant Borel probability measure on Ω. A map f is uniquely ergodic if and only if it has exactly one invariant measure. If f is uniquely ergodic and μ is its invariant measure, then (A.4) holds for all continuous transformations φ and all $x \in \Omega$ [202].

Ergodicity is just but the first step in a series of notions measuring the statistical properties of the orbits generated by the dynamic: ergodicity, mixing, completely positive entropy, etc. Here we will recall only the definition of strong mixing.

Definition 11 The measure-preserving dynamical system $(\Omega, \mathcal{B}, \mu, f)$ is called (strong) mixing if

$$\lim_{n\to\infty} \mu(f^{-n}(A) \cap B) = \mu(A)\mu(B) \tag{A.5}$$

for all $A, B \in \mathcal{B}$.

In contrast to (A.5), f is ergodic if and only if

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(f^{-i}(A) \cap B) = \mu(A)\mu(B) \tag{A.6}$$

for all $A, B \in \mathcal{B}$. Hence mixing is a stronger condition than ergodicity. In practice it suffices to check (A.6) and (A.5) for $A, B \in \mathcal{S}$, a semi-algebra that generates $\mathcal{B}$.

Sufficient conditions for the existence of *ergodic* absolutely continuous invariant measures can be found, e.g., in [52, Chap. 5] . Mixing piecewise C^2 expanding Markov maps have unique ergodic invariant measures [105].

As in any other area of mathematics, the notion of isomorphism is central. It specifies when two dynamical systems are to be considered equivalent from the point of view of the properties that matter in this theory.

Definition 12 Given the measure-preserving dynamical systems $(\Omega_1, \mathcal{B}_1, \mu_1, f_1)$ and $(\Omega_2, \mathcal{B}_2, \mu_2, f_2)$, we say that f_1 is (metrically) isomorphic to f_2 if there exist $B_1 \in \mathcal{B}_1$, $B_2 \in \mathcal{B}_2$ with $\mu_1(B_1) = \mu_2(B_2) = 1$ such that (i) $f_1(B_1) \subset B_1, f_2(B_2) \subset B_2$ and (ii) there is an invertible, measure-preserving map $\phi : B_1 \to B_2$ with $\phi \circ f_1(x) = f_2 \circ \phi(x)$ for all $x \in B_1$.

The dynamical systems $(\Omega_1, \mathcal{B}_1, \mu_1, f_1)$ and $(\Omega_2, \mathcal{B}_2, \mu_2, f_2)$ are said to be isomorphic. Sometimes ϕ is called an isomorphism "modulo 0" or just "mod 0" (shorthand for modulo measure zero sets), but usually we dispense with measure zero sets without stating it explicitly. In the more general case that ϕ is measure preserving but only surjective, $(\Omega_2, \mathcal{B}_2, \mu_2, f_2)$ is called a *factor* of $(\Omega_1, \mathcal{B}_1, \mu_1, f_1)$ (or $(\Omega_1, \mathcal{B}_1, \mu_1, f_1)$ a *cover* of $(\Omega_2, \mathcal{B}_2, \mu_2, f_2)$) via the factor map ϕ. Two isomorphic maps are obtained from each other by a change of coordinates, so that properties that are independent of such changes of coordinates are invariant. Isomorphism invariants include ergodicity and mixing.

There is a broader (and more technical) concept called *conjugacy* that embraces isomorphism. Both concepts are though equivalent in virtually all probability spaces that one encounters in applications (e.g., compact metric spaces). Indeed, as it turns out, there is essentially only one type of probability space, called a *Lebesgue space*, which is characterized as being measure-theoretically isomorphic to the union of an interval of $\mathbb{R}$ endowed with Lebesgue measure, with at most countably many points of positive measure (called atoms) [177, 202]. In a Lebesgue space, set maps are always induced by point maps. Conjugacy and isomorphy coincide for a Lebesgue space, so both terms can be used interchangeably in that case.

Example 24 The symmetric tent map Λ:[0, 1] $\to$ [0, 1],

$$\Lambda(x) = \begin{cases} 2x & 0 \leq x \leq \frac{1}{2} \\ 2 - 2x & \frac{1}{2} \leq x \leq 1 \end{cases}, \tag{A.7}$$

preserves the Lebesgue measure $\lambda(dx) = dx$. If, furthermore, $\mu(dx) = \frac{1}{\pi\sqrt{x(1-x)}}dx$ is the natural invariant measure of the logistic map g:[0, 1] $\to$ [0, 1], $g(x) = 4x(1 - x)$ (see (A.3)), then ϕ:([0, 1], λ) $\to$ ([0, 1], μ) given by

$$\phi(x) = \sin^2\left(\tfrac{\pi}{2}x\right) \tag{A.8}$$

is invertible, measure preserving, and it satisfies $g \circ \phi = \phi \circ \Lambda$. Hence, Λ and g are conjugate.

Exercise 16 Show that

$$x_k = \sin^2(2^k\xi),$$

$\xi \in \mathbb{R}$, is a solution of the logistic recursion (or finite difference equation)

$$x_{k+1} = 4x_k(1 - x_k), \quad k \geq 0,$$

$x_k \in [0, 1]$, with initial condition $x_0 = \sin^2 \xi$.

A.2 Shift Systems

Shift systems are dynamical systems which due to their importance as models and prototypes are considered separately in this section. In the simplest and most usual version, the elements of the shift spaces are one-sided or two-sided sequences of N symbols or "letters". Sometimes one has to consider also sequences with elements from an arbitrary (countable or uncountable) "alphabet," and this requires some degree of sophistication. We set out from this more general situation.

First of all, let us recall the definition of a *product measurable space*. For our purposes it is sufficient to consider products of countably many copies of a measurable space $(\Omega, \mathcal{B})$. As index set $\mathbb{K}$ we take without restriction $\mathbb{K} = \mathbb{N}_0 := \{0\} \cup \mathbb{N}$ or $\mathbb{K} = \mathbb{Z}$. Then, $\Pi_{k\in\mathbb{K}}(\Omega, \mathcal{B}) = (\Omega^{\mathbb{K}}, \mathcal{B}_{\Pi}(\Omega))$, where

$$\Omega^{\mathbb{K}} = \{(\omega_k)_{k\in\mathbb{K}}:\omega_k \in \Omega\}$$

is the set of all one-sided sequences

$$(\omega_k)_{k\in\mathbb{N}_0} = \omega_0, \ldots, \omega_k, \ldots$$

if $\mathbb{K} = \mathbb{N}_0$, or the set of all two-sided sequences (also called bisequences or doubly infinite sequences)

$$(\omega_k)_{k\in\mathbb{Z}} = \ldots, \omega_{-k}, \ldots, \omega_0, \ldots, \omega_n, \ldots$$

if $\mathbb{K} = \mathbb{Z}$, and $\mathcal{B}_\Pi(\Omega)$ is the sigma-algebra generated by the semi-algebra $\mathcal{S}$ of *cylinder sets*

$$\prod_{j\in\mathbb{F}} A_j \times \prod_{k\notin\mathbb{F}} \Omega = \{(\omega_k)_{k\in\mathbb{K}} : \omega_j \in A_j \text{ for } j \in \mathbb{F}\}, \tag{A.9}$$

where $\mathbb{F} \subset \mathbb{K}$ is finite and $A_j \in \mathcal{B}$ for $j \in \mathbb{F}$. If $\mathbb{K} = \mathbb{N}_0$ (correspondingly, $\mathbb{K} = \mathbb{Z}$), then we can take $\mathbb{F} = \{0, 1, \ldots, n\}$ (correspondingly, $\mathbb{F} = \{-n, \ldots, 0, \ldots, n\}$), $n \in \mathbb{N}_0$, in (A.9) without restriction.

In most applications we have in mind (for instance, to information theory), $(\Omega, \mathcal{B}) = (S, 2^S)$ with $S = \{0, \ldots, N{-}1\}$, $N \geq 2$, and 2^S denoting as usual the family of all subsets of S. In this case, the set of all one-sided sequences of the symbols $0, 1, \ldots, N{-}1$,

$$S^{\mathbb{N}_0} = \{(s_n)_{n\in\mathbb{N}_0} : s_n \in S\}, \tag{A.10}$$

is called the (one-sided) *sequence space on N symbols*. Depending on the context, the set of symbols S may receive different names. In the setting of information theory, S is called an *alphabet*, its elements are called *letters*, and sequences $\mathbf{s} = (s_n)_{n\in\mathbb{N}_0}$ are called *messages*. In dynamics, S is sometimes called the state space and its elements, states. Segments (or *words*) of symbols of length L, like $s_k, s_{k+1}, \ldots, s_{k+L-1}$, will be shortened as s_k^{k+L-1}.

If S is thought to be a topological space (eventually endowed with the discrete topology), then $S^{\mathbb{N}_0}$ can be promoted to a topological space by means of the product topology, which is generated by the corresponding cylinder sets

$$C_{a_0,\ldots,a_n} = \{\mathbf{s} \in S^{\mathbb{N}_0} : s_k = a_k, 0 \leq k \leq n\}, \tag{A.11}$$

where $a_0, \ldots, a_n \in S$. (The general definition (A.9) with $A_j = \{a_j\}$ leads to the same topology.) The product topology makes $S^{\mathbb{N}_0}$ compact, perfect (i.e., it is closed and all its points are accumulation points), and totally disconnected. Such topological spaces are sometimes called *Cantor sets* because they are homeomorphic to Cantor's ternary set in the unit interval. By definition, the product sigma-algebra, $\mathcal{B}_\Pi(S)$, is generated by the cylinder sets (A.11) and comprises all Borel sets of $S^{\mathbb{N}_0}$.

Moreover, $S^{\mathbb{N}_0}$ is a metrizable space. In fact, there are several (non-equivalent) metrics compatible with the topology of $S^{\mathbb{N}_0}$, the perhaps most popular being

$$d_K(\mathbf{s},\mathbf{s}') = \sum_{n=0}^{\infty} \frac{\delta(s_n, s'_n)}{K^n}, \tag{A.12}$$

where $\delta(s_n, s'_n) = 1$ if $s_n \neq s'_n$, $\delta(s_n, s_n) = 0$ and $K > 2$. Observe that given $\mathbf{s} \in C_{a_0,\ldots,a_n}$, then $d_K(\mathbf{s},\mathbf{s}') < \frac{1}{K^n}$ if $\mathbf{s}' \in C_{a_0,\ldots,a_n}$, and $d_K(\mathbf{s},\mathbf{s}') \geq \frac{1}{K^n}$ if $\mathbf{s}' \notin C_{a_0,\ldots,a_n}$, thus $C_{a_0,\ldots,a_n} = B_{d_K}(\mathbf{s}; \frac{1}{K^n})$, the open ball of radius K^{-n} and center $\mathbf{s}$ in the metric space $(S^{\mathbb{N}_0}, d_K)$. Moreover, every point in $B_{d_K}(\mathbf{s}; \frac{1}{K^n})$ is a center, a property known from non-Archimedean normed spaces (e.g., the rational numbers with p-adic norms [115]).

Exercise 17 1. Prove that the cylinder sets (thus the open balls) are also closed in the product topology. Open and closed sets are sometimes called clopen sets.
2. Prove that the cylinder sets are not connected (i.e., they can be written as a disjoint union of open sets).

Shifting all the symbols of a one-sided sequence to the left one place and dropping the first symbol define a self-map of one-sided sequence spaces which plays an important role in both theory and applications. Formally, the (one-sided) *shift* $\Sigma: S^{\mathbb{N}_0} \to S^{\mathbb{N}_0}$ is defined as

$$\Sigma(s_0, s_1, s_2, \ldots) = (s_1, s_2, s_3, \ldots), \tag{A.13}$$

that is, $\Sigma(\mathbf{s}) = \mathbf{s}'$ with $s'_n = s_{n+1}$. Since $\Sigma^{-1} C_{a_0,\ldots,a_n} = \cup_{a \in S} C_{a,a_0,\ldots,a_n}$, Σ is continuous on $(S^{\mathbb{N}_0}, d_K)$, each point $\mathbf{s} \in S^{\mathbb{N}_0}$ having exactly N preimages under Σ. Furthermore, Σ has N fixed points: $\mathbf{s} = a_0^{\infty}$, $0 \leq a \leq N-1$.

In order to make a measure-preserving dynamical system out of $S^{\mathbb{N}_0}$, $\mathcal{B}_{\Pi}(S)$, and Σ, only a Σ-invariant measure is missing. All probability measures on $(S^{\mathbb{N}_0}, \mathcal{B}_{\Pi}(S))$ that make Σ a measure-preserving transformation are obtained in the following way [202]. For any $n \geq 0$ and $a_i \in S$, $0 \leq i \leq n$, let a real number $p_n(a_0, \ldots, a_n)$ be given such that (i) $p_n(a_0, \ldots, a_n) \geq 0$, (ii) $\sum_{a_0 \in S} p_0(a_0) = 1$, and (iii) $p_n(a_0, \ldots, a_n) = \sum_{a_{n+1} \in S} p_{n+1}(a_0, \ldots, a_n, a_{n+1})$. If we define now

$$m(C_{a_0,\ldots,a_n}) = p_n(a_0, \ldots, a_n),$$

then m can be extended to a probability measure on $(S^{\mathbb{N}_0}, \mathcal{B}_{\Pi}(S))$. The resulting dynamical system $(S^{\mathbb{N}_0}, \mathcal{B}_{\Pi}(S), m, \Sigma)$ is called the *one-sided shift system*.

If instead of considering (one-sided) sequences $s = (s_n)_{n \in \mathbb{N}_0}$, $s_n \in S = \{0, \ldots, N-1\}$, we consider two-sided sequences $\mathbf{s} = (s_n)_{n \in \mathbb{Z}}$, we are in the realm of the *two-sided sequence spaces on N symbols*,

$$S^{\mathbb{Z}} = \{(s_n)_{n \in \mathbb{Z}} : s_n \in S\}.$$

The corresponding (invertible) *two-sided shift* on $S^{\mathbb{Z}}$ is defined as $\Sigma: \mathbf{s} \mapsto \mathbf{s}'$ with $s'_n = s_{n+1}$, $n \in \mathbb{Z}$. (Although not strictly correct, we use the same letter Σ for one-sided and two-sided shifts.) The cylinder sets are given now as

$$C_{a_{-n},\ldots,a_0,\ldots,a_n} = \{\mathbf{s} \in S^{\mathbb{Z}}: s_k = a_k, |k| \le n\}$$

and

$$d_K(\mathbf{s}, \mathbf{s}') = \sum_{n\in\mathbb{Z}} \frac{\delta(s_n, s'_n)}{K^{|n|}},$$

$K > 3$, is a metric for $S^{\mathbb{Z}}$. The dynamical system $(S^{\mathbb{Z}}, \mathcal{B}_{\Pi}(S), m, \Sigma)$ is called the *two-sided shift system.*

Exercise 18 Prove that the cylinder set $C_{a_{-n},\ldots,a_0,\ldots,a_n}$ of $S^{\mathbb{Z}}$ coincides with the open ball $B_{d_K}(\mathbf{s}; K^{1-n})$, where $\mathbf{s}$ is any point of $C_{a_{-n},\ldots,a_0,\ldots,a_n}$.

