Effects of Kolmogorov Complexity Present in Inductive Inference as Well

Andris Ambainis[1], Kalvis Apsītis[2], Cristian Calude[3], Rūsiņš Freivalds[1], Marek Karpinski[4], Tomas Larfeldt[5], Iveta Sala[1], Juris Smotrovs[1]

[1] Department of Computer Science, University of Latvia, Raiņa bulvāris 29, LV-1459, Riga, Latvia[§], {ambainis,rusins}@cclu.lv, iveta@di.lv, smotrovs@cclu.lv

[2] Department of Computer Science, University of Maryland, College Park, MD 20742, U.S.A. kalvis@cs.umd.edu [¶]

[3] Computer Science Department, University of Auckland, Private Bag 92109, Auckland, New Zealand, cristian@cs.auckland.ac.nz [‖]

[4] Department of Computer Science, University of Bonn, 53117, Bonn, Germany, marek@cs.bonn.edu [**]

[5] Mälardalens Högskola, Institutionen för matematik och fysik, 63105 Eskilstuna, Sweden, tomas.larfeldt@mdh.se

Abstract. For all complexity measures in Kolmogorov complexity the effect discovered by P. Martin-Löf holds. For every infinite binary sequence there is a wide gap between the supremum and the infimum of the complexity of initial fragments of the sequence. It is assumed that that this inevitable gap is characteristic of Kolmogorov complexity, and it is caused by the highly abstract nature of the unrestricted Kolmogorov complexity.

We consider the complexity of inductive inference for recursively enumerable classes of total recursive functions. This object is considered as a rather simple object where no effects from highly abstract theories are likely to be met. Here, similar gaps were discovered. Moreover, the existence of these gaps is proved by an explicit use of the theorem by P. Martin-Löf.

In our paper, we study a modification of inductive inference complexity. The complexity is usually understood as the maximum of mindchanges over the functions defined by the first n indices of the numbering. Instead we consider the mindchange complexity as the maximum over the first n functions in the numbering (disregarding the repeated functions). Linear upper and lower bounds for the mindchange complexity are proved. However, the gap between bounds for all n and bounds for infinitely many n remains.

[§] Supported by Latvia Science Council Grant 96.0282

[¶] Supported by NSF Grant 9301339.

[‖] Supported by Auckland University Research Grants A18/XXXXX/62090/3414012, A18/XXXXX/62090/F3414022

[**] Supported by the Internatonal Computer Science Institute, Berkeley, California, by the DFG grant KA 673/4-1, and by the ESPRIT BR Grants 7079 and ECUS030

1 Introduction

1.1 Kolmogorov complexity

Consider $K(n)$, $C(n)$ and other functions which measure various versions of Kolmogorov complexity for words, natural numbers and other objects. These functions are not computable. Moreover, even if we are not so much interested in the Kolmogorov complexity of individual objects but we wish only to understand the order of magnitude of the growth of these functions, we discover with some surprise that even functions like

$$K^{max}(n) = \max_{0 \leq i \leq n} K(i)$$

are highly chaotic. The first result of this kind was the theorem by P. Martin Löf [15]. Let $h(n)$ be an arbitrary total recursive function such that the series

$$\sum 2^{-h(n)}$$

diverges. Then for every 0-1 valued function f it is true that for infinitely many values of n

$$K_B(f^{[n]}) \leq n - h(n)$$

This theorem showed that there does not exist a maximally complicated binary sequence every initial fragment of which would have the complexity

$$K_B(f^{[n]}) = n$$

If we consider a complicated binary sequence, at the best we have

$$K_B(f^{[n]}) = n$$

for infinitely many values of n but for infinitely many other values of n we have

$$K_B(f^{[n]}) \leq n - h(n)$$

This implies that for all the functions expressing Kolmogorov complexity we are to consider seperately the complexity *for nearly all n* (or more precisely, *for all n but a finite number of them*) and the complexity *for infinitely many n* .

Indeed, J. Bārzdiņš [2] proved many results on Kolmogorov complexity of initial fragments of binary sequences describing recursively enumerable sets, and in all these results the complexity of the recursively enumerable set *for nearly all n* was essentially higher than the complexity *for infinitely many n* . Various modifications of the problem were considered and various versions of the Kolmogorov complexity were used.

We consider the mindchange complexity in Inductive Inference. This notion seems to be much simpler than the unrestricted Kolmogorov complexity. Hence, one can expect that the gaps between the upper and lower bounds *for nearly all n* and *for infinitely many n* may not exist. However they exist! (cf. [4])

This is not a coincidence. The proofs of the lower bounds make use of the theorem by P. Martin-Löf in an explicit way.

The unexpected nature of this relation is exactly in the fact that Inductive Inference of recursively enumerable families of total recursive functions is usually considered as something very simple while effects like the one discovered by P. Martin-Löf are expected only in highly complicated abstract worlds. What really happens in the proofs of these bounds is a reduction from the Martin-Löf effect in Kolmogorov complexity theory to Inductive Inference.

A posteriori the possibility to perform such a reduction is not at all surprising. It remains to remember that the very first A.N.Kolmogorov's paper on Kolmogorov complexity [14] was named "Three approaches to the quantitative definition of information". One of these three approaches was based on predictability which is very much related to Inductive Inference. However, we are somewhat far from a complete understanding of this relation between Kolmogorov complexity and Inductive Inference.

The way to better understanding of these phenomena leads through more detailed study of both fields. One of directions for such a study is defining refinements of both notions and analysing relations between them. The first success in this direction was [1] where a different type of Kolmogorov complexity was applied to inductive inference from imprecise information to obtain new results.

In this paper, we define a new variation of inductive inference complexity. Both upper and lower bounds differ considerably from bounds for complexity measures studied before. Logarithmic bounds of [4] are replaced by linear bounds. However, the gap between upper and lower bounds still remains and is even bigger!

