Perpetuities in Fair Leader Election Algorithms

by Ravi Kalpathy

M.S. in Electrical Engineering, August 2005, The George Washington University
M.S. in Statistics, May 2007, The George Washington University

A Dissertation submitted to

The Faculty of
The Columbian College of Arts and Sciences
of The George Washington University
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

August 31, 2013

Dissertation directed by

Hosam M. Mahmoud
Professor of Statistics

The Columbian College of Arts and Sciences of The George Washington University certifies that Ravi Kalpathy has passed the Final Examination for the degree of Doctor of Philosophy as of May 3, 2013. This is the final and approved form of the dissertation.

Perpetuities in Fair Leader Election Algorithms

Ravi Kalpathy

Dissertation Research Committee:

Hosam M. Mahmoud, Professor of Statistics,
Dissertation Director

Reza Modarres, Professor of Statistics,
Committee Member

Srinivasan Balaji, Assistant Professor of Statistics,
Committee Member

Acknowledgments

It is impossible for me to explain in words about how much I appreciate the guidance, knowledge, support, and encouragement I received from Professor Mahmoud and my family. Thank you so much.

I would like to thank Professor Ward who has been a good friend and a great mentor. Mark, I greatly appreciate your guidance on everything—Mathematics, LaTeX, MAPLE, and the list goes on. Many thanks to the Ward family for your kind hospitality during my several visits to Purdue University.

Special thanks to my examiners and readers: Professor Rosenkrantz, Professor Barta, Professor Modarres, Professor Balaji, and Professor Larsen for serving on my examination committee. Your suggestions and feedback greatly improved this dissertation.

Finally, I would like to thank the GW CCAS and Statistics department for all the logistical support, GW OGSAF for the Dissertation Fellowship, and my friends and colleagues: David, Gabe, Dan, Nicole, Ana, Evgeniya, SiHun, Alessandro, Sanaa, Ghassan, Vicky, Nathan, Cat, Clément, Dani, Bijit, Anna, Owen, Sertan, Sharmeen, Bumsoo, Nassos, Wenjing, Isaac, Sabrina, Jason, Ben, Somak, Pinar, Erika, Hareen, Tong, Haojin, Iveta, Melinda, Swetha, Calley, Fowzia, Max, Laura, Behram, Alicia, Jared, Samantha, Erin, Paige, Chris, Tyler, Courtnay, Hugo (maybe I should stop here and say), and 8233 others for all the great memories!

Washington, DC *Ravi Kalpathy*

May 2013

Abstract

Perpetuities in Fair Leader Election Algorithms

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 two steps are commonly referred to as the contraction method.

Our main contribution is a unifying treatment for leader election algorithms and see how perpetuities (a construct that appears in insurance and the mathematics of finance) naturally come about. In the past, leader election algorithms have been discussed via a variety of methods, such as analytic techniques, poissonization, integral transforms, etc., which may work for some splitting protocols but not the others. Our treatment covers one-sided leader election algorithms represented by a certain stochastic recurrence with linearly bounded toll functions. Our methodology can be extended, with some adaptation, to other one-sided algorithms of similar structure.

We give several examples to illustrate the asymptotic theory presented. In the first example, we study the scenario when the size of the set produced at each stage follows a discrete uniform distribution. In this example we go a little beyond asymptotics

into areas such as exact moments and rates of convergence. In the second example, we obtain the limit distributions for the duration and the overall cost for the power distributions. The third example is the classic binomial case, where the leader election is done via coin flipping. We obtain the geometric limit distribution for the duration and further show that the lower order asymptotics may have fluctuations. For the total cost we obtain the first-order and second-order asymptotics. We also specify a rate of convergence for the weak law in the form of a central limit theorem. The fourth example explores the asymptotic distribution for the duration and cost when the advancing set has a size which has a distribution involving atoms. In the final example, we obtain the limit distributions for the duration and cost when the selection process is carried out in the form of a ladder, also known as tournament brackets.

For the duration of a contestant, the only aspect of the limiting distribution of the selection algorithm (also called splitting protocol) that enters the picture in the geometric limit is its mean. All distributions of the splitting protocols that have the same limiting mean, will have the same geometric limit law. This shows that fair leader election is a robust algorithm across a wide variety of splitting protocols. For instance, the uniform splitting, binomial splitting (with an unbiased coin), and certain ladder contests, albeit remarkable differences among these splitting protocols, all have the same geometric limit for the duration of participants.

Findings in the dissertation appear in Kalpathy, Mahmoud, and Ward [55], and Kalpathy and Mahmoud [53]. A preliminary draft (Kalpathy and Mahmoud [52]) was presented at ANALCO'13 meeting and an extended abstract appears in the Proceedings of the Tenth ACM-SIAM Workshop on Analytic Algorithmics and Combinatorics (ANALCO).

Contents

List of Figures

Chapter 1

Introduction

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. Leader election plays an important role in distributed systems, particularly in the context of wireless and wired communication networks, where a leader needs to be chosen to supervise synchronization and communication. In these networks, failures tend to occur frequently because of rapidly changing network topology, hardware and software problems, node deployment issues, etc. If we look at the big picture, randomized divide and conquer algorithms have numerous manifestations. In the context of distributed computing, leader election algorithms play a vital role. The cost measures that appear in these algorithms have several important applications that can impact communication and control in information processing systems. For instance, the underlying tree structure that accompanies these algorithms (the size, height, depth, profile, etc.) directly translates to memory usage and computational time. In the language of probability, these manifestations appear as stochastic recursive distributional equations for the cost measures.

1.1 Background

The common scenario in leader election situations is the following. There is a number of contestants who will compete fairly to elect a winner (and in some variations they may all lose the election, resulting in none). 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 any contestant to win is the same.

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 re*trie*val tree (also called the trie, see Mahmoud [69]) 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 itself. 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 incomplete tree forms the backbone for many one-sided algorithms, such as for example the one that underlies the algorithm Find in Rösler [85], which identifies order statistics in a data set.

There is a plethora of research work on leader election algorithms, each dealing with a specific instance of the problem. Prodinger [78] found the average behavior of several characteristic properties for a certain leader election algorithm that flips unbiased coins. He coined the terminology *incomplete trie* for the tree structure underlying the elimination process. Using analytic methods, the exact and asymptotic average for the size of the tree, i.e. the number of nodes, the depth (also known as the height in the literature) or the number of rounds, and the cost measured in terms of the total number of coin flips were obtained. Fill, Mahmoud, and Szpankowski [28] uses analytic and probabilistic methods to obtain the oscillating distribution of the height of a random incomplete trie constructed using unbiased coins. Janson and Szpankowski [46] follows up and analyzed the height for biased coins using analytic techniques. Mohamed [73] also investigated the biased-case scenario for the height, but used probabilistic methods. More recently, Louchard and Prodinger [66] used analytic methods to study the number of rounds in a coin flipping selection algorithm that occurs in the presence of a demon. The demon randomly eliminates some contestants, and the algorithm stops, if all contestants flip tails and are declared winners (survivors) of the competition. Louchard, Prodinger, and Ward [68] follows up with a sequel complementing the above mentioned paper where they precisely analyze the distribution and all moments of the number of survivors in a selection process that occurs in the presence of a demon. Louchard, Martínez, and Prodinger [67] study another variant called the Swedish leader election protocol, where they analyze several parameters like the probability of success, the expected number of rounds, the expected number of players still playing by the time the protocol fails, etc. using analytic methods. Kalpathy, Mahmoud, and Ward [55] uses analytic and probabilistic techniques to study a leader election algorithm using biased coins for the duration a particular player survives in the competition and the total cost involved in the selection process.

As we can see, in the analysis of several specific variants of leader election algorithms, the typical parameter that research focused on is the number of rounds till termination. However, several other properties are also of interest, and more recently the research branched out to consider other parameters. 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. There is a need for a broader framework to establish results for classes of such algorithms. The recent work by Janson, Lavault, and Louchard [47] 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 dissertation 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. By duration we mean the prospect 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.

Another aspect of this dissertation is about perpetuities. There is a whole body of literature on this topic and our goal is to provide a unified treatment for the limit distributions of cost measures that appear in one-sided divide and conquer algorithms (specifically speaking, leader election algorithms in theoretical computer science) and see how perpetuities naturally come about.

In insurance and mathematical finance, sums of the type

$$S := B_1 + A_1B_2 + A_1A_2B_3 + A_1A_2A_3B_4 + \cdots ,$$

where $\{(A_i, B_i)\}_{i=1}^{\infty}$ are i.i.d. $\mathbb{R}^2$-valued random variables on some probability space $(\Omega, \mathcal{F}, P)$ satisfying some regularity conditions are called *perpetuities*. The A_i's are interpreted as discount factors, and B_i's represent the payments. These stochastically discounted sums represents the present value of a future long-term guaranteed payment, i.e. *perpetual payment streams*, and hence the name *perpetuity*. The sum S also solves a stochastic fixed point equation

$$S \stackrel{\mathcal{L}}{=} AS + B,$$

with S independent of (A, B) in the sense that the distribution of S is the fixed point solution of the above recursive distributional equation. By this we mean,

$$\mathcal{L}(S) = \mathcal{L}(AS + B),$$

where for a random variable Z, $\mathcal{L}(Z)$ is the probability law (distribution) of Z. In a general setting, S represents the current quantity of a fluctuating resource, B_i is the initial quantity at each stage, A_i is the intrinsic rate of increment or decay due to a change in the environment, and both A_i and B_i are subject to stochastic fluctuations. See Kesten [58], Vervaat [91], Embrechts and Goldie [26], Goldie and Grübel [33], Devroye [20], Knape and Neininger [61], Alsmeyer, Iksanov, and Rösler [2], and Hitczenko and Wesołowski [41] for a general background.

In the grand scheme of things, perpetuities are an interesting class in probability theory because it has a long and diverse lineage. Vervaat [91] studies the relation between perpetuities and *infinitely divisible distributions*. Jurek [48] studies the relationship between perpetuities and *self-decomposable distributions*. It is important to note that the latter class extends the well studied class of *stable distributions*. See Jurek [48], and Iksanov, Jurek, and Schreiber [44] for the relationship. See Nolan [75] for a comprehensive treatment on stable distributions. Letac [64], Chamayou and Letac [15], and Diaconis and Freedman [23] studies the link between perpetuities and stationary distributions of recursive *Markov chains* obtained by iterating random functions.

Perpetuities and stochastic fixed point equations appear and play an important role in several applications. We list a few applications and a few references (the list is not exhaustive):

(1) Agriculture: Puri [79], Rachev and Todorovic [83], and Todorovic and Gani [90].

(2) Analytic and probabilistic number theory: Chamayou [14], de Bruijn [19], and Donnelly and Grimmett [22].

(3) Biology and sociology: Cavalli-Sforza [12], and Cavalli-Sforza and Feldman [11].

(4) Combinatorial structures: Arratia, Barbour, and Tavaré [3].

(5) Communication and control theory: Kalman [50], and Maulik and Zwart [72].

(6) Environmental studies: Rachev and Samorodnitsky [80], Rachev and Todorovic [83], Todorovic [89], and Todorovic and Gani [90].

(7) Fractals: Diaconis and Freedman [23].

(8) Hydrology: Lawrance and Kottegoda [63], and Weiss [93].

(9) Insurance and mathematical finance: Dufresne [24], Embrechts, Klüppelberg, and Mikosch [27], and Geman and Yor [31].

(10) Neuroscience: Hitczenko and Medvedev [40].

(11) Nuclear technology: Chamayou and Schorr [16], and Paulson and Uppuluri [76].

(12) Physics: Carmona, Petit, and Yor [10], and Chandrasekhar and Münch [17].

(13) Probability and probabilistic analysis of algorithms: Alsmeyer, Iksanov, and Rösler [2], Blanchet and Sigman [8], Brandt, Franken, and Lisek [9], Devroye [20], Devroye and Fawzi [21], Elmasry and Mahmoud [25], Embrechts and Goldie [26], Fill and Huber [29], Goldie [32], Goldie and Grübel [33], Goldie and Maller [34], Grübel and Rösler [38], Hitczenko [39], Hitczenko and Wesołowski [41], Hwang and Tsai [42], Iksanov and Jurek [43], Jurek [48], Knape and Neininger [61], Kuba and Panholzer [62], Mahmoud [70], Mahmoud, Modarres, and Smythe [71], Rachev and Samorodnitsky [80], Rachev and Rüschendorf [81], Rachev and Rüschendorf [82], Rösler [84], Rösler and Rüschendorf [86], and Vervaat [91].

(14) Queueing theory: Lindley [65], and Preater [77].

(15) Random matrix theory: Chamayou and Letac [15], Grintsevichyus [36], Kesten [58], and Letac [64].

(16) Random walk in random environment: Solomon [87].

(17) Shot-noise processes in electronics: Iksanov and Jurek [43], and Takács [88]

(18) Statistics: Aebi, Embrechts, and Mikosch [1], Baron and Rukhin [5], Embrechts, Klüppelberg, and Mikosch [27], Grübel and Pitts [37], and Kagan, Linnik, and Rao [49].

1.2 Motivation and overview

Our quest for a broad theory is essentially driven by the following questions: What if the algorithm for leader election is not based on coin flipping? What if the size of the advancing set produced at each stage is general, not necessarily binomial? Do we expect comparable results? What methods succeed? Our main contribution is to present a unifying treatment for leader election algorithms, and see how perpetuities naturally come about. Using the contraction method and associated techniques we bring together a variety of different leader election algorithms under one umbrella. From a theoretical viewpoint, the main advantage is the portability of this methodology because it encapsulates a wide variety of scenarios. From a practical viewpoint, algorithmic portability is a key issue in cost reduction.

In the past, leader election algorithms have been discussed via a variety of methods, such as analytic techniques, poissonization, integral transforms, etc., which may work for some splitting protocols but not the others. 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, as was done in Kirschenhofer and Prodinger [60], Rösler [85], Mahmoud [70], and Elmasry and Mahmoud [25]. A variety of other one-sided algorithms may be approached by the methods we discuss in this dissertation, such as the random walk on interval trees in Itoh and Mahmoud [45], which have a continuous flavor.

1.3 Outline

The dissertation is organized as follows. In Chapter 2 we review some of the tools and techniques used in our study. These include the contraction method and the associated probability metrics. Some additional tools are discussed in the chapters as needed. Chapter 2 has two sections. Section 2.1 overviews the contraction method, which is a basic tool in this dissertation. This section further has two subsections: Subsection 2.1.1 is for definitions, and Subsection 2.1.2 is for Banach's fixed point theorem. Section 2.2 overviews the Wasserstein metric, and its two Subsections are about definitions and properties. In Chapter 3 we develop some general theory which formalizes the leader election problem in terms of recursive distributional equations. In particular, we define and study two parameters associated with our problem—the duration of an individual contestant and the total time or cost of the competition. In Section 3.3, we give the main results in two subsections. Subsection 3.3.1 is for a set of regularity conditions to derive the main theorems, which are presented in Subsection 3.3.2. Section 3.4 is dedicated to the technical proofs of the main results. In Chapter 4, we present five illustrative examples in five sections, one example per section. We end the dissertation in Chapter 5, where future work is outlined. Proofs of some technical points appear in the appendices.

