A LIMIT THEORY FOR RANDOM SKIP LISTS

Luc Devroye
School of Computer Science
McGill University

Abstract. The skip list was introduced by Pugh in 1989 as a data structure for dictionary operations. Using a binary tree representation of skip lists, we obtain the limit law for the path lengths of the leaves in the skip list. We also show that the the height (maximal path length) of a skip list holding n elements is in probability asymptotic to $c \log_{1/p} n$, where c is the unique solution greater than one of the equation $\log(1-p) = \log(c-1) - \frac{c}{c-1} \log c$, and $p \in (0,1)$ is a design parameter of the skip list.

Keywords and phrases. Skip list, data structures, probabilistic analysis, weak convergence, height of a tree, branching processes.

CR Categories: 3.74, 5.25, 5.5.

Authors' address: School of Computer Science, McGill University, 3480 University Street, Montreal, Canada H3A 2K6. The authors' research was sponsored by NSERC Grant A3456 and FCAR Grant 90-ER-0291.

Introduction.

A skip list is as a fast probabilistic data structure for dictionary operations (insert, delete, search, member, sort) introduced by Pugh (1989). The object of this note is to analyze the expected time behavior of the data structure. We first describe a random tree that could be of interest in its own right. We will then see how it relates to the skip list.

The **basic random tree** holding n elements is defined as follows. Each element in the tree can be visualized as "living" at one of the grid points in the integer grid consisting of $\{(i,j) | \, 0 \le i \le n, \, j \ge 0\}$. Edges in the tree only run horizontally or vertically. The randomness is introduced by generating n i.i.d. geometric $(1-p)$ random variables $G_1, \ldots, G_n$, i.e. each G_i is distributed as G where

$$\mathbf{P}\{G = i\} = p^i(1-p), i \ge 0,$$

and $p \in (0,1)$ is a design parameter. The nodes alive in the random tree are of the form (i,j) with $1 \le i \le n$, $0 \le j \le G_i$, or $(0,j)$, with $0 \le j \le \max_{1 \le i \le n} G_i$. We say that node (i,j) lives at level j. First we introduce all the edges between nodes (i,j) and $(i,j+1)$ if both nodes exist in the structure. These vertical edges constitute all the "left" edges in a binary tree. The "right" edges are obtained by horizontally connecting (i,G_i) with (k,G_i), where $0 < k < i$, and k is the largest integer less than i (if it exists) with the property that $G_k \ge G_i$. If such a k does not exist, (i,G_i) is connected with $(0,G_i)$.

Put Figure 1 about here.

As shown in Figure 1, the G_i's can be considered as upright poles. Each top of a pole is connected horizontally to the left to the nearest pole that is at least as high. All horizontal connections are stopped, if need be, at a pole with first coordinate zero. In binary tree lingo, a horizontal edge corresponds to a father-right child relationship, the father being on the left (with smaller first coordinate), and a vertical edge represents a father-left child connection, the father having the larger second coordinate. The root of the tree is at position $(0,K_n)$. The unique path to the root from a node $(i,0)$ has length D_i, and hits the pole at coordinate zero for the first time at $(0,K_i)$. The path distance between $(i,0)$ and $(0,K_i)$ is called P_i, so that

$$D_i = P_i + (K_n - K_i) \ .$$

The quantities we will study in this note are the D_i's, and the height of the basic random tree,

$$H_n = \max_{1 \le i \le n} D_i \ .$$

In particular, we will obtain a limit law for D_i when i changes with n, and a law of large numbers for H_n. The latter result is obtained by constructing a particular branching process.

Skip lists.

