1999

Limit Laws for Heights in Generalized Tries and PATRICIA Tries

Charles Knessl

Wojciech Szpankowski
Purdue University, spa@cs.purdue.edu

Report Number:
99-011
LIMIT LAWS FOR HEIGHTS IN GENERALIZED TRIES AND PATRICIA TRIES

Charles Knessl
Wojciech Szpankowski

CSD-TR #99-011
April 1999
Limit Laws for Heights in Generalized Tries and
PATRICIA Tries

April 7, 1999

Charles Knessl
Dept. Mathematics, Statistics & Computer Science
University of Illinois at Chicago
Chicago, Illinois 60607-7045
U.S.A
knessl@uic.edu

Wojciech Szpankowski
Department of Computer Science
Purdue University
W. Lafayette, IN 47907
U.S.A
spa@cs.purdue.edu

Abstract

We consider digital trees such as (generalized) tries and PATRICIA tries, built from \( n \) random strings generated by an unbiased memoryless source (i.e., all symbols are equally likely). We study limit laws of the height which is defined as the longest path in such trees. It turns out that this height also represents the number of random questions required to recognize \( n \) distinct objects. We shall identify three natural regions of the height distributions. For tries, in the region where most of the probability mass is concentrated, the asymptotic distribution is of extreme value type (i.e., double exponential distribution). Surprisingly enough, the height of the PATRICIA trie behaves quite differently in this region; it exhibits an exponential of a Gaussian distribution (with an oscillating term) around the most probable value \( k_1 = \lceil \log_2 n + \sqrt{2 \log_2 n - \frac{3}{2}} \rceil + 1 \). In fact, the asymptotic distribution of PATRICIA height concentrates on one or two points. For most \( n \) all the mass is concentrated at \( k_1 \), however, there exist subsequences of \( n \) such that the mass is on the two points \( k_1 - 1 \) and \( k_1 \), or \( k_1 \) and \( k_1 + 1 \). We derive these results by a combination of analytic methods such as generating functions, Mellin transform, the saddle point method and ideas of applied mathematics such as linearization, asymptotic matching and the WKB method. We present also some numerical verification of our results.

Key Words: digital trees, \( b \)-tries, PATRICIA trie, height distribution, Mellin transform, saddle point method, matched asymptotics, WKB method.
1 Introduction

Data structures and algorithms on words have experienced a new wave of interest due to a number of novel applications in computer science, communications, and biology. These include dynamic hashing, partial match retrieval of multidimensional data, searching and sorting, pattern matching, conflict resolution algorithms for broadcast communications, data compression, coding, security, genes searching, DNA sequencing, genome maps, and so forth. To satisfy these diversified demands various data structures were proposed for these algorithms. Undoubtedly, the most popular data structures for algorithms on words are digital trees [18, 21] (e.g., tries, PATRICIA tries, digital search trees), and suffix trees [14, 30]. The importance of digital trees stem from their applications in sorting and searching [5, 10, 15, 18, 21, 23, 26, 28, 29, 31], data compression [16, 30, 32, 33] pattern matching [14], the shortest common superstring problem [12], searching for a leader [9], estimating the number of questions necessary to identify many distinct objects [25], prediction [8], and so forth. These problems recently became very important due to the need for efficient storage and transmission of multimedia, and applications to DNA sequencing (cf. [14]).

The most basic digital tree is known as a trie (the name comes from retrieval). The primary purpose of a trie is to store a set $S$ of strings (words, keys), say $S = \{X_1, \ldots, X_n\}$. Each word $X = x_1x_2x_3 \ldots$ is a finite or infinite string of symbols taken from a finite alphabet. Throughout the paper, we deal only with the binary alphabet $\{0, 1\}$, but all our results should be extendable to a general finite alphabet. A string will be stored in a leaf of the trie. The trie over $S$ is built recursively as follows: For $|S| = 0$, the trie is, of course, empty. For $|S| = 1$, trie($S$) is a single node. If $|S| > 1$, $S$ is split into two subsets $S_0$ and $S_1$ so that a string is in $S_j$ if its first symbol is $j \in \{0, 1\}$. The tries trie($S_0$) and trie($S_1$) are constructed in the same way except that at the $k$-th step, the splitting of sets is based on the $k$-th symbol of the underlying strings. Figure 1 illustrates such a construction.

There are many possible variations of the trie. One such variation is the $b$-trie, in which a leaf is allowed to hold as many as $b$ strings (cf. [21, 30]). A second variation of the trie, the PATRICIA trie, eliminates the waste of space caused by nodes having only one branch. This is done by collapsing one-way branches into a single node. In a digital search tree (in short DST) strings are directly stored in nodes, and hence external nodes are eliminated. The branching policy is the same as in tries. These digital trees are also shown in Figure 1. The reader is referred to [14, 18, 21] for a detailed description of digital trees.

In this paper, we consider tries and PATRICIA tries built over $n$ randomly generated strings of binary symbols. We assume that every symbol is equally likely, thus we are within
Figure 1: A trie, PATRICIA trie and a digital search tree (DST) built from the following four strings \( X_1 = 11100 \ldots \), \( X_2 = 10111 \ldots \), \( X_3 = 00110 \ldots \), and \( X_4 = 00001 \ldots \).

the framework of the so called symmetric Bernoulli model. In other words, the strings are emitted by an unbiased memoryless source. Our interest lies in establishing asymptotic distributions of the heights for random tries and PATRICIA tries. The height is the longest path in such trees, and its distribution is of considerable interest for several applications.

Let us further motivate our study by describing a generalization of Rényi’s problem [27] analyzed by Pittel and Rubin [25]. The problem is to identify \( n \) distinct objects by asking random questions. More formally, let \( \Phi : \mathcal{X} \to \mathcal{A} \) be a bijective mapping from set \( \mathcal{X} \) into set \( \mathcal{A} \), both sets having the same cardinality \( n \). The goal is to identify all elements of \( \mathcal{X} \) by asking questions like: for a given \( \mathcal{B} \in \mathcal{A} \) what is \( \Phi^{-1}(\mathcal{B}) \). If \( \mathcal{B} \) is chosen uniformly at random, how many questions must one ask to recognize \( \mathcal{X} \)? What happens if we do not allow \( \mathcal{B} \) to be empty, or to be the same as in the previous attempt? Pittel and Rubin [25] proved that the former problem is equivalent to estimating the height in a trie, while the latter can be reduced to the height in a PATRICIA trie.

We now summarize our main results. As mentioned before, we are aiming at establishing asymptotic distributions for the \( b \)-trie height \( \mathcal{H}_n^T \) and the PATRICIA height \( \mathcal{H}_n^P \). We obtain asymptotic expansions of the distributions \( \Pr\{\mathcal{H}_n^T \leq k\} \) and \( \Pr\{\mathcal{H}_n^P \leq k\} \) for three ranges of \( n \) and \( k \). For \( b \)-tries we consider: (i) the “right-tail region” \( k \to \infty \) and \( n = O(1) \); (ii) the “central region” \( n, k \to \infty \) with \( \xi = n2^{-k} \) and \( 0 < \xi < b \); and (iii) the “left-tail region” \( k, n \to \infty \) with \( n - b2^k = O(1) \). We prove that most probability mass is concentrated in between the right tail and the central region. In particular, for real \( x \)

\[
\Pr\left\{ \mathcal{H}_n^T \leq \frac{1 + b}{b} \log_2 n + x \right\} \sim \exp\left( -\frac{1}{(b + 1)!} 2^{-\frac{b+1}{b}(\frac{1+b}{b} \log_2 n + x)} \right),
\]

where \( \langle r \rangle = r - \lfloor r \rfloor \) is the fractional part of \( r \).\(^1\) In words, the asymptotic distribution

\(^1\)The fractional part \( \langle r \rangle \) is often denoted as \( \{ r \} \), but in order to avoid confusion we adopt the above
of b-tries height around its most likely value \( \frac{1+\delta}{6} \log_2 n \) resembles a double exponential (extreme value) distribution. In fact, due to the oscillating term \( (\frac{1+\delta}{6} \log_2 n + x) \) the limiting distribution does not exist, but one can find lim inf and lim sup of the distribution (cf. Corollary 1 and Figure 2 in the next section).

The height of PATRICIA tries behaves differently in the central region (i.e., where most of the probability mass is concentrated). It is concentrated at or near the most likely value \( k_1 = [\log_2 n + \sqrt{2\log_2 n} - \frac{3}{2}] + 1 \). We shall prove that the asymptotic distribution around \( k_1 \) resembles an exponential of a Gaussian distribution, with an oscillating term (cf. Theorem 3). In fact, there exist subsequences of \( n \) such that the asymptotic distribution of PATRICIA height concentrates only on \( k_1 \), or on \( k_1 \) and one of the two points \( k_1 - 1 \) or \( k_1 + 1 \). Later, we characterize precisely these subsequences.

With respect to previous results, Devroye [5] and Pittel [24] established the asymptotic distribution in the central regime for tries and b-tries, respectively, using probabilistic tools. Jacquet and Régnier [15] obtained similar results by analytic methods. The most probable value, \( \log_2 n \), of the height for PATRICIA was first proved by Pittel [23]. This was then improved to \( \log_2 n + \sqrt{2\log_2 n}(1 + o(1)) \) by Pittel and Rubin [25], and independently by Devroye [6]. No results concerning the asymptotic distribution for PATRICIA height were reported.

Finally, we say a few words about our method of derivation, and put our results in a larger perspective. From a mathematical view point, we study two non-linear recurrence equations. The distribution \( \tilde{h}_n^k = \Pr\{H_n^T \leq k\} \) of the height of b-tries satisfies

\[
\tilde{h}_n^{k+1} = 2^{-n} \sum_{i=0}^{n} \binom{n}{i} \tilde{h}_n^k \tilde{h}_n^{k-i}, \quad k \geq 0
\]

with the initial condition \( \tilde{h}_0^0 = 1 \) for \( n = 0, 1, 2, \ldots, b \) and \( \tilde{h}_n^0 = 0 \) for \( n > b \). For PATRICIA tries the distribution \( h_n^k = \Pr\{H_n^P \leq k\} \) satisfies

\[
h_n^{k+1} = 2^{-n+1} h_n^{k+1} + 2^{-n} \sum_{i=1}^{n-1} \binom{n}{i} h_i^k h_{n-i}^k, \quad k \geq 0
\]

with the initial conditions \( h_0^0 = h_1^0 = 1 \) and \( h_n^0 = 0 \) for \( n \geq 2 \). More importantly, the PATRICIA recurrence satisfies an additional boundary condition, namely, \( h_n^k = 1 \) for \( k \geq n - 1 \), which does not hold for tries. This boundary condition seems to have an enormous impact on the final solution, as shown by our results (cf. Theorem 2 and Theorem 3).

We use two different methodologies to solve these recurrences. The trie recurrence is first analyzed by analytic methods (i.e., generating functions, Mellin transform, saddle notation).
point method). We then re-derive the results by methods of applied mathematics such as matched asymptotics and the WKB method. These are also analytic methods and are especially suitable for problems that cannot be solved exactly by transform methods, such as the PATRICIA model. The approach we suggest works also for other problems (e.g., height of digital search trees and quicksort [19]).

The paper is organized as follows. In the next section, we present and discuss our main results for tries (cf. Theorem 1), b-trees (cf. Theorem 2), and PATRICIA tries (cf. Theorem 3). In Section 3 we derive the results for tries and b-trees, while in Section 4 we deal with the PATRICIA trie. Finally, the last section presents some numerical results.

2 Summary of Results

As before, we let \( \mathcal{H}^T_n \) and \( \mathcal{H}^P_n \) denote, respectively, the height of a b-trie and a PATRICIA trie. Their probability distributions are

\[
\tilde{h}^k_n = \Pr\{\mathcal{H}^T_n \leq k\} \quad \text{and} \quad h^k_n = \Pr\{\mathcal{H}^P_n \leq k\}. \tag{2.1}
\]

We note that for tries \( \tilde{h}^k_n = 0 \) for \( n > \delta 2^k \) (corresponding to a balanced tree), while for PATRICIA tries \( h^k_n = 0 \) for \( n > 2^k \). In addition, for PATRICIA we have the following boundary condition: \( h^k_n = 1 \) for \( k \geq n \). It asserts that the height in a PATRICIA trie cannot be bigger than \( n \) (due to the elimination of all one-way branches).

The distribution of b-trees satisfies the recurrence relation

\[
\tilde{h}^{k+1}_n = 2^{-n} \sum_{i=0}^{n} \binom{n}{i} \tilde{h}^k_i \tilde{h}^k_{n-i}, \quad k \geq 0 \tag{2.2}
\]

with the initial condition(s)

\[
\tilde{h}^0_n = 1, \quad n = 0, 1, 2, \ldots, \delta; \quad \text{and} \quad \tilde{h}^0_n = 0, \quad n > b. \tag{2.3}
\]

This follows from \( \mathcal{H}^T_n = \max\{\mathcal{H}^{LT}_i, \mathcal{H}^{KT}_{n-i}\} + 1 \), where \( \mathcal{H}^{LT}_i \) and \( \mathcal{H}^{KT}_{n-i} \) denote, respectively, the left subtree and the right subtree of sizes \( i \) and \( n - i \), which happens with probability \( 2^{-n} \binom{n}{i} \). Similarly, for PATRICIA tries we have

\[
h^{k+1}_n = 2^{-n+1} h^{k+1}_n + 2^{-n} \sum_{i=1}^{n-1} \binom{n}{i} h^k_i h^k_{n-i}, \quad k \geq 0 \tag{2.4}
\]

with the initial conditions

\[
h^0_0 = h^0_1 = 1 \quad \text{and} \quad h^0_n = 0, \quad n \geq 2. \tag{2.5}
\]
Unlike $b$-tries, in a PATRICIA trie the left and the right subtrees cannot be empty (which occurs with probability \(2^{-n+1}\)).

We shall analyze these problems asymptotically, in the limit \(n \to \infty\). Despite the similarity between (2.2) and (2.4), we will show that even asymptotically the two distributions behave very differently.

We first consider ordinary tries (i.e., 1-tries). It is relatively easy to solve (2.2) and (2.3) explicitly and obtain the integral representation (cf. Section 3)

\[
\mathbb{h}_n^k = \frac{n!}{2\pi i} \oint (1 + z2^{-k})^{2k} z^{-n-1} dz \tag{2.6}
\]

\[
= \begin{cases} 
0, & n > 2^k \\
\frac{(q^k)^n}{z^n(z^{2^k-n})!}, & 0 \leq n \leq 2^k.
\end{cases}
\]

Here the loop integral is for any closed circle surrounding \(z = 0\).

Using asymptotic methods for evaluating integrals, or applying Stirling’s formula to the second part of (2.6), we obtain the following.

