Noisy Inference and Oracles

Frank Stephan *
Universität Karlsruhe

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.

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

*Institut für Logik, Komplexität und Deduktionssysteme, Universität Karlsruhe, 76128 Karlsruhe, Germany, email ⟨fstephan@ira.uka.de⟩. Supported by the Deutsche Forschungsgemeinschaft (DFG) grant Me 672/4-2.

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.

Each correct item is presented to the learner infinitely often, but in addition the learner has to cope with some incorrect information, called the noise. Thus the noisy inference considered here can be put into Jain's first category (a), i.e., learning with additional faulty information. 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 $M(w_0w_1 \dots w_n) = e$ 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.

There are two further important concepts of learning theory in this paper: 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 either some $\sigma \in W$ with $\sigma \preceq A$ ("A meets W") or there is some $\sigma \notin W$ with $\sigma \preceq A$ and $\tau \notin W$ for all $\tau \succeq \sigma$ ("A strongly avoids W"). In this definition, "recursive" may be replaced by "r.e." without changing the concept. 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 $\{B : \text{NoisyInf}[B] = \text{NoisyInf}[A]\}$ [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 mind change 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 $\alpha_n = (U(0), U(1), \ldots, U(n))$ and $\sigma_0 = (0,0)$. There are strings τ_n for $n = 1, 2, \ldots$ such that

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

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 $\subseteq$ FinTxt$[K]$.

FinInf$[K]$ $\subseteq$ NoisyInf: Let N^K be a FinInf$[K]$ IIM for some family $\mathcal{L}$. Given any string σ, the NoisyInf IIM first computes a set U_σ as follows:

$$U_\sigma = \{x : (x,1) \text{ occurs in } \sigma \text{ more often than } (x,0)\}$$

Here is the only place where the algorithm for arbitrary functions slightly differs: Inferring the function f, $f_\sigma(x) = y$ for that y with the most occurrences (x,y) in σ and $f_\sigma(x) = 0$ if there is no such unique y.

Now $M(\sigma) = N^{K_{|\sigma|}}(\alpha)$ for the shortest $\alpha \preceq U_\sigma$ such that $N^{K_{|\sigma|}}(\sigma)\downarrow \neq ?$ within $|\sigma|$ steps. Let $\alpha \preceq K$ be the shortest string with $N^K(\alpha) = e \neq ?$. Then $\alpha \preceq U_\sigma$, $N^{K_{|\sigma|}}(\alpha) = e$ and $N^{K_{|\sigma|}}(\beta) = ?$ for all $\beta \prec \alpha$ and almost all $\sigma \preceq T$. Thus $M(\sigma) = e$ for almost all $\sigma \preceq T$ and M infers $\mathcal{L}$ from noisy informant. ▮

The proof easily relativizes. Since FinInf$[A]$ $\subseteq$ 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) $\text{NoisyInf}[A] = \text{FinInf}[A']$.
(b) $\text{NoisyInf}[A] \subseteq \text{NoisyInf}[B]$ *iff* $A' \leq_T B'$.
(c) $\text{FinInf} \subset \text{NoisyInf} \subset \text{LimInf}$.

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

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

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

3 Inference From 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. We do not give the proof of an existence of a locking sequence, this proof is almost identical to that in the case of learning from text. But the next two theorems make use of this fact.

Theorem 3.1 *Let $\mathcal{L} \in$ NoisyTxt. Then there is a recursive* IIM *which learns $\mathcal{L}$ from noisy text and for each $L \in \mathcal{L}$ there is a unique index e for L such that every noisy text T has an locking sequence $\sigma \preceq T$ with $M(\sigma) = e$.*

Proof: The proof is similar to the locking-sequence-hunting construcion of [17, Proposition 4.3.6A]. Let M be an given IIM; the new IIM N is defined inductively. Given the input $w_0 w_1 \ldots w_n$, let m denote the maximal number such that either $N(w_0 w_1 \ldots w_m) \neq N(w_0 w_1 \ldots w_{n-1})$ or $N(w_0 w_1 \ldots w_m) = ?$, further $m = 0$ if $n = 0$; note that $0 \leq m \leq n$. Now N looks for the first string σ such that

- $\sigma \in \{w_0, w_1, \ldots, w_n\}^*$ and $|\sigma| \leq n$.
- $range(\sigma) \subseteq W_{e,n}$ for $e = M(\sigma)$.
- $M(\sigma) = M(\sigma\tau)$ for all $\tau \in \{w_m, w_{m+1}, \ldots, w_n\}^*$ with $|\tau| \leq n$.

