Forschungsberichte
Mathematische Logik

Noisy Inference and Oracles

Frank Stephan

Bericht Nr. 19 Januar 1996

Universität Heidelberg
Mathematisches Institut

Home page: http://math.uni-heidelberg.de/logic

Noisy Inference and Oracles

Frank Stephan*

Abstract

A learner noisily infers a function or set, if every correct item is presented infinitely often while in addition some incorrect data ("noise") is presented a finite number of times. It is shown that learning from a noisy informant is equal to finite learning with K-oracle from a usual informant. This result has several variants for learning from text and using different oracles. Furthermore, partial identification of all r.e. sets can cope also with noisy input.

1 Introduction

Many scientific or mathematical problems are only solved numerically and the correctness of the solution depends on the computer power available. Scientists therefore have to trust the data, which may be incorrect; they can not wait until better computer are available 10 years later. So they make up their current theories from uncertain data.

Modeling the development of science as a long process, it has to be taken into account that errors in today's simulation-data are discovered in 10 or 20 years, when the experiments will be done using more powerful computers, which e.g. enable to work with smaller grids and higher precision. Discovering an error results in a revision of the theory, if necessary. So there is an infinite sequence of data, where each correct item, say each correct outcome of a given simulation process, occurs infinitely often, while the incorrect items only occur a finite number of times; from this sequence of data scientists generate a finite sequence of theories; the last of these theories should be correct.

---

*Mathematisches Institut, Ruprecht-Karls-Universität Heidelberg, Im Neuenheimer Feld 294, 69120 Heidelberg, Germany, EU, `fstephan@math.uni-heidelberg.de`. Supported by the Deutsche Forschungsgemeinschaft (DFG) grant Am 60/9-1.

In inductive inference, the topic of exploration is typically modeled by a function or a set to be learned. And a given simulation is modeled as a pair (x, y) where x is the input and y the output. The supplied information (x, y) is correct iff $f(x) = y$. The sequent theories guessed by the scientists are modeled as programs, which are intended to compute f. There are various approaches to inference from faulty data [2, 5, 7, 17]; Jain [7] distinguishes three basic types of concepts for learning in the limit from faulty data: (a) the learner receives some faulty data together with the information on the concept to be learned, (b) all data is correct but some information on the concept is never presented and (c) the combination of both concepts.

Many concepts of learning from faulty data have the disadvantage, that it is impossible to define the object to be learned only from the input to the learner. If e.g. in case (a) the informations $(0, 0)$ and $(0, 1)$ are both supplied to the learner, then it is impossible to know which one is correct, i.e., whether $f(0) = 0$ or $f(0) = 1$. The same holds if according to (b) no statement of the form $(0, y)$ is made at all. The learner therefore has to overcome this gap by a priori knowledge about f, e.g. that always $f(0) = f(1)$ and $f(1)$ is specified uniquely on the information supplied to the learner.

The model considered here solves this problem by presenting the correct information infinitely often while the incorrect one occurs only finitely often, i.e., $f(x) = y$ iff (x, y) occurs infinitely often on the learner's input-tape. During the inference process, the learner still has the problem not to know whether the current input is correct, but in the limit it turns out which data is correct and which is incorrect; so the learner needs less a priori knowledge for learning in the limit.

So the noisy inference considered here can be put into Jain's first category (a), i.e., learning with additional faulty information. The noisy text in this context are a combination of the intrusion texts as defined by Osherson, Stob and Weinstein [17, Exercise 5.4.1 E] which may contain finite additional false words and the fat texts [17, Section 5.5.4] in which each item appears infinitely often. Learning from texts and learning from fat texts are equivalent models [17, Proposition 5.5.4 A]. But the intrusion texts are more restrictive than noisy texts as defined below since the class $\{\{0\}, \{1\}\}$ can be learned from very noisy text but not from intrusion texts. Now the concepts are presented in detail:

Definition 1.1 A *noisy informant* for a function f is an infinite sequence T such that every pair $(x, f(x))$ occurs infinitely often in this sequence while for each x pairs (x, y) with $y \neq f(x)$ occur only finitely often. A noisy informant for a set is a noisy informant for its characteristic function.

A *noisy text* for a set L is an infinite sequence $T = \{w_i\}_{i \in \omega}$ in which every $x \in L$ occurs infinitely often, i.e., $(\forall x \in L)\,(\exists^\infty i)\,[w_i = x]$, while only finitely often some $x \notin L$ occurs, i.e., $(\forall^\infty i)\,[w_i \in L]$.

An IIM M infers $\mathcal{L}$ noisily (from text or informant), iff for every $L \in \mathcal{L}$ and on every noisy text/informant $T = \{w_i\}_{i \in \omega}$ for L, M converges to an index for L, i.e., iff $M(w_0 w_1 \ldots w_n) = e$ for some index e for L and for almost all n. These criteria are denoted by NoisyTxt and NoisyInf.

A *very noisy text* for a set L is an infinite sequence T such that $x \in L$ iff x occurs infinitely often in T. A *very noisy informant* is a very noisy text for the graph of a function. Every noisy text is also very noisy but not vice versa.

A further important concept of learning theory in this paper is that of learning without mindchanges: An IIM learns *finitely* (FIN) a language or function iff it makes exactly one guess during the inference process and this guess is correct.

An IIM M learns *dual strongly monotonic* ($\mathrm{SMon}^{\mathrm{d}}$) iff the guessed languages form a descending sequence, i.e., $W_{M(\sigma)} \supseteq W_{M(\sigma\tau)}$ for all strings σ, τ [11, 14].

Osherson, Stob and Weinstein [17] give an overview and further details on inductive inference.

The main recursion-theoretic definitions and notations can be found in the books of Odifreddi [16] and Soare [18]. Nevertheless some basic facts are included for the convenience of the reader:

A set is recursive enumerable (r.e.) iff there is an algorithm which outputs a sequence just containing all elements of the set, i.e., which outputs a text T of the set – this text may contain the symbol # to avoid undefined output in the case of $\emptyset$. RE denotes the class of all r.e. sets, REC that of all recursive functions.

A set A is Turing reducible to B ($A \leq_T B$) if A can be computed via a machine which knows B, i.e., which has an infinite database which supplies for each x the information whether x belongs to B or not. Such a database is called an *oracle* and the question "$x \in B$?" a *query to* B. The class $\{A : A \equiv_T B\}$ is called the Turing degree of B where $A \equiv_T B$ means that both, $A \leq_T B$ and $B \leq_T A$ hold. Given two sets A and B, the Turing degree of the join $A \oplus B = \{2x : x \in A\} \cup \{2x+1 : x \in B\}$ is the least upper bound of the Turing degrees of A and B. K denotes the halting problem, i.e., the set $\{x : \varphi_x(x)\downarrow\}$. This notion can be relativized: $A' = \{x : \varphi_x^A(x)\downarrow\}$ is the halting problem relative to A where φ_x^A is the x-th recursive function equipped with the oracle A. Also an IIM may use an oracle, e.g., FinTxt$[A]$ denotes the class of all families of languages learnable finitely via an IIM which uses the oracle A.

A set A has high Turing degree if $K' \leq_T A'$, i.e., if the halting problem relative to K can be solved using the halting problem relative to A. The high Turing degrees are also the degrees of the sets A such that there is a function f computable in A which dominates every recursive function g, i.e., which satisfies $(\forall^\infty x)\,[g(x) < f(x)]$. Some kind of counterpart are the hyperimmune-free degrees: they are the degrees of all sets A such that any function f computable relative to A is dominated by a recursive function, i.e., $(\forall f \leq_T A)\,(\exists g \in \mathrm{REC})\,(\forall x)\,[f(x) \leq g(x)]$.

The 1-generic sets are those that either meet or strongly avoid each recursive set of strings: If A is 1-generic and W is a recursive set of strings, then there is a prefix $\sigma \preceq A$ such that either $\sigma \in W$ ("A meets W") or $\sigma\tau \notin W$ for all τ ("A strongly avoids W"). The interested reader may find more information about 1-generic sets in Jockusch's paper [10] or Soare's book [18, A.VI.3.6-9].

Section 2 deals with learning from noisy informant; this concept can be identified with finite learning from informant: NoisyInf = FinInf$[K]$. This connection relativizes and motivates looking for similar relations w.r.t. noisy learning and oracles. Further it provides an easy characterization for the inference degrees of noisy inference — the inference degree of an oracle A is $\{B : \mathrm{NoisyInf}[B] = \mathrm{NoisyInf}[A]\}$ [4]. In particular section 3 provides many connections for learning from noisy text, but it does not find an equivalence between noisy learning and an already well-known concept. Section 4 looks for connections between the text and informant version of noisy learning. Section 5 considers the case, where the family to be learned is uniformly recursive. Section 6 deals with partially identification. Osherson, Stob and Weinstein [17] showed that the class of all r.e. sets is partially identifiable from text; the same holds for noisy text and noisy informant, but not for very noisy text.

2 Inference From Informant

The main result of this section is, that finite learning from informant with K-oracle equals learning from noisy informant. This relation motivates the study of connections between noisy inference and finite inference with oracles.

Theorem 2.1 FinInf$[K]$ = NoisyInf.

Proof: NoisyInf $\subseteq$ FinInf$[K]$: Assume that M is a recursive IIM which learns a family $\mathcal{L}$ of sets from noisy informant. A string σ is called α-consistent iff

$$(\forall x < |\alpha|)\,[((x,y) \text{ occurs in } \sigma) \;\Rightarrow\; y = \alpha(x)].$$

Now the following FinInf$[K]$ IIM N infers $\mathcal{L}$:

$$N(\alpha) = \begin{cases} e & \text{if } e = N(\sigma\tau) \text{ for some string } \sigma \text{ of length up to } |\alpha| \\ & \text{and for all } \alpha\text{-consistent strings } \tau; \\ ? & \text{otherwise, i.e., there is no such } \sigma. \end{cases}$$

In short: N searches — using K-oracle — some kind of locking sequence σ and then outputs $e = M(\sigma)$.

Assume now that on the inference of U, $N(\alpha)$ outputs $e = N(\sigma)$ for some $\alpha \preceq U$. Let $w_{\langle x,y\rangle} = (x, U(x))$ for all x, y. Now $\sigma w_0 w_1 w_2 \ldots$ is a noisy informant for U and thus M has to converge on this informant to a correct index. Since all strings $\tau_n = w_0 w_1 \ldots w_n$ are α-consistent, $e = N(\sigma\tau_n)$ and e is an index for U. Thus the first guess is already correct and no mindchange is necessary.

