Partial Quicksort and weighted branching processes

Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Mahmoud Ragab

Kiel, 2011

Referent:	Prof. Dr. Uwe Roesler
Korreferent:	Prof. Dr. Albrecht Irle
Tag der mündlichen Prüfung:	06.12.2011
Zum Druck genehmigt: Kiel,	06.12.2011

Der Dekan, gez. Prof. Dr. Lutz Kipp

Acknowledgments

First and foremost, I would like to express my sincere appreciation to my supervisor Prof. Dr. Uwe Roesler for suggesting the topics of this thesis and valuable guidance throughout the development of this work. I am indebted to him for his great encouragement and endless help, kindness and his fatherly guidance as well as for his patience and his generosity with ideas. He is surely a model supervisor.

Furthermore, I want to express my deepest thanks to my colleagues, the staff members of the “Arbeitsgruppe Stochastik” for their collaboration, practical assistance and friendship. Special thanks go to Soeren Christensen, you are always friendly and helpful. Jonas Kauschke is always nice, patient and understanding. I enjoyed our discussions when we both worked in the same office.

A special hugs and kisses goes to both my sons, Ahmad and Yusof.

Last but not least, I would like to thank my beloved wife, Sara. Clearly, there are not enough words to express my gratitude for your patience, encouragement, continuous support during my work, and for so much more than I can write down. I love you and look forward to spend the rest of my life with you. I dedicate this thesis to you.

Partial Quicksort and Weighted Branching Process

Mahmoud Ragab

Abstract

In this dissertation we look at different two models of sorting algorithms based on divide-and-conquer algorithms. Quicksort algorithm, sort an unsorted array of n distinct elements. Partial Quicksort sorts the l smallest elements in a list of length n. Both stochastic divide-and-conquer algorithms are widely studied. Our algorithm Quicksort on the fly provides online the first smallest, then second smallest and so on. If we stop at the l-th smallest, we obtain Partial Quicksort. We analyze the running time performance Y_n of Quicksort on the fly using the parameter l as time index. We show that, the process Y_n converges not only in distribution, but also uniformly as $n \to \infty$ almost everywhere in D to a random variable Y. The distribution of Y is characterized as a solution of a stochastic fixed point equation

$$Y \stackrel{\mathcal{D}}{=} \sum_{i \in \mathbb{N}} A_i Y_i \circ B_i + C$$

with values in the space D of cadlag functions on the unit interval. This result includes the performance of Quicksort via $l = n$, respectively $Y(1)$ in the limit.

Contents

Chapter 1

Introduction

In this thesis we focus on the Partial Quicksort algorithm, first proposed and analyzed by Martinez [32]. The input is a list of n distinct reals and the output of the algorithm are the sorted l smallest elements. Basically find with Quickselect [27] first the list of l smallest elements and then sort them by Quicksort [27]. It works as follow

- Choose a pivot with random with a uniform distribution;
- Find the list of strictly smaller numbers and the list of strictly larger numbers;
- If the rank of the pivot is smaller than l
  - Sort the left sub-array using 'Quicksort';
  - Apply Partial Quicksort to the right sub-array.
- If the rank of the pivot is greater than l
  - Apply Partial Quicksort to the left sub-array.
- Recall Partial Quicksort till termination.

We introduce a nice version of Partial Quicksort, called Quicksort on the fly. Our algorithm is a suitable rearrangement of Partial Quicksort, it provides first the smallest, then second smallest, and so on until the largest element of a list of n different reals. Quicksort on the fly works as follow

- Pick by random with uniform distribution an element of the list as a pivot.
- Compare all others to the pivot and form the list of strictly smaller numbers, pivot, and the list of strictly larger numbers in this order.
- Recall independently the algorithm for the left sided list until termination.
- If necessary recall the algorithm for the right sided list.

For given array with large size it can be assumed that the running time of the algorithm is proportional to the number of comparisons needed to sort the array. Let $X(n, l)$ be the number of comparisons done by Partial Quicksort to sort the l smallest elements out of n elements. The average number [32] is

$$E(X(n,l)) = 2n + 2(n+1)H_n - 2(n+3-l)H_{n+1-l} - 6l + 6$$

With the normalization

$$Y_n(\frac{l}{n}) = \frac{X(n,l) - E(X(n,l))}{n} \tag{1.1}$$

$Y_n(\frac{l_n}{n})$ will converge in distribution to some $Y(t)$ as $\frac{l_n}{n} \to t$, [33]. Further there is a characterization of the one-dimensional marginals via a unique solution Y with values in the space D, of cadlag functions, satisfying some fixed point equation [33].
In our paper we consider l not as a fixed value as above but as a variable and we consider Y_n as a D valued precess. Our first main result is the following.

Theorem 1.1. *The process $(Y_n(\frac{l}{n}))_{l\in\{0,1,2,\cdots,n\}}$ converges as $n \to \infty$ in distribution to a process $Y = (Y(t))_{t\in[0,1]}$ with values in D satisfying the stochastic fixed point equation*

$$(Y(t))_t \stackrel{\mathcal{D}}{=} (\mathbb{1}_{t\geq U}\left(UY^1(1) + (1-U)Y^2(\frac{t-U}{1-U})\right) + \mathbb{1}_{t<U}UY^1(\frac{t}{U}) + C(U,t))_t \tag{1.2}$$

Here U is a uniformly distributed random variable on the unit interval $[0, 1]$. The random variables Y^1, Y^2, U are independent. The random variables Y^1 and Y^2 have the same distribution as Y with values in D. The cost function $C = C(\cdot,\cdot)$ is given by

$$\begin{aligned} C(x,t) &= 1 + 2x\ln x + 2(1-x)\ln(1-x) + 2\mathbb{1}_{t\leq x}((1-t)\ln(1-t) \\ &- (1-x)\ln(1-x) - (x-t)\ln(x-t) - (1-x)) \end{aligned}$$

Our second main result provides a very specific version of Y via the weighted branching process.

Theorem 1.2. *There exists random variables Y, Y^1, Y^2, U and C such that equation* (1.2) *holds even almost everywhere.*

In our work, we use a smart approach to show the existence of equation (1.2) via the weighted branching process [41]. This approach is inspired by the analysis of Quickselect [17, 24, 33], where fixed point equations on D were considered for the first time.
Let $(V, ((T_1, T_2), C), (D, *), (D_\uparrow, \otimes))$ be the corresponding weighted branching process, which we will discuss in Chapter 3. Then the total weight up to generation n for the tree $\mathbb{V}$ with root v

$$R_n^v := \sum_{v\in\mathbb{V}_{<n}} L_w^v \otimes C^{vw}$$

will converge not only in distribution, but also uniformly as $n \to \infty$ almost everywhere in D to a random variable R^v. For all $v \in \mathbb{V}$ the family $R^v, v \in \mathbb{V}$ satisfies

$$R^v = \sum_{i=1}^{2} T_i^v R^{vi} + C^v$$

almost everywhere. This is exactly the equation (1.2) for random variables. Moreover, for every $p > 1$ holds

$$\| \|R_n\|_\infty \|_p \leq \frac{8 + (\frac{1}{p+1})^{\frac{1}{p}} \|Q\|_p}{1 - k_p}$$

where $k_p = (\frac{2}{p+1})^{\frac{1}{p}}$ and Q is a random variable with the Quicksort distribution.
The almost everywhere convergence is much stronger than convergence in distribution. The interplay of convergence of measures and of random variables, forward and backward view, is the key to our main result. As a motivation of these results, we give in the following section a short overview on the previous related work including some basic facts. In Section 1.2 we describe briefly the outline of this thesis.

1.1 A brief overview

Our approach was inspired by the methods used for the analysis of Quicksort and Quickselect. We present a short overview of the ongoing, since it is similar to our more refined approach.

Quicksort sorts an array of n distinct reals and is a well known and popular sorting algorithm. Quicksort was invented by Hoare [20] and based on a divide and conquer strategy, described in Chapter 4. Quicksort is considered as one of the ten algorithms with the greatest influence on the development and practical impact of science and engineering in the 20th century [14]. There is a large number of contributions to the Quicksort analysis, for example we mention [10], [19], [37], [39], [30], [11], [36], [8], [23] and [15].
Let X_n be the number of comparisons used by Quicksort to sort a list of size n. The random variable X_n is basically proportional to the running time of Quicksort [45], which depends (a little bit) on the implementation and computer hardware. The average number $E(X_n)$ of comparisons [27] is $E(X_n) \approx n \ln n$. The first complete running time analysis for a random divide and conquer was for Quicksort [39]. The random variable

$$Y_n = \frac{X_n - E(X_n)}{n}$$

converges in distribution to a random variable Y, which distribution is characterized as the unique solution of the stochastic fixed point equation

$$Y \stackrel{\mathcal{D}}{=} UY^1 + (1-U)Y^2 + C(U)$$

with expectation 0 and finite variance. Here U is uniformly distributed on $[0,1]$ and U, Y^1, Y^2 are independent. Y^1 and Y^2 have the same distribution and C is given by

$$C(x) := 2x \ln x + 2(1-x)\ln(1-x) + 1, \qquad x \in [0,1]$$

Quickselect or FIND, introduced by Hoare [28] in 1961 is a search algorithm widely used for finding the l-th smallest element out of n distinct numbers. Most of the mathematical results on the complexity of Quickselect are about expectations or distributions for the number of comparisons needed to complete its task by the algorithm [18]. A pivot is uniformly chosen at random from the available n elements, and compares the $n-1$ remaining elements against it.

Let $X_n(k)$ be the number of comparisons needed to find the l-th smallest out of n. The running time of this algorithm is always a random variable either by random input or internal randomness. The expectation of $X_n(k)$ is explicitly known [26]

$$E(X_n(k)) = 2(n + 3 + (n+1)H_n - (k+2)H_k - (n+3-k)H_{n+1-k})$$

for $1 \leq k \leq n$, and H_k denotes the k-th harmonic number. An asymptotic approximation as $n \to \infty$ is

$$\frac{E(X_n(k))}{n} \approx 2 - 2t\ln t - 2(1-t)\ln(1-t)$$

for $0 \leq t = \frac{k}{n} \leq 1$ [18]. The asymptotic variance $Var(X_n(k))$ was derived by Kirschenhofer and Prodinger [22] using combinatorial and generating function methods. Furthermore, [9], [18], [31], [28], [17] and [42] studied the limiting distribution of $X_n(k)$ or X_n as a process.
A major tool are fixed point equation and the contraction method for operator K like

$$K(\mu) \stackrel{\mathcal{D}}{=} \sum_{i \in \mathbb{N}} A_i Y_i + C \tag{1.3}$$

The random variables $(A, B, C), Y_i, i \in \mathbb{N}$ are independent and the random variables Y_i have the same distribution μ. In a more general form of (1.2), Knof and Roesler [25] considered general recurrence

$$Y^n \stackrel{\mathcal{D}}{=} \sum_{i \in \mathbb{N}} A_i^n Y_i^{I_n} \circ B_i^n + C^n \tag{1.4}$$

on the set D of cadlag functions on the unit interval $[0,1]$. Here $((A_i^n, B_i^n, I_i^n)_i, C^n), Y_k^j$, $i, j, k \in \mathbb{N}$ are all independent. Y_i^j, A_i^n, C^n have values in the set D. The random variables B_i^n take values in $D_\uparrow$, the set of all maps from the unit interval to itself and piecewise increasing.

Under some assumptions they showed the existence of solutions of (1.4) via the weighted branching process, and Y^n converges in distribution to Y satisfying

$$Y \stackrel{\mathcal{D}}{=} \sum_{i \in \mathbb{N}} A_i Y_i \circ B_i + C$$

The contraction method [40] invented for the analysis of Quicksort, proved to be very successful for other algorithms [38, 34]. The contraction method is a general method to derive convergence in distribution of recursive structures. This method was pioneered by Roesler [39, 40] and later by Rachev and Rueschendorf [36] and Neininger [34]. This method was explained in the context of several divide and conquer algorithms in [38]. Knof [24] studied the finite dimensional distributions of D-valued processes Y^n by the contraction method. He introduced a suitable complete metric space and showed convergence of all finite dimensional distributions. His results include Quicksort. Gruebel and Roesler [18] used a nice version of the Quickselect processes and showed the convergence in other topology to a limiting process Y which is a fixed point of the map K [35].

1.2 Thesis outline

In view of the above brief overview, the thesis is organized as follows:

Chapter 2 provides general well known results on the space $D = D[0,1]$ of cadlag functions on the unit interval $[0,1]$, following Billingsley [4].

In Chapter 3 we introduce the weighted branching process $(\mathbb{V}, (T, C), (G, *), (H, \otimes))$ which was first introduced by Roesler [40] and plays a crucial role in our work. After that, we give two extreme examples, first a pure branching process and a second one with a pure multiplicative structure.

In Chapter 4 we embed the Quicksort process into a weighted branching process and discuss some properties of the limiting distribution.

In Chapter 5 we introduce in detail the Partial Quicksort algorithm and our version Quicksort on the fly. We show the distribution of the number $X(S, l)$ of comparisons to sort the l smallest elements out of the list S depends only on $|S|$ and l. This provides recursive equations for $X(|S|, l)$, which are essential for our work.

In Chapter 6 we present the two main results in this thesis. We suggest a normalization of the running time of Partial Quicksort using l as time index. Further we derive asymptotic formulas for the limiting process as an application of the weighted branching process. Finally we choose a very specific version with nice additional properties to ensure almost everywhere convergence of random variables. That implies the desired distributional convergence of processes (1.2).

In the Appendix the reader will see some basic facts about stochastic processes and their convergence.

In the paper we defined Y_n slightly different than in (1.1). The reasons are only notational ones. For example we divided by $n+1$ instead of n to assure a divisor unequal 0 without saying this all the time. In probability the space D of right continuous functions with left hand limits is (for good reasons) preferred to the space H of left continuous functions with existing right limits. In our problem the use of H would be more natural by the example in mind. Both spaces are equivalent in the sense of a bijection preserving the structure (The functions here have only discontinuities of the first kind). But these are only technical considerations not inflicting the semantic statements.

Chapter 2

Metric spaces

Metrics generalize the notion of distance on the real line, which is a useful tool for more general applications. We will introduce in this chapter an important tool in the theory of metric spaces, called Banach's fixed point theorem. Following we will construct a special metric space with respective properties which we need in our work.

Definition 2.1. (metric space) Let E be a non-empty set and suppose there is a function $d : E \times E \to [0, \infty)$, and for all $x, y, z \in E$ the following three properties holds:

(i) (positivity) $d(x, y) \geq 0$
(ii) (Non- degenerated) $d(x, y) = 0 \Leftrightarrow x = y$
(iii) (symmetry) $d(x, y) = d(y, x)$
(iv) (Triangle inequality) $d(x, y) \leq d(x, z) + d(z, y)$.

Then the tuple (E, d) is called metric space and d is called metric.

Definition 2.2. (Cauchy sequences)
A sequence $(x_n), n \in \mathbb{N}$ in a metric space (E, d) is called a Cauchy sequence, if for all $\epsilon > 0$, there exist an $n_0 \in \mathbb{N}$ such that

$$\forall n, m \geq n_0, \quad d(x_n, x_m) < \epsilon.$$

Definition 2.3. (complete metric space)
A metric space (E, d) is said to be complete if every Cauchy sequence in E converges in E.
In many fields, stability is a fundamental concept that can be described in terms of fixed point. Before we begin the special theory, we introduce a general view on the fixed point problem.

Definition 2.4. (fixed point)
A fixed point of a function $f : E \to E$ is an element $x \in E$ such that $f(x) = x$.

2.1 Banach's contraction principle

In our work we are interested in the most important result in fixed point theory. We state and prove Banach's contraction principle, also called the contraction mapping theorem. It is due to Banach and was first appeared in his Ph.D. thesis (1920, published in 1922)[6]. It states that every strict contraction on a complete metric space has a unique fixed point. This is an important tool in the theory of metric spaces. For national notations we denote the n-th iterates of $f : E \to E$ by $f^n, n \in \mathbb{N}$, the n times composition of f.

Definition 2.5. Let (X, d) be a metric space. A map $f : X \to X$ is called Lipschitz map (contraction) if there exists a real positive constant α such that

$$d\left(f(x), f(y)\right) \leq \alpha \cdot d\left(x, y\right), \forall x, y \in X.$$

The constant α is called Lipschitz constant for f. If $\alpha \leq 1$, f is called a contraction and α is said to be a contraction constant of f. We use strict contraction if $\alpha < 1$. If $\alpha > 1$, we say f is non expansive.

Theorem 2.6. *(Banach's contraction principle) Let (E, d) be a complete metric space and let $f : E \to E$ be a strict contraction. Then f has a unique fixed point z, and for each $z \in E$, we have*

$$\lim_{n\to\infty} f^n(x) = z$$

Proof. First we show the existence. Let $f : E \to E$ be a strict contraction with a contraction constant α. For any fixed $x_0 \in E$, define the sequence x_k by settings

$$x_{k+1} = f(x_k), \quad k \in \mathbb{N}$$

$$x_1 = f(x_0), x_2 = f(x_1) = f^2(x_0)\ , ..., \ x_k = f(x_{k-1}) = f^k(x_0).$$

From the fact that

$$d(f(x), f(y)) \leq \alpha \cdot d(x, y),$$

conclude by induction for every integer $j \geq 1$

$$d(x_{j+1}, x_j) \leq \alpha^j \cdot d(x_1, x_0) \tag{2.1}$$

For $1 \leq m \leq n$, the triangle inequality gives

$$\begin{aligned}
d(x_n, x_m) &= d(f^n(x_0), f^m(x_0)) \\
&\leq \alpha^m d(f^{n-m}(x_0), x_0) \\
&\leq \alpha^m [d(f^{n-m}(x_0), d(f^{n-m-1}(x_0)) + d(f^{n-m-1}(x_0), d(f^{n-m-2}(x_0)) \\
&+ \cdots + d(f(x_0), x_0)] \\
&\leq \alpha^m (\sum_{j=0}^{n-m-1} \alpha^j) d(x_1, x_0) \\
&\leq \alpha^m (\sum_{j=0}^{\infty} \alpha^j) d(x_1, x_0) \\
&\leq \frac{\alpha^m}{1-\alpha} d(x_1, x_0).
\end{aligned}$$

Since $\lim_{m\to\infty} \alpha^m = 0$ implies $d(x_n, x_m) \to 0$, i.e. $(x_n)_{n\in\mathbb{N}}$ is Cauchy in the metric space (E, d). Since E is complete this sequence converges to a point $z \in E$, $\lim_{n\to\infty} f^n(x) = z$. z is a fixed point since since f is continuous,

$$z = \lim_{n\to\infty} f^{n+1}(x) = \lim_{n\to\infty} f\left(f^n(x)\right) = f(z),$$

For the uniquenesses, let $z_1, z_2 \in E$ be both fixed points of f, i.e. $z_1 = f(z_1)$ and $z_2 = f(z_2)$. then from equation (2.1)

$$d(z_1, z_2) = d(f(z_1), f(z_2)) \leq \alpha \cdot d(z_1, z_2),$$

Since $0 \leq \alpha < 1$ implies $d(z_1, z_2) = 0$, and therefore we must have that $z_1 = z_2$, f has at most one fixed point. □

Corollary 2.7. *Let (E, d) be a complete metric space and let $f : E \to E$ be a strict contraction with a contraction constant α and fixed point z. Then for any $x_0 \in E$ we have the estimates*

$$d(x_m, z) \leq \frac{\alpha^m}{1-\alpha} d(x_0, f(x_0)), \tag{2.2}$$

$$d(x_m, z) \leq \alpha \cdot d(x_{m-1}, z), \tag{2.3}$$

$$d(x_m, z) \leq \frac{\alpha}{1-\alpha} d(x_{m-1}, x_m). \tag{2.4}$$

Proof. From the Theorem 2.6, and $1 \leq m \leq n$ provides

$$d(x_n, x_m) \leq \frac{\alpha^m}{1-\alpha} d(x_1, x_0) = \frac{\alpha^m}{1-\alpha} d(x_1, f(x_0))$$

This implies (2.2) since

$$\lim_{n\to\infty} d(x_m, x_n) = d(x_m, z).$$

Since z is a fixed point, we obtain (2.3) by

$$d(x_m, z) = d(f(x_{m-1}), f(z)) \leq \alpha \cdot d(x_{m-1}, z),$$

$m \geq 1$. By (2.3) the triangle inequality provides

$$d(x_m, z) \leq \alpha(d(x_{m-1}, x_m) + d(x_m, z)),$$

and therefore the equation (2.4) is obtained. □

2.2 The space $D\,[0,1]$

In this section we introduce the space $D = D[0,1]$ of cadlag functions on $[0,1]$ which are important in the study of stochastic processes. We introduce a metric and topology on D under which the space becomes a separable metric space. We will give briefly some properties of that space. For more details we refer to [4].

Definition 2.8. (cadlag functions) Let $D = D[0,1]$ be the space of real functions f on $[0,1]$ that are right-continuous and have left-hand limits:
(i) For $0 \leq t \leq 1$, $\quad f(t+) = \lim_{s\downarrow t} f(s)$ exists and $f(t+) = f(t)$.
(ii) For $0 \leq t \leq 1$, $\quad f(t-) = \lim_{s\uparrow t} f(s)$ exists.

Functions having these two properties are called cadlag functions.

For $f \in D[0,1]$ and $T_0 \subseteq [0,1]$, let

$$w_f(T_0) := \sup\left\{|f(t) - f(s)| : t, s \in T_0\right\},$$

and for $\delta > 0$

$$w_f(\delta) := \sup_{0\leq t\leq 1-\delta} w_f([t, t+\delta]).$$

A continuous function is uniformly continuous on a compact interval. The following lemma provides a similar property for the cadlag functions on $D[0,1]$.

Proposition 2.9. *For each $f \in D$ and for each $\epsilon > 0$ there exist points $t_0, t_1, \cdots t_m$ such that*

$$0 = t_0 < t_1 < \cdots t_m = 1$$

and

$$w_f([t_{i-1}, t_i)) < \epsilon, \quad i = 1, 2, \cdots, m.$$

From the above lemma we note that for each given real $r > 0$, there can be at most finitely many points $t \in [0,1]$ such that $|f(t) - f(t-)| > r$. Therefore f has at most countably many discontinuities and it follow also that f is bounded [4], i.e.

$$\sup_t |f(t)| < \infty.$$

Proposition 2.10. *Let* $f : [0,1] \to \mathbb{R}^+$ *be The map* $f(x) = x \ln x$ *and let* a *be a constant,* $0 \leq a \leq \frac{1}{e}$*. Then*

$$w_f(a) \leq 4a \left|\ln a\right|.$$

Proof. Let $|x - y| \leq a$ for $x, y \in [0,1]$ and $0 \leq a \leq \frac{1}{e}$. We consider the different cases. Case 1. If $x, y \geq a$. then using Taylor expansion of the function f is $f(y) - f(x) = (y - x)\acute{f}(\xi)$ for some value Ξ between x and y. We obtain

$$|f(y) - f(x)| \leq |y - x| \sup_{z \geq a} \left|\acute{f}(z)\right| \leq a(1 + |\ln a|) \leq 2a \left|\ln a\right|.$$

Case 2. If $x, y \leq a$. Then

$$|f(y) - f(x)| \leq 2 \sup_{z \leq a} |f(z)| \leq 2a \left|\ln a\right|.$$

Case 3. If $x \vee y \leq a$ and $x \wedge y > a$. Without loss of generality let $x \leq a$ and $y \geq a$. Then

$$|f(y) - f(x)| \leq |f(y) - f(a)| + |f(a) - f(x)| \leq 4a \left|\ln a\right|.$$

□

Definition 2.11. For $0 < \delta < 1$ and $f \in D$ define

$$\tilde{w}_f(\delta) = \inf_{\{t_i\}} \max_{1 \leq i \leq r} w_f \left[t_{i-1}, t_i\right),$$

where the infimum is over all finite sets $t_i \in [0,1]$ and the set $\{t_i\}$ satisfy that $0 = t_0 < t_1 < \cdots < t_r = 1$ and $t_i - t_{i-1} > \delta, \quad \forall i = 1, 2, \cdots r$. Notice, the Lemma 2.9 is just equivalent to assertion that

$$\lim_{\delta \to 0} \tilde{w}_f(\delta) = 0 \quad \text{for all} \quad f \in D.$$

We will introduce a norm ρ on D to become a Banach space. For all $f, g \in D\left[0,1\right]$ define a map $\rho : D \times D \to [0, \infty)$ by

$$\rho(f, g) := \sup_{t \in [0,1]} \left|f(t) - g(t)\right|. \tag{2.5}$$

Lemma 2.1. *Let* ρ *be defined as in* (2.5)*. Then* $(D\left[0,1\right], \rho)$ *is a complete metric space.*

Proof. We show the properties of the metric. ρ is well defined. Since for all $f, g \in D$ it is trivial to show $\rho(f,g) \geq 0, \rho(f,g) = 0 \Leftrightarrow f = g$ and $\rho(f,g) = \rho(g,h)$, we show the triangle inequality. For $f, g, h \in D$ and for all $t \in [0,1]$ then

$$\begin{aligned} \rho(f,g) &= \sup_{t\in[0,1]} |f(t) - g(t)| \\ &\leq \sup_{t\in[0,1]} (|f(t) - h(t)| + |h(t) - g(t)|) \\ &\leq \sup_{t\in[0,1]} |f(t) - h(t)| + \sup_{t\in[0,1]} |h(t) - g(t)| \\ &= \rho(f,h) + \rho(h,g). \end{aligned}$$

Therefore ρ is a metric on $D\,[0,1]$. To prove the completeness, let $(f_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in $D\,[0,1]$. Let $\epsilon_n > 0, n \in \mathbb{N}$ be decreasing to 0 and $\sum_{n\in\mathbb{N}} \epsilon_n < \infty$.

