Discrete Mathematics and Theoretical Computer Science **AC**, 2003, 265–276

Non-crossing trees revisited: cutting down and spanning subtrees

Alois Panholzer[†]

Institut für Algebra und Computermathematik, Technische Universität Wien, Wiedner Hauptstraße 8–10, A-1040 Wien, Austria.
`Alois.Panholzer@tuwien.ac.at`

Here we consider two parameters for random non-crossing trees: (i) the number of random cuts to destroy a size-n non-crossing tree and (ii) the spanning subtree-size of p randomly chosen nodes in a size-n non-crossing tree. For both quantities, we are able to characterise for $n \to \infty$ the limiting distributions. Non-crossing trees are almost conditioned Galton-Watson trees, and it has been already shown, that the contour and other usually associated discrete excursions converge, suitable normalised, to the Brownian excursion. We can interpret parameter (ii) as a functional of a conditioned random walk, and although we do not have such an interpretation for parameter (i), we obtain here limiting distributions, that are also arising as limits of some functionals of conditioned random walks.

Keywords: Non-crossing trees, generating function, limiting distributions

1 Introduction

A non-crossing tree is a tree drawn on the plane having as vertices a set of points on the boundary of a circle, whose edges are straight line segments that do not cross. We consider the vertices labelled clockwise from 1 to n, where the root of the tree is vertex 1. Fig. 1 gives an example of a non-crossing tree.

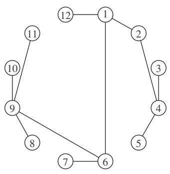

Fig. 1: A non-crossing tree of size 12.

It is well known, that the number T_n of different non-crossing trees of size n is equal to the number of ternary trees of size $n-1$. Therefore it holds

$$T_n = \frac{1}{2n-1}\binom{3n-3}{n-1}. \tag{1}$$

[†]Supported by the "Stiftung Aktion Österreich-Ungarn" within the project "Applications of Trees in Computer Science".

In this paper, we are considering two parameters for random non-crossing trees, as they are described in the following subsections: the number of random cuts to destroy a random size-n non-crossing tree, and the spanning subtree-size of p randomly chosen nodes in a random size-n non-crossing tree. The size of a tree T is here always measured by the number of nodes of T. Although the here treated quantities are not directly a study of a random walk, there are close relations to discrete random walks and might thus still of interest.

First, in [MP02] it is shown, that the analysed structure, the non-crossing trees, are almost conditioned Galton-Watson trees. They behave in many parameters like conditioned Galton-Watson trees, or equivalently, like simply generated trees (see e. g. [Ald91] for a description of these tree models). But for Galton-Watson trees, it is well known, that the contour and other usually associated discrete excursions converge, suitable normalised, to the Brownian excursion. Thus parameters for non-crossing trees are often related to parameters of a conditioned random walk, as it is described in [MP02].

Second, as limit law for the number of random cuts to destroy a random size-n non-crossing tree we obtain a Rayleigh-distribution, and as limit law for the size of the spanning subtree of p randomly chosen nodes in a random size-n non-crossing tree we obtain for $p=2$ also a Rayleigh-distribution and for $p>2$ generalised Gamma-distributions. Thus we obtain limiting distributions, that are arising also as limits of some functionals of conditioned random walks.

For general Galton-Watson trees, the parameter random cuts to destroy a non-crossing tree is still not analysed, but quite recently in [Pan03], it was at least possible to obtain limiting distribution results for tree families, that preserve randomness when cutting-off a random edge. The parameter spanning subtree-size of p randomly chosen nodes was already analysed for Galton-Watson trees in [Pan02]. As an insightful referee remarked, it must be possible for non-crossing trees (and also for general Galton-Watson trees) to express the limiting distribution of this quantity as a functional of the Brownian excursion. Following Aldous' construction in [Ald91], it is indeed possible to describe for non-crossing trees this relation as

$$\frac{Y_{n,p}}{\sqrt{n}} \xrightarrow{(d)} \sqrt{\tfrac{8}{3}}\Big(\sum_{J\subseteq\{1,\dots,p\},J\neq\emptyset} (-1)^{|J|+1} \min\{e(t) : t\in[\min\{U_j : j\in J\},\max\{U_j : j\in J\}]\} \\ - \min\{e(t) : t\in[\min\{U_1,\dots,U_p\},\max\{U_1,\dots,U_p\}]\}\Big),$$

where $Y_{n,p}$ is defined in Subsection 1.2, $e(t)$, $0\le t\le 1$ denotes the normalized Brownian excursion and $U_1,\dots,U_p$ are independent random variables, that are uniformly distributed on $[0,1]$. As usual $\xrightarrow{(d)}$ denotes convergence in distribution. Nevertheless, although this gives a relation to conditioned random walks, it is not obvious how to compute the limiting distribution directly from this description. Thus it might be interesting, to present and discuss the here obtained results in the context of random walks.

1.1 Random cuts to destroy a non-crossing tree

We are considering here the following "Cutting-down"-procedure. Starting with a tree T of size n, we choose one of the $n-1$ edges in the tree and remove this edge of T. The tree falls into two subtrees, where one of them, let us call it $\tilde{T}$ contains the root and has size k with $1\le k\le n-1$. We can now iterate the edge-removing procedure with $\tilde{T}$ and choose one of the $k-1$ edges in $\tilde{T}$, select the subtree containing the root, and so on. Finally we will obtain a tree consisting only of the root itself and we stop. An example for cutting-down a non-crossing tree is given in Figure 2.

