LAWS OF LARGE NUMBERS AND TAIL INEQUALITIES FOR RANDOM TRIES AND PATRICIA TREES

Luc Devroye
School of Computer Science
McGill University
Montreal, Canada H3A 2K6
luc@cs.mcgill.ca

June 25, 2001

ABSTRACT. We consider random tries and random PATRICIA trees constructed from n independent strings of symbols drawn from any distribution on any discrete space. If H_n is the height of this tree, we show that $H_n/\mathbb{E}\{H_n\}$ tends to one in probability. Additional tail inequalities are given for the height, depth, size, and profile of these trees and ordinary tries that apply without any conditions on the string distributions—they need not even be identically distributed.

KEYWORDS AND PHRASES. Trie, PATRICIA tree, probabilistic analysis, law of large numbers, concentration inequality, height of a tree.

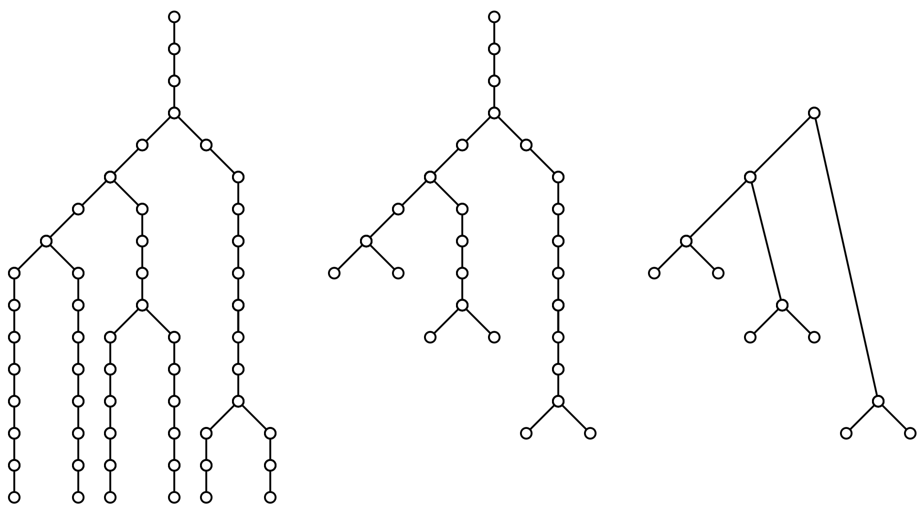
CR CATEGORIES: 3.74, 5.25, 5.5.

1991 MATHEMATICS SUBJECT CLASSIFICATIONS: 60D05, 68U05.

Author' address: School of Computer Science, McGill University, 3480 University Street, Montreal, Canada H3A 2K6. The author' research was sponsored by NSERC Grant A3456 and FCAR Grant 90-ER-0291. FAX number: 1–514–398-3883.

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 i-th 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 won't 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}-st 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 representring 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. Thus, it necessarily has n leaves. Each non-leaf (or internal) node has two or more children.



The left figure shows an infinite binary trie. In the middle, the suffixes are trimmed away to obtain a six string trie, the "finite trie". Removing the one-child nodes yields the PATRICIA tree on the right.

The purpose of this short note is to draw attention to a few specialized concentration inequalities that may be used to obtain powerful universal results for random tries and random PATRICIA trees with almost no work. The heights and the profiles of these trees are taken as prototype examples to make that point. For example, we will show that PATRICIA trees have a remarkable universal property, namely that

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

in probability as $n \to \infty$, regardless of the string distribution, where H_n denotes the height of the PATRICIA tree. We will not be concerned with the computation of $\mathbb{E}\{H_n\}$, as this depends very heavily on the string distribution. The modern concentration inequalities are mainly due to Talagrand (1988, 1989, 1990, 1991a-b, 1993a-b, 1994, 1995, 1996a-b) and Ledoux (1996a-b), as surveyed by McDiarmid (1998). An interesting inequality by Boucheron, Lugosi and Massart (2000), extended below in Lemma 1, will be helpful in the development of the results.

Boucheron-Lugosi-Massart inequality

The following inequalities will be fundamental for the remainder of the paper. Lemma 1 is an almost trivial extension of a similar inequality due to Boucheron, Lugosi and Massart (2000). Its proof is based on logarithmic Sobolev inequalities developed in part by Ledoux (1996a).