**Theorem 1** The distribution of the height of tries has the following asymptotic expansions:

(i) **Right-Tail Region:** \( k \to \infty, \ n = O(1) \)

\[
\Pr\{\mathcal{H}_n^T \leq k\} = \mathbb{h}_n^k = 1 - n(n - 1)2^{-k-1} + O(2^{-2k}).
\]

(ii) **Central Region:** \( k, n \to \infty \) with \( \xi = n2^{-k}, \ 0 < \xi < 1 \)

\[
\mathbb{h}_n^k \sim A(\xi) e^{n\phi(\xi)},
\]

where

\[
\phi(\xi) = \left(1 - \frac{1}{\xi}\right) \log(1 - \xi) - 1,
\]

\[
A(\xi) = (1 - \xi)^{-1/2}.
\]

(iii) **Left-Tail Region:** \( k, n \to \infty \) with \( 2^k - n = j = O(1) \)

\[
\mathbb{h}_n^k \sim n^j e^{-n-j} \frac{j!}{j^{j}} \sqrt{2\pi n}.
\]

This shows that there are three ranges of \( k \) and \( n \) where the asymptotic form of \( \mathbb{h}_n^k \) is different.
We next consider the "asymptotic matching" (see [20]) between the three expansions.
If we expand (i) for \( n \) large, we obtain 
\[ 1 - h^k_n \sim n^2 2^{-k/2}. \]
For \( \xi \to 0 \) we have \( A(\xi) \sim 1 \) and 
\( \phi(\xi) \sim -\xi/2 \) so that the result in (ii) becomes
\[ A(\xi)e^{\phi(\xi)} \sim e^{-n\xi/2} = \exp \left( -\frac{1}{2} n^2 2^{-k} \right) \sim 1 - \frac{1}{2} n^2 2^{-k} \] (2.7)
where the last approximation assumes that \( n, k \to \infty \) in such a way that \( n^2 2^{-k} \to 0 \).
Since (2.7) agrees precisely with the expansion of (i) as \( n \to \infty \), we say that (i) and (ii)
asymptotically match. To be precise, we say they match the leading order; higher order
matchings can be verified by computing higher order terms in the asymptotic series in (i)
and (ii). We can easily show that the expansion of (ii) as \( \xi \to 1^- \) agrees with the expansion
of (iii) as \( j \to \infty \), so that (ii) and (iii) also asymptotically match. The matching verifications
imply that, at least to leading order, there are no "gaps" in the asymptotics. In other words,
one of the results in (i)-(iii) applies for any asymptotic limit which has \( k \) and/or \( n \) large.
We recall that \( h^k_n = 0 \) for \( n > 2^k \) so we need only consider \( k \geq \log_2 n \).
The asymptotic limits where (i)-(iii) apply are the three "natural scales" for this problem.
We can certainly consider other limits (such as \( k, n \to \infty \) with \( k/n \) fixed), but the
expansions that apply in these limits would necessarily be limiting cases of one of the three
results in Theorem 1. In particular, if we let \( k, n \to \infty \) with \( k - 2 \log_2 n = O(1) \), we are led to
\[ h^k_n \sim \exp \left( -\frac{1}{2} n^2 2^{-k} \right) = \exp \left( -\frac{1}{2} \exp(-k \log 2 + 2 \log n) \right). \] (2.8)
This result is well-known (see [5, 15]) and corresponds to a limiting double exponential (or
extreme value) distribution. However, according to our discussion, \( k = 2 \log_2 n + O(1) \) is
not a natural scale for this problem. The scale \( k = \log_2 n + O(1) \) (where (ii) applies) is a
natural scale, and the result in (2.8) may be obtained as a limiting case of (ii), by expanding
(ii) for \( \xi \to 0 \).
To obtain Theorem 1, we have identified the natural scales from the representations of
\( h^k_n \) in (2.6), and made no use of asymptotic matching. However, for problems such as (2.4)
(with (2.5)), which have not been solved exactly, we shall show that asymptotic matching is
a very useful tool for analyzing the model as \( k, n \to \infty \). For such problems we must identify
the natural scales using only the recursion equation, such as (2.4). We discuss this in more
deepth in Sections 3 and 4.
We next generalize Theorem 1 to arbitrary \( b_i \) and obtain the following. The proof can
be found in Section 3.
**Theorem 2** The distribution of the height of \( b \)-tries has the following asymptotic expansions for fixed \( b \):

(i) **RIGHT-TAIL REGION:** \( k \to \infty, \ n = O(1) \):

\[
\Pr\left( \mathcal{H}^b_n \leq k \right) = \tilde{h}_n^k \sim 1 - \frac{n!}{(b + 1)!(n - b - 1)!} 2^{-kb}.
\]

(ii) **CENTRAL REGIME:** \( k, n \to \infty \) with \( \xi = n^{2-k}, \ 0 < \xi < b \):

\[
\tilde{h}_n^k \sim A(\xi; b) e^{n\phi(\xi; b)},
\]

where

\[
\phi(\xi; b) = -1 - \log \omega_0 + \frac{1}{\xi} \left( b \log(\omega_0 \xi) - \log b! - \log \left( 1 - \frac{1}{\omega_0} \right) \right),
\]

\[
A(\xi; b) = \frac{1}{\sqrt{1 + (\omega_0 - 1)\left( \xi - b \right)}}.
\]

In the above, \( \omega_0 = \omega_0(\xi; b) \) is the solution to

\[
1 - \frac{1}{\omega_0} = \frac{(\omega_0 \xi)^b}{b! \left( 1 + \omega_0 \xi + \frac{\omega_0^2 \xi^2}{2!} + \cdots + \frac{\omega_0^b \xi^b}{b!} \right)}.
\]

(iii) **LEFT-TAIL REGION:** \( k, n \to \infty \) with \( j = b2^k - n \)

\[
\tilde{h}_n^k \sim \sqrt{2\pi n} \frac{\xi^j}{j!} b^n \exp \left( -(n+j) \left( 1 + b^{-1} \log b! \right) \right)
\]

where \( j = O(1) \).

We note that when \( b = 1 \) or 2, we can obtain \( \omega_0 \) explicitly:

\[
\omega_0(\xi; 1) = \frac{1}{1 - \xi}, \quad (2.9)
\]

\[
\omega_0(\xi; 2) = \frac{2}{1 - \xi + \sqrt{1 + 2\xi - \xi^2}}. \quad (2.10)
\]

For arbitrary \( b \), we have \( \omega_0 \to \infty \) as \( \xi \to b^- \) and \( \omega_0 \to 1 \) as \( \xi \to 0^+ \). More precisely,

\[
\omega_0 = 1 - \frac{\xi^b}{b^b} + O(\xi^{b+1}), \quad \xi \to 0 \quad (2.11)
\]

\[
\omega_0 = \frac{1}{b - \xi} + \frac{b - 1}{b} + O(b - \xi), \quad \xi \to b. \quad (2.12)
\]

When \( b = 1 \) we can easily show (using (2.9)) that Theorem 2 reduces to Theorem 1.
Using (2.11) we can also show that the three parts of Theorem 2 asymptotically match. In particular, by expanding part (ii) as \( \xi \to 0 \) we obtain
\[
\Pr\{H_n \leq k\} \sim A(\xi)e^{n\phi(\xi)} \sim \exp \left( -\frac{n\xi}{b(1+b)!} \right) \quad \xi \to 0
\]
\[
= \exp \left( -\frac{n^{1+b/2-bb}}{b(1+b)!} \right).
\]

This yields the well-known (see [15, 24]) asymptotic distribution of \( b \)-tries. We note that, for \( k, n \to \infty \), (2.13) is \( O(1) \) for \( k - (1 + 1/b) \log_2 n = O(1) \). More precisely, let us estimate the probability mass of \( H_n^T \) around \( (1 + 1/b) \log_2 n + x \) where \( x \) is a real fixed value. We observe from (2.13) that
\[
\Pr\{H_n^T \leq (1 + 1/b) \log_2 n + x\} = \Pr\{H_n^T \leq [(1 + 1/b) \log_2 n + x]\}
\sim \exp \left( -\frac{1}{(1+b)!}2^{-bx+b((1+b)/b)\log_2 n+x} \right),
\]
where \( \langle x \rangle \) is the fractional part of \( x \), that is, \( \langle x \rangle = x - [x] \). Since \( \langle \log_2 n \rangle \) is dense in \([0,1]\) but not uniformly dense, the limit of (2.14) does not exist as \( n \to \infty \). We can, however, conclude the following.

**Corollary 1** While the limiting distribution of the height for \( b \)-tries does not exist, the following lower and upper envelopes can be established
\[
\liminf_{n \to \infty} \Pr\{H_n^T \leq (1 + 1/b) \log_2 n + x\} = \exp \left( -\frac{1}{(1+b)!}2^{-b(x-\langle x \rangle)} \right),
\]
\[
\limsup_{n \to \infty} \Pr\{\mathcal{H}_n^T \leq (1 + 1/b) \log_2 n + x\} = \exp \left(-\frac{1}{(1+b)!} 2^{-bx}\right)
\]

for fixed real \(x\).

In Figure 2 we plot the asymptotic distribution (2.14) (the stair-wise function) together with the lower and the upper envelopes for \(b = 2\) and \(b = 10\). We observe that for large \(b\) the asymptotic distribution becomes more concentrated on one or two points.

We next turn our attention to PATRICIA tries. Using ideas of applied mathematics, such as linearization and asymptotic matching, we obtain the following. The derivation of this result is presented in Section 4 where we make certain assumptions about the forms of the asymptotic expansions, as well as the asymptotic matching between the various scales.

**Theorem 3** The distribution of PATRICIA tries has the following asymptotic expansions:

(i) **Right-Tail Regime:** \(k, n \to \infty\) with \(n - k - j = O(1), j \geq 2\)

\[
\Pr\{\mathcal{H}_n \leq n - j\} = h_n^{n-j} \sim 1 - \rho_0 K_j \cdot n! \cdot 2^{-n^2/2+2j-3/2}n^j,
\]

where

\[
K_j = \frac{1}{j!} 2^{-j/2} 2^{3j/2} C_j, \quad (2.15)
\]

\[
C_j = \frac{j!}{2^j} \int \frac{z^{1-j/2} e^z}{2} \prod_{m=0}^{\infty} \left(1 - \exp\left(-\frac{m}{2}\right)\right) dz, \quad (2.16)
\]

and \(\rho_0 = \prod_{t=2}^{\infty} (1 - 2^{-t})^{-1} = 1.73137\ldots\)

(ii) **Central Regime:** \(k, n \to \infty\) with \(\xi = n2^{-k}, 0 < \xi < 1\)

\[
h_n^k \sim \sqrt{1 + 2\xi \Phi'(\xi) + \xi^2 \Phi''(\xi) e^{-n\Phi(\xi)}}.
\]

We know \(\Phi(\xi)\) analytically only for \(\xi \to 0\) and \(\xi \to 1\). In particular, for \(\xi \to 0\)

\[
\Phi(\xi) \sim \frac{1}{2} \rho_0 e^{\varphi(\log_2 \xi)} \xi^{3/2} \exp \left(-\frac{\log^2 \xi}{2\log 2}\right), \quad \xi \to 0^+,
\]

with

\[
\varphi(x) = \frac{\log 2}{2} x(x+1) + \sum_{t=0}^{\infty} \log \left(1 - \exp(-2^{x-t})\right) + \sum_{t=1}^{\infty} \log(1 - \exp(-2^{x-t}))
\]

\[
= \Psi(x) - \frac{\log 2}{12} + \frac{1}{\log 2} \left(\frac{\gamma^2}{2} + \gamma(1 - \frac{\pi^2}{12})\right), \quad (2.18)
\]

\[
\Psi(x) = \sum_{t=0}^{\infty} \frac{1}{2\pi i t} \Gamma \left(1 - \frac{2\pi i t}{\log 2}\right) \left(1 - \frac{2\pi i t}{\log 2}\right) e^{2\pi i t}. \quad (2.19)
\]
In the above, $\Gamma(\cdot)$ is the Gamma function, $\zeta(\cdot)$ is the Riemann zeta function, $\gamma = -\Gamma'(1)$ is the Euler constant, and $\gamma(1)$ is defined by the Laurent series $\zeta(s) = 1/(s - 1) + \gamma - \gamma(1)(s - 1) + O((s - 1)^2)$. The function $\Psi(x)$ is periodic with a very small amplitude, i.e., $|\Psi(x)| < 10^{-5}$. Moreover, for $\xi \to 1$ the function $\Phi(\xi)$ becomes

$$\Phi(\xi) \sim D_1 + (1 - \xi) \log(1 - \xi) - (1 - \xi)(1 + \log D_2), \quad \xi \to 1^-$$

where $D_1 = 1 + \log(K_0^*)$ and $D_2 = K_1^* K_0^*/e$ with

$$K_0^* = \prod_{t=1}^{\infty} \left(1 - 2^{-2t+1}\right)^{2^{-t}} = .68321974 \ldots$$

$$K_1^* = \prod_{t=1}^{\infty} \prod_{m=1}^{\infty} \left(1 - 2^{-2t+1+2}\right)^{-1} \left[1 - 2^{-2t+2m+1}\right]^{2^{-m}} = 1.2596283 \ldots$$

(iii) Left-Tail Regime: $k, n \to \infty$ with $2^k - n = M = O(1)$

$$h_n^k \sim \frac{\sqrt{2\pi}}{M!} D_2^M n^{M+1/2} e^{-D_1 n}$$

where $D_1$ and $D_2$ are defined above.

The expressions for $h_n^k$ in parts (i) and (iii) are completely determined. However, the expression in part (ii) involves the function $\Phi(\xi)$. We have not been able to determine this function analytically, except for its behaviors as $\xi$ approaches 0 or 1. In Section 5, we discuss the numerical computation of $\Phi(\cdot)$ and sketch this function in Figure 5. The behavior of $\Phi(\xi)$ as $\xi \to 1^-$ implies the asymptotic matching of parts (ii) and (iii), while the behavior as $\xi \to 0^+$ implies the matching of (i) and (ii). As $\xi \to 0$, this behavior involves the periodic function $\varphi(x)$, which satisfies $\varphi(x+1) = \varphi(x)$. In part (ii) we give two different representations for $\varphi(x)$; the latter (which involves $\Psi(x)$) is a Fourier series. In Appendix A we show the equivalence of these two representations.

