



Central and local limit theorems for RNA structures

Emma Y. Jin, Christian M. Reidys*

Center for Combinatorics, LPMC-TJKLC, Nankai University, Tianjin 300071, PR China

Received 29 July 2007; received in revised form 15 September 2007; accepted 17 September 2007

Abstract

A k -noncrossing RNA pseudoknot structure is a graph over $\{1, \dots, n\}$ without 1-arcs, i.e. arcs of the form $(i, i + 1)$ and in which there exists no k -set of mutually intersecting arcs. In particular, RNA secondary structures are 2-noncrossing RNA structures. In this paper we prove a central and a local limit theorem for the distribution of the number of 3-noncrossing RNA structures over n nucleotides with exactly h bonds. Our analysis employs the generating function of k -noncrossing RNA pseudoknot structures and the asymptotics for the coefficients. The results of this paper explain the findings on the number of arcs of RNA secondary structures obtained by molecular folding algorithms and are of relevance for prediction algorithms of k -noncrossing RNA structures.

© 2007 Published by Elsevier Ltd.

Keywords: k -noncrossing RNA structure; Pseudoknot; Generating function; Singularity; Central limit theorem; Local limit theorem

1. Introduction

An RNA molecule consists of the primary sequence of the four nucleotides **A**, **G**, **U** and **C** together with the Watson–Crick (**A–U**, **G–C**) and (**U–G**) base pairing rules. The latter specify the pairs of nucleotides that can potentially form bonds. Single stranded RNA molecules form helical structures whose bonds satisfy the above base pairing rules and which, in many cases, determine their function. For instance RNA ribosomes are capable of catalytic activity, cleaving other RNA molecules. Not all possible bonds are realized, though. Due to bio-physical constraints and the chemistry of Watson–Crick base pairs there exist rather severe constraints on the bonds of an RNA molecule. In light of this three decades ago Waterman et al. pioneered the concept of RNA secondary structures (Schmitt and Waterman, 1994; Waterman, 1978), being subject to the most strict combinatorial constraints. Any structure can be represented by drawing the primary sequence horizontally, ignoring all chemical bonds of its backbone, see Fig. 1. Then one draws all bonds, satisfying the Watson–Crick (**A–U**, **G–C**) and (**U–G**) base pairing rules as arcs in the upper half-plane,

effectively identifying structure with the set of all arcs. In this representation, RNA secondary structures have no 1-arcs, i.e. arcs of the form $(i, i + 1)$ and no two arcs (i_1, j_1) , (i_2, j_2) , where $i_1 < j_1$ and $i_2 < j_2$ with the property $i_1 < i_2 < j_1 < j_2$. In other words there exist no two arcs that cross in the diagram representation of the structure. It is well known that there exist additional types of nucleotide interactions (RNA, 2005). These bonds are called pseudoknots (Westhof and Jaeger, 1992) and occur in functional RNA (RNaseP, Loria and Pan, 1996), ribosomal RNA (Konings and Gutell, 1995) and are conserved in the catalytic core of group I introns. Pseudoknots appear in plant viral RNAs pseudoknots and in *in vitro* RNA evolution (Tuerk et al., 1992) experiments have produced families of RNA structures with pseudoknot motifs, when binding HIV-1 reverse transcriptase. Important mechanisms like ribosomal frame shifting (Chamorro et al., 1991) and telomerase RNA as part of the telomere RNP (Chen et al., 2000) also involve pseudoknot interactions. Detailed structural information on RNA pseudoknots can be found in the PseudoBase database (van Batenburg et al., 2001). The concept of k -noncrossing RNA structures introduced in Jin et al. (2007) captures these pseudoknot bonds and generalizes the concept of the RNA secondary structures in a natural way. In the diagram representation k -noncrossing

*Corresponding author. Tel.: +86 22 2350 6800; fax: +86 22 2350 9272.

E-mail address: reidys@nankai.edu.cn (C.M. Reidys).

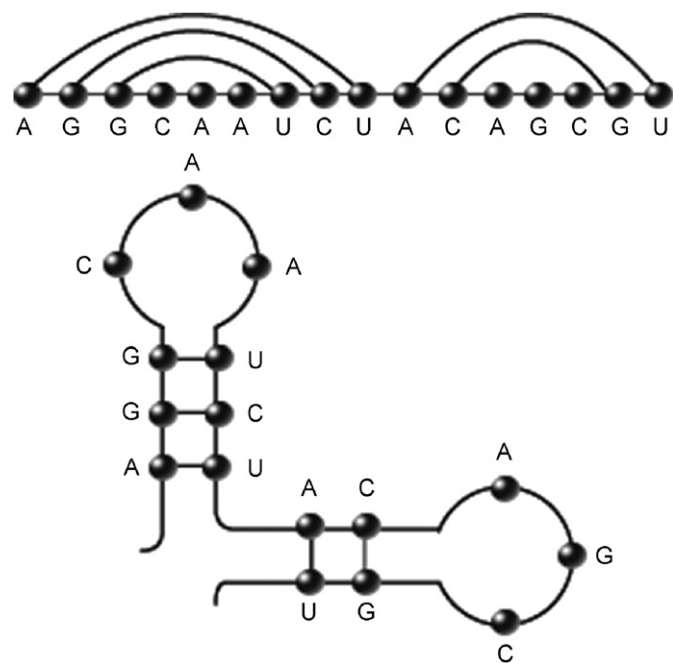


Fig. 1. RNA secondary structures. Diagram representation (top): the primary sequence, AGGCAAUCUACAGCGU, is drawn horizontally and its backbone bonds are ignored. All bonds are drawn in the upper half-plane. Secondary structures have the property that no two arcs intersect and all arcs have minimum length 2. Outer planar graph representation (bottom).

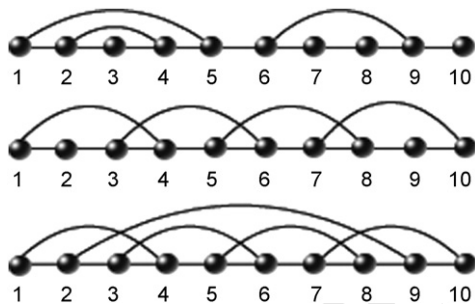


Fig. 2. k -noncrossing RNA structures: (a) secondary structure (with isolated labels 3, 7, 8, 10), (b) planar 3-noncrossing RNA structure, 2, 9 being isolated and (c) the smallest nonplanar 3-noncrossing structure.

RNA structure has no 1-arcs and contains at most $k - 1$ mutually crossing arcs (Fig. 2).

The starting point of this paper was the experimental finding that 3-noncrossing RNA structures for random sequences of length 100 over the nucleotides A, G, U and C exhibited sharply concentrated number of arcs (centered at 39). It was furthermore intriguing that the numbers of arcs was significantly higher than those in RNA secondary structures. Since all these quantities were via the generating functions for k -noncrossing RNA structures in Jin et al. (2007) explicitly known we could easily confirm that the number of 3-noncrossing RNA structures with exactly h arcs, $S'_3(n, h)$ satisfies indeed almost “perfectly” a Gaussian distribution with a mean of 39, see Fig. 3. We also found that a central limit theorem holds for RNA secondary

structures with h arcs, see Fig. 4. These observation motivated us to understand how and why these limit distributions arise, which is what the present paper is about. Let n be the sequence length and $S'_3(n, h)$ and $S_3(n)$ denote the number of 3-noncrossing RNA structures with exactly h arcs and the total number of 3-noncrossing RNA structures, respectively. Our main results can be summarized as follows:

Theorem 1.1. *The random variable X_n having distribution $\mathbb{P}(X_n = h) = S'_3(n, h)/S_3(n)$ satisfies a central and a local limit theorem with mean $0.39089n$ and variance $0.041565n$. Equivalently, in the limit of long sequences ($n \rightarrow \infty$) X_n is Gaussian distributed with mean $0.39089n$ and variance $0.041565n$.*

Our particular strategy is rooted in our recent work on asymptotic enumeration of k -noncrossing RNA structures (Jin and Reidys, 2007) and a paper of Bender (Bender, 1973) who showed how such central limit theorems arise in case of singularities that are poles. In order to obtain our results into context let us provide some background on central and local limit theorems. Suppose we are given a set A_n (of size a_n). For instance let A_n be the set of subsets of $\{1, \dots, n\}$. Suppose further we are given $A_{n,k}$ (of size $a_{n,k}$), $k \in \mathbb{N}$ representing a disjoint set partition of A_n . For instance let $A_{n,k}$ be the number of subsets of $\{1, \dots, n\}$ with exactly k elements. Consider the random variable ξ_n having the probability distribution $\mathbb{P}(\xi_n = k) = a_{n,k}/a_n$, then the corresponding probability generating function is given by

$$\sum_{k \geq 0} \mathbb{P}(\xi_n = k) w^k = \sum_{k \geq 0} \frac{a_{n,k}}{a_n} w^k = \frac{\sum_{k \geq 0} a_{n,k} w^k}{\sum_{k \geq 0} a_{n,k} 1^k}.$$

Let $\varphi_n(w) = \sum_{k \geq 0} a_{n,k} w^k$, then $\frac{\varphi_n(w)}{\varphi_n(1)}$ is the probability generating function of ξ_n and

$$f(z, w) = \sum_{n \geq 0} \varphi_n(w) z^n = \sum_{n \geq 0} \sum_{k \geq 0} a_{n,k} w^k z^n$$

is called the bivariate generating function. For instance, in our example we have $\mathbb{P}(\xi_n) = \binom{n}{2}$ and the resulting bivariate generating function is

$$\sum_{n \geq 0} \sum_{k \leq n} \binom{n}{k} w^k z^n = \frac{1}{1 - z(1 + w)}. \tag{1.1}$$

The key idea consists in considering $f(z, w)$ as being parameterized by w and to study the change of its singularity in an ε -disc centered at $w = 1$. Indeed the moment generating function is given by

$$E(e^{s\xi_n}) = \sum_{k \geq 0} \frac{a_{n,k}}{a_n} e^{sk} = \frac{\varphi_n(e^s)}{\varphi_n(1)} = \frac{[z^n]f(z, e^s)}{[z^n]f(z, 1)}$$

and $\frac{[z^n]f(z, e^{it})}{[z^n]f(z, 1)} = E(e^{it\xi_n})$ is the characteristic function of ξ_n . This shows that the coefficients of $f(z, w)$ control the distribution, which can, for large n , be obtained via singularity analysis. The resulting analysis can be amazingly simple. Let us illustrate this in the case of the

