The contraction method for recursive algorithms

Uwe Rösler
University of Kiel*

Ludger Rüschendorf
University of Freiburg†

Abstract

In this paper we give an introduction to the analysis of algorithms by the contraction method. By means of this method several interesting classes of recursions can be analyzed as particular cases of our general framework. We introduce the main steps of this technique which is based on contraction properties of the algorithm with respect to suitable probability metrics. Typically the limiting distribution is characterized as a fixed point of a limiting operator on the class of probability distributions. We explain this method in the context of several "divide and conquer" algorithms. In the second part of the paper we introduce a new quite general model for branching dynamical systems and explain that the contraction method can be applied in this model. This model includes many classical examples of random trees and gives a general frame for further applications.

1 Introduction

In its basic form the contraction method for the asymptotic analysis of a recursive algorithm uses the following steps. First typically by studying first moments one has to find the correct normalization of the algorithm. The normalized version again is given in a recursive form. Then often one can guess a fixed point equation for the limiting equation of the algorithm. This limiting fixed point equation is formally described by an equation of the form $K\mu = \mu$ for a limiting operator K on the set of all probability distributions. In the next step one has to prove existence of a solution of this fixed point equation. This is done by choosing a suitable probability metric on the space of probability measures such that K is a contraction map with respect to this metric and the Banach Fixed Point Theorem is applicable. In the final step one has to prove that the algorithm converges (in distribution) to this fixed point. This is done by introducing accompanying sequences of distributions which resemble the algorithm and also reflect the structure of the limiting equation.

The first example where this method has been applied was the Quicksort algorithm in a paper by Rösler (1991). In its general form this method has been

*Ludewig-Meyn-Str. 4, D-24118 Kiel; e-mail: roesler@math.uni-kiel.de
†Eckerstr. 1, D-79104 Freiburg; e-mail: ruschen@pollux.mathematik.uni-freiburg.de

developed independently in Rösler (1992) and Rachev, Rüschendorf (1995)[1]. In the meantime many algorithms have been analyzed by this method (see the reference list) and it turned out that this probabilistic method very well supplements the combinatorial and analytic methods for the analysis.

The contraction method is natural from a mathematical point of view. A classical subject of probability theory is the central limit theorem for sequences of sums of independent random variables. These sum sequences have an easy recursive structure. Probability metrics and contraction techniques have been introduced in particular in the Russian school of probability to prove the central limit theorem. From this context many useful probability metrics, its regularity properties and relations between them have been studied intensively (Rachev (1991), Rachev, Rüschendorf (1995), Zolotarev (1997)). These regularity properties are also useful for the analysis of general recursive algorithms.

In the first part of this paper we explain the basic steps of the analysis of algorithms by the contraction method by several examples. We also explain some essential modifications of the basic method like the problem of choice of the probability metric and the case of contraction coefficients converging to one, and we remark on some techniques to prove convergence to the limiting equation. In particular we give a convergence result for a general version of divide and conquer algorithms and for a multivariate branching type recursion (in the setting of weighted branching processes and branching dynamical systems).

In the second part of this paper we introduce a quite general branching dynamical system on a tree which serves as a basic reference model for random trees and for the limiting equations for many recursive algorithms. The analysis of this branching dynamical system can be given by the contraction method. For concrete applications a lot of detailed work has to be invested, however and the analysis of this system is a challenge for the future.

2 Analysis of divide and conquer algorithms

Let $X_n, n \in \mathrm{N}$, denote some performance costs like time or resources of a stochastic divide and conquer algorithm. The random variable (r.v.) X_n respectively the distribution of X_n is recursively given by

$$X_n \stackrel{\mathcal{D}}{=} \sum_{i \in \mathrm{N}} T_{n,i} X^{(i)}_{Z_{n,i}} + C_n \tag{2.1}$$

for $n \geq n_0$. Here $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution, n_0 is a natural number, and the r.v.'s (C_n, Z_n, T_n), $X_j^{(i)} \in \mathrm{N}$, $0 \leq j \leq n-1$, are independent. We use the notation $Z_n = (Z_{n,1}, Z_{n,2}, \ldots)$, $T_n = (T_{n,1}, T_{n,2}, \ldots)$. $Z_{n,i}$ takes values in $\{0, \ldots, n-1\}$. The distribution of (C_n, Z_n, T_n) is given. The distribution of X_j, $0 \leq j \leq n_0 - 1$ is given in advance, and the distribution of $X_j^{(i)}$ is that of X_j and given recursively. For notational reasons we set $X_0 \equiv 0$.

[1]This paper was submitted in 1990 but published only in 1995.

The analysis of X_n is done in a series of steps.

First step: The Expectation. *Find the expectation* $a_n := EX_n$. The expectation satisfies the recursive equation, $n_0 \leq n$,

$$a_n = \sum_{i \in \mathbb{N}} ET_{n,i} a_{Z_{n,i}} + c_n \tag{2.2}$$

where $c_n := EC_n$. The standard form of the equation is

$$a_n = \sum_{j=0}^{n-1} \nu_n(j) a_j + c_n. \tag{2.3}$$

Here ν_n is some signed measure on $\{0, \ldots, n-1\}$, not necessarily a probability measure.

For our purpose we are interested in the asymptotic behaviour of a_n in n. Standard examples, unless a_n converges to 0, are are a_n of order (const) $n \ln n +$(const) $n + o(n)$ or (const) $n + o(n)$ or of order $\ln n$. There are many methods and techniques available, such as a proof by an Ansatz (guess), generating functions and some algebraic equation, difference equation and differential equations in the limit, probabilistic interpretation via Markov or renewal processes by the Cauchy or Dirichlet problem, etc. A general method from renewal theory for these equations arising in the analysis of algorithms is given in Rösler (1998b). (The technical details are too involved for this paper.)

If ν_n has only mass at the point $n-1$ then the equations

$$a_n = \nu_n(n-1) a_{n-1} + c_n \tag{2.4}$$

are usually easier to solve, e.g. by direct recursion, or difference equations or in the limit differential equations. In the following we assume sufficient knowledge on the asymptotics of a_n, if needed.

Second Step: The Normalization. *Guess the right normalization, such that the normalized version* Y_n *of* X_n *converges to a limit* W.

Consider

$$Y_n := \frac{X_n - a_n}{b_n}$$

with some normalization $b_n > 0$. We can not give exact general rules for an appropriate choice of b_n. One indication is to choose b_n as small as possible, such that the normalized version C_n^* of C_n defined below converges to a limit and, if possible, to a non degenerate limit. If the variance is finite and can be calculated then choose $b_n = \sqrt{\operatorname{Var} X_n}$. Typical examples are b_n of the order n or $\sqrt{n}$ or $\ln n$.

The basic equation (2.1) becomes

$$Y_n \stackrel{\mathcal{D}}{=} \sum_{i \in \mathbb{N}} T_{n,i} \frac{b_{Z_{n,i}}}{b_n} Y^{(i)}_{Z_{n,i}} + \sum_{i \in \mathbb{N}} \frac{T_{n,i} a_{Z_{n,i}} - ET_{n,i} a_{Z_{n,i}}}{b_n} + C_n - EC_n. \tag{2.5}$$

We call this equation the normalized basic equation. This equation (2.5) has the same structure as the basic equation (2.1) with $T_{n,i}, C_n$ replaced by

$$T^*_{n,i} := T_{n,i}\frac{b_{Z_{n,i}}}{b_n} \qquad C^*_n := \sum_{i\in\mathrm{N}} \frac{T_{n,i}a_{Z_{n,i}} - ET_{n,i}a_{Z_{n,i}}}{b_n} + C_n - EC_n.$$

The main difference from equation (2.1) is that the expectation of all Y_i are zero.

Third step: The Limiting distribution. *Guess the limiting distribution W of Y_n.*

In many situations the limiting distribution can be guessed as the fixed point of an operator K. Consider the normalized equation (2.5). If Y_n converges to W, $Z_{n,i}$ converges to ∞ and in some nice sense (C^*_n, T^*_n) converges to (C, T) then we obtain a limiting equation by

$$\begin{array}{ccccccc} Y_n & \stackrel{\mathcal{D}}{=} & \sum\limits_{i\in\mathrm{N}} & T^*_{n,i} & Y^{(i)}_{Z_{n,i}} & + & C^*_n. \\ \downarrow & & & \downarrow & \downarrow & & \downarrow \\ W & \stackrel{\mathcal{D}}{=} & \sum\limits_{i\in\mathrm{N}} & T_i & W_i & + & C \end{array}$$

Here $(C, T), W_i, i \in \mathrm{N}$, are independent and all W_i have the same distribution as W. The distribution of W should be a solution of the above fixed point equation. Sometimes this fixed point equation characterizes the limiting distribution and gives a good hint.

In order to highlight this last equation, we would like to formulate this in terms of distributions and in operator setting. Define the operator K on a subset M of probability measures to itself via

$$K(\mu) := \mathcal{L}\Big(\sum_{i\in\mathrm{N}} T_i Y_i + C\Big). \tag{2.6}$$

Here $(C, T), Y_i,\ i \in \mathrm{N}$, are independent, Y_i has the distribution μ, and the distribution of (C, T) is from the limiting equation. We call K the *limiting operator*. A fixed point of K is a probability measure ν satisfying $K(\nu) = \nu$. The description via r.v.'s,

$$W \stackrel{\mathcal{D}}{=} \sum_{i\in\mathrm{N}} T_i W_i + C, \tag{2.7}$$

is equivalent. The distribution of W is a fixed point. We call equation (2.7) the fixed point equation and a solution a fixed point (of the limiting operator K). In Section 4 we give hints for solutions of the fixed point equation.

Fourth step: The Contraction method. *This method is a tool to prove the convergence of the normalized sequence Y_n to W, the basic idea is a contraction argument on the space of distributions.*

Define the maps $K_n,\ n \geq n_0$, on M^n to M defined via

$$K_n(\mu_0, \dots, \mu_{n-1}) := \mathcal{L}\left(\sum_{i \in \mathrm{N}} T^*_{n,i} Y^{(i)}_{Z_{n,i}} + C^*_n\right). \tag{2.8}$$

Here $(C^*_n, Z_n, T^*_n), Y^{(i)}_j,\ 0 \leq j \leq n,\ i \in \mathrm{N}$, are independent. $Y^{(i)}_j$ is distributed as μ_j and the distribution of (C^*_n, Z_n, T^*_n) is given as above.

Starting with some probability measures $\mu_0, \dots, \mu_{n_0-1}$ we obtain a sequence

$$\begin{aligned}
\mu_{n_0} &= K_{n_0}(\mu_0, \dots, \mu_{n_0-1}) \\
\mu_{n_0+1} &= K_{n_0+1}(\mu_0, \dots, \mu_{n_0}) \\
\mu_{n_0+2} &= K_{n_0+2}(\mu_0, \dots, \mu_{n_0+1}) \\
&\ \ \vdots \\
\mu_n &= K_n(\mu_0, \dots, \mu_{n-1})
\end{aligned}$$

The distribution μ_n is the distribution of Y_n as before if $\mu_0, \dots, \mu_{n_0-1}$ are the distributions of the $X_0, \dots, X_{n_0-1}$.

On a subspace M of probability measures consider a metric d. (For simplicity it is common to use the notation $d(X, Y) := d(\mathcal{L}(X), \mathcal{L}(Y))$. The reader is advised to distinguish clearly between metrics on the space of distributions and on the space of r.v.'s.) Usually the metrics considered have some desirable properties.

Let ν_n be a sequence of distributions converging to a known distribution ν. Usually ν_n satisfies a recurrence equation similar to the μ_n, formulated in terms of an operator L_n. We call the sequence (ν_n) the accompanying sequence. The contraction method is a method to show that the distance between μ_n and ν_n converges to 0 (after a good guess of L_n and the sequence ν_n). This would prove that the distribution of Y_n converges to the distribution ν. Often the sequence ν_n is the sequence identical to the fixed point ν of the limiting operator.

