Acta Cybernetica, Vol. 11, No. 3, Szeged, 1994

Measure of Infinitary Codes

Nguyen Huong Lam * Do Long Van *

Abstract

An attempt to define a measure on the set A^N of infinite words over an alphabet A starting from any Bernoulli distribution on A is proposed. With respect to this measure, any recognizable (in the sense of Büchi-McNaughton) language is measurable and the Kraft-McMillan inequality holds for measurable infinitary codes. Nevertheless, we face some "anomalies" in contrast with ordinary codes.

1 Introduction

In this paper we need only very basic concepts and facts from the formal language theory and the theory of codes, for which we always refer to [Ei] and [Be-Pe]. Let A be a finite or countable alphabet and A^* be the set of (finite) words on A (that is A^* is the free monoid with base A) with the empty word (the unit of A^*) denoted by ϵ. The set of nonempty words is denoted by $A^+ = A^* - \epsilon$. The product of two words u and v is the concatenation uv of them.

A *factorization* of a word w on a given subset X of A^* is a sequence $u_1, \ldots, u_n$ of words of X such that $w = u_1 \ldots u_n$. A subset X of A^* is a *code* if every word of A^* has at most one factorization on X.

Intuitively, a code may not contain too many words and this idea has been stated mathematically in the remarkable Kraft-McMillan inequality. Let us mention it now.

A *Bernoulli distribution* on A is a function

$$p : A \to R_+$$

associating with each *letter* a nonnegative real number such that

$$\sum_{a \in A} p(a) = 1.$$

A distribution p is *positive* if $p(a) > 0$ for all $a \in A$. We extend p in a natural way to a word $u = a_1 \ldots a_n$ of A^* ($a_1, \ldots, a_n$ are letters) by

$$p(u) = \prod_{i=1}^{n} p(a_i)$$

*Institute of Mathematics, P. O. Box 631, 10 000 Hanoi, Vietnam.

and then to a subset X of A^* by

$$p(X) = \sum_{u \in X} p(u).$$

The value $p(X)$ is called the *measure* of X, which may be finite or infinite. If finite, the measure is the sum of an absolutely convergent numerical series, so the order of summation is not important and the definition is correct.

The well-known in the information theory Kraft-McMillan inequality ([Mc] or [Be-Pe]) says that:

For any Bernoulli distribution, the measure of any code does not exceed 1.

The presentation that follows is an attempt to resolve a question, quite natural, in the mainstream of extensive studies on infinite words: how can one define a measure (in some sense) on the set of infinite words A^N so that this measure should be well compatible with the measure structure and properties of languages in A^*? Besides, we want this measure to satisfy our own demand: to prove something like the Kraft-McMillan inequality for infinitary codes, introduced in [Va]. To do this we come to the theory of measure, making use of its very basic concepts (Lebesgue extension of measures, infinite product of probability spaces) and we also exploit some techniques suggested by [Sm].

2 Measure Theory

2.1 Basic

We give a brief survey of facts for furthergoing treatment. For more details the reader is referred to [Ha]. Let X be any fixed set; we always deal with subsets of X, so in the sequel sets always mean subsets of this "base" set. Also we use the Euler fraktur alphabet to indicate classes (collections) of sets, for example, $\mathfrak{P}(X)$ is the class of all subsets of X (the power set). A class $\mathfrak{R}$ is called a (Boolean) *ring* of sets provided for any $E, F \in \mathfrak{R}$ the set-theoretic difference $E - F$ and union $E \cup F$ are also in $\mathfrak{R}$. A ring is called *σ-ring* if $\mathfrak{R}$ is closed under the formation of countable unions, i.e., $\cup_{i=1}^{\infty} E_i$ is in $\mathfrak{R}$ for any countable sequence of sets $E_1, E_2, \ldots$ of $\mathfrak{R}$. A ring (σ-ring) containing the base set X, is said to be an *algebra* (a *σ-algebra* resp.). Since $E \cap F = E \cup F - ((E - F) \cup (F - E))$ and $\cap_{i=1}^{\infty} E_i = X - \cup_{i=1}^{\infty}(X - E_i)$, we see that a ring is also closed under the formation of finite, and moreover if it is a σ-algebra, of countable intersections. Since the intersection of any number of rings (σ-rings) is also a ring (σ-ring), for any class $\mathfrak{C}$ there exists the smallest ring (σ-ring) containing it, which is called the ring (σ-ring) *generated* by $\mathfrak{C}$ and denoted by $R(\mathfrak{C})$ ($S(\mathfrak{C})$ resp.). We say that $\mathfrak{C}$ is a *hereditary class* if for every $E \in \mathfrak{C}$, $F \subseteq E$ implies $F \in \mathfrak{C}$. Clearly, the hereditarity of classes is preserved under any intersection therefore we can say of the smallest hereditary class $H(\mathfrak{C})$ containing a given class $\mathfrak{C}$.

