

Conditional Densities of Regular Languages

Jakub Kozik ¹

*Jagiellonian University, Institute of Computer Science
ul. Nawojki 11, 30-072 Kraków, Poland*

Abstract

We define a density of a given language S in a given language L as an asymptotic probability that a randomly and uniformly chosen word of length n from L belongs to S . There are languages for which densities do not exist. We show that a problem of checking whether one regular language has a density in another regular language is decidable.

Keywords: asymptotic evaluation, probabilistic density, regular language

1 Introduction

It is known that it is impossible to define an uniformly distributed probabilistic measure on a set of words over a finite alphabet. Instead, we can consider a density of a given language L over a finite alphabet Σ as

$$d(L) = \lim_{n \rightarrow \infty} \frac{\text{Card}(L \cap \Sigma^n)}{\text{Card}(\Sigma^n)},$$

which denotes an asymptotic probability that a randomly and uniformly chosen word of length n belongs to L . Simple example of a language without a density is $((a+b)^2)^*$ (over the alphabet $\Sigma = \{a, b\}$). It is easy to observe that a density is always nonnegative providing it exists and that $d(\Sigma^*) = 1$. Moreover, if disjoint languages L_1, L_2 have densities, then $d(L_1 \cup L_2) = d(L_1) + d(L_2)$. A density is not countably additive on disjoint sets. A notion of a density presented above was studied in [1] and [4] using formal power series. Another approach, based on the Markov chains theory, was presented in [2].

¹ email: jkozik@ii.uj.edu.pl

Given two languages $S \subset L$ such that L contains words of almost every length, we define a (conditional) density of the language S in the language L as

$$d(S|L) = \lim_{n \rightarrow \infty} \frac{\text{Card}(S \cap \Sigma^n)}{\text{Card}(L \cap \Sigma^n)}.$$

In the present paper we investigate conditional densities for regular languages.

Example 1.1 Let us represent the positive integer numbers as words over the binary alphabet $\{1, 0\}$. Then the language of all positive numbers will be $L = 1(1 + 0)^*$, and the language of even numbers will be $S = 1(1 + 0)^*0$. The conditional density of S in L gives us the density of the odd numbers in all numbers, which is $\frac{1}{2}$ as one would expect.

In the example above, the language L has a positive density in the full language.

Example 1.2 If $L = (a + bb)^*$ and $S = a(a + bb)^*$, then we get $d(L) = 0$ and $d(S|L) = \frac{\sqrt{5}-1}{2}$.

In our paper, we focus on the problem of deciding if, for given regular languages $S \subset L$ such that L contains words of almost every length, a density of S in L exists, and – in positive case – whether it is positive. The assumption that $S \subset L$ is not essential, since regular languages are closed under intersection.

2 Preliminary definitions

Let Σ be a finite fixed alphabet and let ε denote the empty word. For any language $L \subset \Sigma^*$ let $L(n)$ denote a number of words from L of length n . We will say that a language S has a (positive) density in a bigger language $L \supset S$ whenever the $\lim_{n \rightarrow \infty} \frac{S(n)}{L(n)}$ exists (and is positive). We will consider the situation where both languages are regular.

Definition 2.1 A complex analytic function $l(x)$ is a generating function of the length distribution of L (a generating function for L), if $l(x) = \sum_{n=0}^{\infty} L(n) \cdot x^n$ for every x for which this series converges. The radius of convergence of $\sum_{n=0}^{\infty} L(n) \cdot x^n$ is denoted by λ_L .

Let us note that $\lambda_L \geq \frac{1}{\text{Card}(\Sigma)}$, since $L \subset \Sigma^*$.

Definition 2.2 A deterministic finite automaton (DFA) is a tuple (Q, E, i, T) , where Q is a finite set of states, $E : Q \times \Sigma \rightarrow Q$ is a transition function (can

be considered as a set of labeled edges), $i \in Q$ is a distinguished initial state and $T \subset Q$ is a set of terminal states.

