On the decomposition of $k$-noncrossing RNA structures

Emma Y. Jin, Christian M. Reidys

Center for Combinatorics, College of Life Sciences, LPMC-TJKLC, Nankai University, Tianjin 300071, PR China

A R T I C L E   I N F O

Article history:
Received 19 January 2009
Accepted 22 March 2009
Available online 16 July 2009

MSC:
05A16

Keywords:
Singularity analysis
$k$-Noncrossing diagram
$k$-Noncrossing RNA structure
Irreducible substructure
Nontrivial return

A B S T R A C T

A $k$-noncrossing RNA structure can be identified with a $k$-noncrossing diagram over $[n]$, which in turn corresponds to a vacillating tableau having at most $(k - 1)$ rows. In this paper we derive the limit distribution of irreducible substructures via studying their corresponding vacillating tableaux. Our main result proves, that the limit distribution of the numbers of irreducible substructures in $k$-noncrossing, $\sigma$-canonical RNA structures is determined by the density function of a $\Gamma\left(\ln \frac{\tau_k}{\tau_k - 1}, 2\right)$-distribution for some $\tau_k > 1$.

$\copyright$ 2009 Published by Elsevier Inc.

1. Introduction and background

In this paper, we analyze the number of irreducible substructures of $k$-noncrossing, $\sigma$-canonical RNA structures. We prove that the numbers of irreducible substructures of $k$-noncrossing, $\sigma$-canonical RNA structures are, in the limit of long sequence length, given via the density function of a $\Gamma\left(\ln \frac{\tau_k}{\tau_k - 1}, 2\right)$-distribution.

An RNA structure is the helical configuration of its primary sequence, i.e. the sequence of nucleotides $A$, $G$, $U$ and $C$, together with Watson–Crick $(A-U$, $G-C)$ and $(U-G)$ base pairs. As RNA structure is oftentimes tantamount to its function, it is of key importance. The concept of irreducibility in RNA structures is of central importance since the computation of the minimum free energy (mfe) configuration of a given RNA molecule is determined by its largest, irreducible substructure.

Three decades ago, Waterman [20,28,27,10,29] pioneered the combinatorics of RNA secondary structures, an RNA structure class exhibiting only noncrossing bonds. Secondary structures can readily be identified with Motzkin-paths satisfying some minimum height and plateau-length, see Fig. 1. The
Fig. 1. The phenylalanine tRNA secondary structure, as generated by the computer folding algorithm cross [11], represented as planar graph, diagram and Motzkin-path. The structure has arc-length $\geq 8$ and stack-length $\geq 3$ and uniquely corresponds to a Motzkin-path with minimum height 3 and minimum plateau-length 7.

latter restrictions arise from biophysical constraints due to mfe and the limited flexibility of chemical bonds. It is clear from the particular bijection, that irreducible substructures in RNA secondary structures are closely related to the number of nontrivial returns, i.e. the number of nonendpoints, for which the Motzkin-path meets the $x$-axis.

For Dyck-paths, this question has been studied by Shapiro [6], who showed that the expected number of nontrivial returns of Dyck-paths of length $2n$ equals $\frac{2n-2}{n+2}$. Subsequently, Shapiro and Cameron [1] derived expectation and variance of the number of nontrivial returns for generalized Dyck-paths from $(0,0)$ to $((t+1)n,0)$

$$
\mathbb{E}[\xi_t] = \frac{2n - 2}{tn + 2} \quad \text{and} \quad \mathbb{V}[\xi_t] = \frac{2tn(n - 1)((t + 1)n + 1)}{(tn + 2)^2(tn + 3)}.
$$

The bijection between a Dyck-path of length $2n$ and a unique triangulation of the $(n+2)$-gon, due to Stanley [24], implies a combinatorial proof for $\mathbb{E}[\xi_1]$. An alternative approach is to employ the Riordan matrix [22], an infinite, lower triangular matrix $L = (l_{n,k})_{n,k \geq 0} = (g,f)$, where $g(z) = \sum_{n \geq 0} g_n z^n$, $f(z) = \sum_{n \geq 0} f_n z^n$ with $f_0 = 0$, $f_1 \neq 0$, such that $\sum_{n \geq k} l_{n,k} z^n = g(z)f^k(z)$. Clearly,

$$
C(z) = \sum_{n \geq 0} C_n z^n = \frac{1 - \sqrt{1 - 4z}}{2z} \quad \text{where} \quad C_n = \frac{1}{n+1} \binom{2n}{n}.
$$
is the generating function of Dyck-paths and let $\zeta_{n,j}$ denote the number of Dyck-paths of length $2n$ with $j$ nontrivial returns. We consider the Riordan matrix $L = (\zeta_{n,j})_{n,j \geq 0} = (zC(z), zC(z))$ and extract the coefficients $\zeta_{n,j}$ from its generating function $(zC(z))^{j+1}$ by Lagrange inversion. Setting $f(z) = zG(f(z))$ with $f(z) = C(z) - 1$ and $G(z) = (1 + z)^2$, we obtain

$$\zeta_{n,j} = \left(2^n - j - 1\right) \binom{n}{j},$$

where $\sum_{j \geq 0} \zeta_{n,j} = C_n$. From this we immediately compute $\mathbb{E}[\xi_1] = \sum_{j \geq 1} j \cdot \frac{\zeta_{n,j}}{C_n}$ and $\mathbb{V}[\xi_1] = \sum_{j \geq 1} j^2 \cdot \frac{\zeta_{n,j}}{C_n} - \left(\sum_{j \geq 1} j \cdot \frac{\zeta_{n,j}}{C_n}\right)^2$, from which the expression of Eq. (1.1), for $t = 1$ follows.

In Section 3, we consider the bivariate generating function directly, which relates to the Riordan matrix in case of generalized Dyck-paths as follows

$$\sum_{n \geq 0} \sum_{j \geq 0} \zeta_{n,j} w^j z^n = \sum_{j \geq 0} z^{j+1} C(z)^{j+1} w^j = \frac{zC(z)}{1 - wzC(z)}.$$

Our main idea is to derive the bivariate generating function from the Riordan matrix employing irreducible paths and to establish via singularity analysis a discrete limit law. This is done, however, for the far more general class of $\mathcal{C}$-tableaux introduced in Section 2: in Theorem 7 we show that the limit distribution of nontrivial returns for these vacillating tableaux is given in terms of the density function of a $\Gamma(\lambda, r)$-distribution, which is, already for Motzkin-paths, a new result. For restricted Motzkin-paths satisfying specific height and plateau-lengths, the Riordan matrix Ansatz does not work “directly”, since the inductive decomposition of restricted Motzkin-paths is incompatible. Instead we introduce the notion of irreducible paths and express the Riordan matrix in terms of the latter, see Lemma 2. This Ansatz allows us to compute the generating function of irreducible paths via setting one indeterminate of the bivariate generating function to one. The framework developed in Section 3 and Section 4, in fact works as long as the generating function of the particular path-class has a singular expansion and is explicitly known. We have, for instance, for nontrivial returns of Motzkin-paths with height $\geq 3$ and plateau length $\geq 3$: $\lim_{n \to \infty} \mathbb{E}[\eta_n] \approx 5.4526$ and $\lim_{n \to \infty} \mathbb{V}[\eta_n] \approx 20.3180$.

