INFORMATION VS DIMENSION – AN ALGORITHMIC PERSPECTIVE

ABSTRACT. This paper surveys work on the relation between fractal dimensions and algorithmic information theory over the past thirty years. It covers the basic development of prefix-free Kolmogorov complexity from an information theoretic point of view, before introducing Hausdorff measures and dimension along with some important examples. The main goal of the paper is to motivate and develop the informal identity "entropy = complexity = dimension" from first principles. The last section of the paper presents some new observations on multifractal measures from an algorithmic viewpoint.

1. INTRODUCTION AND PRELIMINARIES

Starting with the work by Ryabko [41, 41] and Staiger [47, 48, 49] in the 1980s, over the past 30 years researchers have investigated a strong relation between fractal dimension and algorithmic information theory. At the center of this relation is a *pointwise* version of Hausdorff dimension (due to Lutz [29, 30]), that is, a notion of dimension that is defined for individual points in a space instead of subsets. This is made possible by effectivizing the notion of measure (in the sense of Martin-Löf [33]), which restricts the collection of nullsets to a countable family, thereby allowing for singleton sets not to be null. Such singletons are considered *random* in this framework. The correspondence between randomness and the various flavors of Kolmogorov complexity (a cornerstone of the theory of algorithmic randomness) then re-emerges in form of an asymptotic information density that bears close resemblance to the entropy of a stationary process. Effective dimension makes the connections between entropy and fractal dimension that have been unearthed by many authors (arguably starting with Billingsley [2] and Furstenberg [17]) very transparent. Moreover, effective dimension points to new ways to compute fractal dimensions, by investigating the algorithmic complexities of single points in a set (e.g. Lutz and Lutz [31]), in particular when studied in relation to other *pointwise* measures of complexity, such as irrationality exponents in diophantine approximation [1].

This survey tries to present these developments in one coherent narrative. Throughout, we mostly focus on Hausdorff dimension, and on a small fraction of results on effective dimension and information, in order to streamline the presentation. For other notions such as packing or box counting dimension, refer to [15, 10] for the classical theory, and to [37, 53, 13, 52] for many results concerning effective dimension.

The outline of the paper is as follows. In Section 2, we cover the basic theory of information measures, introducing entropy and Kolmogorov complexity, highlighting the common foundations of both. In Section 3, we briefly introduce Hausdorff measures and dimension, along with some fundamental examples. Arguably the central section of the paper, Section 4 develops the theory of effective dimension, establishes the connection with asymptotic Kolmogorov complexity, and describes the fundamental instance of entropy vs dimension – the Shannon-Macmillan-Breiman Theorem. Finally, in Section 5, we present some new observations concerning multifractal measures. It turns out that the basic idea of pointwise dimension emerges here, too.

It is the goal of this paper to be accessible for anyone with a basic background in computability theory and measure theory. Where the proofs are easy to give and of reasonable length, we include them. In all other cases we refer to the literature.

2. Information Measures and Entropy

Suppose X is a finite, non-empty set. We choose an element $a \in X$ and want to communicate our choice to another person, through a binary channel, i.e. we can only transmit bits 0 and 1. How many bits are needed to transmit our choice, if nothing else is known about X but its cardinality? To make this possible at all, we assume that we have agreed on a numbering of $X = \{a_1, \ldots, a_n\}$ that is known to both parties. We can then transmit the *binary code* for the index of a, which requires at most $\log_2 n = \log_2 |X|$ bits. Of course, if we had an N-ary channel available, we would need $\log_N |A|$ bits. We will mostly work with binary codes, so log will denote the binary logarithm $\log_2$.

Often our choice is guided by a probabilistic process, i.e. we have X given as a *discrete random variable* X has countably many outcomes, and we denote the range of X by $\{a_0, a_1, a_2, \ldots\}$. Suppose that we repeatedly choose

from X at random and want to communicate our choice to the other party. Arguably the central question of information theory is:

> *How do we minimize the expected number of bits we need to transmit to communicate our choice?*

Depending on the distribution of the random variable, we can tweak our coding system to optimize the length of the code words.

Definition 2.1. A *binary code* for a countable set X is a one-one function $c : \{x_i \colon i \in \mathbb{N}\} \to 2^{<\mathbb{N}}$.

Here, $2^{<\mathbb{N}}$ denotes the set of all finite binary strings. In the non-probabilistic setting, when X is a finite set, we essentially use a *fixed-length code* $c : X \to \{0,1\}^n$, by mapping

$$x_i \mapsto \text{ binary representation of } i.$$

In the probabilistic setting, if the distribution of the random variable X does not distinguish between the outcomes, i.e. if X is equidistiributed, this will also be the best we can do.

However, if the possible choices have very different probabilities, we could hope to save on the *expected* code word length

$$\sum_i \text{len}(c(a_i)) \cdot \mathbb{P}(X = a_i),$$

where $\text{len}(c(a_i))$ is the length of the string $c(a_i)$, by assigning shorts code words to a of high probability. The question then becomes how small the expected code word length can become, and how we would design a code to minimize it.

Another way to approach the problem of information transmission is by trying to measure the *information gained* by communicating only partial information. Suppose our random variable X is represented by a partition

$$[0,1] = \dot{\bigcup_i} X_i,$$

where each X_i is an interval such that $\mathbb{P}(X = a_i) = |X_i|$, i.e. the lengths of the intervals mirror the distribution of X. Pick, randomly and uniformly, an $x \in [0,1]$. If we know which X_i x falls in, we gain knowledge of about the first $-\log |X_i|$ bits of the binary expansion of x. Therefore, on average, we gain

$$H(X) = -\sum \mathbb{P}(X = a_i) \cdot \log \mathbb{P}(X = a_i) = -\sum |X_i| \cdot \log |X_i|$$

bits of information. We put $0 \cdot \log 0 =: 0$ to deal with $\mathbb{P}(X = a_i) = 0$. $H(X)$ is called the *entropy* of X. We will apply H not only to random variables, but also to measures in general. If $\vec{p} = (p_1, \ldots, p_n)$ is a finite probability vector, i.e. $p_i \geq 0$ and $\sum_i p_i = 1$, then we write $H(\vec{p})$ to denote $\sum_i p_i \cdot \log(p_i)$. Similarly, for $p \in [0, 1]$, $H(p)$ denotes $p \cdot \log p + (1 - p) \cdot \log(1 - p)$.

The emergence of the term $-\log \mathbb{P}$ is no coincidence. Another, more axiomatic way to derive it is as follows. Suppose we want to measure the information gain of an event by a function $I : [0, 1] \to \mathbb{R}^{\geq 0}$, that is, I depends only on the probability of an event. We require the following properties of I:

(I1) $I(1) = 0$. An almost sure event gives us no new information.
(I2) I is decreasing. The lower the probability of an event, the more information we gain from the knowledge that it occurred.
(I3) $I(x \cdot y) = I(x) + I(y)$. If X and Y are independent, then $\mathbb{P}(X \cap Y) = \mathbb{P}(X) \cdot \mathbb{P}(Y)$, and hence $I(\mathbb{P}(X \cap Y)) = I(\mathbb{P}(X)) + I(\mathbb{P}(Y))$. In other words, information gain is additive for independent events.

Proposition 2.2. *If $I : [0, 1] \to \mathbb{R}^{\geq 0}$ is a function satisfying (I1)-(I3), then there exists a constant c such that*

$$I(x) = -c \log(x).$$

In this sense, $-\log$ is the only reasonable information function, and the entropy H measures the expected gain in information.

It is also possible to axiomatically characterize entropy directly. Many such characterizations have been found over the years, but the one by Khinchin [23] is arguably still the most popular one.

We will see next how the entropy of X is tied to the design of an optimal code.

2.1. **Optimal Prefix Codes.** We return to the question of how to find an optimal code, i.e. how to design a binary code for a random variable X such that its average code word length is minimal. While a binary code guarantees that a single code word can be uniquely decoded, we would like to have a similar property for *sequences of code words*.

Definition 2.3. Given a binary code c for a countable set X, its *extension* c^* to $X^{<\mathbb{N}}$ is defined as

$$c^*(x_0 x_1 \ldots x_n) = c(x_0)^\frown c(x_1)^\frown \ldots {}^\frown c(x_n),$$

where $\frown$ denotes concatenation of strings. We say a binary code c is *uniquely decodable* if its extension is one-one.

One way to ensure unique decodability is to make c *prefix free*. A set $S \subseteq 2^{<\mathbb{N}}$ is prefix free if no two elements in S are prefixes of one another. A code is prefix free if its range is. It is not hard to see that the extension of a prefix free code is indeed uniquely decodable.

Binary codes live in $2^{<\mathbb{N}}$. The set $2^{<\mathbb{N}}$ is naturally connected to the space of infinite binary sequences $2^{\mathbb{N}}$. We can put a metric on $2^{\mathbb{N}}$ by letting

$$d(x, y) = \begin{cases} 2^{-N} & \text{if } x \neq y \text{ and } N = \min i \colon x(i) \neq y(i) \\ 0 & \text{otherwise.} \end{cases}$$

Given $\varepsilon > 0$, the ε-ball $B(x, \varepsilon)$ around x is the so-called *cylinder set*

$$[\![\sigma]\!] = \{y \in 2^{\mathbb{N}} \colon \sigma \subset y\},$$

where $\subset$ denotes the prefix relation between strings (finite or infinite), and $\sigma = x{\restriction}_n$ is the length-n initial segment of x with $n = \lceil -\log \varepsilon \rceil$.

