A NOTE ON THE PROBABILISTIC ANALYSIS OF PATRICIA TREES

Luc Devroye

School of Computer Science
McGill University
3450 University Street
Montreal H3A 2K6
Canada

September 8, 2008

The author's research was sponsored by NSERC Grant A3456 and FQRNT Grant 90-ER-0291.

ABSTRACT. We consider random PATRICIA trees constructed from n i.i.d. sequences of independent equiprobable bits. We study the height H_n (the maximal distance between the root and a leaf), and the minimal fill-up level F_n (the minimum distance between the root and a leaf). We give probabilistic proofs of

$$\frac{H_n - \log_2 n}{\sqrt{2\log_2 n}} \to 1 \quad \text{almost surely}$$

and

$$\frac{F_n - \log_2 n}{\log_2 \log n} \to -1 \quad \text{almost surely.}$$

KEYWORDS AND PHRASES. Trie, PATRICIA tree, probabilistic analysis, strong convergence, height of a tree.

CR CATEGORIES: 3.74, 5.25, 5.5.

1991 MATHEMATICS SUBJECT CLASSIFICATIONS: 60D05, 68U05.

Introduction.

Tries are efficient data structures that were initially developed and analyzed by Fredkin (1960) and Knuth (1973). The tries considered here are constructed from n independent infinite binary strings $X_1, \ldots, X_n$. Each string defines an infinite path in a binary tree: a 0 forces a move to the left, and a 1 forces a move to the right. For storage purposes, n nodes are identified, one per path, which will represent the n infinite strings; we say that X_i is stored at node i. The tree is now pruned so that it has just n leaves at the n representative nodes. Observe that no representative node is allowed to be an ancestor of any other representative node. The trie is the minimal tree of the type defined above. This implies that every internal (non-leaf) node has at least two leaves in its collection of descendants.

In the **uniform trie model**, the bits in the string X_1 are i.i.d. Bernoulli random variables with success probability $p = 0.5$. For other models, we refer to Devroye (1982, 1984), Régnier (1988), Szpankowski (1988) and Pittel (1985).

The number of steps required to locate a leaf is equal to the length of the path linking X_i and the root. We call this distance the **depth** D_{ni} of node i in a trie of size n. When we want to give guarantees to a potential user about the time required for a look-up, then we should really refer to the **height** $H_n \stackrel{\text{def}}{=} \max_i D_{ni}$. Another quantity of interest to the user is the lower bound on time required to access an element in the sructure, i.e. $F_n \stackrel{\text{def}}{=} \min_i D_{ni}$.

The asymptotic behavior of tries under the uniform trie model is well-known. The height is studied by Régnier (1981), Mendelson (1982), Flajolet and Steyaert (1982), Flajolet (1983), Devroye (1984), Pittel (1985, 1986), and Szpankowski (1988,1989). For the depth of a node, see e.g. Pittel (1986), Jacquet and Régnier (1986), Flajolet and Sedgewick (1986), Kirschenhofer and Prodinger (1986), and Szpankowski (1988). For example, it is known that

$$H_n / \log_2 n \to 2 \text{ almost surely .}$$

The limit law of H_n was obtained in Devroye (1984), and laws of the iterated logarithm for the difference $H_n - 2\log_2 n$ can be found in Devroye (1990).

PATRICIA is a space efficient improvement of the classical trie discovered by Morrison (1968) and first studied by Knuth (1973). It is simply obtained by removing from the trie all internal nodes with one child. Thus, it necessarily has n leaves and $n-1$ internal nodes. The trie from which it is deduced is called the associated trie. All parameters of PATRICIA such as H_n and F_n improve over those for of the associated trie: Pittel (1985) has shown that $H_n / \log_2 n \to 1$ almost surely, which constitutes a 50% improvement over the trie. For other properties, see Knuth (1973), Flajolet and Sedgewick (1986), Kirschenhofer and Prodinger (1986), Szpankowski (1988), and Kirschenhofer, Prodinger and Szpankowski (1989). Recently, Pittel and Rubin (1990) and Pittel (1991) showed that

$$\frac{H_n - \log_2 n}{\sqrt{2\log_2 n}} \to 1 \text{ almost surely.}$$

This result was obtained by a profound combinatorial analysis based on generating functions. Aldous and Shields (1988) showed that the same property holds true for the digital search tree, another modification of the trie with properties typically similar to those of PATRICIA trees. Interestingly, their proof was purely probability theoretical. This led us to believe that the asymptotic behavior of H_n and F_n should be obtainable for PATRICIA trees by purely probabilistic methods as well. The main results can be formulated as follows.

