The Annals of Applied Probability
1992, Vol. 2, No. 2, 402–434

A STUDY OF TRIE-LIKE STRUCTURES UNDER THE DENSITY MODEL[1]

By Luc Devroye

McGill University

We consider random tries constructed from sequences of i.i.d. random variables with a common density f on $[0, 1]$ (i.e., paths down the tree are carved out by the bits in the binary expansions of the random variables). The depth of insertion of a node and the height of a node are studied with respect to their limit laws and their weak and strong convergence properties. In addition, laws of the iterated logarithm are obtained for the height of a random trie when $\int f^2 < \infty$. Finally, we study two popular improvements of the trie, the PATRICIA tree and the digital search tree, and show to what extent they improve over the trie.

Introduction. *Tries* are efficient data structures that were developed and modified by Fredkin (1960), Knuth (1973), Larson (1978), Fagin, Nievergelt, Pippenger and Strong (1979), Litwin (1981), Aho, Hopcroft and Ullman (1983) and others. The tries considered here are constructed from n independent infinite binary strings $X_1, \dots, X_n$. 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. Clearly, there are infinitely many possible trees. We define the trie as the minimal tree of the type defined above. This implies that every internal (nonleaf) 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 $p = 0.5$. The X_i's can also be considered as random variables on $[0, 1]$ when the bits in the strings are just the fractional binary expansions. Thus, in the uniform trie model, the X_i's are i.i.d. uniform $[0, 1]$ random variables [see, e.g., Knuth (1973) or Aho, Hopcroft and Ullman (1983)]. Other models have been proposed in the literature: In the *density model*, the X_i's are i.i.d. with density f on $[0, 1]$ [Devroye (1982, 1984)]. In this case, the bits are no longer independent. It is this model that

Received January 1990; revised November 1990.

[1]The author's research was sponsored by NSERC Grant A3456 and FCAR Grant EQ-1678. Part of this research was carried out while visiting the Division of Statistics, University of California at Davis.

AMS 1980 *subject classifications.* 60D05, 68U05.

Key words and phrases. Trie, digital search tree, probabilistic analysis, strong convergence, height of a tree.

will be dealt with in the paper. Others have considered the *singular continuous model*, in which the strings form an m-ary expansion, and the symbols in the string (items in the expansion) occur independently with probabilities $p_0, \dots, p_{m-1}$. When the probabilities are unequal and none is 1, then the X_i's are singular continuous random variables. Noting that alphabetic data rarely follow the uniform or singular continuous models, Régnier (1988) and Szpankowski (1988b) considered the *Markovian model*, in which the strings of symbols form a Markov sequence. A *strongly mixing model* has been studied by Pittel (1985).

The random tries thus constructed are used in computer science applications when data need to be stored and the whole is to be regarded as a dictionary; that is, we can insert new elements, delete certain elements, look up information stored at certain elements and modify information stored at certain elements. If element X_i is stored in node i, we usually associate with node i additional information regarding X_i that is of no concern to us here; just think of it as the definition of X_i in a dictionary. To look this information up forces us to access the root, and then to follow a path down the tree as indicated by the bit string in X_i, until we reach the node at which X_i is stored. The number of steps 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 time required to insert a new element in the dictionary. This is clearly seen to be proportional to the depth of node $n + 1$ in a trie of size $n + 1$. We will use the notation D_{n+1}. The above quantities have a direct relationship to the time required to carry out certain operations. Other key quantities not studied here include: the *conditional depth of insertion* $C_n \stackrel{\text{def}}{=} \mathbf{E}\{D_{n+1} | X_1, \dots, X_n\}$ (which measures the depth of insertion in a given trie when averaged over all possible random variables that have to be inserted), the *average depth* $A_n = (1/n)\sum_{i=1}^{n} D_{n,i}$ and the *size* S_n, the number of nodes in the trie (which can be greater than n since only leaves represent elements). The size of the trie can be superlinear for some densities; yet, for two simple modifications, PATRICIA and the digital search tree, the size is guaranteed to be $O(n)$ in the worst case; therefore, a study of its properties is less important at this stage.

The asymptotic behavior of tries under the singular continuous and uniform models is well known. It should be clear that tries grow unboundedly as $n \to \infty$. Yet, alphabetic data are by definition of limited length. Thus the asymptotic analysis for the singular continuous model may not reflect what is observed in practice for such tries. If we use the trie to store real numbers, however (e.g., times of events in discrete event simulation; positions of particles in a physics simulation), then the asymptotic analysis for the density model does indeed have a direct relationship with the "real world." Hence the need to understand the properties of the density model.

Most of the key properties of tries under the singular continuous model are well understood. The height is studied by Régnier (1981), Mendelson (1982), Flajolet and Steyaert (1982), Flajolet (1983), Devroye (1984), Pittel (1985, 1986) and Szpankowski (1988a, 1989). For the depth of a node, see, for example, Pittel (1986), Jacquet and Régnier (1986), Flajolet and Sedgewick (1986), Kirschenhofer and Prodinger (1986) and Szpankowski (1988b). The size is analyzed in Jacquet and Régnier (1986), and the average depth in Kirschenhofer, Prodinger and Szpankowski (1989b). See also Flajolet and Puech (1983, 1986), Flajolet (1983) and Flajolet, Régnier and Sotteau (1985). To put tries into a broader context, see Vitter and Flajolet (1990).

In this study, the depth and the height of a trie under the density model are studied in more detail. Arguments include Schur convexity, the Lebesgue density theorem and the Hardy–Littlewood maximal function. Furthermore, it is shown how two related improvements of tries, the PATRICIA tree [Knuth (1973)] and the digital search tree [Coffman and Eve (1970)], behave under the density model. Another variant, the DISC [Luccio, Régnier and Schott (1989)], will not be discussed here.

The main message in the paper is that the trie and its variants are remarkably robust under departures from uniformity. The asymptotic behavior of D_n and H_n is basically like that of the uniform trie model; typically, the density f only affects the second term in the asymptotics. We first show that for any density, $D_{n+1} - \log_2 n$ has a limit law depending upon f. The same is true for $H_n - 2\log_2 n$ when $\int f^2 < \infty$. In both cases, the density affects the "constant" term in the asymptotics. The factor $\int f^2$, which is an indicator of the peakedness of f, plays a key role in the analysis of the height. This should come as no surprise, as growing tries are bound to uncover the finer detail of densities. Interestingly, the depth D_n is mainly influenced by the *entropy* $-\int f \log f$. In the context of coding, Rényi (1959) and Csiszár (1969) already noted the importance of the entropy in problems involving partitions of the unit interval. The strong behavior and some laws of the iterated logarithm complete the study of H_n. When $\int f^2 = \infty$, the height is no longer concentrated about $2\log_2 n$, but can grow at any prespecified rate. This dependence is investigated. Similarly, we also look at how the density influences $\mathbf{E}D_{n+1}$. We conclude the study by looking at the same questions for PATRICIA and digital search trees. We will find, for example, that in both cases, $H_n/\log_2 n \to 1$ almost surely when $\int f^p < \infty$ for all $p \geq 1$. This improves over the behavior of H_n for the ordinary trie by about 50%. A final word about the notation. The *dyadic intervals* of $[0,1]$ are denoted by $I_{i,k}$, $1 \leq i \leq 2^k$, $k \geq 0$, where $I_{i,k} = [(i-1)/2^k, i/2^k)$. For $x \in [0,1)$, let $A_{x,k}$ be the unique interval in the collection of $I_{i,k}$'s to which x belongs. We define $q_{i,k} = \int_{I_{i,k}} f$ and $q_{x,k} = \int_{A_{x,k}} f$.

Depth of a trie: A limit law. Consider a trie built up on the basis of n i.i.d. random variables $X_1, \ldots, X_n$, drawn from density f. Clearly, all the $D_{n,i}$'s are identically distributed, and thus we can and do write D_n for the marginal random variable. We begin with the following fundamental property.

LEMMA D1.

$$\mathbf{P}\{D_{n+1} \le k\} = \mathbf{E}\left\{\left(1 - \int_{I_{X,k}} f\right)^n\right\} = \sum_{i=1}^{2^k} \int_{I_{i,k}} f\left(1 - \int_{I_{i,k}} f\right)^n$$
$$= \sum_{i=1}^{2^k} q_{i,k}(1 - q_{i,k})^n,$$

where X is a random variable with density f. Also,

$$\mathbf{P}\{D_{n+1} > k\} = \sum_{i=1}^{2^k} q_{i,k}\left(1 - (1 - q_{i,k})^n\right) \le n \sum_{i=1}^{2^k} q_{i,k}^2 \le n 2^{-k} \int f^2.$$

PROOF. Let $X_{n+1} = X$. Then X_{n+1} has depth of insertion D_{n+1} less than or equal to k if and only if $I_{X,k}$ contains none of the X_i's with $1 \le i \le n$. The last statement follows from Jensen's inequality. □

We have the following limit law.

THEOREM D2. *Under the density model, for all constants M,*

$$\lim_{n \to \infty} \sup_{u:\, u + \log_2 n \text{ integer},\, |u| \le M} \left| \mathbf{P}\{D_{n+1} \le \log_2 n + u\} - \int f(y) e^{-f(y)2^{-u}}\, dy \right| = 0,$$

and for all fixed u,

$$\lim_{n \to \infty} \left| \mathbf{P}\{D_{n+1} \le \lfloor \log_2 n + u \rfloor\} - \int f(y) e^{-f(y)2^{-u^*}}\, dy \right| = 0,$$

where $u^* = \lfloor u + \log_2 n \rfloor - \log_2 n$.

PROOF. Note that for almost all $x \in [0,1)$, we have $\int_{A_{x,k}} f \sim 2^{-k} f(x)$ as $k \to \infty$, where 2^k is the number of intervals into which the unit interval is partitioned. This is a special version of the Lebesgue density theorem [Wheeden and Zygmund (1977)]. Thus, if $k = \log_2 n + u$, we have $(1 - \int_{A_{x,k}} f)^n \to e^{-f/2^u}$. This concludes the first part of the theorem. The second part is immediate from the first one. □

The odd format of the limit law is due to some discretization problems. Ignoring these for a moment, we could say that the limit distribution function is close to

$$F(u) = \int f(y) e^{-f(y)2^{-u}}\, dy.$$

Note that by the Lebesgue dominated convergence theorem, it is easy to verify that F is a bona fide distribution function. It depends in an intricate way upon f. From Theorem D2, we also have the following law of large numbers [see also Devroye (1982)], where a sequence of random variables Y_n is said to be $O_p(1)$ when $\lim_{M \to \infty} \limsup_{n \to \infty} \mathbf{P}\{|Y_n| > M\} = 0$.

THEOREM D3. *For any density f, $D_n - \log_2 n = O_p(1)$.*

Thus D_n shows a remarkable robustness to nonuniformity. In fact, we cannot expect to find a structure with better asymptotic performance, since complete binary trees need about $\log_2 n$ levels of nodes. It is nevertheless odd that for any $u > 0$, no matter how large,

$$\liminf_{n \to \infty} \mathbf{P}\{D_{n+1} \le \log_2 n - u\} > 0.$$

In other words, we beat the "optimal" (complete binary tree) value of $\log_2 n$ by any large fixed amount with positive probability.

REMARK 1 (Singular continuous distributions). Assume for the moment that the distribution of X_1 is continuous with an absolutely continuous part and singular continuous part with support set $S \subseteq [0, 1]$, where the set S is the collection of all x for which $\mu(x - h, x + h)/2h \to \infty$ as $h \downarrow 0$, and μ is the probability measure for X_1. Using arguments as in the proof of Theorem D2, we note that for any u, no matter how large,

$$\liminf_{n \to \infty} \mathbf{P}\{D_{n+1} > \log_2 n + u\} \ge \mu(S).$$

If there is an atomic part in the distribution of X_1, then $\mathbf{P}\{D_2 = \infty\} = \sum_i p_i^2$, where the p_i's form the sequence of probabilities of the atoms. We thus conclude that $D_{n+1} - \log_2 n = O_p(1)$ if and only if μ is absolutely continuous.

