Probabilistic analysis of algorithms, stochastic fixed-point equations and ideal metrics

Ralph Neininger*
Institut für Stochastik und Mathematische Informatik
Johann Wolfgang Goethe-Universität
Frankfurt am Main

Lecture Notes: Summer School on
Probabilistic Methods in Combinatorics
Graz, July 17-19

*Supported by an Emmy Noether-fellowship of the DFG.

Contents

1 Introduction

In the area of algorithms and data structures many recursive structures and algorithms appear that are commonly used in practical applications in Computer Science. Various trees used for the organization of data, algorithms for searching and sorting problems, algorithms on graphs and algorithms on sequences (DNA sequences, searching the internet) as well as problems from combinatorial optimization have a recursive structure. To measure the complexity of such algorithms certain elementary operations are counted to measure the time needed by the algorithm. The basic interest consists in quantitative information about complexity measures so that algorithms can be compared and efficient algorithms can be identified.

Since complexity measures usually depend on the particular input of the algorithm, in Computer Science one mainly does a "worst case" analysis or an "average case" analysis. While for the worst case one just takes the supremum of the complexity over all possible inputs (of a given size) for an average case analysis a certain probability measure is assumed on the set of possible inputs. This is often the uniform distribution if the possible inputs form a finite set.

The "Average Case Analysis of Algorithms" was started by D. E. Knuth in 1963 and has developed into a mathematical field, where generating function play a dominant role. An encyclopedic discussion of this approach to the analysis of algorithms is given by Knuth's (1969a, 1969b, 1973) books "The Art of Computer Programming".

Since the 80s of the last century the whole distributions of complexity measures have more regularly been studied to obtain more refined information on the behavior of the algorithms. As in the "average case analysis of algorithms" a stochastic model for the input is assumed for this. Then, beyond averages one tries to approximate and describe the distributions, for example with respect to large deviations and limit laws. Large deviations are of interest for Computer Science, since such bounds quantify the probability of bad behavior of the algorithm, something one usually wants to be controlled.

Here, we are not hunting for large deviations but for limit laws. For the study of weak convergence of complexity measures of recursive algorithms and data structure various techniques are in use:

- **moment generating functions and saddle point methods**, cf., e.g., Flajolet and Odlyzko (1982, 1990), Pittel (1999), Drmota (1997), Szpankowski (2001) and the references therein.

- **moments method (and cumulants)**, cf. Hwang (2003), Janson, Łuczak and Ruciński (2000, chapter 6) and the references therein.

- **martingale methods**, cf. Régnier (1989), Chauvin and Rouault (2004) and the references therein.

- **shortcuts to asymptotic normality** (using representations of independent or weakly dependent random variables, Stein's method, Berry-Esseen methods), cf. Devroye (2002/03), Barbour, Holst and Janson (1992), Janson, Łuczak and Ruciński (2000, chapter 6) and the references therein.

- **contraction method**, cf. Rösler (1991, 1992), Rachev and Rüschendorf (1995), Rösler and Rüschendorf (2001) and Neininger and Rüschendorf (2004a, 2004b) and the references therein.

In this text we will discuss the approach by the contraction method. The method is tailored to derive convergence in distribution for parameters of recursive structures. It was introduced in Rösler (1991) and later independently extended in Rösler (1992) and Rachev and Rüschendorf (1995) and has been developed to a fairly general tool during the last years.

By this method one starts with a recurrence satisfied by the quantities of interest and, based on information on the first two moments, does a proper normalization of the quantities. The recurrence for the scaled quantities leads to a fixed-point equation and a potential limit distribution is characterized as the fixed-point of a measure valued map.

In probability theory the concept of limit laws aims to approximate (convergent) sequences of distributions by their limit distribution. The contraction method can be considered a one step towards this aim. Often, beforehand asymptotic expansion of moments are needed which in more difficult cases are often derived using generating functions. Based on moments the contraction method can be used to show convergence in distribution. However, the limit distribution typically cannot directly be used to approximate the quantities as it is only given implicitly by some fixed-point property. In a final step one has to extract information about the limit distribution from the fixed-point property.

Here, we survey aspects of the method which have proven useful for the analysis of many concrete problems form Computer Science and related fields. For this, we start in section 2 with a general type of recurrence for distributions together with a dozen of application from various areas. In section 3 the idea of the method is explained. Later on, the realization of the idea depends on the choice of a suitable probability metric.

Until 2001, almost every practical application in Computer Science was derived in an L_2 setting based on the minimal L_2 metric. This realization of the general idea is discussed in section 4 together with applications.

In section 5 some fundamental limitations of the L_2 settings, mainly related to asymptotic normality, are discussed. These problems, together with applications, are studied in sections 6-8, where the contraction method is developed with the use of Zolotarev's metric following Neininger and Rüschendorf (2004a, 2004b).

In section 6 properties of ideal metrics and, in particular, of the Zolotarev metric are briefly introduced. Section 7 has a general convergence theorem with respect to the Zolotarev metric, that is specialized in different directions. In this section it is also discussed up to which extend information on the expansion of moments is needed to apply the method. Various contraction conditions are compared. Section 8 contains a universal limit law for quantities with a variance that is slowly varying at infinity. This covers cases in applications that lead to so-called degenerate fixed-point equations.

In section 9 some selected related problems are discussed, in particular rates of convergence, large deviations, the characterization of the set of solutions of a fixed-point equation, properties and perfect simulation of fixed-points, and recurrences, where we have a maximum instead of a sum.

The theory is illustrated a lot with applications. The reason for this is to show the richness of problems Computer Science has to offer and to capture the various different mathematical phenomenon that appear in the analysis of these problems.

2 Recursive sequences of distributions

In the section we specify a general recursive sequence of distributions and give various examples from the area of algorithms and data structures, that are special cases of the general setting. We develop the theory for random vectors in $\mathbb{R}^d$. For most applications however, it is sufficient to consider the univariate case $d = 1$.

2.1 A general type of recurrence

We consider a sequence of random, d-dimensional vectors $(Y_n)_{n\in\mathbb{N}_0}$ which satisfy the recurrence

$$Y_n \stackrel{d}{=} \sum_{r=1}^{K} A_r(n) Y^{(r)}_{I_r^{(n)}} + b_n, \quad n \geq n_0, \tag{1}$$

where $(A_1(n), \ldots, A_K(n), b_n, I^{(n)}), (Y_n^{(1)}), \ldots, (Y_n^{(K)})$ are independent, $A_1(n), \ldots, A_K(n)$ are random $d \times d$ matrices, b_n is a random d-dimensional vector, $I^{(n)}$ is a vector of random integers $I_r^{(n)} \in \{0, \ldots, n\}$, and $(Y_n^{(1)}), \ldots, (Y_n^{(K)})$ are identical distributed as (Y_n). The symbol $\stackrel{d}{=}$ denotes, that left and right hand side of equation (1) are identically distributed. We have $n_0 \geq 1$ and $Y_0, \ldots, Y_{n_0-1}$ are given initializing random vectors. The number $K \geq 1$ is deterministic. For random K, $K = K_n$ dependent on n or $K_n \to \infty$ the results presented subsequently can be generalized.

2.2 Examples from Computer Science

In this section we are looking at some applications from Computer Science and related areas, that are covered by the general equation (1) and to which we will come back later. We only give the relevant recursive equations and refer to where the algorithms and data structures are introduced in detail.

2.2.1 Quicksort

The number of key comparisons needed by the sorting algorithm Quicksort (Hoare (1962)) when applied to a uniform random permutation of length n (or applied to an arbitrary permutation with random uniform choice of the pivot element) satisfies recurrence (1) with $d = 1$, $K = 2$, $A_1(n) = A_2(n) = 1$, $I_1^{(n)}$ uniformly distributed on $\{0, \ldots, n-1\}$, $I_2^{(n)} = n - 1 - I_1^{(n)}$ and $b_n = n - 1$, cf. Mahmoud (2000). For this example the contraction method was originally introduced by Rösler (1991). The number of key exchanges of Quicksort when applied to a uniform random permutation of length n also satisfies (1), where the parameters d, K, $A_1(n)$. $A_2(n)$, $I^{(n)}$ are gives as for the key comparisons, however b_n now depends on $I^{(n)}$ as

$$\mathbb{P}(b_n = j \,|\, I_1^{(n)} = k) = \frac{\binom{k}{j}\binom{n-1-k}{j}}{\binom{n-1}{k}}, \quad 0 \le j \le k < n,$$

cf. Sedgewick (1980, page 55) and Hwang and Neininger (2002, section 6).

2.2.2 Valuations of random binary search trees

For the probabilistic analysis of binary search trees it is usually assumed that the tree is generated from an equiprobable random permutation. This is the model of the *random binary search tree*, cf. Mahmoud (1992, chapter 2). The internal path length of a random binary search tree with n internal nodes has the same distribution as the number of key comparisons of Quicksort and hence satisfies the recurrence of type (1) specified in section 2.2.1. It turns out that a couple of other parameters of random binary search trees satisfy the same recurrence, the only difference being the so-called toll function b_n which is specific for each parameter. For this reason Devroye (2002/03) and Hwang and Neininger (2002) consider general valuations of random binary search trees with n internal nodes, i.e., recurrence (1) with $d = 1$, $K = 2$, $A_1(n) = A_2(n) = 1$, $I_1^{(n)}$ uniformly distributed on $\{0, \dots, n-1\}$, $I_2^{(n)} = n - 1 - I_1^{(n)}$ and variable toll function b_n, which is only allowed to depend on $I^{(n)}$. Numerous quantities of random binary search trees with direct algorithmic interpretation, that are covered by this type of recurrence, are discussed in Hwang and Neininger (2002, section 6).

2.2.3 Depth of nodes in random binary search trees

The depth of a random (uniformly chosen) node in a random binary search tree describes the complexity for a typical (successful) search in the tree. The depth (in a tree with n internal nodes) satisfies recurrence (1) with $d = 1$, $K = 1$, $A_1(n) = 1$, $b_n = 1$ and

$$\mathbb{P}(I_1^{(n)} = k) = \begin{cases} \frac{1}{n} & \text{for } k = 0, \\ \frac{2k}{n^2} & \text{for } 1 \le k \le n-1, \end{cases}$$

cf. Mahmoud (1992, section 2.5) and Cramer and Rüschendorf (1996).

2.2.4 Wiener index of random binary search tree

The Wiener index of a connected graph is the sum of the distances between all pairs of nodes in the graph, where distance is the minimal number of edges connecting the nodes in the graph. The Wiener index has its origin in mathematical chemistry but has independently been investigated in graph theory, cf. Gutman and Polansky (1986), Trinajstić (1992) and Dobrynin, Entringer and Gutman (2001). The Wiener index of a random binary search tree does not satisfy recurrence (1) for (Y_n) with dimension $d = 1$. However, it can be covered by (1) in dimension $d = 2$ as follows, cf. Neininger (2002): We choose $d = 2$, $K = 2$, $I_1^{(n)}$ uniformly distributed on $\{0, \dots, n-1\}$, $I_2^{(n)} = n - 1 - I_1^{(n)}$ as well as

$$A_1(n) = \begin{bmatrix} 1 & n - I_1^{(n)} \\ 0 & 1 \end{bmatrix}, \quad A_2(n) = \begin{bmatrix} 1 & n - I_2^{(n)} \\ 0 & 1 \end{bmatrix}, \quad b_n = \begin{pmatrix} 2I_1^{(n)} I_2^{(n)} + n - 1 \\ n - 1 \end{pmatrix}.$$

The first component of Y_n then has the distribution of the Wiener index, the second component has the distribution of the internal path length of a random binary search tree with n internal nodes.

2.2.5 Size of random m-ary search trees

The size Y_n of random m-ary search trees, $m \geq 3$, (cf. Mahmoud (1992, chapter 3)) with n data inserted satisfies recurrence (1) with $d = 1$, $K = m$, $A_1(n) = \cdots = A_m(n) = 1$ and $b_n = 1$. We denote by $V = (U_{(1)}, U_{(2)} - U_{(1)}, \dots, 1 - U_{(m-1)})$ the vector of spacings between independent, uniform on $[0,1]$ distributed random $U_1, \dots, U_{m-1}$. With the order statistics $U_{(1)}, \dots, U_{(m-1)}$ we can hence write $S_1 = U_{(1)}$, $S_2 = U_{(2)} - U_{(1)}, \dots, S_{m-1} = U_{(m)} - U_{(m-1)}$, $S_m = 1 - U_{(m)}$. With this notation for $u \in [0,1]^m$ with $\sum_{r=1}^m u_r = 1$ the conditional distribution of $I^{(n)}$ given $V = u$ is multinomial:

$$\mathbb{P}^{I^{(n)} \mid V = u} = M(n - (m-1), u),$$

where $M(n, u)$ denotes the multinomial distribution with parameters n and u, cf. Mahmoud and Pittel (1989), Lew and Mahmoud (1994) and Chern and Hwang (2001b).