The function E can be extended in an unambiguous way to a function from $Q \times \Sigma^*$ such that

$$E(q, x \cdot \omega) = E(E(q, x), \omega) \text{ for } q \in Q, x \in \Sigma, \omega \in \Sigma^*$$

$$E(q, \varepsilon) = q \text{ for } q \in Q.$$

The language accepted (recognized) by DFA A (denoted by $L(A)$) is a set of all words ω such that $E(i, \omega) \in T$. In other words, it is a set of words determined by all paths from i to T in the graph (Q, E) .

Definition 2.3 A language $L \subset \Sigma^*$ is regular if and only if there exists a DFA A , such that $L = L(A)$.

Our approach to computing conditional densities is based on generating functions. An introduction to the generating functions theory can be found in [6]. All the functions that we consider are complex functions of one complex variable. A function is rational, if it is a quotient of polynomials with integral coefficients. The proposition below is well-known (see [4], [3]).

Proposition 2.4 Generating functions for regular languages are rational.

Let us note that the poles of a rational function are the roots of its denominator (we will always assume that the denominator of a rational function has no common roots with its numerator). A pole has multiplicity k , if it is a root of multiplicity k of the denominator.

It is easy to observe, that for every regular language L the radius of convergence of the series $\sum_{n=0}^{\infty} L(n) \cdot x^n$ is equal to the least positive real pole of its (rational) generating function.

It is not true that every rational function is a generating function for some regular language. We will cite a characterization of such functions according to [4], which is based on the localization of their poles.

Definition 2.5 A rational function has a dominating root if it is a polynomial or if it has a real positive pole which is strictly smaller than modulus of any other pole.

Definition 2.6 For a positive integer k a rational function has k -dominating root if has a real positive pole of multiplicity k , and any other pole has strictly greater modulus or equal modulus but strictly smaller multiplicity.

The theorem of Soittola (see [4]) characterizes rational functions which are generating functions for regular languages.

Theorem 2.7 (Soittola) A rational function f is a generating function for

some regular language if and only if one of the following cases holds:

- (i) f has a dominating root
- (ii) there exist an integer $\nu > 1$ and rational functions $f_0, \dots, f_{\nu-1}$ with dominating roots such that

$$f(x) = \sum_{i=0}^{\nu-1} x^i \cdot f_i(x^\nu).$$

In the following sections we measure the density of a regular language S in a regular language L . Without loss of generality we assume that $S \subset L \subset \Sigma^*$.

3 A density in the full language

In this section we assume that the set Σ has two elements. Let us start with a simple case when the bigger language is Σ^* . Its generating function is $\frac{1}{1-2x}$ and it has exactly one pole in $\frac{1}{2}$.

Let $S \subset \Sigma^*$ be regular and let $s(x) = \frac{p_s(x)}{q_s(x)}$ be the generating function for S .

Suppose that λ_S (the radius of convergence of series of length distribution of S) is strictly greater than $\frac{1}{2}$. It means that the series $\sum_{n=0}^{\infty} \frac{S(n)}{2^n}$ converges and we obtain

$$\lim_{n \rightarrow \infty} \frac{S(n)}{2^n} = 0.$$

Assume that $\lambda_S = \frac{1}{2}$. The function $s(x)$ can not have a pole of multiplicity greater than 1 on the circle of radius $\frac{1}{2}$, since this would imply that $S(n) > 2^n$ for sufficiently large n (compare Section 5). We consider two cases.

Suppose first, that $s(x)$ has a single pole of multiplicity 1 in $\frac{1}{2}$. Thus $s(x)$ decomposes as follows:

$$s(x) = \frac{\varrho_s}{1-2x} + \frac{\bar{p}_s(x)}{\bar{q}_s(x)},$$

where $\frac{\bar{p}_s(x)}{\bar{q}_s(x)}$ is bounded in the closed disc $|x| \leq \frac{1}{2}$. It implies that

$$S(n) = \varrho_s 2^n + o(2^n)$$

and that the density of the language S is equal to ϱ_s . Let us note that ϱ_s can be also determined by computing

$$\lim_{x \rightarrow \frac{1}{2}} \frac{s(x)}{\frac{1}{1-2x}} = \bar{s}\left(\frac{1}{2}\right),$$

where \bar{s} is obtained by eliminating the factor $1-2x$ from the denominator of s .

