Program Synthesis in the Presence of Infinite Number of Inaccuracies

Sanjay Jain
Department of Information Systems and Computer Science
Lower Kent Ridge Road
National University of Singapore
Singapore 119260, Republic of Singapore
Email: sanjay@iscs.nus.sg

Abstract

Most studies modeling inaccurate data in Gold style learning consider cases in which the number of inaccuracies is finite. The present paper argues that this approach is not reasonable for modeling inaccuracies in concepts that are infinite in nature (for example, graphs of computable functions).

The effect of infinite number of inaccuracies in the input data in Gold's model of learning is considered in the context of identification in the limit of computer programs from graphs of computable functions. Three kinds of inaccuracies, namely, noisy data, incomplete data, and imperfect data, are considered. The amount of each of these inaccuracies in the input is measured using certain density notions. A number of interesting hierarchy results are shown based on the densities of inaccuracies present in the input data. Several results establishing tradeoffs between the density and type of inaccuracies are also derived.

1 Introduction

Consider the scenario in which a subject is attempting to learn its environment. At any given time, the subject receives a finite piece of data about its environment, and based on this finite information, conjectures an explanation about the environment. The subject is said to *learn* its environment just in case the explanations conjectured by the subject become fixed over time, and this fixed explanation is a correct representation of the subject's environment. Computational learning theory provides a framework for the study of the above scenario when the subject is an algorithmic device. The above model of learning is based on the work initiated by Gold [9] and has been used in inductive inference of both functions and languages. We refer the reader to [1, 3, 6, 14, 12] for background material in this field.

Most learning situations involve the presence of inaccuracies in the data presented to a learner. In the context of linguistic development, children are likely to face both ungrammatical intrusions and omission of some grammatical sentences from the ambient language; it is to be expected that minor perturbations of this kind would not influence the outcome of linguistic development. Similarly, in the context of scientific discovery, the business of science progresses despite experimental errors and unfeasibility of performing certain experiments. Several attempts have been made to model inaccuracies in Gold's paradigm [7, 8, 14, 18]. Each of these studies, however, only consider cases in which the number of inaccuracies is finite. Now, this may be a suitable approach if the data available about the concepts to be learned is finite in nature, but not when the nature of data is infinite.

A problem of interest is identifying in the limit computer programs from graphs of computable functions. Now, the graph of a computable function is an infinite set of ordered pairs. Considering only finite number of errors in the graph is not a very realistic model of inaccuracies because this may imply that all the inaccuracies are, in some sense, restricted to some small region of the graph. A more suitable model would allow for the inaccuracies to be spread throughout the graph of the function such that "density" of these errors is bounded.

The present paper investigates precisely such models of identification from inaccurate data. To measure the amount of inaccuracy present in the input data when they might be infinite in number we use notions of density from [17] (see also [19]).

We discuss three forms of inaccuracies that may be present in the input. For each of these we give a hierarchy of inference criteria based on the density of inaccuracy present in the input. We also give results comparing the three types of inaccuracies with each other. Even though some of our results and arguments also apply to language identification we will mainly be concerned with function inference in this paper.

In section 2 we discuss notation and fundamental inference paradigms. In section 3 we discuss inaccurate information sequences and inference paradigms based on them. In sections 4 and 5 we give our results.

2 Preliminaries

2.1 Notation

Recursion-theoretic concepts not explained below are treated in [16]. $\mathbb{N}$ denotes the set of natural numbers, $\{0, 1, 2, 3, \ldots\}$, and $\mathbb{N}^+$ denotes the set of positive integers, $\{1, 2, 3, \ldots\}$. $\in$,

$\subseteq$, and $\subset$ denote, respectively, membership, containment, and proper containment for sets (including sets of ordered pairs). e, i, j, k, l, m, n, r, s, x, y, z, with or without decorations[1], range over $\mathbb{N}$. $*$ is a non-member of $\mathbb{N}$ satisfying $(\forall n \in \mathbb{N})[n < * < \infty]$. a, b, c, with or without decorations, range over $\mathbb{N} \cup \{*\}$. d, with or without decorations, ranges over the real interval $[0,1]$. We let A, B, R, S, W, X, Y, Z, with or without decorations, range over subsets of $\mathbb{N}$. $\mathrm{card}(S)$ denotes the cardinality of S. So then, '$\mathrm{card}(S) \leq *$' means that $\mathrm{card}(S)$ is finite. $\min(S)$ and $\max(S)$ respectively denote the minimum and maximum element in S. We take $\min(\emptyset)$ to be ∞ and $\max(\emptyset)$ to be 0. $S_1 \Delta S_2$ denotes $(S_1 - S_2) \cup (S_2 - S_1)$, the symmetric difference of S_1 and S_2.

Let $\lambda x, y . \langle x, y \rangle$ denote a fixed pairing function (a recursive, bijective mapping: $\mathbb{N} \times \mathbb{N} \to \mathbb{N}$) [16]. $\langle \cdot, \cdot \rangle$ can be extended to pairing function for multiple arguments in a natural way.

f, g, h, p, F, with or without decorations, range over total functions. $\mathcal{C}$ and $\mathcal{S}$, with or without decorations, range over sets of total functions. $\mathrm{graph}(f)$ denotes the set $\{(x, f(x)) \mid x \in \mathbb{N}\}$. η and ξ range over partial functions. For $a \in \mathbb{N} \cup \{*\}$, $\eta_1 =^a \eta_2$ means that $\mathrm{card}(\{x \mid \eta_1(x) \neq \eta_2(x)\}) \leq a$. $\eta_1 \neq^a \eta_2$ means $\neg[\eta_1 =^a \eta_2]$. $\mathrm{domain}(\eta)$ and $\mathrm{range}(\eta)$ respectively denote the domain and range of partial function η. $f(A) = y$ is used as a shorthand for $(\forall x \in A)[f(x) = y]$. $\downarrow$ denotes defined and $\uparrow$ denotes undefined.

We fix φ to be an *acceptable programming system* [15, 16, 13] for the partial recursive functions: $\mathbb{N} \to \mathbb{N}$. φ_i denotes the partial recursive function computed by φ-program i. $\mathcal{R}$ denotes the class of all total recursive functions. Let Φ be an arbitrary Blum complexity measure [4] associated with acceptable programming system φ; such measures exist for any acceptable programming system [4]. $\mathrm{MinProg}(f)$ denotes $\min(\{i \mid \varphi_i = f\})$

The quantifiers '$\overset{\infty}{\forall}$' and '$\overset{\infty}{\exists}$' mean 'for all but finitely many' and 'there exist infinitely many,' respectively.

2.2   Information Sequences and Learning Machines

An *information sequence* is a mapping from $\mathbb{N}$ or an initial segment of $\mathbb{N}$, into $\{(x,y) \mid x, y \in \mathbb{N}\}$. We let G and T, with or without decorations, range over infinite information sequences. We let σ, τ range over finite information sequences. By $\sigma \subseteq \tau$ we mean that σ is an initial sequence of τ. $G[n]$, denotes the initial sequence of G of length n. $|\sigma|$ denotes the length of σ. $f[n]$, denotes the finite information sequence σ such that

$$\sigma(x) = \begin{cases} (x, f(x)), & \text{if } x < n; \\ \uparrow, & \text{otherwise.} \end{cases}$$

The content of an information sequence G, denoted $\mathrm{content}(G)$ is $\mathrm{range}(G)$. $\mathrm{content}(\sigma)$ is defined similarly. An information sequence, G, is for a function f if $\mathrm{content}(G) = \mathrm{graph}(f)$.

An *inductive inference machine* (IIM) is an algorithmic mapping from finite information sequences into $\mathbb{N}$. We let $\mathbf{M}$, with or without decorations, range over IIMs. $\sigma_1 \diamond (x,y)$ denotes the *concatenation* of (x,y) at the end of the information sequence σ_1; i.e. $\sigma = \sigma_1 \diamond (x,y)$ is defined as follows:

$$\sigma(i) = \begin{cases} \sigma_1(i), & \text{if } i < |\sigma_1|; \\ (x,y), & \text{if } i = |\sigma_1|; \\ \uparrow, & \text{otherwise.} \end{cases}$$

---

[1]Decorations are subscripts, superscripts, primes and the like.

