Electronic Colloquium on Computational Complexity, Report No. 106 (2008)

A Divergence Formula for Randomness and Dimension

Jack H. Lutz*
Department of Computer Science
Iowa State University
Ames, IA 50011, USA
lutz@cs.iastate.edu

Abstract

If S is an infinite sequence over a finite alphabet Σ and β is a probability measure on Σ, then the *dimension* of S with respect to β, written $\dim^\beta(S)$, is a constructive version of Billingsley dimension that coincides with the (constructive Hausdorff) dimension $\dim(S)$ when β is the uniform probability measure. This paper shows that $\dim^\beta(S)$ and its dual $\mathrm{Dim}^\beta(S)$, the *strong dimension* of S with respect to β, can be used in conjunction with randomness to measure the similarity of two probability measures α and β on Σ. Specifically, we prove that the *divergence formula*

$$\dim^\beta(R) = \mathrm{Dim}^\beta(R) = \frac{\mathcal{H}(\alpha)}{\mathcal{H}(\alpha) + \mathcal{D}(\alpha||\beta)}$$

holds whenever α and β are computable, positive probability measures on Σ and $R \in \Sigma^\infty$ is random with respect to α. In this formula, $\mathcal{H}(\alpha)$ is the Shannon entropy of α, and $\mathcal{D}(\alpha||\beta)$ is the Kullback-Leibler divergence between α and β. We also show that the above formula holds for all sequences R that are α-normal (in the sense of Borel) when $\dim^\beta(R)$ and $\mathrm{Dim}^\beta(R)$ are replaced by the more effective finite-state dimensions $\dim^\beta_{\mathrm{FS}}(R)$ and ${\mathrm{Dim}_{\mathrm{FS}}}^\beta(R)$. In the course of proving this, we also prove finite-state compression characterizations of $\dim^\beta_{\mathrm{FS}}(S)$ and ${\mathrm{Dim}_{\mathrm{FS}}}^\beta(S)$.

1 Introduction

The constructive dimension $\dim(S)$ and the constructive strong dimension $\mathrm{Dim}(S)$ of an infinite sequence S over a finite alphabet Σ are constructive versions of the two most important classical fractal dimensions, namely, Hausdorff dimension [9] and packing dimension [22, 21], respectively. These two constructive dimensions, which were introduced in [13, 1], have been shown to have the useful characterizations

$$\dim(S) = \liminf_{w \to S} \frac{\mathrm{K}(w)}{|w| \log |\Sigma|} \tag{1.1}$$

and

$$\mathrm{Dim}(S) = \limsup_{w \to S} \frac{\mathrm{K}(w)}{|w| \log |\Sigma|}, \tag{1.2}$$

where the logarithm is base-2 [16, 1]. In these equations, $\mathrm{K}(w)$ is the Kolmogorov complexity of the prefix w of S, i.e., the *length in bits of the shortest program* that prints the string w. (See section

*This research was supported in part by National Science Foundation Grants 9988483, 0344187, 0652569, and 0728806 and by the Spanish Ministry of Education and Science (MEC) and the European Regional Development Fund (ERDF) under project TIN2005-08832-C03-02.

ISSN 1433-8092

2.6 or [11] for details.) The numerators in these equations are thus the *algorithmic information content* of w, while the denominators are the "naive" information content of w, also in bits. We thus understand (1.1) and (1.2) to say that $\dim(S)$ and $\mathrm{Dim}(S)$ are the lower and upper *information densities* of the sequence S. These constructive dimensions and their analogs at other levels of effectivity have been investigated extensively in recent years [10].

The constructive dimensions $\dim(S)$ and $\mathrm{Dim}(S)$ have recently been generalized to incorporate a probability measure ν on the sequence space Σ^∞ as a parameter [14]. Specifically, for each such ν and each sequence $S \in \Sigma^\infty$, we now have the constructive dimension $\dim^\nu(S)$ and the constructive strong dimension $\mathrm{Dim}^\nu(S)$ of S with respect to ν. (The first of these is a constructive version of Billingsley dimension [2].) When ν is the uniform probability measure on Σ^∞, we have $\dim^\nu(S) = \dim(S)$ and $\mathrm{Dim}^\nu(S) = \mathrm{Dim}(S)$. A more interesting example occurs when ν is the product measure generated by a nonuniform probability measure β on the alphabet Σ. In this case, $\dim^\nu(S)$ and $\mathrm{Dim}^\nu(S)$, which we write as $\dim^\beta(S)$ and $\mathrm{Dim}^\beta(S)$, are again the lower and upper information densities of S, but these densities are now measured with respect to unequal letter costs. Specifically, it was shown in [14] that

$$\dim^\beta(S) = \liminf_{w \to S} \frac{\mathrm{K}(w)}{\mathcal{I}_\beta(w)} \tag{1.3}$$

and

$$\mathrm{Dim}^\beta(S) = \limsup_{w \to S} \frac{\mathrm{K}(w)}{\mathcal{I}_\beta(w)}, \tag{1.4}$$

where

$$\mathcal{I}_\beta(w) = \sum_{i=0}^{|w|-1} \log \frac{1}{\beta(w[i])}$$

is the Shannon self-information of w with respect to β. These unequal letter costs $\log(1/\beta(a))$ for $a \in \Sigma$ can in fact be useful. For example, the complete analysis of the dimensions of individual points in self-similar fractals given by [14] requires these constructive dimensions with a particular choice of the probability measure β on Σ.

In this paper we show how to use the constructive dimensions $\dim^\beta(S)$ and $\mathrm{Dim}^\beta(S)$ in conjunction with randomness to measure the degree to which two probability measures on Σ are similar. To see why this might be possible, we note that the inequalities

$$0 \le \dim^\beta(S) \le \mathrm{Dim}^\beta(S) \le 1$$

hold for all β and S and that the maximum values

$$\dim^\beta(R) = \mathrm{Dim}^\beta(R) = 1 \tag{1.5}$$

are achieved whenever the sequence R is random with respect to β. It is thus reasonable to hope that, if R is random with respect to some other probability measure α on Σ, then $\dim^\beta(R)$ and $\mathrm{Dim}^\beta(R)$ will take on values whose closeness to 1 reflects the degree to which α is similar to β.

This is indeed the case. Our first main theorem says that the *divergence formula*

$$\dim^\beta(R) = \mathrm{Dim}^\beta(R) = \frac{\mathcal{H}(\alpha)}{\mathcal{H}(\alpha) + \mathcal{D}(\alpha||\beta)} \tag{1.6}$$

holds whenever α and β are computable, positive probability measures on Σ and $R \in \Sigma^\infty$ is random with respect to α. In this formula, $\mathcal{H}(\alpha)$ is the Shannon entropy of α, and $\mathcal{D}(\alpha||\beta)$ is the Kullback-Leibler divergence between α and β. When $\alpha = \beta$, the Kullback-Leibler divergence $\mathcal{D}(\alpha||\beta)$ is 0, so (1.6) coincides with (1.5). When α and β are dissimilar, the Kullback-Leibler divergence $\mathcal{D}(\alpha||\beta)$ is large, so the right-hand side of (1.6) is small. Hence the divergence formula tells us that, when R is α-random, $\dim^\beta(R) = \mathrm{Dim}^\beta(R)$ is a quantity in $[0, 1]$ whose closeness to 1 is an indicator of the similarity between α and β.