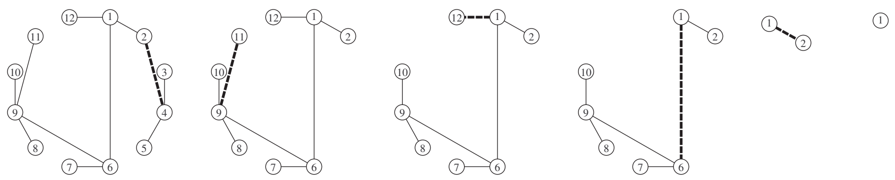

Fig. 2: Destroying in 5 steps a non-crossing tree of size 12 (dotted lines mark edges, that will be removed next).

A natural question is now: "How many edges will be removed of a size-n tree, until the root is isolated, or equivalently, how many steps are done until the tree is destroyed?"
We are considering this problem here for non-crossing trees, that means, we study under the probability model, described below, the random variable X_n, which counts the number of edges that will be removed from a non-crossing tree of size n until the root is isolated. We assume here, that all T_n non-crossing trees of size n are chosen for the cutting down procedure with equal probability, and also that the edges which will be removed are at each stage chosen at random from the remaining subtree containing the root.

This problem was originally studied by Meir and Moon [MM70] for random Cayley-trees (random labelled unordered trees), where exact and asymptotic results for the first two moments for the number of random cuts to destroy a tree are obtained. Chassaing and Marchand [CM02] were even able to characterise the limiting distribution of the suitably normalised parameter for random Cayley-trees, where they used an interpretation in terms of hashing tables. It turned out, that the number of random cuts, that are needed to destroy the tree, are asymptotically Rayleigh distributed.

In Section 3 it is shown, how we can obtain a corresponding limiting distribution result for random non-crossing trees. Since Chassaing and Marchand used in their proof relations between Cayley-trees and hashing tables, we cannot apply their ideas, but we will use the recursive nature of the cutting-down procedure. By solving the obtained recurrence via generating functions, we can describe the asymptotic behaviour of every moment of X_n:

Theorem 1. *The r-th moments $\mathbb{E}\{X_n^r\} := \sum_{m\geq 0} m^r \mathbb{P}\{X_n = m\}$ ($r = 1,2,\dots$) of the number of random cuts X_n, that will be done to destroy a random non-crossing tree of size n, are asymptotically for fixed r and $n \to \infty$ given by*

$$\mathbb{E}\{X_n^r\} = 3^{\frac{r}{2}} \Gamma\left(\frac{r}{2}+1\right) n^{\frac{r}{2}} \left(1 + O\left(\tfrac{\log n}{n}\right)\right).$$

If $X_{[\beta]}$ denotes a random variable, that is Rayleigh distributed with parameter β, it has the density function

$$f_{[\beta]}(x) = \frac{x}{\beta^2} e^{-\frac{1}{2}\left(\frac{x}{\beta}\right)^2}, \ \text{ for } \ x \geq 0, \ \text{ and } \ f_{[\beta]}(x) = 0 \ \text{ otherwise.}$$

The r-th moments $\mathbb{E}\{X_{[\beta]}^r\}$ are then given by

$$\mathbb{E}\{X_{[\beta]}^r\} = 2^{\frac{r}{2}} \beta^r \Gamma\left(\frac{r}{2}+1\right).$$

Since the Rayleigh distribution is fully characterised by its moments, we can apply the Theorem of Fréchet and Shohat (see e. g. [Fis63]) and obtain, that the limiting distribution of the normalised random variable $\frac{X_n}{\sqrt{n}}$ is a Rayleigh distribution with parameter $\sqrt{3/2}$. This is described by the following corollary.

Corollary 2. *The number of random cuts X_n, that will be done to destroy a random non-crossing tree of size n are asymptotically for $n \to \infty$ Rayleigh distributed:*

$$\frac{X_n}{\sqrt{n}} \xrightarrow{(d)} X_{[\sqrt{3/2}]}.$$

1.2 Spanning subtree-size in non-crossing trees

Given a tree T and selecting p nodes $\{v_1,\ldots,v_p\}$ in the tree, we obtain the spanning subtree of $\{v_1,\ldots,v_p\}$ as the smallest subtree of T containing these p nodes. We are interested here in the size of the spanning subtree of $\{v_1,\ldots,v_p\}$. If we choose two nodes $\{v_1,v_2\}$, then this parameter is nothing else than the node-distance between v_1 and v_2, and the spanning subtree-size generalises therefore the parameter distance to more than two nodes.

We treat here this quantity for non-crossing trees: we study the random variable $Y_{n,p}$, which counts the size of the spanning subtree of p randomly chosen nodes in a random non-crossing tree of size n. It is here assumed, that all T_n non-crossing trees of size n can occur equally likely and also, that all $\binom{n}{p}$ possibilities to select p nodes in the tree of size n are equally likely.

This parameter was already analysed in [Pan02] for simply generated trees, and we will use some lemmata, which appeared in the analysis described therein. We obtain then the following limiting distribution result for non-crossing trees, which is proven in Section 4.

Theorem 3. *The size $Y_{n,p}$ of the spanning subtree of p randomly chosen nodes in a random non-crossing tree of size n converges for $n \to \infty$ and fixed $p \geq 2$ to a generalised Gamma-distribution:*

$$\frac{Y_{n,p}}{\sqrt{n}} \xrightarrow{(d)} Y_p,$$

where Y_p is a random variable with density function

$$f_p(x) = \frac{2x^{2p-3}}{(p-2)!}\left(\frac{3}{4}\right)^{p-1} e^{-\frac{3}{4}x^2}, \ \ for \ x \geq 0, \ \ and \ \ f_p(x) = 0 \ \ otherwise.$$

We get in particular, that $Y_{n,2}$, and thus the distance between two random nodes in a size-n non-crossing tree, is asymptotically for $n \to \infty$ Rayleigh distributed with parameter $\sqrt{2/3}$.

