ANALYSIS OF AN ASYMMETRIC LEADER ELECTION ALGORITHM

September 3, 1996

Svante Janson
Department of Mathematics
Uppsala University
75106 Uppsala
Sweden
svante.janson@math.uu.se

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

Abstract

We consider a leader election algorithm in which a set of distributed objects (people, computers, etc.) try to identify one object as their leader. The election process is randomized, that is, at every stage of the algorithm those objects that survived so far flip a *biased* coin, and those who received, say a tail, survive for the next round. The process continues until only one objects remains. Our interest is in evaluating the limiting distribution and the first two moments of the number of rounds needed to select a leader. We establish precise asymptotics for the first two moments, and show that the asymptotic expression for the duration of the algorithm exhibits some periodic fluctuations and consequently no limiting distribution exists. These results are proved by analytical techniques of the precise analysis of algorithms such as: analytical poissonization and depoissonization, Mellin transform, and complex analysis.

*The work of this author was supported by NSF Grants NCR-9206315 and NCR-9415491, and NATO Collaborative Grant CRG.950060. The second author also thanks INRIA, Sophia Antipolis, project MISTRAL for hospitality and support during the summer of 1996 when this paper was completed.

1 Introduction

Consider a group of n people (users, computers, objects, etc.) sharing a scarce resource (e.g., channel, CPU, etc.). The following elimination process can be used to find a "winner" or a "leader" that has undisputed and uncontested access to the resource (cf. [2, 5, 19]): All objects involved toss a *biased* coin, and all players to throw heads are losers while those who throw tails remain candidate winners and flip the coins again until a single winner (leader) is identified. If all players throw heads at any stage, the toss is inconclusive and all players participate again in the contest. How many tosses are needed to identify a winner? The problem was posed for a *fair* (unbiased) coin tossing process by Prodinger [19] (cf. also [10]), who provided the first non-trivial analysis. Recently, for the same fair coin model, Fill *et. al.* [5] find the limiting distribution for the number of rounds. In this paper, we analyze the same problem but when the coins involved are *biased*, that is, the probability p of throwing a head is not equal to one half ($p \neq \frac{1}{2}$). In passing, we should mention that such a randomized elimination algorithm has many applications, notably in electing a "leading" computer after a synchronization is lost in a distributed computer network (e.g., token lost in a token passing ring network). We also remark that a formula for the exact distribution has been given by Fill *et. al.* [5] for the fair model and by Fill [6] for the biased case.

The above elimination process can be represented as a incomplete trie (cf. [5, 18, 19]) in which only one side of the trie is developed while the other side is pruned (all those players who throw heads do not participate any more in the process). Therefore, the number of throws needed to find the winner (leader) is equivalent to the height in such a incomplete trie. Accordingly, we shall call the duration of the above elimination process as height, and we study asymptotics of its moments and the limiting distribution, if it exists.

Tries have been extensively analyzed in the past including the height. The reader is referred to Knuth [16] and Mahmoud [18] for updated account on recent developments in this area. In fact, tries and other digital trees were used as a testbed for the "precise analytical analysis of algorithms". Several new analytical techniques were developed in the process of analyzing different parameters of digital trees (cf. [4, 5, 10, 13, 14, 16, 20, 21, 22]). Recently, the focus of the research was moved towards developing analytical techniques that can handle limiting distributions and large deviations results (cf. [5, 11, 12, 14, 15]).

In this paper, we continue recent lines of research and establish asymptotic distribution together with the first two moments of the height. The novelty of this work lies in deriving an asymptotic solution to a certain functional equation that often arises in the analysis of algorithms and data structures (cf. [4, 21]). Namely, we consider functional equations of the following type:

$$f(z) = f(pz) + f(qz)e^{-pz} + a(z) \tag{1}$$

where $p + q = 1$ and $a(z)$ is a given function. The point to observe is that there *is* a coefficient depending on z in front of $f(qz)$ which makes the problem interesting (otherwise a standard approach can be applied; cf. [7]). While a first-order asymptotic for a such equations, when $z \to \infty$ in a cone around the positive axis, is rather easy to obtain, second-order asymptotics are more challenging. This demands an evaluation of some constants for which a closed-form solution does not exist. We provide a quickly converging numerical procedure to assess these constants. We must mention that functional equations of type (1) could be alternatively treated by the method proposed in [4] (cf. [21]), however, it seems to us that our method is more straightforward. In addition, in [4] the problem of evaluating the constants was not discussed.

When dealing with the limiting distribution, we use a two steps approach recently advocated in some papers (notably: [5, 11, 12, 14]): That is, we first *poissonize* the problem and then *depoissonize* it. By poissonization we mean to replace the fixed size population model (i.e., fixed n) by a model in which the number of persons involved is Poisson distributed with mean n. Such a model leads to a functional equation of type (1): More precisely, for all integer $k \geq 0$

$$f_{k+1}(z) = f_k(pz) + e^{-pz} f_k(qz) \ .$$

This equation is solved inside a cone, and then depoissonized in order to obtain an asymptotic distribution of the original fixed size model. Actually, during the course of establishing the limiting distribution we realize that its asymptotic expression exhibits some fluctuations leading us to a conclusion that the height does not possess a limiting distribution. This was already observed for the height of tries (cf. [3]) and symmetric (unbiased coin tossing) incomplete trie (cf. [5]).