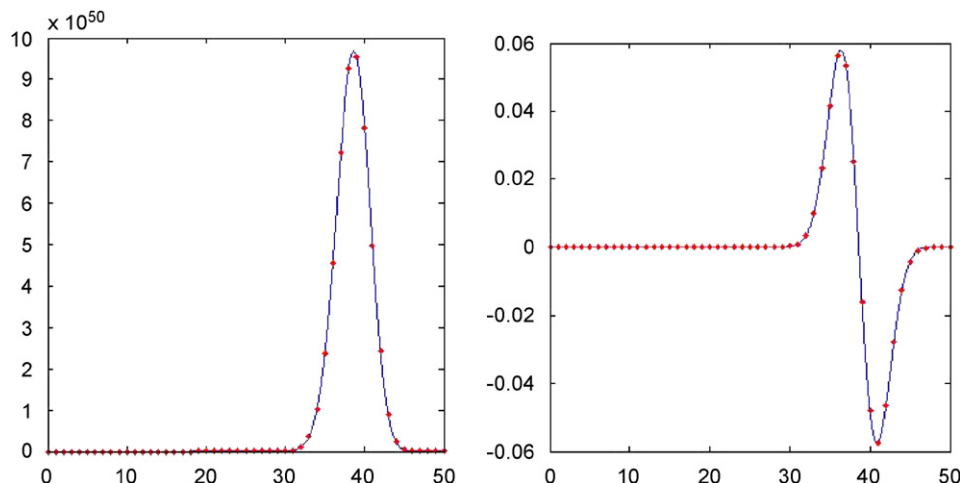


Fig. 3. Central limit theorem and local limit theorem for 3-noncrossing RNA structures of length $n = 100$ with exactly h arcs: we display the central limit theorem (left) for $S'_3(100, h)$, $h = 0, 1, 2, \dots, 50$ (labeled by red dots) with mean $0.39089 \times 100 = 39.089$ and variance $0.041565 \times 100 = 4.1565$, and for the local limit theorem (right), we display the difference $\sqrt{4.1565} \mathbb{P}\left(\frac{X_n - 39.089}{\sqrt{4.1565}} = x\right) - \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ which is maximal close to the peak of the distribution.

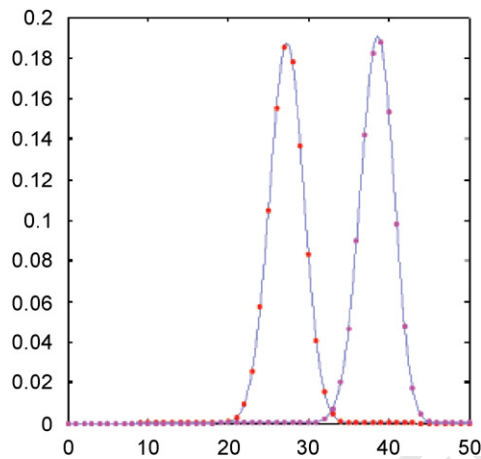


Fig. 4. Central limit theorem of 2-noncrossing and 3-noncrossing RNA structures: both random variables are normalized to $S'_2(n, h)/S_2(n)$ and $S'_3(n, h)/S_3(n)$, respectively. In case of $n = 100$, for 2-noncrossing RNA structures we have a mean of $0.276393n = 27.6393$ and variance $0.044721n = 4.4721$ (left curve), while for 3-noncrossing RNA structures mean $0.39089n = 39.089$ and variance $0.041565n = 4.1565$ (right curve). The red and magenta dots represent the values $S'_2(n, h)/S_2(n)$ and $S'_3(n, h)/S_3(n)$, respectively.

binomial distribution. Here we have the bivariate generating function $\sum_{n \geq 0} \sum_{k \leq n} \binom{n}{k} w^k z^n = \frac{1}{1-z(1+w)}$, Eq. (1.1). The simple pole $r(s)$ of $f(z, e^s)$ is $\frac{1}{1+e^s}$. Observe that $\frac{\varphi_n(e^s)}{\varphi_n(1)} \sim \left(\frac{r(0)}{r(s)}\right)^n$ holds for s uniformly in a neighborhood of 0, and Taylor expansion shows

$$\frac{\varphi_n(e^{it})}{\varphi_n(1)} \sim \exp\left(i \cdot \frac{n}{2} \cdot t - \frac{1}{2} \cdot \frac{n}{4} \cdot t^2 + O(t^3)\right)$$

uniformly for t for any arbitrary finite interval. It remains to apply the Lévy–Cramér theorem (Theorem 4.2) to the normalized characteristic function of the random variable $\eta_n = \frac{\xi_n - n/2}{\sqrt{n/4}}$, which yields the asymptotic normality of η_n .

Thus, $\binom{n}{k}$ is asymptotically normal distributed with mean $\frac{n}{2}$ and variance $\frac{n}{4}$. As it turns out we will have to work a bit harder to prove our main result. The complication is due to the fact that the generating function for 3-noncrossing RNA structures is much more complex (and fascinating) than the bivariate function of Eq. (1.1) which has a simple pole as dominant singularity. For instance, the singularity of the generating function for 3-noncrossing RNA structures is *not* a pole but of algebraic-logarithmic type (Flajolet et al., 2005; Gao and Richmond, 1992; Odlyzko, 1995).

Our two main results, Theorem 4.3 in Section 4 and Theorem 5.1 in Section 5 shed light on the distribution of 3-noncrossing RNA structures from a global and local perspective. A central limit theorem represents the global perspective on the limiting distribution of some random variable X_n :

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{X_n - \mu_n}{\sigma_n} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt.$$

Bender (1973) observed that a central limit theorem combined with certain smoothness conditions on the coefficients $a_{n,k}$ implies a local limit theorem which considers the difference between $\mathbb{P}(x \leq \frac{X_n - \mu_n}{\sigma_n} < x + 1)$ and $\frac{1}{\sqrt{2\pi}} \int_x^{x+1} e^{-t^2/2} dt$ as n tends to infinity. To be precise, X_n satisfies a local limit theorem for some $S = \{x \in \mathbb{R} | x = o(\sqrt{n})\}$ if and only if

$$\lim_{n \rightarrow \infty} \sup_{x \in S} \left| \sigma_n \mathbb{P}\left(\frac{X_n - \mu_n}{\sigma_n} = x\right) - \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right| = 0$$

holds and we say X_n satisfies a local limit theorem with mean μ_n and variance σ_n for some $S \subset \mathbb{R}$. Why is the smoothness of the $a_{n,k}$ so important? Suppose $a_{n,k} = \binom{n}{k} + (-1)^k 2 \binom{n}{k}$, then it follows in analogy to our above argument that a central limit theorem with mean $\frac{1}{2}n$ and variance $\frac{1}{4}n$ holds. However, ξ_n does not satisfy a local

limit theorem, since

$$\begin{aligned} & \left| \sigma_n \mathbb{P} \left(\frac{\xi_n - \mu_n}{\sigma_n} = x \right) - \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right| \\ &= \frac{1}{2} \sqrt{n} \mathbb{P} \left(\frac{\xi_n - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}} = x \right) - \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \end{aligned}$$

and for $S = \{\frac{\sqrt{n}}{2} | n = 1, 2, \dots\}$, we have $|\frac{1}{2}\sqrt{n}\mathbb{P}(\xi_n = \frac{3}{4}n) - \frac{1}{\sqrt{2\pi}}e^{-n/8}| \rightarrow 0$, the key point being that $a_{n,k}$ flips between $-\binom{n}{k}$ and $3\binom{n}{k}$.

All results of this paper hold for 2-noncrossing RNA structures, i.e. RNA secondary structures. This is a consequence of an analogous analysis of their respective bivariate generating function. In this case, however, no singular expansion is necessary as the generating function itself can be used. They also give rise to obtain the asymptotic results on RNA secondary structures of Hofacker et al. (1998) on a new level. We can pass from computing exponential growth rates to computing distributions for RNA secondary structures with specific properties. To be precise, let n be the sequence length and $S'_2(n, h)$ and $S_2(n)$ denote the number of RNA secondary structures with exactly h arcs and the total number of secondary structures, respectively. Then we have for RNA secondary structures.

Theorem 1.2. *The random variable Y_n having distribution $\mathbb{P}(Y_n = h) = S'_2(n, h)/S_2(n)$ satisfies a central and local limit theorem with mean $0.27639n$ and variance $0.04472n$. Equivalently, in the limit of long sequences ($n \rightarrow \infty$) Y_n is Gaussian distributed with mean $0.27639n$ and variance $0.04472n$.*

In particular the theorem predicts a sharp concentration of the number of RNA secondary structures with 55.278% paired bases which agrees with the statistics of RNA secondary structures obtained by folding algorithms (Zuker and Sankoff, 1984; Hofacker et al., 1998; Waterman and Smith, 1986; Tacker et al., 1996, 1994; McCaskill, 1990). Let us finally remark that much more holds: due to the determinant formula for k -noncrossing matchings and the functional identity of Lemma 3.1 in Section 3 our results can be generalized to k -noncrossing RNA structures, where k is arbitrary. Why this is of interest can be seen in Fig. 4. For higher k the mean of the central limit theorems for k -noncrossing RNA structures will shift towards the maximum combinatorially possible number of arcs. We speculate that each increase in k will basically cut the distance to the maximum arc number in half. This is the work in progress.

The paper is structured as follows: In Section 2 we provide some background on k -noncrossing RNA structures and all generating functions involved. In Section 3 we give a functional equation for the bivariate generating function of $S'_3(n, h)$ via 3-noncrossing matchings proved in Jin and Reidys (2007). We have included its proof in the appendix in order to keep the paper self-contained. This

functional identity plays a key role in proving the central limit and local limit theorem in Sections 4 and 5, respectively. The central limit theorem is proved by analyzing the singular expansion of analytic function of power series $\sum_{n \geq 0} \sum_{h \leq n/2} S'_3(n, h) w^h z^n$ and using transfer theorems (Flajolet et al., 2005; Gao and Richmond, 1992; Odlyzko, 1995) and to prove the local limit theorem, we use a theorem of Hwang (1998) and build on our proof of the central limit theorem.

2. RNA structures

Let us begin by illustrating the concept of RNA structures. Suppose we are given the primary sequence

AACCAUGUGGUACUUGAUGGCGAC.

Structures are combinatorial graphs over the labels of the nucleotides of the primary sequence. These graphs can be represented in several ways. In Fig. 5 we represent a 3-noncrossing RNA structure with loop-loop interactions in two ways: first we display the structure as a planar graph and secondly as a diagram, where the bonds are drawn as arcs in the positive half-plane.

In the following we will consider structures as diagram representations of digraphs. A digraph D_n is a pair of sets V_{D_n}, E_{D_n} , where $V_{D_n} = \{1, \dots, n\}$ and $E_{D_n} \subset \{(i, j) | 1 \leq i < j \leq n\}$. V_{D_n} and E_{D_n} are called vertex and arc set, respectively. A k -noncrossing digraph is a digraph in which all vertices have degree ≤ 1 and which does not contain a k -set of arcs that are mutually intersecting, i.e.

$$\begin{aligned} & \exists (i_{r_1}, j_{r_1}), (i_{r_2}, j_{r_2}), \dots, (i_{r_k}, j_{r_k}); \\ & i_{r_1} < i_{r_2} < \dots < i_{r_k} < j_{r_1} < j_{r_2} < \dots < j_{r_k}. \end{aligned} \quad (2.1)$$

We will represent digraphs as a diagrams (Fig. 5) by representing the vertices as integers on a line and connecting any two adjacent vertices by an arc in the upper half-plane. The direction of the arcs is implicit in the linear ordering of the vertices and accordingly omitted.

