A Note on the Height of Binary Search Trees

LUC DEVROYE

McGill University, Montreal, Canada

Abstract. Let H_n be the height of a binary search tree with n nodes constructed by standard insertions from a random permutation of $1, \ldots, n$. It is shown that $H_n/\log n \to c = 4.31107\ldots$ in probability as $n \to \infty$, where c is the unique solution of $c \log((2e)/c) = 1$, $c \geq 2$. Also, for all $p > 0$, $\lim_{n\to\infty} E(H_n^p)/\log^p n = c^p$. Finally, it is proved that $S_n/\log n \to c^* = 0.3733\ldots$, in probability, where c^* is defined by $c \log((2e)/c) = 1$, $c \leq 1$, and S_n is the saturation level of the same tree, that is, the number of full levels in the tree.

Categories and Subject Descriptors: E.1 **[Data]**: Data Structures—*trees*; F.2.2 **[Analysis of Algorithms and Problem Complexity]**: Nonnumerical Algorithms and Problems—*sorting and searching*

General Terms: Algorithms, Theory, Verification

Additional Key Words and Phrases: Analysis of algorithms, binary search tree, branching random walk, data structures, expected height of a tree

1. *Introduction*

The *height* of *random binary trees* (with various definitions of "randomness") has been analyzed by a variety of authors (see, e.g., the recent work of Flajolet and Odlyzko [5] and the references found there). The following types of binary trees have received special attention: *tries* (digital search trees) [4, 9, 17], *planted plane trees* [3], *planar trees* [7], *labeled nonplanar trees* [12], and *rooted free trees* [16].

In this note, we consider *binary search trees* with the usual randomization; that is, the binary search tree is constructed in the standard fashion (n consecutive insertions) from a random permutation of $\{1, \ldots, n\}$, where each permutation is equally likely. The tree has n nodes, and its height H_n is the number of nonempty levels minus one. We have $H_0 = H_1 = 0$, and $H_n \geq \text{int}(\log_2 n)$ where $\text{int}(\cdot)$ is the integer part of $(\cdot)$. The random variable H_n has been studied by a variety of authors, but the first nontrivial result goes back to Robson [13] who proved that

$$3.6 \log n + o(\log n) \leq E(H_n) \leq 4.31107 \cdots \log n + o(\log n).$$

Robson also showed that the limit of $E(H_n)/\log n$ exists [14], and G. Gonnet (personal communication) indicated that this limit is the constant 4.31107 . . . . The purpose of this note is to prove this result and to indicate the power of some probability theoretical tools in the analysis of algorithms.

The author's research was sponsored in part by NSERC grant A3456 in Canada and by the Mathematical Sciences Research Center at Australian National University, Canberra.

Author's address: School of Computer Science, McGill University, 805 Sherbrooke Street West, Montreal, Canada H3A 2K6.

2. *The Fundamental Inequalities*

In this section, we establish a vital link, via not-too-crude probability inequalities, with *trees of random variables* and *branching random walks.* We consider in particular a complete binary tree T_k with k full levels of edges (the total number of edges is thus $2^1 + 2^2 + \cdots + 2^k = 2^{k+1} - 2$). We use the symbol p for a path from root to leaf (there are 2^k such paths). With each edge i we associate a random variable X_i in the following manner: Consider all edges levelwise and from left to right, and identify $X_1, X_2, \ldots, X_{2^{k+1}-2}$ with U_1, $1 - U_1$, U_2, $1 - U_2, \ldots, U_{2^k-1}$, $1 - U_{2^k-1}$, where $U_1, \ldots, U_{2^k-1}$ are independent uniform [0, 1] random variables. Define

$$Z_k = \begin{cases} 1, & k = 0, \\ \max\limits_p \prod\limits_{i \in p} X_i, & k > 0. \end{cases}$$

LEMMA 2.1. *Let $n \geq 1$, $k \geq 0$ be integers. Then*

$$P\left(Z_k \geq \frac{1 + k}{n}\right) \leq P(H_n \geq k) \leq P\left(Z_k \geq \frac{1}{n}\right).$$

PROOF. We proceed by induction on k. For $k = 0$, we have equality in both the upper and the lower bound. We assume that the inequalities hold up to $k - 1$ and for all n. First, we show that the upper bound remains valid for k and all $n \geq 1$.

