Algorithmica (2005) 42: 11–29
DOI: 10.1007/s00453-004-1137-7

Algorithmica

Universal Asymptotics for Random Tries and PATRICIA Trees[1]

Luc Devroye[2]

Abstract. We consider random tries and random PATRICIA trees constructed from n independent strings of symbols drawn from any distribution on any discrete space. We show that many parameters Z_n of these random structures are universally stable in the sense that $Z_n/\mathbf{E}\{Z_n\}$ tends to one probability. This occurs, for example, when Z_n is the height, the size, the depth of the last node added, the number of nodes at a given depth (also called the profile), the search time for a partial match, the stack size, or the number of nodes with k children. These properties are valid without any conditions on the string distributions.

Key Words. Trie, PATRICIA tree, Probabilistic analysis, Law of large numbers, Concentration inequality, Height of a tree.

1. 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 strings $X_1, \ldots, X_n$, each drawn from $\prod_{i=1}^{\infty} \Omega_i$, where Ω_i, the ith alphabet, is a countable set. By appropriate mapping, we can and do assume that for all i, $\Omega_i = \mathcal{Z}$. In practice, the alphabets are often $\{0, 1\}$, but that will not even be necessary for the results in this paper. Each string $X_i = (X_{i1}, X_{i2}, \ldots)$ defines an infinite path in a tree: from the root, we take the X_{i1}st child, then its X_{i2}nd child, and so forth. The collection of nodes and edges visited by the union of the n paths is the infinite trie. If the X_i's are different, then each infinite path ends with a suffix path that is traversed by that string only. If this suffix path for X_i starts at node u, then we may trim it by cutting away everything below node u. This node becomes the leaf representing X_i. If this process is repeated for each X_i, we obtain a finite tree with n leaves, called the trie. 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. While it still has n leaves, each nonleaf (or internal) node has two or more children.

This article attempts to unify the concentration results for most trie parameters, and is a continuation of an earlier paper of the author (2002), where universal concentration results were obtained for the height, size, and profile of random PATRICIA trees. A real-valued trie parameter Z_n is said to be *concentrated* if $Z_n/\mathbf{E}\{Z_n\} \to 1$ in probability. It is *universally concentrated* if this holds for all string distributions, as long as the

[1] This research was sponsored by NSERC Grant A3456 and FCAR Grant 90-ER-0291.
[2] School of Computer Science, McGill University, 3480 University Street, Montreal, Quebec, Canada H3A 2K6. luc@cs.mcgill.ca.

Received November 18, 2002; revised March 28, 2004. Communicated by P. Bose and P. Morin.
Online publication February 16, 2005.

strings are independent. If it holds whenever the strings are independent and identically distributed, then we say that Z_n is *universally concentrated for i.i.d. input*. These results are obtained by powerful exponential inequalities developed most recently by Boucheron et al. (2000, 2003), which are related to but more easily applicable than their ancestors, the inequalities of Hoeffding (1963), Talagrand (1988, 1989, 1990, 1991a,b, 1993a,b, 1994, 1995, 1996a,b), Ledoux (1996a,b), Azuma (1967), and McDiarmid (1989, 1998). For the practicing computer scientist, they imply that analyzing $\mathbf{E}\{Z_n\}$ often suffices, as the random variable Z_n is highly likely to be close to its mean most of the time. In addition, the inequalities give explicit numerical support for that closeness, not hidden behind $O(\cdot)$ terms, and not requiring often tedious calculations of variances (denoted here by $\mathbf{V}\{Z_n\}$).

2. String Models. We refer to a number of string models in this paper. The oldest model is that of the i.i.d. symbols: each symbol is drawn from a symbol distribution $(p_0, p_1, p_2, \ldots)$ over the nonnegative integers. We call this the *i.i.d. symbol model*.

Text files have given rise to the *Markovian string model*, in which the symbols in a string form a Markov chain on the nonnegative integers. This is completely characterized by the distribution of the first symbol, and the transition probability matrix over $\mathcal{Z} \times \mathcal{Z}$ (Régnier, 1988; Szpankowski, 1988; Jacquet and Szpankowski, 1991; and Pittel, 1985).

Flajolet and Vallée (1998) and their co-workers have considered various models for storing real numbers. Their base model has as symbols the coefficients in the continued fraction expansion of a random variable X, drawn from a density on [0, 1]. We call this the continued fraction trie and PATRICIA tree. Alternatively, Devroye (1984, 1992a,b) takes the k-ary expansion of X, and lets the symbols be the digits in the expansion. This is the so-called *density model*. Just as with the *continued fraction model*, the symbols are in general dependent. However, in the density model, they are asymptotically uniformly distributed on Z_k. In the continued fraction model, the distribution on the positive integers is asymptotically fixed as well. Thus, both models should in many cases yield results that are valid for most or all densities of X.

Clément et al. (1998, 2001) and Flajolet and Vallée (1998) consider dynamic sources. Here, the real number X starts off functional iterations, leading to $X, f(X), f(f(X)), \ldots$, where $f : [0, 1] \to [0, 1]$ is a mapping. There is a second mapping $s : [0, 1] \to Z_+$, the nonnegative integers, and the string symbols are

$$s(X), s(f(X)), s(f(f(X))), \ldots.$$

For example, by taking $f(x) = kx \bmod 1$, and defining $s(\cdot)$ by partitioning the unit interval equally among $0, 1, \ldots, k-1$, we obtain the density model. The continued fraction and Markovian models can be obtained as special cases as well. The first-order parameters of the resulting tries were studied by Flajolet and Vallée (1998) and Clément et al. (2001). For random PATRICIA trees, they can be found in the thesis of Bourdon (2002).

Tries are also used as multidimensional data structures. They were first introduced by Orenstein (1982) for database applications. Related ideas had earlier been proposed by Bentley and Burkhard (1976). If X is a random vector of $[0, 1]^d$, then any of the ways of transforming a real number into a string of integers described above can be used. The simplest model uses the k-ary expansions of each component of X to generate a string of symbols, each taking values in Z_k^d. The resulting tries have a possible fanout of k^d. With $k = 2$, they are called *quadtries*, useful data structures for the compaction of multidimensional (geometric, video) information. In the present paper, parameters such as height, depth, profile, and size are dealt with uniformly for all models of tries. The multidimensional tries become interesting in their own right when one considers multivariate operations such as partial match. Puech and Yahia (1985) provide the first analytical study.

3. Boucheron–Lugosi–Massart Inequality.

The following inequalities are fundamental for the remainder xof the paper. Lemma 1 is an almost trivial extension of a similar inequality due to Boucheron et al. (2000). Its proof is based on logarithmic Sobolev inequalities developed in part by Ledoux (1996a).

