Identification in the limit of Probabilistic Non Deterministic Automata and Undecidable problem for Multiplicity Automata

François Denis, Yann Esposito

LIF-CMI, UMR 6166, 39, rue F. Joliot Curie 13453 Marseille Cedex 13 FRANCE,
`fdenis,esposito@cmi.univ-mrs.fr`

Abstract : Probabilistic finite automata (PFA) model stochastic languages, i.e. probability distributions over strings. Inferring PFA from stochastic data is an open field of research. We show that PFA are identifiable in the limit with probability one. Multiplicity automata (MA) are another device which can be used to represent stochastic languages. We show that MA generate strictly more stochastic languages than PFA, but we show also that it is undecidable whether an MA generates a stochastic language. Moreover, stochastic languages generated from MA cannot be described by a recursively enumerable subset of MA.

Topics: algorithmic learning theory, identification in the limit, grammatical inference, stochastic languages, probabilistic automata, multiplicity automata.

1 Introduction

Probabilistic automata (PFA) are formal objects which model *stochastic languages*, i.e. probability distributions over words (1). They are composed of a *structure* which is a finite automaton (NFA) and of *parameters* associated with states and transitions which represent the probability for a state to be initial, terminal or the probability for a transition to be chosen. Given the structure of a probabilistic automaton A and a sequence of words $u_1, \ldots, u_n$ independently distributed according to a probability distribution P, computing parameters for A which maximize the likelihood of the observation is NP-hard (2). However in practical cases, algorithms based on EM (*Expectation-Maximization*) method (3) can be used to compute approximate values. On the other hand, inferring a probabilistic automaton (structure and parameters) from a sequence of words is a widely open field of research. In some applications, prior knowledge may help to choose a structure (for example, the standard model for biological sequence analysis (4)). Without prior knowledge, a complete graph structure can be chosen. But it is likely that in general, inferring both appropriate structure and parameters from data would provide better results (see for example (5)).

Several learning frameworks can be considered to study inference of PFA, which often consist in adaptations to the stochastic case of classical learning models. Here, we consider a variant of the identification in the limit model of Gold (6), adapted to the stochastic case in (7). Given a PFA A and a sequence $u_1, \ldots, u_n, \ldots$ independently drawn according to the associated distribution P_A, an inference algorithm must compute a PFA A_n from each subsequence $u_1, \ldots, u_n$ such that with probability one, the support of A_n is stationary from some index n and P_{A_n} converges to P_A; moreover, when parameters of the target A are rational numbers, it can be requested that A_n itself is stationary from some index. It has been shown that the set of deterministic probabilistic automata (PDFA), i.e. PFA whose structure is deterministic, is identifiable in the limit with probability one (8; 9; 10), the identification being exact when the parameters of the target are rational numbers. However, PDFA are far less expressive than PFA, i.e. the set of probability distributions associated with deterministic probabilistic automata is stricly included in the set of distributions generated from general probabilistic automata. This result has been extended to the class of Probabilistic Residual Finite Automata (PRFA), i.e. PFA A whose states generate a residual language of P_A (11; 12).

Here, we show that the whole class of PFA is identifiable in the limit, the identification being exact when the parameters of the target are rational numbers (Section 3).

Multiplicity automata (MA) are devices which model a set $\mathcal{F}_{\mathrm{MA}}$ of functions from Σ^* to $\mathbb{R}$. It has been shown that $\mathcal{F}_{\mathrm{MA}}$ is very efficiently learnable in a variant of the exact learning model of Angluin, where the learner can ask *equivalence* and *extended membership queries*(13; 14; 15). As PFA are particular MA, they are learnable in this model. However, the learning is improper in the sense that the output function is not a PFA but a multiplicity automaton. We show that the class of MA is maybe not a very suitable representation scheme to represent stochastic languages if the goal is to learn them from stochastic data. First, representation by MA is not robust, i.e. there are MA which does not compute a stochastic language and which are arbitrarily close to a given PFA. Second, we show that it is undecidable whether a MA generates a stochastic language (this problem was left open in (1)). That is, given a MA computed from stochastic data: it is possible that it does not compute a stochastic language and there are maybe no ways to detect it! Finally, let $\mathcal{S}_{\mathrm{MA}}(\Sigma)$ be the set of stochastic languages that can be computed from MA. We show that no recursively enumerable subset of MA can generate $\mathcal{S}_{\mathrm{MA}}(\Sigma)$. As a corollary, MA can compute stochastic languages that cannot be computable by PFA.

2 Preliminaries

2.1 Automata and Languages

Let Σ be a finite *alphabet*, and Σ^* be the set of words on Σ. The empty word is denoted by ε and the length of a word u is denoted by $|u|$. We denote by $<$ the length-lexicographic order on Σ^*. A *language* is a subset of Σ^*.

A *non deterministic finite automaton (NFA)* is a 5-tuple $A = \langle \Sigma, Q, Q_0, F, \delta \rangle$ where Q is a finite set of states, $Q_0 \subseteq Q$ is the set of initial states, $F \subseteq Q$ is the set of terminal states, δ is the *transition* function defined from $Q \times \Sigma$ to 2^Q. Let δ also denote the extension of the transition function defined from $2^Q \times \Sigma^*$ to 2^Q. An NFA is *deterministic (DFA)* if $\mathrm{Card}\,(Q_0) = 1$ and if $\forall q \in Q$, $\forall x \in \Sigma$, $\mathrm{Card}\,(\delta(q,x)) \leq 1$. An NFA is *trimmed* if for any state q, $q \in \delta(Q_0, \Sigma^*)$ and $\delta(q, \Sigma^*) \cap F \neq \emptyset$.

Let $A = \langle \Sigma, Q, Q_0, F, \delta \rangle$ be an NFA. A word $u \in \Sigma^*$ is *recognized* by A if $\delta(Q_0, u) \cap F \neq \varnothing$. The language recognized by A is $L_A = \{u \in \Sigma^* \mid \delta(Q_0, u) \cap F \neq \varnothing\}$.

2.2 Multiplicity Automata, Probabilistic Automata and Stochastic Languages

A *multiplicity automaton (MA)* is a 5-tuple $\langle \Sigma, Q, \varphi, \iota, \tau \rangle$ where Q is a finite set of states, $\varphi : Q \times \Sigma \times Q \to \mathbb{R}$ is the transition function, $\iota : Q \to \mathbb{R}$ is the initialization function and $\tau : Q \to \mathbb{R}$ is the termination function. We extend the transition function φ to $Q \times \Sigma^* \times Q$ by $\varphi(q, wx, r) = \sum_{s \in Q} \varphi(q, w, s)\varphi(s, x, r)$ where $x \in \Sigma$ and $\varphi(q, \varepsilon, r) = 1$ if $q = r$ and 0 otherwise. We extend again φ to $Q \times 2^{\Sigma^*} \times 2^Q$ by $\varphi(q, U, R) = \sum_{w \in U, r \in R} \varphi(q, w, r)$. Let $A = \langle \Sigma, Q, \varphi, \iota, \tau \rangle$ be an MA. Let P_A be the function defined by: $P_A(u) = \sum_{q, r \in Q} \iota(q)\varphi(q, u, r)\tau(r)$. The *support* of A is the NFA $\langle \Sigma, Q, Q_I, Q_T, \delta \rangle$ where $Q_I = \{q \in Q \mid \iota(q) \neq 0\}$, $Q_T = \{q \in Q \mid \tau(q) \neq 0\}$ and $\delta(q, x) = \{r \in Q \mid \varphi(q, x, r) \neq 0\}$ for any state q and any letter x. An MA is said to be *trimmed* if its support is a trimmed NFA.