The paper is organized as follows. The next section presents our main results: In Theorem 1 we discuss asymptotics of the mean and the variance of the height. The next Theorem 2 provides an asymptotic expression for the distribution of the height. We close this section with a brief discussion of main consequences of our results. Section 3 contains the proofs of both Theorem 1 and Theorem 2. Since, as we already mentioned above, we work on the Poisson model instead of the original model, we need a tool of depoissonization. For the completeness of our presentation, we briefly discuss a depoissonization lemma of Jacquet and Szpankowski [15] in Section 3.1. Then, Theorem 1 is proved in Section 3.2, and Theorem 2 in Section 3.3.

2 Main Results

In this section, we present our main results. We start with a brief description of the elimination process, and introduce some additional notation. To recall, n people use the following randomized elimination algorithm to identify a leader: Players toss a *biased* coin, and those who throw heads are losers while those who throw tails remain candidate winners and flip the coins again until a single winner (leader) is identified. If *all* players throw heads at any stage, the toss is inconclusive and all players participate again in the contest. Let p be the probability of throwing a tail, that is probability of survival. We also write $q = 1-p$. By H_n we denote the number of tosses needed to identify the winner.

As mentioned before, the elimination process can be represented as an incomplete trie. Having this in mind, one can easily derive the basic recurrence equation for the generating function of H_n. Indeed, let for $n \geq 1$, $G_n(u) = \mathbf{E}u^{H_n} = \sum_{k\geq 0}\mathbf{P}(H_n = k)z^k$ be the probability generating function of H_n, where u is a complex number. We further let $G_0(u) = 0$ for convenience. (This corresponds to defining $H_0 = \infty$; as pointed out by Jim Fill [6], this convention is reasonable since we never succeed to choose a leader without any candidates.)

Then, $G_1(u) = 1$ and for $n \geq 2$

$$G_n(u) = u\sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k}G_k(u) + uq^n G_n(u)\ . \tag{2}$$

The first term of the above is a consequence of the Bernoulli-like split (after the first round) of n players into those who still stay in the game. Clearly, the remaining players have $H_n - 1$ tosses to finish the game. The second term of the above, takes care of the inconclusive throw (when all plays throw heads).

In this paper, we derive the distribution of H_n as well as the first two moments, that is, $\mathbf{E}H_n$ and $\mathbf{Var}\ H_n$. We use the following abbreviated notation: $x_n = \mathbf{E}H_n$ and $w_n = \mathbf{E}H_n(H_n-1)$. Observing that $x_n = G_n'(1)$ and $w_n = G_n''(1)$, we derive from (2):

$$x_n = 1 + q^n x_n + \sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k}x_k\ , \qquad n\geq 2\ , \tag{3}$$

$$w_n = 2(x_n-1) + q^n w_n + \sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k}w_k\ , \qquad n\geq 2\ , \tag{4}$$

with $x_0 = x_1 = w_0 = w_1 = 0$.

In the next section, we solve asymptotically the above recurrence equations using poissonization, Mellin transform and depoissonization. This results in our first main finding.

Theorem 1 *Let $P := 1/p$ and $\chi_k := 2\pi i k/\ln P$. Then:*

(i) *The mean $\mathbf{E}H_n$ of the height admits the following asymptotic formula*

$$\mathbf{E}H_n = \log_P n + \frac{1}{2} - \frac{1-\gamma-T_1^*(0)}{\ln P} + \delta_1(\log_P n) + O(1/n) \tag{5}$$

where $\gamma = 0.577\ldots$ is the Euler constant, and

$$T_1^*(0) = \sum_{n=2}^{\infty} \frac{x_n q^n}{n} \ , \tag{6}$$

where x_n must be computed from (3) (observe that the series converges geometrically fast). The function $\delta_1(x)$ is periodic function of small magnitude (e.g., for $p = 0.5$ one proves $|\delta_1(x)| \le 2 \times 10^{-5}$) given by

$$\delta_1(x) = -\sum_{k \ne 0} \alpha_k e^{-2\pi i k x} \tag{7}$$

where

$$\alpha_k = \frac{(1+\chi_k)\Gamma(\chi_k) - T_1^*(\chi_k)}{\ln P} \ ,$$

$\Gamma(s)$ is the Euler gamma function (cf. [1]) and $T_1^*(s)$ is given by (37).

(ii) *The variance $\mathbf{Var}\ H_n = \mathbf{E}H_n(H_n-1) + \mathbf{E}H_n - (\mathbf{E}H_n)^2$ satisfies*

$$\begin{aligned}\mathbf{Var}\ H_n \ &= \ \frac{\pi^2/6 - 1 + 2(1-\gamma)T_1^*(0) - 2T_1^{*\prime}(0) - (T_1^*(0))^2}{\ln^2 P} + \frac{2T_1^*(0) + T_2^*(0)}{\ln P} + \frac{1}{12} \\ &\quad - [\delta_1^2]_0 + \delta_2(\log_P n) + O\left(\frac{\ln n}{n}\right) \end{aligned} \tag{8}$$

where

$$T_1^{*\prime}(0) = \sum_{n=2}^{\infty} \frac{x_n q^n}{n!}\Gamma'(n) = \sum_{n=2}^{\infty} \frac{x_n q^n}{n}\Psi(n) \ , \tag{9}$$

where $\Psi(z) = \Gamma'(z)/\Gamma(z)$ is the psi-function. Observe that for natural n we have $\Psi(n) = -\gamma + \mathcal{H}_{n-1}$ where $\mathcal{H}_n$ is the Harmonic number. The constant $T_2^*(0)$ can be computed as

$$T_2^*(0) = \sum_{n=2}^{\infty} \frac{w_n q^n}{n}$$

where w_n is given by the recurrence (4). Finally, $\delta_2(x)$ is a periodic continuous function of zero mean and small amplitude. The constant $[\delta_1^2]_0 = \sum_{k \ne 0} |\alpha_k|^2$ is the zeroth term of $\delta_1^2(x)$ and its value is extremely small (e.g., for $p = 0.5$ one proves that $[\delta_1^2]_0 \le \sup|\delta_1(x)|^2 \le 4 \times 10^{-10}$).