2.3   Fundamental Function Identification Paradigms

In Definition 1 below we spell out what it means for an IIM to converge (in the limit) on an information sequence.

Definition 1 Suppose $\mathbf{M}$ is an IIM and G is an information sequence. $\mathbf{M}(G)\!\downarrow$ (read: $\mathbf{M}(G)$ *converges*; $\mathbf{M}$ converges on G) $\iff$ $(\exists i)(\overset{\infty}{\forall} n)\,[\mathbf{M}(G[n]) = i]$. If $\mathbf{M}(G)\!\downarrow$, then $\mathbf{M}(G)$ is defined $=$ the unique i such that $(\overset{\infty}{\forall} n)[\mathbf{M}(G[n]) = i]$, otherwise we say that $\mathbf{M}(G)$ diverges (written: $\mathbf{M}(G)\!\uparrow$).

We now introduce two different criteria for an IIM to successfully infer a function.

Definition 2 [9, 3, 6] Let $a \in \mathbb{N} \cup \{*\}$.
(a) $\mathbf{M}$ $\mathbf{Ex}^a$-*identifies* f (written: $f \in \mathbf{Ex}^a(\mathbf{M})$) $\iff$ ($\forall$ information sequences G for $f)(\exists i \mid \varphi_i =^a f)[\mathbf{M}(G)\!\downarrow = i]$.
(b) $\mathbf{Ex}^a = \{\mathcal{C} \mid (\exists \mathbf{M})[\mathcal{C} \subseteq \mathbf{Ex}^a(\mathbf{M})]\}$.

$\mathbf{Ex}$ in the above definition stands for *explanatory*.

Case and Smith [6] introduced another infinite hierarchy of identification criteria which we describe below. "$\mathbf{Bc}$" stands for *behaviorally correct*. Barzdin [2] essentially introduced $\mathbf{Bc}^0$.

Definition 3 [6] Let $a \in \mathbb{N} \cup \{*\}$.
(a) $\mathbf{M}$ $\mathbf{Bc}^a$-*identifies* f (written: $f \in \mathbf{Bc}^a(\mathbf{M})$) $\iff$ ($\forall$ information sequences G for $f)(\overset{\infty}{\forall} n)[\varphi_{\mathbf{M}(G[n])} =^a f]$.
(b) $\mathbf{Bc}^a = \{\mathcal{C} \mid (\exists \mathbf{M})[\mathcal{C} \subseteq \mathbf{Bc}^a(\mathbf{M})]\}$.

We usually write $\mathbf{Ex}$ for $\mathbf{Ex}^0$ and $\mathbf{Bc}$ for $\mathbf{Bc}^0$. For function identification with accurate data, identification (for criteria of inference discussed in this paper) from arbitrary information sequences is equivalent to identification from the canonical information sequence. Theorem 4 below describes some of the basic results about the two kinds of function identification criteria described above.

Theorem 4 *For all $a \in \mathbb{N}$,*
   (a) $\mathbf{Ex}^a \subset \mathbf{Ex}^{a+1}$.
   (b) $\bigcup_{a \in \mathbb{N}} \mathbf{Ex}^a \subset \mathbf{Ex}^*$.
   (c) $\mathbf{Ex}^* \subset \mathbf{Bc}$.
   (d) $\mathbf{Bc}^a \subset \mathbf{Bc}^{a+1}$.
   (e) $\bigcup_{a \in \mathbb{N}} \mathbf{Bc}^a \subset \mathbf{Bc}^*$.
   (f) $\mathcal{R} \in \mathbf{Bc}^*$.

Parts (a), (b), (d), and (e) are due to Case and Smith [6]. John Steel first observed that $\mathbf{Ex}^* \subseteq \mathbf{Bc}$ and diagonalization in part (c) is due to Harrington and Case [6]. Part (f) is due to Harrington [6]. Blum and Blum [3] first showed that $\mathbf{Ex} \subset \mathbf{Ex}^*$. Barzdin [2] independently showed $\mathbf{Ex} \subset \mathbf{Bc}$.

4

3 Inaccurate Data

We consider three kinds of inaccuracies that could creep into natural environments of learners.

- **Noisy data:** Ungrammatical intrusions into the language presented to the child is a very reasonable assumption about a child's environment. Similarly, experimental error caused by a faulty equipment could result in spurious data that is not representative of the reality under investigation.

- **Incomplete data:** Natural linguistic environments may omit sentences from the ambient language, and it is possible that the child's learning function can identify a natural language despite the systematic omission of sentences from its environment. Similarly, some experiments cannot be performed either due to technological limitations or due to ethical considerations.

- **Imperfect data:** Most natural linguistic environments are likely to be victims of both ungrammatical intrusions and omission of sentences from the ambient language. Such environments that contain a mixture of noisy and incomplete inaccuracies are referred to as environments with *imperfect* data. Similarly, in most experimental investigations, the inaccuracies are a mixture of both noisy and incomplete data.

The three kinds of inaccuracies discussed above yield three kinds of information sequences—noisy, incomplete, and imperfect. However, a further distinction is made based on whether the number of inaccuracies in an information sequence is finite or infinite. In [8, 14, 18] the case of finite number of inaccuracies was discussed. In this paper we examine the case when inaccuracies are infinite in number. We first introduce the definitions related with inference from finitely inaccurate information sequence.

It should be noted that the inaccuracies discussed here model spurious data and unavailability of data; they don't say anything about situations like "data is correct within 10% of actual value." Moreover, for our identification criteria in presence of inaccuracies, we require identification on all input information sequences where the amount of inaccuracies is bounded appropriately: thus the inaccuracies are not random but can be considered to be generated by an adversary.

3.1 Information Sequences with Finite Number of Inaccuracies

Pursuant to the classification of inaccuracies, we define three kinds of inaccurate information sequences for functions.

Definition 5 [8, 14] Let $a \in \mathbb{N} \cup \{*\}$.
(a) An information sequence G is *a-noisy* for f $\iff$ graph$(f) \subseteq$ content(G) and card(content$(G) -$ graph$(f)) \leq a$.
(b) An information sequence G is *a-incomplete* for f $\iff$ content$(G) \subseteq$ graph(f) and card(graph$(f) -$ content$(G)) \leq a$.
(c) An information sequence G is *a-imperfect* for f $\iff$ card(graph$(f)\mathbf{\Delta}$content$(G)) \leq a$.

An a-noisy information sequence for f can be viewed as an information sequence for f into which up to a "extra" pairs have been inserted. Note that any single such intrusion may occur infinitely often in G. Similarly, a-incomplete information sequences, have at most a pairs removed from them and a-imperfect information sequences have at most a pairs inserted/deleted from them.

In the above definitions, $a = *$ case implies that the number of inaccuracies is any finite number. The other $a \in \mathbb{N}$ cases model situations when a scientist may be aware, apriori, of an upper bound on the number of inaccuracies infesting its environment; possible sources of such information could be previous experience and nature of instruments used.

Note that in the case of noisy information sequences for functions, two incorrect values for $f(n)$ count as two distinct noise points, i.e., if the correct value of $f(n) = x$ and both (n, y) and (n, z), where x, y and z are distinct, are present in an inaccurate information sequence for f, then the data points (n, y) and (n, z) contribute separately to noise count. Also, if the actual value of $f(n) = y$, but (n, y) doesn't appear in an information sequence and instead (n, z), $y \neq z$, appears, then these contribute two to imperfection count.

We now introduce the learning criteria based on finite number of inaccuracies in the input.