Example 25 (a) Let $\mathbf{p} = (p_0, p_1, \ldots, p_{N-1})$, $N \ge 2$, be a probability vector with non-zero entries (i.e., $p_i > 0$ and $\sum_{i=0}^{N-1} p_i = 1$). Set

$$p_n(a_0, a_1, \ldots, a_n) = p_{a_0} p_{a_1} \cdots p_{a_n}.$$

The resulting measure on $(S^{\mathbb{K}}, \mathcal{B}_{\Pi}(S))$ is called the Bernoulli measure defined by $\mathbf{p}$. The dynamical system $(S^{\mathbb{K}}, \mathcal{B}_{\Pi}(S), m, \Sigma)$, where m is the Bernoulli measure defined by the probability vector $\mathbf{p}$, is called a one-sided (if $\mathbb{K} = \mathbb{N}_0$) or two-sided (if $\mathbb{K} = \mathbb{Z}$) $\mathbf{p}$-*Bernoulli shift*.

(b) Let $\mathbf{p} = (p_0, p_1, \ldots, p_{N-1})$ be a probability vector as in (a) and $P = (p_{ij})_{0\le i,j\le N-1}$ an $N \times N$ stochastic matrix (i.e., $p_{ij} \ge 0$ and $\sum_{j=0}^{N-1} p_{ij} = 1$) such that $\sum_{i=0}^{N-1} p_i p_{ij} = p_j$. Set then

$$p_n(a_0, a_1, \ldots, a_n) = p_{a_0} p_{a_0 a_1} p_{a_1 a_2} \cdots p_{a_{n-1} a_n}.$$

The resulting measure on $(S^{\mathbb{K}}, \mathcal{B}_{\Pi}(S))$ is called the Markov measure defined by $(\mathbf{p}, P)$. The dynamical system $(S^{\mathbb{K}}, \mathcal{B}_{\Pi}(S), m, \Sigma)$, where m is the Markov measure defined by the probability vector $\mathbf{p}$ and the stochastic matrix P, is called a one-sided (if $\mathbb{K} = \mathbb{N}_0$) or two-sided (if $\mathbb{K} = \mathbb{Z}$) $(\mathbf{p}, P)$-*Markov shift.* A $\mathbf{p}$-*Bernoulli shift* can be considered as a $(\mathbf{p}, P)$-Markov shift by taking $p_{ij} = p_j$.

Simple as they might seem, one-sided and two-sided shifts exhibit most of the basic properties of ergodic theory, like ergodicity and strong mixing. In particular, they are easily shown to be *chaotic* in the sense of Devaney [69], i.e., they are sensitive to initial conditions, are strong mixing, and their periodic points are dense. Let us recall at this point the notion of sensitivity to initial conditions.

Definition 13 Given a metric space (M, d), a map $f: M \to M$ is said to be sensitive to initial conditions if there exists $\delta > 0$, called a sensitivity constant, such that for every $x \in \Omega$ and $\varepsilon > 0$ there exists $y \in \Omega$ with $d(x, y) < \varepsilon$ and $d(f^n(x), f^n(y)) \ge \delta$ for some $n \in \mathbb{N}$.

Equivalently, a continuous self-map of a compact metric space is said to be chaotic if it is topologically transitive (that is, it has a dense orbit) and its periodic points are dense [91].

Exercise 19 Prove that the one- and two-sided shifts on N symbols are sensitive to initial conditions, are topological transitive, and their periodic points are dense.

Example 26 Let $\Omega = [0,1]$, $\mathcal{B}$ the Borel sigma-algebra of $[0,1]$, λ the corresponding Lebesgue measure, and $E_2{:}x \mapsto 2x \pmod 1$ the so-called dyadic map. The dynamical system $([0,1], \mathcal{B}, \lambda, E_2)$ is then isomorphic (up to a measure zero set) to the one-sided $(\frac{1}{2}, \frac{1}{2})$-Bernoulli shift on the symbols $\{0,1\} = S$. An isomorphism $\phi{:}S^{\mathbb{N}_0} \to [0,1]$ is given by

$$(x_0, x_1, \ldots, x_k, ..) \mapsto \sum_{k=0}^{\infty} x_k 2^{-(k+1)}. \tag{A.14}$$

Of course, the map ϕ is not injective in strict sense because the sequences $(x_0, \ldots, x_{n-1}, 0, 1^\infty)$ and $(x_0, \ldots, x_{n-1}, 1, 0^\infty)$ are sent to the same point (the upper label "∞" means indefinite repetition); indeed,

$$\sum_{k=0}^{n-1} x_k 2^{-(k+1)} + \sum_{k=n+1}^{\infty} 2^{-(k+1)} = \sum_{k=0}^{n-1} x_k 2^{-(k+1)} + 2^{-(n+1)}.$$

However, since the set of sequences eventually terminating in an infinite string of 0's or 1's is countable, we conclude that $(S^{\mathbb{N}_0}, \mathcal{B}_\Pi(S), m, \Sigma)$ and $([0,1], \mathcal{B}, \lambda, E_2)$ are conjugate modulo 0, i.e., the diagram

$$\begin{array}{ccc} \Sigma{:}\{0,1\}^{\mathbb{N}_0} & \to & \{0,1\}^{\mathbb{N}_0} \\ \phi \downarrow & & \downarrow \phi \\ E_2{:}[0,1] & \to & [0,1] \end{array}$$

is commutative almost everywhere: $E_2 = \phi \circ \Sigma \circ \phi^{-1}$. Observe that there is otherwise a topological obstruction that prevents $S^{\mathbb{N}_0}$ and $[0,1]$ from being homeomorphic: the first is (homeomorphic to) a Cantor set while, certainly, the second is not.

Exercise 20 Prove that the map $\phi{:}S^{\mathbb{N}_0} \to [0,1]$ defined in (A.14) is measure preserving, i.e., $m(\phi^{-1}(I)) = \lambda(I)$ for any interval $I \subset [0,1]$. It suffices to consider "dyadic" intervals, i.e., intervals of the forms $[0, k_2/2^n]$ and $(k_1/2^n, k_2/2^n]$, $0 \le k_1 < k_2 \le 2^n$, $n \in \mathbb{N}$.

Let us mention in passing the dyadic map $x \mapsto 2x \pmod 1$ is just the first member of the family of *expanding maps* of the circle:

$$E_N{:}x \mapsto Nx \pmod 1,$$

where N is an integer of absolute value greater than 1. In a way similar to Example 26 one can show that $([0,1], \mathcal{B}, \lambda, E_N)$ and the $(\frac{1}{N}, \ldots, \frac{1}{N})$-Bernoulli shift are conjugate for $N \ge 2$. In this case, map (A.14) is replaced by $(x_0, x_1, \ldots) \mapsto \sum_{k=0}^{\infty} x_k N^{-(k+1)}$.

Exercise 21 What transformation induces on the sequence space $\{0, 1\}^{\mathbb{N}_0}$ the expanding map E_{-2} via map (A.14)?

A.3 Stochastic Processes and Sequence Spaces

A stochastic (or random) process is a mathematical model for the occurrence of random phenomena as time goes on. This is the case, for example, when a random experiment is repeated over and over again. Put in a formal way, a *stochastic process* is a collection of random variables $\mathbf{X} = \{X_t\}_{t\in\mathcal{T}}$ on a common probability space $(\Omega, \mathcal{B}, \mu)$, called the *sample space*, taking on values in a measurable space $(S, \mathcal{A})$, called the *state space*. Technically this means that $X_t{:}\Omega \to S$ is a measurable map for all $t \in \mathcal{T}$, i.e., $X_t^{-1}(A) \in \mathcal{B}$ for all $A \in \mathcal{A}$. The index $t \in \mathcal{T}$ is conveniently interpreted as time, the usual choices for $\mathcal{T}$ being (i) $\mathcal{T} = \mathbb{R}$ or $\mathbb{R}_+ = [0, \infty]$, in which case $\mathbf{X}$ is called a continuous-time stochastic process or (ii) $\mathcal{T} = \mathbb{N}_0$ or $\mathbb{Z}$, in which case $\mathbf{X}$ is called a discrete-time stochastic process. The map $t \mapsto X_t(\omega)$ is the realization (sample path, trajectory, etc.) of the process $\mathbf{X}$ associated with the fixed sample point $\omega \in \Omega$. As usual in probability theory and statistics, a realization of a random variable X will be denoted by the same letter in small caps: $X(\omega) = x$.

The stochastic process $\mathbf{X}$ is characterized by its joint (finite-dimensional) probability distributions

$$\mu\{\omega \in \Omega{:}X_{t_1}(\omega) \in A_1, \ldots, X_{t_r}(\omega) \in A_r\} = \Pr\{X_{t_1} \in A_1, \ldots, X_{t_r} \in A_r\},$$

where $r \geq 1$, $t_1, \ldots, t_r \in \mathcal{T}$ and $A_1, \ldots, A_r \in \mathcal{A}$. If, furthermore, $\mathcal{T}$ is such that $\mathcal{T} + t \in \mathcal{T}$ for any $t \in \mathcal{T}$ (think of $\mathcal{T} = [0, \infty)$ or $\mathcal{T} = \mathbb{N}_0$) and the distribution of the random vector $(X_{t_1+t}, X_{t_2+t}, \ldots, X_{t_r+t})$ does not depend on t for any $r \geq 1$, $t_1, \ldots, t_r \in \mathcal{T}$, then the process $\mathbf{X}$ is called *stationary*. Stationary stochastic processes are also called *information sources* because they are used in information theory to model data sources.

In this book we consider mostly discrete-time, finite-state, one-sided stochastic processes modeling, say, finite-alphabet information sources or arising as symbolic dynamics after dividing the state space of a dynamical system. In this case we use the following notation for the joint probability distributions of the *discrete* random variables $X_0, \ldots, X_n$ with states in (without restriction) $S = \{0, 1, \ldots, N-1\}$:

$$\begin{aligned}\mu\{\omega \in \Omega{:}X_0(\omega) = x_0, \ldots, X_n(\omega) = x_n\} &= \Pr\{X_0 = x_0, \ldots, X_n = x_n\} \\ &= p(x_0, \ldots, x_n), \end{aligned} \tag{A.15}$$

and the corresponding notations for the conditional probabilities, etc. Occasionally, these finite-state processes will arise as discretizations or quantizations $\mathbf{X}^{\Delta}$ of processes $\mathbf{X}$ taking values in a finite interval $I \subset \mathbb{R}^q$ endowed with the Lebesgue measure. Formally this means that there exists a (usually uniform) partition $\delta = \{\Delta_1, \ldots, \Delta_{|\delta|}\}$ of I into a finite number of Lebesgue-measurable subsets (say, subintervals), such that $X_n^{\Delta} = a_j$ if $X_n^{\Delta} \in \Delta_j$, where $a_j \in \Delta_j$ is usually set by the precision with which the outputs of $\mathbf{X}$ are measured.

Example 27 A finite-state stochastic process $\mathbf{X} = \{X_n\}_{n\in\mathbb{N}_0}$ is called a Markov process or Markov chain if

$$\Pr\{X_n = x_n | X_{n-1} = x_{n-1}, \ldots, X_0 = x_0\} = \Pr\{X_n = x_n | X_{n-1} = x_{n-1}\},$$

$n \geq 1$, where $x_0, \ldots, x_n \in S = \{0, \ldots, N-1\}$. If, moreover, the conditional probability $\Pr\{X_n = x_n | X_{n-1} = x_{n-1}\}$ does not depend on n, then the Markov process $\mathbf{X}$ is called time homogeneous or time invariant. In this case,

$$P_{i,j} := \Pr\{X_n = j | X_{n-1} = i\},$$

$0 \leq i,j \leq N-1$, is called the transition matrix. We call a probability vector $\mathbf{p} = (p_0, \ldots, p_{N-1})$ an invariant, stationary, or equilibrium probability for $\mathbf{X}$ if $\mathbf{p} = \mathbf{p}P$, that is, if $\mathbf{p}$ is a left eigenvector of P with eigenvalue 1.

Any *stationary* discrete-time stochastic process $\mathbf{X} = \{X_n\}_{n\in\mathbb{K}}$ on a probability space $(\Omega, \mathcal{B}, \mu)$ with state space $(S, \mathcal{A})$ corresponds in a standard way to a shift system $(S^{\mathbb{K}}, \mathcal{B}_{\Pi}(S), m, \Sigma)$, where $(S^{\mathbb{K}}, \mathcal{B}_{\Pi}(S))$ is the product measurable space $\Pi_{k\in\mathbb{K}}(S, \mathcal{A})$, via the map $\Phi\colon\Omega \to S^{\mathbb{K}}$ defined by $(\Phi(\omega))_n = X_n(\omega)$. Here the measure m is the induced or transported probability on the space of possible outputs, $\mathcal{B}_{\Pi}(S)$, of the random process $\mathbf{X}$:

$$m(B) = \mu(\Phi^{-1}B), \quad B \in \mathcal{B}_{\Pi}(S), \tag{A.16}$$

that is, $m = \mu \circ \Phi^{-1}$ (note that $\Phi^{-1}B \in \mathcal{B}$ because each X_n is measurable). Moreover, because of the stationarity of $\mathbf{X}$, the probability measure m is shift invariant on cylinder sets and hence on all of $\mathcal{B}_{\Pi}(S)$.

We will also refer to the shift systems $(S^{\mathbb{K}}, \mathcal{B}_{\Pi}(S), m, \Sigma)$ as the (sequence space) *model* of the stochastic process or information source $\mathbf{X}$; if S is finite, then we may speak of a *sequence space model*. Models allow to focus on the random process itself as given by the probability distribution of its outputs, dispensing with a perhaps complicated underlying probability space. Depending on the setting or the process being modeled, some particular choices for S and/or $\mathbb{K}$ may be more convenient. For instance, one-sided random processes (i.e., $\mathbb{K} = \mathbb{N}_0$) provide better models than the two-sided processes $\{X_n\}_{n\in\mathbb{Z}}$ for physical information sources that must be turned on at some time. Also, if the source is digital, a finite state space S is the right choice.

Finally, since each information source has associated a dynamical system—its sequence space model—we can eventually assign dynamical properties to the sources. Thus, we say that a source $\mathbf{X}$ is *ergodic*, *mixing*, etc., if its sequence space model $(S^{\mathbb{K}}, \mathcal{B}_{\Pi}(S), m, \Sigma)$ possesses those properties.

Annex B
Entropy

In this annex we review only the Shannon, Kolmogorov–Sinai, and topological entropies. Standard references include [91, 169, 202].