Let us introduce the following notation: U is a uniform [0, 1] random variable, independent of all the other random variables that we shall mention. Let N, N^* be the number of nodes in the left and right subtrees of the root (i.e., $N + N^* = n - 1$), and let Z_{k-1}, and Z^*_{k-1} be defined as above and independent of each other. These two random variables will be associated with the subtrees of the root. If H_N and $H^*_{N^*}$ are the heights of the left and right subtrees, respectively, we have

$$P(H_n \geq k) = P(H_N \geq k - 1, \text{ or } H^*_{N^*} \geq k - 1).$$

But, by our induction hypothesis and various independencies,

$$\begin{aligned} P(H_N \geq k - 1 \text{ or } H^*_{N^*} \geq k - 1 \mid N) &= 1 - P(H_N < k - 1, H^*_{N^*} < k - 1 \mid N) \\ &= 1 - P(H_N < k - 1 \mid N)P(H^*_{N^*} < k - 1 \mid N) \\ &\leq 1 - P\left(Z_{k-1} < \frac{1}{N} \,\middle|\, N\right)P\left(Z^*_{k-1} < \frac{1}{N^*} \,\middle|\, N\right) \\ &= 1 - P(\max(NZ_{k-1}, N^*Z^*_{k-1}) < 1 \mid N). \end{aligned}$$

We note that (N, N^*) is distributed as $(\text{int}(nU), \text{int}(n(1 - U)))$, and that, with this embedding, $N \leq nU$, $N^* \leq n(1 - U)$. Thus,

$$\max(NZ_{k-1}, N^*Z^*_{k-1}) \leq n \max(UZ_{k-1}, (1 - U)Z^*_{k-1}) = nZ_k.$$

We conclude that $P(H_n \geq k) \leq P(Z_k \geq 1/n)$.

For the lower bound, we obtain, as before, the intermediate result

$$P(H_n \geq k) \geq E(1 - P(\max(NZ_{k-1}, N^*Z^*_{k-1}) < 1 + (k - 1) \mid N)).$$

Applying the same embedding procedure, we observe that $N \geq nU - 1$, and that $N^* \geq n(1 - U) - 1$. Thus,

$$\begin{aligned} \max(NZ_{k-1}, N^*Z^*_{k-1}) &\geq \max(nUZ_{k-1} - Z_{k-1}, n(1 - U)Z^*_{k-1} - Z^*_{k-1}) \\ &\geq \max(nUZ_{k-1}, n(1 - U)Z^*_{k-1}) - 1 \\ &= nZ_k - 1. \end{aligned}$$

Clearly,

$$P(H_n \geq k) \geq E(1 - P(nZ_k - 1 < k \mid N)) = P(nZ_k \geq 1 + k),$$

and we are done. □

Remark. What we have done in the proof of Lemma 2.1 amounts to an association between the sizes of the subtree of a node and the spacings defined by a uniform splitting of [0, 1]. Thus, $(U, 1 - U)$ are the sizes of the intervals after splitting, and they contain, roughly speaking, nU and $n(1 - U)$ "nodes." Each of the subintervals is further subdivided, giving rise to a tree of a product of uniform random variables. Each of the intervals after several iterations "represents" a tree with approximately n times the length of the interval number of nodes. Lemma 2.1 now connects the height of a tree with n nodes with the size of the largest uniform spacing after k iterations of splitting.

3. *An Upper Bound*

The purpose of this section is to establish a firm upper bound for $P(H_n \geq k)$, in preparation of our main theorem in Section 5.

