

STATISTICAL CHARACTERIZATION OF LEARNABLE SEQUENCES

P.H. Fuchs

Fachbereich Mathematik, Universität Frankfurt a.M.

Throughout this paper X denotes the binary alphabet $\{0,1\}$, X^* the set of (finite) words over X , X^∞ the set of infinite binary sequences; Λ denotes the empty word and $|x|$ denotes the length of $x \in X^*$.

First we present some informal considerations concerning learnability of infinite 0-1-sequences. Let $R \subseteq X^* \times X^*$ be a rec. en. relation (an effective system of descriptions; sometimes called operator); we say y describes x iff yRx . We suppose that R is an universal rec. en. relation, i.e. if S is any rec. en. relation, there is $w \in X^*$ such that $\forall y, x: ySx \Rightarrow wyRx$. Now, if $z \in X^\infty$ is a recursive sequence, there is a "best" way to describe z : there is $v \in X^*$ such that $\forall n \in \mathbb{N}: vRz^n$ (where z^n denotes the initial segment $z_1 \dots z_n$ of z). But usually (if z is not recursive) the length of the description of z^n will become larger as n grows, because more and more additional information must be provided by the description. We call z learnable, if there is a "best" way to describe all the initial segments z^n of z . The problem to characterize learnable sequences is due to R.P. Daley [1].

Let us pass to the formal definitions. Following Levin [3], we restrict the systems of descriptions to monotonic operators. A rec. en. relation $A \subseteq X^* \times X^*$ is called monotonic operator iff

$$(1) \quad \forall (y,x) \in A: \forall (v,u) \in A: y \sqsubseteq v \Rightarrow x \sqsubseteq u \vee u \sqsubseteq x.$$

(We write $y \sqsubseteq v$ iff $v \in yX^*$)

There is an universal monotonic operator U ; that means: if A is any monotonic operator, then $w \in X^*$ can effectively be found such that $\forall (y,x) \in A: (wy,x) \in U$. We fix an universal monotonic operator U , a partial recursive function $\varphi: X^* \times X^* \rightarrow X$ such that $\text{domain}(\varphi) = U$ and a running time function Φ for φ . For technical reasons we assume

$$(2) \quad \forall y, x: \Phi(y, x) \geq \log |y|.$$

Notations:

$$(3) \quad K_m(x) := \min \{ |y| \mid y \in X^* \wedge yUx \}; K_m \text{ is called } \underline{\text{monotonic operator complexity}}.$$

(4) Let T be a growth function (i.e. T is recursive, isotonic, unbounded).

$$M_x^T := \{y \in X^* \mid \forall i \leq |x| : \exists k \leq |y| : y^k \cup x^i \wedge \phi(y^k, x^i) \leq T(i)\}.$$

Condition (2) ensures that there is a finite set $\tilde{M}_x^T \subseteq X^*$ such that $M_x^T = \tilde{M}_x^T \cup X^*$ and $y \in \tilde{M}_x^T \Rightarrow \log |y| \leq T(|x|)$.

(5) $Km^T(x) := \min \{|y| \mid y \in M_x^T\}$.

Km^T is recursive for sufficiently large functions T .

In order to give a short definition of learnability we introduce a relation \leq^* for (arbitrary) functions $f_1, f_2: \mathbb{N} \rightarrow \mathbb{N}$. We write $f_1 \leq^* f_2$ iff $\forall g$ growth funct.: $\forall^\infty n: f_1(n) \leq f_2(n) + g(n)$.

Definition:

$z \in X^\infty$ is called learnable iff there is a growth function T such that the function $n \rightarrow Km^T(z^n)$ is \leq^* -minimal among $(n \rightarrow Km^t(z^n) \mid t \text{ growth funct.})$.

An equivalent, but perhaps more intuitive definition of learnability makes use of the notion of a recursive coding. A recursive function $\psi: X^* \rightarrow X^*$ is called a recursive coding, iff $\forall x \in X^*: \psi(x) \cup x$. Thus $\psi(x)$ describes x ; and $z \in X^\infty$ is learnable iff there is a recursive coding ψ such that $n \rightarrow |\psi(z^n)|$ is \leq^* -minimal among $(n \rightarrow |\psi'(z^n)| \mid \psi' \text{ rec. coding})$.

There is a rec. en. sequence z which is not learnable [2], [4]. Moreover this sequence z can be chosen such that it satisfies the following i.o. cut-down property.

For every growth function T there exists a growth function T' such that $\forall q < 1: \exists^\infty n: Km^{T'}(z^n) \leq Km^T(z^n) - q \cdot n$.

The attempt to characterize learnable sequences statistically arises from the following idea. Consider random sequences with respect to the equiprobability distribution $\bar{\mu}$. These sequences are completely irregular; we suppose that the best way to describe their initial segments is to describe them by themselves. Recursive sequences are learnable but not at all μ -random. Therefore we first generalize the concept of randomness in order to cover both cases.