So it remains to verify that N always converges. Assume that N does not converge, i.e., for every σ and every $\alpha \preceq U$ there is an α-consistent τ with $M(\sigma\tau) \neq M(\sigma)$. Let $\sigma_0 = (0,0)$. For $n = 1, 2, \ldots$, there are $(U(0), U(1), \ldots, U(n))$-consistent strings τ_n such that

$$\begin{aligned} M(\sigma_n) &\neq M(\sigma_{n-1}) \text{ where} \\ \sigma_n &= \sigma_{n-1}\,(0, U(0))\,(1, U(1))\,\ldots\,(n, U(n))\,\tau_n. \end{aligned}$$

It follows that $T = \lim_n \sigma_n$ is a noisy informant for U and that M diverges on T, a contradiction. Thus N infers every $U \in \mathcal{L}$ and NoisyInf $\subseteq$ FinInf$[K]$.

FinInf$[K]$ $\subseteq$ NoisyInf: Let N^K be a FinInf$[K]$ IIM for some family $\mathcal{L}$. A string α is called σ-consistent iff for all $x < |\alpha|$ the pair $(x, \alpha(x))$ occurs in σ at least as often as any other pair (x, y). Now

$$M(\sigma) = \begin{cases} N^{K_{|\sigma|}}(\alpha) & \text{for the shortest } \sigma\text{-consistent } \alpha \\ & \text{which satisfies } N^{K_{|\sigma|}}(\alpha)\downarrow \neq ? \text{ within } |\sigma| \text{ steps.} \\ ? & \text{otherwise, i.e., there is no such } \alpha. \end{cases}$$

M NoisyInf infers $\mathcal{L}$: Let $\alpha \preceq U$ be the shortest string with $N^K(\alpha) = e \neq ?$. Then α is σ-consistent for almost all $\sigma \preceq T$ of any given noisy informant T for U. Since any α' which is σ-consistent for infinitely many $\sigma \preceq T$ either satisfies $\alpha' \prec \alpha$ or $\alpha' \succ \alpha$, either $N^{K_{|\sigma|}}(\alpha')\downarrow = ?$ holds for almost all these σ or

α' is not considered since already $\alpha \prec \alpha'$ satisfies all necessary requirements. Thus M converges on every noisy informant T of U to the same index e which N infers. ∎

It is easy to see that the proof holds as well for learning functions as well for learning r.e. sets. Further the proof relativizes. Since $\mathrm{FinInf}[A] \subseteq \mathrm{FinInf}[B] \Leftrightarrow A \leq_T B$ [4, Theorem 6.36], the relativized version of this theorem also characterizes the inference degrees for noisy informant. In the non-relativized world, NoisyInf is between FinInf and LimInf.

Corollary 2.2
(a) $\mathrm{NoisyInf}[A] = \mathrm{FinInf}[A']$.
(b) $\mathrm{NoisyInf}[A] \subseteq \mathrm{NoisyInf}[B]$ *iff* $A' \leq_T B'$.
(c) $\mathrm{FinInf} \subset \mathrm{NoisyInf} \subset \mathrm{LimInf}$.

While for sets the definitions of noisy informant and very noisy informant are equivalent (data (x, y) with $y > 1$ can be ignored), this equivalence does not hold in the field of inferring functions. But there remains a connection:

Theorem 2.3 *If $\mathcal{L}$ can be learned from noisy informant and some K-recursive function f bounds all functions $g \in \mathcal{L}$, then $\mathcal{L}$ can also be learned from very noisy informant.*

Proof: Let M be an IIM which infers $\mathcal{L}$ from noisy informant and let f_s be a uniform recursive sequence of functions which approximate f in the limit: $(\forall x)\,(\forall^\infty s)\,[f(x) = f_s(x)]$. Since f only has to be an upper bound, w.l.o.g. the f_s approximate f from below.

Every very noisy informant $T = w_0 w_1 \ldots$ for $g \in \mathcal{L}$ can be translated into a new noisy informant $T' = v_0 v_1 \ldots$ as follows:

$$v_s = \begin{cases} w_s & \text{if } w_s = (x, y) \text{ and } y \leq f_s(x); \\ \# & \text{otherwise.} \end{cases}$$

Also in T' every pair $(x, g(x))$ occurs infinitely often since $(x, g(x))$ occurs infinitely often in T, $g(x) \leq f(x)$ and therefore $g(x) \leq f_s(x)$ for almost all s. On the other hand, if $y > f(x)$, then $y > f_s(x)$ for all x and therefore (x, y) never occurs in T'. Further if $y \leq f(x)$ and $y \neq g(x)$, then (x, y) occurs only finitely often in T and therefore also only finitely often in T'. Thus T' is a noisy informant for g. Since this translation is computable and can be done on all finite initial segments of T, M can infer g from T'. ∎

The converse does not hold. For example the family $\{e1^e0^\infty : e \in \omega\}$ can be learned from very noisy informant, but it has no bound on $f(0)$ at all. On the other hand, the condition, that f is K-recursive can not be weakened, since the family

$$\{0^x y 0^\infty : x \in \omega \wedge 1 \leq y \leq f(x)\}$$

can be learned from very noisy informant iff some K-recursive function majorizes f.

3 Inference From Text

Comparing the definition for learning from noisy informant with those for learning from noisy text and from very noisy text, the second seems more to fit to its counterpart than the first one. But it turns out that learning from very noisy text is a very restrictive concept since here two restrictions add - that of texts (compared to informant) and that of noise. Indeed the class of all singleton sets can only be learned from noisy text and not from very noisy text as similarly the class of all constant functions can only be learned from noisy informant but not from very noisy informant. So the next theorem indicates why noisy text is more interesting than very noisy text.

Theorem 3.1 *The class L containing all singleton sets $\{x\}$ can be learned from noisy text but not from very noisy text.*

Proof: There is an easy algorithm to infer L from noisy text: For each input σw the learner just guesses $\{w\}$. Since for each given noisy text $w_0 w_1 \ldots$ almost all w_i are the single word x of the singleton language $\{x\}$ to be learned, this algorithm is correct.

Assume by the way of contradiction that M learns all singleton sets from very noisy text. Then let σ_0 be the empty string and $\sigma_{n+1} = \sigma_n 0 n^k$ for the first k with $M(\sigma_n 0 n^k)$ outputting an index for $\{n\}$. Such a k must exist since $\sigma_n 0 n^\infty$ is a very noisy text for $\{n\}$. The limit of all these σ_n is a very noisy text for $\{0\}$ since 0 occurs infinitely often (in each σ_n exactly n times) and each n occurs only the k times for the k in the definition of σ_{n+1}. But M does not converge on this very noisy text and so M does not infer the family from very noisy text. ∎

Locking sequences are an important tool in learning from text. Therefore it is useful to define them also for inference from noisy text. Let M be an IIM which infers L. σ is called a locking sequence for L iff

- $M(\sigma) = e$ with $W_e = L$ and
- $M(\sigma\tau) = M(\sigma)$ for all $\tau \in L^*$.

Since $L = W_e$, σ is also called a locking sequence for the index e. The proof, that a locking sequence exists, is almost identical to the one in the case of learning from text and is therefore omitted. Using the concept of locking sequences, the next theorem shows, that it is impossible to learn a class of sets from noisy text, if some of the sets is a proper subset of some other.

Theorem 3.2 *If $L' \subset L$ then $\{L', L\} \notin$ NoisyTxt.*

Proof: Let M be an IIM which infers at least L and has a locking sequence σ for L. Further let $w_0 w_1 \ldots$ be an enumeration of L' in which every element of L' occurs infinitely often. Now $\sigma w_0 w_1 \ldots$ is a noisy text for L', but since $e = M(\sigma)$ is an index for L and $M(\sigma w_0 w_1 \ldots w_n) = e$ for all n (by $w_0, w_1, \ldots \in L$), M does not infer L' from noisy text. ∎

The severe restriction from Theorem 3.2 contrasts the fact, that if the sets to be learned are the graphs $\mathcal{G}$ of a set of functions, then there is no difference between noisy and non-noisy text, so learning from noisy text is in general not so restrictive as learning from noisy informant.

Theorem 3.3 *Let $\mathcal{G}$ be a the set of the graphs of some set of total recursive functions. Then $\mathcal{G} \in$ NoisyTxt $\Leftrightarrow \mathcal{G} \in$ LimTxt.*

Proof: If $w_0 w_1 \ldots$ is a noisy text for the graph of a function g, then it contains only finitely many (x, y) with $y \neq g(x)$ while each pair $(x, g(x))$ occurs infinitely often in $w_0 w_1 \ldots$. There is a first k such that all w_i with $i \geq k$ are of the form $(x, g(x))$ for some x. So $w_k w_{k+1} w_{k+2} \ldots$ is a text for *graph*(g) which is not noisy. Some IIM M infers $\mathcal{G}$ from text. The following IIM N infers $\mathcal{G}$ from noisy text:

> On input $w_0 w_1 \ldots w_n$, N searches the least $m \leq n$ such that the information $w_m, w_{m+1}, \ldots, w_n$ is not contradictory, i.e.,
>
> $$(\forall i, j)\,[m \leq i \leq j \leq n \wedge w_i = (x, y) \wedge w_j = (x, z) \Rightarrow y = z],$$
>
> and then N outputs $M(w_m w_{m+1} \ldots w_n)$.

