A NOTE ON LANGUAGE MEASURABILITY memo: ONGOING DRAFT

RYUYA HORA

Abstract. This is a short note on a σ-additive measure extending Sin'ya's Reg-measurability of formal languages. memo: This work (in progress) was originally intended to be a note exclusively for discussion with Sin'ya Ryoma. Any comments, especially on errors or preceding work, would be greatly appreciated.

Contents

1. Introduction: σ-additive density of languages

For a formal language L, the notion of *density* $\mathsf{dty}(L)$ is, intuitively, a real number in $[0,1]$ that represents the probability that a random word belongs to L. Density turns out to be a much more powerful concept than this naive definition might suggest, and already in the earliest stages of its study, connections to the computational hierarchy were established [Sch65; Ber72]. For instance, the density of a regular language is always a rational number in $[0,1]\cap\mathbb{Q}$, and the density of an unambiguous context-free language is an algebraic number in $[0,1]\cap\overline{\mathbb{Q}}$. Therefore, this tool can be used to prove that a given language is *not* an unambiguous context-free language.

Then, we mathematicians would ask: how "natural" is language density as a mathematical object? Put less pedantically: how does language density connect to modern mathematical theories, and how might it develop using modern mathematical tools? Of course, the most naive expectation is a link to probability theory. Density is certainly probabilistic in nature, and recent work on the measurability of languages by Sin'ya / Sin'ya Yuyama [SY25] implicitly suggests that collections of languages can be embedded into an appropriate function space. Also, probability measures on a free monoid might remind one of Haar measure on a locally compact group. These ideas suggest connections with harmonic analysis on languages. The recent works [Hor24; Ura25] might suggest a further connection with profinite Galois groups, which is one of my personal motivations.

However, to the best of the author's knowledge, people have long believed that there is a basic and decisive gap between language density and modern probability theory, which is the problem of σ-additivity. Namely, density is at most finitely additive and does not satisfy σ-additivity (unlike the standard settings in probability theory). This seems deeply tied to the nature of languages because there are only countably many words. If we put a σ-additive measure on a countable set A^* such that every single word has measure 0, then clearly every subset must have measure 0 as well. Therefore, it is often said that trying to fit language density into the paradigm of σ-additive probability measures is misguided.

This paper provides a way to treat language density in a σ-additive manner using profinite words. We define a σ-additive probabilistic Borel measure on $\widehat{A^*}$, which is known to be the Stone dual to the boolean algebra of regular languages Reg. memo: See [Pin22] and papers cited therein. Since this profinite completion is no longer a countable infinite set, it can carry σ-additive probability measures that were impossible before.

Let us explain why the profinite space $\widehat{A^*}$ is a natural subject to consider. It is not only because it admits the σ-additive measure, but also because it has exactly the same information as the densities of regular languages, which have the canonical notion of density. memo: Densities can be defined a priori in may ways. But they tend to coincide and define the canonical notion of density for regular languages. Furthermore, the σ-additive probability on $\widehat{A^*}$ captures the notion of Reg-measurability, which clarifies an analytic aspect of the asymptotic approximation of languages [Sin21]. memo: As mentioned above, one of the motivations is to prepare a bridge to modern mathematics—especially geometric group theory and harmonic analysis on groups.

2. Densities as finitely additive probabilistic measures

2.1. The classical notion of density. We fix a non-empty finite set of alphabet A and write A^* for the free monoid it generates.

Definition 2.1.1 (density)**.** For a language $L \subset A^*$, its *density* is defined to be the Cesàro mean of $\frac{\#L\cap A^k}{\#A^k}$

$$\mathsf{dty}(L) := \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \frac{\#(L\cap A^k)}{\#A^k}.$$

We say that a language L *has density* if the limit converges.

Remark 2.1.2 (Historical remark). memo: ... possibly [Sch65] To the author's knowledge, the above definition first appears in the first ICALP [Ber72] as a special case of *densité asymptotique* $d^\eta(L)$ for the monoid homomorphism $\eta\colon A^* \to \mathbb{R}_>$ that sends every character $a \in A$ to $\frac{1}{n}$.

Fact 2.1.3. *Every regular language has density.*

3. A general theory for Boolean algebras with probability

3.1. Preliminary: Stone duality. See Johnstone's textbook [Joh82] for detailed explanations of Stone duality.

Notation 3.1.1. Let X be a topological space.

- We write $\Delta(X)$ for the boolean algebra of all clopen subsets of X, and $\mathcal{B}_\Delta(X)$ for the σ-algebra generated by $\Delta(X)$.