LEMMA 1 (Boucheron et al., 2000). *Let $\Omega = \mathcal{Z}^n$. Let $f \geq 0$ be a function on Ω, let $c \geq 0$ be a constant, and let g be a real-valued function on $\mathcal{Z}^{n-1}$ satisfying the following properties for every $x = (x_1, \ldots, x_n) \in \Omega$:*

$$0 \leq f(x) - g(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \leq 1, \qquad 1 \leq i \leq n;$$

$$\sum_{i=1}^{n} (f(x) - g(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)) \leq f(x) + c.$$

Then for any $X = (X_1, \ldots, X_n)$ with independent components $X_i \in \mathcal{Z}$, and all $t \geq 0$,

$$\mathbf{P}\{f(X) \geq \mathbf{E}\{f(X)\} + t\} \leq \exp\left(-\frac{t^2}{2\mathbf{E}\{f(X) + c\} + 2t/3}\right)$$

and

$$\mathbf{P}\{f(X) \leq \mathbf{E}\{f(X)\} - t\} \leq \exp\left(-\frac{t^2}{2\mathbf{E}\{f(X) + c\}}\right).$$

The most outstanding application area for these inequalities are Talagrand's configuration functions. However, as we need to define g on a space of dimension one less than n, it is best to reformulate things in terms of "properties". Assume that we have a property P defined over the union of all finite products $\mathcal{Z}^k$. Thus, if $i_1 < \cdots < i_k$, we have an indicator function that decides whether $(x_{i_1}, \ldots, x_{i_k}) \in \mathcal{Z}^k$ satisfies property P. We assume that P is hereditary in the sense that if $(x_{i_1}, \ldots, x_{i_k})$ satisfies P, then so does any subsequence $(x_{j_1}, \ldots, x_{j_\ell})$ where $\{j_1, \ldots, j_\ell\} \subseteq \{i_1, \ldots, i_k\}$, with the j_m's increasing. The configuration function $f_n(x_{i_1}, \ldots, x_{i_n})$ gives the size of the largest subsequence of $x_{i_1}, \ldots, x_{i_n}$ satisfying P. Any subsequence of maximal length satisfying property P is called a witness. In Lemma 1 we can set $f(x_1, \ldots, x_n) = f_n(x_1, \ldots, x_n)$ and $g(x_1, \ldots, x_{n-1}) = f_{n-1}(x_1, \ldots, x_{n-1})$. Clearly, the first condition of Lemma 1 is satisfied, as adding a point to a sequence can only increase the value of the configuration function (so, $f \geq g$), but by not more than one. To verify the second condition, let $\{x_{i_1}, \ldots, x_{i_k}\} \subseteq \{x_1, \ldots, x_n\}$ be a witness of the fact that $f(x_1, \ldots, x_n) = k$. For

$i \le n$ and $x_i \notin \{x_{i_1}, \ldots, x_{i_k}\}$, we have $f(x_1, \ldots, x_n) = g(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$, and thus, the difference between f and g in the second condition can only be one if $x_i \in \{x_{i_1}, \ldots, x_{i_k}\}$. Therefore, the sum in that condition is at most $k = f(x_1, \ldots, x_n)$. Properties P include being monotonically increasing, being in convex position, and belonging to a given set S.

We also need some exponential versions of the Efron–Stein inequality. To this end, let $X_1, \ldots, X_n$ be independent random variables in a measurable space, and let $X_1', \ldots, X_n'$ be an independent copy. Let f be a measurable mapping, and set

$$Z = f(X_1, \ldots, X_n)$$

and

$$Z_i = f(X_1, \ldots, X_{i-1}, X_i', X_{i+1}, \ldots, X_n).$$

We write $X = (X_1, \ldots, X_n)$. Conditional on X, Z_i is thus a function of X_i' only. Define

$$V_+ = \mathbf{E}\left\{ \sum_{i=1}^n (Z - Z_i)^2 \mathbf{1}_{[Z > Z_i]} \Big| X \right\}$$

and

$$V_- = \mathbf{E}\left\{ \sum_{i=1}^n (Z - Z_i)^2 \mathbf{1}_{[Z < Z_i]} \Big| X \right\}.$$

The Efron–Stein inequality (Efron and Stein, 1981) states that

$$\mathbf{V}\{Z\} \le \mathbf{E}\{V_+\} = \mathbf{E}\{V_-\}.$$

By Chebyshev's inequality, we see that if $Z \ge 0$ is a positive random variable, and $\varepsilon > 0$, then

$$\mathbf{P}\{|Z - \mathbf{E}\{Z\}| \ge \varepsilon \mathbf{E}\{Z\}\} \le \frac{\mathbf{V}\{Z\}}{\varepsilon^2 \mathbf{E}^2\{Z\}} \le \frac{\mathbf{E}\{V_+\}}{\varepsilon^2 \mathbf{E}^2\{Z\}} .$$

In many cases (examples will follow), this ratio tends to zero. However, it does so rather slowly, and the resulting bounds are sometimes unsatisfactory. In this sense, the following inequality is very useful:

LEMMA 2 (Boucheron et al., 2003). *Assume $Z \ge 0$. If $V_+ \le aZ + b$ for some nonnegative constants a, b, then for all $t > 0$,*

$$\mathbf{P}\{Z \ge \mathbf{E}\{Z\} + t\} \le \exp\left(-\frac{t^2}{4a\mathbf{E}\{Z\} + 4b + 2at} \right).$$

If there exists a nondecreasing function h such that

$$V_- \le h(Z),$$

then

$$\mathbf{P}\{Z \le \mathbf{E}\{Z\} - t\} \le \exp\left(-\frac{t^2}{4\mathbf{E}\{h(Z)\}} \right).$$

Finally, the following inequality is useful whenever only V_+ is easy to bound.

LEMMA 3 (Boucheron et al., 2003). *Assume $Z \geq 0$ and assume that for all i, $|Z - Z_i| \leq \alpha$. If $V_+ \leq \alpha^2 h(Z/\alpha)$ for some nondecreasing function h, then for all $0 < t < (e-1)\alpha \mathbf{E}\{h(Z/\alpha)\}$,*

$$\mathbf{P}\{Z \leq \mathbf{E}\{Z\} - t\} \leq \exp\left(-\frac{t^2}{4\alpha^2(e-1)\mathbf{E}\{h(Z/\alpha)\}}\right).$$

4. Occupancy Problems. Consider a very general bin model in which we have n balls thrown independently into a countable number of bins, where each ball has its own distribution over the bins. Let $N_1, N_2, \ldots$ be the numbers of balls in the bins. Quantities of interest in certain applications include $M_n = \max_i N_i$, the maximum number of balls in a single bin, and $O_n = \sum_i \mathbf{1}_{N_i > 0}$, the number of occupied bins. If we throw one less ball, then M_n and O_n both decrease by at most one. Thus, uniformly over distributions, by the bounded difference inequality (Azuma, 1967; McDiarmid, 1989), we have

$$\mathbf{P}\{|O_n - \mathbf{E}\{O_n\}| \geq t\} \leq 2e^{-t^2/2n}.$$

Also,

$$\mathbf{P}\{|M_n - \mathbf{E}\{M_n\}| \geq t\} \leq 2e^{-t^2/2n}.$$

These results are sometimes unsatisfactory, as t needs to be at least $\Omega(\sqrt{n})$ for the inequalities to kick in. Note however that both O_n and M_n may be cast in the format of Lemma 1, with M_n being the configuration function for the hereditary property "belonging to the same bin", and O_n being the configuration function for the hereditary property "belonging to different bins". Thus, by Lemma 1,

$$\mathbf{P}\{O_n \geq \mathbf{E}\{O_n\} + t\} \leq \exp\left(-\frac{t^2}{2\mathbf{E}\{O_n\} + 2t/3}\right), \qquad t \geq 0,$$

and

$$\mathbf{P}\{O_n \leq \mathbf{E}\{O_n\} - t\} \leq \exp\left(-\frac{t^2}{2\mathbf{E}\{O_n\}}\right), \qquad t \geq 0.$$