Since we deal with linear structured operators K some metrics with special properties can be shown to perform better than others. In an abstract setting see Zolotarev (1997) and Rachev (1991) for these properties and an account and Rachev, Rüschendorf (1995) for an explicit account in the context of algorithms. Instead of listing these properties it seems worthwhile to give some intuition of the structure, what the metrics 'really' do. In an example we will show the essentials. The main 'trick' is the choice of the r.v.'s close (= coupled) together.

Example: Contraction of affine functions. We choose the minimal $\ell_p, 1 \leq p < \infty$ metric (also called Mallows metric (Bickel, Freedman (1981)):

$$\ell_p(\mu, \nu) = \inf\{\|X - Y\|_p\}.$$

The infimum is over all r.v.'s X with distribution μ and all r.v.'s Y with distribution ν on some non-atomic probability space. The symbol $\|.\|_p$ denotes the usual L_p-norm. The common notation for the metric is ℓ_p. The infimum is attained and given by

$$\ell_p(\mu, \nu) = \|F_\mu^{-1}(U) - F_\nu^{-1}(U)\|_p$$

(DallAglio (1956)). Here U is uniformly distributed on $[0,1]$ and F_μ is the distribution function of μ.

The metric ℓ_p has the following properties

$$\begin{aligned} \ell_p(aX, aY) &\leq |a|\ell_p(X,Y), \\ \ell_p\Big(\sum_i X_i, \sum_i Y_i\Big) &\leq \sum_i \ell_p(X_i, Y_i), \\ \ell_p(AX, AY) &\leq \|A\|_p \ell_p(X,Y). \end{aligned}$$

where $a \in \mathrm{R}$, $(X_i, Y_i), i \in \mathrm{N}$, are independent r.v.'s, and A is independent of (X,Y).

Let M_p be the space of all distributions μ with finite absolute p-th moment, $\int |x|^p \mu(dx) < \infty$. The space space (M_p, ℓ_p) is a complete separable metric space (Bickel, Freedman (1981)).

We choose as example the following weighted branching process, (see also Section 4). Let K be the operator corresponding to $K(\mu) = \mathcal{L}(\sum_i T_i Y_i + C)$, where (Y_i) are independent with distribution μ, $Y_i \stackrel{d}{=} \mu$. Let (C_n, T_n) have the same distribution as (C,T) and let $Z_{n,i} \equiv n-1$. Notice the simplification

$$K_n(\mu_0, \ldots, \mu_{n-1}) = \mu_n = K(\mu_{n-1}) = \ldots = K^n(\mu_0).$$

Here is the contraction lemma, which is indeed remarkable with its beauty and simplicity. Let $\mu, \nu \in M_2$ be two probability measures with $\int x\mu(dx) = 0 = \int x\nu(dx)$.

Lemma 1 (The Contraction Lemma)

$$\ell_2(K(\mu), K(\nu)) \leq \sqrt{E\Big(\sum_j T_j^2\Big)}\, \ell_2(\mu, \nu).$$

Proof: Choose independent r.v.'s $(C,T), U_i, i \in \mathrm{N}$ on some probability space. The U_i have uniform distribution on $[0,1]$. Define $X_i := F_\mu^{-1}(U_i), Y_i = F_\nu^{-1}(U_i)$. The r.v. $\sum_i T_i X_i + C$ has distribution $K(\mu)$ and $\sum_i T_i Y_i + C$ has distribution $K(\nu)$. Notice we take the *same* (C,T) which is important. Then

$$\begin{aligned} \ell_2^2(K(\mu), K(\nu)) &\leq \Big\| \sum_i T_i X_i + C - \sum_i T_i Y_i - C \Big\|_2^2 \\ &= E\Big(\sum_i T_i (X_i - Y_i)\Big)^2 \\ &= \sum_i E(T_i(X_i - Y_i))^2 \\ &= \sum_i E(T_i^2) E(X_i - Y_i)^2 \\ &= \sum_i E(T_i^2) \ell_2^2(\mu, \nu) \end{aligned}$$

□

Next Banach's Fixed Point Theorem provides existence and uniqueness of solutions of the fixed point equation (Rösler (1992,1993), Rachev, Rüschendorf (1995), Cramer, Rüschendorf (1996b)).

Theorem 2 (Banach's Fixed Point Theorem)
Let (E, d) be a complete metric space. Let $K : E \mapsto E$ be a strict contraction with contraction constant $k < 1$. Then the sequence $K^n(e)$, $n \in \mathrm{N}$, converges exponentially fast with the order k^n to the unique fixed point of K for every $e \in E$.

Theorem 3 *Assume $EC^2 < \infty$, $EC = 0$, $\sum_i ET_i^2 < 1$. Then the sequence $K^n(\mu)$, $n \in \mathrm{N}$, $\mu \in M^0$ converges exponentially fast in ℓ_2-metric to the unique fixed point of K in M^0.*

We used M^0 as the set of $\mu \in M_2$ satisfying $\int x\mu(dx) = 0$. This result provides a characterization of the limiting distribution W.

Remark: Equation (2.1) can be rewritten in the form:

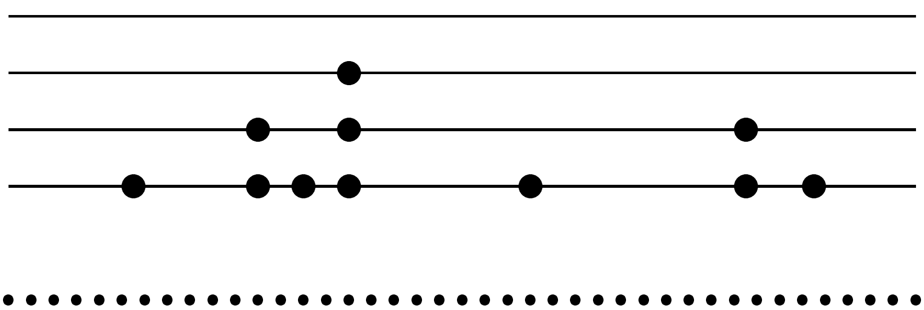
$$X_n \stackrel{\mathcal{D}}{=} \sum_{i \in \mathrm{N}} T_{n,i} X^{(i)}_{Z_{n,i}} + C_n \stackrel{\mathcal{D}}{=} \sum_{i,j} \overline{T}^{(i)}_{n,j} X^{(i)}_j + \overline{C}_n.$$

Here $\left(\overline{T}^{(i)}_{n,j}, \overline{C}_n\right)$, $X^{(i)}_j$, $i \in \mathrm{N}, j \leq n-1$ are independent. The easiest way to see that is to condition the original recursive equation to $(T_{n,i}, Z_{n,i}, C_n)$. For fixed values, by the independence assumptions, the sum is a sum of independent r.v.'s. In the latter form there is no random indexing.

3 Quicksort Type Recursions

In this section we present examples of divide and conquer algorithms whose analysis can be given by a similar scheme. The prototype is Quicksort (Rösler (1991)).

Quicksort: Quicksort (Hoare (1961)) is a sorting algorithms for numbers (or objects with a strict order). In its randomized form, as already described by Hoare (1962), the recursive step consists of choosing at random with a uniform distribution a number (object) and forming three lists: the list of numbers strictly smaller than the chosen one, the list of the chosen number, and the list of strictly larger numbers. After some iteration the algorithm terminates and provides an ordered list. In symbolic form



(The lines represent the lists in different stages during the sorting. The $\bullet$ gives the position of the chosen element after comparisons. All elements left of the $\bullet$ are strictly smaller, all elements right to the $\bullet$ strictly larger. Finally the procedure provides an ordered list.)

Let X_n denote the number of comparisons made by the algorithm Quicksort in order to sort n given different numbers. The basic equation is

$$X_n \stackrel{\mathcal{D}}{=} X_{Z_n} + \overline{X}_{n-Z_n-1} + n - 1$$

for $n \geq 1$. Here $Z_n, X_., \overline{X}_.$ are independent r.v.'s. The size of the subset of smaller elements Z_n has a uniform distribution on $\{0, \ldots, n-1\}$. The X_j have the same distribution as the $\overline{X}_j$ and are given recursively.

The expectation satisfies the relation

$$a_n = \frac{2}{n} \sum_{i=0}^{n-1} a_i + n - 1.$$

This reduces to

$$\frac{a_n}{n+1} = \frac{a_{n-1}}{n} + \frac{2(n-1)}{n(n+1)} = \ldots = 2 \sum_{k=1}^{n+1} \frac{1}{k} + \frac{2}{n+1} - 4.$$

Therefore a_n has the asymptotic expansion

$$a_n = 2n \ln n + n(2\gamma - 4) + 2 \ln n + 2\gamma + 1 + o(n^{-1} \ln n)$$

with $\gamma = 0.57721\ldots$ the Euler constant.

We start with a "guess", $b_n = n$, of the normalization which is the order of the standard variation. The choice of this normalization $b_n = n$ yields convergence of the remainder term C_n to C. Our guess for the limiting distribution W, following the general suggestion, is the solution of the fixed point equation

$$W \stackrel{\mathcal{D}}{=} \tau W + (1-\tau)\overline{W} + C. \tag{3.1}$$

Here $\tau, W, \overline{W}$ are independent. τ has a uniform distribution, W and $\overline{W}$ have the same distribution and C is a function of τ given by

$$C = 1 + 2\tau \ln \tau + 2(1-\tau) \ln(1-\tau).$$

The existence and uniqueness of W follows by the contraction method applied to the weighted branching process (see Sections 2, 5). This is done with respect to the minimal ℓ_p metrics in Rösler (1991). In the context of weighted branching processes a related contraction argument is given in Rösler (1992) and Rachev, Rüschendorf (1995).

Consequences of the fixed point equation: Hennequin (1989) considered moments of Quicksort and conjectured an explicit form proved later on in his dissertation. He used combinatorial arguments but did not make use of the fixed point equation. The fixed point equation provides a recursive formula for the moments. It should be possible, even simpler, to derive the moments directly by the fixed point equation. See the paper Dobrow, Fill (1998) on a similar equation. As a consequence of the convergence results for X_n one obtains that the variance is of order n^2.

McDiarmid, Hayward (1996) gave the right exponential bounds (moments). Again exponential results (Rösler (1992)) should be easier to obtain via the fixed point equation. Knessl, Szpankowski (1998) took this view and analyzed the corresponding generating function. They obtained asymptotic expansions, e.g. on the tail of the distribution.

As a consequence of the fixed point equation one obtains that the Quicksort distribution has a Lebesgue density (Tan, Hadjicostas (1995)). The best known approximation of the Quicksort distribution is by a lognormal (Cramer (1996)). The existence of a limiting distribution was proved indepently by Régnier (1989) (martingale argument) and Rösler (1991) (contraction argument).

The Median version of Quicksort: In this version of Quicksort choose $2t+1$ elements and use the median for the comparison. This algorithm was suggested by Hoare (1962), the analysis is in Rösler (1998b). Quicksort is the special case $t = 0$.

Let X_n denote the number of comparisons performed by the $2t+1$ median version of Quicksort in order to sort n given different numbers. The basic equation is

$$X_n \stackrel{\mathcal{D}}{=} X_{Z_n} + \overline{X}_{n-Z_n-1} + C_n$$

for $n \geq 2t+1$. Here $Z_n, X_\cdot, \overline{X}_\cdot$ are independent r.v.'s. The Z_n r.v.'s have a distribution on $\{t, \ldots, n-t-1\}$. The distribution of Z_n is

$$P(Z_n = i) = \frac{\binom{i}{t}\binom{n-i-1}{t}}{\binom{n}{2t+1}}.$$

The X_j have the same distribution as the $\overline{X}_j$ r.v.'s and are given recursively. C_n involves also the costs for finding the median. We drop a description of the algorithm for list sizes strictly smaller than $n_0 = 2t+1$.

