THE SATURATION LEVEL IN BINARY SEARCH TREE

MICHAEL DRMOTA

ABSTRACT. Let H'_n denote the saturation level in binary search trees. It is shown that $\mathbf{E}H'_n = c' \log n + \mathcal{O}\left((\log n)^{1/2} \log\log n\right)$, where $c' = 0.373\ldots < 2$ is the solution of the equation $\left(\frac{2e}{c'}\right)^{c'} = e$. The methods used rely on the analysis of a *retarded* differential equation of the form $\Phi'(u) = -\alpha^{-2}\Phi(u/\alpha)^2$ with $\alpha > 1$.

1. Introduction

Let us consider the usual probabilistic model for binary search trees (see Mahmoud [8] for a description and an overview of the state of the art). The saturation level H'_n of a binary search tree is defined to be maximal level h' such that for all levels h up to h' there are no external nodes, i.e. the binary search has 2^h (internal) nodes for all levels $h \le h'$ but less than $2^{h'+1}$ (internal) nodes at level $h'+1$. For example, if follows from Biggins [1] that

$$\frac{H'_n}{c' \log n} \to 1 \quad a.s.,$$

where $c' = 0.373\ldots < 2$ is the solution of the equation $\left(\frac{2e}{c'}\right)^{c'} = e$. (It can also be shown that $\mathbf{E}H'_n \sim c' \log n$, compare with Mahmoud [8].) Quite recently, the author [7] could prove the following property for $\mathbf{E}H'_n$ and for $\mathbf{E}|H'_n - \mathbf{E}H'_n|^k$.

Theorem 1. *Let $z_h(x)$ $(h \ge 0)$ be recursively defined by $z_0(x) \equiv 1$ and by*

$$z_{h+1}(x) := 1 + \int_0^x y_h(t)\left(\frac{2}{1-t} - z_h(t)\right) dt \qquad (h \ge 0).$$

Then the expected value $\mathbf{E}H'_n$ of the saturation level of binary search trees is given by

$$\mathbf{E}H'_n = \max\{h : z_h(1-n^{-1}) \le n/2\} + \mathcal{O}(1) \qquad (n \to \infty) \tag{1}$$

and all centralized moments of H'_n are bounded:

$$\mathbf{E}|H'_n - \mathbf{E}H'_n|^k = \mathcal{O}(1) \qquad (n \to \infty). \tag{2}$$

Note that (1) is quite implicit. It does not reprove the limiting relation $\mathbf{E}H'_n \sim c' \log n$. For this purpose we have to discuss the functions $z_h(x)$ in more detail. (Interestingly, it is not necessary to know the asymptotic behaviour of $\mathbf{E}H'_n$ to prove boundedness of the variance etc., compare with [7].) The purpose of this paper is to extends the methods of [7] to obtain proper bounds for $z_h(x)$ and to prove the following quantified limiting relation.

Theorem 2. *The expected value $\mathbf{E}H'_n$ of the saturation level of binary search trees is given by*

$$\mathbf{E}H'_n = c' \log n + \mathcal{O}\left((\log n)^{1/2} \log\log n\right). \tag{3}$$

Date: March 15, 2000.

Key words and phrases. binary search tree, average case analysis.

Institut für Geometrie, TU Wien, Wiedner Hauptstrasse 8-10/118, A-1040 Wien, Austria, email: drmota@tuwien.ac.at.

It is very likely that the error term $\mathcal{O}\left((\log n)^{1/2}\log\log n\right)$ in (3) is not optimal. It might be expected that there is a constant d' such that

$$\mathbf{E}H_n' = c'\log n + d'\log\log n + \mathcal{O}(1).$$

The method presented below is adapted from that used in [6] and [7] to treat the height of binary search trees. (For results and references concerning the height of binary search trees we refer to [2, 3, 4, 6, 9].)

Before we start with the proof we want to mention that the functions $z_h(x)$ introduced in Theorem 1 are just the generating functions of the probabilities $\mathbf{P}[H_n' \le n]$:

$$z_h(x) = \sum_{n\ge 0} \mathbf{P}[H_n' \le n]\, x^n.$$

The paper is organized in the following way. In section 2 we prove an upper bound for $z_h(x)$ yielding an lower bound for $\mathbf{E}H_n'$. Sections 3 and 4 are devoted to the analysis of an integral equation. Those results will be then applied in section 5 to provide a lower bound for $z_h(x)$ which finally yields an upper bound for $\mathbf{E}H_n'$

2. A lower bound for $\mathbf{E}H_n'$.

Interestingly, a lower bound for $\mathbf{E}H_n'$ is much easier to obtain than an upper bound. We can use the following lemma.