Definition 6 [8, 14] Let $a, b \in \mathbb{N} \cup \{*\}$.
(a.1) $\quad$ **M** $\quad$ $\mathbf{N}^a\mathbf{Ex}^b$-*identifies* $\quad$ f $\quad$ (written: $\quad$ f $\quad$ $\in$ $\quad$ $\mathbf{N}^a\mathbf{Ex}^b(\mathbf{M})$) $\quad\quad$ $\Longleftrightarrow$
($\forall$ a-noisy information sequences G for f)$[\mathbf{M}(G)\downarrow \ \wedge \ \varphi_{\mathbf{M}(G)} =^b f]$.
(a.2) $\mathbf{N}^a\mathbf{Ex}^b = \{\mathcal{C} \mid (\exists \mathbf{M})[\mathcal{C} \subseteq \mathbf{N}^a\mathbf{Ex}^b(\mathbf{M})]\}$.
(b.1) $\quad$ **M** $\quad$ $\mathbf{In}^a\mathbf{Ex}^b$-*identifies* $\quad$ f $\quad$ (written: $\quad$ f $\quad$ $\in$ $\quad$ $\mathbf{In}^a\mathbf{Ex}^b(\mathbf{M})$) $\quad\quad$ $\Longleftrightarrow$
($\forall$ a-incomplete information sequences G for f)$[\mathbf{M}(G)\downarrow \ \wedge \ \varphi_{\mathbf{M}(G)} =^b f]$.
(b.2) $\mathbf{In}^a\mathbf{Ex}^b = \{\mathcal{C} \mid (\exists \mathbf{M})[\mathcal{C} \subseteq \mathbf{In}^a\mathbf{Ex}^b(\mathbf{M})]\}$.
(c.1) $\quad$ **M** $\quad$ $\mathbf{Im}^a\mathbf{Ex}^b$-*identifies* $\quad$ f $\quad$ (written: $\quad$ f $\quad$ $\in$ $\quad$ $\mathbf{Im}^a\mathbf{Ex}^b(\mathbf{M})$) $\quad\quad$ $\Longleftrightarrow$
($\forall$ a-imperfect information sequences G for f)$[\mathbf{M}(G)\downarrow \ \wedge \ \varphi_{\mathbf{M}(G)} =^b f]$.
(c.2) $\mathbf{Im}^a\mathbf{Ex}^b = \{\mathcal{C} \mid (\exists \mathbf{M})[\mathcal{C} \subseteq \mathbf{Im}^a\mathbf{Ex}^b(\mathbf{M})]\}$.

Similar to the above definitions one can define the function identification paradigms: $\mathbf{N}^a\mathbf{Bc}^b$, $\mathbf{In}^a\mathbf{Bc}^b$, $\mathbf{Im}^a\mathbf{Bc}^b$.

We now turn our attention to potentially infinite number of inaccuracies.

3.2 Information Sequences with Infinite Number of Inaccuracies

We first define density notions needed to measure the amount of inaccuracy in the input. These notions of "density" are from [17]. Similar notions were also used by Smith and Velauthapillai [19] in the context of inductive inference.

Definition 7 (S. Tennenbaum: see page 156 in [16], [17])
(a) Suppose that $A \subseteq \mathbb{N}$ and that B is a finite, nonempty subset of $\mathbb{N}$. We define the *density of A in B* (denoted: $\mathbf{den}(A; B)$) as $\text{card}(A \cap B)/\text{card}(B)$.
(b) The *density* of a set A (denoted: $\mathbf{den}(A)$) is $\lim_{n \to \infty} \inf(\{\mathbf{den}(A; \{z \mid z \leq x\}) \mid x \geq n\})$.

Intuitively, $\mathbf{den}(A; B)$ can be thought of as the probability of selecting an element of A when choosing an arbitrary element from B.

Note that, even if $\mathbf{den}(A)$ is 1, A may have "large holes". To overcome this situation, we consider the notion of "uniform density" from [17].

Definition 8 [17] The *uniform density* of a set A in intervals of length $\geq n$ (denoted: $\mathbf{uden}_n(A)$) is $\inf(\{\mathbf{den}(A; \{z \mid x \leq z \leq y\}) \mid x, y \in \mathbb{N}$ *and* $y - x \geq n\})$. *Uniform density* of A (denoted: $\mathbf{uden}(A)$) is $\lim_{n \to \infty} \mathbf{uden}_n(A)$.

We now define the inaccurate information sequences with certain density.

Definition 9 Suppose $0 \leq d \leq 1$. An information sequence G is d-Dnoisy for a total function f if

(a) $\mathrm{graph}(f) \subseteq \mathrm{content}(G)$,
(b) $\mathbf{den}(\mathbb{N} - \{x \mid (\exists y)[(x, y) \in \mathrm{content}(G) - \mathrm{graph}(f)]\}) \geq 1 - d$ and
(c) $(\forall x)[\mathrm{card}(\{(x, y) \mid (x, y) \in \mathrm{content}(G) - \mathrm{graph}(f)\}) < \infty]$.

D in Dnoisy indicates that the density of inaccuracy is considered. Note the difference in the way the inaccuracies in the information sequence are counted for finite inaccuracies and infinite inaccuracies. Instead of the definition used in clause (b) above we may want to define the density of the noise in an information sequence as the limiting value of the ratio:
$$\frac{\text{number of erroneous elements in the information sequence for inputs}_{\leq x}}{x+1}.$$
We feel that this is not a natural definition for infinite inaccuracies, since, for such a definition, the density of noise can be infinite. Clause (c) has been added since we believe that the number of possible outcomes, even allowing for errors, in any particular experiment is bounded. Some of our proofs are dependent on clause (c).

The following definitions give the corresponding notions for incomplete and imperfect information sequences.

Definition 10 Suppose $0 \leq d \leq 1$. An information sequence G is d-Dincomplete for a total function f if

(a) $\mathrm{content}(G) \subseteq \mathrm{graph}(f)$ and
(b) $\mathbf{den}(\mathbb{N} - \{x \mid (\exists y)[(x, y) \in \mathrm{graph}(f) - \mathrm{content}(G)]\}) \geq 1 - d$.

Definition 11 Suppose $0 \leq d \leq 1$. An information sequence G is d-Dimperfect for a total function f if $\mathbf{den}(\mathbb{N} - \{x \mid (\exists y)[(x, y) \in \mathrm{content}(G)\mathbf{\Delta}\mathrm{graph}(f)]\}) \geq 1 - d$.

Note that the equivalent of clause (c) in the definition of Dnoisy information sequence is not necessary for the definition of Dincomplete and Dimperfect information sequences and thus has been dropped.

Similarly by considering uniform density one can define d-UDnoisy, d-UDincomplete, and d-UDimperfect information sequences.

3.3 Identification Criteria on Infinitely Inaccurate Information Sequences

We now define the corresponding notions of function identification.

Definition 12 Suppose $0 \leq d \leq 1$ and $a \in \mathbb{N} \cup \{*\}$.
(a.1) $\mathbf{M}$ $\mathbf{DN}^d\mathbf{Ex}^a$ identifies f (written: $f \in \mathbf{DN}^d\mathbf{Ex}^a(\mathbf{M})$) iff $(\forall\ d\text{-Dnoisy information sequences } G \text{ for } f)[\mathbf{M}(G){\downarrow} \wedge \varphi_{\mathbf{M}(G)} =^a f]$.
(a.2) $\mathbf{DN}^d\mathbf{Ex}^a = \{\mathcal{C} \mid (\exists \mathbf{M})[\mathcal{C} \subseteq \mathbf{DN}^d\mathbf{Ex}^a(\mathbf{M})]\}$.
(b.1) $\mathbf{M}$ $\mathbf{DIn}^d\mathbf{Ex}^a$ identifies f (written: $f \in \mathbf{DIn}^d\mathbf{Ex}^a(\mathbf{M})$) iff

($\forall$ d-Dincomplete information sequences G for f)$[\mathbf{M}(G){\downarrow}\ \wedge\ \varphi_{\mathbf{M}(G)}=^a f]$.

(b.2) $\mathbf{DIn}^d\mathbf{Ex}^a = \{\mathcal{C} \mid (\exists \mathbf{M})[\mathcal{C} \subseteq \mathbf{DIn}^d\mathbf{Ex}^a(\mathbf{M})]\}$.

