View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by AIR Universita degli studi di Milano

Frequency of pattern occurrences in Mozkin words

Massimiliano Goldwurm+, Jean-Guy Penaud♯,
Nasser Saheb-Djahromi♯

+ Dip. Scienze dell'Informazione, Università degli Studi di Milano
Via Comelico 39/41, 20135 Milano – Italy, goldwurm@dsi.unimi.it

♯ Laboratoire Bordelais de Recherche en Informatique, Université Bordeaux I,
351 cours de la Libération, 33405 Talence – France, {penaud,saheb}@labri.fr

June 2005

Abstract

In this work we show that the number of horizontal steps in a Motzkin word of length n, drawn at random under uniform distribution, has a Gaussian limit distribution. We also prove a local limit property for the same random variable which stresses its periodic behaviour. Similar results are obtained for the number of peaks in a word of given length drawn at random from the same language.

1 Introduction

The major problem in pattern statistics is to estimate the frequency of pattern occurrences in a random text. A formal model to define such a statistics is given by a finite alphabte Σ, a language $R \subseteq \Sigma^*$ of patterns and stochastic model $\mathcal{P}$ for the generation of a random word $x \in \Sigma^*$ of length n. The associated statistics is defined as the number of (positions of) occurrences of strings of R in x. This problems has a variety of applications (see for instance [12]) and it has been studied mainly for Markovian models $\mathcal{P}$ [11], or when $\mathcal{P}$ is a rational model defined by a weighted finite automaton over Σ [2, 3]. Gaussian limit distributions have been obtained both in the global and in the local sense for pattern statistics in rational models defined by powers of primitive rational formal series [3]. These results are obtained by applying general criteria for establishing global and local limit distribution of Gaussian type, based on the properties of moment generating functions [9, 6, 3].

In this work we study the same problem assuming a simple algebraic model defined by the traditional language of Motzkin words. We show that the number of horizontal steps in a Motzkin word of length n, drawn at random under uniform distribution, has a Gaussian limit distribution. We also prove a local limit property for the same random variable which stresses its periodic behaviour. Analogously we consider the statistics representing the number of peaks in a Motzkin word of length n, drawn at random under uniform distribution. Also in this case we prove a Gaussian limit distribution and a corresponding local limit property.

The main goal of this note is to apply the general analytic criteria used in [3] for the analysis of pattern statistics in rational models, to a simple algebraic model. The results we obtain are in line with a more general approach to the symbol frequency problem in context-free languages presented in [4] (see also [6, Sec VII]).

2 Gaussian limit distributions

In this section we recall a simple general criterion to prove that a sequence of random variables has a Gaussian limit distribution.

Consider an nonnegative integer random variable (r.v.) X, i.e. a random variable taking on values in $\mathbb{N}$, and for every $j \in \mathbb{N}$, set $p_j = \Pr\{X = j\}$. The moment generating function of X is defined as

$$\Psi_X(z) = \mathbb{E}(e^{zX}) = \sum_{k\in\mathbb{N}} p_k e^{zx_k} ,$$

where z is a complex variable. This function is related to the moments of X; in particular we have $\mathbb{E}(X) = \Psi'_X(0)$, $\mathbb{E}(X^2) = \Psi''_X(0)$. Moreover, it can be used to show convergence in distribution: given a sequence of random variables $\{X_n\}_n$ and a random variable X, if Ψ_{X_n} and Ψ_X are defined all over $\mathbb{C}$ and $\Psi_{X_n}(z)$ tends to $\Psi_X(z)$ for every $z \in \mathbb{C}$, then X_n converges to X in distribution (see for instance [7] or [6]).

The characteristic function of X is the restriction of $\Psi_X(z)$ to the immaginary axis, that is the function $\Psi_X(i\theta)$, where $\theta \in \mathbb{R}$. Such a function is well-defined for every $\theta \in \mathbb{R}$ and, as we deal with integer r.v.'s, it is periodic of period 2π.

