ELSEVIER

Discrete Mathematics 139 (1995) 143–154

DISCRETE MATHEMATICS

Random palindromes: multivariate generating function and Bernoulli density

Ph. Dumas[a,*], L. Thimonier[a,b]

[a] *Algorithms Project, INRIA Rocquencourt BP 105, 78153 Le Chesnay Cedex, France*
[b] *Laboratoire de Mathématiques et d'Informatique Fondamentale, Université de Picardie, 80039 Amiens Cedex, France*

Received 10 July 1992; revised 29 November 1993

Abstract

Consider a finite alphabet with a probability distribution $\boldsymbol{p}$. We study the probability $\delta(\boldsymbol{p})$ of obtaining a palindrome in a finite time by independent draws. Using a Mahler equation for an associated generating function, we give a closed-form expression for $\delta(\boldsymbol{p})$. Moreover we describe completely the cases where $\delta(\boldsymbol{p})$ has value less than 1, in connection with the singularities of the generating function. Except for the case of a one or two letters alphabet it is found that $\delta(\boldsymbol{p})$ is always less than 1.

1. Introduction

Consider a finite alphabet $\mathscr{A}$ and a language $\mathscr{L}$ over $\mathscr{A}$. Fix a probability law $\boldsymbol{p}$ on $\mathscr{A}$ and choose randomly and independently letters $w_1, w_2, \ldots, w_t, \ldots$ from $\mathscr{A}$. We look at the first time T when we obtain a word which belongs to $\mathscr{L}$. This time T is a finite integer or infinite. We denote $\delta(\mathscr{L}, \boldsymbol{p})$ the probability that T is finite. By definition [3, 9] the Bernoulli density of $\mathscr{L}$ is the function $\delta(\mathscr{L}, \boldsymbol{p})$ of $\boldsymbol{p}$.

It must be noticed that some classical random allocation problems like the birthday paradox (occurring in collisions in hashing methods) or the coupon collector problem arise in this context [4, p. 297]. There have been only few general attacks for these random words problems until recently, when the use of regular languages equipped with shuffle was proposed [6, 9].

Taking into account that the word obtained at time T has no proper left factor in $\mathscr{L}$, this notion of a density leads us to study the language $\mathscr{F}$ whose elements are the words from $\mathscr{L}$ but without proper prefix in $\mathscr{L}$. We call these words prefix-free words and $\mathscr{F}$ the prefix-free language associated with $\mathscr{L}$.

* Corresponding author. E-mail: Philippe.Dumas@inria.fr

SSDI 0012-365X(94)00130-B

Our basic tool is generating functions: with a language $\mathscr{G}$ we associate its multivariate generating functions

$$G(z) := \sum_{\alpha \in \mathbb{N}^m} g_\alpha z^\alpha.$$

In the expression, m is the number of letters in the alphabet $\mathscr{A} = \{a_1, \ldots, a_m\}$, z is an m-tuple of indeterminates $(z_i)_{1 \leqslant i \leqslant m}$, the indeterminate z_i marks the letter a_i and, for each multi-index α, g_α is the number of words in $\mathscr{G}$ for which the letter a_i occurs α_i times.

Frequently we want to take into account the length of words and we bring in an additional indeterminate u. Then the generating series is

$$G(u,z) := G(uz) = \sum_{\alpha \in \mathbb{N}^m} g_\alpha u^{|\alpha|} z^\alpha.$$

We equip the algebras of formal power series $\mathbb{Q}[[z]]$ and $\mathbb{Q}[[u,z]]$ with the classical ultrametric, which make them complete spaces.

When $\mathscr{F}$ is the prefix-free language associated with $\mathscr{L}$ we have the equality, with $F(u,z)$ the generating function of $\mathscr{F}$,

$$\delta(\mathscr{L}, \boldsymbol{p}) = F(1, \boldsymbol{p})$$

since the word $a_{i_1} \ldots a_{i_T}$ appears with probability $p_{i_1} \cdots p_{i_T}$.

A palindrome is a word $w = w_1 \ldots w_l$ of length l greater than or equal to 2, which is equal to its reversal $\tilde{w} = w_l \cdots w_1$. We now specialize $\mathscr{L}$ to be the language of palindromes and we look for a closed-form expression and some properties of its density $\delta(\boldsymbol{p}) := \delta(\mathscr{L}, \boldsymbol{p})$. To this end we use the language $\mathscr{F}$ of the palindromes without proper prefix in $\mathscr{L}$. So we have $\delta(\boldsymbol{p}) = F(1, \boldsymbol{p})$.