Indeed, RNA structures are far more complex than secondary structures: they exhibit additional, cross-serial nucleotide interactions [21]. These interactions were observed in natural RNA structures, as well as via comparative sequence analysis [30]. They are called pseudoknots, see Fig. 2, and widely occur in functional RNA, like for instance, eP RNA [17] as well as ribosomal RNA [16]. RNA pseudoknots are conserved also in the catalytic core of group I introns. In plant viral RNAs, pseudoknots occur in functional RNA, like for instance, eP RNA [17] as well as ribosomal RNA [16]. RNA pseudoknots mimic tRNA structure and in vitro RNA evolution [25], experiments have produced families of RNA structures with pseudoknot motifs, when binding HIV-1 reverse transcriptase.

Combinatorially, cross serial interactions are tantamount to crossing bonds. To this end, RNA pseudoknot structures have been modeled via $k$-noncrossing diagrams [12], i.e. labeled graphs over the vertex set $[n] = \{1, \ldots, n\}$ with degree $\leq 1$. Diagrams are represented by drawing their vertices $1, \ldots, n$ in a horizontal line and its arcs $(i, j)$, where $i < j$, in the upper half plane. In the following, the degree of $i$ refers to the number of nonhorizontal arcs incident to $i$, i.e. the backbone of the primary sequence is not accounted for. The vertices and arcs correspond to nucleotides and Watson–Crick (A-U, G-C) and (U-G) base pairs, respectively, see Fig. 2.

Natural RNA pseudoknot structures are typically 3-noncrossing [9]. However, relatively high numbers of pairwise crossing bonds occur in natural RNA structures. For instance, the gag-pro ribosomal frame-shift signal of the simian retrovirus-1 [5], which is a 10-noncrossing RNA structure motif, displayed in Fig. 4.

Diagrams are characterized via their maximum number of mutually crossing arcs, $k - 1$, their minimum arc-length, $\lambda$, and their minimum stack-length, $\sigma$. A $k$-crossing is a set of $k$ distinct arcs $(i_1, j_1), (i_2, j_2), \ldots, (i_k, j_k)$ with the property $i_1 < i_2 < \cdots < i_k < j_1 < j_2 < \cdots < j_k$. A diagram without any $k$-crossings is called a $k$-noncrossing diagram. The length of an arc $(i, j)$ is $j - i$ and a stack of length $\sigma$ is a sequence of “parallel” arcs of the form

$$[\zeta_{n,j}] = \left(2^n - j - 1\right) \binom{n}{j},$$

where $\sum_{j \geq 0} \zeta_{n,j} = C_n$. From this we immediately compute $\mathbb{E}[\xi_1] = \sum_{j \geq 1} j \cdot \frac{\zeta_{n,j}}{C_n}$ and $\mathbb{V}[\xi_1] = \sum_{j \geq 1} j^2 \cdot \frac{\zeta_{n,j}}{C_n} - \left(\sum_{j \geq 1} j \cdot \frac{\zeta_{n,j}}{C_n}\right)^2$, from which the expression of Eq. (1.1), for $t = 1$ follows.

In Section 3, we consider the bivariate generating function directly, which relates to the Riordan matrix in case of generalized Dyck-paths as follows

$$\sum_{n \geq 0} \sum_{j \geq 0} \zeta_{n,j} w^j z^n = \sum_{j \geq 0} z^{j+1} C(z)^{j+1} w^j = \frac{zC(z)}{1 - wzC(z)}.$$

Our main idea is to derive the bivariate generating function from the Riordan matrix employing irreducible paths and to establish via singularity analysis a discrete limit law. This is done, however, for the far more general class of $\mathcal{C}$-tableaux introduced in Section 2: in Theorem 7 we show that the limit distribution of nontrivial returns for these vacillating tableaux is given in terms of the density function of a $\Gamma(\lambda, r)$-distribution, which is, already for Motzkin-paths, a new result. For restricted Motzkin-paths satisfying specific height and plateau-lengths, the Riordan matrix Ansatz does not work “directly”, since the inductive decomposition of restricted Motzkin-paths is incompatible. Instead we introduce the notion of irreducible paths and express the Riordan matrix in terms of the latter, see Lemma 2. This Ansatz allows us to compute the generating function of irreducible paths via setting one indeterminate of the bivariate generating function to one. The framework developed in Section 3 and Section 4, in fact works as long as the generating function of the particular path-class has a singular expansion and is explicitly known. We have, for instance, for nontrivial returns of Motzkin-paths with height $\geq 3$ and plateau length $\geq 3$: $\lim_{n \to \infty} \mathbb{E}[\eta_n] \approx 5.4526$ and $\lim_{n \to \infty} \mathbb{V}[\eta_n] \approx 20.3180$.

Indeed, RNA structures are far more complex than secondary structures: they exhibit additional, cross-serial nucleotide interactions [21]. These interactions were observed in natural RNA structures, as well as via comparative sequence analysis [30]. They are called pseudoknots, see Fig. 2, and widely occur in functional RNA, like for instance, eP RNA [17] as well as ribosomal RNA [16]. RNA pseudoknots are conserved also in the catalytic core of group I introns. In plant viral RNAs, pseudoknots mimic tRNA structure and in vitro RNA evolution [25], experiments have produced families of RNA structures with pseudoknot motifs, when binding HIV-1 reverse transcriptase.

Combinatorially, cross serial interactions are tantamount to crossing bonds. To this end, RNA pseudoknot structures have been modeled via $k$-noncrossing diagrams [12], i.e. labeled graphs over the vertex set $[n] = \{1, \ldots, n\}$ with degree $\leq 1$. Diagrams are represented by drawing their vertices $1, \ldots, n$ in a horizontal line and its arcs $(i, j)$, where $i < j$, in the upper half plane. In the following, the degree of $i$ refers to the number of nonhorizontal arcs incident to $i$, i.e. the backbone of the primary sequence is not accounted for. The vertices and arcs correspond to nucleotides and Watson–Crick (A-U, G-C) and (U-G) base pairs, respectively, see Fig. 2.

Natural RNA pseudoknot structures are typically 3-noncrossing [9]. However, relatively high numbers of pairwise crossing bonds occur in natural RNA structures. For instance, the gag-pro ribosomal frame-shift signal of the simian retrovirus-1 [5], which is a 10-noncrossing RNA structure motif, displayed in Fig. 4.

