Fractal Dimension and Logarithmic Loss Unpredictability*

John M. Hitchcock
Department of Computer Science
Iowa State University
jhitchco@cs.iastate.edu

Abstract

We show that the Hausdorff dimension equals the logarithmic loss unpredictability for any set of infinite sequences over a finite alphabet. Using computable, feasible, and finite-state predictors, this equivalence also holds for the computable, feasible, and finite-state dimensions. Combining this with recent results of Fortnow and Lutz (2002), we have a tight relationship between prediction with respect to logarithmic loss and prediction with respect to absolute loss.

1 Introduction

We establish a fundamental relationship between logarithmic loss and Hausdorff dimension, central ideas in two very active research areas. Hausdorff dimension [7] is a refinement of Lebesgue measure that has become a powerful tool in fractal geometry [3]. Logarithmic loss (also known as self-information loss) is very important in the theory of prediction. The survey by Merhav and Feder on universal prediction [9] contains historical references and a thorough discussion of the motivation and significance of logarithmic loss.

Given a set X of infinite sequences over a finite alphabet, consider the problem of designing a single predictor that performs well on all sequences in X. Informally, we define the unpredictability of X as the minimal average loss, with respect to a given loss function, that a predictor can achieve on all members of X. (Our prediction model is standard. Technical definitions will be given in the body of the paper.)

Relationships between prediction and Hausdorff dimension have been investigated by Fortnow and Lutz [6], Ryabko [10, 11, 12], and Staiger [14]. In this note we show that unpredictability with respect to logarithmic loss corresponds *exactly* to Hausdorff dimension. For any set X of sequences, writing $\dim_{\mathrm{H}}(X)$ for the Hausdorff dimension of X and $\mathrm{unpred}^{\log}(X)$ for the logarithmic loss unpredictability of X, we prove that

$$\dim_{\mathrm{H}}(X) = \mathrm{unpred}^{\log}(X). \tag{1.1}$$

Fortnow and Lutz [6] defined the predictability of a set of sequences as the maximum "success" achievable by any predictor on the set. This notion of predictability corresponds with absolute loss

*This research was supported in part by National Science Foundation Grant 9988483.

unpredictability. Specifically, if we denote the absolute loss unpredictability of X by $\mathrm{unpred}^{\mathrm{abs}}(X)$ and the Fortnow-Lutz predictability of X by $\mathrm{pred}(X)$, we have

$$\mathrm{pred}(X) = 1 - \mathrm{unpred}^{\mathrm{abs}}(X) \tag{1.2}$$

for every set X.

For any set X of *binary* sequences, Fortnow and Lutz proved that

$$2(1 - \mathrm{pred}(X)) \leq \dim_{\mathrm{H}}(X) \leq \mathcal{H}(\mathrm{pred}(X)), \tag{1.3}$$

where $\mathcal{H}$ is the binary entropy function. Combining (1.1), (1.2), and (1.3), we have a relationship between unpredictability with respect to absolute loss and unpredictability with respect to logarithmic loss:

$$2 \cdot \mathrm{unpred}^{\mathrm{abs}}(X) \leq \mathrm{unpred}^{\log}(X) \leq \mathcal{H}(\mathrm{unpred}^{\mathrm{abs}}(X)) \tag{1.4}$$

for any set X of binary sequences. Fortnow and Lutz also stated that the bounds in (1.3), and hence in (1.4), are tight in a very strong sense. Using this result, for any $\alpha \in [0, \frac{1}{2}]$ and $\beta \in [2 \cdot \alpha, \mathcal{H}(\alpha)]$, there is a set X of binary sequences with $\mathrm{unpred}^{\mathrm{abs}}(X) = \alpha$ and $\mathrm{unpred}^{\log} = \beta$. (Fortnow and Lutz also gave tight bounds for sequences over non-binary alphabets; in this introduction we only state the binary case for simplicity.)

Our main result, that (1.1) holds for all sets X, is proved using Lutz's characterization of Hausdorff dimension by gales [8]. Lutz introduced gales, betting functions that are generalizations of martingales, in order to effectivize Hausdorff dimension. Our proof of (1.1) makes use of a natural correspondence between gales and predictors.