A *semi Probabilistic Finite Automaton (semi-PFA)* is an MA such that ι, φ and τ take their values in $[0,1]$, such that $\sum_{q \in Q} \iota(q) \leq 1$ and for any state q, $\tau(q) + \varphi(q, \Sigma, Q) \leq 1$. A *Probabilistic Finite Automaton (PFA)* is a trimmed semi-PFA such that $\sum_{q \in Q} \iota(q) = 1$ and for any state q, $\tau(q) + \varphi(q, \Sigma, Q) = 1$. A *Probabilistic Deterministic Finite Automaton (PDFA)* is a PFA whose support is deterministic.

A *stochastic language* on Σ is a probability distribution over Σ^*, i.e. a function P defined from Σ^* to $[0,1]$ such that $\sum_{u \in \Sigma^*} P(u) = 1$. The function P_A associated with a PFA A (resp. a semi-PFA A) is a stochastic language (resp. satisfies $\sum_{u \in \Sigma^*} P_A(u) \leq 1$). Let us denote by $\mathcal{S}\,(\Sigma)$ the set of all stochastic languages on Σ. Let $P \in \mathcal{S}\,(\Sigma)$ and let $\mathrm{res}(P) = \{u \in \Sigma^* | P(u\Sigma^*) \neq 0\}$. Let $u \in \mathrm{res}(P)$, the *residual language* of P associated with u is the stochastic language $u^{-1}P$ defined by $u^{-1}P(w) = P(uw)/P(u\Sigma^*)$. Let $\mathrm{Res}\,(P) = \{u^{-1}P | u \in \mathrm{res}\,(P)\}$. It can be shown that $\mathrm{Res}\,(P)$ spans a finite dimensional vector space iff P can be generated by an MA. Let $\mathrm{MA}_{\mathcal{S}}$ be the set composed of MA which generate stochastic languages.

A *Probabilistic Residual Finite Automaton (PRFA)* is a PFA $A = \langle \Sigma, Q, \varphi, \iota, \tau \rangle$ whose states define residual languages of P_A, i.e. such that $\forall q \in Q$, $\exists u \in \Sigma^*$, $P_{A,q} = u^{-1}P_A$, where $P_{A,q}$ denotes the stochastic language generated by $< \Sigma, Q, \varphi, \iota_q, \tau >$ where $\iota_q(q) = 1$ (12).

Let us denote by $\mathcal{S}_{\mathrm{MA}}\,(\Sigma)$ (resp. $\mathcal{S}_{\mathrm{PFA}}\,(\Sigma)$, $\mathcal{S}_{\mathrm{PRFA}}\,(\Sigma)$, $\mathcal{S}_{\mathrm{PDFA}}\,(\Sigma)$) the set of stochastic languages generated by MA (resp. PFA, PRFA, PDFA). It has been shown that $\mathcal{S}_{\mathrm{PDFA}}\,(\Sigma) \subsetneq \mathcal{S}_{\mathrm{PRFA}}\,(\Sigma) \subsetneq \mathcal{S}_{\mathrm{PFA}}\,(\Sigma)$ (11). We show in Section 4 that $\mathcal{S}_{\mathrm{PFA}}\,(\Sigma) \subsetneq \mathcal{S}_{\mathrm{MA}}\,(\Sigma)$.

Let $R \subseteq \mathrm{MA}$. Let us denote by $R[\mathbb{Q}]$ the set of elements of R, the parameters of which are all in $\mathbb{Q}$.

2.3 Learning Stochastic languages

We are interested in learnable subsets of $\mathrm{MA}_{\mathcal{S}}$. Several learning model can be used to study inference of stochastic languages. We consider two of them.

2.3.1 Identification in the limit with probability 1.

The identification in the limit learning model of Gold (6) can be adapted to the stochastic case (7).

Let $P \in \mathcal{S}(\Sigma)$ and let S be a finite sample drawn according to P. For any $X \subseteq \Sigma^*$, let $P_S(X) = \frac{1}{\mathrm{Card}(S)} \sum_{x \in S} \mathbb{1}_{x \in X}$ be the empirical distribution associated with S. A *complete presentation* of P is an infinite sequence S of words generated according to P. We denote by S_n the sequence composed of the n first words (not necessarily different) of S and we write $P_n(X)$ instead of $P_{S_n}(X)$.

Definition 1
Let $\mathcal{R} \subset MA_{\mathcal{S}}$. $\mathcal{R}$ is said to be identifiable in the limit with probability one *if there exists a learning algorithm $\mathcal{L}$ such that for any $R \in \mathcal{R}$, with probability 1, for any complete presentation S of P_R, $\mathcal{L}$ computes for each S_n given as input, a hypothesis $R_n \in \mathcal{R}$ such that the support of R_n is stationary from some index n^* and such that $P_{R_n} \rightarrow P_R$ as $n \rightarrow \infty$. Moreover, $\mathcal{R}$ is* strongly identifiable in the limit with probability one *if P_{R_n} is also stationary from some index.*

Remark.
Unfortunately, this model is too weak as non polynomial time learning algorithms could be used: Let $\mathcal{L}'$ be an algorithm which on input S_n, runs the n first steps of $\mathcal{L}$ on each sample $S_1, \ldots, S_n$; if no $\mathcal{L}(S_i)$ terminates within n steps, $\mathcal{L}'$ outputs a default hypothesis ; otherwise, $\mathcal{L}'$ outputs R_m where m is the last index such that $\mathcal{L}(S_m)$ terminates within n steps. See (16) for an extensive study.

It has been shown that the class of PDFA is identifiable in the limit with probability one (8; 9) and that PDFA$[\mathbb{Q}]$ is strongly identifiable in the limit (10). It has been shown that the class of PRFA is identifiable in the limit with probability one and that PRFA$[\mathbb{Q}]$ is strongly identifiable in the limit (17).

We show in Section 3 that the class of PFA is identifiable in the limit with probability one and that PFA$[\mathbb{Q}]$ is strongly identifiable in the limit.

2.3.2 Learning using queries

The MAT model of Angluin (18), which allows to use *membership queries* (MQ) and *equivalence queries* (EQ) has been extended to functions computed by MA. Let P be the target function, let u be a word and let A be an MA. The answer to the query $MQ(u)$ is the value $P(u)$; the answer to the query $EQ(A)$ is YES if $P_A = P$ and NO otherwise. Functions computed by MA can be learned exactly within polynomial time provided that the learning algorithm can make extended membership queries and equivalence queries. Therefore, any stochastic language that can be computed by an MA can be learned by this algorithm.