The case $m = 1$ is evident; there is only one prefix-free palindrome, aa.

The case $m = 2$ is well known: $\mathscr{F}$ is a regular language, the prefix-free palindromes over a two letters alphabet being of the form $abb \cdots bba$. The generating series is then rational:

$$F(u,z) = \frac{u^2 z_1^2}{1 - u z_2} + \frac{u^2 z_2^2}{1 - u z_1}.$$

The case $m \geqslant 3$ has been studied by Beauquier and Thimonier but with univariate generating series. Hence they only deal with the uniform probability distribution case. In Part 2 we use a functional equation to derive a closed-form expression of F as a series of rational fractions.

In the very simple case $m = 2$, the density is constant and equal to 1. Hence it is natural to ask if we may have $\delta(\boldsymbol{p}) = 1$ when $m \geqslant 3$. Part 3 is devoted to this question. The expression of $\delta(\boldsymbol{p})$ obtained as a by-product of Part 2 is an alternating sum and we cannot estimate it directly. To achieve our goal we relate the density to the radius of convergence of the power series $F(u, \boldsymbol{p})$. The classification of the singularities of $F(u, \boldsymbol{p})$

according to the location of $\boldsymbol{p}$ gives us a very simple answer: we have $\delta(\boldsymbol{p})=1$ only when $m \leqslant 2$.

2. Multivariate generating series

To obtain an expression of the generating series $\mathscr{F}$ of the language of the prefix-free palindromes, we use a lemma from Beauquier and Thimonier [3].

Lemma 1. *If a palindrome has a prefix in $\mathscr{F}$ with length k, its length is greater than or equal to $2k-1$.*

The application of the lemma requires a few supplementary notations. First if k is an integer, we put

$$z^k := (z_1^k, \ldots, z_m^k)$$

and

$$S_k := \sum_{i=1}^{m} z_i^k.$$

Next we introduce the language $\mathscr{S}$ of the symmetrical words: a word is symmetrical if it is equal to its reversal $\tilde{w} = w_l \cdots w_1$. Accordingly a palindrome is a symmetrical word of length at least 2. We introduce also for each $i = 1 \cdots m$, the language $\mathscr{L}_i$ of the palindromes starting with a_i and the language $\mathscr{F}_i$ of the prefix-free palindromes starting with a_i. We obtain immediately

$$S(z) = \frac{1+S_1}{1-S_2},$$

$$L_i(z) = z_i^2 S(z),$$

$$L(z) = S_2 S(z)$$

and we are looking for

$$F(z) = \sum_{i=1}^{m} F_i(z).$$

The preceding lemma translates into a functional equation à la Mahler [8, p. 132].

Lemma 2. *The generating series F_i of the language of the prefix-free palindromes starting with a_i satisfies a Mahler equation,*

$$L_i(z) = F_i(z) + \left(\frac{1}{z_i} + S(z)\right) F_i(z^2).$$

Proof. If w is a palindrome starting with a_i: (i) either it is prefix free, (ii) or it has a unique proper left factor $v=a_iv'a_i$, which is prefix free and has length k. In the latter case: (a) either w has length $2k-1$ and is written $v=a_iv'a_iv'a_i$, (b) or w has length $2k+l$ ($l\geqslant 0$) and is written vuv, where u is a symmetrical word of length l. This immediately generates the equation above, by translating union and concatenation of languages (operating unambiguously) into sum and product of generating series. $\square$

Proposition 1. *The generating series F_i has an explicit expression,*

$$F_i(z)=\sum_{n\geqslant 0}(-1)^n z_i^{2^n+1}\frac{1+S_{2^n}}{1-S_{2^{n+1}}}\prod_{0\leqslant k<n}\left(1+z_i^{2^k}\frac{1+S_{2^k}}{1-S_{2^{k+1}}}\right).$$

Proof. Let $G_i(z)$ be the series defined by $F_i(z)=z_i^2\,G_i(z)$. This is licit because the letter a_i occurs at least twice in every word of $\mathscr{F}_i$. The series G_i satisfies a fixed-point equation,

$$G_i(z)=S(z)-z_iG_i(z^2)(1+z_iS(z)).$$

The mapping $H(z)\mapsto S(z)-z_iH(z^2)(1+z_iS(z))$ is a contraction in the complete space $\mathbb{Q}[[z]]$, whence the result by iteration. $\square$

By rearranging the preceding expression and summing over i, we can prove the next theorem. It shows clearly the symmetrical aspect of the series F with respect to the indeterminates $z_1,\dots,z_m$. Recall that the indeterminate u labels the length of the words. Moreover $\lg x$ is the logarithm base 2 of x. Last, the support of an integer s, $\operatorname{supp} s$, is the set of the indexes for which the corresponding bit of its binary expansion equals 1. For example the support of $13=2^0+2^2+2^3$ is $\{0,2,3\}$.