B.1 Shannon Entropy

One of the most important characterizations one can attach to a random variable and to a stochastic process is its entropy and entropy rate, respectively. We refer to Annex A, Sect. A.3, for the basics of random processes.

B.1.1 The Entropy of a Discrete Random Variable

Let X be a random variable with sample space $(\Omega, \mathcal{B}, \mu)$ and finite state space S. If φ is a real-valued map on S, $\varphi: S \to \mathbb{R}$, then $\varphi \circ X = \varphi(X)$ is a random variable with finitely many states $\varphi(S) \subset \mathbb{R}$. The expectation value or average of $\varphi(X)$ will be denoted by $\mathbb{E}\varphi(X)$,

$$\mathbb{E}\varphi(X) = \sum_{x \in S} p(x)\varphi(x),$$

where $p(x)$ is the probability function of X (see (B.21) with $n = 0$).

Definition 14 The (Shannon) entropy of a discrete random variable X on a probability space $(\Omega, \mathcal{B}, \mu)$ is defined by

$$H(X) = -\sum_{x \in S} p(x) \log p(x) = \mathbb{E} \log \frac{1}{p(X)}. \tag{B.1}$$

Whenever convenient, we will write $H_\mu(X)$ to make clear which measure enters into the definition of entropy. Alternatively, one may write $H(p)$ since the entropy depends actually on the probability function $p(x)$ and not on the values taken by X.

(The previous observations hold also for the definitions of different kinds of entropy we will encounter in the sequel.) The logarithm in (B.1) may be taken to any base greater than 1. If the base 2 is used, the entropy comes in units of *bits* (shorthand for "binary digits"). Another usual choice for the logarithm base is Euler's number $e \approx 2.7182818\ldots$, in which case the units of the entropy are called *nats*. Unless otherwise stated, we will henceforth assume the entropy to be in units of bits. Recall that one can change from one logarithmic base a to another base b by means of the formula $\log_b p = \log_b a \log_a p$. By convention, $0 \times \log 0 := \lim_{x\to 0+} x \log x = 0$. Note that $H(X) \geq 0$ because $0 < p(x) \leq 1$ implies $-\log p(x) = \log \frac{1}{p(x)} \geq 0$. On the other hand if $|S|$ denotes the cardinality of the state space S, then $H(X) \leq \log |S|$, as can be easily proved, e.g., using Lagrange multipliers, the highest entropy corresponding to random variables with equiprobable outcomes, that is, $p(x) = 1/|S|$ for all $x \in S$. Observe that Boltzmann's equation (6.1) is nothing else but the entropy for such a flat probability function, $H(X) = \log |S|$, except for the notation (S means entropy in (6.1), while we use S to denote the state space throughout the book) and the physical constant k_B.

Example 28 Suppose that a random variable X takes values $0, 1$ with probabilities $p(0) = p$, $p(1) = 1 - p(0) = 1 - p$. Then

$$H(X) = -p \log p - (1 - p) \log (1 - p) = H(p). \tag{B.2}$$

The function $H(p)$ is plotted in Fig. B.1. We see that $H(p)$ vanishes when $p = 0$ or $p = 1$, i.e., when the outcome is certain, and it is maximal when $p = 1/2$, i.e., when the uncertainty about the outcome is maximal: $H(1/2) = \log 2 = 1$ bit.

The entropy of a discrete random variable can be given different meanings; see [22] for three interesting interpretations. In information theory one defines

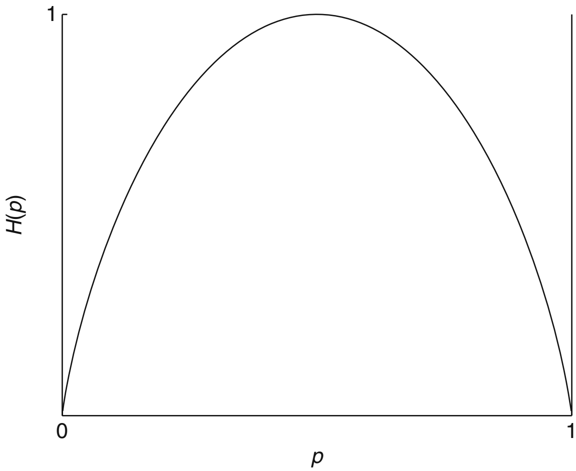

Fig. B.1 The function $H(p)$, (B.2)

$I(X) = -\log p(X)$ to be the *information* of a random variable X with probability function $p(x)$, $-\log p(x)$ being the information conveyed by the outcome $X = x$. Observe that the more rare the event x (that is, the more unlikely the observation of the event x), the more information is gained from its occurrence; one can argue that the most probable events are the less informative ones since their occurrence comes as no surprise. According to Definition 14, $H(X)$ is then the expected value of the information of X: $H(X) = \mathbb{E}I(X)$. Furthermore, if we agree that uncertainty means lack of information, then the entropy can be interpreted as the average uncertainty associated with a random variable or random experiment. In this light, equiprobable events correspond to maximal uncertainty about the outcome.

We turn now to the problem of characterizing the uncertainty associated with more than one random variable.

The *relative entropy* or *Kullback–Leibler distance* between two probability mass functions $p(x)$ and $q(x)$, $x \in S$, is defined as

$$D(p \parallel q) = \sum_{x \in S} p(x) \log \frac{p(x)}{q(x)}. \tag{B.3}$$

In this definition, the convention (based on continuity arguments) that $0 \log \frac{0}{q} = 0$ and $p \log \frac{p}{0} = \infty$ is used. From definition (B.3) it follows that $D(p \parallel q) \geq 0$ and $D(p \parallel q) = 0$ if and only if $p = q$ [59]. On the other hand (and despite of its name), $D(p \parallel q)$ is not symmetric in p, q and does not satisfy the triangle inequality. Nonetheless, it is often useful to think of $D(p \parallel q)$ as a "distance" between the distributions p and q. The relative entropy $D(p \parallel q)$ is a measure of the inefficiency of assuming that the distribution of the random variable X is q when the true distribution is p. For example, if we knew the true distribution p of X, then we could construct a code with average code-word length $H(p)$ (see Sect. 1.1.1, (1.2)). If, instead, we use the code for a distribution q, we would need $H(p) + D(p \parallel q)$ bits on the average to describe the random variable X.

Let X and Y be two random variables on a common sample space $(\Omega, \mathcal{B}, \mu)$ but, in general, with different finite state spaces S_1 and S_2, respectively. This corresponds to a situation where two different observations or measurements (with finite precision) are made at the same random experiment. If X and Y have the joint probability function

$$p(x, y) = \mu\{\omega \in \Omega : X(\omega) = x, Y(\omega) = y\} = \Pr(X = x, Y = y)$$

$(x \in S_1, y \in S_2)$, then the *joint entropy* of X and Y is defined as

$$H(X, Y) = -\sum_{x \in S_1} \sum_{y \in S_2} p(x, y) \log p(x, y) = \mathbb{E} \log \frac{1}{p(X, Y)}. \tag{B.4}$$

It is easy to prove that

$$H(X,Y) \leq H(X) + H(Y).$$

The generalization of (B.4) to $n \geq 2$ random variables is straightforward and needs no further elaboration.

The joint probability function $p(x,y)$ and the conditional probability function

$$p(y\,|x) = \frac{p(x,y)}{p(x)}$$

allow the definition of two instrumental concepts in information theory: the conditional entropy and the mutual information. The *conditional entropy* of Y given X is

$$H(Y\,|X) = -\sum_{x\in S_1}\sum_{y\in S_2} p(x,y)\log p(y\,|x) = \mathbb{E}\log\frac{1}{p(Y\,|X)}, \tag{B.5}$$

and the *mutual information* of X and Y is

$$\begin{aligned} I(X;Y) &= H(X) - H(X\,|Y) = H(Y) - H(Y\,|X) \\ &= H(X) + H(Y) - H(X,Y) \\ &= I(Y;X), \end{aligned} \tag{B.6}$$

where we have used the so-called *chain rule* [59]:

$$H(X,Y) = H(X) + H(Y\,|X)\,. \tag{B.7}$$

Note that $H(Y\,|X)$ is the average of the uncertainties

$$H(Y\,|X=x) = -\sum_{y\in S_2} p(y\,|x)\log p(y\,|x)$$

weighted with the probabilities $p(x)$, $x \in S_1$. As for the mutual information of two random variables, $I(X;Y)$ is the information about X conveyed by Y (i.e., the information about the realization of X knowing the realization of Y), which is the same as the information about Y conveyed by X, (B.6). Alternatively,

$$I(X;Y) = \mathbb{E}\log\frac{p(X,Y)}{p(X)p(Y)}.$$

Let us mention in passing that the *capacity* of a discrete memoryless channel with input X, output Y, and transition probability $p(Y\,|X)$ is defined as

$$C = \max_{p(x)} I(X;Y),$$

where the maximum is taken over all possible input distributions $p(x)$.

Again, the generalization of these concepts to $n_1 + n_2$ random variables $X_0, \ldots, X_{n_1-1}$ and $Y_0, \ldots, Y_{n_2-1}$ is straightforward. In particular, the (joint) entropy of the random vector $X_0^{n-1} = X_0, \ldots, X_{n-1}$, where, say, all components can take the same states $x_i \in S$, is given by

$$H(X_0, \ldots, X_{n-1}) = -\sum_{x_0,\ldots,x_{n-1}\in S} p(x_0, \ldots, x_{n-1}) \log p(x_0, \ldots, x_{n-1})$$
$$= \mathbb{E} \log \frac{1}{p(X_0, \ldots, X_{n-1})},$$

where $p(x_0, \ldots, x_{n-1})$ is the joint probability function of $X_0, \ldots, X_{n-1}$.

Exercise 22 By iteration of the two-variable rules $p(X, Y) = p(X)p(Y|X)$ and (B.7) prove the general chain rule for the joint entropy: given the random variables $X_0, \ldots, X_{n-1}$ with a joint probability function $p(x_0, \ldots, x_{n-1})$, then

$$p(X_0, \ldots, X_{n-1}) = \prod_{i=0}^{n-1} p(X_i |X_{i-1}, \ldots, X_0) \tag{B.8}$$

and

$$H(X_0, \ldots, X_{n-1}) = \sum_{i=0}^{n-1} H(X_i |X_{i-1}, \ldots, X_0), \tag{B.9}$$

with the conventions $p(X_0 |X_{-1}) := p(X_0)$ and $H(X_0 |X_{-1}) := H(X_0)$.

B.1.2 The Entropy Rate of a Discrete-Time Finite-State Stochastic Process

Definition 15 The entropy rate of a finite-state random process $\mathbf{X} = \{X_n\}_{n\in\mathbb{N}_0}$ on a probability space $(\Omega, \mathcal{B}, \mu)$ is defined by

$$h(\mathbf{X}) = \lim_{n\to\infty} \frac{1}{n} H(X_0, \ldots, X_{n-1}), \tag{B.10}$$

provided the limit exists.

Sometimes the terms

$$h(X_0, \ldots, X_{n-1}) = \frac{1}{n} H(X_0, \ldots, X_{n-1})$$

($n \geq 2$) are called the *entropy rates of order n* of $\mathbf{X}$. Hence, $h(X_0, \ldots, X_{n-1})$ or, more compactly written, $h(X_0^{n-1})$ is the average uncertainty per symbol (time unit, channel use, etc. depending on the interpretation of n) about n consecutive outcomes of the random experiment modeled by $\mathbf{X}$. If we repeat the experiment an arbitrarily long number of times, these average uncertainty rates eventually converge to a limit—Shannon's entropy rate $h(\mathbf{X})$.

Although $h(X_0^{n-1})$ and, consequently, $h(\mathbf{X})$ are actually entropy *rates*, the term "rate" is generally omitted—also in other types of entropy. We follow sometimes this common usage, since this does not lead to misunderstandings.

Lemma 12 *For a stationary stochastic process* $\mathbf{X} = \{X_n\}_{n\in\mathbb{N}_0}$, *the sequence of conditional entropies* $H(X_n\,|X_{n-1},\ldots,X_0)$ *is decreasing.*

Proof Indeed,

$$\begin{aligned} H(X_{n+1}\,|X_n,\ldots,X_1,X_0) &\le H(X_{n+1}\,|X_n,\ldots,X_1) \\ &= H(X_n\,|X_{n-1},\ldots,X_0)\,, \end{aligned}$$

where the inequality follows from the fact that conditioning reduces uncertainty, and the equality follows from the stationarity of $\mathbf{X}$. □

Theorem 22 *For a stationary stochastic process* $\mathbf{X} = \{X_n\}_{n\in\mathbb{N}_0}$,

$$h(\mathbf{X}) = \lim_{n\to\infty} H(X_n\,|X_{n-1},\ldots,X_0)\,. \tag{B.11}$$

Proof First of all, limit (B.11) converges because, according to Lemma 12, the positive sequence $H(X_n\,|X_{n-1},\ldots,X_0)$ is decreasing. Furthermore, by the chain rule (B.9),

$$h(X_0,\ldots,X_n) = \frac{1}{n+1}\sum_{i=0}^{n} H(X_i\,|X_{i-1},\ldots,X_0)\,.$$

By Cesáro's mean theorem ("If $a_n \to a$ and $b_n = \frac{1}{n+1}\sum_{i=0}^{n} a_i$, then $b_n \to a$"),

$$h(\mathbf{X}) = \lim_{n\to\infty} h(X_0,\ldots,X_n) = \lim_{n\to\infty} H(X_n\,|X_{n-1},\ldots,X_0)\,.$$

□

From Lemma 12 it follows that the convergence of the entropy rates of order n, $h(X_0,\ldots,X_{n-1})$, to $h(\mathbf{X})$ is monotonically decreasing:

$$h(X_0) \ge h(X_0,X_1) \ge \cdots \ge h(X_0,\ldots,X_{n-1}) \ge \cdots\,. \tag{B.12}$$

Thus, when estimating the entropy rate of a stationary random process by its entropy rate of order n, the estimation always exceeds the true value. Intuitively speaking, with increasing n we see more and more correlations among the variables $X_0,\ldots,X_{n-1}$ and this reduces our uncertainty about the next observation X_n. We turn back to this point in Example 31.

In an information-theoretical setting and in applications (Sect. A.3), one can think of a stationary stochastic process $\mathbf{X} = \{X_n\}_{n \in \mathbb{N}_0}$ as a data source. Its realizations are then the messages output by the source. This is illustrated in Fig. B.2. Here x_0 can be considered the current and last letter of the message, the other letters having been output in the past, the greater the index, the earlier in time.