However, using MA to represent stochastic languages involves some serious drawbacks: first, this representation is not robust, i.e. an MA may compute a stochastic language for a given set of parameters θ_0 and computes a function which is not a stochastic language for any $\theta \neq \theta_0$; moreover, we show in Section 4 that given an MA, it is undecidable whether it computes a stochastic language. That is, by using MA to represent stochastic languages, a learning algorithm relying on approximate data might infer an MA which does not compute a stochastic language and with no means to detect it. We also show that $\mathrm{MA}_{\mathcal{S}}$ contains no recursively enumerable subset sufficient to generate $\mathcal{S}_{\mathrm{MA}}(\Sigma)$.

3 Identifying $\mathcal{S}_{\mathrm{PFA}}(\Sigma)$ in the limit.

We show in this Section that the set of stochastic languages which can be generated by PFA is identifiable in the limit with probability one. Moreover, the identification is strong when the target can be generated by a PFA whose parameters are rational numbers.

3.1 Weak identification

Let P be a stochastic language over Σ, let $\mathcal{A} = (A_i)_{i \in I}$ be a family of subsets of Σ^*, let S be a finite sample drawn according to P, and let P_S be the empirical distribution associated with S. It can be shown (19; 20) that for any confidence parameter δ, with a probability greater than $1-\delta$, for any $i \in I$,

$$|P_S(A_i) - P(A_i)| \leq c\sqrt{\frac{\mathrm{VC}(\mathcal{A}) - \log\frac{\delta}{4}}{\mathrm{Card}(S)}} \quad (1)$$

where $\mathrm{VC}(\mathcal{A})$ is the dimension of Vapnik-Chervonenkis of $\mathcal{A}$ and where c is an universal constant.

When $\mathcal{A} = (\{w\})_{w \in \Sigma^*}$, $\mathrm{VC}(\mathcal{A}) = 1$. Let $\Psi(\epsilon, \delta) = \frac{c^2}{\epsilon^2}(1 - \log\frac{\delta}{4})$.

Lemma 1
Let $P \in \mathcal{S}(\Sigma^*)$ be a stochastic language and let S be a complete presentation of P. For any precision parameter ϵ, any confidence parameter δ, any $n \geq \Psi(\epsilon, \delta)$, with a probability greater than $1-\delta$, $|P_n(w) - P(w)| \leq \epsilon$ for all $w \in \Sigma^*$.

Proof. Use Inequality (1). □

For any integer k, let $Q_k = \{1, \ldots, k\}$ and let $\Theta_k = \{\iota_i, \tau_i, \varphi_{i,j}^x | i, j \in Q_k, x \in \Sigma\}$ be a set of variables. We consider the following set of constraints C_k on Θ_k:

$$C_k = \begin{cases} 0 \leq \iota_i, \tau_i, \varphi_{i,j}^x \leq 1 \text{ for any } i, j \in Q_k, x \in \Sigma \\ \sum_{i \in Q_k} \iota_i \leq 1 \\ \tau_i + \sum_{j \in Q_k, x \in \Sigma} \varphi_{i,j}^x \leq 1 \text{ for any } i \in Q_k \end{cases}.$$

Any assignment θ of these variables satisfying C_k is said to be *valid*; any valid assignement θ defines a semi-PFA A_k^θ by letting $\iota(i) = \iota_i$, $\tau(i) = \tau_i$ and $\varphi(i, x, j) = \varphi_{i,j}^x$ for any states i and j and any letter x. We simple denote by P_θ the function $P_{A_k^\theta}$ associated with A_k^θ. Let V_k be the set of valid assignments. For any $\theta \in V_k$, let θ^t be the associated trimmed assignment which set to 0 every parameter which never contributes to the probability $P_\theta(w)$ of some word w. Clearly, θ^t is valid and $P_\theta = P_{\theta^t}$.

For any word w, $P_\theta(w)$ can be seen as a function whose variables are elements of Θ_k: $P_\theta(w)$ is a polynomial and is therefore continuous. Moreover, for any valid assignment θ, $\sum_w P_\theta(w) \leq 1$. On the other hand, the series $\sum_w P_\theta(w)$ are convergent but not uniformly convergent and $P_\theta(w\Sigma^*)$ is not a continous function of θ (see Fig. 1). However, we show below that the function $(\theta, w) \to P_\theta(w)$ is uniformly continuous.

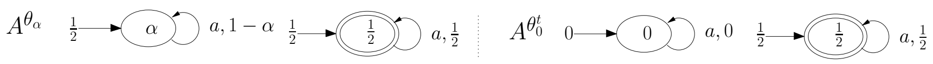

Figure 1: $P_{\theta_0}(\epsilon) = P_{\theta_0^t}(\epsilon) = 1/4$ and $P_\theta(\epsilon) = 1/4 + \alpha/2$; $P_{\theta_0}(\Sigma^*) = P_{\theta_0^t}(\Sigma^*) = 1/2$ and $P_\theta(\Sigma^*) = 1$ when $\alpha > 0$.

Proposition 1
For any integer k, the function $(\theta, w) \to P_\theta(w)$ is uniformly continuous, that is,

$$\forall \epsilon, \exists \alpha, \forall w \in \Sigma^*, \forall \theta, \theta' \in V_k, ||\theta - \theta'|| < \alpha \Rightarrow |P_\theta(w) - P_{\theta'}(w)| < \epsilon \quad .$$

Proof. We prove the proposition in several steps.

1. Let $A = \langle \Sigma, Q, \varphi, \iota, \tau \rangle$ be a semi-PFA. It can easily be shown by induction on n that for any integer n and any state $q \in Q$, $\varphi(q, \Sigma^n, Q) \leq 1$. Now, let w be a word and q' be state such that $\varphi(q, w, q') \neq 0$ and $\tau(q') \neq 0$. Then, for any integer $n > |w|$, $\varphi(q, \Sigma^n, Q) \leq 1 - \varphi(q, w, q')\tau(q')$. Proof by induction on $|w|$:

   - If $w = \varepsilon$, $q = q'$, $\tau(q) > 0$ and
   $$\varphi(q, \Sigma^n, Q) = \sum_{q_1 \in Q} \varphi(q, \Sigma, q_1)\varphi(q_1, \Sigma^{n-1}, Q) \leq \sum_{q_1 \in Q} \varphi(q, \Sigma, q_1) \leq 1 - \tau(q).$$

   - In the general case,
   $$\begin{aligned} \varphi(q, \Sigma^n, Q) &= \sum_{q_1 \neq q'} \varphi(q, \Sigma^{|w|}, q_1)\varphi(q_1, \Sigma^{n-|w|}, Q) \\ &\quad + \varphi(q, \Sigma^{|w|}, q')\varphi(q', \Sigma^{n-|w|}, Q) \\ &\leq \sum_{q_1 \neq q'} \varphi(q, \Sigma^{|w|}, q_1) + \varphi(q, \Sigma^{|w|}, q')(1 - \tau(q')) \\ &\leq \sum_{q_1} \varphi(q, \Sigma^{|w|}, q_1) - \varphi(q, w, q')\tau(q') \\ &\leq 1 - \varphi(q, w, q')\tau(q'). \end{aligned}$$