LEMMA 3.1. *For integer $k \geq max(1, log\ n)$, we have*

$$P(H_n \geq k) \leq \frac{1}{n}\left(\frac{2e\ log\ n}{k}\right)^k.$$

For $\epsilon > 0$ and $c \geq 2$, we have

$$P(H_n \geq (c + \epsilon)log\ n) \leq n^{(c+\epsilon)log(2e/(c+\epsilon))-1}.$$

In particular, the right-hand side of this inequality tends to 0 for all $\epsilon > 0$ when c is the solution of

$$c\ log\left(\frac{2e}{c}\right) = 1, \qquad c \geq 2.$$

Finally, note that for all $p \geq 1$, $X_n = H_n^p/(log\ n)^p$ is uniformly integrable, that is,

$$\lim_{t\to\infty} \sup_{n\geq 1} E(|X_n| I_{[|X_n|\geq t]}) = 0.$$

PROOF. We repeatedly use the fact that a uniform [0, 1] random variable is distributed as e^{-Y}, where Y is an exponential random variable (i.e., Y has density e^{-y}, $y > 0$), and that the sum of k independent exponential random variables has a gamma (k) density, $y^{k-1}e^{-y}/(k - 1)!$.

Now,

$$\begin{aligned}
P(H_n \geq k) &\leq P\left(Z_k \geq \frac{1}{n}\right) && \text{(Lemma 2.1)}\\
&\leq 2^k P\left(\prod_{i=1}^{k} U_i \geq \frac{1}{n}\right) && \text{(by Bonferroni's inequality; here } U_1, \dots, U_k \text{ are independent uniform [0, 1] random variables)}\\
&= 2^k P(ne^{-G_k} \geq 1) && \text{(where } G_k \text{ is a gamma (}k\text{) random variable)}\\
&\leq 2^k E(n^t e^{-tG_k}) && \text{(by Chebyshev's inequality, } t > 0)\\
&= 2^k n^t (t + 1)^{-k} &&\\
&= \frac{1}{n}\left(\frac{2e \log n}{k}\right)^k && \text{(upon taking } t + 1 = k/\log n, \text{ a choice that minimizes the bound).}
\end{aligned}$$

This concludes the proof of the first inequality of Lemma 3.1. For the second inequality, we note first that $((2e \log n)/k)^k$ decreases in k for $k \geq 2 \log n$. Thus, since $P(H_n \geq (c+\epsilon)\log n) = P(H_n \geq \text{ceil}((c+\epsilon)\log n))$ where ceil$(\cdot)$ is the ceiling function, it is allowed to replace k in the first inequality by $(c+\epsilon)\log n$, which gives us our second inequality.

For the last statement, we rewrite the upper bound in the second inequality as follows:

$$n^{(c+\epsilon)\log(2e/(c+\epsilon))-c\log(2e/c)}.$$

Now, the function $u \log(2e/u)$ has derivative $\log(2e/u) - 1 = \log(2/u)$, which decreases from 0 (at $u = 2$) to $\log(2/c)$ (at $u = c$) to $-\infty$ (as $u \to \infty$), and this in a monotone manner. Therefore, $(c+\epsilon)\log(2e/(c+\epsilon) - c\log(2e/c) \leq \epsilon \log(2/c)$, and

$$P(H_n \geq (c+\epsilon)\log n) \leq n^{\epsilon\log(2/c)}.$$

The uniform integrability is quickly verified. □

4. *A Lower Bound*

The upper bound of Section 3 required very little in terms of technical machinery. In this section, we prove Lemma 4.1, and rely very heavily upon profound results from the theory of branching random walks.

LEMMA 4.1. *Let $\epsilon > 0$ be arbitrary, and let $c = 4.31107\ldots$ be the unique solution of $c\ log(2e/c) = 1$. Then*

$$\lim_{n\to\infty} P(H_n < (c-\epsilon) log\ n) = 0.$$

PROOF. Lemma 4.1 follows directly from Example 4.1 and Lemma 4.2, which are stated below.

Consider once again a complete binary tree T_k with edge-associated random variables X_i, $1 \leq i \leq 2^{k+1} - 2$, where now the X_i's are independent and identically distributed (and not necessarily uniformly distributed on [0, 1]). We define the quantity

$$\bar{X}_k = \max_p \sum_{i\in p} X_i,$$

where the maximum is taken over all 2^k paths p in T_k.