2.2.6 Size and path length of random tries

The size Y_n of a random trie with n data inserted satisfies in a standard model $Y_0 = 0$ and recurrence (1) with $d = 1$, $K = 2$, $A_1(n) = A_2(n) = 1$, and $I_1^{(n)}$ binomial $B(n, p)$ distributed and $I_2^{(n)} = n - I_1^{(n)}$. For $p = 1/2$ this is the symmetric Bernoulli model, for $p \neq 1/2$ the asymmetric Bernoulli model. For the (external) path length of a random trie we have the same recurrence as for the size, only $b_n = 1$ has to be changed to $b_n = n$. As for random binary search trees in section 2.2.2 we can consider general valuations of random tries (i.e. variable b_n) and cover further quantities relevant in applications, cf. Schachinger (2001). Parameters for digital search trees and Patricia tries can be covered by similar recurrences, also being of type (1), cf. Szpankowski (2001).

2.2.7 Mergesort

The number of key exchanges of the sorting algorithm Mergesort (in its "top-down" variant) applied to an equiprobable random permutation of length n satisfies recurrence (1) with $d = 1$, $K = 2$, $A_1(n) = A_2(n) = 1$, $I_1^{(n)} = \lceil n/2 \rceil$, $I_2^{(n)} = n - I_1^{(n)}$ and b_n a random variable, that is described in Knuth (1973, section 5.2.4), cf. Flajolet and Golin (1994), Hwang (1996, 1998), Cramer (1997) and Chen, Hwang and Chen (1999).

2.2.8 Randomized game tree evaluation

The complexity of algorithms to evaluate game trees is usually measured by the number of external nodes read by the algorithm, cf. Motwani and Raghavan (1995, chapter 2). For the randomized algorithm of Snir (1985) to evaluate game trees it can be shown that there are certain inputs, for which the complexity is maximized in stochastic order, cf. Ali Khan and Neininger (2004). This worst case complexity can be described by recurrence (1) with $d = 2$, $K = 4$,

$$A_1(n) = A_2(n) = \mathrm{Id}_2, \quad A_3(n) = \begin{bmatrix} B_1 B_2 & 0 \\ 1 - B_2 & 0 \end{bmatrix}, \quad A_4(n) = \begin{bmatrix} 0 & B_1 \\ B_1 & 0 \end{bmatrix},$$

where Id_2 is the 2×2 unity matrix and B_1, B_2 are independent Bernoulli $B(1/2)$ distributed random variables, $b_n = 0$ and $I_r^{(n)} = n/4$ for $r = 1, \ldots, 4$. The second component of Y_n then has the distribution of the worst case complexity in a binary game tree of corresponding height. The optimality of Snir's algorithm is discussed in Saks and Wigderson (1986), an alternative approach via 2-type Galton Watson processes can be found in Karp and Zhang (1995).

2.2.9 Maxima in right triangles

We consider the number of maxima of n independent, uniformly distributed points in the right triangle in $\mathbb{R}^2$ with vertices $(0,0), (1,0), (0,1)$. A point is maximal in a set of points, if there is no other point in the set with larger x and y coordinate. The number of maximal point satisfies recurrence (1) with $d = 1$, $K = 2$, $A_1(n) = A_2(n) = 1$, $b_n = 1$ and $I_1^{(n)}$, $I_2^{(n)}$ are given as follows as the first two components of a mixture of trinomial distributions: We denote by (U_n, V_n) the point in the given set of points, that maximizes the sum of its components. The vector $I^{(n)} = (I_1^{(n)}, I_2^{(n)}, I_3^{(n)})$ conditioned on $(U_n, V_n) = (u, v)$ has the trinomial distribution

$$\mathbb{P}^{I^{(n)} \,|\, (U_n, V_n) = (u,v)} = M\left(n-1, \frac{u^2}{(u+v)^2}, \frac{v^2}{(u+v)^2}, \frac{2uv}{(u+v)^2}\right),$$

cf. Bai et al. (2001, 2003). The case of the right triangle is crucial as more general convex polygons "without an upper right vertex" can be reduced to the case of the right triangle.

2.2.10 Size of critical Galton Watson processes

Yaglom's (1947) exponential limit law for the size of a critical Galton Watson process conditioned on surviving of the population can be covered with a variant of recurrence (1), where instead of K a random number K_n of summands, depending on n, is used with $((A_r(n))_{r\geq 1}, b_n, I^{(n)}, K_n), (Y_n^{(1)}), (Y_n^{(2)}), \dots$ being independent. The size of generation n (conditioned on survival) satisfies this variant of recurrence (1) with $d = 1$, $A_1(n) = A_2(n) = \cdots = 1$ and $b_n = 0$. Denote T_n the generation of the most recent common ancestor of the n-th population. Then K_n is the number of children of this ancestor, which have offspring that survives until generation n, and we have $I_1^{(n)} = I_2^{(n)} = \cdots = n - T_n$. This recursive approach has been developed for the case of finite variance of the offspring distribution in Geiger (2000), the case of an offspring distribution being in the domain of attraction of a α-stable law, $1 < \alpha \leq 2$, (first treated by Slack (1968) using generating functions), can be found in Kauffmann (2003, section 3.2).

2.2.11 Broadcast communication models

In Chen and Hwang (2003) cost parameters of two algorithms to identify maxima in broadcast communication models with n processors are analyzed. The running time of their algorithm B satisfies recurrence (1) with $d = 1$, $K = 1$, $A_1(n) = 1$, $I_1^{(n)}$ uniformly distributed on $\{0, \dots, n-1\}$ and b_n itself is the running time of another algorithm to solve the "leader election" algorithm. This "leader election" algorithm has been studied in Prodinger (1993) and Fill, Mahmoud, and Szpankowski (1996) and can also be covered by recurrence (1).

The number of key comparisons of algorithm A in Chen and Hwang (2003) satisfies recurrence (1) with $d = 1$, $K = 2$, $A_1(n) = A_2(n) = 1$, $(I_1^{(n)}, I_2^{(n)})$ has the distribution

$$\mathbb{P}\Big((I_1^{(n)}, I_2^{(n)}) = (j,k)\Big) = \begin{cases} 2^{-n}, & (j,k) = (0,0), \\ \binom{n-k-1}{j-1} 2^{-n}, & k \geq 0,\ 1 \leq j \leq n-k, \end{cases}$$

and we have $b_n = n - I_1^{(n)}$.

Further parameters of these two algorithms in Chen and Hwang (2003) can also be covered by recurrence (1).

2.2.12 Profile of random binary search trees

The profile of random binary search trees can be described with another variant of recurrence (1), that is given here explicitly. The profile $Y_{n,k}$, $n \geq 0$, $0 \leq k \leq n$ of random binary search trees with n data is the number of external nodes, that have distance k to the root of the tree. We have $Y_{n,0} = \delta_{n0}$ with Kronecker's δ and

$$Y_{n,k} \stackrel{d}{=} Y^{(1)}_{I_1^{(n)},k-1} + Y^{(2)}_{I_2^{(n)},k-1}, \quad n \geq 1,\ 1 \leq k \leq n,$$

where $I_1^{(n)}$ is uniformly distributed on $\{0, \ldots, n-1\}$, $I_2^{(n)} = n - 1 - I_1^{(n)}$ and independence properties as in (1), cf. Chauvin, Drmota and Jabbour-Hattab (2001), Drmota and Hwang (2005) and Fuchs, Hwang and Neininger (2004).

2.2.13 Further examples

Further important quantities in Computer Science and related areas, that can be covered by recurrence (1), are parameter of the selection algorithm Quickselect (also called Find, cf. Grübel and Rösler (1996), Kodaj and Móri (1997), Mahmoud et al. (1995), Hwang and Tsai (2001) and for further references see the survey article Rösler (2004)), multiple Quickselect (Mahmoud and Smythe (1998)) and Bucket Selection as well as the sorting algorithm Bucket Sorting (Mahmoud et al. (2000)), secondary cost measures of Quicksort (number of recursive calls, stack pushs and pops, cf. Neininger and Hwang (2002)), and variants of Quicksort (Chern et al. (2002)) as well as parameters of Quicksort with erroneous key comparisons (cf. Alonso et al. (2004)), parameters of random skip lists (cf. Pugh (1989), Papadakis et al. (1990) and Devroye (1992)), algorithms for listing the ideals of random posets (Janson (2002)), distances (and size of minimal spanning trees) in random binary search trees (cf. Mahmoud and Neininger (2003), Panholzer and Prodinger (2004) and Devroye and Neininger (2004)), number of occurring patterns in random binary search trees (cf. Devroye (1991) and Flajolet et al. (1997)), the length of a random external branch in a coalescent tree (Durrett (2002, page 162)), as well as all parameters mentioned before for binary search trees in other search trees, in particular in random recursive trees (cf. Smythe and Mahmoud (1995)), quad trees (cf. Flajolet et al. (1995)), m-ary search trees (cf. Chern and Hwang (2001b)), median of $(2t+1)$ search trees (cf. Chern and Hwang (2001a)), simplex trees (cf. Devroye (1999)), Catalan trees (cf. Flajolet and Odlyzko (1982), Kolchin (1986, chapter 2) and Fill and Kapur (2004a)) and universal split tree models (cf. Devroye (1999)).

3 The idea of the contraction method

We rescale the random vector Y_n in (1) by

$$X_n := C_n^{-1/2}(Y_n - M_n), \quad n \geq 0, \tag{2}$$

where $M_n \in \mathbb{R}^d$ and C_n is a symmetric, positive-definite $d \times d$ matrix. If the first two moments of Y_n are finite, then M_n and C_n are typically of the order of the expectation and the covariance matrix of Y_n respectively. For X_n from recurrence (1) we obtain the modified recurrence

$$X_n \stackrel{d}{=} \sum_{r=1}^{K} A_r^{(n)} X_{I_r^{(n)}}^{(r)} + b^{(n)}, \quad n \geq n_0, \tag{3}$$

with

$$A_r^{(n)} := C_n^{-1/2} A_r(n) C_{I_r^{(n)}}^{1/2}, \quad b^{(n)} := C_n^{-1/2}\left(b_n - M_n + \sum_{r=1}^{K} A_r(n) M_{I_r^{(n)}}\right) \tag{4}$$

and independence properties as in (1).

The contraction method aims to provide assertions of the following type:

> Appropriate convergence of the coefficients
>
> $$A_r^{(n)} \to A_r^*, \quad b^{(n)} \to b^*, \quad (n \to \infty) \tag{5}$$
>
> implies convergence in distribution of the quantities (X_n) to a limit X. The limit distribution $\mathcal{L}(X)$ is characterized by a fixed-point equation, which is obtained from the modified recurrence by letting formally $n \to \infty$:
>
> $$X \stackrel{d}{=} \sum_{r=1}^{K} A_r^* X^{(r)} + b^*. \tag{6}$$
>
> Here, $(A_1^*, \ldots, A_K^*, b^*)$, $X^{(1)}, \ldots, X^{(K)}$ are independent and $X^{(r)} \stackrel{d}{=} X$ for $r = 1, \ldots, K$.

To reformulate the fixed-point property, we denote by $\mathcal{M}^d$ the space of all probability measures on $\mathbb{R}^d$ and by T the measure valued map

$$T : \mathcal{M}^d \to \mathcal{M}^d, \quad \mu \mapsto \mathcal{L}\left(\sum_{r=1}^{K} A_r^* Z^{(r)} + b^*\right), \tag{7}$$

where $(A_1^*, \dots, A_K^*, b^*)$, $Z^{(1)}, \dots, Z^{(K)}$ are independent and $\mathcal{L}(Z^{(r)}) = \mu$ for $r = 1, \dots, K$. Then, X is a solution of the fixed-point equation (6) if and only if its distribution $\mathcal{L}(X)$ is a fixed-point of the map T.

Maps of type (7) often do not have unique fixed-points in the space of all probability distributions, and the characterization of the set of all fixed-points is up to a few special cases an open and important problem, cf. section 9.3.

Limit distributions that appear as such fixed-points in the analysis of algorithms are often distinguished by having finite absolute moments of some order. To make this more precise, we define the following subsets of $\mathcal{M}^d$:

$$\mathcal{M}_s^d := \{\mu \in \mathcal{M}^d : \|\mu\|_s < \infty\}, \quad s > 0, \tag{8}$$

$$\mathcal{M}_s^d(M) := \{\mu \in \mathcal{M}_s^d : \mathbb{E}\,\mu = M\}, \quad s \geq 1, \tag{9}$$

$$\mathcal{M}_s^d(M, C) := \{\mu \in \mathcal{M}_s^d(M) : \operatorname{Cov}(\mu) = C\}, \quad s \geq 2, \tag{10}$$

where $M \in \mathbb{R}^d$ and C is a symmetric, positive-definite $d \times d$ matrix, and where $\|\mu\|_s$, $\mathbb{E}\,\mu$ and $\operatorname{Cov}(\mu)$ denote the s-th absolute moment, expectation and covariance matrix of a random variable with distribution μ respectively.