Since $\Phi(\xi) > 0$, we see that in (ii) and (iii), the distribution is exponentially small in $n$, while in (i), $1 - h_n^k$ is super-exponentially small (the dominant term in $1 - h_n^k$ is $2^{-n^{3/2}}$). Thus, (i) applies in the right tail of the distribution while (ii) and (iii) apply in the left tail. We wish to compute the range of $k$ where $h_n^k$ undergoes the transition from $h_n^k \approx 0$ to $h_n^k \approx 1$, as $n \to \infty$. This must be in the asymptotic matching region between (i) and (ii). We shall show in Section 4 that $C_j$, defined in Theorem 3(i), becomes as $j \to \infty$

$$C_j \sim \frac{\pi^{5/2}}{2} e^{\varphi(\alpha)} \exp \left( -\frac{1}{2 \log 2} \right), \quad (2.20)$$
where \( \alpha = (\log_2 j) \) is the fractional part. With (2.20), we can verify the matching between parts (i) and (ii), and the limiting form of (ii) as \( \xi \to 0^+ \) is

\[
\mathcal{H}_n \sim \exp \left( -\frac{\rho_0}{2} e^{\alpha n} \exp \left( -\frac{\log 2}{2} \left( \left( k + \frac{3}{2} - \log_2 n \right)^2 - 2 \log_2 n - \frac{9}{4} \right) \right) \right)
\]

\[
= \exp \left( -\frac{\rho_0}{2} \exp (\log_2 n) \exp \left( -\log_2 (k + 1.5 - \log_2 n)^2 \right) \right)
\]

(2.21)

\[
= \exp \left( -\rho_0 \cdot \exp \left( -\frac{\log 2}{2} (k + 1.5 - \log_2 n)^2 + \theta + \Psi(\log_2 n) \right) \right)
\]

(2.22)

where \( \rho_0 \) is defined in Theorem 3(i) and

\[
\theta = \frac{1}{\log 2} \left( \frac{\gamma^2}{2} + \gamma(1) - \frac{\pi^2}{12} \right) + \frac{\log 2}{24} = -1.022401 \ldots
\]

while \( |\Psi(\log_2 n)| \leq 10^{-5} \). We have written (2.21) in terms of \( k \) and \( n \), recalling that \( \xi = n 2^{-k} \). We also have used \( \sqrt{1 + 2\xi \Phi'(\xi) + \xi^2 \Phi''(\xi)} \sim 1 \) as \( \xi \to 0 \).

We now set, for an integer \( \ell \),

\[
k_\ell = \left\lfloor \log_2 n + \sqrt{2 \log_2 n - \frac{3}{2}} + \ell \right\rfloor \tag{2.23}
\]

\[
= \log_2 n + \sqrt{2 \log_2 n - \frac{3}{2}} + \ell - \beta_n,
\]

where

\[
\beta_n = \left( \log_2 n + \sqrt{2 \log_2 n - \frac{3}{2}} \right) \in [0, 1). \tag{2.24}
\]

In terms of \( \ell \) and \( \beta_n \), (2.22) becomes

\[
\Pr \{ \mathcal{H}_n \leq \lfloor \log_2 n + \sqrt{2 \log_2 n - 1.5} + \ell \rfloor \} \sim \exp \left( -\rho_0 e^{\theta + \Psi(\log_2 n) 2 - (\ell - \beta_n)^2 / 2 - (\ell - \beta_n) \sqrt{2 \log_2 n}} \right).
\]

(2.25)

For \( 0 < \beta_n < 1 \) and \( n \to \infty \) the above is small for \( \ell \leq 0 \), and it is close to one for \( \ell \geq 1 \). This shows that asymptotically, as \( n \to \infty \), all the mass accumulates when \( k = k_1 \) is given by (2.23) with \( \ell = 1 \). Now suppose \( \beta_i = 0 \) for some \( n \), or more generally that we can find a sequence \( n_i \) such that \( n_i \to \infty \) as \( i \to \infty \) but \( \sqrt{2 \log_2 n_i} \left( \log_2 n_i + \sqrt{2 \log_2 n_i - \frac{3}{2}} \right) \) remains bounded. Then, the expression in (2.25) would be \( O(1) \) for \( \ell = 0 \) (since \( \beta_n \sqrt{2 \log_2 n} = O(1) \)). For \( \ell = 1 \), (2.25) would then be asymptotically close to 1. Thus, now the mass would accumulate at two points, namely, \( k_0 = k_1 - 1 \) and \( k_1 \). Finally, if \( \beta_n = 1 - o(1) \) such that \( (1 - \beta_n) \sqrt{2 \log_2 n} = O(1) \), then the probability mass is concentrated on \( k_1 \) and \( k_1 + 1 \).

In order to verify the latter assertions, we must either show that \( \beta_n = 0 \) for an integer \( n \) or that there is a subsequence \( n_i \) such that \( \sqrt{2 \log_2 n_i} \beta_n = O(1) \). The former is false,
Figure 3: The function $R(n) = \sqrt{2 \log_2 n} (\log_2 n + \sqrt{2 \log_2 n} - \frac{3}{2})$ versus $n.$

while the latter is true. To prove that $\beta_n = 0$ is impossible for integer $n$, let us assume the contrary. If there exists an integer $N$ such that

$$\log_2 n + \sqrt{2 \log_2 n} - \frac{3}{2} = N,$$

then

$$n = 2^{N+6/2-\sqrt{4+2N}}.$$

But this is impossible since this would require that $4 + 2N$ is odd.

To see that there exists a subsequence such that $R(n_i) = \beta_n; \sqrt{2 \log_2 n_i} = O(1)$, we first refer to Figure 3 which indicates that the function $R(n)$ fluctuates from zero to $\sqrt{2 \log_2 n}$.

In Section 5, we show that if $n_i = [2^{i+6/2-\sqrt{2i+4}^2} + 1$, then $R(n_i) \to 0$ as $i \to \infty$. Note that this subsequence corresponds to the minima of $R(n)$ in Figure 3.

We summarize our findings in the following corollary.

**Corollary 2** The asymptotic distribution of PATRICIA height is concentrated among the three points $k_1 - 1$, $k_1$ and $k_1 + 1$ where $k_1 = \lfloor \log_2 n + \sqrt{2 \log_2 n - \frac{3}{2}} \rfloor + 1$, that is,

$$\Pr\{H_n^P = k_1 - 1 \text{ or } k_1 \text{ or } k_1 + 1\} = 1 - o(1)$$

as $n \to \infty$. More precisely: (i) there are subsequences $n_i$ for which $\Pr\{H_{n_i}^P = k_1\} = 1 - o(1)$
Figure 4: Asymptotic distributions for the height and their corresponding lower and upper bound for PATRICIA tries with $n = 6.2 \cdot 10^8$ (one-point distribution) and $n = 5 \cdot 10^8$ (two-points distribution).

provided that

$$R(n) = \sqrt{2 \log_2 n \left( \log_2 n + \sqrt{2 \log_2 n - \frac{3}{2}} \right)} \to \infty$$

as $i \to \infty$; (ii) there are subsequences $n_i$ for which $\Pr\{H_{n_i}^P = k_1 - 1 \text{ or } k_1\} = 1 - o(1)$ provided that $R(n_i) = O(1)$; (iii) finally, there are subsequences $n_i$ for which $\Pr\{H_{n_i}^P = k_1 \text{ or } k_1 + 1\} = 1 - o(1)$ provided that $\sqrt{2 \log_2 n_i} - R(n_i) = O(1)$.

As in the case of $b$-tries, we can study the asymptotic distribution of $\Pr\{H_n^P \leq \log_2 n + \sqrt{2 \log_2 n - 1.5 + x}\}$ for $x$ real. We plot the asymptotic distribution of $H_n^P$ in Figure 4 and its lower and upper envelopes. The figure illustrates the concentration at one point or two points (cf. also Table 4 and Table 5). We also observe that the asymptotic distribution for the PATRICIA height qualitatively resembles the asymptotic distribution for $b$-tries with large $b$ (cf. Figure 2 and Figure 4).

To summarize the PATRICIA results, we have obtained an explicit formula (i.e., (2.21) or (2.25)) that can be used to compute the distribution $h_n^k$ where there is appreciable mass, while Theorem 3 also describes in detail the left and right tails of the distribution. In Section 5 we give detailed numerical comparisons between our asymptotic results and the exact values of $h_n^k$, obtained by numerically iterating the recursion (2.4).
3 Asymptotics of Tries and b-Tries

We analyze (2.2) and (2.3) for $k$ and/or $n \to \infty$ by two independent approaches. First, we solve the recurrence exactly and then evaluate the result asymptotically. Second, we try to obtain asymptotic information using only the recurrence (2.2). The latter method will form the basis for analyzing the PATRICIA model.

3.1 Transform Method

To solve (2.2), let us define $H^k(z) = \sum_{n \geq 0} \frac{\bar{h}_n z^n}{n!}$. Then, (2.2) implies that

$$H^k(z) = \left(H^0(z2^{-k})\right)^{2^k}$$

with $H^0(z) = 1 + z + \cdots + z^b/b!$. By Cauchy's formula, we obtain

$$\bar{h}_n = \frac{n!}{2\pi i} \oint \left(1 + z2^{-k} + \frac{z^2}{2!}2^{-2k} + \cdots + \frac{z^b}{b!}2^{-bk}\right)^{2^k} z^{-n-1}dz$$

(3.1)

where the contour integral is a loop around the origin in the $z$-plane. When $b = 1$, (3.1) reduces to (2.6) and then the integral may be explicitly evaluated.

Let us define

$$F(t) \equiv 1 + t + \frac{t^2}{2!} + \cdots + \frac{t^b}{b!} = e^t \int_t^\infty e^{-w} \frac{w^b}{b!} dw$$

(3.2)

and we note that $F(t) - F'(t) = \frac{t^b}{b!}$. From (3.2) it follows that

$$\log[F(t)] - t = \log \left(1 - \int_0^t e^{-w} \frac{w^b}{b!} dw\right) = -\frac{t^{b+1}}{(b+1)!} + O(t^{b+2}), \quad t \to 0. \quad (3.3)$$

We first consider the limit $n, k \to \infty$, with $n2^{-k} \equiv \xi$ fixed and $0 < \xi < b$. Scaling $z = n\omega$, (3.1) becomes

$$\bar{h}_n = \frac{n^!}{n^! 2\pi i} \oint e^{n f(\omega; \xi)} \frac{1}{\omega} d\omega$$

(3.4)

where

$$f(\omega; \xi) = -\log \omega + \frac{1}{\xi} \log \left(1 + \omega \xi + \frac{\omega^2}{2!} \xi^2 + \cdots + \frac{\omega^b}{b!} \xi^b\right). \quad (3.5)$$

We evaluate (3.4) by the saddle point method (cf. [3, 22]). We can easily show that the equation $\frac{d}{dw} f(\omega; \xi) = 0$, i.e.,

$$-\frac{1}{\omega} + 1 + \omega \xi + \cdots + (\omega \xi)^{b-1}/(b-1)! + 1 + \omega \xi + \cdots + (\omega \xi)^{b}/b! = 0, \quad (3.6)$$

15
has a unique solution on the real $\omega$-axis. We call it $\omega_0 = \omega_0(\xi; b)$. It satisfies

$$
\omega_0 \sim 1 - \frac{\xi^b}{b!} \to 1 \quad \xi \to 0,
$$

$$
\omega_0 \sim \frac{1}{b - \xi} \to \infty \quad \xi \to b.
$$

Using Stirling's formula to simplify $n!$ in (3.4) and evaluating the integral by the standard saddle point method (cf. [3]) yields

$$
\tilde{h}_n^k \sim \sqrt{2\pi n} e^{-n} \left[ 1 - \frac{1}{2\pi \omega_0} \sqrt{\frac{2\pi}{f''(\omega_0; \xi)n}} e^{nf(\omega_0; \xi)} \right].
$$

(3.7)

Also, by (3.4) and (3.5) we can show that

$$
f''(\omega_0; \xi) = \frac{1}{\omega_0^2} + \left(\frac{1}{\omega_0} - 1\right)(\frac{\xi - b}{\omega_0}) - \xi \left(1 - \frac{1}{\omega_0}\right)^2 = \frac{1}{\omega_0^2} \left[ 1 + (\xi - b)(\omega_0 - 1) \right] > 0
$$

(3.8)

with which (3.7) becomes the same as Theorem 2, part (ii).

Next we consider the limit $n = O(1)$ with $k \to \infty$. In view of (2.3) we have $\tilde{h}_n^k = 1$ for $0 \leq n \leq b$ and any $k \geq 0$. We return to (3.1) and note that for $k \to \infty$

$$
H^k(z) = \exp \left( 2^k \log[F(z^{2-k})] \right)
$$

(3.9)

$$
= \exp \left( 2^k \left[ \log \left( \int e^{w^b/\tau} dw \right) \right] \right)
$$

$$
= \exp \left( 2^{k+1} \left[ (b+1)!2^{-k} + O(2^{k+2}\tau^{-k+1}) \right] \right)
$$

where we have used (3.3). Using the last expression of (3.9) in (3.1) and evaluating explicitly the resulting integral(s) leads to Theorem 2, part (i).

Finally, we consider the limit $k, n \to \infty$ with $2^k - n$ fixed. Setting $2^k - n = j$ and scaling $z = (n+j)\tau/b = 2^k\tau/b$ in (3.1) leads to

$$
\tilde{h}_n^k \sim n! \left( \frac{n+j}{b} \right)^{-n} \frac{1}{2\pi i} \int \tau^{j-1} \left( \frac{1}{\tau^b} + \frac{1}{\tau^{b-1}} + \cdots + \frac{1}{(b-1)!} \right) \frac{1}{\tau} e^{n-j} d\tau.
$$

(3.10)

We furthermore expand the integrand for $\tau$ large (to be precise, for $\tau = O(n)$) to get

$$
\tilde{h}_n^k \sim n! \left( \frac{n+j}{b} \right)^{-n} \frac{1}{2\pi i} \int \tau^{j-1} \exp \left( \frac{n+j}{b} \log \left[ \frac{1}{b!} \left( \frac{1}{(b-1)!} \tau + \cdots + \frac{1}{(b-1)!} \right) \right] + O(\tau^{-2}) \right) d\tau
$$

$$
\sim n! \left( \frac{n+j}{b} \right)^{-n} \exp \left( -\frac{n+j}{b} \log (bl) \right) \frac{1}{2\pi i} \int \tau^{j-1} \exp \left( \frac{n+j}{\tau} \right) d\tau
$$

$$
\sim n! \left( \frac{n+j}{b} \right)^n \exp \left( -\frac{n+j}{b} \log (bl) \right) \frac{n^j}{j!}.
$$

(3.11)
Theorem 2 part (iii) follows from (3.11) and Stirling’s formula. In summary, to obtain (ii) from (3.1) the appropriate scaling was \( z = O(n) \), the scaling for (iii) was \( z = O(n^2) \).

