Prediction and Dimension

Lance Fortnow

Department of Computer Science, University of Chicago, 1100 E 58th Street, Chicago, IL 60637, fortnow@cs.uchicago.edu

Jack H. Lutz

Department of Computer Science, Iowa State University, Ames, IA 50010, lutz@cs.iastate.edu

Abstract

Given a set X of sequences over a finite alphabet, we investigate the following three quantities.

(i) The *feasible predictability* of X is the highest success ratio that a polynomial-time randomized predictor can achieve on all sequences in X.

(ii) The *deterministic feasible predictability* of X is the highest success ratio that a polynomial-time deterministic predictor can achieve on all sequences in X.

(iii) The *feasible dimension* of X is the polynomial-time effectivization of the classical Hausdorff dimension ("fractal dimension") of X.

Predictability is known to be *stable* in the sense that the feasible predictability of $X \cup Y$ is always the minimum of the feasible predictabilities of X and Y. We show that deterministic predictability also has this property if X and Y are computably presentable. We show that deterministic predictability coincides with predictability on singleton sets. Our main theorem states that the feasible dimension of X is bounded above by the maximum entropy of the predictability of X and bounded below by the segmented self-information of the predictability of X, and that these bounds are tight.

[1] This author's work was done while a senior research scientist at NEC Research Institute.

[2] This author's research was supported in part by National Science Foundation Grants 9988483 and 0344187. Much of the work was done while this author was on sabbatical at NEC Research Institute.

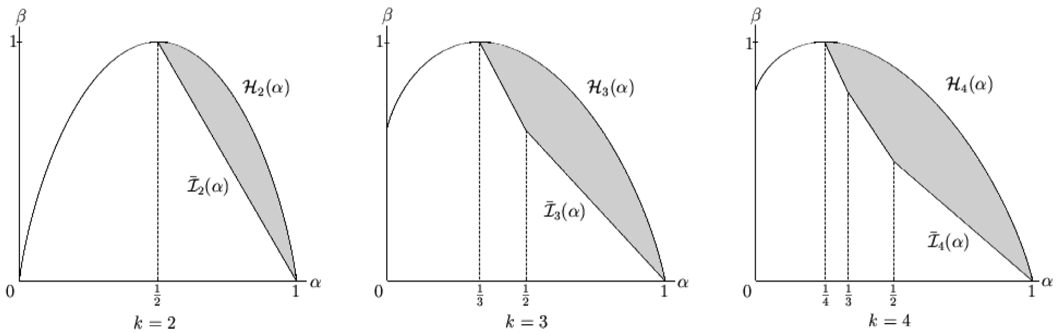

Fig. 1.1. Prediction-dimension diagrams for $k = 2, 3, 4$.

1 Introduction

The relationship between prediction and gambling has been investigated for decades. In the 1950s, Shannon [28] and Kelly [13] studied prediction and gambling, respectively, as alternative means of characterizing information. In the 1960s, Kolmogorov [14] and Loveland [15] introduced a strong notion of unpredictability of infinite binary sequences, now known as *Kolmogorov-Loveland stochasticity.* In the early 1970s, Schnorr [26,27] proved that an infinite binary sequence is *random* (in the sense of Martin-Löf [19]) if and only if no constructive gambling strategy (martingale) can accrue unbounded winnings betting on the successive bits of the sequence. It was immediately evident that every random sequence is Kolmogorov-Loveland stochastic, but the converse question remained open until the late 1980s, when Shen′ [29] established the existence of Kolmogorov-Loveland stochastic sequences that are not random, i.e., sequences that are unpredictable but on which a constructive gambling strategy can accrue unbounded winnings. This result gave a clear qualitative separation between unpredictability and randomness, and hence between prediction and gambling. However, the precise quantitative relationship between these processes has not been elucidated. Given the obvious significance of prediction and gambling for computational learning [3,4,32] and information theory [8,9] this situation should be remedied.

Recently, Lutz [18,16] has defined computational effectivizations of classical Hausdorff dimension ("fractal dimension") and used these to investigate questions in computational complexity and algorithmic information theory. These effectivizations are based not on Hausdorff's 1919 definition of dimension [11,7], but rather on an equivalent formulation in terms of gambling strategies called *gales*. These gales (defined precisely in section 4 below) give a convenient way of quantifying the discount rate against which a gambling strategy can succeed. (The equivalence of the gale formulation with Hausdorff's original definition was proven in [18]. Ryabko [23–25] and Staiger [30,31] have conducted related investigations of classical Hausdorff dimension in terms of the rate at which a gambling strategy can succeed in the absence of discounting, and the gale characterization of Hausdorff dimension can easily be reformulated in these terms [16,1].) The *feasible dimension* $\dim_{\mathrm{p}}(X)$ of a set X of sequences is then defined in terms of the maximum discount rate against which a feasible gambling strategy can succeed.

In this paper we use feasible dimension as a model of feasible gambling, and we compare $\dim_{\mathrm{p}}(X)$ quantitatively with the *feasible predictability* $\mathrm{pred}_{\mathrm{p}}(X)$ of X, which is the highest success ratio that a polynomial-time randomized predictor (defined precisely in section 3 below) can achieve on all sequences in X. Our main theorem, described after this paragraph, gives precise bounds on the relationship between $\mathrm{pred}_{\mathrm{p}}(X)$ and $\dim_{\mathrm{p}}(X)$. We also investigate the *deterministic feasible predictability* $\mathrm{dpred}_{\mathrm{p}}(X)$, in which the predictor is required to commit to a single outcome. We use the probabilistic method to prove that $\mathrm{dpred}_{\mathrm{p}}(X) = \mathrm{pred}_{\mathrm{p}}(X)$ whenever X consists of a single sequence, and we show that deterministic feasible predictability is *stable* on computably presentable sets, i.e., that $\mathrm{dpred}_{\mathrm{p}}(X \cup Y) = \min\{\mathrm{dpred}_{\mathrm{p}}(X), \mathrm{dpred}_{\mathrm{p}}(Y)\}$ whenever the sets X and Y are computably presentable. (Feasible predictability is known to be stable on arbitrary sets [3].)

To describe our main theorem precisely, we need to define two information-theoretic functions, namely, the *k-adic segmented self-information* function $\overline{\mathcal{I}_k}$ and the *k-adic maximum entropy* function $\mathcal{H}_k$.

The *k-adic self-information* of a real number $\alpha \in (0, 1]$ is $\mathcal{I}_k(\alpha) = \log_k \frac{1}{\alpha}$. This is the number of symbols from a k-element alphabet that would be required to represent each of $\frac{1}{\alpha}$ equally probable outcomes (ignoring the fact that $\frac{1}{\alpha}$ may not be an integer). The *k-adic segmented self-information* function $\overline{\mathcal{I}_k} : [\frac{1}{k}, 1] \to [0, 1]$ is defined by setting $\overline{\mathcal{I}_k}(\frac{1}{j}) = \mathcal{I}_k(\frac{1}{j})$ for $1 \le j \le k$ and interpolating linearly between these points.