(c.1) $\mathbf{M}$ $\mathbf{DIm}^d\mathbf{Ex}^a$ identifies f (written: $f \in \mathbf{DIm}^d\mathbf{Ex}^a(\mathbf{M})$) iff ($\forall$ d-Dimperfect information sequences G for f)$[\mathbf{M}(G){\downarrow}\ \wedge\ \varphi_{\mathbf{M}(G)}=^a f]$.

(c.2) $\mathbf{DIm}^d\mathbf{Ex}^a = \{\mathcal{C} \mid (\exists \mathbf{M})[\mathcal{C} \subseteq \mathbf{DIm}^d\mathbf{Ex}^a(\mathbf{M})]\}$.

We can similarly define $\mathbf{UDN}^d\mathbf{Ex}^a, \mathbf{UDIn}^d\mathbf{Ex}^a, \mathbf{UDIm}^d\mathbf{Ex}^a$, $\mathbf{DN}^d\mathbf{Bc}^a$, $\mathbf{DIn}^d\mathbf{Bc}^a, \mathbf{DIm}^d\mathbf{Bc}^a, \mathbf{UDN}^d\mathbf{Bc}^a, \mathbf{UDIn}^d\mathbf{Bc}^a$, and $\mathbf{UDIm}^d\mathbf{Bc}^a$.

4 Hierarchy Results

The following theorem demonstrates the disadvantages of increasing the density of noise. It establishes that there are collections of functions that can be **Ex**-identified with noise of a particular density, but cannot be identified if the density of noise is increased, even if the noise is of uniform type and a more liberal criterion of success is used.

Theorem 13 *Suppose* $0 \le d_1 < d_2 \le 1$. $\mathbf{DN}^{d_1}\mathbf{Ex} - [\mathbf{UDN}^{d_2}\mathbf{Ex}^* \cup \bigcup_{l\in\mathbb{N}} \mathbf{UDN}^{d_2}\mathbf{Bc}^l] \ne \emptyset$.

Corollary 14 *Suppose* $0 \le d_1 < d_2 \le 1$. $\mathbf{DN}^{d_1}\mathbf{Ex} - \mathbf{UDN}^{d_2}\mathbf{Ex}^* \ne \emptyset$.

PROOF OF THEOREM 13. Without loss of generality assume $d_1 = j/n$, $d_2 = (j+4)/n$, where $n > j+4$ and $j, n \in \mathbb{N}$. Consider the following classes of functions

$\mathcal{C}_0 = \{f \mid \varphi_{f(0)} = f\ \wedge\ (\overset{\infty}{\exists} x)[f(x) \ne 0]\}$, and $\mathcal{C}_1 = \{f \mid (\overset{\infty}{\forall} x)[f(x) = 0]\}$.

It was shown in [6] that $\mathcal{C}_0 \cup \mathcal{C}_1 \notin \mathbf{Ex}^* \cup \bigcup_{l\in\mathbb{N}} \mathbf{Bc}^l$. We will use a modification of $\mathcal{C}_0 \cup \mathcal{C}_1$ as our diagonalizing class.

Let $N_0 = 0$ and, for $i > 0$, $N_i = n^i$.

Let $X = \{x \mid (\exists r)[N_{2r} \le x < N_{2r+1}]\}$, $X_0 = X \cap \{x \mid (x \bmod n) < (j+4)\}$, and $X_1 = X - X_0$.

Let $R_{k,j} = \{x \mid [N_{2\cdot\langle k,j\rangle+1} \le x < N_{2\cdot\langle k,j\rangle+2}]\}$, and $S_k = \bigcup_{j\in\mathbb{N}} R_{k,j}$.

Now for $f \in \mathcal{C}_0 \cup \mathcal{C}_1$ define a function F_f as follows.

$F_f(S_k) = f(k)$.

$F_f(X_1) = 0$.

$F_f(X_0) = 0$, if $f \in \mathcal{C}_0$; $F_f(X_0) = 1$, otherwise.

Let $\mathcal{C} = \{F_f \mid f \in \mathcal{C}_0 \cup \mathcal{C}_1\}$.

Claim 15 $\mathcal{C} \in \mathbf{DN}^{d_1}\mathbf{Ex}$.

PROOF. Suppose G is a d_1-Dnoisy information sequence for $F_f \in \mathcal{C}$.

Thus we have:

1. $(\forall x \in N)[(x, F_f(x)) \in \text{content}(G)]$
2. $(\forall y)[\ [(\forall x \in X_0)[(x, y) \in \text{content}(G)]] \Rightarrow F_f(X_0) = y\ ]$

   (This holds because: for large enough r, the fraction of noisy points less than N_{2r+1}, is bounded by $(j+1)/n$. Now since $X_0 \cap \{x \mid x < N_{2r+1}\}$, consists of at least $(j+4)/n$ fraction of points less than N_{2r+1}, there exists a point in X_0 which is noise free.)

8

3. $(\overset{\infty}{\forall} \langle j,k\rangle)(\forall y)[\ [(\forall x \in R_{j,k})[(x,y) \in \text{content}(G)]] \Rightarrow F_f(S_j) = y\ ]$.

   (Let $\langle j,k\rangle$ be large enough, such that the fraction of noisy points below $\max(R_{j,k})$ is bounded by $(j+1)/n$. Now since $R_{j,k}$ consists of $(n-1)/n$ fraction of points below $\max(R_{j,k})$, there exists a point in $R_{j,k}$ which is noise free.)

Since (1, 2) hold, it is easy to determine $F_f(X_0)$ in the limit from G. Also, since (1, 3) hold, it is easy to determine $F_f(S_0)$ in the limit. Now if $F_f(X_0) = 0$, then i defined as follows is a program for F_f,

$$\varphi_i(x) = \begin{cases} 0, & \text{if } x \in X_0; \\ 0, & \text{if } x \in X_1; \\ \varphi_{F_f(S_0)}(k), & \text{if } x \in S_k. \end{cases}$$

If $F_f(X_0) = 1$, then $W = \{k \mid F_f(S_k) \neq 0\}$ can be determined in the limit (since (1, 3) hold). A program for F_f can then easily be constructed from W. Thus $\mathcal{C} \in \mathbf{DN}^{d_1}\mathbf{Ex}$. $\square$

Claim 16 $\mathcal{C} \notin [\mathbf{UDN}^{d_2}\mathbf{Ex}^* \cup \bigcup_{l \in \mathbb{N}} \mathbf{UDN}^{d_2}\mathbf{Bc}^l]$.

PROOF. For $f \in \mathcal{C}_0 \cup \mathcal{C}_1$, we will show (i) how to convert an information sequence for f to an d_2-UDnoisy information sequence for F_f and (ii) (for $a \in \mathbb{N} \cup \{*\}$) how to convert an a-error program for F_f into an a-error program for f.

Assuming this we have $[\mathcal{C} \in \mathbf{UDN}^{d_2}\mathbf{Ex}^* \cup \bigcup_{l \in \mathbb{N}} \mathbf{UDN}^{d_2}\mathbf{Bc}^l] \Rightarrow [\mathcal{C}_0 \cup \mathcal{C}_1 \in \mathbf{Ex}^* \cup \bigcup_{l \in \mathbb{N}} \mathbf{Bc}^l]$. Since, $[\mathcal{C}_0 \cup \mathcal{C}_1 \notin \mathbf{Ex}^* \cup \bigcup_{l \in \mathbb{N}} \mathbf{Bc}^l]$ we conclude that $\mathcal{C} \notin \mathbf{UDN}^{d_2}\mathbf{Ex}^* \cup \bigcup_{l \in \mathbb{N}} \mathbf{UDN}^{d_2}\mathbf{Bc}^l$.

From an information sequence, G, for f a d_2-UDnoisy information sequence can be constructed for F_f, by forming an information sequence G_{F_f} such that $\text{content}(G_{F_f}) = \{(x,0) \mid x \in X\} \cup \{(x,1) \mid x \in X_0\} \cup \{(x,f(k)) \mid x \in S_k\}$. Note, that this construction can be done effectively.