Let $\mathfrak{C}$ be any class of sets. A *set function* on $\mathfrak{C}$ is a mapping

$$f : \mathfrak{C} \to R_+ \cup \infty$$

defined on $\mathfrak{C}$, taking real nonnegative values including infinity. A set function f is called

— *additive*, if for any disjoint sets E_1, E_2 of $\mathfrak{C}$ such that $E_1 \cup E_2 \in \mathfrak{C}$

$$f(E_1 \cup E_2) = f(E_1) + f(E_2);$$

— *countably additive*, or *σ-additive*, if for any countable sequence of mutually disjioint sets $E_1, E_2, \ldots$ of $\mathfrak{C}$ such that $\cup_{i=1}^{\infty} E_i \in \mathfrak{C}$

$$f(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} f(E_i).$$

A σ-additive set function μ on a ring $\mathfrak{R}$ is said to be a *measure* (on $\mathfrak{R}$). The value $\mu(E)$ is the *measure* of E. A measure μ is *finite* if every E of $\mathfrak{R}$ has finite measure and is *σ-finite* if every E of $\mathfrak{R}$ is a countable union of sets of $\mathfrak{R}$, all of them having finite measure.

2.2 Lebesgue Extension of Measures

Let μ_1, μ_2 be measures respectively on the rings $\mathfrak{R}_1$ and $\mathfrak{R}_2$ with $\mathfrak{R}_1 \subseteq \mathfrak{R}_2$, then μ_2 is an *extension* of μ_1 if restricted to $\mathfrak{R}_1$, μ_2 is equal to μ_1.

Provided the σ-additivity of the measure μ on some ring $\mathfrak{R}$, we can extend it considerably further to a σ-ring which is in some sense maximal as follows.

Let $H(\mathfrak{R})$ be the smallest hereditary σ-ring containing $\mathfrak{R}$. For any set $E \in H(\mathfrak{R})$, we define the *outer measure* of E

$$\mu^*(E) = \inf\{\sum_{i=1}^{\infty} \mu(E_i) : E \subseteq \bigcup_{i=1}^{\infty} E_i, E_i \in \mathfrak{R} \}.$$

Indeed, $\mu^*(E) = \mu(E)$ for $E \in \mathfrak{R}$. Following [Ko-Fo], a set $E \in H(\mathfrak{R})$ is called *measurable* if for any $\epsilon > 0$ there exist $E_0 \in \mathfrak{R}$ such that

$$\mu^*(E \triangle E_0) < \epsilon,$$

where $E \triangle E_0 = (E - E_0) \cup (E - E_0)$ is the symmetric difference of E and F.

It is proved that the class $\mathfrak{M}$ of all measurable sets is a σ-ring and the function μ^* is σ-additive on it and $S(\mathfrak{R}) \subseteq \mathfrak{M}$ [Ko-Fo].

Thus the measure μ on $\mathfrak{R}$ has been extended to the measure μ^* on the σ-ring $S(\mathfrak{R})$ generated by $\mathfrak{R}$ and certainly $\mu^*(E) = \mu(E)$ when $E \in \mathfrak{R}$. Usually, the triple $(X, \mathfrak{M}, \mu)$ consisting of the base set X, a σ-ring $\mathfrak{M}$ of subsets of X and a measure μ on $\mathfrak{M}$ is called a *measure space*; when $X \in \mathfrak{M}$ and $\mu(X) = 1$ the measure space is called a *probability space*.

We now make a remark that will be useful in the sequel. Sometimes, the starting point is not the ring $\mathfrak{R}$ itself, but some subclass $\mathfrak{S}$ such that it can generates $\mathfrak{R}$ and the latter is easily constructed from $\mathfrak{S}$. An example of such classes are semirings, considered in [Ko-Fo]: a class $\mathfrak{S}$ is a *semiring* provided, first, it is closed under the formation of finite intersections and, second, if $E, F \in \mathfrak{S}, E \subseteq F$ then F splits into a finite number of mutually disjoint subsets $E_0, E_1, \ldots, E_n$ of $\mathfrak{S}$ such that $E = E_0$: $F = \bigcup_{i=0}^{n} E_i$. If $\mathfrak{S}$ is a semiring, $R(\mathfrak{S})$ is then the class of all finite unions of subsets of $\mathfrak{S}$. It is easy to see also that if μ is σ-additive on $\mathfrak{S}$, so is in $R(\mathfrak{S})$.

