On the Distribution for the Duration of a Randomized Leader Election Algorithm

James Allen Fill*
Department of Mathematical Sciences
The Johns Hopkins University
Baltimore, MD 21218–2689

Hosam M. Mahmoud†
Department of Statistics
The George Washington University
Washington, D.C. 20052

Wojciech Szpankowski‡
Department of Computer Science
Purdue University
West Lafayette, IN 47907

March 1996

Abstract

We investigate the duration of an elimination process for identifying a winner by coin tossing, or, equivalently, the height of a random incomplete trie. Applications of the process include the election of a leader in a computer network. Using direct probabilistic arguments we obtain exact expressions for the discrete distribution and the moments of the height. Elementary approximation techniques then yield asymptotics for the distribution. We show that no limiting distribution exists, as the asymptotic expressions exhibit periodic fluctuations.

In many similar problems associated with digital trees, no such exact expressions can be derived. We therefore outline a powerful general approach, based on the analytic techniques of Mellin transforms, Poissonization, and de-Poissonization, from which distributional asymptotics for the height can also be derived. In fact, it was this complex variables approach that led to our original discovery of the exact distribution. Complex analysis methods are indispensable for deriving asymptotic expressions for the mean and variance, which also contain periodic terms of small magnitude.

Key words and Phrases: Random trees, tries, height, distributed computing, leader election, asymptotic distribution, Poissonization, de-Poissonization.

AMS 1991 Classification Codes. Primary: 05C05, 60F05; secondary: 05C80, 60G70.

*This author's research has been supported by NSF grant DMS–9311367.

†This author's research has been supported by NSA grant MDA904–92–H3086.

‡This author's research has been supported by NSF grants NCR–9206315, NCR–9415491, and CCR–9201078.

1 Overview

1.1 Introduction

The following elimination process has several applications, such as the election of a leader in a computer network. A group of n people play a game to identify a winner by tossing fair coins. All players who throw heads are losers; those who throw tails remain candidate winners and flip their coins again. The process is repeated among candidate winners until a single winner is identified. If at any stage all remaining candidate winners throw heads, the tosses are deemed inconclusive and all remaining players participate again as candidate winners in the next round of coin tossing.

We investigate the distribution of the height of a random incomplete trie, the discrete structure that underlies the elimination process described above. Height distributions of random digital trees have usually been attacked by purely probabilistic methods that only identify the leading terms [cf. Devroye (1992), Flajolet (1983), Mendelson (1982), and Pittel (1985, 1986)]. Our approach here is mainly analytic and provides a mechanical way of computing moments and the asymptotic distribution. Having found the asymptotic distribution via this approach, we were able to go back and find a direct probabilistic approach, also discussed in this paper. However, it appears that the analytic approach must be used in order to find asymptotic moments, whose calculation we also discuss.

The elimination process described above provides the foundation for an efficient randomized distributed algorithm for leader election in a computer network. A computer network comprising n identical processors needs to have one of the processors acting as the leader to supervise communication and synchronization in the network. Communication is accomplished by exchanging messages and waiting an allotted amount of time for a response. Control messages of a special kind (called *tokens*) are sometimes passed within the network.

Because of routine hardware failures (such as the development of bad sectors on disks) or software failures (such as losing the token or degradation of synchronization) a leader may temporarily go out of service, in which case the remaining active processors need to agree on a new leader. The processors have identification numbers, say $1, 2, \ldots, n$. The failure of the leader may be detected when some other processor j (say) sends a message to the leader but receives no answer in the amount of time allotted for a response. Processor j then initiates an election by sending a message to all the other processors. (It is possible that several processors detect the leader's failure simultaneously, in which case all processors that encounter the failure initiate the election message simultaneously.) Every active processor receiving the election message suspends its routine computing and simulates the entire elimination process locally by generating an unbiased Bernoulli random variable for each coin flip. All the simulations are identical if all processors use the same random number generator and all start from the same seed value [see Devroye's (1986) encyclopedia on random number generation]. The winner broadcasts its success by sending a message to all processors and rewrites the list of active processors by including only those processors that send a congratulatory acknowledgement. The updated list is sent to all processors along with the seed of the random number generator, because it is possible that some processors have come back into operation since the last election. Those processors have lost track of the current seed and the current list of active processors.

The distributed algorithm discussed above determines a leader in average time of logarithmic order in n, as implied in Prodinger (1993). Thus, the average duration of the distributed algorithm based on simulating the elimination process is better than the usual linear-time deterministic algorithms currently in use for leader election [see Brassard and Bratley (1988)]. Our results will also reveal a sharp concentration around the average; the variance is only of constant order, giving rise to a very narrow probability profile around the mean.

Our algorithm (and any other leader election algorithm, for that matter) will need also to resolve the situation where the elected leader itself goes down during the election process. In this case a reelection must be held. Thus the algorithm discussed above is only the basic building block of a more elaborate algorithm for leader election. This building block alone will handle the election most of the time in a reliable network with low failure rates, i.e., when the average wait until the next failure of any computer is much larger than the average time it takes to simulate the leader election algorithm. This is a reasonable assumption in a modern computer network of moderate size, where the average wait until the next failure is on the order of a few hours, whereas the average time of local simulation is only a fraction of a second.

1.2 The Height of a Random Incomplete Trie

A binary tree structure underlies the elimination process we have discussed. At the root of the tree we have one node labelled with all participants. After all the participants flip their coins for the first time, losers (if any) are placed in a *leaf* node that is attached to the root as a right child, and all candidate winners are placed in a non-leaf node that is attached as a left child. Leaf nodes are terminal nodes that are not developed any further. (If at some stage all remaining candidate winners flip heads, they are all placed in a non-leaf node attached as a right child.) The process repeats recursively on every non-leaf node until a single winner is identified. The node containing the winner is also considered a leaf, as it is terminal. Figure 1 illustrates the discrete structure underlying the elimination process. In Figure 1, leaf nodes are represented as rectangles; all the other nodes of the tree have an oval shape. An edge of the tree in Figure 1 leading from a parent node to its child is labelled with H (head) or T (tail) according to the result obtained by the group within the child node. This random discrete structure is similar in some aspects to the random

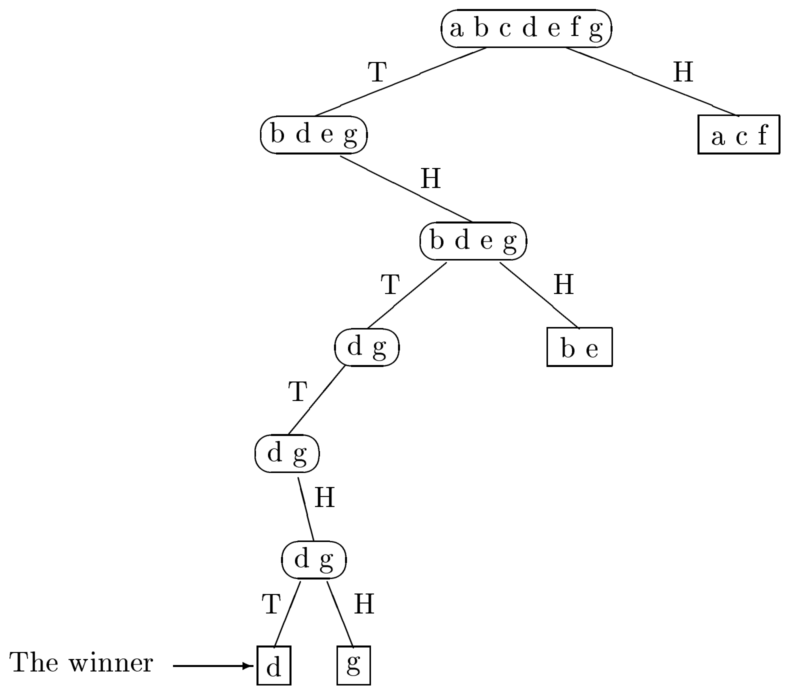

Figure 1. An incomplete trie for the elimination process starting with 7 players.

trie structure, a classical data structure for digital data [see Knuth (1973b) or Mahmoud (1992)]. The difference between the discrete structure of the elimination process and the standard trie is that in the trie all the nodes are further developed if they contain more than two data items so that each datum is eventually in a node by itself. Thus, in a sense, the tree structure underlying the elimination process is an *incomplete trie* and will be so called in this paper. The terminology was coined by Prodinger (1993), who introduced this tree structure and found the average behavior of several of its characteristic properties. Grabner (1993) generalized the process to that of identifying several winners instead of only one. Grabner (1993) found the average behavior of some of the characteristic properties of this more general incomplete trie.

