Adv. Appl. Prob. **46,** 203–216 (2014)
Printed in Northern Ireland

PERPETUITIES IN FAIR LEADER ELECTION ALGORITHMS

RAVI KALPATHY * ** AND
HOSAM MAHMOUD,* *** *The George Washington University*

Abstract

We consider a broad class of fair leader election algorithms, and study the duration of contestants (the number of rounds a randomly selected contestant stays in the competition) and the overall cost of the algorithm. We give sufficient conditions for the duration to have a geometric limit distribution (a perpetuity built from Bernoulli random variables), and for the limiting distribution of the total cost (after suitable normalization) to be a perpetuity. For the duration, the proof is established via convergence (to 0) of the first-order Wasserstein distance from the geometric limit. For the normalized overall cost, the method of proof is also convergence of the first-order Wasserstein distance, augmented with an argument based on a contraction mapping in the first-order Wasserstein metric space to show that the limit approaches a unique fixed-point solution of a perpetuity distributional equation. The use of these two steps is commonly referred to as the contraction method.

Keywords: Leader election; recurrence; functional equation; fixed point; contraction method; metric space; perpetuity; weak convergence

2010 Mathematics Subject Classification: Primary 60C05
Secondary 60F05; 68W40

1. Introduction

There is a plethora of research work on leader election algorithms, each dealing with a specific instance of the problem (see, e.g. [5], [7], [9], [11], [12], [13], and [16]). There is a need for a broader framework to establish results for classes of such algorithms. The recent work [8] gives a theory for the cost associated with the number of rounds (equivalently the height of the underlying incomplete tree). It is our purpose in this paper to give a parallel set of conditions to obtain results for the duration of contestants and overall cost for a broad class of leader election algorithms.

Fair leader election algorithms are used in numerous applications including the selection of a winner of a contest, a loser of bets, or a coordinator of a security system in the case of failure of the existing central coordinator. The common scenario in these situations is the following. There are a number of contestants who will compete fairly to elect a winner (and in some variations they may all lose the election, resulting in no winner). The contestants go through elimination rounds in which they generate events that decide whether or not they advance to the next round, or alternatively a moderator generates these events to fairly elect the candidates at the next round. The concept of fairness will be implied throughout, according to which the chance (probability) of winning is the same for all contestants.

Received 1 June 2012; revision received 29 April 2013.
* Postal address: Department of Statistics, The George Washington University, Washington, DC 20052, USA.
** Email address: kalpathy@gwu.edu
*** Email address: hosam@gwu.edu

https://doi.org/10.1239/aap/1396360110 Published online by Cambridge University Press

In the analysis of several specific variants of leader election algorithms, the typical parameter that researchers have focused on has been the number of rounds till termination (see [5] and [7]). An equally important parameter is the total cost. For instance, in variants in which the elimination is determined by coin flips, the cost can be taken to be the total number of coin flips till termination. We also study the prospects of a particular contestant, as represented by the distribution of the number of rounds she stays, which is an important measure from the point of view of an individual contestant.

To embed the fair leader election in a broader scope of algorithms, we consider an underlying one-sided tree structure (also called an incomplete tree). Consider, for instance, the classic case of n contestants flipping coins, and only those who flip Heads (with probability p) advance to the next round. Those who flip Tails (with probability q) are out of the competition, unless all the contestants flip Tails, in which case the coin tosses are deemed inconclusive and all the contestants try again. Rounds of coin tossing are repeated among those who advance till one winner is elected as a leader. A path in a binary tree (also called the trie) underlies this elimination process. If we develop both sides of the tree, we would get the full binary tree with each contestant residing in a leaf by herself. However, as we eliminate the losers by pruning the branches leading to them, we trim the tree down to a path joining the root to the single leaf containing the winner. Such an incomplete tree forms the backbone for many one-sided algorithms, such as the tree that underlies the algorithm Find [19], which identifies order statistics in a data set.

Our main contribution is to present a unifying treatment for leader election algorithms, and to see how perpetuities naturally come about. In the past, leader election algorithms have been discussed via a variety of methods, such as analytic techniques, Poissonization, integral transforms, and others, which may work for some splitting protocols but not for all. Our treatment covers one-sided leader election algorithms represented by a certain stochastic recurrence with linearly bounded toll functions. The methodology can be extended, with some adaptation, to other one-sided algorithms of similar structure; see, e.g. [3], [10], [14], and [19]. A variety of other one-sided algorithms may be approached by the methods we discuss in the present paper, such as the random walk on interval trees [6], which have a continuous flavor.

2. The technical setup

Assume that there are n contestants competing. They (or a contest moderator on their behalf) generate(s) a certain number $K_n \in \{0, 1, \ldots, n\}$, possibly deterministic or random, of candidates who remain in the contest and the rest of the contestants are eliminated. The algorithm is then applied recursively on the remaining set of contestants, until a leader is elected or no one wins the contest. The generated events and the moderators are fair, in the sense that all contestants have the same chance of winning. The cost (number of steps) of the operations for generating the set of candidates who advance to the next round is a *toll* T_n. It is natural to consider efficient algorithms, where T_n is of order n. Thus, it costs T_n to generate a specific subset of candidates, of size K_n. We call the selection algorithm a *splitting protocol*, and call the set chosen to proceed to the next round the *advancing set*. We shall study the distribution of the duration of an individual contestant and the total time or cost of the competition. We define these two parameters in the two following subsections.

2.1. Duration of a contestant

Let $D_{n,j}$ be the number of rounds, or the time duration, that the jth contestant survives. Under a fair splitting policy, all contestants are equally likely to be selected to advance, and we have $D_{n,j} \overset{\mathrm{D}}{=} D_{n,1}$ (here $\overset{\mathrm{D}}{=}$ denotes equality in distribution). So, we shall develop results for $D_n := D_{n,1}$, dropping the second subscript to keep the notation simple. Thus, D_n has the distribution of the duration of a randomly selected contestant too.