In Pugh's skip list (1989), the nodes at $(i, 0)$, $1 \le i \le n$, represent elements to be stored in a data structure, which we shall call $x_1, \ldots, x_n$. These will be stored in ascending order: $x_1 < x_2 < \cdots < x_n$. All other nodes in the structure are auxiliary nodes holding only pointer information. All the poles are nothing but linked lists or doubly linked lists. Horizontally, we store a linked list for every level, which means that at every level, we connect all the nodes that exist at that level. The rightmost nodes at every level have nil pointers. This is where the original skip list deviates from the tree representation of Figure 1, as additional horizontal connections are needed. However, it turns out that all the skip list operations can be carried out equally efficiently if we had just stored all the horizontal pointers in the tree of Figure 1. With each vertical linked list of nodes (i, j), we associate the value of x_i. With the pole at zero, we associate the value $-\infty$. By "association", we mean that a pointer to the location of the value x_i is stored; hence a quick look-up reveals the value x_i, for any position (i, j) in the skip list.

When we search for element x_i, we go first to the root at $(0, K_n)$, which is specially marked as such. With the aid of comparisons with the values associated with a node and with its right child, we will either move to the left child or the right child. If a node has no right child, we automatically move to the left child. In this manner, we will move down the tree, following the unique path linking the root with $(i, 0)$.

Inserting an element is equally simple. It requires the generation by means of a random number generator of an independent geometric random variable. We will assume that we have at our disposal a source capable of producing an i.i.d. sequence of uniform $[0, 1]$ random variables. The operation insert is like the operation search described above, but along the way we will have to modify some pointers. Deletion is like undoing an insertion. Note that there is no choice as to how a deletion is carried out because a skip list is uniquely determined by its collection of geometric random variables.

The time taken by an operation involving x_i is proportional to D_i, the depth of node $(i, 0)$ in the basic random tree. Recall that i refers to the rank of the present element among the elements stored in the skip list, and not its time of insertion. To have uniform performance guarantees over all elements of all ranks, it is useful to know how H_n behaves.

Before proceeding with the analysis, it is useful to note that this structure does not impose any conditions on the data. All its good qualities are entirely due to the randomness artificially introduced. We also note that when the linked list for level 0 is emptied, then the elements are encountered in order, so constructing the skip list from scratch by consecutive insertions among other things sorts the data. The **skip list** is thus a generalization of the

linked list in which the data are linked by pointers and elements are sorted as we move from the header to the tail.

The following results are known.

- Papadakis, Munro and Poblete (1990) showed that in fact

$$\mathbf{E}D_i = (1/p)\log_{1/p} i + \log_{1/p}(n/i) + O(1) \ .$$

  They also obtained precise expressions for the $O(1)$ term. Earlier, Pugh (1989) proved the explicit inequality

$$\mathbf{E}D_i \leq (1/p)\log_{1/p} i + \log_{1/p}(n/i) + 1/(1-p) + 1 \ .$$

- Devroye (1990) proved that $D_n/\log_{1/p} n \to 1/p$ in probability and in the mean. The optimal value for p minimizing the asymptotic value for $\mathbf{E}D_n$ is $p = 1/e$. In that case, $\mathbf{E}D_n \sim e \log n$.

The purpose of this paper is to get more refined results. We present a limit law for the D_i's, as well as a large numbers for H_n. The latter result requires a delicate reduction of the problem to one involving the survival of a particular branching process.

THEOREM 1. *Let $i = i(n)$ be a sequence of integers varying with n in such a manner that $i \to \infty$ as $n \to \infty$. Then*

$$\frac{D_i - \log_{1/p}(n/i) - (1/p)\log_{1/p} i}{\sqrt{(1-p)p^{-2}\log_{1/p} i}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \ ,$$

where $\xrightarrow{\mathcal{L}}$ denotes convergence in distribution, and $\mathcal{N}$ is the normal distribution.

COROLLARY. *If i is as in Theorem 1, then*

$$\frac{D_i - \log_{1/p}(n/i)}{\log_{1/p} i} \to \frac{1}{p} \quad \textit{in probability} \ .$$

THEOREM 2. *$H_n/\log_{1/p} n \to c$ in probability, where c is the unique solution greater than one of the equation*

$$\log(1-p) = \log(c-1) - \frac{c}{c-1}\log c \ .$$

Proof of Theorem 1.