In the second case $s(x)$ has many poles of multiplicity 1 on the radius of $\frac{1}{2}$. We will show that $\frac{S(n)}{2^n}$ diverges. Soittolla's theorem provides us with $s_0, \dots, s_{\nu-1}$ with dominating roots and such that

$$(1) \quad s(x) = \sum_{i=0}^{\nu-1} s_i(x^\nu) \cdot x^i.$$

Let us assume, for a contradiction, that S has a density. Then

$$(2) \quad S(n) = \varrho_S 2^n + o(2^n).$$

Therefore for each $i = 0, \dots, \nu - 1$ we have

$$(3) \quad \lim_{k \rightarrow \infty} \frac{S_i(k)}{2^{k \cdot \nu + i}} = \varrho_S.$$

We can deduce that each $s_i(y)$ factorizes as

$$s_i(y) = \frac{\varrho_i}{1 - 2^\nu \cdot y} + r_i(y),$$

where $r_i(y)$ is bounded for $y \leq \frac{1}{2^\nu}$ (if $r_i(y)$ were not bounded than $s_i(y)$ would not have dominating root). Then

$$S(k \cdot \nu + i) = \varrho_i \cdot 2^{k \cdot \nu} + o(2^{k \cdot \nu + i})$$

and from (2) we obtain

$$\varrho_i = \varrho_S \cdot 2^i.$$

Therefore

$$s_i(y) = \frac{\varrho_S \cdot 2^i}{1 - 2^\nu \cdot y} + r_i(y)$$

and substituting to (1) we obtain

$$\begin{aligned} s(x) &= \sum_{i=0}^{\nu-1} s_i(x^\nu) \cdot x^i = \sum_{i=0}^{\nu-1} x^i \cdot \left(\frac{\varrho_S \cdot 2^i}{1 - (2x)^\nu} + r_i(x^\nu) \right) \\ &= \varrho_S \frac{\sum_{i=0}^{\nu-1} (2x)^i}{1 - (2x)^\nu} + \sum_{i=0}^{\nu-1} x^i \cdot r_i(x^\nu) \\ &= \frac{\varrho_S}{1 - 2x} + r(x), \end{aligned}$$

where $r(x) = \sum_{i=0}^{\nu-1} x^i \cdot r_i(x^\nu)$ is bounded for $|x| \leq \frac{1}{2}$. Which shows that $s(x)$ has at most one pole on the radius $\frac{1}{2}$, contradiction.

We can summarize the above observations in the following proposition:

Proposition 3.1 *If $s(x)$ is a generating function for some regular language S over a binary alphabet, then exactly one of the following is true:*

- (i) $s(x)$ has no poles on the radius $\frac{1}{2}$ and the density of S is 0;

- (ii) $s(x)$ has a pole in $\frac{1}{2}$, and no other pole in the closed disc of the radius of $\frac{1}{2}$ and S has a positive density equal to $\lim_{x \rightarrow \frac{1}{2}} s(x) \cdot (1 - 2x)$;
- (iii) $s(x)$ has more than one pole on the radius of $\frac{1}{2}$ and S has no density.

Let us note that the proposition above solves a problem of finding a conditional density if the bigger language has a positive density in the full language.

4 A language L has a dominating root

Let us assume that the bigger language L has a dominating root. As usual, let λ_L denote the radius of convergence of the series of the length distribution of L . This case is very similar to the one tackled in Section 3.

Proposition 4.1 *Let L be a regular language, $l(x)$ be a generating function for L with dominating root λ_L . Let $S \subset L$ be a regular language and $s(x)$ a generating function for S . Exactly one of the following is true:*

- (i) $s(x)$ has no poles on the radius λ_L and the density of S in L is 0;
- (ii) $s(x)$ has a pole in λ_L , and no other pole in the closed disc of radius λ_L and S has a positive density in L equal to $\lim_{x \rightarrow \lambda} \frac{s(x)}{l(x)}$;
- (iii) $s(x)$ has more than one pole on the radius of λ_L and S has no density in L .

Proof. The cases (i) and (iii) are analogous to the cases from Proposition 3.1 (substituting $\frac{1}{2}$ with λ_L).