3.2 Direct Method

What part of Theorem 2 can be obtained directly from the recurrence (2.2)? We first consider the limit \( n = O(1) \), \( k \to \infty \), and gradually increase the relative size of \( n \) to \( k \). We note that (2.3) implies that \( \tilde{h}^k_n = 1 \) for \( 0 \leq n \leq b \), for all \( k \geq 0 \), and we set \( \tilde{h}^k_n = 1 - G^k_n \). Then \( G^k_n \) satisfies

\[
G^k_{n+1} = 2 \left( \frac{1}{2} \right)^n \sum_{i=0}^{n} \binom{n}{i} G^k_{n-i} - \left( \frac{1}{2} \right)^n \sum_{i=0}^{n} \binom{n}{i} G^k_{i} G^k_{n-i}.
\]

(3.12)

Here we have used the fact that \( G^k_{n} = 0 \) for \( 0 \leq n \leq b \). For \( b+1 \leq n \leq 2b+1 \), the non-linear term in (3.12) vanishes and we are left with a linear recurrence. We can solve (3.12) exactly by first solving the linear problem for \( n \in [b+1, 2b+1] \). Then, using this solution to compute explicitly the nonlinear term in (3.12) for \( n \in [2b+2, 3b+2] \), we obtain an inhomogeneous linear problem. Thus we can solve the nonlinear problem (2.2) by solving a sequence of linear problems involving the \( n \)-intervals \([0,b], [b+1, 2b+1], [2b+2, 3b+2], \ldots \). However, the resulting expressions become complicated and may not be preferable to solving (2.2) by the transform method.

Our focus is on the asymptotics of \( G^k_n \) (hence \( \tilde{h}^k_n \)). For ranges of \( k,n \) where \( \tilde{h}^k_n \) is asymptotically close to 1, we can replace (3.12) by the asymptotic relation

\[
G^k_{n+1} \sim 2^{1-n} \sum_{i=0}^{n} \binom{n}{i} G^k_{n-i}.
\]

(3.13)

This is "constant-coefficient" in \( k \) so we seek solutions of the form \( G^k_n = a^k f(n) \). Replacing \( \sim \) by \( = \) in (3.13) we obtain

\[
a f(n) = 2^{1-n} \sum_{i=b+1}^{n} \binom{n}{i} f(i), \quad n \geq b + 1.
\]

(3.14)

Setting \( n = b + 1 \) determines \( a \) as \( a = 2^{-b} \) and then (3.14) becomes

\[
f(n) = 2^{b+1-n} \sum_{i=b+1}^{n} \binom{n}{i} f(i), \quad n \geq b + 1
\]

(3.15)

whose solution is
\[ f(n) = \binom{n}{b+1} = \frac{n!}{(b+1)!((n-b-1)!}. \]  

(3.16)

Here we have used \( f(b+1) = 1 \), which follows from (2.3). We have thus obtained Theorem 2(i) using only the recurrence relation. We also note that for \( b+1 \leq n \leq 2b+1 \), the exact answer is

\[ G_n^k = \frac{n!}{b!} \sum_{\ell=0}^{n-b-1} \frac{(-1)^\ell}{\ell!(n-b-1-\ell)!} \frac{2^{-k(\ell+b)}}{\ell + 1 + b} \]

which certainly satisfies \( G_n^k \sim 2^{-k}f(n) \) as \( k \to \infty \).

Next we consider \( k \) and \( n \) simultaneously large. We argue that the asymptotic relation \( \bar{h}_n^k \sim 1 - 2^{-kb}f(n) \) becomes invalid when \( n \) is sufficiently large so to make \( 2^{-kb}f(n) = O(1) \). From (3.16) we see that \( f(n) = O(n^{b+1}) \) as \( n \to \infty \) and thus \( 2^{-kb}f(n) = O(1) \) for \( n^{b+1}2^{-kb} = O(1) \), or \( k = (1 + 1/b) \log_2 n + O(1) \). Let us set

\[ k = \frac{b+1}{b} \log_2 n + \beta, \quad \beta = O(1) \]

and re-examine (2.2) on the \( \beta \)-scale. Here we expect that it will no longer be permissible to neglect the non-linear term in (3.12) (and thus to linearize (2.2)).

We set

\[ \bar{h}_n^k = F(\beta; n) = F(k - (1 + 1/b) \log_2 n; n) \]

and obtain from (2.2)

\[ F(\beta + 1; n) = \left( \frac{1}{2} \right)^n \sum_{i=0}^{n} \binom{n}{i} \cdot F\left( \beta - \left( 1 + \frac{1}{b} \right) \log_2 \left( \frac{i}{n} \right); n-i \right) \]

(3.19)

\[ \cdot F\left( \beta - \left( 1 + \frac{1}{b} \right) \log_2 \left( 1 - \frac{i}{n} \right) ; n-i \right). \]

Now we postulate that \( F(\beta; n) \) assumes an expansion of the form

\[ F(\beta; n) = F_0(\beta) + \frac{1}{n}F_1(\beta) + O(n^{-2}). \]

(3.20)

Then (3.19) yields the following equation for the leading term in (3.20)

\[ F_0(\beta + 1) = [F_0(\beta + 1 + 1/b)]^2. \]

(3.21)

Here we have used the fact that \( 2^{-n}\binom{n}{i} \sim \delta(i, n/2) \) as \( n \to \infty \) (cf. (3.27) for a more precise estimate), where \( \delta \) is the Kronecker delta function. Note that in order to obtain (3.21), the assumption (3.20) can be weakened to \( F(\beta; n) \to F_0(\beta) \) as \( n \to \infty \).

On the \( \beta \)-scale, we have thus approximated the non-linear problem (2.2) by another non-linear problem (3.21), which is much easier to solve. By using (3.20) and refining the
approximation $2^{-n} \binom{n}{i} \sim \delta(i, n/2)$, as in (3.27), we can obtain equations for the correction terms $F_\ell(\beta)$, $\ell \geq 1$, in (3.20). These will be linear equations and the linear operator that will appear will be the linearization of (3.21).

The general solution of (3.21) is

$$F_0(\beta) = \exp(-c2^{-b\beta}) = \exp(-ce^{-b\beta \log 2})$$

(3.22)

where $c$ is a constant. To determine this constant we require that (3.20), which applies for $\beta = O(1)$, asymptotically matches to $\bar{h}^k_n \sim 1 - 2^{-kb}f(n)$, which applies for $k \to \infty$ and $n = O(1)$. Here, we use the principle of matched asymptotic expansions (cf. [20]). Symbolically, we write this condition as

$$F_0(\beta)|_{\beta \to \infty} \sim 1 - 2^{-kb}f(n)|_{n \to \infty}.$$ 

(3.23)

The left side of (3.23) is, using (3.17), $1 - c2^{-b\beta} = 1 - cn^{b+1}2^{-bK}$. Since $f(n) \sim n^{b+1}/(b+1)!$ as $n \to \infty$, the matching is accomplished with

$$c = \frac{1}{(b+1)!}.$$ 

(3.24)

The matching condition was necessary to uniquely determine the expansion on the $\beta$-scale. We also comment that the most general solution to (3.21) is $F_0(\beta) = \exp(-c(\beta)2^{-b\beta})$ where $c(\beta)$ is a periodic function with period $1/b$ (i.e., $c(\beta + 1/b) = c(\beta)$). However, the matching condition implies that $c(\beta)$ is in fact a constant. In summary, our analysis of the $\beta$-scale yielded the asymptotic distribution in (2.8) and (2.13), with no recourse to the exact solution.

We next consider $k, n \to \infty$ with $\xi = n2^{-k}$ fixed. We set

$$\bar{h}^k_n = G(\xi; n) = G(n2^{-k}; n)$$

and note that $\bar{h}^{k-1}_i = G(2i\xi/n; i)$ and $\bar{h}^{k-1}_{n-i} = G(2i(1-i/n); n-i)$. From (2.2) we obtain

$$G(\xi; n) = \left(\frac{1}{2}\right)^n \sum_{i=0}^{n} \binom{n}{i} G\left(\frac{2i}{n}; i\right) G\left(2\left(1 - \frac{i}{n}\right); \xi; n-i\right).$$

(3.25)

Note that the initial condition (2.3) does not apply on the $\xi$-scale, since $k$ is assumed large.

We analyze (3.25) by the WKB method [2, 13]. That is, we seek an asymptotic solution of the form

$$G(\xi; n) \sim e^{n\phi(\xi)} \left[A(\xi) + \frac{1}{n}A^{(1)}(\xi) + \frac{1}{n^2}A^{(2)}(\xi) + \cdots\right]$$

(3.26)
The ansatz (3.26) may be viewed as a generalized saddle point approximation. By symmetry, the major contribution to the sum will come from \( i \approx n/2 \). We also note that Stirling's formula yields, for \( i = xn \) \((0 < x < 1)\),

\[
2^{-n} \binom{n}{i} = \frac{e^{i f_0(x)}}{\sqrt{2\pi n}} \frac{1}{\sqrt{x(1-x)}} \left[ 1 + \frac{1}{12n} \left( \frac{1}{x} - \frac{1}{1-x} \right) + O(n^{-2}) \right],
\]

where \( f_0(x) = -\log 2-x \log x-(1-x) \log(1-x) \). For \( x = 1/2+y/\sqrt{n} \) (i.e., \( i = n/2 + O(\sqrt{n}) \)), (3.27) simplifies to the Gaussian form

\[
2^{-n} \binom{n}{n/2+y\sqrt{n}} = \sqrt{\frac{2}{\pi n}} e^{-2y^2} \left[ 1 + \frac{1}{n} \left( -\frac{1}{4} + 2y^2 - \frac{4}{3}y^4 \right) + O(n^{-2}) \right].
\]

Using (3.28) and (3.26) in (3.25) and retaining only leading order terms, we are led to

\[
A_1(\xi) e^{n\phi(\xi)} \sim \sum_{i=0}^{n} \sqrt{\frac{2}{\pi n}} e^{-2y^2} A \left( \frac{2i}{n} \xi \right) A \left( \frac{2(n-i)}{n} \xi \right) \exp \left[ i\phi \left( \frac{2i}{n} \xi \right) + (n-i)\phi \left( 2 \left( 1 - \frac{i}{n} \right) \xi \right) \right].
\]

Now we set \( \psi(x) = x\phi(2x\xi) + (1-x)\phi(2(1-x)\xi) \) and expand this function about \( x = 1/2 \). We have \( \psi(1/2) = \phi(\xi) \), \( \psi(1/2) = 0 \) and \( \psi''(1/2) = 8\xi\phi'(\xi) + 4\xi^2\phi''(\xi) \). Approximating the sum in (3.29) by an integral and using the Laplace method to evaluate the integral for \( n \rightarrow \infty \), we obtain

\[
e^{n\phi(\xi)} A_1(\xi) = \sqrt{\frac{2}{\pi}} e^{n\phi(\xi)} A_2(\xi) \int_{-\infty}^{\infty} \exp \left( -2y^2 + y^2 (4\xi\phi'(\xi) + 2\xi^2\phi''(\xi)) \right) dy.
\]

In (3.30) the exponential factors \( e^{n\phi} \) cancel and we have

\[
1 = \sqrt{\frac{2}{\pi}} A_1(\xi) \sqrt{\frac{\pi}{2 - 4\xi\phi'(\xi) - 2\xi^2\phi''(\xi)}}.
\]

We have thus determined the function \( A_1(\xi) \) in terms of \( \phi(\xi) \), though we have not determined the latter. By continuing the expansion of (3.25) (using (3.26) and (3.28)) to higher orders we can express \( A^{(1)} \) in terms of \( A \) and \( \phi \), then \( A^{(2)} \) in terms of \( A^{(1)} \), \( A \) and \( \phi \), etc. Thus, the asymptotic series (3.26) is known, up to the function \( \phi(\xi) \). It does not seem to be possible to determine \( \phi \) using only the recursion (2.2). This function is apparently very sensitive to the initial condition(s) (2.3). By comparing Theorems 1 and 2, we see that \( \phi \) depends on the parameter \( b \), which enters the problem only through the initial condition (2.3).

We verify that \( A_1(\xi) = \sqrt{1 - 2\xi\phi'(\xi) - \xi^2\phi''(\xi)} \) is consistent with Theorem 2(ii). By differentiating \( \phi(\xi; b) \) in Theorem 2(ii) we obtain

\[
(\xi\phi)' = -1 - \log \omega_0 + \frac{b}{\xi} + \left( b - \xi - \frac{1}{\omega_0 - 1} \right) \frac{\omega_0'}{\omega_0}.
\]

\[\text{20}\]
By differentiating equation (3.6) we obtain, after some simplification,

$$\omega'_0 = \left(\frac{b}{\xi} - 1\right)(\omega_0 - 1)\omega_0 + \xi \omega'_0).$$

(3.33)

From (3.32) and (3.33) it follows that

$$\phi(\xi') = -\log \omega_0$$

(3.34)

so that

$$\phi(\xi') = 2\phi' + \xi \phi'' = -\omega'_0/\omega_0$$

and thus

$$1 - 2\xi \phi' - \xi^2 \phi'' = 1 + \xi \frac{\omega'_0}{\omega_0} = [1 + (\omega_0 - 1)(\xi - b)]^{-1},$$

which agrees with $A(\xi; b)$ of Theorem 2(ii) and verifies the relation between $A(\xi)$ and $\phi(\xi)$ in (3.31).

Finally, we comment that it is possible to obtain the (complete) result in Theorem 2(iii) using only the recurrence (2.2). We omit that analysis since it is completely analogous to that presented for PATRICIA tries (for $k - n = O(1)$) in Section 4. By using this result and asymptotic matching, we can infer the behavior of $\phi(\xi)$ as $\xi \to b$. By matching (3.26) to either the expansion on the $\beta$-scale, or the result for $k \to \infty$, $n = O(1)$, we can obtain the behavior of $\phi(\xi)$ as $\xi \to 0$.

To summarize, we have obtained a fair amount of asymptotic information for $h^k_n$ using only (2.2) and (2.3), along with some mild assumptions about the forms of the expansions in the various regions of $(n, k)$ space. The main deficiency in this method is the failure to determine the function $\phi$, except for its behaviors as $\xi \to 0$ and $\xi \to b$. However, for ranges of $k$ and $n$ where appreciable mass accumulates (i.e., where $h^k_n$ undergoes the transition from 0 to 1), we have been able, using asymptotic matching, to completely determine the asymptotic solution (cf. (3.22) and (3.24)).

4 Asymptotics of PATRICIA Tries

We analyze the PATRICIA model asymptotically for $n \to \infty$, and also give some exact results when $n$ is close to $k$ or $2^k$. Since we do not have an exact expression for $h^k_n$, we use the ideas developed in the previous section and analyze the recurrence (2.4). We first discuss the right tail asymptotics and then deal with the left tail approximation.

4.1 Right Tail Asymptotics

From the definition of the PATRICIA model, it follows that $\Pr\{h^k_n = k\} = h^k_n = 1$ for $k \geq n - 1$ and $h^k_n = 0$ for $n > 2^k$. It thus suffices to consider the range $k + 2 \leq n \leq 2^k$. 