Recall [5] that the *k-adic entropy* of a probability measure p on a discrete sample space X is

$$H_k(p) = \sum_{x \in X} p(x) \log_k \frac{1}{p(x)}.$$

This is the expected value of $\mathcal{I}_k(p(x))$, i.e., the average number of symbols from a k-element alphabet that is required to represent outcomes of the experiment (X, p) reliably. The *k-adic maximum entropy* function $\mathcal{H}_k : [0, 1] \to [0, 1]$ is defined by

$$\mathcal{H}_k(\alpha) = \max_p H_k(p),$$

where the maximum is taken over all probability measures p on a k-element alphabet Σ such that $p(a) = \alpha$ for some $a \in \Sigma$. This maximum is achieved when the other $k-1$ elements of Σ are equally probable, so

$$\mathcal{H}_k(\alpha) = \alpha \log_k \frac{1}{\alpha} + (1-\alpha) \log_k \frac{k-1}{1-\alpha}.$$

Our main theorem says that for every set $X \subseteq \Sigma^\infty$,

$$\overline{\mathcal{I}_k}(\mathrm{pred}_{\mathrm{p}}(X)) \le \dim_{\mathrm{p}}(X) \le \mathcal{H}_k(\mathrm{pred}_{\mathrm{p}}(X)).$$

That is, the feasible dimension of any set of sequences is bounded below by the k-adic segmented self-information of its feasible predictability and bounded above by the k-adic maximum entropy of its feasible predictability. Graphically, this says that for every set X of sequences, the ordered pair $(\mathrm{pred}_{\mathrm{p}}(X), \dim_{\mathrm{p}}(X))$ lies in the region R_k bounded by the graphs of $\overline{\mathcal{I}_k}$ and $\mathcal{H}_k$. The regions R_2, R_3, and R_4 are depicted in Figure 1. It will be shown in the companion paper [10] that these bounds are tight in the strong sense that for every $k \ge 2$ and every point $(\alpha, \beta) \in R_k$, there is a set X of sequences over a k-element alphabet such that $\mathrm{pred}_{\mathrm{p}}(X) = \alpha$ and $\dim_{\mathrm{p}}(X) = \beta$. Our main theorem is thus a precise statement of the quantitative relationship between feasible predictability and feasible dimension. Since dimension is defined in terms of the achievable success rates of gambling strategies, this can also be regarded as a precise statement of the quantitative relationship between prediction and gambling.

We note that Hitchcock [12] has very recently proven that the feasible dimension $\dim_{\mathrm{p}}(X)$ can be *completely* characterized in terms of the logarithmic loss model of prediction. This enabled him to reinterpret our main theorem as a precise statement of the quantitative relationship between absolute loss prediction and logarithmic loss prediction. (See [20] for a survey of these prediction models.) We also refer the reader to [22,21] for recent work relating Hausdorff dimension to prediction.

For brevity and clarity, our results are stated in terms of feasible (i.e., polynomial-time) prediction and dimension. However, our results generalize to other levels of complexity, ranging from finite-state computation through polynomial-space and unrestricted algorithmic computation and beyond to prediction by arbitrary mathematical functions and classical Hausdorff dimension. At the finite-state level, Feder, Merhav, and Gutman [9] have derived a graph comparing predictability to compressibility for binary sequences. This graph (Figure 3 in [9]) is equivalent to the finite-state version of our $k = 2$ graph in Figure 1. (This equivalence follows from the recent proof by Dai, Lathrop, Lutz, and Mayordomo [6] of the equivalence of finite-state dimension and finite-state compressibility.)

2 Preliminaries

We work in an arbitrary finite alphabet Σ with cardinality $|\Sigma| \geq 2$. When convenient, we assume that Σ has the form $\Sigma = \{0, 1, \ldots, k-1\}$. A *sequence* is an element of Σ^∞, i.e., an infinite sequence of elements of Σ. Given a sequence $S \in \Sigma^\infty$ and natural numbers $i, j \in \mathbb{N}$ with $i \leq j$, we write $S[i..j]$ for the string consisting of the i^{th} through j^{th} symbols of S and $S[i]$ for the i^{th} symbol in S. (The leftmost symbol of S is $S[0]$.) We say that a string $w \in \Sigma^*$ is a *prefix* of S and we write $w \sqsubseteq S$, if $w = S[0..|w| - 1]$.

Given a time bound $t : \mathbb{N} \to \mathbb{N}$, we define the complexity class $\mathrm{DTIME}_\Sigma(t(n))$ to consist of all sequences $S \in \Sigma^\infty$ such that the n^{th} symbol in S can be computed in $O(t(\log n))$ steps. We are especially interested in the classes $\mathrm{DTIME}_\Sigma(2^{cn})$ for fixed $c \in \mathbb{N}$ and the class $\mathrm{E}_\Sigma = \cup_{c\in\mathbb{N}} \mathrm{DTIME}_\Sigma(2^{cn})$. Note that if $S \in \mathrm{E}_\Sigma$, then the time required to compute the n^{th} symbol of S is exponential in the length of the binary representation of n and polynomial in the number n itself.

If D is a discrete domain, then a real-valued function $f : D \to \mathbb{R}$ is *polynomial-time computable* if there is a polynomial-time computable, rational-valued function $\hat{f} : D \times \mathbb{N} \to \mathbb{Q}$ such that for all $x \in D$ and $r \in \mathbb{N}$, $|\hat{f}(x, r) - f(x)| \leq 2^{-r}$.

3 Prediction

Our models of deterministic and randomized prediction are very simple. In both cases, there is a given alphabet Σ containing two or more symbols. Having seen a string $w \in \Sigma^*$ of symbols, a predictor's task is to predict the next symbol.

Definition A *deterministic predictor* on an alphabet Σ is a function

$$\pi : \Sigma^* \to \Sigma.$$

Intuitively, $\pi(w)$ is the symbol that π predicts will follow the string w. This prediction is well-defined and unambiguous, and it is either correct or incorrect. In contrast, a randomized predictor is allowed to simply state the probabilities with which it will predict the various symbols in Σ.

Notation We write $\Delta(\Sigma)$ for the set of all probability measures on Σ, i.e., all functions $p : \Sigma \to [0, 1]$ satisfying $\sum_{a\in\Sigma} p(a) = 1$.

Definition A *(randomized) predictor* on an alphabet Σ is a function

$$\pi : \Sigma^* \to \Delta(\Sigma).$$

Intuitively, having seen the string $w \in \Sigma^*$, a randomized predictor π performs a random experiment in which each symbol $a \in \Sigma$ occurs with probability $\pi(w)(a)$. The outcome of this experiment is the symbol that π predicts will follow w. It is evident that π will be correct with probability $\pi(w)(a)$, where a is the symbol that does in fact follow w.

It is natural to identify each deterministic predictor π on Σ with the randomized predictor $\pi' : \Sigma^* \to \Delta(\Sigma)$ defined by

$$\pi'(w)(a) = \begin{cases} 1 & \text{if } a = \pi(w) \\ 0 & \text{if } a \neq \pi(w). \end{cases}$$

Using this identification, a deterministic predictor is merely a special type of randomized predictor. Thus, in our terminology, a *predictor* is a randomized predictor, and a predictor π is *deterministic* if $\pi(w)(a) \in \{0, 1\}$ for all $w \in \Sigma^*$ and $a \in \Sigma$.

Definition Let π be a predictor on Σ.

