Asymptotic Divergences and Strong Dichotomy

Xiang Huang
Le Moyne College, Syracuse, NY 13214, USA
Iowa State University, Ames, IA 50011, USA
huangx@iastate.edu

Jack H. Lutz
Iowa State University, Ames, IA 50011, USA
lutz@iastate.edu

Elvira Mayordomo
Departamento de Informática e Ingeniería de Sistemas,
Instituto de Investigación en Ingeniería de Aragón,
Universidad de Zaragoza, 50018 Zaragoza, Spain
elvira@unizar.es

Donald M. Stull
Iowa State University, Ames, IA 50011, USA
dstull@iastate.edu

Abstract

The *Schnorr-Stimm dichotomy theorem* [31] concerns finite-state gamblers that bet on infinite sequences of symbols taken from a finite alphabet Σ. The theorem asserts that, for any such sequence S, the following two things are true.

(1) If S is not normal in the sense of Borel (meaning that every two strings of equal length appear with equal asymptotic frequency in S), then there is a finite-state gambler that wins money at an infinitely-often exponential rate betting on S.

(2) If S is normal, then any finite-state gambler betting on S loses money at an exponential rate betting on S.

In this paper we use the Kullback-Leibler divergence to formulate the *lower asymptotic divergence* $\mathrm{div}(S||\alpha)$ of a probability measure α on Σ from a sequence S over Σ and the *upper asymptotic divergence* $\mathrm{Div}(S||\alpha)$ of α from S in such a way that a sequence S is α-normal (meaning that every string w has asymptotic frequency $\alpha(w)$ in S) if and only if $\mathrm{Div}(S||\alpha) = 0$. We also use the Kullback-Leibler divergence to quantify the *total risk* $\mathrm{Risk}_G(w)$ that a finite-state gambler G takes when betting along a prefix w of S.

Our main theorem is a *strong dichotomy theorem* that uses the above notions to *quantify* the exponential rates of winning and losing on the two sides of the Schnorr-Stimm dichotomy theorem (with the latter routinely extended from normality to α-normality). Modulo asymptotic caveats in the paper, our strong dichotomy theorem says that the following two things hold for prefixes w of S.

(1′) The infinitely-often exponential rate of winning in 1 is $2^{\mathrm{Div}(S||\alpha)|w|}$.

(2′) The exponential rate of loss in 2 is $2^{-\mathrm{Risk}_G(w)}$.

We also use (1′) to show that $1 - \mathrm{Div}(S||\alpha)/c$, where $c = \log(1/\min_{a\in\Sigma}\alpha(a))$, is an upper bound on the finite-state α-dimension of S and prove the dual fact that $1 - \mathrm{div}(S||\alpha)/c$ is an upper bound on the finite-state strong α-dimension of S.

2012 ACM Subject Classification Theory of computation; Theory of computation → Formal languages and automata theory

Keywords and phrases finite-state dimension, finite-state gambler, Kullback-Leibler divergence, normal sequences

Digital Object Identifier 10.4230/LIPIcs.STACS.2020.51

Funding *Xiang Huang*: This research was supported in part by National Science Foundation Grants 1247051, 1545028, and 1900716.
Jack H. Lutz: This research was supported in part by National Science Foundation Grants 1247051, 1545028, and 1900716.

37th International Symposium on Theoretical Aspects of Computer Science (STACS 2020).
Editors: Christophe Paul and Markus Bläser; Article No. 51; pp. 51:1–51:15

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Elvira Mayordomo: Part of this work was supported by a grant from the Spanish Ministry of Science, Innovation and Universities (TIN2016-80347-R) and was partly done during a research stay at the Iowa State University supported by National Science Foundation Research Grant 1545028.
Donald M. Stull: This research was supported in part by National Science Foundation Grants 1247051, 1545028, and 1900716.

Acknowledgements We thank students in Iowa State University's fall, 2017, advanced topics in computational randomness course for listening to a preliminary version of this research. We thank three anonymous reviewers of an earlier draft of this paper for useful comments and corrections.

1 Introduction

An infinite sequence S over a finite alphabet is *normal* in the 1909 sense of Borel [7] if every two strings of equal length appear with equal asymptotic frequency in S. Borel normality played a central role in the origins of measure-theoretic probability theory [6] and is intuitively regarded as a weak notion of randomness. For a masterful discussion of this intuition, see section 3.5 of [22], where Knuth calls normal sequences "∞-distributed sequences."

The theory of computing was used to make this intuition precise. This took place in three steps in the 1960s and 1970s. First, Martin-Löf [28] used constructive measure theory to give the first successful formulation of the randomness of individual infinite binary sequences. Second, Schnorr [30] gave an equivalent, and more flexible, formulation of Martin-Löf's notion in terms of gambling strategies called *martingales*. In this formulation, an infinite binary sequences S is random if no lower semicomputable martingale can make unbounded money betting on the successive bits of S. Third, Schnorr and Stimm [31] proved that an infinite binary sequence S is normal if and only if no martingale that is computed by a finite-state automaton can make unbounded money betting on the successive bits of S. That is, *normality is finite-state randomness.*

This equivalence was a breakthrough that has already had many consequences (discussed later in this introduction), but the Schnorr-Stimm result said more. It is a *dichotomy theorem* asserting that, for any infinite binary sequence S, the following two things are true.

1. If S is not normal, then there is a finite-state gambler that makes money at an infinitely-often exponential rate when betting on S.
2. If S is normal, then every finite-state gambler that bets infinitely many times on S loses money at an exponential rate.

The main contribution of this paper is to *quantify* the exponential rates of winning and losing on the two sides (1 and 2 above) of the Schnorr-Stimm dichotomy.

To describe our main theorem in some detail, let Σ be a finite alphabet. It is routine to extend the above notion of normality to an arbitrary probability measure α on Σ. Specifically, an infinite sequence S over Σ is *α-normal* if every finite string w over Σ appears with asymptotic frequency $\alpha^{|w|}(w)$ in S, where α^ℓ is the natural (product) extension of α to strings of length ℓ. Schnorr and Stimm [31] correctly noted that their dichotomy theorem extends to α-normal sequences in a straightforward manner, and it is this extension whose exponential rates we quantify here.

The quantitative tool that drives our approach is the Kullback-Leibler divergence [23], also known as the relative entropy [12]. If α and β are probability measures on Σ, then the *Kullback-Leibler divergence of β from α* is

$$D(\alpha||\beta) = E_\alpha \log \frac{\alpha}{\beta},$$

i.e., the expectation with respect to α of the random variable $\log \frac{\alpha}{\beta} : \Sigma \to \mathbb{R} \cup \{\infty\}$, where the logarithm is base-2. Although the Kullback-Leibler divergence is not a metric on the space of probability measures on Σ, it does quantify "how different" β is from α, and it has the crucial property that $D(\alpha||\beta) \geq 0$, with equality if and only if $\alpha = \beta$.