2.3 Infinite Product Measure

Another fundamental construction we need here is the infinite product measure. More specifically, we treate only the countable product.

Let $(X_i, \mathfrak{M}_i, \mu_i), i = 1, 2, \dots$ be a countable collection of probability spaces, i.e. measure spaces with $X_i \in \mathfrak{M}_i$ and $\mu_i(X_i) = 1$. Further, let $X = \prod_{i=1}^{\infty} X_i$ be the set-theoretic Cartesian product of the sets $X_1, X_2, \dots$. A subset A of X of the form

$$A = \prod_{i=1}^{\infty} A_i, \qquad A_i \in \mathfrak{M}_i$$

and $A_i = X_i$ for almost all i, is called a *measurable rectangle*. The class of measurable rectangles is obviously a semiring and is denoted by $\mathfrak{A}$. Let us denote $\mathfrak{M} = S(\mathfrak{A})$ the σ-ring generated by the measurable rectangles. Theorem 2 of [Ha, Chapter VII, §38] states, in fact, that there exists uniquely a measure μ on $\mathfrak{M}$ such that if

$$A = A_1 \times \dots \times A_n \times X_{n+1} \times X_{n+2} \times \dots$$

is a measurable rectangle then

$$\mu(A) = \mu_1(A_1) \dots \mu_n(A_n).$$

Since $\mu_i(X_i) = 1$ for all i, μ is well-defined on $\mathfrak{A}$ and $\mu(X) = 1$. Therefore, the triple $(X, \mathfrak{M}, \mu)$ is a probability space that is called the *product measure space* of spaces $(X_i, \mathfrak{M}_i, \mu_i)$ and the measure μ on $\mathfrak{M}$ is then called the *product measure* of measures μ_i.

This construction ensures the existence of a measure on the set of infinite words, which we shall consider in the next section.

3 Measure on A^N

An infinite word α on the alphabet A is an infinite sequense of letters indexed by natural numbers

$$\alpha = a_1 a_2 \dots.$$

The set of all infinite words on A is denoted by A^N. We consider also the set $A^\infty = A^* \cup A^N$, on which we define the monoid structure as follows [Va]: for $\alpha, \beta \in A^\infty$, if $\alpha \in A^*$ then the product $\alpha \cdot \beta$ is the concatenation $\alpha\beta$ of α and β; otherwise, if $\alpha \in A^N, \alpha \cdot \beta$ is defined to be α. Naturally, the product of words can be extended for *languages*, i.e. subsets of A^∞: $XY = \{\alpha \cdot \beta | \alpha \in X \subseteq A^\infty, \beta \in Y \subseteq A^\infty\}$. Not to be too strict, in the folowing, we omit the dot in the product of words and when a set is a singleton we frequently identify it with its element.

Let now p be any Bernoulli distribution on A, as before extended to A^*; then $(A, \mathfrak{P}(A), p)$ actually forms a probability space, where $\mathfrak{P}(A)$ is the set of all subsets of A. Next, we can view A^N as the Cartesian product of ω (the cardinality of N) copies of A

$$A^N = \prod_{i \in N} A$$

and we can say of the class $\mathfrak{A}$ of measurable rectangles R

$$R = \prod_{i=1}^{\infty} A_i, \quad A_i \in \mathfrak{M}_i$$

with $A_i = A$ for almost all i, which is, needless to say, a semiring. We define a set function μ on $\mathfrak{A}$ by

$$\mu(R) = \prod_i^{\infty} p(A_i).$$

Clearly, by consideration of product measure in 2.3, μ is σ-additive on $\mathfrak{A}$ and thus is so on $\mathfrak{R} = R(\mathfrak{A})$. Now we can extend μ further to a σ-algebra $\mathfrak{M} = S(\mathfrak{R}) = S(\mathfrak{A})$ by measure extension procedure.

Beside measurable rectangles we also consider a subclass $\mathfrak{S}$ of measurable rectangles S of the special form

$$S = (a_1, \ldots, a_n, A, A, \ldots), \qquad a_i \in A,\ n \geq 1$$

which are nothing but the subset wA^N of A^N, where $w = a_1 \ldots a_n \in A^*$. Clearly, each measurable rectangle of $\mathfrak{A}$ is a union no more than countable of sets from $\mathfrak{S}$, and consequently $S(\mathfrak{S}) = S(\mathfrak{A}) = \mathfrak{M}$.

