Atti Accademia Nazionale dei Lincei

Classe Scienze Fisiche Matematiche Naturali

Rendiconti

Dan A. Simovici

On Some Measures on Free Semigroups Induced by Semiautomata

Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali. Rendiconti, Serie 8, Vol. **55** (1973), n.1-2, p. 4–9.

Accademia Nazionale dei Lincei

<http://www.bdim.eu/item?id=RLINA_1973_8_55_1-2_4_0>

Algebra. — *On Some Measures on Free Semigroups Induced by Semiautomata.* Nota (*) di DAN A. SIMOVICI, presentata dal Socio B. SEGRE.

RIASSUNTO. — In questa Nota si studia la relazione esistente fra la nozione di rappresentabilità in semiautomati e la misurabilità rispetto ad alcune misure sui semigruppi liberi. Se ne deduce quindi un teorema di caratterizzazione delle misure esteriori indotte da semiautomati su semigruppi liberi.

I. INTRODUCTION

A semiautomaton (in the common meaning of this word) is a 4-uple $\mathbf{A} = (I, S, f; s_0)$, where I, S are respectively the input alphabet and the set of states, $f: S \times I \to S$ is the transition function and $s_0 \in S$ is the initial state of the semiautomaton. Let I^* be the free semigroup generated by the input alphabet I.

The mapping f can be extended until $S \times I^*$, taking inductively on the length of the words from I^*:

$$f(s, e) = s, \qquad \forall s \in S$$

$$f(f(s, p), i) = f(s, pi), \qquad \forall p \in I^*, \quad i \in I,$$

e being the null word.

A semiautomaton homomorphism between the semiautomata $\mathbf{A}_j = (I, S_j, f_j, s_{0j})$, $j = 1, 2$ is a mapping $h: S_1 \to S_2$ for which $h(s_{01}) = s_{02}$ and $h(f_1(s, p)) = f_2(h(s), p)$, $\forall s \in S_1, p \in I^*$.

The language $E \subseteq I^*$ is represented in the semiautomaton $\mathbf{A} = (I, S, f, s_0)$ if there exists $S_1 \subseteq S$ so that $E = \{p \mid p \in I^*, f(s_0, p) \in S_1\}$. S_1 is the set of final states for the language E.

We shall consider a new notion which allows a new characterization of the languages which can be represented in semiautomata and the inference of some other properties of these languages.

A *measure semiautomaton* is a 5-uple $\mathbf{A} = (I, S, f, m; s_0)$, where I, S, f, s_0 have the previous significance and $m: \mathcal{S}(S) \to \overline{R}_+$ is a measure. Here $\mathcal{S}(S)$ is the set of the subsets of the set S. If m is a finite measure (positive measure-which means that $m(S_1) = 0 \to S_1 = \varnothing$) the measure semiautomaton is called measure-finite (positive).

The notion of representable language for measure semiautomata is the same as for common semiautomata.

(*) Pervenuta all'Accademia il 31 luglio 1973.

An *outer measure* on the set M is a mapping $\mu: \mathcal{S}(M) \rightarrow \overline{R}_+$ for which:

1) $M_1 \subseteq M_2$ implies $\mu(M_1) \leq \mu(M_2)$, $\forall M_1, M_2 \subseteq M$;

2) $\mu\left(\bigcup_{j=1}^{\infty} M_j\right) \leq \sum_{j=1}^{\infty} \mu(M_j)$, for every family $\{M_j \mid j \in N\} \subseteq \mathcal{S}(M)$.

Following Carathéodory, the μ–measurable sets are those sets M_1 for which:

$$\mu(L) = \mu(L \cap M_1) + \mu(L \cap M_1^c), \qquad \forall L \subseteq M.$$

It is a well known fact that the set of all μ–measurable sets is a σ–complete Boolean algebra with respect to the set-theoretical operations $(\cup, \cap, ^c)$ [3].