Define

$$n_i := \inf\left\{\bar{n} : \forall n > m \geq \bar{n},\ \rho(f_m, f_n) < \epsilon_i\right\}.$$

To show that $n_i, i \in \mathbb{N}$ is increasing, consider

$$A_i := \left\{\bar{n} : \forall n > m \geq \bar{n},\ \rho(f_m, f_n) < \epsilon_i\right\},$$

and notice $n_i = \inf A_i$. Since for $i \leq j$ we have $\epsilon_i \geq \epsilon_j$, it follows $A_i \supseteq A_j$. This implies $n_i \leq n_j$ and therefore $n_i, i \in \mathbb{N}$ is increasing.

The sequence $(f_{n_i})_{i\in\mathbb{N}}$ is a Cauchy sequence in (D, ρ) Since for $i_0 \leq i \leq j$

$$\begin{aligned} \rho(f_{n_j}, f_{n_i}) &\leq \sum_{k=i}^{j-1} \rho(f_{n_{k+1}}, f_{n_k}) \\ &\leq \sum_{k=i}^{j-1} \epsilon_k \leq \sum_{k\geq i_0} \epsilon_k \to 0 \quad \text{as} \quad i_0 \to \infty. \end{aligned}$$

For $n_{i_0} \leq m \leq n$ and i, j such that $n_i \leq m \leq n_{i+1}$ and $n_j \leq n \leq n_{j+1}$. The sequence $(f_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in (D, ρ) since

$$\rho(f_m, f_n) \leq \rho(f_m, f_{n_i}) + \rho(f_{n_i}, f_{n_j}) + \rho(f_{n_j}, f_n).$$

□

In our work we use the supremum norm $\|f\|_\infty = \sup_t |f|$ and sometimes we use $\|f\|_\infty$ instead of using $\rho(f, 0)$.

2.3 Skorodhod topology

The Skorodhod space provides a natural and convenient formalism for describing trajectories of stochastic processes admitting jumps. The Theorem 2.1 introduced the Banach space (D, ρ) but it is easy to see that it is non-separable, what is always disadvantageous in probability theory. To overcome this inconvenience, A.V. Skorodhod introduced a metric and topology under which the space becomes a separable metric space.

Definition 2.12. Let Λ denote the space of strictly increasing and continuous mappings $\lambda : [0,1] \to [0,1]$ with $\lambda(0) = 0$ and $\lambda(1) = 1$. Let $\mathcal{I} \in \Lambda$, $\mathcal{I} : [0,1] \to [0,1]$ be the identity map. The Skorodhod metric on $D[0,1]$ is defined by

$$d(f,g) := \inf \{\epsilon > 0 : \exists \lambda \in \Lambda \quad \text{with} \quad \rho(\lambda, \mathcal{I}) < \epsilon \quad \textit{and} \quad \rho(f, g \circ \lambda) < \epsilon\}, \tag{2.6}$$

for all $f, g \in D$.

This metric generates the Skorodhod topology on D, see [1]. The $\sigma-$filed $\sigma(D)$ is generated by the Skorodhod topology on D. The $\sigma-$field $\sigma(D)$ is isomorphic to the product $\sigma-$field $\mathbb{R}^T$ intersected with D, where T is a dense subset of $[0,1]$ containing 1. We refer to books by Billingsley [4] and Aldous [1] for exhaustive discussion of limit theorem results.

Lemma 2.2. *A sequence $(f_n)_{n\in\mathbb{N}} \subset D[0,1]$ converges in the Skorodhod topology to a limit $f \in D[0,1]$ if and only if there exist functions $\lambda_n \in \Lambda$ such that*

$$\lim_{n\to\infty} \sup_t \rho(\lambda_n, \mathcal{I}) = 0,$$

and

$$\lim_{n\to\infty} \rho(f_n \circ \lambda_n, f) = 0.$$

Remark 2.13. From the definitions (2.5) and (2.6), it is obvious that $d(f,g) \geq \rho(f,g)$. By Lemma 2.2 if f_n converges uniformly to f then f_n converges in the Skorodhod topology to f. Since

$$|f_n(t) - f(t)| \leq |f_n(t) - f(\lambda_n(t))| + |f(\lambda_n(t)) - f(t)|, \quad \lambda_n \in \Lambda. \tag{2.7}$$

The Skorodhod convergence does imply that $f_n(t) \to f(t)$ holds for continuity points t of f and hence for all but countably many t. From (2.7), if f is uniformly continuous on $[0,1]$, then

$$\|f_n - f\| \leq \|f_n - f\lambda_n\| + w_f(\|\lambda_n - \mathcal{I}\|).$$

Therefore the Skorodhod convergence implies uniform convergence. The Skorodhod topology is finer than the uniform topology.

The finite-dimensional sets play an important role in D. For $0 \leq t_1 < \cdots < t_k \leq 1$ and $f \in D$ define the projection

$$\pi_{t_1 \cdots t_k} : D \to \mathbb{R}^k$$

by

$$\pi_{t_1 \cdots t_k}(f) = (f(t_1), \cdots f(t_k)) \, .$$

It is obviously by the remark 2.13 that the zero and identity functions in Λ, called 0 and 1 respectively are elements in D and π_0 and π_1 are continuous. If f is not continuous at t, then f_n converges in the Skorodhod topology to f but $f_n(t)$ does not converge to $f(t)$. And therefore if $t \in (0,1)$, then π_t is continuous in the Skorodhod topology at f if and only if f is continuous at t. We state the next two lemmas without proofs and for more details we refer to [4].

Lemma 2.3. *1. The projections π_0 and π_1 are continuous and for $t \in (0,1)$, the projection π_t is continuous in the Skorodhod topology at f if and only if f is continuous at t.*
2. Each π_t is measurable with respect to $\sigma(D) - \mathcal{B}(\mathbb{R})$, and each $\pi_{t_1 \cdots t_d}$ is measurable with respect to $\sigma(D) - \mathcal{B}(\mathbb{R}^d)$.

Lemma 2.4. *Let $(\Omega, \mathcal{A}, P)$ be a probability space and $X : \Omega \to D$. Then X is a random variable with values in $(D, \sigma(D))$ if and only if for all $t \in [0,1]$ the projections.*

$$X_t : \omega \to \mathbb{R}, \quad \omega \to X(\omega)(t)$$

is a random variable on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Let $D_\uparrow$ be the subset of all functions $f : [0,1] \to [0,1]$ in D such that there exists $0 = t_0 < t_1 < \cdots < t_r = 1$ satisfying f is increasing on the interval $[t_{1-1}, t_i)$ for $i = 1, 2, \cdots, r$. In the following we consider the composition of random variables and give some results needed in our work.

Lemma 2.5. *Let X be a random variable with values in D and let B be a random variable with values in $D_\uparrow$. Then $X \circ B$ is a random variable with values in D.*

Proof. Let $X : \Omega \to D$, and $B : \Omega \to D_\uparrow$. Since for all $\omega \in \Omega$, $B(\omega) \in D_\uparrow$, then the function $(X \circ B)(\omega) \in D$. It is sufficient by Lemma 2.4 to show $X \circ B(t) : \Omega \to \mathbb{R}$ is measurable for all $t \in [0,1]$.
For all $t \in [0,1]$ define $B_j : \Omega \to \mathbb{R}, j \in \mathbb{N}$ by

$$B_j(\omega) := \frac{\lceil B(\omega)(t) \cdot j \rceil}{j}$$

we will approximate the $X \circ B$ by the random variables $X \circ B_j$ using the discretization on the values $0, \frac{1}{j}, \frac{2}{j} \cdots, 1 - \frac{1}{j}$. $B_j(t)$ is a measurable for all $\omega \in \Omega$ since

$$(B_j(t))^{-1}([a,1)) = (B(t))^{-1}\left([\frac{\lceil a \cdot j \rceil}{j}, 1)\right) \in \mathcal{A},$$

for all $j \in \mathbb{N}$ and $t \in [0,1]$. By the definition of B_j, $B_j(\omega) \geq B(\omega)(t)$ and $\lim_j B_j(\omega) = B(\omega)(t)$. Furthermore

$$X \circ B(\omega)(t) = X(\lim_j B_j(\omega)(t)) = \lim_j X(B_j(\omega)(t)) = \lim_j (X \circ B_j(\omega)(t))$$

Then $\lim_{j\to\infty}(X \circ B_j(t))$ is measurable since $X \circ B_j(t)$ is measurable. This implies that $X \circ B(t)$ is measurable for all $t \in [0,1)$ with respect to $\mathcal{B}(\mathbb{R})$. □

2.4 The metric $\|\cdot\|_{p,\tilde{D}}$

Let

$$\mathcal{F}(D) := \{X : \Omega \to D : X \text{ is measurable}\}, \tag{2.8}$$

be the space of all measurable functions X with values in D. For $1 \leq p < \infty$, define the subspace

$$\mathcal{F}_p(D) := \left\{X \in \mathcal{F}(D) : \|\|X\|_\infty\|_p < \infty\right\}.$$

Here $\|\cdot\|_p$ denotes the usual L_p-norm for random variables and $\|\cdot\|_\infty$ refers to the supremum norm.
For $1 \leq p < \infty$, define

$$\|\cdot\|_{p,D} : \mathcal{F}_p(D) \to \mathbb{R}$$

by

$$\|X\|_{p,D} := \|\|X\|_\infty\|_p. \tag{2.9}$$

Proposition 2.14. *For $1 \leq p < \infty$ the map $\|\cdot\|_{p,D}$ defined above is a pseudo-metric on $\mathcal{F}_p(D)$.*

Definition 2.15. For all $X, Y \in \mathcal{F}(D)$ and $1 \leq p < \infty$, define the binary relation $\sim$ by

$$X \sim Y :\Leftrightarrow P(X \neq Y) = 0.$$

Proposition 2.16. *The relation $\sim$ defined above is an equivalence relation on $\mathcal{F}(D)$.*

For the equivalence relation $\sim$ on $\mathcal{F}(D)$, define

$$[f]_p := \{g \in \mathcal{F}_p(D) : f \sim g\},$$

the equivalence class of an element $f \in \mathcal{F}(D)$. Let

$$\tilde{\mathcal{F}}_p(D) := \left\{[f]_p : f \in \mathcal{F}_p(D)\right\},$$

be the space of equivalence classes $[f]_p$.

For all $1 \leq p < \infty, t \in [0,1]$ and $f, g \in \tilde{\mathcal{F}}_p(D)$ define

$$\tilde{d}_p : \tilde{\mathcal{F}}_p(D) \times \tilde{\mathcal{F}}_p(D)$$

by

$$\tilde{d}_p([f], [g]) = \|\rho(f,g)\|_{p,D}.$$

Proposition 2.17. *For all $1 \leq p < \infty$, $\tilde{d}_p$ defined above is well defined.*

Proof. Let $f, g \in \mathcal{F}_p(D)$ and $f_1 \in [f], g_1 \in [g]$. We will show $\|\rho(f_1,g_1)\|_{p,D} = \|\rho(f,g)\|_{p,D}$. By the triangle inequality

$$\|\rho(f_1,g_1)\|_{p,D} \leq \|\rho(f_1,f)\|_{p,D} + \|\rho(f,g)\|_{p,D} + \|\rho(g,g_1)\|_{p,D}$$

$$\begin{aligned}
\|\rho(f_1,f)\|_{p,D} &= \left(E\left(\rho^p(f_1,f)(\mathbb{1}_{f=f_1} + \mathbb{1}_{f\neq f_1})\right)^{\frac{1}{p}}\right) \\
&= \left(E\left(\rho^p(f_1,f)\mathbb{1}_{f=f_1}\right)\right)^{\frac{1}{p}} + \left(E\left(\rho^p(f_1,f)\mathbb{1}_{f\neq f_1}\right)\right)^{\frac{1}{p}} \\
&= 0.
\end{aligned}$$

In the last equality the first term is zero since ρ is a metric, and the second term is zero since $P(f_1 \neq f) = 0$. Analogue $\|\rho(g,g_1)\|_{p,D} = 0$ and therefore we have

$$\|\rho(f_1,g_1)\|_{p,D} \leq \|\rho(f,g)\|_{p,D}. \tag{2.10}$$

On the other hand

$$\begin{aligned}
\|\rho(f,g)\|_{p,D} &\leq \|\rho(f,f_1)\|_{p,D} + \|\rho(f_1,g_1)\|_{p,D} + \|\rho(g_1,g)\|_{p,D} \\
&\leq \|\rho(f_1,g_1)\|_{p,D}. \qquad (2.11)
\end{aligned}$$

Combining equations (2.10) and (2.11), then $\|\rho(f_1,g_1)\|_{p,D} = \|\rho(f,g)\|_{p,D}$. And therefore $\|\rho(f_1,g_1)\|_{p,D}$ is independent of the choice f_1, g_1. □

Proposition 2.18. $\left(\tilde{\mathcal{F}}_p(D), \tilde{d}_p\right)$, $1 \leq p < \infty$, *is a Banach space with the usual addition and multiplication*

$$[f] + [g] := [f+g],$$

$$c \cdot [f] := [c \cdot f],$$

$$\|[f]\|_{p,D} := \|f\|_{p,D},$$

for $f, g \in \mathcal{F}_p(D), c \in \mathbb{R}$.

Chapter 3

Weighted branching process

The purpose of this chapter is to introduce the weighted branching process (WBP), which was first introduced by Roesler [40], and may be viewed as a generalization of the classical Galton-Watson process. The importance of the weighted branching processes does not only come from their natural relevance in the general theory of branching processes which are very common in nature and real world, but these processes include various applications, especially data structures and the analysis of random algorithms in computer science. First we specify some notation for trees.

3.1 The Ulam-Harris tree

Consider the infinite Ulam-Harris tree

$$\mathbb{V} := \bigcup_{n\in\mathbb{N}_0} \mathbb{N}^n$$

be the infinite tree rooted at $\{\phi\}$ where $\mathbb{N} = \{1, 2, \dots\}$ denotes the set of positive integers and by convention $\mathbb{N}^0 := \{\phi\}$ contains the null sequence ϕ. Each $v = (v_1, v_2, \dots, v_n) \in \mathbb{V}$ is called a node or vertex which we also write as $v_1v_2\dots v_n$. The vertex v is uniquely connected to the root ϕ by the path

$$\phi \to v_1 \to v_1v_2 \to \cdots \to v_1 \dots v_n.$$

The length of v is denoted by $|v|$, thus $|v_1 \cdots v_n| = n$ and in particularly we use $|\phi| := 0$. For all $v \in \mathbb{V}$ and for every $k \in \mathbb{N}$ define

$$v_{|k} := \begin{cases} \phi & k = 0 \\ v_1 \cdots v_k & k < |v| \\ v & k \geq |v| \end{cases}$$

Further we use the notations for $w = w_1 \dots w_m$

$$vw := v_1v_2 \dots v_nw_1w_2 \dots w_m,$$

$$\phi v := v \quad \text{and} \quad v\phi := v$$

In our work we suppress if possible the root ϕ. We introduce the definition of an ordering on $\mathbb{V}$ which reflects the kinship of its vertices when interpreted as individuals of a genealogical tree. We now define some useful relations on $\mathbb{V}$.

Definition 3.1. (relation on $\mathbb{V}$)
For all $v, w \in \mathbb{V}$, the prefix order $\preceq$ on $\mathbb{V}$ is given by

$$v \preceq w :\Leftrightarrow \exists u \in \mathbb{V} : vu = w.$$

The node v is called an ancestor or progenitor of w and conversely w is called a descendant of v. If $u \in \mathbb{N}$, then v is also called a mother of w and, conversely, w is called a child or offspring of v.
We further define

$$v \prec w :\Leftrightarrow v \preceq w \wedge v \neq w.$$

In the context of branching tree, the relation $v < w$ may be interpreted as, the node v is strictly older than w in terms of generations. We extend this definition to subsets of $\mathbb{V}$. For all $L \subseteq \mathbb{V}$ and $v \in \mathbb{V}$, Define

$$v \preceq L :\Leftrightarrow \exists w \in \mathbb{V} : vw \in L,$$

$$v \prec L :\Leftrightarrow \exists w \in \mathbb{V} \backslash \{\phi\} : vw \in L$$

and

$$L \preceq v :\Leftrightarrow \exists w \in L \quad \text{and} \quad w \preceq v,$$

$$L \prec v :\Leftrightarrow \exists w \in L \quad \text{and} \quad w \prec v.$$

Remark 3.2. We use the m-ary tree $\{1, \dots m\}^* := \bigcup_{n \in \mathbb{N}_0} \{1, \dots m\}^n$. In the case that $\mathbb{V} := \bigcup_{n \in \mathbb{N}_0} \{1, 2\}^n$, the tree is called binary tree.

3.2 Weighted branching process

Let $(\Omega, \mathcal{A}, P)$ be a probability space, rich enough to carry all occurring random variables in our work. Let $(G, *)$ be a measurable semigroup ($* : G \times G \to G, (g, h) = g * h$ associative and measurable) with a grave $\Delta, (\forall g \in \Delta : \Delta * g = \Delta = g * \Delta\ )$ and a neutral element e ($\forall g \in G : e * g = g = g * e\ $).

The semigroup $(G, *)$ operates transitive and measurable on the measurable space H via $\otimes : G \times H \to H$ and $G \times H$ is endowed with the product σ-field. Let $T : \Omega \to G^{\mathbb{N}}$ be a random variable relative to the product space.
We use the notation $T = (T_1, T_2, \cdots)$, where $T_i : \Omega \to G$ is the i-th projection of T. Let $C : \Omega \to H$ be a random variable with values in a measurable semigroup H.

Let $(T^v, C^v), v \in \mathbb{V}$, be independent copies of (T, C) on the same probability space $(\Omega, \mathcal{A}, P)$. We call T_i^v, the weight attached to the edge (v, vi) connecting v and vi. C^v is called the weight of the vertex v. The interpretation of C^v is as a cost function on a vertex $v \in \mathbb{V}$.

Definition 3.3. (weighted branching process)
A tuple $(\mathbb{V}, (T, C), (G, *), (H, \otimes))$ as above is called a weighted branching process (WBP).

For a weighted branching process without costs we write also $(\mathbb{V}, T, (G, *))$. We shall use freely other trees such as m-ary trees $\{1, 2, \cdots m\}^*$ of all sequences in an appropriate sense. The interpretation of G is as maps from H to H. If H has additional structure then we might enlarge G to have the induced structure.

For example if H is a vector space or an ordered set, we may extended G to a vector space of maps or ordered sets via the natural extension.

Definition 3.4. (path weight)
Define recursively a family $L := (L_v)_{v \in \mathbb{V}}$ of random variables $L_v : \Omega \to G$ by

$$L_\phi := e, \; L_{vi} := L_v * T_i^v \quad \text{for all} \quad v \in \mathbb{V}, i \in \mathbb{N}.$$

We call L_v the path weight from the root ϕ to the node v.

Similarly, we define recursively for all $v \in \mathbb{V}$, the family of path weights from v to vw $L^v := (L_w^v)_{w \in \mathbb{V}}$ by

$$L_\phi^v := e, \quad L_{wi}^v := L_w^v * T_i^{vw} \quad \text{for all} \quad w \in \mathbb{V}, i \in \mathbb{N}. \tag{3.1}$$

The path weight L_v has the following product representation

$$\begin{aligned} L_v &= T_{v_1}^\phi * T_{v_2}^{v_1} * T_{v_3}^{v_1 v_2} * \cdots * T_{v_n}^{v_1 v_2 \ldots v_{n-1}} \\ &=: \prod_{k=0}^{n-1} T_{v_{k+1}}^{v|k} \end{aligned} \tag{3.2}$$

for $v = v_1 v_2 \cdots v_n$. Hence L_v is just the accumulated multiplicative weight along the path connecting the root ϕ with the node v. L_v forms the total weight of the branch starting from the root ϕ to node v accumulated under operation * of the edge weights.

An individual or a node v is called alive, if $L_v \neq \Delta$, otherwise the node is called dead. In particular all nodes with weight Δ are skipped in pictures. Figure 3.1 shows the first few individuals of a family tree. We draw only the living individuals.

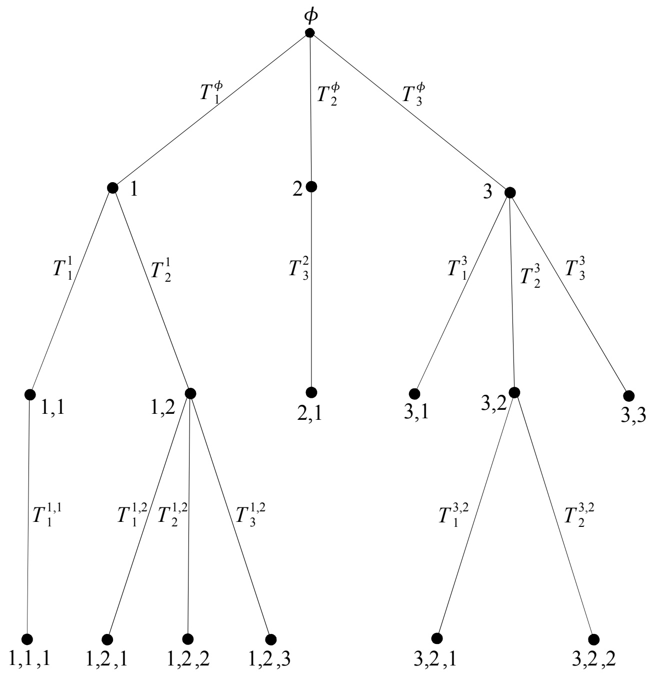


Figure 3.1: Family tree with weights

Define the total weight (cost) regarded up to the n-th generation by

$$R_n := \sum_{|v|<n} L_v \otimes C^v, \quad n \in \mathbb{N} \tag{3.3}$$

Because we deal only with positive values, everything will be well defined in our examples. We explain the forward and backward view via a weighted branching process on $\mathbb{R}^+$ for simplicity. The same argument will hold later for the the weighted branching process $(\mathbb{V}, (T, C), (D, *), (D_\uparrow, \otimes))$. Consider a weighted branching process $(\mathbb{V}, (T, C), (\mathbb{R}^+, \cdot), (\mathbb{R}, \cdot))$.

Define

$$R_n^v := \sum_{|w|<n} L_w^v \cdot C^{vw}$$

for $v \in \mathbb{V}, n \in \mathbb{N}$. Since H is a vector space and $\mathbb{R}$ is a lattice, we will embed G to maps H^H and use freely the induced structures $+$, $\cdot$ and $\vee$. Let $R_n^i, i \in \mathbb{N}$ are independent copies of R_n, then we have the following lemma

Lemma 3.1. *For each $v \in \mathbb{V}$, the sequence $(R_n^v), n \in \mathbb{N}$ is a WBP and satisfies the backward equation*

$$R_n^v = C^v + \sum_{i \in \mathbb{N}} T_i^v \otimes R_{n-1}^{vi}, \tag{3.4}$$

$n \in \mathbb{N}_0$.

