# Predictive Complexity for Games with Finite Outcome Spaces *

Yuri Kalnishkan[†]

### Abstract

Predictive complexity is a generalisation of Kolmogorov complexity motivated by an on-line prediction scenario. It quantifies "unpredictability" of a sequence in a particular prediction environment. This paper surveys key results on predictive complexity for games with finitely many outcomes. The issues of existence, non-existence, uniqueness, and linear inequalities are covered.

## 1   Introduction

We consider the following on-line learning scenario: given a sequence of previous outcomes $x_1, x_2, \ldots, x_{n-1}$, a prediction strategy is required to output a prediction $\gamma_n$ for the next outcome $x_n$. A loss function $\lambda(\gamma, x)$ is used to measure the discrepancy between predictions and actual outcomes. The performance of the strategy is measured by the cumulative loss $\sum_{i=1}^{n} \lambda(\gamma_i, x_i)$ and the learner's goal is to make the loss as small as possible.

Different aspects of this prediction problem have been extensively studied; a key feature of the approach of this paper is that no mechanism generating the outcomes is postulated. This approach is closely related to (and shares concepts and methods with) prediction with expert advice; see [CBL06] for an overview.

We want to define a measure of "(un)predictability" of a string. "Simple" regular strings are easily predictable with small cumulative loss and "complex" ones defy prediction algorithms. However implementing this idea is not straightforward.

One may want to define complexity of a sequence as the loss of an optimal prediction strategy. However in natural cases there is no such strategy: every strategy is greatly outperformed by some other strategy on some sequences (see Remark 1 below).

In [KVV14] the difficulty is resolved by considering asymptotic loss per element. This allows us to define complexity of infinite sequences and infinite sets of finite sequences, but not of individual finite sequences.

The approach suggested in [VW98] is to extend the class of strategies to some superstrategies, which often have optimal elements. This results in the notion of predictive complexity of a sequence. Predictive complexity is defined up to a constant (or a function sublinear in the length of the sequence) and is conceptually similar to Kolmogorov complexity, which quantifies "incompressibility" of a sequence (Kolmogorov complexity is surveyed in Chap. 3 of this book). Predictive complexity for a particular environment, the logarithmic complexity, coincides with a version of Kolmogorov complexity, negative logarithm of Levin's apriori semimeasure (see Sect. 3 for references).

Section 3 contains proofs for the existence of basic predictive complexity for mixable games after [VW98] and weak complexity up to $\sqrt{T}$ for convex bounded games essentially extending the argument from [KV08].

After Sect. 4 describes the connections between predictive complexity and generalised entropy, Sect. 5 surveys negative results on the existence of complexity (after [KVV04b] and [KVV04a]) and Sect. 6 proves the uniqueness theorem from [KVV04b]. Section 7 discusses linear inequalities between different predictive complexities after [Kal02]. In Sect. 8 some more references are given.

---

# 2   Preliminaries

We denote the set of positive integers $\{1, 2, \ldots\}$ by $\mathbb{N}$.

## 2.1   Games and Losses

The concept of a game describes the prediction environment. A *game* $\mathfrak{G}$ is a triple $\langle \Omega, \Gamma, \lambda \rangle$, where $\Omega$ is an *outcome space*, $\Gamma$ is a *prediction space*, and $\lambda : \Omega \times \Gamma \to [0, +\infty]$ is a *loss function*.

We assume that $\Omega$ is a finite set of cardinality $M < +\infty$, i.e.,

$$\Omega = \left\{ \omega^{(0)}, \omega^{(1)}, \ldots, \omega^{(M-1)} \right\} \; .$$

If $M = 2$, then $\Omega$ may be identified with $\mathbb{B} = \{0, 1\}$; we will call games with two outcomes *binary*.

Bold letters $\boldsymbol{x}$, $\boldsymbol{y}$ etc. are used to refer to finite sequences of outcomes. By $|\boldsymbol{x}|$ we denote the length of a finite sequence $\boldsymbol{x}$, i.e., the number of elements in it. The set of sequences of length $n$ is denoted by $\Omega^n$, $n = 0, 1, 2, \ldots$. We denote the set of all finite sequences of elements of $\Omega$ by $\Omega^*$. We will also be using the notation $\sharp_i \boldsymbol{x}$ for the number of $\omega^{(i)}$s among elements of $\boldsymbol{x}$. Clearly, $\sum_{i=0}^{M-1} \sharp_i \boldsymbol{x} = |\boldsymbol{x}|$ for any finite sequence $\boldsymbol{x}$.

We assume that $\Gamma \subseteq \mathbb{R}^K$ is compact and $\lambda$ is continuous w.r.t. the topology of the extended half-line $[0, +\infty]$. We treat $\Omega$ as a discrete space and thus continuity of $\lambda$ in two arguments is the same as continuity in the second argument.

In order to take some important games into account we must allow $\lambda$ to attain the value $+\infty$. However we assume that the set $\Gamma_{\mathrm{fin}} = \{\gamma \in \Gamma \mid \max_{\omega \in \Omega} \lambda(\omega, \gamma) < +\infty\}$ is dense in $\Gamma$. In other words, every prediction $\gamma_0$ leading to infinite loss can be approximated by predictions giving finite losses.

The following are examples of *binary games* with $\Omega = \mathbb{B}$ and $\Gamma = [0, 1]$: the *square-loss* game with the loss function $\lambda(\omega, \gamma) = (\omega - \gamma)^2$, the *absolute-loss* game with the loss function $\lambda(\omega, \gamma) = |\omega - \gamma|$, and the *logarithmic* game with

$$\lambda(\omega, \gamma) = \left\{ \begin{array}{lll} -\log_2(1 - \gamma) & \text{if} & \omega = 0 \\ -\log_2 \gamma & \text{if} & \omega = 1 \end{array} \right. \; .$$

A *prediction strategy* $\mathfrak{S} : \Omega^* \to \Gamma$ maps a finite sequence of outcomes to a prediction. We say that on a finite sequence $\boldsymbol{x} = \omega_1 \omega_2 \ldots \omega_T \in \Omega^T$ the strategy $\mathfrak{S}$ suffers loss

$$\mathrm{Loss}_{\mathfrak{S}}^{\mathfrak{G}}(\boldsymbol{x}) = \sum_{t=1}^{T} \lambda(\mathfrak{S}(\omega_1 \omega_2 \ldots \omega_{t-1}), \omega_t) \; .$$

The first term in this sum involves $\mathfrak{S}(\Lambda)$, where $\Lambda$ is the sequence of length 0. If $\boldsymbol{x} = \Lambda$, the sum has no terms and we assume $\mathrm{Loss}_{\mathfrak{A}}^{\mathfrak{G}}(\Lambda) = 0$. The upper index $\mathfrak{G}$ will be omitted if it is clear from the context.

## 2.2   Superpredictions

Given a prediction $\gamma \in \Gamma$ one can consider a function $\lambda(\gamma, \cdot) : \Omega \to [0, +\infty]$. If $\Omega$ is finite, this function can be identified with a point from $[0, +\infty]^{|\Omega|}$.

A *generalised prediction* is a function $g : \Omega \to [0, +\infty]$. *Superpredictions* are generalised predictions minorated by predictions; in other terms, $s : \Omega \to [0, +\infty]$ is a superprediction if there is a prediction $\gamma \in \Gamma$ such that for all $\omega \in \Omega$ we have $s(\omega) \geq \lambda(\gamma, \omega)$.

For $x = (x_1, x_2, \ldots, x_K), y = (y_1, y_2, \ldots, y_K) \in \mathbb{R}^K$ we will be writing $x < y$ if the inequality holds coordinate-wise, i.e., $x_k < y_k$ for $k = 1, 2, \ldots, K$; the same applies to inequalities $x > y$, $x \leq y$, and $x \geq y$. Thus $s \in [0, +\infty]^{|\Omega|}$ is a superprediction if there is $p = \left( \lambda\left(\gamma, \omega^{(0)}\right), \lambda\left(\gamma, \omega^{(1)}\right), \ldots, \lambda\left(\gamma, \omega^{(M-1)}\right) \right)$, where $\gamma \in \Gamma$, such that $s \geq p$.

## 2.3   Computability

We use the computational model from [Ko91].

A *dyadic* rational number is a number of the form $n/2^m$, where $n$ and $m$ are integers. We call a triple $r = \langle b, \boldsymbol{x}, \boldsymbol{y} \rangle$, where $b \in \mathbb{B}$ is a bit and $\boldsymbol{x} = (x_1 x_2 \ldots x_u), \boldsymbol{y} = (y_1 y_2 \ldots y_v) \in \mathbb{B}^*$ are binary strings, a *representation of a dyadic number* $d$ if $x_1 = 1$ and

$$d = s \left( \sum_{i=0}^{u-1} x_{u-i} 2^i + \sum_{i=1}^{v} y_i 2^{-i} \right) \ , \tag{1}$$

where $s = 1$ if $b = 1$ and $s = -1$ if $b = 0$. Intuitively, $b$ represents the sign of $d$ and $\boldsymbol{x}.\boldsymbol{y}$ is the finite binary expansion of $|d|$. We will denote the set of correctly formed triples by $\mathbb{D}$. Let $\mathrm{d}$ map correctly formed triples into dyadic numbers according to (1). We will call $v$ the *precision* of the triple $r$ and write $v = \mathrm{prec}(r)$.

For every $x \in \mathbb{R}$ let $\mathrm{CF}_x$ be the set of sequences of triples, i.e., functions $\phi$ from positive integers to representations of dyadic numbers, such that $\mathrm{prec}(\phi(m)) = m$ and $|\mathrm{d}(\phi(m)) - x| \leq 2^{-m}$ for all $m = 1, 2, \ldots$. Any element of $\mathrm{CF}_x$ can be thought of as a representation of $x$. A number $x \in \mathbb{R}$ is *computable* if $\mathrm{CF}_x$ contains a computable function $\phi$.

