SEARCH TREES: METRIC ASPECTS AND STRONG LIMIT THEOREMS

BY RUDOLF GRÜBEL

Leibniz Universität Hannover

We consider random binary trees that appear as the output of certain standard algorithms for sorting and searching if the input is random. We introduce the subtree size metric on search trees and show that the resulting metric spaces converge with probability 1. This is then used to obtain almost sure convergence for various tree functionals, together with representations of the respective limit random variables as functions of the limit tree.

1. Introduction. A sequential algorithm transforms an input sequence $t_1, t_2, \ldots$ into an output sequence $x_1, x_2, \ldots$ where, for all $n \in \mathbb{N}$, x_{n+1} depends on x_n and t_{n+1} only. Typically, the output variables are elements of some combinatorial family $\mathbb{F}$, each $x \in \mathbb{F}$ has a size parameter $\phi(x) \in \mathbb{N}$ and x_n is an element of the set $\mathbb{F}_n := \{x \in \mathbb{F} : \phi(x) = n\}$ of objects of size n. In the probabilistic analysis of such algorithms, one starts with a stochastic model for the input sequence and is interested in certain aspects of the output sequence. The standard input model assumes that the t_i's are the values of a sequence $\eta_1, \eta_2, \ldots$ of independent and identically distributed random variables. For random input of this type, the output sequence then is the path of a Markov chain $X = (X_n)_{n \in \mathbb{N}}$ that is adapted to the family $\mathbb{F}$ in the sense that

$$(1) \qquad P(X_n \in \mathbb{F}_n) = 1 \qquad \text{for all } n \in \mathbb{N}.$$

Clearly, X is highly transient—no state can be visited twice.

The special case we are interested in, and which we will use to demonstrate an approach that is generally applicable in the situation described above, is that of binary search trees and two standard algorithms, known by their acronyms BST (binary search tree) and DST (digital search tree). These are discussed in detail in the many excellent texts in this area, for example in [23, 24] and [13]. Various functionals of the search trees, such as the height [10], the path length [27, 28], the node depth profile [5–7, 14, 18, 20], the subtree size profile [9, 17], the Wiener index [25] and the silhouette [19] have been studied, with methods spanning the wide range from generatingfunctionology to martingale methods to contraction arguments on metric spaces of probability distributions (neither of these lists is

complete). Many of the results are asymptotic in nature, where the convergence obtained as $n \to \infty$ may refer to the distributions or to the random variables themselves. As far as strong limit theorems are concerned, a significant step toward a unifying approach was made in the recent paper [16], where methods from discrete potential theory were used to obtain limit results on the level of the combinatorial structures themselves: In a suitable extension of the state space $\mathbb{F}$, the random variables X_n converge almost surely as $n \to \infty$, and the limit generates the tail σ-field of the Markov chain. The results in [16] cover a wide variety of structures; search trees are a special case. It should also be mentioned here that the use of boundary theory has a venerable tradition in connection with random walks; see [22] and [29].

Our aims in the present paper are the following. First, we use the algorithmic background for a direct proof of the convergence of the BST variables X_n, as $n \to \infty$, to a limit object X_∞, and we obtain a representation of X_∞ in terms of the input sequence $(\eta_i)_{i \in \mathbb{N}}$. Second, we introduce the subtree size metric on finite binary trees. This leads to a reinterpretation of the above convergence in terms of metric trees. We also introduce a family of weighted variants of this metric, with parameter $\rho \geq 1$, and then identify the critical value ρ_0 with the property that the metric trees converge for $\rho < \rho_0$ and do not converge if $\rho > \rho_0$. The value ρ_0 turns out to also be the threshold for compactness of the limit tree. Third, we use convergence at the tree level to (re)obtain strong limit theorems for three tree functionals—the path length, the Wiener index and a metric version of the silhouette.

These topics are treated in the next three sections, where each has its own introductory remarks.

2. Binary search trees. We first introduce some notation, mostly specific to binary trees, then discuss the two search algorithms and the associated Markov chains and finally recall the results from [16] related to these structures, including an alternative proof of the main limit theorem.

2.1. *Some notation.* We write $\mathcal{L}(X)$ for the distribution of a random variable X and $\mathcal{L}(X|Y = k)$, $\mathcal{L}(X|Y)$, $\mathcal{L}(X|\mathcal{F})$ for the various versions of the conditional distribution of X given (the value of) a random variable Y or a σ-field $\mathcal{F}$. Further, δ_c is the one-point mass at c, 1_A is the indicator function of the set A [so that $1_A(c) = \delta_c(A)$], $\mathrm{Bin}(n, p)$ denotes the binomial distribution with parameters $n \in \mathbb{N}$ and $p \in (0, 1)$, $\mathrm{Beta}(\alpha, \beta)$ is the beta distribution with parameters $\alpha, \beta > 0$ and $\mathrm{unif}(0, 1) = \mathrm{Beta}(1, 1)$ is the uniform distribution on the unit interval. We also write $\mathrm{unif}(M) = (\#M)^{-1} \sum_{c \in M} \delta_c$ for the uniform distribution on a finite set M.

With $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ let

$$\mathbb{V}_k := \{0, 1\}^k, \qquad \mathbb{V} := \bigsqcup_{k \in \mathbb{N}_0} \mathbb{V}_k, \qquad \partial\mathbb{V} := \{0, 1\}^\infty$$

be the set of 0–1 sequences of length k, $k \in \mathbb{N}_0$, the set of all finite 0–1 sequences and the set of all infinite 0–1 sequences, respectively. The set $\mathbb{V}_0$ has $\varnothing$, the "empty sequence," as its only element, and $|u|$ is the length of $u \in \mathbb{V}$, that is, $|u| = k$ if $u \in \mathbb{V}_k$. For each node $u = (u_1, \ldots, u_k) \in \mathbb{V}$ we use

$$u0 := (u_1, \ldots, u_k, 0),$$

$$u1 := (u_1, \ldots, u_k, 1),$$

$$\bar{u} := (u_1, \ldots, u_{k-1}) \qquad \text{if } k \geq 1,$$

to denote its left and right direct descendant (child) and its direct ancestor (parent). We write $u \leq v$ for $u = (u_1, \ldots, u_k) \in \mathbb{V}$, $v = (v_1, \ldots, v_l) \in \mathbb{V}$ if $k \leq l$ and $u_j = v_j$ for $j = 1, \ldots, k$, that is, if u is a prefix of v; the extension to $v \in \partial\mathbb{V}$ is obvious. The prefix order is a partial order only, but there exists a unique minimum $u \wedge v$ to any two nodes $u, v \in \mathbb{V}$, their last common ancestor; again, this can be extended to elements of $\partial\mathbb{V}$. Another ordering on $\mathbb{V}$ can be obtained via the function $\beta : \mathbb{V} \to [0, 1]$,

$$(2) \qquad \beta(u) := \frac{1}{2} + \sum_{j=1}^{k} \frac{2u_j - 1}{2^{j+1}}, \qquad u \in \mathbb{V}.$$

This will be useful in various proofs, and also in connection with illustrations.

By a *binary tree* we mean a subset x of the set $\mathbb{V}$ of nodes that is prefix stable in the sense that $u \in x$ and $v \leq u$ implies that $v \in x$. Informally, we regard the components $u_1, \ldots, u_k$ of u as a routing instruction leading to the vertex u, where 0 means a move to the left, 1 a move to the right and the empty sequence is the root node. The edges of the tree x are the pairs $(\bar{u}, u)$, $u \in x, u \neq \varnothing$. A node is external to a tree if it is not one of its elements, but its direct ancestor is; we write $\partial x := \{u \in \mathbb{V} : \bar{u} \in x, u \notin x\}$ for the set of external nodes of x. Finally,

$$(3) \qquad \sigma(x, u) := \#\{v \in x : u \leq v\}$$

is the size of the subtree of x rooted at u (or the number of descendants of u in x, including u).

Let $\mathbb{B}$ denote the (countable) set of finite binary trees, $\mathbb{B}_n := \{x \in \mathbb{B} : \#x = n\}$ those of size (number of nodes) n. The single element of $\mathbb{B}_1$ is $\{\varnothing\}$, the tree that consists of the root node only.

2.2. *Search algorithms and Markov chains.* Let $(t_i)_{i \in \mathbb{N}}$ be a sequence of pairwise distinct real numbers. The BST (binary search tree) algorithm stores these sequentially into labeled binary trees (x_n, L_n), $n \in \mathbb{N}$, with $x_n \in \mathbb{B}_n$ and $L_n : x_n \to \{t_1, \ldots, t_n\}$. For $n = 1$ we have $x_1 = \{\varnothing\}$ and $L_1(\varnothing) = t_1$. Given (x_n, L_n), we construct (x_{n+1}, L_{n+1}) as follows: Starting at the root node we compare the next input value t_{n+1} to the value $L_n(u)$ attached to the node u under consideration, and move to $u0$ if $t_{n+1} < L_n(u)$ and to $u1$ otherwise, until an "empty" node u (necessarily

an external node of x_n) is found. Then $x_{n+1} := x_n \cup \{u\}$ and $L_{n+1}(u) := t_{n+1}$, $L_{n+1}(v) := L_n(v)$ for all $v \in x_n$.

Now let $(\eta_i)_{i \in \mathbb{N}}$ be a sequence of independent random variables with $\mathcal{L}(\eta_i) =$ unif$(0, 1)$ for all $i \in \mathbb{N}$, and let X_n be the random binary tree associated with the first n of these. By construction, the label functions L_n are monotone with respect to the β-order of the tree nodes, that is, with β as in (2),

$$(4) \qquad \beta(u) \le \beta(v) \Rightarrow L_n(u) \le L_n(v) \qquad \text{for all } n \text{ with } \{u, v\} \subset X_n.$$

In particular, if we number the external nodes of X_n from the left to the right, then the number of the node that receives η_{n+1} is the rank of this value among $\{\eta_1, \dots, \eta_n\}$, hence uniformly distributed on $\{1, \dots, n + 1\}$. This shows that the (deterministic) BST algorithm, when applied to the (random) input $(\eta_i)_{i \in \mathbb{N}}$, results in a Markov chain $(X_n)_{n \in \mathbb{N}}$ with state space $\mathbb{B}$, start at $X_1 \equiv \{\varnothing\}$ and transition probabilities

$$(5) \qquad Q(x, x \cup \{u\}) = \begin{cases} 1/(1 + \#x), & \text{if } u \in \partial x, \\ 0, & \text{otherwise.} \end{cases}$$

In words: We obtain X_{n+1} by choosing one of the $n + 1$ external nodes of X_n uniformly at random and joining it to the tree. We refer to this construction as the *BST chain*.