LEMMA 1. *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 $x = (x_1, \ldots, x_n) \in \Omega$:*

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

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

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

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

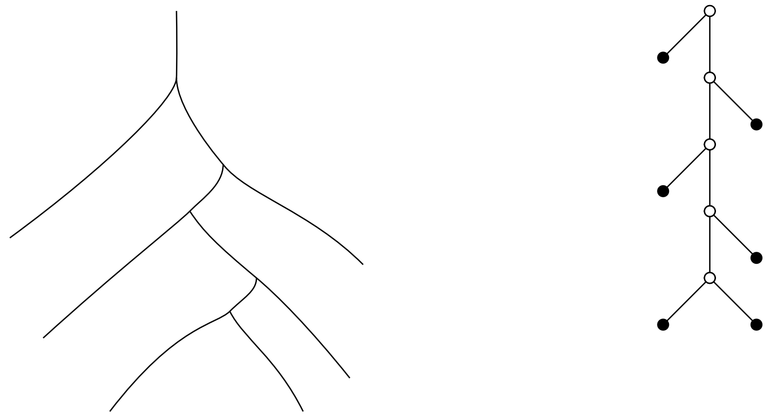
PROOF. In the proof of Theorem 6 of Boucheron, Lugosi and Massart (1999), note that in (16), it suffices to replace v by $v + c$. $\square$

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 $i \leq n$ 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.

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. This property is clearly hereditary, and $H_n + 1$ is thus a configuration function.



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

We have

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

and

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

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 stable and well-behaved:

THEOREM 1. *For any* PATRICIA *tree constructed by using n independent strings, if* $\lim_{n\to\infty} \mathbb{E}\{H_n\} = \infty$, *then*

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

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

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

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

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

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

THE CONDITION ON $\mathbb{E}\{H_n\}$. In PATRICIA trees of bounded degree, it is clear that $\mathbb{E}\{H_n\} \to \infty$. 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 i-th string be $(i, 0, 0, 0, \ldots)$.

BIBLIOGRAPHIC REMARKS: STRING MODELS. In the **uniform trie model**, the bits in the string X_1 are i.i.d. Bernoulli random variables with success probability $p = 0.5$. In a **non-uniform trie model**, the symbols in the string X_1 are i.i.d. $\mathcal{Z}$-valued random variables, with $\mathbb{P}\{\text{symbol} = j\} = p_j$. In the **density model**, X_1 consists of the bits in the binary expansion of a $[0, 1]$-valued random variable X (Devroye, 1982, 1984). In the **Markov model**, the symbols themselves form a Markov chain with a given fixed transition matrix over $\mathcal{Z} \times \mathcal{Z}$, and with a fixed distribution for the first symbol (Régnier (1988), Szpankowski (1988), Jacquet and Szpankowski (1991) and Pittel (1985)). More exotic models were studied by Clément, Flajolet and Vallée (1999), who considered strings of partial quotients in the continued fractions expansion of certain random variables (this creates a peculiarly dependent sequence). Theorem 1 above applies to all models described above.

BIBLIOGRAPHIC REMARKS: 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), Kirschenhofer and Prodinger (1986) and Szpankowski (1990, 1991). Pittel and Rubin (1990), Pittel (1991) and Devroye (1992) showed that

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

More refined results for general multi-branching PATRICIA trees and tries are given by Szpankowski and Knessl (2000). For the non-uniform trie model, we have $\mathbb{E}\{H_n\} \sim c \log n$, where $c = 2/\log_2(1/\sum_j p_j^2)$.

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:

THEOREM 2. *For any PATRICIA tree constructed by using n independent strings, if x is a string such that $\lim_{n\to\infty} \mathbb{E}\{D_n(x)\} = \infty$, then*

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

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

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

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

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

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 non-binary 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 independent identically distributed symbols, or when the tree is of bounded fan-out). Even though these results do not require Lemma 1, they appear to be new.