(1) The *success rate* of π on a nonempty string $w \in \Sigma^+$ is

$$\pi^+(w) = \frac{1}{|w|} \sum_{i=0}^{|w|-1} \pi(w[0..i-1])(w[i]).$$

(2) The *success rate* of π on a sequence $S \in \Sigma^\infty$ is

$$\pi^+(S) = \limsup_{n\to\infty} \pi^+(S[0..n-1]).$$

(3) The *(worst-case) success rate* of π on a set $X \subseteq \Sigma^\infty$ is

$$\pi^+(X) = \inf_{S \in X} \pi^+(S).$$

Note that $\pi^+(w)$ is the expected fraction of symbols in w that π predicts correctly. In particular, if π is deterministic, then $\pi^+(w)$ is the fraction of symbols in w that π predicts correctly.

We say that a predictor $\pi : \Sigma^* \to \Delta(\Sigma)$ is *feasible* provided that the associated function $\pi' : \Sigma^* \times \Sigma \to [0, 1]$ defined by $\pi'(w, a) = \pi(w)(a)$ is computable in polynomial time. We say that π is *exactly feasible* if the values of π' are rational and can be computed exactly in polynomial time.

Definition Let Σ be an alphabet, and let $X \subseteq \Sigma^\infty$.

(1) The *(randomized feasible) predictability* of X is

$$\mathrm{pred_p}(X) = \sup\{\pi^+(X) | \pi \text{ is a feasible predictor on } \Sigma\}.$$

(2) The *deterministic (feasible) predictability* of X is

$$\mathrm{dpred_p}(X) = \sup\{\pi^+(X) | \pi \text{ is a deterministic feasible predictor on } \Sigma\}.$$

It is clear that

$$0 \leq \mathrm{dpred_p}(X) \leq \mathrm{pred_p}(X)$$

and

$$\frac{1}{|\Sigma|} \leq \mathrm{pred_p}(X) \leq 1$$

for all $X \subseteq \Sigma^\infty$. As the following example shows, all these inequalities can be proper.

Example 3.1 *If*

$$X = \left\{ S \in \{0,1\}^\infty \middle| (\forall n)\left[S[2n] = 1 \text{ or } S[2n+1] = 1\right] \right\},$$

then the reader may verify that

$$\mathrm{dpred_p}(X) = \frac{1}{2} < \frac{5}{8} = \mathrm{pred_p}(X).$$

It is clear that predictability is *monotone* in the sense that

$$X \subseteq Y \Rightarrow \mathrm{pred_p}(X) \geq \mathrm{pred_p}(Y)$$

and

$$X \subseteq Y \Rightarrow \mathrm{dpred_p}(X) \geq \mathrm{dpred_p}(Y)$$

for all $X, Y \subseteq \Sigma^\infty$. Very roughly speaking, the smaller a set of sequences is, the more predictable it is. The following theorem shows that, for fixed $c \in \mathbb{N}$, the set $\mathrm{DTIME}_\Sigma(2^{cn})$ is "completely predictable," while the set E_Σ is "completely unpredictable."

Theorem 3.2 *(1) For each* $c \in \mathbb{N}$,

$$\mathrm{dpred_p}(\mathrm{DTIME}_\Sigma(2^{cn})) = \mathrm{pred_p}(\mathrm{DTIME}_\Sigma(2^{cn})) = 1.$$

(2) $\mathrm{dpred_p}(\mathrm{E}_\Sigma) = 0$, *and* $\mathrm{pred_p}(\mathrm{E}_\Sigma) = \frac{1}{|\Sigma|}$.

Proof. (Sketch.)

(1) For fixed c, there is an n^{c+1}-time-computable function $g : \mathbb{N} \times \Sigma^* \to \Sigma$ such that $\mathrm{DTIME}_\Sigma(2^{cn}) = \{S_0, S_1, \ldots\}$, where $g(k, S_k[0..n-1]) = S_k[n]$ for all $k, n \in \mathbb{N}$. The deterministic predictor $\pi : \Sigma^* \to \Sigma$ defined by $\pi(w) = g(k_w, w)$, where

$$k_w = \min\{k \in \mathbb{N} | (\forall n < |w|) g(k, w[0..n-1]) = w[n]\},$$

is then computable in polynomial time and satisfies $\pi^+(\mathrm{DTIME}_\Sigma(2^{cn})) = 1$.

(2) For any feasible predictor π there is an adversary sequence $S \in \mathrm{E}_\Sigma$ that minimizes the value of $\pi^+(S[0..n])$ at every step n. If π is deterministic, then $\pi^+(S) = 0$. In any case, $\pi^+(S) \leq \frac{1}{|\Sigma|}$.

□

Definition If π_1 and π_2 are predictors on Σ, then the *distance* between π_1 and π_2 is

$$d(\pi_1, \pi_2) = \sup_{w \in \Sigma^*} \max_{a \in \Sigma} |\pi_1(w)(a) - \pi_2(w)(a)|.$$

Observation 3.3 *If π_1 and π_2 are predictors on Σ, then for all $S \in \Sigma^\infty$, $|\pi_1^+(S) - \pi_2^+(S)| \leq d(\pi_1, \pi_2)$.*

Definition Let π be a predictor on Σ, and let $l \in \mathbb{N}$. Then π is *l-coarse* if $2^l \pi(w)(a) \in \mathbb{N}$ for all $w \in \Sigma^*$ and $a \in \Sigma$.

That is, a predictor π is l-coarse if every probability $\pi(w)(a)$ is of the form $\frac{m}{2^l}$ for some $m \in \mathbb{N}$. Note that every l-coarse predictor is $(l+1)$-coarse and that a predictor is deterministic if and only if it is 0-coarse.

Lemma 3.4 (Coarse Approximation Lemma) *For every feasible predictor π on Σ and every $l \in \mathbb{N}$, there is an exactly feasible l-coarse predictor π' such that $d(\pi', \pi) \leq 2^{1-l}$.*

Proof. Let π be a feasible predictor, and let $l \in \mathbb{N}$. Let $c = 1 + \lfloor \log k \rfloor$, where $\Sigma = \{0, 1, \ldots, k-1\}$. Since π is feasible, there is a function $\hat{\pi} : \Sigma^* \times \Sigma \times \mathbb{N} \to \mathbb{Q} \cap [0, 1]$ such that $\hat{\pi}$ is computable in polynomial time and, for all $w \in \Sigma^*$, $a \in \Sigma$, and $r \in \mathbb{N}$, $|\hat{\pi}(w, a, r) - \pi(w)(a)| \leq 2^{-r}$. For each $w \in \Sigma^*$ and $a \in \Sigma$, let

$$m_w(a) = \max\{m \in \mathbb{N} | m \cdot 2^{-l} \leq \hat{\pi}(w, a, l + c)\},$$

and define $\pi' : \Sigma^* \to \Delta(\Sigma)$ by

$$\pi'(w)(a) = \begin{cases} 2^{-l}[m_w(a) + 1] & \text{if } a + \sum_{b \in \Sigma} m_w(b) < 2^l \\ 2^{-l} m_w(a) & \text{if } a + \sum_{b \in \Sigma} m_w(b) \geq 2^l. \end{cases}$$

It is clear that π' is exactly feasible and l-coarse, provided that it is a predictor. Also, for all $w \in \Sigma^*$ and $a \in \Sigma$,

$$2^{-l} m_w(a) \leq \hat{\pi}(w, a, l+c) \leq 2^{-l}[m_w(a) + 1],$$

so

$$|\pi'(w)(a) - \hat{\pi}(w, a, l+c)| \leq 2^{-l},$$

so

$$\begin{aligned} |\pi'(w)(a) - \pi(w)(a)| &\leq 2^{-l} + |\hat{\pi}(w, a, l+c) - \pi(w)(a)| \\ &\leq 2^{-l} + 2^{-(l+c)} \\ &< 2^{1-l}. \end{aligned}$$

It follows that $d(\pi', \pi) \leq 2^{1-l}$.