The idea of the contraction method consists of endowing an appropriate subset $\mathcal{M}^* \subset \mathcal{M}^d$, e.g., on of the sets in (8)–(10), with a complete metric δ, such that the restriction of T to $\mathcal{M}^*$ is a contraction on the metric space $(\mathcal{M}^*, \delta)$ in the sense of Banach's fixed-point theorem. This implies the existence of a fixed-point $\mathcal{L}(X)$ of T being unique in $\mathcal{M}^*$. In a second step one shows convergence of the rescaled quantities $\mathcal{L}(X_n)$ to $\mathcal{L}(X)$ in the metric δ, $\delta(\mathcal{L}(X_n), \mathcal{L}(X)) \to 0$ for $n \to \infty$, based on appropriate convergence of the coefficients as in (5). If δ is chosen such that convergence in δ implies weak convergence, then the desired convergence in distribution follows.

4 The $\mathbf{L_2}$ realization of the idea

In this section we describe the realization of the idea of the contraction method using the minimal L_2 metric ℓ_2. This has been done for the one-dimensional case in Rösler (1991, 2001) and extends directly to general dimension d. Then, we come back to the examples from section 2.2 and discuss them as far as they can be treated with the L_2 realization of the contraction method.

4.1 A general $\mathbf{L_2}$ convergence theorem

The minimal L_p metrics ℓ_p are given for $p > 0$ by

$$\ell_p(\mu, \nu) := \inf \left\{ \|X - Y\|_p \ : \ \mathcal{L}(X) = \mu,\, \mathcal{L}(Y) = \nu \right\}, \quad \mu, \nu \in \mathcal{M}_p^d,$$

where $\|X\|_p := (\mathbb{E}\,\|X\|^p)^{(1/p)\wedge 1}$ denotes the L_p norm of a random vector X and $\|X\|$ denotes its Euclidean norm.

The spaces $(\mathcal{M}_p^d, \ell_p)$ for $p > 0$ as well as $(\mathcal{M}_p^d(M), \ell_p)$ for $M \in \mathbb{R}^d$, $p \geq 1$ are complete metric spaces and convergence in ℓ_p is equivalent to weak convergence plus convergence of the p-th absolute moment. For $\mu, \nu \in \mathcal{M}_p^d$ there always exist vectors X, Y on a joint probability space with $\mathcal{L}(X) = \mu$, $\mathcal{L}(Y) = \nu$ and $\ell_p(\mu, \nu) = \|X - Y\|_p$. Such vectors are called optimal couplings of μ and ν. For these and further properties of the minimal L_p metric ℓ_p see Dall'Aglio (1956), Major (1978), Bickel and Freedman (1981), Rachev (1991), and Rachev and Rüschendorf (1998).

In order to obtain contraction properties of the map (7) we denote by $\|A\|_{\mathrm{op}} := \sup_{\|x\|=1} \|Ax\|$ the operator norm of a square matrix A and by A^t the transposed of the matrix A.

Lemma 4.1 *Let $(A_1^*, \ldots, A_K^*, b^*)$ be an L_2-integrable vector of random $d \times d$ matrices $A_1^* \ldots, A_K^*$ and a random d-dimensional vector b^* with $\mathbb{E}\, b^* = 0$ and assume that T is as in (7). Then, the restriction of T to $\mathcal{M}_2^d(0)$ is Lipschitz continuous in ℓ_2, and for the Lipschitz constant* $\mathrm{lip}(T)$ *we have*

$$\mathrm{lip}(T) \leq \left\| \mathbb{E} \sum_{r=1}^{K} (A_r^*)^t A_r^* \right\|_{\mathrm{op}}^{1/2}. \qquad (11)$$

The proof can be given using optimal couplings, cf. Rösler (1992, 2001) for $d = 1$, Burton and Rösler (1995) for $K = 1$ and Neininger (2001, Lemma 3.1). If the right hand side of (11) is less than 1 then T has a unique fixed-point in $\mathcal{M}_2^d(0)$.

The second step of the contraction method is to show convergence in ℓ_2 for sequences $(\mathcal{L}(X_n))$ of form (3). This can be obtained under the following conditions; cf. Rösler (2001, Theorem 3.1) and Neininger (2001, Theorem 4.1).

Satz 4.2 *Assume (X_n) is a sequence of centered d-dimensional, L_2-integrable random vectors, satisfying the recurrence (3) with L_2-integrable random $d \times d$ matrices and a random L_2-integrable centered vector $b^{(n)}$. Assume that we have*

$$\left(A_1^{(n)}, \ldots, A_K^{(n)}, b^{(n)}\right) \xrightarrow{L_2} (A_1^*, \ldots, A_K^*, b^*), \quad (n \to \infty), \qquad (12)$$

$$\mathbb{E} \sum_{r=1}^{K} \left\|(A_r^*)^t A_r^*\right\|_{\mathrm{op}} < 1, \qquad (13)$$

$$\mathbb{E}\left[\mathbf{1}_{\{I_r^{(n)} \leq \ell\} \cup \{I_r^{(n)} = n\}} \left\|(A_r^{(n)})^t A_r^{(n)}\right\|_{\mathrm{op}}\right] \to 0, \quad (n \to \infty), \qquad (14)$$

for all $\ell \in \mathbb{N}$ and $r = 1, \ldots, K$. Then we have

$$\ell_2(\mathcal{L}(X_n), \mathcal{L}(X)) \to 0, \quad (n \to \infty),$$

where $\mathcal{L}(X)$ is the in $\mathcal{M}_2^d(0)$ unique fixed-point of map T in (7).

Condition (12) means that the convergence of the coefficients in (5) has to hold in L_2. For this we are allowed to construct $(A_1^*, \dots, A_K^*, b^*)$ according to $(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)})$ on a joint probability space, i.e., (12) means

$$\ell_2(\mathcal{L}(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)}), \mathcal{L}(A_1^*, \dots, A_K^*, b^*)) \to 0. \tag{15}$$

Condition (13), by Jensen's inequality, is stronger than the contraction condition

$$\left\| \mathbb{E} \sum_{r=1}^{K} (A_r^*)^t A_r^* \right\|_{\mathrm{op}} < 1 \tag{16}$$

from Lemma 4.1. Whether condition (13) in Theorem 4.2 can be replaced by the weaker condition (16) is unknown, cf. also the discussion in Neininger and Rüschendorf (2006).

Condition (14) is a technical condition, which in applications is usually easy to verify.

For the application of Theorem 4.2 to recursive sequences (Y_n) as in (1) one has to note, that for the scaling in (2) we have to choose $M_n = \mathbb{E}\, Y_n$ in order to guarantee the conditions $\mathbb{E}\, X_n = 0$ and $\mathbb{E}\, b^{(n)} = 0$. Since, on the other hand $b^{(n)}$ in (4) contains the quantities M_n and in (12) we need to derive a limit for $b^{(n)}$, this implies that for the application of Theorem 4.2 an asymptotic expansion of the mean $\mathbb{E}\, Y_n$ has to be known. In contrast, the covariance matrix $\mathrm{Cov}(Y_n)$ can be guessed in its first order asymptotic expansion such that Theorem 4.2 applies. Since convergence in ℓ_2 implies convergence of the second moment, Theorem 4.2 then automatically implies an asymptotic expansion of the covariance matrix $\mathrm{Cov}(Y_n)$.

4.2 Applications of the $\mathbf{L_2}$ setting

Now, we can discuss a couple of applications from section 2.2. As explained before in section 4.1, we will need an expansion of the first moment beforehand to do so.

4.2.1 Quicksort

For the number of key comparisons Y_n of Quicksort for the model discussed in section 2.2.1 we have

$$\mathbb{E}\, Y_n = 2(n+1)H_n - 4n = 2n \log n + c_p n + o(n), \tag{17}$$

where $H_n = \sum_{k=1}^{n} 1/k$ denotes the n-th harmonic number, log the natural logarithm and $c_p = 2\gamma - 4$ with the Euler-Mascheroni constant γ. This expansion of the expectation is sufficient for the application of Theorem 4.2. After rescaling and deriving the limits of the coefficients we obtain

$$\frac{Y_n - \mathbb{E}\, Y_n}{n} \xrightarrow{d} X,$$

where the limit equation for X is given by (6) and (7) respectively with $A_1^* = U$, $A_2^* = 1-U$ and $b^* = 1+2\mathcal{E}(U)$ with U uniformly distributed on $[0,1]$ and $\mathcal{E}(U) := U\log U + (1-U)\log(1-U)$, cf. Rösler (1991), and for alternative approaches Hennequin (1989, 1991) and Régnier (1989). The number of key exchanges can be treated analogously and leads to a limit equation, where, compared to the limit equation for the key comparisons, only b^* is changed to $b^* = U(1-U) + \mathcal{E}(U)/3$.

4.2.2 Valuations of random binary search trees

For the valuations Y_n of random binary search trees described in section 2.2.2 the application of Theorem 4.2 depends on the order of the toll term b_n. If $\mathbb{E}\, b_n = n^\alpha L(n) + o(n^\alpha L(n))$ for some $\alpha > 1/2$ and L a function being slowly varying at infinity, then the expectation of Y_n can be identified sufficiently precise to apply Theorem 4.2 under weak additionally conditions on b_n. This leads to a limit distribution that depends only on α, cf. Hwang and Neininger (2002). For smaller toll functions, e.g., with $\mathbb{E}\, b_n = O(\sqrt{n})$, Theorem 4.2 cannot be applied, cf. section 7.2.5.

4.2.3 Wiener index of random binary search trees

For the Wiener index W_n of random binary search trees from section 2.2.3 we have

$$\mathbb{E}\, W_n = 2n^2 H_n - 6n^2 + 8nH_n - 10n + 6H_n = 2n^2 \log n + c_w n^2 + o(n^2), \quad (18)$$

with $c_w = 2\gamma - 6$. This leads, together with the expansion of the mean of the internal path length in (17), to the applicability of Theorem 4.2 with limit equation (6) and (7) respectively given by

$$A_1^* = \begin{bmatrix} (1-U)^2 & U(1-U) \\ 0 & 1-U \end{bmatrix}, \quad A_2^* = \begin{bmatrix} (1-U)^2 & U(1-U) \\ 0 & 1-U \end{bmatrix},$$

$$b^* = \begin{pmatrix} 6U(1-U) + 2\,\mathcal{E}(U) \\ 1 + 2\,\mathcal{E}(U) \end{pmatrix},$$

where U and $\mathcal{E}(U)$ are as in section 4.2.1 , cf. Neininger (2002).

Interestingly, the contraction method can only be applied in this case since we have the relation $c_w = c_p - 2$ for the constants c_p and c_w from (17) and (18).

This suggests a similar relation for the corresponding constants for the universal split tree model in Devroye (1999):

$$c_w = c_p - \frac{b\,\mathbb{E}\,V^2}{1 - b\,\mathbb{E}\,V^2}, \tag{19}$$

where b denotes the branching degree and V the splitter (cf. Devroye (1999)). Relation (19) is conjectured for splitters that are "not too concentrated" and is motivated in Neininger (2002, page 596 top) by the fact that it is the only relation that would allow to apply the contraction method. One may be able to decide the validity of the conjecture in special cases using analytic tools, e.g., for random (point) quad trees with the tools in Flajolet et al. (1995), for m-ary search trees see Chern and Hwang (2001b), and for median of $2t+1$ trees see Chern and Hwang (2001a).

For the stochastic analysis of the Wiener index in other families of random trees (the "simply generated random trees") see Janson (2003).

4.2.4 Randomized game tree evaluation

For Y_n arising in the context of the worst case complexity of Snir's randomized game tree evaluation algorithm in section 2.2.8 the expectation can be derived exactly. The second component $Y_{n,2}$ of Y_n describes the worst case complexity and satisfies

$$\mathbb{E}\,Y_{n,1} = c_1 n^\alpha - c_2 n^\beta$$

with

$$\alpha = \log_2 \frac{1+\sqrt{33}}{4}, \quad \beta = \log_2 \frac{\sqrt{33}-1}{4}, \quad c_1 = \frac{1}{2} + \frac{7}{2\sqrt{33}}, \quad c_2 = c_1 - 1.$$

Together with $\mathbb{E}\,Y_{n,2}$ this is sufficient to apply Theorem 4.2, cf. Ali Khan and Neininger (2004), where mainly tail bounds of $Y_{n,1}$ are studied.

4.2.5 Size of critical Galton Watson trees

The approach to the size of critical Galton Watson trees (conditioned on survival) from section 2.2.10 via a variant of recurrence (1) leads in the case of finite variance of the offspring distribution to the applicability of an appropriate extension of Theorem 4.2. The exponential distribution (with parameter 1) comes up in this context as the unique fixed-point in $\mathcal{M}_2^1(1)$ of map (7) with $K = 2$, $A_1^* = A_2^* = U$ and $b^* = 0$ with U being uniformly distributed on $[0,1]$, cf. Geiger (2000), where also the ℓ_2 metric is used to show convergence. Offspring distributions, that are in the domain of attraction of an α-stable distribution with $1 < \alpha < 2$, cannot be covered by the L_2 setting. We will come back to this in section 7.2.6.