- We write $\mathcal{O}(X)$ for the set of all open subsets of X, and $\mathcal{B}(X)$ for the σ-algebra generated by $\mathcal{O}(X)$, i.e., the set of all Borel sets of X.

Fact 3.1.2 (Stone duality [Sto36; Sto37; Doc64])**.** *The category of Boolean algebras and the category of profinite spaces are contravariantly equivalent via the functor* Δ.

$$\mathbf{ProfSp}^{\mathrm{op}} \xrightarrow[\simeq]{\Delta} \mathbf{BoolAlg}$$

We write Sp for the (essentially unique) converse equivalence $\mathrm{Sp}\colon \mathbf{BoolAlg}^{\mathrm{op}} \to \mathbf{ProfSp}$

3.2. Boolean algebra with measure.

Definition 3.2.1. A *Boolean algebra with measure* (in short, *BwM*) is a pair (B, μ) of a Boolean algebra B and a function $\mu\colon B \to [0, \infty)$ such that

- for any $x, y \in B$, $x \wedge y = \bot$ implies $\mu(x \vee y) = \mu(x) + \mu(y)$.

Lemma 3.2.2 (Basic properties)**.** *For any BwM* (B, μ)*, the following conditions hold.*

(1) $\mu : B \to [0, \infty)$ *is order-preserving.*
(2) $\mu(\bot) = 0$
(3) *For any* $x, y \in B$, $\mu(x \vee y) \leq \mu(x) + \mu(y)$

Proof. We prove them one by one.

(1) Assuming $x \leq y$, we have $y = y \wedge (x \vee \neg x) = x \vee (y \wedge \neg x)$. Since $x \wedge (y \wedge \neg x) = \bot$, we have $\mu(y) = \mu(x) + \mu(y \wedge \neg x) \geq \mu(x)$. This proves that μ is order-preserving.
(2) Since $\bot = \bot \wedge \bot$, we have $\mu(\bot) = \mu(\bot \wedge \bot) = \mu(\bot) + \mu(\bot)$, which implies $\mu(\bot) = 0$.
(3) As $x \vee y = x \vee (\neg x \wedge y)$, we have $\mu(x \vee y) = \mu(x) + \mu(\neg x \wedge y) \leq \mu(x) + \mu(y)$.

$\square$

For a topological space, we call a finitely additive measure on $\Delta(X)$ a Δ*-measure*. A (finitely or σ-)additive measure is *finite* if the measure of the whole set is finite. The next proposition immediately follows from the argument above and the isomorphism $\Delta(\mathrm{Sp}(B)) \cong B$ given by the Stone duality.

Proposition 3.2.3. *For any BwM* (B, μ)*, the function* $\Delta(\mathrm{Sp}(B)) \cong B \xrightarrow{\mu} [0, \infty)$ *defines a finitely additive finite* Δ*-measure on* $\mathrm{Sp}(B)$*. Furthermore, this correspondence provides a bijective correspondence between BwM structures on a Boolean algebra* B *and finitely additive finite* Δ*-measure on its Stone dual* $\mathrm{Sp}(B)$*.*

$$\text{BwMs on a Boolean algebra } B \xleftrightarrow{1:1} \text{(fin.add.) finite } \Delta\text{-measures on its Stone dual } \mathrm{Sp}(B)$$

3.3. Extension via Hopf-Kolmogorov.

Then, we expand our consideration from $\Delta(X)$ to $\mathcal{B}_\Delta(X)$ (Notation 3.1.1).

Proposition 3.3.1. *memo: This should be already known For any Boolean algebra* B*, there is a natural one-to-one correspondence between BwM structures on* B *and finite* $\mathcal{B}_\Delta$*-measures on its stone dual* $\mathrm{Sp}(B)$*.*

$$\text{BwMs on a Boolean algebra } B \xleftrightarrow{1:1} (\sigma\text{-add.) finite } \mathcal{B}_\Delta\text{-measures on its Stone dual } \mathrm{Sp}(B)$$

Our proof requires the following classical fact.

Fact 3.3.2 (Hopf-Kolmogorov extension)**.** *memo: Possible citations include [Coh13, Exercise 1.3.5, 1.3.6] A finitely additive measure* (X, B, μ) *admits an extension to* σ*-additive measure on the measureble space* $(X, \sigma(B))$ *if and only if it satisfies the* σ*-additivity in the following sense.*

σ-additivity: *For any mutually disjoint family* $\{A_i\}_{i=0}^{\infty}$ *of elements of* B*, if* $\bigcup_{i=0}^{\infty} A_i$ *happens to be an element of* B*, then* $\mu\left(\bigcup_{i=0}^{\infty} A_i\right) = \sum_{i=0}^{\infty} \mu(A_i)$ *holds.*

Furthermore, the finitely additive measure space (X, B, μ) is finite, i.e. $\mu(X) < \infty$, then the extension of (X, B, μ) to $(X, \sigma(B))$ is unique.