Also, since $f(k) = F_f(S_k) = F_f(N_{2\cdot\langle k,j\rangle+1})$, it is easy to convert an a-error program for F_f into an a-error program for f. This completes the proof of the claim $\square$ ∎

It can be similarly shown that $\mathcal{C}$ defined in the above proof is in $\mathbf{DIn}^{d_1}\mathbf{Ex} - [\mathbf{UDIn}^{d_2}\mathbf{Ex}^* \cup \bigcup_{l \in \mathbb{N}} \mathbf{UDIn}^{d_2}\mathbf{Bc}^l]$ and $\mathbf{DIm}^{d_1/2}\mathbf{Ex} - [\mathbf{UDIm}^{d_2/2}\mathbf{Ex}^* \cup \mathbf{UDIm}^{d_2/2}\mathbf{Bc}^l]$. Thus we have,

Theorem 17 *Suppose* $0 \leq d_1 < d_2 \leq 1$.
$[\mathbf{DN}^{d_1}\mathbf{Ex} \cap \mathbf{DIn}^{d_1}\mathbf{Ex} \cap \mathbf{DIm}^{d_1/2}\mathbf{Ex}] -$
$[(\mathbf{UDN}^{d_2}\mathbf{Ex}^* \cup \mathbf{UDIn}^{d_2}\mathbf{Ex}^* \cup \mathbf{UDIm}^{d_2/2}\mathbf{Ex}^*) \cup \bigcup_{l \in \mathbb{N}} (\mathbf{UDN}^{d_2}\mathbf{Bc}^l \cup \mathbf{UDIn}^{d_2}\mathbf{Bc}^l \cup \mathbf{UDIm}^{d_2/2}\mathbf{Bc}^l)] \neq \emptyset$.

Corollary 18 *Suppose* $0 \leq d_1 < d_2 \leq 1$. $\mathbf{DIn}^{d_1}\mathbf{Ex} - \mathbf{UDIn}^{d_2}\mathbf{Ex}^* \neq \emptyset$.

Corollary 19 *Suppose* $0 \leq d_1 < d_2 \leq 1/2$. $\mathbf{DIm}^{d_1}\mathbf{Ex} - \mathbf{UDIm}^{d_2}\mathbf{Ex}^* \neq \emptyset$.

Thus identification criteria based on inaccurate information form a strict hierarchy based on the density of inaccuracy.

Let $\mathcal{C}$ be a class of functions such that there exist functions f_1, f_2, such that $f_1 \neq^{2a} f_2$. Then it is easy to see that $\mathcal{C} \notin \mathbf{UDIm}^{1/2}\mathbf{Bc}^a$ (since the input information sequence may be $1/2$-UDimperfect for both f_1 and f_2). For $d < 1/2$ we do not know if $\mathbf{DIm}^d\mathbf{Ex} - \mathbf{DN}^{2d}\mathbf{Ex}$ is empty or not.

We now consider the advantages of a uniformity restriction on the density of inaccuracies over the situation where such a restriction is not there. The next result demonstrates this advantage in the context of noise by showing that there are collections of functions that can be **Ex**-identified on information sequences with a uniform noise density < 1, but cannot be **Ex***-identified even with 0-density noise if the uniformity constraint on the density of noise is removed.

Theorem 20 *Suppose $0 \leq d < 1$.* $\mathbf{UDN}^d\mathbf{Ex} - \mathbf{DN}^0\mathbf{Ex}^* \neq \emptyset$.

PROOF. Without loss of generality let $d = (n-2)/n$, where $n > 2$. Let $\mathcal{C}_0$ and $\mathcal{C}_1$ be as defined in the proof of Theorem 13. We will use a modification of $\mathcal{C}_0 \cup \mathcal{C}_1$ as our diagonalizing class. Let $N_0 = 0$, $N_{2i+1} = N_{2i} + (i+1) * n$, and $N_{2i+2} = N_{2i+1} * n$.
Let $X = \{x \mid (\exists j)[N_{2j} \leq x < N_{2j+1}]\}$, $R_{j,k} = \{x \mid N_{2 \cdot \langle j,k \rangle + 1} \leq x < N_{2 \cdot \langle j,k \rangle + 2}\}$, and $S_j = \bigcup_{k \in \mathbb{N}} R_{j,k}$.
Now for $f \in \mathcal{C}_0 \cup \mathcal{C}_1$, define F_f as follows. $F_f(S_j) = f(j)$. $F_f(X) = 0$, if $f \in \mathcal{C}_0$; $F_f(X) = 1$ otherwise.
Let $\mathcal{C} = \{F_f \mid f \in \mathcal{C}_0 \cup \mathcal{C}_1\}$.
Now proceeding in a way similar to that of Theorem 13 it can be shown that $\mathcal{C} \in \mathbf{UDN}^d\mathbf{Ex} - \mathbf{DN}^0\mathbf{Ex}^*$. ∎

We can similarly show that

Theorem 21 *Suppose $0 \leq d < 1$.*
$[\mathbf{UDN}^d\mathbf{Ex} \cap \mathbf{UDIn}^d\mathbf{Ex} \cap \mathbf{UDIm}^{d/2}\mathbf{Ex}] -$
$[(\mathbf{DN}^0\mathbf{Ex}^* \cup \mathbf{DIn}^0\mathbf{Ex}^*) \cup \bigcup_{l \in \mathbb{N}}(\mathbf{DN}^0\mathbf{Bc}^l \cup \mathbf{DIn}^0\mathbf{Bc}^l)] \neq \emptyset$.

Corollary 22 *Suppose $0 \leq d < 1$.* $\mathbf{UDIn}^d\mathbf{Ex} - \mathbf{DIn}^0\mathbf{Ex}^* \neq \emptyset$.

Corollary 23 *Suppose $0 \leq d < 1/2$.* $\mathbf{UDIm}^d\mathbf{Ex} - \mathbf{DIm}^0\mathbf{Ex}^* \neq \emptyset$.

The next result compares the case of finite number of errors with infinite number of errors. It demonstrates that there are collections of functions that can be **Ex**-identified from information sequences that contain a finite number of imperfections, but cannot be learned from uniform 0-density noise even if a more liberal criterion of identification is used.

Theorem 24 $\mathbf{Im}^*\mathbf{Ex} - [\mathbf{UDN}^0\mathbf{Ex} \cup \bigcup_{l \in \mathbb{N}} \mathbf{UDN}^0\mathbf{Bc}^l]$.

PROOF. Let $\mathcal{C}_0$ and $\mathcal{C}_1$ be as defined in the proof of Theorem 13.
We will use a modification of $\mathcal{C}_0 \cup \mathcal{C}_1$ as our diagonalizing class.
Let $X = \{2^n + 1 \mid n \geq 1\}$. Let $S_k = \{2 \cdot \langle k, z \rangle \mid z \in \mathbb{N}\}$. Let $Z = \mathbb{N} - (X \cup \bigcup_k S_k)$.
For $f \in \mathcal{C}_0 \cup \mathcal{C}_1$, define a function F_f as follows.
$F_f(S_k) = f(k)$,
$F_f(Z) = 0$,
$F_f(X) = 0$, if $f \in \mathcal{C}_0$; $F_f(X) = 1$, otherwise.
Let $\mathcal{C} = \{F_f \mid f \in \mathcal{C}_0 \cup \mathcal{C}_1\}$.

Claim 25 $\mathcal{C} \in \mathbf{Im}^*\mathbf{Ex}$.

PROOF. Suppose G is a $*$-imperfect information sequence for $F_f \in \mathcal{C}$. Thus, it is easy to determine $F_f(X)$ and $F_f(S_0)$, in the limit from G. Now if $F_f(X) = 0$, then i defined as follows is a program for F_f,

$$\varphi_i(x) = \begin{cases} 0, & \text{if } x \in X; \\ 0, & \text{if } x \in Z; \\ \varphi_{F_f(S_0)}(k), & \text{if } x \in S_k. \end{cases}$$

If $F_f(X) = 1$, then $W = \{k \mid F_f(S_k) \neq 0\}$ can be determined in the limit (since G is $*$-imperfect information sequence for F_f). A program for F_f can then easily be constructed from W. Thus $\mathcal{C} \in \mathbf{Im}^*\mathbf{Ex}$. $\square$

Claim 26 $\mathcal{C} \notin [\mathbf{UDN}^0\mathbf{Ex} \cup \bigcup_{l \in \mathbb{N}} \mathbf{UDN}^0\mathbf{Bc}^l]$.