2. Let $\theta_0 \in V_k$, let $A_k^{\theta_0^t} = \langle \Sigma, Q_k, \varphi_0, \iota_0, \tau_0 \rangle$ and let $\beta_0 = Max\{\varphi_0(q, \Sigma^k, Q_k) | q \in Q_k\}$. As θ_0^t is trimmed, for any state q such that $\varphi_0(q, \Sigma^k, Q_k) > 0$, there exists a word v and a state q' such that $\varphi_0(q, v, q') \neq 0$ and $\tau_0(q') \neq 0$. As any path in A of length $\geq k$ passes through the same states at least twice, there exists a word w of length $< k$ such that $\varphi_0(q, w, q') \neq 0$. Hence, $\varphi(q, \Sigma^k, Q) < 1$ and $\beta_0 < 1$.

3. For any integer n and any state q, $\varphi_0(q, \Sigma^{nk}, Q_k) \leq \beta_0^n$. Easy proof by induction on n.

4. For any integer n, $P_{\theta_0}(\Sigma^{nk}\Sigma^*) \leq \sum_{q \in Q_k} \iota_0(q)\varphi_0(q, \Sigma^{nk}, Q_k) \leq \beta_0^n$.

5. For any state q,

$$\begin{aligned} \varphi_0(q, \Sigma^*, Q_k) &= \sum_{n \in \mathbb{N}, 0 \leq m < k} \varphi_0(q, \Sigma^{nk+m}, Q_k) \\ &\leq \sum_{n \in \mathbb{N}, 0 \leq m < k, q' \in Q_k} \varphi_0(q, \Sigma^m, q')\varphi_0(q', \Sigma^{nk}, Q_k) \\ &\leq \sum_{n \in \mathbb{N}, 0 \leq m < k, q' \in Q_k} \beta_0^n \varphi_0(q, \Sigma^m, q') \leq k/(1 - \beta_0). \end{aligned}$$

6. Let α_0 be the minimal non null parameter in θ_0^t, let $\alpha < \alpha_0/2$, let θ be a valid assignement such that $||\theta - \theta_0|| < \alpha$ and let $A_k^{\theta^t} = \langle \Sigma, Q_k, \varphi, \iota, \tau \rangle$. Note that any non null parameter in θ_0^t corresponds to a non null parameter in θ^t but that the converse is false (see Fig. 1). Let θ' be the assignment obtained from θ^t by setting to 0 every parameter which is null in θ_0^t, let $A_k^{\theta'} = \langle \Sigma, Q_k, \varphi', \iota', \tau' \rangle$ and let $\beta' = Max\{\varphi'(q, \Sigma^k, Q_k) | q \in Q_k\}$. As θ' and θ_0^t have the same set of non null parameters, there exists $\alpha_1 < \alpha_0/2$ such that $||\theta - \theta_0|| < \alpha_1$ implies $\beta' < (1 + \beta_0)/2$. Let $\beta_1 = (1 + \beta_0)/2$.

7. Let w be a word of length $\geq nk$. There are two categories of derivations of w in $A_k^{\theta^t}$:

   - those which exist in $A_k^{\theta'}$. Their contribution to $P_{\theta^t}(w)$ is not greater than β_1^n.
   - those which do not entirely exist in $A_k^{\theta'}$ and one parameter of which is $\leq \alpha_1$. Let $q_0, \ldots, q_{|w|}$ be such a derivation. Either $\iota(q) \leq \alpha_1$, either $\tau(q_{|w|}) \leq \alpha_1$, or there exists a first state q_i such that $q_0, \ldots, q_i$ is a derivation in $A_k^{\theta'}$ and $\varphi(q_i, w_i, q_{i+1}) \leq \alpha_1$, where w_i is the ith letter of w. The contribution of these derivations to $P_{\theta^t}(w)$ is bounded by

$$\sum_{q, \iota(q) \leq \alpha_1} \alpha_1 \varphi(q, w, Q) + \sum_{q, q', \iota(q') \leq \alpha_1} \iota(q)\varphi(q, w, q')\alpha_1 + \sum_{q_0, q_i \in Q_k} \iota'(q_0)\varphi'(q_0, \Sigma^*, q_i)\alpha_1 \leq \alpha_1(k + 1 + k/(1 - \beta_1)).$$

That is,

$$P_{\theta^t}(w) \leq \beta_1^n + \alpha_1(k+1+k/(1-\beta_1)).$$

8. Let $\epsilon > 0$. Let $\alpha_2 = Min(\alpha_1, \epsilon/[4(k+1+k/(1-\beta_1))])$ and let N be such that $\beta_1^N < \epsilon/4$. As for any fixed w, $P_\theta(w)$ is continuous, there exists $\alpha \leq \alpha_2$ such that $||\theta - \theta_0|| < \alpha$ implies that for any $w \in \Sigma^{\leq N}$, $|P_{\theta_0}(w) - P_\theta(w)| < \epsilon$. As, $P_{\theta_0}(w) \leq \epsilon/2$ and $P_\theta(w) \leq \epsilon/2$ when $|w| \geq N$, we conclude that for all words w, $|P_{\theta_0}(w) - P_\theta(w)| < \epsilon$.

9. We have shown that

$$\forall \epsilon, \forall \theta_0 \in V_k, \exists \alpha, \forall w \in \Sigma^*, \forall \theta \in V_k, ||\theta - \theta_0|| < \alpha \Rightarrow |P_\theta(w) - P_{\theta_0}(w)| < \epsilon.$$

Now, suppose that

$$\exists \epsilon, \forall n \in \mathbb{N}, \exists w_n \in \Sigma^*, \exists \theta_n,\ \theta'_n \in V_k \text{ s.t. },\quad ||\theta_n - \theta'_n|| < 1/n \text{ and } |P_{\theta_n}(w_n) - P_{\theta'_n}(w_n)| \geq \epsilon.$$

As valid assignments are elements of a compact set, it would exist a valid assignement θ_0 such that $\theta_{\sigma(n)} \rightarrow \theta_0$ and $\theta'_{\sigma(n)} \rightarrow \theta_0$ (for some subsequence $\sigma(n)$). We know that there exists $\alpha > 0$ such that $||\theta - \theta_0|| < \alpha$ implies that for all w, $|P_{\theta_0}(w) - P_\theta(w)| < \epsilon/2$. When $1/n < \alpha$, the hypothesis leads to a contradiction. □

Let $P \in \mathcal{S}(\Sigma^*)$ be a stochastic language and let S be a complete presentation of P. For any integers n and k and for any $\epsilon > 0$, let $I_{\Theta_k}(S_n, \epsilon)$ be the following system

$$I_{\Theta_k}(S_n, \epsilon) = C_k \cup \{|P_\theta(w) - P_n(w)| \leq \epsilon \text{ for } w \in S_n\}.$$

Lemma 2
Let $P \in \mathcal{S}(\Sigma^*)$ be a stochastic language and let S be a complete presentation of P. Suppose that there exists an integer k and a PFA $A_k^{\theta_0}$ such that $P = P_{\theta_0}$. Then, for any precision parameter ϵ, any confidence parameter δ and any $n \geq \Psi(\epsilon/2, \delta)$, with a probability greater than $1 - \delta$, $I_{\Theta_k}(S_n, \epsilon)$ has a solution that can be computed.