For the DST (digital search tree) algorithm, the input values are infinite 0–1 sequences, that is, elements of $\partial \mathbb{V}$. Given $t_1, t_2, \dots \in \partial \mathbb{V}$ we again obtain a sequence $x_1, x_2, \dots$ of labeled binary trees, but now we use the components $t_{n+1,k}$, $k \in \mathbb{N}$, of the next input value t_{n+1} as a routing instruction through x_n, moving to $u0$ from an occupied node $u \in \mathbb{V}_k$ if $t_{n+1,k+1} = 0$ and to $u1$ otherwise. As in the BST case we assume that the t_i's are the values of a sequence of independent and identically distributed random variables η_i, where the distribution of the η_i's is now a probability measure μ on the measurable space $(\partial \mathbb{V}, \mathcal{B}(\partial \mathbb{V}))$, with $\mathcal{B}(\partial \mathbb{V})$ the σ-field generated by the projections on the sequence elements, $\partial \mathbb{V} \ni t = (t_k)_{k \in \mathbb{N}} \mapsto t_i$, $i \in \mathbb{N}$. This σ-field is also generated by the sets

$$(6) \qquad A_u := \{v \in \partial \mathbb{V} : v \ge u\}, \qquad u \in \mathbb{V}.$$

It is easy to check that the intersection of two such sets is either empty or again of this form. This implies that μ is completely specified by its values $\mu(A_u)$, $u \in \mathbb{V}$, and the DST analogue of (5) then is

$$(7) \qquad Q(x, x \cup \{u\}) = \begin{cases} \mu(A_u), & \text{if } u \in \partial x, \\ 0, & \text{otherwise.} \end{cases}$$

By the *DST chain with driving distribution* μ we mean a Markov chain $(X_n)_{n \in \mathbb{N}}$ with state space $\mathbb{B}$, start at $\{\varnothing\}$ and transition mechanism given by (7).

2.3. *Doob–Martin compactification.* We refer the reader to Doob's seminal paper [12] and to the recent textbook [30] for the main results of, background on and further references for the boundary theory for transient Markov chains. For the BST chain the Doob–Martin compactification has recently been obtained in [16]: It can be described as the closure $\bar{\mathbb{B}}$ of the embedding of $\mathbb{B}$ into the compact space $[0, 1]^{\mathbb{V}}$, endowed with pointwise convergence, that is given by the standardized *subtree size functional*

$$\mathbb{B} \ni x \mapsto \left(\mathbb{V} \ni u \mapsto \frac{\sigma(x, u)}{\#x} \right)$$

with σ as defined in (3). Further, the elements of the boundary $\partial\mathbb{B}$ may be represented by probability measures μ on $(\partial\mathbb{V}, \mathcal{B}(\partial\mathbb{V}))$, with convergence $x_n \to \mu$ of a sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathbb{B}$ meaning that

$$\mu(A_u) = \lim_{n\to\infty} \frac{\sigma(x_n, u)}{\#x_n} \qquad \text{for all } u \in \mathbb{V},$$

and $\mu_n(A_u) \to \mu(A_u)$ for all $u \in \mathbb{V}$ if we have a sequence $(\mu_n)_{n\in\mathbb{N}}$ of elements of $\partial\mathbb{B}$ instead.

The general theory implies that X_n converges almost surely to a limit X_∞ with values in $\partial\mathbb{B}$; [16] also contains a description of $\mathcal{L}(X_\infty)$. The proof given there does not make use of the algorithmic background, but takes the transition mechanism (5) as its starting point. We now show that this background leads to a direct proof of $X_n \to X_\infty$, and to a representation of X_∞ in terms of the input sequence.

We need some more notation. On $\mathbb{V}$ we define a metric $d_\mathbb{V}$ by

$$(8) \qquad d_\mathbb{V}(u, v) := 2^{-|u \wedge v|} - \tfrac{1}{2}(2^{-|u|} + 2^{-|v|}), \qquad u, v \in \mathbb{V}.$$

On $\mathbb{V}$ itself this gives the discrete topology, and the completion of $\mathbb{V}$ with respect to $d_\mathbb{V}$ leads to $\bar{\mathbb{V}} := \mathbb{V} \cup \partial\mathbb{V}$, a compact and separable metric space. This is also the ends compactification if we regard $\mathbb{V}$ as the complete rooted binary tree. We extend the A_u's to $\bar{\mathbb{V}}$ by

$$\bar{A}_u := \{v \in \bar{\mathbb{V}} : v \geq u\}, \qquad u \in \mathbb{V}.$$

Because of

$$\bar{A}_u := \{v \in \bar{\mathbb{V}} : d_\mathbb{V}(u, v) < 2^{-|u|}\} = \{v \in \bar{\mathbb{V}} : d_\mathbb{V}(u, v) \leq 2^{-|u|-1}\}$$

these sets are open and closed. Further,

$$\{u\} = \bar{A}_u \setminus (\bar{A}_{u0} \cup \bar{A}_{u1}), \qquad \bar{A}_u \cap \bar{A}_v = \begin{cases} \bar{A}_u, & \text{if } u \leq v, \\ \bar{A}_v, & \text{if } u \geq v, \\ \varnothing, & \text{otherwise,} \end{cases}$$

hence $\{\bar{A}_u : u \in \mathbb{V}\}$ is a π-system that generates $\mathcal{B}(\bar{\mathbb{V}})$. Together these facts imply that weak convergence of probability measures μ_n to a probability measure μ on $(\bar{\mathbb{V}}, \mathcal{B}(\bar{\mathbb{V}}))$ is equivalent to

$$(9) \qquad \lim_{n\to\infty} \mu_n(\bar{A}_u) = \mu(\bar{A}_u) \qquad \text{for all } u \in \mathbb{V}.$$

In view of

$$\frac{1}{n}\sigma(X_n, u) = \text{unif}(X_n)(\bar{A}_u)$$

and $X_\infty(\mathbb{V}) = 0$ convergence in the Doob–Martin topology is therefore equivalent to the weak convergence of probability measures on the metric space $(\bar{\mathbb{V}}, d_{\mathbb{V}})$ if we represent finite subsets M of $\mathbb{V}$ by the uniform distribution $\text{unif}(M)$ on $(\bar{\mathbb{V}}, \mathcal{B}(\bar{\mathbb{V}}))$.

Moreover, any sequence $(\mu_n)_{n \in \mathbb{N}}$ of probability measures on $(\bar{\mathbb{V}}, \mathcal{B}(\bar{\mathbb{V}}))$ is tight, as $\bar{\mathbb{V}}$ is compact, and therefore has a limit point by Prohorov's theorem [2], page 37. If $(\mu_n(\bar{A}_u))_{n \in \mathbb{N}}$ is a convergent sequence for each $u \in \mathbb{V}$, then there is only one such limit point, which means that μ_n converges weakly to some probability measure μ and that (9) holds. Finally, let

$$(10) \qquad \tau(u) := \inf\{n \in \mathbb{N} : X_n \ni u\}, \qquad u \in \mathbb{V},$$

be the time that the node u becomes an element of the BST sequence. It is easy to see that the $\tau(u)$'s are finite with probability 1.

THEOREM 1. *Let $(X_n)_{n \in \mathbb{N}}$ be the sequence of binary trees generated by the BST algorithm with input a sequence $(\eta_i)_{i \in \mathbb{N}}$ of independent and identically distributed random variables with $\mathcal{L}(\eta_1) = \text{unif}(0, 1)$.*

(a) *With probability 1 the sequence $\text{unif}(X_n)$ converges weakly to a random probability measure X_∞ on $(\partial\mathbb{V}, \mathcal{B}(\partial\mathbb{V}))$ as $n \to \infty$.*

(b) *For each $u \in \mathbb{V}$, $u \neq \varnothing$, with $i := \tau(u) - 1$, τ as in (10), and*

$$0 =: \eta_{(i:0)} < \eta_{(i:1)} < \cdots < \eta_{(i:i)} < \eta_{(i:i+1)} := 1$$

the augmented order statistics associated with $\eta_1, \ldots, \eta_i$, we have

$$X_\infty(A_u) = \eta_{(i:j+1)} - \eta_{(i:j)} \qquad \text{with } \eta_{(i:j)} < \eta_{i+1} < \eta_{(i:j+1)}.$$

(c) *The random variables*

$$\xi_u := \frac{X_\infty(A_{u0})}{X_\infty(A_u)}, \qquad u \in \mathbb{V},$$

are independent, and $\mathcal{L}(\xi_u) = \text{unif}(0, 1)$ for all $u \in \mathbb{V}$.

PROOF. Let u, $\tau(u)$, i and $\eta_{(i:j)}$, $j = 0, \ldots, i+1$, be as in part (b) of the theorem. The order property (4) of the labeled binary search trees implies that for a node v with label η_k, $k > i$, the relation $v \geq u$ is equivalent to $\eta_{(i:j)} < \eta_k < \eta_{(i:j+1)}$. Hence, by the law of large numbers,

$$\begin{aligned}
\lim_{n\to\infty} \text{unif}(X_n)(\bar{A}_u) &= \lim_{n\to\infty} \frac{\#\{v \in X_n : v \geq u\}}{n} \\
&= \lim_{n\to\infty} \frac{\#\{i < k \leq n : \eta_k \in (\eta_{(i:j)}, \eta_{(i:j+1)})\}}{n} \\
&= \eta_{(i:j+1)} - \eta_{(i:j)}
\end{aligned}$$

with probability 1 for every $u \in \mathbb{V}$. In view of

$$\{u\} = \{v \in \bar{\mathbb{V}} : d(u, v) < 2^{-|u|-1}\} \qquad \text{for all } u \in \mathbb{V},$$

the one-point sets with elements from $\mathbb{V}$ are open in the topology on $\bar{\mathbb{V}}$. As $\text{unif}(X_n)$ assigns at most the value $1/n$ to such a set, it follows with the portmanteau theorem [2], page 11, that any limit point of this sequence is concentrated on $\partial \mathbb{V}$. Parts (a) and (b) of the theorem now follow with the above general remarks on weak convergence of probability measures on $(\bar{\mathbb{V}}, \mathcal{B}(\bar{\mathbb{V}}))$.

For the proof of (c) we use the following well-known fact: The conditional distribution of η_{i+1}, given $\eta_1, \ldots, \eta_i$ and given that the value lands in an interval $I = (\eta_{(i:j)}, \eta_{(i:j+1)})$ of the augmented order statistics, is the uniform distribution on I, which implies that $\text{unif}(0, 1)$ is the distribution of the normalized distance ξ_u to the left endpoint of I. For different η-values these relative insertion positions are independent, hence ξ_u, $u \in \mathbb{V}$, are independent and uniformly distributed on the unit interval. $\square$

We note the following consequence of the representation in part (c) of the theorem: For a fixed $u \in \mathbb{V}$ let

$$\varnothing = u(0) < u(1) < \cdots < u(k) = u$$

with $|u(j)| = j$ for $j = 0, \ldots, k$ be the path that connects u to the root node. We then have

$$X_\infty(A_u) = \prod_{j=0}^{k-1} \tilde{\xi}_{u(j)}$$

(11)

$$\text{with } \tilde{\xi}_{u(j)} := \begin{cases} \xi_{u(j)}, & \text{if } u(j+1) = u(j)0, \\ 1 - \xi_{u(j)}, & \text{if } u(j+1) = u(j)1. \end{cases}$$

Note that the factors $\tilde{\xi}_{u(j)}$, $j = 0, \ldots, k-1$, are independent and that they all have distribution $\text{unif}(0, 1)$.