In the case (ii) $s(x)$ decomposes as:

$$s(x) = \frac{\varrho_S}{1 - \lambda_L^{-1}x} + \frac{\bar{p}_s(x)}{\bar{q}_s(x)},$$

where $\frac{\bar{p}_s(x)}{\bar{q}_s(x)}$ is bounded for $|x| \leq \lambda_L$. Analogously

$$l(x) = \frac{\varrho_L}{1 - \lambda_L^{-1}x} + \frac{\bar{p}_l(x)}{\bar{q}_l(x)}.$$

We immediately obtain

$$S(n) = \varrho_S \cdot \lambda_L^n + o(\lambda_L^n).$$

Finally

$$L(n) = \varrho_L \cdot \lambda_L^n + o(\lambda_L^n).$$

Therefore

$$d(S|L) = \frac{\varrho_S}{\varrho_L}$$

and

$$\lim_{x \rightarrow \lambda_L} \frac{s(x)}{l(x)} = \lim_{x \rightarrow \lambda_L} \frac{\frac{\varrho_S}{1 - \lambda_L^{-1}x} + \frac{\bar{p}_s(x)}{\bar{q}_s(x)}}{\frac{\varrho_L}{1 - \lambda_L^{-1}x} + \frac{\bar{p}_l(x)}{\bar{q}_l(x)}} = \lim_{x \rightarrow \lambda_L} \frac{\varrho_S + (1 - \lambda_L^{-1}x) \cdot \frac{\bar{p}_s(x)}{\bar{q}_s(x)}}{\varrho_L + (1 - \lambda_L^{-1}x) \cdot \frac{\bar{p}_l(x)}{\bar{q}_l(x)}} = \frac{\varrho_S}{\varrho_L}.$$

The last equality holds since $\frac{\bar{p}_l(x)}{\bar{q}_l(x)}$ and $\frac{\bar{p}_s(x)}{\bar{q}_s(x)}$ are bounded for $|x| < \lambda_L$. \square

5 A language L has k -dominating root

Let L be a regular language and $l(x)$ be a (rational) generating function for L . Let us assume that $l(x)$ has a k -dominating root λ_L . We know that λ_L is a root of the denominator of multiplicity k , therefore

$$l(x) = \frac{C}{(\lambda_L - x)^k} + r(x) = \frac{C \cdot \lambda_L^{-k}}{(1 - \lambda_L^{-1} \cdot x)^k} + r(x),$$

where $\lim_{x \rightarrow \lambda_L} r(x) \cdot (\lambda_L - x)^k = 0$ (i.e. $r(x)$ does not have a pole of multiplicity k on the radius λ_L). Hence

$$\begin{aligned} (4) \quad l(x) &= (C \cdot \lambda_L^{-k}) \cdot \sum_{n=0}^{\infty} \binom{n+k-1}{k-1} \cdot x^n \cdot \lambda_L^{-n} + r(x) \\ &= \frac{C \cdot \lambda_L^{-k}}{(k-1)!} \left(\sum_{n=0}^{\infty} \frac{n^{k-1}}{\lambda_L^n} \cdot x^n + \sum_{n=0}^{\infty} \frac{w(n)}{\lambda_L^n} \cdot x^n \right) + r(x), \end{aligned}$$

where $w(n)$ is a polynomial of degree strictly smaller than $k-1$. Therefore

$$(5) \quad L(n) = \varrho_L \frac{n^{k-1}}{\lambda_L^n} + o(\lambda_L^{-n} \cdot n^{k-1}),$$

where $\varrho_L = \frac{C \cdot \lambda_L^{-k}}{(k-1)!}$.