Theorem 1. *The generating series F satisfies*

$$F(u,z)=\sum_{s\geqslant 1}(-1)^{\lfloor \lg s\rfloor}u^{s+1}S_{s+1}\prod_{j\in\operatorname{supp} s}\frac{1+u^{2^j}S_{2^j}}{1-u^{2^{j+1}}S_{2^{j+1}}}.$$

This formula allows us to compute the first few terms of F:

$$F(u,z)=S_2u^2+(S_1S_2-S_3)u^3+(S_2^2-S_4)u^4+(S_1S_2^2-S_1S_4-S_2S_3+S_5)u^5+\cdots.$$

Proof of Theorem 1. This result directly follows from Proposition 1. It suffices to apply the formulae

$$\prod_{0\leqslant j<n}(1+a_jx^{2^j})=\sum_{0\leqslant s<2^n}x^s\prod_{k\in\operatorname{supp} s}a_k,$$

$$\sum_{n\geqslant 0}(-1)^n a_nx^{2^n}\prod_{0\leqslant j<n}(1+a_jx^{2^j})=\sum_{s\geqslant 1}(-1)^{\lfloor \lg s\rfloor}x^s\prod_{k\in\operatorname{supp} s}a_k. \qquad \square$$

3. Radius of convergence, Bernoulli density

Until now series were considered as formal power series. Henceforth we think of them as analytic functions. We call $R(\boldsymbol{p})$ the radius of convergence of $F(u,\boldsymbol{p})$. Its value is greater than or equal to 1 since $F(u,\boldsymbol{p})$ is as a generating function in the sense of probability theory.

We use some open disks; we denote $\Delta(0,\rho)$ the open disk with center 0 and radius ρ.

The density of the language of palindromes is the function $\delta:\boldsymbol{p}\mapsto F(1,\boldsymbol{p})$. It is defined on the standard simplex of dimension $m-1$,

$$\mathrm{Simp}_{m-1}: \quad p_i \geqslant 0 \ (i=1\cdots m), \quad \sum_{i=1}^{m} p_i = 1.$$

With a $\boldsymbol{p}$ from Simp_{m-1}, we associate $p_+ = \max_i p_i$, the Euclidean norm $\|\boldsymbol{p}\|$ of $\boldsymbol{p}$ and more generally the norm $\|\boldsymbol{p}\|_{2^k}$ defined by

$$\|\boldsymbol{p}\|_{2^k} = \left(\sum_{i=1}^{m} p_i^{2k}\right)^{1/2^k}.$$

Theorem 2. *Let $\boldsymbol{p}$ be a point of* Simp_{m-1}.

1. *If all the p_i but one are zero ($\boldsymbol{p}$ is an extremal point of* Simp_{m-1}*), we have $\delta(\boldsymbol{p})=1$ and $R(\boldsymbol{p})=+\infty$.*
2. *If only two p_i are nonzero ($\boldsymbol{p}$ is on an edge of* Simp_{m-1} *but is not a vertex), we have $\delta(\boldsymbol{p})=1$ and $R(\boldsymbol{p})=1/p_+$.*
3. *In the other cases ($\boldsymbol{p}$ is not on an edge of* Simp_{m-1}*), we have $\delta(\boldsymbol{p})<1$ and $R(\boldsymbol{p})=1/\|\boldsymbol{p}\|$. Moreover $F(u,\boldsymbol{p})$ extends to the disk $\Delta(0,1/p_+)$ as a meromorphic function, whose poles are all simple and equal to $\omega/\|\boldsymbol{p}\|_{2^k}$, with k a positive integer and ω a 2^kth root of unity.*

Proof. The first two cases of the theorem have been seen in the introduction.

The proof follows from four lemmas. To avoid calculations with infinity, we suppose in the sequel that $\boldsymbol{p}$ is not an extremal point of the simplex. □

Lemma 3. *The radius of convergence of $F(u,\boldsymbol{p})$ satisfies*

$$\frac{1}{\|\boldsymbol{p}\|} \leqslant R(\boldsymbol{p}) \leqslant \frac{1}{p_+}.$$

Proof. The prefix-free palindromes are palindromes, therefore $L(u,\boldsymbol{p})$ is a majorizing series for $F(u,\boldsymbol{p})$. It follows that $R(\boldsymbol{p})$ is greater than or equal to the radius of convergence of

$$L(u,\boldsymbol{p})=u^2\,\|\boldsymbol{p}\|^2\frac{1+u}{1-u^2\,\|\boldsymbol{p}\|^2};$$

that is to say $R(\boldsymbol{p})\geqslant 1/\|\boldsymbol{p}\|$.