Chapter 2

Mathematical preliminaries

In this chapter we review some of the mathematical tools that we shall use in later chapters of this dissertation.

2.1 Contraction method

The asymptotic analysis of algorithms could be carried out using several methods and one useful technique is the contraction method. Rösler [84], Rachev and Rüschendorf [81], Rösler and Rüschendorf [86], and Neininger and Rüschendorf [74] are some key references for this method. The method is based on contraction properties of the stochastic recursive algorithm using a suitable probability metric. Toward a simple exposition we present the mechanics of this method restricted to the context of the type of stochastic recurrence that appears in leader election algorithms. The technique goes through the following steps:

Step 1: We find the correct normalization of the algorithm and determine the stochastic recurrence for the normalized algorithm. For instance, starting with

$$X_n \stackrel{\mathcal{L}}{=} X_{K_n} + T_n,$$

we obtain the stochastic recurrence,

$$X_n^* := \frac{X_n}{n} \stackrel{\mathcal{L}}{=} \frac{K_n}{n} \times \frac{X_{K_n}}{K_n} + \frac{T_n}{n} = K_n^* X_{K_n}^* + T_n^*.$$

The normalization could be a centering and scaling or just a scaling as shown above. The center and scale values are usually obtained by solving a recurrence equation for the moments, or by analytic techniques, or using renewal theory.

Step 2: Assuming K_n^* converges to K^* and T_n^* converges to T^* (at some appropriate rate), we guess the limiting structure of the normalized algorithm such that the normalized version X_n^* converges to a nondegenerate limit X^*. It is plausible to see convergence according to the following chart:

$$\begin{array}{ccccc} X_n^* & \overset{\mathcal{L}}{=} & K_n^* \, X_{K_n}^* + T_n^*, \\ \downarrow & & \downarrow \quad \downarrow \qquad \downarrow \\ X^* & \overset{\mathcal{L}}{=} & K^* \;\; X^* + T^*, \end{array} \tag{2.1}$$

with X^* independent of (K^*, T^*). So, X^* has a distribution that solves (2.1).

Step 3: We confirm the guess by actually proving the convergence of X_n^* to X^* which needs some technical work. The basic idea is to show that for a suitable choice of a probability metric, we have a contraction on the space of distribution functions. The metric has to reflect the structure of the algorithm. Since we are dealing with an affine function,[1] we should choose metrics with some desirable properties, so that the random variables could be coupled[2] together. The Wasserstein metric described in section 2.2 serves our purpose. This is the technically involved step in the methodology. We accomplish this step using some regularity conditions augmented with an induction argument.

Step 4: One can view the right-hand side of (2.1) as a mapping from the space of distribution functions to itself. The idea is to show that this mapping is contracting, hence by the Banach's fixed point theorem, there is a unique fixed point that solves

[1] An affine function is a linear function plus a translation.

[2] Two random variables are said to be coupled, if they are defined on the same probability space.

this equation.

2.1.1 Definitions

Before we get to Banach's fixed point theorem, we need the following definitions. In what follows X is a nonempty set, and d is a distance on it.

Definition 2.1 *Let (X, d) be a metric space and $F : X \to X$ be a mapping. The point $x \in X$ is called a fixed point for F, if $x = F(x)$.*

Definition 2.2 *A metric space (X, d), in which every Cauchy sequence is convergent, is said to be a complete metric space.*

Definition 2.3 *Let (X, d) be a metric space. A map $F : X \to X$ is called a Lipschitzian on X, if $\exists\ \beta \geq 0 :\ d(F(x), F(y)) \leq \beta d(x, y),\ \forall\ x, y \in X$. The smallest β is called the Lipschitz constant of F. If this constant is strictly less than 1, we say that the map F is a contraction.*

Definition 2.4 *Let $F : X \to X$ be a map of X into itself. For any given $x \in X$ and $n \in \{0, 1, \ldots\}$, we define $F^n(x)$, the n^{th} iterate of x under F, inductively by $F^0(x) = x$, and $F^{n+1}(x) = F(F^n(x))$.*

2.1.2 Banach's fixed point theorem

We now state the theorem from Granas and Dugundji [35].

Theorem 2.1 *Let (X, d) be a complete metric space, and let $F : X \to X$ be a contraction. Then F has a unique fixed point $u \in X$, and $F^n(x) \to u$ for any $x \in X$.*

Intuitively, since all distances are reduced by a factor strictly less than 1, repeated application of F will contract the distances so that the existence of a unique fixed point becomes reasonable.

In our context of leader election algorithms, our complete metric space is a space of distribution functions, and the metric would be the Wasserstein metric, discussed next.

2.2 Wasserstein metric

The above framework needs suitable metrics with nice properties. One metric that proved effective in this area is the Wasserstein metric, which we discuss briefly in this section. Bickel and Freedman [6] provides a survey.

2.2.1 Definition

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 having the respective distribution functions F and G (with $||\,.\,||_k$ being the usual $\mathcal{L}_k$ norm). In what follows we use F_Y for the distribution function of a random variable Y.

2.2.2 A convergence property

The metric Δ_k metrizes both weak convergence and convergence of k^{th} absolute moments. That is, for a sequence of random variables W_n, the convergence of k^{th} order Wasserstein distances between F_{W_n} and F_W to 0 implies convergence in law (or in distribution), which we denote by $W_n \xrightarrow{\mathcal{L}} W$, as well as convergence of the k^{th} absolute moments, i.e., $\mathbf{E}[|W_n|^k] \to \mathbf{E}[|W|^k]$.

Chapter 3

A theory for leader election

Assume 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 moderator are fair, in the sense that, given $K_n = k$, all subsets of size k are equally likely choices. The cost (number of steps) of the operations for generating the 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.

3.1 Duration of a contestant

Let $D_{n,j}$ be the number of rounds or the time duration the jth contestant survives. Under a fair splitting policy, all contestants are equally likely to be selected to advance, and we have $D_{n,j} \stackrel{\mathcal{L}}{=} D_{n,1}$ (here $\stackrel{\mathcal{L}}{=}$ denotes equality in law). So, we shall develop results for $D_n := D_{n,1}$ and drop 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 $\frac{\binom{n-1}{K_n-1}}{\binom{n}{K_n}} = \frac{K_n}{n}$, 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:

$$D_n \stackrel{\mathcal{L}}{=} \begin{cases} 1 + D_{K_n}, & \text{with probability } \dfrac{K_n}{n}; \\ \\ 1, & \text{with probability } 1 - \dfrac{K_n}{n}. \end{cases}$$

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

$$D_n \stackrel{\mathcal{L}}{=} \mathbf{1}_{\{U < \frac{K_n}{n}\}} D_{K_n} + 1, \tag{3.1}$$

here, for any $n \geq 2$, K_n is a random variable with a given distribution on $0, \ldots, n$, and K_n, U, D_i are independent, for all $i < n$; all the random variables are defined on the same probability space. So, if $\mathbf{Prob}(K_n = n) < 1$, for all $n \geq 1$, this gives a recursive definition for the distribution of D_n: In this case the distribution of D_n is nontrivial and, for $\ell \geq 1$, is determined by

$$\begin{aligned} \mathbf{Prob}(D_n = \ell) &= \sum_{k=0}^{n} \mathbf{Prob}(\mathbf{1}_{\{U < \frac{K_n}{n}\}} D_{K_n} + 1 = \ell \,|\, K_n = k)\, \mathbf{Prob}(K_n = k) \\ &= \sum_{k=0}^{n} \mathbf{Prob}(\mathbf{1}_{\{U < \frac{k}{n}\}} D_k = \ell - 1)\, \mathbf{Prob}(K_n = k), \end{aligned}$$

which upon reorganization becomes

$$\begin{aligned}\mathbf{Prob}(D_n = \ell) &= \mathbf{Prob}(D_n = \ell - 1)\,\mathbf{Prob}(K_n = n) \\ &\quad + \sum_{k=0}^{n-1} \frac{k}{n}\,\mathbf{Prob}(D_k = \ell - 1)\,\mathbf{Prob}(K_n = k),\end{aligned}$$

providing an inductive definition of the distribution of D_n.

3.2 The total cost

Let X_n be the total cost till termination (i.e., till a winner is chosen or the moderator declares 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 specific contexts in the sequel.

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

$$X_n \stackrel{\mathcal{L}}{=} X_{K_n} + T_n, \tag{3.2}$$

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 $\mathbf{Prob}(K_n = n) < 1$, this gives a recursive definition for the distribution of X_n, in a manner very similar to how we showed that the distribution of D_n is well defined.

Under suitable normalization, a random variable satisfying a distributional recurrence of the type (3.2) usually leads to a stochastic fixed point equation

$$X^* \overset{\mathcal{L}}{=} K^* X^* + T^*,$$

with X^* on the right-hand side independent of (K^*, T^*), and the latter pair has the limiting distribution of $(K_n/n, T_n/n)$; in the context of leader election, T^* is integrable. In this context, X^* will be integrable, too. A random variable satisfying the latter distributional equation is called a *perpetuity*, a construct that appears in insurance and the mathematics of finance (see Embrechts, Klüppelberg, and Mikosch [27]), in probability theory (see Alsmeyer, Iksanov, and Rösler [2], which is quite relevant to our work), and in many other areas, as discussed in section 1.1.

3.3 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.

3.3.1 A set of regularity conditions

When we say 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 $\mathbf{E}[|Y_n|] \le C|g(n)|$, 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 (see Appendix A):

(i) The advancing set size satisfies

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

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

(ii) The toll function satisfies

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

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 $\mathcal{L}_1$ 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 only considering 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 3.1: *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 Chapter 4.*

Remark 3.2: *Note that regularity condition (i) implies that, for some* $\nu \geq 1$, $\sup_{n\geq\nu} \mathbf{E}[(K_n^*)^{1-\alpha}] < 1$, *as follows: From condition (i), we also have* $K_n^* \xrightarrow{\mathcal{L}} K^*$. *For* $\alpha \in (0,1)$, *the function* $x \mapsto x^{1-\alpha}$ *is bounded and continuous on* $[0,1]$. *Hence weak convergence implies* $\mathbf{E}[(K_n^*)^{1-\alpha}] \to \mathbf{E}[(K^*)^{1-\alpha}]$. *Since* $\mathbf{E}[K^*] < 1$, *and* $0 \leq K^* \leq 1$, *we obtain* $\mathbf{E}[(K^*)^{1-\alpha}] < 1$. *Hence, the convergence* $\mathbf{E}[(K_n^*)^{1-\alpha}] \to \mathbf{E}[(K^*)^{1-\alpha}]$ *implies* *the existence of a* $\nu \geq 1$*, such that* $\sup_{n\geq\nu} \mathbf{E}[(K_n^*)^{1-\alpha}] < 1$*. This technical point is instrumental in building induction proofs of convergence.*

3.3.2 The main theorems

The first main result is represented using the notation $\mathrm{Geo}(p)$, for the geometric random variable with success probability p. As before, we use the symbol $\xrightarrow{\mathcal{L}}$ to denote convergence in law (distribution). Any of the splitting protocols we consider takes only n as input, and produces a random number K_n and a random subset of $\{1, \dots, n\}$ of size K_n, in such a way that all subsets of size K_n are equally likely. We shall use the adjective "fair" to mean that conditionally, if $K_n = k$, all subsets of size k are equally likely. A particular leader election algorithm therefore is associated with sequences of distributions for K_n and T_n. We shall assume in what follows that these sequences are known in advance.

Theorem 3.1 *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, till one leader or none is elected. Assume* K_n *follows regularity condition (i). Let* D_n *be the duration (number of rounds) a contestant stays in the competition. We then have*

$$D_n \xrightarrow{\mathcal{L}} \mathrm{Geo}\big(1 - \mathbf{E}[K^*]\big).$$

Remark 3.3: *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*[1] *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, albeit remarkable differences among these splitting protocols, all have* $\mathrm{Geo}(\frac{1}{2})$ *as limit for the duration of participants.*

Theorem 3.2 *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, till one leader or none is elected. Assume K_n follows regularity condition (i). Suppose, moreover, 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*

$$X_n^* := \frac{X_n}{n} \xrightarrow{\mathcal{L}} X^*,$$

where X^* is a perpetuity given by

$$X^* \overset{\mathcal{L}}{=} S_1^* + \sum_{j=1}^{\infty} S_{j+1}^* \prod_{i=1}^{j} V_i^*,$$

with $\{(V_i^*, S_i^*)\}_{i=1}^{\infty}$ being a totally independent set of random vectors (for a standard definition of total independence, see the classic textbook by Chung [18]) distributed like (K^*, T^*).

3.4 Proofs

Proof of Theorem 3.1. In view of regularity condition (i), the structure of the stochastic recurrence (3.1) suggests that D_n converges to a limiting random variable, say D^*,

[1] In this dissertation we are not using "robust" in the statistical sense. Rather, it is used in the algorithmic sense, meaning that the runtime does not change much by changing the splitting protocols.

satisfying the distributional equation

$$D^* \overset{\mathcal{L}}{=} \mathbf{1}_{\{U<K^*\}} D^* + 1, \tag{3.3}$$

with D^* on the right-hand side and $\mathbf{1}_{\{U<K^*\}}$ being independent and U and K^* are also independent.

The strategy of the proof is to first show that the first-order Wasserstein distance between D_n and D^* converges to 0. We then elicit the nature of the limit, which turns out to be a geometric random variable.

Theorem 3.3 *(The coupling theorem).*

$$D_n \xrightarrow{\mathcal{L}} D^*,$$

where D^* satisfies (3.3) and the qualifying statement about independence following it.

Proof. Let (D_n, D^*) be optimal couplings[2], for all $n \ge 0$. Let

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

We shall show that $b_n \to 0$; subsequently, we have $\Delta_1(F_{D_n}, F_{D^*}) \le b_n \to 0$, i.e., $D_n \xrightarrow{\mathcal{L}} D^*$. 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), $\mathbf{1}_{\{U<\frac{K_n}{n}\}} = \mathbf{1}_{\{U<K^*\}} + O_{\mathcal{L}_1}(n^{-\alpha})$ (for justification see Appendix B), so that

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

[2] A coupling is said to be optimal, if it is the one that achieves the infimum in the definition of the Wasserstein distance.