Proposition 5.1 *Fix arbitrary regular languages $S \subset L$. If the generating function for L has a k -dominating root, then one of the following is true:*

- (i) *the radius of convergence of $s(x)$ is strictly greater than the radius of convergence of $l(x)$ and the density of S in L is 0;*
- (ii) *both radii are equal, but $s(x)$ does not have a pole of multiplicity k on the radius λ_L and the density of S in L is 0.*
- (iii) *both radii are equal, and s has k -dominating root and S has a positive density in L ;*
- (iv) *s has many poles of multiplicity k on the radius λ_L and S does not have a density (i.e. the sequence $\frac{S(n)}{L(n)}$ diverges).*

Proof. The case (i) is analogous to the case (i) from Proposition 4.1. In the case ii, when $s(x)$ does not have a pole on the radius λ_L of multiplicity k , we can find S' such that $S'(n) \geq S(n)$ (for all n) and $S'(n) = \varrho_{S'} \frac{n^{k-2}}{\lambda_L^n} + o(\lambda_L^{-n} \cdot n^{k-2})$. Hence the density of L in S is 0.

Let us consider case (iii) when s has a dominating root of multiplicity k which is equal to the L 's dominating root. Analogously to (5) we have

$$S(n) = \varrho_S \frac{n^{k-1}}{\lambda_L^n} + o(\lambda_L^{-n} \cdot n^{k-1})$$

and therefore the sequence $\frac{S(n)}{L(n)}$ converges to $\frac{\varrho_S}{\varrho_L}$.

In the case (iv) when s has many roots of multiplicity k on the radius λ , we will show that the sequence $\frac{S(n)}{L(n)}$ diverges. Let us assume, for a contradiction, that there exists ϱ_S such that

$$(6) \quad \lim_{n \rightarrow \infty} \frac{S(n)}{L(n)} = \frac{\varrho_S}{\varrho_L}.$$

It means that

$$(7) \quad S(n) = \varrho_S \cdot \frac{n^{k-1}}{\lambda_L^n} + o(n^{k-1} \cdot \lambda_L^{-n}).$$

According to Soittola's Theorem (2.7) there are functions $s_0, \dots, s_{\nu-1}$ such that

$$s(x) = \sum_{i=0}^{\nu-1} s_i(x^\nu) \cdot x^i$$

and each $s_i(x)$ has exactly one pole on its radius of convergence. If some function $s_i(x)$ does not have a pole in λ_L^ν , it would mean that $\frac{S(n)}{L(n)}$ diverges (because at least one has a pole in λ_L^ν). If all radii are greater than λ_L^ν , then the function $s(x)$ does not have a pole in λ_L .

Let us assume that all the radii are equal to λ_L^ν . Therefore, analogously to (4), we obtain

$$\begin{aligned} s_i(x^\nu) &= \frac{S_i}{(\lambda_L^\nu - x^\nu)^k} + r_i(x^\nu) \\ &= \frac{S_i}{(k-1)! \cdot \lambda_L^{\nu \cdot k}} \left(\sum_{n=0}^{\infty} \frac{n^{k-1}}{\lambda_L^{n \cdot \nu}} \cdot x^{n \cdot \nu} + \sum_{n=0}^{\infty} \frac{w_i(n)}{\lambda_L^{n \cdot \nu}} \cdot x^{n \cdot \nu} \right) + r_i(x^\nu), \end{aligned}$$

where w_i is a polynomial of degree strictly smaller than $k-1$, and S_i is a nonnegative real number. Then

$$x^i \cdot s_i(x^\nu) = \frac{S_i}{(k-1)! \cdot \lambda_L^{\nu \cdot k}} \left(\sum_{n=0}^{\infty} \frac{n^{k-1}}{\lambda_L^{n \cdot \nu}} \cdot x^{n \cdot \nu + i} + \sum_{n=0}^{\infty} \frac{w_i(n)}{\lambda_L^{n \cdot \nu}} \cdot x^{n \cdot \nu + i} \right) + r_i(x^\nu) \cdot x^i$$

$$= \frac{S_i \cdot \lambda_L^i}{(k-1)! \cdot \lambda_L^{\nu \cdot k} \cdot \nu^{k-1}} \left(\sum_{n=0}^{\infty} \frac{(\nu n)^{k-1}}{\lambda_L^{n \cdot \nu + i}} \cdot x^{n \cdot \nu + i} + \sum_{n=0}^{\infty} \frac{\overline{w}_i(n)}{\lambda_L^{n \cdot \nu + i}} \cdot x^{n \cdot \nu + i} \right) + r_i(x^\nu) \cdot x^i$$