Table 1: Numerical evaluation of the constants $T_1^*(0)$, $T_1^{*\prime}(0)$, $T_2^*(0)$, and the variance $\mathbf{Var}\ H_n$ for various $p \in [0.2..0.8]$

p	$T_1^*(0)$	$T_1^{*\prime}(0)$	$T_2^*(0)$	$\mathbf{Var}\ H_n$
0.2	2.36	2.38	9.32	5.83
0.3	1.22	1.09	3.41	3.58
0.4	0.70	0.56	1.64	2.97
0.5	0.42	0.30	0.95	3.12
0.6	0.25	0.17	0.62	4.07
0.7	0.15	0.09	0.45	6.68
0.8	0.08	0.04	0.35	14.84

In Table 1 we present numerical values of the constants $T_1^*(0)$, $T_1^{*\prime}(0)$, $T_2^*(0)$, and the variance $\mathbf{Var}\ H_n$ given by (8) (for large n) as a function of p. In particular, we verify that our formula (8) on the variance agrees with that of Fill *et al.* [5] for $p = 0.5$, where the exact value $1 - \gamma = 0.422\ldots$ is given.

In order to formulate our next result concerning the distribution of H_n we need a new definition. Let a measure μ be defined on the positive real axis as follows: Partition the positive real axis into an infinite sequence of *consecutive* intervals $I_0, I_1, \ldots$ such that I_k has length $(q/p)^{s(k)}$, where $s(k)$ is the number of 1's in the binary expansion of k. Thus, $I_0 = [0,1], I_1 = [1, 1+q/p]$, etc. Note that the total length of the first 2^m intervals $I_0, ..., I_{2^m-1}$ is p^{-m}, and that these 2^m intervals are obtained by repeated subdivisions of $[0, p^{-m}]$, each time dividing each interval in the proportions $p : q$. Given these intervals, define μ by putting a point mass $|I_k|$ at the right endpoint of I_k, for each $k = 0, 1, ...$ Note that for $p = q = 1/2$, μ consists of a unit mass at each positive integer.

Now, we are in a position to present our second main finding:

Theorem 2 *The following holds, uniformly for all integers k,*

$$\mathbf{P}(H_n \le k) = F(p^k n) + O(n^{-1}), \tag{10}$$

where

$$F(x) = x \int_0^\infty e^{-xt} d\mu(t) = \int_0^\infty e^{-t} d\mu_x(t), \tag{11}$$

with μ_x denoting the dilated measure defined as above for the intervals $xI_0, xI_1, \ldots$.
In particular, when $k = \lfloor \log_P n \rfloor + \kappa$ where κ is an integer, then for large n the following

asymptotic formula is true uniformly over κ

$$\mathbf{P}(H_n \le \lfloor \log_P n \rfloor + \kappa) = p^{\kappa - \{\log_P n\}} \int_0^\infty e^{-tp^{\kappa - \{\log_P n\}}} d\mu(t) + O\left(\frac{1}{n}\right), \qquad (12)$$

where $\{\log_P n\} = \log_P n - \lfloor \log_P n \rfloor$.

Remarks (i) *Limiting Distribution Does Not Exist.* The fractional part $\{\log_P n\}$ appearing in Theorem 2 is dense in the interval $[0, 1)$ and does not converge. Thus, the limiting distribution of $H_n - \lfloor \log_P n \rfloor$ does not exist. In fact, we observe that:

$$\begin{aligned} \liminf_{n\to\infty} \mathbf{P}(H_n \le \lfloor \log_P n \rfloor + \kappa) &\le p^{\kappa-1} \int_0^\infty e^{-tp^{\kappa-1}} d\mu(t) , \\ \limsup_{n\to\infty} \mathbf{P}(H_n \le \lfloor \log_P n \rfloor + \kappa) &\ge p^{\kappa} \int_0^\infty e^{-tp^{\kappa}} d\mu(t) . \end{aligned}$$

(ii) *Symmetric Case* $p = q = 0.5$. We observe that for $p = q = 0.5$ we obtain

$$F(x) = x \sum_{j=1}^{\infty} e^{-jx} = \frac{x}{e^x - 1} ,$$

and our results coincide with those of [5].

(iii) It is easily seen that $\lim_{x\to 0} F(x) = 1$ and $\lim_{x\to\infty} F(x) = 0$. We conjecture that $F(x)$ is always decreasing, as it is for $p = 0.5$ by the explicit formula in (ii). If $F(x)$ is decreasing, then $F(p^x)$ is a distribution function, and if Z is a random variable with this distribution, then (10) can be written

$$\mathbf{P}(H_n \le k) = \mathbf{P}(Z + \log_P n \le k) + O(n^{-1}).$$

Hence, in this case, the distribution of H_n is well approximated by the distribution of $\lceil Z + \log_P n \rceil$; for example it follows that the total variation distance between the two distributions tends to 0 as $n \to \infty$, which is a substitute for the failing limit distribution.

(iv) It is possible to obtain further terms in the asymptotic formulae in Theorems 1 and 2 using the same methods.