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 (3.3) that D^* has mean $1/(1 - \mathbf{E}[K^*])$. Condition (i) guarantees that this mean is finite. By regularity condition (i), there exist a positive integer n_0, and a positive real constant A, such that

$$\mathbf{E}\big[|K_n^* - K^*|\big] \le \frac{A}{n^\alpha}, \qquad \text{for all } n \ge n_0. \tag{3.4}$$

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

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

According to Remark 3.2, the probability $\mathbf{Prob}(K_n = n)$ is less than 1, for all $n \ge \nu \ge 1$. So, for all $n \ge n_0' = \max\{\nu, n_0\}$, we can now write

$$b_n \le \frac{\displaystyle \frac{1}{n} \sum_{k=0}^{n-1} k b_k \mathbf{Prob}(K_n = k) + \frac{A'}{n^\alpha}}{1 - \mathbf{Prob}(K_n = n)}.$$

Let us start an induction[3] to show that $b_n \le h/n^\alpha$, for some constant $h > 0$. For $1 \le n \le n_0'$, we have

$$b_n \le \max_{1 \le j \le n} b_j \le \frac{(n_0')^\alpha \displaystyle\max_{1 \le j \le n_0} b_j}{n^\alpha} =: \frac{h_1}{n^\alpha}.$$

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

[3]One may be able to weaken the regularity conditions to ones without rates. Assuming rates of convergence leads to constructive and transparent proofs in this context. It may be possible to find a proof (probably more involved) that does not assume such rates.

some $n \geq n_0'$, and all $k \in \{1, \ldots, n-1\}$, that $b_k \leq h/k^\alpha$. We have

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

The induction step will be complete, if, for all $n \geq n_0'$,

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

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

$$h \geq \frac{A'}{1 - \mathbf{E}[(K_n^*)^{1-\alpha}]}, \qquad \text{for all } n \geq n_0',$$

which is the case, if

$$h \geq \frac{A'}{1 - \sup_{n \geq n_0'} \mathbf{E}[(K_n^*)^{1-\alpha}]} =: h_2.$$

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

Lemma 3.1 *A random variable D^* satisfying (3.3) and the qualifying statement about independence following it, has a geometric distribution with parameter $1-\mathbf{E}[K^*]$.*

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

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

The unique solution to this equation is

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

which is the moment generating function of $\text{Geo}(1 - \mathbf{E}[K^*])$. □

Theorem 3.3 and Lemma 3.1 establish a proof for Theorem 3.1. □

The tool we shall use to prove Theorem 3.2 is the contraction method. This method was introduced by Rösler [84] 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 Neininger and Rüschendorf [74], an exposition in the context of recursive algorithms is given in Rachev and Rüschendorf [81], and Rösler and Rüschendorf [86] provides a survey.

The proof will be in three parts (organized as a theorem, a lemma 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 of in the proof of Theorem 3.3. In forthcoming Lemma 3.2, 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 similar to the one in Lemma 3.1, 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 (3.3) 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 3.1 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.

Proof of Theorem 3.2. Let us write the recurrence equation (3.2) in normalized form:

$$X_n^* := \frac{X_n}{n} \overset{\mathcal{L}}{=} \frac{K_n}{n} \times \frac{X_{K_n}}{K_n} + \frac{T_n}{n} = K_n^* X_{K_n}^* + T_n^*. \tag{3.5}$$

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 the distributional equation

$$X^* \overset{\mathcal{L}}{=} K^* X^* + T^*, \tag{3.6}$$

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

Theorem 3.4

$$X_n^* \xrightarrow{\mathcal{L}} X^*,$$

where X^* satisfies (3.6) and the qualifying statement about independence following it.

Proof. Let us take the pairs (K_n^*, K^*), for $n \geq 0$, 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 such that (X_n^*, X^*) are optimal couplings for all $n \geq 0$. Use these variables as realizations in the right-hand sides of (3.5) and (3.6). We then have

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

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

$$\mathbf{E}\Big[|T_n^* - T^*|\Big] \le \frac{A''}{n^\alpha}, \qquad \text{for all } n \ge n_0''.$$

Recall here the bound on $\mathbf{E}[|K_n^* - K^*|]$ (established in (3.4)). In addition, the conditions

$$\mathbf{E}[|K^*|^s] < 1, \qquad \text{and} \qquad \mathbf{E}[|T^*|^s] < \infty,$$

for some $s \in [1, \infty)$, guarantee $\mathbf{E}[|X^*|^s] < \infty$; see [27], 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. Thus, $\mathbf{E}[X^*]$ is a well defined positive number, and is given by $\mathbf{E}[X^*] = \mathbf{E}[T^*]/(1 - \mathbf{E}[K^*])$. Combining the bounds, we see that

$$\begin{aligned}
g_n &\le \mathbf{E}\Big[\Big|K_n^* X_{K_n}^* - \Big(K_n^* - O_{\mathcal{L}_1}\Big(\frac{1}{n^\alpha}\Big)\Big)X^*\Big|\Big] + \frac{A''}{n^\alpha} \\
&\le \mathbf{E}\big[\big|K_n^*(X_{K_n}^* - X^*)\big|\big] + \mathbf{E}\Big[\Big|\,X^* O_{\mathcal{L}_1}\Big(\frac{1}{n^\alpha}\Big)\Big|\Big] + \frac{A''}{n^\alpha} \\
&\le \mathbf{E}\big[\big|K_n^*(X_{K_n}^* - X^*)\big|\big] + \mathbf{E}[\,X^*]\,\mathbf{E}\Big[O_{\mathcal{L}_1}\Big(\frac{1}{n^\alpha}\Big)\Big] + \frac{A''}{n^\alpha} \\
&\le \mathbf{E}\big[\big|K_n^*(X_{K_n}^* - X^*)\big|\big] + \frac{\mathbf{E}[T^*]\,A}{(1 - \mathbf{E}[K^*])\,n^\alpha} + \frac{A''}{n^\alpha},
\end{aligned}$$

for $n \ge \max\{n_0, n_0''\}$ (where n_0 is the one we used in the proof of Theorem 3.3). Recall that T^* is integrable, and let $A''' = \frac{\mathbf{E}[T^*]}{1-\mathbf{E}[K^*]}\,A + A''$, and we have

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

According to Remark 3.2, the probability $\mathbf{Prob}(K_n = n)$ 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

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

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

$$\begin{aligned} h' &\ge \frac{A'''}{1 - \mathbf{E}[(K_n^*)^{1-\alpha}]}, \qquad \text{for all } n \ge n_0''' \\ &\ge \frac{A'''}{1 - \sup\limits_{n \ge n_0'''} \mathbf{E}[(K_n^*)^{1-\alpha}]}, \end{aligned}$$

and by Remark 3.2, 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 $X_n^* \xrightarrow{\mathcal{L}} X^*$. $\square$

Lemma 3.2 *(Contraction). There is a unique distribution satisfying (3.6) and the qualifying statement about independence following it.*

Proof. Let us recall that the conditions

$$\mathbf{E}[|K^*|^s] < 1, \qquad \text{and} \qquad \mathbf{E}[|T^*|^s] < \infty,$$

for some $s \in [1, \infty)$, guarantee $\mathbf{E}[|X^*|^s] < \infty$; see [27], 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 (3.6) 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 (3.6), with distribution functions F_{X^*} and F_{Y^*} such that X^* is independent of (K^*, T^*), and Y^* is independent of (K^*, T^*). Start with the calculation

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

Condition (i) ensures that $\mathbf{E}[K^*]$ 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 (3.6) and the qualifying statement about independence following it. □

Having established the uniqueness of the distribution of X^* in Lemma 3.2, we proceed to argue that X^* is a perpetuity. It is shown in [91] and [27], pp. 457–459 that distributional equations of the form (3.6) unwind into a sum of products of independent random variables (i.e. a perpetuity) as in Theorem 3.2, provided that $-\infty \le \mathbf{E}[\ln |K^*|] < 0$, and $\mathbf{E}[\ln^+ |T^*|] < \infty$. 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 3.4, Lemma 3.2, and the above concluding argument complete the proof of Theorem 3.2. □

Remark 3.4: *We can iterate (3.3) to obtain the perpetuity*

$$D_n \xrightarrow{\mathcal{L}} 1 + B_1 + B_1B_2 + B_1B_2B_3 + \cdots ,$$

where $\{B_i\}_{i=1}^{\infty}$ are totally independent identically distributed Bernoulli random variables with success probability $\mathbf{E}[K^*]$. Of course, the representation of the limit as a geometric random variable in Theorem 3.1 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. Also it is interesting to note that there are only a few examples of exact solutions to perpetuity equations in the literature and ours is one of them.

Remark 3.5: *Our regularity conditions restrict the splitting protocols to be in a class of well-behaved algorithms (i.e. producing well-behaved distributions of K_n), for which we can get nice convergence results. They limit the class of distributions to sequences that have common threads (such as the uniform and the binomial examples), where the generating mechanism produces distributions that change smoothly, for example their averages increase at decent rates. If instead the sequence of distributions of K_n is too erratic (oscillatory, for example), similar results cannot be expected. For instance, K_n may be $\frac{n}{2}$, for n even and $\lfloor \frac{\ln n}{3} \rfloor$, for n odd. In this case D_n cannot converge.*

Remark 3.6: *We are concerned only with efficient algorithms, which are those that do not perform superfluous operations. It is clear that any algorithm deciding on the future success of n contestants in a tournament will have to have $O(n)$ running time in the first elimination round. For instance, we can determine the advancing set by having contestants advance only if they flip Heads, using identically distributed coins. So, there are n coin flips inducing $O(n)$ running time. Such an algorithm is in our repertoire of fair leader election algorithms that we analyze. On the other hand, a coin flipping algorithm that operates as follows is excluded from our analysis: Each contestant is asked to throw the same coin, and each of them flips it $(2n+1)^2$ times. If there are more Heads than Tails in a contestant's set of coin flips, that player advances, otherwise she is eliminated. Clearly, the first round of flips consumes $O(n^3)$ time. This algorithm produces the same fair election results (distributionally) as one that advances or eliminates a contestant after only one flip, it just demands many more coin flips without providing any advantage. We also exclude from our analysis those algorithms that can be unfair, such as one based on coin flips that are not identical. For example, an algorithm may assign a different coin to each contestant, and the* *coins do not have the same probabilities of producing Heads. Such an algorithm favors contestants holding coins with higher probability of Heads*, which would render it an unfair election algorithm.

Chapter 4

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, asymptotic fluctuations, or rates of convergence.

4.1 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 $\{1, 2, \ldots, n\}$. In this example, we can find some exact moments and say a word about rates of convergence. We derive a functional equation for $\phi_n(t)$, the moment generating function of D_n, from the distributional equation (3.1):

$$\mathbf{E}\big[e^{tD_n} \mid K_n\big] = e^t\Big(1 - \frac{K_n}{n}\Big) + e^t e^{tD_{K_n}}\Big(\frac{K_n}{n}\Big),$$

with an unconditional expectation satisfying

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

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

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

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

$$\phi_n(t) = \frac{(n-1)^2\phi_{n-1}(t) + (n-1)e^t}{n(n-e^t)}, \tag{4.1}$$

valid for $n \geq 3$. The following lemma gives a solution.

Lemma 4.1 *For $n \geq 2$, $t < \ln 2$, and under the initial condition $\phi_2(t) = \dfrac{e^t}{2-e^t}$,*

$$\phi_n(t) = \frac{e^t}{2-e^t} - \frac{e^t}{n(2-e^t)} + \frac{e^t\,\Gamma\big(2-e^t\big)\,\Gamma\big(n\big)}{n\Gamma\big(n+1-e^t\big)}.$$

Proof. We show the above result by induction. For the induction hypothesis, assume

$$\begin{aligned}
\phi_k(t) &= \frac{e^t}{2-e^t} - \frac{e^t}{k(2-e^t)} + \frac{e^t\,\Gamma\big(2-e^t\big)\,\Gamma\big(k\big)}{k\Gamma\big(k+1-e^t\big)} \\
&= \frac{e^t}{2-e^t} - \frac{e^t}{k(2-e^t)} + \frac{e^t\,\Gamma\big(2-e^t\big)\,\Gamma\big(k\big)}{k\Gamma\big(k-e^t\big)(k-e^t)},
\end{aligned}$$

valid for $k \geq 2$, and $t < \ln 2$. This assumption is first conjectured with the aid of a symbolic computation system such as MAPLE. From (4.1), we have

$$\begin{aligned}
\phi_{k+1}(t) &= \frac{k^2\phi_k(t) + ke^t}{(k+1)(k+1-e^t)} \\
&= \frac{\dfrac{k^2e^t}{2-e^t} - \dfrac{ke^t}{(2-e^t)} + \dfrac{ke^t\,\Gamma\big(2-e^t\big)\,\Gamma\big(k\big)}{\Gamma\big(k-e^t\big)(k-e^t)} + ke^t}{(k+1)(k+1-e^t)} \\
&= \frac{k^2e^t + ke^t - ke^{2t}}{(2-e^t)(k+1)(k+1-e^t)} + \frac{ke^t\,\Gamma\big(2-e^t\big)\,\Gamma\big(k\big)}{(k+1)(k+1-e^t)\Gamma\big(k-e^t\big)(k-e^t)} \\
&= \frac{e^t}{2-e^t} - \frac{e^t}{(k+1)(2-e^t)} + \frac{e^t\,\Gamma\big(2-e^t\big)\,\Gamma\big(k+1\big)}{(k+1)\Gamma\big(k+1-e^t\big)(k+1-e^t)}.
\end{aligned}$$

By the principle of induction, the proof is complete. □

Derivatives of $\phi_n(t)$, 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 4.1 *For $n \ge 3$, we have*

$$\begin{aligned}
\mathbf{E}[D_n] &= 2 + \frac{H_n}{n} - \frac{1}{n} - \frac{1}{n^2}, \\
\mathbf{Var}[D_n] &= 2 - \frac{1}{n} - \frac{2}{n-1} + \frac{2}{n^2} - \frac{2}{n^3} - \frac{1}{n^4} - \frac{3H_n}{n} + \frac{2H_n}{n-1} + \frac{2H_n}{n^3} \\
&\qquad - \frac{2H_{n-1}}{n(n-1)} + \frac{H_n^2}{n} - \frac{H_n^2}{n^2} + \frac{H_n^{(2)}}{n}.
\end{aligned}$$

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 regularity condition (i) is satisfied (we can take $\alpha = \frac{1}{2}$). And so, $K^* = U$, with mean $\mathbf{E}[K^*] = \frac{1}{2}$. According to Theorem 3.1, we have

$$D_n \xrightarrow{\mathcal{L}} \mathrm{Geo}\Big(\frac{1}{2}\Big).$$

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:

$$X_n^* \xrightarrow{\mathcal{L}} U_1 + U_1U_2 + U_1U_2U_3 + \cdots ,$$

and $\{U_i\}_{i=1}^{\infty}$ are totally independent and identically distributed continuous Uniform (0,1) random variables. The above limiting perpetuity is known as the Dickman distribution.