for some polynomial $\overline{w}_i(n) = w_i(n) \cdot \nu^{k-1}$. Hence, because $(\nu \cdot n)^{k-1} = (\nu \cdot n + i)^{k-1} - v_i(n)$ for some polynomial v_i such that $\deg(v_i) < k-1$, the following equality holds:

$$x^i \cdot s_i(x^\nu) = \frac{S_i \cdot \lambda_L^i}{(k-1)! \cdot \lambda_L^{\nu \cdot k} \cdot \nu^{k-1}} \left(\sum_{n=0}^{\infty} \frac{(\nu n + i)^{k-1}}{\lambda_L^{n \cdot \nu + i}} x^{n \cdot \nu + i} + \sum_{n=0}^{\infty} \frac{\overline{w}_i(n)}{\lambda_L^{n \cdot \nu + i}} x^{n \cdot \nu + i} \right) + r_i(x^\nu),$$

where $\overline{w}_i(n) = w_i(n) \cdot \nu^{k-1} - v_i(n)$ and hence its degree is strictly smaller than $k-1$. Therefore

$$S(\nu \cdot n + i) = \frac{S_i \cdot \lambda_L^i}{(k-1)! \cdot \lambda_L^{\nu \cdot k} \cdot \nu^{k-1}} \cdot \frac{(\nu \cdot n + i)^{k-1}}{\lambda_L^{n \cdot \nu + i}} + o(\lambda_L^{-n \cdot \nu + i} \cdot (\nu \cdot n + i)^{k-1}).$$

Comparing to (7) we obtain

$$(8) \quad S_i = \varrho_S \cdot \lambda_L^{\nu \cdot k - i} \cdot \nu^{k-1} \cdot (k-1)!$$

Since every $s_i(x)$ has a dominating root, we have

$$s(x) = \sum_{i=0}^{\nu-1} x^i \cdot \frac{S_i}{(\lambda_L^\nu - x^\nu)^k} + r(x),$$

where $r(x)$ does not have a pole of multiplicity k on the radius λ and its radius of convergence is not smaller than λ . Then, from (8), we have

$$\begin{aligned} s(x) &= \varrho_S \cdot \lambda_L^{\nu \cdot (k-1)} \cdot \nu^{k-1} \cdot (k-1)! \cdot \sum_{i=0}^{\nu-1} \frac{x^i \cdot \lambda_L^{\nu-i}}{(\lambda_L^\nu - x^\nu)^k} + r(x) \\ &= \varrho_S \cdot \lambda_L^{\nu \cdot (k-1)} \cdot \nu^{k-1} \cdot (k-1)! \cdot \frac{\lambda_L}{\lambda_L - x} \cdot \frac{1}{(\lambda_L^\nu - x^\nu)^{k-1}} + r(x), \end{aligned}$$

which shows that there is exactly one pole of multiplicity k on the radius λ_L , a contradiction. \square

6 L has many poles of multiplicity k on the radius of convergence

Let us consider the case when L has many poles of maximal multiplicity (denoted by k) on its radius of convergence λ_L . Let $l_0, \dots, l_{\nu-1}$ be functions

with dominating roots such that

$$l(x) = \sum_{i=0}^{\nu-1} x^i \cdot l_i(x^\nu).$$

For a regular language $S \subset L$, we decompose $s(x)$ to functions $s_0, \dots, s_{\nu-1}$ such that:

$$s(x) = \sum_{i=0}^{\nu-1} x^i \cdot s_i(x^\nu).$$

For each function l_i we decide whether $\frac{L_i(n)}{S_i(n)}$ converges according to the result of previous sections. This is a necessary condition for the language S to have a density in L . If this condition holds following situations can occur:

- (i) every $\frac{S_i(n)}{L_i(n)}$ converges to 0 and then $d(S|L) = 0$
- (ii) there exist i, j such that $\lim_{n \rightarrow \infty} \frac{S_i(n)}{L_i(n)} = 0$ and $\lim_{n \in \mathbb{N}} \frac{S_j(n)}{L_j(n)} \neq 0$ - in this case there S has no density in L
- (iii) for all i $\lim_{n \in \mathbb{N}} \frac{L_i(n)}{S_i(n)} \neq 0$.