Proof. From Lemma 1, with a probability greater than $1 - \delta$, we have $|P_{\theta_0}(w) - P_n(w)| \leq \epsilon/2$ for all $w \in S_n$. For any $w \in S_n$, $P_\theta(w)$ is a polynomial in θ whose coefficients are all equal to 1. A bound M_w of $||\frac{dP_\theta(w)}{d\theta}||$ can easily be computed. We have

$$|P_\theta(w) - P_{\theta'}(w)| \leq M_w ||\theta - \theta'||.$$

Let $\alpha = \inf\{\frac{\epsilon}{2M_w} | w \in S_n\}$. If $||\theta - \theta'|| < \alpha$, $|P_\theta(w) - P_{\theta'}(w)| \leq \epsilon/2$ for all $w \in S_n$. So, we can compute a finite number of assignments: $\theta_1^\alpha, \ldots \theta_{N_\alpha}^\alpha$ such that for all valid assignment θ, there exists $1 \leq i \leq N_\alpha$ such that $||\theta - \theta_i^\alpha|| \leq \alpha$. Let i be such that $||\theta_0 - \theta_i^\alpha|| \leq \alpha$: θ_i^α is a solution of $I_{\Theta_k}(S_n, \epsilon)$. □

The Borel-Cantelli Lemma is often used to show that a given property holds with probability one: let $(A_n)_{n\in\mathbb{N}}$ be a sequence of events such that $\sum_{n\in\mathbb{N}} P(A_n) < \infty$; then, the probability that a finite number of A_n occur is 1.

For any integer n, let $\epsilon_n = n^{-\frac{1}{3}}$ and $\delta_n = n^{-2}$. Clearly, $\epsilon_n \to 0$ and $\sum_{n\in\mathbb{N}} \delta_n < \infty$. Moreover, there exists an integer N such that $\forall n > N$, $n \geq \psi(\epsilon_n/2, \delta_n)$.

Proposition 2
Let P be a stochastic language and let S be a complete presentation of P. Suppose that there exists an integer k and a PFA $A_k^{\theta_0}$ such that $P = P_{\theta_0}$. With probability 1 there exists an integer N such that for any $n > N$, $I_{\Theta_k}(S_n, \epsilon_n)$ has a solution θ_n and $\lim_{n\to\infty} P_{\theta_n}(w) \to P(w)$ uniformly in w.

Proof. The Borel-Cantelli Lemma entails that with probability 1 there exists an integer N such that for any $n > N$, $I_{\Theta_k}(S_n, \epsilon_n)$ has a solution θ_n. Now suppose that

$$\exists \epsilon, \forall N, \exists n \geq N, \exists w_n \in \Sigma^*, |P_{\theta_n}(w_n) - P(w_n)| \geq \epsilon.$$

Let $(\theta_{\sigma(n)})$ be a subsequence of (θ_n) such that for every integer n, $\sigma(n) \geq n$, there is $|P_{\theta_{\sigma(n)}}(w_{\sigma(n)}) - P(w_{\sigma(n)})| \geq \epsilon$ and $\theta_{\sigma(n)} \to \theta$. As each $\theta_{\sigma(n)}$ is a solution of $I_{\Theta_k}(S_{\sigma(n)}, \epsilon_{\sigma(n)})$, θ is a valid assignement such that for all w such that $P(w) \neq 0$, $P(w) = P_\theta(w)$. As P is a stochastic language, we must have $P(w) = P_\theta(w)$ for every word w, i.e. $P = P_\theta$. From Proposition 1, $P_{\theta_{\sigma(n)}}$ converges uniformy to P, which contradicts the hypothesis.

It remains to show that when the target cannot be expressed by a PFA on k states, the system $I_{\Theta_k}(S_n, \epsilon_n)$ has no solution from some index.

Proposition 3
Let P be a stochastic language and let S be a complete presentation of P. Let k be an integer such that there exist no $\theta \in V_k$ satisfying $P = P_\theta$. Then, with probability 1, there exist an integer N such that for any $n > N$, $I_{\Theta_k}(S_n, \epsilon_n)$ has no solution.

Proof.
Suppose that $\forall N \in \mathbb{N}$, $\exists n \geq N$ such that $I_{\Theta_k}(S_n, \epsilon_n)$ has a solution. Let $(n_i)_{i\in\mathbb{N}}$ be an increasing sequence such that $I_{\Theta_k}(S_{n_i}, \epsilon_{n_i})$ has a solution θ_i and let (θ_{k_i}) be a subsequence of (θ_i) that converges to a limit value $\overline{\theta}$.

Let $w \in \Sigma^*$ be such that $P(w) \neq 0$. We have

$$|P_{\overline{\theta}}(w) - P(w)| \leq |P_{\overline{\theta}}(w) - P_{\theta_i}(w)| + |P_{\theta_i}(w) - P_{n_i}(w)| + |P_{n_i}(w) - P(w)|$$

for any integer i.

With probability one, the last term converges to 0 as i tends to infinity (Lemma 1). With probability one, there exists an index i such that $w \in S_{n_i}$. From this index, the second term is less than ϵ_{n_i} which tends to 0 as i tends to infinity. Now, as $P_\theta(w)$ is a continuous function of θ, the first term tends to 0 as i tends to infinity. Therefore, $P_{\overline{\theta}}(w) = P(w)$ and $P_{\overline{\theta}} = P$, which contradicts the hypothesis.

□

Theorem 1
$\mathcal{S}_{\mathrm{PFA}}(\Sigma)$ is identifiable in the limit with probability one.

Proof. Consider the following algorithm $\mathcal{A}$:

```
Input:  A stochastic sample S_n of length n.
for k = 1 to n do
    compute α and θ_1^α, ... θ_{N_α}^α  as in Lemma 2
    if ∃1 ≤ i ≤ N_α s.t.  θ_i^α is a solution of I_{Θ_k}(S_n, ε_n) then
        return the smallest solution (in some order) A_k^{θ_i^α} and exit
    endif
endfor
return a default hypothesis; if none solution has been found
Output:  A.
```

Let P be the target and let $A_k^{\theta_0}$ be a minimal state PFA which computes P. Previous propositions prove that with probability one, from some index N, the algorithm shall output a PFA $A_k^{\theta_n}$ such that P_{θ_n} converges uniformly to P. □

3.2 Strong identification

When the target can be computed by a PFA whose parameters are in
Q, an equivalent PFA can be identified in the limit with probability 1.