We have a stochastic recurrence for D_n, ensuing as follows. Contestant 1 either advances to the next round with probability (conditioned on K_n) equal to $\binom{n-1}{K_n-1}/\binom{n}{K_n} = K_n/n$, by the fairness of the selection algorithm, or loses the contest and gets eliminated in one step, if she is not selected in the advancing set (with conditional probability $1 - K_n/n$). The algorithm repeats recursively on the set chosen to advance. Thus, for $n \geq 2$, and for a given K_n, we have a stochastic recurrence equation:

$$D_n \overset{\mathrm{D}}{=} \begin{cases} 1 + D_{K_n} & \text{with probability } K_n/n, \\ 1 & \text{with probability } 1 - K_n/n. \end{cases}$$

The initial conditions are $D_0 = D_1 = 0$. Let U be a random variable uniformly distributed on $(0, 1)$. Equivalently, we can write the latter recurrence as

$$D_n \overset{\mathrm{D}}{=} \mathbf{1}_{\{U<K_n/n\}} D_{K_n} + 1, \tag{1}$$

where, for any $n \geq 2$, K_n is a random variable with a given distribution on $0, \ldots, n$, and K_n, U, and D_i are independent for all $i < n$. All the random variables are defined on the same probability space. So, if $\mathbb{P}(K_n = n) < 1$, this gives a recursive definition for the distribution of D_n.

2.2. The total cost

Let X_n be the total cost till termination (i.e. till a winner is chosen or the moderators declare there are no winners). It takes a toll of T_n in the first round to produce the set that will advance to the second round, and then the algorithm is applied recursively on this set of remaining contestants. The toll can be taken to be the time it takes to generate the advancing set measured in suitable units, such as the number of algorithmic steps or machine instructions when the algorithm is executed on a computer, or the number of coin flips in coin flipping variants. We shall say more about these tolls in particular contexts in the following text.

For $n \geq 2$, we have a stochastic recurrence equation for the total underlying cost till termination:

$$X_n \overset{\mathrm{D}}{=} X_{K_n} + T_n, \tag{2}$$

with initial conditions $X_0 = X_1 = 0$. Here, for any $n \geq 2$, again K_n is a random variable with a given distribution in the set $\{0, \ldots, n\}$, and (K_n, T_n) is independent of $X_0, \ldots, X_{n-1}$. All the random variables are defined on the same probability space. So, if $\mathbb{P}(K_n = n) < 1$, this gives a recursive definition for the distribution of X_n.

Under suitable normalization, a random variable satisfying a distributional recurrence of type (2) usually leads to a stochastic fixed-point equation

$$X^* \overset{\mathrm{D}}{=} K^* X^* + T^*,$$

with X^* independent of (K^*, T^*), and with the latter pair having the limiting distribution of $(K_n/n, T_n/n)$. In the context of a leader election, T^* is integrable and X^* will also be integrable. A random variable satisfying the latter distributional equation is called a *perpetuity*, a construct that appears in insurance and the mathematics of finance [4], in stochastic recursive algorithms (see [1], which is quite relevant to our work), and in many other areas.

The rest of the paper is organized as follows. In Section 3 we give the main results in two subsections: Subsection 3.1 is for a set of regularity conditions to derive the main theorems, which are presented in Subsection 3.2. Section 4 is dedicated to the technical proofs of the main results. We end the paper in Section 5 with five illustrative examples, presented in five subsections, one example per subsection.

3. Main results

We impose sufficient *regularity conditions*, of broad applicability, to derive a geometric limit distribution for D_n, and a perpetuity representation for a suitably scaled version of X_n.

3.1. A set of regularity conditions

When we say that a sequence of random variables Y_n is $O_{\mathcal{L}_1}(g(n))$, we mean there exist a positive constant C and a positive integer n_0 such that $\mathbb{E}[|Y_n|] \le C|g(n)|$ for all $n \ge n_0$. We shall assume the following set of regularity conditions, for some $\alpha \in (0, 1)$ and random variables all defined on the same probability space:

(i) The advancing set size satisfies

$$K_n^* := \frac{K_n}{n} = K^* + O_{\mathcal{L}_1}\left(\frac{1}{n^\alpha}\right)$$

for some limiting random variable K^*, with distribution supported on $[0, 1]$, and mean $\mathbb{E}[K^*] < 1$.

(ii) The toll function satisfies

$$T_n^* := \frac{T_n}{n} = T^* + O_{\mathcal{L}_1}\left(\frac{1}{n^\alpha}\right)$$

for some integrable limiting random variable T^*.

The rationale for condition (i) is clear—the size K_n of the advancing set is always a proportion of n, and we deal with cases where K_n/n converges in the $\mathcal{L}_1$ norm to a limit K^* at a fast enough rate to aid the convergence of the duration and scaled cost. If K_n/n does not converge, there may be no convergence at all for the distributions of D_n and X_n.

The rationale for condition (ii) is that we are considering only efficient selection algorithms that do not perform superfluous steps. It is possible for most familiar distributions of K_n to generate equally likely subsets of size K_n to advance to the next round (with cost T_n) in time asymptotically proportional to K_n, as we shall see in a number of illustrating examples. In general, it is possible to generate the advancing set in $O(n)$ time. For the total cost, we require convergence of T_n/n for reasons similar to the rationale of (i).

Remark 1. The regularity conditions (i) and (ii) are not too restrictive in practice. Natural splitting protocols easily meet these constraints, as we shall see in several illustrative examples in Section 5.

Remark 2. Regularity condition (i) implies that $\sup_{n\ge\nu} \mathbb{E}[(K_n^*)^{1-\alpha}] < 1$ for some $\nu \ge 1$, a technical point that is instrumental in building induction proofs of convergence.

3.2. The main theorems

The first main result is represented using the notation Geo(p), for the geometric random variable with success probability p. We use the symbol $\xrightarrow{\mathrm{D}}$ to denote convergence in distribution.