Also, for fixed $t > 0$, if $\mathbf{E}\{O_n\} \to \infty$,

$$\mathbf{P}\left\{\left|\frac{O_n - \mathbf{E}\{O_n\}}{\sqrt{\mathbf{E}\{O_n\}}}\right| \geq t\right\} \leq 2\exp\left(-\frac{t^2}{2 + o(1)}\right).$$

Precisely the same inequalities hold when O_n is replaced by M_n throughout. Note that these inequalities are strong enough to imply the following:

$$\frac{O_n}{\mathbf{E}\{O_n\}} \to 1$$

in probability whenever $\mathbf{E}\{O_n\} \to \infty$, and the result is true over a triangular array of bin distributions. Also, we have

$$\frac{M_n}{\mathbf{E}\{M_n\}} \to 1$$

in probability whenever $\mathbf{E}\{M_n\} \to \infty$.

In data structures these results are relevant for hashing with chaining with equal or unequal probabilities. The maximal chain length satisfies the law of large numbers regardless of how the table size changes with n. For M_n, if the number of bins equals the number of balls, then $M_n \sim \log n / \log\log n$ if each bin has equal probability of receiving a ball. The inequalities at the top of the section would not allow one to obtain a law of large numbers. However, Lemma 1, as shown above, suffices to obtain it. See Gonnet (1981), Devroye (1985), or Knuth (1973) for more on the maximum chain length.

It is interesting to note that problems on tries can often be regarded as problems on bins. In fact, we consider levels of nested bins, with the zeroth level consisting of one bin, corresponding to the root, the first level having one bin for each child of the root, and so forth. A string $X_i = (X_{i1}, X_{i2}, \ldots)$ thus drops a ball in the bin characterized by X_{i1} of the level 1 collection, a ball in the bin characterized by (X_{i1}, X_{i2}) of the level 2 collection, and so forth. We call bin $(X_{i1}, \ldots, X_{i,j+1})$ a child bin of $(X_{i1}, \ldots, X_{i,j})$. In a trie, a bin that receives at least two balls corresponds to an internal node. The number of nodes is n (the number of leaves) plus the number of internal nodes, which is a random variable. In a PATRICIA tree, a bin with at least two balls, whose parent bin has more balls, corresponds to an internal node. These observations permit us to make observations about the size and the profile of random tries and PATRICIA trees.

5. Size of a PATRICIA Tree. Let S_n be the number of internal nodes, and let $T_n = S_n + n$ be the total number of nodes in a PATRICIA tree for n strings. Note that for binary PATRICIA trees, $S_n = n - 1$, so only nonbinary trees have random sizes. Adding a string increases T_n by one and S_n by one or zero. Thus, if the strings are independent (but not necessarily identically distributed), by the bounded difference inequality (McDiarmid, 1989),

$$\mathbf{P}\{|S_n - \mathbf{E}\{S_n\}| \geq t\} = \mathbf{P}\{|T_n - \mathbf{E}\{T_n\}| \geq t\} \leq 2\exp\left(-\frac{t^2}{2n}\right).$$

The fanout and string distributions do not figure in the bound. We immediately have

$$\frac{T_n}{\mathbf{E}\{T_n\}} \to 1$$

almost surely (as $T_n \geq n$), and

$$\frac{S_n}{\mathbf{E}\{S_n\}} \to 1$$

in probability whenever $\mathbf{E}\{S_n\}/\sqrt{n} \to \infty$ (which is satisfied, for example, if the strings consist of i.i.d. symbols, or when the tree is of bounded fanout). Even though these results do not require Lemma 1, they appear to be new. Still, we would like to offer a stronger result.

Let $Z = S_n$ be the parameter of interest. Observe that removing a string X_i causes Z to shrink by at most one node, so that $|Z - Z_i| \leq 1$. Furthermore,

$$\sum_i (Z - Z_i)^2 \mathbf{1}_{[Z > Z_i]} \leq \sum_i \mathbf{1}_{[Z > Z_i]} \leq 2Z$$

as each internal node can be removed only when it has two balls and one of its two balls are removed. It cannot be removed when it has more than two balls. Thus, $V_+ \le 2Z$. We are in a position to apply Lemma 2 with $(a, b) = (2, 0)$, and obtain

$$\mathbf{P}\{S_n \ge \mathbf{E}\{S_n\} + t\} \le \exp\left(-\frac{t^2}{8\mathbf{E}\{S_n\} + 4t}\right).$$

Furthermore, we can apply Lemma 3 with $h(u) = 2u$, $\alpha = 1$ and obtain

$$\mathbf{P}\{S_n \le \mathbf{E}\{S_n\} - t\} \le \exp\left(-\frac{t^2}{8(e-1)\mathbf{E}\{S_n\}}\right), \qquad 0 < t.$$

By setting $t = \varepsilon\mathbf{E}\{S_n\}$, we conclude immediately that

$$\frac{S_n}{\mathbf{E}\{S_n\}} \to 1 \qquad \text{in probability}$$

whenever $\mathbf{E}\{S_n\} \to \infty$. Only pathological examples do not satisfy this condition. For example, this would occur if $X_i = (i, i, i, \ldots)$, so that the PATRICIA tree consists of one root and n child leaves, and thus $S_n = 1$.

When the strings are i.i.d., and the symbols in the strings are i.i.d. as well, the size S_n of a random trie has been studied by Régnier and Jacquet (1989) and Jacquet and Régnier (1989). The expected value $\mathbf{E}\{S_n\}$ is close to $n/\mathcal{H}$, where $\mathcal{H}$ is the entropy of a random symbol. (If a symbol has distribution $(p_j)_{j\ge 0}$, then the entropy $\mathcal{H}_k$ of order k is $\sum_j p_j \log^k(1/p_j)$. The entropy is the entropy of order one.) There are string distributions that cause $\mathbf{E}\{S_2\} = \infty$: just let each string be all zeros, with ones added in at places $W_1, W_1 + W_2, \ldots$, and the W_i's are i.i.d. with the same distribution on the positive integers. Then $S_2 \ge \min(W_1, W_1')$, the minimum of two independent copies of W_1. Thus, $\mathbf{P}\{\min(W_1, W_1') \ge t\} = (\mathbf{P}\{W_1 \ge t\})^2$ and so, whenever $\mathbf{P}\{W_1 \ge t\} \ge C/\sqrt{t}$, for some positive C, we have $\mathbf{E}\{S_2\} = \infty$.