3 Analysis and Proofs

In this section, we prove Theorems 1 and 2 using an analytical approach. In the next subsection, we transform the problem to the Poisson model (i.e., poissonize it), which is easier to solve. Then, we apply Mellin transform (cf. Section 3.2) and a simple functional analysis (cf. Section 3.3) to obtain asymptotic solution for the poissonized moments and the poissonized distribution for the height. Finally, we depoissonize these findings to recover our results for the original model.

3.1 Poissonization and Depoissonization

It is well known that often poissonization leads to a simpler solution due to unique properties of the Poisson distribution (cf. [9]). *Poissonization* is a technique which replaces the *fixed population* model (sometimes called the Bernoulli model) by a model in which the population varies according to the Poisson law (hence, Poisson model). In the case of the leader election algorithm, we replace n by a random variable N distributed according to Poisson with mean equal n. We shall apply *analytical poissonization* (cf. [10, 11, 13, 14, 20]) that makes use of the Poisson transform (i.e., exponential generating function as shown below). One must observe, however, that after solving the Poisson model (in most cases we can only solve it asymptotically!), we must *depoissonize* to recover the Bernoulli model results. In this subsection, we first derive functional equations for the Poisson model, and then present a general depoissonization lemma of Jacquet and Szpankowski [15] (cf. also [5, 11, 12, 13, 14, 20]) that we apply throughout the paper.

We now build the Poisson model. Let us define

$$
\begin{aligned}
\widetilde{G}(z,u) &= \sum_{n=0}^{\infty} G_n(u)\frac{z^n}{n!}e^{-z} , \\
\widetilde{X}(z) &= \sum_{n=0}^{\infty} x_n\frac{z^n}{n!}e^{-z} , \\
\widetilde{W}(z) &= \sum_{n=0}^{\infty} w_n\frac{z^n}{n!}e^{-z} ,
\end{aligned}
$$

where $G_n(u)$, x_n and w_n are expressed as (2)–(4), respectively. They are poissonized versions of the corresponding quantities in the Bernoulli model.

Remark If $z \geq 0$, then $\widetilde{G}(z,\cdot)$ is the probability generating function of $H_{N(z)}$, where the population size $N(z)$ is random with the Poisson distribution $\mathrm{Po}(z)$. Note, however, that because of our convention $G_0 = 0$ (or $H_0 = \infty$), $\widetilde{G}(z,\cdot)$ is a defective probability generating function. This could be rectified by instead defining $H_0 = 0$, but our choice is more convenient for us. Similarly, $\widetilde{X}(z) = \frac{\partial}{\partial u}\widetilde{G}(z,u)|_{u=1}$ is for $z \geq 0$ the expectation $\mathbf{E}H_{N(z)}$ of the height when the population is random $\mathrm{Po}(z)$, provided we here use the convention $H_0 = 0$.

To see the achieved simplifications, we observe that the recurrences (2)–(4) now become:

$$
\begin{aligned}
\widetilde{G}(z,u) &= u\widetilde{G}(pz,u) + u\widetilde{G}(qz,u)e^{-pz} + (1-u)ze^{-z} \ , && (13)\\
\widetilde{X}(z) &= \widetilde{X}(pz) + \widetilde{X}(qz)e^{-pz} + 1 - (1+z)e^{-z} \ , && (14)\\
\widetilde{W}(z) &= \widetilde{W}(pz) + \widetilde{W}(qz)e^{-pz} + 2\widetilde{X}(z) + 2\left((e^{-z}-1) + ze^{-z}\right) && (15)
\end{aligned}
$$

for a complex z. The above functional equations have a simpler form than their corresponding Bernoulli model equations, but they are far from being trivial. The main difficulty lies in the fact that there is a factor e^{-pz} in front of $\widetilde{G}(qz,u)$, $\widetilde{X}(qz)$ and $\widetilde{W}(qz)$. Observe that in the symmetric case (i.e., $p = q = 0.5$) these functional equations reduce to the one analyzed in Szpankowski [22] (cf. also [5, 7, 16]). We solve these functional equations asymptotically (see the next two subsections) for z large and real. The next step is a depoissonization of these results, and we present now a general depoissonization result of Jacquet and Szpankowski [15] that generalize previous depoissonization lemmas of [11, 12, 13, 20]. Recall that a measurable function $\psi\colon (0,\infty) \to (0,\infty)$ is *slowly varying* if $\psi(tx)/\psi(x) \to 1$ as $x \to \infty$ for every fixed $t > 0$.

Lemma 1 [DEPOISSONIZATION LEMMA] *Assume that $\widetilde{G}(z) = \sum_{n=0}^{\infty} g_n \frac{z^n}{n!} e^{-z}$ is an entire functions of a complex variable z. Suppose that there exist real constants $a < 1$, β, $\theta \in (0, \pi/2)$, c_1, c_2, and z_0, and a slowly varying function ψ such that the following conditions hold, where S_θ is the cone $S_\theta = \{z : \ |\arg(z)| \le \theta\}$:*

(I) *For all $z \in S_\theta$ with $|z| \ge z_0$,*

$$|\widetilde{G}(z)| \le c_1 |z|^\beta \psi(|z|); \tag{16}$$

(O) *For all $z \notin S_\theta$ with $|z| \ge z_0$,*

$$|\widetilde{G}(z)e^z| \le c_2 e^{a|z|}. \tag{17}$$

Then for $n \ge 1$,

$$g_n = \widetilde{G}(n) + O\left(n^{\beta-1}\psi(n)\right) \ . \tag{18}$$

More precisely,

$$g_n = \widetilde{G}(n) - \tfrac{1}{2} n \widetilde{G}''(n) + O\left(n^{\beta-2}\psi(n)\right) \ . \tag{19}$$

The "Big-Oh" terms in (18) and (19) are uniform for any family of entire functions $\widetilde{G}$ that satisfy the conditions with the same a, β, θ, c_1, c_2, z_0 and ψ.