Theorem 1. *Suppose we conduct a leader election among n contestants, in which a fair selection of a subset of contestants of a random size K_n advance to the next round, and the algorithm is applied recursively on that subset until one leader is elected or no one is elected. Assume that K_n follows regularity condition (i). Let D_n be the duration (number of rounds) a contestant stays in the competition. We then have*

$$D_n \xrightarrow{\mathrm{D}} \mathrm{Geo}(1 - \mathbb{E}[K^*]).$$

Remark 3. The only aspect of K^* that enters the picture in the limit is its mean. All distributions of K^* that have the same mean, will have the same limit geometric distribution for the duration of a contestant. This shows that fair leader election is a robust algorithm across a wide variety of splitting protocols. For instance, we shall see that uniform splitting, binomial splitting (with an unbiased coin), and certain ladder contests, in spite of remarkable differences among these splitting protocols, all have Geo$(\frac{1}{2})$ as limit for the duration of participants.

Theorem 2. *Suppose we conduct a leader election among n contestants, in which a fair selection of a subset of contestants of a random size K_n advance to the next round, and the algorithm is applied recursively on that subset until one leader is elected or no one is elected. Assume that K_n follows regularity condition (i). Suppose, moreover, that generating the advancing set of size K_n costs T_n, with T_n satisfying regularity condition (ii). Let X_n be the total cost of the algorithm (over all rounds till termination). We then have*

$$\frac{X_n}{n} \xrightarrow{\mathrm{D}} X^*,$$

where X^* is a perpetuity given by

$$X^* \stackrel{\mathrm{D}}{=} S_1^* + \sum_{j=1}^{\infty} S_{j+1}^* \prod_{i=1}^{j} V_i^*,$$

with $\{(V_i^*, S_i^*)\}_{i=1}^{\infty}$ being a totally independent set of random vectors distributed like (K^*, T^*).

4. Proofs

The *Wasserstein distance of order k* between two distribution functions F and G is defined by

$$\Delta_k(F, G) = \inf \|W - Z\|_k,$$

where the infimum is taken over all coupled random variables W and Z (defined on the same probability space) having the respective distribution functions F and G (with $\|\cdot\|_k$ being the usual $\mathcal{L}_k$ norm). In what follows we use F_Y for the distribution function of a random variable Y. It is well known [2] that, for a sequence of random variables W_n, the convergence of first-order Wasserstein distances between F_{W_n} and F_W to 0 implies the convergence $W_n \xrightarrow{\mathrm{D}} W$, as well as convergence of the first absolute moment.

https://doi.org/10.1239/aap/1396360110 Published online by Cambridge University Press

Before giving the proof of Theorem 1, we note that, in view of regularity condition (i), the structure of the stochastic recurrence (1) suggests that D_n converges to a limiting random variable, say D^*, satisfying the distributional equation

$$D^* \overset{\mathrm{D}}{=} \mathbf{1}_{\{U<K^*\}} D^* + 1, \tag{3}$$

with D^* and $\mathbf{1}_{\{U<K^*\}}$ being independent and U and K^* also independent.

The strategy of the proof is to first show, in Theorem 3, that the first-order Wasserstein distance between D_n and D^* converges to 0. Then, in Lemma 1, we elicit the nature of the limit, which turns out to be a geometric random variable.

Theorem 3. (The coupling theorem.) *Let D_n be as in Theorem 1. Then*

$$D_n \overset{\mathrm{D}}{\to} D^*,$$

where D^* satisfies (3).

Proof. Let (D_n, D^*) be optimal couplings for all $n \geq 0$. Let

$$b_n := \mathbb{E}[|(\mathbf{1}_{\{U<K_n/n\}} D_{K_n} + 1) - (\mathbf{1}_{\{U<K^*\}} D^* + 1)|].$$

We show that $b_n \to 0$; subsequently, we have $\Delta_1(F_{D_n}, F_{D^*}) \leq b_n \to 0$, i.e. $D_n \overset{\mathrm{D}}{\to} D^*$. Since D_n and D^* are an optimal coupling defined on the same space, they have a joint distribution, and b_n is well defined. By regularity condition (i), $\mathbf{1}_{\{U<K_n/n\}} = \mathbf{1}_{\{U<K^*\}} + O_{\mathcal{L}_1}(n^{-\alpha})$, so

$$\begin{aligned} b_n &\leq \mathbb{E}[|\mathbf{1}_{\{U<K_n/n\}}(D_{K_n} - D^*)|] + \mathbb{E}\bigg[\bigg|D^* \times O_{\mathcal{L}_1}\bigg(\frac{1}{n^\alpha}\bigg)\bigg|\bigg] \\ &= \mathbb{E}[|\mathbf{1}_{\{U<K_n/n\}}(D_{K_n} - D^*)|] + \mathbb{E}[D^*] \times \mathbb{E}\bigg[O_{\mathcal{L}_1}\bigg(\frac{1}{n^\alpha}\bigg)\bigg]; \end{aligned}$$

the separation of the expectations of the $O_{\mathcal{L}_1}$ term (coming from $K_n^* - K^*$) and D^* follows from their independence. It is immediate from (3) that D^* has mean $1/(1 - \mathbb{E}[K^*])$. Condition (i) guarantees that this mean is finite. By regularity condition (i), there exists a positive integer n_0 and a positive real constant A, such that

$$\mathbb{E}[|K_n^* - K^*|] \leq \frac{A}{n^\alpha} \quad \text{for all } n \geq n_0. \tag{4}$$

By the finiteness of $\mathbb{E}[D^*]$, we see that $A' = A/(1 - \mathbb{E}[K^*])$ is a positive number. We use conditional independence to write

$$\begin{aligned} b_n &\leq \sum_{k=0}^{n} \frac{k}{n} \mathbb{E}[|D_k - D^*|]\mathbb{P}(K_n = k) + \frac{A'}{n^\alpha} \\ &= b_n \mathbb{P}(K_n = n) + \frac{1}{n}\sum_{k=0}^{n-1} k b_k \mathbb{P}(K_n = k) + \frac{A'}{n^\alpha}. \end{aligned}$$