Lutz [8] used feasible (polynomial-time) and computable gales to define feasible and computable dimensions. Subsequently, Dai, Lathrop, Lutz, and Mayordomo used gales induced by finite-state gamblers to define finite-state dimension [2]. By using feasible, computable, and finite-state predictors, we can also define feasible, computable, and finite-state unpredictability. The results mentioned in this introduction extend to the feasible, computable, and finite-state settings.

This note is organized as follows. We define gales and predictors in Section 2 and briefly review the definitions of the Hausdorff, computable, feasible, and finite-state dimensions. In Section 3 we define unpredictability with respect to general loss functions and prove the equivalence of logarithmic loss unpredictability and dimension. A full comparison of absolute loss and logarithmic loss unpredictability via the work of Fortnow and Lutz [6] is given in Section 4.

2 Gales, Predictors, and Dimension

In this section we define gales and predictors and briefly review Hausdorff dimension, computable dimension, feasible dimension, and finite-state dimension. The book by Falconer [3] is an excellent reference on Hausdorff dimension. Further details on feasible and computable dimension are available in Lutz's introductory paper [8]. Finite-state dimension was introduced by Dai, Lathrop, Lutz, and Mayordomo [2]. (The latter two references [2, 8] only address the binary alphabet but are easily extended to arbitrary finite alphabets.)

Throughout the paper we let Σ be a k-symbol alphabet where $2 \leq k < \infty$. We write Σ^* for the set of all finite strings over Σ and $\Sigma^{<n}$ for the set of strings of length less than n. The empty string is denoted by λ. The set of all infinite sequences over Σ is Σ^∞. For a string or sequence $\omega \in \Sigma^* \cup \Sigma^\infty$, we write $\omega[0..n-1]$ for the length n prefix of ω and $\omega[n]$ for the n^{th} symbol of ω.

For each $n \in \mathbb{N}$, let $\mathcal{A}_n$ be the set of all prefix sets $A \subseteq \Sigma^*$ such that $A \cap \Sigma^{<n} = \emptyset$. ($A$ is a prefix set if no element of A is a prefix of another element of A.) For each $X \subseteq \Sigma^\infty, s \in [0, \infty)$, and $n \in \mathbb{N}$, we define

$$H_n^s(X) = \inf \left\{ \sum_{w \in A} k^{-s|w|} \;\middle|\; A \in \mathcal{A}_n \text{ and } X \subseteq \bigcup_{w \in A} w \cdot \Sigma^\infty \right\}$$

and

$$H^s(X) = \lim_{n \to \infty} H_n^s(X).$$

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

$$\dim_{\mathrm{H}}(X) = \inf \{ s \in [0, \infty) \,|\, H^s(X) = 0 \}.$$

For each $X \subseteq \Sigma^\infty$, it holds that $0 \leq \dim_{\mathrm{H}}(X) \leq 1$.

2.1 Gales

Lutz [8] proved an alternative characterization of Hausdorff dimension using functions called gales.

Definition. Let $s \in [0, \infty)$. A function $d : \Sigma^* \to [0, \infty)$ is an *s-gale* if for all $w \in \{0, 1\}^*$,

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

Intuitively, a gale is viewed as a function betting on an unknown sequence. If w is a prefix of the sequence, then the capital of the gale after placing its first $|w|$ bets is given by $d(w)$. Assuming that w is a prefix of the sequence, the gale places bets on wa also being a prefix for each $a \in \Sigma$. The parameter s determines the fairness of the betting; as s decreases the betting is less fair. The goal of a gale is to bet successfully on sequences.

Definition. Let $s \in [0, \infty)$ and let d be an s-gale.