The proof of (1.6) serves as an outline of our other, more challenging task, which is to prove that the divergence formula (1.6) also holds for the much more effective *finite-state β-dimension* $\dim^\beta_{\mathrm{FS}}(R)$ and *finite-state strong β-dimension* $\mathrm{Dim}_{\mathrm{FS}}{}^\beta(R)$. (These dimensions, defined in section 2.5, are generalizations of finite-state dimension and finite-state strong dimension, which were introduced in [6, 1], respectively.)

With this objective in mind, our second main theorem characterizes the finite-state β-dimensions in terms of finite-state data compression. Specifically, this theorem says that, in analogy with (1.3) and (1.4), the identities

$$\dim^\beta_{\mathrm{FS}}(S) = \inf_C \liminf_{w \to S} \frac{|C(w)|}{\mathcal{I}_\beta(w)} \tag{1.7}$$

and

$$\dim^\beta_{\mathrm{FS}}(S) = \inf_C \limsup_{w \to S} \frac{|C(w)|}{\mathcal{I}_\beta(w)} \tag{1.8}$$

hold for all infinite sequences S over Σ. The infima here are taken over all information-lossless finite-state compressors (a model introduced by Shannon [20] and investigated extensively ever since) C with output alphabet $0, 1$, and $|C(w)|$ denotes the number of bits that C outputs when processing the prefix w of S. The special cases of (1.7) and (1.8) in which β is the uniform probability measure on Σ, and hence $\mathcal{I}_\beta(w) = |w| \log |\Sigma|$, were proven in [6, 1]. In fact, our proof uses these special cases as "black boxes" from which we derive the more general (1.7) and (1.8).

With (1.7) and (1.8) in hand, we prove our third main theorem. This involves the finite-state version of randomness, which was introduced by Borel [3] long before finite-state automata were defined. If α is a probability measure on Σ, then a sequence $S \in \Sigma^\infty$ is *α-normal* in the sense of Borel if every finite string $w \in \Sigma^*$ appears with asymptotic frequency $\alpha(w) \in S$, where we write

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

(See section 2.6 for a precise definition of asymptotic frequency.) Our third main theorem says that the *divergence formula*

$$\dim^\beta_{\mathrm{FS}}(R) = \mathrm{Dim}_{\mathrm{FS}}{}^\beta(R) = \frac{\mathcal{H}(\alpha)}{\mathcal{H}(\alpha) + \mathcal{D}(\alpha||\beta)} \tag{1.9}$$

holds whenever α and β are positive probability measures on Σ and $R \in \Sigma^\infty$ is α-normal.

In section 2 we briefly review ideas from Shannon information theory, classical fractal dimensions, algorithmic information theory, and effective fractal dimensions that are used in this paper. Section 3 outlines the proofs of (1.6), section 4 outlines the proofs of (1.7) and (1.8), and section 5 outlines the proof of (1.9). Various proofs are consigned to a technical appendix.

2 Preliminaries

2.1 Notation and setting

Throughout this paper we work in a finite alphabet $\Sigma = \{0, 1, \dots, k-1\}$, where $k \geq 2$. We write Σ^* for the set of (finite) *strings* over Σ and Σ^∞ for the set of (infinite) *sequences* over Σ. We write $|w|$ for the length of a string w and λ for the empty string. For $w \in \Sigma^*$ and $0 \leq i < |w|$, $w[i]$ is the ith symbol in w. Similarly, for $S \in \Sigma^\infty$ and $i \in \mathbb{N}$ $(= \{0, 1, 2, \dots\})$, $S[i]$ is the ith symbol in S. Note that the leftmost symbol in a string or sequence is the 0th symbol.

A *prefix* of a string or sequence $x \in \Sigma^* \cup \Sigma^\infty$ is a string $w \in \Sigma^*$ for which there exists a string or sequence $y \in \Sigma^* \cup \Sigma^\infty$ such that $x = wy$. In this case we write $w \sqsubseteq x$. The equation $\lim_{w \to S} f(w) = L$ means that, for all $\epsilon > 0$, for all sufficiently long prefixes $w \sqsubseteq S$, $|f(w) - L| < \epsilon$. We also use the limit inferior,

$$\liminf_{w \to S} f(w) = \lim_{w \to S} \inf \{ f(x) \mid w \sqsubseteq x \sqsubseteq S \},$$

and the limit superior

$$\limsup_{w \to S} f(w) = \lim_{w \to S} \sup \{ f(x) \mid w \sqsubseteq x \sqsubseteq S \}.$$

2.2 Probability measures, gales, and Shannon information

A *probability measure* on Σ is a function $\alpha : \Sigma \to [0, 1]$ such that $\sum_{a \in \Sigma} \alpha(a) = 1$. A probability measure α on Σ is *positive* if $\alpha(a) > 0$ for every $\alpha \in \Sigma$. A probability measure α on Σ is *rational* if $\alpha(a) \in \mathbb{Q}$ (i.e., $\alpha(a)$ is a rational number) for every $a \in \Sigma$.

A *probability measure* on Σ^∞ is a function $\nu : \Sigma^* \to [0, 1]$ such that $\nu(\lambda) = 1$ and, for all $w \in \Sigma^*$, $\nu(w) = \sum_{a \in \Sigma} \nu(wa)$. (Intuitively, $\nu(w)$ is the probability that $w \sqsubseteq S$ when the sequence $S \in \Sigma^\infty$ is "chosen according to ν.") Each probability measure α on Σ naturally induces the probability measure α on Σ^∞ defined by

$$\alpha(w) = \prod_{i=0}^{|w|-1} \alpha(w[i]) \tag{2.1}$$

for all $w \in \Sigma^*$.

We reserve the symbol μ for the *uniform probability measure* on Σ, i.e.,

$$\mu(a) = \frac{1}{k} \text{ for all } a \in \Sigma,$$

and also for the *uniform probability measure* on Σ^∞, i.e.,

$$\mu(w) = k^{-|w|} \text{ for all } w \in \Sigma^*.$$

If α is a probability measure on Σ and $s \in [0, \infty)$, then an *s-α-gale* is a function $d : \Sigma^* \to [0, \infty)$ satisfying

$$d(w) = \sum_{a \in \Sigma} d(wa) \alpha(a)^s \tag{2.2}$$

for all $w \in \Sigma^*$. A 1-α-gale is also called an *α-martingale*. When $\alpha = \mu$, we omit it from this terminology, so an s-μ-gale is called an *s-gale*, and a μ-martingale is called a *martingale*.

We frequently use the following simple fact without explicit citation.