THEOREM 4.1. ([1, 2]; for special cases, see [6] and [8]). *Assume that*

$$m(\theta) = 2E(exp(\theta X_1)) < \infty \qquad for\ some \quad \theta > 0.$$

Define

$$\mu(t) = inf(exp(-t\theta)m(\theta):\theta > 0),$$
$$\gamma = inf(t:\mu(t) > 1).$$

Then

$$\frac{\bar{X}_k}{k} \to \gamma \qquad almost\ surely\ as \quad k \to \infty.$$

Theorem 4.1 is in fact only a special form, adapted for our cause, of the powerful Biggins–Kingman–Hammersley theorem. We illustrate its use in Example 4.1.

Example 4.1. Let X_1 be minus an exponential random variable. Then, in the notation of Theorem 4.1,

$$m(\theta) = 2E(\exp(\theta X_1)) = 2\int_0^\infty \exp(-\theta x)\exp(-x)\,dx = \frac{2}{\theta+1}, \qquad \theta > 0.$$

Also,

$$\mu(t) = \inf\left(\exp(-t\theta)\cdot\frac{2}{\theta+1} : \theta > 0\right).$$

For $t < -1$, the infimum is reached at $\theta = 0$ and equals 2. For $t > 0$, the infimum is obviously zero, and for $-1 \le t \le 0$, the infimum is reached for $\theta = -1 - (1/t)$, and its value there is

$$\mu(t) = -2t\exp(t+1).$$

Thus, γ is well defined (it always is) and can be found by solving the equation

$$-2t\exp(t+1) = 1, \qquad -1 \le t \le 0.$$

We note here that $\gamma = 1/c$, where c is the constant of Lemmas 3.1 and 4.1. This concludes the example. □

Let us return now to our binary search tree, and in particular to the random variable Z_k defined in Section 2. The only problem that bothers us in the definition of Z_k is the awkward dependence between the X_i's (consecutive X_i's are distributed as $(U, 1-U)$ where U is a uniform [0, 1] random variable). To be able to apply Theorem 4.1, we need at the very least a tree of independent random variables. For example, it would be nice if we could show that Z_k is close to Z_k^*, where Z_k^* is formally defined as Z_k, except that the defining X_i sequence consists of independent uniform [0, 1] random variables. In Lemma 4.2, we show that Z_k is stochastically greater than Z_k^*.

LEMMA 4.2. *For all real y, and all integers k,*

$$P(Z_k \ge y) \ge P(Z_k^* \ge y).$$

PROOF. Once again, we proceed by induction on k. For $k = 1$, we must prove that $\max(U, 1-U)$ is stochastically greater than $\max(U_1, U_2)$ where U, U_1, U_2 are independent uniform [0, 1] random variables. But this follows from the fact that for $\frac{1}{2} \le y \le 1$,

$$P(\max(U, 1-U) \ge y) = 2(1-y) \ge 1 - y^2 = P(\max(U_1, U_2) \ge y).$$

We assume next that our Lemma is true up to $k-1$ for fixed $k \ge 2$. The proof is complete if we can show that for all y,

$$P(\max(UZ_{k-1}(1), (1-U)Z_{k-1}(2)) < y) \le P(\max(U_1Z_{k-1}^*(1), U_2Z_{k-1}^*(2)) < y),$$

where $Z_{k-1}(1)$, $Z_{k-1}(2)$ are independent copies of Z_{k-1}, and $Z_{k-1}^*(1)$, $Z_{k-1}^*(2)$ are independent copies of Z_{k-1}^*, and all random variables involved are independent of each other. First, we have by our induction hypothesis,

$$\begin{aligned}&P(\max(UZ_{k-1}(1), (1-U)Z_{k-1}(2)) < y)\\ &\quad\le P(\max(UZ_{k-1}^*(1), (1-U)Z_{k-1}^*(2)) < y), \qquad y \in R.\end{aligned}$$

If F_{k-1} is the σ-algebra generated by $Z^*_{k-1}(1)$, $Z^*_{k-1}(2)$, then

$$\begin{aligned}
&P(\max(UZ^*_{k-1}(1), (1 - U)Z^*_{k-1}(2)) < y \mid F_{k-1}) \\
&= P\left(1 - \frac{y}{Z^*_{k-1}(2)} < U \leq \frac{y}{Z^*_{k-1}(1)} \;\middle|\; F_{k-1}\right) \\
&\leq P\left(U < \frac{y}{Z^*_{k-1}(1)} \;\middle|\; F_{k-1}\right) P\left(1 - U < \frac{y}{Z^*_{k-1}(2)} \;\middle|\; F_{k-1}\right) \\
&= P(U_1 Z^*_{k-1}(1) < y,\ U_2 Z^*_{k-1}(2) < y \mid F_{k-1}).
\end{aligned}$$

In the inequality, we used the fact that U is independent of F_{k-1} and uniformly distributed on $[0, 1]$.