5 Limitations of the $\mathbf{L_2}$ setting

In this section we consider the univariate case $d = 1$. The contraction condition (16) and condition (13) are identical for $d = 1$:

$$\mathbb{E} \sum_{r=1}^{K} (A_r^*)^2 < 1. \tag{20}$$

Numerous applications lead to limit equations (6) of the form

$$X \stackrel{d}{=} \sum_{r=1}^{K} A_r^* X^{(r)} + b^*,$$

with

$$b^* = 0 \quad \text{and} \quad \sum_{r=1}^{K} (A_r^*)^2 = 1, \tag{21}$$

e.g., for the problems discussed in sections 2.2.2, 2.2.5–2.2.7 and 2.2.9. The coefficients $A_1^*, \ldots, A_K^*$ usually are still random quantities, however the sum of their squares is almost surely 1.

From the convolution property of the normal distribution follows directly, that normal distributed random variables are solutions of the limit equation (6) under (21).

If one excludes the degenerate cases, where the coefficients $A_1^*, \ldots, A_K^*$ do only have the values 0 and 1, then the set of fixed points of map T in (7) is exactly the set

$$\mathcal{F} := \{\mathcal{N}(0, \sigma^2) \;:\; \sigma \geq 0\},$$

where $\mathcal{N}(0, \sigma^2)$ denotes the centered normal distribution with variance σ^2. (For $\sigma = 0$ this is interpreted as the Dirac measure in 0.) Since the conditions (20) and (21) cannot hold simultaneously, Theorem 4.2 cannot directly be applied. This is a fundamental problem of the contraction method: Fixed-points of a map T as in (7) with (21) are not unique in $\mathcal{M}_2^1(0)$, since $\mathcal{F} \subset \mathcal{M}_2^1(0)$. Hence T cannot be a contraction on $(\mathcal{M}_2^1(0), \delta)$ for any metric δ on $\mathcal{M}_2^1(0)$.

For the case of limit equation (6) with (21) it was known for some algorithmic problems (and proved by different methods) that in fact the normal distribution is the limit distribution. In section 7.1 we describe how a universal normal limit law (Corollary 7.5) in this direction can be obtained. For this, the space $\mathcal{M}_2^1(0)$ has to be refined in order to guarantee the uniqueness of fixed-points, and alternative metrics have to be used being ideal of an order larger than 2 so that contraction properties can be obtained.

Another problem of the L_2 setting consists in the fact that concrete problems in Computer Science and other fields lead to limit distributions that do not have a finite second moment. Examples are the profile of random binary search trees (section 2.2.12) or the size of critical Galton Watson trees conditioned on survival with an offspring distribution with infinite variance (section 2.2.10). This problem cannot be surmounted by using instead of the ℓ_2 metric a minimal L_p metric ℓ_p with some $p < 2$, since the corresponding limit maps T have no contraction properties in these metrics. The approach developed in section 7.1 based on ideal metrics is sufficiently flexible to cover applications that have emerged so far, cf. sections 7.2.6 and 7.2.7.

A further principal problem of the contraction method that is independent of the metric used are degenerate limit equations. To see this we consider the simplest case $d = 1$, $K = 1$ and $A_1(n) = 1$, which often appears in applications, cf. for example section 2.2.3. The recurrence (1) for (Y_n) now has, with $I_n = I_1^{(n)}$, the form

$$Y_n \stackrel{d}{=} Y_{I_n} + b_n, \quad n \geq n_0.$$

Rescaling $X_n := (Y_n - \mu_n)/\sigma_n$ as in (2) yields the modified recurrence

$$X_n \stackrel{d}{=} \frac{\sigma_{I_n}}{\sigma_n} X_{I_n} + b^{(n)}, \quad n \geq n_0,$$

with

$$b^{(n)} := \frac{1}{\sigma_n} \left(b_n - \mu_n + \mu_{I_n} \right).$$

Typically, with a reasonable choice of μ_n and σ_n, the limits of the coefficients can be derived as

$$\frac{\sigma_{I_n}}{\sigma_n} \to A^*, \quad b^{(n)} \to b^*$$

leading to the limit equation

$$X \stackrel{d}{=} A^* X + b^*,$$

with (A^*, b^*) and X being independent.

In a series of applications one will be led to the case $A^* = 1$ and $b^* = 0$, i.e., one has the limit equation

$$X \stackrel{d}{=} X. \tag{22}$$

The degenerate limit equation does not give any information about a potential limit distribution and the concept of the contraction method needs to be significantly extended to deal with such cases. In the analysis of algorithms these cases occur naturally when $\mathrm{Var}(Y_n) = L(n) + o(L(n))$ with a function L being slowly varying at infinity. Since we essentially have to choose $\sigma_n = \sqrt{\mathrm{Var}(Y_n)}$, one obtains

$$\frac{\sigma_{I_n}}{\sigma_n} = \sqrt{\frac{L(I_n)}{L(n)}} + o(1) \to A^* = 1, \quad (n \to \infty),$$

almost surely under fairly general conditions on I_n. If, furthermore, b_n is sufficiently small, then $b^{(n)} = \frac{1}{\sigma_n}(b_n - \mu_n + \mu_{I_n}) \to 0$ almost surely and we are led to a degenerate limit equation. We discuss in section 8 a universal central limit law that can be applied to some problems occurring in applications from Computer Science.

6 Ideal metrics — the Zolotarev metric

A probability metric $\tau(X, Y)$ is defined for random vectors X, Y and depends on the joint distribution $\mathcal{L}(X, Y)$ of X and Y. A probability metric is called simple, if $\tau(X, Y) = \tau(\mathcal{L}(X), \mathcal{L}(Y))$ does only depend on the marginals $\mathcal{L}(X)$, $\mathcal{L}(Y)$ of X and Y.

Simple probability metrics on the space of pairs of random variable induce a metric on $\mathcal{M}^d$ or on subspaces, where they are finite. Only such metrics will be considered subsequently. A simple probability metric τ is called $(s, +)$ ideal (or ideal of order $s > 0$), if

$$\tau(X + Z, Y + Z) \le \tau(X, Y)$$

for all Z being independent of (X, Y) and

$$\tau(cX, cY) = c^s \tau(X, Y).$$

for all $c > 0$.

The following realization of the idea of the contraction method is based on $(s, +)$ ideal metrics. It turns out that the flexibility in the index s allows to solve various of the problems described in section 5. Also numerous problems from Computer Science can universally be treated by these metrics.

Zolotarev (1976) constructed for random vectors X, Y in $\mathbb{R}^d$ the probability metrics

$$\zeta_s(X, Y) = \sup_{f \in \mathcal{F}_s} |\,\mathbb{E}\,[f(X) - f(Y)]| \tag{23}$$

where $s = m + \alpha$ with $0 < \alpha \le 1$, $m \in \mathbb{N}_0$, and

$$\mathcal{F}_s := \{f \in C^m(\mathbb{R}^d, \mathbb{R}) : \|f^{(m)}(x) - f^{(m)}(y)\| \le \|x - y\|^\alpha\},$$

the space of m times continuously differentiable functions from $\mathbb{R}^d$ to $\mathbb{R}$ such that the m-th derivative is Hölder continuous of order α. We use the short notation

$\zeta_s(\mathcal{L}(X), \mathcal{L}(Y)) = \zeta_s(X, Y)$. We have that $\zeta_s(X, Y) < \infty$, if all mixed moments of orders $1, \ldots, m$ of X and Y are equal and if the s-th absolute moments of X and Y are finite. Furthermore, $(\mathcal{M}_s^d, \zeta_s)$ for $0 < s \leq 1$, $(\mathcal{M}_s^d(M), \zeta_s)$ for $1 < s \leq 2$ and $(\mathcal{M}_s^d(M, C), \zeta_s)$ for $2 < s \leq 3$ are metric spaces with $M \in \mathbb{R}^d$ and C begin a symmetric, positive definite $d \times d$ matrix. The metric ζ_s is $(s, +)$ ideal and convergence in ζ_s implies weak convergence. For $d \times d$ matrices A we have

$$\zeta_s(AX, AY) \leq \|A\|_{\text{op}}^s \zeta_s(X, Y).$$

For additional properties of the Zolotarev metrics, in particular lower and upper bounds, see Zolotarev (1976, 1977), Rachev (1991) and Neininger and Rüschendorf (2004a).

Subsequently we will exclusively use the Zolotarev metrics since for this metric useful properties are established in the literature. However, crucial for the approach is that the metric is $(s, +)$ ideal. For this reason we could in principle use other $(s, +)$ ideal metrics; for examples for the application of alternative ideal metrics in the context of the analysis of algorithms see Rachev and Rüschendorf (1995), Cramer (1995) and Hwang and Neininger (2002).

7 The ζ_s realization of the idea

In this section the idea of the contraction method is developed based on the Zolotarev metric as developed in Neininger and Rüschendorf (2004a). First, a general contraction theorem is discussed that later is specified in different directions. With this, in particular the problem of asymptotic normality as discussed in section 5 will be resolved. Further examples of section 2.2 will then be discussed.

7.1 A general contraction theorem in ζ_s

First we develop conditions for contraction of map T in (7) with respect to the Zolotarev metric.

Lemma 7.1 *Assume that $(A_1^*, \ldots, A_K^*, b^*)$ is an L^s-integrable vector, $s > 0$, of random $d \times d$ matrices $A_1^* \ldots, A_K^*$ and a random d-dimensional vector b^*. Denote by T the map given in (7). For $\mu, \nu \in \mathcal{M}_s^d$ with identical mixed moments of orders $1, \ldots, m$ we have*

$$\zeta_s(T(\mu), T(\nu)) \leq \left(\mathbb{E} \sum_{r=1}^{K} \|A_r^*\|_{\text{op}}^s \right) \zeta_s(\mu, \nu).$$

In order to obtain contraction of T on a metric space $(\mathcal{M}^*, \zeta_s)$ with $\mathcal{M}^* \subset \mathcal{M}^d$ we have to ensure that ζ_s is finite on $\mathcal{M}^*$ and that we have $T(\mathcal{M}^*) \subset \mathcal{M}^*$.

For the finiteness of ζ_s we note that be the rescaling in (2) we can only control the first two (mixed) moments. Thus, finiteness of $\zeta_s(X_n, X)$ with a general X_n as in (2) and a fixed-point $\mathcal{L}(X)$ of map T can in general only be achieved for $0 < s \leq 3$.

To show that T maps $\mathcal{M}^*$ into itself we make use of the spaces in (8)–(10) as in section 6: $\mathcal{M}^* = \mathcal{M}_s^d$ for $0 < s \leq 1$, $\mathcal{M}^* = \mathcal{M}_s^d(M)$ for $1 < s \leq 2$ and $\mathcal{M}^* = \mathcal{M}_s^d(M, C)$ for $2 < s \leq 3$. For the property that T maps $\mathcal{M}^*$ into itself we then need for the cases $1 < s \leq 2$ and $2 < s \leq 3$ conditions on M, C and $(A_1^*, \dots, A_K^*, b^*)$; cf. Neininger and Rüschendorf (2004a, Lemma 3.2). By choosing M_n and C_n in (2) appropriately we can always reach the case $M = 0$ and $C = \mathrm{Id}_d$ where Id_d denotes the d-dimensional unity matrix. Then as a result of these considerations we obtain the following Lemma on contraction of T in ζ_s and the existence of fixed-points of T (Neininger and Rüschendorf (2004a, Corollary 3.4)):

Lemma 7.2 *Assume $(A_1^*, \dots, A_K^*, b^*)$ and T are as in (7) with L^s-integrable $(A_1^*, \dots, A_K^*, b^*)$, $0 < s \leq 3$, and $\mathbb{E} \sum_{r=1}^K \|A_r^*\|_{\mathrm{op}}^s < 1$. If*

$$\begin{cases} \mathbb{E}\, b^* = 0 & \text{for } 1 < s \leq 2, \\ \mathbb{E}\, b^* = 0 \text{ and } \mathbb{E}\,[b^*(b^*)^t] + \mathbb{E} \sum_{r=1}^K A_r^*(A_r^*)^t = \mathrm{Id}_d & \text{for } 2 < s \leq 3, \end{cases}$$

then T has a fixed-point, which is unique in

$$\begin{cases} \mathcal{M}_s^d & \text{for } 0 < s \leq 1, \\ \mathcal{M}_s^d(0) & \text{for } 1 < s \leq 2, \\ \mathcal{M}_s^d(0, \mathrm{Id}_d) & \text{for } 2 < s \leq 3. \end{cases}$$

This Lemma reduces in the univariate case for $s = 2$ to the assertion of Lemma 4.1. The advantage of the additional flexibility in $0 < s \leq 3$ will appear later.