REMARK 2 (Smooth limit distribution). It is instructive to look at the smooth (i.e., nondiscretized) distribution function

$$F(u) = \int f(y) e^{-f(y)2^{-u}}\, dy.$$

The mean of the extreme value distribution function $e^{-e^{-u}}$ is $\gamma = 0.57722\ldots$ (Euler's constant) and its variance is $\pi^2/6 = 1.64493\ldots$. It is easy to verify that F has mean $(\gamma - H)/\log 2$ and variance $(\pi^2/6 + H^{(2)})/\log^2 2$, where $H = -\int f \log f$ is the *entropy*, and $H^{(2)} = \int f \log^2 f - \int^2 f \log f$. The entropy H of a density f on $[0, 1]$ is always nonpositive. It is maximal and 0 for the uniform density. The quantity $H^{(2)}$ governing the variance is minimal and 0 for the uniform distribution as well.

REMARK 3 (Uniform density). Pittel (1986) obtained the limit law D2 for the uniform density. His result states that $\mathbf{P}\{D_{n+1} \le \log_2 n + u\} \to \exp(-2^{-u})$. This coincides with our result, except for the fact that his statement does not seem to require the discretization format of D2. Without the discretization adjustment, Pittel's result is only valid when n and u vary in such a manner that $\log_2 n + u$ is an integer. To see this, note that $\mathbf{P}\{D_{n+1} \le \log_2 n + u\} = \mathbf{P}\{D_{n+1} \le \log_2 n + v\}$ when $\lfloor \log_2 n + u \rfloor = \lfloor \log_2 n + v \rfloor$. Thus, if $|u - v| < 1$, then, along an infinite subsequence, the difference between the two probabilities is 0. This contradicts Pittel's statement, according to which the difference is asymptotically nonzero whenever $u \neq v$.

Height of a trie. The height of the trie can be studied via a Poissonization argument along the lines of Devroye (1984). Some key lemmas from that reference allow us to present a very short proof of the limit law for H_n. Theorem H1 shows that the distribution of $H_n - 2\log_2 n$ is close to a suitably discretized version of the extreme-value distribution $e^{-e^{-x}}$. Ignoring small oscillations due to discretization, we have

$$\mathbf{P}\left\{H_n \le 2\log_2 n + \frac{x - \log 2}{\log 2} + \log_2 \int f^2\right\} \approx e^{-e^{-x}}$$

when f is square integrable. It is interesting to note that f influences the height in the second term only, the main term being $2\log_2 n$, precisely double the main term $\log_2 n$ for D_{n+1}. The uniform density version of Theorem H1 is due to Mendelson (1982).

THEOREM H1. *Assume that $\int f^2 < \infty$. For all $x \in R$ and $k = \lfloor 2\log_2 n + x \rfloor$,*

$$\lim_{n\to\infty} \left|\mathbf{P}\{H_n \le k\} - e^{(-n^2 2^{-k}\int f^2)/2}\right| = 0.$$

Also, $H_n - 2\log_2 n = O(1)$ in probability.

PROOF. We first introduce a Poissonization argument. Let $N(\lambda)$ denote a Poisson random variable with parameter λ. Let n_1 and n_2 be real numbers such that $0 < n_1 < n < n_2 < \infty$, and let S be the collection of all 2^k intervals into which $[0,1)$ can be partitioned. For $1 \le i \le 2^k$, we denote by $q_{i,k}$ the probability mass $\int_{I_{i,k}} f$. Then, by Lemma 2 of Devroye (1984),

$$\mathbf{P}\{H_n \le k\} \le \mathbf{P}\{N(n_1) \ge n\} + \prod_{i=1}^{2^k} (1 + n_1 q_{i,k}) e^{-n_1 q_{i,k}}$$

and

$$\mathbf{P}\{H_n \le k\} \ge \prod_{i=1}^{2^k} (1 + n_2 q_{i,k}) e^{-n_2 q_{i,k}} - \mathbf{P}\{N(n_2) \le n\}.$$

For $\varepsilon \in (0, 1/2)$, $n_1 = n(1-\varepsilon)$ and $n_2 = n(1+\varepsilon)$, the following exponential inequalities were proved in Devroye (1984), Lemma 4:

$$\mathbf{P}\{N(n_2) \le n\} \le e^{-n\varepsilon^2/4},$$

$$\mathbf{P}\{N(n_1) \ge n\} \le e^{-n\varepsilon^2/2}.$$

If we combine these inequalities and use $(1+u)e^{-u} \le e^{-u^2/(2(1+u))}$, $u > 0$

[Devroye (1984) Lemma 3], then

$$\mathbf{P}\{H_n \le k\} \le e^{-n\varepsilon^2/2} + \exp\left(-\sum_{i=1}^{2^k} (n_1 q_{i,k})^2/(2(1+n_1 q_{i,k}))\right)$$
$$\le e^{-n\varepsilon^2/2} + \exp\left(-\sum_{i=1}^{2^k} (n_1 q_{i,k})^2/(2(1+\varepsilon))\right)$$
$$\times \exp\left(\frac{1}{2}\sum_{i=1}^{2^k} (n_1 q_{i,k})^2 I_{[n_1 q_{i,k} > \varepsilon]}\right).$$

But $\sum_{i=1}^{2^k} q_{i,k}^2 \sim 2^{-k}\int f^2$ as $n \to \infty$ (Lemma E2). Also,

$$\sum_{i=1}^{2^k} (n_1 q_{i,k})^2 I_{[n_1 q_{i,k} > \varepsilon]} \le n^2 \sum_{i=1}^{2^k} 2^{-k} \int_{I_{i,k}} f^2 I_{[n q_{i,k} > \varepsilon]}$$
$$\le n^2 2^{-k} \int f^2(x) I_{[n\int_{x-2^{-k}}^{x+2^{-k}} f > \varepsilon]}\,dx = o(n^2 2^{-k})$$

when $k \to \infty$ in such a way that $|k - 2\log_2 n| \le M < \infty$ for some constant M. To get this, we used the fact that $\int f^2 < \infty$, the Lebesgue dominated convergence theorem and the observation that for almost all x,

$$2^{k-1}\int_{x-2^{-k}}^{x+2^{-k}} f(y)\,dy \to f(x).$$

Consequently,

$$\mathbf{P}\{H_n \le k\}$$
$$\le o(1) + \exp\left(-(1-\varepsilon+o(1))^2 n^2 2^{-k}\int f^2/(2(1+\varepsilon)) + o(n^2 2^{-k})\right).$$

Choose $k = \lfloor 2\log_2 n + x \rfloor$, so that $2^{1-x} \ge n^2 2^{-k} \ge 2^{-x}$. Then

$$\mathbf{P}\{H_n \le k\} \le (1+o(1))\exp\left(-(1-\varepsilon)^2 n^2 2^{-k}\int f^2/(2(1+\varepsilon))\right),$$

which is as close to $e^{(-n^2 2^{-k}\int f^2)/2}$ as desired by our choice of ε and the fact that $n^2 2^{-k}$ remains bounded away from 0 and ∞. Similarly, for the lower bound, using $(1+u)e^{-u} \ge e^{-u^2/2}$ $u > 0$, we have

$$\mathbf{P}\{H_n \le k\} \ge \prod_{i=1}^{2^k} (1 + n_2 q_{i,k}) e^{-n_2 q_{i,k}} - e^{-n\varepsilon^2/4}$$
$$\ge \prod_{i=1}^{2^k} e^{-(n_2 q_{i,k})^2/2} - o(1)$$
$$= \exp\left(-\sum_{i=1}^{2^k} (n_2 q_{i,k})^2/2\right) - o(1)$$
$$\ge \exp\left(-(1+\varepsilon)^2 n^2 2^{-k}\int f^2/2\right) - o(1),$$

which once again is as close to $e^{-(n^2 2^{-k}\int f^2)/2}$ as desired by our choice of ε. $\square$

THEOREM H2. *For any density f with $\int f^2 < \infty$, $\mathbf{E}H_n = 2\log_2 n + O(1)$.*

PROOF. Mimic the proof of Theorem 4 of Devroye (1984) and combine it with the exponential bounds given in the proof of Property H1, provided that one takes $k = [(2 - \delta)\log_2 n]$ and $k = [(2 + \delta)\log_2 n]$ respectively, for arbitrary small $\delta > 0$. □

For the class of densities with $\int f^p < \infty$, $1 < p < 2$, we have:

THEOREM H3. *Assume $\int f^p < \infty$ for some $p \in (1, 2]$. Then*

$$\mathbf{P}\{D_{n+1} > k\} \leq Cn2^{-k(2(p-1)/p)},$$

where $C = (\int f^p)^{2/p}$. Also,

$$\lim_{n \to \infty} \mathbf{P}\left\{H_n > \left(\frac{p}{p-1} + \varepsilon\right)\log_2 n\right\} = 0$$

for all $\varepsilon > 0$.

PROOF. From Theorem E4 and Lemma E2,

$$\mathbf{P}\{D_{n+1} > k\} \leq n\sum_{i=1}^{2^k} q_{i,k}^2 \leq n2^{-2k(p-1)/p}\left(\int f^p\right)^{2/p}.$$

This proves the first part of the theorem. The second part follows from

$$\mathbf{P}\{H_n > k\} \leq n\mathbf{P}\{D_{n+1} > k\} \leq Cn^2 2^{2k(1-p)/p},$$

which tends to 0 when we choose $k \geq (p/(p-1) + \varepsilon)\log_2 n$ for some $\varepsilon > 0$. □

We can push things a bit further in the direction of more peaked densities.

THEOREM H4. *Let $\int f \log^{1+a}(1 + f) < \infty$ for some $a > 0$. Then $H_n = O_p(n^{2/(1+a)})$.*

PROOF. Let Ψ: $[0, \infty) \to [0, \infty)$ a strictly increasing convex function. Then, by Jensen's inequality, $q_{i,k} \leq 2^{-k}\Psi^{\text{inv}}(2^k \int \Psi(f))$. Thus

$$\mathbf{P}\{D_{n+1} > k\} \leq n2^{-k}\Psi^{\text{inv}}\left(2^k \int \Psi(f)\right).$$

With $\Psi(u) \equiv u\log^{1+a}(1 + u)$, $a > 0$, we can verify that $\mathbf{P}\{D_{n+1} > k\} \leq cnk^{-(1+a)}$ for some constant c. Therefore, $\mathbf{P}\{H_n > k\} \leq cn^2k^{-(1+a)}$. This tends to 0 when $k/n^{2/(1+a)} \to \infty$. □

We finally note that $\mathbf{E}H_n = \infty$ for all $n \geq 2$ if and only if $\mathbf{E}D_2 = \infty$, so that the pathological cases are again described by Theorem E3. Otherwise (i.e., if $\mathbf{E}D_2 < \infty$), we have $\mathbf{E}H_n \leq n\mathbf{E}D_{n+1} = o(n^2)$. If f is nonincreasing with distribution function F, then trivial arguments show that $\mathbf{P}\{H_n > \log_2(1/F^{\text{inv}}(2/\lfloor n/2 \rfloor))\} \geq 1/2$ for all $n \geq 4$, so that it is impossible to have any kind of universal upper bound on the weak convergence rate of H_n. For example, there exists an f such that for all $n \geq 4$, $\mathbf{P}\{H_n > 2^{2^{2^n}}\} \geq 1/2$. This should be contrasted with the observation that for any f, $D_n/\log_2 n \to 1$ in probability.

Strong convergence results. Assume that $\int f^2 < \infty$. The objective of this section is to show that H_n does not deviate a lot from $2\log_2 n$ as we let the trie grow ($n \to \infty$). But the same is not true for D_n. We recall from Theorem D3 that $\liminf D_n/\log_2 n \leq 1$ almost surely. However, $\limsup D_n/\log_2 n = 2$ almost surely, and for the uniform density, $\liminf D_n/\log_2 n = 1$ almost surely. Similar results for the alphabetic model were obtained by Pittel (1985). In fact, the almost sure behavior of $H_n/\log_2 n$ and of $D_n/\log_2 n$ (lim sup only) matches that of the uniform density model, as long as $\int f^2 < \infty$.

THEOREM S1. *Assume that* $\int f^2 < \infty$. *Then* $\lim_{n\to\infty} H_n/\log_2 n = 2$ *almost surely and* $\limsup_{n\to\infty} D_n/\log_2 n = 2$ *almost surely.*

Theorem S1 follows from Lemmas S2, S4 and S6. Note that the limit infimum of $D_n/\log_2 n$ is related to the behavior of the density f when $f(x)$ is near 0. It is less important since it furnishes virtually no information about the average or worst-case behavior of random tries.