According to Remark 2, the probability $\mathbb{P}(K_n = n)$ is less than 1 for all $n \geq \nu \geq 1$. So, for all $n \geq n_0' = \max\{\nu, n_0\}$, we can now write

$$b_n \leq \frac{(1/n)\sum_{k=0}^{n-1} k b_k \mathbb{P}(K_n = k) + A'/n^\alpha}{1 - \mathbb{P}(K_n = n)}.$$

Let us start an induction to show that $b_n \leq h/n^\alpha$ for some constant $h > 0$. (One may be able to weaken the regularity conditions to ones without rates (in this context, assuming rates of convergence leads to constructive and transparent proofs). It may be possible to find a proof (probably more involved) that does not assume such rates.) For $1 \leq n \leq n'_0$, we have

$$b_n \leq \max_{1\leq j\leq n} b_j \leq \frac{(n'_0)^\alpha \max_{1\leq j\leq n'_0} b_j}{n^\alpha} =: \frac{h_1}{n^\alpha}.$$

This can be a basis of induction, if h is taken at least as large as h_1.

Assume, for some $n \geq n'_0$ and all $k \in \{1, \dots, n-1\}$, that $b_k \leq h/k^\alpha$. We have

$$\begin{aligned} b_n &\leq \frac{(1/n)\sum_{k=0}^{n} kh\mathbb{P}(K_n = k)/k^\alpha - h\mathbb{P}(K_n = n)/n^\alpha + A'/n^\alpha}{1 - \mathbb{P}(K_n = n)} \\ &= \frac{h\mathbb{E}[(K_n/n)^{1-\alpha}]/n^\alpha - h\mathbb{P}(K_n = n)/n^\alpha + A'/n^\alpha}{1 - \mathbb{P}(K_n = n)}. \end{aligned}$$

The induction step will be complete if, for all $n \geq n'_0$,

$$\frac{h\mathbb{E}[(K_n^*)^{1-\alpha}] - h\mathbb{P}(K_n = n) + A'}{1 - \mathbb{P}(K_n = n)} \leq h.$$

Indeed, such a bound holds if h is chosen high enough. Specifically, after rearrangement, we see that the bound holds if

$$h \geq \frac{A'}{1 - \mathbb{E}[(K_n^*)^{1-\alpha}]} \quad \text{for all } n \geq n'_0,$$

which is the case if

$$h \geq \frac{A'}{1 - \sup_{n\geq n'_0} \mathbb{E}[(K_n^*)^{1-\alpha}]} =: h_2.$$

Take $h = \max\{h_1, h_2\}$, and by induction $b_n \leq h/n^\alpha$ for all $n \geq 1$. Thus, $\Delta_1(F_{D_n}, F_{D^*}) \leq b_n \to 0$. This completes the proof of Theorem 3.

Lemma 1. *A random variable D^* satisfying (3) has a geometric distribution with parameter $1 - \mathbb{E}[K^*]$.*

Proof. Let $\phi^*(t) = \mathbb{E}[\mathrm{e}^{D^*t}]$ be the moment generating function of D^*, with $t < \ln(1/\mathbb{E}[K^*])$. Condition on the indicator random variable to write

$$\begin{aligned} \phi^*(t) &= \mathbb{E}[\mathrm{e}^{(\mathbf{1}_{\{U<K^*\}}D^*+1)t}] \\ &= \mathrm{e}^t\mathbb{P}(\mathbf{1}_{\{U<K^*\}} = 0) + \mathbb{E}[\mathrm{e}^{(D^*+1)t}]\mathbb{P}(\mathbf{1}_{\{U<K^*\}} = 1) \\ &= \mathrm{e}^t(1 - \mathbb{E}[K^*]) + \mathrm{e}^t\phi^*(t)\mathbb{E}[K^*]. \end{aligned}$$

The solution to this equation is

$$\phi^*(t) = \frac{(1 - \mathbb{E}[K^*])\mathrm{e}^t}{1 - \mathbb{E}[K^*]\mathrm{e}^t},$$

which is the moment generating function of $\mathrm{Geo}(1 - \mathbb{E}[K^*])$ and so Lemma 1 holds.

Proof of Theorem 1. Theorem 3 and Lemma 1 establish a proof for Theorem 1.

The tool we shall use to prove Theorem 2 is the contraction method. This method was introduced by Rösler [18] to analyze the Quick Sort algorithm. Soon thereafter, it became a popular method because of the transparency of structure it provides in the limit for processes with complicated distributional recurrences. A broad theory is developed in [15], an exposition in the context of recursive algorithms is given in [17], and [20] provides a survey.

The proof will be in three parts: Theorem 4, Lemma 2, and a concluding argument. In the theorem we prove that X_n^* converges to a limit. The proof of this theorem runs along very similar lines to those in the proof of Theorem 3. We shall be brief in our sketch, and argue points only where there is a divergence from the proof of Theorem 3. In forthcoming Lemma 2 we shall show that the limit represents a contraction mapping in the first-order Wasserstein metric space. Thus, it must have a unique fixed point (distribution function). We cannot use a direct technique like that in Lemma 1 because K^* and T^* are, in general, dependent and it is not straightforward to get an explicit solution for the functional equation of the moment generating function, whereas in (3) the counterpart of T^* is 1, which is independent of $\mathbf{1}_{\{U<K^*\}}$, and an explicit unique solution for the moment generating function is attainable. A concluding argument establishes the unique limit as a perpetuity. In other words, Theorem 1 did not need the full power of the contraction method. It does not need a uniqueness argument at the end, as the uniqueness is in the nature of the limiting equation itself.

Before giving the proof of Theorem 2, let us write recurrence equation (2) in normalized form:

$$X_n^* := \frac{X_n}{n} \stackrel{\mathrm{D}}{=} \frac{X_{K_n}}{K_n}\frac{K_n}{n} + \frac{T_n}{n} = K_n^* X_{K_n}^* + T_n^*. \tag{5}$$

In view of regularity conditions (i) and (ii), the structure of this normalized equation suggest that X_n^* converges to a limiting random variable, say X^*, satisfying distributional equation

$$X^* \stackrel{\mathrm{D}}{=} K^* X^* + T^*, \tag{6}$$

with X^* independent of (K^*, T^*), and the latter pair has the limiting distribution of (K_n^*, T_n^*).