Diagrams are characterized via their maximum number of mutually crossing arcs, $k - 1$, their minimum arc-length, $\lambda$, and their minimum stack-length, $\sigma$. A $k$-crossing is a set of $k$ distinct arcs $(i_1, j_1), (i_2, j_2), \ldots, (i_k, j_k)$ with the property $i_1 < i_2 < \cdots < i_k < j_1 < j_2 < \cdots < j_k$. A diagram without any $k$-crossings is called a $k$-noncrossing diagram. The length of an arc $(i, j)$ is $j - i$ and a stack of length $\sigma$ is a sequence of “parallel” arcs of the form
Fig. 2. The hepatitis delta virus (HDV)-pseudoknot structure and its diagram representation. Top: the structure as folded by cross [11] for $k = 3$ and minimum stack size 4 and the corresponding diagram representation (bottom).

Fig. 3. $k$-Noncrossing diagrams: we display a $4$-noncrossing, arc-length $\lambda \geq 4$ and $\sigma \geq 1$ diagram (top), where the edge set $\{(1, 7), (3, 9), (5, 10)\}$ is a $3$-crossing, the arc $(2, 6)$ has length $4$ and $(5, 10)$ has stack-length $1$. Below, we display a $3$-noncrossing, $\lambda \geq 4$ and $\sigma \geq 2$ (lower) diagram, where $(2, 6)$ has arc-length $4$ and the stack $((2, 6), (1, 7))$ has stack-length $2$.

Fig. 4. The proposed SRV-1 frame-shift [5] is a $10$-noncrossing RNA structure motif.

A subdiagram of a $k$-noncrossing diagram is a subgraph over a subset $M \subset [n]$ of consecutive vertices that starts with an origin and ends with a terminus of some arc. Let $(i_1, \ldots, i_m)$ be a sequence of isolated points, and $(j_1, j_2)$ be an arc. We call $(i_1, \ldots, i_m)$ interior if and only if there exists some arc $(j_1, j_2)$ such that $j_1 < i_1 < i_m < j_2$ holds and exterior, otherwise. By abuse of language, a gap either contains no vertices, or is any exterior sequence of consecutive, isolated vertices. A diagram or subdiagram is called irreducible, if it cannot be decomposed into a sequence of gaps and subdiagrams, see Fig. 5. Accordingly, any $k$-noncrossing diagram can be uniquely decomposed into an alternating sequence of gaps and irreducible subdiagrams. In fact, irreducibility is quite common for natural RNA pseudoknot structures, see for instance, Fig. 6.

We call a $k$-noncrossing diagram with arc-length $\geq 4$ and stack-length $\geq \sigma$, a $k$-noncrossing, $\sigma$-canonical RNA structure, see Fig. 3. We accordingly adopt the notions of gap, substructure and irreducibility for RNA structures.
Subdiagram 1

Subdiagram 2

Gap

Irreducible subdiagram

Fig. 5. Subdiagrams, gaps and irreducibility: Subdiagram 1 is decomposed into the irreducible subdiagram over (1, 6), the gap (7, 8) and the irreducible subdiagram over (9, 12). Subdiagram 2 is decomposed into the irreducible subdiagram over (1, 5), the (empty) gap and the irreducible subdiagram over (6, 8). Finally we display a gap and an irreducible diagram over (1, 12) (bottom).

Fig. 6. mRNA-Ecα: the irreducible pseudoknot structure of the regulatory region of the α ribosomal protein operon.

Our main result is Theorem 6, which proves that the numbers of irreducible substructures are, in the limit of long sequence length, given via the density function of a \( \Gamma(\ln \frac{n}{\tau_k}, 2) \)-distribution. Furthermore, we show that the probability generating function of the limit distribution is given by

\[
q(u) = \frac{u}{(1-u)\tau_k + u}^2,
\]

where \( \tau_k \) is expressed in terms of the generating function of \( k \)-noncrossing, \( \sigma \)-canonical RNA structures [18] and its dominant singularity \( \alpha_k \). In Fig. 7 we compare our analytic results with mfe secondary and 3-noncrossing structures generated by computer folding algorithms [26,11], respectively. The data indicate that already for \( n = 75 \), the limit distribution of Theorem 6 provides a good fit for both structure classes.

The paper is organized as follows: in Section 2 we recall some basic combinatorial background. Of particular importance here is the bijection between \( k \)-noncrossing diagrams and vacillating tableaux of Theorem 1 with at most \( (k-1) \) rows [4]. In Section 3, we present all the key ideas and derive the limit distribution of \( * \)-tableaux. In Section 4, we study the limit distribution of nontrivial returns using the framework developed in Section 3.

2. Some basic facts

A Ferrers diagram (shape) is a collection of squares arranged in left-justified rows with weakly decreasing number of boxes in each row. A standard Young tableau (SYT) is a filling of the squares by numbers which is strictly decreasing in each row and in each column. We refer to standard Young tableaux as Young tableaux, see Fig. 8. A vacillating tableau \( V^n_2 \) of shape \( \lambda \) and length \( 2n \) is a sequence of Ferrers diagrams \( (\lambda^0, \lambda^1, ..., \lambda^{2n}) \) of shapes such that (i) \( \lambda^0 = \emptyset \) and \( \lambda^{2n} = \lambda \), and (ii) \( (\lambda^{2i-1}, \lambda^{2i}) \) is derived from \( \lambda^{2i-2} \), for \( 1 \leq i \leq n \), by one of the following operations. \((\emptyset, \emptyset)\): do nothing twice; \((-\square, \emptyset)\): first remove a square then do nothing; \((\emptyset, +\square)\): first do nothing then adding a square;
For \( n = 75 \) the lhs displays the distribution of irreducible substructures obtained by folding \( 10^4 \) random sequences into their RNA secondary structures [26] (dashed), and the scaled density function of a \( \Gamma(\ln(1.2888), 2) \)-distribution (solid) sampled at the positive integers. The rhs shows this distribution obtained by folding \( 9 \times 10^3 \) random sequences into 3-noncrossing, 3-canonical structures [11] (dashed) and the scaled density function of a \( \Gamma(\ln(0.0170), 2) \)-distribution (solid) derived from Theorem 6.

**Fig. 8.** Ferrers diagram and Young tableau.

<table>
<thead>
<tr>
<th>Ferrers diagram</th>
<th>Young tableau</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Ferrers Diagram" /></td>
<td><img src="image2" alt="Young Tableau" /></td>
</tr>
</tbody>
</table>

\( (\pm \Box, \pm \Box) \): add/remove a square at the odd and even steps, respectively. We denote the set of vacillating tableaux by \( \mathcal{V}^{2n}_2 \). The **RSK-algorithm** is a process of row-inserting elements into a Young tableau. Suppose we want to insert \( q \) into a standard Young tableau of shape \( \lambda \). Let \( \lambda_{i,j} \) denote the element in the \( i \)-th row and \( j \)-th column of the Young tableau. Let \( j \) be the largest integer such that \( \lambda_{i,j-1} \leq q \). (If \( \lambda_{1,1} > q \), then \( j = 1 \).) If \( \lambda_{1,j} \) does not exist, then simply add \( q \) at the end of the first row. Otherwise, if \( \lambda_{1,j} \) exists, then replace \( \lambda_{1,j} \) by \( q \). Next insert \( \lambda_{1,j} \) into the second row following the above procedure and continue until an element is inserted at the end of a row. As a result, we obtain a new standard Young tableau with \( q \) included. For instance, inserting the sequence 5, 2, 4, 1, 6, 3, starting with an empty shape yields the standard Young tableaux displayed in Fig. 10.