The recursive equation for the expectation (2.3) is

$$a_n = \sum_{i=0}^{n-1} 2P(Z_n = i)a_i + c_n.$$

The asymptotics of a_n are

$$a_n = e_t n \ln n + f_t n + o(n).$$

f_t is an unknown constant and e_t is given by

$$e_t := \frac{-1}{\int (x \ln x + (1-x)\ln(1-x))\, g_t(x) dx},$$

(Bruhn (1996), Rösler (1998b), Sedgewick (1980)). The distribution of $\frac{Z_n}{n}$ converges 'nicely' to the distribution with Lebesgue density g_t, where g_t is the density of the $t+1$-st order statistic of $2t+1$ independent r.v.'s with a uniform distribution. This is a beta(t+1,t+1) distribution on $[0,1]$. See also van Emden (1970).

As in Quicksort we start with the normalizing factor $b_n = n$, which will turn out to be the right one. Our guess for the limiting distribution W, following the general suggestion, is the solution of the fixed point equation

$$W \stackrel{\mathcal{D}}{=} \tau W + (1-\tau)\overline{W} + C. \tag{3.2}$$

Here $\tau, W, \overline{W}$ are independent. τ has density g_t as above, W and $\overline{W}$ have the same distribution. C is given by

$$C = e_t(\tau \ln \tau + (1-\tau)\ln(1-\tau)) + 1.$$

Existence and uniqueness of a solution of (3.2) is proved as suggested in Sections 2, 5 using the contraction argument for the minimal ℓ_2-metric. Finally one has to prove the convergence of the normalized sequence Y_n to W (Rösler (1998b)).

Random recursive trees: Let X_n denote the total path-length of a random recursive tree, (Dobrow, Fill (1998), Mahmoud (1991), Smythe, Mahmoud (1995)). The basic equation is

$$X_n \stackrel{\mathcal{D}}{=} X_{Z_n} + \overline{X}_{n-Z_n} + Z_n,$$

$n \geq n_0 = 2$, with Z_n uniformly distributed on $\{1,\ldots,n-1\}$. The recursive equation for the expectation is

$$a_n = \sum_{i=1}^{n-1} \frac{2}{n-1} a_i + \frac{n}{2}.$$

This is rewritten in the form

$$\frac{a_n}{n} = \frac{a_{n-1}}{n-1} + \frac{1}{n} = \ldots = \frac{a_1}{1} + \sum_{i=2}^{n} \frac{1}{i}.$$

With the normalization $b_n = n$ the limiting fixed point equation becomes

$$W \stackrel{\mathcal{D}}{=} \tau W + (1-\tau)\overline{W} + C. \tag{3.3}$$

Here $\tau, W, \overline{W}$ are independent, τ has a uniform distribution, and W and $\overline{W}$ have the same distribution. C is given as a function of τ,

$$C = \tau + \tau \ln \tau + (1-\tau)\ln(1-\tau).$$

Dobrow, Fill (1998) give explicit structural formulas for the moments and prove convergence of X_n to W.

Quad Trees: Quad trees are data structures to store and retrieve data from some multidimensional data space that extends the binary search tree for one-dimensional data. In a standard model with independent identically distributed data points the total path length of a quad tree satisfies the recursion

$$X_n \stackrel{\mathcal{D}}{=} \sum_{k=0}^{2^d-1} X^{(k)}_{Z_{n,k}} + n \tag{3.4}$$

where $\left(X_i^{(k)}\right)$ are independent copies of X_i, and $Z_{n,k}$ are independent conditionally multinomial distributed. The condition is given by the first key U to be inserted and conditionally on $U = u$ the parameter of the multinomial is given by $\langle u \rangle$ the vector of volumes of the quadrants in $[0,1]^d$ generated by U when cutting parallel to the axes (see Mahmoud (1992)).

Flajolet, Labelle, Laforest, Salvy (1995) proved that

$$EX_n = \frac{2}{d} n \ln n + \mu_d n + o(n). \tag{3.5}$$

Using contraction arguments, Neininger, Rüschendorf (1998) showed that the normalized version Y_n of X_n

$$Y_n = \frac{X_n - EX_n}{n} \tag{3.6}$$

converges to the unique solution of the fixed point equation

$$Y \stackrel{d}{=} \sum_{k=0}^{2^d-1} \langle U \rangle_k Y^{(k)} + C(\langle U \rangle) \tag{3.7}$$

where $C(x) = 1 + \frac{2}{d} \sum_{i=0}^{2^d-1} x_i \ln x_i$ is the entropy functional. Furthermore in the same work it is shown that convergence of all moments and of the Laplace-transform holds. In particular, the asymptotics of first order of the variance of the internal path length is given by

$$\mathrm{Var}(X_n) \sim v_d n^2, \tag{3.8}$$

where v_d is given by

$$v_d = \frac{21 - 2\pi^2}{9d} \, \frac{1}{1 - \left(\frac{2}{3}\right)^d}, \tag{3.9}$$

in particular $v_1 = 0.4202\ldots$, $v_2 = 0.1260\ldots$, $v_3 = 0.0663\ldots$

Quickselect or Find: Mahmoud, Modarres, Smythe (1995) considered the asymptotics of Quickselect (= Find) for a random order statistic M_n which is discrete uniform on $1, \ldots, n$. Let $C_n^{(M_n)}$ denote the number of comparisons between list elements by Quickselect. Then

$$EC_n^{(M_n)} = 3n - 8H_n + 13 - \frac{8H_n}{n},$$

and

$$\mathrm{Var}\, C_n^{(M_n)} \sim n^2. \tag{3.10}$$

The normalized sequence

$$Y_n = \frac{C_n^{(M_n)} - EC_n^{(M_n)}}{n} \tag{3.11}$$

satisfies the recursion

$$Y_n \stackrel{d}{=} 1_{\{Z_n > M_n\}} \frac{Z_n - 1}{n} (Y_{Z_n - 1} + 3) + 1_{\{Z_n < M_n\}} \frac{n - Z_n}{n} (Y_{n - Z_n} + 3) + \frac{n-1}{n} - 3 + o(1), \tag{3.12}$$

where Z_n is the position to which the pivot moves. Z_n has a discrete uniform distribution on $1, \dots, n$. In Mahmoud, Modarres, Smythe (1995), it is proved (using a contraction argument) that Y_n converges to the solution Y of the fixed point equation,

$$Y \stackrel{d}{=} X(Y+3) - 2, \tag{3.13}$$

where X has density $f(x) = 2x$, $0 < x < 1$. The characteristic function of Y has an explicit representation

$$\varphi_Y(t) = \exp\left(2 \int_0^t \frac{e^{iu} - 1 - iu}{u} \, du\right). \tag{3.14}$$

An asymptotic convergence result, in the case of finding the k-th order statistic for a fixed number k, is also obtained by the contraction method. Note that $\frac{k}{n} \to 0$. The analysis of the general case where $\frac{k}{n} \to t$ needs a different framework and will be analyzed in Section 9.

Mergesort: Mergesort is a sorting algorithm. The basic step is to merge two lists, both already sorted in increasing order, of length r and s. We consider here only the linear merging, others are similar (Mahmoud (1998)). Compare the smallest element of the two lists with each other, remove the smaller from the list and add it to the merged list. Repeat this recursively, until one of the lists is empty. The list of stored elements in that order and the remaining elements are ordered.

Let X_n denote the random number to sort a list of length n different numbers. X_n satisfies a recursive equation

$$X_n \stackrel{\mathcal{D}}{=} X_{Z_n} + \overline{X}_{n - Z_n} + C_n(Z_n).$$

Here $Z_n, C_n(i), X_i, \overline{X}_i,\ i \le n-1$ are independent r.v.'s. The distribution of $C_n(i)$ is given by

$$P(C_n(i) = m) = \frac{\binom{m-1}{i-1} + \binom{m-1}{n-i-1}}{\binom{n}{i}}$$

for $\inf\{i, n-i\} \le m \le n-1$. ($C_n(i)$ denotes the merging cost for lists of length i and $n-i$.) Z_n takes values in $\{1, \dots, n-1\}$.

Common choices for Z_n are

- $Z_n = \lfloor \frac{n}{2} \rfloor$ (half-half rule) This is the top-down mergesort analyzed by Flajolet, Golin (1994) and Cramer (1997c).

- $Z_n = 2^{\lceil \log_2 \frac{n}{2} \rceil}$ (max power-of-two-rule) This is bottom up mergesort analyzed by Panny, Prodinger (1995) .
- $Z_n = 2^{\lfloor \log_2 \frac{2n}{3} \rfloor}$ (balanced power-of-two rule) This is queue mergesort analyzed by Golin, Sedgewick (1993) .

See the paper by Chen, Hwang, Chen (1998) for a comparison of the three versions and the analytical techniques. Cramer (1995) and Schimmler (1997) used the contraction method to obtain similar results.

In all the above cases the splitting Z_n is deterministic. Therefore, applying the recursive formula, we obtain X_n as a sum of independent r.v.s. For these objects (see e.g. Gnedenko, Kolmogorov (1949)) strong (and equivalent) results for the convergence to a normal distribution are known. Both the Lyapunov condition and the Lindeberg condition are easy to establish. This reduces the convergence problem to first and second order (or higher) moment considerations, which are a lot easier.

4 Weighted Branching Processes

In this and the following section we introduce general weighted branching processes. They provide a suitable framework for fixed point equations as in (2.7) and their solutions. The contraction method is a major tool to analyze these structures.

Many recursive structures as random recursive trees and splitting type algorithms are in this setting. All contraction arguments and examples given here can be seen in the setting of weighted branching processes (with slight modifications). There is a forward and a backward recursive description of weighted branching processes with two different perspectives. The backward view leads to the existence and uniqueness of solutions of the fixed point equation in the general setting of weighted branching process.

Trees: We consider the graph of a tree with countable many branches. The set of vertices V is $\bigcup_{n=0}^{\infty} \mathrm{N}^n$. N denotes the natural numbers $1, 2, 3, \ldots$ and N^0 is the set of the empty set $\{\emptyset\}$ by convention. The name of a vertex $v = (v_1, v_2, \ldots, v_n) = v_1 v_2 \ldots v_n \in \mathrm{N}^n$ corresponds to the path starting at the root $\emptyset$ to the vertex. We suppress if possible the root $\emptyset$. For $m \le |v|$ we use the restriction $v_{|m} = v_1 \ldots v_m$.

For v as above and and $w = w_1 \ldots w_m \in V$ we use $vw = v_1 \ldots v_n w_1 \ldots w_m$. The edges of the tree are the elements (v, vi). We will not stress that.

Other names for a vertex, depending on the context, are *knot*, *individual*, *particle* or *mother*. $n = |v|$ is the generation of the individual. vi is a *child* or *offspring* of v and v is the *mother* of vi. vw is a *descendent* of v and v is an *ancestor* of vw.

Many authors use different index systems for various kinds of trees, e.g. binary trees. Formally we can embed smaller trees via a function $\varphi : V \mapsto \{0, 1\}$ satisfying: If v is an ancestor of w and $\varphi(v) = 0$ then $\varphi(w) = 0$.

We find it easiest to think of v as alive if $\varphi(v) \neq 0$ and dead if $\varphi(v) = 0$. Then the alive individuals form the desired tree. We return to this construction via the 'grave', see later on.