For almost all n, this m (depending on n) coincides with k and therefore

$$(\forall^\infty n)\,[N(w_0 w_1 \ldots w_n) = M(w_k w_{k+1} \ldots w_n)].$$

Thus N on the noisy text $w_0 w_1 \ldots$ and M on the text $w_k w_{k+1} \ldots$, both converge to the same index for *graph*(g). ∎

Some families of functions can be inferred in the limit, but are not in FinInf$[A]$ for any oracle A. The family

$$\{f : (\forall^\infty x)\,[f(x) = 0]\} \in \text{LimInf} \Leftrightarrow \text{FinInf}[A]$$

is an example. So their graphs are LimTxt and NoisyTxt learnable, but not FinInf$[A]$ and FinTxt$[A]$ learnable for any oracle A. Therefore inference from noisy text is not contained in finite inference relative to any oracle:

Corollary 3.4 NoisyTxt $\not\subseteq$ FinTxt$[A]$ *for all oracles A.*

So in contrary to the case of the informant, the classes FinTxt$[K]$ and NoisyTxt do not coincide. Indeed Theorem 3.7 will show, that FinTxt$[A]$ and NoisyTxt are incomparable for all oracles A. Since Theorem 3.7 also studies the connections FinTxt$[A] \subseteq$ NoisyTxt$[B]$ it is worth to look first at the inference degrees with respect to learning from noisy text:

Theorem 3.5 *The following holds for all oracles A and B:*
(a) *If A is r.e. then* NoisyTxt$[A] \subseteq$ NoisyTxt$[B] \Leftrightarrow A \leq_T B$.
(b) *If $A, B \geq_T K$ then* NoisyTxt$[A] \subseteq$ NoisyTxt$[B] \Leftrightarrow A' \leq_T B'$.
(c) NoisyTxt$[A]$ = NoisyTxt *iff $A \leq_T K$ and A has recursive or 1-generic degree.*

Proof: (a): Obviously $A \leq_T B \Rightarrow$ NoisyTxt$[A] \subseteq$ NoisyTxt$[B]$ holds. For the converse consider the family $\mathcal{L}$ consisting of the r.e. set A and all sets $\{x\}$ with $x \notin A$. The IIM $M(w_0 w_1 \dots w_n)$ outputs an index of the set

$$\begin{array}{ll} \{w_n\} & \text{if } w_n \notin A; \\ A & \text{otherwise, i.e., if } w_n \in A. \end{array}$$

M is obviously A-recursive; further if $w_0 w_1 \dots$ is a noisy text for $\{x\}$, then $w_i = x$ for almost all i and M converges to an index for $\{x\}$. If $w_0 w_1 \dots$ is a noisy text for A then $w_i \in A$ for almost all i and the M almost always outputs the same index for A.

On the other hand, assume that M is B-recursive and infers $\mathcal{L}$. A has a locking sequence σ. If $x \notin A$, then M converges on σx^∞ to an index of $\{x\}$, thus $M(\sigma x^n) \neq M(\sigma)$ for some n. If $x \in A$, then $M(\sigma x^n) = M(\sigma)$ for all n since σ is a locking sequence for A. In short

$$x \in A \;\Leftrightarrow\; (\forall n)\,[M(\sigma x^n) = M(\sigma)]$$

and the r.e. set A is co-r.e. relative to B. Thus $A \leq_T B$.

(b): Let $A, B \geq_T K$, $A' \leq_T B'$ and $\mathcal{L} \in \mathrm{NoisyTxt}[A]$ via M. The set of all locking sequences σ for some W_e is recursive in A' by the formula

$$E = \{(\sigma, e) : (\forall \tau \in {W_e}^*) [M(\sigma\tau) = M(\sigma)]\}.$$

Thus E has a B-recursive approximation E_s such that w.l.o.g. no E_s is void. The new B-recursive IIM N infers $L \in \mathcal{L}$ from the text $w_0 w_1 \ldots$ as follows:

> $N(w_0 w_1 \ldots w_n) = e$ where there are m and $\sigma = v_0 v_1 \ldots v_m$ such that $(\sigma, e) \in E_n$ and the norm
>
> $$e + m + v_0 + v_1 + \ldots + v_m + |\{i \leq n : w_i \notin W_e\}|$$
>
> of (σ, e) w.r.t. the current input $w_0 w_1 \ldots w_n$ is minimal among the norms of all $(\sigma', e') \in E_n$ w.r.t. $w_0 w_1 \ldots w_n$.

Since $B \geq_T K$ the W_e are uniformly decidable relative to B. Since for each s some pair $(\sigma, e) \in E_s$ the algorithm finds at least one e and furthermore it has to compare the pair (σ, e) only with a finite number of other pairs (σ', e') since almost all pairs (σ', e') have a higher norm than (σ, e). Thus the algorithm terminates using the B-oracle.

Since for every set $L \in \mathcal{L}$ there is a pair $(\sigma, e) \in E$ with $W_e = L$, this pair is found for sufficient long n and either the algorithm converges to this e for the pair (σ, e) for L or to e' for some other pair $(\sigma', e') \in E$. Assume by the way of contradiction, that the algorithm takes the second case for some e' with $W_{e'} \neq L$. If there is some $w \in L \Leftrightarrow W_{e'}$, then this w occurs infinitely often. While the norm of (σ, e) w.r.t. each input $w_0 w_1 \ldots w_n$ is bounded by a constant c, the norm of (σ', e') is greater than the number of occurrences of w in the so far seen input and so the norm of (σ', e') is almost always greater than c and greater than the norm of (σ, e). From this contradiction it follows that the algorithm takes e' only if $L \subset W_{e'}$. Since $(\sigma', e') \in E$, it follows that $M(\sigma'\tau) = e'$ for all $\tau \in W_{e'}$ and in particular $M(\sigma'\tau) = e'$ for all $\tau \in L^*$. Since M converges to e' on some noisy text $T \in \sigma L^\infty$, e' must be an index for L, a contradiction. So this case also fails and N infers $\mathcal{L}$.

For the other way round, let C be a retraceable set of degree A', which is co-r.e. in A. Now let $\mathcal{L}$ consist of the sets

$$\begin{array}{lll} \{x, 0\} & \text{iff} & x > 0 \text{ and } x \in C, \\ \{x\} & \text{iff} & x > 0 \text{ and } x \notin C. \end{array}$$

Further C_s denotes an A-recursive approximation of C. Now given any input σ, let $x(\sigma)$ denote the last $y > 0$ which occurs in σ, i.e., $x(\sigma) = y \Leftrightarrow$

$\sigma \in \omega^* y 0^*$. If $\sigma \in 0^*$ then $x(\sigma) = 0$. Now $M(\sigma)$ outputs an index of the set

$$\begin{array}{ll} \{x(\sigma), 0\} & \text{if } x(\sigma) \in C_{|\sigma|} \\ \{x(\sigma)\} & \text{otherwise.} \end{array}$$

If T is a noisy text for $\{0, x\}$ or $\{x\}$, then $x(\sigma) = x$ for almost all $\sigma \preceq T$. Further $x \in C_{|\sigma|}$ iff $x \in C$ for almost all $\sigma \preceq T$. Thus $\mathcal{L} \in \text{NoisyTxt}[A]$ via M.

Thus $\mathcal{L} \in \text{NoisyTxt}[B]$ via some B-recursive N. If $x \in C$ then there is a locking sequence σ such that $N(\sigma\tau) = e$ for some index e of $\{0, x\}$ and all $\tau \in \{0, x\}^*$. On the other hand if $x \notin C$ then N converges on every text σx^∞ to an index for $\{x\}$. Thus

$$x \in C \Leftrightarrow (\exists \sigma)\,(\exists e)\,(\forall n)\,[0 \in W_e \wedge N(\sigma x^n) = e].$$

Therefore C is r.e. in B'; since C is retraceable, C is even recursive in B' and $A' \leq_T B'$ follows.

(c): The proof of this fact is similar to that of [13, Theorem 9.5] concerning LimTxt inference degrees. ∎

Theorem 3.5 also holds with LimTxt instead of NoisyTxt [13, Theorems 9.2, 9.4 and 9.5]. So it is likely, that the structures of the LimTxt and NoisyTxt inference degrees coincide and the following conjecture holds:

Conjecture 3.6 $\text{NoisyTxt}[A] \subseteq \text{NoisyTxt}[B] \Leftrightarrow \text{LimTxt}[A] \subseteq \text{LimTxt}[B]$.

The next result deals with the relation between FinTxt$[A]$ and NoisyTxt$[B]$.

Theorem 3.7 $\text{FinTxt}[A] \subseteq \text{NoisyTxt}[B] \Leftrightarrow K \leq_T B \wedge (A \oplus K)' \leq_T B'$.

Proof: The proof consists of three parts:
(a) If FinTxt $\subseteq$ NoisyTxt$[B]$ then $K \leq_T B$.
(b) If FinTxt$[A] \subseteq$ NoisyTxt$[B]$ then $(A \oplus K)' \leq_T B'$.
(c) If $(A \oplus K)' \leq_T B'$ and $K \leq_T B$ then FinTxt$[A] \subseteq$ NoisyTxt$[B]$.

(a): Let $\mathcal{L}$ contain K plus all singletons $\{x\}$ with $x \notin K$. There is a recursive function f such that

$$W_{f(x)} = \begin{cases} K & \text{if } x \in K; \\ \{x\} & \text{if } x \notin K. \end{cases}$$

Now the FinTxt IIM waits for the first x to appear on the input, outputs the guess $W_{f(x)}$ and terminates. The proof of Theorem 3.5:(a) shows that

$\mathcal{L} \in \text{NoisyTxt}[B]$ only if $K \leq_T B$.

(b): Let $\text{FinTxt}[A] \subseteq \text{NoisyTxt}[B]$. Let C be a retraceable set of degree $(A \oplus K)'$ which is co-r.e. in $A \oplus K$. So $\overline{C}$ is the domain of the partial function $\psi^{A \oplus K}$. Let U_y contain all x such that the computation of $\psi^{A \oplus K_y}(x)$ terminates within y steps and equals to that relative $A \oplus K$: whenever an odd number $2z+1$ is queried, then either $z \in K_y$ or $z \notin K$. Furthermore, all queries are made to numbers below y. Note that $U_y \subseteq \overline{C}$. Further each $x \notin C$ is in almost all sets U_y. The sets U_y are uniformly co-r.e. in A and there is a recursive function h such that $\overline{U_y} = W_{h(y)}^{\{z \in A : z < y\}}$. Now let $\mathcal{L}$ consist of the sets

$$\begin{array}{lll} \{2x, 1, 3, 5, 7, \ldots\} & \text{iff} & x \in C, \\ \{2x, 2y+1\} & \text{iff} & x \in U_y. \end{array}$$

First $\mathcal{L} \in \text{FinTxt}[A]$ is shown. The IIM waits until the even number $2x$ and an odd number $2y+1$ are in the input. Then it outputs the index $f(x,y)$ where

$$W_{f(x,y)} = \begin{cases} \{2x, 2y+1\} & \text{if } x \in U_y, \text{ i.e., if } x \notin W_e^{\{z \in A : z < y\}}; \\ \{2x, 1, 3, 5, 7, \ldots\} & \text{otherwise, i.e., if } x \in W_e^{\{z \in A : z < y\}}. \end{cases}$$

The function f is A-recursive and queries A only below y. $f(x,y)$ contains a table of $A(0), A(1), \ldots, A(y)$ and first enumerates $2x$ and $2y+1$ into $W_{f(x,y)}$. Then the machine emulates the enumeration of $W_{h(y)}^{\{z \in A : z < y\}}$ until x is enumerated into this set; if this happens then all odd numbers are enumerated into $W_{f(x,y)}$. So the IIM guesses $\{2x, 2y+1\}$ if $x \in U_y$ and guesses $\{2x, 1, 3, 5, 7, \ldots\}$ if $x \notin U_y$, in particular if $x \in C$. Thus $\mathcal{L} \in \text{FinTxt}[A]$ and $\mathcal{L} \in \text{NoisyTxt}[B]$ via some B-recursive M.