The outer measure induced by the semiautomaton $\mathbf{A} = (I, S, f, m; s_0)$ is given by:

$$\mu_{\mathbf{A}}(E) = m(\{s \mid f(s_0, p) = s, p \in E\}).$$

II. LANGUAGES REPRESENTABLE IN FINITE–MEASURE SEMIAUTOMATA

The connection between representable languages in semiautomata and measurable languages is described in the next two propositions.

THEOREM 1. *If E is a language which can be represented in the semiautomaton $\mathbf{A} = (I, S, f, m; s_0)$, then E is a $\mu_{\mathbf{A}}$–measurable language.*

Proof. If K is a language, $K \subseteq I^*$, let us denote $f(s_0, K) = \{s \mid \exists p \in K, s = f(s_0, p)\}$. Since E is a representable language, we have $f(s_0, E) \cap \cap f(s_0, E^c) = \varnothing$ and it follows:

$$\mu_{\mathbf{A}}(E_1) = m(f(s_0, E_1)) = m(f(s_0, E_1 \cap E) \cup f(s_0, E_1 \cap E^c)) =$$
$$= m(f(s_0, E_1 \cap E)) + m(f(s_0, E_1 \cap E^c)) =$$
$$= \mu_{\mathbf{A}}(E_1 \cap E) + \mu_{\mathbf{A}}(E_1 \cap E^c), \qquad \text{for every } E_1 \subseteq I^*.$$

Hence E_1 is a $\mu_{\mathbf{A}}$–measurable language.

THEOREM 2. *Let $\mathbf{A}$ be a finite-measure semiautomaton. A $\mu_{\mathbf{A}}$–measurable language differs from a representable language only by a $\mu_{\mathbf{A}}$–negligible set.*

Proof. Let E_1 be a $\mu_{\mathbf{A}}$–measurable language and $\bar{E}_1 = \{p \mid f(s_0, p) \in \in f(s_0, E_1)\}$. It is obvious that $\bar{E}_1$ is represented in the semiautomaton $\mathbf{A}$ and $E_1 \subseteq \bar{E}_1$. Since $\bar{E}_1$ is $\mu_{\mathbf{A}}$–measurable (in view of Theorem 1) we can write:

$$\mu_{\mathbf{A}}(\bar{E}_1 \setminus E_1) = \mu_{\mathbf{A}}(\bar{E}_1) - \mu_{\mathbf{A}}(E_1) = m(f(s_0, \bar{E}_1)) - m(f(s_0, E_1)) = 0,$$

which shows that E_1 differs from $\bar{E}_1$ by a $\mu_{\mathbf{A}}$–negligible set.

COROLLARY. *If $\mathbf{A}$ is a positive, finite-measure semiautomaton the $\mu_{\mathbf{A}}$–measurable languages coincide with the languages which are represented in the same measure semiautomaton.*

REMARK. The existence of a finite, positive-measure $m: \mathcal{S}(S) \to R_+$ implies immediately that $|S| \leq \aleph_0$. (Here $|S|$ is the cardinal number of the set S). Hence, only for semiautomata having the set of states of cardinality less or equal than $\aleph_0$, the representable languages can coincide with the measurable languages.

Let now μ be an outer measure on I^* and $M(\mu)$ the Boolean algebra of μ-measurable sets. Suppose that $M(\mu)$ is an atomic Boolean algebra and let $\mathfrak{A}_\mu$ be the set of its atoms. If $E_1, E_2 \in \mathfrak{A}_\mu$, we have $E_1 \cap E_2 = \varnothing$ and $\cup\{E \mid E \in \mathfrak{A}_\mu\} = I^*$. Hence the family of sets $\{E \mid E \in \mathfrak{A}_\mu\}$ is a partition of I^*.

Let $\zeta_\mu: \mathcal{S}(I^*) \to \mathcal{S}(I^*)$ be the mapping given by

$$\zeta_\mu(K) = \cup\{E \mid E \in \mathfrak{A}_\mu, E \cap K = \varnothing\}.$$

It is clear that ζ_μ is a closure operator [1].