Hence any string σ corresponds to a basic open cylinder $[\![\sigma]\!]$ with diameter $2^{-\operatorname{len}(\sigma)}$. This induces a Borel measure λ on $2^{\mathbb{N}}$, $\lambda[\![\sigma]\!] = 2^{-\operatorname{len}(\sigma)}$, which is the natural analogue to Lebesgue measure on $[0, 1]$.

Which code lengths are possible for prefix free codes? The *Kraft inequality* gives a fundamental restraint.

Theorem 2.4 (Kraft inequality)**.** *Let $S \subseteq 2^{<\mathbb{N}}$ be prefix free. Then*

$$\sum_{\sigma \in S} 2^{-\operatorname{len}(\sigma)} \leq 1.$$

Conversely, given any sequence $l_0, l_1, l_2, \ldots$ of non-negative integers satisfying

$$\sum_i 2^{-l_i} \leq 1,$$

there exists a prefix-free set $\{\sigma_0, \sigma_1, \ldots\} \subseteq 2^{<\mathbb{N}}$ of code words such that $\operatorname{len}(\sigma_i) = l_i$.

Proof sketch. ($\Rightarrow$) Any prefix free set S corresponds to a disjoint union of open cylinders

$$U = \dot{\bigcup_{\sigma \in S}} [\![\sigma]\!].$$

The Lebesgue measure λ of $[\![\sigma]\!]$ is $2^{-\operatorname{len}(\sigma)}$. Hence by the additivity of measures,

$$1 \geq \lambda(U) = \sum_{\sigma \in S} 2^{-\operatorname{len}(\sigma)}.$$

($\Longleftarrow$) We may assume the l_i are given in non-decreasing order:

$$l_0 \leq l_1 \leq \ldots,$$

which implies the sequence (2^{-l_i}) is non-increasing. We construct a prefix code by "shaving off" cylinder sets from the left. We choose as σ_0 the leftmost string of length l_0, i.e. 0^{l_0}. Suppose we have chosen $\sigma_0, \ldots, \sigma_k$ such that $\operatorname{len}(\sigma_i) = l_i$ for $i = 0, \ldots, k$. The measure we have left is

$$1 - (2^{-l_0} + \cdots + 2^{-l_k}) \geq 2^{-l_{k+1}},$$

since $\sum_i 2^{-l_i} \leq 1$. And since $2^{-l_0} \geq \cdots \geq 2^{-l_k} \geq 2^{-l_{k+1}}$, the remaining measure is a multiple of $2^{-l_{k+1}}$. Therefore, we can pick σ_{k+1} to be the leftmost string of length l_{k+1} that does not extend any of the $\sigma_0, \ldots, \sigma_k$.

This type of code is known as the *Shannon-Fano code*.

$\square$

Subject to this restraint, what is the minimum expected length of a code for a random variable X? Suppose X is discrete and let $p_i = \mathbb{P}(X = x_i)$. We want to find code lengths l_i such that

$$L = \sum p_i l_i$$

is minimal, where the l_i have to satisfy

$$\sum 2^{-l_i} \leq 1,$$

the restriction resulting from the Kraft inequality.

The following result, due to Shannon, is fundamental to information theory.

Theorem 2.5. *Let L be the expected length of a binary prefix code of a discrete random variable X. Then*

$$L \geq H(X).$$

Furthermore, there exists a prefix code for X whose expected code length does not exceed $H(X) + 1$.

To see why no code can do better than entropy, note that

$$L - H(X) = \sum p_i l_i + \sum p_i \log(p_i)$$
$$= -\sum p_i \log 2^{-l_i} + \sum p_i \log(p_i).$$

Put $c = \sum 2^{-l_i}$ and $r_i = 2^{-l_i}/c$. $\vec{r}$ is the probability distribution induced by the code lengths. Then

$$L - H(X) = \sum p_i \log p_i - \sum p_i \log \left(\frac{2^{-l_i}}{c} c \right)$$
$$= \sum p_i \log \frac{p_i}{r_i} - \log c$$

Note that $c \leq 1$ and hence $\log c \leq 0$. The expression

$$\sum p_i \log \frac{p_i}{r_i}$$

is known as the *relative entropy* or *Kulback-Leibler (KL) distance* $D(\vec{p} \parallel \vec{r})$ of distributions $\vec{p}$ and $\vec{r}$. It is not actually a distance as it is neither symmetric nor does it satisfy the triangle inequality. But it nevertheless measures the difference between distributions in some important aspects. In particular, $D(\vec{p} \parallel \vec{r}) \geq 0$ and is equal to zero if and only if $\vec{p} = \vec{r}$. This can be inferred from Jensen's inequality.

Theorem 2.6 (Jensen's inequality). *If f is a convex function and X is a random variable, then*

$$\mathbb{E}f(X) \geq f(\mathbb{E}X).$$

In particular, it follows that $L - H(X) \geq 0$. This can also be deduced directly using the concavity of the logarithm in the above sums, but both KL-distance and Jensen's inequality play a fundamental, almost "axiomatic" role in information theory, so we reduced the problem to these two facts.

To see that there exists a code with average code length close to entropy, let

$$l_i := \lceil \log p_i \rceil.$$

Then $\sum 2^{-l_i} \leq \sum p_i = 1$, and by Theorem 2.4, there exists a prefix code $\vec{\sigma}$ such that $\text{len}(\sigma_i) = l_i$. Hence $x_i \mapsto \sigma_i$ will be a prefix code for X with expected code word length

$$\sum p_i l_i = \sum p_i \lceil \log p_i \rceil \leq \sum p_i (\log p_i + 1) = H(X) + 1.$$

2.2. **Effective coding.** The transmission of information is an inherently computational act. Given X, we should be able to assign code words to outcomes of X effectively. In other words, if $\{x_i\}$ collects the possible outcomes of X, and $\{\sigma_i\}$ is a prefix code for X, then the mapping $i \mapsto \sigma_i$ should be computable.

The construction of a code is implicit in the proof of the Kraft inequality (Theorem 2.4), and is already of a mostly effective nature. The only non-effective part is the assumption that the word lengths l_i are given in increasing order. However, with careful book keeping, one can assign code words in the order they are requested. This has been observed independently by Levin [25], Schnorr [44], Chaitin [6].

Theorem 2.7 (Effective Kraft inequality). *Let $L : \mathbb{N} \to \mathbb{N}$ be a computable mapping such that*

$$\sum_{i \in \mathbb{N}} 2^{-L(i)} \leq 1.$$

Then there exists a computable, one-one mapping $c : \mathbb{N} \to 2^{<\mathbb{N}}$ such that the image of c is prefix free, and for all i, $\mathrm{len}(c(i)) = L(i)$.

A proof can be found, for example, in the book by Downey and Hirschfeldt [13]. Let us call c an effective prefix-code. As c is one-one, given $\sigma \in \mathrm{range}(c)$, we can compute i such that $c(i) = \sigma$. And since the image of c is prefix-free, if we are given a sequence of concatenated code words

$$\sigma_1 {}^\frown \sigma_2 {}^\frown \sigma_3 \cdots = c(i_1) {}^\frown c(i_2) {}^\frown c(i_3) {}^\frown \ldots,$$

we can effectively recover the source sequence $i_1, i_2, i_3 \ldots$. Hence, every effective prefix code c comes with an effective *decoder*.

Let us therefore switch perspective now: Instead of mapping events to code words, we will look at functions mapping code words to events. An *effective prefix decoder* is a partial computable function d whose domain is a prefix-free subset of $2^{<\mathbb{N}}$. Since every partial recursive function is computable by a Turing machine, we will just use the term *prefix-free machine* to denote effective prefix decoders. The range of a prefix-free machine will be either $\mathbb{N}$ or $2^{<\mathbb{N}}$. Since we can go effectively from one set to the other (by interpreting a binary string as the dyadic representation of a natural number or by enumerating $2^{<\mathbb{N}}$ effectively), in this section it does not really matter which one we are using. In later sections, however, when we look at infinite binary sequences, it will be convenient to have $2^{<\mathbb{N}}$ as the range.

If M is a prefix-free machine and $M(\sigma) = m$, we can think of σ as an M-code of m. Note that by this definition, m can have multiple codes, which is not provided for in our original definition of a binary code (Definition 2.1). So a certain asymmetry arises when switching perspective. We can eliminate this asymmetry by considering only the *shortest* code. The length of this code will then, with respect to the given machine, tell us how far the information in σ can be *compressed*.

Definition 2.8. Let M be a prefix-free machine. The M-complexity $\mathrm{K}_M(m)$ of a string σ is defined as

$$\mathrm{K}_M(m) = \min\{\mathrm{len}(\sigma) \colon M(\sigma) = m\},$$

where $\mathrm{K}_M(m) = \infty$ if there is no σ with $M(\sigma) = m$.

K_M may not be computable anymore, as we are generally not able to decide whether M will halt on an input σ and output m. But we can *approximate* K_M *from above*: Using a bootstrapping procedure, we run M simultaneously on all inputs. Whenever we see that $M(\sigma) \downarrow = m$, we compare σ with our current approximation of $\mathrm{K}_M(m)$. If σ is shorter, $\mathrm{len}(\sigma)$ is our new approximation of $\mathrm{K}_M(m)$. Eventually, we will have found the shortest M-code for m, but in general, we do not know whether our current approximation is actually the shortest one.

Alternatively, we can combine the multiple codes into a single one probabilistically. Every prefix-free machine M induces a distribution on $\mathbb{N}$ by letting

$$(2.1) \qquad Q_M(m) = \sum_{M(\sigma)\downarrow = m} 2^{-\mathrm{len}(\sigma)}.$$

Q_M is not necessarily a probability distribution, however, for it is not required that the domain of M covers all of $2^{<\mathbb{N}}$. Nevertheless, it helps to think of $Q_M(m)$ as the probability that M on a randomly chosen input halts and outputs m.

