Discrete Mathematics and Theoretical Computer Science DMTCS vol. **(subm.)**, by the authors, 1–8

Note on the weighted internal path length of b-ary trees

Ludger Rüschendorf[1] and Eva-Maria Schopp[1]

[1] *Department of Mathematical Stochastics, University of Freiburg, Eckerstraße 1, 79104 Freiburg, Germany*

received ???, revised 9th January 2007, accepted ???.

In a recent paper Broutin and Devroye (2005) have studied the height of a class of edge-weighted random trees. This is a class of trees growing in continuous time which includes many well known trees as examples. In this paper we derive a limit theorem for the internal path length for this class of trees. The application of this limit theorem to concrete examples depends upon the possibility to obtain an expansion of the mean of the path length. For the proof we extend a limit theorem in Neininger and Rüschendorf (2004) to recursive sequences of random variables with continuous time parameter.

Keywords: b-ary random trees, continuous time tree models, contraction method, internal path length

1 Introduction

In this paper we derive a limit theorem for the internal path length of edge-weighted b-ary random trees. For this class of trees which includes as particular cases many well known types of random trees as e.g. random binary search trees, random recursive trees, random split trees and others in a recent paper by Broutin and Devroye (2005), a general law of large numbers for the weighted height was established.

The weighted b-ary tree is defined as follows. Let T_∞ denote an infinite complete rooted b-ary tree. To each node u in T_∞ independently a random vector $((Z_1, E_1), \dots, (Z_b, E_b))$ is assigned corresponding to the b outgoing edges of u where $Z_i, E_i \geq 0$, each pair (Z_i, E_i) is identically distributed as (Z, E) and where Z, E have finite expectations. We also assume that (Z_i) and (E_i) are independent. Z_e assigns a weight and E_e an age to edge e. Then for node u

$$G_u := \sum_{e \in \pi(u)} E_e \quad \text{is the age of } u \tag{1}$$

$$D_u := \sum_{e \in \pi(u)} Z_e \quad \text{is the weighted depth of } u. \tag{2}$$

Here $\pi(u)$ is the set of edges in T_∞ on the path of the root to node u. The b-ary tree of age $\leq t$ then is defined in continuous time $t \geq 0$ by

$$T_t := \{u \in T_\infty; G_u \leq t\}. \tag{3}$$

Important parameters of T_t are the height of T_t

$$H_t := \max\{D_u : u \in T_t\} \tag{4}$$

the path length of T_t,

$$Y_t := \sum_{u \in T_t} D_u \tag{5}$$

and $V_t = |T_t|$ the random number of nodes of T_t. Broutin and Devroye (2005) proved a strong law for the height assuming that Z_i, E_i are independent. More precisely they established

$$\frac{H_n}{n} \underset{P}{\longrightarrow} c, \tag{6}$$

where $c = \operatorname{argmax}\{\frac{\alpha}{\rho}; (\rho, \alpha) \in C_{Z,E}\}$, with $C_{Z,E} = \{(\rho, \alpha); \Lambda^*_{Z,E}(\alpha, \rho) = \log b, \rho \leq E(E), \alpha \geq E(Z)\}$. Here $\Lambda^*_{Z,E} = \sup_{\lambda \in \mathbb{R}^2}\{\langle t, \lambda\rangle - \log E e^{\langle \lambda, (Z,E)\rangle}\}$ denotes the Cramer function of (Z, E). This result applies to rBST, random recursive trees, plane oriented trees, oriented trees, split trees and others. Their proof was based on Chernoff's theorem and extends earlier results of Biggins and Grey (1997) and Biggins (1977, 1978) using branching random walks. The upper bound in (6) can be extended to dependent reproduction based on the Gärtner–Ellis theorem (see Schopp (2005)). The lower bound however needs a new Galton–Watson type result for the case of dependent reproduction which seems to be not available in sufficient generally. The application of our limit theorem to concrete examples depends upon an expansion of the first moment of the path length resp. in some cases of the first two moments.