In order to show a similar property for PDFA, a method based on tree of Stern-Broco was used in (10). Here we use the representation of real numbers by continuous fractions (our main reference is (21)).

Let x be a non negative real number. Define $x_0 = x$, $a_0 = \lfloor x_0 \rfloor$ and while $x_n \neq a_n$, $x_{n+1} = 1/(x_n - a_n)$ and $a_{n+1} = \lfloor x_n \rfloor$. The sequences (x_n) and (a_n) are finite iff $x \in$
Q.

Suppose from now on that $x \in$
Q, let N be the greatest index such that $x_n \neq a_n$, and for any $n \leq N$, let

$$p_n/q_n = a_0 + 1/(a_1 + 1/(\ldots(a_{n-1} + 1/a_n)\ldots))$$

where $gcd(p_n, q_n) = 1$. The fraction p_n/q_n is called the nth *convergent* of x.

Lemma 3
(21) We have $x = \frac{p_N}{q_N}$ and $\forall n < N$, $\left|x - \frac{p_n}{q_n}\right| \leq \frac{1}{q_n q_{n+1}} < \frac{1}{q_n^2}$. If a and b are two integers such that $\left|\frac{a}{b} - x\right| < \frac{1}{2b^2}$, then there is an integer $n \leq N$ such that $\frac{a}{b} = \frac{p_n}{q_n}$. For any integer A, there exists only a finite number of rational numbers $\frac{p}{q}$ such that $\left|x - \frac{p}{q}\right| \leq \frac{A}{q^2}$.

Let $x = 5/14$. We have $p_0/q_0 = 0$, $p_1/q_1 = 1/2$, $p_2/q_2 = 1/3$ and $p_3/q_3 = x$.

Lemma 4
(17) Let (ϵ_n) be a sequence of non negative real numbers which converges to 0, let $x \in \mathbb{Q}$, let (y_n) be a sequence of elements of $\mathbb{Q}$ such that $|x - y_n| \leq \epsilon_n$ for all but

finitely many n. Let $\frac{p_m^n}{q_m^n}$ the convergents associated with y_n. Then, there exists an integer N such that, for any $n \geq N$, there is an integer m such that $x = \frac{p_m^n}{q_m^n}$. Moreover, $\frac{p_m^n}{q_m^n}$ is the unique rational number such that $\left|y_n - \frac{p_m^n}{q_m^n}\right| \leq \epsilon_n \leq \frac{1}{(q_m^n)^2}$.

Example. If $y_n = \frac{1}{2} - \frac{1}{n}$ and $\epsilon_n = \frac{1}{n}$, we have $y_3 = \frac{1}{6}$, $y_4 = \frac{1}{4}$, $y_5 = \frac{3}{10}$, $y_6 = \frac{1}{3}$, $y_7 = \frac{5}{14}$. The first natural number n for which $\left|y_n - \frac{p_m^n}{q_m^n}\right| \leq \frac{1}{n} \leq \frac{1}{(q_m^n)^2}$ has a solution is $n = 4$. Let z_n be the first solution. We have $z_4 = \frac{1}{4}$, $z_5 = \frac{1}{3}$, $z_6 = \frac{1}{3}$ and $z_n = \frac{1}{2}$ after $n = 7$.

Theorem 2
$\mathcal{S}_{\text{PFA}}(\Sigma)[\mathbb{Q}]$ is strongly identifiable in the limit with probability one.

Proof.
Let θ be a valid assignment and let $\epsilon > 0$. Suppose that for every parameter α of θ, there exists integers p_α and q_α such that $|\alpha - p_\alpha/q_\alpha| \leq \epsilon \leq 1/q_\alpha^2$ and suppose that replacing each α with p_α/q_α defines a valid assignment. Then, let $\text{frac}(\theta, \epsilon)$ be such an assignment.

We slightly modify the algorithm $\mathcal{A}$ in computing $\text{frac}(\theta, \epsilon_n)$ for each assignment θ_i and in keeping a list L of all correct assignments computed during the previous steps.

Input: A stochastic sample S_n of length n, $n-1$ lists $L_1, \ldots, L_{n-1}$ of correct rational assignments computed by the algorithm at previous steps.
```
L_n ← empty list
for k = 1 to n do
    compute α and θ_1^α, ... θ_{N_α}^α as in Lemma 2
    for i = 1 to N_α do
        if θ' = frac(θ_i^α, ε_n) is a solution of I_{Θ_k}(S_n, ε_n) then
            push(L_n, θ')
        endif
    endfor
    if L_n is not empty then
        let p the smallest integer such that ∩_{i=p}^n L_i is not empty
        let A be the first element of ∩_{i=p}^n L_i (in some order); exit
    endif
A ← A_k^{θ_def} a default hypothesis;
```
Output: A and L.

Let θ_0 be a rational assignment which computes the target. There is some step n from where $\theta_0 = \text{frac}(\theta_i^\alpha, \epsilon_n)$ is in L_n. Either the algorithm identifies a previous solution, or it identifies θ_0. □

4 $\text{MA}_\mathcal{S}$ is not a suitable class of representation for learning stochastic languages.

The representation of stochastic languages by MA is not robust. Fig. 2 shows two MA which depend on parameter x. They define a stochastic language when $x = 0$ but not

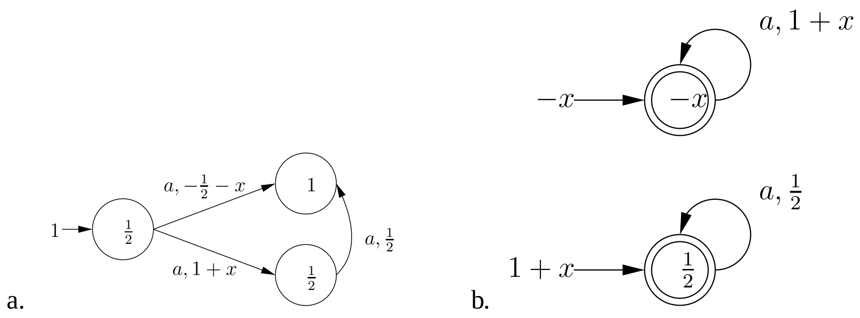

Figure 2: Two MA generating stochastic language if $x = 0$. If $x > 0$, the first generates negative values and the second unbounded values.

when $x > 0$. When $x > 0$, the first one generates negative values, and the second one generates unbounded values.

Let P be a target stochastic language and let A be the MA generating P which is output by the exact learning algorithm defined in (14). A sample S drawn according to P defines an empiric distribution P_S that could be used by some variants of this learning algorithm. In the best case, this variant is expected to output a hypothesis $\hat{A}$ having the same support as A and with approximated parameters close from those of A. But there is no garanty that $\hat{A}$ defines a stochastic language. More serious, we show below that it is undecidable whether a given MA generates a stochastic language. The conclusion is that MA representation of stochastic languages is maybe not suitable to learn stochastic languages.

4.1 Membership to $\mathrm{MA}_{\mathcal{S}}$ is undecidable

We show that the membership problem for $MA_{\mathcal{S}}$, which was left open in (1), is undecidable. We use a reduction to a decision problem about *acceptor PFA*.