Rais et al. (1993) studied S_n for PATRICIA trees when the symbols are independent. Bourdon (2001, 2002) studied S_n for PATRICIA trees under a general string distribution model. It is noted that in the former case, $\mathbf{E}\{S_n\}$ is about $(1-\mathcal{H}_2/(2\mathcal{H}^2))n/\mathcal{H}$. With more general definitions of entropy for dependent sequences of symbols (as in Markovian sources), the same result was obtained by Bourdon. For the continued fraction trie, Bourdon showed that $\mathbf{E}\{S_n\}/n$ is about

$$\frac{6\log 2}{\pi^2}$$

(this is the inverse of Lévy's constant), and for the random PATRICIA tree, he showed that it is about equal to the same constant multiplied by

$$\left(1 - \frac{12\gamma\log 2}{\pi^2} - \frac{9\log^2 2}{\pi^2} + \frac{72\zeta'(2)\log 2}{\pi^4} + \frac{1}{2}\right) \approx 0.87.$$

Here γ is Euler's constant, and ζ is Riemann's zeta function. The study of variances and second-order properties of S_n for PATRICIA trees was left as a major open problem in the thesis of Bourdon (2002). For symbol distributions that are "periodic", the expected value of S_n/n oscillates forever for random tries and PATRICIA trees, so the limiting constants mentioned above are not true "limits". However, as we showed in this paper, even in those cases, $S_n/\mathbf{E}\{S_n\} \to 1$ in probability. We did not have to worry about the oscillation phenomenon. We observe in fact, that for all models, without exception,

$$\mathbf{V}\{S_n\} \le \mathbf{E}\{V_+\} \le 2\mathbf{E}\{S_n\}.$$

6. Multidimensional Tries and the Partial Match Operation. We can define a *general query* in a trie or PATRICIA tree as follows: it is defined by a sequence of sets $S_1, S_2, S_3, \ldots$ of symbols, and asks for all the leaves (if any) of the trie or PATRICIA tree that are in the infinite query tree as well, where the query tree consists of a root, all nodes at level one indexed by $s_1 \in S_1$, all nodes at level 2 indexed by $(s_1, s_2) \in S_1 \times S_2$, and so forth. Thus, the query tree can be viewed as an infinite trie. The time needed to collect all the leaves that need reporting is proportional to the number N of nodes in the intersection of the query tree and the original trie or PATRICIA tree. In the case of a PATRICIA, care must be taken to identify nodes in the tree not by their position but by their index. Some examples follow that show that most parameters in a random trie can be viewed as special cases of N.

EXAMPLE 1. If $S_1 = S_2 = \cdots$ is the full symbol set, then we must report all leaves, and the time N in that case is the *size of the random trie*.

EXAMPLE 2. If $S_i = \{y_i\}$ for all i, then the query tree is the infinite path $(y_1, y_2, y_3, \ldots)$, and thus N is the size of the path followed when we search for the string $(y_1, y_2, y_3, \ldots)$ in the trie or PATRICIA tree. This is equal to the depth of the unique leaf on that path plus one. Thus, *search times for individual strings* are obtained as special cases.

EXAMPLE 3. In the multivariate random trie and random PATRICIA tree, with a fanout of k^d (see the Introduction for the notation), we may describe each symbol as an element of Z_k^d. A *partial match query* can be regarded as a general query in which $S_i = S_{i1} \times \cdots \times S_{id}$, $S_{ij} = Z_k$ for all $j \in J \subseteq \{1, 2, \ldots, d\}$, the so-called set of wild cards. For $j \notin J$, we have $S_{ij} = \{y_{ij}\}$, a singleton set. The partial match is thus determined by all the values $y_{ij}, i \ge 1, j \notin J$. Usually, the interpretation is easy in term of a vector $y \in [0, 1]^d$, whose jth component has a k-ary expansion $(y_{1j}, y_{2j}, \ldots)$. Thus, the pair (y, J) uniquely defines a partial match query. We may thus write $N(y, J)$ to denote the size of the intersection of our two trees for the query given by (y, J). If J is the full set, then the query tree is the full tree, and we are reduced to Example 1. If J is empty, then the query tree reduces to a path, as in Example 2. The only interesting cases appear when $0 < |J| < d$, and in that case we speak of a proper partial match. Few results are known, and they all relate to random tries with $k = 2$, when each of the strings in the random trie is generated by an independently drawn random vector uniformly distributed on $[0, 1]^d$. In that case all symbols in all strings are independent and uniformly distributed on Z_2^d. Consider a proper partial match. Flajolet and Puech (1986) showed that

$$\mathbf{E}\{N(y, J)\} = \tau(\log_2 n) n^{|J|/d} + o\left(n^{|J|/d}\right),$$

where τ is a continuous positive periodic function. Kirschenhofer et al. (1993) showed that for $k = 2$, $d = 2$, $|J| = 1$,

$$\mathbf{V}\{N(y, J)\} \sim \tau(\log_2 n)\sqrt{n},$$

where τ is again a continuous positive periodic function. This was generalized to $d > 2$ by Schachinger (1995). In 2000, Schachinger proved that

$$\frac{N(y, J)}{\mathbf{E}\{N(y, J)\}} \to 1$$

in probability as $n \to \infty$, and

$$\frac{N(y, J) - \mathbf{E}\{N(y, J)\}}{\sqrt{\mathbf{V}\{N(y, J)\}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1),$$

where $\xrightarrow{\mathcal{L}}$ denotes convergence in distribution. Using concentration inequalities, Devroye and Zamora-Cura (2002) were able to show that

$$\frac{\sup_y N(y, J)}{\inf_y N(y, J)} \to 1$$

in probability, which shows a remarkable stability of all partial match query times: every partial match query time is with high probability close to $\mathbf{E}\{N(y_0, J)\}$, where y_0 is the vector of all zeros. The asymmetric Bernoulli model refers to strings with i.i.d. symbols, and each symbol consists of d independent components, with each component drawn from some nonuniform distribution of Z_2. For this model, some asymptotics for $\mathbf{E}\{N(y, J)\}$ are obtained by Kirschenhofer et al. (1993), and, with y replaced by a random Y, by Schachinger (2000). In the latter paper it is shown that $N(y, J)/\mathbf{E}\{N(Y, J)\} \xrightarrow{\mathcal{L}} Z_p$ under an idealized partial match model that does not correspond to our definition, where the distribution of Z_p depends upon the probability p only. More recently, Schachinger has studied the asymptotic behavior of the ratio $\log N(y, J)/\mathbf{E}\{\log N(y, J)\}$.

EXAMPLE 4. If we take all S_i, $1 \le i \le k$, equal to the full set of symbols, and all S_i, $i > k$, empty, then we obtain the cumulative size of the tree up to level k. It is also called the *cumulative profile*. In fact, the profile would be obtained by considering all the nodes at one level. Strictly speaking, the profile does not follow the subtree analogy set up in this section, so we deal with it below.