Proof. See Roesler and Rueschendorf [38] □

The backward equation can be simply written as

$$R_n = C + \sum_{i \in \mathbb{N}} T_i \otimes R_{n-1}^i, \tag{3.5}$$

where the random variables (T, C) and $R_{n,i}, i \in \mathbb{N}$ are independent. The distribution of (T, C) is given. $R_{n,i}, i \in \mathbb{N}$ are independent copies of R_n and $R_0 = 0, R_1 = C_\phi$. Here R_n^i denotes the R_n random variable for the tree with root i.

Remark 3.5. We took in the backward equation (3.4) the almost everywhere version of the random variables. One of the major applications is equation (3.4) in distribution. This provides a recursive definition of the distribution of R_n, since (T^v, C^v) is independent of $(R_{n-1}^{vi})_{i \in \mathbb{N}}$ for every $v \in \mathbb{V}$.

3.3 Examples

Weighted branching processes have a branching structure and a multiplicative structure. We give two extreme examples, first example is a pure branching process and the second is a pure multiplicative structure.

3.3.1 Galton-Watson branching process

The simple Galton-Watson branching process (GWP) is a classical example. A Galton-Watson branching process [2] is defined as a process $(\bar{Z}_n)_{n \in \mathbb{N}_0}$ with values in $\mathbb{N}_0$ is introduced according to the recurrence formula

$$\tilde{Z}_1 = 1, \quad \tilde{Z}_n = \sum_{i=1}^{\tilde{Z}_{n-1}} \tilde{X}_{n,i},$$

where $\tilde{X}_{n,i}, n, i \in \mathbb{N}_0$ are independent and identically random variables with values in $\mathbb{N}_0$. The random number of offspring per individual always has the same distribution p, called offspring distribution and

$$p_k := P\left(\tilde{X}_{n,i} = k\right), \quad \forall k \in \mathbb{N}.$$

Let $G = \{0, 1\}$ with usual multiplication. Galton-Watson branching processes are special cases of Weighted branching processes without costs.

Let $T : \Omega \to \{0,1\}^{\mathbb{N}}$ and we allow only two possible cases for T_i, that T_i are either 0 or 1. The individual v has the path weight 1 if alive or 0 if dead . The process

$$Z_n = \sum_{|v|=n} L_v$$

gives the number of living individuals in the n-th generation, which is the total weight in the n-th generation. $(Z_n)_{n\in\mathbb{N}_0}$ forms an ordinary Galton-Watson branching process to Z_n to

$$p_k = P(\sum_{i=1}^{\infty} T_i = k).$$

The process $(Z_0 = 1, Z_1, \cdots)$ has the same distribution as the process $(\tilde{Z}_0 = 1, \tilde{Z}_1, \cdots)$ above.

3.3.2 Branching random walk

Another example is the multiplicative random walk. Let $Y_n, n \in \mathbb{N}$ be independent and identically distributed random variables with values in $\mathbb{R}^+$. Define a process $(W_n)_{n\in\mathbb{N}}$ on the multiplicative group $(\mathbb{R}^+, \cdot)$ by

$$W_0 = 1, \quad W_n = \prod_{i=1}^{n} Y_i.$$

Such process is called multiplicative random walk, ([44], [41]). The multiplicative random walk corresponds to a WBP with exactly one individual as successor. i.e.

$$P\left(\sum_i \mathbb{1}_{T_i\neq 0} = 1\right) = 1$$

. Without loss of generality, $T_2 = 0 = T_3 = \dots$ and T_1 has the same distribution as Y_1. So we could give an pointwise embedding via random variables

$$W_0 = 1 = L_\phi, \quad Y_1 = T_1^{\phi}, \quad Y_2 = T_1^{1}, \quad Y_3 = T_1^{11}, \quad \cdots$$

Then W_n is the product of n independent copies of T_1. The process $(W_n)_{n\in\mathbb{N}}$ is pointwise equal to a weighted branching process $(Z_n)_{\mathbb{N}_0}$.

3.4 The operator K

In this section, we introduce an operator K from probability measure on D into itself and give the connection to a weighted branching process $(\mathbb{V}, (T, C), (D, \cdot)), (D_{\uparrow}, \cdot)$.

Let $A = (A_i)_{i\in\mathbb{N}}, B = (B_i))_{i\in\mathbb{N}}$,C be random variables. The random variables $A_i, i \in \mathbb{N}$ and C take values in D and the random variables $B_i, i \in \mathbb{N}$ take values in $D_{\uparrow}$. The distribution of (A, B, C) is assumed to be fixed. Let $\mathcal{M}(D)$ be the set of all probability measures defined on D.

Define a map $K : \mathcal{M}(D) \to \mathcal{M}(D)$ by

$$K(\mu) \stackrel{\mathcal{D}}{=} \sum_{i\in\mathbb{N}} A_i \cdot X_i \circ B_i + C, \tag{3.6}$$

where the random variables $(A, B, C), X_i, i \in \mathbb{N}$ are independent. The random variables X_i have the same distribution μ. We always assume that the right hand side of equation (3.6) is well defined.

Lemma 3.2. *(Knof)*
Let $(X_i)_{i\in\mathbb{N}}$ be a family of random variables with values in D. Then

$$\left(\sum_{i\in\mathbb{N}} A_i(t) \cdot X_i\left(B_i(t)\right) + C(t)\right)_{t\in[0,1]}$$

are random variables with values in D.

Lemma 3.3. *(Knof)*
Let $(X_i)_{i\in\mathbb{N}}$ and $(Y_i)_{i\in\mathbb{N}}$ be two families of independent and identically distributed random variables with values in D. Let $\left((A_i)_{i\in\mathbb{N}}, (B_i)_{i\in\mathbb{N}}, C\right), (X_i)_{i\in\mathbb{N}}, (Y_i)_{i\in\mathbb{N}}$ be independent random variables. Suppose for all $i \in \mathbb{N}$, that X_i and Y_i have the same distribution. Then

$$\left(\sum_{i\in\mathbb{N}} A_i(t) \cdot X_i \circ B_i(t) + C(t)\right)_{t\in[0,1]} \stackrel{\mathcal{D}}{=} \left(\sum_{i\in\mathbb{N}} A_i(t) \cdot Y_i \circ B_i(t) + C(t)\right)_{t\in[0,1]}.$$

We are interested in fixed points of the operator K obtained by an iteration of K. Let the starting measure μ_0 be the point measure on the function $\underline{0} \in D$ identical 0. Define $\mu_i, i \in \mathbb{N}$ by

$$\begin{aligned}
\mu_1 := K(\mu_0) &= \mathcal{L}\left(\sum_{i\in\mathbb{N}} A_i \cdot X_i \circ B_i + C\right) \\
\mu_2 := K \circ K(\mu_0) &= \mathcal{L}\left(\sum_i A_i \left(\sum_j (A_{ij} \circ B_i \cdot X^1_{ij} \circ B_{ij} \circ B_i) + C_i \circ B_i\right) + C\right) \\
\mu_3 := K \circ K \circ K(\mu_0) &= \mathcal{L}\Bigg(\sum_i A_i \sum_j A_{ij} \circ B_i(\sum_k A_{ijk} \circ B_{ij} \circ B_i \cdot X^2_{ijk} \circ Bijk \circ B_{ij} \circ B_i \\
&+ \; C_{ij} \circ B_{ij} \circ B_i) + \sum_i A_i \cdot C_i \circ B_i + C\Bigg) \\
&. \\
&. \\
&. \\
\mu_n := K^n(\mu_0).
\end{aligned}$$

Here $\mathcal{L}(X)$ denotes the distribution of X. The random variables $X_0^{\cdot}$ are independent with distribution μ_0 and independent of all A, B, C random variables. It is obviously that

$$K(\mu_0) = \mathcal{L}(C).$$

In the next chapter we will give in some examples some connection between the operator K and well defined weighted branching processes.

Chapter 4

Quicksort

In the short history of computer science, the Quicksort algorithm is one of the fastest, best known and most widely studied sorting algorithms. It was invented and developed by Hoare in 1961 [20]. Most real-world sorting is done by Quicksort. It is the default sorting scheme in some operating systems, such as UNIX, and it is not difficult to implement. It has been studied extensively by Sedgewick [45], Knuth [27] and others. Quicksort sorts an array of distinct numbers based on a divide and conquer strategy. The basic idea of Quicksort algorithm to sort an array of n distinct elements is quite simple. The idea can be described as follows

- Choose one element of the array as the pivot.
- Divide the array of elements (except the pivot) into two nonempty subarrays.
  - All elements in the left partition are strictly less than the pivot.
  - All elements in the right partition are strictly greater than the pivot.
- Recall the procedure recursively to sort both subarrays.

What is the worst-case performance for these pivot selection mechanisms? In this work we will ignor all aspects of the Quicksort algorithm except the number of comparisons. Let X_n denote the random number of comparisons required by Quicksort to sort n elements and suppose that $X_0 = 0$. Then X_n satisfies the recurrence relation

$$X_n = \begin{cases} X_{U_n - 1} + \bar{X}_{n-U_n} + n - 1 & \text{for} \quad n \geq 2 \\ 0 & \text{for} \quad n = 0, 1 \end{cases} \tag{4.1}$$

The random variable U_n is uniformly distributed on $\{1, ..., n\}$ if we pick the pivot by random with uniform distribution. $X, \bar{X}$ have the same distribution and $U_n, X, \bar{X}$ are all independent. To understand above recurrence (4.1), we must consider it term by term. The term $n-1$ is needed since every splinting uses $n-1$ comparisons. $U_n - 1$ and $n - U_n$ represent the sizes of the left and right subarrays. The sorting of the subarrays are independent. In other words, the distribution $\mathcal{L}(X_n)$ of the random variable X_n satisfies the distributional recurrence relation $X_0 = 0, X_1 = 0, X_2 = 1$,

$$\mathcal{L}(X) = \mathcal{L}(X_{U_n-1} + \bar{X}_{n-U_n} + n - 1), \qquad n \geq 2. \tag{4.2}$$

4.1 Performance of Quicksort

For the comparison-based algorithms like Quicksort, we express running time in terms of the number of comparisons done. So, the running time of the Quicksort algorithm depends on the position of selected pivot. In this section we discuss three possible cases for the Quicksort: worst-case, best-case and the average-case.

4.1.1 The worst-case

The worst case scenario for the random standard Quicksort algorithm appears if we pick a pivot always as the largest or smallest value in the array. The case occurs when the two subsequences are as unbalanced as they can be one sequence has all the remaining elements and the other has none.

Proposition 4.1. *The maximal value of X_n is bounded below by $\frac{(n-1)^2}{2}$ and above by $\frac{n^2}{2}$.*

Proof. Let $a_n := ess\sup_{n\in\mathbb{N}} X_n$, the maximum value of X_n. We have to show the inequality

$$\frac{(n-1)^2}{2} \leq a_n \leq \frac{n^2}{2}$$

for $n \in \mathbb{N}$. We show the first inequality by an induction on n. The induction start $n = 1$ is obvious. We show only the induction step $n - 1$ to n.

$$\begin{aligned} a_n &= \sup_{1\leq u\leq n, u\in\mathbb{N}} (a_{n-1} + a_{n-u} + n - 1) \\ &\leq \sup_{1\leq u\leq n, u\in\mathbb{R}} (\frac{(u-1)^2}{2} + \frac{(n-u)^2}{2} + n - 1) \end{aligned}$$

The supremum is attained at the boundaries of the interval $[1, n]$. The last inequality provides $a_n \leq \frac{n^2}{2}$. For the second inequality we take $u = 1$ and argue

$$\begin{aligned} a_n &\geq a_0 + a_{n-1} + n - 1 \\ &\geq \frac{(n-2)^2}{2} + n - 1 \\ &\geq \frac{(n-2)^2}{2}. \end{aligned}$$

□

Remark 4.2. The worst case occurs when the partition produces one two subproblems, one has none elements and other has all the $n-1$ remaining elements. In that case the recursion depth is $n-1$ levels and the Quicksort runs in time $O(n^2)$. It is no better than simple quadratic time algorithms like straight insertion sort.

4.1.2 The best-case

The good case behavior of the Quicksort algorithm occurs when the pivot is the median. Therefore in each recursion step the partitioning produces exactly two subarrays of almost equal length each have size $\frac{n}{2}$. Let $b_n := ess \inf_{1\leq u\leq n, u\in\mathbb{N}} X_n$ refers to the minimal value of X_n. We show the next two propositions

Proposition 4.3. *The minimal value b_n of X_n is bounded below by $n\ln n$.*

Proof. We show the inequality $n\ln n \leq b_n$ for $n\in\mathbb{N}$ by an induction on n. For the induction step $n-1$ to n, notice $x \to x\ln x$ is convex.

$$\begin{aligned}
b_n &= \inf_{1\leq u\leq n, u\in\mathbb{N}} (b_{n-1} + b_{n-u} + n - 1) \\
&\geq \inf_{1\leq u\leq n, u\in\mathbb{R}} ((u-1)\ln(u-1) + (n-u)\ln(n-u) + n - 1) \\
&\geq \inf_{1\leq u\leq n, u\in\mathbb{R}} \left(\left[(n-1)\frac{u-1}{n-1}\ln(\frac{u-1}{n-1}) + \frac{n-u}{n-1}\ln(\frac{n-u}{n-1}) + \ln(n-1) \right] + 1 \right) \\
&\geq (n-1)\ln(n-1) + (n-1)(\ln(\frac{1}{2}) + 1) \\
&\geq (n-1)\ln(n-1).
\end{aligned}$$

□

Proposition 4.4. *The mimimal value b_n of X_n is bounded above by $2n\ln_2 n$. Moreover*

$$b_{2^m-1} \leq 2^m(m-1)$$

for $m\in\mathbb{N}$.

Proof. Define c_n, $n\in\mathbb{N}$ by $c_1 = 0$ and recursively

$$c_n := c_{\lceil \frac{n}{2} \rceil} + c_{\lfloor \frac{n}{2} \rfloor} + n - 2, \quad \text{for} \quad n \geq 2.$$

Consider

$$d_m := \frac{c_{2^m}}{2^m}, \quad m\in\mathbb{N}_0.$$

It is obvious that $d_0 = 0$ and

$$c_{2^m} = 2(c_{2^{m-1}} + 2^{m-1}) \quad \text{for} \quad m \geq 1.$$

Hence

$$\begin{aligned} d_m &= \frac{2}{2^m}(c_{2^{m-1}} + 2^{m-1}) \\ &= \frac{c_{2^{m-1}}}{2^{m-1}} + 1 \\ &= d_{m-1} + 1 \\ &\vdots \\ &= m - 1. \end{aligned}$$

Therefore $c_{2^m} = 2^m(m-1)$. Now we prove first the second statement of the proposition by an induction on m. The induction start $n = 1$ is obvious since $c_1 = 0$. We show only the induction step $n-1$ to n. Choose $u = \frac{\lfloor n \rfloor}{2}$, we have

$$\begin{aligned} b_{n-1} &= \inf_{1 \leq u \leq n-1, u \in \mathbb{N}} (b_{n-2} + b_{n-u-1} + n - 2) \\ &\leq c_{\lceil \frac{n}{2} \rceil} + c_{\lfloor \frac{n}{2} \rfloor} + n - 2 \\ &\leq c_n. \end{aligned}$$

Then the second statement follows easily. For the first statement, let $2^m \leq n < 2^{m+1}$. Then by the monotonicity

$$b_n \leq b_{2^{m+1}-1} \leq 2^{m+1} m \leq 2n \ln_2 n.$$

□

Remark 4.5. If the Quicksort always picks median as pivot at every step then the Quicksort runs in time $O(n \ln n)$. In the best case, this is excellent essentially as good as any comparison sorting algorithm can be.

4.1.3 The average-case

Since the best case is $O(n \ln n)$ and the worst case is $O(n^2)$, then the average case must be between the best case and the worst case. Certainly, on random data, the position of the pivot could be the first one, the second one, etc.

In practice, the implementations of Quicksort algorithm often chooses a pivot randomly each time. This method is seen to work excellently in the practical implementations of Quicksort. This choice reduces the chance that the worst-case ever occurs.
The other technique, which deterministically prevents the worst case from ever occurring, pick a pivot as the median of the array each time.

When we pick the pivot, we compare all other elements to it and therefore we perform $n-1$ comparisons to split the array. Depending on the pivot, we might split the array into one sub array of size 0 and one sub array of size $n-1$, or into a sub array of size 1 and other one of size $n-2$, and so on up to a sub array of size $n-1$ and one of size 0. There are n possible positions and each one is equally likely with probability $\frac{1}{n}$.

Therefore to determine the average running time for the Quicksort let $\pi_{n,j}$ denote the probability that the chosen pivot is the jth smallest element among n distinct elements and in the standard case, we assume that $\pi_{n,k} = P(U_n = k) = \frac{1}{n}$.

Let $a_n = E(X_n)$, the average number of comparisons to sort n distinct elements by Quicksort.

Lemma 4.1. *For $n \geq 1$, the average number of comparisons to sort n distinct elements made by randomized Quicksort is at most $2n \ln n$.*

Proof. By the recurrence (4.2) we have

$$E(X_n) = n - 1 + \sum_{j=1}^{n} P(U_n = j)(E(X_{j-1}) + E(X_{n-j}))$$

If we consider the standard case, when $P(U_n = k) = \frac{1}{n}$, then

$$a_n = n - 1 + \frac{1}{n} \sum_{j=1}^{n} (a_{j-1} + a_{n-j})$$

Since $\sum_{j=1}^{n} a_{j-1} = \sum_{j=1}^{n} a_{n-j}$ are the same,

$$a_n = n - 1 + \frac{2}{n} \sum_{j=1}^{n} a_{j-1}$$

Multiply both sides by n and subtract the same equation for $n-1$, then divide by $n(n+1)$ and obtain

$$\frac{a_n}{n+1} = \frac{a_{n-1}}{n} + \frac{2(n-1)}{n(n+1)}$$

Iterating this recursion provides

$$\begin{aligned}
\frac{a_n}{n+1} &= \frac{a_{n-1}}{n} + \frac{2(n-1)}{n(n+1)} = \sum_{j=1}^{n-1} \frac{2j}{(j+1)(j+2)} \\
&= \sum_{j=1}^{n-1} \frac{2((j+2)-2)}{(j+1)(j+2)} = \sum_{j=1}^{n-1} \frac{2}{j+1} - \sum_{j=1}^{n-1} \frac{4}{(j+1)(j+2)} \\
&= 2\sum_{j=2}^{n} \frac{1}{j} - 4\sum_{j=2}^{n} \frac{1}{(j)(j+1)} \\
&= 2\sum_{j=2}^{n} \frac{1}{j} - 4\sum_{j=2}^{n} \frac{1}{j} + 4\sum_{j=2}^{n} \frac{1}{j+1} \\
&= 2\sum_{j=1}^{n} \frac{1}{j} - 4\sum_{j=1}^{n} \frac{1}{j} + 4\sum_{j=1}^{n+1} \frac{1}{j} - 4 \\
&= 4H_{n+1} - 2H_n - 4 = 2H_{n+1} + \frac{2}{n+1} - 4.
\end{aligned}$$

where $H_n = \sum_{k=1}^{n} \frac{1}{k}$, $n \in \mathbb{N}$ denote the n-th harmonic number. The values of the sequence $(H_n - \ln n)$ decrease monotonically towards the limit:

$$\lim_{n\to\infty} (H_n - \ln n) = \gamma,$$

where $\gamma = 0.5772156649\ldots$, is called Euler constant. The corresponding asymptotic expansion is

$$H_n = \ln n + \gamma + \frac{1}{2}n^{-1} - \frac{1}{12}n^{-2} + \frac{1}{120}n^{-4} + O(n^{-6}),$$

see [27]. Therefore $\forall n \in \mathbb{N}$, a_n has the asymptotic expansion

$$\begin{aligned}
a_n &= 2(n+1)H_{n+1} + 2 - 4(n+1) = 2(n+1)H_n - 4n \\
&\approx 2(n+1)(\gamma + \ln n + O(n^{-1})) \\
&\approx 2n\ln n + n(2\gamma - 4) + 2\ln n + 2\gamma + 1 + O(n^{-1}).
\end{aligned}$$

□

4.2 How to choose a pivot?

In the preceding section we show that selecting a good pivot is important. A poor choice of a pivot could give a running time quadratic proportional to the number of elements squared. Therefore selecting a good pivot greatly improves the speed of the Quicksort algorithm. Many people just use the first element in the list as the pivot, however this causes the sort to perform very badly if the data is already sorted. There are several methods to avoid the worst case in practical solutions. Unix uses the median of the first the last and the element in the middle. Therefore, when choosing the pivot we need to be more careful. For example, when Quicksort is used in web services, it is possible for an attacker to intentionally exploit the worst case performance and choose data which will cause a slow running time or maximize the chance of running out of stack space. The choice of a good pivot greatly improves the speed of the Quicksort algorithm.

The simplest way is to choose an arbitrary element say the first for example as pivot, this does not avoid the worst case. Instead of using the first element, a much better method is called median of three Quicksort. In that method choose the pivot of each recursive stage as the median of a sample of three elements. Other method is to take tree samples, each sample contain 3 elements, take the median for each sample and choose the median of three medians as a pivot,this method called pseudomedian of 9 Quicksort.

To make sure to avoid any kind of presorting it is better to use the median element of the first, middle, and the last element as a pivot. To optimize the algorithm, for an array smaller than 7, the pivot is chosen as the middle key or sort with the standard Quicksort, for mid-sized arrays(for an array of size between 8 and 39) the pivot is chosen using the median-of-three Quicksort, and finally for larger arrays use the pseudomedian of 9 Quicksort. This helps some but unfortunately simple anomalies happens [27].

4.3 Normalization of X_n

Let X_n be the random variable defined as in (4.2) representing the number of comparisons in an array of size n. Consider the corresponding normalized random variable

$$Y_n := \frac{X_n - EX_n}{n}, \quad n \geq 2. \tag{4.3}$$

The equation (4.2) rewrites in terms of Y_n, and $Y_0 = 0 = Y_1$ implies the recursion

$$\mathscr{L}(Y_n) = \mathscr{L}\left(Y_{U_n-1}\frac{U_n - 1}{n} + \bar{Y}_{n-U_n}\frac{n - U_n}{n} + C_n(U_n)\right), \quad n \geq 2 \tag{4.4}$$

where for any fixed n, the random variables $Y_i, \bar{Y}_i, U_n,\ 1 \leq i \leq n$ are independent. The random variable U_n is uniformly distributed on $\{1, ..., n\}$. C_n is a function on $\{1, 2, \cdots, n\}$ defined by

$$C_n(i) = \frac{n-1}{n} + \frac{1}{n}\left(E(X_{i-1}) + E(X_{n-i}) - E(X_n)\right), \tag{4.5}$$

It is obvious that $E(Y_n) = 0$ and $E(C_n(U_n)) = 0$. In order to estimate $C_n(j)$ defined by (4.5), we need some explicit bounds on a_n by the harmonic numbers.

As $n \to \infty$, $\frac{U_n}{n}$ converges in distribution to some random variable U, which is uniformly distributed on the unit interval $[0,1]$. Furthermore the random variable $C_n(n \cdot \frac{U_n}{n})$ converges to $C(u)$, when $U_n \to U$ as $n \to \infty$,

$$C(u) := 2u \ln u + 2(1-u)\ln(1-u) + 1, \qquad u \in [0,1], \tag{4.6}$$

with $C(u) := 1$ for $u = 0, 1$.
By using martingales, Roesler [40, 43] and Régnier [37] used different methods showed that Y_n converges in distribution to some random variable Y. Moreover, by formally taking limits in (4.4), Y satisfying the distributional equation of the form

$$Y \stackrel{\mathcal{D}}{=} UY + (1-U)\bar{Y} + C(U) \tag{4.7}$$

where $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution, and U, Y,and $\bar{Y}$ are independent. Y and $\bar{Y}$ have the same distribution; and U is uniformly distributed on $[0,1]$. In the following we discuss the existence of some random variable Y satisfying (4.7) by a fixed point argument.