Any function $Q : \mathbb{N} \to \mathbb{R}^{\geq 0}$ with

$$\sum_m Q(m) \leq 1$$

is called a *discrete semimeasure*. The term 'discrete' is added to distinguish these semimeasures from the *continuous semimeasures* which will be introduced in Section 5. In this section, all semimeasures are discrete, so we will refer to them just as 'semimeasures'.

If a semimeasure Q is induced by a prefix-free machine, it moreover has the property that for any m, the set

$$\{q \in \mathbb{Q} \colon q < Q(m)\}$$

is recursively enumerable, uniformly in m. In general, semimeasures with this property are called *left-enumerable*.

We can extend the effective Kraft Inequality to left-enumerable semimeasures. We approximate a code enumerating better and better dyadic lower bounds for Q. As we do not know the true value of Q, we cannot quite match the code lengths prescribed by Q, but only up to an additive constant.

Theorem 2.9 (Coding Theorem). *Suppose Q is a left-enumerable semimeasure. Then there exists a prefix-free machine M_Q such that for some constant d and for all $m \in \mathbb{N}$,*

$$\mathrm{K}_{M_Q}(m) \leq - \log Q(m) + d$$

For a proof see Li and Vitányi [26]. Since the shortest M-code contributes $2^{-\mathrm{K}_M(m)}$ to $Q_M(m)$, it always holds that

$$\mathrm{K}_{M_Q}(m) \leq \mathrm{K}_M(m) + d$$

for all m and for some constant d. However, if we consider universal machines, the two processes will essentially yield the same code.

Definition 2.10. A prefix-free machine U is *universal* if for each partial computable prefix-free function $\varphi : 2^{<\mathbb{N}} \to \mathbb{N}$, there is a string σ_φ such that for all τ,

$$U(\sigma_\varphi {}^\frown \tau) = \varphi(\tau)$$

Universal prefix-free machines exist. The proof is similar to the proof that universal (plain) Turing machines exist (see e.g. [13]).

Fix a universal prefix-free machine U and define

$$\mathrm{K}(m) = \mathrm{K}_U(m).$$

$\mathrm{K}(m)$ is called the *prefix-free Kolmogorov complexity* of m. The universality of U easily implies the following.

Theorem 2.11 (Invariance Theorem). *For any prefix-free machine M there exists a constant d_M such that for all m,*
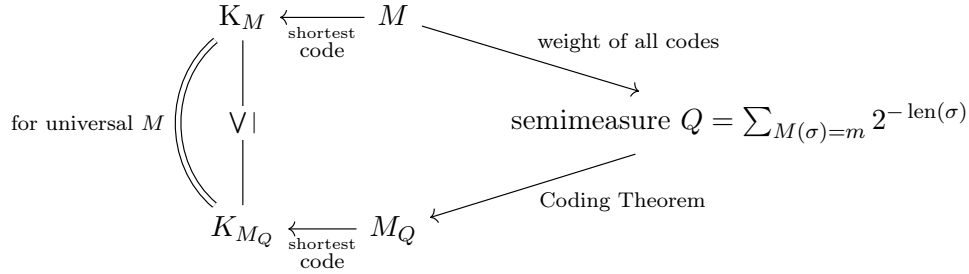
$$\mathrm{K}(m) \leq \mathrm{K}_M(m) + d_M.$$

FIGURE 1. Relating complexity and semimeasures

Theorems 2.9 and 2.11 together imply that up to an additive constant,

$$\mathrm{K} = \mathrm{K}_U \leq \mathrm{K}_{Q_U} \leq \mathrm{K}_U = \mathrm{K},$$

that is, up to an additive constant, *both coding methods agree* (see Figure 1).

The Coding Theorem also yields a relation between K and Shannon entropy H. Suppose X is a random variable with a computable distribution p. As K is the length of a prefix code on $2^{<\mathbb{N}}$, Shannon's lower bound (Theorem 2.5) implies that $H(X) \leq \sum_m \mathrm{K}(m)p(m)$. On the other hand, it is not hard to infer from the Coding Theorem that, up to a multiplicative constant, $2^{-\mathrm{K}(m)}$ dominates every computable probability measure on $\mathbb{N}$. Therefore,

$$H(X) = \sum_m p(m)(-\log p(m)) \geq \sum_m p(m)\,\mathrm{K}(m) - d$$

for some constant d depending only on p. In fact, the constant d depends only on the length of a program needed to compute p (which we denote $\mathrm{K}(p)$).

Theorem 2.12 (Cover, Gács, and Gray [7])**.** *There exists a constant d such that for every discrete random variable X with computable probability distribution p on $\mathbb{N}$ with finite Shannon entropy $H(X)$,*

$$H(X) \leq \sum_\sigma \mathrm{K}(\sigma)p(\sigma) \leq H(X) + \mathrm{K}(p) + d.$$

In other words, probabilistic entropy of a random variable X taking outcomes in N is equal, within an additive constant, to the expected value of complexity of these outcomes. This becomes interesting if $\mathrm{K}(p)$ is rather low compared to the length of strings p puts its mass on, for example, if p is the uniform distribution on strings of length n.

3. Hausdorff dimension

The origins of Hausdorff dimension are geometric. It is based on finding an optimal scaling relation between the diameter and the measure of a set.

Let (Y, d) be a separable metric space, and suppose s is a non-negative real number. Given a set $U \subseteq Y$, the diameter of U is given as

$$d(U) = \sup\{d(x, y) \colon x, y \in U\}.$$

Given $\delta > 0$, a family $(U_i)_{i \in I}$ of subsets of Y with $d(U_i) \leq \delta$ for all $i \in I$ is called a δ-*family*.

The *s-dimensional outer Hausdorff measure* of a set $A \subseteq Y$ is defined as

$$\mathcal{H}^s(A) = \lim_{\delta \to 0} \left(\inf \left\{ \sum d(U_i)^s \colon A \subseteq \bigcup U_i, \ (U_i) \ \delta\text{-family} \right\} \right).$$

Since there are fewer δ-families available as $\delta \to 0$, the limit above always exists, but may be infinite. If $s = 0$, $\mathcal{H}$ is a counting measure, returning the number of elements of A if A is finite, or ∞ if A is infinite. If Y is a Euclidean space $\mathbb{R}^n$, then $\mathcal{H}^n$ coincides with Lebesgue measure, up to a multiplicative constant.

It is not hard to see that one can restrict the definition of $\mathcal{H}^s$ to δ-families of open sets and still obtain the same value for $\mathcal{H}^s$.

In $2^{\mathbb{N}}$, it suffices to consider only δ-families of cylinder sets, since for covering purposes any set $U \subseteq 2^{\mathbb{N}}$ can be replaced by $[\![\sigma]\!]$, where σ is the longest string such that $\sigma \subset x$ for all $x \in U$ (provided U has more than one point). Then $d([\![\sigma]\!]) = d(U)$. It follows that δ-covers in $2^{\mathbb{N}}$ can be identified with sets of strings $\{\sigma_i \colon i \in \mathbb{N}\}$, where $\operatorname{len}(\sigma_i) \geq \lceil \log \delta \rceil$ for all i.

In $\mathbb{R}$, one can obtain a similar simplification by considering the *binary net measure* $\mathcal{M}^s$, which is defined similar to $\mathcal{H}^s$, but only coverings by dyadic intervals of the form $[r2^{-k}, (r+1)2^{-k})$ are permitted. For the resulting outer measure $\mathcal{M}^s$, it holds that, for any set $E \subseteq \mathbb{R}$,

$$(3.1) \qquad\qquad \mathcal{H}^s E \leq \mathcal{M}^s E \leq 2^{s+1} \mathcal{H}^s E.$$

Instead of dyadic measures, one may use other g-adic net measures and obtain a similar relation as in (3.1), albeit with a different multiplicative constant (see e.g. [15]).

For each s, $\mathcal{H}$ is a *metric outer measure*, that is, if $A, B \subseteq Y$ have positive distance $(\inf\{d(x, y) \colon x \in A, y \in B\} > 0)$, then

$$\mathcal{H}^s(A \cup B) = \mathcal{H}^s(A) + \mathcal{H}^s(B).$$

This implies that all Borel sets are $\mathcal{H}^s$-measurable [40].

It is not hard to show that

$$(3.2) \qquad \mathcal{H}^s(A) < \infty \;\Rightarrow\; \mathcal{H}^t(A) = 0 \text{ for all } t > s,$$

and

$$(3.3) \qquad \mathcal{H}^s(A) > 0 \;\Rightarrow\; \mathcal{H}^t(A) = \infty \text{ for all } t < s.$$

Hence the number

$$\dim_H(A) = \inf\{s \colon \mathcal{H}(A) = 0\} = \sup\{s \colon \mathcal{H}(A) = \infty\}$$

represents the "optimal" scaling exponent (with respect to diameter) for measuring A. $\dim_H(A)$ is called the *Hausdorff dimension* of A. By (3.1), working in $\mathbb{R}$, we can replace $\mathcal{H}^s$ by the binary net measure $\mathcal{M}^s$ in the definition of Hausdorff dimension. On of the theoretical advantages of Hausdorff dimension over other concepts that measure fractal behavior (such as box counting dimension) is its *countable stability*.

Proposition 3.1. *Given a countable family $\{A_i\}$ of sets, it holds that*

$$\dim_H \bigcup A_i = \sup\left\{\dim_H A_i \colon i \in \mathbb{N}\right\}.$$

Examples of Hausdorff dimension.