21
We analyze here the "right tail" (we think of plotting $h_n^k$ as a function of $k$ for a fixed $n$), where $h_n^k$ is asymptotically close to 1. We set

$$H_n^k = 1 - h_n^k,$$  \hspace{1cm} (4.1)

and then $H_n^k \equiv L_{n-k}(n) = L_j(n)$ for $j = n - k$. Using (4.1) in (2.4) we obtain

$$(2 - 2^n)L_{j-1}(n) + 2 \sum_{i=1}^{j-2} \binom{n}{i} L_{j-i}(n-i) = \sum_{i=n-j+2}^{j-2} \binom{n}{i} L_{i+j-n}(i)L_{j-i}(n-i).$$  \hspace{1cm} (4.2)

The above holds for $n \geq j$ and we have used the fact that $L_1(n) = L_2(n) = L_1(n) = L_2(n) = \cdots = 0$. The boundary condition (2.5) implies that $H_n^0 = 0$ for $n = 0, 1$ and $H_n^0 = 1$ for $n \geq 2$. Hence, the boundary condition for $L_j(n)$ is

$$L_n(n) = 0, \quad n = 0, 1;$$  \hspace{1cm} (4.3)

$$L_n(n) = 1, \quad n \geq 2.$$  \hspace{1cm} (4.4)

We first compute $L_j(n)$ exactly for $j = 2, 3$ and 4. We set $j = 3$ in (4.2) and obtain

$$(2 - 2^n)L_3(n) + 2nL_3(n - 1) = 0, \quad n \geq 3.$$  \hspace{1cm} (4.5)

Since $L_2(2) = 1$, we can easily solve this linear recurrence and thus obtain

$$L_2(n) = n!2^{-n^2/2}2^{n/2} \prod_{m=3}^{n} \left( \frac{1}{1 - 2^{1-m}} \right), \quad n \geq 2.$$  \hspace{1cm} (4.6)

If the upper limit in any product exceeds the lower limit, we define the product to be 1.

Setting $j = 4$ in (4.2) we see that the sum in the right-hand side is void for $n \geq 5$, which yields

$$(2 - 2^n)L_4(n) + 2nL_3(n - 1) + 2 \binom{n}{2} L_2(n - 2) = 0, \quad n \geq 5$$  \hspace{1cm} (4.7)

and when $n = j = 4$ we obtain (using $L_2(2) = L_3(3) = 1$)

$$(2 - 2^4)L_3(4) + 20 = 6.$$  \hspace{1cm} \text{(4.7)}

It follows that $L_3(4) = 1$ and then (4.7) is readily solved, using (4.6), to give

$$L_3(n) = n!2^{-n^2/2}2^{3n/2} \left( \frac{1}{4} - \frac{n2^{-n}}{2} - \frac{1}{4}2^{-n} \right) \prod_{m=3}^{n} \left( \frac{1}{1 - 2^{1-m}} \right), \quad n \geq 4.$$  \hspace{1cm} (4.8)

Next we set $j = 5$ in (4.2) and note that the sum is void for $n \geq 7$. By examining (4.2) with $j = 5$ and $n = 5, 6$ we find that $L_4(5) = L_4(6) = 1$. For $n \geq 7$ we solve (4.2) and
obtain, after some calculation,
\[
L_j(n) = n!2^{-n^{7/2}}2^{5n/2} \cdot \left[ \frac{5}{288} + \frac{1 - n}{8} \left( \frac{1}{2} \right)^n + \left( \frac{n^2}{8} - \frac{n}{12} - \frac{1}{2} + \frac{1}{72} \right) \left( \frac{1}{4} \right)^n \right] 
\]
\[
\cdot \prod_{m=3}^{n} \left( \frac{1}{1 - 2^{-1-m}} \right), \quad n \geq 6
\]

Using this method we can solve for \( L_j(n) \) for any fixed \( j \), but the calculations become tedious as \( j \) become large. For \( n \geq 2j - 3 \) the recurrence (4.2) becomes linear:
\[
(2 - 2^n)L_{j-1}(n) + 2 \sum_{i=1}^{j-2} \binom{n}{i} L_{j-i}(n-i) = 0, \quad n \geq 2j - 3. \tag{4.10}
\]

We can (in theory) use (4.10) to express \( L_j(n) \) in terms of \( L_j(2j - 2) \), but the latter is unknown.

Now consider arbitrary \( j \) and the limit \( n \to \infty \). From (4.6), (4.8) and (4.9) it is easy to see that
\[
L_j(n) \sim \rho_0 K_j n!2^{-n^{7/2}}2^{j-3/2}n, \quad \text{fixed}, \quad \rho_0 = \prod_{\ell=2}^{\infty} (1 - 2^{-\ell})^{-1}.
\]

To compute the constants \( K_j \), we use (4.11) in (4.10) and obtain, to leading order in \( n \),
\[
K_j (1 - 2^{2-j}) = \sum_{\ell=1}^{j-2} \frac{K_{j-\ell}}{(\ell + 1)!} 2^{2+\ell}2^{(\ell+3)/2-((\ell+1)/2)}j
\]
for \( j \geq 3 \) with \( K_2 = 1 \), in view of (4.6). The recurrence (4.12) may be somewhat simplified by setting
\[
K_j = \frac{1}{j!} 2^{-(j+7/2)}2^{3j/2}C_j \tag{4.13}
\]
which leads to
\[
C_j = 4 \sum_{m=2}^{j} \binom{j}{m} \frac{2^{-j}}{j-m+1} C_m, \quad j \geq 2 \tag{4.14}
\]
with \( C_2 = 1 \).

To solve (4.14) we introduce the exponential generating function
\[
C(z) = \sum_{j=0}^{\infty} \frac{z^j}{j!} C_j \tag{4.15}
\]
to obtain
\[
C(z) = 4C \left( \frac{z}{2} \right) \frac{e^{z/2} - 1}{z/2}. \tag{4.16}
\]
Setting $\tilde{C}(z) = e^{-z}C(z)$ (so that $\tilde{C}(z)$ is the Poisson transform of $C_j$), (4.16) becomes

$$\tilde{C}(z) = \frac{8}{z}(1 - e^{-z/2})\tilde{C}\left(\frac{z}{2}\right)$$  \hspace{1cm} (4.17)$$

where $\tilde{C}(z) \sim z^2/2$ as $z \to 0$. The latter follows from $C_2 = 1$ and we define $C_0 = C_1 = 0$. Next we set $\tilde{C}(z) = z^2G(z)/2$ and replace $z$ by $2z$ in (4.17), which gives

$$G(2z) = G(z)\left(\frac{1 - e^{-z}}{z}\right)$$  \hspace{1cm} (4.18)$$

with $G(0) = 1$. Taking the logarithm of (4.18) with $F(z) = \log G(z)$, we obtain

$$F(2z) - F(z) = \log\left(\frac{1 - e^{-z}}{z}\right),$$  \hspace{1cm} (4.19)$$

with $F(0) = 0$.

Functional equations of the type (4.19) are often encountered in the analysis for algorithms and they are usually handled by the Mellin transform [11]. The interested reader can find more on Mellin transform in a recent survey [11]. The Mellin transform $F^*(s)$ of a real valued function $F(z)$ is defined as

$$F^*(s) = \int_0^\infty F(z)z^{s-1}dz = \mathcal{M}[F; s]$$

and its inverse is

$$F(z) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} z^{-s}F^*(s)ds = \mathcal{M}^{-1}[F^*; z]$$

where $c$ belongs to the so called fundamental strip where the Mellin transform is analytic (cf. [11]).

Taking the Mellin transform of (4.19) yields

$$F^*(s) = \frac{2^s}{1 - 2^s} \int_0^\infty z^{s-1}\log\left(\frac{1 - e^{-z}}{z}\right)dz$$  \hspace{1cm} (4.20)$$

for $\Re(s) \in (-1, 0)$. We now evaluate the inverse Mellin transform of $F^*(s)$ by two methods. First, we observe that

$$\mathcal{M}^{-1}\left[\frac{2^s}{1 - 2^s}; z\right] = \sum_{m=0}^{\infty} 2^{m+1}\delta(z - 2^{m+1})$$  \hspace{1cm} (4.21)$$

where $\delta(\cdot)$ is the Dirac delta function. The convolution theorem for Mellin transforms $f^*(s)$ and $g^*(s)$ of functions $f(z)$ and $g(z)$ is

$$\mathcal{M}^{-1}[f^*g^*; z] = \int_0^\infty f(\xi)g(z/\xi)\frac{d\xi}{\xi}.$$  \hspace{1cm} (4.22)$$
We apply (4.22) with \( f \) given by the integral in (4.20) and \( g = 2^s/(1 - 2^s) \). Using (4.21) then yields

\[
F(z) = \int_0^\infty \frac{1}{\xi} \log \left( \frac{1 - e^{-\xi}}{\xi} \right) \sum_{m=0}^\infty 2^{m+1} \delta \left( \frac{z}{\xi} - 2^{m+1} \right) d\xi
\]

(4.23)

\[
= \int_0^\infty \frac{1}{u} \log \left( \frac{1 - e^{-z/u}}{z/u} \right) \sum_{m=0}^\infty 2^{m+1} \delta(u - 2^{m+1}) du
\]

\[
= \sum_{m=0}^\infty \log \left( \frac{1 - \exp(-z2^{m-1})}{z2^{-m-1}} \right).
\]

With (4.23), we can easily compute \( C(z) = e^z z^2 e^{F(z)}/2 \) and then invert the generating function in (4.15) by Cauchy's theorem to get

\[
C_j = \frac{j!}{2\pi i} \oint \frac{z^{-j-1} z^2}{2} e^z \prod_{m=0}^\infty \left( \frac{1 - \exp(-z2^{-m-1})}{z2^{-m-1}} \right) dz,
\]

(4.24)

which establishes (2.16) of Theorem 3(i).

An alternate expression for \( C_j \) can be established by evaluating the integral in (4.20) in a different way. Integrating by parts, we find

\[
M \left[ \log \left( \frac{1 - e^{-z}}{z} \right) \right] = \int_0^\infty z^{s-1} \log \left( \frac{1 - e^{-z}}{z} \right) dz
\]

(4.25)

\[
= -\frac{1}{s} \int_0^\infty z^{s-1} \left( \frac{z}{e^z - 1} - 1 \right) dz,
\]

where the last integral converges for \(-1 < \Re(s) < 0\). We now use

\[
M \left( \frac{z}{e^z - 1} \right) = \Gamma(s+1)\zeta(s+1), \quad 0 < \Re(s) < \infty
\]

(4.26)

where \( \Gamma(\cdot) \) and \( \zeta(\cdot) \) are the Gamma and Riemann zeta functions, respectively. Shifting the fundamental strip to \( \Re(s) \in (-1,0) \) we obtain (cf. [11])

\[
M \left[ \frac{z}{e^z - 1} \right] = \Gamma(s+1)\zeta(s+1), \quad -1 < \Re(s) < 0.
\]

(4.27)

We use (4.27) to evaluate the right side of (4.25) and then (4.20) becomes

\[
F^*(s) = -\frac{\Gamma(s+1)\zeta(s+1)}{s(2^{-s} - 1)}
\]

for \( \Re(s) \in (-1,0) \). It follows that an alternate representation for the Poisson transform \( \tilde{C}(z) \) of \( C_j \) is

\[
\tilde{C}(z) = \frac{z^2}{2} \exp \left( \frac{1}{2\pi i} \int_{-\frac{1}{2}-i\infty}^{-\frac{1}{2}+i\infty} \frac{z^{-s} \Gamma(s+1)\zeta(s+1)}{s(1-2^{-s})} ds \right).
\]

(4.28)
In Appendix A we prove that (4.28) and (4.24) are equivalent.

In view of (4.11) we see that \(1 - h_n^k = 1 - h_n^{n-j}\) is exponentially small (roughly of order \(O(2^{-n^2/2})\)) for \(n \to \infty\) and \(j\) fixed. We next study the asymptotics of \(C_j\) (and hence \(K_j\) and \(L_j(n)\)) for \(j \to \infty\). If \(j\) is sufficiently large so as to make \(h_n^{n-j} = O(1)\), then this will give us a rough estimate of the range of \(k\) where \(h_n^k\) changes from \(h_n^k \approx 0\) to \(h_n^k \approx 1\).

To accomplish the above goal, we apply the "principle of matched asymptotics" and assume that (4.24) holds also for \(j \to \infty\). The Poisson transform \(\tilde{G}(z)\) of \(G_j\) is given in (4.28). Our goal is to "depoissonize" it, that is, to extract \(C_j\) for \(j \to \infty\) from the behavior of \(\tilde{G}(z)\) as \(z \to \infty\) in a cone around real axis. We expect that, under some mild growth assumptions of \(\tilde{G}(z)\), \(C_j \sim \tilde{C}(j)\) since the Poisson process is well concentrated around its mean \(z = j\). To be more precise, we appeal to recent depoisnization results of Jacquet and Sspankowski [17], applying the following.

**Theorem 4 (Jacquet and Sspankowski 1998)** Let \(g_n\) be a sequence whose Poisson transform is \(\tilde{G}(z) = e^{-z} \sum_{n \geq 0} g_n z^n\) where \(z\) is complex. Consider a linear cone \(S_\theta = \{z : \arg(z) \leq \theta, |\theta| < \pi/2\}\). Assume for \(z \to \infty\) that:

1. For \(z \in S_\theta\)
   \[|\tilde{G}(z)| \leq A \exp(B|z|^\beta)\]  \hspace{1cm} (4.29)

where \(0 \leq \beta < 1/2\), and \(A, B > 0\) are constants.

2. For \(z \notin S_\theta\)
   \[|\tilde{G}(z)| \leq A_1 \exp(\omega|z|)\]  \hspace{1cm} (4.30)

for \(\omega < 1\) and \(A_1 > 0\). Then

\[g_n = \tilde{G}(n) + O\left(n^{-(1-2\beta)} \exp(Bn^\beta)\right)\]  \hspace{1cm} (4.31)

for \(n \to \infty\).

