Symmetry Properties of Binary Branching Trees *

Götz Kersting
University of Frankfurt am Main †

June 1998

Abstract

In this paper we generalize the notion of a binary branching tree to that of a "uniform tree". This class contains also conditioned and size–biased branching trees. It is shown that uniform trees fulfill the following symmetry properties: i) The subtree, spanned by the root and k leaves, chosen purely at random, is a uniform tree, too. ii) If we divide a uniform tree into three pieces at a branching vertex, chosen purely at random, then the parts are uniform, too.

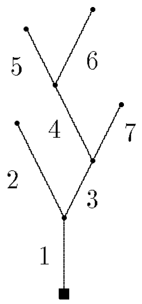
AMS classification numbers: Primary 60J80, Secondary 60J15
Key words: exponential random walks, Galton-Watson trees, random walk bridges, random walk excursions
Running title: Symmetries of Trees

1 Introduction and main result

The shape of a binary branching tree possesses properties of symmetry, which are hardly shared by other trees. In this paper we discuss two of such properties. We are dealing with random binary, ordered, rooted trees. Let us first describe this class of trees more precisely.

We consider trees T, which can be thought of as describing the whole progeny of some individual, the founding ancestor. The individuals are identified with the edges of the tree. Each edge e possesses a length L_e, which can be regarded as the lifetime of the individual. Among the vertices there is the root, to which the edge of the common ancestor is attached. Going away from the root through the tree we allow, that an edge either splits into 2 new edges or else ends in a leaf. In other words, an individual has 2 children or none. Here is an example with 7 individuals, 3 of them possessing children:

*partially supported by the German Research Foundation DFG
†Fachbereich Mathematik, Postfach 11 19 32, D–60054 Frankfurt/Main



We assume that the trees are ordered, which means that among siblings there is an older and a younger one. This allows to embed the tree into the plane (older brothers to the left, younger brothers to the right). The number of branching vertices is denoted by N. Then, as is easy to see, there are $N+1$ leaves and $2N+1$ edges. In the terminology of Aldous [2] we deal with ordered $(N+1)$–trees.

These assumptions allow to label the edges with the numbers $1, 2, \ldots,$ $2N+1$, and thus to distinguish them. We use the unique labeling, characterized by the following property: e has a smaller label than e' in the following two cases

– e is an ancestor of e' (e belongs to the path connecting e' with the root),

– e stems from the older child and e' from the younger child of their most recent common ancestor (e is located left to the line connecting e' with the root).

The above example is labeled in this manner, known as "depth–first–search".

We call T a binary branching tree (or a binary Galton–Watson tree), if the following holds:

– Any individual has two children with probability p and no child with probability $q = 1 - p$, independent of the other individuals.

– The lifetimes are identically distributed random variables with exponential distribution (expectation 1, say), independent from each other and from the splitting mechanism.

It is well–known that in case $p \leq 1/2$ such a tree T is finite with probability one. In order to explore its geometric shape, we apply different procedures to T.

P_1 : **Spanned subtrees :** Choose K leaves purely at random. Let T_K be the tree spanned by the root and these leaves. Then T_K is again a binary rooted tree. Its edges are built up from edges in T, whose lengths sum up to the edge–lengths of T_K.

P_2 : **Cut and paste :** Cut T up into three trees T_c, T_c', T_c'' at a branching vertex in T chosen purely at random. T_c contains the root of T, whereas T_c' and T_c'' obtain the chosen vertex as their root. Next join T_c, T_c' and T_c'' together to another tree T_p by attaching the roots of T_c' and T_c'' to a leaf of T_c, chosen purely at random.

We shall show that all these trees have the same shape as T. In fact we shall see that T_p is again a binary branching tree. This is not exactly true for the other trees, however, all trees belong to a class of trees, which slightly generalizes the concept of a branching tree. Briefly we consider such random trees, which coincide in distribution with a binary branching tree, if both trees are conditioned on a given length and a given number of vertices.

To define this class of "uniform trees" more precisely, we need some notation. Let

$$X_i = \sum_{j=1}^{i} L_j \ , \quad i \le 2N+1 \ ,$$

where L_j is the length of the edge with label j. Thus $0 = X_0 < X_1 < \ldots < X_{2N+1} = L$. L is called the **length** of T, and N the **size** of T. Further we consider the **topological type** $[T]$ of T. $[T]$ results in identifying all those trees, which differ only in the lengths of their edges but otherwise cannot be distinguished. $[T]$ is determined by N and by specifying, which of the individuals $1, \ldots, 2N+1$ have children and which do not. The set of topological types of all trees with n branching vertices is denoted by $\mathcal{T}_n$. It is a well–known fact that $\mathcal{T}_n$ contains $\frac{1}{n+1}\binom{2n}{n}$ different elements. Clearly T is completely determined by $[T]$ and $X_1, \ldots, X_{2N+1}$.