Taking expectations on left- and right-hand sides to get rid of the conditioning gives us our lemma. □

But now we are ready to apply the Biggins–Kingman–Hammersley theorem, because, by the connection between uniform and exponential random variables, Z^*_k is distributed as $\exp(\bar{X}_k)$ where $\bar{X}_k$ is the random variable of Theorem 4.1 with as underlying distribution the flipped exponential distribution discussed in Example 4.1. Therefore, we know that

$$\frac{1}{k}\log(Z^*_k) \to -\frac{1}{c} \quad \text{almost surely.}$$

Via Lemma 4.2, we can now argue directly—for example, Lemma 4.1 follows from the following chain of inequalities:

$$\begin{aligned}
P(H_n \geq (c-\epsilon)\log n) &= P(H_n \geq k)(k = \text{ceil}((c-\epsilon)\log n)) \\
&\geq P\left(Z_k \geq \frac{1 + (c-\epsilon)\log n}{n}\right) && \text{(Lemma 2.1)} \\
&\geq P\left(Z^*_k \geq \frac{1 + (c-\epsilon)\log n}{n}\right) && \text{(Lemma 4.2)} \\
&= P\left(\frac{1}{k}\log(Z^*_k) \geq \frac{1}{k}\log\left(\frac{1 + (c-\epsilon)\log n}{n}\right)\right) \\
&\to 1 \quad \text{as} \quad k \to \infty.
\end{aligned}$$

The last line follows from the fact that

$$\left(\frac{1}{k}\right)\log\left(\frac{1 + (c-\epsilon)\log n}{n}\right) \sim -\frac{1}{c-\epsilon} < -\frac{1}{c}.$$

5. *The Main Result*

THEOREM 5.1. *Let $c = 4.31107\ldots$ be the unique solution of the equation*

$$c\ log\left(\frac{2e}{c}\right) = 1, \quad c \geq 2.$$

Then:

A. $H_n/log\ n \to c$ in probability as $n \to \infty$.
B. For all $p > 0$, $E(H^p_n) \sim E^p(H_n) \sim (c\ log\ n)^p$.

Before we give the technical details we should observe that in many a situation, a statement of type A is more valuable than one of type B. Both A and B confirm in our case that H_n is indeed very close to $c \log n$. Robson [15] has in fact obtained evidence that $E(\,|H_n - E(H_n)|\,)$ is uniformly bounded in n. From work related to Theorem 4.1, one can infer that $H_n - c \log n - c^* \log\log n$ tends to a limit distribution for some constant c^*. The arguments given in this short note are not fine enough to compute the coefficient of the log log n term.

Mahmoud and Pittel [10] have shown that $\limsup H_n/\log n \le c$ almost surely, and Pittel [11] has shown that $H_n/\log n \to \alpha$ almost surely for some constant α. In view of Theorem 5.1, this constant is $\alpha = c$.

PROOF. Statement A follows by combining Lemmas 3.1 and 4.1. Furthermore, $E(H_n^p/(\log n)^p) \to c^p$ by part A and the uniform integrability of $\{H_n^p/(\log n)^p\}$ (Lemma 3.1). This concludes the proof of part B. □

6. *The Saturation Level*

We now illustrate the power of Theorem 4.1 on a second (albeit less important) characteristic of the random binary search tree: the *saturation level* S_n, that is, the number of full levels of nodes in the tree. Thus, $S_n = j$ if and only if levels 1 (root level) through j are full, and level $j + 1$ is not full (it has less than 2^j nodes). By convention, we set $S_0 = 0$. Note that for $n \ge 1$, $S_n \ge 1$. The development parallels the development for the height, and we proceed with fewer explanations.