If such a σ exists, $N(w_0 w_1 \ldots w_n) = e$, otherwise $N(w_0 w_1 \ldots w_n) = ?$. The verification is similar to the in the proof of [17, Proposition 4.3.6A]; only the second condition is due to the fact, that a noisy text $w_0 w_1 \ldots$ for some $L \in \mathcal{L}$ may contain finitely many elements w_i outside L. ∎

Learning from text is impossible, if the language contains an infinite set and all its finite subsets. The following theorem shows, that learning from noisy text is even more restrictive:

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 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 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 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 $N(w_0 w_1 \ldots w_n) = M(w_k w_{k+1} \ldots w_n)$ for almost all 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 $\mathrm{FinInf}[A]$ for any oracle A. The family

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

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

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

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

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

Proof: (a): Obviously $A \leq_T B \Rightarrow \mathrm{NoisyTxt}[A] \subseteq \mathrm{NoisyTxt}[B]$ 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 $M(w_0 w_1 \ldots w_n)$ 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 $w_0 w_1 \ldots$ is a noisy text for $\{x\}$, then $w_i = x$ for almost all i and M converges to an index for $\{x\}$. If $w_0 w_1 \ldots$ 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 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 $\mathcal{L} \in \text{NoisyTxt}[A]$ via M. W.l.o.g. for every $L \in \mathcal{L}$ and every noisy text T for L, there is some locking sequence $\sigma \preceq T$ for 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; further A has an B-recursive approximation A_s. Now the new B-recursive IIM N infers $L \in \mathcal{L}$ from the text $w_0 w_1 \ldots$ as follows:

> $N(w_0 w_1 \ldots w_n)$ searches for the first m such that $w_m, w_{m+1}, \ldots, w_n \in W_{e_m}$ and $e_m \in E_n$ for $e_m = M^{A_n}(w_0 w_1 \ldots w_m)$.

There is a first m such that $w_n \in W_e$ for all $n \geq m$, $w_0 w_1 \ldots w_n$ is a locking sequence for W_e and $M^A(w_0 w_1 \ldots w_m) = e$. For all sufficiently large n, $M^{A_n}(w_0 w_1 \ldots w_m) = e$, $(w_0 w_1 \ldots w_m, e) \in E_n$ and for all $i < n$ with $e_i = M^{A_n}(w_0 w_1 \ldots w_i)$, either $(w_0 w_1 \ldots w_i, e_i) \notin E_n$ or some $w_j \notin W_{e_i}$ where $i \leq j \leq n$. Thus $N(w_0 w_1 \ldots w_n) = e$ for almost all n and also 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 a 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 FinTxt$[A] \subseteq$ 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 $(A \oplus B)' \leq_T B'$.
(c) If $(A \oplus B)' \leq_T B'$ and $K \leq_T B$ then FinTxt$[A] \subseteq$ NoisyTxt$[B]$.

(a): Let $\mathcal{L}$ contain K plus all sigletons $\{x\}$ 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 $\mathcal{L} \in$ NoisyTxt$[B]$ only if $K \leq_T B$.

(b): Let FinTxt$[A] \subseteq$ 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, there is a recursive function h such that the $\overline{U_y} = W_{h(y)}^{\{z \in A: z < y\}}$. The condition is only co-r.e. in A because of the last condition which enforces $\psi^{A \oplus K_y}(x) = \psi^{A \oplus K}(x)$ for $x \in U_y$. All other conditions are recursive in A. 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 $\mathcal{L} \in$ 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 uses A only to give the information, which of the $z < y$ are in A. Thus the IIM always guesses $\{2x, 1, 3, 5, 7, \ldots\}$ if $x \notin U_y$, in particular if $x \in C$.

Thus $\mathcal{L} \in$ FinTxt$[A]$ and $\mathcal{L} \in$ NoisyTxt$[B]$ via 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. Thus C is r.e. in B'. Since C is retraceable, $C \leq_T B'$ and $(A \oplus B)' \leq_T B'$.