Theorem 1 confirms the view expressed in [30], pages 191 and 218, that in specific cases embeddings (or boundaries) can generally be obtained directly on using the then available additional structure; here this turns out to be the algorithmic representation of the Markov chain. However, there are two additional benefits of the general theory: First, because of the space–time property (1) the limit X_∞ generates the tail σ-field

$$\mathcal{T} := \bigcap_{n=1}^{\infty} \sigma(\{X_m : m \geq n\})$$

associated with the sequence $(X_n)_{n \in \mathbb{N}}$. This may serve as a starting point for the unification of strong limit theorems for functionals $(Y_n)_{n \in \mathbb{N}}$, $Y_n = \Psi(X_n)$ of the discrete structures: If Y_n converges to Y_∞ in a "reasonable" space, then the

limit Y_∞, which is $\mathcal{T}$-measurable, must be a function of X_∞; see, for example, [21], Lemma 1.13. The second general result is extremely useful in the context of the calculations that arise in specific applications of the theory: The conditional distribution of the chain $(X_n)_{n \in \mathbb{N}}$ given the value of X_∞ is again a Markov chain, where the new transition probabilities can be obtained from the limit value and the old transition probabilities by a procedure that is known as Doob's h-transform. In the present situation it turns out that the conditional distribution of the BST chain, given $X_\infty = \mu$, is the same as that of the DST chain driven by μ. We refer the reader to [16] for details; the last statement appears there only for a specific μ, but the generalization to an arbitrary probability measure μ in the boundary is straightforward. Roughly, the embedded jump chains at the individual nodes are Pólya urns; for these the boundary has been obtained in [3], and from the general construction of the Doob–Martin boundary it is clear that the outcome is unaffected by the step from a Markov chain to its embedded jump chain. We collect some consequences in the following proposition, where

$$(12) \qquad \mathcal{F}_n := \sigma(X_1, \ldots, X_n), \qquad n \in \mathbb{N},$$

are the elements of the natural filtration of the BST chain.

PROPOSITION 2.   *With the notation and assumptions as in Theorem* 1,

$$(13) \qquad \mathcal{L}\big(\sigma(X_n, u0)|\sigma(X_n, u) = k, \xi_u = p\big) = \mathrm{Bin}(k - 1, p) \qquad if\ k > 0,$$

and, for all $i, j \in \mathbb{N}_0$,

$$(14) \qquad \mathcal{L}\big(\xi_u|\sigma(X_n, u0) = i, \sigma(X_n, u1) = j\big) = \mathrm{Beta}(i + 1, j + 1).$$

Further, the variables $(\xi_u)_{u \in \mathbb{V}}$ *are conditionally independent given* $\mathcal{F}_n$.

3. Metric aspects.   All trees in this paper are subgraphs of the complete binary tree, which has $\mathbb{V}$ as its set of nodes and $\{(\bar{u}, u) : u \neq \varnothing\}$ as its set of edges; in particular, our trees are specified by their node sets x. In a tree metric d the distance of any two nodes u, v is the sum of the distances between successive nodes on the unique path from u to v, which means that such a metric is given by its values $d(\bar{u}, u)$, $u \in x$, $u \neq \varnothing$. For example, the metric $d_\mathbb{V}$ in Section 2.3 has $d_\mathbb{V}(\bar{u}, u) = 2^{-|u|-1}$, and the canonical tree distance d_{can} is given by $d_{\mathrm{can}}(\bar{u}, u) = 1$. For our trees the prefix order further leads to

$$(15) \qquad d(u, v) = d(u, \varnothing) + d(v, \varnothing) - 2d(u \wedge v, \varnothing) \qquad for\ all\ u, v \in x.$$

Metric trees may also be interpreted as graphs with edge weight, where the edge $(\bar{u}, u)$ receives the weight $d(\bar{u}, u)$.

Our aim in this section is to rephrase the convergence of the BST sequence as a convergence of metric trees, and to show that this view leads to convergence with respect to stronger topologies. The situation here is much simpler than for Aldous's

continuum random tree where the Gromov–Hausdorff convergence of equivalence classes of metric trees is used; see [15] and the references given there. In fact, the search trees considered here have node sets that grow monotonically to the full $\mathbb{V}$, so we may define convergence of a sequence $((x_n, d_n))_{n \in \mathbb{N}}$ of metric binary trees to $(\mathbb{V}, d_\infty)$ to mean that

$$(16) \qquad \lim_{n \to \infty} d_n(u, v) = d_\infty(u, v) \qquad \text{for all } u, v \in \mathbb{V},$$

which of course is equivalent to $\lim_{n \to \infty} d_n(\bar{u}, u) = d_\infty(\bar{u}, u)$ for all $u \in \mathbb{V}, u \neq \varnothing$. Note that $d_{\mathbb{V}}$ and d_{can} are both local metrics in the sense that $d(u, v)$ does not depend on the tree x as long as $u, v \in x$.

Motivated by the view in Section 2.3 of finite and infinite binary trees as probability measures μ on $(\bar{\mathbb{V}}, \mathcal{B}(\bar{\mathbb{V}}))$, we now introduce the (relative) *subtree size metric*, which assigns $\mu(\bar{A}_u)$ to the distance of $\bar{u}$ and u, that is,

$$d_x(\bar{u}, u) = \frac{\sigma(x, u)}{\sigma(x, \varnothing)} \qquad \text{for all } u \in x, u \neq \varnothing,$$

if $x \in \mathbb{B}$, and

$$d_\mu(\bar{u}, u) = \mu(A_u) \qquad \text{for all } u \in \mathbb{V}, u \neq \varnothing$$

for the complete tree and a probability measure μ on $(\partial \mathbb{V}, \mathcal{B}(\partial \mathbb{V}))$, where we assume that $\mu(A_u) > 0$ for all $u \in \mathbb{V}$. Again, there is an algorithmic motivation: In terms of the BST mechanism, the weight of an edge $(\bar{u}, u)$ is the (relative) number of times this edge has been traversed in the construction of the tree. These metrics depend on their tree in a global manner.

With this terminology in place we may now rephrase the convergence in Theorem 1 as the convergence in the sense of (16) of the finite metric trees (X_n, d_{X_n}) to the infinite metric tree $(\mathbb{V}, d_{X_\infty})$, almost surely and as $n \to \infty$.

By construction the Doob–Martin compactification is the weakest topology that allows for a continuous extension of the functions $\mathbb{B} \ni x \mapsto \sigma(x, u)/\sigma(x, \varnothing)$, $u \in \mathbb{V}$. For the analysis of tree functionals stronger modes of convergence turn out to be useful; for example, do we have uniform convergence in (16)? Also, subtree sizes decrease along paths leading away from the root node, so we may consider a weight factor for the distance of a node to its parent that depends on the depth of the node: For all $\rho \geq 1$, we define the *weighted subtree size metric* with *weight parameter ρ* by

$$d_{x, \rho}(\bar{u}, u) := \rho^{|u|} d_x(\bar{u}, u), \qquad d_{\mu, \rho}(\bar{u}, u) := \rho^{|u|} d_\mu(\bar{u}, u),$$

in the finite and infinite case, respectively. Of course, with $\rho = 1$ the subtree size metric reappears.

THEOREM 3. *Let $\rho_0 = 1.26107 \cdots$ be the smaller of the two roots of the equation $2e \log(\rho) = \rho$, $\rho > 0$. Let X_n, $n \in \mathbb{N}$ and X_∞ be as in Theorem 1.*

(a) *For $\rho < \rho_0$, the metric space $(\mathbb{V}, d_{X_\infty, \rho})$ is compact with probability* 1.

(b) *For $\rho > \rho_0$, the metric space $(\mathbb{V}, d_{X_\infty, \rho})$ has infinite diameter with probability* 1.

(c) *For $\rho < \rho_0$, the metric spaces $(X_n, d_{X_n, \rho})$ converge uniformly to $(\mathbb{V}, d_{X_\infty, \rho})$ as $n \to \infty$ in the sense of*

$$(17) \quad \sup_{u, v \in X_n} \left| d_{X_n, \rho}(u, v) - d_{X_\infty, \rho}(u, v) \right| \to 0 \qquad \textit{almost surely and in mean.}$$

(d) *For $\rho > \rho_0$, and with $d_{X_n, \rho}(\bar{u}, u) := 0$ for $u \notin X_n$,*

$$\sup_{u, v \in \mathbb{V}} \left| d_{X_n, \rho}(u, v) - d_{X_\infty, \rho}(u, v) \right| = \infty \qquad \textit{with probability} \ 1.$$

PROOF.   We embed the metric trees into the linear space $\mathbb{L}(0)$ of all functions $f : \mathbb{V} \setminus \{\varnothing\} \to \mathbb{R}$ via

$$x \mapsto f := \big(u \mapsto d_x(\bar{u}, u) \big), \qquad x \in \mathbb{B};$$

probability measures μ on $(\bar{\mathbb{V}}, \mathcal{B}(\bar{\mathbb{V}}))$ become elements of $\mathbb{L}(0)$ by identifying μ with the function $u \mapsto \mu(A_u)$. In particular, we now write $X_\infty(u)$ instead of $X_\infty(A_u)$. For $\rho \geq 1$ let $\mathbb{L}(\rho)$ be the set of all $f \in \mathbb{L}(0)$ with

$$\| f \|_\rho := \sum_{k=1}^{\infty} \rho^k \max_{|u|=k} \big| f(u) \big| < \infty.$$

Clearly, this gives a family of nested separable Banach spaces, with

$$\mathbb{B} \hookrightarrow \mathbb{L}(\gamma) \subset \mathbb{L}(\rho) \subset \mathbb{L}(0) \qquad \text{for } 1 \leq \rho < \gamma.$$

We now show that, with the above identification,

$$(18) \qquad\qquad\qquad E \| X_\infty \|_\rho < \infty \qquad \text{if } \rho < \rho_0,$$

$$(19) \qquad\qquad P\Big(\sup_{u \in \mathbb{V}} \rho^{|u|} X_\infty(u) = \infty \Big) = 1 \qquad \text{if } \rho > \rho_0$$

and that, for $\rho < \rho_0$ and as $n \to \infty$,

$$(20) \qquad\qquad \| X_n - X_\infty \|_\rho \to 0 \qquad \text{almost surely and in mean.}$$

Clearly, (18) implies that $X_\infty \in \mathbb{L}(\rho)$ with probability 1 if $\rho < \rho_0$.

