Chapter 1
What Is This All About?

This introductory chapter is meant as a tour of the main topics in this book: patterns, ordinal relations, complexity, and entropy. The approach is mostly informal; for the technicalities behind the different notions met on the way, the reader is referred to Annex A and Annex B.

1.1 Patterns, Complexity, and Entropy

Pattern is an abstract concept with different acceptations. In the context of dynamical systems, information theory, and computer science (the ones we are interested in), a pattern is a finite string of symbols, eventually chosen with some criterion. In the next sections we will meet some familiar instances of patterns in those contexts. Contrary to the concept of pattern, complexity does not lend itself to a short definition (would this be not a contradiction otherwise?) but, like poetry, it is very easy to recognize. For a panorama of complexity, see [77] or, at an introductory level, [158]. A third and also recurrent issue in the next pages will be entropy, one of the most important quantities when dealing with complexity in deterministic and random dynamical systems. Indeed, no matter how one counts the diversity of patterns generated by a data source, entropy enters the scene in some of its many disguises: Shannon entropy, metric entropy, topological entropy, etc.

1.1.1 Information Theory

Consider an information source outputting symbols or letters, one at a time, from a finite alphabet $S = \{s_1, \ldots, s_{|S|}\}$ (i.e., $|S|$ is the cardinality of S). Formally, an information source is a discrete-time, stationary stochastic process $\mathbf{X} = \{X_n\}_{n \in \mathbb{N}_0}$, where $\mathbb{N}_0 = \{0, 1, \ldots\}$ and X_n are random variables on a common probability space, taking on values in S. For the time being, we will dispense with the underlying probability space. A realization of $\mathbf{X}$ is a one-sided sequence, $x_0^\infty := (x_n)_{n \in \mathbb{N}_0}$, called[1] a

[1] The symbol ":=" means that the left side is defined by the right one; a corresponding meaning holds for "=:".

J.M. Amigó, *Permutation Complexity in Dynamical Systems,*
Springer Series in Synergetics, DOI 10.1007/978-3-642-04084-9_1,

message. Correspondingly, the symbols $x_n \in S$ are sometimes called *letters*. A finite segment of a message, say, $x_k^{k+L-1} := x_k x_{k+1} \dots x_{k+L-1}$ is called a *word* of length L. If $p(x_0^{L-1})$ denotes the probability of the word x_0^{L-1} to be output, then the (Shannon) entropy rate (or just *entropy*) of the data source $\mathbf{X}$ is defined as

$$h(\mathbf{X}) = -\lim_{L\to\infty} \frac{1}{L} \sum p(x_0^{L-1}) \log p(x_0^{L-1}), \tag{1.1}$$

where log usually stands for logarithm to base 2 ($h(\mathbf{X})$ is then measured in bits per symbol), and the sum is over all possible words of length L, numbering $|S|^L$, with the convention $0 \times \log 0 = \lim_{x\to 0+} x \log x = 0$. To indicate that a logarithm is to base e, we will write ln instead of log ($h(\mathbf{X})$ is then measured in nats per symbol). The convergence of limit (1.1) is proven in Sect. B.1.2.

In an information-theoretical setting, $\log p(x_0^{L-1})$ is the information conveyed by the output x_0^{L-1}, hence $h(\mathbf{X})$ is the average information per symbol conveyed by the messages of the information source $\mathbf{X}$ in the limit of arbitrarily long messages.

When the random variables X_n are independent, or (more often) intersymbol dependency is neglected for simplicity or limited influence, the information source is called *memoryless*. In this case $h(\mathbf{X})$ coincides with the entropy $H(X)$ of a random variable X with outcomes $x \in S$ and probabilities $p(x)$:

$$H(X) = -\sum_{x\in S} p(x) \log p(x).$$

Compression is any procedure that reduces the data requirements of a message without, in principle, losing information—although it can be acceptable as a trade-off between data reduction and information degradation. The idea of using codes or dictionaries for compression of information originates with the invention of the telegraph, since users were charged by the number of letters in the message. It is clear that data compression can be achieved by assigning short words to the most frequent outcomes of the information source. For example, in the Morse code, the most frequent symbol in English, namely the letter e, is represented by a single dot. This intuition is the guiding principle in the construction of the celebrated Huffman code for memoryless sources. Suppose that code words $w_1, \dots, w_{|S|}$ of lengths $l_1, \dots, l_{|S|}$, respectively, are assigned to the values $s_1, \dots, s_{|S|}$ taken on by a random variable X with probabilities $p(s_1), \dots, p(s_{|S|})$. The code words are combinations of characters taken from an alphabet $a_1, \dots, a_D$, usually 0, 1 ($D = 2$) in modern communications. Then the Huffman code is a uniquely decipherable code that minimizes the average code-word length $\bar{l} = \sum_{n=1}^{|S|} p(s_n) l_n$, which according to the *noiseless coding theorem* is known to satisfy [22]

$$H(X) \le \bar{l} < H(X) + 1, \tag{1.2}$$

where the logarithm of $H(X)$ is taken to base D. But how to compress a message, say a digital picture to be sent by electronic mail or a text file written in a foreign language, if the probabilities of the corresponding symbols are not known? This feat requires a universal compressor.

Universal compressors are based on the fact that natural languages are not completely random but repeat patterns from time to time. In 1976 and 1978, A. Lempel and J. Ziv published two simple algorithms for universal data compression [137, 211], which work by parsing an input string of finite length into successive phrases. Some variants of the second (LZ78) are implemented in the most popular compressors currently used in electronic editing (like WinZip or pdf). For our purposes it is sufficient to consider the first scheme (LZ76); also, we will emphasize the interplay between complexity and entropy rather than the compression-related aspects.

In the LZ76, the message is sequentially parsed into strings that have not appeared so far in the initial segment ending at (and excluding) the current letter. For example, the binary word $x_0^{19} = 01011010001101110010$ is parsed as

$$0, 1, 011, 0100, 011011, 1001, 0. \tag{1.3}$$

If, say, x_k is the first bit after a comma, then we check whether x_k appears in x_0^{k-1}. If it does not, then we write a comma after x_k and start a new block (this is the case for $k = 1$ in (1.3)). Otherwise, we check whether $x_k x_{k+1}$ appears in x_0^k; in negative case, we write a comma after x_{k+1}, otherwise the process continues till a pattern $x_k x_{k+1} \ldots x_{k+l}$ repeats (or the sequence finishes). The number of patterns found in the parsing of a word x_0^{L-1} is called its Lempel–Ziv (LZ) complexity, $C(x_0^{L-1})$. In example (1.3), $C(x_0^{19}) = 7$. Words x_0^{L-1} with a general alphabet S are parsed in an analogous way.

The formal definition of $C(x_0^{L-1})$ is recursive. A *block* of length l ($1 \le l \le L$) is just a segment of x_0^{L-1} of length l, i.e., a string of l consecutive letters, say $x_k^{k+l-1} = x_k x_{k+1} \ldots x_{k+l-1}$ ($0 \le k \le L - l$). In particular, letters are blocks of length 1. Set $B_0 = x_0$ and suppose that after $k \ge 1$ steps, we have parsed x_0^{L-1} as

$$B_0, B_1, \ldots, B_{k-1},$$

where $B_1 = x_1^{n_1}, \ldots, B_{k-1} = x_{n_{k-2}+1}^{n_{k-1}}$, and $n_{i-1} + 1 \le n_i < L - 1$ for $i = 1, \ldots, k-1$ (with $n_0 = 0$). Define

$$B_k := x_{n_{k-1}+1}^{n_k} \quad (n_{k-1} + 1 \le n_k \le L - 1),$$

to be the shortest block such that it does not occur in the sequence $x_0^{n_k - 1}$. (In the LZ78 algorithm, one checks instead whether the current block $x_{n_{k-1}+1}^{n_k}$ coincides with one of the previous blocks, $B_0, B_1, \ldots, B_{k-1}$.) Proceeding in this way, we obtain a (uniquely defined) decomposition of x_0^{L-1} in "minimal" blocks, say

$$x_0^{L-1} = B_0, B_1, \ldots, B_{p-1}, \tag{1.4}$$

in which only the last block can occasionally appear twice. Then,

$$C(x_0^{L-1}) := p.$$

For computational efficiency, one uses the well-known "suffix-tree" data structure and search algorithms for quickly finding substrings of the input string.

From the foregoing description, we may say that $C(x_0^{L-1})$ measures the complexity of the word x_0^{L-1}; words with a periodic or almost periodic structure have a small LZ complexity, while those displaying a random-looking structure have a high count of distinct patterns, hence a great LZ complexity. It can be proven [211] that if the source $\mathbf{X}$ is ergodic (i.e., the probability of any length-L word equals its frequency in a single, "typical" sequence), then

$$\limsup_{L\to\infty} \frac{C(x_0^{L-1})}{L/\log_{|S|} L} = h(\mathbf{X}) \tag{1.5}$$

with probability 1. The normalization factor in (1.5) is the LZ complexity of a memoryless, equidistributed source. Let us mention in passing that (1.5) shows that the ideal compression factor of the LZ76 algorithm, in the limit of long messages, is $h(\mathbf{X})$. The same is true for the LZ78 scheme.

Equations (1.2) and (1.5) provide examples in which the concepts of complexity (here related to "incompressibility") and entropy (here related to "uncertainty") are linked in a perhaps unexpected way. As a by-product, LZ complexity can be used as an estimator of the entropy. A principal advantage of this approach is that the LZ algorithm is entirely automatic with no free parameters (unlike naive plug-in methods or methods which estimate $h(\mathbf{X})$ via block entropies; see [167] and Sect. 2.1). Another practical issue is the convergence speed with L: the normalized LZ76 complexity converges to the entropy faster than the LZ78, what makes it a better choice in practice [6]. A variance estimator for the entropy estimation by means of the LZ76 complexity can be found in [9].

1.1.2 Symbolic Dynamics

Symbolic dynamics, first proposed by Morse and Hedlund [160], is an approach to complex dynamics that aims to capture the essential aspects of complexity by studying conceptually simple models. As it often happens in mathematics, symbolic dynamics has developed in short time from an auxiliary tool to an independent field [139, 123], with applications to the study of formal languages. As a result, dynamical systems connect through symbolic dynamics to computer science, information theory, and automata.