If $x \in C$ then M infers $V_x = \{2x, 1, 3, 5, 7, \ldots\}$ and V_x has a locking sequence. If $x \notin C$, then $x \in U_y$ for some y. Then M infers $\{2x, 2y+1\}$ and V_x has no locking sequence since $\{2x, 2y+1\} \subset V_x$. So the equivalences

$$\begin{array}{rcl} x \in C & \Leftrightarrow & V_x \text{ has a locking sequence} \\ & \Leftrightarrow & (\exists \sigma)\,(\exists e)\,(\forall \tau \in {V_x}^*)\,[M(\sigma\tau) = e \wedge |W_e| > 2] \end{array}$$

hold. C is r.e. in B'. Since C is retraceable, $C \leq_T B'$ and $(A \oplus K)' \leq_T B'$.

(c): If $B \geq_T K$ and $(A \oplus K)' \leq_T B'$, then $\text{NoisyTxt}[A \oplus K] \subseteq \text{NoisyTxt}[B]$. So it remains to show that $\text{FinTxt}[A] \subseteq \text{NoisyTxt}[A \oplus K]$.

Let $\mathcal{L} \in \text{FinTxt}[A]$. Form the definition of finite learning follows, that there are A-recursive functions f, g such that for every $L \in \mathcal{L}$:

- If $D_{f(i)} \subseteq L$ then $W_{g(i)} = L$;

- There is some i with $D_{f(i)} \subseteq L$.

Such a sequence can be obtained by A-recursively enumerating all strings σ on which a given FinTxt$[A]$ IIM M outputs some $e \neq ?$. Then for the i-th such string σ_i, let $D_{f(i)} = range(\sigma_i)$ and $g(i) = M(\sigma_i)$. W.l.o.g. $\mathcal{L} \neq \{\emptyset\}$ and therefore $D_{f(i)} \neq \emptyset$ for all i. Now the following IIM N infers $\mathcal{L}$ from noisy text:

- For all $i \leq |\sigma|$, N calculates c_i which is the maximal number y such that every $x \in D_{f(i)}$ occurs y times in σ.
- N finds the least i with $c_i \geq c_j$ for all $j \leq |\sigma|$.
- N outputs $g(j)$ for the least j with $D_{f(j)} \subseteq W_{g(i)}$ and $D_{f(i)} \subseteq W_{g(j)}$.

In a given text T for L, only finitely often, say k times, occurs some $x \notin L$. On the other hand each $x \in L$ occurs infinitely often in T. There is a minimal j with $D_{f(j)} \subseteq L$ and $W_{g(j)} = L$. Every $x \in D_{f(j)}$ occurs at least $k+1$ times in almost all $\sigma \preceq T$, thus for almost all $\sigma \preceq T$, the i computed in the second step satisfies $W_{g(i)} = L$. Then $D_{f(j)} \subseteq W_{g(i)}$ and $D_{f(i)} \subseteq W_{g(j)}$ and further that j is the minimal index with this property. So $N(\sigma) = j$ for almost all $\sigma \preceq T$ and N infers $\mathcal{L}$ from noisy text. ∎

So the only relation is FinTxt$[K] \subset$ NoisyTxt$[K]$ and there is no equivalent statement to FinInf$[K]$ = NoisyInf. The family

$$\{\omega \Leftrightarrow \{i\} : i \in \omega\}$$

is learnable from very noisy text but not FinTxt$[A]$ learnable for any oracle A. On the other hand there is a nice characterization of FinTxt$[K]$ using monotonicity notions:

Kapur [11] introduced (in the restricted context of section 5) the notion of strongly dual monotonic inference, i.e., whenever the IIM makes a mindchange from e to e', then the guessed language must be more special: $W_{e'} \subseteq W_e$. Jain and Sharma [8] and Kinber and Stephan [12] generalized this and other notions of monotonic inference to learning r.e. languages. While the class FinTxt$[K]$ can not be characterized in terms of noisy inference, it turned out to be equivalent with strongly dual monotonic inference without oracle. The reader may find more information on the field of monotonic learning in [9, 11, 14, 19, 20].

Theorem 3.8 *$\mathcal{L} \in$ FinTxt$[K]$ iff $\mathcal{L}$ can be learned via a recursive and strongly dual monotonic machine.*

Proof: $\text{SMon}^{\text{d}}\text{Txt} \subseteq \text{FinTxt}[K]$: Assume that M $\text{SMon}^{\text{d}}\text{Txt}$ infers $\mathcal{L}$. Then an IIM N $\text{FinTxt}[K]$ infers $\mathcal{L}$ as follows:

$$N(\sigma) = \begin{cases} e & \text{if there is a locking sequence } \tau \text{ for } W_e \text{ with } M(\tau) = e, \\ & |\tau| \leq |\sigma| \text{ and } range(\tau) \subseteq range(\sigma); \\ ? & \text{otherwise.} \end{cases}$$

Further N is required to make no further mindchange if it once has made a guess. Since N has only to check the strings τ in a finite set whether they are locking sequences for $W_{M(\tau)}$ or not, this can be done with K-oracle: τ is a locking sequence iff $M(\tau\eta) = M(\tau)$ for all $\eta \in {W_{M(\tau)}}^*$. Since during the $\text{SMon}^{\text{d}}\text{Txt}$ inference of a language L, all guesses $W_{M(\tau)}$ contain L, N never falsely suggests a τ being locking sequence. On the other hand there is a locking sequence τ and whenever σ is long enough, i.e., $range(\sigma) \supseteq range(\tau)$ and $|\sigma| \geq |\tau|$, the locking sequence is discovered.

$\text{FinTxt}[K] \subseteq \text{SMon}^{\text{d}}\text{Txt}$: This proof is similar to the corresponding part of Theorem 2.1. Given the $\text{FinTxt}[K]$ IIM M, the guess $N(\sigma)$ of the new $\text{SMon}^{\text{d}}\text{Txt}$ IIM is calculated as follows:

- Let $s = |\sigma|$. N searches for the shortest $\tau \preceq \sigma$ with $M^{K_s}(\tau) \neq ?$.
- If there is no such τ, then N outputs an index of Σ^*.
- Otherwise N computes $e = M^{K_s}(\tau)$ and outputs an index $f(e)$ of the set

  $$W_{f(e)} = \begin{cases} W_e & \text{if } M^{K_t}(\eta) = M^{K_s}(\eta) \text{ for all } \eta \preceq \tau \text{ and } t \geq s; \\ \Sigma^* & \text{otherwise, i.e., if } M^{K_t}(\eta) \neq M^{K_s}(\eta) \\ & \qquad \text{for some } \eta \preceq \tau \text{ and } t \geq s. \end{cases}$$

  The condition in the "otherwise"-case is r.e., thus an uniform algorithm for $W_{f(e)}$ first enumerates W_e until it discovers that the condition in the "otherwise"-case holds and then enumerates the whole set Σ^*. So f is recursive.

The inference process converges to the guess e of M and all previous guesses are changed to Σ^* at the moment that an error in the estimation M^{K_s} is discovered. ∎

One might ask, if this theorem relativizes. It does not relativize in the obvious way; the relativization needs the concept of inferring with finitely many queries. An IIM M $\text{SMon}^{\text{d}}\text{Txt}[A*]$ infers L iff M is strongly dual monotonic and L has a locking sequence σ such that $M(\sigma\tau) = M(\sigma)$ for all $\tau \in W_{M(\sigma)}{}^*$ and M makes the same oracle queries while calculating $M(\sigma)$ and $M(\sigma\tau)$. An equivalent definition is that M on every text for L makes only finitely many queries to A. See [4, Definition 2.23 and Section 5.2] for more information. Now the relativizations are:

Theorem 3.9
(a) $\mathrm{SMon^dTxt}[A*] = \mathrm{FinTxt}[A \oplus K]$.
(b) $\mathrm{SMon^dTxt}[A] \subseteq \mathrm{FinTxt}[A']$.
(c) $\mathrm{SMon^dTxt}[A] = \mathrm{FinTxt}[A']$ *for 1-generic sets* A.
(d) $\mathrm{SMon^dTxt}[K] \subset \mathrm{FinTxt}[K']$.

Proof: The proofs of (a) and (b) follow the corresponding parts of Theorem 3.8. (c) follows from the fact, that $A' \equiv_T A \oplus K$ for every 1-generic set A. The inclusion in (d) follows from (b) and the family $\mathcal{L}$ containing the sets

$$\begin{aligned} \{x\} \quad &\text{iff} \quad x > 0 \ \text{ and } x \in K''; \\ \{0, x\} \quad &\text{iff} \quad x > 0 \ \text{ and } x \notin K''; \end{aligned}$$

witnesses that the inclusion $\mathrm{SMon^dTxt}[K] \subset \mathrm{FinTxt}[K']$ is proper: The proof of Theorem 3.7 shows, that $\mathcal{L} \in \mathrm{FinTxt}[K']$ since K'' is r.e. in K'. To show that $\mathcal{L} \notin \mathrm{SMon^dTxt}[K]$ assume by the way of contradiction, that a K-recursive IIM M infers $\mathcal{L}$ dual monotonically from text. If $x \in K''$, then M infers $\{x\}$ form the text x^∞ and there is an n such that $M(x^n)$ outputs an index for $\{x\}$. Otherwise ($x \notin K''$) the IIM M must identify $\{0, x\}$ on each text $x^{n+1}0^\infty$ and therefore $M(x^n)$ always outputs a language which not only contains x but also 0. Thus

$$x \in K'' \ \Leftrightarrow \ (\exists n)\,[0 \notin W_{M(x^n)}].$$

Since the computation of $M(x^n)$ and the test, whether $0 \notin W_{M(x^n)}$, are recursive in K, K'' would be r.e. in K, which is obviously not possible. Thus such an M does not exist and the inclusion is proper. ∎

4 Informant Versus Text

It follows immediately from the definition that every family of r.e. sets, which is learnable from text, is also learnable from informant. But this does not hold in the case of noisy inference, since the definitions of noisy text and noisy informant do not match so good as in the standard case. So the following holds:

Theorem 4.1 NoisyInf$[A]$ *and* NoisyTxt$[B]$ *are incomparable for all oracles* A *and* B.