(1) Any finite set has Hausdorff dimension zero.

(2) Any set of non-zero n-dimensional Lebesgue measure in $\mathbb{R}^n$ has Hausdorff dimension n. Hence all "regular" geometric objects of full topological dimension (e.g. squares in $\mathbb{R}^2$, cylinders in $\mathbb{R}^3$) have full Hausdorff dimension, too.

(3) The *middle-third Cantor set*

$$C = \{x \in [0,1] \colon x = \sum_{i \geq 1} t_i \cdot 3^{-i},\ t_i \in 0, 2\}$$

has Hausdorff dimension $\ln(2)/\ln(3)$.

(4) Inside Cantor space, we can define a "Cantor-like" set by letting

$$D = \{x \in 2^{\mathbb{N}} \colon x(2n) = 0 \text{ for all } n\}.$$

Then $\dim_H(D) = \frac{1}{2}$.

Generalized Cantor sets. Later on, a more general form of Cantor set will come into play. Let $m \geq 2$ be an integer, and suppose $0 < r < 1/m$. Put $I_0^0 = [0,1]$. Given intervals $I_n^0, \ldots, I_n^{m^n - 1}$, define the intervals $I_{n+1}^0, \ldots, I_{n+1}^{m^{n+1} - 1}$ by replacing each I_n^j by m equally spaced intervals of length $r|I_n^j|$. The Cantor set $C_{m,r}$ is then defined as

$$C_{m,r} = \bigcap_n \bigcup_i I_n^i.$$

It can be shown that

$$\dim_H C_{m,r} = \frac{\ln(m)}{-\ln(r)},$$

see [15, Example 4.5]. In fact, it holds that $0 < \mathcal{H}^{\ln(m)/-\ln(r)}(C_{m,r}) < \infty$. An upper bound on $\mathcal{H}^{\ln(m)/-\ln(r)}$ can be obtained by using the intervals $I_n^0, \ldots, I_n^{m^n - 1}$ of the same level as a covering. To show $\mathcal{H}^{\ln(m)/-\ln(r)}(C_{m,r}) > 0$, one can show a mass can be distributed along $C_{m,r}$ that disperses sufficiently fast. A *mass distribution* on a subset E of a metric space Y is a measure on Y with support[1] contained in Y such that $0 < \mu E < \infty$.

Lemma 3.2 (Mass Distribution Principle)**.** *Suppose Y is a metric space and μ is a mass distribution on a set $E \subseteq Y$. Suppose further that for some s there are numbers $c > 0$ and $\varepsilon > 0$ such that*

$$\mu U \leq c \cdot d(U)^s$$

for all sets U with $d(U) < \varepsilon$. Then $\mathcal{H}^s \geq \mu(E)/c$.

For a proof, refer to [15]. The mass distribution principle has a converse, *Frostman's Lemma*, which we will address in Section 4.

4. Hausdorff dimension and information

A 1949 paper by Eggleston [14] established a connection between Hausdorff dimension and entropy. Let $\vec{p} = (p_1, \ldots, p_N)$ be a probability vector, i.e. $p_i \geq 0$ and $\sum p_i = 1$. Given $(p_1, \ldots, p_N)$, $i < N$, and $x \in [0,1]$, let

$$\#(x, i, n) = \# \text{ occurrences of digit } i \text{ among the first } n \text{ digits}$$

$$\text{of the } N\text{-ary expansion of } x$$

and

$$D(\vec{p}) = \left\{ x \in [0,1] : \lim_{n \to \infty} \frac{\#(x, i, n)}{n} = p_i \text{ for all } i \leq N \right\}.$$

---

[1] Recall that the support of a measure on a metric space Y is the largest closed subset F of Y such that for every $x \in F$, if $U \ni x$ is an open neighborhood of x, $\mu U > 0$

Theorem 4.1 (Eggleston)**.** *For any* $\vec{p}$,

$$\dim_H D(\vec{p}) = -\sum_i p_i \log_N p_i = \log_2 N \cdot H(\vec{p}).$$

Using an effective version of Hausdorff dimension, we can see that Eggleston's result is closely related to a fundamental theorem in information theory.

4.1. **Hausdorff dimension and Kolmogorov complexity.** In the 1980s and 1990s, a series of papers by Ryabko [41, 42] and Staiger [47, 48, 49, 50] exhibited a strong connection between Hausdorff dimension, Kolmogorov complexity, and entropy. A main result was that if a set has high Hausdorff dimension, it has to contain elements of high incompressibility ratio. In this section, we focus on the space $2^{\mathbb{N}}$ of infinite binary sequences. It should be clear how the results could be adapted to sequence spaces over arbitrary finite alphabets. We will later comment how these ideas can be adapted to $\mathbb{R}$.

Definition 4.2. Given $x \in 2^{\mathbb{N}}$, we define the *lower incompressibility ratio* of x as

$$\underline{\kappa}(x) = \liminf_{n \to \infty} \frac{K(x \restriction n)}{n}$$

respectively. The lower incompressibility ratio of a subset $A \subseteq 2^{\mathbb{N}}$ is defined as

$$\underline{\kappa}(A) = \sup\{\underline{\kappa}(x) \colon x \in A\}$$

Theorem 4.3 (Ryabko)**.** *For any set* $A \subseteq 2^{\mathbb{N}}$,

$$\dim_H A \leq \underline{\kappa}(A).$$

Staiger was able to identify settings in which this upper bound is also a lower bound for the Hausdorff dimension, in other words, the dimension of a set equals the least upper bound on the incompressibility ratios of its individual members. In particular, he showed it holds for any Σ_2^0 (lightface) subset of $2^{\mathbb{N}}$.

Theorem 4.4 (Staiger)**.** *If* $A \subseteq 2^{\mathbb{N}}$ *is* Σ_2^0, *then*

$$\dim_H A = \underline{\kappa}(A).$$

Using the countable stability of Hausdorff dimension, this result can be extended to arbitrary countable unions of Σ_2^0 sets.

The ideas underlying Ryabko's and Staiger's results were fully exposed when Lutz [29, 30] extended Martin-Löf's concept of randomness tests for individual sequences [33] to Hausdorff measures. Lutz used a variant of martingales, so called s-gales[2]. In this paper, we will follow the approach using variants of Martin-Löf tests. We will define a rather general notion of Martin-Löf test which can be used both for probability measures and Hausdorff measures.

4.2. Effective nullsets.

Definition 4.5. A *premeasure* on $2^{\mathbb{N}}$ is a function $\rho : 2^{<\mathbb{N}} \to [0, \infty)$.

Definition 4.6. Let ρ be a computable premeasure on $2^{\mathbb{N}}$. A ρ-test is a sequence $(W_n)_{n \in \mathbb{N}}$ of sets $W_n \subseteq 2^{<\mathbb{N}}$ such that for all n,

$$\sum_{\sigma \in W_n} \rho(\sigma) \leq 2^{-n}.$$

A set $A \subseteq 2^{\mathbb{N}}$ is *covered by a ρ-test* (W_n) if

$$A \subseteq \bigcap_n \bigcup_{\sigma \in W_n} [\![\sigma]\!],$$

in other words, A is contained in the G_δ set induced by (W_n).

Of course, whether this is a meaningful notion depends on the nature of ρ. We consider two families of premeasures, *Hausdorff premeasures* of the form $\rho(\sigma) = 2^{-\operatorname{len}(\sigma)s}$, and *probability premeasures*, which satisfy

(4.1) $$\rho(\varnothing) = 1,$$

(4.2) $$\rho(\sigma) = \rho(\sigma^\frown 0) + \rho(\sigma^\frown 1).$$

All premeasures considered henceforth are assumed to be either Hausdorff premeasures or probability premeasures. The latter we will often simply refer to as *measures*.

Hausdorff premeasures are completely determined by the value of s, and hence in this case we will speak of s-tests. By the Carathéodory Extension Theorem, probability premeasures extend to a unique Borel probability measure on $2^{\mathbb{N}}$. We will therefore also identify Borel probability measures μ on $2^{\mathbb{N}}$ with the underlying premeasure $\rho_\mu(\sigma) = \mu[\![\sigma]\!]$.

Even though the definition of Hausdorff measures is more involved than that of probability measures, if we only are interested in Hausdorff nullsets

---

[2]A variant of this concept appeared, without explicit reference to Hausdorff dimension, already in Schnorr's fundamental book on algorithmic randomness [43]

(which is the case if we only worry about Hausdorff dimension), it is sufficient to consider s-tests.

Proposition 4.7. $\mathcal{H}^s(A) = 0$ *if and only if A is covered by an s-test.*

The proof is straightforward, observing that any set $\{W_n\}$ for which $\sum_{\sigma \in W_n} 2^{-\mathrm{len}(\sigma)s}$ is small admits only cylinders with small diameter.

Tests can be made *effective* by requiring the sequence $\{W_n\}$ be uniformly recursively enumerable in a suitable representation [12, 39]. To avoid having to deal with issues arising from representations of non-computable measures, we restrict ourselves here to computable premeasures.

We call the resulting test a *Martin-Löf ρ-test*.

If A is covered by a Martin-Löf ρ-test, we say that A is *effectively ρ-null*. There are at most countably many Martin-Löf ρ-tests. Since for the premeasures we consider countable unions of null sets are null, for every premeasure ρ there exists a non-empty set of points not covered by any ρ-test. In analogy to the standard definition Martin-Löf random sequences for Lebesgue measure λ, it makes sense to call such sequences *ρ-random*.

Definition 4.8. Given a premeasure ρ, a sequence $x \in 2^{\mathbb{N}}$ is *ρ-random* if $\{x\}$ is not effectively ρ-null.