To motivate symbolic dynamics, consider the dynamics generated by a self-map f of a set Ω. Of course, the dynamics is introduced in the *state space* Ω via the repeated action of f on Ω. Given $x \in \Omega$, the *orbit* or *trajectory* of x under f is defined as $\mathcal{O}_f(x) = \{f^n(x) : n \in \mathbb{N}_0\}$, where $f^0(x) := x$ and $f^n(x) := f(f^{n-1}(x))$. If f is invertible, then one can distinguish between the *full orbit* $\mathcal{O}_f(x) = \{f^n(x) : n \in \mathbb{Z}\}$ and the *forward orbit* $\mathcal{O}_f^+(x) = \{f^n(x) : n \in \mathbb{N}_0\}$. The name "orbit" clearly hints to the interpretation of the iteration index n as discrete time: each application of f on the point $x_n = f(x_{n-1})$ updates the "movement" of the *initial condition* x in Ω. If the resulting dynamics is complicated, we might content ourselves with a "blurred" picture of the orbit behavior. This can be done as follows. Divide Ω into a finite number of disjoint pieces A_i, $i = 0, 1, \ldots, k-1$, and keep track of the trajectory of $x \in \Omega$ with the precision set by the decomposition $\alpha = \{A_0, \ldots, A_{k-1}\}$. (We reserve the name partition for a measurable decomposition, provided Ω is endowed with a sigma algebra; see below.) Specifically, we assign to x a (one-sided) sequence[2] $\Phi(x) = (\xi_0, \xi_1, \ldots, \xi_n, \ldots)$, the nth entry $\xi_n \in \{0, 1, \ldots, k-1\}$ telling us in which element of α the iterate $f^n(x)$ is to be found. When f is invertible, we can also assign a two-sided sequence $\Phi(x) = (\ldots, \xi_{-1}, \xi_0, \xi_1, \ldots, \xi_n, \ldots)$, the entries with negative indices corresponding to the locations of $f^{-n}(x)$, $n \geq 1$. For brevity we focus on the general case. We call Φ a *coding map*, and $\Phi(x)$ the *itinerary* of x with respect to the decomposition α. Formally,

$$\Phi_n(x) = i \text{ iff } f^n(x) \in A_i, \tag{1.6}$$

where $n \in \mathbb{N}_0$ and $\Phi_n(x)$ denotes the nth component of the sequence $\Phi(x)$.

Let us reformulate this simple idea in a more general way. Given the finite alphabet $S = \{0, 1, \ldots, k-1\}$, denote by $S^{\mathbb{N}_0}$ the space of one-sided sequences of symbols from S:

$$S^{\mathbb{N}_0} = \{(\xi_n)_{n\in\mathbb{N}_0} = (\xi_0, \xi_1, \ldots, \xi_n, \ldots): \xi_n \in S\}.$$

Hence, $\Phi(x) \in S^{\mathbb{N}_0}$. The space $S^{\mathbb{N}_0}$ (and also $S^{\mathbb{Z}}$) is generically referred to as a *sequence* or *symbolic space*. One can put on a sequence space different (non-equivalent) metrics d making it a compact space. For example,

$$d((\xi_n)_{n\in\mathbb{N}_0}, (\eta_n)_{n\in\mathbb{N}_0}) = \begin{cases} 0 & \text{if } \xi_n = \eta_n \text{ for all } n \in \mathbb{N}_0, \\ 2^{-N} & \text{if } \xi_n = \eta_n \text{ for } n < N \text{ and } \xi_N \neq \eta_N. \end{cases} \tag{1.7}$$

Thus, two one-sided sequences are apart 2^{-N} in this metric if their first N entries coincide (and the $(N+1)$th ones do not). In $S^{\mathbb{Z}}$, two sequences $(\xi_n)_{n\in\mathbb{Z}}$ and $(\eta_n)_{n\in\mathbb{Z}}$ are at distance 2^{-N} if their entries coincide from $-(N-1)$ to $N-1$, i.e., if $\xi_n = \eta_n$ for $|n| < N$. In Annex A.2 we consider other metrics.

Having introduced the sequence spaces, observe now that the action of f on the orbit of $x \in \Omega$, namely, $f^n(x) \mapsto f(f^n(x)) = f^{n+1}(x)$, translates into the action $(\Phi(x))_n \mapsto (\Phi(x))_{n+1}$ on the components of the itineraries. For this reason one introduces the (one-sided) *shift transformation* (or just *shift*) $\Sigma: S^{\mathbb{N}_0} \to S^{\mathbb{N}_0}$ as follows:

$$\Sigma: (\xi_0, \xi_1, \ldots, \xi_n, \ldots) \mapsto (\xi_1, \xi_2, \ldots, \xi_{n+1}, \ldots). \tag{1.8}$$

[2] The dependence of $\Phi(x)$ on f and α is not made explicit in order to keep the notation simple.

In words, Σ deletes the first component of $(\xi_n)_{n\in\mathbb{N}_0}$ and shifts the other components one position to the left. It is easily shown that Σ is a continuous transformation. As observed above, the diagram

$$\begin{array}{ccc} \Omega & \xrightarrow{f} & \Omega \\ \Phi\downarrow & & \downarrow\Phi \\ S^{\mathbb{N}_0} & \xrightarrow{\Sigma} & S^{\mathbb{N}_0} \end{array}$$

commutes, i.e., $\Phi \circ f = \Sigma \circ \Phi$. Note that Σ is not invertible (indeed, it is a k-to-1 map), although f might be invertible—unless two-sided itineraries are used.

As a simple illustration (see Fig. 1.1), consider the *sawtooth* (also called *dyadic*, *shift*, etc.) *map* E_2:$[0, 1] \to [0, 1]$, defined as

$$E_2(x) = 2x \bmod 1,$$

and decompose $[0, 1]$ into the intervals $A_0 = [0, \frac{1}{2})$ and $A_1 = [\frac{1}{2}, 1]$, so the alphabet is $S = \{0, 1\}$. In this case, the orbit $E_2^n(x)$, $n \in \mathbb{N}_0$, is coded to an infinitely long 0–1 string $\Phi(x)$, where

$$(\Phi(x))_n = \begin{cases} 0 & \text{if } E_2^n(x) \in A_0, \\ 1 & \text{if } E_2^n(x) \in A_1. \end{cases}$$

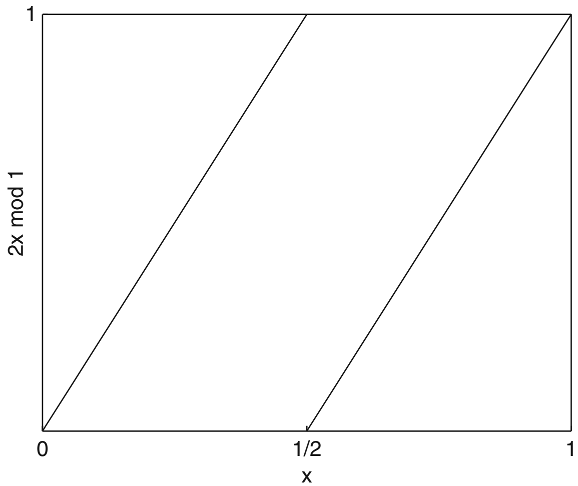

Fig. 1.1 The function $E_2(x) = 2x \bmod 1$

Let

$$x = \frac{b_0}{2} + \frac{b_1}{2^2} + \cdots + \frac{b_k}{2^{k+1}} + \cdots = \sum_{k=0}^{\infty} b_k 2^{-(k+1)} =: 0.b_0 b_1 \ldots b_k \ldots,$$

$b_n \in \{0, 1\}$, be a binary expansion of $x \in [0, 1]$. Then

$$E_2(0.b_0 b_1 \ldots b_k \ldots) = 0.b_1 b_2 \ldots b_{k+1} \ldots$$

for $x \in [0, 1)$ and $E_2(1) = E_2(0.1^\infty) = 0 = 0.0^\infty$, where here and throughout the upper label "∞" attached to a symbol means indefinite repetition of that symbol. The *dyadic rationals* in $(0, 1)$ (i.e., numbers of the form $m/2^n$, $m = 1, 2, \ldots, 2^n - 1$) are characterized by possessing two binary expansions: one terminating with 0^∞ and other terminating with 1^∞. Indeed, $0.10^\infty = 0.01^\infty$ and $0.b_0 \ldots b_{k-1}10^\infty = 0.b_0 \ldots b_{k-1}01^\infty$, $k \geq 1$, since

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

If $x = 0.b_0 b_1 \ldots \in (0, 1)$ is not a dyadic rational, then

$$E_2^n(x) = 0.b_n b_{n+1} \ldots \in A_i \quad \text{iff } b_n = i \in \{0, 1\},$$

hence

$$\Phi(x) = (b_n)_{n \in \mathbb{N}_0} = (b_0, b_1, \ldots, b_n, \ldots). \tag{1.9}$$

Furthermore, $\Phi(0) = (0^\infty)$ and $\Phi(1) = (1, 0^\infty)$. If $x \in (1, 0)$ is a dyadic rational, then x is a preimage of 0 under E_2, thus (1.9) is fulfilled provided $(b_n)_{n \in \mathbb{N}_0}$ corresponds to the binary expansion of x ending with 0^∞. We conclude that given any binary sequence $(b_n)_{n \in \mathbb{N}_0}$ not terminating with 1^∞, there exists always $x \in [0, 1]$, namely $x = 0.b_0 b_1 \ldots$, such that its itinerary with respect to the decomposition $\alpha = \{A_0, A_1\}$ under E_2 is precisely that sequence. In particular, given a finite word b_0^n, there exist infinitely many points in $[0, 1]$, to wit:

$$\begin{aligned} x &\in \left[\frac{b_0 2^n + b_1 2^{n-1} + \cdots + b_n}{2^{n+1}}, \frac{b_0 2^n + b_1 2^{n-1} + \cdots + b_n + 1}{2^{n+1}} \right) \\ &= [0.b_0 \ldots b_n,\ 0.b_0 \ldots b_n + 2^{-(n+1)}), \end{aligned} \tag{1.10}$$

whose itineraries $\Phi(x)$ "realize" the pattern b_0^{L-1} in the sense that $\Phi(x)_0^n = b_0^n$. The fact that all finite words of a symbolic space can be materialized as segments of itineraries for a wide class of maps (Sect. 3.1) contrasts with the situation we shall come upon when studying the so-called ordinal patterns in Sect. 1.2.

Shifts are a special instance of the so-called subshifts. If K is a closed and Σ-invariant (i.e., $\Sigma(K) \subset K$) subset of $S^{\mathbb{N}_0}$, the restriction of the shift transformation to K, written as $\Sigma|_K$, is called a *subshift*. Sometimes Σ is called a *full* shift to distinguish it from the subshifts proper ($K \neq S^{\mathbb{N}_0}$).