Proof: The family $\{\emptyset, \Sigma^*\}$ is finitely learnable from noisy informant, but not learnable from noisy text by Theorem 3.2. The family mentioned to prove Corollary 3.4 is in NoisyTxt$[B]$ for all oracles B, but not in FinInf$[A']$ for any oracle A, in particular not in NoisyInf$[A]$. ∎

So it is better to look for inclusions which hold under additional constraints. The first is to consider very noisy text versus (very) noisy informant; note that in the case of characteristic functions of sets, there is no difference between noisy and very noisy informant. Given a noisy informant $T = (w_0, b_0), (w_1, b_1), \ldots$ for a set L, the sequence T containing all w_i with $b_i = 1$ is a very noisy text for L: w_i occurs in T' infinitely often iff $(w_i, 1)$ occurs in T infinitely often iff $w_i \in L$. Thus one can translate every (very) noisy informant into a very noisy text and simulate the IIM learning from very noisy text. Thus the following theorem holds (and also relativizes to every oracle):

Theorem 4.2 *Every class of sets learnable from very noisy text is also learnable from noisy informant.*

While NoisyInf$[A] \not\subseteq$ NoisyTxt$[B]$ for all oracles A and B, there is a connection if the IIM learns from text without any noise:

Theorem 4.3 NoisyInf$[A] \subseteq$ LimTxt$[B]$ $\Leftrightarrow$ $A' \leq_T B'$.

Proof: ($\Rightarrow$): Let NoisyInf$[A] \subseteq$ LimTxt$[B]$. Further let C be a retraceable set of degree A' and let the class $\mathcal{L}$ contain the sets

$$\begin{array}{lcll} X_x & = & \{x, x+1, x+2, \ldots\} & \text{iff} \quad x \in C; \\ X_{x,y} & = & \{x, x+1, x+2, \ldots, x+y\} & \text{iff} \quad x \notin C \quad \text{and} \quad y \in \omega. \end{array}$$

The class has a FinInf$[A']$ IIM which on input σ outputs indices of the following sets:

$$\begin{array}{ll} X_x & \text{if } x \in C \quad \text{and} \quad \sigma \succeq 0^x 1; \\ X_{x,y} & \text{if } x \notin C \quad \text{and} \quad \sigma \succeq 0^x 1 \cdot 1^y 0; \\ ? & \text{otherwise.} \end{array}$$

The IIM makes only one guess and is recursive in C, i.e., recursive in A'. From $\text{FinInf}[A'] = \text{NoisyInf}[A]$ follows, that $\mathcal{L} \in \text{LimTxt}[B]$ via some N.

If $x \in C$ then N has a locking sequence σ for the set X_x. If $x \notin C$ then there is no locking sequence $\sigma \in X_x^*$: The range of σ is finite and there is some $y > \max(range(\sigma))$ such that $M(\sigma)$ is not an index for $X_{x,y}$. Therefore there is some $\tau \in X_{x,y}^*$ with $M(\sigma\tau) \neq M(\sigma)$. So the equivalences

$$\begin{aligned} x \in C \quad &\Leftrightarrow \quad N \text{ has a locking sequence on the set } X_x \\ &\Leftrightarrow \quad (\exists \sigma \in X_x^*)\,(\forall \tau \in X_x^*)\,[N(\sigma\tau) = N(\sigma)] \end{aligned}$$

hold and show that C is r.e. in B'. Since C is retraceable, C is recursive in B' and $A' \leq_T B'$.

($\Leftarrow$): From $\mathcal{L} \in \text{NoisyInf}[A]$ and $A' \leq_T B'$, it follows by Corollary 2.2 (a) that $\mathcal{L} \in \text{NoisyInf}[B]$ via some $M \leq_T B$. Now a LimTxt learner $N \leq_T B$ just translates the given text $w_0\, w_1 \ldots$ into a noisy informant $v_0\, v_1 \ldots$ for M and emulates M:

From input $w_0\, w_1 \ldots w_n$ compute $v_0, v_1, \ldots, v_n$ via

$$v_{\langle i,j \rangle} = \begin{cases} (i,1) & \text{if } i \in \{w_0, w_1, \ldots, w_j\}; \\ (i,0) & \text{otherwise } (i \notin \{w_0, w_1, \ldots, w_j\}); \end{cases}$$

and output $M(v_0\, v_1 \ldots v_n)$.

Since $j \leq \langle i,j \rangle$ the values $v_0, v_1, \ldots, v_n$ are computed without accessing the input-text beyond w_n, thus the computation is well-defined. Furthermore the whole sequence $v_0\, v_1 \ldots$ is a noisy informant for L: if $i \notin L$ then i does not occur in the sequence $w_0\, w_1 \ldots$ and thus only $(i,0)$ occurs in the informant. If $i \in L$ then $w_n = i$ for some n and $v_{\langle i,j \rangle} = (i,1)$ for all $j \geq n$, i.e., $(i,1)$ occurs infinitely often in the noisy informant and $(i,0)$ only finitely often (at most n times). So N behaves on the text $w_0\, w_1 \ldots$ exactly as M on the noisy informant $v_0\, v_1 \ldots$ and thus $\mathcal{L} \in \text{LimTxt}$ via N. ∎

5 Learning Uniformly Recursive Families

Angluin [1] introduced the concept of learning, where the class $\mathcal{L}$ to be learned must have a uniformly recursive representation. Zeugmann's Habilitationsschrift [20] gives an overview on this field of learning theory. There are three well-known forms of learning uniformly recursive family $\{L_0, L_1, \ldots\}$ of languages:

Exact Learning: The learner outputs indices of the original uniformly recursive family $\{L_0, L_1, \ldots\}$.

Class Preserving Learning: The learner outputs indices of some uniformly recursive family $\{H_0, H_1, \ldots\}$ with $\{L_0, L_1, \ldots\} = \{H_0, H_1, \ldots\}$.

Class Comprising Learning: The learner outputs indices of some uniformly recursive family $\{H_0, H_1, \ldots\}$ with $\{L_0, L_1, \ldots\} \subseteq \{H_0, H_1, \ldots\}$.

In the context of noisy inference these three notions turn out to be equivalent. Furthermore, they are very restrictive, therefore the results, in particular the relativization, are different from those in section 3.

Theorem 5.1 *For a uniformly recursive family $\mathcal{L} = \{L_i\}$ the following is equivalent:*
(a) $(\forall i, j)\,[L_i \subseteq L_j \Rightarrow L_i = L_j]$.
(b) *$\mathcal{L}$ is exactly learnable from noisy text.*
(c) *$\mathcal{L}$ is class preserving learnable from noisy text.*
(d) *$\mathcal{L}$ is class comprising learnable from noisy text.*

Proof: (b $\Rightarrow$ c) and (c $\Rightarrow$ d) are obvious. Further (d $\Rightarrow$ a) follows from Theorem 3.2.

(a $\Rightarrow$ b): Let $\mathcal{L}$ fulfill the requirement from (a). It is shown that the following machine M infers $L \in \mathcal{L}$ from the text $w_0 w_1 \ldots$ as follows:

$M(w_0 w_1 \ldots w_n) = i$ for the first i such that

$$\text{norm}(i, w_0 w_1 \ldots w_n) = i + |\{m \leq n : w_m \notin L_i\}|$$

is minimal among all values $\text{norm}(j, w_0 w_1 \ldots w_n)$ for $j \leq i + n$.

Let $T = w_0\, w_1\, \ldots$ be a noisy text for $L \in \mathcal{L}$ and let i be the minimal index of L in the given enumeration. There are k numbers m such that $w_m \notin L_i$. Furthermore for each $j \leq k + j$ with $L_j \not\subseteq L_i$ there is some $x_j \in L_j \Leftrightarrow L_i$. For almost all n the k elements $w_m \notin L_i$ are on the initial segment $w_0\, w_1\, \ldots\, w_n$ and also each of the x_j occurs at least $i + k + 1$ times for those $j \leq i + k$ with $L_i \not\subseteq L_j$. So it holds that

$$\begin{array}{lll} \text{norm}(j, w_0 w_1 \ldots w_n) = j + k & \text{if} & L_j = L_i \text{ and therefore } j \geq i; \\ \text{norm}(j, w_0 w_1 \ldots w_n) > i + k + j & \text{if} & L_j \neq L_i \text{ and } j \leq i + k; \\ \text{norm}(j, w_0 w_1 \ldots w_n) > i + k & \text{if} & j > i + k. \end{array}$$

Thus M outputs the correct value i for all sufficient long $w_0 w_1 \dots w_n \preceq T$ and learns $\mathcal{L}$ exactly from noisy text. ▮

It is easy to see that Theorem 3.2 holds also in a relativized world, i.e., that for any oracle A, $L \subset L' \Rightarrow \{L, L'\} \notin \text{NoisyTxt}[A]$. Since avoiding inclusions is the only restriction to $\mathcal{L}$ and this restriction can not be overcome, oracles do not help to increase the learning power:

Theorem 5.2 *If $\mathcal{L}$ is a uniformly recursive family which is* NoisyTxt$[A]$ *learnable for some oracle A, then $\mathcal{L}$ is already learnable from noisy text without any oracle.*

The theorem needs that $\mathcal{L}$ is uniformly recursive. Note that this is totally different in the case of learning arbitrary families of r.e. languages since by Theorem 3.7, there is even no greatest inference degree and the jump of an oracle always supplies more learning power: $\text{NoisyTxt}[A] \subset \text{NoisyTxt}[A']$.

Theorem 5.3 *For a uniformly recursive family $\mathcal{L} = \{L_i\}$ the following is equivalent:*
(a) $(\forall i)\,(\exists D)\,(\forall j)\,[D \subseteq L_j \Leftrightarrow L_i = L_j]$.
(b) *$\mathcal{L}$ is exactly* FinTxt$[K]$ *learnable.*
(c) *$\mathcal{L}$ is class preserving* FinTxt$[K]$ *learnable.*
(d) *$\mathcal{L}$ is class comprising* FinTxt$[K]$ *learnable.*

Proof: (b $\Rightarrow$ c) and (c $\Rightarrow$ d) are obvious.

(a $\Rightarrow$ b): Let $\mathcal{L}$ fulfill the requirement from (a). The FinTxt$[K]$ IIM asks on input σ with range D always iff D has two incomparable extensions in $\mathcal{L}$. Or more formally, the IIM asks the query

$$(\exists i, j, x)\,[D \subseteq L_i \wedge D \subseteq L_j \wedge (x \in L_i \not\Leftrightarrow L_j \vee x \in L_j \not\Leftrightarrow L_i)].$$

Since D is a fixed finite set, the query is K-recursive. By condition (a) after finite time the query receives a negative answer. Then the IIM has only to output the first index i with $D \subseteq L_i$; this index exists since σ is part of a text of some language L_i.