An MA $\langle \Sigma, Q, \varphi, \iota, \tau \rangle$ is an *acceptor PFA* if φ, ι and τ are non negative functions, $\sum_{q \in Q} \iota(q) = 1$, $\forall q \in Q, \forall x \in \Sigma, \sum_{r \in Q} \varphi(q, x, r) = 1$ and if there exists a unique terminal state t such that $\tau(t) = 1$.

Theorem 3
(22) Given an acceptor PFA A whose parameters are in $\mathbb{Q}$ and $\lambda \in \mathbb{Q}$, it is undecidable whether there exists a word w such that $P_A(w) < \lambda$.

The following lemma shows some constructions on MA.

Lemma 5
Let A and B be two MA and let $\lambda \in \mathbb{Q}$. We can construct:

1. *an MA I_λ such that $\forall w \in \Sigma^*, P_{I_\lambda}(w) = \lambda$,*
2. *an MA $A + B$ such that $P_{A+B} = P_A + P_B$,*

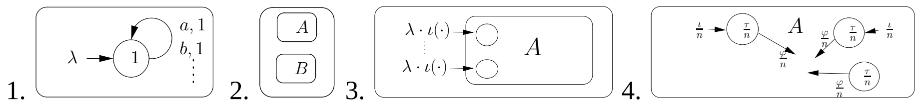

Figure 3: How to construct I_λ, $A+B$, $\lambda \cdot A$ and $\mathrm{tr}(A)$, where $n = |\Sigma| + 1$.

3. *an MA $\lambda \cdot A$ such that $P_{\lambda \cdot A} = \lambda P_A$,*

4. *an MA $\mathrm{tr}(A)$ such that for any word w, $P_{\mathrm{tr}(A)}(w) = \frac{P_A(w)}{(|\Sigma|+1)^{|w|+1}}$*

Proof. Proofs are omitted. See Fig. 3.

Note that when A is an acceptor PFA, $\mathrm{tr}(A)$ is a semi-PFA.

Lemma 6
Let $A = \langle \Sigma, Q, \varphi, \iota, \tau \rangle$ be a semi-PFA, let Q^t be the set of states $q \in Q$ such that $\varphi(Q_I, \Sigma^*, q) > 0$ and $\varphi(q, \Sigma^*, Q_T) > 0$. Let $A^t = \left\langle \Sigma, Q^t, \varphi_{|Q_t}, \iota_{|Q_t}, \tau_{|Q_t} \right\rangle$. Then, A^t is a trimmed semi-PFA such that $P_A = P_{A^t}$ and which can be constructed from A.

Proof. Straightforward.

Lemma 7
Let A be a trimmed semi-PFA, we can compute $P_A(\Sigma^*)$.

Proof. Let M be the square matrix $[\varphi(q, \Sigma, r)]_{(q,r) \in Q^2}$, T be the column vector $[\tau(q)]_{q \in Q}$ and X be the column vector $[P_{A,q}(\Sigma^*)]_{q \in Q}$. We have $X = T + MX$.

Let $k = Card(Q^t)$. Remark that $M^k = \left[\varphi(q, \Sigma^k, r)\right]_{(q,r) \in Q^2}$ and that $\sum_{r \in Q^t} \varphi(q, \Sigma^k, r) = \varphi(q, \Sigma^k, Q) < 1$ (see Prop. 1, item 1 and 2), since A is trimmed. Therefore, $\sum_{k \in \mathbb{N}} M^k$ is convergent, $(I - M)$ is inversible and $X = T(I - M)^{-1}$. Let J be the row vector $[\iota(q)]_{q \in Q}$. We have $P_A(\Sigma^*) = JX$. □

Proposition 4
It is undecidable whether an MA generates a stochastic language.

Proof. Let A be an acceptor PFA on Σ whose parameters are in $\mathbb{Q}$ and $\lambda \in \mathbb{Q}$.

For every word w, we have

$$P_{\mathrm{tr}(A - I_\lambda)}(w) = (|\Sigma| + 1)^{-(|w|+1)} (P_A(w) - \lambda) = P_{\mathrm{tr}(A)}(w) - \lambda(|\Sigma| + 1)^{-(|w|+1)}$$

and therefore $P_{\mathrm{tr}(A - I_\lambda)}(\Sigma^*) = P_{\mathrm{tr}(A)}(\Sigma^*) - \lambda$.

- If $P_{\mathrm{tr}(A)}(\Sigma^*) = \lambda$ then either $\exists w$ s.t. $P_A(w) < \lambda$ or $\forall w, P_A(w) = \lambda$. Let B be the PFA such that $P_B(w) = 1$ if $w = \epsilon$ and 0 otherwise. We have, $P_{B + \mathrm{tr}(A - I_\lambda)}(\Sigma^*) = 1$. Therefore,

  $\forall w, P_A(w) \geq \lambda$ iff $P_A(\epsilon) \geq \lambda$ and $B + \mathrm{tr}(A - I_\lambda)$ generates a stochastic language.

- If $P_{\mathrm{tr}(A)}(\Sigma^*) \neq \lambda$, let

$$B = \left|P_{\mathrm{tr}(A)}(\Sigma^*) - \lambda\right|^{-1} \cdot \mathrm{tr}\,(A - I_\lambda)\,.$$

  Check that B is computable from A, that $P_B(\Sigma^*) = 1$ and that

$$P_B(w) = \left|P_{\mathrm{tr}(A)}(\Sigma^*) - \lambda\right|^{-1} \left(\mathrm{Card}\,(\Sigma + 1)^{|w|+1}\right)^{-1} (P_A(w) - \lambda)\,.$$

  So,

$$\exists w \in \Sigma^*, P_A(w) < \lambda \text{ iff } B \text{ does not generate a stochastic language.}$$

In both cases, we see that deciding whether an MA generates a stochastic language would solve the decision problem on PFA acceptors. □

Remark that in fact, we have proved a stronger result: it is undecidable whether a multiplicity automaton $A \in MA[\mathbb{Q}]$ such that $\sum_{w\in\Sigma^*} P_A(w) = 1$ generates a stochastic langage.

This negative result is not sufficient yet to give up MA. It could be possible that $\mathrm{MA}_{\mathcal{S}}$ contains a recursively enumerable subset sufficient to generate $\mathcal{S}_{\mathrm{MA}}(\Sigma)$. We show in next section that such a subset does not exist.

4.2 MAs which generate stochastic languages cannot be enumerated

We show that the set $\mathrm{MA}_{\mathcal{S}}[\mathbb{Q}]$ composed of multiplicity automata whose coefficients are in $\mathbb{Q}$ and which generate stochastic languages is not recursively enumerable.

Theorem 4
$MA_{\mathcal{S}}[\mathbb{Q}]$ *is not recursively enumerable.*

Proof. The proof uses a technical result which can be found in the document `http://www.cmi.univ-mrs.fr/~esposito/pub/cap04VL.pdf`: given an MA A with rational coefficients, it is decidable whether $\sum_k P_A(\Sigma^k)$ converges and if the answer is yes, the sum $P_A(\Sigma^*)$ can be computed. This result generalizes Lemma 7.