Moment generating functions can be also used to prove Gaussian limit laws. The following property is a simplification of the so called "quasi–power" theorem introduced in [9] and implicitly used in the previous literature [1] (see also Section IX.5 in [6]).

Theorem 1 *Let $\{X_n\}$ be a sequence of non-negative integer random variables and assume the following conditions hold true:*

C1 *There exist two functions $r(z)$, $y(z)$, both analytic at $z = 0$ where they take the value $r(0) = y(0) = 1$, and a positive constant c, such that for every $|z| < c$*

$$\Psi_{X_n}(z) = r(z) \cdot y(z)^n \left(1 + O(n^{-1})\right); \qquad (1)$$

C2 *The constant $\sigma = y''(0) - (y'(0))^2$ is strictly positive (variability condition).*

Also set $\mu = y'(0)$. Then $\frac{X_n - \mu n}{\sqrt{\sigma n}}$ converges in distribution to a normal random variable of mean 0 and variance 1, i.e., for every $x \in \mathbb{R}$

$$\lim_{n\longrightarrow+\infty} Pr\left\{\frac{X_n - \mu n}{\sqrt{\sigma n}} \le x\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt .$$

The main advantage of this theorem, with respect to other classical statements of this kind, is that it does not require any condition of independence concerning the random variables X_n. For instance, the standard central limit theorems assume that each X_n is a partial sum of the form $X_n = \sum_{j\le n} U_j$, where the U_j's are independent random variables [7].

We recall that the convergence in law of a sequence of r.v.'s $\{X_n\}$ does not yield an approximation of the probability that X_n has a specific value. Theorems concerning approximations for expressions of the form $\Pr\{X_n = x\}$ are usually called *local limit theorems*. A typical example is given by the so-called de Moivre–Laplace Local Limit Theorem [7], which intuitively states that, for a sequence of binomial random variables $\{X_n\}$, up to a factor $\Theta(1/\sqrt{n})$ the probability that X_n takes on a value x approximates a Gaussian density at x.

As for convergence in distribution, also for local limit properties general criteria can be established and several of them appear in the literature. For instance a theorem of this kind is given in [6, Sect. IX.9] and a deeper one is presented in [10]. In this work we use the following result, which yields a natural extention of Theorem 1 to local limit properties of lattice random variables.

We recall that, given $d, \rho \in \mathbb{N}$ such that $0 \le \rho < d$, a lattice random variable X of period d and initial value ρ is an integer r.v. with values in the set $\{x \in \mathbb{Z} \mid x \equiv \rho(\bmod d)\}$. It is well-known that an integer random variable X is a lattice r.v. of period d if and only if $\Psi_X(i2\pi/d) = 1$.

Theorem 2 *[3, Th. 13] Given a positive integer d, let $\{X_n\}$ be a sequence of lattice random variables of period d such that, for every n, X_n takes on values in the interval $[0, n]$ and has initial value ρ_n, for some integer $0 \leq \rho_n < d$. Let Conditions* **C1** *and* **C2** *of Theorem 1 hold true and let μ and σ be the positive constants defined therein. Moreover assume the following property:*

C3 *For all $0 < t < \pi/d$* $$\lim_{n\to+\infty} \left\{ \sqrt{n} \sup_{|\theta|\in[t,\pi/d]} |\Psi_{X_n}(i\theta)| \right\} = 0$$

Then, as n grows to $+\infty$ the following relation holds uniformly for every $j = 0, 1, \ldots, n$.

$$Pr\{X_n = j\} = \begin{cases} \dfrac{de^{-\frac{(j-\mu n)^2}{2\sigma n}}}{\sqrt{2\pi\sigma n}} \cdot (1 + o(1)) & \text{if } j \equiv \rho_n \text{ (mod } d) \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Condition **C3** states that, for every constant $0 < t < \pi/d$, as n grows to $+\infty$, the value $\Psi_{X_n}(i\theta)$ is of the order $o(n^{-1/2})$ uniformly with respect to $\theta \in [-\pi/d, -t] \cup [t, \pi/d]$. Note that relation (2) is meaningful for j lying in an interval $\mu n - a\sqrt{\sigma n} \leq j \leq \mu n + a\sqrt{\sigma n}$, where $a > 0$ is a constant.