LEMMA S2. *If* $\int f^2 < \infty$, *then* $\limsup_{n\to\infty} H_n/\log_2 n \leq 2$ *almost surely.*

PROOF. If $a_n = \lceil (2+\varepsilon)\log_2 n \rceil$ with $\varepsilon > 0$, we have from Theorem E4,

$$\begin{aligned}
\mathbf{P}\left\{\bigcup_{n=N}^{\infty} [H_n > a_n]\right\} &\leq \mathbf{P}\{H_N > a_N\} + \sum_{n=N}^{\infty} \mathbf{P}\{H_{n+1} > a_{n+1}, H_n \leq a_n\} \\
&\leq N\mathbf{P}\{D_N > a_N\} + \sum_{n=N}^{\infty} \mathbf{P}\{D_{n+1} > a_{n+1}\} \\
&\leq N\mathbf{P}\{D_{N+1} > a_N\} + \sum_{n=N+1}^{\infty} \mathbf{P}\{D_{n+1} > a_n\} \\
&\leq N^2 2^{-a_N} \int f^2 + \sum_{n=N+1}^{\infty} n 2^{-a_n} \int f^2 \\
&\leq N^{-\varepsilon} \int f^2 + \sum_{n=N+1}^{\infty} n^{-(1+\varepsilon)} \int f^2 \\
&\leq (1 + 1/\varepsilon) N^{-\varepsilon} \int f^2. \qquad \square
\end{aligned}$$

The explicit inequality obtained in the proof above may be of interest in its own right. Note also that a simple Borel–Cantelli-type argument applied to bounds for $\mathbf{P}\{H_n > a_n\}$ would only yield $H_n/\log_2 n \leq 3$ almost surely. The same technique coupled with Lemma H3 shows that when $\int f^p < \infty$, then

$\limsup H_n/\log_2 n \le p/(p-1)$ almost surely for all $p \in (1,2]$. Similarly, when $\int f \log^{1+a} f < \infty$ for some $a > 0$, then, almost surely, $\limsup \log H_n/\log n \le 2/(1+a)$, that is H_n is almost surely smaller than $n^{(2+\varepsilon)/(1+a)}$ for any $\varepsilon > 0$ and all n large enough. The lower bound to complement the upper bound of Lemma S2 can be obtained via a Poissonization argument as in Theorem H1, but we prefer to give a different, more instructive proof, which yields useful information even for densities with $\int f^2 = \infty$. First, we recall an inequality for unions of events in a form due to Chung and Erdös (1952).

LEMMA S3. *Let $\{A_i\}$ be a sequence of events. Then*

$$\mathbf{P}\left\{\bigcup_i A_i\right\} \ge \frac{(\Sigma_i \mathbf{P}\{A_i\})^2}{\Sigma_i \mathbf{P}\{A_i\} + \Sigma_{i \ne j} \mathbf{P}\{A_i \cap A_j\}}.$$

LEMMA S4. $\limsup_{n\to\infty} D_n/\log_2 n = \limsup_{n\to\infty} H_n/\log_2 n$.

PROOF. $D_{n+1} = H_{n+1}$ almost surely when $H_{n+1} > H_n$, which happens i.o. almost surely. Thus

$$\limsup_{n\to\infty} \frac{D_n}{\log_2 n} = \limsup_{n\to\infty} \frac{H_n}{\log_2 n}. \qquad \square$$

LEMMA S5. *When $\varepsilon \in (0,1)$ and $\int f^2 < \infty$, then*

$$\mathbf{P}\{H_n \le (2-\varepsilon)\log_2 n\} \le \frac{(\sqrt{32} + o(1))n^{-\varepsilon/2}}{\sqrt{\int f^2}}.$$

When $\int f^2 = \infty$, then $\mathbf{P}\{H_n \le (2-\varepsilon)\log_2 n\} = o(n^{-\varepsilon/2})$.

PROOF. First, assume $\int f^2 < \infty$. Define $k = \lceil (2-\varepsilon)\log_2 n \rceil$, and for two indices $i \ne j$, $1 \le i, j \le n$, let A_{ij} be the event that the trie formed by X_i and X_j has height greater than k; that is, it is the event that the first k bits in the expansions of X_i and X_j are identical. We have the fundamental identity

$$\mathbf{P}\{H_n > k\} = \mathbf{P}\left\{\bigcup_{i \ne j} A_{ij}\right\}.$$

A lower bound for this is obtained via Lemma S3. From the proof of Theorem H1, we recall that

$$p \stackrel{\text{def}}{=} \mathbf{P}\{A_{ij}\} = \sum_{i=1}^{2^k} \left(\int_{I_{i,k}} f\right)^2 \sim 2^{-k} \int f^2.$$

Also, if the indices i, j, l, m are all different, then $\mathbf{P}\{A_{ij} \cap A_{lm}\} = p^2$. Furthermore, if i, j and l are different, Lemma E2 implies that

$$\mathbf{P}\{A_{ij} \cap A_{il}\} = \sum_{i=1}^{2^k} \left(\int_{I_{i,k}} f\right)^3 \le p^3/2.$$

Also,

$$|\{(i,j): i \neq j, 1 \leq i,j \leq n\}| = n(n-1),$$
$$|\{(i,j,l,m): i,j,l,m \text{ all different}, 1 \leq i,j,l,m \leq n\}|$$
$$= n(n-1)(n-2)(n-3),$$
$$|\{(i,j,l): i,j,l \text{ all different}, 1 \leq i,j,l \leq n\}| = n(n-1)(n-2).$$

By Lemma S3 and a combinatorial argument,

$$\begin{aligned}\mathbf{P}\left\{\bigcup_{i \neq j} A_{ij}\right\} &\geq (n(n-1)p)^2 \\ &\qquad \times\{n(n-1)p + n(n-1)(n-2)(n-3)p^2 \\ &\qquad\quad + 2n(n-1)p + 4n(n-1)(n-2)p^3/2\}^{-1} \\ &= \frac{n(n-1)p}{3 + (n-2)(n-3)p + 4(n-2)\sqrt{p}}.\end{aligned}$$

Thus, since $pn^{2-\varepsilon}$ remains bounded away from 0 and ∞,

$$\begin{aligned}\mathbf{P}\{H_n \leq k\} &\leq \frac{3 + (n-2)(n-3)p + 4(n-2)\sqrt{p} - n(n-1)p}{3 + (n-2)(n-3)p + 4(n-2)\sqrt{p}} \\ &\sim \frac{4}{n\sqrt{p}} \leq \frac{4 + o(1)}{n\sqrt{\int f^2}} 2^{k/2} \leq \frac{\sqrt{32} + o(1)}{\sqrt{\int f^2}} n^{-\varepsilon/2}.\end{aligned}$$

Assume next $\int f^2 = \infty$. With k and ε as before and using the fact that $pn^{2-\varepsilon} \to \infty$ and $p = o(1)$, we obtain without effort that $P\{H_n \leq k\} = o(n^{-\varepsilon/2})$. □

LEMMA S6. *For any f, $\liminf_{n \to \infty} H_n/\log_2 n \geq 2$ almost surely.*

PROOF. Fix $\varepsilon \in (0,1)$. Assume first that $\int f^2 < \infty$. We use a simple dyadic argument and the bound obtained in Lemma S5: $\mathbf{P}\{H_n \leq (2-\varepsilon)\log_2 n\} \leq Cn^{-\varepsilon/2}$ for some constant $C > 0$ depending upon f and ε only. Let N be so large that $N \geq 2^{(2-2\varepsilon)/\varepsilon}$. Then, by the monotonicity of H_n,

$$\begin{aligned}\mathbf{P}\left\{\bigcup_{n \geq N} [H_n \leq (2-2\varepsilon)\log_2 n]\right\} &\leq \mathbf{P}\left\{\bigcup_{i=0}^{\infty} [H_{N2^i} \leq (2-2\varepsilon)\log_2(N2^i + 1)]\right\} \\ &\leq \mathbf{P}\left\{\bigcup_{i=0}^{\infty} [H_{N2^i} \leq (2-\varepsilon)\log_2(N2^i)]\right\} \\ &\leq \sum_{i=0}^{\infty} \mathbf{P}\{H_{N2^i} \leq (2-\varepsilon)\log_2(N2^i)\} \\ &\leq \sum_{i=0}^{\infty} C(N2^i)^{-\varepsilon/2} \\ &= CN^{-\varepsilon}(1 - 2^{-\varepsilon/2})^{-1}.\end{aligned}$$

This tends to 0 with N. □

Large deviation results. We will require sharp estimates of the large deviation type for the tails of the distribution of H_n. The following result suffices.

THEOREM L1. *Consider an integer sequence $k = k_n \to \infty$ for which $n2^{-k} \to 0$ and $n^2 2^{-k} \to \infty$, then, if $\int f^2 < \infty$, we have*

$$\mathbf{P}\{H_n \le k_n\} = \exp\left(-\left(\frac{1}{2}\int f^2 + o(1)\right)n^2 2^{-k}\right).$$

If $k = k_n \to \infty$ in such a way that $n^2 2^{-k} \to 0$ and if $\int f^2 < \infty$, then

$$\mathbf{P}\{H_n > k_n\} = \left(\frac{1}{2}\int f^2 + o(1)\right)n^2 2^{-k}.$$

PROOF. We argue as in the proof of Theorem H1, where we choose $\varepsilon = \varepsilon_n$ in such a way that $\varepsilon \to 0$ and $n\varepsilon_n^2/(n^2 2^{-k}) \to \infty$. In trying to find asymptotics for $\mathbf{P}\{H_n \le k\}$, we verify that

$$\sum_{i=1}^{2^k} (n_1 q_{i,k})^2 I_{[n_1 q_{i,k} > \varepsilon]} \le n^2 2^{-k} \int_{x:\, n\int_{x-2^{-k}}^{x+2^{-k}} f > \varepsilon} f^2(x)\, dx = o(n^2 2^{-k}),$$

since for almost all x, $2^{k-1}\int_{x-2^{-k}}^{x+2^{-k}} f \to f(x)$ provided that $\int f^2 < \infty$, and since $n2^{-(k+1)}/\varepsilon \to 0$. Thus

$$\mathbf{P}\{H_n \le k\} = \exp(-n^2 2^{-k}/o(1)) + \exp\left(-(1+o(1))\frac{1}{2}\int f^2 n^2 2^{-k}\right).$$

Furthermore,

$$\mathbf{P}\{H_n > k\} \le \binom{n}{2}\mathbf{P}\{H_2 > k\} \le \frac{1}{2}n^2 \sum_{i=1}^{2^k} q_{i,k}^2 \le \frac{1}{2}n^2 2^{-k}\int f^2.$$

Finally, return again to the proof of Theorem H1 and pick $\varepsilon = \varepsilon_n$ in such a way that $\varepsilon \to 0$, $\varepsilon^2/(n2^{-k}) \to \infty$. Then, using the fact that $\log(1+u) - u \le -u^2/2 + u^3/3$ for $u \ge 0$, we have

$$\begin{aligned}\mathbf{P}\{H_n > k\} &\ge 1 - \prod_{i=1}^{2^k}(1 + n_1 q_{i,k})e^{-n_1 q_{i,k}} - \mathbf{P}\{N(n_1) \ge n\}\\ &\ge 1 - \prod_{i=1}^{2^k} e^{-(n_1 q_{i,k})^2/2 + (n_1 q_{i,k})^3/3} - e^{-n_1\varepsilon^2/2}\\ &= 1 - \exp\left(-\sum_{i=1}^{2^k}\frac{1}{2}(n_1 q_{i,k})^2\right)\exp\left(\sum_{i=1}^{2^k}\frac{1}{3}(n_1 q_{i,k})^3\right) - e^{-n_1\varepsilon^2/2}\\ &\overset{\text{def}}{=} 1 - e^{-a_n}e^{b_n} - e^{-c_n}.\end{aligned}$$

Obviously, the lower bound is approximately a_n when $a_n > 0$, $b_n = o(a_n)$ and $a_n/c_n \to 0$. But we already know that $a_n \sim \frac{1}{2}n^2 2^{-k}\int f^2 \to 0$. Also, by our choice of ε, we verify that $n\varepsilon^2/(n^2 2^{-k}) \to \infty$, so that $a_n = o(c_n)$. By Lemma E2,

$$3b_n \le n^3 \sum_{i=1}^{2^k} q_{i,k}^3 \le n^3 \left(2^{-k} \int f^2\right)^{3/2} \sim a_n^{3/2},$$

and thus $b_n = o(a_n)$. We conclude that $\mathbf{P}\{H_n > k\} \ge (1 + o(1))\frac{1}{2}\int f^2 n^2 2^k$. □

Laws of the iterated logarithm. It is interesting to note that our problem has so much structure that we are able to obtain laws of the iterated logarithm for H_n, enabling us therefore to tell how wide the "swings" are of H_n as n grows large. In this section, we only assume that $\int f^2 < \infty$. For sequences a_n and b_n that increase to ∞, we need to be able to tell whether $H_n > a_n$ infinitely often (i.o.) or finitely often (f.o.), and whether $H_n \le b_n$ i.o. or f.o. Hence the need to consider four distinct problems. We begin by noting that the upper-class behavior of H_n is not affected by the density at all, as long as $\int f^2 < \infty$.