To obtain this result, we will in analogy to the study of the corresponding quantity for simply generated trees first study the auxiliary parameter size of the ancestor-tree. Selecting p nodes $\{v_1,\ldots,v_p\}$ in a tree T, the ancestor-tree of $\{v_1,\ldots,v_p\}$ is the subtree of T, which is spanned by the root and these p nodes or equivalently defined as the tree, containing all ascendants of the p chosen nodes. If we select only one node v_1, then the size of the ancestor tree is exactly the depth of this node in the tree and the ancestor-tree size generalises this parameter to more nodes.

We denote by $X_{n,p}$ the random variable, which counts the size of the ancestor-tree of p randomly chosen nodes in a random non-crossing tree of size n, where it is again assumed, that all T_n non-crossing trees of size n can occur equally likely and also, that all $\binom{n}{p}$ possibilities to select p nodes in the tree of size n are equally likely. We can prove in Section 4 the following theorem.

Theorem 4. *The size $X_{n,p}$ of the ancestor-tree of p randomly chosen nodes in a random non-crossing tree of size n converges for $n \to \infty$ and fixed $p \geq 1$ to a generalised Gamma-distribution:*

$$\frac{X_{n,p}}{\sqrt{n}} \xrightarrow{(d)} Y_{p+1},$$

where Y_p denotes the random variable defined in Theorem 3.

Thus $X_{n,1}$, the depth of a random node in a random size-n non-crossing tree, is asymptotically for $n \to \infty$ Rayleigh distributed.

An example of a non-crossing tree with the parameters ancestor-tree size and spanning subtree size of some selected nodes is given in Figure 3.

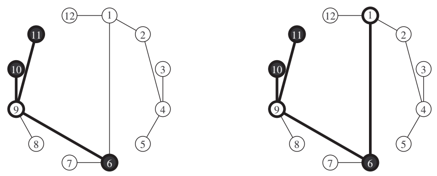

Fig. 3: Selecting the nodes $\{6, 10, 11\}$, we get that the ancestor-tree has size 5 and the spanning subtree has size 4.

2 Mathematical Preliminaries

In the mathematical analysis of the studied parameters, we will use a generating functions approach. The basic decomposition of non-crossing trees, which is described in [FN99], can be translated into equations for suitable chosen generating functions for the here considered parameters. Following [FN99], a non-crossing tree consists of a root, which is attached to a (possibly empty) sequence of butterflies, where a butterfly is a (ordered) pair of non-crossing trees, that share a common root (see Figure 4).

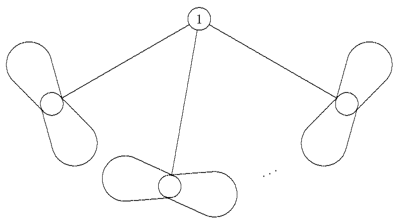

Fig. 4: The combinatorial decomposition of a non-crossing tree.

This combinatorial decomposition can be translated immediately into the following system of equations for the generating functions $T(z) = \sum_{n \geq 1} T_n z^n$ and $B(z) = \sum_{n \geq 1} B_n z^n$ of the number T_n of non-crossing trees of size n resp. the number B_n of butterflies of size n:

$$T(z) = \tfrac{z}{1-B(z)}, \quad B(z) = \tfrac{T^2(z)}{z}.$$

From this equation follows also, that $B(z) = \frac{z}{(1-B(z))^2}$, and thus, that the family of the butterflies (but not the non-crossing trees) are simply generated trees. One gets easily, that the number of butterflies of size n are given by

$$B_n = \frac{1}{n}\binom{3n-2}{n-1}. \tag{2}$$

One obtains further, that the generating functions $T(z)$ and $B(z)$ have their only dominant singularities (= singularities of smallest modulus) at $z = \frac{4}{27}$ with the local expansions

$$T(z) = \frac{2}{9} - \frac{2\sqrt{3}}{27}\sqrt{1-\frac{27}{4}z} + O\Big(1-\frac{27}{4}z\Big), \quad B(z) = \frac{1}{3} - \frac{2\sqrt{3}}{9}\sqrt{1-\frac{27}{4}z} + O\Big(1-\frac{27}{4}z\Big).$$

Either from the closed forms (1) and (2) or from these expansions of the generating functions via singularity analysis (see [FO90]), we obtain the asymptotic expansions

$$T_n = \frac{\sqrt{3}}{27\sqrt{\pi}}\left(\frac{27}{4}\right)^n n^{-\frac{3}{2}}\left(1+O\Big(\frac{1}{n}\Big)\right), \quad B_n = \frac{\sqrt{3}}{9\sqrt{\pi}}\left(\frac{27}{4}\right)^n n^{-\frac{3}{2}}\left(1+O\Big(\frac{1}{n}\Big)\right). \tag{3}$$

3 Proofs for results concerning the number of random cuts

The basic idea in our approach is to study the recurrence

$$\mathbb{P}\{X_n = m\} = \sum_{k=1}^{n-1} p_{n,k}\,\mathbb{P}\{X_k = m-1\}, \quad \text{for} \quad n \geq 2,\ m \geq 1, \tag{4}$$

with the initial values $\mathbb{P}\{X_1 = 0\} = 1$ and $\mathbb{P}\{X_n = 0\} = 0$ for $n \geq 2$. Here $p_{n,k}$ denotes the probability, that by choosing a random non-crossing tree of size n and removing a random edge, the remaining subtree containing the root, is of size k.

In order to reduce this problem to study (4), it is of course necessary to show, that randomness is preserved by cutting-off a random edge. This means, that after removing a randomly selected edge of a size-n tree, the remaining subtree containing the root, which is assumed to have size k, is after a natural order-preserving relabelling of the nodes with labels $\{1,\ldots,k\}$, again a *random* non-crossing tree of size k. Preserving randomness by cutting-off a random edge is a strong property, which is by far not satisfied for every tree family; one can easily see, that e. g. Binary search trees or Motzkin-trees do not have this property, and can't be attacked by this simple approach.

