Journal of Universal Computer Science, vol. 7, no. 9 (2001), 794-815
submitted: 16/7/01, accepted: 6/2/01, appeared: 28/9/01 © Springer Pub. Co.

Efficient Measure Learning

Sandra Fontani
(University of Siena, ITALY
fontanis@unisi.it)

Abstract: We study the problem of efficient identification of particular classes of p-time languages, called uniform. We require the learner to identify each language of such a class by constantly guessing, after a small number of examples, the same index for it. We present three identification paradigms based on different kind of examples: identification on informant (positive and negative information), measure identification (positive information in a probabilistic setting), identification with probability (positive and negative information in a probabilistic setting). In each case we introduce two efficient identification paradigms, called *efficient* and *very efficient identification* respectively. We characterize efficient identification on informant and with probability and, as a corollary, we show that the two identification paradigms are equivalent. A necessary condition is shown for very efficient identification on informant, which becomes sufficient if and only if $\mathcal{P} = \mathcal{NP}$. The same condition is sufficient for very efficient identification with probability if and only if $\mathcal{NP} = \mathcal{RP}$. We show that (very) efficient identification on informant and with probability are strictly stronger than (very) efficient measure identification.
Key Words: Learning Theory
Category: I.2.6

1 Introduction

In this paper we are concerned with the problem of identifying a class of languages in efficient time. The main idea is very similar to that adopted in [Fontani 00] for efficient identification of classes of functions. Efficiency will be synonymous with polynomial time, and so we will only consider the class of polynomial-time (p-time) languages, $P_{\mathcal{L}}$. Informally, given a class of p-time languages together with a class of representations (*indexes*) for them, we will require the learner to be a p-time function and to become successful in a number of guesses polynomially bounded in the (*length* of the) least index of the unknown language in the chosen representation class.

Indexes for p-time languages will be given by an *acceptable indexing* for P, the class of p-time functions [Fontani 00]. In fact, since a language is p-time if and only if its characteristic function is p-time, each p-time language will be assigned to the same indexes its characteristic function has with respect to any such indexing.

If we want a "polynomial time" identification, it is reasonable that the learner could check in polynomial time the consistency of an index h for a p-time language in a given class with a sample S. In the case of functions, this was possible for a uniform class (i.e. a class for which there exists a p-time algorithm -*universal function*- which uniformly computes all its functions). The same property continues to hold for a class of languages with an associated uniform class of characteristic functions. So we will consider only such classes of languages, ones that we will call *uniform*. The fixed universal function will attribute to each language of a uniform class, indexes with respect to any particular acceptable indexing for P, so it will completely specify the indexes for these languages.

We will define two efficient identification paradigms for uniform classes of languages. Namely, for every uniform class $\mathcal{W}$ with representation class $\mathcal{R}$, we say that $\mathcal{W}$ is *efficiently identifiable* if the learner becomes successful in a number of steps polynomially bounded by the least index of the unknown language in $\mathcal{R}$. We will say that $\mathcal{W}$ is *very efficiently identifiable*, if the learner becomes successful in a number of steps polynomially bounded by the *length* of the least index of the unknown language in $\mathcal{R}$. Obviously very efficient identification will imply efficient identification.

However a question comes immediately to mind: can different kinds of information affect the "efficiency" of identification (in the previous sense) and what kinds of information are more reasonable to choose for our purposes? The classical approach considers identification of languages on positive data (elements of the unknown language) or on positive and negative data (values assumed by the characteristic function of the unknown language). The second seems very adaptable to the efficient case: it is entirely analogous to the efficient identification of a function. But the first approach is very difficult to treat in an efficient setting. The reason is that, if in the fixed class there are two languages L and L' having a common element i which is given consecutively infinitely many times, then the learner cannot distinguish in a short time L from L'. A way to overcome this problem is to consider elements of the unknown language randomly generated by a probability distribution which satisfies certain properties avoiding the previous situation. To conclude, we will consider "efficient identification" of uniform classes of languages in the following environments for the unknown language L:

(1) The learner receives at each time n a piece of positive information ("0") if $n \in L$ or a piece of negative information ("1") about L if $n \notin L$. The infinite string so generated is called an *informant for* L.

(2) The learner receives at each time n a piece of positive information about L (an element of L) randomly generated by a "probability distribution" on L. The infinite string so generated will be called a *text for* L and we will speak about *measure identification.*

(3) The learner receives at each time n a couple of values in $\{0, 1\}$, the first of which is the n-th element of the informant of L and the second is generated according to a random coin drawn. In this case, we will speak about *identification with probability.*

In Section 4 we introduce the notions of (*very*) *efficient identification on informant* of uniform classes. We characterize efficient identification. We find a necessary condition for very efficient identification whose sufficiency turns out to be equivalent to $\mathcal{P} = \mathcal{NP}$.

In Section 5 we define (*very*) *efficient measure identification* of uniform classes. In both efficient settings, we prove that identification on informant implies measure identification, but the reverse is not true.

In Section 6 we study (*very*) *efficient identification with probability* of uniform classes. We show that it is equivalent to identification on informant in the efficient case. We find a necessary condition for very efficient identification, which becomes sufficient if and only if $\mathcal{NP} = \mathcal{RP}$.

2 Preliminaries

ω, $\emptyset$ denote, respectively, the set of natural numbers and the emptyset. $\omega^{<\omega}$ and $2^{<\omega}$ denote the set of strings of natural numbers and the set of binary strings respectively. For every $a_0, ..., a_n \in \{0,1\}$, $a_0 a_1 ... a_n$ denotes the binary string whose elements are (in the given order) $a_0, a_1, ..., a_n$. Analogously we interpret $(\tau_0, ..., \tau_n)$, if $\tau_0, ..., \tau_n \in \omega$ or $\tau_0, ..., \tau_n \in 2^{<\omega}$. We use $f, g, h, \varphi, ..., \Psi, \Phi, \Gamma, \Delta, ...$ for recursive functions. We omit the arity of a function when it is clear from the context. We write $\lambda x_i.f(x_0, ..., x_n)$ $(i \leq n)$ to mean that f depends on variable x_i only.

We use $p, q, p', q', \overline{p}, \overline{q}...$ to indicate polynomials with positive integer coefficients. In the case of polynomials in one variable we sometimes omit the argument.

We write $min\, x[\cdots x \cdots]$ or $\mu x[\cdots x \cdots]$ ($max\, x[\cdots x \cdots]$) to denote the least (the greatest) natural number for which the expression $[\cdots x \cdots]$ is true when "x" assumes this value. If S is a set, $minS$ ($maxS$) denotes the least (the greatest) element of S, while $card(S)$ denotes the cardinality of S.

2.1 Coding sequences

For every $a \in \omega$, we code a by its binary expansion, so, if $a = \sum_{i=0}^{n} \alpha_i 2^i$, for some $n \in \omega$, $\alpha_i \in \{0,1\}$, where either $a = n = \alpha_0 = 0$ or $\alpha_n \neq 0$, we consider $a = \alpha_n \alpha_{n-1} ... \alpha_0$.

The length of a, $|a|$, is the number of bits in its binary expansion. So $|a| = \lceil log_2(a+1) \rceil$ (the least integer $\geq log_2(a+1)$), and we approximate $|a|$ by $log_2(a)$ (note that $|0| = 0$)).

The code for the numerical sequence $\overline{a} = (a_0, ..., a_n)$ is constructed by the following procedure. We write the a_i's in binary notation, obtaining a string of 0, 1 and commas. We write such a string in reverse order. We replace each 0 by "10", each 1 by "11" and each comma by "00". The resulting string is the binary representation of the code of $\overline{a}$, which we denote by $< a_0, ..., a_n >$. For example, the code of $(3, 2, 4)$ is the number whose binary expansion is 101011001011001111. The code of (a) is the binary expansion of a and the code of the empty sequence is 0.

Notice that, for every $(a_0, ..., a_n)$, $| < a_0, ..., a_n > | = 2(|a_0| + ... + |a_n| + n)$. Moreover there is a uniform effective method for checking if a number is the code of a finite sequence, which also works in time polynomial in the length of the input sequence. The set of codes of finite numerical sequences and the set of codes of finite sequences of binary strings are denoted by Seq and $Bseq$ respectively.

We use the symbol $*$ to indicate the concatenation of codes of finite numerical sequences, i.e. $< a_0, ..., a_n > * < b_0, ..., b_m > = < a_0, ..., a_n, b_0, ..., b_m >$. Moreover we write $< a_0, ..., a_n > \subseteq < b_0, ..., b_m >$ if $n \leq m$, and, for every $i \leq n$, $a_i = b_i$. We denote by $^{\#}\!< a_0, ..., a_n >$ the sequence $(a_0, ..., a_n)$.

If Σ is a finite alphabet of symbols and Σ^* is the set of finite sequences of symbols in Σ, we can codify each $\sigma \in \Sigma^*$ in a similar manner. We associate to each $a \in \Sigma$ a number (different numbers for different symbols). The code of each $\sigma \in \Sigma^*$ is the code of the numerical sequence associated to it. Consider for example $\Sigma = \{x, |, \wedge, \vee, \_, (,)\}$. Each propositional formula in conjunctive normal form ($A \in CNF$) can be represented as a sequence of symbols of Σ. Associate to each symbol in Σ the following numbers:

x	$\vert$	$\wedge$	$\vee$	$\_$	(	)
1	2	3	4	5	6	7

Hence, if $A \equiv x_1 \vee (\overline{x}_2 \wedge \overline{x}_1)$, we write it as $A \equiv x| \vee (\_x|| \wedge \_x|)$ and the code of A is the code of $(1,2,4,6,5,1,2,2,3,5,1,2,7)$. We denote by $lth(A)$ the length of A in Σ and by $\lceil A \rceil$ the code of A. Note that, if $A \in CNF$, then $|\lceil A \rceil| \leq 8n-2$.

2.2 The class P

Throughout the paper we denote by P the class of p-time functions (for a standard definition see [Buss 86]). Without loss of generality, we consider as a model of computation deterministic Turing machines (T.m.). We adopt the convention that all functions have domain ω^k and codomain ω. If M is a T.m. and $t:\omega^{n+1} \to \omega$ is any function, then, for all $a_0,...,a_n \in \omega$, we write $M(a_0,...,a_n)\downarrow \leq t(a_0,...,a_n)$ if M on input $(a_0,...,a_n)$ converges within $t(a_0,...,a_n)$ steps of computation. In particular, M is deterministic polynomial time if and only if $M(a_0,...,a_n)\downarrow \leq t(a_0,...,a_n)$ where $t(a_0,...,a_n) = p(|a_0|,...,|a_n|)$ for some polynomial p. Sometimes, given a function $f:\omega^{n+1} \to \omega$ for which we have fixed a T.m. M that computes it, we write, by abuse of language, $f(a_0,...,a_n)\downarrow \leq t(a_0,...,a_n)$ instead of $M(a_0,...,a_n)\downarrow \leq t(a_0,...,a_n)$. We recall the following p-time functions:

(1) $\lfloor \frac{x}{2} \rfloor$: the integer part of $\frac{x}{2}$ (the greatest integer $\leq \frac{x}{2}$).

(2) $s(x) = x+1$: the successor function.

(3) $[x]_i$, $i \leq |x|-1$: the i-th bit of the binary representation of x ($[x]_i$ is arbitrarily defined if $i \geq |x|$).

(4)
$$\beta(i, <a_0,...,a_n>) = \begin{cases} n+1 & \text{if } i=0 \\ a_{i-1} & \text{if } 0 < i \leq n+1 \end{cases}$$

β is arbitrarily defined if $i > n+1$ or if the second argument is not the code of any sequence [Buss 86]. For simplicity, we write $n+1$ instead of $\beta(0, <a_0,...,a_n>)$ and, if $x \leq n$, we write a_x instead of $\beta(s(x), <a_0,...,a_n>)$. Moreover, for every $n \in \omega$, we let $(n)_x = \beta(s(x), n)$ and $lth(n) = \beta(0, n)$. Note that, in particular, if $\sigma \in Seq$ ($\sigma \in Bseq$) and $\sigma = <a_0,...,a_n>$, then $(\sigma)_x = (<a_0,...,a_n>)_x = a_x$, $lth(\sigma) = lth(<a_0,...,a_n>) = n+1$.