Balls in urns and hashing

Consider a very general urn model in which we have n balls thrown independently into a countable number of urns, where the i-th urn has probability p_i of receiving a ball. Let $N_1, N_2, \ldots$ be the numbers of balls in the urns. Quantities of interest in certain applications include $M_n = \max_i N_i$, the maximum number of balls, and $O_n = \sum_i 1_{N_i > 0}$, the number of occupied urns. If we throw one less ball, then M_n and O_n both decrease by at most one. Thus, uniformly over all urn probabilities, 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 urn", and O_n being the configuration function for the hereditary property "belonging to different urns". Thus, by Lemma 1,

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

and

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

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

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

And 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 urns (in which the p_i's are allowed to change with n). 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 urns equals the number of balls, then $M_n \sim \log n / \log\log n$ if each urn 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.

Profile of a trie

Consider an infinite trie constructed based on n infinite strings with symbols drawn from an arbitrary alphabet. At level m, or distance m from the root, we count the number N_m of nodes that are visited by at least one string. Clearly, N_m is a random monotone function in m, increasing from $N_0 = 1$ to (usually) n. Let Q_m be the number of nodes at level m that are visited by at least two strings. We note that Q_m is the number of internal trie nodes at level m in the finite trie. Also, $L_m \stackrel{\text{def}}{=} N_m - N_{m-1}$ is the number of leaves at level m in the finite trie. The number of nodes at level m is thus $Q_m + (N_m - N_{m-1})$. As a function of m, this is a random sequence usually called the profile. We note that Lemma 1 is applicable to the quantities Q_m and N_m. This then yields very simple inequalities and proofs for the behavior of these quantities.

We note here the analogy with urns. Consider the m-prefixes of the strings $X_1, \ldots, X_n$. Each m-prefix takes values in Ω^m, where Ω is the symbol alphabet. The probability of each element of Ω^m is thus fixed once and for all. Each of the n strings is associated with such an element, very much the way we drop balls in urns (elements of Ω^m) of unequal probability. Clearly, N_m counts the number of occupied urns. If $f(X_1, \ldots, X_n) = N_m$, and $g_i(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ is similarly defined for $n-1$ strings, then $0 \le f - g_i \le 1$, and $\sum_i (f - g_i) \le f$, so the conditions of Lemma 1 are satisfied. We thus have

$$\mathbf{P}\{N_m \ge \mathbf{E}\{N_m\} + t\} \le \exp\left(-\frac{t^2}{2\mathbf{E}\{N_m\} + 2t/3}\right) \ , \ t \ge 0 \ ,$$

and

$$\mathbf{P}\{N_m \le \mathbf{E}\{N_m\} - t\} \le \exp\left(-\frac{t^2}{2\mathbf{E}\{N_m\}}\right) \ , \ t \ge 0 \ .$$

This leads to laws for the profile of the infinite trie. The profile of any trie is close to $\mathbf{E}\{N_m\}$ for a wide range of levels m. This, again, is true regardless of the distribution of X_1, and regardless of the fanout of the trie.

Consider the number of leaves L_m at level m. Because $N_{m-1} \le N_m$,

$$\begin{aligned}\mathbf{P}\{L_m \ge \mathbf{E}\{L_m\} + 2t\} &\le \mathbf{P}\{N_m \ge \mathbf{E}\{N_m\} + t\} + \mathbf{P}\{N_{m-1} \le \mathbf{E}\{N_{m-1}\} - t\} \\ &\le 2\exp\left(-\frac{t^2}{2\mathbf{E}\{N_m\} + 2t/3}\right) \\ &\le 2\exp\left(-\frac{t^2}{2n + 2t/3}\right) \ .\end{aligned}$$

Similarly,

$$\begin{aligned}\mathbf{P}\{L_m \le \mathbf{E}\{L_m\} - 2t\} &\le \mathbf{P}\{N_m \le \mathbf{E}\{N_m\} - t\} + \mathbf{P}\{N_{m-1} \ge \mathbf{E}\{N_{m-1}\} + t\} \\ &\le 2\exp\left(-\frac{t^2}{2\mathbf{E}\{N_m\} + 2t/3}\right) \\ &\le 2\exp\left(-\frac{t^2}{2n + 2t/3}\right) \ .\end{aligned}$$

These are indeed universal inequalities. Without further work, we have

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

in probability for all $m = m(n)$ when $\mathbf{E}\{L_m\}/\sqrt{n} \to \infty$.