1.2 Inductive Inference

The complexity of Inductive Inference is usually measured in mindchanges. Mindchange complexity is a natural complexity measure, no less natural than time or tape complexity in the Theory of Computation. This way, we can consider in Inductive Inference essentially the same complexity problems as in the Theory of Computation. For instance, there are languages recognizable in polynomial time,exponential time, etc., and there are classes of total recursive functions identifiable with no mindchanges, with one mindchange, ..., with finitely many mindchanges.

Unfortunately for Inductive Inference, time complexity classe can be named by many types of functions (polynomial, exponential, double-exponential, etc.), but it is rather difficult to name complexity classes in inductive inference using functions different from constants. One can say that a class U of total recursive functions is identifiable with 7 mindchanges, but what meaning has a "logarithmic number of mindchanges"? Logarithmic of what?

For this reason, the researhers in Inductive Inference have mainly focused on the investigation of constant bounds. The first result in this direction was that[5], for any $a \in \mathbb{N}$, there is a set of functions that can be identified with $a+1$ mindchanges but cannot be identified with a mindchanges. It was followed by many other results on constant bounds which can be found in the proceedings of ALT(Algorithmic Learning Theory) and COLT(Computational Learning Theory) conferences (cf. [8, 11, 13]).

In this paper, we consider nonconstant bounds. For recursively enumerable classes U, it is easy to define complexity classes with nonconstant complexity bounds. In this case $U = \{\tau_i\}$ and "identification with logarithmic number of mindchanges" means "every $f \in U$ is identified with at most $\log n$ mindchanges, where n is the minimal τ-index of f." Complexity of identification of recursively enumerable classes was first considered in [4]. The results from [4] and many other related results are contained in the survey [6]. Recent results on this topic can be found in [1, 7].

In the general case the number of mindchanges can be a rather complicated function of the τ-index i. There can be three different kinds of lower bounds:

1) "for all n, the number of mindchanges in identification of τ_n exceeds $g(n)$,"
2) "for all n, there is a function τ_i with $i \leq n$ such that the number of mindchanges in identification of τ_i exceeds $g(n)$,"
3) "for infinitely many n, there is a function τ_n such that the number of mindchanges in identification of τ_n exceeds $g(n)$."

The lower bounds of type 1) are impractical. If at least one function f has infinitely many τ-indices, there is no hope to prove a lower bound exceeding constants. Bounds of type 2) and 3) are two different complexity measures and both of them are interesting. Each of them has a matching upper bound type. Together, we get 4 meaningful types of bounds: 2 types of upper bounds and 2 types of lower bounds.

The main results in [6] were a logarithmic upper bound of the mindchanges for arbitrary recursively enumerable class U of total recursive functions and large gaps between (upper and lower) bounds *for all initial segments of natural numbers* and *for infinitely many initial segments of natural numbers.* The most surprising effect in [6] was the fact that these large gaps were direct corollaries of a similar effect in Kolmogorov complexity proved by P. Martin-Löf [15]. (For more information on Kolmogorov complexity, refer to the textbook [14].)

We noticed that in Theorem 9 below the number of mindchanges depends not on the τ-indices involved but rather on *how many distinct functions there are among the first n functions* $\tau_1, \tau_2, \ldots, \tau_n$. However, the more sophisticated algorithms used to prove the upper bounds in [4] have this property no more. Hence, we decided to study all 4 bounds (lower and upper bounds both for all initial segments and for infinitely many initial segments) in the case when we understand the "initial sgment of length n" not as $\langle \tau_1, \ldots, \tau_n \rangle$ but rather as the minimal $\langle \tau_1, \ldots, \tau_N \rangle$ containing at least n distinct functions.

Our results can be characterized as follows. The logarithmic upper bounds[4] for the old notion of initial segments are no more valid, they are replaced by linear bounds. However, the gaps between "for all initial segments" and "for infinitely many initial segments" are even larger.

2 Technical Preliminaries

$\mathbb{N} = \{0, 1, \ldots\}$ denotes the set of natural numbers. $card(S)$ denotes the cardinality of a set S. We use standard recursion-theoretic notation (cf. textbooks [17, 18]).

We consider recursively enumerable classes U of total recursive functions with some given numbering τ.

Definition 1. A class of total recursive functions U is called recursively enumerable, if there exists a recursive function $g(i, x)$ such that

1. For each $f \in U$ there exists i such that $g(i, x) = f(x)$ for all x.
2. For each positive integer i there exists a function $f \in U$ such that $g(i, x) = f(x)$ for all x.

The function $g(i, x)$ introduces the numbering τ of the class U: $\tau = \{\tau_i\}$, where $\tau_i(x) = g(i, x)$.

An *enumerated class* is the pair (U, τ) of a class U and a numbering τ.

We fix some numbering of all tuples of non-negative integers using non-negative integers. The number of $(x_1, \ldots, x_n)$ is denoted by $\langle x_1, \ldots, x_n \rangle$. For a function f, f^n denotes $\langle f(0), \ldots, f(n) \rangle$.

The inductive inference of total recursive functions works as follows.

A black box computes some function $f \in U$ and outputs $f(0), f(1), \ldots$. An inductive inference algorithm or *strategy* F receives f^n and outputs some hypothesis. The hypothesis is either a program computing some function or an index of the function in the numbering τ (cf. Definition 1). We expect that the sequence of hypotheses produced on f^0, f^1, $f^2, \ldots$ converges to a correct Gödel number for $f(x)$ (or correct τ-index).

Definition 2. A strategy is an arbitrary partial recursive function of one argument.