Observation 2.1. *Let α and β be positive probability measures on Σ, and let $s,t \in [0,\infty)$. If $d : \Sigma^* \to [0,\infty)$ is an s-α-gale, then the function $\tilde{d} : \Sigma^* \to [0,\infty)$ defined by*

$$\tilde{d}(w) = \frac{\alpha(w)^s}{\beta(w)^t} d(w)$$

is a t-β-gale.

Intuitively, an s-α-gale is a strategy for betting on the successive symbols in a sequence $S \in \Sigma^\infty$. For each prefix $w \sqsubseteq S$, $d(w)$ denotes the amount of capital (money) that the gale d has after betting on the symbols in w. If $s = 1$, then the right-hand side of (2.2) is the conditional expectation of $d(wa)$, given that w has occurred, so (2.2) says that the payoffs are fair. If $s < 1$, then (2.2) says that the payoffs are unfair.

Let d be a gale, and let $S \in \Sigma^\infty$. Then d *succeeds* on S if $\limsup_{w\to S} d(w) = \infty$, and d *succeeds strongly* on S if $\liminf_{w\to S} d(w) = \infty$. The *success set* of d is the set $S^\infty[d]$ of all sequences on which d succeeds, and the *strong success set* of d is the set $S^\infty_{\mathrm{str}}[d]$ of all sequences on which d succeeds strongly.

The *Shannon entropy* of a probability measure α on Σ is

$$\mathcal{H}(\alpha) = \sum_{a\in\Sigma} \alpha(a) \log \frac{1}{\alpha(a)},$$

where $0 \log \frac{1}{0} = 0$. (unless otherwise indicated, all logarithms in this paper are base-2.) The *Kullback-Leibler divergence* between two probability measures α and β on Σ is

$$\mathcal{D}(\alpha||\beta) = \sum_{a\in\Sigma} \alpha(a) \log \frac{\alpha(a)}{\beta(a)}.$$

The Kullback-Leibler divergence is used to quantify how "far apart" the two probability measures α and β are. The *Shannon self-information* of a string $w \in \Sigma^*$ with respect to a probability measure β on Σ is

$$\mathcal{I}_\beta(w) = \log \frac{1}{\beta(w)} = \sum_{i=0}^{|w|-1} \log \frac{1}{\beta(w[i])}.$$

Discussions of $\mathcal{H}(\alpha)$, $\mathcal{D}(\alpha||\beta)$, $\mathcal{I}_\beta(w)$ and their properties may be found in any good text on information theory, e.g., [5].

2.3 Hausdorff, packing, and Billingsley dimensions

Given a probability measure β on Σ, each set $X \subseteq \Sigma^\infty$ has a *Hausdorff dimension* $\dim(X)$, a *packing dimension* $\mathrm{Dim}(X)$, a *Billingsley dimension* $\dim^\beta(X)$, and a *strong Billingsley dimension* $\mathrm{Dim}^\beta(X)$, all of which are real numbers in the interval $[0,1]$. In this paper we are not concerned with the original definitions of these classical dimensions, but rather in their recent characterizations (which may be taken as definitions) in terms of gales.

Notation. For each probability measure β on Σ and each set $X \subseteq \Sigma^\infty$, let $\mathcal{G}^\beta(X)$ (respectively, $\mathcal{G}^{\beta,\mathrm{str}}(X)$) be the set of all $s \in [0,\infty)$ such that there is a β-s-gale d satisfying $X \subseteq S^\infty[d]$ (respectively, $X \subseteq S^\infty_{\mathrm{str}}[d]$).

Theorem 2.2 (gale characterizations of classical fractal dimensions)**.** *Let β be a probability measure on Σ, and let $X \subseteq \Sigma^\infty$.*

1. [12] $\dim(X) = \inf \mathcal{G}^{\mu}(X)$.

2. [1] $\mathrm{Dim}(X) = \inf \mathcal{G}^{\mu,\mathrm{str}}(X)$.

3. [14] $\dim^{\beta}(X) = \inf \mathcal{G}^{\beta}(X)$.

4. [14] $\mathrm{Dim}^{\beta}(X) = \inf \mathcal{G}^{\beta,\mathrm{str}}(X)$.

2.4 Randomness and constructive dimensions

Randomness and constructive dimensions are defined by imposing computability constraints on gales.

A real-valued function $f : \Sigma^* \to \mathbb{R}$ is *computable* if there is a computable, rational-valued function $\hat{f} : \Sigma^* \times \mathbb{N} \to \mathbb{Q}$ such that, for all $w \in \Sigma^*$ and $r \in \mathbb{N}$,

$$|\hat{f}(w,r) - f(w)| \leq 2^{-r}.$$

A real-valued function $f : \Sigma^* \to \mathbb{R}$ is *constructive*, or *lower semicomputable*, if there is a computable, rational-valued function $\hat{f} : \Sigma^* \times \mathbb{N} \to \mathbb{Q}$ such that

(i) for all $w \in \Sigma^*$ and $t \in \mathbb{N}$, $\hat{f}(w,t) \leq \hat{f}(w,t+1) < f(w)$, and

(ii) for all $w \in \Sigma^*$, $f(w) = \lim_{t\to\infty} \hat{f}(w,t)$.

The first successful definition of the randomness of individual sequences $S \in \Sigma^{\infty}$ was formulated by Martin-Löf [15]. Many characterizations (equivalent definitions) of randomness are now known, of which the following is the most pertinent.

Theorem 2.3 (Schnorr [17, 18])**.** *Let α be a probability measure on Σ. A sequence $S \in \Sigma^{\infty}$ is* random *with respect to α (or, briefly, α-*random*) if there is no constructive α-martingale that succeeds on S.*

Motivated by Theorem 2.2, we now define the constructive dimensions.

Notation. We define the sets $\mathcal{G}^{\beta}_{\mathrm{constr}}(X)$ and $\mathcal{G}^{\beta,\mathrm{str}}_{\mathrm{constr}}(X)$ to be like the sets $\mathcal{G}^{\beta}(X)$ and $\mathcal{G}^{\beta,\mathrm{constr}}(X)$ of section 2.3, except that the β-s-gales are now required to be constructive.

Definition. Let β be a probability measure on Σ, let $X \subseteq \Sigma^{\infty}$, and let $S \in \Sigma^{\infty}$.

1. [13] The *constructive dimension* of X is $\mathrm{cdim}(X) = \inf \mathcal{G}^{\mu}_{\mathrm{constr}}(X)$.
2. [1] The *constructive strong dimension* of X is $\mathrm{cDim}(X) = \inf \mathcal{G}^{\mu,\mathrm{str}}_{\mathrm{constr}}(X)$.
3. [14] The *constructive β-dimension* of X is $\mathrm{cdim}^{\beta}(X) = \inf \mathcal{G}^{\beta}_{\mathrm{constr}}(X)$.
4. [14] The *constructive strong β-dimension* of X is $\mathrm{cDim}^{\beta}(X) = \inf \mathcal{G}^{\beta,\mathrm{str}}_{\mathrm{constr}}(X)$.
5. [13] The *dimension* of S is $\dim(S) = \mathrm{cdim}(\{S\})$.
6. [1] The *strong dimension* of S is $\mathrm{Dim}(S) = \mathrm{cDim}(\{S\})$.
7. [14] The *β-dimension* of S is $\dim^{\beta}(S) = \mathrm{cdim}^{\beta}(\{S\})$.
8. [14] The *strong β-dimension* of S is $\mathrm{Dim}^{\beta}(S) = \mathrm{cDim}^{\beta}(\{S\})$.