The elimination process to identify a single or several winners also has the spirit of a class of problems posed by Rényi (1961) in his lecture series at Michigan State University. Pittel and Rubin (1992) find connections between one of Rényi's interesting questions and the notion of a PATRICIA (for "Practical Algorithm To Retrieve Information Coded In Alphanumeric") tree, a kind of trie with path compression for faster data retrieval [Knuth (1973b)].

The *height* of an incomplete trie is the length of the path from the root to the winner, which is the longest root-to-leaf path in the tree. We shall denote the height of an incomplete trie underlying the elimination process beginning with n players by H_n. This quantity is the number of elimination rounds until the winner is identified, which is a measure of the time duration of the elimination process, if all the coin tosses at any stage are carried out simultaneously (as is the case in the leader election distributed algorithm).

In this paper we investigate the asymptotic distribution of H_n. This random variable has a very wide range. For $n \geq 2$, it can assume any value in $\{1, 2, \ldots\} \cup \{\infty\}$. We shall develop asymptotics for the distribution function of a centered version of H_n. Even with the proper centering, we shall see that no limit distribution exists. However, the distribution function of the centered H_n oscillates between well-defined extremes. More specifically, the periodic function[1]

$$\alpha(n) := \lg n - \lfloor \lg n \rfloor$$

appears in the distribution of the centered height $H_n - \lfloor \lg n \rfloor$. Corresponding to values of n that are integer powers of 2, the upper extreme is a discrete distribution function that coincides at the integer points with the continuous distribution function

$$\frac{2^{-x}}{\exp(2^{-x}) - 1}.$$

As n gradually increases, the periodic effect of $\alpha(n)$ on the distribution is to lower the staircase distribution down (at any fixed argument) until it comes very close to the other extreme discrete distribution function, which is also a staircase that coincides at the integer points with the continuous distribution function

$$\frac{2^{1-x}}{\exp(2^{1-x}) - 1},$$

[1]In this paper, lg denotes logarithm with base 2; the natural logarithm is denoted, as usual, by ln.

before it "wraps around" to approach the upper continuous distribution function when n becomes a power of 2 again. This behavior is formally expressed as a corollary to Theorem 2.

The plan of this paper is as follows. We present the main results of this paper in Section 2. The results come in two flavors: exact (Theorem 1) and asymptotic (Theorem 2). The rest of Section 2 discusses several ramifications of the main results and connections to the digital tree structures used for computer data storage.

In Section 3 the exact results and their proofs are fully developed. In Section 3.1, the exact distribution and exact mean and variance are found. Then, in Section 3.2, the exact distribution is manipulated by elementary asymptotic techniques to yield an asymptotic approximation and tight rates of approach, from which a Berry–Esséen type result follows.

The exact distribution was actually first obtained by an analytic method based on a generating function approach. Once we had obtained the exact distribution, we realized that it can also be derived by the directly probabilistic arguments of Section 3.1.

The analytic formulation is the focus of Section 4. Moment calculations are known to be somewhat more tractable under Poissonization, that is, when the number of relevant objects (the number of players, in our case) is assumed to follow a Poisson distribution instead of being a fixed number. A subsequent step of de-Poissonization transforms the results back to the fixed-population model, the probability model of prime interest. This method, requiring a foray into the complex domain, has been successfully applied in average-case analyses, as well as for a few variance calculations, in digital methods. We apply the method in Section 4.2 to derive asymptotic expressions for the mean and variance of height. It is a main objective of this paper to show (see Section 4.1) that the method extends beyond mean and variance calculations to asymptotic distributions (in the context of order statistics of dependent random variables).

2 Main Results

In this section we present our main results for the height H_n and conclude with some remarks. Our first main theorem is concerned with exact values, which involve Bernoulli numbers. For background on these numbers, see, e.g., Knuth (1973a). In Section 3.1 [Lemma 2(ii)] we shall in fact find exact expressions for all the moments.

Theorem 1. *Consider an incomplete binary trie with $n \geq 1$, and let B_j denote the jth Bernoulli number. Then:*

(i) *For any integer* $k \geq 0$,

$$\mathbf{Prob}\{H_n \leq k\} = \frac{n}{2^{kn}} \sum_{j=0}^{2^k-1} j^{n-1} = \sum_{j=0}^{n-1} \binom{n}{j} \frac{B_j}{2^{kj}}. \tag{1}$$

(ii) [Prodinger (1993)]. *The average height* $\mathbf{E}[H_n]$ *is given exactly by*

$$\mathbf{E}[H_n] = -\sum_{j=1}^{n-1} \binom{n}{j} \frac{B_j}{1-2^{-j}}.$$

(iii) *The variance* $\mathbf{Var}[H_n]$ *of the height is given exactly by*

$$\mathbf{Var}[H_n] = -\sum_{j=1}^{n-1} \binom{n}{j} B_j \frac{1+2^{-j}}{(1-2^{-j})^2} - \mathbf{E}^2[H_n].$$

Computationally, the formulas of Theorem 1(ii) and 1(iii) become unwieldy as n becomes large. The Bernoulli numbers vanish for odd index ≥ 3, but for even index they increase in magnitude very rapidly and alternate in sign. It is desirable then to have asymptotic approximations involving only elementary functions. The next theorem gives such accurate asymptotic approximations.

Theorem 2. *Define* $L := \ln 2$ *and* $\chi_k := 2\pi i k/L$. *Then:*

(i) *Uniformly over all integers* k,

$$\mathbf{Prob}\{H_n \leq \lfloor \lg n \rfloor + k\} = \frac{2^{\alpha(n)-k}}{\exp(2^{\alpha(n)-k}) - 1} + O\left(\frac{1}{\sqrt{n}}\right)$$

as $n \to \infty$.

(ii) [Prodinger (1993)]. *The average height* $\mathbf{E}[H_n]$ *satisfies*

$$\mathbf{E}[H_n] = \lg n + \frac{1}{2} - \delta_1(\lg n) + O\left(\frac{1}{n}\right),$$

where $\delta_1(\cdot)$ *is a periodic function of magnitude* $\leq 2 \times 10^{-5}$ *given by*

$$\delta_1(x) := \frac{1}{L} \sum_{\substack{-\infty<k<\infty \\ k\neq 0}} \zeta(1-\chi_k)\,\Gamma(1-\chi_k)\,e^{2\pi i k x}, \tag{2}$$

and $\zeta(\cdot)$ *and* $\Gamma(\cdot)$ *denote Riemann's zeta function and Euler's gamma function, respectively.*

(iii) *The variance* $\mathbf{Var}[H_n]$ *of the height satisfies*

$$\begin{aligned}\mathbf{Var}[H_n] &= \frac{\pi^2}{6L^2} + \frac{1}{12} - \frac{2\gamma_1}{L^2} - \frac{\gamma^2}{L^2} + \delta_2(\lg n) + O\left(\frac{\ln n}{n}\right) \\ &= 3.116695\ldots + \delta_2(\lg n) + O\left(\frac{\ln n}{n}\right).\end{aligned}$$

Here the constants $(-1)^k \gamma_k / k!$, $k \geq 0$, *are the so-called Stieltjes constants, with*

$$\gamma_k := \lim_{m\to\infty} \left(\sum_{i=1}^{m} \frac{\ln^k i}{i} - \frac{\ln^{k+1} m}{k+1} \right);$$

in particular, $\gamma_0 = \gamma = 0.577215\ldots$ *is Euler's constant and* $\gamma_1 = -0.072815\ldots$. *The periodic function* $\delta_2(\cdot)$ *has magnitude* $\leq 2 \times 10^{-4}$.

Remark 1. (a) The function $\delta_2(\cdot)$ appearing in the asymptotic formula for the variance is explicitly given by

$$\delta_2(x) := \frac{2}{L^2} \sum_{\substack{-\infty<k<\infty \\ k\neq 0}} d_2(\chi_k)\, e^{2\pi i k x} - \delta_1^2(x), \tag{3}$$

where

$$d_2(s) := \zeta(1-s)\,\Gamma(-s) - \zeta'(1-s)\, s\,\Gamma(-s) - s\,\zeta(1-s)\,\psi(-s)\,\Gamma(-s).$$

Here $\psi(\cdot)$ is the classical psi (or digamma) function $\psi(s) := \Gamma'(s)/\Gamma(s)$.