A special class of subshifts are of great interest in applications. Let $A = (a_{ij})_{0 \leq i,j \leq k-1}$ be a $k \times k$ matrix of 0's and 1's and define

$$S_A^{\mathbb{N}_0} = \left\{ (\xi_n)_{n\in\mathbb{N}_0} \in S^{\mathbb{N}_0} : a_{\xi_n \xi_{n+1}} = 1 \text{ for all } n \in \mathbb{N}_0 \right\}.$$

Put in simple terms, the matrix A determines which letters $\xi_{n+1} \in S = \{0, 1, \ldots, k-1\}$ may follow the letter ξ_n in the word $(\xi_n)_{n\in\mathbb{N}_0}$. Thus $S_A^{\mathbb{N}_0}$ is a closed and Σ-invariant subset of the sequence space $S^{\mathbb{N}_0}$ that contains all well-formed or *admissible* sequences. Alternatively, one can also describe $S_A^{\mathbb{N}_0}$ by listing the forbidden words. This explains the connection between symbolic dynamics and the theory of formal languages we mentioned above. The restriction of Σ to $S_A^{\mathbb{N}_0}$, written as Σ_A, is called a *subshift of finite type*, *Markov subshift*, or a *topological Markov chain*. If $a_{ij} = 1$ for every $0 \leq i, j \leq k-1$, we recover the full shift. At the opposite end, $S_A^{\mathbb{N}_0}$ may be empty. This happens if and only if the matrix A is nilpotent (i.e., $A^n = 0$ for some $n \in \mathbb{N}$).

As way of example, take $k = 2$ and

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

Since $a_{11} = 0$, this means that the binary sequence $(\xi_n)_{n\in\mathbb{N}_0}$ is admissible if and only if it does not contain two consecutive 1's. In this case, the only forbidden block of length 2 is 11.

Let $\mathbb{K} = \mathbb{N}_0$ or $\mathbb{Z}$, and $(S_A^{\mathbb{K}}, \Sigma_A)$, $(T_B^{\mathbb{K}}, \Sigma_B)$ be two subshifts of finite type possibly with different alphabets S and T, respectively. Suppose $F : S_A^{\mathbb{K}} \to T_B^{\mathbb{K}}$ is a shift-commuting map, that is, $F \circ \Sigma_A = \Sigma_B \circ F$. The continuous, shift-commuting maps from a subshift of finite type $S_A^{\mathbb{K}}$ to another $T_B^{\mathbb{K}}$ were characterized in [92] as those maps for which there exist integers $l \leq r$ and a "local rule" $f : S^{r-l+1} \to T$ such that for any $\xi = (\xi_n)_{n\in\mathbb{K}} \in S_A^{\mathbb{K}}$ and $i \in \mathbb{K}$,

$$F(\xi)_i = f(\xi_{i+l}, \ldots, \xi_{i+r}). \tag{1.11}$$

If F is not the constant map, then a maximal l and a minimal r with this property exist; they are called left and right radii of F, respectively. If $\mathbb{K} = \mathbb{N}_0$, then $l \geq 0$. When $\mathbb{K} = \mathbb{Z}$, $p = \max\{-l, r\}$ is called the *radius* of F. In this case,

$$F(\xi)_i = f(\xi_{i-p}, \ldots, \xi_i, \ldots \xi_{i+p}),$$

where $\xi = (\xi_n)_{n\in\mathbb{Z}}$. A map between two subshifts of finite type of the form (1.11) is called a *block map* [123]. Block maps provide the mathematical underpinnings of cellular automata (Sect. 1.5).

Markov subshifts not only do provide conceptually simple prototypes for important dynamical properties, but they are basic components of some physical systems (e.g., think of Smale's horseshoes in Hamiltonian dynamical systems). To be more specific, we point out next that Markov subshifts can exhibit all properties of low-dimensional chaos.

Let us recall some basic definitions first. A 0–1 matrix A is said to be *transitive* if A^m is positive (i.e., all its entries are positive) for some $m \in \mathbb{N}$. A continuous self-map f of a metric space M is *topologically transitive* if there exists $x \in M$ such that $\mathcal{O}_f(x) = (f^n)_{n\in\mathbb{N}_0}$ is dense in M; if f is invertible, then the requirement for topological transitivity is that $\mathcal{O}_f(x) = (f^n)_{n\in\mathbb{Z}}$ is dense in M for some $x \in M$. It holds [91] that if A is a transitive $k \times k$ matrix, then the topological Markov chain Σ_A is topologically transitive and its periodic orbits are dense in $S_A^{\mathbb{N}_0}$ ($S = \{0, 1, \ldots, k-1\}$), therefore Σ_A is chaotic in the sense of Devaney [69]; in particular, Σ_A has sensitive dependence on initial conditions (see Sect. A.2). This result includes the full shifts. The corresponding statements for f invertible and $M = S_A^{\mathbb{Z}}$ hold true as well.

1.1.3 Dynamical Systems

We shall encounter two kinds of dynamical systems in this book. A *continuous* (or *topological*) *dynamical system* consists of a topological space (e.g., a metrical space) M and a continuous map $f{:}M \to M$. This being the case, these systems will be denoted by the pair (M, f). Subshifts are examples of continuous systems, (K, Σ_K). A *measure-theoretical dynamical system* is comprised of a *measurable space* $(\Omega, \mathcal{B})$, a measurable map $f{:}\Omega \to \Omega$, and a *non-singular measure* μ on $(\Omega, \mathcal{B})$. Thus, Ω is a non-empty set, $\mathcal{B}$ is a sigma-algebra of subsets of Ω, $f^{-1}B \in \mathcal{B}$ for all $B \in \mathcal{B}$, and $B \in \mathcal{B}$ is a μ-zero set iff $f^{-1}B$ is a μ-zero set. Only finite-measure spaces will be considered henceforth. Therefore, $(\Omega, \mathcal{B}, \mu)$ may be assumed without restriction to be a probability space, with μ being a probability on the space of "events" $(\Omega, \mathcal{B})$. Measure-theoretical systems will be denoted by $(\Omega, \mathcal{B}, \mu, f)$. To promote a continuous system (M, f) to a measure-theoretical one, it suffices to endow the topological space M with its Borel sigma-algebra (i.e., the sigma-algebra generated by the open sets), and the corresponding Lebesgue measure. In topological dynamics, the attention focuses on continuous systems. In ergodic theory, the framework is set by *measure-preserving* self-maps of (usually) probability spaces. We say that $f{:}\Omega \to \Omega$ preserves a measure μ on $(\Omega, \mathcal{B})$, if $\mu(f^{-1}B) = \mu(B)$ for all $B \in \mathcal{B}$. Alternatively, we say that the measure-theoretical system $(\Omega, \mathcal{B}, \mu, f)$ is μ-preserving, or that μ is f-invariant. Sometimes, measure-preserving, invertible maps are called *automorphisms*, while the name *endomorphisms* is reserved for the non-invertible ones.

The dynamical complexity of a measure-preserving system $(\Omega, \mathcal{B}, \mu, f)$ can be quantified by its metric entropy. So to speak, the metric entropy measures the uncertainty of the forward evolution of the system when the initial condition is not exactly known —the higher the uncertainty, the greater the complexity. The original proposal of A. Kolmogorov (later completed by Y. Sinai) amounts to the following recipe: coarse-grain the state space of the dynamical system and calculate the Shannon entropy of the resulting stochastic process. Let us follow this path.

A *partition* of a measure space $(\Omega, \mathcal{B}, \mu)$ (or just Ω for brevity) is a disjoint family of elements of $\mathcal{B}$, called atoms, whose union is Ω. Partitions will be denoted by small Greek letters. Two extreme examples of partitions of Ω are the *trivial partition* $\{\emptyset, \Omega\}$ and the *point partition* (or partition of Ω into separate points)

$$\epsilon = \{\{x\}:x \in \Omega\}. \tag{1.12}$$

Except for ϵ, we consider only finite partitions, i.e., partitions with a finite number of atoms. If, furthermore, Ω is a compact metric space with metric d, then the "size" or "coarseness" of a partition $\alpha = \{A_0, A_1, \ldots, A_{|\alpha|-1}\}$ is measured by its *norm* (sometimes also called *diameter*),

$$\|\alpha\| = \sup_{0 \le k \le |\alpha|-1} \{d(x,y):x, y \in A_k\}. \tag{1.13}$$

We saw already in the last section that a discretization of the state space Ω may provide useful insights into a complicated dynamic. In measure-preserving systems this is even more certain since, as we are going to see presently, partitions allow establishing a connection with stochastic and information theory.

Given a finite partition $\alpha = \{A_0, A_1, \ldots, A_{|\alpha|-1}\}$ of $(\Omega, \mathcal{B}, \mu)$, the maps[3] $X_n:\Omega \to S = \{0, 1, \ldots, |\alpha| - 1\}$, $n \in \mathbb{N}_0$, defined as

$$X_n(x) = i \quad \text{iff } f^n(x) \in A_i$$

are random variables on the probability space $(\Omega, \mathcal{B}, \mu)$. Indeed,

$$X_n^{-1}(i) = f^{-n}(A_i) \in \mathcal{B}$$

because f is measurable. Observe that $X_n(x)$ is the nth component of the itinerary of x with respect to α. The difference now with respect to the itineraries of Sect. 1.1.2 is the existence of an invariant measure, which allows to promote $\mathbf{X} = \{X_n\}_{n \in \mathbb{N}_0}$ to a stationary stochastic process. In fact

(i) The probability (mass) function of X_n is given by

---

[3] The dependence of X_n on α is not made explicit here in order to keep the notation simple.

$$\Pr\{X_n = i\} = \mu\left\{x \in \Omega{:}f^n(x) \in A_i\right\} = \mu(f^{-n}A_i) = \mu(A_i),$$

because f is μ-preserving. As for the joint probability function of $X_0, \ldots, X_n = X_0^n$,

$$\begin{aligned}\Pr\left\{X_0^n = i_0, \ldots, i_n\right\} &= \mu\left\{x \in \Omega{:}x \in A_{i_0}, \ldots, f^n(x) \in A_{i_n}\right\} \\ &= \mu\left(A_{i_0} \cap \ldots \cap f^{-n}A_{i_n}\right).\end{aligned}$$

(ii) The stochastic process $\{X_n{:}n \in \mathbb{N}_0\}$ is stationary:

$$\begin{aligned}\Pr\left\{X_k^{k+n} = i_0, \ldots, i_n\right\} &= \mu\left\{x \in \Omega{:}f^k(x) \in A_{i_0}, \ldots, f^{k+n}(x) \in A_{i_n}\right\} \\ &= \mu\left(f^{-k}(A_{i_0} \cap \cdots \cap f^{-n}A_{i_n})\right) \\ &= \mu\left(A_{i_0} \cap \cdots \cap f^{-n}A_{i_n}\right)\end{aligned}$$

because f is μ-preserving. Therefore,

$$\Pr\{X_k = i_0, \ldots, X_{k+n} = i_n\} = \Pr\{X_0 = i_0, \ldots, X_n = i_n\}$$

for every $n, k \in \mathbb{N}_0$.