1. We say d *succeeds on* a sequence $S \in \Sigma^\infty$ if

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

2. The *success set* of d is

$$S^\infty[d] = \{ S \in \Sigma^\infty | d \text{ succeeds on } S \}.$$

Theorem 2.1. (Lutz [8]) *For any $X \subseteq \Sigma^\infty$,*

$$\dim_{\mathrm{H}}(X) = \inf \left\{ s \;\middle|\; \begin{array}{l} \textit{there exists an } s\textit{-gale } d \\ \textit{for which } X \subseteq S^\infty[d] \end{array} \right\}.$$

2.2 Predictors

Consider predicting the symbols of an unknown infinite sequence. Given an initial finite segment of the sequence, a predictor specifies a probability distribution over Σ. We may think of the probability that the algorithm assigns to each character as representing the predictor's confidence of that character occurring next in the sequence. Formally, we define a predictor as follows.

Definition. A function $\pi : \Sigma^* \times \Sigma \to [0, 1]$ is a *predictor* if for all $w \in \Sigma^*$, $\sum_{a\in\Sigma} \pi(w, a) = 1$.

Here we interpret $\pi(w, a)$ as the predictor π's estimation of the likelihood that the character immediately following the string w is a. There is a natural correspondence between predictors and gales. (An early reference for the following type of relationship between prediction and gambling is [1].)

Notation.

1. A predictor π *induces* for each $s \in [0, \infty)$ an s-gale $d_\pi^{(s)}$ defined by the recursion

$$d_\pi^{(s)}(\lambda) = 1$$

$$d_\pi^{(s)}(wa) = k^s d_\pi^{(s)}(w)\pi(w, a)$$

   for all $w \in \Sigma^*$ and $a \in \Sigma$; equivalently

$$d_\pi^{(s)}(w) = k^{s|w|} \prod_{i=0}^{|w|-1} \pi(w[0..i-1], w[i])$$

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

2. An s-gale d with $d(\lambda) = 1$ is *induced by* the predictor π_d defined by

$$\pi_d(w, a) = \begin{cases} k^{-s}\frac{d(wa)}{d(w)} & \text{if } d(w) \neq 0 \\ \frac{1}{k} & \text{otherwise} \end{cases}$$

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

2.3 Feasible and Computable Dimension

The characterization of Hausdorff dimension by gales motivated the following definitions of feasible and computable dimensions. We say that a real-valued function $f : \Sigma^* \to [0, \infty)$ is *computable* if there is a computable function $\hat{f} : \mathbb{N} \times \Sigma^* \to [0, \infty) \cap \mathbb{Q}$ satisfying $|\hat{f}(r, w) - f(w)| \leq 2^{-r}$ for all $w \in \Sigma^*$ and $r \in \mathbb{N}$. We say that f is *feasible* if there is a function $\hat{f}$ approximating f in the same way that is computable in time polynomial in $|w| + r$.

Definition. Let $X \subseteq \Sigma^\infty$.

1. The *feasible dimension* of X is

$$\dim_{\mathrm{p}}(X) = \inf\left\{ s \;\middle|\; \begin{array}{l} \text{there exists a feasible} \\ s\text{-gale } d \text{ for which } X \subseteq S^\infty[d] \end{array} \right\}.$$

2. The *computable dimension* of X is

$$\dim_{\mathrm{comp}}(X) = \inf\left\{ s \;\middle|\; \begin{array}{l}\text{there exists a computable}\\ s\text{-gale } d \text{ for which } X \subseteq S^\infty[d]\end{array} \right\}.$$

We say that a function $f : \Sigma^* \to [0,\infty) \cap \mathbb{Q}$ is *exactly feasible* (or *exactly computable*) if f itself is polynomial-time computable (or computable). The following result known as the Exact Computation Lemma shows that feasible and computable dimension can be equivalently defined using exactly feasible and exactly computable gales.

Lemma 2.2. (Lutz [8]) *Let k^s be rational.*

1. *For any feasible s-gale d there exists an exactly feasible s-gale d' such that $S^\infty[d] \subseteq S^\infty[d']$.*
2. *For any computable s-gale d there exists an exactly computable s-gale d' such that $S^\infty[d] \subseteq S^\infty[d']$.*

The following observation is simple but useful.