The cases (i) and (ii) are trivial. In the third case we have to compare the limits $\lim_{n \in \mathbb{N}} \frac{L_i(n)}{S_i(n)}$ (which will be called partial densities). S has a density in L if all the limits are equal.

7 Decidability

We present a procedure of deciding whether a regular generating function has a dominating root of multiplicity 1. The theorem below is a consequence of the Tarski theorem [5] on the quantifier elimination in the ordered field of real numbers.

Theorem 7.1 *The first order theory of the ordered field of real numbers is decidable.*

Let $f(x) = \frac{p(x)}{q(x)}$ be the generating function for some regular language ($p(x)$ and $q(x)$ are polynomials with rational coefficients). We can assume that the polynomials $p(x)$ and $q(x)$ have no common roots. Roots of $q(x)$ are poles of f . The function f has a dominating root of multiplicity 1 if and only if q has a real positive root which is strictly smaller than the modulus of every other root. That can be formulated in the first order language over the field of complex numbers in the following way:

$$\exists_x (q(x) = 0 \wedge q'(x) \neq 0 \wedge \forall_y ((q(y) = 0 \wedge |y| \leq |x|) \Rightarrow y = x)).$$

Theorem 2.7 guarantees that such x is a real number. The problem whether the above formula is true for the polynomial q is decidable by the Tarski theorem (we can translate the formula above to an equivalent formula over the reals).

Analogously, for the function $f(x) = \frac{p(x)}{q(x)}$ the property of having a dominating root of multiplicity d can be formulated as follows:

$$\begin{aligned} & \exists_x [q(x) = 0 \wedge q'(x) = 0 \wedge \dots \wedge q^{(d)}(x) \neq 0 \wedge \\ & \wedge \forall_y ((q(x) = 0 \wedge q'(x) = 0 \wedge \dots \wedge q^{(d)}(x) \neq 0 \wedge |y| \leq |x|) \Rightarrow y = x)]. \end{aligned}$$

Since the multiplicity of a dominating root is bounded by a degree of the polynomial $q(x)$, the property of having dominating root of any multiplicity is decidable, and in the positive case it is possible to find that multiplicity. Let us note that for two functions with dominating roots of the same multiplicity it is possible to write a formula which compares those roots.

Using the methods sketched above we can distinguish all the cases from propositions 3.1, 4.1, 5.1 and the cases (i) and (ii) from Section 6. In the remaining case (iii) in Section 6 we have to compare partial limits to decide whether the language S has a density in L . We present a procedure of comparing partial limits that answers this question.

Given two pairs of rational functions with dominating roots $(l_1, s_1), (l_2, s_2)$, we have to check whether $\lim_{n \rightarrow \infty} \frac{L_1(n)}{S_1(n)} = \lim_{n \rightarrow \infty} \frac{L_2(n)}{S_2(n)}$. Without loss of generality we can assume that both limits are not 0. Let $l_1(x) = \frac{p_1^l(x)}{q_1^l(x)(\lambda_1 - x)^{k_1}}$ (respectively $s_1(x) = \frac{p_1^s(x)}{q_1^s(x)(\lambda_1 - x)^{k_1}}$) such that k_1 is multiplicity of the dominating root λ_1 of l_1 . Let us note that

$$\lim_{n \rightarrow \infty} \frac{L_1(n)}{S_1(n)} = \lim_{x \rightarrow \lambda_1} \frac{l_1(x)}{s_1(x)} = \frac{p_1^s(\lambda_1) \cdot q_1^l(\lambda_1)}{p_1^l(\lambda_1) \cdot q_1^s(\lambda_1)}$$

and the function $g_1(x) = \frac{p_1^s(x) \cdot q_1^l(x)}{p_1^l(x) \cdot q_1^s(x)}$ can be computed by eliminating common roots from the numerator and the denominator of $\frac{s_1(x)}{l_1(x)}$. Let $l_2(x) = \frac{p_2^l(x)}{q_2^l(x)(\lambda_2 - x)^{k_2}}$ (resp. $g_1(x) = \frac{p_2^s(x)}{q_2^s(x)(\lambda_2 - x)^{k_2}}$). Analogously we compute the rational function $g_2(x)$ for (l_2, s_2) . We can define an equality of partial densities as:

$$\begin{aligned} & \exists_{\lambda_1} \exists_{\lambda_2} : [\lambda_1 \text{ is a dominating root of multiplicity } k_1 \text{ of } l_1 \wedge \\ & \lambda_2 \text{ is a dominating root of multiplicity } k_2 \text{ of } l_2 \wedge \\ & g_1(\lambda_1) = g_2(\lambda_2)]. \end{aligned}$$

So that equality of all partial densities is decidable as well.