THEOREM I1. *Let a_n be monotone $\uparrow$ and assume that $\int f^2 < \infty$ and $n^2 2^{-a_n} \to 0$. Then*

$$\mathbf{P}\{H_n > a_n \ i.o.\} = \begin{cases} 0, & \text{if } \sum_n n 2^{-a_n} < \infty, \\ 1, & \text{if } \sum_n n 2^{-a_n} = \infty. \end{cases}$$

If we write $\log_2^{(p)}$ for the p times iterated logarithm base 2, then we have

$$\mathbf{P}\left\{H_n > 2\log_2 n + \log_2^{(2)} n + \cdots + (1+\delta)\log_2^{(m)} n \ i.o\right\} = 0 \text{ or } 1$$

according to $\delta > 0$ or $\delta = 0$. In particular,

$$\limsup_{n\to\infty} \frac{H_n - 2\log_2 n}{\log_2 \log n} = 1 \quad \text{almost surely.}$$

PROOF. Since $H_n \uparrow$, we have for integer N,

$$\bigcup_{n \ge N} [H_n > a_n] = \bigcup_{n \ge N} [D_n > a_n] \cup [H_N > a_N].$$

If $D_n > a_n$ i.o. almost surely, then $H_n > a_n$ i.o. almost surely. If $D_n > a_n$ f.o. almost surely, then $H_n > a_n$ f.o. almost surely. Then $H_n > a_n$ i.o. almost surely if and only if $D_n > a_n$ i.o. almost surely. In particular,

$$\frac{\limsup H_n}{a_n} = \frac{\limsup D_n}{a_n} \quad \text{almost surely.}$$

If $\sum_n \mathbf{P}\{D_n > a_n\} < \infty$, then $\mathbf{P}\{H_n > a_n \text{ i.o.}\} = 0$. By Theorem E4,

$$\mathbf{P}\{D_n > a_n\} \le n 2^{-a_n} \int f^2.$$

If this is summable in n, then we have that $H_n > a_n$ f.o. with probability 1.

Assume next that $\sum_n n2^{-a_n} = \infty$ and that $a_n \uparrow \infty$ in such a way that $n^2 2^{-a_n} \to 0$. Let us split the data sequence into parts of sizes 1, 2, 4, 8 and so forth, and consider the sequence of (independent) tries formed in this manner. The heights of these tries are denoted by Z_1, Z_2 and so forth, so that Z_k is distributed as $H_{2^{k-1}}$. Also, if $n_k = 2^k - 1$, we see that $H_{n_k} \geq \max_{1 \leq i \leq k} Z_i \geq Z_k$. By the Borel–Cantelli lemma, $H_n > a_n$ i.o. almost surely if

$$\sum_{k=1}^{\infty} \mathbf{P}\{Z_k > a_{n_k}\} = \infty.$$

This is equivalent to

$$\sum_{k=1}^{\infty} \mathbf{P}\{H_{2^{k-1}} > a_{2^k-1}\} = \infty.$$

If $2^{2(k-1)} 2^{-a_{2^k-1}} \to 0$ (which holds in view of $n^2 2^{-a_n} \to 0$), then we can apply the large deviation estimate of Theorem L1, and reduce the above condition to

$$\sum_{k=1}^{\infty} 2^{2(k-1)} 2^{-a_{2^k-1}} = \infty.$$

We now show that this is indeed satisfied. By the monotonicity of a_n,

$$\sum_{k=1}^{\infty} 2^{2k} 2^{-a_{2^k-1}} = \sum_{k=1}^{\infty} \sum_{j=2^k}^{2^{k+1}-1} 2^k 2^{-a_{2^k-1}} \geq \sum_{k=1}^{\infty} \sum_{j=2^k}^{2^{k+1}-1} j 2^{-a_j} = \sum_{j=2}^{\infty} j 2^{-a_j} = \infty. \qquad \square$$

THEOREM I2. *Assume that $\int f^2 < \infty$. Then, for all $\varepsilon > 0$,*

$$\mathbf{P}\left\{ H_n \leq \left\lfloor 2 \log_2 n - \log_2 \log\log n - \log_2\left((1+\varepsilon) 2 \Big/ \int f^2 \right) \right\rfloor \ i.o. \right\} = 0.$$

Thus, almost surely,

$$\liminf_{n \to \infty} (H_n - 2\log_2 n - \log_2 \log\log n) \geq \log_2\left(2 \Big/ \int f^2 \right) - 1.$$

PROOF. Let $X_n \uparrow$ be a sequence of monotone random variables. Note that for all N,

$$\bigcup_{n \geq N} [X_n \leq a_n] \subseteq \bigcup_{n > N:\, a_n > a_{n-1}} [X_{n-1} \leq a_n] \cup [X_N \leq a_N].$$

Take probabilities and let $N \to \infty$ to conclude that $\mathbf{P}\{X_n \leq a_n \text{ i.o.}\} = 0$ if $\mathbf{P}\{X_n \leq a_n\} \to 0$ and $\sum_{n:\, a_n > a_{n-1}} \mathbf{P}\{X_{n-1} \leq a_n\} < \infty$.

Define $a_n = \lfloor 2\log_2 n - \log_2 \log\log n - \log_2((1+\varepsilon)2/\int f^2) \rfloor$ and assume that n is large enough so that this is well defined. We have $\mathbf{P}\{H_n \leq a_n \text{ i.o.}\} = 0$ if $\mathbf{P}\{H_n \leq a_n\} \to 0$ and $\sum_{n:\, a_n > a_{n-1}} \mathbf{P}\{H_{n-1} \leq a_n\} < \infty$. With our choice of a_n, we have $n^2 2^{-a_n} \to \infty$ and $n 2^{-a_n} \to 0$. From Theorem L1, we thus have the estimate

$$\mathbf{P}\{H_{n-1} \leq a_n\} \leq (\log n)^{-1-\varepsilon-o(1)}.$$

We need only verify that this is summable over all n with $a_n > a_{n-1}$. Let $\{n_j\}$ be the smallest index such that $a_n = j$. We verify easily that there are positive constants c, d such that $c \le n_j/(2^j/2\log j) \le d$. Thus we need to check the summability of

$$\sum_j (\log n_j)^{-1-\varepsilon-o(1)}.$$

But this is clearly the case for any fixed $\varepsilon > 0$. □

THEOREM I3. *Assume that* $\int f^2 < \infty$. *Then, for all* $\varepsilon > 0$,

$$\mathbf{P}\left\{H_n \le \left\lceil 2\log_2 n - \log_2 \log\log n - \log_2\left(2\Big/\int f^2\right) + \varepsilon \right\rceil \ i.o.\right\} = 1.$$

Thus, almost surely,

$$\liminf_{n\to\infty} (H_n - 2\log_2 n - \log_2 \log\log n) \le \log_2\left(2\Big/\int f^2\right) + 1.$$

PROOF. Define $n_j \underset{\text{def}}{=} j^{2j}$. We will show that almost surely, $H_{n_j} \le a_{n_j}$ infinitely often, where $a_n = \lceil 2\log_2 n - \log_2 \log\log n - \log_2(2/\int f^2) + \varepsilon\rceil$. We also need random variables V_j and W_j defined as follows: V_j is the height formed by the trie based upon all data points X_i with $n_{j-1} < i \le n_j$, and W_j is the maximal depth of insertion of the elements X_i, $i > n_{j-1}$, in the trie formed by $X_1, \dots, X_{n_{j-1}}$ (thus insert each of these elements and delete it immediately). Let Z_j be the height of the latter trie. We have $H_{n_j} = \max(V_j, W_j, Z_j)$. For $H_{n_j} \le a_{n_j}$ infinitely often, it suffices that $V_j \le a_{n_j}$ infinitely often and $W_j > a_{n_j}$ finitely often and $Z_j > a_{n_j}$ finitely often. By three applications of the Borel–Cantelli lemma, we see that this is true if

$$\sum_j \mathbf{P}\{V_j \le a_{n_j}\} = \infty,$$

$$\sum_j \mathbf{P}\{W_j > a_{n_j}\} < \infty,$$

$$\sum_j \mathbf{P}\{Z_j > a_{n_j}\} < \infty.$$

Note that

$$\begin{aligned}
\mathbf{P}\{W_j > a_{n_j}\} &\le n_{j-1} n_j \mathbf{P}\{H_2 > a_{n_j}\} \\
&\le n_{j-1} n_j 2^{-a_{n_j}} \int f^2 \qquad \text{(Theorem E4)} \\
&\le \frac{n_{j-1}}{n_j} 2\log\log n_j \\
&\le \frac{(j-1)^{2j-2}}{j^{2j}} 2\log(2j\log j) \\
&\le \frac{2\log(2j\log j)}{(j-1)^2},
\end{aligned}$$

which is summable in $j \geq 2$. Also, $\mathbf{P}\{Z_j > a_{n_j}\} \leq n_{j-1}^2 \mathbf{P}\{H_2 > a_{n_j}\}$, and this too is summable in j by the argument used for W_j. Finally, since $n_{j-1} = o(n_j)$, $n^2 2^{-a_n} \to \infty$ and $n 2^{-a_n} \to 0$, we can apply Theorem L1 and obtain

$$\begin{aligned}
\mathbf{P}\{V_j \leq a_{n_j}\} &= \mathbf{P}\{H_{n_j - n_{j-1}} \leq a_{n_j}\} \\
&\geq \exp\left(- \left(\frac{1}{2} \int f^2 + o(1) \right) (n_j - n_{j-1})^2 2^{-a_{n_j}} \right) \\
&= \exp\left(- \left(\frac{1}{2} \int f^2 + o(1) \right) n_j^2 2^{-a_{n_j}} \right) \\
&\geq e^{-(1+o(1)) \log\log n_j 2^{-\varepsilon}} \\
&= e^{-(2^{-\varepsilon}+o(1))} \log(2j \log j) \\
&= (2j \log j)^{-(2^{-\varepsilon}+o(1))},
\end{aligned}$$

and this is not summable in j, as required. $\square$

In the lower-class behavior of H_n, we observe that the density f affects the constant term only. We have for all square integrable f,

$$\liminf_{n \to \infty} \frac{H_n - 2\log_2 n}{\log_2 \log\log n} = -1 \quad \text{almost surely}.$$

Also, if $\int f^2 = \infty$, $\liminf(H_n - 2\log_2 n + \log_2 \log\log n) = \infty$ almost surely. Finally, Theorems I2 and I3 together pin down the lower classes for H_n modulo unavoidable discretization factors due to the fact that H_n is integer valued. Summarizing, we note that the upswings of $H_n - 2\log_2 n$ are about $\log_2 \log n$, and the downswings about $-\log_2 \log\log n$.

PATRICIA. PATRICIA is a space-efficient improvement of the classical trie discovered by Morrison 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. Also, $H_n \leq n/2$. The trie from which it is deduced is called the associated trie. Also, all parameters of PATRICIA such as D_n and H_n have to improve over those of the associated trie, regardless of which density drives the data. The object of our analysis then is to show how much PATRICIA improves over the trie. For the uniform density, Pittel (1985) has shown that $H_n/\log_2 n \to 1$ almost surely, which constitutes a 50% improvement over the trie. Thus it is of interest to see for just how large a class of densities H_n is close to $\log_2 n$.

For the uniform density, $\mathbf{E}D_{n+1}$ was studied in Knuth (1973), Flajolet and Sedgewick (1986), Kirschenhofer and Prodinger (1986) and Szpankowski (1988). The variance of D_{n+1} was studied in the latter two references, while the variance of the external path length was obtained in Kirschenhofer, Prodinger and Szpankowski (1989a). The structure has never been studied under the density model. The first remarkable property of PATRICIA is that $\mathbf{E}D_{n+1} = o(n)$ for any density. Recall that for tries, we can have $\mathbf{E}D_2 = \infty$.