Definition A finite random tree T is called a **uniform tree**, if it satisfies the following conditions:

i) Conditioned on the event $\{N = n, L = l\}$, $n \in \mathbb{N}$, $l \in \mathbb{R}_+$, $[T]$ is uniformly distributed on the set $\mathcal{T}_n$.

ii) Given the event $\{N = n, L = l\}$, $(X_1, \ldots, X_{2n})$ is the ordered sample of $2n$ independent random variables, distributed uniformly on $[0, l]$.

iii) Given (N, L), $[T]$ and $(X_1, \ldots, X_{2N})$ are independent.

Thus a uniform tree T is completely determined by the distribution of (N, L). The main parameter is the distribution of the size N. The total length L of T is of secondary importance, it plays the role of a (random) scaling parameter. By rescaling the edge–lengths with a common (random) factor, it can always be achieved that $L_1, \ldots, L_{2N+1}$ become i.i.d. with exponential distribution. Here are some examples of uniform trees T:

1) Branching trees with $p \le 1/2$: Then for $[t] \in \mathcal{T}_n$, $n \in \mathbb{N}$

$$P\left(N = n, L_j \in [l_j, l_j + dl_j], j = 1, \dots, 2n+1, [T] = [t]\right)$$
$$= p^n q^{n+1} \exp(-(l_1 + \dots + l_{2n+1})) dl_1 \dots dl_{2n+1} \ ,$$

Thus

$$P(N = n, L \in [l, l+dl]) = \frac{1}{n+1} \binom{2n}{n} p^n q^{n+1} \frac{l^{2n}}{(2n)!} \ e^{-l} dl$$

and

$$P(N = n) = \frac{1}{n+1} \binom{2n}{n} p^n q^{n+1} \ .$$

2) Branching trees, conditioned on fixed size or fixed length.

3) Size–biased branching trees : These trees emerge, if one passes over from the probability measure P, belonging to branching trees, to the probability measure $\widehat{P}$, given by $d\widehat{P} = \frac{N}{EN}\, dP$. This requires $EN < \infty$, which is true, if $p < 1/2$. Thus we give new probability weights to the trees, which favour the trees according to their number of branching vertices. It follows

$$\widehat{P}(N = n) = \frac{n}{(n+1)EN} \binom{2n}{n} p^n q^{n+1} \ .$$

As we shall see later, this implies independence of the trees T_c', T_c'', T_c. — Instead one can bias according to the number of leaves; then the corresponding formula reads

$$\widehat{P}(N = n) = \frac{1}{EN+1} \binom{2n}{n} p^n q^{n+1} \ .$$

Another possibility is to bias with respect to the length, i.e. to pass over to $d\widehat{P} = \frac{L}{EL}\, dP$. A simple calculation shows that then

$$\widehat{P}(N = n) = \frac{2n+1}{EL}\ P(N = n) = \frac{2n+1}{2EN+1}\ P(N = n) \ .$$

Thus biasing according to the edge number $2N+1$ and to the length L amounts to the same thing. For the appealing properties of biased trees we refer the reader to Lyons, Pemantle, Peres [10] and Geiger [8].

$\square$

The main result of our paper now can be summarized as follows.

Theorem 1 : If the procedures P_1 or P_2 are applied to a uniform tree T, then all resulting trees $T_K, T_c, T_c', T_c'', T_p$ are uniform, too.

This will be shown in more detail in the following sections. In section 2 we analyse procedure P_1 by means of an algorithm to generate uniform trees. This kind of algorithm was introduced by Aldous [1] for a somewhat different purpose. Compare also Aldous [2] for spanned subtrees.

In section 3 we use path representations of uniform trees, which prove useful for procedure P_2. This representation, which is inspired by the papers of Le Gall [6] and Neveu, Pitman [12], relates the trees to excursions resp. bridges of exponential random walks. Also in different contexts random walks have proven to be very useful for the analysis of branching trees (compare the recent papers [4], [5], [7], [9] and the papers cited therein).

A result of similar spirit as theorem 1 is due to Neveu [11], who showed, that the tree, resulting from erasing parts of a branching tree (with general offspring distribution), remains a branching tree.