(b) The bound on δ_1 was obtained by using

$$|\delta_1(x)| \leq \frac{1}{L} \sum_{\substack{-\infty<k<\infty \\ k\neq 0}} |\zeta(1-\chi_k)\,\Gamma(1-\chi_k)|$$

and numerically evaluating the series on the right, which converges very rapidly. The bound on δ_2 was obtained in a similar fashion.

(c) While the function $\delta_1(\cdot)$ fluctuates symmetrically about 0, the same is not quite true for $\delta_2(\cdot)$. In fact, one sees from (2) and (3) that the fluctuations of $\delta_2(\cdot)$ are symmetric about the constant

$$-\frac{1}{L^2} \sum_{k=1}^{\infty} |\zeta(1-\chi_k)\,\Gamma(1-\chi_k)|^2$$

whose value is approximately -10^{-10}.

(d) A starting point for background on the zeta and gamma functions is Abramowitz and Stegun (1972). The proof in Section 4.2 will show how these arise naturally in our calculations. More information on the Stieltjes constants can be found in Berndt (1985), pp. 164–165. These arise in our calculations as a result of the Laurent series expansion

$$\zeta(s) = \frac{1}{s-1} + \sum_{k=0}^{\infty} \frac{(-1)^k \gamma_k}{k!} (s-1)^k,$$

valid for all complex $s \neq 1$.

The sequence $\alpha(n)$ appearing in Theorem 1 is dense on the interval $[0, 1)$ (though not uniformly dense): see Kuipers and Niederreiter (1974). This consideration leads us to conclude the following.

Corollary 1. *A limit distribution does not exist for $H_n - \lfloor \lg n \rfloor$. However, for each fixed integer k,*

$$\begin{aligned}
\liminf_{n\to\infty} \mathbf{Prob}\{H_n \le \lfloor \lg n \rfloor + k\} &= \frac{2^{-k+1}}{\exp(2^{-k+1}) - 1}, \\
\limsup_{n\to\infty} \mathbf{Prob}\{H_n \le \lfloor \lg n \rfloor + k\} &= \frac{2^{-k}}{\exp(2^{-k}) - 1}.
\end{aligned}$$

Now we are in position to discuss some consequences of our main results and to offer some additional remarks. Corollary 1 indicates that the asymptotic distribution of $H_n - \lfloor \lg n \rfloor$ lies between two well-defined extremes. Each extreme is a discretized version of a simple continuous distribution function. That is, each extreme is a staircase that rises at the integer points to agree with the corresponding continuous distribution function. The distribution function of $H_{20} - \lfloor \lg 20 \rfloor$ is sketched in Figure 2 and illustrates the relationship between a "typical" distribution and the two extreme continuous distribution functions described in Section 1.2.

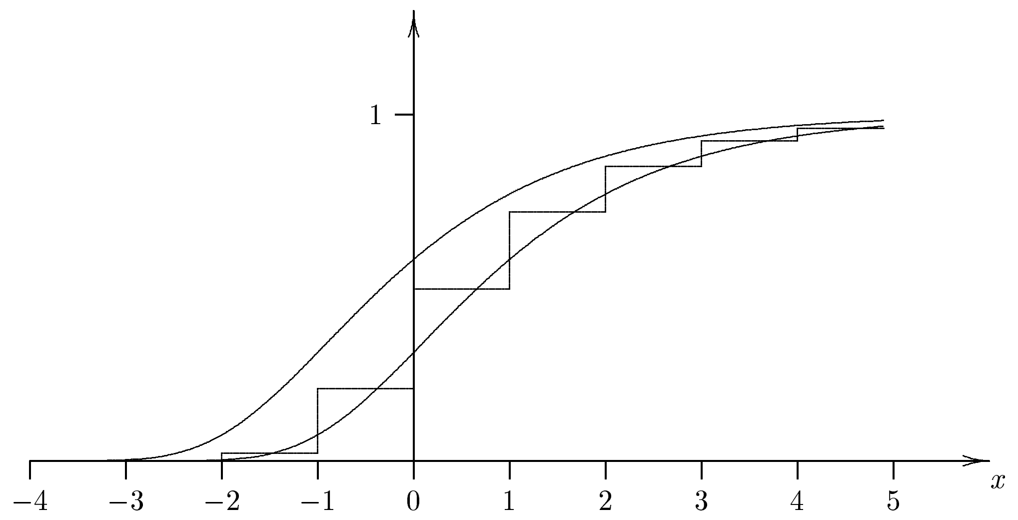

Figure 2. The distribution function of $H_{20} - \lfloor \lg 20 \rfloor$ and the two continuous extremes.

Observe that the asymptotic distribution [Theorem 2(i)] can be expanded as a Taylor series involving doubly exponential terms which resemble the extreme value distribution (i.e., $e^{-e^{-x}}$) that often appears as the limiting distribution of the maximum of n *continuous* i.i.d. random variables [see Galambos (1987)]. However, Anderson (1970) observed that such a limiting distribution may not exist for n *discrete* i.i.d. random variables; see also D'Aristotile et al. (1988), Section 5C. That this phenomenon occurs for the height of an incomplete trie should not be surprising, since the height H_n can be represented as an extreme statistic [cf. Jacquet and Szpankowski (1991) and Pittel (1986) for analogous connections between order statistics and characteristics of digital tries]. To see this, let L_k be the number of candidate winners after k rounds of elimination. Further, suppose that the leaves are numbered $1, 2, \ldots, K$ from right to left. Observe that K, the number of leaves, is a random variable. Let C_j be the length of the path from the root to the lowest common ancestor (that is, the common ancestor farthest from the root) of the jth leaf and the leaf containing the winner. Clearly, $0 \le C_1 \le C_2 \le \ldots \le C_K$, and

$$H_n = 1 + \max\{j \ge 0 : L_j > 1\} = 1 + \max_{1 \le j < K}\{C_j\} = 1 + C_{K-1}.$$

We can relate the above to properties of the standard trie (in the context of identification by coin tossing, even those who tossed heads continue the process until one player is left in each leaf). In such a standard digital tree, let $\tilde{C}_{ij}$ be the length of a path from the root to the lowest common ancestor of the ith and jth leaves. We denote by $\tilde{D}_n(i)$ the length of a path from the root to the ith leaf, and by $\tilde{H}_n$ the longest root-to-leaf path in such a tree. It is an easy exercise [Jacquet and Szpankowski (1991)] to see that

$$\tilde{D}_n(i) = 1 + \max_{\substack{1 \le j \le n \\ j \ne i}}\{\tilde{C}_{ij}\},$$

and

$$\tilde{H}_n = 1 + \max_{1 \le i < j \le n}\{\tilde{C}_{ij}\}.$$

Observe that each $\tilde{C}_{ij}$ is geometrically distributed.

Had $\tilde{C}_{ij}$ been i.i.d. random variables, then from standard extreme distribution theory we could immediately conclude that there exists a sequence $\tilde{a}_n$ of order $\log n$ such that $\mathbf{Prob}\{\tilde{H}_n - \tilde{a}_n \le k\}$ oscillates between $e^{-2^{-(k-1)}}$ and $e^{-2^{-k}}$. However, $\tilde{C}_{ij}$ are *not inde-*

pendent. Nevertheless $\tilde{D}_n(i)$ and $\tilde{H}_n$ still exhibit this sort of behavior[2] [cf. Jacquet and Szpankowski (1991), Pittel (1986)].

3 A Probabilistic Approach

In this section we prove all the exact results of Theorem 1 and derive from Theorem 1(i) the asymptotics of Theorem 2(i). The arguments of this section have some resemblance to those used in the analysis of probabilistic counting [Flajolet and Martin (1985)].