2 Limit theorem for the weighted internal path length

For the internal path length of b-ary weighted trees as introduced in Section 1 we obtain the following recursive equation in continuous time which arises when splitting the tree at the root:

$$Y_t \stackrel{d}{=} \sum_{i=1}^{b} Y_{t-E_i}^{(i)} \mathbb{1}_{\{E_i \le t\}} + b_t, \qquad t > 0 \tag{7}$$

where $b_t = \sum_{i=1}^{b} Z_i V_{t-E_i}^{(i)}$, defining $Y_0 := 0$, $V_s := 0$ for $s \le 0$. Here $(Y_t^{(i)})$ are independent copies of the internal path length process Y_t, $V_{t-E_i}^{(i)}$ is the number of nodes in subtree i with age $\le t - E_i$, $(V_t^{(i)})$ are independent copies of each other. To argue for (7) let $u_1, \dots, u_b$ the b nodes of T_t below the root with corresponding ages $E_1, \dots, E_b$. If $E_1 > t$, then V_{t-E_1} the number of nodes in the subtree with root u_1, is zero and we get no contribution of this subtree to the internal path length. Only the nodes in the subtree of u_1 of age less than $t - E_1$ contribute to the internal path length. For each of them we have to add the weight Z_1 of the edge from the root to u_1, i.e. $Z_1 V_{t-E_1}^{(1)}$. Similarly, the contribution of the other subtrees is accounted in (7) yielding the recursion

$$Y_t \stackrel{d}{=} \sum_{i=1}^{b} Y_{t-E_i}^{(i)} \mathbb{1}_{\{E_i \le t\}} + b_t. \tag{8}$$

To deal with the recursive random variables (Y_t) with continuous time parameter t as in (8) we derive in the following an extension to continuous time of the contraction method as developed in Neininger and Rüschendorf (2004) (see also Rösler and Rüschendorf (2001)). Let $0 < s \le 3$, let Y_t be s-integrable for all t, and consider the normalized version X_t of Y_t defined by

$$X_t := \frac{Y_t - M_t}{\sqrt{C_t}}, \tag{9}$$

where for $1 < s \le 3$, $M_t := EY_t$ and for $2 < s \le 3$, $C_t := \mathrm{Var}(Y_t)$, $C_t > 0$ else (for the motivation of this normalization see Neininger and Rüschendorf (2004)). Convergence will be formulated w.r.t. the Zolotarev metric

$$\zeta_s(X, Y) = \sup_{f \in \mathcal{F}_s} |E(f(X) - f(Y))|, \tag{10}$$

where $s = m + \alpha$, $0 < \alpha \le 1$, $m = \lceil s \rceil - 1 \ge 0$ is an integer and $\mathcal{F}_s = \{f \in C^m(\mathbb{R}, \mathbb{R});\ \|f^{(m)}(x) - f^{(m)}(y)\| \le |x - y|^\alpha\}$ denotes the space of m-fold continuously differentiable real functions on $\mathbb{R}^1$ with a Hölder condition for the m-th derivative. $\zeta_s(X, Y)$ is finite if X, Y have finite absolute moments of order s and the moments of order $1, \dots, m$ of X and Y coincide. ζ_s is an ideal metric of order s, i.e. for Z independent of X, Y and any $c \in \mathbb{R}$ holds

$$\zeta_s(X + Z, Y + Z) \le \zeta_s(X, Y), \quad \zeta_s(cX, cY) = |c|^s \zeta_s(X, Y). \tag{11}$$

The normalized version X_t of Y_t satisfies a recursive equation of a form similar to (8):

$$X_t \stackrel{d}{=} \sum_{r=1}^{b} A_r^{(t)} X_{t-E_r}^{(r)} + b^{(t)}, \tag{12}$$

where $A_r^{(t)} := \mathbb{1}_{\{E_r \le t\}} \sqrt{\frac{C_{t-E_r}}{C_t}}$ and

$$b^{(t)} := \frac{1}{\sqrt{C_t}} \left(b_t - M_t + \sum_{r=1}^{b} \mathbb{1}_{\{E_r \le t\}} M_{t-E_r} \right). \tag{13}$$

Theorem 1 *Let $0 < s \le 3$ und $X_t \in L^s$ satisfy the recursive equation (12) and assume that $\|A_r^{(t)}\|_s < \infty$, $\|b^{(t)}\|_s < \infty$ and $\sup_{0 \le u \le t} \|X_u\|_s < \infty$ for all $t > 0$. Assume further that*

1) $A_r^{(t)} \xrightarrow{L^s} A_r^*$, $b^{(t)} \xrightarrow{L^s} b^*$ *as* $t \to \infty$ $\quad (14)$

2) $E \sum_{r=1}^{b} |A_r^*|^s < 1$ $\quad (15)$

3) For all $\tau > 0$ holds $E \sum_{r=1}^{b} \mathbb{1}_{\{t-E_r < \tau\}} |A_r^{(t)}|^s \to 0.$ $\quad (16)$

Then X_t converges in distribution to a limit X,

$$\zeta_s(X_t, X) \to 0 \text{ as } t \to \infty \tag{17}$$

and X is in law the unique solution of the fixpoint equation

$$X \stackrel{d}{=} \sum_{r=1}^{b} A_r^* X^{(r)} + b^* \tag{18}$$

in L^s with $EX = 0$ for $1 < s \le 3$ and $\operatorname{Var} X = 1$ for $2 < s \le 3$.