The outer measures induced by measure semiautomata on their free input semigroup are characterized by the following:

THEOREM 3. *The outer measure* $\mu: \mathcal{S}(I^*) \to R_+$ *is induced by a positive, finite-measure semiautomaton iff there are satisfied the following conditions*:

1) *the algebra* $M(\mu)$ *is an atomic one,*

2) $\forall E \in \mathfrak{A}_\mu$, $p \in I^*$, $\exists E' \in \mathfrak{A}_\mu$ *so that* $Ep \subseteq E'$,

3) *the outer measure* μ *is invariant with respect to the closure* ζ_μ, *which means that*:

$$\mu(E) = \mu(\zeta_\mu(E)), \qquad \forall E \in \mathcal{S}(I^*).$$

Proof. Suppose that the outer measure μ is induced by the semiautomaton $\mathbf{A} = (I, S, f, m; s_0)$ positive and finite in measure, which means that $\mu = \mu_{\mathbf{A}}$.

Let us consider the languages:

$$E_s = \{p \mid f(s_0, p) = s\}, \qquad s \in S$$

and let E be a measurable language. In view of Theorem 2, E is representable in the semiautomaton $\mathbf{A} = (I, S, f, m; s_0)$. It is obvious that $E = \cup\{E_s \mid s \in f(s_0, E)\}$; E_s are also representable languages in the same semiautomaton and are the minimal languages having this property. Hence $M(\mu)$ is an atomic Boolean algebra.

The atoms of the algebra $M(\mu)$ are necessarily languages having the form E_s. It follows that, if $r \in E_s$ than $s = f(s_0, r)$ and $f(s_0, rp) = f(s, p)$, from which $rp \in E_{f(s,p)}$, hence $E_s p \subseteq E_{f(s,p)}$ which points the validity of the second property.

Finally, we have

$$\mu(E_1) = \mu_{\mathbf{A}}(E_1) = m(f(s_0, E))$$

and

$$\mu(\zeta_\mu(E_1)) = \mu(\cup\{E \mid E \in \mathfrak{A}_\mu, E_1 \cap E \neq \varnothing\}).$$

The languages $\{E \mid E \in \mathfrak{A}_{\mu}, E_1 \cap E \neq \varnothing\}$ are mutually disjoint. Since these languages are representable in the semiautomaton **A**, they are $\mu_{\mathbf{A}}$–measurable languages. Hence:

$$\mu(\zeta_{\mu}(E_1)) = \Sigma\{\mu(E) \mid E \in \mathfrak{A}_{\mu}, E_1 \cap E \neq \varnothing\} =$$

$$= \Sigma\{m(\{s\}) \mid s \in f(s_0, E), E \in \mathfrak{A}_{\mu}, E_1 \cap E \neq \varnothing\} =$$

$$= m(f(s_0, E_1)) = \mu(E_1).$$

and we have proved the third property.

Let μ be a finite, outer measure which fulfils the conditions 1)–3) and the semiautomaton $\mathbf{A} = (I, \mathfrak{A}_{\mu}, \varphi, m; E_0)$, E_0 being the atom which contains the null word e of the free semigroup I^*.

Here $\varphi(E, i) = E'$ (where E' is the atom for which $Ep \subseteq E'$) and $m(\{E_j \mid j \in J\}) = \Sigma\{\mu(E_j) \mid j \in J\}$. The mapping φ is well defined for, if there existed $E', E'' \in \mathfrak{A}_{\mu}$ so that $Ep \subseteq E', Ep \subseteq E''$ with $E' \neq E''$, there would result $Ep \subseteq E' \cap E'' = \varnothing$, which is absurd.

We claim that $\mu = \mu_{\mathbf{A}}$. Indeed, we can write $\mu(K) = \mu(\zeta(K)) = \mu(\{\cup H \mid H \in \mathfrak{A}_{\mu}, H \cap K \neq \varnothing\})$.