We apply Theorem 4 to find \(C_j\) for large \(j\). We present two alternate derivations. Let us start with (4.28). Using (4.14) and the method of "increasing domains" proposed in [17] we can easily show that condition (4.30) of Theorem 4 is satisfied. To verify condition (4.29), let us evaluate \(\tilde{G}(z)\) asymptotically for \(z \to \infty\) in the cone \(S_\theta\). We first compute asymptotically the exponent of (4.28), that is,

\[F(z) = \frac{1}{2\pi i} \int_{-\frac{1}{2} - i\infty}^{\frac{1}{2} + i\infty} \frac{z^{-s} \Gamma(s+1) \zeta(s+1)}{s(1-2^{-s})} ds\]  \hspace{1cm} (4.32)
as $z \to \infty$. We note that the integrand in (4.32) has a triple pole at $s = 0$ and simple poles along the imaginary axis, at $s = -2\pi ik/\log(2) = s_k$, $k = \pm 1, \pm 2, \ldots$. A lengthy computation shows that the negative of the residue at $s = 0$ is

$$-\frac{\log^2(z)}{2 \log 2} + \frac{1}{2} \log(z) - \frac{1}{12} \log(2) + \frac{1}{\log 2} \left( \gamma(1) + \frac{1}{2} \gamma^2 - \frac{\pi^2}{12} \right),$$

and the residues at $s = s_k$ are

$$\text{Res}_{s=s_k}[F(z)] = -\frac{1}{2\pi i k} \Gamma(1 + s_k) \zeta(1 + s_k).$$

It follows that

$$\exp[F(z)] \sim \sqrt{2} 2^{-1/12} \exp \left( \frac{\gamma(1) + \gamma^2/2 - \pi^2/12}{\log 2} \right) \exp \left[ -\frac{1}{2} \log^2(z) + \Psi(\log_2 z) \right]$$

where

$$\Psi(\log_2 z) = \sum_{\ell=-\infty}^{\infty} \frac{1}{2\pi i \ell} \Gamma \left( 1 - \frac{2\pi i \ell}{\log 2} \right) \zeta \left( 1 - \frac{2\pi i \ell}{\log 2} \right) e^{2\pi i \ell \log_2 z}.$$

Thus, $\tilde{C}(z) = O(z^{5/2} \exp(\log^2 z))$ in the cone $S_0$ (by analytic continuation). In view of this, we can apply Theorem 4 to get

$$C_j \sim \frac{1}{2} j^{5/2 - 1/12} \exp \left( \frac{\gamma(1) + \gamma^2/2 - \pi^2/12}{\log 2} \right) \exp \left[ -\frac{1}{2} \log^2(j) + \Psi(\alpha) \right] \quad (4.33)$$

where $\alpha = \{\log_2 j\}$ with $\{\cdot\}$ denoting the fractional part.

An alternate representation for the above can be obtained by using (4.24). Following the footsteps of the depoissonization verification above, we can show that (I) and (O) of Theorem 4 hold, and hence

$$C_j \sim \frac{j^3}{2} \exp \left[ \sum_{m=1}^{\infty} \log \left( \frac{1 - \exp(-j2^{-m})}{j2^{-m}} \right) \right], \quad j \to \infty. \quad (4.34)$$

We can further simplify the above by setting

$$m = \log_2(j) - \alpha + \ell, \quad (4.35)$$

where $\alpha = \{\log_2 j\}$. Using (4.35) in (4.34) we have

$$\sum_{\ell=1-\lceil\log_2(j)\rceil}^{\infty} \log \left( \frac{1 - \exp(-2^{\alpha-\ell})}{2^{\alpha-\ell}} \right) = \sum_{\ell=0}^{\infty} \log \left( \frac{1 - \exp(-2^{\alpha-\ell})}{2^{\alpha-\ell}} \right) \quad (4.36)$$

$$+ \sum_{\ell=1}^{\lceil\log_2(j)\rceil-1} \log[1 - \exp(-2^{\alpha+\ell})]$$

$$- \sum_{N=1}^{N+\alpha} (N + \alpha) \log(2) + O(\rho^j)$$

27
for some $\rho < 1$. The last sum in the right-hand side of (4.36) is

$$
-\alpha \log 2([\log_2(j)] - 1) - \frac{1}{2} \log 2([\log_2(j)] - 1) [\log_2(j)]
$$

(4.37)

$$
= -\frac{1}{2} \log^2(j) + \frac{1}{2} \log(j) + \frac{1}{2} \log 2\alpha(\alpha + 1).
$$

Using (4.36) and (4.37) in (4.34) we find that

$$
C_j \sim \frac{1}{2} j^{5/2} \exp \left( -\frac{\log^2(j)}{2 \log 2} \right) e^{\varphi(\alpha)}, \quad \alpha = \{\log_2(j)\}
$$

(4.38)

where $\varphi(\alpha + 1) = \varphi(\alpha)$ is given by

$$
\varphi(\alpha) = \frac{1}{2} \alpha(\alpha + 1) \log 2 + \sum_{t=0}^{\infty} \log \left( \frac{1 - \exp(-2^{\alpha+t})}{2^{\alpha+t}} \right) + \sum_{t=1}^{\infty} \log \left[ 1 - \exp(-2^{\alpha+t}) \right].
$$

(4.39)

By comparing (4.38) and (4.33), it follows that $\varphi(\alpha)$ and $\psi(\alpha)$ are related as in Theorem 3(ii).

We are now in position to continue our analysis and extend the validity of (4.11) to $j \to \infty$. Using (4.38) and (4.13), we see that the right side of (4.11) becomes for $j \to \infty$

$$
L_j(n) = \frac{n!}{j!} 2^{-(j-n)^2/2} 2^{3(j-n)^2} j^{5/2} \exp \left( -\frac{\log^2(j)}{2 \log 2} \right) e^{\varphi(\alpha)}.
$$

(4.40)

We set $j = n - k$ and find for what range of $k$ is (4.40) $O(1)$ as $n \to \infty$. Taking the logarithm of (4.40), this condition is the same as

$$
n \log n - n + \frac{1}{2} \log n - j \log j + j + 2 \log j + \frac{3}{2} (\log 2)(j - n) - \frac{1}{2} (\log 2)(j - n)^2
$$

$$
- \frac{1}{2} \log 2 \log^2(j) = O(1).
$$

Using $j \log j = (n - k) \log(n - k) = (n - k) [\log n - k/n + O(k^2 n^{-2})]$, the above becomes

$$
k \log n - \frac{\log 2}{2} k^2 - \frac{3}{2} (\log 2) k - \frac{\log^2 n}{2 \log 2} + \frac{5}{2} \log n = O(1)
$$

and this implies that

$$
k = \log_2 n + \sqrt{2 \log_2 n - \frac{3}{2}} + o(1).
$$

Thus, we expect that if $k$ satisfies the above condition, then the asymptotic expression for $h^k_n$ in Theorem 3(i) breaks down, as the second term becomes comparable in magnitude to the first term. This completes our discussion of the right tail of $h^k_n$, where $h^k_n$ is asymptotically close to 1.
4.2 Left Tail Asymptotics

We next consider the left tail. Here \( h_n^k \) will be asymptotically small. We first consider the limit where \( k, n \to \infty \) with \( 2^k - n = M = O(1) \) and \( M \geq 0 \). This corresponds to the "left-most" tail of the distribution, since \( h_n^k = 0 \) for \( 2^k < n \). We shall derive exact results for \( h_n^k \) if \( M = 0 \) or 1 and then establish part (iii) of Theorem 3, which applies for arbitrary \( M = O(1) \).

Let us define the exponential generating function \( H^k(z) \) of \( h_n^k \) as

\[
H^k(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!} h_n^k
\]

with which (2.4) becomes

\[
H^{k+1}(z) = 2H^{k+1} \left( \frac{z}{2} \right) + \left[ H^k \left( \frac{z}{2} \right) \right]^2 - 2H^k \left( \frac{z}{2} \right), \quad k \geq 0
\]

and \( H^0(z) = 1 + z \) by (2.5). We can simplify (4.42) by setting \( H^k(z) = 1 + \tilde{H}^k(z) \) to get

\[
\tilde{H}^{k+1}(z) - 2\tilde{H}^{k+1} \left( \frac{z}{2} \right) = \left[ H^k \left( \frac{z}{2} \right) \right]^2, \quad k \geq 0
\]

and \( \tilde{H}^0(z) = z \).

From our previous discussion, we can truncate the sum in (4.41) at \( n = 2^k \) so that \( H^k(z) \) (and thus \( \tilde{H}^k(z) \)) will be a polynomial of degree \( 2^k \). We will identify the two leading coefficients in this polynomial by writing

\[
\tilde{H}^k(z) = a(k)z^{2^k} + b(k)z^{2^k-1} + \cdots + z
\]

where

\[
a(k) = \frac{h_{2^k}^k}{(2^k)!}, \quad b(k) = \frac{h_{2^k-1}^k}{(2^k-1)!}.
\]

From (4.44) it follows that

\[
\left[ H^k \left( \frac{z}{2} \right) \right]^2 = \frac{a^2(k)}{2^{2^{k+1}}} z^{2^{k+1}} + \frac{4a(k)b(k)}{2^{2^{k+1}}} z^{2^{k+1}-1} + \cdots + \frac{z^2}{4}
\]

and

\[
\tilde{H}^{k+1}(z) - 2\tilde{H}^{k+1} \left( \frac{z}{2} \right) = \left( 1 - \frac{2}{2^{2^{k+1}}} \right) a(k+1)z^{2^{k+1}} + \left( 1 - \frac{4}{2^{2^{k+1}}} \right) b(k+1)z^{2^{k+1}-1} + O(z^{2^{k+1}-2}).
\]
By comparing (4.45) to (4.46) we obtain the recurrences

\[
\alpha(k + 1) \left[ 1 - 2 \left( \frac{1}{2} \right)^{2^{k+1}} \right] = \alpha^2(k) \frac{1}{2} 2^{2k+1},
\]

(4.47)

and

\[
b(k + 1) \left[ 1 - 4 \left( \frac{1}{2} \right)^{2^{k+1}} \right] = \frac{4}{2^{2k+1}} \alpha(k) b(k).
\]

(4.48)

Solving (4.47) subject to \(a(0) = 1\) yields

\[
a(k) = 2^{-k2^k} \prod_{\ell=1}^{k} \left( 1 - 2 \left( \frac{1}{2} \right)^{2^{\ell}} \right)^{-2^{k-\ell}}.
\]

(4.49)

It follows that

\[
h_{2^k}^k = (2^k)! 2^{-k2^k} \exp \left[ -2^k \sum_{\ell=1}^{k} 2^{-\ell} \log \left( 1 - 2 \left( \frac{1}{2} \right)^{2^{\ell}} \right) \right].
\]

(4.50)

Evaluating (4.50) as \(k \to \infty\) yields

\[
h_{2^k}^k \sim \sqrt{2\pi} \sqrt{k} \exp[2^k(-1 - \log K_0^*)] \quad k \to \infty,
\]

(4.51)

where

\[
K_0^* = \prod_{\ell=1}^{\infty} \left[ 1 - 2 \left( \frac{1}{2} \right)^{2^\ell} \right]^{2^{-\ell}}
\]

and the numerical value of \(K_0^*\) is given in Theorem 3.

Having computed \(a(k)\) we can easily solve the linear recurrence (4.48) for \(b(k)\), subject to \(b(1) = 1\). This yields

\[
b(k) = \prod_{\ell=1}^{k-1} \frac{4}{1 - 42^{-2^\ell} 2^{2^{\ell+1}}} \alpha(\ell).
\]

(4.52)

The expression (4.49) may be rewritten as

\[
a(k) = 2^{-k2^k} \left( \frac{1}{K_0^*} \right)^{2^k} \prod_{m=1}^{\infty} \left( 1 - 2 \left( \frac{1}{2} \right)^{2^{k+m}} \right)^{2^{-m}}
\]

with which (4.52) becomes

\[
b(k) = \prod_{\ell=1}^{k-1} \frac{4 \cdot 2^{-\ell2^\ell(K_0^*)^{-2^\ell}}}{1 - 4 \cdot 2^{2^{\ell+1}} 2^{2^{2\ell+1}}} \prod_{m=1}^{\infty} \left( 1 - 2 \left( \frac{1}{2} \right)^{2^{k+m}} \right)^{2^{-m}}
\]

(4.53)

\[
= \prod_{\ell=1}^{k-1} \frac{4 \cdot 2^{-(\ell+2)2^\ell}}{1 - 4 \cdot 2^{2^{\ell+1}}} \prod_{m=1}^{\ell} \left( \frac{1}{1 - 2 \cdot 2^{-2m}} \right)^{2^{\ell-m}}.
\]
As \( k \to \infty \) we obtain from (4.53)

\[
b(k) \sim 2^{2k}2^{-k^2} (K_0^*)^{2-2^k} K_1^*, \quad k \to \inftyan

(4.54)

where

\[
K_1^+ = \prod_{\ell=1}^{\infty} \prod_{m=1}^{\infty} \frac{(1 - 2 \cdot 2^{-2^\ell+m})2^{-m}} {1 - 4 \cdot 2^{-2^\ell+1}} = \prod_{N=2}^{\infty} \frac{(1 - 2 \cdot 2^{-2^N})1-2^{1-N}} {1 - 4 \cdot 4^{-2^N}}.
\]

Using (4.54) and approximating \((2^k - 1)!\) by Stirling's formula we are led to

\[
h_{2k-1}^k \sim \sqrt{2\pi} 2^k \sqrt{2^k K_1^* (K_0^*)^2} \exp[2^k (-1 - \log K_0^*)], \quad k \to \infty.
\]

(4.55)

We next consider (2.4) for values of \( n \) that are close to \( 2^k \). We set \( M = 2^k - n \), with

\[
h_n^k = W(2^k - n; n) = W(M; n)
\]

(4.56)

and note that

\[
h_{i-1}^k = W \left( \frac{M}{2} + \frac{n}{2} - i; i \right) \quad \text{and} \quad h_{n-i}^k = W \left( \frac{M}{2} - \frac{n}{2} + i; n - i \right)
\]

Replacing \( k \) by \( k - 1 \) in (2.4) and dropping the first term in the right-hand side (which is clearly negligible as \( n \to \infty \)) we obtain

\[
W(M; n) \sim \sum_{i=1}^{n-1} \binom{n}{i} 2^{-n} W \left( \frac{M + n}{2} - i; i \right) W \left( \frac{M - n}{2} + i; n - i \right)
\]

(4.57)

\[
= \sum_{i=(n-M)/2}^{(n+M)/2} \binom{n}{i} 2^{-n} W \left( \frac{M + n}{2} - i; i \right) W \left( \frac{M - n}{2} + i; n - i \right).
\]

Here we have used the fact that \( W(M; n) = 0 \) for \( M < 0 \) (since \( h_n^k = 0 \) for \( 2^k < n \)) to truncate the limits on the summation in (4.57). For \( i = n/2 + O(1) \) we have

\[
\binom{n}{i} 2^{-n} \sim \frac{\sqrt{2}}{\pi n}
\]

so that (4.57) can be rewritten as

\[
W(M; n) \sim \sqrt{\frac{2}{\pi n}} \sum_{\ell=0}^{M} W \left( \ell; \frac{n + M}{2} - \ell \right) W \left( \ell; \frac{M - n}{2} + \ell \right).
\]

(4.58)

Setting \( M = 0 \), (4.58) becomes

\[
W(0; n) \sim \sqrt{\frac{2}{\pi n}} \left[ W \left( 0; \frac{n}{2} \right) \right]^2
\]

31
that admits an asymptotic solution in the form
\[ W(0; n) \sim \sqrt{2\pi}ne^{-D_1n} \tag{4.59} \]
where \( D_1 \) is at this point undetermined. By considering \( M = 1 \) and proceeding inductively, we find that (4.58) admits an asymptotic solution
\[ W(M; n) \sim n^{M+1/2}A(M)e^{-D_1n} \tag{4.60} \]
where \( A(M) \) satisfies
\[ M \frac{A(M)}{A(M)} \sim TM A(f)A(M-f), \quad M \geq 0. \tag{4.61} \]
The solution to (4.61) is
\[ A(M) = \frac{1}{\sqrt{2\pi}}2^{-M} \sum_{\ell=0}^{M} A(\ell)A(M-\ell), \quad M \geq 0. \tag{4.62} \]
where \( D_2 \) is also undetermined. Combining (4.59) – (4.62) yields
\[ h_n^k \sim \frac{\sqrt{2\pi}}{M!} (D_2)^M n^{M+1/2}e^{-D_1n}, \quad M = 2^k - n = O(1). \tag{4.63} \]
It remains only to determine \( D_1 \) and \( D_2 \).