As a result we can distinguish all the cases of the classification described in Sections 4–6 and hence

Theorem 7.2 *It is decidable whether, given two regular languages $S \subset L$, S has a density in L , and in positive case whether that density is 0.*

7.1 Complexity

We assume that languages are given by DFAs. The procedure sketched above can be easily improved to have a complexity of verifying validity of a first order formula in the ordered field of real numbers. Best known algorithms for that problem work in a double exponential time. However, using simpler methods (like the Sturm series) we can solve simple cases (i.e. all except the one when L has many poles of maximal multiplicity on the radius of convergence) in a polynomial time.

8 Balanced densities

In the definitions above we use *length* to measure the complexity of words. We can generalize it by adding a weight to each letter. Let $\Sigma = \{a, b\}$ and ϱ_a and ϱ_b denote weights of the letters. For each word $\omega \in \Sigma^*$ let $|\omega|_x$ denote the number of occurrences of a letter x in ω . We define a **balanced complexity** of the word ω as

$$c_\varrho(\omega) = \varrho_a \cdot |\omega|_a + \varrho_b \cdot |\omega|_b.$$

We define a **balanced density** of $L \subset \Sigma^*$ as

$$d_\varrho(L) = \lim_{n \rightarrow \infty} \frac{\text{Card}(L \cap c_\varrho^{-1}(n))}{\text{Card}(c_\varrho^{-1}(n))}.$$

It is reasonable to choose weights for the letters in such a way that they have no common divisor, to avoid infinitely many fractions of kind $\frac{0}{0}$. To illustrate that the balanced density can be not equivalent to the density, let us consider the language $((a + b)^2)^*$, which does not have a density, but when we assign the following weights to the letters $\varrho_a = 1$, $\varrho_b = 2$ it has a balanced density $\frac{1}{2}$.

The problem of deciding whether a regular language S has a (positive) balanced density in a regular language L is not harder than the one for a density. We can construct languages S' and L' by substituting each occurrence of each letter x in the languages by x^{ϱ_x} (languages S' and L' are still regular). Then we have:

$$d_\varrho(S|L) = d(S'|L').$$

9 Conclusions

We proved that the problem whether one regular language has a (positive) density in other regular language is decidable and can be solved in a double exponential time, provided that the languages are given by a deterministic finite automata. That computational complexity is determined by the complexity of general procedure of quantifier elimination in the ordered field of real numbers. On the other hand, the formulae that are produced by the algorithm presented in this paper are of the special kind and it seems that a better complexity bound can be obtained. However, we anticipate that in general the problem cannot be solved in polynomial time.

In further work we want to focus on densities of unambiguous context-free languages for which generating functions are algebraic.

References

- [1] Berstel J., *Sur la densite de langages formels*, ICALP, 1972, 345-358.
- [2] Bodirsky M., and Gärtner T., and von Oertzen T., and Schwinghammer J., *Efficiently Computing the Density of Regular Languages*, Proceedings of Latin American INformatics (LATIN'04), 262-270.
- [3] Flajolet P., and R. Sedgewick, “Analytic Combinatorics”, URL: <http://algo.inria.fr/flajolet/Publications/AnaCombi1to9.pdf>.
- [4] Salomaa A., and M. Soittola, “Automata: Theoretic Aspects of Formal Power Series”, Springer-Verlag, Berlin, 1978.
- [5] Tarski A., *A decision method for elementary algebra and geometry*, 2nd edition, Univ. Calif. Press, Berkeley, 1951.
- [6] Wilf H., “generatingfunctionology”, second edition. New York: Academic Press, 1993.