2 An algorithm to generate uniform trees

It is no loss of generality, if one fixes the length l and size n of a uniform tree. In this section we give a probabilistic construction of a uniform tree of size n and length l, which additionally to the order assigns random labels to the leaves. In the limit this construction is Aldous' cut–and–paste construction of the compact continuous random tree (see [1]). We proceed as follows:

– Take independent random variables $U_1, \dots, U_n, V_1, \dots, V_n$, uniformly distributed on the interval $[0, l]$, and $S_1, \dots, S_n$ with $P(S_j = 1) = P(S_j = -1) = 1/2$.

– Let $0 = Y_0 < Y_1 < \dots < Y_n < Y_{n+1} = l$ be the ordered sample of the random variables $\max(U_1, V_1), \dots, \max(U_n, V_n)$. If $Y_j = \max(U_k, V_k)$, let $Z_j = \min(U_k, V_k)$; thus $Z_j < Y_j$.

– Use the points $Y_j \in [0, l)$ to divide the interval $[0, l)$ into $n+1$ pieces, and the points $Z_j \in [0, l)$ to join the pieces together to a tree. Start with the segment $[0, Y_1)$ and make 0 the root. Grow the tree inductively by adding $[Y_j, Y_{j+1})$ as a new branch of length $Y_{j+1} - Y_j$, connected to the point Z_j. If $S_j = 1$, let this branch grow to the right, otherwise to the left.

In this construction each edge gets a length. The branching vertices correspond to $Z_1, \dots, Z_n$, and the the leaves to $Y_1, \dots, Y_{n+1}$. The final step of the algorithm is

– Assign the label j to the leaf, coming from Y_j.

Altogether we end up with a tree T of size n, whose leaves are labeled with the numbers $1, \dots, n+1$. It is possible to reconstruct $Y_1, \dots, Y_n, Z_1,$

$\ldots, Z_n, S_1, \ldots, S_n$ from the tree. Y_1 is the distance between the root and the leaf with label 1. Z_1 is the distance between the root and the most recent common ancestor of the leaves with labels 1 and 2. S_1 is 1 or -1, depending on which of these two leaves is left and which is right. The reader will easily complete this procedure.

The topological type of T is determined by the order of $Y_1, \ldots, Y_n$, $Z_1, \ldots, Z_n$. Since $Z_1 < Y_1$, there is 1 possibility to position Z_1 relatively to $Y_1 < \ldots < Y_n$. Given $Y_1, \ldots, Y_n, Z_1$, there are 3 possible positions for Z_2 relatively to Y_1 and Z_1, since $Z_2 < Y_2 < \ldots < Y_n$. More generally, given $Y_1, \ldots, Y_n, Z_1, \ldots, Z_{j-1}$ there are $2j-1$ possible positions of Z_j. Altogether we can arrange $Y_1, \ldots, Y_n, Z_1, \ldots, Z_n$ is $1 \cdot 3 \cdot \ldots \cdot (2n-1)$ different ways, such that $Y_1 < \ldots < Y_n$ and $Z_j < Y_j$ for all j. Since $(S_1, \ldots, S_n)$ may take 2^n different values, we end up with

$$1 \cdot 3 \cdot \ldots \cdot (2n-1) \cdot 2^n = n! \binom{2n}{n}$$

different outcomes, each corresponding to a specific topological type and labeling of the leaves. Another way to see this is to note, that there are $\frac{1}{n+1}\binom{2n}{n}$ ordered trees of size n and $(n+1)!$ possibilities to label the leaves.

Next we show that each possibility occurs with the same probability. To see this note, that for independent random variables U, V with uniform distribution in $[0, l]$ $(\max(U, V), \min(U, V))$ is uniformly distributed in $\{(y, z) : 0 \leq z \leq y \leq l\}$. Therefore $(\max(U_1, V_1), \ldots, \max(U_n, V_n)$, $\min(U_1, V_1), \ldots, \min(U_n, V_n))$ is uniformly distributed in $\{(y_1, \ldots, y_n, z_1,$ $\ldots, z_n) : 0 \leq z_j \leq y_j \leq l\}$, and $(Y_1, \ldots, Y_n, Z_1, \ldots, Z_n)$ is uniformly distributed on the set

$$S(n, l) = \{(y_1, \ldots, y_n, z_1, \ldots, z_n) : 0 \leq y_1 \leq \ldots \leq y_n \leq l, 0 \leq z_j \leq y_j\} .$$

This set is the union of the simplices $\{(c_{\pi(1)}, \ldots, c_{\pi(2n)}) :$ $0 \leq c_1 \leq \ldots \leq c_{2n} \leq l\}$, where π is a permutation of $1, 2, \ldots, 2n$ with the property $\pi(1) < \ldots < \pi(n)$, $\pi(n+j) < \pi(j)$. Each such π corresponds to a specific order of $y_1, \ldots, y_n$, $z_1, \ldots, z_n$. Since these simplices all have the same volume, our claim follows.