These constants are fixed by comparing (4.63) for \( M = 0 \) and \( M = 1 \) to (4.51) and (4.55). Setting \( M = 0 \) in (4.63) and noting that \( 2^k = n \), we see that (4.51) agrees with (4.63) provided that
\[ D_1 = 1 + \log(K_0^*). \tag{4.64} \]
Setting \( M = 1 \) in (4.63), noting that \( 2^k = n + 1 \) and comparing to (4.55) determines \( D_2 \) as
\[ D_2 = K_1^*K_0^*/e. \tag{4.65} \]
This completes the analysis of the scale \( M = O(1) \).

We have thus obtained the asymptotics of \( h_n^k \) for \( j = n-k = O(1) \) (the right tail) and for \( M = 2^k - n = O(1) \) (the left tail). However, these expansions do not asymptotically match, which indicates there must be at least one additional natural scale in the problem, as was the case for tries. The recurrences (2.2) and (2.4) differ only slightly. Furthermore, as \( n \to \infty \), the term \( 2^{1-n}h_n^{k+1} \) in (2.4) is exponentially small compared to the left side (\( = h_n^{k+1} \)) of the equation. The two boundary terms in the sum in (2.4) (i.e., \( i = 0 \) and \( i = n \)) are absent, but our analysis of the scale \( M = O(1) \) shows that they are asymptotically negligible (and in some cases \textit{exactly} equal to zero).
As in the analysis of tries, we consider the scale $n, k \to \infty$ with $\xi = n^{2^{-k}}$ fixed and $0 < \xi < 1$. We set $h_n^k = F(\xi; n) = F(n^{2^{-k}}; n)$ and assume an asymptotic solution of (2.4) in the WKB form

$$F(\xi; n) \sim e^{-n\Phi(\xi)} \left[ A(\xi) + \frac{1}{n} A^{(1)}(\xi) + \cdots \right].$$  \hspace{1cm} (4.66)

The calculation is essentially identical to that in Section 3, and we find that

$$A(\xi) = \sqrt{1 + 2\xi\Phi'(\xi) + \xi^2\Phi''(\xi)},$$  \hspace{1cm} (4.67)

which expresses $A$ in terms of $\Phi$. Once again it seems that $\Phi$ is very sensitive to the initial conditions(s) (2.5), and we cannot analytically determine this function. However, in Section 5 we discuss the numerical calculation of this function; the numerical results also provide partial justification of the "ansatz" (4.66).

We next show that (4.66) can asymptotically match to the expansions on the $j$ and $M$ scales. This will also yield the local behavior of $\Phi(\xi)$ as $\xi \to 0^+$ and $\xi \to 1^-$. The matching of the $M$ and $\xi$ scales requires that

$$\left. \frac{\sqrt{2\pi}}{M!} (D_2)^M e^{-D_1 n_{1/2} M} \right|_{M \to \infty} \sim A(\xi) e^{-n\Phi(\xi)} \right|_{\xi \to 1^-}.$$  \hspace{1cm} (4.68)

The left side of (4.68) is easily evaluated. We simply expand $M!$ by Stirling's formula and note that $n/M = \xi/(1 - \xi)$, which yields

$$\sqrt{\frac{\xi}{1 - \xi}} \exp \left( n \left[ \frac{M}{n} \log \left( \frac{n}{M} \right) + \frac{M}{n} \log(D_2) - D_1 \right] \right)$$

so that the matching condition implies that

$$\Phi(\xi) \sim D_1 + (1 - \xi) \log(1 - \xi) - (1 - \xi)(1 + \log(D_2)), \quad \xi \to 1$$  \hspace{1cm} (4.69)

and

$$A(\xi) \sim (1 - \xi)^{-1/2}, \quad \xi \to 1.$$  \hspace{1cm} (4.70)

Given (4.69), we can also use (4.67) to infer the behavior of $A(\xi)$ as $\xi \to 1$. We have $\Phi'(\xi) \sim -\log(1 - \xi)$ and $\Phi''(\xi) \sim 1/(1 - \xi)$ with which (4.67) implies (4.70). It follows that $\Phi$ is finite at $\xi = 1$ (with $\Phi(1) = D_1$), but its derive has a logarithmic singularity at $\xi = 1$.

Now consider the matching of the $j$ and $\xi$ scales. This requires that the large $j$ asymptotics of $h_n^k = 1 - L_j(n)$ agrees with the expansion of $A e^{-n\Phi}$ as $\xi \to 0^+$. We must have $A \to 1$ and $\Phi \to 0$ as $\xi \to 0^+$. The asymptotic matching region between the $j$ and $\xi$ scales
will be where the probability mass accumulates as $n \to \infty$. If we let $\xi \to 0^+$ in such a way that $n\Phi(\xi) \to 0$, then $A e^{-n\Phi} \sim 1 - n\Phi$ and the matching is satisfied provided that

$$n\Phi(\xi)|_{\xi \to 0^+} \sim L_j(n)|_{j \to \infty}. $$

We have already computed the right side of (4.71) (cf. (4.11), (4.13) and (4.38)). In (4.71) $L_j(n)$ is understood to be replaced by its large $n$ expansion (cf. (4.11)), which is then evaluated for $j \to \infty$. The matching condition thus becomes

$$n\Phi(\xi)|_{\xi \to 0^+} \sim \frac{n!}{j! 2^{-(n-j)/2} 2^j (j-n)/2} \left( \frac{1}{2} \rho_0 \varphi(\alpha) \right) j^{5/2} \exp \left[ \frac{-\log^2(j)}{2 \log 2} \right].$$

To evaluate (4.72) we recall that $k = n - j$ and $\log \xi = \log n - k \log 2$. For $n \to \infty$ with $k = o(n)$ we can replace $\alpha = \log_2(j) = \log_2(n - k)$ by $\log_2(n) + o(1)$, since $\varphi(\alpha)$ is periodic with period one. Also, by periodicity $\varphi(\log_2 n) = \varphi(\log_2 n) = \varphi(k + \log_2 \xi) = \varphi(\log_2 \xi)$. By expanding $n!$ and $j!$ in (4.72) by Stirling’s formula and rewriting the result in terms $\xi$, we see that the matching condition is satisfied provided that

$$n\Phi(\xi) \sim \frac{1}{2} \rho_0 \varphi(\log_2(\xi)) n\xi^{3/2} \exp \left( \frac{-\log^2(\xi)}{2 \log 2} \right), \quad \xi \to 0^+. $$

This yields the behavior of $\Phi(\xi)$ as $\xi \to 0$ in Theorem 3, whose derivation is now complete.

It follows that $\Phi(\xi)$ and all its derivatives vanish as $\xi \to 0^+$. Also, (4.73) shows that the structure of $\Phi$ for the PATRICIA model is more complicated than the corresponding function $(-\phi)$ for $b$-tries, as the latter did not have the oscillatory behavior of the former. The behavior as $\xi \to 1^-$ is similar to that of $b$-tries, except that the constants $D_1$ and $D_2$ differ for the two models.

5 Numerical Results

In this section we discuss some numerical calculations. They will give an idea of the accuracy of the asymptotic formulas for $h_n^k$ for the PATRICIA model, and also will verify some of the assumptions we made in our analysis.

We define, for some integer $\ell$,

$$k^* = \begin{cases} \log_2 n & \text{if } n = 2^\ell \\ \lfloor \log_2 n \rfloor + 1 & \text{if } n \neq 2^\ell \end{cases} $$

and

$$k_1 = \left\lfloor \log_2 n + \sqrt{2\log_2 n - \frac{3}{2}} \right\rfloor + 1, $$

34
Table 1: Right-Tail Comparison

<table>
<thead>
<tr>
<th>n</th>
<th>j = n - k</th>
<th>1 - h_n^k (Numerical)</th>
<th>1 - h_n^k (Theorem 3(i))</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2</td>
<td>1.782(10^{-7})</td>
<td>1.786(10^{-7})</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4.469(10^{-5})</td>
<td>4.571(10^{-5})</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3.040(10^{-3})</td>
<td>3.251(10^{-3})</td>
</tr>
<tr>
<td></td>
<td>5 (= n - k_1)</td>
<td>7.071(10^{-2})</td>
<td>8.322(10^{-2})</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>2.684(10^{-39})</td>
<td>2.684(10^{-39})</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1.343(10^{-24})</td>
<td>1.343(10^{-24})</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>3.463(10^{-8})</td>
<td>3.517(10^{-8})</td>
</tr>
<tr>
<td></td>
<td>14 (= n - k_1)</td>
<td>0.2040</td>
<td>0.2676</td>
</tr>
<tr>
<td>30</td>
<td>2</td>
<td>5.176(10^{-99})</td>
<td>5.176(10^{-99})</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2.781(10^{-75})</td>
<td>2.781(10^{-75})</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>8.200(10^{-44})</td>
<td>8.200(10^{-44})</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>5.411(10^{-21})</td>
<td>5.417(10^{-21})</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>2.050(10^{-6})</td>
<td>2.108(10^{-6})</td>
</tr>
<tr>
<td></td>
<td>23 (= n - k_1)</td>
<td>0.1288</td>
<td>0.1616</td>
</tr>
</tbody>
</table>

as in Section 2. For a fixed n, h_n^k will satisfy 0 < h_n^k < 1 for k^* ≤ k ≤ n - 2, and our analysis predicts that

\[
\begin{align*}
    h_n^k &\to 1 \quad \text{as} \quad n \to \infty \quad \text{if} \quad k \geq k_1, \quad (5.3) \\
    h_n^k &\to 0 \quad \text{as} \quad n \to \infty \quad \text{if} \quad k < k_1 - 1, \quad (5.4)
\end{align*}
\]

i.e., all the mass concentrates at k = k_1 if \( \sqrt{2 \log_2 n (\log_2 n + \sqrt{2 \log_2 n - 1.5})} \to \infty \). If \( \sqrt{2 \log_2 n (\log_2 n + \sqrt{2 \log_2 n - 1.5})} = O(1) \) or \( \sqrt{2 \log_2 n - \sqrt{2 \log_2 n (\log_2 n + \sqrt{2 \log_2 n - 1.5})} = O(1) \), then the probability mass is concentrated on two points: either k_1 - 1 and k_1 or k_1 and k_1 + 1, as discussed in Corollary 2.

We first consider the right tail of the distribution, where Theorem 3(i) applies. In Table 1 we compare the exact (numerical) values of 1 - h_n^k to the asymptotic formula in Theorem 3(i). To evaluate the latter we computed C_j recursively using (4.14). We consider n = 10, 20 and 30. For each value of n, we start with j = n - k = 2 and increase j to n - k_1. For  j > n - k_1, 1 - h_n^k may become negative, as then we are clearly out of the range of validity of this asymptotic result. Table 1 shows good agreement between asymptotic
Table 2: Left-Tail Comparison

<table>
<thead>
<tr>
<th>n</th>
<th>k</th>
<th>M</th>
<th>$h_n^k$ (Numerical)</th>
<th>$h_n^k$ (Theorem 3(iii))</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>3</td>
<td>0</td>
<td>5.062(10^{-2})</td>
<td>5.010(10^{-2})</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>0</td>
<td>5.032(10^{-4})</td>
<td>5.006(10^{-4})</td>
</tr>
<tr>
<td>15</td>
<td>4</td>
<td>1</td>
<td>4.734(10^{-3})</td>
<td>4.275(10^{-3})</td>
</tr>
<tr>
<td>32</td>
<td>5</td>
<td>0</td>
<td>3.544(10^{-8})</td>
<td>3.534(10^{-8})</td>
</tr>
<tr>
<td>31</td>
<td>5</td>
<td>1</td>
<td>6.667(10^{-7})</td>
<td>6.341(10^{-7})</td>
</tr>
<tr>
<td>30</td>
<td>5</td>
<td>2</td>
<td>6.197(10^{-6})</td>
<td>5.324(10^{-6})</td>
</tr>
<tr>
<td>64</td>
<td>6</td>
<td>0</td>
<td>1.248(10^{-15})</td>
<td>1.246(10^{-16})</td>
</tr>
<tr>
<td>63</td>
<td>6</td>
<td>1</td>
<td>4.694(10^{-15})</td>
<td>4.579(10^{-15})</td>
</tr>
<tr>
<td>62</td>
<td>6</td>
<td>2</td>
<td>8.780(10^{-14})</td>
<td>8.148(10^{-14})</td>
</tr>
<tr>
<td>61</td>
<td>6</td>
<td>3</td>
<td>1.089(10^{-12})</td>
<td>9.353(10^{-13})</td>
</tr>
</tbody>
</table>

and exact results. As expected, the further we get into the tail, the more accurate is the asymptotic formula.

Next, we test the left tail approximation from Theorem 3(iii). For a fixed large $n$, the condition $2^k - n = O(1)$ cannot be satisfied unless $n$ is close to a power of 2. However, for a fixed large $k$, this condition can be satisfied. It is thus convenient to do the comparisons when $n$ is a power of 2, and then decrease $n$. In Table 2 we consider $k \in [3, 6]$ and various values of $M = 2^k - n$. For a fixed $M$ we see that the accuracy of the asymptotic result increases with $n$. Also, as $n$ becomes larger, we can allow for larger values of $M$ and still get good agreement.