The basic identity is

$$D_i = P_i + (K_n - K_i) \ .$$

We will show the following things:

A. $K_n - K_i - \log_{1/p}(n/i) = O_p(1)$, where a sequence of random variables X_n is $O_p(1)$ when $\forall \epsilon > 0 \ \exists M > 0$ such that $\sup_n \mathbf{P}\{|X_n| > M\} < \epsilon$.

B. $(P_i - (1/p)\log_{1/p} i)/\sqrt{(1-p)p^{-2}\log_{1/p} i} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$.

The limit law follows immediately from these two statements. By this device, we have split the problem cleanly into two subproblems, and we won't have to worry about the dependence between P_i and $K_n - K_i$.

Since

$$K_i = \max_{1 \le j \le i} G_j \ ,$$

and $i \to \infty$, we note that $K_i - \log_{1/p} i = O_p(1)$. This standard fact is easily obtained as follows: if $k = \log_{1/p} i + a_i$, and $|a_i - a| < 1$ for some integer a,

$$\begin{aligned}
\mathbf{P}\{K_i \le k\} &= (\mathbf{P}\{G_1 \le k\})^i \\
&= (1 - \mathbf{P}\{G_1 > k\})^i \\
&= \left(1 - p^{k+1}\right)^i \\
&= \left(1 - p^{a_i+1}/i\right)^i \\
&\sim \exp(-p^{a_i+1}) \ .
\end{aligned}$$

We obtain the result by taking a very small and very large. We conclude that $K_n - K_i - \log_{1/p}(n/i) = O_p(1)$.

To handle part B, extend the basic random tree infinitely far to the left, and associate i.i.d. random variables G_i with elements $(i,0)$, $i = \ldots, -2, -1, 0, 1, \ldots, n$. This creates an infinite random tree, as it is uniquely determined by a left-infinite sequence of geometric random variables. There is no root — the root "escapes" to infinity so to speak. The path starting at $(i,0)$ still passes through $(0, K_i)$, as it does in the basic random tree. Thus, P_i remains unchanged, as does $K_n - K_i$. Let N_j be the number of nodes at level j on the path in question. Then it is easy to see that the $N_j - 1$'s are i.i.d. geometric p random variables:

$$\mathbf{P}\{N_j = i\} = (1-p)^{i-1} p \ , \ (i \ge 1) \ .$$

Also, the truncation of the tree at $(0, K_i)$ implies the following:

$$\sum_{j=0}^{\infty} N_j I_{[j<K_i]} < P_i \le \sum_{j=0}^{\infty} N_j I_{[j\le K_i]} \ .$$

As we know from above, $K_i = \log_{1/p} i + O_p(1)$, so that there is very little variation in the number of levels that we need consider. Thus, a simple splitting argument allows us to handle the limit law for P_i. We define the integers

$$k = \lfloor \log_{1/p} i + \log\log i \rfloor \ ,$$
$$l = \lceil \log_{1/p} i - \log\log i \rceil \ .$$

Since the N_j's have mean $1/p$ and variance $(1-p)/p^2$, we have by the central limit theorem, as $i \to \infty$,

$$\frac{\sum_{j=0}^{k} N_j - (k+1)/p}{\sqrt{(1-p)p^{-2}(k+1)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \ ,$$

and similarly with k replaced by l.

We have

$$P_i \le \sum_{j=0}^{k} N_j + \sum_{j=k+1}^{\infty} N_j I_{[j\le K_i]} \ ,$$

so that, with

$$z_i \stackrel{\text{def}}{=} (1/p)\log_{1/p} i + u\sqrt{(1-p)p^{-2}\log_{1/p} i} \ ,$$

and $u \in R$ and $\epsilon > 0$ fixed,

$$\mathbf{P}\{P_i \ge z_i\} \le \mathbf{P}\left\{\sum_{j=0}^{k} N_j \ge z_i\right\} + \mathbf{P}\left\{k+1 \le K_i\right\} \ .$$

The first term on the right-hand-side tends to $1-\Phi(u)$ by the central limit theorem alluded to above, where Φ is the standard normal distribution function. The second term is $o(1)$ because $K_i = \log_{1/p} i + O_p(1)$. This shows that $\liminf_{i\to\infty} \mathbf{P}\{P_i \le z_i\} \ge \Phi(u)$.