3.1 Exact Distribution and Moments by a Probabilistic Argument

We begin by proving Theorem 1(i). The second equality in (1) follows from a standard identity [e.g., Knuth (1973a), Exercise 1.2.11.2-4]:

$$n\sum_{j=0}^{J-1} j^{n-1} = \mathcal{B}_n(J) - B_n = \sum_{j=0}^{n-1}\binom{n}{j}B_j J^{n-j},$$

where $\mathcal{B}_n$ denotes the nth Bernoulli polynomial. We shall prove the first equality in an alternative form. Let $[k] := \{1,\ldots,k\}$. We first note the identity

$$\frac{1}{2^{kn}}\sum_{j=0}^{2^k-1} j^{n-1} = \frac{1}{2^k}\sum_{S\subseteq[k]}\left(1-\frac{1}{2^k}-\sum_{s\in S}\frac{1}{2^s}\right)^{n-1}.$$

Indeed, by reversing the order of summation we establish the second equality in

$$\frac{1}{2^{kn}}\sum_{j=0}^{2^k-1} j^{n-1} = \frac{1}{2^k}\sum_{j=0}^{2^k-1}\left(\frac{j}{2^k}\right)^{n-1} = \frac{1}{2^k}\sum_{j=0}^{2^k-1}\left(1-\frac{1}{2^k}-\frac{j}{2^k}\right)^{n-1}.$$

Finally, by considering binary expansions, as S ranges over the subsets of $[k]$ the expression $\sum_{s\in S}2^{-s}$ranges over the dyadic rationals of rank k in $[0,1)$. So we shall complete the proof of Theorem 1(i) by showing

$$\mathbf{Prob}\{H_n\le k\} = \frac{n}{2^k}\sum_{S\subseteq[k]}\left(1-\frac{1}{2^k}-\sum_{s\in S}\frac{1}{2^s}\right)^{n-1}.$$

Arbitrarily identify the players with labels 1 through n. It is sufficient to show that the probability that player 1 is chosen as the winner by the completion of the kth round equals

$$\frac{1}{2^k}\sum_{S\subseteq[k]}\left(1-\frac{1}{2^k}-\sum_{s\in S}\frac{1}{2^s}\right)^{n-1}.$$

[2] The behavior of the depth is not robust: in the case of biased coins the centered depth has a *Gaussian* limiting distribution!

We may imagine that every player continues tossing coins through the kth round, even if that player has been eliminated or declared the winner in an earlier round. Then it is enough to show that, through k rounds of play, the conditional probability that player 1 wins, given that the set of rounds on which player 1 flips heads is precisely S, equals $\left(1 - 2^{-k} - \sum_{s \in S} 2^{-s}\right)^{n-1}$. Since the various players' flips are mutually independent, the following lemma is sufficient.

Lemma 1. *The conditional probability that player 2 is eliminated in the course of the first k rounds, given that player 1 flips heads on rounds $s \in S$ and tails on rounds $s \in [k] - S$, equals*

$$1 - 2^{-k} - \sum_{s \in S} 2^{-s}.$$

Proof. All probabilities in this proof are computed conditionally given that player 1 flips heads on rounds $s \in S$ and tails on rounds $s \in [k]-S$. Write $S = \{s_1, \ldots, s_r\}$ with $0 \le r \le k$ and $s_1 < \cdots < s_r$. For convenience, set $s_0 := 0$ and $s_{r+1} := k+1$. Let E denote the event that player 2 is eliminated in the course of the first k rounds. For $i = 0, \ldots, r$ let E_i denote the event that the round-by-round flips for player 2 agree with those of player 1 through round s_i but not through round $s_{i+1} - 1$. Then it is easy to see that E is the disjoint union of $E_0, \ldots, E_r$, and that

$$\mathbf{Prob}(E_i) = 2^{-s_i} - 2^{-(s_{i+1}-1)}.$$

It follows that

$$\begin{aligned} \mathbf{Prob}(E) &= \sum_{i=0}^{r} \left[2^{-s_i} - 2^{-(s_{i+1}-1)}\right] = \sum_{i=0}^{r} 2^{-s_i} - 2\sum_{i=1}^{r+1} 2^{-s_i} \\ &= 2^{-s_0} - 2^{-(s_{r+1}-1)} - \sum_{i=1}^{r} 2^{-s_i} = 1 - 2^{-k} - \sum_{s \in S} 2^{-s}. \quad \blacksquare \end{aligned}$$

Remark 2. Suppose that our standard fair-coin game is modified by specifying that the probability of heads is $p \in (0,1)$ and the probability of tails is $q = 1-p$; thus when $p \neq q$ we introduce bias. The argument used to prove Theorem 1(i) leads to the following extension to incomplete binary tries arising from flipping a biased coin.

Proposition 1. *Consider a biased incomplete binary trie, as discussed above. Then for any integers $n \ge 1$ and $k \ge 0$,*

$$\mathbf{Prob}\{H_n \le k\} = n \sum_{r=0}^{k} p^r q^{k-r} \sum_{1 \le s_1 < \cdots < s_r \le k} \left[1 - p^r q^{k-r} - \sum_{j=1}^{r} p^{j-1} q^{s_j - (j-1)}\right]^{n-1}.$$

The asymptotic behavior of H_n in the biased-coin case is the subject of ongoing research.

The probability mass function and (factorial) moments of H_n are easily calculated from the exact distribution function in Theorem 1(i). The result is stated in the next lemma, from which parts (ii) and (iii) of Theorem 1 follow immediately.

Lemma 2. *Consider an incomplete binary trie with $n \geq 2$. Then:*

(i) *We have* $\mathbf{Prob}\{H_n = 0\} = 0$ *and*

$$\mathbf{Prob}\{H_n = k\} = -\sum_{j=1}^{n-1} \binom{n}{j} \frac{2^j - 1}{2^{kj}} B_j$$

for $k \geq 1$.

(ii) *For any integer $r \geq 1$, the rth factorial moment of H_n is given by*

$$\mathbf{E}[(H_n)_r] := \mathbf{E}\left[H_n(H_n-1)\cdots(H_n-(r-1))\right] = -r! \sum_{j=1}^{n-1} \binom{n}{j} \frac{2^j}{(2^j-1)^r} B_j.$$

3.2 Derivation of the Asymptotic Distribution from the Exact

In this section we use the rearrangement

$$\mathbf{Prob}\{H_n \leq k\} = \frac{n}{2^k} \sum_{j=1}^{2^k} \left(1 - \frac{j}{2^k}\right)^{n-1}, \quad k = 0, 1, \ldots, \tag{4}$$

of the first equality of (1) to prove the following quantified improvement to Theorem 2(i).

Proposition 2. *For any integers $n \geq 1$ and $-\infty < k < \infty$,*

$$-\frac{9}{\sqrt{n}} \leq \mathbf{Prob}\{H_n \leq \lfloor \lg n \rfloor + k\} - \frac{2^{\alpha(n)-k}}{\exp\left(2^{\alpha(n)-k}\right) - 1} \leq \frac{8}{n}. \tag{5}$$

The proof of Proposition 2 will make use of the following two calculus facts:

Lemma 3. *Given $c \geq 2/3$, define*

$$f(x) := \frac{x^2 e^x}{(e^{cx}-1)(e^x-1)}, \qquad x > 0.$$

Then $f(x) < 8$ for all $x > 0$.

Proof. For $x \le 1$ we have

$$f(x) \le \frac{x^2 e^x}{(cx)x} = \frac{e^x}{c} \le \frac{e}{c} \le \frac{3e}{2} < 8.$$

For $x \ge 1$ we have

$$f(x) = \frac{x^2}{(e^{cx}-1)(1-e^{-x})} \le \frac{x^2}{\left(\frac{1}{2}c^2x^2\right)(1-e^{-1})} = \frac{2e}{e-1}c^{-2} \le \frac{9e}{2(e-1)} < 8. \quad \blacksquare$$

Lemma 4. *For all $0 \le x \le c \le 3/5$,*

$$1 - x \ge e^{-(1+c)x}.$$

Proof. It is sufficient to show that

$$g(x) := 1 - x - e^{-x(1+x)}, \qquad x \in (0,1),$$

is non-negative for $x \in (0, 3/5)$. Indeed, it is straightforward to show that g is unimodal (i.e., g' switches sign once, from positive to negative) on $(0,1)$, and $g(3/5) > 0.01 > 0$. $\blacksquare$

Proof *of Proposition 2.* Since the result is easily verified for $n = 1, 2, 3$, we may suppose $n \ge 4$ throughout.