The condition 2) shows that the equivalence α related to the partition $\mathfrak{A}_{\mu}$ is a right congruence on the semigroup I^*. Let us denote by $[p]_{\alpha}$ the set $E \in \mathfrak{A}_{\mu}$ which contains p. The third condition implies that $\mu(\{p\}) = \mu([p]_{\alpha})$.

We have:

$$\mu_{\mathbf{A}}(K) = \mu(\varphi([e]_{\alpha}, K)) = \mu(\{[p]_{\alpha} \mid p \in K\}) = \mu(K)$$

and the proof is completed.

REMARK. Let $\mu_{\mathbf{A}}$ be an outer measure induced by the semiautomaton $\mathbf{A} = (I, S, f, m; s_0)$. The condition 2) implies that, if $E \in \mathfrak{A}_{\mu}$, $p \in I^*$ and $Ep \cap E \neq \varnothing$ then $Ep \subseteq E$.

Hence, for $p, pq, r \in E$ we have, $rq \in E$. This is the condition obtained by A. Salomaa in [4] for languages which can be represented by one state in a semiautomaton.

Conversely, if this condition holds for the language E there exists a semiautomaton $\mathbf{A} = (I, S, f; s_0)$ and a state $s \in S$ so that $E = E_s$. If the set S is endowed with a measure $m: \mathcal{P}(S) \to R_+$, it follows that E is an atom of the Boolean algebra $M(\mu)$ which corresponds to the outer measure induced by the measure semiautomaton $\mathbf{A}_m = (I, S, f, m; s_0)$ on I^*.

Hence the condition of Salomaa characterizes also the atoms of the Boolean algebras of $\mu_{\mathbf{A}}$ measurable sets.

Let μ be a finite outer measure $\mu: \mathcal{P}(I^*) \to R_+$. The outer measure induced by μ on the set of states of a semiautomaton $\mathbf{A} = (I, S, f; s_0)$ is given by:

$$m_{\mu}(S_1) = \mu(\{p \mid f(s_0, p) \in S_1\}).$$

We have the following "regularity" theorem:

THEOREM 4. *Let* $\mathbf{A} = (I, S, f, m; s_0)$ *be a measure semiautomaton and* $m_\mu : \mathcal{S}(S) \to R_+$ *be the outer measure induced on the set S by the outer measure* $\mu = \mu_{\mathbf{A}}$. *Then* $m = m_\mu$.

Let $\mathbf{A} = (I, S, f; s_0)$ *be a semiautomaton. If the languages which are representable in this semiautomaton are* μ*–measurable, then the outer measure* $m_\mu : \mathcal{S}(S) \to R_+$ *is a measure and the outer measure* μ_1 *induced by the measure semiautomaton* $\mathbf{A} = (I, S, f, m_\mu; s_0)$ *coincides with* μ *on the set of representable languages.*

Proof. In view of the definition of the measure m_μ we can write

$$m_\mu(S_1) = \mu(\{p \mid f(s_0, p) \in S_1\}) =$$
$$= m(\{s \mid f(s_0, p) = s, s \in S_1\}) = m(S_1)$$

and the first part of the theorem is proved.

Let now $\mathbf{A} = (I, S, f; s_0)$ be a semiautomaton, $S_1, S_2 \subseteq S$, $S_1 \cap S_2 = \varnothing$. We have

$$m_\mu(S_1 \cup S_2) = \mu(\{p \mid f(s_0, p) \in S_1 \cup S_2\}) =$$
$$= \mu(\{p \mid f(s_0, p) \in S_1\} \cup \{p \mid f(s_0, p) \in S_2\}).$$

Since the languages $E_j = \{p \mid f(s_0, p) \in S_j\}$, $j = 1, 2$ are represented in the semiautomaton $\mathbf{A} = (I, S, f; s_0)$, they are $\mu_{\mathbf{A}}$–measurable languages and being disjoint sets we obtain:

$$m_\mu(S_1 \cup S_2) = \mu(\{p \mid f(s_0, p) \in S_1\}) + \mu(\{p \mid f(s_0, p) \in S_2\}) =$$
$$= m_\mu(S_1) + m_\mu(S_2)$$

hence m_μ is a measure.