The basis for our proof of (18) and (19) is the connection of BST trees to branching random walks, a connection that has previously been used by several authors, especially for the analysis of the height of search trees; see the survey [11] and the references given there. Let $u(k, j)$, $j = 1, \ldots, 2^k$, be a numbering of the nodes from $\mathbb{V}_k$ such that

$$\beta\big(u(k, 1) \big) < \beta\big(u(k, 2) \big) < \cdots < \beta\big(u(k, 2^k) \big),$$

with β as defined in (2). The key observation is that the variables

$$Y_{k,j} := -\log X_\infty(u(k,j)), \qquad j = 1, \ldots, 2^k,$$

are the positions of the members of the kth generation in a branching random walk with offspring distribution δ_2 and with

$$Z := \delta_{-\log \xi} + \delta_{-\log(1-\xi)}, \qquad \mathcal{L}(\xi) = \mathrm{unif}(0,1)$$

for the point process of the positions of the children relative to their parent. Biggins [1] obtained several general results for such processes that we now specialize to the present offspring distribution and point process of relative positions. Let

$$m(\theta) := E\left(\int e^{-\theta t} Z(dt)\right) = \frac{2}{1+\theta}$$

and

$$(21) \qquad \tilde{m}(a) := \inf\{e^{\theta a} m(\theta) : \theta \geq 0\} = 2a e^{1-a}.$$

Note that

$$(22) \qquad \tilde{m}(a) = m(\theta(a)) \qquad \text{with } \theta(a) = \frac{1}{a} - 1,$$

and that, by definition of ρ_0,

$$(23) \qquad \rho < \rho_0 \quad \Longleftrightarrow \quad \tilde{m}(\log \rho) < 1.$$

Finally, let $Z^{(k)}(t)$ be the number of particles in generation k that are located to the left of t.

Now suppose that $\rho < \rho_0$. Let $\alpha := (\rho + \rho_0)/2$ and $\eta := \log(\alpha)$. We adapt the upper bound argument in [1] to our present needs: For all $\theta > 0$ and $C > 1$, with $\gamma := \log(C)$,

$$P\left(\alpha^k \max_{|u|=k} X_\infty(u) > C\right) = P\left(\min_{1 \leq j \leq 2^k} Y_{k,j} \leq k\eta - \gamma\right)$$

$$\leq E Z^{(k)}\left(k\left(\eta - \frac{\gamma}{k}\right)\right)$$

$$\leq \exp\left(k\left(\eta - \frac{\gamma}{k}\right)\theta\right) m(\theta)^k$$

$$= C^{-\theta}\left(e^{\eta\theta} m(\theta)\right)^k.$$

By (23), $\tilde{m}(\eta) < 1$. Choosing the optimal $\theta = \theta(\eta)$, which with (22) is easily seen to be greater than 1, leads to

$$E\left(\alpha^k \max_{|u|=k} X_\infty(u)\right) \leq 1 + \int_1^\infty P\left(\alpha^k \max_{|u|=k} X_\infty(u) > x\right) dx$$

$$\leq 1 + \tilde{m}(\eta)^k \int_1^\infty x^{-\theta(\eta)} dx \leq c,$$

with a finite constant c that does not depend on k. Hence

$$\sum_{k=1}^{\infty} \rho^k E\Big(\max_{|u|=k} X_\infty(u)\Big) \le c \sum_{k=1}^{\infty} \Big(\frac{\rho}{\alpha}\Big)^k < \infty,$$

which in turn implies (18) by monotone convergence.

Suppose now that $\rho > \rho_0$, so that $\tilde{m}(\eta) > 1$ by (23) for $\eta := \log \rho$. By [1], Theorem 2,

$$\lim_{k \to \infty} \frac{1}{k} \log\big(\#\{1 \le j \le 2^k : Y_{k,j} \le k\eta\}\big) = \log \tilde{m}(\eta) > 0$$

with probability 1. In particular, and again with probability 1,

$$\exists k_0 \; \forall k \ge k_0 \; \exists u \in \mathbb{V}_k \qquad -\log X_\infty(u) \le k \log \rho.$$

Clearly, this implies (19).

For the proof of (20) we first consider the random variables $\sigma(X_n, u)$, $n \in \mathbb{N}$, for some fixed $u \in \mathbb{V}$. We wish to relate these to $E[X_\infty(u)|\mathcal{F}_n]$, with $\mathcal{F}_n$ as in (12). For this, we use the representation of X_∞ in terms of $(\xi_u)_{u \in \mathbb{V}}$ given in Section 2.3, together with Proposition 2. We may assume that $k := |u| > 0$.

The representation (11), the conditional independence of the $\tilde{\xi}$-variables given $\mathcal{F}_n$, and the well-known formula for the first moment of beta distributions together lead to

$$E[X_\infty(u)|\mathcal{F}_n] = \prod_{j=0}^{k-1} E[\tilde{\xi}_{u(j)}|\mathcal{F}_n] = \prod_{j=0}^{k-1} \frac{\sigma(X_n, u(j+1)) + 1}{\sigma(X_n, u(j)0) + \sigma(X_n, u(j)1) + 2}.$$

In view of

$$\sigma(x, u0) + \sigma(x, u1) + 1 = \begin{cases} \sigma(x, u), & \text{if } u \in x, \\ 1, & \text{if } u \notin x, \end{cases}$$

the product telescopes to

$$(24) \qquad E[X_\infty(u)|\mathcal{F}_n] = \frac{\sigma(X_n, u) + 1}{n + 1} \qquad \text{for all } u \in X_n.$$

We now introduce

$$Z_n : \mathbb{V} \to \mathbb{R}, \qquad u \mapsto E[X_\infty(u)|\mathcal{F}_n].$$

Then $(Z_n, \mathcal{F}_n)_{n \in \mathbb{N}}$ is a vector-valued martingale. For $\rho < \rho_0$ we have by part (a) of the theorem that $X_\infty \in \mathbb{L}(\rho)$ with probability 1 and that $E\|X_\infty\|_\rho < \infty$, hence $Z_n \to X_\infty$ almost surely and in mean in $\mathbb{L}(\rho)$ by Proposition V-2-6 in [26].

In our present representation of trees as functions on $\mathbb{V}$ we have

$$X_n(u) = \begin{cases} \dfrac{n+1}{n} Z_n(u) - \dfrac{1}{n}, & \text{if } u \in X_n, \\ 0, & \text{if } u \notin X_n, \end{cases}$$

which implies that $0 \leq X_n \leq (1 + n^{-1})Z_n$ for all $n \in \mathbb{N}$. As $X_n \to X_\infty$ pointwise with probability 1 by Theorem 1 we can now use a suitable version of the dominated convergence theorem, such as that given in [21], Theorem 1.21, to obtain that X_n converges to X_∞ in $\mathbb{L}(\rho)$ as $n \to \infty$, again almost surely and in mean.

It remains to show that the tree statements in the theorem follow from the linear space statements (18), (19) and (20).

For (a) we prove that the limiting metric space is totally bounded. From (18) and the definition of the norm we obtain for any given $\varepsilon > 0$ a $k = k(\varepsilon) \in \mathbb{N}$ such that

$$\sum_{j=k}^{\infty} \rho^j \max_{|u|=j} X_\infty(u) < \varepsilon,$$

which by the definition of the weighted subtree size metric means that all nodes v with $|v| \geq k$ have a distance from their predecessor at level k that is less than ε. As there are only finitely many nodes of level less than k this shows that the whole of $\bar{\mathbb{V}}$ may be covered by a finite number of ε-balls. Of course, this argument is meant to be applied to each element of a suitable set of probability 1 separately.

For (b) we simply note that (19) implies that, with probability 1,

$$\sup_{u \in \mathbb{V}, u \neq \varnothing} d_{X_\infty, \rho}(\bar{u}, u) = \infty$$

if $\rho > \rho_0$. This also gives (d).

Finally, for all $u \in \mathbb{V}$, $u \neq \varnothing$,

$$\left| d_{X_n, \rho}(u, \varnothing) - d_{X_\infty, \rho}(u, \varnothing) \right| \leq \sum_{\varnothing \neq v \leq u} \left| d_{X_n, \rho}(\bar{v}, v) - d_{X_\infty, \rho}(\bar{v}, v) \right|$$

$$\leq \sum_{k=1}^{|u|} \rho^k \max_{|v|=k} \left| X_n(v) - X_\infty(v) \right|$$

$$\leq \| X_n - X_\infty \|_\rho.$$

The upper bound does not depend on u, hence (c) follows on using (15). $\quad\square$

We note that the convergence of metric trees considered in Theorem 3 implies the convergence with respect to the Gromov–Hausdorff distance of the corresponding equivalence classes of metric trees; see [4], Section 7.3.3.

The subtree size metric also leads to a visualization of search trees: We use the function β defined in (2) to map nodes to points in the unit interval, and above the x-coordinate $\beta(u)$ we draw a line parallel to the y-axis from $d_{X_n}(\bar{u}, \varnothing)$ to $d_{X_n}(u, \varnothing)$. In order to obtain a visually more pleasing result we may add lines that run parallel to the x-axis, connecting nodes with the same parent. In Figure 1 we have carried this out for the trees arising from two separate input sequences for the BST algorithm, with the data obtained from alternating blocks of
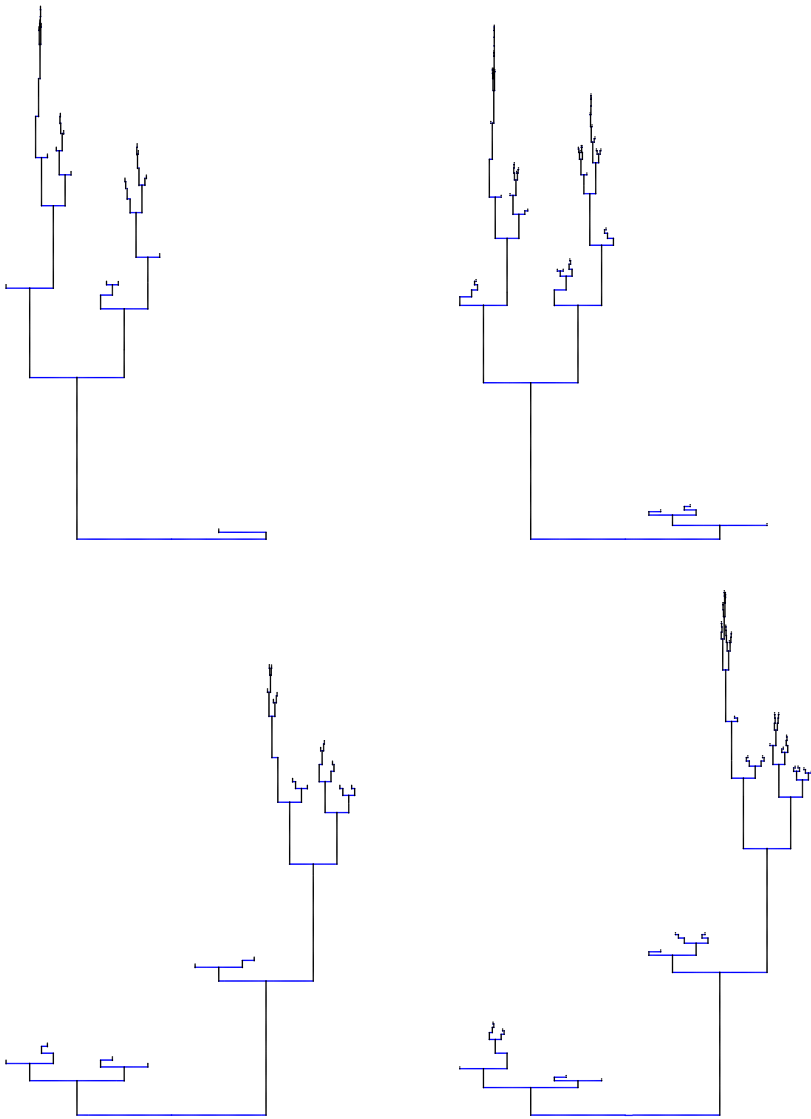
FIG. 1.   *The metric tree for the odd* (*upper part*) *and even* (*lower part*) π-*data, for n* = 50 (*left*) *and* *n* = 100 (*right*), *respectively*; *see text for details.*

length 10 of digits in the decimal expansion of $\pi - 3$. The upper part refers to the odd and the lower to the even numbered blocks. In both cases we have given the trees for $n = 50$ and $n = 100$, and with $\rho = 1$. Vertically, the trees are from the same distribution; moving horizontally to the right, we have almost sure convergence.