A point $x = (x_1, x_2, \ldots, x_K) \in \mathbb{R}^K$ is computable if all its coordinates $x_1, x_2, \ldots, x_K$ are computable.

Let $M$ be a finite set. The elements of $M$ can be identified with letters in a finite alphabet and passed to a Turing machine directly. A function $f : M^* \to \mathbb{R}$ is *computable* if there is a Turing machine that given a finite string $\boldsymbol{x} = x_1 x_2 \ldots x_n \in M^*$ and non-negative integer precision $m$ outputs a representation $r$ of a dyadic number such that $\mathrm{prec}(r) = m$ and $|f(\boldsymbol{x}) - \mathrm{d}(r)| \leq 2^{-m}$. In other words, for every $\boldsymbol{x} \in M^*$ the machine calculates a function from $\mathrm{CF}_{f(\boldsymbol{x})}$. A function $f = (f_1, f_2, \ldots, f_K) : M^* \to \mathbb{R}^K$ is computable if all its components $f_1, f_2, \ldots, f_K$ are computable.

A function $f : M \to \mathbb{R}$, where $M \subseteq \mathbb{R}$, is computable if there is an oracle Turing machine that given a non-negative integer precision $m$ and an oracle evaluating some $\phi \in \mathrm{CF}_x$ outputs a representation $r$ of a dyadic number such that $\mathrm{prec}(r) = m$ and $|f(x) - \mathrm{d}(r)| \leq 2^{-m}$. It is easy to see that the class of computable functions is closed under composition.

Computable functions on $M \subseteq \mathbb{R}^K$ and $M \subseteq \mathbb{R}^*$ to $\mathbb{R}$ and $\mathbb{R}^N$ are defined in a similar fashion. For convenience of notation we extend $\mathrm{d}$ to $\mathbb{D}^K$ in the natural way and refer to elements of $\mathbb{D}^K$ as dyadic vectors. By the absolute value $|d|$ of $d = (d_1, d_2, \ldots, d_K)$ we mean the 1-norm $\sum_{k=1}^{K} |d_k|$.

## 2.4 Computable Games

In this section we define a *computable game* $\mathfrak{G} = \langle \Omega, \Gamma, \lambda \rangle$.

We assume that $\Omega$ is a finite set and its elements can be passed to a Turing machine directly. Let $\Gamma$ be a subset of $\mathbb{R}^K$. We require that the function $e^{-\lambda(\gamma, \omega)} : \Gamma \times \Omega \to [0, 1]$ should be computable[1]. We do not require computability of $\lambda$ itself so as not to exclude games with unbounded loss functions (such as the logarithmic game) from consideration.

We also require that dyadic vectors are dense in $\Gamma$ and moreover there is an effective sequence of dyadic vectors $\gamma_i \in \Omega$, $i \in \mathbb{N}$, that is dense in $\Omega$.

# 3 Definition and Existence

## 3.1 Superloss Processes

Consider a game $\mathfrak{G} = \langle \Omega, \Gamma, \lambda \rangle$. A *process* is a function $L : \Omega^* \to [0, +\infty]$. A process $L$ is a *superloss process* if for every $\boldsymbol{x} \in \Omega^*$ there is $\gamma \in \Gamma$ satisfying

$$L(\boldsymbol{x}\omega) - L(\boldsymbol{x}) \geq \lambda(\gamma, \omega) \tag{2}$$

for all $\omega \in \Omega$. In other terms, the difference $L(\boldsymbol{x}\omega) - L(\boldsymbol{x})$ considered as a function of $\omega$ is a superprediction.

A process $L : \Omega^* \to [0, +\infty]$ is *upper semi-computable* if its epigraph is enumerable. Let us expand this definition. If $\Omega$ is finite, then $L : \Omega^* \to [0, +\infty]$ is upper semi-computable if and only if the set $\{(\boldsymbol{x}, d) \in \Omega^* \times \mathbb{D} \mid \mathrm{d}(d) > L(\boldsymbol{x})\}$ is enumerable, i.e., coincides with the image of a computable function

---

[1]According to Theorem 2.31 in [Ko91] the domain of a partial computable function is open; therefore any machine calculating $e^{-\lambda(\gamma, \omega)}$ on $\Gamma$ for a fixed $\omega$ will also calculate an extension of this function to some open set.

on $\mathbb{N}$. It is easy to see that an equivalent requirement is the existence of a partial computable function $f : \Omega^* \times \mathbb{N} \to \mathbb{D}$ such that $L(\boldsymbol{x}) = \inf_{n \in \mathbb{N}} \mathrm{d}(f(\boldsymbol{x}, n))$, where undefined values of $f$ are taken to be $+\infty$.

Yet another (arguably more intuitive) equivalent definition may be given. A process $L : \Omega^* \to [0, +\infty]$ is upper semi-computable if there is a total computable function $f : \Omega^* \times \mathbb{N} \to \overline{\mathbb{D}}$, where $\overline{\mathbb{D}}$ is $\mathbb{D}$ appended with a symbol for $+\infty$, such that $\mathrm{d}(f(\boldsymbol{x}, n)) \geq \mathrm{d}(f(\boldsymbol{x}, n+1))$ for all $\boldsymbol{x} \in \Omega^*$ and $n \in \mathbb{N}$ and $L(\boldsymbol{x}) = \lim_{n \to \infty} \mathrm{d}(f(\boldsymbol{x}, n))$. This definition implies that an upper semi-computable process is the monotonous limit of computable upper bounds.

The concept of a superloss process generalises the concept of a *loss process*. A loss process $L : \Omega^* \to [0, +\infty]$ is the loss of a prediction strategy, i.e.,

$$L(\omega_1 \omega_2 \ldots \omega_T) = \sum_{t=1}^{T} \lambda(\mathfrak{S}(\omega_1 \omega_2 \ldots \omega_{t-1}), \omega_t)$$

for some strategy $\mathfrak{S}$. Clearly, every loss process is a superloss process; for a loss process, inequality (2) holds, because $\gamma = \mathfrak{S}(\boldsymbol{x})$ turns it into an equality.

Let us show that if a strategy $\mathfrak{S}$ is computable, then the corresponding loss process $L$ is upper semi-computable. We will define the partial function $f(\boldsymbol{x}, n)$ such that $L(\boldsymbol{x}) = \inf_{n \in \mathbb{N}} \mathrm{d}(f(\boldsymbol{x}, n))$ as the result of the following calculation. Take $\boldsymbol{x} = \omega_1 \omega_2 \ldots \omega_T$ and $n \in \mathbb{N}$. Let us calculate the values of

$$e^{-\lambda(\gamma_t, \omega_t)} = e^{-\lambda(\mathfrak{S}(\omega_1 \omega_2 \ldots \omega_{t-1}), \omega_t)} \quad ,$$

$t = 1, 2, \ldots, T$, accurate to within $2^{-n}$. If any of the resulting values $l_t$ is less than or equal to $2^{-n}$, the calculation diverges (e.g., falls into an infinite loop) and outputs nothing. Otherwise we calculate the sum $-\sum_{t=1}^{T} \ln l_t$ with the maximum possible accuracy given that $l_t$ are accurate to within $2^{-n}$ (we allow extra loss of accuracy of $2^{-n}$ to cater for non-dyadic results). Let the result $l$ be accurate to within a dyadic rational $\varepsilon$. We then output a dyadic representation of $l + \varepsilon$.

For most natural games the set of upper semi-computable superloss processes is enumerable, i.e., there is a total computable function on $\mathbb{N}$ outputting programs enumerating epigraphs of superloss processes such that for every upper semi-computable processes at least one program enumerating its epigraph will be output. Equivalently, there is a partial computable function $f(\boldsymbol{x}, n, i) : \Omega^* \times \mathbb{N} \times \mathbb{N} \to \mathbb{D}$ such that for every $i \in \mathbb{N}$ the process $k_i(\boldsymbol{x}) = \inf_{n \in \mathbb{N}} f(\boldsymbol{x}, n, i)$ is a superloss process and every upper semi-computable superloss process coincides with some $k_i$. In the Appendix we formulate a sufficient condition for the set of upper semi-computable superloss processes to be enumerable.

## 3.2 Basic Predictive Complexity

If for a game $\mathfrak{G} = \langle \Omega, \Gamma, \lambda \rangle$ there is an upper semi-computable superloss process minimal up to a constant, we call it *(basic) predictive complexity*. In other terms, an upper semi-computable superloss process $\mathcal{K}$ is called basic predictive complexity if for any other upper semi-computable superloss process $L$ there is a constant $C$ such that

$$\mathcal{K}(\boldsymbol{x}) \leq L(\boldsymbol{x}) + C$$

for all $\boldsymbol{x} \in \Omega^*$. One can say that basic predictive complexity is defined up to a constant (cf. the situation with Kolmogorov complexity described in Sect. 3.3 of this book).

A fundamental theorem proven in [VW98] states that predictive complexity exists for mixable games.

Let $|\Omega| = M$ and $S \subseteq [0, +\infty]^M$ be the set of superpredictions w.r.t. $\mathfrak{G}$. For every $\eta > 0$ consider the transformation $\mathfrak{B}_\eta : [0, +\infty]^M \to [0, 1]^M$ given by

$$\mathfrak{B}_\eta(x_0, x_1, \ldots, x_{M-1}) = (e^{-\eta x_0}, e^{-\eta x_1}, \ldots, e^{-\eta x_{M-1}}) \quad .$$

The game $\mathfrak{G}$ is called $\eta$-*mixable* for $\eta > 0$ if the set $\mathfrak{B}_\eta(S)$ is convex. The game $\mathfrak{G}$ is called *mixable* if there is $\eta > 0$ such that the game is $\eta$-mixable.

It is easy to check (e.g., by differentiation) that the square-loss game is mixable for $\eta \leq 2$ and the logarithmic-loss game is mixable for $\eta \leq \ln 2$ while the absolute-loss game is not mixable.