Finally note, that for our tree $X_1 < \ldots < X_{2n}$ are obtained by ordering $U_1, \ldots, U_n, V_1, \ldots, V_n$, and thus is the ordered sample of independent, uniformly distributed random variables. Clearly the ordered sample is independent of the relative order of $U_1, \ldots, U_n, V_1, \ldots, V_n$. Altogether we can state:

Theorem 2 : The above algorithm generates a uniform tree T of size n and length l. The labeling of the leaves is independent of T and chosen purely at random.

The theorem shows that our algorithm generates uniform trees by constructing subtrees, which in the end turn out to be spanned from leaves, chosen purely at random. In fact: $[0, Y_1)$ is the subtree, spanned by the root and the leaf with label 1, which is distributed uniformly among the leaves of T. Next $[Y_1, Y_2)$ is joined to $[0, Y_1)$ at point Z_1. The resulting tree is the subtree, spanned by the root and the leaves with labels 1 and 2, which again in the end are located purely at random among all leaves, and so forth.

This observation makes it easy to analyse subtrees of a uniform tree, spanned by the root and k random leaves. We just have to analyse the tree T_k, resulting from $Y_1, \ldots, Y_k, Z_1, \ldots, Z_{k-1}$ in our construction. It is immediate, that given Y_k (the length of T_k), then $(Y_1, \ldots, Y_{k-1}, Z_1, \ldots, Z_{k-1})$ is uniformly distributed on the set $S(k-1, Y_k)$. Thus it is obvious, that T_k is a uniform tree, too. Nothing changes, if T or T_k have random size. This proves the first half of theorem 1.

There is another way to look at our construction. $Y_1 < \ldots < Y_n$ arise by ordering the independent random variables $\max(U_1, V_1), \ldots, \max(U_n, V_n)$. It is easy to see that $\max(U, V)^2$ is uniform on $[0, l^2]$. As shown above, $(Z_1, \ldots, Z_n)$ has uniform distribution on $[0, Y_1] \times \ldots \times [0, Y_n]$, given $Y_1, \ldots, Y_n$. In other words: $Z_1/Y_1, \ldots, Z_n/Y_n$ are independent random variables with uniform distribution on $[0, 1]$. Therefore we may reformulate our construction as follows:

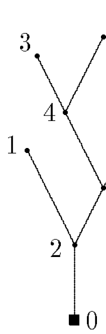
- Let $\eta_1, \ldots, \eta_n$ be independent random variables, uniform on $[0, l^2]$. Let $0 < Y_1 < \ldots < Y_n < l$ be the ordered sample of $\eta_1^{1/2}, \ldots, \eta_n^{1/2}$.
- Let $\zeta_1, \ldots, \zeta_n$ be independent random variables, independent of $\eta_1, \ldots, \eta_n$, with uniform distribution in $[0, 1]$, and let $Z_j = \zeta_j Y_j$.
- Now proceed as above (with additional random variables $S_1, \ldots, S_n$).

This version makes it easy to describe our algorithm in the limit $n \to \infty$. Let $l^2 = n$, which means, that edges have length of order $n^{-1/2}$. Then $0 < Y_1^2 < Y_2^2 < \ldots$ converges to a standard Poisson process, as $n \to \infty$. Therefore $0 < Y_1 < Y_2 < \ldots$ forms in the limit an inhomogeneous Poisson process with intensity measure $d\mu = 2y\, dy$. Thus it turns out, that in the limit we end up with Aldous' cut–and–paste construction. Clearly also the spanned subtrees T_k converge to limiting trees.

3 Path representation of trees

In order to analyse the cut–and–paste procedure P_2 we represent uniform trees by random paths. This approach goes back to Le Gall's paper [6], where random walk excursions are used for this purpose. We shall utilize "bridges" instead of "excursions", as this is done by Bennies [3] and Bennies, Kersting [4].

Now it is convenient to label the vertices with the numbers $0, 1, \dots, 2N+1$ in the following way. The $N+1$ leaves are ordered from left to right in the tree, they get the labels $1, 3, 5, \dots, 2N+1$ according to this order. The even label j is attributed to the branching vertex, which is the most recent common ancestor of the two leaves with labels $j-1$ and $j+1$. The root gets the label 0. It is easy to see that in this way each vertex receives exactly one label. Here is an example:



Next let $D_{i,j}$ be the distance (the length of the connecting path) between the vertices with labels i and j. Then the tree is completely determined by the sequence of numbers $D_{0,1}, D_{1,2}, \dots, D_{2N,2N+1}$ (in this order). This is easy to see. Build up T from left to right as follows: Draw the segment of the tree, connecting the root and the leaf with label 1; it has length $D_{0,1}$. Use $D_{1,2}$ to find the position of the vertex with label 2 on this segment. Now add the branch from vertex 2 to the leaf 3 to the right; it has length $D_{2,3}$. Use $D_{3,4}$ to locate vertex 4 on the branch from 0 to 4, and so forth. This construction shows that the length of the tree is given by

$$L = D_{0,1} + D_{2,3} + \dots + D_{2N,2N+1} \ .$$

Equally we can build up T the other way round, from right to left. Then we end up with the alternative formula

$$\begin{aligned} L &= D_{0,2N+1} + D_{2N,2N-1} + \dots + D_{2,1} \\ &= D_{1,2} + D_{3,4} + \dots + D_{2N+1,2N+2} \ , \end{aligned}$$

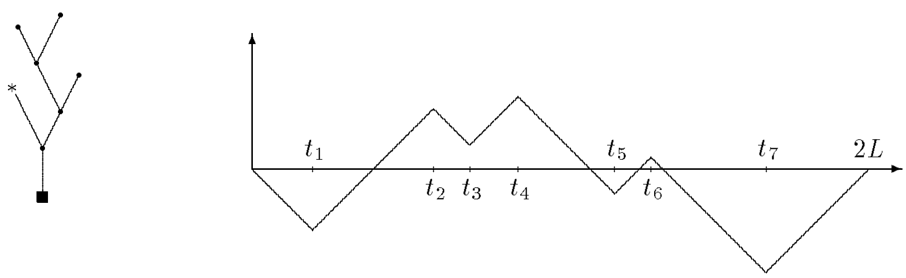
where we use the convention $D_{2N+1,2N+2} = D_{2N+1,0}$.

This can be extended to marked trees. By a **marked tree** (T, M) we mean a tree T together with the label M of some specified vertex in T. (T, M) is completely determined by the sequence $(-1)^k D_{k,k+1}$, $k = M, \ldots, M + 2N$, respectively by the numbers

$$\begin{aligned} D_{j+M,0} - D_{M,0} &= \sum_{i=0}^{j-1} (D_{i+1+M,0} - D_{i+M,0}) \\ &= \sum_{k=M}^{M+j-1} (-1)^k D_{k,k+1} \end{aligned}$$

(with $D_{2N+2+i,2N+2+i+1} = D_{i,i+1}$, $D_{2N+2+i,0} = D_{i,0}$).
Note that $D_{j+M,0} - D_{M,0}$ takes its minimum at $j = 2N + 2 - M$. Therefore we can identify M from this sequence, as well as the sequence $D_{0,1}, D_{1,2}, \ldots, D_{2N,2N+1}$ and thereby T.

It is helpful to visualize this procedure by a path, as in the picture below. It takes the value $D_{j+M,0} - D_{M,0}$ at the points $t_j = D_{M,M+1} + \ldots + D_{M+j-1,M+j}$, $j = 1, \ldots, 2N + 2$, in between it is obtained by linear interpolation. At $t_{2N+2} = 2L$ it returns to zero.



Its slope is ± 1. Local maxima and minima correspond to leaves and branching vertices (respectively), the global minimum refers to the root. If $M = 0$, then the path takes only positive values, i.e. is a positive excursion.

We want to show that for a uniform tree this path is distributed uniformly, too. To this end we describe the path by the quantities

$$\begin{aligned} Y_j &= D_{M,M+1} + D_{M+2,M+3} + \ldots + D_{M+2j-2,M+2j-1} \ , \\ Z_j &= D_{M+1,M+2} + D_{M+3,M+4} + \ldots + D_{M+2j-1,M+2j} \ , \\ S &= (-1)^M \ . \end{aligned}$$

It is easy to reconstruct the path from these numbers: $t_{2j+1} - t_{2j} = Y_{j+1} - Y_j$, $t_{2j} - t_{2j-1} = Z_j - Z_{j-1}$ and S is the slope of the path's leftmost line segment (since this slope is negative, iff M labels a leaf). Our discussion shows that $(Y_1, \ldots, Y_N, Z_1, \ldots, Z_N, S)$ together with L determines (T, M)

completely. Further $0 < Y_1 < \ldots < Y_N < L$, $0 < Z_1 < \ldots < Z_N < L$. Therefore, given that $N = n$ and $L = l$, $(Y_1, \ldots, Y_N, Z_1, \ldots, Z_N, S)$ takes values in $Si(n, l) \times Si(n, l) \times \{1, -1\}$, where

$$Si(n, l) = \{(y_1, \ldots, y_n) : 0 < y_1 < \ldots < y_n < l\} .$$

Theorem 3 : Suppose that T is a uniform tree of given size n and length l. Let M be independent of T and uniformly distributed in $\{0, 1, \ldots, 2n+1\}$. Then $(Y_1, \ldots, Y_n, Z_1, \ldots, Z_n, S)$ is uniformly distributed in $Si(n, l) \times Si(n, l) \times \{1, -1\}$.