Also observe that if $d = 1$, i.e. the values of X_n have no periodicity, then relation (2) reduces to state that, as $n \to +\infty$,

$$\Pr\{X_n = j\} = \frac{e^{-\frac{(j-\mu n)^2}{2\sigma n}}}{\sqrt{2\pi\sigma n}} (1 + \mathrm{o}(1)) \tag{3}$$

holds uniformly for every $j = 0, 1, \ldots, n$. Moreover, if the X_n's are independent binomial random variables of parameter n, p, then Theorem 2 coincides with the classical de Moivre-Laplace Local Limit Theorem.

3 Symbol frequency in a Motzkin word

Let us consider the language $L \subseteq \{a, b, \bar{b}\}$ defined by the grammar

$$S = \varepsilon + aS + bS\bar{b}S \tag{4}$$

We want to study the frequency of of occurrences of a in a word drawn at random under uniform distribution in the set of all strings of length n in L. To this end we first study the asymptotic behaviour of the associated single and bivariate sequences $\{m_n\}$ and $\{r_{nk}\}$ defined by

$$m_n = \sharp\{\omega \in L \mid |\omega| = n\} \tag{5}$$

and

$$r_{nk} = \sharp\{\omega \in L \mid |\omega| = n, |\omega|_a = k\} \; . \tag{6}$$

The analysis of these sequences is based on the corresponding generating functions and can be achieved by applying analytic methods presented in [5, 6]).

The context-free grammar (4) can be trasformed into an algebraic equation by the map ϕ associating the terminal symbols $a, b, \bar{b}$ with the complex variable z. This equation is given by

$$1 + (z - 1)S + z^2S^2 = 0 \tag{7}$$

and it implicitely defines an algebraic function S of the variable z. Such a function has a unique branch non-singular at $z = 0$ given by

$$S(z) = \frac{1 - z - \sqrt{(1 + z)(1 - 3z)}}{2z^2} \; , \tag{8}$$

and hence, since grammar (4) is unambiguous, its power series expansion at the point $z = 0$ is $S(z) = \sum_{n=1}^{+\infty} m_n z^n$.

Then an asymptotic expression for $\{m_n\}$ can be obtained by applying a well-known analytic method (here called *transfer method*) based on the behaviour of generating functions near the singularities of smallest modulus [5] (see also [6, Ch. VI]). This method relies on classical Cauchy's formula for Taylor coefficients of an analytic function and on the choice of a particular contour of the corresponding integral. It allows us to translate the expansion of a function around a dominant singularity into an asymptotic expression for its Taylor coefficients. If the function we consider is a branch of an algebraic curve the asymptotic expansion around a singular point is called *Puiseux expansion* and a general procedure is known to derive its main terms from the polynomial equation defining the function [6, Sec. VII.4].

In our case, the branch $S(z)$ has a unique singularity of smallest modulus at $z = 1/3$. At that point $S(z)$ admits the Puiseux expansion

$$S(z) = 3 - 3\sqrt{3}\sqrt{1-3z} + \mathrm{O}\left((1-3z)^{3/2}\right) \ .$$

This allows us to transfer the well-known series expansion

$$(1-u)^{1/2} = 1 - \sum_{n\geq 1} \binom{2n-2}{n-1} \frac{u^n}{n2^{2n-1}} = 1 - \sum_{n\geq 1} \frac{1+\mathrm{O}(1/n)}{2\sqrt{\pi n^3}} u^n \qquad (|u|<1) \tag{9}$$

and get

$$m_n = 3\sqrt{\frac{3}{4\pi}} \cdot \frac{3^n}{n^{3/2}} \left(1 + \mathrm{O}(1/n)\right) \ . \tag{10}$$

