Theoretical Computer Science 412 (2011) 166–177

Contents lists available at ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

A divergence formula for randomness and dimension

Jack H. Lutz *

Department of Computer Science, Iowa State University, Ames, IA 50011, USA

ARTICLE INFO

Keywords:
Constructive dimension
Finite-state dimension
Kolmogorov complexity
Kullback–Leibler divergence
Randomness
Shannon entropy

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*

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

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 $\mathcal{D}(\alpha \parallel \beta)$ 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 $\dim_{\mathrm{FS}}^{\beta}(R)$ and ${\mathrm{Dim}_{\mathrm{FS}}}^{\beta}(R)$. In the course of proving this, we also prove finite-state compression characterizations of $\dim_{\mathrm{FS}}^{\beta}(S)$ 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 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,

* Tel.: +1 515 294 9941; fax: +1 515 294 0258.
E-mail address: lutz@cs.iastate.edu.

0304-3975/$ – see front matter © 2010 Published by Elsevier B.V.
doi:10.1016/j.tcs.2010.09.005

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 if (but not only if) 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* $\dim_{\mathrm{FS}}^\beta(R)$ 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

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

and

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