Fig. B.2 A data source **X** outputs a message x_0^∞

B.2 Kolmogorov–Sinai Entropy

B.2.1 Deterministic Systems

A partition of a probability space $(\Omega, \mathcal{B}, \mu)$ is a collection $\alpha = (A_i)_{i \in J}$ of disjoint sets $A_i \in \mathcal{B}$, with a countable index set J, such that $\bigcup_{i \in J} \mu(A_i) = 1$. If J is finite, α is called a finite partition. If α is a finite partition of $(\Omega, \mathcal{B}, \mu)$, then the collection of all elements of $\mathcal{B}$ which are unions of elements of α is a finite sub-sigma-algebra of $\mathcal{B}$ which we denote by $\mathcal{B}(\alpha)$. We write $\alpha \leq \beta$, where α, β are two finite partitions of $(\Omega, \mathcal{B}, \mu)$, to mean that each element of α is a union of elements of β. In this case, β is called a *refinement* of α. We have $\alpha \leq \beta$ iff $\mathcal{B}(\alpha) \subset \mathcal{B}(\beta)$.

Definition 16 Let $\alpha = \{A_1, \ldots, A_{|\alpha|}\}$ be a finite partition of $(\Omega, \mathcal{B}, \mu)$. The entropy of the partition α is the number

$$H_\mu(\alpha) = -\sum_{i=1}^{|\alpha|} \mu(A_i) \log \mu(A_i).$$

The same considerations concerning the base of the logarithm we made after the definition of Shannon's entropy, Definition 14, apply here as well. By the same token, $H(\alpha)$ is a measure of the information gained (or the uncertainty removed) by performing a random experiment whose outcomes have probabilities $\mu(A_1), \ldots, \mu(A_{|\alpha|})$.

Sometimes it is convenient to quantify the "coarseness" of a partition. Roughly speaking, if we assign a "size" to each $A \in \alpha$, then we can take the maximum of those sizes as the coarseness of α. The resulting parameter is called the *norm of the partition* α and denoted by $\|\alpha\|$. In metric spaces (X, d), one can take $\|\alpha\| = \max_{A \in \alpha} \mathrm{diam}(A)$, where $\mathrm{diam}(A) = \sup\{d(x, y) : x, y \in A\}$ is called the "diameter" of A.

If $f : \Omega \to \Omega$ is a measure-preserving function on the probability space $(\Omega, \mathcal{B}, \mu)$, we denote by $f^{-n}\alpha$ the partition $\{f^{-n}A_1, \ldots, f^{-n}A_{|\alpha|}\}$. Furthermore, given two finite partitions $\alpha = \{A_1, \ldots, A_{|\alpha|}\}$ and $\beta = \{B_1, \ldots, B_{|\beta|}\}$ of $(\Omega, \mathcal{B}, \mu)$, we denote by $\alpha \vee \beta$ their *least common refinement*,

$$\alpha \vee \beta = \{A \cap B : A \in \alpha, B \in \beta, \mu(A \cap B) > 0\}.$$

More general refinements, like

$$\alpha \vee f^{-1}\alpha \vee \cdots \vee f^{-(n-1)}\alpha = \bigvee_{i=0}^{n-1} f^{-i}\alpha,$$

are defined recursively.

Definition 17 Let $(\Omega, \mathcal{B}, \mu, f)$ be a measure-preserving dynamical system. If α is a finite partition of $(\Omega, \mathcal{B}, \mu)$, then

$$h_\mu(f, \alpha) = \lim_{n \to \infty} \frac{1}{n} H_\mu \left(\bigvee_{i=0}^{n-1} f^{-i}\alpha \right) \tag{B.13}$$

is called the metric entropy of f with respect to α.

In this setting, consider now a finite-state random process $\mathbf{X}^\alpha = \{X_n^\alpha\}_{n \in \mathbb{N}_0}$, with $X_n^\alpha : \Omega \to S = \{0, \ldots, |\alpha| - 1\}$, defined as follows:

$$X_n^\alpha(\omega) = i \quad \text{iff} \quad f^n(\omega) \in A_i \in \alpha. \tag{B.14}$$

Note that $X_{n+1} = X_n \circ f$, thus $X_n = X_n \circ f^n$. Then

$$\begin{aligned} \Pr\left\{X_0^\alpha = i_0, \ldots, X_n^\alpha = i_n\right\} &= \mu\left\{\omega \in \Omega : \omega \in A_{i_0}, f(\omega) \in A_{i_1}, \ldots, f^n(\omega) \in A_{i_n}\right\} \\ &= \mu\left\{A_{i_0} \cap \cdots \cap f^{-n} A_{i_n}\right\}, \end{aligned} \tag{B.15}$$

$n \geq 0$, and similarly,

$$\begin{aligned} \Pr\left\{X_k^\alpha = i_0, \ldots, X_{n+k}^\alpha = i_n\right\} &= \mu\left\{f^{-k}(A_{i_0} \cap \cdots \cap f^{-n} A_{i_n})\right\} \\ &= \Pr\left\{X_0^\alpha = i_0, \ldots, X_n^\alpha = i_n\right\} \end{aligned}$$

because of the f-invariance of μ. We conclude that $\mathbf{X}^\alpha$ is a stationary process, which is called the *symbolic dynamics* of $(\Omega, \mathcal{B}, \mu, f)$ with respect to the partition ("coarse graining" or "quantization") α. Depending on the context, $\mathbf{X}^\alpha$ is also called a *coding map* (dynamical systems) or a collection of *simple observations* with respect to f with precision $\|\alpha\|$ (information theory). Moreover, it follows from (B.15) that

$$h_\mu(f, \alpha) = h_\mu(\mathbf{X}^\alpha). \tag{B.16}$$

This not only proves that limit (B.13) does exist but also that the *entropy rates of order n* of f with respect to α,

$$h_\mu^{(n)}(f,\alpha) = \frac{1}{n} H_\mu \left(\bigvee_{i=0}^{n-1} f^{-i}\alpha \right),$$

decrease to $h_\mu(f,\alpha)$ when $n \to \infty$ (remember (B.12)).

Definition 18 Let $(\Omega, \mathcal{B}, \mu, f)$ be a measure-preserving dynamical system and α a finite partition of $(\Omega, \mathcal{B}, \mu)$. Then,

$$h_\mu(f) = \sup_\alpha h_\mu(f,\alpha) \tag{B.17}$$

is called the metric entropy (or just, the entropy) of the map f with respect to μ.

Sometimes $h_\mu(f)$ is called the Kolmogorov–Sinai entropy or the measure-theoretic entropy too. To streamline the notation, the subscript μ may be dropped from $H_\mu(\alpha)$, $h_\mu(f,\alpha)$, and $h_\mu(f)$, as we generally do, if the probability measure is clear from the context.

The isomorphic invariance is one of the fundamental properties of entropy.

Theorem 23 *(a) If the dynamical systems* $(\Omega_1, \mathcal{B}_1, \mu_1, f_1)$ *and* $(\Omega_2, \mathcal{B}_2, \mu_2, f_2)$ *are isomorphic, then* $h(f_1) = h(f_2)$.
(b) If $(\Omega_2, \mathcal{B}_2, \mu_2, f_2)$ *is a factor of* $(\Omega_1, \mathcal{B}_1, \mu_1, f_1)$, *then* $h(f_2) \leq h(f_1)$.

It should be obvious from definitions (B.13) and (B.17) that the exact calculation of $h(f)$ from scratch is, in general, unfeasible. There are though a few results that, depending on the specifics of the dynamical system in question, can come to the rescue. We mention a few next.

A finite partition α of $(\Omega, \mathcal{B}, \mu)$ is called a *generating partition* or a *generator for a* μ*-preserving transformation* $f : \Omega \to \Omega$ if (i)

$$\bigvee_{n=-\infty}^{\infty} f^{-n}\mathcal{B}(\alpha) = \mathcal{B} \text{ (modulo } \mu\text{-zero sets)} \tag{B.18}$$

when f is invertible (i.e., f is an automorphism) or (ii)

$$\bigvee_{n=0}^{\infty} f^{-n}\mathcal{B}(\alpha) = \mathcal{B} \text{ (modulo } \mu\text{-zero sets)} \tag{B.19}$$

when f is non-invertible (i.e., f is an endomorphism). This means that for any $B \in \mathcal{B}$, there is a $B' \in \bigvee_{n=-\infty}^{\infty} f^{-n}\mathcal{B}(\alpha)$ or $B' \in \bigvee_{n=0}^{\infty} f^{-n}\mathcal{B}(\alpha)$, respectively, such that $\mu(B \triangle B') = 0$. If f is invertible and the stronger condition (B.19) holds, then α is called a strong or one-sided generator for f. Equivalent definitions of generators and one-sided generators by means of partition refinements converging to the point partition $\epsilon = \{\{x\} : x \in \Omega\}$ were given in Sect. 1.3.

Example 29 Since the sigma-algebra $\mathcal{B}_\Pi(S)$ of the one-sided and two-sided shift spaces are generated by the cylinder sets

$$C_{a_0,\ldots,a_k} = \{\mathbf{s} = (s_n)_{n\in\mathbb{N}_0} : s_0 = a_0, \ldots, s_k = a_k\} = \bigcap_{i=0}^{k} \Sigma^{-i} C_{a_i}$$

and

$$C_{a_{-k},\ldots,a_0,\ldots,a_k} = \{\mathbf{s} = (s_n)_{n\in\mathbb{Z}} : s_{-k} = a_{-k}, \ldots, s_k = a_k\} = \bigcap_{i=-k}^{k} \Sigma^{-i} C_{a_i},$$

respectively, it follows that the partition

$$\gamma = \{C_a : a \in S\}$$

is a generator of both the one-sided and two-sided shifts.

Generating partitions can be found numerically; see, e.g., [40] for a general method based on relaxation algorithms. For higher dimensional maps, numerical techniques have been proposed for the dissipative Henón map [87], the standard map [53], two-dimensional hyperbolic maps [26], etc. A method based on unstable period orbits was proposed in [63]. The construction of one-dimensional maps possessing generating partitions was studied in [99].

Theorem 24 *(Kolmogorov–Sinai Theorem) Let* $(\Omega, \mathcal{B}, \mu, f)$ *be a dynamical system.*

(a) If f *is an automorphism and* α *is a generator or a one-sided generator for* f*, then* $h(f) = H(f, \alpha)$*.*
(b) If f *is an endomorphism and* α *is a generator for* f*, then* $h(f) = H(f, \alpha)$*.*

The case of automorphisms with one-sided generators is uninteresting since then one can show that $h(f) = 0$ [202]. More interestingly, *Krieger's theorem* states that if f is an ergodic automorphism with $h(f) < \infty$, then f has a generator [67, 130, 169]. Although Krieger's proof is non-constructive, Smorodinsky [191] and Denker [65] provided methods to construct a two-sided generator for ergodic and aperiodic automorphisms. Denker's construction could even be extended by Grillenberger [66] to all aperiodic automorphisms. The existence of generators for endomorphisms was proved by Kowalski under different assumptions [128, 129]. At variance with the previous case, the construction of one-sided generators for endomorphisms remains an open problem till this very day; see [182] for some progress in this issue.

Example 30 Using the fact that the cylinder sets C_a are generators for the one-sided and two-sided $(\mathbf{p}, P)$-Markov shifts Σ on N symbols, one can prove

$$h_\mu(\Sigma) = -\sum_{i,j=1}^{N} p_i P_{ij} \log P_{ij}, \tag{B.20}$$

where μ is the Markov measure defined by $(\mathbf{p}, P)$ (see Example 25 (b)). Upon substituting $P_{ij} = p_j$ in (B.20), we get for **p**-Bernoulli shifts

$$h_\mu(\Sigma) = -\sum_j^N p_j \log p_j,$$

where μ is the Bernoulli measure defined by **p** (see Example 25 (a)).

A second practical way of calculating (or, at least, estimating) the entropy is provided by the following theorem.

Theorem 25 [169, Ch. 5, Prop. 3.6] *Let* $(\Omega, \mathcal{B}, \mu, f)$ *be a measure-preserving dynamical system. If* $\alpha_0 \leq \alpha_1 \leq \cdots$ *is an increasing sequence of finite partitions of* $(\Omega, \mathcal{B}, \mu)$ *and* $\vee_{n=0}^{\infty} \mathcal{A}(\alpha_n) = \mathcal{B}$ *up to sets of measure* 0*, then*

$$\lim_{n\to\infty} h_\mu(f, \alpha_n) = h_\mu(f).$$

A third practical method calls for Pesin's theorem and Lyapunov exponents. Since this topic would take us too far away, we refer the interested reader to the specialized literature [142, 52, 72]. Due to the important role that the Lyapunov exponent(s) play in nonlinear dynamics, several numerical schemes have been developed to calculate them [193]. On the other hand, Pesin's theorem and its generalizations require the invariant measure to possess some properties—but invariant measures are in many interesting cases unknown. This fact limits the application of this method. For the calculation of the metric entropy in some one-dimensional systems, see [105].

B.2.2 Random Systems

Let $\mathbf{X} = \{X_n\}_{n\in\mathbb{N}_0}$ be a stationary stochastic process on a probability space $(\Omega, \mathcal{B}, \mu)$, taking on values in $S = \{0, \ldots, N-1\}$. In Sect. A.3 it is shown that $\mathbf{X}$ can be associated in a canonical way with a shift system $(S^{\mathbb{N}_0}, \mathcal{B}_\Pi(S), m, \Sigma)$, called its sequence space model, via $\Phi{:}\Omega \to S^{\mathbb{N}_0}$, $(\Phi(\omega))_n = X_n(\omega)$. The joint probability function $p(x_0, \ldots, x_{n-1})$ of the random process $\mathbf{X}$ is related to the measure of the cylinder sets $C_{x_0,\ldots,x_{n-1}}$, $x_0, \ldots, x_{n-1} \in S$, of the sequence space model in the following way:

$$\begin{aligned} p(x_0, \ldots, x_{n-1}) &= \mu\,\{\omega \in \Omega{:}X_0(\omega) = x_0, \ldots, X_{n-1}(\omega) = x_{n-1}\} \\ &= \mu\left\{\Phi^{-1}\left\{\mathbf{s} \in S^{\mathbb{N}_0}{:}s_0 = x_0, \ldots, s_{n-1} = x_{n-1}\right\}\right\} \\ &= m\left\{C_{x_0,\ldots,x_{n-1}}\right\} \\ &= m\{C_{x_0} \cap \cdots \cap \Sigma^{-(n-1)} C_{x_{n-1}}\}. \end{aligned}$$

Since the partition $\gamma = \{C_{x_0}{:}x_0 \in S\}$ is a generator of Σ (Example 29), we have

$$\begin{aligned}
h_\mu(\mathbf{X}) &= -\lim_{n\to\infty}\frac{1}{n}\sum_{x_0,\ldots,x_{n-1}\in S} p(x_0,\ldots,x_{n-1})\log p(x_0,\ldots,x_{n-1}) \\
&= -\lim_{n\to\infty}\frac{1}{n}H_m\left(\bigvee_{i=0}^{n-1}\Sigma^{-i}\gamma\right) \\
&= h_m(\Sigma,\gamma) \\
&= h_m(\Sigma)
\end{aligned}$$

by Theorem 24 (b). In words, the Shannon entropy rate of a stochastic process $\mathbf{X} = \{X_n\}_{n\in\mathbb{N}_0}$ coincides with the Kolmogorov–Sinai entropy rate of its sequence space model.