To see that π' is a predictor, let $w \in \Sigma^*$. A straightforward inspection of the definition of π' shows that

$$\sum_{a \in \Sigma} \pi'(w)(a) = 2^{-l} \left[\sum_{a \in \Sigma} m_w(a) + \min\left\{ k, 2^l - \sum_{a \in \Sigma} m_w(a) \right\} \right].$$

This is clearly 1 if

$$2^l - k \leq \sum_{a \in \Sigma} m_w(a) \leq 2^l, \tag{3.1}$$

so it suffices to establish (3.1).

By the definition of $m_w(a)$ and our choice of c,

$$\begin{aligned} \sum_{a \in \Sigma} m_w(a) &\leq 2^l \sum_{a \in \Sigma} \hat{\pi}(w, a, l+c) \\ &\leq 2^l \sum_{a \in \Sigma} [\pi(w)(a) + 2^{-(l+c)}] \\ &= 2^l + k2^{-c} \\ &< 2^l + 1. \end{aligned}$$

Since each $m_w(a)$ is an integer, this establishes the right-hand inequality in

(3.1). By the maximality of each $m_w(a)$,

$$\begin{aligned}\sum_{a\in\Sigma} m_w(a) &> \sum_{a\in\Sigma}[2^l\hat{\pi}(w,a,l+c)-1]\\ &= 2^l\sum_{a\in\sigma}\hat{\pi}(w,a,l+c)-k\\ &\geq 2^l\sum_{a\in\Sigma}[\pi(w)(a)-2^{-(l+c)}]-k\\ &= 2^l-k2^{-c}-k\\ &> 2^l-(k+1).\end{aligned}$$

Since each $m_w(a)$ is an integer, this implies the first inequality in (3.1). $\square$

We now use the probabilistic method to show that deterministic predictability coincides with predictability on singleton sets.

Theorem 3.5 *For all $S \in \Sigma^\infty$, $\mathrm{dpred_p}(\{S\}) = \mathrm{pred_p}(\{S\})$.*

Proof. Let $S \in \Sigma^\infty$, and let $\alpha < \mathrm{pred_p}(\{S\})$. It suffices to show that $\mathrm{dpred_p}(\{S\}) > \alpha$.

Let $\epsilon = \frac{\mathrm{pred_p}(\{S\})-\alpha}{2}$, and choose $l \in \mathbb{N}$ such that $2^{1-l} < \epsilon$. Since $\alpha+\epsilon < \mathrm{pred_p}(\{S\})$, there is a feasible predictor π' such that $\pi'^{+}(S) > \alpha+\epsilon$. By the Coarse Approximation Lemma, there is an exactly feasible l-coarse predictor π such that $d(\pi,\pi') \leq 2^{1-l} < \epsilon$. It follows by Observation 3.3 that $\pi^+(S) > \alpha$.

For each $w \in \Sigma^*$ and $a \in \Sigma$, define an interval $I(w,a) = [x_a, x_{a+1}) \subseteq [0,1)$ by the recursion

$$x_a = 0, \quad x_{a+1} = x_a + \pi(w)(a).$$

Given $\rho \in [0,1)$, define a deterministic predictor π_ρ on Σ by

$$\pi_\rho(w)(a) = \begin{cases} 1 & \text{if } \rho \in I(w,a)\\ 0 & \text{if } \rho \notin I(w,a).\end{cases}$$

Since π is l-coarse, we have

$$\lfloor 2^l\rho\rfloor = \lfloor 2^l\rho'\rfloor \Rightarrow \pi_\rho = \pi_{\rho'} \tag{3.2}$$

for all $\rho \in [0,1)$. If we choose ρ probabilistically according to the uniform probability measure on [0,1) and E_ρ denotes the expectation with respect to this experiment, then Fatou's lemma tells us that (writing $w_i = S[0..i-1]$)

$$
\begin{aligned}
\mathrm{E}_\rho \pi_\rho^+(S) &= \mathrm{E}_\rho \limsup_{n\to\infty} \pi_\rho^+(w_n) \\
&\geq \limsup_{n\to\infty} \mathrm{E}_\rho \pi_\rho^+(w_n) \\
&= \limsup_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \mathrm{E}_\rho \pi_\rho(w_i)(S[i]) \\
&= \limsup_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \Pr_\rho[\pi_\rho(w_i)(S[i]) = 1] \\
&= \limsup_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \mathrm{length}(I(w_i, S[i])) \\
&= \limsup_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \pi(w_i)(S[i]) \\
&= \limsup_{n\to\infty} \pi^+(w_n) \\
&= \pi^+(S).
\end{aligned}
$$

It follows that there exists $\rho \in [0,1)$ such that $\pi_\rho^+(S) \geq \pi^+(S) > \alpha$. Hence by (3.2) there is a rational $\rho' \in [0,1)$ for which $\pi_{\rho'}^+(S) > \alpha$. Since $\pi_{\rho'}$ is a feasible deterministic predictor, this implies that $\mathrm{dpred}_\mathrm{p}(\{S\}) > \alpha$. □

An important property of predictability is its *stability*, which is the fact that the predictability of a union of two sets is always the minimum of the predictabilities of the sets. (The term "stability" here is taken from the analogous property of dimension [7].) The stability of predictability follows from the (much stronger) main theorem of Cesa-Bianchi, Freund, Helmhold, Haussler, Schapire, and Warmuth [3]. For deterministic predictability, we have the following partial result.

Recall [2] that a set $X \subseteq \Sigma^\infty$ is *computably presentable* if $X = \emptyset$ or there is a computable function $f : \mathbb{N} \to \mathbb{N}$ such that $X = \{L(M_{f(i)}) | i \in \mathbb{N}\}$, where $M_0, M_1, \ldots$ is a standard enumeration of all Turing machines over the alphabet Σ and $M_{f(i)}$ decides the sequence $L(M_{f(i)})$ for all $i \in \mathbb{N}$. Deterministic predictability is stable on sets that are computably presentable.