Proof : As in section 1 we represent T by $[T]$ and $X_1, \ldots, X_{2n}$. By assumption $(M, [T], X_1, \ldots, X_{2n})$ is uniformly distributed in

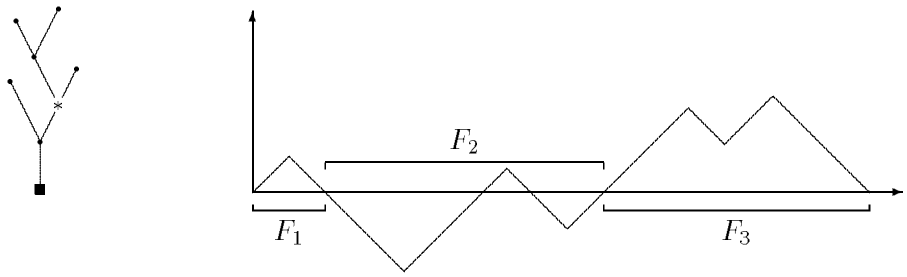
$$A_n = \{0, 1, \ldots, 2n + 1\} \times \mathcal{T}_n \times Si(2n, l) .$$

Let φ be the mapping from A_n into $Si(n, l) \times Si(n, l) \times \{1, -1\}$ with $\varphi(M, [T], X_1, \ldots, X_{2n}) = (Y_1, \ldots, Y_n, Z_1, \ldots, Z_n, S)$. This mapping is, as shown above, injective. We shall demonstrate, that it preserves the volume. Note that $Y_1, \ldots, Y_n, Z_1, \ldots, Z_n$ are obtained by adding up the edge lengths $L_i = X_i - X_{i-1}$, $i = 1, \ldots, 2n + 1$. The way, this is done, is determined by M and $[T]$. Thus given m and $[t]$, $\varphi(m, [t], x_1, \ldots, x_{2n})$ is affin–linear in $x_1, \ldots, x_{2n}$, the coefficients being integer–valued. Therefore, given m and $[t]$, the Jacobian of φ is an integer. Since φ is injectiv, the Jacobian cannot be zero. Furthermore A_n is built up from $(2n+2) \cdot \mathrm{card}\ \mathcal{T}_n = (2n+2) \cdot \frac{1}{n+1} \binom{2n}{n}$ simplices in $\mathbb{R}^{2n}$ of volume $l^{2n}/(2n)!$. This adds up to $2l^{2n}/(n!)^2$. On the other hand $Si(n, l) \times Si(n, l) \times \{1, -1\}$ consists of two subsets of $\mathbb{R}^{2n}$ with volume $l^{2n}/(n!)^2$, which amounts to $2l^{2n}/(n!)^2$, too. Therefore the Jacobian of φ, given m and $[t]$, can take the value 1 or -1 only. (This can be shown also by analyzing φ explicitly.) Consequently φ preserves the volume, and our claim follows. □

Different variants of the theorem follow as well. For example: If M is chosen uniformly from the leaves, i.e. from $\{1, 3, \ldots, 2n+1\}$, then $S = -1$ and $(Y_1, \ldots, Y_n, Z_1, \ldots, Z_n)$ is uniformly distributed in $Si(n, l) \times Si(n, l)$. Another instance: If $M = 0$, then the path is a positive excursion, which means that $Z_j < Y_j$ for all j. Now $(Y_1, \ldots, Y_n, Z_1, \ldots, Z_n)$ is uniform in

$$E(n, l) = \{(y_1, \ldots, y_n, z_1, \ldots, z_n) \in Si(n, l) \times Si(n, l) : z_j < y_j\} .$$

We use these facts to analyse procedure P_2. In the sequel we assume that M is drawn at random from the branching vertices, in other words, M is an even number, unequal to zero. For the corresponding path this means that its leftmost linear segment has slope 1 and the linear segment at the other end of the path has slope -1. Since $M \neq 0$, the minimal value of the path is strictly negative (as in the picture). Therefore we may divide the path into three fragments, which we denote by F_1, F_2, F_3.