2.3 Identification paradigm

Consider some acceptable indexing $\pi_0,...,\pi_n,...$ for the class of partial recursive functions [Shoenfield 58]. We recall a very natural identification paradigm for classes of total recursive functions (see [Odifreddi 99]).

Definition 1 Let $\mathcal{C}$ be a class of total recursive functions. We say that $\mathcal{C}$ is EX-*identifiable* if there exists a total recursive function g such that, for every $f \in \mathcal{C}$,

$$(\exists n_0)(\forall n \geq n_0) g(<f(0),...,f(n)>) = i$$

for some $i \in \omega$ such that $f = \pi_i$.

3 Efficient identification: basic concepts

In [Fontani 00] we introduced a particular indexing for the class P which can be adopted as an indexing for the class of all p-time languages, $P_{\mathcal{L}}$, as well. We recall the definition of such an indexing.

Definition 2 An *acceptable indexing for P* is an enumeration $\varphi_0, ..., \varphi_n$ of (all of) the p-time functions that meets the following conditions:

(i) There exist a function $\Phi(i,x)$ and a polynomial $p(i,x)$ such that, for every $i,x \in \omega$:

- $\Phi(i,x) \downarrow \leq p(i,x)$ $(\lambda ix.\Phi(|i|,|x|) \in P)$
- $\Phi(i,x) = \varphi_i(x)$.

(ii) For every $\Psi(i,x) \in P$ there exists $h \in P$ strictly increasing such that, for every $i,x \in \omega$, $\Psi(i,x) = \Phi(h(i),x) = \varphi_{h(i)}(x)$.

Such a function $\Phi(i,x)$ will be called *a universal function for P.*

Let L be a language and let f be its characteristic function. Since $L \in P_{\mathcal{L}}$ if and only if $f \in P$, it is reasonable to associate to L all (and only) those indexes that f has with respect to some acceptable indexing.

Definition 3 Let $\Phi(i,x)$ be an acceptable indexing for P. Let $L \in P_{\mathcal{L}}$ and let f be its characteristic function. If $f = \lambda x.\Phi(i,x)$, then we say that *i is an index for L* and we write $L = W_i$.

In the case of efficient identification of p-time functions we restricted our attention to the so called uniform classes of functions, i.e. classes whose functions are computable in a uniform and "efficient" way. This definition can be extended in a very natural way to classes of p-time languages too.

Definition 4 Let $\mathcal{C} \subseteq P$. We say that $\mathcal{C}$ is *uniform* if there exists $\Psi(i,x)$ such that:

(i) $\lambda ix.\Psi(i,x) \in P$.

(ii) For every $i \in \omega$, $\lambda x.\Psi(i,x) \in \mathcal{C}$.

(iii) For every $f \in \mathcal{C}$, there exists $i \in \omega$ such that $f = \lambda x.\Psi(i,x)$.

Such a function $\Psi(i,x)$ is called *a universal function for $\mathcal{C}$.*

Definition 5 Let $\mathcal{W} \subseteq P_{\mathcal{L}}$. We say that $\mathcal{W}$ is *uniform* if and only if the class $\mathcal{C}$ of the characteristic functions of languages in $\mathcal{W}$ is uniform. If $\Psi(i,x)$ is a universal function for $\mathcal{C}$, we say that $\Psi(i,x)$ is a *universal function for $\mathcal{W}$.*

Notation 1 (1) By Definition 2 (ii), if $\mathcal{W}$ is a uniform class with universal function $\Psi(i,x)$, then there exists $h \in P$ strictly increasing such that, for every $i,x \in \omega$, $\Psi(i,x) = \varphi_{h(i)}(x)$. We call h a *Ψ-indexing* for $\mathcal{W}$ and we write $\mathcal{W} = \mathcal{W}_h = \{W_{h(i)} : i \in \omega\}$.
In this case, the class $\mathcal{C}$ of the characteristic functions of languages in $\mathcal{W}$ is denoted by $\mathcal{L}_h = \{\varphi_{h(i)} : i \in \omega\}$ ($\varphi_{h(i)}$ is the characteristic function of $W_{h(i)}$).
(2) Let $\mathcal{W}_h = \{W_{h(i)} : i \in \omega\}$ be a uniform class. For every $i \in \omega$, we define:

$$m(i) = min\{j \in \omega : W_{h(j)} = W_{h(i)}\}.$$

In other words, for every $W_{h(i)} \in \mathcal{W}_h$, $W_{h(i)} = W_{h(m(i))}$ and the first occurrence of $W_{h(i)}$ in the enumeration of the class induced by h is at step $m(i)$ (we refer to $m(i)$ as to the "least index" of $W_{h(i)}$ in $\mathcal{W}_h$).

We are now ready to define "efficient identification" of uniform classes of languages. In each of the next three sections we will present different efficient identification paradigms according to the type of data the learner receives. All these definitions will however share the same idea for "efficient identification". In fact the learner will have to make, at each step, a guess in time polynomial in the length of the input sequence and to stabilize on a correct index within a number of guesses polynomially bounded in the (length of the) least index of the unknown language in the given class.

4 Identification on informant

We consider the case in which the learner receives both positive and negative information about the unknown language. More precisely, if $\mathcal{W}$ is a uniform class and $L \in \mathcal{W}$, at each step n the learner is given the value "0" if $n \in L$, "1" if $n \notin L$.

Definition 6 An *informant* I is an infinite binary sequence ($I \in 2^\omega$). If L is a language, we say that $I \in 2^\omega$ is an *informant for* L if $\{n \in \omega : I_n = 0\} = L$, where I_n is the n-th bit of I.

Obviously, for every language L, there exists a unique informant I for L.

Notation 2 Let $\mathcal{W}_h = \{W_{h(i)} : i \in \omega\}$ be a uniform class. For every $i \in \omega$, if I is the informant for $W_{h(i)}$, we write $I = I^{h(i)}$.

Let $\mathcal{W}_h = \{W_{h(i)} : i \in \omega\}$ be a uniform class. In the previous section we noted that $\mathcal{L}_h = \{\varphi_{h(i)} : i \in \omega\}$ is the class of the characteristic functions of languages in $\mathcal{W}_h$ and that, for every $i \in \omega$, $\varphi_{h(i)}$ is the characteristic function of $W_{h(i)}$. Moreover the informant for $W_{h(i)}$ is just the infinite sequence of values assumed by its characteristic function. This suggests that identification on informant of $\mathcal{W}_h$ is equivalent to the identification of $\mathcal{L}_h$. Such an observation motivates the following definition.

Definition 7 Let $\mathcal{W}_h = \{W_{h(i)} : i \in \omega\}$ be a uniform class. We say that:

(i) $\mathcal{W}_h$ is *EX-efficiently identifiable on informant* ($\mathcal{W}_h \in EX^{eff}_{inf}$) if there exist $g \in P$ and a polynomial p such that, for every $W_{h(i)} \in \mathcal{L}_h$, g EX-identifies $W_{h(i)}$ on $I^{h(i)}$ in at most $p(m(i))$ guesses, i.e.:

$$(\exists n_0 < p(m(i)))(\forall n \geq n_0)(g(< \varphi_{h(i)}(0), ..., \varphi_{h(i)}(n) >) = i')$$

for some $i' \in \omega$ such that $W_{h(i)} = W_{h(i')}$.

(ii) $\mathcal{W}_h$ is *EX-very efficiently identifiable on informant* ($\mathcal{W}_h \in EX^{v\text{-}eff}_{inf}$) if there exist $g \in P$ and a polynomial p such that, for every $W_{h(i)} \in \mathcal{L}_h$, g EX-identifies $W_{h(i)}$ on $I^{h(i)}$ in at most $p(|m(i)|)$ guesses, i.e.:

$$(\exists n_0 < p(|m(i)|))(\forall n \geq n_0)(g(< \varphi_{h(i)}(0), ..., \varphi_{h(i)}(n) >) = i')$$

for some $i' \in \omega$ such that $W_{h(i)} = W_{h(i')}$.

Obviously, for every uniform class $\mathcal{W}_h$, if $\mathcal{W}_h \in EX^{v\text{-}eff}_{inf}$ then $\mathcal{W}_h \in EX^{eff}_{inf}$.

Examples 1 (1) Let $Cof = \{L : L \text{ is a cofinite language}\}$. Let $\Psi(n, x)$ be such that, for every $n, x \in \omega$:

$$\Psi(n,x) = \begin{cases} [n]_{x+1} & \text{if } n \text{ is odd, } n \neq 1 \text{ and } x < |n| - 1 \\ 0 & \text{otherwise} \end{cases}$$

$\Psi(n, x)$ is a universal function for Cof such that, if h is a Ψ-indexing, for n odd ($n \neq 1$), $W_{h(n)} = \{x \in \omega : (x < |n| - 1 \wedge [n]_{x+1} = 0) \vee (x \geq |n| - 1)\}$, for n even or $n = 1$, $W_{h(n)} = \omega$. On the other hand, if $L \in Cof$ $(L \neq \omega)$, $\sigma = max\{j \in \omega : j \notin L\}$ and $a_0 a_1 ... a_\sigma$ is the initial segment of length $\sigma + 1$ of the informant for L, then $L = W_{h(n)}$ where $n = 1a_0...a_\sigma$ (if $L = \omega$, then $L = W_{h(n)}$ for every even n or $n = 1$). So $Cof = \mathcal{W}_h = \{W_{h(n)} : n \in \omega\}$. Let g be such that, for every $a_0, ..., a_n \in \omega$:

$$g(< a_0, ..., a_n >) = \begin{cases} 1a_0...a_i & \text{if } i = max\{j \leq n : a_j = 1\}, \\ & \quad \text{if such } i \text{ exists} \\ 0 & \text{otherwise} \end{cases}$$

It is easy to verify that $g \in P$. Let $W_{h(n)} \in \mathcal{W}_h$. For n even or $n = 1$, g EX-identifies on informant $W_{h(n)}$ after the first guess. For n odd ($n \neq 1$), g EX-identifies on informant $W_{h(n)}$ in at most $|n| - 1$ guesses. So $\mathcal{W}_h \in EX_{inf}^{v\text{-}eff}$.
(2) Let $Fin = \{L : L \text{ is a finite language}\}$. If $\Psi(n, x)$ is a universal function for Cof, let, for every $n, x \in \omega$, $\overline{\Psi}(n, x) = 1 - \Psi(n, x)$. Obviously $\overline{\Psi}(n, x)$ is a universal function for Fin. Moreover, if $\overline{h}$ is a $\overline{\Psi}$-indexing, $Fin = \mathcal{W}_{\overline{h}}$ and it is readily seen that $\mathcal{W}_{\overline{h}} \in EX_{inf}^{v\text{-}eff}$.
(3) For every finite uniform class $\mathcal{W}_h$, $\mathcal{W}_h \in EX_{inf}^{v\text{-}eff}$, since every finite uniform class of functions is EX-very efficiently identifiable [Fontani 00].

4.1 Main results for EX_{inf}^{eff} and $EX_{inf}^{v\text{-}eff}$

Efficient identification on informant of uniform classes of languages is a particular case of efficient identification of uniform classes of functions (functions assuming 0–1 values). Hence, in both efficient paradigms, all results obtained for identification of uniform classes of functions remain true for identification on informant of uniform classes of languages. If $\Psi(i, x)$ is a universal function for $\mathcal{W}_h$ ($\mathcal{L}_h$), consider the following:

$$(\exists \text{ polynomial } p)(\forall i)(\forall j < m(i))(\exists x \leq p(m(i)))(\varphi_{h(j)}(x) \neq \varphi_{h(i)}(x)) \qquad (*)$$

$$(\exists \text{ polynomial } p)(\forall i)(\forall j < m(i))(\exists x \leq p(|m(i)|))(\varphi_{h(j)}(x) \neq \varphi_{h(i)}(x)). \qquad (*')$$

In [Fontani 00] we proved that $(*)$ characterizes EX-efficient identification of uniform classes of functions, while $(*')$ is a necessary condition for EX-very efficient identification of uniform classes of functions, whose sufficiency is equivalent to $\mathcal{P} = \mathcal{NP}$. Moreover, we there proved that P is not EX-efficiently identifiable with respect to any acceptable indexing. Hence the following theorems are immediately estabilished.