In a similar vein, we have

$$P_i \ge \sum_{j=0}^{l} N_j - \sum_{j=0}^{l} N_j I_{[j\ge K_i]} \ .$$

With the same choice of z_i as above, we have

$$\begin{aligned}
\mathbf{P}\{P_i \le z_i\} &\le \mathbf{P}\{\sum_{j=0}^{l} N_j \le z_i\} + \mathbf{P}\{l \ge K_i\} \\
&\le (1+o(1))\Phi(u) + \mathbf{P}\{K_i \le l\} \\
&= (1+o(1))\Phi(u) + (1-p^{l+1})^i
\end{aligned}$$

$$\leq (1+o(1))\Phi(u) + e^{-ip^{l+1}}$$
$$= (1+o(1))\Phi(u) + o(1) \ .$$

We conclude that $\limsup_{i\to\infty} \mathbf{P}\{P_i \leq z_i\} \leq \Phi(u)$. Thus, $\mathbf{P}\{P_i \leq z_i\} \to \Phi(u)$. This concludes the proof of Theorem 1. $\square$

The height of the skip list.

When studying the height of the skip tree, it is convenient to construct a certain branching process. This will be done in this section and in Lemma 1, whose proof is based on arguments that go back to Biggins (1976, 1977) and Devroye (1987). Another proof can be given that uses the theory of extrema of branching random walks as developed for example by Kingman (1975), Hammersley (1974) and Biggins (1976, 1977), but the present proof seems more intuitive and direct.

We begin by defining a right-infinite extension of the basic random tree. The construction is as before, but with every positive integer i we associate a geometric $(1-p)$ random variable G_i, and these form an i.i.d. sequence. The root of this tree is at $(0,\infty)$, and the path from $(i,0)$ to the root still passes through $(0,K_i)$, as it does for the (finite) basic random tree. The collection of descendants of a node $(0,k)$ includes nodes $(1,0)$, ..., $(F_k,0)$, where the letter F refers to the fact that this is the final node among the $(i,0)$'s that can be reached from $(0,k)$. For example, it is clear that $F_{K_n} \geq n$. In other words, by duality, $F_k \geq n$ if $K_n \geq k$.

The basic random tree and its right-infinite extension have the interesting property that every node in it, except those with coordinate of the form $(i,0)$, has a left child. And every node has a right child with probability $1-p$. Moreover, the presence of each of these right edges is decided independently of all the other edges. Indeed, (i,j) has a right child if it is true that the next $m > i$ with $G_m \geq G_i$ is such that $G_m = G_i$. But given that $G_m \geq G_i$, by the memoryless property of the geometric distribution, $G_m = G_i$ with probability $1-p$.

We also note that the subtree of the infinite tree restricted to only those nodes that are descendants of $(0,K_n)$, and are ascendants of one of the nodes $(i,0)$, $i \leq n$, is our original finite basic random tree.

The properties of K_n were studied in the proof of Theorem 1. We need to study the properties of the subtree of $(0,k)$ as k increases. Let T_k be the maximal path distance between $(0,k)$ and any node of the type $(i,0)$. The fundamental auxiliary result of this paper is Lemma 1:

LEMMA 1. *$T_k/k \to c$ in probability, where c is the unique solution greater than one of the equation*

$$\log(1-p) = \log(c-1) - \frac{c}{c-1}\log c \ .$$

PROOF. The proof that $T_k/k \geq c+\epsilon$ finitely often almost surely follows from simple inequalities. Since this result is not needed further on, it will not be shown. To show that $\mathbf{P}\{T_k < (c-\epsilon)k\} \to 0$ for any $0 < \epsilon < c$, we will first prove that it suffices to show that

$$\liminf_{k\to\infty} \mathbf{P}\{T_k \geq (c-\epsilon/2)k\} \geq \beta > 0$$

for some positive β. Consider the maximal path distances in the trees rooted at the right children (if they exist) of $(0,k)$, $(0,k-1)$, $\ldots$, $(0,k-r)$. Call these V_k, $\ldots$, V_{k-r}, where, by convention, the maximal path distance is -1 if the tree is empty. Obviously, since these trees are disjoint, Each tree is empty with probability p.

$$T_k \geq \max_{0\leq j\leq r} \{j+1+V_{k-j}\} \ .$$

Thus, by independence of the V_j's,

$$\begin{aligned}
\mathbf{P}\{T_k < (c-\epsilon)k\} &\leq \prod_{j=0}^{r} \mathbf{P}\{j+1+V_{k-j} < (c-\epsilon)k\} \\
&\leq \prod_{j=0}^{r} (p+(1-p)\mathbf{P}\{j+1+T_{k-j} < (c-\epsilon)k\}) \\
&\leq \prod_{j=0}^{r} (p+(1-p)\mathbf{P}\{T_{k-j} < (c-\epsilon)(k-j)+r(c-\epsilon-1)\}) \\
&\leq \prod_{j=0}^{r} (p+(1-p)\mathbf{P}\{T_{k-j} < (c-\epsilon/2)(k-j)\}) \quad \text{(for } k \text{ large enough)} \\
&= \prod_{j=0}^{r} (1-(1-p)\mathbf{P}\{T_{k-j} \geq (c-\epsilon/2)(k-j)\}) \\
&\leq e^{-(1-p)\sum_{j=0}^{r} P\{T_{k-j}\geq(c-\epsilon/2)(k-j)\}} \\
&\leq e^{-(1-p)(r+1)\alpha/2}
\end{aligned}$$

when $k-r$ is large enough. By choice of a large fixed r, we can make the upper bound as small as desired. Thus, we need only prove that

$$\liminf_{k\to\infty} \mathbf{P}\{T_k \geq (c-\epsilon/2)k\} > 0 \ .$$

To show this, we construct a fictitious branching process. Let $\rho < c$ be a constant. The pater familias of the population is the root node at position $(0,k)$. We carefully choose a large but fixed integer l according to a recipe to be given below. In the infinite associated binary tree, we consider all nodes at distance l from the root, and keep only those whose right-distance is $\geq \lceil \alpha l \rceil$, i.e., the number of right-edges on the path to the root is $\geq \lceil \alpha l \rceil$. The constant α is in $(0,1)$. These nodes are called the offspring of the pater familias. We choose l and α in such a manner that the expected number of offspring is $m > 1$. Each offspring in turn has offspring at distance l, and in this manner, we create a Galton-Watson branching process. The fundamental theorem of branching processes says that the population survives forever with probability $q > 0$. In particular, with probability at least q, there exists a node at distance $\lfloor \rho k/l \rfloor l$ from the root, and at right-distance at least $\lceil \alpha l \rceil \lfloor \rho k/l \rfloor$. Its left-distance is not greater than

$$\lfloor \rho k/l \rfloor (l - \lceil \alpha l \rceil) \leq (\rho k/l)(l - \alpha l) = \rho k(1-\alpha) = k$$

if we have $\rho = 1/(1-\alpha)$. Since the left-distance is $\leq k$, the node actually is a real node, as its level is greater than 0. So, assume the node in question is at coordinates (i,j). Then the node at $(i,0)$ is at even greater distance from the root. This distance is at least

$$l \lfloor \rho k/l \rfloor \geq l(\rho k/l - 1) = \rho k - l \ .$$

Thus,

$$\liminf_{k\to\infty} \mathbf{P}\{T_k \geq \rho k - l\} \geq q > 0 \ .$$

Since l is a constant, and ρ is an arbitrary positive constant less than c, we see that for arbitrary $\epsilon > 0$,