Theorem 3.6 *For all computably presentable sets $X, Y \subseteq \Sigma^\infty$,*

$$\mathrm{dpred}_\mathrm{p}(X \cup Y) = \min\{\mathrm{dpred}_\mathrm{p}(X), \mathrm{dpred}_\mathrm{p}(Y)\}.$$

Proof. Let $X, Y \subseteq \Sigma^\infty$ be computably presentable. Then there exist computable functions $f, g : \mathbb{N} \to \Sigma$ such that if we let $S_k[n] = f(k,n)$ and $T_k[n] = g(k,n)$ for all $k, n \in \mathbb{N}$, then $X = \{S_0, S_1, S_2, \ldots\}$ and $Y = \{T_0, T_1, T_2, \ldots\}$. Fix arbitrary reals $\alpha < \mathrm{dpred}_\mathrm{p}(X)$ and $\beta < \mathrm{dpred}_\mathrm{p}(Y)$. It suffices to show that

$$\mathrm{dpred}_\mathrm{p}(X \cup Y) > \min\{\alpha, \beta\}. \tag{3.3}$$

By our choice of α and β, there exist deterministic feasible predictors π_X and π_Y such that for all $S \in \Sigma^\infty$,

$$S \in X \Rightarrow \pi_X^+(S) > \alpha \tag{3.4}$$

and

$$S \in Y \Rightarrow \pi_Y^+(S) > \beta. \tag{3.5}$$

To prove (3.3) it suffices to construct a deterministic feasible predictor π such that for all $S \in \Sigma^\infty$,

$$S \in X \cup Y \Rightarrow \pi^+(S) > \min\{\alpha, \beta\}. \tag{3.6}$$

The idea of the construction of π is simple. Given a prefix w of a sequence $S \in \Sigma^\infty$, π attempts to predict the next symbol of S. For each such w, π has a *working hypothesis* concerning the identity of S. This working hypothesis is formally a nonnegative integer $h(w)$. Intuitively, the working hypothesis "$S = S_k$" is represented by the condition $h(w) = 2k$, while the working hypothesis "$S = T_k$" is represented by the condition $h(w) = 2k + 1$. Our predictor π is then defined by

$$\pi(w) = \begin{cases} \pi_X(w) & \text{if } h(w) \text{ is even} \\ \pi_Y(w) & \text{if } h(w) \text{ is odd.} \end{cases} \tag{3.7}$$

We define h so that

$$h \text{ is feasible} \tag{3.8}$$

and for all $S \in \Sigma^\infty$,

$$S \in X \cup Y \Rightarrow \begin{cases} \text{for every sufficiently long prefix } w \sqsubseteq S, \\ h(w) \text{ is the least correct working hypothesis.} \end{cases} \tag{3.9}$$

It is clear that (3.7) and (3.8) imply that π is feasible. It is also clear that (3.7), (3.9), (3.4), and (3.5) imply (3.6). Thus it suffices to define h so that (3.8) and (3.9) hold.

The function h is computed by the following "sudden death" algorithm.

input $w \in \Sigma^*$;
$h(w) := 0$;
for $|w|$ computation steps **do**
  while true **do**
  begin
    $\phi :=$ **if** $h(w)$ is even **then** f **else** g;
    $k := \left\lfloor \frac{h(w)}{2} \right\rfloor$;
    if $(\exists n \in \mathbb{N})\phi(k, n) \neq w[n]$

then $h(w) := h(w) + 1$**;**
end;
output $h(w)$.

A few remarks on this algorithm are in order. The while-loop would be non-terminating were it not for the "sudden death condition" that its execution is terminated after a total $|w|$ computation steps. Typically this sudden death termination occurs part of the way through a computation of some value of $\phi(k, n)$. In any event, the value of $h(w)$ at the time of this sudden death termination is the final output.

We stipulate that the if-test is evaluated by checking successive values of n (starting at 0 during each iteration of the while-loop) until either the if-condition is determined to be true or the sudden death termination occurs. If the if-test is true, then the working hypothesis $h(w)$ is incorrect (because $w \sqsubseteq S$) and is thus incremented. The final output $h(w)$ is thus the least working hypothesis that is not discovered to be incorrect within $|w|$ computation steps.

It is clear that (3.8) holds. To see that (3.9) holds, let $S \in X \cup Y$. Then $S \in X$ or $S \in Y$, so there exists a working hypothesis "$S = S_k$" or "$S = T_k$" that is correct. Let $m \in \mathbb{N}$ be the least correct working hypothesis. Then there is a prefix $w_0 \sqsubseteq S$ such that every working hypothesis $m' < m$ is discovered to be incorrect within $|w_0|$ computation steps. Since m is correct, it follows that for all w such that $w_0 \sqsubseteq w \sqsubseteq S$, we have $h(w) = m$. Thus (3.9) holds. □

An earlier draft of this paper conjectured, but did not prove, that deterministic predictability is not stable on arbitrary sets. We thank an anonymous referee for proving this conjecture via the following simple example.

Example 3.7 *For each $a \in \Sigma$, let π_a be the deterministic predictor that always predicts a, and let*

$$X_a = \{S \in \Sigma^\infty | \pi^+(S) \geq \frac{1}{k}\}.$$

Then each $\mathrm{dpred_p}(X_a) = \frac{1}{k}$*, but*

$$\mathrm{dpred_p}\left(\bigcup_{a \in \Sigma} X_a\right) = \mathrm{dpred_p}(\Sigma^\infty) = 0$$

by Theorem 3.2.

4 Dimension

In this section we sketch the elements of feasible dimension in Σ^∞, where Σ is a finite alphabet. Without loss of generality, we let $\Sigma = \{0, 1, \ldots, k-1\}$, where $k \geq 2$.

Definition Let $s \in [0, \infty)$.

(1) An *s-gale* over Σ is a function $d : \Sigma^* \to [0, \infty)$ that satisfies the condition

$$d(w) = k^{-s} \sum_{a \in \Sigma} d(wa) \tag{4.1}$$

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

(2) An s-gale d *succeeds* on a sequence $S \in \Sigma^\infty$, and we write $S \in S^\infty[d]$, if

$$\limsup_{n \to \infty} d(S[0..n-1]) = \infty.$$

(3) An s-gale is *feasible* if it is computable in polynomial time.

(4) An s-gale is *exactly feasible* if its values are rational and can be computed exactly in polynomial time.

(5) For $X \subseteq \Sigma^\infty$, we let

$$\mathcal{G}(X) = \{ s \mid \text{there is an } s\text{-gale } d \text{ such that } X \subseteq S^\infty[d] \},$$
$$\mathcal{G}_p(X) = \{ s \mid \text{there is a feasible } s\text{-gale } d \text{ such that } X \subseteq S^\infty[d] \}.$$

The gale characterization of classical Hausdorff dimension [18] shows that the classical Hausdorff dimension $\dim_{\mathrm{H}}(X)$ of a set $X \subseteq \Sigma^\infty$ is given by the equation

$$\dim_{\mathrm{H}}(X) = \inf \mathcal{G}(X).$$

This motivates the following.

Definition The *feasible dimension* of a set $X \subseteq \Sigma^\infty$ is

$$\dim_{\mathrm{p}}(X) = \inf \mathcal{G}_p(X).$$

It is easy to see that $0 \leq \dim_{\mathrm{H}}(X) \leq \dim_{\mathrm{p}}(X) \leq 1$ for all $X \subseteq \Sigma^\infty$ and that feasible dimension is *monotone* in the sense that $X \subseteq Y$ implies $\dim_{\mathrm{p}}(X) \subseteq \dim_{\mathrm{p}}(Y)$ for all $X, Y \subseteq \Sigma^\infty$. It is shown in [18] that feasible dimension is *stable* in the sense that

$$\dim_{\mathrm{p}}(X \cup Y) = \max\{\dim_{\mathrm{p}}(X), \dim_{\mathrm{p}}(Y)\}$$

for all $X, Y \subseteq \Sigma^\infty$. The following result is the dimension-theoretic analog of Theorem 3.2.