4. Tree functionals. In this section we show how the above results can be used in connection with the asymptotic analysis of tree functionals. Here is the recipe: We start with a functional $Y_n = \Psi_n(X_n)$ of the trees, with (deterministic) functions Ψ_n on $\mathbb{B}_n$ that have values in some separable Banach space $(\mathbb{L}, \|\cdot\|)$. We suspect that Y_n converges almost surely to some limit variable Y_∞ as $n \to \infty$. We know that if this is the case, then $Y_\infty = \Psi(X_\infty)$ for some Ψ defined on $\partial\mathbb{B}$ (as always, almost surely). We do not know what Ψ is, but if we manage to rewrite the Ψ_n's in terms of subtree sizes, then Theorem 1 may lead to an educated guess. On that basis we next consider $\Phi_n(X_n) = E[\Psi(X_\infty)|\mathcal{F}_n]$, assuming that $E\|\Psi(X_\infty)\| < \infty$. This gives an $\mathbb{L}$-valued martingale. By the associated convergence theorem we then have that $\tilde{Y}_n := \Phi_n(X_n)$ converges to Y_∞ almost surely and in mean. Finally, a simple inspection of $\Phi_n - \Psi_n$ may reveal that $\tilde{Y}_n - Y_n$ is asymptotically negligible—indeed, if Y_n converges to Y_∞, then $\tilde{Y}_n - Y_n$ *must* tend to 0.

In the first three subsections we work out the details of the above strategy for path lengths, for a tree index and for an infinite dimensional tree functional. The final subsection is a collection of remarks on other functionals and related tree structures, indicating further applications of the method, but also its limitations. The potential-theoretic approach can provide additional insight; for example, we will relate a martingale introduced in connection with tree profiles to Doob's h-transform.

Throughout this section we abbreviate $X_\infty(A_u)$ to $X_\infty(u)$.

4.1. *Path length.* The first tree functional we consider is the *internal path length*,

$$\text{IPL}(x) := \sum_{u \in x} |u|, \qquad x \in \mathbb{B}, \tag{25}$$

which may be rewritten as

$$\text{IPL}(x) = \sum_{u \in x, u \neq \varnothing} \sigma(x, u) = \sum_{u \in x} \sigma(x, u) - \#x. \tag{26}$$

Let

$$H(0) := 0, \qquad H(n) := \sum_{i=1}^{n} \frac{1}{i} \qquad \text{for all } n \in \mathbb{N},$$

be the harmonic numbers. It is well known that

$$\lim_{n \to \infty} (H(n) - \log n) = \gamma,$$

where $\gamma \approx 0.57722$ is Euler's constant. We need two auxiliary statements; we omit the (easy) proofs.

LEMMA 4. *For all $i, j \in \mathbb{N}_0$,*

$$\frac{\Gamma(i+j+2)}{\Gamma(i+1)\Gamma(j+1)} \int_0^1 x^i (1-x)^j \log(x) \, dx = H(i) - H(i+j+1).$$

For a random variable η with distribution Beta$(i+1, j+1)$ Lemma 4 leads to

$$(27) \qquad E(\eta \log(\eta)) = \frac{i+1}{i+j+2}(H(i+1) - H(i+j+2)).$$

The next lemma is a summation by parts formula for binary trees.

LEMMA 5. *For any function $\psi : \mathbb{V} \to \mathbb{R}$,*

$$\sum_{u \in x} \big(\psi(u) - \psi(u0) - \psi(u1)\big) = \psi(\varnothing) - \sum_{u \in \partial x} \psi(u) \qquad \text{for all } x \in \mathbb{B}.$$

Major parts of the following theorem are known; we will give details later in order to be able to refer to the proof for a comparison of the methods used. Let $(X_n)_{n \in \mathbb{N}}$ be the BST chain, and let X_∞ be its limit, as in Theorem 3.

THEOREM 6. *Let $C : (0, 1) \to \mathbb{R}$ be defined by*

$$C(s) := 1 + 2\big(s \log(s) + (1-s) \log(1-s)\big).$$

(a) *The limit*

$$Y_\infty := \sum_{u \in \mathbb{V}} X_\infty(u) C\left(\frac{X_\infty(u0)}{X_\infty(u)}\right)$$

exists almost surely and in quadratic mean.

(b) *As $n \to \infty$,*

$$(28) \qquad \frac{1}{n} \operatorname{IPL}(X_n) - 2 \log n \to 2\gamma - 4 + Y_\infty,$$

almost surely and in quadratic mean.

PROOF. From the representation of X_∞ given in Section 2.3, we know that the random variables

$$\xi_u := \frac{X_\infty(u0)}{X_\infty(u)}, \qquad u \in \mathbb{V},$$

are independent and uniformly distributed on the unit interval, and that $X_\infty(u)$ is a function of the ξ_v's with $v < u$. In particular, for all nodes u, the two factors in the sum appearing in the definition of Y_∞ are independent. Let $\mathcal{G}_k$ be the σ-field generated by the ξ_u's with $|u| \le k$, and put

$$Y_k := \sum_{u \in \mathbb{V}, |u| \le k} X_\infty(u) C\left(\frac{X_\infty(u0)}{X_\infty(u)}\right), \qquad k \in \mathbb{N}.$$

Then these properties lead to

$$E[Y_{k+1}|\mathcal{G}_k] = Y_k + E\left[\sum_{|u|=k+1} X_\infty(u)C\left(\frac{X_\infty(u0)}{X_\infty(u)}\right)\Big|\mathcal{G}_k\right]$$

$$= Y_k + \sum_{|u|=k+1} X_\infty(u)EC(\xi_u)$$

$$= Y_k,$$

where we have used the fact that $EC(\xi_u) = 0$. Further, with the same arguments,

$$E[(Y_{k+1} - Y_k)^2|\mathcal{G}_k] = E\left[\left(\sum_{|u|=k+1} X_\infty(u)C(\xi_u)\right)^2\Big|\mathcal{G}_k\right]$$

$$= \sum_{|u|=k+1} X_\infty(u)^2 EC(\xi_u)^2,$$

so that

$$E(Y_{k+1} - Y_k)^2 = \sum_{|u|=k+1} EX_\infty(u)^2 EC(\xi_u)^2.$$

We also have $\kappa := EC(\xi_u)^2 < \infty$, and using (11) we get

$$EX_\infty(u)^2 = \left(E\xi_\varnothing^2\right)^k = 3^{-k},$$

so that

(29) $$E(Y_{k+1} - Y_k)^2 = 2^k 3^{-k} \kappa \qquad \text{for all } k \in \mathbb{N}.$$

Taken together these calculations show that $(Y_k, \mathcal{G}_k)_{k \in \mathbb{N}}$ is an L^2-bounded martingale, and an appeal to the corresponding martingale limit theorem completes the proof of (a). In particular, Y_∞ is well defined, and even has finite second moment.

For the proof of (b) let $Z_n := E[Y_\infty|\mathcal{F}_n]$, $n \in \mathbb{N}$, so that $(Z_n, \mathcal{F}_n)_{n \in \mathbb{N}}$ is again a martingale bounded in L^2. Our plan is to show that Z_n is sufficiently close to the transformed internal path length that appears in (28).

Using again the stochastic structure of X_∞ we are thus led to consider the conditional expectations $E[X_\infty(u)|\mathcal{F}_n]$ and $E[C(\xi_u)|\mathcal{F}_n]$, $u \in \mathbb{V}$ and $n \in \mathbb{N}$. From Proposition 2 we know that, for all $u \in X_n$,

$$\mathcal{L}(\xi_u|\mathcal{F}_n) = \text{Beta}(\sigma(X_n, u0) + 1, \sigma(X_n, u1) + 1),$$

and that the ξ_u's are conditionally independent given $\mathcal{F}_n$. Hence Lemma 4 can be applied [see also (27)], resulting in

(30)
$$E[C(\xi_u)|\mathcal{F}_n] = 1 + \frac{2\tau(X_n, u0) + 2\tau(X_n, u1)}{\sigma(X_n, u0) + \sigma(X_n, u1) + 2}$$
$$- 2H(\sigma(X_n, u0) + \sigma(X_n, u1) + 2),$$

where the function $\tau : \mathbb{B} \times \mathbb{V} \to \mathbb{R}$ is given by

$$\tau(x, u) := (\sigma(x, u) + 1)H(\sigma(x, u) + 1).$$

For each fixed $n \in \mathbb{N}$, almost sure convergence of $E[Y_k|\mathcal{F}_n]$ to $E[Y_\infty|\mathcal{F}_n]$ as $k \to \infty$ follows from

$$\big\|E[Y_k|\mathcal{F}_n] - E[Y_\infty|\mathcal{F}_n]\big\|_2 \leq \|Y_k - Y_\infty\|_2,$$

the upper bound in (29) and the Borel–Cantelli lemma. Together with the conditional independence of $X_\infty(u)$ and $C(\xi_u)$ given $\mathcal{F}_n$, this leads to

$$(31) \qquad Z_n = \sum_{u \in \mathbb{V}} E\big[X_\infty(u)|\mathcal{F}_n\big]E\big[C(\xi_u)|\mathcal{F}_n\big].$$

From (30) we obtain $E[C(\xi_u)|\mathcal{F}_n] = 0$ for $u \notin X_n$, and, clearly,

$$(32) \qquad \sigma(X_n, u0) + \sigma(X_n, u1) + 1 = \sigma(X_n, u) \qquad \text{for all } u \in X_n.$$

Taken together, (24), (30), (31) and (32) lead to

$$Z_n = \sum_{u \in X_n} \frac{\sigma(X_n, u) + 1}{n + 1}\left(1 + \frac{2\tau(X_n, u0) + 2\tau(X_n, u1)}{\sigma(X_n, u) + 1} - 2H\big(\sigma(X_n, u) + 1\big)\right),$$

which in turn gives

$$Z_n = \frac{1}{n + 1}\big(\text{IPL}(X_n) + 2n\big) - \frac{2}{n + 1}\sum_{u \in X_n}\big(\tau(X_n, u) - \tau(X_n, u0) - \tau(X_n, u1)\big).$$

Lemma 5 can be applied to the second sum, and the assertion finally follows from $\tau(X_n, \varnothing) = (n + 1)H(n + 1)$ and $\tau(X_n, u) = 1$ for $u \in \partial X_n$. $\quad\square$

Almost sure convergence of the standardized internal path length for the BST sequence has been obtained in [27], and convergence in distribution, together with a fixed point relation for the limit distribution, in [28]. Our method may been seen as an amalgamation of Régnier's martingale approach and Rösler's approach, where the latter has come to be known as the contraction method in the analysis of algorithms: We obtain a strong limit, but we do not need to "find the martingale" (a task familiar to many an applied probabilist). The approach suggested in the present paper, to look at convergence of the full objects via a suitable completion of the state space of the underlying combinatorial Markov chain, leads to a representation of the almost sure limit. This gives the martingale by projection via conditional expectations, and from the representation one can also read off a fixed point relation for the distribution of the limit.