So, we consider N for a general query, freed from the multidimensional confines. This N is our random variable to which we want to apply some concentration inequalities. When a string is removed, in the PATRICIA version, N decreases by at most two, as one leaf disappears and possibly one internal node. Thus, denoting by N_i the size of the intersection tree when string X_i is removed but an independent copy X_i' is added, we see that $|N - N_i| \le 2$. Furthermore,

$$\sum_i (N - N_i)^2 \mathbf{1}_{[N > N_i]} \le 4 \sum_i \mathbf{1}_{[N > N_i]} \le 8N$$

as each internal node can be removed only when it has two balls and one of its two balls is removed. It cannot be removed when it has more than two balls. Thus, $V_+ \le 8N$. We are in a position to apply Lemma 2 with $(a, b) = (8, 0)$, and obtain

$$\mathbf{P}\{N \ge \mathbf{E}\{N\} + t\} \le \exp\left(-\frac{t^2}{32\mathbf{E}\{N\} + 16t}\right).$$

Furthermore, we can apply Lemma 3 with $h(u) = 4u$, $\alpha = 2$ and obtain

$$\mathbf{P}\{N \le \mathbf{E}\{N\} - t\} \le \exp\left(-\frac{t^2}{32(e-1)\mathbf{E}\{N\}}\right), \qquad 0 < t.$$

By setting $t = \varepsilon\mathbf{E}\{N\}$, we conclude immediately that

$$\frac{N}{\mathbf{E}\{N\}} \to 1 \qquad \text{in probability}$$

whenever $\mathbf{E}\{N\} \to \infty$.

Finally, the Efron–Stein inequality implies that for all models, without exception, and for any kind of general query,

$$\mathbf{V}\{N\} \le \mathbf{E}\{V_+\} \le 8\mathbf{E}\{N\}.$$

It is interesting to note that the addition of one string to a trie can cause a trie to grow by many nodes, and that this growth cannot be a priori bounded. The concentration inequalities given here thus cannot be used directly. However, if X_i is removed and replaced by string X_i', we have

$$\sum_i (N - N_i)^2 \mathbf{1}_{[N > N_i]} \le (1 + H_n)^2 \sum_i \mathbf{1}_{[N > N_i]}, \le 2(1 + H_n)^2 N$$

where H_n is the height of the random trie, as $|N - N_i| \le 1 + H_n$. Thus, $V_+ \le 2(1+H_n)^2 N$. In most, but not all, models, H_n is of the order of $\log n$, so that one can expect to make good use of this if one has further information on the height H_n, and this requires, unfortunately, some information on the string distributions. We draw attention though to an inequality of Boucheron et al. (2003) that deals precisely with the case that $V_+ \le WZ$, where Z is the random variable of interest, and W is another random variable, whose tail can be bounded sufficiently tightly. This way, one can obtain good concentration inequalities, although not universal ones, for N in random tries.

The second way in which N can be dealt with in random tries is by decomposing $N = \sum_\ell N^{(\ell)}$, where $N^{(\ell)}$ is the number of nodes in the intersection with the query tree that are at level ℓ. As $N^{(\ell)}$ is easy to deal with, one can obtain a concentration inequality for it (see Section 9). It is also true that in many models, N is nearly equal to the sum of $N^{(\ell)}$ with ℓ ranging over just a few levels. In those cases, concentration for N follows. An example of this approach is worked out by Devroye and Zamora-Cura (2004) for partial match queries in multidimensional tries in which all symbols are i.i.d. and uniformly distributed on Z_2^d.

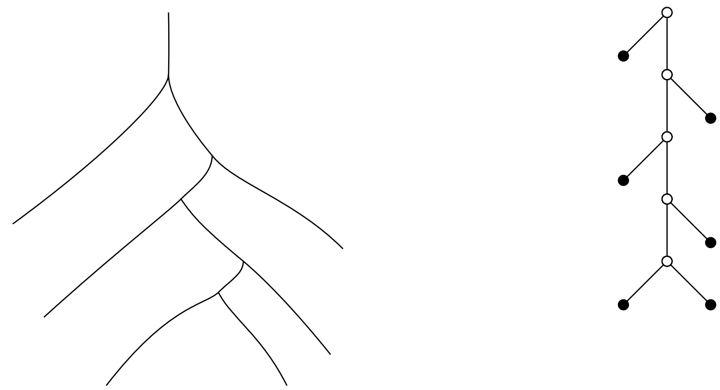

Fig. 1. Six strings with the PATRICIA property. Each (black) leaf represents a contracted infinite string. The height is five.

7. Height of a PATRICIA Tree. Given are n independent infinite strings $X_1, \ldots, X_n$ (if they are not infinite, pad them by some designated character, repeated infinitely often), each drawn from a distribution on $\mathcal{Z}$. The height of the PATRICIA tree is denoted by H_n. If (deterministic) strings $x_1, \ldots, x_k$ induce a PATRICIA tree of height $k-1$, then the PATRICIA tree can have only one configuration, namely, it consists of a chain of length $k-1$ from the root on down, with every node of this chain receiving one leaf, except the furthest node, which receives two leaves. We say that such a collection of strings has the PATRICIA property (Figure 1). This property is clearly hereditary, and $H_n + 1$ is thus a configuration function.

We have

$$\mathbf{P}\{H_n \geq \mathbf{E}\{H_n\} + t\} \leq \exp\left(-\frac{t^2}{2\mathbf{E}\{(H_n+1)\} + 2t/3}\right), \qquad t \geq 0,$$

and

$$\mathbf{P}\{H_n \leq \mathbf{E}\{H_n\} - t\} \leq \exp\left(-\frac{t^2}{2\mathbf{E}\{(H_n+1)\}}\right), \qquad t \geq 0.$$

We stress that the individual strings may have any distribution. The symbols themselves need not be independent or identically distributed, and the strings need not be identically distributed. All PATRICIA trees, without exception, are thus stable and well behaved. For any PATRICIA tree constructed by using n independent strings, if $\lim_{n\to\infty} \mathbf{E}\{H_n\} = \infty$, then

$$\frac{H_n}{\mathbf{E}\{H_n\}} \to 1$$

in probability as $n \to \infty$, and

$$\frac{H_n - \mathbf{E}\{H_n\}}{\sqrt{\mathbf{E}\{H_n\}}} = O(1)$$

in probability in this sense: for fixed $t > 0$,

$$\mathbf{P}\left\{\left|\frac{H_n - \mathbf{E}\{H_n\}}{\sqrt{\mathbf{E}\{H_n\}}}\right| \geq t\right\} \leq 2\exp\left(-\frac{t^2}{2+o(1)}\right).$$

The last inequality remains valid whenever $0 < t = o(\mathbf{E}\{H_n\})$.

The Condition on $\mathbf{E}\{H_n\}$. In PATRICIA trees of bounded degree, it is clear that $\mathbf{E}\{H_n\} \to \infty$. In unbounded degree trees, this is also true provided that the strings are identically distributed and the probability of two identical strings is zero. However, without the identical distribution constraint, PATRICIA trees may have $H_n = 1$ for all n: just let the ith string be $(i, 0, 0, 0, \ldots)$.