An important property of ergodic processes is the so-called *asymptotic equipartition property* or *Shannon–McMillan–Breiman theorem*.

Theorem 26 *(Shannon–McMillan–Breiman) If $\mathbf{X} = \{X_n\}_{n\in\mathbb{N}_0}$ is a finite-valued stationary ergodic process, then $-\frac{1}{n}\log p(X_0,\ldots,X_{n-1})$ converges in probability to the entropy rate $h(\mathbf{X})$.*

Example 31 The sequence space model of a finite-state, time-homogeneous Markov chain $\mathbf{X} = \{X_n\}_{n\in\mathbb{N}_0}$ (Example 27) with transition matrix $P_{i,j}$, $0 \le i,j \le N-1$, and stationary probability vector $\mathbf{p}$ is the one-sided $(\mathbf{p},P)$-Markov shift $\Sigma_{\mathbf{p},P}$. Therefore,

$$h(\mathbf{X}) = h(\Sigma_{\mathbf{p},P}) = -\sum_{i,j=0}^{N-1} p_i P_{ij}\log P_{ij}.$$

For the specific case

$$P = \begin{pmatrix} 1-p_{01} & p_{01} \\ p_{10} & 1-p_{10} \end{pmatrix} = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix},$$

the stationary probability is

$$\mathbf{p} = \left(\frac{p_{10}}{p_{01}+p_{10}}, \frac{p_{01}}{p_{01}+p_{10}}\right) = \left(\frac{1}{2}, \frac{1}{2}\right).$$

The upper curve in Fig. B.3 shows the entropy rates of order n, $h(X_0,\ldots,X_{n-1})$, closing in on the true value $h(\mathbf{X}) = 0.469$ bits/symbol (horizontal line). The lower curve shows what happens in practice when $h(\mathbf{X})$ is estimated numerically in a naive way. Here the probabilities $p(x_0,\ldots,x_{n-1})$ were estimated by the frequencies of the word $x_0,\ldots,x_{n-1}$ in a sequence of 10,000 draws. In the left part of the experimental curve, we see the entropy rates of successive order $n = 1,2,\ldots$ converging from above to the true value. For $n \approx 20$, the numerical values provide accurate estimates of the entropy. For greater lengths, the estimates tend toward zero along the parabola

$$h(n) = \frac{\log (N - n + 1)}{n}$$

due to undersampling.

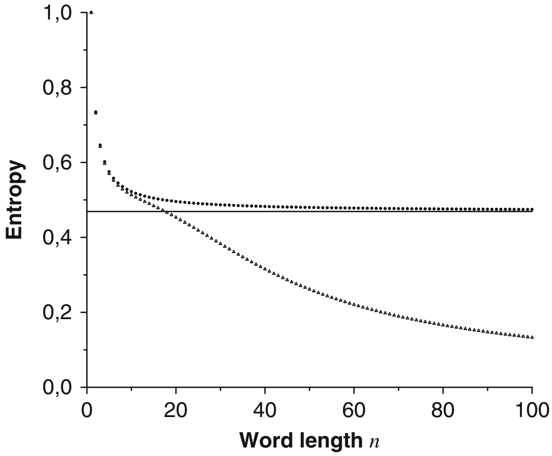

Fig. B.3 The *upper dotted line* shows the convergence of the entropy rate of order n to the true value, 0.469 bits/symbol (*horizontal line*), for an arbitrarily long sequence generated by a two-state Markov chain with transition probabilities $p_{01} = p_{10} = 0.1$. The *lower dotted line* shows what happens in practice due to undersampling

A particular case is of interest. Consider now not a general stationary stochastic process but the symbolic dynamics $\mathbf{X}^{\alpha} = \{X_n^{\alpha}\}_{n \in \mathbb{N}_0}$ of the system $(\Omega, \mathcal{B}, \mu, f)$ with respect to a partition $\alpha = \{A_1, \ldots, A_{|\alpha|}\}$ (see (B.14)), and let $(S^{\mathbb{N}_0}, \mathcal{B}_{\Pi}(S), m, \Sigma)$ be the sequence space model of $\mathbf{X}^{\alpha}$; hence $S = \{1, \ldots, |\alpha|\}$ and

$$m(C_{a_0, a_1, \ldots, a_n}) = \mu(A_{a_0} \cap f^{-1} A_{a_1} \cap \cdots \cap f^{-n} A_{a_n}$$

for any cylinder set $C_{a_0, \ldots, a_n} = \{\mathbf{s} \in S^{\mathbb{N}_0}: s_0 = a_0, \ldots, s_n = a_n\}$, with $a_0, \ldots, a_n \in S$. In this setting, the following question arises. When are the dynamical systems $(\Omega, \mathcal{B}, \mu, f)$ and $(S^{\mathbb{N}_0}, \mathcal{B}_{\Pi}(S), m, \Sigma)$ isomorphic (via $\Phi^{\alpha}: \Omega \to S^{\mathbb{N}_0}$, $(\Phi^{\alpha}(\omega))_n = X_n^{\alpha}(\omega)$)? Since $\{C_a: a \in S\}$ is a generator for Σ and $(\Phi^{\alpha})^{-1} C_a = A_a$ for every $a \in S$, we need clearly that

$$\{(\Phi^{\alpha})^{-1} C_a: 1 \le a \le |\alpha|\} = \{A_a: 1 \le a \le |\alpha|\} = \alpha$$

is also a generator for f. In other words, a generator for f gives a natural isomorphism between $(\Omega, \mathcal{B}, \mu, f)$ and the sequence space model associated with its symbolic dynamics. By Krieger's theorem we conclude that any ergodic, invertible dynamical system with finite entropy can be represented as a two-sided shift system. This result is useful in that it provides prototypes of ergodic, finite-entropy systems.

B.3 Topological Entropy

Topological entropy for continuous self-maps of compact topological spaces was introduced by Adler, Koheim, and McAndrews by means of open covers [3]. Later Dinaburg [70] and Bowen [36] found alternative approaches via separating and spanning sets in (not necessarily compact) metric spaces.

B.3.1 Generalities

Recall that a continuous or topological dynamical system is a pair (M,f), where M is a topological space and $f{:}M \to M$ is a continuous map. As compared to measure-theoretical dynamical systems, there is here no measurable structure involved (although M can be thought to be endowed with the Borel sigma-algebra); instead, continuity enters the scenario. Sometimes, continuity is weakened to piecewise continuity, especially in conjunction with other properties like piecewise monotonicity.

Furthermore, in this section (M, d) denotes a metric space and $f{:}M \to M$ a uniformly continuous map. If, moreover, M is compact, then f needs only to be continuous (since every continuous self-map of a compact space is uniformly continuous).

Definition 19 Let K be a compact topological space, α an open cover of K, and $N(\alpha)$ the number of sets in a finite subcover of α with smallest cardinality. The entropy of the cover α is then defined as $H(\alpha) = \log N(\alpha)$.

If α is an open cover of K and $f{:}K \to K$ is continuous, then $f^{-1}\alpha$ is the open cover consisting of all sets $f^{-1}A$, $A \in \alpha$.

Definition 20 If α is an open cover of the compact space K and $f{:}K \to K$ is continuous, then the entropy of f relative to α is given by

$$h(f, \alpha) = \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} f^{-i}\alpha\right) \tag{B.21}$$

and the topological entropy of f is given by

$$h(f) = \sup_{\alpha} h(f, \alpha). \tag{B.22}$$

It can be proved that the limit in (B.21) exists and the supremum in (B.22) can be taken over *finite* open covers of K.

In a metric space (M, d), the alternative definitions of topological entropy via spanning and separating sets may be more useful.

Definition 21 Let $n \in \mathbb{N}$, $\varepsilon > 0$, and $K \subset M$ compact. A subset $A \subset M$ is said to (n, ε)-span K with respect to $f{:}M \to M$ if for each $x \in K$ there exists $y \in A$ such that

$$\max_{0\leq i\leq n-1} d(f^i(x), f^i(y)) \leq \varepsilon.$$

Furthermore, let $r_n(\varepsilon, K)$ denote the smallest cardinality of any (n, ε)-spanning set for K with respect to f.

Definition 22 The topological entropy of $f{:}M \to M$ is

$$h_d(f) = \sup_K \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log r_n(\varepsilon, K), \tag{B.23}$$

where the supremum is taken over all compact subsets of M.

The definition of topological entropy by means of separating sets is as follows.

Definition 23 Let $n \in \mathbb{N}$, $\varepsilon > 0$, and $K \subset M$ compact. A subset $A \subset K$ is said to be (n, ε)-separated with respect to $f{:}M \to M$ if $x, y \in A$, $x \neq y$, implies

$$\max_{0\leq i\leq n-1} d(f^i(x), f^i(y)) > \varepsilon.$$

Furthermore, let $s_n(\varepsilon, K)$ denote the largest cardinality of any (n, ε)-separated subset of K with respect to f.

Thus, an (n, ε)-separated subset of Ω is a kind of microscope that allows us to distinguish orbits of length n up to a precision ε.

Definition 24 The topological entropy of $f{:}M \to M$ is

$$h_d(f) = \sup_K \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log s_n(\varepsilon, K), \tag{B.24}$$

where the supremum is taken over all compact subsets of M.

If M is compact, then $h_d(f)$ can be shown [202] not to depend on the metric d (thus, it will be denoted by $h_{\text{top}}(f)$) and, moreover, definitions (B.23) and (B.24) can be simplified to

$$h_{\text{top}}(f) = \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log r_n(\varepsilon, M) = \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log s_n(\varepsilon, M). \tag{B.25}$$

Both $r_n(\varepsilon, M)$ and $s_n(\varepsilon, M)$ can be interpreted as the number of orbits of length n up to an error ε. For $\varepsilon \ll 1$,

$$e^{nh(f)} \sim r_n(\varepsilon, M) \quad \text{and} \quad e^{nh(f)} \sim s_n(\varepsilon, M),$$

where $\sim$ stands for "asymptotically as $n \to \infty$" (assuming the convergence of $\frac{1}{n} \log r_n(\varepsilon, M)$ and $\frac{1}{n} \log s_n(\varepsilon, M)$ in this limit), so the topological entropy measures the asymptotic exponential growth rate with n of the number of orbits of length n, up to error ε.

Definition 25 Let $f_1:M_1 \to M_1$ and $f_2:M_2 \to M_2$ be continuous maps of metric spaces and suppose that there exists a continuous surjective map $\phi:M_1 \to M_2$ such that $\phi \circ f_1 = f_2 \circ \phi$. Then we say that f_1 is topologically semiconjugate to f_2 or that f_2 is a factor of f_1 via the topological semi-conjugacy or factor map ϕ. In the case that ϕ is a homeomorphism, then f_1 and f_2 are said to be topologically conjugate and ϕ is said to be a topological conjugacy.

In particular, if two maps are metrically conjugate via a (measure-preserving) homeomorphism, then they are also topologically conjugate. Such is the case of the logistic and symmetric tent maps via the homeomorphism A.8 (Example 24). The qualifiers "topological" and "topologically" may be dropped if it is clear that they refer to a topological system.

Thus, conjugate maps are obtained from each other by a continuous change of coordinates. Therefore, properties that are independent of such changes of coordinates will be invariant under topological conjugacy, e.g., sensitivity to initial conditions, topological transitivity, number of periodic orbits of a given period.

Just as metric entropy is an invariant of metric conjugacy, so is topological entropy an invariant of topological conjugacy.

Theorem 27 *Let f_1 and f_2 be continuous self-maps of compact spaces. If f_1 and f_2 are topologically conjugate, then $h(f_1) = h(f_2)$. More generally, if f_2 is a factor of f_1, then $h(f_2) \leq h(f_1)$.*

Exercise 23 Show that the quadratic transformations $f_1(x) = vx(1-x)$ on $[0, 1]$, $0 < v \leq 4$, and

$$f_2(y) = \frac{1}{2}(y^2 - v^2 + 2v)$$

on $[-v, v]$ are topologically conjugate via the homeomorphism

$$\phi(x) = v(1 - 2x) = f_1'(x).$$

In spite of not involving a measure-theoretical structure, topological entropy is tightly related to metric entropy through the following *variational principle*.

Theorem 28 *Let M be a compact metric space endowed with the Borel sigma-algebra $\mathcal{B}$, and $f:M \to M$ a continuous map. Then*

$$h_{top}(f) = \sup h_\mu(f), \tag{B.26}$$

where the supremum is taken over all f-invariant measures μ on the measurable space $(M, \mathcal{B})$.

Note that the set of f-invariant measures invoked in the variational principle (B.26) is non-empty by Theorem 20. Moreover, the supremum in (B.26) can be restricted to ergodic measures [202],

$$h_{\text{top}}(f) = \sup_{\mu \in E(M,f)} h_\mu(f), \tag{B.27}$$

where $E(M,f)$ is the set of f-invariant, ergodic measures on $(M, \mathcal{B})$. Measures μ such that $h_{\text{top}}(f) = h_\mu(f)$ are called *measures with maximal entropy* for obvious reasons.

In Sect. A.1 we defined the concept of generator of a measure-preserving transformation. In topological dynamics, there is also a concept of generator that plays a similar role with respect to the topological entropy. Given a compact metric space M and a map $f{:}M \to M$, a finite open cover $\alpha = \{A_1, \ldots, A_{|\alpha|}\}$ of M is said to be a *generator* for f if

(a) in case f is invertible, for any bisequence $(a_i)_{i\in\mathbb{Z}}$, $1 \le a_i \le |\alpha|$, the intersection

$$\bigcap_{i=-\infty}^{\infty} f^{-i} A_{a_i}$$

contains at most one point or

(b) in case f is non-invertible, for any sequence $(a_i)_{i\in\mathbb{N}_0}$, $1 \le a_i \le |\alpha|$, the intersection

$$\bigcap_{i=0}^{\infty} f^{-i} A_{a_i}$$

contains at most one point.

The topological dynamical systems that admit a generator have a simple characterization.

Definition 26 Let M be a compact metric space. A homeomorphism (correspondingly, a continuous map) $f{:}M \to M$ is said to be *expansive* if there exists $\delta > 0$, called an *expansivity constant* for f, such that

$$d(f^n(x), f^n(y)) \le \delta$$

for all $n \in \mathbb{Z}$ (correspondingly, $n \in \mathbb{N}_0$) implies $x = y$. Expansive non-invertible maps and homeomorphisms for which the expansiveness condition holds already for non-negative iterates are collectively called positively expansive maps.