It is clear that definitions 1, 2, 5, and 6 above are the special case $\beta = \mu$ of definitions 3, 4, 7, and 8, respectively. It is known that $\mathrm{cdim}^{\beta}(X) = \sup_{S\in X} \dim^{\beta}(S)$ and that $\mathrm{cDim}^{\beta}(X) = \sup_{S\in X} \mathrm{Dim}^{\beta}(S)$ [14]. Constructive dimensions are thus investigated in terms of the dimensions of individual sequences. Since one does not discuss the classical dimension of an individual sequence (because the dimensions of section 2.3 are all zero for singleton, or even countable, sets), no confusion results from the notation $\dim(S)$, $\mathrm{Dim}(S)$, $\dim^{\beta}(S)$, and $\mathrm{Dim}^{\beta}(S)$.

2.5 Normality and finite-state dimensions

The preceding section developed the constructive dimensions as effective versions of the classical dimensions of section 2.3. We now introduce the even more effective finite-state dimensions.

Notation. $\Delta_{\mathbb{Q}}(\Sigma)$ is the set of all rational-valued probability measure on Σ.

Definition ([19, 8, 6])**.** A *finite-state gambler* (*FSG*) is a 4-tuple

$$G = (Q, \delta, q_0, B),$$

where Q is a finite set of *states*, $\delta : Q \times \Sigma \to Q$ is the *transition function*; $q_0 \in Q$ is the *initial state*, and $B : Q \to \Delta_{\mathbb{Q}}(\Sigma)$ is the *betting function*.

The transition structure (Q, δ, q_0) here works as in any deterministic finite-state automaton. For $w \in \Sigma^*$, we write $\delta(w)$ for the state reached by starting at q_0 and processing w according to δ.

Intuitively, if the above FSG is in state $q \in Q$, then, for each $a \in \Sigma$, it bets the fraction $B(q)(a)$ of its current capital that the next input symbol is an a. The payoffs are determined as follows.

Definition. Let $G = (Q, \delta, q_0, B)$ be an FSG.

1. The *martingale* of G is the function $d_G : \Sigma^* \to [0, \infty)$ defined by the recursion

$$d_G(\lambda) = 1,$$

$$d_G(wa) = kd_G(w)B(\delta(w))(a)$$

for all $w \in \Sigma^*$ and $a \in \Sigma$.

2. If β is a probability measure on Σ and $s \in [0, \infty)$, then the *s-β-gale* of G is the function $d^{(s)}_{G,\beta} : \Sigma^* \to [0, \infty)$ defined by

$$d^{(s)}_{G,\beta}(w) = \frac{\mu(w)}{\beta(w)^s} d_G(w)$$

for all $w \in \Sigma^*$.

It is easy to verify that $d_G = d^{(1)}_{G,\mu}$ is a martingale. It follows by Observation 2.1 that $d^{(1)}_{G,\beta}$ is an s-β-gale.

Definition. A *finite-state s-β-gale* is an s-β-gale of the form $d^{(s)}_{G,\beta}$ for some FSG G.

Notation. We define the sets $\mathcal{G}^{\beta}_{\mathrm{FS}}(X)$ and $\mathcal{G}^{\beta,\mathrm{str}}_{\mathrm{FS}}(X)$ to be like the sets $\mathcal{G}^{\beta}(X)$ and $\mathcal{G}^{\beta,\mathrm{str}}(X)$ of section 2.3, except that the s-β-gales are now required to be finite-state.

Definition. Let β be a probability measure on Σ, and let $S \in \Sigma^\infty$.

1. [6] The *finite-state dimension* of S is $\dim_{\mathrm{FS}}(S) = \inf \mathcal{G}^{\mu}_{\mathrm{FS}}(\{S\})$.

2. [1] The *finite-state strong dimension* of S is $\mathrm{Dim}_{\mathrm{FS}}(S) = \inf \mathcal{G}^{\mu,\mathrm{str}}_{\mathrm{FS}}(\{S\})$.

3. The *finite-state β-dimension* of S is $\dim^{\beta}_{\mathrm{FS}}(S) = \inf \mathcal{G}^{\beta}_{\mathrm{FS}}(\{S\})$.

4. The *finite-state strong β-dimension* of S is ${\mathrm{Dim}_{\mathrm{FS}}}^{\beta}(S) = \inf \mathcal{G}^{\beta,\mathrm{str}}_{\mathrm{FS}}(\{S\})$.

We now turn to some ideas based on asymptotic frequencies of strings in a given sequence. For nonempty strings $w, x \in \Sigma^*$, we write

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

for the number of block occurrences of w in x. For each sequence $S \in \Sigma^\infty$, each positive integer n, and each nonempty $w \in \Sigma^{<n}$, the nth *block frequency* of w in S is

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

Note that, for each n and l, the restriction $\pi^{(l)}_{S,n}$ of $\pi_{S,n}$ to Σ^l is a probability measure on Σ^l.

Definition. Let α be a probability measure on Σ, let $S \in \Sigma^\infty$, and let $0 < l \in \mathbb{N}$.

1. S is *α-l-normal* in the sense of Borel if, for all $w \in \Sigma^l$, $\lim\limits_{n\to\infty} \pi_{S,n}(w) = \alpha(w)$.
2. S is *α-normal* in the sense of Borel if S is α-l-normal for all $0 < l \in \mathbb{N}$.
3. [3] S is *normal* in the sense of Borel if S is μ-normal.
4. S has *asymptotic frequency* α, and we write $S \in \mathrm{FREQ}^\alpha$, if S is α-1-normal.

Theorem 2.4 ([19, 4])**.** *For each probability measure α on Σ and each $S \in \Sigma^\infty$, the following three conditions are equivalent.*

(1) S is α-normal.

(2) No finite-state α-martingale succeeds on S.

(3) $\dim^\alpha_{\mathrm{FS}}(S) = 1$.

The equivalence of (1) and (2) where $\alpha = \mu$ was proven in [19]. The equivalence of (2) and (3) when $\alpha = \mu$ was noted in [4]. The extensions of these facts to arbitrary α is routine.

For each $S \in \Sigma^\infty$ and $0 < l \in \mathbb{N}$, the lth *normalized lower and upper block entropy* rates of S are

$$\mathrm{H}^-_l(S) = \frac{1}{l \log k} \liminf_{n\to\infty} \mathcal{H}(\pi^{(l)}_{S,n})$$

and

$$\mathrm{H}^+_l(S) = \frac{1}{l \log k} \limsup_{n\to\infty} \mathcal{H}(\pi^{(l)}_{S,n}),$$

respectively.