Theorem 1 *Let $\mathcal{W}_h$ be a uniform class. Then:*

(i) $\mathcal{W}_h \in EX_{inf}^{eff} \Leftrightarrow \mathcal{W}_h$ *satisfies* $(*)$.

(ii) $\mathcal{W}_h \in EX_{inf}^{v\text{-}eff} \Rightarrow \mathcal{W}_h$ *satisfies* $(*')$.

Theorem 2 *$P_{\mathcal{L}}$ is not EX-efficiently identifiable on informant with respect to any acceptable indexing.*

Theorem 3 *The following are equivalent:*

(i) Every uniform class of languages satisfying $(*')$ is EX-very efficiently identifiable on informant.

(ii) $\mathcal{P} = \mathcal{NP}$.

Condition $(*)$ is very useful to show that a uniform class of languages is not EX-efficiently (hence not EX-very efficiently) identifiable on informant.

Example 1 Let $\mathcal{W} = \{L_i : i \in \omega\}$ such that $L_0 = \omega$ and, for every $i \in \omega \setminus \{0\}$, $L_i = \omega \setminus \{2^i\}$. Let:

$$\Psi(i, x) = \begin{cases} 1 & \text{if } i \neq 0 \text{ and } x = 2^i \\ 0 & \text{otherwise} \end{cases}$$

It is readily seen that $\Psi(i, x)$ is a universal function for $\mathcal{W}$ and, if h is a Ψ-indexing, then $\mathcal{W} = \mathcal{W}_h$ and $\mathcal{W}_h$ does not satisfy $(*)$. Hence $\mathcal{W}_h \notin EX_{inf}^{eff}$.

5 Efficient measure learning

As discussed in the introduction, the problem of efficient identification of uniform classes of languages on positive data becomes meaningful in a probabilistic setting. The main idea is the following: given a uniform class $\mathcal{W}$ and $L \in \mathcal{W}$, a learner g receives, at each step n, an element of L randomly generated by a probability distribution δ_L according to which each element of L is generated with positive probability, while other elements are generated with probability zero (δ_L is called a *distribution for* L). With high probability, after infinite extractions, all (and only all) of the elements of L will have been generated and g can learn from the infinite list of such elements (the *text for* L). We will say that g *(very) efficiently measure identifies* L if, for every $\varepsilon \in [0,1]$, the set of texts so generated on which g stabilizes on an index for L in a number of guesses bounded by a polynomial in the least index (in the length of the least index) of L in $\mathcal{W}$ and $\frac{1}{\varepsilon}$, has measure $> 1-\varepsilon$. Before making this notion precise, we briefly recall some concepts of Measure Theory (see [Oxtoby 71]).

5.1 Basic concepts

Definition 8 Let A be a set and let $\mathcal{S} \subseteq \mathcal{P}(A)$ be a subalgebra of the Boolean algebra $\mathcal{P}(A)$ of the subsets of A. We call a *probability measure* (in short: *probability*) *on* A a function $\mu : S \rightarrow [0,1]$ such that $\mu(A)=1$ and, for every $B, C \in S$ such that $B \cap C = \emptyset$, $\mu(B \cup C) = \mu(B) + \mu(C)$.

Definition 9 Let A be a set and let $\mathcal{S} \subseteq \mathcal{P}(A)$ be a subalgebra of the Boolean algebra $\mathcal{P}(A)$ closed under countable unions. We say that a measure μ on A is *σ-additive* if, for every family of mutually disjoint elements of S, $\{B_n : n \in \omega\}$, $\mu(\bigcup_{n\in\omega} B_n) = \sum_{n\in\omega} \mu(B_n)$.

Definition 10 We say that $\delta : \omega \rightarrow [0,1]$ is a *probability distribution* (in short: *a distribution*) *on* ω if $\sum_{n\in\omega} \delta(n) = 1$ and, for every $n \in \omega$, $\delta(n) > 0$.

Definition 11 Let L be a language and let δ be a distribution on ω. We say that δ is *for* L if, for every $n \in \omega$, $n \in L$ if and only if $\delta(n) > 0$.

Examples 2 (1) Let $L \neq \emptyset$ be a finite language and $K = card(L)$. Consider:

$$\delta^{1,L}(n) = \begin{cases} \frac{1}{K} & \text{if } n \in L \\ 0 & \text{otherwise} \end{cases}$$

$\delta^{1,L}$ is a distribution on L called *uniform*, since each element of L is generated with the same probability with respect to it.
(2) Let $L \neq \emptyset$ be a language and $K = \sum_{n\in L} \frac{1}{2^{2|n|+1}}$. Consider:

$$\delta^{2,L}(n) = \begin{cases} \frac{1}{2^{2|n|+1}} \cdot \frac{1}{K} & \text{if } n \in L \\ 0 & \text{otherwise} \end{cases}$$

$\delta^{2,L}$ is a distribution for L.

Definition 12 Let $\mathcal{W}_h = \{W_{h(i)} : i \in \omega\}$ be a uniform class and let $\mathcal{D}$ be a class of probability distributions on ω. We say that $\mathcal{D}$ is *for* $\mathcal{W}_h$ if the following conditions hold:

(i) For every $W_{h(i)} \in \mathcal{W}_h$ there exists $\delta \in \mathcal{D}$ such that δ is for $W_{h(i)}$.

(ii) For every $\delta \in \mathcal{D}$ there exists $W_{h(i)} \in \mathcal{W}_h$ such that δ is for $W_{h(i)}$.

Definition 13 A *text* t is an infinite sequence of natural numbers ($t \in \omega^\omega$). The set of numbers occurring in a text t is denoted by $rng(t)$. Moreover if $S \subseteq \omega$, we say that a text t is *for* S if and only if $rng(t) = S$.

We are almost ready to formalize the identification paradigms informally described at the beginning of this section. We have only to make precise how we will "measure" the sets of texts for a certain language L on which the learner "efficiently identifies" L. There is a very natural choice. It is in fact well-known that, given a language L and a distribution δ for L, letting for every $A \subseteq \omega$, $\Delta_\delta(A) = \sum_{n \in A} \delta(n)$, Δ_δ is a σ-additive measure on ω. Moreover Δ_δ induces a product measure μ_δ on ω^ω which is σ-additive and such that, if for every $\sigma \in \omega^{<\omega}$, we let $B_\sigma = \{t \in \omega^\omega : \sigma \subseteq t\}$, then:

- For every $\sigma \in \omega^{<\omega}$, if $\sigma = (a_0, ..., a_n)$, then $\mu_\delta(B_\sigma) = \prod_{i < lth(\sigma)} \delta(a_i)$.
- For every $\sigma \in \omega^{<\omega}$, $rng(\sigma) \subseteq L$ if and only if $\mu_\delta(B_\sigma) > 0$.
- $\mu_\delta(\{t \in \omega^\omega : rng(t) = L\}) = 1$.

Definition 14 Let $\mathcal{W}_h = \{W_{h(i)} : i \in \omega\}$ be a uniform class and let $g(x, y)$ be a total recursive function. For every $\varepsilon \in [0, 1]$, $W_{h(i)} \in \mathcal{W}_h$, $t \in \omega^\omega$ and $k \in \omega$, we say that g *EX–converges on* t, ε *to* $W_{h(i)}$ *in at most* k *steps* (in short: *in* $< k$ *steps*) if:

$$(\exists n_0 < k)(\forall n \geq n_0)(g(t_{|n}, \varepsilon) = i')$$

for some $i' \in \omega$ such that $W_{h(i)} = W_{h(i')}$.

Remark 1 In the previous definition we can suppose that $\varepsilon \in [0, 1]$ is of the form $\frac{1}{n}$, where n is a positive natural number. Under this assumption $g(t_{|n}, \varepsilon)$ really depends on natural numbers, $t_{|n}$ and n, according to the initial requirement that $g(x, y)$ is a total recursive function.

Definition 15 Let $\mathcal{W}_h = \{W_{h(i)} : i \in \omega\}$ be a uniform class and let $\mathcal{D}$ be a class of distributions for $\mathcal{W}_h$. We say that $\mathcal{W}_h$ is *EX-efficiently measure identifiable with respect to* $\mathcal{D}$ ($\mathcal{W}_h \in EX^{eff}_{meas}(\mathcal{D})$) if there exist a function $g(x, y) \in P$ and a polynomial $p(x, y)$ such that, for every $\varepsilon \in [0, 1]$, for every $W_{h(i)} \in \mathcal{W}_h$ and for every $\delta \in \mathcal{D}$ for $W_{h(i)}$, g EX-measure identifies $W_{h(i)}$ in at most $p(m(i), \frac{1}{\varepsilon})$ guesses with respect to δ, i.e.:

$$\mu_\delta(\{t \in \omega^\omega : g\ EX\text{−converges on } t, \varepsilon \text{ to } W_{h(i)} \text{ in} < p(m(i), \tfrac{1}{\varepsilon}) \text{ steps}\}) > 1 - \varepsilon.$$

Definition 16 Let $\mathcal{W}_h = \{W_{h(i)} : i \in \omega\}$ be a uniform class and let $\mathcal{D}$ be a class of distributions for $\mathcal{W}_h$. We say that $\mathcal{W}_h$ is *EX-very efficiently measure identifiable with respect to* $\mathcal{D}$ ($\mathcal{W}_h \in EX^{v\text{-}eff}_{meas}(\mathcal{D})$) if there exist a function $g(x, y) \in P$ and a polynomial $p(x, y)$ such that, for every $\varepsilon \in [0, 1]$, for every $W_{h(i)} \in \mathcal{W}_h$ and for every $\delta \in \mathcal{D}$ for $W_{h(i)}$, g EX-measure identifies $W_{h(i)}$ in at most $p(|m(i)|, \frac{1}{\varepsilon})$ guesses with respect to δ, i.e.:

$$\mu_\delta(\{t \in \omega^\omega : g\ EX\text{−converges on } t, \varepsilon \text{ to } W_{h(i)} \text{ in} < p(|m(i)|, \tfrac{1}{\varepsilon}) \text{ steps}\}) > 1 - \varepsilon.$$

Obviously, for every uniform class of languages $\mathcal{W}_h$ and for every distribution class $\mathcal{D}$ for $\mathcal{W}_h$, if $\mathcal{W}_h \in EX^{v\text{-}eff}_{meas}(\mathcal{D})$ then $\mathcal{W}_h \in EX^{eff}_{meas}(\mathcal{D})$.

Angluin showed that a class of languages is measure identifiable if and only if it is identifiable (on texts) [Angluin 88]. Uniform classes of languages efficiently measure identifiable with respect to every class of distributions are even rarer. It is in fact straightforward to verify that, for every uniform class $\mathcal{W}_h$, if $\mathcal{W}_h$ is EX-efficiently measure identifiable with respect to every class of distributions for it, then $\mathcal{W}_h$ contains only mutually disjoint languages. This is the reason why we will be mainly concerned with measure identification (in both efficient paradigms) of uniform classes with respect to classes of distributions satisfying particular properties.

Notation 3 Let L be a language and let δ be a distribution for L. For every $n \in L$ and $k \in \omega$, we denote by $Pr_\delta^{\leq k}(n)$ the probability that n occurs within k steps (with a number of extractions $\leq k$) with respect to δ. $Pr_\delta^{>k}(n)$ denotes the probability that n does not occur within k steps with respect to δ.