It follows that the stochastic process $\mathbf{X} = \{X_n\}_{n\in\mathbb{N}_0}$ is an information source with alphabet $S = \{0, 1, \ldots, |\alpha| - 1\}$. The *metric entropy of f with respect to the partition α* is defined to be the Shannon entropy (rate) of $\mathbf{X}$:

$$\begin{aligned}h_\mu(f, \alpha) &= -\lim_{n\to\infty} \frac{1}{n} \sum \Pr\{X_0^{n-1} = i_0, \ldots, i_{n-1}\} \log \Pr\{X_0^{n-1} = i_0, \ldots, i_{n-1}\} \\ &= -\lim_{n\to\infty} \frac{1}{n} \sum \mu(A_{i_0} \cap \cdots \cap f^{-n}A_{i_n}) \log \mu(A_{i_0} \cap \cdots \cap f^{-n}A_{i_n}),\end{aligned}$$

where the summation is over all $i_0, \ldots, i_{n-1} \in S$. If we define the *refinement*

$$\bigvee_{i=0}^{n-1} f^{-i}\alpha = \{A_{j_0} \cap f^{-1}A_{j_1} \cap \cdots \cap f^{-(n-1)}A_{j_{n-1}} : 0 \le j_0, \ldots, j_{n-1} \le |\alpha| - 1\}$$

of the partition $\alpha = \{A_0, \ldots, A_{|\alpha|-1}\}$, and the function

$$H_\mu(\beta) = -\sum_{j=0}^{|\beta|-1} \mu(B_j) \log (B_j)$$

for any partition $\beta = \{B_0, \ldots, B_{|\beta|-1}\}$ of $(\Omega, \mathcal{B}, \mu)$, then we recover the usual expression of $h_\mu(f, \alpha)$:

$$h_\mu(f,\alpha) = \lim_{n\to\infty} \frac{1}{n} H_\mu\left(\bigvee_{i=0}^{n-1} f^{-i}\alpha\right). \tag{1.14}$$

The convergence of this limit is proven in Sect. B.2.

If an application of f is interpreted as a passage of one unit of time, then $\bigvee_{i=0}^{n-1} f^{-i}\alpha$ represents the combined experiment of performing n consecutive times the original experiment, represented by α. Then $h_\mu(f,\alpha)$ is the average information per unit of time that one gets from performing the original experiment every unit of time [202].

The metric (Kolmogorov–Sinai or measure-theoretical) entropy of f is then the supremum of $h_\mu(f,\alpha)$ over all finite partitions of $(\Omega, \mathcal{B}, \mu)$:

$$h_\mu(f) = \sup_\alpha h_\mu(f,\alpha). \tag{1.15}$$

Continuing with the previous information-theoretical interpretation, $h_\mu(f)$ provides the maximum average information per unit of time obtainable by performing the same experiment every unit of time.

In general there are several obstacles preventing an exact calculation of $h(f)$. First, except in simple cases limit (1.14) itself is not computable, so we must be content with an evaluation of $\frac{1}{n}H_\mu\left(\bigvee_{i=0}^{n-1} f^{-i}\alpha\right)$ for some large value of n. Second, considerable computation is necessary to identify the elements of the refined partitions $\bigvee_{i=0}^{n-1} f^{-i}\alpha$, the computational effort being exponential in n. Third, the measure μ is usually unknown to us in closed form. Fortunately, there are exceptions, for instance, when one can find a partition α for which $h_\mu(f,\alpha) = h_\mu(f)$. Such partitions are called generators or *generating partitions* with respect to f. A finite partition α is a *one-sided generator* for f if

$$\bigvee_{i=0}^{\infty} f^{-i}\alpha = \epsilon, \tag{1.16}$$

where ϵ is the point partition of Ω (see (1.12)). Moreover, if f is even an automorphism and $\bigvee_{i=-\infty}^{\infty} f^{-i}\alpha = \epsilon$, then α is called a *two-sided generator* or just a generator for f. Automorphisms may have not only generators but also one-sided generators. According to the Kolmogorov–Sinai theorem (Annex B.13), if α is a generator (one-sided or not) for f, then $h_\mu(f,\alpha) = h_\mu(f)$.

As way of illustration, consider the *symmetric tent map* $\Lambda{:}[0,1] \to [0,1]$ defined as (Fig. 1.2)

$$\Lambda(x) = 1 - |1-2x| = \begin{cases} 2x & \text{if } 0 \le x \le \frac{1}{2}, \\ 2(1-x) & \text{if } \frac{1}{2} \le x \le 1. \end{cases} \tag{1.17}$$

If we equip $[0,1]$ with the Borel sigma-algebra (generated by the intersections of open intervals of $\mathbb{R}$ with $[0,1]$), then Λ is easily seen to preserve the Lebesgue

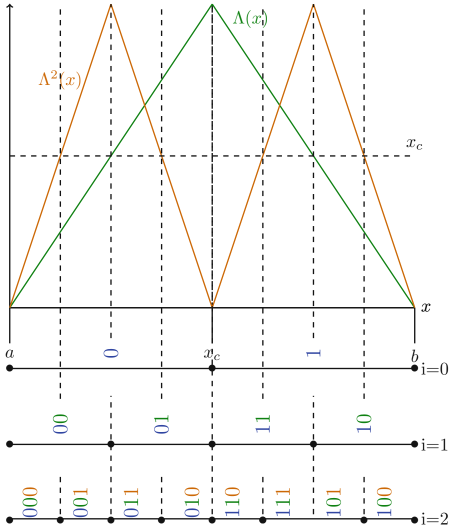

Fig. 1.2 Symbolic intervals generated by the symmetric tent map Λ and its second iterate $\Lambda^2(x_c = \frac{1}{2})$

measure. As in the previous section, let $\alpha = \{A_0, A_1\}$, where

$$A_0 = [0, \tfrac{1}{2}),\ A_1 = [\tfrac{1}{2}, 1].$$

Then,

$$\Lambda^{-1}A_0 = [0, \tfrac{1}{4}) \cup (\tfrac{3}{4}, 1],\ \Lambda^{-1}A_1 = [\tfrac{1}{4}, \tfrac{3}{4}].$$

Hence

$$\alpha \cap \Lambda^{-1}\alpha = \{A_{00}, A_{01}, A_{11}, A_{10}\},$$

with

$$A_{00} = A_0 \cap \Lambda^{-1}A_0 = [0, \tfrac{1}{4}), \quad A_{01} = A_0 \cap \Lambda^{-1}A_1 = [\tfrac{1}{4}, \tfrac{1}{2}),$$

$$A_{11} = A_1 \cap \Lambda^{-1}A_1 = [\tfrac{1}{2}, \tfrac{3}{4}], \quad A_{10} = A_1 \cap \Lambda^{-1}A_0 = (\tfrac{3}{4}, 1].$$

The sets of $\alpha, \alpha \cap \Lambda^{-1}\alpha$, and $\alpha \cap \Lambda^{-1}\alpha \cap \Lambda^{-2}\alpha$ are shown in Fig. 1.2. In general,

$$\bigcap_{i=0}^{k} \Lambda^{-i}\alpha = \left\{A_{b_0 b_1 \ldots b_k} : b_0, b_1, \ldots, b_k \in \{0, 1\}\right\},$$

where the 2^{k+1} disjoint sets

$$A_{b_0 b_1 \ldots b_k} = A_{b_0} \cap \Lambda^{-1} A_{b_1} \cap \cdots \cap \Lambda^{-k} A_{b_k} \tag{1.18}$$

build a family of ever-shorter intervals that covers uniformly the unit interval. As a matter of fact, the sets $A_{b_0 b_1 \ldots b_k}$ are a permutation of the dyadic intervals (1.10), except eventually for the endpoints. It follows that $\bigcap_{i=0}^{k} \Lambda^{-i}\alpha$ converges to the point partition of $[0, 1]$, hence α is a one-sided generator for Λ. If λ denotes the Lebesgue measure, $\lambda(dx) = dx$, then

$$\begin{aligned} h_\lambda(\Lambda) &= -\lim_{n\to\infty} \frac{1}{n} \sum_{b_0 \ldots b_{n-1} \in \{0,1\}} \lambda(A_{b_0 \ldots b_{n-1}}) \log \lambda(A_{b_0 \ldots b_{n-1}}) \\ &= -\lim_{n\to\infty} \frac{1}{n} \sum_{b_0 \ldots b_{n-1} \in \{0,1\}} 2^{-n} \log 2^{-n} \\ &= \log 2. \end{aligned}$$

A similar argument can be applied to other maps, like the *logistic map* $g{:}[0, 1] \to [0, 1]$,

$$g(x) = 4x(1 - x). \tag{1.19}$$

In this case, the absolutely continuous measure[4]

$$\mu(dx) = \frac{dx}{\pi\sqrt{x(1-x)}} \tag{1.20}$$

is g-invariant. This measure is called the *natural* or *physical invariant measure* of g because it is the one obtained in numerical experiments [72].

Since $(\Omega, \mathcal{B}, \mu)$ is a probability space, dynamical complexity can be given a probabilistic meaning. In this sense we can say that the entropy $h_\mu(f)$ (or other related concepts, like the Lyapunov exponents greater than 1) measures the randomness or, rather, the pseudo-randomness of the dynamic induced by the map f.

The complexity of continuous dynamical systems is usually measured by the topological entropy. As we shall presently see, this quantity is related to the periodic structure in some relevant systems. Rather than going into the definition of topological entropy, which is quite technical (see Sect. B.3), we only recall here its expression for a one-sided or two-sided Markov subshift Σ_A. It can be shown [91] that

$$h_{\text{top}}(\Sigma_A) = \limsup_{n\to\infty} \frac{1}{n} \log^+ P_n(\Sigma_A),$$

[4] Absolute continuity of measures will refer to the Lebesgue measure throughout this book.

where $h_{\text{top}}(\Sigma_A)$ is the topological entropy of Σ_A (in general, $h_{\text{top}}(f)$ stands for the topological entropy of a continuous self-map f), $P_n(\Sigma_A)$ is the number of periodic points of period n of Σ_A, and $\log^+ x = \log x$ if $x \geq 1$, and 0 otherwise. To explicitly calculate the right-hand side of this expression, we need the following two properties: (i) If B is a non-negative matrix, then there exists an eigenvalue $\lambda_{\max} \geq 0$ such that no other eigenvalue of B has absolute value greater than $\lambda_{\max}$ (this is part of the Perron–Frobenius theorem [202]) and (ii) the number of periodic points of period $p \in \mathbb{N}$ of a Markov subshift Σ_A is the trace of A^p (i.e., the sum of the diagonal elements), denoted as $\operatorname{tr} A^p$. For the full shift on k symbols, $(A^n)_{ij} = k^{n-1}$, for all $0 \leq i,j \leq k-1$, hence the trace of A^n is k^n. This yields

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

In general, $\operatorname{tr} A^p = \lambda_1^p + \cdots + \lambda_k^p$, where λ_i are the k eigenvalues (eventually repeated) of the matrix A. It follows that [91]

$$h_{\text{top}}(\Sigma_A) = \log^+ \lambda_{\max}.$$

1.1.4 Computer Science

The origin of algorithmic complexity has to be sought in the efforts of R. Solomonoff, A. Kolmogorov, and G. Chaitin to define the elusive concept of “randomness” of finite-alphabet sequences [79, 133, 201]. The basic intuition is that random sequences are “patternless,” hence there is no efficient way to describe them other than giving the sequence itself. The *algorithmic complexity* of a string $s_0^{n-1} = s_0 s_1 \ldots s_{n-1}$, written as $K(s_0^{n-1})$, can be consistently defined as the length of the shortest binary program that, run on a universal prefix-free Turing machine, outputs s_0^{n-1} and halts [59, 67, 138]. As in the case of information theory, this definition of complexity is linked to the general concept of compressibility, this time with respect to all possible algorithms that produce the sequence in question.