3.2 Analysis of Moments

We now prove Theorem 1 using the Mellin transform and depoissonization techniques. We thus begin by studying the functions $\widetilde{G}(z,u)$, $\widetilde{X}(z)$ and $\widetilde{V}(z)$ defined above, which satisfy the functional equations (13)–(15). We write $f^*(s)$ or $\mathcal{M}(f,s)$ for the Mellin transform of a function $f(x)$ of real parameter x, that is,

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

provided the above integral converges. A beautiful survey on Mellin transform can be found in [8], and we refer the reader to this paper for details concerning Mellin transform.

The Poisson mean $\widetilde{X}(z)$ and second factorial moment $\widetilde{W}(z)$ satisfy function equations (14) and (15), respectively. We observe that from the recurrence equations (3) and (4) we immediately prove that $x_n = O(\ln(n+1))$ and $w_n = O(\ln^2(n+1))$. It follows that $\widetilde{X}$ and $\widetilde{W}$ are entire functions. Moreover, it follows easily that $\widetilde{X}(x) = O(\ln(x+1))$ for $x > 0$. In order to apply the depoissonization lemma we have to extend this estimate to complex arguments in a cone S_θ.

Thus fix $\theta = \pi/4$, say; we claim that

$$|\widetilde{X}(z)| = O(\ln(|z|+1)), \qquad z \in S_\theta. \tag{20}$$

This is proved by induction along increasing domains (cf. [14]) as follows: Let $\rho = \max(p,q)^{-1} > 1$. Suppose that R and A are such that

$$|\widetilde{X}(z)| \le A\ln(|z|), \qquad z \in S_\theta,\ 2 \le |z| \le R. \tag{21}$$

If now $z \in S_\theta$ with $R \le |z| \le \rho R$, then the recursion relation (14) yields, provided $R\min(p,q) \ge 2$,

$$\begin{aligned} |\widetilde{X}(z)| &\le |\widetilde{X}(pz)| + |\widetilde{X}(qz)|e^{-p|z|\cos\theta} + 1 + (1+|z|)e^{-|z|\cos\theta} \\ &\le A\ln(|z|) + A\ln(p) + A\ln(R)e^{-pR\cos\theta} + 2 + (\cos\theta)^{-1}. \end{aligned} \tag{22}$$

Now choose $R_0 \ge 2/\min(p,q)$ such that $\ln(p) + \ln(R)e^{-pR\cos\theta} \le -\delta < 0$ for $R \ge R_0$. If $A \ge 3/\delta\cos\theta$ and $R \ge R_0$, then (22) shows that (21) holds also for $z \in S_\theta$ with $R \le |z| \le \rho R$. Since clearly (21) holds for $R = R_0$ and a suitable large A, (21) holds by induction for $R = \rho^n R_0$ for every $n \ge 0$ (with the same A) and (20) follows for $|z| \ge 2$; for small $|z|$ we use $\widetilde{X}(z) = O(|z|^2)$, $|z| \le 2$, because $x_0 = x_1 = 0$.

Similarly one proves, using (15) and (20),

$$|\widetilde{W}(z)| = O(\ln^2(|z|+1)), \qquad z \in S_\theta. \tag{23}$$

In particular, (20) and (23) hold for real $x > 0$. It follows that the Mellin transforms $X^*(s)$ and $W^*(s)$ exist (and are analytic) in the strip $-1 < \Re s < 0$. (In fact, since $x_1 = w_1 = 0$, they exist for $-2 < \Re s < 0$, but we do not need this.)

Let us now concentrate on the first moment. Define

$$T_1(z) = \widetilde{X}(qz)e^{-pz} \ . \tag{24}$$

Then, $T_1(z)$ is an entire function and the Mellin transform $T_1^*(s)$ exists at least for $-2 < \Re s < \infty$. Indeed, since every $x_n \geq 0$, we have

$$|\widetilde{X}(z)e^z| \leq \widetilde{X}(|z|)e^{|z|} \tag{25}$$

and thus $|\widetilde{X}(z)| \leq \widetilde{X}(|z|)e^{|z|-\Re z}$. Hence, if $x > 0$ and $|z - x| < px/4$,

$$|T_1(z)| \leq \widetilde{X}(q|z|)e^{q|z|-\Re z} \leq \widetilde{X}(q|z|)e^{qx-x+2|z-x|} \leq \widetilde{X}(q|z|)e^{-px/2} = O(e^{-px/2}\ln(1+x)).$$

Thus, by Cauchy's estimate, for every $m \geq 0$,

$$T_1^{(m)}(x) = O(x^{-m}e^{-px/2}\ln(1+x)), \qquad x > 0.$$

Since further $T_1^{(m)}(x)$ is bounded for $0 \leq x \leq 1$, the Mellin transform $T_1^{(m)*}(s)$ exists at least for $0 < \Re s < \infty$, and is bounded on each line $\Re s = \sigma > 0$.

Integration by parts yields $s(s+1)\cdots(s+m-1)T_1^*(s) = (-1)^m T_1^{(m)*}(s+m)$ and thus the estimate

$$|T_1^*(\sigma + i\tau)| \leq \frac{C(\sigma, m)}{(1+|\tau|)^m} \tag{26}$$

for each $m \geq 2$ and $-2 < \sigma < \infty$; $C(\sigma, m)$ is bounded for σ in a compact interval of $(-2, \infty)$ and m fixed. In particular, $T_1^*(\sigma + i\tau)$ is integrable in τ for each $\sigma > -2$.