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

Let $\mathcal{L} \in$ 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 $D_{f(i)} = range(\sigma)$ and $g(i) = M(\sigma)$. Now the following IIM N infers $\mathcal{L}$ from noisy text:

- For all $i \leq |\sigma|$, N calculates c_i which is the minimal number y such that some $x \in D_{f(i)}$ 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. ∎

Kapur [11] introduced (in the restricted context of section 5) the notion of strongly dual monotonic inference, i.e., whenever the IIM makes a mind change from e to e', then the guesses must fulfill the requirement $W_{e'} \subseteq W_e$. Jain and Sharma [8] and Kinber and Stephan [12] generalized this and other notions of monotonic inference to arbitrary inference. 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.

Theorem 3.8 SMon$^{\mathrm{d}}$Txt $=$ FinTxt$[K]$.

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

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

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

  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 finitely many queries. An IIM M $\mathrm{SMon}^{\mathrm{d}}\mathrm{Txt}[A*]$ 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) $\mathrm{SMon}^{\mathrm{d}}\mathrm{Txt}[A*] = \mathrm{FinTxt}[A \oplus K]$.
(b) $\mathrm{SMon}^{\mathrm{d}}\mathrm{Txt}[A] \subseteq \mathrm{FinTxt}[A']$.
(c) $\mathrm{SMon}^{\mathrm{d}}\mathrm{Txt}[A] = \mathrm{FinTxt}[A']$ *for 1-generic sets A.*
(d) $\mathrm{SMon}^{\mathrm{d}}\mathrm{Txt}[K] \subset \mathrm{FinTxt}[K']$.

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

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

witnesses that the inclusion $\mathrm{SMon}^{\mathrm{d}}\mathrm{Txt}[K] \subset \mathrm{FinTxt}[K']$ 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 $\mathcal{L} \notin \mathrm{SMon}^{\mathrm{d}}\mathrm{Txt}[K]$ 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 informant, but not learnable from noisy text by Theorem 3.2. The same holds for the family consisting of Σ^* and all finite sets: it is even in the absence of any noise not learnable from text.

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: Let $\mathcal{L} \in$ NoisyInf$[A]$, i.e., $\mathcal{L} \in$ FinInf$[A']$, and let $A' \leq_T B'$. There is an A'-recursive sequence (α_n, e_n) such that for every $L \in \mathcal{L}$, $L = W_{e_n}$ for the first n with $\alpha_n \preceq L$. The sequence (α_n, e_n) has an B-recursive approximation $(\alpha_{n,s}, e_{n,s})$, i.e., $\alpha_{n,s} = \alpha_n$ and

$e_{n,s} = e_n$ for almost all s. Now let $w_0 w_1 \ldots$ be a text for $L \in \mathcal{L}$. The LimTxt$[B]$ IIM works as follows:

$$M(w_0 w_1 \ldots w_s) = \begin{cases} e_{n,s} & \text{for the first } s \leq n \text{ with } \alpha_{n,s} \preceq \{w_0, w_1, \ldots, w_s\}; \\ ? & \text{otherwise, i.e., if there is no such } s. \end{cases}$$

Let n be the smallest number such that $\alpha_n \preceq L$ and let $k = |\alpha_0|+|\alpha_1|+\ldots+|\alpha_n|$. Further let t be the first stage such that $x \in \{w_0, w_1, \ldots, w_s\}$, $\alpha_{m,s} = \alpha_m$ and $e_{m,s} = e_m$ for all $m \leq n$, $s \geq t$ and $x \in L \cap \{0, 1, \ldots, k\}$. It is easy to verify that $M(w_0 w_1 \ldots w_s) = e_n$ for all $s \geq t$ and M infers L from text.

For the other way round, 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

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

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

$$\begin{array}{ll} \{x, x{+}1, x{+}2, \ldots\} & \text{if } x \in C \;\text{ and }\; \sigma \succeq 0^x 1; \\ \{x, x{+}1, x{+}2, \ldots, x{+}y\} & \text{if } x \notin C \;\text{ and }\; \sigma \succeq 0^x 1 \cdot 1^y 0; \\ ? & \text{otherwise.} \end{array}$$