We have

$$\mu_1(E) = m_\mu(\{s \mid f(s_0, p) = s, p \in E)\} =$$
$$= \mu(\{q \mid f(s_0, q) \in f(s_0, E)\}).$$

If the language E is represented in the semiautomaton $\mathbf{A}$ we have $E = \{q \mid f(s_0, q) \in f(s_0, E)\}$. Hence, for these languages $\mu_1(E) = \mu(E)$.

III. Cardinal measure semiautomata

In this paragraph we shall consider a special class of measure semiautomata.

A *cardinal measure semiautomaton* is a measure semiautomaton $\mathbf{A} = (I, S, f, m; s_0)$ for which $m(S_1) = |S_1|$, $\forall S_1 \subseteq S$.

We can relate to every semiautomaton, in a natural manner, its cardinal measure semiautomaton. We shall use the same capital letter to denote a semiautomaton and its cardinal measure semiautomaton.

The outer measure induced on the free input semigroup by a cardinal measure semiautomaton characterizes the isomorphism type of the semiautomaton, if this semiautomaton is connected by its initial state.

Let $\mathbf{A}_j = (I, S_j, f_j; s_{0j})$ $j = 1, 2$, be two semiautomata connected by their initial states. We have:

THEOREM 5. *The following assertions are equivalent*:

i) *There exists a surjective homomorphism* $h: S_1 \to S_2$ *between* $\mathbf{A}_1$ *and* $\mathbf{A}_2$;

ii) $\mu_{\mathbf{A}_1}(E) \geq \mu_{\mathbf{A}_2}(E)$, $\quad \forall E \in \mathcal{S}(I^*)$;

iii) $\mu_{\mathbf{A}_1}(E) \geq \mu_{\mathbf{A}_2}(E)$ *for the languages* $E \in \mathcal{S}(I^*)$ *for which* $|E| \leq 2$.

COROLLARY. *If* $\mathbf{A}_j = (I, S_j, f_j; s_{0j})$, $j = 1, 2$ *are two connected semiautomata by their initial states and* $\mu_{\mathbf{A}_1}(E) = \mu_{\mathbf{A}_2}(E)$ *for every* $E \in \mathcal{S}(I^*)$, $|E| \leq 2$ *then* $\mathbf{A}_1$ *and* $\mathbf{A}_2$ *are isomorphic semiautomata.*

Let now $E \in \mathcal{S}(I^*)$ be a language with $e \in E$. We shall consider the set of languages $\{E^k \mid k \in N\}$ where $E^0 = \{e\}$, $E^1 = E$, $E^{k+1} = \{pq \mid p \in E^k, q \in E\}$. It is clear that $E^0 \subseteq E^1 \subseteq \cdots E^k \subseteq \cdots$.

Suppose that $\mathbf{A}$ is a finite semiautomaton.

THEOREM 6. *If* $\mu_{\mathbf{A}}(E^k) = \mu_{\mathbf{A}}(E^{k+1})$ *then* $\mu_{\mathbf{A}}(E^k) = \mu_{\mathbf{A}}(E^{k+j})$, $\forall j \in N$.

In a next Note we shall study the behaviour of $\mu_{\mathbf{A}}$–measurable languages to homomorphisms of the free semigroup I^*.

REFERENCES

[1] P. M. COHN, *Universal Algebra* (Harper and Row, Evanston and London 1965).

[2] A. GINZBURG, *Algebraic Theory of Automata* (Academic Press, New York, London 1968).

[3] P. R. HALMOS, *Measure Theory* (Van–Nostrand, Princeton 1950).

[4] A. SALOMAA, *Theorems on the representation of events in Moore Automata*, « Ann. Univ. Turku », ser. A. I., 69, 3–14 (1964).