Here we use the empirical frequencies of symbols in S to define the *asymptotic lower divergence* $\mathrm{div}(S||\alpha)$ of α from S and the *asymptotic upper divergence* $\mathrm{Div}(S||\alpha)$ of α from S in a natural way, so that S is α-normal if and only if $\mathrm{Div}(S||\alpha) = 0$.

The first part of our *strong dichotomy theorem* says that the infinitely-often exponential rate that can be achieved in 1 above is essentially at least $2^{\mathrm{Div}(S||\alpha)|w|}$, where w is the prefix of S on which the finite-state gambler has bet so far. More precisely, it says the following.

1′. If S is not α-normal, then, for every $\gamma < 1$, there is a finite-state gambler G such that, when G bets on S with payoffs according to α, there are infinitely many prefixes w of S after which G's capital exceeds $2^{\gamma\,\mathrm{Div}(S||\alpha)|w|}$.

The second part of our strong dichotomy theorem, like the second part of the Schnorr-Stimm dichotomy theorem, is complicated by the fact that a finite-state gambler may, in some states, decline to bet. In this case, its capital after a bet is the same as it was before the bet, regardless of what symbol actually appears in S. Once again, however, it is the Kullback-Leibler divergence that clarifies the situation. As explained in section 3 below, in any particular state q, a finite-state gambler's betting strategy is a probability measure $B(q)$ on Σ. If $B(q) = \alpha$, then the gambler does not bet in state q. We thus define the *risk* that the gambler G takes in state q to be

$$\mathrm{risk}_G(q) = D(\alpha||B(q)),$$

i.e., the divergence of $B(q)$ from not betting. We then define the *total risk* that the gambler takes along a prefix w of the sequence S on which it is betting to be the sum $\mathrm{Risk}_G(w)$ of the risks $\mathrm{risk}_G(q)$ in the states that G traverses along w. The second part of our strong dichotomy theorem says that, if S is α-normal and G is a finite-state gambler betting on S, then after each prefix w of S, the capital of G on prefixes w of S is essentially bounded above by $2^{-\mathrm{Risk}_G(w)}$. In some sense, then, G loses all that it risks. More precisely, the second part of our strong dichotomy says the following.

2′. If S is α-normal, then, for every finite-state gambler G and every $\gamma < 1$, after all but finitely many prefixes w of S, the gambler G's capital is less than $2^{-\gamma\,\mathrm{Risk}_G(w)}$.

A routine ergodic argument, already present in [31], shows that, if a finite-state gambler G bets on an α-normal sequence S, then every state of G that occurs infinitely often along S occurs with positive frequency along S. Hence 2 above follows from 2′ above.

Our strong dichotomy theorem has implications for finite-state dimensions. For each probability measure α on Σ and each sequence S over Σ, the *finite-state α-dimension* $\dim^\alpha_{\mathrm{FS}}(S)$ and the *finite-state strong α-dimension* $\mathrm{Dim}^\alpha_{\mathrm{FS}}(S)$ (defined in section 4 below) are finite-state versions of Billingsley dimension [5, 10] introduced in [26]. When α is the uniform probability measure on Σ, these are the finite dimension $\dim_{\mathrm{FS}}(S)$, introduced in [14] as a finite-state version of Hausdorff dimension [20, 17], and the finite-state strong dimension $\mathrm{Dim}_{\mathrm{FS}}(S)$, introduced in [2] as a finite-state version of packing dimension [35, 34, 17]. Intuitively, $\dim^\alpha_{\mathrm{FS}}(S)$ and $\mathrm{Dim}^\alpha_{\mathrm{FS}}(S)$ measure the lower and upper asymptotic α-densities of the finite-state information in S.

Here we use part 1 of our strong dichotomy theorem to prove that, for every positive probability measure α on Σ and every sequence S over Σ,

$$\dim^{\alpha}_{\mathrm{FS}}(S) \leq 1 - \mathrm{Div}(S||\alpha)/c,$$

where $c = \log(1/\min_{a\in\Sigma} \alpha(a))$. We also establish the dual result that, for all such α and S,

$$\mathrm{Dim}^{\alpha}_{\mathrm{FS}}(S) \leq 1 - \mathrm{div}(S||\alpha)/c.$$

Research on normal sequences and normal numbers (real numbers whose base-b expansions are normal sequences for various choices of b) connected the theory of normal numbers so directly to the theory of computing. Further work along these lines has been continued in [21, 29, 3, 33]. After the discovery of algorithmic dimensions in the present century [24, 25, 14, 2], the Schnorr-Stimm dichotomy led to the realization [8] that the finite-state world, unlike any other known to date, is one in which maximum dimension is not only necessary, but also sufficient, for randomness. This in turn led to the discovery of nontrivial extensions of classical theorems on normal numbers [11, 36] to new quantitative theorems on finite-state dimensions [19, 16], a line of inquiry that will certainly continue. It has also led to a polynomial-time algorithm [4] that computes real numbers that are provably absolutely normal (normal in every base) and, via Lempel-Ziv methods, to a nearly linear time algorithm for this [27]. In parallel with these developments, connections among normality, Weyl equidistribution theorems, and Diophantine approximation have led to a great deal of progress surveyed in the books [15, 9]. This paragraph does not begin to do justice to the breadth and depth of recent and ongoing research on normal numbers and their growing involvement with the theory of computing. It is to be hoped that our strong dichotomy theorem and the quantitative methods implicit in it will further accelerate these discoveries.

2 Divergence and normality

This section reviews the discrete Kullback-Leibler divergence, introduces asymptotic extensions of this divergence, and uses these to give useful characterizations of Borel normal sequences.

2.1 The Kullback-Leibler divergence

We work in a finite alphabet Σ with $2 \leq |\Sigma| < \infty$. We write Σ^{ℓ} for the set of strings of length ℓ over Σ, $\Sigma^* = \bigcup_{\ell=0}^{\infty} \Sigma^{\ell}$ for the set of (finite) *strings* over Σ, Σ^{ω} for the set of (infinite) *sequences* over Σ, and $\Sigma^{\leq\omega} = \Sigma^* \cup \Sigma^{\omega}$. We write λ for the empty string, $|w|$ for the length of a string $w \in \Sigma^*$, and $|S| = \omega$ for the length of a sequence $S \in \Sigma^{\omega}$. For $x \in \Sigma^{\leq\omega}$ and $0 \leq i < |x|$, we write $x[i]$ for the i-th symbol in x, noting that $x[0]$ is the leftmost symbol in x. For $x \in \Sigma^{\leq\omega}$ and $0 \leq i \leq j < |x|$, we write $x[i..j]$ for the string consisting of the i-th through j-th symbols in x. Specially, we write $x \upharpoonright n$ to mean $x[0..n-1]$. For $x,\ y \in \Sigma^{\leq\omega}$, we write $x \sqsubseteq y$ if x is a prefix of y. We write $x \sqsubset y$ to denote x being a strict prefix of y, which excludes the case $x = y$.