If ρ is a probability premeasure, we obtain the usual notion of Martin-Löf randomness for probability measures. For Hausdorff measures, similar to (3.2) and (3.3), we have the following (since every s-test is a t-test for all rational $t \geq s$).

Proposition 4.9. *Suppose s is a rational number. If x is not s-random, then it is not t-random for all rational $t > s$. And if x is s-random, then it is t-random for all rational $0 < t < s$.*

We can therefore define the following *pointwise* version of Hausdorff dimension.

Definition 4.10 (Lutz)**.** The *effective* or *constructive* Hausdorff dimension of a sequence $x \in 2^{\mathbb{N}}$ is defined as

$$\dim_H x = \inf\{s > 0 \colon x \text{ not } s\text{-random }\}.$$

We use the notation $\dim_H x$ to stress the pointwise analogy to (classical) Hausdorff dimension.

4.3. **Effective dimension and Kolmogorov complexity.** Schnorr [see 13] made the fundamental observation that Martin-Löf randomness for $\lambda(\sigma) = 2^{-\operatorname{len}(\sigma)}$ and incompressibility with respect to prefix-free complexity coincide. Gács [18] extended this to other computable measures.

Theorem 4.11 (Gács [18]). *Let μ be a computable measure on $2^{\mathbb{N}}$. Then x is μ-random if and only of there exists a constant c such that*

$$\forall n \ K(x{\restriction}_n) \geq -\log \mu[\![x{\restriction}_n]\!].$$

One can prove a similar result for s-randomness.

Theorem 4.12. *A sequence $x \in 2^{\mathbb{N}}$ is s-random if and only if there exists a c such that for all n,*

$$K(x{\restriction}_n) \geq sn - c.$$

In other words, being s-random corresponds to being incompressible only to a factor of s. For this reason s-randomness has also been studied as a notion of *partial randomness* (for example, [5])

Proof. ($\Leftarrow$) Assume x is not s-random. Thus there exists an s-test (W_n) covering x. Define functions $m_n : 2^{<\mathbb{N}} \to \mathbb{R}$ by

$$m_n(\sigma) = \begin{cases} n2^{-\operatorname{len}(\sigma)s-1} & \text{if } \sigma \in W_n, \\ 0 & \text{otherwise,} \end{cases}$$

and let

$$m(\sigma) = \sum_{n=1}^{\infty} m_n(\sigma).$$

All m_n and thus m are enumerable from below. Furthermore,

$$\sum_{\sigma \in 2^{<\mathbb{N}}} m(\sigma) = \sum_{\sigma \in 2^{<\mathbb{N}}} \sum_{n=1}^{\infty} m_n(\sigma) = \sum_{n=1}^{\infty} n \sum_{\sigma \in W_n} 2^{-\operatorname{len}(w)s-1} \leq \sum_{n=1}^{\infty} \frac{n}{2^{n+1}} \leq 1.$$

Hence m is a semimeasure and enumerable from below. By the Coding Theorem (Theorem 2.9), there exists a constant d such that for all σ,

$$K(\sigma) \leq -\log m(\sigma) + d.$$

Now let $c > 0$ be any constant. Set $k = \lceil 2^{c+d+1} \rceil$. As x is covered by (W_n) there is some $\sigma \in C_k$ and some $l \in \mathbb{N}$ with $\sigma = x{\restriction}_l$. This implies

$$K(x{\restriction}_l) = K(\sigma) \leq -\log m(\sigma) + d = -\log(k2^{\operatorname{len}(\sigma)s}) + d \leq sl - c.$$

As c was arbitrary, we see that $K(x{\restriction}n)$ does not have a lower bound of the form $sn - \text{constant}$.

($\Rightarrow$) Assume for every k there exists an l such that $K(x{\restriction}l) < sl - k$. Let

$$W_k := \{\sigma \colon K(\sigma) \le s\,\text{len}(\sigma) - k\}.$$

Then (W_k) is uniformly enumerable, and each W_k covers x. Furthermore,

$$\sum_{\sigma \in W_k} 2^{-\text{len}(\sigma)s} \le \sum_{\sigma \in W_k} 2^{-K(\sigma)-k} \le 2^{-k}.$$

Therefore, (W_k) is an s-test witnessing that x is not s-random. $\qquad\square$

As a corollary, we obtain the "fundamental theorem" of effective dimension [34, 51].

Corollary 4.13. *For any $x \in 2^{\mathbb{N}}$,*

$$\dim_H x = \underline{\kappa}(x).$$

Proof. Suppose first $\dim_H x < s$. This means x is not s-random and by Theorem 4.12, for infinitely many n, $K(x{\restriction}n) \le sn$. Thus $\underline{\kappa}(x) \le s$. On the other hand, if $\underline{\kappa}(x) < s$, there exists $\varepsilon > 0$ such that $K(x{\restriction}n) < (s - \varepsilon)n$ for infinitely many n. In particular, $sn - K(x{\restriction}n)$ is not bounded from below by a constant, which means x is not s-random. Therefore, $\dim_H x \le s$. $\qquad\square$

As a further corollary, we can describe the correspondence between Hausdorff dimension and Kolmogorov complexity (Theorem 4.4) as follows.

Corollary 4.14. *For every Σ^0_2 $A \subseteq 2^{\mathbb{N}}$,*

$$\dim_H A = \sup_{x \in A} \dim_H x.$$

The general idea of these results is that if a set of a rather simple nature in terms of definability, then set-wise dimension can be described as the supremum of pointwise dimensions. For more complicated sets, one can use relativization to obtain a suitable correspondence. One can relativize the definition of effective dimension with respect to an oracle $z \in 2^{\mathbb{N}}$. Let us denote the corresponding notion by $\dim^z_H$. If we relativize the notion of Kolmogorov complexity as well, we can prove a relativized version of Corollary 4.13.

Lutz and Lutz [31] were able to prove a most general set-point correspondence principle that is based completely on relativization, in which the Borel complexity does not feature anymore.

Theorem 4.15 (Lutz and Lutz). *For any set $A \subseteq 2^{\mathbb{N}}$,*

$$\dim_H A = \min_{z \in 2^{\mathbb{N}}} \sup_{x \in A} \dim_H^z x.$$

In other words, the Hausdorff dimension of a set is the minimum among its pointwise dimensions, taken over all relativized worlds.

These point-vs-set principles are quite remarkable in that they have no direct classical counterpart. While Hausdorff dimension is *countably stable* (Proposition 3.1), stability does not extend to arbitrary unions, since singleton sets always have dimension 0.

4.4. **Effective dimension vs randomness.** Let μ be a computable probability measure on $2^{\mathbb{N}}$. By Theorem 4.11, for a μ-random $x \in 2^{\mathbb{N}}$, the prefix-free complexity of $x{\restriction}_n$ is bounded from below by $\log \mu[\![x{\restriction}_n]\!]$. If we can bound $\log \mu[\![x{\restriction}_n]\!]$ asymptotically from below by a bound that is linear in n, we also obtain a lower bound on the effective dimension of x. This can be seen as an effective/pointwise version of the mass distribution principle: If a point "supports" measure that decays sufficiently fast along it, we can bound its effective dimension from below.

Consider, for example, the *Bernoulli measure* μ_p on $2^{\mathbb{N}}$. Given $p \in [0,1]$, μ_p is induced by the premeasure

$$\rho_p(\sigma{}^\frown 1) = \rho_p(\sigma) \cdot p \qquad \rho_p(\sigma{}^\frown 0) = \rho_p(\sigma) \cdot (1-p).$$

As we will see later, if x is μ-random,

$$(4.3) \qquad \lim_{n \to \infty} \frac{-\log \mu_p[\![x{\restriction}_n]\!]}{n} = H(p),$$

and therefore, $\dim_H x \geq H(p)$ for any μ-random x. It follows that the "decay" of μ along x is bounded by $2^{-H(p)n}$.

Does the reverse direction hold, too? If x has positive effective dimension, is x random for a suitably fast dispersing measure?

Theorem 4.16 (Reimann [38]). *Suppose $x \in 2^{\mathbb{N}}$ is such that $\dim_H x > s$. Then there exists a probability measure μ on $2^{\mathbb{N}}$ and a constant c such that for all $\sigma \in 2^{<\mathbb{N}}$,*

$$\mu[\![\sigma]\!] \leq c2^{-sn}.$$

This is an effective/pointwise version of *Frostman's Lemma* [16], which in Cantor space essentially says that if a Borel set has Hausdorff dimension $> s$, it supports a measures that disperses at least as fast as 2^{-sn} (up to a multiplicative constant).

4.5. **The Shannon-Macmillan-Breiman Theorem.** In (4.3), we already stated a special case of the *Shannon-McMillan-Breiman Theorem*, also known as the *Asymptotic Equipartition Property* or *Entropy Rate Theorem*. It is one of the central results of information theory.

Let $T : 2^{\mathbb{N}} \to 2^{\mathbb{N}}$ be the *shift map*, defined as

$$T(x)_i = x_{i+1},$$

where y_i is the i-th bit of $y \in 2^{\mathbb{N}}$. A measure μ on $2^{\mathbb{N}}$ is *shift-invariant* if for any Borel set A, $\mu T^{-1}(A) = \mu A$. A shift-invariant measure μ is *ergodic* if $T^{-1}(A) = A$ implies $\mu A = 0$ or $\mu A = 1$. Bernoulli measures μ_p are an example of shift-invariant, ergodic measures. For background on ergodic measures on sequence spaces, see [45].