Definition 2.1. An RNA structure (of pseudoknot type $k - 2$), $S_{k,n}$, is a digraph over $\{1, \dots, n\}$ in which all vertices have degree ≤ 1 that does not contain a k -set of mutually intersecting arcs and 1-arcs, i.e. arcs of the form $(i, i + 1)$ for $1 \leq i \leq n - 1$, respectively. We denote the number of k -noncrossing RNA structures by $S_k(n)$ and the number of k -noncrossing RNA structures with exactly ℓ isolated vertices and with h arcs by $S_k(n, \ell)$ and $S'_k(n, h)$, respectively. Note that $S'_k(n, h) = S_k(n, n - 2h)$.

Let $f_k(n, \ell)$ denote the number of k -noncrossing digraphs with ℓ isolated points. We have shown in (Jin et al., 2007) that

$$f_k(n, \ell) = \binom{n}{\ell} f_k(n - \ell, 0) \quad (2.2)$$

$$\det[I_{i-j}(2x) - I_{i+j}(2x)]_{i,j=1}^{k-1} = \sum_{n \geq 1} f_k(n, 0) \cdot \frac{x^n}{n!} \quad (2.3)$$

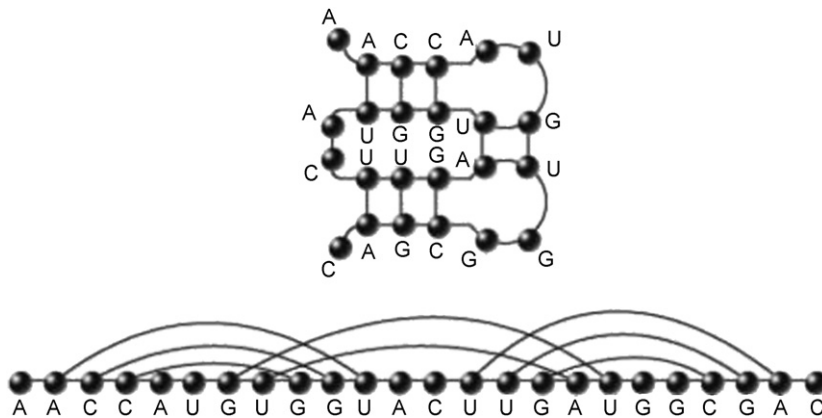


Fig. 5. A 3-noncrossing RNA structure, as a planar graph (top) and as a diagram (bottom).

$$\begin{aligned}
 & e^x \det[I_{i-j}(2x) - I_{i+j}(2x)]_{i,j=1}^{k-1} \\
 &= \left(\sum_{\ell \geq 0} \frac{x^\ell}{\ell!} \right) \left(\sum_{n \geq 1} f_k(n, 0) \frac{x^n}{n!} \right) \\
 &= \sum_{n \geq 1} \left\{ \sum_{\ell=0}^n f_k(n, \ell) \right\} \cdot \frac{x^n}{n!}.
 \end{aligned} \tag{2.4}$$

In particular we obtain for $k = 2$ and $k = 3$

$$\begin{aligned}
 f_2(n, \ell) &= \binom{n}{\ell} C_{(n-\ell)/2} \quad \text{and} \\
 f_3(n, \ell) &= \binom{n}{\ell} \left[C_{\frac{n-\ell}{2}+2} C_{\frac{n-\ell}{2}} - C_{\frac{n-\ell}{2}+1}^2 \right],
 \end{aligned} \tag{2.5}$$

where C_m denotes the m th Catalan number. The derivation of the generating function of k -noncrossing RNA structures, given in Theorem 2.1, uses advanced methods and novel constructions of enumerative combinatorics due to Chen et al. (2007) (Gessel and Zeilberger, 1992) and Stanley’s mapping between matchings and oscillating tableaux, i.e. families of Young diagrams, in which any two consecutive shapes differ by exactly one square. The enumeration is obtained using the reflection principle due to Gessel and Zeilberger (1992) and Lindstroem (1973) combined with an inclusion–exclusion argument in order to eliminate the arcs of length 1. In Jin et al. (2007) generalizations to restricted (i.e. where arcs of the form $(i, i + 2)$ are excluded) and circular RNA structures are given. The following theorem provides all data on number of k -noncrossing RNA structures with h arcs and the number of all k -noncrossing RNA structures.

Theorem 2.1 (Jin et al., 2007). *Let n be the sequence length, $k \in \mathbb{N}$, $k \geq 2$, and let C_m denote the m th Catalan number. Furthermore, let $f_k(n, \ell)$ be the number of k -noncrossing digraphs over n vertices with exactly ℓ isolated vertices. Then the number of k -noncrossing RNA structures with ℓ isolated vertices, $S_k(n, \ell)$, is given by*

$$S_k(n, \ell) = \sum_{b=0}^{(n-\ell)/2} (-1)^b \binom{n-b}{b} f_k(n-2b, \ell), \tag{2.6}$$

where $f_k(n-2b, \ell)$ is given by the generating function in Eq. (2.3). Furthermore, the number of k -noncrossing RNA structures, $S_k(n)$, is

$$S_k(n) = \sum_{b=0}^{\lfloor n/2 \rfloor} (-1)^b \binom{n-b}{b} \left\{ \sum_{\ell=0}^{n-2b} f_k(n-2b, \ell) \right\} \tag{2.7}$$

where $\{ \sum_{\ell=0}^{n-2b} f_k(n-2b, \ell) \}$ is given by the generating function in Eq. (2.4).

In principle, Theorem 2.1 contains all information about the number of k -noncrossing RNA structures. However, due to the inclusion–exclusion structure of its coefficients it is, however, difficult to interpret and to express their behavior for large n . Subsequent asymptotic analysis (Jin and Reidys, 2007) produced the following simple formula

Theorem 2.2 (Jin and Reidys, 2007). *The number of 3-noncrossing RNA structures is asymptotically given by*

$$S_3(n) \sim \frac{10.4724 \cdot 4!}{n(n-1) \dots (n-4)} \left(\frac{5 + \sqrt{21}}{2} \right)^n.$$

3. A functional equation

We have shown in the introduction that the bivariate generating function is the key to prove the central and local limit theorems. The following lemma, whose proof is given in the appendix, rewrites this bivariate generating function as a composition of two “simple” functions. This is crucial for the singularity analysis insofar as we can use a phenomenon known as persistence of the singularity of the “outer” function (the *supercritical case*) (Flajolet et al., 2005). It basically means that the type of the singularity is determined by the generating function of k -noncrossing matchings.

Lemma 3.1 (Jin and Reidys, 2007). *Let x be an indeterminate over \mathbb{R} and $w \in \mathbb{R}$ a parameter. Let $\rho_k(w)$ denote the radius of convergence of the power series $\sum_{n \geq 0} [\sum_{h \leq n/2} S'_k(n, h) w^{2h}] x^n$. Then for $|x| < \rho_k(w)$*

$$\sum_{n \geq 0} \sum_{h \leq n/2} S'_k(n, h) w^{2h} x^n$$

$$= \frac{1}{w^2 x^2 - x + 1} \sum_{n \geq 0} f_k(2n, 0) \left(\frac{wx}{w^2 x^2 - x + 1} \right)^{2n} \quad (3.1)$$

holds. In particular we have for $w = 1$

$$\sum_{n \geq 0} S_k(n) z^n = \frac{1}{z^2 - z + 1} \sum_{n \geq 0} f_k(2n, 0) \left(\frac{z}{z^2 - z + 1} \right)^{2n} \quad (3.2)$$

for $z \in \mathbb{C}$ with $|z| < \rho_k(1)$.

To keep the paper self-contained we give the proof of Lemma 3.1 in the appendix. While (3.1) can only be proved on the level of formal power series for real variables, complex analysis, i.e. the interpretation of these generating functions as analytic functions, allows to extend the equality to arbitrary complex variables.

Lemma 3.2. Suppose $\varepsilon > 0$, $k \in \mathbb{N}$, $k \geq 2$ and $w = e^{s/2}$, where $|s| < \varepsilon$ and $\varphi_{n,k}(s) = \sum_{h \leq n/2} S'_k(n, h) e^{hs}$. Let $\rho_k(s) \in \mathbb{R}^+$ denote the radius of convergence of $\sum_{n \geq 0} \varphi_{n,k}(s) z^n$ parameterized by s . Then we have

$$\forall s, z \in \mathbb{C}; \quad |s| < \varepsilon, |z| < \rho_k(s); \quad \sum_{n \geq 0} \varphi_{n,k}(s) z^n$$

$$= \frac{1}{e^s z^2 - z + 1} \sum_{n \geq 0} f_k(2n, 0) \left(\frac{e^{s/2} z}{e^s z^2 - z + 1} \right)^{2n}. \quad (3.3)$$

Furthermore, $\sum_{n \geq 0} \varphi_{n,3}(s) z^n$ has an analytic continuation, $\Xi_3(z, s)$. For ε sufficiently small and $|s| < \varepsilon$, $\Xi_3(z, s)$ has exactly six singularities, four of which have distinct moduli.

Proof. We first prove Eq. (3.3). For this purpose we observe that

$$\forall |s| < \varepsilon, |z| < \rho_k(s) \quad G(z, s)$$

$$= \frac{1}{e^s z^2 - z + 1} \sum_{n \geq 0} f_k(2n, 0) \left(\frac{e^{s/2} z}{e^s z^2 - z + 1} \right)^{2n} \quad (3.4)$$

considered as a power series in $e^{1/2s}$ is analytic in a neighborhood of $s = 0$, since $G(z, 0)$ is analytic for $|z| < \rho_k(0)$. In addition, we can interpret $\sum_{n \geq 0} \varphi_{n,k}(s) z^n$ as a power series in $e^{1/2s}$:

$$\sum_{n \geq 0} \left(\sum_{h \leq n/2} S'_k(n, h) e^{hs} \right) z^n = \sum_{h \geq 0} \left(\sum_{n \geq 2h} S'_k(n, h) z^n \right) (e^s)^h$$

$$= \sum_{h \geq 0} \psi_h(z) (e^{1/2s})^{2h}. \quad (3.5)$$

Therefore, $G(z, s)$ and the power series $\sum_{n \geq 0} \varphi_{n,k}(s) z^n$ are analytic in the indeterminate $e^{1/2s}$ in an ε -disc centered at 0. Lemma 3.1 implies that for $s \in]-\varepsilon, \varepsilon[$ the analytic functions $G(z, s)$ and $\sum_{n \geq 0} \varphi_{n,k}(s) z^n$ are equal. Since any two functions that are analytic at 0 and coincide on the interval $]-\varepsilon, \varepsilon[$ are identical, we obtain

$$\forall |s| < \varepsilon, |z| < \rho_k(s) \quad G(z, s) = \sum_{n \geq 0} \varphi_{n,k}(s) z^n. \quad (3.6)$$

Claim 1. Suppose $|s| < \varepsilon$. Then $\sum_{n \geq 0} \varphi_{n,3}(s) z^n$ has an analytic continuation, $\Xi_3(z, s)$, which has exactly six singularities four of which have distinct moduli.