Alternatively, if $x \ne y$ and δ is an expansivity constant for f, then there exists $n \in \mathbb{Z}$ (correspondingly, $n \in \mathbb{N}_0$) with $d(f^n(x), f^n(y)) > \delta$. Notice that expansiveness differs from sensitive dependence in that *all* nearby points eventually separate by at least δ (for sensitive dependence it suffices this to occur for a single point in each neighborhood of the other). Intuitively, the orbits of an expansive map f can be resolved to any desired precision by taking n sufficiently large. Expansive maps

f have some nice properties like having a countable number of periodic points, and at least one invariant measure with maximal entropy [202]. Examples of expansive maps include the shift transformations and the hyperbolic toral automorphisms. On the other hand, there are no expansive maps of closed one-dimensional intervals [19, Thm. 2.2.31] nor expansive homeomorphisms of the circle [202]. Expansiveness and positively expansiveness are topological conjugacy invariants.

Theorem 29 *Let $f{:}M \to M$ be a map of the compact metric space (M,d). Then f is expansive if and only if f has a generator.*

Observe that the cylinder sets C_a are generators both in the measure-theoretical and in the topological senses because, among other considerations, they build a partition and an open cover at the same time. Therefore, shifts on sequence spaces are expansive transformations. Expansiveness is an invariant of topological conjugacy.

Theorem 30 *If $f{:}M \to M$ be an expansive map of the compact metric space (M,d) and α is a generator for f, then $h_{top}(f) = h(f,\alpha)$.*

Example 32 Let $S = \{0,\ldots,k-1\}$ and Σ be the shift on the bisequence space $S^{\mathbb{Z}} = \{(s_n)_{n\in\mathbb{Z}}\}$. Then Σ has topological entropy $\log N$. Indeed, apply Theorem 30 with α comprising the cylinder sets $C_j = \{(x_n)_{n\in\mathbb{Z}}{:}x_0 = j\}$ to obtain

$$h_{top}(\Sigma) = \lim_{n\to\infty} \frac{1}{n} \log N\left(\bigvee_{i=0}^{n-1} \Sigma^{-i}\alpha\right) = \lim_{n\to\infty} \frac{1}{n} \log k^n = \log k.$$

Thus, if μ_0 is the Bernoulli measure on $(S^{\mathbb{Z}}, \mathcal{B}_\Pi(S))$ defined by the probability vector $\mathbf{p}_0 = (\frac{1}{k},\ldots,\frac{1}{k})$, we have

$$h_{\mu_0}(\Sigma) = \log k = h_{top}(\Sigma).$$

This illustrates the existence of (in this case, unique) measures of maximal entropy. The result in the one-sided case is the same.

Example 33 Let $S = \{0,\ldots,k-1\}$, $A = (a_{ij})_{i,j=0}^{k-1}$ be a $k \times k$ matrix whose entries a_{ij} are either 0's or 1's, and

$$\Omega_A = \{\omega \in S^{\mathbb{Z}}{:}a_{\omega_n\omega_{n+1}} = 1 \text{ for } \forall n \in \mathbb{Z}\}.$$

The space Ω_A is closed and shift invariant. The restriction

$$\Sigma_A := \Sigma|_{\Omega_A}$$

is called the two-sided *topological Markov chain* determined by the matrix A, a *Markov subshift*, or a *subshift of finite type* (see Sect. 1.1.2). One-sided topological Markov chains are defined analogously over $S^{\mathbb{N}_0}$. The matrix A is said to be irreducible if for any pair i,j there is $n > 0$ such that $a_{ij}^{(n)} > 0$, where $a_{ij}^{(n)}$ are the entries of A^n. If A is irreducible and Σ_A is a one-sided or two-sided topological Markov chain, then [202]

$$h_{top}(\Sigma_A) = \log\lambda, \tag{B.28}$$

where λ is the largest positive eigenvalue of A. A topological Markov chain Σ_A has a unique measure (called its Parry measure) of maximal topological entropy.

It can be proved [31, Sect. 4.3] that a C^2 piecewise expanding Markov map f is topologically conjugate (modulo 0) to the one-sided topological Markov chain Σ_A, where A is the transition matrix for f. Therefore, piecewise expanding Markov maps admit a symbolic description.

Example 34 Consider the *rooftop map* f defined by

$$f(x) = \begin{cases} ax + c & \text{if } 0 \le x \le c, \\ (1-b)x & \text{if } c \le x \le 1, \end{cases}$$

$a > 1$, $b > 1$, and $c = \frac{1}{1+a}$; see Fig. B.4. Set $I_1 = [0, c)$ and $I_2 = [c, 1]$. Then f is C^∞ on I_1 and I_2 (lateral derivatives at the endpoints),

$$|f'(x)| = \begin{cases} a & \text{if } x \in I_1, \\ b & \text{if } x \in I_2, \end{cases}$$

and

$$f(\mathring{I}_1) = \mathring{I}_2, \quad f(\mathring{I}_2) \supset \mathring{I}_1 \cup \mathring{I}_2.$$

It follows that f is a smooth piecewise expanding Markov map with transition matrix

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix},$$

see (B.2). Finally, from (B.28) we get

$$h_{top}(f) = h_{top}(\Sigma_A) = \log\tfrac{1+\sqrt{5}}{2}.$$

B.3.2 Topological Entropy of One-Dimensional Maps

Topological entropy, as metric entropy, is in general difficult to calculate and even to estimate. An exception worth mentioning because of its importance in applications is the case of one-dimensional interval maps.

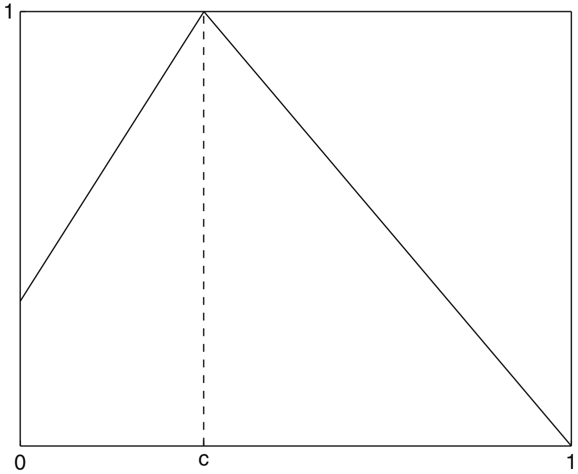

Fig. B.4 Rooftop map

Definition 27 Given an interval $I \subset \mathbb{R}$, a map $f{:}I \to I$ is said to be piecewise monotone if there is a finite partition of I into subintervals, such that f is continuous and monotone on each of those subintervals.

If $f{:}I \to I$ is piecewise monotone, there are different expressions for its topological entropy $h_{\text{top}}(f)$ that allow calculating it analytically in many cases. For instance [4, 155],

$$h_{\text{top}}(f) = \lim_{n\to\infty} \frac{1}{n} \log \operatorname{lap}(f^n) \tag{B.29}$$

and

$$h_{\text{top}}(f) = \lim_{n\to\infty} \frac{1}{n} \log \left|\{x \in I{:}f^n(x) = x\}\right|, \tag{B.30}$$

where $\operatorname{lap}(f^n)$ is the number of pieces of monotonicity of f^n (called laps of f^n) and $|\cdot|$ stands for the cardinality.

Other expressions of $h(f)$ are related to the notion of variation [4, 155]:

$$h_{\text{top}}(f) = \lim_{n\to\infty} \frac{1}{n} \log^+ \operatorname{var}(f^n), \tag{B.31}$$

where, as usual, $\log^+ x = \max\{0, \log x\}$. Let us recall that the variation of a function $\varphi{:}I \to \mathbb{R}$ is given as

$$\operatorname{var}(\varphi) = \sup\left\{\sum_{i=1}^{s} |\varphi(x_i) - \varphi(x_{i-1})|\right\},$$

where the supremum is taken over all finite sequences $x_0 < x_1 < \cdots < x_s$ of elements of I. If φ is piecewise monotone, then (i) $\mathrm{var}(\varphi) < \infty$, (ii) φ has finite derivative φ' almost everywhere on I, and (iii) φ' is integrable on I [95]. In this case,

$$\mathrm{var}(\varphi) = \int_I \left|\varphi'(x)\right| dx. \tag{B.32}$$

Note that, for a piecewise monotone map φ, $\mathrm{var}(\varphi)$ is closely related to the length of the graph of φ,

$$\mathrm{len}(\varphi) = \int_I \sqrt{1 + |\varphi'(x)|^2} dx.$$

Indeed, since

$$\left|\varphi'(x)\right| < \sqrt{1 + |\varphi'(x)|^2} \leq \left|\varphi'(x)\right| + 1 \tag{B.33}$$

for all $x \in I$, we have

$$\mathrm{var}(\varphi) < \mathrm{len}(\varphi) \leq \mathrm{var}(\varphi) + \mathrm{len}(I), \tag{B.34}$$

upon integration of (B.33) over the interval I ($\mathrm{len}(I)$ denotes the length of I). It follows

$$\lim_{n\to\infty} \frac{1}{n} \log^+ \mathrm{len}(f^n) = \lim_{n\to\infty} \frac{1}{n} \log^+ \mathrm{var}(f^n) = h(f), \tag{B.35}$$

since $\lim_{n\to\infty} \frac{1}{n} \log^+ \mathrm{len}(I) = 0$.

Corollary 10 *If f is a continuous, piecewise monotone interval map of constant slopes $\pm s$, then*

$$h_{top}(f) = \log^+ s.$$

This result is very interesting for the following reason. If f is a continuous, piecewise monotone interval map and $h_{\mathrm{top}}(f) = \log \beta > 0$, then f is semiconjugate to some continuous, piecewise monotone interval map of constant slopes $\pm\beta$ (via a non-decreasing map) [4]. If, moreover, f is topologically transitive, then "semiconjugate" can be replaced by "conjugate" in the previous statement (and the condition $h_{\mathrm{top}}(f) > 0$ can be dropped because it is automatically satisfied).

Finally, let us mention that there are efficient algorithms for the numerical estimation of the topological entropy of piecewise monotone interval maps; see, for example, [27] for an algorithm that converges rapidly and provides both upper and lower bounds.

References

1. H.D.I. Abarbanel, *Analysis of Observed Chaotic Data*. Springer, New York, 1996.
2. M. Abraimowitz and I.A. Stegun (Eds.), *Handbook of Mathematical Functions*. Dover, New York, 1972.
3. R.L. Adler, A.G. Koheim, and M.H. McAndrews, Topological entropy, *Transactions of the American Mathematical Society* **114** (1965) 309–319.
4. L. Alsedà, J. Llibre, and M. Misiurewicz, *Combinatorial Dynamics and Entropy in Dimension One*. World Scientific, Singapore, 2000.
5. G. Alvarez, M. Romera, G. Pastor, and F. Montoya, Gray codes and 1D quadratic maps, *Electronic Letters* **34** (1998) 1304–1306.
6. J.M. Amigó, J. Szczepanski, E. Wajnryb, and M.V. Sanchez-Vives, Estimating the entropy of spike trains via Lempel-Ziv complexity, *Neural Computation* **16** (2004) 717–736.
7. J.M. Amigó, M.B. Kennel, and L. Kocarev, The permutation entropy rate equals the metric entropy rate for ergodic information sources and ergodic dynamical systems, *Physica D. Nonlinear Phenomena* **210** (2005) 77–95.
8. J.M. Amigó, L. Kocarev, and J. Szczepanski, Order patterns and chaos, *Physics Letters A* **355** (2006) 27–31.
9. J.M. Amigó and M.B. Kennel, Variance estimators for the Lempel-Ziv entropy rate estimators, *Chaos* **16** (2006) 043102.
10. J.M. Amigó, L. Kocarev, and I. Tomovski, Discrete entropy, *Physica D. Nonlinear Phenomena* **228** (2007) 77–85.
11. J.M. Amigó and M.B. Kennel, Topological permutation entropy, *Physica D. Nonlinear Phenomena* **231** (2007) 137–142.
12. J.M. Amigó, S. Zambrano, and M.A.F. Sanjuán, True and false forbidden patterns in deterministic and random dynamics, *Europhysics Letters* **79** (2007) 50001.
13. J.M. Amigó, L. Kocarev, and J. Szczepanski, Discrete Lyapunov exponent and resistance to differential cryptanalysis, *IEEE Transactions on Circuits and Systems II* **54** (2007) 882–886.
14. J.M. Amigó, S. Elizalde, and M.B. Kennel, Forbidden patterns and shift systems, *Journal of Combinatorial Theory, Series A* **115** (2008) 485–504.
15. J.M. Amigó, S. Zambrano, and M.A.F. Sanjuán, Combinatorial detection of determinism in noisy time series, *Europhysics Letters* **83** (2008) 60005.
16. J.M. Amigó and M.B. Kennel, Forbidden ordinal patterns in higher dimensional dynamics, *Physica D. Nonlinear Phenomena* **237** (2008) 2893–2899.
17. J.M. Amigó, The ordinal structure of the signed shift transformations, *International Journal of Bifurcation and Chaos* **19** (2009) 3311–3327.
18. C. Anteneodo, A.M. Batista, and R.L. Viana, Synchronization threshold in coupled logistic map lattices, *Physica D. Nonlinear Phenomena* **223** (2006) 270–275.
19. N. Aoki and K. Hiraide, *Topological theory of dynamical systems*. North Holland, Amsterdam, 1994.