PROOF. For $f \in \mathcal{C}_0 \cup \mathcal{C}_1$, we will show (i) how to convert an information sequence for f to an 0-UDnoisy information sequence for F_f and (ii) (for $a \in \mathbb{N} \cup \{*\}$) how to convert an a-error program for F_f into an a-error program for f.

Assuming this we have $[\mathcal{C} \in \mathbf{UDN}^0\mathbf{Ex}^* \cup \bigcup_{l \in \mathbb{N}} \mathbf{UDN}^0\mathbf{Bc}^l] \Rightarrow [\mathcal{C}_0 \cup \mathcal{C}_1 \in \mathbf{Ex}^* \cup \bigcup_{l \in \mathbb{N}} \mathbf{Bc}^l]$. Since, $[\mathcal{C}_0 \cup \mathcal{C}_1 \notin \mathbf{Ex}^* \cup \bigcup_{l \in \mathbb{N}} \mathbf{Bc}^l]$ we conclude that $\mathcal{C} \notin \mathbf{UDN}^0\mathbf{Ex}^* \cup \bigcup_{l \in \mathbb{N}} \mathbf{UDN}^0\mathbf{Bc}^l$.

From an information sequence, G, for f a 0-UDnoisy information sequence can be constructed for F_f, by forming an information sequence G_{F_f} such that content$(G_{F_f}) = \{(x, 0) \mid x \in X \cup Z\} \cup \{(x, 1) \mid x \in X\} \cup \{(x, f(k)) \mid x \in S_k\}$. Note, that this construction can be done effectively.

Also, since $f(k) = F_f(S_k) = F_f(2 \cdot \langle k, 0 \rangle)$, it is easy to convert an a-error program for F_f into an a-error program for f. This completes the proof of the claim $\square$ ∎

The next result parallels the above theorem for incomplete data.

Theorem 27 $\mathbf{Im}^*\mathbf{Ex} - [\mathbf{UDIn}^0\mathbf{Ex} \cup \bigcup_{l \in \mathbb{N}} \mathbf{UDIn}^0\mathbf{Bc}^l]$.

As a corollary to results in this section and results from [8] we have,

Corollary 28 *Suppose* $0 \leq d < d' \leq 1$, $a \in \mathbb{N} \cup \{*\}$. *Then,* $\mathbf{Ex}^a \supset \mathbf{N}^1\mathbf{Ex}^a \supset \cdots \supset \mathbf{N}^*\mathbf{Ex}^a \supset \mathbf{DN}^0\mathbf{Ex}^a \supset \mathbf{DN}^d\mathbf{Ex}^a \supset \mathbf{DN}^{d'}\mathbf{Ex}^a$.

Similar corollaries can be obtained for incomplete information, imperfect information and for uniform inaccuracies.

5 Comparison Between Different Types of Inaccuracies

We now compare the effects of different kinds of inaccuracies. The following theorem demonstrates the advantages of noise over missing data by establishing that there are collections of functions that can be **Ex**-identified with high noise density but cannot be identified from information sequences in which a single data is missing even if the final program is allowed to make a finite number of errors.

Theorem 29 *Suppose* $0 \leq d < 1$. $\mathbf{DN}^d\mathbf{Ex} - \mathbf{In}^1\mathbf{Ex}^* \neq \emptyset$.

PROOF. Without loss of generality let $d = (n-2)/n$.

For $i \in \mathbb{N}$, let $N_i = n^i$.

Let $R_{j,k} = \{x \mid N_{\langle j,k \rangle} \le x < N_{\langle j,k \rangle+1}\}$, and $S_j = \bigcup_{k \in \mathbb{N}} R_{j,k}$.

Now for $f \in \mathcal{R}$, define F_f as follows. $F_f(S_j) = f(j)$. $F_f(0) = \langle i, \langle \mathrm{err}_0, \mathrm{err}_1, \mathrm{err}_2, \ldots, \mathrm{err}_{i-1} \rangle \rangle$, where $i = \mathrm{MinProg}(f)$, and for $j < i$, $\mathrm{err}_j = \min(\{x \mid \varphi_j(x) \ne f(x)\})$.

Let $\mathcal{C} = \{F_f \mid f \in \mathcal{R}\}$.

Claim 30 $\mathcal{C} \notin \mathbf{In^1 Ex^*}$.

PROOF. For $f \in \mathcal{R}$, we will show below (i) how to convert an information sequence for f to an 1-incomplete information sequence for F_f, and (ii) how to convert a $*$-error program for F_f to a $*$-error program for f. From this it follows that, $\mathcal{C} \in \mathbf{In^1 Ex^*} \Rightarrow \mathcal{R} \in \mathbf{Ex^*}$. Since $\mathcal{R} \notin \mathbf{Ex^*}$, we conclude that $\mathcal{C} \notin \mathbf{In^1 Ex^*}$.

Given an information sequence G for $f \in \mathcal{R}$, let G' be an information sequence such that $\mathrm{content}(G') = \{(x, f(k)) \mid x \in S_k\}$. Note that such a G' can be effectively computed from G. Also since $f(k) = F_f(S_k) = F_f(N_{\langle k,0 \rangle})$, a $*$-error program for F_f can be easily converted to a $*$-error program for f. $\square$

Claim 31 $\mathcal{C} \in \mathbf{DN^d Ex}$.

PROOF. We describe an IIM $\mathbf{M}$ which $\mathbf{DN^d Ex}$-identifies $\mathcal{C}$.

Suppose $f \in \mathcal{R}$ and G is an d-Dnoisy information sequence for F_f ($\in \mathcal{C}$). We describe how $\mathbf{M}$ computes its output on $G[n]$. For this we first describe, $X_n, Y_n^j, Z_n, e_n, z_n$ (which depend on G, n). Let

$X_n = \{x \mid (0, x) \in \mathrm{content}(G[n])\}$,

$Y_n^j = \{\langle k, y \rangle \mid (\forall x \in R_{j,k})[(x, y) \in \mathrm{content}(G[n])]\}$,

$Z_n = \{(j, y) \mid (\exists k)[\langle k, y \rangle = \max(Y_n^j)]\}$.

Note that $\frac{\mathrm{card}(R_{j,k})}{1 + \max(R_{j,k})} = \frac{N_{\langle j,k \rangle+1} - N_{\langle j,k \rangle}}{N_{\langle j,k \rangle+1}} = \frac{n-1}{n}$. Thus, for all but finitely many $\langle j, k \rangle$, if $(\forall x \in R_{j,k})[(x, y) \in \mathrm{content}(G[n])]$, then $f(j)$ must be equal to y. Thus, for large enough n, $Z_n \subseteq \mathrm{graph}(f)$.

Let $e_n = \max(\{i \mid (\exists err_0, err_1, \ldots, err_{i-1} \mid \langle i, \langle err_0, err_1, err_2, \ldots, err_{i-1} \rangle \rangle \in X_n)[(\forall j < i)[\Phi_j(err_j) > n \lor (err_j, \varphi_j(err_j)) \notin Z_n]]\})$.

It is easy to see that, for large enough n, $e_n = \mathrm{MinProg}(f)$.

Let $z_n = \langle e_n, \langle err_0^n, \ldots, err_{e_n-1}^n \rangle \rangle$,

where, for $j < e_n$, $err_j^n = \min(\{n\} \cup \{x < n \mid \Phi_j(x) > n \lor \Phi_{e_n}(x) > n \lor \varphi_j(x) \ne \varphi_{e_n}(x)\})$. From the definition of F_f, it follows that for large enough n, $z_n = F_f(0)$.