Lemma 1. *For $0 \le x < 1$ we have*

$$z_h(x) \le 2^h \sum_{k=0}^{h} \frac{\left(\log\frac{1}{1-x}\right)^k}{k!}.$$

Proof. Obviously, we have equality for $h = 0$. Now we use the inequality

$$z_{h+1}'(x) = z_h(x)\left(\frac{2}{1-x} - z_h(x)\right) \le \frac{2}{1-x} z_h(x) \tag{4}$$

and proceed by induction. □

Corollary . *Let $h = c'\log n + \frac{c'}{2(1-c')}\log\log n + r$. Then*

$$z_h\left(1-\frac{1}{n}\right) \ll n\left(\frac{2}{c'}\right)^r$$

uniformly for $r = \mathcal{O}(1)$.

Thus, it follows from Theorem 1 that

$$\mathbf{E}H_n' \ge c'\log n + \frac{c'}{2(1-c')}\log\log n - C_1 \tag{5}$$

for some constant $C_1 > 0$.

3. An Integral Equation

Our aim is to find proper solutions $\Phi(u)$ of the (retarded) differential equation

$$\Phi'(u) = -\frac{1}{\alpha^2}\Phi(u/\alpha)^2 \tag{6}$$

with

$$\alpha \ge \alpha_0 = e^{1/c'} = 14.59\ldots$$

Interestingly there is an explicit solution

$$\Phi(u) = \frac{1+u^{1/4}}{u} e^{-u^{1/4}} \tag{7}$$

with $\alpha = 16$. This function was used to prove Theorem 1 in [7]. However, if we want to get more then we need solutions for every $\alpha > \alpha_0$.

Instead of solving (6) directly we consider a related integral equation. In fact, we will prove the following relations.

Lemma 2. *Let $\alpha > \alpha_0 = e^{1/c'}$ and let $\beta > c' - 1$ be the solution of the equation*

$$2\alpha^\beta = \beta + 1.$$

Then there exists a function $\Psi(y)$, $y \ge 0$ with the following properties:

1. $\Psi(y) = 1 - y^\beta + \mathcal{O}((1-(\alpha/\alpha_0)^{c'-1})^{-1} y^{c'-1})$ *as $y \to \infty$ with an $\mathcal{O}$-constant independent of α.*
2. $\Psi(y) \le e^{-C(\alpha) y^{-\gamma}}$ *for $\gamma = \log 2/(\log(\alpha/2))$ and $y \le y_0(\alpha)$, where $C(\alpha)$ and $y_0(\alpha)$ can be chosen in a way that $C(\alpha) = \exp(-c_1(\alpha-\alpha_0)^{-1/2}\log((\alpha-\alpha_0)^{-1}))$ and $y_0(\alpha) = \exp(c_2(\alpha-\alpha_0)^{-1/2}\log((\alpha-\alpha_0)^{-1}))$ for some constants $c_1, c_2 > 0$.*
3. $0 \le \Psi(y) \le 1$, $0 \le y < \infty$, *is strictly increasing.*
4. $\displaystyle\int_0^y \Psi(z)\Psi(y-z)\,dz = y\Psi(y/\alpha)$, $(0 \le y < \infty)$.

Proof. We first observe that for $\alpha > \alpha_0 = e^{1/c'}$ the equation

$$2\alpha^\beta = \beta + 1 \tag{8}$$

has two negative solutions β_1, β_2 with $\beta_1 < \beta_0 = c' - 1 = -0.626\cdots < \beta_2$. The solution we are interested in is $\beta = \beta_2 > \beta_0$. For $\alpha = \alpha_0$ there is only one solution $\beta = \beta_0 < 0$ and for $\alpha < \alpha_0$ there are no negative solutions. It is also an easy exercise to show that for $\alpha > \alpha_0$ we have

$$\beta = \beta_0 + \sqrt{\frac{2(c')^2(1-c')}{\alpha_0}}\sqrt{\alpha-\alpha_0} + \mathcal{O}(\alpha-\alpha_0).$$

Now let $\mathcal{F}$ denote the set of functions $\Psi(y)$, $y > 0$, with the following properties:

1. $\Psi(y) = 1 - y^\beta + \mathcal{O}(y^{\beta_0})$ as $y \to \infty$.
2. $\Psi(0) = 0$.
3. $0 \le \Psi(y) \le 1$, $0 \le y < \infty$.
4. $\Psi(y)$, $0 \le y < \infty$, is strictly decreasing.