$F(\langle f(0), \ldots, f(n) \rangle)$ (or $F(f^n)$) denotes the hypothesis issued by the strategy F after reading $f(0), \ldots, f(n)$.

We shall consider three types of identification:

1. Identification of Gödel numbers.
   Informally, Gödel numbers[16, 17] are programs in some universal programming language (Gödel numbering[16, 17]). The task of the strategy is to find a program computing the input function f. Before finding the correct program, the strategy is allowed to make arbitrary (but finite) number of incorrect guesses. The strategy makes a *mindchange* when it replaces one hypothesis with another. More formally, we have

   Definition 3. [9, 10] A strategy F identifies a Gödel number of a function f, if there exist N and t such that $F(f^n) = t$ for all $n > N$ and t is a correct Gödel number of $f(x)$.

Definition 4. A strategy F identifies a class U if F identifies all functions $f \in U$.

Definition 5. The number of mindchanges of F on a function f is the number of n such that $F(f^n) \neq F(f^{n+1})$. It is denoted $F^{EX}(f)$.

We define
$$F^{EX}_{U,\tau}(n) = max\{F^{EX}(\tau_i) | i \in \{1, \ldots, m\}\},$$
where m is the minimal number of the n-th different function in numbering τ. We shall consider $F^{EX}_{U,\tau}$ as a measure of complexity for strategy F.

2. Prediction (or NV-identification).
In prediction, the task of a strategy is slightly different. Instead of finding a correct program for f, the strategy has to predict the next value $f(n+1)$ of f after reading $f(0)$, ..., $f(n)$. The strategy succeds on f if it predicts correctly all but finitely many values $f(n+1)$. More formal definition follows.

Definition 6. [4, 3] A strategy F predicts a function f in the limit, if
(a) For all $n \geq 0$, $F(f^n)$ is defined.
(b) There exists an N such that for all $n > N$, $F(f^n) = f(n+1)$.

Definition 7. A strategy F predicts in the limit a class U if it predicts in the limit all functions $f \in U$.

Definition 8. The number of errors of a strategy F on a function f is the number of integers i such that $F(f^i) \neq f(i+1)$. It is denoted $F^{NV}(f)$.

Similarly, as we defined $F^{EX}_{U,\tau}$ using $F^{EX}(f)$, we now define $F^{NV}_{U,\tau}$ using $F^{NV}(f)$.

3. Identification of τ-indices.
The definitions are similar to identification of Gödel numbers. The only difference is that we require that the hypothesis $F(\langle f(0), \ldots, f(n)\rangle)$ should be an index of f in the numbering τ, not a Gödel number. We also define the counterpart of $F^{EX}_{U,\tau}$ and denote it $F^{\tau}_{U,\tau}$.

A straight-forward identification by the enumeration [9, 10] gives the first (trivial) result.

Theorem 9. *For every enumerated class (U, τ) there exists a strategy F such that $F^{\tau}_{U,\tau}(n) \leq n-1$.*

A similar result holds for identification of Gödel numbers and prediction.

Lemma 10. *For each strategy F identifying τ-indices of class (U, τ) there exists a strategy G identifying Gödel numbers such that for all n*
$$G^{EX}_{U,\tau}(n) \leq F^{\tau}_{U,\tau}(n)$$

Proof. If a τ-index is known, then Gödel number can be computed by some algorithm. So, the strategy G can work as the strategy F, compute the τ-index and then obtain the Gödel number from the τ-index. □

Lemma 11. *[6] For each strategy F predicting a class (U, τ) there exists a strategy G identifying Gödel numbers such that for all n*

$$G^{EX}_{U,\tau}(n) \leq F^{NV}_{U,\tau}(n)$$

These two lemmas show that from a complexity viewpoint, the identification of Gödel numbers is the simplest of three identification types. If we wish to prove lower complexity bounds for all three types, it is enough to prove them for the identification of Gödel numbers.

3 Linear Lower Bounds

Theorem 12. *There exists an enumerated class of recursive functions (U, τ), such that for each strategy F which identifies class (U, τ) there exist numbers c_1, c_2 $(c_1 > 0)$ such that for all n*

$$F^{EX}_{U,\tau}(n) \geq c_1 n - c_2$$

Sketch of proof. We use a subroutine which takes some strategy F and enumerates 2^m functions such that there are at most $m+1$ different functions among these 2^m, and the strategy F makes at least m mindchanges on one of these functions (or F does not identify one of these functions). This subroutine works as follows:

It simulates the strategy F on function $f(x) = 0$ and, while F has issued no conjecture, it defines the values of all 2^m functions equal to 0. (After the first step of simulation it defines $f_1(0) = \ldots = f_{2^m}(0) = 0$, after the second step $f_1(1) = \ldots = f_{2^m}(1) = 0$, and so on.)

After F has issued its first conjecture, the subroutine defines the first free value of functions equal to 0 for half of the functions and equal to 1 for the other half of the functions:

$$f_1(k) = \ldots = f_{2^{m-1}}(k) = 0$$

$$f_{2^{m-1}+1}(k) = \ldots = f_{2^m}(k) = 1.$$

Then the subroutine simulates F on these functions and, while F does not change its conjecture, defines the values of $f_1, \ldots, f_{2^m}$ equal to 0.

If F does not change its conjecture, then the conjecture is wrong either on the functions $f_1, \ldots, f_{2^{m-1}}$ or on the functions $f_{2^{m-1}+1}, \ldots, f_{2^m}$. So, if F does not change its conjecture, it does not identify some of the functions $f_1, \ldots, f_{2^m}$.

If F changes its conjecture on the functions $f_1, \ldots, f_{2^{m-1}}$, we define

$$f_{2^{m-1}+1}(x) = \ldots = f_{2^m}(x) = 0$$

for all x for which the values of $f_{2^{m-1}+1}, \ldots, f_{2^m}$ are not defined yet. So, it turns out that all 2^{m-1} functions $f_{2^{m-1}+1}, \ldots, f_{2^m}$ on which F does not change its conjecture are equal.