Bibliographic Remark: *Height of* PATRICIA *Trees*. All parameters of a PATRICIA tree such as H_n improve over those of the associated trie: for the uniform trie model, 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), Aldous and Shields (1988), Kirschenhofer and Prodinger (1986), Kirschenhofer el al. (1989a), and Szpankowski (1990, 1991). Pittel and Rubin (1990), Pittel (1991), and Devroye (1992a,b) showed that

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

More refined results for general multibranching PATRICIA trees and tries are given by Szpankowski and Knessl (2000). For the independent symbol trie model with symbol probabilities p_j, we have $\mathbf{E}\{H_n\} \sim c\log n$, where $c = 2/\log_2(1/\sum_j p_j^2)$.

8. Depth Along a Given Path in a PATRICIA Tree. Consider a string x that defines an infinite path in a trie. We define the depth of the path x, denoted by $D_n(x)$ in the PATRICIA tree, as the depth (distance to the root) of the leaf that corresponds to x in the PATRICIA tree for $X_1, \ldots, X_n, x$. We say that strings $x_1, \ldots, x_k$ have the x-property if the prefixes $x \cap x_1, \ldots, x \cap x_k$ are strictly nested. That is, there is a reordering $x_1', \ldots, x_k'$ of the strings such that the common prefix of x_1' and x is strictly contained in that of x_2' and x, and so forth. In that case the distance of the leaf of x from the root of the PATRICIA tree for $x_1, \ldots, x_k, x$ is precisely k. The function $D_n(x) = f(x_1, \ldots, x_n)$ that describes the length of the longest subset of $x_1, \ldots, x_n$ with the x-property is clearly a configuration function, to which Lemma 1 may be applied. Thus, we conclude as in the previous section:

For any PATRICIA tree constructed by using n independent strings, if x is a string such that $\lim_{n\to\infty} \mathbf{E}\{D_n(x)\} = \infty$, then

$$\frac{D_n(x)}{\mathbf{E}\{D_n(x)\}} \to 1$$

in probability as $n \to \infty$, and

$$\frac{D_n(x) - \mathbf{E}\{D_n(x)\}}{\sqrt{\mathbf{E}\{D_n(x)\}}} = O(1)$$

in probability in this sense: for fixed $t > 0$,

$$\mathbf{P}\left\{\left|\frac{D_n(x) - \mathbf{E}\{D_n(x)\}}{\sqrt{\mathbf{E}\{D_n(x)\}}}\right| \geq t\right\} \leq 2\exp\left(-\frac{t^2}{2+o(1)}\right).$$

9. General Profiles. This section is about the profile in random tries. The profile of a trie is the sequence $(P_1, P_2, P_3, \ldots)$, where P_ℓ is the number of nodes at level ℓ (at distance ℓ from the root). This is in fact the number of nodes that have either at least leaves in their subtrees or correspond to leaves. Let $P_{\ell i}$ denote the value when string X_i is deleted and string X_i' is added. It is easy to see that $|P_\ell - P_{\ell i}| \leq 1$. Furthermore,

$$\sum_i (P_\ell - P_{\ell i})^2 \mathbf{1}_{[P_\ell > P_{\ell i}]} \leq \sum_i \mathbf{1}_{[P_\ell > P_{\ell i}]} \leq 2P_\ell$$

because only internal nodes with two balls can be deleted when X_i is deleted from the list of strings. Summed over all i, this can happen just twice. So, we have a situation exactly like that for the size S_n of a PATRICIA tree. In particular, $V_+ \leq 2P_\ell$. We apply Lemma 2 with $(a, b) = (2, 0)$:

$$\mathbf{P}\{P_\ell \geq \mathbf{E}\{P_\ell\} + t\} \leq \exp\left(-\frac{t^2}{8\mathbf{E}\{P_\ell\} + 4t}\right).$$

Furthermore, we can apply Lemma 3 with $h(u) = 2u$, $\alpha = 1$ and obtain

$$\mathbf{P}\{P_\ell \leq \mathbf{E}\{P_\ell\} - t\} \leq \exp\left(-\frac{t^2}{8(e-1)\mathbf{E}\{P_\ell\}}\right), \qquad 0 < t < 2(e-1)\mathbf{E}\{P_\ell\}.$$

By setting $t = \varepsilon\mathbf{E}\{P_\ell\}$, we conclude immediately that

$$\frac{P_\ell}{\mathbf{E}\{P_\ell\}} \to 1 \qquad \text{in probability}$$

whenever $\mathbf{E}\{P_\ell\} \to \infty$, regardless of whether ℓ is fixed or varies with n. Finally,

$$\mathbf{V}\{P_\ell\} \leq 2\mathbf{E}\{P_\ell\}.$$

All of the above remains valid for the *general profile*, which for each level counts the number of trie nodes (internal or leaves) that belong to a certain set S on the space of all indices.

10. Height of a Random Trie: Lower Bound. Note the duality

$$[H_n \geq \ell] = [P_\ell > 0].$$

Thus,

$$\mathbf{P}\{H_n < \ell\} = \mathbf{P}\{P_\ell = 0\} = \mathbf{P}\{P_\ell - \mathbf{E}\{P_\ell\} \leq -\mathbf{E}\{P_\ell\}\} \leq \exp\left(-\frac{\mathbf{E}\{P_\ell\}}{8(e-1)}\right).$$

This inequality is valid without any conditions. For random tries, we have a simple universal lower bound on the tail of the height H_n in terms of the first moment of P_ℓ only. Since we also have

$$\mathbf{P}\{H_n \geq \ell\} = \mathbf{P}\{P_\ell \geq 1\} \leq \mathbf{E}\{P_\ell\},$$

we can conclude that if $\ell = \ell(n)$ is a function of n, then

$$\lim_{n\to\infty} \mathbf{P}\{H_n \geq \ell(n)\} = \begin{cases} 0 & \text{if } \lim_{n\to\infty} \mathbf{E}\{P_{\ell(n)}\} = 0, \\ 1 & \text{if } \lim_{n\to\infty} \mathbf{E}\{P_{\ell(n)}\} = \infty. \end{cases}$$

First-order behavior of the height of all random tries reduces thus simply to the study of the expected profile.

Bibliographic Remark: Height of Random Tries. The asymptotic behavior of tries under the uniform trie model is well known. For example, it is known that

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

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 (1992b). The height for other models was 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), Kirschenhofer et al. (1989b), and Szpankowski (1988).

11. Height of a Random Trie: Upper Bound. It is possible to deal with upper bounds for the tail of H_n in random tries in some way, despite the instability of this parameter for some string distributions. The route suggested here is based on Talagrand's q-points inequality and a subadditivity argument due to van der Vaart and Wellner (1996). Assume that random variables $X_i \in \Omega$ are given that are i.i.d. Let Ω^n be the product space.