4.4 Existence of a fixed point

Let $\mathcal{M}$ be the set of distributions on $\mathbb{R}$. Define the set of distributions F on $\mathbb{R}$ with finite p-th moment by

$$\mathcal{M}_p = \left\{ F \in \mathcal{M} : \int |x|^p dF(x) < \infty \right\}, \qquad 1 \le p < \infty.$$

Let

$$\mathcal{M}_{p,0} = \left\{ F \in \mathcal{M} : \int x dF(x) = 0 \right\}.$$

Define a map $S : \mathcal{M}_{2,0} \to \mathcal{M}_{2,0}$ by

$$S(\nu) := \mathscr{L}(UY + (1-U)\bar{Y} + C(U))$$

where U is uniformly distributed on $[0,1]$, the random variables $U, Y, \bar{Y}$ are independent and Y, $\bar{Y}$ have the same distribution ν and $C : (\mathbb{0}, \mathbb{1}) \to \mathbb{R}$ is defined as in (4.6). Let

$$A_1 = U, A_2 = (1-U), A_3 = A_4 = \cdots, C := C(t),$$

Roesler [39] showed the following theorem:

Theorem 4.6. *(Roesler) (i) The Function S is a strict contraction with respect to the Wasserstein d_2-metric, and has a unique fixed point in $\mathcal{M}_{2,0}$*
(ii)Any sequence $\nu, S(\nu), S^2(\nu), \ldots$ where $\nu \in \mathbb{D}$ converges in the Wasserstein d_2-metric to a unique fixed point $\mathscr{L}(Y)$ of the operator

$$\nu = \mathscr{L}(Y) \to S(\nu) := \mathscr{L}(UY + (1-U)\bar{Y} + C(U))$$

(iii) The fixed point of S is the weak limit of the Y_n of (4.3).

By the equation (4.3), $E(Y_n) = 0$. U is uniformly distributed on $[0,1]$, U, Y,and $\bar{Y}$ are independent and Y and $\bar{Y}$ have the same distribution.

$$\begin{aligned} E(C(U)) &= 2E(U \ln U) + 2E\left((1-U)\ln(1-U)\right) + 1 \\ &= 2\int_0^1 x \ln x dx + 2\int_0^1 (1-x)\ln(1-x) dx + 1 \\ &= 2(-\frac{1}{4}) + 2(-\frac{1}{4}) + 1 = 0. \end{aligned}$$

$$\begin{aligned} E\left((C(U)^2\right) &= 1 + 8 \cdot E(U \ln U) + 8 \cdot E\left(U^2 (\ln U)^2\right) + 8 \cdot E\left(U(1-U)\ln U \cdot \ln(1-U)\right) \\ &= 1 + 8 \cdot \left(\int_0^1 x \ln x dx + \int_0^1 x^2 (\ln x)^2 dx + \int_0^1 x \ln x \cdot (1-x)\ln(1-x) dx\right) \\ &= 1 + 8 \cdot \left(-\frac{1}{4} + \frac{2}{27} + \frac{37 - 3\pi^2}{108}\right) \\ &= \frac{21 - 2\pi^2}{9} \approx 0.14231. \end{aligned}$$

Let us calculate the standard deviation of the random variables Y_n and X_n.

$$\begin{aligned} Var\,(Y) &= E\left(Y^2\right) = E\left((UY + (1-U)\bar{Y} + G(U))^2\right) \\ &= E\left(U^2 Y^2 + (1-U)^2 \bar{Y}^2 + (C(U))^2\right) \\ &= \frac{2}{3} E(Y^2) + E\left((C(U))^2\right), \end{aligned}$$

Then

$$\begin{aligned} Var\,(Y) &= 3 \cdot E\left((C(U))^2\right) \\ &= 3 \cdot \frac{21 - 2\pi^2}{9} = 7 - \frac{2}{3}\pi^2 \approx 0.4269. \end{aligned}$$

By the equation (4.3), and the Theorem 4.6

$$Var\,(X_n) = n^2 Var\,(Y_n) \approx n^2 Var\,(Y) \approx 0.4269 \cdot n^2,$$

and therefore $\sigma_{Y_n} \approx 0.652276$ and $\sigma_{X_n} \approx 0.652276 \cdot n$.

4.5 Quicksort as a weighted branching process

In the above section, we explained the idea of the Quicksort algorithm. We will now introduce the Quicksort process as weighted branching process. We use the same notations we discussed before.

We consider the binary tree $\mathbb{V} = \{1,2\}^*$. Let $U^v, v \in \mathbb{V}$ be independent real random variables with a uniform distribution on $[0,1]$. Let E be the set of the intervals $[a,b), 0 < a < b \leq 1$. Define the map

$$T^v = (T_0^v, T_1^v) : \Omega \times E \to E \times E$$

by

$$T^v([a,b)) := (T_1^v([a,b)), T_2^v([a,b))) = ([a, a + U^v(b-a)), [a + U^v(b-a), b)).$$

For all $v \in \mathbb{V}$ and $i = 0,1$ define a map $L : \mathbb{V} \to E$ by

$$L_{vi} = T_i(L_v) \quad \text{with} \quad L_\phi = [0,1).$$

L_v is interpreted as the weight of the path from the root ϕ to the vertex $v \in \mathbb{V}$.
For all $t \in [0,1)$ define the process $Z_n : \Omega \times [0,1] \to \mathbb{R}$ by

$$Z_n(t) := \sum_{|v| \leq n} \mathbb{1}_{t \in L_v} |L_v|,$$

where $Z_n(1) = \lim_{t \uparrow 1} Z_n(t)$. We define the limiting process $Z : \Omega \times [0,1] \to \mathbb{R}$ by

$$Z(t) := \lim_{n \to \infty} Z_n(t) = \sum_{v \in \mathbb{V}} \mathbb{1}_{t \in L_v} |L_v|.$$

Let $G = \mathbb{R}$ be the multiplicative semigroup with neutral element $e = 1$ and the grave $\Delta = 0$. G operates on $H = \mathbb{R}$ transitive by usual multiplication.

Define the weights of the edges as follow

$$\begin{aligned} T_1^v &= U^v, \\ T_2^v &= 1 - U^v, \\ T_3^v &= T_4^v = \cdots = 0, \end{aligned}$$

the weights from the root ϕ to the node v. The path weight L_v from the root ϕ to the node v is recursively by

$$\begin{aligned} L_{vi} &= L_v T_i^v, \\ L_\phi &= 1. \end{aligned}$$

Define $C : [0,1] \to \mathbb{R}$ by

$$C(x) := 1 + 2x \ln x + 2(1-x)\ln(1-x) \quad \text{for all} \quad x \in [0,1]\,.$$

And define $C^v := C(U^v)$.

According to our settings, then the tuple $(\{1,2\}^*, (T^v, C^v)_{v\in\mathbb{V}}, (\mathbb{R},\cdot,\mathbb{R},\cdot))$ is a weighted branching process. Since H is an ordered vector space, we extend G with the interpretation of maps to the ordered vector space generated by the maps $x \to ax$, $a \in \mathbb{R}$.

The total weight cost up to $n-1$ generation is

$$R_n := \sum_{|v|<n} L_v C^v.$$

R_n is an L_2 martingale and converges in L_2 and almost everywhere to Q [39]. The distribution of Q is called the Quicksort distribution. Moreover the Quicksort distribution is uniquely characterized [39, 13] as the solution Q with $E(Q) = 0, Var(Q) < \infty$ of the stochastic fixed points equation

$$Q \stackrel{\mathcal{D}}{=} UQ_1 + (1-U)Q_2 + C(U). \tag{4.8}$$

Here $\stackrel{\mathcal{D}}{=}$ denote the equality in distribution. The random variables U, Q_1 and Q_2 are independent and Q_1, Q_2 have the same distribution as Q.

By the almost everywhere convergence of $R_n^v := \sum_{|w|<n} L_w^v C^{vw}$, the random variables

$$Q^v := \sum_{w\in\mathbb{V}} L_v C^{vw}, \quad v \in \mathbb{V}$$

exists and satisfy almost everywhere

$$Q^v = U^v Q^{v1} + (1-U^v)Q^{v2} + C(U^v).$$

4.6 Convergence of the discrete Quicksort distributions

Now the abstract embedding into a WBP with an additional parameter $n \in \mathbb{N}$. Let

$$H := \{h : \mathbb{N}_0 \to \mathbb{R}\},$$

and

$$G_2 := \{g : \mathbb{N}_0 \to \mathbb{N}_0\},$$

with $g(0) = 0$ and $g(n) < n$ for all $n \in \mathbb{N}$. Define G as the set $H \times G_1$. The semigroup structure $*$ is given by

$$(f_1, g_1) * (f_2, g_2) = f_1 \circ g_2 \cdot f_2 \cdot h \circ g_2 \circ g_1$$

And the operation $\otimes$ on H is given by

$$(f, g) \otimes h = f \cdot h \circ g.$$

Here $\circ$ denotes composition and $\cdot$ denotes the usual multiplication on $\mathbb{R}$. The interpretation of $(f, g) \in G$ is as a map on H with f is a multiplicative factor and g an index transformation. The operation $*$ corresponds to the convolution of maps on H. Since H is a vector space, we may enlarge G naturally to a vector space.

Consider a binary tree $\mathbb{V} = \{1, 2\}^*$. Let $U^v, v \in \mathbb{V}$ be independent random variables with a uniform distribution on $[0, 1]$. Let $I_n^v := \lceil nU^v \rceil$. Define the edge weights on the edges (v, v_1) and (v, v_2) by

$$\begin{aligned}
J_1^v(n) &:= I_n^v - 1 \\
J_2^v(n) &:= n - I_n^v \\
T_1^v(n) &:= (\frac{I_n^v - 1}{n}, J_1^v) \\
T_2^v(n) &:= (1 - \frac{I_n^v}{n}, J_2^v).
\end{aligned}$$

And define the vertex weight $C^v(n) := C_n(I_n^v)$. The total weight cost up to $m - 1$ generation is

$$R_m^v := \sum_{v \in \mathbb{V}_{<m}} L_w^v \otimes C^{vw}.$$

As $m \to \infty$, the random variable R_m^v converges almost every where and in L_2 to some limit R_∞^v. The random variable R_∞^v satisfying the fixed point equation

$$R_m^v := \sum_i T_i^v R_{m-1}^{iv} + C^v$$

Remark 4.7. The connection to the previous description is

$$Y_n \stackrel{\mathcal{D}}{=} R_\infty(n).$$

$R^v_\infty(n)$ converges for every $v \in \mathbb{V}$ in L_2 to the random variable Q^v, the Quicksort distribution [39].

4.7 Properties of the limiting distribution

In this section we introduce the basic facts about the limiting distribution of Quicksort. We give a short definition and we show some concepts for the convergence.

4.7.1 Bounds on characteristic function

Define the characteristic function of Y for all real t by $\phi(t) = E\left(e^{itY}\right)$. No closed form is known for the moment generating function. Because of the fixed point properties according to the Theorem 4.6. In 1994 Eddy and Schervish [11] applied a successive substitution method to obtain a numerical approximation of the characteristic function $\phi(t)$, see [46].
Specifically by conditioning on U, an estimate of characteristic function of X was obtained from the fundamental relation

$$\phi(t) = \int_0^1 \phi(ut)\phi((1-u)t)e^{itG(u)}du$$

and thus the estimate

$$|\phi(t)| \leq \int_0^1 |\phi(ut)|\,|\phi((1-u)t)|\,du.$$

Fill and Janson [13] proved that the characteristic function of $S(\nu)$ is bounded by $2\left|t^{-\frac{1}{2}}\right|$ for each Y and $\bar{Y}$ and therefore improved the exponent to show the following theorem on super polynomial decay of the characteristic function of the limit variable Y .

Theorem 4.8. *(Fill) For every real $r \geq 0$ and for all $t \in \mathbb{R}$ there exist a smallest constant $0 \leq c_r \leq \infty$ such that the characteristic function of the limit variable Y satisfies*

$$|\phi(t)| \leq c_r|t|^{-r}$$

If $0 < r < 1$, then

$$c_{2r} \leq \frac{[\Gamma(1-r)]^2}{\Gamma(2-2r)}c_r^2.$$

Some explicit bounds on c_r are given for some special cases when $c_r < \infty$ for r as large as possible. This mean the characteristic function $\phi(t)$ belongs to the class of infinitely differentiable functions with derivatives of all order, decrease more rapidly than any power. Roesler [39] showed that the moment generating function of Y is everywhere finite, i.e Y has finite moments of all orders and the characteristic function $\phi(t)$ is infitely differentiable. Using Theorem 4.8 leads to a rapid decrease of all derivatives, so for each real $r \geq 0$ and integer $k \geq 0$, there exist a constant $c_{r,k}$ such that

$$\left|\phi^{(k)}(t)\right| \leq c_{r,k}|t|^{-r}.$$

Consider an estimate of $\phi(t)$ by the recurrence

$$\phi_{n+1}(t) = \int_0^1 \phi_n(ut)\phi_n((1-u)t)e^{itG(u)}du, n = 0, 1, 2, ... \tag{4.9}$$

Tan and Hadjicostas [46] used the Theorem 4.6 and proved the following result:

Theorem 4.9. *Let $\phi_0(t)$ be the characteristic function of ν with zero mean and finite second moment. Define recursively by* (4.9) *a sequence of complex valued functions ϕ_n. Then for each n, ϕ_n is the characteristic function of $S^n(\nu) \in D$.*

Moreover, by the Theorem 4.6, $S^n(\nu)$ converges in distribution to the fixed point ν of S, and therefore

$$\lim_{n\to\infty} \phi_n(t) = \phi(t)$$

for all real $t \in D$.

4.7.2 Bounds on Density function and derivatives

For every fixed $y, \bar{y}$, the random variable $h_{Y,\bar{Y}}(U)$ defined as in (4.7) is absolutely continuous. Given $(y, \bar{y}) \in \mathbb{R}^2$, Tan and Hadjicostas [46] considered a map related to the fixed point equation (4.7), $h_{y,\bar{y}} : [0,1] \to \mathbb{R}$ by

$$h_{y,\bar{y}} = \begin{cases} y+1 & \text{if } \ u = 1 \\ uy + (1-u)\bar{y} + G(u) & \text{if } \ u \in (0,1) \\ \bar{y}+1 & \text{if } \ u = 0 \end{cases} \tag{4.10}$$

where $G(u)$ is given by (4.6). If U is uniformly distributed on the unit interval $[0,1]$, then the function $h_{Y,\bar{Y}}(u)$ has a density with respect to the Lebesgue measure[46]. So, the limiting distribution of the random number of comparisons required by Quicksort to sort a list of n distinct elements has a density, and that density is positive almost everywhere.

From the Theorem 4.8, if $r = 0$ and $r = 2$, assume that the characteristic function is integrable over the real line, then according to the Fourier inversion theorem in [12], for all real x,, Y has a bounded continuous density $f(x)$ given by

$$f(x) = \frac{1}{2\pi} \int_{t=-\infty}^{\infty} e^{-itx} \phi(t) dt$$

By the Theorem 4.8, in the case of $r = k + 2$, then $t^k e^{-itx} \phi(t)$, and since the moment generating function of Y is everywhere finite [39], then the density f has derivatives of all orders [12] given by

$$f^{(k)}(x) = \frac{1}{2\pi} \int_{t=-\infty}^{\infty} (-it)^k \phi(t) dt, k \geq 0, x \in \mathbb{R}.$$

Therefore the derivatives are bounded,

$$\sup_x \left| f^{(k)}(x) \right| \leq \frac{1}{2\pi} \int_{t=-\infty}^{\infty} |t|^k |\phi(t)| \, dt, k \geq 0$$

Using theorem 4.8, [14] improved the result by [46] on the existence of a density and showed the next theorem

Theorem 4.10. *For each integer $k \geq 0$, the Quicksort limiting distribution has an k times continuously differentiable density function f for all $n \geq k+3$. For all $x \in \mathbb{R}$, there exist a constant C_k independent of n such that $\left| f^{(k)}(x) \right| \leq C_k$. Explicitly $|f(x)| \leq 16$ when $n \geq 5$, and $|f(x)| \leq 2466$.*

Chapter 5

Partial Quicksort

5.1 Introduction

The partial sort problem which appear in many applications is, given an array of size n, sort the l smallest elements. Sorting the whole array is an obvious solution, but it clearly too much. By using the divide and conquer strategy, the Quickselect algorithm solves the problem of finding only the l-th smallest element in an array of n [20]. The partial sorting algorithm named Partial Quicksort which is a modification of Quickselect answers this question.

Partial Quicksort was first proposed and analyzed by Martinez [32]. It is a simple variant of the Quicksort algorithm that solves the partial sorting problem, by combining selection and sorting into a single algorithm. It finds the l smallest elements of a given array containing n distinct elements and sorts them. If the array contains one or no elements, then we are done. Otherwise, the algorithm Partial Quicksort uses the principles of Quicksort, and works as follows.

- Choose a pivot by random with a uniform distribution;
- Find the list of strictly smaller numbers and the list of strictly larger numbers;
- If the rank of the pivot is smaller than l
  - Sort the left sub-array using 'Quicksort';
  - Apply Partial Quicksort to the right sub-array.
- If the rank of the pivot is greater than l
  - Apply Partial Quicksort to the left sub-array.
- Recall Partial Quicksort till termination.

The partitioning step needs exactly $n-1$ element comparisons. After that we face two lists under conditions like the initial. When $l=n$ the Partial Quicksort algorithm behaves exactly as Quicksort.

We introduce a nice version of Partial Quicksort, called Quicksort on the fly. Our algorithm is a suitable rearrangement of Partial Quicksort, it provides first the smallest, then second smallest, and so on until the largest element of a list of n different reals. Quicksort on the fly works as follow

- Pick by random with uniform distribution an element of the list as a pivot.
- Compare all others to the pivot and form the list of strictly smaller numbers, pivot, and the list of strictly larger numbers in this order.
- Recall independently the algorithm for the left sided list until termination.
- If necessary recall the algorithm for the right sided list.

For given array with large size it can be assumed that the running time of the algorithm is proportional to the number of comparisons needed to sort the array.

5.2 Running time analysis

In computer science, one of the major goals is to understand how to solve problems with computers. The efficiency of the running time analysis provides theoretical estimates for the resources needed by the Partial Quicksort, the time needed to sort l smallest elements of a given array containing n distinct elements. The running time of an algorithm is stated as a function depending on the computer, the number of steps and so on. Basically it suffices to consider only the number of comparisons done by Partial Quicksort. The running time will be proportional with a factor depending on the actual performing computer and program code. The running time of Partial Quicksort is proportional to the number of comparisons performed during the execution.

In more detail, let S be the set of distinct n elements ordered randomly. One of its elements is chosen as the pivot usually randomly with a uniform distribution. Split the list S into the left sub-array $S_<$, strictly smaller ones and the right sub-array $S_>$, strictly larger ones. If the l-th is in $S_<$ then all the required elements are in the left subarray, so we only need to make a recursive call for $S_<$. If the $l-th$ is in $S_>$ or the pivot then sort $S_<$ by the Quicksort algorithm and recall recursively the Partial Quicksort algorithm for $S_>$.

Now define the random variable, $X(S,l)$ of comparisons required to sort l smallest elements in an array of distinct $n=|S|$ numbers. The random variable X satisfies the recursive formula

$$(X(S,l))_l = \left(|S| - 1 + \mathbb{1}_{l \leq I} X^1(S_<, l) + \mathbb{1}_{l > I} \left(X^1(S_<, I-1) + X^2(S_>, l-I)\right)\right)_l \quad (5.1)$$

for $l = 1, 2, \cdots |S|$. Here the random variable $I(S) = |S_<| + 1$ represents the position of the rank of the pivot after comparisons and takes values in $\{1, 2, \ldots n\}$ with a uniform distribution. The random variables X^1, X^2 denote the number of comparisons on both subarrays using recursively the same algorithm. For given $S_<$ and $S_>$, the random variables X^1, X^2 are independent. $X(S_1, \cdots)$ satisfies a similar recursion and so on. The random variable I is independent of the random variables X^1 and X^2.

Notice the distribution of $X_1(S_<, |S_<|)$ in (5.1) is the Quicksort distribution sorting all $|S_<|$ numbers by some specified Quicksort version. The equation (5.1) determines recursively the distribution of $(X(S,l))_l$ in order to find the sorted l smallest elements out of the set S of distinct numbers.

Proposition 5.1. *Let $S, \overline{S}$ be two sets of n different reals , and let $l = 1, 2, \cdots n$. Then*

$$\mathcal{L}\left(X(S,l)\right) = \mathcal{L}\left(X(\bar{S},l)\right).$$

Proof. By induction on n. It is true for $n = 1$ and we are done. Assume it is true for $k \leq n$, so we use the notation

$$\mathcal{L}(X(S,l))_l = \mathcal{L}((X(|S|, l))_l \quad \text{for} \quad |S| \leq k. \quad (5.2)$$

The random variable $I(S) = I(|S|)$ is uniformly distributed on $\{1, 2, \cdots k\}$. Let $(X^1(k,l))_{l \in \{1,2,\cdot,k\}}$, $(X^2(k,l))_{l \in \{1,2,\cdot,k\}}$ be independent random variables independent of I and with the distribution given in (5.2). Since $|S_<|, |S_>|, \left|\bar{S}_<\right|, \left|\bar{S}_>\right| \leq n$ then

$$\mathcal{L}(X^1(S_<, l))_l = \mathcal{L}((X(\bar{S}_<, l))_l$$

and

$$\mathcal{L}(X^1(S_>, l))_l = \mathcal{L}((X(\bar{S}_>, l))_l$$

Now let $|S| = \left|\bar{S}\right| = n + 1$. Then

$$\begin{aligned}
\mathcal{L}(X(S,l))_l &= \sum_{i=1}^{k+1} P(I=i) \mathcal{L}(k - 1 + \mathbb{1}_{l \leq i} X^1(S_<, l) \\
&+ \mathbb{1}_{l > i}(X^1(S_<, i-1) + X^2(S_>, l-i)))_l \\
&= \sum_{i=1}^{k+1} P(I=i) \mathcal{L}(k - 1 + \mathbb{1}_{l \leq i} X^1(i-1, l) \\
&+ \mathbb{1}_{l > i}(X^1(i-1, i-1) + X^2(n-i, l-i)))_l \\
&= \sum_{i=1}^{k+1} P(I=i) \mathcal{L}(k - 1 + \mathbb{1}_{l \leq i} X^1(\bar{S}_<, l) \\
&+ \mathbb{1}_{l > i}(X^1(\bar{S}_<, i-1) + X^2(\bar{S}_>, l-i)))_l \\
&= \mathcal{L}(X(\bar{S}, l))_l.
\end{aligned}$$

$$\mathcal{L}(X^1(S_<, I-1)) = \mathcal{L}(X^1(|S_<|, I-1)),$$

and

$$\mathcal{L}(X^2(S_>, l-I)) = \mathcal{L}(X^2(|S_>|, l-I)).$$

Then

$$\mathcal{L}(X(S,l)) = \mathcal{L}(X(k+1,l)).$$

Then the statement is true for $n = k+1$ and therefore true for all $n \in \mathbb{N}$. $\square$

Remark 5.2. The above Proposition is true states that the distribution of $X(S,l)$ depends only on $|S|$ and l. It is true since we use the internal randomness. So we will use $X(S,l) \stackrel{\mathcal{D}}{=} X(|S|, l)$ for any input S (see [33]).

The equation (5.1) determines recursively the distribution of comparisons $X(n,l)$ needed to sort the l smallest elements in an array of size n. Using the Proposition 5.1, we obtain the recursion for $X(n,\cdot)$

$$\begin{aligned} X(n,l) &\stackrel{\mathcal{D}}{=} n-1+\mathbb{1}_{l<I_n}X^1(I_n-1,l)+\mathbb{1}_{l\geq I_n}X^1(I_n-1,I_n-l) \\ &+ \mathbb{1}_{l\geq I_n}X^2(n-I_n,l-I_n) \end{aligned} \tag{5.3}$$

The random variables $I_n, X^1(i,j)$ and $X^2(i,j)$, $i,j \in \mathbb{N}$, $j \leq i < n$ are independent. The random variables $X^i(j,.)$ have the same distribution recursively as $X(j,.)$. The random variable $I_n = I(n,l)$ is uniformly distributed on $\{1,2,\dots n\}$. We put for natural reasons the boundary conditions,

$$X(0,0) \equiv 0, X(1,0) \equiv 0 \equiv X(1,1).$$

Notice that $X(j,j)$, the random number of comparisons required by to sort j elements is the cost of the algorithm Quicksort and $X(n,n) = X(n,n-1)$.