Example 2 Consider Fin, the class of finite languages without the empty language, with the following representation:

$$\Psi(i,x) = \begin{cases} 0 & \text{if } (i \neq 0,\ x < |i| \text{ and } [i]_{|i|-1-x} = 1) \text{ or} \\ & \quad (i = 0 \text{ and } x = 0) \\ 1 & \text{otherwise} \end{cases}$$

Obviously $\Psi(i,x)$ is a universal function for Fin which we refer to as "canonical". In fact, if h is a Ψ-indexing and $L \in Fin$, say $L = \{\sigma_1, ..., \sigma_r\}$, then, for every $i \in \omega \setminus \{0\}$, $W_{h(i)} = L$ if and only if $i = 2^{\sigma_1} + ... + 2^{\sigma_r}$. Moreover, for $i \neq 1$, $m(i) = i$, while for $i = 1$, $m(1) = 0$. Let $\mathcal{D}$ be a class of distributions for $\mathcal{W}_h$ for which there exists a polynomial $\overline{p}(x)$ such that, for every $W_{h(i)} \in \mathcal{W}_h$, for every $n \in W_{h(i)}$, for every $\delta \in \mathcal{D}$ for $W_{h(i)}$:

$$\delta(n) \geq \tfrac{1}{\overline{p}(\sigma_{W_{h(i)}})} \qquad \text{II}$$

where $\sigma_{W_{h(i)}} = max W_{h(i)}$. We prove that $\mathcal{W}_h \in EX_{meas}^{v\text{-}eff}(\mathcal{D})$.

Let $W_{h(i)} \in \mathcal{W}_h$, say $W_{h(i)} = \{\sigma_1, ..., \sigma_r\}$, and let $\sigma \in \mathcal{D}$ be a distribution for $W_{h(i)}$. It is readily seen that, for every $k \in \omega$, the probability that some $\sigma_j \in W_{h(i)}$ $(1 \leq j \leq r)$ does not occur within k steps is bounded by $\sum_{1 \leq j \leq r} Pr_\delta^{>k}(\sigma_j)$ and

$$\textstyle\sum_{1 \leq j \leq r} Pr_\delta^{>k}(\sigma_j) \leq r \left(1 - \frac{1}{\overline{p}(|i|)}\right)^k. \qquad (1)$$

Moreover, by condition $(1-x) < e^{-x}$, since $r \leq |i|$, then $r(1 - \frac{1}{\overline{p}(|i|)})^k \leq |i| \cdot e^{-\frac{k}{\overline{p}(|i|)}}$. But, for every $\varepsilon \in [0,1]$,

$$|i| \cdot e^{-\frac{k}{\overline{p}(|i|)}} \leq \varepsilon \;\Leftrightarrow\; k \geq \overline{p}(|i|)(ln|i| + ln\tfrac{1}{\varepsilon}). \qquad (2)$$

So, if $p(x,y) = \overline{p}(x)(x+y)$, for every $k \geq p(|i|, \frac{1}{\varepsilon})$, by (2) we have $|i| \cdot e^{-\frac{k}{\overline{p}(|i|)}} \leq \varepsilon$ and, by the previous considerations and (1), $\sum_{1 \leq j \leq r} Pr_\delta^{>k}(\sigma_j) \leq \varepsilon$. It follows that every σ_j $(1 \leq j \leq r)$ is generated within k steps (in particular within $p(|i|, \frac{1}{\varepsilon})$ steps) with respect to δ with probabilty $> 1 - \varepsilon$.

For every $\sigma \in Seq$, if $\sigma = < a_0, ...a_n >$, let $a = max\{a_j : j \leq n, a_j \leq n\}$. Let $s_\sigma = b_a ... b_0$ be such that, for every $x \leq a$, if $x = a_j$ for some $j \leq n$, then $b_x = 1$, otherwise $b_x = 0$. Note that, for every $\sigma \in Seq$, s_σ can be constructed in a number of steps polynomial in $lth(\sigma)$. For every $\varepsilon \in [0,1]$, for every $\sigma \in Seq$, let $g(\sigma, \varepsilon) = s_\sigma$. Obviously $g \in P$. Let $W_{h(i)} \in \mathcal{W}_h$, say $W_{h(i)} = \{\sigma_1, ..., \sigma_r\}$, and let t be a text for $W_{h(i)}$ randomly generated by some $\delta \in \mathcal{D}$ for $W_{h(i)}$. For every $\varepsilon \in [0,1]$, the probability according to which each σ_j $(1 \leq j \leq r)$ occurs within $p(|i|, \frac{1}{\varepsilon})$ steps with respect to δ is $> 1 - \varepsilon$, so:

$$\mu_\delta(\{t \in \omega^\omega : (\forall n)(n \in W_{h(i)} \rightarrow n \in rng(t_{|p(|i|, \frac{1}{\varepsilon})-1}))\}) > 1 - \varepsilon.$$

Moreover $\sigma_j \leq p(|i|, \frac{1}{\varepsilon})$. Hence, for every $k \geq p(|i|, \frac{1}{\varepsilon}) - 1$, $s_{t_{|k}} = i$ and $g(t_{|k}, \varepsilon) = i$. If $\mathcal{A} = \{t \in \omega^\omega : g\, EX-\text{converges on } t, \varepsilon \text{ to } W_{h(i)} \text{ in } < p(|i|, \frac{1}{\varepsilon}) \text{ steps}\}$, then

$$\mathcal{A} \supseteq \{t \in \omega^\omega : (\forall n)(n \in W_{h(i)} \rightarrow n \in rng(t_{|p(|i|, \frac{1}{\varepsilon})-1}))\},$$

so $\mu_\delta(\mathcal{A}) > 1 - \varepsilon$. It follows that g EX-measure identifies $W_{h(i)}$ in at most $p(|i|, \frac{1}{\varepsilon})$ guesses (with respect to δ) and $\mathcal{W}_h \in EX_{meas}^{v\text{-}eff}(\mathcal{D})$.

Remark 2 Consider $\mathcal{D}_1=\{\delta^{1,L} : L\in Fin\}$ and $\mathcal{D}_2=\{\delta^{2,L} : L\in Fin\}$ where, for every $L\in Fin$, $\delta^{1,L}$ and $\delta^{2,L}$ are the probability distributions defined in Examples 2. It is easy to verify that $\mathcal{D}_1$ and $\mathcal{D}_2$ satisfy (II). Hence, if $\Psi(i,x)$ is the canonical universal function for Fin and h is a Ψ-indexing ($Fin = \mathcal{W}_h$), then $\mathcal{W}_h \in EX^{v\text{-}eff}_{meas}(\mathcal{D}_1)$ and $\mathcal{W}_h \in EX^{v\text{-}eff}_{meas}(\mathcal{D}_2)$.

5.2 Main results for EX^{eff}_{meas} and $EX^{v\text{-}eff}_{meas}$

In the present section we give an answer to the following question: what relationships hold between identification on informant and measure identification of uniform classes of languages in both efficient paradigms? As stated before, such a problem becomes meaningful only if we refer to particular, even though quite natural, classes of distributions.

Definition 17 Let $\mathcal{W}_h = \{W_{h(i)} : i \in \omega\}$ be a uniform class and let $\mathcal{D}$ be a class of distributions for $\mathcal{W}_h$. We say that $\mathcal{D}$ is *polynomially bounded* if there exists a polynomial $\overline{p}(x)$ such that, for every $W_{h(i)} \in \mathcal{W}_h$, for every $n \in W_{h(i)}$, for every $\delta \in \mathcal{D}$ for $W_{h(i)}$, $\delta(n) \geq \frac{1}{\overline{p}(n)}$. In this case we say that $\mathcal{D}$ is *polynomially bounded by $\overline{p}$.*

Remark 3 Let $\mathcal{W}_h = \{W_{h(i)} : i \in \omega\}$ be a uniform class. For every polynomial $\overline{p}(x)$ of degree ≥ 2 there exists a class of distributions for $\mathcal{W}_h$ polynomially bounded by $\overline{p}$. Let, for all $i\in\omega$:

$$\delta^{1,i}_{\overline{p}}(n) = \begin{cases} \frac{1}{\sum_{j\in W_{h(i)}} \frac{1}{\overline{p}(j)}} \cdot \frac{1}{\overline{p}(n)} & \text{if } n\in W_{h(i)} \\ 0 & \text{otherwise} \end{cases}$$

It is readily seen that $\delta^{1,i}_{\overline{p}}$ is a distribution for $W_{h(i)}$ and that $\mathcal{D}^1_{\overline{p}}=\{\delta^{1,i}_{\overline{p}} : i\in\omega\}$ is a class of distributions for $\mathcal{W}_h$ polynomially bounded by $\overline{p}$. Note that, for every $W_{h(i)} \in \mathcal{W}_h$ there exists a unique $\delta\in\mathcal{D}^1_{\overline{p}}$ for $W_{h(i)}$, namely $\delta=\delta^{1,i}_{\overline{p}}$.

Notation 4 Let $\mathcal{W}_h$ be a uniform class. We will write $\mathcal{W}_h \in EX^{v\text{-}eff}_{meas}(\mathcal{D}^1)$ if, for every polynomially bounded class of distributions $\mathcal{D}$ for $\mathcal{W}_h$, $\mathcal{W}_h$ is EX-very efficiently measure identifiable with respect to $\mathcal{D}$.

The next theorem states that EX-very efficient identification on informant is stronger than EX-very efficient measure identification.

Theorem 4 *Let $\mathcal{W}_h=\{W_{h(i)} : i\in\omega\}$ be a uniform class. Then:*

$$\mathcal{W}_h \in EX^{v\text{-}eff}_{inf} \Rightarrow \mathcal{W}_h \in EX^{v\text{-}eff}_{meas}(\mathcal{D}^1).$$

The proof of Theorem 4 descends from the definitions, the constructions and the technical results given below. In this part we will refer to a generic uniform class $\mathcal{W}_h = \{W_{h(i)} : i \in \omega\}$ and to a generic class of distributions $\mathcal{D}$ for $\mathcal{W}_h$ polynomially bounded by $\overline{p}$ and, for every $W_{h(i)}\in\mathcal{W}_h$, $n\in W_{h(i)}$, $\delta\in\mathcal{D}$ for $W_{h(i)}$, we will indicate by (†) the condition $\delta(n)\geq\frac{1}{\overline{p}(n)}$.

Fact 1 *There exists a polynomial $p(x,y)$ such that, for every $\varepsilon \in [0,1]$, for every $W_{h(i)} \in \mathcal{W}_h$, for every $\delta \in \mathcal{D}$ for $W_{h(i)}$, for every $j,n \in \omega$, if $j \in W_{h(i)}$, $j \leq n$, then, for every $k \geq p(n, \frac{1}{\varepsilon})$, $Pr_{\delta}^{\leq k}(j) > 1 - \frac{\varepsilon}{(n+1)2^{n+1}}$.*

Proof Let $\varepsilon \in [0,1]$, $W_{h(i)} \in \mathcal{W}_h$, $\delta \in \mathcal{D}$ for $W_{h(i)}$, $j,n \in \omega$ with $j \in W_{h(i)}$, $j \leq n$. By (†), for every $k \in \omega$, $Pr_{\delta}^{>k}(j) \leq (1 - \frac{1}{\overline{p}(n)})^k$. By condition $(1-x) < e^{-x}$, it follows that $(1 - \frac{1}{\overline{p}(n)})^k \leq e^{-\frac{k}{\overline{p}(n)}}$. Moreover it is readily seen that

$$e^{-\frac{k}{\overline{p}(n)}} \leq \frac{\varepsilon}{(n+1)2^{n+1}} \Leftrightarrow k \geq \overline{p}(n)[ln\tfrac{1}{\varepsilon} + ln(n+1) + a(n+1)]$$

where $a = \frac{1}{log_2 e}$. So, letting $p(x,y) = \overline{p}(x)[y + (a+1)(x+1)]$, for every $k \geq p(n, \frac{1}{\varepsilon})$, we have $Pr_{\delta}^{\leq k}(j) > 1 - \frac{\varepsilon}{(n+1)2^{n+1}}$.

q.e.d.