Observation 2.3. 1. *Let s be rational and π be a predictor. If π is feasible (or computable), then $d_\pi^{(s)}$ is feasible (or computable).*

2. *Let k^s be rational and d be an s-gale. If d is exactly feasible (or exactly computable), then π_d is exactly feasible (or exactly computable).*

2.4 Finite-State Dimension

Finite-state gamblers [4, 13] are used to define finite-state dimension.

Definition. A *finite-state gambler (FSG)* is a tuple $G = (Q, \delta, \beta, q_0)$ where Q is a nonempty, finite set of states, $\delta : Q \times \Sigma \to Q$ is the transition function, $\beta : Q \times \Sigma \to \mathbb{Q} \cap [0,1]$ is the betting function, and $q_0 \in Q$ is the initial state. The betting function satisfies $\sum_{a\in\Sigma} \beta(q,a) = 1$ for each $q \in Q$.

An FSG $G = (Q, \delta, \beta, q_0)$ defines a predictor π_G by

$$\pi_G(w,a) = \beta(\delta^*(w),a)$$

for all $w \in \Sigma^*$ and $a \in \Sigma$. Here $\delta^* : \Sigma^* \to Q$ is the standard extension of δ to strings defined by the recursion

$$\delta^*(\lambda) = q_0$$
$$\delta^*(wa) = \delta(\delta^*(w),a).$$

We say that a predictor π is *finite-state* if $\pi = \pi_G$ for some FSG G and that an s-gale d is *finite-state* if $d = d_\pi^{(s)}$ for some finite-state predictor π. Note that π_d is finite-state if d is finite-state.

Dai, Lathrop, Lutz, and Mayordomo [2] used finite-state gales to define finite-state dimension.

Definition. The finite-state dimension of a set $X \subseteq \Sigma^\infty$ is

$$\dim_{\mathrm{FS}}(X) = \inf\left\{ s \;\middle|\; \begin{array}{l}\text{there exists a finite-state}\\ s\text{-gale } d \text{ for which } X \subseteq S^\infty[d]\end{array} \right\}.$$

(Note: In [2], finite-state dimension was defined using multi-account finite state gamblers, and the above single-account definition was shown to be equivalent.)

From Theorem 2.1 and the definitions of the computable, feasible, and finite-state dimensions, we observe that

$$0 \leq \dim_{\mathrm{H}}(X) \leq \dim_{\mathrm{comp}}(X) \leq \dim_{\mathrm{p}}(X) \leq \dim_{\mathrm{FS}}(X) \leq 1$$

for all $X \subseteq \Sigma^\infty$.

2.5 Unified Notation

We now introduce some unified notation to simplify the statement of our results. For this, we define the following sets.

$$\begin{aligned} \mathrm{all} &= \{\text{all gales and predictors}\} \\ \mathrm{comp} &= \{\text{all computable gales and predictors}\} \\ \mathrm{p} &= \{\text{all feasible gales and predictors}\} \\ \mathrm{FS} &= \{\text{all finite-state gales and predictors}\} \end{aligned}$$

For any $\Delta \in \{\mathrm{all}, \mathrm{comp}, \mathrm{p}, \mathrm{FS}\}$ and $X \subseteq \Sigma^\infty$, we then define

$$\dim_\Delta(X) = \inf\left\{ s \,\middle|\, \begin{array}{l} \text{there exists an } s\text{-gale } d \in \Delta \\ \text{for which } X \subseteq S^\infty[d] \end{array} \right\}.$$

Using this notation, $\dim_{\mathrm{all}}$ represents Hausdorff dimension. The notation for the computable, feasible, and finite-state dimensions remains the same.

3 Unpredictability and Dimension

In this section we use a standard prediction model to define the unpredictability of sets of sequences with respect to a loss function. Under the logarithmic loss we obtain an equivalent definition of Hausdorff dimension.