Now, let us use the same method to study the generating function of the bivariate sequence $\{r_{nk}\}$. To this end, let us consider the map associating the symbol a with the monomial xz and both b and $\bar{b}$ with z. This map transforms grammar (4) into the algebraic equation

$$z^2 S^2 + (xz-1)S + 1 = 0 \tag{11}$$

which implicitely defines an algebraic function S of the complex variables x, z. It is easy to see that, for every x, the only branch of S that is non-singular at $z = 0$ is

$$S(x,z) = \frac{1 - xz - \sqrt{(1+(2-x)z)(1-(2+x)z)}}{2z^2} \tag{12}$$

Again, since grammar (4) is unambiguous, for every x, $S(x,z)$ admits at the point $z = 0$ the series expansion $S(x,z) = \sum_{n,k} r_{nk} x^k z^n$. Observe that, for every $x \neq 2, -2$, $S(x,z)$ is singular only at the points $z = (x-2)^{-1}$ and $z = (2+x)^{-1}$. In particular, for x near 1, the singularity of smallest modulus is $z = (2+x)^{-1}$, where $S(x,z)$ admits the Puiseux expansion

$$S(x,z) = 2 + x - (2+x)^{3/2}\sqrt{1-(2+x)z} + \mathrm{O}\left((1-(2+x)z)^{3/2}\right)$$

Then, applying the transfer method we get the power series expansion

$$S(x,z) = \sum_{n\geq 0} S_n(x) z^n = 2 + x - (2+x)^{3/2} \left(1 - \sum_{n\geq 1} \frac{(2+x)^n}{2\sqrt{\pi n^3}} \left(1+\mathrm{O}(1/n)\right) z^n\right) \ .$$

This proves the following

Proposition 3 *For every constant x near 1, function $S(x,z)$ admits at $z = 0$ an expansion $S(x,z) = \sum_{n\geq 0} S_n(x)z^n$ such that*

$$S_n(x) = (2+x)^{3/2} \frac{(2+x)^n}{2\sqrt{\pi n^3}} \ (1 + O(1/n)) \ .$$

Now, let us study the limit distribution of the sequence of random variables $\{Y_n\}$, where each Y_n is the number of occurrences of a in a word drawn at random in the set $L \cap \{a, b, \bar{b}\}^n$ under uniform distribution. This means that such a word is generated in the probabilistic model defined by the characteristic series $\chi_L \in \mathbb{R}_+\langle\langle a, b, \bar{b}\rangle\rangle$. Then, for every $n \in \mathbb{N}$ and each $k = 0, 1, \ldots, n$, we have

$$\Pr\{Y_n = k\} = \frac{r_{nk}}{m_n}$$

The moment generating function of Y_n is

$$\Psi_{Y_n}(u) = \sum_{k=0}^{n} \frac{r_{nk} e^{ku}}{m_n} = \frac{S_n(e^u)}{m_n}$$

As a consequence, by Proposition 3 end Equation (10), $\Psi_{Y_n}(u)$ admits at the point $u = 0$ the expansion

$$\Psi_{Y_n}(u) = \left(\frac{2+e^u}{3}\right)^{3/2} \ \left(\frac{2+e^u}{3}\right)^{n} \ (1 + \mathrm{O}(1/n)) \ , \tag{13}$$

which allows us to apply Theorem 1 where

$$y(u) = \left(\frac{2+e^u}{3}\right) \quad \text{and} \quad r(u) = \left(\frac{2+e^u}{3}\right)^{3/2}$$

Clearly, here $r(u)$ is the "positive" branch of the corresponding algebraic function and hence $r(0) = 1$. Also note that the main part of $\Psi_{Y_n}(u)$, i.e. $y(u)^n$, is the moment generating function of the sum of n independent, identically distributed Bernoullian random variables, having success probability $1/3$: an evocative interpretation of this property is that in a long Motzkin word the occurrence of an horizontal step in a given position can be simulated by a simple (biased) coin tossing. Then, applying Theorem 1 we get the following

Theorem 4 *As n tends to $+\infty$ the random variable*

$$\frac{Y_n - \frac{1}{3}n}{\sqrt{\frac{2}{9}n}}$$

has a Gaussian limit distribution of mean value 0 and variance 1.