Let h be a recursive function such that, for all e, z, j, $\varphi_{h(e,z)}(0) = z$ and $\varphi_{h(e,z)}(S_j) = \varphi_e(j)$. Let $\mathbf{M}(G[n]) = h(e_n, z_n)$. It is easy to see that, $F_f \in \mathbf{DN^d Ex}(\mathbf{M})$. Since F_f was an arbitrary member of $\mathcal{C}$, we have $\mathcal{C} \subseteq \mathbf{DN^d Ex}(\mathbf{M})$. $\square$ $\blacksquare$ (Theorem 29)

Similarly we can establish the following result.

Theorem 32 *Suppose* $0 \le d < 1$. $\mathbf{DN^d Ex} - \bigcup_{j \in \mathbb{N}} \mathbf{In^1 Bc^j} \ne \emptyset$.

The following theorem shows the advantages of incomplete information sequences over imperfect information sequences.

12

Theorem 33 *Suppose $0 \leq d < 1$.* $\mathbf{DIn}^d\mathbf{Ex}^* - (\mathbf{Im}^*\mathbf{Ex}^* \cup \bigcup_{j\in\mathbb{N}} \mathbf{Im}^*\mathbf{Bc}^j) \neq \emptyset$.

PROOF. Consider the following class of functions:

$\mathcal{C} = \{f \mid (\forall y \in range(f))[\varphi_y =^* f]\}$.

Clearly $\mathcal{C} \in \mathbf{DIn}^d\mathbf{Ex}^*$. We argue that $\mathcal{C} \notin \mathbf{Im}^*\mathbf{Ex}^*$. (The argument that $\mathcal{C}$ is not in any of the $\mathbf{Im}^*\mathbf{Bc}^j$ classes is similar and we omit the details). Suppose by way of contradiction that $\mathbf{M}$ $\mathbf{Im}^*\mathbf{Ex}^*$-identifies $\mathcal{C}$. Then by Operator Recursion Theorem [5] there exists a 1-1 recursive p such that the functions $\varphi_{p(\cdot)}$ may be described as follows. Let x_s denote the least x such that $\varphi_{p(0)}(x)$ is not defined before stage s. Let $\varphi_{p(0)}(0) = p(0)$. Let $\sigma_1 = ((0, p(0)))$. The following properties will be maintained in the construction.

(A) At the beginning of stage s, $\varphi_{p(2s)}$ and $\varphi_{p(2s+1)}$ are not defined on any input.
(B) $\sigma_s = \varphi_{p(0)}[x_s]$.
(C) For each i, if $p(i)$ is in the range of $\varphi_{p(0)}$, then $\varphi_{p(i)} =^* \varphi_{p(0)}$.

Go to stage 1.

Stage s

1. For all $x < x_s$, let $\varphi_{p(2s)}(x) = p(2s)$.
   Dovetail steps $2, 3, 4$ until step 3 or 4 succeeds. If step 3 succeeds before step 4 does, if ever, then go to step 5. If step 4 succeeds before step 3 does, if ever, then go to step 6.
2. Let $z = x_s$. Go to substage 0.
   Substage s'
        Let $\varphi_{p(2s)}(z) = p(2s)$.
        Let $z = z + 1$.
        Go to substage $s' + 1$.
   End substage s'.
3. Search for $y > x_s$ such that $\varphi_{\mathbf{M}(\sigma_s)}(y)\downarrow = p(2s)$.
4. Search for $y > x_s$ such that $\mathbf{M}(\sigma_s) \neq \mathbf{M}(\sigma_s \diamond (x_s, p(2s)) \diamond \ldots \diamond (y, p(2s)))$.
5. Let y be as found in step 3.
   For $x_s \leq x \leq y$, let $\varphi_{p(0)}(x) = p(2s + 1)$.
   For $x \leq y$, let $\varphi_{p(2s+1)}(x) = \varphi_{p(0)}(x)$. Let $\varphi_{p(2s+1)}$ follow $\varphi_{p(0)}$ from now on (i.e. whenever $\varphi_{p(0)}(x)$ gets defined for $x > y$, let $\varphi_{p(2s+1)}(x) = \varphi_{p(0)}(x)$).
   Let $\sigma_{s+1} = \sigma_s \diamond (x_s, p(2s + 1)) \diamond \ldots \diamond (y, p(2s + 1))$.
   Go to stage $s + 1$.
   (Note that if $p(i)$ is in the range of $\varphi_{p(0)}$, then $\varphi_{p(i)} =^* \varphi_{p(0)}$).
6. Let y be as found in step 4. Let z be as at the beginning of the last substage executed in step 2.
   Let $y' = \max(\{y, z\})$.
   For $x_s \leq x \leq y'$, let $\varphi_{p(0)}(x) = \varphi_{p(2s)}(x) = p(2s)$.
   Let $\varphi_{p(2s)}$ follow $\varphi_{p(0)}$ from now on.
   Let $\sigma_{s+1} = \sigma_s \diamond (x_s, p(2s)) \diamond \ldots \diamond (y', p(2s))$.
   Go to stage $s + 1$.
   (Note that if $p(i)$ is in the range of $\varphi_{p(0)}$, then $\varphi_{p(i)} =^* \varphi_{p(0)}$).

End Stage s

13

Now consider the following cases:

Case 1: All stages halt.

In this case let $f = \varphi_{p(0)}$. Clearly, $f \in \mathcal{C}$.

Case 1a: $\mathbf{M}$ does not converge on $G = \bigcup_{s \in \mathbb{N}} \sigma_s$, which is an information sequence for f.

In this case $\mathbf{M}$ does not $\mathbf{Im}^*\mathbf{Ex}^*$ identify f.

Case 1b: $\mathbf{M}$ on $G = \bigcup_{s \in \mathbb{N}} \sigma_s$ converges.

In this case the only way infinitely many stages can exist is by execution of step 5 infinitely often. But then $\varphi_{\mathbf{M}(G)}$ is infinitely different from f.

Case 2: Stage s starts but never halts.

In this case let $f = \varphi_{p(2s)}$. Clearly, $f \in \mathcal{C}$. Let $G = \sigma_s \diamond (x_s, p(2s)) \diamond (x_s + 1, p(2s)) \ldots$.
Thus G is a $*$-imperfect information sequence for f. But $\mathbf{M}$ on G converges to $\mathbf{M}(\sigma_s)$, and, for all but finitely many x, $\varphi_{\mathbf{M}(\sigma_s)}(x) \neq p(2s)$. Thus $\mathbf{M}$ does not $\mathbf{Im}^*\mathbf{Ex}^*$-identify f.

From the above cases we have that $\mathbf{M}$ does not $\mathbf{Im}^*\mathbf{Ex}^*$ identify $\mathcal{C}$. ■

We do not know if the above theorem can be improved. However we would like to bring to the reader's attention the following theorem from [10] (which limits how much the above theorem can be improved).

Theorem 34 [10] $(\forall i, j \in \mathbb{N})[\mathbf{In}^{4i}\mathbf{Ex}^j \subseteq \mathbf{Im}^{2i}\mathbf{Ex}^{2j}]$.

Now we consider the possibility of whether noisy data can hurt more than incomplete data. The following theorem almost answers the question negatively. It shows that if a class of functions can be identified from incomplete information sequences then it can also be identified from noisy information sequences as long as the density of inaccuracies is slightly reduced.

Theorem 35 *Suppose* $0 \leq d_1 < d_2 \leq 1$. $(\forall a \in \mathbb{N} \cup \{*\})[\mathbf{DIn}^{d_2}\mathbf{Ex}^a \subseteq \mathbf{DN}^{d_1}\mathbf{Ex}^a]$.

PROOF. This proof is a complex modification of the proof of Theorem 14 in [8]. In this case we cannot just try to remove the multiple valued points from the input information sequence (as done in the proof of Theorem 14 in [8]), since the number of such points may be infinite. However we know that if we wait long enough, all initial segments can be made noise free. Moreover the density of noise beyond a certain point is always smaller than d_2. We use these facts to simulate an IIM $\mathbf{M}$, which $\mathbf{DIn}^{d_2}\mathbf{Ex}^a$ identifies $\mathcal{C}$.