Hence the pathological case corresponding to densities described in Theorem E3 no longer exists.

THEOREM P1. *For any density f,* PATRICIA *behaves such that* $\mathbf{E}H_n = o(n)$ *and* $H_n/n \to 0$ *almost surely.*

PROOF. Since $H_n \le n$, $\mathbf{E}H_n = o(n)$ whenever $H_n/n \to 0$ in probability. So we consider only the strong convergence. Let $N_{i,k}$ ($N_{x,k}$) denote the number of X_i's belonging to $I_{i,k}$ ($A_{x,k}$). Let $\varepsilon > 0$ be arbitrary and let M be an integer picked so large that $\sup_x q_{x,M} \le \varepsilon/3$. Let n be so large that $M \le \varepsilon n/3$. We apply an inequality due to Bennett (1962) [see also Hoeffding (1963)] which states that for a binomial (n, p) random variable Z,

$$\mathbf{P}\{Z - \mathbf{E}Z \ge n\varepsilon\} \le e^{-n\varepsilon((1+p/\varepsilon)\log(1+\varepsilon/p)-1)}.$$

For $p \le \varepsilon$, the upper bound does not exceed $(e/4)^{n\varepsilon}$. For all x, $\mathbf{E}N_{x,M} = n\sup_x q_{x,M} \le n\varepsilon/3$. Thus

$$\begin{aligned}\mathbf{P}\{H_n > n\varepsilon\} &\le \mathbf{P}\left\{\bigcup_{i=1}^{2^M} I_{[N_{i,M} \le n\varepsilon - M]}\right\}\\ &\le \sum_{i=1}^{2^M} \mathbf{P}\{N_{i,M} \ge n\varepsilon - M\} \le 2^M \sup_x \mathbf{P}\{N_{x,M} \ge 2n\varepsilon/3\}\\ &\le 2^M \sup_x \mathbf{P}\{N_{x,M} - \mathbf{E}N_{x,M} \ge n\varepsilon/3\} \le 2^M (e/4)^{n\varepsilon/3}.\end{aligned}$$

The upper bound is summable in n, so we can conclude by the Borel–Cantelli lemma. $\square$

The optimality of Theorem P1 will be established below. It is helpful to have convenient representations of PATRICIA trees such as the one given below. We define the sequence of neighboring buckets by $L_{x,k}$, where $L_{x,k}$ is the kth level bucket to which x would belong if we flipped its kth bit. Observe that $A_{x,k}$ and $L_{x,k}$ are always adjacent buckets. Also, for any x, $\bigcup_{k=1}^{\infty} L_{x,k} = [0,1]$. Let N_k denote the number of X_i's with $1 \le i \le n$ belonging to $L_{k,X_{n+1}}$. A simple argument in terms of internal nodes with two children shows that

$$D_{n+1} = \sum_{k=1}^{\infty} I_{[N_k > 0]}.$$

Note in passing that $D_2 \equiv 1$, and compare with the possibility of $\mathbf{E}D_2 = \infty$ for an ordinary trie. Also,

$$\mathbf{P}\{N_k > 0 | X_{n+1} = x\} = 1 - \left(1 - \int_{L_{x,k}} f\right)^n$$

and

$$\mathbf{E}D_{n+1} = \int f(x) \sum_{k=1}^{\infty} \left(1 - \left(1 - \int_{L_{x,k}} f\right)^n\right) dx.$$

The following obvious lower bound is valid for all f:

$$\mathbf{E}D_{n+1} \geq \sum_{k=1}^{\infty} \int_{I_{1,k}} f\left(1 - \left(1 - \int_{I_{2,k}} f\right)^n\right).$$

Consider the decreasing density $f(x) = c/(x \log^{1+a}(1/x))$, where $a > 0$ is a parameter, $0 \leq x \leq e^{-(1+a)}$ and $c > 0$ is a normalization constant. We recall that for k large enough, $\int_{I_{1,k}} f = bk^{-a}$, where $b = c/(a \log^a 2)$. Also, if $a < 1$,

$$\int_{I_{2,k}} f = b(k-1)^{-a}\left(1 - \left(1 - \frac{1}{k}\right)^a\right) \geq b(k-1)^{-a} a k^{-1} \geq bak^{-(a+1)}.$$

Thus, when K denotes a large integer,

$$\mathbf{E}D_{n+1} \geq \frac{1}{2} \sum_{k \geq K:\, \int_{I_{2,k}} f \geq \log 2/n} \int_{I_1,k} f \geq \frac{b + o(1)}{2(1-a)} \times \left(\frac{abn}{\log 2}\right)^{(1-a)/(1+a)}.$$

The lower bound can be made larger than $n^{1-\varepsilon}$ for any small $\varepsilon > 0$, merely by choosing a small enough. Similarly, for any $a_n \downarrow 0$, however slowly, it is possible to find a density f such that for n large enough, $\mathbf{E}D_{n+1} \geq na_n$. This concludes the proof of the optimality of Theorem P1.

Now for the main result of this section: If $\int f^p < \infty$ for all $p \geq 1$, then $H_n/\log_2 n \to 1$ almost surely. Thus, for all bounded densities and for many unbounded densities, the asymptotic behavior of PATRICIA's height is like that for the uniform density, and improves dramatically (50%) over the associated trie. Theorem P2 below also bounds H_n for those densities for which $\int f^p < \infty$ for some, but not all, p. Observe in particular the improvement over the associated trie, where for all square integrable f (regardless of whether $\int f^{500} < \infty$ for example), $\limsup_{n \to \infty} H_n/\log_2 n = 2$ almost surely.

THEOREM P2. *If $\int f^p < \infty$ for fixed integer $p > 1$, then*

$$\limsup_{n \to \infty} \frac{H_n}{\log_2 n} \leq \frac{p}{p-1}$$

almost surely. In particular, if f is such that $\int f^p < \infty$ for all $p \geq 1$, then $\lim H_n/\log_2 n = 1$ almost surely.

PROOF. We follow a simple argument due to Pittel (1985), page 426. Again, we argue in terms of the original unimproved trie, not PATRICIA. The event $[H_n > k + l]$ implies that there exists a set of l data points X_i with $1 \leq i \leq n$, such that all of them share the same first k bits in their binary expansion. By

Lemma E2,

$$\mathbf{P}\{H_n > k + l\} \le \binom{n}{l}\mathbf{P}\{X_1, \dots, X_l \text{ belong to the same interval in the } k\text{th level partition}\}$$

$$\le \frac{n^l}{l!}\sum_{i=1}^{2^k}\int_{I_{i,k}} f\left(\int_{I_{i,k}} f\right)^{l-1}$$

$$\le \frac{n^l}{l!}2^{-k(l-1)}\int f^l.$$

Consider the monotone random variable H_n and define

$$a_n = \left\lceil (1+\varepsilon)\frac{l}{l-1}\log_2 n \right\rceil.$$

We have

$$\mathbf{P}\left\{\bigcup_{n\ge N}[H_n > a_n]\right\} \le \sum_{n\ge N:\, a_{n+1}>a_n} \mathbf{P}\{H_n > a_n\} + \mathbf{P}\{H_N > a_N\}.$$

The last probability tends to 0 with N in view of the inequality derived above. Let n_j be the largest index such that $a_n < j$. Then it is easy to see that $a \le n/2^{j(l-1)/((1+\varepsilon)l)} \le b$ for some $0 < a < b < \infty$. Also, for all j at least equal to a large constant J, we have $a_{n_j} = j - 1$. Thus

$$\sum_{n:\, a_{n+1}>a_n} \mathbf{P}\{H_n > a_n\} \le J + \sum_{j=J}^{\infty}\mathbf{P}\{H_{n_j} > a_{n_j}\}$$

$$\le J + \sum_{j=J}^{\infty}\frac{n_j^l}{l!}2^{-(a_{n_j}-1-l)(l-1)}\int f^l$$

$$\le J + \sum_{j=J}^{\infty}\frac{b2^{j(l-1)/(1+\varepsilon)}}{l!}2^{-(j-1-1-l)(l-1)}\int f^l$$

$$\le J + \frac{b2^{(l-1)(l+2)}\int f^l}{l!}\sum_{j=J}^{\infty}2^{-\varepsilon(l-1)^j/(1+\varepsilon)} < \infty.$$

Thus $H_n > a_n$ finitely often almost surely. Finally, use the fact that $H_n \ge \lfloor \log_2 n \rfloor$. □

Theorem P2 covers the least peaked densities. For the very peaked densities, we could present a myriad of results, all pointing to the improvement over the ordinary trie. To make the point, we will just present a result for the class of densities with $\int f \log^a(f+1) < \infty$, where $a > 0$ is a fixed parameter (see also Theorem H4). In the trie, we have $H_n = O_p(n^{2/a})$. In contrast, PATRICIA has $H_n = O_p(n^{1/(1+a)})$. For finite entropy densities (case $a = 1$), $H_n = O_p(\sqrt{n})$.

THEOREM P3. *Let* $\int f \log^a(f+1) < \infty$ *for some* $a > 0$. *Then* $\limsup_{n\to\infty} H_n/n^{1/(1+a)} < \infty$ *almost surely.*

PROOF. Let us follow the argument and notation of Theorem P1. By using an inequality from Theorem H4, we have $\sup_x q_{x,k} \le Ak^{-a}$ for some constant A. Take $k = \lceil n\varepsilon \rceil$, where ε depends upon n and will be picked later. To apply the bound derived in the proof of Theorem P1, we need M so large that $A/M^a \le \varepsilon/3$, and n so large that $M \le n\varepsilon/3$. Take $M = \lceil 3A/\varepsilon \rceil^{1/a}$ and note that both requirements are met for all n large enough when $\varepsilon = un^{-a/(1+a)}$ for fixed $u > 3A^{1/(1+a)}$. Thus $k \sim un^{1/(1+a)}$. Then

$$\mathbf{P}\{H_n > k\} \le 2^M (e/4)^{n\varepsilon} \le 2^{(\log_2 e - 5/3)n\varepsilon} \le 2^{-0.22un^{1/(1+a)}}.$$

Therefore, by the Borel–Cantelli lemma, almost surely,

$$\limsup_{n\to\infty} H_n/n^{1/(1+a)} \le 3A^{1/(1+a)}. \qquad \square$$

Digital search trees. Digital search trees are constructed from our data sequence $X_1, \dots, X_n$ by repeated insertion into an initially empty tree. A node travels to the first unoccupied slot (thus X_1 is the root). When a node travels down the tree, its kth bit in its binary expansion determines whether it should go left (0) or right (1). First suggested by Coffman and Eve (1970), it can be represented in a different manner. Let the data be i.i.d. random variables $(X_1, T_1), \dots, (X_n, T_n)$, where the T_i's are independent of the X_i's and are uniform [0, 1] time stamps. The point with the smallest time stamp forms the root. The next point to be inserted is the one with the next smallest time stamp and so forth. This model will be called the random digital search tree (RDST). But we may also consider other models such as the random incremental digital search tree (RIDST) in which $T_i \equiv i$; this is the model typically considered in the literature. It is also possible to consider models in which the T_i's depend upon the data; for example, consider the case in which the data are inserted in order of increasing values. Then we could set $T_i = k$ if X_i is the kth smallest among the X_j's.

A random digital search tree can be obtained from the trie defined by $X_1, \dots, X_n$ as follows: Declare all nodes "unmarked"; grab the leaf with the smallest T_i value and move it toward the root as far as possible without hitting a marked node; mark the node where the point comes to rest (so that it is either the root or its father is a marked node); next, grab the leaf with the smallest T_i value from among the leaves not considered earlier; and repeat the same process until all n leaves are treated. The resulting tree of n marked nodes is a subtree of the original trie. It has $n-1$ edges, so that storagewise, the digital search tree is optimal. Furthermore, for all i, D_{ni} is smaller than or equal to the corresponding quantity in the startup (or *associated*) trie. Similarly, we can define the associated PATRICIA, defined on the same $X_1, \dots, X_n$.