**Theorem 1.** *If a game $\mathfrak{G}$ is mixable and there is an effective enumeration of upper semi-computable superloss processes w.r.t. $\mathfrak{G}$, then there is basic predictive complexity w.r.t. $\mathfrak{G}$.*

*Proof.* Let $k_i$, $i \in \mathbb{N}$, be an enumeration of upper semi-computable superloss processes and let $f(\boldsymbol{x}, n, i) :$ $\Omega^* \times \mathbb{N} \times \mathbb{N} \to \mathbb{D}$ be a partial computable function such that $k_i(\boldsymbol{x}) = \inf_n f(\boldsymbol{x}, n, i)$.

Let $\mathfrak{G}$ be $\eta$-mixable. Put

$$\mathcal{K}(\boldsymbol{x}) = -\frac{1}{\eta} \ln \sum_{i=1}^{\infty} \frac{1}{2^i} e^{-\eta k_i(\boldsymbol{x})} \ . \tag{3}$$

Let us show that it is an upper semi-computable superloss process.

We get

$$\mathcal{K}(\boldsymbol{x}\omega) - \mathcal{K}(\boldsymbol{x}) = -\frac{1}{\eta} \ln \sum_{i=0}^{\infty} \frac{e^{-\eta k_i(\boldsymbol{x})}/2^i}{\sum_{j=1}^{\infty} e^{-\eta k_j(\boldsymbol{x})}/2^j} e^{-\eta(k_i(\boldsymbol{x}\omega) - k_i(\boldsymbol{x}))}$$

$$= -\frac{1}{\eta} \sum_{i=0}^{\infty} p_i e^{-\eta(k_i(\boldsymbol{x}\omega) - k_i(\boldsymbol{x}))} \ ,$$

where $\sum_{i=1}^{\infty} p_i = 1$. All the series in the formula converge absolutely because the exponents are upper bounded by 1.

Let $S$ be the set of superpredictions w.r.t. $\mathfrak{G}$. The points

$$s_i = \left( k_i \left( \boldsymbol{x}\omega^{(0)} \right) - k_i(\boldsymbol{x}), k_i \left( \boldsymbol{x}\omega^{(1)} \right) - k_i(\boldsymbol{x}), \dots, k_i \left( \boldsymbol{x}\omega^{(M-1)} \right) - k_i(\boldsymbol{x}) \right)$$

are superpredictions, $i \in \mathbb{N}$. Since the set $\mathfrak{B}_\eta(S)$ is convex, the infinite convex combination $\sum_{i=1}^{\infty} p_i \mathfrak{B}_\eta(s_i)$ belongs to $\mathfrak{B}_\eta(S)$ (see [BG54], Theorem 2.4.1). Thus the inverse image $\mathfrak{B}_\eta^{-1} \left( \sum_{i=1}^{\infty} p_i \mathfrak{B}_\eta(s_i) \right)$ is a superprediction and $\mathcal{K}$ is a superloss process.

In order to see that $\mathcal{K}$ is upper semi-computable, note that by truncating the infinite sum in (3) we get an upper bound on $\mathcal{K}$. Then upper bounds on $k_i$ will lead to upper bounds on $\mathcal{K}$.

Finally if we drop all terms in (3) except for one, we get

$$\mathcal{K}(\boldsymbol{x}) \leq k_i(\boldsymbol{x}) + \frac{i \ln 2}{\eta}$$

and thus $\mathcal{K}$ is minimal up to a constant. $\hfill\square$ $\hfill\square$

Since the square-loss and logarithmic games are mixable, there are basic predictive complexities $\mathcal{K}^{\mathrm{sq}}$ and $\mathcal{K}^{\mathrm{log}}$ w.r.t. them. The logarithmic game coincides with a variant of Kolmogorov complexity known as the negative logarithm of Levin's apriori semimeasure and denoted by KM (see [ZL70] and [LV08], Definition 4.5.8).

**Remark 1.** *One may wonder why the construction based on (3) does not lead to the existence of an optimal strategy. Let us show first that in natural cases there are no optimal strategies.*

*We will consider an example of the square-loss game. Pick a computable strategy $\mathfrak{S} : \{0,1\}^* \to [0,1]$. We will build a computable sequence $\omega_1, \omega_2, \dots$ and a computable strategy $\mathfrak{S}'$ such that $\mathfrak{S}'$ greatly outperforms $\mathfrak{S}$ on the sequence. The construction is by induction. Suppose that $\boldsymbol{x}_t = \omega_1 \omega_2 \dots \omega_t$ has been constructed. Let us run $\mathfrak{S}(\boldsymbol{x}_t)$ and calcualte the result $\gamma_t$ accurate to within $1/8$. We then will be able to guarantee that one of the inequalities hold: $\gamma_t \geq 1/4$ or $\gamma_t \leq 1/2$. In the former case let $\mathfrak{S}'(\boldsymbol{x}_t) = \omega_{t+1} = 0$ and in the later $\mathfrak{S}'(\boldsymbol{x}_t) = \omega_{t+1} = 1$ (if the prediction calculates at precisely $3/8$ let us go for the first option for the sake of being definite). Clearly,*

$$\mathrm{Loss}_{\mathfrak{S}'}(\boldsymbol{x}_t) = 0 \ ,$$

$$\mathrm{Loss}_{\mathfrak{S}}(\boldsymbol{x}_t) \geq \frac{t}{16}$$

*for all $t = 1, 2, \dots$ by construction. Thus there is no optimal computable strategy (up to a constant or even a function $\alpha(T)$ that is $o(T)$ as $T \to +\infty$, cf. Sect. 3.3) w.r.t. the square-loss game.*

*On the other hand, the construction from (3) provides a universal strategy for any enumerable set of strategies. One may conclude that computable strategies w.r.t. the square-loss game are not enumerable. This parallels a result from the theory of recursive functions stating that total computable functions from $\mathbb{N}$ to $\mathbb{N}$ are not enumerable.*

*This observation can be generalised to a large class of games in a straightforward fashion.*

## 3.3 Weak Complexity

We can relax the definition of predictive complexity by allowing $C$ to grow with the length of $\boldsymbol{x}$.

Take a function $\alpha : \mathbb{N} \to (0, +\infty)$. An upper semi-computable superloss process $\mathcal{K}$ is called *(weak) predictive complexity up to* $\alpha(T)$ if for any other upper semi-computable superloss process $L$ there is a constant $C$ such that

$$\mathcal{K}(\boldsymbol{x}) \leq L(\boldsymbol{x}) + C\alpha(|\boldsymbol{x}|)$$

for all $\boldsymbol{x} \in \Omega^*$.

We call a game $\mathfrak{G} = \langle \Omega, \Gamma, \lambda \rangle$ with $|\Omega| = M$ *convex* if the "finite" part of its set of superpredictions $S \cap \mathbb{R}^M$ is convex. A game is *bounded* if the loss function $\lambda$ is bounded.

The following theorem essentially reproduces Lemma 11 from [KV08].

**Theorem 2.** *If a game $\mathfrak{G}$ is convex and bounded and there is an effective enumeration of superloss processes w.r.t. $\mathfrak{G}$, then there is predictive complexity up to $\sqrt{T}$ w.r.t. $\mathfrak{G}$.*

*Proof.* Let $L$ be a dyadic upper bound on $\lambda$, i.e., $\lambda(\gamma, \omega) \leq L$ for all $\gamma \in \Gamma$ and $\omega \in \Omega$. Let $k_i$, $i \in \mathbb{N}$, be an effective enumeration of upper semi-computable superloss processes.

Take $\beta_t = e^{-1/\sqrt{t}}$ and

$$\mathcal{K}(\boldsymbol{x}) = \sum_{t=1}^{|\boldsymbol{x}|} \frac{L^2}{2\sqrt{t}} + \log_{\beta_{|\boldsymbol{x}|}} \sum_{i=1}^{\infty} \frac{1}{2^i} \beta_{|\boldsymbol{x}|}^{k_i(\boldsymbol{x})} \quad . \tag{4}$$

Let us show that $\mathcal{K}$ is an upper semi-computable superloss process.

**Lemma 1.** *For all sequences $p_i$ and $x_i$, $i \in \mathbb{N}$, such that $p_i \geq 0$ for all $i \in \mathbb{N}$ and $\sum_{i=1}^{\infty} p_i = 1$ and for all $0 < \beta_1 < \beta_2 < 1$ we have*

$$\log_{\beta_1} \sum_{i=1}^{\infty} p_i \beta_1^{x_i} \leq \log_{\beta_2} \sum_{i=1}^{\infty} p_i \beta_2^{x_i} \quad .$$

*Proof.* We have

$$\beta_2^{\log_{\beta_1} \sum_{i=1}^{\infty} p_i \beta_1^{x_i}} = \beta_1^{\left(\log_{\beta_1} \beta_2\right) \log_{\beta_1} \sum_{i=1}^{\infty} p_i \beta_1^{x_i}}$$

$$= \left(\sum_{i=1}^{\infty} p_i \beta_1^{x_i}\right)^{\log_{\beta_1} \beta_2}$$

$$\geq \sum_{i=1}^{\infty} p_i \beta_2^{x_i} \quad .$$

(The inequality follows by concavity of the function $x^{\delta}$ with $\delta \in (0,1)$ in $x$.) Taking the logarithm to the base $\beta_2$ yields the desired inequality. $\qquad\square\qquad\qquad\qquad\square$