Theorem 4.1 ([18])

(1) For each $c \in \mathbb{N}$, $\dim_{\mathrm{p}}(\mathrm{DTIME}_\Sigma(2^{cn})) = 0$.

(2) $\dim_{\mathrm{p}}(\mathrm{E}_\Sigma) = 1$

The following example establishes the existence of sets of arbitrary feasible dimension between 0 and 1.

Example 4.2 ([18]) *Let q be a feasible probability measure on Σ, and let $\mathrm{FREQ}(q)$ be the set of all sequences $S \in \Sigma^\infty$ such that each $a \in \Sigma$ has asymptotic frequency $q(a)$ in S. Then*

$$\dim_{\mathrm{p}}(\mathrm{FREQ}(q)) = H_k(q).$$

5 Prediction versus Dimension

This section develops our main theorem, which gives precise quantitative bounds on the relationship between predictability and dimension. As before, let $\Sigma = \{0, 1, \ldots, k-1\}$ be an alphabet with $k \geq 2$. Recall the k-adic segmented self-information function $\overline{\mathcal{I}_k}$ and the k-adic maximum entropy function $\mathcal{H}_k$ defined in section 1.

Theorem 5.1 (Main Theorem) *For all $X \subseteq \Sigma^\infty$,*

$$\overline{\mathcal{I}_k}(\mathrm{pred}_{\mathrm{p}}(X)) \leq \dim_{\mathrm{p}}(X) \leq \mathcal{H}_k(\mathrm{pred}_{\mathrm{p}}(X)).$$

The rest of this section is devoted to proving Theorem 5.1.

Construction 5.2 *Given an alphabet Σ with $|\Sigma| = k \geq 2$, a predictor π on Σ, and rational numbers $\beta, s \in (\frac{1}{k}, 1)$, we define an s-gale*

$$d = d(\pi, \beta, s) : \Sigma^* \to [0, \infty)$$

by the recursion

$$d(\lambda) = 1,$$
$$d(wa) = k^s \mathrm{bet}_w(a) d(w),$$

where $\mathrm{bet}_w(a)$, the amount that d bets on a having seen w, is defined as follows. If π were to deterministically predict b (i.e., $\pi(w)(b) = 1$), then the amount that d would bet on a is

$$\gamma(a, b) = \begin{cases} \beta & \text{if } a = b \\ \frac{1-\beta}{k-1} & \text{if } a \neq b. \end{cases}$$

However, π is a randomized predictor that predicts various b according to the probability measure $\pi(w)$, so d instead uses the quantity

$$\begin{aligned}\gamma_w(a) &= \prod_{b\in\Sigma} \gamma(a,b)^{\pi(w)(b)} \\ &= \beta^{\pi(w)(a)} \left(\frac{1-\beta}{k-1}\right)^{1-\pi(w)(a)},\end{aligned}$$

which is the geometric mean of the bets $\gamma(a,b)$, weighted according to the probability measure $\pi(w)$. The amount that d bets on a is then the normalization

$$\mathrm{bet}_w(a) = \frac{\gamma_w(a)}{\sigma_w},$$

where

$$\sigma_w = \sum_{a\in\Sigma} \gamma_w(a).$$

Observation 5.3 *In Construction 5.2, $0 < \sigma_w \le 1$ for all $w \in \Sigma^*$.*

Observation 5.4 *In Construction 5.2, d is an s-gale, and d is p-computable if π is feasible.*

Lemma 5.5 *In Construction 5.2,*

$$\log_k d(w) \ge |w| \left(s + \log_k \frac{1-\beta}{k-1} + \pi^+(w) \log_k \frac{\beta(k-1)}{1-\beta} \right)$$

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

Proof. Let $w \in \Sigma^*$, and let $n = |w|$. For each $0 \le i < n$, write $\pi_i = \pi(w[0..i-1])(w[i])$. By the construction of d and Observation 5.3,

$$\begin{aligned}d(w) &= k^{sn} \prod_{i=0}^{n-1} \mathrm{bet}_{w[0..i-1]}(w[i]) \\ &= k^{sn} \prod_{i=0}^{n-1} \frac{\gamma_{w[0..i-1]}(w[i])}{\sigma_{w[0..i-1]}} \\ &\ge k^{sn} \prod_{i=0}^{n-1} \gamma_{w[0..i-1]}(w[i]) \\ &= k^{sn} \prod_{i=0}^{n-1} \beta^{\pi_i} \left(\frac{1-\beta}{k-1}\right)^{1-\pi_i}.\end{aligned}$$

It follows that

$$
\begin{aligned}
\log_k d(w) &\geq sn + \sum_{i=0}^{n-1} \left[\pi_i \log_k \beta + (1 - \pi_i) \log_k \frac{1-\beta}{k-1} \right] \\
&= n \left(s + \log_k \frac{1-\beta}{k-1} \right) + \log_k \frac{\beta(k-1)}{1-\beta} \sum_{i=0}^{n-1} \pi_i \\
&= n \left(s + \log_k \frac{1-\beta}{k-1} + \pi^+(w) \log_k \frac{\beta(k-1)}{1-\beta} \right).
\end{aligned}
$$

□

We can now prove an upper bound on dimension in terms of predictability.

Theorem 5.6 *If Σ is an alphabet with $|\Sigma| = k \geq 2$, then for all $X \subseteq \Sigma^\infty$,*

$$
\dim_{\mathrm{p}}(X) \leq \mathcal{H}_k(\mathrm{pred}_{\mathrm{p}}(X)).
$$

Proof. Let $X \subseteq \Sigma^\infty$, and let $\alpha = \mathrm{pred}_{\mathrm{p}}(X)$. If $\mathcal{H}_k(\alpha) = 1$ then the result holds trivially, so assume that $\mathcal{H}_k(\alpha) < 1$, i.e., $\alpha \in \left(\frac{1}{k}, 1\right]$. Choose a rational number $s \in (\mathcal{H}_k(\alpha), 1]$. It suffices to show that $\dim_{\mathrm{p}}(X) \leq s$.

By our choice of s, there is a rational number $\beta \in \left(\frac{1}{k}, \alpha\right)$ such that $\mathcal{H}_k(\beta) \in (\mathcal{H}_k(\alpha), s)$. Since $\beta < \alpha$, there is a feasible predictor π such that $\pi^+(X) > \beta$. Let $d = d(\pi, \beta, s)$ be the s-gale of Construction 5.2. By Observation 5.4, it suffices to show that $X \subseteq S^\infty[d]$. To this end, let $S \in X$. For each $n \in \mathbb{N}$, let $w_n = S[0..n-1]$. Then the set

$$
J = \{ n \in \mathbb{Z}^+ | \pi^+(w_n) \geq \beta n \}
$$

is infinite, and Lemma 5.5 tells us that for each $n \in J$,

$$
\begin{aligned}
\log_k d(w_n) &\geq n \left(s + \log_k \frac{1-\beta}{k-1} + \pi^+(w_n) \log_k \frac{\beta(k-1)}{1-\beta} \right) \\
&\geq n \left(s + \log_k \frac{1-\beta}{k-1} + \beta \log_k \frac{\beta(k-1)}{1-\beta} \right) \\
&= n(s - \mathcal{H}_k(\beta)).
\end{aligned}
$$

Since $s > \mathcal{H}_k(\beta)$, this implies that $S \in S^\infty[d]$. □

The lower bound on dimension is a function of predictability whose graph is not a smooth curve. It is thus instructive to *derive* this bound rather than to simply assert and prove it. As before, let Σ be an alphabet with $|\Sigma| \geq 2$.