We note first that for both the RDST and the RIDST, H_n has the same distribution. Also, D_{ni} in both models in stochastically smaller than D_{nn} in the RIDST. Thus, for weak convergence results, it suffices to study H_n and D_{nn} in the RIDST. Several results are known for the uniform distribution on [0, 1] in the RIDST. Konheim and Newman (1973) and Knuth (1973) showed that the expected value of $(1/n)\sum_{i=1}^{n} D_{ni}$ is $\log_2 n + O(1)$. Among other things, Flajolet and Sedgewick (1986) showed that the average number of leaves is $(0.3720468\ldots + \omega_n)n$, where $|\omega_n| \le 10^{-6}$ is an oscillating function. Similarly to PATRICIA, Pittel (1985) showed that $H_n/\log_2 n \to 1$ almost surely.

LEMMA T1. *Let D_n be D_{nn} for the* RIDST. *Then for integer $k \ge 1$, we have*

$$\mathbf{P}\{D_{n+1} \ge k\} \le \inf_{k \ge l \ge 1} \frac{2^{l^2}\int f^{l+1}}{l!} (n2^{-k})^l.$$

PROOF. The proof hinges on the following inclusion of events: Given $X_{n+1} = x$,

$$[D_{n+1} \ge k] \subseteq \bigcap_{l=0}^{k} [\mathrm{Card}(A_{x,k-1}) \ge l],$$

where Card is the cardinality function: $\mathrm{Card}(A) = \sum_{j=1}^{n} I_{[X_j \in A]}$. We fix a positive integer $l \le k$, and note that by Lemma E2,

$$\mathbf{P}\{D_{n+1} \ge k\} \le \mathbf{P}\left\{\sum_{j=1}^{n} I_{[X_j \in I_{x_{n+1},k-l}]} \ge l\right\} \le \mathbf{E}\binom{n}{l}\left(\int_{I_{x_{n+1},k-1}} f\right)^l$$

$$\le \frac{n^l}{l!}\sum_{i=1}^{2^l}\left(\int_{I_{i,k-1}} f\right)^{l+1} \le \frac{2^{l^2}\int f^{l+1}}{l!}(n2^{-k})^l. \qquad \square$$

If we take $l = 1$, Lemma T1 is strong enough to imply that whenever $\int f^2 < \infty$, then $D_{n+1} = \log_2 n + O_p(1)$. This result was already known, of course (Theorem E4). From Lemma T1, we have the following theorem without much work, just as for PATRICIA.

THEOREM T2. *For any density f, the* RIDST *and* RDST *have $\mathbf{E}H_n = o(n)$ and $H_n/n \to 0$ almost surely. If $\int f^p < \infty$ for fixed integer $p > 1$, then*

$$\limsup_{n\to\infty} \frac{H_n}{\log_2 n} \le \frac{p}{p-1}$$

almost surely, both for the RDST *and the* RIDST. *In particular, if f is* *such that* $\int f^p < \infty$ *for all* $p \geq 1$, *then* $\lim H_n/\log_2 n = 1$ *almost surely. If* $\int f \log^a(f+1) < \infty$ *for some* $a > 0$, *then* $\limsup_{n\to\infty} H_n/n^{1/(1+a)} < \infty$ *almost surely.*

PROOF. The first statement is obtainable by mimicking the proof of Theorem P1. For the RIDST, we have $H_{n+1} = \max(H_n, D_{n+1})$. Thus we can argue as in Lemma I1 and conclude the proof of the first statement if for every $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} \mathbf{P}\left\{D_{n+1} > \left(\frac{p}{p-1} + \varepsilon\right)\log_2 n\right\} < \infty.$$

By Lemma T1, the nth term in the summation does not exceed a constant depending upon p only times $n^{-(p-1)(p/(p-1)+\varepsilon-1)} = n^{-(p-1)\varepsilon-1}$. This is summable in n. The second statement follows from the first one and the observation that $H_n \geq \lfloor \log_2 n \rfloor$.

For the RDST, H_n does not increase in the same simple manner. Still, H_n increases monotonically, so that we may argue as in the proof of Theorem P2; we also need the fact that $\mathbf{P}\{H_n > k\} \leq n\mathbf{P}\{D_{nn} > k\} \leq n\mathbf{P}\{D_{n+1} > k\}$, where D_{nn} and D_{n+1} refer to the RIDST. □

The expected depth of insertion. There is an essential difference between the weak convergence of D_n and the convergence of $\mathbf{E}D_n$. Indeed, for many densities with infinite peaks, $D_n/\mathbf{E}D_n \to 0$ in probability or almost surely, while only for "nice" densities we have the law of large numbers: $D_n/\mathbf{E}D_n \to 1$ in probability. In view of this discrepancy, $\operatorname{Var} D_n$ is not a good measure of the "spread" of the distribution of D_n. We will thus not focus on the variance. In this section, we take a closer look at all the possible rates at which $\mathbf{E}D_{n+1}$ can diverge. We begin with lower bounds.

LEMMA E1. *If* $\int f \log(f+1) = \infty$, *then both* $\mathbf{E}\{D_{n+1} - \log_2 n\}$ *and* $\mathbf{E}\{D_{n+1} - \log_2 n\}_+$ *tend to* ∞. *In all cases,*

$$\liminf_{n\to\infty} \mathbf{E}\{D_n - \log_2 n\} \geq -3 + \frac{\gamma - H}{\log 2}$$

and $\mathbf{E}D_n \geq \lfloor \log_2 n \rfloor$.

PROOF. The function $u(1-u)^n$ never exceeds the value $1/(n+1)$ for $0 \leq u \leq 1$. Thus, from Lemma D1, we have $\mathbf{P}\{D_{n+1} \leq k\} \leq 2^k/(n+1)$. Thus also,

$$\sum_{k=-\infty}^{\lfloor \log_2 n \rfloor} \mathbf{P}\{D_{n+1} \leq k\} \leq \sum_{k=0}^{\lfloor \log_2 n \rfloor} \frac{2^k}{n+1} \leq \frac{2}{n+1} 2^{\lfloor \log_2 n \rfloor} \leq 2.$$

Thus $\mathbf{E}\{D_{n+1} - \log_2 n\}_- \geq -2$. Let F be the smooth limit distribution of

Remark 2 of the previous section. By Fatou's lemma and Theorem D2,

$$\begin{aligned}
\liminf_{n\to\infty} \mathbf{E}\{D_n - \log_2 n\}_+ &= \liminf_{n\to\infty} \int_0^\infty \mathbf{P}\{D_n - \log_2 n > u\}\, du \\
&\geq \int_0^\infty \liminf_{n\to\infty} \mathbf{P}\{D_n > \log_2 n + u\}\, du \\
&\geq \int_0^\infty \left(1 - \int f(y) e^{-f(y)2^{-u-1}}\, dy\right) du \\
&= \int_0^\infty (1 - F(u+1))\, du = \int_1^\infty (1 - F(u))\, du \\
&\geq \int_0^\infty (1 - F(u))\, du - 1 \\
&\geq \frac{\gamma - H}{\log 2} - 1.
\end{aligned}$$

The last line follows from properties of the smooth limit distribution F defined in Remark 2. It remains valid even if $H = -\infty$. For the last part of the theorem, standard combinatorial arguments show that $(1/n)\sum_{i=1}^n D_{n,i} \geq \lfloor \log_2 n \rfloor$. But $\mathbf{E}D_n = \mathbf{E}D_{n,i}$ for all i, so the same lower bound is valid for $\mathbf{E}D_n$. □

LEMMA E2 (Properties of $q_{i,k}$). *For any nonnegative function ϕ with the property that $u\phi(u)$ is convex, we have*

$$\sum_{i=1}^{2^k} q_{i,k}\phi(q_{i,k}) \leq \int f\phi(2^{-k}f).$$

In particular, for any $\alpha \geq 0$,

$$1 - o(1) \leq \frac{\sum_{i=1}^{2^k} q_{i,k}^{1+\alpha}}{2^{-k\alpha}\int f^{1+\alpha}} \leq 1.$$

Furthermore, for any $p \geq 2$,

$$\sum_{i=1}^{2^k} q_{i,k}^p \leq \left(\sum_{i=1}^{2^k} q_{i,k}^2\right)^{p/2} \leq \left(2^{-k}\int f^2\right)^{p/2},$$

and for $1 \leq p \leq 2$,

$$\sum_{i=1}^{2^k} q_{i,k}^2 \leq \left(\sum_{i=1}^{2^k} q_{i,k}^p\right)^{2/p} \leq \left(2^{-k(p-1)}\int f^p\right)^{2/p}.$$

PROOF. The first inequality follows directly from Jensen's inequality applied for each term $q_{i,k}\phi(q_{i,k})$. The particular case of this inequality is obvious. Next, the asymptotic result follows from Fatou's lemma and the Lebesgue density theorem:

$$\begin{aligned}\liminf_{k\to\infty}\sum_{i=1}^{2^k} 2^{k\alpha}q_{i,k}^{1+\alpha} &= \liminf_{k\to\infty}\int f(x)2^{k\alpha}q_{x,k}^{\alpha}\,dx\\ &\geq \int f(x)\Bigl(\liminf_{k\to\infty} 2^{k\alpha}q_{x,k}^{\alpha}\Bigr)\,dx\\ &= \int f(x)^{1+\alpha}\,dx.\end{aligned}$$

The last two statements follow from the observation that $(\sum_i q_{i,k}^p)^{1/p}$ is nonincreasing in p. □

The class of all densities is dichotomized into two sets A and B, where on A, $\mathbf{E}D_n = o(n)$, while on B, $\mathbf{E}D_{n+1} = \infty$ for all n. There is no intermediate result; for example, there does not exist a density for which $\infty > \mathbf{E}D_{n+1} \geq n$ for all n.

THEOREM E3. *If* $\sum_{k=1}^{\infty}\sum_{i=1}^{2^k}(\int_{I_{i,k}} f)^2 = \infty$, *then* $\mathbf{E}D_{n+1} \geq \mathbf{E}D_2 = \infty$ *for all* $n \geq 1$. *Otherwise*, $\mathbf{E}D_{n+1} = o(n)$.

PROOF. Assume first that $\sum_{k=1}^{\infty}\sum_{i=1}^{2^k}q_{i,k}^2 < \infty$. Let $\varepsilon > 0$ be arbitrary and let K be such a large integer that $\sum_{k=K}^{\infty}\sum_{i=1}^{2^k}q_{i,k}^2 < \varepsilon$. Then, from Lemma D1,

$$\begin{aligned}\mathbf{E}D_{n+1} &= \sum_{k=0}^{\infty}\mathbf{P}\{D_{n+1} > k\} \leq K + \sum_{k=K}^{\infty}\sum_{i=1}^{2^k} q_{i,k}\bigl(1-(1-q_{i,k})^n\bigr)\\ &\leq K + \sum_{k=K}^{\infty}\sum_{i=1}^{2^k} nq_{i,k}^2\\ &\leq K + \varepsilon n.\end{aligned}$$

This proves the second part of the theorem. Next, assume that

$$\sum_{k=1}^{\infty}\sum_{i=1}^{2^k} q_{i,k}^2 = \infty.$$

Observe that $\mathbf{E}D_{n+1} \geq \mathbf{E}D_2$ for all densities and that

$$\mathbf{E}D_2 = \sum_{k=1}^{\infty}\sum_{i=1}^{2^k} q_{i,k}^2 = \infty. \qquad \square$$

The discriminating double sum in Theorem E3 is just the value of $\mathbf{E}D_2$. The class of densities of interest to us is the one with $\mathbf{E}D_2 < \infty$. Unfortunately, this class is not easily characterized in terms of the entropy $H = -\int f \log f$. To get a clearer picture, we offer the following range of analyses:

1. When f is bounded, $\mathbf{E}D_{n+1} - \log_2 n = O(1)$. In fact, the same is true if merely $\int f^2 < \infty$. This case already captures most unbounded densities as well.
2. When $\int f \log^{1+a}(1+f) < \infty$ for some $a > 0$, then $\mathbf{E}D_{n+1} = \log_2 n + O(1)$. This covers all but the most peaked densities.
3. When the entropy is finite, $\mathbf{E}D_{n+1} = O(\log_2 n)$. Otherwise, $\mathbf{E}(D_{n+1} - \log_2 n\} = \infty$.
4. To look at the boundary outlined in Theorem E3, we will consider in more detail the family of decreasing densities on $[0,1]$. For monotone densities, a complete characterization of all densities for which $\mathbf{E}D_2 = \infty$ is provided. Densities will be constructed with $H = -\infty$, $\mathbf{E}D_2 < \infty$, for which $\mathbf{E}D_{n+1}$ grows at any polynomial rate n^b, where $b \in (0,1)$.