Lemma 5. *Let us assume, that we choose a random non-crossing tree of size n and also one of its $n-1$ edges at random, and after removing this edge, the obtained subtree containing the root is of size k. Then it holds, that this subtree is (after the order-preserving relabelling of the nodes) a random non-crossing tree of size k.*

Proof. If we consider a size-n non-crossing tree and remove one of its $n-1$ edges, we obtain a subtree of size $1 \leq k \leq n-1$ which contains the root and another subtree of size $n-k$. After the order-preserving relabellings, we can consider the subtree of size k as a non-crossing tree, whereas the other subtree is a butterfly of size $n-k$. Now we want to count, how often we obtain a particular pair of a non-crossing tree T of size k and a butterfly B of size $n-k$, when removing one edge of a non-crossing tree of size n. It will turn out, that this quantity depends only on k, not on the particular chosen trees T and B, and the lemma will be proven. To show this, we are considering the other way around and ask, in how many ways

$w(T)$ (it is apparently independent of B) can we add an edge to the pair of T and B, such that, after the order-preserving relabelling with labels $\{1,\dots,n\}$ we obtain a size-n non-crossing tree.

If the root of T has degree $d(root)$, then we can add an edge to the root and append B in $d(root)+1$ different ways. If we consider a non-root node v of T with out-degree $d^+(v)=i$, which has j left subtrees and $i-j$ right subtrees, then we can add an edge in $j+1$ different ways as a left subtree and in $i-j+1$ different ways as a right subtree and thus in $(j+1)+(i-j+1)=i+2$ different ways to append B. If we denote by k_i the number of non-root nodes in T with out-degree i, then we obtain

$$w(T)=d(root)+1+\sum_{i\geq 0}(i+2)k_i.$$

Counting the nodes resp. the edges in a tree, we obtain the equations

$$k=1+\sum_{i\geq 0}k_i \quad \text{resp.} \quad k-1=d(root)+\sum_{i\geq 0}ik_i.$$

With $d(root)=\sum_{i\geq 0}(1-i)k_i$, we get therefore

$$w(T)=1+\sum_{i\geq 0}(1-i+i+2)k_i=1+3\sum_{i\geq 0}k_i=3k-2. \tag{5}$$

Thus the quantity $w(T)$ depends only on the size of the remaining non-crossing tree T and is independent of the structure, and the lemma is proven. $\square$

Due to the bijective proof of above Lemma 5, we obtain further the equation $\sum_{k=1}^{n-1}(3k-2)T_kB_{n-k}=(n-1)T_n$, which leads thus to the following formula for the probabilities $p_{n,k}$, that appear in (4):

$$p_{n,k}=\frac{(3k-2)T_kB_{n-k}}{(n-1)T_n}=\frac{(3k-2)(2n-1)\binom{3k-3}{k-1}\binom{3(n-k)-2}{(n-k)-1}}{(n-1)(2k-1)(n-k)\binom{3n-3}{n-1}}. \tag{6}$$

To solve the resulting recurrence

$$\mathbb{P}\{X_n=m\}=\sum_{k=1}^{n-1}\frac{(3k-2)T_kB_{n-k}}{(n-1)T_n}\mathbb{P}\{X_k=m-1\}, \quad n\geq 2,\ m\geq 1, \tag{7}$$

we introduce the bivariate generating function $M(z,v)=\sum_{n\geq 1}\sum_{m\geq 0}T_n\mathbb{P}\{X_n=m\}z^nv^m$.

The recurrence (7) can then be translated into the linear first order differential equation

$$\frac{\frac{\partial}{\partial z}M(z,v)}{M(z,v)}=\frac{1-2vB(z)}{z(1-3vB(z))}. \tag{8}$$

To integrate the right side of (8), we change variables, and obtain

$$\begin{aligned}\log(M(z,v))&=\int^z\frac{1-2vB(t)}{t(1-3vB(t))}dt=\int^{B(z)}\frac{(1-2vB)(1-3B)}{B(1-B)(1-3vB)}dB\\&=\log(B(z))+\frac{2(1-2v)}{3v-1}\log\left(\frac{1}{1-B(z)}\right)+\frac{v-1}{3v-1}\log\left(\frac{1}{1-3vB(z)}\right)+\tilde{C}(v).\end{aligned}$$

Adapting to the initial value $\frac{\partial}{\partial z} M(z,v)\Big|_{z=0} = 1$, we obtain finally as solution for $M(z,v)$:

$$M(z,v) = B(z)(1-B(z)) \left(\frac{1-B(z)}{1-3vB(z)}\right)^{\frac{v-1}{3v-1}}. \tag{9}$$

In order to establish the asymptotic behaviour of every factorial moment $\mathbb{E}\{X_n^{\underline{r}}\} := \sum_{m\geq 0} m^{\underline{r}} \mathbb{P}\{X_n = m\}$, with $m^{\underline{r}} := m(m-1)\cdots(m-r+1)$, we substitute $w := v-1$ and extract coefficients at w^r, where we use

$$[w^r]M(z,v) = \sum_{n\geq 1} \sum_{m \geq r} T_n \frac{m^{\underline{r}}}{r!} \mathbb{P}\{X_n = m\} z^n. \tag{10}$$

Expanding (9) leads to

$$M(z,v) = B(z)(1-B(z)) e^{\frac{w}{2}\left[\sum_{k\geq 0}(-1)^k \left(\frac{3}{2}\right)^k w^k\right] \cdot \left[\log\left(\frac{1-B(z)}{1-3B(z)}\right) + \sum_{k\geq 1}\left(\frac{3B(z)}{1-3B(z)}\right)^k \frac{w^k}{k}\right]},$$

and thus to

$$[w^r]M(z,v) = B(z)(1-B(z)) \sum_{l=0}^{r} [w^{r-l}] \frac{1}{2^l} \left(\sum_{k\geq 0}(-1)^k \left(\tfrac{3}{2}\right)^k w^k\right)^l \left(\log\left(\tfrac{1-B(z)}{1-3B(z)}\right) + \sum_{k\geq 1}\left(\tfrac{3B(z)}{1-3B(z)}\right)^k \tfrac{w^k}{k}\right)^l.$$