LEMMA 6.1 (FUNDAMENTAL INEQUALITIES FOR S_n). *For all* $n \ge 1$, $k \ge 1$,

$$P\left(Z_{k-1} > \frac{1 + (k-1)}{n}\right) \le P(S_n \ge k) \le P\left(Z_{k-1} > \frac{1}{n}\right),$$

where

$$Z_k = \begin{cases} 1, & k = 0, \\ \min\limits_{p} \prod\limits_{i \in p} X_i, & k > 0, \end{cases}$$

and $\{X_i,\ 1 \le i \le 2^{k+1} - 2\}$ *are random variables defined as in Section* 2. *The minimum is taken over all paths p in the tree* T_k *of Section* 2.

PROOF. We inherit the notation from the proof of Lemma 2.1. Note that the inequalities are valid for $k = 1$. We mimic the induction proof of Lemma 2.1, and note that for $k > 1$,

$$\begin{aligned}
P(S_n < k) &= P(S_N < k - 1 \text{ or } S_{N^*}^* < k - 1) \\
&= 1 - P(S_N \ge k - 1, S_{N^*}^* \ge k - 1) \\
&= 1 - E(P(S_N \ge k - 1 \mid N) P(S_{N^*}^* \ge k - 1 \mid N)) \\
&\ge 1 - E\left(P\left(Z_{k-2} > \frac{1}{N} \,\middle|\, N\right) P\left(Z_{k-2}^* > \frac{1}{N^*} \,\middle|\, N\right)\right) && \text{(by our induction hypothesis)} \\
&= 1 - P(\min(NZ_{k-2}, N^*Z_{k-2}^*) > 1) && \text{(by independence)} \\
&\ge 1 - P(nZ_{k-1} > 1). && \left\{\begin{array}{l}\text{(since} \\ \min(NZ_{k-1}, N^*Z_{k-1}^*) \\ \le nZ_{k-1}).\end{array}\right.
\end{aligned}$$

This proves the upper bound of Lemma 6.1. The lower bound is obtained as in Lemma 2.1, and is based on the observation that $\min(NZ_{k-2}, N^*Z_{k-2}^*) \ge nZ_{k-1} - 1$, $k > 1$. □

LEMMA 6.2 (A LOWER BOUND). *For $k \geq 1$, $\log n > k + \log(k + 1)$,*

$$P(S_n < k + 1) \leq \left(\frac{k+1}{n}\right)\left(\frac{2e}{k}\log\left(\frac{n}{k+1}\right)\right)^k.$$

If $c^* = 0.3733\ldots$ is the unique solution of $c\log((2e)/c) = 1$, $c < 1$, then $P(S_n < (c^* - \epsilon)\log n) = o(1)$, all $\epsilon > 0$.

PROOF. The following chain of inequalities is valid for $0 < t < 1$:

$$\begin{aligned}
P(S_n < k+1) &\leq P\left(Z_k < \frac{k+1}{n}\right) && \text{(Lemma 6.1, } k \geq 1\text{)}\\
&\leq 2^k P\left(\prod_{i=1}^{k} U_i < \frac{k+1}{n}\right)\\
&= 2^k P(n\exp(-G_k) < k+1) && \text{(notation of Lemma 3.1)}\\
&\leq 2^k\left(\frac{k+1}{n}\right)^t E(e^{tG}k) && \text{(Chebyshev's inequality)}\\
&= 2^k\left(\frac{k+1}{n}\right)^t\left(\frac{1}{1-t}\right)^k\\
&= \left(\frac{k+1}{n}\right)\left(\frac{2e}{k}\log\left(\frac{n}{k+1}\right)\right)^k && \left(\text{upon taking } \frac{1}{1-t} = \frac{1}{k}\log\left(\frac{n}{k+1}\right),\right.\\
& && \text{a choice that minimizes the}\\
& && \text{bound).}
\end{aligned}$$