$$\liminf_{k\to\infty} \mathbf{P}\{T_k \geq (c-\epsilon)k\} > 0 \ .$$

Lemma 1 follows if we can establish the existence of constants $\alpha \in (0,1)$ and l (integer) such that $m > 1$. To see this, note that the pater familias node has $\binom{l}{i}$ possible offspring at distance l and right-distance i. A given node with these properties actually exists with probability $(1-p)^i$ because on the path to the root we must meet precisely i right edges at given locations, and these occur independently with probability $1-p$. The expected number of offspring at distance l and right-distance at least $\lceil \alpha l \rceil$ is

$$\begin{aligned}
\sum_{i \geq \lceil \alpha l \rceil} \binom{l}{i} (1-p)^i &\geq \binom{l}{\lceil \alpha l \rceil} (1-p)^{\lceil \alpha l \rceil} \\
&\geq \frac{1}{\sqrt{3\pi}} \sqrt{\frac{l}{\lceil \alpha l \rceil (l - \lceil \alpha l \rceil)}} \frac{l^l (1-p)^{\lceil \alpha l \rceil}}{(\lceil \alpha l \rceil)^{\lceil \alpha l \rceil} (l - \lceil \alpha l \rceil)^{l - \lceil \alpha l \rceil}} \\
&\quad \text{(for all } l \text{ large enough, by Stirling's formula)} \\
&\geq \frac{1}{e\sqrt{4\pi(\alpha l + 1)(1-\alpha)\alpha l}} \left(\frac{(1-p)^\alpha}{\alpha^\alpha (1-\alpha)^{1-\alpha}} \right)^l \\
&\quad \text{(for all } l \text{ large enough)} \\
&> 1
\end{aligned}$$

for all l large enough, provided that

$$\alpha \log(1-p) > \alpha \log(\alpha) + (1-\alpha)\log(1-\alpha) \ .$$

Translated in terms of ρ, this is equivalent to asking that $\rho > 1$ and

$$(\rho-1)\log(1-p) + \rho \log \rho - (\rho-1)\log(\rho-1) > 0 \ .$$

With equality instead of inequality, this is precisely the definition of c. For $1 < \rho < c$, the left-hand side expression is indeed positive. $\square$

Proof of Theorem 2.

In this proof, T_k, D_n, N_j, P_n and K_n keep their meaning from above. In particular, N_j is distributed like one plus a geometric p random variable. Observing that $T_1 \le T_2 \le \cdots \le T_k \le \cdots$, we have the following implication of events, where $k > 0$ is a fixed integer and $\epsilon \in (0, c)$ is a constant:

$$\begin{aligned}
[H_n \le (c-\epsilon)k] &\subseteq [T_{K_n-1} \le (c-\epsilon)k] \\
&\subseteq [[T_k \le (c-\epsilon)k] \cap [K_n - 1 \ge k]] \cup [K_n - 1 < k] \\
&\subseteq [T_k \le (c-\epsilon)k] \cup [K_n < k+1] \ .
\end{aligned}$$

Take $k = \lceil (1-\epsilon)\log_{1/p} n \rceil$. We have seen in the previous section that $K_n / \log_{1/p} n \to 1$ in probability, so that $\mathbf{P}\{K_n < k+1\} \to 0$. Also, by Lemma 1, $\mathbf{P}\{T_k \le (c-\epsilon)k\} \to 0$. We conclude that

$$\lim_{n\to\infty} \mathbf{P}\{H_n \le (c-\epsilon)\log_{1/p} n\} = 0 \ .$$

The theorem follows if we can show that

$$\lim_{n\to\infty} \mathbf{P}\{H_n \ge (c+\epsilon)\log_{1/p} n\} = 0 \ .$$

Define $l = \lfloor \log_{1/p} n - \sqrt{\log_{1/p} n} \rfloor$. Since $H_n = \max_{1\le i \le n} D_i$, and each D_i is stochastically smaller than D_n, we have, with $k = \lfloor \theta \log_{1/p} n \rfloor$, $\theta > c$,

$$\begin{aligned}
\mathbf{P}\{H_n > k\} &\le \mathbf{P}\{K_n > l\} + n\mathbf{P}\{D_n > k, K_n \le l\} \\
&= o(1) + n\mathbf{P}\{P_n > k, K_n \le l\} \\
&\le o(1) + n\mathbf{P}\left\{ \sum_{j=0}^{l} N_j > k \right\} \ .
\end{aligned}$$