Theorem 4. *Let X_n^* be as defined in (5). Then*

$$X_n^* \xrightarrow{\mathrm{D}} X^*,$$

where X^* satisfies (6).

Proof. Let us take the pairs (K_n^*, K^*), for $n \ge 0$, to be independent. We consider the same for (T_n^*, T^*) and for (X_n^*, X^*). We assume these conditions, together with regularity conditions (i) and (ii), and also that (X_n^*, X^*) are optimal couplings for all $n \ge 0$. Use these variables as realizations in the right-hand sides of (5) and (6). We have

$$\begin{aligned} g_n &:= \mathbb{E}[|(K_n^* X_{K_n}^* + T_n^*) - (K^* X^* + T^*)|] \\ &\le \mathbb{E}[|K_n^* X_{K_n}^* - K^* X^*|] + \mathbb{E}[|T_n^* - T^*|]. \end{aligned}$$

Since all the random variables are defined on the same space, and we are dealing with an optimal coupling, X_n and X^* have a joint distribution, and g_n is well defined. By regularity condition (ii), there exist a positive integer n_0'' and a positive real constant A'' such that

$$\mathbb{E}\big[|T_n^* - T^*|\big] \le \frac{A''}{n^\alpha} \quad \text{for all } n \ge n_0''.$$

https://doi.org/10.1239/aap/1396360110 Published online by Cambridge University Press

Recall here the bound on $\mathbb{E}[|K_n^* - K^*|]$ (established in (4)). In addition, from (6), $\mathbb{E}[X^*] = \mathbb{E}[T^*]/(1 - \mathbb{E}[K^*])$, which, in view of the regularity conditions, is a well-defined positive number. Combining the bounds, we see that

$$\begin{aligned} g_n &\le \mathbb{E}\biggl[\biggl|K_n^* X_{K_n}^* - \biggl(K_n^* - O_{\mathcal{L}^1}\biggl(\frac{1}{n^\alpha}\biggr)\biggr)X^*\biggr|\biggr] + \frac{A''}{n^\alpha} \\ &\le \mathbb{E}[|K_n^*(X_{K_n}^* - X^*)|] + \mathbb{E}\biggl[\biggl|X^* O_{\mathcal{L}^1}\biggl(\frac{1}{n^\alpha}\biggr)\biggr|\biggr] + \frac{A''}{n^\alpha} \\ &= \mathbb{E}[|K_n^*(X_{K_n}^* - X^*)|] + \mathbb{E}[X^*]\mathbb{E}\biggl[O_{\mathcal{L}^1}\biggl(\frac{1}{n^\alpha}\biggr)\biggr] + \frac{A''}{n^\alpha} \\ &\le \mathbb{E}[|K_n^*(X_{K_n}^* - X^*)|] + \frac{\mathbb{E}[T^*]A}{(1 - \mathbb{E}[K^*])n^\alpha} + \frac{A''}{n^\alpha} \end{aligned}$$

for $n \ge \max\{n_0, n_0''\}$ (where n_0 is the integer we used in the proof of Theorem 3). Recall that T^* is integrable, and let $A''' = A\mathbb{E}[T^*]/(1 - \mathbb{E}[K^*]) + A''$. We have

$$g_n \le g_n \mathbb{P}(K_n = n) + \frac{1}{n}\sum_{k=0}^{n-1} k g_k \mathbb{P}(K_n = k) + \frac{A'''}{n^\alpha}.$$

According to Remark 2, the probability $\mathbb{P}(K_n = n)$ is less than 1 for all $n \ge \nu$. So, for all $n \ge n_0''' = \max\{\nu, n_0, n_0''\}$, we can now write

$$g_n \le \frac{(1/n)\sum_{k=0}^{n-1} k g_k \mathbb{P}(K_n = k) + A'''/n^\alpha}{1 - \mathbb{P}(K_n = n)}.$$

Next we carry out an induction to show that, $g_n \le h'/n^\alpha$, if h' is chosen such that, for all $n \ge n_0'''$,

$$\begin{aligned} h' &\ge \frac{A'''}{1 - \mathbb{E}[(K_n^*)^{1-\alpha}]} \\ &\ge \frac{A'''}{1 - \sup_{n \ge n_0'''} \mathbb{E}[(K_n^*)^{1-\alpha}]}, \end{aligned}$$

and by Remark 2, the right-hand side in the last inequality is a well-defined positive number. For large enough h', covering initial conditions at the basis, the induction is complete. Hence, $X_n^* \xrightarrow{\mathrm{D}} X^*$. This completes the proof of Theorem 4.

Lemma 2. (Contraction.) *There is a unique distribution satisfying (6).*

Proof. The conditions

$$\mathbb{E}[|K^*|^s] < 1 \quad \text{and} \quad \mathbb{E}[|T^*|^s] < \infty$$

for some $s \in [1, \infty)$ guarantee $\mathbb{E}[|X^*|^s] < \infty$; see [4, p. 458]. In our case, these conditions are satisfied for $s = 1$. Regularity condition (i) guarantees the former, and the fact that our class of toll functions T_n is $O(n)$ guarantees the latter.