Somewhat paradoxically, algorithmic complexity is not a computable quantity. Then suppose that K_n is claimed to be the complexity of a length-n string s_0^{n-1}. In order to check this, we remove one bit from the hypothetically shortest program and let it run. There are two possibilities: either the $(K_n - 1)$-bit program outputs a string different from s_0^{n-1} and halts or else it runs longer than we have time to wait. In the second case, there is no way to know whether the program will halt (this is the famous Turing's halting problem), eventually revealing the actual complexity to be $K_n - 1$.

Any finite sequence s_0^{n-1} can be certainly output by the copy program: “PRINT $s_0, \ldots, s_n$.” Without loss of generality, we may restrict to binary sequences for the time being. Since patternless n-bit sequences cannot be computed by any algorithm significantly shorter than the copy program, their complexity is given by $K_n \leq n + C$, where C is a constant that accounts for the computational overhead (like the operating system). At the opposite end stands the sequences consisting of a repeated bit, say 0. The complexity of the program "PRINT 0, n TIMES" can be bounded as $K_n \leq \log_2 n + C'$, where $\log_2 n$ is the number of bits needed to specify the length n and, again, C' is the computational overhead. Observe that if these programs are run on a computer other than a universal Turing machine, the constants C and C' may depend on the machine, but they are independent of the actual sequence being calculated. In the limit of very long sequences, the algorithmic complexity will practically range between $\log_2 n$ and n. This being the case, one may state that the binary sequence s_0^{n-1} is random if $K(s_0^{n-1}) \simeq n$. (In the non-binary case, $K(s_0^{n-1}) \simeq nb$ for random sequences, where b is the minimal number of bits needed to code the symbols s_i, $0 \leq i \leq n-1$.) Formally, a sequence $(s_n) \in S^{\mathbb{N}_0}$ is said to be *incompressible* when there exists a constant C such that

$$K(s_0^{n-1}) \geq n - C$$

for all $n \geq 1$.

Randomness can also be defined as *typicality*, meaning that typical sequences have no feature that makes them special in any sense. This was the path taken by Martin-Löf to come to grips with the concept of random sequence. Rather than addressing the technicalities of this approach, which are beyond the scope of this book, we will proceed directly to the conclusions: random sequences are realizations of stochastic processes.

Let $(\Omega, \mathcal{B}, \mu)$ be a probability space. The realizations of a stochastic process $\{X_n\}_{n \in \mathbb{N}_0}$ on $(\Omega, \mathcal{B}, \mu)$ with a finite number of possible outcomes can be identified with the elements of a (one-sided) sequence space. Specifically, if $X_n : \Omega \to S$ with $S = \{s_1, \ldots, s_{|S|}\}$ for every $n \in \mathbb{N}_0$, then $(X_n(\omega))_{n \in \mathbb{N}_0} \in S^{\mathbb{N}_0}$ for every $\omega \in \Omega$. The general method to place a probability m on $S^{\mathbb{N}_0}$ induced by the probability μ is explained in Sect. A.3. At present we only need to resort to the so-called (p, q)-Bernoulli shifts or systems on two symbols, which are measure-preserving systems $(S^{\mathbb{N}_0}, \mathcal{B}, m, \Sigma)$, where

(i) $S = \{0, 1\}$,

(ii) $\mathcal{B}$ is the sigma-algebra generated by the so-called *cylinder sets*,

$$C_{s_0 \ldots s_{n-1}} = \{\xi_0^\infty \in S^{\mathbb{N}_0} : \xi_0 = s_0, \ldots, \xi_{n-1} = s_{n-1}\},$$

(iii) the probability m of the binary string $s_0^{n-1} = s_0 s_1 \ldots s_{n-1}$ is defined as

$$m(s_0^{n-1}) = m(C_{s_0 \ldots s_{n-1}}) = p^k q^{n-k},$$

where $p + q = 1$, k is the number of 1's in s_0^{n-1}, and $n - k$ is the number of 0's, and

(iv) Σ is the shift transformation on $S^{\mathbb{N}_0}$.

In the language of probability theory, the cylinder sets correspond to the elementary events; in the language of computer science, $C_{s_0 \ldots s_{n-1}}$ comprises all sequences with the prefix $w = s_0, \ldots, s_{n-1}$. The (p, q)-Bernoulli system models an independent, dichotomous process, one outcome (say, "success") having probability p to occur and the other ("failure") probability $q = 1 - p$. Think, for example, of a random experiment consisting in tossing forever a coin with the odds p for head and q for tail. The shift Σ corresponds to the "time" translation $n \mapsto n + 1$. The fact that Σ preserves m (or, equivalently, that m is Σ-invariant) accounts for the probabilities being the same in every draw.

In particular, the $(\frac{1}{2}, \frac{1}{2})$-Bernoulli system is a model for the tossing of a fair coin. If $0.b_0b_1 \ldots b_n \ldots$ is a binary expansion and $\Phi:[0, 1] \to \{0, 1\}^{\mathbb{N}_0}$ is the map

$$\Phi:0.b_0b_1 \ldots b_n \ldots \mapsto (b_0, b_1, \ldots, b_n, \ldots)$$

we met already in (1.9), then

$$\Phi([0.b_0b_1 \ldots b_n,\ 0.b_0b_1 \ldots b_n + 2^{-(n+1)})) = C_{b_0b_1 \ldots b_n}.$$

Thus, Φ allows to identify the cylinder set $C_{b_0b_1 \ldots b_n}$ of $\{0, 1\}^{\mathbb{N}_0}$ with the interval $[0.b_0b_1 \ldots b_{n-1}, 0.b_0b_1 \ldots b_{n-1} + 2^{-n})$ of $[0, 1]$. But even more is true. If m denotes the measure of the $(\frac{1}{2}, \frac{1}{2})$-Bernoulli system and λ the Lebesgue measure of $[0, 1]$, then

$$m(C_{b_0b_1 \ldots b_{n-1}}) = \frac{1}{2^n} = \lambda([0.b_0b_1 \ldots b_{n-1},\ 0.b_0b_1 \ldots b_{n-1} + 2^{-n})).$$

Since the cylinder sets generate the sigma-algebra of the Bernoulli systems and the semi-open dyadic intervals do the same for the Borel sigma-algebra of $[0, 1]$, we conclude $m = \lambda \circ \Phi^{-1}$, i.e., m corresponds to the Lebesgue (or uniform) measure on $[0, 1]$.

Levin, Schnorr, and Chaitin proved that a binary sequence is typical with respect to the $(\frac{1}{2}, \frac{1}{2})$-Bernoulli measure (i.e., it can be considered the result of tossing a fair coin indefinitely) if and only if it is incompressible. In this way, two seemingly different concepts of randomness incompressibility and typicality are shown to coincide in a natural setting.

Remarkably enough, this result is not the only achievement connecting concepts related to complexity but stemming from different areas. Let us provide another one in which algorithmic complexity and metric entropy are brought together.

Given a measure-preserving dynamical system $(\Omega, \mathcal{B}, \mu, f)$, each $x \in \Omega$ generates an infinitely long sequence under the action of f, namely, its (forward) orbit $\mathcal{O}_f(x) = \{f^n(x):n \in \mathbb{N}_0\}$. Let $s_0^\infty = s_0^\infty(x, \alpha)$ be the itinerary of x with respect to the partition $\alpha = \{A_0, \ldots, A_{|\alpha|-1}\}$ of Ω, that is, $s_k = i$ iff $f^k(x) \in A_i$, $i \in \{0, \ldots, |\alpha| - 1\}$. The *algorithmic complexity* of $\mathcal{O}_f(x)$, written as $k(f, x)$, is measured by the largest algorithmic complexity per symbol of $s_0^\infty(x, \alpha)$ over all possible finite partitions α:

$$k(f, x) = \sup_\alpha \limsup_{n \to \infty} \frac{1}{n} K(s_0^{n-1}(x, \alpha)).$$

Of course, one expects that random-like trajectories are computationally more difficult to reproduce than the regular ones. This expectation can be rigorously proved under the proviso that f is ergodic with respect to the invariant measure μ. In this case [39],

$$k(f, x) = h_\mu(f) \quad \mu\text{-almost everywhere.}$$

1.1.5 Cellular Automata

A cellular automaton is a discrete-time dynamical system with discrete space and discrete states. The state variables are defined on the sites of a D-dimensional regular lattice ($\mathbb{Z}^D$)—the cells of the D-dimensional automaton—taking on values in a finite alphabet $S = \{0, 1, \ldots, k-1\}$. The set of all possible states (formally the set of all possible mappings $\mathbb{Z}^D \to S$) is called the *configuration space*. For numerical simulations it is convenient that the lattice of sites is finite or has a non-trivial topology, like a circle or a 2-torus; these requirements can be implemented with quiescent cells or with periodic conditions, respectively. In order to accommodate this disparity of possibilities, the configuration space will be denoted by a neutral Ω. The states of the cells evolve synchronously in discrete time steps according to identical rules. But what makes cellular automata special is the evolution rule: the state of a particular cell is determined by the previous states of a neighborhood of cells around it.