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

If $x \in C$ then N has a locking sequence σ for the set $\{x, x{+}1, x{+}2, \ldots\}$. If $x \notin C$ then there is no locking sequence $\sigma \in \{x, x{+}1, x{+}2, \ldots\}^*$: The range of σ is finite and there is some $y > \max(range(\sigma))$ such that $M(\sigma)$ is not an index for $\{x, x{+}1, x{+}2, \ldots, x{+}y\}$. Therefore there is some $\tau \in \{x, x{+}1, x{+}2, \ldots, x{+}y\}^*$ with $M(\sigma\tau) \neq M(\sigma)$. So the equivalences

$$\begin{array}{lll} x \in C & \Leftrightarrow & N \text{ has a locking sequence on the set } \{x, x{+}1, x{+}2, \ldots\} \\ & \Leftrightarrow & (\exists \sigma \in \{x, x{+}1, x{+}2, \ldots\}^*)(\forall \tau \in \{x, x{+}1, x{+}2, \ldots\}^*)[N(\sigma\tau) = N(\sigma)] \end{array}$$

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

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 [19] gives an overview on this field of learning theory. This notion is 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: ($\mathtt{b} \Rightarrow \mathtt{c}$) and ($\mathtt{c} \Rightarrow \mathtt{d}$) are obvious. Further ($\mathtt{d} \Rightarrow \mathtt{a}$) follows from Theorem 3.2.

($\mathtt{a} \Rightarrow \mathtt{b}$): Let $\mathcal{L}$ fulfill the requirement from ($\mathtt{a}$). It has to be shown that $\mathcal{L} \in$ NoisyTxt via an IIM that outputs the indices relative to the numbering $\{L_i\}$. For each σ and i let $error(\sigma, i)$ denote the number of $w \in \sigma$ with $w \notin L_i$. Now

$$M(\sigma) = \begin{cases} i & \text{if } i \text{ is the smallest } j \leq |\sigma| \text{ such that every } x \leq error(\sigma, j) \\ & \text{with } x \in L_j \text{ occurs at least } error(\sigma, j) \text{ times in } \sigma; \\ ? & \text{if there is no such } j \leq |\sigma|. \end{cases}$$

Let T be a noisy text for $L \in \mathcal{L}$. If $L_i \neq L$ then there are $x \in L - L_i$ and $y \in L_i - L$. x occurs infinitely often in T while the y occurs only finitely often in T, say k times. Every sufficient large initial segment $\sigma \preceq T$ contains x more than $x + y + k$ times, thus $error(\sigma, i) > x + y + k$. Since $y \leq error(\sigma, i)$ and y occurs less than $error(\sigma, i)$ times in L_i, $M(\sigma) \neq i$.

Now let i be the minimal index of L, i.e., i is the minimal j with $L = L_j$. For almost all $\sigma \preceq T$, $M(\sigma) \notin \{0, 1, \ldots, i-1\}$. Only finitely often, say k times, occurs some element outside L in T. Thus $error(\sigma, i) \leq k$ for all $\sigma \preceq L$. Further all $x \in L_i$ with $x \leq k$ occur in all sufficient long $\sigma \preceq T$ at least $k + 1$ times, thus $M(\sigma) = i$ for almost all $\sigma \preceq T$ and M converges to i. ∎

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 the avoidence of 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.*

This is a consequence of $\mathcal{L}$ being uniformly recursive; in the general context there is even no omniscient inference degree since by Theorem 3.7, NoisyTxt$[A] \subset$ NoisyTxt$[A']$ holds for all oracles A.

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

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

($\mathtt{a} \Rightarrow \mathtt{b}$): Let $\mathcal{L}$ fulfill the requirement from ($\mathtt{a}$). 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

$$(\exists i, j, x)\,[D \subseteq L_i \wedge D \subseteq L_j \wedge (x \in L_i - L_j \vee x \in L_j - L_i)].$$

Since D is a fixed finite set, the query is K-recursive. By condition ($\mathtt{a}$) 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.

($\mathtt{d} \Rightarrow \mathtt{a}$): If $\mathcal{L}$ is class comprising $\mathrm{FinTxt}[K]$ learnable, then for each L_i there is some string σ such that the $\mathrm{FinTxt}[K]$ IIM M makes a guess, which of course is correct, say $M(\sigma) = i$. Therefore if $D = range(\sigma) \subseteq L_j$ then $L_j = L_i$ since σ is also an initial segment of some text for L_j and there is no mindchange possible after M suggested L_i. ∎

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* $\{(x, y) : f(x) < y\}$ *is r.e.; further consider the inference degrees with respect to learning uniformly recursive families. Now the following is equivalent:*
($\mathtt{a}$) $\mathrm{F}[A] \subseteq \mathrm{F}[B]$.
($\mathtt{b}$) $\mathrm{FinTxt}[A] \subseteq \mathrm{FinTxt}[B]$.
($\mathtt{c}$) $\mathrm{FinInf}[A] \subseteq \mathrm{FinInf}[B]$.