A *(discrete) probability measure* on a nonempty finite set Ω is a function $\pi : \Omega \to [0, 1]$ satisfying

$$\sum_{\omega\in\Omega} \pi(\omega) = 1. \tag{2.1}$$

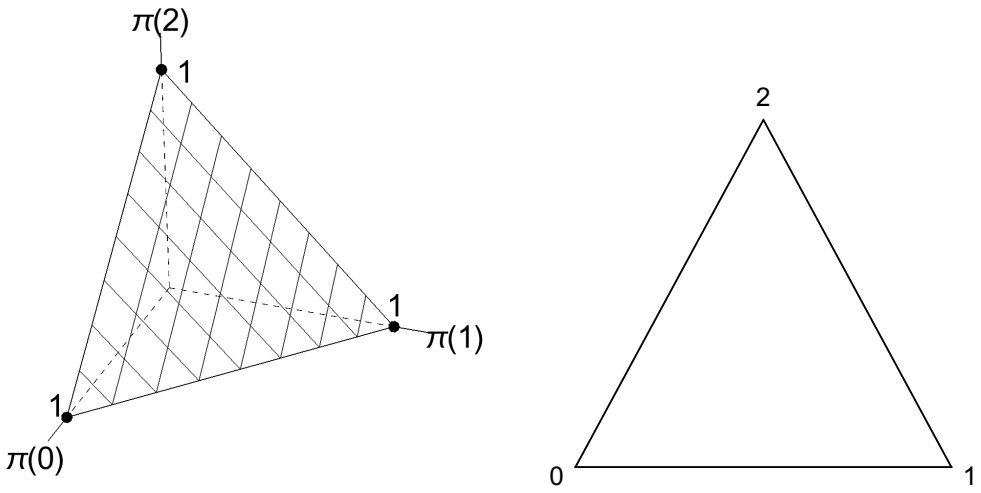

Figure 1 Two views of the simplex $\Delta(\{0,1,2\})$.

We write $\Delta(\Omega)$ for the set of all probability measures on Ω, $\Delta^+(\Omega)$ for the set of all $\pi \in \Delta(\Omega)$ that are *strictly positive* (i.e., $\pi(\omega) > 0$ for all $\omega \in \Omega$), $\Delta_{\mathbb{Q}}(\Omega)$ for the set of all $\pi \in \Delta(\Omega)$ that are rational-valued, and $\Delta^+_{\mathbb{Q}}(\Omega) = \Delta^+(\Omega) \cap \Delta_{\mathbb{Q}}(\Omega)$. In this paper we are most interested in the case where $\Omega = \Sigma^\ell$ for some $\ell \in \mathbb{Z}^+$.

Intuitively, we identify each probability measure $\pi \in \Delta(\Omega)$ with the point in $\mathbb{R}^{|\Omega|}$ whose coordinates are the probabilities $\pi(\omega)$ for $\omega \in \Omega$. By (2.1) this implies that $\Delta(\Omega)$ is the $(|\Omega| - 1)$-dimensional simplex in $\mathbb{R}^{|\Omega|}$ whose vertices are the points at 1 on each of the coordinate axes. (See Figure 1 for an illustration with $|\Omega| = 3$.) For each $\omega \in \Omega$, the vertex on axis ω is the degenerate probability measures π_ω with $\pi_\omega(\omega) = 1$. The centroid of the simplex $\Delta(\Omega)$ is the uniform probability measure on Ω, and the (topological) interior of $\Delta(\Omega)$ is $\Delta^+(\Omega)$. We write $\partial\Delta(\Omega) = \Delta(\Omega) \setminus \Delta^+(\Omega)$ for the boundary of $\Delta(\Omega)$.

▶ **Definition.** ([23]). Let $\alpha, \beta \in \Delta(\Omega)$, where Ω is a nonempty finite set. The *Kullback-Leibler divergence* (or *KL-divergence*) *of* β *from* α is

$$D(\alpha||\beta) = E_\alpha \log \frac{\alpha}{\beta}, \tag{2.2}$$

where the logarithm is base-2.

Note that the right-hand side of (2.2) is the α-expectation of the random variable

$$\log \frac{\alpha}{\beta} : \Omega \longrightarrow \mathbb{R}$$

defined by

$$\left(\log \frac{\alpha}{\beta}\right)(\omega) = \log \frac{\alpha(\omega)}{\beta(\omega)}$$

for each $\omega \in \Omega$. Hence (2.2) is a convenient shorthand for

$$D(\alpha||\beta) = \sum_{\omega \in \Omega} \alpha(\omega) \log \frac{\alpha(\omega)}{\beta(\omega)}.$$

Note also that $D(\alpha||\beta)$ is infinite if and only if $\alpha(\omega) > 0 = \beta(\omega)$ the some $\omega \in \Omega$.

The Kullback-Leibler divergence $D(\alpha||\beta)$ is a useful measure of how different β is from α. It is not a metric (because it is not symmetric and does not satisfy the triangle inequality), but it has the crucial property that $D(\alpha||\beta) \geq 0$, with equality if and only if $\alpha = \beta$. The two most central quantities in Shannon information theory, entropy and mutual information, can both be defined in terms of divergence as follows.

1. *Entropy is divergence from certainty.* The *entropy* of a probability measure $\alpha \in \Delta(\Omega)$, conceived by Shannon [32] as a measure of the uncertainty of α, is

$$H(\alpha) = \sum_{\omega \in \Omega} \alpha(\omega) D(\pi_\omega||\alpha), \tag{2.3}$$

   i.e., the α-average of the divergences of α from the "certainties" π_ω.
2. *Mutual information is divergence from independence.* If α, $\beta \in \Delta(\Omega)$ have a joint probability measure $\gamma \in \Delta(\Omega \times \Omega)$ (i.e., are the marginal probability measures of γ), then the *mutual information* between α and β, conceived by Shannon [32] as a measure of the information shared by α and β, is

$$I(\alpha\,;\,\beta) = D(\alpha\beta||\gamma), \tag{2.4}$$

   i.e., the divergence of γ from the probability measure in which α and β are independent.

Two additional properties of the Kullback-Leibler divergence are useful for our asymptotic concerns. First, the divergence $D(\alpha||\beta)$ is continuous on $\Delta(\Omega)^2$ (as a function into $[0, \infty]$). Hence, if $\alpha_n \in \Delta(\Omega)$ for each $n \in \mathbb{N}$ and $\lim_{n\to\infty} \alpha_n = \alpha$ in the sense of the Euclidean metric on the simplex $\Delta(\Omega)$, then $\lim_{n\to\infty} D(\alpha_n||\alpha) = \lim_{n\to\infty} D(\alpha||\alpha_n) = 0$. Second, the converse holds. It is well known [12] that

$$D(\alpha||\beta) \geq \frac{1}{2\ln 2}\|\alpha - \beta\|_1^2,$$

where $\|\alpha - \beta\|_1 = \sum_{\omega\in\Omega}|\alpha(\omega) - \beta(\omega)|$ is the $\mathcal{L}_1$-norm. Hence, if either $\lim_{n\to\infty} D(\alpha_n||\alpha) = 0$ or $\lim_{n\to\infty} D(\alpha||\alpha_n) = 0$, then $\lim_{n\to\infty} \alpha_n = \alpha$.