We re-write (14) as follows:

$$\widetilde{X}(z) = \widetilde{X}(pz) + T_1(z) - (e^{-z} - 1) - ze^{-z} \ .$$

Taking the Mellin transform of the above we have, for $-1 < \Re s < 0$,

$$X^*(s) = p^{-s}X^*(s) + T_1^*(s) - \Gamma(s) - \Gamma(s+1), \tag{27}$$

where $\Gamma(\cdot)$ is the Euler gamma function. Now, we can solve (27) to get

$$X^*(s) = \frac{\Gamma(s) + \Gamma(s+1) - T_1^*(s)}{(1/p)^s - 1}, \qquad -1 < \Re s < 0. \tag{28}$$

The right hand side extends to a meromorphic function in the half plane $-1 < \Re s < \infty$, with poles at $\chi_k = 2\pi i k/\ln(1/p)$, $k = \ldots, -1, 0, 1, 2, \ldots$. All poles are simple except the one at 0 ($k = 0$), which is double.

It follows from (28) and (26) that for every $\sigma \in (-1,0) \cup (0,\infty)$, $|X^*(\sigma + i\tau)| = O((1+\tau^2)^{-1}) \in L^1(d\tau)$. The Mellin (Fourier) inversion formula thus yields for $x > 0$ the following.

$$\widetilde{X}(x) = \frac{1}{2\pi i} \int_{-1/2-i\infty}^{-1/2+i\infty} x^{-s} X^*(s) ds. \tag{29}$$

Shifting the line of integration to $\Re s = R > 0$ (using the Cauchy residue theorem) we obtain for any large R,

$$\widetilde{X}(x) = O(x^{-R}) - \sum_{-\infty}^{\infty} \operatorname{Res}_{\chi_k}(x^{-s} X^*(s)). \tag{30}$$

Let $\alpha_k = \operatorname{Res}_{\chi_k} X^*(s)$. If $k \neq 0$, then

$$\alpha_k = ((1+\chi_k)\Gamma(\chi_k) - T_1^*(\chi_k))/\ln(1/p);$$

in particular this implies $\alpha_k = O(|k|^{-n})$ for each $n > 0$. Moreover, for $k \neq 0$,

$$\operatorname{Res}_{\chi_k}(x^{-s} X^*(s)) = x^{-\chi_k} \operatorname{Res}_{\chi_k} X^*(s) = \alpha_k e^{-2\pi i k \log_{1/p} x}.$$

For $k = 0$, we obtain

$$\operatorname{Res}_0(x^{-s} X^*(s)) = -\ln(x)/\ln(1/p) + \operatorname{Res}_0 X^*(s) = -\log_{1/p} x - \frac{1}{2} + \frac{1-\gamma-T_1^*(0)}{\ln(1/p)}.$$

Consequently, for every $R > 0$,

$$\widetilde{X}(x) = \log_{1/p} x + \frac{1}{2} - \frac{1-\gamma-T_1^*(0)}{\ln(1/p)} + \delta_1(\log_{1/p} x) + O(x^{-R}), \tag{31}$$

where $\delta_1(t) = -\sum_{k\neq 0} \alpha_k e^{-2\pi i k t}$ is a periodic function with mean 0.

We now apply the depoissonization lemma. We have already verified condition (I) in (20), with $\beta = 0$ and $\psi(x) = \ln(x+1)$. Condition (O) can be verified similarly, but it is also possible to avoid induction and use the recursion just once as follows. First, by (25) and (20), $|\widetilde{X}(z)e^z| \leq \widetilde{X}(|z|)e^{|z|} \leq Ce^{|z|}\ln(|z|+1)$ for every z. Next, (14) can be written

$$e^z \widetilde{X}(z) = e^{qz} e^{pz} \widetilde{X}(pz) + e^{qz} \widetilde{X}(qz) + e^z - 1 - z,$$

which thus yields

$$|e^z \widetilde{X}(z)| \leq C \ln(1+|z|) \left(e^{p|z|+q\Re z} + e^{q|z|}\right) + e^{\Re z} + 1 + |z|,$$

and (O) follows.

Depoissonization Lemma now gives immediately, by (18), the first moment

$$\mathbf{E}H_n = x_n = \log_P n + \frac{1}{2} - \frac{1-\gamma-T_1^*(0)}{\ln P} + \delta_1(\log_P n) + O(\ln n/n). \tag{32}$$

The error term can be improved to $O(1/n)$, which yields (5), by instead using (19) and verifying that $\widetilde{X}''(x) = O(x^{-2})$, $x > 0$. The latter estimate is easily obtained by differentiating (29) twice (moving the derivatives inside the integral) and estimating the integral by residue calculus as above. The details can be found in [14, 15].