4.2. *The Wiener index.* The canonical graph distance $d_{\text{can}}(u, v)$ of any two nodes u and v in a finite connected graph G with node set V is the minimum length of a path (sequence of edges) that connects u and v in G. The sum of these distances is the *Wiener index* of the graph,

$$\text{(33)} \qquad \text{WI}(G) := \frac{1}{2} \sum_{(u,v) \in V \times V} d_{\text{can}}(u, v),$$

introduced by the chemist H. Wiener. Some background together with pointers to the literature is given in [25], which is also our main reference in this subsection. Among other results it is shown in [25] that for the BST sequence $(X_n)_{n \in \mathbb{N}}$ the rescaled Wiener indices,

$$W_n := \frac{1}{n^2} \text{WI}(X_n) - 2 \log n,$$

converge in distribution as $n \to \infty$.

Again, we project a suitable functional $\Psi(X_\infty)$ of the limit tree X_∞ to a function $E[\Psi(X_\infty)|\mathcal{F}_n]$ of X_n that is sufficiently close to W_n. This will give a strong limit theorem, that is, it turns out that the rescaled Wiener indices in fact converge almost surely for the random binary trees generated by the BST algorithm for i.i.d. input, and it will also lead to a representation of the limit W_∞ as a function of X_∞.

We begin by rewriting the Wiener index in terms of subtree sizes, similar to the transition from (25) to (26) in the analysis of the internal path length. For a binary tree x,

$$\text{(34)} \qquad \sum_{(u,v) \in x \times x} |u \wedge v| = \sum_{u \in x} \sigma(x, u)^2.$$

This may be proved by induction, using the left and right subtrees in the induction step; see [8], page 70. Using (15), (26), (33) and (34) we now obtain

$$\text{(35)} \qquad \text{WI}(X_n) = n \, \text{IPL}(X_n) + n^2 - \sum_{u \in X_n} \sigma(X_n, u)^2.$$

It is a benefit of working with almost sure convergence that we can deal with the constituents on the right-hand side of (35) separately (which means that we can make use of Theorem 6), whereas in connection with convergence in distribution one needs to consider the joint distribution of $\text{IPL}(X_n)$ and $\text{WI}(X_n)$; see [25].

THEOREM 7. *The series*

$$\text{(36)} \qquad Z_\infty := \sum_{u \in \mathbb{V}} X_\infty(u)^2$$

converges almost surely and in quadratic mean, and, as $n \to \infty$,

$$\text{(37)} \qquad \frac{1}{n^2} \text{WI}(X_n) - 2 \log n \to W_\infty$$

again almost surely and in quadratic mean, where the limit is given by

$$(38) \qquad\qquad W_\infty := 2\gamma - 3 + Y_\infty - Z_\infty,$$

with Y_∞ as in Theorem 6.

PROOF.   Almost sure convergence in (36) follows with Theorem 3, and the moment calculations below show that $EZ_\infty^2 < \infty$. In particular,

$$Z_n := E[Z_\infty | \mathcal{F}_n] \to Z_\infty$$

almost surely and in quadratic mean. Again, the Markov property implies that Z_n can be written as a function of X_n. In order to obtain this function we first consider a fixed node $u \in \mathbb{V}$.

From (11) we get

$$X_\infty(u)^2 = \prod_{j=0}^{k-1} \tilde{\xi}_{u(j)}^2.$$

From (14) and the known formula for the second moment of beta distributions we obtain, considering the cases $u(j+1) = u(j)0$ and $u(j+1) = u(j)1$ separately,

$$E\big[\tilde{\xi}_{u(j)}^2 | \mathcal{F}_n\big]$$

$$= \frac{(\sigma(X_n, u(j+1)) + 1)(\sigma(X_n, u(j+1)) + 2)}{(\sigma(X_n, u(j)0) + \sigma(X_n, u(j)1) + 2)(\sigma(X_n, u(j)0) + \sigma(X_n, u(j)1) + 3)}.$$

Using the conditional independence statement in Proposition 2, we see that we have a telescoping product again, so that

$$E\big[X_\infty(u)^2 | \mathcal{F}_n\big] = \frac{(\sigma(X_n, u) + 1)(\sigma(X_n, u) + 2)}{(n+1)(n+2)} \qquad \text{for all } u \in X_n \cup \partial X_n.$$

The set $\mathbb{V} \setminus X_n$ can be written as the disjoint union of the subtrees rooted at the $n+1$ external nodes of X_n, and we have

$$E\big[\xi_u^2 | \mathcal{F}_n\big] = E\big[(1 - \xi_u)^2 | \mathcal{F}_n\big] = \tfrac{1}{3} \qquad \text{for all } u \notin X_n.$$

Therefore,

$$\sum_{u \notin X_n} E\big[X_\infty(u)^2 | \mathcal{F}_n\big]$$

$$= \frac{1}{(n+1)(n+2)} \sum_{u \in \partial X_n} (\sigma(X_n, u) + 1)(\sigma(X_n, u) + 2) \sum_{v \in \mathbb{V}, v \geq u} \left(\frac{1}{3}\right)^{|v|-|u|}$$

$$= \frac{2}{(n+1)(n+2)} \sum_{u \in \partial X_n} \sum_{k=0}^{\infty} 2^k \left(\frac{1}{3}\right)^k$$

$$= \frac{6}{n+2}.$$

in view of $\sigma(X_n, u) = 0$ for $u \in \partial X_n$. Taken together this gives

$$Z_n = \frac{1}{(n+1)(n+2)} \sum_{u \in X_n} (\sigma(X_n, u) + 1)(\sigma(X_n, u) + 2) + \frac{6}{n+2}.$$

Using (26) we get

$$\sum_{u \in X_n} (\sigma(X_n, u) + 1)(\sigma(X_n, u) + 2) = \sum_{u \in X_n} \sigma(X_n, u)^2 + 3 \cdot \mathrm{IPL}(X_n) + 5n$$

so that, with (35),

$$\frac{1}{(n+1)(n+2)} \mathrm{WI}(X_n) = \frac{n}{(n+1)(n+2)} \mathrm{IPL}(X_n) + \frac{n^2}{(n+1)(n+2)} - Z_n + R_n,$$

where R_n tends to 0 almost surely and in quadratic mean. From Theorem 6 we know that

$$\frac{1}{n} \mathrm{IPL}(X_n) - 2 \log n \to 2\gamma - 4 + Y_\infty$$

in the same sense. Combining the last two statements we obtain (37), with W_∞ as in (38). □

4.3. *Metric silhouette.* In our third application we consider an infinite-dimensional tree functional.

Each element $v = (v_k)_{k \in \mathbb{N}}$ of $\partial \mathbb{V}$ defines a path through a binary tree via the sequence $(v(k))_{k \in \mathbb{N}}$ of nodes given by $v(k) = (v_1, \ldots, v_k)$, $k \in \mathbb{N}$. In [19] the "silhouette" $\mathrm{Sil}(x)$ of $x \in \mathbb{B}$ was introduced in an attempt to obtain a search tree analogue of the famous Harris encoding of simply generated trees: with each path v, we record its exit level when passing through x, that is,

$$\mathrm{Sil}(x)(v) := \min\{k \in \mathbb{N} : v(k) \notin x\}, \qquad v \in \partial \mathbb{V}.$$

The tree silhouette can be visualized as a function on the unit interval via the binary expansion

$$(39) \quad \Phi : [0, 1) \to \partial \mathbb{V}, \qquad t \mapsto (v_k)_{k \in \mathbb{N}} \qquad \text{with } v_k := \lceil 2^{k+1} t \rceil - 2 \lceil 2^k t \rceil.$$

It was shown in [19] that for the BST chain $(X_n)_{n \in \mathbb{N}}$ some smoothing is necessary to obtain an interesting limit for the stochastic processes $(\mathrm{Sil}(X_n)(\Phi(t)))_{0 \le t < 1}$ as $n \to \infty$.

We have seen in the previous sections that for search trees it makes sense to replace the canonical tree distance implicit in the above definition of $\mathrm{Sil}(x)$ by the subtree size metric. A corresponding variant of the silhouette is the *metric silhouette*,

$$\mathrm{mSil}(x)(v) := \sum_{k=1}^{\infty} \sigma(x, v(k)), \qquad v \in \partial \mathbb{V}.$$

Again, our aim is to obtain a strong limit theorem in the BST situation, together with a representation of the limit as a function of X_∞. In addition, and going beyond the individual arguments $v \in \partial\mathbb{V}$, we regard $\mathrm{mSil}(X_n)$ as a random function on $\partial\mathbb{V}$. With $d_\mathbb{V}$ as in (8) this is a compact and separable metric space ($\mathbb{V}$ is open in the completion $\bar{\mathbb{V}}$ that we introduced in Section 2.3). We write $C(\partial\mathbb{V}, d_\mathbb{V})$ for the space of continuous functions $f : \partial\mathbb{V} \to \mathbb{R}$. Together with

$$\|f\|_\infty := \sup_{v \in \partial\mathbb{V}} |f(v)|,$$

this is a separable Banach space.

Remember that the values of X_∞ are probability measures on $(\partial\mathbb{V}, \mathcal{B}(\partial\mathbb{V}))$. Let $\Sigma_\infty : \partial\mathbb{V} \to [0, \infty]$ be defined by

$$\Sigma_\infty(v) := - \int_{\partial\mathbb{V}} \log_2(d_\mathbb{V}(u, v)) X_\infty(du), \qquad v \in \partial\mathbb{V}.$$

This is the logarithmic potential of the random measure X_∞ with respect to $d_\mathbb{V}$; see [29], page 62. Finally, we recall that a real function f on the metric space $(\partial\mathbb{V}, d_\mathbb{V})$ is said to be (globally) Hölder continuous with exponent α if there exists a constant $C < \infty$ such that

$$|f(u) - f(v)| \leq C d_\mathbb{V}(u, v)^\alpha \qquad \text{for all } u, v \in \partial\mathbb{V}.$$

THEOREM 8.   *Let* $\alpha_0 := \log_2 \rho_0 = 0.33464\ldots$ *with* ρ_0 *as in Theorem* 3.