Via Chernoff's exponential bounding method (Chernoff, 1953), we obtain for any $t > 0$,

$$\mathbf{P}\left\{ \sum_{j=0}^{l} N_j \ge k \right\} \le e^{-tk} (\mathbf{E}N_1)^{l+1}$$

$$
\begin{aligned}
&\le e^{-tk}(\mathbf{E}e^{tN_1})^{l+1} \\
&= e^{-tk}\left(\frac{pe^t}{1+pe^t-e^t}\right)^{l+1} \\
&\le e^{-tk}\left(\frac{pe^t}{1+pe^t-e^t}\right)^{l} \\
&= \frac{k^k}{l^l(k-l)^{k-l}}(1-p)^{k-l}p^l \quad \left(\text{take } e^t = \frac{k-l}{k(1-p)}\right) \\
&= \left\{\left(\frac{1-p}{1-u}\right)^{1-u}\left(\frac{p}{u}\right)^{u}\right\}^k
\end{aligned}
$$

where $u = l/k \in (0,1)$. We note that as $n \to \infty$, $u \to 1/\theta$. Collecting bounds, we obtain

$$
\begin{aligned}
\mathbf{P}\{H_n > k\} &\le o(1) + n\left\{\left(\frac{1-p}{1-u}\right)^{1-u}\left(\frac{p}{u}\right)^{u}\right\}^k \\
&= o(1) + \exp\Big(\log_{1/p} n(\log(1/p) + \theta((1-u)\log(1-p) \\
&\qquad - (1-u)\log(1-u) + u\log(p) - u\log(u)))\Big) \\
&= o(1)\ ,
\end{aligned}
$$

provided that the coefficient of $\log_{1/p} n$ in the exponent is negative for all n large enough. This leads to the requirement that

$$
(\theta-1)\log(1-p) - (\theta-1)\log\left(\frac{\theta-1}{\theta}\right) + \log(\theta) < 0\ .
$$

This in turn is equivalent to

$$
(\theta-1)\log(1-p) - (\theta-1)\log(\theta-1) + \theta\log(\theta) < 0\ .
$$

With equality instead of inequality, the equation has a unique solution greater than one, c. For $\theta > c$, the left-hand side is negative, as it is asymptotic to $\theta\log(1-p)$. This shows that $\mathbf{P}\{H_n/\log_{1/p} n > c+\epsilon\} \to 0$. $\square$

Remarks, improvements, and extensions.

A The optimal p. We have seen that $D_n \sim (p\log(1/p))^{-1}\log(n)$ in probability, and that $H_n \sim C_p\log(n)$ in probability, where $C_p = c/\log(1/p)$, and c is as in Theorem 2. In Figure 2, we have sketched the coefficients of $\log(n)$ as a function of p. The minimal value of $1/p\log(1/p)$ occurs for $p = 1/e$: at this value, we have $D_n \sim e\log n$ in probability. This was also observed by Pugh (1989, Table 1). Interestingly, the value of p that minimizes the coefficient C_p is very different: the minimal value $C_p = 6.1593\ldots$ is obtained for $p = 0.59139\ldots$. Figure 2 also shows the relative insensitivity of the constants $1/p\log(1/p)$ and C_p to the value of p when p is in the range $0.3\ldots0.6$.

Put Figure 2 about here.