The appearing functions $[w^r]M(z,v)$ are therefore a finite sum of functions, that have their only dominant singularities at $z = \frac{4}{27}$. Since $1 - 3B(z) \sim \frac{2\sqrt{3}}{3}\sqrt{1 - \frac{27}{4}z}$, we expand around $B(z) = \frac{1}{3}$ and obtain

$$\begin{aligned}
[w^r]M(z,v) = {} & \tfrac{B(z)(1-B(z))}{2} [w^{r-1}] \left(\sum_{k\geq 0}(-1)^k \left(\tfrac{3}{2}\right)^k w^k\right) \left(\log\left(\tfrac{1-B(z)}{1-3B(z)}\right) + \sum_{k\geq 1}\left(\tfrac{3B(z)}{1-3B(z)}\right)^k \tfrac{w^k}{k}\right) \\
& + \tfrac{B(z)(1-B(z))}{4} [w^{r-2}] \left(\sum_{k\geq 0}(-1)^k \left(\tfrac{3}{2}\right)^k w^k\right)^2 \left(\log\left(\tfrac{1-B(z)}{1-3B(z)}\right) + \sum_{k\geq 1}\left(\tfrac{3B(z)}{1-3B(z)}\right)^k \tfrac{w^k}{k}\right)^2 \\
& + O\left(\log^2\left(\tfrac{1}{1-3B(z)}\right) \tfrac{1}{(1-3B(z))^{r-3}}\right).
\end{aligned}$$

This gives further

$$\begin{aligned}
[w^r]M(z,v) &= \frac{B(z)(1-B(z))}{2(r-1)} \left(\frac{3B(z)}{1-3B(z)}\right)^{r-1} + O\left(\log\left(\frac{1}{1-3B(z)}\right) \frac{1}{(1-3B(z))^{r-2}}\right), \quad \text{for } r \geq 3, \\
[w^2]M(z,v) &= \frac{3B(z)^2(1-B(z))}{2(1-3B(z))} - \frac{3B(z)(1-B(z))}{4} \log\left(\frac{1-B(z)}{1-3B(z)}\right) + \frac{B(z)(1-B(z))}{8} \log^2\left(\frac{1-B(z)}{1-3B(z)}\right), \\
[w^1]M(z,v) &= \frac{B(z)(1-B(z))}{2} \log\left(\frac{1-B(z)}{1-3B(z)}\right).
\end{aligned}$$

Via singularity analysis, we obtain thus the following asymptotic expansions.

$$[z^n w^r]M(z,v) = \frac{\left(\frac{\sqrt{3}}{2}\right)^{r-1}}{9(r-1)} \left(\frac{27}{4}\right)^n \frac{n^{\frac{r-3}{2}}}{\Gamma(\frac{r-1}{2})} \left(1 + O\Big(\frac{\log n}{\sqrt{n}}\Big)\right), \quad \text{for } r \geq 3, \tag{11}$$

$$[z^n w^2]M(z,v) = \frac{\sqrt{3}}{18} \left(\frac{27}{4}\right)^n \frac{n^{-\frac{1}{2}}}{\Gamma(\frac{1}{2})} \left(1 + O\Big(\frac{\log n}{\sqrt{n}}\Big)\right), \ [z^n w^1]M(z,v) = \frac{1}{18} \left(\frac{27}{4}\right)^n \frac{1}{n} \left(1 + O\Big(\frac{\log n}{\sqrt{n}}\Big)\right).$$

By applying the reflection law of the Gamma-function $\Gamma\left(\frac{r}{2}+1\right)\Gamma\left(\frac{r-1}{2}\right)=\frac{r!\sqrt{\pi}}{(r-1)2^{r-1}}$, we obtain from (11) for fixed $r \geq 1$ (the constant in the O-remainder term depends of course on r):

$$[z^n w^r]M(z,v) = \frac{3^{\frac{r-1}{2}}\Gamma(\frac{r}{2}+1)}{9\sqrt{\pi}r!}\left(\frac{27}{4}\right)^n n^{\frac{r-3}{2}}\left(1+O\Big(\frac{\log n}{\sqrt{n}}\Big)\right). \tag{12}$$

With the asymptotic formula (3), we obtain by using (10) an asymptotic expansion of every factorial moment:

$$\mathbb{E}\{X_n^{\underline{r}}\} = 3^{\frac{r}{2}}\Gamma\left(\frac{r}{2}+1\right)n^{\frac{r}{2}}\left(1+O\Big(\frac{\log n}{n}\Big)\right).$$

Since it holds $\mathbb{E}\{X_n^r\} = \mathbb{E}\{X_n^{\underline{r}}\} + O\big(\mathbb{E}\{X_n^{\underline{r-1}}\}\big)$, we obtain Theorem 1 and also Corollary 2.

4 Proofs for results concerning the spanning subtree-size

We start by analysing the auxiliary parameter $X_{n,p}$, where we introduce also the corresponding random variable $\tilde{X}_{n,p}$ for butterflies: $\tilde{X}_{n,p}$ counts the size of the ancestor-tree of p randomly selected nodes in a random butterfly of size n. With generating functions

$$G(z,u,v) = \sum_{n\geq 1}\sum_{0\leq p\leq n}\sum_{m\geq 0} T_n\binom{n}{p}\mathbb{P}\{X_{n,p}=m\}z^n u^p v^m, \quad \tilde{G}(z,u,v) = \sum_{n\geq 1}\sum_{0\leq p\leq n}\sum_{m\geq 0} B_n\binom{n}{p}\mathbb{P}\{\tilde{X}_{n,p}=m\}z^n u^p v^m, \tag{13}$$

we can translate the combinatorial decomposition described in Section 2 into the following equations:

$$G(z,u,v) = \frac{zv(1+u)}{1-\tilde{G}(z,u,v)} + (1-v)T(z), \quad \tilde{G}(z,u,v) = \frac{zv(1+u)}{(1-\tilde{G}(z,u,v))^2} + (1-v)B(z). \tag{14}$$

The extra terms $(1-v)T(z)$ resp. $(1-v)B(z)$ are resulting from the case $p=0$.