It is clear that $\mathcal{F}$ with the distance

$$d(\Psi_1, \Psi_2) := \sup_{y>0} |(\Psi_1(y) - \Psi_2(y))y^{-\beta_0}|$$

is a complete metric space. Now we show that the operator I, defined by

$$(I\Psi)(y) := \frac{1}{\alpha y}\int_0^{\alpha y} \Psi(z)\Psi(\alpha y - z)\,dz$$

is a contraction on $\mathcal{F}$.

Firstly, we prove that $I\Psi \in \mathcal{F}$ for all $\Psi \in \mathcal{F}$. Suppose that $\Psi \in \mathcal{F}$. Then

$$\Psi(z)\Psi(\alpha y - z) = 1 - (z^\beta + (\alpha y - z)^\beta) + \mathcal{O}(z^{\beta_0}) + \mathcal{O}((\alpha y - z)^{\beta_0}).$$

Since

$$\int_0^{\alpha y} z^\beta\,dz = \frac{(\alpha y)^{\beta+1}}{\beta+1}$$

it immediately follows that

$$\begin{aligned}(I\Psi)(y) &= \frac{1}{\alpha y}\int_0^{\alpha y} \Psi(z)\Psi(\alpha y - z)\,dz \\ &= 1 - 2\frac{(\alpha y)^{\beta+1}}{\alpha y(\beta+1)} + \mathcal{O}(y^{\beta_0}) \\ &= 1 - y^\beta + \mathcal{O}(y^{\beta_0}).\end{aligned}$$

Furthermore, it is clear that $(I\Psi)(0) = 0$ and $0 \le (I\Psi)(y) \le 1$. Finally, by using the representation

$$(I\Psi)(y) = \int_0^1 \Psi(\alpha y x)\Psi(\alpha y(1-x))\,dx$$

it is also clear that $(I\Psi)(y)$ is increasing (if $\Psi(y)$ is increasing).

Now suppose that $\Psi_1, \Psi_2 \in \mathcal{F}$ with $d(\Psi_1, \Psi_2) = \delta$. Then it follows from $0 \le \Psi_j(y) \le 1$ that

$$\begin{aligned} |\Psi_1(z)\Psi_1(\alpha y - z) - \Psi_2(z)\Psi_2(\alpha y - z)| &\le |\Psi_1(z) - \Psi_2(z)| + |\Psi_1(\alpha y - z) - \Psi_2(\alpha y - z)| \\ &\le \delta\left(z^{\beta_0} + (\alpha y - z)^{\beta_0}\right) \end{aligned}$$

and consequently

$$\begin{aligned} |(I\Psi_1)(y) - (I\Psi_2)(y)| &\le \delta \frac{2}{\alpha y}\int_0^{\alpha y} z^{\beta_0}\,dz \\ &= \delta \frac{2\alpha^{\beta_0}}{\beta_0 + 1} y^{\beta_0} \end{aligned}$$

which implies

$$d(I\Psi_1, I\Psi_2) \le L\, d(\Psi_1, \Psi_2)$$

with

$$L = \frac{2\alpha^{\beta_0}}{\beta_0 + 1} = \left(\frac{\alpha}{\alpha_0}\right)^{\beta_0} < 1.$$

Thus, $I : \mathcal{F} \to \mathcal{F}$ is as contraction.

By Banach's fixed point theorem there exists a unique fixed point $\Psi \in \mathcal{F}$. By definition, this fixed points satisfies properties 1., 3., and 4. of Lemma 2.

If we start with

$$\Psi_0(y) = \max\{1 - y^\beta, 0\}$$

and define iteratively $\Psi_{k+1} := I\Psi_k$ then the fixed point is the limit $\Psi = \lim_{k\to\infty} \Psi_k$ and we have

$$d(\Psi_0, \Psi) = \mathcal{O}\left(\frac{1}{1-L}\right),$$

which directly translates to the precise estimate in 1.