Cellular automata were introduced by Ulam [199] and von Neumann [161] as simple models of universal computation and machine self-reproduction, respectively. Indeed, a remarkable property of cellular automata is their ability to simulate other symbol processors. Another one is self-organization, even when started from disordered configurations. Two-dimensional cellular automata became quite popular in the 1970s thanks to the article that Martin Gardner devoted to John Conway's *Game of Life* in his section "Mathematical Games" of *Scientific American* [84]. A purely mathematical approach was initiated by Hedlund and collaborators, who studied the endomorphisms and automorphisms of the shift dynamical system [92]. Apart from the many subsequent papers on their dynamical and ergodic properties from this point of view, cellular automata have also been the object of intensive study in mathematical physics, computer science, biology, etc. [207]. Being at the crossroads of symbolic dynamical systems and computation, it is not surprising that the theory of cellular automata benefits from both areas, at the same time that cross-pollinate them, as we try to show in the next lines. For a readable account on cellular automata and their remarkable performance in physical modeling, see, e.g., [198].

For simplicity we will consider only one-dimensional cellular automata. In this case, the configuration space is the two-sided sequence space $S^{\mathbb{Z}}$. One-sided sequences or even finite sequences, corresponding to lattices adequately flanked by quiescent cells, may also be considered along the same lines. A *neighborhood* of size $l \geq 1$ of the cell $i \in \mathbb{Z}$, written as $\mathcal{U}_l(i)$, is the set of $2l + 1$ cells

$$i-l,\ i-l+1,\ldots,i,\ldots,i+l.$$

The state of cell i at time $t \geq 0$ will be denoted as $s_t(i)$. At each time step $t+1$, the previous state at each cell i, $s_t(i) \in S$, is updated according to the states of $\mathcal{U}_l(i)$ by a *local rule* $f{:}S^{2l+1} \to S$ of the form

$$s_{t+1}(i) = f(s_t(i-l), s_t(i-l+1), \ldots, s_t(i+l)).$$

Note that f does not depend on i nor t, but only on the states of $\mathcal{U}_l(i)$; if f is allowed to depend on i, then one speaks of *hybrid* cellular automata.

The local rule f leads to a *global transition map* of the configuration space, $F{:}\Omega \to \Omega$, defined in the obvious way:

$$\begin{aligned} F(\ldots, s_t(i), \ldots) &= (\ldots, f(s_t(i-l), s_t(i-l+1), \ldots, s_t(i+l)), \ldots) \\ &= (\ldots, s_{t+1}(i), \ldots). \end{aligned}$$

Observe that F is a block map from a full shift to itself of radius l. As pointed out in Sect. 1.1.2, it follows that F is continuous and shift-commuting. (This characterization generalizes to D-dimensional cellular automata just by replacing the sequence space $S^{\mathbb{Z}}$ by $S^{\mathbb{Z}^D}$.)

As way of illustration, Fig. 1.3 depicts the time evolution of a one-dimensional, binary cellular automaton with periodic boundary conditions: $s_t(N+1) = s_t(1)$ and $s_t(0) = s_t(N)$ for all $t \geq 0$. Here $N = 250$, the horizontal axis represents space (label i), and time (label t) elapses along the vertical direction, from top to bottom. Once the initial configuration has been fixed, the global map F determines the dynamics of the automaton on the configuration space.

The relation between the properties of the local rule f and the properties of the global transition map F is one of the most important and difficult problems in the

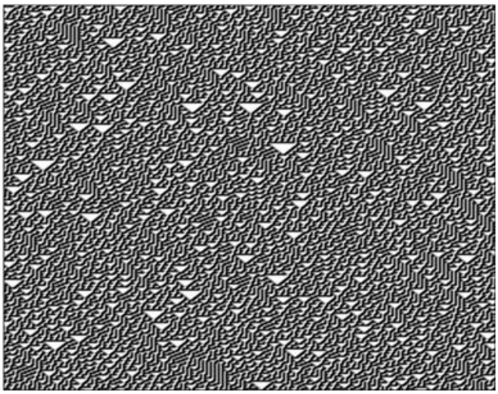

Fig. 1.3 A typical space–time evolution diagram of a one-dimensional cellular automaton with 250 sites and periodic boundary conditions. Time elapses from top to bottom

theory of cellular automata. This problem has been proved to be algorithmically unsolvable for some properties (surjectivity and injectivity for dimension $D > 1$, nilpotency for $D \geq 1$, etc.), and it is believed to be unsolvable for others (ergodicity, sensitivity, etc.).

On a more practical level, hybrid cellular automata with binary state variables and null boundaries (i.e., the cells delimiting the site lattice are permanently in the 0-state) have been explicitly shown to emulate linear feedback shift registers (LFSRs), which are widely used in cryptography as pseudo-random bit generators for stream ciphers. Specifically, given the primitive polynomial of an LFSR [151], then the algorithm given in [48] allows to "synthesize" a null-boundary, hybrid binary cellular automaton that emulates the said LFSR using only the local rules $f(p,q,r) = p + r \bmod 2 \equiv p \oplus q$ and $f(p,q,r) = p + q + r \bmod 2 \equiv p \oplus q \oplus r$. Most importantly, the same is true for the so-called self-shrunken LSFRs [149], which are nonlinear structures featured in some designs of stream ciphers. Since the previous local rules are linear, this fact allows to cryptanalize such ciphers using cellular automata.

Suppose that the configuration space Ω is $S^{\mathbb{Z}}$. In the topology induced by the cylinder sets

$$C_{s_{-n},\ldots,s_0,\ldots,s_n} = \{\xi_0^\infty \in S^{\mathbb{Z}} : \xi_k = s_k, |k| \leq n\},$$

the global transition map $F:\Omega \to \Omega$ that updates the states of the cellular automaton is continuous, which makes (Ω, F) a continuous dynamical system. Hence, we can measure the complexity of its time evolution with the topological entropy $h_{\text{top}}(F)$; see Sect. B.3 for different ways of calculating the topological entropy of a continuous dynamical system. Alternatively, let $R(w,t)$ be the number of distinct rectangles of width w and height (temporal extent) t occurring in a space–time evolution diagram of (Ω, F); see Fig. 1.4. Then [62]

$$h_{\text{top}}(F) = \lim_{w\to\infty} \lim_{t\to\infty} \frac{1}{t} \log R(w,t). \tag{1.21}$$

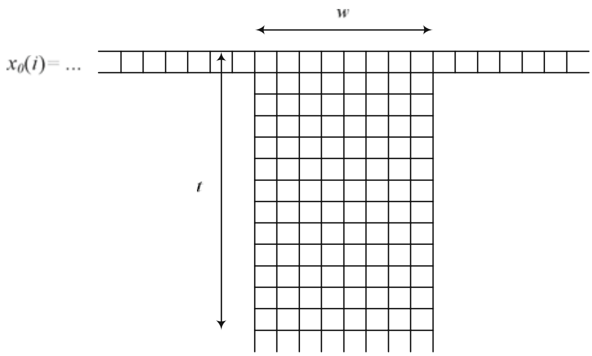

Fig. 1.4 Geometrical illustration of the rectangles $R(w,t)$ used in (1.21)

Therefore, the complexity of (Ω, F) can be measured by the number of distinct words or patterns per time unit generated by the global transition map F as time evolves. It follows that

$$h_{\text{top}}(F) \leq 2l \log k,$$

where l is the neighborhood size of the automaton and $k = |S|$.

Topological entropy belongs also to the dynamical properties that cannot be algorithmically computed for general cellular automata [101]. More generally, whether metric and/or topological entropy is effectively computable (i.e., can be approximated with an arbitrary small error) is an open question for most dynamical systems.

1.2 Admissible and Forbidden Ordinal Patterns

The concept of ordinal pattern of length L only demands a totally ordered set $(\Omega, \leq)$. Let us caution the reader that there are several definitions of ordinal patterns in the literature; the one used in this book follows Bandt et al. [28, 29]. In the simplest setting, the ordinal pattern defined by the elements $x_0, \ldots, x_{L-1} \in \Omega$ can be viewed as the permutation π of $\{0, 1, \ldots, L-1\}$ that arrange those elements according to their order in Ω: $x_{\pi_0} < x_{\pi_1} < \cdots < x_{\pi_{L-1}}$. In case $x_i = x_j$, we agree that $x_i < x_j$ if $i < j$. We write $\pi = \langle \pi_0, \pi_1, \ldots, \pi_{L-1} \rangle$ to summarize that x_{π_0} is the smallest element, x_{π_1} is the second smallest element, etc., in the length-L sequence $x_0, \ldots, x_{L-1}$. For example, if $\Omega = \mathbb{R}$ (endowed with the standard order), and $x_0 = \sqrt{3}$, $x_1 = e$, $x_2 = 2$, and $x_3 = -1.7$, then $\pi = \langle 3, 0, 2, 1 \rangle$. In an extended setting where we have a self-map f of Ω, the sets of points to be arranged by π are naturally provided by the initial segments of the f-orbits: $x_n = f^n(x)$, $0 \leq n \leq L-1$. In this case, one usually dispenses with periodic orbits of period smaller than L. The set of ordinal L-patterns will be denoted by $\mathcal{S}_L$ throughout this book. Ordinal patterns are sometimes called permutations.

As a minor technical point, let us mention that a permutation $\tau{:}i \mapsto \tau(i)$, $i \in \{0, 1, \ldots, L-1\}$, is written in combinatorics as

$$\begin{pmatrix} 0 & 1 & \ldots & L-1 \\ \tau(0) & \tau(1) & \ldots & \tau(L-1) \end{pmatrix} =: [\tau(0), \tau(1), \ldots, \tau(L-1)]. \tag{1.22}$$

Observe that an ordinal pattern $\pi = \langle \pi_0, \ldots, \pi_{L-1} \rangle$ does not correspond—as one might think— to the permutation $[\pi_0, \ldots, \pi_{L-1}]$, but rather to its inverse: $\pi_0 \mapsto 0, \ldots, \pi_{L-1} \mapsto L-1$, i.e.,

$$\langle \pi_0, \ldots, \pi_{L-1} \rangle = \begin{pmatrix} \pi_0 & \pi_1 & \ldots & \pi_{L-1} \\ 0 & 1 & \ldots & L-1 \end{pmatrix} = \left[\pi_0, \ldots, \pi_{L-1}\right]^{-1}. \tag{1.23}$$

For example, the ordering $x_2 < x_0 < x_1$ defines the ordinal pattern $\langle 2, 0, 1 \rangle$ but the permutation $0 = \pi_1 \mapsto 1$, $1 = \pi_2 \mapsto 2$, and $2 = \pi_0 \mapsto 0$, which in the conventional notation reads

$$[1,2,0] = [2,0,1]^{-1}.$$

In sum, an ordinal pattern $\pi \in \mathcal{S}_L$ corresponds actually to the permutation $\pi_i \mapsto i$, $0 \le i \le L-1$, which will be denoted as $[\pi]^{-1}$ whenever needed:

$$[\pi]^{-1} = [\pi_0, \pi_1, \ldots, \pi_{L-1}]^{-1}. \tag{1.24}$$

Furthermore, if $\pi = \langle \pi_0, \ldots, \pi_{L-1} \rangle$ and $\pi' = \left\langle \pi'_0, \ldots, \pi'_{L-1} \right\rangle$, a (non-commutative) product $\pi \circ \pi'$ can be defined in $\mathcal{S}_L$ via composition

$$\begin{aligned} \pi \circ \pi' &= \begin{pmatrix} \pi_0 & \pi_1 & \ldots & \pi_{L-1} \\ 0 & 1 & \ldots & L-1 \end{pmatrix} \begin{pmatrix} \pi'_0 & \pi'_1 & \ldots & \pi'_{L-1} \\ 0 & 1 & \ldots & L-1 \end{pmatrix} \\ &= \begin{pmatrix} \pi'_{\pi_0} & \pi'_{\pi_1} & \ldots & \pi'_{\pi_{L-1}} \\ 0 & 1 & \ldots & L-1 \end{pmatrix} \\ &= \langle \pi'_{\pi_0}, \pi'_{\pi_1}, \ldots, \pi'_{\pi_{L-1}} \rangle. \end{aligned} \tag{1.25}$$

Endowed with this product, $\mathcal{S}_L$ becomes a non-Abelian group of order $L!$. The neutral element of the group $(\mathcal{S}_L, \circ)$ is the identity permutation $\langle 0, 1, \ldots, L-1 \rangle$. Ordinal patterns will be studied in detail in Chap. 3.