Then we choose k such that $f_1(k), \dots f_{2^{m-1}}(k)$ are not defined yet and define $f_i(k) = 0$ for half of the functions $f_1(k), \dots, f_{2^{m-1}}(k)$ and $f_i(k) = 1$ for the other half. We wait until F changes its conjecture on some of $f_1(k), \dots, f_{2^{m-1}}(k)$. Then we define equal to 0 all undefined values of all functions on which F does not change its conjecture. Functions on which F changes conjecture are split into two halves once again. We repeat such splitting m times and obtain 2^m functions such that there are $m+1$ different functions among them and F makes at least m mindchanges on one of them.

If F does not change its conjecture on functions $f_1, \dots, f_{2^{m-1}}$, but changes it on $f_{2^{m-1}+1}, \dots, f_{2^m}$, we proceed similarly.

Next, we define our recursively enumerable class (U, τ) on which each strategy F makes a linear number of mindchanges.

We split the set $\mathbb{N} = \{0, 1, \dots\}$ into segments $S_0, S_1, \dots$. The segment S_0 consists of the first 2^{2^0} consecutive numbers, the segment S_1 consists of the next 2^{2^1} consecutive numbers and so on. The segment S_i consists of 2^{2^i} consecutive numbers.

We take some numbering of all partial recursive strategies $F_0, F_1, \dots$. We associate the segments S_0, S_2, S_4, ... with the strategy F_0, the segments S_1, S_5, S_9, ... with the strategy F_1, the segments S_3, S_{11}, S_{19}, ... with the strategy F_2 and so on.

2^{2^i} functions $\tau_{i'}, \dots, \tau_{i'+2^{2^i}-1}$ with indices from the segment $S_i = \{i', \dots, i'+2^{2^i}-1\}$ are defined as 2^{2^i} functions constructed by the subroutine described above for the strategy F_j which is associated with S_i.

So, among these 2^{2^i} functions there are at most 2^i+1 different ones and F_j makes at least 2^i mindchanges on one of them.

If we take any n which is so large that, among first n different functions, there are all functions with indices from some S_i associated with F_j, we can conclude that F_j makes at least $n/(2^{2^j+1})$ mindchanges on one of the first n different functions. We omit the details of the proof here. □

Lemmas 10 and 11 imply that similar results hold for prediction and identification of τ-indices, too.

Similarly to Theorem 12, Theorem 13 can be proved.

Theorem 13. *There exists an enumerated class of recursive functions (U, τ), such that for each strategy F which identifies class (U, τ) there exist infinitely many n such that*

$$F_{U,\tau}^{EX}(n) \geq n - o(n)$$

Sketch of proof. The proof of this theorem is similar to the proof of Theorem 12. The only difference is that, to prove this theorem, we should take segments S_i of length $2^{2^{n^2}}$, not 2^{2^n}. □

4 Improved Inference Algorithms

Theorems 12 and 13 show that the linear upper bound given by Theorem 9 cannot be considerably improved. Theorem 13 implies that the upper bound of

Theorem 9 is the best possible upper bound "for all initial segments".

However, if we consider the upper bounds "for infinitely many initial segments", some improvements are possible. Namely, the fact that there exists a class (U, τ) such that for infinitely many "bad" n the inequality $F^{EX}_{U,\tau}(n) \geq n - o(n)$ holds (Theorem 13) does not mean that there does not exist such "good" n that $F^{EX}_{U,\tau}(n)$ is less than $\frac{n}{2}, \frac{n}{3}$ or even $\frac{n}{1000}$. We have proved

Theorem 14. *Let (U, τ) be an enumerated class. There exists a strategy G that predicts U so that*

$$G^{NV}_{U,\tau}(n) \leq \frac{n}{2} + o(n)$$

for infinitely many n.

Proof. We start with defining an auxilary function M.

Lemma 15. *There is a recursive function $M(x)$ such that*

1. *$M(0) = 1$;*
2. *For each $n \geq 0$, there are at least $M(n)^2$ different functions among $\tau_1, \ldots, \tau_{M(n+1)}$.*

Proof. $M(n+1)$ can be computed from $M(n)$ as follows:

Take some number $M > M(n)$ and compute how many different tuples are among $\langle \tau_1(0), \ldots, \tau_1(M) \rangle, \ldots, \langle \tau_M(0), \ldots, \tau_M(M) \rangle$. If there are at least $M(n)^2$ different tuples, then $M(n+1) = M$. Otherwise, increase M by 1 and try again.

To compute $M(n)$, we first compute $M(0)$, then compute $M(1)$ from $M(0)$, then compute $M(2)$ and so on, until $M(n)$ is computed. □

We use the function $M(n)$ to define 2 identification strategies:

The strategy F_1. Find the smallest i such that $f(0) = \tau_i(0), \ldots, f(n) = \tau_i(n)$. and output $\tau_i(n+1)$ as a prediction of $f(n+1)$.

The strategy F_2.

1. Find the smallest i such that $f(0) = \tau_i(0), \ldots, f(n) = \tau_i(n)$.
2. Find an m such that $M(m) < i \leq M(m+1)$.
3. Search for $j \leq M(m+1)$ such that

$$\langle \tau_j(0), \ldots, \tau_j(n) \rangle = \langle f(0), \ldots, f(n) \rangle$$

and $\tau_i(n+1) \neq \tau_j(n+1)$. If there is no such j, then give $\tau_i(n+1)$ as the prediction for $f(n+1)$. Otherwise, give $\tau_j(n+1)$ as the prediction.