LEMMA 4 (Van Der Vaart and Wellner, 1996). *Assume that Ω is Euclidean space, and that $f_n : \Omega^n \to \mathbf{R}$ is a permutation-symmetric function, satisfying the following monotonicity and subadditivity conditions:*

$$\begin{aligned} f_n(x) &\leq f_{n+m}(x, y), \qquad x \in \Omega^n, \quad y \in \Omega^m; \\ f_{n+m}(x, y) &\leq f_n(x) + f_m(y), \qquad x \in \Omega^n, \quad y \in \Omega^m. \end{aligned}$$

Assume furthermore that $0 \leq f_n \leq n$. Let $X = (X_1, \ldots, X_n)$ have i.i.d. coordinates in Ω^n. Then

$$\mathbf{P}\{f_n(X_1, \ldots, X_n) \geq t\} \leq \exp\left(-\frac{t}{2}\log\left(\frac{t}{12\max(1, \mathbf{E}\{f_n(X_1, \ldots, X_n)\})}\right)\right)$$

for all $t > 0, n \geq 1$.

Consider a random trie based upon n i.i.d. strings $X_1, \ldots, X_n$ with string symbols from an arbitrary alphabet, and let the string distribution be arbitrary. Then the height H_n satisfies all the required conditions, except possibly the requirement that $H_n \leq n$. However, $\min(H_n, n)$ is fully compliant. The Euclidean space requirement is fulfilled as we can study $\min(H_n, n)$ by just considering string collections truncated to their length n prefixes. Set $h_n = \mathbf{E}\{\min(H_n, n)\}$. Note that for $n \geq 2$, $h_n \geq 1$. Then we have for $n \geq 2$, $0 \leq t \leq n$,

$$\mathbf{P}\{H_n \geq t\} = \mathbf{P}\{\min(H_n, n) \geq t\} \leq \exp\left(-\frac{t}{2}\log\left(\frac{t}{12h_n}\right)\right).$$

Thus, if $h_n \to \infty$, yet $h_n = o(n)$, as is usually the case, we see that for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbf{P}\{H_n \geq (12+\varepsilon)h_n\} = 0$$

in all generality. It is difficult to imagine that one can state such universal results without access to the powerful machinery provided by Talagrand's inequalities.

12. External Path Length of a PATRICIA Tree. The external path length I_n of a PATRICIA tree or trie is the sum of the distances from the leaves to the root. In the case of a trie, it is an upper bound for the time needed to construct the trie. Upper bounds for the tail of I_n are thus useful. When the ith string is deleted in a PATRICIA tree, then I_n decreases by D_i, the distance from the ith leaf to the root, and, in case one internal is removed in the process, by the number of leaves, i excepted, in the subtree of the node that disappeared (as all those leaves decrease their depth by one). That number of such leaves is denoted by N_i. Note that $N_i = 0$ if no leaf changes depth. Now add a new independent string X_i'. Denote by I_{ni} the new external path length. We thus have

$$\sum_i (I_n - I_{ni})^2 \mathbf{1}_{[I_n > I_{ni}]} \leq \sum_i (D_i + N_i)^2 \leq n \sum_i (D_i + N_i) \leq 2nI_n,$$

where we used the fact that $\sum_i N_i$ is bounded from above by I_n, as each leaf contributes at most its depth towards the total sum. Thus, we are in a position to apply Lemma 2 with $(a, b) = (2n, 0)$. We see that for all $t > 0$,

$$\mathbf{P}\{I_n \geq \mathbf{E}\{I_n\} + t\} \leq \exp\left(-\frac{t^2}{8n\mathbf{E}\{I_n\} + 4nt}\right).$$

In particular, for $s > 0$,

$$\mathbf{P}\{I_n \geq (1+s)\mathbf{E}\{I_n\}\} \leq \exp\left(-\frac{s^2\mathbf{E}\{I_n\}}{8n + 4ns}\right).$$

This probability tends to zero whenever $\mathbf{E}\{I_n\}/n \to \infty$, a condition that is satisfied for all models with a finite symbol alphabet, and nearly all models with an infinite symbol alphabet as well. In fact, $\mathbf{E}\{I_n\} = \sum_i \mathbf{E}\{D_i\} = n\mathbf{E}\{D_1\}$, so that the condition holds if and only if $\mathbf{E}\{D_1\} \to \infty$.

13. Other Parameters. We can deal with other parameters of tries and PATRICIA trees with equal ease. Easiest among these is the Horton–Strahler number, which measures the minimum number of registers needed for evaluating an expression when the tree is considered as an expression tree. This quantity is always bounded by $O(\log_2 n)$ and can change by at most one when any string is deleted or added. It is thus stable in a strong sense, and concentration inequalities are easily derived. In the context of tries, this quantity is however less important, as most models of random tries do not qualify as good models of expression trees.

The stack size of a tree is the maximal stack size needed when traversing a tree in preorder, while the last subtree (the last trie symbol) is not put on the stack. It is bounded by the height of the tree, and is analyzed for a number of random trie models by Bourdon et al. (2001). For the PATRICIA tree, this parameter can be dealt with using the configuration function method used for the height. The details are left to the reader.

If N is the number of nodes in a PATRICIA tree with k children, then adding a new string can increase N by at most one. The Azuma–McDiarmid inequality thus immediately implies

$$\mathbf{P}\{|N - \mathbf{E}\{N\}| \geq t\} \leq 2e^{-t^2/2n} .$$

This should suffice for most situations. In this case, V_+ is bounded by the number of nodes with $k-1$ children, and, thus, a different argument than that provided by Lemmas 2 and 3 is needed, if one wants to improve on the Azuma–McDiarmid inequality.

Our analysis still requires the strings to be independent. It is an interesting challenge to deal with dependent strings, such as those occurring in suffix tries and suffix trees. In those cases we will invariably have to place conditions on the string symbols themselves, as most concentration inequalities, at their core, require independence. This will be dealt with elsewhere.

Acknowledgment. The author thanks both referees for their help.

References

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

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

J. L. Bentley and W. A. Burkhard (1976), Heuristics for partial-match retrieval in database design, *Information Processing Letters*, vol. 4, no. 5, pp. 132–135.

S. Boucheron, G. Lugosi, and P. Massart (2000), A sharp concentration inequality with applications in random combinatorics and learning, *Random Structures and Algorithms*, vol. 16, pp. 277–292.

S. Boucheron, G. Lugosi, and P. Massart (2003), Concentration inequalities using the entropy method, *The Annals of Probability*, vol. 31, pp. 1583–1614.

J. Bourdon (2001), Size and path length of Patricia tries: dynamical sources context, *Random Structures and Algorithms*, vol. 19, pp. 289–315.

J. Bourdon (2002), Analyse dynamique d'algorithmes: exemples en arithmétique et en théorie de l'information, Thèse de doctorat, Université de Caen Basse-Normandie.

J. Bourdon, M. Nebel, and B. Vallée (2001), On the stack size of general tries, *Theoretical Informatics and Applications*, vol. 35, pp. 163–185.

J. Clément, P. Flajolet, and B. Vallée (1998), The analysis of hybrid trie structures, in: *Proceedings of the Ninth Annual ACM–SIAM Symposium on Discrete Algorithms* (*San Francisco*, *CA*, 1998), pp. 531–539, ACM, New York.