Now, the second moment. Let $T_2(z) = \widetilde{W}(qz)e^{-pz}$; then the Mellin transform $T_2^*(s)$ exists for $-2 < \Re s < \infty$ and (15) yields, for $-1 < \Re s < 0$,

$$W^*(s) = p^{-s}W^*(s) + T_2^*(s) + 2X^*(s) + 2\Gamma(s) + 2\Gamma(s+1), \tag{33}$$

and thus

$$\begin{aligned} W^*(s) &= \frac{2X^*(s)}{1-p^{-s}} + \frac{2\Gamma(s)+2\Gamma(s+1)+T_2^*(s)}{1-p^{-s}} \\ &= \frac{-2\Gamma(s)-2\Gamma(s+1)+2T_1^*(s)}{((1/p)^s-1)^2} - \frac{2\Gamma(s)+2\Gamma(s+1)+T_2^*(s)}{(1/p)^s-1}. \end{aligned} \tag{34}$$

As above, we can obtain an asymptotic expansion of $\widetilde{W}(z)$ by finding the inverse of the Mellin transform. Thus, the Poisson variance $\widetilde{V}(z) = \widetilde{W}(z) + \widetilde{X}(z) - \widetilde{X}^2(z)$ becomes, after some lengthy but elementary calculations,

$$\begin{aligned} \widetilde{V}(x) &= \frac{\pi^2}{6\ln^2 p} + \frac{1}{12} + \frac{2T_1^*(0)+T_2^*(0)}{\ln(1/p)} \\ &+ \frac{-1-2T_1^{*\prime}(0)+2(1-\gamma)T_1^*(0)-(T_1^*(0))^2}{\ln^2 p} + \delta_3(\log_{1/p} x) + O(x^{-R}) \end{aligned} \tag{35}$$

where $\delta_3(t) = \delta_2(t) - [\delta_1^2]_0$ is a small fluctuating function. Applying now the Depoissonization Lemma to $\widetilde{W}$ (verifying (O) as for $\widetilde{X}$), we easily obtain (8). In fact, it follows as above, using (19), that $w_n = \widetilde{W}(n) + O(\ln n/n)$ ((18) would give $O(\ln^2 n/n)$), and this together with the already proven $x_n = \widetilde{X}(n) + O(1/n)$ yields $\mathbf{Var}\ \ H_n = w_n + x_n - x_n^2 = \widetilde{V}(n) + O(\ln n/n)$ (cf. [14, 15]).

To complete the proof of Theorem 1, we need a method of evaluating the constants $T_1^*(0)$, $T_2^*(0)$, and $T_1^{*\prime}(0)$. Let again $x_n = \mathbf{E}H_n$ which we can compute for any n from the recurrence above. We need an evaluation of the Mellin of $\widetilde{X}(z) = \sum_{n\geq 2} x_n \frac{z^n}{n!}e^{-z}$ since $x_0 = x_1 = 0$. Thus noting that $\mathcal{M}(e^{-z}, s) = \Gamma(s)$ for $\Re(s) > 0$, and furthermore $\mathcal{M}(z^n e^{-z}, s) = \Gamma(s+n)$ for $\Re(s) > -n$, we immediately derive

$$X^*(s) = \sum_{n=2}^{\infty} \frac{x_n}{n!}\mathcal{M}(z^n e^{-z}, s) = \sum_{n=2}^{\infty} \frac{x_n}{n!}\Gamma(s+n) \tag{36}$$

provided $\Re(s) \in (-2, 0)$. Observe that the series converges absolutely in this range by the estimate of x_n above.

Moreover,

$$T_1(z) = \widetilde{X}(qz)e^{-pz} = \sum_{n\geq 2} x_n \frac{(qz)^n}{n!} e^{-z} = \sum_{n\geq 2} \frac{x_n q^n}{n!} z^n e^{-z}$$

and thus, similarly,

$$T_1^*(s) = \sum_{n=2}^{\infty} \frac{x_n q^n}{n!} \Gamma(s+n) \tag{37}$$

provided $-2 < \Re(s) < \infty$. In particular,

$$T_1^*(0) = \sum_{n=2}^{\infty} \frac{x_n q^n}{n!} \Gamma(n) = \sum_{n=2}^{\infty} \frac{x_n q^n}{n}, \tag{38}$$

which has an exponential rate of convergence.

Now, we can move on and estimate $T_1^{*\prime}(0)$. Taking the derivative of (37) at $s = 0$ and arguing as before we arrive at the following formula

$$T_1^{*\prime}(0) = \sum_{n=2}^{\infty} \frac{x_n q^n}{n!} \Gamma'(n) = \sum_{n=2}^{\infty} \frac{x_n q^n}{n} \Psi(n), \tag{39}$$

where $\Psi(s) = \Gamma'(s)/\Gamma(s)$ is the Psi function; recall that for $n \geq 2$ we have $\Psi(n) = -\gamma + \mathcal{H}_{n-1}$ (where $\mathcal{H}_n$ here is the Harmonic number).

Of course, $T_2^*(0)$ can be computed in a similar fashion.

$$T_2^*(0) = \sum_{n=2}^{\infty} \frac{w_n q^n}{n}, \tag{40}$$

This completes the proof of Theorem 1.

3.3 Analysis of the Distribution

We now prove Theorem 2. We start with the functional equation (13). After defining $\widetilde{H}(z,u) = \widetilde{G}(z,u)/(1-u)$ we obtain

$$\widetilde{H}(z,u) = u\widetilde{H}(pz,u) + u\widetilde{H}(qz,u)e^{-pz} + ze^{-z} \ . \tag{41}$$

Let now $\widetilde{G}_k(z) = \sum_{n=1}^{\infty} \mathbf{P}(H_n \leq k)\frac{z^n}{n!}e^{-z}$. Then, $\widetilde{H}(z,u) = \sum_{k=0}^{\infty} u^k \widetilde{G}_k(z)$ and an identification of the coefficients of u in (41) yield

$$\begin{aligned} \widetilde{G}_0(z) &= ze^{-z} \\ \widetilde{G}_{k+1}(z) &= \widetilde{G}_k(pz) + e^{-pz}\widetilde{G}_k(qz) \ , \qquad k \geq 0. \end{aligned} \tag{42}$$