For all our purposes the distribution of the X random variables is important, not the realization as random variables. For that reason we consider the equation (5.3) for distributions. In our version of Quicksort that we discussed in Chapter 4, we use the internal randomness by picking the pivot with a uniform distribution. Like in standard Quicksort, we could use external randomness instead of internal randomness. Choose as input a uniform distribution on all permutation π on order n and pick as pivot any, for example always the first in the list.

Now $X(\pi,\cdot)$ for input π we face the same distribution as with internal randomness. The main advantage using internal randomness is the same distribution of X for every input. Alternatively we could start with independent and identically distributed random variables uniformly on $[0,1]$ and choose as pivot always the first element of the list. Again X is deterministic for given independent and identically distributed sequence like for random input.

From the recursive formula (5.3) we obtain the best and worst performance of the Partial Quicksort algorithm. The best behavior is by picking incidentally the l-th largest for the Find procedure and then doing Quicksort in its best. When the pivot is the largest all the time, then we face the worst performance of the algorithm [33].

5.3 Average number of comparisons

From equation (5.3), we obtain a recursion for the expected number of comparisons $a(n,l) = E(X(n,l))$

$$\begin{aligned} a(n,l) &= n-1+E\left(\mathbb{1}_{l+1\geq I}(X_1(I-1,I-1)+X_1(n-I,l-I))\right)+E\left(\mathbb{1}_{I>l+1}X_2(I-1,l)\right) \\ &= n-1+\sum_{k=1}^{l}\pi_{n,k}\left(a(k-1,k-1)+a(n-k,l-k)\right)+\sum_{k=l+1}^{n}\pi_{n,k}a(k-1,l) \end{aligned}$$

for $n \in \mathbb{N}$, $l = 1, 2, \cdots, n$. Here $\pi_{n,k}$ denote the probability that the chosen pivot is the k-th element among the n given distinct elements.

Up to now, everything holds no matter which pivot selection scheme we do use. It is usual in the analysis of comparison-based sorting algorithms, to assume that the probability of any permutation of the given distinct n elements is equally the same. So we take $\pi_{n,k} = \frac{1}{n}$ for all $1 \leq k \leq n$ in the standard case and therefore we get the recursion

$$\begin{aligned} a(n,l) &= n-1+\frac{1}{n}\sum_{k=1}^{l}(a(k-1,k-1)+a(n-k,l-k)) \\ &+ \frac{1}{n}\sum_{k=l+1}^{n}a(k-1,l), l>0 \end{aligned} \tag{5.4}$$

with the boundary conditions

$$\forall n \in \mathbb{N} \quad a(n,0)=0 \quad \text{and} \quad \forall l \in \mathbb{N} \quad a(0,l)=0.$$

The recurrence (5.4) reflects the cost for splitting the array $a(k-1,k-1), a(n-k,l-k)$ and the cost for sorting $a(k-1,l)$. In this recurrence, the number of comparisons done to choose the pivot is $n-1$, and k refers to the position of the pivot chosen with probability $\frac{1}{n}$.

When k is smaller than l, the algorithm fully sorts the entire left subarray, and partial quicksort call itself recursively for the right subarray of size $n-k$. In the other case

when k is greater than l, the algorithm sorts the l smallest elements in the left subarray of size $k-1$.
Let $t_{n,l}$ denote the toll function given by

$$t_{n,l} = n - 1 + \frac{1}{n}\sum_{k=1}^{l} a(k-1, k-1)$$

Hence, we get the recurrence

$$a(n,l) = t_{n,l} + \frac{1}{n}\sum_{k=1}^{l} a(n-k, l-k) + \frac{1}{n}\sum_{k=l+1}^{n} a(k-1, l) \tag{5.5}$$

The recurrence (5.5) is the recurrence for the average number of comparisons made by Quickselect. The next lemma represents the toll function expressed in terms of the harmonic number.

Lemma 5.1. *The toll function defined above is*

$$t_{n,l} = n - 1 + \frac{1}{n}\left(l(l+1)H_l + \frac{l}{2}(1-5l)\right).$$

Proof. By using the well known properties of the nth harmonic number

$$\sum_{k=1}^{n}\binom{k}{m} H_k = \binom{n+1}{m+1}\left(H_{n+1} - \frac{1}{m+1}\right)$$

and for m=1 we get

$$\sum_{k=1}^{n} kH_k = \binom{n+1}{2}\left(H_{n+1} - \frac{1}{2}\right)$$

also $\sum_{k=1}^{n} H_k = (n+1)H_n - n$, see [16]

$$\begin{aligned}
\sum_{k=1}^{l} a(k-1, k-1) &= \sum_{k=1}^{l}(2kHk - 1 - 4(k-1)) \\
&= 2\sum_{k=1}^{l} kH_k - 2l - 4\sum_{k=1}^{l}(k-1) \\
&= l(l+1)H_l + l - \frac{5}{2}l(l+1) + 2l \\
&= l(l+1)H_l + \frac{l}{2}(1-5l).
\end{aligned}$$

□

The solution of the equation (5.5) is solved exactly using different techniques. Martinez [32] translated the equation (5.5) into a functional relation over the bivariate generating functions. He used bivariate generating functions as the main tool associated to the quantities $a(n,l)$ and $t_{n,l}$ and solving the arising differential equation. Kuba [29] connected the Partial Quicksort algorithm with the Multiple Quickselect algorithm. He introduced a general method to find the solutions for a class of recurrences related to the Multiple Quickselect algorithm which is again based on a difference argument of Knuth.

The second approach is much easier using a general theorem and the Lemma 5.1. Since $a(n,l)$ is a function of n then all $a(i,\cdot), i < n$ is uniquely determined under some boundary conditions. So if the rank of the pivot I_n is uniformly distributed on $\{1, 2, \ldots n\}$, the next Lemma [32] give an explicit solution for the equation (5.5)

Lemma 5.2. *The average number of comparisons done by Partial Quicksort to sort the l smallest elements out of n elements is given by*

$$a(n,l) = 2n + 2(n+1)H_n - 2(n+3-l)H_{n+1-l} - 6l + 6$$

if $1 \leq l \leq n$ and 0 otherwise.

Here $H_n = \sum_{k=1}^{n} \frac{1}{k}$, $n \in \mathbb{N}$ denotes the n-th harmonic number and the values of the sequence $(H_n - \ln n)$ decrease monotonically towards the limit:

$$\lim_{n\to\infty} (H_n - \ln n) = \gamma,$$

where $\gamma = 0.5772156649\ldots$, is called Euler constant [27].
As we have already studied in the previous chapter, when $n = l$ we recognize that the partial Quicksort works exactly the same way as Quicksort. So that

$$a(n,n) = a_n = 2(n+1)H_n - 4n.$$

Chapter 6

Partial Quicksort as a process

The aim of this chapter is to introduce the Partial Quicksort process or simply Quicksort process. Asymptotic running times of algorithms are widely studied. We focus here on the distribution of comparisons done by Partial Quicksort to sort the l smallest elements out of n elements, which is proportional to the running time of the algorithm discussed in the previous chapter. We will specify the Partial Quicksort process as a weighted branching process, by using the weighted branching process notation. The main result in that chapter is the convergence almost everywhere of the sequence $(R_n)_{n\in\mathbb{N}}$ to a limit called R. Moreover the distribution of the limit R will be a fixed point.
In order to study the convergence behavior of $X(n,l)$ as s process as $n\to\infty$, we suggest a suitable normalization of $X(n,l)$ as

$$Y_n(\frac{l}{n}) := \frac{X(n,l+1)-a(n,l+1)}{n+1} \tag{6.1}$$

were $a(n,j)$ is given as in Lemma 5.2 for $1\leq j\leq n$. Assume for simplicity that the rank $I(n,l)$ of the pivot is uniformly distributed on $\{1,2,\cdots,n\}$. $I=I(n,l)$ depends only on l and on the size n of the list. We have a recursion given by the following Lemma

Lemma 6.1. *For $l=0,1,\cdots,n-1, n\in\mathbb{N}$ and $Y_n(1)=Y_n(\frac{n-1}{n})$, the random variable $Y_n(\cdot)$ satisfying the recursion*

$$\begin{aligned}\left(Y_n(\frac{l}{n})\right)_l &\stackrel{\mathcal{D}}{=} (\mathbb{1}_{l+1\leq I_n}\frac{I_n}{n+1}Y^1_{I_n-1}(\frac{l}{I_n-1})\\ &+ \mathbb{1}_{l+1>I_n}\left(\frac{I_n}{n+1}Y^1_{I_n-1}(1)+(1-\frac{I_n}{n+1})Y^2_{n-I_n}(\frac{l-I_n}{n-I_n})\right)\\ &+ C(n,l,I_n))_l\end{aligned}$$

where

$$\begin{aligned}C(n,l,i) &= \frac{1}{n+1}(\mathbb{1}_{l+1<i}(a(i-1,l+1)+\mathbb{1}_{l+1=i}a(i-1,i-1))\\ &+ \mathbb{1}_{l+1>i}(a(i-1,i-1)+a(n-i,l+1-i)))_{0\leq l\leq n}.\end{aligned} \tag{6.2}$$

Proof. Straight forward by using (6.1), (5.3) and Lemma 5.2. □

In the next lemma we will give an explicit formula of the function $C(n, l, i)$. For the proof we need the well known properties of the n-th harmonic number [16]

$$\sum_{k=1}^{n} kH_k = \binom{n+1}{2}\left(H_{n+1} - \frac{1}{2}\right),$$

$\sum_{k=1}^{n} H_k = (n+1)H_n - n$ and $nH_{n-1} = nH_n - 1$.

Lemma 6.2. *The function C as defined in* (6.2) *has the explicit representation*

$$\begin{aligned}
C(n,l,i) &= 2\frac{i}{n+1}(H_i - H_{n+1}) + 2(1 - \frac{i}{n+1})(H_{n+1-i} - H_{n+1}) \\
&+ 2\mathbb{1}_{l\le i}(-\frac{i+2-l}{n+1}(H_{i+2-l} - H_{n+1}) \\
&- (1 - \frac{i}{n+1})(H_n + 1 - i) - H_{n+1} + (n+3-l)(H_{n+1-l} - H_{n+1}) \\
&- n + i - 2 + \frac{1}{(n+1)(i+1+l)} - \frac{1}{(n+1)(n+2-l)}).
\end{aligned}$$

Proof. From equation (6.2), Lemma 5.2 and using the properties of the n-th harmonic number we have

$$\begin{aligned}
(n+1)C(n,l,i) &= n - 1 - a(n,l) + \mathbb{1}_{l<i}a(i-1,l) + \mathbb{l} = \mathbb{i}a(i-1,i-1) \\
&+ \mathbb{1}_{l>i}(a(i-1,i-1) + a(n-i,l-i)) \\
&= -n - 1 - 2(n+1)H_n + 2(n+3-l)H_{n+1-l} + 6l - 6 \\
&+ \mathbb{1}_{l<i}(2(i-1) + 2iH_{i-1} - 2(i+2-l)H_{i-1} - 6l + 6) \\
&+ \mathbb{1}_{l=i}(2iH_{i-1} - 4(i-1)) + \mathbb{1}_{l<i}(2(n-i) + 2(n-i+1)H_{n-i}) \\
&- 2(n+3-l)H_{n+1-i} - 6(l-i) + 6 \\
&= -n - 1 - 2(n+1)H_{n+1} + 2(n+3-l)H_{n+1-l} + 2iH_i \\
&+ 2\mathbb{1}_{l<i}(-(i+2-l)H_{i-1} + i - 1) \\
&+ 2\mathbb{1}_{l=i}(l-1) + 2\mathbb{1}_{l>i}((n+1-i)H_{n+1-i} - 2(n+3-l)H_{n+1-l} + n + 1) \\
&= 2iH_i + 2(n+1-i)H_{n+1-i} - 2(n+1)H_{n+1} + n + 1 \\
&+ 2\mathbb{1}_{l<i}(-n + i - 2 - (i+2-l)H_{i-l} - (n+1-i)H_{n+1-i} \\
&+ (n+3-l)H_{n+1-l}) \\
&+ 2\mathbb{1}_{l=i}(-n + l - 2 - (n+1-i)H_{n+1-i} + 2(n+3-i)H_n + 1 - l) \\
&= 2iH_i + 2(n+1-i)H_{n+1-i} - 2(n+l)H_{n+1} + n + 1 \\
&+ 2\mathbb{1}_{l\le i}(-n + i - 2 - (i+2-l)H_{i-1} - (n+1-i)H_{n+1-i} \\
&+ (n+3-l)H_{n+1-l}) \\
&= 2iH_i + 2(n+1-i)H_{n+1-i} - 2(n+l)H_{n+1} + n + 1 \\
&+ 2\mathbb{1}_{l\le i}(-n + i - 2 - (i+2-l)H_{i+2-l} - (n+1-i)H_{n+1-i} \\
&+ (n+3-l)H_{n+3-l} - n + i - 2 + \frac{1}{i+1-l} - \frac{1}{n+2-l}).
\end{aligned}$$

□

Notice Y_n is well defined and there are no boundary conditions besides $Y_0 = 0 = Y_1$. The equation (6.2) allows an analysis of the asymptotic of $Y_n(\cdot)$ in n.

Proposition 6.1. *Let $l = l_n$ and $i = i_n$ be two sequences such that $\frac{l}{n} \to t \in [0,1]$ and $\frac{i_n}{n} \to x \in [0,1]$ as $n \to \infty$. Then the cost function $C(n, l_n, i_n)$ defined as in (6.2) converges almost everywhere to a function $C(x,t)$, given by*

$$\begin{aligned} C(x,t) &= 1 + 2x\ln x + 2(1-x)\ln(1-x)) + 2\mathbb{1}_{t\leq x}((1-t)\ln(1-t) \\ &- (1-x)\ln(1-x) - (x-t)\ln(x-t) - (1-x)) \end{aligned} \tag{6.3}$$

Proof. The cost function $C(n, l_n, i_n)$ takes the form

$$\begin{aligned} C(n,l,i) &= \mathbb{1}_{l\geq i}\left(\frac{a(i-1,i-1)}{n} + \frac{a(n-i_n, l-i_n)}{n} + \frac{n-1}{n} - \frac{a(n,l)}{n}\right) \\ &+ \mathbb{1}_{l<i}\left(\frac{a(i-1,l)}{n} + \frac{n-1}{n} - \frac{a(n,l)}{n}\right) \end{aligned} \tag{6.4}$$

Using lemma 5.2, and as $n \to \infty$ then $\frac{l_n}{n}$ converges to some $t \in [0,1]$ and $\frac{i_n}{n}$ converges to some $x \in [0,1]$ and $t \neq x$

For the first term in the equation (6.4)

$$\begin{aligned} \lim_{n\to\infty} &\left(\frac{a(i-1,i-1)}{n} + \frac{a(n-i,l-i)}{n} + \frac{n-1}{n} - \frac{a(n,l)}{n}\right) = \\ = \quad & 2u\ln u + 2u\ln n - 4n + 2(1-u) + 2(1-u)\ln(1-u) + 2(1-u)\ln n \\ - \quad & 2(1-t)\ln(1-t) - 2(1-t)\ln n - 6(t-u) - 2 - 2t\ln n \\ + \quad & 2(1-t)\ln(1-t) + 6t + 1 \\ = \quad & 2u\ln u + 2(1-u)\ln(1-u) + 1 \end{aligned} \tag{6.5}$$

and for the second term in equation (6.4)

$$\begin{aligned} \lim_{n\to\infty} &\left(\frac{a(i-1,l)}{n} + \frac{n-1}{n} - \frac{a(n,l)}{n}\right) = \\ = \quad & 2u + 2u\ln u + 2u\ln n - 2(u-t)\ln(u-t) - 2(u-t)\ln n \\ - \quad & 6t - 2 - 2t\ln n + 2(1-t)\ln(1-t) + 6t + 1 \\ = \quad & 2u + 2u\ln u + 2(1-t)\ln(1-t) - 2(u-t)\ln(u-t) - 1 \end{aligned} \tag{6.6}$$

Combining equation (6.5) with equation (6.6) finishes the proof.

□

Extend $Y_n(\cdot)$ nicely to a suitable right continuous step function with values in function space $D = D[0,1]$, the space of cadlag functions on the unit interval $[0,1]$.

We shall use the extension

$$Y_n(t) := Y_n(\frac{\lfloor nt \rfloor}{n}). \tag{6.7}$$

Here $\lfloor x \rfloor$ denotes to the largest integer less than or equal to x. The process Y_n is continuous at 1 and satisfying the recursion

$$\begin{aligned} Y_n(t) &\stackrel{\mathcal{D}}{=} (\mathbb{1}_{t<U_n} \frac{I_n}{n+1} Y^1_{I_n-1}(\frac{nt}{I_n-1} \wedge 1) \\ &+ \mathbb{1}_{t\geq U_n} \left(\frac{I_n}{n+1} Y^1_{I_n-1}(1) + (1 - \frac{I_n}{n+1}) Y^2_{n-I_n}(\frac{t-U_n}{1-U_n}) \right) \\ &+ C(n, \lfloor nt \rfloor, I_n) \end{aligned}$$

for $n \geq 1$ with the boundary conditions

$$\forall n \in \mathbb{N} \quad Y_n(0) = 0 \quad \text{and} \quad \forall t \in [0,1] \quad Y_0(t) = 0.$$

In short notations the above equation can be written as

$$Y_n \stackrel{\mathcal{D}}{=} \varphi_n(U_n, (Y^1_k)_{k<n}, (Y^2_k)_{k<n}) \tag{6.8}$$

for a suitable $\varphi_n : [0,1] \times D^n \times D^n \to D$.

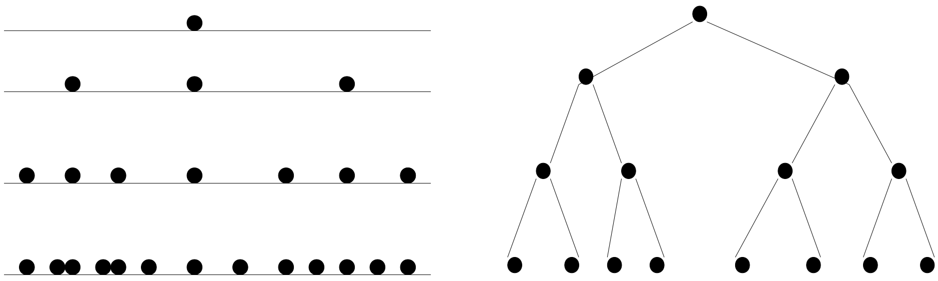
If $n \to \infty$ U_n converges weakly to a random variable U with uniform distribution on $[0,1]$. Under appropriate conditions the next theorem show that Y_n converges weakly to a aprocess Y with values in D. The next theorem is one of our major results.

Theorem 6.2. *The process $(Y_n(\frac{l}{n}))_{l\in\{0,1,2,\cdots,n\}}$ converges as $n \to \infty$ in distribution to a process $Y = (Y(t))_{t\in[0,1]}$ with values in D satisfying the stochastic fixed point equation*

$$(Y(t))_{t\in[0,1]} \stackrel{\mathcal{D}}{=} (\mathbb{1}_{t\geq U} \left(UY^1(1) + (1-U)Y^2(\frac{t-U}{1-U}) \right) + \mathbb{1}_{t<U} UY^1(\frac{t}{U}) + C(U,t))_{t\in[0,1]}$$

Here U is a uniformly distributed random variable on the unit interval $[0,1]$. The random variables Y^1, Y^2, U are independent. The random variables Y^1 and Y^2 have the same distribution as Y with values in D. The cost function $C = C(\cdot,\cdot)$ is given by equation (6.3).

Figure 6.1.a illustrate the sorting by the Partial Quicksort and Figure 6.1.b illustrate the corresponding binary tree.



Figur 6.1.a   Figur 6.1.b

Figure 6.1: Partial Quicksort and the corresponding binary tree.

6.1 Model description

Let $G := D \times D_{\uparrow}$ and let $H := D$. For all $f, f_1, f_2 \in D$ and $g, g_1, g_2 \in D_{\uparrow}$ define the operation $* : G \times G \to G$ by

$$*\left((f_1, g_1), (f_2, g_2)\right) := (f_1, g_1) * (f_2, g_2) = (f_1 \cdot f_2 \circ g_1, g_2 \circ g_1)\,. \tag{6.9}$$

where $\circ$ denotes the convolution and $\cdot$ is the pointwise multiplication in D.

For all $f, f_1, h \in D$ and $g, g_1 \in D_{\uparrow}$, define the operation $\otimes : G \times H \to H$ by

$$\otimes\left((f, g), h\right) := (f, g) \otimes h = f \cdot h \circ g.$$

For all $f, f_1, f_2, \acute{f}, \acute{f_1}, \acute{f_2} \in D$,the operation $*$ is bilinear in the first coordinate on $D \times D_{\uparrow}$,

$$(f_1 + f_2, g) * (\acute{f}, \acute{g}) = (f_1, g) * (\acute{f}, \acute{g}) + (f_2, g) * (\acute{f}, \acute{g})$$
$$(f, g) * ((\acute{f_1} + \acute{f_2}, \acute{g}) = (f, g) * (\acute{f_1}, \acute{g}) + (f, g) * (\acute{f_2}, \acute{g}).$$

And for all $f, \acute{f_1}, \acute{f_2}, h \in D$ $g, \in D_{\uparrow}$

$$(f_1 + f_2, g) \otimes h = (f_1, g) \otimes h + (f_2, g) \otimes h,$$
$$(f, g) \otimes (h_1 + h_2) = (f, g) \otimes h_1 + (f, g) \otimes h_2.$$

The tuple $(f, g) \in G$ has the interpretation of a map $M_{f,g} : H \to H$ acting as

$$(M_{f,g}(h))(t, n) = f(t, n)h(g(t, n)). \tag{6.10}$$

The first coordinate f is a space transformation and the second coordinate g is a time and index transformation. The semigroup structure $*$ is the composition of the corresponding maps. Since H is a vector space and $\mathbb{R}$ is a lattice, we will embed G to maps H^H and use freely the induced structures $+$,$\cdot$ and $\vee$.

$$(M_{f,g} + M_{f_1,g_1})(h) = M_{f,g}(h) + M_{f_1,g_1}(h),$$
$$a \cdot (M_{f,g})(h) = (a \cdot M_{f,g})(h),$$

and

$$(M_{f,g} \vee M_{f_1,g_1})(h) = ((M_{f,g})(h)) \vee (M_{f_1,g_1}(h)).$$

It is easy to see the equation (6.9) as follow

$$\begin{aligned} M_{f,g} \circ M_{f_1,g_1}(h)(t,n) &= M_{f,g}(f_1(t,n), h(g_1(t,n))) \\ &= f(t,n) \cdot (f_1, h(g_1))(g(t,n)) \\ &= f(t,n) f_1(g(t,n) h(g_1(g(t,n))) \\ &= M_{f \cdot f_1 \circ g, g_1 \circ g}(h)(t,n). \end{aligned}$$

Proposition 6.3. $(G,*)$ *is a measurable semigroup with neutral element* $(\underline{1}, \underline{id})$, *and the grave grave* Δ *in* G *is the element* $(\underline{0}, \underline{0})$ *where* $\underline{id}$ *is the identity and* $\underline{1}$, $\underline{0}$ *denote the function constant 1 and 0 respectively.* G *operates via* $\otimes$ *transitive and measurable on the measurable space* H.

To specify the Partial Quicksort process discussed in the previous chapter as a weighted branching process we translate the above recurrence into weighted branching process notations. The recurrence of the Partial Quicksort process see [33] takes the form

$$(Y^v(t))_{t\in[0,1]} = \left(\mathbb{1}_{U^v \le t}(1-U^v) \cdot Y^{v2}\left(\frac{t-U^v}{1-U^v}\right) + \mathbb{1}_{U^v > t} U^v \cdot Y^{v1}\left(\frac{t}{U^v}\right) + C(U^v, t) \right)_{t\in[0,1]},$$

where the cost function C is given by

$$\begin{aligned} C(U^v,t) &= 2U^v \ln U^v + \mathbb{1}_{U^v \le t}(1 + U^v Q^{v1} + 2(1-U^v)\ln(1-U^v)) \\ &+ \mathbb{1}_{U^v > t}\left(2U^v - 1 + 2(1-t)\ln(1-t) - 2(U^v - t)\ln(U^v - t)\right). \end{aligned} \tag{6.11}$$

Here $U^v, v \in \mathbb{V}$ is uniformly distributed random variable on the unit interval $[0,1]$. The random variable Q^v has a limiting Quicksort distribution. The random variables Y^1, Y^2, U^v and Q^v are independent. The random variables Y^1 and Y^2 have the same distribution as Y with values in D. In the next section we will construct the corresponding binary tree.