THEOREM E4. *When $\int f^{1+\alpha} < \infty$ for some $\alpha \in (0,1]$, then*

$$\mathbf{P}\{D_{n+1} > k\} \le \left(n2^{-k}\right)^{\alpha} \int f^{1+\alpha}.$$

Also,

$$\mathbf{E}D_{n+1} \le \log_2 n + 2 + \frac{1}{\alpha}\log_2\left(\int f^{1+\alpha}(1-2^{-\alpha})^{-1}\right).$$

PROOF. From Lemmas D1 and E5,

$$\begin{aligned}\mathbf{P}\{D_{n+1} > k\} &= \sum_{i=1}^{2^k} q_{i,k}\left(1-(1-q_{i,k})^n\right) \le \sum_{i=1}^{2^k} q_{i,k}\min(nq_{i,k},1)\\ &\le \sum_{i=1}^{2^k} q_{i,k}^{1+\alpha}n^{\alpha} \le \left(n2^{-k}\right)^{\alpha}\int f^{1+\alpha}.\end{aligned}$$

Also, for integer M to be picked later,

$$\begin{aligned}\mathbf{E}D_{n+1} &= \sum_{k=0}^{\infty}\mathbf{P}\{D_{n+1} > k\} \le M + \sum_{k=M}^{\infty}\left(n2^{-k}\right)^{\alpha}\int f^{1+\alpha}\\ &= M + n^{\alpha}\int f^{1+\alpha}2^{-M\alpha}(1-2^{-\alpha})^{-1}.\end{aligned}$$

Take

$$M = \left\lceil \alpha^{-1}\log_2\left(n^{\alpha}\int f^{1+\alpha}(1-2^{-\alpha})^{-1}\right)\right\rceil.$$

Then

$$\mathbf{E}D_{n+1} \le 2 + \alpha^{-1}\log_2\left(n^{\alpha}\int f^{1+\alpha}(1-2^{-\alpha})^{-1}\right). \qquad \square$$

The bound in Theorem E4 shows again the remarkable robustness of the trie with respect to deviations from uniformity. The square integrability of f is crucial in the study of the height of the trie; Theorem E4 will thus be useful there. When $f \notin L_p$ for any $p > 1$, it is still possible to have $\mathbf{E}D_{n+1} = \log_2 n + O(1)$. The class of densities for which this happens is described in Lemma E5. We introduce the Hardy–Littlewood maximal function f^* defined by

$$f^*(x) = \sup_{r>0} \frac{1}{2r} \int_{x-r}^{x+r} f(y)\, dy.$$

We have $f \le f^*$ almost everywhere, and for every $p > 1$, $\int f^{*p} \le 10p/(p-1)\int f^p$ [see, e.g., Stein (1970), page 7, or De Guzman (1975, 1981)]. In particular,

$$\int_{A_{x,k}} f \le \int_{x-2^{-k}}^{x+2^{-k}} f \le 2^{1-k} f^*(x).$$

The finiteness of $\int f \log(f^* + 1)$ is crucial in the arguments that follow. It follows trivially (since $f \le f^*$ almost everywhere) from the finiteness of $\int f^* \log(f^* + 1)$, which in turn is equivalent to the finiteness of $\int f \log^2(f + 1)$ [see, e.g., Stein (1970), page 23]. However, this is much too strong a condition in the present context. We were not able to find a proof of the relatively straightforward result in the literature on maximal functions, so we include a proof here. It should be noted that the class F^* includes all f satisfying, for some $\varepsilon > 0$, one of the following conditions:

$$\int f \log^{1+\varepsilon}(f+1) < \infty,$$

$$\int f \log(f+1) \log^{1+\varepsilon} \log(f+e) < \infty,$$

$$\int f \log(f+1) \log\log(f+e) \log^{1+\varepsilon} \log\log(f+e^e) < \infty.$$

There still is a tiny gap between F^* and the class of all densities with finite entropy: $\int f \log(f+1) < \infty$.

Lemma E5. *Let F^* be the class of densities satisfying the following property: There exists a positive convex function ψ, with $\psi(1) \ge 1$, $\psi(u)/u \uparrow$ for $u \ge 1$, $\psi^{inv}(u)/u \downarrow$, such that $\int_1^\infty 1/\psi(u)\, du < \infty$, and $\int \psi(f) < \infty$. Then, if f^* is the maximal function for f, it follows that $\int f \log(f^* + 1) < \infty$ whenever $f \in F^*$.*

Proof. Let ψ be a function satisfying the conditions mentioned in the definition of F^*. We will need the fact that if A_t is the set of all x with $f^*(x) > e^t - 1$, then

$$\int_{A_t} dx \le \frac{5}{e^t - 1}, \qquad t > 0$$

[see, e.g., Stein (1970), page 5]. Writing ψ^{inv} for the inverse of ψ, and K for

$\int \psi(f)$, we have the following chain of inequalities:

$$\begin{aligned}
&\int f \log(f^* + 1) \\
&\quad = \int_0^\infty \int_{A_t} f(x)\,dx\,dt \\
&\quad \le 1 + \int_1^\infty \int f(x) I_{A_t}\,dx\,dt \\
&\quad \le 1 + \int_1^\infty \int_{A_t} dx\, \psi^{\mathrm{inv}}\left(\int \psi(f) \Big/ \int_{A_t} dx \right) dt \qquad \text{(Jensen's inequality)} \\
&\quad = 1 + \frac{\int_1^\infty \psi^{\mathrm{inv}}(K/\int_{A_t} dx)\,dt}{K/\int_{A_t} dx} \\
&\quad \le 1 + \int_1^\infty \frac{\psi^{\mathrm{inv}}(K(e^t - 1)/5)}{K(e^t - 1)/5}\,dt \\
&\quad = 1 + \int_{K/5(e-1)}^\infty \frac{\psi^{\mathrm{inv}}(v)}{v(K + 5v)}\,dv,
\end{aligned}$$

which is finite if $\int_1^\infty \psi^{\mathrm{inv}}(v)/v^2\,dv < \infty$. Use the transformation $v := \psi(u)$, partial integration, the monotonicity of $\psi(u)/u$ and the finiteness of $\int_1^\infty 1/\psi(u)\,du$ to verify the finiteness of the integral in question. □

THEOREM E6. *If $f \in F^*$ (see Lemma E5 for definition), we have $\mathbf{E}D_{n+1} = \log_2 n + O(1)$. In particular,*

$$\mathbf{E}D_{n+1} \le \log_2 n + 5 + \int f \log_2(f^* + 1).$$

PROOF. From the proof of Theorem E4, we recall the following:

$$\begin{aligned}
\mathbf{E}D_{n+1} &= \int f(x) \sum_{k=0}^\infty \left(1 - \left(1 - \int_{I_{i,k}} f\right)^n\right) dx \\
&\le \int f(x) \inf_{M \text{ integer}} \left(1 + M + n \sum_{k=M}^\infty 2^{1-k} f^*(x)\right) dx \\
&= \int f(x) \inf_{M \text{ integer}} (1 + M + 4n 2^{-M} f^*(x))\,dx \\
&\le \int f(x)(3 + \log_2(4n(f^*(x) + 1)))\,dx \\
&\qquad\qquad [\text{take } M = \lfloor \log_2(4n(f^*(x) + 1)) \rfloor] \\
&\le 5 + \log_2 n + \int f \log_2(f^* + 1).
\end{aligned}$$

The proof is complete in view of Lemma E5. □

Let us finally look at densities that cause us problems because $\mathbf{E}D_2 = \infty$. Note that a sufficient condition for this is that

$$\sum_{k=1}^{\infty} \left(\sup_i \int_{I_{i,k}} f \right)^2 = \infty.$$

This is often a necessary condition as well. To better illustrate the matter, we consider monotone densities f.

THEOREM E7. *For decreasing densities f on* $[0,1]$, *we have* $\mathbf{E}D_2 = \infty$ *if and only if*

$$\sum_{k=1}^{\infty} \left(\int_0^{2^{-k}} f \right)^2 = \infty.$$

PROOF. From Theorem E6, we note that

$$\mathbf{E}D_2 = \sum_{k=1}^{\infty} \sum_{i=1}^{2^k} \left(\int_{I_{i,k}} f \right)^2 \geq \sum_{k=1}^{\infty} \left(\int_0^{2^{-k}} f \right)^2.$$

Also,

$$\begin{aligned}
\mathbf{E}D_2 &\leq \sum_{k=1}^{\infty} \left(\int_0^{2^{-k}} f \right)^2 + \sum_{k=1}^{\infty} \sum_{i=2}^{2^k} \int_{I_{i,k}} f^2/2^k \\
&= \sum_{k=1}^{\infty} \left(\int_0^{2^{-k}} f \right)^2 + \sum_{k=1}^{\infty} \int_{2^{-k}}^{1} f^2/2^k \\
&= \sum_{k=1}^{\infty} \left(\int_0^{2^{-k}} f \right)^2 + \int_0^1 f^2(x) \sum_{k=1}^{\infty} I_{[k > \log_2(1/x)]}/2^k \, dx \\
&\leq \sum_{k=1}^{\infty} \left(\int_0^{2^{-k}} f \right)^2 + \int_0^1 2xf^2(x)\, dx \\
&\leq \sum_{k=1}^{\infty} \left(\int_0^{2^{-k}} f \right)^2 + 2.
\end{aligned}$$

The inequalities here are obtained by the observation that for monotone densities, $xf(x) \leq 1$, and by an association inequality: If g, h are increasing positive functions, then for any random variable X, $\mathbf{E}h(X)\mathbf{E}g(X) \leq \mathbf{E}h(X)g(X)$ [see, e.g., Joag-Dev and Proschan (1983)]. □

THEOREM E8. *For monotone f, we have* $\mathbf{E}D_{n+1} = \log_2 n + O(1)$ *if and only if* $\int f \log(f+1) < \infty$.

PROOF. If f is a nonincreasing density on [0, 1], then $\int f(x)\log_2(1/x)\,dx < \infty$ if and only if $\int f \log(f+1) < \infty$. In fact,

$$\int_0^1 f \log(f+1) \le \int_0^1 f(x)\log\left(1+\frac{1}{x}\right)dx$$
$$\le \frac{4}{e} + \int_0^1 f \log(f+1).$$

This can best be seen as follows: In view of $xf(x) \le 1$, the leftmost implication is immediate. The rightmost inequality is a Young-type bound found, for example, in Hardy, Littlewood and Pólya (1952), Theorem 239.

Let $A_n \overset{\text{def}}{=} \sum_{k=0}^{\infty} \int_{I_{1,k}} f(1-(1-\int_{I_{1,k}} f)^n)$. Note that

$$A_n \le \sum_{k=0}^{\infty} \int_{I_{1,k}} f = \int_0^1 f(x) \sum_{k=0}^{\infty} I_{[x<2^{-k}]}\,dx \le 1 + \int_0^1 f(x)\log_2(1/x)\,dx.$$

We now show that for any decreasing density,

$$A_n \le \mathbf{E}D_{n+1} \le A_n + 2 + \log_2 n + \int f \log_2 f.$$

The lower bound is obvious since

$$\mathbf{E}D_{n+1} = A_n + \sum_{k=1}^{\infty}\sum_{i=2}^{2^k} \int_{I_{i,k}} f\left(1-\left(1-\int_{I_{i,k}} f\right)^n\right).$$

Using the fact that $\int_{I_{i,k}} f \le 2^{-k} f(x-2^{-k})$ for any $x \in I_{i,k}$, we have

$$\begin{aligned}
\mathbf{E}D_{n+1} &\le A_n + \sum_{k=1}^{\infty} \int_{2^{-k}}^1 f(x)\left(1-\left(1-2^{-k}f(x-2^{-k})\right)^n\right)dx \\
&\le A_n + \sum_{k=1}^{\infty} \int_0^1 f(x)\left(1-\left(1-2^{-k}f(x)\right)^n\right)dx \\
&\le A_n + \int_0^1 f(x)\left(\sum_{k\ge1:\,k\le\log_2(nf(x))} 1 + \sum_{k\ge1:\,k>\log_2(nf(x))} 2^{-k}nf(x)\right)dx \\
&\le A_n + \int_0^1 f(x)(\log_2(nf(x))+2)\,dx \\
&= A_n + \int f\log_2 f + \log_2 n + 2.
\end{aligned}$$

Theorem E8 now follows from Lemma E1 and the estimates obtained above. □

EXAMPLE. In this example we consider the family of densities F_a, where $f \in F_a$ if as $x \downarrow 0$, $f(x) \sim 1/(x \log^{1+a}(1/x))$, where $a > 0$ is a parameter. Note that for $a \le 0$, f cannot possibly be a density. Also, $\int_0^{2^{-k}} f \sim 1/(a(k\log 2)^a)$. Thus $\mathbf{E}D_2 = \infty$ if and only if $0 < a \le 1/2$. Consider next $1/2 < a \le 1$. These densities still have $H = -\infty$, but now $\mathbf{E}D_2 < \infty$. In view of

$$\frac{1}{2} \sum_{k \geq 1:\, \int_0^{2^{-k}} f \geq (1/n)\log 2} \int_0^{2^{-k}} f \leq A_n - 1$$

$$\leq \sum_{k \geq 1:\, \int_0^{2^{-k}} f \geq (1/n)\log 2} \int_0^{2^{-k}} f$$

$$+ \sum_{k \geq 1:\, \int_0^{2^{-k}} f \leq (1/n)\log 2} n \left(\int_0^{2^{-k}} f \right)^2,$$

we have $c_1 n^{(1-a)/a} \leq A_n \leq c_2 n^{(1-a)/a}$ for some constants c_1, c_2. For $1/2 < a < 1$, $\mathbf{E}D_{n+1}$ grows polynomially with n at any sublinear rate one desires to attain. For $a = 1$, we still have $\int f \log f = \infty$, but $A_n = O(\log n)$ so that $\mathbf{E}D_{n+1} = O(\log n)$. For $a > 1$, the entropy is finite, $A_n = O(1)$ and $\mathbf{E}D_{n+1} = \log_2 n - O(1)$. □

We conclude this section by arguing that in a certain sense, monotone densities constitute the worst case, and so we have the following theorem.