Proof: Note that by the normalization for $1 < s \le 2$ X_t is centered, thus $Eb^{(t)} = 0$. For $2 < s \le 3$ $EX_t = 0$, $\operatorname{Var}(X_t) = 1$ and thus $Eb^{(t)} = 0$ and $E(b^{(t)})^2 + E\sum_{r=1}^{b} (A_r^{(t)})^2 = 1$. Thus from assumption (14) we obtain $Eb^* = 0$, $1 < s \le 2$ and

$$Eb^* = 0, \quad E(b^*)^2 + E\sum_{r=1}^{b} (A_r^*)^2 = 1, \text{ for } 2 < s \le 3. \tag{19}$$

This implies by Corollary 3.4 of Neininger and Rüschendorf (2004) existence and uniqueness of a solution of (17) in $M_s(0,1)$, the class of distributions of all $X \in L^s$ with moments as specified above. We introduce as in the discrete time case an accompanying sequence Q_t of X_t by

$$Q_t := \sum_{r=1}^{b} A_r^{(t)} \left(\mathbb{1}_{\{0 \le t-E_r < \tau\}} X_{t-E_r}^{(r)} + \mathbb{1}_{\{t-E_r \ge \tau\}} X^{(r)} \right) + b^{(t)}, \quad t > 0, \tag{20}$$

where $(X^{(r)})$, $(X_t^{(r)})$ are independent copies of X, X_t and τ is some suitable positive number specified later in the proof. Then for $2 < s \le 3$ $\operatorname{Var}(Q_t) = \operatorname{Var}(X_t)$ and thus $Q_t \in M_s(0,1)$ and the distance between X_t, Q_t, X w.r.t. the Zolotarev metric is finite for all $t > 0$.

By the triangle inequality holds

$$d_t := \zeta_s(X_t, X) \le \zeta_s(X_t, Q_t) + \zeta_s(Q_t, X). \tag{21}$$

As in the discrete case we obtain that the remainder term $r_t := \zeta_s(Q_t, X) \to 0$ as $t \to \infty$ using (16) and the condition $\sup_{0<t\le\tau} \|X_t\|_s < \infty$. Further, using the ideality properties of the ζ_s-metric we obtain

$$\zeta_s(X_t, Q_t) \le E \sum_{r=1}^{b} \mathbb{1}_{\{0 \le t - E_r > \tau\}} |A_r^{(t)}|^s d_{t-E_r}$$

and thus by (21) for $t \ge \tau$

$$\begin{aligned} d_t \;&\le\; E \sum_{r=1}^{b} \mathbb{1}_{\{0 \le t - E_r \ge \tau\}} |A_r^{(t)}|^s d_{t-E_r} + r_t \\ &\le\; E \sum_{r=1}^{b} \mathbb{1}_{\{0 \le t - E_r \ge \tau\}} |A_r^{(t)}|^s \sup_{\tau \le u \le t} d_u + r^*, \end{aligned} \tag{22}$$

where $r^* = \sup_{\tau \le t} r_t < \infty$. By an inequality due to Zolotarev

$$d_t^* := \sup_{\tau \le u \le t} d_u \le C\Big(\|X\|_s^s + \sup_{\tau \le u \le t} \|X_u\|_s^s \Big) < \infty$$

with some constant $C > 0$. Thus we obtain by assumption (15) from (22)

$$d_t \le \eta \cdot d_t^* + r^*, \quad t > \tau \tag{23}$$

for some $\eta < 1$ if τ is chosen large enough. This implies $d_t^* \le \eta d_t^* + r^*$ by monotonicity of d_t^*, i.e. $d_t^* \le \frac{r^*}{1-\eta}$ for all $t > \tau$. Thus we get that d_t is bounded.

Now we refine the estimate as in the discrete case to obtain that $d_t \to 0$. Let $\overline{d} := \limsup_{t\to\infty} d_t$. Then for any $\epsilon > 0$ holds $d_t \le \overline{d} + \epsilon$ for all $t \ge \tau_1$ and thus by (20), (21) with $d_\infty^* = \sup\limits_t d_t^*$

$$d_t \le E \sum_{r=1}^{b} \mathbb{1}_{\{\tau \le t - E_r \le \tau_1\}} |A_r^{(t)}|^s d_\infty^* + E \sum_{r=1}^{b} \mathbb{1}_{\{t - E_r > \tau_1\}} |A_r^{(t)}|^s (\overline{d} + \epsilon) + r_t. \tag{24}$$

Using assumptions (15), (16), this implies $\overline{d} \le \xi(\overline{d} + \epsilon)$ where $\xi = E \sum_{r=1}^{b} |A_r^*|^s < 1$ a contradiction for $\epsilon \le \epsilon_0$.