Likewise, among the prefix free palindromes, there are those in which there occur only two distinct letters. Their generating function is

$$\Phi(u,\boldsymbol{p})=\sum_{i<j}\frac{u^2p_i^2}{1-up_i}+\frac{u^2p_j^2}{1-up_j}.$$

Hence we have a minorant series for $F(u,\boldsymbol{p})$.

We call H the set of indices i which satisfy $p_i=p_+$ and h the cardinality of H. Using the classical notation $[u^n]f(u)$ for the coefficient of u^n in the power series $f(u)$, we may write

$$[u^n]\Phi(u,\boldsymbol{p})=\sum_{i\neq j}p_i^2p_j^{n-2}\geqslant\sum_{i\neq j,\,j\in H}p_i^2p_j^{n-2}=\sum_{j\in H}(\|\boldsymbol{p}\|^2-p_+^2)p_+^{n-2}$$

and consequently

$$[u^n]F(u,\boldsymbol{p})\geqslant[u^n]\Phi(u,\boldsymbol{p})\geqslant h(\|\boldsymbol{p}\|^2-p_+^2)p_+^{n-2}.$$

Thanks to Hadamard's formula regarding the radius of convergence of power series, we get $1/R(\boldsymbol{p})\geqslant p_+$. □

We have to be more precise on the value of $R(\boldsymbol{p})$. To this goal we associate with $\boldsymbol{p}$ another point of the simplex, $\boldsymbol{p}'=\boldsymbol{p}^2/\|\boldsymbol{p}\|^2$.

Lemma 4. *The density at $\boldsymbol{p}'$ and the radius of convergence at $\boldsymbol{p}$ are related.*

1. *If $\delta(\boldsymbol{p}')=1$ then $R(\boldsymbol{p})=1/p_+$.*
2. *If $\delta(\boldsymbol{p}')<1$ then $R(\boldsymbol{p})=1/\|\boldsymbol{p}\|$ and $F(u,\boldsymbol{p})$ extends as a meromorphic function over the disk $\Delta(0,1/p_+)$. The singularities of this function are the simple poles $\omega/\|\boldsymbol{p}\|_{2^k}$, with $k>0$ and ω a 2^kth root of unity.*

Proof. When $|u|\cdot\|\boldsymbol{p}\|<1$ we may write Lemma 2 as follows:

$$F_i(u,\boldsymbol{p})=-\frac{F_i(u^2,\boldsymbol{p}^2)}{up_i}+\frac{1+u}{1-u^2\,\|\boldsymbol{p}\|^2}(u^2p_i^2-F_i(u^2,\boldsymbol{p}^2)).$$

Summing from $i=1$ to $i=m$, we obtain

$$F(u,\boldsymbol{p})=-H(u,p)+\frac{1+u}{1-u^2\,\|\boldsymbol{p}\|^2}(u^2\,\|\boldsymbol{p}\|^2-F(u^2,\boldsymbol{p}^2)).$$

The functions $u\mapsto H(u,\boldsymbol{p})$ and $u\mapsto F(u^2,\boldsymbol{p}^2)$ are analytic on the disk $\Delta(0,1/\|\boldsymbol{p}\|_4)$, because $|u^2|\,\|\boldsymbol{p}^2\|<1$ iff $|u|\,\|\boldsymbol{p}\|_4<1$, and the disk $\Delta(0,1/\|\boldsymbol{p}\|_4)$ contains the disk $\Delta(0,1/\|\boldsymbol{p}\|)$. Accordingly there are only two possibilities.

In the first case the function of u

$$\frac{u^2\|\boldsymbol{p}\|^2-F(u^2,\boldsymbol{p}^2)}{1-u^2\|\boldsymbol{p}\|^2}$$

is analytic on the disk $\Delta(0,1/\|\boldsymbol{p}\|_4)$, which means that

$$F\left(\frac{1}{\|\boldsymbol{p}\|^2},\boldsymbol{p}^2\right)=1 \quad \text{hence } \delta(\boldsymbol{p}')=1.$$

By recurrence on k, $F(u,\boldsymbol{p})$ is analytic on the disks $\Delta(0,1/\|\boldsymbol{p}\|_{2^k})$ and then on the disk $\Delta(0,1/p_+)$, since the sequence $(\|\boldsymbol{p}\|_{2^k})$ is strictly decreasing and converges to the limit p_+ [7, pp. 15, 26]. As a result the radius of convergence satisfies $R(\boldsymbol{p})\geqslant 1/p_+$. From Lemma 3 we conclude $R(\boldsymbol{p})=1/p_+$ in the case under consideration.