Weighted branching processes: Let $(\Omega, \mathcal{F}, P)$ be a probability space, which is always assumed to be rich enough. Let $T(v) : \Omega \mapsto \mathrm{R}^{\mathrm{N}}$, $v \in V$, be independent identically distributed random variables. We use also $T(v) = (T_1(v), T_2(v), \ldots)$ with $T_i(v) : \Omega \mapsto \mathrm{R}$ the i-th projection. We drop the index v if superfluous. Notice that we allow arbitrary dependence of the components $T_i(v)$, $i \in \mathrm{N}$, for fixed v. The families are independent, but within a family the offspring of the mother may depend on each other. A process $((\Omega, \mathcal{F}, P), V, T)$ as above is called a *weighted branching process.*

Define recursively the *weight* or *length* or *type* or *state* or *layer* $\ell(v) : \Omega \to \mathrm{R}$ of the vertex v by

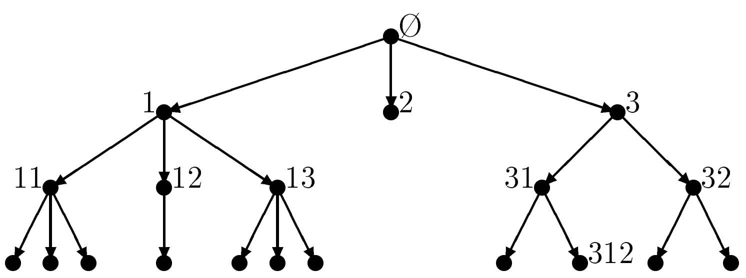
$$\ell(vj) := T_j(v)\ell(v) \tag{4.1}$$

$j \in \mathrm{N}$. The starting r.v. $\ell(\emptyset)$ is independent of the T. (and plays the role of a starting distribution like in Markov chains). For simplicity we take $\ell(\emptyset) \equiv 1$. The weight $\ell(v)$ has a product representation

$$\begin{aligned}
\ell(v) &= T_{v_n}(v_1 v_2 \ldots v_{n-1}) T_{v_{n-1}}(v_1 v_2 \ldots v_{n-2}) \ldots T_{v_1}(\emptyset)\ \ell(\emptyset) \\
&= T_{v_n}(v_{|n-1}) T_{v_{n-1}}(v_{|n-2}) \ldots T_{v_1}(\emptyset)\ \ell(\emptyset)
\end{aligned}$$

We call an individual v *observable* or *alive*, if $\ell(v) \neq 0$ and *non-observable* or *dead* if $\ell(v) = 0$. Notice that any offspring of a non-observable individual is not observable.

The following picture shows all living individuals in a realization.



Objects of interest: The n-th generation consists of a collection $\ell(v)$, $|v| = n$ of weights. A handy description is via discrete measures

$$\xi_n = \sum_{|v|=n} \delta_{\ell(v)}.$$

Here δ_x denotes a point measure on $x \in \mathrm{R}$. These objects are also used in branching process theory.

$$\xi_n := \sum_{|v|=n} \delta_{\ell(v)\ell(\emptyset)}$$

$n \in \mathrm{N}$, is called a branching process with type space the real line (Athreya, Ney (1972), Jagers (1974)) .

Typical questions (for divide and conquer algorithms) concern the size

$$Z_n := \int x\xi_n(dx) = \sum_{|v|=n} \ell(v)$$

of the n-th generation or

$$R_n := \sum_{|v|\leq n} \ell(v)C(v)$$

in the first n generations. Here $(T(v), C(v))$, $v \in V$, are i.i.d. r.v.'s. (Some interpretation in the algorithmic examples is possible, like $l(v)C(v)$ as cost depending on the size and R_n as total costs up to the n-th generation.)

For some applications one is more interested in the maximal (or minimal) height of a tree

$$\inf\{n : \forall |v| = n,\ \ell(v) \leq a\}$$

for a suitable constant a.

Let us give some examples in order to show the variety of included processes. Weighted branching processes have a branching structure and an iterative structure. We point this out by the two 'extreme' settings, one with no branching at all and one with only two states, dead or alive.

No branching: In this setting every mother has exactly one (living) child, let us say the first child is alive. The map T will be

$$T(v) = (T_1(v), 0, 0, 0).$$

If we start with one individual then in the n-th generation there is exactly one child alive, the child with name $v = (1, \ldots, 1)$, $|v| = n$. If we use $\Psi_i = T_1(\underbrace{1, \ldots, 1}_{i \text{ times}})$ for a moment, then

$$\ell(v) = \Psi_{n-1}\Psi_{n-2}\ldots\Psi_0$$

is the product of n independent identically distributed random variables (i.i.d. r.v.s). These are well studied objects.

Trivial weights: In this setting we allow only two possible states, alive ($= 1$) and dead ($= 0$). The living particles form in a natural way a random tree.

In the n-th generation live

$$Z_n = \sum_{|v|=n} \ell(v)$$

individuals. This process Z_n is known as a Galton-Watson process Athreya, Ney (1972), Asmussen, Hering (1983)). More precisely, commonly a Galton-Watson process $\widetilde{Z}_n$ is introduced as $\widetilde{Z}_0 = 1$ and

$$\widetilde{Z}_n = \sum_{i=1}^{\widetilde{Z}_{n-1}} X_{n,i}.$$

The r.v.'s $X_{n,i},\ n, i \in \mathrm{N}$, are i.i.d. with values in Z^+. Then the process $(Z_0 = 1, Z_1, Z_2, \ldots)$ has the same distribution as the process $(\widetilde{Z}_0 = 1, \widetilde{Z}_1, \widetilde{Z}_2, \ldots)$. The offspring distribution $p_k = P(X_{n,i} = k)$ of any mother (the i-th mother in n-th generation) corresponds to $p_k = P(\sum_{i=1}^{\infty} T_i(v) = k)$ (of mother v).

5 View forward and backward in time

Consider the composition

$$(f_n \circ f_{n-1} \circ \ldots) \circ f_1 = f_n \circ f_{n-1} \circ \ldots \circ f_1 = f_n \circ (f_{n-1} \circ \ldots \circ f_1)$$

of n functions f_1 to f_n. We used the brackets to show the different views. The left composition is the view backward in time, the right composition the view forward in time (Letac (1986)). The wording forward and backward originates from Markov process theory with forward and backward generating operators and has a natural meaning (Breiman (1968)).

There is some advantage in both views and switching from one to the other. See papers on random affine transformations (Kesten (1973)) (time reversion) for beautiful application of the principle and (Burton, Rösler (1995)) using the weighted branching process and the contraction method.

We are interested in ξ_{n+1}, the members of the $n+1$-st generation. The **forward view** looks first to the n-th generation ξ_n and then takes the offspring of every individual there,

$$\xi_{n+1} = \sum_{|v|=n} \sum_{j \in \mathrm{N}} \delta_{\ell(vj)}.$$

The **backward view** looks back to the ancestor of the individuals in the very first generation. The measure $\xi_n = \xi_n(\ell(\emptyset))$ is a function of the starting value $\ell(\emptyset)$. Every individual j of the first generation generates independently of the others some offspring within n generations, described by the independent copy $\xi_{n,j}$ of ξ. However, the starting value has changed to $T_j(\emptyset)\ell(\emptyset)$. Therefore ξ_{n+1} has the representation

$$\xi_{n+1}(\cdot) \stackrel{\mathcal{D}}{=} \sum_{j \in \mathrm{N}} \xi_{n,j}(T_j(\emptyset)(\cdot)).$$

So far this looks very innocent. To show the potential and application of the different views let us consider the total mass Z_n in the n-th generation starting with $\ell(\emptyset) \equiv 1$. The *forward equation* is

$$Z_{n+1} = \sum_{|v|=n} \Big(\ell(v) \sum_{j \in \mathrm{N}} T_j(v) \Big). \tag{5.1}$$

The *forward structure* is a martingale. The weighted branching process, normalized by the exponential factor m^n, $m = E \sum_i T_i$, is a martingale. Every positive martingale converges almost everywhere,

$$\frac{Z_n}{m^n} \to W \quad \text{a.e.} \tag{5.2}$$

The *forward result* is a probabilistic result, an a.e. result for r.v.s.

The *backward equation* is

$$Z_{n+1} = \sum_j T_j(\emptyset) \sum_{|v|=n} \frac{\ell(jv)}{T_j(\emptyset)}. \tag{5.3}$$

We prefer to write this in terms of distributions

$$Z_{n+1} \stackrel{\mathcal{D}}{=} \sum_j T_j(\emptyset) Z_{n,j}. \tag{5.4}$$

The r.v.

$$Z_{n,j} := \sum_{|v|=n} \frac{\ell(jv)}{T_j(\emptyset)}$$

describes the total mass in the $n+1$-st generation of the offspring of the j-th individual in the first generation. The distribution is the same as that of Z_n.

The *backward structure* is an iteration on distributions. Define the map K of probability measures on the real line to probability measures on the real line by

$$K(\mu) \stackrel{\mathcal{D}}{=} \sum_{i=1}^{\infty} T_i X_i, \tag{5.5}$$

where (X_i) are independent, $X_i = \mu$.

Notice (5.4) is of the form (5.5) when seen for distributions. The distribution of Z_{n+1} is the map K applied to the distribution of Z_n. By iteration we obtain

$$K^n(\mu_0) \stackrel{\mathcal{D}}{=} Z_n \tag{5.6}$$

where μ_0 denotes the starting distribution (of $\ell(\emptyset)$).

The *backward result* is a functional analytic result on measures. It deals with fixed points $\nu = K(\nu)$ of the function K. In probabilistic form (cf. (2.7))

$$W \stackrel{\mathcal{D}}{=} \sum_j T_j W_j. \tag{5.7}$$

Here W_j are independent copies of W and $T, W_.$ are independent.

The forward view of the R_n process is

$$R_{n+1} = R_n + \sum_{|v|=n+1} \ell(v) C(v) \tag{5.8}$$

and the backward view is,

$$K(\mu) \stackrel{\mathcal{D}}{=} \sum_j T_j X_j + C \tag{5.9}$$

or written in more intuitive terms of r.v.s,

$$R_{n+1} \stackrel{\mathcal{D}}{=} \sum_j T_j(\emptyset) R_{n,j} + C(\emptyset) \tag{5.10}$$

Here $(T, C), R_{n,j}, j \in \mathrm{N}$, are independent r.v.s. The distribution of (T, C) is given in (2.7) and $R_{n,j}$ are independent copies of R_n (having distribution μ).

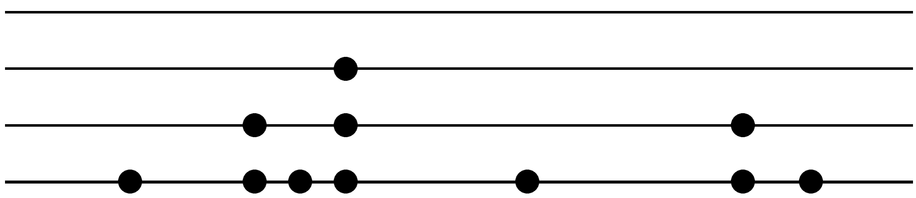
If the R_n converges in some sense to W then one expects, by (2.7),

$$W \stackrel{\mathcal{D}}{=} \sum_j T_j W_j + C. \tag{5.11}$$

Here W_j are independent copies of W and $(T, C), W_i$ are independent.

Equation (5.10) turns up in several different settings. We shall start with the easier equation (2.7), i.e. possible fixed points of K. Let us look again to Quicksort.

Quicksort: One way to think of the limiting distribution of Quicksort is via a binary tree.



(The lines denote the lists in order of appearance. The $\bullet$ gives the position of the chosen element after comparison and divides the list into two parts. The weights correspond to the list lengths. A list of length 0 is non observable.)

In our notation as a weighted branching process we get a representation of Quicksort with $T(v) = (T_1(v), 1 - T_1(v), 0, \ldots)$ where $T_1(v) : \Omega \mapsto \mathrm{R}$ is uniformly distributed. Then the total in the first n generations R_n as defined above (with $C(v)$ given by $C(v) = f(T_1(v))$, $f(x) := 2x \ln x + 2(1-x)\ln(1-x) + 1$) converges to some W and this W satisfies the fixed point equation for Quicksort (cf. (3.1)). Also the median version of quicksort fits this scheme.