The lemma implies that

$$\mathcal{K}(\boldsymbol{x}) \leq \sum_{t=1}^{|\boldsymbol{x}|} \frac{L^2}{\sqrt{t}} + \log_{\beta_{|\boldsymbol{x}|+1}} \sum_{i=1}^{\infty} \frac{1}{2^i} \beta_{|\boldsymbol{x}|+1}^{k_i(\boldsymbol{x})} \quad .$$

and

$$\mathcal{K}(\boldsymbol{x}\omega) - \mathcal{K}(\boldsymbol{x}) \geq \frac{L^2}{2\sqrt{|\boldsymbol{x}|+1}} + \log_{\beta_{|\boldsymbol{x}|+1}} \frac{\sum_{i=1}^{\infty} \frac{1}{2^i} \beta_{|\boldsymbol{x}|+1}^{k_i(\boldsymbol{x}\omega)}}{\sum_{i=1}^{\infty} \frac{1}{2^i} \beta_{|\boldsymbol{x}|+1}^{k_i(\boldsymbol{x})}}$$

$$\geq \frac{L^2}{2\sqrt{|\boldsymbol{x}|+1}} + \log_{\beta_{|\boldsymbol{x}|+1}} \sum_{i=1}^{\infty} \frac{\beta_{|\boldsymbol{x}|+1}^{k_i(\boldsymbol{x})}/2^i}{\sum_{j=1}^{\infty} \beta_{|\boldsymbol{x}|+1}^{k_j(\boldsymbol{x})}/2^j} \beta_{|\boldsymbol{x}|+1}^{k_i(\boldsymbol{x}\omega)-k_i(\boldsymbol{x})}$$

$$\geq \frac{L^2}{2\sqrt{|\boldsymbol{x}|+1}} + \log_{\beta_{|\boldsymbol{x}|+1}} \sum_{i=1}^{\infty} p_i \beta_{|\boldsymbol{x}|+1}^{\lambda(\gamma_i, \omega)} \quad ,$$

where

$$p_i = \frac{\beta_{|\boldsymbol{x}|+1}^{k_i(\boldsymbol{x})}/2^i}{\sum_{j=1}^{\infty} \beta_{|\boldsymbol{x}|+1}^{k_j(\boldsymbol{x})}/2^j}$$

and $\gamma_i \in \Gamma$ is such that $k_i(\boldsymbol{x}\omega) - k_i(\boldsymbol{x}) \geq \lambda(\gamma_i, \omega)$ for all $\omega \in \Omega$.

**Lemma 2.** *For all sequences $p_i$ and $x_i$, $i \in \mathbb{N}$, such that $p_i \geq 0$ and $|x_i| \leq L$ for all $i \in \mathbb{N}$ and $\sum_{i=1}^{\infty} p_i = 1$ and for all $0 < \beta < 1$ we have*

$$\sum_{i=1}^{\infty} p_i x_i - \log_\beta \sum_{i=1}^{\infty} p_i \beta^{x_i} \leq -\frac{L^2}{2} \ln \beta \ .$$

(Compare with Lemma A.3 from [CBL06].)

*Proof.* Applying the inequality $\ln x \leq x - 1$ yields

$$-\log_\beta \sum_{i=1}^{\infty} p_i \beta^{x_i} \leq -\frac{1}{\ln \beta} \left( \sum_{i=1}^{\infty} p_i \beta^{x_i} - 1 \right) \ . \tag{5}$$

Taylor's decomposition implies

$$\beta^x = e^{x \ln \beta} \leq 1 + x \ln \beta + \frac{(L \ln \beta)^2}{2} \ . \tag{6}$$

Substituting (6) into (5) completes the proof. □ □

The lemma implies that

$$\mathcal{K}(\boldsymbol{x}\omega) - \mathcal{K}(\boldsymbol{x}) \geq \sum_{i=1}^{\infty} p_i \lambda(\gamma_i, \omega_i) \ .$$

The convexity of $S \cap \mathbb{R}^M$ implies that there is $\gamma \in \Gamma$ such that $\mathcal{K}(\boldsymbol{x}\omega) - \mathcal{K}(\boldsymbol{x}) \geq \lambda(\gamma, \omega)$ for all $\omega \in \Omega$. Thus $\mathcal{K}$ is a superloss process.

One can show that $\mathcal{K}$ is upper semi-computable in the same way as in the proof of Theorem 1.

By dropping all terms except for one in the infinite series (4) one gets

$$\mathcal{K}(\boldsymbol{x}) = \sum_{t=1}^{|\boldsymbol{x}|} \frac{L^2}{2\sqrt{t}} + i\sqrt{|\boldsymbol{x}|} \ln 2 + k_i(\boldsymbol{x}) \ .$$

It remains to note that

$$\sum_{t=1}^{T} \frac{1}{\sqrt{t}} \leq \int_0^T \frac{1}{\sqrt{t}} = 2\sqrt{T} \ .$$

The theorem follows. □ □

The question of the existence of predictive complexity for unbounded convex games is an open problem. (Note that the case of unbounded mixable games, such as the logarithmic game, is straightforward.)

# 4 Generalised Entropy and Asymptotic Expectations

In this section we discuss an important technical lemma linking asymptotic properties of predictive complexity with geometric properties of the game.

Consider a game $\mathfrak{G} = \langle \Omega, \Gamma, \lambda \rangle$, where $\Omega = \left\{ \omega^{(0)}, \omega^{(1)}, \ldots, \omega^{(M-1)} \right\}$. The simplex

$$\Delta_M = \{(p_0, p_1, \ldots, p_{M-1}) \in [0,1]^M \mid \sum_{m=0}^{M-1} p_m\}$$

can be identified with the set of all probability distribution on $\Omega$. The *generalised entropy* $H_{\mathfrak{G}} : \Delta_M \to \mathbb{R}$ is defined by

$$H_{\mathfrak{G}}(p) = \inf_{\gamma \in \Gamma} \sum_{m=0}^{M-1} p_m \lambda\left(\gamma, \omega^{(m)}\right) \ ,$$

where $p = (p_0, p_1, \ldots, p_{M-1}) \in \Delta_M$. It follows from the compactness of $\Gamma$ and continuity of $\lambda$ that the infimum in the definition is always achieved and can be replaced by the minimum (cf. [KVV14]). Proposition 10 from [KVV14] shows that $H_{\mathfrak{G}}$ is a continuous function on $\Delta_M$.

The definition of entropy can also be rewritten as

$$H_{\mathfrak{G}}(p) = \inf_{x = \left(x_0, x_1, \ldots, x_{M-1}\right) \in \Sigma_{\mathfrak{G}}} \sum_{m=0}^{M-1} p_m x_m \ , \tag{7}$$

where $\Sigma_{\mathfrak{G}}$ is the set of superpredictions w.r.t. $\mathfrak{G}$. We can restrict the infimum to $x = (x_0, x_1, \ldots, x_{M-1}) \in \Sigma_{\mathfrak{G}} \cap \mathbb{R}^M$ because the predictions leading to finite losses are dense.

**Lemma 3.** *If the game $\mathfrak{G}$ is computable and $\mathcal{K}$ is predictive complexity w.r.t. $\mathfrak{G}$ up to $\alpha(t)$ such that $\alpha(t) = o(t)$ as $t \to \infty$, then for every $p = (p_0, p_1, \ldots, p_{M-1}) \in \Delta_M$ we have*

$$\lim_{T \to \infty} \frac{1}{T} \mathbf{E}\mathcal{K}\left(\xi_1^{(p)} \xi_2^{(p)} \ldots \xi_T^{(p)}\right) = H_{\mathfrak{G}}(p) \ ,$$

*where $\xi_t^{(p)}$ are independent identically distributed Bernoulli trials such that*

$$\Pr\left(\xi_t^{(p)} = \omega^{(m)}\right) = p_m \ ,$$

$m = 0, 1, \ldots, M-1$, $t = 1, 2, \ldots$.

*Proof.* For every computable prediction $\gamma \in \Gamma$ one can consider the strategy $\mathfrak{G}_\gamma$ predicting $\gamma$ on every step. We have $\mathcal{K}(\boldsymbol{x}) \leq \mathrm{Loss}_{\mathfrak{G}_\gamma}(\boldsymbol{x}) + c\alpha(|\boldsymbol{x}|)$ and

$$\frac{1}{T} \mathbf{E}\,\mathrm{Loss}_{\mathfrak{G}_\gamma}\left(\xi_1^{(p)} \xi_2^{(p)} \ldots \xi_T^{(p)}\right) = \sum_{m=0}^{M-1} p_m \lambda\left(\gamma, \omega^{(m)}\right) \ ;$$

thus

$$\limsup_{T \to \infty} \frac{1}{T} \mathbf{E}\mathcal{K}\left(\xi_1^{(p)} \xi_2^{(p)} \ldots \xi_T^{(p)}\right) \leq \sum_{m=0}^{M-1} p_m \lambda\left(\gamma, \omega^{(m)}\right)$$

for every computable $\gamma$. Since computable predictions are dense in $\Gamma$ and $\lambda$ is continuous, we get

$$\limsup_{T \to \infty} \frac{1}{T} \mathbf{E}\mathcal{K}\left(\xi_1^{(p)} \xi_2^{(p)} \ldots \xi_T^{(p)}\right) \leq H_{\mathfrak{G}}(p) \ .$$

In order to obtain a lower bound, notice that

$$\mathbf{E}\mathcal{K}\left(\xi_1^{(p)} \xi_2^{(p)} \ldots \xi_T^{(p)}\right) = \sum_{t=1}^{T} \mathbf{E}\left(\mathcal{K}\left(\xi_1^{(p)} \xi_2^{(p)} \ldots \xi_t^{(p)}\right) - \mathcal{K}\left(\xi_1^{(p)} \xi_2^{(p)} \ldots \xi_{t-1}^{(p)}\right)\right)$$

and

$$\mathbf{E}\left(\mathcal{K}\left(\boldsymbol{x}\xi_t^{(p)}\right) - \mathcal{K}(\boldsymbol{x})\right) \geq \sum_{m=0}^{M-1} p_m \lambda\left(\gamma, \omega^{(m)}\right)$$

for some $\gamma \in \Gamma$. Hence

$$\mathbf{E}\left(\mathcal{K}\left(\boldsymbol{x}\xi_t^{(p)}\right) - \mathcal{K}(\boldsymbol{x})\right) \geq H_{\mathfrak{G}}(p)$$

and

$$\mathbf{E}\mathcal{K}\left(\xi_1^{(p)} \xi_2^{(p)} \ldots \xi_T^{(p)}\right) \geq T H_{\mathfrak{G}}(p) \ . \tag{8}$$

Therefore

$$\liminf_{T \to \infty} \frac{1}{T} \mathbf{E}\mathcal{K}\left(\xi_1^{(p)} \xi_2^{(p)} \ldots \xi_T^{(p)}\right) \geq H_{\mathfrak{G}}(p) \ .$$

The theorem follows. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$ $\qquad\qquad\qquad\qquad\square$

In [KVV04b] this lemma is strengthened. It can be shown that under general conditions the ratio $\mathcal{K}\left(\xi_1^{(p)}\xi_2^{(p)}\ldots\xi_T^{(p)}\right)/T$ converges to $H(p)$ almost surely. In [GN12] the result is generalised to ergodic sequences of outcomes.