6.1.1 Binary tree

Consider the binary tree

$$\mathbb{V} = \{1,2\}^{\mathbb{N}}.$$

For $v = v_1 v_2 \cdots v_n \in \mathbb{V}, n \in \mathbb{N}$ and for $m \in \mathbb{N}$, $m \le n$, $v|m = v_1 v_2 \cdots v_m$ denote to the m-th coordinate of v. Let $U^v : \Omega \to [0,1], v \in \mathbb{V}$ be independent and identically uniformly distributed random variables on the unit interval $[0,1]$. Let $Q^v, v \in \mathbb{V}$ be the random variable has a limiting Quicksort distribution.

Define a map $T_i^v : \Omega \to G$, the weights on the edges $(v, vi), v \in \mathbb{V}, i \in \{1,2\}$ by

$$T_i^v := (A_i^v, B_i^v) \quad \text{for all} \quad v \in \mathbb{V}$$

and

$$T^v = (T_1^v, T_2^v, 0, \cdots)$$

For all $v \in \mathbb{V}$, define the following parameters

$$\begin{aligned}
A_1^v(t) &:= \mathbb{1}_{U^v > t} U^v, \\
A_2^v(t) &:= \mathbb{1}_{U^v \leq t}(1 - U^v), \\
A_3^v(t) &:= 0 = A_4^v(t) = \cdots \\
B_1^v(t) &:= \left(\frac{t}{U^v} \wedge 1\right)_{t \in [0,1]} \\
B_2^v(t) &:= \left(\frac{t - U}{1 - U} \vee 0\right)_{t \in [0,1]} \\
B_3^v(t) &:= 0 = B_4^v(t) = \cdots .
\end{aligned} \tag{6.12}$$

Define a map $C^v : \Omega \to H$, the vertex weight by

$$C^v := C(U^v, .) + \mathbb{1}_{U^v \leq t} U^v Q^{v1}, \tag{6.13}$$

where C is given in (6.3).

Remark 6.4. Here $(A^v, B^v, C^v), v \in \mathbb{V}$ in terms of U^v are iid copies of (A, B, C). The edge weight T_i^v attached to the edge (v, vi) is given by (A_i^v, B_i^v). The tuple $(\mathbb{V}, (T^v, C^v)_{v \in \mathbb{V}}, (G, *, H, \otimes))$ is a weighted branching process.

Consider the weighted branching process as given above and for all $v \in \mathbb{V}$, define the sequence

$$R_n^v := \sum_{w \in \mathbb{V}_{<n}} L_w^v \otimes C^{vw}, \quad R_0^v = 0, \tag{6.14}$$

where the family of path weights $L^v := (L_w^v)_{w \in \mathbb{V}}$ from the node v is given recursively by (3.1).

Lemma 6.3. *Let U be a uniformly distributed random variable on $[0,1]$ and C defined as in* (6.13)*. Then for all $t \in [0,1)$ $E(C) = 0$ and $Var(C) < \infty$.*

Proof. From equations (6.11), we have

$$\begin{aligned} E(C) &= 2\int_0^1 u\ln u du + E\int_0^t (1+uQ+2(1-u)\ln(1-u))du \\ &+ \int_t^1 (2u-1+2(1-t)\ln(1-t)-2(u-t)\ln(u-t))du. \end{aligned}$$

It follows

$$\int_0^1 u\ln u du = \int_0^1 u^2(\ln u)^2 du = -\frac{1}{4},$$

$$\int_0^1 (1-u)^2(\ln(1-u))^2 du = \frac{2}{27},$$

$$\int_0^t (1-u)\ln(1-u)du = \frac{1}{2}\left(-(1-t)^2\ln(1-t)+\frac{t^2}{2}-t\right),$$

$$\int_t^1 (u-t)\ln(u-t)du = \frac{1}{2}\left((1-t)^2\ln(1-t)-\frac{1}{2}(1-t)^2\right),$$

$$\int_0^t (t-u)\ln(t-u)du = \frac{1}{2}\left(t^2\ln t-\frac{t^2}{2}\right),$$

and

$$\int_0^1 u(1-u)(\ln u)\ln(1-u)du = \frac{1}{108}\left(37-3\pi^2\right) \approx 0,068437 < 1.$$

Therefore

$$E(C)) = 2\int_0^1 u(\ln u)du + \int_0^t (1+uQ+2(1-u)\ln(1-u))du = 0.$$

Moreover

$$\begin{aligned} E\left(C^2\right) &\leq 1+4\int_0^1 u\ln u du + 4\int_0^1 u^2(\ln u)^2 du \\ &+ 4\int_0^1 (1-u)\ln(1-u)du \\ &+ 4\int_0^1 (1-u)^2(\ln(1-u))^2 du \\ &+ 8\int_0^1 u(1-u)\ln u\ln(1-u)du \\ &+ \frac{1}{3}E(Q^2) \\ &< \infty. \end{aligned}$$

□

Our first main result is the following statement.

Theorem 6.5. *Let $(V, ((T_1, T_2), C), (G, *), (H, \otimes))$ be the weighted branching process defined as above. Then R_n^v converges uniformly as $n \to \infty$ almost everywhere in D to a random variable R^v for all $v \in \mathbb{V}$. The family $R^v, v \in \mathbb{V}$ satisfies*

$$R^v = \sum_{i=1}^{2} T_i^v R^{vi} + C^v \tag{6.15}$$

almost everywhere. Moreover, for every $p > 1$ holds

$$\| \|R\|_\infty \|_p \leq \frac{8 + (\frac{1}{p+1})^{\frac{1}{p}} \|Q\|_p}{1 - k_p} \tag{6.16}$$

where $k_p = (\frac{2}{p+1})^{\frac{1}{p}}$ and Q is a random variable with the Quicksort distribution.

Before we prove the above theorem we state the following consequence.

Lemma 6.4. *The random variables R_n^v defined as in (6.14) satisfies the backward recursion*

$$R_m^v = \sum_{i=1}^{2} T_i^v \otimes R_{m-1}^{vi} + C^v$$

for all $v \in \mathbb{V}$ and $m \in \mathbb{N}$.

Proof. The sum R_m^v is well defined and by equations (6.14) and (3.2), we have

$$\begin{aligned}
R_m^v &= \sum_{|w|<m} (T_{w_1}^v * T_{w_2}^{vw_1} * \cdots T_{w_k}^{vw_1 \cdot w_{k-1}}) \otimes C^{vw} \\
&= \sum_{k=0}^{m} \sum_{w \in V_k} (T_{w_1}^v * T_{w_2}^{vw_1} * \cdots T_{w_k}^{vw_1 \cdot w_{k-1}}) \otimes C^{vw} \\
&= C^v + \sum_{k=1}^{m} \sum_{i=1}^{2} \sum_{x \in V_{k-1}} (T_i^v * T^{vix_1} * \cdots T^{vix_1 \cdots x_{k-2}}) \otimes C^{vix} \\
&= C^v + \sum_{i=1}^{2} T_i^v * (\sum_{k=1}^{m} \sum_{x \in V_{k-1}} T^{vix_1} * \cdots T^{vix_1 \cdots x_{k-2}}) \otimes C^{vix} \\
&= C^v + \sum_{i=1}^{2} T_i^v * \sum_{|x|<n-1} L_x^{vi} \otimes C^{vix} \\
&= C^v + \sum_{i=1}^{2} T_i^v * R_{m-1}^{vi}.
\end{aligned}$$

□

The connection between the operator K introduced in Chapter 3 and the weighted branching process R_n, in the case of the Partial Quicksort is given by the following Corollary.

Corollary 6.6. *Let the starting measure μ_0 be the point measure on the function $\underline{0} \in D$ identical 0. Then the random variables R_n defined as in (6.14) satisfies*

$$R_{n+1} = C^{\phi} + \sum_{i \in \mathbb{N}} A_i^{\phi} \cdot R_n^i \circ B_i^{\phi}, \tag{6.17}$$

where R_n^i denotes the the random variable R_n for the tree with root i. The distribution of R_n is $K^n(\mu_0)$.

Proof. In the positive case all R_n and $n-$fold iterates $K^n(\mu_0)$ are well defined [25]. By Lemma 6.4, the sequence R_n satisfies the backward recursion

$$R_{n+1} = C_{\phi} + \sum_{i \in \mathbb{N}} T_i^{\phi} \otimes R_n^i.$$

If $T_i^v := (A_i^v, B_i^v)$ for all $v \in \mathbb{V}, i \in \mathbb{N}$ and $T^v = (T_1^v, T_2^v)$, then by the settings in Section 6.1 we have

$$\begin{aligned} R_n &= C^{\phi} + \sum_{i \in \mathbb{N}} \left(A_i^{\phi}, B_i^{\phi}\right) \otimes R_n^i \\ &= C^{\phi} + \sum_{i \in \mathbb{N}} A_i^{\phi} \cdot R_n^i \circ B_i^{\phi}. \end{aligned}$$

To prove the distributional result on R_n we use the mathematical induction on n. The induction base case when $n = 1$ is true since $R_1 = C^{\phi}$ has the distribution $K(\mu_0)$. For the inductive step n to $n+1$ argue by the backward recursion

$$\begin{aligned} R_{n+1} &= C^{\phi} + \sum_{i \in \mathbb{N}} A_i^{\phi} \cdot R_n^i \circ B_i^{\phi} \\ &= K(\mathcal{L}(R_n)) \\ &= K(K^n(\mu_0)) \\ &= K^{n+1}(\mu_0). \end{aligned}$$

Here the random variables $\left(A_i^{\phi}, B_i^{\phi}, C^{\phi}\right), R_n^i, i \in \mathbb{N}$ are independent. □

For all $v \in \mathbb{V}, n \in \mathbb{N}$ define $S_n^v : \Omega \to D$ by

$$S_n^v := \sum_{|w|=n} L_w^v \otimes C^{vw}, \quad S_0^v = C^v. \tag{6.18}$$

The equation (6.18) defines the total weight in the n-th generation. Our interest concentrates on the total weight (cost) regarded up to the n-th generation. From equations (6.18) and (3.3), we have

$$S_n^v = R_{n+1}^v - R_n^v \tag{6.19}$$

Using equation (3.5) it is easy to show that for all $\omega \in \Omega$ and $S_n(\omega) \in D$,

$$S_n^v = \sum_{i\in\mathbb{N}} T_i^\phi * S_{n-1}^{vi} \tag{6.20}$$

where T_i defined as in (6.12) and S_n^1 and S_n^2 are the S_n random variable for the tree with root i.

Using the equation (6.12) and (6.20), for all $t \in [0,1]$ we obtain

$$\begin{aligned} S_n^v(t) &= \sum_{i\in\mathbb{N}} \left(A_i^v \cdot S_{n-1}^{vi} \circ B_i^v\right)(t) \\ &= \left(A_1^v \cdot S_{n-1}^{v1} \circ B_1^v + A_2^v \cdot S_{n-1}^{v2} \circ B_2^v\right)(t) \end{aligned}$$

and therefore

$$S_n^v(t) = \mathbb{1}_{U^v>t} U^v S_{n-1}^{v1}(\frac{t}{U^v}) + \mathbb{1}_{U^v\leq t}(1-U^v) S_{n-1}^{v2}(\frac{t-U^v}{1-U^v}). \tag{6.21}$$

Lemma 6.5. *Let $S_n^v, n \in \mathbb{N}, v \in \mathbb{V}$ be as above. Then for all $n \in \mathbb{N}$, $E(\|S_n\|_\infty) = 0$ and $Var(\|S_n\|_\infty)$ converges exponentially fast to 0, as $n \to \infty$.*

Proof. Notice $(S_n^v)_n \stackrel{\mathcal{D}}{=} (S_n^\phi)_n$. By equation (6.21)

$$\begin{aligned} \|S_n\|_{2,D}^2 &= \|T_i S_{n-1}\|_{2,D}^2 \\ &= E\left(\sup_t \left|\mathbb{1}_{t<U} U S_{n-1}^1(\frac{t}{U}) + \mathbb{1}_{t\geq U}(1-U) S_{n-1}^2(\frac{t-U}{1-U})\right|^2\right) \\ &= E\sup_t\left(\left(\mathbb{1}_{t<U} U S_{n-1}^1(\frac{t}{U})\right)^2 + \left(\mathbb{1}_{t\geq U}(1-U) S_{n-1}^2(\frac{t-U}{1-U})\right)^2\right). \end{aligned}$$

All mixed terms are zero. The first term is

$$\begin{aligned} E\sup_t\left(\mathbb{1}_{t<U} U S_{n-1}^1(\frac{t}{U})\right)^2 &= E\left(U^2 \sup_{t\leq u}(S_{n-1}^1(\frac{t}{U}))^2\right) \\ &\leq E\left(U^2 \left\|S_{n-1}^1\right\|_\infty^2\right) \\ &\leq E(U^2) E\left(\left\|S_{n-1}^1\right\|_\infty^2\right) \end{aligned} \tag{6.22}$$

And the second term is

$$\begin{aligned} E\sup_t\left(\mathbb{1}_{t\geq U}(1-U) S_{n-1}^2(\frac{t-U}{1-U})\right)^2 &= E\left((1-U)^2 \sup_{t\geq U} S_{n-1}^2(\frac{t-U}{1-U})^2\right) \\ &\leq E\left((1-u)^2 \left\|S_{n-1}^2\right\|_\infty^2\right) \\ &\leq E\left(1-u)^2\right) E\left(\left\|S_{n-1}^2\right\|_\infty^2\right) \end{aligned} \tag{6.23}$$

Then we have

$$\begin{aligned} \|S_n\|_{2,D}^2 &\leq \left(EU^2 + E(1-U)^2\right) E\left(\left\|S_{n-1}^2\right\|_\infty^2\right) \\ &\leq \frac{2}{3} E\left(\left\|S_{n-1}^2\right\|_{2,D}^2\right). \end{aligned}$$

For $n \in \mathbb{N}_0$, let $b_n := \|S_{n-1}\|_{2,D}$. Notice b_n does not depend on the vertex v. Then

$$b_n^2 \leq \frac{2}{3} b_{n-1}^2,$$

and therefore by iteration of the inequality

$$b_n^2 \leq \left(\frac{2}{3}\right)^n b_0^2 < \infty.$$

Where $b_0 = \|S_0\|_{2,D} = \|C(U,.)\|_{2,D} < \infty$. □

Let $A(t) := \sum_i |A_i(t)|$ for all $t \in [0,1]$ and note that $E\|A\|_\infty = E(U) + E(1-U) = 1$ and

$$\sum_{i=1,2} E \sup_{t\in[0,1]} |A_i(t)|^2 = EU^2 + E(1-U)^2 = 2E(U^2) = \frac{2}{3} < 1.$$

Lemma 6.6. *For fixed $t \in [0,1]$ and $n \in \mathbb{N}$, $Var(R_n(t))$ converges exponentially fast to 0, as $n \to \infty$.*

Proof. The random variables $S_j, \in \mathbb{N}$ are pointwise well defined and measurable. By Theorem 6.5 we have

$$\begin{aligned} E\left(R_n(t)^2\right) &= E\left(\left(\sum_{j=0}^{n-1} S_j(t)\right)^2\right) \\ &= E\left(\sum_i \sum_j S_j(t) S_j(t)\right) \\ &= \sum_i \sum_j E\left(S_i(t) S_j(t)\right) \\ &= \sum_{i=0}^{n-1} E\left((S_i(t))^2\right) + \sum_{i=0}^{n-1} \sum_{i\neq j} E\left(S_i(t) S_j(t)\right). \end{aligned}$$

By the Cauchy-Schwarz inequality, $S_i(t)S_j(t)$ is integrable. For $i,j \in \mathbb{N}$ and $i \leq j$, define

$$\mathcal{B}_i := \sigma\left((T^v, C^v)_{|v|\leq n}\right).$$

By conditional expectation we have for $i \leq j$

$$\begin{aligned} E\left(S_i(t)S_j(t)\right) &= E\left[E\left(S_i(t)S_j(t)|\mathcal{B}_i\right)\right] \\ &= E\left[S_i(t)E\left(S_j(t)|\mathcal{B}_i\right)\right] = 0 \end{aligned}$$

Therefore

$$E\left((R_n(t))^2\right) = \sum_{i=0}^{n-1} E\left((S_i(t))^2\right)$$

And therefor Lemma 6.5 finishes the proof. □

Lemma 6.7. *For all* $t \in [0,1]\,, n \in \mathbb{N}$

$$\sum_n \|S_n\|_\infty < \infty \quad \textit{almost everywhere.}$$

Proof. By equation (2.9) and lemma 6.5, we get

$$\begin{aligned} E(\sum_{n\in\mathbb{N}} \|S_n\|_\infty) &= \sum_{n\in\mathbb{N}} E(\|S_n\|_\infty) \\ &= \sum_{n\in\mathbb{N}} \|\|S_n\|_\infty\|_1 \\ &\leq \sum_{n\in\mathbb{N}} \|\|S_n\|_\infty\|_2 \\ &\leq \sum_{n\in\mathbb{N}} \|S_n(t)\|_{2,D} \\ &\leq \sum_{n\in\mathbb{N}} \left(\frac{2}{3}\right)^{\frac{n}{2}} b_0 < \infty \end{aligned}$$

as $n \to \infty$. □

Lemma 6.8. *The sequence* $(R_n)_{n\in\mathbb{N}_0}$ *is a Cauchy sequence in* $\left(\mathcal{F}_2(D), \|\cdot\|_{2,D}\right)$.

Proof. We have to show $\forall \epsilon > 0, \exists n_0 \in \mathbb{N}$ such that

$$\forall n > m \geq n_0, \quad \|R_n - R_m\|_{2,D} < \epsilon.$$

By Lemma 2.18 and Lemma 6.5, the triangle inequality holds for all $n > m \geq n_0$

$$\begin{aligned}
\|R_n - R_m\|_{2,D} &= \left\|\sum_{l=m}^{n-1} S_l\right\|_{2,D} \\
&\leq \sum_{l=m}^{n-1} \|S_l\|_{2,D} \\
&\leq \sum_{l\geq n_0} b_l \\
&\leq \left(\frac{2}{3}\right)^{\frac{n_0}{2}} b_0 \cdot \left(1-\sqrt{\frac{2}{3}}\right)^{-1} \\
&\to 0 \quad \text{as} \quad n_0 \to \infty.
\end{aligned}$$

Here for given ϵ use n_0 such that

$$n_0 = \inf\left\{m \in \mathbb{N} : \sum_{j\geq m} \|S_j\|_{2,D} < \epsilon\right\}.$$

□

Now we come back to the proof of the Theorem 6.5.

Proof. of the Theorem 6.5
From the above Lemma 6.1.1, the Cauchy sequence $R_n, n \in \mathbb{N}$ converges to some $R = R_\infty$. Since $R_n(t) = \sum_{n\in\mathbb{N}} S_n(t)$, R is the point wise limit of $\sum_{n\in\mathbb{N}} S_n(t)$ for all $t \in [0,1]$. From the Lemma 6.1.1

$$\begin{aligned}
E\|R - R_n\|_\infty &\leq E\left\|\sum_{i\geq n} S_i\right\|_\infty \\
&\leq \sum_{i\geq n} E\|S_i\|_\infty \\
&\to 0 \quad as \quad n \to \infty.
\end{aligned}$$

Then R is well defined almost everywhere and for every realization $\omega \in \Omega$ and every $t \in [0,1]$, the limit $R(\omega)(t)$ exists. By the backward view in Lemma 6.4 going to the limit as $n \to \infty$ and by the almost everywhere convergence R satisfies the equation (6.15). For the equation (6.16), by the backward equation in Lemma 6.4 we have

$$\|R_n\|_\infty = \left\|T_1 R_{n-1}^1 \vee T_2 R_{n-1}^2\right\|_\infty + \|C\|_\infty.$$

Consider the weighted branching process as above but with cost function

$$\bar{C}^v := 8 + U^v |Q^v|, \quad v \in \mathbb{V}.$$

Define

$$\bar{R}_n^v := \sum_{j<n} \bigvee_{w \in V_j} L_w^v \bar{C}^{vw}.$$

Then $\bar{R}_0 = 0$ and $\bar{R}_n \uparrow \bar{R}$ point wise,

$$\bar{R}_n^v = T_1^v \bar{R}_{n-1}^{v1} \vee T_2^v \bar{R}_{n-1}^{v2} + \bar{C}^v$$

for $n \in \mathbb{N}$. By an induction on n we show $\left\|R_{n+1}^v\right\|_\infty \leq \bar{R}_{n+1}^v$. If $n = 1$, $\|R_1^v\|_\infty = \|C^v\|_\infty \leq 8$ then $\|R_1^v\|_\infty \leq \bar{R}_1^v$. For the induction step n to $n+1$ argue by

$$\begin{aligned} \left\|R_{n+1}^v\right\|_\infty &\leq \left\|T_1 R_n^{v1} \vee T_2 R_n^{v2}\right\|_\infty + \|C^v\|_\infty \\ &\leq \left\|T_1 R_n^{v1} \vee T_2 R_n^{v2}\right\|_\infty + \bar{C}^v \\ &\leq T_1 R_n^{v1} \vee T_2 R_n^{v2} + \bar{C}^v = \bar{R}_{n+1}^v. \end{aligned}$$

Then for $p > 1$

$$\begin{aligned} \left\|\bar{R}_n\right\|_\infty &\leq \left\|T_1 \bar{R}_n^{v1} \vee T_2 \bar{R}_n^{v2}\right\|_p + \left\|\bar{C}\right\|_p \\ &\leq (E\left|T_1 \bar{R}_{n-1}^1\right|^p + E\left|T_2 \bar{R}_{n-1}^2\right|^p)^{\frac{1}{p}} + 8 + \|UQ\|_p \\ &\leq \|R_{n-1}\|_p (E|T_1|^p + E|T_2|^p)^{\frac{1}{p}} + 8 + \|UQ\|_p \\ &\leq k_p \|R_{n-1}\|_p + 8 + \|UQ\|_p \\ &\leq \sum_{i=0}^n k_p^i \left(\|U\|_p \|Q\|_p + 8\right) \\ &\leq \frac{8 + \|U\|_p \|Q\|_p}{1 - k_p}. \end{aligned}$$

where

$$\begin{aligned} k_p &= E|T_1|^p + E|T_2|^p \leq E|U|^p + E|1-U|^p \\ &\leq 2E(U^p) \leq \frac{2}{p+1}. \end{aligned}$$

And therefore

$$\left\|\bar{R}\right\|_p = \lim_{n\to\infty} \left\|\bar{R}_n\right\|_p = \frac{8 + (\frac{1}{p+1})^{\frac{1}{p}} \|Q\|_p}{1 - k_p}.$$

□

6.2 Convergence of the discrete Quicksort process

In this section we prove the convergence of finite dimensional marginals of Y_n to Y. We will define a nice version of Y_n such that $Y_n(t)$ converges in L_2-norm to $Y(t)$ for every

$t \in [0,1]$. This requires to define a nice family $(Y_n^v)_{n\in\mathbb{N}}$ of random variables with values in D, indexed by the tree $\mathbb{V}$. We will include the Partial Quicksort process via the index ∞, compare this with the Weighted branching processes in Chapters 4 and 6.