Several further examples of weighted branching processes (and algorithms) can be analyzed similarly (cf. Rösler (1992)). The contraction method provides the fixed points as in Lemma 1. A related branching type recursion in the context of random fractal measures is discussed in Hutchinson, Rüschendorf (1998). Different metrics are useful and provide various related convergence results also in the nonstationary case after normalization (Cramer (1997a), Cramer, Rüschendorf (1996b)). This point of view is detailed in the following section.

6 Alternative probability metrics

In Sections 2, 3, and 5 we stated some conditions which imply convergence of a random divide and conquer algorithm to the fixed point of an affine linear stochastic equation. One of the conditions is that the limiting operator for this equation is a contraction with respect to the minimal ℓ_p-metric (the Mallows-metric). A well-known characteristic equation of the normal distribution is the fixed point equation

$$X \stackrel{d}{=} \sum_{i=1}^{d} \tau_i^{1/2} X_i, \tag{6.1}$$

where X_i are i.i.d. copies of X, $0 \le \tau_i \le 1$, $\sum \tau_i = 1$. The $(\tau_1, \ldots, \tau_d)$, X_i, $1 \le i \le d$, are independent random variables. It is easy to check that the operator corresponding to equation (6.1) has contraction factor 1 with respect to the ℓ_2-metric and is not a contraction with respect to any ℓ_p-metric. Therefore, in order to prove asymptotic normality of a recursive algorithm we have to follow metrics different from the ℓ_p-metrics (if we are going to follow our approach in section 1).

Consider a recursion of the form

$$X_n \stackrel{d}{=} 1 + X_{Z_n} + \overline{X}_{n-Z_n}, \tag{6.2}$$

where $(\overline{X}_k)$ are independent copies of (X_k), and Z_n denotes the subgroup size. In the classical examples of the size of tries and of the communication resolution interval (CRI), Z_n is binomial distributed $Z_n \stackrel{d}{=} \mathcal{B}(n, \vartheta)$ (see e.g. Mahmoud (1991)). Consider the probability metric $\vartheta_{1,\alpha}$ for $r = 2 + \frac{1}{\alpha}$, $1 \le \alpha < \infty$, $\frac{1}{\alpha} + \frac{1}{\beta} = 1$ defined by

$$\vartheta_{1,\alpha}(X, Y) = \sup\{|Ef(X) - Ef(Y)|;\ f \in C^{(2)}, ||f^{(2)}||_\beta \le 1\}. \tag{6.3}$$

Then $\vartheta_{1,\alpha}(X, Y) = \vartheta_{1,\alpha}(\mathcal{L}(X), \mathcal{L}(Y))$ depends only on the distributions of X and Y and has the important property of regularity of order r:

$$\vartheta_{1,\alpha}(cX, cY) = c^r \vartheta_{1,\alpha}(X, Y)$$

for $c > 0$ (For further properties see Rachev (1991)).

Convergence with respect to $\vartheta_{1,\alpha}$ implies convergence in distribution.

Let $a_n = EX_n$, $Y_n = \frac{X_n - a_n}{\sqrt{n}}$ and $\sigma_n^2 = EY_n^2$. Then consider the following assumptions

A1) $\sigma_n^2 \to \sigma^2 > 0$, $\sup E|Y_n|^r < \infty$, and

A2) $\dfrac{Z_n}{n} \to \tau$ stochastically, $E\tau > 0$ and $\overline{\lim}\, E\left(\left(\dfrac{Z_n}{n}\right)^{r/2} + \left(\dfrac{n - Z_n}{n}\right)^{r/2}\right) < 1.$

Condition A1) says that the asymptotic order of the variance of X_n is n and the normalization by $\sqrt{n}$ is correct. Condition A2) implies that the two subgroups of size Z_n, $n - Z_n$ are in some sense not too extreme. These conditions have been checked for the binomial case by analytic methods. The following convergence result has been proved in Rachev, Rüschendorf (1995).

Theorem 4 (Asymptotic normality)
Under conditions A1) *and* A2) *the sequence* (Y_n) *is asymptotically normal,*

$$Y_n \xrightarrow{\mathcal{D}} N(0, \sigma^2). \tag{6.4}$$

Sketch of the proof: The normalized sequence (Y_n) satisfies the recursion

$$Y_n \stackrel{d}{=} \left(\frac{Z_n}{n}\right)^{1/2} Y_{Z_n} + \left(\frac{n - Z_n}{n}\right)^{1/2} \overline{Y}_{n-Z_n} + C_n, \tag{6.5}$$

where $C_n = c_n(Z_n) = \frac{1}{\sqrt{n}} + \frac{a_{Z_n}}{\sqrt{n}} + \frac{a_{n-Z_n}}{\sqrt{n}} - \frac{a_n}{\sqrt{n}}$ and Z_n is the subgroup size.

If we can prove that $C_n \to 0$ then the limiting form of equation (6.5) is given by

$$Y \stackrel{d}{=} \tau^{1/2} Y + (1 - \tau)^{1/2} \overline{Y} \tag{6.6}$$

with $\tau = \lim \frac{Z_n}{n}$ independent of $Y, \overline{Y}$. (6.6) characterizes the normal distribution, while A2) implies that the operator corresponding to equation (6.6) is a contraction with respect to $\vartheta_{1,\alpha}$.

In order to prove convergence of Y_n we introduce the accompanying sequence,

$$N_n^* = \left(\frac{Z_n}{n}\right)^{1/2} N_{Z_n} + \left(\frac{n - Z_n}{n}\right)^{1/2} \overline{N}_{n-Z_n} + C_n, \tag{6.7}$$

where (N_n) are independent, $N_n \stackrel{d}{=} N(0, \sigma_n^2)$ and $(\overline{N}_n)$ are independent copies of (N_n). To prove $a_n = \vartheta_{1,\alpha}(Y_n, N_n^*) \to 0$ the essential argument is given by the following bounds:

$$\begin{aligned}
\vartheta_{1,\alpha}(Y_n, N_n^*) &\leq \sum_{k=0}^{n} P(Z_n = k)\, \vartheta_{1,\alpha}\left(\left(\frac{k}{n}\right)^{1/2} N_k + \left(\frac{n-k}{n}\right)^{1/2} \overline{N}_{n-k} + C_n,\right. \\
&\qquad\qquad \left.\left(\frac{k}{n}\right)^{1/2} Y_k + \left(\frac{n-k}{n}\right)^{1/2} \overline{Y}_{n-k} + C_n\right) \\
&\leq \sum_{k=0}^{n} P(Z_n = k) \left(\vartheta_{1,\alpha}\left(\left(\frac{k}{n}\right)^{1/2} N_k, \left(\frac{k}{n}\right)^{1/2} Y_k\right)\right. \\
&\qquad\qquad \left. + \vartheta_{1,\alpha}\left(\left(\frac{n-k}{n}\right)^{1/2} \overline{N}_{n-k}, \left(\frac{n-k}{n}\right)^{1/2} \overline{Y}_{n-k}\right)\right) \\
&= \sum_{k=0}^{n} P(Z_n = k) \left(\vartheta_{1,\alpha}(N_k, Y_k) \left(\frac{k}{n}\right)^{r/2} + \vartheta_{1,\alpha}(N_{n-k}, Y_{n-k}) \left(\frac{n-k}{n}\right)^{r/2}\right) \\
&\leq a \left[E\left(\frac{Z_n}{n}\right)^{r/2} + E\left(\frac{n - Z_n}{n}\right)^{r/2}\right].
\end{aligned}$$

Here $a = \sup \vartheta_{1,\alpha}(Y_n, N_n)$ can be shown to be finite by moment estimates for $\vartheta_{1,\alpha}$ from above. For the first inequality we used conditioning for the second and third step we applied the regularity properties of the metric. From properties of the metric we obtain that $\vartheta_{1,\alpha}(N_n, N_n^*) \to 0$. Furthermore, at this step we simultaneously obtain that $C_n \to 0$. Arguing as in Sections 2 and 4 we obtain $a_n \to 0$. Note that the order of regularity $r > 2$ is essential for the convergence proof. □

The proof of Theorem 4 has been extended to more extreme splittings leading to stable limit distributions. The assumptions on the variance have been also weakened to regularly varying variances (σ_n), as well as to multivariate versions of the form

$$X_n \stackrel{d}{=} 1 + \sum_{i=1}^{M} X^{(i)}_{Z_{n,i}+L_i} \tag{6.8}$$

with immigrations L_i (Feldman, Rachev, Rüschendorf (1995,1997)).

For recursive equations of the type

$$X_n \stackrel{d}{=} X_{Z_n} \vee \overline{X}_{n-Z_n} + C_n, \quad \vee = \max, \tag{6.9}$$

or of the type

$$X_n \stackrel{d}{=} X_{n-1} \vee \overline{X}_{n-1} + Y_n, \tag{6.10}$$

it turns out that weighted sup-norm metrics

$$\varrho_r(X, Y) = \sup_x |x|^r |F_X(x) - F_Y(x)| \tag{6.11}$$

have suitable regularity properties. This type of recursion arises in problems like max search length or simultaneous merge sort (parallel computing). For some examples see Cramer (1997b) and Rachev, Rüschendorf (1995).

7 Logarithmic rates and weak contraction factors

For some recursion sequences (X_n) the standard form of the contraction analysis as described in sections 1, 2 is not applicable. Take for example the case of the classical central limit theorem where S_n is a normalized sum of independent identically distributed centered random variables, and $0 < \alpha \leq 2$,

$$S_n = \frac{1}{n^{1/\alpha}} \sum_{i=1}^{n} Y_i. \tag{7.1}$$

Then (S_n) satisfies the recursion

$$S_{n+1} \stackrel{d}{=} \left(\frac{n}{n+1}\right)^{1/\alpha} S_n + \left(\frac{1}{n+1}\right)^{1/\alpha} Y_{n+1}. \tag{7.2}$$

This leads to the trivial limiting equation

$$S \stackrel{d}{=} S + 0, \tag{7.3}$$

i.e. we do not get a hint on the limiting equation in a formal way. A second problem which can be seen also in the recursion (7.2) is that the contraction factors between S_n and S_{n+1} converge to one so we can not apply the Banach Fixed Point Theorem for proving convergence. Finally, a third problem is the fact that probability metrics do not easily handle logarithmic normalizations or logarithmic convergence rates which are quite common in search type algorithms. In the following simple example we show that these problems can be handled by some variation of the contraction method. For this and some related examples see Cramer, Rüschendorf (1996a).

Let X_n denote the number of records in a random permutation read from the left to the right. Then X_n satisfies the simple recursion,

$$X_n \stackrel{d}{=} X_{n-1} + C_n, \qquad X_1 := 0, \tag{7.4}$$

where $C_n \stackrel{d}{=} \mathcal{B}(1, \frac{1}{n})$ and X_{n-1}, C_n are independent (see Hofri (1987, p. 112)). From (7.4),

$$\begin{aligned} EX_n &= H_n^{(1)} - 1\,, & \operatorname{Var} X_n &= H_n^{(1)} - H_n^{(2)} \\ H_n^{(k)} &:= \sum_{j=1}^{n} \frac{1}{j^k}\,, & H_n^{(1)} &= \ln n + \gamma + O\left(\frac{1}{n}\right), \end{aligned} \tag{7.5}$$

where $\gamma = 0.57722\ldots$ is the Euler constant and $H_n^{(2)} \to \zeta(2) = \frac{\pi^2}{6}$. The normalized version Y_n is defined by

$$Y_n = \frac{X_n - EX_n}{\sqrt{\ln n}}. \tag{7.6}$$

The variance of X_n is of order $\ln n$ and as in (7.2) we obtain a trivial limiting equation. For the analysis we use the Zolotarev-metric

$$\zeta_3(X,Y) = \sup\left\{|Ef(X) - Ef(Y)|;\ f \in C^{(2)} \text{ and } |f^{(2)}(x) - f^{(2)}(y)| \le |x-y|\right\}. \tag{7.7}$$

Note that ζ_3 is regular of order 3 and allows upper bounds in terms of moments.