Proof of Proposition 3.3.1. Due to Proposition 3.2.3 and Fact 3.3.2, it suffices to prove that any finitely additive Δ-measure on $\mathrm{Sp}(B)$ is σ-additive. Let $\{A_i\}_{i=0}^{\infty}$ be a mutually disjoint family of elements of B such that $A \coloneqq \bigcup_{i=0}^{\infty} A_i \subset \mathrm{Sp}(B)$ is an element of $\Delta(\mathrm{Sp}(B))$. Since $\mathrm{Sp}(B)$ is compact, every closed (in particular, clopen) subset is also compact. Therefore, the equation $A \coloneqq \bigcup_{i=0}^{\infty} A_i$ provides a open cover of a compact subspace A, and hence there exists a finite open subcover. As the family $\{A_i\}_{i=0}^{\infty}$ is mutually disjoint, we conclude that all but finite components are empty subsets. Thus, the required equality $\mu\left(\bigcup_{i=0}^{\infty} A_i\right) = \sum_{i=0}^{\infty} \mu(A_i)$ is reduced to the finite additivity of μ. □

Remark 3.3.3 (Completeness in the above proof). To the question of why σ-additivity can be proved from purely finite data, the following answer can be given. For a Boolean algebra, its Stone dual is the space obtained by adjoining all points that exist virtually. In this sense, Stone duality can be regarded as a kind of existence theorem for points (or completeness theorem). What makes the crucial assumption of σ-additivity in the Hopf–Kolmogorov theorem trivial in our case is precisely an existence theorem asserting that for any equation of the form $A \coloneqq \bigcup_{i=0}^{\infty} A_i$, there exists a point refuting it.

3.4. Comparison with Borel measures.

Corollary 3.4.1. *If a Boolean algebra B is countable, there is a one-to-one correspondence between BwM structures on B and Borel measures on its Stone dual $\mathrm{Sp}(B)$.*

$$\text{BwMs on a Boolean algebra } B \xleftrightarrow{1:1} (\sigma\text{-add.}) \text{ finite Borel-measures on its Stone dual } \mathrm{Sp}(B)$$

Fact 3.4.2. *Any profinite space is zero-dimensional, i.e., every open subset is a sum of clopen subsets.*

Proof. memo: This follows since every profinite space is a flitered limit of finite discrete topological spaces. □

Lemma 3.4.3. *If a Boolean algebra B is countable, then the $\mathcal{B}_{\Delta}(\mathrm{Sp}(B))$ coincides with the Borel σ-algebra $\mathcal{B}(\mathrm{Sp}(B))$.*

Proof. As the Borel σ-algebra $\mathcal{B}(\mathrm{Sp}(B))$ is generated by open subsets $\mathcal{O}(\mathrm{Sp}(B))$, it suffices to prove $\mathcal{O}(\mathrm{Sp}(b)) \subset \mathcal{B}_{\Delta}(\mathrm{Sp}(B))$. The countability of B and Fact 3.4.2 implies that every open subset is a sum of a necessarily countable number of clopen subsets. □

Remark 3.4.4 (Countability assumption is necessary). We cannot remove the countability assumption in Lemma 3.4.3. Let I be an uncountable set, and B_I be the free Boolean algebra generated by I. Then, its Stone dual $\mathrm{Sp}(B_I)$ is (homeomorphic to) the product space $2^I \coloneqq \{f \colon I \to \{0,1\} \text{ function}\}$. We call a subset $A \subset 2^I$ $\aleph_0$-*bounded* if there exists a countable subset $J \subset I$ such that A is an inverse image of a subset of 2^J along the restriction function $2^I \twoheadrightarrow 2^J$. For any countable family of $\aleph_0$-bounded subsets $\{A_i\}_{i=0}^{\infty}$ witnessed by $\{J_i\}_{i=0}^{\infty}$, their countable sum or intersection is also $\aleph_0$-bounded witnessed by $\bigcup_{i=0}^{\infty}$. With easier observations on finite Boolean operations, we can conclude that the set of all bounded subsets, denoted by $\mathcal{B}_0(2^I)$, is a σ-additive family. Furthermore, one can check $\Delta(2^I) \subset \mathcal{B}_0(2^I)$, which implies $\mathcal{B}_{\Delta}(2^I) \subset \mathcal{B}_0(2^I)$.