As an immediate consequence of the existence of the product measure on A^N, we have

Theorem 1 *If $X \subseteq A^*$ is a code of A^* such that $A^N = XA^N$, then X is a prefix code and for any Bernoulli distribution p on A, $p(X) = 1$, so X is a maximal code.*

Proof. Set $X' = X - XA^+$. Then X' is a prefix code and $A^N = XA^N = X'A^N = \cup_{w \in X'} wA^N$. The union is certainly countable and disjoint, therefore

$$1 = \mu(A^N) = \mu(\bigcup_{w \in X'} wA^N) = \sum_{w \in X'} \mu(wA^N) = \sum_{w \in X'} p(w) = p(X') \leq p(X).$$

But X is a code, by the Kraft-McMillan inequality, $p(X) \leq 1$, which implies $p(X') = p(X) = 1$ and $X = X'$ is a maximal prefix code. $\square$

For any subset $X \subseteq A^N$, a *cover* of X is a finite or countable collection $\mathfrak{C}$ of sets from $\mathfrak{R}$ such that $X \subseteq \cup_{E \in \mathfrak{C}} E$. Since every set of $\mathfrak{R}$ is a finite or countable union of sets of $\mathfrak{S}$, so we can assume that a cover is always a countable collection of sets from $\mathfrak{S}$ and we write $\mathfrak{C} = \{w_i A^N : i \in I\}$, where $I \subseteq N$. From $\mathfrak{C}$ we discard the redundant subsets, that is, the subsets having no intersection with $X = \emptyset$ or containing another subset $\mathfrak{C}$ to obtain a subclass $\mathfrak{C}' = \{w'A^N : w' \in J \subseteq I\}$ which, evidently, is still a cover of X and besides $\{w' : w'A^N \in \mathfrak{C}'\}$ is a prefix subset of A^*. From now on, speaking of covers, we always mean covers with these properties. Obviously, the outer measure of X is

$$\mu^*(X) = \inf_{\mathfrak{C}} \sum_{wA^N \in \mathfrak{C}} \mu(wA^N) = \inf_{\mathfrak{C}} \sum_{wA^N \in \mathfrak{C}} p(w).$$

We prove now one simple property of the measure μ^*.

Proposition 2 *For any set $X \subseteq A^N$ and $w \in A^*, \mu^*(wX) = p(w)\mu^*(X)$.*

Proof. For any $\epsilon > 0$ let $\mathfrak{C} = \{w_i A^N : i \in I\}$ be a cover of X such that

$$\mu^*(X) \leq \sum_{i \in I} \mu(w_i A^N) = \sum_{i \in I} p(w_i) < \mu^*(X) + \epsilon$$

then $\mathfrak{C}' = \{ww_iA^N : i \in I\}$ is a cover of wX and

$$\begin{aligned}\mu^*(wX) \le \sum_{i\in I}\mu(ww_iA^N) &= \sum_{i\in I}p(ww_i)\\ = p(w)\sum_{i\in I}p(w_i) = p(w)\sum_{i\in I}\mu(w_iA^N) &< p(w)(\mu^*(X)+\epsilon)\end{aligned}$$

that means $\mu^*(wX) \le p(w)\mu^*(X)$.

For the reverse inequality, suppose that $\mathfrak{C} = \{w_iA^N : i \in I\}$ is a cover of wX,

$$wX \subseteq \bigcup_{i\in I} w_iA^N \tag{1}$$

such that

$$\mu^*(wX) \le \sum_{i\in I}\mu(w_iA^N) < \mu^*(wX)+\epsilon. \tag{2}$$

If $w = w_iw'$ for some i and $w' \in A^+$, then, in fact, $\mathfrak{C}$ must be a singleton class, $I = \{i\}$, hence

$$\mu^*(wX)+\epsilon > p(w_i) \ge p(w) \ge p(w)\mu^*(X).$$

If now for all i, w is a prefix of w_i, $w_i = ww_i'$, from (1) we have

$$X \subseteq \bigcup_{i\in I} w_i'A^N$$

that means $\mathfrak{C}' = \{w_i'A^N : i \in I\}$ is a cover, for which from (2) we get

$$\begin{aligned}p(w)\mu^*(X) \le p(w)\sum_{i\in I}\mu(w_i'A^N) &= \sum_{i\in I}\mu(ww_i'A^N)\\ = \sum_{i\in I}\mu(w_iA^N) &< \mu^*(wX)+\epsilon.\end{aligned}$$