3.1 Local limit property

To determine a local limit property first observe that Y_n is a lattice random variable of period 2 and initial value $[n]_2$ (and hence $\Psi_{Y_n}(i\pi) = 1$). Our aim is to apply Theorem 2. Note that Equation (13) is a local property of $\Psi_{Y_n}(u)$ for u near 0 and hence it cannot be used to prove Condition **[C3]** of Theorem 2, which concernes the values of $\Psi_{Y_n}(i\theta)$ for $0 \leq \theta \leq 2\pi$.

Proposition 5 *For every $0 < t < \pi/2$ there exists $0 < \varepsilon < 1$ such that as $n \to +\infty$,*

$$\sup_{t\leq|\theta|\leq\pi-t} |\Psi_{Y_n}(i\theta)| = O(\varepsilon^n)$$

Proof. We study the singular points of the branch $S(x, z)$ defined by (12) for $x = e^{i\theta}$, $0 \leq \theta \leq 2\pi$. For every such θ, the singularities of $S(e^{i\theta}, z)$ are at $z = (e^{i\theta} - 2)^{-1}$ and $z = (e^{i\theta} + 2)^{-1}$. We distinguish three cases:

1. $\theta \in [0, \pi/2) \cup (3\pi/2, 2\pi]$,
2. $\theta \in (\pi/2, 3\pi/2)$,
3. $\theta \in \{\pi/2, 3\pi/2\}$.

In the first interval $(e^{i\theta} + 2)^{-1}$ is the singularity of smallest modulus and we can can reason as in the proof of equation (13), getting the relation

$$\Psi_{Y_n}(i\theta) = \left(\frac{2 + e^{i\theta}}{3}\right)^{3/2} \left(\frac{2 + e^{i\theta}}{3}\right)^n (1 + \mathrm{O}(1/n)) \ .$$

Now, given $0 < t < \pi/2$, in the set $\{\theta \in \mathbb{R} \mid t \leq |\theta| < \pi/2\}$ function $|e^{i\theta} + 2|$ attains the maximum value at points $\theta = t$ and $\theta = -t$. This implies

$$\sup_{t \leq |\theta| < \pi/2} |\Psi_{Y_n}(i\theta)| = \mathrm{O}\left(\frac{|e^{it} + 2|}{3}\right)^n = \mathrm{O}(\varepsilon^n) \tag{14}$$

for some $0 < \varepsilon < 1$.

In the second interval $\theta \in (\pi/2, 3\pi/2)$, $S(e^{i\theta}, z)$ has the smallest singularity in modulus at $z = (e^{i\theta} - 2)^{-1}$, where it admits a Puiseux expansion

$$S(e^{i\theta}, z) = 2 - e^{i\theta} - (e^{i\theta} - 2)^{3/2}\sqrt{1 - (e^{i\theta} - 2)z} + \mathrm{O}\left((1 - (e^{i\theta} - 2)z)^{3/2}\right)$$

Applying the transfer method from the previous equation we get

$$S_n(e^{i\theta}) = (e^{i\theta} - 2)^{3/2} \frac{(e^{i\theta} - 2)^n}{2\sqrt{\pi n^3}} (1 + \mathrm{O}(1/n))$$

and hence

$$\Psi_{Y_n}(i\theta) = \frac{S_n(e^{i\theta})}{m_n} = \left(\frac{e^{i\theta} - 2}{3}\right)^{3/2} \left(\frac{e^{i\theta} - 2}{3}\right)^n (1 + \mathrm{O}(1/n)) \ .$$