$N(m)$ denotes the number of different functions among $\tau_1, \tau_2, \ldots, \tau_{M(m)}$.

Lemma 16. *For any $i \in \{1, 2\}$ and $m \in \mathbb{N}$, F_i makes at most $N(m)$ prediction errors on any of the functions $\tau_1, \ldots, \tau_{M(m)}$.*

Proof. If F_1 or F_2 predicts $\tau_i(n)$ incorrectly, there exists an another function τ_j such that $j \leq M(m)$, $\tau_i(0) = \tau_j(0)$, $\tau_i(1) = \tau_j(1)$, …, $\tau_i(n-1) = \tau_j(n-1)$, but $\tau_i(n) \neq \tau_j(n)$.

There are $N(m)$ different functions among $\tau_1, \tau_2, \ldots, \tau_{M(m)}$. So, there are at most $N(m)$ numbers n such that one of our strategies makes an error predicting $\tau_i(n)$. This proves the lemma. □

Lemma 17. *For each m one of F_1 and F_2 makes at most $\frac{N(m)}{2} + N(m-1)$ errors on each of the functions $\tau_1, \ldots, \tau_{M(m)}$.*

Proof. Let $i \in \{1, \ldots, M(m)\}$. If $i \leq M(m-1)$, each strategy makes at most $N(m-1)$ errors (Lemma 16). It remains to consider $i > M(m-1)$.

Let $n(i)$ be the largest number such that

$$\tau_i(0) = \tau_j(0), \ldots, \tau_i(n(i)) = \tau_j(n(i)).$$

for some $j \leq M(m-1)$. By Lemma 16, each strategy makes at most $N(m-1)$ errors on τ_j. Hence, each strategy makes at most $N(m-1)$ errors predicting $\tau_i(0)$, $\tau_i(1)$, …, $\tau_i(n(i))$. It remains to estimate the number of errors while $\tau_i(n(i)+1)$, $\tau_i(n(i)+2)$, … are predicted. We consider two cases:

1. For each $i \in \{1, \ldots, M(m)\}$, F_1 makes at most $\frac{N(m)}{2}$ errors when $\tau_i(n(i)+1)$, $\tau_i(n(i)+2)$, … are predicted.
   Lemma is evident in this case.
2. There exists an $i \in \{1, \ldots, M(m)\}$ such that F_1 makes more than $\frac{N(m)}{2}$ errors predicting $\tau_i(n(i)+1), \tau_i(n(i)+2), \ldots$.
   Let E be the number of errors of F_1 on $\tau_i(n(i)+1)$, $\tau_i(n(i)+2)$, …. Let $n_1, \ldots, n_E$ be the numbers (in increasing order) such that $\tau_i(n_1)$, … $\tau_i(n_E)$ are predicted incorrectly. We have $E > \frac{N(m)}{2}$.
   F_1 searches for the smallest j satisfying $\tau_j(0) = \tau_i(0)$, $\tau_j(1) = \tau_i(1)$, …, $\tau_j(n-1) = \tau_i(n-1)$ and outputs $\tau_i(n)$ as its prediction. So, if it makes predicts $\tau_i(n_k)$ incorrectly, there exists $j_k < i$ such that $\tau_i(0) = \tau_{j_k}(0)$, …, $\tau_i(n_k - 1) = \tau_{j_k}(n_k - 1)$, but $\tau_i(n_k) \neq \tau_{j_k}(n_k)$. We find the smallest such j_k for any $k \in \{1, \ldots, E\}$. Then, $j_1 < j_2 < \ldots < j_E$.

   Proposition 18. *Let $k \in \{1, \ldots, E\}$, $n > n(i)$ and $n \neq n_k$. If F_2 predicts $\tau_{j_k}(n)$ incorrectly, there exists $j \in \{1, \ldots, M(m)\}$ such that*
   (a) $\tau_j(0) = \tau_{j_k}(0)$, …, $\tau_j(n-1) = \tau_{j_k}(n-1)$, but $\tau_j(n) \neq \tau_{j_k}(n)$, and
   (b) τ_j is not equal to any of $\tau_{j_1}, \ldots, \tau_{j_E}$.

   Proof. We consider two cases:
   (a) j_k is the smallest τ-index consistent with $\tau_{j_k}(0), \ldots, \tau_{j_k}(n-1)$.
   If F_2 makes an error, there exists $j \leq M(m)$ such that $\tau_j(0) = \tau_{j_k}(0)$, …, $\tau_j(n-1) = \tau_{j_k}(n-1)$, but $\tau_j(n) \neq \tau_{j_k}(n)$.
   Then, $j > j_k$. $\tau_j \neq \tau_{j_1}$, …, $\tau_j \neq \tau_{j_{k-1}}$, because $j_1 < \ldots < j_{k-1} < j_k$.
   By the definition of $j_1, \ldots j_E$, $\tau_{j_k}(n_k)$ is the first different value of τ_{j_k} and $\tau_{j_{k+1}}$, …, τ_{j_E}. (It follows from $\tau_{j_k}(n_k) \neq \tau_i(n_k)$ and $\tau_i(n_k) = \tau_{j_{k+1}}(n_k) = \ldots = \tau_{j_E}(n_k)$.)

However, n is the first different value of τ_i and τ_{j_k} and $n \neq n_k$. Hence, $\tau_j \neq \tau_{j_{k+1}}, \ldots, \tau_j \neq \tau_{j_E}$.