4.2 Power distribution splitting

This example explores the asymptotic distribution of D_n and X_n, when the splitting protocol K_n follows a (discrete) power distribution. This distribution with parameter $\theta \geq 0$ has a probability mass function

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

The discrete uniform distribution (discussed in section (4.1)) 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, which has a mean value of $(\theta + 1)/(\theta + 2)$. On a suitable probability space we can take $K_n^* = V + O(1/n)$, and V is a Beta$(\theta + 1, 1)$ random variable. All the regularity conditions for Theorem 3.1 hold, and as $n \to \infty$, we have

$$D_n \xrightarrow{\mathcal{L}} \text{Geo}\Big(\frac{1}{\theta + 2}\Big).$$

A set of size with a power distribution can be generated in time linear in the size plus a constant overhead (and as before we take the linearity constant to be 1). So, $T_n/n = V + O(1/n)$, All the regularity conditions for Theorem 3.2 check out, and as $n \to \infty$, we have

$$X_n^* \xrightarrow{\mathcal{L}} V_1 + V_1 V_2 + V_1 V_2 V_3 + \cdots ,$$

and $\{V_i\}_{i=1}^{\infty}$ are independent and identically distributed Beta$(\theta + 1, 1)$ random variables.

4.3 Binomial splitting

This is a classic example and several of its properties have been studied. For instance, see Prodinger [78], Fill, Mahmoud, and Szpankowski [28], Janson and Szpankowski [46], and Kalpathy, Mahmoud, and Ward [55]. In this example, contestants flip coins that produce Heads (with probability p) and Tails (with probability $q = 1 - p$). At the start we have n contestants, of which ultimately one or none is declared the winner after a fair competition. At the first round, those who flip Tails are eliminated from the competition, and those who flip Heads stay to compete in further rounds. This elimination process goes on until either one contestant is declared a winner, or all contestants are eliminated. See Figures 4.1, 4.2 and 4.3 for an illustration. Thus, K_n has the binomial distribution underlying n independent identically distributed experiments, with success probability p per experiment. Hence K_n/n converges in $\mathcal{L}_1$ to p. This convergence occurs at the rate $O_{\mathcal{L}_1}(1/\sqrt{n}\,)$, as can be found out, for instance, from the standard approximation of the binomial distribution by the normal. The conditions for Theorem 3.1 are all met and, as $n \to \infty$, we have

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

This result appears in Kalpathy, Mahmoud, and Ward [55], where the authors further show that the lower-order asymptotics in the distribution function may have oscillations. In this example, we are able to produce the exact probability distribution.

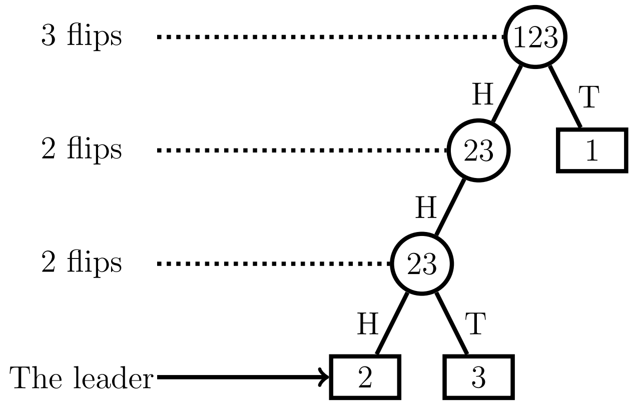

Figure 4.1: Example with 3 contestants. Here we have $D_{3,1} = 1$, $D_{3,2} = 3$, $D_{3,3} = 3$, and $X_3 = 7$.

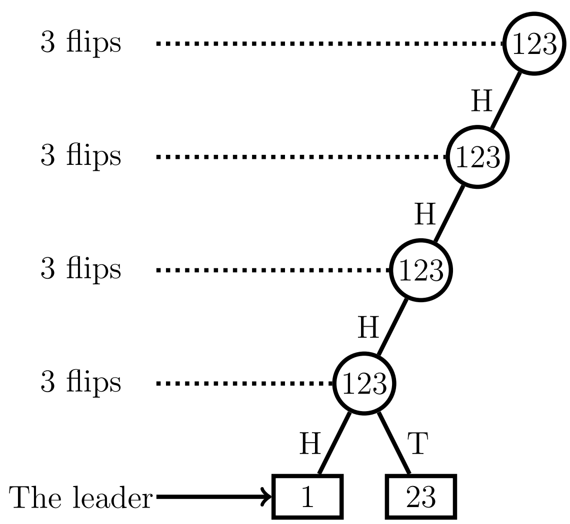

Figure 4.2: Another example with 3 contestants. Here we have $D_{3,1} = 4$, $D_{3,2} = 4$, $D_{3,3} = 4$, and $X_3 = 12$.

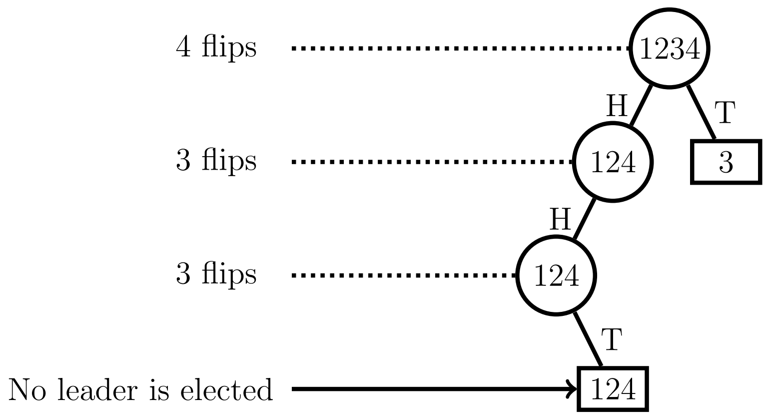

Figure 4.3: Example with 4 contestants. Here we have $D_{4,1} = 3$, $D_{4,2} = 3$, $D_{4,3} = 1$, $D_{4,4} = 3$, and $X_4 = 10$.

Interestingly, the asymptotic distribution of D_n has fluctuations, which are captured by Rice's integral method—we will introduce this method in the sequel.

4.3.1 Exact distribution of the duration

Theorem 4.1 *The probability mass function for D_n is exactly*

$$\textbf{Prob}(D_n = k) = p^{k-1}\big(q - (1 - p^{k-1})^{n-1} + p(1 - p^k)^{n-1}\big),$$

for $n \geq 2$, and $k \geq 1$.

Proof. The duration of the first contestant is k in one of two ways: She either loses in k rounds (event $\mathcal{L}_k$), or wins in k rounds (event $\mathcal{W}_k$). The event $\mathcal{L}_k$ occurs, if she flips $k - 1$ Heads followed by a Tail in the kth flip, provided that at least one other contestant flips $k-1$ Heads (otherwise, the contest would have stopped sooner). Thus,

$$\textbf{Prob}(\mathcal{L}_k) = p^{k-1}q\big(1 - (1 - p^{k-1})^{n-1}\big).$$

The event $\mathcal{W}_k$ occurs, if the first contestant flips k Heads at the start, provided that none of the other contestant flips k Heads at the start, but at least one of them flips

$k-1$ Heads at the start (otherwise, the winning contestant would have won sooner). Thus,

$$\mathbf{Prob}(\mathcal{W}_k) = p^k\big((1-p^k)^{n-1} - (1-p^{k-1})^{n-1}\big).$$

Whence, $\mathbf{Prob}(D_n = k) = \mathbf{Prob}(\mathcal{L}_k) + \mathbf{Prob}(\mathcal{W}_k)$, and the theorem follows. □

Corollary 4.1 *As $n \to \infty$,*

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

4.3.2 Oscillating moments via Rice's integral method

The asymptotic moments can be found from the exact distribution. Of course, in the limit they all are the moments of a Geo(q) random variable. What is interesting here is the asymptotic oscillations that ultimately disappear. We illustrate this phenomenon for the mean and variance.

Corollary 4.2 *As $n \to \infty$,*

$$\begin{aligned}
\mathbf{E}[D_n] &= \frac{1}{q} + \Big(\frac{1}{\ln p}\sum_{k=-\infty}^{\infty}\Gamma\Big(1-\frac{2\pi ik}{\ln p}\Big)e^{-2\pi ik\log_{1/p} n}\Big)\frac{1}{n} + O\Big(\frac{1}{n^2}\Big),\\
\mathbf{Var}[D_n] &= \frac{p}{q^2} - \Big(\frac{2}{(\ln p)^2}\sum_{k=-\infty}^{\infty}\Gamma\Big(1-\frac{2\pi ik}{\ln p}\Big)e^{-2\pi ik\log_{1/p} n}\Big)\frac{\ln n}{n} + O\Big(\frac{1}{n}\Big).
\end{aligned}$$

The error terms include small oscillations.

Proof. We use the exact form of the distribution of D_n in Theorem 4.1 as a starting point. This yields the exact mean for $n \ge 2$:

$$\mathbf{E}[D_n] = \sum_{k=0}^{\infty} k\,\mathbf{Prob}(D_n = k) = \frac{1}{q} - \sum_{k=0}^{\infty} kp^{k-1}(1-p^{k-1})^{n-1} + \sum_{k=0}^{\infty} kp^k(1-p^k)^{n-1}.$$

After shifting the index of k in the first sum, we have $\mathbf{E}[D_n] = \frac{1}{q} - \sum_{k=0}^{\infty} p^k(1-p^k)^{n-1}$. Using the binomial expansion $(1-p^k)^{n-1} = \sum_j \binom{n-1}{j}(-1)^j p^{kj}$, and simplifying the resulting geometric sum, we conclude that

$$\mathbf{E}[D_n] = \frac{1}{q} - \sum_{j=0}^{n-1} \binom{n-1}{j} \frac{(-1)^j}{1-p^{j+1}}.$$

The alternating sum in the mean can be handled by Rice's integral method. See Mahmoud [69] and Flajolet and Sedgewick [30]. The fundamental idea is to recognize that such an alternating sum is exactly the same as

$$-\frac{1}{2\pi i} \oint_{C_1} \frac{\beta(n,-z)}{1-p^{z+1}}\, dz,$$

where $\beta(.,.)$ is the Beta function, and the line integral is taken over a closed contour C_1 surrounding the integers $0, 1, \ldots, n-1$ and no other integers (see Mahmoud [69], p. 275). For instance, C_1 can be the contour consisting of a rectangle connecting the four corners at $-\frac{1}{2} \pm i$ and $n - \frac{1}{2} \pm i$. The link is that the integrand has simple poles at the integer points $0, 1, \ldots, n-1$, and their residues coincide with the negative of the summands. The integral is then evaluated via the residues of the poles outside C_1 and a small error. This is done by deforming C_1 into a larger contour $C_2(a, M)$, say a rectangle connecting the four corners at $-a \pm \pi(2M+1)i$ and $n + b \pm \pi(2M+1)i$, for $a, b > 1$, and large positive integer M. The two integrals over C_1 and $C_2(a, M)$ differ by the residues of the poles enclaved between the two contours (see Figure 4.4). Outside C_1, there are simple poles at $z_k = -1 - 2\pi k i/\ln p$, for $k = 0, \pm 1, \pm 2, \ldots$. Thus, as we let $M \to \infty$, the rectangle $C_2(a, M)$ grows to encompass all the poles outside C_1, and

$$-\frac{1}{2\pi i} \oint_{C_1} \frac{\beta(n,-z)}{1-p^{z+1}}\, dz = -\frac{1}{2\pi i} \lim_{M\to\infty} \oint_{C_2(a,M)} \frac{\beta(n,-z)}{1-p^{z+1}}\, dz + \sum_{k=-\infty}^{\infty} \operatorname*{Res}_{z=z_k} \frac{\beta(n,-z)}{1-p^{z+1}}.$$

By Stirling's approximation to the Gamma function, we have the residue calculation

$$\operatorname*{Res}_{z=z_k} \frac{\beta(n,-z)}{1-p^{z+1}} = -\frac{\Gamma(n)\,\Gamma(-z_k)}{\Gamma(n-z_k)\,\ln p} = -\frac{n^{-1+2\pi k i/\ln p}}{\ln p}\,\Gamma\Big(1 - \frac{2\pi k i}{\ln p}\Big)\Big(1 + O\Big(\frac{1}{n}\Big)\Big).$$

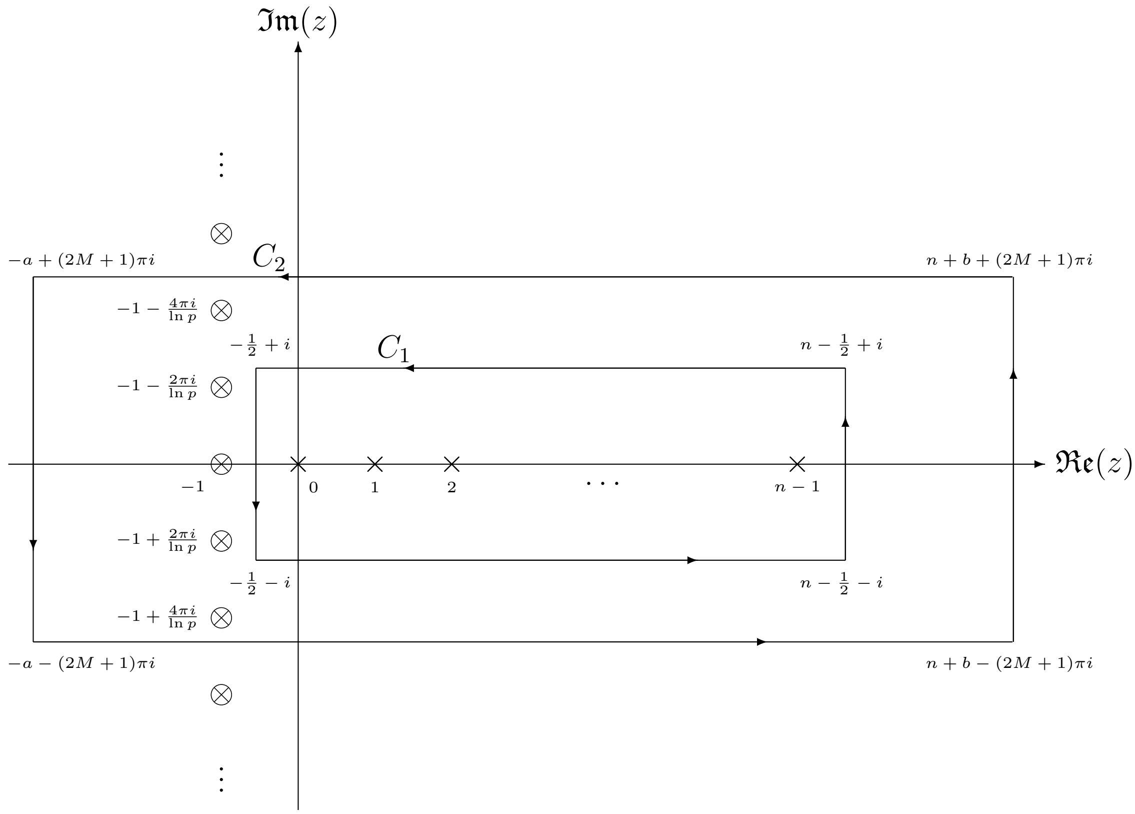

Figure 4.4: The integration contours and the poles of the integrand; $\times$ indicates simple poles coming from $\beta(n,-z)$ and $\otimes$ indicates simple poles coming from $(1-p^{z+1})$.