At the left side there is a positive section. It ends, where the path starts taking negative values. This is F_1. At the right side there is a similar positive section F_3. In between it remains part F_2, which starts with slope -1 and ends with slope 1. Now a bit of reflection shows that F_1 describes the tree T_c' in just the same way as T is described by the whole path. Note that the vertices belonging to T_c' have the labels $M, M+1, \ldots, M+2N'+1$, where N' is the size of T_c'. Similarly T_c'' is built up from the vertices $M, M-1, \ldots, M-2N''-1$, and thus is described by F_3. These parts both are positive excursions, which reflects the fact, that vertex M becomes the root of T_c' and T_c''. The intermediate part F_2 represents T_c, more precisely the marked tree (T_c, M). Vertex M turns to a leaf of T_c, which corresponds to the fact, that F_2 starts with slope $S_c = -1$.

Suppose now that T is conditioned on fixed size n and fixed length l. Then $(M, [T], X_1, \ldots, X_{2n})$ is uniformly distributed in $\{2, 4, \ldots, 2n\} \times \mathcal{T}_n \times Si(2n, l)$. Thus $S = 1$ and in view of theorem 3 $(Y_1, \ldots, Y_n, Z_1, \ldots, Z_n)$ is uniformly distributed in $Si(n,l) \times Si(n,l) - E(n,l)$. (Positive paths are excluded because of $M \neq 0$). Suppose further that T_c', T_c'' and T_c are conditioned on fixed sizes n', n'' and n_c and fixed lengths l', l'' and l_c. The branching vertices of T_c', T_c'' and T_c arise from those of T, apart from vertex M. Therefore $n' + n'' + n_c = n - 1$, further $l' + l'' + l_c = l$. Now there arise additional restrictions on the values of $(Y_1, \ldots, Y_n, Z_1, \ldots, Z_n)$. F_1 is a positive excursion of length $2l'$. Therefore $Z_j < Y_j$ for $j \leq n'$, $Y_{n'+1} = l'$ and $Z_{n'+1} > l'$. The path $(Y_1', \ldots, Y_{n'}', Z_1', \ldots, Z_{n'}') = (Y_1, \ldots, Y_{n'}, Z_1, \ldots, Z_{n'})$ belonging to T_c' takes values in $E(n', l')$. Similarly T_c'' is described by the path $(Y_1'', \ldots, Y_{n''}'', Z_1'', \ldots, Z_{n''}'') = (Y_{n-n''+1} - l + l'', \ldots, Y_n - l + l'', Z_{n-n''+1} - l + l'', \ldots, Z_n - l + l'')$ with values in $E(n'', l'')$. Here the restrictions are $Z_{n-n''} = l - l'', Y_{n-n''} < l - l''$ and $Z_j < Y_j$ for $j > n - n''$. Finally the path, describing T_c is given by $(Y_1^c, \ldots, Y_{n_c}^c, Z_1^c, \ldots, Z_{n_c}^c) = (Z_{n'+1} - l', \ldots, Z_{n'+n_c} - l', Y_{n'+2} - l', \ldots, Y_{n'+n_c+1} - l')$ with values in

$Si(n_c, l_c) \times Si(n_c, l_c)$, together with a negative slope $S_c = -1$ at the beginning.

Though these considerations are somewhat involved in detail, it is clear, that by the indicated conditioning the property of uniformity remains unaffected. Note that the restrictions on the Y_i, Z_i of T_c', T_c'' and T_c do not interfere. Therefore we can state that $(Y_1', \ldots, Y_{n'}', Z_1', \ldots, Z_{n'}')$ is uniformly distributed in $E(n', l')$. This means that T_c' is a uniform tree. For T_c'' the same conclusion holds. Furthermore $(Y_1^c, \ldots, Y_{n^c}^c, Z_1^c, \ldots, Z_{n^c}^c)$ is uniform in $Si(n_c, l_c) \times Si(n_c, l_c)$. Since $S_c = -1$, this means, as shown above, that T_c is a uniform tree, too, and M is located purely at random in the leaves of T_c and independent of T_c. Therefore, if we choose another leaf from T_c purely at random and attach T_c' and T_c'' to this leaf, then the resulting tree T_p coincides with T in distribution. This proves the second part of theorem 1.

Example: Size–biased branching trees

The trees T_c', T_c'', T_c are in general dependent. There is an interesting exception: For a size–biased branching tree T_c', T_c'' and T_c are independent. To show this, we carry out some calculations.