THEOREM 1. *In a sequence of* PATRICIA *trees constructed from an i.i.d. sequence* $X_1, X_2, \ldots$, *we have*

$$\frac{H_n - \log_2 n}{\sqrt{2\log_2 n}} \to 1 \quad \textit{almost surely.}$$

THEOREM 2. *In a sequence of* PATRICIA *trees constructed from an i.i.d. sequence* $X_1, X_2, \ldots$, *we have*

$$\frac{F_n - \log_2 n}{\log_2 \log n} \to -1 \quad \textit{almost surely.}$$

The point of this paper is that both results can be obtained by standard probabilistic methods, such as Poissonization, exponential inequalities, and various embeddings.

Trees and cell occupancies.

It helps to think in terms of an infinite binary tree in which each X_i carves out an infinite path, left edges corresponding to zeros, and right edges to ones. The length of the longest common prefix of two infinite strings x and y is denoted by $l(x,y)$. The collection of all infinite paths y for which $l(x,y) = k$ is denoted by $L(x,k)$. It is clear that for a given x, the sets $L(x,k)$, $k \geq 0$, are disjoint. The point of this is that the depth of X_1 in the PATRICIA tree can be characterized in terms of the occupancies of the sets $L(X_1,k)$. For later reference, $|L(x,k)|$ is the number of X_j's, $1 \leq j \leq n$, that belong to $L(x,k)$. Thus, $\sum_k |L(x,k)| \leq n$, with equality occurring if and only if x is not one of the X_i's. Also, $O(x,k) \stackrel{\text{def}}{=} I_{|L(x,k)|>0}$. Note, in particular, that $|.|$ is not the ordinary cardinality operator.

All D_{ni}'s are identically distributed. From elementary considerations, we have

$$D_{n1} = \sum_{k=0}^{\infty} I_{O(X_1,k)} \ . \tag{1}$$

Without work, we conclude

$$\begin{aligned}
\mathbf{E}D_{n1} &= \sum_{k=0}^{\infty} \mathbf{P}\{|L(X_1,k)| > 0\} \\
&= \sum_{k=0}^{\infty} \Big(1 - \mathbf{P}^{n-1}\{X_2 \notin L(X_1,k) | X_1\}\Big) \\
&= \sum_{k=0}^{\infty} \Big(1 - (1 - 1/2^{k+1})^{n-1}\Big)
\end{aligned}$$

because for any x, $\mathbf{P}\{X_1 \in L(x,k)\} = 1/2^{k+1}$. This suffices to establish that $\mathbf{E}D_{n1} = \log_2 n + O(1)$. It is equally easy to prove that $\mathbf{V}D_{n1} = O(1)$.

For the study of H_n, we need rather exact information regarding the upper tail of the distribution of D_{n1}. The following basic inequality is helpful in this respect.

PROPOSITION 1. *For* $a \geq 1$ *and* $n \geq 2$, *we have*

$$\mathbf{P}\{D_{n1} \geq \log_2 n + \sqrt{a \log_2 n}\} \leq e^{9/2}(n-1)^{-a/2} \ .$$

Proof. Take $t = \log_2(n-1) + \sqrt{a \log_2(n-1)}$. From (1) and Chernoff's bounding method (Chernoff, 1952) we see that for $\lambda > 0$,

$$
\begin{aligned}
\mathbb{P}\{D_{n1} \geq t | X_1\} &\leq e^{-\lambda t} \mathbb{E}\{\prod_{k=0}^{\infty} e^{\lambda I_{O(X_1,k)}} | X_1\} \\
&\leq e^{-\lambda t} \prod_{k=0}^{\infty} \mathbb{E}\{e^{\lambda I_{O(X_1,k)}} | X_1\}
\end{aligned}
$$