As $M\to\infty$, the integral over the limiting contour $\lim_{M\to\infty} C_2(a,M)$ gives a correction error of the order n^{-a}. As we already took only one term in Stirling's approximation, it does not help to take a much greater than 1. Let us take $a=2$ and obtain

$$\mathbf{E}[D_n]=\frac{1}{q}+\Big(\frac{1}{\ln p}\sum_{k=-\infty}^{\infty}\Gamma\Big(1-\frac{2\pi ik}{\ln p}\Big)e^{-2\pi ik\log_{1/p} n}\Big)\frac{1}{n}+O\Big(\frac{1}{n^2}\Big).$$

We can improve the overall error by taking more terms in Stirling's approximation, and choosing larger a.

For the variance calculation, we look at the exact second factorial moment:

$$\mathbf{E}\big[D_n\big(D_n-1\big)\big]=\sum_{k=0}^{\infty}k(k-1)\,\mathbf{Prob}(D_n=k)=2\sum_{j=1}^{n-1}\binom{n-1}{j}\frac{(-1)^{j+1}p^{j+1}}{(1-p^{j+1})^2}.$$

We manipulate the formula for the exact probability to be in the form of an alternating sum, which we then handle via Rice's integral method.

$$\begin{aligned}\mathbf{E}[D_n(D_n-1)] &= -2\sum_{j=0}^{n-1}\binom{n-1}{j}\frac{(-1)^j p^{j+1}}{(1-p^{j+1})^2}+\frac{2p}{q^2}\\ &= \frac{2}{2\pi i}\oint_{C_1}\frac{\beta(n,-z)p^{z+1}}{(1-p^{z+1})^2}\,dz+\frac{2p}{q^2}.\end{aligned}$$

The exact second moment is

$$\begin{aligned}\mathbf{E}[{D_n}^2] &= \frac{2}{2\pi i}\oint_{C_1}\frac{\beta(n,-z)p^{z+1}}{(1-p^{z+1})^2}\,dz+\frac{2p}{q^2}+\mathbf{E}[D_n]\\ &= \frac{2-q}{q^2}-\frac{2\gamma}{n(\ln p)^2}+\Big(\frac{2}{(\ln p)^2}\sum_{k=-\infty}^{\infty}\Gamma\Big(1-\frac{2\pi ik}{\ln p}\Big)e^{-2\pi ik\log_{1/p}n}\Big[\Psi\Big(1-\frac{2\pi ik}{\ln p}\Big)\\ &\qquad -\ln n+\frac{1}{2}\ln p\Big]\Big)\frac{1}{n}+O\Big(\frac{1}{n}\Big).\end{aligned}$$

where $\gamma = 0.577215\ldots$ is Euler's constant, and $\Psi(.)$ being the digamma function. We then obtain the following expression for the variance

$$\mathbf{Var}[D_n]=\frac{p}{q^2}-\Big(\frac{2}{(\ln p)^2}\sum_{k=-\infty}^{\infty}\Gamma\Big(1-\frac{2\pi ik}{\ln p}\Big)e^{-2\pi ik\log_{1/p}n}\Big)\frac{\ln n}{n}+O\Big(\frac{1}{n}\Big).\quad\square$$

4.3.3 Total cost

The total cost or speed of the algorithm is measured by the number of independent coin flips. In a parallel environment (where there are enough coins to give one to each contestant), the number of rounds (height of the underlying incomplete tree) is a measure of the total cost till termination of the contest. This measure has been analyzed in Fill, Mahmoud, and Szpankowski [28] for unbiased coins, and in Janson and Szpankowski [46] for biased coins. However, in a serial environment, where all the contestants share one coin that 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 3.2 are met. We have the first-order asymptotic

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

as in Kalpathy, Mahmoud, and Ward [55]. Second-order asymptotics are also given in that paper, specifying a rate of convergence for this weak law in the form of a central limit theorem. The second order theory is developed using the second order Wasserstein distance.

4.3.4 A second order theory for the total cost

Let $W_n := K_n$ be the number of candidate winners who move to the next round. The number W_n is a binomially distributed random variable counting the number of successes in n independent identically distributed trials with rate of success p per trial. For $n \geq 2$, we have

$$X_n \stackrel{\mathcal{L}}{=} X_{W_n} + n; \tag{4.2}$$

the initial conditions are $X_1 = X_0 = 0$. Note that

$$X_n = D_{n,1} + D_{n,2} + \cdots + D_{n,n}. \tag{4.3}$$

The random durations are identically distributed but are dependent (in fact they are exchangeable). This adds complexity to the study of the distribution of X_n. However, this does not introduce an added difficulty for calculation of the mean:

$$\mathbf{E}[X_n] = \sum_{j=1}^{n} \mathbf{E}[D_{n,j}] = n\mathbf{E}[D_n] = \frac{n}{q} + \frac{1}{\ln p} \sum_{k=-\infty}^{\infty} \Gamma\Big(1 - \frac{2\pi i k}{\ln p}\Big) e^{-2\pi i k \log_{1/p} n} + O\Big(\frac{1}{n}\Big).$$

In view of the dependence, higher moments are harder to compute directly from recurrences derived from (4.3). However, the form of (4.2) is amenable to convergence methods in metric spaces, as simple limits can be guessed. The binomial distribution of W_n can be approximated by a normal distribution with mean np and variance npq.

(In what follows a normal random variate with mean μ and variance σ^2 will be denoted by $\mathcal{N}(\mu, \sigma^2)$.) This suggests that $\sqrt{n}$ is the right scale factor for limits to appear on the right-hand side. Centering and scaling (4.2), we get

$$X_n^* := \frac{X_n - n/q}{\sqrt{n}} \stackrel{\mathcal{L}}{=} \frac{X_{W_n} - W_n/q}{\sqrt{W_n}} \sqrt{\frac{W_n}{n}} + \frac{n - n/q + W_n/q}{\sqrt{n}}. \tag{4.4}$$

Thus, in normalized form we have the distributional equation

$$X_n^* \stackrel{\mathcal{L}}{=} X_{W_n}^* \sqrt{\frac{W_n}{n}} + \frac{n - n/q + W_n/q}{\sqrt{n}}.$$

Now, if $X_n^* \xrightarrow{\mathcal{L}} X^*$, so will $X_{W_n}^*$ because $W_n \xrightarrow{a.s.} \infty$. Also $\sqrt{W_n/n} \xrightarrow{P} \sqrt{p}$. It would then follow from Slutsky's theorem (see Karr [57], p. 147) that the combination $X_{W_n}^* \sqrt{W_n/n}$ converges in distribution to $X^* \sqrt{p}$. Furthermore, $\frac{n-n/q+W_n/q}{\sqrt{n}} \xrightarrow{\mathcal{L}} Z\sqrt{p/q}$, where $Z \stackrel{\mathcal{L}}{=} \mathcal{N}(0,1)$. Hence, if X_n^* has a limit, we would expect it to satisfy the distributional equation

$$X^* \stackrel{\mathcal{L}}{=} \sqrt{p}\, X^* + \sqrt{\frac{p}{q}}\, Z.$$

To rigorously prove all this, it suffices to show that the second-order Wasserstein distance $\Delta_2(F_n^*, F^*)$ converges to 0, as $n \to \infty$, where F_n^* is the distribution function of X_n^* and F^* is the distribution function of X^*.

Recall from Section 2.2, the Wasserstein distance of order 2 between two distribution functions F_n^* and F^* is defined by

$$\Delta_2(F_n^*, F^*) := \inf_{V_n^*, V^*} \{ ||W_n^* - W||_2 : W_n^* \stackrel{\mathcal{L}}{=} V_n^*, W \stackrel{\mathcal{L}}{=} V^* \}, \tag{4.5}$$

where V_n^* is of distribution F_n^*, V^* is of distribution F^*, and $||\,.\,||_2$ is the $\mathcal{L}_2$ norm. We show in Lemma 4.3 that indeed $\Delta_2(F_n^*, F^*) \to 0$. Consequently, $X_n^* \xrightarrow{\mathcal{L}} X^*$, and converges in the first two moments, too. See Bickel and Freedman [6] and Barbour, Holst, and Janson [4]. The next lemma characterizes the limit.

Lemma 4.2 *Let Z be a standard normal random variable. If a random variable X^* is independent from Z and satisfies the distributional equation*

$$X^* \stackrel{\mathcal{L}}{=} \sqrt{p}\, X^* + \sqrt{\frac{p}{q}}\, Z,$$

then X^* has the distribution of $\mathcal{N}(0, p/q^2)$.[1]

Proof. Let $\eta^*_X(t)$ be the characteristic function of X^*. Both sides of the distributional equation have the same characteristic function:

$$\begin{aligned}
\eta_{X^*}(t) &= \eta_{X^*}(p^{1/2}t)\exp\Big(-\frac{pt^2}{2q}\Big) \\
&= \eta_{X^*}(p^{2/2}t)\exp\Big(-\frac{pt^2}{2q} - \frac{p^2t^2}{2q}\Big) \\
&\ \ \vdots \\
&= \eta_{X^*}(p^{k/2}t)\exp\Big(-\frac{pt^2}{2q}\Big(1 + p + p^2 + \cdots + p^{k-1}\Big)\Big).
\end{aligned}$$

Take the limit, as $k \to \infty$. By dominated convergence, the function $\eta_{X^*}(p^{k/2}t)$ has the limit $\eta_{X^*}(0) = 1$. Hence,

$$\eta_{X^*}(t) = \exp\left(-\frac{pt^2}{2q^2}\right),$$

and the right-hand side is the characteristic function of $\mathcal{N}(0, p/q^2)$. □

The convergence of X^*_n to a limit X^* satisfying the distributional equation in Lemma 4.2 is demonstrated in the next lemma.

Lemma 4.3 *As $n \to \infty$,*

$$X^*_n \xrightarrow{\mathcal{L}} X^*,$$

and the limit X^* satisfies the distributional functional equation

$$X^* \stackrel{\mathcal{L}}{=} \sqrt{p}\, X^* + \sqrt{\frac{p}{q}}\, Z,$$

for a standard normal random variable Z, where X^* and Z are independent.

[1] In fact, this is a perpetuity with building blocks $\Big(\sqrt{p}, \sqrt{p/q}\, Z\Big)$.

Proof. Let $F_n^*(x)$ be the distribution function of X_n^*, and $F^*(x)$ be the distribution function of X^*. We shall show that the second-order Wasserstein distance between $F_n^*(x)$ and $F^*(x)$ converges to 0; and consequently $X_n^* \xrightarrow{\mathcal{L}} X^*$.

To be able to measure differences between random variables, they must be defined on the same probability space. By the standard approximation of the binomial distribution by the normal, we have

$$\frac{W_n - pn}{\sqrt{pqn}} \xrightarrow{\mathcal{L}} Z,$$

and $Z = \mathcal{N}(0,1)$. According to Skorohod's representation (see Billingsley [7], p. 333), there is a probability space on which we can define $W_n' \stackrel{\mathcal{L}}{=} W_n$, $Z' \stackrel{\mathcal{L}}{=} Z$, and $W_n' \xrightarrow{a.s.} Z'$ in that space. We can show in that space that

$$W_n' = pn + \sqrt{pqn}\, Z' + O_{\mathcal{L}_1}(1).$$

Let V_n^* and V^* be versions of these variables in that space, and let

$$\begin{aligned}
b_n &:= \mathbf{E}\big[(V_n^* - V^*)^2\big] \\
&= \mathbf{E}\Big[\Big(\Big(V_{W_n'}^* \sqrt{\frac{W_n'}{n}} + \frac{n - n/q + W_n'/q}{\sqrt{n}}\Big) - \Big(\sqrt{p}\, V^* + \sqrt{p/q}\, Z'\Big)\Big)^2\Big] \\
&= \mathbf{E}\Big[\Big(\Big(V_{W_n'}^* \sqrt{p + O_{\mathcal{L}_1}(n^{-1/2})} + \frac{\sqrt{pqn}\, Z' + O_{\mathcal{L}_1}(1)}{q\sqrt{n}}\Big) \\
&\qquad - \Big(\sqrt{p}\, V^* + \sqrt{p/q}\, Z'\Big)\Big)^2\Big] \\
&= \mathbf{E}\Big[\Big(\sqrt{p}\big(X_{W_n'}^* - V^*\big) + V_{W_n'}^* O_{\mathcal{L}_1}(n^{-1/2}) + O_{\mathcal{L}_1}(n^{-1/2})\Big)^2\Big].
\end{aligned}$$

By the Cauchy-Schwarz inequality, we bound the expectation of $V_{W_n'}^* O_{\mathcal{L}_1}(n^{-1/2})$ by $O(n^{-1/2})$. The other cross-product terms are bounded similarly. And so, when we square out and take expectation we get

$$b_n \;=\; p\mathbf{E}[b_{W_n'}] + O(n^{-1/2}).$$

Condition on W'_n (which has a binomial distribution) to get a recurrence for b_n:

$$b_n \;=\; p\sum_{k=0}^{n} b_k p^k q^{n-k}\binom{n}{k} + O(n^{-1/2}).$$

From the last recurrence we show next (by induction on n) that b_n is bounded from above by $Cn^{-1/2}$, for some positive constant C. The O in the recurrence means that the last term in the recurrence is bounded from above by $\delta n^{-1/2}$, for some positive constant δ, and all $n \ge n'_0$, for some $n'_0 \ge 1$.