In order to prove Claim 1 we observe that the power series $\sum_{n \geq 0} f_3(2n, 0) y^n$ has the analytic continuation $\Psi(y)$ (obtained by MAPLE sumtools) given by

$$\Psi(y) = \frac{-(1 - 16y)^{3/2} P_{3/2}^{-1} \left(\frac{16y + 1}{16y - 1} \right)}{16y^{5/2}}, \quad (3.7)$$

where $P_v^m(x)$ denotes the Legendre Polynomial of the first kind with the parameters $v = \frac{3}{2}$ and $m = -1$. According to Eq. (3.6) we have

$$\sum_{n \geq 0} \varphi_{n,k}(s) z^n = \frac{1}{e^s z^2 - z + 1} \sum_{n \geq 0} f_k(2n, 0) \left(\frac{e^{s/2} z}{e^s z^2 - z + 1} \right)^{2n} \quad (3.8)$$

which implies that $\sum_{n \geq 0} \varphi_{n,3}(s) z^n$ has the analytic continuation

$$\forall |s| < \varepsilon, \quad \Xi_3(z, s) = \frac{1}{e^s z^2 - z + 1} \Psi \left(\frac{e^{1/2s} z}{e^s z^2 - z + 1} \right)^2. \quad (3.9)$$

In particular for $s = 0$, $\Xi_3(z, 0)$ is the analytic continuation of the power series $\sum_{n \geq 0} S_3(n) z^n$. We proceed by showing that $\Xi_3(z, s)$ has exactly six singularities and four of them have different moduli in \mathbb{C} parameterized by s . Two singularities are given by the roots of $e^s z^2 - z + 1$ are $\zeta_1(s) = \frac{1 - \sqrt{1 - 4e^s}}{2e^s}$ and $\zeta_2(s) = \frac{1 + \sqrt{1 - 4e^s}}{2e^s}$. Observe that $|\zeta_1(0)| = |\zeta_2(0)| = 1$ and polynomial $e^s z^2 - z + 1$ depends continuously on $e^{s/2}$; therefore, $\zeta_1(s)$ and $\zeta_2(s)$ could potentially have equal modulus for $|s| < \varepsilon$. The remaining four singularities are induced by the the unique dominant singularity $\alpha_1 = \frac{1}{16}$ of analytic function $\Psi(y)$. The function $\Psi(y)$ has three singularities, two of them $\alpha_1 = \frac{1}{16}$ and $\alpha_2 = +\infty$ are branch points and the other $\alpha_3 = 0$ is a removable singularity. The function $g(z) = \left(\frac{e^{s/2} z}{e^s z^2 - z + 1} \right)^2$ with $g(0) = 0$ has a radius of convergence of 1 as s tends to 0. Therefore, the singularity type only depends on $\Psi(y)$ (this is the *supercritical case* in Flajolet et al., 2005). The singularity $\alpha_1 = \frac{1}{16}$ gives rise to the equations

$$0 = e^s z^2 - (1 + 4e^{1/2s})z + 1 \quad \text{and}$$

$$0 = e^s z^2 + (4e^{1/2s} - 1)z + 1$$

and setting $\mu_+(s) = 1 + 4e^{1/2s}$, $\mu_-(s) = 1 - 4e^{1/2s}$ and $\theta(s) = \sqrt{12e^s + 8e^{1/2s} + 1}$ its roots are given by

$$\zeta_3(s) = \frac{\mu_+(s) - \theta(s)}{2e^s}, \quad \zeta_4(s) = \frac{\mu_+(s) + \theta(s)}{2e^s},$$

$$\zeta_5(s) = \frac{\mu_-(s) + \theta(s)}{2e^s} \quad \text{and} \quad \zeta_6(s) = \frac{\mu_-(s) - \theta(s)}{2e^s},$$

respectively. Observe that for $|s| < \varepsilon$, $e^{s/2}$ is in a neighborhood of 1 over \mathbb{C} ; hence, $\theta(s) \neq 0$. That leads to four distinct

roots $\zeta_3(s), \zeta_4(s), \zeta_5(s), \zeta_6(s)$ over $|s| < \varepsilon$, all of them have distinct moduli for s being a sufficiently small neighborhood of 0. Indeed, for $s = 0$ we have four distinct real-valued roots

$$\zeta_3(0) = \frac{5 - \sqrt{21}}{2}, \quad \zeta_4(0) = \frac{5 + \sqrt{21}}{2},$$

$$\zeta_5(0) = \frac{-3 + \sqrt{5}}{2} \quad \text{and} \quad \zeta_6(0) = \frac{-3 - \sqrt{5}}{2}$$

and the polynomials $e^s z^2 - (1 + 4e^{1/2s})z + 1$, $e^s z^2 + (4e^{1/2s} - 1)z + 1$ and $e^s z^2 - z + 1$ depend continuously on the parameter $e^{1/2s}$, whence Claim 1 and the lemma follow. \square

4. The central limit theorem

In this section we prove a central limit theorem for the number of 3-noncrossing RNA structures with h arcs. We will analyze for fixed but arbitrary n the distribution of $S'_3(n, h)$. Let us first prepare some methods and results used in the proof of Theorem 4.3. $[z^n]f(z)$ denotes the coefficient of z^n in the power series expansion of $f(z)$ around 0. The scaling property of Taylor coefficients

$$\forall \gamma \in \mathbb{C} \setminus \{0\}, \quad [z^n]f(z) = \gamma^n [z^n]f\left(\frac{z}{\gamma}\right), \tag{4.1}$$

shows that w.l.o.g. any singularity analysis can be reduced to the case where 1 is the dominant singularity. We will be interested in the behavior of an analytic function “locally”, i.e. around a certain singularity ρ . For this purpose we use the notation

$$f(z) = O(g(z)) \text{ as } z \rightarrow \rho \iff f(z)/g(z) \text{ is bounded as } z \rightarrow \rho \tag{4.2}$$

and if we write $f(z) = O(g(z))$ it is implicitly assumed that z tends to a (unique) singularity. Given two numbers ϕ, R , where $R > |\rho| > 0$ and $0 < \phi < \frac{\pi}{2}$ and $\rho \in \mathbb{C}$ the open domain $\Delta_\rho(\phi, R)$ is defined as

$$\Delta_\rho(\phi, R) = \{z \mid |z| < R, z \neq \rho, |\text{Arg}(z - \rho)| > \phi\}. \tag{4.3}$$

A domain is a Δ_ρ -domain if it is of the form $\Delta_\rho(\phi, R)$ for some R and ϕ . A function is Δ_ρ -analytic if it is analytic in some Δ_ρ -domain. We use $U(a, r) = \{z \in \mathbb{C} \mid |z - a| < r\}$ to denote the open neighborhood of a in \mathbb{C} . From the following theorem we can extract the coefficients of analytic functions provided these functions satisfy certain “local” properties.

Theorem 4.1 (Flajolet et al., 2005). Let $r \in \mathbb{Z}_{\geq 0}$ and $f(z, e^s)$ be a $\Delta_{\rho(s)}$ -analytic function parameterized by s , which satisfies in the intersection of a neighborhood of $\rho(s)$ with its $\Delta_{\rho(s)}$ -domain

$$f(z, e^s) = b_0(s) + b_1(s)(z - \rho(s)) + A(s)(\rho(s) - z)^r \ln\left(\frac{1}{\rho(s) - z}\right) + R(z, s) \tag{4.4}$$

where $A(s), b_0(s)$ and $b_1(s)$ are analytic in $|s| < \varepsilon$ and $|R(z, s)| \leq c|\rho(s) - z|$ for some absolute constant $c \in \mathbb{C}$. That is, we have $f(z, e^s) = O((\rho(s) - z)^r \ln(\frac{1}{\rho(s) - z}))$ with uniform error bound as s in a neighborhood of 0. Then we have

$$[z^n]f(z, e^s) = A(s)(-1)^r \frac{r!}{n(n-1)\dots(n-r)} \times \left(1 - O\left(\frac{1}{n}\right)\right) \text{ for some } A(s) \in \mathbb{C}, \tag{4.5}$$

where the error term is again uniform for s from a neighborhood of origin, i.e. $R(s) \leq c|s|$, where $c > 0$.

Remark. The equivalence between Eq. (4.4) and $f(z, e^s) = O((\rho(s) - z)^r \ln(\frac{1}{\rho(s) - z}))$ for $r \in \mathbb{Z}_{\geq 0}$ can be seen as follows: by definition of $f(z, e^s) = O((\rho(s) - z)^r \ln(\frac{1}{\rho(s) - z}))$ there exist $A(z, s)$ and $B(z, s)$, such that $f(z, e^s) = B(z, s) + A(z, s)(\rho(s) - z)^r \ln(\frac{1}{\rho(s) - z})$, where $A(z, s)$ and $B(z, s)$ are analytic in a neighborhood of $\rho(s)$. Taylor expansion of $A(z, s)$ and $B(z, s)$ at $z = \rho(s)$ produces

$$f(z, s) = B(z, s) + A(z, s)(\rho(s) - z)^r \ln\left(\frac{1}{\rho(s) - z}\right)$$

$$= b_0(s) + b_1(s)(z - \rho(s)) + \dots + (a_0(s) + a_1(s)(z - \rho(s)) + \dots) \times (\rho(s) - z)^r \ln\left(\frac{1}{\rho(s) - z}\right)$$

$$= b_0(s) + b_1(s)(z - \rho(s)) + a_0(s)(\rho(s) - z)^r \ln\left(\frac{1}{\rho(s) - z}\right) + R(z, s)$$

where $R(z, s) = O((\rho(s) - z)^{r+1} \ln(\frac{1}{\rho(s) - z}))$. For $r \in \mathbb{Z}_{\geq 0}$, $\frac{|R(z, s)|}{|\rho(s) - z|} = O(|\rho(s) - z|^r \ln(\frac{1}{|\rho(s) - z|}))$ is bounded by an absolute constant as z tends to $\rho(s)$, which implies that the error bound is uniform.

The following Theorem is a classic result on limit distributions which allows us to prove our main result via characteristic functions, i.e. explicitly by showing $\lim_{n \rightarrow \infty} \varphi_n(t) = \varphi(t)$ uniformly over an arbitrary finite interval enclosing the origin.