The formula that applies for $k \approx k_1$, which is where there is significant mass, is given by (2.21), which corresponds to $A(\xi)e^{-n\Phi(n)}$, with $\Phi(\xi)$ replaced by its small $\xi$ expansion and $A(\xi) \sim 1$. We can refine this by using $A(\xi) = \sqrt{1 + 2\xi \Phi''(\xi) + \xi^2 \Phi''''(\xi)}$ with $\Phi(\xi)$ computed from (4.73). In Table 3 we compare the exact values of $h_n^k$ to the expression in (2.21) and also the refinement $Ae^{-n\Phi}$ discussed above, for $n = 10, 20, 30, 50, 100, 150, 250$ and 500. We also tabulated $\xi = n2^{-k}$, since (2.21) assumes that $\xi$ is small. For each $n$ we consider $k = k_1 - 1, k_1$ and $k_1 + 1$. When $n = 10$ the probability mass at $k_1$ is about .49, and this increases to .71 when $n = 150$. Our asymptotic result predicts the values .53 and .73 when $n = 10$ and 150, respectively. Table 3 is consistent with our prediction that the mass accumulates at $k_1$, but the agreement between the exact and asymptotic result (2.21) is not
Table 3: Central Regime Comparison

<table>
<thead>
<tr>
<th>n</th>
<th>k</th>
<th>$\xi_{n} (\approx n2^{-k})$</th>
<th>$h_{n}^{\xi}$ (Numerical)</th>
<th>$h_{n}^{\xi}$ (2.21)</th>
<th>$h_{n}^{\xi}$ (2.17), $\xi \to 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>4</td>
<td>.63</td>
<td>.43975</td>
<td>.30024</td>
<td>.37067</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.31</td>
<td>.92929</td>
<td>.82862</td>
<td>.87673</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>.16</td>
<td>.99696</td>
<td>.98542</td>
<td>.99145</td>
</tr>
<tr>
<td>20</td>
<td>5</td>
<td>.63</td>
<td>.12641</td>
<td>.09015</td>
<td>.11129</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>.31</td>
<td>.79599</td>
<td>.68661</td>
<td>.72648</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>.16</td>
<td>.98564</td>
<td>.97105</td>
<td>.97699</td>
</tr>
<tr>
<td>30</td>
<td>6</td>
<td>.47</td>
<td>.24414</td>
<td>.16249</td>
<td>.18627</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>.23</td>
<td>.87115</td>
<td>.80820</td>
<td>.82855</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>.12</td>
<td>.99283</td>
<td>.98760</td>
<td>.98948</td>
</tr>
<tr>
<td>50</td>
<td>7</td>
<td>.39</td>
<td>.23770</td>
<td>.15836</td>
<td>.17423</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>.20</td>
<td>.87927</td>
<td>.83529</td>
<td>.84681</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>.10</td>
<td>.99401</td>
<td>.99125</td>
<td>.99210</td>
</tr>
<tr>
<td>100</td>
<td>8</td>
<td>.39</td>
<td>.04845</td>
<td>.02508</td>
<td>.02759</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>.20</td>
<td>.75483</td>
<td>.60772</td>
<td>.70733</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>.10</td>
<td>.99623</td>
<td>.98259</td>
<td>.98342</td>
</tr>
<tr>
<td>150</td>
<td>9</td>
<td>.29</td>
<td>.15798</td>
<td>.10126</td>
<td>.10619</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>.15</td>
<td>.87248</td>
<td>.84558</td>
<td>.84961</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>.07</td>
<td>.99491</td>
<td>.99388</td>
<td>.99409</td>
</tr>
<tr>
<td>250</td>
<td>10</td>
<td>.24</td>
<td>.18706</td>
<td>.12842</td>
<td>.13209</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>.12</td>
<td>.89950</td>
<td>.88226</td>
<td>.88426</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>.06</td>
<td>.99669</td>
<td>.99618</td>
<td>.99627</td>
</tr>
<tr>
<td>500</td>
<td>11</td>
<td>.24</td>
<td>.03336</td>
<td>.01649</td>
<td>.01696</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>.12</td>
<td>.80512</td>
<td>.77838</td>
<td>.78014</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>.06</td>
<td>.99313</td>
<td>.99238</td>
<td>.99247</td>
</tr>
</tbody>
</table>
Table 4: Probability mass at $k_0$ and $k_1$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\ell$</th>
<th>$k_n^i$ (Numerical)</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>0</td>
<td>0.6507</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.9675</td>
</tr>
<tr>
<td>23</td>
<td>0</td>
<td>0.6432</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.9674</td>
</tr>
<tr>
<td>39</td>
<td>0</td>
<td>0.6162</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.9665</td>
</tr>
<tr>
<td>66</td>
<td>0</td>
<td>0.6095</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.9692</td>
</tr>
<tr>
<td>113</td>
<td>0</td>
<td>0.6016</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.9719</td>
</tr>
</tbody>
</table>

particularly good. However, the refined results are better as long as $\xi$ is reasonable small. Also, in order to see the importance of the left tail approximation in parts (ii) and (iii) of Theorem 3, we need for $k_1 - k^*$ to be fairly large. However, when $n = 150$, $k_1 = 10$ and $k^* = 8$ so that the "left tail" really consists of the two points $k = 8, 9$. In order to have $k_1 - k^*$ as large as, say, 10, $n$ would have to be about $10^{20}$ (cf. (5.1) and (5.2)).

To better see the convergence of mass to one or two points, it is useful to consider subsequences $n_i$ of $n$ that correspond to $\beta_n$ nearly constant. To optimize the mass at $k_0 = k_1 - 1$ we want to minimize $\beta_n$. From Figure 3 we see that the local minima of $R(n)$ (and $\beta_n$) occur at

$$n_i = \left[2^{i+5/2-\sqrt{2i+4}}\right] + 1.$$ (5.5)

The first few integers in this subsequence are: 3, 4, 6, 9, 14, 23, 39, 66, 113 and 195. By using (5.5) in (2.23) and evaluating the result asymptotically, we obtain

$$R(n_i) = \sqrt{2\log_2 n_i \beta_{n_i} \left(1 - I_i \sqrt{2x + \frac{4}{\log 2}} 2^{-i-\frac{5}{2}+\sqrt{2i+4}}(1 + O(n_{i-1}^{-1}))\right)}$$ (5.6)

where

$$I_i = \left(2^{i+5/2-\sqrt{2i+4}}\right).$$

It follows that along $n_i$, $R(n_i)$ is not only bounded, but approaches zero, roughly like $O(n_{i-1}^{-1})$. 38
Thus along this subsequence (2.25) becomes, for $\ell = 0$ and $n \to \infty$

$$
\Pr\{\mathcal{H}_n^P \leq \lfloor \log_2 n + \sqrt{2\log_2 n - 1.5} \rfloor \} \sim \exp \left( -\rho_0 e^{\theta + \Psi(\log_2 n)} \right).
$$

(5.7)

While the above does not approach a limit as $n \to \infty$, the numerically small value of $\Psi(\cdot)$ allows us to approximate (5.7) by

$$
\Pr\{\mathcal{H}_n^P \leq \lfloor \log_2 n + \sqrt{2\log_2 n - 1.5} \rfloor \} \approx 0.536426...
$$

which is obtained by neglecting $\Psi(\cdot)$. This yields the optimal mass at $k_0 = k_1 - 1$ and the remaining mass ($\approx 0.464$) will be at $k_1$. In Table 4 we compare the exact values of $h_n^k$ along this subsequence, for $k = k_0$ and $k_1$. We see that $h_n^{k_0}$ is slowly "converging" to the theoretical value.

By choosing the subsequence $n_i - 1$, we achieve local maxima of $\beta_n$ and $R(n)$. We can similarly show that $\beta_{n_i - 1} \to 1^-$ and furthermore $\sqrt{2\log_2 (n_i - 1) - R(n_i - 1)} \to 0$. Now, we will have about .536 of the mass at $k_1$ and the remaining .464 at $k_2 = k_1 + 1$. By choosing other subsequences we can achieve any value of mass at $k_3$ in the range $[0, 0.536...]$ and any value of mass at $k_2$ in the range $[0, 0.463...]$.

For most $n_i$ the mass will be at a single point $k_1$. To optimize the convergence we choose $n_i$ so that $\beta_{n_i} \approx 0.5$. This simultaneously avoids the mass at $k_1 - 1$ and $k_1 + 1$. We accomplish this by selecting $\{n_i^*\}_{i=1}^{\infty}$ such that

$$
n_i^* = \left[2^{i+3} - \sqrt{275}\right].
$$

In Table 5 we consider a few values of $n_i^*$ and show the exact $h_n^k$ for $k = k_0$ and $k = k_1$. When $n = 10(= n_{11}^*)$ the mass at $k_1$ is about 0.49 and it gradually increases to 0.76 when $n = 446(= n_{11}^*)$. We also note that if $\beta_n = 0.5$ and $\ell = 0$, then (2.25) becomes

$$
\Pr\{\mathcal{H}_n^P \leq \log_2 n + \sqrt{2\log_2 n - 2} \} \sim \exp \left( -\rho_0 e^{\theta + \Psi(\log_2 n)} \right)^{2^{-1/8} \sqrt{2\log_2 n} / 2} \quad (5.8)
$$

where in the approximation we have neglected the oscillatory term. To make (5.8) less than 0.01 requires $n > 288000$. The last two entries in Table 5 are computed according to the approximate formula (5.8). These results confirm our prediction that all mass concentrates at the one point $k_1$.

Finally, we discuss the numerical computation of the function $\Phi(\xi)$. We define $\Phi_{NUM}$ by

$$
\Phi_{NUM}(\xi; k) = \Phi_{NUM}(n2^{-k}; k) \equiv -\frac{1}{n} \log(h_n^k).
$$

(5.9)
Table 5: Probability mass at $k_1$ only.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\ell$</th>
<th>$h_n^k$ (Numerical)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0</td>
<td>.4397</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.9293</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>.3559</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.9049</td>
</tr>
<tr>
<td>29</td>
<td>0</td>
<td>.2946</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.8904</td>
</tr>
<tr>
<td>49</td>
<td>0</td>
<td>.2674</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.8904</td>
</tr>
<tr>
<td>85</td>
<td>0</td>
<td>.2169</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.8839</td>
</tr>
<tr>
<td>147</td>
<td>0</td>
<td>.1847</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.8847</td>
</tr>
<tr>
<td>256</td>
<td>0</td>
<td>.1546</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.8663</td>
</tr>
<tr>
<td>446</td>
<td>0</td>
<td>.1340</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.8916</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>80226</td>
<td>0</td>
<td>.0161</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.9839</td>
</tr>
<tr>
<td>1571598</td>
<td>0</td>
<td>.0051</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.9948</td>
</tr>
</tbody>
</table>
According to our analysis, as $k \to \infty$ for each fixed $0 < \xi \leq 1$, we should have $\Phi_{NUM}(\xi; k) \to \Phi(\xi)$. We have also analytically computed

$$\Phi(1) = 1 + \log(K_0) = .61906125 \ldots .$$

(5.10)

In Table 6 we evaluate $\Phi_{NUM}(1; k)$ for $k = 4, 5, 6$ and 7. This sequence certainly appears to be converging to the theoretical value in (5.10). Except for $\xi = 1$ we do not have the exact values of $\Phi(\xi)$, however, it is clear from Figure 5 that the sequence of functions $\Phi_{NUM}(\xi; k)$ is indeed converging to a limit. To give an approximation to $\Phi(\xi)$ we plot $\Phi_{NUM}(\xi, k)$ in Figure 5 for $k = 4, 5$ and 6.

A Proof of the equivalence of (4.24) and (4.28)

In this Appendix we prove the equivalence of the two representations (4.24) and (4.28) for $C_j$.

The expression in (4.28) may be simplified by closing the contour of integration in the left half-plane and noting that the integrand has simple poles at $s = -1, -2, \ldots$. Denoting the integral by $F(z)$, we have

$$F(z) = \frac{1}{2\pi i} \int_{-\frac{1}{2} + \infty}^{-\frac{1}{2} - \infty} z^{-s} \Gamma(s + 1) \zeta(s + 1) ds = \sum_{m=1}^{\infty} \frac{z^m \zeta(1-m) (-1)^{m+1}}{m(2^m - 1)(m-1)!},$$

(A.1)
Table 6: Numerical Evaluation of $\Phi(l)$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$k$</th>
<th>$\Phi_{NUM}(1; k)$</th>
<th>$\Phi(1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>4</td>
<td>.47466</td>
<td>.61906</td>
</tr>
<tr>
<td>32</td>
<td>5</td>
<td>.53611</td>
<td>.61906</td>
</tr>
<tr>
<td>64</td>
<td>6</td>
<td>.57219</td>
<td>.61906</td>
</tr>
<tr>
<td>128</td>
<td>7</td>
<td>.59292</td>
<td>.61906</td>
</tr>
</tbody>
</table>

for $z$ sufficiently small. But, $\zeta(0) = -1/2$ and

$$\zeta(-2n) = 0, \; n = 1, 2, \ldots \; \zeta(1-2n) = \frac{-B_{2n}}{2n}, \; n = 1, 2, \ldots$$

where $B_n$ are the Bernoulli numbers (cf. [1]). Thus, (A.1) becomes

$$F(z) = \mathcal{M}^{-1} \left[ \frac{\Gamma(s + 1) \zeta(s + 1)}{s(1 - 2^{-s})} \right] = -\frac{z}{2} + \sum_{t=1}^{\infty} \frac{z^{2t}}{(2t)!} \frac{1}{2^{2t - 1}}.$$

To show the equivalence of (4.24) and (4.28), we represent the Bernoulli numbers as integrals:

$$\frac{B_{2t}}{(2t)!} = \frac{1}{2\pi i} \int \frac{1}{e^t - 1} \frac{1}{t^{2t}} dt, \quad 0 < |t| < 2\pi.$$

Using the above, expanding $1/(2^{2t} - 1) = \sum_{k=0}^{\infty} 2^{-2t(k+1)}$, and noting that $B_3 = B_5 = \cdots = 0$, we obtain

$$F(z) = \sum_{k=0}^{\infty} \sum_{m=1}^{\infty} z^m 2^{-m(k+1)} \left[ \frac{1}{2\pi i} \int \frac{t^{-m}}{e^t - 1} dt \right]$$

$$= \frac{1}{2\pi i} \int \sum_{k=0}^{\infty} \log(1 - z 2^{-k-1}/t) dt.$$  

Setting $t = z 2^{-k-1} \xi$ and integrating by parts yields

$$F(z) = \frac{1}{2\pi i} \sum_{k=0}^{\infty} \int \frac{1}{\xi(\xi - 1)} \log \left( \frac{1 - \exp(-z \xi 2^{-k-1})}{z \xi 2^{-k-1}} \right) d\xi$$  \hspace{1cm} (A.2)

where $|\xi| > 1$ on the loop of the integration. For any $k$ the integrand in (A.2) is analytic at $\xi = 0$ and has a simple pole at $\xi = 1$. By evaluating the residue at $\xi = 1$, we see that $\exp[F(z)]$ becomes the same as the infinite product in (4.24). This establishes the equivalence of (4.24) and (4.28).
References