More extensive discussions of the Kullback-Leibler divergence appear in [12, 13].

2.2 Asymptotic divergences

For nonempty strings $w, x \in \Sigma^*$, we write

$$\#_\square(w, x) = \left|\{m \leq \frac{|x|}{|w|} - 1 \;\middle|\; x[m|w|..(m+1)|w| - 1] = w\}\right|$$

for the *number of block occurrences* of w in x. Note that $0 \leq \#_\square(w, x) \leq \frac{|x|}{|w|}$.

For each $S \in \Sigma^\omega$, $n \in \mathbb{Z}^+$, and $\lambda \neq w \in \Sigma^*$, the n-th *block frequency* of w in S is

$$\pi_{S,n}(w) = \frac{\#_\square(w, S[0..n|w|-1])}{n} \tag{2.5}$$

Note that (2.5) defines, for each $S \in \Sigma^\omega$ and $n \in \mathbb{Z}^+$, a function

$$\pi_{S,n} : \Sigma^* \smallsetminus \{\lambda\} \longrightarrow \mathbb{Q}.$$

For each such S and n and each $\ell \in \mathbb{Z}^+$, let $\pi_{S,n}^{(\ell)} = \pi_{S,n} \upharpoonright \Sigma^\ell$ be the restriction of the function $\pi_{S,n}$ to the set Σ^ℓ of strings of length ℓ.

▶ **Observation 2.1.** For each $S \in \Sigma^\omega$ and $n, \ell \in \mathbb{Z}^+$,

$$\pi^{(\ell)}_{S,n} \in \Delta_{\mathbb{Q}}(\Sigma^\ell),$$

i.e., $\pi^{(\ell)}_{S,n}$ is a rational-valued probability measure on Σ^ℓ.

We call $\pi^{(\ell)}_{S,n}$ the n-th *empirical probability measure on* Σ^ℓ *given by* S.

A probability measure $\alpha \in \Delta(\Sigma)$ naturally induces, for each $\ell \in \mathbb{Z}^+$, a probability measure $\alpha^{(\ell)} \in \Delta(\Sigma^\ell)$ defined by

$$\alpha^{(\ell)}(w) = \prod_{i=0}^{|w|-1} \alpha(w[i]). \tag{2.6}$$

The empirical probability measures $\pi^{(\ell)}_{S,n}$ provide a natural way to define useful empirical divergences of probability measures from sequences.

▶ **Definition.** Let $\ell \in \mathbb{Z}^+$, $S \in \Sigma^\omega$, and $\alpha \in \Delta(\Sigma)$.

1. The *lower ℓ-divergence of α from S* is $\mathrm{div}_\ell(S||\alpha) = \liminf_{n\to\infty} D(\pi^{(\ell)}_{S,n}||\alpha^{(\ell)})$.
2. The *upper ℓ-divergence of α from S* is $\mathrm{Div}_\ell(S||\alpha) = \limsup_{n\to\infty} D(\pi^{(\ell)}_{S,n}||\alpha^{(\ell)})$.
3. The *lower divergence of α from S* is $\mathrm{div}(S||\alpha) = \sup_{\ell\in\mathbb{Z}^+} \mathrm{div}_\ell(S||\alpha)/\ell$.
4. The *upper divergence of α from S* is $\mathrm{Div}(S||\alpha) = \sup_{\ell\in\mathbb{Z}^+} \mathrm{Div}_\ell(S||\alpha)/\ell$.

A similar approach gives useful empirical divergences of one sequence from another.

▶ **Definition.** Let $\ell \in \mathbb{Z}^+$ and $S, T \in \Sigma^\omega$.

1. The *lower ℓ-divergence of T from S* is $\mathrm{div}_\ell(S||T) = \liminf_{n\to\infty} D(\pi^{(\ell)}_{S,n}||\,\pi^{(\ell)}_{T,n})$.
2. The *upper ℓ-divergence of T from S* is $\mathrm{Div}_\ell(S||T) = \limsup_{n\to\infty} D(\pi^{(\ell)}_{S,n}||\,\pi^{(\ell)}_{T,n})$.
3. The *lower divergence of T from S* is $\mathrm{div}(S||T) = \sup_{\ell\in\mathbb{Z}^+} \mathrm{div}_\ell(S||T)/\ell$.
4. The *upper divergence of T from S* is $\mathrm{Div}(S||T) = \sup_{\ell\in\mathbb{Z}^+} \mathrm{Div}_\ell(S||T)/\ell$.

2.3 Normality

The following notions are essentially due to Borel [7].

▶ **Definition.** Let $\alpha \in \Delta(\Sigma)$, $S \in \Sigma^\omega$, and $\ell \in \mathbb{Z}^+$.

1. S is *α-ℓ-normal* if, for all $w \in \Sigma^\ell$,

$$\lim_{n\to\infty} \pi_{S,n}(w) = \alpha^{(\ell)}(w).$$

2. S is *α-normal* if, for all $\ell \in \mathbb{Z}^+$, S is α-ℓ-normal.
3. S is *ℓ-normal* if S is μ-ℓ-normal, where μ is the uniform probability measure on Σ.
4. S is *normal* if, for all $\ell \in \mathbb{Z}^+$, S is ℓ-normal.

▶ **Lemma 2.2.** *For all $\alpha \in \Delta(\Sigma)$, $S \in \Sigma^\omega$, and $\ell \in \mathbb{Z}^+$, the following four conditions are equivalent.*

(1) *S is α-ℓ-normal.*
(2) $\mathrm{Div}_\ell(S||\alpha) = 0$.
(3) *For every α-ℓ-normal sequence $T \in \Sigma^\omega$, $\mathrm{Div}_\ell(S||T) = 0$.*
(4) *There exists an α-ℓ-normal sequence $T \in \Sigma^\omega$ such that $\mathrm{Div}_\ell(S||T) = 0$.*

Lemma 2.2 (proved in the appendix) immediately implies the following.

▶ **Theorem 2.3** (divergence characterization of normality). *For all $\alpha \in \Delta(\Sigma)$ and $S \in \Sigma^\omega$, the following conditions are equivalent.*

(1) *S is α-normal.*
(2) $\mathrm{Div}(S||\alpha) = 0$.
(3) *For every α-normal sequence $T \in \Sigma^\omega$,* $\mathrm{Div}(S||T) = 0$.
(4) *There exists an α-normal sequence $T \in \Sigma^\omega$ such that* $\mathrm{Div}(S||T) = 0$.

3 Strong Dichotomy

This section presents our main theorem, the strong dichotomy theorem for finite-state gambling. We first review finite-state gamblers.

Fix a finite alphabet Σ with $|\Sigma| \geq 2$.

▶ **Definition** ([31, 18, 14]). A *finite-state gambler (FSG)* is a 4-tuple

$$G = (Q, \delta, s, B),$$

where Q is a finite set of *states*, $\delta : Q \times \Sigma \to Q$ is the *transition function*, $s \in Q$ is the *start state*, and $B : Q \to \Delta_{\mathbb{Q}}(\Sigma)$ is the *betting function*.