In the second case the application is not analytic on $\Delta(0,1/\|\boldsymbol{p}\|_4)$. We have $\delta(\boldsymbol{p}')<1$ and $R(\boldsymbol{p})=1/\|\boldsymbol{p}\|$ according to Lemma 3. The function $F(u,\boldsymbol{p})$ extends to the disk $\Delta(0,1/\|\boldsymbol{p}\|_4)$ and its extension has two simple poles, $\pm 1/\|\boldsymbol{p}\|$. By recurrence it extends to the disk $\Delta(0,1/p_+)$ as described in the lemma. □

To conclude we need a lemma of independent interest. One may see it as a convexity-like property.

Lemma 5. *Let $\sum_\alpha c_\alpha z^\alpha$ be a power series in m variables with nonnegative coefficients. Assume that the power series in one variable u:*

$$\sum_n u^n \sum_{|\alpha|=n} c_\alpha \boldsymbol{a}^\alpha,$$

has a radius of convergence $R(\boldsymbol{a})>0$ when $\boldsymbol{a}$ lies in a convex set A from $\mathbb{R}_+^m$. Then for $\boldsymbol{a}$ and $\boldsymbol{b}$ in A, we have the inequality

$$\frac{1}{R(\lambda\boldsymbol{a}+(1-\lambda)\boldsymbol{b})}\geqslant\begin{cases}\dfrac{1-\lambda}{R(\boldsymbol{b})} & \text{if } \lambda\in[0,1/2[,\\[2mm] \dfrac{1}{2}\left(\dfrac{1/2}{R(\boldsymbol{a})}+\dfrac{1/2}{R(\boldsymbol{b})}\right) & \text{if } \lambda=1/2,\\[2mm] \dfrac{\lambda}{R(\boldsymbol{a})} & \text{if } \lambda\in]1/2,1].\end{cases}$$

Proof. For $\lambda\in[0,1]$, α a multi-index and $\boldsymbol{a}$, $\boldsymbol{b}\in A$, we have, according to the binomial formula,

$$\begin{aligned}(\lambda\boldsymbol{a}+(1-\lambda)\boldsymbol{b})^\alpha &= \prod_{1\leqslant i\leqslant m}(\lambda a_i+(1-\lambda)b_i)^{\alpha_i}\\ &\geqslant \prod_{1\leqslant i\leqslant m}(\lambda^{\alpha_i}a_i^{\alpha_i}+(1-\lambda)^{\alpha_i}b_i^{\alpha_i})\\ &\geqslant \lambda^{|\alpha|}\prod_{1\leqslant i\leqslant m}a_i^{\alpha_i}+(1-\lambda)^{|\alpha|}\prod_{1\leqslant i\leqslant m}b_i^{\alpha_i}.\end{aligned}$$

We collect these inequalities for all the α such that $|\alpha|=n$ and we use the concavity of the nth root function to obtain

$$\left(\sum_{|\alpha|=n} c_\alpha(\lambda \boldsymbol{a}+(1-\lambda)\boldsymbol{b}^\alpha\right)^{1/n} \geqslant (\lambda^n+(1-\lambda)^n)^{1/n}$$
$$\times\left(\frac{\lambda^n}{\lambda^n+(1-\lambda)^n}\left(\sum_{|\alpha|=n} c_\alpha \boldsymbol{a}^\alpha\right)^{1/n}+\frac{(1-\lambda)^n}{\lambda^n+(1-\lambda)^n}\left(\sum_{|\alpha|=n} c_\alpha \boldsymbol{b}^\alpha\right)^{1/n}\right).$$

According to Hadamard's formula it suffices to compute the upper limit [1] in order to exhibit the formula. □

We arrive at the last step of our proof. In the simplex Simp_{m-1} we consider a radius from the center of gravity $\boldsymbol{eq}_m=(1/m,\ldots,1/m)$, which corresponds to the equally likely case. This radius is parametrically defined by $\boldsymbol{p}=\boldsymbol{eq}_m+t\boldsymbol{q}$, $t\in[0,t_+]$. Here $\boldsymbol{q}$ is a vector which satisfies $\sum_i q_i=0$, $\|\boldsymbol{q}\|=1$. The number t_+ has value $-1/(mq_-)$, where $q_-=\min_i q_i$ to ensure that the point $\boldsymbol{p}=\boldsymbol{eq}_m+t\boldsymbol{q}$ lies on the simplex.