In the sequel, we will use, that the butterflies are simply generated trees: $B(z) = z\tilde{\varphi}\big(B(z)\big)$, with $\tilde{\varphi}(t) = \frac{1}{(1-t)^2}$, and therefore the results from [Pan02] are applicable with this degree-generating function. With $\tau = \frac{1}{3}$ being the only solution of smallest modulus of the equation $\tilde{\varphi}(t) = t\tilde{\varphi}'(t)$ and $\rho = \frac{\tau}{\tilde{\varphi}(\tau)} = \frac{4}{27}$ being the only dominant singularity of $B(z)$, one obtains the following lemma, where we used the abbreviations D_u for the differential operator w. r. t. u and N_u for the operator, that evaluates at $u=0$.

Lemma 6 (proven in [Pan02]). *For $p \geq 1$ we have the following representations*

$$N_u D_u^p \tilde{G}(z,u,v) = \sum_{k=1}^{2p-1}\tilde{\alpha}_{p,k}(z)\left(\frac{vz}{1-vz\tilde{\varphi}'\big(B(z)\big)}\right)^k, \quad N_u D_u^p \tilde{\varphi}\big(\tilde{G}(z,u,v)\big) = \sum_{k=1}^{2p-1}\tilde{\beta}_{p,k}(z)\left(\frac{vz}{1-vz\tilde{\varphi}'\big(B(z)\big)}\right)^k,$$

with functions $\tilde{\alpha}_{p,k}(z)$, $\tilde{\beta}_{p,k}(z)$, that are analytic with their dominant singularities at $z=\rho$ and expansions $\tilde{\alpha}_{p,k}(z) = \tilde{\alpha}_{p,k} + O(\sqrt{\rho-z})$, $\tilde{\beta}_{p,k}(z) = \tilde{\beta}_{p,k} + O(\sqrt{\rho-z})$ in a neighborhood of $z=\rho$.

Furthermore, in the expansion $\tilde{\alpha}_{p,2p-1}(z) = \tilde{\alpha}_{p,2p-1} + O(\sqrt{\rho-z})$, it holds for $p \geq 1$:

$$\tilde{\alpha}_{p,2p-1} = \frac{(p-1)!}{\tilde{\varphi}''(\tau)2^{p-1}}\binom{2(p-1)}{p-1}\big(\tilde{\varphi}''(\tau)\tilde{\varphi}(\tau)\big)^p.$$

To evaluate the coefficients asymptotically, we will also use the following lemma:

Lemma 7 (proven in [Pan02]). *Let $\tilde{\alpha}(z)$ be analytic with its dominant singularity at $z=\rho$, where in a neighborhood of $z=\rho$ the expansion $\tilde{\alpha}(z)=\tilde{\alpha}+O(\sqrt{\rho-z})$ holds. Then we have for $n\to\infty$ and $m\le C\sqrt{n}$ with an arbitrary but fixed $C>0$ the following asymptotic expansion*

$$[z^n]\tilde{\alpha}(z)\left(z\tilde{\varphi}'\big(B(z)\big)\right)^m=\frac{\tilde{\alpha}\sqrt{2\rho\tau\tilde{\varphi}''(\tau)}m}{2\sqrt{\pi}\rho^n n^{\frac{3}{2}}}e^{-\frac{\rho\tau\tilde{\varphi}''(\tau)m^2}{2n}}\left(1+O\Big(\frac{1}{\sqrt{n}}\Big)+O\Big(\frac{m}{n}\Big)\right).$$

We obtain now from equation (14) for $p\ge 1$:

$$N_uD_u^pG(z,u,v)=zv\left(N_uD_u^p\frac{1}{1-\tilde{G}(z,u,v)}+p\,N_uD_u^{p-1}\frac{1}{1-\tilde{G}(z,u,v)}\right), \tag{15}$$

$$N_uD_u^p\frac{1}{1-\tilde{G}(z,u,v)}=\tilde{\varphi}\big(B(z)\big)N_uD_u^p\tilde{G}(z,u,v)+\sum_{l=1}^{p-1}\binom{p-1}{l}N_uD_u^l\tilde{\varphi}\big(\tilde{G}(z,u,v)\big)N_uD_u^{p-l}\tilde{G}(z,u,v). \tag{16}$$

From (15) and (16) follows then with Lemma 6 by induction for $p\ge 2$ the representation

$$N_uD_u^pG(z,u,v)=vz\sum_{k=1}^{2p-1}\alpha_{p,k}(z)\left(\frac{vz}{1-vz\tilde{\varphi}'\big(B(z)\big)}\right)^k,\ \text{ with }\ N_uD_uG(z,u,v)=vz\left(\frac{\alpha_{1,1}(z)\,vz}{1-vz\tilde{\varpi}'\big(B(z)\big)}+\frac{1}{1-B(z)}\right),$$

where around $z=\rho$ the expansions $\alpha_{p,k}(z)=\alpha_{p,k}+O(\sqrt{\rho-z})$ and in particular $\alpha_{p,2p-1}(z)=\tilde{\varphi}\big(B(z)\big)\tilde{\alpha}_{p,2p-1}(z)$, with $\tilde{\alpha}_{p,2p-1}(z)$ as in Lemma 6, hold.