(d $\Rightarrow$ a): If $\mathcal{L}$ is class comprising FinTxt$[K]$ learnable, then for each L_i there is some string σ such that the FinTxt$[K]$ IIM M makes a guess, which of course is correct, i.e., $M(\sigma)$ guesses L_i. Assume that $D = range(\sigma) \subseteq L_j$. Then on one hand σ is also a prefix of some text for L_j and on the other hand M does not change its mind after guessing L_i on input σ. It follows that $L_j = L_i$ and for each i there is a D satisfying condition (a). ▮

The degree-structure of the FinTxt and FinInf inference degrees relative to learning uniform recursive families of sets is different from the degree structure of learning arbitrary families of r.e. sets. Fortnow et al. [4, Theorem 6.36] showed that the latter coincides with the Turing degrees.

Theorem 5.4 *Let* $\mathrm{F}[A]$ *be the set of all functions* f *which are majorized by an* A*-recursive function and for which the set* $\{(x,y) : y < f(x)\}$ *is r.e.; further consider the inference degrees with respect to learning uniformly recursive families. Now the following is equivalent:*
(a) $\mathrm{F}[A] \subseteq \mathrm{F}[B]$.
(b) $\mathrm{FinTxt}[A] \subseteq \mathrm{FinTxt}[B]$.
(c) $\mathrm{FinInf}[A] \subseteq \mathrm{FinInf}[B]$.

Proof: (a $\Rightarrow$ b): Let $\mathrm{F}[A] \subseteq \mathrm{F}[B]$ and $\mathcal{L} = \{L_i\} \in \mathrm{FinTxt}[A]$ be a uniformly recursive family. W.l.o.g. if $i \neq j$ then $L_i \neq L_j$. Now for each i let $w_{i,x} = x$ if $x \in L_i$ and $w_{i,x} = \#$ otherwise ($x \notin L_i$). Further let $f_A(i)$ be the first x such that $M(w_{i,0}w_{i,1}\ldots w_{i,x}) \neq ?$. Certainly $range(w_{i,0}w_{i,1}\ldots w_{i,x}) = \{y \in L_i : y \leq x\} \not\subseteq L_j$ for every set $L_j \neq L_i$. Thus f_A dominates the function $f_{\mathcal{L}}$ given by

$$f_{\mathcal{L}}(i) = \min\{x : (\forall j \neq i)\,(\exists y \leq x)\,[y \in L_i \Leftrightarrow L_j]\}.$$

The set $\{(i,y) : y < f_{\mathcal{L}}(i)\}$ is r.e. and $f_{\mathcal{L}} \in \mathrm{F}[A]$. From the hypothesis (a) follows, that a B-recursive function f_B majorizes $f_{\mathcal{L}}$. The new IIM M works as follows:

$$M(\sigma) = \begin{cases} i & \text{if } i \leq |\sigma| \text{ and } (\forall x \leq f_B(i))\,[x \in L_i \Leftrightarrow x \in range(\sigma)]; \\ ? & \text{otherwise.} \end{cases}$$

Since $\{x \in L_i : x \leq f_B(i)\} \not\subseteq L_j$ for all $j \neq i$, the i in the expression is unique and M is well-defined. Whenever M infers L_i then M outputs ? until it has seen all elements in $\{x \in L_i : x \leq f_B(i)\}$; then it begins to output its only guess i. So $\mathcal{L} \in \mathrm{FinTxt}[B]$ via M.

(b $\Rightarrow$ c): Note that $\mathcal{L} \in \mathrm{FinInf}[A] \Leftrightarrow \mathcal{L}' = \{L \oplus \overline{L} : L \in \mathcal{L}\} \in \mathrm{FinTxt}[A]$. Thus $\mathcal{L} \in \mathrm{FinInf}[A] \Rightarrow \mathcal{L}' \in \mathrm{FinTxt}[A] \Rightarrow \mathcal{L}' \in \mathrm{FinTxt}[B] \Rightarrow \mathcal{L} \in \mathrm{FinInf}[B]$ and therefore $\mathrm{FinInf}[A] \subseteq \mathrm{FinInf}[B]$.

(c $\Rightarrow$ a): This is shown by contraposition, let $f \in \mathrm{F}[A] \Leftrightarrow \mathrm{F}[B]$ and let the A-recursive function f_A majorize f. Since f has a recursive approximation from below, the family

$$\mathcal{L} = \{D : (\exists i \in D)\,[D \subseteq \{i, i+1, i+2, \ldots, i+f(i)\}\,]\,\}$$

is uniformly recursive. M finitely infers $\mathcal{L}$ relative to A as follows:

- If $\sigma = 0^i 1\tau$ and $|\tau| > f_A(i)$ then M outputs an index for $\{x < |\sigma| : \sigma(x) = 1\}$.
- Otherwise M makes no guess, i.e., $M(\sigma) = ?$.

On the other hand assume that $\mathcal{L} \in \mathrm{FinInf}[B]$ via N and let

$$f_B(i) = \min\{|\tau| : N(0^i 1\tau) \neq ?\}.$$

Since no B-recursive function majorizes f, there is some i with $f_B(i) < f(i)$. Thus there is $D \in \mathcal{L}$ such that $i = \min(D)$ and inferring D, N makes its guess before seeing whether $i + f(i) \in D$ or not. N fails to infer either $D \cup \{i + f(i)\}$ or $D \Leftrightarrow \{i + f(i)\}$, but both sets are in $\mathcal{L}$. ∎

Corollary 5.5 *For the inference degrees of* FinTxt *or* FinInf *learning uniformly recursive families, the following holds:*
(a) *All oracles of hyperimmune-free degree are in the least inference degree.*
(b) *All 1-generic oracles are in the least inference degree.*
(c) *If A is r.e., then A's inference degree is below that of B iff $A \leq_T B$.*
(d) *$\{A : A \geq_T K\}$ is the greatest inference degree.*

Proof: (a): If A is of hyperimmune-free degree then $\mathrm{F}[A] = \mathrm{F}[\emptyset]$ since any A-recursive function is majorized by a recursive one. Thus all sets of hyperimmune-free degrees belong to the least inference degree.

(b): Let A be a 1-generic set. Consider any $f \in \mathrm{F}[A]$ and let the A-recursive function $f_A = \{e\}^A$ majorize f. The set

$$B = \{\eta : (\exists x)\,[\{e\}^\eta(x)\!\downarrow < f(x)\,]\,\}$$

is r.e.; since $\{e\}^A(x)\!\downarrow \geq f(x)$, no string in B is a prefix of A. Since A is 1-generic, there is a string $\sigma \preceq A$ such that no extension of σ is in B. Now let

$$g(x) = \{e\}^\eta(x) \text{ for the first } \eta \succeq \sigma \text{ such that } \{e\}^\eta(x)\!\downarrow \text{ within } |\eta| \text{ steps.}$$

g is recursive and majorizes f. Thus $f \in \mathrm{F}[\emptyset]$, i.e., $\mathrm{F}[A] = \mathrm{F}[\emptyset]$.

(c): Let A_s be a recursive enumeration of A and $\mathrm{F}[A] \subseteq \mathrm{F}[B]$. Now

$$f(x) = \begin{cases} s & \text{for the first } s \text{ with } x \in A_s; \\ 0 & \text{otherwise } (x \notin A, \text{ i.e., there is no such } s); \end{cases}$$

is a function in F$[A]$ and some B-recursive function g majorizes f. Then $x \in A \Leftrightarrow x \in A_{g(x)}$ and $A \leq_T B$.

(d): The greatest degree can only contain degrees $A \geq_T K$ since K is r.e.; so it remains to show that $\mathrm{F}[A] \subseteq \mathrm{F}[K]$ for all oracles A. But this follows from the fact, that each function $f \in \mathrm{F}[A]$ is already K-recursive since $\{(x,y) : y < f(x)\}$ is an r.e. set. ∎

Theorem 5.6 $\mathrm{FinTxt}[K] \subset \mathrm{NoisyTxt}$ *in the context of uniformly recursive families.*

Proof: Assume that $\mathcal{L}$ satisfies the condition (a) of Theorem 5.3. Then $\mathcal{L}$ also satisfies condition (a) of Theorem 5.1: If $L_i \subseteq L_j$ then there is some $D \subseteq L_i$ such that all $L_j \supseteq D$ are equal to L_i. Then in particular, $L_i = L_j$.

The family $\mathcal{L} = \{\omega \Leftrightarrow \{i\} : i \in \omega\}$ of all sets whose complement has cardinality 1 witnesses that the inclusion in proper. ∎

Sometimes the addition of an oracle allows to overcome the difference between two concepts. An analogous result from the area of uniform recursive languages to the result $\mathrm{FinTxt}[K] = \mathrm{SMon}^{\mathrm{d}}\mathrm{Txt}$ from the general context is the following one, where Angluin [1] introduced the notion *conservative*: An IIM is conservative iff every mindchange is motivated by a counterexample to the previous conjecture, i.e., iff j is guessed on input $\sigma\tau$ after i was guessed on input σ then $range(\sigma\tau) \not\subseteq L_i$.

Theorem 5.7 $\mathrm{ConsvTxt}[K] = \mathrm{LimTxt}$ *in the context of uniformly recursive families.*

Proof: ($\Rightarrow$): Let $\mathcal{L} = \{L_0, L_1, \ldots\} \in \mathrm{Consv}[K]$ via M^K and T be a text for some $L \in \mathcal{L}$. Since M is conservative, the following holds for all oracles A: If $i = M^A(\sigma) \neq j = M^A(\sigma\tau)$ then $range(\sigma\tau) \not\subseteq L_i$. That means that M regardless of the oracle postpones any mindchange until a witness is seen that makes it necessary. In other words: M is conservative for any oracle. Since the sets $L_0, L_1, \ldots$ are uniform recursive, this postponing does not need the oracle. On each input σ the LimTxt learner guesses $N(\sigma) = M^{K_{|\sigma|}}(\sigma)$. M converges on some $\tau \preceq T$ to an index i for the language to be learned. Now for sufficient long $\sigma \in \tau \cdot L_i^*$ it holds that $M^{K_{|\sigma|}}(\tau) = M^K(\tau) = i$ and therefore also $M^{K_{|\sigma|}}(\sigma) = i$ by the conservativeness of the machine M^{K_σ}. So $N(\sigma) = i$ for all sufficient long σ and N infers $\mathcal{L}$.