That is, in both cases, ϵ abitrarily small, we have $p(w)\mu^*(X) \le \mu^*(wX)$ that concludes the proof. □

For any word $w \in A^\infty$ and any subset $E \subseteq A^\infty$ we define

$$\begin{aligned}w^{-1}E &= \{\beta \in A^\infty : (w\beta \in E)\&(w \in A^N) \Rightarrow \beta = \epsilon\};\\ Ew^{-1} &= \{\alpha \in A^\infty : (\alpha w \in E)\&(\alpha \in A^N) \Rightarrow w = \epsilon\}.\end{aligned}$$

The fisrt set is clear; the last one has the following meaning: empty word is the only one to be allowed to cut on the right of an infinite word in E. For any subset $F \subseteq A^\infty$, we write

$$F^{-1}E = \bigcup_{w\in F} w^{-1}E, \qquad EF^{-1} = \bigcup_{w\in F} Ew^{-1}.$$

Further on, p is assumed to be positive.

Proposition 3 *Let X be a subset of A^N and w a finite word of A^*. Then X is measurable if and only if wX is measurable and $\mu(wX) = p(w)\mu(X)$.*

Proof. It is easy to check that

$$w(X \triangle E) = (wX \triangle wE) \tag{3}$$

for any subset $E \subseteq A^N$. Set $E_1 = w^{-1}E$, we have

$$\begin{aligned} wX - wE_1 &= wX - E, \\ wE_1 - wX &\subseteq E - wX. \end{aligned}$$

Hence

$$w(X \triangle E_1) = (wX \triangle wE_1) \subseteq (wX \triangle E). \tag{4}$$

Proposition 2, monotonicity of μ^*, (3) and (4) imply that

$$\begin{aligned} p(w)\mu^*(X \triangle E) &= \mu^*(wX \triangle wE), \\ p(w)\mu^*(X \triangle E_1) &\leq \mu^*(wX \triangle E). \end{aligned}$$

Note that if $E \in \mathfrak{R}$ then $wE, w^{-1}E \in \mathfrak{R}$, so X is measurable iff wX is measurable. The second claim immediately follows from Proposition 2. □

Any language $X \subseteq A^\infty$ is a disjoint union of its finitary part $X_{\text{fin}} = X \cap A^*$ and its infinitaty part $X_{\text{inf}} = X \cap A^N$:

$$X = X_{\text{fin}} \cup X_{\text{inf}}.$$

For a langague of finite words $X \subseteq A^*$, commonly, X^* denotes its Kleene closure, that is $X^* = \{\epsilon\} \bigcup_{i=1}^{\infty} X^i$, or in other words, X^* is the smallest submonoid of A^* (thus of A^∞) containing X. We can extend this notion for any language X of A^∞, namely, X^* by definition is the smallest submonoid of A^∞ containing X, which, as one can easily verify, is $X^*_{\text{fin}} \cup X^*_{\text{fin}} X_{\text{inf}}$.

We recall now the concept of codes on A^∞ [Va]. Given any language X of A^∞ and a word $w \in A^\infty$, a *factorization* of w on X is a finite sequence of words $x_1, \ldots, x_{n-1}, x_n$ such that $x_1, \ldots, x_{n-1} \in X_{\text{fin}}, x_n \in X$ and $w = x_1 \ldots x_{n-1} x_n$. X is said to be an *infinitary code*, or *code* for short, if every word of A^∞ has at most one factorization on X. Clearly, if restricted to A^*, the infinitary codes are just the ordinary ones.

Naturally, we say that a subset $X \subseteq A^\infty$ is *measurable* if its infinitary part X_{inf} is measurable, and the measure $\mu(X)$ is defined to be

$$\mu(X) = p(X_{\text{fin}}) + \mu(X_{\text{inf}}).$$

Now we are in a position to prove the Kraft-McMillan inequality for infinitary codes.

Theorem 4 (Kraft-McMillan Inequality) *For any measurable code X of A^∞, $\mu(X) \leq 1$.*

Proof. Set $f = p(X_{\text{fin}}), i = \mu(X_{\text{inf}})$. We have $f \leq 1$ by Kraft-McMillan Inequality for ordinary codes. Since X is an infinitary code, the union

$$X^*_{\text{fin}} X_{\text{inf}} = \bigcup_{w \in X^*_{\text{fin}}} wX_{\text{inf}}$$

is disjoint. Therefore, by Proposition 2

$$\begin{aligned}\mu(X_{\mathrm{fin}}^* X_{\mathrm{inf}}) &= \sum_{w\in X_{\mathrm{fin}}^*} \mu(wX_{\mathrm{inf}}) = \sum_{w\in X_{\mathrm{fin}}^*} p(w)\mu(X_{\mathrm{inf}}) = \\ &= p(X_{\mathrm{fin}}^*)\mu(X_{\mathrm{inf}}) \le 1 = \mu(A^N).\end{aligned}$$