For the logarithmic game the generalised entropy is the Shannon entropy; the connections between Shannon entropy and Kolmogorov complexity are explored in [LV08], Sect. 2.8.1.

For connections between the generalised entropy and a related asymptotic complexity see [KVV14].

# 5 Negative Results on the Existence of Predictive Complexity

The following simple theorem shows that convexity of the game is a necessary condition for the existence of complexity. We therefore concentrate on convex games throughout the paper.

**Theorem 3.** *If a computable game $\mathfrak{G}$ is not convex, there is no predictive complexity w.r.t. $\mathfrak{G}$ up to $\alpha(T)$ such that $\alpha(T) = o(T)$ as $T \to \infty$.*

**Lemma 4.** *The set of superpredictions $\Sigma \subseteq [-\infty, +\infty]^M$ w.r.t. a game $\mathfrak{G}$ is closed w.r.t. the extended Euclidean topology.*

*Proof.* Take $\mathfrak{G} = \langle \Omega, \Gamma, \lambda \rangle$ with $\Omega = \left\{ \omega^{(0)}, \omega^{(1)}, \ldots, \omega^{(M-1)} \right\}$.

Suppose that a sequence of superpredictions $s_i = \left( s_i^{(0)}, s_i^{(1)}, \ldots, s_i^{(M-1)} \right)$, $i \in \mathbb{N}$, converge to $s = \left( s^{(0)}, s^{(1)}, \ldots, s^{(M-1)} \right)$ w.r.t. the extended topology. For every $i \in \mathbb{N}$ there is $\gamma_i \in \gamma$ such that $\lambda\left( \gamma_i, \omega^{(m)} \right) \leq s_i^{(m)}$, $m = 0, 1, \ldots, M-1$. Since $\Gamma$ is compact, there a subsequence of $\gamma_i$ converging to $\gamma_0 \in \Gamma$ and continuity of $\lambda$ implies $\lambda\left( \gamma_0, \omega^{(m)} \right) \leq s^{(m)}$, $m = 0, 1, \ldots, M-1$. □ □

*of Theorem 3.* Let $\Sigma$ be the set of superpredictions w.r.t. $\mathfrak{G} = \langle \Omega, \Gamma, \lambda \rangle$ with $\Omega = \left\{ \omega^{(0)}, \omega^{(1)}, \ldots, \omega^{(M-1)} \right\}$. If $\mathfrak{G}$ is not convex, there are points $b' = (b'_0, b'_1, \ldots, b'_{M-1})$ and $b'' = (b''_0, b''_1, \ldots, b''_{M-1}) \in \Sigma \cap \mathbb{R}^M$ such that the line segment connecting them is not contained in $\Sigma \cap \mathbb{R}^M$, i.e., there is $\beta \in (0,1)$ such that $a' = \beta b' + (1-\beta) b'' \notin \Sigma$. Since $\Sigma$ is closed, there is a vicinity of $a'$ outside of $\Sigma$ and $a = (a_0, a_1, \ldots, a_{M-1}) > a'$ such that $a \notin \Sigma$.

Since $b'$ and $b''$ are superpredictions, there are predictions $\gamma'$ and $\gamma''$ from $\Gamma$ such that $\lambda\left( \gamma', \omega^{(m)} \right) \leq b'_m$ and $\lambda\left( \gamma'', \omega^{(m)} \right) \leq b''_m$, $m = 0, 1, \ldots, M-1$. Since computable predictions are dense in $\Gamma$, $b'$ and $b''$ can be slightly moved to ensure that $\gamma'$ and $\gamma''$ can be chosen to be computable and the line segment connecting them is still not contained in $\Sigma$. Let us assume $b'$ and $b''$ are of this king without restricting the generality.

Assume that $\mathcal{K}$ is predictive complexity w.r.t. $\alpha(T)$ such that $\alpha(T) = o(T)$ w.r.t. $\mathfrak{G}$. By considering the two strategies, one predicting $\gamma'$ and another predicting $\gamma''$ all the time, we get the following inequalities for all $\boldsymbol{x} \in \Omega^*$:

$$\mathcal{K}(\boldsymbol{x}) \leq \sum_{m=0}^{M-1} b'_m \sharp_m \boldsymbol{x} + C\alpha(|\boldsymbol{x}|) \tag{9}$$

$$\mathcal{K}(\boldsymbol{x}) \leq \sum_{m=0}^{M-1} b''_m \sharp_m \boldsymbol{x} + C\alpha(|\boldsymbol{x}|) \tag{10}$$

where $C \in \mathbb{R}$. Taking a convex combination of (9) and (10) we get

$$\mathcal{K}(\boldsymbol{x}) \leq \sum_{m=0}^{M-1} a'_m \sharp_m \boldsymbol{x} + C\alpha(|\boldsymbol{x}|) \ . \tag{11}$$

Since $a \notin \Sigma$, no point $s \in \Sigma$ satisfies $s \leq a$ and for every $s = (s_0, s_1, \ldots, s_{M-1}) \in \Sigma$ there is $i$ such that $s_i > a_i$. By induction we construct a sequence $\omega_1, \omega_2, \ldots$, such that for every $T$ we have

$$\mathcal{K}(\omega_1\omega_2\ldots\omega_T\omega_{T+1}) - \mathcal{K}(\omega_1\omega_2\ldots\omega_T) > a_i$$
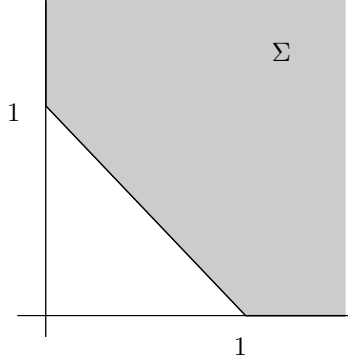
9

Figure 1: The set of superpredictions for the absolute-loss game.

where $\omega_T = \omega^{(i)}$ or, in other terms,

$$\mathcal{K}(\omega_1\omega_2\ldots\omega_T) > \sum_{m=0}^{M-1} a_m \sharp_m(\omega_1\omega_2\ldots\omega_T) \ . \tag{12}$$

Combining this with (11) we get

$$\sum_{m=0}^{M-1} (a_m - a'_m)\sharp_m(\omega_1\omega_2\ldots\omega_T) \leq o(T)$$

as $T \to \infty$. Taking $\delta = \min_{m=0,1,\ldots,M-1}(a_m - a'_m) > 0$ we get

$$\delta T \leq o(T)$$

as $T \to \infty$, which is a contradiction. $\qquad\qquad\square \qquad\qquad\qquad\qquad \square$

For the absolute-loss game we can prove a matching lower bound.

**Theorem 4.** *There is no predictive complexity w.r.t. the absolute-loss game up to $\alpha(T)$ such that $\alpha(T) = o(\sqrt{T})$ as $T \to \infty$.*

*Proof.* The set of superpredictions for the absolute-loss game is shown in Fig. 1. If $\mathcal{K}$ is predictive complexity w.r.t. the absolute loss game, then

$$\mathcal{K}\left(\xi_1^{(1/2)}\xi_2^{(1/2)}\ldots\xi_T^{(1/2)}\right) \geq \frac{T}{2} \ , \tag{13}$$

where $\xi_t^{(1/2)}$ are independent Bernoulli trials such that

$$\Pr\left(\xi_t^{(1/2)} = 0\right) = \Pr\left(\xi_t^{(1/2)} = 1\right) = 1/2$$

(recall inequality (8)).

Consider two strategies, $\mathfrak{S}_1$ that always predicts 1 and $\mathfrak{S}_0$ that always predicts 0. If $\mathcal{K}$ is predictive complexity up to $\alpha(T)$, then there is a constant $C$ such that

$$\mathcal{K}(\boldsymbol{x}) \leq \mathrm{Loss}_{\mathfrak{S}_1}(\boldsymbol{x}) + C\alpha(T) = \sharp_0\boldsymbol{x} + C\alpha(|\boldsymbol{x}|) \ , \tag{14}$$
$$\mathcal{K}(\boldsymbol{x}) \leq \mathrm{Loss}_{\mathfrak{S}_0}(\boldsymbol{x}) + C\alpha(T) = \sharp_1\boldsymbol{x} + C\alpha(|\boldsymbol{x}|) \tag{15}$$

for every $\boldsymbol{x} \in \mathbb{B}^*$. Since $\min(\sharp_0\boldsymbol{x}, \sharp_1\boldsymbol{x}) \leq |\boldsymbol{x}|/2$, combining the inequalities yields

$$\mathcal{K}(\boldsymbol{x}) \leq \frac{|\boldsymbol{x}|}{2} + C\alpha(|\boldsymbol{x}|) \ . \tag{16}$$

10

Consider the set $\Xi_T \subseteq \mathbb{B}^T$ of strings $\boldsymbol{x}$ with $T/2 + \sqrt{T}$ or more zeroes. The DeMoivre-Laplace limit theorem (see, e.g., [Fel68]) implies that

$$\Pr\left(\xi_1^{(1/2)}\xi_2^{(1/2)}\ldots\xi_T^{(1/2)} \in \Xi_T\right) = \delta + o(1)$$

as $T \to \infty$, where $\delta > 0$. It follows from (14) that for every $\boldsymbol{x} \in \Xi_T$ the inequality

$$\mathcal{K}(\boldsymbol{x}) \leq \frac{T}{2} - \sqrt{T} + C\alpha(|\boldsymbol{x}|)$$

holds. Combining this with inequality (16) for $\boldsymbol{x} \in \mathbb{B}^T \setminus \Xi_T$ we get

$$\mathbf{E}\mathcal{K}\left(\xi_1^{(1/2)}\xi_2^{(1/2)}\ldots\xi_n^{(1/2)}\right) \leq \left(\frac{T}{2} - \sqrt{T}\right)\Pr(\Xi_T) + \frac{T}{2}(1 - \Pr(\Xi)) + C\alpha(T)$$

$$= \frac{T}{2} - \sqrt{T}(\delta + o(1)) + C\alpha(T) \ .$$

Comparing this with (13) completes the proof. $\qquad\qquad\square\qquad\qquad\qquad\qquad\qquad\square$

Obtaining tight lower bounds of this kind in the general case is an open problem. In [KVV04a] the argument from Theorem 4 is developed to show that there is no basic predictive complexity for non-mixable binary games. In [KV02] some lower bounds for general convex binary games are obtained.