Theorem 4.2 (Lévy–Cramér). Let $\{\xi_n\}$ be a sequence of random variables and let $\{\varphi_n(x)\}$ and $\{F_n(x)\}$ be the corresponding sequences of characteristic and distribution functions. If there exists a function $\varphi(t)$, such that $\lim_{n \rightarrow \infty} \varphi_n(t) = \varphi(t)$ uniformly over an arbitrary finite interval enclosing the origin, then there exists a random variable ξ with distribution function $F(x)$ such that

$$F_n(x) \implies F(x)$$

uniformly over any finite or infinite interval of continuity of $F(x)$.

We now consider the random variable X_n having the distribution $\mathbb{P}(X_n = h) = S'_3(n, h)/S_3(n)$, where $h = 0, 1, \dots, \lfloor \frac{n}{2} \rfloor$. The key point in the proof of Theorem 4.3 is to compute the coefficients of the bivariate generating

function whose variable, s , is considered as a parameter. Intuitively the particular distribution is a result of how the singularity shifts as a function of this parameter. As a result the proof is somewhat “nonprobabilistic” and has two distinct parts: (a) the analytic combinatorics of the bivariate generating function and (b) the computation of the characteristic function with subsequent application of the Lévy–Cramér Theorem.

Theorem 4.3. The random variable $\frac{X_n - \mu n}{\sqrt{\sigma^2 n}}$ has asymptotically normal distribution with parameter $(0, 1)$, i.e.

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{X_n - \mu n}{\sqrt{\sigma^2 n}} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-1/2t^2} dt, \quad (4.6)$$

and μ and σ^2 are given by

$$\begin{aligned} \mu &= -\frac{-\frac{3}{2} + \frac{13}{42}\sqrt{21}}{\frac{5}{2} - \frac{1}{2}\sqrt{21}} = 0.39089 \quad \text{and} \\ \sigma^2 &= \mu^2 - \frac{1 - \frac{94}{441}\sqrt{21}}{(5 - \sqrt{21})/2} = 0.041565. \end{aligned} \quad (4.7)$$

Proof. We set $w = e^{1/2s}$ and $\varphi_{n,3}(s) = \sum_{h \leq n/2} \mathbf{S}'_3(n, h) e^{hs}$. Since

$$\sum_{n \geq 0} \varphi_{n,3}(s) z^n = \sum_{n \geq 0} \left(\sum_{h \leq n/2} \mathbf{S}'_3(n, h) e^{hs} \right) z^n, \quad (4.8)$$

we can consider the double generating function $\sum_{n \geq 0} \sum_{h \leq n/2} \mathbf{S}'_3(n, h) w^{2h} z^n$ as a power series in the complex indeterminate z , parameterized by s .

Claim 2.

$$\begin{aligned} \Psi(z) &= O\left((1 - 16z)^4 \ln\left(\frac{1}{1 - 16z}\right)\right) \text{ holds uniformly for} \\ \forall z \in \Delta_{1/16}(\phi, R) \cap U\left(\frac{1}{16}, \varepsilon\right); \end{aligned} \quad (4.9)$$

$\Psi(z)$ is $\Delta_{1/16}(\phi, R)$ -analytic and has the singular expansion $(1 - 16z)^4 \ln\left(\frac{1}{1 - 16z}\right)$ in the intersection of $U\left(\frac{1}{16}, \varepsilon\right)$ with the $\Delta_{1/16}$ -domain, where $\Delta_r(\phi, R) = \{z \mid |z| < R, z \neq r, |\text{Arg}(z - r)| > \phi\}$ for some $R > r$. First $\Delta_{1/16}(\phi, R)$ -analyticity of the function $(1 - 16z)^4 \ln\left(\frac{1}{1 - 16z}\right)$ is obvious. We proceed by proving that $(1 - 16z)^4 \ln\left(\frac{1}{1 - 16z}\right)$ is the singular expansion of $\Psi(z)$. The above-mentioned scaling property of Taylor coefficients allows us to consider the power series $\sum_{n \geq 0} f_3(2n, 0) \left(\frac{z}{16}\right)^n$ over the Δ -domain $\Delta_1(\phi, R)$ for some $R > 1$. Using the notation of falling factorials $(n - 1)_4 = (n - 1)(n - 2)(n - 3)(n - 4)$ we observe

$$\begin{aligned} f_3(2n, 0) &= C_{n+2} C_n - C_{n+1}^2 \\ &= \frac{1}{(n - 1)_4} \frac{12(n - 1)_4 (2n + 1)}{(n + 3)(n + 1)^2 (n + 2)^2} \binom{2n}{n}. \end{aligned}$$

With this expression for $f_3(2n, 0)$ we arrive at the formal identity

$$\begin{aligned} &\sum_{n \geq 5} 16^{-n} f_3(2n, 0) z^n \\ &= O\left(\sum_{n \geq 5} \left[16^{-n} \frac{1}{(n - 1)_4} \frac{12(n - 1)_4 (2n + 1)}{(n + 3)(n + 1)^2 (n + 2)^2} \binom{2n}{n} \right. \right. \\ &\quad \left. \left. - \frac{4!}{(n - 1)_4} \frac{1}{\pi n} \right] z^n + \sum_{n \geq 5} \frac{4!}{(n - 1)_4} \frac{1}{\pi n} z^n\right), \end{aligned} \quad (59-63)$$

where $f(z) = O(g(z))$ denotes that the limit $f(z)/g(z)$ is bounded for $z \rightarrow 1$, Eq. (4.2). It is clear that the error bound

$$\begin{aligned} &\sum_{n \geq 5} \left[16^{-n} \frac{1}{(n - 1)_4} \frac{12(n - 1)_4 (2n + 1)}{(n + 3)(n + 1)^2 (n + 2)^2} \binom{2n}{n} \right. \\ &\quad \left. - \frac{4!}{(n - 1)_4} \frac{1}{\pi n} \right] z^n \\ &\sim \sum_{n \geq 5} \left[16^{-n} \frac{1}{(n - 1)_4} \frac{12(n - 1)_4 (2n + 1)}{(n + 3)(n + 1)^2 (n + 2)^2} \binom{2n}{n} \right. \\ &\quad \left. - \frac{4!}{(n - 1)_4} \frac{1}{\pi n} \right] < \kappa \end{aligned} \quad (64-79)$$

holds uniformly for z in $\Delta_1(\phi, R) \cap U(1, \varepsilon)$ and some absolute $\kappa < 0.0784$. Therefore, we can conclude

$$\sum_{n \geq 5} 16^{-n} f_3(2n, 0) z^n = O\left(\sum_{n \geq 5} \frac{4!}{(n - 1)_4} \frac{1}{\pi n} z^n\right). \quad (4.10)$$

We proceed by interpreting the power series on the rhs, observing

$$\begin{aligned} \forall n \geq 5; \quad [z^n] \left((1 - z)^4 \ln \frac{1}{1 - z} \right) \\ = \frac{4!}{(n - 1) \dots (n - 4) n}, \end{aligned} \quad (4.11)$$

whence $((1 - z)^4 \ln \frac{1}{1 - z})$ is the unique analytic continuation of $\sum_{n \geq 5} \frac{4!}{(n - 1)_4} \frac{1}{\pi n} z^n$. Using the scaling property of Taylor coefficients $[z^n] f(z) = \gamma^n [z^n] f\left(\frac{z}{\gamma}\right)$

$$\begin{aligned} \Psi(z) &= O\left((1 - 16z)^4 \ln\left(\frac{1}{1 - 16z}\right)\right) \\ &\text{holds uniformly for } \forall z \in \Delta_{1/16}(\phi, R) \cap U\left(\frac{1}{16}, \varepsilon\right) \end{aligned} \quad (4.12)$$

Therefore, we have proved that $(1 - 16z)^4 \ln\left(\frac{1}{1 - 16z}\right)$ is the singular expansion of $\Psi(z)$ at $z = \frac{1}{16}$, whence Claim 2. Our next step consists in verifying that when passing from $\Psi(z)$ to the bivariate generating function $\Psi(z, s) = \Psi\left(\left(\frac{wz}{w^2 z^2 - z + 1}\right)^2\right)$, then there exists a singular expansion of the form $O\left(\left(1 - \frac{z}{\rho_3(s)}\right)^4 \ln\left(\frac{1}{1 - \frac{z}{\rho_3(s)}}\right)\right)$, parameterized in s .

Claim 3. Let $0 < \varepsilon < 1$, then for any $|s| < \varepsilon$ and $z \in \Delta_{\rho_3(s)}(\phi, R)$, we have $\Psi(z, s) = O\left(\left(1 - \frac{z}{\rho_3(s)}\right)^4 \ln\left(\frac{1}{1 - \frac{z}{\rho_3(s)}}\right)\right)$, and the error bound is uniform for s in a neighborhood of 0.

To prove the claim we first observe that Claim 2 implies

$$\Psi(z) = \kappa(1 - 16z)^4 \ln\left(\frac{1}{1 - 16z}\right) + R(z) \tag{4.13}$$

for some absolute constant κ and $R(z)$ is the uniform error bound for $z \in \Delta_{1/16}(\phi, R) \cap U(\frac{1}{16}, \varepsilon)$. That is, for $z \in \Delta_{1/16}(\phi, R) \cap U(\frac{1}{16}, \varepsilon)$, there exists some absolute constant c , such that $|R(z)| \leq c \cdot |1 - 16z|$ holds. According to Lemma 3.2 we have

$$\begin{aligned} \Xi_3(z, s) &= \frac{1}{e^s z^2 - z + 1} O\left(\left(1 - 16\left(\frac{e^{1/2s} z}{e^s z^2 - z + 1}\right)^2\right)^4 \right. \\ &\quad \left. \times \ln \frac{1}{\left(1 - 16\left(\frac{e^{1/2s} z}{e^s z^2 - z + 1}\right)^2\right)} \right) \\ &= \frac{\kappa}{e^s z^2 - z + 1} \left(1 - 16\left(\frac{e^{1/2s} z}{e^s z^2 - z + 1}\right)^2\right)^4 \\ &\quad \times \ln \left(\frac{1}{1 - 16\left(\frac{e^{1/2s} z}{e^s z^2 - z + 1}\right)^2}\right) + R(z, e^s). \end{aligned}$$