(b) j_k is not the smallest τ-index consistent with $\tau_{j_k}(0), \ldots, \tau_{j_k}(n-1)$.
If F_2 makes an error in this case, there exist two different functions τ_{i_1}, τ_{i_2} with τ-indices smaller than j_k such that both τ_{i_1} and τ_{i_2} are consistent with the fragment of τ_j seen by the strategy and $\tau_{i_1}(n) \neq \tau_{i_2}(n)$. We select two such functions with the smallest i_1 and i_2. Without the loss of generality, we assume that $i_1 < i_2$.
F_2 predicts $\tau_{i_2}(n)$. This prediction is incorrect. τ_{i_2} is the sought function τ_j. τ_{i_2} is not equal to $\tau_{j_{k+1}}, \ldots, \tau_{j_E}$, because $i_2 < j_k$, but $j_k < j_{k+1} < \ldots$.
Next, we prove that τ_{i_2} is not equal to $\tau_{j_1}, \ldots, \tau_{j_{k-1}}$.
If $i < k$, then the first different values of τ_{j_k} and τ_{j_i} are $\tau_{j_k}(n_i)$ and $\tau_{j_i}(n_i)$. So, if $\tau_{i_2} = \tau_{j_i}$, then $n = n_i$. By the definition of j_i, j_i is the smallest τ-index consistent with $\tau_{j_i}(0), \ldots, \tau_{j_i}(n_i - 1)$. Hence, if $n = n_i$, then the function τ_{i_1} is equal to τ_{j_i}, but τ_{i_2} is not.

□

So, the number of errors of F_2 on τ_{j_k} is less than or equal to the number of different functions among $\tau_1, \ldots, \tau_{M(m)}$ such that none of them is equal to $\tau_{j_1}, \ldots, \tau_{j_E}$ plus 1. Hence, the number of errors on $\tau_{j_k}(n(i)+1)$, $\tau_{j_k}(n(i)+2), \ldots$ is at most

$$N(m) - E + 1 < N(m) - \frac{N(m)}{2} + 1 = \frac{N(m)}{2} + 1.$$

If $i \in \{1, \ldots, M(m)\}$, but $i \neq j_k$, the proof that F_2 makes at most $\frac{N(m)}{2}$ errors is similar.

□

By Lemma 17, there is $i \in \{1, 2\}$ such that

$$(F_i)^{NV}_{U,\tau}(N(m)) \leq \frac{N(m)}{2} + N(m-1)$$

for infinitely many m. By the definition of $M(n)$,

$$N(n) \geq M^2(n-1) \geq N^2(n-1),$$

$$(F_i)^{NV}_{U,\tau}(N(m)) \leq \frac{N(m)}{2} + N(m-1) \leq \frac{N(m)}{2} + \sqrt{N(m)} = \frac{N(m)}{2} + o(N(m)).$$

We have proved that, for infinitely many m, F_i makes at most $\frac{N(m)}{2} + o(N(m))$ prediction errors on the first $N(m)$ different functions from $\tau_1, \tau_2, \ldots$. □

Corollary 19. *For each class (U, τ) there exists a strategy F such that*

$$F^{EX}_{U,\tau}(n) \leq \frac{n}{2} + o(n)$$

for infinitely many n.

Proof. Follows from Theorem 14 and Lemma 11. □

For τ-indices the proof desribed above is not valid. As we show in Section 5, the argument similar to Theorem 14 cannot give similar result for τ-indices.

5 Lower Bounds for Sets of Strategies

In proof of Theorem 9 we constructed a strategy which identified any class (U, τ) making no more than n mindchanges on each of first n functions. In Theorem 14 we constructed 2 strategies, one of which identified class (U, τ) making no more than $\frac{n}{2} + o(n)$ mindchanges. What can we do with 3 or more strategies? Can we achieve better upper bounds?

No, we cannot. Methods from proof of Theorems 9 and 14 cannot be improved so far that they will give better upper bounds. To prove this, we need to change the definition of strategy.

Strategy, as we defined earlier, is a partial recursive function of one argument. Identification algorithms of Theorems 9 and 14 are more general, because they use two arguments: the class (U, τ) (by looking up the values of $\tau_i(j)$) and the previous values of function $f(x)$ ($\langle f(0), \ldots, f(n)\rangle$). The formal definition in this case is as follows:

Definition 20. A uniform strategy is partial recursive function of two arguments $F(l, x)$. The argument l is a Gödel number of the function $g(i, x)$ (Definition 1), the argument x is $\langle f(0), \ldots, f(n)\rangle$.

This definition allows the strategy not only to compute the values of $\tau_i(j)$, it also allows to analyse the algorithm according to which these values are computed. However, this does not help in improving Theorem 14.

Theorem 21. *For each finite set of uniform strategies $F_1, \ldots, F_m$ identifying Gödel numbers, there exists a class (U, τ) containing infinitely many functions such that for each strategy F_i $(i = 1, \ldots, m)$ which identifies class U, $(F_i)^{EX}_{U,\tau}(n) \geq \frac{n}{2} - o(n)$ for all n.*

Proof. Suppose a team of general strategies $F_1, \ldots, F_m$ is given.

At first we construct a class (U, τ) by giving an algorithm for a function $g(i, x, l)$ with an additional parameter l (see Definition 1). We give l as the first argument to the strategies - $F_k(l, \langle f^{[n]}\rangle)$, $1 \leq k \leq m$.

We divide the functions $\tau_0, \tau_1, \ldots$ in two infinite disjoint subclasses: $\rho_0, \rho_1, \ldots$ and $\sigma_0, \sigma_1, \ldots$. σ_i is $\tau_{2^{2^i}}$, the remaining are ρ-functions, with growing order of indices.

The algorithm for the class is as follows:

INIT.

$F \leftarrow \{1, \ldots, m\}$

For all $S \subseteq F$ assign

$$Prev[S] \leftarrow \begin{cases} 0, \text{if } card(S) \geq m/2, \text{ or } card(S) = m/2 \text{ and } 1 \in S \\ 1, \text{otherwise} \end{cases}$$

$t \leftarrow 0$

$x_0 \leftarrow 0$

$I \leftarrow \{0, 1, 2, \ldots\}$

Go to STEP 0.