Theorem 4.17. *Let μ be a shift-invariant ergodic measure on $2^{\mathbb{N}}$. Then there exists a non-negative number h such that almost surely,*

$$\lim_{n \to \infty} -\frac{\log \mu [\![x \lceil n]\!]}{n} = h$$

For a proof, see for example [45]. The number h is called the *entropy rate* of μ and also written as $h(\mu)$. It is also possible to define the entropy of the underlying $\{0, 1\}$-valued process. First, let

$$H(\mu^{(n)}) = -\sum_{\sigma \in A^n} \mu [\![\sigma]\!] \log \mu [\![\sigma]\!].$$

One can show that this is *subadditive* in the sense that

$$H(\mu^{(n+m)}) \le H(\mu^{(n)}) + H(\mu^{(m)}),$$

which implies that

$$H(\mu) = \lim \frac{1}{n} H(\mu^{(n)})$$

exists. $H(\mu)$ is called the *process entropy* of μ. It is clear that for i.i.d. $\{0, 1\}$-valued processes, $H(\mu)$ agrees with the entropy of the distribution on $\{0, 1\}$ as defined in Section 2.

Entropy rate is a local measure, as it follows the behavior of a measure along a typical point, while process entropy captures the entropy over finer and finer partitions of the whole space $2^{\mathbb{N}}$. Process entropy in turn is a special case of a general definition of entropy in measure-theoretic dynamical systems, known as *Kolmogorov-Sinai entropy* (see for example [56]).

The Shannon-Macmillan-Breiman Theorem states that the pointwise entropy rate

$$\lim_{n \to \infty} -\frac{\log \mu [\![x \lceil n]\!]}{n}$$

not only exists almost surely, but also is constant up to a set of measure zero. Furthermore, it can be shown that this constant entropy rate coincides with the process entropy $H(\mu)$. This fundamental principle has been established in much more general settings, too (e.g. for amenable groups, see [35],[27]).

4.6. **The effective SMB-theorem and dimension.** We have already seen a connection between randomness and complexity in Theorems 4.11 and 4.12.

If μ is a computable ergodic measure, this immediately connects the complexity of infinite sequences to entropy via the Shannon-Macmillan-Breiman Theorem. Since μ-almost every sequence is μ-random, we obtain that almost surely,

$$\dim_H x = \liminf_{n \to \infty} \frac{K(x \lceil_n)}{n} \geq \frac{-\log \mu[\![x \lceil_n]\!]}{n} = H(\mu).$$

Zvonkin and Levin [58] and independently Brudno [3] showed that for μ-almost every sequence,

$$\lim_{n \to \infty} \frac{K(x \lceil_n)}{n} \text{ exists,}$$

which implies in turn the asymptotic compression ratio of almost every real equals the metric entropy of μ. Analyzing the proof of the SMB-Theorem by Ornstein and Weiss [35], V'yugin [55] was able to establish, moreover, that for *all* μ-random sequences x,

$$\limsup_{n \to \infty} \frac{K(x \lceil_n)}{n} = H(\mu).$$

Finally, Hoyrup [20] established that the lim inf also equals $H(\mu)$, thereby showing that the SMB-Theorem is effective in the sense of Martin-Löf randomness.

Theorem 4.18 (Effective Shannon-Macmillan-Breiman Theorem [55, 20]). *Let μ be a computable, ergodic, shift-invariant measure. Then, for each $x \in 2^{\mathbb{N}}$ that is random with respect to μ,*

$$\dim_H x = \lim_{n \to \infty} -\frac{\log \mu[\![x \lceil_n]\!]}{n} = H(\mu).$$

Note that effective dimension as a pointwise version of Hausdorff dimension is a natural counterpart to the pointwise nature of entropy rate, the asymptotic compression ratio (in terms of Kolmogorov complexity) forming the bridge between the two, or, to state it somewhat simplified:

$$dimension = complexity = entropy$$

4.7. **Subshifts.** The principle above turns out to be quite persistent, even when one passes from effective dimension of sequences (points) to (classical) dimension of sets. A *subshift* of $2^{\mathbb{N}}$ is a closed subset $X \subseteq 2^{\mathbb{N}}$ invariant under the shift map on $2^{\mathbb{N}}$.

Subshifts are a *topological dynamical system*, but they can also carry an invariant measure. There is a topological variant of entropy, which for subshifts is given as

$$h_{\text{top}}(X) = \lim_{n \to \infty} \frac{\log |\{\sigma \colon |\sigma| = n \,\&\, \sigma \subset X\}|}{n},$$

that is, h_{top} measures the relative numbers of strings present in X as the length goes to infinity.

Furstenberg [17] showed that for subshifts $X \subseteq 2^{\mathbb{N}}$,

$$h_{\text{top}}(X) = \dim_H X.$$

Simpson [46] extended this result to multidimensional one- and two-sided subshifts. Furthermore, he established that for *any* such subshifts, the coincidence between entropy, complexity, and dimension is complete in the sense that

$$h_{\text{top}}(X) = \dim_H X = \underline{\kappa}(X).$$

Most recently, Day [11] further extended this identity to computable subshifts of A^G, where G is a computable amenable group with computable bi-invariant tempered Følner sequence.

Staiger [48, 49] showed that Furstenberg's result also holds for other families of sets: closed sets definable by finite automata, and for ω-powers of languages definable by finite automata.

4.8. **Application: Eggleston's theorem.** Using the dimension-complexity-entropy correspondence, we can also give a short proof of Eggleston's Theorem (Theorem 4.1). We focus on the binary version. The proof for larger alphabets is similar.

Let $p \in [0, 1]$ be computable, and consider the Cantor space equivalent of D_p,

$$\overline{D}_p = \{x \in 2^{\mathbb{N}} \colon \lim \frac{\#(x, 0, n)}{n} = p\},$$

where $\#(x, i, n)$ now denotes the number of occurrences of digit i among the first n digits of x.

All μ_p-random sequences satisfy the Law of Large Numbers. Therefore, there exists a Π_1^0 class $P \subset \overline{D}_p$ consisting of only μ_p-random sequences. By

Theorem 4.4 and Theorem 4.18

$$\dim_H D_p \geq \dim_H P = \underline{\kappa}(P) = H(\mu_p) = H(p).$$

On the other hand, for every $x \in D_p$,

$$\underline{\kappa}(x) \leq H(p),$$

as we can always construct a two-part code for $x \in D_p$, by giving the number of 1's in $x{\upharpoonright}_n$, and then the position of $x{\upharpoonright}_n$ in a lexicographic enumeration of all binary strings of length n with that number of 1's. As $\log_2 \binom{n}{k} \approx nH(k/n)$, this gives the desired upper bound for $\underline{\kappa}(x)$. Therefore, by Theorem 4.3,

$$\dim_H D_p \leq H(p).$$

For non-computable p, one can use relativized versions of the results used above.

Cantor space vs the real line. The reader will note that Eggleston's Theorem, as originally stated, applies to the unit interval $[0, 1]$, while the version proved above is for Cantor space $2^{\mathbb{N}}$. It is in fact possible to develop the theory of effective dimension for real numbers, too. One way to do this is to use net measures $\mathcal{M}^s$ as introduced in Section 3. Dyadic intervals correspond directly to cylinder sets in $2^{\mathbb{N}}$ via the finite-to-one mapping

$$x \in 2^{\mathbb{N}} \mapsto \sum_n x_n 2^{-(n+1)}.$$

If we identify a real with its dyadic expansion (which is unique except for dyadic rationals), we can speak of Kolmogorov complexity of (initial segments of) reals etc. Furthermore, the connection between effective dimension and Kolmogorov complexity carries over to the setting in $[0, 1]$. For a thorough development of effective dimension in Euclidean (and other) spaces, see [28].

5. Multifractal measures

A measure can have a fractal nature by the way it spreads its mass. To prove a lower bound on the Hausdorff dimension of a Cantor set, we distribute a mass uniformly along it (and appeal to the mass distribution principle, Lemma 3.2). In other words, we define a measure supported on a fractal. But a measure can exhibit fractal properties also through *the way it spreads its mass over its support*, which is not necessarily uniform. Think of earthquake

distributions along fault systems. The fault systems themselves are usually of a fractal nature. Moreover, there often seems to be a non-uniformity in the distribution of earthquakes along fault lines. Some "hot spots" produce much more earthquakes than other, more quiet sections of a fault (see for example Kagan and Knopoff [22], Kagan [21], Mandelbrot [32]). The underlying mathematical concept has become known as *multifractality of measures*. It is widely used in the sciences today. It would go far beyond the scope of this survey to introduce the various facets of multifractal measures here. Instead, we focus on one aspect of multifractality that appears to be a natural extension of the material presented in Section 4. The interested reader may refer to a forthcoming paper [36].

5.1. **The dimension of a measure.** Unless explicitly noted, in the following *measure* will always mean Borel probability measure on $[0, 1]$.

First, we define the *dimension of a measure*, which reflects the fractal nature of its support.

Definition 5.1 (Young [57])**.** Given a Borel probability measure μ on $\mathbb{R}$, let

$$\dim_H \mu = \inf\{\dim_H E \colon \mu E = 1\}$$

For example, if we distribute a unit mass uniformly along a Cantor set $C_{m,r}$, we obtain a measure $\mu_{m,r}$ with

$$\dim_H \mu_{m,r} = -\ln(m)/\ln(r).$$

Instead by a single number, one can try to capture the way a measure spreads its mass among fractal sets by means of a distribution.