Let G be a d_1-Dnoisy information sequence for $f \in \mathcal{C}$. Without loss of generality we can assume that, for all n, content$(G[2n]) \subseteq \{(x, y) \mid x \leq n\}$ and $(\forall x \leq n)(\exists y)[(x, y) \in$ content$(G[2n])]$ (otherwise such a G can be effectively constructed from the input information sequence). Let $S = \{(x, z) \mid \text{card}(\{(x, y) \in \text{content}(G) \mid y \in \mathbb{N}\}) = 1 \wedge (x, z) \in \text{content}(G)\}$. Let G' be the subsequence of G formed by deleting from G all elements not in S. Now if both (x_1, y_1) and (x_2, y_2) are in content(G'), where $x_1 < x_2$, then (x_1, y_1) appears before (x_2, y_2) in G'. Let σ_n denote the smallest initial segment of G' such that, for all $x \leq n$, $(x, y) \in \text{content}(G') \Rightarrow (x, y) \in \text{content}(\sigma_n)$.

Let Good(n) be true iff, for all $n' > n$, there exists a subsequence τ (extending σ_n) of $\sigma_{n'}$ such that,

(A) card$(\{x \mid (\exists y)[(x, y) \in \text{content}(\sigma_n)]\})/(n + 1) > (1 - (d_1 + d_2)/2)$

(B) For $n \leq n'' \leq n'$, card$(\{x \leq n'' \mid (\exists y)[(x, y) \in \text{content}(\tau)]\})/(n'' + 1) > (1 - (d_1 + d_2)/2)$
and

14

(C) $\mathbf{M}(\sigma_n) = \mathbf{M}(\tau')$ for $\sigma_n \subseteq \tau' \subseteq \tau$.

Clearly, there exists an n such that $\mathrm{Good}(n)$. Also given σ_n, one can determine if $\mathrm{Good}(n)$ holds in the limit (from G). To see this, first note that, if for some $\sigma'_{n'}$, a supersequence of $\sigma_{n'}$ the following are satisfied:

(i) $\sigma'_{n'}$ extends σ_n, and
(ii) no subsequence τ, extending σ_n, of $\sigma'_{n'}$ satisfies (A), (B), and (C),

then $\neg\mathrm{Good}(n)$. Thus, given σ_n, if $\neg\mathrm{Good}(n)$ then one can find a witness for this using G and $\mathbf{M}$. It follows that one can determine in the limit if $\mathrm{Good}(n)$ holds.

Let n_0 be the least n such that $\mathrm{Good}(n)$. Based on the discussion above, an IIM $\mathbf{M}'$ can determine n_0, and thus $\mathbf{M}(\sigma_{n_0})$, in the limit. We now claim that $\mathbf{M}(\sigma_{n_0})$ is a program for an a-variant of f. To see this consider the tree formed by considering all subsequences τ of G' (with a corresponding n') extending σ_{n_0} such that (B) and (C) are satisfied. Clearly this tree is infinite. Moreover the branching factor, in this tree, at any particular node is finite (due to density constraint in (B)). Thus there exists an infinite branch in this tree. Let G'' be the information sequence formed using this infinite branch. Clearly G'' is d_2-Dincomplete for f. Also $\mathbf{M}(G'') = \mathbf{M}(\sigma_{n_0})$. Thus $\mathbf{M}(\sigma_{n_0})$ is a program for an a-variant of f. ∎

Similarly we also have

Theorem 36 *Suppose* $0 \le d_1 < d_2 \le 1$. $(\forall a \in \mathbb{N} \cup \{*\})[\mathbf{UDIn}^{d_2}\mathbf{Ex}^a \subseteq \mathbf{UDN}^{d_1}\mathbf{Ex}^a]$.

We leave it as an open question whether, for $a \in \mathbb{N} \cup \{*\}$ and $d \in [0,1)$, $[\mathbf{DIn}^d\mathbf{Ex}^a \subseteq \mathbf{DN}^d\mathbf{Ex}^a]$ and $[\mathbf{UDIn}^d\mathbf{Ex}^a \subseteq \mathbf{UDN}^d\mathbf{Ex}^a]$. We also leave it open whether, $[\mathbf{DIn}^d\mathbf{Bc}^a \subseteq \mathbf{DN}^d\mathbf{Bc}^a]$ and $[\mathbf{UDIn}^d\mathbf{Bc}^a \subseteq \mathbf{UDN}^d\mathbf{Bc}^a]$.

6 Conclusions

In this paper we considered the effects of infinite number of inaccuracies in the input data on the learning power of IIMs. For $d < 1/2$ it is open whether $\mathbf{DIm}^d\mathbf{Ex} - \mathbf{DN}^{2d}\mathbf{Ex}$ (or $\mathbf{DIn}^{2d}\mathbf{Ex}$) is empty or not. It is also open whether, for $a \in \mathbb{N} \cup \{*\}$ and $d \in [0,1)$, $[\mathbf{DIn}^d\mathbf{Ex}^a \subseteq \mathbf{DN}^d\mathbf{Ex}^a]$ or $[\mathbf{UDIn}^d\mathbf{Ex}^a \subseteq \mathbf{UDN}^d\mathbf{Ex}^a]$.

7 Acknowledgements

References

[1] D. Angluin and C. Smith. A survey of inductive inference: Theory and methods. *Computing Surveys*, 15:237–289, 1983.

[2] J. M. Barzdin. Two theorems on the limiting synthesis of functions. *In Theory of Algorithms and Programs, Latvian State University, Riga*, 210:82–88, 1974. In Russian.

[3] L. Blum and M. Blum. Toward a mathematical theory of inductive inference. *Information and Control*, 28:125–155, 1975.

[4] M. Blum. A machine-independent theory of the complexity of recursive functions. *Journal of the ACM*, 14:322–336, 1967.

[5] J. Case. Periodicity in generations of automata. *Mathematical Systems Theory*, 8:15–32, 1974.

[6] J. Case and C. Smith. Comparison of identification criteria for machine inductive inference. *Theoretical Computer Science*, 25:193–220, 1983.

[7] M. A. Fulk and S. Jain. Learning in the presence of inaccurate information. In R. Rivest, D. Haussler, and M. K. Warmuth, editors, *Proceedings of the Second Annual Workshop on Computational Learning Theory, Santa Cruz, California*, pages 175–188. Morgan Kaufmann Publishers, Inc., August 1989.

[8] M. A. Fulk and S. Jain. Learning in the presence of inaccurate information. *Theoretical Computer Science A*, 161(1–2):235–261, 15 July 1996.

[9] E. M. Gold. Language identification in the limit. *Information and Control*, 10:447–474, 1967.

[10] S. Jain. *Learning in the Presence of Additional Information and Inaccurate Information*. PhD thesis, University of Rochester, 1990.

[11] S. Jain. Program synthesis in the presence of infinite number of inaccuracies. In S.Arikawa and K. P. Jantke, editors, *Proceedings of the Fifth International Workshop on Algorithmic Learning Theory, Reinhardsbrunn Castle, Germany*, pages 333–348, October 1994.

[12] R. Klette and R. Wiehagen. Research in the theory of inductive inference by GDR mathematicians – A survey. *Information Sciences*, 22:149–169, 1980.

[13] M. Machtey and P. Young. *An Introduction to the General Theory of Algorithms*. North Holland, New York, 1978.

[14] D. Osherson, M. Stob, and S. Weinstein. *Systems that Learn, An Introduction to Learning Theory for Cognitive and Computer Scientists*. MIT Press, Cambridge, Mass., 1986.

[15] H. Rogers. Gödel numberings of partial recursive functions. *Journal of Symbolic Logic*, 23:331–341, 1958.

[16] H. Rogers. *Theory of Recursive Functions and Effective Computability*. McGraw-Hill, New York, 1967. Reprinted by MIT Press, Cambridge, Massachusetts in 1987.

[17] J. Royer. Inductive inference of approximations. *Information and Control*, 70:156–178, 1986.

[18] G. Schäfer-Richter. Some results in the theory of effective program synthesis - learning by defective information. *Lecture Notes in Computer Science*, 215:219–225, 1986.

[19] C. Smith and M. Velauthapillai. On the inference of approximate programs. *Theoretical Computer Science*, 77:249–266, 1990.