THEOREM E9. *For any density with* $\int f \log(f+1) < \infty$, *we have* $\mathbf{E}D_{n+1} \leq A \log_2 n + B \int f \log(f+1) + C$, *where* $A = 1 + e^{-1}$, $B = 2 + 2e^{-1}$ *and* $C = 4 + 9e^{-1} + 8e^{-2}$.

PROOF. The function $g(u) = u(1-(1-u)^n)$ is important in the study of $\mathbf{E}D_{n+1}$. Unfortunately, while it is monotone, it is not convex. It is convex on $[0, 2/(n+1)]$ and concave on $[2/(n+1), \infty)$. Clearly, it is bounded from above by the following convex function:

$$h(u) = \begin{cases} nu^2, & 0 \leq u \leq \dfrac{2}{n+1}, \\ \left(u - \dfrac{2}{n+1}\right)\left(1 + \dfrac{1}{e}\right) + n\left(\dfrac{2}{n+1}\right)^2, & u \geq \dfrac{2}{n+1}. \end{cases}$$

This is seen by noting that

$$g'\left(\frac{2}{n+1}\right) = 1 + \left(1 - \frac{2}{n+1}\right)^{n-1} \leq 1 + \frac{1}{e}$$

for all n. Define $p_i = \int_{I_{i,k}} f$ and $q_i = \int_{I_{i,k}} \bar{f}$, where $\bar{f}$ is the rearranged monotone version of f; that is, it is a nonincreasing density with the property that $\int_{\bar{f} > u} \bar{f} \equiv \int_{f > u} f$ for all u. Then it is clear that the vector of p_i's is stochastically majorized by the vector of q_i's; that is, if both vectors are sorted from small to large, then $q_1 + q_2 + \cdots + q_i \leq p_1 + p_2 + \cdots + p_i$ for all i. Thus, by some results on Schur convexity [see, e.g., Marshall and Olkin (1979)], $\sum_i h(p_i) \leq \sum_i h(q_i)$. The function $h(u)$ in turn does not exceed the more practical function $\bar{h}(u) = \min(nu^2, (1 + 1/e)u)$. Combining all this shows that

$$
\begin{aligned}
&\mathbf{E}D_{n+1} - 1 \\
&\quad \le \sum_{k=1}^{\infty} \sum_{i=1}^{2^k} \bar{h}\left(\int_{I_{i,k}} \bar{f}\right) \\
&\quad = \sum_{k=1}^{\infty} \sum_{i=1}^{2^k} \min\left(n\left(\int_{I_{i,k}} \bar{f}\right)^2, (1 + 1/e)\int_{I_{i,k}} \bar{f}\right) \\
&\quad \le 2\sum_{k=1}^{\infty} (1 + 1/e)\int_{I_{1,k}} \bar{f} + \sum_{k=2}^{\infty} \sum_{i=3}^{2^k} \min\left(n\left(\int_{I_{i,k}} \bar{f}\right)^2, (1 + 1/e)\int_{I_{i,k}} \bar{f}\right) \\
&\quad \le 2(1 + 1/e)\int_0^1 \bar{f}(x)\log_2(1/x)\,dx \\
&\qquad + \sum_{k=2}^{\infty} \int_{2^{-k}}^1 \bar{f}(x)\min\left(n2^{-k}\bar{f}(x), 1 + 1/e\right) dx \\
&\quad \le 2(1 + 1/e)\left(\frac{4}{e} + \int \bar{f}\log(\bar{f} + 1)\right) \\
&\qquad + \int_0^1 \bar{f}(x)\left(\sum_{k \ge 2:\, k \le \log_2(n\bar{f}(x))} (1 + 1/e) I_{[x \ge 2^{-k}]}\right. \\
&\qquad\qquad \left. + \sum_{k \ge 2:\, k > \log_2(n\bar{f}(x))} 2^{-k} n\bar{f}(x)\right) dx \\
&\quad \le 2(1 + 1/e)\left(\frac{4}{e} + \int \bar{f}\log(\bar{f} + 1)\right) \\
&\qquad + \int_0^1 \bar{f}(x)\Big(\big(1 + \log_2(nx\bar{f}(x))\big)(1 + 1/e) + 2\Big)\,dx \\
&\quad \le 2(1 + 1/e)\left(\frac{4}{e} + \int f\log(f + 1)\right) \\
&\qquad + 3 + 1/e + (1 + 1/e)\log_2 n.
\end{aligned}
$$

Here we followed arguments from the proof of Theorem E8. □

Acknowledgments. The author would like to thank both referees and an Associate Editor for their comments.

REFERENCES

AHO, A. V., HOPCROFT, J. E. and ULLMAN, J. D. (1983). *Data Structures and Algorithms*. Addison-Wesley, Reading, Mass.

BENNETT, G. (1962). Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.* **57** 33–45.

CHUNG, K. L. and ERDŐS, P. (1952). On the application of the Borel–Cantelli lemma. *Trans. Amer. Math. Soc.* **72** 179–186.

COFFMAN, E. G. and EVE, J. (1970). File structures using hashing functions. *Comm. ACM* **13** 427–436.

CSISZÁR, I. (1969). On generalized entropy. *Studia. Math. Hungar.* **4** 401–419.

DE GUZMAN, M. (1981). *Real Variable Methods in Fourier Analysis*. North-Holland, Amsterdam.

DE GUZMAN, M. (1975). *Differentiation of Integrals in R^n. Lecture Notes in Math.* **481**. Springer, Berlin.

DEVROYE, L. (1982). A note on the average depth of tries. *Computing* **28** 367–371.

DEVROYE, L. (1984). A probabilistic analysis of the height of tries and of the complexity of triesort. *Acta Inform.* **21** 229–237.

FAGIN, R., NIEVERGELT, J., PIPPENGER, N. and STRONG, H. R. (1979). Extendible hashing—a fast access method for dynamic files. *ACM Trans. Database Systems* **4** 315–344.

FLAJOLET, P. (1983). On the performance evaluation of extendible hashing and trie search. *Acta Inform.* **20** 345–369.

FLAJOLET, P. and PUECH, C. (1983). Tree structures for partial match retrieval. In *Proc. 24th IEEE Symp. Foundations of Computer Science* 282–288. IEEE, New York.

FLAJOLET, P. and PUECH, C. (1986). Tree structure for partial match retrieval. *J. ACM* **33** 371–407.

FLAJOLET, P., RÉGNIER, M. and SOTTEAU, D. (1985). Algebraic methods for trie statistics. *Ann. Discrete Math.* **25** 145–188.

FLAJOLET, P. and SEDGEWICK, R. (1986). Digital search trees revisited. *SIAM J. Comput.* **15** 748–767.

FLAJOLET, P. and STEYAERT, J. M. (1982). A branching process arising in dynamic hashing, trie searching and polynomial factorization. *Lecture Notes in Comput. Sci.* **140** 239–251. Springer, New York.

FREDKIN, E. H. (1960). Trie memory. *Comm. ACM* **3** 490–500.

HARDY, G. H., LITTLEWOOD, J. E. and PÓLYA, G. (1952). *Inequalities*. Cambridge Univ. Press.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

JACQUET P. and RÉGNIER, M. (1986). Trie partitioning process: Limiting distributions. *Lecture Notes in Comput. Sci.* **214** 196–210. Springer, Berlin.

JOAG-DEV, K. and PROSCHAN, F. (1983). Negative association of random variables, with applications. *Ann. Statist.* **11** 286–295.

KIRSCHENHOFER, P. and PRODINGER, H. (1986). Some further results on digital trees. *Lecture Notes in Comput. Sci.* **226** 177–185. Springer, Berlin.

KIRSCHENHOFER, P., PRODINGER, H. and SZPANKOWSKI, W. (1989a). On the balance property of PATRICIA tries: External path length viewpoint. *Theoret. Comput. Sci.* **68** 1–17.

KIRSCHENHOFER, P., PRODINGER, H. and SZPANKOWSKI, W. (1989b). On the variance of the external path length in a symmetric digital trie. *Discrete Appl. Math.* **25** 129–143.

KNUTH, D. E. (1973). *The Art of Computer Programming. Sorting and Searching* **3**. Addison-Wesley, Reading, Mass.

KONHEIM, A. G. and NEWMAN, D. J. (1973). A note on growing binary trees. *Discrete Math.* **4** 57–63.

LARSON, P. A. (1978). Dynamic hashing. *BIT* **18** 184–201.

LITWIN, W. (1981). Trie hashing. In *Proc. ACM-SIGMOD Conf. MOD*. Ann Arbor, Mich.

LUCCIO, F., RÉGNIER, and SCHOTT, R. (1989). Discs and other related data structures. *Algorithms and Data Structures. Lecture Notes in Comput. Sci.* **382** 192–205. Springer, Berlin.

MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications*. Academic, New York.

MENDELSON, H. (1982). Analysis of extendible hashing. *IEEE Trans. Software Engrg.* **8** 611–619.

PITTEL, B. (1985). Asymptotical growth of a class of random trees. *Ann. Probab.* **13** 414–427.

PITTEL, B. (1986). Path in a random digital tree: Limiting distributions. *Adv. in Appl. Probab.* **18** 139–155.

RÉGNIER, M. (1981). On the average height of trees in digital searching and dynamic hashing. *Inform. Process. Lett.* **13** 64–66.

RÉGNIER, M. (1988). Trie hashing analysis. In *Proc. Fourth Internat. Conf. Data Engineering* 377–381. IEEE, Los Angeles.

RÉNYI, A. (1959). On the dimension and entropy of probability distributions. *Acta Math. Acad. Sci. Hungar.* **10** 193–215.

STEIN, E. M. (1970). *Singular Integrals and Differentiability Properties of Functions*. Princeton Univ. Press.

SZPANKOWSKI, W. (1988a). On the height of digital trees and related problems. Technical Report CSD-TR-816, Dept. Computer Science, Purdue Univ.

SZPANKOWSKI, W. (1988b). Some results on *V*-ary asymmetric tries. *J. Algorithms* **9** 224–244.

SZPANKOWSKI, W. (1989). Digital data structures and order statistics. *Algorithms and Data Structures. Lecture Notes in Comput. Sci.* **382** 206–217. Springer, Berlin.

VITTER, J. S. and FLAJOLET, P. (1990). Average-case analysis of algorithms and data structures. In *Algorithms and Complexity* (J. Van Leeuwen, ed.) 432–524. MIT Press.

WHEEDEN, R. L. and ZYGMUND, A. (1977). *Measure and Integral*. Dekker, New York.

SCHOOL OF COMPUTER SCIENCE
MCGILL UNIVERSITY
MONTREAL, QUEBEC
CANADA H3A 2A7