Theorem 5 *Let $\sigma_n^2 = \operatorname{Var} Y_n$, $N_n \stackrel{d}{=} N(0,\sigma_n^2)$, then $\zeta_3(Y_n, N_n) \to 0$.*

Proof: Y_n satisfies the recursion,

$$Y_n \stackrel{d}{=} \sqrt{\frac{\ln(n-1)}{\ln n}}\, Y_{n-1} + \widetilde{C}_n \tag{7.8}$$

with $\widetilde{C}_n = \frac{C_n - EC_n}{\sqrt{\ln n}}$. Consider the accompanying 'normal' copy of Y_n $N_n^* := \sqrt{\frac{\ln(n-1)}{\ln n}}\, N_{n-1} + \widetilde{C}_n$. Since $N_n = \sqrt{\frac{\ln(n-1)}{\ln n}}\, N_{n-1} + V_n$, where $V_n \stackrel{d}{=} N(0,\tau_n^2)$ with

$\tau_n^2 = \sigma_n^2 - \frac{\ln(n-1)}{\ln n}\sigma_{n-1}^2 = \mathrm{Var}\,\widetilde{C}_n$, we obtain using the regularity of order 3 of the Zolotarev metric as in Section 6

$$\begin{aligned}\zeta_3(Y_n, N_n) &\leq \zeta_3(Y_n, N_n^*) + \zeta_3(N_n^*, N_n) \\ &\leq \left(\frac{\ln(n-1)}{\ln n}\right)^{3/2} \zeta_3(Y_{n-1}, N_{n-1}) + \zeta_3(V_n, \widetilde{C}_n).\end{aligned} \tag{7.9}$$

This yields by iteration the inequality

$$\zeta_3(Y_n, N_n) \leq \left(\frac{\ln 2}{\ln n}\right)^{3/2} \zeta_3(Y_2, N_2) + \sum_{i=3}^{n} \left(\frac{\ln i}{\ln n}\right)^{3/2} \zeta_3(V_i, \widetilde{C}_i). \tag{7.10}$$

Next, $\zeta_3(V_i, \widetilde{C}_i) \leq \frac{C}{i(\ln i)^{3/2}}$ and $E|V_i - \frac{1}{i}|^3 \leq \frac{1}{i}$ and then from (7.10), by straightforward calculations, we get

$$\zeta_3(Y_n, N_n) \leq \frac{1}{3\sqrt{\ln n}} + O\left((\ln n)^{-3/2}\right). \qquad \square$$

Equation (7.9) shows that we have a contraction factor of order $\left(\frac{\ln(n-1)}{\ln n}\right)^{3/2}$ converging to one in this case. The convergence result is of logarithmic order while typically it is of exponential order in the standard contraction method. From a related convergence result for the supremum distance and by the Lyapunov Central Limit Theorem, it is to be expected that the rate of convergence $\frac{1}{\sqrt{\ln n}}$ is exact in this case. Furthermore, in this example the limiting normal distribution can be expected since the normal distribution allows us to mimic the recursion in (7.8) (see the sequences N_n^*, N_n). In fact, this is the basic principle to construct the accompanying sequence N_n^*. It would be desirable, to obtain general rules to "predict" the limiting distribution.

In probability theory there has been developed a very effective induction method for proving the central limit theorem for sums, including a rate of convergence result. The analysis is based on contraction techniques and is known as Bergströms induction technique (Rachev (1991), Rachev, Rüschendorf (1994a), Rachev, Rüschendorf (1994b)). It should be a useful technique for more general classes of recursive algorithms.

8 Branching Dynamical Systems

The algorithm FIND, a basic random divide and conquer algorithm, does not fit into the scheme of a weighted branching process. Its recursive structure involves the whole process (see Section 9). This is the reason, why we extend the frame further, to branching dynamical systems. For these processes we have a view forward and backward in time and the contraction method is applicable, with many interesting metrics and new results. In this setting of branching dynamical systems we have many examples, like iterated function systems, image compression, branching Markov processes or the algorithm FIND.

This framework is a suitable setting to deal with fixed point equations (on more abstract spaces). Solutions of fixed point equations serve as candidates for the limiting distribution in the given examples. The contraction method plays a major role in finding solutions and in the convergence to the solution.

Branching Dynamical Systems: Let V be a tree as before. Let $(\Omega, \mathcal{F}, P)$ be a probability space, which is always assumed to be rich enough. Let $(E, \mathcal{E})$ be any non empty measurable space. Let $T(v) : \Omega \mapsto F(E, E^{\mathrm{N}}),\ v \in V$, be independent identically distributed random variables with values in the set $F(E, E^{\mathrm{N}})$ of measurable functions from E to E^{N} (endowed with the product $\sigma-$field). We suppress the ω whenever possible. We use $T(v) = (T_1(v), T_2(v), \ldots)$ as before with $T_i(v) : \Omega \mapsto F(E, E)$ the i-th projection. We drop the index v if superfluous. Notice that we allow arbitrary dependence of the components $T_j(v),\ j \in \mathrm{N}$, for fixed v. The families are independent, but within a family the offspring of the mother depend on each other.

The weights $\ell : \Omega \times V \mapsto E^E$ are defined by

$$\ell(vi)(\cdot) := T_i(v)(\ell(v)(\cdot)).$$

We call a tupel $((\Omega, \mathcal{F}, P), V, E, T)$, as above a *random branching dynamical system.*

If we are only interested in the collection of weights $\ell(v)$ in the n-th generation $|v| = n$, without the specific enumeration by the v, then we use the name *branching process with an arbitrary type space* (Athreya, Ney (1972), Asmussen, Hering (1983)). (These objects are studied in the field of branching processes under that name. E is the type space. See also under branching Markov process and branching random walk.)

Branching Markov process: As a dynamical system consider $\ell(v)(e_0)$ as it evolves in time (which corresponds to the tree) for a fixed e_0. This process has a Markov property in the sense, that the weight of a children $vi, i \in \mathrm{N}$, depends only on the weight of the mother and some stochastic manipulation. It is independent of all the other history. (For example, at every v throw a specific dice, the choice depends only on $\ell(v)(e_0)$, and the outcome is the weight of the child.) This model as described is a *branching Markov process.* e_0 has the interpretation of a starting distribution δ_{e_0}. Thinking in terms of random processes, we may use e_0 also as a r.v., independent of everything else so far, with values in E. For simplicity we drop e_0 (and ω) whenever possible.

The grave: An element $\triangle \in E$ is called *grave* if it satisfies

$$T_j(v)(\triangle) = \triangle$$

$v \in V,\ j \in \mathrm{N}$. The notation $\triangle$ is standard for the grave in Markov Process theory and used here throughout. Assume E has a *grave* $\triangle$. (Otherwise add one.) Call every individual v *observable* or *alive*, if $\ell(v) = \ell(v, e_0, \omega)) \neq \triangle$ and *non-observable* if $\ell(v) = \triangle$. Notice that any offspring of a non-observable individual is non-observable. Non-observable indivuduals are not shown in our picture.

Depending on the dynamics T any mother might have only finitely many children, or none or even infinitely many. Playing around with this construction (maybe adding a grave) we can imbed branching processes, see the examples, with a finite number of offspring into our setup of a branching dynamical system. We could for example construct in this way a process on a binary tree $\{0,1\}^* = \cup_{n=0}^{\infty}\{0,1\}^n$ by assigning some values different from the grave to the first two children and the grave to all the other children. This way of interpretation and consideration has some advantages. See the picture of the weighted branching processes for a realization.

Objects of interest: The n-th generation consists of a collection $\ell(v),\ |v| = n$ of weights. A handy description is via discrete measures

$$\xi_n = \sum_{|v|=n} \delta_{\ell(v)}.$$

Here δ_x denotes a point measure on $x \in E^E$ respectively E for given e_0.

Typical questions concern the type or behaviour of the n-th generation. Let u be a (measurable) function from E to some E'. E' is some space, where a sum makes sense, like a group, the reals, some function space, a vectorspace, a positive cone, a monoid etc. An interesting object is the total weight

$$Z_n(e_0) := \int u(f(e_0))\xi_n(df) = \sum_{|v|=n} u(\ell(v)(e_0))$$

in the n-th generation. ($e_0 \in E$ or random.) If a semigroup G acts on E', $E' \times G \mapsto E'$, for example a field on a vector space over that field, we consider also

$$R_n(e_0) := \sum_{|v|\leq n} u(\ell(v)(e_0))C(v)$$

depending on the first n generations. Here $(T(v), C(v)),\ v \in V$, are i.i.d. r.v.s and $C(v)$ is an r.v. with values in the semigroup G.

For some applications one is more interested in the maximal (or minimal) height of a tree

$$\inf\{n : \forall |v| = n,\ u(\ell(v)(e_0)) \geq 1\}$$

for a suitable function u.

Let us give some examples in order to show the variety of included processes. Branching dynamical systems have a branching structure and an iterative structure. We point this out by the two 'extreme' settings, one with no branching at all and one with only two states, dead or alive.

No branching: In this setting every mother has exactly one (living) child, let us say the first child is alive. The map T will be

$$T(v)(e) = (T_1(v)(e), \triangle, \triangle, \ldots).$$

If we start with one indivual, then in the n-th generation there is exactly one child alive, the child with name $(1, \dots, 1) =: v(n)$, $|v(n)| = n$. If we use $\Psi_i = T_1(v(i))$ for a moment, then

$$\ell(v) = \Psi_{n-1} \circ \Psi_{n-2} \circ \dots \circ \Psi_0$$

is the composition of n independent identically distributed random variables (i.i.d. r.v.s) with value in E^E. These objects are well studied in random dynamical systems (or Markov processes).

Trivial weights: In this setting we allow only two possible states, alive $(= 1)$ and dead $(= 0)$. 0 serves as a grave. This is the Galton-Watson process as described earlier.

Weighted branching process: Let $E = \mathrm{R}$ and $T_i(v)$ be a multiplication, $T_i(v)(e) = ae$ for some $a \in \mathrm{R}$. We use, this is a slight abuse of notation, $T_i(v)$ as the factor a. This branching dynamical system is a weighted branching process and every weighted branching process is a branching dynamical system. The 0 serves as a grave.

Iterated function systems: Iterated function systems are branching dynamical systems. As a special case take E a Hilbert space or R^d and T_1 a random affine transformation, $T_1(\cdot)(e) = Ae + b$ with A a random linear operator matrix (matrix) and b a vector (see Kesten (1973)), and for a treatment based on contraction arguments (see Burton, Rösler (1995), Rachev, Rüschendorf (1995)). Iterated function systems are thus related to products of i.i.d. matrices. the limit of a iterated function system represents a coding of a picture.

The fixed point equation

$$W \stackrel{d}{=} T_1 W_1 + C,$$

is also of importance in financial mathematics and insurance (Dufresne (1990), Embrechts, Goldie (1994)).

This example is representative for many others in the setup of self similarity and contraction of set valued functions like Graf (1987) and Arbeiter (1991) or in the context of random fractal measures in Hutchinson, Rüschendorf (1998). Concerning the aspect of self similarity of this equation see a series of papers by Mandelbrot on fractals and some models, which he called "canonique". Inspired by these views, Kahane, Peyrière (1976) and Guivarch (1990) considered positive solutions of the fixed point equation in the set-up of a fixed number of i.i.d. components T_j of T. In the context of infinite particle systems Durrett, Liggett (1983), gave in the above setting a complete description of all positive solutions of the fixed point equation. Rösler (1998a) considered all solution in case T_i are constants. The general solution is unknown. A discussion of these results would unduly expand this paper.