Recall that $b_0 = ||V^*_0 - V^*||_2 = \mathbf{E}[(0-V^*)^2] = \mathbf{E}[(X^*)^2] = p/q^2$. For any fixed p (and hence fixed q), the function $1 - p^{x+1} - pb_0q^x = 1 - p^{x+1} - p^2q^{x-2}$ approaches 1 (from below), as $x \to \infty$. It then follows that for n large enough, say $n \ge n''_0 \ge 1$,

$$1 - p^{n+1} - p^2 q^{n-2} \ge\; p^2\sqrt{1+\frac{q}{n}} + q^2. \tag{4.6}$$

Take $n_0 = \max\{n'_0, n''_0\}$. Note that

$$b_j \le \max\{b_1, b_2, \ldots, b_{n_0}\} \le \frac{\max\{b_1, b_2, \ldots, b_{n_0}\}\sqrt{n_0}}{\sqrt{j}}, \qquad \text{for } j = 1, 2, \ldots, n_0.$$

Thus, if $C > \max\{b_1, b_2, \ldots, b_{n_0}\}\sqrt{n_0}$, this guarantees the upper bound $b_n \le Cn^{-1/2}$, at $n = 1, \ldots, n_0$. We take

$$C > \max\Big\{\frac{\delta}{q^2}, \max\{b_1, b_2, \ldots, b_{n_0}\}\sqrt{n_0}\Big\}.$$

Assume the induction hypothesis holds from 1 up to $n-1 \ge n_0$. Then

$$\begin{aligned} b_n(1 - p^{n+1} - p^2q^{n-2}) \;&\le\; p\sum_{j=1}^{n-1}\frac{C}{\sqrt{j}}\,p^j q^{n-j}\binom{n}{j} + \frac{\delta}{\sqrt{n}} \\ &= \frac{p^2 C}{n}\sum_{j=1}^{n-1}\sqrt{j}\,p^{j-1}q^{n-j}\binom{n-1}{j-1} + \frac{\delta}{\sqrt{n}}. \end{aligned} \tag{4.7}$$

We bound the sum via its connection to a binomial random variable B_n on $n-1$ trials with rate of success p per trial:

$$\sum_{j=1}^{n-1}\sqrt{j}\,p^{j-1}q^{n-j}\binom{n-1}{j-1} \le \sum_{s=0}^{n-1}\sqrt{s+1}\,p^s q^{n-1-s}\binom{n-1}{s} = \mathbf{E}[\sqrt{B_n+1}\,].$$

By Jensen's inequality

$$\mathbf{E}[\sqrt{B_n+1}\,] \le \sqrt{\mathbf{E}[B_n+1]} = \sqrt{p(n-1)+1}.$$

It then follows from the last inequality, (4.6) and (4.7) that, for all $n \ge 1$,

$$b_n \le \frac{p^2C\sqrt{1+q/n}+\delta}{\sqrt{n}\,(1-p^{n+1}-p^2q^{n-2})} \le \frac{p^2C\sqrt{1+q/n}+q^2C}{\sqrt{n}\,(1-p^{n+1}-p^2q^{n-2})} \le \frac{C}{\sqrt{n}},$$

completing the induction.

This induction demonstrates that

$$\Delta_2^2(F_n^*, F^*) \le b_n \to 0, \qquad \text{as } \; n \to \infty, \tag{4.8}$$

which is sufficient to establish convergence of X_n^* to X^* in distribution and in the first two moments. □

The convergence of the distribution of X_n^* in Lemma 4.3, together with the characterization in Lemma 4.2, establish the following weak law in the form of a central limit theorem.

Theorem 4.2 *As $n \to \infty$,*

$$\frac{X_n - \dfrac{n}{q}}{\sqrt{n}} \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{N}\Big(0, \frac{p}{q^2}\Big).$$

4.4 An example with a splitting distribution with atoms

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

$$K_n = \begin{cases} 0, & \text{with probability } \dfrac{1}{3}; \\[2ex] k \in \{1, 2, \ldots n-1\}, & \text{with probability } \dfrac{1}{3(n-1)}; \\[2ex] n, & \text{with probability } \dfrac{1}{3}. \end{cases}$$

Let U be a continuous uniform(0,1) random variable. So, $\frac{K_n}{n}$ 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

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

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

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

The limit distribution K^* has atoms (i.e., jumps in the distribution function) of magnitude $\frac{1}{3}$ at 0 and at 1. Here $\mathbf{E}[K^*]$ is $\frac{1}{2}$. The conditions for Theorem 3.1 are met, and we have, as $n \to \infty$,

$$D_n \xrightarrow{\mathcal{L}} \text{Geo}\Big(\frac{1}{2}\Big).$$

We also have $T_n = K_n + c$ time (where c is constant overhead). That is, $T_n^* = K_n^* + O(1/n) = K^* + O(1/n)$. The conditions for Theorem 3.2 are met, giving a limiting perpetuity:

$$X_n^* \xrightarrow{\mathcal{L}} V_1 + V_1 V_2 + V_1 V_2 V_3 + \cdots,$$

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

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

4.5 An almost-deterministic ladder example

Many real-life tournaments such as Wimbledon Championship, NCAA March Madness, and FIFA World Cup are organized in ladders a.k.a. tournament brackets. 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 n is even, a ladder can be created by asking contestants, i and $i+1$ to compete and only one of them advances, for $i = 1, 2, \ldots, \frac{1}{2}n$. 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 by without competing (sometimes called a *bye* or *wild card*), and the contestants are renumbered $1, \ldots, n-1$ (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 equal chance to win the contest. In this example, the size of the advancing set is deterministic, but elements of randomness come 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 3.1 are met; the duration of any contestant in the competition converges in distribution:

$$D_n \xrightarrow{\mathcal{L}} \text{Geo}\Big(\frac{1}{2}\Big).$$

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 3.2 are met. The scaled overall cost converges to a perpetuity:

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

Chapter 5

Future directions

In this section, we summarize four future research projects. The first and third are independent research work, the second is joint work with Professor Mark Daniel Ward from Purdue University, and the fourth is joint work with Professor Walter Rosenkrantz from University of Massachusetts Amherst and Professor Hosam Mahmoud.

5.1 Continued fraction structure of perpetuities

This is an ongoing independent research work from Kalpathy [51], where we are studying the relationship between perpetuities and random continued fractions. We are not aware of any previous studies about this problem. The author is investigating this problem after several discussions and personal communication with some experts like Professor Daniel Panario (Carleton University), The Late Dr. Philippe Flajolet (Institut National de Recherche en Informatique et en Automatique), Professor Jim Fill (Johns Hopkins University), and Professor Dr. Ralph Neininger (Universität Frankfurt). Recently the author learned from a personal communication with Professor Jonathon Peterson (Purdue University) that random continued fractions also arise in the study of large deviations for a random walk in random environment (see

Comets, Gantert, and Zeitouni [13]). The main intuition behind our study is that some one-sided divide and conquer algorithms can be translated using the contraction principle into fixed point equations and perpetuities which in turn can be viewed as a random continued fraction using tools from number theory. The goal is to connect these three dots and make use of the rich and diverse tools available in these three areas to study their interplay.

5.1.1 Continued fractions

A finite continued fraction is an expression of the form

$$\alpha = [a_0; a_1, a_2, a_3, \ldots, a_n] = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{\ldots + \cfrac{1}{a_n}}}}}. \tag{5.1}$$

The variables $a_0, a_1, a_2, a_3, \ldots, a_n$ are called the *partial quotients* or *elements* of the continued fraction. In general, these elements may be assumed to be integers, real or complex numbers, functions of one or several variables, and so on. Continued fractions have many remarkable properties. One fundamental property is that they can be used as an apparatus for representing real numbers (see Khinchin [59]). According to that property every real number α can be uniquely represented by a continued fraction with a value equal to α. This continued fraction will be a finite continued fraction as illustrated by equation (5.1), if α is rational, and an infinite continued fraction, if α is irrational:

$$\alpha = [a_0; a_1, a_2, a_3, \ldots] = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ddots}}}.$$

Here, $\alpha = \lim_{n\to\infty} \frac{p_n}{q_n}$ where the fraction $\frac{p_n}{q_n} = [a_0; a_1, a_2, a_3, \ldots, a_n]$ is irreducible (that is, p_n and q_n are relatively prime), and $q_n > 0$.

A theoretical advantage of such a representation is that it reflects the properties of the number that it represents in a pure form. A practical advantage is that, it provides best approximations, when we want to find values that approximate the number with any arbitrary degree of accuracy. Note that a continued fraction can also be written in other equivalent forms, one such form, which is useful for our analysis, is described below.

5.1.2 Random continued fractions and perpetuities

The main observation is the following. Suppose we have an infinite random continued fraction

$$\cfrac{1}{1 - \cfrac{R_1^*}{1 + R_1^* - \cfrac{R_2^*}{1 + R_2^* - \ddots}}},$$

where we allow the elements of an infinite continued fraction to be random with $\{R_i^*\}_{i=1}^{\infty}$ being totally independent random variables. We observe that the above structure has a perpetuity distribution. This is true because of Euler's continued fraction formula (see Wall [92]). The perpetuity is given by

$$1 + \sum_{j=1}^{\infty} \left(\prod_{i=1}^{j} R_i^* \right),$$

and it is characterized by the following stochastic fixed point equation, that could arise in a one-sided tree algorithm

$$D^* \stackrel{\mathcal{L}}{=} R^* D^* + 1.$$

Alternatively if we have

$$\cfrac{R_0^*}{1 - \cfrac{R_1^*}{1 + R_1^* - \cfrac{R_2^*}{1 + R_2^* - \ddots}}},$$

where $\{R_i^*\}_{i=0}^{\infty}$ are totally independent random variables then the transformed version is a perpetuity given by

$$\sum_{j=0}^{\infty} \left(\prod_{i=0}^{j} R_i^* \right),$$

and it is characterized by the stochastic fixed point equation

$$X^* \stackrel{\mathcal{L}}{=} R^* X^* + R^*.$$

We are investigating whether distributional properties of perpetuities (accompanying the underlying one-sided tree structure) could be studied by expressing them as a random continued fraction using Euler's continued fraction formula. It is also interesting to see how much computational gain or loss one would incur in such analysis.

5.2 Truncated geometric and truncated Poisson splitting protocols

An example where our proposed theory produces trivial asymptotic results for the depth and the cost is when the splitting protocol K_n follows a truncated geometric or a truncated Poisson distribution. In these scenarios, K_n/n converge in distribution to 0. For these type of distributions, the exact calculations are more interesting than the asymptotics. The following sequel is from Kalpathy and Ward [56], which is joint work with Professor Mark Daniel Ward from Purdue University, and is currently in preparation.

5.2.1 Duration D_n for truncated geometric splitting

We use D_n to denote the depth of a specific contestant (say Bob) when starting election with exactly n contestants. We set $D_0 = D_1 = 0$. At the start of a round, if n contestants are present, then K_n contestants advance to the next round. We handle the case where K_n is a "truncated geometric" with parameters p and $q := 1 - p$. We have,

$$\mathbf{Prob}(K_n = \ell) = cpq^\ell, \qquad \text{for some constant } c,$$

we solve for the constant by

$$1 = c \sum_{\ell=0}^{n} \mathbf{Prob}(K_n = \ell) = c \sum_{\ell=0}^{n} pq^\ell = c(1 - q^{n+1}),$$

so $c = \frac{1}{1-q^{n+1}}$. Thus the mass of K_n is

$$\mathbf{Prob}(K_n = \ell) = \frac{pq^\ell}{1 - q^{n+1}}, \qquad \text{for } \ell = 0, 1, \ldots, n.$$

Given the value of K_n, we can set up a conditional equation for the value of D_n. We have:

- If $n = 0$, then $K_0 = 0$ (always), because $D_0 = 0$ in this case (the rationale is: Nobody advances to the next round, so no additional levels were needed in the election).
- If $n = 1$, then K_1 is 0 or 1, because $D_1 = 0$ in this case (the rationale is: Only one person is present at the start of the round, so no additional levels were needed in the election).
- If $n \geq 2$, then K_n is between 0 and n (inclusive). Then:
  - Given the value of K_n, Bob is one of the K_n participants with probability K_n/n, so the *conditional distribution* of D_n (under these conditions) is the

same as the unconditional distribution of $1 + D_{K_n}$ (because one round was used, and K_n contestants will participate in the next round, including the specific contestant (Bob)).

– Given the value of K_n, Bob fails to be one of the K_n participants with probability $1 - K_n/n$, so the *conditional distribution* of D_n (under these conditions) is 1 (because one round was used, and then the specific contestant (Bob) is removed from the rest of the election).

Now we can write a recursion for $n \geq 2$, giving

$$\begin{aligned}
\phi_n(t) &= \mathbf{E}[e^{tD_n}] \\
&= \sum_{\ell=0}^{n} \mathbf{E}[e^{tD_n} \mid K_n = \ell]\,\mathbf{Prob}(K_n = \ell) \\
&= \sum_{\ell=0}^{n} \left(\mathbf{E}[e^{t(1+D_\ell)}]\,\frac{\ell}{n} + \mathbf{E}[e^{t(1)}]\left(1 - \frac{\ell}{n}\right)\right)\mathbf{Prob}(K_n = \ell) \\
&= e^t \sum_{\ell=0}^{n} \left(\mathbf{E}[e^{tD_\ell}]\,\frac{\ell}{n} + 1 - \frac{\ell}{n}\right)\mathbf{Prob}(K_n = \ell).
\end{aligned}$$

Thus, for $n \geq 2$, we have

$$\begin{aligned}
n\phi_n(t) &= e^t \sum_{\ell=0}^{n} \left(\mathbf{E}[e^{tD_\ell}]\,\ell + n - \ell\right)\mathbf{Prob}(K_n = \ell) \\
&= ne^t\phi_n(t)\frac{pq^n}{1 - q^{n+1}} + e^t \sum_{\ell=0}^{n-1} \left(\phi_\ell(t)\ell + n - \ell\right)\frac{pq^\ell}{1 - q^{n+1}}.
\end{aligned}$$

So, for $n \geq 2$, we have

$$(1 - q^{n+1})n\phi_n(t) = ne^t\phi_n(t)pq^n + e^t \sum_{\ell=0}^{n-1} \left(\phi_\ell(t)\ell + n - \ell\right)pq^\ell,$$

and thus

$$n(1 - q^n(q + e^t p))\phi_n(t) = e^t \sum_{\ell=0}^{n-1} \left(\phi_\ell(t)\ell + n - \ell\right)pq^\ell.$$

Hence, for $n \geq 3$, we have

$$(n-1)(1-q^{n-1}(q+e^t p))\phi_{n-1}(t) = e^t \sum_{\ell=0}^{n-2} (\phi_\ell(t)\ell + n - 1 - \ell)\, pq^\ell.$$

It follows that, for $n \geq 3$, we have

$$n(1-q^n(q+e^t p))\phi_n(t) - (n-1)(1-q^{n-1}(q+e^t p))\phi_{n-1}(t)$$
$$= e^t (\phi_{n-1}(t)(n-1))\, pq^{n-1} + e^t(1-q^n).$$

Rearranging the above equation, we get

$$n(1-q^n(q+e^t p))\phi_n(t) = (n-1)\phi_{n-1}(t)(1-q^n) + e^t(1-q^n).$$

Consequently for $n \geq 3$, we have

$$\phi_n(t) = \frac{(n-1)(1-q^n)}{n(1-q^n(q+e^t p))}\phi_{n-1}(t) + \frac{e^t(1-q^n)}{n(1-q^n(q+e^t p))}.$$