We use the following result in section 5.

Theorem 2.5 ([4])**.** *Let $S \in \Sigma^\infty$.*

1. $\dim_{\mathrm{FS}}(S) = \inf_{0<l\in\mathbb{N}} \mathrm{H}^-_l(S)$.

2. $\mathrm{Dim}_{\mathrm{FS}}(S) = \inf_{0<l\in\mathbb{N}} \mathrm{H}^+_l(S)$.

2.6 Kolmogorov complexity and finite-state compression

We now review known characterizations of constructive and finite-state dimensions that are based on data compression ideas.

The *Kolmogorov complexity* $\mathrm{K}(w)$ of a string $w \in \Sigma^*$ is the minimum length of a program $\pi \in \{0,1\}^*$ for which $U(\pi) = w$, where U is a fixed universal self-delimiting Turing machine [11].

Theorem 2.6. *Let β be a probability measure on Σ, and let $S \in \Sigma^\infty$.*

1. *[16]* $\dim(S) = \liminf_{w \to S} \frac{\mathrm{K}(w)}{|w| \log k}$.
2. *[1]* $\mathrm{Dim}(S) = \limsup_{w \to S} \frac{\mathrm{K}(w)}{|w| \log k}$.
3. *[14]* $\dim^\beta(S) = \liminf_{w \to S} \frac{\mathrm{K}(w)}{\mathcal{I}_\beta(w)}$.
4. *[14]* $\mathrm{Dim}^\beta(S) = \limsup_{w \to S} \frac{\mathrm{K}(w)}{\mathcal{I}_\beta(w)}$.

Definition ([20])**.**

1. A *finite-state compressor* (FSC) is a 4-tuple

$$C = (Q, \delta, q_0, \nu),$$

where Q, δ, and q_0 are as in the FSG definition, and $\nu : Q \times \Sigma \to \{0,1\}^*$ is the output function.

2. The *output* of an FSC $C = (Q, \delta, q_0, \nu)$ on an input $w \in \Sigma^*$ is the string $C(w) \in \{0,1\}^*$ defined by the recursion

$$C(\lambda) = \lambda,$$
$$C(wa) = C(w)\nu(\delta(w), a)$$

for all $w \in \Sigma^*$ and $a \in \Sigma$.

3. An *information-lossless finite-state compressor* (ILFSC) is an FSC for which the function

$$\Sigma^* \to \{0,1\}^* \times Q$$
$$w \mapsto (C(w), \delta(w))$$

is one-to-one.

Theorem 2.7. *Let $S \in \Sigma^\infty$.*

1. *[6]* $\dim_{\mathrm{FS}}(S) = \inf_C \liminf_{w \to S} \frac{|C(w)|}{|w| \log k}$.
2. *[1]* $\mathrm{Dim}_{\mathrm{FS}}(S) = \inf_C \limsup_{w \to S} \frac{|C(w)|}{|w| \log k}$.

3 Divergence formula for randomness and constructive dimensions

This section proves the divergence formula for α-randomness, constructive β-dimension, and constructive strong β-dimension. The key point here is that the Kolmogorov complexity characterizations of these β-dimensions reviewed in section 2.6 immediately imply the following fact.

Lemma 3.1. *If α and β are computable, positive probability measure on Σ, then, for all $S \in \Sigma^\infty$,*

$$\liminf_{w \to S} \frac{\mathcal{I}_\alpha(w)}{\mathcal{I}_\beta(w)} \leq \frac{\dim^\beta(S)}{\dim^\alpha(S)} \leq \limsup_{w \to S} \frac{\mathcal{I}_\alpha(w)}{\mathcal{I}_\beta(w)},$$

and

$$\liminf_{w \to S} \frac{\mathcal{I}_\alpha(w)}{\mathcal{I}_\beta(w)} \leq \frac{\mathrm{Dim}^\beta(S)}{\mathrm{Dim}^\alpha(S)} \leq \limsup_{w \to S} \frac{\mathcal{I}_\alpha(w)}{\mathcal{I}_\beta(w)}.$$

The following lemma is crucial to our argument, both here and in section 5.

Lemma 3.2 (frequency divergence lemma)**.** *If α and β are positive probability measures on Σ, then, for all $S \in \mathrm{FREQ}^\alpha$,*

$$\mathcal{I}_\beta(w) = (\mathcal{H}(\alpha) + \mathcal{D}(\alpha||\beta))|w| + o(|w|)$$

as $w \to S$.

The next lemma gives a simple relationship between the constructive β-dimension and the constructive dimension of any sequence that is α-1-normal.

Lemma 3.3. *If α and β are computable, positive probability measures on Σ, then, for all $S \in \mathrm{FREQ}^\alpha$,*

$$\dim^\beta(S) = \frac{\dim(S)}{\mathcal{H}_k(\alpha) + \mathcal{D}_k(\alpha||\beta)},$$

and

$$\mathrm{Dim}^\beta(S) = \frac{\mathrm{Dim}(S)}{\mathcal{H}_k(\alpha) + \mathcal{D}_k(\alpha||\beta)}.$$

We now recall the following constructive strengthening of a 1949 theorem of Eggleston [7].

Theorem 3.4 ([13, 1])**.** *If α is a computable probability measure on Σ, then, for every α-random sequence $R \in \Sigma^\infty$,*

$$\dim(R) = \mathrm{Dim}(R) = \mathcal{H}_k(\alpha).$$

The main result of this section is now clear.

Theorem 3.5 (divergence formula for randomness and constructive dimensions)**.** *If α and β are computable, positive probability measures on Σ, then, for every α-random sequence $R \in \Sigma^\infty$,*

$$\dim^\beta(R) = \mathrm{Dim}^\beta(R) = \frac{\mathcal{H}(\alpha)}{\mathcal{H}(\alpha) + \mathcal{D}(\alpha||\beta)}.$$

Proof. This follows immediately from Lemma 3.3 and Theorem 3.4. □

We note that $\mathcal{D}(\alpha||\mu) = \log k - \mathcal{H}(\alpha)$, so Theorem 3.4 is the case $\beta = \mu$ of Theorem 3.5.

4 Finite-state dimensions and data compression

This section proves finite-state compression characterizations of finite-state β-dimension and finite-state strong β-dimension that are analogous to the characterizations given by parts 3 and 4 of Theorem 2.6. Our argument uses the following two technical lemmas, which are proven in the technical appendix.