Clearly,

$$\mathcal{A} = \{A \in \mathrm{MA}[\mathbb{Q}] | P_A(\Sigma^*) = 1\}$$

and

$$\mathcal{B} = \{A \in \mathcal{A} | \exists\, w \in \Sigma^* P_A(w) < 0\}$$

can be enumerated. Therefore, as $\mathrm{MA}_{\mathcal{S}}[\mathbb{Q}] = \mathcal{A} \setminus \mathcal{B}$, if $\mathrm{MA}_{\mathcal{S}}[\mathbb{Q}]$ was recursively enumerable, then $\mathrm{MA}_{\mathcal{S}}[\mathbb{Q}]$ would be recursive, which is false. □

Corollary 1
$\mathrm{MA}_{\mathcal{S}}[\mathbb{Q}]$ *contains no recursively enumerable subset sufficient to generate* $\mathcal{S}_{\mathrm{MA}}(\Sigma)$.

Proof. Given two MA A and B, it is possible to decide whether $P_A = P_B$: it is shown in (1) that $P_A = P_B$ iff $P_A(w) = P_B(w)$ for all words w of length $< |Q_A| + |Q_B|$ where Q_A and Q_B are the set of states of A and B.

Suppose that there exists an enumerable subset $\mathcal{R}$ of $MA_{\mathcal{S}}[\mathbb{Q}]$ sufficient to generate $\mathcal{S}_{\text{MA}}(\Sigma)$. Then, we could enumerate $\mathcal{R}$ and $\text{MA}[\mathbb{Q}]$ in parallel and test whether elements of the second set are equivalent to at least one lement of the first set. This procedure yields an enumeration of $\text{MA}_{\mathcal{S}}[\mathbb{Q}]$. $\square$

Corollary 2
$\mathcal{S}_{\text{PFA}}(\Sigma) \subsetneq \mathcal{S}_{\text{MA}}(\Sigma)$.

Proof. Straightforward since $PFA[\mathbb{Q}]$ is recursively enumerable. So, $\mathcal{S}_{\text{MA}}(\Sigma) \neq \mathcal{S}_{\text{PFA}}(\Sigma)$. $\square$

5 Conclusion

We have shown that PFA are identifiable in the limit with probability one. However, our learning algorithm is far from being efficient while algorithms that identifies PDFA or PRFA in the limit can also be used in practical learning situations (ALERGIA, RLIPS (8; 9), MDI (10)); work in progress for PRFA. We do not have model that describe algorithms "that can be used in practical cases": identification in the limit model is clearly too weak, exact learning via queries is irrealistic, PAC-model is maybe too strong (PDFA are not PAC-learnable (23)). Identifiability in the limit of PFA can be interpreted as: there are no information-theoretic properties which forbid to look for subclasses of PFA, as rich as possible and having good empirical learnability properties.

On the other hand, we have shown that representing stochastic languages by using Multiplicity Automata presents some serious drawbacks. The subclass of stochastic languages which has one of the simplest characterization (the residual languages must span a finite dimensional vector space) yields to a very complicated subset of MA. We feel that this representation scheme is not very suitable to represent stochastic languages if the goal is to learn them from stochastic data.

References

Paz, A.: Introduction to probabilistic automata. Academic Press, London (1971)

Abe, N., Warmuth, M.: On the computational complexity of approximating distributions by probabilistic automata. Machine Learning **9** (1992) 205–260

Dempster, A., Laird, N.M., Rubin, D.B.: Maximum likelyhood from incomplete data via the em algorithm. Journal of the Royal Statistical Society **39** (1977) 1–38

Baldi, P., Brunak, S.: Bioinformatics: The Machine Learning Approach. MIT Press (1998)

Freitag, D., McCallum, A.: Information extraction with HMM structures learned by stochastic optimization. In: AAAI/IAAI. (2000) 584–589

Gold, E.: Language identification in the limit. Inform. Control **10** (1967) 447–474

Angluin, D.: Identifying languages from stochastic examples. Technical Report YALEU/DCS/RR-614, Yale University, New Haven, CT (1988)

Carrasco, R., Oncina, J.: Learning stochastic regular grammars by means of a state merging method. In: International Conference on Grammatical Inference, Heidelberg, Springer-Verlag (1994) 139–152

Carrasco, R.C., Oncina, J.: Learning deterministic regular grammars from stochastic samples in polynomial time. RAIRO (Theoretical Informatics and Applications) **33** (1999) 1–20

de la Higuera, C., Thollard, F.: Identification in the limit with probability one of stochastic deterministic finite automata. Volume 1891 of Lecture Notes in Artificial Intelligence., Springer (2000) 141–156

Denis, F., Esposito, Y.: Residual languages and probabilistic automata. In: 30th International Colloquium, ICALP 2003. Number 2719 in LNCS, SV (2003) 452–463

Esposito, Y., Lemay, A., Denis, F., Dupont, P.: Learning probabilistic residual finite state automata. In: ICGI'2002, 6th International Colloquium on Grammatical Inference. LNAI, Springer Verlag (2002)

Bergadano, F., Varricchio, S.: Learning behaviors of automata from multiplicity and equivalence queries. In: Italian Conference on Algorithms and Complexity. (1994)

Beimel, A., Bergadano, F., Bshouty, N.H., Kushilevitz, E., Varricchio, S.: On the applications of multiplicity automata in learning. In: IEEE Symposium on Foundations of Computer Science. (1996) 349–358

Beimel, A., Bergadano, F., Bshouty, N.H., Kushilevitz, E., Varricchio, S.: Learning functions represented as multiplicity automata. Journal of the ACM **47** (2000) 506–530

Pitt, L.: Inductive Inference, DFAs, and Computational Complexity. In: Proceedings of AII-89 Workshop on Analogical andInductive Inference; Lecture Notes in Artificial Intelligence 397, Heidelberg, Springer-Verlag (1989) 18–44

Denis, F., Esposito, Y.: Identification à la limite d'automates probabilistes avec probabilité de 1. In: CAP 2003, Presses Universitaires de Grenoble (2003) 249–264

Angluin, D.: Queries and concept learning. Machine Learning **2** (1988) 319–342

Vapnik, V.N.: Statistical Learning Theory. John Wiley (1998)

Lugosi, G.: Pattern classification and learning theory. In: Principles of Nonparametric Learning. Springer (2002) 1–56

Hardy, G.H., Wright, E.M.: An introduction to the theory of numbers. Oxford University Press (1979)

Blondel, V., Canterini, V.: Undecidable problems for probabilistic automata of fixed dimension (2001)

Kearns, M., Mansour, Y., Ron, D., Rubinfeld, R., Schapire, R.E., Sellie, L.: On the learnability of discrete distributions. (1994) 273–282

Blondel, V.D., Tsitsiklis, J.N.: A survey of computational complexity results in systems and control. Automatica **36** (2000) 1249–1274