J. Clément, P. Flajolet, and B. Vallée (2001), Dynamical sources in information theory: a general analysis of trie structures, *Algorithmica*, vol. 29, pp. 307–369.

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

L. Devroye (1985), The expected length of the longest probe sequence when the distribution is not uniform, *Journal of Algorithms*, vol. 6, pp. 1–9.

L. Devroye (1992a), A note on the probabilistic analysis of PATRICIA trees, *Random Structures and Algorithms*, vol. 3, pp. 203–214.

L. Devroye (1992b), A study of trie-like structures under the density model, *Annals of Applied Probability*, vol. 2, pp. 402–434.

L. Devroye (2002), Laws of large numbers and tail inequalities for random tries and Patricia trees, *Journal of Computational and Applied Mathematics*, vol. 142, pp. 27–37.

L. Devroye and C. Zamora-Cura (2004), Expected worst-case time for partial match in random quadtries, *Discrete Applied Mathematics*, vol. 141, pp. 103–117.

B. Efron and C. Stein (1981), The jackknife estimate of variance, *Annals of Statistics*, vol. 9, pp. 586–596.

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

P. Flajolet and C. Puech (1986), Tree structure for partial match retrieval, *Journal of the ACM*, vol. 33, pp. 371–407.

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

P. Flajolet and J. M. Steyaert (1982), 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, Berlin.

P. Flajolet and B. Vallée (1998), Continued fraction algorithms, functional operators, and structure constants, *Theoretical Computer Science*, vol. 194, pp. 1–34.

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

G. H. Gonnet (1981), Expected length of the longest probe sequence in hash code searching, *Journal of the ACM*, vol. 28, pp. 289–304.

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

P. Jacquet and M. Régnier (1986), Trie partitioning process: limiting distributions, in: *CAAP* 86 (edited by P. Franchi-Zannettacci), pp. 196–210, Lecture Notes in Computer Science, vol. 214, Springer-Verlag, Berlin.

P. Jacquet and M. Régnier (1989), Normal limiting distribution for the size and the external path length of tries, Technical Report, INRIA, Rocquencourt.

P. Jacquet and W. Szpankowski (1991), Analysis of digital tries with Markovian dependency, *IEEE Transactions on Information Theory*, vol. IT-37, pp. 1470–1475.

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

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

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

P. Kirschenhofer, H. Prodinger, and W. Szpankowski (1993), Multidimensional digital searching and some new parameters in tries, *International Journal of Foundations of Computer Science*, vol. 4, pp. 69–84.

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

M. Ledoux (1966a), On Talagrand's deviation inequalities for product measures, *ESAIM*: *Probability and Statistics*, vol. 1, pp. 63–87.

M. Ledoux (1966b), Isoperimetry and gaussian analysis, in: *Lectures on Probability Theory and Statistics* (edited by P. Bernard), pp. 165–294, Ecole d'Eté de Probabilités de St-Flour XXIV-1994, Springer-Verlag, Berlin.

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

C. McDiarmid (1985), Concentration, in: *Probabilistic Methods for Algorithmic Discrete Mathematics* (edited by M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, and B. Reed), pp. 195–248, Springer-Verlag, New York.

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

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

J. A. Orenstein (1982), Multidimensional tries used for associative searching, Technical Report, School of Computer Science, McGill University, Montreal.

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

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

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

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

C. Puech and H. Yahia (1985), Quadtrees, octrees, hyperoctrees: a unified analytical approach to tree data structures used in graphics, geometric modeling and image processing, in: *Proceedings of the Symposium on Computational Geometry*, pp. 272–280, ACM, New York.

B. Rais, P. Jacquet, and W. Szpankowski (1993), A limiting distribution for the depth in Patricia tries, *SIAM Journal on Discrete Mathematics*, vol. 6, pp. 197–213.

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

M. Régnier (1981), Trie hashing analysis, in: *Proceedings of the Fourth International Conference on Data Engineering*, pp. 377–381, IEEE, Los Angeles, CA.

M. Régnier and P. Jacquet (1989), New results on the size of tries, *IEEE Transactions on Information Theory*, vol. IT-35, pp. 203–205.

W. Schachinger (1995), The variance of a partial match retrieval in a multidimensional symmetric trie, *Random Structures and Algorithms*, vol. 7, pp. 81–95.

W. Schachinger (2000), Limiting distributions for the costs of partial match retrievals in multidimensional tries, *Random Structures and Algorithms*, vol. 17, pp. 428–459.

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

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

W. Szpankowski (1990), Patricia trees again revisited, *Journal of the ACM*, vol. 37, pp. 691–711.

W. Szpankowski (1991), On the height of digital trees and related problems, *Algorithmica*, vol. 6, pp. 256–277.

W. Szpankowski and C. Knessl (2000), Heights in generalized tries and PATRICIA tries, in: *LATIN* '2000, pp. 298–307, Lecture Notes in Computer Science, vol. 1776, Springer-Verlag, Berlin.

M. Talagrand (1998), An isoperimetric theorem on the cube and the Kintchine–Kahane inequalities, *Proceedings of the American Mathematical Society*, vol. 104, pp. 905–909.

M. Talagrand (1989), Isoperimetry and integrability of the sum of independent Banach-space valued random variables, *Annals of Probability*, vol. 17, pp. 1546–1570.

M. Talagrand (1990), Sample unboundedness of stochastic processes under increment conditions, *Annals of Probability*, vol. 18, pp. 1–49.

M. Talagrand (1991a), A new isoperimetric inequality and the concentration of measure phenomenon, in: *Geometric Aspects of Functional Analysis* (edited by J. Lindenstrauss and V. D. Milman), pp. 4–137, Lecture Notes in Mathematics, vol. 146, Springer-Verlag, New York.

M. Talagrand (1991b), A new isoperimetric inequality for product measure and the tails of sums of independent random variables, *Geometric and Functional Analysis*, vol. 1, pp. 211–223.

M. Talagrand (1993a), Isoperimetry, logarithmic Sobolev inequalities on the discrete cube, and Margulis' graph connectivity theorem, *Geometric and Functional Analysis*, vol. 3, pp. 295–314.

M. Talagrand (1993b), New gaussian estimates for enlarged balls, *Geometric and Functional Analysis*, vol. 3, pp. 502–526.

M. Talagrand (1994), Sharper bounds for gaussian and empirical processes, *Annals of Probability*, vol. 22, pp. 28–76.

M. Talagrand (1995), Concentration of measure and isoperimetric inequalities in product spaces, *Institut des Hautes Etudes Scientifiques*. *Publications Mathématiques*, vol. 81, pp. 73–205.

M. Talagrand (1996a), New concentration inequalities in product spaces, *Inventiones Mathematicae*, vol. 126, pp. 505–563.

M. Talagrand (1996b), A new look at independence, *Annals of Probability*, vol. 24, pp. 1–34.

A. W. van der Vaart and J. A. Wellner (1996), *Weak Convergence and Empirical Processes*, Springer-Verlag, New York.