Lemma 4.1. *Let β be a positive probability measure on Σ, and let C be an ILFSC. Assume that $I \subseteq \Sigma^*$, $s > 0$, and $\epsilon > 0$ have the property that, for all $w \in I$,*

$$s \geq \frac{|C(w)|}{\mathcal{I}_\beta(w)} + \epsilon. \tag{4.1}$$

Then there exist an FSG G and a real number $\delta > 0$ such that, for all sufficiently long strings $w \in I$,

$$d_{G,\beta}^{(s)}(w) \geq 2^{\delta|w|}. \tag{4.2}$$

Lemma 4.2. *Let β be a positive probability measure on Σ, and let G be an FSG. Assume that $I \subseteq \Sigma^*$, $s > 0$, and $\epsilon > 0$ have the property that, for all $w \in I$,*

$$d_{G,\beta}^{(s-2\epsilon)}(w) \geq 1. \tag{4.3}$$

Then there is an ILFSC C such that, for all $w \in I$,

$$|C(w)| \leq s\mathcal{I}_\beta(w). \tag{4.4}$$

We now prove the main result of this section.

Theorem 4.3 (compression characterizations of finite-state β-dimensions)**.** *If β is a positive probability measure on Σ, then, for each sequence $S \in \Sigma^\infty$,*

$$\dim_{\mathrm{FS}}^{\beta}(S) = \inf_{C} \liminf_{w \to S} \frac{|C(w)|}{\mathcal{I}_\beta(w)}, \tag{4.5}$$

and

$$\mathrm{Dim_{FS}}^{\beta}(S) = \inf_{C} \limsup_{w \to S} \frac{|C(w)|}{\mathcal{I}_\beta(w)}, \tag{4.6}$$

where the infima are taken over all ILFCSs C.

Proof. Let β and S be as given. We first prove that the left-hand sides of (4.5) and (4.6) do not exceed the right-hand sides. For this, let C be an ILFSC. It suffices to show that

$$\dim_{\mathrm{FS}}^{\beta}(S) \leq \liminf_{w \to S} \frac{|C(w)|}{\mathcal{I}_\beta(w)} \tag{4.7}$$

and

$$\mathrm{Dim_{FS}}^{\beta}(S) \leq \limsup_{w \to S} \frac{|C(w)|}{\mathcal{I}_\beta(w)}. \tag{4.8}$$

To see that (4.7) holds, let s exceed the right-hand side. Then there exist an infinite set I of prefixes of S and an $\epsilon > 0$ such that (4.1) holds for all $w \in I$. It follows by Lemma 4.1 that there exist an FSG G and a $\delta > 0$ such that, for all sufficiently long $w \in I$, $d^{(s)}_{G,\beta}(w) \geq 2^{\delta|w|}$. Since I is infinite and $\delta > 0$, this implies that $S \in S^\infty[d^{(s)}_{G,\beta}]$, whence $\dim^\beta_{\mathrm{FS}}(S) \leq s$. This establishes (4.7).

The proof that (4.8) holds is identical to the preceding paragraph, except that I is now a cofinite set of prefixes of S, so $S \in S^\infty_{\mathrm{str}}[d^{(s)}_{G,\beta}]$.

It remains to be shown that the right-hand sides of (4.5) and (4.6) do not exceed the left-hand sides. To see this for (4.5), let $s > \dim^\beta_{\mathrm{FS}}(S)$. It suffices to show that there is an ILFSC C such that

$$\liminf_{w \to S} \frac{|C(w)|}{\mathcal{I}_\beta(w)} \leq s. \tag{4.9}$$

By our choice of s there exists $\epsilon > 0$ such that $s - 2\epsilon > \dim^\beta_{\mathrm{FS}}(S)$. This implies that there is an infinite set I of prefixes of S such that (4.3) holds for all $w \in I$. Choose C for G, I, S, and ϵ as in Lemma 4.2. Then

$$\liminf_{w \to S} \frac{|C(w)|}{\mathcal{I}_\beta(w)} \leq \inf_{w \in I} \frac{|C(w)|}{\mathcal{I}_\beta(w)} \leq s \tag{4.10}$$

by (4.4), so (4.9) holds.

The proof that the right-hand side of (4.6) does not exceed the left-hand side is identical to the preceding paragraph, except that the limits inferior in (4.9) and (4.10) are now limits superior, and the set I is now a cofinite set of prefixes of S. □

5 Divergence formula for normality and finite-state dimensions

This section proves the divergence formula for α-normality, finite-state β-dimension, and finite-state strong β-dimension. As should now be clear, Theorem 4.3 enables us to proceed in analogy with section 3.

Lemma 5.1. *If α and β are positive probability measures on Σ, then, for all $S \in \Sigma^\infty$,*

$$\liminf_{w \to S} \frac{\mathcal{I}_\alpha(w)}{\mathcal{I}_\beta(w)} \leq \frac{\dim^\beta_{\mathrm{FS}}(S)}{\dim^\alpha_{\mathrm{FS}}(S)} \leq \limsup_{w \to S} \frac{\mathcal{I}_\alpha(w)}{\mathcal{I}_\beta(w)}, \tag{5.1}$$

and

$$\liminf_{w \to S} \frac{\mathcal{I}_\alpha(w)}{\mathcal{I}_\beta(w)} \leq \frac{{\mathrm{Dim}_{\mathrm{FS}}}^\beta(S)}{{\mathrm{Dim}_{\mathrm{FS}}}^\alpha(S)} \leq \limsup_{w \to S} \frac{\mathcal{I}_\alpha(w)}{\mathcal{I}_\beta(w)}. \tag{5.2}$$

Lemma 5.2. *If α and β are positive probability measures on Σ, then, for all $S \in \mathrm{FREQ}^\alpha$,*

$$\dim^\beta_{\mathrm{FS}}(S) = \frac{\dim_{\mathrm{FS}}(S)}{\mathcal{H}_k(\alpha) + \mathcal{D}_k(\alpha||\beta)},$$

and

$${\mathrm{Dim}_{\mathrm{FS}}}^\beta(S) = \frac{\mathrm{Dim}_{\mathrm{FS}}(S)}{\mathcal{H}_k(\alpha) + \mathcal{D}_k(\alpha||\beta)}.$$

We next prove a finite-state analog of Theorem 3.4.

Theorem 5.3. *If α is a probability measure on Σ, then, for every α-normal sequence $R \in \Sigma^\infty$,*

$$\dim_{\mathrm{FS}}(R) = \mathrm{Dim}_{\mathrm{FS}}(R) = \mathcal{H}_k(\alpha).$$

We now have our third main theorem.

Theorem 5.4 (divergence theorem for normality and finite-state dimensions)**.** *If α and β are positive probability measures on Σ, then, for every α-normal sequence $R \in \Sigma^\infty$,*

$$\dim_{\mathrm{FS}}^{\beta}(R) = \mathrm{Dim}_{\mathrm{FS}}{}^{\beta}(R) = \frac{\mathcal{H}(\alpha)}{\mathcal{H}(\alpha) + \mathcal{D}(\alpha||\beta)}.$$

Proof. This follows immediately from Lemma 5.2 and Theorem 5.3. □

We again note that $\mathcal{D}(\alpha||\beta) = \log k - \mathcal{H}(\alpha)$, so Theorem 5.3 is the case $\beta = \mu$ of Theorem 5.4.

Acknowledgments. I thank Xiaoyang Gu and Elvira Mayordomo for useful discussions.