(a) $E\|\Sigma_\infty\|_\infty < \infty$.

(b) *With probability* 1, Σ_∞ *is Hölder continuous with exponent* α *for all* $\alpha < \alpha_0$.

(c) *As* $n \to \infty$,

$$\left\|\frac{1}{n} \mathrm{mSil}(X_n) - \Sigma_\infty\right\|_\infty \to 0 \qquad \text{almost surely and in mean.}$$

PROOF.   Because of $d_\mathbb{V}(u, v) = 2^{-|u \wedge v|}$ for all $u, v \in \partial\mathbb{V}$ we have $-\log_2 d_\mathbb{V}(u, v) \in \mathbb{N}_0$ and

$$-\log_2 d_\mathbb{V}(u, v) \geq k \quad \Longleftrightarrow \quad u \in A_{v(k)}$$

for all $k \in \mathbb{N}$ so that

$$(40) \qquad \Sigma_\infty(v) = \sum_{k=1}^\infty X_\infty(v(k)) \qquad \text{for all } v \in \partial\mathbb{V}.$$

Now let α be as in the statement of the theorem; we may assume that $\alpha > 0$. Let $\rho := 2^\alpha$. By Theorem 3 there exists a set of probability 1 such that for all ω in this set, $C(\omega) := \|X_\infty(\omega)\|_\rho < \infty$. We fix such an ω and drop it from

the notation. Because of $X_\infty(u) \leq C\rho^{|u|}$ for all $u \in \mathbb{V}$ and (40), we then have $\Sigma_\infty(v) \leq C \sum_{k=1}^{\infty} \rho^{-k}$ for all $v \in \partial\mathbb{V}$, which implies

$$\|\Sigma_\infty\|_\infty \leq \frac{1}{\rho - 1}\|X_\infty\|_\rho.$$

In particular, $E\|\Sigma_\infty\|_\infty < \infty$ by (18) in the proof of Theorem 3.

Similarly, if $u, v \in \partial\mathbb{V}$ are such that $|u \wedge v| = k$, then

$$\left|\Sigma_\infty(u) - \Sigma_\infty(v)\right| = \sum_{j=k+1}^{\infty} X_\infty(u(j)) + \sum_{j=k+1}^{\infty} X_\infty(v(j))$$

$$\leq 2C \sum_{j=k+1}^{\infty} \rho^{-j} = \frac{2C\rho^{-k}}{\rho - 1} \leq \frac{2C}{\rho - 1}d_\mathbb{V}(u, v)^\alpha$$

by definition of ρ. This proves (b).

For the proof of (c) we first consider the random functions Σ_n defined by

$$\Sigma_n(v) := E\left[\Sigma_\infty(v)|\mathcal{F}_n\right], \qquad v \in \partial\mathbb{V}.$$

With $J := \mathrm{Sil}(X_n)(v)$ we get, using monotone convergence for conditional expectations and $\mathcal{F}_n$-measurability of J,

$$\Sigma_n(v) = \sum_{k=1}^{J} E\left[X_\infty(v(k))|\mathcal{F}_n\right] + \sum_{k=J+1}^{\infty} E\left[X_\infty(v(k))|\mathcal{F}_n\right]$$

$$= \sum_{k=1}^{J} \frac{\sigma(X_n, v(k)) + 1}{n + 1} + \frac{\sigma(X_n, v(J)) + 1}{n + 1} \sum_{k=J+1}^{\infty} \left(\frac{1}{2}\right)^{k-J}$$

$$= \frac{1}{n + 1}\left(\mathrm{mSil}(X_n)(v) + J\right) + \frac{1}{n + 1}.$$

Here we have used our formula (24) for $E[\Sigma_\infty(u)|\mathcal{F}_n]$ and its extension to nodes outside X_n that can be obtained as in the proof of Theorem 7.

Let $h(x) = \max\{|u| : u \in x\}$ be the height of $x \in \mathbb{B}$. Taking the supremum over $v \in \partial\mathbb{V}$ we get

$$\left\|\Sigma_n - \frac{1}{n + 1}\,\mathrm{mSil}(X_n)\right\|_\infty \leq \frac{h(X_n) + 1}{n + 1}.$$

It is easy to show that the right-hand side converges to 0 with probability 1 (see [10] for techniques and results on the height), so it remains to prove that Σ_n converges almost surely and in mean to Σ_∞ in the separable Banach space $(C(\partial\mathbb{V}, d_\mathbb{V}), \|\cdot\|_\infty)$. This, however, is again immediate from the vector-valued martingale convergence theorem given in [26], page 104. $\square$

Figure 2 shows the metric silhouette for the trees in Figure 1. Note that the continuity in Theorem 8 refers to the space $(\partial\mathbb{V}, d_\mathbb{V})$; for example, $(t_n)_{n \in \mathbb{N}}$ with
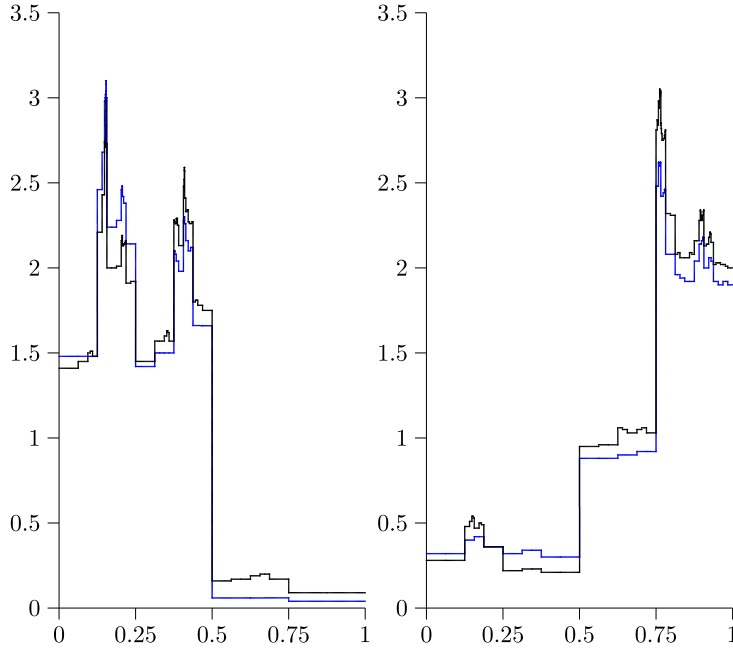
FIG. 2.   *The metric silhouette for the odd* (left) *and even* (right) π-*data, with* $n = 50$ (blue) *and* $n = 100$ (black).

$t_n = \frac{1}{2} + (-1)^n \frac{1}{n}$ for all $n \in \mathbb{N}$ is a Cauchy sequence with respect to euclidean distance, but its inverse under the function Φ defined in (39) that we used for the illustration is not a Cauchy sequence in $(\partial \mathbb{V}, d_{\mathbb{V}})$. Loosely speaking, the function β "flattens" the node set $\mathbb{V}$.

4.4. *Other functionals and tree structures.*   The fill (or saturation) level $F(x)$ and height $H(x)$ of a tree $x \in \mathbb{B}$ are defined by

$$F(x) = \max\{k \in \mathbb{N}_0 : \{0, 1\}^k \subset x\}, \qquad H(x) = \max\{|u| : u \in x\},$$

respectively. For these tree functionals, the following asymptotic results are well known:

$$(41) \qquad \frac{F(X_n)}{\log n} \to \alpha_-, \qquad \frac{H(X_n)}{\log n} \to \alpha_+ \qquad \text{as } n \to \infty,$$

both almost surely. Here $\alpha_- = 0.373\ldots$ and $\alpha_+ = 4.311\ldots$ are the two solutions of the equation $x \log(2e/x) = 1$. The survey [11] gives details and references, and explains the relation to branching processes.

In situations such as these, where the almost sure limit is a constant, projection on the sub-σ-fields $\mathcal{F}_n$ would simply return the constant, hence no simplification arises.

Both the fill level and height of a tree as well as its path length (see Section 4.1) can be written as functionals of the tree's node profile. Recall that $|u|$ denotes the length of $u \in \mathbb{V}$. Let

$$v(x, k) := \#\{u \in \partial x : |u| = k\}, \qquad w(x, k) := \#\{u \in x : |u| = k\}$$

be the number of external (resp., internal) nodes of $x \in \mathbb{B}$ at depth k. Applied to the BST sequence $(X_n)_{n \in \mathbb{N}}$, this gives sequences $(V_n)_{n \in \mathbb{N}}$ and $(W_n)_{n \in \mathbb{N}}$ of random functions on the nonnegative integers via $V_n(k) = v(X_n, k)$ and $W_n(k) = w(X_n, k)$, the external and internal node profile of the binary search tree. Clearly,

$$F(X_n) = \min\{k \in \mathbb{N} : V_n(k) > 0\} - 1,$$
$$H(X_n) = \max\{k \in \mathbb{N} : W_n(k) > 0\}$$

and

$$\mathrm{IPL}(X_n) = \sum_{k=1}^{\infty} k W_n(k),$$

so such profiles go some way toward a unifying approach to tree functionals and indeed, they have been studied extensively; see [5–7, 14, 18, 20]. A crucial role in [5–7, 20] is played by a parametrized family of martingales introduced in [20]. In order to connect this to the point of view of the present paper we rephrase the basic idea using our terminology and notation.

Fix some $z \in \mathbb{R}_+$. The external profile V_n of X_n can be regarded as the counting density of a random finite measure on $\mathbb{N}$ with total mass $n + 1$ and value

$$Y_n := \sum_{k=1}^{\infty} V_n(k) z^k = \sum_{u \in \partial X_n} z^{|u|}$$

at z of its generating function. Let v be the random node that is added to X_n to obtain X_{n+1}. It is easy to see that

$$Y_{n+1} = Y_n + z^{|v|}(2z - 1).$$

In the BST mechanism the node v is chosen uniformly at random from the $n + 1$ external nodes of X_n, hence

$$E[Y_{n+1}|\mathcal{F}_n] = E\left[\sum_{v \in \partial X_n} \left(Y_n + z^{|v|}(2z - 1)\right) 1_{\{X_{n+1} = X_n \cup \{v\}\}} \Big| \mathcal{F}_n\right]$$

$$= Y_n E\left[\sum_{v \in \partial X_n} 1_{\{X_{n+1} = X_n \cup \{v\}\}} \Big| \mathcal{F}_n\right]$$

$$+ (2z - 1) E\left[\sum_{v \in \partial X_n} z^{|v|} 1_{\{X_{n+1} = X_n \cup \{v\}\}} \Big| \mathcal{F}_n\right]$$

$$= Y_n + \frac{2z-1}{n+1} \sum_{v \in \partial X_n} z^{|v|}$$

$$= \frac{n+2z}{n+1} Y_n,$$

which means that $(M_n, \mathcal{F}_n)_{n \in \mathbb{N}}$ with

$$M_n := C(n)Y_n, \qquad C(n) := \prod_{k=1}^{n-1} \frac{k+1}{k+2z} \text{ for all } n \in \mathbb{N},$$