Upper bound: To prove the upper bound, we may also suppose that $k \ge -\lfloor \lg n \rfloor$. In that case we may apply (4):

$$\begin{aligned}
&\mathbf{Prob}\{H_n \le \lfloor \lg n \rfloor + k\} = \mathbf{Prob}\{H_n \le \lg n - \alpha(n) + k\} \\
&= \; 2^{\alpha(n)-k} \sum_{j=1}^{n \times 2^{k-\alpha(n)}} \left(1 - \frac{j}{n \times 2^{k-\alpha(n)}}\right)^{n-1} \\
&\le \; 2^{\alpha(n)-k} \sum_{j=1}^{n \times 2^{k-\alpha(n)}} \exp\left(-\frac{j(n-1)}{n} \times 2^{\alpha(n)-k}\right) \\
&\le \; 2^{\alpha(n)-k} \sum_{j=1}^{\infty} \exp\left(-\frac{j(n-1)}{n} \times 2^{\alpha(n)-k}\right) = \frac{2^{\alpha(n)-k}}{\exp\left(2^{\alpha(n)-k} \times \frac{n-1}{n}\right) - 1} \\
&= \; 2^{\alpha(n)-k} \left[\frac{1}{\exp\left(2^{\alpha(n)-k}\right) - 1} + \frac{1}{\exp\left(2^{\alpha(n)-k} \times \frac{n-1}{n}\right) - 1} - \frac{1}{\exp\left(2^{\alpha(n)-k}\right) - 1}\right] \\
&= \; 2^{\alpha(n)-k} \left[\frac{1}{\exp\left(2^{\alpha(n)-k}\right) - 1} + \frac{\exp\left(2^{\alpha(n)-k}\right) - \exp\left(2^{\alpha(n)-k} \times \frac{n-1}{n}\right)}{\left[\exp\left(2^{\alpha(n)-k} \times \frac{n-1}{n}\right) - 1\right]\left[\exp\left(2^{\alpha(n)-k}\right) - 1\right]}\right] \\
&\le \; 2^{\alpha(n)-k} \left[\frac{1}{\exp\left(2^{\alpha(n)-k}\right) - 1} + \frac{n^{-1} 2^{\alpha(n)-k} \exp\left(2^{\alpha(n)-k}\right)}{\left[\exp\left(2^{\alpha(n)-k} \times \frac{n-1}{n}\right) - 1\right]\left[\exp\left(2^{\alpha(n)-k}\right) - 1\right]}\right],
\end{aligned}$$

where for the final inequality we have employed the mean value theorem. The desired upper bound now follows from Lemma 3.

Lower bound: First, if $k < \alpha(n) + 1 - \frac{1}{2}\lg n$, then

$$\frac{2^{\alpha(n)-k}}{\exp\left(2^{\alpha(n)-k}\right) - 1} \le \frac{\frac{1}{2}\sqrt{n}}{\exp\left(\frac{1}{2}\sqrt{n}\right) - 1} \le \frac{\frac{1}{2}\sqrt{n}}{\frac{1}{2}\sqrt{n} + \frac{1}{8}n} \le \frac{4}{\sqrt{n}} < \frac{9}{\sqrt{n}},$$

whence the lower bound in (5) holds trivially. So we may assume $k \ge \alpha(n) + 1 - \frac{1}{2}\lg n \ge -\lg n$. Then we may again apply (4):

$$\begin{aligned}
&\mathbf{Prob}\{H_n \le \lfloor \lg n \rfloor + k\} \\
&= \quad 2^{\alpha(n)-k} \sum_{j=1}^{n \times 2^{k-\alpha(n)}} \left(1 - \frac{j}{n \times 2^{k-\alpha(n)}}\right)^{n-1} \\
&\ge \quad 2^{\alpha(n)-k} \sum_{j=1}^{\lfloor \sqrt{n} \times 2^{k-\alpha(n)} \rfloor} \left(1 - \frac{j}{n \times 2^{k-\alpha(n)}}\right)^{n} \\
&\ge \quad 2^{\alpha(n)-k} \sum_{j=1}^{\lfloor \sqrt{n} \times 2^{k-\alpha(n)} \rfloor} \exp\left[-j\left(1 + n^{-1/2}\right) \times 2^{\alpha(n)-k}\right] \qquad \text{by Lemma 4} \\
&= \quad 2^{\alpha(n)-k} \times \frac{1 - \exp\left[-\left(1 + n^{-1/2}\right)\left\lfloor \sqrt{n} \times 2^{k-\alpha(n)} \right\rfloor \times 2^{\alpha(n)-k}\right]}{\exp\left[\left(1 + n^{-1/2}\right) \times 2^{\alpha(n)-k}\right] - 1}.
\end{aligned}$$

The exponential term appearing in the numerator of the last fraction is bounded above by

$$\exp\left[-\left(\sqrt{n} \times 2^{k-\alpha(n)} - 1\right) 2^{\alpha(n)-k}\right] = \exp\left[-\sqrt{n} + 2^{\alpha(n)-k}\right] \le \exp\left(-\frac{1}{2}\sqrt{n}\right) \le \frac{2}{e\sqrt{n}},$$

and, proceeding as for the upper bound,

$$\begin{aligned}
&\frac{1}{\exp\left[2^{\alpha(n)-k}\left(1 + n^{-1/2}\right)\right] - 1} \\
&\ge \quad \frac{1}{\exp\left(2^{\alpha(n)-k}\right) - 1} - \frac{n^{-1/2} 2^{\alpha(n)-k} \exp\left[2^{\alpha(n)-k}\left(1 + n^{-1/2}\right)\right]}{\left[\exp\left(2^{\alpha(n)-k}\right) - 1\right]\left[\exp\left[2^{\alpha(n)-k}\left(1 + n^{-1/2}\right)\right] - 1\right]} \\
&\ge \quad \frac{1}{\exp\left(2^{\alpha(n)-k}\right) - 1} - \frac{8 \times 2^{k-\alpha(n)}}{\sqrt{n}} \qquad \text{by Lemma 3.}
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbf{Prob}\{H_n \le \lfloor \lg n \rfloor + k\} \quad &\ge \quad \left(1 - \frac{2}{e\sqrt{n}}\right)\left[\frac{2^{\alpha(n)-k}}{\exp\left(2^{\alpha(n)-k}\right) - 1} - \frac{8}{\sqrt{n}}\right] \\
&\ge \quad \frac{2^{\alpha(n)-k}}{\exp\left(2^{\alpha(n)-k}\right) - 1} - \frac{9}{\sqrt{n}}. \quad \blacksquare
\end{aligned}$$

4 Distribution by an Analytic Approach

We apply an analytic approach based on Poissonization and de-Poissonization to rederive exact and asymptotic distributions for H_n. One of the earliest applications of this idea can be traced back to the work of Kac (1949). Over the past few decades the idea has resurfaced in various guises, and the difficult step has always been in carrying results from the Poisson model back to the fixed-population model, the model of prime interest. A variety of ad hoc solutions have been proposed. Perhaps the earliest in digital problems is due to Jacquet and Régnier (1986). Poissonization and de-Poissonization have also been considered by some as a mathematical transform followed by an inversion operation. The first to treat it as such were Gonnet and Munro (1984) [Poblete (1987) follows up on the inversion of the transform]. Other useful related results may be found in Holst (1986), Aldous (1989), Rais et al. (1993), Arratia and Tavaré (1994), and Jacquet and Szpankowski (1995).

Poissonization is carried out as follows. Instead of having a population of fixed size, we first determine the number of players participating in the elimination contest by a draw from a Poisson distribution with parameter w. For subsequent de-Poissonization, we even formally allow w to be a complex number in order to manipulate the resulting generating function by considering its analytic continuation to the w complex plane. The de-Poissonization process is then achieved by a combination of Mellin transform methods and asymptotic integration over a circle of radius n in the w complex plane. The general idea in Poissonization is that the behavior of a fixed-population problem should be close to that of the same problem under a Poisson model having the fixed-population problem size as its mean. The rationale behind the mechanics of the approximation is discussed following the De-Poissonization Lemma.

The Mellin transform [see, e.g., Mahmoud (1992), Section 1.5 for an introduction or Flajolet et al. (1995) for a survey] plays a prominent role in this process because the functional equations for the probability generating functions in digital problems under the Piosson model usually have a harmonic nature. That is, when solved by iteration the solution is a harmonic sum of the general form

$$\sum_{k=0}^{\infty} \lambda_k f(a_k x). \tag{6}$$

Such a harmonic function has a simple Mellin transform. The formula for the transform f^* of a function f is

$$f^*(s) = \int_0^{\infty} f(x) x^{s-1}\, dx.$$

Applied to a harmonic sum of a base function f with different scales as in (6), the Mellin transform yields

$$f^*(s) \sum_{k=0}^{\infty} \lambda_k a_k^{-s}.$$

With the proper choice of a fundamental strip (usually a vertical strip of finite width or a semi-infinite strip) for the complex variable s, convergence of the series is achieved as well as the existence of $f^*(s)$. One then resorts to contour integration to evaluate the inversion integral

$$f(x) = \frac{1}{2\pi i} \int x^{-s} \Big(f^*(s) \sum_{k=0}^{\infty} \lambda_k a_k^{-s} \Big) \, ds,$$

which is taken over an infinite vertical line in the fundamental strip. That is, the integral is really done by a residue calculation and an asymptotic solution is recovered for the functional equation of the probability generating function under the Poisson model.