Definition 5.2 (Cutler [10])**.** The *dimension distribution* of a measure μ is given by

$$\mu_{\dim}[0, t] = \sup\{\mu D \colon \dim_H D \le t, \ D \subseteq \mathbb{R} \text{ Borel}\}.$$

As $\dim_H$ is countably stable (see Section 4.3), $\mu_{\dim}$ extends to a unique Borel probability measure on $[0, 1]$. The dimension of a measure can be expressed in terms of $\mu_{\dim}$:

$$\dim_H \mu = \inf\{\alpha \colon \mu_{\dim}[0, \alpha] = 1\}.$$

For many measures, the dimension distribution does not carry any extra information. For example, if λ is Lebesgue measure on $[0, 1]$, the dimension

distribution is the *Dirac point measure* δ_1, where δ_α is defined as

$$\delta_\alpha A = \begin{cases} 1 & \alpha \in A, \\ 0 & \alpha \notin A. \end{cases}$$

Measures whose dimension distribution is a point measure are called *exact dimensional*. This property also to uniform distributions on Cantor sets.

Proposition 5.3. *Let $\mu_{m,r}$ be the uniform distribution along the Cantor set $C_{m,r}$. Then $\mu_{\dim} = \delta_\alpha$, where $\alpha = -\ln(m)/\ln(r)$.*

It is not completely obvious that $\mu_{m,r}$ has no mass on any set of Hausdorff dimension $< -\ln(m)/\ln(r)$. One way to see this is by connecting the dimension distribution of a measure to its *pointwise dimensions*, which we will introduce next.

5.2. **Pointwise dimensions.**

Definition 5.4. Let μ be a probability measure on a separable metric space. The *(lower) pointwise dimension* of μ at a point x is defined as

$$\delta_\mu(x) = \liminf_{\varepsilon \to 0} \frac{\log \mu B(x, \varepsilon)}{\log \varepsilon}.$$

(Here $B(x, \varepsilon)$ is the ε-ball around x.)

Of course, one can also define the upper pointwise dimension by considering $\limsup$ in place of $\liminf$, which is connected to packing measures and packing dimension similar to the way lower pointwise dimension is connected to Hausdorff measures and dimension. As this survey is focused on Hausdorff dimension, we will also focus on the lower pointwise dimensions and simply speak of "pointwise dimension" when we mean lower pointwise dimension.

We have already encountered the pointwise dimension of a measure when looking at the Shannon-Macmillan-Breiman Theorem (Theorem 4.17), which says that for ergodic measures on $2^{\mathbb{N}}$ μ-almost surely the pointwise dimension is equal to the entropy of μ.

We have also seen (Corollary 4.13) that the effective dimension of a real x is its pointwise dimension with respect to the semimeasure $\widetilde{Q}(\sigma) = 2^{-K(\sigma)}$.

We can make this analogy even more striking by considering a different kind of universal semimeasure that gives rise to the same notion of effective dimension.

Recall that a semimeasure is a function $Q : \mathbb{N} \to \mathbb{R}^{\geq 0}$ such that $\sum Q(m) \leq 1$. Even when seen as a function on $2^{<\mathbb{N}}$, such a semimeasure is of a discrete

nature. In particular, it does not take into account the partial ordering of strings with respect to the prefix relation. The notion of a *continuous semimeasure* does exactly that. As the prefix relation corresponds to a partial ordering of basic open sets, and premeasures are defined precisely on those sets, continuous semimeasures respects the structure of $2^{\mathbb{N}}$. Compare the following definition with the properties of a probability premeasure (4.1), (4.2).

Definition 5.5. (i) A *continuous semimeasure* is a function $M : 2^{<\mathbb{N}} \to \mathbb{R}^{\geq 0}$ such that

$$M(\varnothing) \leq 1,$$
$$M(\sigma) \geq M(\sigma^\frown 0) + M(\sigma^\frown 1).$$

(ii) A continuous semimeasure M is *enumerable* if there exists a computable function $f : \mathbb{N} \times 2^{<\mathbb{N}} \to \mathbb{R}^{\geq 0}$ such that for all n, σ,

$$f(n, \sigma) \leq f(n + 1, \sigma) \quad \text{and} \quad \lim_{n \to \infty} f(n, \sigma) = M(\sigma).$$

(iii) An enumerable continuous semimeasure M is *universal* if for every enumerable continuous semimeasure P there exists a constant c such that for all σ,

$$P(\sigma) \leq c\, M(\sigma).$$

Levin [58] showed that a universal semimeasure exists. Let us fix such a semimeasure $\mathbf{M}$. Similar to K, we can introduce a complexity notion based on $\mathbf{M}$ by letting

$$\mathrm{KM}(\sigma) = -\log \mathbf{M}(\sigma).$$

Some authors denote KM by KA and refer to it as *a priori complexity*. KM is closely related to K.

Theorem 5.6 (Gács [19],Uspensky and Shen [54])**.** *There exist constants* c, d *such that for all* σ,

$$\mathrm{KM}(\sigma) - c \leq \mathrm{K}(\sigma) \leq \mathrm{KM}(\sigma) + \mathrm{K}(\mathrm{len}(\sigma)) + d.$$

When writing $K(n)$, we identify n with is binary representation (which is a string of length approximately $\log(n)$). Since $\mathrm{K}(n)/n \to 0$ for $n \to \infty$, we obtain the following alternative characterization of effective dimension.

Corollary 5.7. *For any* $x \in 2^{\mathbb{N}}$,

$$\dim_H x = \liminf_{n \to \infty} \frac{\mathrm{KM}(x{\upharpoonright}n)}{n} = \liminf_{n \to \infty} \frac{-\log \mathbf{M}(x{\upharpoonright}n)}{n}.$$

In other words, the effective dimension of a point x in Cantor space (and hence also in $[0,1)$) is its pointwise dimension with respect to a universal enumerable continuous semimeasure $\mathbf{M}$.

KM "smoothes" out a lot of the complexity oscillations K has. This has a nice effect concerning the characterization of random sequences. On the one hand, $\mathbf{M}$ is universal among enumerable semimeasures, hence in particular among computable probability measures. Therefore, for any $x \in 2^{\mathbb{N}}$, and any computable probability measure μ,

$$-\log \mathbf{M}(x{\restriction}_n) - c \leq -\log \mu[\![x{\restriction}_n]\!] \quad \text{ for all } n,$$

where c is a constant depending only on μ.

On the other hand, Levin [24] showed that $x \in 2^{\mathbb{N}}$ is random with respect to computable probability measure μ if and only if for some constant d,

$$\mathrm{KM}(x{\restriction}_n) \geq -\log \mu[\![x{\restriction}_n]\!] - d \quad \text{ for all } n.$$

Thus, along μ-random sequences, $-\log \mathbf{M}$ and $-\log \mu$ differ by at most a constant. This gives us the following.

Proposition 5.8. *Let μ be a computable probability measure on $2^{\mathbb{N}}$. If $x \in 2^{\mathbb{N}}$ is μ-random, then*

$$\dim_H x = \delta_\mu(x).$$

Given a measure μ and $\alpha \geq 0$, let

$$D_\mu^\alpha = \{x \colon \delta_\mu(x) \leq \alpha\}.$$

If μ is computable, then since $\mathbf{M}$ is a universal semimeasure,

$$D_\mu^\alpha \subseteq \{x \colon \dim_H x \leq \alpha\}.$$

Let us denote the set on the right hand side by $\dim_{\leq \alpha}$.

Theorem 5.9 (Cai and Hartmanis [4], Ryabko [41])**.** *For any $\alpha \geq 0$,*

$$\dim_H(\dim_{\leq \alpha}) = \alpha.$$

It follows that $\dim_H D_\mu^\alpha \leq \alpha$. This was first shown, for general μ, by Cutler [8, 9]. Cutler also characterized the dimension distribution of a measure through its pointwise dimensions.

Theorem 5.10 (Cutler [8, 9])**.** *For each $t \geq 0$,*

$$\mu_{\dim}[0, t] = \mu(D_\mu^t).$$

By the above observation, for computable measures, we can replace D_μ^α by $\dim_{\leq t}$. The theorem explains why $\dim_H \mu_{m,r} = \delta_{-\ln(m)/\ln(r)}$: Since the mass is distributed uniformly over the set $C_{m,r}$, $\delta_{\mu_{m,r}}(x)$ is almost surely constant.

5.3. **Multifractal spectra.** We have seen in the previous section that uniform distributions on Cantor sets result in dimension distributions of the form of Dirac point measures δ_α. We can define a more "fractal" measure by biasing the distribution process. Given a probability vector $\vec{p} = (p_1, \ldots, p_m)$, the measure $\mu_{m,r,\vec{p}}$ is obtained by splitting the mass in the interval construction of $C_{m,r}$ not uniformly, but according to $\vec{p}$. The resulting measure is still exact dimensional (see [10] – this is similar to the SMB Theorem). However, the measures $\mu_{m,r,\vec{p}}$ exhibit a rich fractal structure when one considers the *dimension spectrum* of the pointwise dimensions (instead of its dimension distribution).

Definition 5.11. The *(fine) multifractal spectrum* of a measure μ is defined as

$$f_\mu(\alpha) = \dim_H\{x \colon \delta_\mu(x) = \alpha\}.$$

For the measures $\mu_{m,r,\vec{p}}$, it is possible to compute the multifractal spectrum using the *Legendre transform* (see [15]). It is a function continuous in α with $f_\mu(\alpha) \leq \alpha$ and maximum $\dim_H \mu$. In a certain sense, the universal continuous semimeasure $\mathbf{M}$ is a "perfect multifractal", as, by the Cai-Hartmanis result, $f_{\mathbf{M}}(\alpha) = \alpha$ for all $\alpha \in [0, 1]$.