($\Leftarrow$): Let $\mathcal{L} = \{L_0, L_1, \ldots\} \in \mathrm{LimTxt}$ via N and $w_0 w_1 \ldots$ be a text for some

$L \in \mathcal{L}$. With K-oracle it is possible to test whether a given sequence σ is a locking-sequence for N. The ConsvTxt[K]-algorithm defines inductively (using the K-oracle) a new text $w_0\sigma_0w_1\sigma_1\ldots$ and emulates N on this text:

If there are i and τ such that

- $i + |\tau| \leq n$ and $\tau \in \{w_0, w_1, \ldots, w_n\}^*$;
- $w_0, w_1, \ldots, w_n \in L_i$;
- $(\forall \eta \in L_i^*)\,[N(w_0\sigma_0w_1\sigma_1\ldots w_n\tau\eta) = i]$;

then let $\sigma_n = \tau$, $M^K(w_0w_1\ldots w_n) = N(w_0\sigma_0w_1\sigma_1\ldots w_n\sigma_n) = i$;
else let $\sigma_n = \lambda$, $M^K(w_0w_1\ldots w_n) = ?$.

The algorithm works with K-oracle, since the search for the τ is bounded. If $w_0w_1\ldots$ is a text for L_i then $w_0\sigma_0w_1\sigma_1\ldots$ is also a text for L_i. N converges on this text to i and so M^K converges on the text $w_0w_1\ldots$ also to i. Furthermore if $M^K(w_0w_1\ldots w_n) = j \neq M^K(w_0w_1\ldots w_m) = i$ with $m > n$ then $N(w_0\sigma_0w_1\sigma_1\ldots w_n\sigma_n) = j$, $N(w_0\sigma_0w_1\sigma_1\ldots w_m\sigma_m) = i$ and $N(w_0\sigma_0w_1\sigma_1\ldots w_n\sigma_n\eta) = j$ for all $\eta \in L_j^*$. Thus $w_0\sigma_0w_1\sigma_1\ldots w_m\sigma_m \notin L_j^*$ and since $\sigma_k \in \{w_0, w_1, \ldots, w_m\}^*$ for $k \leq m$ it follows that some $w_k \notin L_j$ for $k \leq m$. So the mindchange from j to i was induced by a counterexample and M^K is conservative (using the definition that outputting ? does not count as a mindchange). ∎

The proof even relativizes to LimTxt[A] = ConsvTxt[A'] which shows that the inference-degrees w.r.t. learning uniform recursive families is quite different to the degree-structure w.r.t. learning arbitrary families of r.e. sets: In the latter case the low r.e. oracles all belong to different inference-degrees. Furthermore if $L_i \not\subset L_j$ for all i, j then the family $\mathcal{L}$ can be learned conservatively: On input σ the IIM just guesses the first index i with $range(\sigma) \subseteq L_i$. So in the context of learning uniformly recursive sets the following holds for all oracles A and B:

Corollary 5.8 FinTxt $\subset$ FinTxt[K] $=$ FinTxt[$A \oplus K$] $\subset$ NoisyTxt $\subset$ ConsvTxt $\subset$ ConsvTxt[K] = LimTxt $\subset$ LimTxt[K] = LimTxt[$B \oplus K$].

6 Behaviorally Correct and Partial Identification

Behavioral Correct identification means that an IIM outputs an infinite sequence of hypothesis which almost all compute the correct function or generate the correct set. It turns out that learning functions from noisy informant, there is no difference between behaviorally correct and explanatory inference (NoisyInf):

Theorem 6.1 *The following three statements are equivalent for any class* $\mathcal{L} \subseteq \mathrm{REC}$*:*
(a) $\mathcal{L}$ *can be learned finitely from informant using* K*-oracle.*
(b) $\mathcal{L}$ *can be learned in the limit from noisy informant.*
(c) $\mathcal{L}$ *can be learned behaviorally correct from noisy informant.*

Proof: Since convergence in the limit always implies behaviorally correct convergence, obviously (b $\Rightarrow$ c) holds. (a $\Rightarrow$ b) is shown in Theorem 2.1, part $\mathrm{FinInf}[K] \subseteq \mathrm{NoisyInf}$. The remaining implication (c $\Rightarrow$ a) is an adapted version of Theorem 2.1, part $\mathrm{NoisyInf} \subseteq \mathrm{FinInf}[K]$:

Assume that M is a recursive IIM which learns the family $\mathcal{L}$ of sets behaviorally correct from noisy informant. Recall that a string σ is called α-consistent iff all (x, y) occurring in σ with $x < |\alpha|$ satisfy $y = \alpha(x)$. Now the following $\mathrm{FinInf}[K]$ IIM N infers $\mathcal{L}$:

> On input α, N checks using the K-oracle whether there is a string σ of length up to $|\alpha|$ such that for all α-consistent strings τ and τ' relation
>
> $$(\forall x)[\varphi_{M(\sigma\tau)}(x)\downarrow \wedge \varphi_{M(\sigma\tau')}(x)\downarrow \Rightarrow \varphi_{M(\sigma\tau)}(x) = \varphi_{M(\sigma\tau')}(x)]$$
>
> holds. If yes, then no two guesses $M(\sigma\tau)$ and $M(\sigma\tau')$ contradict each other and $N(\alpha)$ converges to an index e of the amalgamation of all functions $\varphi_{M(\sigma\tau)}$ with τ ranging over all α-consistent strings. If not, then $N(\alpha) = ?$.

Let $f \in \mathcal{L}$ and M behaviorally correct infer f. Then there is some $\alpha \preceq f$ and some string σ such that $M(\sigma\tau)$ is an index for f for all α-consistent strings τ – otherwise it could be shown as in Theorem 2.1 that there is a noisy informant from which M does not learn f behaviorally correct. W.l.o.g. assume that $|\sigma| \leq |\alpha|$. Then $N(\alpha)$ outputs an index e of the amalgamation of the functions $\varphi_{M(\sigma\tau)}$; it is easy to see that $\varphi_e = f$.

So it remains to show that N does not output an other false index before finding e, i.e., that already the first index output by N is correct. So let $\alpha \preceq f$ satisfy $N(\alpha) = e \neq ?$. Take the σ from the definition of $N(\alpha)$. Let T enumerate all pairs $(x, f(x))$ infinitely often without any noise. Now σT is obviously a noisy informant for f and there is a $\tau \preceq T$ such that $f = \varphi_{M(\sigma\tau)}$.

By choice, τ is α-consistent. So $\varphi_e(x)\downarrow = \varphi_{M(\sigma\tau)}(x)\downarrow$ for all x and $\varphi_e = f$. It follows that inferring any function $f \in \mathcal{L}$ the first guess of N is already correct and w.l.o.g. N makes no mindchanges. ∎

NoisyBC denotes the concept of inferring behaviorally correct from noisy text. The non-inclusion FinTxt $\not\subseteq$ NoisyTxt does not generalize behaviorally correct inference:

Theorem 6.2 FinTxt $\subset$ NoisyBC.

Proof: Let M infer finitely a class $\mathcal{L}$ of languages from text, in particular M guesses ? until it outputs a guess e and then keeps this output e for ever. Now consider N given by

$$N(w_0 w_1 \ldots w_n) = \begin{array}{l} M(w_m w_{m+1} \ldots w_n) \text{ for the maximal } m \leq n \\ \text{with } M(w_m w_{m+1} \ldots w_n) \neq ? \end{array}$$

and let $w_0 w_1 \ldots$ be a noisy text for L. Since there is a maximal k with $w_k \notin L$, each sequence $w_m w_{m+1} \ldots$ with $m > k$ is a text for L. In particular for all $n \geq m$, $M(w_m w_{m+1} ... w_n)$ is either ? or an index for L. Since N outputs $M(w_m w_{m+1} ... w_n)$ for the maximal m such that $M(w_m w_{m+1} \ldots w_n) \neq ?$, these m satisfy $m > k$ for almost all input $w_0 w_1 \ldots w_n$; then $w_m, w_{m+1}, \ldots, w_n \in L$ and since M finitely learns L, $M(w_m w_{m+1} \ldots w_n)$ is always an index for L.

The properness of the inclusion follows from NoisyTxt $\not\subseteq$ FinTxt (Corollary 3.4) and the obvious fact that NoisyTxt $\subseteq$ NoisyBC. ∎

Osherson, Stob and Weinstein [17, Exercise 7.5A] introduced the notion of partial identification from text and showed that the family of all r.e. languages can be learned from text under this criterion. The concept directly transfers to noisy learning:

Definition 6.3 A machine M partially identifies $\mathcal{L}$ from noisy text iff for every $L \in \mathcal{L}$ and every noisy text T for L there is a unique index e such that M outputs e infinitely often on input T and $W_e = L$. Partial identification from very noisy text and very noisy informant is defined analogously.

Let REC denote the class of all total recursive functions and RE that of all r.e. sets. The result of Osherson, Stob and Weinstein generalizes for learning from very noisy informant and from noisy text:

Theorem 6.4 REC *is partially identifiable from very noisy informant.*
RE *is partially identifiable from noisy informant.*
RE *is partially identifiable from noisy text.*

Proof: REC is partially identifiable from very noisy informant:

Let $\{\varphi_{h(e)}\}_{e\in\omega}$ be a Friedberg numbering of all partial recursive functions, h is total recursive. Further let T be a noisy informant for f. M may be specified only by stating how often M outputs an index $h(e)$ on text T since it does not matter when these outputs occur and identification only depends on how often M outputs an index.

> M outputs $h(e)$ at least n times iff for $x = 0, 1, \ldots, n$ the following two conditions are satisfied:
> - $\varphi_{h(e)}(x)\downarrow$,
> - $(x, \varphi_{h(e)}(x))$ occurs at least n times in T.

So M reads longer and longer initial segments and whenever M notices that it has put out less than n times $h(e)$ while the conditions above demand to output $h(e)$ at least n times, M's next output is $h(e)$.

There is an unique index e with $f = \varphi_{h(e)}$. For each x, the pairs $(x, f(x))$ occur infinitely often in T and furthermore, $\varphi_{h(e)}(x)\downarrow = f(x)$ for all x. Thus the conditions are satisfied for each n and M outputs $h(e)$ infinitely often.

Now consider any $e' \neq e$. There is some x such that either $\varphi_{h(e')}(x)\uparrow$ or $\varphi_{h(e')}(x) \neq f(x)$. In the latter case, $(x, \varphi_{h(e')}(x))$ occurs only finitely often, say m times in T. Thus for all $n > x$ – with additionally $n > m$ in the second case – M outputs the index $h(e')$ less than n times, in particularly only finitely often. Therefore M partially identifies REC from very noisy informant.