Lemma 6. *If $m\geqslant 3$ the radius of convergence satisfies*

$$R(\boldsymbol{p})=1/\|\boldsymbol{p}\|$$

for arbitrary $\boldsymbol{p}\in[\boldsymbol{eq}_m,\boldsymbol{eq}_m+t_+\boldsymbol{q}[$.

Proof. By restriction to the segment $[\boldsymbol{eq}_m,\boldsymbol{eq}_m+t_+\boldsymbol{q}]$, the three functions of the point $\boldsymbol{p}$, namely $1/R(\boldsymbol{p})$, $\|\boldsymbol{p}\|$, and p_+, give us three functions of t. We call them respectively $f(t)$, $g(t)$ and $h(t)$. We know that $f(t)$ can take only two values: $g(t)$ or $h(t)$. Our goal is to prove that $f(t)=g(t)$ for all $t\in[0,t_+[$.

We introduce the set $C=\{t\in[0,t_+[,\ f(t)=g(t)\}$. We will verify that 0 lies in C and C is both open and closed. As $[0,t_+[$ is connected, it will yield $C=[0,t_+[$, as desired.

First $0\in C$. Indeed we use one more time the fact that prefix free palindromes are palindromes, but some palindromes are not prefix free. Hence $\delta(\boldsymbol{eq}_m)=F(1,\boldsymbol{eq}_m)<L(1,\boldsymbol{eq}_m)$, that is to say

$$\delta(\boldsymbol{eq}_m)<\frac{2}{m-1}\leqslant 1.$$

The last inequality follows from the assumption $m\geqslant 3$. According to Lemma 4, we have $R(\boldsymbol{eq}_m)=1/\|\boldsymbol{eq}_m\|$ since $\boldsymbol{eq}'_m=\boldsymbol{eq}_m$. Consequently $f(0)=g(0)$.

Secondly C is open. Let $t_0\neq 0$ be a point from C. (The treatment of 0 is similar.) According to Lemma 5, we have the inequalities

$$f(t)\geqslant\frac{t}{t_0}g(t_0)\quad\text{if } t\in]t_0/2,t_0[,$$

$$f(t)\geqslant\frac{t_+-t}{t_+-t_0}g(t_0)\quad\text{if } t\in]t_0,(t_0+t_+)/2[.$$

This permits us to define a continuous piecewise linear function $\varphi(t)$ which is a lower bound for $f(t)$ on the segment $[t_0/2, (t_+ - t_0)/2]$. The difference $\varphi(t) - h(t)$ is continuous and takes a positive value at t_0. By continuity it takes positive values on an open interval I around t_0. For t in I, we have $f(t) \geqslant \varphi(t) > h(t)$, therefore $f(t) = g(t)$. Hence I is a subset of C and C is open.

Finally C is closed. Let t_1 be a point adherent to C. We use an integer $N \geqslant 2$, which will be made precise later. Let ε be a positive number that satisfies

$$\varepsilon < \min\left(\frac{t_1}{N}, \frac{t_+ - t_1}{N}, \frac{g(t_1)^2}{2N}\right).$$

We choose a point t_0 from C such that $|t_1 - t_0| < \varepsilon$. In the first place we prove that

$$f(t_1) > \frac{N}{N+1} g(t_0).$$

(i) If $t_1 \leqslant t_0$, we have $t_1 \in]t_0/2, t_0]$ since $|t_1 - t_0| < \varepsilon$ and $\varepsilon < t_1/2 \leqslant t_0/2$. From Lemma 5 we may write $f(t_1) \geqslant (t_1/t_0) g(t_0)$, but $t_0 - t_1 < \varepsilon < t_1/N$ hence $t_1/t_0 > N/(N+1)$. This gives us $f(t_1) > N/(N+1) g(t_0)$.

(ii) If $t_0 < t_1$, we successively derive by the same method $t_1 \in [t_0, (t_0 + t_+)/2[$, $f(t_1) \geqslant (t_+ - t_1)/(t_+ - t_0) g(t_0)$ and again $f(t_1) > N/(N+1) g(t_0)$.