Proof: ($\mathtt{a} \Rightarrow b$): Let $\mathrm{F}[A] \subseteq \mathrm{F}[B]$ and $\mathcal{L} = \{L_i\}$ 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 $M(w_0 w_1 \ldots w_x) \neq ?$. Certainly $range(w_0 w_1 \ldots w_x) = \{y \in L_i : y \leq x\} \not\subseteq L_j$ for every set $L_j \neq L_i$. Thus f_A dominates the function $f_{\mathcal{L}}$ given by

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

Clearly $\{(i, y) : y < f_{\mathcal{L}}(i)\}$ is an r.e. set and $f_{\mathcal{L}} \in \mathrm{F}[A]$. From the hypothesis ($\mathtt{a}$) 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.

($\mathtt{b} \Rightarrow \mathtt{c}$): 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 $\mathrm{FinInf}[A] \subseteq \mathrm{FinInf}[B]$.

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

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

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 $D - \{i + f(i)\}$, 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 have trivial inference degree.*
(b) *All 1-generic oracles have trivial 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 omniscient 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 trivial 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

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

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 $\mathrm{F}[A]$ 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 omniscient 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 *Concerning learning uniformly recursive families,* $\mathrm{FinTxt}[K] \subset \mathrm{NoisyTxt}$.

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, $L_i = L_j$.

To see the properness of the inclusion, consider the family which contains the graphs of all primitive recursive functions. ▮

6 Partial Identification

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 6.1 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 6.2 REC *is partially identifiable from very noisy informant.*
RE *is partially identifiable from noisy informant.*
RE *is partially identifiable from noisy text.*

Proof: For the first part 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.

In the second case 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 $(x, W_{h(e),s}(x))$ occurs 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 ouput $h(e')$ for any $n > x + m$. Thus M partially identifies L from T.

The third case 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 be a noisy text for the r.e. language L. For the ease of notation let w.l.o.g. $T(0) \notin L$.

M outputs $g(e, k)$ at least n times iff there is $\sigma \preceq T$ and $s \geq n$ such that the following two conditions are satisfied:
- if $x < n$ and $x \in W_{g(e,k),s}$ then x occurs in σ at least n times,
- $\sigma(k)$ is the last occurence of some $x \notin W_{g(e,k),s}$ in σ.

Let e denote the index with $W_{g(e,k)} = L$ and $k = \max\{k' : T(k') \notin L\}$. If $n \geq k$ then there is some $\sigma \preceq T$ longer than n such that each $x \in L$ with $x \leq n$ occurs at least n times in σ. At some stage $s \geq n$ all $x \in L$ which occur in σ are enumerated into $W_{g(e,k),s}$. In particular all $\sigma(k')$ with $k < k' < |\sigma|$ are enumerated into $W_{g(e,k),s}$. Now both conditions are satisfied and M outputs $g(e, k)$ at least n times. Since this holds for infinitely many n, 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 $x \in W_{e',g}$ occurs infinitely often in T since each such x is enumerated into $W_{e',g}$ 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 $x \notin W_{e',g}$ then x must not occur in T beyond the k-th position and therefore $x \notin L$. Therefore $W_{e',g} = L$. If $k' > k$ then $T(k')$ is enumerated into $W_{e,k'}$ at some stage s. Thus for no $n \geq s + k'$ the learner M outputs the guess $g(e, k')$. Therefore $k' < k$. If $n \geq k$, then $|\sigma| > k$ and $\sigma(k) \notin W_{g(e,k),s}$. Therefore M does not guess $g(e, k')$ more than k times and 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:

Theorem 6.3 *The family $\mathcal{L}$ containing $\emptyset$ and $\{x, x+1, x+2, x+3, \ldots\}$ for all x is not partially identifiable from very noisy text.*

Proof: By the way of contradiction assume that M partially identifies $\mathcal{L}$, 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 6.4 *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., 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*, to appear. An extended extract appears in "Proceedings of the 8th Annual ACM Conference on Computational Learning Theory," Santa Cruz, July 1995.

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