For our convergence result that will extend Theorem 4.2, we consider a sequence (Y_n) as in (1) with scaled quantities (X_n) as in (2). We have to distinguish the three cases $0 < s \leq 1$, $1 < s \leq 2$ and $2 < s \leq 3$. For the case $2 < s \leq 3$ we assume additionally that Y_n has a regular covariance matrix for all $n \geq n_1$ with an $n_1 \geq n_0$. For the scaling in (2) we assume subsequently that

$$\left.\begin{cases} C_n \text{ symmetric, positive definite} & \text{for } 0 < s \leq 1, \\ M_n = \mathbb{E}\, Y_n \text{ and } C_n \text{ symmetric, positive definite} & \text{for } 1 < s \leq 2, \\ M_n = \mathbb{E}\, Y_n \text{ and } C_n = \mathrm{Cov}(Y_n) \text{ for } n \geq n_1 & \text{for } 2 < s \leq 3. \end{cases}\right\} \quad (24)$$

Then, the following theorem holds (Neininger and Rüschendorf (2004a, Theorem 4.1)):

Satz 7.3 *Assume that $0 < s \leq 3$ and (Y_n) is a sequence of L^s-integrable random vectors as in (1) with all random matrices and vectors there being L^s-integrable. Assume that (X_n) is the rescaled sequence according to (2) with condition (24) If*

$$\left(A_1^{(n)}, \ldots, A_K^{(n)}, b^{(n)}\right) \xrightarrow{L_s} \left(A_1^*, \ldots, A_K^*, b^*\right), \quad (n \to \infty) \tag{25}$$

$$\mathbb{E} \sum_{r=1}^{K} \|A_r^*\|_{\text{op}}^s < 1, \tag{26}$$

$$\mathbb{E}\left[\mathbf{1}_{\{I_r^{(n)} \leq \ell\} \cup \{I_r^{(n)} = n\}} \|A_r^{(n)}\|_{\text{op}}^s\right] \to 0, \quad (n \to \infty) \tag{27}$$

for all $\ell \in \mathbb{N}$ and $r = 1 \ldots, K$, then we have

$$\zeta_s(\mathcal{L}(X_n), \mathcal{L}(X)) \to 0, \quad (n \to \infty),$$

where $\mathcal{L}(X)$ is the in Lemma 7.2 described unique fixed-point of map T in (7).

Condition (25) is meant analogously to condition (12), see (15).

The advantage of Theorem 7.3 is the flexibility in the parameter s. From the point of view of applications the three cases $0 < s \leq 1$, $1 < s \leq 2$ and $2 < s \leq 3$ are substantially different. In case $2 < s \leq 3$ the condition $\mathcal{L}(X_n) \in \mathcal{M}_s^d(0, \mathrm{Id}_d)$ requires that the original sequence (Y_n) in (2) is scaled with the exact mean and covariance matrix, cf. (24). In order to obtain the limits of $(A_1^{(n)}, \ldots, A_K^{(n)}, b^{(n)})$ in (25) one has to draw back to the representations of $A_r^{(n)}$ and $b^{(n)}$ in (4) which involve the quantities M_n and C_n. Hence, for the application of Theorem 7.3 with $2 < s \leq 3$ expansions of $\mathbb{E}\, Y_n$ and $\mathrm{Cov}(Y_n)$ need to be known in advance. This is different from the cases $s \leq 2$. For $1 < s \leq 2$ we only need to ensure $\mathcal{L}(X_n) \in \mathcal{M}_s^d(0)$. This implies that for the application of the theorem only an expansion of the mean $M_n = \mathbb{E}\, Y_n$ has to be known in advance. An expansion of the covariance matrix can be guessed in first order and can be verified by Theorem 7.3 together with weak convergence. If Theorem 7.3 can be applied with $s \leq 1$ then no information about moments needs to be known in advance. An appropriate application of the theorem with $s = 1$ implies an expansion of the expectation.

To get a first glance of the applicability of Theorem 7.3 we consider in recurrence (1) for (Y_n) as special case the univariate case $d = 1$ with $A_1(n) = \cdots = A_K(n) = 1$. This case is frequent in applications,

$$Y_n \stackrel{d}{=} \sum_{r=1}^{K} Y_{I_r^{(n)}}^{(r)} + b_n, \quad n \geq n_0, \tag{28}$$

with conditions as in (1) and $\mathrm{Var}(Y_n) > 0$ for all $n \geq n_1 \geq n_0$. We assume that there are functions $f, g : \mathbb{N}_0 \to \mathbb{R}_0^+$ with $g(n) > 0$ for all sufficiently large n such that

$$\left(\frac{g(I_r^{(n)})}{g(n)}\right)^{1/2} \xrightarrow{L_s} A_r^* \text{ for } r = 1, \ldots, K \text{ with } \mathbb{E} \sum_{r=1}^{K} (A_r^*)^s < 1, \tag{29}$$

and

$$\frac{1}{g^{1/2}(n)} \left(b_n - f(n) + \sum_{r=1}^{K} f(I_r^{(n)}) \right) \xrightarrow{L_s} b^*. \tag{30}$$

Satz 7.4 *Assume that* $0 < s \leq 3$ *and that* (Y_n) *is a sequence of* L^s*-integrable random variables with (28) and* b_n *there is* L^s*-integrable. Assume that* f *and* g *are functions with (29) and (30). If*

$$\begin{cases} \mathbb{E}\, Y_n = f(n) + o(g^{1/2}(n)) & \text{for } 1 < s \leq 2, \\ \mathbb{E}\, Y_n = f(n) + o(g^{1/2}(n)) \text{ and } \mathrm{Var}(Y_n) = g(n) + o(g(n)) & \text{for } 2 < s \leq 3, \end{cases}$$

then

$$\frac{Y_n - f(n)}{g^{1/2}(n)} \xrightarrow{d} X,$$

where $\mathcal{L}(X)$ *is the in Lemma 7.2 described unique fixed-point of map* T *in (7) and* $\xrightarrow{d}$ *denotes convergence in distribution.*

In Theorem 7.4 we see, up to which extent expansions of the moments of Y_n are needed in advance in the cases $2 < s \leq 3$ and $1 < s \leq 2$. A further specialization yields a solution to one of the problems of the L_2 setting in section 5.

Korollar 7.5 *Assume that* (Y_n) *is a sequence of* L^s*-integrable random variables as in (28) with* b_n *there being* L^s*-integrable. Assume that we have* $\mathbb{E}\, Y_n = f(n) + o(g^{1/2}(n))$ *and* $\mathrm{Var}(Y_n) = g(n) + o(g(n))$ *and for some* $2 < s \leq 3$

$$\left(\frac{g(I_r^{(n)})}{g(n)}\right)^{1/2} \xrightarrow{L_s} A_r^*, \quad \sum_{r=1}^{K} (A_r^*)^2 = 1, \quad \mathbb{P}\left(\bigcap_{r=1}^{K} \{A_r^* \in \{0,1\}\} \right) < 1$$

and

$$\frac{1}{g^{1/2}(n)} \left(b_n - f(n) + \sum_{r=1}^{K} f(I_r^{(n)}) \right) \xrightarrow{L_s} 0.$$

Then we have

$$\frac{Y_n - f(n)}{g^{1/2}(n)} \xrightarrow{d} \mathcal{N}(0,1).$$

It turns out that a large number of applications in the analysis of algorithms and random trees that leads to asymptotic normality can be covered by Corollary 7.5, cf. section 7.2. These applications lead in Corollary 7.5 to an application of the ζ_s metric for $2 < s \leq 3$. As explained before, this requires that asymptotic expansions for mean and variance are known in advance. This somehow corresponds to a heuristic principle formulated by Pittel (1999) "Normal convergence problem? Two moments and a recurrence may be the clues".

The value of different metrics from the point of view of applications to recurrences of type (1) can be measured by the corresponding contraction properties of the map in (7). In dimension $d = 1$, e.g., we have for ℓ_2 as contraction condition in Theorem 4.2

$$\mathbb{E} \sum_{r=1}^{K} (A_r^*)^2 < 1.$$

For general ℓ_p, $p > 0$, we could formulate a corresponding result under the contraction condition

$$\sum_{r=1}^{K} \|A_r^*\|_p < 1. \tag{31}$$

It is known that the map in (7) on $\mathcal{M}_p^1(0)$ also is a contraction with respect to ℓ_p under condition

$$\mathbb{E} \sum_{r=1}^{K} |A_r^*|^p < 1. \tag{32}$$

However, it is unknown (but likely) whether this condition also implies a convergence result corresponding to Theorem 4.2.

For the Zolotarev metric ζ_s we have contraction in the mentioned spaces under the contraction condition

$$\mathbb{E} \sum_{r=1}^{K} |A_r^*|^s < 1. \tag{33}$$

In typical algorithmic applications we have $A_1(n) = \cdots = A_K(n) = 1$ and the variance of Y_n is often (up to periodic prefactors) monotone increasing. In these cases this implies that the corresponding limit equation has coefficients $A_1^*, \ldots, A_K^*$ with $0 \leq A_r^* \leq 1$ for $r = 1, \ldots, K$. The contraction conditions (32) and (33) hence are becoming weaker as s and p respectively are increased. As noted after Theorem 7.3, for more comfortable contraction conditions for increasing s we have to pay with information about the moments of Y_n that has to be known in advance to apply the method. Hence, there is a tradeoff between contraction condition and input of moments. From this perspective the condition (31) is less useful. The L_2 setting of section 4.1 in dimension $d = 1$ is about a mighty as Theorem 7.3 with $s = 2$; for $d \geq 2$ however there are differences, cf. Neininger and Rüschendorf (2003).

An interesting limit equation with coefficients $A_1^*, \ldots, A_K^*$ taking values larger than 1, will appear in (37) for the profile of random binary search trees in section 7.2.7. The special feature in that recurrence is the presence of a second index k. In the area of the analysis of algorithms profiles of random split tree models are the first examples of such limit equations.

7.2 Applications of the ζ_s setting

With Theorem 7.3 we can discuss further examples from section 7.3. We start with cases leading to asymptotic normality that are covered by Corollary 7.5.

7.2.1 Size of random m-ary search trees

For the size Y_n of random m-ary search trees, expectation and variance satisfy, for the cases $3 \leq m \leq 26$, the expansions

$$\mathbb{E}\, Y_n = \frac{1}{2(H_m - 1)} n + O(1 + n^{\alpha - 1}), \quad \mathrm{Var}(Y_n) = \gamma_m n + o(n),$$

with $\gamma_m > 0$ and $\alpha < 3/2$ depending on m. Corollary 7.5 can be applied with $f(n) = \frac{1}{2(H_m - 1)} n$ and $g(n) = \gamma_m n$. This yields a central limit law first obtained in Mahmoud and Pittel (1989) and Lew and Mahmoud (1994). For $m > 26$ Hwang and Chern (2001) showed, that the exactly rescaled quantities $(Y_n - \mathbb{E}\, Y_n)/\sqrt{\mathrm{Var}(Y_n)}$ do not converge in distribution at all due to periodic phenomena, cf. Chauvin and Pouyanne (2004) and Fill and Kapur (2004b).

7.2.2 Size and path length of random tries

For the quantities Y_n of random tries described in section 2.2.6 we have

$$\mathbb{E}\, Y_n = n\varpi_1(\log_2 n) + O(1), \quad \mathrm{Var}(Y_n) = n\varpi_2(\log_2 n) + O(1),$$

where ϖ_1, ϖ_2 are positive, infinitely differentiable functions with period 1. The conditions in Corollary 7.5 can be verified by direct computations, cf. Neininger and Rüschendorf (2004a, pages 403-406). This implies a central limit law due to Jacquet and Régnier (1986).

A more systematic study of the application of Corollary 7.5 shows that the properties of the periodic functions ϖ_1, ϖ_2 needed are only that they are strictly positive, that ϖ_1 is continuously differentiable and that ϖ_2 is continuous. Hence Corollary 7.5 can also be applied to other parameters of digital structures (i.e., tries, digital search trees, Patricia tries), where corresponding periodic functions appear. In particular the external (resp. internal) path length of tries (external), digital search trees (internal), where expansions of the first two moments are of this type can be covered by Corollary 7.5. For different approaches to these problems see Jacquet and Régnier (1986, 1988) and Schachinger (2001).

7.2.3 Mergesort

For the number of key comparisons of Mergesort discussed in section 2.2.7 we have the expansions

$$\begin{aligned} \mathbb{E}\, Y_n &= n \log_2 n + n\varpi_3(\log_2 n) + O(1), \\ \mathrm{Var}(Y_n) &= n\varpi_4(\log_2 n) + o(n), \end{aligned}$$

where ϖ_3, ϖ_4 are continuous functions with period 1, ϖ_4 is strictly positive, however, ϖ_3 is not differentiable. We have

$$\varpi_3(u) = C + \frac{1}{\log 2} \sum_{k \in \mathbb{Z}\setminus\{0\}} \frac{1 + \Psi(\chi_k)}{\chi_k(\chi_k + 1)} e^{2k\pi i u}, \quad u \in \mathbb{R}, \tag{34}$$

with a constant $C \in \mathbb{R}$, a complex function Ψ being $O(1)$ on the imaginary line $\Re(s) = 0$ and

$$\chi_k = \frac{2\pi i k}{\log 2}, \quad k \in \mathbb{Z},$$

cf. Flajolet and Golin (1994), where also a central limit law for Y_n is shown. Although ϖ_3 is not differentiable, we can apply Corollary 7.5 with

$$f(n) = n \log_2(n) + n\varpi_3(\log_2 n), \quad g(n) = n\varpi_4(\log_2 n).$$

For this, the special representation of ϖ_3 given in (34) can be exploited, cf. Neininger and Rüschendorf (2004a, pages 408-409).