The transition structure (Q, δ, s) here works as in any deterministic finite-state automaton. For $w \in \Sigma^*$, we write $\delta(w)$ for the state reached from s by processing w.

Intuitively, a gambler $G = (Q, \delta, s, B)$ bets on the successive symbols of a sequence $S \in \Sigma^\omega$. The payoffs in the betting are determined by a *payoff probability measure* $\alpha \in \Delta(\Sigma)$. (We regard α and S as external to the gambler G.) We write $d_{G,\alpha}(w)$ for the gambler G's capital (amount of money) after betting on the successive bits of a prefix $w \sqsubseteq S$, and we assume that the initial capital is $d_{G,\alpha}(\lambda) = 1$.

The meaning of the betting function B is as follows. After betting on a prefix $w \sqsubseteq S$, the gambler is in state $\delta(w) \in Q$. The betting function B says that, for each $a \in \Sigma$, the gambler bets the fraction $B(\delta(w))(a)$ of its current capital $d_{G,\alpha}(w)$ that $wa \sqsubseteq S$, i.e., that the next symbol of S is an a. If it then turns out to be the case that $wa \sqsubseteq S$, the gambler's capital will be

$$d_{G,\alpha}(wa) = d_{G,\alpha}(w)\frac{B(\delta(w))(a)}{\alpha(a)}. \tag{3.1}$$

(Note: If $\alpha(a) = 0$ here, we may define $d_{G,\alpha}(wa)$ however we wish.)

The payoffs in (3.1) are fair with respect to α, which means that the conditional α-expectation

$$\sum_{a \in \Sigma} \alpha(a) d_{G,\alpha}(wa)$$

of $d_{G,\alpha}(wa)$, given that $w \sqsubseteq S$, is exactly $d_{G,\alpha}(w)$. This says that the function $d_{G,\alpha}$ is an *α-martingale*.

If $\delta(w) = q$ is a state in which $B(q) = \alpha$, then (3.1) says that, for each $a \in \Sigma$, $d_{G,\alpha}(wa) = d_{G,\alpha}(w)$. That is, the condition $B(q) = \alpha$ means that G *does not bet* in state q. Accordingly, we define the *risk* that G takes in a state $q \in Q$ to be

$$\mathrm{risk}_G(q) = D(\alpha||B(q)).$$

i.e., the divergence of $B(q)$ from not betting. We also define the *total risk* that G takes *along* a string $w \in \Sigma^*$ to be

$$\mathrm{Risk}_G(w) = \sum_{u \sqsubset w} \mathrm{risk}_G(\delta(u)).$$

We now state our main theorem.

▶ **Theorem 3.1** (strong dichotomy theorem). *Let $\alpha \in \Delta(\Sigma)$, $S \in \Sigma^\omega$, and $\gamma < 1$.*

1. *If S is not α-normal, then there is a finite-state gambler G such that, for infinitely many prefixes $w \sqsubseteq S$,*

$$d_{G,\alpha}(w) > 2^{\gamma\, \mathrm{Div}(S||\alpha)|w|}.$$

2. *If S is α-normal, then, for every finite-state gambler G, for all but finitely many prefixes $w \sqsubseteq S$,*

$$d_{G,\alpha}(w) < 2^{-\gamma\, \mathrm{Risk}_G(w)}.$$

Proof. To prove the first part, let S be a non-normal sequence. Then by Theorem 2.3, we know that $\mathrm{Div}(S||\alpha) > 0$. Let $r < 1$ and let $\epsilon > 0$. By the definition of $\mathrm{Div}(S||\alpha)$ there must exist ℓ such that

$$\mathrm{Div}_\ell(S||\alpha)/\ell > r\, \mathrm{Div}(S||\alpha). \tag{3.2}$$

That is

$$\limsup_{n\to\infty} D(\pi^{(\ell)}_{S,n}||\alpha^{(\ell)}) > \ell r\, \mathrm{Div}(S||\alpha).$$

We can pick a subsequence of indices n_k's, such that $\lim_{k\to\infty} D(\pi^{(\ell)}_{S,n_k}||\alpha^{(\ell)}) = \mathrm{Div}_\ell(S||\alpha)$. Therefore by inequality (3.2)

$$D(\pi^{(\ell)}_{S,n_k}||\alpha^{(\ell)}) > \ell r\, \mathrm{Div}(S||\alpha) \tag{3.3}$$

for sufficiently large k. In particular, by compactness of $[0,1]^{|\Sigma^\ell|}$ equipped with $\mathcal{L}_1$-norm, we can further request that

$$\lim_{k\to\infty} \pi^{(\ell)}_{S,n_k} \quad \text{exists.} \tag{3.4}$$

Let $\pi_0 = \pi_0(r,m) \in \Delta_{\mathbb{Q}}(\Sigma^\ell)$ be the m-th $\pi^{(\ell)}_{S,n_k}$ that satisfies (3.3), indexed by k. By the way we define π_0, we have

$$D(\pi_{S,n_k}||\alpha^{(\ell)}) \geq D(\pi_0||\alpha^{(\ell)}) > \ell r\, \mathrm{Div}(S||\alpha), \tag{3.5}$$

and

$$\|\pi_0 - \pi^{(\ell)}_{S,n_k}\| \to 0, \quad \text{as } m \to \infty \text{ and } k \to \infty, \tag{3.6}$$

whence D's continuity in section 2.1 tells us that

$$D(\pi^{(\ell)}_{S,n_k}||\pi_0) \to 0, \quad \text{as } m \to \infty \text{ and } k \to \infty. \tag{3.7}$$

Also note that,

$$\lim_{m\to\infty} D(\pi_0||\alpha^{(\ell)}) = \lim_{k\to\infty} D(\pi^{(\ell)}_{S,n_k}||\alpha^{(\ell)}) = \mathrm{Div}_\ell(S||\alpha) > 0. \tag{3.8}$$

For a fixed $\pi_0 = \pi_0(r, m)$, by the definition, for any n_k sufficiently large, we have

$$D(\pi^{(\ell)}_{S,n_k} || \alpha^{(\ell)}) > D(\pi_0 || \alpha^{(\ell)})(1-\epsilon) > 0. \tag{3.9}$$

By doing the above we pick a probability measure π_0 that is "far" away from $\alpha^{(\ell)}$, we now hard code π_0 in a gambler $G = (Q, \delta, s, B)$, where

$$Q = \Sigma^{\leq \ell-1}, \qquad \delta(w,a) = \begin{cases} wa & \text{if } |wa| < \ell \\ \lambda & \text{if } |wa| = \ell \end{cases}, \qquad s = \lambda, \qquad \textit{and} \qquad B(w)(a) = \pi_0(a|w),$$

where $\pi_0(a|w)$ describes the conditional probability (induced by π_0) of occurrence of an a after $w \in Q$, and is defined by $\pi_0(a|w) = \pi_0(wa)/\pi_0(w)$, where for $u \in Q$, the notation $\pi_0(u)$ is defined recursively by $\pi_0(w) = \sum_{a\in\Sigma} \pi_0(wa)$.