The RSK-insertion algorithm has an inverse [4], see Lemma 1 below, which will be of central importance for constructing a vacillating tableaux from a tangled diagram.

**Lemma 1.** Suppose we are given two shapes \( \lambda^n \subseteq \lambda^{i-1} \), which differ by exactly one square. Let \( T_{i-1} \) and \( T_i \) be SYT of shape \( \lambda^{i-1} \) and \( \lambda^i \), respectively. Given \( \lambda^i \) and \( T_{i-1} \), then there exists a unique \( j \) contained in \( T_{i-1} \) and a unique tableau \( T_i \) such that \( T_{i-1} \) is obtained from \( T_i \) by inserting \( j \) via the RSK-algorithm.
Fig. 10. RSK-insertion of the elements 5, 2, 4, 1, 6, 3. The insertion of the above sequence successively constructs a standard Young tableau.

Fig. 11. How Lemma 1 works: given the Young tableau, $T_{i-1}$ and the shape $\lambda_i$, we show how to find the unique $j$ (note here we have $j=x_1=1$) such that $T_{i-1}$ is obtained from $T_i$ by inserting 1 via the RSK-algorithm.

Fig. 12. Tangled diagrams: the first tangled diagram represents the key bonds of the hammerhead ribosome and the second tangle represents key bonds of the catalytic core region of the Group I self-splicing intron [2].

Fig. 13. The inflation of the first tangled diagram displayed in Fig. 12.

In addition, Lemma 1 explicitly constructs this unique $j$ such that $T_{i-1}$ is obtained from $T_i$ by inserting $j$ via the RSK-algorithm, see Fig. 11. Here $r$ denotes the label of the row out of which the entry $x_r$ has been extracted (which in turn is inserted into row $r-1$, leading to the extraction of $x_{r-1}$). As for Fig. 11, we begin by extracting 4 out of the third row. Next we remove the square in this row and replace the entry 2 of the second row (being maximal subject to the condition $x_2 < 4$) by 4. We proceed accordingly, extracting 1 from the first row.

2.1. From diagrams to vacillating tableaux and back

RNA tertiary interactions, in particular the interactions between helical and nonhelical regions give rise to consider tangled diagrams [4]. The key feature of tangled diagrams (tangles) is to allow for two interactions: one being Watson–Crick or G-U and the other being a hydrogen bond for each nucleotide. A tangled diagram, $G_n$, over $[n]$ is obtained by drawing its arcs in the upper halfplane having vertices of degree at most two and a specific notion of crossings and nestings [4]. The inflation, of a tangle is a diagram, obtained by "splitting" each vertex of degree two, $j$, into two vertices $j$ and $j'$ having degree one, see Fig. 13. Accordingly, a tangled diagram with $\ell$ vertices of degree two is expanded into a diagram over $n+\ell$ vertices. Obviously, the inflation has its unique inverse, obtained by simply identifying the vertices $j$, $j'$. By construction, the inflation preserves the maximal number of mutually crossing and nesting arcs [4]. Given a $k$-noncrossing tangle, we can construct a vacillating tableaux, using the following algorithm: starting from right-to-left, we take three types of actions: we either RSK-insert, extract (via Lemma 1) or do nothing, depending on whether we are given a terminus, origin or isolated point of the inflated tangle. In fact, each arc in the inflated structure is considered twice: for its terminus and origin, respectively, see Fig. 14.

To be explicit: we first read the vertices of the inflated tangle from right-to-left. For an inflated tangle having $n$ vertices, we will construct the sequence of Young tableaux $\{T_m\}_{m=0}^{2n}$. To this
We only inflate the vertex adjacent to two arcs, i.e., 2, 4.

We next read the inflated structure from right to left.

end, we set $T_{2n} = \emptyset$. Starting from vertex $i = n, n - 1, \ldots, 1$ we derive inductively the sequence $(T_{2n}, T_{2n-1}, \ldots, T_0)$ as follows:

(1) If the $i$-th vertex, $x$, is a terminus of an arc $(j, x)$: set $T_{2i-1} = T_{2i}$. Next we derive $T_{2i-2}$ from $T_{2i-1}$ by adding $j$ via the RSK-algorithm to $T_{2i-1}$.
Theorem 1. There exists a bijection between \( k \)-noncrossing tangled diagrams and vacillating tableaux of type \( \lambda^{2n} \) having shapes \( \lambda^1 \) with less than \( k \) rows.

The combinatorics of \( k \)-noncrossing RNA pseudoknot structures has been derived in [12,13]. The set (number) of \( k \)-noncrossing, \( \sigma \)-canonical RNA structures is denoted by \( T_{k,\sigma}(n) \) (\( T_{k,\sigma}(n) \)) and let \( f_k(n, \ell) \) denote the number of \( k \)-noncrossing diagrams with arbitrary arc-length and \( \ell \) isolated vertices over \( n \). It follows from Theorem 1, that the number of \( k \)-noncrossing matchings on \( [2n] \) equals the number of walks from \( (k-1, k-2, \ldots, 1) \) to itself that stay inside the Weyl chamber \( x_1 > x_2 > \cdots > x_k-1 > 0 \) with steps \( \pm e_i, 1 \leq i \leq k-1 \). The latter is given by Grabiner and Magyar [8]. It is exactly the situation \( \eta = \lambda = (k-1, k-2, \ldots, 1) \) of Eq. (38) in [8]. As shown in detail in [12, Lemma 2]

\[
\sum_{n \geq 0} f_k(n, 0) \cdot \frac{x^n}{n!} = \det \left[ I_{i-j}(2x) - I_{i+j}(2x) \right]_{i,j=1}^{k-1}, \tag{2.1}
\]

\[
\sum_{n \geq 0} \left( \sum_{\ell=0}^{n} f_k(n, \ell) \right) \cdot \frac{x^n}{n!} = e^x \det \left[ I_{i-j}(2x) - I_{i+j}(2x) \right]_{i,j=1}^{k-1}. \tag{2.2}
\]
where $I_r(2x) = \sum_{j \geq 0} \frac{x^{2j+r}}{2^j (2j+r)!}$ denotes the hyperbolic Bessel function of the first kind of order $r$. In particular for $k = 2$ and $k = 3$ we have the formulas

$$f_2(n, \ell) = \binom{n}{\ell} C_{(n-\ell)/2} \quad \text{and} \quad f_3(n, \ell) = \binom{n}{\ell} \left[ C_{\frac{n-\ell}{2} + 2} C_{\frac{n-\ell}{2}} - C_{\frac{n+\ell+1}{2}}^2 \right]. \quad (2.3)$$