Here are more examples from computer science, especially divide and conquer algorithms.

Quicksort: Quicksort can be seen as a branching dynamical system. In our notation, $E = \mathrm{Z}^+$, $\triangle = 0$, $n_0 = 1$, $T_1(v)(e) : \Omega \mapsto \{0, 1, \dots, e-1\}$, $e \geq 2$, has a uniform distribution, $T_2(v)(e) = e - T_1(v)(e),\ T_3 = 0 = T_4 = \ldots$. The state 0 corresponds to the grave. We start the process in some state n, representing the number of different objects to be sorted. After some running time, at least in n-th generation, the living individuals (v for the corresponding list) all have at most the weight 1, this means there is at most one element in the list.

Now take $n \to \infty$. The whole branching dynamical system, after normalization, converges to a limiting branching dynamical system, the weighted branching process associated to the fixed point equation.

In the same way Quad Trees, Median Quicksort, and Merge Sort are special cases. An example which needs the full generality of branching dynamical systems is the FIND algorithm which is analyzed in the following section.

9 Quickselect or Find

The algorithm Find or Quickselect (Hoare (1961)) gives as output the k-th largest element out of the input of n (different) numbers. We partition recursively like in Quicksort and use the same tree setting. The k-th largest element is for every partitioning step in exactly one of the resulting lists, the list of strictly smaller numbers, the list consisting of the chosen one and the list of strictly larger numbers. We have to follow only that branch. Finding the k-th largest corresponds to travelling down the tree by a specific path until we hit a list (v) with only one element ($\ell(v) = 1$).

Let $X_n(k)$ be the number of comparisons in order to find the k-th largest in a set of n different numbers. The key equation for Find in this settting is for $n \geq 2$

$$X_n(k) = n - 1 + 1_{\{k<U_n\}} X_{U_n - 1}(k) + 1_{\{k>U_n\}} \overline{X}_{n-U_n}(k - U_n).$$

Here $X, \overline{X}, U_n$ are independent and the distribution of U_n is uniform on $1, 2, \ldots, n$. The distribution of $X_i, \overline{X}_i$ is given recursively.

Introduce $Y_n : \Omega \to \mathrm{R}^{[0,1]}$ by

$$Y_n(t) = \frac{1}{n} X_n(\lceil tn \rceil)$$

and $Y_n(0) = \lim_{t\to 0} Y_n(t)$. Notice $Y_n(\frac{k}{n}) = \frac{1}{n} X_n(k)$. Then the above key equation becomes

$$\begin{aligned} Y_n(t) \ &\stackrel{\mathcal{D}}{=}\ 1 - \frac{1}{n} + 1_{\{\lceil tn \rceil < U_n\}} \frac{U_n - 1}{n} Y_{U_n - 1}\left(\frac{\lceil tn \rceil}{U_n - 1}\right) \\ &\qquad + 1_{\{\lceil tn \rceil > U_n\}} \frac{n - U_n}{n} \overline{Y}_{n-U_n}\left(\frac{\lceil tn \rceil - U_n}{n - U_n}\right). \end{aligned}$$

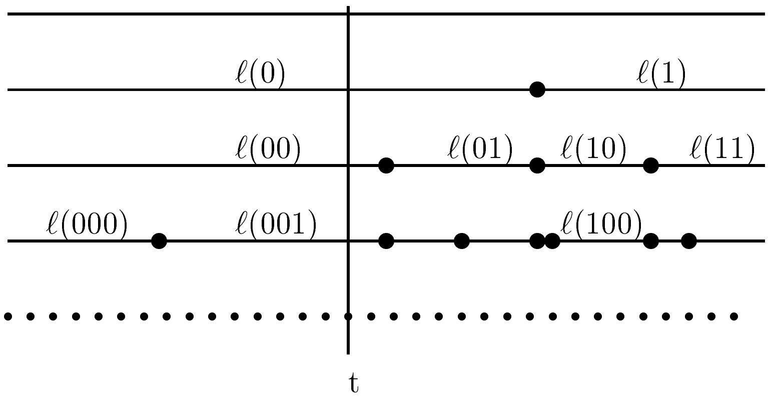
From this equation one might guess, that Y_n as a process converges to some process Z, called the Find process, satisfying the recursive equation

$$Z \stackrel{\mathcal{D}}{=} 1 + 1_{[0,U)} U Z\left(\frac{\cdot}{U}\right) + 1_{[U,1]}(1 - U)\overline{Z}\left(\frac{\cdot - U}{1 - U}\right). \tag{9.1}$$

Here $U, Z, \overline{Z}$ are independent. U has a uniform distribution, Z and $\overline{Z}$ have the same distribution and are r.v.s with values in $\mathrm{R}^{[0,1]}$. In the following we introduce the Find process. We do not discuss in this review the convergence of Y_n to Z in a suitable topology (Grübel, Rösler (1996)), since that is more technical. We instead obtain a regular version of the limiting process in $D[0,1]$ by an explicit construction.

The Find process: Let, on a sufficiently rich probability space $(\Omega, \mathcal{F}, P)$, $U(v), v \in V$ be independent random variables uniformly distributed on the unit interval. Define the factors $T_1(v) = U(v),\ T_2(v) = 1 - U(v),\ T_3(v) = 0 = T_4(v) = T_5(v) = \dots$ and the weights recursively by $\ell(vi) = \ell(v)T_i(v),\ \ell(\emptyset) = 1$. The process $((\Omega, \mathcal{F}, P),$

$V, T)$ is a weighted branching process.



Define the map ψ on $[0,1) \times \Omega$ to the set of infinite $0-1$ sequences in $\{0,1\}^{\mathrm{N}}$ with infinitely many zeros by

$$t = \sum_{n=1}^{\infty} \psi(t)|_n \ell((\psi(t)|_{n-1}, 0)). \tag{9.2}$$

The map ψ is well defined. Notice $\psi(t)_n$ is either 0 or 1 according to whether t is in the left or right branch of the n-th generation of the corresponding knot in the $n-1$-st generation. (The map ψ is for almost all $\omega \in \Omega$ a bijective map.) We call the sequence $\psi(t)$ the *random binary representation* of a real number in the unit interval. Notice the analogy to the binary representation with a weight function $\overline{\ell}(v) := 2^{-|v|}, t = \sum_n t_n 2^{-n}$.

Define the process

$$Z_n(t) := 1 + \sum_{j=1}^{n} \ell(\psi(t)|_j), \tag{9.3}$$

$t \in [0,1)$, $n \in \mathrm{Z}^+$. The r.v. $Z_n(t)$ counts the length of the halfopen intervals $[.,.)$ up to the n-th generation, which contain the real number t. Define $Z_n(1) := \lim_{t\uparrow 1} Z_n(t)$.

For every ω the function $Z_n : [0,1] \mapsto \mathrm{R}$ is pointwise increasing to some limiting function Z. Since the Z_n satisfy the backward recursive equation, (5.10)

$$Z_{n+1} \stackrel{\mathcal{D}}{=} 1_{[0,U)} U Z_n \left(\frac{\cdot}{U}\right) + 1_{[U,1]}(1-U) Z_n \left(\frac{\cdot - U}{1-U}\right) + 1,$$

best seen by the picture, the limiting function Z satisfies the Find key equation (9.1). Further all paths of Z are in $D([0,1])$. (This process is unique as a measure on $D([0,1])$. $D([0,1])$ is the set of right continuous function $[0,1] \mapsto \mathrm{R}$ such that the left limits exist.) The Find process was introduced in Grübel, Rösler (1996).

Find as a branching dynamical system: Let us first consider the Find process as a special branching dynamical system. As state space E choose the function space $\mathrm{R}^{[0,1]}$. The function identically 0 serves as a grave. The maps T_j are

$$T_1(v)(e) = 1_{[0,U(v))} U(v) e \left(\frac{\cdot}{U(v)}\right) \qquad T_2(v)(e) = 1_{[U(v),1]}(1-U(v)) e \left(\frac{\cdot - U(v)}{1-U(v)}\right)$$

and all $T_3(v)(e) \equiv 0 \equiv T_4(v)(e) \equiv \ldots$ The $U(v)$, $v \in V$, are i.i.d. r.v.'s with a uniform distribution. Choose $C(v) \equiv 1$. Then $R_n := \sum_{|v| \le n} \ell(v) C(v)$ satisfies the backward equation

$$R_{n+1} \stackrel{\mathcal{D}}{=} \sum_j T_j R_{n,j} + 1$$

as in (5.10). The map K on distributions is written as

$$K(\mu) \stackrel{\mathcal{D}}{=} 1_{[0,U)} U X_1 \left(\frac{\cdot}{U}\right) + 1_{[U,1]}(1-U) X_2 \left(\frac{\cdot - U}{1-U}\right) + 1, \tag{9.4}$$

where U, X_1, X_2 are independent r.v.'s. U has a uniform distribution, the processes X_i have distribution μ. But this is exactly the key equation for the Find process.

In order to see K as a contraction, let us consider the metric

$$d(\mu, \nu) := \inf \| \, \|X - Y\|_{L_p([0,1],\lambda)} \|_{L_p(\Omega,P)}.$$

The infimum is over all processes X, Y with the given distribution. The symbol $\| \cdot \|_{L_p(\Omega,P)}$ denotes the $L_p(\Omega, P)$ norm and the symbol $\| \cdot \|_{L_p([0,1],\lambda)}$ denotes the L_p norm on $[0,1]$ with respect to the Lebesgue measure λ,

$$\|X - Y\|_{L_p([0,1],\lambda)} = \left(\int_0^1 |X(t) - Y(t)|^p dt \right)^{1/p}.$$

We use $\| \cdot \|_p$ if possible. (Notice, that we may view an r.v. $X : \Omega \mapsto \mathrm{R}^{[0,1}$ also on the product space $X : \Omega \times [0,1] \mapsto \mathrm{R}$. The above norm is exactly the $L_p(\Omega \times [0,1), P \otimes \lambda)$ norm.)

d is a metric on the set M of measures μ of $L_p(\Omega \times [0,1), P \otimes \lambda)$. K is a strict contraction with respect to that metric with contraction constant

$$\|U^2\|_p + \|(1-U)^2\|_p < 1.$$

The contraction method provides convergence of the distribution of Y_n to some measure μ on the function space $\mathrm{R}^{[0,1]}$. Additional arguments are needed to establish the existence of a regular version of a process corresponding to μ.

Moments: The fixed point equation allows the calculation of moments of the Find process $EZ^m(t)$ for natural m. The moments are again fixed points of some operators. This approach was chosen by Paulsen (1997) for $p \in \mathrm{N}$ and an explicit formula was given there for $p = 2$. The first moment $p = 1$ was established earlier by Grübel, Rösler (1996) with the contraction method. Knuth (1973) gave already a closed form for the expectation of $X_n(k)$. The first and second moment of $X_n(k)$ were calculated explicitly by Kirschenhofer, Prodinger (1998) by combinatorial arguments. Their results also provide the asymptotic behaviour. For the speed of convergence see Kodaj, Móri (1997). See Lent, Mahmoud (1996) and Mahmoud, Modarres, Smythe (1995) for further results using the contraction method.

Acknowledgement: The authors thank two reviewers for very detailed and helpful comments on the paper.

References

[1] M. Arbeiter, Random recursive construction of self-similar fractal measures. The noncompact case, *Probability Theory and Related Fields*, **88** (1991), 497-520.

[2] K.B. Athreya and P. Ney, *Branching Processes,* Die Grundlehren der mathematischen Wissenschaften, Band 196, Springer Verlag, New York-Heidelberg (1972), XI, p. 287.

[3] S. Asmussen and H. Hering, *Branching Processes,* Progress in Probability and Statistics, 1983, Birkhäuser, Boston-Basel-Stuttgart.

[4] P.J. Bickel and D.A. Freedman, Some asymptotic theory for the bootstrap, *Ann. Statist.*, **9** (1981), 1196-1217.