View the right-hand side of (6) as a mapping from the Wasserstein metric space of order 1 (of distribution functions under the Wasserstein distance of order 1) into itself. Let X^* and Y^* be two (integrable) random variables satisfying (6), with distribution functions F_{X^*} and F_{Y^*}

https://doi.org/10.1239/aap/1396360110 Published online by Cambridge University Press

such that X^* is independent of (K^*, T^*) and Y^* is independent of (K^*, T^*). Start with the calculation

$$\mathbb{E}[|(K^*X^* + T^*) - (K^*Y^* + T^*)|] = \mathbb{E}[K^*]\mathbb{E}[|X^* - Y^*|].$$

Condition (i) ensures that $\mathbb{E}[K^*]$ is strictly less than 1. Taking the infimum of every version X^* and Y^* having the respective distribution functions F_{X^*} and F_{Y^*}, we find that

$$\Delta_1(F_{K^*X^*+T^*}, F_{K^*Y^*+T^*}) < \Delta_1(F_{X^*}, F_{Y^*}).$$

Thus, the mapping is contracting. As the mapping is a contraction in a complete metric space, there is a unique fixed point (i.e. distribution function) satisfying (6). This completes the proof of Lemma 2.

Proof of Theorem 2. Having established, in Lemma 2, the uniqueness of the distribution of X^*, we proceed to argue that X^* is a perpetuity. It is shown in [4, pp. 457–459] and [21] that distributional equations of the form (6) unwind into a sum of products of independent random variables (i.e. a perpetuity) as in Theorem 2, provided that $-\infty \le \mathbb{E}[\ln|K^*|] < 0$ and $\mathbb{E}[\ln^+|T^*|] < \infty$. These conditions are satisfied in our case because K^* is supported on $[0, 1]$ and is not a mass at 1, and we are considering only efficient splitting, dealing with, at most, linear toll functions. Theorem 4, Lemma 2, and the above concluding argument complete the proof of Theorem 2.

Remark 4. We can iterate (3) to obtain the perpetuity

$$D_n \xrightarrow{\mathrm{D}} 1 + B_1 + B_1B_2 + B_1B_2B_3 + \cdots,$$

where $\{B_i\}_{i=1}^{\infty}$ are totally independent identically distributed (i.i.d.) Bernoulli random variables with success probability $\mathbb{E}[K^*]$. Of course, the representation of the limit as a geometric random variable in Theorem 1 appeals to a more standard distribution. However, the latter perpetuity representation reinforces the notion that perpetuities may come about in the distribution of several characteristics of leader election algorithms, like D_n and X_n.

5. Examples

We give a few examples arising from practical applications that illustrate the asymptotic theory presented. For some of these examples we shall be able to say a word that goes a bit beyond asymptotics into areas such as exact moments or rates of convergence. We shall present the first example in some detail. The presentation will be briefer for the rest.

5.1. Uniform splitting

In this example we study the behavior of D_n and X_n when the splitting protocol generates a set of size K_n following a discrete uniform distribution on the set $\{1, 2, \dots, n\}$. In this example, we can find some exact moments and say something about rates of convergence. We derive a functional equation for $\phi_n(t)$, the moment generating function of D_n, from distributional equation (1)

$$\mathbb{E}[\mathrm{e}^{tD_n} \mid K_n] = \mathrm{e}^t\left(1 - \frac{K_n}{n}\right) + \mathrm{e}^t\mathrm{e}^{tD_{K_n}}\left(\frac{K_n}{n}\right),$$

with an unconditional expectation satisfying

$$\phi_n(t) := \mathbb{E}[\mathrm{e}^{tD_n}] = \frac{\mathrm{e}^t}{n}\mathbb{E}[K_n\mathrm{e}^{tD_{K_n}}] + \mathrm{e}^t - \frac{\mathrm{e}^t}{n}\mathbb{E}[K_n],$$

valid for $n \geq 2$. Using the fact that K_n is uniformly distributed on the set $\{1, 2, \dots, n\}$, we have the recurrence

$$\phi_n(t) := \frac{\mathrm{e}^t}{n} \sum_{k=1}^{n} k\phi_k(t) + \mathrm{e}^t - \frac{1}{2}(n+1)\mathrm{e}^t.$$

This full-history recurrence involves a telescopic sum. The recurrence is simplified if we subtract from it a version written with $n-1$ replacing n, yielding

$$\phi_n(t) = \frac{(n-1)^2\phi_{n-1}(t) + (n-1)\mathrm{e}^t}{n(n-\mathrm{e}^t)},$$

valid for $n \geq 3$. Under the initial condition $\phi_2(t) = \mathrm{e}^t/(2-\mathrm{e}^t)$, it can be shown by induction that

$$\phi_n(t) = \frac{\mathrm{e}^t}{2-\mathrm{e}^t} - \frac{\mathrm{e}^t}{n(2-\mathrm{e}^t)} + \frac{\mathrm{e}^t\Gamma(2-\mathrm{e}^t)\Gamma(n)}{n\Gamma(n+1-\mathrm{e}^t)}, \tag{7}$$

valid for $n \geq 2$, and $t < \ln 2$. (Such results are first conjectured with the aid of a symbolic manipulation system such as MAPLE®. Once the form is obtained, induction follows easily.)

Derivatives of (7), evaluated at $t = 0$, give us exact moments. The following proposition uses the notation $H_n^{(r)} = \sum_{s=1}^{n} 1/s^r$ for the nth harmonic number of order r (with $H_n^{(1)}$ written customarily as H_n).

Proposition 1. *For $n \geq 3$, we have*

$$\mathbb{E}[D_n] = 2 + \frac{H_n}{n} - \frac{1}{n} - \frac{1}{n^2},$$

$$\operatorname{var}[D_n] = 2 - \frac{1}{n} - \frac{2}{n-1} + \frac{2}{n^2} - \frac{2}{n^3} - \frac{1}{n^4} - \frac{3H_n}{n} + \frac{2H_n}{n-1} + \frac{2H_n}{n^3} - \frac{2H_{n-1}}{n(n-1)} + \frac{H_n^2}{n} - \frac{H_n^2}{n^2} + \frac{H_n^{(2)}}{n}.$$