In judging the performance of a predictor, it is instructive to consider its "loss" on individual symbols of the sequence. One natural way to do this is to measure the absolute loss. If the probability that the predictor assigned to the correct symbol is p, then we assign the *absolute loss*

$$loss^{\mathrm{abs}} : [0,1] \to [0,1]$$

$$loss^{\mathrm{abs}}(p) = 1 - p$$

to that prediction. That is, the loss is equal to the probability that the predictor did not assign to the correct symbol. Another common, but more severe, measure of loss is the *logarithmic loss*

$$loss^{\log} : [0,1] \to [0,\infty]$$

$$loss^{\log}(p) = \log_k \frac{1}{p}.$$

(Recall that $k = |\Sigma|$.) If $p = 1$, then both the absolute and logarithmic losses are 0. As p approaches 0, the logarithmic loss becomes infinite, while the absolute loss only goes to 1.

The cumulative loss of a predictor on a finite string is the sum of the losses it incurs while predicting the individual symbols. We now formally define this as well as the asymptotic loss rate on infinite sequences and sets of sequences. Here $loss^\tau$ may be absolute or logarithmic loss or any other loss function.

Definition. Let π be a predictor and let $loss^\tau : [0,1] \to [0,\infty]$ be a loss function.

1. The τ-*cumulative loss* of π on a string $w \in \Sigma^*$ is

$$\mathcal{L}_\pi^\tau(w) = \sum_{i=0}^{|w|-1} loss^\tau(\pi(w[0..i-1], w[i])).$$

2. The τ-*loss rate* of π on a sequence $S \in \Sigma^\infty$ is

$$\mathcal{L}_\pi^\tau(S) = \liminf_{n\to\infty} \frac{\mathcal{L}_\pi^\tau(S[0..n-1])}{n}.$$

3. The τ-*loss rate* of π on a set $X \subseteq \Sigma^\infty$ is

$$\mathcal{L}_\pi^\tau(X) = \sup_{S\in X} \mathcal{L}_\pi^\tau(S).$$

The unpredictability of a set of sequences is defined as the infimum of the loss rates that a predictor can guarantee on the set.

Definition. Let $loss^\tau : [0,1] \to [0,\infty]$ be a loss function and let $\Delta \in \{\text{all}, \text{comp}, \text{p}, \text{FS}\}$. For any $X \subseteq \Sigma^\infty$, the Δ-τ-*unpredictability* of X is

$$\text{unpred}_\Delta^\tau(X) = \inf\{\mathcal{L}_\pi^\tau(X) | \pi \text{ is a predictor in } \Delta\}.$$

Observe that

$$0 \le \text{unpred}_{\text{all}}^{\log}(X) \le \text{unpred}_{\text{comp}}^{\log}(X) \le \text{unpred}_{\text{p}}^{\log}(X) \le \text{unpred}_{\text{FS}}^{\log}(X) \le 1$$

and

$$0 \le \text{unpred}_{\text{all}}^{\text{abs}}(X) \le \text{unpred}_{\text{comp}}^{\text{abs}}(X) \le \text{unpred}_{\text{p}}^{\text{abs}}(X) \le \text{unpred}_{\text{FS}}^{\text{abs}}(X) \le \frac{k-1}{k}$$

for all $X \subseteq \Sigma^\infty$. Here the upper bounds of 1 and $\frac{k-1}{k}$ are witnessed by the predictor π defined by $\pi(w,a) = \frac{1}{k}$ for all $w \in \Sigma^*$ and $a \in \Sigma$.

3.1 Logarithmic Loss and Dimension

We now show that logarithmic loss unpredictability equals dimension.

Theorem 3.1. *For any $X \subseteq \Sigma^\infty$ and $\Delta \in \{\text{all}, \text{comp}, \text{p}, \text{FS}\}$,*

$$\dim_\Delta(X) = \text{unpred}_\Delta^{\log}(X).$$

Proof. Let k^s be rational and assume that d is an s-gale succeeding on X. Assume without loss of generality that $d(\lambda) = 1$. Let $\pi_d : \Sigma^* \times \Sigma \to [0,1]$ be the predictor inducing d as defined in Section 2. For any $w \in \Sigma^*$ with $d(w) > 0$,

$$\begin{aligned}
\mathcal{L}^{\log}_{\pi_d}(w) &= \sum_{i=0}^{|w|-1} \log_k \frac{1}{\pi_d(w[0..i-1], w[i])} \\
&= -\log_k \prod_{i=0}^{|w|-1} \pi_d(w[0..i-1], w[i]) \\
&= -\log_k k^{-s|w|} d(w) \\
&= s|w| - \log_k d(w).
\end{aligned}$$