Let $W_{h(i)} \in \mathcal{W}_h$ and let $\delta \in \mathcal{D}$ for $W_{h(i)}$. From a text t for $W_{h(i)}$ randomly generated by δ, we want to construct the informant $I^{h(i)}$ for $W_{h(i)}$. Informally, at each step k, we create a binary sequence I of length $k+1$ in which we put positive information ($I_n = 0$, $n \leq k$) if $n \in rng(t_{|k})$, and negative information ($I_n = 1$, $n \leq k$) if $n \notin rng(t_{|k})$. For this purpose, if $p(x,y)$ is the polynomial defined in Fact 1, consider the following functions:

- $H : Seq \times \omega \rightarrow Bseq$ such that, for every $\sigma \in Seq$ and $n \in \omega$:

$$H(\sigma, 0) = \begin{cases} <0> & \text{if } 0 \in rng(\sigma) \\ <1> & \text{otherwise} \end{cases}$$

$$H(\sigma, n+1) = \begin{cases} H(\sigma,n) * <0> & \text{if } n+1 \in rng(\sigma) \\ H(\sigma,n) * <1> & \text{otherwise} \end{cases}$$

- $\alpha : \omega \times [0,1] \rightarrow \omega$ such that, for every $k \in \omega$ and $\varepsilon \in [0,1]$, if $k_{0,\varepsilon} = p(0, \frac{1}{\varepsilon})$,

$$\alpha(k,\varepsilon) = \begin{cases} max\{r \in \omega : p(r, \frac{1}{\varepsilon}) < k\} & \text{if } k \geq k_{0,\varepsilon} \\ 0 & \text{otherwise} \end{cases}$$

- $\beta : \omega \times [0,1] \rightarrow \omega$ such that, for every $n \in \omega$ and $\varepsilon \in [0,1]$,

$$\beta(n,\varepsilon) = \left\lfloor p\left(n, \tfrac{1}{\varepsilon}\right) + 1 \right\rfloor.$$

- $I : Seq \times [0,1] \rightarrow Bseq$ such that, for every $\sigma \in Seq$ and $\varepsilon \in [0,1]$:

$$I(\sigma,\varepsilon) = H(\sigma, \alpha(lth(\sigma) - 1, \varepsilon)).$$

Remark 4 (i) For every $\sigma \in Seq$ and $\varepsilon \in [0,1]$, $I(\sigma,\varepsilon)$ is the code of the initial segment of an informant of length $\alpha(lth(\sigma)-1,\varepsilon)+1$, i.e. $I(\sigma,\varepsilon) = < i_0, ..., i_{\alpha(lth(\sigma)-1,\varepsilon)} >$, where, for every $j \leq \alpha(lth(\sigma)-1,\varepsilon)$, $i_j \in \{0,1\}$, $i_j = (H(\sigma,j))_j$ and $(I(\sigma,\varepsilon))_j = 0$ if and only if $j \in rng(\sigma)$.

(ii) Let $W_{h(i)} \in \mathcal{W}_h$, let t be a text randomly generated by $\delta \in \mathcal{D}$ for $W_{h(i)}$ and $\varepsilon \in [0,1]$. It is easy to see that, for every $k \in \omega$, $I(t_{|k},\varepsilon)$ has length $\alpha(k,\varepsilon)+1$ and contains as positive information elements of $W_{h(i)}$ and as negative information elements which do not occur in $rng(t_{|k})$ yet and which with "high" probability ($>1-\varepsilon$) are not in $W_{h(i)}$.

In the following, for every $\varepsilon \in [0,1]$ and $t \in \omega^\omega$, we let:

$$I(t,\varepsilon) = \bigcup_{k \in \omega} {}^{\#}(I(t_{|k},\varepsilon)),$$

where, if $\sigma = < a_0, ..., a_n >$, then ${}^{\#}\sigma = (a_0, ..., a_n)$.

Fact 2 *(i)* $H(\sigma,n)\in P$.

(ii) For every $\varepsilon\in[0,1]$, $\alpha(k,\varepsilon)$ *is a function increasing in* k *such that, for every* $n\in\omega$, $p(\alpha(n,\varepsilon),\frac{1}{\varepsilon})<n$ *and* $\alpha(\beta(n,\varepsilon),\varepsilon)=n$.

(iii) $I(\sigma,\varepsilon)\in P$.

(iv) For every $\varepsilon\in[0,1]$, *for every* $t\in\omega^\omega$, $\{j\in\omega : I(t,\varepsilon)_j=0\}=rng(t)$.

Proof Immediate by the previous definitions.

q.e.d.

Definition 18 Let $\varepsilon\in[0,1]$, $W_{h(i)}\in\mathcal{W}_h$, $t\in\omega^\omega$ randomly generated by some $\delta\in\mathcal{D}$ for $W_{h(i)}$, $k\in\omega$. We say that $I(t_{|k},\varepsilon)$ *is wrong,* if there exists $j\in W_{h(i)}$, $j\le\alpha(k,\varepsilon)$, such that j does not occur within $k+1$ steps with respect to δ, i.e. $j\notin rng(t_{|k})$. If $I(t_{|k},\varepsilon)$ is not wrong, we say that $I(t_{|k},\varepsilon)$ *is correct.*

Note that, if $I(t,\varepsilon)$ is not an informant for $W_{h(i)}$, then there are $j\in W_{h(i)}$ and $k\in\omega$ such that $j\le\alpha(k,\varepsilon)$ and $(I(t_{|k},\varepsilon))_j=1$. This happens only if $j\notin rng(t_{|k})$. So $I(t_{|k},\varepsilon)$ is "wrong" according to Definition 18.

Definition 19 Let $\varepsilon\in[0,1]$, $W_{h(i)}\in\mathcal{W}_h$, $t\in\omega^\omega$ randomly generated by some $\delta\in\mathcal{D}$ for $W_{h(i)}$, $n\in\omega$. Define:

$$E_{n,\varepsilon}\equiv(\exists k)[\alpha(k,\varepsilon)=n\wedge I(t_{|k},\varepsilon)\text{ wrong}].$$

The following results are easily estabilished (proofs are left as exercise).

Fact 3 *For every* $n\in\omega$, $E_{n,\varepsilon}$ *holds if and only if* $I(t_{|\beta(n,\varepsilon)},\varepsilon)$ *is wrong.*

Fact 4 *The following are equivalent:*

(i) For every $n\in\omega$, $E_{n,\varepsilon}$ *does not hold.*

(ii) For every $k\in\omega$, $I(t_{|k},\varepsilon)$ *is correct.*

Fact 5 $I(t,\varepsilon)$ *is the informant for* $W_{h(i)}$ *with probability* $>1-\varepsilon$.

Proof By construction, $I(t,\varepsilon)$ is an informant (Remark 3). Moreover, for every $j\in\omega$, $I(t,\varepsilon)_j=0$ if and only if $j\in W_{h(i)}$ (Fact 2(*iv*)). It remains to show that the probability that $I(t,\varepsilon)$ is wrong is $\le\varepsilon$. Let $j,n\in\omega$ be such that $j\le n$. Since $p(j,\frac{1}{\varepsilon})<\beta(n,\varepsilon)$, if $j\in W_{h(i)}$, then

$$Pr_\delta^{>\beta(n,\varepsilon)+1}(j)\le\tfrac{\varepsilon}{(n+1)2^{n+1}}$$

(Fact 1). But $\alpha(\beta(n,\varepsilon),\varepsilon)=n$, so the probability that $I(t_{|\beta(n,\varepsilon)},\varepsilon)$ is wrong is bounded by $\sum_{j\le n}Pr_\delta^{>\beta(n,\varepsilon)+1}(j)$ and

$$\textstyle\sum_{j\le n}Pr_\delta^{>\beta(n,\varepsilon)+1}(j)\le\frac{\varepsilon}{2^{n+1}}.$$

Therefore the probability that $E_{n,\varepsilon}$ holds is $\le\frac{\varepsilon}{2^{n+1}}$ (Fact 3) and the probability that there exists $n\in\omega$ such that $E_{n,\varepsilon}$ holds is bounded by $\sum_{n\in\omega}\frac{\varepsilon}{2^{n+1}}=\varepsilon$. Hence, for every $k\in\omega$, $I(t_{|k},\varepsilon)$ is a correct initial segment of $I^{h(i)}$ with probability $>1-\varepsilon$ (Fact 4), and $I(t,\varepsilon)=I^{h(i)}$ with probability $>1-\varepsilon$.

q.e.d.

We are now in a position to prove Theorem 4.

Theorem 4 *Let $\mathcal{W}_h = \{W_{h(i)} : i \in \omega\}$ be a uniform class. Then:*

$$\mathcal{W}_h \in EX^{v\text{-}eff}_{inf} \Rightarrow \mathcal{W}_h \in EX^{v\text{-}eff}_{meas}(\mathcal{D}^1).$$

Proof Let $\mathcal{D}$ be a class of distributions for $\mathcal{W}_h$ polynomially bounded by $\overline{p}$. Assume $\mathcal{W}_h \in EX^{v\text{-}eff}_{inf}$ and let $g \in P$ and the polynomial q be such that, for every $W_{h(i)} \in \mathcal{W}_h$, g EX-identifies on informant $W_{h(i)}$ in at most $q(|m(i)|)$ guesses. Let $\overline{q}(x,y) = p(q(x),y)+2$ (p defined in Fact 1). Let g be such that, for every $\varepsilon \in [0,1]$, for every $a_0, ...a_n \in \omega$,

$$g'(< a_0, ..., a_n >, \varepsilon) = g(I(< a_0, ..., a_n >, \varepsilon)).$$

Obviously $g' \in P$. Let $\varepsilon \in [0,1]$, $W_{h(i)} \in \mathcal{W}_h$ and let t be a text for $W_{h(i)}$ randomly generated by some $\delta \in \mathcal{D}$ for $W_{h(i)}$. Define:

$$\mathcal{A} = \{t \in \omega^\omega : g'\, EX - \textit{converges on } t, \varepsilon \textit{ to } W_{h(i)} \textit{ in } < \overline{q}(|m(i)|, \tfrac{1}{\varepsilon}) \textit{ steps}\}$$
$$\mathcal{A}' = \{t \in \omega^\omega : g\, EX - \textit{identifies } W_{h(i)} \textit{ on } I(t,\varepsilon) \textit{ in } < q(|m(i)|) \textit{ steps}\}.$$

It is readily seen that $\mathcal{A} \supseteq \mathcal{A}'$. So, for every $k \geq \overline{q}(|m(i)|, \frac{1}{\varepsilon}) - 1$, $I(t_{|k}, \varepsilon)$ is (the code of) a binary sequence of length $\geq q(|m(i)|)$. Moreover $\mathcal{A}' \supseteq \{t \in \omega^\omega : I(t,\varepsilon) = I^{h(i)}\}$. But $\mu_\delta(\{t \in \omega^\omega : I(t,\varepsilon) = I^{h(i)}\}) > 1 - \varepsilon$ (Fact 5), so $\mu_\delta(\mathcal{A}) > 1 - \varepsilon$. It follows that g' EX-measure identifies $W_{h(i)}$ in at most $\overline{q}(|m(i)|, \frac{1}{\varepsilon})$ guesses with respect to δ, hence $\mathcal{W}_h \in EX^{v\text{-}eff}_{meas}(\mathcal{D})$.

q.e.d.

Theorem 5 *Let $\mathcal{W}_h$ be a uniform class. Then:*

$$\mathcal{W}_h \in EX^{eff}_{inf} \Rightarrow \mathcal{W}_h \in EX^{eff}_{meas}(\mathcal{D}^1).$$

Proof Analogous to Theorem 4.

q.e.d.

Corollary 1 *Let $\mathcal{W}_h$ be a uniform class. Then:*

$$\mathcal{W}_h \textit{ satisfies } (*) \Rightarrow \mathcal{W}_h \in EX^{eff}_{meas}(\mathcal{D}^1).$$

Proof Immediate from Theorem 1(i) and Theorem 5.

q.e.d.

Unfortunately, in both efficient paradigms, measure identification does not imply identification on informant.