# 6 Uniqueness Theorem

In this section we show that predictive complexity uniquely determines the set of superpredictions; different games with the same set of superpredictions can be thought as mere parametrisations.

We start by considering generalised entropies.

**Lemma 5.** *Let $H_1$ and $H_2$ be generalised entropies w.r.t. convex games $\mathfrak{G}_1$ and $\mathfrak{G}_2$ with the same finite outcome space $\Omega$ of size $M < +\infty$ and sets of superpredictions $\Sigma_1$ and $\Sigma_2$, respectively. Then $H_1(p) \leq H_2(p)$ for all $p \in \Delta_M$ if and only if $\Sigma_2 \subseteq \Sigma_1$.*

*Proof.* It immediately follows from (7) that $\Sigma_2 \subseteq \Sigma_1$ implies $H_1(p) \leq H_2(p)$ for all $p \in \Delta_M$.

Let $\Sigma_2 \nsubseteq \Sigma_1$ and there is $s = (s_0, s_1, \ldots, s_{M-1}) \in \Sigma_2 \setminus \Sigma_1$. Lemma 4 and the assumption that predictions leading to finite losses are dense imply that without restricting the generality we can assume $s \in \mathbb{R}^M$. Corollary 11.4.2 from [Roc70] implies that closed convex sets $\{s\}$ and $\Sigma_1 \cap \mathbb{R}^M$ can be strongly separated by a hyperplane, i.e., there is a linear function $h : \mathbb{R}^M \to \mathbb{R}$ given by $h(x_0, x_1, \ldots, x_{M-1}) = \alpha_0 x_0 + \alpha_1 x_1 + \ldots + \alpha_{M-1} x_{M-1}$ such that

$$\alpha(s) < \inf_{x \in \Sigma_1} h(x) \ . \tag{17}$$

If there is $i$ such that $\alpha_i < 0$, then (17) cannot hold. Indeed, take a point $x \in \Sigma_1$. For every $y > 0$ the point $x + (0, \ldots, 0, y, 0, \ldots, 0)$, where $y$ is in position $i$, belongs to $\Sigma_1$ and $h(x + (0, \ldots, 0, y, 0, \ldots, 0)) = h(x) + \alpha_i y \to -\infty$ as $y \to +\infty$.

Therefore the numbers $p_i = \alpha_i / \sum_{j=0}^{M-1} \alpha_j \geq 0$, $i = 0, 1, \ldots, M-1$, form a distribution

$$p = (p_0, p_1, \ldots, p_{M-1}) \in \Delta_M \ .$$

We get

$$H_2(p) \leq \sum_{i=0}^{M-1} p_i s_i < \inf_{(x_0, x_1, \ldots, x_{M-1}) \in S \cap \mathbb{R}^M} \sum_{i=0}^{M-1} p_i x_i = H_1(p) \ .$$

This completes the proof. $\qquad\qquad\qquad\square\qquad\qquad\qquad\qquad\qquad\square$

**Corollary 1.** *Let $H_1$ and $H_2$ be generalised entropies w.r.t. convex games $\mathfrak{G}_1$ and $\mathfrak{G}_2$ with the same finite outcome space $\Omega$ of size $M < +\infty$ and sets of superpredictions $\Sigma_1$ and $\Sigma_2$, respectively. Then $H_1(p) = H_2(p)$ for all $p \in \Delta_M$ if and only if $\Sigma_1 = \Sigma_2$.*

**Theorem 5.** *Let $\mathcal{K}_1$ be predictive complexity w.r.t. $\mathfrak{G}_1$ up to $\alpha_1(T)$ such that $\alpha_1(T) = o(T)$ as $T \to \infty$ and $\mathcal{K}_2$ be predictive complexity w.r.t. $\mathfrak{G}_2$ up to $\alpha_2(T)$ such that $\alpha_2(T) = o(T)$ as $T \to \infty$. If the games $\mathfrak{G}_1$ and $\mathfrak{G}_2$ are computable, have the same finite set of outcomes $\Omega$, and there is a function $\alpha_3(T)$ such that $|\mathcal{K}_1(\boldsymbol{x}) - \mathcal{K}_2(\boldsymbol{x})| \leq \alpha_3(|\boldsymbol{x}|)$ and $\alpha_3(T) = o(T)$ as $T \to \infty$, then the games $\mathfrak{G}_1$ and $\mathfrak{G}_2$ have the same set of superpredictions.*

*Proof.* The proof is by combining Lemma 3 with Corollary 1. □ □

The theorem implies the following observation. There are several versions of Kolmogorov complexity, plain, prefix, and monotonic complexities, all within a term not exceeding $C \ln(|\boldsymbol{x}|)$ but not within a constant term from complexity KM equal to the logarithmic complexity. One may wonder if these complexities can be predictive complexities for some other games. Theorem 5 implies that this is not the case.

# 7   Inequalities

In this section we consider inequalities between predictive complexities for different games.

**Theorem 6.** *Let $\mathcal{K}_1$ be predictive complexity w.r.t. $\mathfrak{G}_1$ up to $\alpha_1(T)$ such that $\alpha_1(T) = o(T)$ as $T \to \infty$ and $\mathcal{K}_2$ be predictive complexity w.r.t. $\mathfrak{G}_2$ up to $\alpha_2(T)$ such that $\alpha_2(T) = o(T)$ as $T \to \infty$. If the games $\mathfrak{G}_1$ and $\mathfrak{G}_2$ are computable and have the same finite set of outcomes*

$$\Omega = \left\{ \omega^{(0)}, \omega^{(1)}, \ldots, \omega^{(M-1)} \right\} \ ,$$

*then the following conditions are equivalent:*

  i. *there is a function $\alpha(T)$ such that $\alpha(T) = o(T)$ as $T \to \infty$ and for all $\boldsymbol{x} \in \Omega^*$ we have*

$$\mathcal{K}_1(\boldsymbol{x}) + \alpha(|\boldsymbol{x}|) \geq \mathcal{K}_2(\boldsymbol{x}) \ ;$$

  ii. *there is a constant $C \in \mathbb{R}$ such that for all $\boldsymbol{x} \in \Omega^*$ we have*

$$\mathcal{K}_1(\boldsymbol{x}) + C\alpha_2(|\boldsymbol{x}|) \geq \mathcal{K}_2(\boldsymbol{x}) \ ; \tag{18}$$

  iii. *$\Sigma_1 \subseteq \Sigma_2$, where $\Sigma_1$ and $\Sigma_2$ are the sets of superpredictions w.r.t. $\mathfrak{G}_1$ and $\mathfrak{G}_2$, respectively;*

  iv. *for all $p \in \Delta_M$ we have $H_2(p) \leq H_1(p)$, where $H_1$ and $H_2$ are the generalised entropies w.r.t. $\mathfrak{G}_1$ and $\mathfrak{G}_2$, respectively;*

  v. *for every distribution $p \in \Delta_M$ there is a function $\alpha_p(T)$ such that $\alpha_p(T) = o(T)$ as $T \to \infty$ and*

$$\mathbf{E}\mathcal{K}_1 \left( \xi_1^{(p)} \ldots \xi_T^{(p)} \right) + \alpha_p(T) \geq \mathbf{E}\mathcal{K}_2 \left( \xi_1^{(p)} \ldots \xi_T^{(p)} \right) \tag{19}$$

*for every $T \in \mathbb{N}$, where $\xi_t^{(p)}$ are independent identically distributed Bernoulli trials such that*

$$\Pr \left( \xi_t^{(p)} = \omega^{(m)} \right) = p_m \ ,$$

*$m = 0, 1, \ldots, M - 1$, $t = 1, 2, \ldots$;*

  vi. *there is a constant $C \in \mathbb{R}$ such that*

$$\mathbf{E}\mathcal{K}_1 \left( \xi_1^{(p)} \ldots \xi_T^{(p)} \right) + C\alpha_2(T) \geq \mathbf{E}\mathcal{K}_2 \left( \xi_1^{(p)} \ldots \xi_T^{(p)} \right)$$

*for every $T \in \mathbb{N}$, where $\xi_t^{(p)}$ are as above.*

*Proof.* Some implications on the list are trivial: ii implies i while vi implies v. By taking expectations we can go from ii to vi and from i to v. Items iii and iv are equivalent by Lemma 5.

By dividing (19) by $T$, letting $T$ go to infinity and invoking Lemma 3, which connects expectations and entropies, one can show that v implies iv.