In view of $f_k(n, \ell) = \binom{n}{k} f_k(n-\ell, 0)$, everything can be reduced to matchings, where we have the following situation: there exists an asymptotic approximation of the determinant of hyperbolic Bessel function for general order $k$ due to [15] and employing the subtraction of singularities-principle [19] one can prove [15]

$$\forall k \in \mathbb{N}; \quad f_k(2n, 0) \sim c_k n^{-(k-1)^2+(k-1)/2} (2k-1)^{2n}, \quad \text{where } c_k > 0. \quad (2.4)$$

Let $F_k(z) = \sum_{n \geq 0} f_k(2n, 0) z^{2n}$ denote the generating function of $k$-noncrossing matchings. Setting

$$w_0(x) = \frac{\sqrt{x^{2\sigma-2}}}{1 - x^2 + x^{2\sigma}} \quad \text{and} \quad v_0(x) = 1 - x + w_0(x)x^2 + w_0(x)x^3 + w_0(x)x^4,$$

we can now state the following result [18].

**Theorem 2.** Let $k, \sigma \in \mathbb{N}$, where $k \geq 2$, $\sigma \geq 3$, let $x$ be an indeterminate and $\rho_k = \frac{1}{2(k-1)}$ the dominant, positive real singularity of $F_k(z)$. Then $T_{k,\sigma}(x)$, the generating function of $k$-noncrossing, $\sigma$-canonical structures, is given by

$$T_{k,\sigma}(x) = \frac{1}{v_0(x)} F_k \left( \frac{\sqrt{w_0(x)}x}{v_0(x)} \right). \quad (2.5)$$

Furthermore,

$$T_{k,\sigma}(n) \sim c_k n^{-(k-1)^2-(k-1)/2} \left( \frac{1}{\gamma_{k,\sigma}} \right)^n, \quad \text{for } k = 2, 3, 4, \ldots, 9, \quad (2.6)$$

holds, where $\gamma_{k,\sigma}$ is the minimal positive real solution of the equation $\frac{\sqrt{w_0(x)}x}{v_0(x)} = \rho_k = \frac{1}{2(k-1)}$.

Via Theorem 1, each $k$-noncrossing, $\sigma$-canonical structure corresponds to a unique $\dagger$-tableau. We refer to the set of these tableaux as $\mathcal{E}$-tableaux.

2.3. Singularity analysis

In view of Theorem 2, it is of interest to deduce relations between the coefficients from the equality of generating functions. The class of theorems that deal with this deduction are called transfer-theorems [7]. We use the notation

$$(f(z) = O(g(z)) \quad \text{as } z \to \rho) \quad \iff \quad (f(z)/g(z) \quad \text{is bounded as } z \to \rho) \quad (2.7)$$

and if we write $f(z) = O(g(z))$ it is implicitly assumed that $z$ tends to a (unique) singularity. $[z^n]f(z)$ denotes the coefficient of $z^n$ in the power series expansion of $f(z)$ around 0.
Theorem 3. (See [7].) Let \( f(z), g(z) \) be D-finite functions with unique dominant singularity \( \rho \) and suppose 
\[ f(z) = O(g(z)) \text{ for } z \to \rho. \]
Then we have
\[
[z^n] f(z) = K \left( 1 - O \left( \frac{1}{n} \right) \right) [z^n] g(z),
\]
where \( K \) is some constant.

Theorem 3 and Eq. (2.4) imply
\[
F_k(z) = \begin{cases} 
O((1 - \frac{z}{\rho_k})^{(k-1)^2 + (k-1)/2 - 1} \ln(1 - \frac{z}{\rho_k})) & \text{for } k \text{ odd}, z \to \rho_k, \\
O((1 - \frac{z}{\rho_k})^{(k-1)^2 + (k-1)/2 - 1}) & \text{for } k \text{ even}, z \to \rho_k,
\end{cases}
\]
in accordance with basic structure theorems for singular expansions of D-finite functions [7]. Furthermore, Theorem 3, Eq. (2.4) and the so called supercritical case of singularity analysis [7, VI.9, p. 411], imply the following result tailored for our functional equations [14]. Let \( \rho_k \) denote the dominant positive real singularity of \( F_k(z) \).

Theorem 4. Suppose \( \vartheta_\sigma(z) \) is algebraic over \( K(z) \), analytic for \( |z| < \delta \) and satisfies \( \vartheta_\sigma(0) = 0 \). Suppose further \( \gamma_{k,\sigma} \) is the real unique solution with minimal modulus \( < \delta \) of the two equations \( \vartheta_\sigma(z) = \rho_k \) and \( \vartheta_\sigma(z) = -\rho_k \). Then
\[
[z^n] F_k(\vartheta_\sigma(z)) \sim c_k n^{-(k-1)^2 + (k-1)/2} (\gamma_{k,\sigma}^{-1})^n\).
\]

The below continuity theorem of discrete limit laws will be used in the proofs of Theorems 6 and 7. It ensures that under certain conditions the point-wise convergence of probability generating functions implicates the convergence of its coefficients.

Theorem 5. Let \( u \) be an indeterminate and \( \Omega \) be a set contained in the unit disc, having at least one accumulation point in the interior of the disc. Assume 
\[
P_n(u) = \sum_{k \geq 0} p_{n,k} u^k \quad \text{and} \quad q(u) = \sum_{k \geq 0} q_k u^k
\]
such that \( \lim_{n \to \infty} P_n(u) = q(u) \) for each \( u \in \Omega \) holds. Then we have for any finite \( k \),
\[
\lim_{n \to \infty} p_{n,k} = q_k \quad \text{and} \quad \lim_{n \to \infty} \sum_{j \leq k} p_{n,j} = \sum_{j \leq k} q_j.
\]

3. Irreducible substructures

In the following, we shall identify a \( \mathcal{C} \)-tableaux with the subsequence of even-indexed shapes, i.e. the sequence \( (\lambda^2, \ldots, \lambda^{2n-2}) \). Subsequences of two or more consecutive \( \emptyset \)-shapes result from the elementary move \( (\emptyset, \emptyset) \). For instance, consider the \( \mathcal{C} \)-tableaux

\[
\lambda^0 \quad \emptyset (\emptyset, \emptyset) \quad \lambda^2 \quad \emptyset (\emptyset, +\emptyset) \quad \lambda^4 \quad (\emptyset, +\emptyset) \quad \lambda^6 \quad (\emptyset, \emptyset) \quad \lambda^8 \quad (-\emptyset, \emptyset) \quad \lambda^{10} \quad (-\emptyset, \emptyset) \quad \emptyset
\]