STEP r.
Start the following parallel processes:

1. Define $\rho_{2^{r+1}j+t}(x_0) = 0$ and $\rho_{2^{r+1}j+t+2^r}(x_0) = 1$ for all j.
   Define $\rho_{2^r j+t}(x) = 0$ for $x = x_0 + 1, x_0 + 2, \ldots$.
   $A \leftarrow \emptyset$
   Go to 2.
2. On functions $\{\sigma_j \mid j \in I\}$ construct a counter-class to the strategies $F_k \in A$, similar to the class we are building now. Counter-class to an empty set of strategies is a class of constant 0 functions. Constructing counter-class, record σ-indices of those functions that are already defined on at least one point in set J.
   Go to 3.
3. For each $k \notin A$ start computing, on which of the two branches $\langle \rho_t^{[x]} \rangle$ and $\langle \rho_{t+2^r}^{[x]} \rangle$, $x = x_0, x_0 + 1, \ldots$, strategy F_k changes its current hypothesis (on STEP 0 we also compute the hypothesis made by F_k on the empty segment $\langle \rangle$, to establish a mindchange).
   If that happens for some $F_{k'}$, then assign to $y_{k'}$ the value of argument x, on which it made the mindchange, and go to 4.
4. Cancel the construction of counter-class on σ-functions.
   Define $\sigma_j(x) = 0$ for all $j \in J$ and for all x, if $\sigma_j(x)$ is undefined yet.
   $I \leftarrow I \setminus J$
   $A \leftarrow A \cup \{k'\}$
   If $A = F$, go to 5. Otherwise go to 2.
5. Cancel definition of $\rho_{2^r j+t}(x)$ for $x > x_0$.
   Assign to y_0 the maximal value of argument, for which at least one of these functions is already defined.
   $B_0 \leftarrow \left\{ k \mid F_k \text{ made the last mindchange on the branch } \langle \rho_t^{[y_k]} \rangle \right\}$
   $B_1 \leftarrow A \setminus B_0$
   If $Prev[B_0] = 0$, then $u \leftarrow t + 2^r$, otherwise $u \leftarrow t$ and $t \leftarrow t + 2^r$.
   Define $\rho_{2^{r+1}j+u}(x) = 0$ for all j and for all $x > x_0$, if these values are undefined yet.
   $Prev[B_0] \leftrightarrow Prev[B_1]$
   Go to 6.
6. $y \leftarrow 1 + \max\limits_{0 \le k \le m} y_k$
   Define $\rho_{2^{r+1}j+t}(x) = 0$ for all j and $x_0 < x < y$.
   $x_0 \leftarrow y$
   Go to STEP $r + 1$.

It follows from the recursion theorem that there exists l' such that $g(i, x, l') = \varphi_{l'}(i, x)$, where φ is the Gödel numbering used in Definition 20. Thus this function (defining the class (U, τ)) gives its own Gödel number as the first parameter to the strategies.

In this algorithm F is the set of the indices of all strategies; A enumerates those strategies that have made a mindchange at the current step; B_0 enumerates those strategies that made this mindchange on the branch that had $\rho_j(x_0) = 0$; B_1 – the same for the branch that had $\rho_j(x_0) = 1$. I is the set of those σ-indices that we have not spoiled at previous steps; it always contains all the natural numbers, except a finite amount of them. In J we record those σ-indices, whose values we have defined and thus spoiled, if we have to cancel the current process on σ-functions. By y we establish that the next x_0 is large enough, so that all the fixed mindchanges have happened on smaller initial segments and all values of $\rho_{2^{r+1}j+t}$ are undefined at this point.

At each step we choose one of the two sets B_0 and B_1, and make the strategies with indices from it to do a new mindchange on $\rho_{2^{r+1}j+t}$, i. e. on all the different functions, considered at this step, except one that equals to $\rho_{2^{r+1}j+u}$ for all j. In the array *Prev* we record for each subset of F, if it was chosen the last time it was encountered, and we choose the subset iff it was not chosen this last time. If such subsets are encountered the first time, we choose the largest one.

Suppose each of m strategies makes a mindchange on at least one of two branches at each step. As in two subsequent choices for the same subsets all m strategies are chosen, and at the first choice for these subsets the largest subset is chosen, we conclude that the mean of the number of strategies chosen at one step is at least $m/2$. As at each new choice we introduce exactly one new different function, the total number of mindchanges made by all m strategies on the first $n+1$ different ρ-functions is at least $\frac{m}{2} \cdot n$. One strategy can have greater number of mindchanges than another only on account of those dividings in subsets that have occurred an odd number of times. As the number of dividings of an m element set in two subsets is 2^{m-1}, each strategy makes at least

$$\frac{(n \cdot m)/2}{m} - 2^{m-1} = \frac{n}{2} - 2^{m-1}$$

mindchanges on one of the first $n+1$ ρ-functions for all n. As at each choice every second of the considered ρ-functions is made to be one and the same function, but other functions (that will be considered at the next step) are made different from it, the number of different functions among the first n ρ-functions is $\lfloor \log_2 n \rfloor + 1$ or $\lceil \log_2 n \rceil + 1$, while the number of σ-functions inbetween them is $o(\log_2 n)$, thus we get the needed estimation.

Now suppose that at some step not all of the strategies make a mindchange on at least one of the branches. Then algorithm remains at this step forever. The ρ-functions are gradually defined equal to 0 for all $x > x_0$, so there is only a finite amount of different functions among them, and the complexity function is decided on σ-functions, on which a similar class is built for a smaller amount of strategies (for those that *made* a mindchange). Thus we get the needed, applying this theorem recursively (the basis for recursion is given by 0 strategies and a class consisting of constant 0 functions). □

From Lemma 11 it follows that similar theorem holds for the prediction, too. For the identification of τ-indices even the stronger result holds.