RE is partially identifiable from noisy informant:

Note that for characteristic functions, the notions noisy informant and very noisy informant are the same. So the statement is equivalent to saying that RE can be partially identified from very noisy informant. Now let $\{W_{h(e)}\}_{e\in\omega}$ be a Friedberg numbering of all r.e. sets and let T be a noisy informant for some r.e. set L. This inference process is similar to the previous one.

> M outputs $h(e)$ at least n times iff there is some $s \geq n$ such that the pairs $(x, W_{h(e),s}(x))$ occur at least n times in T for $x = 0, 1, \ldots, n$.

Let e be the index of L, i.e., $L = W_{h(e)}$. For each n there is $s \geq n$ such that $W_{h(e)}(x) = W_{h(e),s}(x)$ for all $x \leq n$. Thus $(x, W_{h(e),s}(x))$ occurs in T infinitely many times for these x and M outputs $h(e)$ at least n times, therefore even infinitely often.

Let $e' \neq e$. There is some x with $W_{h(e)}(x) \neq W_{h(e')}(x)$. There is some m such that $(x, W_{h(e')}(x))$ does not occur in T more than m times and $W_{h(e'),s}(x) = W_{h(e')}(x)$ for all $s \geq m$. Then M does not output $h(e')$ for any $n > x + m$. Thus M partially identifies L from T.

RE is partially identifiable from noisy text:

To proof this, one needs a padded version of the Friedberg numbering. So let $W_{g(e,k)} = W_{h(e)}$ for an injective recursive function g and the Friedberg numbering h of all r.e. sets from the second part. Let $T = w_0 w_1 w_2 \ldots$ be a noisy text for the r.e. language L.

> M outputs $g(e,k)$ at least n times iff the the following three r.e. conditions are satisfied:
> - $w_k, w_{k+1}, \ldots, w_{k+n} \in W_{h(e)}$;
> - $k = 0$ or $w_{k-1} \notin W_{h(e),n}$;
> - Each $x \in W_{h(e),n}$ occurs at least n times in T.

Let e denote the index with $W_{g(e,k)} = L$ and $k = \min\{l : (\forall m \geq l)\,[w_m \in L]\}$. k exists since T is a noisy text for L and so $w_k, w_{k+1}, \ldots, w_{k+n} \in W_{h(e)}$ for all n. Either $k = 0$ or $w_{k-1} \notin W_{h(e)}$ (and therefore $w_{k-1} \notin W_{h(e),n}$). Each $x \in W_{h(e)}$ occurs infinitely often in T. So all three conditions are satisfied for each n and M outputs $g(e,k)$ infinitely often.

Assume by the way of contradiction that M outputs a further index $g(e',k')$ infinitely often on text T. Then each $x \in W_{g(e',k')}$ occurs infinitely often in T since each such x is enumerated into $W_{g(e',k')}$ at some stage s and for all $n > x + s$, if M outputs $g(e',k)$ at least n times then x occurs in T at least n times. Thus $x \in L$. If $x \notin W_{g(e',k')}$ then x must not occur in T beyond the k-th position and therefore $x \notin L$. Therefore $W_{g(e',k')} = L$ and $e' = e$. If $k' > k$ then $w_{k'-1}$ is enumerated into $W_{h(e)}$ at some stage s. Thus for no $n \geq s + k'$ the learner M outputs the guess $g(e,k')$. If $k' < k$ (and thus $k > 0$) then $g(e,k')$ is not output more than k times since $w_{k-1} \notin W_{h(e)}$. So M does not output any index $g(e',k') \neq g(e,k)$ infinitely often. ∎

While RE is partially identifiable from noisy text, RE is not partially identifiable from very noisy text as the following example shows:

Example 6.5 Let $\mathcal{L}$ contain all sets $\{x, x+1, x+2, x+3, \ldots\}$.
$\mathcal{L}$ is partially identifiable from very noisy text.
$\mathcal{L} \cup \{\emptyset\}$ is not partially identifiable from very noisy text.

Proof: Since each set in $\mathcal{L}$ is co-finite, every very noisy text for some $L \in \mathcal{L}$ is already a noisy text: each number not in L occurs only finitely often and since there are only finitely many numbers outside L, only finitely many items of a very noisy text for L are not in L: Thus the text is already noisy. Since every class of languages can be partially identified from noisy text, $\mathcal{L}$ can be identified from very noisy text.

By the way of contradiction assume that M partially identifies $\mathcal{L} \cup \{\emptyset\}$, M may be even nonrecursive. Further let $e_0, e_1, e_2, \ldots$ be the list of all indices of the empty set. Now the following sequence $T = \sigma_0\sigma_1\sigma_2\ldots$ is constructed inductively:

For each n select a string $\sigma_n \in \{n, n+1, n+2, \ldots\}^*$ such that

$$(\forall \tau \in \{n, n+1, n+2, \ldots\}^*)\, [M(\eta_n\sigma_n\tau) \neq e_n],$$

where $\eta_0 = \lambda$ and $\eta_n = \sigma_0\sigma_1\ldots\sigma_{n-1}$ for $n > 0$.

This construction works, because if σ_n would not exist there would be a noisy text $T_n \in \eta_n\{n, n+1, n+2, \ldots\}^\infty$ for $\{n, n+1, n+2, \ldots\}$ on which M infinitely often outputs e_n and then M would not partially identify $\{n, n+1, n+2, \ldots\}$ since e_n is an index of $\emptyset$.

So by construction, $M(\tau) \neq e_n$ whenever $\eta_n\sigma_n \preceq \tau \preceq T$, thus M outputs e_n on input T only finitely often. Further each number n occurs only in the strings σ_m for $m \leq n$, thus each number n occurs only finitely often in T. So T is a very noisy text for $\emptyset$ but M does not partially identify $\emptyset$ from T. ∎

Since $\mathcal{L}$ is learnable in the limit from text by guessing $\emptyset$ if $range(\sigma) = \emptyset$ and guessing the set $\{n, n+1, n+2, \ldots\}$ if $range(\sigma)$ is not empty and has minimum n, $\mathcal{L}$ is a witness for the fact, that LimTxt does not imply partially identifiability from very noisy text. On the other hand the class of all graphs of recursive functions is partial identifiable from very noisy text without being learnable in the limit from text or informant.

Corollary 6.6 *Learning in the limit from text and partially identification from very noisy text are incomparable concepts.*

Acknowledgments I would like to thank John Case, Susanne Kaufmann and Efim Kinber for helpful discussions, proofreading and introducing me to noisy inference.

References

[1] Angluin, D. (1980), Inductive inference of formal languages from positive data, *Information and Control* **45**, pp. 117–135.

[2] Baliga, G., Jain, S., and Sharma, A. (1992), Learning from Multiple Sources of Inaccurate Data, *in* "Proceedings of the International Workshop on Analogical and Inductive Inference in Dagstuhl Castle, Germany", October 1992, pp. 108–128.

[3] Blum, M., and Blum, L. (1975), Towards a mathematical theory of inductive inference, *Information and Control*, **28**, pp. 125–155.

[4] Fortnow, L., Gasarch, W. I., Jain, S., Kinber, E., Kummer, M., Kurtz, S., Pleszkoch, M., Slaman, T., Solovay, R., Stephan, F. (1994), Extremes in the degrees of inferability. *Annals of Pure and Applied Logic*, **66**, pp. 231–276.

[5] Fulk, M., and Jain, S. (1989), Learning in the presence of inaccurate information, *in* "Proceedings of the 2nd Annual ACM Conference on Computational Learning Theory," Santa Cruz, July 1989, pp. 175–188, Morgan Kauffmann Publishers.

[6] Gold, E.M. (1967), Language identification in the limit, *Information and Control* **10**, pp. 447–474.

[7] Jain, S. (1994), Program Synthesis in the Presence of Infinite Number of Inaccuracies, *in* "Proceedings of the 5th Workshop on Algorithmic Learning Theory", October 1994,

[8] Jain, S., and Sharma, A. (1994), On monotonic strategies for learning r.e. languages, *in* "Proceedings of the 5th Workshop on Algorithmic Learning Theory", October 1994, pp. 349–364.

[9] Jantke, K.P. (1991) Monotonic and non-monotonic inductive inference, *New Generation Computing* **8**, pp. 349–360.

[10] Jockusch, C. (1981) Degrees of generic sets. *London Mathematical Society Lecture Notes 45*, pp. 110–139.

[11] Kapur, S. (1992), Monotonic language learning, *in* "Proceedings of the 3rd Workshop on Algorithmic Learning Theory", October 1992, Tokyo, JSAI, pp. 147–158.

[12] KINBER, E., AND STEPHAN, F. (1995), Language Learning from Texts: Mind Changes, Limited Memory and Monotonicity, *Information and Computation* **123**, pp. 224–241.

[13] KUMMER, M., AND STEPHAN, F. (1993) On the structure of degrees of inferability, *in* "Proceedings of the 6th Annual ACM Conference on Computational Learning Theory," Santa Cruz, July 1993, pp. 117–126, ACM Press, New York.

[14] LANGE, S., ZEUGMANN, T., AND KAPUR, S. (1992), Monotonic and dual monotonic language learning, GOSLER-Report 14/94, TH Leipzig, FB Mathematik und Informatik, August 1992.

[15] SCHÄFER, G. (1995), Some results in the theory of effective program synthesis – learning by defective information. *Lecture Notes in Computer Science 225*, pp. 219–225.

[16] ODIFREDDI, P. (1989), "Classical Recursion Theory", North-Holland, Amsterdam.

[17] OSHERSON, D., STOB, M., AND WEINSTEIN, S. (1986), "Systems that Learn, An Introduction to Learning Theory for Cognitive and Computer Scientists," MIT-Press, Cambridge, Massachusetts.

[18] SOARE, R. (1987), "Recursively Enumerable Sets and Degrees", Springer-Verlag, Heidelberg.

[19] WIEHAGEN, R. (1990), A thesis in inductive inference, *in* "Proceedings First International Workshop on Nonmonotonic and Inductive Logic" Karlsruhe, Dec. 1990, Lecture Notes in Artificial Intelligence 534, pp. 184–207, Springer Verlag, Heidelberg.

[20] ZEUGMANN, T. (1993), Algorithmisches Lernen von Funktionen und Sprachen. Habilitationsschrift, Technische Hochschule Darmstadt, Fachbereich Informatik.