In order to prove 2. we set

$$\overline{\Psi}(y) := e^{-Cy^{-\gamma}}$$

with $\gamma = \log 2/\log(\alpha/2)$. Since $\gamma > 1$ we surely have

$$\begin{aligned} \frac{1}{y}\int_0^y \overline{\Psi}(z)\overline{\Psi}(y-z)\,dz &\le \frac{1}{y}\int_0^y \overline{\Psi}(y/2)^2\,dz \\ &= \overline{\Psi}(y/2)^2 \\ &= \overline{\Psi}(y/\alpha). \end{aligned}$$

Thus, if we know that $\Psi_k(y) \le \overline{\Psi}(y)$ for $y \le y_0$ then it follows that $(I\Psi_k)(y) = \Psi_{k+1}(y) \le \overline{\Psi}(y)$ for $y \le y_0/\alpha$. However, there is an a-priori bound for all Ψ_k of the form

$$\Psi_k(y) \le 1 - y^\beta + C'(1 - (\alpha/\alpha_0)^{c'-1})^{-1} y^{\beta_0}.$$

Hence, if we choose C and y_0 appropriately (i.e., as proposed) then we can assure that $\Psi_0(y) \le \overline{\Psi}(y)$ for $y \le y_0$ and (a-priori) $\Psi_k(y) \le \overline{\Psi}(y)$ for $y_0/\alpha \le y \le y_0$. Thus, 2. follows by induction. □

Corollary . *Let $\Psi(y)$ be as in Lemma 2 then the Laplace transform*

$$\Phi(u)=\int_0^\infty \Psi(y)e^{-uy}\,dy$$

is an analytic function for $\Re u>0$ and satisfies the following properties.

1. *$\Phi(u)$ and and $u\Phi(u)$ are decreasing for real $u>0$.*
2. *$1-u\Phi(u)=\Gamma(\beta+1)u^{-\beta}+\mathcal{O}\left((1-(\alpha/\alpha_0)^{c'-1})^{-1}u^{1-c'}\right)$ as $u\to 0+$.*
3. *$\Phi(u)=\mathcal{O}(e^{-C(\alpha)u^{\gamma'}})$ as $u\to\infty$, where $\gamma'>0$ and $C(\alpha)$ is of the form stated in Lemma 2.*
4. *$\Phi'(u)=-\dfrac{1}{\alpha^2}\Phi\left(\dfrac{u}{\alpha}\right)^2$.*

Proof. It is clear that the Laplace transform $\Phi(u)$ is an analytic function for $\Re u>0$. Moreover, since $\Phi(y)$ is non-negative, it follows by definition that $\Phi(u)$ is decreasing.

By partial integration we get for any $u>0$

$$u\Phi(u)=-\int_0^\infty \Psi'(y)e^{-uy}\,dy.$$

Since $\Psi'(y)>0$ for $y>0$ it also follows that $u\Phi(u)$ is decreasing for $u>0$, too.

Next, the expansion $\Phi(y)=1-y^\beta+\mathcal{O}(y^{\beta_0})$ as $y\to\infty$ directly translates to

$$\Phi(u)=\frac{1}{u}-\frac{\Gamma(\beta+1)}{u^{\beta+1}}+\mathcal{O}\left((1-(\alpha/\alpha_0)^{c'-1})^{-1}\frac{1}{u^{\beta_0+1}}\right)$$

as $u\to 0+$.

Finally the integral equation for $\Psi(y)$ induces the proposed differential equation for $\Phi(u)$. □

4. Auxilary Functions

We will now work with the auxiliary functions

$$\tilde z_h(x):=\frac{1}{1-x}-\tilde y_h(x)=\frac{1}{1-x}\left(1-\alpha^h(1-x)\Phi(\alpha^h(1-x))\right), \tag{9}$$

where h is a real parameter.

The properties of $\Phi(u)$ can be translated to corresponding properties of $\tilde z_h(x)$. The proof is immediate. The idea behind is that theses functions imitate the original functions $z_h(x)$.

Lemma 3. *The functions $\tilde z_h(x)$, $h\ge 0$, $0\le x<1$, defined by (9) satisfy*

1. *$0<\tilde z_h(0)<1$.*
2. *$0\le \tilde z_h(x)\le 1/(1-x)$ for $0\le x<1$.*
3. *$\tilde z_h(x)$ is strictly increasing for $0\le x<1$.*
4. *$\tilde z'_{h+1}(x)=\left(\dfrac{2}{1-x}-\tilde z_h(x)\right)\tilde z_h(x)$.*

We will also make use of the following property of the functions $z_h(x)$.