Let $p: X^* \rightarrow]0, 1[\cap \mathbb{Q}$ be a recursive function. p defines a recursive probability measure on X^∞ (i.e. on the σ -algebra generated by the sets xX^∞) as follows.

$$\bar{p}(xX^\infty) := \prod_{x_{i+1}=1} p(x^i) \cdot \prod_{x_{i+1}=0} (1-p(x^i)).$$

We call p a recursive probability measure (r.p.m.), too. μ denotes the equiprobability distribution $\forall x \in X^*: \mu(x) = 2^{-|x|}$.

A recursive p-martingale is a recursive function $V: X^* \rightarrow Q_+$ satisfying $\forall x \in X^*: V(x) = p(x)V(x1) + (1-p(x))V(x0)$. In analogy to the μ -case [5] we define p -randomness.

Definition:

Let p be a r.p.m.. $z \in X^\infty$ is not p-random iff $\exists V$ rec. p -martingale: $\exists h$ growth function: $\exists^\infty n: V(z^n) \geq h(n)$.

We call $\mathcal{K} := \{z \in X^\infty \mid \exists p \text{ r.p.m.: } z \text{ is } p\text{-random}\}$ the set of general random sequences.

Let z be a recursive sequence. We define

$$p(z^{n-1}) = \begin{cases} (2^n+1) \cdot (2^{n+2})^{-1} & , \text{ if } z_n=1 \\ (2^n+2)^{-1} & , \text{ if } z_n=0 \end{cases}$$

It is easily seen that z is p -random. Thus, every recursive sequence is a general random sequence. Martingales and monotonic operator complexity are strongly related as the following theorem shows.

Theorem 1.

(a) Let T be a growth function. There is a recursive μ -martingale $V > 0$ such that $\forall x \in X^*: \log V(x) \geq |x| - Km^T(x)$.

(b) Let V be a recursive μ -martingale, $V > 0$. There is a growth function T and a constant c such that $\forall x \in X^*: \log V(x) \leq |x| - Km^T(x) + c$.

Proof:

(a) Define $\bar{V}(x) := \bar{\mu}(M_x^T X^\infty) \cdot 2^{|x|}$. From condition (2) we conclude that \bar{V} is recursive. \bar{V} is a recursive μ -submartingale, i.e. $\forall x \in X^*: \bar{V}(x) \geq 2^{-1} \cdot (\bar{V}(x0) + \bar{V}(x1))$. Therefore we can construct a recursive μ -martingale $V \geq \bar{V}$. V satisfies (a): from $y \in M_x^T$, $|y| = Km^T(x)$ we conclude $\bar{\mu}(M_x^T X^\infty) \geq 2^{-|y|}$ and this implies $\log V(x) \geq |x| - Km^T(x)$.

(b) We define a r.p.m. p as follows: $p(x) := V(x1) / 2 \cdot V(x)$. It suffices to find a growth function T such that $Km^T(x) \leq |\log \bar{p}(x)| + c$, for we have $V(x) = 2^{|x|} \bar{p}(x) (V(\Lambda))^{-1}$.

First we define a mapping $F: X^* \rightarrow X^*$ using the following notations. For $i=1,2,3,\dots$ we divide the interval $[0,1]$ by 2^{i-1} equidistant points into 2^i open intervals of length 2^{-i} ; these intervals are denoted by I_k^i ($k=1,\dots,2^i$). To every interval of this kind we associate a finite binary sequence by the function B in the following way:

$B(I_k^i)$ = the (lexicographically) k -th sequence in X^i .

Fact 1: $I_k^i \supseteq I_r^i \iff B(I_k^i) \subset B(I_r^i)$.

To define F at $x \in X^*$ we consider the interval $I_x = \left[\sum_{y < x} \bar{p}(y), \sum_{y \leq x} \bar{p}(y) \right]$,
 $y \in X^n$ $y \in X^n$

where $x < y$ denotes the lexicographical ordering and $\bar{p}(y) := \bar{p}(yX^\infty)$.

Let I_k^i be an interval of maximal length contained in I_x , i.e. $I_k^i \subset I_x \wedge \forall j < i: \forall r (0 < r \leq 2^j): I_r^j \not\subset I_x$. There is one and only one such I_k^i . Then we set $F(x) := B(I_k^i)$.

Fact 2: $\forall x: |F(x)| \leq \lceil -\log \bar{p}(x) \rceil + 1$.

The length of I_x is $\bar{p}(x)$; therefore $i \leq \lceil -\log \bar{p}(x) \rceil + 1$.

Fact 3: $F: X^* \rightarrow X^*$ is recursive.

Fact 4: F^{-1} is a monotonic operator.