20. D.K. Arrowsmith and C.M. Place, *Dynamical Systems*. Chapman and Hall, Boca Raton, 1996.
21. D. Arroyo, G. Alvarez, and J.M. Amigó, Estimation of the control parameter from symbolic sequences: Unimodal maps with variable critical point, *Chaos* **19** (2009) 023125.
22. R.B. Ash, *Information Theory*. Dover Publications, New York, 1990.
23. H. Atmanspacher and H. Scheingraber, Inherent global stabilization of unstable local behavior in coupled map lattices, *International Journal of Bifurcation and Chaos* **15** (2005) 1665–1676.
24. N. Ay and J.P. Crutchfield, Reductions of hidden information sources, *Journal of Statistical Physics* **120** (2005) 659–684.
25. E. Babson and E. Steingrímsson, Generalized permutation patterns and a classification of the Mahonian statistics, Séminaire Lotharingien de Combinatoire **44** (2000), Article B44b, 18.
26. A. Bäcker and N. Chernov, Generating partitions for two-dimensional hyperbolic maps, *Nonlinearity* **11** (1998) 79–87.
27. N.J. Balmforth, E.A. Spiegel, and C. Tresser, Topological entropy of one-dimensional maps: Approximations and bounds, *Physical Review Letters* **72** (1994) 80–83.
28. C. Bandt and B. Pompe, Permutation entropy: A natural complexity measure for time series, *Physical Review Letters* **88** (2002) 174102.
29. C. Bandt, G. Keller, and B. Pompe, Entropy of interval maps via permutations. *Nonlinearity* **15** (2002) 1595–1602.
30. C. Bandt and F. Shiha, Order patterns in time series, *Journal of Time Series Analysis* **28** (2007) 646–665.
31. A. Berger, *Chaos and Chance*, Walter de Gruyter, Berlin, 2001.
32. G.D. Birkhoff, Proof of a recurrence theorem for strongly transitive systems, *Proceedings of the National Academy of Science* **17** (1931) 650.
33. F. Blanchard, P. Kurka, and A. Maass, Topological and measure-theoretical properties of one-dimensional cellular automata, *Physica D. Nonlinear Phenomena* **103** (1997) 86–99.
34. S. Boccaletti and D.L. Valladares, Characterization of intermittent lag synchronization, *Physical Review E* **62** (2000) 7497–7500.
35. L. Boltzmann, Über the mechanischen Analogien des zweiten Hauptsatzes der Thermodynamik, *Journal für reine und angewandte Mathematik* **100** (1887) 201.
36. R.E. Bowen, Entropy for group endomorphisms and homogeneous spaces, *Transactions of the American Mathematical Society* **153** (1971) 401–414.
37. A. Boyarsky and P. Gora, *Laws of Chaos*. Birkhäuser, Boston, 1997.
38. W.A. Brock, W.D. Dechert, J.A. Scheinkman, and B. LeBaron, A test for independence based on the correlation dimension, *Econometrics Reviews* **15** (1996) 197–235.
39. A.A. Brudno, Entropy and complexity of trajectories of a dynamical system. *Transactions of the Moscow Mathematical Society* **44** (1983) 127–151.
40. M. Buhl and M.B. Kennel, Statistically relaxing to generating partitions for observed time-series data, *Physical Review E* **71** (2005) 046213: 1–14.
41. J. Bunge and M. Fitzpatrick, Estimating the Number of Species: A Review, *Journal of the American Statistical Association* **88** (1993) 364–373.
42. L.A. Bunimovich and Y.G. Sinai, Space-time chaos in coupled map lattices, *Nonlinearity* **1** (1988) 491–518.
43. L.A. Bunimovich and Y.G. Sinai, Statistical mechanics of coupled map lattices, In: K.Kaneko (Ed.), *Theory and Applications of Coupled Map Lattices*. Wiley, New York, 1993.
44. L.A. Bunimovich, Coupled map lattices: Some topological and ergodic properties, *Physica D. Nonlinear Phenomena* **103** (1997) 1–17.
45. Y. Cao, W. Tung, J.B. Gao, V.A. Protopopescu, and L.M. Hively, Detecting dynamical changes in time series using the permutation entropy, *Physical Review E* **70** (2004) 046217.
46. R. Carretero-González, Low dimensional travelling interfaces in coupled map lattices, *International Journal of Bifurcations and Chaos* **7** (1997) 2745–2754.

47. R. Carretero-González, D.K. Arrowsmith, and F. Vivaldi, One-dimensional dynamics for traveling fronts in coupled map lattices, *Physical Review E* **61** (2000) 1329–1336.
48. K. Cattell and J.C. Muzio, Synthesis of one-dimensional linear hybrid cellular automata, *IEEE Transactions on Computer-Aided Design of Integrated circuits and Systems* **15** (1996) 325–335.
49. A. Chao, Nonparametric estimation of the number of classes in a population, *Scandinavian Journal of Statistics, Theory and Applications* **9** (1984) 265–270.
50. H. Chaté and P. Manneville, Coupled map lattices as cellular automata, *Journal of Statistical Physics* **56** (1989) 357–370.
51. B.V. Chirikov and F. Vivaldi, An algorithmic view of pseudochaos, *Physica D. Nonlinear Phenomena* **129** (1999) 223–235.
52. G.H. Choe, *Computational Ergodic Theory*. Springer Verlag, Berlin, 2005.
53. F. Christiansen and A. Politi, Generating partition for the standard map, *Physical Review E* **51** (1995) R3811.
54. L.O. Chua, V.I. Sbitnev, and S. Yoon, A nonlinear dynamics perspective of Wolfram's New Kind of Science –Part II: Universal neuron, *International Journal of Bifurcation and Chaos* **13** (2003) 2377–2491.
55. L.O. Chua, V.I. Sbitnev, and S. Yoon, A nonlinear dynamics perspective of Wolfram's new kind of science –Part IV: From Bernoulli shift to $1/f$ spectrum, *International Journal of Bifurcation and Chaos* **15** (2005) 1045–1183.
56. R.W. Clarke, M.P. Freeman, and N.W. Watkins, Application of computational mechanics to the analysis of natural data: An example in geomagnetism, *Physical Review E* **67** (2003) 016203.
57. P. Collet and J.P. Eckmann, *Iterated Maps on the Interval as Dynamical Systems*, 5th printing. Birkhäuser, Boston, 1997.
58. M. Courbage, D. Mercier, and S. Yasmineh, Traveling waves and chaotic properties in cellular automata, *Chaos* **9** (1999) 893–901.
59. T.M. Cover and J.A. Thomas, *Elements of Information Theory*, 2nd edition. New York, John Wiley & Sons, 2006.
60. J.P. Crutchfield and K. Young, Inferring statistical complexity, *Physical Review Letters* **63** (1989) 105–108.
61. R. Dahlhaus, J. Kurths, P. Maass, and J. Timmer, *Mathematical Methods in Time Series Analysis and Digital Image Processing*. Springer Verlag, Berlin, 2008.
62. M. D'amico, G. Manzini, and L. Margara, On computing the entropy of cellular automata, *Theoretical Computer Science* **290** (2003) 1629–1646.
63. R. Davidchack, Y.C. Lai, E.M. Bollt, and M. Dhamala, Estimating generating partitions by unstable periodic orbits, *Physical Review E* **61** (2000) 1353–1356.
64. K. Denbigh, How subjective is entropy. In: H.S. Leff and A.F. Rex (Ed.), *Maxwell's Demon, Entropy, Information, Computing*, pp. 109–115. Princeton University Press, Princeton,1990.
65. M. Denker, Finite generators for ergodic, measure-preserving transformations, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **29** (1974) 45–55.
66. M. Denker, C. Grillenberger, and K. Sigmund, *Ergodic Theory on Compact Spaces*. Springer Lecture Notes in Math. **527**, Springer Verlag, Berlin, 1976.
67. M. Denker and W.A. Woyczynski, *Introductory Statistics and Random Phenomena*. Birkhäuser, Boston, 1998.
68. M. Denker, *Einführung in die Analysis Dynamischer Systeme*. Springer Verlag, Berlin, 2005.
69. R.L. Devaney, *Chaotic Dynamical Systems* (2nd edition). Westview Press, Boulder, 2003.
70. E.I. Dinaburg, The relation between topological entropy and metric entropy, *Soviet Mathematics* **11** (1970) 13–16.
71. Y. Dobyns and H. Atmanspacher, Characterizing spontaneous irregular behavior in coupled map lattices, *Chaos, Solitons & Fractals* **24** (2005) 313–327.
72. J.P. Eckmann and D. Ruelle, Ergodic theory of chaos and strange attractors, *Review of Modern Phys*ics **57** (1985) 617–656.

73. J.P. Eckmann, S.O. Kamphorst, and D. Ruelle, Recurrence plots of dynamical systems, *Europhysics Letters* **4** (1987) 973–977.
74. S. Elizalde and M. Noy, Consecutive patterns in permutations, *Advances in Applied Mathematics* **30** (2003) 110–125.
75. S. Elizalde, Asymptotic enumeration of permutations avoiding generalized patterns, *Advances in Applied Mathematics* **36** (2006) 138–155.
76. S. Elizalde, The number of permutations realized by a shift, *SIAM Journal of Discrete Mathematics* **23** (2009) 765–786.
77. R. Érdi, *Complexity Explained*. Springer Verlag, Berlin, 2007.
78. A. Fernández, J. Quintero, R. Hornero, P. Zuluaga, M. Navas, C. Gómez, J. Escudero, N. García-Campos, J. Biederman, and T. Ortiz, Complexity analysis of spontaneous brain activity in attention-deficit/hyperactivity disorder: Diagnosis implications, *Biological Psychiatry* **65** (2009) 571–577.
79. J. Ford, G. Mantica, and G.H. Ristow, The Arnold's cat: Failure of the correspondence principle, *Physica D. Nonlinear Phenomena* **50** (1991) 493–520.
80. A.M. Fraser and H.L. Swinney, Independent coordinates for strange attractors from mutual information, *Physical Review A* **33** (1986) 1134–1140.
81. J.B. Gao and H.Q. Cai, On the structures and quantification of recurrence plots, *Physics Letters A* **270** (2000) 75–87.
82. Y. Gao, I. Kontoyiannis, and E. Bienenstock, Estimating the entropy of binary time series: Methodology, some theory and a simulation study, *Entropy* **10** (2008) 71–99.
83. J. García-Ojalvo, J.M. Sancho, and L. Ramírez-Piscina, Generation of spatiotemporal colored noise, *Physical Review A* **46** (1992) 4670–4675.
84. M. Gardner, The fantastic combinations of John Conway's new solitaire game "life", *Scientific American* **223** (1970) 120–123.
85. A. Golestani, M.R. Jahed Motlagh, K. Ahmadian, A.H. Omidvarnia, and N. Mozayani, A new criterion to distinguish stochastic and deterministic time series with the Poincaré section and fractal dimension, *Chaos* **19** (2009) 013137.
86. S.W. Golomb, *Bulletin of the American Mathematical Society* **70** (1964) 747 (research problem 11).
87. P. Grassberger and H. Kantz, Generating partitions for the dissipative Hénon map, *Physics Letters A* **113** (1985) 235–238.
88. P. Grassberger, Finite sample corrections to entropy and dimension estimates, *Physics Letters A* **128** (1988) 369–373.
89. R.M. Gray, *Entropy and Information Theory*. Springer Verlag, New York, 1990.
90. F. Gu, X. Meng, E. Shen, and Z. Cai, Can we measure consciousness with EEG complexities?, *International Journal of Bifurcations and Chaos* **13** (2003) 733–742.
91. B. Hasselbaltt and A. Katok, *A First Course in Dynamics*. Cambridge University Press, Cambridge, 2003.
92. G.A. Hedlund, Endomorphisms and automorphisms of the shift dynamical system, *Mathematical Systems Theory* **3** (1969) 320–375.
93. H. Herzel, Complexity of symbol sequences, *Systems, Analysis, Modelling, Simulations* **5** (1988) 435–444.
94. H. Herzel, A.O. Schmitt, and W. Ebeling, Finite sample effects in sequence analysis, *Chaos, Solitons & Fractals* **4** (1994) 97–113.
95. E. Hewitt and K. Stromberg, *Real and Abstract Analysis*. Springer Verlag, New York 1965.
96. F.C. Hoppensteadt, *Analysis and Simulation of Chaotic Systems* (2nd edition). Springer Verlag, New York, 2000.
97. K. Hiraide, Nonexistence of positively expansive maps on compact connected manifolds with boundary, *Proceedings of the American Mathematical Society* **110** (1990) 565–568.
98. M.W. Hirsch, S. Smale, and R.L. Devaney, *Differential Equations, Dynamical Systems, and an Introduction to Chaos*. Academic Press, San Diego, 2003.

99. C.S. Hsu and M.C. Kim, Construction of maps with generating partitions for entropy evaluation, *Physical Review A* **31** (1985) 3253–3265.
100. J. Hughes, J. Hellman, T.H. Rickets, and B.J.M. Bohannan, Counting the uncountable: Statistical approaches to estimating microbial diversity, *Applied and Environ. Microbiology* **67** (2001) 4399–4406.
101. L.P. Hurd, J. Kari, and K. Culik, The topological entropy of cellular automata is uncomputable, *Ergodic Theory and Dynamical Systems* **12** (1992) 255–265.
102. Y. Ishii and D. Sands, Monotonicity of the Lozi family near the tent-maps, *Communications in Mathematical Physics* **198** (1998) 397–406.
103. N. Israeli and N. Goldenfeld, Coarse-graining of cellular automata, emergence, and the predictability of complex systems, *Physical Review E* **73** (2006) 1–17.
104. S. Jalan, J. Jost, and F.M. Atay, Symbolic synchronization and the detection of global properties of coupled dynamics from local information, *Chaos* **16** (2006) 033124.
105. O. Jenkinson and M. Pollicott, Entropy, exponents and invariant densities for hyperbolic systems: Dependence and computation. In: M. Brin, B. Hasselblatt, and Y. Pesin (Eds.), *Modern Dynamical Systems and Applications*. pp. 365–384 Cambridge University Press, Cambridge, 2004.
106. K. Kaneko, Transition from torus to chaos accompanied by frequency lockings with symmetry breaking, *Progress in Theoretical Physics* **69** (1983) 1427–1442.
107. K. Kaneko, Period-doubling of kink-antikink patterns, quasiperiodicity in anti-ferro-like structures and spatial intermittency in coupled logistic lattice, *Progress in Theoretical Physics* **72** (1984) 480–486.
108. K. Kaneko, Pattern dynamics in spatiotemporal chaos, *Physica D. Nonlinear Phenomena* **34** (1989) 1–41.
109. K. Kaneko, Spatiotemporal chaos in one- and two-dimensional coupled map lattices, *Physica D. Nonlinear Phenomena* **37** (189) 60–82.
110. K. Kaneko, Chaotic traveling waves in a coupled map lattice, *Physica D. Nonlinear Phenomena* **68** (1993) 299–317.
111. H. Kantz, Quantifying the closeness of fractal measures, *Physical Review E* **49** (1994) 5091–5097.
112. H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*. Cambridge University Press, Cambridge, 1997.
113. N.J. Kasdin, Discrete simulation of colored noise and stochastic processes and $1/f^{\alpha}$ power law noise generation, *Proceedings of the IEEE* **83** (1995) 802–827.
114. A. Katok and B. Hasselbaltt, *Introduction to the Theory of Dynamical Systems*. Cambridge University Press, Cambridge, 1998.
115. S. Katok, *p-adic Analysis compared with real*. American Mathematical Society, Providence, 2007.
116. K. Keller and K. Wittfeld, Distances of time series components by means of symbolic dynamics, *International Journal of Bifurcation and Chaos* **14** (2004) 693–703.
117. K. Keller and M. Sinn, Ordinal analysis of time series, *Physica A* **356** (2005) 114–120.
118. K. Keller, H. Lauffer, and M. Sinn, Ordinal analysis of EEG time series, *Chaos and Complexity Letters* **2** (2007) 247–258.
119. M.B. Kennel and S. Isabelle, Method to distinguish possible chaos from colored noise and to determine embedding parameters, *Physical Review A* **46** (1992) 3111–3118.
120. M.B. Kennel, Statistical test for dynamical nonstationarity in observed time-series data, *Physical Review E* **56** (1997) 316–321.
121. M.B. Kennel and A.I. Mees, Context-tree modeling of observed symbolic dynamics, *Physical Review E* **66** (2002) 056209.
122. M.B. Kennel, J. Shlens, H.D.I. Abarbanel, and E.J. Chichilnisky, Estimating entropy rates with Bayesian confidence intervals, *Neural Computation* **17** (2005) 1531–1576.
123. B.P. Kitchens, *Symbolic Dynamics*. Springer Verlag, Berlin, 1998.