Proposition 1 *There exist a uniform class $\mathcal{W}$ and a universal function $\Psi(i,x)$ for $\mathcal{W}$ such that, if h is a Ψ-indexing, then $\mathcal{W} = \mathcal{W}_h$ and $\mathcal{W}_h \in EX^{v\text{-}eff}_{meas} \setminus EX^{eff}_{inf}$.*

Proof Let $\mathcal{W} = \{L_i : i \in \omega\}$ be such that, for every $i \in \omega$,

- if i is odd: $L_i = \{x \in \omega : x \in Pow_n(i) \textit{ for some } n \geq 1\}$ where, for every $n \geq 1$,

$$x \in Pow_n(i) \iff x = \underbrace{2^{2^{\cdot^{\cdot^{\cdot^{2^i}}}}}}_{n}$$

- if i is even: $L_i = L_1$.

Define for every $i, x \in \omega$:

$$\Psi(i,x) = \begin{cases} 0 & \text{if } i \text{ is odd and } x \in Pow_n(i) \text{ for some } n \geq 1 \\ & \quad \text{or if } i \text{ is even and } x \in Pow_n(1) \text{ for some } n \geq 1 \\ 1 & \text{otherwise} \end{cases}$$

It is easily seen that $\Psi(i,x)$ is a universal function for $\mathcal{W}$ such that, if h is a Ψ-indexing, for i odd $(i\neq 1)$, $W_{h(i)}=L_i$ $(m(i)=i)$, while for i even or $i=1$, $W_{h(i)}=L_1$ $(m(i)=0)$.

$\mathcal{W}_h \notin EX^{eff}_{inf}$. For every polynomial p, for sufficiently large i, $p(i+2)<2^i-1$. For one of these odd i, we cannot distinguish $W_{h(i)}$ from $W_{h(i+2)}$ before 2^i-1 steps. Then $\mathcal{W}_h$ does not satisfy $(*)$ and $\mathcal{W}_h \notin EX^{eff}_{inf}$.

$\mathcal{W}_h \in EX^{v\text{-}eff}_{meas}$. For every $\varepsilon\in[0,1]$ and $a_0,...,a_n\in\omega$ let:

$$g(<a_0,...,a_n>,\varepsilon)=\begin{cases} i & \text{if there exists an odd } i \text{ such that} \\ & \quad a_0\in Pow_n(i), \text{ for some } n\geq 1 \\ 0 & \text{otherwise}\end{cases}$$

Obviously $g\in P$. Moreover, for every class of distributions $\mathcal{D}$ for $\mathcal{W}_h$, $\varepsilon\in[0,1]$ and $W_{h(i)}\in\mathcal{W}_h$, if t is a text for $W_{h(i)}$ randomly generated by some $\delta\in\mathcal{D}$ for $W_{h(i)}$, then $t_0\in Pow_n(i)$ for some $n\geq 1$. Hence g EX-measure identifies $W_{h(i)}$ after the first guess.

q.e.d.

However not all uniform classes of languages are EX-efficiently measure identifiable with respect to $\mathcal{D}^1$ (hence with respect to every class of distributions for it).

Proposition 2 *There exist a uniform class $\mathcal{W}$ and a universal function $\Psi(i,x)$ for $\mathcal{W}$ such that, if h is a Ψ-indexing, then $\mathcal{W}=\mathcal{W}_h$ and $\mathcal{W}_h\notin EX^{eff}_{meas}(\mathcal{D}^1)$.*

Proof Let $\mathcal{W}=\{L_i : i\in\omega\}$ be such that $L_0=\omega$ and, for every $i\in\omega\setminus\{0\}$, $L_i=\omega\setminus\{2^i\}$. Define, for every $i,x\in\omega$:

$$\Psi(i,x)=\begin{cases} 1 & \text{if } i\neq 0 \text{ and } x=2^i \\ 0 & \text{otherwise}\end{cases}$$

It is readily seen that $\Psi(i,x)$ is a universal function for $\mathcal{W}$ such that, if h is a Ψ-indexing, for $i=0$, $W_{h(0)}=\omega$ $(m(0)=0)$, otherwise $W_{h(i)}=\omega\setminus\{2^i\}$ $(m(i)=i)$.

We want now to show the existence of a polynomially bounded class of distributions $\mathcal{D}$ for $\mathcal{W}_h$ such that, for every $g(x,y)\in P$, for every polynomial $q(x,y)$, there exist $\varepsilon\in[0,1]$, $W_{h(i)}\in\mathcal{W}_h$ and $\delta\in\mathcal{D}$ for $W_{h(i)}$ such that:

$$\mu_\delta(\{t\in\omega^\omega : g\, EX-converges\ on\ t,\varepsilon\ to\ W_{h(i)}\ in\ <q(i,\tfrac{1}{\varepsilon})\ steps\})\not> 1-\varepsilon,$$

or, equivalently,

$$\mu_\delta(\{t\in\omega^\omega : g\ does\ not\ EX-converge\ on\ t,\varepsilon\ to\ W_{h(i)}\ in\ <q(i,\tfrac{1}{\varepsilon})\ steps\})\geq\varepsilon.$$

Let $\overline{p}(x)=(x+1)(x+2)$ and consider $\mathcal{D}^{\frac{1}{\overline{p}}}$, the class of distributions for $\mathcal{W}_h$ constructed from $\overline{p}$ as shown in Remark 2. Let $\delta_0\in\mathcal{D}^{\frac{1}{\overline{p}}}$ be for $W_{h(0)}$ $(\delta_0=\delta^{1,0}_{\overline{p}})$. For every $i\in\omega$, for all $x<2^i$, $\delta_0(x)\geq\frac{1}{\overline{p}(x)}$. The probability that, within k steps, only $x<2^i$ are generated by δ_0 is bounded by $\left(\sum_{x<2^i}\delta_0(x)\right)^k$. Moreover it is readily seen that

$$(\textstyle\sum_{x<2^i}\delta_0(x))^k\geq e^{-\frac{2k}{2^i+1}}$$

and, for every $\varepsilon\in[0,1]$,

$$e^{-\frac{2k}{2^i+1}}\geq 2\varepsilon \Leftrightarrow k\leq\tfrac{2^i+1}{2}ln\tfrac{1}{2\varepsilon}.$$

For every $g(x,y)\in P$ and polynomial $q(x,y)$, if $\overline{\varepsilon}=\frac{e^{-2}}{2}$, let $q'(x)=q(x,\frac{1}{\overline{\varepsilon}})$. For sufficiently large i, $q'(i)<2^i$ and $\frac{(2^i+1)}{2}ln\frac{1}{2\overline{\varepsilon}}>q'(i)+1$. Hence

$$e^{-\frac{2(q'(i)+1)}{2^i+1}} \geq 2\overline{\varepsilon}.$$

Consider $W_{h(i)}$ for one such i and let $\delta \in \mathcal{D}^1_{\overline{p}}$ be for $W_{h(i)}$ $(\delta = \delta^{1,i}_{\overline{p}})$. For every $x < 2^i$, $\delta(x) > \delta_0(x)$, so:

$$(\sum_{x<2^i} \delta(x))^{q'(i)+1} > (\sum_{x<2^i} \delta_0(x))^{q'(i)+1} \geq 2\overline{\varepsilon}.$$

Let $\mathcal{B} = \cup\{B_\sigma : lth(\sigma) = q'(i), rng(\sigma) \subseteq \{0, 1, ..., 2^i - 1\}\}$. Note that

$$\mathcal{B} = \{t \in \omega^\omega : (\exists \sigma \in \omega^{<\omega})(lth(\sigma) = q'(i) \wedge rng(\sigma) \subseteq \{0, 1, ..., 2^i - 1\} \wedge t \supseteq \sigma)\}.$$

Hence $\mu_\delta(\mathcal{B}) = (\sum_{x<2^i} \delta(x))^{q'(i)+1}$ and $\mu_{\delta_0}(\mathcal{B}) = (\sum_{x<2^i} \delta_0(x))^{q'(i)+1}$. Moreover it is obvious that:

(1) For every $B_\sigma \in \mathcal{B}$: $\mu_\delta(B_\sigma) > \mu_{\delta_0}(B_\sigma)$ (2) $\mu_\delta(\mathcal{B}) > \mu_{\delta_0}(\mathcal{B}) \geq 2\overline{\varepsilon}$.

Let:

$$\mathcal{B}_1 = \{t \in \mathcal{B} : g(t_{|q'(i)}, \overline{\varepsilon}) \textit{ is an index for } W_{h(0)}\}$$
$$\mathcal{B}_2 = \{t \in \mathcal{B} : g(t_{|q'(i)}, \overline{\varepsilon}) \textit{ is an index for } W_{h(i)}\}$$
$$\mathcal{B}_3 = \mathcal{B} \setminus (\mathcal{B}_1 \cup \mathcal{B}_2).$$

By (1), (2) and the properties of a measure, since $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3$, one of the following cases takes place:

(I) $\mu_\delta(\mathcal{B}_1 \cup \mathcal{B}_3) > \mu_{\delta_0}(\mathcal{B}_1 \cup \mathcal{B}_3) \geq \overline{\varepsilon}$ (II) $\mu_\delta(\mathcal{B}_2 \cup \mathcal{B}_3) > \mu_{\delta_0}(\mathcal{B}_2 \cup \mathcal{B}_3) \geq \overline{\varepsilon}$.

In case (I), if $\mathcal{A} = \{t \in \omega^\omega : g$ does not EX-converge on $t, \overline{\varepsilon}$ to $W_{h(i)}$ in $< q(i, \frac{1}{\overline{\varepsilon}})$ steps$\}$, then $\mathcal{A} \supseteq \mathcal{B}_1 \cup \mathcal{B}_3$, hence $\mu_\delta(\mathcal{A}) > \overline{\varepsilon}$.

In case (II), if $\mathcal{A} = \{t \in \omega^\omega : g$ does not EX-converge on $t, \overline{\varepsilon}$ to $W_{h(0)}$ in $< q(0, \frac{1}{\overline{\varepsilon}})$ steps$\}$, then $\mathcal{A} \supseteq \mathcal{B}_2 \cup \mathcal{B}_3$, hence $\mu_{\delta_0}(\mathcal{A}) \geq \overline{\varepsilon}$.

In each case there exist $W_{h(i)} \in \mathcal{W}_h$ and $\delta \in \mathcal{D}^1_{\overline{p}}$ for $W_{h(i)}$ such that g does not EX-measure identify $W_{h(i)}$ in at most $q(i, \frac{1}{\overline{\varepsilon}})$ guesses with respect to δ, so $\mathcal{W}_h \notin EX^{eff}_{meas}(\mathcal{D}^1)$.

q.e.d.

6 Probabilistic identification

Measure identification is not the only way to treat identification from a probabilistic point of view. For example, if $\mathcal{W}$ is a uniform class and $L \in \mathcal{W}$, we can require the learner g to make, at each step n, a guess on the basis of the initial segment of length n+1 of the informant for L and of a binary sequence of length n+1, which we can imagine generated by random coin drawns (*coin sequence*). This approach aims to represent the idea that the conjectures of a learner can be not only determined by "exact" information on L, but also by random and independent external events. Moreover it "merges" the previous two efficient identification paradigms, since it considers positive and negative information in a probabilistic setting. We will say that g *(very) efficiently identifies* L *with probability* if, for every $\varepsilon \in [0, 1]$, on the basis of the two sequences described above, g stabilizes on an index for L in time polynomial in the (length of the) least index for L in $\mathcal{W}$ with probability $> 1 - \varepsilon$.

Definition 20 (i) A *coin* τ is a text such that $rng(\tau) \subseteq \{0, 1\}$.

(ii) We call $\sigma \in Seq$ a *coin-sequence* if $rng(\sigma) \subseteq \{0, 1\}$.

Recall that, if $\mathcal{W}_h = \{W_{h(i)} : i \in \omega\}$ is a uniform class, for every $i \in \omega$, $I^{h(i)}$ denotes the informant for $W_{h(i)}$.

Definition 21 Let $\mathcal{W}_h = \{W_{h(i)} : i \in \omega\}$ be a uniform class and let $g(x,y)$ be a total recursive function. For every $s \in 2^\omega$, for every $k \in \omega$, for every $W_{h(i)} \in \mathcal{W}_h$, if $t = I^{h(i)}$ we say that g *EX-converges on* t, s *to* $W_{h(i)}$ *in at most* k *steps* (in short: *in* $<k$ *steps*) if:

$$(\exists n_0 < k)(\forall n \geq n_0)(g(t_{|n}, s_{|n}) = i')$$

for some $i' \in \omega$ such that $W_{h(i)} = W_{h(i')}$.