by a property of the multinomial distribution. Indeed, given X_1, $|L(X_1,0)|, |L(X_1,1)|, \ldots$ is multinomially distributed. It is known (see e.g. Esary, Proschan and Walkup, 1967, or Joag-Dev and Proschan, 1983) that for a multinomial random vector $Y_1, Y_2, ...,$

$$
\mathbb{E} \prod_i f_i(Y_i) \leq \prod_i \mathbb{E} f_i(Y_i) \ ,
$$

where the f_i's are increasing positive functions. Taking the expectation with respect to X_1 yields the following bound:

$$
\begin{aligned}
\mathbb{P}\{D_{n1} \geq t\} &\leq e^{-\lambda t} \prod_{k=0}^{\infty} \left(1 + (e^{\lambda} - 1)(1 - (1 - 2^{-(k+1)})^{n-1})\right) \\
&\leq e^{-\lambda t} \prod_{k=0}^{\infty} \left(1 + (e^{\lambda} - 1) \min\left(1, \frac{n-1}{2^{k+1}}\right)\right) \ . \qquad (2)
\end{aligned}
$$

The product will be split into three parts,

$$
\prod_{k=0}^{j-1} \times \prod_{k=j}^{m-1} \times \prod_{k=m}^{\infty} \ ,
$$

where we define

$$
\begin{aligned}
j &= \lceil \log_2(n-1) \rceil \\
m &= \lfloor (\lambda + \log(n-1)) / \log 2 \rfloor \\
\lambda &= t \log 2 - \log(n-1) \ .
\end{aligned}
$$

Note that for $a \geq 1$ and $n \geq 2$, we have $0 \leq j \leq m$. We obtain the following:

$$
\prod_{k=0}^{j-1} \left(1 + (e^{\lambda} - 1)\right) \leq e^{\lambda j} \ ,
$$

$$
\begin{aligned}
\prod_{k=j}^{m-1} \left(1 + (e^{\lambda} - 1)\frac{n-1}{2^{k+1}}\right) &\leq \prod_{k=j}^{m-1} e^{\lambda} \frac{n-1}{2^{k+1}} \left(1 + 2^{k+1}/((n-1)e^{\lambda})\right) \\
&\leq \left((n-1)e^{\lambda}\right)^{m-j} 2^{-m(m+1)/2 + j(j+1)/2} \exp\left(\sum_{k=j}^{m-1} 2^{k+1}/((n-1)e^{\lambda})\right) \\
&\leq 2^{t(m-j) - m^2/2 + j^2/2 - (m-j)/2} \exp\left(2^{m+1}/((n-1)e^{\lambda})\right) \\
&\leq 2^{t(m-j) - m^2/2 + j^2/2 - (m-j)/2} e^2 \ ,
\end{aligned}
$$

$$\prod_{k=m}^{\infty}\left(1+(e^{\lambda}-1)\frac{n-1}{2^{k+1}}\right) \le \prod_{k=m}^{\infty}\left(1+e^{\lambda}\frac{n-1}{2^{k+1}}\right)$$
$$\le \exp\left(\sum_{k=m}^{\infty}(n-1)e^{\lambda}/2^{k+1}\right)$$
$$= \exp\left((n-1)e^{\lambda}/2^{m}\right)$$
$$\le e^2 \ .$$

Combining all this shows that

$$\begin{aligned}
\mathrm{P}\{D_{n1} \ge t\} &\le \exp\left(4+\lambda(j-t)\right)2^{t(m-j)-m^2/2+j^2/2-(m-j)/2} \\
&\le \exp\left(4-(j-t)\log(n-1)\right)2^{t(m-t)-m^2/2+j^2/2} \\
&\le e^4 2^{(t-j)\log_2(n-1)-t^2/2+j^2/2} \\
&\le e^4 2^{-(1/2)\log_2^2(n-1)+1/2+t\log_2(n-1)-t^2/2} \\
&= e^{9/2}2^{-(a/2)\log_2(n-1)} \ ,
\end{aligned}$$

which was to be shown. □

The fill-up level of PATRICIA trees.