If $f < 1$, then

$$p(X_{\mathrm{fin}}^*) = 1 + f + f^2 + \cdots = \frac{1}{1-f}.$$

Consequently, $\frac{i}{1-f} \le 1$, i.e., $\mu(X) = i + f \le 1$. In the case $f = 1$, we show that $i = 0$. In fact, for all n, $p(X_{\mathrm{fin}} \cup \cdots \cup X_{\mathrm{fin}}^n)\mu(X_{\mathrm{inf}}) = ni$. Hence, if $i > 0$, $\mu(X_{\mathrm{fin}}^* X_{\mathrm{inf}}) = \lim_{n\to\infty} ni = \infty$, a contradiction. □

Example 5 A *prefix* of a word $\alpha \in A^\infty$ is a finite word w such that $\alpha = w\beta$ for some $\beta \in A^\infty - \epsilon$; a subset $X \subseteq A^\infty$ is called *prefix* if for any two words in X none of them is a prefix of the other i.e. $X_{\mathrm{fin}}(A^\infty - \epsilon) \cap X = \emptyset$; X is *prefix-maximal* if for any prefix subset $Y, X \subseteq Y$ implies $Y = X$. Evidently, a prefix subset is a code. Every prefix-maximal subset P is measurable and $\mu(P) = 1$. Indeed, since P is prefix-maximal, every word not in P_{inf} has a prefix in P_{fin}, therefore

$$A^N = P_{\mathrm{inf}} \bigcup_{w\in P_{\mathrm{fin}}} wA^N$$

is a disjoint union. Consequently

$$1 = \mu(A^N) = \mu(P_{\mathrm{inf}}) + \sum_{w\in P_{\mathrm{fin}}} \mu(wA^N) = \mu(P_{\mathrm{inf}}) + \sum_{w\in P_{\mathrm{fin}}} p(P_{\mathrm{fin}}) = \mu(P). \;\square$$

When A is a finite alphabet, any recognizable language is measurable, thus we have got a large class of measurable languages, which, by the way, are algorithmically constructible by finite means. Recall that a language $X \subseteq A^N$ is said to be *recognizable* if it is recognized by a finite Büchi automaton [Ei]. It has been well-known that the family Rec A^N of recognizable languages of A^N is the Boolean closure of the family Det A^N of deterministic recognizable ones (Büchi-McNaughton Theorem), i.e. the languages recognized by finite deterministic Büchi automata, which are the finite unions $\bigcup_{i=1}^n B_i C_i^\omega$, where B_i, C_i are (regular) prefix subsets of A^* and C_i^ω stands for the set of infinite words obtained by infinite concatenation of nonempty words of $C_i : C_i^\omega = \{x_1x_2\ldots : x_1, x_2, \ldots \in C_i\}$.

Proposition 6 *Every recognizable language X of A^N is measurable, i.e.* Rec $A^N \subseteq \mathfrak{M}$.

Proof. For any subset $B_iC_i^\omega$ with B_i, C_i prefix subsets of A^* we have

$$B_iC_i^\omega = \bigcap_{n=1}^{\infty} B_iC_i^nA^N.$$

By proposition 2, $B_iC_i^nA^N$ is measurable for all n. Since the σ-algebra $\mathfrak{M}$ of measurable subsets is closed under the formation of Boolean operations, moreover, of countable unions and intersections, $B_i C_i^\omega$ is measurable, hence $\mathrm{Det}A^N \subseteq \mathfrak{M}$ and thus $\mathrm{Rec}A^N \subseteq \mathfrak{M}$. □

We now resume the assumption that A is finite or countable. A code is said to be *maximal* if it cannot be included properly in another code. The existence of a maximal code containing a given code X is easily verified by mean of the Zorn's lemma. A maximal code must has a "nonnegligible" fraction of words in A^N. More precisely, we have

Proposition 7 *For every maximal code X, the outer measure of $X_{\inf}$ is positive: $\mu^*(X_{\inf}) > 0$.*