Observe that, in the set $\{\theta \in \mathbb{R} \mid \pi/2 < |\theta| \leq \pi - t\}$, $|e^{i\theta} - 2|$ takes on the maximum value at $\theta = \pi - t$ and $\theta = -\pi + t$. As a consequence, for some $0 < \varepsilon < 1$, we have

$$\sup_{\pi/2 < |\theta| \leq \pi - t} |\Psi_{Y_n}(i\theta)| = \mathrm{O}\left(\frac{|e^{i(\pi - t)} - 2|}{3}\right)^n = \mathrm{O}(\varepsilon^n) \tag{15}$$

Finally, for $\theta \in \{\pi/2, 3\pi/2\}$, $S(e^{i\theta}, z)$ has two singularities of modulus $5^{-1/2}$, which implies $S_n(e^{i\theta}) = \mathrm{O}(\sqrt{5})^n$ and hence $|\Psi_{Y_n}(i\theta)| = \mathrm{O}(\sqrt{5}/3)^n = \mathrm{O}(\varepsilon^n)$. $\square$

From Theorem 2 and Proposition 5 we get

Theorem 6 *As n grows to $+\infty$ the following relation holds uniformly for every $j = 0, 1, \ldots, n$:*

$$Pr\{Y_n = j\} = \begin{cases} \dfrac{3\ e^{-9\frac{(j - n/3)^2}{4n}}}{\sqrt{\pi n}} \cdot (1 + o(1)) & \text{if } j \equiv n \ (\text{mod } 2) \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

4 Frequency of peaks in Motzkin words

In this section we determine the limit distribution of the number of peaks in a Motzkin word of length n drawn at random under uniform distribution. Representing the peaks by a terminal symbol t, we consider the language $\mathcal{L} \subseteq \{a, b, \bar{b}, t\}^*$ of all words ω obtained from a Motzkin word $x \in L$ by replacing every factor $b\bar{b}$ by $bt\bar{b}$. The language $\mathcal{L}$ is generated by the context-free grammar

$$S = 1 + aS + bT\bar{b}S \tag{17}$$
$$T = t + aS + bT\bar{b}S \tag{18}$$

We now consider the morphism from the free monoid $\{a, b, \bar{b}, t\}^*$ into the commutative monoid $\{x, z\}^{\oplus}$ associating t with x and the symbols $a, b, \bar{b}$ with z. This transforms the previous grammar into the following system of algebraic equations

$$S = 1 + zS + z^2TS \tag{19}$$
$$T = x + zS + z^2TS \tag{20}$$

Note that (a branch of) the solution $S = S(x, z)$ is the generating function of the bivariate sequence $\{r_{nk}\}$ such that $r_{nk} = \sharp\{\omega \in L \mid |\omega| = n,\ |\omega|_{b\bar{b}} = k\}$. Hence we study such a solution to get an asymptotic evaluation of the corresponding sequence.

Eliminating T from the previous system we get the algebraic equation

$$1 + (z - 1 - z^2 + xz^2)S + z^2S^2 = 0$$

yielding the following non-singular (at $z = 0$) branch

$$S(x, z) = \frac{1 - z + (1 - x)z^2 - \sqrt{(1 + z + (1 - x)z^2)(1 - 3z + (1 - x)z^2)}}{2z^2}$$

For $x = 1$ such a function reduces to $N(z)$ studied in the previous section. So we assume $x \neq 1$. In this case $S(x, z)$ is singular at points α, β, γ and δ given by

$$\alpha = \frac{3 - \sqrt{5 + 4x}}{2(1 - x)}\,, \quad \beta = \frac{3 + \sqrt{5 + 4x}}{2(1 - x)}\,, \quad \gamma = \frac{-1 - \sqrt{4x - 3}}{2(1 - x)}\,, \quad \delta = \frac{-1 + \sqrt{4x - 3}}{2(1 - x)} \tag{21}$$