Lemma 4. *For every non-negative integer h and for every (real) D there exists $0<\overline{x}_{h,D}<1$ such that*

$$\tilde z_{h+D}(x)<z_h(x)\qquad (0\le x<\overline{x}_{h,D}) \tag{10}$$

and

$$z_h(x)<\tilde z_{h+D}(x)\qquad (\overline{x}_{h,D}<x<1). \tag{11}$$

Furthermore

$$\overline{x}_{h+1,D}>\overline{x}_{h,D}. \tag{12}$$

Proof. We proceed by induction. Since $\tilde{z}_D(x)$ is strictly increasing and satisfies $0 < \tilde{z}_D(0) < 1$ and $\lim_{x\to 1-} \tilde{z}_D(x) = \infty$ the assertion is surely true for $h = 0$. Now suppose that (10) and (11) are satisfied for some $h \ge 0$, i.e. the difference

$$\delta_{h,D}(x) := z_h(x) - \tilde{z}_{h+D}(x)$$

has a unique zero $\overline{x}_{h,D} > 0$ such that $\delta_{h,D}(x) > 0$ for $0 \le x < x_{h,D}$ and $\delta_{h,D}(x) < 0$ for $x > x_{h,D}$. Now we have

$$\begin{aligned}
\delta'_{h+1,D}(x) &= z'_{h+1}(x) - \tilde{z}'_{h+1+D}(x) \\
&= z_h(x)\left(\frac{2}{1-x} - z_h(x)\right) - \tilde{z}_{h+D}(x)\left(\frac{2}{1-x} - \tilde{z}_{h+D}(x)\right) \\
&= \delta_{h,D}(x)\left(\frac{2}{1-x} - z_h(x) - \tilde{z}_{h+D}(x)\right).
\end{aligned}$$

Hence, $\delta_{h+1,D}(x)$ is increasing for $0 \le x < \overline{x}_{h,D}$ and decreasing for $x > \overline{x}_{h,D}$. Since $\delta_{h+1,D}(0) > 0$ and $\lim_{x\to\infty} \delta_{h+1,D}(x) = -\infty$ there exists a unique zero $\overline{x}_{h+1,D} > \overline{x}_{h,D}$ of $\delta_{h+1,D}(x)$ such that $\delta_{h+1,D}(x) > 0$ for $0 \le x < \overline{x}_{h+1,D}$ and $\delta_{h+1,D}(x) < 0$ for $x > \overline{x}_{h+1,D}$. □

Finally, we can provide a uniform upper bound for $\tilde{z}_h(x)$.

Lemma 5. *Let $\alpha > \alpha_0$ and $z_h(x)$ be defined by (9). Then we have*

$$\begin{aligned}
\tilde{z}_h(x) \le{}& \Gamma(\beta+1)\alpha^{\beta d} \sum_{k\ge 0} \frac{(\beta+1)^k}{k!}\left(\log\frac{1}{1-x}\right)^k \\
&+ \mathcal{O}\left(\frac{1}{1-(\alpha/\alpha_0)^{c'-1}}\alpha^{(c'-1)h} \sum_{k\ge 0}\frac{(c')^k}{k!}\left(\log\frac{1}{1-x}\right)^k\right)
\end{aligned}$$

uniformly for all real h and all $0 \le x < 1$.

Proof. Firstly we note that the inequality

$$1 - u\Phi(u) \le \Gamma(\beta+1)u^{-\beta} + \mathcal{O}\left((1-(\alpha/\alpha_0)^{c'-1})^{-1}u^{1-c'}\right)$$

is valid for all $u \ge 0$. Thus, by definition

$$z_h(x) \le \frac{\Gamma(\beta+1)}{(1-x)^{\beta+1}\alpha^{\beta h}} + \mathcal{O}\left(\frac{(1-(\alpha/\alpha_0)^{c'-1})^{-1}}{(1-x)^{c'}\alpha^{(c'-1)h}}\right),$$

which proves the lemma. □

5. An upper bound for $\mathbf{E}H'_n$.

For every integer $h \ge 0$ we will now use a special chosen

$$\alpha = \alpha_h := \alpha_0 + \frac{1}{h}.$$

This α is associated with

$$\beta = \beta_h = c' - 1 + \frac{d_0}{\sqrt{h}} + \mathcal{O}\left(\frac{1}{h}\right)$$