Proof. Let

$$\mathrm{FD}\,(X_{\inf}) = \{\alpha \in A^N : \exists w \in A^* : w\alpha \in X_{\inf}\}.$$

be the subset of *suffixes* of $X_{\inf}$. Suppose that $\mu^*(X_{\inf}) = 0$, hence $\mu^*(\mathrm{FD}\,(X_{\inf})) = 0$. For any $w \in A^+, w(w^{-1}X_{\inf}) \subseteq X_{\inf}$, we have

$$0 \le \mu^*(w(w^{-1}X_{\inf})) = p(w)\mu^*(w^{-1}X_{\inf}) \le \mu^*(X_{\inf}) = 0,$$

hence $p(w)\mu^*(w^{-1}X_{\inf}) = 0$ and so $\mu^*(w^{-1}X_{\inf}) = 0$. Consequently

$$0 \le \mu^*(\mathrm{FD}\,(X_{\inf})) = \mu^*(\bigcup_{w\in A^*} w^{-1}X_{\inf}) \le \sum_{w\in A^*} \mu^*(w^{-1}X_{\inf}) = 0$$

(subadditivity of μ^*).

On the other hand, being a maximal code, X is *complete* [Va], i.e., $A^N = \mathrm{FD}\,(X_{\mathrm{fin}}^* X_{\inf})$. By $\mu^*(X_{\inf}) = 0$

$$0 \le \mu^*(X_{\mathrm{fin}}^* X_{\inf}) \le \sum_{w\in X_{\mathrm{fin}}^*} \mu^*(wX_{\inf}) = \sum_{w\in X_{\mathrm{fin}}^*} p(w)\mu^*(X_{\inf}) = 0,$$

that is $\mu^*(X_{\mathrm{fin}}^* X_{\inf}) = 0$, therefore

$$\mu^*(\mathrm{FD}\,(X_{\mathrm{fin}}^* X_{\inf})) = 0 = \mu(A^N) = 1,$$

a contradiction. □

Example 8 (a non-measurable subset of A^N) A *suffix* of a word $\alpha \in A^\infty$ is a word β such that $\alpha = w\beta$ for some $w \in A^+$; $X \subseteq A^\infty$ is called a *suffix* subset if there are no words in X one of which is a suffix of the other, i.e. for every $w \in A^+ : X \cap wX = \emptyset$. A suffix set of A^N is called *suffix-maximal* if it is not contained properly in any other suffix subset of A^N. Let S be any suffix-maximal subset of A^N. Suppose that S is measurable; it is easy to see that $S \cup A$ is a code, so we have $\mu(S) = 0$. On the other hand, since $S \cup A$ is even a maximal code, the previous proposition shows that $\mu(S) = \mu^*(S) > 0$. This contradition means that S is not measurable.

In the propositions that follow we prove some properties of codes imposed with special conditions.

Proposition 9 *Let X be a measurable code of A^∞ with $\mu(X) = 1$ and $\mu(X_{\inf}) > 0$, then X_{fin} is a prefix code.*

Proof. We show that X_{fin}^* is *left unitary*, i.e., $X_{\text{fin}}^* = (X_{\text{fin}}^*)^{-1}X_{\text{fin}}^*$, whose base X_{fin} is then a prefix code. Always, $X_{\text{fin}}^* \subseteq (X_{\text{fin}}^*)^{-1}X_{\text{fin}}^*$. For the converse inclusion, we take any nonempty word $w \in (X_{\text{fin}}^*)^{-1}X_{\text{fin}}^*$, so there exist $u, v \in X_{\text{fin}}^*$ such that $uw = v$. Since $\mu(X) = 1, \mu(X_{\text{fin}}^* X_{\text{inf}}) = \frac{i}{1-f} = \frac{i}{i} = 1$, we have $wX_{\text{inf}} \cap X_{\text{fin}}^* X_{\text{inf}} \neq \emptyset$ otherwise

$$\mu(wX_{\text{inf}} \cup X_{\text{fin}}^* X_{\text{inf}}) = \mu(wX_{\text{inf}}) + \mu(X_{\text{fin}}^* X_{\text{inf}}) = p(w)i + 1 > 1$$

that is an obvious contradiction. So there exist $x \in X_{\text{fin}}^*, \alpha, \beta \in X_{\text{inf}}$ such that $w\alpha = x\beta$. Hence $v\alpha = ux\beta$, that implies $v = ux$, as X is a code. Thus $w = x \in X_{\text{fin}}^*$. □

Theorem 10 *If X is a measurable maximal code with $\mu(X) = 1$ then X_{fin} is a prefix code.*

Proof. By Proposition 7, $\mu(X_{\text{inf}}) > 0$ and by the previous proposition the result immediately follows. □

A language $X \subseteq A^\infty$ is called *finite-state* provided the collection $\{w^{-1}X : w \in A^*\}$ is finite. It is not difficult to prove that the family of finite-state languages is closed under the formation of finite unions, of finite intersections and the ω-product. It is noteworthy that $\mathrm{Rec}\, A^N$ is a subfamily of finite-state languages.