Throughout this part we denote by Δ a σ-additive measure on the Cantor space 2^ω.

Definition 22 Let $\mathcal{W}_h = \{W_{h(i)} : i \in \omega\}$ be a uniform class. We say that $\mathcal{W}_h$ is EX-*efficiently identifiable with probability* ($\mathcal{W}_h \in EX^{\mathit{eff}}_{\mathit{prob}}$), if there exist a function $g(x,y) \in P$ and a polynomial $p(x,y)$ such that, for every $W_{h(i)} \in \mathcal{W}_h$, g EX-identifies $W_{h(i)}$ with probability in at most $p(m(i), \frac{1}{\varepsilon})$ guesses, i.e., for every $\varepsilon \in [0,1]$, if $t = I^{h(i)}$:

$$\Delta(\{s \in 2^\omega : g\, EX-\text{converges on } t, s \text{ to } W_{h(i)} \text{ in } < p(m(i), \tfrac{1}{\varepsilon}) \text{ steps}\}) > 1 - \varepsilon.$$

Definition 23 Let $\mathcal{W}_h = \{W_{h(i)} : i \in \omega\}$ be a uniform class. We say that $\mathcal{W}_h$ is EX-*very efficiently identifiable with probability* ($\mathcal{W}_h \in EX^{\mathit{v\text{-}eff}}_{\mathit{prob}}$), if there exist a function $g(x,y) \in P$ and a polynomial $p(x,y)$ such that, for every $W_{h(i)} \in \mathcal{W}_h$, g EX-identifies $W_{h(i)}$ with probability in at most $p(|m(i)|, \frac{1}{\varepsilon})$ guesses, i.e., for every $\varepsilon \in [0,1]$, if $t = I^{h(i)}$:

$$\Delta(\{s \in 2^\omega : g\, EX-\text{converges on } t, s \text{ to } W_{h(i)} \text{ in } < p(|m(i)|, \tfrac{1}{\varepsilon}) \text{ steps}\}) > 1 - \varepsilon.$$

6.1 Main results for $EX^{\mathit{eff}}_{\mathit{prob}}$

It seems reasonable that identification on informant is related to identification with probability, since the latter is based on the informant for the unknown language too. Really the two identification paradigms turns out to be equivalent in the efficient setting. Let $\mathcal{W}_h = \{W_{h(i)} : i \in \omega\}$ be a uniform class, and let $\Psi(i,x)$ and $h(i)$ be a universal function for $\mathcal{W}_h$ and a Ψ-indexing respectively. Recall that $(*)$ denotes the following condition:

$$(\exists \text{ polynomial } p)(\forall i)(\forall j < m(i))(\exists x \leq p(m(i)))(\varphi_{h(j)}(x) \neq \varphi_{h(i)}(x)).$$

Theorem 6 *Let* $\mathcal{W}_h = \{W_{h(i)} : i \in \omega\}$ *be a uniform class. Then:*

$$\mathcal{W}_h \in EX^{\mathit{eff}}_{\mathit{prob}} \Leftrightarrow \mathcal{W}_h\, satisfies\, (*).$$

Proof ($\Rightarrow$) Let $\mathcal{W}_h \in EX^{\mathit{eff}}_{\mathit{prob}}$ by $g \in P$. Suppose by contradiction that $\mathcal{W}_h$ does not satisfy $(*)$, i.e.:

$$(\forall \text{ polynomial } q)(\exists i)(\exists j < m(i))(\forall x \leq q(m(i)))(\varphi_{h(j)}(x) = \varphi_{h(i)}(x)). \tag{1}$$

We want to show that, for every polynomial $q(x,y)$, there exist $\varepsilon \in [0,1]$, $W_{h(i)} \in \mathcal{W}_h$ such that, if $t = I^{h(i)}$:

$$\Delta(\{s \in 2^\omega : g\, EX-converges\ on\ t, s\, to\, W_{h(i)}\ in < q(m(i), \tfrac{1}{\varepsilon})\ steps\}) \not> 1 - \varepsilon$$

or, equivalently,

$$\Delta(\{s \in 2^\omega : g\ does\ not\ EX-converge\ on\, t, s\ to\, W_{h(i)}\ in < q(m(i), \tfrac{1}{\varepsilon})\ steps\}) \geq \varepsilon.$$

For every polynomial $q(x,y)$, for every $\overline{\varepsilon} < \frac{1}{2}$, let $q'(x) = q(x, \frac{1}{\overline{\varepsilon}})$. By (1), corresponding to every q', there exist $i, j \in \omega$ such that $j < m(i)$ and, for every $x \leq q'(m(i))$, $\varphi_{h(j)}(x) = \varphi_{h(i)}(x)$. So, if $t = I^{h(i)}$ and $t' = I^{h(j)}$, $t_{|q'(m(i))} = t'_{|q'(m(i))}$. Consider:

$$\mathcal{B}_1 = \{s \in 2^\omega : g(t_{|q'(m(i))}, s_{|q'(m(i))}) \textit{ is an index for } W_{h(i)}\}$$

$$\mathcal{B}_2 = \{s \in 2^\omega : g(t_{|q'(m(i))}, s_{|q'(m(i))}) \textit{ is an index for } W_{h(j)}\}$$

$$\mathcal{B}_3 = \{s \in 2^\omega : g(t_{|q'(m(i))}, s_{|q'(m(i))}) \textit{ is neither an index for } W_{h(i)} \textit{ nor for } W_{h(j)}\}.$$

If $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3$, then $\Delta(\mathcal{B}) = 1$ and one of the following cases hold:

(I) $\Delta(\mathcal{B}_1 \cup \mathcal{B}_3) \geq \frac{1}{2}$ $\quad$ (II) $\Delta(\mathcal{B}_2 \cup \mathcal{B}_3) \geq \frac{1}{2}$.

In case (I), if $\mathcal{A} = \{s \in 2^\omega : g$ does not EX- converge on t, s to $W_{h(j)}$ in $< q(m(j), \frac{1}{\overline{\varepsilon}})$ steps$\}$, since $q'(m(i)) \geq q(m(j), \frac{1}{\overline{\varepsilon}})$, then $\mathcal{A} \supseteq \mathcal{B}_1 \cup \mathcal{B}_3$ and $\Delta(\mathcal{A}) \geq \frac{1}{2} > \overline{\varepsilon}$.

In case (II), if $\mathcal{A} = \{s \in 2^\omega : g$ does not EX-converge on t, s to $W_{h(i)}$ in $< q(m(i), \frac{1}{\overline{\varepsilon}})$ steps$\}$, since $q'(m(i)) = q(m(i), \frac{1}{\overline{\varepsilon}})$, then $\mathcal{A} \supseteq \mathcal{B}_2 \cup \mathcal{B}_3$ and $\Delta(\mathcal{A}) \geq \frac{1}{2} > \overline{\varepsilon}$.

In each case there exists $W_{h(i)} \in \mathcal{W}_h$ such that g does not EX-identify with probability $W_{h(i)}$ in at most $q(m(i), \frac{1}{\overline{\varepsilon}})$ guesses, hence $\mathcal{W}_h \notin EX^{\mathit{eff}}_{\mathit{prob}}$.

($\Leftarrow$) Suppose that $\mathcal{W}_h$ satisfies $(*)$. By Theorem 1(i), $\mathcal{W}_h \in EX^{\mathit{eff}}_{\mathit{inf}}$ and let $g \in P$ and the polynomial p be such that, for every $W_{h(i)} \in \mathcal{W}_h$, g EX-identifies on informant $W_{h(i)}$ in at most $p(m(i))$ guesses. Define, for every $\sigma, \tau \in Bseq$,

$$g'(\sigma, \tau) = g(\sigma).$$

It is obvious that $g'(\sigma, \tau) \in P$ and, for every $\varepsilon \in [0, 1]$ and $W_{h(i)} \in \mathcal{W}_h$, g' EX-identifies with probability $W_{h(i)}$ in at most $p'(m(i), \frac{1}{\varepsilon})$ guesses, where $p'(x, y) = p(x)$.

q.e.d.

Corollary 2 *Let $\mathcal{W}_h$ be a uniform class. The following are equivalent:*

(i) $\mathcal{W}_h$ *satisfies* $(*)$.

(ii) $\mathcal{W}_h \in EX^{\mathit{eff}}_{\mathit{inf}}$.

(iii) $\mathcal{W}_h \in EX^{\mathit{eff}}_{\mathit{prob}}$.

Proof Immediate from Theorem 1(i) and Theorem 6.

q.e.d.

Theorem 7 *$P_{\mathcal{L}}$ is not EX-efficiently identifiable with probability with respect to any acceptable indexing.*

Proof Immediate from Theorem 2 and Corollary 2.

q.e.d.

Corollary 3 *Let $\mathcal{W}_h$ be a uniform class. Then:*

$$\mathcal{W}_h \in EX^{\mathit{eff}}_{\mathit{prob}} \Leftrightarrow \mathcal{W}_h \in EX^{\mathit{eff}}_{\mathit{inf}} \Rightarrow \mathcal{W}_h \in EX^{\mathit{eff}}_{\mathit{meas}}(\mathcal{D}^1).$$

Proof Immediate from Corollary 2 and Corollary 1.

q.e.d.

6.2 Main results for $EX_{prob}^{v\text{-}eff}$

Let $\mathcal{W}_h = \{W_{h(i)} : i \in \omega\}$ be a uniform class, and let $\Psi(i,x)$ and $h(i)$ be respectively a universal function for $\mathcal{W}_h$ and a Ψ-indexing. We recall that $(*')$ denotes the following condition:

$$(\exists \text{ polynomial } p)(\forall i)(\forall j{<}m(i))(\exists x{\le} p(|m(i)|))(\varphi_{h(j)}(x) \neq \varphi_{h(i)}(x)).$$

Theorem 8 *Let $\mathcal{W}_h = \{W_{h(i)} : i \in \omega\}$ be a uniform class. Then:*

$$\mathcal{W}_h \in EX_{prob}^{v\text{-}eff} \Rightarrow \mathcal{W}_h \, satisfies \, (*').$$

Proof Analogous to Theorem 6 ($\Rightarrow$).

q.e.d.

As in the identification on informant setting, the reverse implication of Theorem 8 is a very hard problem.

Theorem 9 *The following are equivalent:*

(i) *Every uniform class of languages satisfying $(*')$ is EX-very efficiently identifiable with probability.*

(ii) $\mathcal{NP} = \mathcal{RP}$.

The proof of Theorem 9 is based on the following auxiliary proposition.

Proposition 3 *Let $\Psi(i,x) \in P$. For every $\sigma \in Seq$, consider:*

$$f(\sigma) = \begin{cases} \mu i \le \sigma(\forall x < lth(\sigma))((\sigma)_x = \Psi(i,x)) & \text{if such an } i \text{ exists} \\ 0 & \text{otherwise} \end{cases}$$

If $\mathcal{NP} = \mathcal{RP}$, then f is computable by a probabilistic T.m. in time polynomial in $|\sigma|$.

Proof We give a hint of the proof, leaving the reader to complete it in detail. Consider the problem such that, for every $\sigma \in Seq$, for every $y, z \in \omega$,

$$\begin{aligned} R(\sigma,y,z) \equiv\ & (y+z \le \sigma) \wedge (\exists i \le y+z) \\ & [(y \le i) \wedge (\forall x < lth(\sigma))((\sigma)_x = \Psi(i,x))]. \end{aligned} \tag{1}$$

Such a problem is obviously in $\mathcal{NP}$. In fact there exists a nondeterministic T.m. which guesses, for every $\sigma \in Seq$ and $y, z \in \omega$, a number $i \le y+z$ and checks if such an i satisfies the right part of (1) (in polynomial time in $lth(\sigma)$). So, by hypothesis, this problem is in $\mathcal{RP}$ too. Let $\Theta(\sigma,y,z,\tau,\varepsilon)$ be a function that decides $R(\sigma,y,z)$ with probability $\ge 1-\varepsilon$ in time $p(|\sigma|,|y|,|z|,\frac{1}{\varepsilon})$. It is easy to show that f is computable in polynomial time by a probabilistic T.m. by means of Θ.

q.e.d.