After these algebraic prolegomena, consider now a function $f: I \to I$, where I is a closed interval of $\mathbb{R}$. Given the finite orbit $\{f^n(x): 0 \le n \le L-1\}$ of $x \in I$, we say that x defines the ordinal pattern of length L (or ordinal L-pattern) $\pi = \pi(x) = \langle \pi_0, \pi_1, \ldots, \pi_{L-1} \rangle$ if

$$f^{\pi_0}(x) < f^{\pi_1}(x) < \cdots < f^{\pi_{L-1}}(x). \tag{1.26}$$

We say also that π is realized by x or that x is of type π.

If, for example, $I = [0, 1]$ and g is the *logistic map*, $g(x) = 4x(1-x)$, then we find to four digit precision.

$$\mathcal{O}_g(0.6416) = 0.6416,\ 0.9198,\ 0.2951,\ 0.8320,\ 0.5590,\ 0.9861, \ldots$$

hence $x = 0.6416$ is of the types

$$\langle 0, 1 \rangle, \langle 2, 0, 1 \rangle, \langle 2, 0, 3, 1 \rangle, \langle 2, 4, 0, 3, 1 \rangle, \langle 2, 4, 0, 3, 1, 5 \rangle, \ldots$$

Instead of fixing x and varying L, we can do the opposite, as in the following illustration with $L = 3$:

$$\begin{array}{ll}\mathcal{O}_g(0.15) = 0.15, 0.51, 0.9996, \ldots & \text{hence } 0.15 \text{ is of type } \langle 0,1,2\rangle\,, \\ \mathcal{O}_g(0.30) = 0.30, 0.84, 0.5376, \ldots & \text{hence } 0.30 \text{ is of type } \langle 0,2,1\rangle\,, \\ \mathcal{O}_g(0.55) = 0.55, 0.99, 0.0396, \ldots & \text{hence } 0.55 \text{ is of type } \langle 2,0,1\rangle\,, \\ \mathcal{O}_g(0.80) = 0.80, 0.64, 0.9216, \ldots & \text{hence } 0.80 \text{ is of type } \langle 1,0,2\rangle\,, \\ \mathcal{O}_g(0.95) = 0.95, 0.19, 0.6156, \ldots & \text{hence } 0.95 \text{ is of type } \langle 1,2,0\rangle\,.\end{array}$$

Points and ordinal patterns provide complementary perspectives of the same picture. Thus, as in the first instance, one can be more interested in the ordinal patterns defined by a given point or, as in the second instance, in the points that realize a given pattern. In order to introduce the second point of view, we define following [29] the sets

$$P_\pi = \{x \in I{:}x \text{ defines } \pi \in \mathcal{S}_L\}. \tag{1.27}$$

If $P_\pi \neq \emptyset$, then π is said to be an *allowed* or *admissible* (ordinal) *pattern* for f; otherwise π is called a *forbidden* (ordinal) *pattern* for f. In words, $\pi \in \mathcal{S}_L$ is allowed or admissible if there exists $x \in I$ such that x is of type π, whereas it is forbidden if no x is of type π. We will see shortly that maps have forbidden patterns (in fact, infinitely many of them) under quite general assumptions.

The properties of the sets $P_\pi \neq \emptyset$ are closely related to the properties of f. Thus, P_π is a union of open intervals if f is continuous or the union of intervals (including none, one, or both endpoints) if f is piecewise continuous. The endpoints of P_π are determined by the periodic points of f. All these facts can be easily exposed via the graphs of the map and their iterates. First of all, draw the graph of the identity (f^0) in the square $I \times I \subset \mathbb{R}^2$, which is the diagonal $y = x$, $x \in I$, on the Cartesian plane $\{(x, y) \in \mathbb{R} \times \mathbb{R}\}$. Then draw the graphs of the functions $y = f(x), \ldots, y = f^{L-1}(x)$, $x \in I$. The components of the distinct P_π's, $\pi \in \mathcal{S}_L$, are separated by the intersection points of all those graphs. Indeed, if $x \in P_\pi$ "moves" leftward or rightward, it will leave the current component of P_π at the left or right endpoint, respectively, as soon as the condition

$$f^{\pi_i}(x) = f^{\pi_{i+1}}(x) \tag{1.28}$$

holds for some $i = 0, 1, \ldots, L-2$, unless it leaves the interval I before. Note that condition (1.28) implies that $f^{\min\{\pi_i,\pi_{i+1}\}}(x)$ is a periodic point of period $|\pi_i - \pi_{i+1}|$, thus x is a $\min\{\pi_i, \pi_{i+1}\}$th preimage of such a point. In this case, $\min\{\pi_i, \pi_{i+1}\} + |\pi_i - \pi_{i+1}| = \max\{\pi_i, \pi_{i+1}\} \leq L-1$. In particular, if $\pi_i = 0$ or $\pi_{i+1} = 0$, then x is a periodic point.

In short, the endpoints of the intervals $P_\pi \neq \emptyset$, $\pi \in \mathcal{S}_L$, are given by the periodic points of f of periods $p \leq L-1$, and their preimages up to the order $L-2$. We conclude that the admissible ordinal patterns for f are determined by its periodic structure.

As a simple illustration, consider again the logistic map $g(x) = 4x(1-x)$, $0 \leq x \leq 1$. For $L = 2$ we have, see Fig. 1.5,

$$P_{\langle 0,1\rangle} = \left(0, \tfrac{3}{4}\right), \quad P_{\langle 1,0\rangle} = \left(\tfrac{3}{4}, 1\right).$$

The separating point $x = \frac{3}{4}$ between $P_{\langle 0,1\rangle}$ and $P_{\langle 1,0\rangle}$ is given by the condition $g^{\pi_0}(x) = g^{\pi_1}(x)$, where $\pi_0, \pi_1 \in \{0, 1\}$, i.e.,

$$g(x) = x.$$

For $L = 3$ ($g^2(x) = -64x^4 + 128x^3 - 80x^2 + 16x$), Fig. 1.6 shows that

$$\begin{aligned} &P_{\langle 0,1,2\rangle} = \left(0, \tfrac{1}{4}\right), \quad &&P_{\langle 0,2,1\rangle} = \left(\tfrac{1}{4}, \tfrac{5-\sqrt{5}}{8}\right), \quad &&P_{\langle 2,0,1\rangle} = \left(\tfrac{5-\sqrt{5}}{8}, \tfrac{3}{4}\right), \\ &P_{\langle 1,0,2\rangle} = \left(\tfrac{3}{4}, \tfrac{5+\sqrt{5}}{8}\right), \quad &&P_{\langle 1,2,0\rangle} = \left(\tfrac{5+\sqrt{5}}{8}, 1\right). \end{aligned} \tag{1.29}$$

The separating points of the intervals P_π, $\pi \in \mathcal{S}_3$, are given now by the conditions $g^{\pi_i}(x) = g^{\pi_{i+1}}(x)$, $\pi_i, \pi_{i+1} \in \{0, 1, 2\}$, i.e.,

$$g(x) = x, \; g^2(x) = x, \; g^2(x) = g(x).$$

We conclude that the common endpoints of the intervals P_π for $\pi \in \mathcal{S}_3$ are now the points of period 1 (fixed points), period 2, and first preimages of period-1 points. Moreover, when going from $L = 2$ to $L = 3$, we see that $P_{\langle 0,1\rangle}$ splits into the subintervals $P_{\langle 0,1,2\rangle}$, $P_{\langle 0,2,1\rangle}$, and $P_{\langle 2,0,1\rangle}$ at the eventually period-1 point $\frac{1}{4}$ (preimage of the fixed point $\frac{3}{4}$) and at the period-2 point $\frac{5-\sqrt{5}}{8}$. Likewise, $P_{\langle 1,0\rangle}$ splits into $P_{\langle 1,0,2\rangle}$ and $P_{\langle 1,2,0\rangle}$ at the period-2 point $\frac{5+\sqrt{5}}{8}$.

Ordinal patterns are the main ingredient of *permutation entropy* which, as the standard concept of entropy, comes also in metric and topological versions.

Suppose that μ is an f-invariant measure. Then the definition of the *metric permutation entropy* of f is formally similar to the definition of the Shannon entropy of an information source:

$$h^*_\mu(f) = -\lim_{L\to\infty} \frac{1}{L} \sum_{\pi \in \mathcal{S}_L} \mu(P_\pi) \log \mu(P_\pi), \tag{1.30}$$

provided the limit exists. Note that $\mu(P_\pi)$ is the probability for the ordinal L-pattern π to occur (while in the expression for the Shannon entropy, (1.1), the corresponding probabilities refer to length-L blocks x_0^{L-1}). Sometimes the factor $1/(L-1)$ is used instead of $1/L$—of course, this is inconsequential in the limit $L \to \infty$.