Extracting coefficients, we get then for $m\ge 2p$:

$$[v^m]N_uD_u^pG(z,u,v)=\sum_{k=1}^{2p-1}\binom{m-2}{k-1}\frac{\alpha_{p,k}(z)}{\big(\tilde{\varphi}'\big(B(z)\big)\big)^{k+1}}\left(z\tilde{\varphi}'\big(B(z)\big)\right)^m,$$

and further for $m\le C\sqrt{n}$ with a fixed $C>0$ by using Lemma 7

$$\begin{aligned}[z^nv^m]N_uD_u^pG(z,u,v)&=[z^n]\frac{m^{2p-2}}{(2p-2)!}\frac{\alpha_{p,2p-1}(z)}{\big(\tilde{\varphi}\big(B(z)\big)\big)^{2p}}\left(z\tilde{\varphi}'\big(B(z)\big)\right)^m\left(1+O\Big(\frac{1}{m}\Big)\right)\\&=\frac{\tilde{\varphi}(\tau)}{\tilde{\varphi}'(\tau)}\frac{m^{2p-1}(\tilde{\varphi}''(\tau)\tilde{\varphi}(\tau))^p\sqrt{2\rho\tau\tilde{\varphi}''(\tau)}}{n^{\frac{3}{2}}(p-1)!2^p\sqrt{\pi}\rho^{n-2p+1}\tilde{\varphi}''(\tau)}e^{-\frac{\rho\tau\tilde{\varphi}''(\tau)m^2}{2n}}\left(1+O\Big(\frac{1}{\sqrt{n}}\Big)+O\Big(\frac{m}{n}\Big)+O\Big(\frac{1}{m}\Big)\right).\end{aligned}$$

Together with (3) and (13), we obtain for a fixed $C>0$ and $m\le C\sqrt{n}$

$$\mathbb{P}\{X_{n,p}=m\}=\frac{[z^nv^m]N_uD_u^pG(z,u,v)}{p!\binom{n}{p}T_n}=\frac{2m^{2p-1}}{n^p(p-1)!}\left(\frac{\rho\tau\tilde{\varphi}''(\tau)}{2}\right)^pe^{-\frac{\rho\tau\tilde{\varphi}''(\tau)m^2}{2n}}\left(1+O\big(\tfrac{1}{\sqrt{n}}\big)+O\big(\tfrac{m}{n}\big)+O\big(\tfrac{1}{m}\big)\right).$$

Specialising for $\tilde{\varphi}(t)$, ρ and τ, we obtain the following lemma.

Lemma 8. *For fixed $p\ge 1$, it holds for $x>0$ and $m=x\sqrt{n}+o(\sqrt{n})$, that*

$$\sqrt{n}\,\mathbb{P}\{X_{n,p}=m\}\sim x^{2p-1}\frac{2}{(p-1)!}\left(\frac{3}{4}\right)^pe^{-\frac{3}{4}x^2}.$$

Thus the density function of $\frac{X_{n,p}}{\sqrt{n}}$ is characterised, and we obtain immediately Theorem 4.

Now we are going to analyse the parameter $Y_{n,p}$, where we again introduce the corresponding random variable $\tilde{Y}_{n,p}$ for butterflies: $\tilde{Y}_{n,p}$ counts the size of the spanning subtree of p randomly selected nodes in a random butterfly of size n. With generating functions

$$F(z,u,v)=\sum_{n\geq 1}\sum_{0\leq p\leq n}\sum_{m\geq 0}T_n\binom{n}{p}\mathbb{P}\{Y_{n,p}=m\}z^n u^p v^m,\ \ \tilde{F}(z,u,v)=\sum_{n\geq 1}\sum_{0\leq p\leq n}\sum_{m\geq 0}B_n\binom{n}{p}\mathbb{P}\{\tilde{Y}_{n,p}=m\}z^n u^p v^m, \tag{17}$$

we can again translate the described combinatorial decomposition into equations for generating functions. Since a difference to the parameter ancestor-tree size can only occur, if all selected nodes are hanging on the same subtree of the root, we obtain

$$\begin{aligned} F(z,u,v) &= G(z,u,v)-\frac{zv}{(1-B(z))^2}\tilde{G}(z,u,v)+\frac{z}{(1-B(z))^2}\tilde{F}(z,u,v)-\frac{z(1-v)B(z)}{(1-B(z))^2},\\ \tilde{F}(z,u,v) &= \tilde{G}(z,u,v)-\frac{2zv}{(1-B(z))^3}\tilde{G}(z,u,v)+\frac{2z}{(1-B(z))^3}\tilde{F}(z,u,v)-\frac{2z(1-v)B(z)}{(1-B(z))^3}. \end{aligned} \tag{18}$$

Again we will use, that the butterflies are simply generated trees, and therefore the results proven in [Pan02] hold for $\tilde{F}(z,u,v)$ with $\tilde{\varphi}(t)=\frac{1}{(1-t)^2}$.

Lemma 9 (proven in [Pan02]). *For $p\geq 1$ we have with the $\tilde{\alpha}_{p,k}(z)$ as in Lemma 6 the representation*

$$N_u D_u^p \tilde{F}(z,u,v)=\sum_{k=1}^{2p-1}\frac{vz\tilde{\alpha}_{p,k}(z)}{1-z\tilde{\varphi}'\big(B(z)\big)}\left(\frac{vz}{1-vz\tilde{\varphi}'\big(B(z)\big)}\right)^{k-1}.$$