7.2.4 Maxima in right triangles

The number of maxima Y_n in a random point set in a right triangle as described in section 2.2.9 satisfies

$$\mathbb{E}\, Y_n = \sqrt{\pi}\sqrt{n} + O(1), \quad \mathrm{Var}(Y_n) = \sigma^2\sqrt{n} + O(1)$$

with $\sigma^2 = (2\log 2 - 1)\sqrt{\pi}$, cf. Bai et al. (2001). The central limit theorem proven in Bai et al. (2001) by the method of moments can alternatively be obtained by verifying the conditions in Corollary 7.5, cf. Neininger and Rüschendorf (2004a, pages 410–411).

7.2.5 Valuations of random binary search trees

For the valuations of random binary search trees in section 2.2.2 we obtained in section 4.2.2 that for toll terms b_n with $\mathbb{E}\, b_n = n^\alpha L(n) + o(n^\alpha L(n))$, $\alpha > 1/2$ and L slowly varying at infinity the L_2 setting can be applied. For smaller toll terms with $\mathbb{E}\, b_n = O(\sqrt{n})$ we are lead to the fixed-point equation (21) with the problems discussed in section 5. It turns out that for such toll terms mean and variance on Y_n are asymptotically linear with error terms that allow to apply Corollary 7.5. This leads to asymptotic normality of the scaled quantities, cf. Hwang and Neininger (2002). An alternative approach to this problem based on Stein's method can be found in Devroye (2002/03).

7.2.6 Size of critical Galton Watson trees

In section 4.2.5 cases of the size of critical Galton Watson trees (conditioned on survival of the n-th generation) were left open, where the offspring distribution is in the domain of attraction of an α-stable distribution with $1 < \alpha < 2$. These can be treated with a variant of recurrence (1), cf. section 2.2.10 and lead to the fixed-point equation

$$X \stackrel{d}{=} U^{1/(\alpha-1)} \sum_{r=1}^{K} X^{(r)} \tag{35}$$

with U being uniformly distributed on $[0, 1]$ and K being random but independent of U and only depending on α. In particular, we have $\mathbb{E}\, K = \alpha/(\alpha - 1)$. The limit distribution has finite absolute moments only for orders less than α. On the other hand the fixed-point equation (35) leads to a contraction for all $1 < s < \alpha$ on $(\mathcal{M}_s^1(1), \zeta_s)$ since the Lipschitz constant here is bounded by

$$\mathbb{E}\,[KU^{s/(\alpha-1)}] = \frac{\alpha}{\alpha-1}\frac{\alpha-1}{s+\alpha-1} < 1.$$

A generalization of Theorem 7.3 to variants of recurrence (1) with random K, possibly depending on n, has been developed in Neininger and Rüschendorf (2004a, section 4.3). For the verification of the conditions corresponding to (25)–(27) and further details of the approach see Kauffmann (2003, section 3.2).

7.2.7 Profile of random binary search trees

For the profile $Y_{n,k}$ of random binary search trees as described in section 2.2.12 we have

$$\mathbb{E}\, Y_{n,k} = \frac{2^k}{n!} s(n,k) = \frac{(2\log n)^k}{\Gamma(k/\log n) k! n}\left(1 + O\left(\frac{1}{\log n}\right)\right), \tag{36}$$

where $s(n,k)$ are the (sign-less) Stirling numbers of first kind. For $k = \alpha \log n + o(\log n)$ this implies

$$\frac{\log \mathbb{E}\, Y_{n,k}}{\log n} \to \lambda(\alpha) = \alpha - 1 - \alpha \log(\alpha/2).$$

It is known that the saturation level and the height of the random binary search tree are at levels $\alpha_- \log n$ and $\alpha_+ \log n$ respectively, cf. Devroye (1986, 1987), where $0 < \alpha_- < 2 < \alpha_+$ are the solutions of the equation

$$\alpha \log\left(\frac{2e}{\alpha}\right) = 1, \quad \alpha_- \doteq 0.373, \quad \alpha_+ \doteq 4.311.$$

Hence, the profile $Y_{n,k}$ is asymptotically non-deterministic at most for $k = \alpha \log n + o(\log n)$ with $\alpha \in (\alpha_-, \alpha_+)$.

The approach of the contraction method can still be applied. From the asymptotic expansion (36) we obtain for the scaled quantities $Y_{n,k}/\,\mathbb{E}\,Y_{n,k}$ with $k = \alpha \log n + o(\log n)$, $\alpha \in (\alpha_-, \alpha_+)$ the limit equation

$$X_\alpha \stackrel{d}{=} \frac{\alpha}{2} U^{\alpha-1} X_\alpha^{(1)} + \frac{\alpha}{2}(1-U)^{\alpha-1} X_\alpha^{(2)}, \tag{37}$$

where U is uniformly distributed on $[0,1]$.

This equation has a unique solution $\mathcal{L}(X_\alpha)$ in the space $\mathcal{M}_s^1(1)$, where we have $1 < s < \varrho$, and $\varrho = \varrho(\alpha)$ is given by $(\alpha - 1)\varrho + 1 = 2(\alpha/2)^\varrho$ for $\alpha \in (\alpha_-, \alpha_+) \setminus [1,2]$ and $\varrho = \infty$ for $\alpha \in [1,2]$. In particular, we have $\varrho < 2$ for $\alpha \in (\alpha_-, \alpha_+) \setminus [2-\sqrt{2}, 2+\sqrt{2}]$.

The distribution $\mathcal{L}(X_\alpha)$ has finite absolute moments of orders less than ϱ, but there is no finite absolute moment of order ϱ for $\alpha \in (\alpha_-, \alpha_+) \setminus [1,2]$. These properties indicate the ranges where different methods can be applied. The method of moments can be applied in the range $\alpha \in [1,2]$. The L_2 setting covers the range $\alpha \in [2-\sqrt{2}, 2+\sqrt{2}]$. With the methods of section 7.1 the whole range $\alpha \in (\alpha_-, \alpha_+)$ can universally been treated. For details and refined results for the cases $\alpha = 1$ and $\alpha = 2$ cf. Fuchs, Hwang and Neininger (2004), for an earlier approach via martingales see Chauvin, Drmota and Jabbour-Hattab (2001).

8 Degenerate fixed-point equations

The phenomenon of degenerate fixed-point equations is more frequent than indicated in section 5, where we were only looking at cases with variances being slowly varying at infinity. Also, the normal distribution is by far not the only possible limit distribution in this context.

Nevertheless, we discuss a central limit law from Neininger and Rüschendorf (2004b) that covers typical cases as described in section 5 uniformly.

8.1 A general central limit law

We consider recurrence (1) for (Y_n) with $d = 1$, $K = 1$ and $A_1(n) = 1$,

$$Y_n \stackrel{\mathcal{L}}{=} Y_{I_n} + b_n, \quad n \geq n_0, \tag{38}$$

with the notation $I_n = I_1^{(n)}$ and the assumption that $\mathbb{P}(I_n = n) < 1$ for $n \geq n_0$.

We denote $\sigma_n = \sqrt{\operatorname{Var}(Y_n)}$ and $\mu_n = \mathbb{E}\, Y_n$, and write $\log^\alpha n := (\log n)^\alpha$ for $\alpha > 0$ and $n \geq 1$. Then we have (Neininger and Rüschendorf (2004b, Theorem 2.1)):

Satz 8.1 *Assume the sequence $(Y_n)_{n\geq 0}$ satisfies (38) with $\|Y_n\|_3 < \infty$ for all $n \geq 0$ and*

$$\limsup_{n\to\infty} \mathbb{E} \log\left(\frac{I_n \vee 1}{n}\right) < 0, \quad \sup_{n\geq 1} \left\|\log\left(\frac{I_n \vee 1}{n}\right)\right\|_3 < \infty. \tag{39}$$

Assume that there are numbers α, λ, κ with $0 \leq \lambda < 2\alpha$, such that expectation and variance of Y_n have the expansions

$$\|b_n - \mu_n + \mu_{I_n}\|_3 = O\left(\log^\kappa n\right) \quad \textit{and} \quad \sigma_n^2 = \sigma^2 \log^{2\alpha} n + O(\log^\lambda n) \tag{40}$$

with some constant $\sigma > 0$. If

$$\beta := \tfrac{3}{2} \wedge 3(\alpha - \kappa) \wedge 3(\alpha - \lambda/2) \wedge (\alpha - \kappa + 1) > 1,$$

then

$$\frac{Y_n - \mathbb{E}\, Y_n}{\sigma \log^\alpha n} \xrightarrow{d} \mathcal{N}(0,1),$$

with a rate of convergence in the Zolotarev metric

$$\zeta_3\left(\frac{Y_n - \mathbb{E}\, Y_n}{\sqrt{\operatorname{Var}(Y_n)}}, \mathcal{N}(0,1)\right) = O\left(\frac{1}{\log^{\beta-1} n}\right).$$

The first condition in (39) implies, that I_n is not too large, the second implies, that I_n is not too small.

The use of the Zolotarev metric is essential in the proof of the theorem, as it is not only used that ζ_3 is an $(3, +)$ ideal metric. The proof is based in the special definition (23) of the metric. In particular a Taylor expansion of the functions in $\mathcal{F}_3$ is necessary to obtain tight estimates.

A generalization of Theorem 8.1 to recurrences (1) with $K \geq 2$ is given in Neininger and Rüschendorf (2004b, section 5).

8.2 Applications

We discuss applications of Theorem 8.1, where the asymptotic normality can directly be obtained from expansions of the moments as in (40).

8.2.1 Depth of nodes in random binary search trees

For the depth Y_n of a random node in a random binary search tree with n internal nodes as discussed in section 2.2.3 we have

$$\mathbb{E}\, Y_n = 2\log n + O(1), \quad \mathrm{Var}(Y_n) = 2\log n + O(1),$$

cf. Mahmoud (1992). With the notation on Theorem 8.1 we obtain

$$\|b_n - \mu_n + \mu_{I_n}\|_3 = \|2\log(I_n/n) + O(1)\|_3 = O(1).$$

Hence, the parameters of Theorem 8.1 are given by $\alpha = 1/2$, $\kappa = \lambda = 0$ and we obtain $\beta = 3/2$. The technical conditions in (39) are satisfied, since $\log((I_n \vee 1)/n) \to \log U$ in $L_{3/2}$ for a random variable U uniformly distributed on $[0,1]$. Theorem 8.1 yields a central limit theorem with a rate of convergence in ζ_3, which is optimal as shown in Mahmoud and Neininger (2003, Theorem 1). For a direct stochastic argument to the asymptotic normality of the depth see Devroye (1988).

8.2.2 Broadcast communication models

For the time complexity of Algorithm B for finding maxima in broadcast communication models as discussed in section 2.2.11 Chen and Hwang (2003) showed

$$\mathbb{E}\, Y_n = \mu \log^2 n + O(\log n), \quad \mathrm{Var}(Y_n) = \sigma^2 \log^3 n + O(\log^2 n)$$

with positive constants μ and σ, as well as a central limit law. A direct calculation yields

$$\|b_n - \mu_n + \mu_{I_n}\|_3 = O(\log n).$$

Hence, in Theorem 8.1 the parameters are $\alpha = 3/2$, $\kappa = 1$, and $\lambda = 2$, i.e., $\beta = 3/2$. We rederive the asymptotic normality.

For the number of key comparisons of the alternative Algorithm A the mean is asymptotically linear, and we have (see Chen and Hwang (2003))

$$\mathbb{E}\, Y_n = n + \bar{\mu} \ln n + O(1), \quad \mathrm{Var}(Y_n) = \bar{\sigma}^2 \ln n + O(1),$$

with explicitly known constants $\bar{\mu}, \bar{\sigma} > 0$. A generalization of Theorem 8.1 can be applied to this recurrence. The leading linear terms in the mean cancel and we obtain

$$\begin{aligned} \|b_n - \mu_n + \mu_{I_1^{(n)}} + \mu_{I_2^{(n)}}\|_3 &= \|\bar{\mu}\ln((I_1^{(n)} \vee 1)/n) + I_2^{(n)} + \bar{\mu}\ln I_2^{(n)} + O(1)\|_3 \\ &= O(1). \end{aligned}$$

This implies $\alpha = 1/2$ and $\kappa = \lambda = 0$, thus $\beta = 3/2$. We obtain the asymptotic normality from Chen and Hwang (2003) with an additional rate of convergence.

8.2.3 Further applications

The analysis of further parameters of the algorithms discussed in section 8.2.2 also leads to degenerate limit equations,which can be covered by Theorem 8.1, cf. Neininger and Rüschendorf (2004b).

In Mahmoud (2003) one-sided recurrences for random binary search trees are studied. These are quantities that in a recursive formulation only draw back to one of the two subtrees of the root. Mahmoud (2003) demonstrates, how Theorem 8.1 can be used to reduce the problem of asymptotic normality to the calculation of expansions of the first two moments.