Let $u = a_0 \cdots a_{\ell-1}$ be in Σ^ℓ. The following observation captures the above intuition:

$$\frac{B(\lambda)(a_0)\cdots B(u[0..\ell-2])(a_{\ell-1})}{\alpha(a_0)\cdots\alpha(a_{\ell-1})} = \frac{\pi_0(u)}{\alpha^{(\ell)}(u)}.$$

Now let $w = S \upharpoonright n_k$ for some k. We can view w as

$$w = u_0 u_1 \cdots u_{n-1} u_n, \quad \text{where } |u_i| = \ell \text{ for } 0 \leq i \leq n-1 \text{ and } u_n = a_0 \cdots a_m \text{ with } m < \ell.$$

Then we have

$$d_{G,\alpha}(w) = \Big(\prod_0^{n-1} \frac{\pi_0(u_i)}{\alpha^{(\ell)}(u_i)}\Big) \frac{B(\lambda)(a_0)\cdots B(u_n[0..m-1])(a_m)}{\alpha(a_0)\cdots\alpha(a_m)} \geq C_0 \prod_0^{n-1} \frac{\pi_0(u_i)}{\alpha^{(\ell)}(u_i)}, \tag{3.10}$$

where C_0 is the minimum value of $\frac{B(\lambda)(a_0)\cdots B(u_n[0..m-1])(a_m)}{\alpha(a_0)\cdots\alpha(a_{\ell-1})}$, where $u_n = a_0\cdots a_m$ ranges over $\Sigma^{<\ell}$. Taking log on both sides of (3.10) we get

$$\begin{aligned}
\log d_{G,\alpha}(w) - \log C_0 &\geq \sum_{i=0}^{n-1} \log \frac{\pi_0(u_i)}{\alpha^{(\ell)}(u_i)} = \sum_{|u|=\ell} \#_\square(u,w) \log \frac{\pi_0(u)}{\alpha^{(\ell)}(u)}, \\
&= n \sum_{|u|=\ell} \frac{\#_\square(u,w)}{n} \log \frac{\pi_0(u)}{\alpha^{(\ell)}(u)} = n \sum_{|u|=\ell} \pi^{(\ell)}_{S,n}(u) \log \frac{\pi_0(u)}{\alpha^{(\ell)}(u)}, \\
&= n \sum_{|u|=\ell} \Big[\pi^{(\ell)}_{S,n}(u) \log \frac{\pi^{(\ell)}_{S,n}(u)}{\alpha^{(\ell)}(u)} - \pi^{(\ell)}_{S,n}(u) \log \frac{\pi^{(\ell)}_{S,n}(u)}{\pi_0(u)}\Big] \\
&= n\Big(D(\pi^{(\ell)}_{S,n} || \alpha^{(\ell)}) - D(\pi^{(\ell)}_{S,n} || \pi_0)\Big)
\end{aligned} \tag{3.11}$$

Then by (3.9), for $w = S \upharpoonright n_k$ long enough, we have

$$\begin{aligned}
\log d_{G,\alpha}(w) - \log C_0 &\geq n\Big(D(\pi^{(\ell)}_{S,n} || \alpha^{(\ell)}) - D(\pi^{(\ell)}_{S,n} || \pi_0)\Big) \\
&\geq n\Big(D(\pi_0 || \alpha^{(\ell)})(1-\epsilon) - D(\pi^{(\ell)}_{S,n} || \pi_0)\Big) \geq \frac{|w|}{\ell} D(\pi_0 || \alpha^{(\ell)})(1-2\epsilon).
\end{aligned}$$

Therefore, by (3.5) we have

$$d_{G,\alpha}(w) \geq C_0 2^{\frac{|w|(1-2\epsilon)}{\ell} D(\pi_0 || \alpha^{(\ell)})} \geq 2^{|w| r(1-2\epsilon)\, \mathrm{Div}(S||\alpha)}.$$

Since r and $1-2\epsilon$ can be picked arbitrary close to 1, take $r(1-2\epsilon) > \gamma$, then

$$d_{G,\alpha}(w) \geq 2^{\gamma\, \mathrm{Div}(S||\alpha)|w|}$$

for $w = S \upharpoonright n_k$ long enough.

We now prove the second part of the main theorem.

Let S be a normal number, G an arbitrary finite-state gambler. By Proposition 2.5 of [31], $G = (Q, \delta, s, B)$ will eventually reach a *bottom strongly connected component* (a component that has no path to leave) when processing S. A similar argument can also be found in [33]. Without loss of generality, we will therefore assume that every state in G is recurrent in processing S.

Let $w = a_0 \cdots a_{n-1} \sqsubseteq S$. Then

$$\begin{aligned} d_{G,\alpha}(w) &= \frac{B(\delta(\lambda))(a_0) \cdots B(\delta(w[a_0..a_{n-2}]))(a_{n-1})}{\alpha(a_0) \cdots \alpha(a_{n-1})} \\ &= \prod_{q \in Q} \prod_{a \in \Sigma} \left(\frac{B(q)(a)}{\alpha(a)} \right)^{\#_{G,w}(q,a)}, \end{aligned} \tag{3.12}$$

where the notation $\#_{G,w}(q,a)$ denotes the number of times G lands on state q and the next symbol is a while processing w. Similarly, we use the notation $\#_{G,w}(q)$ to denote the number of times G lands on q in the same process.

Taking the logarithm of both sides of (3.12), we have

$$\begin{aligned} \log d_{G,\alpha}(w) &= \sum_{q \in Q} \sum_{a \in \Sigma} \#_{G,w}(q,a) \log \frac{B(q)(a)}{\alpha(a)} \\ &= \sum_{q \in Q} \#_{G,w}(q) \sum_{a \in \Sigma} \frac{\#_{G,w}(q,a)}{\#_{G,w}(q)} \log \frac{B(q)(a)}{\alpha(a)} \end{aligned} \tag{3.13}$$

By a result of Agafonov [1], which extends easily to the arbitrary probability measures considered here, we have that, for every state q, the limit of $\frac{\#_{G,w}(q,a)}{\#_{G,w}(s)}$ along S exists and converges to $\alpha(a)$. That is

$$\lim_{w \to S} \frac{\#_{G,w}(q,a)}{\#_{G,w}(s)} = \alpha(a), \tag{3.14}$$

for every state q.

Therefore, by equations (3.13) and (3.14), and the fact that there are finitely many states, we have

$$\begin{aligned} \log d_{G,\alpha}(w) &\leq \sum_{q \in Q} \#_{G,w}(q) \sum_{a \in \Sigma} (\alpha(a) + o(1)) \log \frac{B(q)(a)}{\alpha(a)} \\ &= \sum_{q \in Q} -\operatorname{risk}_G(q) \#_{G,w}(q) + \sum_{q \in Q} \#_{G,w}(q) \sum_{a \in \Sigma} o(1) \log \frac{B(q)(a)}{\alpha(a)} \\ &= -\operatorname{Risk}_G(w) + \sum_{q \in Q} \#_{G,w}(q) \sum_{a \in \Sigma} o(1) \log \frac{B(q)(a)}{\alpha(a)} \\ &= \operatorname{Risk}_G(w)(-1 + o(1)). \end{aligned}$$

It follows that

$$d_{G,\alpha}(w) \leq 2^{-(1+o(1)) \operatorname{Risk}_B(w)},$$

so part 2 of the theorem holds. ◀

4 Dimension

Finite-state dimensions give a particularly sharp formulation of part 1 of the strong dichotomy theorem, along with a dual of this result.