PROPOSITION 2. *For all $a > 1$,*

$$\mathrm{P}\{F_n < \log_2 n - a\log_2\log n \text{ i.o. }\} = 0 \ .$$

PROOF. Define $k = \lfloor \log_2 n - a\log_2 \log n \rfloor$. If $\min_i D_{ni} < k$, then one of the 2^k possible prefix strings of length k does not occur among $X_1, \ldots, X_n$. Thus, by symmetry,

$$\begin{aligned}
\mathrm{P}\{\min_{1\le i\le n} D_{ni} < k\} &\le 2^k\mathrm{P}\{\text{no } X_i \text{ starts with } k \text{ zeroes}\} \\
&= 2^k(1-1/2^k)^n \\
&\le 2^k e^{-n/2^k} \\
&\le \exp\left(\log n - (\log n)^a - a\log\log n\right) \\
&\to 0 \ .
\end{aligned}$$

The upper bound is also summable in n for all $a > 1$, so Proposition 2 follows by the Borel-Cantelli lemma. □

Hoeffding (1963) has developed useful exponential inequalities for tail probabilities of sums of independent random variables. The generalization of these inequalities to martingales (Hoeffding, 1963, Azuma, 1967) has led to interesting applications in combinatorics and the theory of random graphs (for a survey, see McDiarmid, 1989). The following extension of Hoeffding's inequality is useful for random variables that are complicated functions of independent random variables, and that are relatively robust to individual changes in the values of the random variables.

LEMMA 1. (MCDIARMID, 1989) *Let $X_1, \ldots, X_n$ be independent random variables taking values in a set A, and assume that $f : A^n \to R$ satisfies*

$$\sup_{\substack{x_1,\ldots,x_n \\ x'_1,\ldots,x'_n \in A}} |f(x_1, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x'_i, x_{i+1}, \ldots, x_n)| \le c_i \ , \ 1 \le i \le n \ .$$

Then

$$\mathbb{P}\left\{|f(X_1, \ldots, X_n) - \mathbb{E}f(X_1, \ldots, X_n)| \ge t\right\} \le 2e^{-2t^2 / \sum_{i=1}^n c_i^2} \ .$$

PROPOSITION 3. *For all $a < 1$,*

$$\mathbb{P}\{F_n > \log_2 n - a \log_2 \log n \text{ i.o. }\} = 0 \ .$$

PROOF. Define $k = \lceil \log_2 n - a \log_2 \log n \rceil$. We note that D_{ni} is smaller than the corresponding value in the associated trie. Thus, we need only prove Proposition 3 for the ordinary trie. For the remainder of this proof, D_{ni} thus denotes a depth in the associated trie. Fix an integer $m < n$, and for $1 \le i \le m$ define D^*_{ni} as the depth of node X_i in the trie formed by X_i and $X_{m+1}, \ldots, X_n$. If $l(X_i, X_j) \le k$ for all $1 \le j \le m$, then $D_{ni} > k$ if and only if $D^*_{ni} > k$. Let J be the collection of all indices $i \le m$ such that its prefix of length k has not occurred among $X_1, \ldots, X_{i-1}$. Observe that

$$\begin{aligned}
\mathbb{P}\{\min_{1\le i\le n} D_{ni} > k\} &\le \mathbb{P}\{\min_{1\le i\le n} D^*_{ni} > k\} \\
&\le \mathbb{P}\{\min_{1\le i\le m} D^*_{ni} > k\} \\
&\le \mathbb{P}\{\min_{i\in J} D^*_{ni} > k\} \ .
\end{aligned}$$

We condition on $X_1, \ldots, X_m$, and let $|J|$ be the cardinality of J. Let A_i be the event

$$\max_{m+1\le j\le n} l(X_i, X_j) \ge k \ .$$

Using the association inequality for the multinomial distribution used earlier, we have

$$\begin{aligned}
\mathbb{P}\{\min_{i\in J} D^*_{ni} > k | X_1, \ldots, X_m\} &\le \mathbb{P}\{\cap_{i\in J} A_i | X_1, \ldots, X_m\} \\
&\le \prod_{i\in J} \mathbb{P}\{A_i | X_i\} \\
&= \prod_{i\in J} \left(1 - \mathbb{P}^{n-m}\{l(X_1, X_2) < k\}\right) \\
&= \prod_{i\in J} \left(1 - (1 - 1/2^k)^{n-m}\right) \\
&\le \exp\left(-|J|(1 - 1/2^k)^{n-m}\right) \ .
\end{aligned}$$