In Gnedin, Pitman and Yor (2004) the number of components of a regenerative composition structure are studied. As in Mahmoud (2003), the problem of asymptotic normality is reduced to the calculation of moments based on Theorem 8.1.

9 Related problems

In this section some selected related problems are mentioned. We illustrate these issues at the example of the number of key comparisons of the Quicksort algorithm in section 2.2.1.

9.1 Rates of convergence

For the L_2 setting in section 4.1 and for the ζ_s setting in section 7.1 rates of convergence for the number of key comparisons Y_n of Quicksort have been estimated. Fill and Janson (2002) find for the minimal L_p metrics ℓ_p the estimates

$$\ell_p\left(\mathcal{L}\left(\frac{Y_n - \mathbb{E}\,Y_n}{n}\right), \mathcal{L}(X)\right) = \begin{cases} O(1/\sqrt{n}), & p \geq 1, \\ \Omega(\log(n)/n), & p \geq 2, \end{cases}$$

and for the Kolmogorov metric ϱ

$$\varrho\left(\mathcal{L}\left(\frac{Y_n - \mathbb{E}\,Y_n}{n}\right), \mathcal{L}(X)\right) = \begin{cases} O(n^{-1/2+\varepsilon}), & \varepsilon > 0, \\ \Omega(1/n), & \end{cases}$$

where $\mathcal{L}(X)$ is the fixed-point described in section 4.2.1. In Neininger and Rüschendorf (2002) the rate of convergence for the Zolotarev metric ζ_3 could be identified:

$$\zeta_3\left(\mathcal{L}\left(\frac{Y_n - \mathbb{E}\,Y_n}{\sqrt{\mathrm{Var}(Y_n)}}\right), \mathcal{L}(X/\sigma)\right) = \Theta\left(\frac{\log n}{n}\right),$$

with $\sigma = \sqrt{7 - 2\pi^2/3}$.

It is an open problem whether $\Theta(\log(n)/n)$ is also the correct rate of convergence for the Kolmogorov metric.

9.2 Large deviations

The study of large deviations from the expectation for quantities of type (1) is important for Computer Science, as such bounds quantify the probability of bad behavior of the algorithms. The first bounds for large deviations for the number of key comparisons for Quicksort going beyond trivial Chebychev-bounds based on moments, were given in Rösler (1991). There, the modified recurrence (3) and induction are used to prove a bound for the moment generating function. Then Chernoff's bounding technique implies bounds for large deviations.

Better bounds were given in McDiarmid and Hayward (1996) with a more combinatorial approach using the "method of bounded differences". These bounds were later also derived by Rösler's approach in Fill and Janson (2002), were estimates for the moment generating function are made explicit.

This inductive approach leads in Rösler (1991) and Fill and Janson (2002) to a certain function $f(K, \lambda, n)$, for which bounds are derived by taking multiple derivatives.

Estimating tail bounds of the worst case complexity of randomized game tree evaluation as discussed in section 2.2.8 leads to a similar problem in a two-dimensional setting. For a crucial estimate of the term corresponding to the function $f(K, \lambda, n)$ in Ali Khan and Neininger (2002) an idea of Bennett (1962) was helpful.

9.3 Solutions of fixed-point equations

A difficult problem is the characterization of the set of all fixed-points of a map (7) in the space $\mathcal{M}^d$. For the Quicksort case from section 4.2.1, Fill and Janson (2000b) could completely solve this problem. For general maps of type (7) see Holley and Liggett (1981), Durrett and Liggett (1983), Liu (1997, 1998), Alsmeyer and Rösler (2006), Caliebe (2003), Caliebe and Rösler (2003), Biggins and Kyprianou (2005) and the references in these papers.

9.4 Properties of fixed-points

In a limit law one wants to use the limit distribution to approximate the distributions of the finite problem. For many of the limit laws proved by the contraction method the limit distribution is only (implicitly) given by its fixed-point property. For this reason characteristic quantities of these distributions such as the distribution function or density are not directly amenable. Hence, it is challenging to derive properties of such fixed-points. For the Quicksort limit distribution $\mathcal{L}(X)$ from section 4.2.1 it is shown in Fill and Janson (2000a) that X has a bounded, infinitely differentiable Lebesgue-density that is rapidly decreasing. Furthermore, explicit bounds for the density f_X and its derivatives are derived. It is not known, whether f_X is unimodal, although simulations suggest a unimodal shape. It is also unknown, whether X is infinitely divisible. In Devroye, Fill and Neininger (2000) an algorithm for a uniform approximation of f_X was given, that is too slow to approximate the density practically. For analogous results for more general fixed-point equations see Devroye and Neininger (2002).

9.5 Simulation of fixed-points

For the approximative simulation of distributions given as fixed-points one can iterate the fixed-point equation to obtain a sequence that converges at geometric rate to the fixed-point. However, for maps as in (7) with $K \geq 2$, also the complexity is exponential in the number of iterations of the fixed-point equation.

For perfect simulation of such fixed-points for the example of the Quicksort limit distribution of section 4.2.1 an algorithm was proposed in Devroye, Fill and Neininger (2000). This algorithm is based on von Neumanns "rejection sampling". For the exact simulation of "Perpetuities" related to the selection algorithm Find, see Devroye (2001); for the exact simulation of a larger class of fixed-points, see Devroye and Neininger (2002).

9.6 Recurrences with maxima instead of sums

Many worst case parameters (Y_n) of recursive algorithms and discrete recursive structures satisfy recurrence (1), with the sum appearing there for summing the contributions of the quantities in the sub-structures replaced by a maximum,

$$Y_n \stackrel{d}{=} \bigvee_{r=1}^{K} \left(A_r(n) Y^{(r)}_{I_r^{(n)}} + b_r(n) \right), \quad n \geq n_0.$$

Here, $(A_1(n), \dots, A_K(n), b_1(n), \dots, b_K(n), I^{(n)})$, $(Y_n^{(1)}), \dots, (Y_n^{(K)})$ are independent, $b_1(n), \dots, b_K(n)$ are random variables and all remaining quantities are as in (1) with $d = 1$. A first convergence result in the style of Theorems 4.2 and 7.3, where the minimal L_p metric ℓ_p is used, is given in Neininger and Rüschendorf (2005). The question of the full space of fixed-points of the associated limit equation is discussed in Jagers and Rösler (2004).

References

[] Ali Khan, T. und Neininger, R. (2004) Probabilistic analysis for randomized game tree evaluation. *Mathematics and Computer Science III (Vienna 2004)*, 163-174, Trends in Mathematics, *Birkhäuser, Basel*.

[] Alonso, L., Chassaing, P., Gillet, F., Janson, S., Reingold, E. M. und Schott, R. (2004) Sorting with unreliable comparisons: a probabilistic analysis. *Combin. Probab. Comput.* **13**, 419-449.

[] Alsmeyer, G. und Rösler, U. (2006) A stochastic fixed point equation related to weighed branching with deterministic weights. *Electron. J. Probab.* **11**, 27–56 (electronic).

Preprint.

[] Bai, Z.-D., Hwang, H.-K., Liang, W.-Q. und Tsai, T.-H. (2001) Limit theorems for the number of maxima in random samples from planar regions. *Electron. J. Probab.* **6**, no. 3, 41 Seiten.

[] Bai, Z.-D., Hwang, H.-K. und Tsai, T.-H. (2003) Berry-Esseen bounds for the number of maxima in planar regions, *Electron. J. Probab.* **8**, no. 9, 26 Seiten.

[] Barbour, A. D., Holst, L. und Janson, S. (1992) Poisson approximation. Oxford Studies in Probability, 2. Oxford Science Publications. *The Clarendon Press, Oxford University Press, New York*.

[] Bennett, G. (1962) Probability inequalities for the sum of independent random variabels *J. Amer. Statist. Assoc.* **57**, 33–45.

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

[] Biggins, J. D. und Kyprianou, A. E. (2005) The smoothing transform: the boundary case. *Electron. J. Probab.* **10**, 609–631 (electronic).

Preprint.

[] Burton, R. M. und Rösler, U. (1995) An L_2 convergence theorem for random affine mappings. *J. Appl. Probab.* **32**, 183–192.

[] Caliebe, A. (2003) Symmetric fixed points od a smoothing transformation. *Adv. Appl. Probab.* **35**, 377–394.

[] Caliebe, A. und Rösler, U. (2003) Fixed points with finite variance of a smoothing transformation. *Stoch. Proc. Appl.* **107**, 105–129.

[] Chauvin, B., Drmota, M. und Jabbour-Hattab, J. (2001) The profile of binary search trees. *Ann. Appl. Probab.* **11**, 1042–1062.

[] Chauvin, B. und Pouyanne, N. (2004) m-ary search trees when $m \geq 27$: a strong asymptotics for the space requirements. *Random Structures Algorithms* **24**, 133–154.

[] Chauvin, B. und Rouault, A. (2004) Connecting Yule process, bisection and binary search tree via martingales. *Journal of the Iranian Statistical Society* **3**, 89–108.

[] Chen, W.-M. und Hwang, H.-K. (2003) Analysis in distribution of two randomized algorithms for finding the maximum in a broadcast communication model. *Journal of Algorithms* **46**, 140-177.

[] Chen, W.-M., Hwang, H.-K. and Chen, G.-H. (1999). The cost distribution of queue-mergesort, optimal mergesorts, and power-of-2 rules. *J. Algorithms* **30**, 423–448.

[] Chern, H.-H. und Hwang, H.-K. (2001a) Transitional behaviors of the average cost of quicksort with median-of-(2t+1). *Algorithmica* **29**, 44-69.

[] Chern, H.-H. und Hwang, H.-K. (2001b) Phase changes in random m-ary search trees and generalized quicksort. *Random Structures Algorithms* **19**, 316–358.

[] Chern, H.-H., Hwang, H.-K. und Tsai, T.-H. (2002) An asymptotic theory for Cauchy-Euler differential equations with applications to the analysis of algorithms. *J. Algorithms* **44**, 177-225.

[] Cormen, T. H., Leiserson, C. E. und Rivest, R. L. (1990) *Introduction to Algorithms*, MIT Press.

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

[] Cramer, M. (1997). Stochastic analysis of the Merge-Sort algorithm. *Random Structures Algorithms* **11**, 81–96.

[] Cramer, M. und Rüschendorf, L. (1996) Analysis of recursive algorithms by the contraction method. *Athens Conference on Applied Probability and Time Series Analysis*, Athens, Greece, 1995, Ed. by C. C. Heyde et al., Lecture Notes in Statistics **114**, 18-33.

[] Dall'Aglio, G. (1956). Sugli estremi dei momenti delle funzioni di ripartizione doppia. *Ann. Scuoloa Norm. Sup. Pisa (3) 10*, 35–74.

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

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

[] Devroye, L. (1988) Applications of the theory of records in the study of random trees. *Acta Inform.* **26** (1988), 123–130.

[] Devroye, L. (1991) Limit laws for local counters in random binary search trees. *Random Structures Algorithms* **2**, 303–315.

[] Devroye, L. (1992) A limit theory for random skip lists. *Ann. Appl. Probab.* **2**, 597–609.

[] Devroye, L. (1999) Universal limit laws for depths in random trees. *SIAM J. Comput.* **28**, 409–432.

[] Devroye, L. (2001) Simulating perpetuities. *Methodology Comput. Appl. Prob.* **3**, 97–115.

[] Devroye, L. (2002/03) Limit laws for sums of functions of subtrees of random binary search trees. *SIAM J. Comput.* **32**, 152–171.

[] Devroye, L., Fill, J. A. und Neininger, R. (2000) Perfect simulation from the Quicksort limit idstribution. *Electron. Commun. Probab.* **5**, 95–99.

[] Devroye, L. und Neininger, R. (2002) Density approximation and exact simulation of random variables that are solutions of fixed-point equations. *Adv. Appl. Probab.* **34**, 441–468.

[] Devroye, L. und Neininger, R. (2004) Distances and finger search in random binary search trees. *SIAM J. Comput.* **33**, 647–658.

[] Dobrynin, A. A., Entringer, R. und Gutman, I. (2001) Wiener index of trees: Theory and applications. *Acta Appl. Math.* **66**, 211–249.

[] Drmota, M. (1997) Systems of functional equations. *Random Structures Algorithms* **10**, 103–124.

[] Drmota, M. und Hwang, H.-K. (2005) Bimodality and phase transition in the profile variance of random binary search trees. *SIAM J. Discrete Math.* **19**, 19-45.

[] Durrett, R. (2002) *Probability models for DNA sequence evolution*, Probability and its Applications (New York). Springer-Verlag, New York.

[] Durrett, R. und Liggett, T. M. (1983) Fixed points of the smoothing transformation. *Z. Wahrsch. Verw. Gebiete* **64**, 275–301.