4.1 Asymptotic Distribution by Poissonization and De-Poissonization

For each $n \geq 0$, let

$$G_n(z) := \sum_{k=0}^{\infty} \mathbf{Prob}\{H_n = k\} z^k$$

denote the probability generating function for the fixed-population height H_n. (We use the conventions $H_1 = 0$ and, since it is not possible to choose a winner from 0 players, $H_0 = \infty$.) The starting point of our analysis is Prodinger's equation [Prodinger (1993)]:

$$G_n(z) = \frac{z}{2^n} \sum_{k=1}^{n} \binom{n}{k} G_k(z) + \frac{z G_n(z)}{2^n}, \tag{7}$$

which is valid for $n \geq 2$, with the boundary values $G_1(z) = 1$ and $G_0(z) = 0$. [Equation (7) is easily derived by conditioning on the number of tails tossed in the first round.] This equation can be used to extract the exact distribution of H_n, as will be sketched later on in this subsection.

We shall call the tree constructed under Poissonization the *Poissonized incomplete trie*. We shall also refer to its properties as Poissonized; in particular we shall call the height of the Poissonized incomplete trie the *Poissonized height* and denote it by $H_{N(w)}$ or simply by H_N, where $N \equiv N(w)$, the Poissonized number of participants, has the Poisson distribution with mean w. Introduce the generating function

$$g(w, z) := \sum_{n=0}^{\infty} \frac{G_n(z) e^{-w} w^n}{n!}.$$

This bivariate generating function has the following interpretation as a probability generating function for the Poissonized height:

$$\begin{aligned}
\mathbf{E}[z^{H_N}] &= \sum_{n=0}^{\infty} \mathbf{E}\left[z^{H_N} | N = n\right] \frac{e^{-w} w^n}{n!} \\
&= \sum_{n=0}^{\infty} \mathbf{E}\left[z^{H_n}\right] \frac{e^{-w} w^n}{n!} \\
&= g(w, z).
\end{aligned}$$

Multiplying both sides of (7) by $e^{-w}w^n/n!$, summing over the range of validity of the recurrence (i.e., $n \geq 2$), and adjusting for the boundary cases $n = 0$ and $n = 1$, we obtain

$$g(w, z) = z(1 + e^{-w/2}) g\left(\frac{w}{2}, z\right) + e^{-w} w(1 - z). \tag{8}$$

To handle this latter recurrence, introduce

$$h(w, z) = \frac{g(w, z)}{1 - e^{-w}},$$

which transforms the recurrence into

$$h(w, z) = zh\left(\frac{w}{2}, z\right) + R(w, z), \tag{9}$$

where

$$R(w, z) := \frac{w(1 - z)}{e^w - 1}.$$

The recurrence (9) can now be solved by direct iteration [see Knuth (1973b) or Szpankowski (1987) for a general solution of this type of equations]. This gives

$$h(w, z) = \sum_{k=0}^{\infty} R\left(\frac{w}{2^k}, z\right) z^k + \lim_{k \to \infty} \left[h\left(\frac{w}{2^k}, z\right) z^k\right]. \tag{10}$$

It is not hard to show that the limit vanishes if we assume that z belongs to the open disc of radius $1/2$ centered at the origin in the z complex plane. Thus, in the assumed domain

$$h(w, z) = \sum_{k=0}^{\infty} R\left(\frac{w}{2^k}, z\right) z^k. \tag{11}$$

The sum in (11) is a harmonic sum and so is well suited to the application of Mellin transform methods [see Flajolet et al. (1995) or Mahmoud (1992) for background]. We denote the Mellin transform of a function $f(w, z)$ with respect to w by $f^*(s, z)$, that is,

$$f^*(s, z) := \int_0^{\infty} f(w, z) w^{s-1} \, dw.$$

Formally,

$$h^*(s, z) = R^*(s, z) \sum_{k=0}^{\infty} (2^s z)^k.$$

This transform calculation is valid provided:

(i) The sum $\sum_{k=0}^{\infty}(2^s z)^k$ converges, which happens if $\Re s < -\lg|z|$.

(ii) The transform $R^*(s,z)$ exists. Using the geometric series representation

$$\frac{1}{e^w - 1} = \sum_{k=1}^{\infty} e^{-wk},$$

and the harmonic sum formula for the Mellin transform [Flajolet et al. (1995)], we find after some simple algebra

$$R^*(s,z) = (1-z)\,\Gamma(s+1)\,\zeta(s+1),$$

provided $\Re s > 1$. This calculation uses the classical representations of the gamma function as an integral and of the zeta function as a sum, both valid over the stated domain.

So the Mellin transform $h^*(s,z)$ exists in the following domain of the s complex plane:

$$1 < \Re s < -\lg|z|.$$

Observe that this fundamental strip is non-empty for $|z| < 1/2$, which is also sufficient to annihilate the limit in (10). Within this fundamental strip, the transform is given by

$$h^*(s,z) = \frac{(1-z)\,\Gamma(s+1)\,\zeta(s+1)}{1-2^s z}.$$

Since we are concerned here with the distribution function of H_N rather than its probability mass function, our interest centers on the transform

$$\frac{h^*(s,z)}{1-z} = \frac{\Gamma(s+1)\,\zeta(s+1)}{1-2^s z}.$$

This expression can be expanded immediately in a power series in z; the coefficient of z^k is

$$2^{ks}\Gamma(s+1)\,\zeta(s+1).$$

Inverting the transform gives

$$\frac{h(w,z)}{1-z} = \sum_{k=0}^{\infty}\left[z^k \times \frac{w/2^k}{\exp(w/2^k) - 1}\right]$$

for $w > 0$. But

$$\sum_{k=0}^{\infty} \mathbf{Prob}\{H_{N(w)} \le k\} z^k = \frac{g(w,z)}{1-z} = \frac{(1-e^{-w})\,h(w,z)}{1-z},$$

so for each $k = 0, 1, 2, \ldots$ and $w > 0$ we have

$$\mathbf{Prob}\{H_{N(w)} \le k\} = (1 - e^{-w}) \frac{w/2^k}{\exp(w/2^k) - 1}. \tag{12}$$

The relation (12) can be manipulated to prove Theorem 1(i). In fact, this is how we originally discovered the exact distribution. Once it became known to us, we went back and devised the direct argument of Section 3.1. Here is the derivation. For any $w > 0$,

$$\begin{aligned}
\mathbf{Prob}\{H_{N(w)} \le k\} &= e^{-w} \frac{w}{2^k} \times \frac{e^w - 1}{\exp(w/2^k) - 1} \\
&= e^{-w} \frac{w}{2^k} \sum_{j=0}^{2^k - 1} \exp(jw/2^k) \\
&= e^{-w} \times \frac{1}{2^k} \sum_{j=0}^{2^k-1} \sum_{n=1}^{\infty} \frac{(j/2^k)^{n-1} w^n}{(n-1)!} \\
&= e^{-w} \sum_{n=1}^{\infty} \frac{w^n}{n!} \left(\frac{n}{2^{kn}} \sum_{j=0}^{2^k-1} j^{n-1} \right),
\end{aligned}$$

from which the first equality in (1) is evident.

We can also use (12) to derive asymptotics, as follows. We use the De-Poissonization Lemma of Rais et al. (1993) to derive an asymptotic expression for the fixed-population probabilities

$$p_{k,n} := \mathbf{Prob}\{H_n \le k\} \tag{13}$$

from an asymptotic development for the Poissonized probabilities

$$P_k(w) := (1 - e^{-w}) \frac{w/2^k}{\exp(w/2^k) - 1}. \tag{14}$$

Although up until now the right side of (14) has arisen only as $\mathbf{Prob}\{H_{N(w)} \le k\}$ for real $w \ge 0$, note that it defines an entire function of the complex variable w. (The singularity at the origin is removable.) We next state the De-Poissonization Lemma of Rais et al. (1993) in a slightly altered form (the proof remains unchanged).