Collecting all this shows that

$$\mathbb{P}\{F_n > k\} \le \mathbb{E}\left\{\exp(-|J|(1-1/2^k)^{n-m})\right\} \ .$$

Note that $|J|$ is a function governed by Lemma 1, provided that we replace n by m and c_i by 1. We have

$$\mathbb{P}\{||J| - \mathbb{E}|J|| > t\} \le 2\exp(-2t^2/m) \ ,$$

and, in particular,

$$\mathbb{P}\{|J| < \mathbb{E}|J|/2\} \le 2\exp(-\mathbb{E}^2|J|/2m) \ .$$

Observe that

$$\mathbb{E}|J| = 2^k\left(1-(1-1/2^k)^m\right) \ge 2^k\left(1-(1-m/2^k+m^2/2^{2k+1})\right) = m\left(1-m/2^{k+1}\right) \ .$$

Thus,

$$\mathbb{P}\{F_n > k\} \le 2\exp(-\mathbb{E}^2|J|/2m) + \exp(-(1/2)\mathbb{E}|J|(1-1/2^k)^{n-m}) \ . \tag{3}$$

We take $m = \lceil n/\log n\rceil$. Since $2^k \ge n/(\log n)^a$, we note that $m/2^k \to 0$. The exponent in the first term of (3) is

$$-\mathbb{E}^2|J|/2m \sim -m/2 \ .$$

The exponent in the second term of (3) is

$$\begin{aligned}
-(1/2)\mathbb{E}|J|(1-1/2^k)^{n-m} &\le -m(1/2+o(1))e^{-(n-m)/(2^k-1)} \\
&\sim -(m/2)e^{-n/2^k} \\
&\le -(1/2)\exp(\log m - (\log n)^a) \\
&= -m^{1+o(1)} \ .
\end{aligned}$$

We have

$$\mathbb{P}\{F_n > k\} \le \exp(-(n/\log n)^{1+o(1)}) \ .$$

The upper bound is summable in n, so that Proposition 3 follows by the Borel-Cantelli lemma. □

Propositions 2 and 3 together imply

$$\frac{F_n - \log_2 n}{\log_2 \log n} \to -1 \ \text{ almost surely.}$$

The height of PATRICIA trees.

In the associated trie, it is very likely that there exists a leaf at distance at least $k = \lfloor \log_2 n + \sqrt{(2-\epsilon)\log_2 n}\rfloor$ from the root with the property that the first k nodes on its path from the root all have two children, and are thus not deleted when the PATRICIA tree is constructed. This argument leads to a lower bound for H_n (proposition 5 below). An upper bound can be obtained trivially from the large deviation result for D_{n1} given in proposition 1. See proposition 4 below.

Proposition 4. *For all $a > 2$,*

$$\mathbf{P}\{H_n \geq \log_2 n + \sqrt{a \log_2 n} \text{ i.o. }\} = 0 \ .$$

Proof. Define $t(n) = \log_2 n + \sqrt{a \log_2 n}$. From Proposition 1,

$$\mathbf{P}\{H_n \geq t(n)\} \leq n\mathbf{P}\{D_{n1} \geq t(n)\} \leq e^{9/2} n(n-1)^{-a/2} = O(n^{(2-a)/2}) \ . \tag{4}$$

Next, with $n_i = 2^i$, the monotonicity of H_n and of $t(n)$ imply that we need only show that $H_{n_{i+1}} > t(n_i)$ finitely often almost surely. For i large enough, we have $t(n_{i+1}) < t(n_i)+2$. Thus, the Proposition follows by the Borel-Cantelli lemma if for all $a > 0$,

$$\sum_{i=1}^{\infty} \mathbf{P}\{H_{n_i} \geq t(n_i)\} < \infty \ .$$

This is immediate from (4). $\square$

Proposition 5. *For all $a < 2$,*

$$\mathbf{P}\{H_n \leq \log_2 n + \sqrt{a \log_2 n} \text{ i.o. }\} = 0 \ .$$

Proof. Assume without loss of generality that $a > 4/3$. Define the following quantities:

$$\begin{aligned} k &= \lfloor \log_2 n - 2\log_2 \log n \rfloor; \\ m &= \lceil n^{a-1} \rceil; \\ l &= \lfloor \sqrt{(3a-4)\log_2 n} \rfloor \ . \end{aligned}$$