As for the *topological permutation entropy* of f, one just counts distinct allowed patterns:

$$h^*_{\text{top}}(f) = -\lim_{L\to\infty} \frac{1}{L} \log |\{P_\pi \neq \emptyset : \pi \in \mathcal{S}_L\}|, \tag{1.31}$$

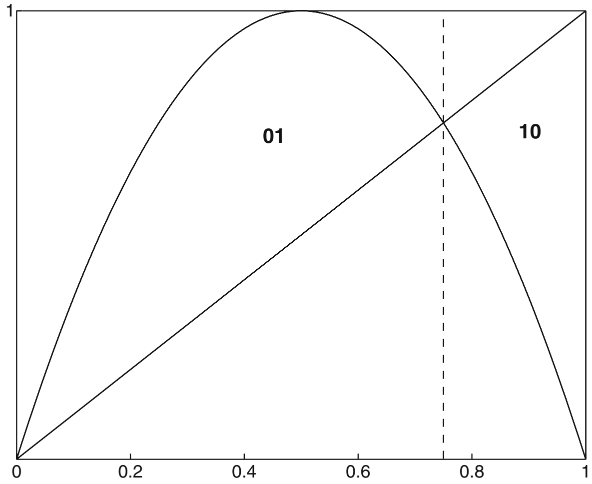

Fig. 1.5 Points in the interval $(0, \frac{3}{4})$ are of type $\langle 0, 1 \rangle$ (shorthanded 01), while points in the interval $(\frac{3}{4}, 0)$ are of type $\langle 1, 0 \rangle$ (shorthanded 10)

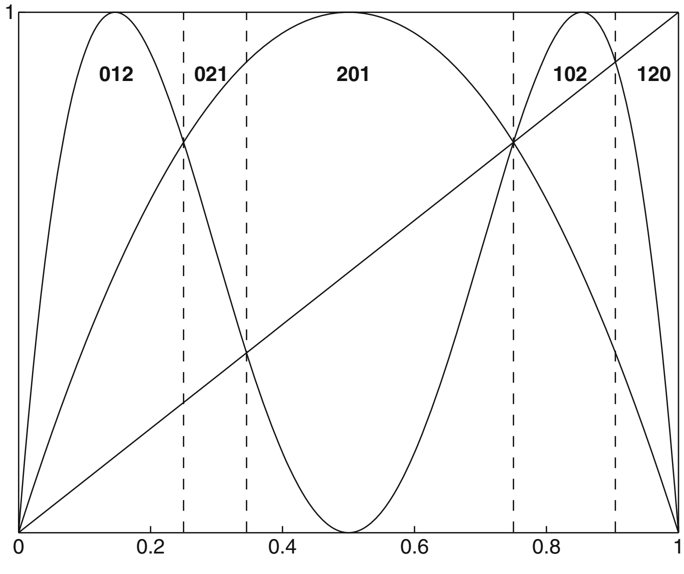

Fig. 1.6 The sets P_π, $\pi \in \mathcal{S}_3$, are graphically obtained by raising vertical lines at the crossing points of the curves $y = x$, $y = f(x)$, and $y = f^2(x)$. The three digits on the upper part of the figure are shorthand for ordinal patterns (e.g., 012 stands for $\langle 0, 1, 2 \rangle$). Observe that $P_{\langle 2,1,0 \rangle} = \emptyset$

where $|\cdot|$ denotes here cardinality. We are assuming again that this limit converges, otherwise $h^*_{\text{top}}(f)$ is not defined.

An interval map $f{:}I \to I$ is called *piecewise monotone* if there is a finite partition of I into intervals, such that f is continuous and monotone on each of those intervals. A nice result of Bandt, Keller, and Pompe [29] states that if f is piecewise monotone, then (i) the metric permutation entropy of f coincides with its metric entropy and (ii) the topological permutation entropy of f coincides with its topological entropy. In mathematical notation:

$$\text{(i) } h^*_\mu(f) = h_\mu(f) \quad \text{and} \quad \text{(ii) } h^*_{\text{top}}(f) = h_{\text{top}}(f). \tag{1.32}$$

From (ii) and (1.31), it follows that if f is piecewise monotone and its topological entropy is finite, then

$$|\{P_\pi \neq \emptyset{:}\pi \in \mathcal{S}_L\}| \sim e^{L h_{\text{top}}(f)}, \tag{1.33}$$

where the symbol $\sim$ stands for "asymptotically as $L \to \infty$." Hence, the number of allowed L-patterns for f grows exponentially with L. On the other hand,

$$|\{P_\pi{:}\pi \in \mathcal{S}_L\}| = L! \sim e^{L(\ln L - 1) + 1/2 \ln 2\pi L}, \tag{1.34}$$

according to Stirling's formula for the factorial of a positive integer. Comparison of (1.33) and (1.34) not only does show that piecewise monotone maps have necessarily forbidden L-patterns for L sufficiently large but also that their number grows superexponentially with L.

From (1.29) we see that already for $L = 3$ there is one forbidden pattern for the logistic map, namely, $\langle 2, 1, 0 \rangle$. But this is not the end of the story. The absence of the ordinal pattern $\pi = \langle 2, 1, 0 \rangle$ triggers, in turn, an avalanche of longer missing patterns. To begin with, all the patterns $\langle *, 2, *, 1, *, 0, * \rangle$ (where the wildcard $*$ stands eventually for any other entries of the pattern) cannot be realized by any $x \in [0, 1]$ since the inequalities

$$\cdots < g^2(x) < \cdots < g(x) < \cdots < x < \cdots \tag{1.35}$$

cannot occur. By the same token, the patterns $\langle *, 3, *, 2, *, 1, * \rangle$, $\langle *, 4, *, 3, *, 2, * \rangle$, and, more generally,

$$\langle *, n+2, *, n+1, *, n, * \rangle \in \mathcal{S}_L,\ 0 \le n \le L-3, \tag{1.36}$$

cannot be realized either for the same reason (replace x by $g^n(x)$ in (1.35)). We conclude that each forbidden pattern generates an infinite trail of ever-longer forbidden patterns. This issue will be revisited in full generality in Chap. 3.

Let us clarify this last point with the logistic map once more and $L = 4$. In Fig. 1.7, which is Fig. 1.6 with the curve

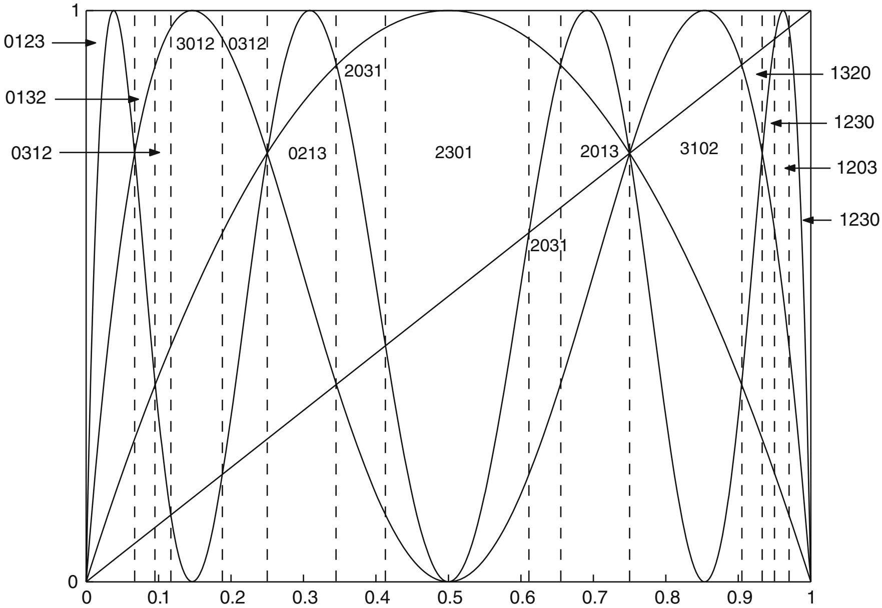

Fig. 1.7 The 12 allowed ordinal 4-patterns for the logistic map. Note the two components of $P_{\langle 0,3,1,2\rangle}$, $P_{\langle 2,0,3,1\rangle}$, and $P_{\langle 1,2,3,0\rangle}$

$$\begin{aligned} y &= g^3(x) \\ &= -16\,384x^8 + 65\,536x^7 - 106\,496x^6 + 90\,112x^5 \\ &\quad -42\,240x^4 + 10\,752x^3 - 1344x^2 + 64x \end{aligned}$$

superimposed, we can see the 12 allowed 4-patterns for the logistic map. Since there are 24 possible patterns of length 4, we conclude that 12 of them are forbidden. Seven forbidden 4-patterns belong to trail (1.36) of $\langle 2,1,0\rangle$ (observe that $\langle 3,2,1,0\rangle$ is repeated):

$$\begin{array}{ll} (n=0) & \langle 3,2,1,0\rangle, \langle 2,3,1,0\rangle, \langle 2,1,3,0\rangle, \langle 2,1,0,3\rangle \\ (n=1) & \langle 0,3,2,1\rangle, \langle 3,0,2,1\rangle, \langle 3,2,0,1\rangle, \langle 3,2,1,0\rangle \end{array}. \tag{1.37}$$

Therefore, the remaining five forbidden 4-patterns,

$$\langle 0,2,3,1\rangle, \langle 1,0,2,3\rangle, \langle 1,0,3,2\rangle, \langle 1,3,0,2\rangle, \langle 3,1,2,0\rangle, \tag{1.38}$$

are seeds for new trails of forbidden patterns of lengths $L \geq 5$ that eventually can overlap.

In Fig. 1.7 one can also follow the first two splittings of the intervals P_π:

$$P_{\langle 0,1\rangle} \to \begin{cases} P_{\langle 0,1,2\rangle} \to P_{\langle 0,1,2,3\rangle}, P_{\langle 0,1,3,2\rangle}, P_{\langle 0,3,1,2\rangle}, P_{\langle 3,0,1,2\rangle}, \\ P_{\langle 0,2,1\rangle} \to P_{\langle 0,2,1,3\rangle}, \\ P_{\langle 2,0,1\rangle} \to P_{\langle 2,0,1,3\rangle}, P_{\langle 2,0,3,1\rangle}, P_{\langle 2,3,0,1\rangle}, \end{cases}$$

$$P_{\langle 1,0\rangle} \to \begin{cases} P_{\langle 1,0,2\rangle} \to P_{\langle 3,1,0,2\rangle}, \\ P_{\langle 1,2,0\rangle} \to P_{\langle 1,2,0,3\rangle}, P_{\langle 1,2,3,0\rangle}, P_{\langle 1,3,2,0\rangle}. \end{cases}$$

The splitting of the intervals P_π can be understood in terms of periodic points and their preimages. Thus, the splitting of $P_{\langle 0,1\rangle}$ is due to the points $\frac{1}{4}$ (first preimage of the period-1 point $\frac{3}{4}$) and $\frac{5-\sqrt{5}}{8}$ (a period-2 point); the second period-2 point, $\frac{5-\sqrt{5}}{8}$, is responsible for the splitting of $P_{\langle 1,0\rangle}$. On the contrary, $P_{\langle 0,2,1\rangle}$ and $P_{\langle 1,0,2\rangle}$ do not split because they contain neither period-3 point nor first preimages of period-2 points nor second preimages of fixed points.