Let $\mathbb{V} = \{1,2\}^{\mathbb{N}}$ be the binary tree and define

$$H := \left\{ h : [0,1] \times \bar{\mathbb{N}}_0 \to \mathbb{R} | \forall n \in \bar{\mathbb{N}}_0 \quad h(\cdot, n) \in D \right\}$$

Here $\bar{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$. Let G_2 be the set of all maps $g : [0,1] \times \bar{\mathbb{N}}_0 \to [0,1] \times \bar{\mathbb{N}}_0$ such that $g(\cdot, 0) = 0, \quad \forall n \in \mathbb{N} \quad \Phi_2(g(\cdot, n)) < n$ and $\Phi_2(g(\cdot, \infty)) = \infty$. Here ϕ_2 denotes the projection to the second coordinate.
Let $G = H \times G_2$ and define the semigroup operation $* : G \times G \to G$ by

$$*\left((f_1, g_1), (f_2, g_2)\right) := (f_1, g_1) * (f_2, g_2) = (f_1 \cdot f_2 \circ g_1, g_2 \circ g_1)$$

for all $f_1, f_2 \in [0,1]$ and $g_1, g_2 \in \bar{\mathbb{N}}_0$. Here $\circ$ denotes the convolution and $\cdot$ is the pointwise multiplication in $[0,1]$. The semigroup $(G, *)$ has the neutral element $(1, id)$, the function identically 1, the identity and the grave is $(0, id)$.
For all $f, f_1, h \in [0,1]$ and $g, g_1 \in G_2$, define the operation $\otimes : G \times H \to H$ by

$$\otimes\left((f,g), h\right) := (f,g) \otimes h = f \cdot h \circ g.$$

For all $f, \acute{f_1}, \acute{f_2}, h \in H$ and $g \in G_2$

$$(f_1 + f_2, g) \otimes h = (f_1, g) \otimes h + (f_2, g) \otimes h,$$

$$(f,g) \otimes (h_1 + h_2) = (f,g) \otimes h_1 + (f,g) \otimes h_2.$$

The tuple $(f,g) \in G$ has the interpretation of a map $M_{f,g} : H \to H$ via

$$(M_{f,g}(h))(t,n) = f(t,n) h(g(t,n)). \tag{6.24}$$

The first coordinate f is a space transformation and the second coordinate g is a time and index transformation. The semi group structure $*$ is the composition of the corresponding maps. Since H is a vector space and $\mathbb{R}$ is a lattice, we will embed G to maps H^H and use freely the induced structures $+$,$\cdot$ and $\vee$.

$$(M_{f,g} + M_{f_1,g_1})(h) = M_{f,g}(h) + M_{f_1,g_1}(h),$$
$$a \cdot (M_{f,g})(h) = (a \cdot M_{f,g})(h),$$

and

$$(M_{f,g} \vee M_{f_1,g_1})(h) = \left((M_{f,g})(h)\right) \vee \left(M_{f_1,g_1}(h)\right).$$

Let $U^v : \Omega \to [0,1]\,, v \in \mathbb{V}$ be independent and identically uniformly distributed random variables on the unit interval $[0,1]$. Let $Q^v, v \in \mathbb{V}$ be a random variable has a limiting Quicksort distribution.

Define $T_i^v : \Omega \to G$, the weights on the edges $(v, vi), v \in \mathbb{V}, i \in \{1,2\}$ by

$$T_i^v := (A_i^v, (B_i^v, J_i^v)) \quad \text{for all} \quad v \in \mathbb{V}, i \in \mathbb{N},$$

and $T^v = (T_1^v, T_2^v)$ with values in G and the vertex weight C^v with values in H by the following coefficients

$$\begin{aligned}
I_n^v &= \lceil nU \rceil \\
U_n^v &= \frac{I_n^v}{n} \\
J_1^v(t,n) &= I_n^v - 1 \\
J_2^v(t,n) &= n - I_n^v \\
A_1^v(t,n) &= \mathbb{1}_{t<U_n^v} \frac{I_n^v - 1}{n} \\
A_2^v(t,n) &= \mathbb{1}_{t \geq U_n^v} (1 - \frac{I_n^v}{n}) \\
B_1^v(t,n) &= 1 \wedge \frac{\lfloor nt \rfloor}{I_n^v - 1} \\
B_2^v(t,n) &= 0 \vee \frac{t - U_n^v}{1 - U_n^v} \\
C^v(t,n) &= C(n, \lfloor nt \rfloor, I_n^v) + \mathbb{1}_{t \geq u_n^v} \frac{I_n^v - 1}{n} Q^{v1}_{I_n^v - 1}
\end{aligned} \tag{6.25}$$

Notice $J_1^v(t,n) + J_2^v(t,n) = n - 1$. For $n \in \mathbb{N}, v \in \mathbb{V}$, with the above coefficients we expect for the limit the following parameters:

$$\begin{aligned}
J_1^v(t,\infty) &= \infty \\
J_2^v(t,\infty) &= \infty \\
A_1^v(t,\infty) &= \mathbb{1}_{t<U^v} U^v \\
A_2^v(t,\infty) &= \mathbb{1}_{t \geq U^v} (1 - U^v) \\
B_1^v(t,\infty) &= 1 \wedge \frac{t}{U^v} \\
B_2^v(t,\infty) &= 0 \vee \frac{t - U^v}{1 - U^v} \\
C^v(t,\infty) &= c(U^v, t) + \mathbb{1}_{t \geq U^v} U^v Q^{v1}
\end{aligned} \tag{6.26}$$

where $t \in [0,1]$ and for $i = 1, 2$, $A_i^v(\cdot, 0)$ is identically 0.

Define

$$R_m^v(t,n) := \sum_{w \in \mathbb{V}_{<m}} (L_w^v \otimes C^{vw})(t,n) \tag{6.27}$$

for $m, n \in \bar{\mathbb{N}}_0, v \in \mathbb{V}, t \in [0,1]$.

Proposition 6.7. *The random variables R_m^v defined as above satisfy the backward recursion*

$$R_m^v(t,n) = \sum_{i=1}^{2} (T_i^v \otimes R_{m-1}^{vi})(t,n) + C^v(t,n)$$

for all $m \in \bar{\mathbb{N}}_0$ and $v \in \mathbb{V}$.

Proof. The sum R_m^v is well defined and by equations (6.27) and (3.2), we have

$$\begin{aligned}
R_m^v(t,n) &= \sum_{|w|<m} ((T_{w_1}^v * T_{w_2}^{vw_1} * \cdots T_{w_k}^{vw_1 \cdot w_{k-1}}) \otimes C^{vw})(t,n) \\
&= \sum_{k=0}^{m} \sum_{w \in V_k} ((T_{w_1}^v * T_{w_2}^{vw_1} * \cdots T_{w_k}^{vw_1 \cdot w_{k-1}}) \otimes C^{vw})(t,n) \\
&= C^v(t,n) + \sum_{k=1}^{m} \sum_{i=1}^{2} \sum_{x \in V_{k-1}} ((T_i^v * T^{vix_1} * \cdots T^{vix_1 \cdots x_{k-2}}) \otimes C^{vix})(t,n) \\
&= C^v(t,n) + \sum_{i=1}^{2} (T_i^v * (\sum_{k=1}^{m} \sum_{x \in V_{k-1}} T^{vix_1} * \cdots T^{vix_1 \cdots x_{k-2}}) \otimes C^{vix})(t,n) \\
&= C^v(t,n) + \sum_{i=1}^{2} (T_i^v * \sum_{|x|<n-1} L_x^{vi} \otimes C^{vix})(t,n) \\
&= C^v(t,n) + \sum_{i=1}^{2} (T_i^v * R_{m-1}^{vi})(t,n).
\end{aligned}$$

□

Lemma 6.9. *For every $n \in \mathbb{N}_0$, the function $R_m^v(\cdot,n)$ converges almost everywhere pointwise as $m \to \infty$ to $R_\infty^v(\cdot,n) = \sum_{w \in \mathbb{V}} L_w^v C^{vw}(\cdot,n)$ for every $v \in \mathbb{V}$. The limit satisfies the equation*

$$R_\infty^v(\cdot,n) = Y_n^v.$$

Proof. By definition $R_m^v(w)(t,n)$ is increasing in m for every realization $\omega \in \Omega$ and every t in the unit interval $[0,1]$. Therefore exists the limit R_∞^v which has only finitely many non zero summands. By equation 6.27 going to the limit as $m \to \infty$ we get

$$R_\infty^v(\cdot,n) = \sum_{w \in \mathbb{V}} L_w^v \otimes C^{vw}(\cdot,n)$$

for every $n \in \mathbb{N}_0, v \in \mathbb{V}$. We prove the second statement by induction on n. The cases $n = 0$ and $n = 1$ are true since $Y_0^v = 0 = R_\infty^v(\cdot,0)$ and $Y_1^v = 0 = R_\infty^v(\cdot,1)$. For the induction step use the representation given in equation (6.8) for Y_n and the backward equation in Proposition 6.7 for R_∞, so we have

$$\begin{aligned}
R^v_\infty(t,n) &= T^v_1 \otimes R^{v1}_\infty(t,n) + T^v_1 \otimes R^{v2}_\infty(t,n) + C^v(t,n) \\
&= A^v_1(t,n)R^{v1}_\infty(B^v_1(t,n), J^v_1(t,n)) + A^v_2(t,n)R^{v2}_\infty(B^v_2(t,n), J^v_2(t,n)) + C^v(t,n) \\
&= A^v_1(t,n)Y^{v1}_{J^v_1(t,n)}(B^v_1(t,n)) + A^v_2(t,n)Y^{v2}_{J^v_2(t,n)}(B^v_2(t,n)) + C^v(t,n) \\
&= \mathbb{1}_{t<U^v_n}\frac{I^v_n - 1}{n}Y^{v1}_{I^v_n - 1}(1 \wedge \frac{\lfloor nt \rfloor}{I^v_n - 1}) + \mathbb{1}_{t \geq U^v_n}(1 - \frac{I^v_n}{n})Y^{v2}_{I^v_n - 1}(0 \vee \frac{t - U^v_n}{1 - U^v_n}) + C^v(t,n) \\
&= Y^v_n.
\end{aligned}$$

□

Remark 6.8. By Lemma 6.9, $R^v(t,n) = R^v_\infty(t,n)$ is the limit of $R^v_m(t,n)$ as $m \to \infty$. Since $R^v_m(t,n) = \sum_{n\in\mathbb{N}_0} S^v_m(t,n)$ then $R(t,n)$ is the point wise limit of $\sum_{m\in\mathbb{N}_0} S^v_m(t,n)$ for all $t \in [0,1], v \in \mathbb{V}, n \in \bar{\mathbb{N}}_0$. Then R is well defined almost everywhere and for every realization $\omega \in \Omega$ and every $t \in [0,1]$, the limit $R^v_\infty(\omega)(t,n)$ exists almost everywhere. By the backward view in Proposition 6.7 going to the limit as $m \to \infty$ and by the almost everywhere convergence R^v_∞ satisfies the equation

$$R^v_\infty(t,n) = \sum_{i=1}^{2}(T^v_i \otimes R^{vi}_\infty)(t,n) + C^v(t,n).$$

For $m, n \in \bar{\mathbb{N}}_0, v \in \mathbb{V}$ and $t \in [0,1]$, define

$$S^v_m(t,n) = R^v_{m+1}(t,n) - R^v_m(t,n).$$

The last equation takes the form

$$S^v_m(t,n) = \sum_{w\in V_m}(L^v_w \otimes C^{vw})(t,n), \quad S^v_0(t,n) = 0.$$

In the following let $v, w, \bar{w} \in \mathbb{V}$, $m, \bar{n} \in \bar{\mathbb{N}}_0$ and $t \in [0,1]$. Notice $L^v_w = (A^v_w, B^v_w, J^v_w)$ acts as a map on H via

$$(L^v_w \otimes h)(t,n) =: A^v_w(t,n)h(B^v_w(t,n), J^v_w(t,n)) \tag{6.28}$$

Define an operator $E(L^v_w)$ acting on H by

$$E\left((L^v_w \otimes h)(t,n)\right) = \left((E(L^v_w))(h)\right)(t,n), \tag{6.29}$$

and define another operator $(L)^2$ on H via

$$((L)^2)(h) = (L(h))^2. \tag{6.30}$$

Let $\mathcal{A}_m$ be the σ-field generated by all random variables $U^v, v \in \mathbb{V}_{<m}, m \in \bar{\mathbb{N}}_0$. The random variable L^v_w is measurable with respect to $\mathcal{A}_{|vw|}$ and C^v is independent of $\mathcal{A}_{|v|}$. Before we look at the second main theorem we have to prove some necessary results.

Proposition 6.9. *For all* $v, w \neq \bar{w} \in \mathbb{V}$, $n \in \bar{\mathbb{N}}_0$ *and* $t \in [0,1]$, *holds*

$$E(L_w^v \otimes C^{vw}(t,n) \cdot L_{\bar{w}}^v \otimes C^{v\bar{w}}(t,n)) = 0 \quad \textit{for} \quad w \neq \bar{w}.$$

Proof. Since L_w^v is measurable with respect to $\mathcal{A}_{|vw|}$ and C^{vw} is independent of $\mathcal{A}_{|vw|}$, then by equation (6.28) we have

$$\begin{aligned} E(L_w^v \otimes C^{vw}(t,n)|\mathcal{A}_{|vw|}) &= E(A_w^v(t,n)C^{vw}((B_w^v, J_w^v)(t,n))|\mathcal{A}_{|vw|}) \\ &= (A_w^v(t,n))E(C^{vw}((B_w^v, J_w^v))(t,n)|\mathcal{A}_{|vw|}) \\ &= 0. \end{aligned}$$

We prove the statement for the possible two cases. For $|w| \neq |\bar{w}|$

$$\begin{aligned} L.H.S. &= E(L_w^v \otimes C^{vw}(t,n) \cdot L_{\bar{w}}^v \otimes C^{v\bar{w}}(t,n)) \\ &= \big(E(A_w^v(t,n)C^{vw}((B_w^v, J_w^v)(t,n))E(A_{\bar{w}}^v(t,n)C^{v\bar{w}}((B_{\bar{w}}^v, J_{\bar{w}}^v)(t,n))|\mathcal{A}_{|v\bar{w}|}))\big) \\ &= E(A_w^v(t,n)A_{\bar{w}}^v(t,n)(C^{vw}((B_w^v, J_w^v)(t,n))E(C^{v\bar{w}}((B_{\bar{w}}^v, J_{\bar{w}}^v)(t,n))|\mathcal{A}_{|v\bar{w}|})) \\ &= E(A_w^v(t,n)A_{\bar{w}}^v(t,n)(C^{vw}((B_w^v, J_w^v)(t,n)))E(C^{v\bar{w}}((B_{\bar{w}}^v, J_{\bar{w}}^v)(t,n))|\mathcal{A}_{|v\bar{w}|})) \\ &= 0. \end{aligned}$$

The last equality follows from equation (6.31). For $|w| = |\bar{w}|$, then we hav

$$\begin{aligned} L.H.S. &= E(L_w^v \otimes C^{vw}(t,n) \cdot L_{\bar{w}}^v \otimes C^{v\bar{w}}(t,n)) \\ &= E\left(E(A_w^v(t,n)C^{vw}((B_w^v, J_w^v)(t,n)) \cdot (A_{\bar{w}}^v(t,n)C^{v\bar{w}}((B_{\bar{w}}^v, J_{\bar{w}}^v)(t,n))|\mathcal{A}_{|vw|}))\right) \\ &= E(A_w^v(t,n)A_{\bar{w}}^v(t,n)E(C^{vw}((B_w^v, J_w^v)(t,n))C^{v\bar{w}}((B_{\bar{w}}^v, J_{\bar{w}}^v)(t,n))|\mathcal{A}_{|vw|})) \\ &= E(A_w^v(t,n)A_{\bar{w}}^v(t,n)(E(C^{vw}((B_w^v, J_w^v)(t,n))|\mathcal{A}_{|vw|})E(C^{v\bar{w}}((B_{\bar{w}}^v, J_{\bar{w}}^v)(t,n))|\mathcal{A}_{|vw|}))) \\ &= 0. \end{aligned}$$

Here we used the independence of C^{vw} and $C^{v\bar{w}}$ given $\mathcal{A}_{|vw|}$. ☐

Lemma 6.10. *For all* $v \in \mathbb{V}$ *and* $m \in \bar{\mathbb{N}}_0$, *we have*

$$\sup_{n\in\bar{\mathbb{N}}_0} \sum_{w\in V_m} E(\sup_{t\in[0,1]} A_w^v(t,n))^2 \leq (\frac{2}{3})^m.$$

Proof. Let $b(n,i) = \frac{i-1}{n} \vee \frac{n-i}{n}$. Since $E(I_n) = \frac{n(n+1)}{n}$ and $E((I_n)^2) = \frac{n(n+1)(2n+1)}{6}$, we have

$$\begin{aligned} E((b(n,I_n^v))^2) &= E(\frac{1}{n^2}(n^2+1+2(I_n^v)^2 - 2nI_n^v - 2I_n^v)) \\ &= \frac{1}{n^2}(n^2+1+2E((I_n^v)^2) - 2nE(I_n^v) - 2E(I_n^v)) \\ &= \frac{2}{3} - \frac{1}{n} + \frac{1}{3n^2} \leq \frac{2}{3}. \end{aligned}$$

By definitions of $A_i^v(t,n), i = 1, 2$ in (6.25), we note that $\sup_i \sup_t A_i^v(t,n) \leq b(n, I_n^v)$ and A_{iw}^v satisfies the recursion

$$A_{iw}^v(t,n) = A_i^v(t,n) A_w^{vi}((B_i^v, J_i^v)(t,n)).$$

Let $\bar{A}_w^v(n) = \sup_t A_w^v(t,n)$, then

$$\begin{aligned}
\sup_t \sum_i A_{iw}^v(t,n) &= \sup_t \sum_i A_i^v(t,n) A_w^{vi}((B_i^v, J_i^v)(t,n)) \\
&\leq \sup_i \sup_t A_i^v(t,n) A_w^{vi}((B_i^v, J_i^v)(t,n)) \\
&\leq \sup_i \sup_t A_i^v(t,n) \sup_j \sup_s A_w^{vi}((B_j^v, J_j^v)(s,n)) \\
&\leq b(n, I_n^v) \sup_j \bar{A}_w^{vi}(J_j^v).
\end{aligned}$$

The last inequality provides

$$\begin{aligned}
E((\bar{A}_{iw}^v(n))^2) &= E(\sup_t A_{iw}^v(t,n))^2 \\
&= E \sup_t \left(A_i^v(t,n) A_w^{vi}((B_i^v, J_i^v)(t,n)) \right)^2 \\
&\leq E \sup_t \left(A_i^v(t,n) A_w^{vi}((B_i^v, J_i^v)(t,n)) \right)^2 \\
&\leq E(b(n, I_n^v)^2 E \sup_t (A_w^{vi}((B_i^v, J_i^v)(t,n)))^2 \\
&\leq E((b(n, I_n^v))^2 \sup_i E(\bar{A}_w^{vi}(J_i^v))^2 | \mathcal{A}_{|v|}) \\
&\leq E((b(n, I_n^v))^2) \sup_i E((\bar{A}_w^{vi}(J_i^v))^2) \\
&\leq \frac{2}{3} \sup_i E((\bar{A}_w^{vi}(J_i^v))^2) \\
&\vdots \\
&\leq (\frac{2}{3})^m
\end{aligned}$$

The last inequality obtained by an induction on the length of w. □

Theorem 6.10. *$R_m^v(t,n)$ is an L_2-martingale in m with respect to $\mathcal{A}_{|v|+m}$ for all $t \in [0,1], n \in \mathbb{N}_0$ and $v \in \mathbb{V}$.*

Proof. Since L_w^v is measurable with respect to $\mathcal{A}_{|vw|}$ and C^{vw} is independent of $\mathcal{A}_{|vw|}$, then by equation (6.28) we have

$$\begin{aligned}
E(L_w^v \otimes C^{vw}(t,n) | \mathcal{A}_{|vw|}) &= E(A_w^v(t,n) C^{vw}((B_w^v, J_w^v)(t,n)) | \mathcal{A}_{|vw|}) \\
&= E(A_w^v(t,n)) E(C^{vw}((B_w^v, J_w^v))(t,n) | \mathcal{A}_{|vw|}) = 0.
\end{aligned}$$

Hence the martingale property follows by

$$E(R_m^v(t,n)|\mathcal{A}_{|v|+m}) = \sum_{w\in V_m} E(L_w^v \otimes C^{vw}(t,n)|\mathcal{A}_{|v|+m}) = 0$$

For the L_2 statement, let $a = \sup_{t\in[0,1]}\sup_{n\in\mathbb{N}_0} E((C(t,n))^2)$. By Proposition 6.9, we have

$$\begin{aligned}
\sup_{(t,n)} E(S_m^v(t,n))^2 &= \sum_{w\in V_m} E((A_m^v(t,n)C^{vw}((B_w^v, J_w^v)(t,n)))^2) \\
&= \sum_{w\in V_m} E(((A_m^v(t,n))^2 E(C^{vw}((B_w^v, J_w^v)(t,n)))^2|\mathcal{A}_{|vw|}) \\
&\leq \sup_{t\in[0,1]}\sup_{n\in\mathbb{N}_0} E((C(t,n))^2) \sum_{w\in V_m} E((A_m^v(t,n))^2) \\
&\leq a \sup_{t\in[0,1]} \sum_{w\in V_m} E((A_m^v(t,n))^2) \\
&\leq a \sum_{w\in V_m} \sup_{t\in[0,1]} E((A_m^v(t,n))^2) \\
&\leq a\cdot(\frac{2}{3})^m < \infty.
\end{aligned}$$

□

From the above Theorem we get the next Corollary

Corollary 6.11. *It holds that*

$$\sup_t\sup_n E(((R_\infty^v - R_m^v)(t,n))^2) \leq a\cdot(\frac{2}{3})^{m-1}, \quad v\in\mathbb{V}, m\in\bar{\mathbb{N}}_0.$$

Proof. From Theorem 6.10, we obtain

$$\begin{aligned}
E(S_m^v S_m^v) &= \sum_{v\in V_m} E((L_w^v \otimes C^{vw})^2) \\
&= E(\sum_{v\in V_m}(L_w^v \otimes C^{vw})^2) \\
&\leq E\sup_{(t,n)}(\sum_{v\in V_m}(L_w^v \otimes C^{vw})^2)(t,n)) \\
&\leq a\cdot(\frac{2}{3})^m
\end{aligned}$$

Therefore by the triangle inequality we get

$$\begin{aligned}
E(((R_\infty^v - R_m^v)(t,n))^2) &\leq \sum_{i\geq m} E(((R_{i+1}^v - R_i^v)(t,n))^2) \\
&\leq \sum_{i\geq m} E((S_m^v)^2) \\
&\leq a\cdot\sum_{i\geq m}(\frac{2}{3})^i = \frac{3}{2}a\cdot(\frac{2}{3})^m.
\end{aligned}$$

□

Now we come to the second main result of this chapter:

Theorem 6.12. *The above version of the Partial Quicksort process $Y_n(t)$ converges in L_2-norm to the limit $Y(t)$ for every $t \in [0,1]$. All finite-dimensional marginals converge and are uniformly integrable.*

Proof. Without loss of generality let t be none of the finitely many splitting points of the tree $v \in \mathbb{V}$ up to depths m. We estimate $\|R^v_m(t,n) - R^v_m(t,\infty)\|_2$ by the finite sum over all $w \in V_m$ of the terms $\|(L^v_w \otimes C^{vw})(t,n) - (L^v_w \otimes C^{vw})(t,\infty)\|_2$. We shall show every such term converges to 0.

$$\begin{aligned}
& \|R^v_m(t,n) - R^v_m(t,\infty)\|_2 = \\
= \; & \| \sum_{|w|<m} (L^v_w \otimes C^{vw})(t,n) - \sum_{|w|<m} (L^v_w \otimes C^{vw})(t,\infty)\|_2 \\
\leq \; & \sum_{|w|<m} \|A^v_w(t,n) C^{vw}((B^v_w, J^v_w)(t,n)) - A^v_w(t,\infty) C^{vw}((B^v_w, J^v_w)(t,\infty))\|_2 \\
\leq \; & \sum_{|w|<m} \|A^v_w(t,n) C^{vw}((B^v_w, J^v_w)(t,n)) - A^v_w(t,\infty) C^{vw}((B^v_w, J^v_w)(t,n)) \\
+ \; & A^v_w(t,\infty) C^{vw}((B^v_w, J^v_w)(t,n)) - A^v_w(t,\infty) C^{vw}((B^v_w, J^v_w)(t,\infty))\|_2
\end{aligned}$$

By the triangle inequality we have

$$\begin{aligned}
& \|R^v_m(t,n) - R^v_m(t,\infty)\|_2 \leq \\
\leq \; & \sum_{|w|<m} \|A^v_w(t,n) C^{vw}((B^v_w, J^v_w)(t,n)) - A^v_w(t,\infty) C^{vw}((B^v_w, J^v_w)(t,n))\|_2 \\
+ \; & \sum_{|w|<m} \|A^v_w(t,\infty) C^{vw}((B^v_w, J^v_w)(t,n)) - A^v_w(t,\infty) C^{vw}((B^v_w, J^v_w)(t,\infty))\|_2 \\
\leq \; & \sum_{|w|<m} \|A^v_w(t,n) - A^v_w(t,\infty)\|_2 \sup_{s\in[0,1]} \sup_i \|C^{vw}(s,i)\|_2 \\
+ \; & \sum_{|w|<m} \|A^v_w(t,\infty)\|_2 \|C^{vw}((B^v_w, J^v_w)(t,n)) - C^{vw}((B^v_w, J^v_w)(t,\infty))\|_2
\end{aligned}$$