The integer m is used to split the data into two parts. We will show that

$$\sum_{n=1}^{\infty} \mathbf{P}\{H_n \leq k + l\} < \infty \ .$$

The proposition then follows by the Borel-Cantelli lemma. Indeed, for fixed $\epsilon > 0$, by choosing a close enough to 2, we can make $k + l$ bigger than $\log_2 n + \sqrt{(2-\epsilon)\log_2 n}$. The proof is greatly simplified if we use an embedding argument. Let $X_1, X_2, \ldots$ be an i.i.d. sequence of infinite strings, defining an infinite sequence of PATRICIA trees and associated tries. It is easy to see that D_{ni} is an increasing function of n for fixed i. Thus, H_n is also increasing in n. Hence, if we let N be a Poisson $((n-m)/2)$ random variable independent of the sequence of X_i's, we note that

$$\begin{aligned} \mathbf{P}\{H_n \leq k + l\} &\leq \mathbf{P}\{H_{m+N} \leq k + l \, , \, N \leq n - m\} + \mathbf{P}\{N > n - m\} \\ &\leq \mathbf{P}\{H_{m+N} \leq k + l\} + \mathbf{P}\{N > n - m\} \\ &\leq \mathbf{P}\{H_{m+N} \leq k + l\} + (\sqrt{e}/2)^{n-m} \ . \end{aligned}$$

The inequality for the Poisson tail follows from Chernoff's bound: for $t > 0$,

$$\mathbf{P}\{N > n - m\} \leq \mathbf{E}\{e^{t(N-n+m)}\} = e^{((n-m)/2)(e^t - 1) - t(n-m)} = (\sqrt{e}/2)^{n-m} \ ,$$

where we took $t = \log 2$. Clearly, we need only show that

$$\sum_{n=1}^{\infty} \mathbb{P}\{H_{m+N} \le k+l\} < \infty \ .$$

Let J be the subset of $\{1, \ldots, m\}$ such that $i \in J$ if and only if $l(X_i, X_j) < k$ for all $j < i$. Thus, $\{X_i, i \in J\}$ is a collection of strings with different prefixes of length k. Let us define the event G (for "good") by

$$G \stackrel{\text{def}}{=} [|J| \ge \mathbb{E}|J|/2] \ .$$

From the proof of Proposition 3, we see that $\mathbb{E}|J| \ge m(1 - m/2^{k+1}) \sim m$, and that

$$\mathbb{P}\{G^c\} \le 2\exp(-\mathbb{E}^2|J|/2m) \le 2\exp(-(1/2 + o(1))m) \ ,$$

which is summable in n. This reduces our problem to that of proving

$$\sum_{n=1}^{\infty} \mathbb{P}\{H_{m+N} \le k+l \, , \, G\} < \infty \ . \tag{5}$$

Let us call $L(x, d)$ the set of all strings y with $l(x, y) = d$, and let $|L(x, d)|$ denote the cardinality of that set, when intersected with $\{X_{m+1}, \ldots, X_{m+N}\}$. We have the following inclusion of events:

$$\begin{aligned}
[H_{m+N} \le\, & k+l] \cap G \\
& \subseteq G \cap \left\{ \cap_{i \in J} \cup_{d=0}^{k+l} [|L(X_i, d)| = 0] \right\} \\
& \subseteq \left\{ \cup_{i=1}^{m} \cup_{d=0}^{k} [|L(X_i, d)| = 0] \right\} \cup \left\{ G \cap \left\{ \cap_{i \in J} \cup_{d=k+1}^{k+l} [|L(X_i, d)| = 0] \right\} \right\} \ .
\end{aligned}$$