We now take $k = \operatorname{int}(s\log n)$, where $s < 1$, $s\log((2e)/s) < 1$. Then, by the previous inequality,

$$\begin{aligned}
P(S_n < k + 1) &\leq \frac{s\log n + 1}{n}\left(\frac{2e\log n}{\operatorname{int}(s\log n)}\right)^{\operatorname{int}(s\log n)}\\
&= O\left(\frac{s\log n}{n}\left(\frac{2e}{s}\right)^{s\log n}\right)\\
&= O(s\log n \cdot n^{s\log((2e)/s)-1})\\
&= o(1).
\end{aligned}$$

Lemma 6.2 is proved, since the conditions on s are satisfied by all $s < c^*$: Indeed, the derivative of $s\log((2e)/s)$ is $\log(2/s)$, and this is positive for $s < 2$. Thus, $s\log((2e)/s) < 1$ for all $s < c^*$. □

LEMMA 6.3 (A STOCHASTIC INEQUALITY). *Let Z_k be the random variable of Lemma 6.1, and let Z_k^* be formally defined in the same way but with independent X_i's, all uniformly distributed on $[0, 1]$. Then, for all $y \in R$,*

$$P(Z_k \geq y) \leq P(Z_k^* \geq y),$$

that is, Z_k is stochastically smaller than Z_k^*.

PROOF. We proceed by induction on k. Let U, U_1, U_2 be independent uniform $[0, 1]$ random variables, and let $0 \le y \le \frac{1}{2}$. Then

$$P(Z_1 \ge y) = P(\min(U, 1 - U) \ge y) = 2\left(\frac{1}{2} - y\right) = 1 - 2y$$
$$\le (1 - y)^2 = P(\min(U_1, U_2) \ge y) = P(Z_1^* \ge y),$$

which proves the inequality for $k = 1$. Assume that we have shown the inequality for all i up to $k - 1$, where $k \ge 2$ is fixed. Then, using self-evident notation as in the proof of Lemma 4.2,

$$\begin{aligned} P(Z_k \ge y) &= P(\min(UZ_{k-1}(1), (1 - U)Z_{k-1}(2)) \ge y) \\ &\le P(\min(UZ_{k-1}^*(1), (1 - U)Z_{k-1}^*(2)) \ge y) \qquad \text{(induction hypothesis)} \\ &= P\left(\frac{y}{Z_{k-1}^*(1)} \le U \le 1 - \frac{y}{Z_{k-1}^*(2)}\right) \\ &\le P(U_1 Z_{k-1}^*(1) \ge y,\ U_2 Z_{k-1}^*(2) \ge y) \qquad \text{(as in Lemma 4.2)} \\ &= P(Z_k^* \ge y). \end{aligned}$$

This concludes the proof of Lemma 6.3. □

THEOREM 6.1. *Let* $c^* = 0.3733\ldots$ *be the unique solution of*

$$c\, log\left(\frac{2e}{c}\right) = 1, \qquad 0 < c < 1.$$

Then $S_n/log\, n \longrightarrow c^*$ *in probability as* $n \longrightarrow \infty$.

PROOF. For $\epsilon > 0$, we know that $P(S_n < (c^* - \epsilon)\log n) = o(1)$ (Lemma 6.2). Also, writing k for ceil$((c^* + \epsilon)\log n)$, we have

$$\begin{aligned} P(S_n \ge (c^* + \epsilon)\log n) &= P(S_n \ge k) \\ &\le P\left(Z_{k-1} \ge \frac{1}{n}\right) \qquad \text{(Lemma 6.1)} \\ &\le P\left(Z_{k-1}^* > \frac{1}{n}\right) \qquad \text{(Lemma 6.3)} \\ &= P\left(\left(\frac{1}{k-1}\right)\log(Z_{k-1}^*) > -\frac{\log n}{k-1}\right) \\ &= o(1), \end{aligned}$$

if the in probability limit of $(1/k)\log(Z_k^*)$ is less than $-1/(c^* + \epsilon)$.