Let $F(x) \subset F(x')$. Applying fact 1 we conclude for the intervals $B^{-1}(F(x))$ and $B^{-1}(F(x'))$: $B^{-1}(F(x)) \supseteq B^{-1}(F(x'))$. Then $x' \supset x$ must hold, for $x' \not\supset x \wedge x \not\supset x'$ would imply $I_x \cap I_{x'} = \emptyset$.

Choose $v \in X^*$ such that $\forall x, y \in X^*: y = F(x) \Rightarrow vyUx$. Define $T(n) := \max \{ \phi(vF(x), x) \mid x \in X^n \}$. Fact 3 implies that T is recursive. From fact 2 we conclude $\forall x \in X^*: Km^T(x) \leq |F(x)| + |v| \leq \lceil -\log \bar{p}(x) \rceil + 1 + |v|$.

Theorem 1 is a counterpart of Levin's theorem 2 [3]; in [4] theorem 1 is used to characterize p -randomness in terms of Km .

Theorem 2. (Statistical characterization of learnable sequences)

For every $z \in X^\infty$ the following conditions are equivalent.

- (a) z is learnable
- (b) z is a general random sequence.

Proof:

"(a) \Rightarrow (b)": Let T be a growth function which learns z . Apply thm.1(a) to T . We find a recursive μ -martingale $V > 0$ s.th. $\log V(x) \geq |x| - Km^T(x)$. Define $p(x) := V(x1) / 2 \cdot V(x)$. p is a r.p.m.. We claim that z is p -random.

Assume z is not p -random. Then there is a recursive p -martingale \bar{V} and a growth function g s.th. $\exists^\infty n: \bar{V}(z^n) \geq g(n)$. We define a recursive μ -martingale $V'(x) := \bar{V}(x) \cdot \bar{p}(x) \cdot 2^{|x|}$. From thm.1(b) we get a growth function T' such that $\log V'(x) \leq |x| - Km^{T'}(x) + c$. Using the identity $\bar{p}(x) = (V(\Lambda))^{-1} \cdot V(x) \cdot 2^{-|x|}$ we obtain $\log \bar{V}(x) + Km^{T'}(x) \leq Km^T(x) + c + \log V(\Lambda)$; and this implies $\exists^\infty n: \log g(n) - c' + Km^{T'}(z^n) \leq Km^T(z^n)$, which is contradictory to the supposition that T learns z .

"(b) \Rightarrow (a)": Let z be p -random. Consider the μ -martingale $V(x) = 2^{|x|} \bar{p}(x)$ and apply thm.1(b). We obtain T and c such that $0 \leq |\log \bar{p}(x)| - Km^T(x) + c$. We show that T learns z . If not, there would be T' and a growth function g such that $\exists^\infty n: Km^{T'}(z^n) + g(n) \leq Km^T(z^n)$. Applying thm.1(a) to T' we obtain a recursive μ -martingale V' such that $\exists^\infty n: g(n) \leq \log V'(z^n) + |\log \bar{p}(z^n)| - n + c$. But this is contradictory to the p -randomness of z , for $\bar{V}(x) := 2^{-|x|} \cdot V'(x) \cdot (\bar{p}(x))^{-1}$ is a recursive p -martingale.

A learnable sequence z may have many algorithmically recognizable regularities; but if T learns z , then all these regularities are found within running time T . Beyond that, z must be quite irregular. The characterization theorem shows that all regularities of a learnable sequence can be completely condensed into a r.p.m. p .

The correspondence between growth functions and r.p.m.'s is effective in the sense that given T (which learns z), one can effectively find a r.p.m. p such that z is p -random and vice versa.

The set of non-learnable sequences, i.e. $X^\infty \setminus \mathcal{K}$ is rather small. In fact, we have $\forall p \text{ r.p.m.} : \bar{p}(X^\infty \setminus \mathcal{K}) = 0$.

References:

- [1] R.P. Daley THE PROCESS COMPLEXITY AND THE UNDERSTANDING OF SEQUENCES. Proceedings of Symposium and Summer School MFCS, High Tatras, September 3-8, 1973.
- [2] P.H. Fuchs PROGRAMMKOMPLEXITÄT, ZUFÄLLIGKEIT, LERNBARKEIT. Dissertation, Frankfurt 1975.
- [3] L.A. Levin ON THE NOTION OF A RANDOM SEQUENCE. Soviet Math. Dokl. 14, 1973.
- [4] C.P. Schnorr THE COINCIDENCE OF LEARNABLE, OF OPTIMALLY COMPRESSIBLE AND OF GENERAL RANDOM SEQUENCES. Preprint, Fachbereich Mathematik, Universität Frankfurt 1975.
- [5] C.P. Schnorr ZUFÄLLIGKEIT UND WAHRSCHEINLICHKEIT. Lecture Notes in Math., vol 218, Berlin and New York, 1971.