12

Let us show that iii implies ii. Since $\mathcal{K}_1$ is a superloss process w.r.t. $\mathfrak{G}_1$, for every $\boldsymbol{x} \in \Omega^*$ the tuple

$$\left( \mathcal{K}_1\left(\boldsymbol{x}\omega^{(0)}\right) - \mathcal{K}_1(\boldsymbol{x}), \mathcal{K}_1\left(\boldsymbol{x}\omega^{(1)}\right) - \mathcal{K}_1(\boldsymbol{x}), \ldots, \mathcal{K}_1\left(\boldsymbol{x}\omega^{(M-1)}\right) - \mathcal{K}_1(\boldsymbol{x}) \right)$$

is a superprediction w.r.t. $\mathfrak{G}_1$ and belongs to $\Sigma_1$. If $\Sigma_1 \subseteq \Sigma_2$ then the tuple belongs to $\Sigma_2$ and therefore $\mathcal{K}_1$ is a superloss process w.r.t. $\mathfrak{G}_2$. Since the process $\mathcal{K}_1$ is upper semi-computable, by the definition of predictive complexity there is a constant $C$ such that (18) holds for all $\boldsymbol{x} \in \Omega^*$. $\qquad\square\qquad\qquad\square$

The theorem covers inequalities $\mathcal{K}_1(\boldsymbol{x}) + \alpha(|x|) \geq \mathcal{K}_2(\boldsymbol{x})$ with $\alpha(T) = o(T)$ as $T \to \infty$. What if $\alpha$ grows linearly? We need the following notation. For $M \subseteq [-\infty, +\infty]^K$, $c \in \mathbb{R}$ and $s = (s_1, s_2, \ldots, s_K) \in \mathbb{R}^K$ we let

$$cM + s = \{(cx_1 + s_1, cx_2 + cs_2, \ldots, cx_K + s_K) \mid (x_1, x_2, \ldots, x_K) \in M\} \ .$$

**Corollary 2.** *Under the conditions of Theorem 6, if $a \in (0, +\infty)$ and $b \in \mathbb{R}$ are computable, then there is $C > 0$ such that*

$$a\mathcal{K}_1 + b|\boldsymbol{x}| + C\alpha_2(\boldsymbol{x}) \geq \mathcal{K}_2(\boldsymbol{x})$$

*for all $\boldsymbol{x} \in \Omega$ if and only if $a\Sigma_1 + (b, b, \ldots, b) \subseteq \Sigma_2$.*

In [Kal02] there is a discussion of inequalities of the type $a_1\mathcal{K}_1(\boldsymbol{x}) + a_2\mathcal{K}_2(\boldsymbol{x}) \leq b|\boldsymbol{x}|$ for binary games with $\Gamma = [0, 1]$ that are symmetric (i.e., $\lambda(\gamma, 0) = \lambda(1 - \gamma, 1)$). There is also a description of the set $(a, b) \in \mathbb{R}^2$ that is the closure of the set of computable $a > 0$ and $b$ such that for some $C \in \mathbb{R}$ the inequality $a\mathcal{K}^{\mathrm{sq}}(\boldsymbol{x}) + b|\boldsymbol{x}| + C \geq \mathcal{K}^{\log}(\boldsymbol{x})$ holds for all $\boldsymbol{x} \in \mathbb{B}^*$, where $\mathcal{K}\mathrm{sq}$ and $\mathcal{K}^{\log}$ are the basic predictive complexities w.r.t. the square-loss and logarithmic games. In [VV02] the behaviour of the ratio $\mathcal{K}^{\log}(\boldsymbol{x})/\mathcal{K}^{\mathrm{sq}}(\boldsymbol{x})$ is analysed.

# 8   Other Results

This section gives references to some more results on predictive complexity.

The result of [KVV05] may be called unpredictability property; this property of predictive complexity parallels the incompressibility property of Kolmogorov complexity (see Sect. 3.15 of this book). Most strings of length $n$ have Kolmogorov complexity close to $n$ and only exponentially few can have lower complexity. In a similar way, most sequences of length $T$ have predictive complexity close to the loss of a trivial minimax strategy and only exponentially few can have smaller complexity.

In [VV05] the concepts of conditional predictive complexity (cf. Sect. 3.12 of this book, where conditional Kolmogorov complexity is discussed) and mutual information are given and analysed.

# Acknowledgement

# Appendix: Enumerating Superloss Processes

In this appendix we will discuss the question of effective enumeration of superloss processes. We reproduce and analyse the construction from [VW98].

A process $L : \Omega^* \to [0, +\infty]$ is *finitary* if the set $\{\boldsymbol{x} \in \Omega^* \mid L(\boldsymbol{x}) < +\infty\}$ is finite. A process $L$ is *dyadic* if its values are dyadic rationals or $+\infty$.

We call a dyadic finitary superloss process $L$ *verifiable* if for every $\boldsymbol{x} \in \Omega$ there is $\gamma \in \Gamma$ such that $L(\boldsymbol{x}\omega) - L(\boldsymbol{x}) > \lambda(\gamma, \omega)$ for all $\omega \in \Omega$. The inequality is equivalent to $e^{L(\boldsymbol{x}) - L(\boldsymbol{x}\omega)} < e^{-\lambda(\gamma, \omega)}$. Since $\lambda$ is continuous, the inequalities will still hold in a small vicinity of $\gamma$. Recall that for computable games we postulated the existence of an effective dense dyadic sequence $\gamma_i$. Thus for a computable game if we are given a finite list of pairs $(\boldsymbol{x}_s, r_s) \in \Omega^* \times \mathbb{D}$, $s = 1, 2, \ldots, S$, such that

$$N(\boldsymbol{x}) = \begin{cases} \min\{\mathrm{d}(r_s) \mid (\boldsymbol{x}, r_s) \text{ is in the list}\} & \text{if } \boldsymbol{x} = \boldsymbol{x}_s \text{ for some} \\ & \qquad\qquad s = 1, 2, \ldots, S \ ; \\ +\infty & \text{otherwise} \end{cases} \qquad (20)$$

13

is a verifiable dyadic finitary superloss process, we will be able to confirm that.

Therefore verifiable dyadic finitary superloss processes can be effectively enumerated. Let $P_i$, $i \in \mathbb{N}$, be an effective enumeration of programs such that each $P_i$ outputs a finite list of pairs $(\boldsymbol{x}_s, r_s) \in \Omega^* \times \mathbb{D}$, $s = 1, 2, \ldots, S_i$ (the program must halt after finitely many steps) defining a verifiable dyadic finitary superloss process $N_i$ as in (20) and every verifiable dyadic finitary superloss process is calculated by some $P_i$.

Pick a universal partial computable function $M(i, j)$ on $\mathbb{N}^2$. Universality means that every partial computable function on integers coincides with some $M(i, \cdot)$. Put $M^*(i, j) = M(i, j)$ if

1. the function $M$ is defined on all pairs $(i, j')$ with $j' \leq j$ and

2. the dyadic finitary superloss process $N_{M*(i,j')}$ never exceeds $N_{M*(i,j'+1)}$ (i.e., for all $\boldsymbol{x} \in \Omega^*$ we have $N_{M*(i,j')}(\boldsymbol{x}) \leq N_{M*(i,j'+1)}(\boldsymbol{x})$), $j' = 1, 2, \ldots, j - 1$.

and let $M^*(i, j)$ be undefined otherwise.

For every $i \in \mathbb{N}$ define a process $k_i$ by

$$k_i(\boldsymbol{x}) = \inf_j N_{M^*(i,j)} \ ,$$

where the infimum is taken over all $j$ such that $M^*(i, j)$ is defined. Clearly, $k_i$ is an upper semi-computable superloss process. Indeed, if $M^*(i, j)$ is undefined from some $j$ on, then $k_i$ is a finitary superloss process. Otherwise each value $k_i(\boldsymbol{x})$ is the limit of a non-increasing sequence of $L_j(\boldsymbol{x}) = N_{M^*(i,j)}(\boldsymbol{x})$. Since each $L_j$ is a superloss process, for every $\boldsymbol{x}$ there is a $\gamma_j \in \Gamma$ such that (2) holds for $L_j$ for all $\omega \in \Omega$. Since $\Gamma$ is compact, there is a converging subsequence of $\gamma_j$ and therefore (2) holds in the limit and thus $k_i$ is a superloss process. Since the partial function $f(\boldsymbol{x}, n) = L_n(\boldsymbol{x})$ is uniformly computable, $k_i$ is upper semi-computable.

In order to show that this construction allows us to enumerate *all* superloss processes, we need to prove that every superloss process is the limit of a uniformly computable sequence of verifiable dyadic finitary superloss processes. We will formulate a sufficient condition for that.

Consider a game $\mathfrak{G} = \langle \Omega, \Gamma, \lambda \rangle$ with $\Omega = \left\{ \omega^{(0)}, \omega^{(1)}, \ldots, \omega^{(M-1)} \right\}$. Consider the partial function $H : [0, +\infty]^M \to \mathbb{R}$ defined by

$$H(x_0, x_1, \ldots, x_{M-1}) = \max\{h \geq 0 \mid \text{there is } \gamma \in \Gamma \text{ such that}$$
$$x_m - h \geq \lambda\left(\gamma, \omega^{(m)}\right), m = 0, 1, \ldots, M-1\}$$

(if the set is empty, the function is undefined). Note that the maximum is achieved because $\Gamma$ is compact and $\lambda$ is continuous. Let us call $\mathfrak{G}$ *a game with effective minoration* if $H$ is computable where it is defined (here we assume that $+\infty$ is given to us as a special symbol).

**Lemma 6.** *Let a computable binary game $\langle \mathbb{B}, [0, 1], \lambda \rangle$ have a monotonous loss function so that $\lambda(\gamma, 0)$ is non-decreasing and $\lambda(\gamma, 1)$ is non-increasing. Then the function $H$ is computable where it is defined.*

This lemma implies that the binary square-loss, absolute-loss, and logarithmic games are games with effective minoration.