For x near 1 (and $x \neq 1$), both modulus of β and γ grow to $+\infty$, while α and δ approach $1/3$ and -1, respectively. Hence, for x near 1, $S(x, z)$ has a unique singularity of smallest modulus at $z = \alpha$, where it admits a Puiseux expansion of the form

$$S(x, z) = \frac{1 - \alpha}{2\alpha^2} + \frac{1 - x}{2} - \sqrt{1 - \frac{z}{\alpha}}\,\frac{\sqrt{(1 + \alpha + (1 - x)\alpha^2)\left(1 - \frac{2(1-x)}{3+\sqrt{5+4x}}\alpha\right)}}{2\alpha^2} + \mathrm{O}\left(1 - \frac{z}{\alpha}\right)^{3/2}$$

This leads to the following expression

$$S(x, z) = \frac{1 - \alpha}{2\alpha^2} + \frac{1 - x}{2} - \left(1 - \frac{3 + \sqrt{5 + 4x}}{2}z\right)^{1/2} F(x) + \mathrm{O}\left(1 - \frac{z}{\alpha}\right)^{3/2}$$

where

$$F(x) = (5 + 4x)^{1/4}\ \frac{3 + \sqrt{5 + 4x}}{2}. \tag{22}$$

Therefore, by the transfer method we obtain the following

Proposition 7 *For every constant x near 1, $x \neq 1$, function $S(x,z)$ admits at $z = 0$ an expansion $S(x,z) = \sum_{n\geq 0} S_n(x)z^n$ such that*

$$S_n(x) = \left(\frac{3+\sqrt{5+4x}}{2}\right)^n \frac{F(x)}{2\sqrt{\pi n^3}} \; (1 + O(1/n)) \; ,$$

where $F(x)$ is defined in (22).

Now, let us study the limit distribution of the sequence of random variables $\{V_n\}$, where each V_n is the number of occurrences of the factor $b\bar{b}$ in a word ω drawn at random in the set $L \cap \{a, b, \bar{b}\}^n$ under uniform distribution. Then, for every $n \in \mathbb{N}$ and each $k = 0, 1, \ldots, n$, we have

$$\Pr\{V_n = k\} = \frac{r_{nk}}{m_n}$$

The moment generating function of V_n is

$$\Psi_{V_n}(u) = \sum_{k=0}^{n} \frac{r_{nk}e^{ku}}{m_n} = \frac{S_n(e^u)}{m_n}$$

As a consequence, by Proposition 7 end Equation (10), $\Psi_{V_n}(u)$ admits at the point $u = 0$ the expansion

$$\Psi_{V_n}(u) = \left(\frac{3+\sqrt{5+4e^u}}{6}\right)^n \frac{F(e^u)}{3\sqrt{3}} \; (1 + \mathrm{O}(1/n)) \; . \tag{23}$$

Such an expansion allows us to apply Theorem 1 with

$$y(u) = \frac{3+\sqrt{5+4e^u}}{6} \quad \text{and} \quad r(u) = \frac{F(e^u)}{3\sqrt{3}}$$

Therefore we get the following

Theorem 8 *As n tends to $+\infty$ the random variable*

$$\frac{V_n - \frac{1}{9}n}{\sqrt{\frac{2}{27}n}}$$

has a Gaussian limit distribution of mean value 0 and variance 1.

4.1 Local limit property

To apply Theorem 2 we compare the modulus of singularities of $S(x,z)$ given in equations (21), for complex values of x such that $|x| = 1$.