On the other hand, $\mathcal{B}(2^I)$ is not included by $\mathcal{B}_b(2^I)$, since the closed subset $\{\mathrm{const}_0\}$, consisting of the unique element of the constant function at 0, is not $\aleph_0$-bounded. Therefore, we have proven $\mathcal{B}_{\Delta}(2^I) \subset \mathcal{B}_b(2^I) \not\supset \mathcal{B}(2^I)$, in particular, $\mathcal{B}_{\Delta}(2^I) \neq \mathcal{B}(2^I)$.

4. On the language measurability

Definitions are due to [SY25] and the papers cited therein.

Words A^*	Profinite words $\widehat{A^*}$	
regular	clopen $\partial C = \emptyset$	classically known Almeida-Pippenger
upper Reg-approximation	closure $\overline{L}$	
lower Reg-approximation	$\widehat{A^*} \setminus \mathrm{Int}(\widehat{A^*} \setminus L)$	
Reg-measurability	null-boundary $\widehat{\delta}(\partial C) = 0$	

TABLE 1. Correspondence between words and profinite words

4.1. σ-additive density of profinite words.

memo: Write contexts: Stone dual between the profinite monoid and the Boolean algebra of regular languages.

Corollary 4.1.1. *There exists a unique probabilistic Borel measure $\widehat{\delta}$ on $\widehat{A^*}$, such that for any clopen subset $C \subset \widehat{A^*}$*

$$\widehat{\delta}(C) = \delta(C \cap A^*).$$

Example 4.1.2 (The singleton case is the pushout of the Haar measure on $\widehat{\mathbb{Z}}$)**.** In the case $\#A = 1$, the profinite space $\widehat{A^*}$ is known to be $\mathbb{N} \sqcup \widehat{\mathbb{Z}}$, where the discrete subspace $\mathbb{N}$ is dense in $\widehat{A^*}$, and its complement subspace $\widehat{\mathbb{Z}}$ is homeomorphic to the usual profinite topology on the profinite integers. As $\widehat{\mathbb{Z}}$ is a profinite group, it admits the unique Haar probabilistic measure. In fact, the canonical measure on $\widehat{A^*}$ coincides with the pushout of this measure along the embedding $\widehat{\mathbb{Z}} \to \widehat{A^*}$. memo: write a proof

memo:

Conjecture 4.1.3. *The Borel measure on $\widehat{A^*}$ is the unique probabilistic measure that is stable under the average of pushouts by alphabets with possibly some additional conditions.*

memo: This might be connected with the theory of μ-stationary measures such as [Kai00]

Definition 4.1.4. We call the completion of the measure $\widehat{\delta}$ the *canonical measure* on $\widehat{A^*}$

4.2. Comparison with the Sin'ya Reg-measurability.

Notation 4.2.1. For a subset S of a topological space X, the *boundary* of S is defined to be the set $\partial S := \overline{S} \setminus \mathrm{Int}(S)$.

Definition 4.2.2. Let (X, μ) be a Borel measure on a topological space X. We say that a subset $S \subset X$ *has a null-boundary* if its boundary ∂S is contained in a μ-null set.

Example 4.2.3. A subset $S \subset X$ is clopen if and only if its boundary ∂S is empty. Therefore, every clopen subset $C \subset X$ has a null-boundary.

Lemma 4.2.4. *For any language $L \subset A^*$, the $\widehat{\delta}$-measure of its closure $\widehat{\delta}(\overline{L})$ coincides with the upper* Reg-*approximation.*

Proposition 4.2.5 (Comparison lemma)**.** *For any subset $S \subset \widehat{A^*}$ with a null-boundary, the intersection with finite words $S \cap A^*$ is* Reg-*measurable (in the sense of [Sin21]). Furthermore, in that case, we have $\widehat{\delta}(S) = \delta(S \cap A^*)$.*

Proof. Let $S \subset \widehat{A^*}$ be a subset with $\widehat{\delta}$-null boundary. Then, we have

$$\widehat{\delta}\left(\overline{S}\right) = \widehat{\delta}(\mathrm{Int}S) + \widehat{\delta}(\partial S) = \widehat{\delta}(\mathrm{Int}S).$$