*Proof.* The system of inequalities

$$x_0 - h \geq \lambda(\gamma, 0) \ ;$$
$$x_1 - h \geq \lambda(\gamma, 1)$$

is equivalent to

$$e^h \leq e^{x_0} e^{-\lambda(\gamma,0)} \ ;$$
$$e^h \leq e^{x_1} e^{-\lambda(\gamma,1)} \ .$$

The maximum $h$ is achieved where the monotonous graphs of the functions on the right-hand side intersect.
□                                                                                                            □

**Lemma 7.** *If $\mathfrak{G}$ is a computable game with effective minoration, then every upper semi-computable superloss process $L$ is the infimum of a non-increasing effective sequence of verifiable finitary superloss processes.*

*Proof.* We say that a process $L_1$ *majorates* a process $L_2$ if $L_1(\boldsymbol{x}) \geq L_2(\boldsymbol{x})$ for all $\boldsymbol{x} \in \Omega^*$. A set of pairs $(\boldsymbol{x}_s, r_s) \in \Omega^* \times \mathbb{D}$, $s = 1, 2, \ldots, S$, *majorates* a process $L_2$ if $\mathrm{d}(r_s) \geq L_2(\boldsymbol{x}_s)$ for all $s = 1, 2, \ldots, S$.

**Lemma 8.** *For a finite set of pairs $(\boldsymbol{x}_s, r_s) \in \Omega^* \times \mathbb{D}$, $s = 1, 2, \ldots, S$, that majorate some superloss process there is the maximum finitary superloss process $N$ that the set majorates, i.e., there is a finitary superloss process $N$ majorated by the set of pairs and majorating every other superloss process majorated by the set of pairs.*

*Proof.* Let $n = \max_s |\boldsymbol{x}_s|$ be the maximum length of a sequence in the set. If $|\boldsymbol{x}| > n$ we let $N(\boldsymbol{x}) = +\infty$. If $|\boldsymbol{x}| = n$ we let $N(\boldsymbol{x})$ to be the minimum of $\mathrm{d}(r_s)$ such that $(\boldsymbol{x}, r_s)$ is in the set or $+\infty$ if there are none. Clearly, for every superloss process $L'$ majorated by the set of pairs, we have $N(\boldsymbol{x}) \geq L'(\boldsymbol{x})$ so far.

For sequences $\boldsymbol{x}$ of smaller length we define $N(\boldsymbol{x})$ by induction from larger lengths to smaller by setting $N(\boldsymbol{x})$ to be the minimum of $\mathrm{d}(r_s)$ such that the pair $(\boldsymbol{x}, r_s)$ is in the set and

$$H\left(N\left(\boldsymbol{x}\omega^{(0)}\right), N\left(\boldsymbol{x}\omega^{(1)}\right), \ldots, N\left(\boldsymbol{x}\omega^{(M-1)}\right)\right) \ .$$

It is easy to see that if for some superloss process $L'$ we have $N\left(\boldsymbol{x}\omega^{(m)}\right) \geq L'\left(\boldsymbol{x}\omega^{(m)}\right)$ for all $m = 0, 1, \ldots, M-1$, then $H \geq L'(\boldsymbol{x})$. $\qquad\square\qquad\qquad\square$

Let $L(\boldsymbol{x}) = \inf_{n \in \mathbb{N}} \mathrm{d}(f(\boldsymbol{x}, n))$ for some partial computable $f : \Omega^* \times \mathbb{N} \to \mathbb{D}$. We keep generating pairs $(\boldsymbol{x}, f(\boldsymbol{x}, n)) \in \Omega^* \times \mathbb{D}$ and every so often (e.g., after every 1000 computation steps) we define a verifiable dyadic finitary superloss process $L_i$; however sometimes we withhold it. The first process $L_1$ is not withheld.

Let $L_j$ be the latest process that was not withheld. The procedure for producing $L_i$ is as follows. Suppose that we have generated $S$ pairs $(\boldsymbol{x}_s, r_s)$. Let $n$ be the largest length of $\boldsymbol{x}_s$. Take a dyadic $\varepsilon = 2^{-i-k}$ where $2^k$ is the minimum power of 2 exceeding $3n$.

There exists the maximum superloss process $N_i(\boldsymbol{x})$ majorated by the set of pairs $(\boldsymbol{x}_s, r_s)$ produced so far. Since $\mathfrak{G}$ is a game with effective minoration, the values of $N_i(\boldsymbol{x})$ are computable. We will now approximate it with a verifiable finitary superloss process $L_i$. If $N_i(\boldsymbol{x}) = +\infty$, we let $L_i(\boldsymbol{x}) = +\infty$. For each $\boldsymbol{x}$ such that $N_i(\boldsymbol{x}) < +\infty$ we can find dyadic numbers $d'_{\boldsymbol{x}}$ and $d''_{\boldsymbol{x}}$ such that $d''_{\boldsymbol{x}} - d'_{\boldsymbol{x}} \leq \varepsilon/2$ and $d'_{\boldsymbol{x}} \leq N_i(\boldsymbol{x}) \leq d''_{\boldsymbol{x}}$. Take $L_i(\boldsymbol{x}) = d''_{\boldsymbol{x}} + 2\varepsilon|\boldsymbol{x}|$.

We have $L_i(\boldsymbol{x}\omega) - L_i(\boldsymbol{x}) \geq N_i(\boldsymbol{x}\omega) - N_i(\boldsymbol{x}) + \varepsilon$. Thus $L_i$ is a verifiable finitary superloss process.

Let us compare $L_i$ with the latest process $L_j$ that was not withheld. If $L_i(\boldsymbol{x}) \leq L_j(\boldsymbol{x})$ for all $\boldsymbol{x} \in \Omega^*$, we output $L_i$; otherwise we withhold it.

We need to show that $L(\boldsymbol{x}) = \inf_i L_i(\boldsymbol{x})$, where the infimum is taken over all $i$ such that $L_i$ is not withheld. First note that $L(\boldsymbol{x}) = \inf_{i \in \mathbb{N}} N_i(\boldsymbol{x})$. Indeed, since $N_i$ is maximal by construction, $L(\boldsymbol{x}) \leq N_i(\boldsymbol{x})$. On the other hand $L(\boldsymbol{x})$ is the infimum of $\mathrm{d}(r_s)$ such that $(\boldsymbol{x}, r_s)$ occur in the enumeration and for every $(\boldsymbol{x}, r_s)$ there is $N_i$ majorated by it. Secondly by construction we have $L_i(\boldsymbol{x}) \leq N_i(\boldsymbol{x}) + 2^{-i}$. Thus $L(\boldsymbol{x}) = \inf_i L_i(\boldsymbol{x})$. Finally note that $L_j(\boldsymbol{x}) \geq N_j(\boldsymbol{x}) + 2^{-j-k+1} \geq L(\boldsymbol{x}) + 2^{-j-k+1}$ for some $k \in \mathbb{N}$, i.e., there is a non-zero gap between $L_j$ and $L$. Therefore infinitely many $L_i$ will not be withheld. $\qquad\square\qquad\qquad\square$

**Corollary 3.** *If $\mathfrak{G}$ is a computable game with effective minoration, then there is an enumeration of upper semi-computable processes w.r.t. $\mathfrak{G}$.*

# References

[BG54]    D. Blackwell and M. A. Girshik. *Theory of Games and Statistical Decisions*. Wiley, 1954.

[CBL06]   N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[Fel68]   W. Feller. *An Introduction to Probability Theory and Its Applications*, volume I. John Wiley & Sons, Inc, 3rd edition, 1968.

[GN12]    M. Ghosh and S. Nandakumar. Predictive complexity and generalized entropy rate of stationary ergodic processes. In *Algorithmic Learning Theory*, volume 7568 of *Lecture Notes in Computer Science*, pages 365–379. Springer, 2012.

[Kal02]   Y. Kalnishkan. General linear relations among different types of predictive complexity. *Theoretical Computer Science*, 271:181–200, 2002.

[Ko91]    Ker-I Ko. *Complexity theory of real functions*. Birkhäuser, 1991.

[KV02]    Y. Kalnishkan and M. V. Vyugin. On the absence of predictive complexity for some games. In *Algorithmic Learning Theory, 13th International Conference, Proceedings*, volume 2533 of *Lecture Notes in Artificial Intelligence*, pages 164–172. Springer, 2002.

[KV08]    Y. Kalnishkan and M. V. Vyugin. The weak aggregating algorithm and weak mixability. *Journal of Computer and System Sciences*, 74(8):1228–1244, 2008.

[KVV04a]  Y. Kalnishkan, V. Vovk, and M. V. Vyugin. A criterion for the existence of predictive complexity for binary games. In *Algorithmic Learning Theory, 15th International Conference, ALT 2004, Proceedings*, volume 3244 of *Lecture Notes in Computer Science*, pages 249–263. Springer, 2004.

[KVV04b]  Y. Kalnishkan, V. Vovk, and M. V. Vyugin. Loss functions, complexities, and the Legendre transformation. *Theoretical Computer Science*, 313(2):195–207, 2004.

[KVV05]   Y. Kalnishkan, V. Vovk, and M. V. Vyugin. How many strings are easy to predict? *Information and Computation*, 201(1):55–71, 2005.

[KVV14]   Y. Kalnishkan, V. Vovk, and M. V. Vyugin. Generalised entropy and asymptotic complexities of languages. *Information and Computation*, 2014. In press.

[LV08]    M. Li and P. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer, 3rd edition, 2008.

[Roc70]   R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[VV02]    M. V. Vyugin and V. V. V'yugin. On complexity of easy predictable sequences. *Information and Computation*, 178(1):241–252, 2002.

[VV05]    M. V. Vyugin and V. V. V'yugin. Predictive complexity and information. *Journal of Computer and System Sciences*, 70:539–554, 2005.

[VW98]    V. Vovk and C. J. H. C. Watkins. Universal portfolio selection. In *Proceedings of the 11th Annual Conference on Computational Learning Theory*, pages 12–23. ACM Press, 1998.

[ZL70]    A. K. Zvonkin and L. A. Levin. The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Math. Surveys*, 25:83–124, 1970.