Let $S \in S^\infty[d]$. Then there exist infinitely many $n \in \mathbb{N}$ such that $d(S[0..n-1]) \geq 1$, and for each of these n we have

$$\begin{aligned}
\frac{\mathcal{L}^{\log}_{\pi_d}(S[0..n-1])}{n} &= \frac{sn - \log_k d(S[0..n-1])}{n} \\
&\leq \frac{sn - \log_k 1}{n} \\
&= s.
\end{aligned}$$

Therefore $\mathcal{L}^{\log}_{\pi_d}(S) \leq s$, so this establishes that $\mathrm{unpred}^{\log}_{\mathrm{all}}(X) \leq \mathcal{L}^{\log}_{\pi_d}(X) \leq s$. If d is exactly computable, then the predictor π_d is exactly computable, so we have $\mathrm{unpred}^{\log}_{\mathrm{comp}}(X) \leq \mathcal{L}^{\log}_{\pi_d}(X) \leq s$. Similarly, if d is exactly feasible or d is finite-state, we have $\mathrm{unpred}^{\log}_{\mathrm{p}}(X) \leq s$ or $\mathrm{unpred}^{\log}_{\mathrm{FS}}(X) \leq s$. For each Δ, we have established that $\mathrm{unpred}^{\log}_{\Delta}(X) \leq s$ for all $s > \dim_\Delta(X)$ such that $k^s \in \mathbb{Q}$. By density of the set $\{s | k^s \in \mathbb{Q}\}$, it follows that $\mathrm{unpred}^{\log}_{\Delta}(X) \leq \dim_\Delta(X)$ for each Δ.

Now let $s > t$ be rational, and assume that π is a predictor for which $\mathcal{L}^{\log}_{\pi}(X) < t$. Let $d^{(s)}_\pi$ be the s-gale induced by π as defined in Section 2. Let $S \in X$. Then there exist infinitely many $n \in \mathbb{N}$ such that

$$\frac{\mathcal{L}^{\log}_{\pi}(S[0..n-1])}{n} \leq t,$$

and for each of these n we have

$$\begin{aligned}
\log_k d^{(s)}_\pi(S[0..n-1]) &= sn + \sum_{i=0}^{n-1} \log_k \pi(w[0..i-1], w[i]) \\
&= sn - \mathcal{L}^{\log}_{\pi}(S[0..n-1]) \\
&\geq sn - tn \\
&= (s-t)n,
\end{aligned}$$

so it follows that $S \in S^\infty[d^{(s)}_\pi]$ and $X \subseteq S^\infty[d^{(s)}_\pi]$. Therefore, $\dim_{\mathrm{all}}(X) \leq s$. If π is feasible (or computable or finite-state), then $d^{(s)}_\pi$ is feasible (or computable or finite-state), so $\dim_{\mathrm{p}}(X) \leq s$ (or $\dim_{\mathrm{comp}}(X) \leq s$ or $\dim_{\mathrm{FS}}(X) \leq s$). For each Δ, we now have $\dim_\Delta(X) \leq s$ for all rational $s > \mathrm{unpred}^{\log}_{\Delta}(X)$. By density of the rationals, we then have $\dim_\Delta(X) \leq \mathrm{unpred}^{\log}_{\Delta}(X)$ for each Δ. □

4 Absolute Loss versus Logarithmic Loss

It is immediate from the definitions that the predictability of Fortnow and Lutz [6] has the following relationship with absolute loss unpredictability.