References

[1] K. B. Athreya, J. M. Hitchcock, J. H. Lutz, and E. Mayordomo. Effective strong dimension, algorithmic information, and computational complexity. *SIAM Journal on Computing*, 37:671–705, 2007.

[2] P. Billingsley. Hausdorff dimension in probability theory. *Illinois Journal of Mathematics*, 4:187–209, 1960.

[3] E. Borel. Sur les probabilités dénombrables et leurs applications arithmétiques. *Rend. Circ. Mat. Palermo*, 27:247–271, 1909.

[4] C. Bourke, J. M. Hitchcock, and N. V. Vinodchandran. Entropy rates and finite-state dimension. *Theoretical Computer Science*, 349(3):392–406, 2005.

[5] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* John Wiley & Sons, Inc., second edition, 2006.

[6] J. J. Dai, J. I. Lathrop, J. H. Lutz, and E. Mayordomo. Finite-state dimension. *Theoretical Computer Science*, 310:1–33, 2004.

[7] H. Eggleston. The fractional dimension of a set defined by decimal properties. *Quarterly Journal of Mathematics*, Oxford Series 20:31–36, 1949.

[8] M. Feder. Gambling using a finite state machine. *IEEE Transactions on Information Theory*, 37:1459–1461, 1991.

[9] F. Hausdorff. Dimension und äusseres Mass. *Mathematische Annalen*, 79:157–179, 1919. English translation.

[10] J. M. Hitchcock. Effective Fractal Dimension Bibliography, http://www.cs.uwyo.edu/ ~jhitchco/bib/dim.shtml (current October, 2008).

[11] M. Li and P. M. B. Vitányi. *An Introduction to Kolmogorov Complexity and its Applications.* Springer-Verlag, Berlin, 1997. Second Edition.

[12] J. H. Lutz. Dimension in complexity classes. *SIAM Journal on Computing*, 32:1236–1259, 2003.

[13] J. H. Lutz. The dimensions of individual strings and sequences. *Information and Computation*, 187:49–79, 2003.

[14] J. H. Lutz and E. Mayordomo. Dimensions of points in self-similar fractals. *SIAM Journal on Computing*, 38:1080–1112, 2008.

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

[16] E. Mayordomo. A Kolmogorov complexity characterization of constructive Hausdorff dimension. *Information Processing Letters*, 84(1):1–3, 2002.

[17] C. P. Schnorr. A unified approach to the definition of random sequences. *Mathematical Systems Theory*, 5:246–258, 1971.

[18] C. P. Schnorr. A survey of the theory of random sequences. In R. E. Butts and J. Hintikka, editors, *Basic Problems in Methodology and Linguistics*, pages 193–210. D. Reidel, 1977.

[19] C. P. Schnorr and H. Stimm. Endliche Automaten und Zufallsfolgen. *Acta Informatica*, 1:345–359, 1972.

[20] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948.

[21] D. Sullivan. Entropy, Hausdorff measures old and new, and limit sets of geometrically finite Kleinian groups. *Acta Mathematica*, 153:259–277, 1984.

[22] C. Tricot. Two definitions of fractional dimension. *Mathematical Proceedings of the Cambridge Philosophical Society*, 91:57–74, 1982.

A Appendix – Various Proofs

Proof of Lemma 3.2. Assume the hypothesis, and let $S \in \mathrm{FREQ}^\alpha$. Then, as $w \to S$, we have

$$\begin{aligned}
\mathcal{I}_\beta(w) &= \sum_{i=0}^{|w|-1} \log \frac{1}{\beta(w[i])} \\
&= \sum_{a\in\Sigma} \#(a,w) \log \frac{1}{\beta(a)} \\
&= |w| \sum_{a\in\Sigma} \mathrm{freq}_a(w) \log \frac{1}{\beta(a)} \\
&= |w| \sum_{a\in\Sigma} (\alpha(a) + o(1)) \log \frac{1}{\beta(a)} \\
&= |w| \sum_{a\in\Sigma} \alpha(a) \log \frac{1}{\beta(a)} + o(|w|) \\
&= |w| \sum_{a\in\Sigma} \left(\alpha(a) \log \frac{1}{\alpha(a)} + \alpha(a) \log \frac{\alpha(a)}{\beta(a)} \right) + o(|w|) \\
&= (\mathcal{H}(\alpha) + \mathcal{D}(\alpha||\beta))|w| + o(|w|).
\end{aligned}$$

□

Proof of Lemma 3.3. Let α, β, and S be as given. By the frequency divergence lemma, we have

$$\begin{aligned}
\frac{\mathcal{I}_\mu(w)}{\mathcal{I}_\beta(w)} &= \frac{|w| \log k}{(\mathcal{H}(\alpha) + \mathcal{D}(\alpha||\beta))|w| + o(|w|)} \\
&= \frac{\log k}{\mathcal{H}(\alpha) + \mathcal{D}(\alpha||\beta) + o(1)} \\
&= \frac{\log k}{\mathcal{H}(\alpha) + \mathcal{D}(\alpha||\beta)} + o(1) \\
&= \frac{1}{\mathcal{H}_k(\alpha) + \mathcal{D}_k(\alpha||\beta)} + o(1)
\end{aligned}$$

as $w \to S$. The present lemma follows from this and Lemma 3.1. □

The following lemma summarizes the first part of the proof of Theorem 2.7.

Lemma A.1 ([6]). *For each ILFSC C there is an integer $m \in \mathbb{Z}^+$ such that, for each $l \in \mathbb{Z}^+$, there is an FSG G such that, for all $w \in \Sigma^*$,*

$$\log d_G^{(1)}(w) \geq |w| \log k - |C(w)| - m(\tfrac{|w|}{l} + l). \tag{A.1}$$

Proof of Lemma 4.1. Assume the hypothesis. Let

$$\delta_\beta = \min_{a\in\Sigma} \log \frac{1}{\beta(a)},$$

noting the following two things.

(i) $\delta_\beta > 0$, because β is positive.

(ii) For all $w \in \Sigma^*$,

$$\mathcal{I}_\beta(w) \geq \delta_\beta |w|. \tag{A.2}$$

Choose m for C as in Lemma A.1, let

$$l = \left\lceil \frac{3m}{\epsilon\delta\beta} \right\rceil, \tag{A.3}$$

and choose G for C, m, and l as in Lemma 4.1. Let

$$\delta = \tfrac{2}{3}\epsilon\delta_\beta,$$

noting that $\delta > 0$ and that

$$\begin{aligned} |w| \geq l^2 \implies & \epsilon\delta_\beta|w| - m(\tfrac{|w|}{l} + l) \\ & = \epsilon\delta_\beta|w| - \tfrac{m}{l}(|w| + l^2) \\ & \geq \epsilon\delta_\beta|w| - \tfrac{2m}{l}|w| \\ & = (\epsilon\delta_\beta - \tfrac{2m}{l})|w| \\ & \geq^{(\mathrm{A.3})} \tfrac{2}{3}\epsilon\delta_\beta|w|, \end{aligned}$$

i.e., that

$$|w| \geq l^2 \implies \epsilon\delta_\beta|w| - m(\tfrac{|w|}{l} + l) \geq \delta|w|. \tag{A.4}$$