The two events on the right-hand side are dealt with separately. First of all,

$$\begin{aligned}
\mathbb{P}\left\{ \cup_{i=1}^{m} \cup_{d=0}^{k} [|L(X_i, d)| = 0] \right\} &\le m(k+1)\mathbb{P}\{|L(X_1, k)| = 0\} \\
&= m(k+1)\exp\left(-\frac{n-m}{2} 2^{-(k+1)} \right) \\
&\le n(\log_2 n + 1)\exp\left(-(1 - m/n)(\log n)^2 \right) \ ,
\end{aligned}$$

which is summable in n, as required. We can use the fundamental property of the Poisson distribution, and the fact that the sets $L(X_i, d)$ for $d \geq k+1$ and $i \in J$ are disjoint. This yields

$$\begin{aligned}
&\mathbb{P}\Big\{G \cap \Big\{\cap_{i\in J} \cup_{d=k+1}^{k+l} [|L(X_i,d)| = 0]\Big\} \Bigm| X_1,\ldots,X_m\Big\} \\
&\quad = I_G \prod_{i\in J} \mathbb{P}\left\{\cup_{d=k+1}^{k+l} [|L(X_i,d)| = 0] \Bigm| X_i\right\} \\
&\quad = I_G \prod_{i\in J} \left(1 - \mathbb{P}\left\{\cap_{d=k+1}^{k+l} [|L(X_i,d)| > 0] \Bigm| X_i\right\}\right) \\
&\quad \leq I_G \exp\left(-\sum_{i\in J} \prod_{d=k+1}^{k+l} \mathbb{P}\{|L(X_i,d)| > 0 | X_i\}\right) \\
&\quad = I_G \exp\left(-|J| \prod_{d=k+1}^{k+l} \left(1 - \exp(-\frac{n-m}{2}/2^{d+1})\right)\right) \\
&\quad \leq \exp\left(-(\mathbb{E}|J|/2) \prod_{d=k+1}^{k+l} (1 - \exp(-(n-m)/2^d))\right) .
\end{aligned}$$

Because the upper bound does not depend upon $X_1, \ldots, X_m$, it remains valid if we take expectations to rid ourselves of the conditioning. Property (5) follows if we can show that this upper bound is summable in n. Call M minus the logarithm of the upper bound. Using $1 - e^{-u} \geq \min(1, u)/2$, valid for $u > 0$, we have

$$\begin{aligned}
M &\geq m2^{-l}\left(\frac{1}{2} + o(1)\right) \prod_{d=k+1}^{k+l} \min(1, (n-m)/2^d) \\
&\sim m2^{-(l+1)-kl} \prod_{d=1}^{l} \min(2^k, (n-m)/2^d) \\
&= m2^{-(l+1)-kl} \prod_{d=1}^{l} ((n-m)/2^d) \times \prod_{1\leq d\leq l; k+d<\log_2(n-m)} (2^{k+d}/(n-m)) \\
&\geq m2^{-(l+1)-kl}(n-m)^l/2^{l(l+1)/2} \times \left(2^{k+1}/(n-m)\right)^{\log_2(n-m)-k} \\
&\geq m2^{-(l+1)-l\log_2 n + l\log_2\log n - l(l+1)/2 + l\log_2(n-m)} \times (\log n)^{2k-2\log_2(n-m)} \\
&= \exp\Big(o(\log n) + \log m - l\log n - l^2\log(2)/2 \\
&\qquad + l\log(n-m) + (2k - 2\log_2(n-m))\log\log n\Big) \\
&\geq \exp\big(o(\log n) + (a-1)\log n - (3a-4)\log(n)/2 \\
&\qquad + l\log(1-m/n) - 2(2\log_2\log n + \log_2(1-m/n))\log\log n\big) \\
&= \exp\left(o(\log n) + (1-a/2)\log n\right) .
\end{aligned}$$

Thus,

$$\sum_n e^{-M} = \sum_n e^{n^{1-a/2+o(1)}} < \infty . \square$$

Propositions 4 and 5 together imply

$$\frac{H_n - \log_2 n}{\sqrt{2 \log_2 n}} \to 1 \quad \text{almost surely.}$$

References.

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

D. Aldous and P. Shields, "A diffusion limit for a class of randomly-growing binary trees," *Probability Theory and Related Fields*, vol. 79, pp. 509–542, 1988.

K. Azuma, "Weighted sums of certain dependent random variables," *Tohoku Mathematical Journal*, vol. 37, pp. 357–367, 1967.

H. Chernoff, "A measure of asymptotic efficiency of 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 average depth of tries," *Computing*, vol. 28, pp. 367–371, 1982.

L. Devroye, "A probabilistic analysis of the height of tries and of the complexity of triesort," *Acta Informatica*, vol. 21, pp. 229–237, 1984.