The above tableaux splits at \( \lambda^2 = \emptyset \) into two \( \mathcal{C} \)-subtableaux, i.e.

\[
\lambda^0 \quad \emptyset (\emptyset, \emptyset) \quad \emptyset (\emptyset, +\emptyset) \quad \lambda^2 \quad \emptyset (\emptyset, +\emptyset) \quad \lambda^4 \quad (\emptyset, +\emptyset) \quad \lambda^6 \quad (\emptyset, \emptyset) \quad \lambda^8 \quad (-\emptyset, \emptyset) \quad \lambda^{10} \quad (-\emptyset, \emptyset) \quad \emptyset
\]
We call a sequence of consecutive ∅-shapes of length \((r + 1)\), \((∅, \ldots, ∅)\) a gap of length \(r\). In particular, the empty gap is the ∅-shape of length 0. Theorem 1 implies that these ∅-gaps correspond uniquely to the gaps of diagrams, introduced in Section 2. A ∗-tableaux is a \(C\)-tableaux, with the property \(λ^i \neq ∅\) for \(2 \leq i \leq 2n - 2\). It is evident that a ∗-tableaux corresponds via the bijection of Theorem 1 to an irreducible \(k\)-noncrossing, \(σ\)-canonical RNA structure. For instance,

\[
\begin{array}{cccccccccccc}
λ^0 & λ^2 & λ^4 & λ^6 & λ^8 & λ^{10} & λ^{12} & λ^{14} & λ^{16} & λ^{18} & λ^{20} & λ^{22} \\
∅ & → & ∅ & → & ∅ & → & ∅ & → & ∅ & → & ∅ & → & ∅ \\
\end{array}
\]

splits into the gap \((λ^0, λ^2)\), the ∗-tableaux over \((λ^2, \ldots, λ^{14})\) and the gap \((λ^{14}, \ldots, λ^{20})\). Let \(δ^{(k)}_{n,j}\) denote the number of \(C\)-tableaux of length \(2n\) with less than \(k\) rows, containing exactly \(j\) ∗-tableaux. Furthermore, let

\[
U_k(z, u) = \sum_{j \geq 0} \sum_{n \geq j} δ^{(k)}_{n,j} u^j z^n,
\]

and \(δ^{(k)}_{n} = \sum_{j \geq 0} δ^{(k)}_{n,j}\). We set \(T_k(z) = T_{k,σ}(z) = \sum_{n \geq 0} δ^{(k)}_{n} z^n\) and denote the generating function of ∗-tableaux by \(R_k(z)\).

**Lemma 2.** The bivariate generating function of the number of ∗-tableaux of length \(2n\) with less than \(k\) rows, which contain exactly \(i\) ∗-tableaux, is given by

\[
U_k(z, u) = \frac{1}{1 - u(1 - \frac{1}{1 - z}T_k(z))}.
\]

**Proof.** Since each ∗-tableau can be uniquely decomposed into a sequence of gaps and ∗-tableaux, we obtain for fixed \(j\)

\[
\sum_{n \geq j} δ_{n,j} z^n = R_k(z)^j \left(\frac{1}{1 - z}\right)^{j+1}.
\]

As a result, the bivariate generating function of \(δ_{n,j}\) is given by

\[
U_k(z, u) = \sum_{j \geq 0} \sum_{n \geq j} δ_{n,j} z^n u^j = \sum_{j \geq 0} R_k(z)^j \left(\frac{1}{1 - z}\right)^{j+1} u^j = \frac{1}{1 - z - uR_k(z)}.
\]

Setting \(u = 1\), we derive
Claim 1. \( \xi_n = \sum_{i=1}^{\infty} u_i = \sum_{i=1}^{\infty} v_i = \sum_{i=1}^{\infty} w_i \), respectively. We denote \( \tau_k \) the real positive dominant singularity of \( u_k \) and call a function \( F(z) = v(z) \) subcritical if and only if \( \tau_k < \rho_v \).

Theorem 6. Let \( \alpha_k \) be the real positive dominant singularity of \( T_k(z) \) and \( \tau_k = (1 - \alpha_k)T_k(\alpha_k) \). Then the r.v. \( \xi_n^{(k)} \) satisfies the discrete limit law

\[
\lim_{n \to \infty} \mathbb{P}(\xi_n^{(k)} = i) = q_i \quad \text{where} \quad q_i = \frac{i}{\tau_k^2} \left( \frac{\tau_k - 1}{\tau_k} \right)^{i-1}.
\]

That is, \( \xi_n^{(k)} \) is determined by the density function of a \(\Gamma(\ln \frac{\tau_k}{\tau_k - 1}, 2)\)-distribution. Furthermore, the probability generating function of the limit distribution \( q(u) = \sum_{i \geq 1} q_i u^i \) satisfies \( q(u) = \frac{u}{(1-u)^{\tau_k} + u} \).

Proof. Since \( g(z) = \frac{1}{1-z} \) and \( h(z) = 1 - \frac{1}{1-\alpha_k T_k(z)} \) have nonnegative coefficients and \( h(0) = 0 \), the composition \( g(h(z)) \) is well defined as a formal power series. According to Eq. (3.7) we may express \( U_k(z, u) = g(z) = g(uh(z)) \).

Claim 1. \( h(z) \) has a singular expansion at its dominant singularity \( z = \alpha_k \) and there exists some constant \( \tilde{c}_k > 0 \) such that

\[
h(z) = \begin{cases} 
\tilde{c}_k (1 - \frac{z}{\alpha_k})^{-\mu} (\ln \frac{1}{1-z/\alpha_k})^{-1} (1 + o(1)) & \text{for } k \equiv 1 \mod 2, \\
\tilde{c}_k (1 - \frac{z}{\alpha_k})^{-\mu} (1 + o(1)) & \text{for } k \equiv 0 \mod 2 
\end{cases}
\]

for \( z \to \alpha_k \) and \( \mu = (k-1)^2 + \frac{k-1}{2} - 1 \).
Since \( F_k(z) \) is \( D \)-finite, the composition \( F_k(\vartheta(z)) \) where \( \vartheta(z) = \frac{v_\infty(z)}{v_0(z)} \) and \( \vartheta(0) = 0 \), is also \( D \)-finite [23]. As a result, \( T_k(z) \) is, being a product of the two \( D \)-finite functions \( \frac{1}{v_0(z)} \) and \( F_k(\vartheta(z)) \), \( D \)-finite. Its \( D \)-finiteness guarantees that \( T_k(z) \) has an analytic continuation \( T_k^*(z) \) for which \( T_k(z) = T_k^*(z) \) holds for some simply connected \( \Delta_{\alpha_k} \)-domain [23]. Eq. (3.4) implies \( T_k^*(z) > 0 \) for \( z \in \Delta_{\alpha_k} \), from which we conclude that

\[
h^*(z) = 1 - \frac{1}{(1-z)T_k^*(z)}
\]

is an analytic continuation of \( h(z) \) to \( \Delta_{\alpha_k} \). As for the order of \( h(z) \) at \( z = \alpha_k \), we recall that \( T_k(z) \) is the composition \( F_k(\vartheta(z)) \). We are given the supercritical case of singularity analysis, i.e. the subexponential factors of the asymptotic expressions of \([z^n]T_k(z)\) coincide with those of \([z^n]F_k(z)\). Theorem 4 implies,

\[
T_k(z) = \begin{cases} 
O((1 - \frac{z}{\alpha_k})^\mu \ln(1 - \frac{z}{\alpha_k})), & k \equiv 1 \mod 2, \ z \to \alpha_k, \\
O((1 - \frac{z}{\alpha_k})^\mu), & k \equiv 0 \mod 2, \ z \to \alpha_k.
\end{cases}
\]