For $n \geq 3$, we obtain the solution

$$\phi_n(t) = \frac{2}{n}\prod_{i=3}^{n} \frac{1-q^i}{1-q^i(q+e^t p)}\phi_2(t) + \sum_{j=3}^{n} \frac{e^t}{n} \prod_{i=j}^{n} \frac{1-q^i}{1-q^i(q+e^t p)},$$

which is obtained by iteration. Also, we know that

$$\phi_2(t) = e^t \frac{p(1+q)}{1-q^2(q+pe^t)} = e^t \frac{1-q^2}{(1-q^2(q+e^t p))}.$$

Therefore, for $n \geq 3$, we find

$$\phi_n(t) = 2\frac{e^t}{n}\prod_{i=2}^{n} \frac{1-q^i}{1-q^i(q+e^t p)} + \sum_{j=3}^{n} \frac{e^t}{n} \prod_{i=j}^{n} \frac{1-q^i}{1-q^i(q+e^t p)}.$$

That is, for $n \geq 3$,

$$\phi_n(t) = \frac{e^t}{n}\prod_{i=2}^{n} \frac{1-q^i}{1-q^i(q+e^t p)} + \sum_{j=2}^{n} \frac{e^t}{n} \prod_{i=j}^{n} \frac{1-q^i}{1-q^i(q+e^t p)}.$$

Note that this holds too for $n = 2$, as we have

$$\phi_2(t) = \frac{e^t}{2}\frac{1-q^2}{1-q^2(q+e^tp)} + \frac{e^t}{2}\frac{1-q^2}{1-q^2(q+e^tp)}.$$

We define the q-Pochhammer symbol as

$$(x)_n := \prod_{j=0}^{n-1}(1-xq^j) = (1-x)(1-xq)(1-xq^2)\cdots(1-xq^{n-1}).$$

Therefore, we have

$$\prod_{i=2}^{n}(1-q^i) = (1-q^2)(1-q^3)\cdots(1-q^n) = \frac{(q)_n}{(1-q)}.$$

For our moment generating function, we have

$$\begin{aligned}\prod_{i=2}^{n}(1-q^i(q+e^tp)) &= (1-q^2(q+e^tp))(1-q^3(q+e^tp))\cdots(1-q^n(q+e^tp)) \\ &= \frac{(q+e^tp)_{n+1}}{(1-(q+e^tp))(1-q(q+e^tp))}.\end{aligned}$$

Note that

$$\prod_{i=j}^{n}(1-q^i) = (1-q^j)(1-q^{j+1})\cdots(1-q^n) = \frac{(q)_n}{(1-q)(1-q^2)\cdots(1-q^{j-1})}.$$

Therefore, we have

$$\begin{aligned}\prod_{i=j}^{n}(1-q^i(q+e^tp)) &= (1-q^j(q+e^tp))(1-q^{j+1}(q+e^tp))\cdots(1-q^n(q+e^tp)) \\ &= \frac{(q+e^tp)_{n+1}}{(1-(q+e^tp))(1-q(q+e^tp))\cdots(1-q^{j-1}(q+e^tp))}.\end{aligned}$$

For $n \geq 2$, we have

$$\begin{aligned}\phi_n(t) = &\frac{e^t}{n}\frac{(q)_n}{(1-q)}\frac{(1-(q+e^tp))(1-q(q+e^tp))}{(q+e^tp)_{n+1}} \\ &+ \sum_{j=2}^{n}\frac{e^t}{n}\frac{(q)_n}{(1-q)(1-q^2)\cdots(1-q^{j-1})} \\ &\times \frac{(1-(q+e^tp))(1-q(q+e^tp))\cdots(1-q^{j-1}(q+e^tp))}{(q+e^tp)_{n+1}}\end{aligned}$$

Therefore, for $n \geq 2$, we have

$$\phi_n(t) = \frac{e^t}{n}\frac{(q)_n}{(q+e^tp)_{n+1}}\Big(\frac{(1-(q+e^tp))(1-q(q+e^tp))}{(1-q)} + \sum_{j=2}^{n}\frac{(1-(q+e^tp))(1-q(q+e^tp))\cdots(1-q^{j-1}(q+e^tp))}{(1-q)(1-q^2)\cdots(1-q^{j-1})}\Big)$$

Now we analyze our moment generating function on the complex domain $\mathbb{C}$. We use

$$(x)_z := \frac{(x)_\infty}{(xq^z)_\infty}.$$

For $n \geq 2$, we have the following expression for the moment generating function:

$$\phi_n(t) = \frac{e^t}{n}\frac{(q)_\infty((q+e^tp)q^{n+1})_\infty}{(q^{n+1})_\infty(q+e^tp)_\infty}\Big(\frac{(1-(q+e^tp))(1-q(q+e^tp))}{(1-q)} + \sum_{j=2}^{n}\frac{(1-(q+e^tp))(1-q(q+e^tp))\cdots(1-q^{j-1}(q+e^tp))}{(1-q)(1-q^2)\cdots(1-q^{j-1})}\Big).$$

Differentiating with respect to t, and setting $t = 0$, yields

$$\begin{aligned}
\mathbf{E}[D_n] &= \left(\frac{d}{dt}\phi_n(t)\right)\Big|_{t=0} \\
&= \Big(\frac{d}{dt}\frac{e^t}{n}\frac{(q)_\infty((q+e^tp)q^{n+1})_\infty}{(q^{n+1})_\infty(q+e^tp)_\infty}\Big(\frac{(1-(q+e^tp))(1-q(q+e^tp))}{(1-q)} \\
&\quad + \sum_{j=2}^{n}\frac{(1-(q+e^tp))(1-q(q+e^tp))\cdots(1-q^{j-1}(q+e^tp))}{(1-q)(1-q^2)\cdots(1-q^{j-1})}\Big)\Big)\Big|_{t=0} \\
&= \Big(\Big(\frac{d}{dt}e^t\Big)\frac{1}{n}\frac{(q)_\infty((q+e^tp)q^{n+1})_\infty}{(q^{n+1})_\infty(q+e^tp)_\infty}\Big(\frac{(1-(q+e^tp))(1-q(q+e^tp))}{(1-q)} \\
&\quad + \sum_{j=2}^{n}\frac{(1-(q+e^tp))(1-q(q+e^tp))\cdots(1-q^{j-1}(q+e^tp))}{(1-q)(1-q^2)\cdots(1-q^{j-1})}\Big) \\
&+ \Big(\frac{d}{dt}((q+e^tp)q^{n+1})_\infty\Big)\frac{e^t}{n}\frac{(q)_\infty}{(q^{n+1})_\infty(q+e^tp)_\infty}\Big(\frac{(1-(q+e^tp))(1-q(q+e^tp))}{(1-q)} \\
&\quad + \sum_{j=2}^{n}\frac{(1-(q+e^tp))(1-q(q+e^tp))\cdots(1-q^{j-1}(q+e^tp))}{(1-q)(1-q^2)\cdots(1-q^{j-1})}\Big) \\
&+ \Big(\frac{d}{dt}\frac{1}{(q+e^tp)_\infty}\Big)\frac{e^t}{n}\frac{(q)_\infty((q+e^tp)q^{n+1})_\infty}{(q^{n+1})_\infty}\Big(\frac{(1-(q+e^tp))(1-q(q+e^tp))}{(1-q)} \\
&\quad + \sum_{j=2}^{n}\frac{(1-(q+e^tp))(1-q(q+e^tp))\cdots(1-q^{j-1}(q+e^tp))}{(1-q)(1-q^2)\cdots(1-q^{j-1})}\Big)
\end{aligned}$$

$$+\frac{e^t}{n}\frac{(q)_\infty((q+e^tp)q^{n+1})_\infty}{(q^{n+1})_\infty(q+e^tp)_\infty}\Big(\frac{d}{dt}\frac{(1-(q+e^tp))(1-q(q+e^tp))}{(1-q)}$$
$$+\frac{d}{dt}\sum_{j=2}^{n}\frac{(1-(q+e^tp))(1-q(q+e^tp))\cdots(1-q^{j-1}(q+e^tp))}{(1-q)(1-q^2)\cdots(1-q^{j-1})}\Big)\Big)\Big|_{t=0}$$

After doing some algebra, the above calculation reduces to the following expression for the mean

$$\begin{aligned}
\mathbf{E}[D_n] &= 1+\sum_{k=0}^{\infty}\frac{-pq^{n+1+k}}{1-q^{n+1+k}}+\sum_{k=1}^{\infty}\frac{pq^k}{1-q^k}+\frac{1}{n}\Big(\frac{-qp}{1-q}+\sum_{j=2}^{n}\sum_{k=1}^{j-1}\frac{-pq^k}{1-q^k}\Big)\\
&= 1+\sum_{k=n+1}^{\infty}\frac{-pq^k}{1-q^k}+\sum_{k=1}^{\infty}\frac{pq^k}{1-q^k}+\frac{1}{n}\Big(-q+\sum_{k=1}^{n-1}(n-k)\frac{-pq^k}{1-q^k}\Big)\\
&= 1+\sum_{k=1}^{n}\frac{pq^k}{1-q^k}+\frac{1}{n}\Big(-q+\sum_{k=1}^{n-1}(n-k)\frac{-pq^k}{1-q^k}\Big)\\
&= 1+\sum_{k=1}^{n}\frac{pq^k}{1-q^k}-\frac{q}{n}+\sum_{k=1}^{n-1}\frac{-pq^k}{1-q^k}+\sum_{k=1}^{n-1}(-k/n)\frac{-pq^k}{1-q^k}\\
&= 1+\frac{1}{n}\sum_{k=2}^{n}\frac{kpq^k}{1-q^k}.
\end{aligned}$$

The above mean calculation illustrates the kinds of results that one obtains with a truncated geometric splitting protocol. We are currently investigating the variance and the moment generating function for this splitting protocol.

Proceeding on similar lines, we would like to investigate other truncated discrete distributions such as a truncated Poisson splitting protocol. The development may involve a q-series analysis or other analytic techniques.

5.3 Dynamic leader election algorithms

The United States presidential election debates has a selection process where at each stage the candidates debate on different policies. Also, in some sports and games like the decathlon and mixed poker games like HORSE, the contestants compete in different games all rolled up into one. We would like to investigate a leader election algorithm having this sort of a flavor wherein candidates are selected at each stage in a dynamic manner, with changing criteria at each level. For example, at the first level the contestants advance based on a binomial splitting protocol, in the second stage the candidate winners from the first stage advance based on a uniform splitting protocol, and at the third stage they may advance based on a power distribution splitting, and so on. We think it will be interesting to study the interplay of these splitting protocols and their relationship to the limiting distribution of the duration of an individual contestant and the limiting behavior of the total time or cost of the competition.

5.4 Survivors in leader election algorithms

The following sequel is from Kalpathy, Mahmoud, and Rosenkrantz [54], which is a joint work with Professor Walter Rosenkrantz from University of Massachusetts Amherst and Professor Hosam Mahmoud. This section was born out of an interesting question that was raised by Professor Rosenkrantz during the author's thesis defense. Here we consider the number of survivors in a broad class of fair leader election algorithms after a number of election rounds. This measure can be important from the point of view of the moderator. Starting with n contestants, we specify a number t (possibly dependent on n), and ask how many survivors are still in contention after t rounds. This parameter may be important in planning ahead of time such resources as the space needed to hold the contestants. We give sufficient conditions for the number of survivors to converge to a product of independent identically distributed random variables. The number of terms in the product is determined by the round number considered. Each individual term in the product is a limit of a scaled random variable associated with the splitting protocol. The proof is established via convergence (to 0) of the first-order Wasserstein distance from the product limit. In a broader context, this work is a case study of a class of stochastic recursive equations.

Bibliography

[1] Aebi, M., Embrechts, P., and Mikosch, T. (1994). Stochastic discounting, aggregate claims, and the bootstrap. *Advances in Applied Probability*, **26**, 183–206.

[2] Alsmeyer, G., Iksanov, A. and Rösler, U. (2009). On distributional properties of perpetuities. *Journal of Theoretical Probability*, **22**, 666–682.

[3] Arratia, R., Barbour, A. and Tavaré, S. (2003). *Logarithmic combinatorial structures: a probabilistic approach*. European Mathematical Society, Zürich.

[4] Barbour, A., Holst, L. and Janson, S. (1992). *Poisson Approximation*. Oxford University Press, Oxford, UK.

[5] Baron, M. and Rukhin, A. (2001). Perpetuities and asymptotic change-point analysis. *Statistics and Probability Letters*, **55**, 29–38.

[6] Bickel, P. and Freedman D. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics*, **9**, 1196–1217.

[7] Billingsley, P. (1995). *Probability and Measure*, Third Edition. Wiley, New York, NY.

[8] Blanchet, J. and Sigman, K. (2011). On exact sampling of stochastic perpetuities. *Journal of Applied Probability Special Volume*, New frontiers in applied probability: A Festschrift for Søren Asmussen, **48A**, 165–182.

[9] Brandt, A., Franken, P. and Lisek, B. (1990). *Stationary Stochastic Models*. Wiley, Chichester, UK.

[10] Carmona, P., Petit, F. and Yor, M. (2001). Exponential functionals of Lévy processes. In *Lévy Processes*, Eds. Barndorff-Nielsen, O. E., Mikosch, T. and Resnick, S. I., 41–55.

[11] Cavalli-Sforza, L. (1975). Cultural and biological evolution: a theoretical inquiry. In *Proceedings of the Conference on Directions for Mathematical Statistics*, Ed. Ghurye, S. G., *Supplement to Advances in Applied Probability*, **7**, 90–99.

[12] Cavalli-Sforza, L. and Feldman, M. (1973). Models for cultural inheritance I. Group mean and within group variation. *Theoretical Population Biology*, **4**, 42–55.

[13] Comets, F., Gantert, N. and Zeitouni, O. (2000). Quenched, annealed and functional large deviations for one-dimensional random walk in random environment. *Probability Theory and Related Fields*, **118**, 65–114.

[14] Chamayou, J. (1973). A probabilistic approach to a differential-difference equation arising in analytic number theory. *Mathematics of Computation*, **27**, 197–203.

[15] Chamayou, J. and Letac, G. (1991). Explicit stationary distributions for compositions of random functions and products of random matrices. *Journal of Theoretical Probability*, **4**, 3–36.

[16] Chamayou, J. and Schorr, B. (1975). On a class of random variables arising in atomic cascade models. Report, European Organization for Nuclear Research, Geneva.

[17] Chandrasekhar, S. and Münch, G. (1950). The theory of the fluctuations in brightness of the Milky Way, I and II. *The Astrophysical Journal*, **112**, 380–398.

[18] Chung, K. (1974). *A Course in Probability Theory*, Second Edition. Academic Press, New York, NY.

[19] de Bruijn, N. (1951). The asymptotic behaviour of a function occurring in the theory of primes. *Journal of the Indian Mathematical Society (N.S.)*, **15**, 25–32.