We expand $(1 - 16(\frac{e^{1/2s} z}{e^s z^2 - z + 1})^2)^4 \ln(\frac{1}{1 - 16(\frac{e^{1/2s} z}{e^s z^2 - z + 1})^2})$ around $z = \rho_3(s)$, where $\rho_3(s)$ is the solution of $\frac{ze^{1/2s}}{e^s z^2 - z + 1} = \frac{1}{4}$ of minimal modulus. Lemma 3.2 implies that

$$\rho_3(s) = \frac{4e^{1/2s} + 1 - \sqrt{12e^s + 8e^{1/2s} + 1}}{2e^s}$$

is the unique dominant singularity. As a function in s we have $\rho_3'(0) = -\frac{3}{2} + \frac{13}{42}\sqrt{21} \neq 0$. The term $\sqrt{12e^s + 8e^{1/2s} + 1}$ in $\rho_3(s)$ produces two branching points parameterized by s , i.e. $w = e^{1/2s} = -\frac{1}{6}$ and $w = e^{1/2s} = -\frac{1}{2}$, or equivalently $s = 2\ln\frac{1}{2} + 2\pi i$ and $s = 2\ln\frac{1}{6} + 2\pi i$, respectively. The interval between $2\ln\frac{1}{6} + 2\pi i$ and $2\ln\frac{1}{2} + 2\pi i$ divides the complex plane of s into two analytic branches. For any $0 < \varepsilon < \min\{|2\ln\frac{1}{2} + 2\pi i|, |2\ln\frac{1}{6} + 2\pi i|\} = 6.4343$, the region $|s| < \varepsilon$ is disjoint to the interval $[(2\ln\frac{1}{6}, 2\pi), (2\ln\frac{1}{2}, 2\pi)]$. Therefore, $\rho_3(s)$ is analytic for $|s| < \varepsilon$. We next consider $q(z, s) = 1 - 16(\frac{e^{1/2s} z}{e^s z^2 - z + 1})^2$ as a function of z and compute the Taylor expansion at $\rho_3(s)$.

$$q(z, s) = \alpha(\rho_3(s) - z) + O(z - \rho_3(s))^2$$

and setting $\alpha = \frac{\sqrt{21}}{5 - \sqrt{21}}$

$$\begin{aligned} &\frac{1}{e^s z^2 - z + 1} \left[q(z, s)^4 \ln \frac{1}{q(z, s)} \right] \\ &= \frac{(\alpha(\rho_3(s) - z) + O(z - \rho_3(s))^2)^4 \ln \frac{1}{\alpha(\rho_3(s) - z) + O(z - \rho_3(s))^2}}{e^s(z - \rho_3(s))^2 + (2\rho_3(s)e^s - 1)(z - \rho_3(s)) - 3\rho_3(s)^2 e^s + \rho_3(s) + 1} \\ &= \frac{\left([\alpha + O(z - \rho_3(s))](\rho_3(s) - z)^4 \ln \frac{1}{[\alpha + O(z - \rho_3(s))](\rho_3(s) - z)} \right)}{O(z - \rho_3(s)) - 3\rho_3(s)^2 e^s + \rho_3(s) + 1} \\ &= O(\rho_3(s) - z)^4 \ln \left(\frac{1}{\rho_3(s) - z} \right). \end{aligned}$$

According to Theorem 4.1 for $r = 4$, we obtain the error term $R(z, e^s)$ in the expansion of $\frac{1}{e^s z^2 - z + 1} [q(z, s)^4 \ln \frac{1}{q(z, s)}]$ is uniform for s in a neighborhood of 0. We observe that the resulting error bound for $\Xi_3(z, s)$ is the sum $R(z, e^s) + R_1(z, e^s)$, where

$$|R(z, e^s)| \leq c \cdot \left| 1 - 16\left(\frac{e^{1/2s} z}{e^s z^2 - z + 1}\right) \right| = O(\rho_3(s) - z).$$

Therefore, the error bound for the expansion of bivariate $\Xi_3(z, s)$ is uniform and Claims 3 is proved. We proceed by using the scaling property of Taylor coefficients $[z^n]f(z) = \gamma^n [z^n]f(\frac{z}{\gamma})$ and apply Theorem 4.1. From Theorem 4.1 we obtain the key information about the coefficients of $\Xi_3(z, s)$ which allows us to substitute $\varphi_{n,3}(\frac{it}{\sigma_n})$ in Eq. (4.17):

$$\begin{aligned} [z^n]\Xi_3(z, s) &= K(s) \frac{4!}{n(n-1)\dots(n-4)} (\rho_3(s)^{-1})^n \\ &\quad \times \left(1 - O\left(\frac{1}{n}\right) \right) \text{ for some } K(s) \in \mathbb{C}, \end{aligned} \tag{4.14}$$

where the error term is again uniform for s from a neighborhood of origin.

Suppose we are given the random variable (r.v.) ξ_n with mean μ_n and variance σ_n^2 . We consider the rescaled r.v. $\eta_n = (\xi_n - \mu_n)\sigma_n^{-1}$ and the characteristic function of η_n :

$$f_{\eta_n}(t) = \mathbb{E}[e^{it\eta_n}] = \mathbb{E}[e^{it(\xi_n/\sigma_n)}] e^{-i(\mu_n/\sigma_n)t}. \tag{4.15}$$

In particular, for $\xi_n = X_n$ we obtain, substituting the term $\mathbb{E}[e^{it\eta_n}]$

$$f_{X_n}(t) = \left(\sum_{h=0}^n \frac{\mathcal{S}'_3(n, h)}{\mathcal{S}_3(n)} e^{it(h/\sigma_n)} \right) e^{-i(\mu_n/\sigma_n)t}. \tag{4.16}$$

Since $\varphi_{n,3}(s) = \sum_{h \leq n/2} \mathcal{S}'_3(n, h) e^{hs}$, we can interpret $\mathcal{S}_3(n) = \sum_{h \leq n/2} \mathcal{S}'_3(n, h)$ as $\varphi_{n,3}(0)$ and $\varphi_{n,3}(\frac{it}{\sigma_n}) = \sum_{h \leq n/2} \mathcal{S}'_3(n, h) e^{h(it/\sigma_n)}$, respectively. Therefore, we have

$$f_{X_n}(t) = \frac{1}{\varphi_n(0)} \varphi_n\left(\frac{it}{\sigma_n}\right) e^{-i(\mu_n/\sigma_n)t}. \tag{4.17}$$

For $|s| < \varepsilon$, Eq. (4.14) yields $\varphi(s) = [z^n]\Xi_3(z, s) \sim K(s) \frac{4!}{n(n-1)\dots(n-4)} (\rho_3(s)^{-1})^n$ with uniform error term and we accordingly obtain

$$f_{X_n}(t) \sim \frac{K\left(\frac{it}{\sigma_n}\right)}{K(0)} \left[\frac{\rho_3\left(\frac{it}{\sigma_n}\right)}{\rho_3(0)} \right]^{-n} e^{-i(\mu_n/\sigma_n)t}. \quad (4.18)$$

where the error term is uniform for t from any bounded interval. Taking the logarithm we obtain

$$\ln f_{X_n}(t) \sim \ln \frac{K\left(\frac{it}{\sigma_n}\right)}{K(0)} - n \ln \frac{\rho_3\left(\frac{it}{\sigma_n}\right)}{\rho_3(0)} - i \frac{\mu_n}{\sigma_n} t. \quad (4.19)$$

Expanding $g(s) = \ln \frac{\rho_3(s)}{\rho_3(0)}$ in its Taylor series at $s = 0$ (note that $g(0) = 0$ holds) yields

$$\ln \frac{\rho_3\left(\frac{it}{\sigma_n}\right)}{\rho_3(0)} = \frac{\rho_3'(0)}{\rho_3(0)} \frac{it}{\sigma_n} - \left[\frac{\rho_3''(0)}{\rho_3(0)} - \left(\frac{\rho_3'(0)}{\rho_3(0)} \right)^2 \right] \frac{t^2}{2\sigma_n^2} + O\left(\left(\frac{it}{\sigma_n}\right)^3\right) \quad (4.20)$$

and therefore,

$$\begin{aligned} \ln f_{X_n}(t) \sim & \ln \frac{K\left(\frac{it}{\sigma_n}\right)}{K(0)} - n \\ & \times \left\{ \frac{\rho_3'(0)}{\rho_3(0)} \frac{it}{\sigma_n} - \frac{1}{2} \left[\frac{\rho_3''(0)}{\rho_3(0)} - \left(\frac{\rho_3'(0)}{\rho_3(0)} \right)^2 \right] \frac{t^2}{\sigma_n^2} \right. \\ & \left. + O\left(\left(\frac{it}{\sigma_n}\right)^3\right) \right\} - \frac{i\mu_n t}{\sigma_n}. \end{aligned} \quad (4.21)$$

Claim 3 implies $\Xi_3(z, s) = O((\rho_3(s) - z)^4 \ln \frac{1}{\rho_3(s) - z})$ is analytic in s where s is contained in a disc of radius ε around 0. Hence, $\Xi_3(z, s)$ is in particular continuous in s for $|s| < \varepsilon$ and we can conclude from Eq. (4.14) for fixed $t \in]-\infty, \infty[$

$$\lim_{n \rightarrow \infty} \left(\ln K\left(\frac{it}{\sigma_n}\right) - \ln K(0) \right) = 0. \quad (4.22)$$

In view of Eq. (4.21) we introduce

$$\mu = -\frac{\rho_3'(0)}{\rho_3(0)}, \quad \sigma = \left\{ \left(\frac{\rho_3'(0)}{\rho_3(0)} \right)^2 - \frac{\rho_3''(0)}{\rho_3(0)} \right\}$$

and Eq. (4.21) becomes

$$\ln f_{X_n}(t) \sim -\frac{t^2}{2} + O\left(\left(\frac{it}{\sigma_n}\right)^3\right) \quad (4.23)$$

with uniform error term for t from any bounded interval. This is equivalent to $\lim_{n \rightarrow \infty} f_{X_n}(t) = \exp(-\frac{t^2}{2})$ with uniform error term. The Lévy–Cramér Theorem (Theorem 4.2) implies now Eq. (4.6) and it remains to compute the values for μ and σ which are given by

$$\mu = -\frac{\rho_3'(0)}{\rho_3(0)} = -\frac{-\frac{3}{2} + \frac{13}{42}\sqrt{21}}{\frac{5}{2} - \frac{1}{2}\sqrt{21}} = 0.39089, \quad (4.24)$$

$$\sigma^2 = \mu^2 - \frac{\rho_3''(0)}{\rho_3(0)} = \mu^2 - \frac{1 - \frac{94}{441}\sqrt{21}}{\frac{5 - \sqrt{21}}{2}} = 0.041565, \quad (4.25)$$

whence Eq. (4.7) and the proof of Theorem 4.3 is complete. \square

5. The local limit theorem

In this section we complement the central limit theorem presented in the previous section

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{X_n - \mu_n}{\sigma_n} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$$

by considering a “local” perspective on the limiting distribution of X_n . For the local limit theorem we analyze the difference between $\mathbb{P}(x \leq \frac{X_n - \mu_n}{\sigma_n} < x + 1)$ and $\frac{1}{\sqrt{2\pi}} \int_x^{x+1} e^{-t^2/2} dt$ as n tends to infinity. X_n satisfies a local limit theorem on some set $S \subset \mathbb{R}$ if and only if

$$\limsup_{n \rightarrow \infty} \sup_{x \in S} \left| \sigma_n \mathbb{P}\left(\frac{X_n - \mu_n}{\sigma_n} = x\right) - \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right| = 0. \quad (5.1)$$

One key condition formulated in Eq. (5.2) of Theorem 5.1 for proving a local limit theorem is given by

$$\frac{\varphi_n(s)}{\varphi_n(0)} \sim \exp(M(s)\beta_n + N(s)),$$

where $M(s)$ is differentiable and $N(s)$ is continuous in some ε -disc centered at 0. In view of Eqs. (4.17) and (4.18) in the proof of the central limit theorem, this condition alone implies the central limit theorem. In other words, the local limit theorem implies the central limit theorem. We have shown in the introduction that a central limit theorem does not imply a local limit theorem. Bender (1973) observed that the central limit theorem combined with certain smoothness conditions does imply the local limit theorem. Accordingly, in order to prove the local limit theorem for 3-noncrossing RNA structures with h arcs our strategy will consist in verifying such smoothness conditions (Hwang, 1998).