We are now ready to prove Theorem 9.

Theorem 9 *The following are equivalent:*

(i) *Every uniform class of languages satisfying $(*')$ is EX-very efficiently identifiable with probability.*

(ii) $\mathcal{NP}=\mathcal{RP}$.

Proof $(i) \Rightarrow (ii)$ Consider SAT, the problem of deciding the satisfiability of propositional formulas A in conjunctive normal form ($A \in CNF$). We want to show that if condition $(*')$ is sufficient for EX-very efficient identifiability with probability of a uniform class of languages, then this problem is solvable by a polynomial time probabilistic algorithm. Since SAT is an $\mathcal{NP}$-complete problem, this will imply $\mathcal{NP}=\mathcal{RP}$, showing the statement. For every $A \in CNF$, we indicate by:

- $Var_A=\{x_i : x_i$ or $\overline{x_i}$ occurs in $A\}$; $k_A=card(Var_A)$, the cardinality of Var_A.
- $\tau : Var_A \rightarrow \{0,1\}$, a truth assignment on A; $Ass_A=\{\tau : \tau$ truth assignment on $A\}$.
- $\tau(A)$, the truth value of A on assignment τ; $|\tau|$, the cardinality of $rng(\tau)$.
- $\lceil A \rceil$ ($\lceil \tau \rceil$), the code of A (τ).

We consider CNF-formulas in the alphabet $\Sigma=\{x, |, \wedge, \vee, \_, (,)\}$ and we suppose them coded as shown in the Preliminaries. Recall that, for every $n \in \omega$, one can check if n is the code of some $A \in CNF$ and, in the positive case, decode it in time polynomial in $|n|$. The same properties hold for $n \in COD$, where $COD=\{<\lceil A \rceil, \lceil \tau \rceil> : A \in CNF, \tau \in Ass_A\}$. Moreover, if $n=<\lceil A \rceil, \lceil \tau \rceil>$, then $|n|=2(|\lceil A \rceil|+|\lceil \tau \rceil|+1)$, so $|\lceil A \rceil| \leq |n| \leq 4|\lceil A \rceil|+2$.

Let $\Psi(n,x)$ be such that, for every $n, x \in \omega$:

- if $n \in COD$, $n=<\lceil A \rceil, \lceil \tau \rceil>$ for some $A \in CNF$, $\tau \in Ass_A$:

$$\Psi(n,x)=\begin{cases} [\lceil A \rceil]_x & \text{if } x<|\lceil A \rceil| \\ 1 & \text{if } x=|\lceil A \rceil|+4 \text{ and } \tau(A)=1 \\ 0 & \text{otherwise} \end{cases}$$

- if $n \notin COD$: $\Psi(n,x)=0$.

Intuitively, if $n \in COD$ and $n=<\lceil A \rceil, \lceil \tau \rceil>$, $\Psi(n,x)$ assumes values:

$$[\lceil A \rceil]_0, ..., [\lceil A \rceil]_{|\lceil A \rceil|-1}, 0, 0, 0, 0, \tau(A), 0, 0.....$$

where $\tau(A)=0$ or $\tau(A)=1$. It is readily seen that $\Psi(n,x) \in P$. So, if $h \in P$ is such that, for every $n, x \in \omega$, $\Psi(n,x)=\varphi_{h(n)}(x)$, then $\mathcal{W}_h=\{W_{h(n)} : n \in \omega\}$ is a uniform class of languages. Moreover it is easy to verify that $\mathcal{W}_h$ satisfies $(*')$. Hence, by hypothesis, $\mathcal{W}_h \in EX^{v\text{-}eff}_{prob}$. Let $g(x,y) \in P$ and $p(x,y)$ be such that, for every $\varepsilon \in [0,1]$ and $W_{h(n)} \in \mathcal{W}_h$, g EX-identifies with probability $W_{h(n)}$ in at most $p(|m(n)|, \frac{1}{\varepsilon})$ guesses. Consider $A \in CNF$, $\varepsilon \in [0,1]$ and let $n=max\{x_i : x_i$ or $\overline{x_i}$ occurs in $A\}$. Define $\sigma^A \in Bseq$ such that, if $1^{n+1}=\underbrace{1, ..., 1}_{n+1}$, then $lth(\sigma^A)=p(|<\lceil A \rceil, 1^{n+1}>|, \frac{1}{\varepsilon})$ and

$$\sigma^A=<[\lceil A \rceil]_0, ..., [\lceil A \rceil]_{|\lceil A \rceil|-1}, 0, 0, 0, 0, 1, 0, ..., 0> .$$

Note that, for $\tau \in Ass_A$, $|m(<\lceil A \rceil, \lceil \tau \rceil>)| \leq |<\lceil A \rceil, 1^{n+1}>|$. Moreover, if $A \in SAT$, $\rho \in Bseq$ and $lth(\rho)=lth(\sigma^A)$, then, for every $\varepsilon \in [0,1]$, with probability $>1-\varepsilon$, $g(\sigma^A, \rho)=m$, where $m=<\lceil A \rceil, \lceil \tau \rceil>$ for some $\tau \in Ass_A$ such that $\tau(A)=1$. On the other hand, if $A \notin SAT$, $\rho \in Bseq$ and $lth(\rho)=lth(\sigma^A)$, then, for every $\varepsilon \in [0,1]$, $g(\sigma^A, \rho)$ is not of the type $<\lceil A \rceil, \lceil \tau \rceil>$ with $<\lceil A \rceil, \lceil \tau \rceil> \in COD$ or $\tau(A)=0$. Hence, given $A \in CNF$, $\varepsilon \in [0,1]$, for every $\rho \in Bseq$, such that $lth(\rho)=lth(\sigma^A)$, let $g(\sigma^A, \rho)=s$.

If $s \neq <\lceil A \rceil, \lceil \tau \rceil>$ with $\tau \in Ass_A$, then $A \notin SAT$. Otherwise, if $\tau=(s)_2$, we can check if τ satisfies A or not ($A=(s)_1$). In the positive case $A \in SAT$, otherwise $A \notin SAT$. Moreover the constructions of σ^A and ρ, the computation of $g(\sigma^A, \rho)$ and other checks can be executed in time polynomial in $lth(A)$, $|\rho|$ and $\frac{1}{\varepsilon}$ (since $|\lceil A \rceil| \leq p'(lth(A))$ for some polynomial p'). It follows that the problem of deciding if $A \in SAT$ is solvable in time polynomial with probability $> 1 - \varepsilon$, hence $SAT \in \mathcal{RP}$ and $\mathcal{NP} = \mathcal{RP}$.

$(ii) \Rightarrow (i)$ Let $\mathcal{W}_h = \{W_{h(i)} : i \in \omega\}$ be a uniform class of languages with universal function $\Psi(i, x)$. Suppose that $\mathcal{W}_h$ satisfies $(*')$ with the polynomial p. Let g be such that, for every $a_0, ..., a_n \in \omega$ and $\tau \in Bseq$:

$$g(< a_0, ..., a_n >, \tau) = \begin{cases} i & \text{where } i = \mu j \leq < a_0, ..., a_n > [(\forall x \leq n) \\ & \quad (\Psi(j, x) = a_x)], \text{ if such an } i \text{ exists} \\ 0 & \text{otherwise} \end{cases}$$

If $\mathcal{NP} = \mathcal{RP}$, then g is computable by a p-time probabilistic T.m. (Proposition 3). Let $W_{h(i)} \in \mathcal{W}_h$. For every $j < m(i)$, there exists $x \leq p(|m(i)|)$ such that $\Psi(j, x) \neq \Psi(i, x)$. Moreover it is readily seen that $m(i) \leq < \varphi_{h(i)}(0), ..., \varphi_{h(i)}(p(|m(i)|)) >$, so, for every $n \geq p(|m(i)|)$, $g(< \varphi_{h(i)}(0), ..., \varphi_{h(i)}(n) >) = m(i)$. Hence, for every $\varepsilon \in [0, 1]$, g EX-identifies with probability $W_{h(i)}$ in at most $p(|m(i)|) + 1$ guesses.

q.e.d.

Theorem 10 *Let $\mathcal{W}_h = \{W_{h(i)} : i \in \omega\}$ be a uniform class. Then:*

$$\mathcal{W}_h \in EX_{inf}^{v\text{-}eff} \Rightarrow \mathcal{W}_h \in EX_{prob}^{v\text{-}eff}.$$

Proof Let $g \in P$ and the polynomial p be such that, for every $W_{h(i)} \in \mathcal{W}_h$, g EX-identifies on informant $W_{h(i)}$ in at most $p(|m(i)|)$ guesses. Let, for every $\sigma, \tau \in Bseq$, $g'(\sigma, \tau) = g(\sigma)$. Clearly $g'(\sigma, \tau) \in P$ and, if $p'(x, y) = p(x)$, for every $\varepsilon \in [0, 1]$ and $W_{h(i)} \in \mathcal{W}_h$, g' EX-identifies with probability $W_{h(i)}$ in at most $p'(|m(i)|, \frac{1}{\varepsilon})$ guesses.

q.e.d.

Remark 5 Note that, if $\mathcal{P} = \mathcal{NP}$, then, for every uniform class $\mathcal{W}_h$:

$$\mathcal{W}_h \text{ satisfies } (*') \Leftrightarrow \mathcal{W}_h \in EX_{inf}^{v\text{-}eff} \Leftrightarrow \mathcal{W}_h \in EX_{prob}^{v\text{-}eff},$$

and $\mathcal{NP} = \mathcal{RP}$ (Theorem 9). On the other hand, if $\mathcal{P} \neq \mathcal{NP}$ and $\mathcal{NP} = \mathcal{RP}$, then:

$$\mathcal{W}_h \in EX_{inf}^{v\text{-}eff} \Rightarrow \mathcal{W}_h \text{ satisfies } (*') \Leftrightarrow \mathcal{W}_h \in EX_{prob}^{v\text{-}eff}.$$

But, under $\mathcal{P} \neq \mathcal{NP}$, condition $(*')$ is not sufficient for EX-very efficient identification on informant; hence, for some uniform class $\mathcal{W}_h$, $\mathcal{W}_h \in EX_{prob}^{v\text{-}eff}$ but $\mathcal{W}_h \notin EX_{inf}^{v\text{-}eff}$.

References

[Angluin 80] Angluin, D.: "Inductive Inference of Formal Languages from Positive Data"; Information and Control, 45 (1980), 117-135.

[Angluin 88] Angluin, D.: "Identifying Languages from Stochastic Examples"; Information and Control (1988).

[Buss 86] Buss, S.R.: "Bounded Arithmetic"; Studies in Proof Theory Lecture Notes (1986).

[Gold 67] Gold, E.M.: "Language Identification in the Limit"; Information and Control, 10 (1967), 447-474.

[Fontani 00] Fontani, S.: "Efficient Identification of Classes of P-Time Functions"; Journal of Universal Computer Science, 6, 8 (2000), 759-780.

[Montagna and Simi 99] Montagna, F., Simi, G.: "Paradigms in Measure Theorethic Learning and in Informant Learning"; Studia Logica, 62, (1999), 243-268.

[Odifreddi 99] Odifreddi, P.: "Classical Recursion Theory II"; Elsevier, Amsterdam, 1999.

[Osherson, Stob and Weinstein 86] Osherson, D.N., Stob, M. and Weinstein, S.: "Systems that Learn. An Introduction to Learning Theory for Cognitive and Computer Scientists"; MIT Press, Cambridge MA (1986).

[Oxtoby 71] Oxtoby, J.C.: "Measure and Category. Graduate Texts in Mathematics"; New York, Springer-Verlag (1971).

[Shoenfield 58] Shoenfield, J. R.: " The Class of Recursive Functions"; Proceedings of the American Mathematical Society, 9, (1958).