We claim that the above functional equation for $\widetilde{G}_k(z)$ is solved by

$$\widetilde{G}_k(z) = p^k z \int_0^{p^{-k}} e^{-p^k z t} d\mu(t) \tag{43}$$

(where μ is defined just above Theorem 2). In fact, the case $k = 0$ is simple, because μ restricted to $[0, 1]$ only consists of a point mass at 1, and thus the integral equals e^{-z}. For $k \geq 1$, we use the fact that the measure μ on $(p^{-k}, p^{-k-1}]$ is obtained from μ on $(0, p^{-k}]$ by a translation and dilation, so that for every function f,

$$\int_{p^{-k}+}^{p^{-k-1}} f(t) d\mu(t) = \frac{q}{p} \int_0^{p^{-k}} f(p^{-k} + \frac{q}{p} t) d\mu(t)$$

and thus

$$\int_0^{p^{-k-1}} f(t) d\mu(t) = \int_0^{p^{-k}} f(t) d\mu(t) + \frac{q}{p} \int_0^{p^{-k}} f(p^{-k} + \frac{q}{p} t) d\mu(t).$$

It is now easy to see that (43) satisfies (42).

Moreover, (43) trivially holds also for $k < 0$, with both sides zero; hence (43) holds for all integers k.

We next observe that, with $F(x)$ defined in (11),

$$|\widetilde{G}_k(n) - F(p^k n)| = p^k n \int_{p^{-k}+}^{\infty} e^{-p^k n t} d\mu(t) \leq p^k n \int_{p^{-k}}^{\infty} e^{-p^k n t} dt = e^{-n} \tag{44}$$

when $k \geq 0$, and similarly $|\widetilde{G}_k(n) - F(p^k n)| = F(p^k n) \leq p^k n e^{-p^k n} + e^{-n} \leq (n+1)e^{-n}$ when $k < 0$. Thus, to complete the proof of Theorem 2 it suffices to apply Depoissonization Lemma, provided we verify conditions (I) and (O) uniformly for all k (and some fixed θ), with $\beta = 0$ and $\psi = 1$. This can be done as above using induction along increasing domains, but we give a simpler argument. Indeed, in order to verify (I) we can use the exact formula (43) and observe that for any decreasing function $f \geq 0$ on $(0, \infty)$, $\int_0^\infty f \, d\mu(t) \leq \int_0^\infty f \, dt$; hence if $\Re z > 0$,

$$|\widetilde{G}_k(z)| \leq p^k |z| \int_0^\infty e^{-p^k \Re z t} \, d\mu(t) \leq p^k |z| \int_0^\infty e^{-p^k \Re z t} \, dt = \frac{|z|}{\Re z}.$$

Consequently (I) holds, uniformly in k, for any $\theta < \pi/2$.

For (O) we first observe that

$$|e^z \widetilde{G}_k(z)| \leq \sum_{n=0}^{\infty} \frac{|z|^n}{n!} = e^{|z|}.$$

Hence (42) yields, for $k \geq 0$,

$$|e^z \widetilde{G}_{k+1}(z)| \leq |e^{qz}||e^{pz} \widetilde{G}_k(pz)| + |e^{qz} \widetilde{G}_k(qz)| \leq e^{q\Re z + p|z|} + e^{q|z|} \leq e^{a|z|}$$

for some $a < 1$ and z sufficiently large. Since further $|e^z \widetilde{G}_0(z)| = |z|$, also (O) holds uniformly in k. This completes the proof of Theorem 2.

ACKNOWLEDGEMENT

We would like to thank Profs. J. Fill and H. Mahmoud for several discussions concerning this paper.

References

[1] Abramowitz, M. and Stegun, I., eds. *Handbook of Mathematical Functions*. Wiley, New York, 1962.

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

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

[4] Fayolle, G., Flajolet, P., and Hofri, M. On a Functional Equation Arising in the Analysis of a Protocol for a Multiaccess Broadcast Channel, *Adv. Appl. Probab.*, 18, 441–472, 1984.

[5] Fill, J. A., Mahmoud, H. M., and Szpankowski, W. On the distribution for the duration of a randomized leader election algorithm. *Ann. Appl. Probab.*, 1996.

[6] Fill, J. Personal Communication.

[7] Flajolet, P., Régnier, M. and Sotteau, D., Algebraic Methods for Trie Statistics, *Annals of Discrete Mathematics*, 25, 145-188, 1985.

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

[9] Gonnet, G. and Munro, J. The Analysis of Linear Probing Sort by the Use of a New Mathematical Transform, *Journal of Algorithms*, 5, 451-470, 1984.

[10] Grabner, P. Searching for losers. *Random Structures and Algorithms*, 4, 99–110, 1993.

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

[12] Jacquet, P. and Régnier, M. Normal Limiting Distribution of the Size of Tries, *Proc. Performance'87*, 209-223, North Holland, Amsterdam 1987

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

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

[15] Jacquet, P. and Szpankowski, W. Analytical Depoissonization Lemma and Its Applications, in preparation.

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

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

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

[19] Prodinger, H. How to select a loser. *Discrete Math.*, 120, 149–159, 1993.

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

[21] Schmid, U. The Average CRI-length of a Tree Collision Resolution Algorithm in Presence of Multiplicity-Dependent Capture Effects, *Proc. ICALP 92*, Vienna, 223-234, 1992.

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