The first term converges as $n \to \infty$ to 0 since $\|A^v_w(t,n) - A^v_w(t,\infty)\|_2$ converges to 0 almost everywhere and is uniformly bounded by 1. Hence we have

$$\|R^v_m(t,n) - R^v_m(t,\infty)\|_2 \leq \sum_{|w|<m} \|C^{vw}((B^v_w, J^v_w)(t,n)) - C^{vw}((B^v_w, J^v_w)(t,\infty))\|_2$$

By the triangle inequality

$$\begin{aligned}
\|R^v_m(t,n) - R^v_m(t,\infty)\|_2 \; & \leq \; \sum_{|w|<m} \|C(U^{vw}_{J^v_w(t,n)}) - C(U^{vw})\|_2 \\
& + \; \sum_{|w|<m} \|\mathbb{1}_{B^v_w(t,n) \geq U^{vw}_{J^v_w(t,n)}} J^v_{w1}(t,n) Q^{vw_1}_{J^v_{w1}(t,n)} - \mathbb{1}_{B^v_w(t,\infty) < U^{vw}} U^{vw} Q^{vw_1}\|_2
\end{aligned}$$

From the Corollary 6.11, U_m^{vw} converges almost everywhere to U^{vw} for $m \to \infty$. Then the dominated convergence provides that the first term converges to 0, since $J_w^v(t,n) \to \infty$ as $n \to \infty$ and the function C is bounded. Estimate the second term as follow

$$\begin{aligned}
\|R_m^v(t,n) - R_m^v(t,\infty)\|_2 &\le \sum_{|w|<m} \|(\mathbb{1}_{B_w^v(t,n) \ge U^{vw}_{J_w^v(t,n)}} - \mathbb{1}_{B_w^v(t,\infty) < U^{vw}}) J_{w_1}^v(t,n) Q^{vw_1}_{J_{w_1}^v(t,n)}\|_2 \\
&+ \sum_{|w|<m} \|\mathbb{1}_{B_w^v(t,\infty) < U^{vw}} (J_{w_1}^v(t,n) - U^{vw}) Q^{vw_1}_{J_{w_1}^v(t,n)}\|_2 \\
&+ \sum_{|w|<m} \|\mathbb{1}_{B_w^v(t,\infty) < U^{vw}} U^{vw} (Q^{vw_1}_{J_{w_1}^v(t,n)} - Q^{vw_1})\|_2 \\
&\le \sum_{|w|<m} \|\mathbb{1}_{B_w^v(t,n) \ge U^{vw}_{J_w^v(t,n)}} - \mathbb{1}_{B_w^v(t,\infty) < U^{vw}}\|_2 \sup_m \|Q_m\|_2 \\
&+ \sum_{|w|<m} \|J_{w_1}^v(t,n) - U^{vw}\|_2 \sup_m \|Q_m\|_2 \\
&+ \sum_{|w|<m} \|Q^{vw_1}_{J_{w_1}^v(t,n)} - Q^{vw_1}\|_2
\end{aligned}$$

The first term converges to 0, since $\|\mathbb{1}_{B_w^v(t,n) \ge U^{vw}_{J_w^v(t,n)}} - \mathbb{1}_{B_w^v(t,\infty) < U^{vw}}\|_2$ is bounded and converges almost everywhere to 0. The second term converges to 0, since $\|J_{w_1}^v(t,n) - U^{vw}\|_2$ is bounded and converges almost everywhere to 0. Hence we get

$$\|R_m^v(t,n) - R_m^v(t,\infty)\|_2 \le \sum_{|w|<m} \|Q^{vw_1}_{J_{w_1}^v(t,n)} - Q^{vw_1}\|_2.$$

Let $b_m := \|Q_m - Q\|_2$. Notice b_m converges almost everywhere to 0 as $m \to \infty$. Then we obtain

$$\begin{aligned}
\|R_m^v(t,n) - R_m^v(t,\infty)\|_2 &\le \sum_{|w|<m} E(E(Q^{vw_1}_{J_{w_1}^v(t,n)} - Q^{vw_1} | \mathcal{A}_{|vw_1|})) \\
&\le E(b_{J_{w_1}^n(t,n)}) \to 0 \quad as \quad n \to \infty.
\end{aligned}$$

And therefore by Lemma 6.9

$$\|Y_n(t) - Y(t)\|_2 \to 0 \quad as \quad n \to \infty$$

for all $t \in [0,1]$. □

Remark 6.13. For a vector $\underline{t} = (t_1, t_2, \cdots, t_k) \in \mathbb{R}^*, k \in \bar{\mathbb{N}}$, and a real valued function $f(t)$, let

$$f(\underline{t}) = (f(t_1), f(t_2), \cdots, f(t_k)))$$

A finite dimensional distribution of a process $X = (X(t))_{t\in[0,1]}$ is the distribution of the $X(\underline{t}), \underline{t} \in [0,1]^*$.

6.3 Conclusion

The online version of Partial Quicksort "Quicksort on the fly" provides returns the input in increasing natural order during the sorting process. The normalized asymptotic number of comparisons needed to sort the l-th smallest out of n appears as a weighted branching process in l. It appear as limiting distributions solutions of some stochastic fixed points equation of the form $Y \stackrel{\mathcal{D}}{=} \sum_i A_i Y_i \circ B_i + C$ with path in the space D of cadlag functions on the unit interval.

Appendix A

Mathematical basics for stochastic processes and their convergence

A.1 Basic definitions

In this section we introduce a short overview of the basic mathematical definitions for stochastic processes and we refer to [2], [7] and [21] for more detailed information. The purpose of this chapter is to give a short overview of the basic terminology used throughout the thesis, and to emphasize certain mathematical results for stochastic processes and its asymptotic properties which are important in the proofs.

A.1.1 Asymptotic notations

The asymptotic notations are mostly used in computer science to give a simple characterization of the algorithms efficiency and allows use to compare the relative performance of the alternative algorithms. Such notations are convenient for as we used for describing the best and worst cases, running time of the Quicksort.
Let $f, g, h : \mathbb{R} \to \mathbb{C}$, and suppose $c \in \mathbb{C}$, we define the following

Big-O

We write $f(n) = O(g(n))$ If

$$\lim_{n\to\infty} \sup \frac{f(n)}{g(n)} < \infty.$$

Little-o

We write $f(n) = o(g(n))$ if

$$\lim_{n\to\infty} \frac{|f(n)|}{|g(n)|} = 0.$$

Asymptotic Equality

We write $f(n) \sim g(n)$ if

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1.$$

Definition A.1. (filtration). Let $(\Omega, \mathcal{A}, P)$ be a probability space. A family of σ-fields $(\mathcal{F}_t)_{t\geq 0}$ is called a filtration, if $\mathcal{F}_s \subseteq \mathcal{F}_t$ for all $s \leq t$. If $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$ for all $t \in [0, \infty)$, $(\mathcal{F}_t)_{t\geq 0}$ is called right continuous filtration. If $\{A \in \mathcal{A} : P(A) = 0\} \subseteq \mathcal{F}_t$ for all t then a filtration $(\mathcal{F}_t)_{t\geq 0}$ is called complete.

For the following let $(E, \sigma(E))$ be a measurable space and $Y : \Omega \times [0, \infty) \to E$ be a stochastic process. Further we assume that $(\mathcal{F}_t)_{t\geq 0}$ is a right continuous and complete filtration.

Definition A.2. (adapted). The stochastic process $Y : \Omega \times [0, \infty) \to E$ is said to be adapted to $(\mathcal{F}_t)_{t\geq 0}$ if $Y_t : \Omega \to E$ is $\mathcal{F}_t$-measurable, i.e. $Y_t^{-1}(C) \in \mathcal{F}_t$ for all $C \in \mathcal{C}$.

Definition A.3. (finite-dimensional distributions). The finite-dimensional distributions of a stochastic process are the distributions of the finite-dimensional vectors

$$(Y_{t_1}, Y_{t_2}, \cdots Y_{t_k}), \quad t_1, t_2, \cdots, t_k \in [a, b]$$

such that $a = t_1 < t_2 < \cdots < t_k = b$ and $k \geq 1$.

A.1.2 (Conditional expectation)

Let $X : \Omega \to E$ be an integrable random variable, i.e. $E|X| < \infty$. The conditional expectation of X given $\mathcal{F}_t$ is a $\mathcal{F}_t$-measurable random variable and denoted by $E(X|\mathcal{F}_t)$ such that

$$\int_A E(X|\mathcal{F}_t) dP = \int_A X dP, \text{ for all } A \in \mathcal{F}_t.$$

The next proposition states the main properties of the conditional expectation.

Proposition A.4. *Let $X_1, X_2, X_3 : \Omega \to E$ be integrable random variables and let $s, t \in [0, \infty)$ with $s < t$. Then the following holds almost everywhere.*

1. $E\left(E(X|\mathcal{F}_t)\right) = E(X)$,
2. *If* $X_1 \leq X_2$ *then* $E(X_1|\mathcal{F}_t) \leq E(X_2|\mathcal{F}_t)$,
3. $E\left(aX_1 + bX_2|\mathcal{F}_t\right) = aE\left(X_1|\mathcal{F}_t\right) + bE\left(X_2|\mathcal{F}_t\right) \quad \forall a, b \in \mathbb{R}$,
4. $E(E(X|\mathcal{F}_t)|\mathcal{F}_s) = (E(X|\mathcal{F}_s)$.
5. $E(XY|\mathcal{F}_t)) = YE(X|\mathcal{F}_t))$ *for all* $\mathcal{F}_t$*-measurable bounded random variables* Y.

Proof. See Kallenberg [21]. □

Definition A.5. (martingale). The process Y is a martingale if it is $(\mathcal{F}_t)_{t\geq 0}$ adapted and integrable such that for all $t \in [0, \infty)$ holds

$$E(Y_t|\mathcal{F}_s) = Y_s \quad \forall s < t.$$

A.2 The Mallow Metric

In this section we introduce the notion Mallows metric on the space $\mathcal{M}_p$ of all one-dimensional distribution functions with existing p-th absolute moment which is used mainly in the statistics literature and in some literature on algorithms. Moreover show connections between the Mallow metric determined by the convergence and other types of convergences.

Definition A.6. For real $1 \leq p < \infty$ and $F, G \in \mathcal{M}_p$, the map $d_p : \mathcal{M}_p \times \mathcal{M}_p \to \mathbb{R}$, defined by

$$d_p(F, G) = \inf_{X \sim F, Y \sim G} \|X - Y\|_p \tag{A.1}$$

is called Mallow-metric d_p. The infimum is taken over all random variable X and Y on any probability space $(\Omega, \mathcal{A}, P)$, with $\mathcal{L}(X) = F$ and $\mathcal{L}(Y) = G$. The next proposition states the important properties of that metric.

Lemma A.1. *For all $F, G, H \in \mathcal{M}_p$ and $1 \leq p < \infty$, the map d_p defined as in equation* (A.1) *has the following properties*
1. $d_p(F, G) = 0$ iff $F = G$ (Identity).
2. $d_p(F, G) = d_p(G, F)$ (Symmetry).
3. $d_p(F, G) \leq d_p(F, H) + d_p(H, G)$ (Triangle inequality).

In addition the infimum is attained for $X = F^{-1}(U), Y = G^{-1}(U)$, where U is uniformly distributed on $[0, 1]$ and F^{-1} is the left continuous inverse of the right continuous distribution function F,

$$F^{-1}(x) = \inf \{y : F(y) \geq x\} \qquad for\ x \in [0, 1], \inf \phi = \infty.$$

And thus

$$d_p^p(F, G) = \left\|F^{-1}(U) - G^{-1}(U)\right\|_p^p = \int_0^1 \left|F^{-1}(u) - G^{-1}(u)\right|^p du.$$

Proof. Let F, G be distribution functions in $\mathcal{M}_p$. For the identity property, if $F = G$, then for a random variable X with $\mathcal{L}(X) = F$, we have

$$0 \leq d_p(F, G) = d_p(F, F) \leq \|X - X\|_p = 0.$$

Conversely, if $d_p(F,G) = 0$ for $F, G \in \mathcal{M}_p$

$$d_p(F,G) = \left\|F^{-1}(U) - G^{-1}(U)\right\|_p = 0$$

Then $F^{-1}(U) = G^{-1}(U)$ almost surely and thus $F = \mathcal{L}(F^{-1}(U)) = \mathcal{L}(G^{-1}(U)) = G$.

The symmetry property of the map d_p follows directly from the definition. For the triangle inequality, consider $F, G \in \mathcal{M}_p$ and U is uniformly distributed random variable on the interval $[0,1]$, then

$$\begin{aligned} d_p(F,G) &= \left\|F^{-1}(U) - G^{-1}(U)\right\|_p \\ &\leq \left\|F^{-1}(U) - H^{-1}(U)\right\|_p + \left\|H^{-1}(U) - G^{-1}(U)\right\|_p \\ &= d_p(F,H) + d_p(H,G). \end{aligned}$$

□

A sequence F_n of points in a metric space $(\mathcal{M}_p, d_p)$ converges to a limit distribution function $F \in \mathcal{M}_p$ if

$$\lim_{n\to\infty} d_p(F_n, F) = \lim_{n\to\infty} \left\|F_n^{-1}(U) - F^{-1}(U)\right\|_p = 0.$$

For $1 \leq p < \infty$, the metric space $(\mathcal{M}_p, d_p)$ is complete. In a natural way the metric d_p on $\mathcal{M}_p$ defines a pseudo-metric for random variables in L_p, by $d_p(X,Y) = d_p(\mathcal{L}(X), \mathcal{L}(Y))$.

Remark A.7. The d_p convergence is the same as weak convergence plus convergence of absolute moment of order p, more details, see [5], [40] and [3].

Notations

$\mathbb{N}_0$	$\{0, 1, 2, 3, ...\}$
$\mathbb{R}$	set of real numbers
$\mathbb{R}^+$	set of nonnegative real numbers
$\mathbb{C}$	set of complex numbers
$x \wedge y$	minimum of x and y
$x \vee y$	maximum of x and y
ln	main branch of complex logarithm
$(\Omega, \mathcal{A}, P)$	the underlying probability space
$f(t-)$, $f(t+)$	left-hand, right-hand limit of a real function f in t
$\mathbb{1}_A$	indicator function of a set A
$\mathscr{L}(X)$	the distribution of the random variable X
$E(X)$	the expected value of the random variable X
$Var(X)$	the variance of the random variable X
$\stackrel{\mathcal{D}}{=}$	equality in distribution
$\lvert x \rvert$	the modulus of $x \in \mathbb{C}$
$\sum_{\lvert v \rvert = n}$	summation over all nodes in the nth generation
$\lfloor x \rfloor$	the largest integer less than or equal to x
$\lceil x \rceil$	thesmallest integer greater than or equal to x
WBP	weighted branching process
GWP	Galton-Watson branching process

Bibliography

[1] David J. Aldous. Book Review: Markov processes: Characterization and convergence. *Bull. Amer. Math. Soc. (N.S.)*, 16(2):315–318, 1987.

[2] K. B. Athreya and P. E. Ney. *Branching processes.* Dover Publications Inc., Mineola, NY, 2004. Reprint of the 1972 original [Springer, New York; MR0373040].

[3] Peter J. Bickel and David A. Freedman. Some asymptotic theory for the bootstrap. *Ann. Statist.*, 9(6):1196–1217, 1981.

[4] Patrick Billingsley. *Convergence of probability measures.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[5] Robert M. Burton and Uwe Roesler. An L_2 convergence theorem for random affine mappings. *J. Appl. Probab.*, 32(1):183–192, 1995.

[6] Krzysztof Ciesielski. On Stefan Banach and some of his results. *Banach J. Math. Anal.*, 1(1):1–10, 2007.

[7] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to algorithms.* MIT Press, Cambridge, MA, second edition, 2001.

[8] Michael Cramer. A note concerning the limit distribution of the quicksort algorithm. *RAIRO Inform. Théor. Appl.*, 30(3):195–207, 1996.

[9] Luc Devroye. Exponential bounds for the running time of a selection algorithm. *J. Comput. System Sci.*, 29(1):1–7, 1984.

[10] Luc Devroye. A note on the height of binary search trees. *J. Assoc. Comput. Mach.*, 33(3):489–498, 1986.

[11] William F. Eddy and Mark J. Schervish. How many comparisons does Quicksort use? *J. Algorithms*, 19(3):402–431, 1995.

[12] William Feller. *An introduction to probability theory and its applications. Vol. II.* John Wiley & Sons Inc., New York, 1966.

[13] James Allen Fill and Svante Janson. Smoothness and decay properties of the limiting Quicksort density function. In *Mathematics and computer science (Versailles, 2000)*, Trends Math., pages 53–64. Birkhäuser, Basel, 2000.

[14] James Allen Fill and Svante Janson. Approximating the limiting Quicksort distribution. *Random Structures Algorithms*, 19(3-4):376–406, 2001. Analysis of algorithms (Krynica Morska, 2000).

[15] James Allen Fill and Svante Janson. The number of bit comparisons used by Quicksort: an average-case analysis. In *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 300–307 (electronic), New York, 2004. ACM.

[16] Daniel H. Greene and Donald E. Knuth. *Mathematics for the analysis of algorithms*. Modern Birkhäuser Classics. Birkhäuser Boston Inc., Boston, MA, 2008. Reprint of the third (1990) edition.

[17] Rudolf Gruebel. Hoare's selection algorithm: a Markov chain approach. *J. Appl. Probab.*, 35(1):36–45, 1998.

[18] Rudolf Gruebel and Uwe Roesler. Asymptotic distribution theory for Hoare's selection algorithm. *Adv. in Appl. Probab.*, 28(1):252–269, 1996.

[19] P. Hennequin. Combinatorial analysis of quicksort algorithm. *RAIRO Inform. Théor. Appl.*, 23(3):317–333, 1989.

[20] C. A. R. Hoare. Quicksort. *Comput. J.*, 5:10–15, 1962.

[21] Olav Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[22] Peter Kirschenhofer and Helmut Prodinger. Comparisons in Hoare's Find algorithm. *Combin. Probab. Comput.*, 7(1):111–120, 1998.

[23] Charles Knessl and Wojciech Szpankowski. Quicksort algorithm again revisited. *Discrete Math. Theor. Comput. Sci.*, 3(2):43–64 (electronic), 1999.

[24] Diether Knof. Struktur von fixpunkten aus prozessgleichungen. *http://www-computerlabor.math.uni-kiel.de/stochastik/knof/Texte/Dissertation.korrigiert.pdf*, 2007.

[25] Diether Knof and Uwe Roesler. The analysis of find or perpetuities on cadlag functions. *Discrete Mathematics and Theoretical Computer Science*, 2010.

[26] Donald E. Knuth. Mathematical analysis of algorithms. In *Information processing 71 (Proc. IFIP Congress, Ljubljana, 1971), Vol. 1: Foundations and systems*, pages 19–27. North-Holland, Amsterdam, 1972.

[27] Donald E. Knuth. *The art of computer programming. Volume 3.* Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1973. Sorting and searching, Addison-Wesley Series in Computer Science and Information Processing.

[28] B. Kodaj and T. F. Móri. On the number of comparisons in Hoare's algorithm "FIND". *Studia Sci. Math. Hungar.*, 33(1-3):185–207, 1997.

[29] Markus Kuba. On Quickselect, partial sorting and multiple Quickselect. *Inform. Process. Lett.*, 99(5):181–186, 2006.

[30] Hosam M. Mahmoud. *Evolution of random search trees.* Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Inc., New York, 1992. A Wiley-Interscience Publication.

[31] Hosam M. Mahmoud, Reza Modarres, and Robert T. Smythe. Analysis of QUICKSELECT: an algorithm for order statistics. *RAIRO Inform. Théor. Appl.*, 29(4):255–276, 1995.

[32] Conrado Martinez. Partial quicksort. *In Proceedings of the First ACM-SIAM Workshop on Analytic Algorithmics and Combinatorics*, 2004.

[33] Conrado Martinez and Uwe Roesler. Partial quicksort and quickpartitionsort. *Discrete Mathematics and Theoretical Computer Science*, (AofA'10):336–42, 2010.

[34] Ralph Neininger and Ludger Rüschendorf. Analysis of algorithms by the contraction method: additive and max-recursive sequences. In *Interacting stochastic systems*, pages 435–450. Springer, Berlin, 2005.

[35] Volkert Paulsen. The moments of FIND. *J. Appl. Probab.*, 34(4):1079–1082, 1997.

[36] S. T. Rachev and L. Rüschendorf. Probability metrics and recursive algorithms. *Adv. in Appl. Probab.*, 27(3):770–799, 1995.

[37] Mireille Régnier. A limiting distribution for quicksort. *RAIRO Inform. Théor. Appl.*, 23(3):335–343, 1989.

[38] U. Roesler and L. Rueschendorf. The contraction method for recursive algorithms. *Algorithmica*, 29(1-2):3–33, 2001. Average-case analysis of algorithms (Princeton, NJ, 1998).

[39] Uwe Roesler. A limit theorem for "Quicksort". *RAIRO Inform. Théor. Appl.*, 25(1):85–100, 1991.

[40] Uwe Roesler. A fixed point theorem for distributions. *Stochastic Process. Appl.*, 42(2):195–214, 1992.

[41] Uwe Roesler. The weighted branching process. In *Dynamics of complex and irregular systems (Bielefeld, 1991)*, Bielefeld Encount. Math. Phys., VIII, pages 154–165. World Sci. Publ., River Edge, NJ, 1993.

[42] Uwe Roesler. Quickselect revisited. *Journal of the Iranian Statistical Institute*, 3(2):271–296, 2004.

[43] U. Rösler. On the analysis of stochastic divide and conquer algorithms. *Algorithmica*, 29(1-2):238–261, 2001. Average-case analysis of algorithms (Princeton, NJ, 1998).

[44] U. Rosler, V. A. Topchiĭ, and V. A. Vatutin. Convergence conditions for branching processes with particles having weight. *Diskret. Mat.*, 12(1):7–23, 2000.

[45] Robert Sedgewick. The analysis of Quicksort programs. *Acta Informat.*, 7(4):327–355, 1976/77.

[46] Kok Hooi Tan and Petros Hadjicostas. Some properties of a limiting distribution in Quicksort. *Statist. Probab. Lett.*, 25(1):87–94, 1995.

Mahmoud Ragab

Personal Data

Name	**Mahmoud Ragab**
Date of birth	**January 23, 1977**
Place of birth	**Mansoura, Egypt**
Nationality	**Egyptian**
marital status	**Married**

Education

2008 **Ph.D. candidate at the mathematics department, Christian-Albrechts-Universität zu Kiel (CAU), Germany**.
Dissertation title: Partial Quicksort and weighted branching process

2006 **M. Sc. Degree in mathematical statistics, mathematics department, Faculty of science, Al Azhar University, Cairo, Egypt**.
Master title: Estimation of cross spectral density processes with irregularly observations

2000 **B. Sc. Degree in statistics and computer science, Faculty of science, Mansoura University, Mansoura, Egypt**.

1995 **General secondary education certificate, science branch, Nabaroh administration, Mansoura, Egypt**.

Experience

since 2008 **Member of "Arbeitsgruppe Stochastik" under the supervision of Prof. Dr. Uwe Roesler, mathematics department, CAU Kiel, Germany**.

2007 **Assistant Lecturer, mathematics department, Faculty of science, Al Azhar University, Cairo, Egypt**.

2006 **The International Computer Driving License (ICDL) issued by the UNESCO**.

2001 **Demonstrator, mathematics department, Faculty of science, Al Azhar University, Cairo, Egypt**.

Ludewig-Meyn-Straße 4 – 24098 Kiel, Germany
017652450242 • ☎ *(0431) 8804393* • ✉ *ragab@math.uni-kiel.de*

Declaration

I declare under penalty of perjury that I have made the present work, apart from the the supervisor‘s guidance the content and design of the essay is all my own work. The thesis has been neither published nor submitted and has not been submitted wholly or partially elsewhere as part of a doctoral degree to another examining body. Furthermore, I certify that the work has been prepared and originated subject to the rules of good scientific practice of the German Research Foundation.

Kiel, December 13, 2011

(Mahmoud Ragab)