Theorem 22. *For each finite set of uniform strategies $F_1, \ldots, F_m$ identifying τ-indices there exists a class (U, τ) containing infinitely many functions such that* *for each strategy F_i $(i = 1, \ldots, m)$ which identifies class U : $F^{\tau}_{U,\tau}(n) \geq n - o(n)$ for all n.*

Sketch of proof. The proof is rather similar to the previous one. Consider uniform strategies $F_1, \ldots, F_m$. The class U, as previously, consists of "usual" ρ_n and "rare" σ_n functions (the amount of rare functions among the first n functions does not exceed $o(n)$ this time), and we use the recursion theorem on it to make the strategies work on itself.

We define $\rho_k(0) = 0$ and $\sigma_k(0) = 1$ for all k, to make these subclasses explicitly different. Then we gradually define $\rho_n(x) = 0$ for $x = 1, 2, \ldots$ and at the same time wait for the strategies to make their hypotheses on the initial segment $\langle 0 \rangle$. While not all of them have given out a hypothesis, we construct a class of the rare functions similar to the one we are defining now - for those strategies that have given out their hypotheses.

Suppose that all m strategies have given out their hypotheses: $h_1, \ldots, h_m$. Then we choose some x large enough and define: $\tau_{h_i}(x) = 1$ for those i, for which τ_{h_i} is a ρ-function. We also define $\rho_k(x) = 1$, where k is the least index, for which we have not yet defined ρ_k to infinity. On all other points we define these functions by 0 (so they are equal). Thus all m strategies have given a correct hypothesis only on one function. On all other functions we repeat this procedure by giving the strategies $0^x, 0^{x+1}, \ldots$ as inputs and waiting on them to change their previous (wrong) hypotheses.

If all strategies change their wrong hypotheses on all initial segments 0^n, then all strategies make at least n on $(n+1)$-st different ρ-function. If some strategy does not change its wrong hypothesis, it does not identify the class U. In this case, the amount of diffrent ρ-functions is finite, and the subclass for the strategies that worked properly on ρ-functions is constructed on σ-functions. □

This theorem shows that the counterpart of Theorem 14 for τ-indices cannot be proved similarly as Theorem 14.

So, attempts to generalize Theorem 14 meet serious difficulties. If we wish to prove $\frac{n}{c}$ upper bound where $c > 2$, we need to consider an infinite number of strategies. In all proofs on the complexity of inductive inference known to us, only finite number of strategies are considered. So, some new methods are needed. On the other hand, if there exists a class such that each strateegy needs $\frac{n}{2}$ mindchanges when working on this class, to prove it we also need some new methods. The idea of proof from Theorem 12 will not work in this case.

References

1. A. Ambainis and J.Smotrovs, *Enumerable classes of total recursive functions: complexity of inductive inference.* Proceedings of the AII'94, Lecture Notes in Computer Science, 872(1994), pp. 10-25
2. J. Bārzdiņš. *On the frequency solution of the membership problem for recursively enumerable sets.* Trudi of V.A.Steklov Mathematics Institute, Moscow, 1973, pp. 133-160

3. L.Blum and M.Blum, *Towards a mathematical theory of inductive inference*, Information and Control, 28(1975), pp.122-155
4. J. Bārzdiņš and R. Freivalds. *On the prediction of general recursive functions.* Soviet Math. Doklady, 13 (1972), pp.1224-1228.
5. J. Case, C. Smith, *Comparison of identification criteria for machine inductive inference.* Theoretical Computer Science, 25(1983), pp.193-220
6. R. Freivalds, J. Bārzdiņš, K. Podnieks. *Inductive inference of recursive functions: complexity bounds.* Baltic Computer Science. Lecture Notes in Computer Science, 502 (1991), pp.111-155.
7. R. Freivalds and A.C. Hoffman. *An inductive inference approach to classification.* Journal of Experimental and Theoretical Artificial Inteligence, 6 (1994), pp.63-72.
8. R. Freivalds, C. H. Smith, *The role of procrastination in machine learning.* Information and Computation, vol. 107(1993), pp. 237-271
9. E.M.Gold, *Limiting recursion.* Journal of Symbolic Logic, 30(1965), pp.28-48
10. E.M.Gold, *Language identification in the limit.* Information and Control, 10(1967), pp.447-474
11. Efim Kinber and Frank Stephan. Language learning from texts: mind changes, limited memory and monotonicity. *Proc. 8th Annu. Conf. on Comput. Learning Theory,ACM Press, New York, NY,* pp. 182–189, 1995
12. A.N. Kolmogorov, *Three approaches to the quantitative definition of information.* Problems of Information Transmission, vol. 1, No. 1, 1965, pp.1-7.
13. S. Lange, T. Zeugmann, *Trading monotonicity demands versus mind changes.* Proceedings of EuroCOLT'95, Lecture Notes in Computer Science, vol. 904(1995), pp. 125-139
14. M. Li, P. Vitanyi, *Introduction to Kolmogorov Complexity and Its Applications.* Springer-Verlag, 1993
15. P. Martin-Löf. *On the definition of random sequence.* Information and Control, 9(1966), pp.602-619
16. H. Rogers, Jr. *Gödel numberings of partial recursive functions.* Journal of Symbolic Logic, 23:331-341, 1958
17. H. Rogers, Jr. *Theory of Recursive Functions and Effective Computability.* McGraw-Hill, 1967. Reprinted, MIT Press, 1987
18. R. Soare, *Recursively Enumerable Sets and Degrees.* Springer-Verlag, 1987