Finite-state dimensions were introduced for the uniform probability measure on Σ in [14, 2] and extended to arbitrary probability measure on Σ in [26]. For each $\alpha \in \Delta(\Sigma)$ and each $S \in \Sigma^\omega$, define the sets

$$\mathfrak{G}^\alpha(S) = \left\{ s \in [0,\infty) \,\middle|\, (\exists \text{FSG } G) \limsup_{w \to S} \alpha^{|w|}(w)^{1-s} d_{G,\alpha}(w) = \infty \right\}$$

and

$$\mathfrak{G}^\alpha_{\text{str}}(S) = \left\{ s \in [0,\infty) \,\middle|\, (\exists \text{FSG } G) \liminf_{w \to S} \alpha^{|w|}(w)^{1-s} d_{G,\alpha}(w) = \infty \right\}$$

The limits superior and inferior here are taken for successively longer prefixes $w \sqsubseteq S$. The "strong" subscript of $\mathfrak{G}_{\text{str}}(S)$ refers to the fact that $\alpha^{|w|}(w)^{1-s} d_{G,\alpha}(w)$ is required to converge to infinity in a stronger sense than in $\mathfrak{G}^\alpha(S)$.

▶ **Definition** ([26]). Let $\alpha \in \Delta(\Sigma)$ and $S \in \Sigma^\omega$.

1. The *finite-state α-dimension* of S is $\dim^\alpha_{\text{FS}}(S) = \inf \mathfrak{G}^\alpha(S)$.
2. The *finite-state strong α-dimension* of S is $\text{Dim}^\alpha_{\text{FS}}(S) = \inf \mathfrak{G}^\alpha_{\text{str}}(S)$

It is easy to see that, for all $\alpha \in \Delta^+(\Sigma)$ and $S \in \Sigma^\omega$, $0 \le \dim^\alpha_{\text{FS}}(S) \le \text{Dim}^\alpha_{\text{FS}}(S) \le 1$.

▶ **Theorem 4.1.** *For all $\alpha \in \Delta(\Sigma)$ and $S \in \Sigma^\omega$ let $c = \log(1/\min_{a\in\Sigma} \alpha(a))$. Then,*

$$\dim^\alpha_{\text{FS}}(S) \le 1 - \text{Div}(S||\alpha)/c$$

and

$$\text{Dim}^\alpha_{\text{FS}}(S) \le 1 - \text{div}(S||\alpha)/c.$$

Proof. Let $t < \text{Div}(S||\alpha)/c$, and let $s = 1 - t$. Fix ℓ such that $\text{Div}_\ell(S||\alpha)/\ell > tc$, then for i.o. n, $D(\pi^{(\ell)}_{S,n}||\alpha^{(\ell)}) > \ell tc$. Note that $\alpha^{|w|}(w) \ge (1/2^c)^{|w|}$ for every $w \in \Sigma^*$.

Define the gambler G be $G = (Q, \delta, s_0, B_n)$, where $Q = \Sigma^{\le \ell-1}$,

$$\delta(w,a) = \begin{cases} wa & \text{if } |wa| < \ell \\ \lambda & \text{if } |wa| = \ell \end{cases}$$

$s_0 = \lambda$, and $B_n(w)(a) = \pi^{(\ell)}_{S,n}(a|w)$, where $\pi^{(\ell)}_{S,n}(a|w)$ describes the conditional probability (induced by $\pi^{(\ell)}_{S,n}$) of occurrence of an a after $w \in Q$.

Let $u = a_0 \cdots a_{\ell-1}$ be in Σ^ℓ.

$$\frac{B_n(\lambda)(a_0)\cdots B_n(u[0..\ell-2])(a_{\ell-1})}{\alpha(a_0)\cdots\alpha(a_{\ell-1})} = \frac{\pi^{(\ell)}_{S,n}(u)}{\alpha^{(\ell)}(u)}.$$

Then for $z \in \Sigma^*$ with $z \sqsubseteq S$ and $|z| = \ell n$, we have

$$\begin{aligned} \alpha^{|z|}(z)^{1-s} d_{G,\alpha}(z) &= \alpha^{|z|}(z)^t d_{G,\alpha}(z) \\ &= \alpha^{|z|}(z)^t \prod_{u\in\Sigma^\ell} \left(\frac{\pi^{(\ell)}_{S,n}(u)}{\alpha^{(\ell)}(u)} \right)^{n\pi^{(\ell)}_{S,n}(u)} \end{aligned}$$

Therefore,

$$\alpha^{|z|}(z)^t d_{G,\alpha}(z) \geq \frac{1}{2^c}^{|z|t} 2^{nD(\pi_{S,n}^{(\ell)}||\alpha^{(\ell)})}$$
$$\geq 2^{-c|z|t+c|z|t}$$

Since the number of states is fixed, this implies $\dim_{\mathrm{FS}}^{\alpha}(S) \leq 1 - \mathrm{Div}(S||\alpha)/c$. The proof of the other case is similar, where we use the fact that, for a.e. n,