Theorem 5.1. Let $\varphi_n(s) = \sum_k a_{n,k} w^k$ and $w = e^s$. Suppose

$$\frac{\varphi_n(s)}{\varphi_n(0)} \sim \exp(M(s)\beta_n + N(s)) \quad (5.2)$$

holds uniformly for $|s| \leq \tau$, $s \in \mathbb{C}$ and $\tau > 0$, where the following conditions are satisfied:

- (i) $M(s)$ is differentiable and $N(s)$ is continuous in $|s| < \varepsilon$ and furthermore, $M(s)$ and $N(s)$ are independent of n .
- (ii) β_n is independent of s , $\beta_n \rightarrow \infty$ and $M''(0) > 0$.
- (iii) There exists constant δ and $c = c(\delta, r) > 0$, where $0 < \delta \leq \tau$ such that

$$\left| \frac{\varphi_n(r + it)}{\varphi_n(r)} \right| = O(\exp(-c\beta_n)) \quad (5.3)$$

holds uniformly for $-\tau \leq r \leq \tau$ and $\delta \leq |t| \leq \pi$ as n tends to infinity.

Then random variable X_n having distribution $\mathbb{P}(X_n = k) = a_{n,k}/a_n$ with mean $M'(0)\beta_n$ and variance $M''(0)\beta_n$ satisfies a local limit theorem on the real set $S = \{x|x = o(\sqrt{\beta_n})\}$, i.e.

$$\lim_{n \rightarrow \infty} \sup_{x \in S} \left| \sigma_n \mathbb{P} \left(\frac{X_n - \mu_n}{\sigma_n} = x \right) - \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right| = 0. \quad (5.4)$$

With the help of Theorem 5.1 we can now prove the local limit theorem for 3-noncrossing RNA structures with h arcs.

Theorem 5.2. Let $S'_3(n, h)$ be the number of 3-noncrossing RNA structures with exactly h arcs. Let X_n be the r.v. having the distribution

$$\forall h = 0, 1, \dots, \lfloor \frac{n}{2} \rfloor, \quad \mathbb{P}(X_n = h) = \frac{S'_3(n, h)}{S_3(n)}. \quad (5.5)$$

Then we have for set $S = \{x|x = o(\sqrt{n})\}$

$$\lim_{n \rightarrow \infty} \sup_{x \in S} \left| \sqrt{\sigma^2 n} \mathbb{P} \left(\frac{X_n - n\mu}{\sqrt{\sigma^2 n}} = x \right) - \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right| = 0, \quad (5.6)$$

where $\mu = 0.39089$ and $\sigma^2 = 0.041565$.

Proof. We will show that $\{ \frac{S'_3(n, h)}{S_3(n)} \}$ satisfies the conditions for Theorem 5.1. For $|s| \leq \varepsilon$, where ε is sufficiently small but fixed. The crucial equation implying the conditions of Theorem 5.1 is Eq. (4.14) of the proof of Theorem 4.3:

$$\varphi_{n,3}(s) = K(s) \frac{4!}{n(n-1) \dots (n-4)} (\rho_3(s)^{-1})^n \times \left(1 - O\left(\frac{1}{n}\right) \right), \quad K(s) \in \mathbb{C},$$

holds uniformly for $|s| < \varepsilon$. Therefore, we have

$$\frac{\varphi_{n,3}(s)}{\varphi_{n,3}(0)} = \frac{K(s)}{K(0)} \left(\frac{\rho_3(0)}{\rho_3(s)} \right)^n \left(1 - O\left(\frac{1}{n}\right) \right) \sim \exp \left(n \ln \left(\frac{\rho_3(0)}{\rho_3(s)} \right) + \ln \left(\frac{K(s)}{K(0)} \right) \right). \quad (5.7)$$

uniformly for $|s| < \varepsilon$. We set

$$\beta_n = n, \quad M(s) = \ln \left(\frac{\rho_3(0)}{\rho_3(s)} \right) \quad \text{and}$$

$$N(s) = \ln \left(\frac{K(s)}{K(0)} \right). \quad (5.8)$$

By construction n is independent of s and clearly $n \rightarrow \infty$ and $M(s)$ is differentiable and $N(s)$ is continuous for all s such that $|s| < \varepsilon$. In addition $M''(0)$ is analytic for $|s| < \varepsilon$ and we have $M''(0) = \mu^2 - \frac{1 - \frac{94}{401}\sqrt{21}}{2 - \sqrt{21}} = 0.041565 > 0$. Let $\delta = \varepsilon = \gamma$ and $-\varepsilon \leq r \leq \varepsilon$, we obtain

$$\frac{\varphi_{n,3}(r+it)}{\varphi_{n,3}(r)} \sim \exp \left(n \ln \left(\frac{\rho_3(r)}{\rho_3(r+it)} \right) + \ln \left(\frac{K(r+it)}{K(r)} \right) \right)$$

uniformly for $-\varepsilon \leq r \leq \varepsilon$ and $\varepsilon \leq |t| \leq \pi$. Since $\frac{K(s)}{K(0)}$ yields a constant factor and $K(s)$ is continuous for $|s| < \varepsilon$, it suffices to analyze $\ln \left(\frac{\rho_3(r)}{\rho_3(r+it)} \right)$. We observe $\rho_3(s) =$

$$\frac{1 + 4e^{s/2} - \sqrt{12e^s - 8e^{s/2} + 1}}{2e^s} \neq 0 \quad \text{for any complex } s \text{ where } |s| < \varepsilon.$$

The singularities of $\ln \left(\frac{\rho_3(0)}{\rho_3(s)} \right)$ correspond to the zeros of $12e^s - 8e^{s/2} + 1 = (2e^{s/2} + 1)(6e^{s/2} + 1)$, that is $e^{s/2} = -\frac{1}{2}$ or $-\frac{1}{6}$. Observe that for $|s| < \varepsilon$, $|e^{s/2}|$ is close to 1. Therefore, $\ln \left(\frac{\rho_3(r)}{\rho_3(r+it)} \right)$ is analytic for any $\varepsilon \leq |t| \leq \pi$ and $r \in]-\varepsilon, \varepsilon[$ and we can conclude

$$\left| \frac{\varphi_{n,3}(r+it)}{\varphi_{n,3}(r)} \right| = O \left(\exp \left(n \cdot \ln \left(\frac{\rho_3(it)}{\rho_3(0)} \right) \right) \right) = O \left(\exp \left(\operatorname{Re} \left(n \cdot \ln \left(\frac{\rho_3(r+it)}{\rho_3(r)} \right) \right) \right) \right)$$

uniformly for $-\varepsilon \leq r \leq \varepsilon$ and $\varepsilon \leq |t| \leq \pi$. Taylor expansion of $\ln \left(\frac{\rho_3(r+it)}{\rho_3(r)} \right)$ at 0 shows (see Eq. (4.20)), that the dominant real part of $\ln \left(\frac{\rho_3(r+it)}{\rho_3(r)} \right)$ is given by

$$\left[\left(\frac{\rho_3'(r)}{\rho_3(r)} \right)^2 - \frac{\rho_3''(r)}{\rho_3(r)} \right] \frac{t^2}{2!} < 0 \quad \text{for } r \in]-\varepsilon, \varepsilon[.$$

Setting $c_1 = \left[\frac{\rho_3'(r)}{\rho_3(r)} - \left(\frac{\rho_3'(r)}{\rho_3(r)} \right)^2 \right] \frac{\pi^2}{2!} > 0$ and $c_2 = \left[\frac{\rho_3''(r)}{\rho_3(r)} - \left(\frac{\rho_3'(r)}{\rho_3(r)} \right)^2 \right] \frac{\delta^2}{2!} > 0$ we can conclude

$$\left| \frac{\varphi_{n,3}(r+it)}{\varphi_{n,3}(r)} \right| = O(\exp(-c \cdot n))$$

for some $0 < c_2 < c < c_1$, uniformly for $-\varepsilon \leq r \leq \varepsilon$ and $\varepsilon \leq |t| \leq \pi$ and Theorem 5.1 applies, whence Theorem 5.2. \square

Acknowledgments

We are grateful to Prof. Jason Gao for helpful 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.

Appendix

Proof of Lemma 3.1. First we observe that for $x, w \in [-1, 1]$ the term $w^2 x^2 - x + 1$ is strictly positive. We set

$$F_k(x, w) = \sum_{n \geq 0} \sum_{h \leq n/2} S'_k(n, h) w^{2h} x^n \quad (6.1)$$

and compute

$$\begin{aligned}
 F_k(x, w) &= \sum_{n \geq 0} \sum_{h \leq n/2} \sum_{j=0}^h (-1)^j \binom{n-j}{j} \\
 &\quad \times \binom{n-2j}{2(h-j)} f_k(2(h-j), 0) w^{2h} x^n \\
 &= \sum_{n \geq 0} \sum_{j \leq n/2} \sum_{h=j}^{n/2} (-1)^j \binom{n-j}{j} \\
 &\quad \times \binom{n-2j}{2(h-j)} f_k(2(h-j), 0) w^{2h} x^n \\
 &= \sum_{j \geq 0} \sum_{n \geq 2j} \sum_{h=j}^{n/2} (-1)^j \binom{n-j}{j} \\
 &\quad \times \binom{n-2j}{2(h-j)} f_k(2(h-j), 0) w^{2h} x^n \\
 &= \sum_{j \geq 0} (-1)^j \frac{(wx)^{2j}}{j!} \sum_{n \geq 2j} (n-j)! \sum_{h=j}^{n/2} \\
 &\quad \times \binom{n-2j}{2(h-j)} f_k(2(h-j), 0) \frac{w^{2(h-j)}}{(n-2j)!} x^{n-2j}.
 \end{aligned}$$