So, the mean and variance of D_n are asymptotically equivalent to 2. Let U be the standard (continuous) uniform random variable on $(0, 1)$. On a suitable probability space, $K_n = \lceil nU \rceil$. Thus, $K_n/n = U + O(1/n)$, and condition (i) is satisfied (we can take $\alpha = \frac{1}{2}$). And so, $K^* = U$ with mean $\mathbb{E}[K^*] = \frac{1}{2}$. According to Theorem 1, we have

$$D_n \xrightarrow{\mathrm{D}} \operatorname{Geo}\left(\tfrac{1}{2}\right).$$

The limit distribution can also be obtained from the exact moment generating function, via Stirling's approximation of the gamma function. This approach gives rates of convergence. For instance, for all $t < \frac{1}{2}\ln 2$, $\phi_n(t)$ approaches the limiting geometric moment generating function at a rate of $O(n^{-(2-\ln 2)})$. Let us take the time of the execution of machine instructions to perform the steps of the loop body of the generating algorithm as the unit of time. Whence, $K_n + c$ can be taken as the toll T_n (to generate a set of size K_n), where c is the fixed overhead of the algorithm (defining variables, setting up loops, etc.). Thus, we have $T_n/n = \lceil nU \rceil/n + c/n = U + O(1/n)$. Note that here $T^* = U$. Condition (ii) is satisfied (with $\alpha = \frac{1}{2}$), yielding a limiting perpetuity

$$X_n^* \xrightarrow{\mathrm{D}} U_1 + U_1U_2 + U_1U_2U_3 + \cdots,$$

and $\{U_i\}_{i=1}^{\infty}$ are totally i.i.d. continuous uniform $(0, 1)$ random variables. (Note, the limiting perpetuity also has Dickman's distribution.)

5.2. Power distribution splitting

The uniform distribution is a member of a class called power distributions. A power distribution with parameter $\theta \geq 0$ has a probability mass function

$$\mathbb{P}(K_n = k) = \frac{k^\theta}{\sum_{j=1}^{n} j^\theta}, \qquad k = 1, 2, \ldots, n.$$

The discrete uniform distribution is in fact a special case of the power distribution with $\theta = 0$. A calculation shows that K_n/n converges in distribution to the continuous random variable K^*, which has a Beta$(\theta + 1, 1)$ distribution, with mean value $(\theta + 1)/(\theta + 2)$. All the regularity conditions for Theorems 1 and 2 are satisfied and, as $n \to \infty$, we have

$$D_n \xrightarrow{\text{D}} \text{Geo}\left(\frac{1}{\theta + 2}\right),$$

$$X_n^* \xrightarrow{\text{D}} V_1 + V_1 V_2 + V_1 V_2 V_3 + \cdots,$$

and $\{V_i\}_{i=1}^{\infty}$ are i.i.d. beta$(\theta + 1, 1)$ random variables.

5.3. Binomial splitting

This is a classic example and several of its properties have been studied; see [5], [7], [9], and [16]. The mechanics of the splitting have been discussed in the introduction. In the crudest variation, if all candidates flip Tails, they are eliminated, resulting in no winner. In this variation, K_n has the binomial distribution underlying n independent identically distributed experiments, with success probability p per experiment. The conditions for Theorem 1 are all met and, as $n \to \infty$, we have

$$D_n \xrightarrow{\text{D}} \text{Geo}(q).$$

This result appears in [9], where the authors further show that the lower-order asymptotics in the distribution function may have oscillations.

The total cost or speed of the algorithm is measured by the number of independent coin flips. In a serial environment, where all the contestants share one coin, which they pass from one to the next, the cost of the first round is n (coin flips). So, $T_n/n \equiv 1$, and all the conditions for Theorem 2 are met. We have the first-order asymptotic

$$\frac{X_n}{n} \xrightarrow{\text{P}} 1 + \sum_{j=1}^{\infty} \prod_{i=1}^{j} p = \frac{1}{q}$$

as given in [9]. Second-order asymptotics are also given in [9], specifying a rate of convergence for this weak law in the form of a central limit theorem.

5.4. An example with a splitting distribution with atoms

Suppose that, after the first round, the advancing set has a size distributed as

$$K_n = \begin{cases} 0 & \text{with probability } \frac{1}{3}, \\ k \in \{1, 2, \ldots n - 1\} & \text{with probability } 1/3(n-1), \\ n & \text{with probability } \frac{1}{3}. \end{cases}$$

Let U be a continuous uniform $(0, 1)$ random variable. So, K_n/n converges in $\mathcal{L}_1$ to K^*, a mixture of three variables of values 0, U, and 1, where each of the three variables has $\frac{1}{3}$

probability of being the outcome. We can generate such a mixture from two independent uniform (0, 1) random variables U and V by letting

$$K_n = \mathbf{1}_{\{1/3<V\le 2/3\}}\lceil (n-1)U\rceil + \mathbf{1}_{\{2/3<V<1\}}n,$$

where $\mathbf{1}_{\mathcal{E}}$ is the indicator function that takes the value 1 if $\mathcal{E}$ occurs, and 0 otherwise. Thus, we have

$$K_n^* = \frac{K_n}{n} = \mathbf{1}_{\{1/3<V\le 2/3\}}U + \mathbf{1}_{\{2/3<V<1\}} + O\left(\frac{1}{n}\right) =: K^* + O\left(\frac{1}{n}\right).$$

The limit distribution K^* has atoms (i.e. jumps in the distribution function) of magnitude $\frac{1}{3}$ at 0 and at 1. Here $\mathbb{E}[K^*]$ is $\frac{1}{2}$. The conditions for Theorem 1 are met, and we have, as $n \to \infty$,

$$D_n \xrightarrow{\mathrm{D}} \mathrm{Geo}\big(\tfrac{1}{2}\big).$$

We also have $T_n = K_n + c$ (where c is a constant overhead). That is, $T_n^* = K_n^* + O(1/n) = K^* + O(1/n)$. The conditions for Theorem 2 are met, giving a limiting perpetuity

$$\frac{X_n}{n} \xrightarrow{\mathrm{D}} V_1 + V_1V_2 + V_1V_2V_3 + \cdots,$$

where $\{V_i\}_{i=1}^{\infty}$ is a set of totally i.i.d. random variables, all having the mixed distribution function

$$F(x) = \begin{cases} 0 & \text{if } x < 0, \\ \frac{1}{3} & \text{if } x = 0, \\ \dfrac{x}{3} + \dfrac{1}{3} & \text{if } 0 < x < 1, \\ 1 & \text{if } x \ge 1. \end{cases}$$