Thereafter we choose N large enough to ensure $f(t_1) > h(t_1)$. Using the fact that $g(t)$ is the Euclidean norm, we have

$$\frac{g(t_0)}{g(t)} = 1 + \frac{t_0 - t}{g(t)^2} + \mathrm{O}((t_0 - t)^2).$$

So we can impose

$$\frac{g(t_0)}{g(t_1)} > 1 - \frac{1}{N+1}$$

if N is large enough, since $|t_0 - t_1| < \varepsilon < g(t_1)^2/(2N)$. Under the assumption we get

$$g(t_0) > \frac{N}{N+1} g(t_1)$$

and further

$$f(t_1) > \left(\frac{N}{N+1}\right)^2 g(t_1).$$

This relation implies $f(t_1) > h(t_1)$ if N is large enough. So $f(t_1) = g(t_1)$, t_1 is a member of C and C is closed. □

Proof of Theorem 2 (*conclusion*). Lemma 6 gives us the theorem. Indeed it shows that $R(\boldsymbol{p}) = 1/\|\boldsymbol{p}\|$ and $\delta(\boldsymbol{p}) < 1$ for all the points $\boldsymbol{p}$ which lie on the interior of a

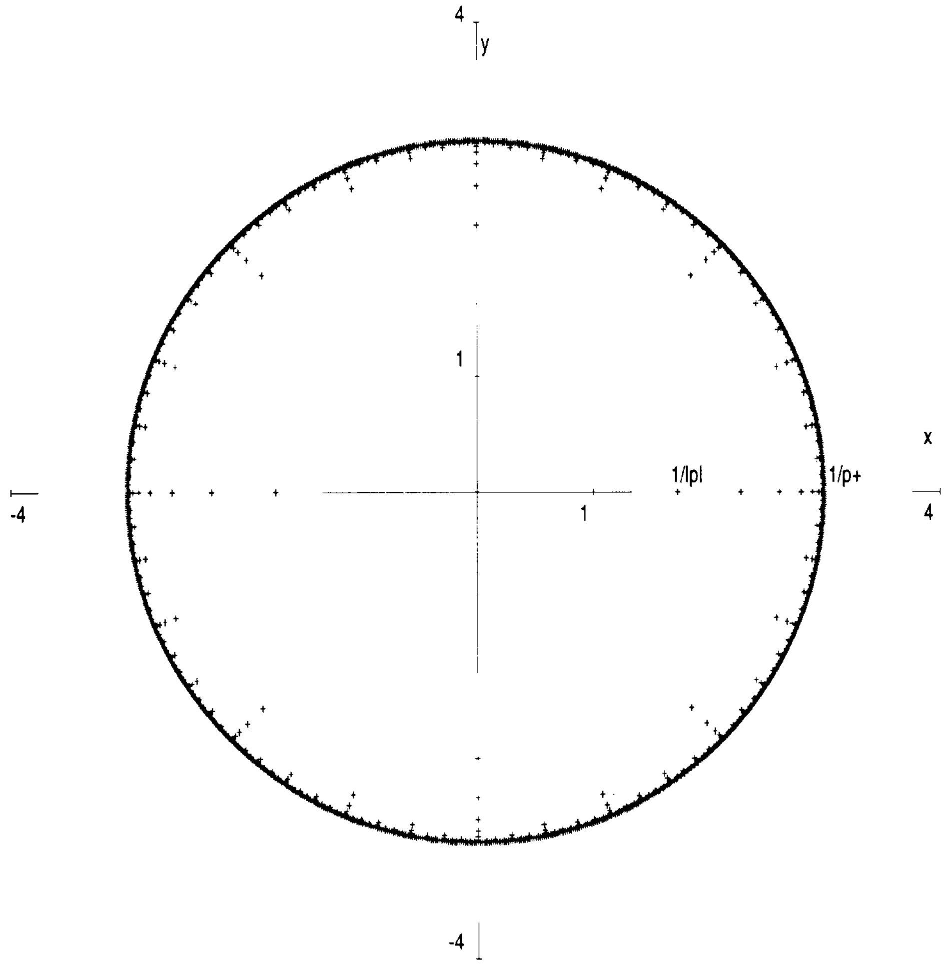

Fig. 1. The poles of $F(u, eq_3)$ agglomerate on the circle of radius 3.

d-dimensional face of the simplex for $d = 2 \cdots m-1$, hence for all the points $\boldsymbol{p}$ which are not on an edge of the simplex. □

4. Conclusion

Except in case m equals 2, the function $F(u, \boldsymbol{p})$ is transcendental because it has infinitely many singularities. As a result the language $\mathscr{F}$ cannot be an unambiguous context-free language [5]. In fact it is not even context free [2].