is a martingale. Obviously, the martingale is strictly positive whenever $z > 0$. Because of the space–time property it can therefore be written as $M_n = h(X_n)$ with some positive harmonic function h on $\mathbb{B}$, which depends on $z > 0$, and which in the present context is given by

$$h(x) = C(\#x) \sum_{u \in \partial x} z^{|u|}.$$

Moreover, the distribution P^h of the corresponding h-transform, which is the Markov chain with transition probabilities

$$p^h(x, y) = \frac{1}{h(x)} p(x, y)h(y), \qquad x, y \in \mathbb{B},$$

is such that for all $n \in \mathbb{N}$ the restriction $P^h_{\mathcal{F}_n}$ of P^h to $\mathcal{F}_n$ has density

$$\frac{dP^h_{\mathcal{F}_n}}{dP_{\mathcal{F}_n}} = \frac{1}{2z} h(X_n)$$

with respect to the restriction $P_{\mathcal{F}_n}$ to $\mathcal{F}_n$ of the distribution P of the original BST chain. Here we have used that both chains start with the tree $X_1 = \{\varnothing\}$, and that $h(\{\varnothing\}) = 2z$. A straightforward calculation yields

$$(42) \qquad p^h(x, x \cup \{v\}) = \frac{1}{n+2z} \frac{\sum_{u \in \partial x} z^{|u|} + z^{|v|}(2z-1)}{\sum_{u \in \partial x} z^{|u|}}$$

for all $x \in \mathbb{B}$, $v \in \partial x$. Note that this agrees with the transition mechanism of the BST chain if $z = 1/2$ or $z = 1$. For general $z > 0$ a corresponding chain may be constructed by a marking mechanism that makes use of an additional spine variable. This idea was introduced in the context of branching processes; for search trees it has been used in [6], to which paper we refer for more details. The following direct construction of a Markov chain with transitions as in (42) may be of interest: Given X_n, we choose an external node u with probability proportional to $z^{|u|}$. With probability $(2z)/(n+2z)$ we then accept u as the node v to be added to X_n; if u is rejected, then v is chosen uniformly at random from the other n external nodes of X_n.

In the first three subsections of the present section we began our analysis by relating the functionals in question to the subtree sizes. As the latter fully describe the tree this must also be possible in the profile context. For $x \in \mathbb{B}$, $z > 0$ let

$$\Psi_z(x) := \sum_{u \in x} \sigma(x, u) z^{|u|}.$$

Each $v \in \mathbb{V}_{k+1}$ with $\bar{v} \in x$ is either an internal or an external node of x, which means that $v(x, k + 1) = 2w(x, k) - w(x, k + 1)$. Also, the number of internal nodes with depth at least k is the sum of all subtree sizes of the nodes with level exactly equal to k, that is, $\sum_{u \in x, |u|=k} \sigma(x, u) = \sum_{j=k}^{\infty} w(x, j)$. Taken together, this gives

$$\sum_{k=1}^{\infty} v(x, k) z^k = \left(2z - 3 + \frac{1}{z}\right) \Psi_z(x) + \left(2 - \frac{1}{z}\right) \#x + 1,$$

which leads to

$$Y_n = \left(2z - 3 + \frac{1}{z}\right) \Psi_z(X_n) + \left(2 - \frac{1}{z}\right) n + 1$$

(note that the bracketed term vanishes for $z = 1/2$ and $z = 1$). This could serve as the basis for an analysis along the lines of the first three subsections. We do not pursue this here but show instead that the general theory can be used to obtain an interpretation of the function that represents the limit of Jabbour's martingale in terms of the Doob–Martin limit of the BST sequence: Recall that $M_n/(2z)$ is the density associated with the change of measure from $P_{\mathcal{F}_n}$ to $P_{\mathcal{F}_n}^h$. If the convergence $M_n \to M_\infty$ is in L^1 (see below), then $M_\infty/(2z)$ is a density of P^h with respect to P. Thus we have $M_\infty = 2z\Psi(X_\infty)$, with Ψ a density of the distribution of X_∞ under the transformed measure P^h with respect to the distribution of X_∞ under the original P.

It is shown in [6] that L^1-convergence holds if and only if the parameter z is inside a specific bounded interval $I = (c_-, c_+)$, that $M_\infty \equiv 0$ if $z \notin I$, and that, with α_+, α_- as in (41), $c_- = \alpha_-/2$ and $c_+ = \alpha_+/2$. These two phase transitions are related to the asymptotics of the maximum and minimum node size respectively at a specific level of the limit X_∞: If z is too small, then nodes close to the root are favored too much by p^h; if z is too large, then too much weight is given to nodes far away from the root. In both cases P^h is then singular with respect to P. For the weighted subtree size metric considered in Section 3 only one of these caveats matters in that node sizes must not be inflated too much. Hence there is only one such phase transition, which should be related to the height constant, and indeed, a straightforward calculation shows that $\rho_0 = (2e)/\alpha_+$.

Finally, let us mention that the approach toward strong asymptotics of dynamic data structures that we have developed in detail for binary search trees should be applicable in many related situations. The necessary modifications may be minor, such as for the discounted path length that appears in [19], or straightforward, as

for the random recursive trees that are often treated in parallel with binary trees (see, e.g., [25] for the Wiener index), or they may be challenging, for example, when we wish to amplify the weak convergence results for node depth profiles obtained in [14] for a wide class of trees to strong limit theorems as we have done for the Wiener index in Section 4.2. Of course, convergence in distribution and convergence along paths are rather different phenomena; see Figures 1 and 2. It is interesting that for a given dynamical structure we may have a strong limit theorem (with nontrivial limit) for some aspects (functionals), but not for others; see [9] for such results in connection with the subtree size profile of binary search trees.

Acknowledgment. I thank the referee for the stimulating comments, and for providing several valuable references.

REFERENCES

[1] BIGGINS, J. D. (1977). Chernoff's theorem in the branching random walk. *J. Appl. Probab.* **14** 630–636. MR0464415

[2] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York. MR0233396

[3] BLACKWELL, D. and KENDALL, D. (1964). The Martin boundary of Pólya's urn scheme, and an application to stochastic population growth. *J. Appl. Probab.* **1** 284–296. MR0176518

[4] BURAGO, D., BURAGO, Y. and IVANOV, S. (2001). *A Course in Metric Geometry. Graduate Studies in Mathematics* **33**. Amer. Math. Soc., Providence, RI. MR1835418

[5] CHAUVIN, B., DRMOTA, M. and JABBOUR-HATTAB, J. (2001). The profile of binary search trees. *Ann. Appl. Probab.* **11** 1042–1062. MR1878289

[6] CHAUVIN, B., KLEIN, T., MARCKERT, J. F. and ROUAULT, A. (2005). Martingales and profile of binary search trees. *Electron. J. Probab.* **10** 420–435 (electronic). MR2147314

[7] CHAUVIN, B. and ROUAULT, A. (2004). Connecting Yule processes, bisection and binary search tree via martingales. *J. Iran. Stat. Soc.* (*JIRSS*) **3** 89–116.

[8] DENNERT, F. (2009). Zufällige binäre Bäume: Algorithmen, Asymptotik und Statistik. Ph.D. thesis, Leibniz Univ. Hannover.

[9] DENNERT, F. and GRÜBEL, R. (2010). On the subtree size profile of binary search trees. *Combin. Probab. Comput.* **19** 561–578. MR2647493

[10] DEVROYE, L. (1986). A note on the height of binary search trees. *J. Assoc. Comput. Mach.* **33** 489–498. MR0849025

[11] DEVROYE, L. (1998). Branching processes and their applications in the analysis of tree structures and tree algorithms. In *Probabilistic Methods for Algorithmic Discrete Mathematics. Algorithms Combin.* **16** 249–314. Springer, Berlin. MR1678582

[12] DOOB, J. L. (1959). Discrete potential theory and boundaries. *J. Math. Mech.* **8** 433–458; erratum 993. MR0107098

[13] DRMOTA, M. (2009). *Random Trees: An Interplay Between Combinatorics and Probability*. Springer, Vienna. MR2484382

[14] DRMOTA, M., JANSON, S. and NEININGER, R. (2008). A functional limit theorem for the profile of search trees. *Ann. Appl. Probab.* **18** 288–333. MR2380900

[15] EVANS, S. N. (2008). *Probability and Real Trees. Lecture Notes in Math.* **1920**. Springer, Berlin. MR2351587

[16] EVANS, S. N., GRÜBEL, R. and WAKOLBINGER, A. (2012). Trickle-down processes and their boundaries. *Electron. J. Probab.* **17** 1–58. MR2869248

[17] FUCHS, M. (2008). Subtree sizes in recursive trees and binary search trees: Berry–Esseen bounds and Poisson approximations. *Combin. Probab. Comput.* **17** 661–680. MR2454562

[18] FUCHS, M., HWANG, H.-K. and NEININGER, R. (2006). Profiles of random trees: Limit theorems for random recursive trees and binary search trees. *Algorithmica* **46** 367–407. MR2291961

[19] GRÜBEL, R. (2009). On the silhouette of binary search trees. *Ann. Appl. Probab.* **19** 1781–1802. MR2569807

[20] JABBOUR-HATTAB, J. (2001). Martingales and large deviations for binary search trees. *Random Structures Algorithms* **19** 112–127. MR1848787

[21] KALLENBERG, O. (1997). *Foundations of Modern Probability*. Springer, New York. MR1464694

[22] KAÏMANOVICH, V. A. and VERSHIK, A. M. (1983). Random walks on discrete groups: Boundary and entropy. *Ann. Probab.* **11** 457–490. MR0704539

[23] KNUTH, D. E. (1973). *The Art of Computer Programming. Vol.* 3: *Sorting and Searching*. Addison-Wesley, Reading, MA. MR0445948

[24] MAHMOUD, H. M. (1992). *Evolution of Random Search Trees*. Wiley, New York. MR1140708

[25] NEININGER, R. (2002). The Wiener index of random trees. *Combin. Probab. Comput.* **11** 587–597. MR1940122

[26] NEVEU, J. (1975). *Discrete-Parameter Martingales*, Revised ed. North-Holland, Amsterdam. MR0402915

[27] RÉGNIER, M. (1989). A limiting distribution for quicksort. *RAIRO Inform. Théor. Appl.* **23** 335–343. MR1020478

[28] RÖSLER, U. (1991). A limit theorem for "Quicksort." *RAIRO Inform. Théor. Appl.* **25** 85–100. MR1104413

[29] WOESS, W. (2000). *Random Walks on Infinite Graphs and Groups. Cambridge Tracts in Mathematics* **138**. Cambridge Univ. Press, Cambridge. MR1743100

[30] WOESS, W. (2009). *Denumerable Markov Chains*: *Generating Functions*, *Boundary Theory*, *Random Walks on Trees*. EMS, Zürich. MR2548569

INSTITUT FÜR MATHEMATISCHE STOCHASTIK
LEIBNIZ UNIVERSITÄT HANNOVER
POSTFACH 6009
30060 HANNOVER
GERMANY
E-MAIL: rgrubel@stochastik.uni-hannover.de