As α, β, γ and δ have the same denominator, we consider the corresponding numerators, say $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$, $\hat{\delta}$, respectively. It is easy to see that, for $|x| = 1$ and $x \neq 1$, we have $0 < |\hat{\alpha}| \leq 2$, $4 \leq |\hat{\beta}| < 6$, $0 < |\hat{\gamma}| \leq 2\sqrt{2}$, $0 < |\hat{\delta}| \leq 2\sqrt{2}$, where the values of $\hat{\gamma}$ and $\hat{\delta}$ interchange while x moves over the circle of radius 1. Hence $|\hat{\alpha}| < |\hat{\beta}|$ for every $|x| = 1$. Moreover, a direct inspection shows that $|\hat{\alpha}| \leq |\hat{\delta}|$ for every $|x| = 1$, the equality being true only for $x = 1$. Since the comparison with $\hat{\gamma}$ is similar, we can state that α is the unique singularity of smallest modulus of $S(x,z)$ for x varying the required domain.

As a consequence Proposition 7 holds for every complex x such that $|x| = 1$ and $x \neq 1$, and we get

$$\Psi_{V_n}(i\theta) = \left(\frac{3+\sqrt{5+4e^{i\theta}}}{6}\right)^n \frac{F(e^{i\theta})}{3\sqrt{3}} \; (1 + \mathrm{O}(1/n))$$

This proves that, for every $0 < t < \pi$,

$$\sup_{t \leq \theta \leq 2\pi - t} \Psi_{V_n}(i\theta) = \left(\frac{3 + \sqrt{5 + 4e^{it}}}{6} \right)^n \mathrm{O}(1) = \mathrm{O}(\varepsilon^n)$$

for some $0 < \varepsilon < 1$. Therefore, Condition **[C3]** of Theorem 2 holds true in our case with $d = 1$ and we can state the following

Theorem 9 *As n grows to $+\infty$ the following relation holds uniformly for every $j = 0, 1, \ldots, n$:*

$$Pr\{V_n = j\} = \frac{3\sqrt{3}\, e^{-27\frac{(j-n/9)^2}{4n}}}{2\sqrt{\pi n}} \cdot (1 + o(1))$$

References

[1] E. A. Bender. Central and local limit theorems applied to asymptotic enumeration. *J. Combin. Theory Ser. A*, 15:91–111, 1973.

[2] A. Bertoni, C. Choffrut, M. Goldwurm, and V. Lonati. On the number of occurrences of a symbol in words of regular languages. *Theoret. Comput. Sci.*, 302(1-3):431–456, 2003.

[3] A. Bertoni, C. Choffrut, M. Goldwurm, and V. Lonati. Local limit properties for pattern statistics and rational models. Rapporto Interno n. 301-04, Dip. Scienze dell'Informazione, Università degli Studi di Milano, November 2004, available at *http://www.dsi.unimi.it/∼goldwurm/index.html*.

[4] M. Drmota. Systems of functional equations. *Random Structures and Algorithms* 10(1-2):103–124, 1997.

[5] P. Flajolet and A. Odlyzko. Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics* 3:216–240, 1990.

[6] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Preliminary version available at *http://algo.inria.fr/flajolet/Publications/books.html*, April 2005.

[7] B.V. Gnedenko. *The theory of probability* (translated by G. Yankovsky). Mir Publishers - Moscow, 1976.

[8] P. Henrici. *Applied and computational complex analysis*, vol. 1, John Wiley, 1974.

[9] H. K.Hwang. Large deviations for combinatorial distributions. I: central limit theorems, *Ann. Appl. Probab.*, vol. 6(1):297–319, 1996.

[10] H. K.Hwang. Large deviations for combinatorial distributions II. Local limit theorems, *Ann. Appl. Probab.*, 8(1):163–181, 1998.

[11] P. Nicodeme, B. Salvy, and P. Flajolet. Motif statistics. *Theoret. Comput. Sci.*, 287(2):593–617, 2002.

[12] M. Régnier and W. Szpankowski. On pattern frequency occurrences in a Markovian sequence. *Algorithmica*, 22 (4):621–649, 1998.

[13] A. Salomaa and M. Soittola. *Automata-Theoretic Aspects of Formal Power Series*, Springer-Verlag, 1978.