124. L. Kocarev and J. Szczepanski, Finite-space Lyapunov exponents and pseudo-chaos, *Physical Review Letters* **93** (2004) 234101.
125. L. Kocarev, J. Szczepanski, J.M. Amigó, and I. Tomovski, Discrete Chaos – Part I: Theory, *IEEE Transactions on Circuits and Systems I* **53** (2006) 1300–1309.
126. A.N. Kolmogorov, Entropy per unit time as a metric invariant of automorphism, *Doklady of Russian Academy of Sciences* **124** (1959) 754–755.
127. I. Kontoyiannis, P.H. Algoet, Y.M. Suhov, and A.J. Wyner, Nonparametric entropy estimation for stationary processes and random fields, with applications to English text. *IEEE Transactions on Information Theory* **44** (1998) 1319–1327.
128. Z.S. Kowalski, Finite generators of ergodic endomorphisms, *Colloquium Mathematicum* **49** (1984) 87–89.
129. Z.S. Kowalski, Minimal generators for aperiodic endomorphisms, *Commentationes Mathematicae Universitatis Carolinae* **36** (1995) 721–725.
130. W. Krieger, On entropy and generators of measure-preserving transformations, *Transactions of the American Mathematical Society* **149** (1970) 453–464.
131. A.P. Kurian and S. Puttusserypady, Self-synchronizing chaotic stream ciphers, *Signal Processing* **88** (2008) 2442–2452.
132. J. Kurths, D. Maraun, C.S. Zhou, G. Zamora-López, and Y. Zou, Dynamics in complex systems, *European Review* **17** (2009), 357–370.
133. J.C. Lagarias, Pseudorandom numbers, *Statistical Science* **8** (1993) 31–39.
134. A. Lasota and J.A. Yorke, On the existence of invariant measures for piecewise monotonic transformations, *Transactions of the American Mathematical Society* **186** (1973), 481–488.
135. A.M. Law and W.D. Kelton, *Simulation, Modeling, and Analysis*, 3rd edition. McGraw-Hill, Boston, 2000.
136. B. LeBaron, A fast algorithm for the BDS statistics, *Studies in Nonlinear Dynamics & Econometrics* **2** (1997) 53–59.
137. A. Lempel and J. Ziv, On the complexity of an individual sequence, *IEEE Transactions on Information Theory* **IT-22** (1976) 75–78.
138. M. Li and P. Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications*. Springer Verlag, New York, 1997.
139. D. Lind and B. Marcus, *Symbolic Dynamics and Coding*. Cambridge University Press, Cambridge, 2003.
140. T. Liu, C.W.J. Granger, and W.P. Heller, Using the correlation exponent to decide whether an economic series is chaotic. *Journal of Applied Econometrics,* Supplement: Special Issue on Nonlinear Dynamics and Econometrics (Dec., 1992) S25–S39.
141. G. Manzini and L. Margara, A complete and efficiently computable topological classification of linear cellular automata over $\mathbb{Z}_m$, *Theoretical Computer Science* **221** (1999) 157–177.
142. R. Mañé, *Ergodic Theory and Differentiable Dynamics*. Springer Verlag, Berlin, 1987.
143. M.T. Martin, A. Plastino, and O.A. Rosso, Generalized statistical complexity measures: Geometrical and analytical properties, *Physica A* **369** (2006) 439–462.
144. N. Marwan, M.C. Romano, M. Thiel, and J. Kurths, Recurrence plots for the analysis of complex systems, *Physics Reports* **438** (2007) 237–329.
145. C. Masoller and A.C. Martí, Random delays and the synchronization of chaotic maps, *Physical Review Letters* **94** (2005) 134102.
146. M. Matilla-García, A non-parametric test for independence based on symbolic dynamics, *Journal of Economic Dynamic & Control* **31** (2007) 3889–3903.
147. M. Matilla-García and M. Ruiz Marín, A non-parametric independence test using permutation entropy, *Journal of Econometrics* **144** (2008) 139–155.
148. M. Matsumoto and T. Nishimura, Mersenne Twister: A 623-dimensionally equidistributed uniform pseudo-random number generator, *ACM Trans. on Modeling and Computer Simulation* **8** (1998) 3–30.
149. W. Meier and O. Staffelbach, The self-shrinking generator. In: Proc. of Eurocrypt'94, Lecture Notes in Computer Science. vol. 950, pp. 205–214. Springer Verlag, Berlin, 1994.

150. W. de Melo and S. van Strien, *One-Dimensional Dynamics*. Springer Verlag, Berlin, 1993.
151. A.J. Menezes, P.C. van Oorschoot, and S.A. Vanstone, *Handbook of Applied Cryptography*. CRC Press, Boca Raton, 1997.
152. M.E. Mera and M. Morán, Geometric noise reduction for multivariate time series, *Chaos* **16** (2006) 013116.
153. N. Metropolis, M. Stein, and P. Stein, On finite limit sets for transformations on the unit interval, *Journal of Combinatorial Theory, Series A* **15**, 25–44 (1973).
154. J. Milnor, Non-expansive Hénon Maps, *Advances in Mathematics* **69** (1988) 109–114.
155. M. Misiurewicz and W. Szlenk, Entropy of piecewise monotone mappings, *Studia Mathematica* **67** (1980) 45–63.
156. M. Misiurewicz, Strange attractors for the Lozi mappings. In: R.G. Helleman (Ed.), *Nonlinear Dynamics*, Vol. **357**, pp. 348–358 The New York Academy of Science, New York 1980.
157. M. Misiurewicz, Permutations and topological entropy for interval maps, *Nonlinearity* **16** (2003) 971–976.
158. M. Mitchell, *Complexity —A Guided Tour*. Oxford University Press, New York, 2009.
159. R. Monetti, W. Bunk, T. Aschenbrenner, and F. Jamitzky, Characterizing synchronization in time series using information measures extracted from symbolic representations, *Physical Review E* **79** (2009) 046207.
160. M. Morse and G.A. Hedlund, Symbolic Dynamics, *American Journal of Mathematics* **60** (1938) 815–866.
161. J. von Neumann, The general and logical theory of automata. In: L.A. Jeffress (Ed.), *Cerebral Mechanisms in Behavior*. Wiley, New York, 1951.
162. M. Newman, A.L. Barabási, and D.J. Watts, *The Structure and Dynamics of Networks*. Princeton University Press, Princeton, 2006.
163. E. Olbrich, N. Bertschinger, N. Ay, and J. Jost, How should complexity scale with system size?, *The European Physical Journal B* **63** (2008) 407–415.
164. G.J. Ortega and E. Louis, Smoothness implies determinism in time series: A measure based approach, *Physical Review Letters* **81** (1998) 4345–4348.
165. E. Ott, *Chaos in Dynamical Systems*. Cambridge University Press, Cambridge, 2002.
166. N.H. Packard, J.P. Crutchfield, J.D. Farmer, and R.S. Shaw, Geometry from a time series, *Physical Review Letters* **45** (1980) 712–716.
167. L. Paninski, Estimation of entropy and mutual information, *Neural Computation* **15** (2003) 1191–1253.
168. H.O. Peitgen, H. Jürgens, and D. Saupe, *Chaos and Fractals*. Springer Verlag, New York, 2004.
169. K. Petersen, *Ergodic Theory*. Cambridge University Press, Cambridge, 1983.
170. S.D. Pethel, N.J. Corron, and E. Bollt, Symbolic dynamics of coupled map lattices, *Physical Review Letters* **96** (2006) 034105.
171. S.D. Pethel, N.J. Corron, and E. Bollt, Deconstructing spatiotemporal chaos using local symbolic dynamics, *Physical Review Letters* **99** (2007) 214101.
172. J. Piepzryk, T. Hardjorno, and J. Seberry, *Fundamentals of Computer Security*. Springer Verlag, Berlin, 2003.
173. W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge, 2007.
174. R.C. Robinson, *An Introduction to Dynamical Systems*. Pearson Prentice Hall, Upper Saddle River NJ, 2004.
175. M.G. Rosenblum, A.S. Pikovsky, and J. Kurths, Phase synchronization of chaotic oscillators, *Physical Review Letters* **76** (1997) 1804–1807.
176. O.A. Rosso, H.A. Larrondo, M.T. Martin, A. Platino, and M.A. Fuentes, Distinguishing noise from chaos, *Physical Review Letters* **99** (2007) 154102.
177. D.J. Rudolph, *Fundamentals of Measurable Dynamics*. Oxford University Press, Oxford, 1990.
178. `http://topo.math.u-psud.fr/ sands/Programs/Lozi/index.html.`

179. A.N. Sarkovskii, Coexistence of cycles of a continuous map of a line into itself, *Ukrainian Mathematical Journal* **16** (1964) 61–71.
180. P.R. Scalassara, M.E. Dajer, C. Dias Maciel, C. Capobianco Guido, and J.C. Pereira, Relative entropy measures applied to healthy and pathological voice characterization, *Applied Mathematics and Computation* **207** (2009) 95–108.
181. A.O. Schmitt, H. Herzel, and W. Ebeling, A new method to calculate higher-order entropies from finite samples, *Europhysics Letters* **23** (1993) 303–309.
182. O. Schmitt, *Remarks on the Generator-Problem* (Thesis). University of Göttingen, 2001.
183. T. Schreiber, Detecting and analyzing nonstationarity in a time series using nonlinear cross predictions, *Physical Review Letters* **78** (1997) 843–846.
184. R. Sexl and J. Blackmore (Eds.), *Ludwig Boltzmann - Ausgewahlte Abhandlungen* (Ludwig Boltzmann Gesamtausgabe, Band 8). Vieweg, Braunschweig, 1982.
185. C.R. Shalizi and J.P. Crutchfield, Computational mechanics: Pattern and prediction, structure and simplicity, *Journal of Statistical Physics* **104** (2001) 817–879.
186. C.E. Shannon, A mathematical theory of communication, *Bell System Technical Journal* **27** (1948) 379–423, 623–653.
187. L.A. Shepp and S.P. Lloyd, Ordered cycle length in a random permutation, *Transactions of the American Mathematical Society* **121** (1966) 340–357.
188. M.A. Shereshevsky, Expansiveness, entropy and polynomial growth for groups acting on subshifts by automorphisms. *Indagationes Mathematicae* **4** (1993) 203–210.
189. Y.G. Sinai, On the Notion of Entropy of a Dynamical System, *Doklady of Russian Academy of Sciences* **124** (1959) 768–771.
190. M. Sinn and K. Keller, Estimation of ordinal pattern probabilities in fractional Brownian motion, arXiv:0801.1598.
191. M. Smorodinsky, *Ergodic Theory, Entropy* (Lectures Notes in Mathematics) Vol. **214**. Springer Verlag, Berlin, 1971.
192. D. Sotelo Herrera and J. San Martín, Analytical solutions of weakly coupled map lattices using recurrence relations, *Physics Letters A* **373** (2009) 2704–2709.
193. J.C. Sprott, *Chaos and Time-Series Analysis*. Oxford University Press, Oxford, 2003.
194. J.C. Sprott, High-dimensional dynamics in the delayed Hénon map. *Electronic Journal of Theoretical Physics* **3** (2006) 19–35.
195. S.P. Strong, R. Koberle, R.R. de Ruyter van Steveninck, and W. Bialek, Entropy and information in neural spike trains. *Physical Review Letters* **80** (1998) 197–200.
196. J. Szczepanski, J.M. Amigó, E. Wajnryb, and M.V. Sanchez-Vives. Application of Lempel-Ziv complexity to the analysis of neural discharges, *Network: Computation in Neural Systems* **14** (2003) 335–350.
197. F. Takens, Detecting strange attractors in turbulence, In: D. Rand and L.S. Young (Eds.), *Dynamical Systems and Turbulence*, Lecture Notes in Mathematics, vol. 898. Springer, Berlin, 1981, pp. 366–381.
198. T. Toffoli and N. Margolus, *Cellular Automata Machines*. The MIT Press, Cambridge MA, 1987.
199. S. Ulam, Random process and transformations, Proceedings of the International Congress of Mathematicians 2 (1952), 264–275.
200. D.B. Vasconcelos, S.R. Lopes, R.L. Viana, and J. Kurths, Spatial recurrence plots, *Physical Review E* **73** (2006) 056207.
201. S.B. Volchan, What is a Random Sequence, *The American Mathematical Monthly* **109** (2002) 46–63.
202. P. Walters, *An Introduction to Ergodic Theory*. Springer Verlag, New York, 2000.
203. L. Wang and N.D. Kazarinoff, On the universal sequence generated by a class of unimodal functions, J*ournal of Combinatorial Theory, Series A* **46** (1987) 39–49.
204. A. Wolf, J.B. Swift, H.L. Swinney, and J.A. Vastano, Determining Lyapunov exponents from a time series, *Physica D. Nonlinear Phenomena* **16** (1985) 285–317.

205. S. Wolfram, Computation theory of cellular automata, *Communications in Mathematical Physics* **96** (1984) 15–57.
206. S. Wolfram, Universality and complexity in cellular automata, *Physica* **10D** (1984) 1–35.
207. S. Wolfram, *A New Kind of Science*. Wolfram Media, Champaign, 2002.
208. X-S. Zhang, R.J. Roy, and E.W. Jensen, EEG complexity as a measure of depth anesthesia for patients, *IEEE Transactions on Biomedical Engineering* **48** (2001) 1424–1433.
209. J. Zhang and M. Small, Complex networks from pseudoperiodic time series: Topology versus dynamics. *Physical Review Letters* **96** (2006) 238701.
210. G.C. Zhuang, J. Wang, Y. Shi, and W. Wang, Phase synchronization and its cluster feature in two-dimensional coupled map lattices, *Physical Review E* **66** (2002) 046201.
211. J. Ziv and A. Lempel, Compression of individual sequences via variable-rate coding *IEEE Transactions on Information Theory* **IT-24** (1978) 530–536.
212. L. Zunino, D.G. Pérez, M.T. Martín, M. Garavaglia, A. Plastino, and O.A. Rosso, Permutation entropy of fractional Brownian motion and fractional Gaussian noise, *Physics Letters A* **372** (2008) 4768–4774.
213. L. Zunino, D.G. Pérez, M.T. Martín, M. Garavaglia, A. Plastino, and O.A. Rosso, Fractional Brownian motion, fractional Gaussian noise, and Tsallis permutation entropy, *Physica A* **387** (2008) 6057–6068.

Index