For Q_m, we argue as we did for the urns. As Q_m is the number of urns that receive at least two strings, we have $Q_m = N_m - O_m$, where O_m is the number of urns receiving precisely one string. Again, with the obvious choices for $f = O_m$ and g_i, we note $0 \le f - g_i \le 1$, and $\sum_i (f - g_i) \le f$. Thus, Lemma 1 is applicable to both N_m and O_m. Therefore, for $t > 0$,

$$\mathbf{P}\{Q_m - \mathbf{E}\{Q_m\} \ge t\} \le \mathbf{P}\{N_m - \mathbf{E}\{N_m\} \ge t/2\} + \mathbf{P}\{O_m - \mathbf{E}\{O_m\} \le -t/2\}$$

and this may be bounded by applying Lemma 1 twice. However, the bounds are unsatisfactory as $\mathbf{E}\{N_m\}$ and $\mathbf{E}\{O_m\}$ are both large and near n for m large enough, and thus much larger than $\mathbf{E}\{Q_m\}$. We might thus as well use the bounded difference method directly on Q_m, after noting that adding one string can increase Q_m by at most one. Thus, directly,

$$\mathbf{P}\{|Q_m - \mathbf{E}\{Q_m\}| \ge t\} \le 2\exp\left(-\frac{t^2}{2n}\right) \ .$$

With $Q_m = f$ put in the framework of Lemma 1, we note that $0 \le f - g_i \le 1$, $\sum_i (f - g_i) \le 2f$ (note the "2"). The $2f$ causes some problems that require a considerable extension of Lemma 1, which will not be done here. Nevertheless, if m is such that $\mathbf{E}\{Q_m\} \to \infty$, then $Q_m/\mathbf{E}\{Q_m\} \to 1$ in probability.

The height of a trie from its profile

With the notation of the previous section, if H_n denotes the height of a random trie for n independent but otherwise arbitrary strings, then $[H_n < m] = [N_m \geq n]$. Thus, we have without further work,

$$\begin{aligned}\mathbb{P}\{H_n < m\} &= \mathbb{P}\{N_m \geq n\} \\ &= \mathbb{P}\{N_m \geq \mathbb{E}\{N_m\} + (n - \mathbb{E}\{N_m\})\} \\ &\leq \exp\left(-\frac{(n - \mathbb{E}\{N_m\})^2}{2\mathbb{E}\{N_m\} + 2(n - \mathbb{E}\{N_m\})/3}\right) \\ &\leq \exp\left(-\frac{(n - \mathbb{E}\{N_m\})^2}{2n}\right) .\end{aligned}$$

This is a remarkable inequality, because the right-hand-side depends solely on $\mathbb{E}\{N_m\}$. It is also valid even if the strings have different distributions! In particular, it implies that if $(n - \mathbb{E}\{N_m\})/\sqrt{n} \to \infty$, then $\mathbb{P}\{H_n < m\} \to 0$. The first moment of N_m suffices to conclude this!

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

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

The limit law of H_n was obtained in Devroye (1984), and laws of the iterated logarithm for the difference $H_n - 2\log_2 n$ can be found in Devroye (1990). 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), and Szpankowski (1988).

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

M. Ledoux, "On Talagrand's deviation inequalities for product measures," *ESAIM: Probability and Statistics*, vol. 1, pp. 63–87, 1996a.

M. Ledoux, "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, 1996b.

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

C. McDiarmid, "Concentration," in: *Probabilistic Methods for Algorithmic Discrete Mathematics*, (edited by M. Habib and C. McDiarmid and J. Ramirez-Alfonsin and B. Reed), pp. 195–248, Springer, New York, 1998.

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

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

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

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

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

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

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

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

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

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

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

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

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

W. Szpankowski and C. Knessl, "Heights in generalized tries and PATRICIA tries," in: *LATIN'2000*, pp. 298–307, Lecture Notes in Computer Science 1776, 2000.

M. Talagrand, "An isoperimetric theorem on the cube and the Kintchine-Kahane inequalities," *Proceedings of the American Mathematical Society*, vol. 104, pp. 905–909, 1988.

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

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

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

M. Talagrand, "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, 1991b.

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

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

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

M. Talagrand, "Concentration of measure and isoperimetric inequalities in product spaces," *Inst. Hautes Etudes Sci. Publ. Math.* , vol. 81, pp. 73–205, 1995.

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

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