Proposition 4.1. *For any* $X \subseteq \Sigma^\infty$ *and* $\Delta \in \{\text{all}, \text{comp}, \text{p}, \text{FS}\}$,

$$\text{pred}_\Delta(X) = 1 - \text{unpred}^{\text{abs}}_\Delta(X).$$

(Fortnow and Lutz only defined pred_Δ for the case $\Delta = \text{p}$. Their definition readily extends to the other Δ.)

Fortnow and Lutz proved very tight bounds between predictability and dimension. (Feder, Merhav, and Gutman [5] previously obtained related results about the finite-state predictability of individual sequences.) To state their results we need the following technical definitions.

1. The *k-adic segmented self-information function* $\overline{\mathcal{I}_k} : [\frac{1}{k}, 1] \to [0, 1]$ is defined by setting $\overline{\mathcal{I}_k}(\frac{1}{j}) = \log_k j$ for $1 \leq j \leq k$ and interpolating linearly between these points.

2. The *k-adic maximum entropy function* $\mathcal{H}_k : [0, 1] \to [0, 1]$ is defined by

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

Theorem 4.2. (Fortnow and Lutz [6]) *For every set* $X \subseteq \Sigma^\infty$ *and* $\Delta \in \{\text{all}, \text{comp}, \text{p}, \text{FS}\}$,

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

(Fortnow and Lutz only presented Theorem 4.2 for the case $\Delta = \text{p}$. Their proof also works for $\Delta \in \{\text{comp}, \text{all}\}$ and can be extended for the case $\Delta = \text{FS}$.)

Combining Theorem 3.1, Proposition 4.1, and Theorem 4.2, we have the following relationship between absolute loss prediction and logarithmic loss prediction.

Theorem 4.3. *For every set* $X \subseteq \Sigma^\infty$ *and* $\Delta \in \{\text{all}, \text{comp}, \text{p}, \text{FS}\}$,

$$\overline{\mathcal{I}_k}(1 - \text{unpred}^{\text{abs}}_\Delta(X)) \leq \text{unpred}^{\log}_\Delta(X) \leq \mathcal{H}_k(1 - \text{unpred}^{\text{abs}}_\Delta(X)).$$

In particular, if Σ *is the binary alphabet, then*

$$2 \cdot \text{unpred}^{\text{abs}}_\Delta(X) \leq \text{unpred}^{\log}_\Delta(X) \leq \mathcal{H}(\text{unpred}^{\text{abs}}_\Delta(X)),$$

where $\mathcal{H}$ *is the binary entropy function.*

Furthermore, Fortnow and Lutz stated that the bounds in Theorem 4.2 are tight for the case $\Delta = \text{p}$ in the strong sense that for any $\alpha \in [\frac{1}{k}, 1]$ and $\beta \in [\overline{\mathcal{I}_k}(\alpha), \mathcal{H}_k(\alpha)]$, there is a set $X \subseteq \Sigma^\infty$ with $\text{pred}_\Delta(X) = \alpha$ and $\dim_\Delta(X) = \beta$. This tightness result also holds for $\Delta \in \{\text{comp}, \text{all}\}$, and holds for $\Delta = \text{FS}$ with a dense subset of such α and β. For each Δ, it is immediate that the bounds in Theorem 4.3 are tight in the same way.

References

[1] T. Cover. Universal gambling schemes and the complexity measures of Kolmogorov and Chaitin. Technical Report 12, Stanford University Department of Statistics, October 1974.

[2] J. J. Dai, J. I. Lathrop, J. H. Lutz, and E. Mayordomo. Finite-state dimension. *Theoretical Computer Science*. To appear.

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

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

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

[6] L. Fortnow and J. H. Lutz. Prediction and dimension. In *Proceedings of the 15th Annual Conference on Computational Learning Theory*, pages 380–395, 2002.

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

[8] J. H. Lutz. Dimension in complexity classes. *SIAM Journal on Computing*. To appear.

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

[10] B. Ryabko. The nonprobabilistic approach to learning the best prediction. Technical Report TR02-011, Electronic Colloquium on Computational Complexity, 2002.

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

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

[13] C. P. Schnorr and H. Stimm. Endliche automaten und zufallsfolgen. *Acta Informatica*, 1:345–359, 1972.

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