It is easiest to first derive a lower bound on predictability in terms of dimension, since this can be achieved by using an s-gale to construct a predictor.

So let s be a positive rational, and let d be a p-computable s-gale over Σ with $d(\lambda) > 0$. The most natural predictor to construct from d is the function $\pi_0 : \Sigma^* \to \Delta(\Sigma)$ defined by

$$\pi_0(w)(a) = \mathrm{bet}_d(wa) \tag{5.1}$$

for all $w \in \Sigma^*$ and $a \in \Sigma$. This is indeed a predictor, and it is clearly feasible. For all $w \in \Sigma^*$, we have

$$\begin{aligned} d(w) &= d(\lambda) k^{s|w|} \prod_{i=0}^{|w|-1} \mathrm{bet}_d(wa) \\ &\leq d(\lambda) k^{s|w|} \left(\frac{1}{|w|} \sum_{i=0}^{|w|-1} \mathrm{bet}_d(wa) \right)^{|w|} \\ &= d(\lambda) \left(k^s \pi_0^+(w) \right)^{|w|} \end{aligned}$$

(because the geometric mean is at most the arithmetic mean), so if $S \in S^\infty[d]$ there must be infinitely many prefixes $w \sqsubseteq S$ for which $\pi_0^+(w) > k^{-s}$. Thus this very simple predictor π_0 testifies that

$$\mathrm{pred}_\mathrm{p}(S^\infty[d]) \geq k^{-s}. \tag{5.2}$$

This establishes the following preliminary bound.

Lemma 5.7 *For all $X \subseteq \Sigma^\infty$,*

$$\dim_\mathrm{p}(X) \geq \mathcal{I}_k(\mathrm{pred}_\mathrm{p}(X)).$$

Proof. The above argument shows that

$$\mathrm{pred}_\mathrm{p}(X) \geq k^{-\dim_\mathrm{p}(X)},$$

whence the lemma follows immediately. □

If we suspect that Lemma 5.7 can be improved, how might we proceed? One approach is as follows. The predictor π_0 achieved (5.2) via the prediction probability (5.1), which is equivalent to

$$\pi_0(w)(a) = k^{-\mathcal{I}_k(\mathrm{bet}_d(wa))}. \tag{5.3}$$

To improve on (5.2), let $f(s) = u - vs$ be a function whose graph is a line intersecting k^{-s} at two points given by $s_0, s_1 \in [0, 1]$. We would like to improve (5.2) to

$$\mathrm{pred}_\mathrm{p}(S^\infty[d]) \geq f(s). \tag{5.4}$$

For what values of s_0 and s_1 can we establish (5.4)?

Guided by (5.3), we set

$$\pi_1(w)(a) = \max\{0, f(\mathcal{I}_k(\mathrm{bet}_d(wa)))\}$$

for all $w \in \Sigma^*$ and $a \in \Sigma$. The function π_1 may not be a predictor because the function $\sigma : \Sigma^* \to [0, \infty)$ defined by

$$\sigma(w) = \sum_{a \in \Sigma} \pi_1(w)(a)$$

may not be identically 1. However, it is clear that $\sigma(w) > 0$ for all $w \in \Sigma^*$, so if we set

$$\pi(w)(a) = \frac{\pi_1(w)(a)}{\sigma(w)}$$

for all $w \in \Sigma^*$ and $a \in \Sigma$, then π is a predictor. For all $w \in \Sigma^+$ we have

$$\begin{aligned}
\pi_1^+(w) &\geq \frac{1}{|w|} \sum_{i=0}^{|w|-1} (u - v\mathcal{I}_k(\mathrm{bet}_d(w[0..i]))) \\
&= u + \frac{v}{|w|} \sum_{i=0}^{|w|-1} \log_k(\mathrm{bet}_d(w[0..i])) \\
&= u + \frac{v}{|w|} \log_k \prod_{i=0}^{|w|-1} \mathrm{bet}_d(w[0..i]) \\
&= u + \frac{v}{|w|} \log_k \left(\frac{d(w)}{k^{s|w|} d(\lambda)} \right) \\
&= u - vs + \frac{1}{|w|} \log_k \frac{d(w)}{d(\lambda)},
\end{aligned}$$

so if $\sigma(w) \leq 1$ and $d(w) > d(\lambda)$, then

$$\pi^+(w) > u - vs = f(s).$$

Thus if s_0 and s_1 are chosen so that $\sigma(w) \leq 1$ for all $w \in \Sigma^*$, then for all $S \in S^\infty[d]$ there exist infinitely many prefixes $w \sqsubseteq S$ for which $\pi^+(w) > f(s)$. This implies that (5.4) holds (provided that π is feasible). Thus the question is how to choose s_0 and s_1 so that $\sigma(w) \leq 1$ for all $w \in \Sigma^*$.

If we let

$$B_w = \{a | f(\mathcal{I}_k(\mathrm{bet}_d(wa))) > 0\},$$

then for all $w \in \Sigma^*$,

$$\begin{aligned}\sigma(w) &= \sum_{a\in B_w} f(\mathcal{I}_k(\mathrm{bet}_d(wa)))\\ &= u|B_w| + v\sum_{a\in B_w}\log_k \mathrm{bet}_d(wa)\\ &= u|B_w| + v\log_k \prod_{a\in B_w}\mathrm{bet}_d(wa)\\ &\leq u|B_w| + v\log_k\left(\frac{1}{|B_w|}\sum_{a\in B_w}\mathrm{bet}_d(wa)\right)^{|B_w|}\\ &\leq |B_w|(u - v\log_k|B_w|)\\ &= g(|B_w|),\end{aligned}$$

where

$$g(x) = xf(\log_k(x)).$$

Since $|B_w| \leq k$ for all $w \in \Sigma^*$, it thus suffices to choose s_0 and s_1 so that

$$g(j) \leq 1 \tag{5.5}$$

for all $1 \leq j \leq k$. Of course we want our lower bound f, and hence the function g, to be as large as possible while satisfying (5.5). Since

$$g'(x) = u - v\left(\frac{1}{\ln k} + \log_k x\right)$$

is positive to the left of some point (namely, $x = \frac{k^{\frac{u}{v}}}{e}$) and negative to the right of this point, (5.5) can be achieved by arranging things so that

$$g(i) = g(i+1) = 1 \tag{5.6}$$

for some (any!) $1 \leq i < k$. Now (5.6) is equivalent to the conditions

$$f(\log_k i) = \frac{1}{i}, \quad f(\log_k(i+1)) = \frac{1}{i+1},$$

which simply say that

$$s_0 = \log_k i, \quad s_1 = \log_k(i+1). \tag{5.7}$$

For $1 \leq i < k$, the predictor π determined by the choice of (5.7) is feasible and thus establishes (5.4). This argument yields the following improvement of Lemma 5.7.