Lemma 5. *Consider an arbitrary collection of sequences $(p_{k,n})_{n \ge 0}$, $k \in K$, and suppose that each Poisson transform*

$$P_k(w) := \sum_{n=0}^{\infty} p_{k,n} \frac{e^{-w} w^n}{n!}$$

can be analytically continued as an entire function of complex w. Fix $\theta \in (0, \pi/2)$ and let S_θ be the cone $\{w : |\arg w| \le \theta\}$. Suppose that there exist constants $a < 1$, c, β_1, β_2, and w_0 such that the following conditions hold for every $k \in K$:

(i) *For all* $w \in S_\theta$ *with* $|w| \geq w_0$,

$$|P_k(w)| \leq \beta_1 |w|^c;$$

(ii) *for all* $w \notin S_\theta$ *with* $|w| \geq w_0$,

$$|P_k(w) e^w| \leq \beta_2 |w|^c e^{a|w|}.$$

Then for large n, *uniformly in* $k \in K$,

$$p_{k,n} = P_k(n) + O\left(n^{c-\frac{1}{2}}\right).$$

Remark 3. The De-Poissonization Lemma says that if conditions (i) and (ii) are satisfied with $c < 1/2$, the fixed-population probabilities are nearly the same as the Poissonized probabilities with n replacing w.

The rough idea behind de-Poissonization is the following. For any fixed n, the coefficient of w^n in the power series expansion of $e^w P_k(w)$ about the origin is $p_{k,n}/n!$. Finding $p_{k,n}$ is then only a matter of extracting a coefficient from a generating function. This is routinely done by considering a contour integral around the origin in the w complex plane. In particular, when we choose the contour to be the circle $|w| = n$, we get

$$p_{k,n} = \frac{n!}{2\pi i} \oint \frac{e^w P_k(w)}{w^{n+1}}\, dw.$$

Such integrals typically have most of their value contributed by a small arc at the intersection of the circle with the positive real line (i.e., at $w = n$); the rest of the contour adds only an ignorable correction. Over this small arc, the value of the function $P_k(w)$ is well approximated by $P_k(n)$. Taking this now-constant factor outside the integral, performing the remaining elementary integration and applying Stirling's approximation to $n!$, we see that all factors cancel out, except $P_k(n)$. The point is that $p_{k,n}$ can be accurately approximated by $P_k(n)$, with a diminishing error as $n \to \infty$. The De-Poissonization Lemma gives sufficient conditions to make this approximation valid.

In order to prove Theorem 2(i), we have two tasks remaining: we must verify that the two conditions of the De-Poissonization Lemma are satisfied by (13) and (14), and we must apply the lemma. Neither task is difficult. We first show that the conditions are met for any $\theta \in (0, \pi/2)$ and $w_0 > 0$ by taking

$$c = 0, \quad a = \cos\theta, \quad \beta_1 = 2\sec\theta, \quad \text{and} \quad \beta_2 = \sec\theta.$$

To verify condition (i), we first observe that if $0 \neq v \in S_\theta$, then $|v| \leq \Re v / \cos\theta$ and

$$|e^v - 1| \geq |e^v| - 1 = \exp(\Re v) - 1,$$

so that

$$\left|\frac{v}{e^v - 1}\right| \leq (\sec\theta) \frac{\Re v}{\exp(\Re v) - 1} \leq \sec\theta.$$

Thus for $0 \neq w \in S_\theta$ we have

$$\begin{aligned}
|P_k(w)| &\leq |1 - e^{-w}| \left|\frac{w/2^k}{\exp(w/2^k) - 1}\right| \\
&\leq (\sec\theta) \left|1 - e^{-w}\right| \\
&\leq (\sec\theta) \left(1 + \left|e^{-w}\right|\right) \\
&\leq (\sec\theta) \left(1 + \exp(-\Re w)\right) \\
&\leq 2\sec\theta.
\end{aligned}$$

To verify condition (ii), we first observe that if $v \notin S_\theta$, then $\Re v \leq |v| \cos\theta$. Thus for $0 \neq w \notin S_\theta$, we have [from (14)]

$$\begin{aligned}
|P_k(w)e^w| &\leq \left|\frac{w}{2^k} \times \frac{e^w - 1}{\exp(w/2^k) - 1}\right| \\
&\leq |w| \times 2^{-k} \sum_{r=0}^{2^k - 1} \left|\exp(rw/2^k)\right| \\
&\leq |w| \times 2^{-k} \sum_{r=0}^{2^k - 1} \exp\left(r 2^{-k} |w| \cos\theta\right) \\
&\leq |w| \int_0^1 \exp(x|w|\cos\theta)\, dx \\
&= (\sec\theta) \left[\exp(|w|\cos\theta) - 1\right] \\
&\leq (\sec\theta) \exp(|w|\cos\theta).
\end{aligned}$$

Finally, we apply the De-Poissonization Lemma to conclude that

$$\mathbf{Prob}\{H_n \leq j\} = p_{j,n} = P_j(n) + O\left(n^{-1/2}\right) \tag{15}$$

holds uniformly in integers $j \geq 0$. But [from (14)]

$$P_j(n) = (1 - e^{-n}) \frac{n/2^j}{\exp(n/2^j) - 1}.$$

Setting $j = \lfloor \lg n \rfloor + k$, we find that

$$P_j(n) = (1 - e^{-n}) \frac{2^{\alpha(n)-k}}{\exp\left(2^{\alpha(n)-k}\right) - 1} = \frac{2^{\alpha(n)-k}}{\exp\left(2^{\alpha(n)-k}\right) - 1} + O\left(e^{-n}\right) \tag{16}$$

holds, uniformly in all integers k ($\geq -\lfloor \lg n \rfloor$). Combining (15) and (16) completes the proof of Theorem 2(i).

Remark 4. We wish to stress the broad applicability of the Poissonization and de-Poissonization approach. The method has been successfully applied to problems where exact results such as our Theorem 1 are unobtainable: see, e.g., Jacquet and Régnier (1986), Rais et al. (1993), and Jacquet and Szpankowski (1995).

4.2 Moments by Poissonization and de-Poissonization

We complete the study of the height in a random incomplete trie by studying its mean and variance. As for the distribution, the strategy is first to compute the mean and variance of the Poissonized height H_N and then to de-Poissonize. The tool here will be a de-Poissonization lemma from Jacquet and Szpankowski (1989). Our presentation will be somewhat informal, as rigorous verification of the conditions required for careful application of this lemma is straightforward but laborious. [One can apply the so-called "increasing domains approach" of Jacquet and Szpankowski (1995) directly to the functional equations (17)–(18) below.] Use of an analytical cousin—Rice's method—of the Poissonization and de-Poissonization method to rigorously establish Theorem 2(ii) and (iii) will be discussed briefly in Section 4.3. Unlike the approach in this subsection, Rice's method requires exact formulas for the moments.

Proof *of Theorem 2 (ii) and (iii).* Let

$$A(w) := \mathbf{E}[H_N; H_N < \infty], \quad M(w) := \mathbf{E}[H_N(H_N - 1); H_N < \infty].$$

(Here the indicators for $H_N < \infty$ are needed only to properly handle the possibility $N = 0$.) ¿From our basic recurrence (8) for the Poisson model, after taking first and second derivatives with respect to z at $z = 1$, we obtain

$$\begin{aligned} A(w) &= A(w/2)(1 + e^{-w/2}) + 1 - e^{-w} - we^{-w}, &(17)\\ M(w) &= M(w/2)(1 + e^{-w/2}) + 2A(w/2)(1 + e^{-w/2}). &(18) \end{aligned}$$

As before, these equations are not yet suitable for direct iteration. Define $a(w) := A(w)/(1 - e^{-w})$ and $m(w) := M(w)/(1 - e^{-w})$. Then the above equations become

$$\begin{aligned} a(w) &= a(w/2) + 1 - \frac{w}{e^w - 1}, \\ m(w) &= m(w/2) + 2a(w/2). \end{aligned}$$

To solve these functional equations, we apply the Mellin transform to each side. It is easy to see that the Mellin transforms $a^*(s)$ and $m^*(s)$ of $a(w)$ and $m(w)$, respectively, exist in the strip $\Re s \in (-1, 0)$. After some algebra we find

$$\begin{aligned} a^*(s) &= \frac{\Gamma(s+1)\,\zeta(s+1)}{2^s - 1}, \\ m^*(s) &= -\frac{2^{s+1}\,\Gamma(s+1)\,\zeta(s+1)}{(2^s-1)^2}. \end{aligned}$$