Moreover, the multifractal spectrum of a computable $\mu = \mu_{m,r,\vec{p}}$ can be expressed as a "deficiency" of multifractality against $\mathbf{M}$.

Theorem 5.12 (Reimann [36]). *For a computable measure* $\mu = \mu_{m,r,\vec{p}}$, *it holds that*

$$f_\mu(\alpha) = \dim_H\left\{x \colon \frac{\dim_H x}{\dim_\mu x} = \alpha\right\}.$$

Here $\dim_\mu$ denotes the *effective Billingsley dimension* of x with respect to μ (see [37]).

References

[1] V. Becher, J. Reimann, and T. A. Slaman. Irrationality exponent, Hausdorff dimension and effectivization. *Monatshefte für Mathematik*, 185 (2):167–188, 2018.

[2] P. Billingsley. *Ergodic theory and information*. John Wiley & Sons Inc., New York, 1965.

[3] A. Brudno. Entropy and the complexity of the trajectories of a dynamic system. *Trudy Moskovskogo Matematicheskogo Obshchestva*, 44:124–149, 1982.

[4] J.-Y. Cai and J. Hartmanis. On Hausdorff and topological dimensions of the Kolmogorov complexity of the real line. *J. Comput. System Sci.*, 49(3):605–619, 1994.

[5] C. Calude, L. Staiger, and S. A. Terwijn. On partial randomness. *Annals of Pure and Applied Logic*, 138(1–3):20–30, 2006.

[6] G. J. Chaitin. A theory of program size formally identical to information theory. *Journal of the ACM*, 22(3):329–340, 1975.

[7] T. M. Cover, P. Gács, and R. M. Gray. Kolmogorov's contributions to information theory and algorithmic complexity. *The Annals of Probability*, 17(3):840–865, 1989.

[8] C. D. Cutler. A dynamical system with integer information dimension and fractal correlation exponent. *Communications in mathematical physics*, 129(3):621–629, 1990.

[9] C. D. Cutler. Measure disintegrations with respect to σ-stable monotone indices and pointwise representation of packing dimension. In *Proceedings of the 1990 Measure Theory Conference at Oberwolfach. Supplemento Ai Rendiconti del Circolo Mathematico di Palermo, Ser. II*, volume 28, pages 319–340, 1992.

[10] C. D. Cutler. A review of the theory and estimation of fractal dimension. In *Dimension estimation and models*, pages 1–107. World Sci. Publ., River Edge, NJ, 1993.

[11] A. Day. Algorithmic randomness for amenable groups. Draft, September 2017.

[12] A. Day and J. Miller. Randomness for non-computable measures. *Trans. Amer. Math. Soc.*, 365:3575–3591, 2013.

[13] R. G. Downey and D. R. Hirschfeldt. *Algorithmic randomness and complexity*. Springer, 2010.

[14] H. G. Eggleston. The fractional dimension of a set defined by decimal properties. *Quart. J. Math., Oxford Ser.*, 20:31–36, 1949.

[15] K. Falconer. *Fractal geometry*. Wiley, 2nd edition, 2003.

[16] O. Frostman. Potentiel d'équilibre et capacité des ensembles avec quelques applications à la théorie des fonctions. *Meddel. Lunds. Univ.*

Mat. Sem., 3:1–118, 1935.

[17] H. Furstenberg. Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation. *Theory of Computing Systems*, 1(1):1–49, 1967.

[18] P. Gács. Exact Expressions for Some Randomness Tests. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 26(25-27): 385–394, 1980.

[19] P. Gács. On the relation between descriptional complexity and algorithmic probability. *Theoretical Computer Science*, 22(1-2):71–93, 1983.

[20] M. Hoyrup. The dimension of ergodic random sequences. In *STACS'12 (29th Symposium on Theoretical Aspects of Computer Science)*, volume 14, pages 567–576, 2011.

[21] Y. Kagan. Fractal dimension of brittle fracture. *Journal of Nonlinear Science*, 1:1–16, 1991.

[22] Y. Kagan and L. Knopoff. Spatial distribution of earthquakes: The two-point correlation function. *Geophysical Journal International*, 62 (2):303–320, 1980.

[23] A. I. Khinchin. *Mathematical foundations of information theory.* Dover Publications, Inc., New York, N. Y., 1957.

[24] L. A. Levin. On the notion of a random sequence. *Soviet Math. Dokl*, 14(5):1413–1416, 1973.

[25] L. A. Levin. Some theorems on the algorithmic approach to probability theory and information theory: (1971 Dissertation directed by A.N. Kolmogorov). *Annals of Pure and Applied Logic*, 162(3):224–235, Dec. 2010.

[26] M. Li and P. Vitányi. *An introduction to Kolmogorov complexity and its applications.* Springer-Verlag, New York, 3rd edition, 2008.

[27] E. Lindenstrauss. Pointwise theorems for amenable groups. *Inventiones Mathematicae*, 146(2):259–295, 2001.

[28] J. Lutz and E. Mayordomo. Dimensions of Points in Self-Similar Fractals. *SIAM Journal on Computing*, 38(3):1080–1112, 2008.

[29] J. H. Lutz. Dimension in complexity classes. In *Proceedings of the Fifteenth Annual IEEE Conference on Computational Complexity*, pages 158–169. IEEE Computer Society, 2000.

[30] J. H. Lutz. The dimensions of individual strings and sequences. *Inform. and Comput.*, 187(1):49–79, 2003.

[31] J. H. Lutz and N. Lutz. Algorithmic information, plane Kakeya sets, and conditional dimension. In *34th Symposium on Theoretical Aspects of Computer Science, STACS 2017, March 8-11, 2017, Hannover, Germany*, pages 1–13, 2017.

[32] B. B. Mandelbrot. Multifractal measures, especially for the geophysicist. *Pure and Applied Geophysics*, 131:5–42, 1989.

[33] P. Martin-Löf. The definition of random sequences. *Information and Control*, 9(6):602–619, 1966.

[34] E. Mayordomo. A Kolmogorov complexity characterization of constructive Hausdorff dimension. *Inform. Process. Lett.*, 84(1):1–3, 2002.

[35] D. Ornstein and B. Weiss. The Shannon-McMillan-Breiman theorem for a class of amenable groups. *Israel Journal of Mathematics*, 44(1): 53–60, 1983.

[36] J. Reimann. Effective mutlifractal spectra. In preparation.

[37] J. Reimann. *Computability and fractal dimension.* PhD thesis, University of Heidelberg, 2004.

[38] J. Reimann. Effectively closed sets of measures and randomness. *Annals of Pure and Applied Logic*, 156:170–182, 2008.

[39] J. Reimann and T. Slaman. Measures and their random reals. *Transactions of the American Mathematical Society*, 367(7):5081–5097, 2015.

[40] C. A. Rogers. *Hausdorff Measures.* Cambridge University Press, 1970.

[41] B. Y. Ryabko. Coding of combinatorial sources and Hausdorff dimension. *Dokl. Akad. Nauk SSSR*, 277(5):1066–1070, 1984.

[42] B. Y. Ryabko. Noise-free coding of combinatorial sources, Hausdorff dimension and Kolmogorov complexity. *Problemy Peredachi Informatsii*, 22(3):16–26, 1986.

[43] C.-P. Schnorr. *Zufälligkeit und Wahrscheinlichkeit. Eine algorithmische Begründung der Wahrscheinlichkeitstheorie.* Springer-Verlag, Berlin, 1971.

[44] C.-P. Schnorr. Process complexity and effective random tests. *J. Comput. System Sci.*, 7:376–388, 1973.

[45] P. C. Shields. *The ergodic theory of discrete sample paths*, volume 13 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1996.

[46] S. G. Simpson. Symbolic dynamics: entropy= dimension= complexity. *Theory of Computing Systems*, 56(3):527–543, 2015.

[47] L. Staiger. Complexity and entropy. In J. Gruska and M. Chytil, editors, *Mathematical Foundations of Computer Science 1981: Proceedings, 10th Symposium Štrbské Pleso, Czechoslovakia August 31 – September 4, 1981*, pages 508–514, Berlin, Heidelberg, 1981. Springer Berlin Heidelberg.

[48] L. Staiger. Combinatorial properties of the Hausdorff dimension. *Journal of Statistical Planning and Inference*, 23(1):95–100, 1989.

[49] L. Staiger. Kolmogorov complexity and Hausdorff dimension. *Inform. and Comput.*, 103(2):159–194, 1993.

[50] L. Staiger. A tight upper bound on Kolmogorov complexity and uniformly optimal prediction. *Theory of Computing Systems*, 31(3):215–229, 1998.

[51] L. Staiger. Constructive dimension equals Kolmogorov complexity. *Information Processing Letters*, 93(3):149–153, 2005.

[52] L. Staiger. The Kolmogorov complexity of infinite words. *Theor. Comput. Sci.*, 383(2-3):187–199, 2007.

[53] S. A. Terwijn. Complexity and randomness. *Rend. Semin. Mat., Torino*, 62(1):1–37, 2004.

[54] V. A. Uspensky and A. Shen. Relations between varieties of Kolmogorov complexities. *Mathematical Systems Theory*, 29:271–292, 1996.

[55] V. V. V'yugin. Ergodic theorems for individual random sequences. *Theoret. Comput. Sci.*, 207(2):343–361, 1998.

[56] P. Walters. *An introduction to ergodic theory*, volume 79 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1982.

[57] L.-S. Young. Dimension, entropy and Lyapunov exponents. *Ergodic theory and dynamical systems*, 2(1):109–124, 1982.

[58] A. K. Zvonkin and L. A. Levin. The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Mathematical Surveys*, 25(6):83–124, 1970.