Theorem 5.8 *For all $X \subseteq \Sigma^\infty$,*

$$\dim_{\mathrm{p}}(X) \geq \overline{\mathcal{I}_k}(\mathrm{pred}_{\mathrm{p}}(X)).$$

Proof. For each $1 \leq i < k$, if we let $f_i(s) = u_i - v_i s$ be the function that agrees with k^{-s} at $\log_k i$ and $\log_k(i+1)$, then the above argument shows that

$$\mathrm{pred_p}(X) \geq f_i(\dim_\mathrm{p}(X)),$$

whence

$$\dim_\mathrm{p}(X) \geq f_i^{-1}(\mathrm{pred_p}(X)).$$

Since $\mathrm{pred_p}(X) \geq \frac{1}{k}$ in any case and f_i^{-1} agrees with $\overline{\mathcal{I}_k}$ on $[\frac{1}{i+1}, \frac{1}{i}]$, this establishes the theorem. □

For each $k \geq 2$, let R_k be the set of all $\alpha, \beta \in [0,1]$ satisfying $\alpha \geq \frac{1}{k}$ and $\overline{\mathcal{I}_k}(\alpha) \leq \beta \leq \mathcal{H}_k(\alpha)$. Thus R_2, R_3, and R_4 are the shaded regions depicted in Figure 1, and Theorem 5.1 says that $(\mathrm{pred_p}(X), \dim_\mathrm{p}(X)) \in R_k$ for all $k \geq 2$ and $X \subseteq \Sigma^\infty$. In fact, Theorem 5.1 is tight in the strong sense that for each $(\alpha, \beta) \in R_k$ there is a set $X \subseteq \mathrm{E}_\Sigma$ such that $\mathrm{pred_p}(X) = \alpha$ and $\dim_\mathrm{p}(X) = \beta$. (A proof using the techniques of the present paper is lengthy and cumbersome. A better proof, using more recent techniques, will appear in the companion paper [10].) Thus R_k is precisely the set of all points of the form $(\mathrm{pred_p}(X), \dim_\mathrm{p}(X))$ for $X \subseteq \Sigma^\infty$ (or, equivalently, for $X \subseteq \mathrm{E}_\Sigma$).

Let R_∞ be the limit of the regions R_k, in the sense that R_∞ consists of all $(\alpha, \beta) \in [0,1]^2$ such that for every $\epsilon > 0$, for every sufficiently large k, there exists $(\alpha', \beta') \in R_k$ such that $(\alpha - \alpha')^2 + (\beta - \beta')^2 < \epsilon$. Then it is interesting to note that R_∞ is the triangular region given by the inequalities $\alpha \geq 0, \beta \geq 0, \alpha + \beta \leq 1$. Thus if the alphabet Σ is very large, then the primary constraint is simply that a set's predictability cannot be significantly greater than 1 minus its dimension.

Acknowledgements

We thank John Hitchcock, Anumodh Abey, and three anonymous referees for useful suggestions.

References

[1] K. Ambos-Spies, W. Merkle, J. Reimann, and F. Stephan. Hausdorff dimension in exponential time. In *Proceedings of the 16th IEEE Conference on Computational Complexity*, pages 210–217, 2001.

[2] J. L. Balcázar, J. Díaz, and J. Gabarró. *Structural Complexity I*. Springer-Verlag, Berlin, second edition, 1995.

[3] N. Cesa-Bianchi, Y. Freund, D. Haussler, D. P. Helmbold, R. E. Schapire, and M. K. Warmuth. How to use expert advice. *Journal of the ACM*, 44(3):427–485, 1997.

[4] N. Cesa-Bianchi and G. Lugosi. On prediction of individual sequences. *Annals of Statistics*, 27(6):1865–1895, 1999.

[5] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., New York, N.Y., 1991.

[6] J. J. Dai, J. I. Lathrop, J. H. Lutz, and E. Mayordomo. Finite-state dimension. *Theoretical Computer Science*. To appear. Preliminary version appeared in *Proceedings of the 28th International Colloquium on Automata, Languages, and Programming*, pages 1028–1039, 2001.

[7] K. Falconer. *Fractal Geometry: Mathematical Foundations and Applications*. John Wiley & Sons, 1990.

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

[9] M. Feder, N. Merhav, and M. Gutman. Universal prediction of individual sequences. *IEEE Transations on Information Theory*, 38:1258–1270, 1992.

[10] L. Fortnow, J. M. Hitchcock, J. H. Lutz, and E. Mayordomo. Prediction and dimension of products and convex combinations. In preparation.

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

[12] J. M. Hitchcock. Fractal dimension and logarithmic loss unpredictability. *Theoretical Computer Science*, 304(1–3):431–441, 2003.

[13] J. Kelly. A new interpretation of information rate. *Bell Systems Technical Journal*, 35:917–926, 1956.

[14] A. N. Kolmogorov. On tables of random numbers. *Sankhyā, Series A*, 25:369–376, 1963.

[15] D. W. Loveland. A new interpretation of von Mises' concept of a random sequence. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 12:279–294, 1966.

[16] J. H. Lutz. The dimensions of individual strings and sequences. *Information and Computation*. To appear. Preliminary version appeared as [17].

[17] J. H. Lutz. Gales and the constructive dimension of individual sequences. In *Proceedings of the 27th International Colloquium on Automata, Languages, and Programming*, pages 902–913, 2000. Revised as [16].

[18] J. H. Lutz. Dimension in complexity classes. *SIAM Journal on Computing*, 32:1236–1259, 2003. To appear. Preliminary version appeared in *Proceedings of the Fifteenth Annual IEEE Conference on Computational Complexity*, pages 158–169, 2000.

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

[20] N. Merhav and M. Feder. Universal prediction. *IEEE Transactions on Information Theory*, 44(6):2124–2147, 1998.

[21] B. Ryabko. The nonprobabilistic approach to learning the best prediction. in: Linkoping Electronic Articles in Computer and Information Science, ISSN 1401-9841 Issue: Vol. 6(2001):nr 016.

[22] B. Ryabko, J. Suzuki, and F. Topsoe. Hausdorff dimension as a new dimension in source coding and predicting. 1999 IEEE Information Theory Workshop, South Africa, 1999, pp.66-68.

[23] B. Ya. Ryabko. Noiseless coding of combinatorial sources. *Problems of Information Transmission*, 22:170–179, 1986.

[24] B. Ya. Ryabko. Algorithmic approach to the prediction problem. *Problems of Information Transmission*, 29:186–193, 1993.

[25] B. Ya. Ryabko. The complexity and effectiveness of prediction problems. *Journal of Complexity*, 10:281–295, 1994.

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

[27] C. P. Schnorr. Zufälligkeit und Wahrscheinlichkeit. *Lecture Notes in Mathematics*, 218, 1971.

[28] C. E. Shannon. Certain results in coding theory for noisy channels. *Bell Systems Technical Journal*, 30:50–64, 1951.

[29] A. Kh. Shen′. On relations between different algorithmic definitions of randomness. *Soviet Mathematics Doklady*, 38:316–319, 1989.

[30] L. Staiger. Kolmogorov complexity and Hausdorff dimension. *Information and Computation*, 103:159–94, 1993.

[31] L. Staiger. A tight upper bound on Kolmogorov complexity and uniformly optimal prediction. *Theory of Computing Systems*, 31:215–29, 1998.

[32] V. Vovk. A game of prediction with expert advice. *Journal of Computer and System Sciences*, pages 153–173, 1998.