In order to derive asymptotics for $a(w)$ and $m(w)$ for large w, we compute inverse Mellin transforms, using the residue theorem to shift the line of integration to the right. Observe also that as $w \to \infty$ in a cone $S_\theta = \{w : |\arg w| \le \theta\}$ with $0 < \theta < \pi/2$, we have $A(w) = a(w)\left(1 + O\left(e^{-a|w|}\right)\right)$ and $M(w) = m(w)\left(1 + O\left(e^{-a|w|}\right)\right)$ with $a = \cos\theta$. For any fixed $R > 0$ we arrive after considerable calculation (aided by MAPLE) at

$$\begin{aligned} A(w) &= \lg w + \frac{1}{2} - \delta_1(\lg w) + O(|w|^{-R}), \\ M(w) &= \lg^2 w + \frac{\pi^2}{6L^2} - \frac{1}{6} - \frac{2\gamma_1}{L^2} - \frac{\gamma^2}{L^2} + \tilde{\delta}_2(w) + O(|w|^{-R}), \end{aligned}$$

with $\delta_1(\cdot)$ and $\delta_2(\cdot)$ as given in Theorem 2 and Remark 1 and

$$\tilde{\delta}_2(x) := -2(\lg x)\delta_1(\lg x) + \delta_2(\lg x) + \delta_1^2(\lg x).$$

Finally, we must de-Poissonize the Poissonized first and second factorial moments to revert to the fixed-population model. In this case we use the result from Jacquet and Szpankowski (1989) to get immediately the fixed-population asymptotics. Define $V(w) := M(w) + A(w) - A^2(w)$, that is, the Poissonized variance. Then the fixed-population mean has the asymptotic expansion $\mathbf{E}[H_n] \sim A(n) - \frac{1}{2}nA''(n)$ and the fixed-population variance satisfies $\mathbf{Var}[H_n] \sim V(n) - n\,(A'(n))^2$ [cf. Régnier and Jacquet (1989) and Jacquet and Szpankowski (1995)]. We can show that $A'(n)$ is $O(1/n)$ and that $A''(n)$ is $O(1/n^2)$. For example, $A'(n) = O(1/n)$ can be established by differentiating (17), which gives a functional equation for $A'(w)$ as a recurrence involving $A'(w/2)$ and $A(w/2)$. This recurrence can then be solved by the Mellin transform and its inverse to yield the desired asymptotic bound. Consequently, the fixed-population moments are obtained from the Poissonized moments by adding to the latter moments corrections only of order $O(1/n)$. [An alternative approach using a bound on $A(w)$ and Cauchy's formula for $A'(w)$ follows Jacquet and Szpankowski (1995).] Completing the algebra, we obtain Theorem 2(ii) and (iii). ∎

4.3 Moments by Rice's Method

Recall from Lemma 2(ii) that for $n \geq 2$ and $r \geq 1$, the rth factorial moment $\mathbf{E}[(H_n)_r]$ of the height H_n is given by

$$\frac{\mathbf{E}[(H_n)_r]}{r!} = \sum_{j=1}^{n-1} \binom{n}{j} (-B_j) \frac{2^j}{(2^j - 1)^r}.$$

As is well known [see, e.g., Chapter 23 in Abramowitz and Stegun (1972)],

$$-B_j = (-1)^j j \zeta(1-j) \quad \text{for } j \geq 1.$$

Thus Rice's method [see, e.g., Section 6.4 in Mahmoud (1992) or Flajolet and Sedgewick (1995)] for handling alternating sums involving binomial coefficients is ideally suited to the asymptotic calculation of moments for the height. An application of the method leads to Theorem 2(ii) and (iii). This approach is outlined in Fill et al. (1996).

References

[1] Abramowitz, M. and Stegun, I., eds. (1972). *Handbook of Mathematical Functions.* Wiley, New York.

[2] Aldous, D. (1989). *Probability Approximations via the Poisson Clumping Heuristic.* Springer–Verlag, New York.

[3] Anderson, C. (1970). Extreme value theory for a class of discrete distributions with applications to some stochastic processes. *J. Appl. Probability*, vol. 7, pp. 99–113.

[4] Arratia, R. and Tavaré, S. (1994). Independent processes approximations for random combinatorial structures. *Advances in Math.*, vol. 104, pp. 90–154.

[5] Berndt, B. (1985). *Ramanujan's Notebook, Part I*. Springer–Verlag, New York.

[6] Brassard, G. and Bratley, P. (1988). *Algorithmics: Theory and Practice*. Prentice Hall, Englewood Cliffs, New Jersey.

[7] D'Aristotile, A., Diaconis, P., and Freedman, D. (1988). On merging of probabilities. *Sankhyā*, vol. 50, ser. A, pt. 3, pp. 363–380.

[8] Devroye, L. (1986). *Non-Uniform Random Variate Generation*. Springer–Verlag, New York.

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

[10] Fill, J. A., Mahmoud, H. M., and Szpankowski, W. (1996). On the distribution for the duration of a randomized leader election algorithm. Technical Report #544, Department of Mathematical Sciences, The Johns Hopkins University.

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

[12] Flajolet, P., Gourdon, X., and Dumas, P. (1995). Mellin transforms and asymptotics: Harmonic sums. *Theoretical Computer Science*, vol. 144, pp. 3–58.

[13] Flajolet, P. and Martin, G. (1985). Probabilistic counting algorithms for data base applications. *J. Computer and System Sciences*, vol. 31, pp. 182–209.

[14] Flajolet, P. and Sedgewick, R. (1995). Mellin transforms and asymptotics: Finite differences and Rice's integrals. *Theoretical Computer Science*, vol. 144, pp. 101–124.

[15] Galambos, H. (1987). *The Asymptotic Theory of Extreme Order Statistics*. Robert E. Krieger Publishing Company, Malabar, Florida.

[16] Gonnet, G. and Munro, J. (1984). The analysis of linear probing sort by the use of a new mathematical transform. *J. of Algorithms*, vol. 5, pp. 451–470.

[17] Grabner, P. (1993). Searching for losers. *Random Structures and Algorithms*, vol. 4, pp. 99–110.

[18] Holst, L. (1986). On birthday, collector's, occupancy and other classical urn problems. *International Stat. Rev.*, vol. 54, pp. 15–27.

[19] Jacquet, P. and Régnier, M. (1986). Trie partitioning process: limiting distributions. *Lecture Notes in Computer Science*, vol. 214, pp. 196–210, Springer, New York.

[20] Jacquet, P. and Szpankowski, W. (1989). Ultimate characterizations of the burst response of an interval searching algorithm. *SIAM J. Computing*, vol. 18, pp. 777–791.

[21] Jacquet, P. and Szpankowski, W. (1991). Analysis of tries with Markovian dependency. *IEEE Trans. Information Theory*, vol. 37, pp. 1470–1475.

[22] Jacquet, P. and Szpankowski, W. (1995). Asymptotic behavior of the Lempel-Ziv parsing scheme and digital search trees. *Theoretical Computer Science*, vol. 144, 161–197.

[23] Kac, M. (1949). On the deviations between theoretical and empirical distributions. *Proc. of the National Academy of Science, U. S. A.*, vol. 35, 252–257.

[24] Knuth, D. (1973a). *The Art of Computer Programming, Vol. 1: Fundamentals of Algorithms.* Addison–Wesley, Reading, Massachusetts.

[25] Knuth, D. (1973b). *The Art of Computer Programming, Vol. 3: Sorting and Searching*. Addison–Wesley, Reading, Massachusetts.

[26] Kuipers, L. and Niederreiter, H. (1974). *Uniform Distribution of Sequences*. Wiley, New York.

[27] Mahmoud, H. (1992). *Evolution of Random Search Trees*. Wiley, New York.

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

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

[30] Pittel, B. (1986). Paths in a random digital tree: Limiting distributions. *Adv. Appl. Probab.*, vol. 18, pp. 139–155.

[31] Pittel, B. and Rubin, H. (1992). How many random questions are necessary to identify n distinct objects. *J. Combinatorial Theory, Ser. A*, vol. 55, pp. 292–312.

[32] Poblete, P. (1987). Approximating functions by their Poisson transform. *Information Proc. Letters*, vol. 23, pp. 127–130.

[33] Prodinger, H. (1993). How to select a loser. *Discrete Math.*, vol. 120, pp. 149–159.

[34] Rais, B., Jacquet, P., and Szpankowski, W. (1993). Limiting distribution for the depth in PATRICIA tries. *SIAM J. on Discrete Math.*, vol. 3, pp. 355–362.

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

[36] Rényi, A. (1961). On random subsets of a finite set. *Mathematica (Cluj)*, vol. 3, pp. 355–362.

[37] Szpankowski, W. (1987). Solution to a linear recurrence equation arising in the analysis of some algorithms. *SIAM J. Alg. Disc. Meth.*, vol. 8, pp. 233–250.