L. Devroye, "A study of trie-like structures under the density model," *Annals of Applied Probability*, vol. 2, pp. 0–0, 1991.

J. D. Esary, F. Proschan, and D. W. Walkup, "Association of random variables, with applications," *Annals of Mathematical Statistics*, vol. 38, pp. 1466–1474, 1967.

P. Flajolet, "On the performance evaluation of extendible hashing and trie search," *Acta Informatica*, vol. 20, pp. 345–369, 1983.

P. Flajolet and R. Sedgewick, "Digital search trees revisited," *Siam Journal on Computing*, vol. 15, pp. 748–767, 1986.

P. Flajolet and J. M. Steyaert, "A branching process arising in dynamic hashing, trie searching and polynomial factorization," in: *Lecture Notes in Computer Science*, vol. 140, pp. 239–251, Springer-Verlag, New York, 1982.

E. H. Fredkin, "Trie memory," *Communications of the ACM*, vol. 3, pp. 490–500, 1960.

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

P. Jacquet and M. Régnier, "Trie partitioning process: limiting distributions," in: *Lecture Notes in Computer Science*, vol. 214, pp. 196–210, 1986.

K. Joag-Dev and F. Proschan, "Negative association of random variables, with applications," *Annals of Statistics*, vol. 11, pp. 286–295, 1983.

P. Kirschenhofer and H. Prodinger, “Some further results on digital trees,” in: *Lecture Notes in Computer Science*, vol. 226, pp. 177–185, Springer-Verlag, Berlin, 1986.

P. Kirschenhofer, H. Prodinger, and W. Szpankowski, “On the balance property of PATRICIA tries: external path length viewpoint,” *Theoretical Computer Science*, vol. 68, pp. 1–17, 1989.

P. Kirschenhofer, H. Prodinger, and W. Szpankowski, “On the variance of the external path length in a symmetric digital trie,” *Discrete Applied Mathematics*, vol. 25, pp. 129–143, 1989.

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

V. F. Kolchin, B. A. Sevast'yanov, and V. P. Chistyakov, *Random Allocations*, V.H. Winston and Sons, Washington, D.C., 1978.

C. McDiarmid, “On the method of bounded differences,” in: *Surveys in Combinatorics*, ed. J. Siemons, vol. 141, pp. 148–188, London Mathematical Society Lecture Note Series, Cambridge University Press, 1989.

H. Mendelson, “Analysis of extendible hashing,” *IEEE Transactions on Software Engineering*, vol. 8, pp. 611–619, 1982.

D. R. Morrison, “PATRICIA — Practical Algorithm To Retrieve Information Coded in Alphanumeric,” *Journal of the ACM*, vol. 15, pp. 514–534, 1968.

B. Pittel, “Asymptotical growth of a class of random trees,” *Annals of Probability*, vol. 13, pp. 414–427, 1985.

B. Pittel, “Path in a random digital tree: limiting distributions,” *Advances in Applied Probability*, vol. 18, pp. 139–155, 1986.

B. Pittel, “On the height of PATRICIA search tree,” ORSA/TIMS Special Interest Conference on Applied Probability in the Engineering, Information and Natural Sciences, Monterey, CA, 1991.

B. Pittel and H. Rubin, “How many random questions are necessary to identify n distinct objects?,” *Journal of Combinatorial Theory*, vol. A55, pp. 292–312, 1990.

M. Régnier, “On the average height of trees in digital searching and dynamic hashing,” *Information Processing Letters*, vol. 13, pp. 64–66, 1981.

M. Régnier, “Trie hashing analysis,” in: *Proceedings of the Fourth International Conference on Data Engineering*, pp. 377–381, IEEE, Los Angeles, 1988.

W. Szpankowski, “On the height of digital trees and related problems,” Technical Report CSD-TR-816, Department of Computer Science, Purdue University, 1988.

W. Szpankowski, “Some results on V-ary asymmetric tries,” *Journal of Algorithms*, vol. 9, pp. 224–244, 1988.

W. Szpankowski, “Digital data structures and order statistics,” in: *Algorithms and Data Structures: Workshop WADS ’89 Ottawa*, vol. 382, pp. 206–217, Lecture Notes in Computer Science, Springer-Verlag, Berlin, 1989.