[20] Devroye, L. (2001). Simulating perpetuities. *Methodology and Computing in Applied Probability*, **3**, 97–115.

[21] Devroye, L. and Fawzi, O. (2010). Simulating the Dickman distribution. *Statistics and Probability Letters*, **80**, 242–247.

[22] Donnelly, P. and Grimmett, G. (1993). On the asymptotic distribution of large prime factors. *Journal of the London Mathematical Society*, **47**, 395–404.

[23] Diaconis, P. and Freedman, D. (1999). Iterated random functions. *SIAM Review*, **41**, 45–76.

[24] Dufresne, D. (1990). The distribution of a perpetuity, with applications to risk theory and pension funding. *Scandinavian Actuarial Journal*, **1990**, 39–79.

[25] Elmasry, A. and Mahmoud, H. (2011). Analysis of swaps in radix selection. *Advances in Applied Probability*, **43**, 524–544.

[26] Embrechts, P. and Goldie, C. (1994). Perpetuities and random equations. In *Asymptotic Statistics: Proceedings of the Fifth Prague Symposium*, Eds. Mandl, P. and Hušková, M., 75–86, Physica-Verlag, Heidelberg.

[27] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events*. Springer-Verlag, Berlin.

[28] Fill, J., Mahmoud, H. and Szpankowski, W. (1996). The distribution for the duration of a randomized leader election algorithm. *The Annals of Applied Probability*, **6**, 1260–1283.

[29] Fill, J. and Huber, M. (2010). Perfect simulation of Vervaat perpetuities. *Electronic Journal of Probability*, **15**, 96–109.

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

[31] Geman, H. and Yor, M. (1993). Bessel processes, Asian options, and perpetuities. *Mathematical Finance*, **3**, 349–375.

[32] Goldie, C. (1991). Implicit renewal theory and tails of solutions of random equations. *The Annals of Applied Probability*, **1**, 126–166.

[33] Goldie, C. and Grübel, R. (1996). Perpetuities with thin tails. *Advances in Applied Probability*, **28**, 463–480.

[34] Goldie, C. and Maller, R. (2000). Stability of perpetuities. *The Annals of Probability*, **28**, 1195–1218.

[35] Granas, A. and Dugundji, J. (2003). *Fixed Point Theory*. Springer-Verlag, New York, NY.

[36] Grintsevichyus, A. (1974). On the continuity of the distribution of a sum of dependent variables connected with independent walks on lines. *Theory of Probability and Its Applications*, **19**, 163–168.

[37] Grübel, R. and Pitts, S. (2000). Statistical aspects of perpetuities. *Journal of Multivariate Analysis*, **75**, 143–162.

[38] Grübel, R. and Rösler, U. (1996). Asymptotic distribution theory for Hoare's selection algorithm. *Advances in Applied Probability*, **28**, 252–269.

[39] Hitczenko, P. (2010). On tails of perpetuities. *Journal of Applied Probability*, **47**, 1191–1194.

[40] Hitczenko, P. and Medvedev, G. (2009). Bursting oscillations induced by small noise. *SIAM Journal on Applied Mathematics*, **69**, 1359–1392.

[41] Hitczenko, P. and Wesołowski, J. (2009). Perpetuities with thin tails revisited. *The Annals of Applied Probability*, **19**, 2080–2101.

[42] Hwang, H. and Tsai, T. (2002). Quickselect and the Dickman function. *Combinatorics, Probability and Computing*, **11**, 353–371.

[43] Iksanov, A. and Jurek, Z. (2002). On fixed points of Poisson shot noise transforms. *Advances in Applied Probability*, **34**, 798–825.

[44] Iksanov, A., Jurek, Z. and Schreiber, B. (2004). A new factorization property of the selfdecomposable probability measures. *The Annals of Probability*, **32**, 1356–1369.

[45] Itoh, Y. and Mahmoud, H. (2003). One-sided variations on interval trees. *Journal of Applied Probability*, **40**, 654–670.

[46] Janson, S. and Szpankowski, W. (1997). Analysis of an asymmetric leader election algorithm. *Electronic Journal of Combinatorics*, **4**, R17.

[47] Janson, S., Lavault, C. and Louchard, G. (2008). Convergence of some leader election algorithms. *Discrete Mathematics and Theoretical Computer Science*, **10**, 171–196.

[48] Jurek, Z. (1999). Selfdecomposability perpetuity laws and stopping times. *Probability and Mathematical Statistics*, **19**, 413–419.

[49] Kagan, A., Linnik, Yu. and Rao, C. (1973). *Characterization Problems in Mathematical Statistics*. Wiley, New York, NY.

[50] Kalman, R. (1962). Control of randomly varying linear dynamical systems. In *Proceedings of Symposia in Applied Mathematics*, **13**, 287–298, American Mathematical Society, Providence, RI.

[51] Kalpathy, R. (2013+). On the relationship between perpetuities and continued fraction distributions *(in preparation)*.

[52] Kalpathy, R. and Mahmoud, H. (2013). Perpetuities in fair leader election algorithm. Extended abstract: In *Proceedings of the Tenth ACM-SIAM Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*, New Orleans. Eds. Nebel, M. and Szpankowski, W., 61–70.

[53] Kalpathy, R. and Mahmoud, H. (2014). Perpetuities in fair leader election algorithms. *Advances in Applied Probability*, **46** *(to appear)*.

[54] Kalpathy, R., Mahmoud, H. and Rosenkrantz, W. (2013+). Survivors in leader election algorithms *(submitted)*.

[55] Kalpathy, R., Mahmoud, H. and Ward, M. (2011). Asymptotic properties of a leader election algorithm. *Journal of Applied Probability*, **48**, 569–575.

[56] Kalpathy, R. and Ward, M. (2013+). Truncated distributions in leader election algorithms *(in preparation)*.

[57] Karr, A. (1993). *Probability*. Springer-Verlag, New York, NY.

[58] Kesten, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, **131**, 207–248.

[59] Khinchin, A. (1964). *Continued Fractions*. The University of Chicago Press, Chicago, IL.

[60] Kirschenhofer, P. and Prodinger, H. (1998). Comparisons in Hoare's Find Algorithm. *Combinatorics, Probability and Computing*, **7**, 111–120.

[61] Knape, M. and Neininger, R. (2008). Approximating perpetuities. *Methodology and Computing in Applied Probability*, **10**, 507–529.

[62] Kuba, M. and Panholzer, A. (2007). On weighted path lengths and distances in increasing trees. *Probability in the Engineering and Informational Sciences*, **21**, 419–433.

[63] Lawrance, A. and Kottegoda, N. (1977). Stochastic modelling of riverflow time series. *Journal of the Royal Statistical Society. Series A (General)*, **140**, 1–47.

[64] Letac, G. (1986) A contraction principle for certain Markov chains and its applications. In *Random Matrices and their Applications. Proceedings of the AMS-IMS-SIAM Joint Summer Research Conference 1984*, Brunswick, Maine. Eds. Cohen, J., Kesten, H. and Newman, C., *Contemporary Mathematics*, **50**, 263–273, American Mathematical Society, Providence, RI.

[65] Lindley, D. (1952). The theory of queues with a single server. In *Mathematical Proceedings of the Cambridge Philosophical Society*, **48**, 277–289.

[66] Louchard, G. and Prodinger, H. (2008) Advancing in the presence of a demon. *Mathematica Slovaca*, **58**, 263–276.

[67] Louchard, G., Martínez, C. and Prodinger, H. (2011). The Swedish leader election protocol: Analysis and variations. In *Proceedings of the Eighth ACM-SIAM Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*, San Francisco. Eds. Flajolet, P. and Panario, D., 127–134.

[68] Louchard, G., Prodinger, H. and Ward, M. (2012). Number of survivors in the presence of a demon. *Periodica Mathematica Hungarica*, **64**, 101–117.

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

[70] Mahmoud, H. (2010). Distributional analysis of swaps in Quick Select. *Theoretical Computer Science*, **411**, 1763–1769.

[71] Mahmoud, H., Modarres, R. and Smythe, R. (1995). Analysis of quickselect: an algorithm for order statistics. *RAIRO Informatique théorique et Applications*, **29**, 255–276.

[72] Maulik, K. and Zwart, B. (2006). Tail asymptotics for exponential functionals of Lévy processes. *Stochastic Processes and their Applications*, **116**, 156–177.

[73] Mohamed, H. (2006). A probabilistic analysis of a leader election algorithm. *DMTCS Proceedings*, **AG**, 225–236.

[74] Neininger, R. and Rüschendorf, L. (2004). A general limit theorem for recursive algorithms and combinatorial structures. *The Annals of Applied Probability*, **14**, 378–418.

[75] Nolan, J. (2013). *Stable Distributions—Models for Heavy Tailed Data*. Birkhäuser, Boston, MA.

[76] Paulson, A. and Uppuluri, V. (1972). Limit laws of a sequence determined by a random difference equation governing a one-compartment system. *Mathematical Biosciences*, **13**, 325–333.

[77] Preater, J. (1997). A perpetuity and the $M/M/\infty$ ranked server system. *Journal of Applied Probability*, **34**, 508–513.

[78] Prodinger, H. (1993). How to select a loser. *Discrete Mathematics*, **120**, 149–159.

[79] Puri, P. (1987). On almost sure convergence of an erosion process due to Todorovic and Gani. *Journal of Applied Probability*, **24**, 1001–1005.

[80] Rachev, S. and Samorodnitsky, G. (1995). Limit laws for a stochastic process and random recursion arising in probabilistic modelling. *Advances in Applied Probability*, **27**, 185–202.

[81] Rachev, S. and Rüschendorf, L. (1995). Probability metrics and recursive algorithms. *Advances in Applied Probability*, **27**, 770–799.

[82] Rachev, S. and Rüschendorf, L. (1998). *Mass Transportation Problems. Vol. II: Applications*. Springer-Verlag, New York, NY.

[83] Rachev, S. and Todorovic, P. (1990). On the rate of convergence of some functionals of a stochastic process. *Journal of Applied Probability*, **28**, 805–814.

[84] Rösler, U. (1991). A limit theorem for “Quicksort”. *RAIRO Informatique théorique et Applications*, **25**, 85–100.

[85] Rösler, U. (2004). QUICKSELECT revisited. *Journal of the Iranian Statistical Society*, **3**, 271–296.

[86] Rösler, U. and Rüschendorf, L. (2001). The contraction method for recursive algorithms. *Algorithmica*, **29**, 3–33.

[87] Solomon, F. (1975). Random walks in a random environment. *The Annals of Probability*, **3**, 1–31.

[88] Takács, L. (1955). On stochastic processes connected with certain physical recording apparatuses. *Acta Mathematica Academiae Scientiarum Hungaricae*, **6**, 363–380.

[89] Todorovic, P. (1987). An extremal problem arising in soil erosion modeling. In *Applied Probability, Stochastic Processes and Sampling Theory*, Eds. MacNeil, I. B. and Umphrey, G. J., 65-73, Reidel, Dordrecht.

[90] Todorovic, P. and Gani, J. (1987). Modeling the effect of erosion on crop production. *Journal of Applied Probability*, **24**, 787–797.

[91] Vervaat, W. (1979). On a stochastic difference equation and a representation of non-negative infinitely divisible random variables. *Advances in Applied Probability*, **11**, 750–783.

[92] Wall, H. (1948). *Analytic Theory of Continued Fractions*. D. Van Nostrand Company, New York, NY.

[93] Weiss, G. (1977). Shot noise models for the generation of synthetic streamflow data. *Water Resources Research*, **13**, 101–108.

Appendix A

For distance calculation to be meaningful, we need versions of D_n and D^* that are defined (coupled) on the same probability space. For instance, even if K_n and K^* are defined on different spaces in regularity condition (i), we can find versions distributed like them on the same space. This is possible in view of regularity condition (i)—since $K_n^* = K_n/n$ converges in $\mathcal{L}_1$ to K^*, it converges in distribution too, and thus we can use Skorohod's representation (see Billingsley [7], p. 333), to find a probability space on which we can define $\tilde{K}_n \stackrel{\mathcal{L}}{=} K_n$ (and so $\tilde{K}_n/n := \tilde{K}_n^* \stackrel{\mathcal{L}}{=} K_n^*$), $\tilde{K}^* \stackrel{\mathcal{L}}{=} K^*$, with $\tilde{K}_n^* \stackrel{a.s.}{\longrightarrow} \tilde{K}^*$. On this space, let the right-hand side of (3.1), written with $\tilde{K}_n^*$ replacing K_n^* be called $\tilde{D}_n$, and let the right-hand side of (3.3), written with $\tilde{K}^*$ replacing K^*, be called $\tilde{D}^*$. Then $\tilde{D}_n$ and $\tilde{D}^*$ are "coupled" versions of D_n and D^* (having the same respective distributions F_{D_n} and F_{D^*}) that are well defined on the same probability space. In that space, $\tilde{D}^*$ is independent of $\tilde{K}_n^*$ and $\tilde{K}^*$. Similar considerations can be taken to have all the random variables in (3.5) and (3.6) defined on the same probability space.

Appendix B

Here we show that a sequence K_n of random variables satisfying regularity condition (i) satisfies

$$\mathbf{1}_{\{U<\frac{K_n}{n}\}} = \mathbf{1}_{\{U<K^*\}} + O_{\mathcal{L}_1}\Big(\frac{1}{n^\alpha}\Big).$$

According to condition (i), we can write $K_n/n = K^* + R_n$, for a random remainder R_n, with $\mathbf{E}[|R_n|] = O\Big(\frac{1}{n^\alpha}\Big)$. We now have

$$\begin{aligned}
\mathbf{E}\Big[\Big|\mathbf{1}_{\{U<\frac{K_n}{n}\}} - \mathbf{1}_{\{U<K^*\}}\Big|\Big] &= \mathbf{E}\Big[\Big|\mathbf{1}_{\{U<K^*+R_n\}} - \mathbf{1}_{\{U<K^*\}}\Big|\Big] \\
&= \int_{k^*=0}^{1}\int_{r=-\infty}^{\infty}\int_{u=0}^{1} \Big|\mathbf{1}_{\{u<k^*+r\}} - \mathbf{1}_{\{u<k^*\}}\Big| \\
&\qquad \times dF_U(u)\, dF_{R_n}(r)\, dF_{K^*}(k^*) \\
&= \int_{k^*=0}^{1}\int_{r=-\infty}^{\infty}\int_{u=k^*}^{k^*+|r|} du\, dF_{R_n}(r)\, dF_{K^*}(k^*) \\
&= \int_{k^*=0}^{1}\int_{r=-\infty}^{\infty} |r|\, dF_{R_n}(r)\, dF_{K^*}(k^*) \\
&= \mathbf{E}\Big[|R_n|\Big] \\
&= O\Big(\frac{1}{n^\alpha}\Big),
\end{aligned}$$

which proves the assertion.