$$D(\pi_{S,n}^{(\ell)}||\alpha^{(\ell)}) > \ell tc.$$ ◀

References

1 Valerii Nikolaevich Agafonov. Normal sequences and finite automata. In *Doklady Akademii Nauk*, volume 179 (2), pages 255–256, 1968.
2 Krishna B. Athreya, John M. Hitchcock, Jack H. Lutz, and Elvira Mayordomo. Effective strong dimension in algorithmic information and computational complexity. *SIAM Journal on Computing*, 37(3):671–705, 2007.
3 Verónica Becher and Pablo Ariel Heiber. Normal numbers and finite automata. *Theoretical Computer Science*, 477:109–116, 2013.
4 Verónica Becher, Pablo Ariel Heiber, and Theodore A. Slaman. A polynomial-time algorithm for computing absolutely normal numbers. *Information and Computation*, 232:1–9, 2013.
5 Patrick Billingsley. Hausdorff dimension in probability theory. *Illinois Journal of Mathematics*, 4(2):187–209, 1960.
6 Patrick Billingsley. *Probability and Measure*. Wiley-Interscience, 1995.
7 Émile Borel. Les probabilités dénombrables et leurs applications arithmétiques. *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, 27(1):247–271, 1909.
8 Chris Bourke, John M. Hitchcock, and N. V. Vinodchandran. Entropy rates and finite-state dimension. *Theoretical Computer Science*, 349(3):392–406, 2005.
9 Yann Bugeaud. *Distribution Modulo One and Diophantine Approximation*. Cambridge University Press, 2012.
10 Helmut Cajar. *Billingsley Dimension in Probability Spaces*. Springer-Verlag, 1981.
11 Arthur H. Copeland and Paul Erdös. Note on normal numbers. *Bulletin of the American Mathematical Society*, 52(10):857–860, 1946.
12 Thomas R. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., 2006.
13 Imre Csiszár and Paul C. Shields. *Information Theory and Statistics: A Tutorial*, volume 1 (4). Now Publishers, Inc., 2004.
14 Jack J. Dai, James I. Lathrop, Jack H. Lutz, and Elvira Mayordomo. Finite-state dimension. *Theoretical Computer Science*, 310(1-3):1–33, 2004.
15 Karma Dajani and Cor Kraaikamp. *Ergodic Theory of Numbers*. Cambridge University Press, 2002.
16 David Doty, Jack H. Lutz, and Satyadev Nandakumar. Finite-state dimension and real arithmetic. *Information and Computation*, 205(11):1640–1651, 2007.
17 Kenneth Falconer. *Fractal Geometry: Mathematical Foundations and Applications*. Wiley, 2014.
18 Meir Feder. Gambling using a finite state machine. *IEEE Transactions on Information Theory*, 37(5):1459–1465, 1991.
19 Xiaoyang Gu, Jack H. Lutz, and Philippe Moser. Dimensions of Copeland–Erdös sequences. *Information and Computation*, 205(9):1317–1333, 2007.
20 F. Hausdorff. Dimension und äußeres maß. *Mathematische Annalen*, 79:157–179, 1919.
21 Teturo Kamae and Benjamin Weiss. Normal numbers and selection rules. *Israel Journal of Mathematics*, 21(2-3):101–110, 1975.

22 Donald E. Knuth. *Art of Computer Programming, volume 2: Seminumerical Algorithms*. Addison-Wesley Professional, 2014.

23 Solomon Kullback and Richard A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.

24 Jack H. Lutz. Dimension in complexity classes. *SIAM Journal on Computing*, 32(5):1236–1259, 2003.

25 Jack H. Lutz. The dimensions of individual strings and sequences. *Information and Computation*, 187(1):49–79, 2003. doi:10.1016/S0890-5401(03)00187-1.

26 Jack H. Lutz. A divergence formula for randomness and dimension. *Theoretical Computer Science*, 412(1-2):166–177, 2011.

27 Jack H. Lutz and Elvira Mayordomo. Computing absolutely normal numbers in nearly linear time. *arXiv preprint*, 2016. arXiv:1611.05911.

28 Per Martin-Löf. The definition of random sequences. *Information and Control*, 9(6):602–619, 1966.

29 Wolfgang Merkle and Jan Reimann. Selection functions that do not preserve normality. *Theory of Computing Systems*, 39(5):685–697, 2006.

30 Claus-Peter Schnorr. A unified approach to the definition of random sequences. *Mathematical Systems Theory*, 5(3):246–258, 1971.

31 Claus-Peter Schnorr and Hermann Stimm. Endliche Automaten und Zufallsfolgen. *Acta Informatica*, 1(4):345–359, 1972.

32 Claude Elwood Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

33 Alexander Shen. Automatic Kolmogorov complexity and normality revisited. In *International Symposium on Fundamentals of Computation Theory*, pages 418–430. Springer, 2017.

34 Dennis Sullivan. Entropy, Hausdorff measures old and new, and limit sets of geometrically finite Kleinian groups. *Acta Mathematica*, 153(1):259–277, 1984.

35 Claude Tricot. Two definitions of fractional dimension. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 91 (1), pages 57–74. Cambridge University Press, 1982.

36 Donald Dines Wall. *Normal numbers*. Ph.D. thesis, University of California, Berkeley, 1949.

A Appendix

The following is a proof for Lemma 2.2.

Proof. Let α, S, and ℓ be as given.

To see that (1) implies (2), assume (1). Then $\lim_{n\to\infty} \pi^{(\ell)}_{S,n} = \alpha^{(\ell)}$, so the continuity of KL-divergence tells us that

$$\mathrm{Div}_\ell(S||\alpha) = \lim_{n\to\infty} D(\pi^{(\ell)}_{S,n}||\alpha^{(\ell)}) = 0,$$

i.e., that (2) holds.

To see that (2) implies (3), assume (2). Then $\lim_{n\to\infty} D(\pi^{(\ell)}_{S,n}||\alpha^{(\ell)}) = \mathrm{Div}_\ell(S||\alpha) = 0$, whence the $\mathcal{L}_1$ bound in section 2.1 tells us that $\lim_{n\to\infty} \pi^{(\ell)}_{S,n} = \alpha^{(\ell)}$. For any α-ℓ-normal sequence $T \in \Sigma^\omega$, we have $\lim_{n\to\infty} \pi^{(\ell)}_{T,n} = \alpha^{(\ell)}$, whence the continuity of KL-divergence tells us that

$$\mathrm{Div}_\ell(S||T) = \lim_{n\to\infty} D(\pi^{(\ell)}_{S,n}||\pi^{(\ell)}_{T,n}) = D(\alpha^{(\ell)}||\alpha^{(\ell)}) = 0,$$

i.e., that (3) holds.

Since α-ℓ-normal sequences exist, it is trivial that (3) implies (4).

Finally, to see that (4) implies (1), assume that (4) holds. Then we have

$$\lim_{n\to\infty} D(\pi_{S,n}^{(\ell)}||\pi_{T,n}^{(\ell)}) = \mathrm{Div}_\ell(S||T) = 0,$$

whence the $\mathcal{L}_1$ bound in section 2.1 tells us that

$$\lim_{n\to\infty} \|\pi_{S,n}^{(\ell)} - \pi_{T,n}^{(\ell)}\|_1 = 0. \tag{A.1}$$

We also have

$$\lim_{n\to\infty} \pi_{T,n}^{(\ell)} = \alpha^{(\ell)},$$

whence

$$\lim_{n\to\infty} \|\pi_{T,n}^{(\ell)} - \alpha^{(\ell)}\|_1 = 0. \tag{A.2}$$

By (A.1), (A.2), and the triangle inequality for the $\mathcal{L}_1$-norm, we have

$$\lim_{n\to\infty} \|\pi_{S,n}^{(\ell)} - \alpha^{(\ell)}\|_1 = 0,$$

whence

$$\lim_{n\to\infty} \pi_{S,n}^{(\ell)} = \alpha^{(\ell)},$$

i.e., (1) holds. ◀

Lemma 2.2 immediately implies the following.

▶ **Theorem 2.3** (divergence characterization of normality). *For all $\alpha \in \Delta(\Sigma)$ and $S \in \Sigma^\omega$, the following conditions are equivalent.*

(1) *S is α-normal.*
(2) $\mathrm{Div}(S||\alpha) = 0$.
(3) *For every α-normal sequence $T \in \Sigma^\omega$, $\mathrm{Div}(S||T) = 0$.*
(4) *There exists an α-normal sequence $T \in \Sigma^\omega$ such that $\mathrm{Div}(S||T) = 0$.*