As $\widehat{A^*}$ has a countable clopen basis, which is given by the Stone dual to regular languages, There exist two sequences of clopen subsets $(C_i)_{i=0}^{\infty}, (D_i)_{i=0}^{\infty}$ such that

$$C_0 \subset C_1 \subset \cdots \subset S \subset \cdots \subset D_1 \subset D_0,$$

and
$$\bigcup_{i=0}^{\infty} C_i = \mathrm{Int}S, \text{ and } \bigcap_{i=0}^{\infty} D_i = \overline{S}.$$
Thus, we have the following situation
$$C_0 \subset C_1 \subset C_2 \subset \cdots \to \overbrace{\mathrm{Int}S \subset S \subset \overline{S}}^{\text{null}} \leftarrow \cdots \subset D_2 \subset D_1 \subset D_0 \quad (\subset \widehat{A^*}),$$
and hence
$$C_0 \cap A^* \subset C_1 \cap A^* \subset C_2 \cap A^* \subset \cdots \subset S \cap A^* \subset \cdots \subset D_2 \cap A^* \subset D_1 \cap A^* \subset D_0 \cap A^* \quad (\subset A^*).$$
by taking their measures, we have
$$\begin{array}{ccccccccccccccccc}
\widehat{\delta}(C_0) & \leq & \widehat{\delta}(C_1) & \leq & \widehat{\delta}(C_2) & \xrightarrow{\sup} & \widehat{\delta}(\mathrm{Int}S) & \leq & \widehat{\delta}(S) & \leq & \widehat{\delta}(\overline{S}) & \xleftarrow{\inf} & \widehat{\delta}(D_2) & \leq & \widehat{\delta}(D_1) & \leq & \widehat{\delta}(D_0) \\
\| & & \| & & \| & & & & & & & & \| & & \| & & \| \\
\delta(C_0 \cap A^*) & \leq & \delta(C_1 \cap A^*) & \leq & \delta(C_2 \cap A^*) & \leq & \cdots & \leq & \delta(S \cap A^*) & \leq & \cdots & \leq & \delta(D_2 \cap A^*) & \leq & \delta(D_1 \cap A^*) & \leq & \delta(D_0 \cap A^*).
\end{array}$$
This proves that $S \cap A^*$ is Reg-measurable and $\widehat{\delta}(S) = \delta(S \cap A^*)$. $\square$

Example 4.2.6. $A^* \subset \widehat{A^*}$ does NOT have a null-boundary.

memo:

5. memo: To be written

5.1. Function space over the profinite words.

5.2. The L^2-space over the profinite words.

5.3. The word actions on the function space.

5.4. The probabilistic independence of languages.

Acknowledgement. The author would like to thank Mao Hoshino, Koki Sakamoto, Ryoma Sin'ya, and Takao Yuyama for their useful discussions. Of course, all mistakes and any immaturity are due to the author, not to them.

References

[Ber72] Jean Berstel. "Sur la densité asymptotique de langages formels". *Automata, Languages and Programming, Colloquium, Paris, France.* 1972, pp. 345–358.
[Coh13] Donald L. Cohn. *Measure Theory.* 2nd ed. Springer, 2013.
[Doc64] Hoshang P. Doctor. "The categories of Boolean lattices, Boolean rings and Boolean spaces". *Canadian Mathematical Bulletin* 7.2 (1964), pp. 245–252.
[Hor24] Ryuya Hora. "Topoi of automata I: Four topoi of automata and regular languages". *arXiv:2411.06358* (2024).
[Joh82] Peter T. Johnstone. *Stone spaces.* Vol. 3. Cambridge university press, 1982.
[Kai00] Vadim A. Kaimanovich. "The Poisson formula for groups with hyperbolic properties". *Annals of Mathematics* (2000), pp. 659–692.

[Pin22] Jean-Éric Pin. *Mathematical foundations of automata theory (version February 18, 2022)*. 2022.

[Sch65] Marcel Paul Schützenberger. "Sur certains sous-monoïdes libres". *Bulletin de la Société mathématique de France* 93 (1965), pp. 209–223.

[Sin21] Ryoma Sin'ya. "Asymptotic Approximation by Regular Languages". *SOFSEM 2021: Theory and Practice of Computer Science*. Springer, 2021, pp. 74–88.

[SY25] Ryoma Sin'ya and Takao Yuyama. "Measure-Theoretic Aspects of Star-Free and Group Languages". *arXiv:2506.14134* (2025).

[Sto36] Marshall H. Stone. "The theory of representation for Boolean algebras". *Transactions of the American Mathematical Society* 40.1 (1936), pp. 37–111.

[Sto37] Marshall H. Stone. "Applications of the theory of Boolean rings to general topology". *Transactions of the American Mathematical Society* 41.3 (1937), pp. 375–481.

[Ura25] Takeo Uramoto. "Semi-galois Categories I: The classical Eilenberg variety theory". *Journal of Pure and Applied Algebra* 229.2 (2025), p. 107863.

Graduate School of Mathematical Sciences, University of Tokyo, Tokyo, Japan
Email address: hora@ms.u-tokyo.ac.jp
Email address: horaryuya38@gmail.com