[] Fill, J. A. und Janson, S. (2000a) Smoothness and decay properties of the limiting quicksort density function. *Mathematics and computer science (Versailles, 2000)*, 53–64, Trends Math., *Birkhäuser, Basel.*

[] Fill, J. A. und Janson, S. (2000b) A characterization of the set of fixed points of the Quicksort transformation. *Electron. Comm. Probab.* **5**, 77–84.

[] Fill, J. A. und Janson, S. (2002) Quicksort asymptotics. *J. Algorithms* **44**, 4–28.

[] Fill, J. A. und Kapur, N. (2004a) Limiting Distributions for Additive Functionals on Catalan Trees. *Theoret. Comput. Sci.*, **326**, 69–102.

[] Fill, J. A. und Kapur, N. (2004b) The Space Requirement of m-ary Search Trees: Distributional Asymptotics for $m \geq 27$. *Proceedings of the 7th Iranian Statistical Conference.*

[] Fill, J. A., Mahmoud, H. M. und Szpankowski, W. (1996) On the distribution for the duration of a randomized leader election algorithm. *Ann. Appl. Probab.* **6**, 1260–1283.

[] Flajolet, P. und Golin, M. (1994). Mellin transforms and asymptotics. The mergesort recurrence. *Acta Inform. 31*, 673–696.

[] Flajolet, P., Gourdon, X. und Martínez, C. (1997) Patterns in random binary search trees. *Random Structures Algorithms*, **11**, 223–244.

[] Flajolet, P., Labelle, G., Laforest, L., und Salvy, B. (1995) Hypergeometrics and the cost structure of quadtrees. *Random Structures Algorithms*, **7**, 117–144.

[] Flajolet, P. and Odlyzko, A. (1982) The average height of binary trees and other simple trees. *J. Comput. System Sci.*, **25**, 171–213.

[] Flajolet, P. and Odlyzko, A. (1990) Singularity analysis of generating functions. *SIAM J. Discrete Math.*, **3**, 216–240.

[] Fuchs, M., Hwang, H.-K. und Neininger, R. (2004) Profiles of random trees: Limit theorems for random recursive trees and random binary search trees. *Algorithmica*, to appear.

[] Geiger, J. (2000) A new proof of Yaglom's exponential limit law. *Mathematics and computer science (Versailles, 2000)*, 245–249, Trends Math., *Birkhäuser, Basel.*

[] Gnedin, A., Pitman, J., and Yor, M. (2004) Asymptotic laws for the regenerative compositions: Gamma subordinators and the like. Preprint, `http://front.math.ucdavis.edu/math.PR/0408071`.

[] Gonnet, G. H. und Baeza-Yates, R. (1991) *Handbook of Algorithms and Data Structures*, Addison-Wesley, Workingham.

[] Grübel, R. und Rösler, U. (1996) Asymptotic distribution theory for Hoare's selection algorithm, *Adv. in Appl. Probab.* **28**, 252–269.

[] Gutman, I. und Polanski, O. E. (1986) Mathematical Concepts in Organic Chemistry, Springer, Berlin.

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

[] Hennequin, P. (1991) *Analyse en moyenne d'algorithme, tri rapide et arbres de recherche*, Ph.D. Thesis, *École Polytechnique.*

[] Hoare, C. A. R. (1962) Quicksort. *Comput. J.*, **5**, 10–15.

[] Holley, R. und Liggett, T. M. (1981) Generalized potlatch and smoothing processes. *Z. Wahrsch. Verw. Gebiete* **55**, 165–195.

[] Hwang, H.-K. (1996). Asymptotic expansions of the mergesort recurrences. *Acta Inform. 35*, 911–919.

[] Hwang, H.-K. (1998). Limit theorems for mergesort. *Random Structures Algorithms 8*, 319–336.

[] Hwang, H.-K. (2003) Second phase changes in random m-ary search trees and generalized quicksort: convergence rates. *Ann. Probab.* **31**, 609–629.

[] Hwang, H.-K. und Neininger, R. (2002) Phase change of limit laws in the quicksort recurrence under varying toll functions. *SIAM J. Comput.* **31**, 1687-1722.

[] Hwang, H.-K. und Tsai, T.-H. (2002) Quickselect and Dickman function, *Combin. Probab. Comput.* **11**, 353–371.

[] Jacquet, P. and M. Régnier (1986). Normal limiting distribution of the size and the external path length of tries. Technical report INRIA-Rocquencourt, RR-0827.

[] Jacquet, P. and M. Régnier (1988). Normal limiting distribution of the size of tries. *Performance '87 (Brussels, 1987)*, 209–223. North-Holland, Amsterdam.

[] Jagers, P. und Rösler, U. (2004) Stochastic fixed points involving the maximum. *Mathematics and computer science. III*, 325–338, Trends Math., *Birkhuser, Basel*, 2004.

[] Janson, S. (2002) Ideals in a forest, one-way infinite binary trees and the contraction method. *Mathematics and computer science, II (Versailles, 2002)*, 393–414, Trends Math., *Birkhäuser, Basel.*

[] Janson, S. (2003) The Wiener index of simply generated random trees. *Random Structures Algorithms* **22**, 337–358.

[] Janson, S., Łuczak, T., and Ruciński, A. (2000) *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization. *Wiley-Interscience, New York.*

[] Karp, R. and Zhang, Y. (1995) Bounded branching process and AND/OR tree evaluation. *Random Structures Algorithms* **7**, 97–116.

[] Kauffmann, L. (2003) Große Stammbäume. Dissertation, *Universität Frankfurt.*

[] Knuth, D. E. (1969a) The art of computer programming. Vol. 1: Fundamental algorithms. *Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont.*

[] Knuth, D. E. (1969b) The art of computer programming. Vol. 2: Seminumerical algorithms. *Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont.*

[] Knuth, D. E. (1973) The art of computer programming. Volume 3. Sorting and searching. Addison-Wesley Series in Computer Science and Information Processing. *Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont.*

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

[] Kolchin, V. F. (1986) Random mappings. With a foreword by S. R. S. Varadhan. Translation Series in Mathematics and Engineering. *Optimization Software, Inc., Publications Division, New York.*

[] Lew, W. and Mahmoud, H. M. (1994). The joint distribution of elastic buckets in multiway search trees. *SIAM J. Comput.* **23**, 1050–1074.

[] Liu, Q. (1997) Sur une équation fonctionnelle et ses applications: une extension du théorème de Kesten-Stigum concernant des processus de branchement. *Adv. in Appl. Probab.* **29**, 353–373.

[] Liu, Q. (1998) Fixed points of a generalized smoothing transformation and applications to the branching random walk. *Adv. in Appl. Probab.* **30**, 85–112.

[] Mahmoud, H. M. (1992) *Evolution of Random Search Trees.* John Wiley & Sons, New York.

[] Mahmoud, H. M. (2000) *Sorting. A Distribution Theory.* Wiley-Interscience, New York.

[] Mahmoud, H. M. (2003) One-sided variations on binary search trees. *Ann. Inst. Statist. Math.* **55**, 885–900.

[] Mahmoud, H. M., Flajolet, P., Jacquet, P. und Régnier, M. (2000) Analytic variations on bucket selection and sorting. *Acta Inform.* **36**, 735–760.

[] Mahmoud, H. M., Modarres, R. und Smythe, R. T. (1995) Analysis of quickselect: An algorithm for order statistics. *RAIRO Inform. Théor. Appl.* **29**, 255-276.

[] Mahmoud, H. M. und Neininger, R. (2003) Distribution of distances in random binary search trees. *Ann. Appl. Probab.* **13**, 253–276.

[] Mahmoud, H. M. und Pittel, B. (1989). Analysis of the space of search trees under the random insertion algorithm. *J. Algorithms* **10**, 52–75.

[] Mahmoud, H. M. und Smythe, R. T. (1998) Probabilistic analysis of Multiple Quick Select, *Algorithmica* **22**, 569–584.

[] Major, P. (1978). On the invariance principle for sums of independent identically distributed random variables. *J. Multivariate Anal.* **8**, 487–517.

[] McDiarmid, C. J. H. und Hayward, R. B. (1996) Large deviations for Quicksort. *J. Algorithms* **21**, 476–507.

[] Morrison, P. and Morrison P. (1999) Scientists' Bookshelf: 100 or so Books that shaped a Century of Science. *American Scientist* **87**, no. 6 (Nov–Dec issue).

[] Motwani, R. and Raghavan, P. (1995) *Randomized algorithms.* Cambridge University Press, Cambridge.

[] Neininger, R. (2001) On a multivariate contraction method for random recursive structures with applications to quicksort. *Random Structures Algorithms* **19**, 498–524.

[] Neininger R. (2002), The Wiener index of random trees. *Combin. Probab. Comput.* **11**, 587-597.

[] Neininger, R. und Rüschendorf, L. (2002) Rates of convergence for quicksort. *J. Algorithms* **44**, 52-62.

[] Neininger, R. und Rüschendorf, L. (2003) Multivariate aspects of the contraction method. Preprint 03-19, Preprintserie des Mathematischen Instituts, *Universität Freiburg.*

[] Neininger, R. und Rüschendorf, L. (2004a) A general limit theorem for recursive algorithms and combinatorial structures. *Ann. Appl. Probab. 14*, 378–418.

[] Neininger, R. und Rüschendorf, L. (2004b) On the contraction method with degenerate limit equation. *Ann. Probab. 32*, 2838-2856.

[] Neininger, R. und Rüschendorf, L. (2005) Analysis of algorithms by the contraction method: additive and max-recursive sequences. *Interacting Stochastic Systems*, 435–450, *Springer, Berlin.*

[] Panholzer, A. und Prodinger, H. (2004) Spanning tree size in random binary search trees. *Ann. Appl. Probab.* **14**, 718–733.

[] Papadakis, T., Munro, J. I. und Poblete, P. (1990) Analysis of the expected search cost in skip lists. *SWAT 90 (Bergen, 1990)*, 160–172, Lecture Notes in Comput. Sci., 447, *Springer, Berlin.*

[] Pittel, B. (1999) Normal convergence problem? Two moments and a recurrence may be the clues. *Ann. Appl. Probab.* **9**, 1260–1302 (1999).

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

[] Pugh, W. (1989) Skip lists: a probabilistic alternative to balanced trees. *Algorithms and data structures (Ottawa, ON, 1989)*, 437–449, Lecture Notes in Comput. Sci., 382, *Springer, Berlin.*

[] Rachev, S. T. (1991) *Probability Metrics and the Stability of Stochastic Models*, John Wiley, New York.

[] Rachev, S. T. und Rüschendorf, L. (1995) Probability metrics and recursive algorithms. *Adv. Appl. Probab.* **27**, 770–799.

[] Rachev, S. T. und Rüschendorf, L. (1998). *Mass Transportation Problems. Vol. I: Theory.* Springer-Verlag, New York.

[] Régnier, M. (1989) A limiting distribution for quicksort. *RAIRO Inform. Théor. Appl.* **23**, 335–343.

[] Rösler, U. (1991) A limit theorem for "Quicksort," *RAIRO Inform. Théor. Appl.* **25**, 85–100.

[] Rösler, U. (1992) A fixed point theorem for distributions. *Stochastic Process. Appl.* **42**, 195–214.

[] Rösler, U. (2001) The analysis of stochastic divide and conquer algorithms. *Algorithmica* **29**, 238–261.

[] Rösler, U. (2004) Quickselect revisited. *Journal of the Iranian Statistical Society* **3**, 273–300.

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

[] Saks, M. and Wigderson, A. (1986) Probabilistic boolean decision trees and the complexity of evaluating game trees. *Proceedings of the 27th Annual IEEE Symposium on Foundations of Computer Science*, 29–38, Toronto, Ontario.

[] Schachinger, W. (2001). Asymptotic normality of recursive algorithms via martingale difference arrays. *Discrete Math. Theor. Comput. Sci.* **4**, 363–398.

[] Sedgewick, R. (1980) *Quicksort.* Gurland Publishing, NY.

[] Slack, R. S. (1968) A branching process with mean one and possibly infinite variance, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **9**, 139–145.

[] Smythe, R. T. und Mahmoud, H. M. (1995) A survey of recursive trees. *Theory Probab. Math. Statist.* **51**, 1–27.

[] Snir, M. (1985) Lower bounds on probabilistic linear decision trees. *Theoret. Comput. Sci. 38*, 69–82.

[] Szpankowski, W. (2001) *Average case analysis of algorithms on sequences.* With a foreword by Philippe Flajolet. Wiley-Interscience Series in Discrete Mathematics and Optimization, *Wiley-Interscience, New York.*

[] Trinajstić, N. (1992) *Chemical Graph Theory*, CRC Press, Boca Raton, FL.

[] Yaglom, A. M. (1947) Certain limit theorems of the theory of branching random processes, *Doklady Akad. Nauk SSSR (N.S.)* **56**, 795–798.

[] Zolotarev, V. M. (1976). Approximation of the distributions of sums of independent random variables with values in infinite-dimensional spaces. *Theor. Probability Appl. 21*, 721–737.

[] Zolotarev, V. M. (1977) Ideal metrics in the problem of approximating distributions of sums of independent random variables, *Theory Probab. Appl.* **22**, 433–449.