The singularities are poles which aggregate on the circle $C(0, 1/\|\boldsymbol{p}\|)$. This phenomenon is illustrated by Fig. 1 in the case $\boldsymbol{p} = eq_3$. It is typical of the solutions of a Mahler

Table 1

m	$\delta(\boldsymbol{eq}_m)$
2	1.0000000000
3	0.7215100801
4	0.5415013253
5	0.4299516130
6	0.3555383290
7	0.3027139845
8	0.2633895810
9	0.2330230426
10	0.2088879991

equation in the simplest case when the function is not rational. The idea we used may be employed for a univariate function when the equation gives poles with modulus less than 1. It must be noticed that the asymptotic behavior of the coefficients is immediate in this case. For example the probability that the waiting time T has value n is asymptotically

$$(1-\delta(\boldsymbol{p}'))\,\|\boldsymbol{p}\|^{2\lfloor n/2\rfloor}$$

and in particular the number of prefix-free palindromes of length n is equivalent to

$$(1-\delta(\boldsymbol{eq}_m))m^{2\lfloor n/2\rfloor}.$$

These formulae can be used to compute $\delta(\boldsymbol{eq}_m)$ with the help of the recurrence corresponding to the Mahler equation,

$$u_n = mu_{n-2} - u_{\lceil n/2\rceil} \quad (n \geqslant 4).$$

For the first few values of m we obtain the following values (Table 1) of $\delta(\boldsymbol{eq}_m)$ which represents the probability to obtain a palindrome in a finite time with a uniform probability distribution over a m letters alphabet.

We can also get $\delta(\boldsymbol{p})$ with a great accuracy since the expansion of F that we obtained converges very rapidly. For example we can compute $\delta(\boldsymbol{p})$ in the case $\boldsymbol{p}=(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$ using the approximation,

$$\delta(\boldsymbol{p}) \simeq \sum_{i=1}^{m}\sum_{n=0}^{N}(-1)^n p_i^{2^n+1}\frac{1+S_{2^n}}{1-S_{2^{n+1}}}\prod_{0\leqslant k<n}\left(1+p_i^{2^k}\frac{1+S_{2^k}}{1-S_{2^{k+1}}}\right).$$

Taking only $N=7$, we get a result which is exact to 38 digits,

$$\delta(1/2, 1/4, 1/4) \simeq 0.75241112528971363575197933398903183045.$$

The path we took to prove Theorem 2 may seem rather indirect; however δ takes values arbitrary near 1 and the elementary methods, which we have employed, obliged us to some tricks. As an extension of our results, we can prove that almost surely the process gives palindromes only a finite number of times if $m \geqslant 2$, thanks to the Borel–Cantelli lemma.

Acknowledgement

We wish to acknowledge the contribution of Philippe Flajolet in offering valuable suggestions and stimulating comments during the course of this work.

References

[1] T.M. Apostol, Mathematical Analysis (Addison-Wesley, 1974).
[2] J. Beauquier, Prefix and perfect languages, in: G. Ausiello and M. Protasi, eds., Trees in Algebra and Programming, Lecture Notes in Computer Science, Vol. 159 (Springer, Berlin 1983) 129–140. Proc. CAAP'83, L'aquila (1983).
[3] J. Beauquier and L. Thimonier, Prefix-free words of length n over m letters: two-sided well-balanced parentheses and palindromes, in: P.L.G. Labelle, ed., Colloque de combinatoire énumérative 1985, Lecture Notes in Mathematics Vol. 1234, (Springer, Berlin, 1986) 27–33, 385.
[4] L. Comtet, Advanced Combinatorics, (Reidel, Dordrecht, 1974).
[5] P. Flajolet, Analytic models and ambiguity of context-free languages, Theoret. Comput. Sci. 49 (1987) 283–309.
[6] P. Flajolet, D. Gardy, and L. Thimonier, Probabilistic languages and random allocations, in: T. Lepistö and A. Salomaa, eds., Automata, Languages and Programming, Lecture Notes in Computer Science, vol. 317 (Springer, Berlin, 1984) 239–253. Proc. 15th ICALP Colloquium, Tempere, Finland, July 1988.
[7] G.H. Hardy, J.E. Littlewood and G. Polya, Inequalities (Cambridge University Press, Cambridge, 2nd ed. 1952).
[8] K. Mahler, Fifty years as a mathematician, J. Number Theory 14 (1982) 121–155.
[9] L. Thimonier, Fonctions génératrices et mots aléatoires, Thèse d'Etat à l'université ParisXI, January 1988.
[10] Ph. Dumas, Récurrences mahlériennes, suites automatiques, études asymptotiques, Thèse à l'université Bordeaux I, September 1993.