Let (T, M) be a uniform tree of fixed size n and fixed length l, with a marked vertex M, taken at random from the branching vertices. Then, as we have seen, $(Y_1, \ldots, Y_n, Z_1, \ldots, Z_n)$ is uniform in $Si(n,l) \times Si(n,l) - E(n,l)$. It is easy to see that $\text{vol } E(n,l) = \frac{1}{n+1} (\text{vol } Si(n,l))^2$. ($E(n,l)$ represents the case $M = 0$, $Si(n,l) \times Si(n,l)$ the cases $M = 0, 2, \ldots, 2n$.) Therefore $(Y_1, \ldots, Y_n, Z_1, \ldots, Z_n)$ has the density $d_{n,l} = \frac{n+1}{n} (\text{vol } Si(n,l))^{-2} = \frac{n+1}{n} \frac{(n!)^2}{l^{2n}}$. Thus it follows from our considerations (recall $Y_{n'+1} = L', Z_{n-n''} = l - L''$)

$$
\begin{aligned}
& P\big(N' = n', N'' = n'', \\
& \qquad L' \in [l', l' + dl'], L'' \in [l'', l'' + dl''] \,\big|\, N = n, L = l\big) \\
&= d_{n,l} \cdot \text{vol } E(n', l') \cdot \text{vol } E(n'', l'') \cdot (\text{vol } Si(n_c, l_c))^2 \cdot dl' dl'' \\
&= \frac{(n+1)(n!)^2}{n(n'+1)(n''+1)(n'!)^2 (n''!)^2 (n_c!)^2} \\
& \qquad \cdot \left(\frac{l'}{l}\right)^{2n'} \left(\frac{l''}{l}\right)^{2n''} \left(\frac{l_c}{l}\right)^{2n_c} \frac{dl'}{l} \cdot \frac{dl''}{l} ,
\end{aligned}
$$

with $n_c = n - 1 - n' - n''$, $l_c = l - l' - l''$. An integration yields the formula

$$
\begin{aligned}
& P(N' = n', N'' = n'', N_c = n_c \,|\, N = n, L = l) \\
& \quad = \frac{n+1}{n} \frac{1}{n'+1} \frac{1}{n''+1} \binom{2n'}{n'} \binom{2n''}{n''} \binom{2n_c}{n_c} \Big/ \binom{2n}{n} .
\end{aligned}
$$

Let now T be size–biased, i.e. $P(N=n)$ is proportional to $\frac{n}{n+1}\binom{2n}{n}p^n q^n$ (compare the introduction). Then

$$\begin{aligned} &P(N'=n', N''=n'', N_c=n_c) \\ &\quad = \text{const}\ \frac{1}{n'+1}\binom{2n'}{n'}(pq)^{n'}\ \frac{1}{n''+1}\binom{2n''}{n''}(pq)^{n''}\binom{2n_c}{n_c}(pq)^{n_c}\ , \end{aligned}$$

which is the independence of N', N'' and N^c. This makes independence of T', T'' and T_c obvious. Further it is seen that T_c' and T_c'' are ordinary branching trees, whereas T_c is a leaf–biased branching tree. $\square$

References

[1] D.J. Aldous. *The continuum random tree I.* Ann. Prob. **19**, 1–28 (1991)

[2] D.J. Aldous. *The continuum random tree III.* Ann. Prob. **21**, 248–289 (1993)

[3] J. Bennies. *Zur Konstruktion von zufälligen Bäumen durch Irrfahrten.* Diplom thesis. Frankfurt/Main (1995)

[4] J. Bennies, G. Kersting. *A Random–Walk Approach to Galton–Watson Trees.* Preprint (1996)

[5] K. Borovkov, V. Vatutin. *On distribution tails and expectations of maxima in critical branching processes.* J. Appl.Prob. **33**, 614-622 (1996)

[6] J.F. Le Gall. *Marches aléatoires, mouvement brownien et processus du branchement.* Seminaire de Probabilités XXIII, 258–274, Lecture Notes in Math. 1372, Springer (1989)

[7] J.F. Le Gall, Y. LeJan. *Branching processes in Lévy-processes: The exploration process.* Ann. Prob. **26**, 213-252 (1998)

[8] J. Geiger. *Size–biased and conditioned random splitting trees.* Stoch. Proc. and Appl. **65**, 187–207 (1996)

[9] G. Kersting. *On the height profile of a conditioned Galton-Watson tree.* preprint (1998)

[10] R. Lyons, R. Pemantle, Y. Peres. *Conceptual proofs of $L\log L$ criteria for mean behaviour of branching processes.* Ann. Prob. **23**, 1125–1138 (1995)

[11] J. Neveu. *Erasing a Branching Tree.* Adv. Appl. Prob. **18** (Supplement), 101–108 (1986)

[12] J. Neveu, J. Pitman. *The branching process in a Brownian excursion.* Seminaire de Probabilités XXIII, 248–257, Lecture Notes in Math. 1372, Springer (1989)