(for some $d_0 > 0$), the solution $> c' - 1$ of the equation $2\alpha^\beta = \beta + 1$. Hence, for every integer $h \ge 0$ there is a function $\Psi(y) = \Psi_h(y)$ (guaranteed by Lemma 2) satisfying

$$\Psi_h(y) = 1 - y^\beta + \mathcal{O}\left(hy^{c'-1}\right) \tag{13}$$

as $y \to \infty$ and and its Laplace transform $\Phi(u) = \Phi_h(u)$ with

$$\Phi_h(u) = 1 - \Gamma(\beta+1)u^{-\beta} + \mathcal{O}\left(hu^{c'-1}\right) \tag{14}$$

as $u \to 0+$ and $\mathcal{O}$-constants which are uniform in h. Thus, (13) and (14) are only significant for

$$y \ge e^{d_1\sqrt{h}\,\log h}$$

resp. for

$$u < e^{-d_2\sqrt{h}\,\log h},$$

where $d_1, d_2 > 0$ are appropriately chosen constants.

The key lemma for the proof of an lower bound of $z_h(x)$ is the following one.

Lemma 6. *For every real number d and for every integer $h \ge 0$ we have*

$$\tilde z_{h+d}(x) - z_h(x) \le \Gamma(\beta_h+1)\left(\frac{2}{\beta_h+1}\right)^{h+d}\sum_{k\ge h}\frac{(\beta_h+1)^k}{k!}\left(\log\frac{1}{1-x}\right)^k$$
$$+\mathcal{O}\left(h\left(\frac{2}{c'}\right)^h \alpha_h^{(1-c')d}\sum_{k\ge h}\frac{(c')^k}{k!}\left(\log\frac{1}{1-x}\right)^k\right)$$

for all $0 \le x < 1$.

Proof. For $h = 0$ this inequality is just Lemma 5. Now we can proceed by induction. We only have to observe that $\tilde z_{h+d}(0) < 1 = z_h(0)$ and that

$$\tilde z'_{h+1+d}(x) - z'_{h+1}(x) = \left(\frac{2}{1-x} - z_{h+d}(x) - z_h(x)\right)(\tilde z_{h+d}(x) - z_h(x))$$

and that

$$\frac{2}{1-x} - z_{h+d}(x) - z_h(x) \le \frac{2}{1-x}.$$

(The proof is quite similar to that of Lemma 1). □

The next lemma uses this estimate to get a lower bound for $z_h(x)$ at a specific value.

Lemma 7. *There exists a constant $D_0 > 0$ such that*

$$z_h\left(1 - e^{-h/(\beta_h+1)}\right) \ge \frac{\Gamma(c')}{3} e^{h/(\beta_h+1)}\left(\frac{\beta_h+1}{2}\right)^{D_0\sqrt{h}\,\log h}.$$

Proof. We use Lemma 6 with $x = 1 - e^{-h/(\beta_h+1)}$ and

$$d = -h\left(1 - \frac{1}{(\beta_h+1)\log\alpha_h}\right) - D_0\sqrt{h}\log h,$$

where $D_0 > 0$ will be chosen in the sequel. We directly get (after some algebra)

$$\left(\frac{2}{\beta_h+1}\right)^{h+d}\sum_{k\ge h}\frac{(\beta_h+1)^k}{k!}\left(\log\frac{1}{1-x}\right)^k$$
$$\sim \frac{1}{2}e^{h/(\beta_h+1)}\left(\frac{\beta_h+1}{2}\right)^{D_0\sqrt{h}\,\log h}$$

and similarly

$$h\left(\frac{2}{c'}\right)^h \alpha_h^{(1-c')d}\sum_{k\ge h}\frac{(c')^k}{k!}\left(\log\frac{1}{1-x}\right)^k$$
$$\ll e^{h/(\beta_h+1)} h\left(\frac{\beta_h+1}{2}\right)^{\frac{c'-1}{\beta_h}(D_0\sqrt{h}\,\log h)}.$$

Since

$$\frac{c'-1}{\beta_h} \ge 1 + \frac{d_4}{\sqrt{h}}$$

(for some constant $c_4 > 0$) it follows that there exists $D_0 > 0$ such that

$$h\left(\frac{\beta_h+1}{2}\right)^{\frac{c'-1}{\beta_h}(D_0\sqrt{h}\,\log h)} = o\left(\left(\frac{\beta_h+1}{2}\right)^{D_0\sqrt{h}\,\log h}\right)$$

as $h \to \infty$.