RELATIONSHIP WITH RANDOM BINARY SEARCH TREES. There are several similarities between random binary search trees (see Aho, Hopcroft and Ullman (1983) for definitions and terminology) and skip lists; in both cases, the distance between the root and the last node inserted is asymptotic (in probability) to $A\log n$ for some constant A, while the height of the tree (skip list) is asymptotic (in probability) to $C\log n$ for some constant $C > A$. The constants are smaller for the randomized binary search tree however, with $A = 2$ (Lynch, 1965; Knuth, 1973; Sedgewick, 1983; Mahmoud and Pittel, 1984) and $C = 4.31107\ldots$ (Devroye, 1986, 1987). As we have seen, for skip lists, $A \geq e$ and $C \geq 6.1593\ldots$ uniformly over all p.

STORAGE REQUIREMENTS. The number of nodes in the basic random tree is $n+1+\sum_{i=1}^{n} G_i + \max_{1\leq i\leq n} G_i$, which is close to $n/(1-p)$ by the law of large numbers. Of these, there are exactly n nodes with no left child, as each horizontal edge corresponds to one pole.

TRIMMED SKIP LISTS. The number of operations needed to locate an element x_i in a skip list can be reduced by eliminating (by-passing) all nodes that are left children, that live at a nonzero level, and that have no right child. This is an idea related to that of PATRICIA TREES (see e.g. Knuth, 1973). What is needed to replace this is extra storage in each node to keep track of the level of the node.

References.

A. V. Aho, J. E. Hopcroft, and J. D. Ullman, *Data Structures and Algorithms*, Addison-Wesley, Reading, Mass., 1983.

J. D. Biggins, "The first and last-birth problems for a multitype age-dependent branching process," *Advances in Applied Probability*, vol. 8, pp. 446–459, 1976.

J. D. Biggins, "Chernoff's theorem in the branching random walk," *Journal of Applied Probability*, vol. 14, pp. 630–636, 1977.

H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Annals of Mathematical Statistics*, vol. 23, pp. 493–507, 1952.

L. Devroye, "A note on the height of binary search trees," *Journal of the ACM*, vol. 33, pp. 489–498, 1986.

L. Devroye, "Branching processes in the analysis of the heights of trees," *Acta Informatica*, vol. 24, pp. 277–298, 1987.

L. Devroye, "Expected time analysis of skip lists," Technical Report, School of Computer Science, McGill University, Montreal, 1990.

J. M. Hammersley, "Postulates for subadditive processes," *Annals of Probability*, vol. 2, pp. 652–680, 1974.

W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, pp. 13–30, 1963.

J. F. C. Kingman, "The first-birth problem for an age-dependent branching process," *Annals of Probability*, vol. 3, pp. 790–801, 1975.

D. E. Knuth, *The Art of Computer Programming, Vol. 3 : Sorting and Searching*, Addison-Wesley, Reading, Mass., 1973.

W. C. Lynch, "More combinatorial problems on certain trees," *Computer Journal*, vol. 7, pp. 299–302, 1965.

H. Mahmoud and B. Pittel, "On the most probable shape of a search tree grown from a random permutation," *SIAM Journal of Algebraic and Discrete Methods*, vol. 5, pp. 69–81, 1984.

T. Papadakis, J. I. Munro, and P. V. Poblete, “Analysis of the expected search cost in skip lists,” in: *SWAT 90*, ed. J. R. Gilbert and R. Karlsson, pp. 160–172, Springer Lecture Notes # 447, Springer-Verlag, 1990.

W. Pugh, “Skip lists: a probabilistic alternative to balanced trees,” in: *Algorithms and Data Structures: Workshop WADS ’89 Ottawa*, vol. 382, pp. 437–449, Lecture Notes in Computer Science, Springer-Verlag, Berlin, 1989.

R. Sedgewick, “Mathematical analysis of combinatorial algorithms,” in: *Probability Theory and Computer Science*, ed. G. Louchard and G. Latouche, pp. 123–205, Academic Press, London, 1983.

D. D. Sleator and R. E. Tarjan, “Self–adjusting binary search trees,” *Journal of the ACM*, vol. 32, pp. 652–686, 1985.