Proposition 11 *If X is a maximal code over A satisfying $(X_{\text{fin}}^*)^{-1}X_{\text{fin}}^* = A^*$, then X_{inf} is not a finite-state language if A consists of at least two elements.*

Proof. Under the assumption $(X_{\text{fin}}^*)^{-1}X_{\text{fin}}^* = A^*$, X is a (maximal) code iff X_{inf} is a suffix(-maximal) set. We show that a suffix-maximal language is not finite-state (the fact that it is not recognizable is shown in Example 8).

Fix $x \in A^*$, for any $r \in A^+$ we take a word

$$\alpha = (A^*(rx)^\omega \cup \mathrm{FD}\,(rx^\omega)) \cap X_{\text{inf}} \neq \emptyset.$$

This can be done, as X_{inf} is suffix-maximal. We write $\alpha = a(rx)^\omega$, where $a \in A^*$, hence $\alpha = arx(rx)^\omega$ and $(rx)^\omega \in (arx)^{-1}X_{\text{inf}}$. Thus for any x, there exists $u \in A^*$ such that $(ux)^{-1}X_{\text{inf}} \neq \emptyset$. Consequently, there exists an infinite sequence $v_1, v_2, \ldots$ such that v_i is a suffix of v_{i+1} and $v_i^{-1}X_{\text{inf}} \neq \emptyset$ for all i. As X_{inf} is a suffix set, $v_i^{-1}X_{\text{inf}} \neq v_j^{-1}X_{\text{inf}}$ for $i \neq j$. □

Proposition 12 *If X is a maximal code with X_{fin} a nonsingleton prefix code, then X_{inf} is not finite-state.*

Proof. Suppose on the contrary that X is finite-state. Consider the subset

$$Y_{\text{inf}} = X_{\text{inf}} \cap X_{\text{fin}}^\omega \subseteq X_{\text{fin}}^\omega \tag{5}$$

which is nonempty, since X is a maximal code. For every $w \in X_{\text{fin}}^*$ it is clear that

$$w^{-1}Y_{\text{inf}} = w^{-1}X_{\text{inf}} \cap X_{\text{fin}}^\omega \subseteq X_{\text{fin}}^\omega. \tag{6}$$

Let now c be a *coding morphism* for X_{fin}

$$c : B \to X_{\text{fin}},$$

where B is an alphabet of the same cardinality as X_{fin}. As X is a prefix code, we may correctly extend c to an injective morphism of monoids

$$c : B^{\infty} \rightarrow X_{\text{fin}}^{\infty},$$

where X_{fin}^{∞} denotes $X_{\text{fin}}^{*} \cup X_{\text{fin}}^{\omega}$. Therefore (5) and (6) and the fact that X is finite-state maximal code imply that $B \cup c^{-1}(Y_{\text{inf}})$ is also a finite-state maximal code on B^{∞} with Card $B \geq 2$ that contradicts Proposition 11. Thus X is not finite-state. □

Putting the propositions 6, 10 and 12 all together, we are lead to a situation quite opposite to the case of ordinary codes

Theorem 13 *Let X be a code on the finite alphabet A with X_{inf} a recognizable language of A^N, then the following two assertions are incompatible*

1. $\mu(X) = 1$
2. *X is a maximal code.*

References

[Sm] M. Smorodinsky, *On Infinite Decodable Codes,* Information and Control **11**(1968), 607–612.

[Ei] S. Eilenberg, *Automata, Languages and Machines,* Vol. **A**, Academic Press, New York, 1974.

[Be-Pe] J. Berstel, D. Perrin, *Theory of Codes,* Academic Press, New York, 1985.

[Va] Do Long Van, *Codes avec des mots infinis,* RAIRO Informatique théorique et applications **16**(1982), 371–386.

[Mc] B. McMillan, *Two Inequalities Implied By Unique Decipherability,* IRE Transactions on Information Theory **IT-2**(1956), 115–116.

[Ha] P. R. Halmos, *Measure Theory,* D. Van Nostrand, New York, 1950; Springer-Verlag, New York, 1974.

[Ko-Fo] A. N. Kolmogorov, S. V. Fomin, *Elements of the Theory of Functions and Functional Analysis,* Nauka, Moscow, 1981. (in Russian)

Received January 30, 1993

Revised February 20, 1994