[5] P. Bougerol and N. Picard, Stationarity of GARCH processes and of some nonnegative time series. *J. Econometrics*, **52** (1992), 125-127.

[6] L. Breiman, *Probability,* Addison-Wesley Publishing Company, Reading, Mass.-London-Don Mills, Ont., 1968.

[7] V. Bruhn, Eine Methode zur asymptotischen Behandlung einer Klasse von Rekursionsgleichungen mit einer Anwendung in der stochastischen Analyse des Quicksort-Algorithmus, *Dissertation 1996*, Mathematisches Seminar der Universität zu Kiel.

[8] R. Burton and U. Rösler, An L_2 convergence theorem for random affine mappings, *Journal of Applied Probability*, **32** (1995), 183-192.

[9] W. M. Chen, H. K. Hwang, and G. H. Chen, The cost distribution of queue-mergesort, optimal mergesorts, and power-of-two rules, *Preprint* (1998).

[10] M. Cramer, Stochastische Analyse rekursiver Algorithmen mit idealen Metriken, *Dissertation 1995*, Freiburg.

[11] M. Cramer, A note concerning the limit distribution of quicksort algorithm, *Theoretical Informatics and Applications*, **30** (1996), 195-207.

[12] M. Cramer, Convergence of a branching type recursion with non-stationary immigration, *Metrika*, **46** (1997a), 187-211.

[13] M. Cramer, Stochastic analysis of 'simultaneous merge-sort', *Adv. Appl. Prob.*, **29** (1997b), 669-694.

[14] M. Cramer, Stochastic analysis of the Merge-Sort algorithm, *Random Structure and Algorithms*, **11** (1997c), 81-96.

[15] M. Cramer and L. Rüschendorf, Analysis of recursive algorithms by the contraction method, *Athens conference on applied probability and time series analysis, Athens, Greece, March 22–25, 1995*, Vol. I, 18-33, Lecture Notes in Statistics, **114** (1996a), Springer, New York.

[16] M. Cramer and L. Rüschendorf, Convergence of a branching type recursion, *Annales de l'Institut Henri Poincaré. Probabilités et Statistiques*, **32** (1996b), 725-741.

[17] M. Cramer and L. Rüschendorf, Convergence of two-dimensional branching recursions, *Preprint* (1998).

[18] G. Dall'Aglio, Sugli estremi dei momenti delle funzioni di ripartizione doppia, *Ann. Scuola Norm Sup. Pisa, Sci. Fis. Math.*, III. Ser., **10** (1956), 35-74.

[19] L. Devroye, A note on the expected height of binary search trees, *Journal of the ACM*, **33** (1986), 489-498.

[20] L. Devroye, Branching processes in the analysis of the heights of trees, *Acta Informatica*, **24** (1987), 277-298.

[21] L. Devroye, Universal limit laws for depth in random trees, *Preprint* (1998).

[22] L. Devroye and L. Laforest, An analysis of random d-dimensional quad trees, *SIAM Journal on Computing*, **19** (1990), 821-832.

[23] R.P. Dobrow and J.A. Fill, Total path length for random recursive trees, *Preprint* (1998).

[24] D. Dubischar, The representation of Markov processes by random dynamical systems. *Universität Bremen*, http://www.math.uni-bremen.de/rds/preprints, *Report 393*, (1997).

[25] D. Dufresne, The distribution of a perpetuity, with applications to risk theory and pension funding, *Scandinavian Actuarial Journal*, **1/2** (1990), 39-79.

[26] R. Durrett and M. Liggett, Fixed points of the smoothing transformation, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **64** (1983), 275-301.

[27] W. Eddy and M. Schervish, How many comparisons does quicksort use? *Journal of Algorithms*, **19** (1995), 402-431.

[28] P. Embrechts and Ch. Goldie, Perpetuities and random equations, in *Asymptotic Statistics (Prague, 1993), Contributions to Statistics,* Physica, Heidelberg, 1994, 75-86.

[29] M.H. van Emden, Increasing the efficiency of quicksort, *Communications of the ACM*, **13** (1970), 563-567.

[30] P. Feldman, S.T. Rachev, and L. Rüschendorf, Limit theorems for recursive algorithms, *J. Comp. Appl. Mathematics*, **56** (1995), 169-182.

[31] P. Feldman, S.T. Rachev, and L. Rüschendorf, Limiting distribution of the collision resolution interval, *Statistica Neerlandica*, **51** (1997), 1-22.

[32] R.A. Finkel and J.L. Bently, Quad trees, a data structure for retrieval on composite keys, *Acta informatica*, **4** (1974), 1-9.

[33] P. Flajolet and M. Golin, Mellin transforms and asymptotics: the mergesort recurrence, *Acta Informatica*, **31** (1994), 673-696.

[34] P. Flajolet, G. Gonnet, C. Puech, and J.M. Robson, Analytic variations on quadtrees, *Algorithmica*, **10** (1993), 473-500.

[35] P. Flajolet, G. Labelle, L. Laforest, and B. Salvy, Hypergeometrics and the cost structure of Quadtrees, *Random Struct. Algorithms*, **7** (1995), 117-144.

[36] P. Flajolet and T. Lafforgue, Search costs in quadtrees and singularity pertubation asymptotics, *Discrete and Computational Geometry*, **12** (1994), 151-175.

[37] H. Fürstenberg and H. Kesten, Products of random matrices, *Ann. Math. Statist.*, **31** (1960), 457-469.

[38] B. W. Gnedenko and A. N. Kolmogorov, Limit Distributions for Sums of Independnet Random Variables, *Gostekhizdat* 1949.

[39] M. Golin and R. Sedgewick, Queue-mergesort, *Information Processing Letters*, **48** (1993), 253-259.

[40] S. Graf, Statistically self-similar fractals, *Prob. Th. Rel. Fields*, **74** (1987), 357-392.

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

[42] Y. Guivarch, Sur une extension de la notion de loi semi-stable. *Ann. Inst. H. Poincaré*, **26** (1990), 261-286.

[43] P. Hennequin, Combinatorial analysis of quicksort algoithm, *Theoretical Informatics and Applications*, **23** (1989), 317-333.

[44] C.A.R. Hoare, Algorithm 63, Partition; Algorithm 64, Quicksort; Algorithm 65, Find, *Communicatons of the ACM*, **4** (1961), 321-322.

[45] C.A.R. Hoare, Quicksort, *Comp. J.*, **5** (1962), 10-15.

[46] M. Hofri, *Probabilistic analysis of algorithms. On computing methodologies for computer algorithms performance evaluation*, Springer, New York, 1987.

[47] I. Hutchinson and L. Rüschendorf, Random fractal measures and probability metrics, *Preprint* (1998).

[48] P. Jagers, Galton-Watson processes in varying environments, *J. Appl. Probab.*, **11**, (1974), 174-178.

[49] J.P. Kahane and J. Peyrière, Sur certaines martingales de Benoit Mandelbrot. *Adv. Math.*, **22** (1976), 131-145.

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

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

[52] Ch. Knessl and W. Szpankowski, Quicksort algorithm again revisited, *Preprint* (1998).

[53] D.E. Knuth, *The Art of Computer Programming, Fundamental Algorithms,* (Vol. 1), Addison-Wesley, 2. ed. (1973).

[54] B. Kodaj and T.F. Móri, On the number of comparisons in Hoare's algorithm "Find", *Studia Sci. Math. Hungar.* **33** (1997), 185-207.

[55] J. Lent and H.M. Mahmoud, Average-case analysis of multiple Quickselect: An algorithm for finding order statistics, *Stat. Prob. Letters*, **28** (1996), 299-310.

[56] G. Letac, A contraction principle for certain Markov chains and its applications, *Random matrices and their applications (Brunswick, Maine, 1984)*, Comtemporary Mathematics **50** (1986), 263-273, Amer. Math. Soc., Providence, R.I.

[57] H. Mahmoud, Limiting distributions for path lengths in recursive trees, *Probability Eng. Info. Sci.*, **5** (1991), 53-59.

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

[59] H.M. Mahmoud, R. Modarres, and R.T. Smythe, Analysis of Quickselect: an algorithm for order statistics, *RAIRO, Inform. Theor. Appl.*, **29** (1995), 255-276.

[60] H.M. Mahmoud, Sorting: A Distribution Theory, *Preprint* 1998.

[61] B. Mandelbrot, Multiplications aléatoires itérées et distributions invariantes par moyenne pondérée aléatoire, *C. R. Acad. Sci. Paris*, **278** (1974), 289-292.

[62] C.J. McDiarmid and R. Hayward, Large deviation for quicksort, *Journal Algorithms*, **21** (1996), 476-507.

[63] R. Neininger and L. Rüschendorf, On the internal path length of d-dimensional quad trees, *Preprint* (1998).

[64] W. Panny and H. Prodinger, Bottom-up mergesort–a detailed analysis, *Algorithmica*, **14** (1995) 340-354.

[65] V. Paulsen, The moments of Find, *Journal of Applied Probability*, **34** (1997), 1079-1082.

[66] S.T. Rachev, *Probability metrics and the Stability of Stochastic Models*, Wiley 1991.

[67] S.T. Rachev and L. Rüschendorf, Propagation of chaos and contraction of stochastic mappings, *Siberian Advances in Mathematics*, **1** (1994a), 114-150.

[68] S.T. Rachev and L. Rüschendorf, On the rate of convergenc in the CLT with respect to the Kantorovich metric, in *Proceedings from the 9th international conference on probability in Banach spaces*, (J. Hoffmann-Joergensen, J. Kuelbs, and M. Marcus, eds.), held at Sandbjerg, Denmark, August 16-21, 1993, Birkhaeuser, Prog. Probab., **35** (1994b), 193-208.

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

[70] M. Régnier, A limiting distribution for quicksort, *RAIRO, Theoretical Informatics and Applications*, **23** (1989), 335-343.

[71] W.T. Rhee and M. Talagrand, A sharp deviation inequality for the stochastic traveling salesman problem, *Ann. Prob.*, **17** (1989), 1-8.

[72] U. Rösler, A limit theorem for "Quicksort", *RAIRO, Theoretical Informatics and Applications*, **25** (1991), 85-100.

[73] U. Rösler, A fixed point theorem for distributions, *Stochastic Processes and their Applications*, **37** (1992), 195-214.

[74] U. Rösler, The weighted branching process, in *Dynamics of complex and irregular systems (Bielefeld, 1991)*, Bielefeld Encounters in Mathematics and Physics VIII, World Science Publishing, River Edge, NJ, 1993, pp. 154-165.

[75] U. Rösler, A fixed point equation for distributions, *Berichtsreihe des Mathematischen Seminars Kiel*, Christian-Albrechts-Universität zu Kiel, http://www.numerik.uni-kiel.de/reports/1998/98-7.html, 1998a.

[76] U. Rösler, On the analysis of stochastic divide and conquer algorithms, *Berichtsreihe des Mathematischen Seminars Kiel*, Christian-Albrechts-Universität zu Kiel, http://www.numerik.uni-kiel.de/reports/1998, 1998b.

[77] J. Schimmler, Stochastische Analyse des Mergesort-Algorithmus, *Diplomarbeit Kiel* 1997.

[78] R. Sedgewick, Quicksort, *Stanford Computer Science Report STAN-CS-75-492, Ph.d. thesis*, 1975. Also published by Garland, Pub. Co., New York, 1980.

[79] R.T. Smythe and H. Mahmoud, A survey of recursive trees, *Theory Prob. Math. Stat.*, **51** (1995), 1-27.

[80] K.H. Tan and P. Hadjicostas, Some properties of a limiting distribution of quicksort, *Statistics & Probability Letters*, **25** (1995), 87-94.

[81] V.M. Zolotarev, *Modern Theory of Summation of Random variables.* VSP, Utrecht, 1997.