5.5. An almost-deterministic ladder example

Many real-life tournaments, such as the Wimbledon Championship, are organized in ladders. Of course, in these tournaments, advancing to later rounds is based on skill. However, many other tournaments and games in local communities are organized in ladders, where advancing to later rounds is based on luck. In these contests, n is not always guaranteed to be a power of 2. An instance of such a ladder is the following. Suppose that n is even, then a ladder can be created by asking contestants i and $i+1$ to compete, for $i = 1, 3, \dots, n-1$, with only one of them advancing to the next round. For instance, say a moderator flips an unbiased coin and chooses contestant i if the toss is heads, and chooses $i+1$ if the toss is tails. If n is odd, one contestant gets through first without competing (sometimes called a *bye*), and the contestants are renumbered $1, \dots, n-1$ (even), and the procedure above for an even number of contestants is applied. As we are committed in this manuscript to fair leader election algorithms, we construct our ladder to advance $\lceil \frac{1}{2}n\rceil$ contestants, and if n is odd, the bye is generated by a moderator uniformly at random from among the n participants, and thus all contestants have an equal chance to win the contest. In this example, the size of the advancing set is deterministic, but elements of randomness appear in the content of that set.

In our fair ladder, $K_n = \lceil \frac{1}{2}n\rceil$, and so $K_n/n = \frac{1}{2} + O(1/n)$. It is easy to check that all the conditions for Theorem 1 are met; the duration of any contestant in the competition converges in distribution

$$D_n \xrightarrow{\mathrm{D}} \mathrm{Geo}\big(\tfrac{1}{2}\big).$$

The advancing set is created in time $T_n = \lfloor \frac{1}{2}n \rfloor + O(1)$. Thus, $T_n/n = \frac{1}{2} + O(1/n)$, and all the conditions for Theorem 2 are met. The scaled overall cost converges to a perpetuity

$$\frac{X_n}{n} \xrightarrow{\text{P}} \frac{1}{2} + \frac{1}{2}\sum_{j=1}^{\infty}\prod_{i=1}^{j}\frac{1}{2} = 1.$$

Acknowledgements

Professor Dr Uwe Rösler has patiently read several drafts of this manuscript. The authors are very thankful to Uwe, whose many insightful comments not only added accuracy and rigor but also improved the presentation and rid the manuscript of some ambiguous notation. Uwe carefully guided us to fine-tune some of the techniques, for which we are most grateful. We also thank an anonymous referee for many constructive comments.

References

[1] Alsmeyer, G., Iksanov, A. and Rösler, U. (2009). On distributional properties of perpetuities. *J. Theoret. Prob.* **22,** 666–682.
[2] Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9,** 1196–1217.
[3] Elmasry, A. and Mahmoud, H. (2011). Analysis of swaps in radix selection. *Adv. Appl. Prob.* **43,** 524–544.
[4] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events*. Springer, Berlin.
[5] Fill, J. A., Mahmoud, H. M. and Szpankowski, W. (1996). On the distribution for the duration of a randomized leader election algorithm. *Ann. Appl. Prob.* **6,** 1260–1283.
[6] Itoh, Y. and Mahmoud, H. M. (2003). One-sided variations on interval trees. *J. Appl. Prob.* **40,** 654–670.
[7] Janson, S. and Szpankowski, W. (1997). Analysis of an asymmetric leader election algorithm. *Electron. J. Combin.* **4,** Research Paper 17.
[8] Janson, S., Lavault, C. and Louchard, G. (2008). Convergence of some leader election algorithms. *Discrete Math. Theoret. Comput. Sci.* **10,** 171–196.
[9] Kalpathy, R., Mahmoud, H. M. and Ward, M. D. (2011). Asymptotic properties of a leader election algorithm. *J. Appl. Prob.* **48,** 569–575.
[10] Kirschenhofer, P. and Prodinger, H. (1998). Comparisons in Hoare's Find algorithm. *Combin. Prob. Comput.* **7,** 111–120.
[11] Louchard, G. and Prodinger, H. (2009). The asymmetric leader election algorithm: another approach. *Ann. Comb.* **12,** 449–478.
[12] Louchard, G., Martínez, C. and Prodinger, H. (2011). The Swedish leader election protocol: analysis and variations. In ANALCO11—Workshop on Analytic Algorithmics and Combinatorics, SIAM, Philadelphia, PA, pp. 127–134.
[13] Louchard, G., Prodinger, H. and Ward, M. D. (2012). Number of survivors in the presence of a demon. *Period. Math. Hungar.* **64,** 101–117.
[14] Mahmoud, H. M. (2010). Distributional analysis of swaps in Quick Select. *Theoret. Comput Sci.* **411,** 1763–1769.
[15] Neininger, R. and Rüschendorf, L. (2004). A general limit theorem for recursive algorithms and combinatorial structures. *Ann. Appl. Prob.* **14,** 378–418.
[16] Prodinger, H. (1993). How to select a loser. *Discrete Math.* **120,** 149–159.
[17] Rachev, S. T. and Rüschendorf, L. (1995). Probability metrics and recursive algorithms. *Adv. Appl. Prob.* **27,** 770–799.
[18] Rösler, U. (1991). A limit theorem for 'Quicksort'. *RAIRO Inform. Théor. Appl.* **25,** 85–100.
[19] Rösler, U. (2004). QUICKSELECT revisited. *J. Iranian Statist. Soc.* **3,** 271–296.
[20] Rösler, U. and Rüschendorf, L. (2001). The contraction method for recursive algorithms. *Algorithmica* **29,** 3–33.
[21] Vervaat, W. (1979). On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables. *Adv. Appl. Prob.* **11,** 750–783.