In a similar way we can discuss $\tilde z_{h+d}(x)$. By definition (and after some algebra) we get

$$\begin{aligned}\tilde z_{h+d}(x) &= \Gamma(\beta_h+1)e^h\left(\frac{2}{\beta_h+1}\right)^{h+d} + \mathcal{O}\left(he^{hc'/(\beta_h+1)}\alpha_h^{(1-c')(h+d)}\right)\\ &= \Gamma(\beta_h+1)e^{h/(\beta_h+1)}\left(\frac{\beta_h+1}{2}\right)^{D_0\sqrt{h}\,\log h}(1+o(1))\end{aligned}$$

if D_0 is sufficiently large. Consequently

$$\begin{aligned}\tilde z_h(x) &\ge \Gamma(\beta_h+1)\left(1-\frac{1}{2}\right)e^{h/(\beta_h+1)}\left(\frac{\beta_h+1}{2}\right)^{D_0\sqrt{h}\,\log h}(1+o(1))\\ &\ge \frac{\Gamma(c')}{3}e^{h/(\beta_h+1)}\left(\frac{\beta_h+1}{2}\right)^{D_0\sqrt{h}\,\log h}\end{aligned}$$

as proposed. □

Corollary . *Set*

$$h' = \frac{h}{(\beta_h+1)\log\alpha_h} - D_0\sqrt{h}\,\log h - 1.$$

Then we have

$$z_h(x) \ge \tilde z_{h'}(x)$$

for $x \ge 1 - e^{-h/(\beta_h+1)}$.

Proof. As in the proof of Lemma 7 it follows that

$$\begin{aligned}\tilde z_{h'}\left(1-e^{-h/(\beta_h+1)}\right) &= \frac{\beta_h+1}{2}\Gamma(\beta_h+1)e^{h/(\beta_h+1)}\left(\frac{\beta_h+1}{2}\right)^{D_0\sqrt{h}\,\log h}(1+o(1))\\ &\le \frac{\Gamma(\beta_h+1)}{3}e^{h/(\beta_h+1)}\left(\frac{\beta_h+1}{2}\right)^{D_0\sqrt{h}\,\log h}.\end{aligned}$$

Thus,

$$\tilde z_{h'}\left(1-e^{-h/(\beta_h+1)}\right) \le z_h\left(1-e^{-h/(\beta_h+1)}\right)$$

and consequently, by Lemma 4

$$\tilde z_{h'}(x) \le z_h(x)$$

for $x \ge 1 - e^{-h/(\beta_h+1)}$. □

Now we can complete the proof of Theorem 2. For given $n \ge 1$ choose a constant $D_1 > 0$ such that

$$h := [c'\log n + D_1(\log n)^{1/2}\log\log n]$$

is large enough that

$$u\Phi_h(u) \le \frac{1}{4}$$

for

$$u = \frac{1}{n}\alpha_h^{h/((\beta_h+1)\log\alpha_h) - D_0\sqrt{h}\,\log h}.$$

(Here we use 3. of the Corollary of Lemma 2.) Hence it follows that for $h' = h/((\beta_h+1)\log\alpha_h) - D_0\sqrt{h}\,\log h$

$$z_h(1-n^{-1}) \ge z_{h'}(1-n^{-1}) = n(1-u\Phi_h(u)) \ge \frac{3}{4}n.$$

Thus, be Theorem 1 we get

$$\mathbf{E}H_n' \le c' \log n + D_1 (\log n)^{1/2} \log\log n.$$

This completes the proof of Theorem 2.

References

[1] J. D. Biggins, *How fast does a general branching random walk spread ?*, in: IMA Vol. Math. Appl. **84** (1997), 19–39.

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

[3] L. Devroye and B. Reed, *On the variance of the height of random binary search trees*, SIAM J. Comput. **24** (1995), 1157–1162.

[4] L. Devroye and B. Reed, manuscript.

[5] M. Drmota, *An Analytic Approach to the Height of Binary Search Trees*, Algorithmica, to appear.

[6] M. Drmota, *The Variance of the Height of Binary Search Trees*, Theoret. Comput. Sci., submitted.

[7] M. Drmota, *An Analytic Approach to the Height of Binary Search Trees. II*, manuscript.

[8] H. M. Mahmoud, *Evolution of Random Search Trees*, John Wiley & Sons, New York, 1992.

[9] B. Reed, *How tall is a tree*, preprint.