It follows that, for all $w \in I$ with $|w| \geq l^2$, we have

$$\begin{aligned} \log d_{G,\beta}^{(s)}(w) &= \log(\tfrac{\mu(w)}{\beta(w)^s} d_G^{(1)}(w)) \\ &= -|w|\log k + s\mathcal{I}_\beta(w) + \log d_G^{(1)}(w) \\ &\geq^{(\mathrm{A.1})} s\mathcal{I}_\beta(w) - |C(w)| - m(\tfrac{|w|}{l} + l) \\ &\geq^{(4.1)} s\mathcal{I}_\beta(w) - m(\tfrac{|w|}{l} + l) \\ &\geq^{(\mathrm{A.2})} \epsilon\delta_\beta|w| - m(\tfrac{|w|}{l} + l) \\ &\geq^{(\mathrm{A.4})} \delta|w|. \end{aligned}$$

Hence (4.2) holds. □

An FSG $G = (Q, \Sigma, \delta, \beta, q_0)$ is *nonvanishing* if all its bets are nonzero, i.e., if $\beta(q)(a) > 0$ holds for all $q \in Q$ and $a \in \Sigma$.

Lemma A.2 ([6])**.** *For each FSG G and each $\delta > 0$, there is a nonvanishing FSG G' such that, for all $w \in \Sigma^*$,*

$$d_{G'}^{(1)}(w) \geq k^{-\delta|w|} d_G^{(1)}(w). \tag{A.5}$$

The following lemma summarizes the second part of the proof of Theorem 2.7.

Lemma A.3 ([6])**.** *For each nonvanishing FSG G and each $l \in \mathbb{Z}^+$, there exists an ILFSC C such that, for all $w \in \Sigma^*$,*

$$|C(w)| \leq (1 + \tfrac{2}{l})|w|\log k - \log d_G^{(1)}(w). \tag{A.6}$$

Proof of Lemma 4.2. Assume the hypothesis. Let

$$\gamma = \log \frac{1}{\beta_{\max}},$$

where

$$\beta_{\max} = \max_{a \in \Sigma} \beta(a).$$

Note that $\gamma > 0$ (because β is positive) and that, for all $w \in \Sigma^*$,

$$\mathcal{I}_\beta(w) \geq \gamma |w|. \tag{A.7}$$

Let

$$\delta = \frac{\gamma\epsilon}{\log k} \tag{A.8}$$

and choose G' for G and δ as in Lemma A.2. Let

$$l = \left\lceil \frac{2\log k}{\gamma\epsilon} \right\rceil, \tag{A.9}$$

and choose C for G' and l as in Lemma A.3. Then, for all $w \in I$,

$$\begin{aligned}
|C(w)| &\leq^{(A.6)} (1 + \tfrac{2}{l})|w| \log k - \log d_{G'}^{(1)}(w) \\
&\leq^{(A.9)} |w|(\gamma\epsilon + \log k) - \log d_{G'}^{(1)}(w) \\
&\leq^{(A.5)} |w|(\gamma\epsilon + \log k) - \log(k^{-\delta|w|} d_G^{(1)}(w)) \\
&= |w|(\gamma\epsilon + \log k + \delta \log k) - \log d_G^{(1)}(w) \\
&= |w|(2\gamma\epsilon + \log k) - \log d_G^{(1)}(w) \\
&= |w|(2\gamma\epsilon + \log k) - \log \left(\frac{\beta(w)^{s-2\epsilon}}{\mu(w)} d_{G,\beta}^{(s-2\epsilon)}(w) \right) \\
&\leq^{(A.3)} |w|(2\gamma\epsilon + \log k) - \log \left(\frac{\beta(w)^{s-2\epsilon}}{\mu(w)} \right) \\
&= |w|(2\gamma\epsilon + \log k) - \log(k^{|w|} \beta(w)^{s-2\epsilon}) \\
&= 2\gamma\epsilon|w| - \log \beta(w)^{s-2\epsilon} \\
&= 2\gamma\epsilon|w| + (s - 2\epsilon)\mathcal{I}_\beta(w) \\
&\leq^{(A.7)} s\mathcal{I}_\beta(w).
\end{aligned}$$

□

Proof of Lemma 5.2. As in the proof of Lemma 3.3, the hypothesis implies that

$$\frac{\mathcal{I}_\mu(w)}{\mathcal{I}_\beta(w)} = \frac{1}{\mathcal{H}_k(\alpha) + \mathcal{D}_k(\alpha||\beta)} + o(1)$$

as $w \to S$. The present lemma follows from this and Lemma 5.1. □

Proof of Theorem 5.3. Assume the hypothesis, and let $l \in \mathbb{Z}^+$. Let $\alpha^{(l)}$ be the restriction of the product probability measure μ^α to Σ^l, noting that $\mathcal{H}(\alpha^{(l)}) = l\mathcal{H}(\alpha)$. We first show that

$$\lim_{n\to\infty} \mathcal{H}(\pi^{(l)}_{R,n}) = \mathcal{H}(\alpha^{(l)}), \tag{A.10}$$

where $\pi^{(l)}_{R,n}$ is the empirical probability measure defined in section 2.5. To see this, let $\epsilon > 0$. By the continuity of the entropy function, there is a real number $\delta > 0$ such that, for all probability measures π on Σ^l,

$$\max_{w\in\Sigma^l} |\pi(w) - \alpha^{(l)}(w)| < \delta \implies |\mathcal{H}(\pi) - \mathcal{H}(\alpha^{(l)})| < \epsilon.$$

Since R is α-normal, there is, for each $w \in \Sigma^l$, a positive integer n_w such that, for all $n \geq n_w$,

$$|\pi^{(l)}_{R,n}(w) - \alpha^{(l)}(w)| = |\pi^{(l)}_{R,n}(w) - \mu^\alpha(w)| < \delta.$$

Let $N = \max_{w\in\Sigma^l} n_w$. Then, for all $n \geq N$, we have $|\mathcal{H}(\pi^{(l)}_{R,n}) - \mathcal{H}(\alpha^{(l)})| < \epsilon$, confirming (A.10).

By Theorem 2.5, we now have

$$\begin{aligned}
\dim_{\mathrm{FS}}(R) &= \mathrm{Dim}_{\mathrm{FS}}(R) \\
&= \inf_{l\in\mathbb{Z}^+} \frac{1}{l\log k} \lim_{n\to\infty} \mathcal{H}(\pi^{(l)}_{R,n}) \\
&= \inf_{l\in\mathbb{Z}^+} \frac{1}{l\log k} \mathcal{H}(\alpha^{(l)}) \\
&= \frac{\mathcal{H}(\alpha)}{\log k} \\
&= \mathcal{H}_k(\alpha).
\end{aligned}$$

□

ECCC ISSN 1433-8092

http://eccc.hpi-web.de/