Consequently, \( h(z) \) has a singular expansion at \( z = \alpha_k \), given by

\[
h(z) = 1 - \frac{1}{(1-z)T_k(z)}
= \begin{cases} 
\tilde{c}_k(1 - \frac{z}{\alpha_k})^{-\mu}(\ln \frac{1}{1 - \frac{z}{\alpha_k}})^{-1}(1 + o(1)), & k \equiv 1 \mod 2, \ z \to \alpha_k, \\
\tilde{c}_k(1 - \frac{z}{\alpha_k})^{-\mu}(1 + o(1)), & k \equiv 0 \mod 2, \ z \to \alpha_k
\end{cases}
\]

and Claim 1 is proved. Note that Claim 1 and Theorem 3 imply

\[
[z^n]h(z) \sim \begin{cases} 
\tilde{c}_k \frac{n^{\mu-1}}{\Gamma(\mu)\ln(m)}\alpha_k^{-n}(1 + o(1)) & \text{for } k \equiv 1 \mod 2, \\
\tilde{c}_k \frac{n^{\mu-1}}{\Gamma(\mu)}\alpha_k^{-n}(1 + o(1)) & \text{for } k \equiv 0 \mod 2.
\end{cases}
\]

Claim 1 implies that \( U_k(z,u) = g(z)g(uh(z)) \), for \( u \in (0,1) \) has the unique dominant singularity \( \alpha_k \) and a singular expansion. Without loss of generality, we restrict our analysis in the following to the case \( k \equiv 1 \mod 2 \). We consider first \( U_k(z,1) = T_k(z) \). For \( k \equiv 1 \mod 2 \), Theorem 3 implies

\[
[z^n]U_k(z,1) = \tilde{c}_k \alpha_k^{-n}n^{-\mu-1}(1 + o(1)). \tag{3.10}
\]

Second we consider the bivariate generating function \( U_k(z,u) \). For any fixed \( u \in (0,1) \), we write

\[
U_k(z,u) = g(z) \cdot v_u(w(z))
\]

where \( v_u(z) = \frac{z}{z-e^{-z}} \) and \( w(z) = (1-z)T_k(z) \). We focus on the composition \( v_u(w(z)) \) which belongs to the subcritical case of singularity analysis [7, VI.9, p. 411]. See also Proposition IX.1, p. 629, therein. In the subcritical case, the inner function, \( w(z) \) has a singular expansion at its unique dominant singularity having strictly smaller modulus than that of the singularity of the outer function, \( v_u \). The singular expansion of \( v_u(w(z)) \) is then given by combining the regular expansion of \( v_u \) with the singular expansion of \( w(z) \) at \( \alpha_k \). Setting \( w = w(z) \) and \( \tau_k = w(\alpha_k) > 1 \) we compute
\[ U_k(z, u) = g(z) \cdot \frac{w(z)}{w(z) - u(w(z) - 1)} \]

\[ = \frac{g(\alpha_k) \cdot \tau_k}{\tau_k - u(\tau_k - 1)} + g(\alpha_k) \frac{d}{dw} \left( \frac{w}{w - u(w - 1)} \right) \bigg|_{w=\tau_k} (w - \tau_k) + \cdots \]

\[ = \frac{g(\alpha_k) \cdot \tau_k}{\tau_k - u(\tau_k - 1)} + g(\alpha_k) \frac{u}{((1-u)\tau_k + u)^2} (w - \tau_k) \left( 1 + o(1) \right). \]

The transfer theorem, Theorem 3, guarantees

\[ [z^n] U_k(z, u) = g(\alpha_k) u^i \frac{1 - \alpha_k}{1 - \alpha_k + u} \left[ z^n \right] T_k(z) \left( 1 + o(1) \right) \]

\[ = \frac{u}{((1-u)\tau_k + u)^2} \tilde{c}_k \alpha_k^{-n} n^{-\mu - 1} \left( 1 + o(1) \right). \]

We consequently arrive at

\[ \lim_{n \to \infty} \frac{[z^n] U_k(z, u)}{[z^n] U_k(z, 1)} = \frac{u}{((1-u)\tau_k + u)^2} = q(u). \quad (3.11) \]

In view of Eq. (3.11) and

\[ [u^i] q(u) = i \frac{\tau_k - 1}{\tau_k} \left( \frac{\tau_k}{\tau_k} - i \right)^{i-1} = q_i, \]

Theorem 5 implies the discrete limit law

\[ \lim_{n \to \infty} P(\xi^{(k)} = i) = \lim_{n \to \infty} \delta^{(k)}_{n,i} = q_i, \quad \text{where} \quad q_i = i \frac{\tau_k - 1}{\tau_k} \left( \frac{\tau_k}{\tau_k} - i \right)^{i-1}. \quad (3.12) \]

Since the density function of a \( \Gamma(\lambda, r) \)-distribution is given by

\[ f_{\lambda, r}(x) = \begin{cases} \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0, \end{cases} \quad (3.13) \]

where \( \lambda > 0 \) and \( r > 0 \), we obtain, setting \( r = 2 \) and \( \lambda = \ln \frac{\tau_k}{\tau_k - 1} > 0 \)

\[ \lim_{n \to \infty} P(\xi^{(k)} = i) = I \left( \frac{\tau_k - 1}{\tau_k} \right)^{i-1} \]

\[ = \frac{1}{\tau_k(\tau_k - 1)} \left( \frac{\ln \tau_k}{\tau_k - 1} \right)^{-2} \left( \frac{\ln \tau_k}{\tau_k - 1} \right)^{2} \cdot i \left( \frac{\tau_k - 1}{\tau_k} \right)^i \]

\[ = \frac{1}{\tau_k(\tau_k - 1)} \left( \frac{\ln \tau_k}{\tau_k - 1} \right)^{-2} f_{\ln \frac{\tau_k}{\tau_k - 1}, 2}(i) \]

and the proof of the theorem is complete. \( \Box \)
4. The limit distribution of nontrivial returns

Let \( \rho_n^{(k)} \) denote the number of \( C \)-tableaux of length \( 2n \), which are in correspondence to \( k \)-noncrossing, \( \sigma \)-canonical RNA structures. Let \( \rho_{n,i}^{(k)} \) denote the number of \( C \)-tableaux of length \( 2n \), having exactly \( i \) \( \varnothing \)-shapes contained in the sequence \( (\lambda^2, \ldots, \lambda^{2n}) \). Let \( W_k(z, u) \) denote the bivariate generating function of \( \rho_{n,i}^{(k)} \). Then \( \rho_{n,i}^{(k)} = [z^n u^i]W_k(z, u) \) and \( W_k(z, u) = \sum_{j \geq 0} \sum_{n \geq j} \beta_{n,j} z^n u^j \). Furthermore, we set \( \beta_{n,k}^{(k)} = [z^n]W_k(z, 1) \).