Since ζ_s-convergence implies weak convergence we obtain the conclusion of the theorem. □

Remark.

a) In order to apply the limit theorem to concrete b-ary weighted trees we have to control the first moment of Y_t for $1 < s \le 2$ and the first and second moment for $2 < s \le 3$ (as typically in the case of normal limits). In the case of discrete time recursive sequences several examples of this type have been given in Neininger and Rüschendorf (2004). Broutin and Devroye (2005) applied their results on the height of b-ary weighted trees to several trees. In the case where the age variables are exponential the corresponding b-ary tree has a Markov structure, the number of nodes V_t can be determined by a law of large numbers and so a transference to e.g. rBST's is possible (see Broutin and Devroye (2005)).

b) Assumption 3) of Theorem 1 can be weakened a bit for the limit theorem (see Schopp (2005)). Also the random *subtree sizes* $t - E_r$ in the recursive equation (12) can be replaced in the formulation of Theorem 1 by general subtree size $I_r(t) \leq t$ as in the discrete time case in Neininger and Rüschendorf (2004). In a recent paper of Janson and Neininger (2006) a similar extension of the limit theorem of Neininger and Rüschendorf (2004) to the continuous time case has been independently established (even in the multivariate case) and has been applied to a fragmentation process.

c) In Theorem 1 we assume finiteness of the s-th absolute moments of the random modified coefficients $(A_r^{(t)})$ and the modified toll terms $(b^{(t)})$. Thus integrability properties of Z, E may have an impact on the applicability of the theorem.

In many applications (see Broutin and Devroye (2005)) Z is a bounded random variable. Thus for the finiteness of the s-th moment of $b^{(t)}$ it suffices in that case to estimate the sth absolute moment of the number of nodes up to time t, since

$$\|b^{(t)}\|_s \leq \frac{1}{\sqrt{C_t}} b \|ZV_t\|_s + c(t),$$

where $c(t)$ is a constant depending on t.

If for example E is Exponential(1)-distributed, then $\{\widetilde{V}_t \leq n\} = \{t_n \geq t\}$, where t_n is the time of the n-th birth and $\widetilde{V}_t$ denote the number of external nodes in the b-ary tree.

As $t_k \stackrel{d}{=} \sum_{i=1}^k \frac{E_i}{1+(i-1)(b-1)}$ we obtain

$$\begin{aligned}
P(\widetilde{V}_t \geq k) &= P\Big(\sum_{i=1}^k \frac{E_i}{1+(i-1)(b-1)} \leq t\Big) \\
&= P\Big(\frac{1}{b-1}\sum_{i=1}^k \frac{E_i}{\frac{1}{b-1}+(i-1)} \leq t\Big) \\
&\leq P\Big(\sum_{i=1}^k \frac{E_i}{i} \leq t(b-1)\Big) \\
&= P\big(\max\{E_1,\dots,E_k\} \leq t(b-1)\big) \\
&= (1-e^{-t(b-1)})^k.
\end{aligned}$$

Therefore the s-th moment of $\widetilde{V}_t$ is bounded by the s-th moment of a geometric $\mathcal{G}(e^{-t(b-1)})$ random variable, and since $P(V_t \geq k) = P(\widetilde{V}_t \geq (b-1)k+1)$ we have also a bound for V_t.

References

J. D. Biggins. Chernoff's theorem in the branching random walk. *J. Appl. Prob.*, 14:630–636, 1977.

J. D. Biggins. The asymptotic shape of the branching random walk. *Adv. Appl. Prob.*, 10:62–84, 1978.

J. D. Biggins and D. R. Grey. A note on the growth of random trees. *Statistics Probab. Letters*, 32: 339–342, 1997.

N. Broutin and L. Devroye. Large deviations for the weighted height of an extended class of trees. Technical report, McGill University, Montreal, Canada, 2005.

S. Janson and R. Neininger. The size of random fragmentation trees. Technical report, Goethe University of Frankfurt, Germany, 2006.

R. Neininger and L. Rüschendorf. A general limit theorem for recursive algorithms and combinatorial structures. *Ann. Appl. Prob.*, 14:378–418, 2004.

U. Rösler and L. Rüschendorf. The contraction method for recursive algorithms. *Algorithmica*, 29:3–33, 2001.

E.-M. Schopp. *Stochastische Fixpunktgleichungen, exponentielle tail Abschätzungen und large deviation für rekursive Algorithmen*. Diplomarbeit, Dept. of. Mathematical Stochastics, University of Freiburg, Germany, 2005.