We shift summation indices $n' = n - 2j$ and $h' = h - j$ and derive for the rhs the following expression

$$\begin{aligned}
 &= \sum_{j \geq 0} (-1)^j \frac{(wx)^{2j}}{j!} \sum_{n' \geq 0} (n' + j)! \sum_{h'=j}^{n'/2} \\
 &\quad \times \binom{n'}{2(h'-j)} f_k(2(h'-j), 0) \frac{w^{2(h'-j)}}{n'!} x^{n'-2j} \\
 &= \sum_{j \geq 0} (-1)^j \frac{(wx)^{2j}}{j!} \sum_{n' \geq 0} (n' + j)! \\
 &\quad \times \left\{ \sum_{h'=0}^{n'/2-j} \binom{n'}{2h'} f_k(2h', 0) w^{2h'} \right\} \frac{x^{n'}}{n'!}.
 \end{aligned}$$

The idea is now to interpret the term $\sum_{h'=0}^{n'/2-j} \binom{n'}{2h'} f_k(2h', 0) w^{2h'} \frac{x^{n'}}{n'!}$ as a product of the two power series e^x and $\sum_{n \geq 0} f_k(2n, 0) \frac{(wx)^{2n}}{(2n)!}$:

$$\begin{aligned}
 &\sum_{\ell \geq 0} \frac{x^\ell}{\ell!} \sum_{n \geq 0} f_k(2n, 0) \frac{(wx)^{2n}}{(2n)!} \\
 &= \sum_{n' \geq 0} \sum_{2n+\ell=n'} \left\{ \frac{1}{\ell!} \frac{1}{(2n)!} f_k(2n, 0) w^{2n} \right\} x^{n'} \\
 &= \sum_{n' \geq 0} \left\{ \sum_{n=0}^{n'/2} \binom{n'}{2n} f_k(2n, 0) w^{2n} \right\} \frac{x^{n'}}{n'!}.
 \end{aligned}$$

We set $\eta_{n'} = \left\{ \sum_{h'=0}^{n'/2} \binom{n'}{2h'} f_k(2h', 0) w^{2h'} \right\}$. By assumption we have $|x| < \rho_k(w)$ and we next derive, using the Laplace transformation and interchanging integration and summation,

$$\sum_{n' \geq 0} (n' + j)! \eta_{n'} \frac{x^{n'}}{n'!} = \int_0^\infty \sum_{n' \geq 0} \eta_{n'} \frac{(xt)^{n'}}{n'!} t^j e^{-t} dt. \quad (6.2)$$

Since $|x| < \rho_k(w)$ the above transformation is valid and using

$$\begin{aligned}
 &\sum_{n' \geq 0} \left\{ \sum_{n=0}^{n'/2} \binom{n'}{2n} f_k(2n, 0) w^{2n} \right\} \frac{x^{n'}}{n'!} \\
 &= \sum_{\ell \geq 0} \frac{x^\ell}{\ell!} \sum_{n \geq 0} f_k(2n, 0) \frac{(wx)^{2n}}{(2n)!}, \quad (6.3)
 \end{aligned}$$

we accordingly obtain

$$\begin{aligned}
 &\sum_{n' \geq 0} \eta_{n'} \frac{(xt)^{n'}}{n'!} t^j e^{-t} dt \\
 &= \int_0^\infty e^{tx} \sum_{n \geq 0} f_k(2n, 0) \frac{(wxt)^{2n}}{(2n)!} t^j e^{-t} dt. \quad (6.4)
 \end{aligned}$$

The next step is to substitute the term $\sum_{n' \geq 0} (n' + j)! \eta_{n'} \frac{x^{n'}}{n'!}$ in Eq. (6.2), whence consequently

$$\begin{aligned}
 F_k(x, w) &= \sum_{j \geq 0} (-1)^j \frac{(wx)^{2j}}{j!} \int_0^\infty e^{tx} \\
 &\quad \times \sum_{n \geq 0} f_k(2n, 0) \frac{(wxt)^{2n}}{(2n)!} t^j e^{-t} dt \\
 &= \int_0^\infty \sum_{j \geq 0} (-1)^j \frac{(wx)^{2j}}{j!} e^{tx} \\
 &\quad \times \sum_{n \geq 0} f_k(2n, 0) \frac{(wxt)^{2n}}{(2n)!} t^j e^{-t} dt.
 \end{aligned}$$

The summation over the index j is just an exponential function and we derive

$$\begin{aligned}
 &= \int_0^\infty e^{-(w^2x^2-x+1)t} \sum_{n \geq 0} f_k(2n, 0) \frac{(wxt)^{2n}}{(2n)!} dt \\
 &= \int_0^\infty e^{-(w^2x^2-x+1)t} \sum_{n \geq 0} f_k(2n, 0) \frac{1}{(2n)!} \\
 &\quad \times \left(\frac{wx}{w^2x^2-x+1} \right)^{2n} ((w^2x^2-x+1)t)^{2n} dt.
 \end{aligned}$$

We proceed by transforming the integral introducing $u = (w^2x^2 - x + 1)t$, i.e. $dt = (w^2x^2 - x + 1)^{-1} du$ and accordingly arrive at

$$\begin{aligned}
 F_k(x, w) &= \sum_{n \geq 0} f_k(2n, 0) \frac{1}{(2n)!} \left(\frac{wx}{w^2x^2-x+1} \right)^{2n} \\
 &\quad \times \int_0^\infty e^{-(w^2x^2-x+1)t} ((w^2x^2-x+1)t)^{2n} dt \\
 &= \sum_{n \geq 0} f_k(2n, 0) \frac{1}{(2n)!} \left(\frac{wx}{w^2x^2-x+1} \right)^{2n} \\
 &\quad \times \frac{1}{w^2x^2-x+1} (2n)! \\
 &= \frac{1}{w^2x^2-x+1} \sum_{n \geq 0} f_k(2n, 0) \left(\frac{wx}{w^2x^2-x+1} \right)^{2n},
 \end{aligned}$$

In particular for $w = 1$,

$$\sum_{n \geq 0} \mathbf{S}_k(n) x^n = \frac{1}{x^2 - x + 1} \sum_{n \geq 0} f_k(2n, 0) \left(\frac{x}{x^2 - x + 1} \right)^{2n} \quad (6.5)$$

holds for any $x \in \mathbb{R}$, satisfying $|x| < \rho_k(1)$, and where $\rho_k(1)$ is the radius of convergence of the power series $\sum_{n \geq 0} \mathbf{S}_k(n) z^n$ over \mathbb{C} , that is, Eq. (6.5) holds for $x \in]-\rho_k(1), \rho_k(1)[$. From complex analysis we know that any two functions that are analytic at 0 and coincide on an open interval which includes 0 are identical. Therefore, Eq. (6.5) holds for $z \in \mathbb{C}$, $|z| < \rho_k(1)$, and the proof of the lemma is complete. \square

References

- Mapping RNA form and function. *Science* 2, 2005.
- Bender, E.A., 1973. Central and local limit theorem applied to asymptotic enumeration. *J. Combin. Theory A* 15, 91–111.
- Chamorro, M., Parkin, N., Varmus, H.E., 1991. An RNA pseudoknot and an optimal heptameric shift site are required for highly efficient ribosomal frameshifting on a retroviral messenger RNA. *Proc. Natl Acad. Sci. USA* 89, 713–717.
- Chen, J., Blasco, M., Greider, C., 2000. Secondary structure of vertebrate telomerase RNP. *Cell* 100, 503–514.
- Chen, W.Y.C., Deng, E.Y.P., Du, R.R.X., Stanley, R.P., Yan, C.H., 2007. Crossings and nestings of matchings and partitions. *Trans. Am. Math. Soc.* 359, 1555–1575.
- Flajolet, P., Fill, J.A., Kapur, N., 2005. Singularity analysis, hadamard products, and tree recurrences. *J. Comput. Appl. Math.* 174, 271–313.
- Gao, Z., Richmond, L.B., 1992. Central and local limit theorems applied to asymptotic enumeration. *J. Appl. Comput. Anal.* 41, 177–186.
- Gessel, I.M., Zeilberger, D., 1992. Random walk in a Weyl chamber. *Proc. Am. Math. Soc.* 115, 27–31.
- Hofacker, I.L., Schuster, P., Stadler, P.F., 1998. Combinatorics of RNA secondary structures. *Discr. Appl. Math.* 88, 207–237.
- Hwang, H.K., 1998. Large deviations of combinatorial distributions. II. local limit theorems. *Ann. Appl. Probab.* 8 (1), 163–181.

- Jin, E.Y., Qin, J., Reidys, C.M., 2007. Combinatorics of RNA structures with pseudoknots. *Bull. Math. Biol.* in press.
- Jin, E.Y., Reidys, C.M., 2007. Asymptotics of rna structures with pseudoknots. *Bull. Math. Biol.*, accepted.
- Konings, D.A.M., Gutell, R.R., 1995. A comparison of thermodynamic foldings with comparatively derived structures of 16s and 16s-like rRNAs. *RNA* 1, 559–574.
- Lindstroem, B., 1973. On the vector representation of induced matroids. *Bull. London Math. Soc.* 5, 85–90.
- Loria, A., Pan, T., 1996. Domain structure of the ribozyme from eubacterial ribonuclease p. *RNA* 2, 551–563.
- McCaskill, J.S., 1990. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers* 29, 1105–1119.
- Odlyzko, A.M., 1995. *Handbook of Combinatorics*. Elsevier, Amsterdam (Chapter 22).
- Schmitt, W.R., Waterman, M.S., 1994. Linear trees and RNA secondary structure. *Discr. Appl. Math.* 51, 317–323.
- Tacker, M., Fontana, W., Stadler, P.F., Schuster, P., 1994. Statistics of RNA melting kinetics. *Eur. Biophys. J.* 23, 29–38.
- Tacker, M., Stadler, P.F., Bauer, E.G., Hofacker, I.L., Schuster, P., 1996. Algorithm independent properties of RNA secondary structure predictions. *Eur. Biophys. J.* 25, 115–130.
- Tuerk, C., MacDougall, S., Gold, L., 1992. RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. *Proc. Natl Acad. Sci. USA* 89, 6988–6992.
- van Batenburg, F.H.D., Gulyaev, A.P., Pleij, C.W.A., 2001. Pseudobase: structural information on RNA pseudoknots. *Nucleic Acids Res.* 29 (1), 194–195.
- Waterman, M.S., 1978. Secondary structure of single-stranded nucleic acids. *Adv. Math. I* 1 (Suppl.), 167–212.
- Waterman, M.S., Smith, T.F., 1986. Rapid dynamic programming algorithms for RNA secondary structure. *Adv. Appl. Math.* 7, 455–464.
- Westhof, E., Jaeger, L., 1992. RNA pseudoknots. *Current Opinion Struct. Biol.* 2, 327–333.
- Zuker, M., Sankoff, D., 1984. RNA secondary structures and their prediction. *Bull. Math. Biol.* 46 (4), 591–621.