**Lemma 3.** The bivariate generating function of the number of \( C \)-tableaux of length \( 2n \), with less than \( k \) rows, containing exactly \( i \) \( \varnothing \)-shapes, is given by

\[
W_k(z, u) = \frac{1}{1 - u(1 - \frac{1}{\tau_k(z)})}.
\]

**Proof.** Suppose the \( C \)-tableaux \( (\lambda^2, \ldots, \lambda^{2n}) \) contains exactly \( i \) \( \varnothing \)-shapes. These \( \varnothing \)-shapes split \( (\lambda^2, \ldots, \lambda^{2n}) \) uniquely into exactly \( i \) \( C \)-subtableaux, each of which either being a gap of length 2 or an irreducible \( \ast \)-tableaux. We conclude from this, that for fixed \( j \),

\[
\sum_{n \geq j} \beta_{n,j} z^n = (z + R_k(z))^j \tag{4.2}
\]

holds. Therefore the bivariate generating function \( W_k(z, u) \) satisfies

\[
W_k(z, u) = \sum_{j \geq 0} \sum_{n \geq j} \beta_{n,j} z^n u^j = \sum_{j \geq 0} (z + R_k(z))^j u^j
\]

\[
= \frac{1}{1 - u(z + R_k(z))}
= \frac{1}{1 - u(1 - \frac{1}{\tau_k(z)})},
\]

where the last equality follows from Eq. (3.5), proving the lemma. \( \Box \)

We set \( g(z) = \frac{1}{1-z}, \; h(z) = 1 - \frac{1}{\tau_k(z)} \) and let \( \eta_n^{(k)} \) denote the random variable having probability distribution \( P(\eta_n^{(k)} = i) = \frac{\rho_{n,i}^{(k)}}{\rho_n^{(k)}} \). In our next theorem, we prove that the limit distribution of \( \eta_n^{(k)} \) is determined by the density function of a \( \Gamma(\lambda, r) \)-distribution.

**Theorem 7.** Let \( \alpha_k \) denote the real, positive, dominant singularity of \( T_k(z) \) and let \( \tau_k = T_k(\alpha_k) \). Then the r.v. \( \eta_n^{(k)} \) satisfies the discrete limit law

\[
\lim_{n \to \infty} P(\eta_n^{(k)} = i) = q_i, \quad \text{where} \quad q_i = \frac{i}{\tau_k} \left( \frac{\tau_k - 1}{\tau_k} \right)^{i-1}. \tag{4.3}
\]

That is, \( \eta_n^{(k)} \) is determined by the density function of a \( \Gamma(\ln \frac{\tau_k}{\tau_k - 1}, 2) \)-distribution and the limit distribution has the probability generating function \( q(u) = \sum_{i \geq 1} q_i u^i = \frac{u}{(\tau_k(1-u) + u)^2} \).
Proof. Since $g(z) = \frac{1}{1-z}$ and $h(z) = 1 - \frac{1}{T_k(z)}$ have nonnegative coefficients and $h(0) = 0$, the composition $g(h(z))$ is again a power series. $W_k(z, u) = g(uh(z))$ has its unique dominant singularity at $z = \alpha_k$. Furthermore we observe, that irrespective of potential singularities arising from $T_k(z) = 0$, the dominant singularity of $h(z) = 1 - \frac{1}{T_k(z)}$ equals the dominant singularity of $T_k(z)$, i.e., $z = \alpha_k$.

Claim 1 implies that, for any fixed $u \in (0, 1)$, $W_k(z, u) = g(uh(z))$ has a singular expansion at its unique dominant singularity $z = \alpha_k$. We proceed by expressing $W_k(z, u) = \nu_u(w(z))$, where $\nu_u(z) = \frac{z}{\tau_k - z}$ and $w(z) = T_k(z)$. Setting $\tau_k = T_k(\alpha_k)$, the singular expansion of $W_k(z, u) = \nu_u(w(z))$ is according to the subcritical paradigm [7] derived by combining the regular expansion of $\nu_u$ and the singular expansion of $w$:

$$W_k(z, u) = \frac{w}{w(1-u) + u} = \frac{\tau_k}{\tau_k(1-u) + u} + \frac{u}{(\tau_k(1-u) + u)^2} \cdot (w - \tau_k)(1 + o(1)).$$

Accordingly, Theorem 3 implies

$$\left[z^n\right]W_k(z, 1) = \tilde{c}_k \alpha_k^{-n} n^{-\mu-1} (1 + o(1)).$$

Consequently we arrive at

$$\lim_{n \to \infty} \frac{\left[z^n\right]W_k(z, u)}{\left[z^n\right]W_k(z, 1)} = \frac{u}{(\tau_k(1-u) + u)^2},$$

where $\tau_k = T_k(\alpha_k)$. In view of $[u^1]q(u) = \frac{1}{\tau_k} \left(\frac{T_k}{\tilde{T}_k}\right)^{-1} = q_i$, Theorem 5 implies the discrete limit law

$$\lim_{n \to \infty} P(\eta_n^{(k)} = i) = \lim_{n \to \infty} \frac{\beta_n^{(k)}}{\beta_n^{(k)}} = q_i.$$

Using Eq. (3.13), setting $r = 2$ and $\lambda = \ln \frac{\tau_k}{\tilde{T}_k} > 0$, we analogously obtain
\[
\lim_{n \to \infty} P(\eta_n = i) = \frac{i}{\tau_k} \left( \frac{\tau_k - 1}{\tau_k} \right)^{i-1} = \frac{1}{\tau_k (\tau_k - 1)} \left( \ln \frac{\tau_k}{\tau_k - 1} \right)^{-2} \left( \ln \frac{\tau_k}{\tau_k - 1} \right)^2 \cdot i \left( \frac{\tau_k - 1}{\tau_k} \right)^i
\]

and Theorem 7 is proved. □

Acknowledgments

We are grateful to W.Y.C. Chen for stimulating discussions. This work was supported by the 973 Project, the PCSIRT Project of the Ministry of Education, the Ministry of Science and Technology, and the National Science Foundation of China.

References