Lemma 10 (proven in [Pan02]). *Let $\tilde{\alpha}(z)$ be analytic with its only dominant singularity at $z=\rho$, where in a neighborhood of $z=\rho$ the expansion $\tilde{\alpha}(z)=\tilde{\alpha}+O(\sqrt{\rho-z})$ holds. Then we have for $n\to\infty$ and $m\leq C\sqrt{n}$ with an arbitrary but fixed $C>0$ the following asymptotic expansion*

$$[z^n]\tilde{\alpha}(z)\frac{\big(z\tilde{\varphi}'\big(B(z)\big)\big)^m}{\sqrt{\rho-z}}=\frac{\tilde{\alpha}}{\rho^{n+\frac{1}{2}}\sqrt{\pi}\sqrt{n}}e^{-\frac{\rho\tau\tilde{\varphi}''(\tau)m^2}{2n}}\left(1+O\left(\tfrac{1}{\sqrt{n}}\right)+O\left(\tfrac{m}{n}\right)\right).$$

Now we consider (18) and obtain for $p\geq 1$

$$N_u D_u^p F(z,u,v)=N_u D_u^p G(z,u,v)-\frac{zv}{(1-B(z))^2}N_u D_u^p\tilde{G}(z,u,v)+\frac{z}{(1-B(z))^2}N_u D_u^p\tilde{F}(z,u,v).$$

Considering $N_u D_u^p G(z,u,v)-\frac{zv}{(1-B(z))^2}N_u D_u^p\tilde{G}(z,u,v)$ in more detail, one gets, that the leading coefficients are annihilated, and we have for $p\geq 2$ the representation

$$N_u D_u^p G(z,u,v)-\frac{zv}{(1-B(z))^2}N_u D_u^p\tilde{G}(z,u,v)=zv\sum_{k=1}^{2p-2}\hat{\alpha}_{p,k}(z)\left(\frac{vz}{1-vz\tilde{\varphi}\big(B(z)\big)}\right)^k,$$

with local expansions $\hat{\alpha}_{p,k}(z)=\hat{\alpha}_{p,k}+O(\sqrt{\rho-z})$. This gives for a fixed $C>0$ and $m\leq C\sqrt{n}$ by using Lemma 7 the bound

$$[z^n v^m]\left(N_u D_u^p G(z,u,v)-\frac{zv}{(1-B(z))^2}N_u D_u^p\tilde{G}(z,u,v)\right)=O\left(\frac{m^{2p-2}}{n^{\frac{3}{2}}}\frac{1}{\rho^n}e^{-\frac{\rho\tau\tilde{\varphi}''(\tau)m^2}{2n}}\right),$$

which turns out to be asymptotically negligible compared to the now stated main term.

Namely, one gets with Lemma 10, that for $p \geq 2$, a fixed $C > 0$ and $m \leq C\sqrt{n}$:

$$\frac{z}{(1-B(z))^2}[z^n v^m] N_u D_u^p \tilde{F}(z,u,v) = \frac{m^{2p-3}}{\sqrt{n}} \frac{\tau\big(\tilde{\varphi}''(\tau)\tilde{\varphi}(\tau)\big)^p}{(p-2)!2^{p-2}\sqrt{\pi}\rho^{n-2p+1}\sqrt{2\rho\tau\tilde{\varphi}''(\tau)}\tilde{\varphi}(\tau)} e^{-\frac{\rho\tau\tilde{\varphi}''(\tau)m^2}{2n}} \times \\ \times \left(1 + O\left(\tfrac{1}{\sqrt{n}}\right) + O\left(\tfrac{m}{n}\right) + O\left(\tfrac{1}{m}\right)\right).$$

Together with the asymptotic expansion (3), we obtain finally for a fixed $C > 0$ and $m \leq C\sqrt{n}$:

$$\mathbb{P}\{Y_{n,p} = m\} = \frac{[z^n v^m] N_u D_u^p F(z,u,v)}{p!\binom{n}{p} T_n} = \frac{2m^{2p-3}}{n^{p-1}(p-2)!}\left(\tfrac{\rho\tau\tilde{\varphi}''(\tau)}{2}\right)^{p-1} e^{-\frac{\rho\tau\tilde{\varphi}''(\tau)m^2}{2n}} \left(1 + O\left(\tfrac{1}{\sqrt{n}}\right) + O\left(\tfrac{m}{n}\right) + O\left(\tfrac{1}{m}\right)\right).$$

Specialising for $\tilde{\varphi}(t)$, ρ and τ, we obtain the following lemma and thus Theorem 3.

Lemma 11. *For fixed $p \geq 2$, it holds for $x > 0$ and $m = x\sqrt{n} + o(\sqrt{n})$, that*

$$\sqrt{n}\,\mathbb{P}\{Y_{n,p} = m\} \sim x^{2p-3} \frac{2}{(p-2)!} \left(\frac{3}{4}\right)^{p-1} e^{-\frac{3}{4}x^2}.$$

References

[Ald91] D. Aldous. The continuum random tree ii: An overview. In *London Mathematical Society Lecture Note Series*, volume 167, pages 23–70, 1991.

[CM02] P. Chassaing and R. Marchand. Cutting a random tree (and union-find algorithms). Talk at the “8th Seminar on the Analysis of Algorithms”, Strobl, 2002.

[Fis63] M. Fisz. *Probability Theory and Mathematical Statistics*. John Wiley, 1963.

[FN99] P. Flajolet and M. Noy. Analytic combinatorics of non-crossing configurations. *Discrete Mathematics*, 204:203–229, 1999.

[FO90] P. Flajolet and A. Odlyzko. Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics*, 3:216–240, 1990.

[MM70] A. Meir and J. W. Moon. Cutting down random trees. *Journal of the Australian Mathematical Society*, 11:313–324, 1970.

[MP02] J.-F. Marckert and A. Panholzer. Noncrossing trees are almost conditioned galton-watson trees. *Random Structures and Algorithms*, 20:115–125, 2002.

[Pan02] A. Panholzer. The distribution of the size of the ancestor-tree and of the induced spanning subtree for random trees. submitted, 2002.

[Pan03] A. Panholzer. Cutting down very simple trees. submitted, 2003.