Thus, we are done if we can show that

$$\left(\frac{1}{k}\right)\log(Z_k^*) \longrightarrow -\frac{1}{c^*} \quad \text{in probability as} \quad k \longrightarrow \infty.$$

Let us take $\bar{X}_k$ as in Theorem 4.1, where in the definition of $\bar{X}_k$, the X_i's are independent exponential random variables. It is clear that we can write

$$Z_k^* = \exp(-\bar{X}_k),$$

so that it suffices to show that $\bar{X}_k/k \longrightarrow 1/c^*$ in probability. This follows with a little work from the Biggins–Kingman–Hammersley theorem: Indeed, in the notation of Theorem 4.1, we have

$$m(\theta) = 2E(\exp(\theta X_1)) = 2\int_0^\infty \exp(\theta x)\exp(-x)\,dx = \frac{2}{1-\theta}, \qquad 0 \le \theta < 1.$$

Also,

$$\mu(t) = \inf\left(\exp(-\theta t)\cdot\frac{2}{1-\theta} : 0 \le \theta < 1\right) = \begin{cases} 2te^{1-t}, & t \ge 1, \\ 2, & t \le 1. \end{cases}$$

Finally, $\gamma = \inf(t{:}\mu(t) > 1)$ is obtainable as the solution of $2te^{1-t} = 1$, $t \ge 1$. It is easy to see that $\gamma = 1/c^*$. This concludes the proof of Theorem 4.1. □

ACKNOWLEDGMENT. I would like to thank Alain Rouault and Mike Robson for helpful discussions, and the two referees for helpful suggestions.

REFERENCES

1. BIGGINS, J. D. The first and last-birth problems for a multitype age-dependent branching process. *Adv. Appl. Probab. 8* (1976), 446–459.
2. BIGGINS, J. D. Chernoff's theorem in the branching random walk. *J. Appl. Probab. 14* (1977), 630–636.
3. DE BRUIJN, N., KNUTH, D. E., AND RICE, O. The average height of planted plane trees. In *Graph Theory and Computing,* R.-C. Read, Ed. Academic Press, Orlando, Fla., 1972, pp. 15–22.
4. DEVROYE, L. A probabilistic analysis of the height of tries and of the complexity of triesort. *Acta Inf. 21* (1984), 229–237.
5. FLAJOLET, P. AND ODLYZKO, A. The average height of binary trees and other simple trees. *J. Comput. Syst. Sci. 25* (1982), 171–213.
6. HAMMERSLEY, J. M. Postulates for subadditive processes. *Ann. Probab. 2* (1974), 652–680.
7. KEMP, R. On the stack size of regularly distributed binary trees. In *Proceedings of the 6th ICALP Conference* (Udine, Italy). 1979.
8. KINGMAN, J. F. C. The first-birth problem for an age-dependent branching process. *Ann. Probab. 3* (1975), 790–801.
9. KNUTH, D. E. *The Art of Computer Programming: Fundamental Algorithms.* Addison-Wesley, Reading, Mass., 1968.
10. MAHMOUD, H., AND PITTEL, B. On the most probable shape of a binary search tree grown from a random permutation. *SIAM J. Algebraic Discrete Meth. 5* (1984), 69–81.
11. PITTEL, B. On growing random binary trees. *J. Math. Anal. Appl. 103* (1984), 461–480.
12. RENYI, A., SZEKERES, G. On the height of trees. *Aust. J. Math. 7* (1967), 497–507.
13. ROBSON, J. M. The height of binary search trees. *Aust. Comput. J. 11* (1979), 151–153.
14. ROBSON, J. M. The asymptotic behaviour of the height of binary search trees. *Aust. Comput. Sci. Commun.* (1982), p. 88 (Also: Tech. Rep. TR-CS-81-15, Dept. of Computer Science, Australian National Univ., Canberra, Australia, 1981).
15. ROBSON, J. M. On the height of binary search trees. Tech. Rep. Dept. of Computer Science, Australian National Univ. Canberra, 1983.
16. STEPANOV, V. E. On the distribution of the number of vertices in strata of a random tree. *Theor. Probab. Appl. 14* (1969), 65–78.
17. YAO, A. C. A note on the analysis of extendible hashing. *Inf. Process. Lett. 11* (1980), 84–86.

RECEIVED JANUARY 1984; REVISED AUGUST 1985; ACCEPTED SEPTEMBER 1985